

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS (CCET)
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Luis Ernesto Bueno Salasar

Eliminação de parâmetros perturbadores em um modelo de captura-recaptura

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

São Carlos
2011

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

S161ep

Salasar, Luis Ernesto Bueno.

Eliminação de parâmetros perturbadores em um modelo de captura-recaptura / Luis Ernesto Bueno Salasar. -- São Carlos : UFSCar, 2012.

67 f.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2011.

1. Estatística. 2. Estimativas de máxima verossimilhança. 3. Tamanho populacional. 4. População fechada. I. Título.

CDD: 519.5 (20^a)

Luis Ernesto Bueno Salasar

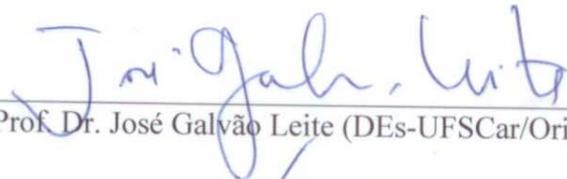
Eliminação de Parâmetros Perturbadores em um Modelo de Captura-recaptura

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

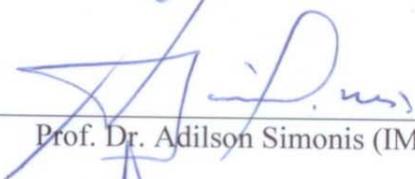
Aprovada em 18 de novembro de 2011.

BANCA EXAMINADORA

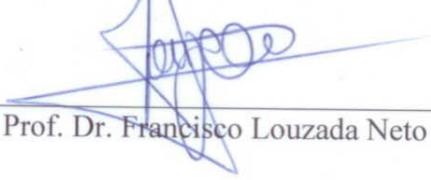
Presidente


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

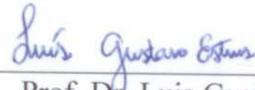
1º Examinador


Prof. Dr. Adilson Simonis (IME-USP)

2º Examinador


Prof. Dr. Francisco Louzada Neto (ICMC-USP)

3º Examinador


Prof. Dr. Luis Gustavo Esteves (IME-USP)

4º Examinador


Profª. Dra. Verónica Andrea González López (IMECC-UNICAMP)

Resumo

O processo de captura-recaptura, amplamente utilizado na estimação do número de elementos de uma população de animais, é também aplicado a outras áreas do conhecimento como Epidemiologia, Linguística, Confiabilidade de Software, Ecologia, entre outras. Uma das primeiras aplicações deste método foi feita por Laplace em 1783, com o objetivo de estimar o número de habitantes da França. Posteriormente, Carl G. J. Petersen em 1889 e Lincoln em 1930 utilizaram o mesmo estimador no contexto de populações de animais. Este estimador ficou conhecido na literatura como o estimador de “Lincoln-Petersen”. Em meados do século XX muitos pesquisadores se dedicaram à formulação de modelos estatísticos adequados à estimação do tamanho populacional, o que causou um aumento substancial da quantidade de trabalhos teóricos e aplicados sobre o tema. Os modelos de captura-recaptura são construídos sob certas hipóteses relativas à população, ao processo de amostragem e às condições experimentais. A principal hipótese que diferencia os modelos diz respeito à mudança do número de indivíduos da população durante o período do experimento. Os modelos que permitem que haja nascimentos, mortes ou migração são chamados de modelos para *população aberta*, enquanto que os modelos em que tais eventos não são permitidos são chamados de modelos para *população fechada*. Neste trabalho, o objetivo é caracterizar o comportamento de funções de verossimilhança obtidas por meio da utilização de métodos de eliminação de parâmetros perturbadores, no caso de modelos para população fechada. Baseado nestas funções de verossimilhança, discutimos métodos de estimação pontual e intervalar para o tamanho populacional. Os métodos de estimação são ilustrados através de um conjunto de dados reais e suas propriedades frequentistas são analisadas via simulação de Monte Carlo.

Palavras-chave: Captura-recaptura. Eliminação de parâmetros perturbadores. Estimação de máxima verossimilhança. Intervalos de confiança. Função de verossimilhança condicional. Função de verossimilhança integrada. Função de verossimilhança perfilada. População fechada. Tamanho populacional.

Abstract

The capture-recapture process, largely used in the estimation of the number of elements of animal population, is also applied to other branches of knowledge like Epidemiology, Linguistics, Software reliability, Ecology, among others. One of the first applications of this method was done by Laplace in 1783, with aim at estimate the number of inhabitants of France. Later, Carl G. J. Petersen in 1889 and Lincoln in 1930 applied the same estimator in the context of animal populations. This estimator has being known in literature as “Lincoln-Petersen” estimator. In the mid-twentieth century several researchers dedicated themselves to the formulation of statistical models appropriated for the estimation of population size, which caused a substantial increase in the amount of theoretical and applied works on the subject. The capture-recapture models are constructed under certain assumptions relating to the population, the sampling procedure and the experimental conditions. The main assumption that distinguishes models concerns the change in the number of individuals in the population during the period of the experiment. Models that allow for births, deaths or migration are called *open population models*, while models that does not allow for these events to occur are called *closed population models*. In this work, the goal is to characterize likelihood functions obtained by applying methods of elimination of nuisance parameters in the case of closed population models. Based on these likelihood functions, we discuss methods for point and interval estimation of the population size. The estimation methods are illustrated on a real data-set and their frequentist properties are analysed via Monte Carlo simulation.

Keywords: Capture-recapture. Closed population. Conditional likelihood function. Confidence intervals. Elimination of nuisance parameters. Integrated likelihood function. Maximum likelihood estimation. Profile likelihood function. Population size.

Sumário

1	Introdução	p. 6
2	Modelo Binomial	p. 8
2.1	Modelo Estatístico	p. 8
2.2	Funções de Verossimilhança	p. 10
2.2.1	Função de Verossimilhança Perfilada	p. 10
2.2.2	Função de Verossimilhança Condicional	p. 11
2.2.3	Função de Verossimilhança Integrada Uniforme	p. 12
2.2.4	Função de Verossimilhança Integrada de Jeffreys	p. 12
3	Estimação de Máxima Verossimilhança e Intervalar	p. 14
3.1	Estimação de Máxima Verossimilhança	p. 14
3.1.1	Função de Verossimilhança Perfilada	p. 14
3.1.2	Função de Verossimilhança Condicional	p. 22
3.1.3	Função de Verossimilhança Integrada Uniforme	p. 30
3.1.4	Função de Verossimilhança Integrada de Jeffreys	p. 37
3.1.5	Exemplo Ilustrativo	p. 42
3.2	Estimação Intervalar	p. 44
3.3	Exemplo Numérico	p. 47
4	Estudo de Simulação	p. 50
4.1	Estimação Pontual	p. 50
4.2	Estimação Intervalar	p. 58

5 Considerações Finais

p. 63

Referências

p. 65

1 *Introdução*

O processo de amostragem conhecido como captura-recaptura é frequentemente usado na estimação do número de indivíduos em uma dada região. Este processo consiste na seleção de um número fixo ou aleatório de indivíduos em diferentes épocas ou ocasiões de amostragem. Na primeira época de amostragem, os indivíduos selecionados são todos marcados, contados e devolvidos à população. Após um período de tempo que permita aos indivíduos marcados se misturarem aos não marcados na população, uma segunda amostra é selecionada na qual conta-se o número de indivíduos com e sem marcas, marca-se os que ainda não possuem marca e todos os indivíduos são devolvidos à população. O procedimento é repetido até se atingir um número fixado de épocas de amostragem.

Uma das primeiras aplicações do método de captura-recaptura foi feita por Laplace (1783) para estimar o tamanho da população da França. Mais tarde, Petersen (1896) aplicou o método para estudar o fluxo migratório de peixes no mar Báltico e, independentemente, Lincoln (1930) aplicou o mesmo método na estimação do número de patos selvagens na América do Norte. O método utilizado por estes pesquisadores baseia-se em um experimento com duas épocas de amostragem e o estimador pontual utilizado para estimar o tamanho populacional ficou conhecido como estimador de “Lincoln-Petersen”. Este estimador é obtido ao se igualar a proporção de indivíduos marcados na segunda amostra com a proporção de indivíduos marcados na população imediatamente antes da segunda seleção.

A partir da década de 1950 foram publicados vários artigos científicos relevantes sobre o tema entre eles Chapman (1954), Darroch (1958), Darroch (1959), Seber (1965), Jolly (1965) e Cormack (1968). Atualmente, os métodos de captura-recaptura tem aplicações nas mais diferentes áreas do conhecimento tais como Confiabilidade de Software (NAYAK, 1988; BASU; EBRAHIMI, 2001), Epidemiologia (SEBER; HUAKAU; SIMMONS, 2000; LEE et al., 2001; CHAO et al., 2001; LEE, 2002), Linguística (BOENDER; RINOY KAN, 1987; THISTED; EFRON, 1987) entre outras. Na literatura é possível encontrar várias revisões dos diferentes modelos e aplicações do processo de captura-recaptura (SEBER, 1982; SEBER,

1986; SEBER, 1992; SCHWARZ; SEBER, 1999; WHITE; GARROT, 1990; POLLOCK, 1991; POLLOCK, 2000; CHAO, 2001; AMSTRUP; MCDONALD; MANLY, 2003).

Os modelos estatísticos para estimação do tamanho populacional a partir do processo de captura-recaptura distinguem-se entre si, principalmente, pelas suposições adotadas. A suposição básica que diferencia os modelos é a de que a população é fechada, isto é, o número de indivíduos não se altera durante o processo de amostragem, não havendo portanto migração, mortes ou nascimentos. Quando esta suposição não é satisfeita, dizemos que a população é aberta. Outras suposições usualmente adotadas pelos modelos mais restritivos são: (i) os indivíduos não perdem suas marcas durante o experimento; (ii) todas as marcas são observadas e registradas corretamente; (iii) os indivíduos tem a mesma probabilidade de serem capturados em qualquer época de amostragem, isto é, o processo de captura e marcação não altera as probabilidades de captura. Otis et al. (1978) discute 8 modelos para população fechada em que a suposição (iii) é flexibilizada, combinando três tipos de variação para as probabilidades de captura: (1) as probabilidades de captura variam de acordo com a época de amostragem; (2) a probabilidade de captura de um indivíduo se altera quando este indivíduo já foi capturado; (3) as probabilidades de capturam variam de indivíduo para indivíduo.

Para muitos modelos de captura-recaptura, o interesse principal reside na estimação do tamanho populacional. Assim, as probabilidades de captura tornam-se parâmetros perturbadores ou *nuisance*, mas que são imprescindíveis na construção dos modelos probabilísticos. Neste contexto, o estudo de funções de verossimilhança que dependam apenas do parâmetro de interesse torna-se importante. O uso de métodos para eliminação de parâmetros perturbadores são amplamente discutidos na literatura estatística (COX, 1975; BERGER; LISEO; WOLPERT, 1999; BASU, 1977).

Neste trabalho, realizamos um estudo das funções de verossimilhança perfilada, condicional, integrada uniforme e integrada de Jeffreys no contexto de um modelo de captura-recaptura para estimação do tamanho populacional. Apresentamos resultados que caracterizam o comportamento destas funções de verossimilhança, permitindo assim a derivação de expressões para os estimadores de máxima verossimilhança e de métodos para a construção de intervalos de confiança baseados somente nas funções de verossimilhança. Os métodos são ilustrados para um conjunto de dados reais amplamente discutido na literatura. As propriedades frequentistas dos estimadores pontuais e intervalares serão comparadas a partir de um estudo de simulação.

2 *Modelo Binomial*

Neste capítulo discutimos como os diferentes métodos de eliminação de parâmetros perturbadores (*nuisance*) considerados no Capítulo 1 podem ser usados na obtenção de estimativas pontuais e intervalares para o tamanho populacional, considerando um modelo de captura-recaptura com marcação para uma população fechada.

2.1 Modelo Estatístico

Consideremos um processo de captura-recaptura em que os indivíduos são selecionados da população em k ocasiões de amostragem, $k \geq 2$. Em cada ocasião conta-se o número de indivíduos recapturados (já possuem marca) e o número total de indivíduos, efetua-se a marcação dos indivíduos sem marca e todos os indivíduos são devolvidos à população. Uma próxima captura é realizada após permitir que os indivíduos marcados e não marcados se distribuam na população.

Denotemos por N o número de indivíduos na população. Suponhamos que, em cada ocasião, um indivíduo seja capturado independentemente e com a mesma probabilidade dos demais e que as ocasiões de captura sejam independentes entre si, isto é, o fato de um indivíduo ser capturado ou não em uma ocasião não influencia na probabilidade de captura em outra ocasião.

Neste contexto, denotemos por

- p_i a probabilidade de um indivíduo ser capturado na i -ésima ocasião;
- n_i o número total de indivíduos capturados na i -ésima ocasião;
- m_i o número de indivíduos recapturados na i -ésima ocasião ($m_1 = 0$);
- $M_i = \sum_{j=1}^{i-1} (n_j - m_j)$ o número de indivíduos distintos marcados presentes na população imediatamente antes da i -ésima captura ($M_1 = 0$);

- $r = M_{k+1} = \sum_{j=1}^k (n_j - m_j)$ o número de indivíduos distintos capturados durante o processo;

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

Portanto, dado o vetor de parâmetros $\theta = (N, \mathbf{p})$, onde N é o parâmetro de interesse e $\mathbf{p} = (p_1, \dots, p_k)$ o vetor de parâmetros *nuisance*, a distribuição de probabilidades dos dados amostrais $\mathbf{n} = (n_1, \dots, n_k)$ e $\mathbf{m} = (m_1, \dots, m_k)$ é dada por

$$\begin{aligned}
P_\theta\{\mathbf{n}, \mathbf{m}\} &= P_\theta\{n_1, m_1\}P_\theta\{n_2, m_2|n_1, m_1\} \cdots P_\theta\{n_k, m_k|n_1, m_1; n_2, m_2; \dots; n_{k-1}, m_{k-1}\} \\
&= \prod_{i=1}^k \binom{N - M_i}{n_i - m_i} p_i^{n_i - m_i} (1 - p_i)^{N - M_i - n_i + m_i} \binom{M_i}{m_i} p_i^{m_i} (1 - p_i)^{M_i - m_i} \\
&= \prod_{i=1}^k \binom{M_i}{m_i} \times \prod_{i=1}^k \binom{N - M_i}{n_i - m_i} p_i^{n_i} (1 - p_i)^{N - n_i}.
\end{aligned} \tag{2.1}$$

Como $M_{i+1} = M_i + n_i - m_i$ para $i = 1, \dots, k$, então

$$\begin{aligned}
\prod_{i=1}^k \binom{N - M_i}{n_i - m_i} &= \prod_{i=1}^k \frac{(N - M_i)!}{(n_i - m_i)!(N - M_i - n_i + m_i)!} \\
&= \prod_{i=1}^k \frac{1}{(n_i - m_i)!} \prod_{i=1}^k \frac{(N - M_i)!}{(N - M_{i+1})!} \\
&= \left[\prod_{i=1}^k \frac{1}{(n_i - m_i)!} \right] \times \frac{N!}{(N - r)!}.
\end{aligned} \tag{2.2}$$

Substituindo a relação (2.2) em (2.1), segue que a função de verossimilhança de $\theta = (N, \mathbf{p})$ é dada por

$$\begin{aligned}
L(N, \mathbf{p}) &= P_\theta\{\mathbf{n}, \mathbf{m}\} \\
&= \left[\prod_{i=1}^k \frac{M_i!}{m_i!(M_i - m_i)!(n_i - m_i)!} \right] \times \frac{N!}{(N - r)!} \prod_{i=1}^k p_i^{n_i} (1 - p_i)^{N - n_i},
\end{aligned} \tag{2.3}$$

$N \geq r$, $0 < p_i < 1$, $i = 1, \dots, k$.

