

George Lucas Moraes Pezzott

Estimação do tamanho populacional a partir
de um modelo de captura-recaptura com
heterogeneidade

São Carlos
2014

George Lucas Moraes Pezzott

Estimação do tamanho populacional a partir
de um modelo de captura-recaptura com
heterogeneidade

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Luis Ernesto Bueno Salasar

Co-orientador: Prof. Dr. José Galvão Leite

São Carlos
2014

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

P522et Pezzott, George Lucas Moraes.
Estimação do tamanho populacional a partir de um modelo de captura-recaptura com heterogeneidade / George Lucas Moraes Pezzott. -- São Carlos : UFSCar, 2014.
91 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2014.

1. Inferência (Estatística). 2. Processo sequencial de captura-recaptura. 3. Modelo de mistura. 4. Em-algoritmo. 5. *Gibbs sampling*. I. Título.

CDD: 519 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Estatística
Via Washington Luís, Km 235 - C.P.676 - CGC 45358058/0001-40
FONE: (016) 3351-8292 – Email: ppgest@ufscar.br
13565-905 - SÃO CARLOS-SP - BRASIL

FOLHA DE APROVAÇÃO

Aluno(a) : George Lucas Moraes Pezzott

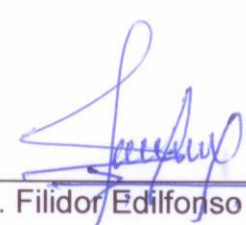
DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 14/03/2014
PELA COMISSÃO JULGADORA:

Presidente




Prof. Dr. Luis Ernesto Bueno Salasar (DEs-UFSCar/Orientador)

1º Examinador



Prof. Dr. Filidor Edilfonso Vilca Labra (UNICAMP)

2º Examinador



Prof. Dr. José Galvão Leite (DEs-UFSCar/Co-orientador)

3º Examinador



Prof. Dr. Luis A. Milan (DEs-UFSCar)

Agradecimentos

Primeiramente, agradeço a Deus por estar sempre comigo.

À minha mãe Edna, meu pai Lauro e meus irmãos Roger e Julio, pela força e incentivo que me deram em toda minha vida.

À minha namorada Vanessa, pela amizade, carinho, companheirismo e apoio nos momentos mais difíceis.

Ao professor Dr. Luis Ernesto pela orientação, ideias, incentivo, paciência, comprometimento e a amizade que construímos durante o mestrado.

Ao meu co-orientador professor Dr. José Galvão Leite, pelas ideias, correções e sugestões feitas neste trabalho e o privilégio de trabalhar com um doutor tão prestigiado na área acadêmica.

Aos meus amigos e colegas do PPGEST, por todo conhecimento compartilhado e principalmente pelo futebol semanal que proporcionava muitas risadas.

Aos professores e funcionários do Departamento de Estatística da Universidade Federal de São Carlos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro concedido durante este trabalho.

Resumo

Neste trabalho, consideramos a estimação do número de erros em um *software* provenientes de uma população fechada. O processo de estimação do tamanho populacional é baseado no método de captura-recaptura, que consiste em examinar o *software*, em paralelo, por certo número de revisores. O modelo probabilístico adotado acomoda situações em que os revisores são independentes e homogêneos (igualmente eficientes) e que cada erro é um elemento que faz parte de uma partição disjunta quanto à sua probabilidade de detecção. Propomos um processo iterativo para obtenção das estimativas de máxima verossimilhança em que utilizamos o algoritmo EM na estimação dos parâmetros perturbadores. As estimativas dos parâmetros populacionais também foram obtidas sob o enfoque Bayesiano, onde utilizamos simulações de Monte Carlo em Cadeias de Markov (MCMC) através do algoritmo Gibbs sampling com a inserção de variáveis latentes nas distribuições condicionais a *posteriori*. As duas abordagens foram aplicadas em dados simulados e em dois conjuntos de dados reais da literatura.

Palavras-chave: Modelo de captura-recaptura, Modelo de mistura, Erros de *software*, Algoritmo EM, Algoritmo Gibbs sampling.

Abstract

In this work, we consider the estimation of the number of errors in a software from a closed population. The process of estimating the population size is based on the capture-recapture method which consists of examining the software, in parallel, by a number of reviewers. The probabilistic model adopted accommodates situations in which reviewers are independent and homogeneous (equally efficient), and each error is an element that is part of a disjoint partition in relation to its detection probability. We propose an iterative process to obtain maximum likelihood estimates in which the EM algorithm is used to the nuisance parameters estimation. The estimates of population parameters were also obtained under the Bayesian approach, in which Monte Carlo on Markov Chains (MCMC) simulations through Gibbs sampling algorithm with insertion of latent variables were used on the conditional posterior distributions. The two approaches were applied to simulated data and in two real data sets from the literature.

Keywords: Capture-recapture model, Mixture model, Software errors, EM algorithm, Gibbs sampling algorithm.

Sumário

1	Introdução	1
2	Modelo de captura-recaptura	5
2.1	Modelo probabilístico com heterogeneidade e independência entre erros e entre revisores	5
2.1.1	Caso particular: homogeneidade entre revisores	8
3	Estimativas de máxima verossimilhança: erros de fácil e de difícil detecção	13
3.1	Estimativas de máxima verossimilhança	13
3.1.1	O Método EM	14
3.1.2	Variáveis latentes e função de verossimilhança completa	17
3.1.3	Implementação do algoritmo	22
3.2	Estimativas de máxima verossimilhanças condicionais	23
3.2.1	Variáveis latentes e função de verossimilhança condicional completa	26
3.2.2	Implementação do algoritmo	28
3.3	Resultados numéricos	29
3.3.1	Dados artificiais	29
3.3.2	Aplicação em dados reais	36
4	Abordagem Bayesiana: erros de fácil e de difícil detecção	40
4.1	Distribuições condicionais <i>a posteriori</i>	42
4.1.1	Distribuições condicionais <i>a posteriori</i> conhecidas para o tamanho populacional	43
4.1.2	<i>Posterioris</i> condicionais completas para os parâmetros perturbadores	45

4.2	Implementação do algoritmo	47
4.3	Resultados numéricos	47
4.3.1	Exemplo 1: Revisão de interruptores AT&T 5 ESS	48
4.3.2	Exemplo 2: Contagem de coelhos	50
5	Modelo probabilístico: caso com l níveis de dificuldade	55
5.1	Estimativas de máxima verossimilhança	56
5.1.1	Implementação do algoritmo	58
5.1.2	Resultados numéricos	59
5.2	Modelo bayesiano	69
5.2.1	Variáveis latentes e <i>posteriori</i> condicional para os parâmetros perturbadores	70
5.2.2	Implementação do Algoritmo	71
5.2.3	Resultados numéricos	72
6	Considerações finais	80
	Referências Bibliográficas	81
A	Apêndice	86

Capítulo 1

Introdução

O método de captura-recaptura é um procedimento estatístico inferencial frequentemente usado quando se deseja estimar o número de indivíduos (elementos) de uma dada população. O método consiste em selecionar um número fixo ou aleatório de indivíduos desta população em diferentes épocas (ocasiões) de amostragem. Na primeira ocasião, uma amostra é retirada, os indivíduos capturados recebem uma marca e, em seguida, todos são devolvidos à população. Após um certo período de tempo, é selecionada uma segunda amostra e realizada a contagem dos indivíduos marcados (recapturas), e aqueles não marcados recebem uma marca, e todos são devolvidos à população. Este procedimento é repetido em k ($k \geq 2$) ocasiões de amostragem, e em cada ocasião é realizada a contagem do número de indivíduos selecionados e daqueles previamente marcados, feita a marcação dos indivíduos não marcados e todos são devolvidos à população. Em cada época, após devolver os indivíduos a população, deixa-se passar um certo intervalo de tempo até a próxima amostragem, permitindo que os indivíduos recém capturados se misturem à população. No final do processo faz-se a inferência sobre os parâmetros populacionais baseada no número de elementos capturados e recapturados nas k amostragens. Esta metodologia é definida por processo de captura-recaptura em vários estágios de marcação ([Castledine, 1981](#)). Se o método for definido em único estágio de marcação, então somente os indivíduos da primeira ocasião de amostragem são marcados e, assim, o procedimento se resume em observar as frequências de capturas e recapturas dos indivíduos nas demais amostras ([Ananda, 1997](#); [Shimizu, 2002](#)).

A população em estudo pode ser fechada ou aberta. A população se diz fechada se não sofre nenhum efeito que leve a uma modificação do tamanho populacional durante o processo de

amostragem como, por exemplo, efeitos de nascimento, morte ou migração. Em contrapartida, a população aberta é aquela cujo tamanho se altera devido alguns destes fatores (Jolly, 1965).

O uso do método de captura-recaptura para estimação de tamanho populacional é antigo. Os primeiros relatos levam a Laplace (1783), que utilizou o método para estimar o tamanho populacional da França e ao dinamarquês Petersen (1896), que foi o primeiro a utilizar o método em ecologia para estudar o fluxo migratório de peixes no mar Báltico. Quase 40 anos depois, Lincoln (1930) utilizou o método para estimação do tamanho populacional de patos selvagens da América do Norte. A partir dos anos 50, houve uma quantidade substancial de trabalhos relacionados ao método, como Chapman (1954), Darroch (1958, 1959), Jolly (1965, 1982), Burnham and Overton (1978, 1979), Seber (1986, 1982, 1965), Pollock (1980, 1991), Parker (1963), Wang (2002). Trabalhos sob o enfoque Bayesiano também são amplamente encontrados como, por exemplo, Hunter and Griffiths (1978), Castledine (1981), Smith (1988, 1991), George (1992), Ananda (1997), Yoshida et al. (1999), Leite et al. (2000).

Apesar do método de captura-recaptura originalmente tratar da inferência em populações animais, esta metodologia é utilizada atualmente nas mais diversas áreas como, por exemplo, na área da saúde na prevalência de doenças (Lee et al., 2001; Lee, 2002) e na área de controle de qualidade, na estimação do número de erros (falhas) em *software* (Nayak, 1988; Basu and Ebrahimi, 2001).

Neste trabalho aplicamos o método na estimação do número de erros em *software*. Desta forma, o procedimento é direcionado no sentido de que as detecções dos erros sejam realizadas por k ($k \geq 2$) examinadores (revisores) distintos. Assim, associamos cada examinador a uma ocasião de amostragem, isto é, as “recapturas” dos erros são dadas quando o erro for detectado por mais do que um revisor.

Basu and Ebrahimi (2001) propuseram a abordagem bayesiana na estimação do número de erros de *software* em um modelo multinomial incompleto e utilizaram uma mistura de duas distribuições binomiais nas probabilidades de captura com o intuito de modelar a heterogeneidade dos erros. Na tentativa de obter estimativas de máxima verossimilhança para este modelo, Silva (2006) optou por utilizar funções de verossimilhanças condicionais (Sanathanan, 1972). Este método consiste em fatorar a função de verossimilhança global em duas funções e utilizá-las para obter as estimativas dos parâmetros perturbadores e do tamanho populacional separadamente,

facilitando a maximização de cada função. Entretanto, a autora considerou alguns parâmetros perturbadores fixos (conhecidos). Neste trabalho consideramos tanto o tamanho populacional quanto os parâmetros perturbadores que indexam o modelo apresentado por [Basu and Ebrahimi \(2001\)](#) desconhecidos e obtemos inferências sobre os parâmetros populacionais sob os pontos de vista clássico e bayesiano.

No Capítulo 2, damos algumas definições sobre o método de captura-recaptura e apresentamos o modelo probabilístico sob suposição de independência e heterogeneidade entre erros e independência e heterogeneidade entre revisores. Em seguida, supomos que os revisores são homogêneos, isto é, a probabilidade de captura depende apenas do erro e não do revisor, e apresentamos o modelo probabilístico adotado por [Basu and Ebrahimi \(2001\)](#).

No Capítulo 3, propusemos a metodologia clássica para obter as estimativas de máxima verossimilhança dos parâmetros. Na Seção 3.1, as estimativas foram obtidas utilizando a função de verossimilhança global e, na Seção 3.2, fazemos uso de funções de verossimilhanças condicionais ([Sanathanan, 1972](#); [Silva, 2006](#)). Em ambas as seções, optamos por utilizar o algoritmo EM ([Dempster et al., 1977](#)) para obtenção das estimativas dos parâmetros perturbadores. A utilização do algoritmo EM neste capítulo foi motivada pelos trabalhos de [Böhning and Schön \(2005\)](#) e [Barger \(2006\)](#). Em [Böhning and Schön \(2005\)](#), o tamanho populacional é estimado a partir de uma função de verossimilhança composta por misturas de distribuições Poisson. Os autores usam o algoritmo EM para obter estimadores para os parâmetros dos componentes de mistura e um estimador não paramétrico para o tamanho populacional e definem um processo iterativo entre estes estimadores. [Barger \(2006\)](#) deriva um algoritmo semelhante para estimar o número de classes não observadas, onde utiliza misturas de distribuições exponenciais para descrever a média de uma distribuição de Poisson. O desempenho dos métodos propostos neste trabalho foi avaliado a partir de um estudo de simulação utilizando dados artificiais. Em seguida, obtemos estimativas de máxima verossimilhança para os parâmetros populacionais em dois conjuntos de dados reais, o primeiro de falhas em interruptores AT&T 5 ([Eick et al., 1993](#)) e o segundo de um estudo de contagem de coelhos da espécie *cottontail* ([Edwards and Eberhardt, 1967](#)).

No Capítulo 4, abordamos o problema sob o enfoque bayesiano. Métodos iterativos de simulações estocásticas via Monte Carlo Markov Chain (MCMC) foram definidos com o uso do algoritmo *Gibbs sampling* ([Casella and George, 1992](#)) utilizando as distribuições condicionais

a posteriori completas com o uso variáveis latentes (Tanner and Wong, 1987; Diebolt and Robert, 1994). A convergência do algoritmo foi monitorada pelo pacote **CODA - Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output** (Best et al., 1995). O algoritmo foi aplicado para se obter estimativas bayesianas dos parâmetros para os dados reais acima citados.

No Capítulo 5, propomos a generalização do modelo definido por Basu and Ebrahimi (2001) para o caso em há mais do que dois componentes de mistura, isto é, dois ou mais níveis de dificuldade de detecção dos erros. Ilustramos a aplicação deste modelo obtendo estimativas de máxima verossimilhança e bayesianas para dados simulados considerando dois e três níveis de dificuldade de detecção. A comparação entre modelos com diferentes níveis foi feita utilizando critérios AIC e BIC na caso clássico e DIC no caso bayesiano.

Por fim, no Capítulo 6, discutimos os resultados obtidos neste trabalho e propomos uma pesquisa futura.

Capítulo 2

Modelo de captura-recaptura

2.1 Modelo probabilístico com heterogeneidade e independência entre erros e entre revisores

Seja N ($N \geq 1$) o número desconhecido de erros em um *software*. Adotamos o método de captura-recaptura para a estimação do parâmetro N de uma população fechada. Desta forma, aplicamos o método com base em vistorias no *software* a partir de k ($k \geq 2$) revisores distintos. O momento de captura é dito quando o erro é detectado por um revisor e as recapturas são dadas pelas detecções feitas por mais do que um revisor.

No processo de amostragem supomos, primeiramente, que os revisores agem de forma independente, ou seja, as detecções realizadas pelo revisor j não interferem nas detecções do revisor j' , para $j \neq j'$. Supomos também independência entre os processos de detecção de cada erro, isto é, a detecção (ou não) do erro i não interfere na probabilidade de detecção do erro i' , para $i \neq i'$. Para facilidade de expressão, vamos denominar estas suposições como “revisores independentes” e “erros independentes”.

Dado N , sejam

$$X_{i,j} = \begin{cases} 1, & \text{se o erro } i \text{ foi detectado pelo revisor } j, \\ 0, & \text{se o erro } i \text{ não foi detectado pelo revisor } j \end{cases}$$

e $\theta = (\theta_{i,j}; 1 \leq i \leq N, 1 \leq j \leq k)$ o vetor de probabilidades de detecção, onde $\theta_{i,j}$ representa a

probabilidade de detecção do erro i pelo revisor j , $\theta_{i,j} \in (0,1)$, $1 \leq i \leq N$, $1 \leq j \leq k$.

Dados N e $\boldsymbol{\theta}$, assumimos

$$X_{i,j} | \theta_{i,j} \stackrel{ind.}{\sim} \text{Bernoulli}(\theta_{i,j}), \quad 1 \leq i \leq N, \quad 1 \leq j \leq k.$$

Dado N , seja o vetor aleatório k -dimensional $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,k})$ associado ao erro i , onde $\mathbf{X}_i \in \{0,1\}^k$, $1 \leq i \leq N$. Note que temos $l = 2^k$ possíveis resultados para $\mathbf{X}_i \in \{0,1\}^k$, $1 \leq i \leq N$. Considere $\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \dots, \boldsymbol{\chi}_l$ uma enumeração dos elementos de $\{0,1\}^k$. De fato, os $\boldsymbol{\chi}_r$'s representam os históricos de leitura (trajetórias) dos erros, uma vez que $\boldsymbol{\chi}_r$, $1 \leq r \leq l$, é uma k -upla de elementos que assumem os valores um ou zero, conforme a detecção ou não do erro por cada revisor. Vamos referenciar $\boldsymbol{\chi}_l = (0,0, \dots, 0)$ o histórico de leitura não observável e $\boldsymbol{\chi}_{l-1} = (1,1, \dots, 1)$ o histórico de leitura do erro observado por todos os revisores.

Dado N , seja δ_i o indicador correspondente à trajetória assumida pelo erro i , isto é, se o erro i assume a trajetória $\boldsymbol{\chi}_r$, então $\delta_i = r$, $1 \leq i \leq N$, $1 \leq r \leq l$. Assim, a probabilidade de um erro i apresentar o histórico de leitura $\boldsymbol{\chi}_{\delta_i} \in \{0,1\}^k$ é dada por

$$\begin{aligned} P(\mathbf{X}_i = \boldsymbol{\chi}_{\delta_i} | N, \boldsymbol{\theta}) &= P(X_{i,1} = \chi_{\delta_i,1}, \dots, X_{i,k} = \chi_{\delta_i,k} | N, \boldsymbol{\theta}) = \prod_{j=1}^k P(X_{i,j} = \chi_{\delta_i,j} | N, \boldsymbol{\theta}) \\ &= \prod_{j=1}^k \theta_{i,j}^{\chi_{\delta_i,j}} (1 - \theta_{i,j})^{1 - \chi_{\delta_i,j}}, \quad 1 \leq i \leq N, \quad 1 \leq \delta_i \leq l. \end{aligned} \quad (2.1)$$

Sejam

- $n_r = \sum_{i=1}^N \mathcal{I}_{\{r\}}(\delta_i)$ o número de erros com histórico de leitura $\boldsymbol{\chi}_r$, $1 \leq r \leq l-1$, ($\mathcal{I}_{\{r\}}$ é a função indicadora que assume valor 1 se $\delta_i = r$, ou valor 0 se $\delta_i \neq r$);
- $n = \sum_{r=1}^{l-1} n_r$ o número de erros distintos observados;
- $n_l = N - n$ é o número de erros não observados, isto é, o número de erros com histórico de leitura $\boldsymbol{\chi}_l$;
- $\mathbf{n} = (n_1, n_2, \dots, n_l) \in \mathbf{A}_n = \left\{ (n_1, n_2, \dots, n_l) \in \{0,1, \dots, N\}^l : n_1 + \dots + n_l = N \right\}$;
- $\mathbf{C}_n = \left\{ (\delta_1, \dots, \delta_N) \in \{0,1\}^N : n_r = \sum_{i=1}^N \mathcal{I}_{\{r\}}(\delta_i), 1 \leq r \leq l \right\}$ o conjunto de todas as combinações dos δ_i 's que resultam, dentre um número N de erros, em n_1 erros com histórico

de leitura χ_1 , n_2 erros com histórico de leitura χ_2, \dots, n_l erros com histórico de leitura χ_l .

Logo, a função distribuição de probabilidades conjunta de (n_1, n_2, \dots, n_l) , dados N e θ , é dada por

$$\begin{aligned}
 p(n_1, n_2, \dots, n_l | N, \theta) &= P \left(\bigcup_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} (\mathbf{X}_1 = \chi_{\delta_1}, \dots, \mathbf{X}_N = \chi_{\delta_N}) | N, \theta \right) \\
 &= \sum_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} P(\mathbf{X}_1 = \chi_{\delta_1}, \dots, \mathbf{X}_N = \chi_{\delta_N} | N, \theta) \\
 &= \sum_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} \prod_{i=1}^N P(\mathbf{X}_i = \chi_{\delta_i} | N, \theta) \\
 &= \sum_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} \prod_{i=1}^N \prod_{j=1}^k \theta_{i,j}^{\chi_{\delta_i,j}} (1 - \theta_{i,j})^{1 - \chi_{\delta_i,j}}, \tag{2.2}
 \end{aligned}$$

$\mathbf{n} = (n_1, \dots, n_l) \in \mathbf{A}_n$. O número de parcelas desta soma é

$$\binom{N}{n_1} \binom{N - n_1}{n_2} \binom{N - n_1 - n_2}{n_3} \dots \binom{N - n_1 - \dots - n_{l-1}}{n_l} = \frac{N!}{n_1! n_2! \dots n_{l-1}! n_l!}.$$

Exemplo. Suponha que $k = 2$ revisores estão dispostos a examinar um *software*. Assim, cada erro pode assumir um dentre os $l = 2^k = 4$ possíveis históricos de leitura, $\chi_1 = (1,0)$: o erro é detectado somente pelo revisor 1; $\chi_2 = (0,1)$: o erro é detectado somente pelo revisor 2; $\chi_3 = (1,1)$: o erro é detectado por ambos revisores e; $\chi_4 = (0,0)$: o erro não é detectado por revisor algum.

Suponhamos agora que o *software* contém exatamente $N = 2$ erros distintos. De (2.1), para $\theta = (\theta_{1,1}, \theta_{1,2}, \theta_{2,1}, \theta_{2,2})$ e $\mathbf{X}_i = (X_{i,1}, X_{i,2})$, $i = 1, 2$, temos as respectivas probabilidades

$$\begin{aligned}
 P(\mathbf{X}_i = \chi_1 | \theta) &= \theta_{i,1}(1 - \theta_{i,2}); & P(\mathbf{X}_i = \chi_2 | \theta) &= (1 - \theta_{i,1})\theta_{i,2}; \\
 P(\mathbf{X}_i = \chi_3 | \theta) &= \theta_{i,1}\theta_{i,2}; & P(\mathbf{X}_i = \chi_4 | \theta) &= (1 - \theta_{i,1})(1 - \theta_{i,2}).
 \end{aligned}$$

Considere o evento $(n_1, n_2, n_3, n_4) = (1, 0, 0, 1)$. Ou seja, apenas um erro foi observado e com histórico de leitura $\chi_1 = (1, 0)$. Ainda temos que há um erro não detectado neste *software*, pois

$N = 2$. Portanto,

$$\begin{aligned} P(n_1 = 1, n_2 = 0, n_3 = 0, n_4 = 1 | N = 2, \boldsymbol{\theta}) &= \\ &= P(\mathbf{X}_1 = \boldsymbol{\chi}_1 | N, \boldsymbol{\theta})P(\mathbf{X}_2 = \boldsymbol{\chi}_4 | N, \boldsymbol{\theta}) + P(\mathbf{X}_1 = \boldsymbol{\chi}_4 | N, \boldsymbol{\theta})P(\mathbf{X}_2 = \boldsymbol{\chi}_1 | N, \boldsymbol{\theta}) \\ &= \theta_{1,1}(1 - \theta_{1,2})(1 - \theta_{2,1})(1 - \theta_{2,2}) + (1 - \theta_{1,1})(1 - \theta_{1,2})\theta_{2,1}(1 - \theta_{2,2}). \end{aligned}$$

Em palavras, a probabilidade de observar o evento $(n_1, n_2, n_3, n_4) = (1, 0, 0, 1)$ é dada pela probabilidade do “erro 1” assumir o histórico $\boldsymbol{\chi}_1$ e o “erro 2” assumir o histórico $\boldsymbol{\chi}_4$, ou vice versa. Note que esta probabilidade é dada por uma soma de $\frac{2!}{1!0!0!1!} = 2$ termos.

Consideramos três casos particulares de (2.2):

- Caso 1. erros heterogêneos e revisores homogêneos: *consideramos que as probabilidades de detecção dos erros são diferentes a cada erro, mas fixado um erro, a eficiência dos revisores na sua detecção é a mesma; assumimos $\theta_{i,j} = \theta_i$;*
- Caso 2. erros homogêneos e revisores heterogêneos: *consideramos que a probabilidade de detecção de um erro depende exclusivamente do revisor, independentemente do erro. Por exemplo, esta suposição pode ser verdadeira em casos de revisores com diferentes níveis de conhecimento (experiência) em detecções cujos erros tem dificuldade uniforme; neste caso $\theta_{i,j} = \theta_j$;*
- Caso 3. erros e revisores homogêneos: *a probabilidade de detecção, independentemente do revisor, é a mesma para todos os erros. Note que este caso é um caso particular de 1 e 2; deste modo, $\theta_{i,j} = \theta$.*

Neste trabalho, vamos considerar o modelo probabilístico sob a suposição dada no Caso 1, uma vez que o Caso 3 tem pouca aplicação prática. Para o modelo sob a suposição dada no Caso 2 veja, por exemplo, [Castledine \(1981\)](#).

2.1.1 Caso particular: homogeneidade entre revisores

Nesta seção, consideramos o modelo sob suposição de independência e homogeneidade entre os revisores e independência e heterogeneidade entre os erros.

Logo, a probabilidade em (2.2) se expressa como

$$p(n_1, \dots, n_l | N, \boldsymbol{\theta}) = \sum_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} \prod_{i=1}^N \theta_i^{\sum_{j=1}^k \chi_{\delta_i, j}} (1 - \theta_i)^{k - \sum_{j=1}^k \chi_{\delta_i, j}}, \quad (2.3)$$

$\mathbf{n} = (n_1, n_2, \dots, n_l) \in \mathbf{A}_n$.

Como o interesse é estimar N , Burnham and Overton (1978) eliminaram a influência de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ em (2.3) supondo, dado N , que $\theta_1, \dots, \theta_N$ são variáveis aleatórias independentes e identicamente distribuídas por uma função distribuição G no suporte $[0, 1]$.

De (2.3), segue que a função distribuição de probabilidades integrada sobre os θ_i 's é dada por

$$\begin{aligned} p(n_1, \dots, n_l | N, G) &= \int_0^1 \cdots \int_0^1 p(n_1, \dots, n_l | N, \boldsymbol{\theta}) dG(\theta_1) \cdots dG(\theta_N) \\ &= \int_0^1 \cdots \int_0^1 \sum_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} \prod_{i=1}^N \theta_i^{\sum_{j=1}^k \chi_{\delta_i, j}} (1 - \theta_i)^{k - \sum_{j=1}^k \chi_{\delta_i, j}} dG(\theta_1) \cdots dG(\theta_N) \\ &= \sum_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} \prod_{i=1}^N \int_0^1 u^{\sum_{j=1}^k \chi_{\delta_i, j}} (1 - u)^{k - \sum_{j=1}^k \chi_{\delta_i, j}} dG(u) \\ &= \sum_{(\delta_1, \dots, \delta_N) \in \mathcal{C}_n} \prod_{i=1}^N \prod_{r=1}^l \left\{ \int_0^1 u^{\sum_{j=1}^k \chi_{r, j}} (1 - u)^{k - \sum_{j=1}^k \chi_{r, j}} dG(u) \right\}^{\mathcal{I}_{\{r\}}(\delta_i)} \\ &= \frac{N!}{n_1! \cdots n_l!} \prod_{r=1}^l \left\{ \int_0^1 u^{\sum_{j=1}^k \chi_{r, j}} (1 - u)^{k - \sum_{j=1}^k \chi_{r, j}} dG(u) \right\}^{n_r}, \end{aligned} \quad (2.4)$$

$\mathbf{n} = (n_1, n_2, \dots, n_l) \in \mathbf{A}_n$.

Portanto,

$$(n_1, \dots, n_l) | N, G \sim \text{Multinomial}(N, \boldsymbol{\xi}), \quad (2.5)$$

em que o vetor de probabilidades $\boldsymbol{\xi} = (\xi_1, \dots, \xi_l)$ tem como elementos

$$\xi_r = \int_0^1 u^{\sum_{j=1}^k \chi_{r, j}} (1 - u)^{k - \sum_{j=1}^k \chi_{r, j}} dG(u), \quad 1 \leq r \leq l.$$

Definição 2.1. *Seja F_y o número de erros detectados por exatamente y revisores, $0 \leq y \leq k$.*

Então,

$$F_y = \sum_{r=1}^l \mathcal{I}_{\{y\}} \left(\sum_{j=1}^k \chi_{r,j} \right) n_r = \sum_{r \in R_y} n_r, \quad (2.6)$$

com $R_y = \left\{ r : \sum_{j=1}^k \chi_{r,j} = y \right\}$, $0 \leq y \leq k$.

