

# Eliminação de Parâmetros Perturbadores na Estimação de Tamanhos Populacionais

ANA CLÁUDIA FESTUCCI

UFSCar - São Carlos/SP

Janeiro/2010

Universidade Federal de São Carlos  
Centro de Ciências Exatas e de Tecnologia  
Departamento de Estatística

# Eliminação de Parâmetros Perturbadores na Estimação de Tamanhos Populacionais

ANA CLÁUDIA FESTUCCI

PROF. DR. JOSÉ GALVÃO LEITE

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.

UFSCar - São Carlos/SP

Janeiro/2010

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

F418ep

Festucci, Ana Cláudia.

Eliminação de parâmetros perturbadores na estimação de tamanhos populacionais / Ana Cláudia Festucci. -- São Carlos : UFSCar, 2010.

97 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2009.

1. Estatística matemática. 2. Função de verossimilhança.  
3. Métodos de estimação. 4. População. I. Título.

CDD: 519.54 (20<sup>a</sup>)

**Ana Cláudia Festucci**

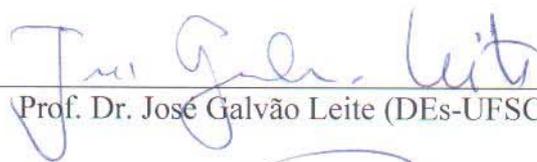
**Eliminação de parâmetros perturbadores na estimação de tamanhos populacionais**

Dissertação apresentada à Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Mestre em Estatística.

Aprovada em 15 de dezembro de 2009.

**BANCA EXAMINADORA**

Presidente



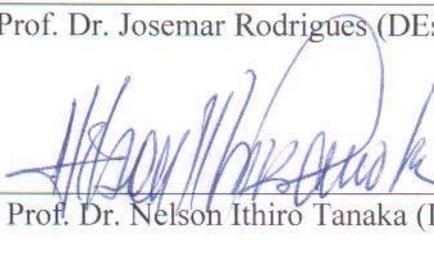
Prof. Dr. José Galvão Leite (DEs-UFSCar/ Orientador)

1º Examinador



Prof. Dr. Josemar Rodrigues (DEs-UFSCar)

2º Examinador



Prof. Dr. Nelson Ithiro Tanaka (IME-USP)

*Dedico esta dissertação aos meus irmãos Nina  
e Car, meus eternos alicerces, por formarem  
comigo a trílice que sustenta nossas vidas.*

# Agradecimentos

Agradeço primeira e principalmente a Deus, por guiar sempre meus passos, pelas oportunidades de crescimento e pelo presente que me destes ao me cercar de pessoas tão especiais.

Ao meu pai e minha mãe, que sempre estiveram presentes (mesmo nas ausências); pelos ensinamentos e pelos exemplos, por me deixarem aprender com seus erros e lidar com os meus próprios; pela pessoa que me tornei, graças à vocês.

Aos meus amados irmãos, Nina, Car e Bia, pelo amor imprescindível, pelo apoio, pela força, pelo incentivo e confiança depositados em mim; pelas brigas e discussões que fizeram como que ficássemos cada vez mais unidos.

À minha madrastra Rosângela, por não me deixar abrir mão do mestrado pelas dificuldades financeiras enfrentadas, pelo apoio a mim dedicado e pelo incentivo sempre presente. Também aos seus filhos, Jaque e Anderson, irmãos que Deus me deu, com os quais sempre pude contar quando precisei.

Às amadas Tia Raquel e Nick, que também sempre estiveram presentes, me dando força, me incentivando; por serem minha família querida, minha segunda mãe e minha irmã de coração, e pelas tardes desestressantes na varanda.

Ao amigo e namorado Leo, por não passar sempre a mão na minha cabeça, mas sim me ajudar a enxergar os principais erros e por estar sempre presente quando precisei.

A todos os meus amigos, em especial Paty, Vi e Zeh, que sempre estiveram carinhosamente presentes incentivando-me e confiando em minha capacidade.

Às minhas queridas amigas de república, Marina, Mari, Pat, Japa, Cami e Sol, que foram minha família, verdadeiras irmãs, com quem pude sempre contar ao longo

desses dois anos.

Aos queridos amigos da pós graduação, Hipólito, Saulo, Bruno, Rubiane e Ernesto, por toda ajuda que me deram no decorrer deste trabalho, mas principalmente pela amizade conquistada e por me aguentarem, é claro.

Ao meu orientador, José Galvão Leite, pela orientação, dedicação, atenção, paciência e, também, pelas puxadas de orelha, quase sempre merecidas, que permitiram a elaboração e conclusão deste trabalho.

Aos professores Luis Aparecido Milan e Josemar Rodrigues, membros da banca do exame de qualificação, pelas sugestões que tanto enriqueceram este trabalho. Também ao professor Nelson Ithiro Tanaka, membro da banca julgadora da defesa, pelas construtivas considerações.

Aos demais professores do DEs-UFSCar, pelos ensinamentos a nós dedicados.

Aos funcionários do DEs-UFSCar pela simpatia e disponibilidade sempre que precisamos, em especial à Tia Luiza, pelas palavras de conforto, de incentivo e pelo carinho.

Ao Departamento de Estatística da UFSCar, pela oportunidade de realização do curso de mestrado.

Finalmente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio concedido para este trabalho.

*"Pedras no caminho? Guardo todas, um dia vou  
construir um castelo..."* (Fernando Pessoa)

# Resumo

Nesta dissertação utilizamos o processo de captura-recaptura para estimar o tamanho de uma população fechada. Analisamos três modelos estatísticos diferentes e, para cada um deles, através de diversas metodologias de eliminação de parâmetros perturbadores, determinamos as funções de verossimilhança e de verossimilhança perfilada, condicional, integrada uniforme, integrada de Jeffreys e integrada generalizada do tamanho populacional, com exceção do último modelo onde determinamos uma função análoga à função de verossimilhança condicional, denominada função de verossimilhança restrita integrada. Em cada capítulo determinamos as respectivas estimativas de máxima verossimilhança e construímos intervalos de confiança empíricos para o tamanho populacional, bem como determinamos os erros quadráticos médios empíricos das estimativas e estudamos, através de dados simulados, as performances dos modelos.

**Palavras-chave:** processo de captura-recaptura, parâmetro perturbador, funções de verossimilhança, verossimilhança integrada uniforme, verossimilhança perfilada, verossimilhança condicional, estimativas de máxima verossimilhança.

# Abstract

In this study, we used the capture-recapture procedure to estimate the size of a closed population. We analysed three different statistics models. For each one of these models we determined - through several methods of eliminating nuisance parameters - the likelihood function and the profile, conditional, uniform integrated, Jeffreys integrated and generalized integrated likelihood functions of the population size, except for the last model where we determined a function that is analogous to the conditional likelihood function, called integrated restricted likelihood function. In each instance we determined the respective maximum likelihood estimates, the empirical confidence intervals and the empirical mean squared errors of the estimates for the population size and we studied, using simulated data, the performances of the models.

**Keywords:** capture-recapture procedure, nuisance parameters, likelihood function, profile, conditional, uniform integrated, maximum likelihood estimates.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Modelo Binomial sem Remoção</b>	<b>7</b>
2.1	Modelo Estatístico e Função de Verossimilhança	7
2.2	Funções de Verossimilhança	8
2.2.1	Função de Verossimilhança Perfilada	8
2.2.2	Função de Verossimilhança Condicional	9
2.2.3	Funções de Verossimilhança Integradas: Uniforme, Jeffreys e Generalizada	11
2.3	Estimativas de Máxima Verossimilhança	14
2.3.1	Solução Gráfica	16
2.3.2	Solução por Continuidade	19
2.3.3	Estudo da Performance do Modelo Binomial sem Remoção	22
2.3.4	Conclusões	32
<b>3</b>	<b>Modelo Binomial com Marcação</b>	<b>35</b>
3.1	Modelo Estatístico e Função de Verossimilhança	36
3.2	Funções de Verossimilhança	38
3.2.1	Função de Verossimilhança Perfilada	38
3.2.2	Função de Verossimilhança Condicional	38

---

3.2.3	Funções de Verossimilhança Integradas: Uniforme, Jeffreys e Generalizada . . . . .	40
3.3	Estimativas de Máxima Verossimilhança . . . . .	43
3.3.1	Solução Gráfica . . . . .	44
3.3.2	Solução por Continuidade . . . . .	46
3.3.3	Estudo da Performance do Modelo Binomial com Marcação . . . . .	48
3.3.4	Conclusões . . . . .	58
<b>4</b>	<b>Modelo Multinomial . . . . .</b>	<b>61</b>
4.1	Modelo Estatístico e a Função de Verossimilhança . . . . .	61
4.2	Funções de Verossimilhança . . . . .	63
4.2.1	Função de Verossimilhança Perfilada . . . . .	63
4.2.2	Função de Verossimilhança Restrita Integrada . . . . .	64
4.2.3	Funções de Verossimilhança Integradas: Uniforme, Jeffreys e Generalizada . . . . .	66
4.3	Estimativas de Máxima Verossimilhança . . . . .	70
4.3.1	Estudo da Performance do Modelo Multinomial . . . . .	72
4.3.2	Conclusões . . . . .	75
<b>5</b>	<b>Considerações Finais . . . . .</b>	<b>77</b>
	<b>Referências . . . . .</b>	<b>79</b>
<b>A</b>	. . . . .	<b>82</b>
<b>B</b>	. . . . .	<b>89</b>
<b>C</b>	. . . . .	<b>95</b>

# Capítulo 1

## Introdução

A motivação para o estudo do tema que resultou nesta dissertação surgiu do artigo Berger *et al.* (1999), onde os autores discutiram métodos de eliminação de parâmetros perturbadores para vários modelos estatísticos, evidenciando a dificuldade em se trabalhar com a função de verossimilhança usual para modelos discretos. Esta dificuldade vem do fato de que os ajustes de modelos a conjuntos de dados discretos são, em geral, inapropriados.

A eliminação de parâmetros perturbadores de uma função de verossimilhança é considerada um problema complexo em inferência estatística. Essa problemática foi evidenciada no século passado com a rejeição de Fisher e Neyman à escola de Bayes-Laplace (Berger *et al.*, 1999), que considerava o uso da função de verossimilhança integrada uniforme um fato óbvio. A partir de então, os pesquisadores tentaram encontrar caminhos alternativos para eliminação de tais parâmetros. Provavelmente, os primeiros exemplos relativos a abordagem frequentista para o problema são encontrados em Fisher (1915, 1921).

Foram inúmeros os esforços para criar uma teoria de verossimilhança para eliminação de parâmetros perturbadores (Barnard *et al.*, 1962). O início da escola de verossimilhança "moderna" pode ser definido talvez a partir de Kalbfleisch & Sprott (1970, 1973), que propuseram um estudo sistemático de vários métodos de eliminação de tais parâmetros (inclusive a função de verossimilhança integrada) e acabaram por abrir caminho para um campo de pesquisa promissor.

Provavelmente a abordagem de verossimilhança mais simples na eliminação de parâmetros perturbadores consiste em substituí-los por seus estimadores de máxima verossimilhança condicionais, conduzindo à função de verossimilhança perfilada, que pode então ser usada como uma função de verossimilhança usual. No entanto, muitos exemplos de comportamento enganoso dessa função (Neyman & Scott, 1948; Cruddas *et al.*, 1989) originaram várias correções dessa metodologia, como as funções de verossimilhança perfilada modificada e de verossimilhança perfilada condicional, levando-se em consideração que a simples substituição do parâmetro perturbador por uma estimativa pontual não necessariamente é uma alternativa eficaz. Excelentes referências incluem Reid (1995, 1996), Fraser & Reid (1989) e, para uma versão bayesiana dessas abordagens, Sweeting (1995a,b, 1996).

Outra abordagem de verossimilhança surge quando um ou mais componentes da estatística suficiente tem distribuição marginal ou condicional que depende do parâmetro de interesse mas não depende do parâmetro perturbador. Nesses casos, essas distribuições são geralmente usadas como uma função de verossimilhança do parâmetro de interesse e são chamadas funções de verossimilhança condicional ou marginal. Como o restante dos termos é ignorado, e estes dependem do parâmetro de interesse, há certa perda de informação. Entretanto, defensores dessa abordagem sugerem que a perda é somente na informação do parâmetro de interesse que é inextricável ao parâmetro perturbador.

Do ponto de vista bayesiano subjetivo, o problema tem uma solução trivial: simplesmente integrar a posteriori conjunta com relação ao parâmetro perturbador e trabalhar com a distribuição à posteriori marginal do parâmetro de interesse resultante.

Algumas questões que podem ser colocadas são o que é função de verossimilhança e porque eliminar parâmetros perturbadores.

Uma definição usual de função de verossimilhança (e geralmente vaga) de Savage (1976) inspirada em Fisher diz que ela é a "probabilidade ou densidade da observação como função do parâmetro". Bayarri *et al.* (1988) argumentam que tal definição não pode ser única e julgam-na como incompleta devido ao fato que ela não orienta sobre como lidar com uma variável não observada (por exemplo uma observação futura). Deve-se condicionar nela como faz-se com os parâmetros desconhecidos ou tratá-la como outra observação? Eles discutiram que a inferência baseada na função de verossimilhança vai

dependem dessa escolha e oferecem exemplos ilustrando que qualquer das escolhas pode ser a melhor delas dependendo de cada caso.

A segunda questão é interessante pois, apesar de ser natural que a inferência final sobre o parâmetro de interesse deva ser livre do parâmetro perturbador, não está claro que seja preciso passar por uma função de verossimilhança livre do mesmo. A maioria das análises não bayesianas passa por algo meio que intermediário, o que não acontece com as análises bayesianas. Por exemplo, hoje em dia muitas análises bayesianas fazem a inferência sobre o parâmetro de interesse baseados diretamente nos valores gerados para tal parâmetro através do método de Monte Carlo, ou seja, não há uma necessidade aparente de considerar explicitamente uma função de verossimilhança livre do parâmetro perturbador (inclusive é comum para bayesianos introduzir parâmetros perturbadores artificiais que simplifiquem o processo). Entretanto, há várias utilidades da função de verossimilhança livre do parâmetro perturbador em estudos bayesianos que justificam o seu uso. Entre elas destacamos:

1. *Informação Científica*: ao apresentar separadamente o gráfico da função de verossimilhança do parâmetro de interesse e o da distribuição à priori utilizada para o mesmo, além de possibilitar a observação do efeito desta priori na função, também permite que outros pesquisadores utilizem suas próprias prioris para o parâmetro de interesse.
2. *Análise de Sensibilidade*: é útil conhecer a função de verossimilhança do parâmetro de interesse ao estudar a sensibilidade da distribuição à priori escolhida.
3. *Custo de Eliciação*: em geral, como é muito dispendioso em termos de tempo e esforço obter distribuições à priori subjetivas, ao eliminar os parâmetros perturbadores pode-se concentrar os esforços apenas na eliciação subjetiva de uma priori para o parâmetro de interesse.
4. *Objetividade*: em algumas aplicações, há uma inegável necessidade de uma análise objetiva e, apesar de muitos estatísticos serem, com razão, céticos quanto a essa possibilidade (Berger & Berry, 1988), o uso da função de verossimilhança do parâmetro de interesse e de uma função peso padrão para o parâmetro perturbador pode satisfazer esta necessidade.

5. *Combinação de Funções de Verossimilhança*: se as informações obtidas sobre o parâmetro de interesse são de fontes diferentes independentes e são originárias de funções de verossimilhança, então é possível sumarizar as informações através do produto das mesmas. Berger *et al.* (1999) alertam sobre os cuidados na multiplicação de funções de verossimilhança.
6. *Prioris Impróprias*: o uso da função de verossimilhança integrada parece reduzir as dificuldades da utilização de prioris impróprias. Berger *et al.* (1999) ilustram tal fato.

Nesta dissertação, fizemos um estudo sobre alguns métodos de eliminação de parâmetros perturbadores (como nos casos das funções de verossimilhança perfilada, condicional e integrada) para três modelos estatísticos, com o objetivo de estimar o tamanho populacional,  $N$ , através das estimativas de máxima verossimilhança (emv).

Assumimos que a população é fechada, ou seja, não há nascimentos, mortes, imigração ou emigração de indivíduos da população durante o processo de estimação. Tal processo, conhecido na literatura por captura-recaptura, consiste na seleção de um número fixado ou aleatório de indivíduos da população em  $k$  ( $k \geq 2$ ) épocas (ocasiões) de amostragem.

O processo de captura-recaptura foi utilizado inicialmente por Laplace (1783) para estimar o tamanho da população da França, mas foi Petersen (1896) que utilizou o método pela primeira vez em ecologia ao estudar o fluxo migratório de peixes no mar Báltico. Um dos primeiros trabalhos aplicando o método no século vinte foi o de Lincoln (1930), que estimou o tamanho da população de patos selvagens na América do Norte e, a partir da década de cinquenta, diversos pesquisadores publicaram trabalhos sobre o assunto, tais como, Chapman (1954), Darroch (1959), Jolly (1965), Burnham & Overton (1978), Seber (1992) e Pollock (1991).

A importância deste método no manejo pesqueiro, no controle de insetos, no impacto da predação de uma espécie carnívora, etc., motivou muitos trabalhos de pesquisa em Estatística.

A inferência bayesiana também fez-se fortemente presente em problemas de captura-recaptura, como visto em trabalhos de pesquisadores como Hunter & Griffiths (1978),

Castledine (1981), Smith (1988, 1991), George & Robert (1992), Ananda (1997) e Yoshida *et al.* (1999).

Embora o processo de captura-recaptura fosse usado originalmente na estimação de população animal, há muitas aplicações desse processo em outras áreas do conhecimento, como por exemplo os artigos de Nayak (1988) e Basu & Ebrahimi (2001) na área de controle de erros de softwares e os de Lee *et al.* (2001) e Lee (2002) na área de controle de doenças não transmissíveis em populações.

No capítulo 2 apresentamos detalhadamente o modelo estatístico referente ao modelo binomial sem remoção. Neste modelo consideramos simplesmente a contagem do número de indivíduos capturados (recapturados) em cada uma das ocasiões de amostragem, sem identificar se o indivíduo já foi amostrado ou não em alguma ocasião anterior. Apresentamos a respectiva função de verossimilhança de  $N$  e  $p$ , probabilidade de captura de qualquer indivíduo da população, em qualquer ocasião, e contruímos as funções de verossimilhança perfilada, condicional, integrada uniforme, integrada de Jeffreys e integrada generalizada de  $N$ .

No segundo modelo, discutido no capítulo 3, consideramos a marcação dos indivíduos capturados pela primeira vez em cada ocasião e a identificação do indivíduo como capturado ou recapturado, de forma que no final do processo amostral temos o total de indivíduos distintos observados (Zacharias, 2000). Novamente apresentamos as respectivas funções de verossimilhança de  $N$  e  $p_1, p_2, \dots, p_k$ , onde  $p_j$  é a probabilidade de captura de qualquer indivíduo da população na ocasião  $j$ , e as funções de verossimilhança perfilada, condicional e integradas uniforme, de Jeffreys e generalizada de  $N$ .

No terceiro e último modelo, apresentado no capítulo 4, utilizamos novamente a técnica de marcação dos indivíduos capturados, mas neste caso consideramos os históricos (trajetórias) de captura de cada indivíduo, isto é, consideramos as sequências de capturas, recapturas e não-recapturas dos indivíduos observados em pelo menos uma ocasião e estimamos  $N$  baseados nos números de indivíduos da população que apresentaram cada histórico de captura. Transcorremos de maneira análoga aos modelos anteriores, com a diferença de que para este modelo apresentamos a função de verossimilhança restrita integrada de  $N$ , análoga à condicional dos modelos anteriores.

Para os três modelos citados realizamos estudos através de dados simulados com o

---

objetivo de obter estimativas do tamanho populacional. Determinamos as emv de  $N$  para cada um dos casos, considerando uma solução gráfica e outra considerando  $N$  como sendo a restrição de uma variável contínua. Além disso, contruímos empiricamente intervalos de confiança para o tamanho populacional e determinamos, ainda empiricamente, os erros quadráticos médios das estimativas.

Ao final de cada capítulo tecemos comentários e conclusões inerentes ao modelo discutido e, no último capítulo, apresentamos uma conclusão geral sobre nosso trabalho conjuntamente com sugestões para trabalhos subsequentes.

# Capítulo 2

## Modelo Binomial sem Remoção

Para a construção do modelo binomial sem remoção consideramos simplesmente a contagem dos indivíduos capturados em cada uma das  $k$  ocasiões amostrais,  $k \geq 2$ , e a simples devolução destes à população. Neste caso, como não há a possibilidade de distinguir os indivíduos recapturados, obtemos apenas as informações sobre o menor número possível de indivíduos distintos observados, ou seja, o maior número entre os números de indivíduos observados em cada uma das ocasiões e do total de indivíduos capturados, independentemente se determinado indivíduo foi ou não capturado em mais de uma ocasião.

Apresentamos neste capítulo as funções de verossimilhança e de verossimilhança perfilada, condicional e integradas para o modelo em questão, bem como as respectivas estimativas do tamanho populacional.

### 2.1 Modelo Estatístico e Função de Verossimilhança

Suponhamos que cada indivíduo seja capturado ou não, em qualquer ocasião, independentemente dos demais elementos da população e das outras ocasiões e que a probabilidade de captura de qualquer indivíduo, em qualquer ocasião, seja a mesma,  $p$ .

Seja  $n_j$  o número de indivíduos capturados na  $j$ -ésima ocasião,  $j = 1, 2, \dots, k$ . Logo  $n_1, n_2, \dots, n_k$  são independentes e identicamente distribuídos com  $n_j | N, p \sim \text{Binomial}$  com parâmetros  $N$  e  $p$ ,  $j = 1, 2, \dots, k$ .

Portanto, a função de verossimilhança de  $N$  e  $p$  é dada por

$$\begin{aligned}
 L(N, p | n_1, n_2, \dots, n_k) &= P(n_1, n_2, \dots, n_k | N, p) \\
 &= \prod_{j=1}^k P(n_j | N, p) \\
 &= \prod_{j=1}^k \binom{N}{n_j} p^{n_j} (1-p)^{N-n_j} \\
 &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] p^{\sum_{j=1}^k n_j} (1-p)^{Nk - \sum_{j=1}^k n_j} \\
 &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] p^n (1-p)^{Nk-n}, \tag{2.1}
 \end{aligned}$$

$N \geq n_0$  e  $0 < p < 1$ , onde  $n_0 = \max \{n_j, j = 1, 2, \dots, k\}$  é o menor número possível de indivíduos distintos observados e  $n = \sum_{j=1}^k n_j$  é o número total de indivíduos capturados (recapturados).

Então, o logaritmo natural da função de verossimilhança é dado por

$$l(N, p | n_1, n_2, \dots, n_k) = \sum_{j=1}^k \log \binom{N}{n_j} + n \log p + (Nk - n) \log(1-p), \tag{2.2}$$

$N \geq n_0$  e  $0 < p < 1$ .

## 2.2 Funções de Verossimilhança

No caso apresentado, o objetivo é a estimação do tamanho da população,  $N$ , que é o parâmetro de interesse e, conseqüentemente,  $p$  é o parâmetro perturbador. Neste contexto apresentamos algumas funções de verossimilhança construídas segundo diferentes metodologias de eliminação de  $p$ .

### 2.2.1 Função de Verossimilhança Perfilada

A função de verossimilhança perfilada consiste em maximizar, para  $N$  fixado, a função de verossimilhança usual, agora considerada como função apenas do parâmetro  $p$ . Então, temos a

**Definição 2.1** A função de verossimilhança perfilada de  $N$  é definida por

$$L^{\mathcal{P}}(N|n_1, n_2, \dots, n_k) = \sup_{p \in (0,1)} \left\{ L(N, p|n_1, n_2, \dots, n_k) \right\},$$

$N \geq n_0$ . Denotemos por  $\hat{p}(N)$  o ponto que maximiza  $L(N, p|n_1, n_2, \dots, n_k)$  para  $N$  fixado. Então,  $L^{\mathcal{P}}(N|n_1, n_2, \dots, n_k) = L(N, \hat{p}(N)|n_1, n_2, \dots, n_k)$ ,  $N \geq n_0$ .

Como maximizar a função de verossimilhança é equivalente à maximizar o seu logaritmo, então para determinar  $\hat{p}(N)$  temos de 2.2 que

$$\frac{\partial l(N, p|n_1, n_2, \dots, n_k)}{\partial p} = \frac{n}{p} - \frac{Nk - n}{1 - p} = 0$$

e

$$\frac{\partial^2 l(N, p|n_1, n_2, \dots, n_k)}{\partial p^2} = - \left[ \frac{n}{p^2} + \frac{Nk - n}{(1 - p)^2} \right] < 0,$$

$N \geq n_0$  e  $0 < p < 1$ , o que implica que  $\hat{p}(N) = \frac{n}{Nk}$ .

Logo, de 2.1 segue que a função de verossimilhança perfilada de  $N$  é dada por

$$\begin{aligned} L^{\mathcal{P}}(N|n_1, n_2, \dots, n_k) &= L(N, \hat{p}(N)|n_1, n_2, \dots, n_k) \\ &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \left( \frac{n}{Nk} \right)^n \left( 1 - \frac{n}{Nk} \right)^{Nk-n} \\ &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \frac{n^n}{(Nk)^n} \frac{(Nk - n)^{Nk-n}}{(Nk)^{Nk-n}} \\ &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \frac{n^n (Nk - n)^{Nk-n}}{(Nk)^{Nk}} \\ &\propto K^{\mathcal{P}}(N|n_1, n_2, \dots, n_k), \end{aligned}$$

onde

$$K^{\mathcal{P}}(N|n_1, n_2, \dots, n_k) = \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \frac{(Nk - n)^{Nk-n}}{(Nk)^{Nk}}, \quad (2.3)$$

$N \geq n_0$ , é o núcleo (kernel) da função de verossimilhança perfilada.

## 2.2.2 Função de Verossimilhança Condicional

Outro tipo de função de verossimilhança surge quando a função de verossimilhança 2.1 pode ser escrita como o produto de duas outras funções de verossimilhança, onde uma delas é uma função apenas de  $N$  e a outra é uma função de  $N$  e  $p$ .

Mais precisamente, temos o seguinte teorema.

**Teorema 2.1** *A função de verossimilhança 2.1 pode ser escrita sob a forma*

$$L(N, p | n_1, n_2, \dots, n_k) = L_1(N | n_1, n_2, \dots, n_k) \cdot L_2(N, p | n),$$

onde  $L_1(N | n_1, n_2, \dots, n_k)$  é a função de verossimilhança correspondente à distribuição de probabilidades condicional de  $n_1, n_2, \dots, n_k$ , dados  $n$ ,  $N$  e  $p$ , e  $L_2(N, p | n)$  corresponde a distribuição de probabilidades de  $n$ , dados  $N$  e  $p$ .

**Demonstração.** Como  $n | N, p \sim \text{Binomial}(N k, p)$  temos

$$\begin{aligned} L_1(N | n_1, n_2, \dots, n_k) &= P(n_1, n_2, \dots, n_k | n, N, p) \\ &= \frac{P(n_1, n_2, \dots, n_k, n | N, p)}{P(n | N, p)} \\ &= \frac{P(n_1, n_2, \dots, n_k | N, p) \mathbf{I}_{\{n\}}(\sum_{j=1}^k n_j)}{P(n | N, p)} \\ &= \frac{\left[ \prod_{j=1}^k \binom{N}{n_j} \right] p^n (1-p)^{N k - n}}{\binom{N k}{n} p^n (1-p)^{N k - n}} \\ &= \frac{\prod_{j=1}^k \binom{N}{n_j}}{\binom{N k}{n}}, \end{aligned}$$

$N \geq n_0$ , onde  $\mathbf{I}_{\{n\}}(\sum_{j=1}^k n_j)$  é a função indicadora que assume o valor 1 se  $n = \sum_{j=1}^k n_j$  e 0, caso contrário.

Notamos que a distribuição de  $n_1, n_2, \dots, n_k$  condicionado à  $n$ ,  $N$  e  $p$  na realidade não depende de  $p$ , o que implica que  $L_1(\cdot | n_1, n_2, \dots, n_k)$  depende somente de  $N$ .