Portanto, o núcleo da função de verossimilhança é dado por

$$K(N, \mathbf{p}) = \frac{N!}{(N - r)!} \prod_{i=1}^k p_i^{n_i} (1 - p_i)^{N - n_i}, \tag{2.4}$$

e seu logaritmo dado por

$$\log(K(N, \mathbf{p})) = \log(N!) - \log((N-r)!) + \sum_{i=1}^k n_i \log(p_i) + \sum_{i=1}^k (N-n_i) \log(1-p_i), \quad (2.5)$$

$$N \geq r, 0 < p_i < 1, i = 1, \dots, k.$$

Salientamos que o núcleo da função de verossimilhança (2.4) depende dos dados amostrais somente através de \mathbf{n} e r , isto é, (n, r) é uma estatística suficiente para (N, p) .

2.2 Funções de Verossimilhança

Nesta seção vamos aplicar os métodos de eliminação de parâmetros perturbadores discutidos no Capítulo 1 à função de verossimilhança (2.3) e obter funções de verossimilhança que dependem exclusivamente do parâmetro de interesse N .

2.2.1 Função de Verossimilhança Perfilada

A função de verossimilhança perfilada de N é obtida a partir da expressão da função de verossimilhança (2.3) substituindo-se o vetor de parâmetros \mathbf{p} por sua estimativa de máxima verossimilhança para cada N fixado. Então, a função de verossimilhança perfilada é obtida como

$$\begin{aligned} L^{\mathcal{P}}(N) &= \sup_{\mathbf{p} \in [0,1]^k} L(N, \mathbf{p}) \\ &= L(N, \hat{\mathbf{p}}(N)) \\ &\propto K(N, \hat{\mathbf{p}}(N)), \end{aligned}$$

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

A partir de (2.5) segue que

$$\frac{\partial \log(K(N, p))}{\partial p_i} = \frac{n_i}{p_i} - \frac{N-n_i}{1-p_i} = 0, \quad i = 1, \dots, k,$$

o que implica

$$\hat{\mathbf{p}}(N) = \left(\frac{n_1}{N}, \dots, \frac{n_k}{N} \right), \quad N \geq r. \quad (2.6)$$

Substituindo (2.6) em (2.4), segue que o núcleo da função de verossimilhança perfilada

é dado por

$$K^{\mathcal{P}}(N) = \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{(N-n_i)^{N-n_i}}{N^{N-n_i}}, \quad N \geq r. \quad (2.7)$$

2.2.2 Função de Verossimilhança Condicional

A função de verossimilhança condicional é obtida a partir da fatoração da função de verossimilhança (2.3) como

$$L(N, \mathbf{p}) = L^{\mathcal{C}}(N) \times L'(N, \mathbf{p}),$$

onde $L^{\mathcal{C}}(N)$ é a função de verossimilhança correspondente à distribuição de probabilidades condicional de $\mathbf{m} = (m_1, m_2, \dots, m_k)$, dado $\mathbf{n} = (n_1, n_2, \dots, n_k)$, e $L'(N, \mathbf{p})$ a função de verossimilhança correspondente à distribuição de probabilidades de (n_1, n_2, \dots, n_k) dados N e \mathbf{p} .

Então, segue que $L'(N, \mathbf{p})$ é dada por

$$\begin{aligned} L'(N, \mathbf{p}) &= P_{\theta}\{n_1, \dots, n_k\} \\ &= \prod_{i=1}^k \binom{N}{n_i} p_i^{n_i} (1-p_i)^{N-n_i}, \end{aligned}$$

$N \geq \max\{n_1, n_2, \dots, n_k\}$, $0 < p_i < 1$, $i = 1, \dots, k$ e segue de (2.3) que $L'(N)$ é dada por

$$\begin{aligned} L'(N) &= P_{\theta}\{m_1, \dots, m_k | n_1, \dots, n_k\} \\ &= \frac{P_{\theta}\{m_1, n_1; \dots; m_k, n_k\}}{P_{\theta}\{n_1, \dots, n_k\}} \\ &= \frac{\left[\prod_{i=1}^k \frac{M_i!}{m_i!(M_i-m_i)!(n_i-m_i)!} \right] \times \frac{N!}{(N-r)!} \prod_{i=1}^k p_i^{n_i} (1-p_i)^{N-n_i}}{\prod_{i=1}^k \binom{N}{n_i} p_i^{n_i} (1-p_i)^{N-n_i}} \\ &= \left[\prod_{i=1}^k \frac{M_i! n_i!}{m_i!(M_i-m_i)!(n_i-m_i)!} \right] \times \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{(N-n_i)!}{N!}, \end{aligned} \quad (2.8)$$

$N \geq r$.

Portanto, segue de (2.8) que o núcleo da função de verossimilhança condicional é dado por

$$K^{\mathcal{C}}(N) = \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{(N-n_i)!}{N!}, \quad N \geq r. \quad (2.9)$$

2.2.3 Função de Verossimilhança Integrada Uniforme

Com o objetivo de eliminar o parâmetro *nuisance* \mathbf{p} , suponhamos que, dado N , p_1, \dots, p_k sejam variáveis aleatórias independentes com mesma distribuição Uniforme $(0, 1)$. Assim, a função de verossimilhança integrada uniforme é obtida a partir da integração da função de verossimilhança (2.3), isto é,

$$\begin{aligned} L^u(N) &= \int_{(0,1)^k} L(N, \mathbf{p}) d\mathbf{p} \\ &\propto \frac{N!}{(N-r)!} \prod_{i=1}^k \int_0^1 p_i^{n_i} (1-p_i)^{N-n_i} dp_i \\ &= \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{\Gamma(n_i+1)\Gamma(N-n_i+1)}{\Gamma(N+2)} \\ &\propto \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{(N-n_i)!}{(N+1)!}, \quad N \geq r. \end{aligned}$$

Portanto, o núcleo da função de verossimilhança integrada uniforme é dado por

$$K^u(N) = \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{(N-n_i)!}{(N+1)!}, \quad N \geq r. \quad (2.10)$$

2.2.4 Função de Verossimilhança Integrada de Jeffreys

Novamente, vamos considerar a eliminação do parâmetro *nuisance* \mathbf{p} via integração. Suponhamos que, dado N , \mathbf{p} seja um vetor aleatório com distribuição de Jeffreys $\pi^J(\mathbf{p}|N)$. A função densidade de probabilidade de Jeffreys é definida como sendo proporcional a raiz quadrada do determinante da matriz de informação esperada de Fisher $I_N(\mathbf{p})$ (supondo N conhecido), isto é,

$$\begin{aligned} I_N(\mathbf{p}) &= \left(\mathbb{E} \left[- \frac{\partial^2 \log L(N, \mathbf{p})}{\partial p_i \partial p_j} \right] \right)_{k \times k} \\ &\propto \left(\mathbb{E} \left[- \frac{\partial^2 \log K(N, \mathbf{p})}{\partial p_i \partial p_j} \right] \right)_{k \times k}. \end{aligned}$$

A partir de (2.5), temos que

$$\frac{\partial \log K(N, \mathbf{p})}{\partial p_i \partial p_j} = \begin{cases} 0, & \text{se } i \neq j, \\ \frac{-n_j + 2n_j p_j - N p_j^2}{p_j^2 (1 - p_j)^2}, & \text{se } i = j, \end{cases} \quad 1 \leq i, j \leq k,$$

o que implica

$$E \left[- \frac{\partial \log L(N, \mathbf{p})}{\partial p_i \partial p_j} \right] = \begin{cases} 0, & \text{se } i \neq j, \\ \frac{N}{p_i (1 - p_i)}, & \text{se } i = j, \end{cases} \quad 1 \leq i, j \leq k.$$

Portanto, a função densidade de probabilidade de Jeffreys é dada por

$$\begin{aligned} \pi^{\mathcal{J}}(\mathbf{p}|N) &\propto \left[\det I_N(\mathbf{p}) \right]^{\frac{1}{2}} \\ &\propto \prod_{i=1}^k \frac{1}{p_i^{1/2} (1 - p_i)^{1/2}}. \end{aligned} \quad (2.11)$$

Portanto, a função de verossimilhança integrada de Jeffreys é dada por

$$\begin{aligned} L^{\mathcal{J}}(N) &= \int_{(0,1)^k} L(N, \mathbf{p}) \pi^{\mathcal{J}}(\mathbf{p}|N) d\mathbf{p} \\ &\propto \frac{N!}{(N-r)!} \prod_{i=1}^k \int_0^1 p_i^{n_i-1/2} (1-p_i)^{N-n_i-1/2} dp_i \\ &= \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{\Gamma(n_i + 1/2) \Gamma(N - n_i + 1/2)}{\Gamma(N + 1)}, \end{aligned}$$

$N \geq r$.

Logo, o núcleo da função de verossimilhança integrada de Jeffreys é dado por

$$K^{\mathcal{J}}(N) = \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{\Gamma(N - n_i + 1/2)}{N!}, \quad N \geq r. \quad (2.12)$$

3 *Estimação de Máxima Verossimilhança e Intervalar*

Neste capítulo apresentamos métodos para obtenção das estimativas de máxima verossimilhança e para a construção de intervalos de confiança para o tamanho populacional, baseados nas funções de verossimilhança apresentadas no Capítulo 2. No que segue vamos supor que o processo consista em pelo menos duas ocasiões de captura ($k \geq 2$) e que, em cada ocasião, haja pelo menos um indivíduo capturado. Denotemos por $m = \max\{n_1, \dots, n_k\}$ o número máximo de indivíduos capturados em uma mesma ocasião e $n = n_1 + \dots + n_k$ o número total de capturas realizadas. Observemos que o número de indivíduos distintos capturados r é tal que $m \leq r \leq n$.

3.1 Estimação de Máxima Verossimilhança

Nesta seção, apresentamos resultados que caracterizam o comportamento das funções de verossimilhança do Capítulo 2, além de resultados a respeito da obtenção das estimativas de máxima verossimilhança.

3.1.1 Função de Verossimilhança Perfilada

Devido ao caráter discreto do parâmetro de interesse N , se existir um ponto de máximo, $(\hat{N}, \hat{\mathbf{p}})$, do núcleo da função de verossimilhança (2.4), então \hat{N} maximiza o núcleo da função de verossimilhança perfilada (2.7).

De fato, suponhamos que $(\hat{N}, \hat{\mathbf{p}})$ maximize o núcleo (2.4) e que, para cada N fixado, $N \geq r$, exista um único ponto, $\hat{\mathbf{p}}(N)$, que maximize a função $h_N(\mathbf{p}) = K(N, \mathbf{p})$, $\mathbf{p} \in [0, 1]^k$. Então, os possíveis candidatos a ponto de máximo de $K(N, \mathbf{p})$ são os pontos do conjunto $\{(N, \hat{\mathbf{p}}(N)) : N \geq r\}$, o que implica que \hat{N} é um ponto de máximo do núcleo da verossimilhança perfilada (2.7).

Portanto, a estimação de máxima verossimilhança do parâmetro de interesse N a partir do núcleo (2.4) é equivalente à obtida a partir do núcleo (2.7).

Proposição 3.1. *O núcleo da função de verossimilhança perfilada, $K^{\mathcal{P}}(N)$, converge para 0 se $m \leq r < n$ e converge para $\exp\{-n\}$ se $r = n$, quando $N \rightarrow \infty$.*

Prova:

A partir de (2.7) segue que

$$\begin{aligned} K^{\mathcal{P}}(N) &= \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{(N-n_i)^{N-n_i}}{N^N} \\ &= N(N-1)\cdots(N-r+1) \prod_{i=1}^k \frac{(N-n_i)^{N-n_i}}{N^{N-n_i} N^{n_i}} \\ &= \frac{N^r}{N^n} \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{r-1}{N}\right) \prod_{i=1}^k \left(1 - \frac{n_i}{N}\right)^N \left(1 - \frac{n_i}{N}\right)^{-n_i}. \end{aligned}$$

Como

$$\lim_{N \rightarrow \infty} \left(1 - \frac{x}{N}\right) = 1, \text{ para todo } x \text{ real} \quad (3.1)$$

e

$$\lim_{N \rightarrow \infty} \left(1 - \frac{x}{N}\right)^N = \exp\{-x\}, \text{ para todo } x > 0, \quad (3.2)$$

então

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

■

Para provar os próximos resultados, vamos utilizar o Lema a seguir.

Lema 3.1. *Para qualquer $y > 0$ e $t \in \{1, 2, \dots\}$, temos*

(A)

$$\log(y+t) - \log(y) > \sum_{i=0}^{t-1} \frac{1}{y+t-i};$$

(B)

$$\begin{aligned}\log(y+t) - \log(y) &< \frac{1}{2(y+t)} + \sum_{i=1}^t \frac{1}{y+t-i} - \frac{1}{2y} \\ &= \frac{1}{2(y+t)} + \sum_{i=1}^{t-1} \frac{1}{y+t-i} + \frac{1}{2y}.\end{aligned}$$

Prova:

Consideremos $y > 0$, fixado.