Proposição 2.2. F_1, \dots, F_k são estatísticas suficientes para o modelo em (2.4).

Demonstração. De (2.5), (2.6) e dado N , temos

$$(F_0, F_1, \dots, F_k) | N, G \sim \text{Multinomial}(N, \boldsymbol{\tau}), \quad (2.7)$$

com $\boldsymbol{\tau} = (\tau_0, \dots, \tau_k)$ e

$$\tau_y = \sum_{r \in R_y} \xi_r = \int_0^1 \binom{k}{y} u^y (1-u)^{k-y} dG(u), \quad 0 \leq y \leq k. \quad (2.8)$$

A distribuição de probabilidades condicional de (n_1, \dots, n_l) , dados (F_1, \dots, F_k) , N e G , é dada por

$$\begin{aligned} p(n_1, \dots, n_l | F_1, \dots, F_k, N, G) &= \frac{p(n_1, \dots, n_l | N, G)}{p(F_0, \dots, F_k | N, G)} \\ &= 1 \times \frac{F_1}{\prod_{r \in R_1} n_r!} \prod_{r \in R_1} \left(\frac{\xi_r}{\tau_1} \right)^{n_r} \times \dots \times \frac{F_{k-1}}{\prod_{r \in R_{k-1}} n_r!} \prod_{r \in R_{k-1}} \left(\frac{\xi_r}{\tau_{k-1}} \right)^{n_r} \times 1 \\ &= \prod_{y=1}^{k-1} \frac{F_y}{\prod_{r \in R_y} n_r!} \left(\frac{1}{\binom{k}{y}} \right)^{F_y}. \end{aligned} \quad (2.9)$$

Portanto, temos que (F_1, \dots, F_k) são estatísticas suficientes, pois (2.9) não depende de N . Note também que a probabilidade acima é livre de G e, conseqüentemente, livre de qualquer parâmetro que indexa esta função. \square

Seja o conjunto das frequências $\mathbf{n}_{(y)} = \{n_r : r \in R_y\}$ com $v_y = \binom{k}{y}$ elementos. De (2.9), para $1 \leq y \leq (k-1)$, temos

$$\mathbf{n}_{(y)} | F_y \sim \text{Multinomial}(F_y, \mathbf{v}_y), \quad (2.10)$$

onde $\mathbf{v}_y = (v_y^{-1}, \dots, v_y^{-1})$ é o vetor $(v_y \times 1)$ de probabilidades multinomiais, cujos elementos são equiprováveis. Vale ressaltar que o evento n_{l-1} , dado F_k , ocorre com probabilidade 1, isto é, dentre os F_k erros detectados por exatamente k revisores, teremos exatamente n_{l-1} erros com histórico de leitura $\boldsymbol{\chi}_{l-1} = (1, 1, \dots, 1)$, $n_{l-1} = f_k$. Em outras palavras, se o erro foi detectado k vezes, então o mesmo só assume um único histórico de leitura, $\boldsymbol{\chi}_{l-1} = (1, 1, \dots, 1)$. Este resultado nos mostra que, dado o conjunto de frequências observadas $\mathcal{D} = (f_1, \dots, f_k)$ de (F_1, \dots, F_k) , qualquer pseudo-amostra $(n_1^*, \dots, n_{l-1}^*)$ simulada utilizando (2.10) terá as mesmas propriedades da amostra original (Casella and Berger, 1990).

Segundo Basu and Ebrahimi (2001), a heterogeneidade dos erros se refere à suposição de que os erros são provenientes de duas populações disjuntas quanto a sua dificuldade de detecção como, por exemplo, erros de fácil e de difícil detecção. Disto, supomos G em (2.7) uma função de probabilidades degenerada nos pontos θ_D e θ_F com probabilidades ω e $(1 - \omega)$, respectivamente, isto é,

$$G(u) = \omega \mathcal{I}_{\{\theta_D\}}(u) + (1 - \omega) \mathcal{I}_{\{\theta_F\}}(u), \quad (2.11)$$

para $0 < \omega < 1$, $0 < \theta_D < 1$ e $0 < \theta_F < 1$. Interpretamos ω como sendo a probabilidade de um erro ser de difícil detecção e θ_D e θ_F como sendo as probabilidades de observar um erro considerado de difícil e de fácil detecção, respectivamente. Seja $\boldsymbol{\phi} = (\omega, \theta_D, \theta_F)$ o vetor de parâmetros perturbadores.

Desta forma, utilizamos (2.11) na solução da integral em (2.8) e temos que as probabilidades multinomiais são dadas por

$$h(y; \boldsymbol{\phi}) = \tau_y = \omega \binom{k}{y} \theta_D^y (1 - \theta_D)^{k-y} + (1 - \omega) \binom{k}{y} \theta_F^y (1 - \theta_F)^{k-y}, \quad 0 \leq y \leq k. \quad (2.12)$$

Logo, temos que (2.7) tem função distribuição de probabilidades dada por

$$p(f_0, \dots, f_k | N, \boldsymbol{\phi}) = \frac{N!}{f_0! \dots f_k!} \prod_{y=0}^k \left[\omega \binom{k}{y} \theta_D^y (1 - \theta_D)^{k-y} + (1 - \omega) \binom{k}{y} \theta_F^y (1 - \theta_F)^{k-y} \right]^{f_y}, \quad (2.13)$$

onde $f_0 = N - n$.

Consideramos o processo inferencial do tamanho populacional (N) através do modelo probabilístico apresentado em (2.13). Apresentamos nos dois próximos capítulos as abordagens clássica e bayesiana na estimação dos parâmetros, respectivamente.

Capítulo 3

Estimativas de máxima

verossimilhança: erros de fácil e de díficil detecção

Neste capítulo, vamos descrever os passos para obter estimativas de máxima verossimilhança para os parâmetros especificados no modelo em (2.13).

Primeiramente, na Seção 3.1, obtemos estimativas de máxima verossimilhança para os parâmetros N e ϕ . Em seguida, na Seção 3.2, descrevemos o método para obter as estimativas dos parâmetros utilizando funções de verossimilhança condicionais.

3.1 Estimativas de máxima verossimilhança

Sejam N e ϕ desconhecidos associados à função de verossimilhança

$$L(N, \phi | \mathcal{D}) \propto \frac{N!}{(N-n)!} \prod_{j=0}^k \left[\omega \theta_D^j (1 - \theta_D)^{k-j} + (1 - \omega) \theta_F^j (1 - \theta_F)^{k-j} \right]^{f_j}, \quad (3.1)$$

para $N \geq n$, $(\omega, \theta_D, \theta_F) \in (0,1)^3$, com $f_0 = N - n$ e o conjunto $\mathcal{D} = (f_1, \dots, f_k)$ de frequências observadas. Note que a função de verossimilhança acima é invariante em relação aos componentes θ_D e θ_F da mistura. Isto é, não conseguimos distinguir a componente “díficil” da componente “fácil”, devido serem permutáveis. Para garantir identificabilidade do modelo, vamos considerar

uma ordenação $\theta_D < \theta_F$. Assim, o modelo é identificável se $\phi \in \Phi$, onde

$$\Phi = \{(\omega, \theta_D, \theta_F) \in (0,1)^3 : \theta_D < \theta_F\}.$$

Por outro lado, nota-se que há uma dificuldade imposta pela função de verossimilhança para se determinar estimadores para os parâmetros $(\omega, \theta_D, \theta_F)$, pelo fato da mesma envolver produto de soma. Nestes casos, em que a função de verossimilhança é composta por distribuições de mistura, o algoritmo EM é uma alternativa interessante do ponto de vista teórico e computacional na tentativa de se determinar estimativas de máxima verossimilhança.

3.1.1 O Método EM

Proposto por [Dempster et al. \(1977\)](#), o método EM é utilizado na determinação de estimativas de máxima verossimilhança quando a maximização da função de verossimilhança é analiticamente complexa. Na maioria dos casos, a razão de dessa dificuldade na maximização está relacionada à necessidade de se ajustar um modelo probabilístico adequado aos dados observados que, por sua vez, são “deficientes” nas informações sobre a população, ocasionadas por limitações no seu processo de observação. Isto é caracterizado como *dados incompletos* do experimento. Em poucas palavras, através do método EM é possível adicionar, aos *dados incompletos*, informações adicionais utilizando *variáveis latentes* e, assim, resultando em uma função de probabilidade mais simplificada em termos de maximização ([McLachlan and Krishnan, 2007](#)).

Sejam y o conjunto de dados observados e x o conjunto dos dados latentes (não observados) cujas respectivas distribuições de probabilidades dependem de θ , $\theta \in \Theta \subseteq \mathbb{R}^q$, $q \geq 1$. Suponha, sem perda de generalidade, θ escalar. Assim, temos que $z = (y, x)$ são os dados completos do experimento e associamos uma função de verossimilhança $L(\theta|z)$ para θ . O método EM é definido em dois passos, o passo E para se obter a esperança condicional e o passo M de maximização. Estes dois passos, repetidos iterativamente, denominamos algoritmo EM e é executado da seguinte forma:

Seja $\theta^{(m)} \in \Theta$ o valor de θ na m -ésima iteração;

Passo E: Dado $\theta^{(m)}$, compute a função

$$\begin{aligned} Q(\theta|\theta^{(m)}) &= \sum_{\mathcal{Z}_{(y)}} [\log L(\theta|z)]g(z|y,\theta^{(m)}) \\ &= \mathbb{E}_{Z|y,\theta^{(m)}} [\log L(\theta|z)], \end{aligned} \quad (3.2)$$

em que a soma dá-se no conjunto $\mathcal{Z}_{(y)} = \{z : g(z|y,\theta) > 0\}$ no caso de Z ser uma variável aleatória discreta.¹ Dizemos que $Z|y,\theta^{(m)}$ é a distribuição condicional dos dados completos aos dados observados;

Passo M: obtenha uma nova estimativa para θ , $\theta^{(m+1)}$, tal que

$$\theta^{(m+1)} = \arg \max_{\theta \in \Theta} Q(\theta|\theta^{(m)}). \quad (3.3)$$

Repita os passos E e M substituindo $m := (m + 1)$ a cada iteração.

O algoritmo EM não define um critério de parada próprio. Usualmente, o processo iterativo é interrompido quando $|\theta^{(m+1)} - \theta^{(m)}| < \epsilon$ ou $|Q(\theta^{(m+1)}|\theta^{(m)}) - Q(\theta^{(m)}|\theta^{(m)})| < \epsilon$, para um valor de ϵ ($\epsilon > 0$) fixo.

Para se fazer uso do algoritmo EM precisamos garantir que o suporte das variáveis latentes não depende de parâmetro algum a ser estimado pelo algoritmo (Gupta and Chen, 2011). No nosso caso, como todas as amostras em processos de captura-recaptura dependem do parâmetro N , não podemos utilizar o método EM diretamente na estimação de N . No entanto, dado N fixo, é possível usar o algoritmo para obter estimativas para os parâmetros perturbadores, $\phi = (\omega, \theta_D, \theta_F)$.

Desta forma, propomos neste trabalho um processo iterativo para obtenção das estimativas de máxima verossimilhança para N e ϕ em duas etapas, dadas a seguir.

Etapa (1): dado um valor para ϕ , obtemos uma estimativa “parcial” de N buscando o ponto mais verossímil de $L(N,\phi|\mathcal{D})$, para este valor de ϕ fixo;

Etapa (2): dado N , obtemos uma estimativa “parcial” ϕ_{EM} para ϕ com o auxílio do algoritmo EM. Assim, temos que ϕ_{EM} maximiza $L(N,\phi|\mathcal{D})$ para este valor (fixo) de N .

¹para Z uma variável aleatória absolutamente contínua, temos que $Q(\theta|\theta^{(t)}) = \int_{\mathcal{Z}_{(y)}} \log L(\theta|z)g(z|y,\theta^{(m)})dz$.

Com este novo valor de ϕ_{EM} , obtemos uma nova estimativa N como feito na *Etapa (1)*. Iterativamente, definimos o algoritmo entre as etapas (1) e (2) até que algum grau de convergência seja satisfeito para as estimativas finais de N e ϕ .

Para executarmos a *Etapa (1)*, segue o seguinte teorema.

Teorema 3.1. *Sejam $\phi \in \Phi$ fixo e*

$$a(n, \phi) = \frac{n}{1 - h(0; \phi)}, \quad (3.4)$$

em que $h(0; \phi)$ é dado em (2.12).

(a) *Se $a(n, \phi)$ não for um número inteiro, então a função de verossimilhança em (3.1) tem um único ponto de máximo,*

$$\hat{N} = [a(n, \phi)],$$

onde $[\bullet]$ é a função que associa o maior inteiro menor que o argumento;

(b) *Se $a(n, \phi)$ for um número inteiro, então a função de verossimilhança em (3.1) tem dois pontos de máximo,*

$$\hat{N}_1 = a(n, \phi) \quad e \quad \hat{N}_2 = a(n, \phi) - 1.$$

Demonstração. Considere a razão

$$R(N) = \frac{L(N, \phi | \mathcal{D})}{L(N-1, \phi | \mathcal{D})} = \frac{N}{N-n} h(0; \phi). \quad (3.5)$$

Para $N \geq n+1$, temos

$$R(N) \begin{cases} > 1 & \text{se } N < a(n, \phi), \\ = 1 & \text{se } N = a(n, \phi), \\ < 1 & \text{se } N > a(n, \phi). \end{cases}$$

No caso (a), observamos de $R(N)$ que $L(N, \phi | \mathcal{D})$ é estritamente crescente para $N < a(n, \phi)$ e decrescente para $N > a(n, \phi)$, o que implica que $\hat{N} = [a(n, \phi)]$ é o único inteiro que maximiza (3.1).

No caso (b), temos que $L(N, \phi | \mathcal{D})$ tem valor máximo nos pontos $\hat{N}_1 = a(n, \phi) - 1$ e $\hat{N}_2 = a(n, \phi)$.

Uma prova análoga pode ser vista em [Bolsoni \(2002\)](#). □

3.1.2 Variáveis latentes e função de verossimilhança completa

Considere um valor fixo e conhecido para N ($N \geq n$). Seja a coleção de variáveis latentes $\mathbf{Z} = \{(T_j, U_j); 0 \leq j \leq k\}$, onde T_j representa o número de “erros difíceis” detectados por exatamente j revisores e U_j representa o número de “erros fáceis” detectados por exatamente j revisores, $0 \leq j \leq k$. Denotamos, assim, $\mathbf{z} = \{(t_j, u_j); 0 \leq j \leq k\}$ como os dados completos referentes às observações dos erros no *software*. Note que não observamos \mathbf{z} diretamente, mas observamos a soma de seus componentes $f_j = t_j + u_j$, $0 \leq j \leq k$, dado N .

Considere

$$\mathbf{Z} | (N, \phi) \sim \text{Multinomial}(N, \mathbf{g}(\phi)), \quad (3.6)$$

onde $\mathbf{g}(\phi) = \{g(j; \omega, \theta_D), g(j; 1 - \omega, \theta_F); 0 \leq j \leq k\}$ é o vetor de probabilidades multinomiais com elementos

$$g(j; \omega, \theta_D) = \omega \binom{k}{j} \theta_D^j (1 - \theta_D)^{k-j}$$

e

$$g(j; 1 - \omega, \theta_F) = (1 - \omega) \binom{k}{j} \theta_F^j (1 - \theta_F)^{k-j}$$

associados ao respectivo par (T_j, U_j) , $0 \leq j \leq k$.

Portanto, de (3.6) e dado N , segue que a função de verossimilhança completa para ϕ é dada por

$$\begin{aligned} L(\phi | \mathbf{z}) &= p(t_0, \dots, t_k, u_0, \dots, u_k | N, \phi) \\ &= \frac{N!}{\prod_{j=0}^k t_j! u_j!} \prod_{j=0}^k \left\{ g(j; \omega, \theta_D) \right\}^{t_j} \left\{ g(j; 1 - \omega, \theta_F) \right\}^{u_j} \\ &\propto \prod_{j=0}^k \left\{ \omega \theta_D^j (1 - \theta_D)^{k-j} \right\}^{t_j} \left\{ (1 - \omega) \theta_F^j (1 - \theta_F)^{k-j} \right\}^{u_j}, \end{aligned} \quad (3.7)$$

para $\phi \in \Phi$. Neste caso, a função log-verossimilhança completa é igual à

$$\begin{aligned} \ell(\phi|\mathbf{z}) = \log L(\phi|\mathbf{z}) &= \sum_{j=0}^k t_j \left[\log \omega + j \log \theta_D + (k-j) \log(1-\theta_D) \right] + \\ &+ \sum_{j=0}^k u_j \left[\log(1-\omega) + j \log \theta_F + (k-j) \log(1-\theta_F) \right] + c, \end{aligned} \quad (3.8)$$

para $\phi \in \Phi$ e c não dependendo de ϕ . Para obter a distribuição condicional necessária para obter a função $Q(\cdot)$ dada na relação (3.2), temos os seguintes cálculos de probabilidades,

$$p(\mathbf{z}|f_1, \dots, f_k, N, \phi) = \frac{p(\mathbf{z}|N, \phi)}{p(f_0, \dots, f_k|N, \phi)} = \frac{L(\phi|\mathbf{z})}{L(N, \phi|\mathcal{D})} = \prod_{j=0}^k \frac{f_j!}{t_j! u_j!} p(j; \phi)^{w_j} [1 - p(j; \phi)]^{u_j}, \quad (3.9)$$

onde $p(j; \phi)$ é a probabilidade de um erro ser de difícil detecção dado que foi detectado por exatamente j revisores,

$$p(j; \phi) = \frac{g(j; \omega, \theta_D)}{h(j; \phi)} = \frac{\omega \theta_D^j (1 - \theta_D)^{k-j}}{\omega \theta_D^j (1 - \theta_D)^{k-j} + (1 - \omega) \theta_F^j (1 - \theta_F)^{k-j}}, \quad 0 \leq j \leq k. \quad (3.10)$$

Portanto, de (3.9), concluímos que

$$T_j | f_j, \phi \stackrel{ind.}{\sim} \text{Binomial}(f_j, p(j; \phi)), \quad (3.11)$$

com valores esperados dados por

$$\mathbb{E}[T_j | f_j, \phi] = f_j p(j; \phi)$$

e

$$\begin{aligned} \mathbb{E}[U_j | f_j, \phi] &= \mathbb{E}[f_j - T_j | f_j, \phi] = f_j - f_j p(j; \phi) \\ &= f_j [1 - p(j; \phi)], \quad 0 \leq j \leq k. \end{aligned}$$

Assim, para um $\phi^{(m)} \in \Phi$ fixo e conhecido, substituímos as variáveis latentes em (3.8) por seus

respectivos valores esperados e obtemos a função

$$\begin{aligned}
 Q(\phi|\phi^{(m)}) &= \sum_{j=0}^k f_j p(j; \phi^{(m)}) \left[\log \omega + j \log \theta_D + (k-j) \log(1 - \theta_D) \right] + \\
 &+ \sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})] \left[\log(1 - \omega) + j \log \theta_F + (k-j) \log(1 - \theta_F) \right] + c,
 \end{aligned} \tag{3.12}$$

onde m corresponde ao índice das iterações do algoritmo EM e c constante em relação à ϕ . Portanto, (3.12) finaliza o passo E do método.

Para o passo M, seja

$$\phi^{(m+1)} = \arg \max_{\phi \in \Phi} Q(\phi|\phi^{(m)}) \tag{3.13}$$

obtido da solução do sistema de equações resultantes de

$$\left. \frac{\partial Q(\phi|\phi^{(m)})}{\partial \phi} \right|_{\phi=\phi^{(m+1)}} = \mathbf{0}.$$

Considere os casos:

(a) Equação para ω :

$$\begin{aligned}
 0 &= \frac{\partial Q(\phi|\phi^{(m)})}{\partial \omega} = \sum_{j=0}^k \frac{f_j p(j; \phi^{(m)})}{\omega} - \sum_{j=0}^k \frac{f_j [1 - p(j; \phi^{(m)})]}{1 - \omega} \Leftrightarrow \\
 \frac{1 - \omega}{\omega} &= \frac{\sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})]}{\sum_{j=0}^k f_j p(j; \phi^{(m)})} \Leftrightarrow \\
 \frac{1}{\omega} &= \frac{\sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})]}{\sum_{j=0}^k f_j p(j; \phi^{(m)})} + 1.
 \end{aligned}$$

Como $\sum_{j=0}^k f_j = N$, temos no ponto $\omega = \omega^{(m+1)}$,

$$\omega^{(m+1)} = \frac{1}{N} \sum_{j=0}^k f_j p(j; \phi^{(m)}). \tag{3.14}$$

(b) Equação para θ_D :

$$\begin{aligned} 0 &= \frac{\partial Q(\phi | \phi^{(m)})}{\partial \theta_D} = \sum_{j=0}^k f_j p(j; \phi^{(m)}) \left[\frac{j}{\theta_D} - \frac{(k-j)}{(1-\theta_D)} \right] \Leftrightarrow \\ \frac{1-\theta_D}{\theta_D} &= \frac{\sum_{j=0}^k (k-j) f_j p(j; \phi^{(m)})}{\sum_{j=1}^k j f_j p(j; \phi^{(m)})} \Leftrightarrow \\ \frac{1}{\theta_D} &= \frac{\sum_{j=0}^k (k-j) f_j p(j; \phi^{(m)})}{\sum_{j=1}^k j f_j p(j; \phi^{(m)})} + 1, \end{aligned}$$

e, resolvendo em $\theta_D = \theta_D^{(m+1)}$, chegamos à

$$\theta_D^{(m+1)} = \frac{\sum_{j=1}^k j f_j p(j; \phi^{(m)})}{k \sum_{j=0}^k f_j p(j; \phi^{(m)})}. \quad (3.15)$$

(c) Analogamente ao item (b), temos

$$\theta_F^{(m+1)} = \frac{\sum_{j=1}^k j f_j [1 - p(j; \phi^{(m)})]}{k \sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})]}. \quad (3.16)$$

Proposição 3.2. Para qualquer $\omega^{(m)} \in (0,1)$, temos que $\theta_D^{(m)} < \theta_F^{(m)} \Rightarrow \theta_D^{(m+1)} < \theta_F^{(m+1)}$.

Demonstração. A prova de que $(\theta_D^{(m+1)}, \theta_F^{(m+1)}) \in (0,1)^2$ não é difícil. Para provar que $\theta_D^{(m+1)} < \theta_F^{(m+1)}$, basta mostrar que

$$\theta_D^{(m+1)} = \frac{\sum_{y=1}^k y f_y p(y; \phi^{(m)})}{k \sum_{j=0}^k f_j p(j; \phi^{(m)})} < \frac{\sum_{y=1}^k y f_y [1 - p(y; \phi^{(m)})]}{k \sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})]} = \theta_F^{(m+1)},$$

ou

$$\sum_{y=1}^k y f_y p(y; \phi^{(m)}) \sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})] < \sum_{y=1}^k y f_y [1 - p(y; \phi^{(m)})] \sum_{j=0}^k f_j p(j; \phi^{(m)}).$$

Então, seja

$$c(j; \phi) = \omega \theta_D^j (1 - \theta_D)^{k-j} + (1 - \omega) \theta_F^j (1 - \theta_F)^{k-j}. \quad (3.17)$$

Abaixo, para facilidade de notação, escrevemos ϕ no lugar de $\phi^{(m)}$, tal como θ_D e θ_F ao invés

de $\theta_D^{(m)}$ e $\theta_F^{(m)}$. Então de (3.10) e (3.17), basta provar que

$$\begin{aligned} & \sum_{y=1}^k y f_y \frac{\theta_D^y (1 - \theta_D)^{k-y}}{c(y; \phi)} \sum_{j=0}^k f_j \frac{\theta_F^j (1 - \theta_F)^{k-j}}{c(j; \phi)} - \sum_{y=1}^k y f_y \frac{\theta_F^y (1 - \theta_F)^{k-y}}{c(y; \phi)} \sum_{j=0}^k f_j \frac{\theta_D^j (1 - \theta_D)^{k-j}}{c(j; \phi)} < 0 \Leftrightarrow \\ & \sum_{y=1}^k y f_y \left\{ \sum_{j=0}^k f_j \left[\frac{\theta_D^y (1 - \theta_D)^{k-y}}{c(y; \phi)} \frac{\theta_F^j (1 - \theta_F)^{k-j}}{c(j; \phi)} - \frac{\theta_F^y (1 - \theta_F)^{k-y}}{c(y; \phi)} \frac{\theta_D^j (1 - \theta_D)^{k-j}}{c(j; \phi)} \right] \right\} < 0 \Leftrightarrow \\ & \sum_{y=1}^k y f_y \left\{ \sum_{j=0}^k f_j d_{y,j}(\phi) \right\} < 0, \end{aligned}$$

onde

$$d_{y,j}(\phi) = \frac{\theta_D^y (1 - \theta_D)^{k-y}}{c(y; \phi)} \frac{\theta_F^j (1 - \theta_F)^{k-j}}{c(j; \phi)} - \frac{\theta_F^y (1 - \theta_F)^{k-y}}{c(y; \phi)} \frac{\theta_D^j (1 - \theta_D)^{k-j}}{c(j; \phi)}.$$

Como $d_{y,j}(\phi) = -d_{j,y}(\phi)$, substituímos $d_{y,j}(\phi)$ por $-d_{j,y}(\phi)$ quando $y > j$. Portanto, temos

$$\begin{aligned} & \sum_{y=1}^k y f_y \left\{ \sum_{j=0}^k f_j d_{y,j}(\phi) \right\} = \sum_{0 \leq j \leq y \leq k} (y - j) f_y f_j d_{y,j}(\phi) < 0 \Leftrightarrow \\ & \sum_{0 \leq j \leq y \leq k} \underbrace{\frac{(y - j) f_y f_j}{c(y; \phi) c(j; \phi)} \left(\frac{\theta_F}{1 - \theta_F} \right)^y \left(\frac{\theta_D}{1 - \theta_D} \right)^j}_{(*)} \left[\underbrace{\left(\frac{\theta_D}{\theta_F} \right)^{y-j} \left(\frac{1 - \theta_F}{1 - \theta_D} \right)^{y-j} - 1}_{(**)} \right] < 0. \quad (3.18) \end{aligned}$$

Como $(*)$ é positivo, para mostrar (3.18) é suficiente que $\left(\frac{\theta_D}{\theta_F} \right)^{y-j} \left(\frac{1 - \theta_F}{1 - \theta_D} \right)^{y-j} < 1$ para todo $0 \leq j < y \leq k$, o que é equivalente a $\theta_D < \theta_F$. \square

Proposição 3.3. *Temos que $\phi^{(m+1)}$ é ponto de máximo de $Q(\phi|\phi^{(m)})$.*

Demonstração. A matriz de derivadas de segunda ordem de $Q(\phi|\phi^{(m)})$ é dada por

$$B = \begin{pmatrix} \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \omega^2} & \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \omega \partial \theta_D} & \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \omega \partial \theta_F} \\ \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_D \partial \omega} & \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_D^2} & \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_D \partial \theta_F} \\ \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_F \partial \omega} & \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_F \partial \theta_D} & \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_F^2} \end{pmatrix} \Big|_{\phi = \phi^{(m+1)}},$$

onde

$$\begin{aligned} \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \omega^2} \Big|_{\phi=\phi^{(m+1)}} &= -\frac{\sum_{j=0}^k f_j p(j; \phi^{(m)})}{\omega^{(m+1)2}} - \frac{\sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})]}{(1 - \omega^{(m+1)})^2} < 0, \\ \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_D^2} \Big|_{\phi=\phi^{(m+1)}} &= \sum_{y=0}^k f_y p(j; \phi^{(m)}) \left[-\frac{y}{\theta_D^{(m+1)2}} - \frac{k-y}{(1 - \theta_D^{(m+1)})^2} \right] < 0, \\ \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_F^2} \Big|_{\phi=\phi^{(m+1)}} &= \sum_{y=0}^k f_y [1 - p(j; \phi^{(m)})] \left[-\frac{y}{\theta_F^{(m+1)2}} - \frac{k-y}{(1 - \theta_F^{(m+1)})^2} \right] < 0, \\ \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_D \partial \theta_F} \Big|_{\phi=\phi^{(m+1)}} &= 0 = \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_D \partial \omega} \Big|_{\phi=\phi^{(m+1)}} = \frac{\partial^2 Q(\phi|\phi^{(m)})}{\partial \theta_F \partial \omega} \Big|_{\phi=\phi^{(m+1)}}. \end{aligned}$$

Seja $\{b_{i,j}\}$ o elemento correspondente à linha i e coluna j de B e $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$, para $\mathbf{x} \neq \mathbf{0}$. Como B é uma matriz diagonal, temos que

$$\mathbf{x}^T B \mathbf{x} = \sum_{i=1}^3 \sum_{j=1}^3 b_{i,j} x_i x_j = \sum_{i=1}^3 b_{i,i} x_i^2 < 0,$$

pois $b_{i,i} < 0$ e $x_i^2 > 0$, para todo $i = 1, 2, 3$. Portanto, como B é uma matriz negativa definida, concluímos que (3.14), (3.15) e (3.16) são conjuntamente pontos de máximo de $Q(\phi|\phi^{(m)})$. \square

Assim, concluímos que o algoritmo EM é determinado por um simples processo iterativo baseado nos resultados obtidos em (3.14), (3.15) e (3.16), dado um valor de N , $N \geq n$, fixo e conhecido. Dizemos que alcançamos a convergência para um ponto estacionário de ϕ a partir do momento em que as estimativas pararem de sofrer (grandes) alterações, isto é, para uma precisão $\varepsilon = (\varepsilon, \varepsilon, \varepsilon)$, $\varepsilon > 0$, interrompemos o processo iterativo quando o critério de parada $|\phi^{(m+1)} - \phi^{(m)}| < \varepsilon$ for satisfeito, para qualquer $m \geq 0$. A implementação do algoritmo EM é apresentada na Seção 3.1.3 juntamente com a estimação de N utilizando o Teorema 3.1.