Por outro lado,

$$L_2(N, p | n) = P(n | N, p) = \binom{N k}{n} p^n (1-p)^{N k - n},$$

$N k \geq n$  e  $0 < p < 1$ , o que implica por 2.1 que

$$\begin{aligned}
L(N, p | n_1, n_2, \dots, n_k) &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] p^n (1-p)^{Nk-n} \\
&= \frac{\prod_{j=1}^k \binom{N}{n_j}}{\underbrace{\binom{Nk}{n}}_{L_1(N|n_1, n_2, \dots, n_k)}} \cdot \underbrace{\binom{Nk}{n} p^n (1-p)^{Nk-n}}_{L_2(N, p|n)},
\end{aligned}$$

$N \geq n_0$  e  $0 < p < 1$ , o que prova o teorema. ■

A função de verossimilhança  $L_1(N|n_1, n_2, \dots, n_k)$  é denominada função de verossimilhança condicional de  $N$  e será denotada por  $L^c(N|n_1, n_2, \dots, n_k)$ . Logo,

$$L^c(N|n_1, n_2, \dots, n_k) = \frac{\prod_{j=1}^k \binom{N}{n_j}}{\binom{Nk}{n}}, \quad (2.4)$$

$N \geq n_0$ .

### 2.2.3 Funções de Verossimilhança Integradas: Uniforme, Jeffreys e Generalizada

A metodologia de verossimilhança integrada consiste na eliminação do parâmetro perturbador  $p$  da função de verossimilhança 2.1 através de sua integração com relação à certa função peso escolhida para  $p$ . Desse modo obtemos uma função de verossimilhança somente em função do parâmetro de interesse. Mais precisamente, temos a seguinte definição.

**Definição 2.2** A função de verossimilhança integrada de  $N$  com relação à função peso  $\pi(p|N)$ ,  $0 < p < 1$ , é dada por

$$L^*(N|n_1, n_2, \dots, n_k) = \int_0^1 L(N, p | n_1, n_2, \dots, n_k) \pi(p|N) dp. \quad (2.5)$$

Notamos que

$$\begin{aligned} L^*(N|n_1, n_2, \dots, n_k) &= P(n_1, n_2, \dots, n_k|N) \\ &= \int_0^1 P(n_1, n_2, \dots, n_k, p|N) \, dp \\ &= \int_0^1 \underbrace{P(n_1, n_2, \dots, n_k|N, p)}_{L(N, p|n_1, n_2, \dots, n_k)} \pi(p|N) \, dp, \end{aligned}$$

$N \geq n_0$ .

Doravante adotamos a linguagem da estatística bayesiana e denominamos  $\pi(p|N)$  de "função densidade condicional à priori de  $p$ , dado  $N$ ".

### (I) Função de Verossimilhança Integrada Uniforme

Um caso particular de função de verossimilhança integrada, definida a seguir, é o caso em que  $\pi(p|N)$  é a função densidade uniforme no intervalo  $(0, 1)$ .

**Definição 2.3** A função de verossimilhança integrada uniforme de  $N$ ,  $L^u(N|n_1, n_2, \dots, n_k)$ , é a função de verossimilhança integrada com relação à priori  $\pi^u(p|N) = \mathbf{I}_{(0,1)}(p)$ , isto é,

$$L^u(N|n_1, n_2, \dots, n_k) = \int_0^1 L(N, p|n_1, n_2, \dots, n_k) \, dp,$$

$N \geq n_0$ . Assim, a função de verossimilhança integrada uniforme para o modelo binomial sem remoção é dada por

$$\begin{aligned} L^u(N|n_1, n_2, \dots, n_k) &= \int_0^1 L(N, p|n_1, n_2, \dots, n_k) \, dp \\ &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \int_0^1 p^n (1-p)^{Nk-n} \, dp \\ &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \frac{\Gamma(n+1) \Gamma(Nk-n+1)}{\Gamma(n+1+Nk-n+1)} \\ &= \left[ \prod_{j=1}^k \frac{N!}{n_j! (N-n_j)!} \right] \frac{n! (Nk-n)!}{(Nk+1)!} \\ &\propto K^u(N|n_1, n_2, \dots, n_k), \end{aligned}$$

onde

$$K^u(N|n_1, n_2, \dots, n_k) = \left[ \prod_{j=1}^k \frac{N!}{(N-n_j)!} \right] \frac{(Nk-n)!}{(Nk+1)!}, \quad (2.6)$$

$N \geq n_0$ , é o núcleo de  $L^u(N|n_1, n_2, \dots, n_k)$ .

## (II) Função de Verossimilhança Integrada de Jeffreys

Outra distribuição à priori geralmente utilizada como função peso para  $p$  é a de Jeffreys dada pela seguinte definição.

**Definição 2.4** A distribuição à priori de Jeffreys para  $p$ , dado  $N$ , é dada por

$$\pi^{\mathcal{J}}(p|N) \propto \left[ I_{(n_1, n_2, \dots, n_k)}(N, p) \right]^{1/2},$$

onde  $I_{(n_1, n_2, \dots, n_k)}(N, p)$  é a medida de informação esperada de Fisher de  $p$  através de  $(n_1, n_2, \dots, n_k)$ , com a suposição de que  $N$  é conhecido.

Então,

$$I_{(n_1, n_2, \dots, n_k)}(N, p) = E \left[ - \frac{\partial^2 l(N, p | n_1, n_2, \dots, n_k)}{\partial p^2} \Bigg|_p \right],$$

logo, de 2.2 segue que

$$\begin{aligned} I_{(n_1, n_2, \dots, n_k)}(N, p) &= E \left[ \frac{n}{p^2} + \frac{Nk - n}{(1-p)^2} \Bigg| N, p \right] \\ &= \frac{E[n|N, p]}{p^2} + \frac{Nk - E[n|N, p]}{(1-p)^2} \\ &= \frac{Nkp}{p^2} + \frac{Nk - Nkp}{(1-p)^2} \\ &= \frac{Nk}{p} + \frac{Nk}{(1-p)} \\ &= Nk \left( \frac{1}{p} + \frac{1}{(1-p)} \right) = \frac{Nk}{p(1-p)}, \end{aligned}$$

o que implica que a priori de Jeffreys para  $p$ , dado  $N$ , é

$$\pi^{\mathcal{J}}(p|N) \propto \left( \frac{Nk}{p(1-p)} \right)^{1/2} \propto p^{-1/2} (1-p)^{-1/2} \mathbf{I}_{(0,1)}(p).$$

Portanto, de 2.1 e 2.5, segue que

$$\begin{aligned} L^{\mathcal{J}}(N | n_1, n_2, \dots, n_k) &= \int_0^1 L(N, p | n_1, n_2, \dots, n_k) \pi^{\mathcal{J}}(p|N) dp \\ &\propto \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \int_0^1 p^{n-1/2} (1-p)^{Nk-n-1/2} dp \\ &= \left[ \prod_{j=1}^k \frac{N!}{n_j! (N-n_j)!} \right] \frac{\Gamma(n+1/2) \Gamma(Nk-n+1/2)}{\Gamma(Nk+1)} \\ &\propto K^{\mathcal{J}}(N | n_1, n_2, \dots, n_k), \end{aligned}$$

onde

$$K^{\mathcal{J}}(N|n_1, n_2, \dots, n_k) = \left[ \prod_{j=1}^k \frac{N!}{(N - n_j)!} \right] \frac{\Gamma(Nk - n + 1/2)}{\Gamma(Nk + 1)}, \quad (2.7)$$

$N \geq n_0$ , é o núcleo de  $L^{\mathcal{J}}(N|n_1, n_2, \dots, n_k)$ .

### (III) Função de Verossimilhança Integrada Generalizada

Mais geralmente, podemos considerar como função peso para  $p$  uma priori  $Beta(\alpha, \beta)$ ,  $\alpha > 0$ ,  $\beta > 0$ , com  $\alpha$  e  $\beta$  conhecidos.

Então,

$$\pi^{\mathcal{S}}(p|N) \propto p^{\alpha-1} (1-p)^{\beta-1} \mathbf{I}_{(0,1)}(p)$$

e, de 2.1 e 2.5, temos

$$\begin{aligned} L^{\mathcal{S}}(N|n_1, n_2, \dots, n_k) &= \int_0^1 L(N, p|n_1, n_2, \dots, n_k) \pi^{\mathcal{S}}(p|N) dp \\ &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \int_0^1 p^{n+\alpha-1} (1-p)^{Nk-n+\beta-1} dp \\ &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \frac{\Gamma(n+\alpha) \Gamma(Nk-n+\beta)}{\Gamma(Nk+\alpha+\beta)} \\ &\propto K^{\mathcal{S}}(N|n_1, n_2, \dots, n_k), \end{aligned}$$

onde

$$K^{\mathcal{S}}(N|n_1, n_2, \dots, n_k) = \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \frac{\Gamma(Nk - n + \beta)}{\Gamma(Nk + \alpha + \beta)}, \quad (2.8)$$

$N \geq n_0$ , é o núcleo da função de verossimilhança integrada generalizada de  $N$ .

Notamos que, se  $\alpha = \beta = 1$  temos a função de verossimilhança integrada uniforme e se  $\alpha = \beta = 1/2$  temos a função de verossimilhança integrada de Jeffreys.

## 2.3 Estimativas de Máxima Verossimilhança

Nesta seção determinamos as estimativas de máxima verossimilhança (emv) de  $N$  correspondentes às diferentes funções de verossimilhança apresentadas anteriormente.

A emv de  $(N, p)$  é o ponto  $(\hat{N}, \hat{p})$  que maximiza a função de verossimilhança,  $L(N, p | n_1, n_2, \dots, n_k)$ , ou o seu logaritmo,  $l(N, p | n_1, n_2, \dots, n_k)$ . Para obter tal ponto, assumimos que  $N$  seja uma variável contínua (Dahiya, 1981) e encontramos  $(\hat{N}, \hat{p})$  através da solução do sistema

$$\begin{cases} \frac{\partial l(N, p | n_1, n_2, \dots, n_k)}{\partial p} = 0, \\ \frac{\partial l(N, p | n_1, n_2, \dots, n_k)}{\partial N} = 0, \end{cases} \quad (2.9)$$

$N \geq n_0$  e  $0 < p < 1$ .

Como de 2.2 temos

$$\begin{aligned} l(N, p | n_1, n_2, \dots, n_k) &= \sum_{j=1}^k \log \frac{N!}{(N - n_j)! n_j!} + n \log p + (N k - n) \log(1 - p) \\ &= k \log(N!) - \sum_{j=1}^k \log[(N - n_j)!] - \sum_{j=1}^k \log(n_j!) + n \log p + (N k - n) \log(1 - p) \\ &= k \log \Gamma(N + 1) - \sum_{j=1}^k \log \Gamma(N - n_j + 1) - \sum_{j=1}^k \log(n_j!) + n \log p + (N k - n) \log(1 - p), \end{aligned}$$

$N \geq n_0$  e  $0 < p < 1$ , então o sistema 2.9 pode ser escrito como

$$\begin{cases} \frac{n}{p} - \frac{N k - n}{1 - p} = 0, \\ k \Psi(N + 1) - \sum_{j=1}^k \Psi(N - n_j + 1) + k \log(1 - p) = 0, \end{cases}$$

$N \geq n_0$  e  $0 < p < 1$ , onde  $\Psi(z) = \frac{d \log \Gamma(z)}{dz}$ ,  $z > 0$ , é a função digama.

Assim,  $(\hat{N}, \hat{p})$  é solução do sistema

$$\begin{cases} p = \frac{n}{N k}, \\ k \Psi(N + 1) - \sum_{j=1}^k \Psi(N - n_j + 1) + k \log(1 - p) = 0, \end{cases}$$

$N \geq n_0$  e  $0 < p < 1$ , o que implica que  $\hat{N}$  é solução da equação

$$k \Psi(N + 1) - \sum_{j=1}^k \Psi(N - n_j + 1) + k \log \left( 1 - \frac{n}{N k} \right) = 0, \quad (2.10)$$

$N \geq n_0$ .

Finalmente, a estimativa do tamanho populacional,  $N$ , como é um valor inteiro, é obtido através do cálculo da função de verossimilhança de  $N$  e  $p$  aplicada nos dois valores inteiros próximos ao determinado em 2.10. A raiz de 2.10, bem como a env inteira de  $N$ , foram determinadas na seção 2.3.3, quando do estudo da performance do modelo, através de um programa implementado em R (versão 2.8.1) e descrito no apêndice A.

### 2.3.1 Solução Gráfica

Para a determinação das env correspondentes às funções de verossimilhança perfilada, condicional e integradas uniforme e de Jeffreys partimos inicialmente de um estudo gráfico. Construímos os gráficos das respectivas funções de verossimilhança (ou de seus núcleos) e, a partir destes, determinamos os pontos de máximo para cada amostra.

Evidentemente o gráfico da respectiva função é baseado em um número finito de valores de  $N$ , o que não nos permite analisar o seu comportamento para valores grandes de  $N$ . Como este comportamento, para todo  $N$  suficientemente grande, depende do limite da respectiva função, quando  $N \rightarrow \infty$ , determinamos os limites de cada uma das funções e os apresentamos nos teoremas que seguem.

**Teorema 2.2** *O núcleo da função de verossimilhança perfilada, dado em 2.3, converge para  $\frac{1}{\prod_{j=1}^k n_j!} \frac{1}{k^n} \exp(-n)$ , quando  $N \rightarrow \infty$ .*

**Demonstração.** De 2.3 temos

$$\begin{aligned}
 K^{\mathcal{P}}(N|n_1, n_2, \dots, n_k) &= \left[ \prod_{j=1}^k \binom{N}{n_j} \right] \frac{(Nk - n)^{Nk-n}}{(Nk)^{Nk}} \\
 &= \frac{1}{\prod_{j=1}^k n_j!} \frac{\prod_{j=1}^k N(N-1) \dots (N - (n_j - 1))}{(Nk - n)^n} \left( \frac{Nk - n}{Nk} \right)^{Nk} \\
 &= \frac{1}{\prod_{j=1}^k n_j!} \frac{\prod_{j=1}^k \frac{N}{N} \left( 1 - \frac{1}{N} \right) \dots \left( 1 - \frac{n_j - 1}{N} \right)}{\frac{1}{N^n} (Nk - n)^n} \left( 1 - \frac{n}{Nk} \right)^{Nk}
 \end{aligned}$$

$$= \frac{1}{\prod_{j=1}^k n_j!} \frac{\prod_{j=1}^k \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{n_j - 1}{N}\right)}{\left(k - \frac{n}{N}\right)^n} \left(1 - \frac{n}{Nk}\right)^{Nk}.$$

Como, para  $N \rightarrow \infty$ , temos

$$\left(1 - \frac{m}{N}\right) \rightarrow 1, \quad m < N, \quad m \text{ fixo}, \quad \left(k - \frac{n}{N}\right)^n \rightarrow k^n \text{ e } \left(1 - \frac{n}{Nk}\right)^{Nk} \rightarrow \exp(-n) \quad (2.11)$$

concluimos que

$$K^{\mathcal{P}}(N|n_1, n_2, \dots, n_k) \xrightarrow{N \rightarrow \infty} \frac{1}{\prod_{j=1}^k n_j!} \frac{1}{k^n} \exp(-n). \quad \blacksquare$$

**Teorema 2.3** A função de verossimilhança condicional, dada em 2.4, converge para  $\frac{n!}{\prod_{j=1}^k n_j!} \frac{1}{k^n}$ , quando  $N \rightarrow \infty$ .

**Demonstração.** De fato, temos de 2.4 que

$$\begin{aligned} L^{\mathcal{C}}(N|n_1, n_2, \dots, n_k) &= \frac{\prod_{j=1}^k \binom{N}{n_j}}{\binom{Nk}{n}} \\ &= \frac{n!}{\prod_{j=1}^k n_j!} \frac{\prod_{j=1}^k N(N-1)\dots(N-n_j+1)}{(Nk)(Nk-1)\dots(Nk-n+1)} \\ &= \frac{n!}{\prod_{j=1}^k n_j!} \frac{\prod_{j=1}^k \frac{N}{N} \left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{n_j-1}{N}\right)}{k \left(k - \frac{1}{N}\right) \dots \left(k - \frac{n-1}{N}\right)}. \end{aligned}$$

Novamente, de 2.11 temos que

$$L^{\mathcal{C}}(N|n_1, n_2, \dots, n_k) \xrightarrow{N \rightarrow \infty} \frac{n!}{\prod_{j=1}^k n_j!} \frac{1}{k^n}. \quad \blacksquare$$

**Teorema 2.4** O núcleo da função de verossimilhança integrada uniforme 2.6 converge para 0, quando  $N \rightarrow \infty$ , isto é, o seu logaritmo converge para  $-\infty$ , quando  $N \rightarrow \infty$ .

**Demonstração.** De 2.6 temos

$$K^{\mathcal{U}}(N|n_1, n_2, \dots, n_k) = \left[ \prod_{j=1}^k \frac{N!}{(N-n_j)!} \right] \frac{(Nk-n)!}{(Nk+1)!}$$

$$\begin{aligned}
& \prod_{j=1}^k N(N-1) \dots (N-n_j+1) \\
&= \frac{\prod_{j=1}^k N(N-1) \dots (N-n_j+1)}{(Nk+1)(Nk)(Nk-1) \dots (Nk-n+1)} \\
&= \frac{\prod_{j=1}^k \frac{N}{N} \left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{n_j-1}{N}\right)}{(Nk+1) \frac{Nk}{N} \left(k - \frac{1}{N}\right) \dots \left(k - \frac{n-1}{N}\right)} \\
&= \frac{1}{Nk+1} \frac{\prod_{j=1}^k \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{n_j-1}{N}\right)}{k \left(k - \frac{1}{N}\right) \dots \left(k - \frac{n-1}{N}\right)}.
\end{aligned}$$

Assim, por 2.11 e pelo fato de que  $\frac{1}{Nk+1} \rightarrow 0$ , quando  $N \rightarrow \infty$ , temos

$$K^u(N|n_1, n_2, \dots, n_k) \xrightarrow{N \rightarrow \infty} 0. \quad \blacksquare$$

**Teorema 2.5** *O núcleo da função de verossimilhança integrada de Jeffreys, dada em 2.7, converge para 0, quando  $N \rightarrow \infty$  e, conseqüentemente, o seu logaritmo converge para  $-\infty$ , quando  $N \rightarrow \infty$ .*

**Demonstração.** De 2.7 temos

$$\begin{aligned}
K^d(N|n_1, n_2, \dots, n_k) &= \prod_{j=1}^k \frac{N!}{(N-n_j)!} \times \\
&\times \frac{(Nk-n+\frac{1}{2}-1)(Nk-n+\frac{1}{2}-2) \dots (Nk-n+\frac{1}{2}-(Nk-n)) \Gamma(1/2)}{(Nk)!} \\
&= \Gamma(1/2) \frac{\prod_{j=1}^k N(N-1) \dots (N-n_j+1)}{(Nk)!} \prod_{l=1}^{Nk-n} (Nk-n+1/2-l) \\
&= \Gamma(1/2) \frac{\prod_{j=1}^k N(N-1) \dots (N-n_j+1)}{(Nk)(Nk-1) \dots (Nk-n+1)} \frac{\prod_{l=1}^{Nk-n} (Nk-n+1/2-l)}{(Nk-n)(Nk-n-1) \dots 3 \cdot 2 \cdot 1} \\
&= \Gamma(1/2) \frac{\prod_{j=1}^k \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{n_j-1}{N}\right)}{k \left(k - \frac{1}{N}\right) \left(k - \frac{2}{N}\right) \dots \left(k - \frac{n-1}{N}\right)} \frac{\prod_{l=1}^{Nk-n} (Nk-n+1/2-l)}{\prod_{l=1}^{Nk-n} (Nk-n+1-l)}
\end{aligned}$$

$$= \Gamma(1/2) \frac{\prod_{j=1}^k \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{n_j - 1}{N}\right)}{k \left(k - \frac{1}{N}\right) \left(k - \frac{2}{N}\right) \cdots \left(k - \frac{n - 1}{N}\right)} \prod_{l=1}^{Nk-n} \left(1 - \frac{1/2}{Nk - n + 1 - l}\right). \quad (2.12)$$

Como,

$$\frac{\prod_{j=1}^k \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{n_j - 1}{N}\right)}{k \left(k - \frac{1}{N}\right) \left(k - \frac{2}{N}\right) \cdots \left(k - \frac{n - 1}{N}\right)} \xrightarrow{N \rightarrow \infty} \frac{1}{k^n} \quad (2.13)$$

e  $1 - x \leq \exp(-x)$ , para todo  $x$  real, temos

$$\prod_{l=1}^{Nk-n} \left(1 - \frac{1/2}{Nk - n + 1 - l}\right) = \prod_{r=1}^{Nk-n} \left(1 - \frac{1}{2r}\right) \leq \prod_{r=1}^{Nk-n} \exp\left(-\frac{1}{2r}\right) = \exp\left(-\frac{1}{2} \sum_{r=1}^{Nk-n} \frac{1}{r}\right).$$

Então, uma vez que

$$\lim_{N \rightarrow \infty} \sum_{r=1}^{Nk-n} \frac{1}{r} = \sum_{r=1}^{\infty} \frac{1}{r} = +\infty,$$

temos

$$0 \leq \lim_{N \rightarrow \infty} \prod_{l=1}^{Nk-n} \left(1 - \frac{1/2}{Nk - n + 1 - l}\right) \leq \exp\left(-\frac{1}{2} \lim_{N \rightarrow \infty} \sum_{r=1}^{Nk-n} \frac{1}{r}\right) = 0. \quad (2.14)$$

O resultado segue de 2.12, 2.13 e 2.14. ■

O conhecimento destes limites não deixa dúvidas em relação ao valor de máximo encontrado da função em questão, nos permitindo atribuir este valor à estimativa procurada.

### 2.3.2 Solução por Continuidade

Uma outra possibilidade para se determinar as emv perfilada, condicional e integradas uniforme e de Jeffreys, aqui denotadas por  $\hat{N}^p$ ,  $\hat{N}^e$ ,  $\hat{N}^u$  e  $\hat{N}^d$ , respectivamente, é a suposição de que  $N$  seja uma variável contínua, como discutido na seção 2.3 para a determinação de  $\hat{N}$ .

**(I) EMV Perfilada**

A emv perfilada de  $N$ , supondo  $N$  uma variável contínua, é solução da equação

$$\frac{\partial \log K^{\mathcal{P}}(N|n_1, n_2, \dots, n_k)}{\partial N} = 0, \quad (2.15)$$

$N \geq n_0$ . Logo, de 2.3, temos

$$\begin{aligned} \log K^{\mathcal{P}}(N|n_1, n_2, \dots, n_k) &= \sum_{j=1}^k \log \binom{N}{n_j} + \log \left[ \frac{(Nk - n)^{Nk-n}}{(Nk)^{Nk}} \right] \\ &= \sum_{j=1}^k \log \left[ \frac{N!}{(N - n_j)! n_j!} \right] + (Nk - n) \log (Nk - n) - Nk \log (Nk) \\ &= k \log \Gamma(N + 1) - \sum_{j=1}^k \log \Gamma(N - n_j + 1) - \sum_{j=1}^k \log (n_j!) \\ &\quad + (Nk - n) \log (Nk - n) - Nk \log (Nk), \end{aligned}$$

$N \geq n_0$ , o que implica que  $\hat{N}^{\mathcal{P}}$  é solução da equação

$$k \Psi(N + 1) - \sum_{j=1}^k \Psi(N - n_j + 1) + k \log \left( 1 - \frac{n}{Nk} \right) = 0, \quad (2.16)$$

$N \geq n_0$ .

Notamos que a equação 2.16 é igual à equação 2.10, isto é, a emv de  $N$  coincide com a emv perfilada de  $N$ .

**(II) EMV Condicional**

Supondo  $N$  uma variável contínua, temos que a emv condicional de  $N$  satisfaz a equação

$$\frac{\partial \log L^c(N|n_1, n_2, \dots, n_k)}{\partial N} = 0, \quad (2.17)$$

$N \geq n_0$ .

Como, de 2.4, temos

$$\begin{aligned} \log L^c(N|n_1, n_2, \dots, n_k) &= \sum_{j=1}^k \log \binom{N}{n_j} - \log \binom{Nk}{n} \\ &= \sum_{j=1}^k \log \left[ \frac{N!}{(N - n_j)! n_j!} \right] - \log \left[ \frac{(Nk)!}{(Nk - n)! n!} \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^k \log(N!) - \sum_{j=1}^k \log[(N - n_j)!] - \sum_{j=1}^k \log(n_j!) - \log[(Nk)!] \\
&\quad + \log[(Nk - n)!] + \log(n!) \\
&= k \log \Gamma(N + 1) - \sum_{j=1}^k \log \Gamma(N - n_j + 1) - \sum_{j=1}^k \log(n_j!) - \log \Gamma(Nk + 1) \\
&\quad + \log \Gamma(Nk - n + 1) + \log(n!),
\end{aligned}$$

$N \geq n_0$ , então de 2.17 temos que  $\hat{N}^c$  é solução da equação

$$k \Psi(N + 1) - \sum_{j=1}^k \Psi(N - n_j + 1) - k \Psi(Nk + 1) + k \Psi(Nk - n + 1) = 0, \quad (2.18)$$

$N \geq n_0$ .

### (III) EMV Integrada Uniforme

Supondo  $N$  uma variável contínua, temos que a emv integrada uniforme de  $N$  é solução da equação

$$\frac{\partial \log K^u(N|n_1, n_2, \dots, n_k)}{\partial N} = 0, \quad (2.19)$$

$N \geq n_0$ . Como de 2.6 temos

$$\begin{aligned}
\log K^u(N|n_1, n_2, \dots, n_k) &= \sum_{j=1}^k \log \frac{N!}{(N - n_j)!} + \log \frac{(Nk - n)!}{(Nk + 1)!} \\
&= \sum_{j=1}^k \log(N!) - \sum_{j=1}^k \log[(N - n_j)!] + \log[(Nk - n + 1)!] - \log[(Nk + 1)!] \\
&= k \log \Gamma(N + 1) - \sum_{j=1}^k \log \Gamma(N - n_j + 1) + \log \Gamma(Nk - n + 1) - \log \Gamma(Nk + 2),
\end{aligned}$$

$N \geq n_0$ , segue de 2.19 que  $\hat{N}^u$  é solução da equação

$$k \Psi(N + 1) - \sum_{j=1}^k \Psi(N - n_j + 1) + k \Psi(Nk - n + 1) - k \Psi(Nk + 2) = 0, \quad (2.20)$$

$N \geq n_0$ .

**(IV) EMV Integrada de Jeffreys**

A emv integrada de Jeffreys de  $N$ , supondo  $N$  uma variável contínua, é solução da equação

$$\frac{\partial \log K^{\mathcal{J}}(N|n_1, n_2, \dots, n_k)}{\partial N} = 0, \quad (2.21)$$

$$N \geq n_0.$$

Como, de 2.7,

$$\begin{aligned} \log K^{\mathcal{J}}(N|n_1, n_2, \dots, n_k) &= \sum_{j=1}^k \log \frac{N!}{(N - n_j)!} + \log \frac{\Gamma(Nk - n + 1/2)}{\Gamma(Nk + 1)} \\ &= \sum_{j=1}^k \log(N!) - \sum_{j=1}^k \log[(N - n_j)!] + \log \Gamma(Nk - n + 1/2) - \log \Gamma(Nk + 1) \\ &= k \log \Gamma(N + 1) - \sum_{j=1}^k \log \Gamma(N - n_j + 1) + \log \Gamma(Nk - n + 1/2) - \log \Gamma(Nk + 1), \end{aligned}$$

$N \geq n_0$ , segue de 2.21 que  $\hat{N}^{\mathcal{J}}$  é solução da equação

$$k \Psi(N + 1) - \sum_{j=1}^k \Psi(N - n_j + 1) + k \Psi(Nk - n + 1/2) - k \Psi(Nk + 1) = 0, \quad (2.22)$$

$$N \geq n_0.$$

**2.3.3 Estudo da Performance do Modelo Binomial sem Remoção**

Com o objetivo de analisar o comportamento das emv de  $N$  com relação às funções de verossimilhança propostas, fizemos um estudo a partir de amostras simuladas deste modelo para  $N = 100$ ,  $p = 0,4$  e  $p = 0,8$  e diversos valores de  $k$ .

Primeiramente apresentamos as amostras simuladas nas tabelas 2.1 e 2.2 e, a partir dessas amostras, determinamos os respectivos valores das estatísticas  $n_0$  e  $n$ , de  $\hat{N}$  e das emv perfilada, condicional e integradas uniforme e de Jeffreys de  $N$ .

Por outro lado, a fim de obter maiores informações a respeito das estimativas, determinamos intervalos de confiança empíricos de 95% para  $N$  e erros quadráticos médios empíricos das estimativas, a partir da geração aleatória de 1000 amostras, entre outras informações.

