

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Programa de Pós Graduação em Estatística
Departamento de Estatística

Uma Revisão do Fator de Bayes com Aplicação à Modelos com
Misturas

Érica Cristina Marins Missão

Orientador: **Prof. Dr. Luis A. Milan**

Dissertação apresentada à CPG do
Departamento de Estatística da
Universidade Federal de São Carlos
como parte dos requisitos necessários
para obtenção do título de Mestre em
Estatística

Projeto financiado pelo CNPq.

São Carlos
Janeiro - 2007

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

M678rf

Missão, Érica Cristina Marins.

Uma revisão do fator de Bayes com aplicação à modelos com misturas / Érica Cristina Marins Missão. -- São Carlos : UFSCar, 2007.

99 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2004.

1. Teoria bayesiana de decisão estatística. 2. Fator de Bayes. 3. Método de Monte Carlo. 4. Gibbs sampling. 5. Modelos com misturas. I. Título.

CDD: 519.542 (20^a)

*“Dedico este trabalho
aos meus pais, João e Dairce,
pois são os melhores pais do mundo.”*

Agradecimentos

Em primeiro lugar agradeço à Deus pois foi a sua presença e seu amor que me deram forças para chegar até aqui.

Agradeço muito aos meus pais por acreditarem e confiarem em mim por todo o tempo de mestrado e também pelo amor e carinho tão presentes por toda a minha vida.

Agradeço muito ao meu orientador Prof. Dr. Luis A. Milan pela paciência, confiança e dedicação contínua em meu trabalho, pela sua grande amizade, e pela atenção dispensada sempre que precisei.

Agradeço também aos professores Josemar e Cristina pelas críticas e sugestões sugeridas neste trabalho.

Agradeço aos professores do departamento de Estatística, Carlos, Neto e Galvão por fazerem parte da minha formação como mestre.

Agradeço especialmente à minha “melhor amiga” Gislaine, a quem admiro muito, pela sua grande amizade, seu ombro amigo, todo seu carinho, compreensão e toda ajuda de que precisei.

Agradeço muito à minha amiga-irmã Silvia a quem tanto me ajudou, pela sua amizade tão presente desde o primeiro instante e também pelo seu carinho tão sincero.

Agradeço ao meu amigo-irmão Juliano a quem também me ajudou muito, pela sua amizade sincera e paciência.

Agradeço à minha grande amiga Karol, pela sua amizade sincera, sua alegria e todo seu carinho comigo.

Agradeço também aos meus avós, pois mesmo na minha ausência torceram por mim a todo momento.

Agradeço às minhas mais novas amigas Graciela e Leiziane pela amizade verdadeira, companheirismo, carinho e pela mão amiga quando precisei, espero retribuir tudo que fizeram e ainda fazem por mim.

Agradeço aos meus eternos amigos Sandra e Luciano pela amizade sincera, companheirismo, carinho e por todos os muitos momentos de alegria que passamos juntos até hoje.

Agradeço muito à todos meus amigos de mestrado por estarem presentes nesta fase da minha vida.

Agradeço aos meus amigos da graduação em Estatística, Julliana (Varejeira), Tio Rô e Tibolão pela amizade e alegria.

Agradeço aos meus grandes e eternos amigos de Rio Preto, Carlão, Kleber, Hassan, Cláudia, Karen, Marquinhos e Altino, pela amizade verdadeira e sincera durante todos esses anos, pelo companheirismo e pelo carinho que sempre tiveram comigo.

Agradeço também aos meus amigos da grande república R&J, especialmente ao Sérgio e ao Tato, pela amizade, carinho e atenção durante esta fase da minha vida.

Agradeço as minhas companheiras de república, Cristiane, Tatiany, Simone, Prescila, Flávia, Beth e Juliana pela paciência, compreensão e amizade.

Agradeço à todos os funcionários do departamento de Estatística, especialmente ao Robson, pois de alguma forma colaboraram para que este trabalho se realizasse.

Agradeço à CNPq pelo apoio financeiro.

“...e nossa história não estará pelo avesso assim sem final feliz,
teremos coisas bonitas pra contar...
e até lá vamos viver, temos muito ainda por fazer,
não olhe pra trás, apenas começamos,
o mundo começa agora, apenas começamos...”
Legião Urbana.

Resumo

O fator de Bayes é uma ferramenta utilizada na seleção de modelos. Neste trabalho fazemos uma revisão abrangente de diversos aspectos do fator de Bayes. Também apresentamos as soluções disponíveis atualmente para os problemas relacionadas à distribuição *a priori* imprópria como o fator de Bayes intrínseco e o fator de Bayes fracional. São apresentados resultados de simulações com o fator de Bayes sendo utilizado na seleção de modelos e uma aplicação a um conjunto de dados reais. Nestas simulações e na aplicação utilizamos o fator de Bayes na seleção de modelos envolvendo misturas. Em ambos os casos utilizamos o fator de Bayes e o fator de Bayes fracional.

Abstract

Bayes factor is a statistical tool used in the model selection. In this work we present a review of several aspects of the Bayes factor. We also present the available solutions for problems related to improper prior distribution as intrinsic Bayes factor and fractional Bayes factor. We develop a simulation study where we verify the performance the Bayes factor for selection of models using mixture. We consider the usual Bayes factor and the fractional Bayes factor. At the end we present a application the Bayes factor to a practical problem.

Índice

1	Introdução	1
2	Fator de Bayes	3
2.1	Definição do fator de Bayes	3
2.2	Interpretação do fator de Bayes	5
2.2.1	Escore Logaritmico	6
2.2.2	Validação Cruzada em densidades preditivas	7
2.3	Aproximação do fator de Bayes	11
2.3.1	Aproximação Assintótica	11
2.3.2	Método de Monte Carlo	15
2.3.3	Método MCMC	16
2.3.4	Comparação dos métodos de aproximação	17
2.4	A escolha da densidade <i>a priori</i>	17
2.5	A incerteza dos modelos	20
2.5.1	Janela de Occam	22
2.5.2	MCMC (Composição de modelo)	23
3	Fator de Bayes Intrínseco	25
3.1	Definição de Fator de Bayes Intrínseco	25
3.2	Fator de Bayes Intrínseco aritmético	27
3.3	Fator de Bayes Intrínseco geométrico	28
4	Fator de Bayes Fracional	29
4.1	Definição de Fator de Bayes Fracional	29
5	Mistura de Distribuições Binomiais	32
5.1	Algoritmo <i>Gibbs Sampling</i>	34
5.1.1	Algoritmo <i>Gibbs Sampling</i> com amostra ampliada	35
5.1.2	Algoritmo <i>Gibbs Sampling</i> para mistura de distribuições binomiais	35
6	Aplicação	40
6.1	Simulação com dados gerados	40
6.1.1	Descrição dos modelos	41
6.1.2	Abordagem Bayesiana	41
6.1.3	Seleção de modelos	42
6.1.4	Conclusão	45
6.2	Aplicação: Jogos de basquete	45
6.2.1	Descrição dos dados	46

6.2.2	Discussão	46
7	Simulação: Modelos com Mistura	47
7.1	Descrição dos dados	47
7.2	Descrição dos modelos	48
7.3	Abordagem Bayesiana	49
7.3.1	Densidades <i>a priori</i>	49
7.3.2	Densidade <i>a posteriori</i>	49
7.3.3	Densidades condicionais	49
7.4	Estudo de simulação	50
7.4.1	1º conjunto de dados	50
7.4.2	2º conjunto de dados	59
7.4.3	3º conjunto de dados	67
7.4.4	Seleção de Modelos	74
7.5	Discussão	78
8	Aplicação: Modelos com Mistura	79
8.0.1	Descrição dos modelos	79
8.0.2	Abordagem Bayesiana	80
8.0.3	Seleção de Modelos	80
8.0.4	Discussão	93
9	Discussão	94
	Referências Bibliográficas	96

Capítulo 1

Introdução

A seleção de modelos é uma importante ferramenta da inferência estatística. Através desta ferramenta podemos selecionar dentre um conjunto de modelos candidatos aquele que melhor se adequa aos dados em análise. Na abordagem Bayesiana, o fator de Bayes é uma ferramenta utilizada para esta finalidade. É possível utilizar o fator de Bayes na escolha de dois ou mais modelos, podendo ser utilizado para vários tipos de modelos, inclusive em modelos não encaixados.

O fator de Bayes pode ser utilizado para verificar presença de observações estranhas, ou *outliers*. Para predizer cada observação da amostra, Gelfand (1995) sugere o uso da validação cruzada, que consiste em retirar uma observação da amostra e fazer a predição desta através das demais observações. Assim, é possível verificar através do fator de Bayes se uma observação se adequa ao modelo considerado. Este método permite avaliar a possibilidade de excluir uma observação, ou de um pequeno conjunto de observações.

É possível também selecionar um modelo através da densidade preditiva dos dados aplicando o escore logaritmico no fator de Bayes, (Kass e Raftery, 1995). Este método consiste em aplicar logaritmo nas probabilidades marginais de cada observação y_i dadas as demais observações, resultando em um escore preditivo, ou seja, uma medida de predição dos dados.

O fator de Bayes pode ser aplicado em casos discretos ou contínuos, com baixa ou alta dimensionalidade. Nos casos em que o parâmetro é n -dimensional, é possível a aproximação do fator de Bayes através de métodos computacionais e assintóticos, vistos na seção 2.3.

O fator de Bayes é influenciado pelas distribuições *a priori* quando estas são informativas, especificamente se as distribuições *a priori* com relação aos parâmetros de um dos modelos em questão são mais informativas que as distribuições *a priori* dos parâmetros do outro modelo. A sugestão de alguns autores para resolver este problema é o uso de distribuições *a priori* não informativas para os parâmetros.

Outro problema referente ao uso do fator de Bayes ocorre quando se utiliza distribuições *a priori* impróprias nos parâmetros dos modelos, pois o resultado do cálculo do fator de Bayes pode ser indeterminado. Berger e Pericchi (1996), sugerem usar partes dos dados, como uma amostra de treinamento, para produzir uma distribuição *a priori* própria e o restante das observações para comparação de modelos propriamente dita. Este método é chamado de fator de Bayes intrínseco. O'Hagan (1995) também propõe uma solução parecida para este problema através do fator de Bayes fracional, que consiste em dividir a amostra completa também em duas partes.

Madigan e Raftery (1994), com o objetivo de reduzir o número de modelos considerados, sugerem o uso da Janela de Occam (*Occam's Window*), que consiste em eliminar do universo de seleção os modelos com menos chance de serem selecionados. O método MCMC (Composição de

modelo) desenvolvido por Madigan e York (1992), também pode ser utilizado com este propósito. Este método consiste em construir uma cadeia de Markov irredutível.

O objetivo da pesquisa apresentada nesta dissertação é fazer uma revisão abrangente de diversos aspectos do fator de Bayes. São tratados aspectos como a sensibilidade à escolha da distribuição *a priori*, se informativa ou não, se imprópria ou não, e seus efeitos na seleção de modelos. São tratados também os aspectos computacionais do fator de Bayes, tais como a aproximação de Laplace, a solução via MCMC e a janela de Occam. Também são apresentadas as soluções disponíveis atualmente para os problemas relacionados à distribuição *a priori* imprópria como o fator de Bayes intrínseco e o fator de Bayes fracional. Finalmente fazemos uma aplicação do fator de Bayes a um problema de mistura de distribuições binomiais.

A dissertação está organizada da seguinte forma.

No capítulo 2, são apresentadas a descrição e definição do fator de Bayes e métodos de aproximação via Laplace e Monte Carlo. São introduzidos também as tabelas de Jeffreys (1961) e de Kass e Raftery (1995). Abordamos no capítulo 3 o fator de Bayes intrínseco, desenvolvido por Berger e Pericchi (1996) para o problema da escolha da distribuição *a priori*. O fator de Bayes fracional é abordado no capítulo 4. No capítulo 5, descrevemos o modelo de mistura de distribuições binomiais e o algoritmo *Gibbs Sampling* com amostra ampliada para k componentes. As aplicações do modelo binomial sem mistura são apresentadas no capítulo 6, sendo considerados dados simulados e dados reais. No capítulo 7 apresentamos simulações para o modelo com misturas de binomiais. No capítulo 8 apresentamos uma aplicação a dados reais. O fator de Bayes e o fator de Bayes fracional também são utilizados para a seleção de modelos. E no capítulo 9 apresentamos a discussão final.

Capítulo 2

Fator de Bayes

O fator de Bayes é um instrumento usado na seleção de modelos, sejam modelos discretos ou contínuos, encaixados ou não encaixados, simples ou complexos.

Quando as densidades são definidas no espaço paramétrico n -dimensional, o fator de Bayes pode ser calculado através de métodos computacionais, como o método de Monte Carlo ou utilizando métodos numéricos viáveis para cada situação diferente, como o método de aproximação de Laplace.

O fator de Bayes está relacionado com o teste da razão de verossimilhança, onde o parâmetro é maximizado, ao invés de integrado. Jeffreys em 1961, introduziu uma tabela de calibragem para a interpretação do fator de Bayes, e a partir daí surgiram novas tabelas, com pequenas diferenças e adaptadas para várias situações. Existem outras interpretações para o fator de Bayes, como a validação cruzada e o escore logarítmico, fornecendo uma medida de predição dos dados e a adequabilidade do modelo em relação aos dados.

2.1 Definição do fator de Bayes

Sejam duas hipóteses, H_0 sob o modelo M_0 , conhecida como hipótese nula e H_1 sob o modelo M_1 , conhecida como hipótese alternativa, e uma sequência de observações, $Y = (y_1, y_2, \dots, y_n)$, com $n \geq 1$, tal que estes dados são provenientes de um dos dois modelos.

As densidades de probabilidades $pr(Y|H_0)$ e $pr(Y|H_1)$ são as distribuições preditivas sob o modelo M_0 e o modelo M_1 , respectivamente. Então

H_0 : hipótese sob o modelo M_0 , e

H_1 : hipótese sob o modelo M_1 .

Inicialmente, existem duas probabilidades *a priori* para os dois modelos considerados, $pr(H_0)$ e $pr(H_1)$.

A partir dos dados existentes, é obtida a probabilidade *a posteriori* sob os modelos das hipóteses H_0 e H_1 respectivamente, $pr(H_0|Y)$ e $pr(H_1|Y)$, ou seja, qualquer opinião inicial é transformada em uma probabilidade *a posteriori*, considerando os dados.

Usando o teorema de Bayes para o caso discreto, segue que

$$pr(H_k|Y) = \frac{pr(Y|H_k)pr(H_k)}{pr(Y|H_0)pr(H_0) + pr(Y|H_1)pr(H_1)}, \quad (2.1)$$

para $k = 0, 1$.

Inicialmente, não se sabe qual dos dois modelos considerados explica melhor os dados, e para a seleção de um deles é usado o fator de Bayes, que consiste na densidade preditiva sob o modelo M_1 , que corresponde à hipótese H_1 em relação à densidade preditiva sob o modelo M_0 que corresponde à hipótese H_0 , ou simplesmente H_1 versus H_0 .

Através do cálculo do fator de Bayes é possível verificar qual modelo se ajusta melhor aos dados. O fator de Bayes é definido como

$$B_{10} = \frac{pr(Y|H_1)}{pr(Y|H_0)}, \quad (2.2)$$

podendo também ser escrito como

$$B_{10} = \frac{pr(H_1|Y)}{pr(H_0|Y)} \times \frac{pr(H_0)}{pr(H_1)}, \quad (2.3)$$

que nada mais é do que a *odds a posteriori* de H_1 em relação à H_0 vezes a *odds a priori* de H_0 em relação à H_1 . Note que a *odds* de um evento qualquer, ϖ , é definida como

$$odds(\varpi) = \frac{pr(\varpi)}{1 - pr(\varpi)}.$$

Quando as hipóteses H_0 e H_1 são igualmente prováveis *a priori*, a distribuição *a priori* é

$$pr(H_0) = 0,5 \text{ e } pr(H_1) = 0,5.$$

Em casos discretos, o fator de Bayes é calculado de forma simples, mesmo quando há um ou mais parâmetros desconhecidos. E em casos contínuos, na maioria das vezes, o cálculo é feito através de integrações numéricas.

As densidades $pr(Y|H_k)$ para $k = 0, 1$, são obtidas integrando, sob o espaço paramétrico θ_k (parâmetro desconhecido), conforme a equação

$$pr(Y|H_k) = \int pr(Y|\theta_k, H_k)\pi(\theta_k|H_k)d\theta_k = E(Y|\theta_k, H_k), \quad (2.4)$$

ou seja, $pr(Y|H_k)$ é a esperança da função de verossimilhança dado o modelo H_k . A quantidade $\pi(\theta_k|H_k)$ é a função de densidade *a priori* para θ_k sob H_k e $pr(Y|\theta_k, H_k)$ é a função de densidade de probabilidade dos dados, dado o valor de θ_k .

A quantidade $pr(Y|H_k)$ pode ser chamada de densidade marginal dos dados, pois é obtida integrando a densidade conjunta de (Y, θ_k) , com relação a θ_k , ou de densidade preditiva dos dados, isto é, a probabilidade de conhecer os dados antes de qualquer observação disponível.

O fator de Bayes, B_{10} , está relacionado com o teste da razão de verossimilhança, com a diferença de que no cálculo do fator de Bayes o parâmetro de interesse é integrado levando à esperança enquanto que o teste da razão de verossimilhança trabalha com a moda, ou seja, trabalhamos com o máximo da função de verossimilhança, (Kass e Raftery, 1995).

No caso de muitas hipóteses envolvidas para o cálculo do fator de Bayes, este pode ser escrito como B_{ij} , que corresponde a hipótese i contra a hipótese j .

2.2 Interpretação do fator de Bayes

O fator de Bayes é uma medida da evidência provida pelos dados a favor de um modelo estatístico. Em 1961, Jeffreys sugeriu interpretar o fator de Bayes através de uma calibragem, ou seja, sugeriu dividir os possíveis valores do cálculo do fator de Bayes em quatro intervalos. A Tabela 1, mostra a divisão destes intervalos.

Tabela 1: Calibragem do fator de Bayes

$\log_{10}(B_{10})$	B_{10}	Evidência a favor de H_1
0 – 0,5	1 – 3,2	Insignificante
0,5 – 1	3,2 – 10	Significativa
1 – 2	10 – 100	Forte
> 2	> 100	Decisiva

Jeffreys (1961)

O logaritmo do fator de Bayes é aplicado a fim de obter números com intervalos menores.

No primeiro intervalo, (1 a 3,2), a evidência a favor da hipótese H_1 é mínima causando dúvidas em relação a escolha do modelo. No segundo intervalo, (3,2 a 10), a evidência a favor da hipótese H_1 aumenta favorecendo a escolha do modelo M_1 . No terceiro intervalo, (10 a 100), a escolha do modelo M_1 pode ser feita com mais precisão pois há uma forte evidência a favor de H_1 . E no quarto intervalo, (> 100), a escolha do modelo M_1 deve ser feita.

Kass e Raftery (1995) propõe a Tabela 2, considerando duas vezes o logaritmo do fator de Bayes. Desta forma, o valor obtido fica na mesma escala do teste da razão de verossimilhança.

Tabela 2: Intervalos construídos a partir da Calibragem de Jeffreys

$2\log_e(B_{10})$	B_{10}	Evidência a favor de H_1
0 – 2	1 – 3	Insignificante
2 – 6	3 – 20	Significativa
6 – 10	20 – 150	Forte
> 10	> 150	Muito forte

Kass e Raftery (1995)

Na Tabela 2, os valores do fator de Bayes também são divididos em 4 intervalos, mas, arredondando e usando 20 em lugar de 10.

No primeiro intervalo, (1 a 3), a evidência a favor da hipótese H_1 é mínima causando dúvidas em relação a escolha do modelo. No segundo intervalo, (3 a 20), a evidência a favor da hipótese H_1 aumenta favorecendo a escolha do modelo M_1 . No terceiro intervalo, (20 a 150), a escolha do modelo M_1 pode ser feita com mais precisão pois há uma forte evidência a favor de H_1 . E no quarto intervalo, (> 150), a escolha do modelo M_1 deve ser feita.

2.2.1 Escore Logaritmico

Nesta seção, é introduzido um método preditivo chamado de escore logaritmico, que avalia o modelo que melhor se adequa aos dados. Este método preditivo consiste em aplicar logaritmo nas probabilidades marginais de cada observação y_i dada as demais observações, resultando em um escore preditivo, ou seja, uma medida de predição dos dados.

A vantagem do escore logaritmico é que não há necessidade de identificar um dos modelos como verdadeiro na interpretação do fator de Bayes.

Suponha $Y = \{y_1, y_2, \dots, y_n\}$, onde para cada y_i é formada uma distribuição preditiva $\hat{pr}_i(\cdot)$ de y_i dado os dados já disponíveis $\{y_1, y_2, \dots, y_{i-1}\}$. A regra de escore logaritmico, $\log \hat{pr}_i(y_i)$, (Good, 1952), avalia o ajuste do modelo aos dados.

Aplicando logaritmo na distribuição preditiva de cada y_i , o escore logaritmico é a soma de todos os $\hat{pr}_i(y_i)$, para $i = 1, \dots, n$.

Assim, o escore global, LS , para qualquer regra que gere tais distribuições preditivas é

$$LS = \sum_{i=1}^n \log \hat{pr}_i(y_i).$$

Em particular, se a regra de predição é derivada do modelo H_k , então

$$\begin{aligned} LS_k &= \log pr(Y|H_k) = \log (pr(y_1, y_2, \dots, y_n|H_k)) \\ &= \log [pr(y_1|H_k) pr(y_2|y_1, H_k) pr(y_3|y_1, y_2, H_k) \dots pr(y_n|y_1, y_2, \dots, y_{n-1}, H_k)] \\ &= \log pr(y_1|H_k) + \log pr(y_2|y_1, H_k) + \log pr(y_3|y_1, y_2, H_k) \dots + \log pr(y_n|y_1, y_2, \dots, y_{n-1}, H_k) \\ &= \sum_{i=1}^n \log pr(y_i|y_1, y_2, \dots, y_{i-1}, H_k). \end{aligned} \quad (2.5)$$

O logarítmo do fator de Bayes é dado pela diferença entre os escores preditivos, isto é

$$\log B_{10} = \log \left(\frac{pr(Y|H_1)}{pr(Y|H_0)} \right).$$

Usando (2.5), temos

$$\begin{aligned} &= \frac{\sum_{i=1}^n \log pr(y_i|y_{i-1}, \dots, y_1, H_1)}{\sum_{i=1}^n \log pr(y_i|y_{i-1}, \dots, y_1, H_0)} \\ &= LS_1 - LS_0, \end{aligned}$$

como em Kass e Raftery (1995).

Deste modo, o fator de Bayes é a diferença entre os escores preditivos do modelo H_1 em relação ao modelo H_0 .

Observe que $\log B_{10}$ é a diferença dos escores logaritmicos de H_1 e H_0 , respectivamente. Se $\log B_{10} > 0$, significa que o modelo que mais se adequa aos dados é H_1 , e se $\log B_{10} < 0$, significa H_0 é o modelo que mais se adequa aos dados.

2.2.2 Validação Cruzada em densidades preditivas

Supondo que Y é um conjunto de variáveis observáveis, $\{y_1, y_2, \dots, y_n\}$, a validação cruzada na densidade preditiva é o conjunto $\{pr(y_r|Y_{(r)}); r = 1, 2, \dots, n\}$, onde $Y_{(r)}$ denota todos os elementos de Y menos o y_r , o r -ésimo observável, isto é, o valor de cada y_r é predito através do conjunto de todos os elementos de Y , exceto o próprio y_r .

O valor observado, $y_{r,obs}$, pode então ser comparado com a densidade preditiva para analisar se esta observação é provável sob um determinado modelo. Isto é feito para verificar se as observações se adequam a este modelo, (Gelfand *et al*, 1992).

A densidade preditiva da r -ésima observação é dada por

$$pr(y_r|Y_{(r)}) = \int pr(y_r|\theta, Y_{(r)})pr(\theta|Y_{(r)})d\theta, \quad (2.6)$$

ou através da equação

$$pr(y_r|Y_{(r)}) = \frac{pr(y_r, Y_{(r)})}{pr(Y_{(r)})} = \frac{pr(Y)}{pr(Y_{(r)})}.$$

No caso de observações condicionalmente independentes dado θ , $pr(y_r|\theta, Y_{(r)}) = pr(y_r|\theta)$, pois,

$$pr(y_r|\theta, Y_{(r)}) = \frac{pr(y_r, Y_{(r)}|\theta)}{pr(Y_{(r)}|\theta)} = \frac{pr(y_r|\theta)pr(Y_{(r)}|\theta)}{pr(Y_{(r)}|\theta)} = pr(y_r|\theta). \quad (2.7)$$

As densidades $\{pr(y_r|Y_{(r)})\}$, $r = 1, \dots, n$, permitem excluir uma observação, (Gelfand, 1995), ou excluir um pequeno conjunto de observações, dois ou três elementos, (Peña e Tiao, 1992).

Selecionando um subconjunto do conjunto de observações, Y , podemos utilizar $pr(y_r|Y_{(r)})$ para o ajuste do modelo.

O produto obtido através da validação cruzada em densidades preditivas é

$$\prod_{r=1}^n pr(y_r|Y_{(r),obs}).$$

Este produto tem sido proposto como um substituto para $pr(Y)$, dando origem ao pseudo-fator de Bayes, (Geisser e Eddy, 1979 e Gelfand, 1995), que é dado por

$$\prod_{r=1}^n \frac{pr(y_{r,obs}|Y_{(r),obs}, H_1)}{pr(y_{r,obs}|Y_{(r),obs}, H_0)}. \quad (2.8)$$

A densidade preditiva *a posteriori*, $pr(Y|Y_{obs})$, é a densidade preditiva de um novo conjunto independente de variáveis observáveis do modelo, dados os valores observados, (Rubin, 1984 e Aitkin, 1991). A distribuição preditiva *a posteriori* é dada por

$$pr(Y|Y_{obs}) = \int pr(Y|\theta)pr(\theta|Y_{obs})d\theta, \quad (2.9)$$

onde $pr(Y|\theta) = pr(Y|\theta, Y_{obs})$.

A densidade $pr(Y) = \int pr(Y|\theta)pr(\theta)d\theta$ é similar a (2.9). A diferença entre estas duas equações é que (2.9) determina a média da densidade condicional dos observáveis, enquanto que $pr(Y)$ informa sobre a média dos não observáveis, (Gelfand, 1995).

A densidade preditiva *a posteriori* oferece uma medida da adequabilidade do modelo. Se as observações, $Y = Y_{obs}$, estão em uma região de probabilidade baixa da densidade $pr(Y|Y_{obs})$, então, provavelmente o modelo não se ajusta aos dados.

Através da distribuição preditiva *a posteriori*, dada pela equação (2.9), é obtida a densidade marginal de y_r ,

$$pr(y_r|Y_{obs}) = \int pr(y_r|\theta)pr(\theta|Y_{obs})d\theta. \quad (2.10)$$

Note que para observações condicionalmente independentes, as equações (2.6) e (2.10) são similares, a menos que, $y_{r,obs}$, seja muito influente. A comparação entre $pr(y_{r,obs}|Y_{obs})$ e $pr(y_{r,obs}|Y_{(r),obs})$ mostra a influência de $y_{r,obs}$ na modelagem.

Para verificar se a validação cruzada fornece realmente o modelo adequado para as observações, são usadas as funções de verificação. Suponha que uma densidade preditiva univariada é selecionada. Para a variável observável, y_r , é possível comparar esta densidade com o valor observado, $y_{r,obs}$, através de uma função de verificação, $f(y_r; y_{r,obs})$, tomando a esperança de

$f(y_r; y_{r,obs})$ em relação a densidade preditiva, $f_{y_r}(y_r) = pr(y_r|Y_{(r),obs})$. Apresentamos a seguir alguns exemplos de função de verificação.

i) Supondo $f_1(y_r; y_{r,obs}) = y_{r,obs} - y_r$, um resíduo, onde $y_{r,obs}$ é o valor observado e y_r é o observável. Assim, aplicando a esperança em $f_1(y_r; y_{r,obs})$, segue que

$$d_{1r} = E(y_{r,obs} - y_r|Y_{(r),obs}) = y_{r,obs} - E(y_r|Y_{(r),obs}), \quad (2.11)$$

é a diferença entre o valor observado e a média dos observáveis sob a distribuição preditiva. Quanto menor for o valor resultante de d_{1r} , melhor será a adequabilidade ao modelo.

Em muitas aplicações y_r é modelado como $y_r = h_r(\theta) + \varepsilon_r$ onde, $h_r(\theta)$ é uma função qualquer do parâmetro e ε_r é algum tipo de erro aleatório com média zero. Considerando a média *a posteriori* do resíduo, segue que

$$d_{1r} = E(y_{r,obs} - h_r(\theta)|Y_{(r),obs}),$$

como em Chaloner e Brant (1988).

A forma padronizada para d_{1r} é

$$d_{1r}^* = \frac{y_{r,obs} - E(y_r|Y_{(r),obs})}{\sqrt{Var(y_r|Y_{(r),obs})}}.$$

ii) Supondo $f_2(y_r; y_{r,obs}) = I(-\infty < y_r \leq y_{r,obs})$, onde I denota a função indicadora, assumindo o valor 1 como verdadeiro e zero caso contrário. Aplicando a esperança em $f_2(y_r; y_{r,obs})$, segue que

$$\begin{aligned} d_{2r} &= E(I(-\infty < y_r \leq y_{r,obs})) \\ &= pr(y_r \leq y_{r,obs}|Y_{(r),obs}). \end{aligned}$$

Assim, d_{2r} mede o quão provável $y_{r,obs}$ é sob a suposição de que o modelo seja $pr(y_r|Y_{(r),obs})$. Valores próximos de zero ou de 1 indicam que a observação $y_{r,obs}$ está em uma região de baixa probabilidade.

Se y_r é discreto, podemos usar

$$pr(y_r = y_{r,obs}|Y_{(r),obs}).$$

iii) Supondo $f_3(y_r; y_{r,obs}) = I(y_r \in B_r)$, onde

$$B_r = \{y_r : pr(y_r|Y_{(r),obs}) \leq pr(y_{r,obs}|Y_{(r),obs})\},$$

e aplicando a esperança em $f_3(y_r; y_{r,obs})$ obtemos

$$d_{3r} = E(I(y_r \in B_r)) = pr(B_r|Y_{(r),obs}).$$

B_r consiste de um conjunto de valores de y_r que apresenta função de densidade menor do que $y_{r,obs}$ sob a densidade preditiva $pr(y_r|Y_{(r),obs})$. Assim, quanto menor for o valor resultante da função d_{3r} , menor será a adequabilidade de $y_{r,obs}$ no modelo.

iv) Supondo $f_4(y_r; y_{r,obs}) = \frac{1}{2\varepsilon} I[y_{r,obs} - \varepsilon \leq y_r \leq y_{r,obs} + \varepsilon]$, uma função indicadora de um pequeno intervalo ao redor de $y_{r,obs}$, onde ε é um valor pequeno qualquer, $\varepsilon > 0$.

Tomando a esperança de $f_4(y_r; y_{r,obs})$ obtemos

$$d_{4r} = \frac{1}{2\varepsilon} pr(y_{r,obs} - \varepsilon \leq y_r \leq y_{r,obs} + \varepsilon).$$

Fazendo $\varepsilon \rightarrow 0$, segue que

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} d_{4r} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} pr(y_{r,obs} - \varepsilon \leq y_r \leq y_{r,obs} + \varepsilon) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} (pr(y_r \leq y_{r,obs} + \varepsilon) - pr(y_r \leq y_{r,obs} - \varepsilon)) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} (F_{y_r}(y_{r,obs} + \varepsilon) - F_{y_r}(y_{r,obs} - \varepsilon)). \end{aligned}$$

Derivando com relação a ε e aplicando a regra de L'Hôpital, temos que

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} d_{4r} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{2} (f_{y_r}(y_{r,obs} + \varepsilon) + f_{y_r}(y_{r,obs} - \varepsilon)) \\ &= \frac{1}{2} \left(\lim_{\varepsilon \rightarrow 0} f_{y_r}(y_{r,obs} + \varepsilon) + \lim_{\varepsilon \rightarrow 0} f_{y_r}(y_{r,obs} - \varepsilon) \right) \\ &= \frac{1}{2} (f_{y_r}(y_{r,obs}) + f_{y_r}(y_{r,obs})) \\ &= f_{y_r}(y_{r,obs}) = pr(y_{r,obs} | Y_{(r),obs}). \end{aligned}$$

Portanto,

$$d_{4r} = pr(y_{r,obs} | Y_{(r),obs}),$$

denominado como *conditional predictive ordinate (CPO)*, (Pettit e Young, 1990). Se d_{4r} tiver um valor pequeno, isto indica que a observação não se ajusta ao modelo.

Aplicando o logaritmo à *CPO* obtemos

$$\begin{aligned} \log(CPO) &= \log(pr(y_{r,obs} | Y_{(r),obs})) \\ &= \log \left(\frac{pr(y_{r,obs}, Y_{(r),obs})}{pr(Y_{(r),obs})} \right) \\ &= \log \left(\frac{pr(Y_{obs})}{pr(Y_{(r),obs})} \right) \\ &= \log pr(Y_{obs}) - \log pr(Y_{(r),obs}). \end{aligned}$$

Assim, o $\log(CPO)$ para a r -ésima observação contrasta a densidade preditiva para todas as observações com a densidade preditiva de todas as observações exceto a r -ésima observação. Esta função verifica a adequabilidade da r -ésima observação ao modelo.

Para o modelo escolhido, considere a razão da *CPO*

$$C_r = \frac{pr(y_{r,obs} | Y_{(r),obs}, M_1)}{pr(y_{r,obs} | Y_{(r),obs}, M_0)}.$$

Note que $\prod_{r=1}^n C_r$ está definido como pseudo-fator de Bayes em (2.8) e

$$\log C_r = \log B_{10} - \log B_{10(r)},$$

pois

$$\begin{aligned} \log B_{10} - \log B_{10(r)} &= \log \frac{pr(Y_{obs}|M_1)}{pr(Y_{obs}|M_0)} - \log \frac{pr(Y_{(r),obs}|M_1)}{pr(Y_{(r),obs}|M_0)} = \\ &= \log \frac{pr(y_{r,obs}|Y_{(r),obs}, M_1)pr(Y_{(r),obs}|M_1)}{pr(y_{r,obs}|Y_{(r),obs}, M_0)pr(Y_{(r),obs}|M_0)} - \log \frac{pr(Y_{(r),obs}|M_1)}{pr(Y_{(r),obs}|M_0)}. \end{aligned}$$

Cancelando os termos iguais, obtemos

$$\log B_{10} - \log B_{10(r)} = \log \frac{pr(y_{r,obs}|Y_{(r),obs}, M_1)}{pr(y_{r,obs}|Y_{(r),obs}, M_0)} = \log C_r,$$

onde $B_{10(r)}$ é o fator de Bayes do modelo 1 em relação ao modelo 0, usando todas as observações exceto a r -ésima observação.

2.3 Aproximação do fator de Bayes

Freqüentemente, as densidades contínuas com espaço paramétrico n -dimensional envolvidas no cálculo do fator de Bayes são difíceis para calcular analiticamente, e nestes casos são necessárias as aproximações. Para a introdução desta abordagem a equação utilizada será

$$pr(Y|H) = \int pr(Y|\theta, H)\pi(\theta|H)d\theta, \quad (2.12)$$

ou seja, a densidade marginal dos dados, considerando o modelo H .

A avaliação analítica dessa integral seria o método mais preciso e mais fácil computacionalmente, mas em geral isto somente é possível em situações simples como no caso das distribuições da família exponencial com distribuições *a priori* conjugadas, (DeGroot, 1970).

A seguir, são descritos alguns métodos para a aproximação do fator de Bayes .