(A)

Para provar esta afirmação basta notar que

$$\begin{aligned}\log(y+t) - \log(y) &= \int_y^{y+t} \frac{1}{x} dx \\ &= \int_y^{y+1} \frac{1}{x} dx + \dots + \int_{y+t-1}^{y+t} \frac{1}{x} dx \\ &> \frac{1}{y+1} + \dots + \frac{1}{y+t}.\end{aligned}$$

(B)

Primeiramente, vamos mostrar a afirmação para $t = 1$, isto é,

$$\log(y+1) - \log(y) < \frac{1}{2} \left(\frac{1}{y+1} + \frac{1}{y} \right).$$

Neste sentido, consideremos a função $f(x) = 1/x$ definida para $x > 0$. Como f é convexa, o segmento de reta que une os pontos $(y, 1/y)$ e $(y+1, 1/(y+1))$ fica acima do gráfico de f no intervalo $(y, y+1)$, isto é,

$$\frac{1}{y+\xi} < \frac{1}{y} + \left(\frac{1}{y+1} - \frac{1}{y} \right) \xi, \quad 0 < \xi < 1.$$

Então,

$$\begin{aligned}
 \log(y+1) - \log(y) &= \int_0^1 \frac{1}{y+\xi} d\xi \\
 &< \int_0^1 \left[\frac{1}{y} + \left(\frac{1}{y+1} - \frac{1}{y} \right) \xi \right] d\xi \\
 &= \frac{1}{2} \left(\frac{1}{y+1} + \frac{1}{y} \right)
 \end{aligned} \tag{3.3}$$

e para o caso em que $t \geq 2$, observemos que

$$\log(y+t) - \log(y) = \sum_{i=1}^t \left[\log(y+t+1-i) - \log(y+t-i) \right].$$

Logo, segue de (3.3) que

$$\begin{aligned}
 \log(y+t) - \log(y) &< \sum_{i=1}^t \left[\frac{1}{2} \left(\frac{1}{y+t-i+1} + \frac{1}{y+t-i} \right) \right] \\
 &= \frac{1}{2(y+t)} + \sum_{i=1}^{t-1} \frac{1}{y+t-i} + \frac{1}{2y} \\
 &= \frac{1}{2(y+t)} + \sum_{i=1}^t \frac{1}{y+t-i} - \frac{1}{2y}.
 \end{aligned}$$

■

Para provar os resultados que seguem, consideramos o logaritmo do núcleo da função de verossimilhança perfilada (2.7), que é dado por

$$\log(K^{\mathcal{P}}(N)) = \sum_{i=0}^{r-1} \log(N-i) + \sum_{i=1}^k \left[(N-n_i) \log(N-n_i) - N \log(N) \right], \quad N \geq r.$$

A extensão desta função para valores reais é dada por

$$f(x) = \sum_{i=0}^{r-1} \log(x-i) + \sum_{i=1}^k \left[(x-n_i) \log(x-n_i) - x \log(x) \right], \quad x \geq r, \tag{3.4}$$

onde supomos $0 \log(0) = 0$.

Observemos que a função definida em (3.4) é contínua e derivável no intervalo (r, ∞) , sendo $\lim_{x \downarrow r} f(x) = f(r)$, isto é, f é contínua à direita em r . Portanto, o estudo do comportamento de $K^{\mathcal{P}}$ com relação ao crescimento ou decrescimento pode ser realizado através

do estudo do sinal da derivada primeira de f , cuja expressão é dada por

$$f'(x) = \sum_{i=0}^{r-1} \frac{1}{x-i} + \sum_{i=1}^k \left[\log(x - n_i) - \log(x) \right], \quad x > r. \quad (3.5)$$

Proposição 3.2. *Se $r = m$, então a função $K^{\mathcal{P}}$ é estritamente decrescente, isto é, $K^{\mathcal{P}}(N + 1) < K^{\mathcal{P}}(N)$, para todo $N \geq r$. Portanto, a estimativa de máxima verossimilhança perfilada é $\widehat{N}^{\mathcal{P}} = m$.*

Prova:

Suponhamos, sem perda de generalidade, que $m = n_1$. Para mostrar que $K^{\mathcal{P}}$ é estritamente decrescente é suficiente mostrar, usando a expressão (3.5) com $r = n_1$, que $f'(x) < 0$ para todo $x > n_1$, ou seja,

$$f'(x) = \sum_{i=0}^{n_1-1} \frac{1}{x-i} + \sum_{i=1}^k \left[\log(x - n_i) - \log(x) \right] < 0, \quad \text{para } x > n_1. \quad (3.6)$$

Do Lema 3.1 (A), segue que

$$\sum_{i=0}^{n_1-1} \frac{1}{x-i} < \log(x) - \log(x - n_1),$$

o que prova o resultado. ■

Proposição 3.3. *Se $r = n$, então a função $K^{\mathcal{P}}$ é estritamente crescente, isto é, $K^{\mathcal{P}}(N + 1) > K^{\mathcal{P}}(N)$ para todo $N \geq r$. Portanto, a estimativa de máxima verossimilhança perfilada é $\widehat{N}^{\mathcal{P}} = +\infty$.*

Prova:

Para mostrar que $K^{\mathcal{P}}$ é estritamente crescente basta mostrar que, usando a expressão (3.5) com $r = n$, $f'(x) > 0$ para todo $x > n$, ou seja,

$$f'(x) = \sum_{i=0}^{n-1} \frac{1}{x-i} + \sum_{i=1}^k \left[\log(x - n_i) - \log(x) \right] > 0, \quad x > n,$$

ou equivalentemente, basta mostrar que, para $k \geq 2$, vale

$$\sum_{i=1}^k \left[\log(x) - \log(x - n_i) \right] < \sum_{i=0}^{n_1+n_2+\dots+n_k-1} \frac{1}{x-i}, \quad (3.7)$$

para quaisquer $n_1, n_2, \dots, n_k \in \mathbb{N}^* = \{1, 2, \dots\}$ e qualquer real $x > n_1 + n_2 + \dots + n_k$. A relação (3.7) será provada por indução finita sobre k .

Primeiramente, observemos pelo Lema 3.1 (B) que

$$\log(x) - \log(x - n_i) < \frac{1}{2x} + \sum_{j=1}^{n_i} \frac{1}{x-j} - \frac{1}{2(x-n_i)} \quad (3.8)$$

$$= \frac{1}{2x} + \sum_{j=1}^{n_i-1} \frac{1}{x-j} + \frac{1}{2(x-n_i)}, \quad (3.9)$$

para $x > n_i$, $i = 1, 2, \dots, k$.

Supondo $k = 2$ e, sem perda de generalidade, que $n_1 \geq n_2$, segue de (3.8) e (3.9) que

$$\begin{aligned} \sum_{i=1}^2 \left[\log(x) - \log(x - n_i) \right] &< \frac{1}{2x} + \sum_{j=1}^{n_1} \frac{1}{x-j} - \frac{1}{2(x-n_1)} \\ &+ \frac{1}{2x} + \sum_{j=1}^{n_2-1} \frac{1}{x-j} + \frac{1}{2(x-n_2)} \\ &= \frac{1}{x} + \sum_{j=1}^{n_1} \frac{1}{x-j} + \sum_{j=1}^{n_2-1} \frac{1}{x-j} + \frac{1}{2} \left(\frac{1}{x-n_2} - \frac{1}{x-n_1} \right) \\ &< \frac{1}{x} + \sum_{j=1}^{n_1} \frac{1}{x-j} + \sum_{j=1}^{n_2-1} \frac{1}{x-n_1-j} \\ &= \sum_{j=0}^{n_1+n_2-1} \frac{1}{x-j}, \end{aligned}$$

$x > n_1 + n_2$, o que resulta (3.7) com $k = 2$.

Agora, suponhamos por indução que a afirmação (3.7) valha para $k = k'$, $k' \geq 2$. Sob esta hipótese, provamos que a afirmação (3.7) vale para $k = k' + 1$.

Sejam $n_1, \dots, n_{k'}, n_{k'+1}$ pertencentes a $\mathbb{N}^* = \{1, 2, \dots\}$ e $x > n_1 + \dots + n_{k'} + n_{k'+1}$. Como $x > n_1 + \dots + n_{k'}$, segue da hipótese de indução, que

$$\sum_{i=1}^{k'} \left[\log(x) - \log(x - n_i) \right] < \sum_{j=0}^{n_1+n_2+\dots+n_{k'}-1} \frac{1}{x-j},$$

e pelo Lema 3.1 (B) segue que

$$\begin{aligned}
\sum_{i=1}^{k'+1} \left[\log(x) - \log(x - n_i) \right] &= \sum_{i=1}^{k'} \left[\log(x) - \log(x - n_i) \right] + \log(x) - \log(x - n_{k'+1}) \\
&< \sum_{j=0}^{n_1+n_2+\dots+n_{k'}-1} \frac{1}{x-j} + \sum_{j=1}^{n_{k'+1}} \frac{1}{x-j} + \frac{1}{2} \left(\frac{1}{x} - \frac{1}{x - n_{k'+1}} \right) \\
&< \sum_{j=0}^{n_1+n_2+\dots+n_{k'}-1} \frac{1}{x-j} + \sum_{j=1}^{n_{k'+1}} \frac{1}{(x - n_1 - \dots - n_{k'} + 1) - j} \\
&= \sum_{j=0}^{n_1+n_2+\dots+n_{k'+1}-1} \frac{1}{x-j},
\end{aligned}$$

o que implica (3.7) para $k = k' + 1$. ■

Proposição 3.4. *Para $m < r < n$, definamos*

$$x_0 = \sup \left\{ x \in (0, 1/r] : \sum_{j=1}^{r-1} \frac{j}{1/x - j} < n - r \right\}.$$

(a) Se $x_0 = 1/r$, então $\widehat{N}^{\mathcal{P}} = r$;

(b) Se $0 < x_0 < 1/r$, então $\widehat{N}^{\mathcal{P}}$ é o ponto de máximo de $K^{\mathcal{P}}(N)$ restrito ao conjunto $\{r, r+1, \dots, [1/x_0] + 1\}$.

Prova:

Observemos que $K^{\mathcal{P}}(N)$, dada em (2.7), pode ser escrita como

$$K^{\mathcal{P}}(N) = \frac{1}{N^{n-r}} \prod_{i=1}^{r-1} \left(1 - \frac{i}{N} \right) \prod_{i=1}^k \left(1 - \frac{n_i}{N} \right)^{N-n_i}, \quad N \geq r.$$

Consideremos a função f definida por

$$f(x) = g_r(x)h(x), \quad 0 < x < 1/r,$$

onde as funções g_r e h são definidas por

$$g_r(x) = x^{n-r} \prod_{i=1}^{r-1} (1 - ix), \quad 0 < x \leq 1/r,$$

$$h(x) = \prod_{i=1}^k (1 - n_i x)^{1/x - n_i}, \quad 0 < x \leq 1/r,$$

o que implica

$$f\left(\frac{1}{N}\right) = g_r\left(\frac{1}{N}\right)h\left(\frac{1}{N}\right) = K^{\mathcal{P}}(N), \quad N \geq r.$$

A função g_r é contínua, estritamente positiva no intervalo $(0, 1/r)$, $\lim_{x \downarrow 0} g_r(x) = 0$ e existe um ponto x_0 , $0 < x_0 \leq 1/r$, tal que $g'_r(x) > 0$ se $0 < x < x_0$. Com efeito,

$$g'_r(x) = (n-r)x^{n-r-1} \prod_{i=1}^{r-1} (1-ix) + x^{n-r} \sum_{j=1}^{r-1} (-j) \prod_{i \neq j} (1-ix)$$

$$= x^{n-r-1} \left[(n-r) \prod_{i=1}^{r-1} (1-ix) + x \sum_{j=1}^{r-1} (-j) \prod_{i \neq j} (1-ix) \right], \quad 0 < x < 1/r,$$

o que implica

$$g'_r(x) > 0 \iff x \sum_{j=1}^{r-1} j \prod_{i \neq j} (1-ix) < (n-r) \prod_{i=1}^{r-1} (1-ix) \iff \sum_{j=1}^{r-1} \frac{j}{1/x - j} < n-r.$$

Observemos que a função $q(x) = \sum_{j=1}^{r-1} \frac{j}{1/x - j}$ é estritamente crescente em $(0, 1/r]$ e $\lim_{x \downarrow 0} q(x) = 0$. Então, $g'_r(x) > 0$, se e somente se, $0 < x < x_0$, onde

$$x_0 = \sup \left\{ x \in (0, 1/r] : \sum_{j=1}^{r-1} \frac{j}{1/x - j} < n-r \right\}.$$

Notemos que $x_0 = 1/r$ se e somente se $q(1/r) \leq n-r$.

Por outro lado, a função h é contínua, positiva e estritamente crescente no intervalo $(0, 1/r]$ e $\lim_{x \downarrow 0} h(x) = e^{-n}$. Para mostrar que $h(x)$ é estritamente crescente basta mostrar que $\log(h(x))$ é crescente em $(0, 1/r]$. Neste sentido, para $0 < x < 1/r$

$$\log(h(x)) = \sum_{i=1}^k \left(1/x - n_i \right) \log(1 - n_i x),$$

o que implica

$$\begin{aligned} \left(\log(h(x)) \right)' &= \sum_{i=1}^k \left\{ -\frac{1}{x^2} \log(1 - n_i x) - \left(\frac{1}{x} - n_i \right) \frac{n_i}{1 - n_i x} \right\} \\ &= -\frac{1}{x^2} \left\{ \sum_{i=1}^k \log(1 - n_i x) + nx \right\} \\ &> -\frac{1}{x^2} \left\{ \sum_{i=1}^k (-n_i x) + \sum_{i=1}^k n_i x \right\} = 0, \end{aligned}$$

onde a última desigualdade segue de $e^{-y} > 1 - y$ para y real.