(I) Estimativas de Máxima Verossimilhança de  $N$ Tabela 2.1: Amostras simuladas da distribuição  $Binomial(100; 0, 4)$ 


---

$k$	Amostra
2	$\mathbf{a}_1 = (46, 38)$
3	$\mathbf{a}_2 = (39, 40, 45)$
5	$\mathbf{a}_3 = (47, 37, 42, 41, 45)$
7	$\mathbf{a}_4 = (43, 39, 38, 29, 42, 36, 48)$
10	$\mathbf{a}_5 = (42, 44, 43, 30, 38, 33, 41, 42, 36, 42)$
15	$\mathbf{a}_6 = (40, 44, 32, 37, 40, 45, 42, 43, 40, 46, 44, 32, 39, 47, 38)$
30	$\mathbf{a}_7 = (37, 42, 36, 41, 43, 35, 40, 42, 33, 35, 43, 38, 41, 46, 39, 52, 37, 32, 41, 43, 42, 47, 50, 45, 46, 43, 37, 38, 52, 39)$
50	$\mathbf{a}_8 = (35, 44, 40, 41, 43, 45, 37, 40, 33, 49, 35, 43, 40, 38, 38, 32, 46, 37, 44, 42, 32, 39, 35, 47, 35, 44, 43, 33, 35, 41, 45, 44, 34, 36, 45, 35, 31, 35, 34, 41, 49, 38, 46, 38, 40, 30, 43, 33, 35, 33)$

---

Tabela 2.2: Amostras simuladas da distribuição *Binomial*(100; 0, 8)

---

$k$	Amostra
2	$\mathbf{a}_9 = (82, 79)$
3	$\mathbf{a}_{10} = (77, 74, 86)$
5	$\mathbf{a}_{11} = (77, 79, 75, 87, 78)$
7	$\mathbf{a}_{12} = (86, 78, 78, 81, 83, 85, 77)$
10	$\mathbf{a}_{13} = (79, 78, 88, 77, 84, 74, 79, 76, 83, 83)$
15	$\mathbf{a}_{14} = (88, 81, 79, 85, 79, 90, 78, 75, 72, 77, 81, 77, 75, 82, 82)$
30	$\mathbf{a}_{15} = (83, 75, 81, 81, 79, 71, 85, 80, 84, 82, 79, 82, 80, 84, 75, 79, 79, 77, 79, 84, 83, 80, 85, 87, 83, 76, 74, 81, 83, 81)$
50	$\mathbf{a}_{16} = (80, 79, 74, 73, 81, 72, 77, 80, 79, 77, 73, 77, 79, 72, 82, 87, 80, 85, 80, 82, 86, 75, 78, 85, 78, 87, 83, 75, 78, 85, 80, 83, 76, 79, 79, 72, 77, 85, 78, 82, 69, 78, 84, 82, 82, 74, 78, 88, 87, 71)$

---

Como mencionamos na seção 2.3, o valor de  $\hat{N}$  foi determinado a partir da solução da equação 2.10 através de uma rotina implementada em R (versão 2.8.1). O mesmo procedimento foi aplicado para a determinação de  $\hat{N}^{\mathcal{P}}$ ,  $\hat{N}^{\mathcal{C}}$ ,  $\hat{N}^{\mathcal{U}}$  e  $\hat{N}^{\mathcal{J}}$  como soluções das equações 2.16, 2.18, 2.20 e 2.22, respectivamente, apresentadas na seção 2.3.2.

Na seção 2.3.1 discutimos uma solução gráfica para as emv perfilada, condicional e integradas uniforme e de Jeffreys. A título de ilustração determinamos  $\hat{N}^{\mathcal{P}}$ ,  $\hat{N}^{\mathcal{C}}$ ,  $\hat{N}^{\mathcal{U}}$  e  $\hat{N}^{\mathcal{J}}$  para a amostra  $a_2$ , utilizando o aplicativo matemático Maple (versão 11.01).

Na figura 2.1(a) temos o gráfico do núcleo da função de verossimilhança perfilada para a amostra  $\mathbf{a}_2 = (39, 40, 45)$ , para  $N$  entre  $n_0 = 45$  e 500. Observamos que a função cresce, em seguida decresce e converge, pelo teorema 2.2, para  $4,8462 \cdot 10^{-264}$ , quando  $N \rightarrow \infty$ . Por outro lado, observando a figura 2.1(b) temos  $\hat{N}^{\mathcal{P}} = 51$ .

Na figura 2.2(a) temos o gráfico da função de verossimilhança condicional, cujo comportamento é análogo ao do núcleo da função de verossimilhança perfilada. Neste caso,  $\hat{N}^{\mathcal{C}} = 56$ .

O mesmo ocorre com as figuras 2.3(a) e 2.4(a) correspondentes aos gráficos dos logaritmos dos núcleos das funções de verossimilhança integradas uniforme e de Jeffreys, respectivamente. As estimativas obtidas são  $\hat{N}^{\mathcal{U}} = 52$  e  $\hat{N}^{\mathcal{J}} = 50$ .

Para os demais casos procedemos de maneira análoga e apresentamos os resultados nas tabelas 2.3 e 2.4, correspondentes aos dados das tabelas 2.1 e 2.2, respectivamente.

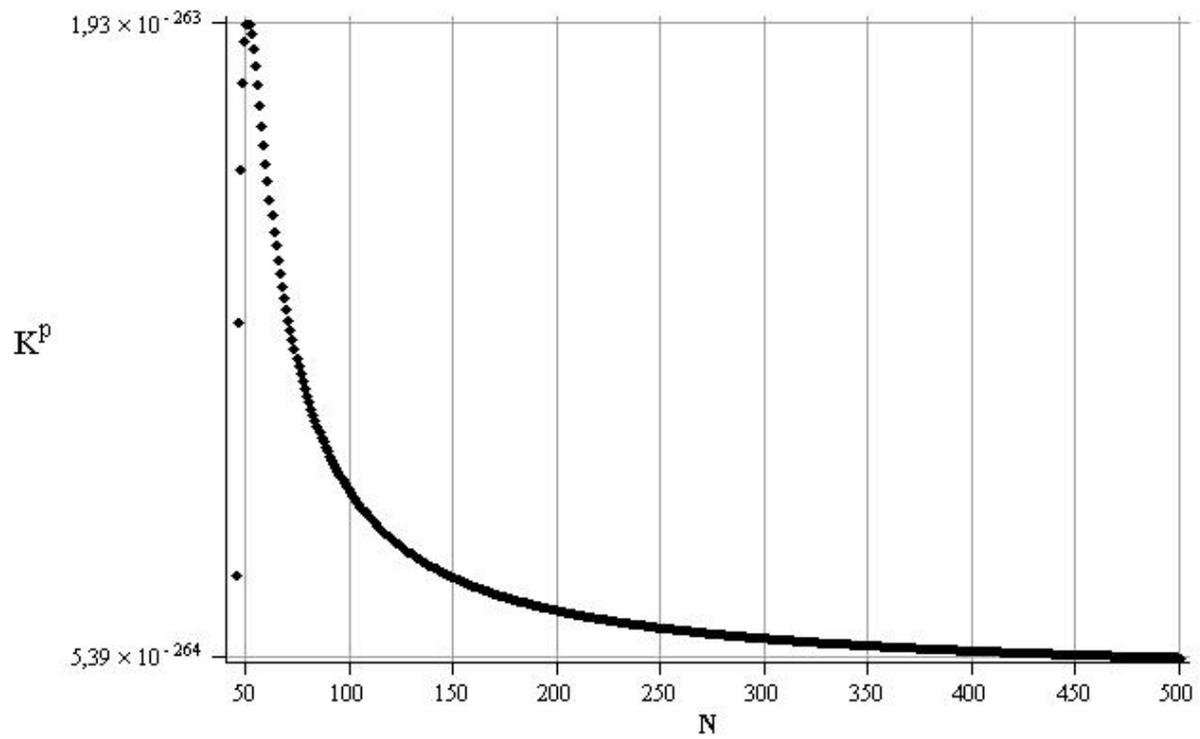
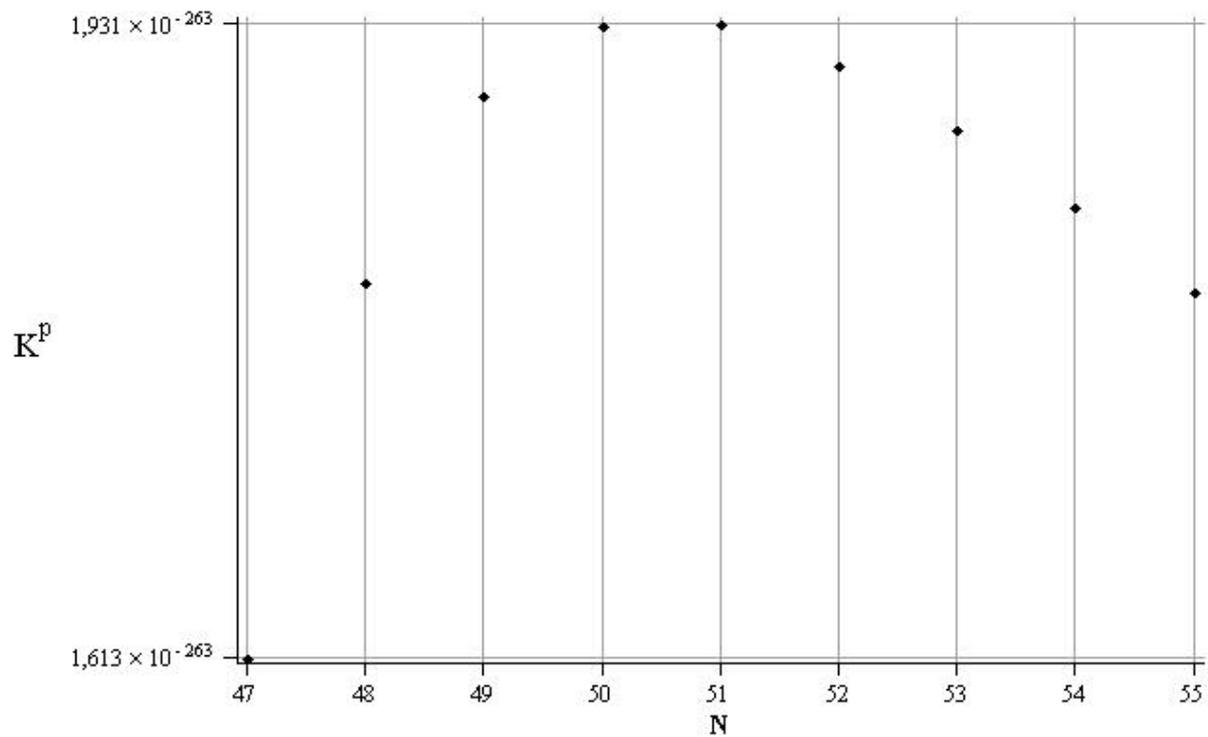
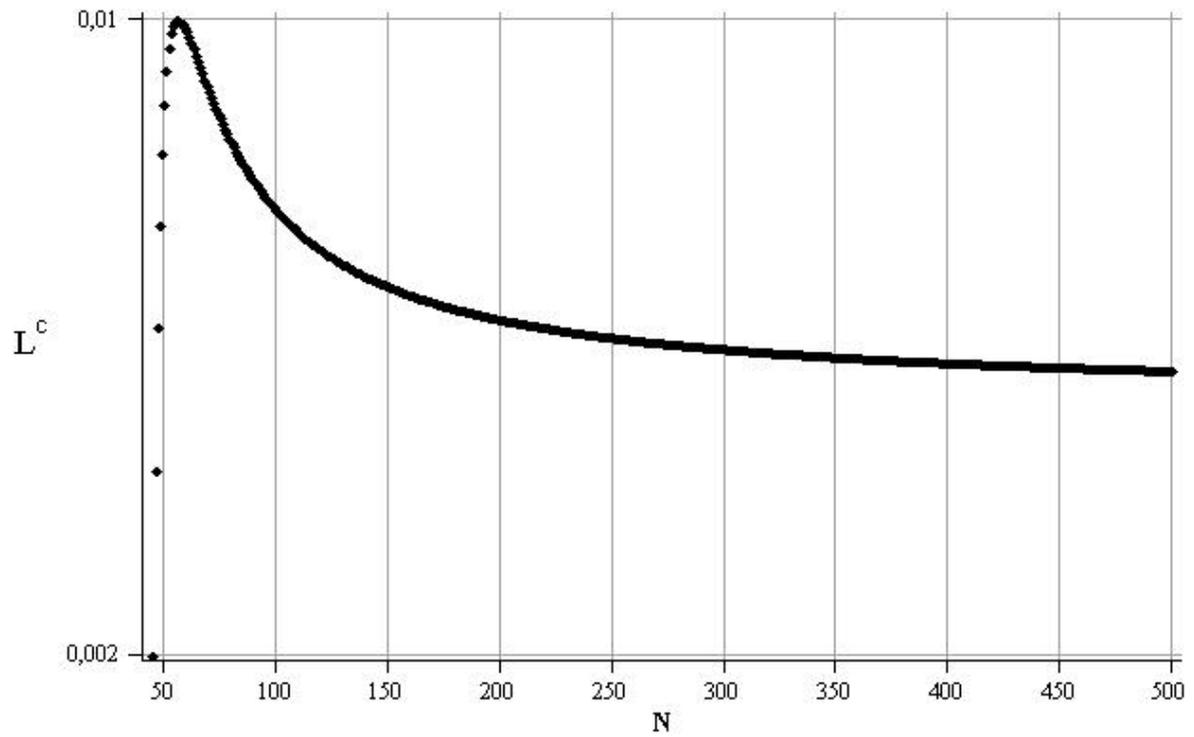
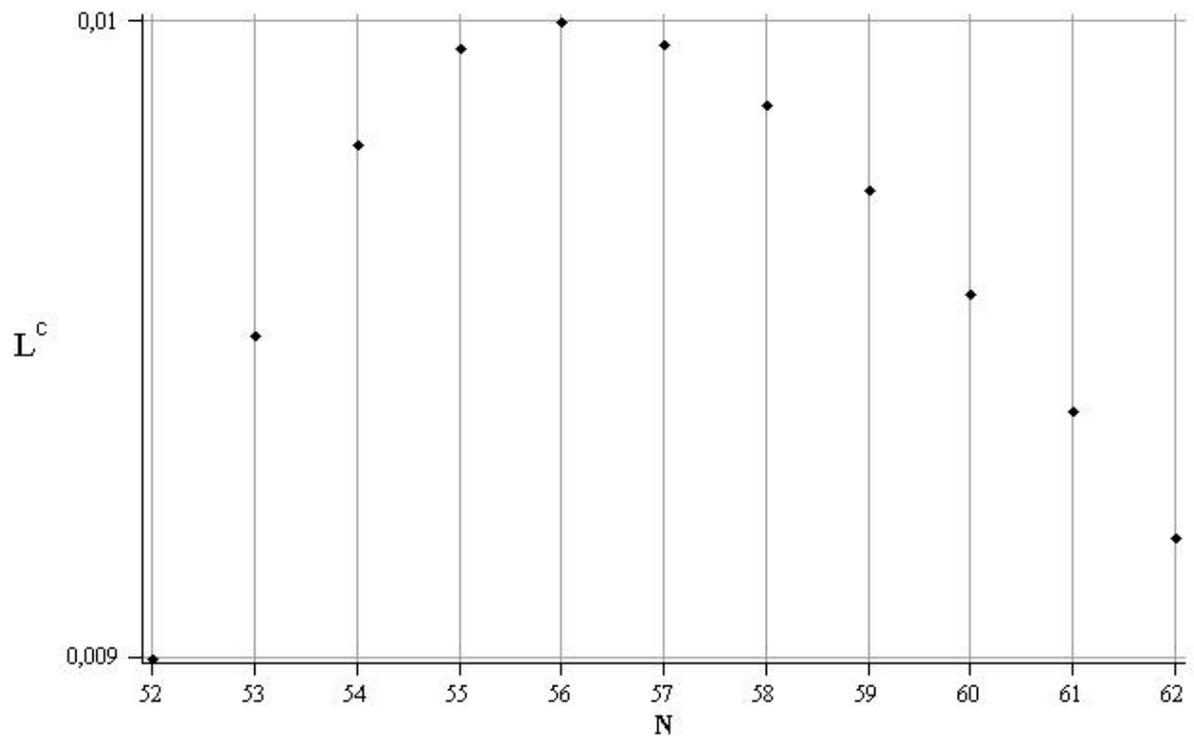
(a)  $45 < N < 500$ (b)  $47 < N < 55$ 

FIGURA 2.1: Gráficos do núcleo da função de verossimilhança perfilada para a amostra

 $a_2$

(a)  $45 < N < 500$ (b)  $52 < N < 62$ FIGURA 2.2: Gráficos da função de verossimilhança condicional para a amostra  $\mathbf{a}_2$

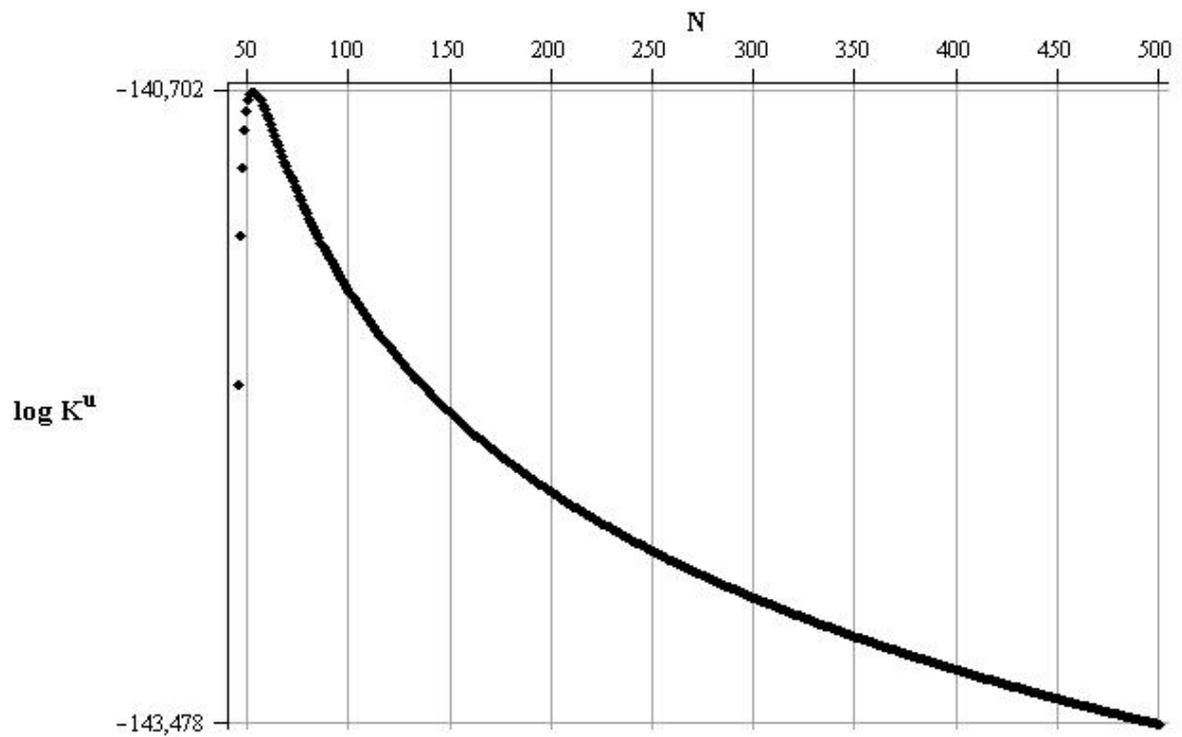
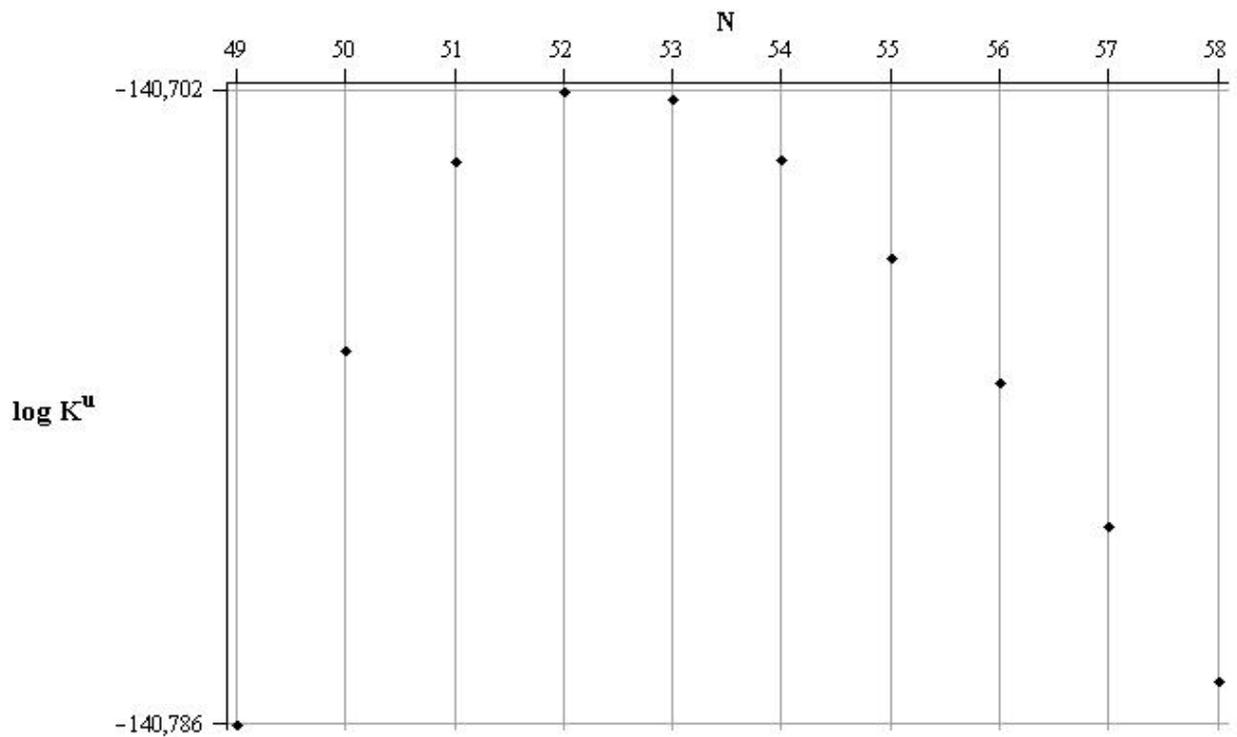
(a)  $45 < N < 500$ (b)  $49 < N < 58$ 

FIGURA 2.3: Gráficos do logaritmo do núcleo da função de verossimilhança integrada uniforme para a amostra  $\mathbf{a}_2$

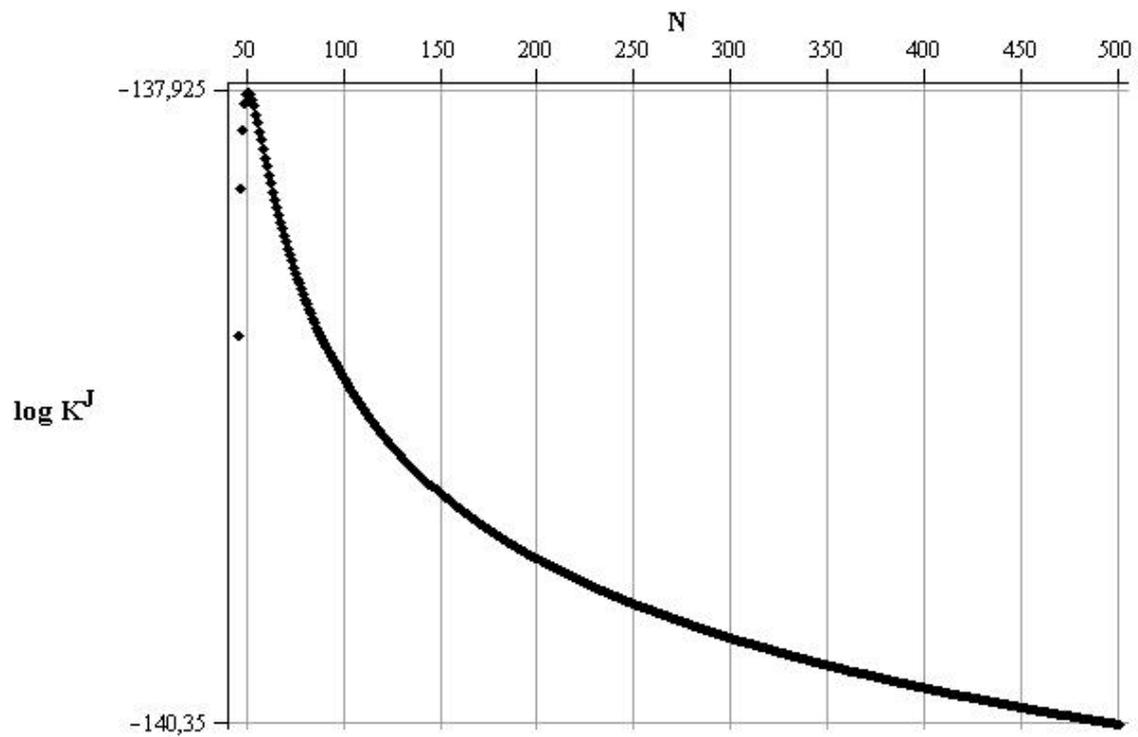
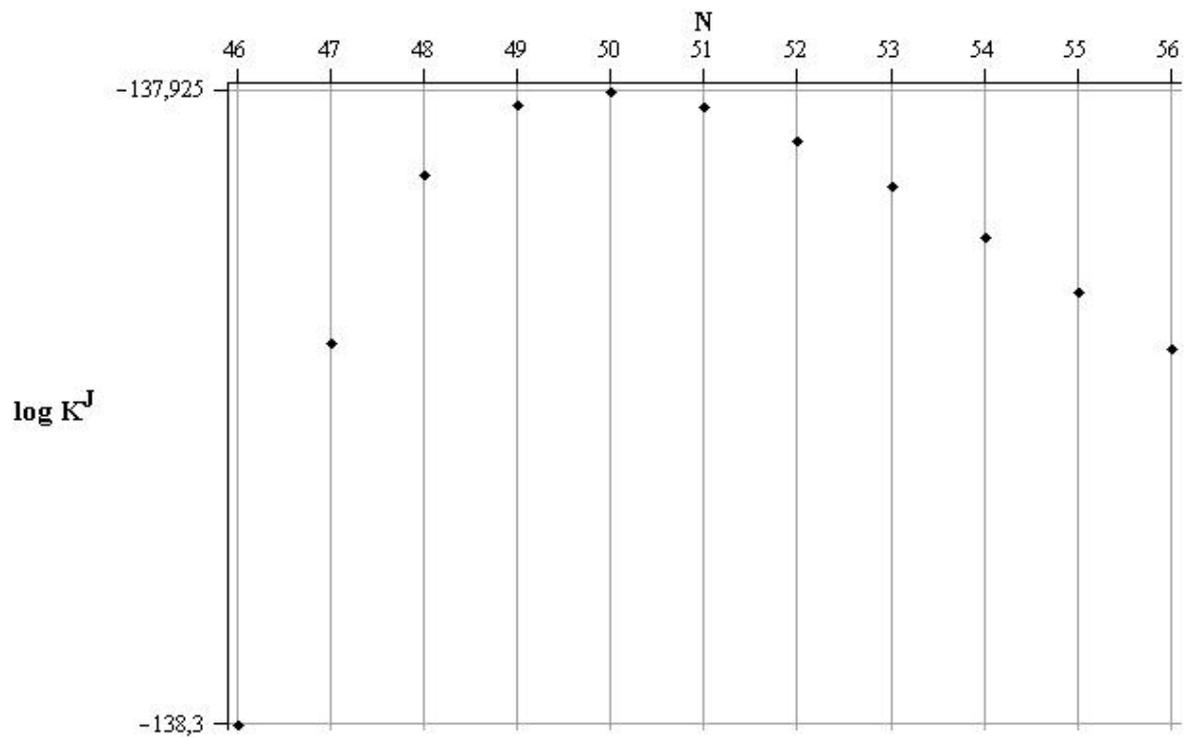
(a)  $45 < N < 500$ (b)  $46 < N < 56$ 

FIGURA 2.4: Gráficos do logaritmo do núcleo da função de verossimilhança integrada de Jeffreys para a amostra  $\mathbf{a}_2$

Tabela 2.3: Estimativas de máxima verossimilhança de  $N$  e de máxima verossimilhança perfilada, condicional e integradas uniforme e de Jeffreys para as amostras da Tabela 2.1

$k$	amostra	$n$	$n_0$	$\hat{N}$	$\hat{N}^{\mathcal{P}}$	$\hat{N}^c$	$\hat{N}^u$	$\hat{N}^j$
2	$\mathbf{a}_1$	84	46	67	67	174	65	61
3	$\mathbf{a}_2$	124	45	51	51	56	52	50
5	$\mathbf{a}_3$	212	47	58	58	64	59	57
7	$\mathbf{a}_4$	275	48	174	174	436	98	104
10	$\mathbf{a}_5$	391	44	74	74	83	70	69
15	$\mathbf{a}_6$	608	47	74	74	79	72	71
30	$\mathbf{a}_7$	1235	52	110	110	116	101	102
50	$\mathbf{a}_8$	1951	49	108	108	112	101	103