3.1.3 Implementação do algoritmo

Inicialize com $N^{(0)} \geq n$, $\phi^{(0)} \in \Phi$ e $(l = 0)$;

(1) Algoritmo EM:

(i) Fixe $N = N^{(l)}$. Faça $(m = 0)$ e $\phi^{(m)} = \phi^{(l)}$.

(ii) Repita iterativamente

$$\theta_D^{(m+1)} = \frac{\sum_{j=1}^k j f_j p(j; \phi^{(m)})}{k \sum_{j=0}^k f_j p(j; \phi^{(m)})}; \quad \theta_F^{(m+1)} = \frac{\sum_{j=1}^k j f_j [1 - p(j; \phi^{(m)})]}{k \sum_{j=0}^k f_j [1 - p(j; \phi^{(m)})]};$$

$$\omega^{(m+1)} = \frac{1}{N} \sum_{j=0}^k f_j p(j; \phi^{(m)});$$

substituindo $m := (m + 1)$ a cada iteração.

(iii) Interrompa (ii) quando $|\phi^{(m+1)} - \phi^{(m)}| < \epsilon$ e vá para o passo (2).

(2) Fazendo $\phi^{(l+1)} = \phi^{(m+1)}$, atualize N :

$$N^{(l+1)} = \text{maior inteiro menor que } \frac{n}{1 - h(0; \phi^{(l+1)})}.$$

Se $N^{(l+1)} = N^{(l)}$, finalize o algoritmo. Caso contrário, faça $l := (l + 1)$ e volte ao passo (1).

Note que o algoritmo precisa de valores iniciais para ambos N e ϕ , embora seja possível inicializá-lo sem necessidade de um valor inicial para N , bastando inverter a ordem dos passos (1) e (2) do algoritmo.

Os resultados numéricos das estimativas dos parâmetros serão apresentados na Seção 3.3, onde comparamos ao método de verossimilhanças condicionais, discutido a seguir.

3.2 Estimativas de máxima verossimilhanças condicionais

Nesta seção, abordamos um método alternativo de obter estimativas de máxima verossimilhança para os parâmetros N e $\phi = (\omega, \theta_D, \theta_F)$ do modelo em (3.1) utilizando funções de verossimilhança condicionais (Sanathanan, 1972).

Note que a função de verossimilhança em (3.1) pode ser reescrita sob a forma

$$\begin{aligned}
 L(N, \phi | \mathcal{D}) &= \binom{N}{n} \frac{n!}{f_1! \cdots f_k!} [1 - h(0; \phi)]^n h(0; \phi)^{N-n} \frac{\prod_{j=1}^k h(j; \phi)^{f_j}}{[1 - h(0; \phi)]^n} \\
 &= \binom{N}{n} [1 - h(0; \phi)]^n h(0; \phi)^{N-n} \frac{n!}{f_1! \cdots f_k!} \frac{\prod_{j=1}^k h(j; \phi)^{f_j}}{[1 - h(0; \phi)]^{\sum_{j=1}^k f_j}} \\
 &= \binom{N}{n} [1 - h(0; \phi)]^n h(0; \phi)^{N-n} \frac{n!}{f_1! \cdots f_k!} \prod_{j=1}^k \left[\frac{h(j; \phi)}{1 - h(0; \phi)} \right]^{f_j} \\
 &= L_1(N, \phi | n) \cdot L_2(\phi | f_1, \dots, f_k, n),
 \end{aligned} \tag{3.19}$$

onde

$$L_1(N, \phi | n) = \binom{N}{n} [1 - h(0; \phi)]^n h(0; \phi)^{N-n} \tag{3.20}$$

e

$$L_2(\phi | f_1, \dots, f_k, n) = \frac{n!}{f_1! \cdots f_k!} \prod_{j=1}^k \left[\frac{h(j; \phi)}{1 - h(0; \phi)} \right]^{f_j} \tag{3.21}$$

para $N \geq n$ e $\phi \in \Phi$.

Teorema 3.4.

- (i) $L_1(N, \phi | n)$ é a função de verossimilhança correspondente à distribuição de probabilidades de n , dado (N, ϕ) ;
- (ii) $L_2(\phi | f_1, \dots, f_k, n)$ é a função de verossimilhança correspondente à distribuição de probabilidades de (f_1, \dots, f_k) , dado (n, ϕ) .

Demonstração. Dado n , considere o conjunto

$$S = \left\{ (f_1, \dots, f_k) \in \{0, 1, \dots, n\}^k : \sum_{j=1}^k f_j = n \right\}.$$

Portanto, de (3.19), segue

$$\begin{aligned}
 p(n|N, \phi) &= \sum_{(f_1, \dots, f_k) \in \mathcal{S}} p(f_0, f_1, \dots, f_k | N, \phi) \\
 &= \sum_{(f_1, \dots, f_k) \in \mathcal{S}} L(N, \phi | \mathcal{D}) \\
 &= L_1(N, \phi | n) \sum_{(f_1, \dots, f_k) \in \mathcal{S}} L_2(\phi | f_1, \dots, f_k, n) \\
 &= L_1(N, \phi | n) \underbrace{\left[\sum_{j=1}^k \frac{h(j; \phi)}{1 - h(0; \phi)} \right]}_{=1}^n = L_1(N, \phi | n),
 \end{aligned}$$

o que prova (i).

De (3.19) e o item (i), temos

$$\begin{aligned}
 p(f_1, \dots, f_k | n, \phi) &= p(f_1, \dots, f_k | n, N, \phi) \\
 &= \frac{p(f_0, f_1, \dots, f_k | N, \phi)}{p(n | N, \phi)} \\
 &= \frac{L(N, \phi | f_1, \dots, f_k)}{L_1(N, \phi | n)} = L_2(\phi | f_1, \dots, f_k, n),
 \end{aligned}$$

o que prova o item (ii). □

Sanathanan (1972) propôs o método de estimação do parâmetro N nos casos em que a função de verossimilhança pode ser fatorada em duas funções, em que uma função depende apenas dos parâmetros perturbadores, como mostra a relação (3.19). Assim, o método de estimação consiste em duas etapas:

Etapa (1): Primeiramente, utilizamos a função de verossimilhança $L_2(\phi | \cdot)$ para obter uma estimativa de máxima verossimilhança condicional (EMVC) para ϕ ;

Etapa (2): Em seguida, determinamos uma EMVC para N com o uso da função de verossimilhança $L_1(N, \hat{\phi} | \cdot)$, em que $\hat{\phi}$ é a EMVC de ϕ obtida na *Etapa (1)*.

Silva (2006) empregou este método para estimar os parâmetros nas abordagens clássica e bayesiana, porém considerou os valores de θ_D e θ_F conhecidos. Em nosso trabalho, optamos pelo auxílio do algoritmo EM para executarmos a *Etapa (1)*.

3.2.1 Variáveis latentes e função de verossimilhança condicional completa

Nesta seção vamos obter a estimativa de máxima verossimilhança condicional para $\phi = (\omega, \theta_D, \theta_F)$ utilizando a função de verossimilhança

$$L_2(\phi|f_1, \dots, f_k, n) = \frac{n!}{f_1! \cdots f_k!} \prod_{j=1}^k \left[\frac{h(j; \phi)}{1 - h(0; \phi)} \right]^{f_j},$$

onde $h(j; \phi) = \omega \binom{k}{j} \theta_D^j (1 - \theta_D)^{k-j} + (1 - \omega) \binom{k}{j} \theta_F^j (1 - \theta_F)^{k-j}$, $0 \leq j \leq k$. Note que, além das probabilidades multinomiais acima serem dadas por uma mistura de duas distribuições binomiais, as mesmas são truncadas no ponto 0.

Assim, para corrigir o problema do “truncamento” utilizando o método EM, adicionamos na função $L_2(\phi|\cdot)$ a variável latente F_0 , que corresponderá ao número de erros não detectados pelos revisores. Logo, como não temos o conhecimento do número total de erros no software (N) em $L_2(\phi|\cdot)$, segue que a variável F_0 é ilimitada superiormente, $F_0 \in \{0, 1, \dots\}$.

Consideramos

$$F_0|n, \phi \sim \text{BinNeg}(n, 1 - h(0; \phi)), \quad (3.22)$$

para $\phi \in \Phi$ e $n \geq 1$ com função de probabilidades dada por

$$P(F_0 = f_0|n, \phi) = \binom{f_0 + n - 1}{f_0} h(0; \phi)^{f_0} [1 - h(0; \phi)]^n, \quad f_0 = 0, 1, \dots$$

O valor esperado é dado por

$$\mathbb{E}[F_0|n, \phi] = n \frac{h(0; \phi)}{1 - h(0; \phi)}. \quad (3.23)$$

Dado n e ϕ , as variáveis (F_1, \dots, F_k) são independentes F_0 . Desta forma, a função de verossimilhança condicional (parcialmente) completa é dada por

$$\begin{aligned} L_2(\phi|F_0, f_1, \dots, f_k, n) &= P(F_0 = f_0, F_1 = f_1, \dots, F_k = f_k|n, \phi) \\ &= P(F_1 = f_1, \dots, F_k = f_k|n, \phi) \cdot P(F_0 = f_0|n, \phi) \\ &\propto \prod_{j=0}^k [h(j; \phi)]^{f_j}, \end{aligned} \quad (3.24)$$

para $\phi \in \Phi$. Note que o “kernel” de (3.24) em ϕ é o mesmo apresentado na relação (3.1). Assim, dados (F_0, f_1, \dots, f_k) , utilizamos novamente as frequências t_j 's definidas na seção 3.1 para sanar o problema da mistura. Neste caso, os dados completos são dados pelo vetor $\mathbf{x} = (f_j, t_j; 0 \leq j \leq k)$. Note que é equivalente escrever os dados completos por \mathbf{x} ou por $\mathbf{z} = (t_j, u_j; 0 \leq j \leq k)$, como anteriormente.

Pelas relações (3.11) e (3.24), segue a função de verossimilhança condicional completa para ϕ ,

$$\begin{aligned}
 L_2(\phi|\mathbf{x}) &= P(T_0 = t_0, \dots, T_k = t_k, F_0 = f_0, \dots, F_k = f_k | n, \phi) \\
 &= P(F_0 = f_0, F_1 = f_1, \dots, F_k = f_k | n, \phi) \cdot P(T_0 = t_0, \dots, T_k = t_k | f_0, \dots, f_k, \phi) \\
 &= L_2(\phi | F_0, f_1, \dots, f_k, n) \cdot \prod_{j=0}^k P(T_j = t_j | f_j, \phi) \\
 &= \frac{n(f_0 + n - 1)!}{\prod_{j=0}^k t_j! (f_j - t_j)!} \prod_{j=0}^k [g(j; \omega, \theta_D)]^{t_j} [g(j; 1 - \omega, \theta_F)]^{f_j - t_j} \\
 &\propto \prod_{j=0}^k \left[\omega \theta_D^j (1 - \theta_D)^{k-j} \right]^{t_j} \left[(1 - \omega) \theta_F^j (1 - \theta_F)^{k-j} \right]^{f_j - t_j}, \tag{3.25}
 \end{aligned}$$

para $\phi \in \Phi$.

Para um valor conhecido $\phi^{(m)} \in \Phi$, temos a função log-verossimilhança esperada,

$$\begin{aligned}
 Q_c(\phi | \phi^{(m)}) &= \sum_{j=0}^k \mathbb{E} \left[T_j | f_1, \dots, f_k, n, \phi^{(m)} \right] \left[\log \omega + j \log \theta_D + (k - j) \log(1 - \theta_D) \right] + \\
 &\quad + \sum_{j=0}^k \mathbb{E} \left[F_j - T_j | f_1, \dots, f_k, n, \phi^{(m)} \right] \left[\log(1 - \omega) + j \log \theta_F + (k - j) \log(1 - \theta_F) \right] \\
 &\quad + \text{constante}, \tag{3.26}
 \end{aligned}$$

para $\phi \in \Phi$.

Por (3.11), para $1 \leq j \leq k$, temos as seguintes esperanças condicionais já definidas,

$$\mathbb{E}[T_j | f_1, \dots, f_k, n, \phi] = \mathbb{E}[T_j | f_j, \phi] = f_j p(j; \phi),$$

e, para $j = 0$, utilizando as relações (3.11) e (3.23), obtemos

$$\begin{aligned}
 \mathbb{E}[T_0|f_1, \dots, f_k, n, \phi] &= \sum_{t_0=0}^{\infty} t_0 P(T_0 = t_0 | f_1, \dots, f_k, n, \phi^{(m)}) \\
 &= \sum_{f_0=0}^{\infty} \sum_{t_0=0}^{f_0} t_0 P(T_0 = t_0, F_0 = f_0 | f_1, \dots, f_k, n, \phi^{(m)}) \\
 &= \sum_{f_0=0}^{\infty} \sum_{t_0=0}^{f_0} t_0 P(T_0 = t_0 | F_0 = f_0, \phi^{(m)}) P(F_0 = f_0 | n, \phi^{(m)}) \\
 &= \sum_{f_0=0}^{\infty} \mathbb{E}[T_0 | f_0, \phi^{(m)}] P(F_0 = f_0 | n, \phi^{(m)}) \\
 &= p(0; \phi^{(m)}) \sum_{f_0=0}^{\infty} f_0 P(F_0 = f_0 | n, \phi^{(m)}) = \mathbb{E}[F_0 | n, \phi^{(m)}] p(0; \phi^{(m)}) \\
 &= \frac{nh(0; \phi^{(m)})}{1 - h(0; \phi^{(m)})} p(0; \phi^{(m)}) = \mathbb{E}[T_0 | n, \phi^{(m)}].
 \end{aligned} \tag{3.27}$$

Seja $\phi_c^{(m+1)} = (\omega_c^{(m+1)}, \theta_{Dc}^{(m+1)}, \theta_{Fc}^{(m+1)})$, em que

$$\phi_c^{(m+1)} = \arg \max_{\phi \in \Phi} Q_c(\phi | \phi^{(m)}).$$

Portanto, temos que $\omega_c^{(m+1)}$, $\theta_{Dc}^{(m+1)}$ e $\theta_{Fc}^{(m+1)}$ serão aproximadamente iguais às expressões dadas em (3.14), (3.15) e (3.16), respectivamente, bastando substituir f_0 por $\mathbb{E}[F_0 | n, \phi^{(m)}]$. Esta maneira direta em obter a atualização $\phi_c^{(m+1)}$ é devido ao fato que as funções de verossimilhança dadas em (3.7) e em (3.25) contém o mesmo “kernel” com relação à ϕ e assim, a maximização das suas respectivas funções log-verossimilhança esperadas são dadas de maneira análoga aos dois casos, para Q em (3.12) e Q_c em (3.26).

Pela Proposição 3.2, temos que se $\theta_{Dc}^{(m)} < \theta_{Fc}^{(m)} \Rightarrow \theta_{Dc}^{(m+1)} < \theta_{Fc}^{(m+1)}$, para todo $m \geq 0$. Pela Proposição 3.3, temos que $\phi_c^{(m+1)}$ é ponto de máximo de $Q_c(\phi | \phi^{(m)})$.

Implementamos a seguir o método iterativo para obtenção da EMVC de ϕ .

3.2.2 Implementação do algoritmo

Inicialize o algoritmo EM com $\phi_c^{(0)} \in \Phi$ e $(m = 0)$;

(i) repita iterativamente

$$\theta_{Dc}^{(m+1)} = \frac{\sum_{j=1}^k j f_j p(j; \phi_c^{(m)})}{k \sum_{j=1}^k f_j p(j; \phi_c^{(m)}) + kn \frac{g(0; \omega_c^{(m)}, \theta_{Dc}^{(m)})}{1 - h(0; \phi_c^{(m)})}},$$

$$\theta_{Fc}^{(m+1)} = \frac{\sum_{j=1}^k j f_j [1 - p(j; \phi_c^{(m)})]}{k \sum_{j=1}^k f_j [1 - p(j; \phi_c^{(m)})] + kn \frac{g(0; 1 - \omega_c^{(m)}, \theta_{Fc}^{(m)})}{1 - h(0; \phi_c^{(m)})}},$$

$$\omega_c^{(m+1)} = g(0; \omega_c^{(m)}, \theta_{Dc}^{(m)}) + \frac{1 - h(0; \phi_c^{(m)})}{n} \sum_{j=1}^k f_j p(j; \phi_c^{(m)}),$$

substituindo $m := (m + 1)$ a cada iteração.

(ii) Interrompa (i) quando $|\phi_c^{(m+1)} - \phi_c^{(m)}| < \epsilon$.

Seja $\hat{\phi}_c$ estimativa de máxima verossimilhança condicional de ϕ obtida pelo algoritmo descrito acima. Note que a razão $R(N)$ em (3.5) será a mesma se utilizarmos a função $L_1(N, \hat{\phi}|n)$ ao invés de $L(N, \hat{\phi}|\mathcal{D})$.

Por este fato, utilizamos o Teorema 3.1 para obter uma EMVC para N, \hat{N}_c , isto é,

(a) \hat{N}_c é o maior inteiro menor que $a(n, \hat{\phi}_c)$, se $a(n, \hat{\phi}_c)$ (veja (3.4)) for não inteiro;

(b) se $a(n, \hat{\phi}_c)$ for um número inteiro, então $\hat{N}_{c1} = a(n, \hat{\phi}_c)$ e $\hat{N}_{c2} = a(n, \hat{\phi}_c) - 1$ são ambos pontos de máximo de $L_1(N, \hat{\phi}_c|n)$.

3.3 Resultados numéricos

Nesta seção, primeiramente, avaliamos a performance dos estimadores através de dados artificiais obtidos via simulação. Em seguida, aplicamos o processo de estimação em dois conjunto de dados reais da literatura.

3.3.1 Dados artificiais

Apresentamos a seguir o método de simulação de um conjunto de dados para o modelo proposto.

Determinamos, de início, os valores “reais” dos parâmetros da população N, ω, θ_D e θ_F , e o número k de revisores a examinar o *software*.

Para cada erro i , o identificamos se é de fácil ou difícil detecção, isto é, simulamos o vetor $\mathbf{x} = (x_1, \dots, x_N)$, onde cada x_i é gerado de uma variável aleatória $X \sim \text{Bernoulli}(\omega)$. Deste modo, temos que um erro i é de difícil detecção ($x_i = 1$) com probabilidade ω e é de fácil detecção ($x_i = 0$), com probabilidade $1 - \omega$.

Dado o vetor \mathbf{x} , simulamos por quantos revisores cada erro i foi detectado, dado sua probabilidade de captura, θ_D ou θ_F . Assim, obtemos o vetor $\mathbf{y} = (y_1, \dots, y_N)$ onde cada y_i é simulado de uma variável aleatória $Y_i \sim \text{Binomial}(k, \theta_D^{x_i} \theta_F^{1-x_i})$, $1 \leq i \leq N$.

Finalmente, obtemos as frequências $f_j = \sum_{i=1}^N \mathcal{I}_{\{j\}}(y_i)$, $1 \leq j \leq k$ e $n = \sum_{j=1}^k f_j$. Note que, se for de interesse adicionar a informação de um $(k+1)$ -ésimo revisor, basta apenas simular as capturas de erros por este novo revisor, i.e., obter via simulação o vetor $\mathbf{t} = (t_1, \dots, t_N)$, onde cada t_i é observado de uma distribuição $\text{Bernoulli}(\theta_D^{x_i} \theta_F^{1-x_i})$. Desta forma, temos que $f_j = \sum_{i=1}^N \mathcal{I}_{\{j\}}(y_i^*)$, onde $y_i^* = y_i + t_i$, para $1 \leq j \leq k+1$.

Avaliamos a desempenho dos algoritmo descritos nas seções 3.1.3 e 3.2.2 utilizando 9 cenários diferentes em um processo de captura-recaptura. Em cada cenário, geramos 1000 amostras para cada valor de k , $k = 5, 6, \dots, 15$. Em todos os casos, os valores reais das probabilidades de captura foram $\theta_D = 0,2$ e $\theta_F = 0,6$. A precisão determinada para o critério de parada dos algoritmos foi $\epsilon = 0,00001$.

- Cenário 1:* Valor de $N = 100$ e $\omega = 0,2$,
- Cenário 2:* Valor de $N = 100$ e $\omega = 0,5$,
- Cenário 3:* Valor de $N = 100$ e $\omega = 0,8$,
- Cenário 4:* Valor de $N = 500$ e $\omega = 0,2$,
- Cenário 5:* Valor de $N = 500$ e $\omega = 0,5$,
- Cenário 6:* Valor de $N = 500$ e $\omega = 0,8$,
- Cenário 7:* Valor de $N = 1000$ e $\omega = 0,2$,
- Cenário 8:* Valor de $N = 1000$ e $\omega = 0,5$,
- Cenário 9:* Valor de $N = 1000$ e $\omega = 0,8$.

Nas Figuras 3.1 à 3.4, ilustramos o comportamento dos estimadores de máxima verossimilhança e máxima verossimilhanças condicionais através do viés das estimativas em cada situação amostral.

Por exemplo, para \hat{N} , calculamos seu viés como

$$\text{Viés}(\hat{N}) = \frac{1}{m} \sum_{i=1}^m \hat{N}_i - N_V,$$

onde m é o número de estimativas, \hat{N}_i é a EMV (ou EMVC) para N na amostra i e N_V é o verdadeiro valor do parâmetro utilizado na geração das amostras.

Nas Figuras 3.5 à 3.8 são exibidos os valores dos erros quadráticos médios dos estimadores em cada um dos cenários descritos acima. O erro quadrático médio para \hat{N} , por exemplo, é calculado por

$$\text{EQM}(\hat{N}) = \frac{1}{m} \sum_{i=1}^m (\hat{N}_i - N_V)^2,$$

onde m é o número de estimativas, \hat{N}_i é a EMV (ou EMVC) para N na amostra i e N_V é o verdadeiro valor do parâmetro utilizado na geração da amostra.

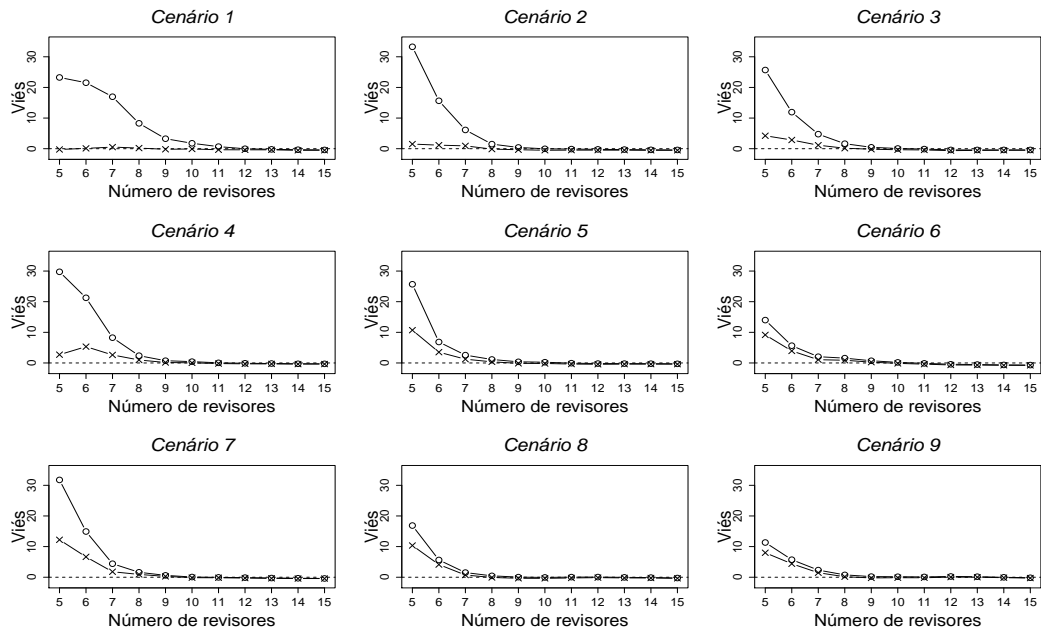


Figura 3.1: Comportamento do viés de \hat{N} e \hat{N}_c para os 9 cenários. *Legenda:* \times EMV; \circ EMVC.

Na Figura 3.1 observamos, em todos os cenários, que o método de verossimilhanças

condicionais apresentou viés maior para o tamanho populacional em comparação ao método utilizando a função de verossimilhança global. Nos *Cenários 1, 2 e 3*, o viés das estimativas de máxima verossimilhança foi quase nulo para todos os valores de k , $5 \leq k \leq 15$. Apesar da grandeza do valor verdadeiro de N nos *Cenários 4 à 9* ($N = 500$ e 1000), as estimativas de máxima verossimilhança para este parâmetro foram razoavelmente boas, em média. Visualmente, notamos que as EMV's e EMVC's para N utilizando mais do que 8 revisores são, em média, não viesadas para as 9 situações amostrais estudadas.

Pela Figura 3.2 notamos que para $\omega = 0.8$ o viés das estimativas foi baixo para qualquer valor de N (*Cenários 3, 6 e 9*). Os piores cenários foram para $N = 100$, $\omega = 0.2$ e 0.5 e para $N = 500$ e $\omega = 0.2$ (*Cenários 1, 2 e 4*).

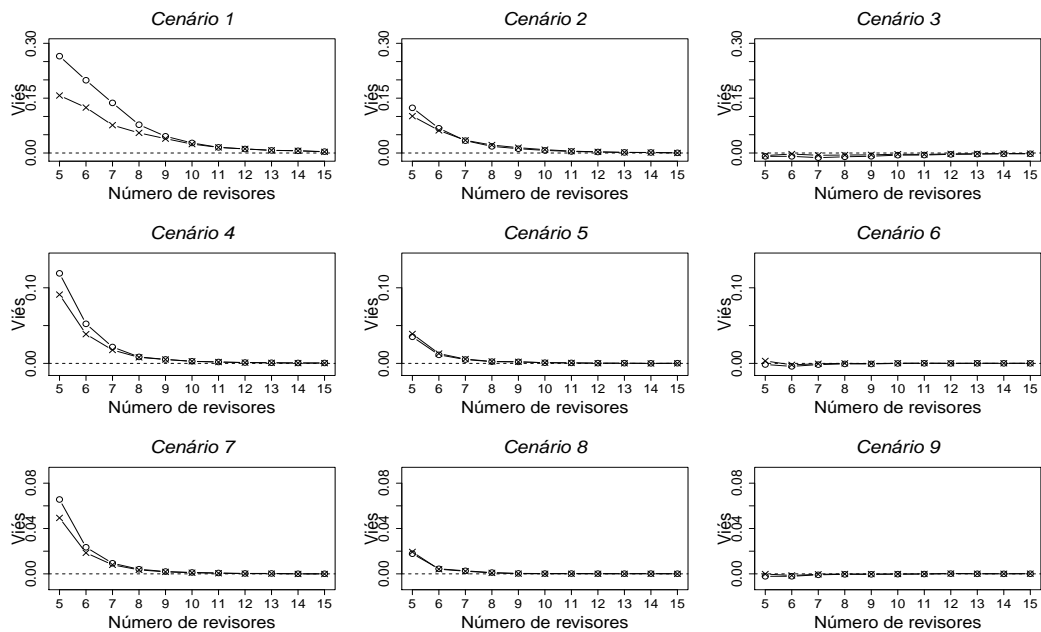


Figura 3.2: Comportamento do viés de $\hat{\omega}$ e $\hat{\omega}_c$ para os 9 cenários. *Legenda:* \times EMV; \circ EMVC.

