

**CRISTIANE ALMEIDA SARAIVA**

**PRIORIS PARA MODELOS PROBABILÍSTICOS DISCRETOS EM  
CIÊNCIAS AGRÁRIAS**

**Recife  
2007**

**CRISTIANE ALMEIDA SARAIVA**

**PRIORIS PARA MODELOS PROBABILÍSTICOS DISCRETOS EM  
CIÊNCIAS AGRÁRIAS**

**Dissertação apresentada ao colegiado do  
Mestrado de Biometria da Universidade  
Federal Rural de Pernambuco, para  
obtenção do título de Mestre em Biometria.**

**Orientador: Prof<sup>o</sup> Dr. Eufrázio de Souza  
Santos.**

**Recife  
2007**

Ficha catalográfica  
Setor de Processos Técnicos da Biblioteca Central – UFRPE

S243p Saraiva, Cristiane Almeida  
Prioris para modelos probalísticos discretos em Ciência  
Agrárias / Cristiane Almeida Saraiva. -- 2007.  
51 f. : il.

Orientador : Eufrázio de Souza Santos  
Dissertação (Mestrado em Biometria) - Universidade  
Federal Rural de Pernambuco. Departamento de Estatística  
e Informática.

Inclui anexo e bibliografia

CDD 519.5

1. Estatística
  2. Inferência Baysiana
- I. Santos, Eufrázio de Souza  
II. Título

Universidade Federal Rural de Pernambuco  
Departamento de Informática e Estatística  
Programa de Pós-Graduação em Biometria

**PRIORIS PARA MODELOS PROBABILÍSTICOS DISCRETOS EM CIÊNCIAS  
AGRÁRIAS**


Cristiane Almeida Saraiva

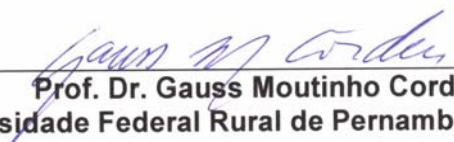
Dissertação foi julgada adequada para obtenção do título de mestre em Biometria,  
defendida e aprovada por unanimidade em 30/03/2007 pela banca examinadora.

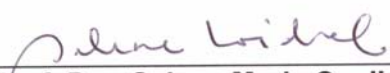
Orientador:

  
\_\_\_\_\_  
**Prof.º Dr. Eufrazio de Souza Santos**  
Universidade Federal Rural de Pernambuco – UFRPE

Banca Examinadora:

  
\_\_\_\_\_  
**Prof. Dr. Borko Stosic**  
Universidade Federal Rural de Pernambuco – UFRPE

  
\_\_\_\_\_  
**Prof. Dr. Gauss Moutinho Cordeiro**  
Universidade Federal Rural de Pernambuco – UFRPE

  
\_\_\_\_\_  
**Prof. Dra. Selene Maria Coelho Loibel**  
Universidade de São Paulo – USP

***Felizes aqueles que se  
divertem com problemas  
que educam a alma  
e elevam o espírito.***

**Fenelon**

35 O Pai ama o Filho, e todas as coisas entregou nas suas mãos.

36 Aquele que crê no Filho tem a vida eterna; mas aquele que não crê no Filho não verá a vida, mas a ira de Deus sobre ele permanece.

João 3, 4.

## AGRADECIMENTOS

A Deus, que me deu forças no momento de angústia e desânimo. Minha fonte para renovar minhas forças e nunca desistir.

A minha família que me deu apoio, sempre participando e me estimulando a ir mais além.

Ao Prof<sup>o</sup>. Eufrázio de Souza Santos pela confiança.

Aos meus colegas Sérgio Paiva, Antônio Lopes, Walter Santos, Ana Virgínia, Simone entre outros.

A minha amiga Adalmeres Cavalcanti que conheci aqui no mestrado e que esteve do meu lado o tempo todo e que sempre me lembrarei com carinho.

E a todos aqueles que sempre me incentivaram a continuar meus sinceros agradecimentos.

**SUMÁRIO**

	<b>página</b>
<b>LISTA DE TABELAS</b> .....	8
<b>RESUMO</b> .....	9
<b>ABSTRACT</b> .....	10
<b>1. INTRODUÇÃO</b> .....	11
<b>2. REVISÃO DE LITERATURA</b> .....	13
2.1 A Estatística Bayesiana .....	18
<b>3. MATERIAL E MÉTODOS</b> .....	22
3.1 Teorema de Bayes .....	22
3.2 Priori Conjugada .....	24
3.3 Método de Jeffreys .....	25
3.4 Método de Laplace .....	27
3.5 Método de Haldane .....	27
3.6 Método da Verossimilhança de dados trasladados .....	28
<b>4. RESULTADOS E DISCUSSÃO</b> .....	34
<b>5. CONCLUSÃO</b> .....	37
<b>6. RECOMENDAÇÕES</b> .....	38
<b>7. REFERÊNCIA BIBLIOGRAFICA</b> .....	39
<b>8. ANEXO</b> .....	42



## LISTA DE TABELAS

	<b>Pág.</b>
Tabela 1. Dados obtidos por telefone de dez granjas de Pernambuco.....	29
Tabela 2. Resultados através do método de Laplace das estimativas do parâmetro $p$ , seu desvio padrão e IC 95%, realizando 500 iterações no Winbugs 14.....	34
.....	34
Tabela 3. Resultados através do método de Jeffreys das estimativas do parâmetro $p$ , seu desvio padrão e IC 95%, realizando 500 iterações no Winbugs 14. ....	35
Tabela 4. Resultados através do método de Haldane das estimativas do parâmetro $p$ , seu desvio padrão e IC 95%, realizando 500 iterações no Winbugs 14. ....	35

## RESUMO

Objetivando selecionar prioris mais adequadas para dados discretos estudamos técnicas para determinação de prioris, tais como métodos de Laplace, método de Jeffreys e método de Haldane em que as prioris são conjugadas. Foi tomada uma amostra de dez granjas dentre as 53 existentes do Estado de Pernambuco com o propósito de estimar a probabilidade de ovos comerciais (grandes). Tendo em vista que os ovos são classificados como industrial, pequeno, médio, grande, extra e jumbo, classificamos os ovos em pequeno e grande. Os ovos industriais, pequenos e médios foram tidos como pequeno e os ovos grandes, extra e jumbo, como grande. Com a suposição de que os dados amostrais seguem uma distribuição binomial e utilizando prioris determinadas pelos métodos acima descritos, utilizamos o software Winbugs 1.4 com o qual foram calculados a média, desvio padrão, intervalo de credibilidade de 95% e sua amplitude. Para cada um dos métodos utilizamos 20.000 iterações das quais as 10.000 primeiras foram descartadas observando-se que o equilíbrio da cadeia iniciou-se com 12.500 iterações. Obtivemos uma estimativa média do parâmetro  $p$  o qual foi semelhante nos métodos de Laplace, Jeffreys e Haldane, correspondendo a aproximadamente  $p = 0,664$ .

## ABSTRACT

With the propose to choose priors more fitted for discrete data, we study technics for determination of priors just as Laplace's Methods, Jeffreys's Methods and Haldane's Methods which are conjugated prior. We take a sample of ten grange among the fifty three ones existent of the Pernambuco's State to estimate the probability of commercial egg (big type). We suppose that the distribution from the sample data is binomial and we use the methods quoted above. The software used for that was the package Winbugs 1.4 where we compute the average, standard deviation, 95% credible interval and their amplitude. For each one of the methods it was observed that 20.000 iterations were sufficient since the balance of the chain already had established with 12.500 iterations. The estimated parameter  $p=0,664$  was obtained by the Laplace's Method, Jeffreys's Method and Haldane's Method.

## 1. INTRODUÇÃO

A inferência estatística trata de problemas onde se têm quantidades de interesse chamadas de parâmetros associadas a algumas características populacionais ou situações práticas onde se deseja fazer afirmações sobre elas com maior precisão possível. Para tal temos a abordagem clássica e a bayesiana. A abordagem clássica permite que afirmações ou inferências sobre estas quantidades sejam feitas a partir de informações fornecidas por uma amostra aleatória selecionada da população enquanto que a abordagem bayesiana dispõe da incorporação de informações obtidas de experimentos passados ou experiências de especialistas. Segundo POLLARD (1986) de forma mais geral, a aproximação bayesiana pode ser vista com uma extensão da aproximação clássica.

Existem várias situações para serem consideradas, tais como: em um casal de  $n$  filhos, a variável de interesse é o número de filhos do sexo masculino ou feminino, pesquisas eleitorais em que os indivíduos na população são ou não favoráveis a determinado partido ou candidato, o número de partículas  $\alpha$  emitidas por uma fonte radioativa em um intervalo de tempo, numa floresta o número de árvores desmatadas em uma região, e o número de mortes em determinado período da vida (idade). Todas essas são situações que produzem dados discretos com os quais trabalharemos para determinar algumas prioris não-informativas tais como: a priori de Jeffreys, verossimilhança de dados transladados, priori conjugada e método de Laplace. Mais especificamente trabalharemos com dados provenientes de uma amostra de dez granjas tomando das 53 existentes em Pernambuco, com o objetivo de estimar a produção de ovos comerciais.

A análise bayesiana com prioris não-informativas é muito comum quando pouco ou nenhuma informação está disponível. Uma das prioris não-informativas mais amplamente usada, introduzida por Laplace (Datta, 1996), é a distribuição uniforme sobre o espaço paramétrico. Outra priori muito citada é a proposta por Jeffreys que é dada pela raiz quadrada do determinante da matriz de informação de Fisher e que possui a propriedade da invariância. Usaremos também a priori de Haldane que é adquirida através da substituição de troca de variável para uma priori uniforme na log-odds.

Utilizamos dados de dez granjas dentre as 53 existentes do Estado de Pernambuco com o propósito de estimar a probabilidade de ovos comerciais

(grandes). Tendo em vista que os ovos são classificados como industrial, pequeno, médio, grande, extra e jumbo, classificamos os ovos em pequeno e grande. Os ovos industriais, pequenos e médios serão tidos como pequeno e os ovos grandes, extra e jumbo, como grande. Supondo que os dados amostrais seguem uma distribuição binomial com parâmetro de sucesso ovos grandes e utilizando prioris determinadas pelos métodos acima descritos utilizamos o software Winbugs 1.4 onde foram calculados a média, desvio padrão, intervalo de credibilidade de 95% e sua amplitude. Para cada um dos métodos utilizamos 20.000 iterações dos quais as 10.000 primeiras foram descartadas observando-se que o equilíbrio da cadeia iniciou com 12.500 iterações. Obtivemos que a estimativa média do parâmetro  $p$  teve valor semelhante nos métodos de Laplace, Jeffreys e Haldane, correspondendo a aproximadamente  $p = 0,664$ . Também obtivemos que esta estimativa está presente em três dos intervalos descritos pelos métodos citados.

## 2. REVISÃO DE LITERATURA

BARTOLUCCI (2006) em uma das aplicações utilizadas, comparou alguns métodos para selecionar um modelo de regressão logístico hierárquico para o número de sobreviventes  $Y$  em uma amostra de 79 indivíduos acometidos de certa doença usando a condição paciente,  $A$ , e o tratamento recebido,  $B$ , como fatores explanatórios. Utilizando cinco modelos possíveis:  $M_1, M_2, M_3, M_4, M_5$ , onde o modelo completo,  $M_5$ , é formulado como:  $Y_{ij} \sim Bi(n_{ij}, p_{ij})$ ,  $\log it(p_{ij}) = \mu + \mu_i^A + \mu_j^B + \mu_{ij}^{AB}$ , com  $i, j = 1, 2$ ,  $Y_{ij}, n_{ij}$  e  $p_{ij}$  são respectivamente o número de sobreviventes, o número total de pacientes e a probabilidade de sobreviventes para o paciente com condição  $i$  que recebeu o tratamento  $j$ . Utilizou uma distribuição a priori normal  $N(0,8)$  para os parâmetros  $\mu, \mu_1^A, \mu_2^B$  e  $\mu_{22}^{AB}$ , que por suposição são prioris independentes.