Tabela 2.4: Estimativas de máxima verossimilhança de  $N$  e de máxima verossimilhança perfilada, condicional e integradas uniforme e de Jeffreys para as amostras da Tabela 2.2

$k$	amostra	$n$	$n_0$	$\hat{N}$	$\hat{N}^{\mathcal{P}}$	$\hat{N}^c$	$\hat{N}^u$	$\hat{N}^j$
2	$\mathbf{a}_9$	161	82	82	82	85	84	82
3	$\mathbf{a}_{10}$	237	86	120	120	159	119	113
5	$\mathbf{a}_{11}$	396	87	104	104	112	106	103
7	$\mathbf{a}_{12}$	568	86	94	94	97	96	94
10	$\mathbf{a}_{13}$	801	88	102	102	105	103	102
15	$\mathbf{a}_{14}$	1201	90	113	113	117	114	112
30	$\mathbf{a}_{15}$	2412	87	94	94	95	95	94
50	$\mathbf{a}_{16}$	3963	88	108	108	109	109	108

**(II) Resumos das Medidas Descritivas Empíricas das Estimativas de  $N$** Tabela 2.5: Resumos das medidas descritivas empíricas das emv de  $N$  para 1000 amostras simuladas da distribuição  $Binomial(100; 0, 4)$ 

	EMV	Média	Erro Quadrático Médio	Intervalo de 95% de Confiança	Amplitude do IC
$k = 5$	Perfilada	91,00987	8203,587	(41,370)	329
	Condicional	107,9094	13559,96	(42,460)	418
	Uniforme	71,573	1548,373	(42,138)	96
	Jeffreys	74,012	2137,028	(42,168)	126
$k = 15$	Perfilada	112,0713	8832,197	(51,401)	350
	Condicional	122,8179	11332,52	(52,424)	372
	Uniforme	87,925	1094,855	(51,174)	123
	Jeffreys	92,65	1860,262	(51,217)	166
$k = 50$	Perfilada	107,6090	3374,083	(64,241)	177
	Condicional	111,2437	3525,097	(65,268)	203
	Uniforme	96,613	710,295	(64,168)	104
	Jeffreys	99,134	1042,146	(64,186)	122

Tabela 2.6: Resumos das medidas descritivas empíricas das emv de  $N$  para 1000 amostras simuladas da distribuição  $Binomial(100; 0, 8)$ 

	EMV	Média	Erro Quadrático Médio	Intervalo de 95% de Confiança	Amplitude do IC
	Perfilada	96,41	279,684	(81, 137)	56
$k = 5$	Condicional	102,4685	642,5706	(81, 166)	85
	Uniforme	97,38	183,532	(81, 130)	49
	Jeffreys	94,926	178,484	(81, 126)	45
	Perfilada	98,471	80,819	(86, 119)	33
$k = 15$	Condicional	100,133	98,559	(86, 123)	37
	Uniforme	99,051	77,391	(86, 119)	33
	Jeffreys	98,034	73,906	(86, 117)	31
	Perfilada	99,565	23,121	(91, 110)	19
$k = 50$	Condicional	100,03	24,32	(91, 111)	20
	Uniforme	99,767	23,075	(91, 110)	19
	Jeffreys	99,438	22,574	(91, 110)	19

### 2.3.4 Conclusões

As estimativas obtidas no estudo feito com dados simulados e descritas nas tabelas 2.3 e 2.4 nos permitiram concluir os seguintes fatos:

1. as emv de  $N$ ,  $\hat{N}$  e  $\hat{N}^{\mathcal{P}}$  resultaram iguais para todas as amostras, como observado na seção 2.3.2(I);
2. no caso das amostras  $a_4$ ,  $a_7$  e  $a_8$ , as emv integradas uniforme e de Jeffreys resultaram mais próximas do valor verdadeiro do parâmetro do que as demais;
3. todas as estimativas apresentaram, de um modo geral, melhores resultados quando o número de ocasiões,  $k$ , aumentou;

4. as emv de  $N$ ,  $\hat{N}$ ,  $\hat{N}^p$  e  $\hat{N}^d$ , resultaram aproximadamente iguais para o caso  $p = 0,8$ ;
5. as emv perfilada, condicional e integradas uniforme e de Jeffreys de  $N$  obtidas por solução por continuidade coincidiram com as estimativas por solução gráfica.

Baseados nos resumos das medidas descritivas empíricas das emv de  $N$  apresentados nas tabelas 2.5 e 2.6, concluímos que

1. como a equação de estimação de  $\hat{N}$  e  $\hat{N}^p$  coincidem, os resumos apresentados nas tabelas 2.5 e 2.6 para as emv perfilada são os mesmos para a emv usual;
2. ao comparar as informações da Tabela 2.5 com as da Tabela 2.6 notamos que as estimativas são mais próximas do valor verdadeiro de  $N$  no segundo caso; além disso, os intervalos de confiança neste caso também são mais satisfatórios; Estes fatos nos mostram que apesar da eliminação do parâmetro perturbador  $p$  da função de verossimilhança, ele ainda influencia fortemente os resultados;
3. de fato, ao analisar a média das estimativas, o eqm e os IC de 95% de  $N$  em ambas as tabelas 2.5 e 2.6, notamos que as estimativas melhoraram quando o número de ocasiões amostrais  $k$  aumentou e, finalmente,
4. como todos os intervalos de confiança contruídos contém o valor verdadeiro de  $N$ , comparamos os desempenhos dos diferentes estimadores através das amplitudes dos IC, dessa forma os emv integradas por apresentarem amplitudes menores são mais satisfatórios do que os emv perfilada e condicional.

Desde o início nossa expectativa era de que esse modelo não produzisse estimativas razoavelmente boas para o tamanho populacional, devido ao fato de que não existe a possibilidade de se obter um número de indivíduos distintos superior a  $n_0$ , mesmo para aquelas amostras em que este número realmente o supera. Este fato, sob o nosso ponto de vista, é o que torna as estimativas não precisas, como por exemplo em alguns casos apresentados nas tabelas 2.3 e 2.4 e nos casos dos extremos dos intervalos de confiança apresentados nas tabelas 2.5 e 2.6. Apesar disso, as médias das estimativas são mais próximas do valor verdadeiro do parâmetro, mesmo nestes casos.

Nos próximos capítulos propomos modelos que eliminam essa característica amostral, nos permitindo identificar os indivíduos distintos observados e obter estimativas mais precisas do tamanho populacional.

## Capítulo 3

# Modelo Binomial com Marcação

Vimos no capítulo 2 para o modelo binomial sem remoção que as emv do tamanho populacional com relação às funções de verossimilhança estudadas, em geral, subestimam o valor do parâmetro,  $N$ . Neste capítulo abordamos outro modelo estatístico que corrige esta deficiência. A essência desse novo modelo reside no fato de que os indivíduos selecionados, em cada ocasião de amostragem, serão reconhecidos se forem selecionados novamente em ocasiões posteriores. Outro ponto a destacar é o fato de considerarmos que as probabilidades de seleção dos indivíduos variam de acordo com as ocasiões de amostragem.

Mais precisamente, neste capítulo consideramos o processo amostral conhecido na literatura como captura-recaptura. Neste processo indivíduos são selecionados da população em  $k$  ocasiões de amostragem,  $k \geq 2$ , e, em cada uma delas, é feita a contagem e marcação dos indivíduos capturados pela primeira vez na ocasião e a contagem dos indivíduos recapturados (ou seja, aqueles capturados já marcados anteriormente). Em nosso estudo as estimativas do tamanho populacional dependem dos números de indivíduos capturados nas diversas ocasiões e do número de indivíduos distintos observados durante todo o processo.

Apresentamos novamente as funções de verossimilhança e de verossimilhança perfilada, condicional e integradas para o presente modelo e as respectivas estimativas do tamanho populacional.

### 3.1 Modelo Estatístico e Função de Verossimilhança

Denotemos por

- $p_j$  a probabilidade de qualquer indivíduo ser capturado na  $j$ -ésima ocasião,  $j = 1, 2, \dots, k$ , e  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ ;
- $n_j$  o número de indivíduos capturados na  $j$ -ésima ocasião;
- $n = \sum_{j=1}^k n_j$  o número total de indivíduos capturados durante todo o processo;
- $m_j$  o número de indivíduos marcados capturados (recapturados) na  $j$ -ésima ocasião ( $m_1 = 0$ );
- $d_j = n_j - m_j$  o número de indivíduos capturados pela primeira vez na  $j$ -ésima ocasião;
- $M_j = \sum_{i=1}^{j-1} d_i$  o número de indivíduos distintos marcados presentes na população imediatamente antes da  $j$ -ésima ocasião,  $j = 2, 3, \dots, k$  ( $M_1 = 0$ );
- $r = \sum_{j=1}^k d_j$  o número total de indivíduos distintos capturados durante todo o processo;
- $D = \{n_1, m_1; n_2, m_2; \dots; n_k, m_k\}$  o conjunto de dados amostrais.

Suponhamos que cada indivíduo da população seja capturado ou não, em cada ocasião, independentemente dos demais e das outras ocasiões. Logo, as amostras obtidas durante o processo são independentes e a distribuição de probabilidades de

$(n_1, m_1; n_2, m_2; \dots; n_k, m_k)$ , dados  $N$  e  $\mathbf{p}$ , é igual a

$$\begin{aligned}
 P(n_1, m_1; n_2, m_2; \dots; n_k, m_k | N, \mathbf{p}) &= \\
 &= P(n_1, m_1 | N, \mathbf{p}) P(n_2, m_2 | n_1, m_1, N, \mathbf{p}) \dots P(n_k, m_k | n_1, m_1; n_2, m_2; \dots; n_{k-1}, m_{k-1}, N, \mathbf{p}) \\
 &= \binom{N}{n_1} p_1^{n_1} (1-p_1)^{N-n_1} \binom{N-M_2}{n_2-m_2} p_2^{n_2-m_2} (1-p_2)^{N-M_2-n_2+m_2} \binom{M_2}{m_2} p_2^{m_2} (1-p_2)^{M_2-m_2} \\
 &\dots \binom{N-M_k}{n_k-m_k} p_k^{n_k-m_k} (1-p_k)^{N-M_k-n_k+m_k} \binom{M_k}{m_k} p_k^{m_k} (1-p_k)^{M_k-m_k}
 \end{aligned}$$

$$\begin{aligned}
&= \binom{N}{n_1} p_1^{n_1} (1-p_1)^{N-n_1} \binom{M_2}{m_2} \binom{N-M_2}{n_2-m_2} p_2^{n_2} (1-p_2)^{N-n_2} \dots \binom{M_k}{m_k} \binom{N-M_k}{n_k-m_k} p_k^{n_k} (1-p_k)^{N-n_k} \\
&= \prod_{j=2}^k \binom{M_j}{m_j} \prod_{j=1}^k \binom{N-M_j}{n_j-m_j} p_j^{n_j} (1-p_j)^{N-n_j} \\
&= \frac{N!}{(N-r)!} \prod_{j=2}^k \binom{M_j}{m_j} \prod_{j=1}^k \frac{1}{(n_j-m_j)!} p_j^{n_j} (1-p_j)^{N-n_j}, \tag{3.1}
\end{aligned}$$

onde a última igualdade é devido ao fato

$$\begin{aligned}
&\prod_{j=1}^k \binom{N-M_j}{n_j-m_j} = \\
&= \frac{N!}{(N-n_1)! n_1!} \frac{(N-M_2)!}{(N-M_2-n_2+m_2)! (n_2-m_2)!} \frac{(N-M_3)!}{(N-M_3-n_3+m_3)! (n_3-m_3)!} \\
&\dots \frac{(N-M_{k-1})!}{(N-M_{k-1}-n_{k-1}+m_{k-1})! (n_{k-1}-m_{k-1})!} \frac{(N-M_k)!}{(N-M_k-n_k+m_k)! (n_k-m_k)!} \\
&= \left[ \prod_{j=1}^k \frac{1}{(n_j-m_j)!} \right] \frac{N!}{(N-M_2)!} \frac{(N-M_2)!}{(N-M_3)!} \frac{(N-M_3)!}{(N-M_4)!} \dots \frac{(N-M_{k-1})!}{(N-M_k)!} \frac{(N-M_k)!}{(N-r)!} \\
&= \left[ \prod_{j=1}^k \frac{1}{(n_j-m_j)!} \right] \frac{N!}{(N-r)!}.
\end{aligned}$$

Portanto, de 3.1 temos que a função de verossimilhança é dada por

$$\begin{aligned}
L(N, \mathbf{p} | D) &= P(n_1, m_1; n_2, m_2; \dots; n_k, m_k | N, \mathbf{p}) \\
&= \frac{N!}{(N-r)!} \prod_{j=2}^k \binom{M_j}{m_j} \prod_{j=1}^k \frac{1}{d_j!} p_j^{n_j} (1-p_j)^{N-n_j}, \tag{3.2}
\end{aligned}$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ .

Então, o logaritmo natural da função de verossimilhança é dado por

$$\begin{aligned}
l(N, \mathbf{p} | D) &= \log \frac{N!}{(N-r)!} + \sum_{j=2}^k \log \binom{M_j}{m_j} + \sum_{j=1}^k \log \frac{1}{d_j!} \\
&\quad + \sum_{j=1}^k n_j \log p_j + \sum_{j=1}^k (N-n_j) \log(1-p_j), \tag{3.3}
\end{aligned}$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ , seu núcleo é

$$K(N, \mathbf{p} | D) = \frac{N!}{(N-r)!} \prod_{j=1}^k p_j^{n_j} (1-p_j)^{N-n_j}, \tag{3.4}$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ , e

$$\log K(N, \mathbf{p} | D) = \log \frac{N!}{(N-r)!} + \sum_{j=1}^k n_j \log p_j + \sum_{j=1}^k (N-n_j) \log(1-p_j), \tag{3.5}$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ .

## 3.2 Funções de Verossimilhança

Novamente construímos funções de verossimilhança segundo diferentes metodologias de eliminação de  $\mathbf{p}$ , como no capítulo 2, com o objetivo de se determinar funções apenas do parâmetro de interesse  $N$ .

### 3.2.1 Função de Verossimilhança Perfilada

A função de verossimilhança perfilada de  $N$ , definida no capítulo 2, é dada por

$$L^{\mathcal{P}}(N|D) = \sup_{\mathbf{p}} \left\{ L(N, \mathbf{p}|D) \right\} = L(N, \hat{\mathbf{p}}(N)|D),$$

onde  $\hat{\mathbf{p}}(N)$  é o ponto de máximo de  $L(N, \mathbf{p}|D)$ , para  $N$  fixado.

Como maximizar a função de verossimilhança é equivalente a maximizar o logaritmo de seu núcleo, então de 3.5 segue que, para  $j = 1, 2, \dots, k$ ,

$$\frac{\partial \log K(N, \mathbf{p}|D)}{\partial p_j} = \frac{n_j}{p_j} - \frac{N - n_j}{1 - p_j} = 0.$$

Assim,  $\hat{\mathbf{p}}(N) = \left( \frac{n_1}{N}, \frac{n_2}{N}, \dots, \frac{n_k}{N} \right)$  e, de 3.2, temos que a função de verossimilhança perfilada de  $N$  é tal que

$$\begin{aligned} L^{\mathcal{P}}(N|D) &= L(N, \hat{\mathbf{p}}(N)|D) \\ &= \frac{N!}{(N-r)!} \prod_{j=2}^k \binom{M_j}{m_j} \prod_{j=1}^k \frac{1}{d_j!} \left( \frac{n_j}{N} \right)^{n_j} \left( 1 - \frac{n_j}{N} \right)^{N-n_j} \\ &\propto K^{\mathcal{P}}(N|D), \end{aligned}$$

onde

$$K^{\mathcal{P}}(N|D) = \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{(N-n_j)^{N-n_j}}{N^N}, \quad (3.6)$$

$N \geq r$ , é o núcleo da função de verossimilhança perfilada,  $L^{\mathcal{P}}(N|D)$ .

### 3.2.2 Função de Verossimilhança Condicional

Como na seção 2.2.2, a função de verossimilhança condicional,  $L^c(N|D)$ , é construída a partir da fatoração da função de verossimilhança 3.2 da seguinte forma:

$$L(N, \mathbf{p}|D) = L_1(N|D) \cdot L_2(N, \mathbf{p}|n_1, n_2, \dots, n_k),$$

onde  $L_1(N|D)$  é a função de verossimilhança correspondente à distribuição de probabilidades de  $(m_1, m_2, \dots, m_k)$ , dados  $(n_1, n_2, \dots, n_k)$ ,  $N$  e  $\mathbf{p}$ , e  $L_2(N, \mathbf{p}|n_1, n_2, \dots, n_k)$  corresponde à distribuição de probabilidades de  $(n_1, n_2, \dots, n_k)$  dados  $N$  e  $\mathbf{p}$ .

De fato, como no teorema 2.1, temos de 3.2

$$\begin{aligned}
L_1(N|D) &= P(m_1, m_2, \dots, m_k | n_1, n_2, \dots, n_k, N, \mathbf{p}) \\
&= \frac{P(n_1, m_1; n_2, m_2; \dots; n_k, m_k | N, \mathbf{p})}{P(n_1, n_2, \dots, n_k | N, \mathbf{p})} \\
&= \frac{\frac{N!}{(N-r)!} \prod_{j=2}^k \binom{M_j}{m_j} \prod_{j=1}^k \frac{1}{d_j!} p_j^{n_j} (1-p_j)^{N-n_j}}{\prod_{j=1}^k \binom{N}{n_j} p_j^{n_j} (1-p_j)^{N-n_j}} \\
&= \frac{\frac{N!}{(N-r)!} \prod_{j=2}^k \binom{M_j}{m_j} \prod_{j=1}^k \frac{1}{d_j!}}{\prod_{j=1}^k \binom{N}{n_j}}, \tag{3.7}
\end{aligned}$$

$N \geq r$ . Por outro lado,

$$\begin{aligned}
L_2(N, \mathbf{p}|n_1, n_2, \dots, n_k) &= P(n_1, n_2, \dots, n_k | N, \mathbf{p}) \\
&= \prod_{j=1}^k \binom{N}{n_j} p_j^{n_j} (1-p_j)^{N-n_j}, \tag{3.8}
\end{aligned}$$

$N \geq \max \{n_1, n_2, \dots, n_k\}$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ , e o resultado segue de 3.2, 3.7 e 3.8.

A função de verossimilhança  $L_1(N|D)$  denomina-se função de verossimilhança condicional de  $N$  e será denotada por  $L^c(N|D)$ . Então,  $L^c(N|D) \propto K^c(N|D)$ , onde

$$K^c(N|D) = \frac{\binom{N}{r}}{\prod_{j=1}^k \binom{N}{n_j}}, \tag{3.9}$$

$N \geq r$ , é o núcleo de  $L_1(N|D)$ .

### 3.2.3 Funções de Verossimilhança Integradas: Uniforme, Jeffreys e Generalizada

Como na seção 2.2.3, apresentamos algumas funções de verossimilhança integradas, com respeito a certas funções peso para  $\mathbf{p}$ , com o objetivo de obter uma função de verossimilhança apenas em função de  $N$ .

#### (I) Função de Verossimilhança Integrada Uniforme

Neste caso, vamos supor que, dado  $N, p_1, p_2, \dots, p_k$  são independentes e identicamente distribuídos com distribuição de probabilidades *Uniforme*(0, 1), o que implica que a função de verossimilhança integrada uniforme para o modelo binomial com marcação, de acordo com 2.5 e 3.2 é dada por

$$\begin{aligned}
 L^u(N|D) &= \int_{(0,1)^k} L(N, \mathbf{p}|D) \, d\mathbf{p} \\
 &\propto \frac{N!}{(N-r)!} \prod_{j=1}^k \int_0^1 p_j^{n_j} (1-p_j)^{N-n_j} \, dp_j \\
 &= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{\Gamma(n_j+1) \Gamma(N-n_j+1)}{\Gamma(N+2)} \\
 &= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{n_j! (N-n_j)!}{(N+1)!} \\
 &\propto K^u(N|D),
 \end{aligned}$$

onde

$$K^u(N|D) = \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{(N-n_j)!}{(N+1)!}, \quad (3.10)$$

$N \geq r$ , é o núcleo de  $L^u(N|D)$ .

#### (II) Função de Verossimilhança Integrada de Jeffreys

Novamente consideramos a priori de Jeffreys como função peso para  $\mathbf{p}$ .

Para a determinação da priori de Jeffreys para  $\mathbf{p}$  vamos utilizar o mesmo raciocínio da definição 2.4, com a definição estendida para o vetor  $\mathbf{p}$ .

**Definição 3.1** A distribuição à priori de Jeffreys para  $\mathbf{p}$ , dado  $N$ , é dada por

$$\pi^{\mathcal{J}}(\mathbf{p}|N) \propto \left[ \det I_D(N, \mathbf{p}) \right]^{\frac{1}{2}},$$

onde  $I_D(N, \mathbf{p}) = \left( E \left[ - \frac{\partial^2 l(N, \mathbf{p}|D)}{\partial p_i \partial p_j} \middle| \mathbf{p} \right] \right)_{1 \leq i, j \leq k}$  é a matriz de informação esperada de Fisher.

Então, de 3.3, temos

$$\frac{\partial l(N, \mathbf{p}|D)}{\partial p_j} = \frac{n_j}{p_j} - \frac{N - n_j}{1 - p_j}, \quad j = 1, 2, \dots, k,$$

as derivadas de segunda ordem de  $l(N, \mathbf{p}|D)$  são

$$\begin{cases} \frac{\partial^2 l(N, \mathbf{p}|D)}{\partial p_i \partial p_j} = 0, & i \neq j, \\ \frac{\partial^2 l(N, \mathbf{p}|D)}{\partial p_j^2} = \frac{-n_j + 2n_j p_j - N p_j^2}{p_j^2 (1 - p_j)^2}, & j = 1, 2, \dots, k, \end{cases}$$

e

$$E \left[ - \frac{\partial^2 l(N, \mathbf{p}|D)}{\partial p_i \partial p_j} \middle| \mathbf{p} \right] = \begin{cases} 0, & i \neq j, \\ \frac{N p_j - 2 N p_j^2 + N p_j^2}{p_j^2 (1 - p_j)^2} = \frac{N}{p_j (1 - p_j)}, & i = j, \end{cases}$$

o que implica

$$I_D(N, \mathbf{p}) = \begin{bmatrix} \frac{N}{p_1 (1 - p_1)} & 0 & 0 & \dots & 0 \\ 0 & \frac{N}{p_2 (1 - p_2)} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{N}{p_k (1 - p_k)} \end{bmatrix}.$$

Portanto, da definição 3.1, temos

$$\begin{aligned} \pi^{\mathcal{J}}(\mathbf{p}|N) &\propto \left[ \det I_D(N, \mathbf{p}) \right]^{\frac{1}{2}} \\ &= \left[ N^k \prod_{j=1}^k p_j^{-1} (1 - p_j)^{-1} \right]^{\frac{1}{2}} \\ &\propto \prod_{j=1}^k p_j^{-\frac{1}{2}} (1 - p_j)^{-\frac{1}{2}}. \end{aligned}$$

Assim, de 2.5 e 3.2, a função de verossimilhança integrada de Jeffreys é dada por

$$\begin{aligned}
L^{\mathcal{J}}(N|D) &= \int_{(0,1)^k} L(N, \mathbf{p}|D) \pi^{\mathcal{J}}(\mathbf{p}|N) d\mathbf{p} \\
&\propto \frac{N!}{(N-r)!} \prod_{j=1}^k \left[ \int_0^1 p_j^{n_j-1/2} (1-p_j)^{N-n_j-1/2} dp_j \right] \\
&= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{\Gamma(n_j + 1/2) \Gamma(N - n_j + 1/2)}{\Gamma(N + 1)} \\
&\propto K^{\mathcal{J}}(N|D),
\end{aligned}$$

onde

$$K^{\mathcal{J}}(N|D) = \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{\Gamma(N - n_j + 1/2)}{N!}, \quad (3.11)$$

$N \geq r$ , é o núcleo da função de verossimilhança integrada de Jeffreys,  $L^{\mathcal{J}}(N|D)$ .

### (III) Função de Verossimilhança Integrada Generalizada

Para obter uma função de verossimilhança integrada generalizada para este modelo suponhamos que, dado  $N, p_1, p_2, \dots, p_k$  são independentes e identicamente distribuídos com  $p_1 \sim \text{Beta}(\alpha, \beta)$ ,  $\alpha > 0, \beta > 0$  e conhecidos.

Então,

$$\pi^{\mathcal{G}}(\mathbf{p}|N) \propto \prod_{i=1}^j p_i^{\alpha-1} (1-p_i)^{\beta-1}$$

e, de 2.5 e de 3.2, segue que

$$\begin{aligned}
L^{\mathcal{G}}(N|D) &= \int_{(0,1)^k} L(N, \mathbf{p}|D) \pi^{\mathcal{G}}(\mathbf{p}|N) d\mathbf{p} \\
&\propto \frac{N!}{(N-r)!} \prod_{j=1}^k \left[ \int_0^1 p_j^{n_j+\alpha-1} (1-p_j)^{N-n_j+\beta-1} dp_j \right] \\
&= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{\Gamma(n_j + \alpha) \Gamma(N - n_j + \beta)}{\Gamma(N + \alpha + \beta)} \\
&\propto K^{\mathcal{G}}(N|D),
\end{aligned}$$

onde

$$K^{\mathcal{G}}(N|D) = \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{\Gamma(N - n_j + \beta)}{\Gamma(N + \alpha + \beta)}, \quad (3.12)$$

$N \geq r$ , é o núcleo de  $L^S(N|D)$ .

Note que se  $\alpha = \beta = 1$  temos função de verossimilhança integrada uniforme e se  $\alpha = \beta = 1/2$  temos a função de verossimilhança integrada de Jeffreys.

### 3.3 Estimativas de Máxima Verossimilhança

Novamente vamos determinar as emv de  $N$  correspondentes às diferentes funções de verossimilhança apresentadas na seção anterior.

Como vimos na seção 2.3, o ponto  $(\hat{N}, \hat{\mathbf{p}})$  que maximiza a função de verossimilhança,  $L(N, \mathbf{p}|D)$ , supondo  $N$  uma variável contínua, é obtido através da solução do sistema

$$\begin{cases} \frac{\partial l(N, \mathbf{p}|D)}{\partial p_j} = 0, & j = 1, 2, \dots, k, \\ \frac{\partial l(N, \mathbf{p}|D)}{\partial N} = 0, \end{cases} \quad (3.13)$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ .

Segue, de 3.3, que

$$\begin{aligned} l(N, \mathbf{p}|D) &= \log \frac{N!}{(N-r)!} + \sum_{j=2}^k \log \binom{M_j}{m_j} + \sum_{j=1}^k \log \frac{1}{d_j!} + \sum_{j=1}^k n_j \log p_j + \sum_{j=1}^k (N-n_j) \log(1-p_j) \\ &= \log \Gamma(N+1) - \log \Gamma(N-r+1) + \sum_{j=2}^k \log \binom{M_j}{m_j} + \sum_{j=1}^k \log \frac{1}{d_j!} + \sum_{j=1}^k n_j \log p_j + \sum_{j=1}^k (N-n_j) \log(1-p_j), \end{aligned}$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ , e reescrevendo o sistema 3.13 como

$$\begin{cases} \frac{n_j}{p_j} - \frac{N-n_j}{1-p_j} = 0, & j = 1, 2, \dots, k, \\ \Psi(N+1) - \Psi(N-r+1) + \sum_{j=1}^k \log(1-p_j) = 0, \end{cases}$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ , onde  $\Psi(z) = \frac{d \log \Gamma(z)}{dz}$ ,  $z > 0$ , é a função digama,

concluimos que  $(\hat{N}, \hat{\mathbf{p}})$  é solução do sistema

$$\begin{cases} p_j = \frac{n_j}{N}, & j = 1, 2, \dots, k, \\ \Psi(N+1) - \Psi(N-r+1) + \sum_{j=1}^k \log(1-p_j) = 0, \end{cases}$$

$N \geq r$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, k$ , o que implica que  $\hat{N}$  é solução da equação

$$\Psi(N+1) - \Psi(N-r+1) + \sum_{j=1}^k \log\left(1 - \frac{n_j}{N}\right) = 0, \quad (3.14)$$

$N \geq r$ .

Novamente, a raiz  $\hat{N}$  da equação 3.14 foi obtida através de um programa implementado em R descrito no apêndice B.