Nas Figuras 3.3 e 3.4, nota-se que os viés das estimativas de máxima verossimilhança e máxima verossimilhanças condicionais de θ_D e θ_F ficam cada vez mais próximos de 0 de acordo com o aumento do número de revisores, como esperado. Além disso, é possível observar, em todos os casos, que os viés variam no intervalo -0.03 a 0.07 . Podemos dizer que, em média, ambos os métodos apresentam boas estimativas para as probabilidades de capturas θ_D e θ_F .

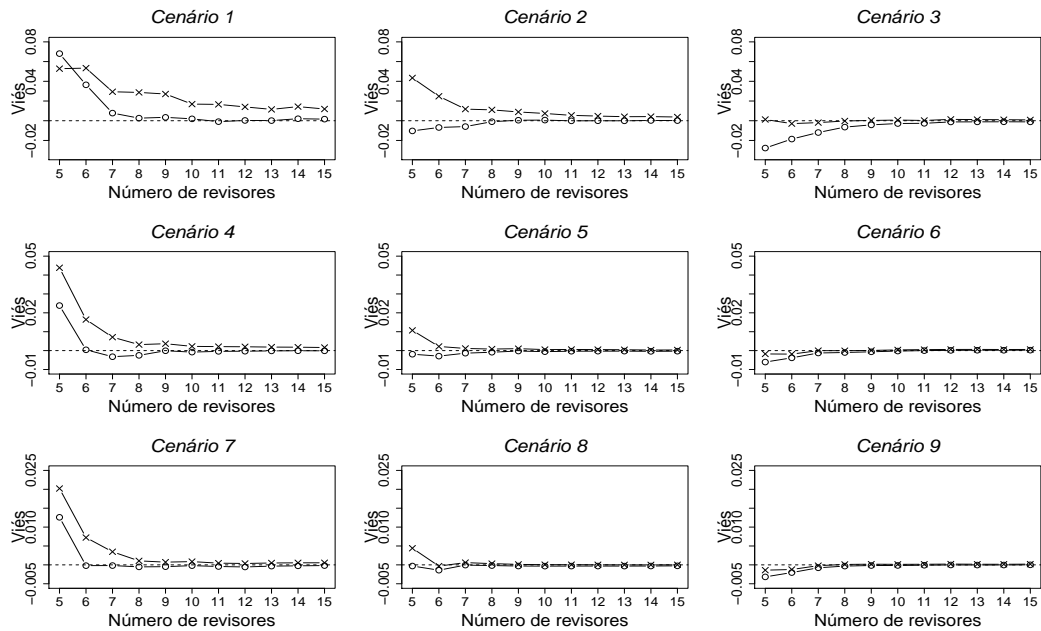


Figura 3.3: Comportamento do viés de $\hat{\theta}_D$ e $\hat{\theta}_{Dc}$ para os 9 cenários. *Legenda:* \times EMV; \circ EMVC.

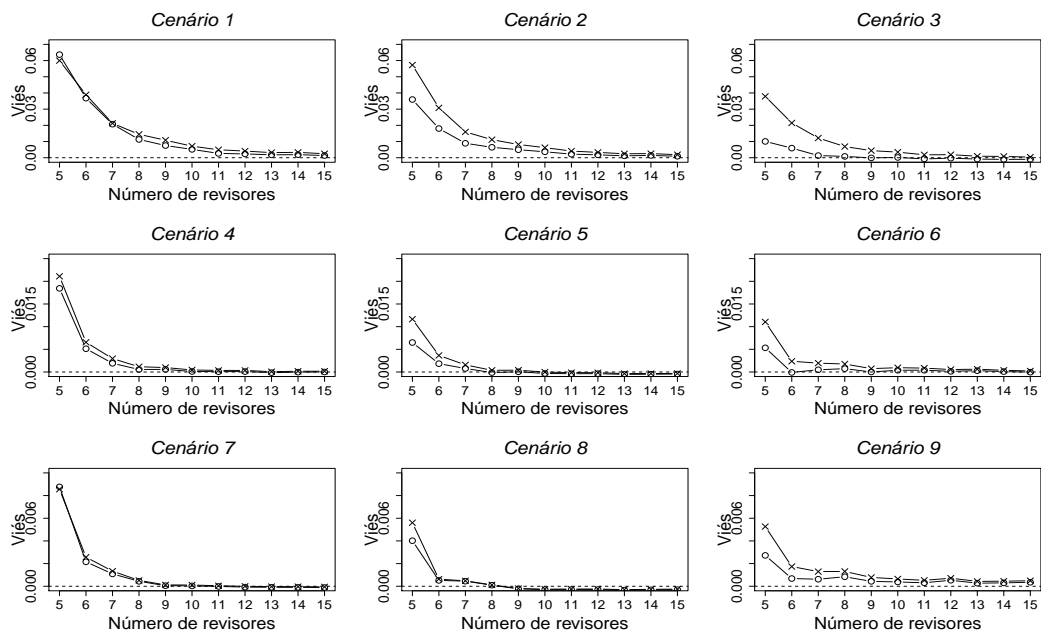


Figura 3.4: Comportamento do viés de $\hat{\theta}_F$ e $\hat{\theta}_{Fc}$ para os 9 cenários. *Legenda:* \times EMV; \circ EMVC.

Na Figura 3.5 observamos que os erros quadráticos médio das estimativas utilizando o método de verossimilhanças condicionais assumem valores mais altos que o método tradicional. Os maiores EQM's são para valores pequenos de k , ($k \leq 8$), valores estes de k dos quais resultaram em maiores viés relativos para as EMV's e EMVC's, conforme visto na Figura 3.1. Assim, concluímos

que os dois métodos produzem resultados semelhantes para os *Cenários 6, 8 e 9*, enquanto para os outros cenários, o método tradicional fornece estimativas para N menos viesadas e erros quadráticos médio menores do que ao método de verossimilhanças condicionais.

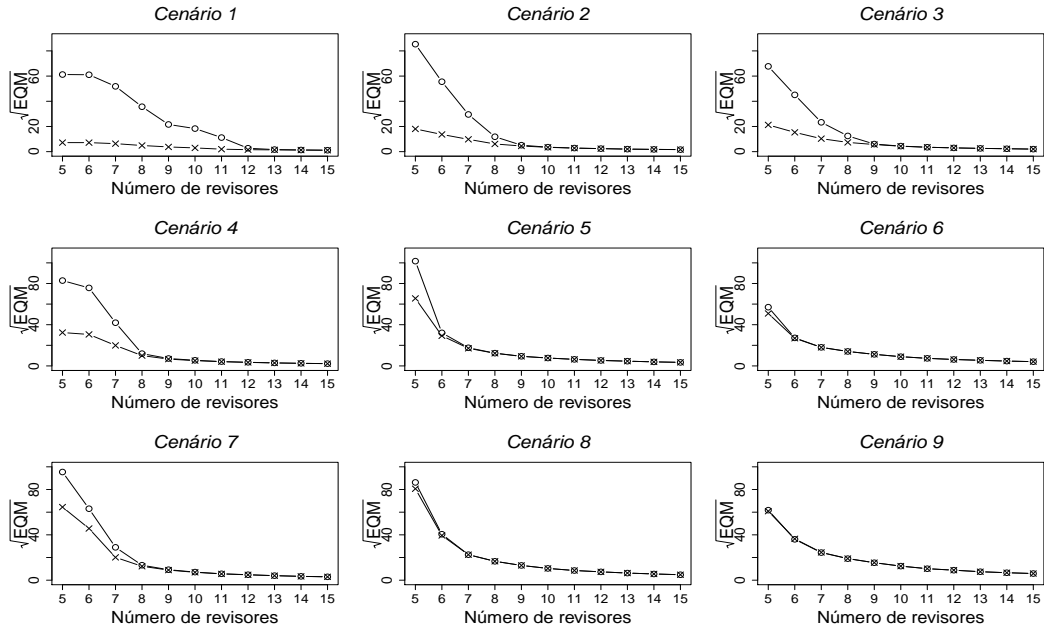


Figura 3.5: Comportamento do EQM de \hat{N} e \hat{N}_c para os 9 cenários. *Legenda:* \times EMV; \circ EMVC.

Notamos nas figuras 3.6, 3.7 e 3.8 que os valores dos erros quadráticos as estimativas para ω , θ_D e θ_F produzidos pelos dois métodos de estimação são bem próximos em todos os casos. Além disso, estas medidas quadráticas tem comportamento decrescente em todos os cenários, conforme o aumento do valor de k . Isto já era o esperado, pois notamos pelas Figuras 3.2, 3.3 e 3.4 que os vícios se aproximam de zero para maiores valores de k .

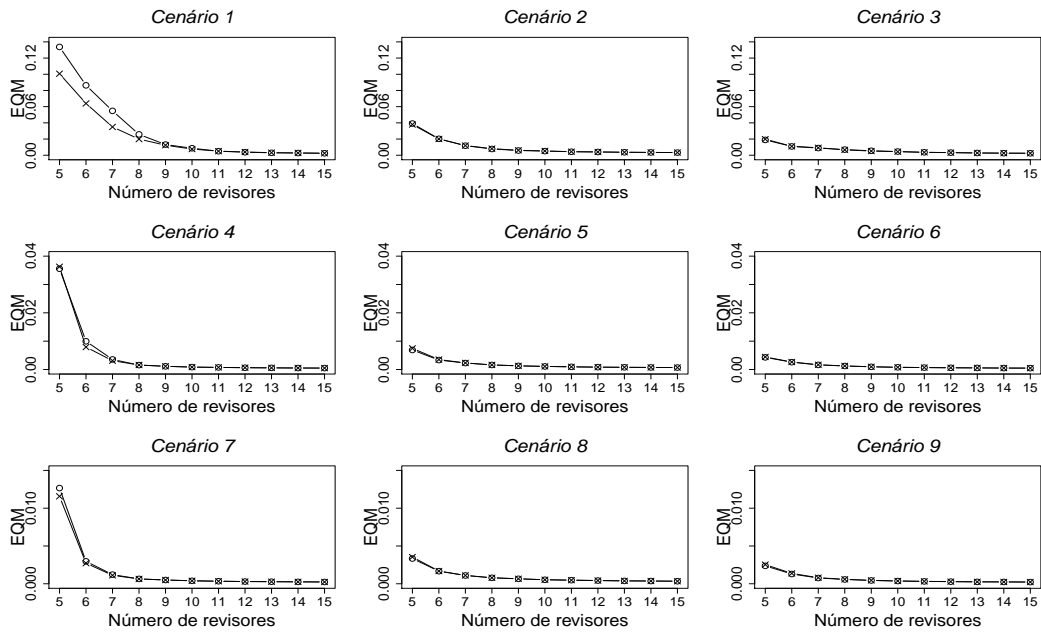


Figura 3.6: Comportamento do EQM de $\hat{\omega}$ e $\hat{\omega}_c$ para os 9 cenários. *Legenda:* × EMV; o EMVC.

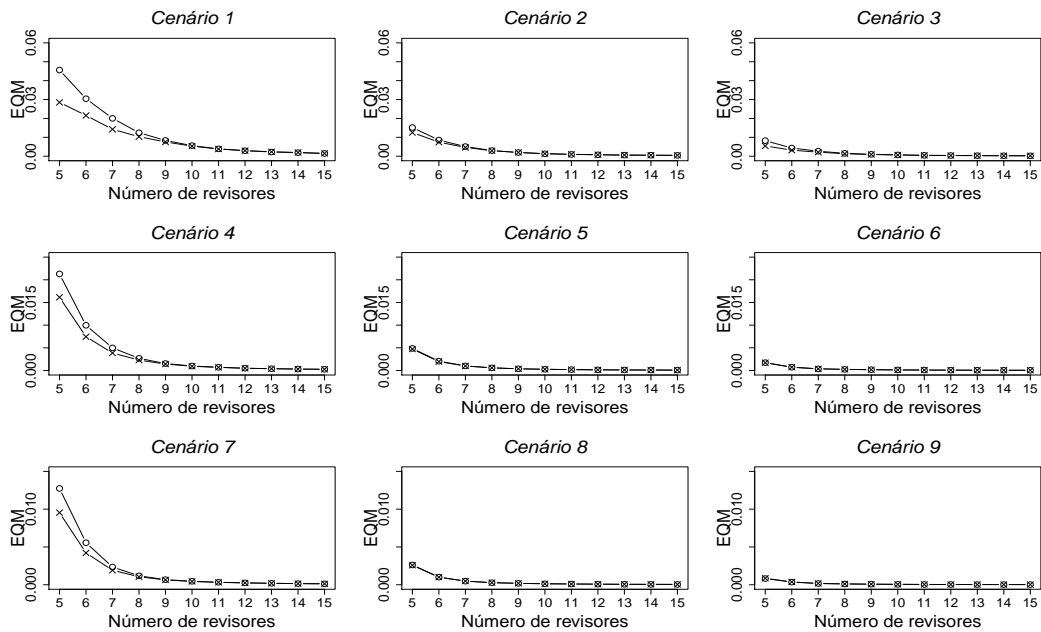


Figura 3.7: Comportamento do EQM de $\hat{\theta}_D$ e $\hat{\theta}_{Dc}$ para os 9 cenários. *Legenda:* × EMV; o EMVC.

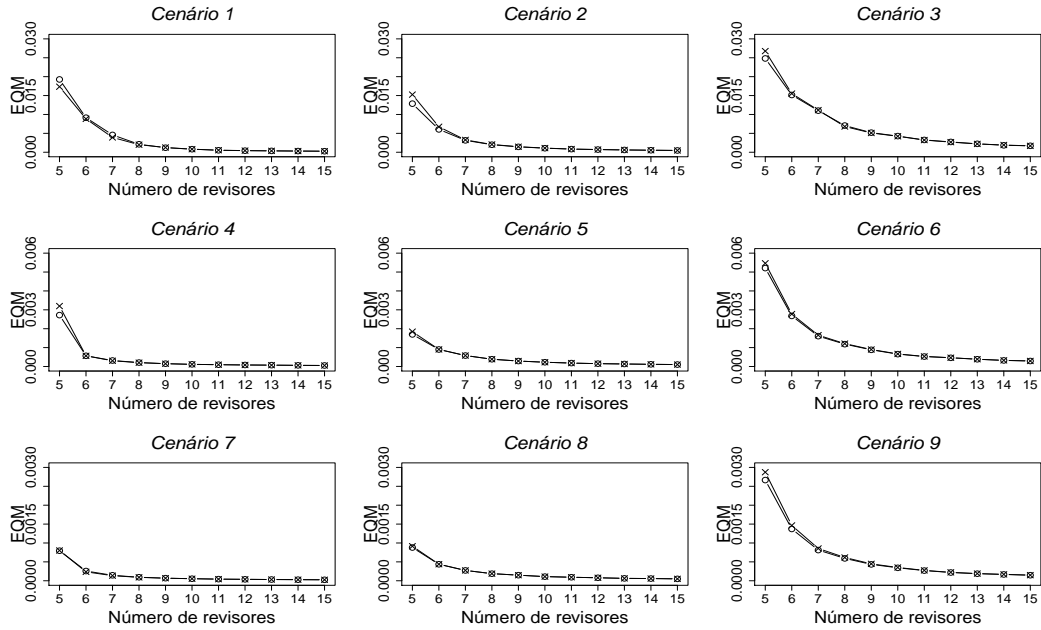


Figura 3.8: Comportamento do EQM de $\hat{\theta}_F$ e $\hat{\theta}_{Fc}$ para os 9 cenários. *Legenda:* \times EMV; \circ EMVC.

3.3.2 Aplicação em dados reais

Exemplo 1: Revisão de interruptores AT&T 5 ESS

Eick et al. (1993) utilizaram dados de captura-recaptura provenientes de erros identificados por uma característica particular em interruptores AT&T 5 ESS. Neste estudo, seis revisores observaram 43 erros distintos. Os dados completos são encontrados em Ebrahimi (1997) e na Tabela 3.1 temos as estatísticas suficientes para o modelo proposto.

Tabela 3.1: Frequências de erros observados em interruptores AT&T 5 ESS.

Frequências	f_1	f_2	f_3	f_4	f_5	f_6
Valores	30	11	1	0	1	0

Após o processo de revisão, 4 erros foram encontrados conjuntamente pelos revisores. Todavia, vamos desconsiderar esta informação na estimação dos parâmetros, pois estes erros não foram detectados de forma independente e, assim, não satisfazendo as suposições do modelo. Para o início dos algoritmos descritos nas Seções 3.1.3 e 3.2.2, determinamos a seguir uma maneira para obter valores iniciais para os parâmetros.

Considere um modelo sob suposição de homogeneidade e independência entre erros e entre revisores, isto é, a probabilidade do revisor j detectar o erro i é θ , para todo $1 \leq j \leq k$, $1 \leq i \leq N$. Neste caso, a função de verossimilhança de N e θ é dada por

$$L(N, \theta | \mathcal{D}) = \frac{N!}{f_1! \dots f_k! (N-n)!} \prod_{j=0}^k \left[\binom{k}{j} \theta^j (1-\theta)^{k-j} \right]^{f_j},$$

e, pelo método de verossimilhanças condicionais (Sanathanan, 1972), temos

$$L_2(\theta | f_1, \dots, f_k, n) \propto \prod_{j=1}^k \left[\frac{\theta^j (1-\theta)^{k-j}}{1 - (1-\theta)^k} \right]^{f_j}.$$

para $N \geq n$ e $\theta \in (0,1)$. A estimativa de máxima verossimilhança condicional para θ utilizando $L_2(\theta | \cdot)$ corresponde a resolver a equação

$$\frac{\theta}{1 - (1-\theta)^k} - \frac{\sum_{j=1}^k j f_j}{kn} = 0$$

em θ . Assim, de posse da EMVC de θ , $\hat{\theta}_c$, e pelo Teorema 3.1, temos que um possível valor inicial para N é dado por

$$N^{(0)} = \left\lceil \frac{n}{1 - (1 - \hat{\theta}_c)^k} \right\rceil.$$

Para os dados da Tabela 3.1, temos $\hat{\theta}_c = 0,1356$, $N^{(0)} = 73$ e $f_0^{(0)} = 30$.

Seja Y_i o número de revisores que detectaram o erro i , $1 \leq i \leq N^{(0)}$. Considere $Y^{(1)}$ o subconjunto dos dados $0 \leq Y_i \leq 3$, com $Y^{(1)} | \theta_D \sim \text{Binomial}(k, \theta_D)$ e $Y^{(2)}$ o subconjunto dos dados $4 \leq Y_i \leq 6$ com $Y^{(2)} | \theta_F \sim \text{Binomial}(k, \theta_F)$. Temos que $n_1 = f_0^{(0)} + f_1 + f_2 + f_3$ e $n_2 = f_4 + f_5 + f_6$ são os números de elementos em $Y^{(1)}$ e $Y^{(2)}$, respectivamente.

Sejam

$$m_D = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i^{(1)} = \frac{\sum_{j=0}^3 j f_j}{\sum_{j=0}^3 f_j} \quad \text{e} \quad m_F = \frac{1}{n_2} \sum_{i=1}^{n_2} y_i^{(2)} = \frac{\sum_{j=4}^6 j f_j}{\sum_{j=4}^6 f_j}$$

as médias amostrais de $Y^{(1)}$ e $Y^{(2)}$, respectivamente. Assim, através do método dos momentos, temos

$$\theta_D^{(0)} = \frac{1}{6} \frac{\sum_{j=0}^3 j f_j}{\sum_{j=0}^3 f_j}, \quad \theta_F^{(0)} = \frac{1}{6} \frac{\sum_{j=4}^6 j f_j}{\sum_{j=4}^6 f_j} \quad \text{e} \quad \omega^{(0)} = \frac{1}{N^{(0)}} \sum_{j=0}^3 f_j.$$

Logo, $\phi^{(0)} = (\omega^{(0)}, \theta_D^{(0)}, \theta_F^{(0)}) = (0,9863014; 0,1273148; 0,8333333)$.

As estimativas de máxima verossimilhança e máxima verossimilhanças condicionais encontram-se na Tabela 3.2. A precisão associada ao critério de parada dos algoritmos foi de $\epsilon = 0,00001$.

Tabela 3.2: EMV e EMVC dos parâmetros para o **Exemplo 1**.

Método	\hat{N}	$\hat{\omega}$	$\hat{\theta}_D$	$\hat{\theta}_F$
EMV	82	0,9873	0,1131	0,8114
EMVC	85	0,9877	0,1090	0,8100

Pela tabela acima, notamos que as estimativas de máxima verossimilhança e máxima verossimilhanças condicionais são bem próximas. Desta forma, utilizando as estimativas de máxima verossimilhanças, por exemplo, inferimos que há 82 erros em interruptores, dos quais 39 não foram observados pelos revisores. Destes erros, aproximadamente 98% são considerados de difícil detecção com probabilidade de 0,1131 de ser observado e que um erro de fácil detecção tem probabilidade 0,8114 de ser observado.

Exemplo 2: Contagem de coelhos

Edwards and Eberhardt (1967) usaram dados de captura-recaptura para estimação do tamanho populacional a partir de um estudo de aprisionamento de coelhos da espécie *cottontail*. O experimento consistiu na liberação de $N = 135$ coelhos em uma área de 40 hectares. Neste recinto não havia nenhuma espécie de coelhos antes do início do estudo. Um processo de 18 capturas-recapturas foi conduzido em dias consecutivos e observaram-se 76 coelhos distintos. As frequências observadas são dadas na Tabela 3.3. A precisão definida ao critério de parada dos dois algoritmos foi de $\epsilon = 0,00001$.

Tabela 3.3: Frequência de capturas de coelhos da espécie *cottontail*.

Frequências	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9
Valores	43	16	8	6	0	2	1	0	0
Frequências	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}
Valores	0	0	0	0	0	0	0	0	0

Utilizamos o mesmo artifício do exemplo anterior para obter valores iniciais para N e ϕ . Consideramos, neste exemplo, $Y^{(1)}$ o subconjunto dos dados $0 \leq Y_i \leq 4$ e $Y^{(2)}$ o subconjunto dos dados $5 \leq Y_i \leq 18$. Os valores iniciais foram $N^{(0)} = 97$, $\omega^{(0)} = 0,9690722$, $\theta_D^{(0)} = 0,07269504$ e $\theta_F^{(0)} = 0,3518519$.

As estimativas dos parâmetros se encontram na tabela 3.4.

Tabela 3.4: EMV e EMVC dos parâmetros para o **Exemplo 2**.

Método	\hat{N}	$\hat{\omega}$	$\hat{\theta}_D$	$\hat{\theta}_F$
EMV	130	0,8504	0,0389	0,1842
EMVC	142	0,8405	0,0328	0,1745

Primeiramente, note que as estimativas para os parâmetros perturbadores foram bem próximas para os dois métodos. Desta forma, inferimos que aproximadamente 80% dos coelhos são capturados com probabilidade 0,0389 e o restante tem probabilidade de captura de 0,1845, segundo as estimativas de máxima verossimilhança.

Ambos os métodos produziram estimativas para o parâmetro N satisfatórias, uma vez que o seu verdadeiro valor é 135 coelhos.

Capítulo 4

Abordagem Bayesiana: erros de fácil e de difícil detecção

Neste capítulo, consideramos a abordagem sob o enfoque Bayesiano da estimação dos parâmetros N e $\phi = (\omega, \theta_D, \theta_F)$ referentes ao modelo probabilístico descrito em (3.1).

Segue que a distribuição *a posteriori* conjunta para N e ϕ é dada por

$$\begin{aligned}\pi(N, \phi | \mathcal{D}) &\propto L(N, \phi | \mathcal{D}) \cdot \pi(N, \phi) \\ &\propto \binom{N}{n} h(0; \phi)^{N-n} \prod_{j=1}^k h(j; \phi)^{f_j} \cdot \pi(N, \phi),\end{aligned}\quad (4.1)$$

onde

$$h(j; \phi) = \binom{k}{j} \omega \theta_D^j (1 - \theta_D)^{k-j} + (1 - \omega) \binom{k}{j} \theta_F^j (1 - \theta_F)^{k-j}, \quad 0 \leq j \leq k,$$

e $\pi(N, \phi)$ é a distribuição *a priori* conjunta de N e ϕ .

Suponhamos independência *a priori* entre os parâmetros N e ϕ , isto é,

$$\pi(N, \phi) = \pi_1(N) \cdot \pi_2(\phi), \quad N \geq 1, \quad \phi \in \Phi, \quad (4.2)$$

onde $\pi_1(N)$ e $\pi_2(\phi)$ são as distribuições *a priori* marginais para N e ϕ , respectivamente.

Neste trabalho, consideramos $\pi_2(\phi)$ uma f.d.p. própria e $\pi_1(N) \propto N^{-r}$, $N \geq 1$, $r \geq 0$. Note que $\pi_1(N)$ não é própria para $0 \leq r \leq 1$. Porém, o teorema a seguir nos fornece uma condição

para existência da distribuição *a posteriori* conjunta de N e ϕ , neste caso.

Teorema 4.1. *Se a distribuição a priori $\pi(N, \phi) \propto \pi_1(N)\pi_2(\phi)$ com $\pi_1(N) = \frac{1}{N^r} \mathcal{I}_{\{N \geq 1\}}$, $r \geq 0$, e $\pi_2(\phi)$ uma f.d.p. própria com suporte $\Phi_{(c)} = \{(\omega, \theta_D, \theta_F) \in (0, 1)^3 : 0 < c < \theta_D < \theta_F \leq 1\}$, para c constante, então a posteriori $\pi(N, \phi | \mathcal{D})$ é própria.*

Demonstração. Considere K a constante normalizadora de $\pi(N, \phi | \mathcal{D})$, isto é,

$$K = \int_{\Phi_{(c)}} \sum_{N \geq n} \frac{1}{N^r} \binom{N}{n} h(0; \phi)^{N-n} \prod_{j=1}^k h(j; \phi)^{f_j} \pi_2(\phi) d\phi. \quad (4.3)$$

Mostremos que $K < \infty$. Para tanto, de (4.3) e de $N^r \geq 1, \forall r \geq 0$, segue que

$$\begin{aligned} K &\leq \int_{\Phi_{(c)}} \sum_{N \geq n} \binom{N}{n} h(0; \phi)^{N-n} \prod_{j=1}^k h(j; \phi)^{f_j} \pi_2(\phi) d\phi \\ &= \int_{\Phi_{(c)}} \prod_{j=1}^k h(j; \phi)^{f_j} \pi_2(\phi) \sum_{y=0}^{\infty} \binom{y+n}{n} h(0; \phi)^y d\phi \\ &= \int_{\Phi_{(c)}} \prod_{j=1}^k h(j; \phi)^{f_j} \pi_2(\phi) \sum_{y=0}^{\infty} \binom{-n-1}{y} [-h(0; \phi)]^y d\phi \\ &= \int_{\Phi_{(c)}} \prod_{j=1}^k h(j; \phi)^{f_j} \pi_2(\phi) [1 - h(0; \phi)]^{-n-1} d\phi \\ &= \int_{\Phi_{(c)}} \frac{\prod_{j=1}^k h(j; \phi)^{f_j} \pi_2(\phi)}{[1 - h(0; \phi)]^n [1 - h(0; \phi)]} d\phi \\ &= \int_{\Phi_{(c)}} \prod_{j=1}^k \left(\frac{h(j; \phi)}{1 - h(0; \phi)} \right)^{f_j} \frac{\pi_2(\phi)}{[1 - h(0; \phi)]} d\phi \\ &= \int_{\Phi_{(c)}} \prod_{j=1}^k \left(\frac{h(j; \phi)}{\sum_{i=1}^k h(i; \phi)} \right)^{f_j} \frac{\pi_2(\phi)}{[1 - h(0; \phi)]} d\phi \\ &\leq \int_{\Phi_{(c)}} \frac{\pi_2(\phi)}{[1 - h(0; \phi)]} d\phi. \end{aligned} \quad (4.4)$$

Observe que $(1 - \theta_D)^k \leq (1 - c)^k$ e $(1 - \theta_F)^k \leq (1 - c)^k$ para quaisquer θ_D e θ_F que satisfaçam $0 < c < \theta_D < \theta_F \leq 1$.