PIRES (2006) considerando a implementação do controle de qualidade de atributos, a eficiência do sistema que classifica os itens manufaturados como conforme ou não conforme precisa ser considerada. Dois tipos de erros podem ocorrer durante a inspeção: o primeiro, conhecido como tipo I, ocorre quando um item conforme é classificado como não conforme; e o segundo, denotado por tipo II, quando um item é dito conforme quando é, na verdade, não conforme. Supondo que uma amostra aleatória de  $n$  unidades, um número  $X$  de itens conformes é observado onde tem distribuição binomial com parâmetros  $(n, p)$ , ou seja,  $X \sim Bin(n, p)$ . Contudo, a presença de erros de classificação no sistema implica numa modificação desta função de probabilidade. Seja  $e_1$  a probabilidade de que um item conforme seja erroneamente classificado como não conforme, e seja  $e_2$  a probabilidade de que um item não conforme seja classificado como conforme. Então a probabilidade de que um item seja classificado como conforme é  $q = p(1 - e_1) + (1 - p)e_2$ , definindo uma variável aleatória  $X$  que tem distribuição binomial com parâmetro  $q$  ao invés de  $p$ . Considerou-se uma distribuição a priori conjunta de  $(p, e_1, e_2)$ , nas quais as variáveis aleatórias são mutuamente independentes a priori, dada por:  $f(p, e_1, e_2) = f_\beta(p | \alpha, \beta) f_\beta(e_1 | \alpha_1, \beta_1) f_\beta(e_2 | \alpha_2, \beta_2)$ , em que  $f_\beta(a | b, c)$  é função densidade de uma distribuição Beta para a variável aleatória  $a$  com parâmetros  $b$  e  $c$ . Porém a ausência de informações suficientes para definir distribuições a priori

informativas para os erros de classificação implica na utilização de distribuições  $U(0,1)$ , caso particular de distribuição Beta, para os parâmetros  $(p, e_1, e_2)$ .

EHLERS (2005) em seu livro texto dá um exemplo onde uma variável tem distribuição Binomial e o parâmetro de interesse assume algumas possíveis distribuições a priori como, uma distribuição normal truncada no intervalo  $(0, 1)$  e também uma distribuição Beta.

BOLSTAD (2004) em seu livro texto tomou como exemplo uma grande população em que o parâmetro a ser estudado  $\pi$  é a proporção de eleitores que tem a intenção de votar no candidato  $A$  para prefeito. Foi tomada uma amostra aleatória da população e considerou  $Y$  o número de eleitores que disse ter a intenção de votar no candidato  $A$ . A função de verossimilhança foi uma binomial  $(n, \pi)$  onde as prioris tomadas foram a uniforme e beta. A uniforme foi tomada com a intenção de não favorecer qualquer possível valor para o parâmetro ou quando se quer ser tão objetivo quanto possível e não colocar sua opinião pessoal. A distribuição beta é a família conjugada para a distribuição binomial e tem a distribuição uniforme  $U \sim (0,1)$  como caso particular.

No trabalho de BORGATTO (2004), três modelos foram ajustados a um conjunto de dados obtidos de um ensaio de controle biológico para *Diatraea saccharalis*, uma praga comum em plantações de cana – de – açúcar. Usando a distribuição binomial como modelo de probabilidade, um ajuste adequado não pode ser obtido devido à superdispersão gerada pela variabilidade dos dados e pelo excesso de zeros. Neste caso, o modelo binomial inflacionado de zeros (ZIB) superdisperso é mais flexível e eficiente para a modelagem desse tipo de dados. Em muitos casos, para dados de contagens, não é possível observar a ocorrência de algum valor ou o pesquisador não tem interesse nesse valor, em geral, o zero. Portanto considera que a variável aleatória tem uma distribuição truncada no zero e distribuição discreta como Poisson e binomial negativa. Nesse trabalho foram utilizados 3 modelos para a modelagem de dados de proporções com superdispersão e excesso de zeros. O modelo binomial inflacionado de zeros apesar de ter sido menos eficiente do que o modelo binomial truncado, quanto ao ajuste desse conjunto de dados, é para ajustar conjuntos de dados, em que o zero é de interesse do pesquisador. A análise Bayesiana dos modelos usados para ajustar a variabilidade e o excesso de zeros permite incorporar aos modelos informações prévias a respeito dos parâmetros, e isso muitas vezes produz inferências mais precisas.

NEVES (2004) em seu trabalho *Graduação bayesiana de Taxas de Mortalidade*, foi implementado modelos de graduação bayesiana de taxas de mortalidade utilizando MCMC, através do WinBUGS. A elaboração de tábuas biométricas de mortalidade bayesianas é realizada por meio de modelos bayesianos dinâmicos e a tábua de melhor ajuste é comparada com tábuas de mortalidade conhecida e utilizada no mercado. Assumindo-se que todos os indivíduos com a mesma idade morrem independentemente e com mesma probabilidade, então o número de mortes observadas –  $d_{x,t}$ , em cada ano de estudo e idade, é Poisson distribuído com média  $e_{x,t} \cdot \mu_{x,t}$ , onde  $\mu_{x,t}$  é a força de mortalidade na idade  $x$  no tempo  $t$  e  $e_{x,t}$  denota a população exposta ao risco, a qual é suposta conhecida. Para descrever a relação entre o número de mortes observadas e as correspondentes idades usou-se o seguinte modelo:  $d_{x,t} | \mu_{x,t} \sim \text{Poisson}(e_{x,t} \cdot \mu_{x,t})$ ,  $x = x_{\text{inf}}, \dots, x_{\text{sup}}$  onde:  $t = 1, \dots, T$ , o número de períodos de observação,  $\mu_{x,t} > 0$  e  $e_{x,t}$  são constantes conhecidas. Para modelar a evolução no tempo das forças de mortalidade utilizamos modelos dinâmicos generalizados considerando  $\mu_{x,t}$  relacionados através de perturbações multiplicativas:  $\log(\mu_{x,t}) = \log(\mu_{x,t-1}) + \omega_t$ , com  $\omega_t \sim \text{Normal}(0, W_t)$ ,  $x = x_{\text{inf}}, \dots, x_{\text{sup}}$  e  $t = 2, \dots, T$  onde  $W_t$  é modelado através de uma distribuição Gama Inversa não informativa.

QUEIROZ (2004) apresentou uma estimativa bayesiana aplicada na identificação de fatores de risco para HIV em doadores de sangue onde utilizou o modelo de regressão binária. O modelo probabilístico para os dados foi o modelo de Bernoulli e a distribuição a priori não informativa para os parâmetros foi  $N(0; 0,001)$ .

ZACHARIAS et al. (2004) apresentou um modelo Bayesiano de captura-recaptura e determinou os estimadores Bayesianos de  $N$ . Considerou prioris não-informativas para os parâmetros populacionais e determinou condições para que a distribuição a posteriori conjunta dos parâmetros exista, bem como expressões para a média e a moda da distribuição a posteriori marginal de  $N$ . Tomando  $N$  o tamanho da população;  $k$  o número de amostras ( $k \geq 2$ );  $p_i$  a probabilidade de que qualquer animal seja capturado na  $i$ -ésima amostra, independentemente dos demais,  $i = 1, 2, \dots, k$ ;  $\mathbf{p} = (p_1, p_2, \dots, p_k)$  o vetor  $k$ -dimensional das probabilidades de captura;  $n_i$  o número de animais capturados na  $i$ -ésima amostra,  $i = 1, 2, \dots, k$ ;  $m_i$  o número de animais marcados capturados na  $i$ -ésima amostra,  $i = 1, 2, \dots, k$ , ( $m_1 = 0$ );



$M_j = \sum_{i=1}^{j-1} (n_i - m_i)$ ,  $j = 2, 3, \dots, k$ , o número de animais marcados presentes na população exatamente antes da seleção da  $i$ -ésima amostra, ( $M_1 = 0$ );  $r = \sum_{i=1}^k (n_i - m_i)$  o número total de animais distintos capturados durante todo o processo das  $k$  amostragens e  $D = n_1, m_1; n_2, m_2; \dots; n_k, m_k$  o conjunto de estatísticas ou dados referentes ao experimento. Se as amostras selecionadas são independentes então a função de verossimilhança pode ser escrita como

$$L(N, \mathbf{p}/D) = P(n_1, m_1; n_2, m_2; \dots; n_k, m_k | N, \mathbf{p}) \propto \binom{N}{r} \prod_{i=1}^k p_i^{n_i} (1 - p_i)^{N - n_i}, \quad N \geq r \quad \text{e}$$

$0 < p_i < 1$ ,  $i = 1, 2, \dots, k$ . A função de probabilidade foi uma binomial as distribuições “a priori” de  $N$  e  $\mathbf{p}$  foram não informativas, e forem “a priori” i.i.d. com distribuição uniforme em  $(0, 1)$ , ou seja,  $\pi(N) = 1$ , para todo  $N$  pertencente aos naturais e

$$\pi(\mathbf{p}) = \prod_{i=1}^k \pi(p_i) = 1, \quad 0 < p_i < 1, \quad i = 1, 2, \dots, k.$$

LEE et al. (2003) considerou a estimação do tamanho de uma população fechada usando modelo de captura-recaptura quando a probabilidade de captura varia com o tempo e com o comportamento resposta. Tomando  $i$  variando de 1 à  $N$  o número de animais em uma população fechada e  $j$  variando de 1 à  $t$  sendo o número de amostras tiradas da população, LEE et al (2003) então considerou como comportamento resposta o fato do  $i$ -ésimo animal já ter sido ou não capturado antes da  $j$ -ésima amostra. Sendo a distribuição dos dados uma binomial os parâmetros foram  $N, P, \phi$  onde tiveram prioris não-informativa exceto para  $\phi$ . A priori para o parâmetro  $P$  é uma distribuição uniforme  $U \sim (0, 1)$ , a priori de  $\phi$  é uma distribuição uniforme  $U \sim (\alpha, \beta)$  e para  $N$  uma priori de Jeffreys.

SOUZA (2003) mostra em um exemplo para a estimação da probabilidade do sétimo filho ser menino dado que são cinco meninos e o sexto é menina, e que todos são filhos do mesmo pai. Em algumas formas de estimar o parâmetro de interesse onde um deles é o método bayesiano. Para os cálculos ele usa uma distribuição a priori beta com três valores diferentes para  $\alpha$  e  $\beta$  para com isso mostrar que quanto a distribuição a priori é vaga, a distribuição a posteriori é baseada na verossimilhança e a medida que se aumenta a informação a priori, mais peso esta terá na distribuição a posteriori e menos dependência dos dados observados.

COELHO (2002) em seu trabalho de abordagem bayesiana na análise genética de populações utilizando dados de marcadores moleculares na avaliação do grau de estruturação da variabilidade genética entre e dentro dos indivíduos e a obtenção de estimativas de parâmetros genéticos indicadores do sistema reprodutivo da espécie assume grande importância. Os parâmetros de maior interesse neste caso são o índice de fixação intrapopulacional ( $f$ ) e a taxa de fecundação cruzada ( $t$ ). Para o parâmetro de fixação intrapopulacional foi utilizada uma distribuição Beta utilizando o algoritmo de Metropolis-Hastings para se obter a distribuição de probabilidade a posteriori empiricamente. Os efeitos da variação na taxa de fecundação cruzada foram avaliados admitindo que esta taxa é uma variável aleatória ( $t$ ) com uma função densidade de probabilidade Beta e o método proposto baseia-se no algoritmo de Monte Carlo baseado em Cadeias de Markov (MCMC) de Metropolis- Hastings para obtenção das distribuições a posteriori.

SILVA (2002) teve como principal objetivo apresentar uma estimativa da densidade populacional do sagüi -do- Nordeste (*Callithrix jacchus*) em uma área de 25ha na Estação Ecológica do Tapacurá. Foi ajustada uma distribuição a priori beta com parâmetros,  $\alpha = 2$  e  $\beta = 5$  para o número total de animais observados concluindo que apenas o conhecimento acerca da composição do grupo não é suficiente para estimar a densidade populacional.

Temos através do trabalho de ABUSEV et al. (2001) que para as distribuições sendo uma binomial ou uma binomial negativa a priori pode ser uma distribuição beta, já para uma função poisson podemos usar uma gama.

FREIRE (2001) fez uma revalidação de alguns métodos criados para corrigir o viés provocado pelo problema da má classificação muito comum em estudos epidemiológicos. Um desses métodos foi o bayesiano onde a distribuição dos dados é uma Poisson e a priori é uma uniforme.

O trabalho de ALEXANDER et. al. (2000) é sobre um parasita que causa a doença filariose linfática em que se trabalhou com a distribuição binomial negativa. Para os parâmetros foram usadas as prioris gama, normal e de Jeffreys.