Para o modelo estatístico apresentado neste capítulo procedemos como no capítulo 2, fazendo um estudo gráfico e outro supondo continuidade da variável  $N$  na determinação dos pontos de máximo das diferentes funções de verossimilhança, exceto para o caso da função de verossimilhança condicional que apresenta uma estimativa explícita de máxima verossimilhança. Segundo Leite *et al.* (1988) a emv condicional de  $N$ ,  $\hat{N}^c$ , existe e é dada por

$$\hat{N}^c = \begin{cases} r, & \text{se } r = n_0; \\ r + x_r - 1, & \text{se } n_0 < r < n; \\ \infty, & \text{se } r = n; \end{cases}$$

onde  $n_0 = \max\{n_1, n_2, \dots, n_k\}$ ,  $n = \sum_{j=1}^k n_j$  e

$x_r = \min\{x \text{ inteiro}, x \geq 1 : (r+x-n_1)(r+x-n_2) \dots (r+x-n_k) < x(r+x)^{k-1}\}$ .

### 3.3.1 Solução Gráfica

Para a determinação das emv perfilada e integradas uniforme e de Jeffreys procedemos da mesma maneira como no capítulo anterior, fazendo inicialmente um estudo gráfico. Na sequência determinamos os limites das funções de verossimilhança citadas, quando  $N \rightarrow \infty$ .

**Teorema 3.1** *O núcleo da função de verossimilhança perfilada, dada em 3.6, converge para 0, quando  $N \rightarrow \infty$ , se  $r < n$ , e converge para  $\exp(-n)$ , se  $r = n$ .*

**Demonstração.** De 3.6 temos

$$\begin{aligned}
K^{\mathcal{P}}(N|D) &= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{(N-n_j)^{N-n_j}}{N^N} \\
&= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{(N-n_j)^{N-n_j}}{N^{n_j} N^{N-n_j}} \\
&= N(N-1) \dots (N-r+1) \prod_{j=1}^k \frac{1}{N^{n_j}} \left( \frac{N-n_j}{N} \right)^{N-n_j} \\
&= \frac{N^r}{N^{\sum_{j=1}^k n_j}} \left( 1 - \frac{1}{N} \right) \dots \left( 1 - \frac{r-1}{N} \right) \prod_{j=1}^k \left( \frac{N-n_j}{N} \right)^N \left( \frac{N-n_j}{N} \right)^{-n_j} \\
&= \frac{N^r}{N^n} \left( 1 - \frac{1}{N} \right) \dots \left( 1 - \frac{r-1}{N} \right) \prod_{j=1}^k \left( 1 - \frac{n_j}{N} \right)^N \left( 1 - \frac{n_j}{N} \right)^{-n_j}
\end{aligned}$$

Como, para  $N \rightarrow \infty$ , temos que

$$\left( 1 - \frac{m}{N} \right) \rightarrow 1, \quad m < N, \quad \text{e} \quad \left( 1 - \frac{n_j}{N} \right)^N \rightarrow \exp(-n_j), \quad (3.15)$$

então,

$$\begin{cases} K^{\mathcal{P}}(N|D) \xrightarrow{N \rightarrow \infty} 0, & \text{se } r < n, \\ K^{\mathcal{P}}(N|D) \xrightarrow{N \rightarrow \infty} \exp(-n), & \text{se } r = n, \end{cases}$$

como queríamos demonstrar. ■

**Teorema 3.2** *O núcleo da função de verossimilhança integrada uniforme, dada em 3.10, converge para 0, quando  $N \rightarrow \infty$ .*

**Demonstração.** De 3.10 temos

$$\begin{aligned}
K^{\mathcal{U}}(N|D) &= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{(N-n_j)!}{(N+1)!} \\
&= N(N-1) \dots (N-r+1) \prod_{j=1}^k \frac{1}{(N+1)N(N-1) \dots (N-n_j+1)}
\end{aligned}$$

$$\begin{aligned}
&= \frac{N^r}{(N+1)^k} \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{r-1}{N}\right) \prod_{j=1}^k \frac{1}{N^{n_j} \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n_j-1}{N}\right)} \\
&= \frac{N^r}{N^n (N+1)^k} \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{r-1}{N}\right) \prod_{j=1}^k \frac{1}{\left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n_j-1}{N}\right)}.
\end{aligned}$$

Logo, de 3.15 e do fato que

$$\frac{N^r}{N^n (N+1)^k} \xrightarrow{N \rightarrow \infty} 0,$$

segue o resultado. ■

**Teorema 3.3** *O núcleo da função de verossimilhança integrada de Jeffreys, dada em 3.11, converge para 0, quando  $N \rightarrow \infty$ .*

**Demonstração.** De 3.11 temos

$$\begin{aligned}
K^{\mathcal{J}}(N|D) &= \frac{N!}{(N-r)!} \prod_{j=1}^k \frac{\Gamma(N - n_j + 1/2)}{N!} \\
&= N(N-1) \cdots (N-r+1) \prod_{j=1}^k \frac{\left[ \prod_{l=1}^{N-n_j} (N - n_j + 1/2 - l) \right] \Gamma(1/2)}{N(N-1) \cdots (N - n_j + 1) (N - n_j) (N - n_j - 1) \cdots 2 \cdot 1} \\
&= \Gamma(1/2)^k N^r \prod_{m=1}^{r-1} \left(1 - \frac{m}{N}\right) \prod_{j=1}^k \left[ \frac{1}{N(N-1) \cdots (N - n_j + 1)} \prod_{l=1}^{N-n_j} \frac{(N - n_j + 1 - 1/2 - l)}{(N - n_j + 1 - l)} \right] \\
&= \Gamma(1/2)^k \frac{N^r}{N^n} \prod_{m=1}^{r-1} \left(1 - \frac{m}{N}\right) \prod_{j=1}^k \left[ \frac{1}{\left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n_j-1}{N}\right)} \prod_{l=1}^{N-n_j} \left(1 - \frac{1/2}{N - n_j + 1 - l}\right) \right]
\end{aligned}$$

Portanto, de 3.15 e 2.14, segue o resultado. ■

### 3.3.2 Solução por Continuidade

Como na seção 2.3.2, podemos determinar as emv perfilada e integradas uniforme e de Jeffreys através da suposição de que  $N$  seja uma variável contínua.

**(I) EMV Perfilada**

A emv perfilada de  $N$ , ao supor  $N$  uma variável contínua, é raiz da equação

$$\frac{\partial \log K^{\mathcal{P}}(N|D)}{\partial N} = 0, \quad (3.16)$$

$N \geq r$ .

Como, de 3.6, temos

$$\begin{aligned} \log K^{\mathcal{P}}(N|D) &= \log \frac{N!}{(N-r)!} + \sum_{j=1}^k \log \left[ \frac{(N-n_j)^{N-n_j}}{N^N} \right] \\ &= \log \Gamma(N+1) - \log \Gamma(N-r+1) + \sum_{j=1}^k (N-n_j) \log(N-n_j) - \sum_{j=1}^k N \log N, \end{aligned}$$

$N \geq r$ , então, de 3.16, segue que  $\hat{N}^{\mathcal{P}}$  é solução da equação

$$\Psi(N+1) - \Psi(N-r+1) + \sum_{j=1}^k \log(N-n_j) - k \log N = 0 \quad (3.17)$$

$N \geq r$ .

Observamos que 3.17 é igual à equação 3.14, ou seja,  $\hat{N}$  coincide com  $\hat{N}^{\mathcal{P}}$ .

**(II) EMV Integrada Uniforme**

Supondo  $N$  uma variável contínua, temos que a emv integrada uniforme,  $\hat{N}^u$ , de  $N$  é solução da equação

$$\frac{\partial \log K^u(N|D)}{\partial N} = 0, \quad (3.18)$$

$N \geq r$ , e de 3.10 temos

$$\begin{aligned} \log K^u(N|D) &= \log \frac{N!}{(N-r)!} + \sum_{j=1}^k \log [(N-n_j)!] - \sum_{j=1}^k \log [(N+1)!] \\ &= \log \Gamma(N+1) - \log \Gamma(N-r+1) + \sum_{j=1}^k \log \Gamma(N-n_j+1) - \sum_{j=1}^k \log \Gamma(N+2), \end{aligned}$$

$N \geq r$ .

Logo, de 3.18, temos que  $\hat{N}^u$  é solução da equação

$$\Psi(N+1) - \Psi(N-r+1) + \sum_{j=1}^k \Psi(N-n_j+1) - k \Psi(N+2) = 0, \quad (3.19)$$

$N \geq r$ .

**(III) EMV Integrada de Jeffreys**

A emv integrada de Jeffreys do tamanho populacional, supondo  $N$  uma variável contínua, é solução da equação

$$\frac{\partial \log K^{\mathcal{J}}(N|D)}{\partial N} = 0, \quad (3.20)$$

$N \geq r$ , e de 3.11 temos

$$\begin{aligned} \log K^{\mathcal{J}}(N|D) &= \log \frac{N!}{(N-r)!} + \sum_{j=1}^k \log \Gamma(N - n_j + 1/2) - \sum_{j=1}^k \log(N!) \\ &= \log \Gamma(N+1) - \log \Gamma(N-r+1) + \sum_{j=1}^k \log \Gamma(N - n_j + 1/2) - k \log \Gamma(N+1), \end{aligned}$$

$N \geq r$ .

Portanto, de 3.20, temos que  $\hat{N}^{\mathcal{J}}$  é solução da equação

$$(1-k)\Psi(N+1) - \Psi(N-r+1) + \sum_{j=1}^k \Psi(N - n_j + 1/2) = 0, \quad (3.21)$$

$N \geq r$ .

**3.3.3 Estudo da Performance do Modelo Binomial com Marcação**

Nesta seção analisamos o comportamento das emv de  $N$  através de um estudo a partir de amostras simuladas para  $N = 100$ ,  $k = 5, 15$  e  $50$  ocasiões amostrais e simulamos valores de  $p_1, p_2, \dots, p_k$  distribuídos uniformemente em subintervalos diferentes do intervalo  $(0, 1)$  da seguinte forma: nos casos onde consideramos as probabilidades muito baixas, geramos os valores de  $p_1, p_2, \dots, p_k$  segundo a distribuição  $Uniforme(0; 0, 2)$  e para efeito de comparação com os resultados obtidos no capítulo 2, geramos os valores de  $p_1, p_2, \dots, p_k$  segundo a distribuição  $Uniforme(0, 3; 0, 5)$  e  $(0, 7; 0, 85)$ , respectivamente.

Novamente apresentamos nas tabelas 3.1, 3.2 e 3.3 as amostras simuladas, onde denotamos por  $\mathbf{m} = (m_1, m_2, \dots, m_k)$  e  $\mathbf{n} = (n_1, n_2, \dots, n_k)$ . Em seguida, determinamos os respectivos valores de  $\hat{N}$  e das emv perfilada,  $\hat{N}^{\mathcal{P}}$ , condicional,  $\hat{N}^{\mathcal{C}}$ , integrada uniforme,  $\hat{N}^{\mathcal{U}}$ , e integrada de Jeffreys,  $\hat{N}^{\mathcal{J}}$  para tais amostras.

Por fim, como discutido no capítulo anterior, apresentamos os resumos das medidas descritivas empíricas das emv de  $N$  que permitem uma melhor comparação entre as

estimativas obtidas a partir do modelo binomial sem remoção e do com marcação.

(I) Estimativas de Máxima Verossimilhança de  $N$

Tabela 3.1: Amostras simuladas com  $k = 5$  ocasiões

$\mathbf{a}_1$	$\mathbf{m} = (0, 0, 1, 2, 2)$ $r = 40$	$\mathbf{n} = (20, 2, 7, 3, 13)$ $n = 45$
$\mathbf{a}_2$	$\mathbf{m} = (0, 13, 29, 29, 21)$ $r = 89$	$\mathbf{n} = (37, 37, 43, 39, 25)$ $n = 181$
$\mathbf{a}_3$	$\mathbf{m} = (0, 55, 79, 85, 83)$ $r = 100$	$\mathbf{n} = (72, 75, 85, 86, 84)$ $n = 402$

A amostra  $\mathbf{a}_1$  foi gerada de  $N = 100$  e  $\mathbf{p} = (0, 17; 0, 02; 0, 03; 0, 02; 0, 19)$ ;  $\mathbf{a}_2$  de  $N = 100$  e  $\mathbf{p} = (0, 43; 0, 33; 0, 43; 0, 44; 0, 32)$  e  $\mathbf{a}_3$  de  $N = 100$  e  $\mathbf{p} = (0, 71; 0, 79; 0, 84; 0, 85; 0, 79)$ .

Tabela 3.2: Amostras simuladas com  $k = 15$  ocasiões

$\mathbf{a}_4$	$\mathbf{m} = (0, 2, 3, 4, 10, 1, 8, 3, 0, 4, 12, 12, 2, 4, 5)$ $\mathbf{n} = (13, 18, 11, 7, 20, 3, 17, 4, 0, 5, 20, 17, 2, 4, 7)$ $r = 78$ $n = 148$
$\mathbf{a}_5$	$\mathbf{m} = (0, 14, 24, 38, 34, 29, 37, 31, 30, 47, 42, 32, 34, 42, 48)$ $\mathbf{n} = (35, 40, 46, 48, 37, 31, 37, 31, 30, 48, 42, 32, 34, 42, 48)$ $r = 99$ $n = 581$
$\mathbf{a}_6$	$\mathbf{m} = (0, 61, 73, 69, 80, 73, 77, 81, 87, 78, 76, 80, 88, 77, 67)$ $\mathbf{n} = (73, 80, 78, 71, 80, 74, 77, 81, 87, 78, 76, 80, 88, 77, 67)$ $r = 100$ $n = 1167$

A amostra  $\mathbf{a}_4$  foi gerada de  $N = 100$  e  $\mathbf{p} = (0, 13; 0, 18; 0, 11; 0, 10; 0, 19; 0, 04; 0, 16; 0, 05; 0, 01; 0, 03; 0, 20; 0, 18; 0, 01; 0, 06; 0, 11)$ ;  $\mathbf{a}_5$  de  $N = 100$  e  $\mathbf{p} = (0, 32; 0, 45; 0, 45; 0, 36; 0, 42; 0, 30;$

0, 39; 0, 42; 0, 32; 0, 49; 0, 47; 0, 41; 0, 35; 0, 46; 0, 44) e  $\mathbf{a}_6$  de  $N = 100$  e  $\mathbf{p} = (0, 76; 0, 79; 0, 76; 0, 71; 0, 82; 0, 73; 0, 78; 0, 83; 0, 81; 0, 77; 0, 80; 0, 80; 0, 84; 0, 80; 0, 75)$ .

Tabela 3.3: Amostras simuladas com  $k = 50$  ocasiões

$\mathbf{a}_7$	$\mathbf{m} = (0, 1, 0, 1, 3, 1, 3, 7, 3, 10, 1, 14, 1, 10, 1, 12, 16, 14, 11, 11, 1, 15, 13, 17, 19,$ $1, 6, 16, 12, 15, 11, 16, 10, 14, 11, 8, 14, 6, 8, 2, 1, 6, 7, 3, 0, 12, 18, 25, 16, 0)$ $\mathbf{n} = (9, 20, 1, 1, 12, 2, 6, 20, 8, 17, 4, 19, 4, 15, 1, 17, 19, 16, 13, 11, 1, 16, 14, 17, 19,$ $1, 6, 18, 12, 15, 11, 16, 10, 14, 11, 8, 14, 7, 8, 2, 1, 6, 7, 3, 0, 12, 18, 25, 16, 0)$ $r = 78 \quad n = 148$
$\mathbf{a}_8$	$\mathbf{m} = (0, 16, 33, 18, 35, 37, 47, 41, 51, 29, 39, 30, 34, 38, 54, 51, 28, 50, 36, 38, 38, 32, 40, 53,$ $35, 37, 44, 31, 31, 45, 41, 38, 43, 37, 36, 34, 40, 28, 50, 34, 51, 41, 27, 41, 31, 42, 43, 32, 47, 37)$ $\mathbf{n} = (29, 46, 50, 30, 40, 40, 48, 42, 53, 29, 39, 30, 34, 38, 54, 51, 28, 50, 36, 38, 38, 32, 40, 53,$ $35, 37, 44, 31, 31, 45, 41, 38, 43, 37, 36, 34, 40, 28, 50, 34, 51, 41, 27, 41, 31, 42, 43, 32, 47, 37)$ $r = 100 \quad n = 1964$
$\mathbf{a}_9$	$\mathbf{m} = (0, 55, 70, 74, 75, 75, 73, 76, 78, 84, 82, 82, 79, 82, 73, 82, 78, 81, 68, 66, 73, 82, 88, 85,$ $80, 78, 82, 81, 71, 69, 75, 88, 81, 80, 76, 88, 72, 76, 85, 69, 71, 74, 79, 72, 69, 66, 72, 84, 84, 74)$ $\mathbf{n} = (79, 70, 74, 76, 75, 75, 73, 76, 78, 84, 82, 82, 79, 82, 73, 82, 78, 81, 68, 66, 73, 82, 88, 85,$ $80, 78, 82, 81, 71, 69, 75, 88, 81, 80, 76, 88, 72, 76, 85, 69, 71, 74, 79, 72, 69, 66, 72, 84, 84, 74)$ $r = 100 \quad n = 3857$

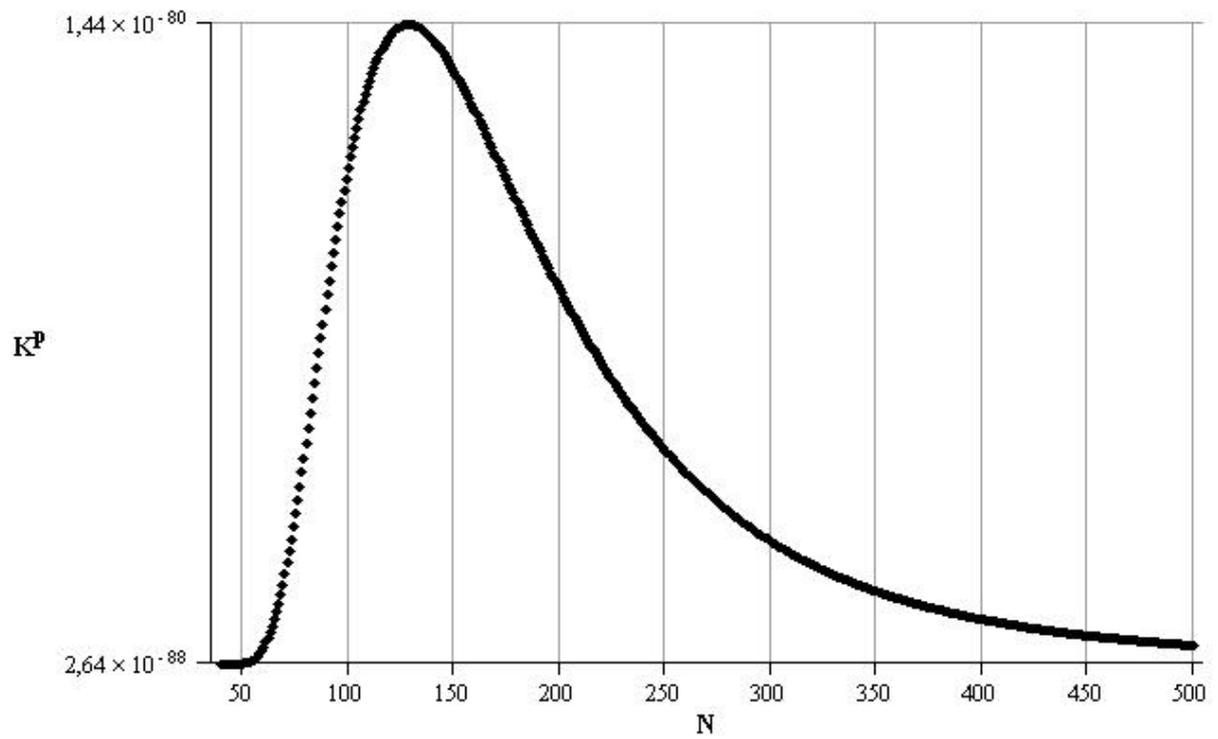
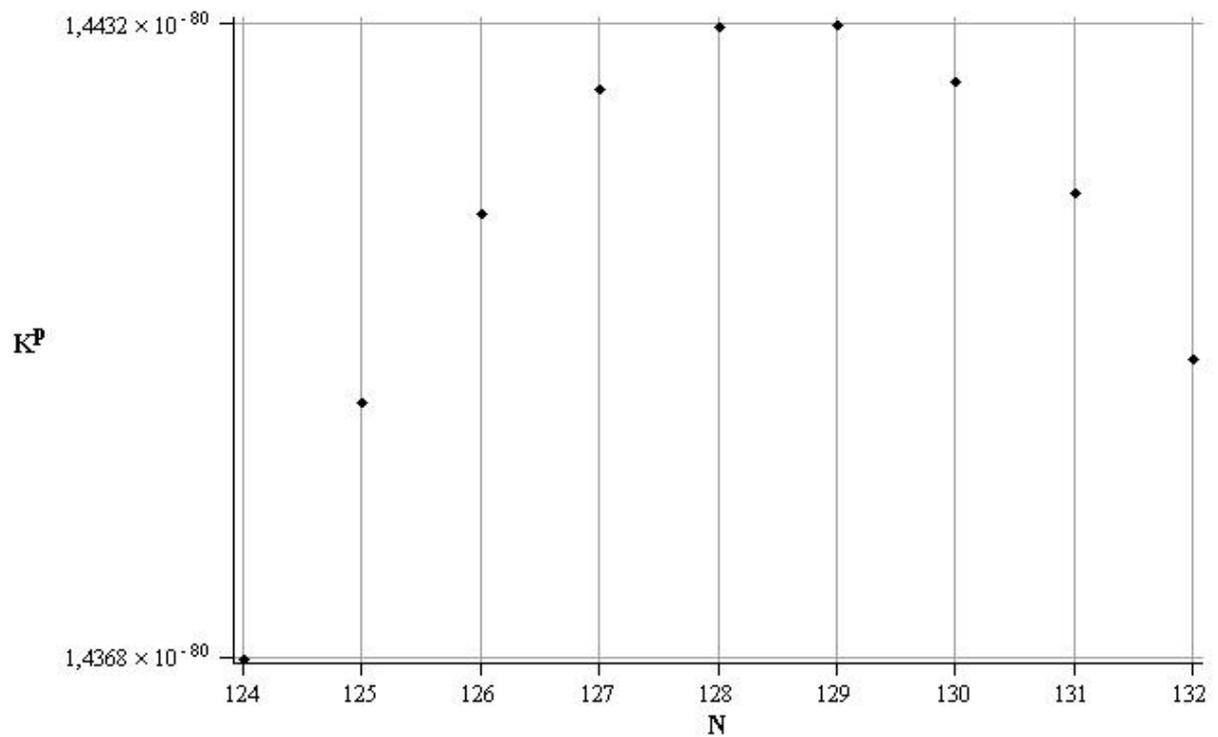
A amostra  $\mathbf{a}_7$  foi gerada de  $N = 100$  e  $\mathbf{p} = (0, 06; 0, 19; 0, 02; 0, 07; 0, 19; 0, 05; 0, 08; 0, 17; 0, 07; 0, 15; 0, 04; 0, 17; 0, 05; 0, 10; 0, 01; 0, 12; 0, 18; 0, 16; 0, 13; 0, 14; 0, 03; 0, 20; 0, 11; 0, 17; 0, 18; 0, 05; 0, 04; 0, 18; 0, 17; 0, 16; 0, 11; 0, 16; 0, 06; 0, 09; 0, 08; 0, 10; 0, 13; 0, 10; 0, 12; 0, 01; 0, 01; 0, 07; 0, 08; 0, 08; 0, 01; 0, 11; 0, 14; 0, 18; 0, 13; 0, 01)$ ;  $\mathbf{a}_8$  de  $N = 100$  e  $\mathbf{p} = (0, 35; 0, 43; 0, 45; 0, 34; 0, 36; 0, 47; 0, 49; 0, 32; 0, 48; 0, 33; 0, 36; 0, 30; 0, 37; 0, 42; 0, 49; 0, 41; 0, 30; 0, 47; 0, 35; 0, 39; 0, 37; 0, 38; 0, 40; 0, 49; 0, 36; 0, 44; 0, 49; 0, 37; 0, 36; 0, 39; 0, 42; 0, 37; 0, 45; 0, 38; 0, 39; 0, 39; 0, 45; 0, 34; 0, 45; 0, 39; 0, 45; 0, 35; 0, 32; 0, 40; 0, 43; 0, 43; 0, 47; 0, 36; 0, 49; 0, 44)$  e  $\mathbf{a}_9$  de  $N = 100$  e  $\mathbf{p} = (0, 80; 0, 70; 0, 78; 0, 75; 0, 75; 0, 75; 0, 82; 0, 76; 0, 83; 0, 81; 0, 78; 0, 82; 0, 79; 0, 83; 0, 76; 0, 82; 0, 81; 0, 84; 0, 71; 0, 71; 0, 71; 0, 79; 0, 83; 0, 81; 0, 78; 0, 73; 0, 77; 0, 79; 0, 78; 0, 72; 0, 71; 0, 83; 0, 81; 0, 84; 0, 73; 0, 85; 0, 73; 0, 76; 0, 83; 0, 73; 0, 75; 0, 73; 0, 73; 0, 75; 0, 74; 0, 75; 0, 73; 0, 84; 0, 81; 0, 81)$ .

Como no capítulo 2, os valores de  $\hat{N}$ , de  $\hat{N}^c$  e das estimativas obtidas supondo continuidade de  $N$  foram determinados através de um programa implementado em R e descrito no apêndice B.

Os gráficos foram construídos no aplicativo matemático Maple e ilustramos o procedimento através das figuras 3.1, 3.2 e 3.3 dos gráficos dos núcleos das funções de verossimilhança perfilada, integrada uniforme e integrada de Jeffreys para a amostra  $a_1$ , respectivamente. Todos os gráficos tem comportamento análogo, isto é, as funções crescem, decrescem e convergem para 0, quando  $N \rightarrow \infty$ , pelos teoremas 3.1, 3.2 e 3.3, respectivamente.

As estimativas do tamanho populacional são dadas observando as figuras 3.1(b), 3.2(b) e 3.3(b) que fornecem  $\hat{N}^p = 129$ ,  $\hat{N}^u = 83$  e  $\hat{N}^j = 97$ , respectivamente.

Procedemos de maneira análoga para os demais casos e apresentamos os resultados na Tabela 3.4.