2.3.1 Aproximação Assintótica

Nesta seção, para aproximar a densidade marginal dos dados, (2.12), é usado o método de Laplace. O método consiste em aproximar a densidade marginal pela distribuição Normal, através da expansão de Taylor. O método de Laplace é de interesse em casos onde existem amostras grandes, pois se trata de um método bem preciso quando é aplicado nestes casos, (Kass e Raftery, 1995).

Método de Laplace

Uma aproximação para a densidade marginal dos dados dada por (2.12), é obtida assumindo que a densidade *a posteriori*, que é proporcional a $pr(Y|\theta, H)\pi(\theta|H)$, é concentrada ao redor do seu máximo, $\hat{\theta}$. Isto normalmente é o caso da função de verossimilhança, $pr(Y|\theta, H)$, que no caso de amostras grandes é concentrada ao redor do seu máximo, $\hat{\theta}$.

Caso unidimensional Considere inicialmente o caso onde θ é unidimensional. A distribuição *a posteriori* é dada por

$$pr(\theta|Y, H) = \frac{pr(Y|\theta, H)\pi(\theta|H)}{\int pr(Y|\theta, H)\pi(\theta|H)d\theta} = \frac{e^{\log pr(Y|\theta, H) + \log \pi(\theta|H)}}{\int e^{\log pr(Y|\theta, H) + \log \pi(\theta|H)} d\theta}.$$

Tachibana (1995) utiliza $\sigma^2 = -\frac{1}{R(\hat{\theta})}$ de forma que substituindo $R(\hat{\theta})$ por $-\frac{1}{\sigma^2}$ na expansão de Taylor, σ^2 apareça no denominador facilitando a aproximação pela distribuição Normal. Seja $R(\theta) = \log(pr(Y|\theta, H)) + \log(\pi(\theta|H))$. Expandindo-se $R(\theta)$ em uma série de Taylor em torno de $\theta = \hat{\theta}$ e ignorando os termos superiores a 2ª ordem, temos que

$$R(\theta) \cong R(\hat{\theta}) + (\theta - \hat{\theta}) R'(\hat{\theta}) + \frac{(\theta - \hat{\theta})^2}{2} R''(\hat{\theta}).$$

A derivada da função $R(\theta)$ no ponto $\hat{\theta}$ é zero, pois $\hat{\theta}$ é o ponto de máximo. Substituindo $R''(\hat{\theta})$ por $-\frac{1}{\sigma^2}$, temos

$$R(\theta) \cong R(\hat{\theta}) - \frac{(\theta - \hat{\theta})^2}{2\sigma^2}.$$

Assim,

$$\int pr(Y|\theta, H)\pi(\theta|H)d\theta = \int e^{R(\theta)} d\theta \cong e^{R(\hat{\theta})} \int e^{-\frac{(\theta - \hat{\theta})^2}{2\sigma^2}} d\theta.$$

Aproximando esta integral por uma distribuição normal com parâmetros $(\hat{\theta}, \sigma^2)$, $N(\hat{\theta}, \sigma^2)$, obtemos

$$e^{R(\hat{\theta})} \int e^{-\frac{(\theta - \hat{\theta})^2}{2\sigma^2}} d\theta = \sqrt{2\pi\sigma} e^{R(\hat{\theta})} = \sqrt{2\pi\sigma} pr(Y|\hat{\theta}, H)\pi(\hat{\theta}|H).$$

Logo, a distribuição *a posteriori* aproximada é dada por

$$pr(\theta|Y, H) \cong \frac{pr(Y|\theta, H)\pi(\theta|H)}{\sqrt{2\pi\sigma} pr(Y|\hat{\theta}, H)\pi(\hat{\theta}|H)},$$

e

$$\widehat{p(Y|H)} = \sqrt{2\pi\sigma} pr(Y|\hat{\theta}, H)\pi(\hat{\theta}|H).$$

Caso d -dimensional Para um vetor $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \dots, \theta_d)$, com dimensão d , a expansão de Taylor até a segunda ordem é

$$R(\boldsymbol{\theta}) \cong R(\tilde{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T R'(\tilde{\boldsymbol{\theta}}) + \frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T [R''(\tilde{\boldsymbol{\theta}})] (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})}{2},$$

onde

$$R''(\tilde{\boldsymbol{\theta}}) = \left[\frac{\partial^2 R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}},$$

e $\boldsymbol{\theta}^T$ é o vetor transposto de $\boldsymbol{\theta}$.

O objetivo é aproximar $\int pr(Y|\boldsymbol{\theta}, H)\pi(\boldsymbol{\theta}|H)d\boldsymbol{\theta} = \int e^{\log(pr(Y|\boldsymbol{\theta}, H)) + \log(\pi(\boldsymbol{\theta}|H))} d\boldsymbol{\theta} = \int e^{R(\boldsymbol{\theta})} d\boldsymbol{\theta}$.

Como no caso unidimensional, a derivada da função $R(\boldsymbol{\theta})$ é zero no ponto de máximo. Substituindo $R''(\tilde{\boldsymbol{\theta}})$ pela matriz das segundas derivadas de $\log(pr(Y|\boldsymbol{\theta}, H)\pi(\boldsymbol{\theta}|H))$, obtemos

$$\int e^{R(\boldsymbol{\theta})} d\boldsymbol{\theta} \cong \int e^{R(\tilde{\boldsymbol{\theta}}) - \frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \Sigma^{-1} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})}{2}} d\boldsymbol{\theta} = e^{R(\tilde{\boldsymbol{\theta}})} \int e^{-\frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \Sigma^{-1} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})}{2}} d\boldsymbol{\theta},$$

onde

$$\Sigma^{-1} = \left[-\frac{\partial^2 R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}.$$

Como se trata de um vetor d -dimensional, a aproximação de $\int pr(Y|\boldsymbol{\theta}, H)\pi(\boldsymbol{\theta}|H)d\boldsymbol{\theta}$ é feita utilizando uma distribuição normal multivariada com parâmetros $(\tilde{\boldsymbol{\theta}}, \Sigma)$, $N(\tilde{\boldsymbol{\theta}}, \Sigma)$,

$$\begin{aligned} e^{R(\tilde{\boldsymbol{\theta}})} \int e^{-\frac{(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T \Sigma^{-1} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})}{2}} d\boldsymbol{\theta} &= e^{R(\tilde{\boldsymbol{\theta}})} (2\pi)^{d/2} |\Sigma|^{1/2} \\ &= pr(Y|\tilde{\boldsymbol{\theta}}, H)\pi(\tilde{\boldsymbol{\theta}}|H)(2\pi)^{d/2} |\Sigma|^{1/2}, \end{aligned}$$

onde $\tilde{\boldsymbol{\theta}}$ é a moda de $\boldsymbol{\theta}$, ou seja, o valor de $\boldsymbol{\theta}$ que maximiza $pr(Y|\boldsymbol{\theta}, H)\pi(\boldsymbol{\theta}|H)$, d é a dimensão de $\boldsymbol{\theta}$, Σ^{-1} é a matriz das segundas derivadas do $\log(pr(Y|\boldsymbol{\theta}, H)\pi(\boldsymbol{\theta}|H))$ avaliada em $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, e Σ é a matriz inversa de Σ^{-1} .

Assim, $p(Y|H) = \int pr(Y|\boldsymbol{\theta}, H)\pi(\boldsymbol{\theta}|H)d\boldsymbol{\theta}$ é aproximada por

$$\widehat{pr(Y|H)} = pr(Y|\tilde{\boldsymbol{\theta}}, H)\pi(\tilde{\boldsymbol{\theta}}|H)(2\pi)^{d/2} |\Sigma|^{1/2}. \quad (2.13)$$

Logo a distribuição *a posteriori* é aproximada por

$$pr(\boldsymbol{\theta}|Y, H) \cong \frac{pr(Y|\boldsymbol{\theta}, H)\pi(\boldsymbol{\theta}|H)}{pr(Y|\tilde{\boldsymbol{\theta}}, H)\pi(\tilde{\boldsymbol{\theta}}|H)2\pi^{d/2} |\Sigma|^{1/2}},$$

ou seja, a distribuição *a posteriori* é aproximada por uma distribuição normal d dimensional, $N_d(\tilde{\boldsymbol{\theta}}, \Sigma)$.

Sob algumas condições, Kass, Tierney e Kadane (1990), constatam que para $n \rightarrow \infty$, $pr(Y|H) = \widehat{pr(Y|H)}(1 + O(n^{-1}))$, isto é, o erro relativo é da ordem de $O(n^{-1})$. Deste modo, quando o método de Laplace é aplicado no numerador e no denominador do fator de Bayes, em (2.2), a aproximação resultante tem o erro relativo da ordem de $O(n^{-1})$.

Em geral, o método prevê aproximações adequadas para problemas de dimensionalidade moderada.

Variantes no método de Laplace

O método de Laplace pode ser aplicado em formas alternativas, como a variante a seguir, apresentada originalmente por Kass e Vaidyanathan (1992),

$$pr(\widehat{Y|H})_{emv} = (2\pi)^{d/2} \left| \widehat{\Sigma} \right|^{1/2} pr(Y|\widehat{\theta}, H)\pi(\widehat{\theta}|H), \quad (2.14)$$

onde $\widehat{\Sigma}^{-1}$ é a matriz de informação observada; isto é, a matriz negativa do logaritmo da verossimilhança avaliada no estimador de máxima verossimilhança, (EMV) , $\widehat{\theta}$. Note que (2.14) difere de (2.13) pelo fato de que $\tilde{\theta}$, que é a moda, é substituído por $\widehat{\theta}$, o valor estimado de θ (EMV). Esta aproximação também tem erro relativo da ordem de $O(n^{-1})$.

Segundo Kass e Raftery, é provável que (2.14) seja menos precisa do que (2.13) quando a distribuição *a priori* é informativa. Contudo apresenta uma vantagem, pode ser facilmente calculada por qualquer programa estatístico que calcule o EMV , a matriz de informação observada (ou a sua inversa) e o valor da verossimilhança maximizada.

O inverso da matriz de informação esperada pode ser usado em lugar de $\widehat{\Sigma}$ em (2.14). A aproximação resultante tem um erro relativo assintótico maior, da ordem de $O(n^{-1/2})$, mas esta aproximação permanece suficientemente precisa para aplicação em muitos problemas, Kass e Raftery (1995).

Agora suponha que existem hipóteses encaixadas, com parâmetros (β, ψ) e que a distribuição *a priori* sob H_1 é $\pi(\beta, \psi|H_1)$. A hipótese nula é dada por $H_0 : \psi = \psi_0$, onde a distribuição *a priori* sob H_0 é $\pi(\beta|H_0)$. Quando (2.14) é aplicado em B_{10} obtemos

$$B_{10} = \frac{(2\pi)^{d_1/2} \left| \widehat{\Sigma}_1 \right|^{1/2} pr(Y|(\widehat{\beta}, \widehat{\psi}), H_1)\pi(\widehat{\beta}, \widehat{\psi}|H_1)}{(2\pi)^{d_0/2} \left| \widehat{\Sigma}_0 \right|^{1/2} pr(Y|(\widehat{\beta}^*, H_0)\pi(\widehat{\beta}^*|H_0)},$$

onde d_k é a dimensão do vetor de parâmetros sob a hipótese H_k e $\widehat{\Sigma}_k$ é a matriz de covariâncias referente às hipóteses H_k , para $k = 0, 1$. O $\widehat{\beta}^*$ denota o EMV sob a hipótese H_0 e $(\widehat{\beta}, \widehat{\psi})$ denota o EMV sob a hipótese H_1 .

Aplicando duas vezes o logaritmo em B_{10} obtemos

$$\begin{aligned} 2 \log B_{10} &= 2 \left(\frac{d_1}{2} \log(2\pi) + \frac{1}{2} \log \left| \widehat{\Sigma}_1 \right| \right) \\ &+ 2 \left(\log pr(Y|(\widehat{\beta}, \widehat{\psi}), H_1) + \log \pi(\widehat{\beta}, \widehat{\psi}|H_1) \right) \\ &- 2 \left(\frac{d_0}{2} \log(2\pi) + \frac{1}{2} \log \left| \widehat{\Sigma}_0 \right| \right) \end{aligned}$$

$$\begin{aligned}
& -2 \left(\log pr(Y|\hat{\beta}^*, H_0) + \log \pi(\hat{\beta}^*|H_0) \right) \\
& = d_1 \log(2\pi) + \log \left| \hat{\Sigma}_1 \right| \\
& + 2 \left(\log pr(Y|(\hat{\beta}, \hat{\psi}), H_1) + \log \pi(\hat{\beta}, \hat{\psi}|H_1) \right) \\
& - d_0 \log(2\pi) - \log \left| \hat{\Sigma}_0 \right| \\
& - 2 \left(\log pr(Y|\hat{\beta}^*, H_0) + \log \pi(\hat{\beta}^*|H_0) \right).
\end{aligned}$$

Portanto,

$$\begin{aligned}
2 \log B_{10} \approx & \Lambda + \log \left| \hat{\Sigma}_1 \right| - \log \left| \hat{\Sigma}_0 \right| + \log \pi(\hat{\beta}, \hat{\psi}|H_1) \\
& - \log \pi(\hat{\beta}^*|H_0) + (d_1 - d_0) \log(2\pi),
\end{aligned} \tag{2.15}$$

onde $\Lambda = 2((\log pr(Y|(\hat{\beta}, \hat{\psi}), H_1) - \log pr(Y|\hat{\beta}^*, H_0))$, isto é, duas vezes o log da razão de verossimilhança, com $(d_1 - d_0)$ graus de liberdade.

Novamente a informação observada ou esperada pode ser usada no cálculo das matrizes de covariância $\hat{\Sigma}_k$ e (2.15) tem um erro relativo da ordem de $O(n^{-1})$ ou $O(n^{-1/2})$.

Jeffreys (1961), propôs uma aproximação para o caso onde $(d_1 - d_0) = 1$, mas com informação esperada ao invés de observada, e Chow (1981), estendeu o resultado de Jeffreys para dimensões mais altas.

2.3.2 Método de Monte Carlo

A integral $pr(Y|H) = \int pr(Y|\theta, H)\pi(\theta|H)d\theta$, em (2.12), pode ser reescrita como

$$pr(Y) = \int pr(Y|\theta)\pi(\theta)d\theta,$$

e pode ser aproximada através do método de Monte Carlo fazendo

$$\widehat{pr(Y)} = \frac{1}{m} \sum_{i=1}^m pr(Y|\theta^{(i)}),$$

onde $\{\theta^{(i)}, i = 1, \dots, m\}$ é uma amostra da distribuição *a priori*, $\pi(\theta)$.

Esta possibilidade foi investigada por Raftery e Banfield (1990) e mais detalhadamente por McCulloch e Rossi (1991).

Uma dificuldade importante nesta estimativa é que se a densidade *a priori* for pouco informativa, a maioria dos $\theta^{(i)}$ têm a verossimilhança pequena de forma que o processo de simulação

neste caso será altamente ineficiente. Deste modo, a estimativa será dominada por uma pequena verossimilhança, resultando em uma variância grande para $\widehat{pr(Y)}$, acarretando uma convergência lenta para o método. Estes problemas aparecem nos exemplos estudados por McCulloch e Rossi (1991).

A precisão do método de Monte Carlo pode ser melhorada pelo método *Importance Sampling*. Este método consiste em gerar amostras $\{\theta^{(i)} : i = 1, \dots, m\}$ de uma densidade $\pi^*(\theta)$ e corrigir o erro da aproximação ponderando através de pesos. Assim, $pr(Y)$ é aproximado por

$$\widehat{pr(Y)} = \frac{\sum_{i=1}^m w_i pr(Y|\theta^{(i)})}{\sum_{i=1}^m w_i}, \quad (2.16)$$

onde $w_i = \pi(\theta^{(i)})/\pi^*(\theta^{(i)})$, (Geweke, 1989).

2.3.3 Método MCMC

Existem vários métodos para a simulação da densidade *a posteriori*, dentre eles, os métodos *Markov Chain Monte Carlo* (MCMC). Dentre os métodos MCMC os mais utilizados são os algoritmos *Metropolis-Hastings* e *Gibbs Sampling*.

O uso de qualquer um destes métodos nos dá uma amostra da densidade *a posteriori*

$$g^*(\theta) = pr(\theta|Y) = pr(Y|\theta)\pi(\theta)/pr(Y).$$

Substituindo $\pi^*(\theta)$ por

$$\frac{pr(Y|\theta)\pi(\theta)}{pr(Y)}$$

na equação (2.16),

$$\frac{\sum_{i=1}^m w_i pr(Y|\theta^{(i)})}{\sum_{i=1}^m w_i}, \quad w_i = \frac{\pi(\theta^{(i)})}{\pi^*(\theta^{(i)})}$$

obtemos

$$\begin{aligned} \widehat{pr(Y)} &= \frac{\sum_{i=1}^m \frac{\pi(\theta^{(i)})}{\pi^*(\theta^{(i)})} pr(Y|\theta^{(i)})}{\sum_{i=1}^m \frac{\pi(\theta^{(i)})}{\pi^*(\theta^{(i)})}} = \frac{\sum_{i=1}^m \frac{\pi(\theta^{(i)})pr(Y)}{pr(Y|\theta^{(i)})\pi(\theta^{(i)})} pr(Y|\theta^{(i)})}{\sum_{i=1}^m \frac{\pi(\theta^{(i)})pr(Y)}{pr(Y|\theta^{(i)})\pi(\theta^{(i)})}} \\ &= \frac{\sum_{i=1}^m pr(Y)}{\sum_{i=1}^m \frac{pr(Y)}{pr(Y|\theta^{(i)})}} = \frac{mpr(Y)}{pr(Y) \sum_{i=1}^m \frac{1}{pr(Y|\theta^{(i)})}} \\ &= \frac{m}{\sum_{i=1}^m \frac{1}{pr(Y|\theta^{(i)})}} = \frac{m}{\sum_{i=1}^m \left(pr(Y|\theta^{(i)}) \right)^{-1}} \\ &= \left[\frac{1}{m} \sum_{i=1}^m \left(pr(Y|\theta^{(i)}) \right)^{-1} \right]^{-1}, \end{aligned} \quad (2.17)$$

isto é, o inverso da média da verossimilhança invertida, ou seja, a média harmônica considerada também como uma estimativa para $pr(Y)$, (Newton e Raftery, 1994). Este resultado converge quase certamente para o valor correto de $pr(Y)$ quando $m \rightarrow \infty$, mas este geralmente não satisfaz o Teorema Central do Limite devido à possibilidade de ocorrência de um valor de $\theta^{(i)}$ com

verossimilhança pequena alterando o resultado final, (Kass e Raftery, 1995). Mas, na maioria das vezes, fornece resultados precisos e suficientes para a interpretação na escala logaritmica da seção 2.2.

Os métodos de Monte Carlo são computacionalmente intensivos, com o tempo de processamento maior que os aproximados, como é o caso do método de Laplace. São mais flexíveis no que se refere à precisão pois o usuário pode obter mais precisão aumentando o número de replicações no processo de simulação.

2.3.4 Comparação dos métodos de aproximação

A avaliação analítica exata é a melhor pois é a mais precisa e também é mais eficiente computacionalmente, mas só é possível para alguns modelos simples.

O método de Laplace produz uma aproximação precisa e usualmente é bastante eficiente computacionalmente, mas a aplicação do método só é possível para uma pequena classe de modelos. Em particular, a aproximação usando a expressão dada em (2.13) pode ser bastante precisa.

A integração de Monte Carlo e o método *Importance Sampling* são mais flexíveis no que se refere à precisão, porém mais exigentes computacionalmente do que os métodos de aproximação. Em geral são os métodos indicados nos casos de modelos complexos.

Os diferentes métodos de cálculo e aproximação do fator de Bayes são comparados por Rosenkrantz (1992) nos contextos de modelos Normais e modelos Poisson-Gamma usando variáveis independentes e *outliers*. No contexto de modelos lineares generalizados a comparação é feita por Raftery (1993).

2.4 A escolha da densidade *a priori*

Para o cálculo do fator de Bayes, devemos expressar a distribuição *a priori* para os modelos e também para cada parâmetro desconhecido de cada modelo. Pode acontecer que um determinado experimento possua dados semelhantes aos de um experimento já analisado anteriormente e neste caso é possível utilizar a distribuição *a posteriori* desse experimento como distribuição *a priori* do experimento em análise. Essa situação é apropriada mas é rara, sendo assim, a escolha das distribuições *a priori* deve ser feita com cautela.

A obtenção da distribuição *a priori* geralmente depende da experiência de um especialista ou de dados relativamente próximos como foi dito antes. A formulação de uma distribuição *a priori* é um problema específico e não existe uma metodologia geral para definir uma distribuição *a priori* correta em cada situação.

O objetivo da distribuição *a priori* é representar o conhecimento sobre o parâmetro desconhecido antes da realização do experimento.

Quando o interesse é estimar um parâmetro no caso discreto ou no caso contínuo, geralmente são usadas distribuições *a priori* escolhidas de acordo com a situação. Se a amostra é muito grande, o efeito da distribuição *a priori* é pequeno e geralmente é escolhida uma distribuição *a priori* não informativa ou de variância grande, isto é, uma constante como a distribuição uniforme com parâmetros (a, b) , $U(a, b)$, com a e b conhecidos ou uma distribuição *a priori* de Jeffreys, $\pi^J(\theta)$. A razão de usar distribuições *a priori* não informativas é permitir que o experimento seja mais informativo do que a distribuição *a priori*.

Nos testes de hipótese, mesmo quando a amostra é grande, pode existir influência da distribuição *a priori*. Note que na expressão (2.14) da ordem de $O(n^{-1})$, existe efeito em $\pi(\hat{\theta}|H)$, mesmo com o aumento da amostra, em contraste com o caso da estimação onde o *EMV* aproxima a média *a posteriori*. Assim, o fator de Bayes tende a ser mais sensível à escolha da distribuição *a priori* do que a probabilidade *a posteriori* de um intervalo, (Kass, 1993 e Kass e Greenhouse, 1989).

Existem algumas contradições com o cálculo do fator de Bayes em testes de hipótese na seleção de modelos. Uma contradição ocorre quando testamos a hipótese $H_0 : \psi = \psi_0$, e se a estimativa de ψ é encontrada muito distante do valor da hipótese nula ψ_0 . Nesta situação, há evidência de que os dados são improváveis sob a hipótese H_0 . Mas, o cálculo do fator de Bayes indica que os dados favorecem a hipótese nula, concluindo assim que os dados são mais improváveis sob a hipótese H_1 do que sob a hipótese H_0 contradizendo os dados. Esta contradição é devido ao uso de uma distribuição *a priori* com variabilidade grande ou não informativa no parâmetro ψ a ser testado sob a hipótese H_1 , forçando o fator de Bayes a favorecer a hipótese H_0 , notificada por Bartlett (1957), e denominado paradoxo de Bartlett. O paradoxo pode ser resolvido evitando o uso de distribuições *a priori* com variabilidade grande nos parâmetros de teste, ψ , neste caso. As distribuições *a priori* em parâmetros de teste devem ser próprias e preferencialmente com pouca variabilidade, (Kass e Raftery, 1995).

Outra dificuldade na escolha da distribuição *a priori* é quando esta for imprópria, ou seja, $\int \pi(\theta) d\theta \rightarrow \infty$. Por outro lado, o interesse está na distribuição *a posteriori* e como esta é, em geral, própria mesmo quando a distribuição *a priori* não é própria. Assim, em alguns casos não se dá importância à eventual impropriedade de distribuições *a priori*. Conforme Migon e Gamerman (1997), a classe de distribuições *a priori* não informativas propostas por Jeffreys é invariante, porém, eventualmente imprópria. Antes de mostrá-la é necessário definir a medida de informação de Fisher,

$$I(\theta) = E_{Y|\theta} \left[-\frac{\partial^2 \log pr(Y|\theta)}{\partial \theta^2} \right], \quad (2.18)$$

quando θ é unidimensional. Se θ for um vetor paramétrico, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)^T$, a matriz de informação esperada de Fisher é definida por

$$I(\boldsymbol{\theta}) = E_{Y|\boldsymbol{\theta}} \left[-\frac{\partial^2 \log pr(Y|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right], \quad (2.19)$$

com elementos $I_{ij}(\boldsymbol{\theta})$ dados por

$$I_{ij}(\boldsymbol{\theta}) = E_{Y|\boldsymbol{\theta}} \left[-\frac{\partial^2 \log pr(Y|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \text{ para } i, j = 1, \dots, p.$$

Jeffreys (1961), utilizou distribuições *a priori* impróprias em parâmetros “de incômodo” existentes em ambas as hipóteses, H_0 e H_1 . O objetivo era testar o valor da média de uma distribuição Normal, $H_0 : \mu = \mu_0$. Utilizando uma distribuição *a priori* própria em μ sob a hipótese H_1 e fixando $\pi(\sigma) = 1/\sigma$, isto é, uma distribuição *a priori* imprópria, o resultado deste teste fornece uma distribuição preditiva imprópria como em (2.4), mas como o parâmetro σ não é o parâmetro de interesse, ou seja, não é o parâmetro estimado, essa impropriedade não foi considerada por Jeffreys.

Na escolha da distribuição *a priori*, também é possível fazer simplificações quando existir modelos encaixados. Mas em alguns casos existem algumas dificuldades quando, por exemplo, é utilizada a transformação de variáveis. Suponha a hipótese $H_0 : \psi = \psi_0$ sendo testada na presença de um parâmetro desconhecido β , assim,

$$\pi(\beta|H_0) = \int \pi(\beta, \psi|H_1) d\psi, \quad (2.20)$$

onde β e ψ são independentes *a priori* sob a hipótese H_1 . Assim, existe a necessidade de escolher uma distribuição *a priori* para β e uma distribuição *a priori* para ψ , ou seja,

$$\pi(\beta|H_0) = \int \pi(\beta|H_1) \pi(\psi|H_1) d\psi. \quad (2.21)$$

Geralmente, costumamos supor que os parâmetros existentes no modelo são independentes como na equação (2.21).

É possível reescrever a equação (2.20) como $\pi(\beta|H_0) = \pi(\beta|\psi = \psi_0, H_1)$.

De acordo com Kass e Raftery (1995), uma reparametrização deve ser feita neste caso. Sejam (ξ, λ) a reparametrização escolhida, aplicada na hipótese H_1 , e a hipótese H_0 especificada por $\lambda = \lambda_0$. Se $\pi(\xi, \lambda|H_1)$ é obtida através de $\pi(\beta, \psi|H_1)$ pela transformação de variáveis, introduzindo o determinante jacobiano apropriado é possível que $\pi(\xi|H_0)$ não seja obtido através de $\pi(\beta|H_0)$ pela mesma transformação de variáveis. O problema com a transformação de variáveis foi tratado em uma conferência em 1963 por L. J. Savage como paradoxo de Borel, e visto também em Dickey (1975).

Portanto, simplificações em modelos encaixados envolvendo distribuições *a priori*, devem ser tratadas com cautela pois podem afetar os resultados na seleção de modelos.

Uma outra dificuldade na escolha da distribuição *a priori* ocorre quando utilizamos distribuições *a priori* não informativas nos parâmetros de interesse de testes de hipóteses. Para exemplificar esta dificuldade, considere o exemplo em Migon e Gamerman (1997), página 148.

Seja $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ uma amostra de uma distribuição Normal com média θ desconhecida e variância σ^2 , $N(\theta, \sigma^2)$.

Suponha que desejamos testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Considere inicialmente o caso em que a variância σ^2 é conhecida. Atribuindo uma distribuição *a priori* p para H_0 e supondo $\theta|H_1 \sim N(\mu, \tau^2)$, uma distribuição *a priori* conjugada, obtemos

$$\begin{aligned} pr(y|H_1) &= \int pr(y, \theta|H_1) d\theta = \int pr(y, \theta|H_1) pr(\theta|H_1) d\theta = \int pr(y|\theta) pr(\theta) d\theta \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2 \right\} \\ &\quad \int \exp \left\{ -\frac{n}{2\sigma^2} (\theta - \bar{y})^2 \right\} \frac{1}{\sqrt{2\pi\tau}} \exp \left\{ -\frac{1}{2\tau^2} (\theta - \mu)^2 \right\} d\theta. \end{aligned}$$

Após algumas substituições e transformações algébricas obtemos

$$pr(y|H_1) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{ns^2}{2\sigma^2} \right\} \left(\frac{\sigma^2/n}{\tau^2 + \sigma^2/n} \right)^{1/2} \exp \left\{ -\frac{1}{2} \frac{(\bar{y} - \mu)^2}{\tau^2 + \sigma^2/n} \right\},$$

onde $s^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / n$.

Então temos o fator de Bayes dado por

$$\begin{aligned}
B_{01} &= \frac{\left(\frac{1}{2\pi\sigma^2}\right)^{-n/2} \exp\left\{-\frac{ns^2}{2\sigma^2}\right\} \exp\left\{-\frac{n}{2\sigma^2}(\bar{y} - \theta_0)^2\right\}}{\left(\frac{1}{2\pi\sigma^2}\right)^{-n/2} \exp\left\{-\frac{ns^2}{2\sigma^2}\right\} \left(\frac{\sigma^2/n}{\tau^2 + \sigma^2/n}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{(\bar{y} - \mu)^2}{\tau^2 + \sigma^2/n}\right\}} \\
&= \left(\frac{\tau^2 + \sigma^2/n}{\sigma^2/n}\right)^{1/2} \exp\left\{\frac{n}{2}\left[\frac{(\bar{y} - \mu)^2}{\sigma^2 + n\tau^2} - \frac{(\bar{y} - \theta_0)^2}{\sigma^2}\right]\right\}. \tag{2.22}
\end{aligned}$$

Como era de se esperar, o fator de Bayes só depende da amostra através de \bar{y} . Para obtermos o valor da amostra que maximiza o fator de Bayes basta obtermos o valor de \bar{y} que o maximiza. Na expressão (2.22), tomando o logaritmo e derivando em relação a \bar{y} obtemos

$$\begin{aligned}
\frac{\partial \log B_{01}}{\partial \bar{y}} &= \frac{n}{2} \left[\frac{2(\bar{y} - \mu)}{\sigma^2 + n\tau^2} - \frac{2(\bar{y} - \theta_0)}{\sigma^2} \right] = 0 \\
\frac{\partial \log B_{01}}{\partial \bar{y}^2} &= \frac{n}{2} \left[\frac{2}{\sigma^2 + n\tau^2} - \frac{2}{\sigma^2} \right] = \frac{n}{2} \left[-\frac{2n\tau^2}{\sigma^2(\sigma^2 + n\tau^2)} \right] < 0
\end{aligned}$$

e portanto a solução da primeira equação fornece o ponto de máximo

$$\bar{y}_{\max} = \theta_0 + \frac{\sigma^2}{n\tau^2}(\theta_0 - \mu),$$

e o valor máximo do fator de Bayes é

$$\exp\left\{\frac{(\theta_0 - \mu)^2}{2\tau^2}\right\} \left(1 + \frac{n\tau^2}{\sigma^2}\right)^{1/2} > 1.$$

Assim, quanto maior o tamanho amostral n , maiores as chances de H_0 e quanto maior τ^2 maior é o valor máximo de B_{01} .

A incerteza *a priori* desempenha um papel fundamental na comparação de hipóteses Bayesianas. No limite, quando $\tau^2 \rightarrow \infty$, temos que $B_{01} \rightarrow \infty$. Esse limite é o resultado obtido quando usamos distribuições *a priori* não informativas para $\theta|H_1$. Esse resultado, notado inicialmente por Lindley (1956) é conhecido como paradoxo de Lindley e tem sido objeto de estudo recente em comparações de abordagem Bayesianas para testes de hipóteses, Migon e Gamerman (1997).

A questão da escolha da distribuição *a priori* para utilização com o fator de Bayes é tratada mais adiante nos capítulos 3 e 4, onde são apresentados os fatores de Bayes intrínseco e fatores de Bayes fracionais. Estes métodos são apresentados em Berger e Pericchi (1993) e (1996) e O'Hagan (1995), respectivamente, e buscam resolver alguns dos problemas tratados aqui.

2.5 A incerteza dos modelos

No cálculo do fator de Bayes, o objetivo é encontrar um modelo que se ajuste melhor aos dados. Pode acontecer que o modelo indicado pelo fator de Bayes não seja o melhor.

Quando o fator de Bayes é aplicado e um único modelo é selecionado, é através deste modelo que a aplicação é realizada, ignorando a presença de incertezas envolvidas na seleção de modelos. Estas incertezas porém podem levar a um modelo falso comprometendo a conclusão final desta aplicação.

A incerteza de modelos, ou seja, o modelo selecionado é ou não o melhor, pode ser resolvida se adotada a abordagem Bayesiana para calcular as probabilidades *a posteriori* de todos os modelos comparados.

Suponha que $k + 1$ modelos, H_0, H_1, \dots, H_k , são considerados. Cada modelo H_i , $i = 1, \dots, k$ é comparado com H_0 fornecendo os fatores de Bayes $B_{10}, B_{20}, \dots, B_{k0}$. Então, de acordo com (2.1) e (2.3), a probabilidade *a posteriori* de cada modelo H_i é

$$pr(H_i|Y) = \frac{\alpha_i \times B_{i0}}{\sum_{r=0}^k \alpha_r B_{r0}},$$

para $i = 1, \dots, k$ e $\alpha_i = pr(H_i)/pr(H_0)$, $i = 0, \dots, k$ que é a *odds a priori* para H_i contra H_0 e $B_{00} = \alpha_0 = 1$, (Kass e Raftery, 1995).

Introduzindo uma quantidade de interesse Δ , que está bem definida para cada modelo, temos a seguinte densidade *a posteriori* dado o modelo H_i

$$pr(\Delta|Y, H_i) = \int pr(\Delta|Y, \theta_i, H_i)pr(\theta_i|Y, H_i)d\theta_i.$$

A quantidade de interesse, $\Delta = h(\theta)$, é uma função dos parâmetros de interesse de cada modelo em questão. Cada modelo considerado pode conter d parâmetros diferentes e não é sempre que estamos interessados em todos estes parâmetros. A quantidade de interesse é introduzida afim de considerar apenas estes parâmetros de interesse.

Esta equação pode ser usada para fazer inferência sobre Δ , condicionando no modelo H_i . A densidade *a posteriori* utilizada é dada por

$$pr(\Delta|Y) = \sum_{i=0}^k pr(\Delta|Y, H_i)pr(H_i|Y), \quad (2.23)$$

ou seja, a média das distribuições *a posteriori* sob cada modelo ponderada pelas probabilidades *a posteriori* de cada modelo H_i . Calculando a média de todos os modelos é possível obter uma predição melhor em relação aos dados.

Na equação (2.23), H_0, H_1, \dots, H_k são os modelos considerados, e

$$pr(H_i|Y) = \frac{pr(Y|H_i)pr(H_i)}{\sum_{i=1}^k pr(Y|H_i)pr(H_i)}, \quad (2.24)$$

onde $pr(Y|H_i) = \int pr(Y|\theta_i, H_i)pr(\theta_i|H_i)d\theta_i$ é a verossimilhança marginal do modelo H_i e θ_i é o vetor paramétrico de H_i . A quantidade $pr(\theta_i|H_i)$ é a distribuição *a priori* de θ_i dado o modelo H_i , $pr(Y|\theta_i, H_i)$ é a verossimilhança e $pr(H_i)$ é a probabilidade *a priori* de H_i .

A média *a posteriori* de Δ é dada por

$$E[\Delta|Y] = \sum_{i=0}^k E[\Delta|Y, H_i]pr(H_i|Y),$$

e a variância *a posteriori* de Δ é dada por

$$Var[\Delta|Y] = E([\Delta|Y]^2) - (E[\Delta|Y])^2.$$

Como $E([\Delta|Y]^2) = Var[\Delta|Y] + (E[\Delta|Y])^2$, substituindo $E[\Delta|Y]$ por $\sum_{i=0}^k E[\Delta|Y, H_i]pr(H_i|Y)$, obtemos

$$Var [\Delta|Y] = \sum_{i=0}^k \left\{ Var [\Delta|Y, H_i] + (E [\Delta|Y, H_i])^2 \right\} pr(H_i|Y) - (E [\Delta|Y])^2,$$

(Kass e Raftery, 1995).