LARSON (1982) em livro texto dá exemplos de modelos probabilísticos discretos, onde um deles em um único lançamento de uma moeda a função de probabilidade para a amostra é uma distribuição de Bernoulli e o autor assume três tipos de prioris. A primeira é uniforme no intervalo entre zero e um, a segunda é uniforme no intervalo de 0,4 a 0,6 a terceira é uma distribuição uniforme discreta no intervalo de 0,4 a 0,6. O outro exemplo é sobre peças defeituosas em um lote, onde

neste caso a função de probabilidade é uma distribuição hipergeométrica e foi tomada para priori uma distribuição binomial para o número de peças defeituosas no lote.

## 2.1 A estatística bayesiana

De acordo com POLLARD (1986) a aproximação bayesiana é um método para pôr num fundo comum a informação a priori e da amostra. Ele especifica como esta informação a priori deve ser corrigida pelos novos dados.

Segundo KASS (1996), o subjetivismo transformou-se no fundamento filosófico dominante para a inferência Bayesiana. Contudo na prática, a maioria das análises Bayesianas é executada com a chamada priori não-informativa, isto é, construídas por regras formais.

Desde Bayes, e especialmente desde Fisher, o espaço e mérito da inferência Bayesiana tem sido debatido. Críticos tendem a ver a determinação de uma priori como um importante, mas controlável detalhe técnico. Como o avanço da computação ampliou o interesse em técnicas Bayesianas, novas aplicações continuam levantando a questão de como prioris são escolhidas.

Ao longo dos anos, vários esquemas foram investigados e com isso vemos que as idéias fundamentais e métodos foram originados com Jeffreys. Uma observação essencial é que o ponto de vista de Jeffreys evoluiu através de prioris tomadas como escolhidas por convenção, melhor do que como representação única de ignorância.

A idéia de “nenhum conhecimento” que é conduzida a uma priori não-informativa, segundo Pollard (1986), tem que ser usada com cautela, pois ela entra em contradição com a sua finalidade. Logo a expressão de conhecimento a priori é melhor vista como uma aproximada representação de conhecimento vago do que uma estrita representação de ignorância a priori.

Segundo PAULINO (2003) as distribuições não-informativas podem desempenhar um papel de referência mesmo que se disponha de fortes crenças a priori como forma de:

- Deduzir as crenças a posteriori para quem parte de um conhecimento escasso (quando a amostra fornece o grosso da informação sobre o parâmetro) e, nessa medida, se acha incapaz de determinar subjetivamente uma distribuição razoável – assim se reconhecendo a sua própria ignorância;
- Permitir a comparação com os resultados da inferência clássica que “só” usa a informação amostral (no todo ou em parte);
- Averiguar a influência nas inferências da distribuição a priori subjetiva que descreve a informação realmente existente, quando confrontada com as que resultam do uso da distribuição a priori de referência.

O argumento primeiramente invocado para gerar distribuições não informativas foi o Princípio da Razão Insuficiente devido a Bayes e Laplace. De acordo com este princípio, na ausência de razão suficiente para privilegiar umas possibilidades em detrimento de outras, decorrente da escassez informativa a priori, deve-se adotar a equiprobabilidade conforme Paulino (2003). A consideração deste argumento como objectivista não é totalmente pacífica já que é questionável falar-se de “ausência de razão suficiente” sem envolvimento de juízos subjetivos.

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

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

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

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

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

De fato para qualquer transformação real injetiva de  $\theta \in \mathfrak{R}$  mostra que a distribuição proposta por Jeffreys para o caso uniparamétrico

$$h(\theta) \propto [I(\theta)]^{\frac{1}{2}},$$

contém a referida propriedade de invariância.

Há quem considere inadequadas as distribuições a priori de Jeffreys pelo fato de dependerem do modelo gerador dos dados e pelas implicações daí advenientes. O argumento avançado é que a representação da ignorância total sobre um parâmetro deve ser a mesma qualquer que seja a natureza da experiência a efetuar. A reação a esta posição crítica apóia-se na argumentação de que as distribuições não-informativas não pretendem representar a ignorância total (cuja existência é questionada por muitos), mas uma informação a priori vaga relativamente à informação que se espera obter da experiência que se projeta realizar, pelo que a sua forma deve depender da respectiva verossimilhança.

BOX e TIAO incluem-se entre os que não reconhecem a existência de um estado de ignorância absoluta e encaram a vaguidade ou informatividade das crenças a priori unicamente em relação à quantidade de informação derivada da experiência geradora dos dados.

Procurando definir critérios que permitissem retratar a vaguidade relativa de informação a priori e, desse modo, gerar distribuições a priori não-informativas, a sua idéia base foi procurar uma reparametrização  $\psi = \psi(\theta)$  onde a respectiva verossimilhança fosse apenas transladada pelos dados. Uma vez detectada uma transformação  $\psi$  deste tipo, considera-se para ela uma distribuição que assegure que a respectiva distribuição a posteriori seja essencialmente a verossimilhança normalizada. Deste modo, as distribuições a posteriori de  $\psi$  para diferentes amostras só diferem umas das outras pela localização.

Não simpatizando com a natureza imprópria das distribuições uniformes em conjuntos não limitados, Box e Tião propuseram para  $\psi$  uma distribuição própria cuja densidade seja sensivelmente constante na gama de valores da função em

questão de modo que aqui nenhum valor de  $\psi$  seja substancialmente mais provável que qualquer outro e que fora dela os seus valores não são tão significativos.

Este tipo de distribuição apresenta uma densidade essencialmente uniforme na gama de valores significativos da verossimilhança, decaindo para zero à medida que se afasta desse gama, o que justifica a designação de *distribuições localmente uniformes*.

### 3. MATERIAL E MÉTODOS

A aproximação bayesiana (Pollard, 1986) é um método que põe num fundo comum priori e informação amostral, e especifica como esta informação a priori será corrigida pelos novos dados. Neste caso há uma relação envolvendo probabilidade condicional que é conhecida como Teorema de Bayes.

#### 3.1 Teorema de Bayes

Tomemos a definição de BOX e TIAO (1973), com relação a natureza da inferência bayesiana. Suponha que  $y' = (y_1, \dots, y_n)$  é um vetor de  $n$  observações cuja distribuição de probabilidade  $p(y|\theta)$  depende dos  $k$  valores do parâmetro  $\theta' = (\theta_1, \dots, \theta_k)$ . Suponha também que a distribuição de probabilidade de  $\theta$  é  $p(\theta)$ . Então,  $p(y|\theta)p(\theta) = p(y, \theta) = p(\theta|y)p(y)$ .

Dado  $y$  observado, a distribuição condicional de  $\theta$  é

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}. \quad (1)$$

Também, nós podemos escrever

$$p(y) = E p(y|\theta) = c^{-1} = \begin{cases} \int p(y|\theta)p(\theta)d\theta & \theta \text{ contínuo} \\ \sum p(y|\theta)p(\theta) & \theta \text{ discreto} \end{cases} \quad (2)$$

onde a soma ou a integral é tomada sobre a amplitude admissível de  $\theta$  e onde  $E[f(\theta)]$  é a esperança matemática de  $f(\theta)$  com relação a distribuição  $p(\theta)$ . Assim nós podemos escrever (1) alternativamente como

$$p(\theta|y) = c p(y|\theta)p(\theta). \quad (3)$$

A condição em (1), ou seu equivalente em (3) é referido como Teorema de Bayes. Nesta expressão,  $p(\theta)$ , diz o que é conhecido a respeito de  $\theta$  sem o conhecimento dos dados, é chamada de distribuição a priori de  $\theta$ . Correspondentemente,  $p(\theta|y)$  diz o que é conhecido a respeito de  $\theta$  dado o conhecimento dos dados, é chamada distribuição a posteriori de  $\theta$  dado  $y$ . A quantidade  $c$  é simplesmente uma constante de “normalização” necessária para assegurar-se de que a distribuição a posteriori  $p(\theta|y)$  integrada ou somada (em caso discreto) é 1.

Para o Teorema de Bayes e a Função de Verossimilhança tomemos as observações dadas  $y$ ,  $p(y|\theta)$  em (3). Elas podem ser consideradas como uma função não de  $y$ , mas de  $\theta$ . Quando for assim considerada, conforme Fisher (1922), será chamada função de verossimilhança de  $\theta$  para  $y$  dado e pode ser escrito  $l(\theta|y)$ . Nós podemos assim escrever a fórmula de Bayes como

$$p(\theta|y) = K \cdot l(\theta|y)p(\theta). \quad (4)$$

Em outras palavras, o Teorema de Bayes diz que a distribuição de probabilidade de  $\theta$  a posteriori para  $y$  dado é proporcional ao produto da distribuição para  $\theta$  a priori para os dados e a verossimilhança de  $\theta$  dado  $y$ . Isto é,

Distribuição a posteriori  $\propto$  verossimilhança  $\times$  distribuição a priori

Quanto à natureza seqüencial do Teorema de Bayes, suponhamos que temos uma amostra inicial  $y_1$  e assim temos pela fórmula:

$$p(\theta|y_1) \propto p(\theta)L(\theta|y_1), \quad (5)$$

supondo uma segunda amostra  $y_2$ , distribuída independentemente da primeira, teremos:

$$p(\theta|y_2, y_1) \propto p(\theta)L(\theta|y_1)L(\theta|y_2). \quad (6)$$



Obviamente este processo pode ter  $n$  repetições independentes que a distribuição a posteriori será recalculada a cada nova observação, de modo que no  $m$ -ésimo estágio a verossimilhança associada à  $m$ -ésima observação seja combinada com a posteriori de  $\theta$  depois de  $m-1$  observações:

$$p(\theta | y_1, \dots, y_m) \propto p(\theta | y_1, \dots, y_{m-1})L(\theta | Y_m), \quad m = 2, \dots, n$$

onde

$$p(\theta | y_1) \propto p(\theta)L(\theta | y_1). \quad (7)$$

### Método para construção de priori referência

Descreveremos alguns métodos que foram propostos para construção de priori referência. Vários dos métodos alternativos conduzem a priori de Jeffreys ou modificação desta.

### 3.2 Priori Conjugada

Segundo LEE (2004) a usual definição adotada é como segue.

Seja  $l$  uma função de verossimilhança  $l(\theta | x)$ . Uma classe  $\Pi$  de distribuições a priori é dita formar uma *família conjugada* se a densidade posteriori

$$p(\theta | x) \propto p(\theta)l(\theta | x)$$

está na classe  $\Pi$  para todo  $x$  sempre que a densidade a priori está em  $\Pi$ .

### 3.3 Método de Jeffreys

O conceito de selecionar uma priori por convenção, como uma “referência de padrão” é devido a Jeffreys. Ele estava crente na existência do estado de ignorância e que o princípio da razão insuficiente era uma maneira formal de expressar tal ignorância.

O ponto de vista de Jeffreys era que a demonstração está certa ou errada como um problema de lógica e não um problema de julgamento pessoal assim como a probabilidade. Em um dado conjunto de dados  $p$  dizemos que uma proposição  $q$  está relacionada a estes dados com uma e somente uma probabilidade. Se qualquer pessoa atribuir uma probabilidade diferente, ela simplesmente está errada, e pela mesma razão como atribuímos no caso do julgamento lógico.

Jeffreys acreditou na existência de um estágio “inicial” em que o investigador é ignorante e achou importante fazer inferência baseada em dados coletados neste estágio. Porém, ele tomou como um conceito primitivo e disse que não havia problema em dizer que a probabilidade expressa um razoável grau de confiança, tendo em vista que a probabilidade não é uma relação lógica unicamente determinada. Assim manteve que sua aproximação não assumiria que somente uma priori fosse logicamente correta.

#### Regras para Prioris em problemas de estimação

Jeffreys considerou diversas situações para formular suas regras tratando cada uma separadamente. O mais simples é o caso de um espaço paramétrico finito em que ele aderiu ao princípio da razão insuficiente a atribuição de probabilidades iguais para cada valor do parâmetro. Então considerou os casos em que o espaço paramétrico fosse: um intervalo limitado, o intervalo  $(-\infty, \infty)$ , ou o intervalo  $(0, \infty)$ . Para intervalos limitados ou para toda linha real, Jeffreys tomou a densidade a priori constante. O segundo caso acarreta que a priori seja imprópria; isto é, que não integra. Ele não considerou este último para não levantar qualquer fundamental dificuldade. Para o terceiro caso, mais comumente associado com um desvio padrão

desconhecido  $\sigma$ , ele usou a priori  $\pi_\sigma(\sigma) = \frac{1}{\sigma}$ . Sua principal justificação para esta escolha foi sua invariância sobre poderosas transformações do parâmetro.