(a)  $40 < N < 500$ (b)  $124 < N < 132$ FIGURA 3.1: Gráficos do núcleo da função de verossimilhança perfilada para a amostra  $a_1$

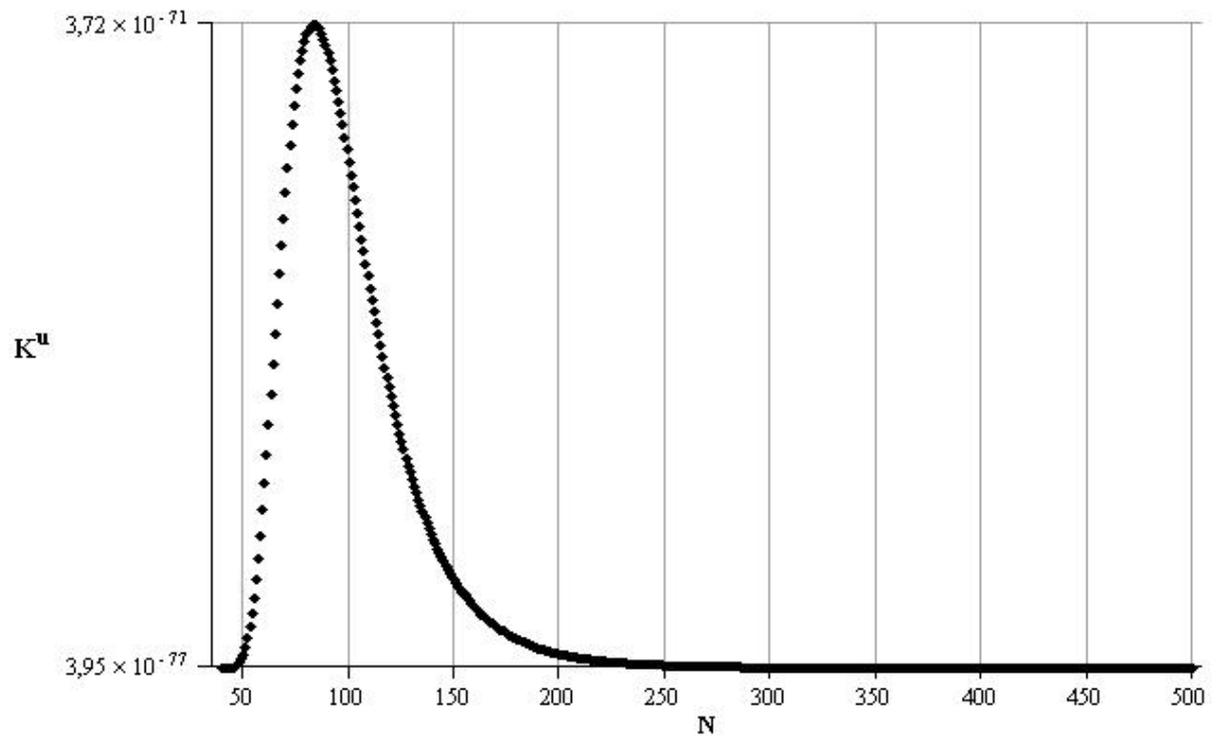
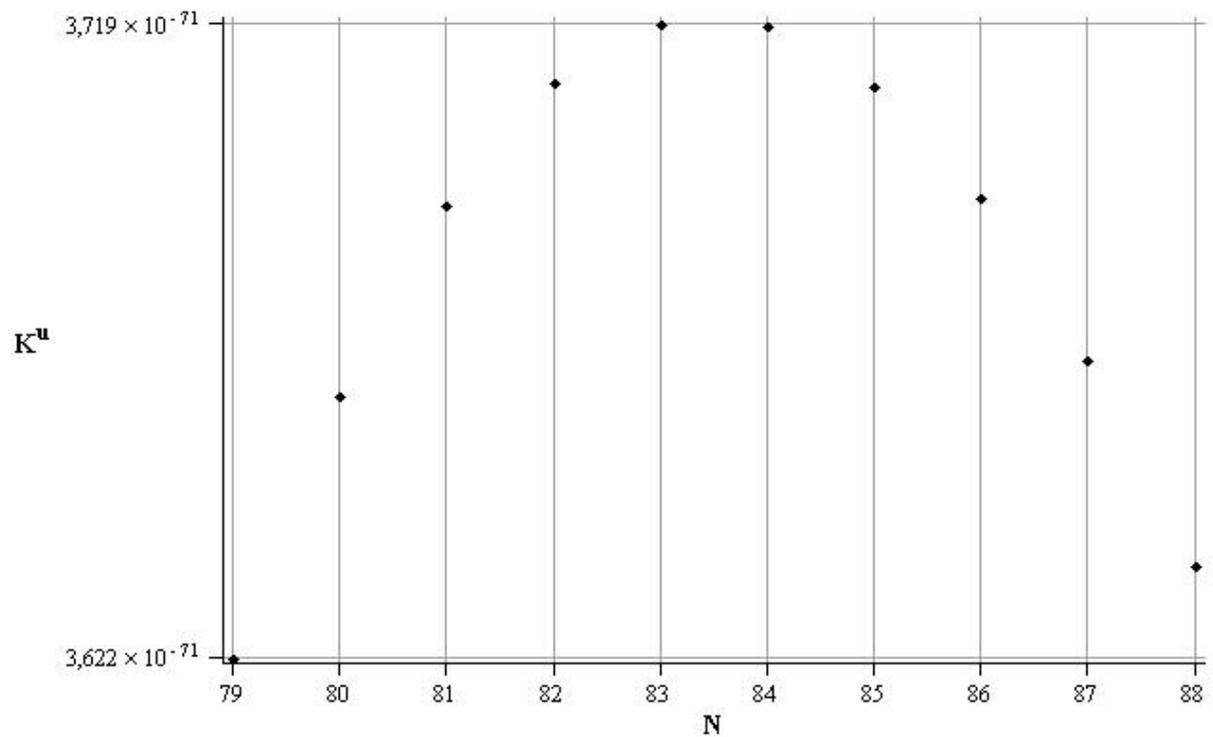
(a)  $40 < N < 500$ (b)  $79 < N < 88$ 

FIGURA 3.2: Gráficos do núcleo da função de verossimilhança integrada uniforme para a amostra  $\mathbf{a}_1$

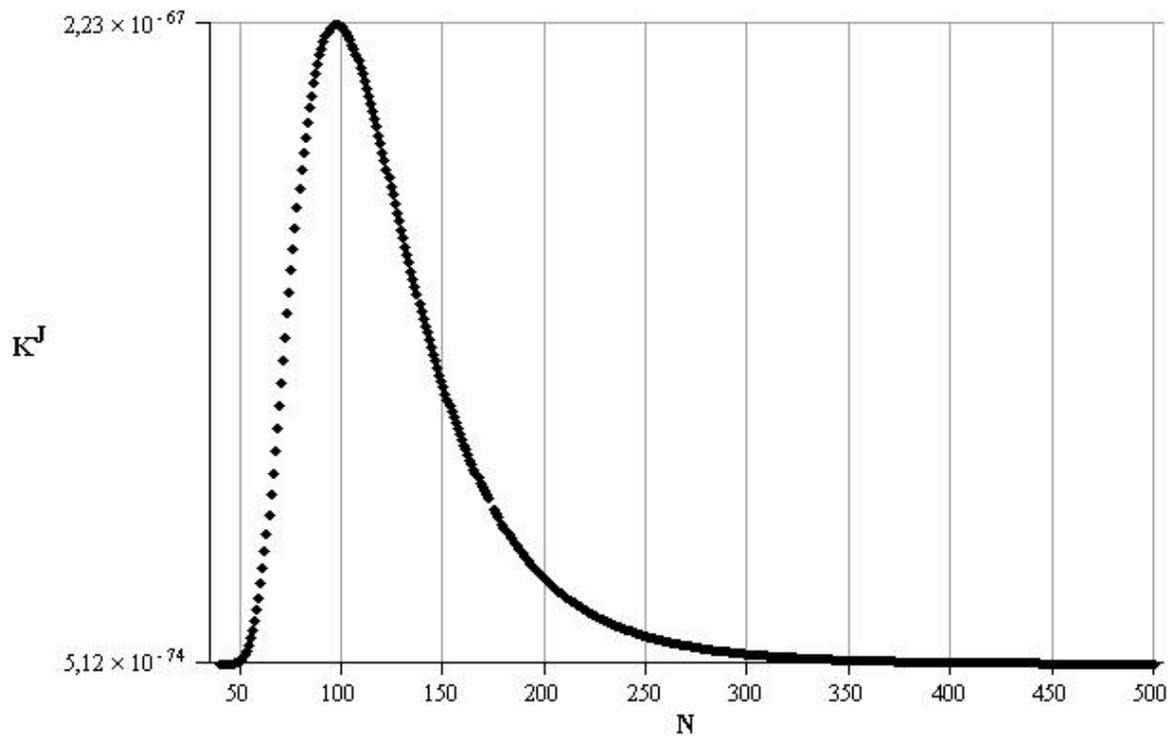
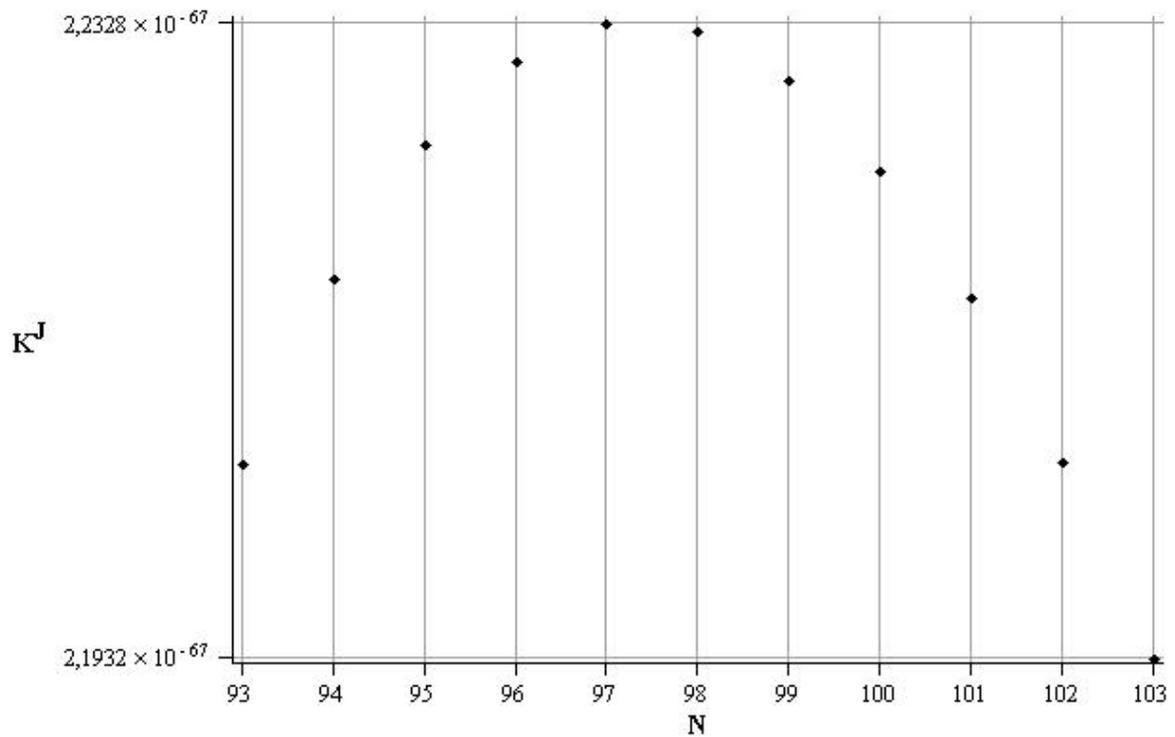
(a)  $40 < N < 500$ (b)  $93 < N < 103$ 

FIGURA 3.3: Gráficos do núcleo da função de verossimilhança integrada de Jeffreys para a amostra  $\mathbf{a}_1$

Tabela 3.4: Estimativas de máxima verossimilhança de  $N$  e de máxima verossimilhança perfilada, condicional e integradas uniforme e de Jeffreys para as amostras das tabelas 3.1, 3.2 e 3.3

$k$	amostra	$r$	$n$	$\hat{N}$	$\hat{N}^{\mathcal{P}}$	$\hat{N}^{\mathcal{C}}$	$\hat{N}^u$	$\hat{N}^{\mathcal{J}}$
5	$\mathbf{a}_1$	40	45	129	129	93	83	97
	$\mathbf{a}_2$	89	181	98	98	89	98	98
	$\mathbf{a}_3$	100	402	100	100	100	100	100
15	$\mathbf{a}_4$	78	148	95	95	78	91	92
	$\mathbf{a}_5$	99	581	99	99	99	99	99
	$\mathbf{a}_6$	100	1167	100	100	100	100	100
50	$\mathbf{a}_7$	100	523	100	100	100	100	100
	$\mathbf{a}_8$	100	1964	100	100	100	100	100
	$\mathbf{a}_9$	100	3857	100	100	100	100	100

**(II) Resumos das Medidas Descritivas Empíricas das Estimativas de  $N$** Tabela 3.5: Resumos das medidas descritivas empíricas das emv de  $N$  para 1000 amostras simuladas com  $p_j \approx 0,2$ ,  $j = 1, 2, \dots, k$ 

	EMV	Média	Erro Quadrático Médio	Intervalo de 95% de Confiança	Amplitude do IC
$k = 5$	Perfilada	108,279	2115,111	(58,228)	170
	Condiciona	111,498	2344,054	(60,237)	177
	Uniforme	71,428	1044,12	(47,105)	58
	Jeffreys	81,832	756,592	(51,131)	80
$k = 15$	Perfilada	99,149	33,745	(88,110)	22
	Condiciona	99,494	33,752	(88,111)	23
	Uniforme	95,108	49,614	(85,104)	19
	Jeffreys	96,789	39,117	(86,107)	21
$k = 50$	Perfilada	99,657	0,461	(98,100)	2
	Condiciona	99,657	0,461	(98,100)	2
	Uniforme	99,657	0,461	(98,100)	2
	Jeffreys	99,657	0,461	(98,100)	2

Tabela 3.6: Resumos das medidas descritivas empíricas das emv de  $N$  para 1000 amostras simuladas com  $p_j \approx 0,4$ ,  $j = 1, 2, \dots, k$ 

	EMV	Média	Erro Quadrático Médio	Intervalo de 95% de Confiança	Amplitude do IC
	Perfilada	99,371	14,925	(91, 107)	16
$k = 5$	Condicional	99,59	14,912	(91, 107)	16
	Uniforme	98,963	14,939	(91, 106)	15
	Jeffreys	99,05	14,996	(91, 107)	16
	Perfilada	99,967	0,035	(99, 100)	1
$k = 15$	Condicional	99,967	0,035	(99, 100)	1
	Uniforme	99,967	0,035	(99, 100)	1
	Jeffreys	99,967	0,035	(99, 100)	1
	Perfilada	100	0	(100, 100)	0
$k = 50$	Condicional	100	0	(100, 100)	0
	Uniforme	100	0	(100, 100)	0
	Jeffreys	100	0	(100, 100)	0

Tabela 3.7: Resumos das medidas descritivas empíricas das emv de  $N$  para 1000 amostras simuladas com  $p_j \approx 0,8$ ,  $j = 1, 2, \dots, k$ 

	EMV	Média	Erro Quadrático Médio	Intervalo de 95% de Confiança	Amplitude do IC
$k = 5$	Perfilada	99,93	0,074	(99, 100)	1
	Condiciona	99,93	0,074	(99, 100)	1
	Uniforme	99,93	0,074	(99, 100)	1
	Jeffreys	99,93	0,074	(99, 100)	1
$k = 15$	Perfilada	100	0	(100, 100)	0
	Condiciona	100	0	(100, 100)	0
	Uniforme	100	0	(100, 100)	0
	Jeffreys	100	0	(100, 100)	0
$k = 50$	Perfilada	100	0	(100, 100)	0
	Condiciona	100	0	(100, 100)	0
	Uniforme	100	0	(100, 100)	0
	Jeffreys	100	0	(100, 100)	0

### 3.3.4 Conclusões

As estimativas obtidas no estudo feito com dados simulados e descritas na Tabela 3.4 nos permitiram concluir que

1. as emv de  $N$ ,  $\hat{N}$  e  $\hat{N}^{\mathcal{P}}$  são iguais para todas as amostras, como observamos na seção 3.3.2(I);
2. todas as estimativas melhoraram quando  $k$  aumentou; particularmente para  $k = 50$  todas as emv resultaram iguais ao valor verdadeiro de  $N$ ;
3. todas as estimativas melhoraram quando as probabilidades de captura aumentaram, particularmente para as amostras  $a_3$  e  $a_6$  todas as estimativas resultaram iguais ao valor verdadeiro de  $N$ ;

4. as emv condicional,  $\hat{N}^c$ , tiveram desempenho pouco inferior em relação às outras nos casos das amostras  $a_2$  e  $a_4$ ;
5. todas as estimativas comportaram-se de maneira semelhante, exceto com relação à primeira amostra;
6. as emv perfilada e integradas uniforme e de Jeffreys de  $N$ , obtidas por solução por continuidade, coincidiram com as estimativas por solução gráfica para as amostras  $a_1$ ,  $a_2$  e  $a_4$ . Para as demais amostras, os núcleos das funções de verossimilhança  $K^p(N|D)$ ,  $K^u(N|D)$  e  $K^j(N|D)$  são funções decrescentes de  $N$  e, naturalmente, as emv são iguais à  $r$ .

Com base nas informações apresentadas nas tabelas 3.5, 3.6 e 3.7, concluímos

1. como a equação de estimação de  $\hat{N}$  e  $\hat{N}^p$  coincidem, os resumos apresentados nas tabelas 3.5, 3.6 e 3.7 para as emv perfilada são os mesmos para a emv usual.
2. comparando os intervalos de confiança apresentados nas tabelas 3.5 e 3.6, como todos eles contém o verdadeiro valor de  $N$ , concluímos que as emv integradas uniforme são mais satisfatórias que as emv integradas de Jeffreys, que por sua vez são mais satisfatórias que as emv perfilada e condicional, ao analisar as respectivas amplitudes dos IC.
3. em condições favoráveis, como probabilidades de captura altas e/ou número de ocasiões de amostragem elevado, todas as estimativas comportaram-se de maneira semelhante, fornecendo médias, eqm, IC e amplitude do IC iguais; além disso, quando o valor de  $k$  e as probabilidades de captura aumentaram, o eqm diminuiu bruscamente, tornando as estimativas mais precisas.
4. novamente, apesar da eliminação dos parâmetros perturbadores da função de verossimilhança, seus valores influenciam fortemente os resultados, uma vez que as estimativas são melhores quando as probabilidades de captura aumentaram.
5. ao comparar as informações contidas na Tabela 3.6 com as da Tabela 2.5, quando a probabilidade de captura em cada ocasião é aproximadamente igual à 0,4, e as da Tabela 3.7 com as da Tabela 2.6, quando a probabilidade de captura em cada ocasião

é aproximadamente igual à 0,8, notamos que o modelo binomial com marcação produziu estimativas mais próximas do valor verdadeiro do tamanho populacional do que o modelo binomial sem remoção.

Neste capítulo introduzimos um modelo estatístico, cujo objetivo era corrigir a má performance apresentada pelo modelo estudado no capítulo 2. Realmente, tal modelo correspondeu a nossa expectativa, produzindo estimativas satisfatórias do tamanho populacional. No próximo capítulo abordamos um outro modelo, onde consideramos os históricos de captura e recaptura dos indivíduos selecionados da população.

# Capítulo 4

## Modelo Multinomial

Apesar do modelo estudado no capítulo 3 ter produzido resultados satisfatórios, consideramos neste capítulo outro modelo que apresenta um pouco mais de informação amostral, além de contemplar, para cada indivíduo, uma dependência entre as várias ocasiões de amostragem. De fato, suponhamos agora que, em cada ocasião, todos os indivíduos capturados marcados (recapturados) e não marcados sejam marcados. Deste modo, no final do processo temos, para cada indivíduo selecionado pelo menos uma vez, o registro da sequência de capturas, recapturas e não recapturas desse indivíduo. Denominamos esta sequência de "históricos de captura" ou "trajetórias" do indivíduo e a estimação do tamanho populacional  $N$  baseia-se nos números de indivíduos que apresentam os diversos históricos de captura.

Como nos capítulos 2 e 3, apresentamos as funções de verossimilhança e de verossimilhança perfilada, restrita integrada (análoga à condicional dos modelos anteriores) e integradas, bem como as estimativas de máxima verossimilhança correspondentes.

### 4.1 Modelo Estatístico e a Função de Verossimilhança

Dado  $N$ , consideremos  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ik})$  o vetor aleatório associado ao  $i$ -ésimo indivíduo da população onde, para  $i = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, k$ , definimos

$$X_{ij} = \begin{cases} 1, & \text{se o } i\text{-ésimo indivíduo for capturado (recapturado) na } j\text{-ésima ocasião;} \\ 0, & \text{caso contrário.} \end{cases}$$

Suponhamos que cada indivíduo da população seja capturado (recapturado) ou não, em qualquer ocasião, independentemente dos demais indivíduos da população. Logo, os vetores aleatórios  $\mathbf{X}_i$ 's são independentes e assumem valores no conjunto

$$\Delta = \{\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_k) \mid \omega_j = 0 \text{ ou } 1, j = 1, 2, \dots, k\}, \text{ com } \#(\Delta) = 2^k = l.$$

Sejam  $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_l$  uma enumeração dos elementos de  $\Delta$ , onde  $\boldsymbol{\omega}_l = (0, 0, \dots, 0)$ . Então  $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_l$  representam os possíveis históricos de captura associados a qualquer um dos indivíduos da população e  $\boldsymbol{\omega}_l$  representa o único histórico não observável.

Supondo que todo indivíduo da população tenha a mesma probabilidade de apresentar um dado histórico de captura, temos que a probabilidade do  $i$ -ésimo indivíduo apresentar o histórico  $\boldsymbol{\omega}_r$  é

$$\begin{aligned} p_r &= P(\mathbf{X}_i = \boldsymbol{\omega}_r) \\ &= P((X_{i1}, X_{i2}, \dots, X_{ik}) = (\omega_{r1}, \omega_{r2}, \dots, \omega_{rk})) \\ &= P(X_{i1} = \omega_{r1}, X_{i2} = \omega_{r2}, \dots, X_{ik} = \omega_{rk}) \end{aligned}$$

$$i = 1, 2, \dots, N; \quad r = 1, 2, \dots, l.$$

Seja  $\mathcal{S} = \{\mathbf{x} = (x_1, x_2, \dots, x_{l-1}) \in (0, 1)^{l-1} : \sum_{j=1}^{l-1} x_j < 1, \}$  o simplex de dimensão  $l - 1$ . Denotemos por  $\mathbf{p} = (p_1, p_2, \dots, p_{l-1})$ ;  $n_r = \sum_{i=1}^N I_{\{\boldsymbol{\omega}_r\}}(\mathbf{X}_i)$  o número registrado de indivíduos com o histórico de captura  $\boldsymbol{\omega}_r$ ,  $r = 1, 2, \dots, l - 1$ ;  $n = \sum_{r=1}^{l-1} n_r$  o número de indivíduos distintos observados durante o processo e  $n_l = N - n$  o número de indivíduos com histórico de captura  $\boldsymbol{\omega}_l$  ou o número de indivíduos não observados.

Segundo as suposições feitas acima, temos que  $(n_1, n_2, \dots, n_{l-1})$ , dados  $N$  e  $\mathbf{p}$ , tem distribuição de probabilidades multinomial com parâmetros  $N$  e  $\mathbf{p}$ , o que implica que a função de verossimilhança é dada por

$$\begin{aligned} L(N, \mathbf{p} \mid n_1, n_2, \dots, n_{l-1}) &= P(n_1, n_2, \dots, n_{l-1} \mid N, \mathbf{p}) \\ &= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N - n)!} \prod_{r=1}^{l-1} p_r^{n_r} p_l^{N-n}, \quad (4.1) \end{aligned}$$

$$N \geq n, \mathbf{p} \in \mathcal{S} \text{ e } p_l = 1 - \sum_{r=1}^{l-1} p_r.$$

Então, o logaritmo natural, o núcleo e o logaritmo do núcleo da função de verossimilhança são dados, respectivamente, por

$$l(N, \mathbf{p} | n_1, n_2, \dots, n_{l-1}) = \log(N!) - \sum_{r=1}^{l-1} \log(n_r!) - \log[(N-n)!] \\ + \sum_{r=1}^{l-1} n_r \log p_r + (N-n) \log p_l; \quad (4.2)$$

$$K(N, \mathbf{p} | n_1, n_2, \dots, n_{l-1}) = \frac{N!}{(N-n)!} \prod_{r=1}^{l-1} p_r^{n_r} p_l^{N-n} \quad (4.3)$$

e

$$\log K(N, \mathbf{p} | n_1, n_2, \dots, n_{l-1}) = \log(N!) - \log[(N-n)!] + \sum_{r=1}^{l-1} n_r \log p_r + (N-n) \log p_l, \quad (4.4)$$

$$N \geq n, \mathbf{p} \in \mathcal{S} \text{ e } p_l = 1 - \sum_{r=1}^{l-1} p_r.$$

## 4.2 Funções de Verossimilhança

Novamente temos interesse em obter funções de verossimilhança apenas do parâmetro de interesse,  $N$ , e o fazemos a partir de diferentes metodologias de eliminação de  $\mathbf{p}$ .

### 4.2.1 Função de Verossimilhança Perfilada

A função de verossimilhança perfilada de  $N$ , definida no capítulo 2, é dada por

$$L^{\mathcal{P}}(N | n_1, n_2, \dots, n_{l-1}) = \sup_{\mathbf{p}} \left\{ L(N, \mathbf{p} | n_1, n_2, \dots, n_{l-1}) \right\} = L(N, \hat{\mathbf{p}}(N) | n_1, n_2, \dots, n_{l-1}),$$

onde  $\hat{\mathbf{p}}(N)$  é o ponto de máximo de  $L(N, \mathbf{p} | n_1, n_2, \dots, n_{l-1})$ , para  $N$  fixado.

Para determinar  $\hat{\mathbf{p}}(N)$  usamos a técnica de multiplicadores de Lagrange e, como maximizar a função de verossimilhança é equivalente a maximizar seu logaritmo, definimos

$$h(\mathbf{p}) = l(N, \mathbf{p} | n_1, n_2, \dots, n_{l-1}) + \lambda \left( 1 - \sum_{r=1}^l p_r \right).$$

Então, de 4.2,  $\hat{\mathbf{p}}(N)$  é a solução do sistema

$$\left\{ \begin{array}{l} \frac{\partial h(\mathbf{p})}{\partial p_r} = \frac{n_r}{p_r} - \lambda = 0, \quad r = 1, 2, \dots, l-1, \\ \frac{\partial h(\mathbf{p})}{\partial p_l} = \frac{N-n}{p_l} - \lambda = 0, \\ \frac{\partial h(\mathbf{p})}{\partial \lambda} = 1 - \sum_{r=1}^l p_r = 0, \end{array} \right. \implies \left\{ \begin{array}{l} p_r = \frac{n_r}{\lambda}, \quad r = 1, 2, \dots, l-1, \\ p_l = \frac{N-n}{\lambda}, \\ \sum_{r=1}^l p_r = 1, \end{array} \right.$$

o que implica

$$\begin{aligned} \sum_{r=1}^l p_r &= \sum_{r=1}^{l-1} \left( \frac{n_r}{\lambda} \right) + \frac{N-n}{\lambda} = 1 \\ &\Rightarrow \frac{1}{\lambda} \left( \underbrace{\sum_{r=1}^{l-1} n_r}_n + N - n \right) = 1 \\ &\Rightarrow \lambda = N. \end{aligned}$$

Logo,  $p_r = \frac{n_r}{N}$ ,  $r = 1, 2, \dots, l-1$ , e  $p_l = \frac{N-n}{N}$ , ou seja,

$$\hat{\mathbf{p}}(N) = \left( \frac{n_1}{N}, \frac{n_2}{N}, \dots, \frac{n_{l-1}}{N} \right)$$

e, de 4.1, a função de verossimilhança perfilada de  $N$  é dada por

$$\begin{aligned} L^{\mathcal{P}}(N|n_1, n_2, \dots, n_{l-1}) &= L(N, \hat{\mathbf{p}}(N)|n_1, n_2, \dots, n_{l-1}) \\ &= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N-n)!} \left[ \prod_{r=1}^{l-1} \left( \frac{n_r}{N} \right)^{n_r} \right] \left( \frac{N-n}{N} \right)^{N-n} \\ &\propto K^{\mathcal{P}}(N|n_1, n_2, \dots, n_{l-1}), \end{aligned}$$

onde

$$K^{\mathcal{P}}(N|n_1, n_2, \dots, n_{l-1}) = \frac{N!}{(N-n)!} \frac{(N-n)^{N-n}}{N^N}, \quad (4.5)$$

$N \geq n$ , é o núcleo de  $L^{\mathcal{P}}(N|n_1, n_2, \dots, n_{l-1})$ .

### 4.2.2 Função de Verossimilhança Restrita Integrada

A função de verossimilhança restrita, analogamente ao que vimos nas seções 2.2.2 e 3.2.2, é definida a partir da fatoração da função de verossimilhança de  $N$  e  $\mathbf{p}$ , que pode

ser escrita na forma

$$L(N, \mathbf{p} | n_1, n_2, \dots, n_{l-1}) = L_1(N, p_l | n_1, n_2, \dots, n_{l-1}) \cdot L_2(\mathbf{p} | n_1, n_2, \dots, n_{l-1}),$$

onde  $L_1(N, p_l | n_1, n_2, \dots, n_{l-1})$  é a função de verossimilhança correspondente à distribuição de probabilidades de  $n$ , dados  $N$  e  $\mathbf{p}$ , e  $L_2(\mathbf{p} | n_1, n_2, \dots, n_{l-1})$  é a função de verossimilhança correspondente à distribuição de probabilidades condicional de  $n_1, n_2, \dots, n_{l-1}$ , dados  $n$ ,  $N$  e  $\mathbf{p}$  (Bolsoni, 2002).