O problema de tomada de decisão é resolvido maximizando a quantidade esperada *a posteriori*, $E [\Delta|Y]$, de cada ação do procedimento considerado. Isto corresponde à média ponderada das utilidades esperadas condicionais *a posteriori* de cada modelo, $E [\Delta|Y, H_i]$, (Kass e Raftery, 1995), com peso igual à probabilidade *a posteriori* do modelo $pr(H_i|Y)$.

Smith (1991) argumenta sobre a possibilidade de escolher um modelo antes de uma decisão. Não seria conveniente escolher um único modelo antes da tomada de decisão pois podem existir incertezas sobre este modelo. Essas incertezas devem ser quantificadas através do cálculo das quantidades esperadas *a posteriori*, $E [\Delta|Y]$.

A equação (2.23) mostra que selecionar um modelo e condicioná-lo em relação aos dados pode diminuir a incerteza de modelos se uma das $pr(H_i|Y)$ se aproxima desta quantidade ou se a soma é dominada por modelos dos quais os valores de $pr(\Delta|Y, H_i)$ são semelhantes.

Desta forma, ao invés de escolher um único modelo dentre vários, é possível fazer uma média ponderada dos modelos, como na equação (2.23), e trabalhar com esta média.

Suponha a existência de três modelos, M_1 , M_2 e M_3 . A probabilidade de cada modelo é 0.3, 0.4 e 0.3, respectivamente, e o interesse é a escolha de um deles para a estimação de um parâmetro qualquer, θ , contido nos 3 modelos, supondo que θ é o mesmo parâmetro em cada modelo. Dentre os três, o modelo M_2 tem a maior probabilidade, porém a probabilidade é pequena. Quando ocorre de nenhum modelo possuir uma probabilidade considerável para ser escolhido, é possível fazer uma média ponderada de $\hat{\theta}$,

$$\hat{\theta} = pr(M_1)\hat{\theta}_1 + pr(M_2)\hat{\theta}_2 + pr(M_3)\hat{\theta}_3,$$

onde θ_i , para $i = 1, 2, 3$, é o parâmetro estimado referente ao i -ésimo modelo.

Usando esta média, nenhum modelo é descartado e possivelmente esta nos fornecerá resultados melhores, ou seja, diminui o risco do modelo escolhido não se ajustar aos dados.

2.5.1 Janela de Occam

Utilizando a quantidade de interesse para a escolha do melhor modelo vista na seção anterior, existem dois problemas considerados importantes.

O primeiro diz respeito ao cálculo do fator de Bayes, quando é necessário introduzir métodos de integração numérica para uma aproximação, quando o espaço paramétrico é n -dimensional. Mas foi visto na seção 2.3 que existem métodos eficientes para tal aproximação.

O segundo é que o número de termos da equação (2.23) pode ser muito grande. Um exemplo é com modelos de regressão linear com n observações e J variáveis candidatas independentes considerando todos os subconjuntos possíveis, a possibilidade de *outliers* e 4 transformações possíveis de cada variável formará um conjunto inicial com $(2^J \times (O \max) \times 4^{J+1})$ modelos, onde $O \max$ é o número máximo de supostos *outliers*, (Madigan e Raftery, 1994). Mesmo no caso de um problema relativamente pequeno, com $n = 40$, $J = 12$, e $O \max = 5$, isto dará aproximadamente 10^{16} modelos.

Dois algoritmos são introduzidos para a resolução do segundo problema referente ao número de termos da equação (2.23). O primeiro é conhecido como Janela de Occam, que foi a proposta

por Madigan e Raftery (1994) e consiste em selecionar um subconjunto de modelos inicialmente considerados. Isto envolve calcular a média para conjunto muito menor de modelos do que na equação (2.23), facilitando a resolução do problema com a incerteza de modelos.

Madigan e Raftery argumentam que se um modelo é muito menos provável *a posteriori* do que um outro modelo, então este é descartado e não mais considerado. Desta forma, são excluídos da equação (2.23) aqueles modelos que não pertencem ao conjunto

$$A' = \left\{ H_i : \frac{\max_l \{pr(H_l|Y)\}}{pr(H_i|Y)} \leq C \right\},$$

para $i = 0, \dots, k$ e $l = 0, \dots, k'$, com $k' < k$. O número C influencia na quantia de modelos selecionados e a sua determinação é empírica. O valor de C , de acordo com Madigan e Raftery deve ser muito maior do que 1 ($\gg 1$). A escolha de C sofre influência dos dados, do número de modelos considerados e do número de parâmetros existentes.

São excluídos também da equação (2.23) modelos complexos que recebem menos suporte dos dados do que suas contrapartes mais simples, isto é, aqueles pertencendo ao conjunto

$$B = \left\{ H_i : \exists H_l \in A, H_l \subset H_i, \frac{pr(H_l|Y)}{pr(H_i|Y)} > 1 \right\},$$

onde A é o conjunto de elementos que pertencem à A' excluindo os elementos do conjunto B , $A = A' - B$. Então, a equação (2.23) é substituída por

$$pr(\Delta|Y) = \frac{\sum_{H_i \in A} pr(\Delta|H_i, Y) pr(Y|H_i) pr(H_i)}{\sum_{H_i \in A} pr(Y|H_i) pr(H_i)},$$

para $i = 0, \dots, k$.

Uma estratégia de busca para identificar os modelos em A , ou seja, o conjunto final de modelos considerados, consiste de uma sequência de pares de comparações de modelos encaixados. Sejam X e Y dois modelos a serem comparados e $X \subset Y$. Se o modelo X é rejeitado em favor do modelo maior, Y , então todos os modelos encaixados dentro de X são rejeitados. Também, se existe uma evidência para a hipótese H_0 , ou seja, o menor modelo, então a hipótese H_1 é rejeitada, isto é, o maior modelo é descartado. Mas, para rejeitar a hipótese H_0 , é exigida uma forte evidência para o maior modelo, o da hipótese H_1 . Se a evidência é inconclusiva (caindo na janela de Occam), então nenhum modelo é rejeitado.

Geralmente o número de termos da equação (2.23) é bastante reduzido. A solução final é independente da classe inicial considerada, mas se a maioria das classes de modelos iniciais conter o conjunto A e for feita a seleção considerando estas classes, darão o mesmo resultado que o conjunto A .

2.5.2 MCMC (Composição de modelo)

Madigan e York (1992) propõem aproximar a equação (2.23) usando o método MCMC. Eles construíram uma cadeia de Markov irreduzível, $\{H(t)_{t=1,2,\dots}\}$, com espaço de estados E , isto é, o espaço do modelo considerado. A cadeia é constituída de modelos que se comunicam através de seus espaços, e com distribuição de equilíbrio $pr(H_i|Y)$, $i = 0, 1, \dots, k$.

Então, para alguma função $u(H_i) = pr(\Delta|Y, H_i)$, se esta cadeia de Markov é simulada para $t = 1, 2, \dots, m$, a média será aproximada por

$$\hat{U} = \frac{1}{m} \sum_{t=1}^m u(H(t)),$$

e \hat{U} converge com probabilidade 1 para $E[u(H)]$ quando $m \rightarrow \infty$.

Portanto, para calcular a equação (2.23), é fixado

$$u(H) = pr(\Delta|H, Y),$$

ou seja, $u(H)$ é igual à probabilidade *a posteriori* de uma quantidade de interesse bem definida para cada modelo considerado.

Para construir a cadeia de Markov, é definido para cada modelo $H_i, i = 0, \dots, k$, uma vizinhança ao redor deste modelo denotada por $v(H_i)$, e os modelos se diferem por um único parâmetro. É definida também uma matriz de transição P e $P(H_i \rightarrow H_j) = 0$ (H_i salta para $H_j, j = 0, \dots, k$ com probabilidade zero) para todos $H_j \notin v(H_i)$ e $P(H_i \rightarrow H_j) = c > 0$ (H_i salta para H_j com probabilidade positiva) para todo $H_j \in v(H_i)$. A matriz P é dada por

$$P = \begin{bmatrix} & H_0 & \dots & H_k \\ H_0 & P_{H_0, H_0} & \dots & P_{H_0, H_k} \\ & & \dots & \\ H_k & P_{H_k, H_0} & \dots & P_{H_k, H_k} \end{bmatrix},$$

e H_j é aceito com probabilidade

$$\min \left\{ 1, \frac{pr(H_j|Y)}{pr(H_i|Y)} \right\},$$

caso contrário a cadeia permanece no estado H_i .

Capítulo 3

Fator de Bayes Intrínseco

No cálculo do fator de Bayes é necessária a especificação das distribuições *a priori* tanto para os modelos existentes como para seus respectivos parâmetros. Em problemas de estimação com baixa dimensionalidade os métodos bayesianos são eficazes, pois a influência das distribuições *a priori* desaparece quando o tamanho da amostra cresce. No caso de seleção de modelos, mesmo quando o tamanho da amostra for relativamente grande, a distribuição *a priori* influencia no cálculo do fator de Bayes causando resultados indeterminados, como foi visto no capítulo 2, seção 2.4. Berger e Pericchi (1996), com o objetivo de eliminar o problema da influência da distribuição *a priori* no cálculo do fator de Bayes, desenvolveram o fator de Bayes intrínseco, discutido a seguir.

3.1 Definição de Fator de Bayes Intrínseco

Suponha que exista k modelos, H_1, H_2, \dots, H_k , a serem comparados para os dados disponíveis $Y = (y_1, y_2, \dots, y_n)$, e que são utilizadas distribuições *a priori* não informativas, $\pi^N(\theta_i)$. Neste caso, o fator de Bayes é dado por

$$B_{ij}^N = \frac{\int pr(Y|\theta_i) \pi^N(\theta_i) d\theta_i}{\int pr(Y|\theta_j) \pi^N(\theta_j) d\theta_j}, \quad (3.1)$$

onde θ_i e θ_j são os parâmetros dos modelos H_i e H_j , respectivamente, com $i, j = 1, \dots, k$, $pr(Y|\theta_i)$ é a verossimilhança sob o modelo H_i e $pr(Y|\theta_j)$ é a verossimilhança sob o modelo H_j , $\pi^N(\theta_i)$ e $\pi^N(\theta_j)$ são as distribuições *a priori* não informativas dos parâmetros θ_i e θ_j , respectivamente.

A densidade marginal dos dados ou densidade preditiva dos dados é dada por

$$pr^N(Y|H_i) = \int pr(Y|\theta_i) \pi^N(\theta_i) d\theta_i,$$

onde θ_i é o parâmetro de cada modelo H_i , $i = 1, \dots, k$; $pr(Y|\theta_i)$ é a verossimilhança sob o modelo H_i e $\pi^N(\theta_i)$ é a distribuição *a priori* não informativa do parâmetro θ_i .

Berger e Pericchi (1996) sugerem como escolha da distribuição *a priori* não informativa em (3.1), a distribuição *a priori* uniforme $\pi^U(\theta_i) = 1$ ou a distribuição *a priori* de Jeffreys $\pi^J(\theta_i) = (\det(I(\theta_i)))^{1/2}$, onde I é a matriz de informação de Fisher associada ao modelo H_i , como foi visto no capítulo 2, seção 2.4, ou a distribuição *a priori* de referência $\pi^R(\theta_i)$, (Bernardo, 1979 e Berger e Bernardo, 1992).

No cálculo do fator de Bayes existe o problema referente às distribuições *a priori*, $\pi^N(\theta_i)$ e $\pi^N(\theta_j)$, pois estas podem ser impróprias e definidas por constantes arbitrárias c_i e c_j . Como o cálculo de B_{ij}^N depende destas constantes arbitrárias, e considerando o fato que as integrais em (3.1) podem ser infinitas, assim, B_{ij}^N será indeterminado.

Uma solução comum para tratar deste problema é dada por Berger e Pericchi (1996). Esta solução consiste em dividir a amostra completa, Y , em duas sub amostras $y(l)$ e $y(-l)$, onde $y(l)$ é uma amostra de treinamento (*training sample*) e $y(-l)$ são as observações restantes. Esta amostra de treinamento combinada com as distribuições *a priori* $\pi^N(\theta_i)$ e $\pi^N(\theta_j)$ resulta em distribuições *a priori* próprias.

As distribuições *a priori* impróprias são, então, transformadas em distribuições *a posteriori* próprias através da relação

$$pr^N(\theta_i|y(l)) = \frac{pr(y(l)|\theta_i)\pi^N(\theta_i)}{pr^N(y(l)|H_i)}, \quad (3.2)$$

onde $pr(y(l)|\theta_i)$ é a densidade marginal de $Y(l)$ sob o modelo H_i e

$$pr^N(y(l)|H_i) = \int pr(y(l)|\theta_i)\pi^N(\theta_i)d\theta_i. \quad (3.3)$$

Os dados restantes são utilizados para calcular o fator de Bayes usando como distribuição *a priori* $pr^N(\theta_i|y(l))$.

Assim, o fator de Bayes, $B_{ij}(l)$, é dado pelo seguinte resultado.

Resultado 1 - O fator de Bayes resultante da solução dada por Berger e Pericchi (1996) para seleção de modelos é dado por

$$B_{ij}(l) = B_{ij}^N B_{ji}^N(l), \quad (3.4)$$

onde

$$B_{ij}^N = \frac{pr(Y|H_i)}{pr(Y|H_j)}, \quad (3.5)$$

e

$$B_{ji}^N(l) = \frac{pr^N(y(l)|H_j)}{pr^N(y(l)|H_i)}, \quad (3.6)$$

e $B_{ji}^N(l)$ é o fator de Bayes intrínseco (*FBI*) associado à amostra de treinamento $y(l)$, e B_{ij}^N é definido por (3.1).

Demonstração: Através da definição do fator de Bayes em (3.1), segue que

$$B_{ij}(l) = \frac{\int pr(y(-l)|\theta_i, y(l)) pr^N(\theta_i|y(l)) d\theta_i}{\int pr(y(-l)|\theta_j, y(l)) pr^N(\theta_j|y(l)) d\theta_j}.$$

Substituindo (3.2) na expressão anterior e considerando que as observações em Y são independentes, segue que

$$\begin{aligned} B_{ij}(l) &= \frac{\int pr(y(-l)|\theta_i) \frac{pr(y(l)|\theta_i)\pi^N(\theta_i)}{pr^N(y(l)|H_i)} d\theta_i}{\int pr(y(-l)|\theta_j) \frac{pr(y(l)|\theta_j)\pi^N(\theta_j)}{pr^N(y(l)|H_j)} d\theta_j} \\ &= \frac{\int pr(y(-l)|\theta_i) pr(y(l)|\theta_i) \pi^N(\theta_i) pr^N(y(l)|H_j) d\theta_i}{\int pr(y(-l)|\theta_j) pr(y(l)|\theta_j) \pi^N(\theta_j) pr^N(y(l)|H_i) d\theta_j} \end{aligned}$$

$$= \frac{\int pr(Y|\theta_i) \pi^N(\theta_i) d\theta_i}{\int pr(Y|\theta_j) \pi^N(\theta_j) d\theta_j} \times \frac{pr^N(y(l)|H_j)}{pr^N(y(l)|H_i)}.$$

Portanto,

$$B_{ij}(l) = B_{ij}^N B_{ji}^N(l).$$

A equação (3.4) remove a arbitrariedade na escolha das constantes de $\pi^N(\theta_i)$ e $\pi^N(\theta_j)$. A razão arbitrária $\frac{c_i}{c_j}$, que multiplicaria B_{ij}^N , é cancelada pela razão $\frac{c_j}{c_i}$, que multiplicaria $B_{ji}^N(l)$. Porém, o uso de uma amostra de treinamento somente faz sentido se $pr^N(y(l)|H_i)$ for finita na equação (3.3). Este fato é formalizado pela seguinte definição.

Definição 3.1.1 *Uma amostra de treinamento $y(l)$ é própria se $pr(y(l)|H_i) < \infty$ para todos os modelos H_i e $y(l)$ é minimal se nenhum subconjunto de $y(l)$ constitui uma distribuição a priori própria.*

Portanto, uma amostra de treinamento satisfazendo as condições da definição acima garante que todas as distribuições *a posteriori* em (3.2) sejam próprias.

O fator de Bayes intrínseco possui várias propriedades de invariância (ver Berger e Pericchi (1993) para discussão destas e justificações relacionadas).

3.2 Fator de Bayes Intrínseco aritmético

Seja $Y = (y_1, y_2, \dots, y_n)$ os valores observáveis e seja $Y_t = \{y(1), y(2), \dots, y(L)\}$ o conjunto de todas as possíveis amostras de treinamento minimais $y(l)$, $l = 1, \dots, L$. Como $B_{ij}(l)$ depende da escolha da amostra de treinamento minimal através da equação (3.4), é interessante eliminar esta dependência. Isto pode ser feito calculando o fator de Bayes para todas possíveis amostras de treinamento $y(l)$, $l = 1, \dots, L$, e tomando a média desses fatores de Bayes. Esta média é chamada de fator de Bayes intrínseco aritmético, $B_{ij}^{(ma)}$, (Berger e Pericchi, 1996).

O fator de Bayes intrínseco aritmético é dado por

$$B_{ij}^{(ma)} = \frac{1}{L} \sum_{l=1}^L B_{ij}(l) \tag{3.7}$$

$$= \frac{1}{L} \sum_{l=1}^L B_{ij}^N B_{ji}^N(l). \tag{3.8}$$

Como B_{ij}^N não depende de l , segue que

$$B_{ij}^{(ma)} = B_{ij}^N \frac{1}{L} \sum_{l=1}^L B_{ji}^N(l),$$

onde $B_{ji}^N(l)$ é definido como na equação (3.6) para todas possíveis amostras de treinamento.

3.3 Fator de Bayes Intrínseco geométrico

Considere novamente os valores observáveis $Y = (y_1, y_2, \dots, y_n)$ e $Y_t = \{y(1), y(2), \dots, y(L)\}$, o conjunto de todas as possíveis amostras de treinamento minimais $y(l)$, $l = 1, \dots, L$. Calculando o fator de Bayes para todo $y(l)$, multiplicando estes fatores e elevando o produto obtido a $1/L$, onde L é o número total de fatores, obtemos a média geométrica dos fatores de Bayes intrínsecos ou fator de Bayes intrínseco geométrico.

O fator de Bayes intrínseco geométrico é dado por

$$\begin{aligned} B_{ij}^{(mg)} &= \left(\prod_{l=1}^L B_{ij}(l) \right)^{1/L} \\ &= \left(\prod_{l=1}^L B_{ij}^N B_{ji}^N(l) \right)^{1/L}. \end{aligned} \quad (3.9)$$

Como B_{ij}^N não depende de l , segue que

$$B_{ij}^{(mg)} = B_{ij}^N \left(\prod_{l=1}^L B_{ji}^N(l) \right)^{1/L}.$$

A média geométrica dos fatores de Bayes intrínsecos, ou fator de Bayes intrínseco geométrico, também tem a função de eliminar do cálculo do fator de Bayes a dependência da escolha das amostras de treinamento minimal, conforme na equação (3.4).

Berger e Pericchi (1996) mostram que as distribuições *a priori* intrínsecas, $\pi^N(\theta_i)$, são próprias em qualquer caso de variância conhecida, e no caso de variâncias desconhecidas quando $\pi^N(\theta_i)$ for uma distribuição *a priori* de referência, $\pi^R(\theta_i)$, ou uma distribuição *a priori* de Jeffreys modificada, $\pi^{J^*}(\theta_i)$. As distribuições *a priori* intrínsecas não são próprias no caso de variância desconhecida se $\pi^N(\theta_i)$ é uma distribuição *a priori* de Jeffreys.

No caso de variância conhecida, a distribuição *a priori* intrínseca, $\pi^N(\theta_i)$, é uma mistura finita de normais, (Berger e Pericchi, 1996).

No caso de variância desconhecida para distribuições *a priori* de referência, $\pi^R(\theta_i)$, e distribuições *a priori* de Jeffreys modificadas, $\pi^{J^*}(\theta_i)$, a distribuição *a priori* intrínseca, $\pi^N(\theta_i)$, se comporta de forma semelhante a uma mistura finita de densidades *t*-Student multivariadas com um grau de liberdade, (Berger e Pericchi, 1996).

Capítulo 4

Fator de Bayes Fracional

Neste capítulo, voltamos a tratar da questão do fator de Bayes no caso de distribuições *a priori* não informativas, vagas ou impróprias.

Quando a informação *a priori* referente ao parâmetro é vaga, eventualmente podem ser utilizadas distribuições *a priori* impróprias para representar a não informação sobre o parâmetro em questão e assim, o fator de Bayes dependerá de constantes arbitrárias indefinidas tornando o seu cálculo indefinido, como discutido no capítulo 3. Também, com o objetivo de eliminar o problema com distribuições *a priori* impróprias, O'Hagan (1991) propõe o uso do método fator de Bayes parcial e posteriormente O'Hagan (1995) propõe o fator de Bayes fracional.

O fator de Bayes parcial foi originalmente proposto por Lempers (1971), e Spiegelhalter e Smith (1982) também trabalham com a idéia de experimentos minimais com a finalidade de eliminar o problema das constantes indefinidas.

O fator de Bayes fracional evita a arbitrariedade de escolha de uma sub amostra como no caso do fator de Bayes intrínseco ou a necessidade de considerar todas as sub amostras possíveis, eliminando também as constantes indefinidas arbitrárias que surgem a partir do uso de distribuições *a priori* impróprias.

4.1 Definição de Fator de Bayes Fracional

Suponha novamente a existência de k modelos, H_1, H_2, \dots, H_k , a serem comparados para os dados disponíveis $Y = (y_1, y_2, \dots, y_n)$. O fator de Bayes neste caso é dado por

$$B_{ij} = \frac{pr(Y|H_i)}{pr(Y|H_j)} = \frac{\int \pi(\theta_i) pr(Y|\theta_i) d\theta_i}{\int \pi(\theta_j) pr(Y|\theta_j) d\theta_j}, \text{ para } i, j = 1, \dots, k,$$

onde θ_i e θ_j são os parâmetros dos modelos H_i e H_j , respectivamente, com $i, j = 1, \dots, k$, $pr(Y|\theta_i)$ é a verossimilhança sob o modelo H_i e $pr(Y|\theta_j)$ é a verossimilhança sob o modelo H_j , $\pi(\theta_i)$ e $\pi(\theta_j)$ são as distribuições *a priori* dos parâmetros θ_i e θ_j , respectivamente.

Suponha que $\pi(\theta_i)$ e $\pi(\theta_j)$ são distribuições *a priori* impróprias, ou seja,

$$\pi(\theta_i) \propto g_i(\theta_i), \text{ onde } \int g_i(\theta_i) d\theta_i \rightarrow \infty,$$

e

$$\pi(\theta_j) \propto g_j(\theta_j), \text{ onde } \int g_j(\theta_j) d\theta_j \rightarrow \infty.$$

Então, $\pi(\theta_i) = c_i g_i(\theta_i)$ e $\pi(\theta_j) = c_j g_j(\theta_j)$ onde c_i e $c_j, i, j = 1, \dots, k$, são constantes finitas arbitrárias e indefinidas. Assim, o fator de Bayes é dado por

$$B_{ij} = \frac{c_i \int \pi(\theta_i) pr(Y|\theta_i) d\theta_i}{c_j \int \pi(\theta_j) pr(Y|\theta_j) d\theta_j},$$

e o seu cálculo depende das constantes finitas indefinidas c_i e c_j .

Uma solução para eliminar esta dependência consiste em dividir a amostra completa, Y , em duas sub amostras x e z , onde x é a amostra de treinamento (*training sample*) e z são as observações restantes. A amostra de treinamento, x , é utilizada para obter a distribuição *a posteriori*, $pr(\theta_i|x)$, a qual é utilizada como distribuição *a priori* para o restante dos dados, z .

A distribuição *a posteriori* obtida a partir da amostra de treinamento é dada pela equação

$$pr(\theta_i|x) = \frac{pr(x|\theta_i) \pi(\theta_i)}{pr(x|H_i)}. \quad (4.1)$$

Portanto, o fator de Bayes para o restante dos dados é dado por

$$B_{ij} = \frac{\int pr(z|\theta_i, x) pr(\theta_i|x) d\theta_i}{\int pr(z|\theta_j, x) pr(\theta_j|x) d\theta_j}, \text{ para } i, j = 1, \dots, k. \quad (4.2)$$

A equação (4.2) é referida por O'Hagan como fator de Bayes parcial, pois é baseada em apenas uma parte dos dados. O fator de Bayes parcial corresponde ao fator de Bayes intrínseco de Berger e Pericchi (1996), a diferença entre eles é que o fator de Bayes intrínseco é baseado em uma amostra minimal enquanto que o fator de Bayes parcial não possui esta exigência.

Para evitar a arbitrariedade de escolher um particular x , ou simplesmente considerar todos os subconjuntos possíveis de amostras de treinamento, O'Hagan (1995) propõe o fator de Bayes fracional. Este método consiste em tomar um b , tal que $b = m/n$, onde m é o tamanho da amostra de treinamento e n é o tamanho da amostra completa. Observe que b sempre é um número entre zero e 1, pois m é sempre menor que n .

O fator de Bayes fracional é definido como

$$B_{ij}^b = \frac{pr(Y, b|H_i)}{pr(Y, b|H_j)},$$

onde

$$pr(Y, b|H_i) = \frac{\int pr(Y|\theta_i) \pi(\theta_i) d\theta_i}{\int \{pr(Y|\theta_i)\}^b \pi(\theta_i) d\theta_i}, \text{ para } i = 1, \dots, k,$$

e

$$pr(Y, b|H_j) = \frac{\int pr(Y|\theta_j) \pi(\theta_j) d\theta_j}{\int \{pr(Y|\theta_j)\}^b \pi(\theta_j) d\theta_j}, \text{ para } j = 1, \dots, k.$$

Se $\pi(\theta_i)$ e $\pi(\theta_j)$ são distribuições *a priori* impróprias, então segue que

$$pr(Y, b|H_i) = \frac{\int pr(Y|\theta_i) c_i g_i(\theta_i) d\theta_i}{\int \{pr(Y|\theta_i)\}^b c_i g_i(\theta_i) d\theta_i},$$

e

$$pr(Y, b|H_j) = \frac{\int pr(Y|\theta_j) c_j g_j(\theta_j) d\theta_j}{\int \{pr(Y|\theta_i)\}^b c_j g_j(\theta_j) d\theta_j},$$

e como as constantes arbitrárias c_i e c_j não dependem respectivamente de θ_i e θ_j , estas se cancelam evitando que o cálculo do fator de Bayes seja indefinido.

Quando elevamos a função $pr(Y|\theta_i)$ ao expoente b , estamos considerando apenas uma fração em relação a esta função. Esta fração é definida de acordo com o tamanho da amostra de treinamento, m , e o tamanho da amostra total, n . Considere, por exemplo, $n = 100$ e $m = 30$. A função $pr(Y|\theta_i)$, então é elevada ao expoente $0,3$, ou seja, $0,3$ é a fração utilizada da função $pr(Y|\theta_i)$, para a obtenção da densidade *a priori* própria. Deste modo, $0,7$ é a fração utilizada das observações para calcular o fator de Bayes fracional.

Para obter o cálculo de B_{ij}^b são necessárias aproximações através de métodos assintóticos ou computacionais vistos no capítulo 2, seção 2.3.

Capítulo 5

Mistura de Distribuições Binomiais

Os modelos Bayesianos, baseados em misturas de distribuições, formam uma classe de modelos estatísticos que são aplicados quando as observações podem vir de mais que uma distribuição. Este tipo de modelagem está sendo utilizada em diversas áreas, especialmente devido a formulação de novos algoritmos computacionais e pela forma natural de interpretar as diferentes situações em que o experimento é realizado.

Considerando uma amostra de tamanho n de uma variável aleatória, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$, e supondo que esta amostra é obtida de uma população formada por k sub-populações (componentes), com densidades $f(y_i|p_j), j = 1, \dots, k$, o modelo com mistura é definido por

$$pr(y_i|\boldsymbol{\gamma}, \mathbf{p}) = \sum_{j=1}^k \gamma_j f(y_i|p_j), i = 1, \dots, n, \quad (5.1)$$

onde $f(y_i|p_j)$ é a j -ésima distribuição binomial, $1 \leq j \leq k$, e o vetor $\mathbf{p} = (p_1, p_2, \dots, p_k)$ é desconhecido. Cada densidade $f(y_i|p_j)$ é ponderada por uma proporção ou componente desconhecida γ_j , com $0 < \gamma_j < 1$, sujeito à restrição $\sum_{j=1}^k \gamma_j = 1$. Uma definição do modelo com mistura pode ser encontrada em Diebolt e Robert (1994).

A verossimilhança dos parâmetros $(\mathbf{p}, \boldsymbol{\gamma})$, pode ser construída através do modelo (5.1). Assim,

$$L(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) = \prod_{i=1}^n pr(y_i|\mathbf{p}, \boldsymbol{\gamma}) = \prod_{i=1}^n \sum_{j=1}^k \gamma_j f(y_i|p_j). \quad (5.2)$$

Existem três pontos importantes definindo os parâmetros de interesse. O primeiro ponto diz respeito à modelagem dos dados, é a definição do número de componentes da mistura, k . Estes parâmetros podem ser identificados por uma densidade *a priori* de acordo com a ocasião. Quando o número de componentes k é desconhecido, o espaço paramétrico tem dimensionalidade infinita, nestes casos outros métodos devem ser utilizados.

O segundo ponto está relacionado à estimação dos parâmetros das componentes $\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_k$, e dos parâmetros $p_1, p_2, p_3, \dots, p_k$. Há a necessidade de desenvolver um algoritmo computacional para a estimação de tais parâmetros.

O terceiro ponto é a questão da classificação, isto é, a determinação do componente para o qual cada $y_i, i = 1, 2, \dots, n$ pertence. Esta classificação é explicada por Diebolt e Robert como uma estrutura “escondida” ou “oculta” do modelo que pode ser vista como “dados perdidos”,

ou seja, cada observação é associada com um indicador não-observado do componente pelo qual é originada.

Os objetivos são obter informações sobre \mathbf{p} baseando-se na amostra Y_1, Y_2, \dots, Y_n , e classificar cada observação como pertencendo a cada uma das k componentes.

Dempster *et al.*(1977) afirmam que um modelo com mistura pode ser escrito em termos destes “dados perdidos”. Isto é possível introduzindo uma variável \mathbf{Z}_i com dimensionalidade k , indicando a componente do modelo da qual a observação y_i é proveniente, tal que $\sum_{j=1}^k Z_{ij} = 1$ e

$$Z_{ij} = \begin{cases} 1, & \text{se a observação } y_i \text{ é gerada pela componente } j \text{ do modelo, e} \\ 0, & \text{caso contrário.} \end{cases} \quad (5.3)$$

A introdução desta variável latente simplifica a função de verossimilhança e por consequência a distribuição *a posteriori*.

Conforme Diebolt e Robert (1994) a distribuição de probabilidade conjunta dos dados ou função de verossimilhança, considerando Y e Z independentes, pode ser escrita como

$$L(\mathbf{p}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{z}) = pr(\mathbf{y}, \mathbf{z} | \mathbf{p}, \boldsymbol{\gamma}) = \prod_{i=1}^n \prod_{j=1}^k \{\gamma_j f(y_i | p_j)\}^{Z_{ij}}. \quad (5.4)$$

O interesse agora é verificar a que distribuição de probabilidade a variável aleatória Y_i é proveniente, através de \mathbf{Z}_i . Uma maneira de identificar a origem de \mathbf{Z}_i é através do método iterativo conhecido como algoritmo EM (ver Celeux e Diebolt (1985/1992)). Este método substitui cada Z_{ij} pela sua média $\alpha_{ij}^{(m)}$, assim,

$$\alpha_{ij}^{(m)} = pr(Z_{ij}^{(m)} = 1 | \mathbf{p}, \boldsymbol{\gamma}, \mathbf{Y}) = \frac{\gamma_j^{(m)} (f(y_i | p_j))^{(m)}}{\sum_{j=1}^k \gamma_j^{(m)} (f(y_i | p_j))^{(m)}}, \quad (5.5)$$

onde m indica a iteração atual e $\gamma_j^{(m)}$ e $(f(y_i | p_j))^{(m)}$ são as avaliações dos parâmetros, ou seja, $Z_i \sim Multinomial(1; \alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik})$ e $Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{ik})$.

Novas estimativas de máxima verossimilhança são então derivadas dos dados pseudo completados $(y_i, Z_i^{(m)})$.

Com o intuito de evitar defeitos do método EM, Celeux e Diebolt (1985/1992) introduziram uma versão estocástica chamada SEM. Ao invés de estimar os dados ampliados, os autores simulam os Z_i 's da distribuição condicional, ou seja, da distribuição Multinomial com os pesos dados em (5.5).

Uma alternativa para estimar os parâmetros desconhecidos em questão é simular \mathbf{p} e \mathbf{Z} através de suas distribuições *a posteriori* condicionais utilizando os métodos MCMC, *Gibbs Sampling* ou *Metropolis-Hasting*.

Pelo algoritmo *Gibbs Sampling*, esta distribuição estacionária é atingida através de simulações sucessivas das densidades condicionais das componentes.

Pelo algoritmo *Metropolis-Hastings* esta distribuição é atingida simulando de uma distribuição proposta conveniente, aceitando o valor proposto com uma probabilidade definida.

A implementação dos algoritmos MCMC são bastante úteis para a estimação com modelos de mistura, estabelecendo uma aproximação precisa.

O procedimento utilizado para simular os parâmetros de interesse será o algoritmo *Gibbs Sampling* descrito na seção a seguir

5.1 Algoritmo *Gibbs Sampling*

O algoritmo *Gibbs Sampling* foi proposto inicialmente por Geman e Geman (1984) no contexto de restauração de imagens. Mas foi Gelfand e Smith (1990) que completaram o seu desenvolvimento para utilização em inferência Bayesiana.

Este método consiste em um esquema markoviano de atualização, que permite a obtenção de amostras de uma distribuição conjunta através de amostragens iterativas das distribuições condicionais.

O grande interesse por este método é devido à sua simplicidade conceitual e facilidade de implementação.

Suponha que a densidade *a posteriori* de interesse é $pr(\boldsymbol{\theta}|y)$, onde $y = (y_1, y_2, \dots, y_n)$ representa os dados e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ o vetor de parâmetros.

Considere que as densidades condicionais completas estão disponíveis e são dadas por

$$pr(\theta_i|y, \theta_{(-i)}), \text{ para } i = 1, 2, \dots, k,$$

onde $\theta_{(-i)} = (\theta_1, \dots, \theta_{(i-1)}, \theta_{(i+1)}, \dots, \theta_k)$.

O algoritmo *Gibbs Sampling*, fornece uma forma de geração baseada em sucessivas gerações das distribuições condicionais completas. Este algoritmo pode ser descrito como,

i) Escolha valores iniciais para os parâmetros

$$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}),$$

de forma que $pr(\boldsymbol{\theta}^{(0)}) > 0$.

ii) Obtenha um novo valor $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})$ a partir de $\boldsymbol{\theta}^{(j-1)}$ através de sucessivas gerações de valores $\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_k^{(j)}$ das distribuições

$$pr(\theta_1|y, \theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)});$$

$$pr(\theta_2|y, \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)});$$

.....

$$pr(\theta_k|y, \theta_1^{(j)}, \theta_3^{(j)}, \dots, \theta_{k-1}^{(j)}).$$

A repetição de (ii) um grande número de vezes produz uma amostra de $\boldsymbol{\theta}$.