Portanto,

$$\begin{aligned}
 h(0; \phi) &= \omega(1 - \theta_D)^k + (1 - \omega)(1 - \theta_F)^k \\
 &\leq \omega(1 - c)^k + (1 - \omega)(1 - c)^k \\
 &= (1 - c)^k \\
 &\leq 1 - c, \quad \forall \phi \in \Phi_{(c)}.
 \end{aligned}$$

Consequentemente, temos que

$$\frac{1}{1 - h(0; \phi)} \leq \frac{1}{c}. \quad (4.5)$$

De (4.4) e (4.5), segue que

$$K \leq \int_{\Phi_{(c)}} \frac{1}{c} \pi_2(\phi) d\phi = \frac{1}{c} < \infty.$$

□

A seguir são apresentadas as distribuições condicionais *a posteriori* dos parâmetros N e $\phi = (\omega, \theta_D, \theta_F)$, as quais serão utilizadas no processo iterativo de simulação de Monte Carlo em Cadeias de Markov através do algoritmo *Gibbs sampling* (Casella and George, 1992).

4.1 Distribuições condicionais *a posteriori*

De (4.1), as distribuições condicionais *a posteriori* de N e ϕ são dadas por

$$\pi(N|\mathcal{D}, \phi) \propto \frac{N!}{(N - n)!} h(0; \phi)^N \frac{1}{N^r}, \quad N \geq n, \quad r \geq 0, \quad (4.6)$$

e, para $f_0 = N - n$,

$$\pi(\phi|\mathcal{D}, N) \propto \prod_{j=0}^k \left[\omega \theta_D^j (1 - \theta_D)^{k-j} + (1 - \omega) \theta_F^j (1 - \theta_F)^{k-j} \right]^{f_j} \cdot \pi(\phi), \quad \phi \in \Phi_{(c)}. \quad (4.7)$$

Pelo fato de (4.7) envolver produto de soma, deduzir estimadores para $\phi = (\omega, \theta_D, \theta_F)$ pode levar a um alto custo computacional e dificilmente poderão ser expressos analiticamente (Diebolt and Robert, 1993, 1994). No entanto, é possível e bastante vantajoso adicionar variáveis latentes, juntamente aos dados observados, na distribuições a *posteriori* no intuito de minimizar a complexidade da função (Celeux and Diebolt, 1985; Tanner and Wong, 1987; Diebolt and Robert, 1994). Conseqüentemente, o conjunto de variáveis completas na distribuição a *posteriori* condicional pode simplificar a implementação do algoritmo *Gibbs sampling* (Diebolt and Robert, 1994).

Desta forma, dados N, ϕ e \mathcal{D} , consideramos o conjunto de variáveis completas por $\mathbf{z} = (t_j, u_j, 0 \leq j \leq k)$ com distribuição condicional $p(\mathbf{z}|N, \phi, \mathcal{D})$, lembrando que t_i é o número de erros difíceis detectados por exatamente j revisores e u_j é o número de erros fáceis detectados por exatamente j revisores, $0 \leq j \leq k$. Uma observação importante é que não condicionaremos a distribuição de N em \mathbf{z} uma vez que, dado \mathbf{z} , o valor de N é determinístico. Isto é, dado \mathbf{z} , a distribuição de N é degenerada no ponto $N = \sum_{j=0}^k (t_j + u_j)$.

Assim, o processo de simulação de N , das variáveis completas (\mathbf{z}) e de ϕ de suas respectivas distribuições condicionais é esquematizado pelo seguinte algoritmo:

- inicialize o contador com $m = 1$ e valores iniciais $N^{(m)} \geq n$ e $\phi^{(m)} \in \Phi_{(c)}$;
- para m -ésima iteração, $m = 2, 3, \dots, M$,
 1. gere $N^{(m)} \sim \pi(N|\mathcal{D}, \phi^{(m-1)})$;
 2. simule as variáveis completas $\mathbf{z} \sim p(\mathbf{z}|\mathcal{D}, N^{(m)}, \phi^{(m-1)})$;
 3. gere $\phi^{(m)} \sim \pi(\phi|\mathcal{D}, N^{(m)}, \mathbf{z})$.

4.1.1 Distribuições condicionais a *posteriori* conhecidas para o tamanho populacional

Observamos que se r assumir os valores 0 ou 1 na distribuição *a priori* $\pi_1(N) \propto N^{-r}$, $N \geq 1$, então teremos as distribuições *a priori* uniforme discreta imprópria e de Jeffreys (Castledine, 1981; Smith, 1991), respectivamente. Assim, os dois teoremas a seguir são relevantes em termos computacionais, uma vez que nos fornecem distribuições condicionais a *posteriori* conhecidas para o parâmetro N nestes dois casos.

Teorema 4.2. *Se $\pi(N) \propto 1, N \geq 1$, então a distribuição condicional a posteriori de N , dados ϕ e \mathcal{D} , tem distribuição igual de uma variável aleatória $X + n$, onde X tem distribuição Binomial Negativa com parâmetros $n + 1$ e $1 - h(0; \phi)$.*

Demonstração. De fato, X tem distribuição Binomial Negativa com parâmetros $n + 1$ e $1 - h(0; \phi)$ se sua função distribuição de probabilidades é dada por

$$P(X = x) = \binom{x+n}{n} h(0; \phi)^x [1 - h(0; \phi)]^{n+1}, \quad x = 0, 1, 2, \dots$$

No entanto,

$$\begin{aligned} \pi(N|\phi, \mathcal{D}) &\propto \frac{N!}{(N-n)!} h(0; \phi)^N \\ &= c_1 \binom{N}{n} h(0; \phi)^N, \end{aligned}$$

onde a constante normalizadora c_1 é tal que

$$\begin{aligned} c_1^{-1} &= \sum_{N=n}^{\infty} \binom{N}{n} h(0; \phi)^N = \sum_{y=0}^{\infty} \binom{y+n}{n} h(0; \phi)^{y+n} \\ &= h(0; \phi)^n \sum_{y=0}^{\infty} \binom{y+n}{n} h(0; \phi)^y \\ &= h(0; \phi)^n \sum_{y=0}^{\infty} \binom{-n-1}{y} [-h(0; \phi)]^y \\ &= h(0; \phi)^n [1 - h(0; \phi)]^{-n-1}. \end{aligned}$$

Logo,

$$\pi(N|\phi, \mathcal{D}) = \binom{N}{n} h(0; \phi)^{N-n} [1 - h(0; \phi)]^{n+1}, \quad N = n, n+1, \dots, \quad (4.8)$$

é equivalente à

$$P(X + n = x) = P(X = x - n) = \binom{x}{n} h(0; \phi)^{x-n} [1 - h(0; \phi)]^{n+1}, \quad x = n, n+1, \dots$$

□

Teorema 4.3. *Se $\pi(N) \propto N^{-1}, N \geq 1$, então a distribuição condicional a posteriori de N , dados \mathcal{D} e ϕ , corresponde à distribuição de uma variável aleatória $Y + n$, onde Y tem uma distribuição Binomial Negativa com parâmetros n e $1 - h(0; \phi)$.*

Demonstração. Temos

$$\begin{aligned}\pi(N|\phi, \mathcal{D}) &\propto \frac{(N-1)!}{(N-n)!} h(0; \phi)^N \\ &= c_2 \binom{N-1}{n-1} h(0; \phi)^N,\end{aligned}$$

em que a constante normalizadora c_2 é tal que

$$\begin{aligned}c_2^{-1} &= \sum_{N=n}^{\infty} \binom{N-1}{n-1} h(0; \phi)^N = \sum_{y=0}^{\infty} \binom{y+n-1}{n-1} h(0; \phi)^{y+n} \\ &= h(0; \phi)^n \sum_{y=0}^{\infty} \binom{-n}{y} [-h(0; \phi)]^y \\ &= h(0; \phi)^n [1 - h(0; \phi)]^{-n}.\end{aligned}$$

Portanto,

$$\pi(N|\phi, \mathcal{D}) = \binom{N-1}{n-1} h(0; \phi)^{N-n} [1 - h(0; \phi)]^n, \quad N = n, n+1, \dots \quad (4.9)$$

Note que

$$P(Y = y) = \binom{y+n-1}{n-1} h(0; \phi)^y [1 - h(0; \phi)]^n, \quad y = 0, 1, 2, \dots$$

Logo,

$$P(Y + n = y) = P(Y = y - n) = \binom{y-1}{n-1} h(0; \phi)^{y-n} [1 - h(0; \phi)]^n, \quad y = n, n+1, \dots$$

□

4.1.2 *Posteriors* condicionais completas para os parâmetros perturbadores

De (3.9), temos a geração das variáveis completas $\mathbf{z} = (t_j, u_j, 0 \leq j \leq k)$ é dada por simular

$$t_j | N, \phi, \mathcal{D} \stackrel{ind.}{\sim} \text{Binomial}(f_j, p(j, \phi)) \quad (4.10)$$

onde $p(j, \phi)$ é definido em (3.10) e tomar $u_j = f_j - t_j$, $0 \leq j \leq k$.

Em seguida, de posse de \mathbf{z} , a função distribuição condicional *a posteriori* de ϕ , dados N , \mathcal{D} e \mathbf{z} , é dada por

$$\begin{aligned} \pi(\phi|N, \mathcal{D}, \mathbf{z}) \propto & \omega^{\sum_{j=0}^k t_j} (1 - \omega)^{\sum_{j=0}^k u_j} \theta_D^{\sum_{j=1}^k j t_j} (1 - \theta_D)^{\sum_{j=0}^k (k-j) t_j} \times \\ & \times \theta_F^{\sum_{j=1}^k j u_j} (1 - \theta_F)^{\sum_{j=0}^k (k-j) u_j} \cdot \pi(\phi), \end{aligned} \quad (4.11)$$

para $\phi \in \Phi_{(c)} = \{(\omega, \theta_D, \theta_F) \in (0,1)^3 : 0 < c < \theta_D < \theta_F \leq 1\}$.

Consideramos, *a priori*, (θ_D, θ_F) independentes de ω , isto é, $\pi(\phi) = \pi(\omega) \cdot \pi(\theta_D, \theta_F)$, onde

$$\omega \sim \text{Beta}(a_1, a_2) \quad (4.12)$$

com $a_1, a_2 > 0$ conhecidos e

$$(\theta_D, \theta_F) \sim \text{Uniforme}((0,1) \times (0,1) \cap \{0 < c < \theta_D < \theta_F \leq 1\}), \quad (4.13)$$

Logo, de (4.11), (4.12) e (4.13), temos

$$\begin{aligned} \pi(\phi|\mathbf{z}, N, \mathcal{D}) \propto & \omega^{\sum_{j=0}^k t_j + a_1 - 1} (1 - \omega)^{\sum_{j=0}^k u_j + a_2 - 1} \times \\ & \times \theta_D^{\sum_{j=1}^k j t_j} (1 - \theta_D)^{\sum_{j=0}^k (k-j) t_j} \times \\ & \times \theta_F^{\sum_{j=1}^k j u_j} (1 - \theta_F)^{\sum_{j=0}^k (k-j) u_j} \end{aligned}$$

para $(\omega, \theta_D, \theta_F) \in \Phi_{(c)} = \{(\omega, \theta_D, \theta_F) \in (0,1)^3 : 0 < c < \theta_D < \theta_F \leq 1\}$.

Portanto, as distribuições condicionais *a posteriori* completa de ω , θ_D e θ_F são dadas por

$$\pi(\omega|\mathbf{z}) \sim \text{Beta} \left(\sum_{j=0}^k t_j + a_1, \sum_{j=0}^k u_j + a_2 \right), \quad (4.14)$$

$$\pi(\theta_D|\theta_F, \mathbf{z}, N) \sim \text{Beta}\mathcal{T}_{(c, \theta_F)} \left(\sum_{j=1}^k j t_j + 1, \sum_{j=0}^k (k-j) t_j + 1 \right), \quad (4.15)$$

$$\pi(\theta_F|\theta_D, \mathbf{z}, N) \sim \text{Beta}\mathcal{T}_{(\theta_D, 1)} \left(\sum_{j=1}^k j u_j + 1, \sum_{j=0}^k (k-j) u_j + 1 \right), \quad (4.16)$$

em que $\text{Beta}\mathcal{T}_{(a,b)}(\cdot, \cdot)$ denota a distribuição $\text{Beta}(\cdot, \cdot)$ truncada no intervalo (a, b) .

O algoritmo *Gibbs sampling* é implementado a seguir e é baseado nas distribuições condicionais determinadas nas relações (4.6), (4.10), (4.14), (4.15) e (4.16).

4.2 Implemetação do algoritmo

- (1) Inicialize com $N^{(1)} \geq n$, $\phi^{(1)} \in \Phi_{(c)}$,
- (2) para a m -ésima iteração, $m = 2, \dots, M$;
 - (i) gere $N^{(m)} \sim \pi(N|\mathcal{D}, \phi^{(m-1)}) \propto \binom{N}{n} h(0; \phi)^N N^{-r}$, $r \geq 0$ (fixo),
 - (ii) tome $f_0 = N^{(m)} - n$, gere $t_j \sim \text{Binomial}(f_j, p(j; \phi^{(m-1)}))$ e faça $u_j = f_j - t_j$, $0 \leq j \leq k$,
 - (iii) gere $\omega^{(m)} \sim \text{Beta}(\sum_{j=0}^k t_j + a_1, \sum_{j=0}^k u_j + a_2)$,
 - (iv) gere $\theta_D^{(m)} \sim \text{Beta}\mathcal{T}_{(c, \theta_F^{(m-1)})}(\sum_{j=1}^k j t_j + 1, \sum_{j=0}^k (k-j) t_j + 1)$,
 - (v) gere $\theta_F^{(m)} \sim \text{Beta}\mathcal{T}_{(\theta_D^{(m)}, 1)}(\sum_{j=1}^k j u_j + 1, \sum_{j=0}^k (k-j) u_j + 1)$.

4.3 Resultados numéricos

Nesta seção, vamos obter estimativas bayesianas pontuais e intervalares para os parâmetros N e $\phi = (\omega, \theta_D, \theta_F)$ nos dois exemplos de dados reais citados na Seção 3.3.2.

Consideramos os valores dos hiperparâmetros $a_1 = a_2 = 1$ das distribuições *a priori* de ω . Desta forma, temos $\omega \sim \mathcal{U}(0,1)$. No entanto, informações sobre os valores a_1 e a_2 podem ser obtidas de especialistas considerando a reparametrização $a_1 = \kappa\rho$ e $a_2 = \kappa[1 - \rho]$, $\rho \in (0,1)$,

$\kappa > 0$, em (4.12). Assim, *a priori*, $\mathbb{E}[\alpha] = \rho$ e $\mathbb{V}[\alpha] = \frac{\rho[1 - \rho]}{\kappa + 1}$. Desta forma, interpretamos ρ como a crença que se tem *a priori* sobre a verdadeira proporção de erros difíceis no *software* e κ o quanto se acredita no valor de ρ . Valores altos de κ torna-se variância *a priori* baixa e, assim, com maior probabilidade em torno de ρ .

Assumimos o valor $c = 0.001$ na distribuição *a priori* $\pi(\theta_D, \theta_F)$, isto é,

$$(\theta_D, \theta_F) \sim \mathcal{U}((0,001; 1] \times (0,001; 1]) \mathcal{I}_{\{\theta_D < \theta_F\}}.$$

Para a distribuição *a priori* de N consideramos quatro valores para r , $(0; 0,5; 1; 1,5)$. Assim, para cada valor de r , simulamos 1000000 valores de cada parâmetro das suas respectivas distribuições condicionais utilizando o algoritmo *Gibbs sampling*, descartamos os 100000 primeiros valores como *burn in* e utilizamos saltos de 10 em 10, restando ao final 90000 valores.

O algoritmo *Gibbs sampling* foi implementado no Software R, cuja convergência foi monitorada pelo pacote **CODA - Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output** (Best et al., 1995). Analisamos a convergência das cadeias através do critério de Gelman and Rubin (1992). Para utilizar este diagnóstico de convergência, geramos duas cadeias em paralelo e, assim, obtivemos duas amostras *a posteriori* para cada parâmetro. Um indicativo de convergência é a estimativa do fator de redução do parâmetro de escala da distribuição marginal *a posteriori* de cada variável. Os diagnósticos apresentados pelo pacote **CODA** são os quantis 50% (mediana) e 97,5% das amostras geradas. Se ambos os quantis estiverem nas proximidades de 1, usualmente $< 1,1$, temos fortes indícios de convergência.

4.3.1 Exemplo 1: Revisão de interruptores AT&T 5 ESS

Sejam os dados referentes à Tabela 3.1. Os valores do diagnóstico de Gelman and Rubin (1992) para os parâmetros são dados na Tabela 4.1. Pela Figura 4.3 no final do capítulo, observamos que os quantis 50% (mediana) e 97,5% se estabilizam em torno de 1,0 para todos os parâmetros, dando indícios de convergência das cadeias.

Tabela 4.1: Diagnóstico de convergência de Gelman e Rubin para o **Exemplo 1**.

Valor de	N	ω	θ_D	θ_F
r à priori	Est. (Q.97,5%)	Est. (Q.97,5%)	Est. (Q.97,5%)	Est. (Q.97,5%)
0	1.0208 (1.0226)	1.0002 (1.0006)	1.0000 (1.0000)	1.0000 (1.0002)
0,5	1.0004 (1.0004)	1.0010 (1.0036)	1.0002 (1.0007)	1.0005 (1.0025)
1	1.0568 (1.0597)	1.0003 (1.0016)	1.0002 (1.0011)	1.0003 (1.0017)
1,5	1.0026 (1.0026)	1.0000 (1.0000)	1.0000 (1.0001)	1.0000 (1.0000)

Os resumos *a posteriori* para o parâmetro N são encontrados na Tabela 4.2, onde Q_1 é o quantil 25%, Q_2 é o quantil 50% e Q_3 é o quantil 75% dos valores do parâmetro N simulados *a posteriori*. O Int. Cred (95%) corresponde o intervalo de credibilidade de 95% para N , isto é, os percentis 2,5% e 97,5% da distribuição *a posteriori* de N . Basu and Ebrahimi (2001) obteve resultados *a posteriori* de 89,8 erros (média), 84 erros (mediana) e intervalo de credibilidade 95% [58 ; 158] para N assumindo a *priori* uniforme discreta imprópria ($r = 0$).

Tabela 4.2: Resumos *a posteriori* para o parâmetro N para o **Exemplo 1**.

Valor de	Estimativas Bayesianas						
	r à priori	Média	Moda	Q_1	Mediana	Q_3	Int.Cred.(N ; 95%)
0		112,94	77	75	88	109	[60 ; 326]
0,5		97,71	75	73	85	102	[59 ; 212]
1		90,37	73	71	82	97	[58 ; 172]
1,5		85,90	74	70	80	94	[57 ; 151]

Na Figura 4.1 são apresentadas as densidades estimadas da distribuição *a posteriori* do parâmetro N de acordo com a escolha do valor de r *a priori*. Como esperado, maior valor de r torna a densidade com massa em valores baixos de N .

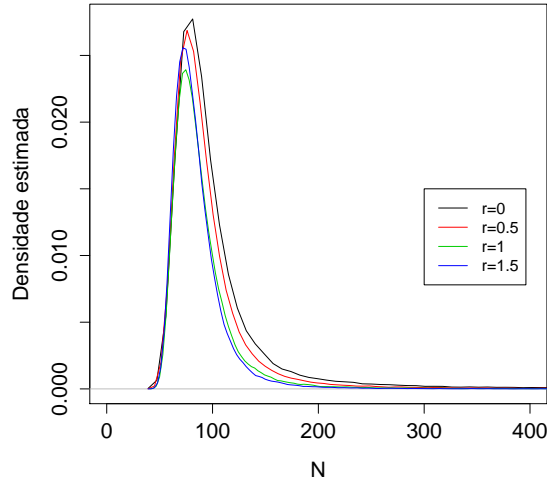


Figura 4.1: Densidades estimadas da distribuição a *posteriori* de N de acordo com o valor de r a *priori* - Exemplo 1.

Na Tabela 4.3 a seguir são apresentados as médias e intervalos de credibilidade *posteriori* dos parâmetros perturbadores do modelo. As estimativas médias obtidas por Basu and Ebrahimi (2001) foram 0,829, 0,098 e 0,450 para ω , θ_D e θ_F , respectivamente. Não notamos influência nas estimativas dos parâmetros perturbadores a partir da escolha do hiperparâmetro r a *priori* para N .

Tabela 4.3: Resumos a *posteriori* para os parâmetros perturbadores para o **Exemplo 1**.

Valor de r à priori	Parâmetro ω		Parâmetro θ_D		Parâmetro θ_F	
	Média	I.Cb. (ω ; 95%)	Média	I.Cb. (θ_D ; 95%)	Média	I.Cb. (θ_F ; 95%)
0	0,8072	(0,1319 ; 0,9949)	0,0884	(0,0056 ; 0,1653)	0,4070	(0,1169 ; 0,9209)
0,5	0,8005	(0,1130 ; 0,9949)	0,0942	(0,0099 ; 0,1688)	0,4169	(0,1178 ; 0,9232)
1	0,7971	(0,1013 ; 0,9950)	0,0990	(0,0138 ; 0,1726)	0,4228	(0,1197 ; 0,9266)
1,5	0,7902	(0,0919 ; 0,9950)	0,1026	(0,0168 ; 0,1742)	0,4228	(0,1205 ; 0,9251)

4.3.2 Exemplo 2: Contagem de coelhos

Na Tabela 4.4 se encontram os resumos a *posteriori* para o parâmetro N com relação aos dados apresentados na Tabela 3.3. Utilizamos neste exemplo as mesmas *prioris* para os parâmetros do

exemplo anterior. As estimativas média e mediana encontradas por [Basu and Ebrahimi \(2001\)](#) para o parâmetro N foram respectivos 146 e 135 coelhos, com intervalo de credibilidade para a média [99 ; 262]. Apresentamos resultados parecidos utilizando a distribuição não informativa $\pi(N) = 1/N$, isto é, ($r = 1$). As melhores estimativas foram considerando *a priori* $r = 1,5$, sabendo que o número verdadeiro de coelhos era de $N = 135$.

Tabela 4.4: Resumos a posteriori para o parâmetro N para o **Exemplo 2**.

Valor de r à priori	Estimativas Bayesianas					
	Média	Moda	Q_1	Mediana	Q_3	Int.Cred.(N ; 95%)
0	177,66	115	119	139	177	[98 ; 538]
0,5	160,43	117	116	134	165	[97 ; 399]
1	145,67	115	114	129	154	[96 ; 295]
1,5	138,20	114	112	126	148	[96 ; 252]

Pela Figura 4.2 temos que as densidades estimadas *a posteriori* do parâmetro N têm aproximadamente a mesma forma para os quatro valores de r *a priori*. No entanto, valores menores de r faz com que a cauda da distribuição *a posteriori* seja mais longa à direita, resultando em valor médio *a posteriori* maior para N nestes casos, como visto na Tabela 4.4.

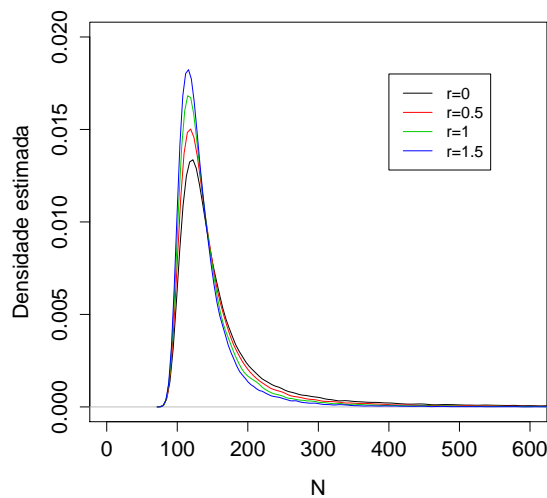


Figura 4.2: Densidades estimadas da distribuição *a posteriori* de N de acordo com o valor de r *a priori* - Exemplo 2.

A Tabela 4.5 mostra os resumos a *posteriori* para os parâmetros perturbadores. As estimativas encontradas para os parâmetros ω , θ_D e θ_F por Basu and Ebrahimi (2001) foram 0,825, 0,036 e 0,181, respectivamente, condizendo com os nossos resultados.

Tabela 4.5: Resumos a *posteriori* para os parâmetros perturbadores - Exemplo 2.

Valor de r à priori	Parâmetro ω		Parâmetro θ_D		Parâmetro θ_F	
	Média	I.Cb. (ω ; 95%)	Média	I.Cb. (θ_D ; 95%)	Média	I.Cb. (θ_F ; 95%)
0	0,8369	(0,6336 ; 0,9607)	0,0340	(0,0038 ; 0,0690)	0,1794	(0,1062 ; 0,3077)
0,5	0,8348	(0,6220 ; 0,9612)	0,0365	(0,0055 ; 0,0707)	0,1832	(0,1068 ; 0,3176)
1	0,8350	(0,6185 ; 0,9625)	0,0390	(0,0084 ; 0,0718)	0,1876	(0,1085 ; 0,3254)
1,5	0,8337	(0,6076 ; 0,9635)	0,0406	(0,0106 ; 0,0726)	0,1899	(0,1083 ; 0,3281)

Pela Tabela 4.6, temos indícios de convergência das cadeias, isto é, estimativas medianas e 97,5% do critério de Gelman Rubin foram menores que 1.1. A representação gráfica é dada pela figura 4.4, que mostra os quantis estabilizados em torno de 1 para todos os parâmetros, o que reforça o indício de convergência do algoritmo.

Tabela 4.6: Diagnóstico de convergência de Gelman e Rubin - Exemplo 2.

Valor de r à priori	N	ω	θ_D	θ_F
	Est. (Q.97,5%)	Est. (Q.97,5%)	Est. (Q.97,5%)	Est. (Q.97,5%)
0	1.0021 (1.0034)	1.0003 (1.0008)	1.0004 (1.0018)	1.0005 (1.0023)
0,5	1.0000 (1.0000)	1.0018 (1.0028)	1.0000 (1.0000)	1.0000 (1.0002)
1	1.0001 (1.0003)	1.0005 (1.0006)	1.0000 (1.0001)	1.0000 (1.0000)
1,5	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0001)	1.0001 (1.0001)

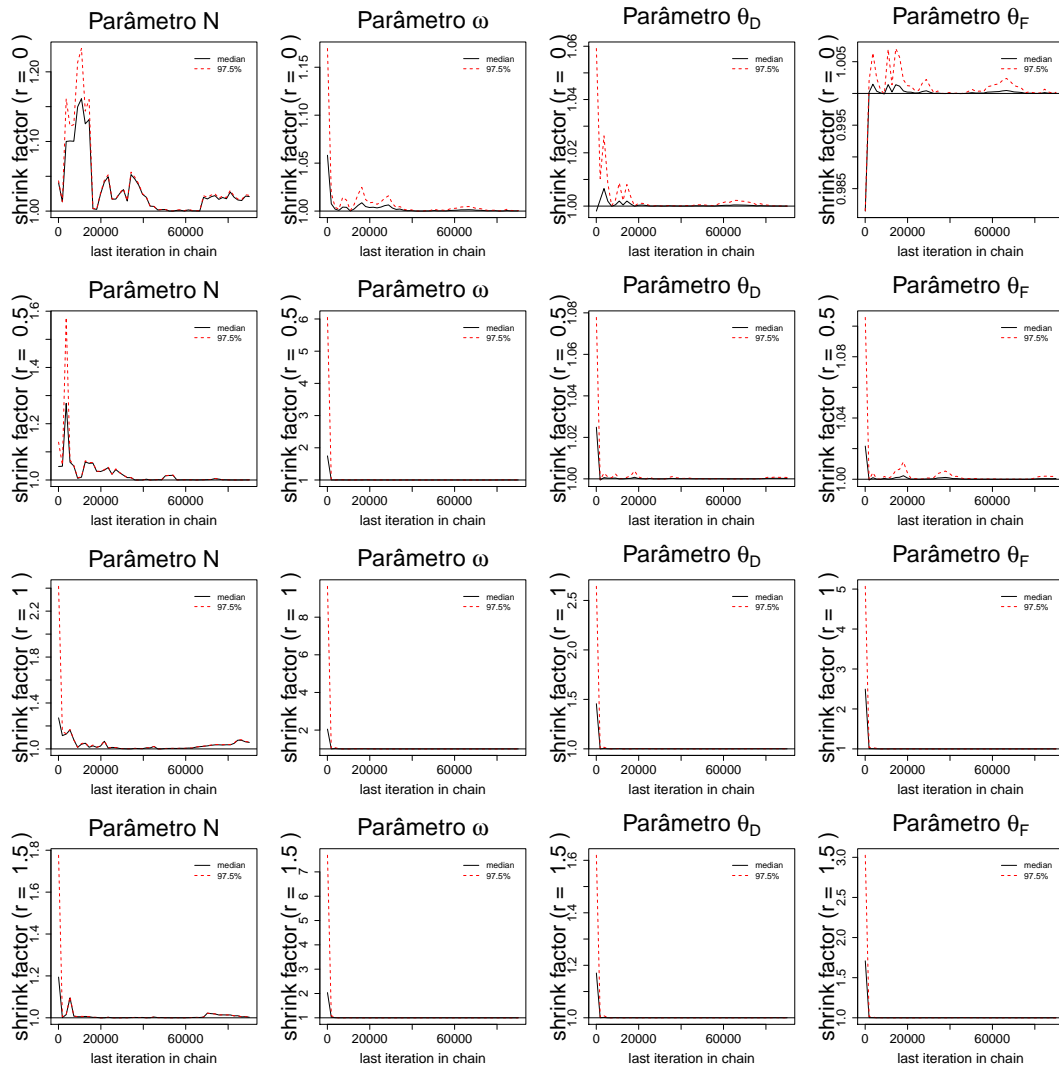


Figura 4.3: Diagnóstico de convergência de Gelman e Rubin - Exemplo 1.