Realmente, como  $N - n$  é o número de indivíduos que apresentam o histórico  $\omega_l$ , temos que  $N - n \sim \text{Binomial}(N, p_l)$ , o que implica

$$\begin{aligned} P(n = n_0 | N, \mathbf{p}) &= P(N - n = N - n_0 | N, \mathbf{p}) \\ &= \binom{N}{N - n_0} p_l^{N - n_0} (1 - p_l)^{n_0} \\ &= \binom{N}{n_0} (1 - p_l)^{n_0} p_l^{N - n_0}, \end{aligned}$$

$$0 \leq n_0 \leq N.$$

Logo, definindo  $L_1(N, p_l | n_1, n_2, \dots, n_{l-1}) = P(n | N, \mathbf{p})$  e  $L_2(\mathbf{p} | n_1, n_2, \dots, n_{l-1}) = P(n_1, n_2, \dots, n_{l-1} | n, N, \mathbf{p})$  temos

$$\begin{aligned} L_1(N, p_l | n_1, n_2, \dots, n_{l-1}) &= P(n | N, \mathbf{p}) \\ &= \binom{N}{n} p_l^{N - n} (1 - p_l)^n, \end{aligned}$$

$N \geq n$ , e de 4.1,

$$\begin{aligned} L_2(\mathbf{p} | n_1, n_2, \dots, n_{l-1}) &= \frac{P(n_1, n_2, \dots, n_{l-1} | n, N, \mathbf{p})}{P(n | N, \mathbf{p})} \\ &= \frac{P(n_1, n_2, \dots, n_{l-1}, n | N, \mathbf{p})}{P(n | N, \mathbf{p})} \\ &= \frac{P(n_1, n_2, \dots, n_{l-1} | N, \mathbf{p}) \mathbf{I}_{\{n\}}(\sum_{j=1}^k n_j)}{P(n | N, \mathbf{p})} \\ &= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N - n)!} \prod_{r=1}^{l-1} p_r^{n_r} p_l^{N - n} \\ &= \frac{\binom{N}{n} p_l^{N - n} (1 - p_l)^n}{\binom{N}{n} p_l^{N - n} (1 - p_l)^n} \\ &= \frac{n!}{n_1! n_2! \dots n_{l-1}!} \prod_{r=1}^{l-1} \left( \frac{p_r}{1 - p_l} \right)^{n_r}, \end{aligned}$$

o que prova o resultado.

Como nosso interesse é construir uma função de verossimilhança em função apenas do parâmetro de interesse,  $N$ , vamos utilizar a metodologia da verossimilhança integrada uniforme em relação à  $L_1(N, p_l | n_1, n_2, \dots, n_{l-1})$ , supondo que, dado  $N$ ,  $p_l$  tem distribuição  $Uniforme(0, 1)$ . Desse modo, definimos a função de verossimilhança restrita integrada de  $N$  por

$$\begin{aligned}
 L^{\mathcal{R}}(N | n_1, n_2, \dots, n_k) &= \int_0^1 L_1(N, p_l | n_1, n_2, \dots, n_{l-1}) dp_l \\
 &= \binom{N}{n} \int_0^1 p_l^{N-n} (1-p_l)^n dp_l \\
 &= \binom{N}{n} \frac{\Gamma(N-n+1) \Gamma(n+1)}{\Gamma(N-n+1+n+1)} \\
 &= \frac{1}{N+1},
 \end{aligned} \tag{4.6}$$

$N \geq n$ .

### 4.2.3 Funções de Verossimilhança Integradas: Uniforme, Jeffreys e Generalizada

Apresentamos, como na seção 2.2.3, as funções de verossimilhança integradas considerando as mesmas funções peso para  $\mathbf{p}$ , com o objetivo de obter uma função de verossimilhança em função apenas de  $N$ . Neste sentido suponhamos que  $\mathbf{p} = (p_1, p_2, \dots, p_{l-1}) | N \sim Dirichlet(\alpha_1, \alpha_2, \dots, \alpha_l)$ ,  $\alpha_r > 0$  e conhecido,  $r = 1, 2, \dots, l$ . Logo,

$$\pi(\mathbf{p} | N) = \frac{\Gamma(\sum_{r=1}^l \alpha_r)}{\prod_{r=1}^l \Gamma(\alpha_r)} \prod_{r=1}^l p_j^{\alpha_r-1} \mathbf{I}_{\mathcal{S}}(\mathbf{p}), \tag{4.7}$$

onde  $\mathbf{I}_{\mathcal{S}}(\cdot)$  é a função indicadora de  $\mathcal{S}$  (simplex de dimensão  $l-1$  definido na seção 4.1).

#### (I) Função de Verossimilhança Integrada Uniforme

Suponhamos que, dado  $N$ ,  $\mathbf{p}$  tem distribuição uniforme em  $\mathcal{S}$ , isto é,  $\mathbf{p} | N \sim Dirichlet(1, 1, \dots, 1)$ . Logo,  $\pi(\mathbf{p} | N) = \Gamma(l) \cdot \mathbf{I}_{\mathcal{S}}(\mathbf{p}) \propto \mathbf{I}_{\mathcal{S}}(\mathbf{p})$  e, de 2.5 e 4.1, segue que a função de verossimilhança integrada uniforme de  $N$  é tal que

$$\begin{aligned}
L^u(N|n_1, n_2, \dots, n_{l-1}) &\propto \int_{\mathcal{S}} L(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1}) d\mathbf{p} \\
&= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N-n)!} \int_{\mathcal{S}} p_1^{n_1} p_2^{n_2} \dots p_l^{N-n} d\mathbf{p} \\
&= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N-n)!} \frac{\prod_{r=1}^{l-1} \Gamma(n_r + 1) \Gamma(N-n+1)}{\Gamma\left[\sum_{r=1}^{l-1} (n_r + 1) + N-n+1\right]} \\
&\propto \frac{N!}{(N-n)!} \frac{\Gamma(N-n+1)}{\Gamma(N+l)} \\
&= \frac{N!}{(N+l-1)!} \\
&= \frac{1}{(N+1)(N+2)\dots(N+l-1)},
\end{aligned}$$

e o núcleo de  $L^u(N|n_1, n_2, \dots, n_{l-1})$  é dado por

$$K^u(N|n_1, n_2, \dots, n_{l-1}) = \frac{1}{\prod_{r=1}^{l-1} (N+r)}, \quad (4.8)$$

$N \geq n$ .

## (II) Função de Verossimilhança Integrada de Jeffreys

Novamente construímos a função de verossimilhança integrada considerando a priori de Jeffreys para  $\mathbf{p}$ , dado  $N$ . Como vimos na definição 3.1 da priori de Jeffreys estendida para o vetor  $\mathbf{p}$ , temos que

$$\pi^{\mathcal{J}}(\mathbf{p}|N) \propto \left[ \det I_{n_1, n_2, \dots, n_{l-1}}(N, \mathbf{p}) \right]^{\frac{1}{2}},$$

onde  $I_{n_1, n_2, \dots, n_{l-1}}(N, \mathbf{p}) = \left( E \left[ - \frac{\partial^2 l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})}{\partial p_i \partial p_j} \Big| \mathbf{p} \right] \right)_{1 \leq i, j \leq l}$  é a matriz de informação esperada de Fisher.

As derivadas de primeira ordem de  $l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})$  são

$$\begin{cases} \frac{\partial l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})}{\partial p_r} = \frac{n_r}{p_r}, & r = 1, 2, \dots, l-1, \\ \frac{\partial l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})}{\partial p_l} = \frac{N-n}{p_l}, \end{cases}$$

e as derivadas de segunda ordem de  $l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})$  são

$$\left\{ \begin{array}{l} \frac{\partial^2 l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})}{\partial p_r^2} = -\frac{n_r}{p_r^2}, \quad r = 1, 2, \dots, l-1, \\ \frac{\partial^2 l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})}{\partial p_l^2} = -\frac{N-n}{p_l^2}, \\ \frac{\partial^2 l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})}{\partial p_i \partial p_j} = 0, \quad i \neq j. \end{array} \right.$$

Então,

$$E \left[ - \frac{\partial^2 l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})}{\partial p_i \partial p_j} \middle| -\mathbf{p} \right] = \left\{ \begin{array}{l} \frac{N p_r}{p_r^2} = \frac{N}{p_r}, \quad r = 1, 2, \dots, l-1, \\ \frac{N - N(1-p_l)}{p_l^2} = \frac{N}{p_l}, \\ 0, \quad i \neq j, \end{array} \right.$$

o que implica

$$I_{n_1, n_2, \dots, n_{l-1}}(N, \mathbf{p}) = \begin{bmatrix} \frac{N}{p_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{N}{p_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{N}{p_l} \end{bmatrix}.$$

Logo, a priori de Jeffreys para  $\mathbf{p}$ , dado  $N$ , é tal que

$$\begin{aligned} \pi^{\mathcal{J}}(\mathbf{p}|N) &\propto \left[ \det I_{n_1, n_2, \dots, n_{l-1}}(N, \mathbf{p}) \right]^{\frac{1}{2}} \\ &= \left[ \frac{N^l}{p_1 p_2 \dots p_l} \right]^{\frac{1}{2}} \\ &\propto p_1^{-1/2} p_2^{-1/2} \dots p_l^{-1/2}. \end{aligned}$$

Desse modo, de 2.5 e 4.1, a função de verossimilhança integrada de Jeffreys é dada por

$$L^{\mathcal{J}}(N|n_1, n_2, \dots, n_{l-1}) = \int_{\mathcal{S}} L(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1}) \pi^{\mathcal{J}}(\mathbf{p}|N) d\mathbf{p}$$

$$\begin{aligned}
&\propto \frac{N!}{n_1! n_2! \dots n_{l-1}! (N-n)!} \int_{\mathcal{S}} p_1^{n_1-1/2} p_2^{n_2-1/2} \dots p_l^{N-n-1/2} d\mathbf{p} \\
&= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N-n)!} \frac{\prod_{r=1}^{l-1} \Gamma(n_r + 1/2) \Gamma(N-n+1/2)}{\Gamma\left[\sum_{r=1}^{l-1} (n_r + 1/2) + N-n+1/2\right]}
\end{aligned}$$

e o núcleo de  $L^{\mathcal{J}}(N|n_1, n_2, \dots, n_{l-1})$  é dado por

$$K^{\mathcal{J}}(N|n_1, n_2, \dots, n_{l-1}) = \frac{N!}{(N-n)!} \frac{\Gamma(N-n+1/2)}{\Gamma(N+l/2)}, \quad (4.9)$$

$N \geq n$ .

### (III) Função de Verossimilhança Integrada Generalizada

Para obter uma função de verossimilhança integrada generalizada para este modelo suponhamos que  $\mathbf{p}|N \sim \text{Dirichlet}(\alpha_1, \alpha_2, \dots, \alpha_l)$ ,  $\alpha_r > 0$  e conhecido,  $r = 1, 2, \dots, l$ .

Então,

$$\pi^{\mathcal{G}}(\mathbf{p}|N) \propto \prod_{r=1}^l p_r^{\alpha_r-1} \cdot \mathbf{I}_{\mathcal{S}}(\mathbf{p})$$

e de 2.5 e 4.1 segue que

$$\begin{aligned}
L^{\mathcal{G}}(N|n_1, n_2, \dots, n_{l-1}) &\propto \int_{\mathcal{S}} L(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1}) \pi^{\mathcal{G}}(\mathbf{p}|N) d\mathbf{p} \\
&= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N-n)!} \int_{\mathcal{S}} p_1^{n_1+\alpha_1-1} p_2^{n_2+\alpha_2-1} \dots p_l^{N-n+\alpha_l-1} d\mathbf{p} \\
&= \frac{N!}{n_1! n_2! \dots n_{l-1}! (N-n)!} \frac{\Gamma(n_1 + \alpha_1) \dots \Gamma(n_{l-1} + \alpha_{l-1}) \Gamma(N-n + \alpha_l)}{\Gamma((n_1 + \alpha_1) + \dots + (n_{l-1} + \alpha_{l-1}) + (N-n + \alpha_l))} \\
&\propto K^{\mathcal{G}}(N|n_1, n_2, \dots, n_{l-1}),
\end{aligned}$$

onde

$$K^{\mathcal{G}}(N|n_1, n_2, \dots, n_{l-1}) = \frac{N!}{(N-n)!} \frac{\Gamma(N-n + \alpha_l)}{\Gamma(N + \sum_{r=1}^l \alpha_r)}, \quad (4.10)$$

$N \geq n$ , é o núcleo de  $L^{\mathcal{G}}(N|n_1, n_2, \dots, n_{l-1})$ .

Note que, se  $\alpha_r = 1$ ,  $r = 1, 2, \dots, l$ , temos a função de verossimilhança integrada uniforme de  $N$  e se  $\alpha_r = 1/2$ ,  $r = 1, 2, \dots, l$ , temos a função de verossimilhança integrada de Jeffreys.

### 4.3 Estimativas de Máxima Verossimilhança

Determinamos novamente as emv de  $N$  correspondentes às diferentes funções de verossimilhança apresentadas anteriormente.

Como vimos na seção 2.3, a emv de  $(N, \mathbf{p})$ ,  $(\hat{N}, \hat{\mathbf{p}})$ , é o ponto que maximiza a função de verossimilhança,  $L(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1})$ , ou o seu logaritmo.

Utilizando a técnica de multiplicadores de Lagrange definimos a função

$$g(\mathbf{p}) = l(N, \mathbf{p}|n_1, n_2, \dots, n_{l-1}) + \lambda \left( 1 - \sum_{r=1}^l p_r \right)$$

e supondo  $N$  uma variável contínua,  $(\hat{N}, \hat{\mathbf{p}})$  é solução do sistema

$$\left\{ \begin{array}{l} \frac{\partial g(\mathbf{p})}{\partial p_r} = 0, \quad r = 1, 2, \dots, l-1, \\ \frac{\partial g(\mathbf{p})}{\partial p_l} = 0, \\ \frac{\partial g(\mathbf{p})}{\partial \lambda} = 0, \\ \frac{\partial g(\mathbf{p})}{\partial N} = 0, \quad N \geq n. \end{array} \right. \quad (4.11)$$

O sistema 4.11 é indeterminado, ou seja, não é possível obter  $(\hat{N}, \hat{\mathbf{p}})$ . Isso ocorre porque o sistema tem  $l$  parâmetros a serem estimados ( $p_1, p_2, \dots, p_{l-1}$  e  $N$ ) e apenas  $l-1$  estatísticas ( $n_1, n_2, \dots, n_{l-1}$ ).

As emv de  $N$  com relação às demais funções de verossimilhança são dadas pelo teorema a seguir.

**Teorema 4.1** *As emv de  $N$  com relação às funções de verossimilhança perfilada,  $\hat{N}^{\mathcal{P}}$ , restrita integrada,  $\hat{N}^{\mathcal{R}}$ , integrada uniforme,  $\hat{N}^{\mathcal{U}}$ , e integrada de Jeffreys,  $\hat{N}^{\mathcal{J}}$ , são iguais à  $n$ .*

**Demonstração.** Com relação à função de verossimilhança perfilada, o resultado  $1 - x < \exp(-x)$ , para todo  $x$  real,  $x \neq 0$ , implica

$$\frac{x}{x+1} = 1 - \frac{1}{x+1} < \exp\left(-\frac{1}{x+1}\right), \quad (4.12)$$

para todo  $x$  real,  $x > 0$ .

Por outro lado, a função  $g(\cdot)$  definida por

$$g(x) = \left( \frac{x+1}{x} \right)^x, \quad (4.13)$$

para todo  $x$  real,  $x > 0$ , é estritamente crescente e  $\lim_{x \rightarrow +\infty} g(x) = \exp(1)$ .

Com efeito, a função  $h(\cdot)$  definida pelo logaritmo natural de  $g(\cdot)$  é

$$h(x) = x (\log(x+1) - \log x),$$

$x > 0$ , e

$$\begin{aligned} \frac{dh(x)}{dx} &= \log(x+1) - \log x + x \left( \frac{1}{x+1} - \frac{1}{x} \right) \\ &= \log(x+1) - \log x - \frac{1}{x+1} \\ &= \log \left( \frac{x+1}{x} \right) - \frac{1}{x+1}, \end{aligned}$$

$x > 0$ .

Então, por 4.12, temos

$$\begin{aligned} \frac{dh(x)}{dx} &= \log \left( \frac{x+1}{x} \right) - \frac{1}{x+1} \\ &= -\log \left( \frac{x}{x+1} \right) - \frac{1}{x+1} > 0, \end{aligned}$$

$x > 0$ , isto é,  $h(\cdot)$  é monótona crescente e, conseqüentemente,  $g(\cdot)$  é monótona crescente.

Logo, de 4.5 e 4.13, temos

$$\begin{aligned} \frac{K^{\mathcal{P}}(N+1 | n_1, n_2, \dots, n_{l-1})}{K^{\mathcal{P}}(N | n_1, n_2, \dots, n_{l-1})} &= \frac{\frac{(N+1)!}{(N+1-n)!} \frac{(N+1-n)^{N+1-n}}{(N+1)^{N+1}}}{\frac{N!}{(N-n)!} \frac{(N-n)^{N-n}}{N^N}} \\ &= \frac{N+1}{N+1-n} \frac{(N+1-n)^{N+1-n}}{(N+1)^{N+1}} \frac{N^N}{(N-n)^{N-n}} \\ &= \left( \frac{N}{N+1} \right)^N \left( \frac{N-n+1}{N-n} \right)^{N-n} \\ &= \frac{g(N-n)}{g(N)} < 1, \end{aligned}$$

$N \geq n + 1$ , e

$$\frac{K^{\mathcal{P}}(n+1|n_1, n_2, \dots, n_{l-1})}{K^{\mathcal{P}}(n|n_1, n_2, \dots, n_{l-1})} = \frac{(n+1)!}{(n+1)^{n+1}} \frac{n^n}{n!} = \left(\frac{n}{n+1}\right)^n < 1,$$

o que implica que a função de verossimilhança perfilada é monótona decrescente de  $N$ , para todo  $N \geq n$ , ou seja,  $\hat{N}^{\mathcal{P}} = n$ .

Nos casos das funções de verossimilhança restrita integrada e integrada uniforme dadas por 4.6 e 4.8, respectivamente, observamos que elas são funções monótonas decrescentes de  $N$ ,  $N \geq n$ . Logo,  $\hat{N}^{\mathcal{R}} = \hat{N}^{\mathcal{U}} = n$ .

Finalmente, com relação à  $\hat{N}^{\mathcal{J}}$  temos, por 4.9,

$$\begin{aligned} \frac{K^{\mathcal{J}}(N+1|n_1, n_2, \dots, n_{l-1})}{K^{\mathcal{J}}(N|n_1, n_2, \dots, n_{l-1})} &= \frac{(N+1)! \Gamma(N+1-n+1/2)}{(N+1-n)! \Gamma(N+1+l/2)} \frac{(N-n)! \Gamma(N+l/2)}{N! \Gamma(N-n+1/2)} \\ &= \frac{(N+1) (N-n+1/2) \Gamma(N-n+1/2)}{(N+1-n) (N+l/2) \Gamma(N+l/2)} \frac{\Gamma(N+l/2)}{\Gamma(N-n+1/2)} \\ &= \frac{N+1}{N+l/2} \frac{N-n+1/2}{N-n+1} < 1, \end{aligned}$$

$N \geq n$ , o que implica que a função de verossimilhança integrada de Jeffreys é uma função monótona decrescente de  $N$ ,  $N \geq n$ . Assim,  $\hat{N}^{\mathcal{J}} = n$ . ■

### 4.3.1 Estudo da Performance do Modelo Multinomial

Novamente analisamos o desempenho das emv de  $N$  através de um estudo a partir de amostras simuladas para  $N = 100$  e  $k = 2$ ,  $k = 3$  e  $k = 4$  ocasiões de amostragem, que correspondem à  $l = 4$ ,  $l = 8$  e  $l = 16$  históricos de captura, respectivamente.

Para obter o vetor  $\mathbf{p}$  geramos  $x_1, x_2, \dots, x_l$  valores da distribuição  $Gama(1,1)$  e, assim, obtivemos  $p_r = \frac{x_r}{\sum_{r=1}^l x_r}$  de forma que  $\mathbf{p} \sim Dirichlet(1, 1, \dots, 1)$ . Consideramos somente amostras que apresentaram  $p_l = 0,01$ ,  $p_l = 0,2$  e  $p_l = 0,5$ , pois o nosso interesse é comparar o desempenho do modelo para  $p_l$  muito pequeno e muito alto. A rotina programada em R é apresentada no apêndice C.

Nas tabelas 4.1, 4.2 e 4.3 apresentamos as amostras obtidas, onde denotamos por  $\mathbf{n} = (n_1, n_2, \dots, n_{l-1})$  e, a partir delas obtemos as emv perfilada, restrita integrada, integrada uniforme e integrada de Jeffreys para  $N$ , todas iguais à  $n$  pelo teorema 4.1.

Em seguida, apresentamos na Tabela 4.5 os resumos das medidas descritivas empíricas das estimativas do tamanho populacional,  $N$ .

### (I) Estimativas de Máxima Verossimilhança de $N$

Tabela 4.1: Amostras simuladas para  $l = 4$  históricos de captura

$\mathbf{a}_1$	$\mathbf{n} = (31, 6, 60)$
	$n = 97$
$\mathbf{a}_2$	$\mathbf{n} = (8, 36, 39)$
	$n = 83$
$\mathbf{a}_3$	$\mathbf{n} = (18, 21, 14)$
	$n = 53$

A amostra  $\mathbf{a}_1$  foi gerada de  $N = 100$  e  $\mathbf{p} = (0, 32; 0, 07; 0, 61; 0, 01)$ ;  $\mathbf{a}_2$  de  $N = 100$  e  $\mathbf{p} = (0, 11; 0, 38; 0, 32; 0, 20)$  e  $\mathbf{a}_3$  de  $N = 100$  e  $\mathbf{p} = (0, 16; 0, 22; 0, 12; 0, 50)$ .

Tabela 4.2: Amostras simuladas para  $l = 8$  históricos de captura

$\mathbf{a}_4$	$\mathbf{n} = (8, 45, 5, 10, 6, 12, 11)$
	$n = 97$
$\mathbf{a}_5$	$\mathbf{n} = (2, 2, 31, 18, 20, 0, 2)$
	$n = 75$
$\mathbf{a}_6$	$\mathbf{n} = (1, 11, 1, 5, 5, 11, 10, 56)$
	$n = 44$

A amostra  $\mathbf{a}_4$  foi gerada de  $N = 100$  e  $\mathbf{p} = (0, 06; 0, 56; 0, 04; 0, 12; 0, 02; 0, 10; 0, 09; 0, 01)$ ;  $\mathbf{a}_5$  de  $N = 100$  e  $\mathbf{p} = (0, 04; 0, 04; 0, 30; 0, 17; 0, 21; 0, 01; 0, 03; 0, 20)$  e  $\mathbf{a}_6$  de  $N = 100$  e  $\mathbf{p} = (0, 04; 0, 12; 0, 03; 0, 07; 0, 06; 0, 07; 0, 11; 0, 50)$ .

Tabela 4.3: Amostras simuladas para  $l = 16$  históricos de captura

$\mathbf{a}_7$	$\mathbf{n} = (5, 7, 3, 3, 1, 11, 10, 4, 5, 15, 14, 5, 1, 13, 0)$	
	$n = 97$	
$\mathbf{a}_8$	$\mathbf{n} = (4, 4, 8, 0, 1, 2, 3, 1, 5, 9, 4, 2, 16, 14, 7)$	
	$n = 80$	
$\mathbf{a}_9$	$\mathbf{n} = (8, 3, 1, 4, 9, 2, 2, 1, 3, 0, 1, 5, 0, 2, 3)$	
	$n = 44$	

A amostra  $\mathbf{a}_7$  foi gerada de  $N = 100$  e  $\mathbf{p} = (0, 08; 0, 06; 0, 03; 0, 05; 0, 01; 0, 10; 0, 07; 0, 08; 0, 04; 0, 16; 0, 13; 0, 04; 0, 01; 0, 12; 0, 00; 0, 01)$ ;  $\mathbf{a}_8$  de  $N = 100$  e  $\mathbf{p} = (0, 03; 0, 04; 0, 05; 0, 02; 0, 01; 0, 03; 0, 04; 0, 03; 0, 06; 0, 04; 0, 04; 0, 04; 0, 20; 0, 12; 0, 07; 0, 20)$  e  $\mathbf{a}_9$  de  $N = 100$  e  $\mathbf{p} = (0, 12; 0, 04; 0, 01; 0, 04; 0, 07; 0, 05; 0, 02; 0, 01; 0, 04; 0, 00; 0, 02; 0, 02; 0, 00; 0, 05; 0, 02; 0, 50)$ .

Tabela 4.4: Estimativas de máxima verossimilhança perfilada, restrita integrada e integradas uniforme e de Jeffreys para as amostras das tabelas 4.1, 4.2 e 4.3

$l$	amostra	n	$\hat{N}^{\mathcal{P}} = \hat{N}^{\mathcal{R}} = \hat{N}^u = \hat{N}^{\mathcal{J}}$
	$\mathbf{a}_1$	97	97
4	$\mathbf{a}_2$	83	83
	$\mathbf{a}_3$	53	53
8	$\mathbf{a}_4$	97	97
	$\mathbf{a}_5$	75	75
	$\mathbf{a}_6$	44	44
16	$\mathbf{a}_7$	97	97
	$\mathbf{a}_8$	80	80
	$\mathbf{a}_9$	44	44

**(II) Resumos das Medidas Descritivas Empíricas das Estimativas de  $N$** 

Tabela 4.5: Resumos das medidas descritivas empíricas das emv perfilada, condicional, integrada uniforme e integrada de Jeffreys, respectivamente

	EMV	Média	Erro Quadrático Médio	Intervalo de 95% de Confiança	Amplitude do IC
$k = 2$	$p_l = 0,01$	98,972	2,118	(96, 100)	4
	$p_l = 0,2$	80,087	412,119	(72, 87)	15
	$p_l = 0,5$	50,083	2515,751	(40, 59)	19
$k = 3$	$p_l = 0,01$	99,016	1,942	(97, 100)	3
	$p_l = 0,2$	79,951	417,683	(72, 87)	15
	$p_l = 0,5$	50,2	2505,338	(40, 60)	20
$k = 4$	$p_l = 0,01$	98,946	2,184	(96, 100)	4
	$p_l = 0,2$	79,715	428,181	(71, 87)	16
	$p_l = 0,5$	49,805	2543,999	(40, 59)	19

**4.3.2 Conclusões**

As estimativas obtidas no estudo feito com dados simulados e descritas na Tabela 4.4 nos permitiram concluir que

1. não existe a estimativa de máxima verossimilhança de  $N$ ,  $\hat{N}$ , para o modelo multinomial;
2. todas as emv de  $N$  são iguais à  $n$ ;
3. claramente, as estimativas são melhores quando  $p_l$  é pequeno;
4. as estimativas não melhoram com o aumento de  $k$ , as amostras  $a_3$  e  $a_9$  ilustram este fato.

Como esperávamos, as informações apresentadas na Tabela 4.5 confirmam que as estimativas são satisfatórias apenas no caso em que  $p_l = 0.01$ , evidenciado pelo fato que os IC de 95% de  $N$  nos outros casos sequer contém o verdadeiro valor do parâmetro. Isso é natural uma vez que, quanto menor for o valor de  $p_l$ , probabilidade de um indivíduo apresentar o histórico não observável, maior a chance de todos os indivíduos da população terem sido observados, portanto a estimativa estará mais próxima do valor verdadeiro do parâmetro.

Neste capítulo propusemos um modelo que apresenta um pouco mais de informação que o modelo considerado no capítulo 3, através dos históricos de captura. Apesar disso, as estimativas não se mostraram melhores que as obtidas pelo modelo binomial com marcação.

# Capítulo 5

## Considerações Finais

No capítulo 2 propusemos o modelo binomial sem remoção, que considera apenas a contagem do número de indivíduos capturados, sem identificar se é a primeira captura do indivíduo ou se é recaptura do mesmo. Como previsto, tal modelo não produziu estimativas satisfatórias de  $N$  por não considerar o número de indivíduos distintos observados.

Com o objetivo de corrigir essa característica amostral, propusemos no capítulo 3 o modelo binomial com marcação, que considera a marcação dos indivíduos capturados, de forma que os indivíduos selecionados, em cada ocasião, serão reconhecidos se forem selecionados novamente em outra ocasião. Tal modelo produziu estimativas realmente satisfatórias do tamanho populacional, mesmo em situações desfavoráveis como probabilidades de captura pequenas e poucas ocasiões amostrais.

O terceiro modelo proposto, apresentado no capítulo 4, supostamente apresenta um pouco mais de informação que o modelo considerado no capítulo 3, através dos históricos de captura. No entanto, notamos que esta informação não é incorporada pela funções de verossimilhança de  $N$  propostas. Claramente, a metodologia de eliminação do parâmetro perturbador para este modelo foi de extrema importância, uma vez que não existem as estimativas de máxima verossimilhança de  $N$  e das probabilidades de ocorrência dos históricos de captura.

Vale ressaltar que, para todos os modelos estudados, apesar das diferentes funções de verossimilhança de  $N$  terem se comportado de maneira semelhante, notamos uma pequena vantagem das funções de verossimilhança integradas (Uniforme e de Jeffreys)

em relação às demais pois, em geral, apresentam erro quadrático médio e amplitudes dos intervalos de confiança menores. Além disso, nos três casos apresentados, apesar da eliminação do parâmetro perturbador, notamos que este ainda influencia fortemente os resultados obtidos.