À medida que o número de iterações aumenta a distribuição da cadeia, gerada pelo algoritmo *Gibbs Sampling*, se aproxima de sua distribuição de equilíbrio, ou seja, sob condições gerais de regularidade, quando $n \rightarrow \infty$ temos que $pr(\theta_1^{(n)}, \theta_2^{(n)}, \dots, \theta_k^{(n)})$ tende para $pr(\theta_1, \theta_2, \dots, \theta_k)$.

Uma consequência importante deste resultado é que as médias ergódicas, $\bar{f}_n = \frac{1}{n} \sum_{j=1}^n f(\boldsymbol{\theta}^{(j)})$, convergem quase certamente para $E[f(\boldsymbol{\theta})]$ quando $n \rightarrow \infty$.

Se a indicação de convergência é obtida na iteração n , m iterações após as n iniciais representam uma amostra da cadeia em equilíbrio, ou seja, uma amostra aleatória de tamanho m de $pr(\boldsymbol{\theta}|y)$.

5.1.1 Algoritmo *Gibbs Sampling* com amostra ampliada

Nesta seção, introduzimos o algoritmo *Gibbs Sampling* para amostra ampliada como proposto por Diebolt e Robert (1994), com o objetivo de obter estimativas para os parâmetros desconhecidos.

Uma aproximação para as distribuições marginais *a posteriori* é dada a partir de um valor inicial para os subvetores, $(p_1^{(m)}, p_2^{(m)}, \dots, p_l^{(m)})$, onde l representa o número de parâmetros desconhecidos a serem estimados, com $l \leq k$. Deste modo, inicializando com valores $(p_1^{(0)}, p_2^{(0)}, \dots, p_l^{(0)})$, o algoritmo na etapa m é dado por

i) Gere $Z_i^{(m)} \sim \text{Multinomial}(1, \alpha_i^{(m)})$, $i = 1, \dots, n$; onde $\alpha_i^{(m)} = (\alpha_{i1}^{(m)}, \dots, \alpha_{ik}^{(m)})$ e

$$\alpha_{ij}^{(m)} = \frac{\gamma_j^{(m)}(f(y_i|p_j))^{(m)}}{\sum_{j=1}^k \gamma_j^{(m)}(f(y_i|p_j))^{(m)}}.$$

ii) Na etapa $m + 1$, gere

$$p_1^{(m+1)} \sim pr(p_1|y_i, z_i^{(m)}, p_2^{(m)}, \dots, p_l^{(m)});$$

.....

$$p_l^{(m+1)} \sim pr(p_l|y_i, z_i^{(m)}, p_1^{(m+1)}, \dots, p_{l-1}^{(m+1)}).$$

Observe que o vetor $(p_1^{(m+1)}, p_2^{(m+1)}, \dots, p_l^{(m+1)})$ é gerado pelas suas respectivas distribuições condicionais. O principal objetivo da implementação do algoritmo *Gibbs Sampling* é a convergência para as distribuições marginais de cada parâmetro desconhecido a ser estimado.

5.1.2 Algoritmo *Gibbs Sampling* para mistura de distribuições binomiais

Desenvolvendo a expressão em (5.4) e substituindo $f(y_i|p_j)$ pelo modelo binomial de k componentes, segue que

$$\begin{aligned} L(\gamma, \mathbf{p}|\mathbf{y}, \mathbf{z}) &= \prod_{i=1}^n \prod_{j=1}^k \{\gamma_j f(y_i|p_j)\}^{Z_{ij}} \\ &\propto \prod_{i=1}^n \left[\gamma_1 (p_1)^{Y_i} (1-p_1)^{(n_i-Y_i)} \right]^{Z_{i1}} \left[\gamma_2 (p_2)^{Y_i} (1-p_2)^{(n_i-Y_i)} \right]^{Z_{i2}} \dots \\ &\quad \left[\gamma_k (p_k)^{Y_i} (1-p_k)^{(n_i-Y_i)} \right]^{Z_{ik}} \\ &= \prod_{i=1}^n \left[\gamma_1^{Z_{i1}} (p_1)^{Y_i Z_{i1}} (1-p_1)^{(n_i-Y_i)Z_{i1}} \right] \left[\gamma_2^{Z_{i2}} (p_2)^{Y_i Z_{i2}} (1-p_2)^{(n_i-Y_i)Z_{i2}} \right] \dots \end{aligned}$$

$$\begin{aligned}
& \left[\gamma_k^{Z_{ik}} (p_k)^{Y_i Z_{ik}} (1 - p_k)^{(n_i - Y_i) Z_{ik}} \right] \\
& = \prod_{i=1}^n \gamma_1^{Z_{i1}} \gamma_2^{Z_{i2}} \dots \gamma_k^{Z_{ik}} (p_1)^{Y_i Z_{i1}} (1 - p_1)^{(n_i - Y_i) Z_{i1}} (p_2)^{Y_i Z_{i2}} (1 - p_2)^{(n_i - Y_i) Z_{i2}} \dots \\
& \quad (p_k)^{Y_i Z_{ik}} (1 - p_k)^{(n_i - Y_i) Z_{ik}} .
\end{aligned}$$

Aplicando o produtório, obtemos a função de verossimilhança baseada em dados ampliados com k componentes dada por

$$\begin{aligned}
L(\boldsymbol{\gamma}, \mathbf{p} | \mathbf{y}, \mathbf{z}) & = \gamma_1^{\sum_{i=1}^n Z_{i1}} \gamma_2^{\sum_{i=1}^n Z_{i2}} \dots \gamma_k^{\sum_{i=1}^n Z_{ik}} (p_1)^{\sum_{i=1}^n Y_i Z_{i1}} (1 - p_1)^{\sum_{i=1}^n (n_i - Y_i) Z_{i1}} \dots \\
& \quad (p_2)^{\sum_{i=1}^n Y_i Z_{i2}} (1 - p_2)^{\sum_{i=1}^n (n_i - Y_i) Z_{i2}} \dots (p_k)^{\sum_{i=1}^n Y_i Z_{ik}} (1 - p_k)^{\sum_{i=1}^n (n_i - Y_i) Z_{ik}} \\
& = \prod_{j=1}^k \gamma_j^{\sum_{i=1}^n Z_{ij}} \prod_{j=1}^k p_j^{Y_i Z_{ij}} (1 - p_j)^{\sum_{i=1}^n (n_i - Y_i) Z_{ij}} . \tag{5.6}
\end{aligned}$$

A densidade *a priori* de $\boldsymbol{\gamma}$ e \mathbf{p} , sugerida de acordo com (5.6) e a densidade *a posteriori* conjunta para $(\boldsymbol{\gamma}, \mathbf{p})$ serão dadas a seguir.

Densidade *a priori* para $\boldsymbol{\gamma}$ e \mathbf{p}

Dados $\beta_j > 0$, assumimos que a densidade *a priori* de $\boldsymbol{\gamma}$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_k)$, é uma distribuição Dirichlet com parâmetros β_j , dada por

$$\pi(\boldsymbol{\gamma} | \beta_1, \beta_2, \dots, \beta_k) = \frac{\Gamma(\beta_1 + \beta_2 + \dots + \beta_k)}{\Gamma(\beta_1) \Gamma(\beta_2) \dots \Gamma(\beta_k)} \left(\prod_{j=1}^k \gamma_j^{(\beta_j - 1)} \right), \tag{5.7}$$

onde $\gamma_j \geq 0$, $j = 1, 2, \dots, k$ e $\sum_{j=1}^k \gamma_j = 1$. A distribuição de Dirichlet é uma generalização de uma distribuição *Beta* (β_1, β_2) .

Para \mathbf{p} , a densidade *a priori* assumida é uma distribuição *Beta* com parâmetros (a, b) , com a e b conhecidos, então

$$\pi(\mathbf{p} | a, b) = \prod_{j=1}^k \frac{\Gamma(a_j + b_j)}{\Gamma(a_j) \Gamma(b_j)} p_j^{a_j - 1} (1 - p_j)^{b_j - 1}, \tag{5.8}$$

onde $0 < p_j < 1$, $j = 1, \dots, k$.

Densidade *a posteriori* para $(\boldsymbol{\gamma}, \mathbf{p})$

A densidade *a posteriori* de $(\boldsymbol{\gamma}, \mathbf{p})$ é obtida multiplicando a verossimilhança de $(\boldsymbol{\gamma}, \mathbf{p})$ pelas distribuições *a priori* de $\boldsymbol{\gamma}$ e de \mathbf{p} , com $\boldsymbol{\gamma}$ e \mathbf{p} independentes e dividindo pela constante de normalização $\int L(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) \pi(\boldsymbol{\gamma}|\beta_1, \beta_2, \dots, \beta_k) \pi(\mathbf{p}|a, b) d\boldsymbol{\gamma} d\mathbf{p}$, isto é

$$pr(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) = \frac{L(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) \pi(\boldsymbol{\gamma}|\beta_1, \beta_2, \dots, \beta_k) \pi(\mathbf{p}|a, b)}{\int L(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) \pi(\boldsymbol{\gamma}|\beta_1, \beta_2, \dots, \beta_k) \pi(\mathbf{p}|a, b) d\boldsymbol{\gamma} d\mathbf{p}},$$

ou

$$pr(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) \propto L(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) \pi(\boldsymbol{\gamma}|\beta_1, \beta_2, \dots, \beta_k) \pi(\mathbf{p}|a, b).$$

A verossimilhança de $(\boldsymbol{\gamma}, \mathbf{p})$, denotada por $L(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y})$ conforme a equação (5.2), é dada por

$$L(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) = \prod_{i=1}^n \sum_{j=1}^k \gamma_j f(y_i|p_j).$$

De (5.7) e (5.8), segue que

$$\begin{aligned} pr(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}) &\propto \left[\prod_{i=1}^n \sum_{j=1}^k \gamma_j f(y_i|p_j) \right] \frac{\Gamma(\beta_1 + \beta_2 + \dots + \beta_k)}{\Gamma(\beta_1) \Gamma(\beta_2) \dots \Gamma(\beta_k)} \left(\prod_{j=1}^k \gamma_j^{(\beta_j-1)} \right) \\ &\frac{\Gamma(a_j + b_j)}{\Gamma(a_j) \Gamma(b_j)} \prod_{j=1}^k p_j^{a_j-1} (1-p_j)^{b_j-1} \\ &\propto \left[\prod_{i=1}^n \sum_{j=1}^k \gamma_j f(y_i|p_j) \right] \left(\prod_{j=1}^k \gamma_j^{(\beta_j-1)} \right) \prod_{j=1}^k p_j^{a_j-1} (1-p_j)^{b_j-1}. \end{aligned}$$

Com o intuito de simplificar o cálculo da distribuição *a posteriori*, é introduzida em uma variável latente Z , vista na seção 5.1, dada por (5.3).

Assim, de acordo com a verossimilhança dada por (5.4), a densidade *a posteriori* de $(\boldsymbol{\gamma}, \mathbf{p})$ baseada em dados ampliados, com $\boldsymbol{\gamma}$ e \mathbf{p} independentes *a priori*, é

$$\begin{aligned} pr(\boldsymbol{\gamma}, \mathbf{p}|\mathbf{y}, \mathbf{z}) &\propto \left[\prod_{i=1}^n \prod_{j=1}^k (\gamma_j f(y_i|p_j))^{Z_{ij}} \right] \left(\prod_{j=1}^k \gamma_j^{(\beta_j-1)} \right) \prod_{j=1}^k p_j^{a_j-1} (1-p_j)^{b_j-1} \\ &= \left[\prod_{i=1}^n \prod_{j=1}^k (\gamma_j^{Z_{ij}}) f(y_i|p_j)^{Z_{ij}} \right] \left(\prod_{j=1}^k \gamma_j^{(\beta_j-1)} \right) \prod_{j=1}^k p_j^{a_j-1} (1-p_j)^{b_j-1} \\ &= \left[\prod_{j=1}^k \gamma_j^{\sum_{i=1}^n Z_{ij}} \prod_{i=1}^n \prod_{j=1}^k f(y_i|p_j)^{Z_{ij}} \right] \left(\prod_{j=1}^k \gamma_j^{(\beta_j-1)} \right) \prod_{j=1}^k p_j^{a_j-1} (1-p_j)^{b_j-1} \\ &= \left[\prod_{j=1}^k \gamma_j^{\sum_{i=1}^n Z_{ij}} \prod_{j=1}^k \gamma_j^{(\beta_j-1)} \right] \left[\left(\prod_{j=1}^k p_j^{a_j-1} (1-p_j)^{b_j-1} \right) \prod_{i=1}^n \prod_{j=1}^k f(y_i|p_j)^{Z_{ij}} \right]. \end{aligned}$$

Logo,

$$pr(\boldsymbol{\gamma}, \mathbf{p} | \mathbf{y}, \mathbf{z}) \propto \left[\prod_{j=1}^k \gamma_j^{\sum_{i=1}^n Z_{ij} + \beta_j - 1} \right] \left[\prod_{j=1}^k p_j^{a_j - 1} (1 - p_j)^{b_j - 1} \prod_{i=1}^n \prod_{j=1}^k f(y_i | p_j)^{Z_{ij}} \right].$$

No caso de $f(y_i | p_j)$ ser a distribuição binomial, $Bin(n_i, p_j)$, $i = 1, \dots, n$ e $j = 1, \dots, k$, temos

$$\begin{aligned} pr(\boldsymbol{\gamma}, \mathbf{p} | \mathbf{y}, \mathbf{z}) &\propto \left[\prod_{j=1}^k \gamma_j^{\sum_{i=1}^n Z_{ij} + \beta_j - 1} \right] \left[\prod_{j=1}^k p_j^{a_j - 1} (1 - p_j)^{b_j - 1} \prod_{i=1}^n \prod_{j=1}^k p_j^{Y_i Z_{ij}} (1 - p_j)^{(n_i - Y_i) Z_{ij}} \right] \\ &= \left[\prod_{j=1}^k \gamma_j^{\sum_{i=1}^n Z_{ij} + \beta_j - 1} \right] \left[\prod_{j=1}^k p_j^{a_j - 1} (1 - p_j)^{b_j - 1} \prod_{j=1}^k p_j^{\sum_{i=1}^n Y_i Z_{ij}} (1 - p_j)^{\sum_{i=1}^n (n_i - Y_i) Z_{ij}} \right] \\ &= \left[\prod_{j=1}^k \gamma_j^{\sum_{i=1}^n Z_{ij} + \beta_j - 1} \right] \left[\prod_{j=1}^k p_j^{a_j + \sum_{i=1}^n Y_i Z_{ij} - 1} (1 - p_j)^{b_j + \sum_{i=1}^n (n_i - Y_i) Z_{ij} - 1} \right]. \end{aligned} \quad (5.9)$$

Distribuições condicionais

A distribuição *a posteriori* condicional de γ_j independente de \mathbf{p} é

$$pr(\gamma_j | \mathbf{p}, \mathbf{y}, \mathbf{z}) \propto Dirichlet(S_1 + \beta_1, S_2 + \beta_2, \dots, S_k + \beta_k), \quad (5.10)$$

onde $S_j = \sum_{i=1}^n Z_{ij}$, $i = 1, \dots, n$ e $j = 1, \dots, k$.

A distribuição *a posteriori* condicional de p_j é dada por

$$pr(p_j | \boldsymbol{\gamma}, \mathbf{y}, \mathbf{z}) \propto Beta\left(a_j + \sum_{i=1}^n Y_i Z_{ij}, b_j + \sum_{i=1}^n (n_i - Y_i) Z_{ij}\right), \quad (5.11)$$

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

A distribuição *a posteriori* condicional de \mathbf{Z}_i é dada por

$$Z_i | \mathbf{y}, \boldsymbol{\gamma}, \mathbf{p} \sim Mult(1; \alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik}), \quad (5.12)$$

para $i = 1, \dots, n$, e $j = 1, \dots, k$ com α_{ij} dada pela equação

$$\alpha_{ij} = \frac{\gamma_j f(y_i | p_j)}{\sum_{j=1}^k \gamma_j (f(y_i | p_j))}. \quad (5.13)$$

O algoritmo para gerar amostras das distribuições *a posteriori* condicionais de $\boldsymbol{\gamma}$ e \mathbf{p} é dado a seguir.

Inicializando com valores $(\boldsymbol{\gamma}^{(0)}, \mathbf{p}^{(0)})$, o algoritmo *Gibbs Sampling* no passo m , é

a) Gerar $Z_i^{(m)} \sim Mult(1; \alpha_{i1}^{(m)}, \alpha_{i2}^{(m)}, \dots, \alpha_{ik}^{(m)})$, $i = 1, \dots, n$

$$\alpha_{ij}^{(m)} = \frac{\gamma_j^{(m)} (f(y_i|p_j))^{(m)}}{\sum_{j=1}^k \gamma_j^{(m)} (f(y_i|p_j))^{(m)}}.$$

b) Na etapa $m + 1$, gerar

$$\boldsymbol{\gamma}^{(m+1)} | y_i, z_i^{(m)} \sim Dirichlet \left(\sum_{i=1}^n Z_{i1}^{(m)} + \beta_1, \sum_{i=1}^n Z_{i2}^{(m)} + \beta_2, \dots, \sum_{i=1}^n Z_{ik}^{(m)} + \beta_k \right),$$

$$p_j^{(m)} | y_i, z_i^{(m)} \sim Beta \left[a_j + \sum_{i=1}^n Y_i Z_{ij}^{(m)}, b_j + \sum_{i=1}^n (n_i - Y_i) Z_{ij}^{(m)} \right].$$

Após a cadeia de Markov atingir a distribuição de equilíbrio, os vetores gerados podem ser considerados aproximadamente como uma amostra das distribuições *a posteriori* de $\boldsymbol{\gamma}$ e \mathbf{p} , respectivamente.

Capítulo 6

Aplicação

Neste capítulo utilizamos o fator de Bayes para verificar a adequabilidade de dois modelos a um conjunto de dados com distribuição binomial.

Um dos modelos considera que todas as variáveis binomiais têm a mesma probabilidade de sucesso, enquanto que o segundo modelo considera que todas as binomiais apresenta probabilidades de sucesso distintas.

A metodologia de seleção de modelos via fator de Bayes é aplicada a dois conjunto de dados.

O primeiro conjunto é gerado artificialmente via computador. Como os parâmetros utilizados na geração são conhecidos, esta aplicação nos permite avaliar a eficiência do fator de Bayes na seleção dos modelos mais adequados, nas condições dos dados gerados.

O segundo conjunto de dados aos quais aplicamos o fator de Bayes para selecionar o modelos mais adequado é um conjunto de dados reais. Estes dados são relativos ao desempenho de um jogador de basquete. As informações que dispomos são o número de arremessos realizados e o número de conversões em uma sequência de 380 jogos. São portanto uma sequência de binomiais. Neste caso, estamos interessados em verificar se a probabilidade de acertar um arremesso é constante durante toda a sequência de jogos ou se temos uma probabilidade diferente a cada jogo.

No próximo capítulo aplicamos ao mesmo conjunto de dados uma abordagem intermediária, permitindo que ocorra uma mistura de binomiais.

6.1 Simulação com dados gerados

Nesta seção, propomos um exemplo com dados simulados para ilustrar a aplicação de algumas metodologias apresentadas nos capítulos anteriores.

Inicialmente estamos interessados em selecionar um entre dois modelos. No primeiro modelo, supomos a igualdade das probabilidades. No segundo modelo, supomos que as probabilidades são todas diferentes.

Estes dados simulam 90 observações geradas de distribuições binomiais com parâmetros (n_i, p_i) , $i = 1, \dots, 90$, de forma que $n_i = 100$ e p_i é tal que

$$p_i = \begin{cases} 0,5 & \text{para } i = 1, \dots, 30 \text{ e } i = 61, \dots, 90; \\ q = 0,4, \dots, 0,6 & \text{para } i = 31, \dots, 60, \end{cases} \quad (6.1)$$

variando de 0,1 em 0,1.

Em nossa estratégia de simulação, mantemos p_i , $i = 1, \dots, 30$, $i = 61, \dots, 90$ e n_i constantes e variamos o valor de $p_i = q$, $i = 31, \dots, 60$. Partimos da condição de igualdade das probabilidades, $q = p_i$, $i = 1, \dots, 90$, e variamos q de forma a afastá-los da condição de igualdade.

A geração dos dados e a estimação dos modelos são realizadas utilizando o programa *R*.

6.1.1 Descrição dos modelos

Nesta aplicação, utilizamos dois modelos, M_0 e M_1 . No modelo M_0 temos y_i com distribuição binomial com parâmetros (n_i, p) , $y_i \sim Bin(n_i, p)$, representando os p_i s iguais, $p_i = p_j$, para $i, j = 1, \dots, 90$. No modelo M_1 temos y_i com distribuição binomial com parâmetros (n_i, p_i) , $y_i \sim Bin(n_i, p_i)$, onde os p_i s podem assumir valores diferentes.

A função de verossimilhança do modelo M_0 é dada por

$$\begin{aligned} L(p|\mathbf{y}, \mathbf{n}, M_0) &= \prod_{i=1}^m \left[\binom{n_i}{y_i} p^{y_i} (1-p)^{n_i-y_i} \right] \\ &= \prod_{i=1}^m \binom{n_i}{y_i} p^{\sum_{i=1}^m y_i} (1-p)^{\sum_{i=1}^m (n_i-y_i)}, \end{aligned} \quad (6.2)$$

onde $\mathbf{n} = (n_1, \dots, n_m)$, $\mathbf{y} = (y_1, \dots, y_m)$ e $m = 90$.

A função de verossimilhança para o modelo M_1 é dada por

$$L(\mathbf{p}|\mathbf{y}, \mathbf{n}, M_1) = \prod_{i=1}^m \left[\binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i-y_i} \right], \quad (6.3)$$

onde $\mathbf{p} = (p_1, \dots, p_m)$.

6.1.2 Abordagem Bayesiana

Nesta seção, apresentamos uma abordagem Bayesiana.

Densidade *a priori*

A densidade *a priori* para p , do modelo M_0 , é uma distribuição beta com parâmetros (a_0, b_0) , $p \sim Beta(a_0, b_0)$. Assim

$$\pi(p) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} \left(p^{a_0-1} (1-p)^{b_0-1} \right). \quad (6.4)$$

As densidades *a priori* para p_i , $i = 1, \dots, m$, no modelo M_1 , são distribuições beta com parâmetros (a_1, b_1) , $p_i \sim Beta(a_1, b_1)$ independentes, portanto a densidade conjunta é dada por

$$\pi(\mathbf{p}) = \prod_{i=1}^m \left(\frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} p_i^{a_1-1} (1-p_i)^{b_1-1} \right). \quad (6.5)$$

Para as simulações iniciais, utilizamos $a_0 = b_0 = a_1 = b_1 = 1$, ou seja,

$$p, p_i \sim U(0, 1), \text{ para } i = 1, \dots, m.$$

Densidade *a posteriori*

A densidade *a posteriori* para p , no modelo M_0 , é dada por

$$\begin{aligned} pr(p|\mathbf{y}, \mathbf{n}, M_0) &\propto \prod_{i=1}^m \binom{n_i}{y_i} p^{\sum_{i=1}^m y_i} (1-p)^{\sum_{i=1}^m (n_i - y_i)} \\ &\propto p^{\sum_{i=1}^m y_i} (1-p)^{\sum_{i=1}^m (n_i - y_i)}. \end{aligned}$$

Portanto

$$p|\mathbf{y}, \mathbf{n}, M_0 \sim \text{Beta}\left(\sum_{i=1}^m y_i + 1, \sum_{i=1}^m (n_i - y_i) + 1\right). \quad (6.6)$$

No modelo M_1 , a densidade *a posteriori* para p_i é dada por

$$\begin{aligned} pr(\mathbf{p}|\mathbf{y}, \mathbf{n}, M_1) &\propto \prod_{i=1}^m \left[\binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i - y_i} \right] \\ &\propto \prod_{i=1}^m (p_i^{y_i} (1-p_i)^{n_i - y_i}). \end{aligned}$$

Portanto

$$p_i|\mathbf{y}, \mathbf{n}, M_1 \sim \text{Beta}(y_i + 1, n_i - y_i + 1), \quad (6.7)$$

onde $i = 1, \dots, m$.

6.1.3 Seleção de modelos

As amostras do modelo M_0 são geradas através da densidade *a posteriori* dada por (6.6). Para a geração de amostras do modelo M_1 utilizamos a densidade *a posteriori* dada por (6.7).

Para o cálculo do fator de Bayes utilizamos uma aproximação baseada em amostras dos parâmetros geradas pelo algoritmo MCMC, utilizando como estimativa a média harmônica dada pela equação (2.17), no capítulo 2.

Na Figura 1 é possível verificar através do gráfico que à medida que o valor de q se aproxima de 0,5 o resultado do fator de Bayes favorece este modelo. Os valores de q variam de 0,4 a 0,6.

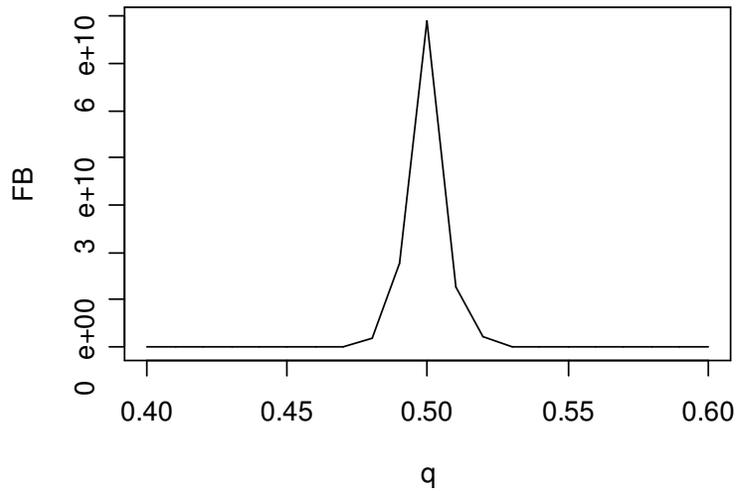


Figura 1: Gráfico de $FB(B_{01})$ versus q

Os resultados da simulação são apresentados na Tabela 3. Como pode ser observado nas colunas 1 a 2 da tabela, mantivemos a probabilidade de sucesso, da distribuição binomial, constante para $p_i = 0,5$, $i = 1, \dots, 30$ e $i = 61, \dots, 90$, e variamos q de 0,4 a 0,6, ou seja, $q = p_i$, $i = 31, \dots, 60$.

Na terceira coluna, apresentamos os valores do fator de Bayes obtido via aproximação pelo método MCMC. Os valores de p e $\mathbf{p} = (p_1, \dots, p_{90})$ foram gerados das distribuições *a posteriori* dadas pelas equações (6.6) e (6.7). O número de repetições de simulação, para obtenção do fator de Bayes, é 2000.

Tabela 3: Valores de B_{01}

$p_i, i = 1, \dots, 30; i = 61, \dots, 90$	$p_i = q, i = 31, \dots, 60$	B_{01}	$2\log(B_{01})$	Modelo
0,5	0,4	$1,74 \times 10^{-10}$	-44,94	M_1
0,5	0,41	$4,17 \times 10^{-6}$	-24,77	M_1
0,5	0,42	$1,15 \times 10^{-3}$	-13,54	M_1
0,5	0,43	2,32	1,68	M_0
0,5	0,44	$3,77 \times 10^3$	16,47	M_0
0,5	0,45	$2,79 \times 10^5$	25,08	M_0
0,5	0,46	$8,3610^5$	27,27	M_0
0,5	0,47	$2,80 \times 10^7$	34,30	M_0
0,5	0,48	$1,06 \times 10^9$	41,58	M_0
0,5	0,49	$2,40 \times 10^9$	43,20	M_0
0,5	0,5	$4,46 \times 10^9$	49,04	M_0
0,5	0,51	$1,75 \times 10^9$	42,56	M_0
0,5	0,52	$1,37 \times 10^9$	42,08	M_0
0,5	0,53	$1,63 \times 10^7$	33,21	M_0
0,5	0,54	$4,29 \times 10^5$	25,94	M_0
0,5	0,55	$4,55 \times 10^5$	26,06	M_0
0,5	0,56	$6,64 \times 10^3$	17,60	M_0
0,5	0,57	4,47	2,99	M_0
0,5	0,58	$2,39 \times 10^{-3}$	-12,07	M_1
0,5	0,59	$9,28 \times 10^{-6}$	-23,17	M_1
0,5	0,6	$4,82 \times 10^{-12}$	-52,12	M_1

Na quarta coluna da tabela está o critério de Kass e Raftery (1995) para comparação de modelos, apresentado na Tabela 2, no capítulo 2. Como podemos observar, para $q < 0,43$ ou $q > 0,57$, a evidência é forte em favor do modelo M_1 , como esperado (p_i s diferentes).

O modelo selecionado é apresentado na quinta coluna da tabela. Observe que para q próximo de 0,5 o modelo selecionado é M_0 . A medida que o fator de Bayes se afasta de 0,5 o modelo selecionado é M_1 , como esperado.

A seguir, apresentamos o gráfico de q versus $2\log(B_{01}) > 0$,

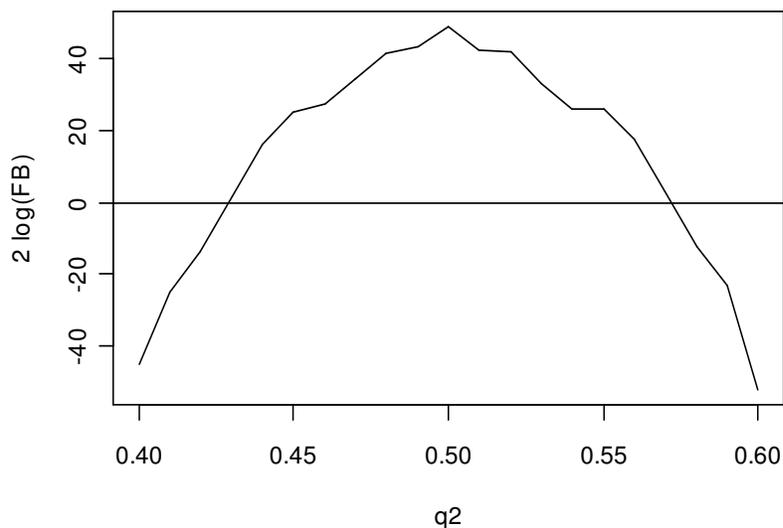


Figura 2: Gráfico de q versus $2\log(B_{01})$

O gráfico da Figura 1 mostra a relação entre q e o fator de Bayes, (B_{01}) . O gráfico da Figura 2 mostra a relação entre q e duas vezes o logaritmo do fator de Bayes, $2\log(B_{01})$. Neste caso os valores de $2\log(B_{01}) < 0$ indicam o modelo M_1 e os valores positivos de $2\log(B_{01})$ indicam o modelo M_0 .

6.1.4 Conclusão

As simulações acima mostram o comportamento do fator de Bayes em condições onde os valores simulados para os parâmetros se afastam do modelo M_0 e gradativamente reforça as condições do modelo M_1 . Observando os resultados, mostramos que quanto mais próximo o valor de q estiver de 0,5, o fator de Bayes privilegia o modelo M_0 , que representa as probabilidades iguais. A medida que q se afasta de 0,5, a evidência em favor de M_0 enfraquece e a evidência em favor de M_1 se fortalece, como é esperado para os modelos em questão.

6.2 Aplicação: Jogos de basquete

Nesta seção, fazemos uma aplicação de seleção de modelos com dados reais. Analisamos a performance de um jogador de basquete buscando verificar se a probabilidade de acerto dos arremessos permanece constante ou se a mesma varia no tempo.

O objetivo desta análise é verificar se, durante o período de observação, a probabilidade de acertos é constante ou se existe variação na probabilidade de acerto de jogo para jogo. Esta aplicação é uma reanálise de um conjunto de dados previamente analisado por Kass e Hsiao (1994).

O fator de Bayes é utilizado como ferramenta para a seleção de modelos.

Posteriormente, o fator de Bayes é utilizado para determinar o número de componentes em um modelo com mistura.

6.2.1 Descrição dos dados

Os dados foram coletados, jogo pós jogo, e referem-se ao aproveitamento do jogador Vinnie Johnson. Estes dados estão disponíveis no site <http://lib.stat.cmu.edu/datasets/Vinnie.Johnson>.

O período referente aos jogos é de 1985 a 1989.

Em um total de m jogos são considerados, em cada jogo, o total de arremessos, denotado por n_i , $i = 1, \dots, m$, e o número de acertos (conversões), denotado por y_i , $i = 1, \dots, m$.

A modelagem utilizada é a mesma apresentada na seção 6.1.2.

A Figura 3 apresenta o gráfico da proporção de acertos, p_i , $i = 1, \dots, 380$, *versus* a sequência das observações. Estas proporções são obtidas a partir da sequência de pares de observações (n_i, y_i) , onde n_i denota o número total de arremessos, y_i denota o número de acertos e $\hat{p}_i = \frac{y_i}{n_i}$.

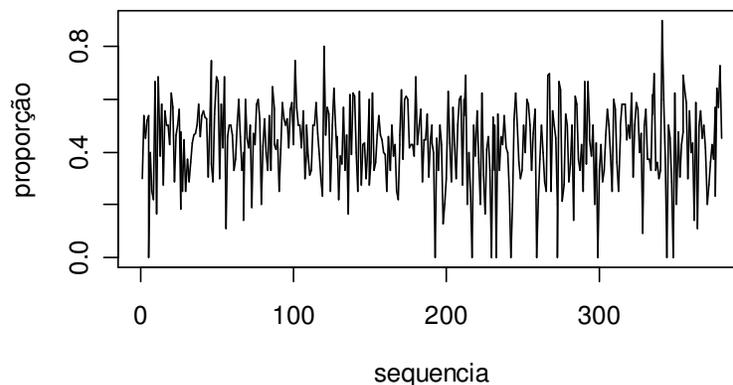


Figura 3: Gráfico de proporções *versus* 380 observações

6.2.2 Discussão

A aplicação do fator de Bayes para a seleção entre os modelos M_0 e M_1 apresentou forte evidência em favor de M_0 , $2\log(B_{01}) = 431$, para 2000 simulações, na aproximação via MCMC.

Este resultado leva à mesma conclusão de Kass e Raftery (1995). A escala de valores dos dois resultados, dos autores e os obtidos aqui, são totalmente diferentes. Esta diferença pode ser em parte explicada pela transformação utilizada por Kass e Raftery, embora este tipo de análise apresenta problemas de instabilidade numérica.

Capítulo 7

Simulação: Modelos com Mistura

Neste capítulo o fator de Bayes é utilizado para seleção de um entre k modelos. Os modelos candidatos são modelos com misturas e as diferenças entre os modelos comparados estão no número de componentes utilizados na mistura, com $1, 2, \dots, k$ componentes.

O primeiro modelo considera todas as variáveis binomiais com a mesma probabilidade de sucesso. Os demais modelos consideram que as variáveis binomiais têm probabilidades distintas, permitindo o uso de modelos com mistura de distribuições binomiais.

Utilizamos três conjuntos de dados para a aplicação do fator de Bayes. Esses conjuntos de dados são gerados através de simulação. A aplicação neste caso nos permite avaliar a precisão do fator de Bayes na identificação do número de componentes nos modelos com mistura, fazendo comparações de dois em dois modelos, incluindo o modelo que considera todas as probabilidades de sucesso iguais, e a probabilidade *a posteriori* para cada modelo.

7.1 Descrição dos dados

Os dados gerados simulam o comportamento de distribuições binomiais. A geração dos dados é semelhante à aplicação do capítulo anterior.