Jeffreys (1946) propõe sua “regra geral”. Escrevendo a matriz de informação de Fisher como  $I(\theta)$ , onde

$$I(\theta)_{ij} = E\left(-\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}\right),$$

em que  $l$  é a log-verossimilhança, a regra é tomar a priori como

$$\pi_\theta(\theta) \propto \det(I(\theta))^{\frac{1}{2}}. \quad (8)$$

Jeffreys notou que esta regra pode contradizer com a regra anterior, que depende do intervalo a qual o parâmetro pertence. Em particular, no caso em que os dados seguem uma distribuição  $N(\mu, \sigma^2)$ , a regra anterior dá  $\pi(\mu, \sigma) = \frac{1}{\sigma}$ , enquanto que a regra geral dá  $\pi(\mu, \sigma) = \frac{1}{\sigma^2}$ . Ele resolveu este problema afirmando que  $\mu$  e  $\sigma$  devem ser julgados independente de uma priori e então trata-los separadamente.

Ele foi além e sugeriu esta modificação para problemas gerais de localização-escala. Também propôs que prioris em problemas envolvendo parâmetros além dos parâmetros de localização e escala sejam tomados tratando parâmetro de localização separadamente do resto. Isto é, se há parâmetros de localização  $\mu_1, \dots, \mu_k$ , e um parâmetro multidimensional adicional  $\theta$ , então a priori que ele recomendou torna-se

$$\pi(\mu_1, \dots, \mu_k, \theta) \propto \det(I(\theta))^{\frac{1}{2}}, \quad (9)$$

onde  $I(\theta)$  é calculado segundo  $\mu_1, \dots, \mu_k$  fixo.

Definição: Nós chamamos (8) e (9) a priori determinada pela regra geral de Jeffreys, tomando o contexto distinto entre os dois casos. Referimos (8) como a priori obtida da regra de não localização de Jeffreys. Assim o que nós chamamos regra de não

locação de Jeffreys é a regra recomendada para não ser aplicada em famílias tendo parâmetro de locação.

### 3.4 Laplace e o Princípio da Razão Insuficiente

Se o espaço paramétrico é finito, então a regra de Laplace ou o princípio da razão insuficiente usa uma priori uniforme que atribua probabilidade igual para cada ponto no espaço paramétrico.

Esta regra é aplicada, mas é sujeita a dualidade: é inconsistente aplicar a regra para todo elemento e refinamento do espaço paramétrico simultaneamente. Shafer deu um exemplo simples. Tome  $\Theta = \{\theta_1, \theta_2\}$ , onde  $\theta_1$  denota o evento que há vida na órbita sobre a estrela Sirius e  $\theta_2$  denota o evento que não há vida. A regra de Laplace dá  $P(\{\theta_1\}) = P(\{\theta_2\}) = \frac{1}{2}$ . Mas agora tome  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ , onde  $\omega_1$  é o evento que há vida em torno de Sirius,  $\omega_2$  é o evento que há planetas mas não há vida, e  $\omega_3$  é o evento que não há planetas. Então a regra de Laplace dá  $P(\{\omega_1\}) = P(\{\omega_2\}) = P(\{\omega_3\}) = \frac{1}{3}$ . O paradoxo é que a probabilidade de vida é  $P(\{\theta_1\}) = \frac{1}{2}$  se nós adotarmos a primeira formulação, mas é  $P(\{\omega_1\}) = \frac{1}{3}$  se adotarmos a segunda formulação.

A dualidade entre conjunto finito e a falta de parametrização invariante estão aproximadamente relacionados. Em ambos os casos nós temos dois espaços,  $\Theta$  e  $\Omega$ , e uma aplicação,  $g: \Omega \rightarrow \Theta$ . Então temos a escolha de adotar uma priori uniforme em  $\Theta$  ou adotar uma priori uniforme  $\mu$  em  $\Omega$ , que então induz uma priori  $\pi$  em  $\Theta$ , onde  $\pi$  é definido por  $\pi(A) = \mu(g^{-1}(A))$ . Em geral,  $\pi$  não será uniforme.

### 3.5 Priori de Haldane

A priori de Haldane é equivalente ao usual argumento da troca de variáveis para uma priori uniforme na log-odds.

Algumas vezes é conveniente trabalhar em termos de odds (probabilidade) de sucesso contra fracasso. E o log-odds é justamente o logaritmo do quociente da probabilidade de sucesso pela probabilidade de fracasso.

### 3.6 Verossimilhança de dados transladados

Segundo BOX e TIAO (1973), o problema é expressar a idéia que pouco é conhecido a priori relativo ao que o dado nos diz sobre um parâmetro  $\theta$ . De um modo geral, se  $\phi(\theta)$  é uma transformação um a um de  $\theta$ , devemos dizer que uma distribuição a priori de  $\theta$  que é localmente proporcional a  $\left| \frac{d\phi}{d\theta} \right|$  é não informativa para o parâmetro  $\theta$  se, em termos de  $\phi$ , a curva de verossimilhança para dados transladados, isto é, os dados servem somente para trocar a locação da verossimilhança  $l(\phi|y)$ . Matematicamente uma verossimilhança de dados transladados será da forma

$$l(\theta|y) = g[\phi(\theta) - f(y)], \quad (12)$$

onde  $g(x)$  é uma função conhecida independente dos dados  $y$  e  $f(y)$  é uma função de  $y$ .

Ainda segundo BOX e TIAO (1973), uma transformação que permita que a verossimilhança seja expressa exatamente na forma (12) geralmente não está disponível. Porém, para amostras de tamanho moderado, por causa da insensibilidade da distribuição a posteriori a menor mudança na priori, tudo o que deve parecer necessário requerer é uma transformação  $\phi(\theta)$  em termos de que a verossimilhança seja aproximadamente dados transladados. Sendo a priori proporcional ao módulo da derivada do parâmetro transformado em relação ao parâmetro de interesse e este proporcional a raiz da informação esperada de Fisher, então implica que a correspondente priori não-informativa para o parâmetro será proporcional a raiz da informação esperada de Fisher.

Visando determinar prioris para casos em que utilize modelos probabilísticos discretos, utilizaremos aqui dados de dez granjas do Estado de Pernambuco obtidos por telefone onde o parâmetro de interesse é a probabilidade de ovos grandes  $\theta$ . Tendo em vista que os ovos são classificados como industrial, pequeno, médio, grande, extra e jumbo, trabalharemos aqui classificando os ovos em pequeno e grande. Os ovos industrial, pequeno e médio serão tidos como pequenos e os ovos grande, extra e jumbo, como grandes.

**Tabela 1.** Dados de ovos de dez granjas de Pernambuco.

	<b>N° de Galinhas</b>	<b>Ovos Pequenos por dia</b>	<b>Ovos Grandes por dia</b>	<b>Total de Ovos por dia</b>
<b>Granja 1</b>	35000	11200	16800	28000
<b>Granja 2</b>	44000	14350	20650	35000
<b>Granja 3</b>	34000	6400	12600	19000
<b>Granja 4</b>	72000	18000	44000	62000
<b>Granja 5</b>	77000	16000	34000	50000
<b>Granja 6</b>	31000	6600	13400	20000
<b>Granja 7</b>	47000	8100	18900	27000
<b>Granja 8</b>	49000	12400	27600	40000
<b>Granja 9</b>	20000	4000	8000	12000
<b>Granja 10</b>	63000	15000	31000	46000

A função de probabilidade neste caso é uma distribuição binomial com a probabilidade de ovos grandes e ovos pequenos,  $\theta$  e  $(1 - \theta)$  respectivamente.

$$p(x | \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad x = 0, 1, \dots, n$$

onde

$$p(x | \theta) \propto \theta^x (1 - \theta)^{n-x}.$$

Se sua priori para  $\theta$  tem a forma

$$p(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1} \quad 0 \leq \theta \leq 1$$

terá uma distribuição Beta  $\theta \sim Be(\alpha, \beta)$ .

A posteriori é

$$p(\theta | x) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1} \theta^x (1-\theta)^{n-x}$$

$$p(\theta | x) \propto \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}$$

isto é,  $\theta | x \sim Be(\alpha + x, \beta + n - x)$ . Isto mostra que a família de distribuição beta é conjugada para verossimilhança binomial.

A distribuição a priori de Laplace é a uniforme sendo assim

$$p(\theta) = \begin{cases} 1 & 0 \leq \theta \leq 1 \\ 0 & \text{caso contr.} \end{cases} \quad p(\theta) = \frac{1}{b-a} \quad ; \quad 0 < \theta < 1.$$

Faremos a construção da priori de Jeffreys

$$L(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad \rightarrow \quad \text{Função de Máxima Verossimilhança}$$

$$l(\theta) = \log[L(\theta)] = \log \left[ \binom{n}{x} \theta^x (1-\theta)^{n-x} \right]$$

$$l(\theta) = \log \left[ \frac{n!}{x!(n-x)!} \right] + \log \theta^x + \log(1-\theta)^{n-x}$$

$$l(\theta) = \log \left[ \frac{n!}{x!(n-x)!} \right] + x \log \theta + (n-x) \log(1-\theta)$$

$l$  é a log-verossimilhança também conhecida como Função Suporte.

$$U(\theta) = \frac{\partial l}{\partial \theta} = \frac{x}{\theta} + \frac{(n-x)}{(1-\theta)} \cdot (-1)$$

$$U(\theta) = \frac{x}{\theta} - \frac{(n-x)}{(1-\theta)} \quad \rightarrow \quad \text{Função Escore}$$

$$J(\theta) = \frac{\partial U}{\partial \theta} = \frac{\theta \cdot 0 - x \cdot 1}{\theta^2} - \frac{(1-\theta) \cdot 0 - (n-x)(-1)}{(1-\theta)^2}$$

$$J(\theta) = -\frac{x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2} \quad \rightarrow \quad \text{Informação Observada de Fisher}$$

Onde temos que  $J(\theta) = \frac{\partial^2 l}{\partial \theta^2}$ .

$$\begin{aligned} I(\theta) &= E\left(-\frac{\partial^2 l}{\partial \theta^2}\right) = E\left[-\left(-\frac{x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2}\right)\right] \\ &= E\left(\frac{x}{\theta^2} + \frac{(n-x)}{(1-\theta)^2}\right) = E\left(\frac{x}{\theta^2}\right) + E\left(\frac{n-x}{(1-\theta)^2}\right) \\ &= \frac{E(x)}{\theta^2} + \frac{n - E(x)}{(1-\theta)^2} = \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1-\theta)^2} \\ &= \frac{n}{\theta} + \frac{n(1-\theta)}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n(1-\theta) + n\theta}{\theta(1-\theta)} \\ &= \frac{n(1-\theta + \theta)}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)} = n\theta^{-1}(1-\theta)^{-1} \end{aligned}$$

$$I(\theta) = n\theta^{-1}(1-\theta)^{-1} \quad \rightarrow \quad \text{Informação Esperada de Fisher}$$

Neste caso a priori de Jeffreys será

$$p(\theta) \propto [I(\theta)]^{\frac{1}{2}} = [\theta^{-1}(1-\theta)^{-1}]^{\frac{1}{2}} = \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$



Para calcularmos a priori pela verossimilhança de dados transladados temos que esta coincide com a priori de Jeffreys já calculada que foi

$$p(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$

onde para sabermos que transformação foi utilizada basta calcular  $\phi \propto \int_{\pi} J^{1/2}(t) dt$ .

$$\int_{\theta} t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}} dt \propto \text{sen}^{-1}\sqrt{\theta} \rightarrow \phi \propto \text{sen}^{-1}\sqrt{\theta}.$$

$$\int_{\theta} t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}} dt = \int_{\theta} \frac{1}{t^{\frac{1}{2}}(1-t)^{\frac{1}{2}}} dt$$

Fazendo a substituição:  $\begin{cases} t^{\frac{1}{2}} = u \\ \frac{1}{2}t^{-\frac{1}{2}} dt = du \end{cases} \rightarrow \begin{cases} t = u^2 \\ \frac{dt}{t^{\frac{1}{2}}} = 2du \end{cases}$

$$= \int_{\theta} \frac{2du}{(1-u^2)^{\frac{1}{2}}} = 2 \int_{\theta} \frac{du}{\sqrt{1-u^2}} = 2 \arcsen(u)^{\theta} = 2 \arcsen\left(t^{\frac{1}{2}}\right)^{\theta}$$

$$\int_{\theta} \frac{dt}{t^{\frac{1}{2}}(1-t)^{\frac{1}{2}}} = 2 \arcsen\sqrt{\theta}$$

$$\therefore \phi = \int_{\theta} t^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}} dt \propto \arcsen\sqrt{\theta} = \text{sen}^{-1}\sqrt{\theta}$$

Esta distribuição é algumas vezes chamada a distribuição arco seno segundo Lee (2004).