Pelas considerações feitas acima, concluímos que $f(x)$ é estritamente crescente no intervalo $(0, x_0)$, pois $f'(x) = g'_r(x)h(x) + g_r(x)h'(x) > 0$, para $0 < x < x_0$, o que implica

$$K^{\mathcal{P}}(N+1) = f\left(\frac{1}{N+1}\right) < f\left(\frac{1}{N}\right) = K^{\mathcal{P}}(N), \quad \text{para } N \geq 1/x_0.$$

Logo, se $x_0 = 1/r$, então $\widehat{N}^{\mathcal{P}} = r$, o que prova o item (a). Se $0 < x_0 < 1/r$, então $\widehat{N}^{\mathcal{P}}$ é o ponto de máximo de $K^{\mathcal{P}}(N)$ restrito ao conjunto $\{r, r+1, \dots, [1/x_0] + 1\}$, o que prova o item (b). ■

3.1.2 Função de Verossimilhança Condicional

Proposição 3.5. *O núcleo da função de verossimilhança condicional, $K^{\mathcal{C}}(N)$, converge para 0 se $r < n$ e converge para 1 se $r = n$, quando $N \rightarrow \infty$.*

Prova:

A partir de (2.9) segue que

$$\begin{aligned} K^{\mathcal{C}}(N) &= N(N-1)\cdots(N-r+1) \prod_{j=1}^k \frac{1}{N(N-1)\cdots(N-n_j+1)} \\ &= \frac{N^r}{N^n} \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{r-1}{N}\right) \prod_{i=1}^k \frac{1}{\left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{n_i-1}{N}\right)}, \end{aligned}$$

para $N \geq r$.

Logo, do fato

$$\lim_{N \rightarrow \infty} \frac{N^r}{N^n} = \begin{cases} 0, & \text{se } r < n, \\ 1, & \text{se } r = n, \end{cases}$$

segue o resultado. ■

No que segue, vamos considerar a razão

$$\begin{aligned} \frac{K^{\mathcal{C}}(N+1)}{K^{\mathcal{C}}(N)} &= \frac{(N+1)}{(N+1-r)} \prod_{i=1}^k \frac{N+1-n_i}{N+1} \\ &= \left[1 - \frac{r}{N+1}\right]^{-1} \prod_{i=1}^k \left(1 - \frac{n_i}{N+1}\right), \quad N \geq r. \end{aligned} \quad (3.10)$$

Proposição 3.6. *Se $r = m$, então a função $K^{\mathcal{C}}$ é estritamente decrescente, isto é, $K^{\mathcal{C}}(N+1) < K^{\mathcal{C}}(N)$, para todo $N \geq r$. Portanto, a estimativa de máxima verossimilhança condicional é $\hat{N}^{\mathcal{C}} = m$.*

Prova:

Suponhamos, sem perda de generalidade, que $m = n_1$. Então, substituindo r por n_1 em (3.10), segue que

$$\frac{K^{\mathcal{C}}(N+1)}{K^{\mathcal{C}}(N)} = \prod_{i=2}^k \left(1 - \frac{n_i}{N+1}\right) < 1, \quad \text{para todo } N \geq r,$$

o que implica resultado. ■

Para provar o próximo resultado utilizamos o seguinte lema.

Lema 3.2. *Para todo inteiro $k \geq 2$ e quaisquer reais x_1, \dots, x_k tais que $0 < x_i < 1$, $1 \leq i \leq k$, temos $\prod_{i=1}^k (1 - x_i) > 1 - \sum_{i=1}^k x_i$*

Prova:

Para provar este resultado utilizamos indução finita sobre $k \geq 2$.

Para $k = 2$, temos que

$$\begin{aligned} (1 - x_1)(1 - x_2) &= 1 - x_1 - x_2 + x_1x_2 \\ &> 1 - x_1 - x_2, \end{aligned}$$

e portanto vale o resultado para $k = 2$.

Suponhamos agora que a tese valha para $k > 2$, isto é,

$$\prod_{i=1}^k (1 - x_i) > 1 - \sum_{i=1}^k x_i.$$

Então,

$$\begin{aligned} \prod_{i=1}^{k+1} (1 - x_i) &= (1 - x_{k+1}) \prod_{i=1}^k (1 - x_i) \\ &> (1 - x_{k+1}) \left(1 - \sum_{i=1}^k x_i \right) \\ &= 1 - \sum_{i=1}^k x_i - x_{k+1} + x_{k+1} \sum_{i=1}^k x_i \\ &> 1 - \sum_{i=1}^{k+1} x_i, \end{aligned}$$

o que prova o resultado. ■

Proposição 3.7. *Se $r = n$, então a função K^c é estritamente crescente, isto é, $K^c(N + 1) > K^c(N)$ para todo $N \geq r$. Portanto, a estimativa de máxima verossimilhança condicional é $\hat{N}^c = \infty$.*

Prova:

A partir do Lema 3.2 temos que

$$\prod_{i=1}^k \left(1 - \frac{n_i}{N + 1} \right) > 1 - \frac{n}{N + 1}, \quad N \geq r,$$

o que permite concluir

$$\frac{K^c(N + 1)}{K^c(N)} > 1, \quad N \geq r. \quad \blacksquare$$

Para demonstrar o próximo resultado, vamos definir a função

$$f_r(x) = (1 - rx)^{-1} \prod_{i=1}^k (1 - n_i x), \quad 0 \leq x < 1/r. \quad (3.11)$$

Observemos que, de acordo com (3.10), f_r satisfaz a seguinte relação

$$f_r\left(\frac{1}{N+1}\right) = \frac{K^c(N+1)}{K^c(N)}, \quad N \geq r. \quad (3.12)$$

Para estudar o crescimento (decréscimo) de K^c necessitamos determinar os valores de N para os quais ocorre $K^c(N+1) > K^c(N)$, $K^c(N+1) = K^c(N)$ ou $K^c(N+1) < K^c(N)$. A partir de (3.12), segue que para realizar este estudo é suficiente determinar para quais valores de x no intervalo $(0, 1/r)$ ocorre $f_r(x) < 1$, $f_r(x) = 1$ ou $f_r(x) > 1$. No entanto, antes de analisar f_r , será útil considerar o seguinte resultado a respeito de funções côncavas.

Lema 3.3. *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ (\mathbb{R} é o conjunto dos números reais) uma função contínua. Se f for côncava no intervalo $[a, b]$, $f(a) > 0$ e $f(b) < 0$, então existe um único ponto x_0 , $a < x_0 < b$ tal que $f(x_0) = 0$, $f(x) > 0$ se $a \leq x < x_0$ e $f(x) < 0$ se $x_0 < x \leq b$. Sob estas condições, x_0 é único.*

Prova:

Pela continuidade da função f , segue que existe x_0 , $a < x_0 < b$, tal que

$$f(x_0) = 0.$$

Pela concavidade de f no intervalo $[a, b]$, temos que o gráfico da função f entre quaisquer y_1 e y_2 , $a \leq y_1 < y_2 \leq b$, não tem nenhum ponto localizado abaixo do segmento de reta que une os pontos $(y_1, f(y_1))$ e $(y_2, f(y_2))$, ou seja,

$$f(y_1 + \alpha(y_2 - y_1)) \geq f(y_1) + \alpha[f(y_2) - f(y_1)], \quad (3.13)$$

para todo $0 < \alpha < 1$.

Logo, tomando $y_1 = a$ e $y_2 = x_0$ em (3.13), temos que

$$\begin{aligned} f(a + \alpha(x_0 - a)) &\geq f(a) - \alpha f(a) \\ &= (1 - \alpha)f(a) \\ &> 0, \quad \text{para todo } 0 < \alpha < 1, \end{aligned}$$

o que implica

$$f(x) > 0, \quad \text{para todo } a \leq x < x_0.$$

Para $x_0 < x < b$ devemos ter $f(x) < 0$. De fato, suponhamos que $f(x) \geq 0$ para algum $x_0 < x < b$. Então, segue de (3.13) com $y_1 = a$ e $y_2 = x$ que

$$\begin{aligned} f(a + \alpha(x - a)) &\geq f(a) + \alpha[f(x) - f(a)] \\ &= (1 - \alpha)f(a) + \alpha f(x) \\ &> 0, \end{aligned}$$

para todo $0 < \alpha < 1$ e para $\alpha = (x_0 - a)/(x - a)$ temos $f(x_0) > 0$, o que é absurdo. ■

Lema 3.4. *Para $m < r < n$, existe x_0 , $0 < x_0 < 1/r$, tal que $f_r(x_0) = 1$, $f_r(x) < 1$ se $0 < x < x_0$ e $f_r(x) > 1$ se $x_0 < x < 1/r$.*

Prova:

Observemos que, definindo

$$\begin{aligned} g(x) &= \prod_{i=1}^k (1 - n_i x), & x \in \mathbb{R} & \quad e \\ h_r(x) &= 1 - rx, & x \in \mathbb{R}, \end{aligned}$$

temos

$$f_r(x) = \frac{g(x)}{h_r(x)}, \quad 0 \leq x < 1/r.$$

As derivadas de primeira e segunda ordem de g são dadas por

$$\begin{aligned} g'(x) &= \sum_{i=1}^k (-n_i) \prod_{\substack{j=1 \\ j \neq i}}^k (1 - n_j x), \\ g''(x) &= \sum_{i=1}^k (-n_i) \sum_{\substack{j=1 \\ j \neq i}}^k (-n_j) \prod_{\substack{s=1 \\ s \neq i, j}}^k (1 - n_s x) \\ &= \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k n_i n_j \prod_{\substack{s=1 \\ s \neq i, j}}^k (1 - n_s x) \end{aligned}$$

e como $1 - n_i x > 0$ para $0 \leq x \leq 1/r$, $i = 1, \dots, n$, temos

$$\begin{aligned} g'(x) &< 0, & \text{para todo } 0 \leq x \leq 1/r, \\ g''(x) &> 0, & \text{para todo } 0 \leq x \leq 1/r. \end{aligned}$$

Portanto, as funções g , g' e g'' são funções contínuas e g também é estritamente decrescente e convexa no intervalo $[0, 1/r]$, com $g(0) = 1$ e $g(1/r) > 0$. Por outro lado, a função h_r é linear e estritamente decrescente no intervalo $[0, 1/r]$ com $h_r(0) = 1$ e $h_r(1/r) = 0$. A Figura 1 ilustra o comportamento das funções g e h_r no intervalo $[0, 1/r]$ e, a partir da observação deste gráfico, podemos concluir que deve existir um único ponto x_0 , $0 < x_0 < 1/r$ tal que $g(x_0) = h_r(x_0)$, $g(x) < h_r(x)$ se $0 < x < x_0$ e $g(x) > h_r(x)$ se $x_0 < x < 1/r$, o que implica a tese. Uma justificativa formal para estas conclusões é dada a seguir.

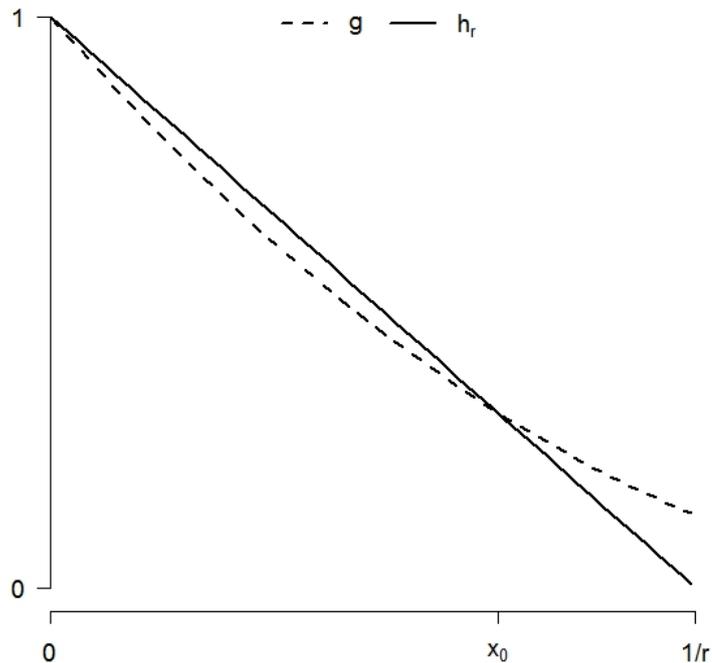


Figura 1. Gráficos das funções g e h_r no intervalo $[0, 1/r]$.

Definamos a diferença $d_r(x) = h_r(x) - g(x)$, $x \in \mathbb{R}$. Então, as funções d_r , d'_r e d''_r são contínuas na reta \mathbb{R} . Além disso, d_r é uma função côncava no intervalo $[0, 1/r]$, pois $d''_r(x) = h''_r(x) - g''_r(x) = -g''_r(x) < 0$ para todo $0 \leq x \leq 1/r$.

Notemos também que

$$d'_r(0) = h'_r(0) - g'(0) = -r + \sum_{i=1}^k n_i > 0,$$

o que implica, pela continuidade de d'_r , que existe $0 < \delta < 1/r$ tal que $d'_r(x) > 0$ para

todo $0 \leq x \leq \delta$. Como $d_r(0) = 0$, segue pelo Teorema do Valor Médio que

$$d_r(x) > 0, \quad \text{para todo } 0 < x \leq \delta. \quad (3.14)$$

Observemos também que $d_r(\delta) > 0$ e $d_r(1/r) = h_r(1/r) - g(1/r) = -g(1/r) < 0$. Logo, pelo Lema 3.3 aplicado à função d_r no intervalo $[\delta, 1/r]$ e por (3.14), segue que existe $0 < x_0 < 1/r$ tal que

$$\begin{cases} d_r(x) > 0, & \text{se } 0 < x < x_0, \\ d_r(x) = 0, & \text{se } x = x_0, \\ d_r(x) < 0, & \text{se } x_0 < x \leq 1/r, \end{cases}$$

o que implica

$$\begin{cases} g(x) < h_r(x), & \text{se } 0 < x < x_0, \\ g(x) = h_r(x), & \text{se } x = x_0, \\ g(x) > h_r(x), & \text{se } x_0 < x < 1/r, \end{cases}$$

e o resultado segue. ■

Portanto, segue do Lema 3.4 que, para $m < r < n$, podemos definir δ_r como sendo o menor número inteiro positivo t tal que

$$f_r\left(\frac{1}{r+t}\right) \leq 1, \quad (3.15)$$

ou seja,

$$\delta_r = \min \left\{ t \in \mathbb{N}^* : \prod_{i=1}^k (r+t-n_i) \leq t(r+t)^{k-1} \right\}.$$

A proposição a seguir caracteriza o comportamento da função $K^{\mathcal{C}}$ quanto ao crescimento (decréscimo) e fornece uma expressão para a estimativa de máxima verossimilhança, no caso em que $m < r < n$.