Em nossa estratégia de simulação utilizamos três conjuntos de dados para averiguar a precisão com que o fator Bayes identifica o modelo correto. Cada conjunto de dados consiste de 100 observações geradas a partir da mistura de distribuições binomiais com parâmetros (n_i, p_i) , de forma que $n_i = 100$ para $i = 1, \dots, 100$ e as probabilidades $p_i = P_1$ para $i = 1, \dots, 50$, P_2 para $i = 51, \dots, 80$ e P_3 para $i = 81, \dots, 100$. Estes P 's são dados de acordo com a Tabela 4.

Tabela 4: Valores de P_1, P_2, P_3

Conj. de Dados	P_1	P_2	P_3
1	0,2	0,5	0,8
2	0,3	0,5	0,7
3	0,4	0,5	0,6

O primeiro conjunto de dados é gerado a partir da mistura de distribuições binomiais com três componentes com parâmetros $(100; 0,2)$, $(100; 0,5)$ e $(100; 0,8)$. Análogo ao primeiro, o segundo e o terceiro conjunto de dados são gerados respectivamente a partir dos parâmetros $(100; 0,3)$, $(100; 0,5)$, $(100; 0,7)$ e $(100; 0,4)$, $(100; 0,5)$, $(100; 0,6)$. Observe que inicialmente

temos intervalos maiores entre as probabilidades no primeiro conjunto de dados e gradativamente diminuimos este intervalo de forma que as probabilidades fiquem cada vez mais próximas. O objetivo é identificar o limite em que o fator de Bayes identifica o número de componentes verdadeiro.

A geração dos dados e a estimação dos modelos são realizadas utilizando o programa *R*.

7.2 Descrição dos modelos

Nesta aplicação estamos interessados em selecionar um entre 4 modelos.

Os modelos utilizados são, M_1, M_2, M_3 e M_4 . No modelo M_1 temos y_i com distribuição binomial com parâmetros (n_i, p) , ou seja, um modelo sem mistura ou com uma componente, no modelo M_2 temos mistura de distribuições binomiais com 2 componentes, no modelo M_3 temos mistura de distribuições binomiais com 3 componentes e no modelo M_4 temos mistura de distribuições binomiais com 4 componentes. Cada M_j contempla um modelo de mistura com j componentes, de forma que, $y_i \sim \sum_{j=1}^k \gamma_j \text{Bin}(n_i, p_j)$ onde $0 < \gamma_j < 1$, sujeito à restrição

$$\sum_{j=1}^k \gamma_j = 1.$$

A função de verossimilhança referente ao modelo M_1 é dada pela equação (6.2), onde $\mathbf{n} = (n_1, \dots, n_m)$, $\mathbf{y} = (y_1, \dots, y_m)$ e $m = 100$, vista no capítulo 6.

As funções de verossimilhança baseadas em modelos de mistura de distribuições binomiais são dadas pela equação (5.2), vista no capítulo 5.

Generalizando para o modelo M_k temos k probabilidades de sucesso diferentes p_1, \dots, p_k , $k > 1$, então a função de verossimilhança do modelo M_k é dada por

$$L(p_1, p_2, \dots, p_k, \gamma_1, \gamma_2, \dots, \gamma_k | \mathbf{y}, \mathbf{n}, M_k) = \prod_{i=1}^{100} \sum_{j=1}^k \gamma_j \binom{n_i}{y_i} p_j^{y_i} (1 - p_j)^{n_i - y_i},$$

onde $\gamma_1, \gamma_2, \dots, \gamma_k$ são as k ponderações do modelo.

Como vimos no capítulo 5, não sabemos a qual componente a observação y_i pertence. De acordo com Dempster *et al* (1977), podemos reescrever o modelo com mistura introduzindo uma variável \mathbf{Z}_i indicando a procedência de y_i , denotada por (5.3) no capítulo 5.

$$Z_{ij} = \begin{cases} 1, & \text{se a observação } y_i \text{ é gerada pela componente } j \text{ do modelo, e} \\ 0, & \text{caso contrário.} \end{cases}$$

A introdução desta variável simplifica a função de verossimilhança e conseqüentemente a densidade *a posteriori*. Assim, a função de verossimilhança para M_k baseada em variáveis latentes de acordo com a equação (5.4) é dada por

$$L(p_1, p_2, \dots, p_k, \gamma_1, \gamma_2, \dots, \gamma_k | \mathbf{y}, \mathbf{n}, M_k) = \prod_{i=1}^{100} \prod_{j=1}^k \left\{ \gamma_j \binom{n_i}{y_i} p_j^{y_i} (1 - p_j)^{n_i - y_i} \right\}^{Z_{ij}}.$$

Utilizamos estimadores bayesianos para os parâmetros obtidos através do método MCMC. Para gerar amostras de \mathbf{Z}_i utilizamos o algoritmo *Gibbs Sampling* visto no capítulo 5, seção 5.1.2.

7.3 Abordagem Bayesiana

7.3.1 Densidades *a priori*

A densidade *a priori* para p , do modelo M_1 , é uma distribuição beta com parâmetros (a_0, b_0) , $p \sim \text{Beta}(a_0, b_0)$, dada pela equação (6.4), vista no capítulo 6.

A densidade *a priori* para $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_k)$ é uma distribuição Dirichlet dada por

$$\pi(\boldsymbol{\gamma} | \beta_1, \beta_2, \dots, \beta_k) = \prod_{j=1}^k \gamma_j^{(\beta_j-1)}, \quad (7.1)$$

onde $\beta_j > 0$ são os parâmetros da distribuição Dirichlet, de acordo com (5.7), no capítulo 5.

A densidade *a priori* para cada p_j de cada modelo M_k , é uma distribuição Beta com parâmetros (a, b) , com a e b conhecidos, dada por

$$\pi(\mathbf{p} | a, b) = \prod_{j=1}^k p_j^{a_j-1} (1 - p_j)^{b_j-1}, \quad (7.2)$$

onde $0 < p_j < 1$, de acordo com (5.8), vista no capítulo 5.

7.3.2 Densidade *a posteriori*

A densidade *a posteriori* para p , relativa ao modelo M_1 , é dada pela equação (6.6), vista no capítulo 6.

A densidade *a posteriori* conjunta para $(\gamma_1, \gamma_2, \dots, \gamma_k, p_1, p_2, \dots, p_k)$ relativa ao modelo M_k é obtida da equação (5.9) conforme as densidades *a priori* de p_j e γ_j dadas respectivamente pelas equações (5.7) e (5.8) e a função de verossimilhança baseada em variáveis latentes dada pela equação (5.4) no capítulo 5, assim

$$\begin{aligned} & pr(\gamma_1, \gamma_2, \dots, \gamma_k, p_1, p_2, \dots, p_k | \mathbf{y}, \mathbf{n}, \mathbf{z}, M_k) \\ & \propto \left[\prod_{j=1}^k \gamma_j^{\sum_{i=1}^{100} Z_{ij}} \prod_{j=1}^k \gamma_j^{(\beta_j-1)} \right] \left[\prod_{j=1}^k p_j^{a_j + \sum_{i=1}^{100} y_i Z_{ij} - 1} (1 - p_j)^{b_j + \sum_{i=1}^{100} (n_i - y_i) Z_{ij} - 1} \right]. \end{aligned} \quad (7.3)$$

7.3.3 Densidades condicionais

A densidade *a posteriori* condicional de $\boldsymbol{\gamma}$ do modelo M_k é dada por

$$pr(\boldsymbol{\gamma} | p_1, p_2, \dots, p_k, \mathbf{y}, \mathbf{n}, \mathbf{z}, M_k) \propto \text{Dirichlet}(S_1 + \beta_1, S_2 + \beta_2, \dots, S_k + \beta_k), \quad (7.4)$$

onde $S_j = \sum_{i=1}^{100} Z_{ij}$, $j = 1, \dots, k$, de acordo com (5.10), no capítulo 5.

A distribuição *a posteriori* condicional de p_j é dada por

$$pr(p_j | \gamma_1, \gamma_2, \dots, \gamma_k, \mathbf{y}, \mathbf{n}, \mathbf{z}, M_k) \propto \text{Beta} \left(a_j + \sum_{i=1}^{100} y_i Z_{ij}, b_j + \sum_{i=1}^{100} (n_i - y_i) Z_{ij} \right), \quad (7.5)$$

para, $j = 1, \dots, k$, de acordo com (5.11), no capítulo 5.

A distribuição *a posteriori* condicional de \mathbf{Z}_i é dada de acordo com (5.12), no capítulo 5, ou seja

$$\mathbf{Z}_i | \mathbf{y}, \mathbf{n}, \gamma_1, \gamma_2, \dots, \gamma_k, p_1, p_2, \dots, p_k \sim \text{Mult}(1; \alpha_i), \quad (7.6)$$

onde $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik})$ são os parâmetros da distribuição Multinomial, para $i = 1, \dots, 100$ com α_{ij} dada pela equação (5.13), também no capítulo 5.

O algoritmo utilizado para gerar amostras das distribuições *a posteriori* condicionais é o *Gibbs Sampling*, visto no capítulo 5, seção 5.1.2.

7.4 Estudo de simulação

Nesta seção realizamos o estudo de simulação referente aos três conjuntos de dados gerados.

7.4.1 1º conjunto de dados

Para o cálculo do fator de Bayes utilizamos uma aproximação baseada em amostras dos parâmetros gerados pelo algoritmo MCMC, utilizando novamente como estimativa a média harmônica dada pela equação (2.17), no capítulo 2. O número de repetições de simulação para obtenção do fator de Bayes é 20000, com *burn in* de 1000 e com saltos de 10 em 10.

Inicialmente utilizamos para análise o conjunto de dados gerados através de mistura de distribuições binomiais com 3 componentes com parâmetros $(100; 0, 2)$, $(100; 0, 5)$ e $(100; 0, 8)$. Através da distribuição binomial com parâmetros $(100; 0, 2)$ geramos 50 observações, através da distribuição binomial com parâmetros $(100; 0, 5)$ geramos 30 observações e através da distribuição binomial $(100; 0, 8)$ geramos 20 observações.

A figura 4 mostra o gráfico do conjunto de dados gerados.

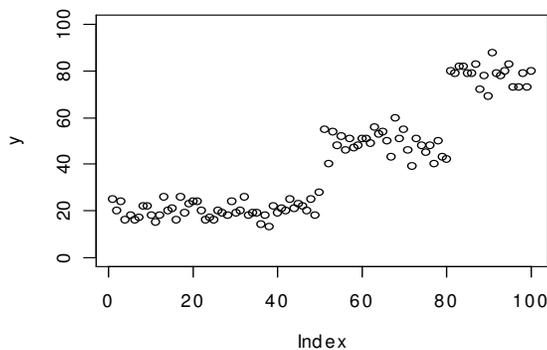


Figura 4: Dados gerados $p = (0, 2; 0, 5; 0, 8)$

Primeiramente ajustamos o primeiro conjunto de dados gerados utilizando o modelo com distribuição binomial com parâmetros (n_i, p) , admitindo que todas as probabilidades são iguais. Os resultados podem ser vistos através da Tabela 5. Na primeira coluna da tabela temos o parâmetro estimado, ou seja, p que é a probabilidades da distribuição binomial. Na segunda e na terceira coluna da tabela temos a média e o desvio de p . Nas demais colunas temos os quantis de p de 2,5%, 25%, 50%, 75% e 97,5%, respectivamente.

Tabela 5: *Resumo a posteriori de p*

Parâmetro	Média	Desvio	2,5%	25%	50%	75%	97,5%
p	0,4935	0,0156	0,4593	0,4767	0,4992	0,5020	0,5103

Através da Tabela 5 podemos observar que a média de p é de aproximadamente 0,5.

O segundo modelo utilizado para o ajuste do primeiro conjunto de dados é o modelo com mistura de distribuições binomiais com dois componentes. Os resultados dessa simulação podem ser vistos através da Tabela 6. Na primeira coluna da tabela temos os parâmetros estimados, ou seja, γ_1, γ_2 , que são as ponderações atribuídas a cada componente da mistura e p_1, p_2 são as probabilidades de cada distribuição binomial. Na segunda e na terceira coluna da tabela temos as médias e os desvios desses parâmetros. Nas demais colunas temos os quantis dos parâmetros de 2,5%, 25%, 50%, 75% e 97,5%, respectivamente.

Tabela 6: *Resumos a posteriori dos parâmetros - dois componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,4801961	0,01023	0,4608	0,4706	0,4804	0,4902	0,5000
γ_2	0,5198039	0,01023	0,5000	0,5098	0,5196	0,5294	0,5392
p_1	0,6156185	0,00830	0,5996	0,6099	0,6155	0,6211	0,6318
p_2	0,2097455	0,00675	0,1967	0,2051	0,2098	0,2144	0,2228

Através da Tabela 6 podemos observar que a média de ambos os gamas permanece próxima de 0,5. A média de p_1 permanece próxima de 0,6 e de p_2 permanece próxima de 0,2, como estes p 's foram estimados de acordo com os dados gerados pela mistura de distribuições binomiais com três componentes com probabilidades 0,2; 0,5 e 0,8 então p_2 identificou a probabilidade de uma componente da mistura 0,2 e p_1 permaneceu entre as probabilidades das outras duas componentes, 0,5 e 0,8. Os resultados das estimativas não foram satisfatórios devido ao ajuste do modelo com mistura com dois componentes, visto que os dados foram gerados a partir do modelo com mistura com três componentes.

Nas Figuras 5 e 6 a seguir mostramos os gráfico das iterações de p_1 e p_2 , respectivamente.

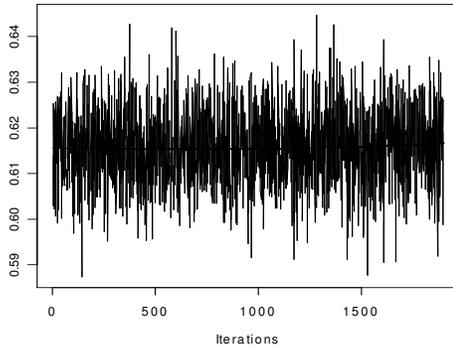


Figura 5: Iterações de p_1

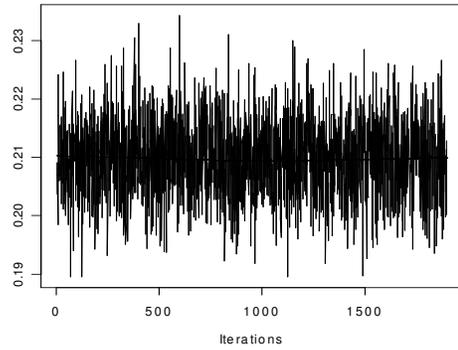


Figura 6: Iterações de p_2

Na Figura 7 mostramos o gráfico das estimativas das densidade de p_1 e na Figura 8 mostramos o gráfico da densidade de p_2 .

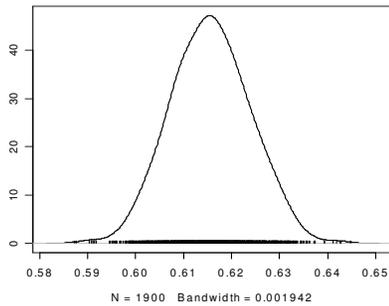


Figura 7: Densidade de p_1

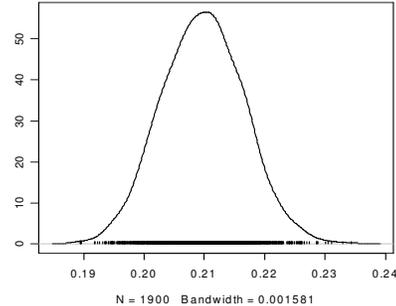


Figura 8: Densidade de p_2

Nas Figuras 9 e 10 são apresentados os gráficos das autocorrelações de p_1 e p_2 , respectivamente.

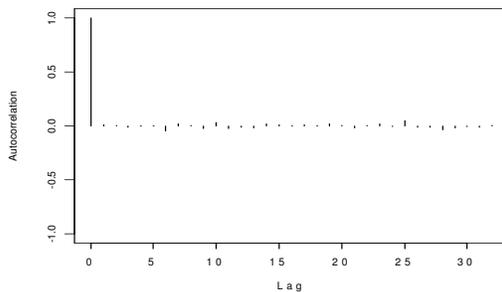


Figura 9: Autocorrelação de p_1

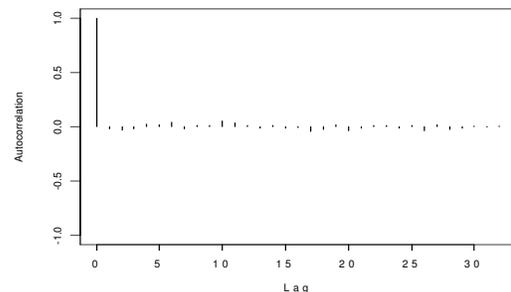


Figura 10: Autocorrelação de p_2

Apenas como ilustração, mostramos a seguir os gráficos das densidades, iterações e autocorrelações referente aos gamas, considerando os mesmos para as demais simulações posteriores. As Figuras 11 e 12 mostram os gráficos das iterações de γ_1 e γ_2 , respectivamente.

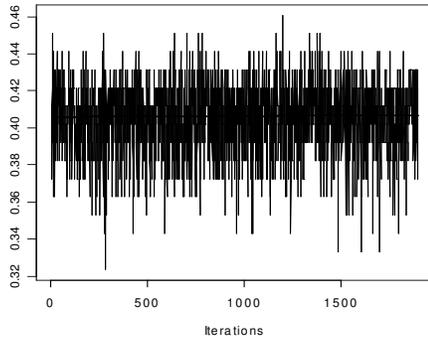


Figura 11: Iterações de γ_1

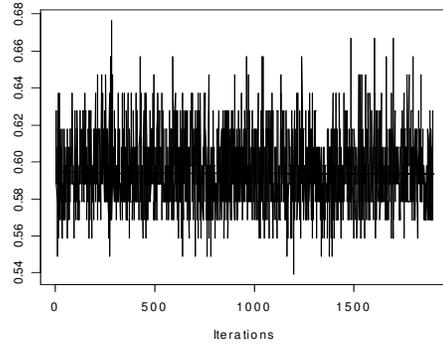


Figura 12: Iterações de γ_2

Nas Figuras 13 e 14 mostramos os gráficos das estimativas das densidades de γ_1 e γ_2 , respectivamente.

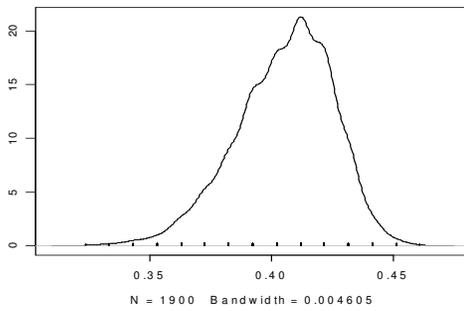


Figura 13: Densidade de γ_1

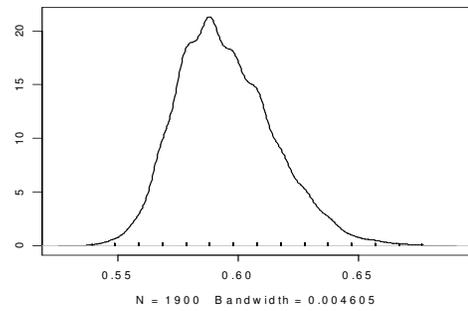


Figura 14: Densidade de γ_2

As Figuras 15 e 16 mostram os gráficos das autocorrelações de γ_1 e γ_2 , respectivamente.

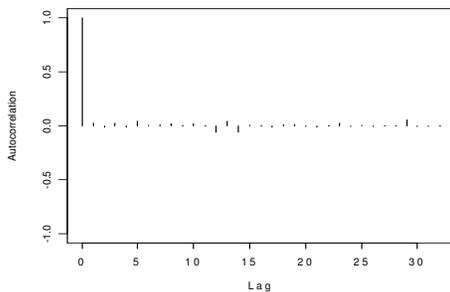


Figura 15: Autocorrelação de γ_1

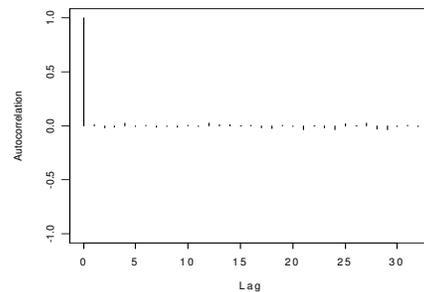


Figura 16: Autocorrelação de γ_2

O terceiro modelo utilizado para o ajuste do primeiro conjunto de dados gerados é o modelo com mistura de distribuições binomiais com três componentes. Podemos observar os resultados da simulação através da Tabela 7. Na primeira coluna da tabela temos os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 7: *Resumo a posteriori dos parâmetros - três componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,2038	0,00116	0,2039	0,2039	0,2039	0,2039	0,2039
γ_2	0,4951	0,00545	0,4951	0,4951	0,4951	0,4951	0,4951
γ_3	0,3010	0,00128	0,301	0,301	0,301	0,301	0,301
p_1	0,7841	0,00918	0,7664	0,7777	0,7840	0,7907	0,8021
p_2	0,2020	0,00565	0,1907	0,1982	0,2022	0,2056	0,2136
p_3	0,4887	0,00903	0,4708	0,4828	0,4887	0,4948	0,5061

Através da Tabela 7 observamos que a média de γ_1 se aproxima de 0,2, o γ_2 permanece próximo de 0,5 e o γ_3 permanece próximo de 0,3, ou seja, atribuiu pesos para cada componente da mistura de acordo com o número de observações geradas por cada componente, satisfazendo nossas expectativas. A média de p_1 permanece próxima de 0,8, a média de p_2 permanece próxima de 0,2 e a média de p_3 permanece próxima de 0,5, satisfazendo também nossas expectativas pois estes p 's foram estimados através da mistura de distribuições binomiais com três componentes com probabilidades 0,2, 0,8 e 0,5.

Nas Figuras 17, 18 e 19 a seguir mostramos os gráficos das iterações de p_1 , p_2 e p_3 , respectivamente.

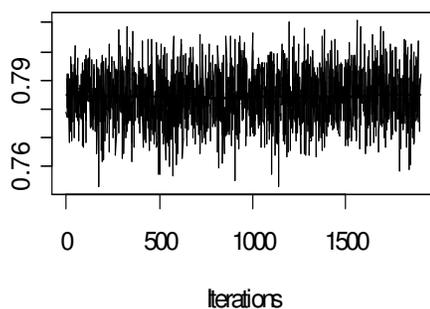


Figura 17: Iterações de p_1

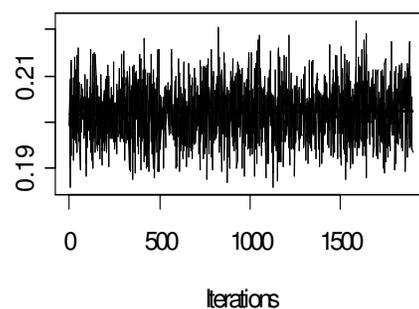


Figura 18: Iterações de p_2

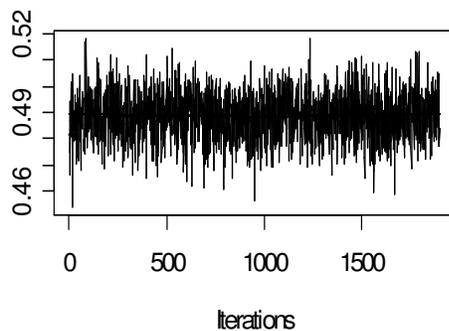


Figura 19: Iterações de p_3

Os gráficos das densidades de p_1, p_2 e p_3 são mostrados abaixo nas Figuras 20, 21 e 22, respectivamente.

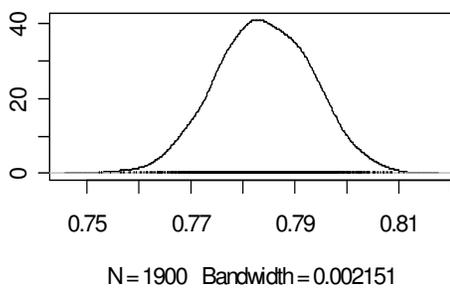


Figura 20: Densidade de p_1

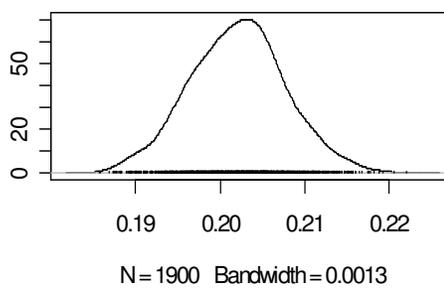


Figura 21: Densidade de p_2

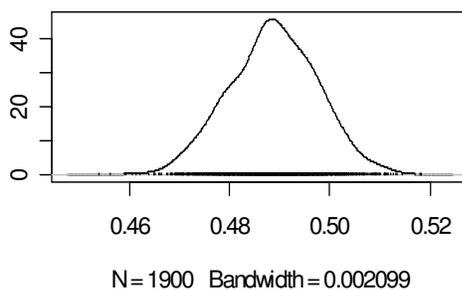


Figura 22: Densidade de p_3

A seguir mostramos os gráficos das autocorrelações de p_1, p_2 e p_3 através das Figuras 23, 24 e 25.

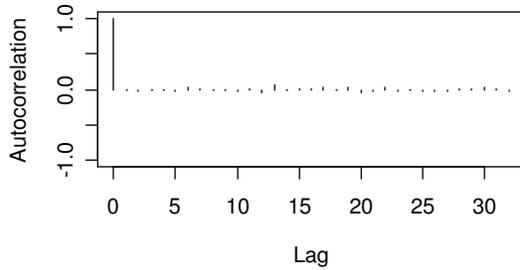


Figura 23: Autocorrelação de p_1

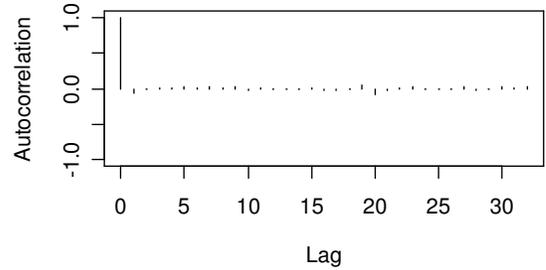


Figura 24: Autocorrelação de p_2

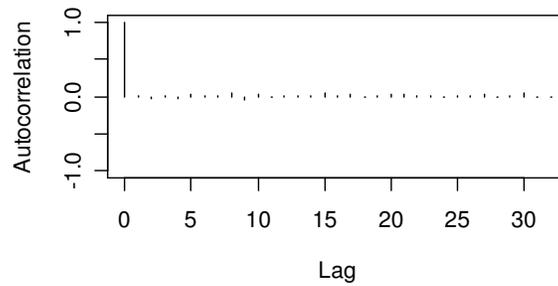


Figura 25: Autocorrelação de p_3

O quarto modelo utilizado para o ajuste do primeiro conjunto de dados gerados é o modelo com mistura de distribuições binomiais com quatro componentes. Os resultados da simulação são vistos através da Tabela 8. Na primeira coluna da tabela temos os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 8: *Resumo a posteriori dos parâmetros - quatro componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,26203	0,14675	0,02885	0,16346	0,20192	0,38462	0,49038
γ_2	0,31224	0,16302	0,02885	0,19231	0,29808	0,49038	0,49038
γ_3	0,20589	0,11837	0,01923	0,11538	0,20192	0,29808	0,49038
γ_4	0,21983	0,11075	0,02885	0,13462	0,21154	0,29808	0,49038
p_1	0,47970	0,24958	0,18646	0,20504	0,48598	0,77709	0,80454
p_2	0,39261	0,25088	0,18519	0,20016	0,2091	0,72606	0,80448
p_3	0,57022	0,223	0,1879	0,4740	0,5154	0,7824	0,8187
p_4	0,51308	0,1832	0,1890	0,4736	0,4898	0,5536	0,8068

Através da Tabela 8 podemos observar que a média dos gamas varia de 0,20589 à 0,31224 e nenhum gama se aproxima de 0,5, somente γ_3 se aproxima de 0,2 e γ_2 se aproxima de 0,3. Ajustamos o modelo com mistura com quatro componentes aos dados gerados, e as estimativas das ponderações para cada componente da mistura não foram satisfatórias visto que os dados foram gerados a partir do modelo com mistura com três componentes.

Em relação aos p 's, note que a média de nenhum p se aproxima de 0,2 e de 0,8, porém temos que as médias de p_1 e de p_4 se aproximam de 0,5. Os resultados das estimativas não foram satisfatórios devido também ao ajuste do modelo com mistura com quatro componentes.

Nas Figuras 26, 27, 28 e 29 a seguir mostramos os gráficos das iterações de p_1 , p_2 , p_3 e p_4 , respectivamente.

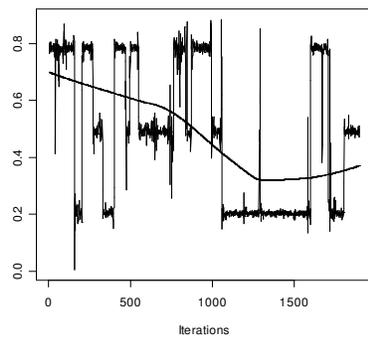


Figura 26: Iterações de p_1

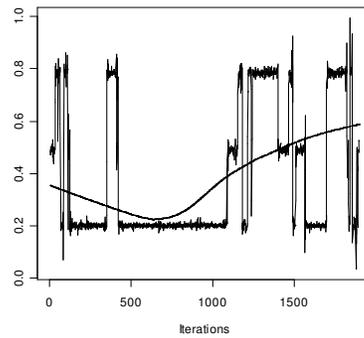


Figura 27: Iterações de p_2

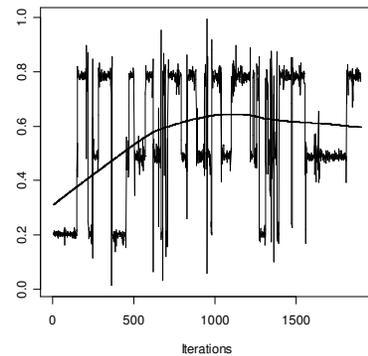


Figura 28: Iterações de p_3

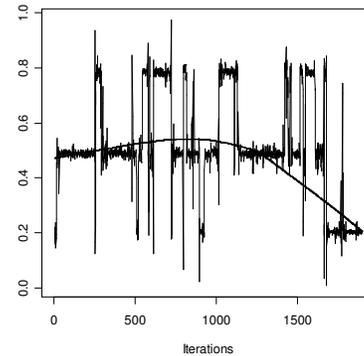


Figura 29: Iterações de p_4

As Figuras 30, 31, 32 e 33 mostram os gráficos das densidades de p_1 , p_2 , p_3 e p_4 , respectivamente.

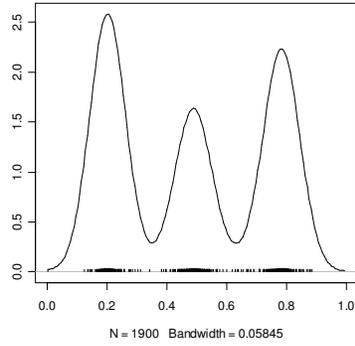


Figura 30: Densidade de p_1

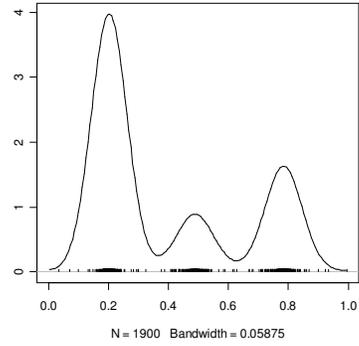


Figura 31: Densidade de p_2

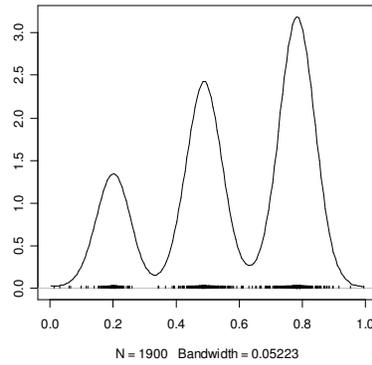


Figura 32: Densidade de p_3

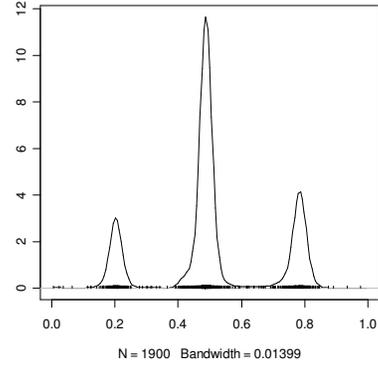


Figura 33: Densidade de p_4

As Figuras 34, 35, 36 e 37 mostram os gráficos das autocorrelações de p_1 , p_2 , p_3 e p_4 , respectivamente.

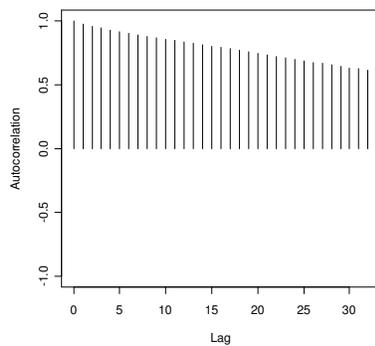


Figura 34: Autocorrelação de p_1

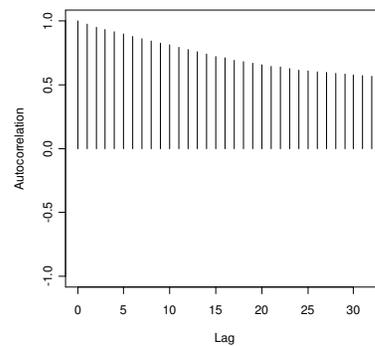


Figura 35: Autocorrelação de p_2

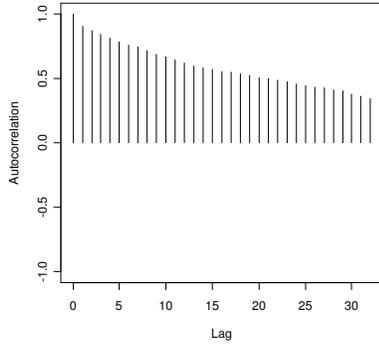


Figura 36: Autocorrelação de p_3

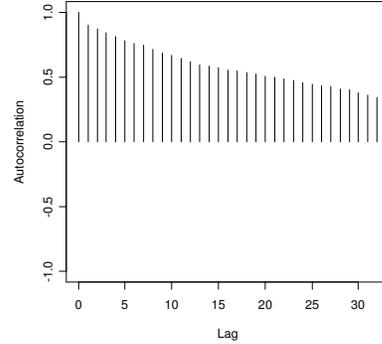


Figura 37: Autocorrelação de p_4

7.4.2 2º conjunto de dados

O segundo conjunto de dados utilizado para análise são os dados gerados através de mistura de distribuições binomiais com 3 componentes com parâmetros $(100; 0, 3)$, $(100; 0, 5)$ e $(100; 0, 7)$. Através da distribuição binomial com parâmetros $(100; 0, 3)$ geramos 50 observações, através da distribuição binomial com parâmetros $(100; 0, 5)$ geramos 30 observações e através da distribuição binomial $(100; 0, 7)$ geramos 20 observações.

A Figura 38 mostra o gráfico deste conjunto de dados.

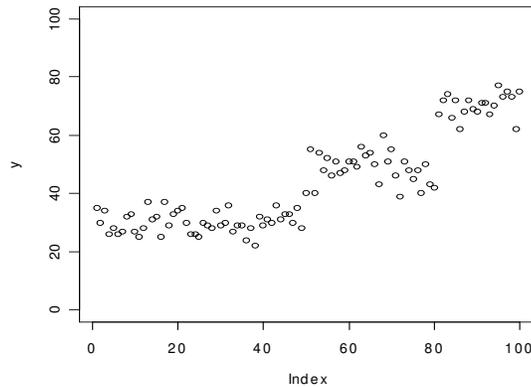


Figura 38: Dados gerados $p = (0, 3; 0, 5; 0, 7)$

Ajustamos o segundo conjunto de dados gerados utilizando o modelo com distribuição binomial com parâmetros (n_i, p) , admitindo que todas as probabilidades são iguais. Os resultados podem ser vistos através da Tabela 9. Na primeira coluna da tabela temos o parâmetro estimado, ou seja, p . Na segunda e na terceira coluna da tabela temos a média e o desvio de p . Nas demais colunas temos os quantis de p .