Nesta dissertação, nos limitamos a encontrar as estimativas do tamanho populacional em cada um dos casos apresentados. No entanto, um estudo mais aprofundado poderia conter análise de sensibilidade das prioris escolhidas, combinações de funções de verossimilhança, entre outras utilidades da eliminação de parâmetros perturbadores, como as apresentadas na introdução, ou ainda uma análise bayesiana plena (full bayesian).

Com relação ao modelo que leva em conta os históricos de captura dos indivíduos, pode ser feito um estudo bayesiano das diversas estimativas de  $N$  propostas pelas diferentes funções de verossimilhança, como discutido em Silva *et al.* (2007) para a função de verossimilhança global.

Uma outra possível pesquisa seria a realização de um estudo bayesiano análogo ao desenvolvido por Berger *et al.* (1999), para as funções de verossimilhança de  $N$  discutidas neste trabalho.

# Referências

- Ananda, M. M. A. (1997). Bayesian methods for mark-resighting surveys. *Communications in Statistics - Theory and Methods*, **26**(3), 685-697.
- Barnard, G. A., Jenkins, G. M. & Winsten, C. B. (1962). Likelihood inference and time series. *Journal of the Royal Statistical Society*, **125**(3), 321-372.
- Basu, S. & Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: heterogeneity and dependence. *Biometrika*, **88**(1), 269-279.
- Bayarri, M. J., DeGroot, M. H. & Kadane, J. B. (1988). What is the likelihood function. *Statistical Decision Theory and Related Topics IV*.
- Berger, J. O. & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, **76**(1), 159-165.
- Berger, J. O., Liseo, B. & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, **14**(1), 1-28.
- Bolsoni, S. B. (2002). *Estimação dos Parâmetros de uma População a partir de Observações Incompletas da Distribuição Multinomial*. Dissertação de mestrado - UFSCar, São Carlos.
- Burnham, K. P. & Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, **65**(3), 625-633.
- Castledine, B. J. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika*, **68**(1), 197-210.
- Chapman, D. G. (1954). The estimation of biological populations. *The Annals of Mathematical Statistics*, **25**(1), 1-15.
- Cruddas, A. M., Reid, N. & Cox, D. R. (1989). A time series illustration of approximate conditional likelihood. *Biometrika*, **76**(2), 231-237.
- Dahiya, R. C. (1981). An improved method of estimating an integer-parameter by maximum likelihood. *The American Statistician*, **35**(1), 34-37.
- Darroch, J. N. (1959). The multiple-recapture census: II. Estimation when there is immigration or death. *Biometrika*, **46**(3/4), 336-351.

- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**(4), 507–521.
- Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.
- Fraser, D. A. S. & Reid, N. (1989). Adjustments to profile likelihood. *Biometrika*, **76**(3), 477–488.
- George, E. I. & Robert, C. P. (1992). Capture-recapture estimation via Gibbs Sampling. *Biometrika*, **79**(4), 677–683.
- Hunter, A. J. & Griffiths, H. J. (1978). Bayesian approach to estimation of insect population size. *Technometrics*, **20**(3), 231–234.
- Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, **52**(1-2), 225–247.
- Kalbfleisch, J. D. & Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society*, **32**(2), 175–208.
- Kalbfleisch, J. D. & Sprott, D. A. (1973). Marginal and conditional likelihoods. *Sankhya: The Indian Journal of Statistics*, **35**(3), 311–328.
- Laplace, P. S. (1783). Sur les naissances, les mariages, et les morts. *Histoire de l'Academie Royale des Sciences*.
- Lee, A. J. (2002). Effect of list errors on the estimation of population size. *Biometrics*, **58**(1), 185–191.
- Lee, A. J., Seber, G. A. F., Holden, J. K. & Huakau, J. T. (2001). Capture-recapture, epidemiology, and list mismatches: several lists. *Biometrics*, **57**(3), 707–713.
- Leite, J. G., Oishi, J. & Pereira, C. A. B. (1988). A note on the exact maximum likelihood estimation of the size of a finite and closed population. *Biometrika*, **75**(1), 178–180.
- Lincoln, F. C. (1930). Calculating waterfowl abundance on the basis of banding returns. *US Department of Agriculture Circular*.
- Nayak, T. K. (1988). Estimating population size by recapture sampling. *Biometrika*, **75**(1), 113–120.
- Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**(1), 1–32.
- Petersen, C. G. (1896). The yearly immigration of young plaice into the Limfjord from the German Sea. *Report of the Danish Biological Station*.
- Pollock, K. H. (1991). 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**(413), 225–238.

- Reid, N. (1995). The roles of conditioning in inference. *Statistical Science*, **10**(2), 138–157.
- Reid, N. (1996). Likelihood and Bayesian approximation methods. *Bayesian Statistics 5*, pages 351–369.
- Savage, L. J. (1976). On rereading R. A. Fisher. *The annals of statistics*, **4**(3), 441–500.
- Seber, G. A. F. (1992). A review of estimating animal abundance II. *International Statistical Review*, **60**(2), 129–166.
- Silva, K. B. R., Leite, J. G. & Tanaka, N. I. (2007). Um modelo Bayesiano para estimar o tamanho de uma população fechada. *Revista Brasileira de Biometria*, **25**(4), 135–156.
- 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**(2), 399–407.
- Sweeting, T. J. (1995a). A framework for Bayesian and likelihood approximations in statistics. *Biometrika*, **82**(1), 1–23.
- Sweeting, T. J. (1995b). A Bayesian approach to approximate conditional inference. *Biometrika*, **82**(1), 25–36.
- Sweeting, T. J. (1996). Approximate Bayesian computation based on signed roots of log-density ratios. *Bayesian Statistics*, (1), 25–36.
- Yoshida, O., Leite, J. G. & Bolfarine, H. (1999). Stochastic monotonicity properties of Bayes estimation of the population size for capture-recapture data. *Statistics and Probability Letters*, **42**(3), 257–266.
- Zacharias, H. P. (2000). *Aplicação do Algoritmo Gibbs Sampling no Processo de Captura-Recaptura*. Dissertação de mestrado - UFSCar, São Carlos.

# Apêndice A

Rotina utilizada para determinar resumos das medidas descritivas empíricas das emv de  $N$  para o capítulo 2. Nesta rotina já estão inseridos os programas utilizados para determinar a emv e as emv perfilada, condicional, integrada uniforme e de Jeffreys de  $N$ , através das equações encontradas na seção 2.3.2.

```
#####  
# FUNÇÕES UTILIZADAS  
#####  
  
# N: tamanho da população  
# n: vetor do nro de indivíduos observados em cada ocasião  
# p: probabilidade de captura  
  
# Log-verossimilhança perfilada  
log.lik_perf <- function(N, n) {  
k <- length(n);  
n_sum <- sum(n);  
result <- k*lgamma(N+1) - sum(lgamma(N - n + 1)) - sum(lgamma(n + 1)) + (N*k - n_sum)*log(N*k  
- n_sum) - (N*k)*log(N*k)  
return(result)  
} # end-function  
  
# Log-verossimilhança condicional  
log.lik_cond <- function(N, n) {  
k <- length(n);  
n_sum <- sum(n);  
result <- k*lgamma(N+1) - sum(lgamma(N - n + 1)) - sum(lgamma(n + 1)) - lgamma(N*k + 1)
```

```

+ lgamma(N*k - n_sum + 1) + lgamma(n_sum + 1)
return(result)
} # end-function

# Log-verossimilhança integrada uniforme
log.lik_uni <- function(N, n) {
k <- length(n);
n_sum <- sum(n);
result <- k*lgamma(N+1) - sum(lgamma(N - n + 1)) - lgamma(N*k + 2) + lgamma(N*k - n_sum
+ 1)
return(result)
} # end-function

# Log-verossimilhança integrada de Jeffreys
log.lik_jeff <- function(N, n) {
k <- length(n);
n_sum <- sum(n);
result <- k*lgamma(N+1) - sum(lgamma(N - n + 1)) - lgamma(N*k + 1) + lgamma(N*k - n_sum
+ 0.5)
return(result)
} # end-function

# Carrega pacote Stats4
require(stats4)

# EQUAÇÃO DE ESTIMAÇÃO (VEROSSIMILHANÇA e VEROSS. PERFILADA)
g2 <- function(N, n) {
k <- length(n)
result <- k*digamma(N + 1) - sum(digamma(N - n + 1)) + k*log(1 - sum(n)/(k*N))
return(result)
}# end-function

# EQUAÇÃO DE ESTIMAÇÃO (VEROSSIMILHANÇA CONDICIONAL)
g3 <- function(N, n) {
k <- length(n);
n_sum <- sum(n);
result <- k*digamma(N + 1) - sum(digamma(N - n + 1)) - k*digamma(N*k + 1) + k*digamma(N*k

```

```

- n_sum + 1)
return(result)
}# end-function

# EQUAÇÃO DE ESTIMAÇÃO (VEROSSIMILHANÇA UNIFORME)
g4 <- function(N, n) {
k <- length(n);
n_sum <- sum(n);
result <- k*digamma(N + 1) - sum(digamma(N - n + 1)) - k*digamma(N*k + 2) + k*digamma(N*k
- n_sum + 1)
return(result)
}# end-function

# EQUAÇÃO DE ESTIMAÇÃO (VEROSSIMILHANÇA de JEFFREYS)
g5 <- function(N, n) {
k <- length(n);
n_sum <- sum(n);
result <- k*digamma(N + 1) - sum(digamma(N - n + 1)) - k*digamma(N*k + 1) + k*digamma(N*k
- n_sum + 0.5)
return(result)
}# end-function

#####
#ROTINA
#####

k <- 5
p <- 0.8
S <- 1000
emv_N.g2 <- rep(0,S)
emv_N.g3 <- rep(0,S)
emv_N.g4 <- rep(0,S)
emv_N.g5 <- rep(0,S)

amostra <- matrix(0, nrow = S, ncol = k)

set.seed(100786)

for (i in 1:S) {

```

```
amostra[i, ] <- rbinom(k,100,p)

# OBTENÇÃO DO ZERO DAS FUNÇÕES
print(paste("Simulação ", i))
zero.g2 <- try(uniroot(g2, lower = max(amostra[i, ]), upper = 1000, n = amostra[i,])$root,
silent = TRUE)
if (class(zero.g2) == "try-error") {
if (log.lik_perf(max(amostra[i, ]), amostra[i, ]) > log.lik_perf(max(amostra[i, ]) + 1,
amostra[i, ])) {
zero.g2 <- max(amostra[i, ])} else {
zero.g2 <- Inf
} # end-if-else
} # end-if

zero.g3 <- try(uniroot(g3, lower = max(amostra[i, ]), upper = 1000, n = amostra[i,])$root,
silent = TRUE)
if (class(zero.g3) == "try-error") {
if (log.lik_perf(max(amostra[i, ]), amostra[i, ]) > log.lik_perf(max(amostra[i, ]) + 1,
amostra[i, ])) {
zero.g3 <- max(amostra[i, ])} else {
zero.g3 <- Inf
} # end-if-else
} # end-if

zero.g4 <- try(uniroot(g4, lower = max(amostra[i, ]), upper = 1000, n = amostra[i,])$root,
silent = TRUE)
if (class(zero.g4) == "try-error") {
if (log.lik_perf(max(amostra[i, ]), amostra[i, ]) > log.lik_perf(max(amostra[i, ]) + 1,
amostra[i, ])) {
zero.g4 <- max(amostra[i, ])} else {
zero.g4 <- Inf
} # end-if-else
} # end-if

zero.g5 <- try(uniroot(g5 , lower = max(amostra[i, ]), upper = 1000, n = amostra[i,])$root,
silent = TRUE)
if (class(zero.g5) == "try-error") {
if (log.lik_perf(max(amostra[i, ]), amostra[i, ]) > log.lik_perf(max(amostra[i, ]) + 1,
amostra[i, ])) {
zero.g5 <- max(amostra[i, ])} else {
```

```
zero.g5 <- Inf
} # end-if-else
} # end-if

# OBTENÇÃO DAS EMV e EMV PERFILADA
if (zero.g2 == Inf) {
  emv_N.g2[i] <- Inf
} else {
  N1.g2 <- floor(zero.g2)
  N2.g2 <- ceiling(zero.g2)
  l_N1.g2 <- log.lik_perf(N1.g2, amostra[i, ])
  l_N2.g2 <- log.lik_perf(N2.g2, amostra[i, ])
  emv_N.g2[i] <- ifelse(l_N1.g2 > l_N2.g2, N1.g2, N2.g2)
} # end-if-else

# OBTENÇÃO DAS EMV CONDICIONAL
if (zero.g3 == Inf) {
  emv_N.g3[i] <- Inf
} else {
  N1.g3 <- floor(zero.g3)
  N2.g3 <- ceiling(zero.g3)
  l_N1.g3 <- log.lik_cond(N1.g3, amostra[i, ])
  l_N2.g3 <- log.lik_cond(N2.g3, amostra[i, ])
  emv_N.g3[i] <- ifelse(l_N1.g3 > l_N2.g3, N1.g3, N2.g3)
} # end-if-else

# OBTENÇÃO DAS EMV INTEGRADA UNIFORME
if (zero.g4 == Inf) {
  emv_N.g4[i] <- Inf
} else {
  N1.g4 <- floor(zero.g4)
  N2.g4 <- ceiling(zero.g4)
  l_N1.g4 <- log.lik_uni(N1.g4, amostra[i, ])
  l_N2.g4 <- log.lik_uni(N2.g4, amostra[i, ])
  emv_N.g4[i] <- ifelse(l_N1.g4 > l_N2.g4, N1.g4, N2.g4)
} # end-if-else
```

```

# OBTENÇÃO DAS EMV INTEGRADA de JEFFREYS
if (zero.g5 == Inf) {
emv_N.g5[i] <- Inf
} else {
N1.g5 <- floor(zero.g5)
N2.g5 <- ceiling(zero.g5)
l_N1.g5 <- log.lik_jeff (N1.g5 , amostra[i, ])
l_N2.g5 <- log.lik_jeff (N2.g5 , amostra[i, ])
emv_N.g5[i] <- ifelse(l_N1.g5 > l_N2.g5 , N1.g5 , N2.g5 )
} # end-if-else
} # end-for

#####
# RESUMO DESCRITIVO
#####

# PERFILADA
ind_inf.g2 <- which(emv_N.g2 == Inf)
if (length(ind_inf.g2) != 0) emv_N.g2 <- emv_N.g2[-ind_inf.g2]
mu_N.g2 <- mean(emv_N.g2); mu_N.g2
eqm.g2 <- mean((emv_N.g2 - 100)^2); eqm.g2
N_perf <- sort(emv_N.g2)
N_perf[floor(0.025*length(emv_N.g2))]; N_perf[ceiling(0.975*length(emv_N.g2))]

# CONDICIONAL
ind_inf.g3 <- which(emv_N.g3 == Inf)
if (length(ind_inf.g3) != 0) emv_N.g3 <- emv_N.g3[-ind_inf.g3]
mu_N.g3 <- mean(emv_N.g3); mu_N.g3
eqm.g3 <- mean((emv_N.g3 - 100)^2); eqm.g3
N_cond <- sort(emv_N.g3)
N_cond[floor(0.025*length(emv_N.g3))]; N_cond[ceiling(0.975*length(emv_N.g3))]

# INTEGRADA UNIFORME
ind_inf.g4 <- which(emv_N.g4 == Inf)
if (length(ind_inf.g4) != 0) emv_N.g4 <- emv_N.g4[-ind_inf.g4]
mu_N.g4 <- mean(emv_N.g4); mu_N.g4
eqm.g4 <- mean((emv_N.g4 - 100)^2); eqm.g4
N_unif <- sort(emv_N.g4)

```

```
N_unif[floor(0.025*length(emv_N.g4));N_unif[ceiling(0.975*length(emv_N.g4))]  
  
# INTEGRADA DE JEFFREYS  
ind_inf.g5 <- which(emv_N.g5 == Inf)  
if (length(ind_inf.g5) != 0) emv_N.g5 <- emv_N.g5[-ind_inf.g5]  
mu_N.g5 <- mean(emv_N.g5); mu_N.g5  
eqm.g5 <- mean((emv_N.g5 - 100)^2); eqm.g5  
N_jeff <- sort(emv_N.g5)  
N_jeff[floor(0.025*length(emv_N.g5));N_jeff[ceiling(0.975*length(emv_N.g5))]]
```

# Apêndice B

Rotina utilizada para determinar resumos das medidas descritivas empíricas das emv de  $N$  para o capítulo 3. Nesta rotina já estão inseridos os programas utilizados para determinar a emv e as emv perfilada, condicional, integrada uniforme e de Jeffreys de  $N$ , através das equações encontradas na seção 3.3.2.

```
#####  
# FUNÇÕES UTILIZADAS  
#####  
  
# N: tamanho da população  
# n:  vetor do nro de indivíduos observados em cada ocasião  
# p:  vetor das probabilidades de captura em cada ocasião  
# M: vetor do nro de indivíduos marcados até a ocasião atual  
# m:  vetor do nro de indivíduos marcados capturados em cada ocasião  
# d:  vetor do nro de indivíduos distintos capturados em cada ocasião  
# r:  nro total de indivíduos distintos observados  
  
# Log-verossimilhança perfilada  
log.lik_perf <- function(N, n, r) {  
  k <- length(n)  
  result <- lgamma(N+1) - lgamma(N - r + 1) + sum((N-n)*log(N-n)) - k*N*log(N)  
  return(result)  
} # end-function  
  
# Log-verossimilhança integrada uniforme  
log.lik_unif <- function(N, n, r) {  
  k <- length(n);
```

```
result <- lgamma(N+1) - lgamma(N - r + 1) + sum(lgamma(N - n + 1)) - k*lgamma(N + 2)
return(result)
} # end-function
```

```
# Log-verossimilhança integrada de Jeffreys
```

```
log.lik_jeff <- function(N, n, r) {
k <- length(n);
result <- lgamma(N+1) - lgamma(N - r + 1) + sum(lgamma(N - n + 0.5)) - k*lgamma(N + 1)
return(result)
} # end-function
```

```
# Carrega pacote Stats4
```

```
require(stats4)
```

```
# EQUAÇÃO DE ESTIMAÇÃO (VEROSSIMILHANÇA e VEROSS. PERFILADA)
```

```
g2 <- function(N, n , r ) {
k <- length(n)
result <- digamma(N + 1) - digamma(N - r + 1) + sum(log(N - n)) - k*log(N)
return(result)
}# end-function
```

```
# EQUAÇÃO DE ESTIMAÇÃO (VEROSSIMILHANÇA UNIFORME)
```

```
g4 <- function(N, n , r ) {
k <- length(n);
result <- digamma(N + 1) - digamma(N - r + 1) + sum(digamma(N - n + 1)) - k*digamma(N +
2)
return(result)
}# end-function
```

```
# EQUAÇÃO DE ESTIMAÇÃO (VEROSSIMILHANÇA de JEFFREYS)
```

```
g5 <- function(N, n , r ) {
k <- length(n);
result <- (1 - k)*digamma(N + 1) - digamma(N - r + 1) + sum(digamma(N - n + 0.5))
return(result)
}# end-function
```

```
#####  
#ROTINA  
#####  
  
k <- 5  
p_inf <- 0  
p_sup <- 0.2  
S <- 1000  
emv_N.g2 <- rep(0,S)  
emv_N.g3 <- rep(0,S)  
emv_N.g4 <- rep(0,S)  
emv_N.g5 <- rep(0,S)  
  
set.seed(100786)  
p <- round(runif(k,p_inf,p_sup),5)  
  
for (i in 1:S) {  
  
# GERAÇÃO DAS AMOSTRAS  
m <- rep(0,k)  
d <- rep(0,k)  
vn <- rep(0,k)  
M <- rep(0,k)  
  
vn[1] <- rbinom(1,100,p[1])  
d[1] = vn[1]  
  
for(j in 2:k){ M[j] <- M[j-1] + d[j-1]  
m[j] <- rbinom(1,M[j],p[j])  
d[j] <- rbinom(1,(100-M[j]),p[j])  
vn[j] = m[j]+d[j]  
} # end-for  
  
n <- sum(vn)  
amostra <- vn  
r_amostra <- sum(d)  
  
# EMV CONDICIONAL  
if (r_amostra == n){  
emv_N.g3[i] <- Inf  
} else {
```

```

if (r_amostra == max(amostra)) {
  emv_N.g3[i] <- r_amostra
} else {
  x <- 1
  while ( prod(x + r_amostra - amostra) > (x*(x + r_amostra)^(k-1)) ){x <- x+1}
  emv_N.g3[i] <- r_amostra + x - 1
} # end-if-else
} # end-if-else

# OBTENÇÃO DO ZERO DAS FUNÇÕES
print(paste("Simulação ", i))
zero.g2 <- try(uniroot(g2, lower = r_amostra, upper = 1000, n = amostra, r = r_amostra)$root,
  silent = TRUE)
if (class(zero.g2) == "try-error") {
  if (log.lik_perf(r_amostra, amostra, r_amostra) > log.lik_perf(r_amostra + 1, amostra, r_amostra))
  {
    zero.g2 <- r_amostra} else {
    zero.g2 <- Inf
  } # end-if-else
} # end-if

zero.g4 <- try(uniroot(g4, lower = r_amostra, upper = 1000, n = amostra, r = r_amostra)$root,
  silent = TRUE)
if (class(zero.g4) == "try-error") {
  if (log.lik_unif(r_amostra, amostra, r_amostra) > log.lik_unif(r_amostra + 1, amostra, r_amostra))
  {
    zero.g4 <- r_amostra} else {
    zero.g4 <- Inf
  } # end-if-else
} # end-if

zero.g5 <- try(uniroot(g5, lower = r_amostra, upper = 1000, n = amostra, r = r_amostra)$root,
  silent = TRUE)
if (class(zero.g5) == "try-error") {
  if (log.lik_jeff(r_amostra, amostra, r_amostra) > log.lik_jeff(r_amostra + 1, amostra, r_amostra))
  {
    zero.g5 <- r_amostra} else {
    zero.g5 <- Inf
  } # end-if-else
} # end-if

```

```
# OBTENÇÃO DAS EMV e EMV PERFILADA
if (zero.g2 == Inf) {
  emv_N.g2[i] <- Inf
} else {
  N1.g2 <- floor(zero.g2)
  N2.g2 <- ceiling(zero.g2)
  l_N1.g2 <- log.lik_perf(N1.g2, amostra, r_amostra)
  l_N2.g2 <- log.lik_perf(N2.g2, amostra, r_amostra)
  emv_N.g2[i] <- ifelse(l_N1.g2 > l_N2.g2, N1.g2, N2.g2)
} # end-if-else

# OBTENÇÃO DAS EMV INTEGRADA UNIFORME
if (zero.g4 == Inf) {
  emv_N.g4[i] <- Inf
} else {
  N1.g4 <- floor(zero.g4)
  N2.g4 <- ceiling(zero.g4)
  l_N1.g4 <- log.lik_unif(N1.g4, amostra, r_amostra)
  l_N2.g4 <- log.lik_unif(N2.g4, amostra, r_amostra)
  emv_N.g4[i] <- ifelse(l_N1.g4 > l_N2.g4, N1.g4, N2.g4)
} # end-if-else

# OBTENÇÃO DAS EMV INTEGRADA de JEFFREYS
if (zero.g5 == Inf) {
  emv_N.g5[i] <- Inf
} else {
  N1.g5 <- floor(zero.g5)
  N2.g5 <- ceiling(zero.g5)
  l_N1.g5 <- log.lik_jeff(N1.g5, amostra, r_amostra)
  l_N2.g5 <- log.lik_jeff(N2.g5, amostra, r_amostra)
  emv_N.g5[i] <- ifelse(l_N1.g5 > l_N2.g5, N1.g5, N2.g5)
} # end-if-else

} # end-for

#####
# RESUMO DESCRITIVO
```

```
#####

# PERFILADA
ind_inf.g2 <- which(emv_N.g2 == Inf)
if (length(ind_inf.g2) != 0) emv_N.g2 <- emv_N.g2[-ind_inf.g2]
mu_N.g2 <- mean(emv_N.g2); mu_N.g2
eqm.g2 <- mean((emv_N.g2 - 100)^2); eqm.g2
N_perf <- sort(emv_N.g2)
N_perf[floor(0.025*length(emv_N.g2))]; N_perf[ceiling(0.975*length(emv_N.g2))]

# CONDICIONAL
ind_inf.g3 <- which(emv_N.g3 == Inf)
if (length(ind_inf.g3) != 0) emv_N.g3 <- emv_N.g3[-ind_inf.g3]
mu_N.g3 <- mean(emv_N.g3); mu_N.g3
eqm.g3 <- mean((emv_N.g3 - 100)^2); eqm.g3
N_cond <- sort(emv_N.g3)
N_cond[floor(0.025*length(emv_N.g3))]; N_cond[ceiling(0.975*length(emv_N.g3))]

# INTEGRADA UNIFORME
ind_inf.g4 <- which(emv_N.g4 == Inf)
if (length(ind_inf.g4) != 0) emv_N.g4 <- emv_N.g4[-ind_inf.g4]
mu_N.g4 <- mean(emv_N.g4); mu_N.g4
eqm.g4 <- mean((emv_N.g4 - 100)^2); eqm.g4
N_unif <- sort(emv_N.g4)
N_unif[floor(0.025*length(emv_N.g4))];N_unif[ceiling(0.975*length(emv_N.g4))]

# INTEGRADA DE JEFFREYS
ind_inf.g5 <- which(emv_N.g5 == Inf)
if (length(ind_inf.g5) != 0) emv_N.g5 <- emv_N.g5[-ind_inf.g5]
mu_N.g5 <- mean(emv_N.g5); mu_N.g5
eqm.g5 <- mean((emv_N.g5 - 100)^2); eqm.g5
N_jeff <- sort(emv_N.g5)
N_jeff[floor(0.025*length(emv_N.g5))];N_jeff[ceiling(0.975*length(emv_N.g5))]
```

# Apêndice C

Rotina utilizada para determinar resumos das medidas descritivas empíricas das emv de  $N$  para o capítulo 4. Nesta rotina já estão inseridos os programas utilizados para gerar as amostras utilizadas no capítulo e para determinar as emv perfilada, condicional, integrada uniforme e de Jeffreys de  $N$ .

```
#####  
# GERAÇÃO DAS PROBABILIDADES  
#####  
  
k <- 2  
l <- 2^k  
p1 <- rep(0,1)  
p2 <- rep(0,1)  
p3 <- rep(0,1)  
  
set.seed(10071986)  
  
#### p1=0.01 ####  
repeat {  
  x <- rgamma(1,1,1)  
  p1 <- round(x/sum(x),2)  
  if(p1[1] == 0.01 & sum(p1) == 1){break}  
}  
  
#### p1=0.2 ####  
repeat {  
  x <- rgamma(1,1,1)  
  p2 <- round(x/sum(x),2)  
  if(p2[1] == 0.2 & sum(p2) == 1){break}  
}
```

```
#### p1=0.5 ####
repeat {
x <- rgamma(1,1,1)
p3 <- round(x/sum(x),2)
if(p3[1] == 0.5 & sum(p3) == 1){break}
}

#####
#ROTINA
#####

S <- 1000

emv_N1 <- rep(0,S)
emv_N2 <- rep(0,S)
emv_N3 <- rep(0,S)

amostra1 <- matrix(0, nrow = S, ncol = 1)
amostra2 <- matrix(0, nrow = S, ncol = 1)
amostra3 <- matrix(0, nrow = S, ncol = 1)

for (i in 1:S) {
print(paste("Simulação ", i))

amostra1[i,] <- rmultinom(1,100,p1)
amostra2[i,] <- rmultinom(1,100,p2)
amostra3[i,] <- rmultinom(1,100,p3)

emv_N1[i] <- sum (amostra1[i,]) - amostra1[i,1]
emv_N2[i] <- sum (amostra2[i,]) - amostra2[i,1]
emv_N3[i] <- sum (amostra3[i,]) - amostra3[i,1]

} # end-for

#####
# RESUMO DESCRITIVO
#####

# EMV para p1=0.01
mu_N1 <- mean(emv_N1); mu_N1
eqm1 <- mean((emv_N1 - 100)^2); eqm1
N1 <- sort(emv_N1)
N1[floor(0.025*length(emv_N1))]; N1[ceiling(0.975*length(emv_N1))]
```

```
# EMV para p1=0.2
mu_N2 <- mean(emv_N2); mu_N2
eqm2 <- mean((emv_N2 - 100)^2); eqm2
N2 <- sort(emv_N2)
N2[floor(0.025*length(emv_N2))]; N2[ceiling(0.975*length(emv_N2))]

# EMV para p1=0.5
mu_N3 <- mean(emv_N3); mu_N3
eqm3 <- mean((emv_N3 - 100)^2); eqm3
N3 <- sort(emv_N3)
N3[floor(0.025*length(emv_N3))]; N3[ceiling(0.975*length(emv_N3))]
```