Proposição 3.8. *Para $m < r < n$, temos que*

- (a) se $\prod_{i=1}^k (r + \delta_r - n_i) < \delta_r (r + \delta_r)^{k-1}$, então $K^{\mathcal{C}}(N+1) < K^{\mathcal{C}}(N)$ para $N \geq r + \delta_r - 1$ e $K^{\mathcal{C}}(N+1) > K^{\mathcal{C}}(N)$ para $r \leq N < r + \delta_r - 1$. Portanto, a estimativa de máxima verossimilhança condicional é única e dada por $\hat{N}^{\mathcal{C}} = r + \delta_r - 1$;
- (b) se $\prod_{i=1}^k (r + \delta_r - n_i) = \delta_r (r + \delta_r)^{k-1}$, então $K^{\mathcal{C}}(N+1) < K^{\mathcal{C}}(N)$ para $N \geq r + \delta_r$, $K^{\mathcal{C}}(N+1) > K^{\mathcal{C}}(N)$ para $r \leq N < r + \delta_r - 1$ e $K^{\mathcal{C}}(r + \delta_r - 1) = K^{\mathcal{C}}(r + \delta_r)$. Neste

caso, há duas estimativas de máxima verossimilhança condicional $\widehat{N}_1^c = r + \delta_r - 1$ e $\widehat{N}_2^c = \widehat{N}_1^c + 1$.

Prova:

(a)

Se $\prod_{i=1}^k (r + \delta_r - n_i) < \delta_r (r + \delta_r)^{k-1}$ temos, pela definição de f_r , dada em (3.11) que $f_r(1/(r + \delta_r)) < 1$. Portanto, segue do Lema 3.4 e da definição de δ_r dada em (3.15) que

$$f_r\left(\frac{1}{r+t}\right) < 1, \text{ se } t \geq \delta_r,$$

$$f_r\left(\frac{1}{r+t}\right) > 1, \text{ se } 1 \leq t < \delta_r,$$

o que implica, pela relação (3.12) que

$$\frac{K^c(N+1)}{K^c(N)} < 1, \text{ se } N \geq r + \delta_r - 1,$$

$$\frac{K^c(N+1)}{K^c(N)} > 1, \text{ se } r \leq N < r + \delta_r - 1,$$

o que prova o resultado.

(b)

Se $\prod_{i=1}^k (r + \delta_r - n_i) = \delta_r (r + \delta_r)^{k-1}$ temos, pela definição de f_r dada em (3.11), que $f_r(1/(r + \delta_r)) = 1$. Logo, pela relação (3.12) segue

$$K^c(r + \delta_r - 1) = K^c(r + \delta_r).$$

e pelo Lema (3.4) e da definição de δ_r , dada em (3.15), segue

$$f_r\left(\frac{1}{r+t}\right) < 1, \text{ se } t \geq \delta_r + 1,$$

$$f_r\left(\frac{1}{r+t}\right) > 1, \text{ se } 1 \leq t < \delta_r,$$

o que implica, pela relação (3.12), que

$$\frac{K^c(N+1)}{K^c(N)} < 1, \text{ se } N \geq r + \delta_r,$$

$$\frac{K^c(N+1)}{K^c(N)} > 1, \text{ se } r \leq N < r + \delta_r - 1,$$

ou seja, $r + \delta_r - 1$ e $r + \delta_r$ maximizam $K^{\mathcal{C}}$. ■

3.1.3 Função de Verossimilhança Integrada Uniforme

Proposição 3.9. *O núcleo da função de verossimilhança integrada uniforme, $K^{\mathcal{U}}(N)$, converge para 0 quando $N \rightarrow \infty$.*

Prova:

A partir de (2.10) temos

$$\begin{aligned} K^{\mathcal{U}}(N) &= \frac{N!}{(N-r)!} \prod_{i=1}^k \frac{(N-n_i)!}{(N+1)!} \\ &= N(N-1)\cdots(N-r+1) \prod_{i=1}^k \frac{1}{(N+1)N(N-1)\cdots(N-n_i+1)} \\ &= \frac{N^r}{(N+1)^k} \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{r-1}{N}\right) \prod_{i=1}^k \frac{1}{N^{n_i} \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n_i-1}{N}\right)} \\ &= \frac{N^r}{N^n (N+1)^k} \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{r-1}{N}\right) \prod_{i=1}^k \frac{1}{\left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n_i-1}{N}\right)}. \end{aligned}$$

Logo, de (3.1) e do fato

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

segue o resultado. ■

No que segue consideremos a razão

$$\begin{aligned} \frac{K^{\mathcal{U}}(N+1)}{K^{\mathcal{U}}(N)} &= \frac{(N+1)}{(N+1-r)(N+2)^k} \prod_{i=1}^k (N+1-n_i) \\ &= \left(\frac{N+1}{N+1-r}\right) \left(\frac{N+1}{N+2}\right)^k \prod_{i=1}^k \left(\frac{N+1-n_i}{N+1}\right) \\ &= \left(1 - \frac{r}{N+1}\right)^{-1} \left(1 + \frac{1}{N+1}\right)^{-k} \prod_{i=1}^k \left(1 - \frac{n_i}{N+1}\right), \quad N \geq r. \quad (3.16) \end{aligned}$$

Proposição 3.10. *Se $r = m$, então $K^{\mathcal{U}}$ é estritamente decrescente, isto é, $K^{\mathcal{U}}(N+1) < K^{\mathcal{U}}(N)$ para todo $N \geq r$. Portanto, a estimativa de máxima verossimilhança integrada uniforme é $\hat{N}^{\mathcal{U}} = m$.*

Prova:

Sem perda de generalidade, suponhamos que $m = n_1$. Logo, para $r = n_1$, segue de (3.16) que

$$\frac{K^{\mathcal{U}}(N+1)}{K^{\mathcal{U}}(N)} = \left(1 + \frac{1}{N+1}\right)^{-k} \prod_{i=2}^k \left(1 - \frac{n_i}{N+1}\right) < 1, \quad N \geq n_1,$$

o que prova o resultado. ■

Para provar o próximo resultado, definamos a função

$$f_r(x) = (1 - rx)^{-1}(1 + x)^{-k} \prod_{i=1}^k (1 - n_i x), \quad 0 \leq x < 1/r. \quad (3.17)$$

Notamos, de acordo com (3.16), que f_r satisfaz a seguinte relação

$$f_r\left(\frac{1}{N+1}\right) = \frac{K^{\mathcal{U}}(N+1)}{K^{\mathcal{U}}(N)}, \quad N \geq r. \quad (3.18)$$

Desse modo, o crescimento (decrecimento) de $K^{\mathcal{U}}$ fica determinado quando conhecermos quais valores de x tornam $f_r(x)$ maior, menor ou igual a 1. O próximo lema caracteriza f_r neste sentido.

Lema 3.5. *Para $r > m$, existe x_0 , $0 < x_0 < 1/r$, tal que $f_r(x_0) = 1$, $f_r(x) < 1$ se $0 < x < x_0$ e $f_r(x) > 1$ se $x_0 < x < 1/r$. Por outro lado, x_0 não é da forma $1/(r+t)$ para nenhum número inteiro positivo t .*

Prova:

Definamos

$$g(x) = \prod_{i=1}^k (1 - n_i x), \quad x \in \mathbb{R}, \quad (3.19)$$

$$h_r(x) = (1 - rx)(1 + x)^k, \quad x \in \mathbb{R}. \quad (3.20)$$

Então,

$$f_r(x) = \frac{g(x)}{h_r(x)}, \quad 0 \leq x < 1/r,$$

e as derivadas de primeira e segunda ordem de g são dadas por

$$\begin{aligned} g'(x) &= \sum_{i=1}^k (-n_i) \prod_{\substack{j=1 \\ j \neq i}}^k (1 - n_j x), \\ g''(x) &= \sum_{i=1}^k (-n_i) \sum_{\substack{j=1 \\ j \neq i}}^k (-n_j) \prod_{\substack{s=1 \\ s \neq i, j}}^k (1 - n_s x) \\ &= \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k n_i n_j \prod_{\substack{s=1 \\ s \neq i, j}}^k (1 - n_s x), \quad \text{para } x \in \mathbb{R}, \end{aligned}$$

o que implica

$$g'(x) < 0, \quad \text{para todo } 0 \leq x \leq 1/r, \quad (3.21)$$

$$g''(x) > 0, \quad \text{para todo } 0 \leq x \leq 1/r. \quad (3.22)$$

Portanto, g , g' e g'' são funções contínuas e g é estritamente decrescente e convexa no intervalo $[0, 1/r]$ com $g(0) = 1$ e $g(1/r) > 0$.

Por outro lado, a primeira derivada de h_r é dada por

$$\begin{aligned} h'_r(x) &= -r(1+x)^k + k(1-rx)(1+x)^{k-1} \\ &= (1+x)^{k-1}[-r(1+x) + k(1-rx)] \\ &= (1+x)^{k-1}[k-r - (k+1)rx], \quad x \in \mathbb{R}, \end{aligned} \quad (3.23)$$

enquanto que a segunda derivada é dada por

$$\begin{aligned} h''_r(x) &= (k-1)(1+x)^{k-2}[k-r - (k+1)rx] - r(k+1)(1+x)^{k-1} \\ &= (1+x)^{k-2}\{(k-1)[k-r - (k+1)rx] - r(k+1)(1+x)\} \\ &= (1+x)^{k-2}\{(k-1)(k-r) - (k-1)(k+1)rx - r(k+1) - r(k+1)x\} \\ &= (1+x)^{k-2}\{(k-1)(k-r) - r(k+1) - k(k+1)rx\} \\ &= (1+x)^{k-2}\{k(k-1) - 2kr - k(k+1)rx\}, \quad x \in \mathbb{R}. \end{aligned} \quad (3.24)$$

Portanto, as funções h_r , h'_r e h''_r são funções contínuas com $h_r(0) = 1$ e $h_r(1/r) = 0$. Notamos que, provar o lema é equivalente a provar que existe x_0 , $0 < x_0 < 1/r$, tal que $g(x_0) = h_r(x_0)$, $g(x) < h_r(x)$ se $0 < x < x_0$ e $g(x) > h_r(x)$ se $x_0 < x < 1/r$. Neste sentido, definamos a diferença $d_r(x) = h_r(x) - g(x)$, $x \in \mathbb{R}$. Vamos mostrar que existe x_0 ,

$0 < x_0 < 1/r$, tal que

$$\begin{cases} d_r(x) > 0, & \text{se } 0 < x < x_0, \\ d_r(x) = 0, & \text{se } x = x_0, \\ d_r(x) < 0, & \text{se } x_0 < x \leq 1/r. \end{cases}$$

Pelas considerações anteriores, segue que d_r é uma função contínua, assim como suas derivadas de primeira e segunda ordem. No que segue, iremos dividir a prova em dois casos distintos: (a) $r \geq k$, (b) $r < k$.

Caso (a) $r \geq k$

Para $r \geq k$, segue das expressões (3.23) e (3.24) que $h_r'(x) < 0$ e $h_r''(x) < 0$, para todo $0 < x < 1/r$. Portanto, h_r é uma função estritamente decrescente e côncava no intervalo $[0, 1/r]$. A Figura 2 ilustra o comportamento das funções g e h_r no intervalo $[0, 1/r]$. A partir da observação deste gráfico, podemos concluir que existe x_0 , $0 < x_0 < 1/r$, tal que $g(x_0) = h_r(x_0)$, $g(x) < h_r(x)$ se $0 < x < x_0$ e $g(x) > h_r(x)$ se $x_0 < x < 1/r$, o que implica a tese. Uma justificativa formal para estas conclusões é dada a seguir.

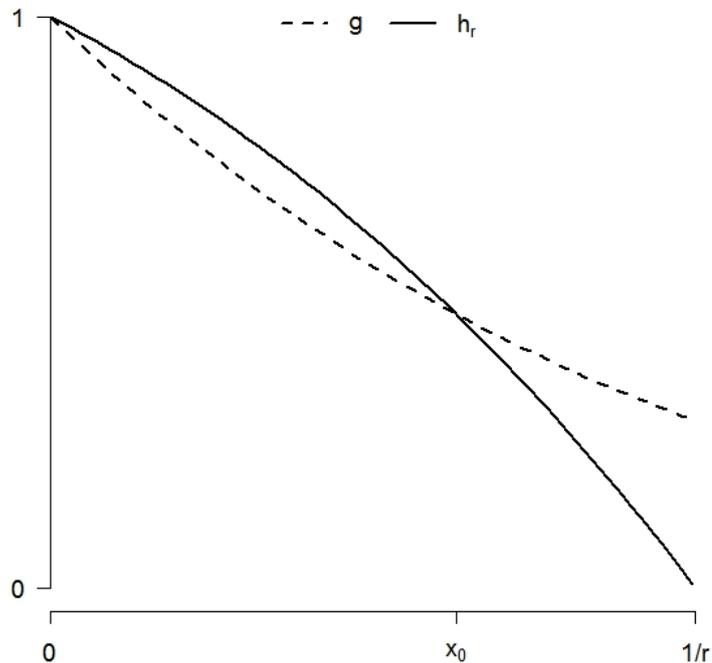


Figura 2. Gráficos das funções g e h_r no intervalo $[0, 1/r]$.

A partir de (3.21) e (3.23), temos

$$d'_r(0) = h'_r(0) - g'(0) = (k - r) + \sum_{i=1}^k n_i = \left(\sum_{i=1}^k n_i - r \right) + k > 0,$$

o que implica pela continuidade de d'_r que existe $0 < \delta < 1/r$ tal que $d'_r(x) > 0$ para todo $0 \leq x \leq \delta$. Como $d_r(0) = 0$, segue do Teorema do Valor Médio que

$$d_r(x) > 0, \quad \text{para todo } 0 < x \leq \delta, \quad (3.25)$$

e de (3.19) e (3.20) temos $d_r(1/r) = h_r(1/r) - g(1/r) < 0$.