Para a priori de Haldane temos a troca de variável

$$\phi = \log\left\{\frac{\theta}{1-\theta}\right\}$$

Ao qual teremos a priori  $p(\theta) \propto \left| \frac{d\phi}{d\theta} \right|$

$$\left| \frac{d\phi}{d\theta} \right| = \frac{1}{\left( \frac{\theta}{1-\theta} \right)} \cdot \frac{(1-\theta) \cdot 1 - \theta(-1)}{(1-\theta)^2}$$

$$= \frac{(1-\theta)}{\theta} \cdot \left\{ \frac{1-\theta+\theta}{(1-\theta)^2} \right\}$$

$$\left| \frac{d\phi}{d\theta} \right| = \theta^{-1}(1-\theta)^{-1}$$

$$p(\theta) \propto \theta^{-1}(1-\theta)^{-1}$$

#### 4. RESULTADOS E DISCUSSÃO

Tendo como objetivo estimar a proporção de ovos comerciais, tomamos uma amostra de dez granjas do Estado de Pernambuco e utilizando uma abordagem bayesiana, obteve-se prioris não-informativas através de quatro métodos: Método de Laplace, Método de Jeffreys, Método da verossimilhança de dados transladados e Método de Haldane. Observamos que os métodos de Jeffreys e verossimilhança de dados transladados são coincidentes utilizou-se para este caso o método de Jeffreys.

Para cada um dos métodos utilizamos 20.000 iterações em que as dez mil primeiras foram descartadas e de 10.001 a 20.000 forneceram as estimativas dos parâmetros.

Pelos gráficos em anexo observa-se que: a moda a posteriori (figura 1a, 1b, 1c) não teve grande variabilidade para as últimas duzentas iterações (19.800; 20.000), e para a distribuição a posteriori do quartil (figura 2a, 2b, 2c) se observou um melhor equilíbrio da cadeia a partir de 12.500 iterações sendo suficiente 20.000 iterações.

Com base na realização de 10.000 iterações foram efetuadas as estimativas das médias, desvio padrão, IC 95% e calculada a amplitude para os métodos de Laplace, Jeffreys e Haldane na tabela 2, 3 e 4 respectivamente.

**Tabela2.** Resultados pelo método de Laplace das estimativas do parâmetro  $p$ , seu desvio padrão e IC 95%, realizando 10000 iterações no Winbugs14.

	Média	Desvio padrão	IC 95% para $p$		Amplitude
			Inferior	Superior	
P1	0,6	0,002901	0,5943	0,6057	0,0114
P2	0,59	0,002608	0,5849	0,5951	0,0102
P3	0,6631	0,003395	0,6566	0,6697	0,0131
P4	0,7097	0,001806	0,7061	0,7131	0,0070
P5	0,68	0,002079	0,6759	0,684	0,0081
P6	0,67	0,003319	0,6634	0,6765	0,0131
P7	0,7	0,002823	0,6945	0,7054	0,0109
P8	0,69	0,002336	0,6854	0,6945	0,0091
P9	0,6666	0,00435	0,6581	0,6751	0,0170
P10	0,674	0,002182	0,6696	0,6782	0,0086

**Tabela3.** Resultados pelo método de Jeffreys das estimativas do parâmetro  $p$ , seu desvio padrão e IC 95%, realizando 10000 iterações no Winbugs14.

	<i>Média</i>	<i>Desvio padrão</i>	<i>IC 95% para p</i>		<i>Amplitude</i>
			<i>Inferior</i>	<i>Superior</i>	
P1	0,6	0,002901	0,5943	0,6057	0,0114
P2	0,59	0,002608	0,5849	0,5951	0,0102
P3	0,6631	0,003395	0,6566	0,6698	0,0132
P4	0,7097	0,001806	0,7061	0,7131	0,0070
P5	0,68	0,002079	0,6759	0,684	0,0081
P6	0,67	0,003319	0,6634	0,6765	0,0131
P7	0,7	0,002824	0,6945	0,7054	0,0109
P8	0,69	0,002336	0,6854	0,6945	0,0091
P9	0,6666	0,00435	0,6582	0,6751	0,0169
P10	0,674	0,002182	0,6696	0,6782	0,0086

**Tabela4.** Resultados pelo método de Haldane das estimativas do parâmetro  $p$ , seu desvio padrão e IC 95%, realizando 10000 iterações no Winbugs14.

	<i>Média</i>	<i>Desvio padrão</i>	<i>IC 95% para p</i>		<i>Amplitude</i>
			<i>Inferior</i>	<i>Superior</i>	
P1	0,6	0,002887	0,5943	0,6056	0,0113
P2	0,59	0,002654	0,5847	0,5952	0,0105
P3	0,6631	0,003426	0,6565	0,6699	0,0134
P4	0,7097	0,001828	0,7061	0,7132	0,0071
P5	0,68	0,002067	0,6759	0,6841	0,0082
P6	0,6699	0,003362	0,6631	0,6765	0,0134
P7	0,7	0,002793	0,6947	0,7056	0,0109
P8	0,69	0,002324	0,6853	0,6945	0,0092
P9	0,6667	0,004309	0,6583	0,6751	0,0168
P10	0,6739	0,002142	0,6697	0,6781	0,0084

Observa-se que a estimativa média do parâmetro  $p$  tem valor semelhante nos métodos de Laplace, Jeffreys e Haldane, correspondendo a aproximadamente  $\bar{p} = 0,664$ . Com isso nota-se que a média pertence aos intervalos das amostras provenientes de P3, P6 e P9 para os três métodos, sendo: P3 (0,6566; 0,6697) , P6 (0,6634; 0,6765) e P9 (0,6581; 0,6751) para o método de Laplace; P3 (0,6566; 0,6698) , P6 (0,6634; 0,6765) e P9 (0,6582; 0,6751) para o método de Jeffreys; e P3 (0,6565; 0,6699) , P6 (0,6631; 0,6765) e P9 (0,6583; 0,6751) para o método de Haldane.

Notamos que os três maiores desvios padrões para os métodos de Laplace e Jeffreys são  $P9 \rightarrow 0,00435$  ,  $P3 \rightarrow 0,003395$  ,  $P6 \rightarrow 0,003319$  ; e para o método de Haldane temos  $P9 \rightarrow 0,004309$  ,  $P3 \rightarrow 0,003426$  ,  $P6 \rightarrow 0,003362$ .

Temos que a média do parâmetro encontra-se nos três intervalos de maior amplitude e maior desvio padrão, para cada método.

Considerando as densidades a posteriori dos parâmetros observamos através dos gráficos do anexo (figura 3a, 3b, 3c) que a distribuição tende para uma distribuição normal.

## 5. CONCLUSÕES

- Utilizando aproximadamente 19% do total de granjas de Pernambuco (10 granjas), a média da probabilidade de ovos comerciais (grandes) varia em torno de 66,4%;
- A probabilidade de ovos comerciais varia de 58% à 71%;
- Concluímos que as estimativas da probabilidade de ovos comerciais (grande) para os três métodos são aproximadamente iguais.
- Os intervalos de credibilidade que estimam a probabilidade de ovos grandes estão entre 0,58 à 0,71.

## 6. RECOMENDAÇÕES

- Recomendamos esses métodos para aplicação em outras áreas.
- Aconselhamos a necessidade de aprofundamento científico através de novas pesquisas incluindo variáveis que possam auxiliar no estudo para determinar a probabilidade de ovos comerciais. Outras variáveis que possam contribuir para melhor estimar tal probabilidade, tais como: A linhagem das aves, o sistema de criação (semi-confinado, em gaiolas, em box), o manejo adotado, e localização.
- Pode-se também utilizar outras prioris não-informativas tais como: as obtidas pelos métodos de entropia máxima, método de Berger-Bernardo, método de Zellner entre outros.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

ABUSEV, R. A.; KOLEGOVA, N. V. Comparison of statistical estimators for the parameters of some distributions. **Journal of Mathematical Sciences**. v. 103, n. 3, p. 449-454, 2001.

ALEXANDER, N. et al. Spatial modelling of individual-level parasite counts using the negative binomial distribution. **Biostatistics**. v. 1, n. 4, p. 453-463, 2000.

BARTOLUCCI, F. Efficient Bayes factor estimation from the reversible jump output. **Biometrika**. v. 93, n. 1, p. 41-52, 2006.

BOLSTAD, W. M. **Introduction to Bayesian Statistics**. John Wiley & Sons, Inc. Publication. 362 p. 2004.

BOX, G. E. P., TIAO, G.C. **Bayesian inference in statistical analysis**. Addison-Wesley, 588p. 1973.

BORGATTO, A. F. **Modelos para proporções com superdispersão e excesso de zeros – um procedimento bayesiano**. São Paulo, 2004. 104p. Tese (Doutorado em Estatística Experimental) – Escola Superior de Agricultura “Luiz de Queiros” da Universidade de São Paulo. Piracicaba.

COELHO, ALEXANDRE S. G. **Abordagem Bayesiana na análise genética de populações utilizando dados de marcadores moleculares**. São Paulo, 2002. 97p. Tese (Doutorado em Estatística Experimental) – Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo. Piracicaba.

DATTA, G. S.; GHOSH, M. On the invariance of noninformative priors. **The Annals of Statistics**. v. 24, n. 1, p. 141-159, 1996.

EHLERS, RICARDO S. **Introdução à Inferência Bayesiana**. Texto publicado pela Universidade Federal do Paraná. Departamento de Estatística, UFPR. 77p. 2005.



FREIRE, J. C. M. **Métodos para estimar “ODDS RATIO” através de abordagens clássica e bayesiana, em estudos epidemiológicos.** Recife, 2001. 62p. Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco.

KASS, R. E; WASSERMAN, L. The selection of prior distributions by formal rules. **Journal of the American Statistical Association.** v. 91, n. 435, p. 1343-1370, 1996.

LARSON, H. J. **Introduction to Probability Theory And Statistical Inference.** John Wiley & Sons. 637p. 1982.

LEE, S.; HWANG, W.; HUANG, L. Bayes estimation of population size from capture-recapture models with time variation and behavior response. **Statistica Sinica.** v. 13, p. 477-494, 2003.

LEE, P. M. **Bayesian Statistics – An Introduction.** Oxford University Press Inc. 351p. 2004.

NEVES, C. R.; MIGON, H. S. Graduação bayesiana de taxas de mortalidade: uma aplicação na cobertura de sobrevivência e na avaliação da provisão matemática. **Revista Brasileira de Risco e Seguro,** v. 1, n. 0, p. 90-111, 2004.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana.** Fundação Calouste Gulbenkian / Lisboa. 446p. 2003.

PIRES, M. C. **Análise Bayesiana Empírica de Dados Dicotômicos com Erros e Classificações Repetidas.** Belo Horizonte, 2006. 18p. Dissertação (Mestrado em Estatística) – Instituto de Ciências Exatas da Universidade Federal de Minas Gerais.

POLLARD, W. E. **Bayesian statistics for evaluation research An Introduction.** Sage Publications New Delhi. 241p. 1986.

SILVA, V. A. **Uma abordagem bayesiana para estimativa da densidade populacional de animais de vida livre.** Recife, 2002. 38p. Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco.

SOUZA, A. D. P. **Fundamentos de Inferência Bayesiana e Aplicações**. Texto dos Cursos de Matemática, Estatística e Ciência da Computação – FCT/UNESP. 35p. 2003.

ZACHARIAS, H. P.; LEITE, J. G.; DINIZ, C. A. R. O Uso de Prioris Não Informativas para Estimação do Tamanho Populacional. **Uma Publicação da Sociedade Brasileira de Matemática Aplicada e Computacional**. N° 1, p. 165–175, 2004.

QUEIROZ, N. **Regressão logística – uma estimativa bayesiana aplicada na identificação de fatores de risco para HIV, em doadores de sangue**. Recife, 2004. 97p. Dissertação (Mestrado em Biometria) – Universidade Federal Rural de Pernambuco.

## 8. ANEXO

Os ovos são classificados através do decreto 99.427 de 1990, publicado em 1991.