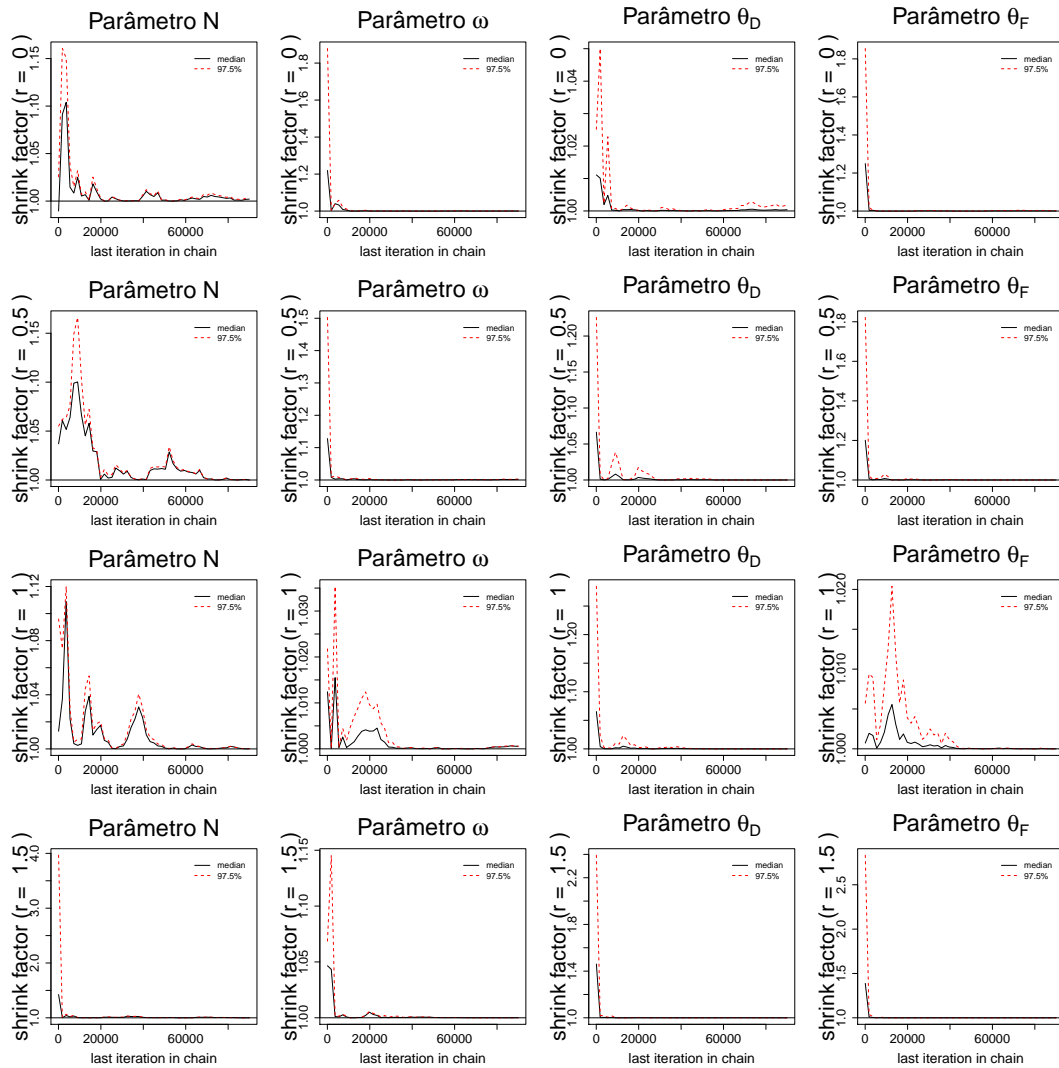


Figura 4.4: Diagnóstico de convergência de Gelman e Rubin - Exemplo 2.

Capítulo 5

Modelo probabilístico: caso com l níveis de dificuldade

Neste capítulo, supomos que a população de erros no *software* pode ser particionada em l , $l \geq 2$, subpopulações disjuntas, em que os erros são classificados de acordo exclusivamente com seus níveis de dificuldade de detecção. Se considerarmos $l = 4$ subpopulações, dizemos que a heterogeneidade da população, por exemplo, é dada por erros de *fácil*, *moderada*, *difícil* e *muito difícil* detecção.

Definimos

- ω_s : a probabilidade de um erro i ($1 \leq i \leq N$) pertencer ao nível de dificuldade s ($1 \leq s \leq l$);
- $\omega \in \Omega$, onde $\omega = (\omega_1, \dots, \omega_l)$ e $\Omega = \{(\omega_1, \dots, \omega_l) \in (0,1)^l : \sum_{s=1}^l \omega_s = 1\}$;
- θ_s : probabilidade de captura de um erro com nível de dificuldade s , ($1 \leq s \leq l$);
- $\theta \in \Theta$, onde $\theta = (\theta_1, \dots, \theta_l)$ e $\Theta = \{(\theta_1, \dots, \theta_l) \in (0,1)^l : \theta_1 < \dots < \theta_l\}$;
- $\psi \in \Psi$, onde $\psi = (\omega, \theta)$ e $\Psi = \Omega \times \Theta$.

Portanto, pela relação (3.1), temos a função de verossimilhança para N e ψ dada por

$$L(N, \psi | \mathcal{D}) \propto \frac{N!}{(N-n)!} \prod_{j=0}^k \left\{ \sum_{s=1}^l \omega_s \theta_s^j (1 - \theta_s)^{k-j} \right\}^{f_j} \quad (5.1)$$

para $N \geq n$, $f_0 = N - n$ e $\psi \in \Psi$.

5.1 Estimativas de máxima verossimilhança

Para obter as estimativas de máxima verossimilhança dos parâmetros N e $\boldsymbol{\psi}$, determinamos um processo iterativo entre seus estimadores, como visto no esquema dado na Seção 3.1.3. Isto é, fixado um valor para $\boldsymbol{\psi}$, obtemos uma estimativa para N , e vice versa, fixado N , obtemos uma estimativa para $\boldsymbol{\psi}$, em que estas etapas são repetidas iterativamente até que algum critério de convergência seja satisfeito.

Assim, dados $\boldsymbol{\psi}$ e n , seja

$$b(n, \boldsymbol{\psi}) = \frac{n}{1 - h'(0; \boldsymbol{\psi})},$$

onde $h'(0; \boldsymbol{\psi}) = \sum_{s=1}^l \omega_s (1 - \theta_s)^k$.

Logo, do Teorema 3.1 e $\boldsymbol{\psi}$ fixo, temos o

Teorema 5.1.

(a) Caso $b(n, \boldsymbol{\psi})$ não seja um número inteiro, então a função de verossimilhança em (5.1) tem um único ponto de máximo,

$$\hat{N} = [b(n, \boldsymbol{\psi})],$$

onde $[\bullet]$ é a função que associa o maior inteiro menor que o argumento;

(b) caso $b(n, \boldsymbol{\psi})$ seja um número inteiro, então a função de verossimilhança em (5.1) tem dois pontos de máximo,

$$\hat{N}_1 = b(n, \boldsymbol{\psi}) \quad e \quad \hat{N}_2 = b(n, \boldsymbol{\psi}) - 1.$$

Demonstração. A demonstração é análoga à do Teorema 3.1. □

Com o auxílio do algoritmo EM e um valor fixo de N usamos a função de verossimilhança completa na estimação de $\boldsymbol{\psi}$.

Dado N , seja $v_{j,s}$: “número de erros pertencentes à população s detectados por exatamente j revisores” e denotamos o vetor de dados completos por $\boldsymbol{v} = (\boldsymbol{v}_0, \boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$, em que $\boldsymbol{v}_j = (v_{j,1}, \dots, v_{j,l})$, $0 \leq j \leq k$, $1 \leq s \leq l$.

Desta forma, a função de verossimilhança completa para $\boldsymbol{\psi}$ é dada por

$$L(\boldsymbol{\psi}|\mathbf{v}) \propto \prod_{j=0}^k \prod_{s=1}^l \left\{ \omega_s \theta_s^j (1 - \theta_s)^{k-j} \right\}^{v_{j,s}}, \quad (5.2)$$

para $\boldsymbol{\psi} \in \boldsymbol{\Psi}$.

Pela relação (3.11), temos que a distribuição das variáveis latentes condicionada aos dados observados é dada por

$$V_{j,1}, \dots, V_{j,l} | f_j, \boldsymbol{\psi} \sim \text{Multinomial}(f_j, \mathbf{q}(j; \boldsymbol{\psi})), \quad (5.3)$$

onde $\mathbf{q}(j; \boldsymbol{\psi}) = (q(j,1; \boldsymbol{\psi}), \dots, q(j,l; \boldsymbol{\psi}))$ e

$$q(j,s; \boldsymbol{\psi}) = \frac{\omega_s \theta_s^j (1 - \theta_s)^{k-j}}{\sum_{i=1}^l \omega_i \theta_i^j (1 - \theta_i)^{k-j}}, \quad 0 \leq j \leq k, \quad 1 \leq s \leq l.$$

Assim, dado um valor de $\boldsymbol{\psi}^{(m)} \in \boldsymbol{\Psi}$, substituímos as informações “ocultas” $v_{j,s}$ ’s por seus respectivos valores esperados, $E[V_{j,s} | f_j, \boldsymbol{\psi}^{(m)}] = q(j,s; \boldsymbol{\psi}^{(m)}) f_j$, no logaritmo de (5.2), e obtemos a função

$$Q(\boldsymbol{\psi} | \boldsymbol{\psi}^{(m)}) = c + \sum_{j=0}^k \sum_{s=1}^l q(j,s; \boldsymbol{\psi}^{(m)}) f_j [\log \omega_s + j \log \theta_s + (k-j) \log(1 - \theta_s)], \quad (5.4)$$

para $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ e c constante em relação à $\boldsymbol{\psi}$.

Seja $\boldsymbol{\psi}^{(m+1)}$ à solução de

$$\left. \frac{\partial Q(\boldsymbol{\psi} | \boldsymbol{\psi}^{(m)})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi} = \boldsymbol{\psi}^{(m+1)}} = \mathbf{0},$$

então, segue por cálculos análogos à (3.15) que

$$\theta_s^{(m+1)} = \frac{\sum_{j=1}^k j q(j,s; \boldsymbol{\psi}^{(m)}) f_j}{k \sum_{j=0}^k q(j,s; \boldsymbol{\psi}^{(m)}) f_j}, \quad 1 \leq s \leq l \quad (5.5)$$

Pela Proposição 3.2, temos que, se $\theta_s^{(m)} < \theta_{s+1}^{(m)}$, então $\theta_s^{(m+1)} < \theta_{s+1}^{(m+1)}$, para todo $1 \leq s \leq (l-1)$.

Por meio de *multiplicadores de Lagrange*, resolvemos

$$\begin{cases} (a) & \frac{\partial Q(\boldsymbol{\psi}, \lambda | \boldsymbol{\psi}^{(m)})}{\partial \boldsymbol{\omega}} = \mathbf{0}, \\ (b) & \text{sujeito à } \sum_{s=1}^l \omega_s = 1, \end{cases} \quad (5.6)$$

onde

$$Q(\boldsymbol{\psi}, \lambda | \boldsymbol{\psi}^{(m)}) = Q(\boldsymbol{\psi} | \boldsymbol{\psi}^{(m)}) - \lambda \left(\sum_{s=1}^l \omega_s - 1 \right),$$

para $\lambda \in \mathbb{R}$. De (a) em (5.6), temos

$$0 = \frac{dQ(\boldsymbol{\psi}, \lambda | \boldsymbol{\psi}^{(m)})}{d\omega_s} = \sum_{j=0}^k q(j, s; \boldsymbol{\psi}^{(m)}) f_j \frac{1}{\omega_s} - \lambda \Rightarrow \omega_s = \frac{1}{\lambda} \sum_{y=0}^k q(j, s; \boldsymbol{\psi}^{(m)}) f_j,$$

para $1 \leq s \leq l$. Para (b) em (5.6), devemos ter $\lambda = N$. Portanto,

$$\omega_s^{(m+1)} = \frac{1}{N} \sum_{j=0}^k q(j, s; \boldsymbol{\psi}^{(m)}) f_j, \quad 1 \leq s \leq l. \quad (5.7)$$

A seguir, pelo Teorema 5.1 e as relações (5.5) e (5.7), descrevemos a implementação do algoritmo para obtenção das estimativas de máxima verossimilhança de N e $\boldsymbol{\psi} = (\omega_s, \theta_s; 1 \leq s \leq l)$ para o modelo em (5.1).

5.1.1 Implementação do algoritmo

Inicialize o algoritmo com $N^{(0)} \geq n$, $\boldsymbol{\psi}^{(0)} \in \boldsymbol{\Psi}$ e $(m = 0)$;

(1) Algoritmo EM:

(i) Fixe $N = N^{(m)}$. Faça $(t = 0)$ e $\boldsymbol{\psi}^{(t)} = \boldsymbol{\psi}^{(m)}$.

(ii) Para $1 \leq s \leq l$, repita iterativamente

$$\theta_s^{(t+1)} = \frac{\sum_{j=1}^k j q(j, s; \boldsymbol{\psi}^{(t)}) f_j}{k \sum_{j=0}^k q(j, s; \boldsymbol{\psi}^{(t)}) f_j} \quad \text{e} \quad \omega_s^{(t+1)} = \frac{1}{N} \sum_{j=0}^k q(j, s; \boldsymbol{\psi}^{(t)}) f_j,$$

substituindo $t := (t + 1)$ a cada iteração.

(iii) Interrompa (ii) quando $|\boldsymbol{\psi}^{(t+1)} - \boldsymbol{\psi}^{(t)}| < \epsilon$ e vá para o passo (2).

(2) Para $\boldsymbol{\psi}^{(m+1)} = \boldsymbol{\psi}^{(t+1)}$, atualize N ,

$$N^{(m+1)} = \text{maior inteiro menor que } \frac{n}{1 - h'(0; \boldsymbol{\psi}^{(m+1)})},$$

$$\text{onde } h'(0; \boldsymbol{\psi}^{(m+1)}) = \sum_{s=1}^l \omega_s^{(m+1)} (1 - \theta_s^{(m+1)})^k.$$

Se $N^{(m+1)} = N^{(m)}$, finalize o algoritmo. Caso contrário, $m := (m + 1)$ e volte ao passo (1).

5.1.2 Resultados numéricos

Nesta seção, obtemos as estimativas de máxima verossimilhança dos parâmetros N e $\boldsymbol{\psi}$ que indexam o modelo em (5.1). A aplicação foi através de dados artificiais obtidos via simulação.

Primeiramente, para fins de ilustração, consideramos o modelo probabilístico no caso em que há três níveis de dificuldade de captura. Simulamos alguns cenários amostrais e avaliamos a precisão dos estimadores através do cálculo do viés e o erro quadrático médio.

Em seguida, foi realizada a comparação entre modelos com diferentes níveis de captura utilizando os critérios AIC (Akaike, 1973) e BIC (Schwarz et al., 1978). Estes critérios de seleção de modelos são determinados por

$$\text{AIC} = -2 \log (L(\hat{N}, \hat{\boldsymbol{\psi}} | \mathcal{D})) + 2p,$$

$$\text{BIC} = -2 \log (L(\hat{N}, \hat{\boldsymbol{\psi}} | \mathcal{D})) + p \log(n),$$

onde p é o número de parâmetros no modelo, \hat{N} e $\hat{\boldsymbol{\psi}}$ são as respectivas estimativas de máxima verossimilhança de N e $\boldsymbol{\psi}$ e n é o tamanho amostral. Segundo Cubaynes et al. (2012), n é o número de indivíduos capturados em pelo menos uma ocasião.

No estudo de simulação, atribuímos valores para k entre 5 e 12 e consideramos os três tamanhos populacionais, $N = 100, 500$ e 1000 . Em todas as simulações, os valores para os parâmetros perturbadores foram fixados em $(\omega_1, \omega_2, \omega_3) = (0,6; 0,3; 0,1)$ e $(\theta_1, \theta_2, \theta_3) = (0,1; 0,5; 0,8)$ e, em cada simulação, 1000 amostras foram simuladas e os parâmetros estimados.

A Figura 5.1 mostra o comportamento do viés (à esquerda) e do erro quadrático médio (à direita) das estimativas do tamanho populacional de acordo com o número de revisores (k) e o valor atribuído para N (100, 500 ou 1000). O pior cenário para o viés foi com $k = 5$ revisores.

No entanto, com seis ou mais revisores os vícios diminuíram significativamente, porém, valores baixos para o EQM só são observados com pelo menos 10 revisores.

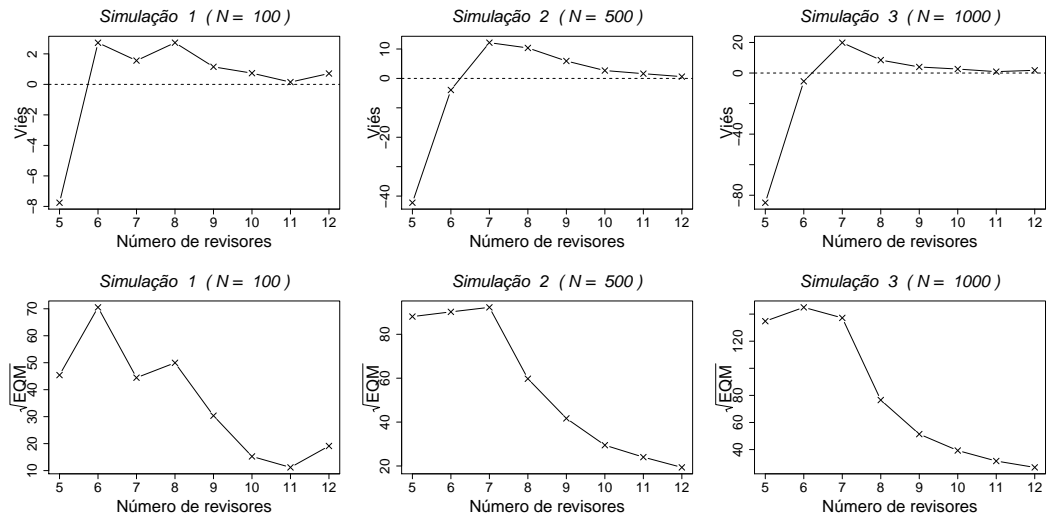


Figura 5.1: Comportamento do Viés e da raiz quadrada do EQM das EMV's de N .

Nas Figuras 5.2 e 5.3, é possível notar que com o aumento do número de revisores, o viés e o erro quadrático médio das estimativas de máxima verossimilhança dos $\hat{\omega}_s$'s e $\hat{\theta}_s$'s tendem a diminuir, independentemente do valor adotado para N (100, 500 ou 1000). Isto já era esperado, pois com mais revisores, se tem mais informações sobre a população e, conseqüentemente, melhores estimativas para os parâmetros. No mais, consideramos satisfatórios os resultados obtidos neste estudo de simulação.

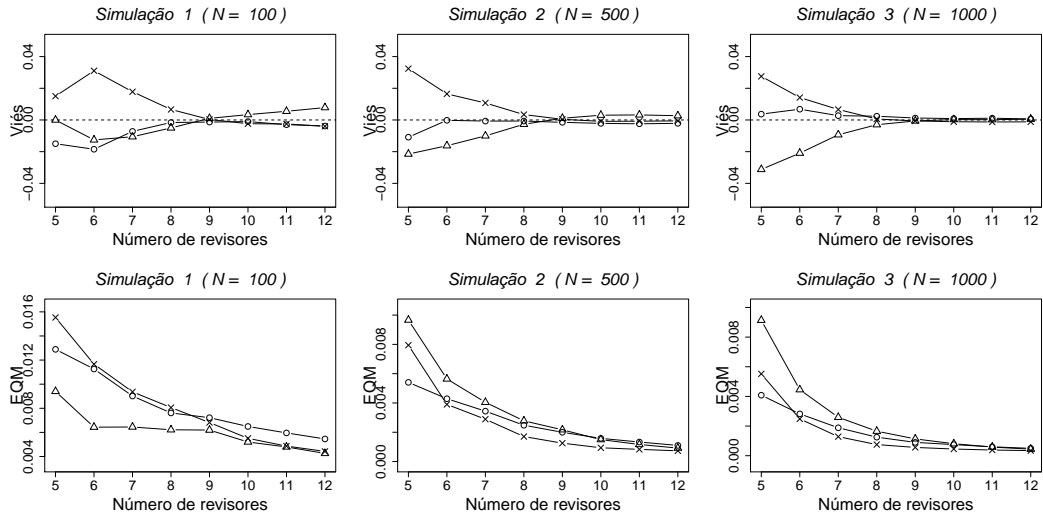


Figura 5.2: Comportamento do Viés e do EQM das EMV's de $(\omega_1, \omega_2, \omega_3)$. *Legenda:* $\times \omega_1$; $\circ \omega_2$; $\Delta \omega_3$.

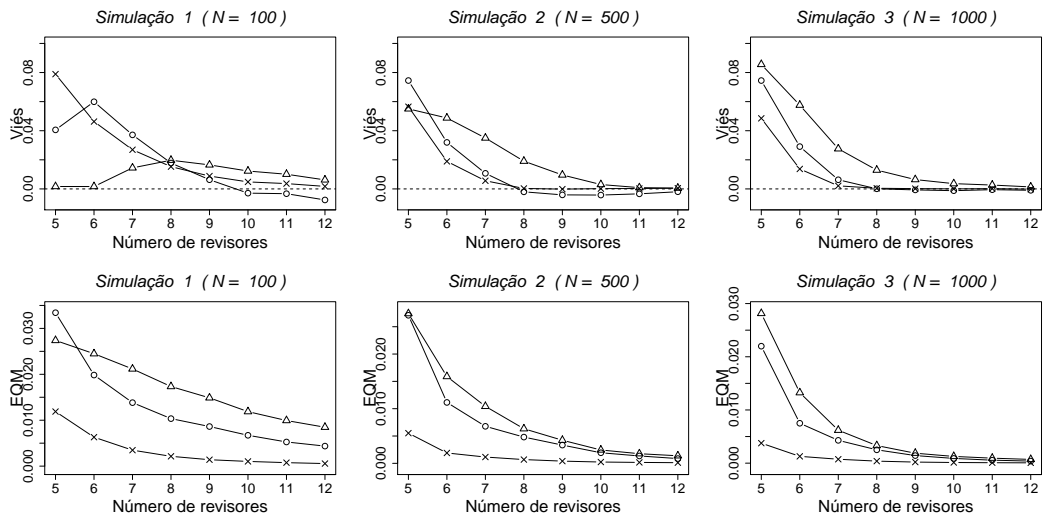


Figura 5.3: Comportamento do Viés e do EQM das EMV's de $(\theta_1, \theta_2, \theta_3)$. *Legenda:* $\times \theta_1$; $\circ \theta_2$; $\Delta \theta_3$.

A segunda etapa do estudo de simulação será dada pela comparação entre modelos utilizando dados artificiais. Para a estimação dos parâmetros, ajustamos aos dados quatro modelos candidatos. São eles:

M1: modelo probabilístico com um nível de dificuldade de detecção (modelo homogêneo);

M2: modelo probabilístico com dois níveis de dificuldade de detecção;

M3: modelo probabilístico com três níveis de dificuldade de detecção;

M4: modelo probabilístico com quatro níveis de dificuldade de detecção.

O modelo M1 é o caso em que a probabilidade de captura é θ , independentemente do erro ou revisor. Associamos a este modelo a função de verossimilhança

$$L(N, \theta | \mathcal{D}) \propto \frac{N!}{(N-n)!} \prod_{j=0}^k [\theta^j (1-\theta)^{k-j}]^{f_j},$$

para $N \geq n$ e $\theta \in (0,1)$, com $f_0 = N - n$. Para os modelos M2, M3 e M4, consideramos os casos de $l = 2, 3$ e 4 em (5.1), respectivamente.

Primeiramente, a simulação das amostras foi a partir de um modelo com dois níveis de dificuldade de captura, fixados $N = 1000$ e $k = 10$.

Simulação 1: fixamos $(\theta_1, \theta_2) = (0,1; 0,4)$ e geramos três amostras considerando os respectivos valores dos parâmetros $(\omega_1, \omega_2) = (0,2; 0,8)$ (amostra 1), $(\omega_1, \omega_2) = (0,5; 0,5)$ (amostra 2) e $(\omega_1, \omega_2) = (0,8; 0,2)$ (amostra 3).

Simulação 2: repetimos o procedimento considerando os valores de $(\theta_1, \theta_2) = (0,2; 0,5)$.

Na Tabela 5.1 se encontram as estimativas do tamanho populacional e os parâmetros perturbadores para os modelos M1, M2, M3 e M4 com respeito à **Simulação 1**. Para a primeira situação, com $(\omega_1, \omega_2) = (0,2; 0,8)$, os quatro modelos apresentaram estimativas razoáveis para N , sabendo que o seu verdadeiro valor é 1000. Note que os modelos M3 e M4 não diferenciam as probabilidades de captura em mais do que níveis, isto é, duas ou mais probabilidades de captura têm aproximadamente o mesmo valor. Nas outras duas situações para (ω_1, ω_2) , é notável a deficiência na estimativa para N em considerar o ajuste do modelo homogêneo (M1) aos dados, mostrando notavelmente a perda de precisão na sua estimação. Os modelos M2, M3 e M4 apresentam boas estimativas para o tamanho populacional.

Tabela 5.1: Estimativas de máxima verossimilhança dos parâmetros dos quatro modelos para a **Simulação 1**.

Simulação (ω_1, ω_2)	Modelo	Estimativas de máxima verossimilhança								
		\hat{N}	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
(0,2;0,8)	M1	933	-	-	-	-	0,3610	-	-	-
	M2	1016	0,1859	0,8141	-	-	0,0741	0,3904	-	-
	M3	996	0,1812	0,4278	0,3911	-	0,0926	0,3926	0,3926	-
	M4	1019	0,1871	0,3925	0,3764	0,0441	0,0719	0,3854	0,3944	0,3959
(0,5;0,5)	M1	841	-	-	-	-	0,2824	-	-	-
	M2	962	0,5369	0,4631	-	-	0,1173	0,3971	-	-
	M3	1036	0,2673	0,3416	0,3912	-	0,0501	0,1662	0,4067	-
	M4	957	0,5398	0,2856	0,1701	0,0046	0,1199	0,3986	0,3986	0,3986
(0,8;0,2)	M1	794	-	-	-	-	0,2005	-	-	-
	M2	939	0,8205	0,1795	-	-	0,1143	0,4223	-	-
	M3	948	0,8070	0,1902	0,0028	-	0,1103	0,4027	0,8184	-
	M4	948	0,8069	0,1645	0,0258	0,0028	0,1103	0,4027	0,4027	0,8183

Os dados das três amostras utilizadas são apresentadas na Tabela 5.2. Segundo os critérios de seleção de modelo AIC e BIC, temos que o modelo M2 foi o que melhor se ajustou aos dados, modelo o qual foi o gerador dos dados. Em cada amostra, obtivemos os valores esperados das frequências segundo cada modelo ajustado. Nas três amostras, os ajustes segundo os modelos M2, M3 e M4 foram bem próximos e razoavelmente boas, segundo a estatística χ^2 , que mede a distância quadrática relativa entre o valor observado e o valor esperado pelo modelo. Esta medida é calculada por

$$\chi^2 = \sum_{j=1}^k \frac{(f_j - e_j)^2}{e_j},$$

onde f_j é a frequência observada e e_j é a frequência esperada para f_j utilizando o modelo ajustado. Apesar dos modelos M3 e M4 apresentarem menores χ^2 , o aumento do número de parâmetros perturbadores não resultou em ganhos significativos na estimação, conforme os critérios AIC e BIC. Isto é, como visto na Tabela 5.2, o modelo com duas probabilidades de captura é suficiente.

Tabela 5.2: Verificação de ajuste dos quatro modelos para a **Simulação 1**.