A partir de (3.22) e (3.24), concluímos que d_r é uma função côncava no intervalo $[0, 1/r]$, pois $d''_r(x) = h''_r(x) - g''(x) < 0$ para todo $0 \leq x \leq 1/r$. Logo, como $d_r(\delta) > 0$ e $d_r(1/r) < 0$, segue do Lema 3.3 e de (3.25) que existe x_0 , $0 < x_0 < 1/r$, tal que

$$\begin{cases} d_r(x) > 0, & \text{se } 0 < x < x_0, \\ d_r(x) = 0, & \text{se } x = x_0, \\ d_r(x) < 0, & \text{se } x_0 < x \leq 1/r. \end{cases}$$

Caso (b) $r < k$

Neste caso, de (3.23) temos

$$h'_r(x) > 0, \quad \text{se } 0 < x < \frac{k-r}{r(k+1)},$$

isto é, h_r é estritamente crescente no intervalo $[0, \frac{k-r}{r(k+1)}]$. Como g é estritamente decrescente e $g(0) = h_r(0) = 1$, então

$$g(x) < 1 < h_r(x), \quad \text{se } 0 < x \leq \frac{k-r}{r(k+1)},$$

o que implica

$$d_r(x) = h_r(x) - g(x) > 0, \quad \text{se } 0 < x \leq \frac{k-r}{r(k+1)}. \quad (3.26)$$

De (3.24), temos

$$h''_r(x) < 0, \quad \text{se } \frac{k-r}{r(k+1)} \leq x \leq \frac{1}{r},$$

o que implica

$$d''_r(x) < 0, \quad \text{se } \frac{k-r}{r(k+1)} \leq x \leq \frac{1}{r},$$

ou seja, d_r é côncava no intervalo $[\frac{k-r}{r(k+1)}, \frac{1}{r}]$.

Portanto, como $d_r(\frac{k-r}{r(k+1)}) > 0$ por (3.26), $d_r(1/r) = h_r(1/r) - g(1/r) = -g(1/r) < 0$

e d_r é côncava em $[\frac{k-r}{r(k+1)}, \frac{1}{r}]$, segue do Lema 3.3 e de (3.26) que existe x_0 , $0 < x_0 < 1/r$ tal que

$$\begin{cases} d_r(x) > 0, & \text{se } 0 < x < x_0, \\ d_r(x) = 0, & \text{se } x = x_0, \\ d_r(x) < 0, & \text{se } x_0 < x \leq 1/r. \end{cases}$$

A Figura 3 ilustra o comportamento das funções g e h_r no intervalo $[0, 1/r]$.

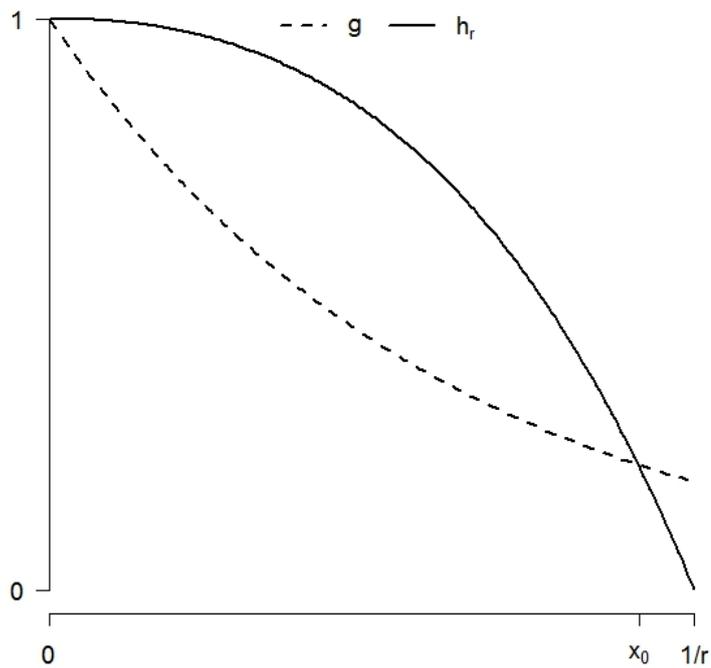


Figura 3. Gráficos das funções g e h_r no intervalo $[0, 1/r]$.

Finalmente, x_0 não é da forma $1/(r+t)$ para nenhum número inteiro positivo t . Com

feito, se $f_r(1/(r+t)) = 1$ para algum número inteiro positivo t , então

$$\begin{aligned}
& \left(1 - \frac{r}{r+t}\right)^{-1} \left(1 + \frac{1}{r+t}\right)^{-k} \prod_{i=1}^k \left(1 - \frac{n_i}{r+t}\right) = 1 \\
\implies & \prod_{i=1}^k \frac{r+t-n_i}{r+t} = \frac{t(r+t+1)^k}{(r+t)^{k+1}} \\
\implies & \prod_{i=1}^k (r+t-n_i) = \frac{(r+t-r)(r+t+1)^k}{r+t} \\
\implies & \prod_{i=1}^k (r+t-n_i) - (r+t+1)^k - \frac{r}{r+t}(r+t+1)^k = 0 \\
\implies & \prod_{i=1}^k (r+t-n_i) - (r+t+1)^k - \frac{r}{r+t} - \frac{r}{r+t} \sum_{i=1}^k \binom{k}{i} (r+t)^i = 0 \\
\implies & \prod_{i=1}^k (r+t-n_i) - (r+t+1)^k - r \sum_{i=1}^k \binom{k}{i} (r+t)^{i-1} = \frac{r}{r+t},
\end{aligned}$$

o que é absurdo, pois o lado esquerdo da última igualdade é um número inteiro, enquanto o lado direito não pode ser inteiro uma vez que $0 < \frac{r}{r+t} < 1$. ■

Portanto, segue do Lema 3.5 que, para $r > m$, podemos definir δ_r como o menor inteiro positivo t tal que

$$f_r\left(\frac{1}{r+t}\right) < 1, \quad (3.27)$$

ou seja,

$$\delta_r = \min \left\{ t \in \mathbb{N}^* : \prod_{i=1}^k (r+t-n_i) < \frac{t(r+t+1)^k}{r+t} \right\}.$$

A proposição a seguir caracteriza o comportamento de $K^{\mathcal{U}}$ quanto ao crescimento (decréscimo) e fornece a estimativa de máxima verossimilhança uniforme de N no caso em que $r > m$.

Proposição 3.11. *Para $r > m$, $K^{\mathcal{U}}(N+1) < K^{\mathcal{U}}(N)$ se $N \geq r + \delta_r - 1$ e $K^{\mathcal{U}}(N+1) > K^{\mathcal{U}}(N)$ se $r \leq N < r + \delta_r - 1$. Portanto, a estimativa de máxima verossimilhança integrada uniforme é $\widehat{N}^{\mathcal{U}} = r + \delta_r - 1$.*

Prova:

Do Lema 3.5 e da definição de f_r dada em (3.27), segue

$$\begin{aligned} f_r\left(\frac{1}{r+t}\right) &< 1, & \text{se } t \geq \delta_r, \\ f_r\left(\frac{1}{r+t}\right) &> 1, & \text{se } 1 \leq t < \delta_r, \end{aligned}$$

o que, pela relação (3.18), implica

$$\begin{aligned} \frac{K^{\mathcal{U}}(N+1)}{K^{\mathcal{U}}(N)} &< 1, & \text{se } N \geq r + \delta_r - 1, \\ \frac{K^{\mathcal{U}}(N+1)}{K^{\mathcal{U}}(N)} &> 1, & \text{se } r \leq N < r + \delta_r - 1, \end{aligned}$$

isto é, $r + \delta_r - 1$ maximiza $K^{\mathcal{U}}$. ■

3.1.4 Função de Verossimilhança Integrada de Jeffreys

Proposição 3.12. *O núcleo da função de verossimilhança integrada de Jeffreys, $K^{\mathcal{J}}(N)$, converge para 0 quando $N \rightarrow \infty$.*

Prova:

De (2.12) segue

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

Como $1 - x \leq \exp(-x)$ para todo x real, então para cada $j = 1, \dots, k$, temos

$$\begin{aligned}
0 &\leq \prod_{l=1}^{N-n_j} \left(1 - \frac{1}{2(N-n_j+1-l)} \right) \\
&\leq \prod_{l=1}^{N-n_j} \exp \left(\frac{-1}{2(N-n_j+1-l)} \right) \\
&= \exp \left(\sum_{l=1}^{N-n_j} \frac{-1}{2(N-n_j+1-l)} \right) \\
&= \exp \left(-\frac{1}{2} \sum_{s=1}^{N-n_j} \frac{1}{s} \right), \tag{3.28}
\end{aligned}$$

e como $\sum_{s=1}^{\infty} 1/s = \infty$ segue de (3.28),

$$\lim_{N \rightarrow \infty} \prod_{l=1}^{N-n_j} \left(1 - \frac{1}{2(N-n_j+1-l)} \right) = 0. \tag{3.29}$$

Portanto, de (3.1), (3.2) e (3.29) segue o resultado. ■

No que segue, vamos considerar a razão

$$\begin{aligned}
\frac{K^{\mathcal{J}}(N+1)}{K^{\mathcal{J}}(N)} &= \frac{N+1}{N+1-r} \prod_{i=1}^k \frac{\Gamma(N+1-n_i+1/2)}{\Gamma(N-n_i+1/2)(N+1)} \\
&= \left[1 - \frac{r}{N+1} \right]^{-1} \prod_{i=1}^k \left(1 - \frac{n_i+1/2}{N+1} \right), \quad N \geq r. \tag{3.30}
\end{aligned}$$

Proposição 3.13. *Para $r = m$, temos que $K^{\mathcal{J}}(N)$ é estritamente decrescente, isto é, $K^{\mathcal{J}}(N+1) < K^{\mathcal{J}}(N)$ para todo $N \geq r$. Portanto, a estimativa de máxima verossimilhança integrada de Jeffreys é $\widehat{N}^{\mathcal{J}} = m$.*

Prova:

Vamos supor, sem perda de generalidade, que $m = n_1$. Então, para $r = n_1$, segue de (3.30)

$$\frac{K^{\mathcal{J}}(N+1)}{K^{\mathcal{J}}(N)} = \frac{N-n_1+1/2}{N-n_1+1} \prod_{i=2}^k \frac{N-n_i+1/2}{N+1} < 1, \quad \text{para todo } N \geq r,$$

pois $N-n_1+1/2 < N-n_1+1$ e $N-n_i+1/2 < N+1$ para $i = 2, \dots, k$. ■

Definamos a função

$$f_r(x) = (1 - rx)^{-1} \prod_{i=1}^k (1 - (n_i + 1/2)x), \quad 0 \leq x < 1/r. \quad (3.31)$$

Então, de (3.30)

$$f_r\left(\frac{1}{N+1}\right) = \frac{K^{\mathcal{J}}(N+1)}{K^{\mathcal{J}}(N)}, \quad N \geq r. \quad (3.32)$$

O comportamento da função f_r é descrito no lema a seguir.

Lema 3.6. *Para $r > m$, existe x_0 , $0 < x_0 < 1/r$, tal que $f_r(x_0) = 1$, $f_r(x) < 1$ para $0 < x < x_0$ e $f_r(x) > 1$ para $x_0 < x < 1/r$. Além disso, x_0 não é da forma $1/(r+t)$ para nenhum número inteiro positivo t .*

Prova:

Com respeito às funções

$$g(x) = \prod_{i=1}^k (1 - (n_i + 1/2)x), \quad \text{para } x \in \mathbb{R},$$

$$h_r(x) = 1 - rx, \quad \text{para } x \in \mathbb{R},$$

temos

$$f_r(x) = \frac{g(x)}{h_r(x)}, \quad 0 \leq x < 1/r.$$

As derivadas de primeira e segunda ordem de g são dadas por

$$g'(x) = \sum_{i=1}^k -(n_i + 1/2) \prod_{\substack{j=1 \\ j \neq i}}^k (1 - (n_j + 1/2)x),$$

$$g''(x) = \sum_{i=1}^k -(n_i + 1/2) \sum_{\substack{j=1 \\ j \neq i}}^k -(n_j + 1/2) \prod_{\substack{s=1 \\ s \neq i, j}}^k (1 - (n_s + 1/2)x)$$

$$= \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k (n_i + 1/2)(n_j + 1/2) \prod_{\substack{s=1 \\ s \neq i, j}}^k (1 - (n_s + 1/2)x),$$

para $x \in \mathbb{R}$ e como $1 - (n_i + 1/2)x > 0$ para $0 \leq x \leq 1/r$, $i = 1, \dots, n$, temos

$$g'(x) < 0, \quad \text{para todo } 0 \leq x \leq 1/r,$$

$$g''(x) > 0, \quad \text{para todo } 0 \leq x \leq 1/r.$$

Portanto, g é função contínua, estritamente decrescente e convexa no intervalo $[0, 1/r]$, com $g(0) = 1$ e $g(1/r) > 0$. Por outro lado, a função h_r é linear e estritamente decrescente no intervalo $[0, 1/r]$, com $h_r(0) = 1$ e $h_r(1/r) = 0$. A Figura 4 ilustra o comportamento das funções f_r e g no intervalo $[0, 1/r]$ e, a partir da observação deste gráfico, podemos deduzir que existe um ponto x_0 , $0 < x_0 < 1/r$ tal que $g(x_0) = h_r(x_0)$, $g(x) < h_r(x)$ se $0 < x < x_0$ e $g(x) > h_r(x)$ se $x_0 < x < 1/r$. Uma justificativa formal para isto é dada a seguir.

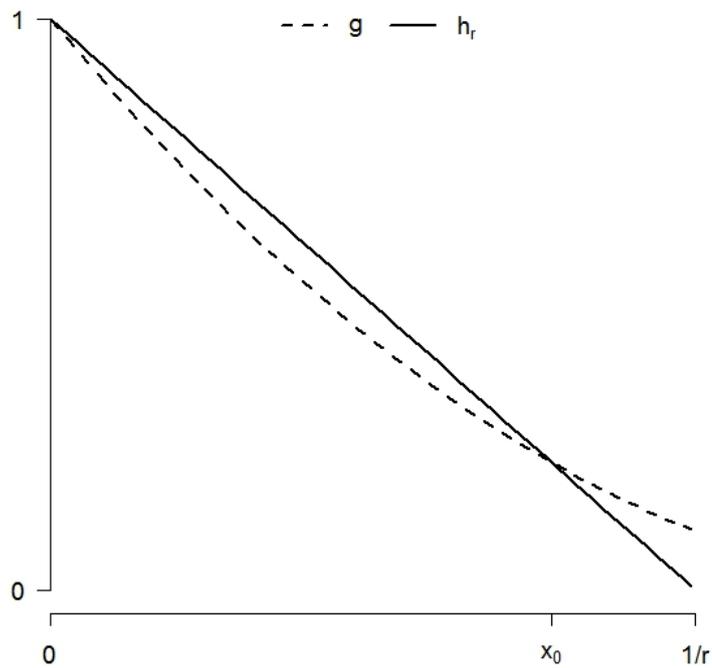


Figura 4. Gráficos das funções g e h_r no intervalo $[0, 1/r]$.