Tabela 9: *Resumo a posteriori de p*

Parâmetro	Média	Desvio	2,5%	25%	50%	75%	97,5%
p	0,4992	0,0104	0,4785	0,4898	0,4995	0,5004	0,5019

Através da Tabela 9 podemos observar que a média de p se aproxima de 0,5.

O segundo modelo utilizado para o ajuste do segundo conjunto de dados é o modelo com mistura de distribuições binomiais com dois componentes. Os resultados dessa simulação podem ser vistos através da Tabela 10. Na Tabela 6 temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 10: *Resumos a posteriori dos parâmetros - dois componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,4051703	0,0196625	0,3627	0,3922	0,4118	0,4216	0,4412
γ_2	0,5948297	0,0196625	0,5588	0,5784	0,5882	0,6078	0,6373
p_1	0,6081627	0,0103533	0,5894	0,6009	0,6077	0,6148	0,6294
p_2	0,3238457	0,0078999	0,3093	0,3182	0,3237	0,3289	0,3397

Através da Tabela 10 podemos observar que nenhuma média de gama se aproximou de 0,5, 0,3 ou de 0,2. Observe que a média de p_1 se aproxima de 0,6 e a média de p_2 se aproxima de 0,3, como estes p 's foram estimados de acordo com os dados gerados pela mistura de distribuições binomiais com três componentes com probabilidades 0,3; 0,5 e 0,7 então p_2 se aproximou de 0,3, e p_1 permaneceu entre as probabilidades das outras duas componentes, 0,5 e 0,7. Os resultados das estimativas não foram satisfatórios devido novamente ao ajuste do modelo com mistura com dois componentes, visto que os dados foram gerados a partir do modelo com mistura com três componentes.

Nas Figuras 39 e 40 a seguir mostramos os gráficos das iterações de p_1 e p_2 , respectivamente.

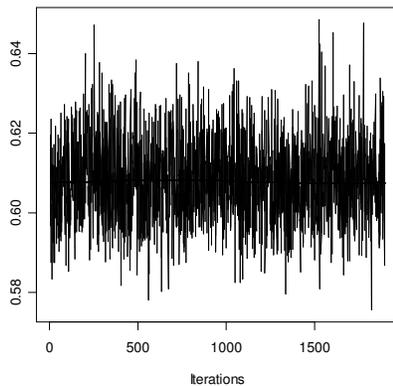


Figura 39: Iterações de p_1

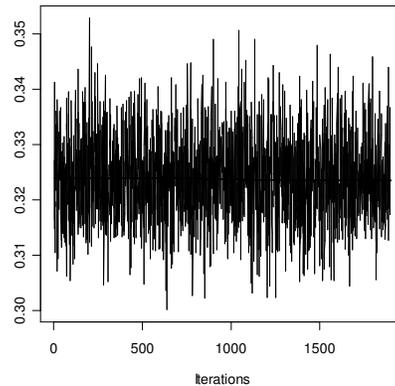


Figura 40: Iterações de p_2

As Figuras 41 e 42 mostram os gráficos das densidades de p_1 e p_2 , respectivamente.

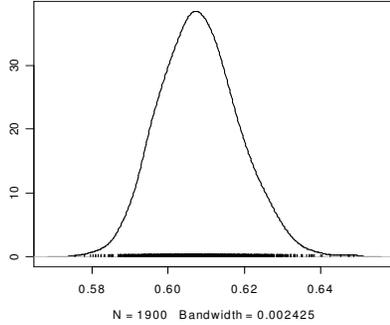


Figura 41: Densidade de p_1

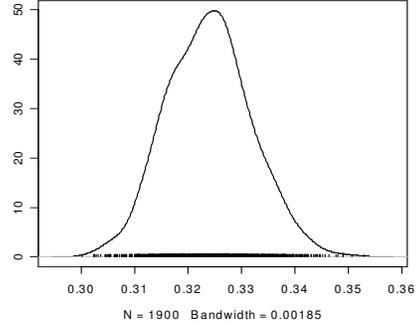


Figura 42: Densidade de p_2

Os gráficos das autocorrelações de p_1 e p_2 são mostrados a seguir através das Figuras 43 e 44, respectivamente.

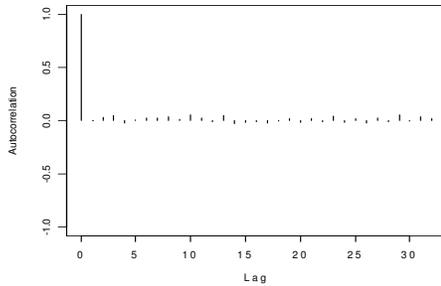


Figura 43: Autocorrelação de p_1

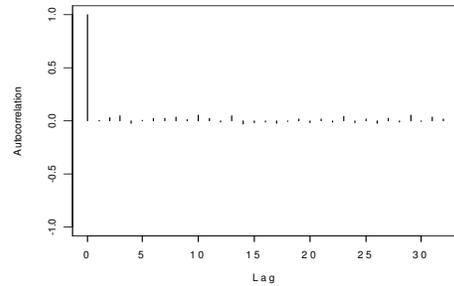


Figura 44: Autocorrelação de p_2

O terceiro modelo utilizado para o ajuste do segundo conjunto de dados gerados é o modelo com mistura de distribuições binomiais com três componentes. Podemos observar os resultados da simulação através da Tabela 11. Análogo à Tabela 7, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 11: *Resumo a posteriori dos parâmetros - três componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,20365	0,00917	0,1845	0,1942	0,2039	0,2136	0,2233
γ_2	0,51383	0,01571	0,4854	0,5049	0,5146	0,5243	0,5437
γ_3	0,28251	0,01652	0,2524	0,2718	0,2816	0,2913	0,3204
p_1	0,70127	0,01118	0,6792	0,6940	0,7014	0,7084	0,7235
p_2	0,30702	0,00689	0,2934	0,3024	0,3069	0,3118	0,3204
p_3	0,49502	0,01228	0,4711	0,4868	0,4951	0,5036	0,5187

Através da Tabela 11 observamos que γ_1 se aproxima de 0,2, γ_2 se aproxima de 0,5 e γ_3 se aproxima de 0,3, satisfazendo nossas expectativas, análogo ao caso do primeiro conjunto de dados ajustando-os ao modelo com mistura com três componentes. A média de p_1 permanece próxima de 0,7, a média de p_2 permanece próxima de 0,3 e a média de p_3 permanece próxima de 0,5, satisfazendo também nossas expectativas pois estes p 's foram estimados através da mistura de distribuições binomiais com três componentes com probabilidades 0,3, 0,5 e 0,7.

Nas Figuras 45, 46 e 47 a seguir mostramos os gráficos das iterações de p_1 , p_2 e p_3 , respectivamente.

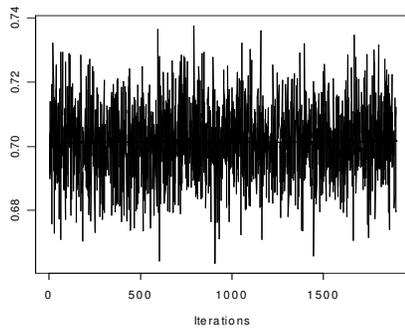


Figura 45: Iterações de p_1

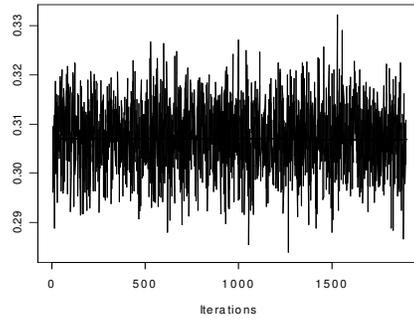


Figura 46: Iterações de p_2

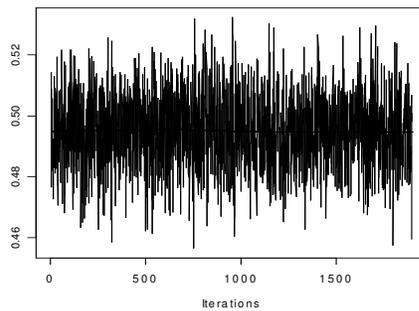


Figura 47: Iterações de p_3

Nas Figuras 48, 49 e 50 mostramos os gráficos referentes às densidades de p_1 , p_2 e p_3 , respectivamente.

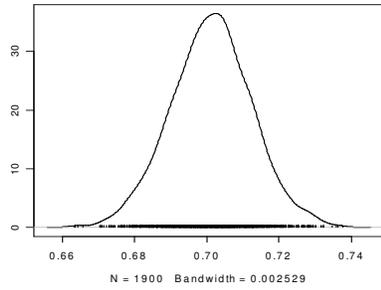


Figura 48: Densidade de p_1

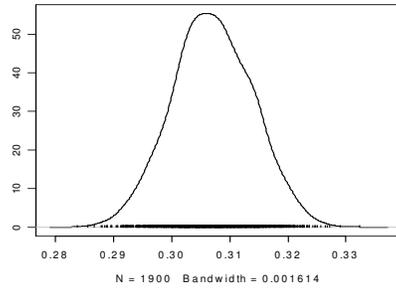


Figura 49: Densidade de p_2

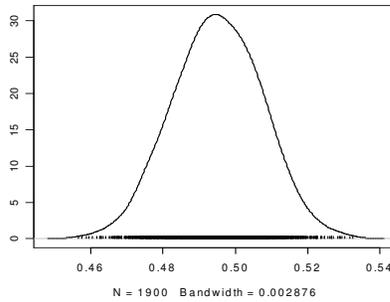


Figura 50: Densidade de p_3

As autocorrelações de p_1, p_2 e p_3 são mostradas nas Figuras 51, 52 e 53, respectivamente.

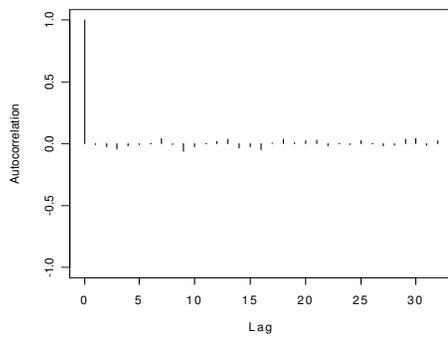


Figura 51: Autocorrelação de p_1

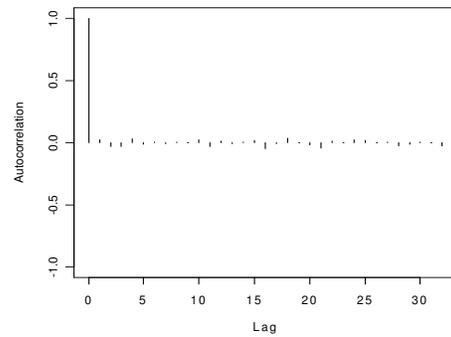


Figura 52: Autocorrelação de p_2

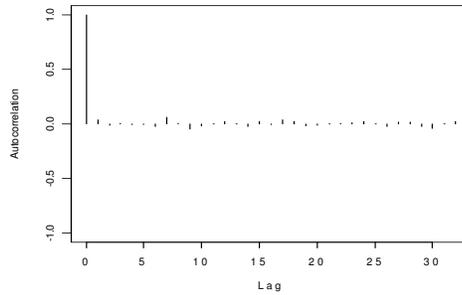


Figura 53: Autocorrelação de p_3

O quarto modelo utilizado para o ajuste do segundo conjunto de dados gerados é o modelo com mistura de distribuições binomiais com quatro componentes. Os resultados da simulação são vistos através da Tabela 12. Como na Tabela 8, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 12: *Resumo a posteriori dos parâmetros - quatro componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,2297	0,11607	0,02885	0,16346	0,21154	0,27885	0,50962
γ_2	0,2560	0,14331	0,02885	0,16346	0,24038	0,31731	0,51923
γ_3	0,2920	0,16233	0,02885	0,18269	0,25000	0,48077	0,51923
γ_4	0,2222	0,10477	0,03846	0,18269	0,21154	0,26923	0,50000
p_1	0,4993	0,15107	0,2878	0,3273	0,4939	0,6840	0,7209
p_2	0,4614	0,15394	0,2829	0,3100	0,4793	0,5313	0,7184
p_3	0,4445	0,17372	0,28580	0,30494	0,32219	0,68418	0,71952
p_4	0,5365	0,14622	0,2914	0,4722	0,5050	0,6962	0,7221

Através da Tabela 12 podemos observar que a média dos gamas varia de 0,2222 à 0,2920, ou seja, as médias permaneceram entre 0,2 e 0,3 e nenhum gama se aproxima de 0,5, como visto anteriormente pelos resultados das estimativas através da Tabela 8. Análogo ao caso com o primeiro conjunto de dados, ajustamos o modelo com mistura com quatro componentes aos dados gerados, e as estimativas das ponderações para cada componente da mistura novamente não foram satisfatórias visto que os dados foram gerados a partir do modelo com mistura com três componentes.

Em relação aos p 's, note que a média de nenhum p se aproxima de 0,3 e de 0,7, porém temos que as médias que mais se aproximam de 0,5 é a média de p_1 e a média de p_4 . Os resultados das estimativas não foram satisfatórios devido também ao ajuste do modelo com mistura com quatro componentes, exatamente como no caso das ponderações.

Nas Figuras 54, 55, 56 e 57 a seguir mostramos os gráficos das iterações de p_1 , p_2 , p_3 e p_4 , respectivamente.

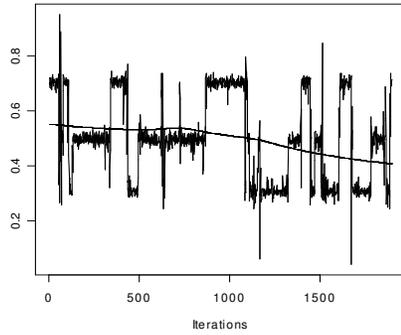


Figura 54: Iterações de p_1

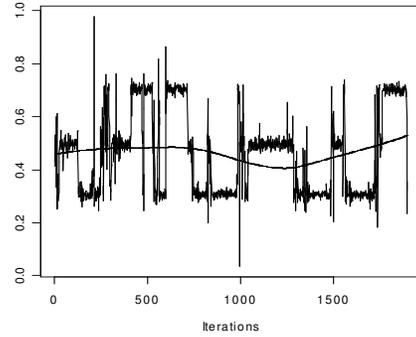


Figura 55: Iterações de p_2

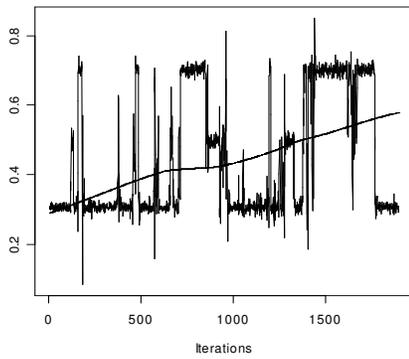


Figura 56: Iterações de p_3

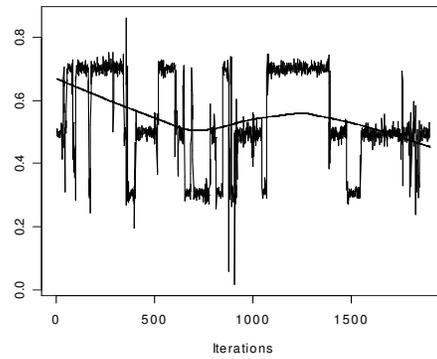


Figura 57: Iterações de p_4

As Figuras 58, 59, 60 e 61 mostram os gráficos das densidades de p_1 , p_2 , p_3 e p_4 , respectivamente.

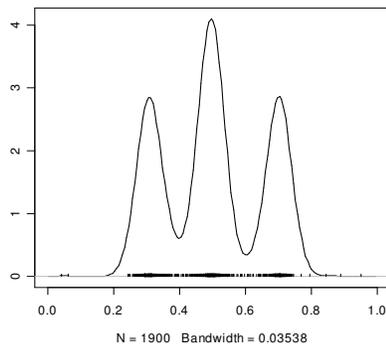


Figura 58: Densidade de p_1

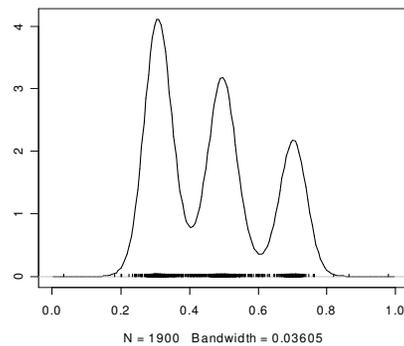


Figura 59: Densidade de p_2

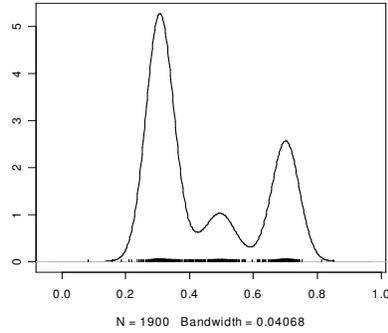


Figura 60: Densidade de p_3

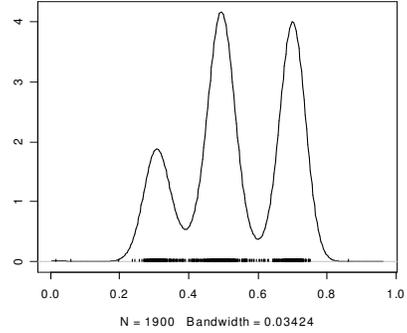


Figura 61: Densidade de p_4

As Figuras 62, 63, 64 e 65 mostram os gráficos das autocorrelações de p_1 , p_2 , p_3 e p_4 , respectivamente.

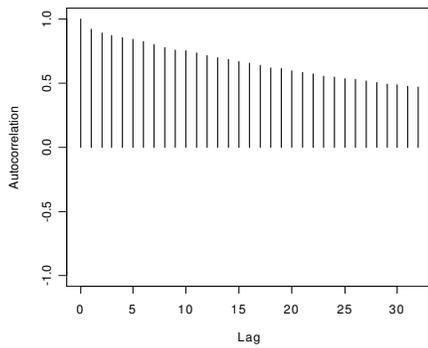


Figura 62: Autocorrelação de p_1

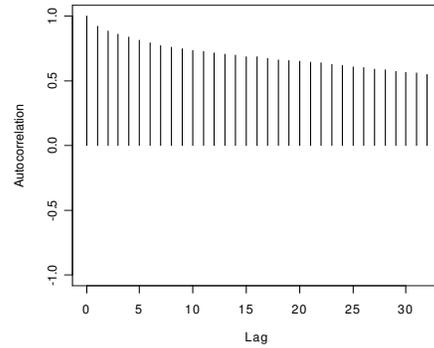


Figura 63: Autocorrelação de p_2

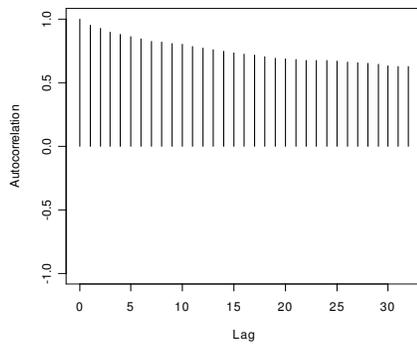


Figura 64: Autocorrelação de p_3

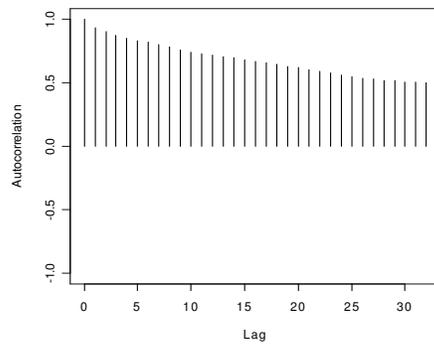


Figura 65: Autocorrelação de p_4

7.4.3 3º conjunto de dados

O terceiro conjunto de dados utilizado para análise são os dados gerados através de mistura de distribuições binomiais com 3 componentes com parâmetros $(100; 0, 4)$, $(100; 0, 5)$ e $(100; 0, 6)$. Através da distribuição binomial com parâmetros $(100; 0, 4)$ geramos 50 observações, através da distribuição binomial com parâmetros $(100; 0, 5)$ geramos 30 observações e através da distribuição binomial $(100; 0, 6)$ geramos 20 observações. A Figura 66 mostra o gráfico deste conjunto de dados.

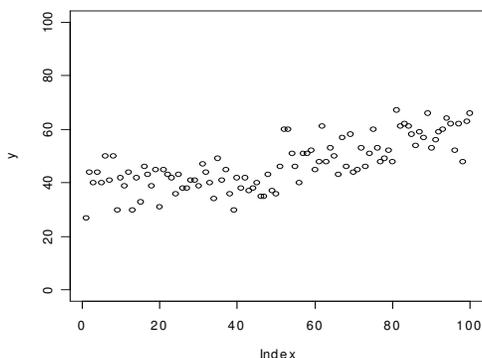


Figura 66: Dados gerados $p = (0, 4; 0, 5; 0, 6)$

Ajustamos o terceiro conjunto de dados gerados utilizando o modelo com distribuição binomial com parâmetros (n_i, p) . Os resultados podem ser vistos através da Tabela 13. Na primeira coluna da tabela temos o parâmetro estimado, ou seja, p . Na segunda e na terceira coluna da tabela temos a média e o desvio de p . Nas demais colunas temos os quantis de p .

Tabela 13: *Resumo a posteriori de p*

Parâmetro	Média	Desvio	2,5%	25%	50%	75%	97,5%
p	0,5023	0,0129	0,4879	0,4949	0,5036	0,5048	0,5060

Através da Tabela 13 podemos observar que a média de p também se aproxima de 0,5, como estes dados foram gerados a partir de mistura de distribuições binomiais com parâmetros e cada componente da mistura com parâmetros 0,4, 0,5 e 0,6, se calcularmos a média destes três parâmetros o resultado é 0,5, como nos resultados dos ajustes com os dois conjuntos de dados anteriores dados pelas Tabelas 5 e 6, assim a estimativa da média também se aproximou da média dos parâmetros das componentes.

O segundo modelo utilizado no ajuste do terceiro conjunto de dados é o modelo com mistura de distribuições binomiais com dois componentes. Os resultados dessa simulação podem ser vistos através da Tabela 14. Análogo a Tabela 10, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 14: Resumos a posteriori dos parâmetros - dois componentes

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,33720	0,03860	0,2647	0,3137	0,3333	0,3627	0,4216
γ_2	0,66279	0,03859	0,5784	0,6373	0,6667	0,6863	0,7353
p_1	0,57608	0,01260	0,5520	0,5676	0,5761	0,5842	0,6007
p_2	0,41712	0,00807	0,4009	0,4115	0,4172	0,4228	0,4326

Através da Tabela 14 podemos observar que as estimativas das médias dos gamas não são satisfatórias devido ao ajuste ao modelo com mistura com dois componentes, como vimos anteriormente no caso do primeiro e do segundo conjunto de dados. Observe que a média de p_1 permanece entre 0,5 e 0,6 e a média de p_2 se aproxima de 0,4. Estes p 's foram estimados de acordo com os dados gerados pela mistura de distribuições binomiais com três componentes com probabilidades 0,4, 0,5 e 0,6, então, não podemos afirmar que p_2 identificou a probabilidade de uma das componente da mistura, pois como visto antes, esta aproximação pode ter ocorrido ao acaso.

Nas Figuras 67 e 68 a seguir mostramos os gráficos das iterações de p_1 e p_2 , respectivamente.

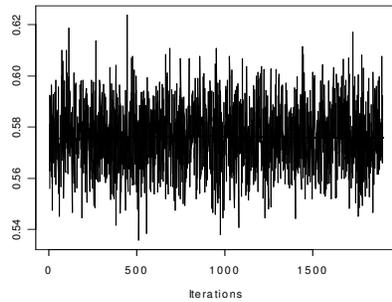


Figura 67: Iterações de p_1

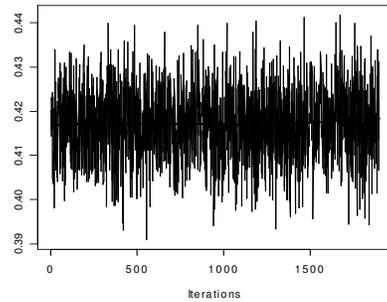


Figura 68: Iterações de p_2

Nas Figuras 69 e 70, mostramos os gráficos referente às densidades de p_1 e p_2 , respectivamente.

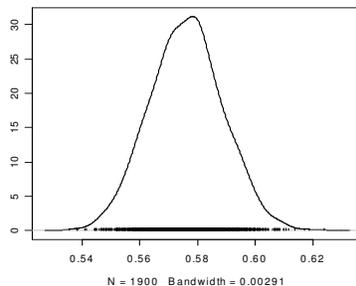


Figura 69: Densidade de p_1

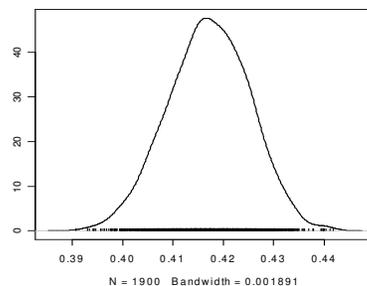


Figura 70: Densidade de p_2

Os gráficos referentes às autocorrelações de p_1 e p_2 são mostrados nas Figuras 71 e 72, respectivamente.

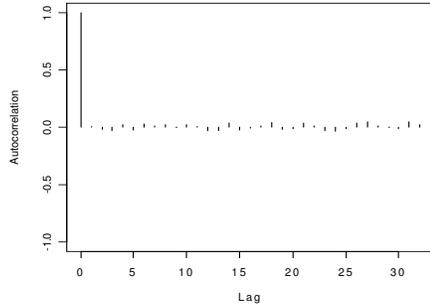


Figura 71: Autocorrelação de p_1

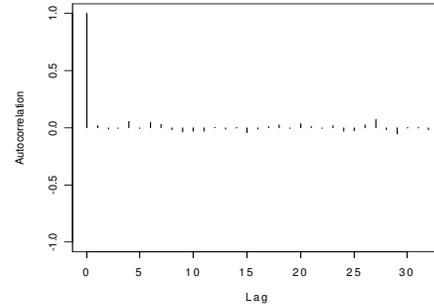


Figura 72: Autocorrelação de p_2

O terceiro modelo utilizado para o ajuste do terceiro conjunto de dados gerados é o modelo com mistura de distribuições binomiais com três componentes. Podemos observar os resultados da simulação através da Tabela 15 Análogo à Tabela 11, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 15: *Resumo a posteriori dos parâmetros - três componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,3756	0,1640	0,0874	0,2528	0,3495	0,5267	0,6505
γ_2	0,2992	0,1242	0,0874	0,2333	0,2815	0,3398	0,6019
γ_3	0,3251	0,1540	0,0680	0,2134	0,2912	0,4369	0,6310
p_1	0,4539	0,0798	0,3157	0,3956	0,4487	0,5722	0,6345
p_2	0,5111	0,0962	0,3290	0,4027	0,5048	0,6245	0,7036
p_3	0,4498	0,0905	0,3104	0,3714	0,4256	0,5012	0,5962

Note que com as médias dos p 's ocorre o mesmo. A média que se aproxima de uma das probabilidades, 0,5, é a média de p_2 . As outras médias permanecem entre 0,4 e 0,5 não satisfazendo nossas expectativas, visto que estamos ajustando a um modelo que possui o mesmo número de componentes dos dados gerados.

Nas Figuras 73, 74 e 75 mostramos os gráficos das iterações de p_1 , p_2 e p_3 , respectivamente.

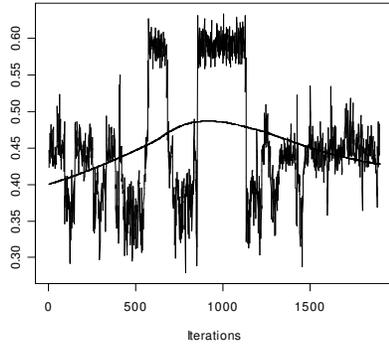


Figura 73: Iterações de p_1

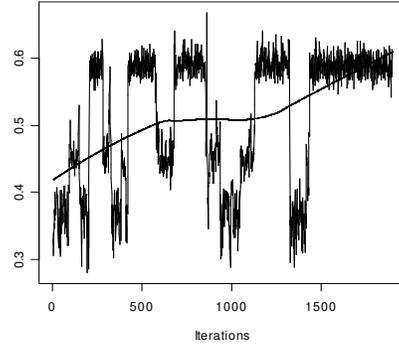


Figura 74: Iterações de p_2

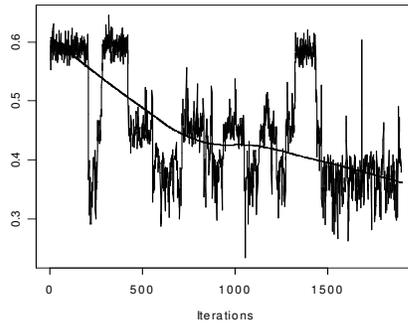


Figura 75: Iterações de p_3

As Figuras 76, 77 e 78 mostram os gráficos das densidades de p_1, p_2 e p_3 , respectivamente.

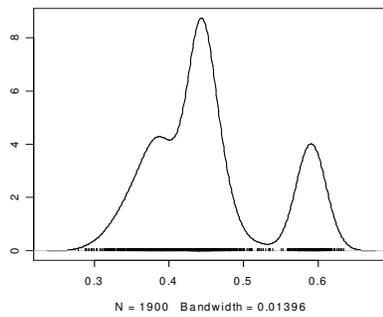


Figura 76: Densidade de p_1

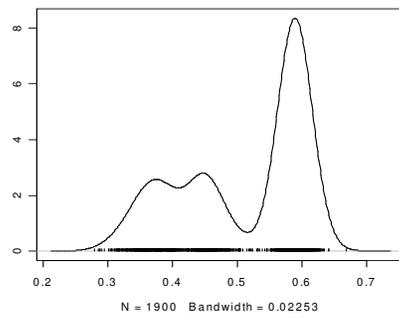


Figura 77: Densidade de p_2

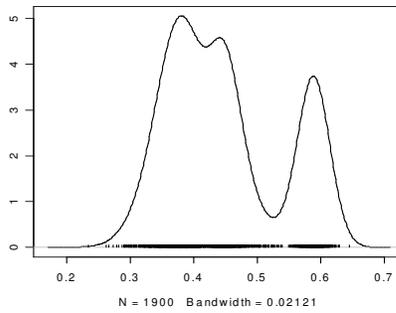


Figura 78: Densidade de p_3

Os gráficos das autocorrelações de p_1, p_2 e p_3 são mostrados nas Figuras 79, 80 e 81, respectivamente.

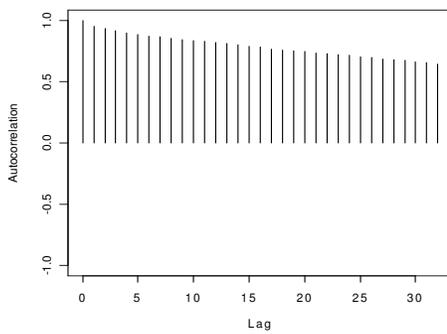


Figura 79: Autocorrelação de p_1

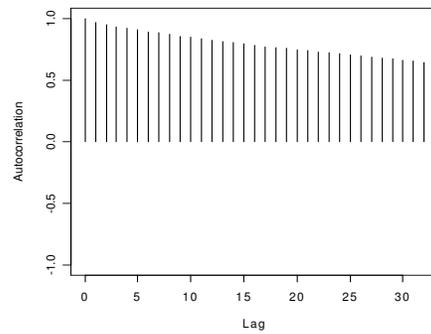


Figura 80: Autocorrelação de p_2

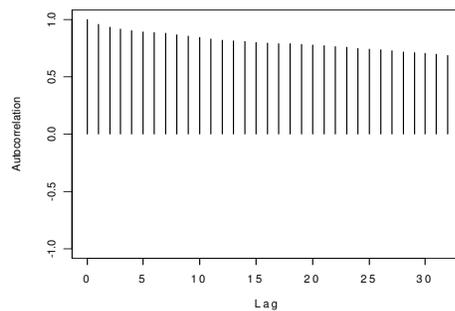


Figura 81: Autocorrelação de p_3

O quarto modelo utilizado para o ajuste do terceiro conjunto de dados gerados é o modelo com mistura de distribuições binomiais com quatro componentes. Os resultados da simulação são vistos através da Tabela 16. Como na Tabela 12, novamente temos na primeira coluna da tabela

os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 16: *Resumo a posteriori dos parâmetros - quatro componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,2557	0,1294	0,0384	0,1538	0,2500	0,3557	0,5000
γ_2	0,2527	0,1295	0,0288	0,1538	0,2500	0,3485	0,5096
γ_3	0,2443	0,1284	0,0288	0,1442	0,2404	0,3365	0,4954
γ_4	0,2472	0,1275	0,0288	0,1442	0,2404	0,3461	0,5000
p_1	0,4891	0,0754	0,3725	0,4146	0,5001	0,5490	0,6178
p_2	0,4944	0,0759	0,3781	0,4224	0,5055	0,5501	0,6202
p_3	0,4834	0,0806	0,3716	0,4109	0,4889	0,5446	0,6248
p_4	0,4856	0,0767	0,3727	0,4170	0,4919	0,5434	0,6199

Através da Tabela 16 podemos observar que a média dos gamas varia de 0,2443 à 0,2557, isto é, todas as médias permanecem entre 0,2 e 0,3, e assim nenhuma média de gama se aproxima de 0,5, 0,2 ou de 0,3. Todas as médias dos p 's se aproximam de 0,5. Novamente como no caso anterior, a geração dos dados é realizada com probabilidades muito próximas, e assim não é possível identificar a qual componente pertence cada observação. Os resultados das estimativas não foram satisfatórios devido também ao ajuste do modelo com mistura com quatro componentes, visto que os dados são gerados de um modelo com três componentes.

Nas Figuras 82, 83, 84 e 85 a seguir mostramos os gráficos das iterações de p_1 , p_2 , p_3 e p_4 , respectivamente.

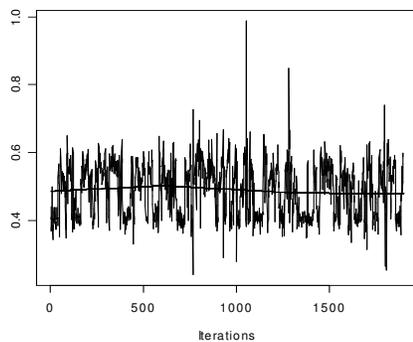


Figura 82: Iterações de p_1

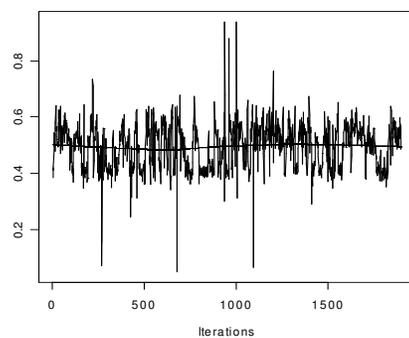


Figura 83: Iterações de p_2

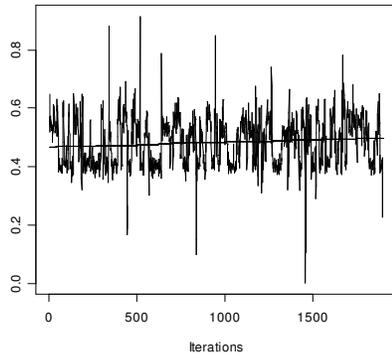


Figura 84: Iterações de p_3

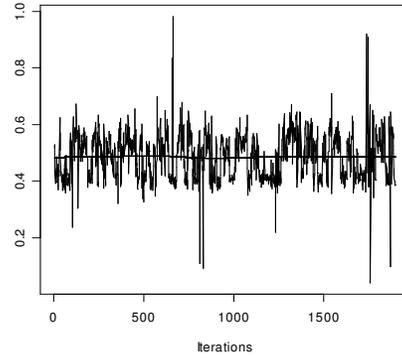


Figura 85: Iterações de p_4

As Figuras 86, 87, 88 e 89 mostram os gráficos das densidades de p_1 , p_2 , p_3 e p_4 , respectivamente.