	Valores atribuídos $\omega = (0,2;0,8)$				Valores atribuídos $\omega = (0,5;0,5)$				Valores atribuídos $\omega = (0,8;0,2)$						
	Freq. Esperada				Freq. Esperada				Freq. Esperada						
	Dados	M1	M2	M3	M4	Dados	M1	M2	M3	M4	Dados	M1	M2	M3	M4
f_1	109	59,8	107,5	105,7	107,8	218	119,8	215,7	216,9	214,3	302	212,5	300,5	301,8	301,7
f_2	128	152,0	133,4	136,8	133	166	212,2	173,1	169,8	174,2	187	239,8	188,3	185,8	185,7
f_3	208	229,1	190,1	189,3	190,2	152	222,7	138,7	142,2	138,9	93	160,4	91,7	92,7	92,7
f_4	180	226,6	207,8	205,9	207,9	106	153,4	121,4	120,9	120,9	52	70,4	55,2	57,0	57,0
f_5	172	153,7	159,2	158,7	159,2	96	72,4	89,8	88,1	89,6	44	21,2	38,8	38,4	38,4
f_6	89	72,4	84,9	85,4	84,9	49	23,8	48,6	48,5	48,7	21	4,4	22,6	20,9	20,9
f_7	30	23,4	31,1	31,5	31,1	21	5,3	18,2	18,8	18,4	7	0,6	9,4	8,4	8,4
f_8	7	5,0	7,5	7,6	7,5	3	0,8	4,5	4,8	4,6	2	0,1	2,6	2,8	2,8
f_9	0	0,6	1,1	1,1	1,1	0	0,1	0,7	0,7	0,7	2	0,0	0,4	1,1	1,1
f_{10}	0	0,0	0,1	0,1	0,1	0	0,0	0,0	0,1	0,0	0	0,0	0,0	0,4	0,4
AIC		107,61	64,70	68,56	72,51		222,75	60,10	63,79	67,91		236,26	57,89	60,06	64,06
BIC		117,26	84,01	97,53	111,13		232,15	78,89	91,98	105,49		245,39	76,15	87,45	100,58
χ^2		65,25	8,06	8,33	8,05		214,23	5,59	5,04	5,60		1495,23	7,81	2,86	2,86

Na Tabela 5.3 são exibidas as estimativas para os parâmetros segundo os quatro modelos candidados utilizando os dados gerados na **Simulação 2**. Neste estudo, todos os modelos apresentaram boas estimativas para o parâmetro N , independente dos valores de (ω_1, ω_2) da população. Do mesmo modo na **Simulação 1**, as estimativas para os parâmetros perturbadores segundo o modelo M2 foram satisfatórias, isto é, bem próximas dos verdadeiros valores dos parâmetros.

Tabela 5.3: Estimativas de máxima verossimilhança dos parâmetros dos quatro modelos para a **Simulação 2**.

Simulação (ω_1, ω_2)	Modelo	Estimativas de máxima verossimilhança								
		\hat{N}	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
(0,2;0,8)	M1	982	-	-	-	-	0,4451	-	-	-
	M2	1004	0,1873	0,8127	-	-	0,1863	0,4944	-	-
	M3	1016	0,1733	0,6688	0,1579	-	0,1492	0,4849	0,5069	-
	M4	1016	0,1714	0,2355	0,5450	0,0482	0,1481	0,4646	0,4979	0,5001
(0,5;0,5)	M1	957	-	-	-	-	0,3536	-	-	-
	M2	1001	0,5267	0,4733	-	-	0,2015	0,4907	-	-
	M3	1070	0,1649	0,4662	0,3690	-	0,0563	0,2537	0,5106	-
	M4	1036	0,1874	0,4400	0,3464	0,0262	0,0960	0,2678	0,5122	0,5122
(0,8;0,2)	M1	957	-	-	-	-	0,2696	-	-	-
	M2	1002	0,8071	0,1929	-	-	0,2006	0,4987	-	-
	M3	1038	0,2524	0,6087	0,1389	-	0,1078	0,2431	0,5290	-
	M4	1111	0,2001	0,6641	0,1196	0,0162	0,0451	0,2287	0,5256	0,5257

Na Tabela 5.4 são apresentadas os dados simulados e as frequências estimadas pelos quatro modelos ajustados. Observa-se que o menor valor de AIC foi para o modelo M2, mostrando-se o melhor modelo. O mesmo resultado é obtido considerando o critério BIC. A partir destas duas simulações, é possível dizer que os critérios selecionaram bem o verdadeiro modelo de duas misturas gerador dos dados.

Tabela 5.4: Verificação de ajuste dos quatro modelos para a **Simulação 2**.

	Valores atribuídos $\omega = (0,2;0,8)$				Valores atribuídos $\omega = (0,5;0,5)$				Valores atribuídos $\omega = (0,8;0,2)$						
	Freq. Esperada				Freq. Esperada				Freq. Esperada						
	Dados	M1	M2	M3	M4	Dados	M1	M2	M3	M4	Dados	M1	M2	M3	M4
f_1	71	21,8	63,5	71,2	71,0	153	66,7	145,6	153,1	152,1	231	152,5	218,2	227,3	230,2
f_2	92	78,7	94,8	90,5	90,5	171	164,1	182,4	170,1	172,1	233	253,5	252,8	240,4	236,4
f_3	123	168,3	134,4	129,9	130,1	167	239,4	166,8	170,9	169,6	192	249,6	186,3	185,9	188,0
f_4	206	236,3	184,8	186,3	186,4	164	229,2	148,0	152,7	152,9	118	161,3	111,6	116,8	117,0
f_5	182	227,4	204,5	207,4	207,1	112	150,4	130,7	127,5	128,2	62	71,5	69,2	69,0	68,4
f_6	173	152,0	164,2	164,6	164,3	101	68,6	96,5	92,9	93,0	48	22,0	44,0	41,7	41,7
f_7	96	69,7	91,5	90,1	90,1	53	21,4	51,9	51,9	51,9	21	4,6	23,1	22,6	22,8
f_8	30	21,0	33,5	32,5	32,6	21	4,4	18,6	19,9	19,9	8	0,6	8,4	9,0	9,1
f_9	7	3,7	7,3	6,9	7,0	3	0,5	4,0	4,6	4,6	3	0,1	1,9	2,2	2,2
f_{10}	0	0,3	0,7	0,7	0,7	0	0,0	0,4	0,5	0,5	0	0,0	0,2	0,2	0,2
AIC		170,50	69,40	72,41	76,21		275,92	66,73	69,32	73,35		215,32	62,51	64,77	68,73
BIC		1180,27	88,95	101,74	115,31		285,62	86,13	98,43	112,16		224,96	81,79	93,69	107,29
χ^2		158,33	8,61	7,25	7,24		298,20	6,68	4,63	4,72		405,42	5,07	2,92	2,49

Simulação 3: Consideramos agora a simulação dos dados a partir de um modelo com três níveis de dificuldade de captura. Fixamos os valores de $(\theta_1, \theta_2, \theta_3) = (0,1; 0,5; 0,9)$ e simulamos três amostras com $(\omega_1, \omega_2, \omega_3) = (0,7; 0,1; 0,2)$ (amostra 1), $(\omega_1, \omega_2, \omega_3) = (0,5; 0,3; 0,2)$ (amostra 2) e $(\omega_1, \omega_2, \omega_3) = (0,2; 0,3; 0,5)$ (amostra 3).

Na Tabela 5.5 nota-se que os modelos M3 e M4 apresentam melhores estimativas para o tamanho populacional nas três amostras. O modelo M3 obtém estimativas para os parâmetros perturbadores próximas aos verdadeiros valores geradores dos dados. Uma observação a ser feita nesta tabela sobre o aumento em um componente sobre o número verdadeiro de misturas é quando utilizamos a amostra 1. O modelo M4 divide o valor verdadeiro de $\omega_3 = 0,2$ para $\hat{\omega}_3 = 0,0958$ e $\hat{\omega}_4 = 0,1166$ e estima $\hat{\theta}_3 = 0,8758$ e $\hat{\theta}_4 = 0,9363$ os quais, em média, é aproximadamente o valor verdadeiro de $\theta_3 = 0,9$.

Tabela 5.5: Estimativas de máxima verossimilhança dos parâmetros dos quatro modelos para a **Simulação 3**.

Simulação $(\omega_1, \omega_2, \omega_3)$	Modelo	Estimativas de máxima verossimilhança								
		\hat{N}	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
(0,7;0,1;0,2)	M1	758	-	-	-	-	0,3961	-	-	-
	M2	895	0,7386	0,2614	-	-	0,1429	0,8798	-	-
	M3	979	0,7064	0,0845	0,2091	-	0,1060	0,4841	0,9133	-
	M4	982	0,7040	0,0836	0,0958	0,1166	0,1048	0,4658	0,8758	0,9363
(0,5;0,3;0,2)	M1	811	-	-	-	-	0,4750	-	-	-
	M2	840	0,6427	0,3573	-	-	0,2498	0,8345-	-	-
	M3	960	0,4650	0,3049	0,2301	-	0,1032	0,4797	0,9002	-
	M4	960	0,3525	0,1125	0,3050	0,2301	0,1031	0,1032	0,4797	0,9002
(0,2;0,3;0,5)	M1	932	-	-	-	-	0,6696	-	-	-
	M2	937	0,4083	0,5917	-	-	0,3457	0,8871	-	-
	M3	1055	0,2250	0,2902	0,4847	-	0,0637	0,4775	0,9026	-
	M4	1029	0,2088	0,0929	0,2020	0,4964	0,0763	0,4798	0,4840	0,9031

A Tabela 5.6 apresenta os valores gerados das amostras na **Simulação 3** e as frequências preditas segundo os quatro modelos ajustados. A partir dos critérios AIC e BIC, temos que o modelo gerador dos dados (M3) foi o que melhor se ajustou as três amostras simuladas.

Tabela 5.6: Verificação de ajuste dos quatro modelos para a **Simulação 3**.

	Valores atribuídos $\omega = (0,7;0,1;0,2)$				Valores atribuídos $\omega = (0,5;0,3;0,2)$				Valores atribuídos $\omega = (0,2;0,3;0,5)$						
	Freq. Esperada				Freq. Esperada				Freq. Esperada						
	Dados	M1	M2	M3	M4	Dados	M1	M2	M3	M4	Dados	M1	M2	M3	M4
f_1	270	32,0	235,8	268,4	268,9	174	11,7	101,5	176,8	176,8	85	0,3	29,1	87,9	84,1
f_2	146	94,6	176,9	147,1	146,2	115	47,5	152,1	105,8	105,7	46	2,7	69,1	43,0	46,1
f_3	56	165,5	78,7	56,1	56,4	56	114,6	135,1	67,5	67,5	46	14,4	97,4	47,2	47,2
f_4	28	190,1	23,0	27,4	27,9	72	181,5	79,3	70,1	70,1	65	51,2	90,2	68,6	67,3
f_5	21	149,6	5,4	21,8	21,4	80	197,1	35,2	72,5	72,5	78	124,4	58,5	75,3	75,0
f_6	18	81,8	5,4	17,4	16,9	59	148,7	24,7	57,4	57,4	70	210,2	34,4	62,0	62,9
f_7	16	30,7	19,9	17,0	17,9	33	76,9	47,7	41,5	41,5	43	243,5	49,0	57,3	58,1
f_8	40	7,5	54,6	36,5	37,3	57	26,1	87,2	52,7	52,7	125	185,1	123,5	106,3	106,2
f_9	74	1,1	88,8	79,1	76,2	88	5,2	97,5	87,6	87,6	185	83,4	213,1	200,1	200,0
f_{10}	85	0,1	65,0	82,7	84,3	76	0,5	49,2	77,4	77,3	189	16,9	167,3	183,7	184,5
AIC		2904,14	139,96	68,90	72,62		2036,73	285,15	77,56	81,37		2155,29	227,67	82,34	85,50
BIC		2913,39	158,46	96,65	109,63		2046,13	303,94	105,75	118,95		2164,97	247,02	111,36	124,20
χ^2		107042,77	106,20	0,86	0,57		15948,42	242,89	5,80	5,80		27475,89	200,27	9,79	9,52

5.2 Modelo bayesiano

Nesta seção, discutimos o modelo e o processo inferencial para os parâmetros N e $\boldsymbol{\psi}$ através da abordagem Bayesiana.

De (5.1), a distribuição *a posteriori* conjunta de N e $\boldsymbol{\psi}$ é dada por

$$\begin{aligned} \pi(N, \boldsymbol{\psi} | \mathcal{D}) &\propto L(N, \boldsymbol{\psi} | \mathcal{D}) \cdot \pi(N, \boldsymbol{\psi}) \\ &\propto \frac{N!}{(N-n)!} \prod_{j=0}^k \left\{ \sum_{s=1}^l \omega_s \theta_s^j (1 - \theta_s)^{k-j} \right\}^{f_j} \cdot \pi(N, \boldsymbol{\psi}) \end{aligned} \quad (5.8)$$

onde $f_0 = N - n$ e $\pi(N, \boldsymbol{\psi})$ é a distribuição *a priori* conjunta de N e $\boldsymbol{\psi}$.

Supomos independência *a priori*, isto é,

$$\pi(N, \boldsymbol{\psi}) = \pi_1(N) \cdot \pi_2(\boldsymbol{\psi}) \quad (5.9)$$

onde $\pi_1(N)$ e $\pi_2(\boldsymbol{\psi})$ são as distribuições *a priori* marginais de N e $\boldsymbol{\psi}$, respectivamente, para $N \geq 1$ e $\boldsymbol{\psi} \in \boldsymbol{\Psi}$.

Sejam $\pi_2(\boldsymbol{\psi})$ uma f.d.p. própria e $\pi_1(N) \propto N^{-r}$, $N \geq 1$, $r \geq 0$. O Teorema 5.2 a seguir, anunciado como uma generalização do Teorema 4.1, assegura a existência da distribuição *a posteriori* conjunta de N e $\boldsymbol{\psi}$ para o caso de $0 \leq r \leq 1$.

Teorema 5.2. *Se a distribuição a priori $\pi(N, \boldsymbol{\psi}) \propto \pi_1(N)\pi_2(\boldsymbol{\psi})$ com $\pi_1(N) = \frac{1}{N^r} \mathcal{I}_{\{N \geq 1\}}$, $r \geq 0$, e $\pi_2(\boldsymbol{\psi})$ uma f.d.p. própria com suporte*

$$\boldsymbol{\Psi}_{(c)} = \left\{ (\omega_s, \theta_s) \in (0,1)^2; 0 \leq s \leq l : 0 < c < \theta_1 < \dots < \theta_l \leq 1 \text{ e } \omega_l = 1 - \omega_1 - \dots - \omega_{l-1} \right\},$$

para c constante, então a posteriori $\pi(N, \boldsymbol{\psi} | \mathcal{D})$ é própria.

Demonstração. A prova é análoga a do Teorema 4.1. □

Desta forma, as distribuições condicionais *a posteriori* necessárias para o algoritmo *Gibbs sampling* são dadas por

$$\pi(N | \mathcal{D}, \boldsymbol{\psi}) \propto \frac{N!}{(N-n)!} \left\{ \sum_{s=1}^l \omega_s (1 - \theta_s)^k \right\}^N \cdot \pi(N), \quad (5.10)$$

$N \geq n$ e para $f_0 = N - n$,

$$\pi(\boldsymbol{\psi}|\mathcal{D}, N) \propto \prod_{j=0}^k \left\{ \sum_{s=1}^l \omega_s \theta_s^j (1 - \theta_s)^{k-j} \right\}^{f_j} \cdot \pi(\boldsymbol{\psi}), \quad (5.11)$$

$\boldsymbol{\psi} \in \boldsymbol{\Psi}_{(c)}$.

Seja $h'(0; \boldsymbol{\psi}) = \sum_{s=1}^l \omega_s (1 - \theta_s)^k$. Note que, se $r = 0$, temos

$$\pi(N|\boldsymbol{\theta}, \mathcal{D}) \propto \frac{N!}{(N - n)!} h'(0; \boldsymbol{\psi})^N$$

e, pelo Teorema 4.2, N , dados $\boldsymbol{\phi}$ e \mathcal{D} , tem distribuição igual de uma variável aleatória $X + n$, onde X tem distribuição Binomial Negativa, isto é,

$$X \sim \text{BinNeg}(n + 1, 1 - h'(0; \boldsymbol{\phi})).$$

Da mesma forma, se $r = 1$, temos

$$\pi(N|\boldsymbol{\psi}, \mathcal{D}) \propto \frac{(N - 1)!}{(N - n)!} h'(0; \boldsymbol{\psi})^N,$$

e, conseqüentemente, do Teorema 4.3, temos a distribuição condicional *a posteriori* para N , dados $\boldsymbol{\phi}$ e \mathcal{D} , corresponde à uma distribuição de uma variável aleatória $Y + n$, onde Y tem distribuição Binomial Negativa,

$$Y \sim \text{BinNeg}(n, 1 - h'(0; \boldsymbol{\phi})).$$

5.2.1 Variáveis latentes e *posteriori* condicional para os parâmetros perturbadores

Para determinarmos as distribuições condicionais *a posteriori* para $\boldsymbol{\omega} = (\omega_1, \dots, \omega_l)$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_l)$, condicionamos os parâmetros no conjunto de variáveis completas, previamente definidos por $\boldsymbol{v} = \{v_{j,s}; 0 \leq j \leq k, 1 \leq s \leq l\}$, dado N . O valor de \boldsymbol{v} será dado, por sua vez, através de simulação utilizando a distribuição definida na relação (5.3).

Assumimos independência *a priori* entre $\boldsymbol{\omega}$ e $\boldsymbol{\theta}$, e consideramos

$$\boldsymbol{\omega} = (\omega_1, \dots, \omega_l) \sim \text{Dirichlet}(\boldsymbol{a} = (a_1, \dots, a_l)), \quad (5.12)$$

para $a_1, \dots, a_l > 0$ com f.d.p. $\pi(\boldsymbol{\omega}) \propto \omega_1^{a_1-1} \dots \omega_l^{a_l-1}$, e

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_l) \sim \text{Uniforme}((0,1)^l \cap \{0 < c < \theta_1 < \dots < \theta_l \leq 1\}), \quad (5.13)$$

com $\pi(\boldsymbol{\theta}) \propto 1$. Desta forma, segue que a distribuição condicional *a posteriori* de $\boldsymbol{\psi}$, dado \mathbf{v} , é

$$\pi(\boldsymbol{\psi}|\mathbf{v}) \propto \prod_{s=1}^l \omega_s^{\sum_{j=0}^k v_{j,s} + a_s - 1} \times \prod_{s=1}^l \theta_s^{\sum_{j=1}^k j v_{j,s}} (1 - \theta_s)^{\sum_{j=0}^k (k-j)v_{j,s}}, \quad (5.14)$$

para $\boldsymbol{\psi} \in \Psi_{(c)}$.

Logo, identificamos de (5.14) que as distribuições condicionais *a posteriori* a serem utilizadas no algoritmo *Gibbs sampling* são dadas por

$$\omega_1, \dots, \omega_l | \mathbf{v} \sim \text{Dirichlet}(\mathbf{a}^* = (a_1^*, \dots, a_l^*)),$$

onde $a_s^* = \sum_{j=1}^k v_{j,s} + a_s$, $1 \leq s \leq l$ e

$$\begin{aligned} \theta_1 | \mathbf{v} &\sim \text{Beta}\mathcal{T}_{(c, \theta_2)} \left(\sum_{j=0}^k j v_{j,1} + 1, \sum_{j=0}^k (k-j)v_{j,1} + 1 \right) \\ \theta_s | \theta_{s-1}, \theta_{s+1}, \mathbf{v} &\sim \text{Beta}\mathcal{T}_{(\theta_{s-1}, \theta_{s+1})} \left(\sum_{j=0}^k j v_{j,s} + 1, \sum_{j=0}^k (k-j)v_{j,s} + 1 \right), \quad 2 \leq s \leq (l-1) \\ \theta_l | \theta_{l-1}, \mathbf{v} &\sim \text{Beta}\mathcal{T}_{(\theta_{l-1}, 1]} \left(\sum_{j=0}^k j v_{j,l} + 1, \sum_{j=0}^k (k-j)v_{j,l} + 1 \right). \end{aligned}$$

em que $\text{Beta}\mathcal{T}_{(a,b)}(\cdot, \cdot)$ denota a distribuição $\text{Beta}(\cdot, \cdot)$ truncada no intervalo (a,b) .

5.2.2 Implementação do Algoritmo

- (1) Inicialize todos os parâmetros, $N^{(1)} \geq n$ e $\boldsymbol{\psi}^{(1)} \in \Psi_{(c)}$.
- (2) Para a m -ésima iteração, $m = 2, \dots, M$,
 - (i) gere $N^{(m)} \sim \pi(N|\mathcal{D}, \boldsymbol{\psi}^{(m-1)})$ e tome $f_0 = N^{(m)} - n$,
 - (ii) simule $v_{j,1}, \dots, v_{j,l} \sim \text{Multinomial}(f_j, \mathbf{q}(j; \boldsymbol{\psi}^{(m-1)}))$; $0 \leq j \leq k$,

- (iii) gere $\omega^{(m)} \sim \text{Dirichlet}((a_1^*, \dots, a_l^*))$, com $a_s^* = \sum_{j=1}^k v_{j,s} + a_s$, $1 \leq s \leq l$,
- (iv) gere $\theta_1^{(m)} \sim \text{Beta}\mathcal{T}_{(c, \theta_2^{(m-1)})} \left(\sum_{j=0}^k j v_{j,1} + 1, \sum_{j=0}^k (k-j) v_{j,1} + 1 \right)$,
- $\theta_s \sim \text{Beta}\mathcal{T}_{(\theta_{s-1}^{(m)}, \theta_{s+1}^{(m-1)})} \left(\sum_{j=0}^k j v_{j,s} + 1, \sum_{j=0}^k (k-j) v_{j,s} + 1 \right)$, $2 \leq s \leq (l-1)$,
- $\theta_l^{(m)} \sim \text{Beta}\mathcal{T}_{(\theta_{l-1}^{(m)}, 1)} \left(\sum_{j=0}^k j v_{j,l} + 1, \sum_{j=0}^k (k-j) v_{j,l} + 1 \right)$.

5.2.3 Resultados numéricos

Designamos esta seção para comparação entre modelos utilizando dados simulados sob os métodos bayesianas visto neste capítulo.

Para ilustração, consideramos o critério de seleção de modelos DIC (*deviance information criterion*) (Spiegelhalter et al., 1998), calculado por

$$DIC = \mathbb{E}[D(\theta)|x] + \{\mathbb{E}[D(\theta)|x] - D(\mathbb{E}[\theta|x])\}$$

onde $D(\theta) = -2 \log L(\theta|x)$ e $\log L(\theta|x)$ é a função log-verossimilhança do modelo probabilístico indexado por um parâmetro θ , $\theta \in \Theta \subseteq \mathbb{R}^q$, $q \geq 1$.

Consideramos quatro modelos candidatos:

M1: modelo probabilístico com um nível de dificuldade de detecção (modelo homogêneo),

M2: modelo probabilístico com dois níveis de dificuldade de detecção,

M3: modelo probabilístico com três níveis de dificuldade de detecção,

M4: modelo probabilístico com quatro níveis de dificuldade de detecção.

Para o modelo M1, temos a posteriori para os parâmetros

$$\pi(N, \theta | \mathcal{D}) \propto \frac{N!}{(N-n)!} \prod_{j=0}^k [\theta^j (1-\theta)^{k-j}]^{f_j} \cdot \pi(N, \theta),$$

para $N \geq n$ e $\theta \in (0,1)$, com $f_0 = N - n$, onde θ é a probabilidade de um erro i ser detectado pelo revisor j , $1 \leq i \leq N$, $1 \leq j \leq k$ e $\pi(N, \theta)$ é a distribuição a priori conjunta para N e θ . Suponhamos independência a priori entre N e θ e as distribuições condicionais a *posteriori* são

dadas por

$$\begin{aligned}\pi(N|\theta, \mathcal{D}) &\propto \frac{N!}{(N-n)!} [(1-\theta)^k]^{N-n} \cdot \pi(N), \\ \pi(\theta|N, \mathcal{D}) &\propto \prod_{j=0}^k [\theta^j (1-\theta)^{k-j}]^{f_j} \cdot \pi(\theta),\end{aligned}$$

para $N \geq n$ e $\theta \sim \mathcal{U}(c, 1)$, $0 < c < 1$.

Para os modelos M2, M3 e M4, consideramos os casos de $l = 2, 3$ e 4 em (5.8), respectivamente, e utilizamos o algoritmo descrito na Seção 5.2.2 para obter as estimativas bayesianas para os parâmetros.

Assumimos para os quatro modelos a distribuição *a priori* não informativa uniforme discreta imprópria para N , $\pi(N) \propto 1$, $N \geq 1$ e $c = 0,001$. Os valores dos hiperparâmetros para $\omega = (\omega_1, \dots, \omega_l)$ foram $a_1 = \dots = a_l = 1$, $l = 2, 3$ e 4 .

Simulação 1: Primeiramente, a simulação foi feita a partir de um modelo com dois níveis de dificuldade de captura. Fixamos $N = 1000$, $k = 10$ e $(\theta_1, \theta_2) = (0,1; 0,5)$ e geramos três amostras considerando os respectivos valores dos parâmetros $(\omega_1, \omega_2) = (0,2; 0,8)$ (amostra 1), $(\omega_1, \omega_2) = (0,5; 0,5)$ (amostra 2) e $(\omega_1, \omega_2) = (0,8; 0,2)$ (amostra 3).

Pela Tabela 5.7, temos os resumos *a posteriori* para o parâmetro N segundo os quatro modelos adotados. Notamos primeiramente que as estimativas médias para o parâmetro N segundo os quatro modelos foram próximas ao verdadeiro valor utilizado na simulação ($N = 1000$), com intervalos de credibilidade contendo seu verdadeiro valor. Observe que, em quase todas as simulações, modelos com mais densidades de mistura obtêm estimativas médias *a posteriori* maiores para o tamanho populacional. Na última coluna da tabela se encontram as estimativas do critério de Gelman and Rubin (1992), em que temos indícios de convergência das cadeias.

Tabela 5.7: Resumos a *posteriori* para o tamanho populacional para os 4 modelos para a **Simulação 1**.

Simulação ($\omega_1; \omega_2$)	Modelo	Estimativas Bayesianas					Int. Cred (95%)	Crit. GR Est. (Q.97,5%)
		Média	Moda	Q_1	Q_2	Q_3		
(0,2;0,8)	M1	954,27	991	953	954	955	[952 ; 958]	1,0001 (1,0007)
	M2	1003,57	954	990	1000	1013	[975 ; 1051]	1,0056 (1,0151)
	M3	1030,66	992	996	1010	1035	[977 ; 1234]	1,0487 (1,0608)
	M4	1047,37	1001	1000	1018	1051	[979 ; 1317]	1,0410 (1,0630)
(0,5;0,5)	M1	850,56	850	848	850	853	[845 ; 857]	0,9998 (0,9999)
	M2	978,16	968	959	976	995	[932 ; 1037]	1,0003 (1,0003)
	M3	1034,62	982	970	993	1026	[939 ; 1415]	1,0194 (1,0461)
	M4	1050,94	996	984	1012	1063	[946 ; 1406]	1,0043 (1,0116)
(0,8;0,2)	M1	747,02	748	741	747	753	[730 ; 766]	1,0005 (1,0027)
	M2	948,49	929	924	947	971	[886 ; 1022]	1,0001 (1,0014)
	M3	1091,17	943	937	965	1007	[895 ; 1565]	1,0184 (1,0268)
	M4	1005,60	960	942	973	1017	[895 ; 1367]	1,0005 (1,0022)

Na Tabela 5.8 seguem as médias a *posteriori* dos parâmetros perturbadores para cada modelo. Note que o modelo M2 obteve estimativas bem próximas aos verdadeiros valores dos parâmetros perturbadores.

Seja DIC_x o valor do critério DIC correspondente ao modelo x , $x \in \{M1, M2, M3, M4\}$. Para a primeira simulação, temos $DIC_{M1} = 205,17$, $DIC_{M2} = 65,15$, $DIC_{M3} = 52,46$ e $DIC_{M4} = 44,37$. Assim, pelo critério DIC temos que o melhor modelo foi o M_4 , o qual apresentou menor valor para o critério. O mesmo modelo foi escolhido por este critério na segunda simulação, onde tivemos $DIC_{M1} = 499,71$, $DIC_{M2} = 66,84$, $DIC_{M3} = 56,77$ e $DIC_{M4} = 53,85$. Já na terceira simulação, o melhor modelo segundo o critério foi o M_3 com $DIC_{M3} = -7,69$. Os outros modelos assumiram $DIC_{M1} = 490,04$, $DIC_{M2} = 59,76$ e $DIC_{M4} = 30,80$. Vale ressaltar que em nenhuma simulação o critério DIC escolheu o modelo gerador dos dados (M2) como o melhor modelo.