Definamos a diferença $d_r(x) = h_r(x) - g(x)$, $x \in \mathbb{R}$. Notemos que a função d_r e suas derivadas de primeira e segunda ordem são contínuas em toda a reta \mathbb{R} e

$$d_r'(0) = h_r'(0) - g'(0) = -r + \sum_{i=1}^k n_i + \frac{k}{2} > 0,$$

o que implica, pela continuidade de d_r' , que existe $0 < \delta < 1/r$ tal que $d_r'(x) > 0$ para todo $0 \leq x \leq \delta$. Portanto, pelo Teorema do Valor Médio, segue

$$d_r(x) > 0, \quad \text{para todo } 0 < x \leq \delta. \quad (3.33)$$

Como $d_r''(x) = h_r''(x) - g_r''(x) = -g_r''(x) < 0$ para todo $0 \leq x \leq 1/r$, concluímos que a função d_r é côncava no intervalo $[0, 1/r]$. Logo, segue de $d_r(\delta) > 0$, $d_r(1/r) = h_r(1/r) - g(1/r) = -g(1/r) < 0$, do Lema 3.3 e de(3.33) que

$$\begin{cases} d_r(x) > 0, & \text{se } 0 < x < x_0, \\ d_r(x) = 0, & \text{se } x = x_0, \\ d_r(x) < 0, & \text{se } x_0 < x < 1/r, \end{cases}$$

ou seja,

$$\begin{cases} g(x) < h_r(x), & \text{se } 0 < x < x_0, \\ g(x) = h_r(x), & \text{se } x = x_0, \\ g(x) > h_r(x), & \text{se } x_0 < x < 1/r. \end{cases}$$

Por outro lado, x_0 não é da forma $1/(r+t)$ para nenhum número inteiro positivo t . De fato, se $f_r(1/(r+t)) = 1$ para algum número inteiro positivo t , então

$$\begin{aligned} & \left(1 - \frac{r}{r+t}\right)^{-1} \prod_{i=1}^k \left(1 - \frac{n_i + 1/2}{r+t}\right) = 1 \\ \implies & \prod_{i=1}^k \left(r+t - n_i - \frac{1}{2}\right) = t(r+t)^{k-1} \\ \implies & \prod_{i=1}^k \left(2(r+t - n_i) - 1\right) = 2^k t(r+t)^{k-1}, \end{aligned}$$

o que é absurdo, pois o lado esquerdo da igualdade acima é número ímpar, enquanto o lado direito é número par. ■

Portanto, a partir do Lema 3.6, podemos definir para $r > m$, δ_r como o menor número inteiro positivo t tal que

$$f_r\left(\frac{1}{r+t}\right) < 1, \quad (3.34)$$

ou seja,

$$\delta_r = \min \left\{ t \in \mathbb{N}^* : \prod_{i=1}^k \left(r+t - n_i - \frac{1}{2}\right) < t(r+t)^{k-1} \right\}.$$

A proposição a seguir caracteriza o comportamento da função $K^{\mathcal{J}}$ quanto ao crescimento (decréscimo) e fornece a estimativa de máxima verossimilhança integrada de Jeffreys de N no caso em que $r > m$.

Proposição 3.14. *Para $r > m$, $K^{\mathcal{J}}(N+1) < K^{\mathcal{J}}(N)$ se $N \geq r + \delta_r - 1$ e $K^{\mathcal{J}}(N+1) > K^{\mathcal{J}}(N)$ se $r \leq N < r + \delta_r - 1$. Portanto, a estimativa de máxima verossimilhança*

integrada de Jeffreys é única e dada por $\widehat{N}^{\mathcal{J}} = r + \delta_r - 1$

Prova:

Se $\prod_{i=1}^k (r + \delta_r - n_i - \frac{1}{2}) < \delta_r (r + \delta_r)^{k-1}$ conclui-se, pela definição de f_r dada em (3.31), que $f_r(1/(r + \delta_r)) < 1$. Portanto, segue do Lema 3.6 e da definição de δ_r dada em (3.34)

$$f_r\left(\frac{1}{r+t}\right) < 1, \quad \text{se } t \geq \delta_r, \quad (3.35)$$

$$f_r\left(\frac{1}{r+t}\right) > 1, \quad \text{se } 1 \leq t < \delta_r, \quad (3.36)$$

o que implica pela relação (3.32) que

$$\begin{aligned} K^{\mathcal{J}}(N+1) &< K^{\mathcal{J}}(N), & \text{se } N \geq r + \delta_r - 1, \\ K^{\mathcal{J}}(N+1) &> K^{\mathcal{J}}(N), & \text{se } r \leq N < r + \delta_r - 1, \end{aligned}$$

ou seja, $r + \delta_r - 1$ maximiza $K^{\mathcal{J}}$. ■

3.1.5 Exemplo Ilustrativo

Para ilustrar os resultados a respeito das funções de verossimilhança estudadas, vamos considerar um exemplo com 5 épocas de amostragem em que o número de indivíduos capturados em cada época de amostragem é dado na Tabela 1 abaixo. Observemos que para esta amostra $m = 3$ e $n = 10$, portanto o número de indivíduos distintos capturados r pode assumir valores entre 3 e 10.

Tabela 1. Número de indivíduos capturados em cada época de amostragem.

n_1	n_2	n_3	n_4	n_5
2	2	1	3	2

A Figura 5 ilustra os gráficos do logaritmo do kernel de cada uma das verossimilhanças para os valores fixados de (a) $r = 3$, (b) $r = 8$ e (c) $r = 10$. Observemos que para as verossimilhanças perfilada e condicional, existem três formas possíveis para o kernel da verossimilhança: (a) estritamente decrescente; (b) crescente e depois decrescente; (c) estritamente crescente. Nos casos das verossimilhanças integradas uniforme e de Jeffreys, existem apenas duas formas possíveis: (a) estritamente decrescente; (b) e (c) crescente e depois decrescente. Observemos que nos casos das verossimilhanças integradas, mesmo

quando o número de indivíduos distintos capturados r é igual a soma do número de indivíduos capturados n (que ocorre quando nenhum indivíduo marcado é recapturado), o kernel da função de verossimilhança não é estritamente crescente, como ocorre nos casos das verossimilhanças perfiladas e condicional.

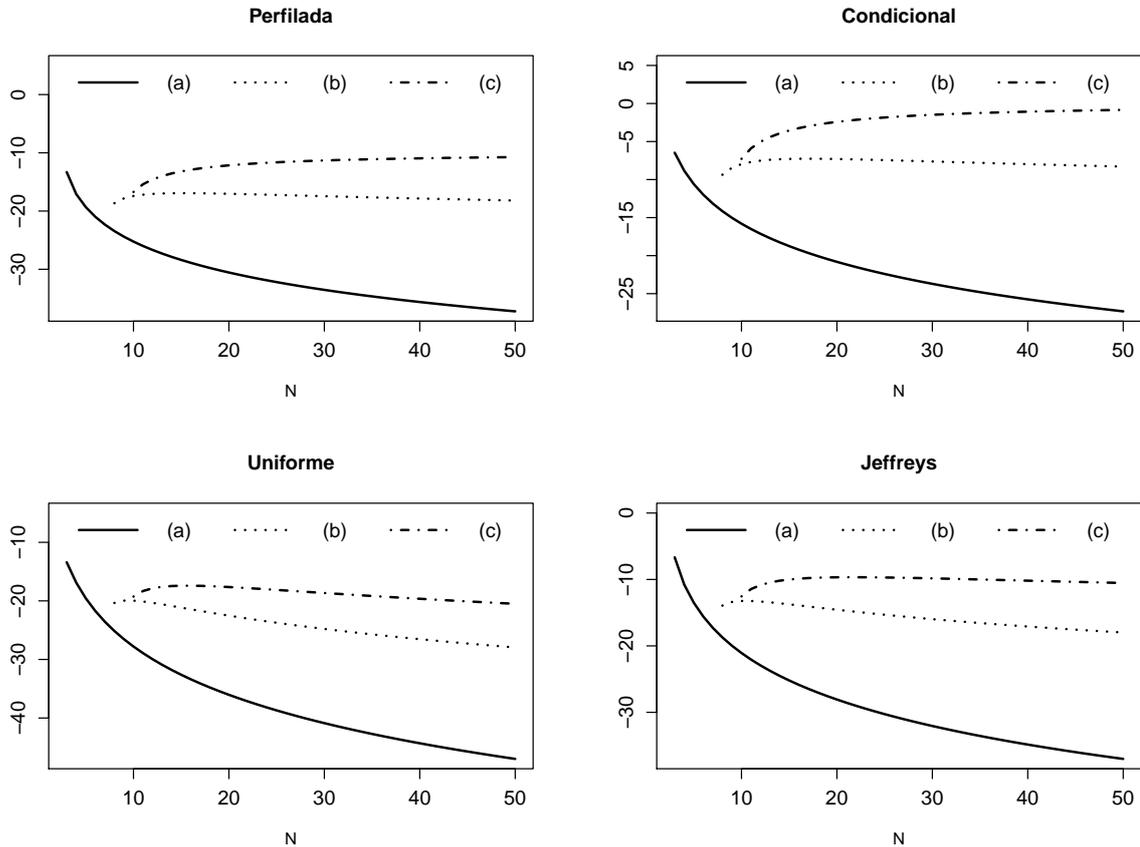


Figura 5. Gráficos do logaritmo do núcleo das funções de verossimilhança perfilada, condicional, integrada uniforme e integrada de Jeffreys para a amostra $n_1 = 2$, $n_2 = 2$, $n_3 = 1$, $n_4 = 3$, $n_5 = 2$, nos casos em que (a) $r = 3$, (b) $r = 8$, (c) $r = 10$.

A Tabela 2 mostra as estimativas de máxima verossimilhança para cada uma das verossimilhanças estudadas para todos os possíveis valores de r , $3 \leq r \leq 10$. Observamos por esta tabela que o aumento no valor de r causa grande impacto sobre as estimativas de máxima verossimilhança perfilada e condicional. No entanto, este impacto é menor nos casos das verossimilhanças integradas uniforme e de Jeffreys.

Tabela 2. Estimativas de máxima verossimilhança perfilada, condicional, integrada uniforme e integrada de Jeffreys.

\mathbf{r}	$\widehat{\mathbf{N}}^{\mathcal{P}}$	$\widehat{\mathbf{N}}^{\mathcal{C}}$	$\widehat{\mathbf{N}}^{\mathcal{U}}$	$\widehat{\mathbf{N}}^{\mathcal{J}}$
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	7	7	6	6
7	10	10	8	8
8	15	17	10	10
9	32	37	12	14
10	∞	∞	16	22

3.2 Estimação Intervalar

No que segue designamos por \mathcal{M} o método de eliminação de parâmetros perturbadores considerado, isto é, \mathcal{M} pode ser igual a \mathcal{P} (Perfilada), \mathcal{C} (Condicional), \mathcal{U} (Integrada Uniforme) ou \mathcal{J} (Integrada de Jeffreys). Baseado em uma função de verossimilhança, $L^{\mathcal{M}}(N)$, livre de parâmetros perturbadores construímos um conjunto de confiança aproximado $100(1 - \alpha)\%$ para N como

$$C^{\mathcal{M}}(1 - \alpha) = \left\{ N : \log \left(K^{\mathcal{M}}(N) \right) \geq \log \left(K^{\mathcal{M}}(\widehat{N}^{\mathcal{M}}) \right) - \frac{1}{2} \chi_1^2(1 - \alpha) \right\}, \quad (3.37)$$

com $\chi_1^2(1 - \alpha)$ representando o quantil $(1 - \alpha)$ da distribuição qui-quadrado com 1 grau de liberdade, $K^{\mathcal{M}}$ o núcleo da função de verossimilhança $L^{\mathcal{M}}$, $\widehat{N}^{\mathcal{M}}$ um valor de N que maximiza $K^{\mathcal{M}}(N)$. Este conjunto de confiança é construído invertendo-se o teste assintótico da razão de verossimilhanças generalizado para uma hipótese nula simples. O uso das funções de verossimilhança livres de parâmetros perturbadores para construção de conjuntos de confiança deste tipo é sugerido por Berger, Liseo e Wolpert (1999) e tem um apelo intuitivo, pois contém os valores de N mais plausíveis de acordo com a respectiva função de verossimilhança.

Para as funções de verossimilhanças perfilada e condicional a estimativa de máxima verossimilhança é igual a $+\infty$, para amostras com $r = n$. Neste caso, adotamos no lugar de $K^{\mathcal{M}}(\widehat{N}^{\mathcal{M}})$ em (3.37) o valor do limite de $K^{\mathcal{M}}(N)$ quando $N \rightarrow +\infty$, dado pela Proposição 3.1 no caso da função de verossimilhança perfilada e pela Proposição 3.5 no caso da função de verossimilhança condicional. Para as funções de verossimilhança

integradas uniforme e de Jeffreys a estimativa de máxima verossimilhança é sempre finita.

Segue dos resultados da seção 3.1 com respeito ao comportamento das funções de verossimilhança, que o conjunto de confiança definido em (3.37) é um intervalo da forma $[N_1^{\mathcal{M}}, N_2^{\mathcal{M}}]$ com $r \leq N_1^{\mathcal{M}} < N_2^{\mathcal{M}} \leq \infty$. Vale observar que o extremo superior $N_2^{\mathcal{M}}$ é igual a ∞ se e somente se $r = n$ e as funções de verossimilhança consideradas forem a perfilada e a condicional. Com efeito, as função de verossimilhança perfilada e condicional são estritamente crescentes se $r = n$ (Proposições 3.3 e 3.7) e, portanto, o intervalo de confiança definido em 3.37 é da forma $[N_1^{\mathcal{M}}, \infty]$ em ambos os casos. Por outro lado, se $r < n$, segue das Proposições 3.1, 3.5, 3.9, 3.12 que o limite de $K^{\mathcal{M}}(N)$ é 0, ou seja, o limite de $\log(K^{\mathcal{M}}(N))$ é igual a $-\infty$ quando $N \rightarrow \infty$. Consequentemente, se $r < n$ a condição sobre N dada em (3.37) será satisfeita apenas para um número finito de valores de N e, portanto, o intervalo de confiança definido em 3.37 é da forma $[N_1^{\mathcal{M}}, N_2^{\mathcal{M}}]$ com $N_2^{\mathcal{M}} < \infty$.