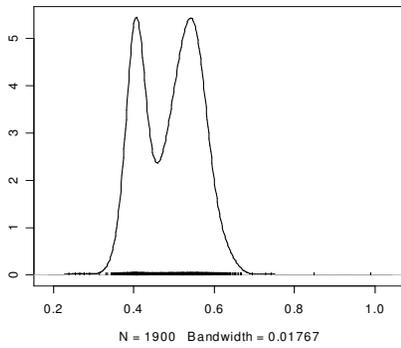


Figura 86: Densidade de p_1

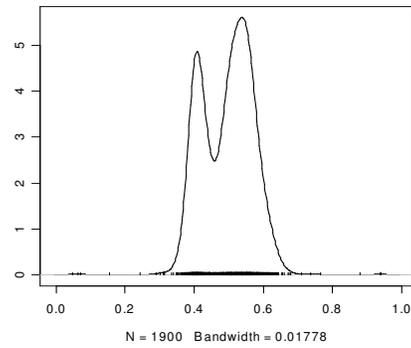


Figura 87: Densidade de p_2

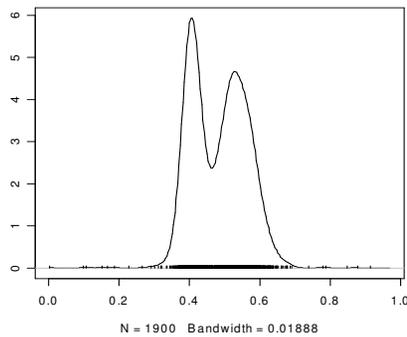


Figura 88: Densidade de p_3

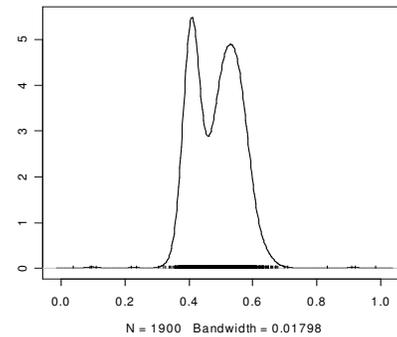


Figura 89: Densidade de p_4

As Figuras 90, 91, 92 e 93 mostram os gráficos das autocorrelações de p_1 , p_2 , p_3 e p_4 , respectivamente.

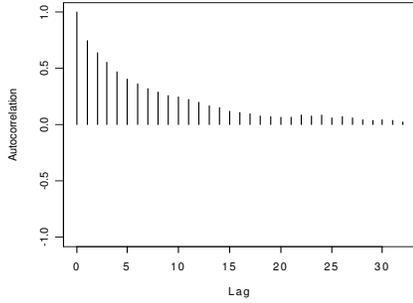


Figura 90: Autocorrelação de p_1

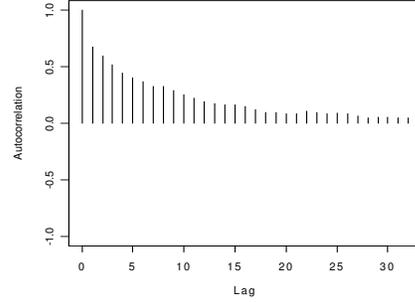


Figura 91: Autocorrelação de p_2

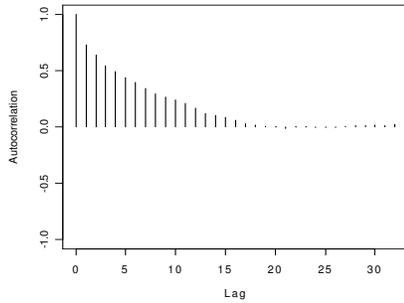


Figura 92: Autocorrelação de p_3

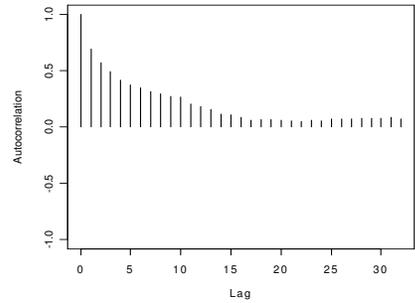


Figura 93: Autocorrelação de p_4

7.4.4 Seleção de Modelos

O próximo passo é utilizar o fator de Bayes como ferramenta para a seleção de modelos. Nas aplicações deste capítulo foram utilizados quatro modelos que denotamos por M_1 : modelo com distribuição binomial, M_2 : modelo com mistura com dois componentes, M_3 : modelo com mistura com três componentes e M_4 : modelo com mistura com quatro componentes.

Pela teoria do fator de Bayes vista no capítulo 2, temos que este é calculado através da expressão

$$B_{ij} = \frac{l_i}{l_j} = \frac{\int pr(y|\theta_i, M_i)\pi(\theta_j|M_j) d\theta_i}{\int pr(y|\theta_j, M_j)\pi(\theta_i|M_i) d\theta_j}, \text{ para } i, j = 1, 2, 3, 4, i \neq j$$

isto é, a densidade preditiva dado o modelo M_i em relação à densidade preditiva dado o modelo M_j . Se o resultado deste cálculo for maior que 1, o modelo escolhido é o modelo do numerador, M_i , se o resultado deste cálculo for menor que 1, o modelo escolhido é o modelo do denominador, M_j .

A Tabela 17 mostra os valores obtidos do cálculo de $I_i = \int pr(y|\theta_i, M_i)\pi(\theta_i|M_i) d\theta_i$, para $i = 1, 2, 3, 4$ para os três conjuntos de dados.

Tabela 17: Valores de I_i

Dados	I_i			
	I_1	I_2	I_3	I_4
1°	0,00518	0,00000086	0,1383	0,1328
2°	0,0879	0,0307	0,1793	0,1689
3°	0,1269	0,2273	0,2522	0,2531

A Tabela 18 mostra os valores dos fatores de Bayes calculados para os modelos utilizados para os três conjuntos de dados.

Tabela 18: Valores dos fatores de Bayes

B_{ij}	Dados		
	1°	2°	3°
B_{12}	6023,25	2,8632	0,5583
B_{13}	0,00374	0,4902	0,5031
B_{14}	0,03900	0,5204	0,5013
B_{23}	0,0000062	0,1712	0,9012
B_{24}	0,0000064	0,1817	0,8980
B_{34}	1,0414	1,061	0,9964

Através da Tabela 18, para o 1° e o 2° conjunto de dados, note que quando comparamos o modelo M_1 versus M_2 o modelo escolhido é M_1 . Na comparação entre M_1 versus M_3 o modelo escolhido é M_3 . Na comparação de M_1 versus M_4 , o modelo escolhido é M_4 . Comparando M_2 com os demais modelos, M_2 sempre é descartado. Na comparação de M_3 versus M_4 , o modelo escolhido é M_3 . Assim, podemos observar que o modelo escolhido para ambos os conjuntos de dados é o modelo com mistura com três componentes, ou seja, M_3 . O modelo menos favorecido foi o modelo M_2 , o qual denota o modelo com mistura com dois componentes. O fator de Bayes no primeiro conjunto de dados favorece fortemente o modelo M_3 em relação aos modelos M_1 e M_2 pelo fato de estarmos trabalhando com um conjunto de dados gerado através de componentes com um intervalo grande entre as probabilidades, assim o fator de Bayes identifica o melhor modelo com precisão.

Para o terceiro conjunto de dados observamos que quando comparamos o modelo M_1 versus M_2 , o modelo escolhido é M_2 . Na comparação entre M_1 versus M_3 o modelo escolhido é M_3 . Na comparação de M_1 versus M_4 , o modelo escolhido é M_4 . Comparando M_2 com os modelos M_3 e M_4 , M_2 sempre é descartado. Na comparação de M_3 versus M_4 , o modelo escolhido é M_4 . Assim, modelo escolhido para o 3° conjunto de dados é M_4 . De fato, os valores resultantes do cálculo de I_3 e I_4 para os modelos M_3 e M_4 , respectivamente são bem próximos, no entanto o cálculo do fator de Bayes privilegia o modelo com mistura com quatro componentes. Isto se deve ao fato do conjunto de dados ser gerado através de componentes com probabilidades muito próximas, dificultando o fator de Bayes na identificação do modelo verdadeiro.

Contudo, para que o fator de Bayes nos forneça resultados precisos, existe a pré condição de convergência, sem a qual não temos as estimativas dos parâmetros. Portanto, neste caso devemos descartar os modelos M_3 e M_4 , o que nos levaria ao modelo M_2 que também não é o correto.

O fator de Bayes fracional também é utilizado para ilustrar a aplicação da metodologia apresentada no capítulo 4.

Vimos que o fator de Bayes fracional é dado pela expressão

$$B_{ij}^b = \frac{I_i^b}{I_j^b} = \frac{pr(Y, b|H_i)}{pr(Y, b|H_j)} = \frac{\int pr(Y|\theta_i)c_i g_i(\theta_i)d\theta_i}{\int \{pr(Y|\theta_i)\}^b c_i g_i(\theta_i)d\theta_i} \cdot \frac{\int pr(Y|\theta_j)c_j g_j(\theta_j)d\theta_j}{\int \{pr(Y|\theta_j)\}^b c_j g_j(\theta_j)d\theta_j}, \text{ para } i, j = 1, 2, 3, 4, \quad (7.7)$$

onde $b = m/n$, sendo m o tamanho da amostra de treinamento e n o tamanho da amostra total. Nesta aplicação os valores utilizados para b são 0, 1, 0, 15, 0, 2, 0, 25, 0, 3 e 0, 35. Na Tabela 19 obtemos os valores resultantes do cálculo de I_i^b , para $i = 1, 2, 3, 4$, referente ao primeiro conjunto de dados.

Tabela 19: Valores de I_i^b para o 1º conj. de dados

I_i^b	b					
	0, 1	0, 15	0, 2	0, 25	0, 3	0, 35
I_1^b	0, 0035	0, 0187	0, 6318	0, 7823	0, 9865	0, 9937
I_2^b	0, 0000905	0, 0104	0, 6071	0, 9962	0, 9995	0, 9995
I_3^b	0, 0188	0, 0229	0, 0279	0, 0342	0, 0419	0, 0515
I_4^b	0, 0183	0, 0225	0, 0277	0, 0341	0, 0410	0, 0528

A Tabela 20 mostra os valores resultantes do cálculo de I_i^b , para $i = 1, 2, 3, 4$, referente ao segundo conjunto de dados.

Tabela 20: Valores de I_i^b para o 2º conj. de dados

I_i^b	b					
	0, 1	0, 15	0, 2	0, 25	0, 3	0, 35
I_1^b	0, 0189	0, 0211	0, 1410	0, 2872	0, 4712	0, 5214
I_2^b	0, 0120	0, 0289	0, 3340	0, 5798	0, 6546	0, 6716
I_3^b	0, 0251	0, 0305	0, 0372	0, 0453	0, 0552	0, 0675
I_4^b	0, 0245	0, 0301	0, 0368	0, 0450	0, 0551	0, 0674

E na Tabela 21 mostramos os valores resultantes do cálculo de I_i^b , referente ao terceiro conjunto de dados.

Tabela 21: Valores de I_i^b para o 3º conj. de dados

I_i^b	b					
	0, 1	0, 15	0, 2	0, 25	0, 3	0, 35
I_1^b	0, 0327	0, 0365	0, 0547	0, 0914	0, 1328	0, 1678
I_2^b	0, 0304	0, 0336	0, 0768	0, 1121	0, 1758	0, 2359
I_3^b	0, 0341	0, 0406	0, 0486	0, 0581	0, 0696	0, 0834
I_4^b	0, 0332	0, 0389	0, 0457	0, 0531	0, 0668	0, 0821

Os resultados do cálculo dos fatores de Bayes fracionais, B_{ij}^b , para $i, j = 1, 2, 3, 4$, podem ser vistos nas Tabelas 22, 23 e 24.

Na Tabela 22 apresentamos os valores obtidos do cálculo do fator de Bayes para o primeiro conjunto de dados.

Tabela 22: Valores dos fatores de Bayes fracionais para o 1º conj. de dados

B_{ij}^b	b					
	0,1	0,15	0,2	0,25	0,3	0,35
B_{12}^b	38,67	1,7980	1,040	0,7853	0,9870	0,9942
B_{13}^b	0,1862	0,8166	22,64	22,87	23,54	19,29
B_{14}^b	0,1912	0,8311	22,81	22,94	24,06	18,82
B_{23}^b	0,0048	0,4541	21,75	29,12	24,85	19,40
B_{24}^b	0,0049	0,4622	21,91	29,21	24,37	18,93
B_{34}^b	1,027	1,0177	1,007	1,002	1,021	0,9753

Através da Tabela 22 é possível observar que o modelo verdadeiro, M_3 , é escolhido apenas quando $b = 0,1$ e $b = 0,15$, quando $b = 0,2$ o modelo escolhido é o modelo com distribuição binomial, M_1 , e quando $b > 0,2$ o modelo escolhido é o modelo com mistura com dois componentes. Com o aumento da amostra de treinamento o cálculo do fator de Bayes favorece um modelo falso. Quando comparamos o modelo M_3 versus o modelo M_4 , o modelo M_3 é o escolhido até $b = 0,3$, e para $b > 0,3$, o fator de Bayes favorece o modelo M_4 .

Na Tabela 23 apresentamos os valores obtidos do cálculo do fator de Bayes para o segundo conjunto de dados.

Tabela 23: Valores do fator de Bayes fracional para o 2º conj. de dados

B_{ij}^b	b					
	0,1	0,15	0,2	0,25	0,3	0,35
B_{12}^b	1,5750	0,7301	0,4221	0,4953	0,7198	0,7763
B_{13}^b	0,7530	0,6918	3,7900	6,3400	8,5362	7,7244
B_{14}^b	0,7714	0,7010	3,8315	6,3822	8,5517	7,7359
B_{23}^b	0,4780	0,9475	8,978	12,80	11,85	9,950
B_{24}^b	0,4897	0,9601	9,076	12,88	11,88	9,964
B_{34}^b	1,0244	1,0133	1,010	1,006	1,001	1,001

Observe que o modelo escolhido continua sendo M_3 apenas quando $b = 0,1$ e $b = 0,15$ e quando $b > 0,15$ o modelo escolhido é o modelo com mistura de dois componentes. Quando comparamos o modelo M_3 versus o modelo M_4 , o modelo M_3 é o modelo escolhido independente do valor de b .

Na Tabela 24 apresentamos os valores obtidos do cálculo do fator de Bayes para o terceiro conjunto de dados.

Tabela 24: Valores do fator de Bayes fracional para o 3º conj. de dados

B_{ij}^b	b					
	0,1	0,15	0,2	0,25	0,3	0,35
B_{12}^b	1,0756	1,0863	0,7122	0,8153	0,7554	0,7113
B_{13}^b	0,9589	0,8990	1,1255	1,5731	1,9080	2,0120
B_{14}^b	0,9849	0,9383	1,1969	1,7213	1,9880	2,0438
B_{23}^b	0,8915	0,8275	1,5802	1,9294	2,5258	2,8285
B_{24}^b	0,9156	0,8637	1,6805	2,111	2,6317	2,8733
B_{34}^b	1,027	1,0437	1,0634	1,0941	1,0403	1,0158

Pela Tabela 24 podemos verificar que novamente o modelo M_3 é o escolhido até $b = 0,15$, para valores de $b > 0,15$, o modelo escolhido é o M_2 , e na comparação de M_3 versus M_4 , M_3 continua sendo escolhido independente do valor de b .

7.5 Discussão

Os resultados das simulações indicam que quando trabalhamos com modelos com misturas conjuntamente com MCMC, a dificuldade na obtenção de convergência ou mesmo a não convergência pode estar associada à escolha do modelo. Através dos resultados podemos observar que os casos de não convergência ocorreram preferencialmente para modelos diferentes daquele utilizado para gerar os dados. Veja por exemplo, quando ajustamos um modelo com mistura com 4 componentes no caso dos três conjuntos de dados gerados, a convergência não ocorreu.

Outro resultado interessante que as simulações apontam é que quando utilizamos o fator de Bayes fracional o valor de b a ser escolhido deve ser menor que 0,15 para os modelos considerados aqui.

Em geral o fator de Bayes e o fator de Bayes fracional (com $b \leq 0,15$) funcionam adequadamente na seleção de modelos com misturas.

A tendência é que o fator de Bayes e o fator de Bayes fracional funcionem melhor quanto mais afastados forem os parâmetros das binomiais envolvidas na mistura e com a performance sendo reduzida a medida que a diferença entre estes parâmetros diminua.

Capítulo 8

Aplicação: Modelos com Mistura

Neste capítulo apresentamos um exemplo com dados reais, os mesmos utilizados no capítulo 6, seção 6.2. Para analisarmos a performance de um jogador de basquete, é utilizada a abordagem Bayesiana com mistura de modelos.

O fator de Bayes é utilizado para identificar o número de componentes nos modelos com mistura, incluindo também o modelo que considera todas as probabilidades de sucesso iguais e a probabilidade *a posteriori* para cada modelo.

Novamente consideramos o total de arremessos realizados e o número de acertos em cada jogo.

No capítulo 6, na aplicação com dados reais, verificamos se a probabilidade de acerto dos arremessos permanecia constante ou se a mesma variava no tempo. Para isso consideramos todas as probabilidades de acerto iguais *versus* todas as probabilidades de acerto diferentes. Nesta aplicação, ao invés de considerarmos todas as probabilidades de acerto diferentes, consideramos k probabilidades de acerto diferentes, com $k \geq 1$.

Na modelagem utilizada para o conjunto de dados, ou seja, mistura de distribuições binomiais, cada componente da mistura considera uma probabilidade de sucesso diferente.

Nesta aplicação consideramos novamente o total de arremessos, denotado por n_i , $i = 1, \dots, 380$, e o número de acertos (conversões), denotado por y_i , $i = 1, \dots, 380$.

8.0.1 Descrição dos modelos

Os modelos utilizados são M_1, M_2, \dots, M_8 ou seja, no modelo M_1 temos y_i com distribuição binomial com parâmetros (n_i, p) , ou seja, um modelo sem mistura ou com uma componente, no modelo M_2 temos mistura de distribuições binomiais com 2 componentes, no modelo M_3 temos mistura de distribuições binomiais com 3 componentes e no modelo M_4 temos mistura de distribuições binomiais com 4 componentes e assim sucessivamente.

Como no capítulo anterior, a função de verossimilhança referente ao modelo M_1 é dada pela equação (6.2), onde $\mathbf{n} = (n_1, \dots, n_m)$, $\mathbf{y} = (y_1, \dots, y_m)$ e $m = 100$. As funções de verossimilhança baseadas em modelos de mistura de distribuições binomiais são dadas de acordo com a equação (5.2), vista no capítulo 5.

Generalizando para o modelo M_k temos k probabilidades de acerto diferentes p_1, p_2, \dots, p_k , $k > 1$, então a função de verossimilhança do modelo M_k é dada por

$$L(p_1, p_2, \dots, p_k, \gamma_1, \gamma_2, \dots, \gamma_k | \mathbf{y}, \mathbf{n}, M_k) = \prod_{i=1}^{380} \sum_{j=1}^k \gamma_j \binom{n_i}{y_i} p_j^{y_i} (1 - p_j)^{n_i - y_i},$$

onde $\gamma_1, \gamma_2, \dots, \gamma_k$ são as k ponderações do modelo.

Introduzimos a variável \mathbf{Z}_i com a finalidade de simplificar a função de verossimilhança e conseqüentemente a densidade *a posteriori*. Assim, a função de verossimilhança para M_k baseada em variáveis latentes de acordo com a equação (5.4) é dada por

$$L(p_1, p_2, \dots, p_k, \gamma_1, \gamma_2, \dots, \gamma_k | \mathbf{y}, \mathbf{n}, M_k) = \prod_{i=1}^{380} \prod_{j=1}^k \left\{ \gamma_j \binom{n_i}{y_i} p_j^{y_i} (1 - p_j)^{n_i - y_i} \right\}^{Z_{ij}}$$

8.0.2 Abordagem Bayesiana

Como no capítulo anterior, a densidade *a priori* para p , do modelo M_1 , é dada pela equação (6.4), vista no capítulo 6.

A densidade *a priori* para γ e cada p_j de cada modelo M_k , são dadas respectivamente pelas equações, (7.1) e (7.2). A densidade *a posteriori* conjunta para $(\gamma_1, \gamma_2, \dots, \gamma_k, p_1, p_2, \dots, p_k)$ do modelo M_k é dada pela equação (7.3), e a densidade *a posteriori* condicional de γ , a distribuição *a posteriori* condicional de p_j e a distribuição *a posteriori* condicional de \mathbf{Z}_i referentes ao modelo M_k , são dadas respectivamente pelas equações (7.4), (7.5) e (7.6), considerando $i = 1, \dots, 380$, em todas as densidades.

Novamente o algoritmo utilizado para gerar amostras das distribuições *a posteriori* condicionais é o *Gibbs Sampling*, utilizado na seção anterior.

8.0.3 Seleção de Modelos

Para gerarmos amostras dos modelos M_1 e M_k , $k > 1$, utilizamos novamente a densidade *a posteriori* de p dada pela equação (6.6), a densidade *a posteriori* condicional de p_j dada pela equação (7.5), a densidade *a posteriori* condicional de γ dada pela equação (7.4) e a densidade *a posteriori* condicional de \mathbf{Z}_i dada pela equação (7.5), respectivamente. O procedimento utilizado para o cálculo do fator de Bayes é análogo ao da seção anterior. Consideramos o mesmo número de repetições do processo de simulação utilizado no capítulo anterior para obtenção do fator de Bayes, com os mesmos números para o *burn in* e para os saltos.

A Figura 3, vista no capítulo 6, mostra o gráfico destes dados reais.

O primeiro modelo a ser ajustado aos dados é o modelo com distribuição binomial, M_1 . Este modelo foi tratado no capítulo 6 e no capítulo 7 e seus resultados serão transferidos para efeito de seleção de modelos. Os resultados dessa simulação podem ser vistos através da Tabela 25. A primeira coluna da tabela temos o parâmetro estimado p que é a probabilidade da distribuição binomial. Na segunda e na terceira coluna da tabela obtemos a média e o desvio de p . Nas demais colunas apresentamos os quantis de p de 2, 5%, 25%, 50%, 75% e 97, 5%, respectivamente.

Tabela 25: *Resumo a posteriori de p*

Parâmetros	Média	Desvio	2, 5%	25%	50%	75%	97, 5%
p	0, 4568	0, 0070	0, 4436	0, 4520	0, 4567	0, 4613	0, 4710

De acordo com a Tabela 25, a média de p é 0, 4568.

O segundo modelo a ser ajustado aos dados é o modelo com mistura de distribuições binomiais com dois componentes, M_2 . Os resultados dessa simulação podem ser vistos através da Tabela 26. A primeira coluna da tabela temos os parâmetros estimados, ou seja, γ_1, γ_2 , que são as ponderações atribuídas a cada componente da mistura e p_1, p_2 são as probabilidades de cada distribuição binomial. Na segunda e na terceira coluna da tabela obtemos as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros de 2,5%, 25%, 50%, 75% e 97,5%, respectivamente.

Tabela 26: *Resumo a posteriori dos parâmetros - dois componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0.4257	0.2697	0,0157	0,1832	0,4201	0,6361	0,9162
γ_2	0.5742	0.2697	0,0837	0,3638	0,5798	0,8167	0,9843
p_1	0.4496	0.0554	0,3102	0,4381	0,4555	0,4700	0,5400
p_2	0.4555	0.0251	0,3991	0,4444	0,4565	0,4675	0,5064

Através da Tabela 26 podemos observar que a média de γ_1 é 0,4257 e a média de γ_2 é 0,5742. Estas estimativas são obtidas de acordo com os dados reais ajustando o modelo com mistura com dois componentes. A média de p_1 é 0,4496 e a média de p_2 é 0,4555.

Nas Figuras 94 e 95 a seguir mostramos os gráficos das iterações de p_1 e p_2 , respectivamente.

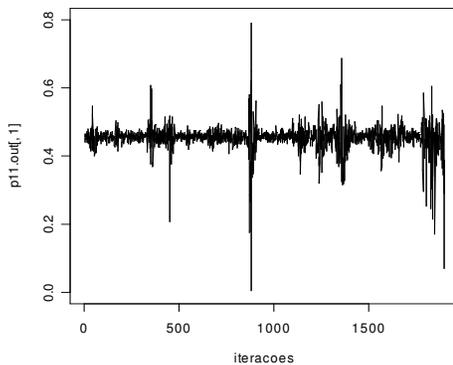


Figura 94: Iterações de p_1

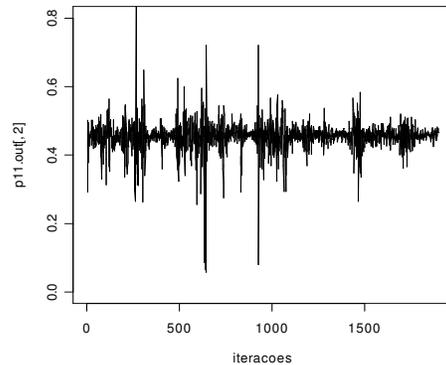


Figura 95: Iterações de p_2

As Figuras 96 e 97 mostram os gráficos das densidades de p_1 e p_2 , respectivamente.

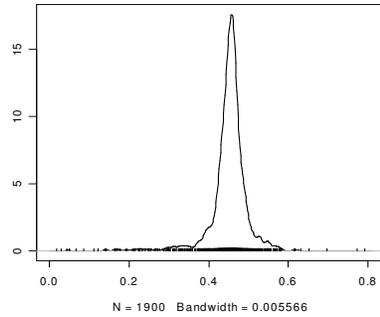


Figura 96: Densidade de p_1

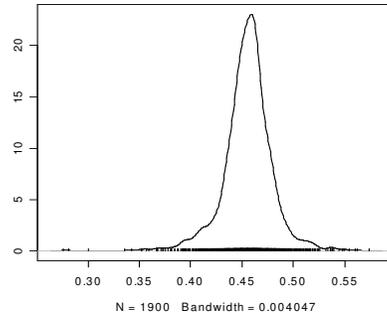


Figura 97: Densidade de p_2

Os gráficos das autocorrelações de p_1 e p_2 são apresentadas pelas Figuras 98 e 99, respectivamente.

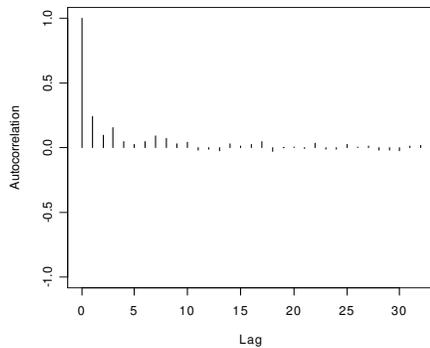


Figura 98: Autocorrelação de p_1

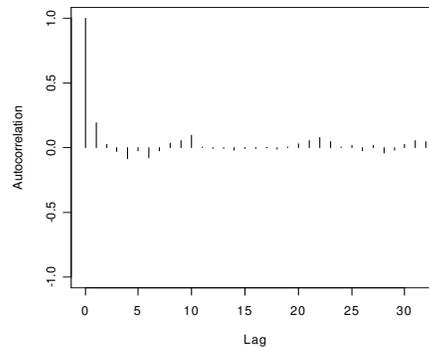


Figura 99: Autocorrelação de p_2

O terceiro modelo utilizado para o ajuste do conjunto de dados reais é o modelo com mistura de distribuições binomiais com três componentes. Podemos observar os resultados da simulação através da Tabela 27. Análogo à Tabela 26, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas apresentamos os quantis dos parâmetros.

Tabela 27: *Resumo a posteriori dos parâmetros - três componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,3157	0,2212	0,0130	0,1279	0,2794	0,4732	0,7937
γ_2	0,3648	0,2287	0,0247	0,1749	0,3420	0,5274	0,8369
γ_3	0,3194	0,2447	0,0157	0,1149	0,2689	0,4621	0,8877
p_1	0,4480	0,0634	0,3092	0,4324	0,4561	0,4756	0,5452
p_2	0,4504	0,0488	0,3434	0,4350	0,4548	0,4720	0,5283
p_3	0,4520	0,0612	0,3046	0,4329	0,4570	0,4768	0,5418

Através da Tabela 27 observamos que a média de γ_1 é 0,3157, a média de γ_2 é 0,3648 e a média de γ_3 é 0,3194. A média de p_1 é 0,4480, a média de p_2 é 0,4504 e a média de p_3 é 0,4520. Nas Figuras 100, 101 e 102 mostramos os gráficos das iterações de p_1 , p_2 e p_3 , respectivamente.

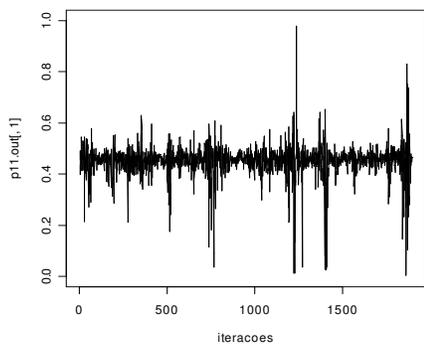


Figura 100: Iterações de p_1

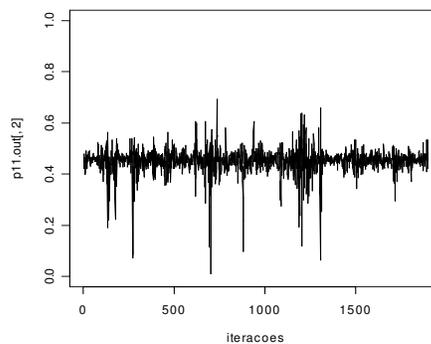


Figura 101: Iterações de p_2

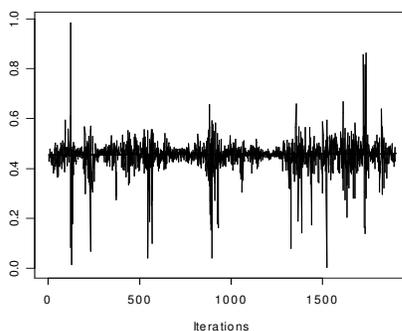


Figura 102: Iterações de p_3

As Figuras 103, 104 e 105 mostram os gráficos das densidades de p_1 , p_2 e p_3 , respectivamente.

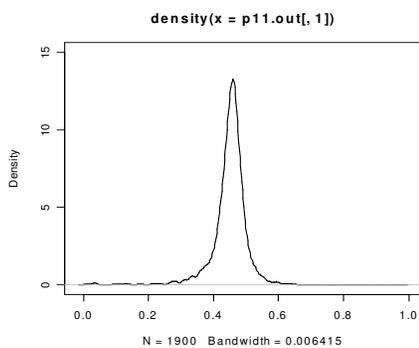


Figura 103: Densidade de p_1

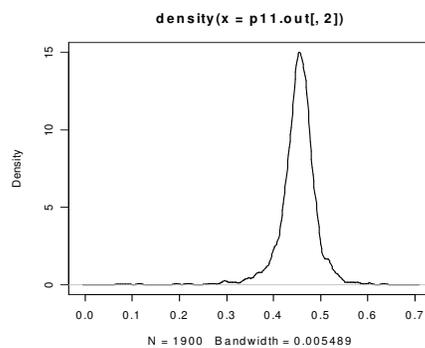


Figura 104: Densidade de p_2

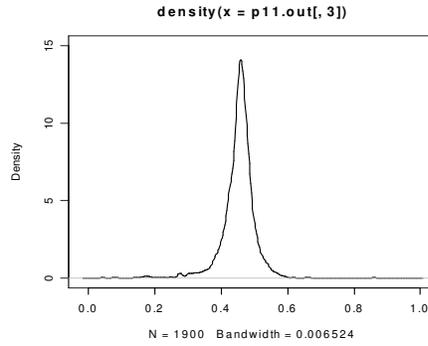


Figura 105: Densidade de p_3

Os gráficos das autocorrelações de p_1, p_2 e p_3 são mostrados nas Figuras 106, 107 e 108, respectivamente.

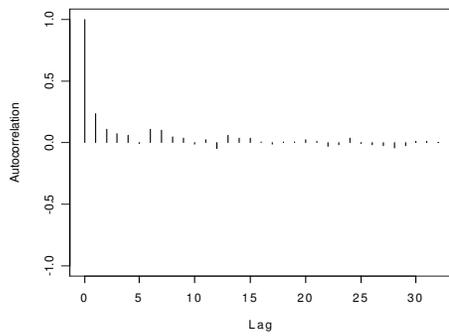


Figura 106: Autocorrelação de p_1

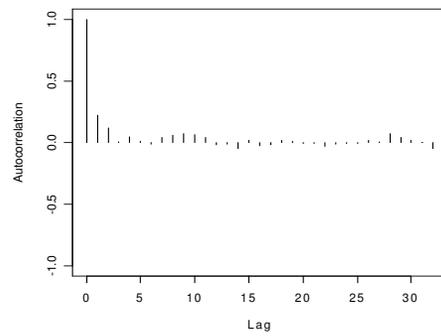


Figura 107: Autocorrelação de p_2

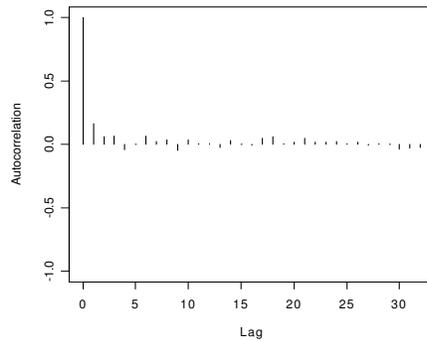


Figura 108: Autocorrelação de p_3

O quarto modelo utilizado para o ajuste do conjunto de dados reais é o modelo com mistura de distribuições binomiais com quatro componentes. Os resultados da simulação são vistos

através da Tabela 28. Como na Tabela 26, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas apresentamos os quantis dos parâmetros.

Tabela 28: *Resumo a posteriori dos parâmetros - quatro componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,2836	0,2040	0,0156	0,1093	0,2474	0,4199	0,7461
γ_2	0,2424	0,1972	0,0104	0,0807	0,1953	0,3463	0,7370
γ_3	0,2331	0,1793	0,0104	0,0859	0,1953	0,3437	0,6368
γ_4	0,2408	0,1750	0,0142	0,1015	0,0182	0,3489	0,6406
p_1	0,4532	0,0562	0,3339	0,4335	0,4568	0,4771	0,5468
p_2	0,4465	0,0666	0,2847	0,4248	0,4552	0,4788	0,5559
p_3	0,4493	0,0664	0,2879	0,4298	0,4563	0,4785	0,5599
p_4	0,4486	0,0608	0,3007	0,4294	0,4547	0,4771	0,5467

Através da Tabela 28 podemos observar que a média dos gamas varia de 0,2331 à 0,2836 e as médias dos p 's variam de 0,4465 à 0,4532.