Na Tabela 5.12 que se encontra no final do capítulo são apresentados os valores do critério de convergência de Gelman and Rubin (1992) para os parâmetros perturbadores na **Simulação 1**.

Tabela 5.8: Médias a *posteriori* dos parâmetros perturbadores para os quatro modelos para a **Simulação 1**.

Simulação		Médias a <i>posteriori</i>							
$(\omega_1; \omega_2)$	Modelo	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
(0,2;0,8)	M1	-	-	-	-	0,4530	-	-	-
	M2	0,1868	0,8132	-	-	0,1265	0,5007	-	-
	M3	0,1661	0,2965	0,5375	-	0,0973	0,3717	0,5273	-
	M4	0,1481	0,1754	0,3014	0,3750	0,0803	0,2886	0,4477	0,5531
(0,5;0,5)	M1	-	-	-	-	0,3621	-	-	-
	M2	0,4868	0,5132	-	-	0,1183	0,5017	-	-
	M3	0,4190	0,2726	0,3084	-	0,0956	0,3518	0,5481	-
	M4	0,3106	0,2038	0,1990	0,2865	0,0738	0,2041	0,3921	0,5531
(0,8;0,2)	M1	-	-	-	-	0,2273	-	-	-
	M2	0,8157	0,1843	-	-	0,1045	0,5120	-	-
	M3	0,4754	0,3663	0,1583	-	0,0754	0,1822	0,5204	-
	M4	0,3584	0,3602	0,1493	0,1322	0,0684	0,1308	0,3048	0,5344

Simulação 2: Definimos a simulação das amostras a partir de um modelo com três níveis de dificuldade de captura, fixados $N = 1000$, $k = 10$ e $(\theta_1, \theta_2, \theta_3) = (0,1; 0,5; 0,9)$ e gerando três amostras considerando os respectivos valores dos parâmetros $(\omega_1, \omega_2, \omega_3) = (0,7; 0,1; 0,2)$ (amostra 1), $(\omega_1, \omega_2, \omega_3) = (0,5; 0,3; 0,2)$ (amostra 2) e $(\omega_1, \omega_2, \omega_3) = (0,2; 0,3; 0,5)$ (amostra 3).

Pela Tabela 5.9 notamos que os intervalos de credibilidade para N segundo os modelos M1 e M2 não contiveram o verdadeiro valor do parâmetro. As inferências sobre N utilizando os modelos M3 e M4 são parecidas, com estimativas médias e intervalos de credibilidade para N bem próximos nas duas últimas amostras simuladas.

Na Tabela 5.10 são apresentadas as médias a *posteriori* dos parâmetros perturbadores de cada modelo. Notamos que para o modelo M3 (modelo gerador dos dados) as estimativas para os parâmetros foram satisfatórias, isto é, bem próximas dos seus verdadeiros valores.

Tabela 5.9: Resumos a *posteriori* para o tamanho populacional para os quatro modelos para a **Simulação 2**.

Simulação		Estimativas Bayesianas						Crit. GR
$(\omega_1; \omega_2)$	Modelo	Média	Moda	Q_1	Q_2	Q_3	Int. Cred (95%)	Est. (Q.97,5%)
(0,7; 0,1; 0,2)	M1	758,98	758	757	759	760	[755 ; 764]	0,9999 (1,0002)
	M2	896,82	884	883	896	909	[862 ; 938]	1,0004 (1,0011)
	M3	984,40	975	959	982	1006	[922 ; 1062]	0,9998 (0,9998)
	M4	1004,06	993	963	990	1020	[918 ; 1156]	1,0333 (1,0805)
(0,5; 0,3; 0,2)	M1	811,29	811	810	811	812	[810 ; 814]	1,0015 (1,0031)
	M2	840,97	841	836	841	846	[828 ; 855]	1,0001 (1,0013)
	M3	964,89	954	943	962	984	[912 ; 1034]	1,0003 (1,0011)
	M4	964,57	952	943	962	983	[910 ; 1035]	1,0027 (1,0140)
(0,2; 0,3; 0,5)	M1	932,01	932	932	932	932	[932 ; 932]	1,0121 (1,0132)
	M2	937,61	937	936	937	939	[933 ; 943]	1,0008 (1,0046)
	M3	1044,11	1011	1012	1035	1065	[983 ; 1162]	1,0075 (1,0077)
	M4	1045,45	1021	1014	1037	1067	[982 ; 1161]	1,0033 (1,0100)

Tabela 5.10: Médias a *posteriori* dos parâmetros perturbadores para os quatro modelos para a **Simulação 2**.

Simulação		Médias a <i>posteriori</i>							
$(\omega_1; \omega_2)$	Modelo	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\omega}_4$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
(0,7; 0,1; 0,2)	M1	-	-	-	-	0,3957	-	-	-
	M2	0,7382	0,2618	-	-	0,1426	0,8791	-	-
	M3	0,7038	0,0878	0,2083	-	0,1050	0,4803	0,9126	-
	M4	0,6358	0,1327	0,0899	0,1416	0,0972	0,3754	0,7302	0,9306
(0,5; 0,3; 0,2)	M1	-	-	-	-	0,4749	-	-	-
	M2	0,6417	0,3583	-	-	0,2493	0,8338	-	-
	M3	0,4662	0,3042	0,2296	-	0,1026	0,4789	0,8996	-
	M4	0,4644	0,2841	0,1326	0,1189	0,1027	0,4717	0,8126	0,9224
(0,2; 0,3; 0,5)	M1	-	-	-	-	0,6695	-	-	-
	M2	0,4085	0,5915	-	-	0,3450	0,8866	-	-
	M3	0,2185	0,2921	0,4894	-	0,0743	0,4814	0,9028	-
	M4	0,2198	0,2840	0,3185	0,1777	0,0745	0,4786	0,8678	0,9342

Para a **simulação 2**, os valores do critério DIC para os modelos foram $DIC_{M1} = 2904,05$, $DIC_{M2} = 139,93$, $DIC_{M3} = 68,90$ e $DIC_{M4} = 44,17$ para a primeira amostra, $DIC_{M1} = 2036,24$, $DIC_{M2} = 285,06$, $DIC_{M3} = 77,51$ e $DIC_{M4} = 74,02$ para a segunda amostra e $DIC_{M1} = 2153,42$, $DIC_{M2} = 227,55$, $DIC_{M3} = 81,06$ e $DIC_{M4} = 81,87$ para a terceira amostra. Nas duas primeiras amostras temos que o modelo M4 se ajustou melhor aos dados. Já na terceira, o modelo gerador dos dados se mostrou melhor (M3), mas o valor do critério foi bem próximo ao obtido pelo modelo M4. Na Tabela 5.12 se encontram os valores do critério de [Gelman and Rubin \(1992\)](#) para os parâmetros perturbadores na **Simulação 2**.

Tabela 5.11: Critérios de Gelman e Rubin para os parâmetros perturbadores para a Simulação 1.

$(\omega_1; \omega_2)$	Modelo	ω_1	ω_2	ω_3	ω_4	θ_1	θ_2	θ_3	θ_4
(0,2;0,8)	M1	-	-	-	-	0,9999(1,0002)	-	-	-
	M2	1,0005(1,0025)	1,0005(1,0025)	-	-	1,0025(1,0097)	1,0014(1,0074)	-	-
	M3	1,0413(1,1005)	1,0097(1,0428)	1,0127(1,0599)	-	1,0208(1,0872)	1,0472(1,0751)	1,0052(1,0242)	-
	M4	1,0520(1,0890)	1,0159(1,0554)	1,0146(1,0657)	1,0004(1,0020)	1,0023(1,0074)	1,0087(1,0296)	1,0054(1,0252)	1,0015(1,0049)
(0,5;0,5)	M1	-	-	-	-	1,0006(1,0034)	-	-	-
	M2	1,0002(1,0017)	1,0002(1,0017)	-	-	1,0006(1,0009)	1,0003(1,0020)	-	-
	M3	1,0361(1,1049)	1,0104(1,0456)	1,0302(1,1001)	-	1,0529(1,0602)	1,0462(1,0602)	1,0334(1,0400)	-
	M4	1,0013(1,0014)	1,0178(1,0684)	1,0118(1,0565)	1,0370(1,0613)	1,0026(1,0132)	1,0078(1,0181)	1,0173(1,0786)	1,0197(1,0716)
(0,8;0,2)	M1	-	-	-	-	0,9999(1,0003)	-	-	-
	M2	1,0001(1,0005)	1,0001(1,0005)	-	-	1,0014(1,0076)	1,0001(1,0008)	-	-
	M3	1,0032(1,0442)	1,0303(1,07346)	1,0240(1,0634)	-	1,0102(1,0458)	1,0159(1,0617)	1,0352(1,0515)	-
	M4	1,0368(1,1005)	1,0360(1,0991)	1,0114(1,0166)	1,0128(1,0400)	1,0043(1,0199)	1,0173(1,0575)	1,0112(1,0494)	1,0223(1,0519)

Tabela 5.12: Critérios de Gelman e Rubin para os parâmetros perturbadores para a Simulação 2.

$(\omega_1; \omega_2, \omega_3)$	Modelo	ω_1	ω_2	ω_3	ω_4	θ_1	θ_2	θ_3	θ_4
(0,7; 0,1; 0,2)	M1	-	-	-	-	1,0000(1,0000)	-	-	-
	M2	0,9999(1,0002)	0,9999(1,0002)	-	-	1,0000(1,0006)	0,9999(1,0000)	-	-
	M3	1,0011(1,0033)	1,0001(1,0008)	1,0004(1,0014)	-	1,0003(1,0003)	0,9999(0,9999)	1,0010(1,0047)	-
	M4	1,0095(1,0373)	1,0111(1,0455)	1,0015(1,0085)	1,0006(1,0035)	1,0281(1,0614)	1,0037(1,0143)	1,0009(1,0024)	0,9998(0,9999)
(0,5; 0,3; 0,2)	M1	-	-	-	-	1,0002(1,0017)	-	-	-
	M2	0,9998(0,9998)	0,9998(0,9998)	-	-	1,0004(1,0028)	1,0005(1,0007)	-	-
	M3	1,0002(1,0006)	1,0009(1,0051)	1,0005(1,0029)	-	1,0000(1,0005)	0,9998(0,9998)	1,0013(1,0018)	-
	M4	1,0003(1,0007)	1,0856(1,0949)	1,0047(1,0056)	1,0171(1,0603)	1,0091(1,0253)	1,0346(1,0820)	1,0550(1,0753)	1,0035(1,0147)
(0,2; 0,3; 0,5)	M1	-	-	-	-	1,0001(1,0004)	-	-	-
	M2	1,0002(1,0012)	1,0002(1,0012)	-	-	1,0015(1,0072)	1,0002(1,0011)	-	-
	M3	1,0026(1,0026)	1,0003(1,0014)	1,0017(1,0022)	-	1,0026(1,0045)	0,9999(1,0004)	1,0013(1,0014)	-
	M4	1,0054(1,0174)	1,0429(1,0800)	1,0229(1,0991)	1,0398(1,0591)	1,0298(1,0299)	1,0040(1,0205)	1,0544(1,0564)	1,0173(1,0808)

Capítulo 6

Considerações finais

O presente trabalho apresentou o modelo de captura-recaptura em que a heterogeneidade entre os indivíduos foi modelada a partir de uma mistura de distribuições binomiais. Tendo visto a complexidade em maximizar a função de verossimilhança, optamos pelo auxílio do algoritmo EM na estimação (via máxima verossimilhança) dos parâmetros perturbadores das componentes da mistura e, com um estimador para o tamanho populacional, obtivemos um algoritmo de fácil implementação. Para a inferência clássica dos parâmetros, podemos dizer que tivemos boas estimativas, tanto no estudo de simulação quanto na aplicação em dados reais. Como pesquisa futura para esta abordagem, buscaremos metodologias para obter intervalos de confiança para os parâmetros populacionais, como o método de reamostragem *bootstrap* (Efron, 1979; Efron and Efron, 1982), por exemplo.

Na abordagem bayesiana para a estimação dos parâmetros, consideramos variáveis latentes na distribuição a *posteriori* e implementamos o algoritmo Gibbs sampling através das distribuições condicionais completas. Consideramos para o tamanho populacional uma distribuição a *priori* não informativa. Em seguida, propusemos o modelo em que considera dois ou mais componentes de mistura nas probabilidades de detecção. Os critérios AIC e BIC se mostraram eficientes em selecionar corretamente o modelo gerador dos dados na abordagem clássica. Já na abordagem bayesiana, a utilização do critério DIC foi ineficiente na escolha do verdadeiro modelo, apesar do mesmo apresentar estimativas pontuais bem próximas aos verdadeiros valores populacionais. Por este fato, consideramos como pesquisa futura a utilização do método *reversible jump* (Green, 1995; Richardson and Green, 1997) para seleção de modelos sob o enfoque bayesiano.

Referências Bibliográficas

- Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, *in* ‘Second international symposium on information theory’, Akademinai Kiado, pp. 267–281.
- Ananda, M. M. (1997), ‘Bayesian methods for mark-resighting surveys’, *Communications in Statistics-Theory and Methods* **26**, 685–697.
- Barger, K. J.-A. (2006), Mixtures of exponential distributions to describe the distribution of Poisson means in estimating the number of unobserved classes, PhD thesis, Cornell University.
- Basu, S. and Ebrahimi, N. (2001), ‘Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence’, *Biometrika* **88**, 269–279.
- Best, N., Cowles, M. K. and Vines, K. (1995), ‘Coda* convergence diagnosis and output analysis software for gibbs sampling output version 0.30’, *MRC Biostatistics Unit, Cambridge* **52**.
- Böhning, D. and Schön, D. (2005), ‘Nonparametric maximum likelihood estimation of population size based on the counting distribution’, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(4), 721–737.
- Bolsoni, S. B. (2002), *Estimação dos parâmetros de uma população a partir de observações incompletas da distribuição multinomial*, Dissertação de Mestrado, Centro de Ciências Exatas e Tecnologia, Departamento de Estatística, Universidade Federal de São Carlos.
- Burnham, K. P. and Overton, W. S. (1978), ‘Estimation of the size of a closed population when capture probabilities vary among animals’, *Biometrika* **65**, 625–633.
- Burnham, K. P. and Overton, W. S. (1979), ‘Robust estimation of population size when capture probabilities vary among animals’, *Ecology* **60**(5), 927–936.

- Casella, G. and Berger, R. L. (1990), *Statistical inference*, Vol. 70, Duxbury Press Belmont, CA.
- Casella, G. and George, E. I. (1992), ‘Explaining the gibbs sampler’, *The American Statistician* **46**(3), 167–174.
- Castledine, B. (1981), ‘A bayesian analysis of multiple-recapture sampling for a closed population’, *Biometrika* **68**, 197–210.
- Celeux, G. and Diebolt, J. (1985), ‘The sem algorithm: a probabilistic teacher algorithm derived from the em algorithm for the mixture problem’, *Computational statistics quarterly* **2**(1), 73–82.
- Chapman, D. G. (1954), ‘The estimation of biological populations’, *The Annals of Mathematical Statistics* pp. 1–15.
- Cubaynes, S., Lavergne, C., Marboutin, E. and Gimenez, O. (2012), ‘Assessing individual heterogeneity using model selection criteria: how many mixture components in capture–recapture models?’, *Methods in Ecology and Evolution* **3**(3), 564–573.
- Darroch, J. (1958), ‘The multiple-recapture census: I. estimation of a closed population’, *Biometrika* pp. 343–359.
- Darroch, J. N. (1959), ‘The multiple-recapture census: Ii. estimation when there is immigration or death’, *Biometrika* **46**, 336–351.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), ‘Maximum likelihood from incomplete data via the em algorithm’, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–38.
- Diebolt, J. and Robert, C. (1993), ‘Discussion of bayesian computations via the gibbs sampler by afm smith and g’, *Roberts. J. Royal Statist. Soc. Series B* **55**, 71–72.
- Diebolt, J. and Robert, C. P. (1994), ‘Estimation of finite mixture distributions through bayesian sampling’, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 363–375.
- Ebrahimi, N. B. (1997), ‘On the statistical analysis of the number of errors remaining in a software design document after inspection’, *Software Engineering, IEEE Transactions on* **23**, 529–532.

- Edwards, W. R. and Eberhardt, L. (1967), 'Estimating cottontail abundance from livetrapping data', *The Journal of Wildlife Management* pp. 87–96.
- Efron, B. (1979), 'Bootstrap methods: another look at the jackknife', *The annals of Statistics* pp. 1–26.
- Efron, B. and Efron, B. (1982), *The jackknife, the bootstrap and other resampling plans*, Vol. 38, SIAM.
- Eick, S. G., Loader, C. R., Vander Wiel, S. A. and Votta, L. G. (1993), 'How many errors remain in a software design document after inspection?', *Computing science and statistics* pp. 195–195.
- Gelman, A. and Rubin, D. B. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical science* pp. 457–472.
- George, E. I. (1992), 'Capture-recapture estimation via gibbs sampling', *Biometrika* **79**, 677–683.
- Green, P. J. (1995), 'Reversible jump markov chain monte carlo computation and bayesian model determination', *Biometrika* **82**(4), 711–732.
- Gupta, M. R. and Chen, Y. (2011), *Theory and Use of the EM Algorithm*, Now Publishers Inc.
- Hunter, A. and Griffiths, H. (1978), 'Bayesian approach to estimation of insect population size', *Technometrics* **20**, 231–234.
- Jolly, G. (1982), 'Mark-recapture models with parameters constant in time.', *Biometrics* **38**(2), 301–321.
- Jolly, G. M. (1965), 'Explicit estimates from capture-recapture data with both death and immigration-stochastic model', *Biometrika* **52**, 225–247.
- Laplace, P. S. (1783), 'Sur les naissances, les mariages et les morts', *Histoire de l'Academie Royale des Sciences* p. 693.
- Lee, A. (2002), 'Effect of list errors on the estimation of population size', *Biometrics* **58**, 185–191.
- Lee, A. J., Seber, G. A., Holden, J. K. and Huakau, J. T. (2001), 'Capture–recapture, epidemiology, and list mismatches: Several lists', *Biometrics* **57**, 707–713.

- Leite, J. G., Rodrigues, J. and Milan, L. A. (2000), 'A bayesian analysis for estimating the number of species in a population using nonhomogeneous poisson process', *Statistics & Probability Letters* **48**(2), 153–161.
- Lincoln, F. C. (1930), 'Calculating waterfowl abundance on the basis of banding returns'.
- McLachlan, G. and Krishnan, T. (2007), *The EM algorithm and extensions*, Vol. 382, John Wiley & Sons.
- Nayak, T. K. (1988), 'Estimating population size by recapture sampling', *Biometrika* **75**, 113–120.
- Parker, R. A. (1963), 'On the estimation of population size, mortality, and recruitment', *Biometrics* **19**(2), 318–323.
- Petersen, C. (1896), 'The yearly immigration of young plaice into the limfjord from the german sea', *Report of the Danish Biological Station* **6**, 1–48.
- Pollock, K. H. (1980), *Capture-recapture models: a review of current methods, assumptions and experimental design*, Citeseer.
- Pollock, K. H. (1991), 'Review papers: Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: Past, present, and future', *Journal of the American Statistical Association* **86**, 225–238.
- Richardson, S. and Green, P. J. (1997), 'On bayesian analysis of mixtures with an unknown number of components (with discussion)', *Journal of the Royal Statistical Society: series B (statistical methodology)* **59**(4), 731–792.
- Sanathanan, L. (1972), 'Estimating the size of a multinomial population', *The Annals of Mathematical Statistics* pp. 142–152.
- Schwarz, G. et al. (1978), 'Estimating the dimension of a model', *The annals of statistics* **6**(2), 461–464.
- Seber, G. (1986), 'A review of estimating animal abundance', *Biometrics* pp. 267–292.
- Seber, G. A. (1982), 'The estimation of animal abundance'.

- Seber, G. A. F. (1965), 'A note on the multiple-recapture census', *Biometrika* pp. 249–259.
- Shimizu, G. Y. (2002), *Estimação do tamanho de uma população fechada: método de captura-recaptura com um único estágio de marcação*, Dissertação de Mestrado, Centro de Ciências Exatas e Tecnologia, Departamento de Estatística, Universidade Federal de São Carlos.
- Silva, K. B. R. d. (2006), *Estimativas de máxima verossimilhança e bayesianas do número de erros de um software.*, Dissertação de Mestrado, Centro de Ciências Exatas e Tecnologia, Departamento de Estatística, Universidade Federal de São Carlos.
- Smith, P. J. (1988), 'Bayesian methods for multiple capture-recapture surveys.', *Biometrics* **44**(4), 1177–1189.
- Smith, P. J. (1991), 'Bayesian analyses for a multiple capture-recapture model', *Biometrika* **78**, 399–407.
- Spiegelhalter, D. J., Best, N., Carlin, B. P. and Van der Linde, A. (1998), Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models, Technical report, Research Report, 98-009.
- Tanner, M. A. and Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American statistical Association* **82**(398), 528–540.
- Wang, X. (2002), *Bayesian Analysis of Capture-recapture Models.*, Ph.D. Dissertation, University of Missouri, Columbia.
- Yoshida, O., Leite, J. G. and Bolfarine, H. (1999), 'Stochastic monotonicity properties of bayes estimation of the population size for capture–recapture data', *Statistics & probability letters* **42**, 257–266.

Apêndice A

Apêndice

Rotina: Estimativas de máxima verossimilhança com l níveis.

```
EMV.Generalizado = function(Dados,c.N,c.a,c.p,epsilon){

#----- Etapas iniciais -----#

c.a = c.a/sum(c.a)      ;
n = sum(Dados)          ; k = length(Dados)
y = seq(0,k,by=1)       ; t = length(c.a)
qjs = matrix(NA,t,(k+1))
f.p = f.a = c()         ; Iteracoes = 0
#-----#

repeat{ #--- start algorithm ---#
Iteracoes = Iteracoes + 1
DadosAtualizados = c(c.N-n,Dados)

repeat{ #--- start EM algorithm ---#

for (j in 0:k){ #--- probabilidades multinomais (Dist. condicional) ---#
```

```

    for (s in 1:t){ qjs[s,j+1] = c.a[s]*dbinom(j,k,c.p[s])  }
    qjs[,j+1] = qjs[,j+1]/sum(qjs[,j+1])
}

for (s in 1:t){ #--- atualização dos parâmetros perturbadores ---#
    f.p[s] = sum(y * qjs[s,] *
                DadosAtualizados)/sum(k* qjs[s,] * DadosAtualizados)
    f.a[s] = sum(qjs[s,] * DadosAtualizados) /c.N
}

Dif = sum(abs(f.p-c.p)) + sum(abs(f.a-c.a)) #--- critério de convergência ---#
if (Dif < epsilon) break
c.p <- f.p ; c.a <- f.a

} #--- end EM algorithm ---#

razao = 0 ; for (s in 1:t) razao = razao + f.a[s] * dbinom(0,k,f.p[s])
f.N = floor(n/(1-razao))
if (f.N == c.N) break
c.N = f.N
} #--- end algorithm ---#

EMV = c(f.N,round(f.a,4),round(f.p,4), Iteracoes)
return(EMV)
}

```


Rotina: Estimativas de máxima verossimilhança condicional com dois tipos de erros

```

EMVC.2M = function(Dados,c.a,c.pd,c.pf,epsilon) {

#----- Etapas Iniciais -----#
k = length(Dados)      ; n = sum(Dados) ;
Iteracoes = 0          ; y = seq(1,k,by=1) ;
#-----#

repeat{ #--- start EM algorithm ---#

  Iteracoes = Iteracoes + 1 ;

  py = c.a*dbinom(y,k,c.pd)/(c.a*dbinom(y,k,c.pd)+(1-c.a)*dbinom(y,k,c.pf))
  p10 = c.a*dbinom(0,k,c.pd)
  p00 = (1-c.a)*dbinom(0,k,c.pf)
  p0 = p10 + p00

  Soma1D = sum(Dados*py)      ;      Soma1F = sum(Dados*(1-py))
  Soma2D = sum(y*Dados*py)    ;      Soma2F = sum(y*Dados*(1-py))

  f.pd = Soma2D / ( k*Soma1D + k*n*p10/(1-p0) )
  f.pf = Soma2F / ( k*Soma1F + k*n*p00/(1-p0) )
  f.a = p10 + (1-p0)/n * Soma1D

  D.a = abs(f.a-c.a)
  D.pd = abs(f.pd-c.pd)
  D.pf = abs(f.pf-c.pf)

  if ( D.a<epsilon & D.pd<epsilon & D.pf<epsilon) break
}

```

```

c.a = f.a ; c.pd = f.pd ; c.pf = f.pf

} #--- end EM algorithm ---#

N.hat = floor(n/(1-(f.a*dbinom(0,k,f.pd)+(1-f.a)*dbinom(0,k,f.pf))))
Estimativas = c(N.hat,round(c(f.a,f.pd,f.pf),4),Iteracoes);
return(Estimativas)
}

```

Rotina: Algoritmo Gibbs sampling com l níveis.

```

GibbsSampling.r = function(L,ChuteN,Chuteomegas,Chutethetas,ValorR,aj,Dados,M) {

n.m = length(Chutepis) ; k = length(Dados) ; n = sum(Dados)
omegas = matrix(NA,L,n.m) ; omegas[1,] = Chuteomegas
thetas = matrix(NA,L,n.m) ; pis[1,] = Chutethetas
N = c() ; N[1] = ChuteN
y = seq(0,k,by=1)
hiperDiri = hiperBeta1 = hiperBeta2 = c()
probs.qj = Vjs = matrix(NA,k+1,n.m)

for (i in 2:L){

DadosC = c(N[i-1]-n,Dados)

for (j in 1:(k+1)) {
probs.qj[j,] = q.js(j-1,1:n.m,omegas[i-1,],thetas[i-1,],k,n.m)
Vjs[j,] = rmultinom(1,DadosC[j],probs.qj[j,])
}
}
}

```

```

for (s in 1:n.m) {
  hiperDiri[s] = sum(Vjs[,s]) + aj[s]
  hiperBeta1[s] = sum(y*Vjs[,s]) +1
  hiperBeta2[s] = sum((k-y)*Vjs[,s])+1
}
omegas[i,] = rdirichlet(1,hiperDiri)

thetas[i,1] = rtrunc(1,"beta", a = 0.001, b =thetas[i-1,2],
  shape1= hiperBeta1[1], shape2 = hiperBeta2[1])
if(n.m>2){
for (s in 2:(n.m-1)) {
  thetas[i,s] = rtrunc(1,"beta", a = thetas[i,s-1],
  b = thetas[i-1,s+1] , shape1= hiperBeta1[s], shape2 = hiperBeta2[s])
}}
thetas[i,n.m] = rtrunc(1,"beta", a = thetas[i,(n.m-1)],
b = 1 , shape1= hiperBeta1[n.m], shape2 = hiperBeta2[n.m])

h0 = sum(omegas[i,1:n.m] * dbinom(0,k,thetas[i,1:n.m]))
if (r==0){
  N[i] = rnbinom(1,n+1,1-h0) + n}
if (r==1){
  N[i] = rnbinom(1,n,1-h0) + n}
if ( r!=1 & r!=0){
  N[i] = SimilarN.cond.Nr(h0,n,M,r) }

} # Fim: Algoritmo Gibbs sampling
Posterioris = cbind(N,alphas,pis)
return(Posterioris)
}

```

Funções auxiliares

```

q.js = function(j,s,alphas,pis,k,n.m){
  constante = 0
  for (l in 1:n.m) constante = constante + alphas[l] * dbinom(j,k,pis[l])
  resul = alphas[s] * dbinom(j,k,pis[s]) / constante
  return(resul)
}

post.cond.Nr = function(N,n,h0,r) {
  value = lgamma(N+1) - lgamma(N-n+1) - lgamma(n+1) + (N-n) * log(h0) - r * log(N)
  return(exp(value))
}

SimilarN.cond.Nr = function(h0,n,M,r){
  x = n:M
  y = round( post.cond.Nr(x,n,h0,r) / sum(post.cond.Nr(x,n,h0,r)) ,10)
  acumulada = c() ; acumulada[1] = y[1]
  for (i in 2:length(y)) acumulada[i] = round( y[i] + acumulada[i-1], 10)
  u = round(runif(1, min = 0.000001, max = 0.999999),10)
  acumulada = c(0,acumulada)
  for (i in 1:(length(acumulada)-1)) {
    if (u >= acumulada[i] & u < acumulada[i+1] ) ValorN = x[i]
  }
  return(ValorN)
}

```