<b>Tipo</b>	<b>Gramas</b>
Jumbo	acima de 66
Extra	60 a 65
Grande	55 a 60
Médio	50 a 55
Pequeno	45 a 50
Industrial	abaixo de 45

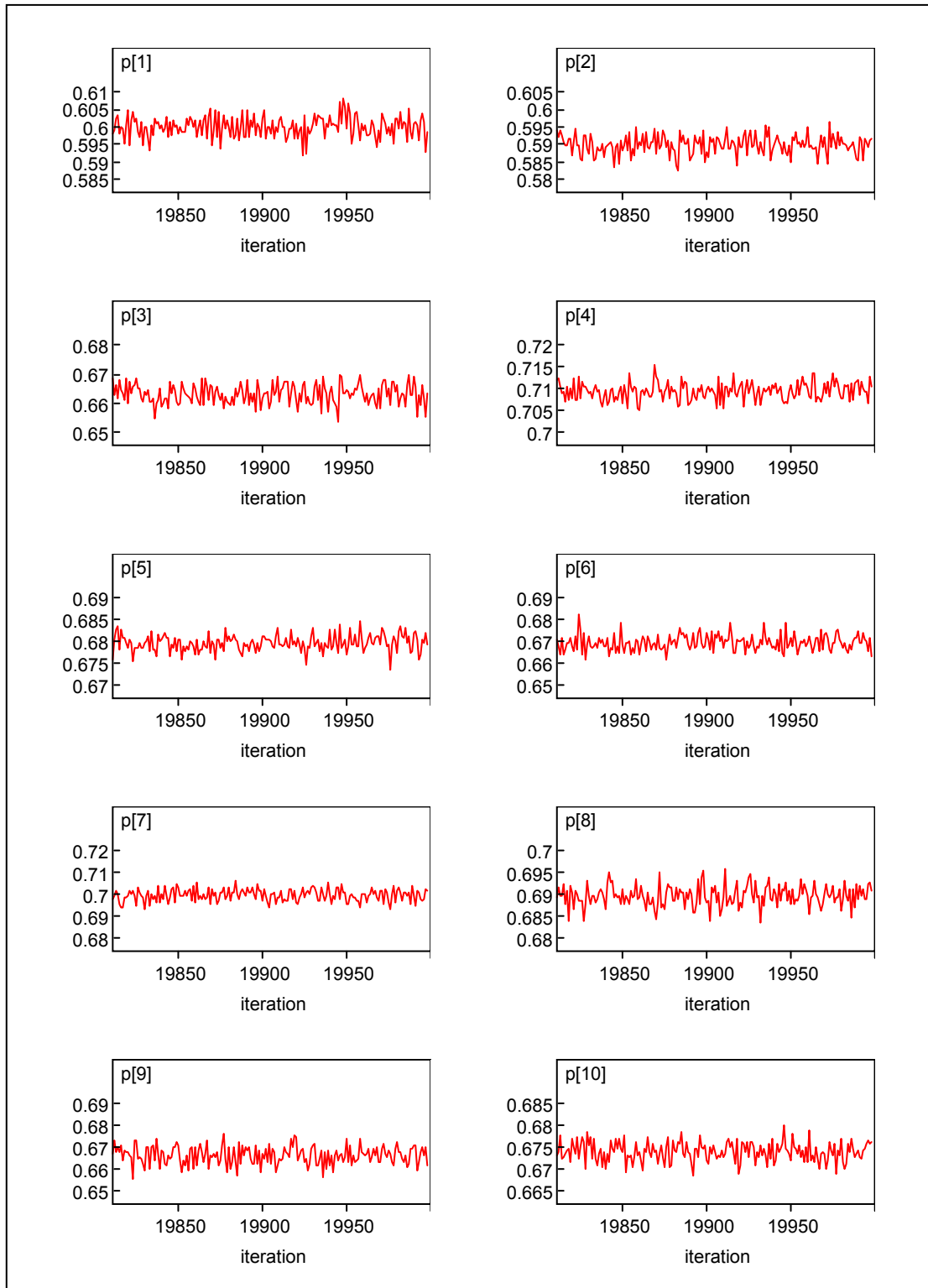
Os ovos deste trabalho foram classificados como:

$$\begin{array}{cc}
 \left. \begin{array}{l} \textit{jumbo} \\ \textit{extra} \\ \textit{grande} \end{array} \right\} \textit{Grande} & \left. \begin{array}{l} \textit{médio} \\ \textit{pequeno} \\ \textit{industrial} \end{array} \right\} \textit{Pequeno} .
 \end{array}$$

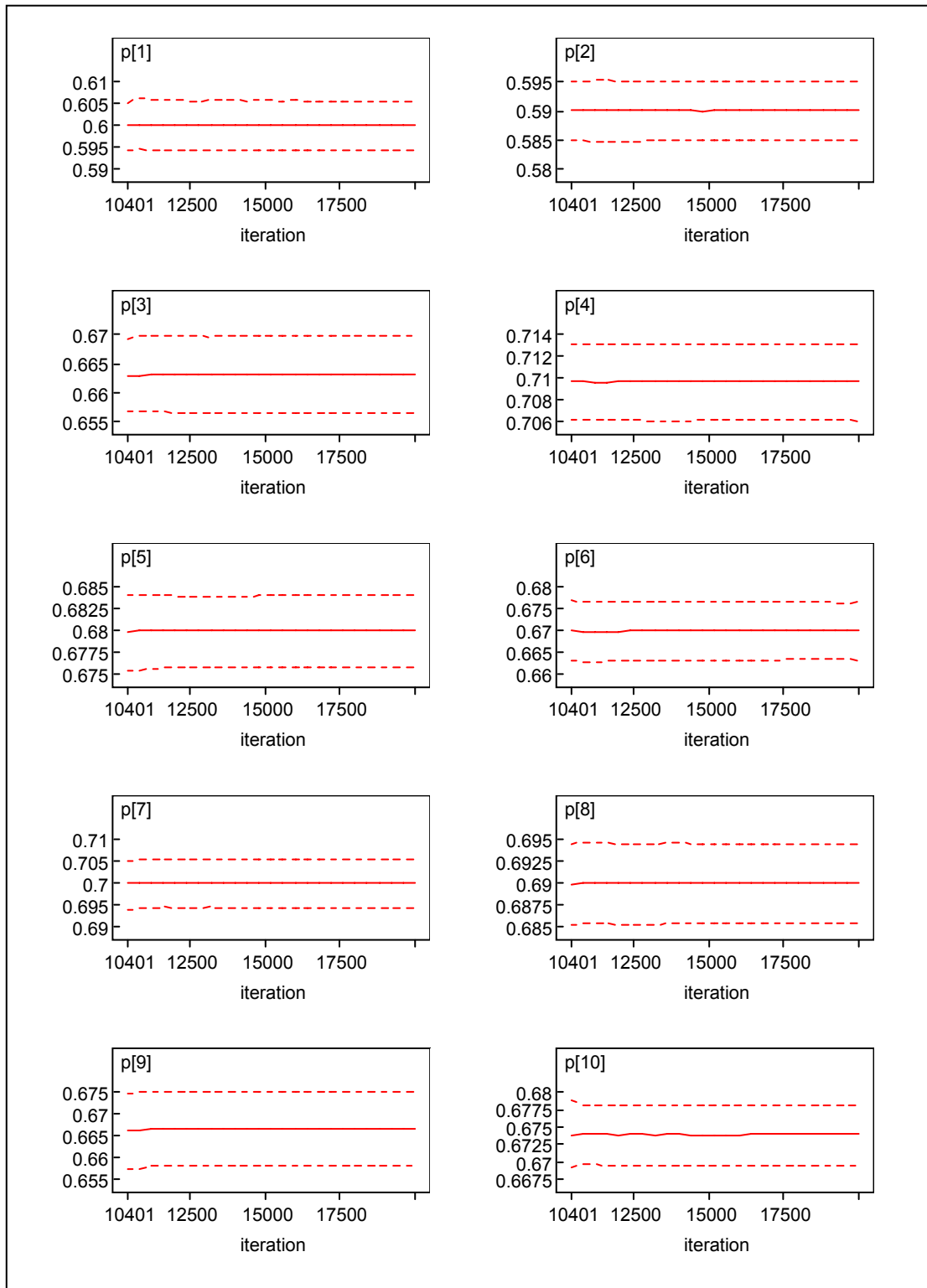
Onde pela classificação temos que até 55 gramas são considerados ovos pequenos e acima de 55 gramas teremos ovos grandes.

Está em anexo os gráficos da densidade a posteriori e as gráficos de convergência: moda e quantis.

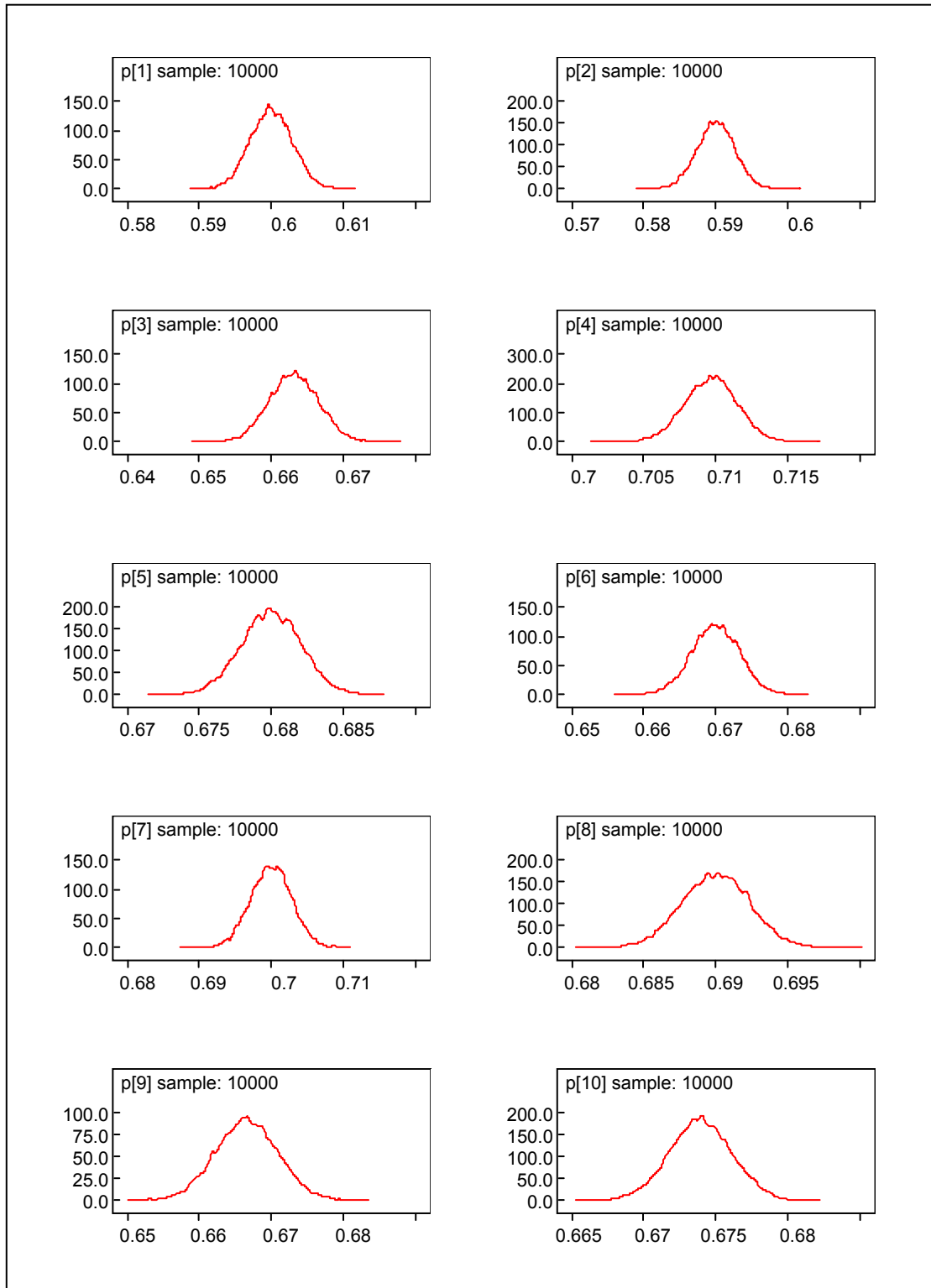
Gráficos da moda (figura 1a), quartil da distribuição a posteriori (figura 2a) e densidade posteriori (figura 3a) pelo método de Laplace.