Os gráficos referentes às iterações, densidades e autocorrelações de p_1, p_2, p_3 e p_4 , são similares aos gráficos das simulações anteriores.

O quinto modelo a ser ajustado aos dados é o modelo com mistura com cinco componentes. Os resultados da simulação são vistos através da Tabela 29. Como na Tabela 26, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela apresentamos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas apresentamos os quantis dos parâmetros.

Tabela 29: *Resumo a posteriori dos parâmetros - cinco componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,2302	0,1685	0,0104	0,0935	0,1948	0,3350	0,6247
γ_2	0,1854	0,1457	0,0104	0,0727	0,1532	0,2597	0,5758
γ_3	0,1904	0,1510	0,0104	0,0727	0,1506	0,2675	0,5766
γ_4	0,1719	0,1411	0,0078	0,0617	0,1350	0,2519	0,5220
γ_5	0,2218	0,1674	0,0130	0,0805	0,1818	0,3299	0,5935
p_1	0,4509	0,0656	0,3072	0,4317	0,4569	0,4788	0,5533
p_2	0,4479	0,0748	0,2760	0,4240	0,4544	0,4807	0,5623
p_3	0,4477	0,0691	0,2897	0,4213	0,4524	0,4793	0,5599
p_4	0,4480	0,0810	0,2477	0,4250	0,4550	0,4840	0,5703
p_5	0,4520	0,0623	0,3167	0,4312	0,4576	0,4596	0,5534

Através da Tabela 29 podemos observar que as médias dos gamas variam de 0,1719 à 0,2302, e as médias dos p 's variam de 0,4477 à 0,4520.

As Figuras 109, 110, 111, 112 e 113 mostram os gráficos das iterações referentes à p_1, p_2, p_3, p_4 e p_5 , respectivamente.

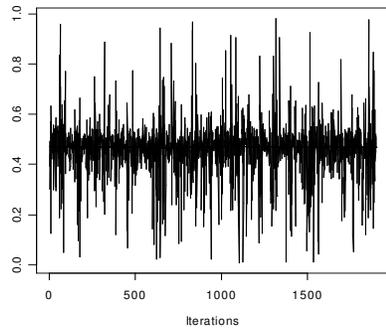


Figura 109: Iterações de p_1

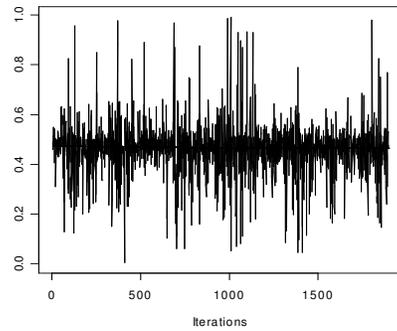


Figura 110: Iterações de p_2

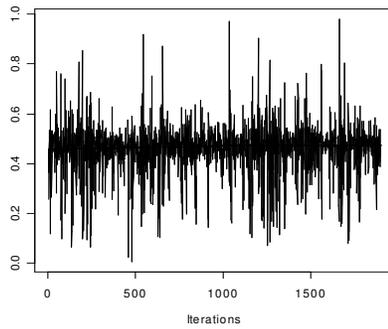


Figura 111: Iterações de p_3

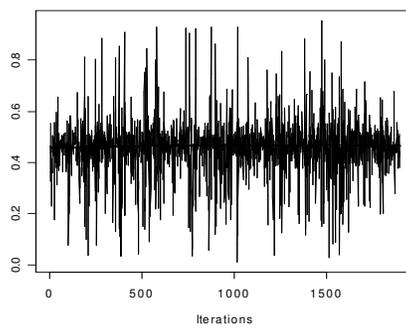


Figura 112: Iterações de p_4

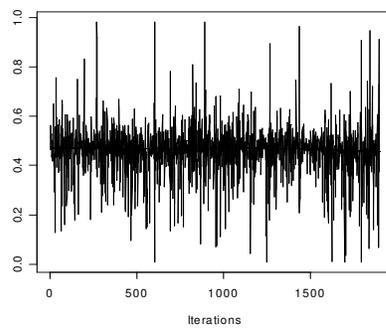


Figura 113: Iterações de p_5

As Figuras 114, 115, 116, 117 e 118 mostram os gráficos das densidades de p_1, p_2, p_3, p_4 e p_5 , respectivamente.

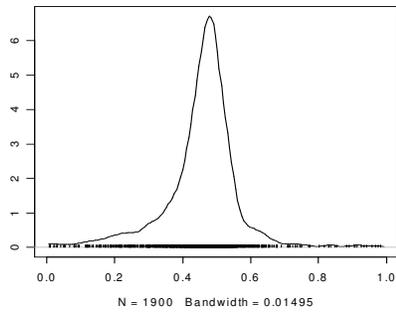


Figura 114: Densidade de p_1

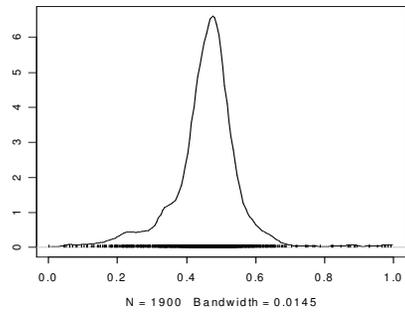


Figura 115: Densidade de p_2

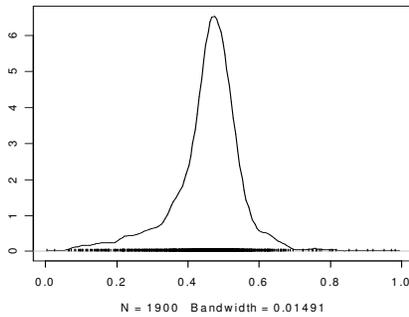


Figura 116: Densidade de p_3

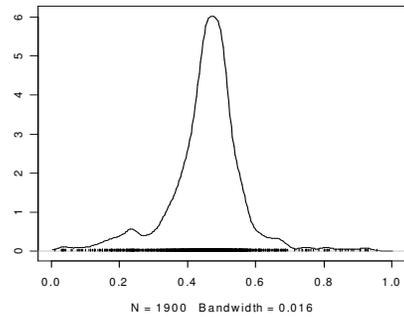


Figura 117: Densidade de p_4

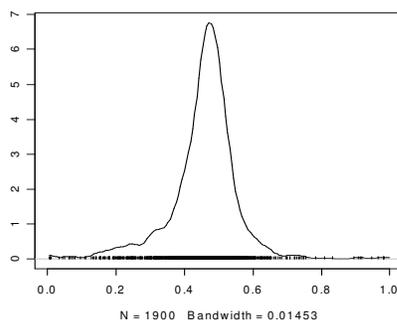


Figura 118: Densidade de p_5

As Figuras 119, 120, 121, 122 e 123 mostram os gráficos referentes às autocorrelações de p_1, p_2, p_3, p_4 e p_5 , respectivamente.

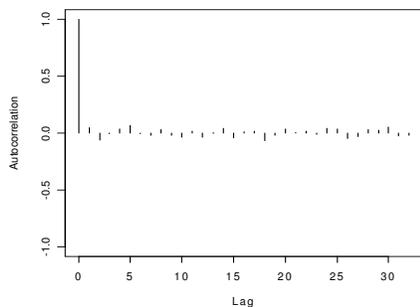


Figura 119: Autocorrelação de p_1

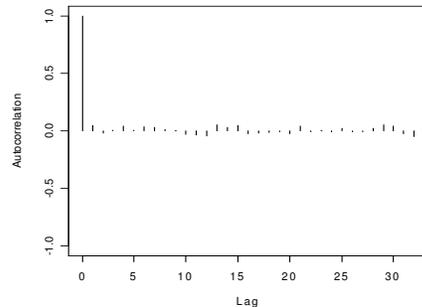


Figura 120: Autocorrelação de p_2

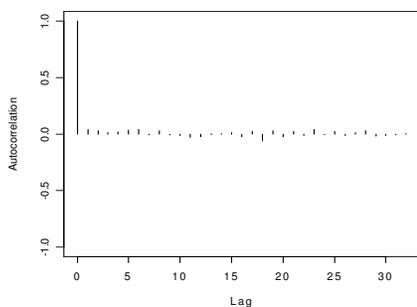


Figura 121: Autocorrelação de p_3

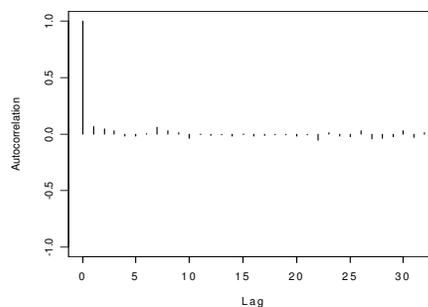


Figura 122: Autocorrelação de p_4

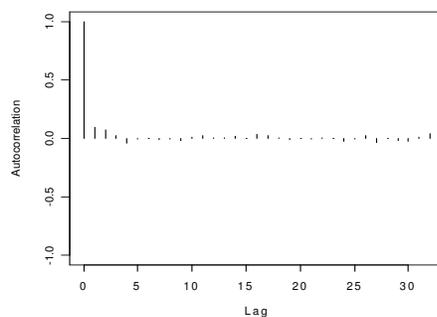


Figura 123: Autocorrelação de p_5

O sexto modelo a ser ajustado aos dados é o modelo com mistura com seis componentes. Os resultados da simulação são vistos através da Tabela 30. Como na Tabela 26, novamente temos na primeira coluna da tabela os parâmetros estimados. Na segunda e na terceira coluna da tabela obtemos respectivamente as médias e os desvios desses parâmetros. Nas demais colunas obtemos os quantis dos parâmetros.

Tabela 30: *Resumo a posteriori dos parâmetros - seis componentes*

Parâmetros	Média	Desvio	2,5%	25%	50%	75%	97,5%
γ_1	0,1758	0,1332	0,0077	0,0673	0,1476	0,2616	0,4820
γ_2	0,1860	0,1501	0,0077	0,0647	0,1502	0,2694	0,5725
γ_3	0,1624	0,1322	0,0052	0,0518	0,1321	0,2409	0,4767
γ_4	0,1450	0,1257	0,0052	0,0466	0,1114	0,2098	0,4703
γ_5	0,1822	0,1478	0,0090	0,0622	0,1477	0,2642	0,5402
γ_6	0,1484	0,1322	0,0077	0,0520	0,1140	0,2046	0,5298
p_1	0,4488	0,0772	0,2476	0,4236	0,4566	0,4848	0,5717
p_2	0,4497	0,0779	0,2319	0,4252	0,4574	0,4846	0,5705
p_3	0,4498	0,0849	0,2300	0,4130	0,4572	0,4860	0,5935
p_4	0,4439	0,0839	0,2243	0,4157	0,4517	0,4830	0,5846
p_5	0,4490	0,0736	0,2643	0,4261	0,4550	0,4822	0,5833
p_6	0,4447	0,0840	0,2314	0,4159	0,4524	0,4840	0,5909

Através da Tabela 30 podemos observar que a média dos gamas varia de 0,1450 à 0,1860 e as médias dos p 's variam de 0,4439 à 0,4498.

Os gráficos referentes às iterações, densidades e autocorrelações de p_1, p_2, p_3, p_4, p_5 e p_6 são similares aos gráficos das simulações feitas com 2 e 3 componentes.

Utilizamos também os modelos com mistura com sete e oito componentes para o ajuste aos dados reais afim de compararmos com o modelo com mistura com cinco componentes, o qual obtemos o melhor ajuste aos dados.

Através das simulações realizadas com os oito modelos considerados, observamos que o melhor ajuste foi com o modelo com mistura com cinco componentes. É possível observar tal fato através do cálculo do fator de Bayes mostrado a seguir.

A Tabela 31 mostra os valores obtidos do cálculo de $I_i = \int pr(y|\theta_i, M_i)\pi(\theta_i|M_i) d\theta_i$, para $i = 1, 2, \dots, 8$, para o conjunto de dados reais.

Tabela 31: *Valores de I_i*

I_i	Resultados
I_1	1,1464
I_2	1,1519
I_3	1,1630
I_4	1,1608
I_5	1,1857
I_6	1,1795
I_7	1,1655
I_8	1,1543

Através da Figura 124 podemos observar o gráfico referente aos valores de I_i .

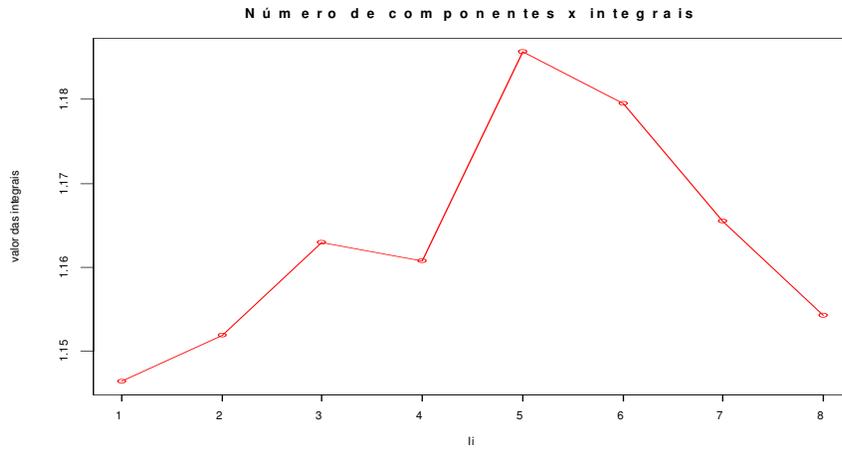


Figura 124: Gráfico dos valores de I_i

Nas Tabelas 32a, 32b, 32c e 32d obtemos os valores dos fatores de Bayes calculados para os oito modelos utilizados no conjunto de dados reais.

Tabela 32a: Valores dos fatores de Bayes

B_{ij}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	B_{17}	B_{18}
Resultados	0,9952	0,9857	0,9104	0,9668	0,9719	0,9836	0,9931

Tabela 32b: Valores dos fatores de Bayes

B_{ij}	B_{23}	B_{24}	B_{25}	B_{26}	B_{27}	B_{28}	B_{34}
Resultados	0,9904	0,9923	0,9714	0,9766	0,9883	0,9979	1,002

Tabela 32c: Valores dos fatores de Bayes

B_{ij}	B_{35}	B_{36}	B_{37}	B_{38}	B_{45}	B_{46}	B_{47}
Resultados	0,9808	0,9860	0,9978	1,0075	0,9790	0,1363	0,9959

Tabela 32d: Valores dos fatores de Bayes

B_{ij}	B_{48}	B_{56}	B_{57}	B_{58}	B_{67}	B_{68}	B_{78}
Resultados	1,0056	1,005	1,017	1,0272	1,012	1,0218	1,0097

Através das Tabelas 32a, 32b, 32c e 32d podemos observar que o modelo escolhido para o conjunto de dados reais é o modelo com mistura com cinco componentes, ou seja, M_5 . O modelo menos favorecido foi o modelo M_1 , o qual denota o modelo com distribuição binomial, ou seja, o modelo sem mistura. Observe na Tabela 31 que os valores de I_i são bem próximos, ou seja, através do cálculo do fator de Bayes, o modelo M_5 é escolhido com uma pequena diferença em relação aos outros modelos.

Na Tabela 33 obtemos os valores resultantes do cálculo de I_i^b , visto na equação (7.7), para $i = 1, 2, \dots, 8$ referente ao conjunto de dados reais.

Tabela 33: Valores de I_i^b

I_i^b	b					
	0, 1	0, 15	0, 2	0, 25	0, 3	0, 35
I_1^b	0, 1321	0, 1457	0, 1609	0, 1779	0, 1969	0, 2182
I_2^b	0, 1327	0, 1464	0, 1616	0, 1787	0, 2153	0, 2190
I_3^b	0, 1340,	0, 1479	0, 1632	0, 1804	0, 1997	0, 2212
I_4^b	0, 1338	0, 1476	0, 1630	0, 1802	0, 1994	0, 2210
I_5^b	0, 1360	0, 1499	0, 1656	0, 1831	0, 2027	0, 2246
I_6^b	0, 1360	0, 1499	0, 1656	0, 1830	0, 2025	0, 2244
I_7^b	0, 1350	0, 1487	0, 1631	0, 1816	0, 1999	0, 2228
I_8^b	0, 1335	0, 1469	0, 1624	0, 1798	0, 2178	0, 2204

Os resultados do cálculo dos fatores de Bayes fracionais, B_{ij}^b , para $i, j = 1, 2, \dots, 8$, podem ser vistos na Tabela 34a e 34b.

Tabela 34a: Valores dos fatores de Bayes fracionais

B_{ij}^b	b					
	0, 1	0, 15	0, 2	0, 25	0, 3	0, 35
B_{12}^b	0, 9954	0, 9952	0, 9956	0, 9955	0, 9145	0, 9963
B_{13}^b	0, 9858	0, 9851	0, 9859	0, 9861	0, 9860	0, 9864
B_{14}^b	0, 9873	0, 9871	0, 9871	0, 9872	0, 9874	0, 9873
B_{15}^b	0, 9713	0, 9720	0, 9716	0, 9716	0, 9713	0, 9715
B_{16}^b	0, 9713	0, 9720	0, 9716	0, 9721	0, 9723	0, 9724
B_{17}^b	0, 9785	0, 9798	0, 9865	0, 9796	0, 9850	0, 9793
B_{18}^b	0, 9895	0, 9918	0, 9907	0, 9894	0, 9040	0, 9900
B_{23}^b	0, 9903	0, 9898	0, 9901	0, 9905	1, 078	0, 9900
B_{24}^b	0, 9917	0, 9918	0, 9914	0, 9916	1, 079	0, 9909
B_{25}^b	0, 9757	0, 9766	0, 9758	0, 9759	1, 062	0, 9750
B_{26}^b	0, 9757	0, 9766	0, 9758	0, 9765	1, 063	0, 9723
B_{27}^b	0, 9830	0, 9945	0, 9865	0, 9840	1, 077	0, 9793
B_{28}^b	0, 9940	0, 9966	0, 9950	0, 9939	0, 9885	0, 9936
B_{34}^b	1, 001	1, 002	1, 001	1, 001	1, 001	1, 000

Tabela 34b: Valores dos fatores de Bayes fracionais (2)

B_{ij}^b	b					
	0,1	0,15	0,2	0,25	0,3	0,35
B_{35}^b	0,9853	0,9866	0,9855	0,9852	0,9852	0,9848
B_{36}^b	0,9853	0,9866	0,9855	0,9858	0,9862	0,9857
B_{37}^b	0,9926	0,9926	1,0060	0,9934	0,9990	0,9928
B_{38}^b	1,0037	1,0068	1,0050	1,0033	0,9169	1,0036
B_{45}^b	0,9838	0,9846	0,9843	0,9841	0,9837	0,9840
B_{46}^b	0,9838	0,9946	0,9843	0,9847	0,9847	0,9848
B_{47}^b	0,9911	0,9926	0,9993	0,9923	0,9975	0,9919
B_{48}^b	1,0022	1,0047	1,0037	1,0022	0,9155	1,0027
B_{56}^b	1,0000	1,0000	1,0000	1,0005	1,0005	1,0009
B_{57}^b	1,0074	1,0080	1,0153	1,0082	1,0082	1,0080
B_{58}^b	1,0187	1,0204	1,0197	1,00183	0,9306	1,0190
B_{67}^b	1,0074	1,0080	1,0153	1,0077	1,0130	1,0072
B_{68}^b	1,0187	1,0204	1,0197	1,0178	0,9297	1,0181
B_{78}^b	1,0112	1,0122	1,0043	1,0100	0,9178	1,01089

Através das Tabelas 35a e 35b observe que o modelo menos favorecido continua sendo o modelo M_1 . Note que quando $0,1 \leq b \leq 0,2$ os modelos escolhidos são M_5 e M_6 , quando $b = 0,25$ e $b = 0,35$, o modelo escolhido é M_5 e quando $b = 0,3$ o modelo escolhido é M_8 .

Note que no cálculo do fator de Bayes usual, o modelo escolhido dentre os oito é o modelo com mistura com cinco componentes, M_5 , e no cálculo do fator de Bayes fracionais o modelo M_5 apenas é escolhido quando $b = 0,25$ e $b = 0,35$, sendo que quando b está entre 0,1 e 0,2, o modelo M_5 é escolhido juntamente com o modelo M_6 .

As Figuras 125, 126, 127, 128 e 129 mostram os gráficos das proporções com que cada variável latente assume o valor 1 ao longo das simulações para cada jogo. Estes gráficos mostram uma estimativa da probabilidade *a posteriori* de que cada jogo esteja associado a cada componente

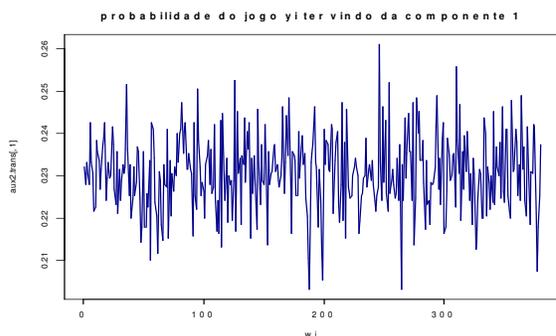


Figura 125: Iterações referente à comp. 1

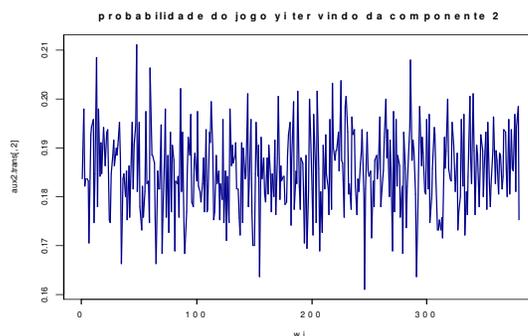


Figura 126: Iterações referente à comp. 2

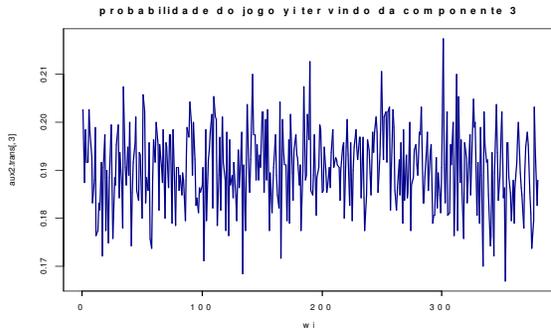


Figura 127: Iterações referente à comp. 3

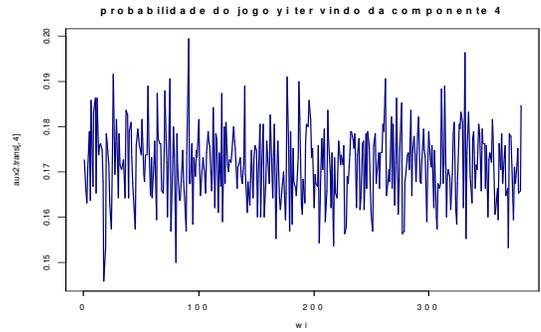


Figura 128: Iterações referente à comp. 4

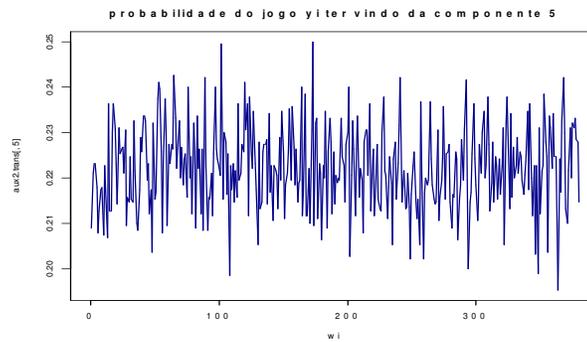


Figura 129: Iterações referentes à comp. 5

8.0.4 Discussão

As Figuras 124, 125, 126, 127 e 128 mostram um resultado claro, indicando o modelo com cinco componentes como sendo o mais adequado para este conjunto de dados. A convergência do método também é um indicador positivo em favor do modelo selecionado.

Entretanto as estimativas, Tabela 30, mostram médias muito próximas para as probabilidades das componentes. Conflitantes com os resultados anteriores.

Capítulo 9

Discussão

Nesta dissertação fizemos uma revisão abrangente dos principais aspectos relacionados ao fator de Bayes.

O fator de Bayes é influenciado pelas distribuições *a priori*. Quando comparamos dois modelos, um com distribuição *a priori* informativa e outro com distribuição *a priori* não informativa para os parâmetros, o primeiro é privilegiado pelo fator de Bayes em função da escolha das distribuições *a priori*.

Os aspectos computacionais do fator de Bayes são tratados no capítulo 2. As soluções geralmente utilizadas são a aproximação de Laplace e os métodos MCMC. Estes métodos são utilizados para aproximar densidades *a posteriori* dos modelos quando não é possível obtê-las analiticamente.

A janela de Occam é utilizada para reduzir o número de modelos inicialmente considerados quando o número de combinações de modelos é muito elevado. O método descarta os modelos com menores chances de serem escolhidos.

Distribuições *a priori* imprópria para os parâmetros dos modelos podem levar à indeterminação do fator de Bayes. Nos capítulos 3 e 4, apresentamos duas soluções para estes problemas, disponíveis atualmente na literatura. São elas, o fator de Bayes intrínseco e o fator de Bayes fracional.

No capítulo 6, apresentamos duas aplicações, uma com um conjunto de dados simulados e outra com dados reais. O fator de Bayes é utilizado como ferramenta para a seleção de modelos nestes casos. Tratamos de uma reanálise de um conjunto de dados relativos ao aproveitamento dos arremessos de um jogador de basquete, Kass e Hsiao (1994).

No capítulo 7 apresentamos uma aplicação com três conjuntos de dados simulados através de misturas de distribuições binomiais com três componentes. Nesta aplicação a finalidade era testar o fator de Bayes afim de que este escolhesse o modelo mais adequado para cada conjunto de dados, assim, verificamos que o fator de Bayes é preciso quando utilizamos, na geração dos dados, parâmetros cujos valores estão razoavelmente afastados. A medida que os dados são gerados através de parâmetros com valores próximos, o fator de Bayes perde especificidade. No capítulo 8 a aplicação é feita com um conjunto de dados reais, o mesmo utilizado no capítulo 6.

O fator de Bayes fracional também foi utilizado na comparação desses modelos. No capítulo 7 verificamos que o fator de Bayes fracional foi preciso apenas quando o valor do expoente b é menor ou igual a 1,5, caso contrário, o fator de Bayes fracional nos evidencia um modelo falso. No capítulo 8, o fator de Bayes fracional nos indica o modelo com mistura com cinco componentes quando $b = 0,25$ e $b = 0,35$, quando $0,1 \leq b \leq 0,2$, o fator de Bayes fracional nos

indica dois modelos, M_5 e M_6 , mas quando o valor do expoente b é igual a 0,3, este nos indica o modelo M_8 .

O resultado da aplicação, capítulo 8, sugerindo um resultado ambíguo, com o fator de Bayes indicando o modelo com cinco componentes e as estimativas dos parâmetros indicando as mesmas probabilidades, aproximadamente 0,44, indicam a necessidade de mais estudos do comportamento do fator de Bayes na seleção de modelos do tipo tratado aqui.

Referências Bibliográficas

- [1] Aitkin, M. (1991), Posterior Bayes Factors, *Journal of the Royal Statistical Society*, série B, **53**, 111-142.
- [2] Bartlett, M. S. (1957), Comment on A Statistical Paradox by D. V. Lindley, *Biometrika*, **44**, 533-534.
- [3] Berger, J. O.; Bernardo, J. M. (1992), On the Development of the Reference Prior Method, *In Bayesian Statistics 4*, (eds. J. M. Bernardo *et al*), London: Oxford University Press, 36-60.
- [4] Berger, J. O.; Pericchi, L. R. (1993), The intrinsic Bayes Factor for Model Selection and Prediction, Rel. Técnico, Purdue University, Depto. Estatística.
- [5] Berger, J. O.; Pericchi, L. R. (1996), The intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the American Statistical Association*, **91**, 109-122.
- [6] Bernardo, J. M.(1979), Reference Posterior Distributions for Bayesian Inference, *Journal of the Royal Statistical Society*, série B, **41**, 113-147.
- [7] Celeux, G.; Diebolt, J. (1985), The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Comput. Statist. Q.*, **2**, 73-82. A stochastic approximation type EM algorithm. Stochastics, to be published, (1992).
- [8] Chaloner, K.; Brant, R. (1988), A Bayesian approach to outlier detection and residual analysis. *Biometrika*, **75**, 651-659.
- [9] Chow, G. C. (1981), A Comparison of the information and Posterior Probability Criteria for Model Selection, *Journal of Econometrics*, **16**, 21-33.
- [10] Dawid, A. P. (1984), Present Position and Potential Developments: Some Personal Views. Statistical Theory. The Prequential Approach, *Journal of the Royal Statistical Society*, série A, **147**, 178-292.
- [11] Dempster, A. P.; Laird, N. M. (1977), Maximum likelihood from incomplete data via the EM algorithm (com discussão), *Journal Royal Statistical Society*, série B, **39**, 1-38.
- [12] Diebolt, J.; Robert, C. P.(1994), Estimation of finite mixture distributions through Bayesian Sampling, *Journal Royal Statistical Society*, série B, **56**, n. 2, 363-375.
- [13] DeGroot, M. H. (1970), Optimal Statistical Decisions. New York: McGraw-Hill

- [14] Dickey, J. M. (1975), Bayesian Alternatives to the F-Test and Least Squares Estimate in the Normal Linear Model, in *Studies in Bayesian Econometrics and Statistics*, ed. S. E. Fieberg and A. Zellner, Amsterdam: North-Holland, pp. 515-554.
- [15] Garcia, C. A. (2001). Métodos numéricos em Inferência Bayesiana, Relatório Final de Pesquisa do Estágio Supervisionado Obrigatório - FCT/UNESP.
- [16] Geisser, S.; Eddy, W. F. (1979), A Predictive Approach to Model Selection, *Journal of the American Statistical Association*, **74**, 153-160.
- [17] Gelfand, A. E. (1995), Model determination using sampling-based methods in *Practical Markov Chain Monte Carlo*, eds. W. R. Gilks, D. J. Spiegelhalter, e S. Richardson, London: Chapman e Hall, pp.145-161.
- [18] Gelfand, A. E.; Dey, D. K. e Chang, H. (1992), Model Determination Using Predictive Distributions with Implementations via Sampling-Based Methods, in *Bayesian Statistics 4*, ed. J. M. Bernardo, Oxford, U. K.: Oxford University Press, pp. 147-167.
- [19] Gelfand, A. E.; Smith, A. F. M. (1990), Sampling - based approaches to calculating marginal densities, *Journal of the American Association*, **85**, 398-409.
- [20] Geman, A. e Geman, D. (1984), Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
- [21] Geweke, J. (1989), Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica*, **57**, 1317-1340.
- [22] Good, I. J. (1952), Rational Decisions, *Journal of the Royal Statistical Society*, série B, **14**, 107-114.
- [23] Hsiao, K. (1994). Bayesian tests of extra-binomial variability with emphasis on the boundary case, Ph.D. dissertation, Carnegie Mellon University, Dept. of Statistics.
- [24] Jeffreys, H. (1961) *Theory of Probability*, London: Oxford University Press.
- [25] Kass, R. E. (1993) Bayes Factors in Practice, *Statistician*, **42**, 551-560.
- [26] Kass, R. E.; Greenhouse, J. B. (1989), Comment on "Investigating Therapies of Potentially Great Benefit: ECMO" by Ware (1989), *Statistical Science*, **4**, 310-317.
- [27] Kass, R. E.; Raftery, A. E. (1995). Bayes Factor, *Journal of the American Statistical Association*, **90**, 773-795.
- [28] Kass, R. E.; Tierney, L.; Kadane, J. B. (1990), The validity of posterior asymptotic expansions based on Laplace's methods, in *Bayesian and Likelihood Methods in Statistics and Econometrics*, eds. S Geisser, J. S. Hodges, S. J. Press e A. Zellner, New York: North-Holland.
- [29] Kass, R. E.; Vaidyanathan, S. (1992), Approximate Bayes Factors and Orthogonal Parameters, *Journal of the Royal Statistical Society*, série B, **54**, 129-144.
- [30] Lempers, F. B. (1971), *Posterior Probabilities of Alternative Linear Models*. Rotterdam: University Press.

- [31] Lindley, D.V. (1957), A statistical paradox, *Biometrika*, **44**, 187-192.
- [32] Lindley, D.V. (1965), Introduction to Probability and Statistics (Partes 1 e 2): Cambridge University Press.
- [33] Madigan, D.; Raftery, A. E. (1994), Model selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window, *Journal of the American Statistical Association*, **89**, 1535-1546.
- [34] Madigan, D.; York, J. (1992), Bayesian graphical models for discrete data, Rel. Técnico, 259, University of Washington, Dept. of Statistics.
- [35] McCulloch, R. E.; Rossi, P. E. (1991), A Bayesian Approach to Testing the Arbitrage Pricing Theory, *Journal of Econometrics*, **49**, 141-168.
- [36] Migon, H. S.; Gamerman, D. (1997), Inferência Estatística: Uma Abordagem Integrada, Textos de Métodos Matemáticos 27, *Instituto de Matemática - UFRJ*.
- [37] Newton, M.; Raftery, A. E. (1994), Approximate Bayesian Inference by the Weighted Likelihood Bootstrap, *Journal of the Royal Statistical Society*, série B, **56**, 3-48.
- [38] O'Hagan, A. (1991), Contribution to the discussion of Posterior Bayes Factor, *Journal of the Royal Statistical Society*, série B, **53**, 137.
- [39] O'Hagan, A. (1995), Fractional Bayes Factors for Model Comparison, *Journal of the Royal Statistical Society*, série B, **57**, 99-138.
- [40] Peña, D.; Tiao, G.C. (1992), Bayesian Robustness Functions for Linear Models in Bayesian Statistics, 4, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith, Oxford: Oxford University Press, pp. 365-388.
- [41] Pettit, L.I.; Young K. D. S. (1990) Measuring the effect of observations on Bayes factors. *Biometrika*, 77, 445-466.
- [42] Raftery, A. E. (1993), Approximate Bayes Factor and Accounting for Model Uncertainty in Generalized Linear Models, Rel. Técnico 255, University of Washington, Dept. de Estatística.
- [43] Raftery, A. E. (1995) Hypothesis Testing and Model Selection with Posterior Simulation in *Practical Markov Chain Monte Carlo*, eds. W. R. Gilks, D. J. Spiegelhalter, e S. Richardson, London: Chapman e Hall.
- [44] Raftery, A. E.; Banfield, J. D. (1990), Stopping the Gibbs Sampler the Use of Morphology and Other Issues in Spatial Statistics, *Annals of the Institute of Statistical Mathematics*, 43, 32-43.
- [45] Rissanen, J. (1987), Stochastic Complexity, *Journal of the Royal Statistical Society*, série B, 49, 223-239.
- [46] Rosenkrantz, S. (1992), The Bayes Factor for Model Evaluation in a Hierarchical Poisson Model for Area Counts, Ph.D. dissertation, University of Washington, 1992, Dept de Bioestatística.

- [47] Rubin, D. B. (1984), Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician, *The Annals of Statistics*, 12, 1151-1172.
- [48] Smith, A. F. M. (1991), Contribution to the discussion of Posterior Bayes Factors, *Journal of the Royal Statistical Society, série B*, **53**, 132-133.
- [49] Spiegelhalter, D. J.; Smith, A. F. M.(1982), Bayes Factors for Linear and Log-Linear Models with Vague Prior Information, *Journal of the Royal Statistical Society, série B*, **44**, 377-387.
- [50] Tachibana, V. M. (1995), Métodos Aproximados em Modelos de Respostas Aleatorizadas e Regressão Logística. Tese de Doutorado, Programa de Engenharia de Produção.