**Figura 1a.** Gráfico da moda dos parâmetros  $p$  para o Método de Laplace, realizando-se 10000 iterações, no Winbugs 14.

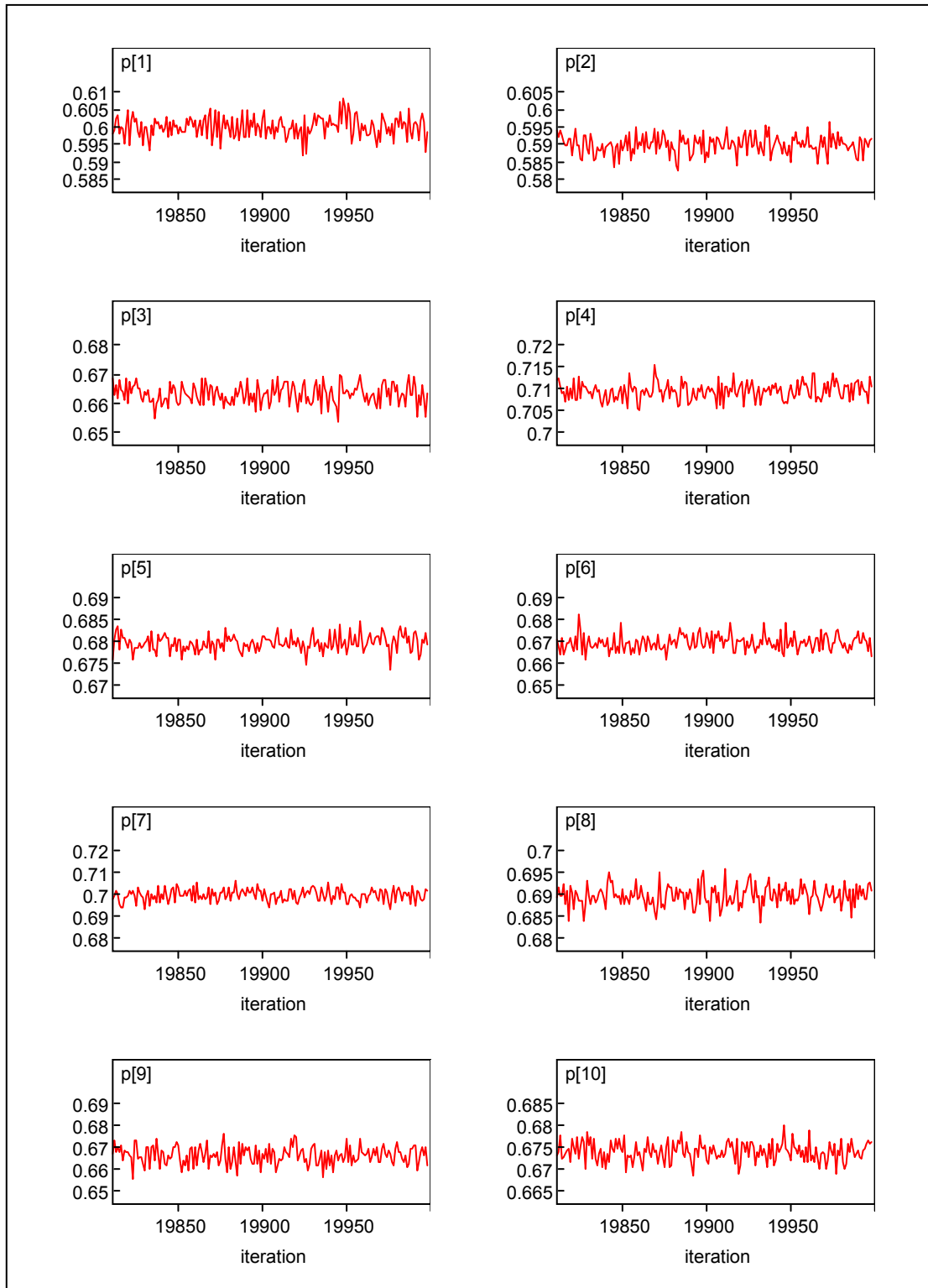


**Figura 2a.** Gráfico do quartil a posteriori dos parâmetros  $p$  para o Método de Laplace, realizando-se 10000 iterações, no Winbugs 14.

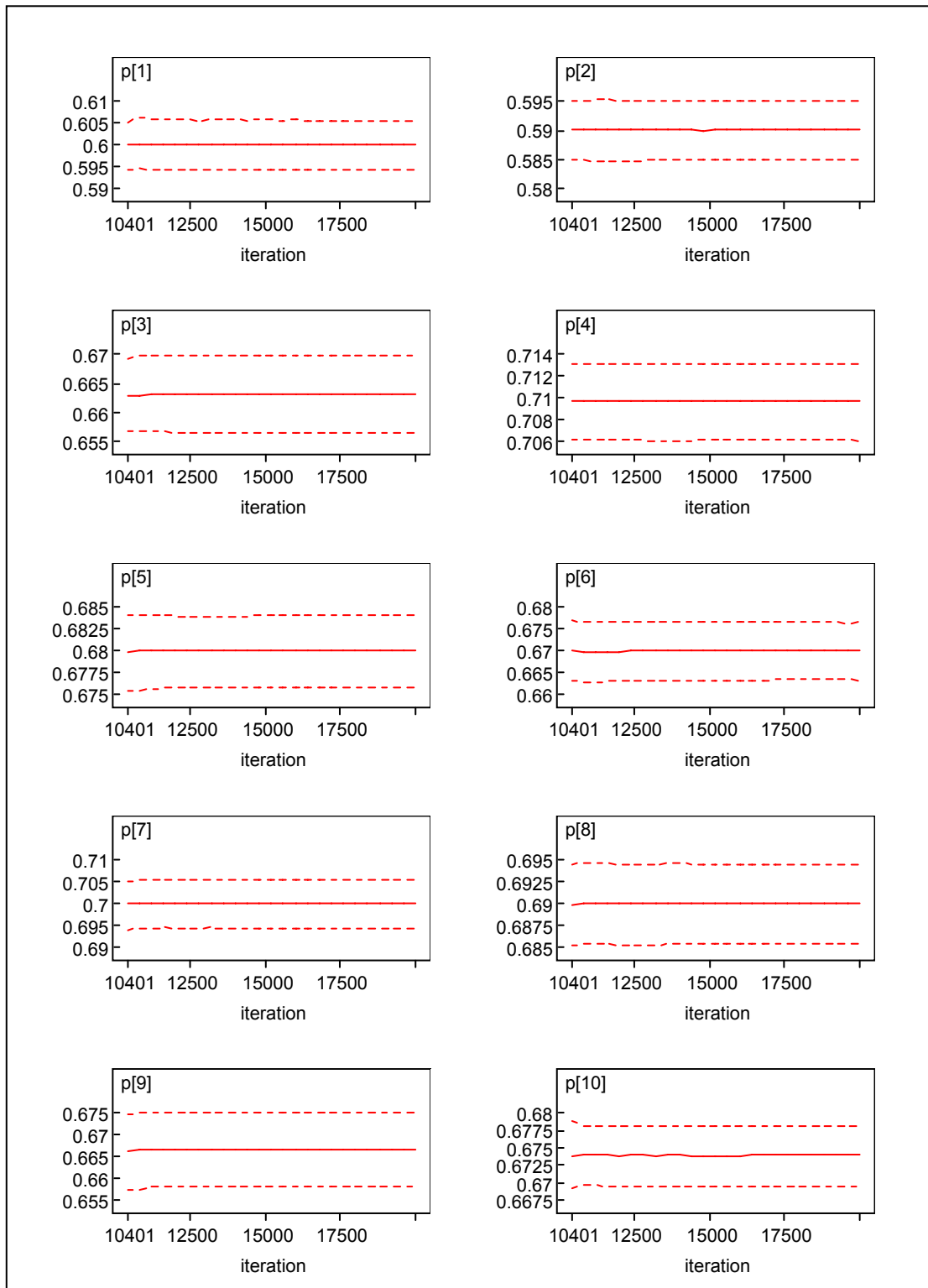


**Figura 3a.** Gráfico da densidade a posteriori dos parâmetros  $p$  para o Método de Laplace, realizando-se 10000 iterações, no Winbugs 14.

Gráfico da moda (figura 1b), quartil da distribuição a posteriori (figura 2b) e densidade posteriori (figura 3b) pelo método de Jeffreys.

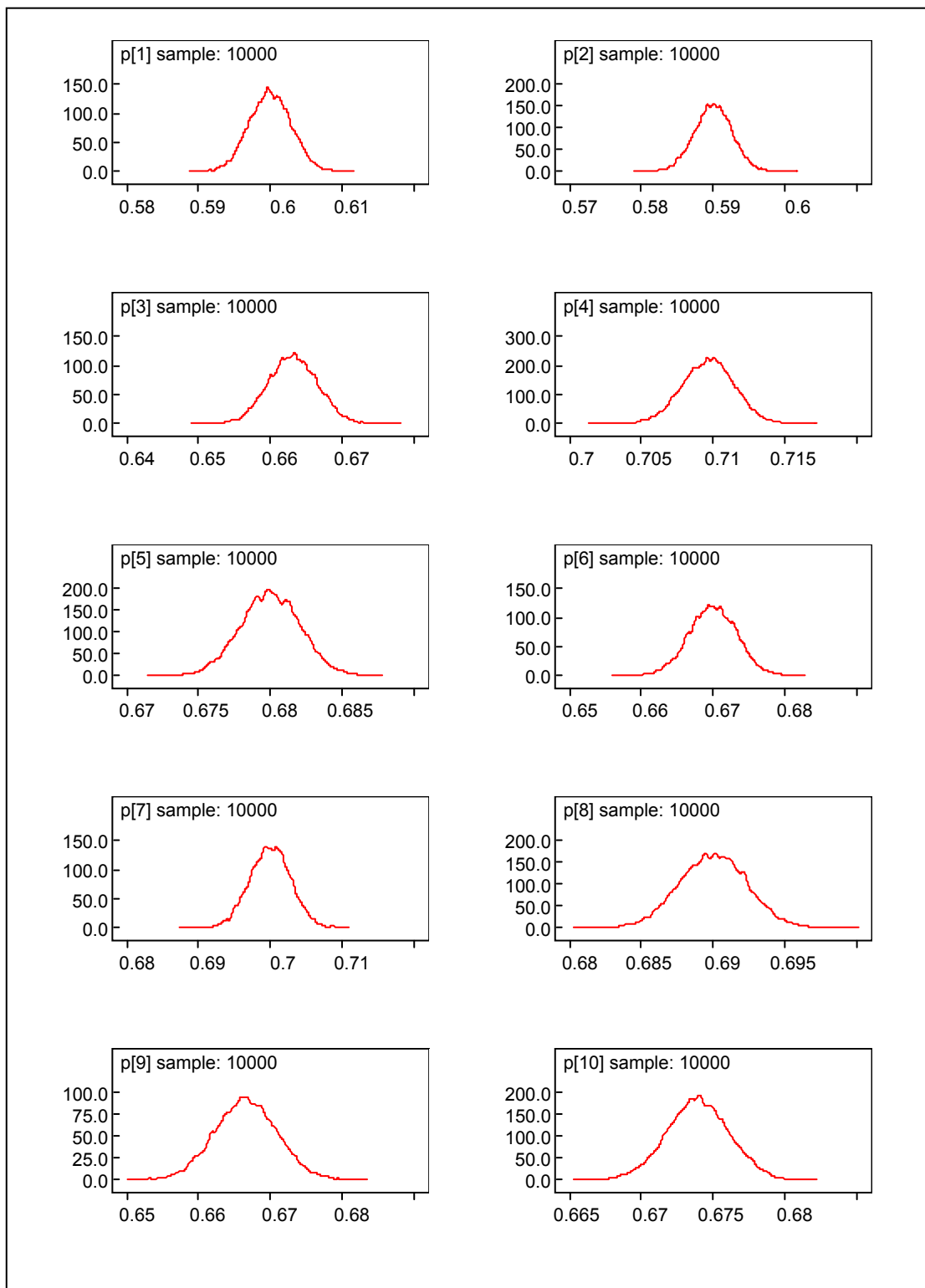


**Figura 1b.** Gráfico da moda dos parâmetros  $p$  para o Método de Jeffreys, realizando-se 10000 iterações, no Winbugs 14.



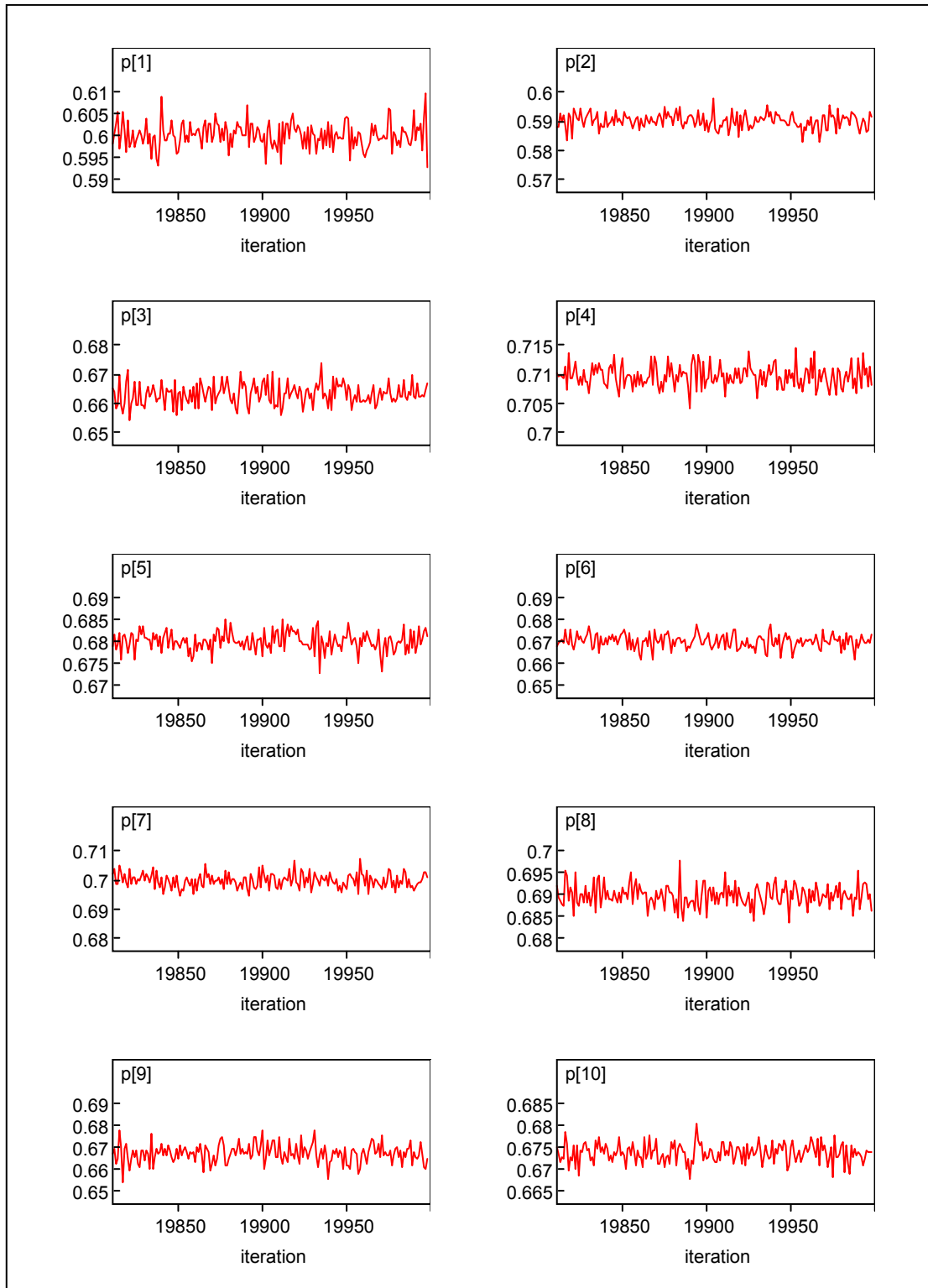
**Figura 2b.** Gráfico do quartil a posteriori dos parâmetros  $p$  para o Método de Jeffreys, realizando-se 10000 iterações, no Winbugs 14.



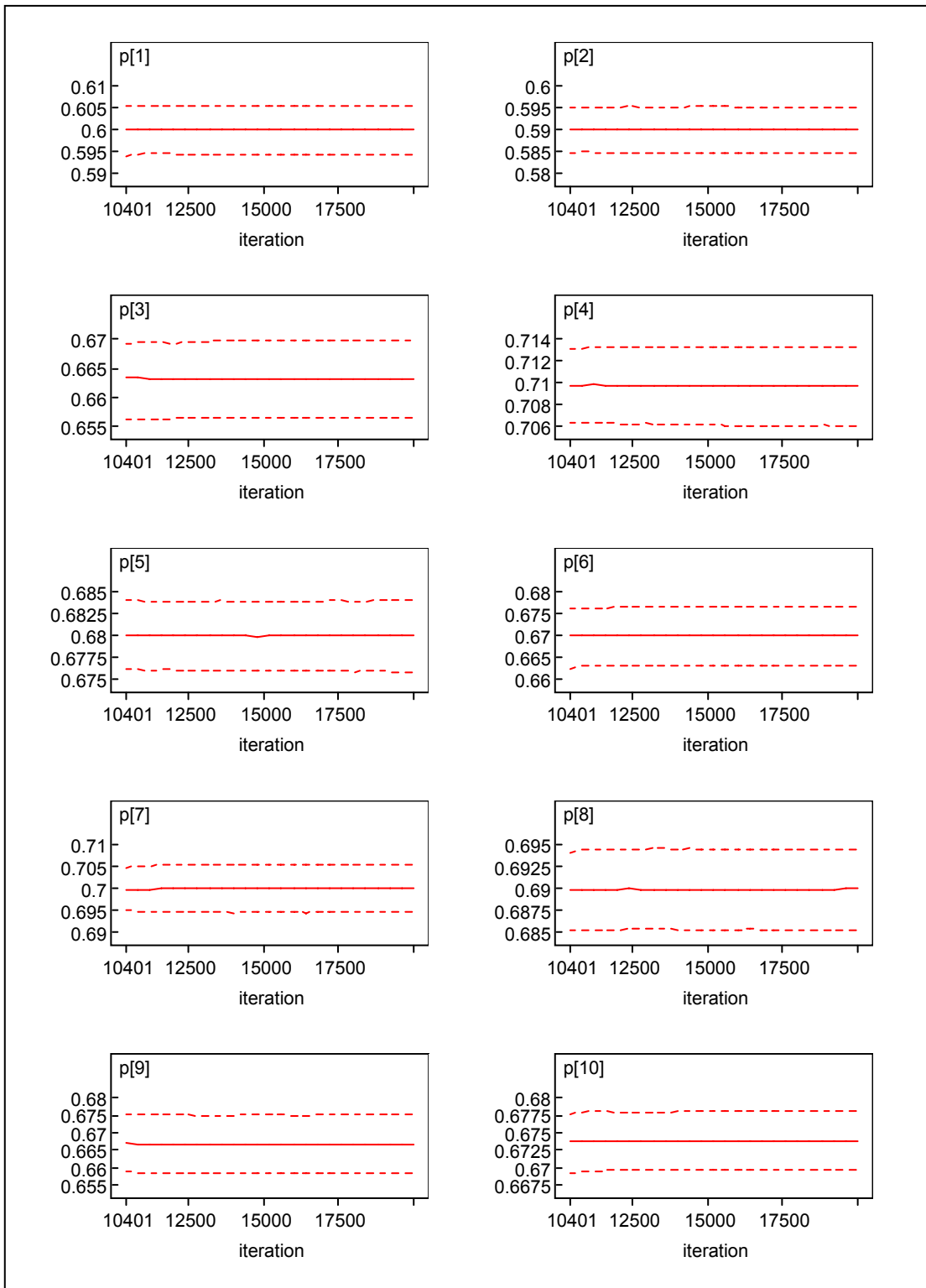


**Figura 3b.** Gráficos das densidades a posteriori dos parâmetros  $p$  para o Método de Jeffreys, realizando-se 10000 iterações, no Winbugs 14.

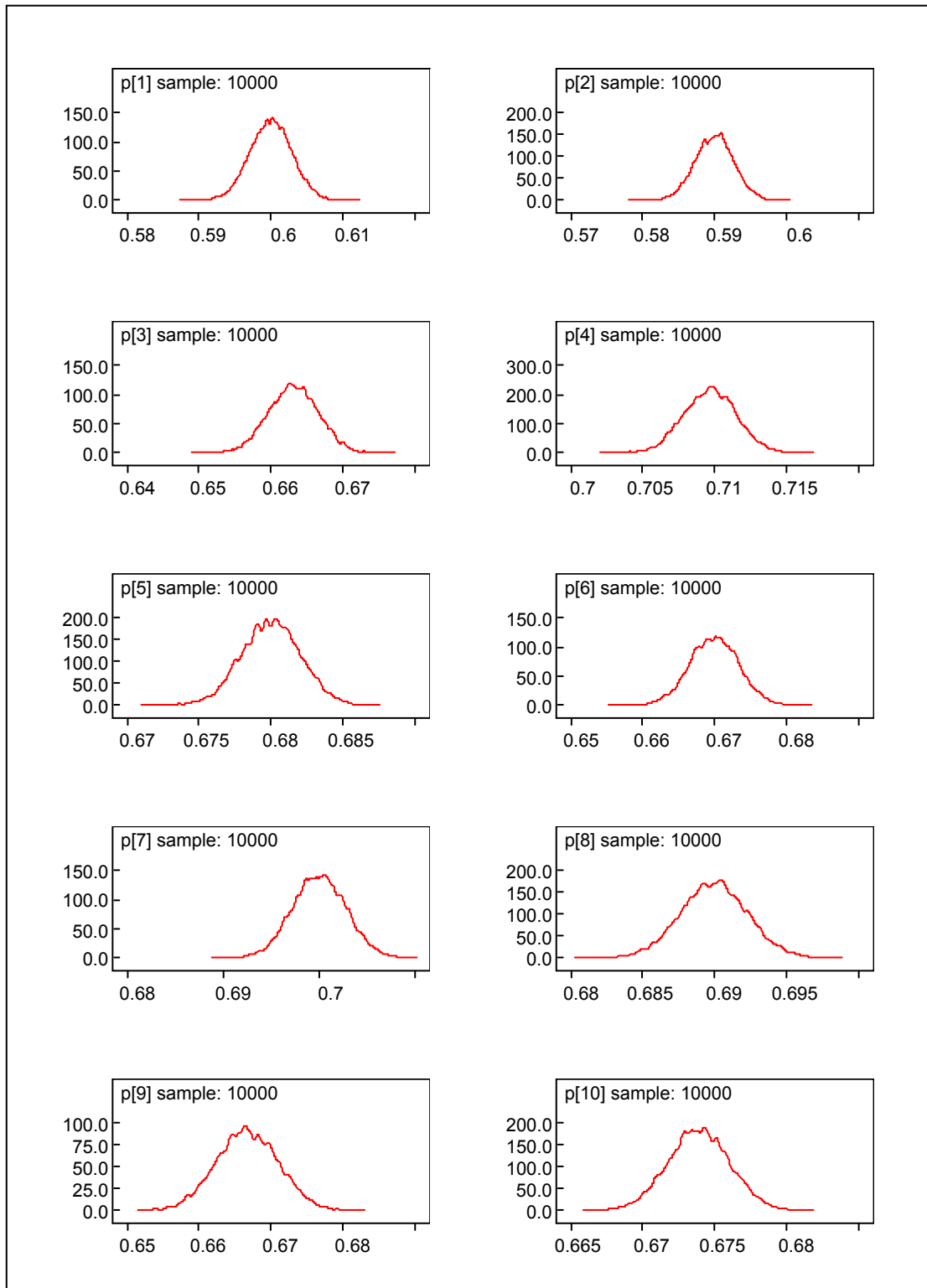
Gráficos da moda (figura 1c), quartil da distribuição a posteriori (figura 2c) e densidade posteriori (figura 3c) pelo método de Haldane.



**Figura 1c.** Gráficos da moda dos parâmetros  $p$  para o Método de Haldane, realizando-se 10000 iterações, no Winbugs 14.



**Figura 2c.** Gráfico do quartil a posteriori dos parâmetros  $p$  para o Método de Haldane, realizando-se 10000 iterações, no Winbugs 14.



**Figura 3c.** Gráficos das densidades a posteriori dos parâmetros  $p$  para o Método de Haldane, realizando-se 10000 iterações, no Winbugs 14.