A Figura 6 ilustra, a partir de um conjunto de dados artificiais, como é obtido o intervalo de confiança 95% para N definido em 3.37 baseado na função de verossimilhança condicional sob três situações distintas: (a) $r = m$; (b) $m < r < n$ e (c) $r = n$. No caso (a), a função de verossimilhança é estritamente decrescente e o intervalo de confiança é da forma $[N_1, N_2]$ com $N_1 = m$. No caso (b), a função de verossimilhança é crescente para $N < \hat{N}$ e decrescente para $N > \hat{N}$, logo o intervalo de confiança é da forma $[N_1, N_2]$. No caso (c), a função de verossimilhança é estritamente crescente e o intervalo de confiança é da forma $[N_1, \infty]$.

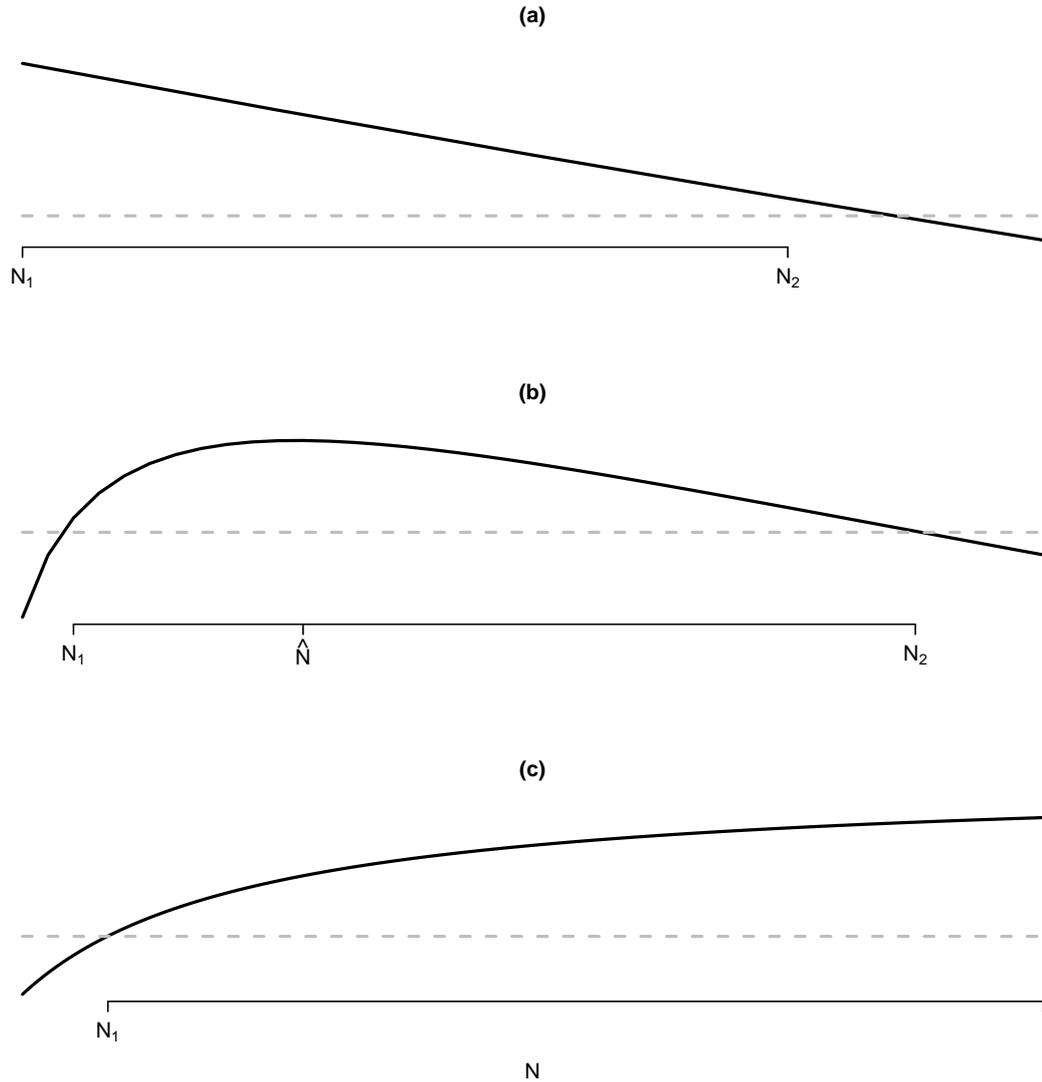


Figura 6. Estimativas intervalares de N baseadas na função de verossimilhança condicional nos casos em que (a) $r = m$, (b) $m < r < n$ e (c) $r = n$. A linha contínua preta representa o valor de $\log(K^C(N))$, enquanto a linha tracejada cinza representa o valor da constante $\log(K^C(\hat{N}^C)) - \chi_1^2(0, 95)/2$.

Os valores de N_1 e N_2 são determinados considerando a forma que assume a função de verossimilhança e determinando, quando for necessário, o valor real x que soluciona, dentro de um intervalo apropriado, a equação

$$\log(K^M(x)) = \log(K^M(\hat{N}^M)) - \frac{1}{2}\chi_1^2(1 - \alpha) = 0,$$

considerando $K^M(x)$ uma extensão a valores reais da função a valores inteiros $K^M(N)$.

Para efeito comparativo, consideraremos também o intervalo de confiança de Wald

para N , largamente usado em aplicações, cujos limites inferior e superior são dados por

$$\begin{aligned}\widehat{N}_L^{\mathcal{W}} &= \widehat{N} - z_{(1-\alpha/2)} \sqrt{\widehat{\text{Var}}(\widehat{N})}, \\ \widehat{N}_U^{\mathcal{W}} &= \widehat{N} + z_{(1-\alpha/2)} \sqrt{\widehat{\text{Var}}(\widehat{N})},\end{aligned}$$

onde $\widehat{\text{Var}}(\widehat{N})$ é uma estimativa da variância \widehat{N} e z_γ é o γ -ésimo quantil da distribuição normal padrão. Consideraremos a estimativa assintótica (DARROCH, 1958)

$$\widehat{\text{Var}}(\widehat{N}) = \widehat{N} \left[\prod_{j=1}^k (1 - \widehat{p}_j)^{-1} + k - 1 - \sum_{j=1}^k (1 - \widehat{p}_j)^{-1} \right]^{-1},$$

onde $\widehat{\mathbf{p}} = (\widehat{p}_1, \widehat{p}_2, \dots, \widehat{p}_k)$ é a estimativa de máxima verossimilhança \mathbf{p} dado por (2.6) substituindo N por \widehat{N} .

3.3 Exemplo Numérico

Nesta seção, ilustramos os métodos de estimação apresentados através de um conjunto de dados reais provenientes de um experimento de captura-recaptura de peixes “Redear Sunfish” realizado no Lago Gordy, Indiana (USA) (GERKING, 1953). A Tabela 3 apresenta o número de indivíduos capturados e recapturados em cada uma das 14 épocas de amostragem. Para este experimento foram observados 138 peixes distintos ($r = 138$). O conjunto de dados apresentado é típico de muitas aplicações ecológicas nas quais há várias épocas de amostragem e o número de recapturas em cada época é pequeno. Estes dados foram usados por Ricker (1975) para ilustrar métodos frequentistas e por Castleline (1981), Smith (1988), Smith (1991), George e Robert (1992) para ilustrar métodos bayesianos, segundo o modelo estatístico descrito na seção 2.1.

Tabela 3. Contagens de peixes no Lago Gordy, Indiana (USA).

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_i	10	27	17	7	1	5	6	15	9	18	16	5	7	19
m_i	0	0	0	0	0	0	2	1	5	5	4	2	2	3

Para estes dados, a Figura 7 ilustra o gráfico do logaritmo do quociente entre o valor da função de verossimilhança e seu valor máximo, para os casos da perfilada, condicional, integrada uniforme e integrada de Jeffreys. A partir destes gráficos, podemos observar um mesmo comportamento para todas as funções de verossimilhança: crescente de r até a estimativa de máxima verossimilhança e decrescente a partir deste valor. Podemos

concluir também que a taxa de decrescimento das funções de verossimilhança perfilada e condicional é menor do que aquelas observadas nas funções de verossimilhança integrada uniforme e de Jeffreys. Esta característica fará com que os intervalos de confiança obtidos a partir das funções de verossimilhança perfilada e condicional tenham amplitude maior do que aqueles obtidos a partir das funções de verossimilhança integrada uniforme e de Jeffreys.

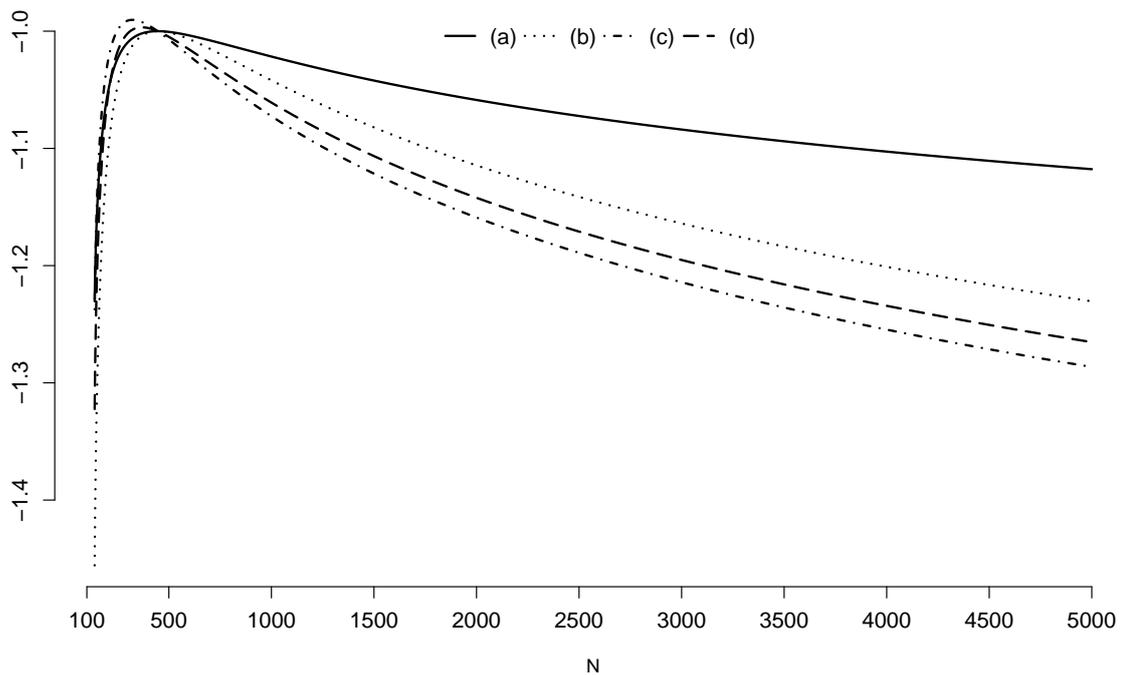


Figura 7. Gráficos do logaritmo do núcleo das funções de verossimilhança (a) perfilada, (b) condicional, (c) integrada uniforme e (d) integrada de jeffreys, para os dados de captura-recaptura de peixes no Lago Gordy.

Para este conjunto de dados, as estimativas de máxima verossimilhança, \hat{N} , e os intervalos de confiança (IC) 95% para cada uma das funções de verossimilhança consideradas encontram-se na Tabela 4. Usando uma abordagem bayesiana Smith (1991) considera uma distribuição *a priori* de Jeffreys para N ($\pi(N) \propto 1/N$) e distribuições *a priori* independentes Beta(a, b) para as probabilidades de captura \mathbf{p} e obtém estimativas *a posteriori* de N . Nos casos das *prioris* Uniforme ($a = 1, b = 1$) e de Jeffreys ($a = 0,5, b = 0,5$) para \mathbf{p} , o autor obtém os intervalos [255, 434] e [281, 520] com credibilidade de 95% para N , respectivamente, semelhantes aos intervalos de confiança 95% das funções de verossimilhança integradas uniforme e de Jeffreys mostrados na Tabela 4, respectivamente.

Tabela 4. Estimativas pontuais e intervalares para os dados do Lago Gordy.

Funções de Verossimilhança	\hat{N}	IC 95 %
Perfilada	445	[325, 654]
Condicional	448	[326, 658]
Integrada Uniforme	322	[254, 428]
Integrada Jeffreys	369	[282, 510]

4 *Estudo de Simulação*

Com o objetivo de comparar o desempenho dos estimadores pontuais e intervalares, realizamos um estudo de simulação com a geração de 1.000 réplicas independentes dos dados $\mathbf{n} = (n_1, n_2, \dots, n_k)$ e $\mathbf{m} = (m_1, m_2, \dots, m_k)$ a partir do modelo dado em (2.1) com os valores dos parâmetros fixados em $N = 50, 100, 200, 400, 800$, $k = 3, 4, \dots, 10$ e as probabilidades de captura geradas de cada uma das populações da Tabela 5 abaixo.

Tabela 5. Populações para as probabilidades de captura.

População	Distribuição
P_1	$p_1, \dots, p_k \sim \text{Uniforme}(0, 05; 0, 10)$
P_2	$p_1, \dots, p_k \sim \text{Uniforme}(0, 10; 0, 15)$
P_3	$p_1, \dots, p_k \sim \text{Uniforme}(0, 15; 0, 20)$
P_4	$p_1, \dots, p_k \sim \text{Uniforme}(0, 20; 0, 25)$

4.1 *Estimação Pontual*

Inicialmente, para se ter ideia do comportamento dos estimadores pontuais considerados, ilustramos graficamente o efeito do aumento das probabilidades de captura sobre o comportamento das estimativas de máxima verossimilhança de cada um dos métodos considerados, construindo um diagrama de caixa (box-plot) das estimativas obtidas para diferentes combinações do tamanho populacional N e do número de épocas de captura k . É importante observar que os diagramas construídos levam em conta apenas as estimativas finitas, fato importante no caso das funções de verossimilhança perfilada e condicional, pois as proporções de estimativas iguais a infinito não são desprezíveis em muitos casos. A Tabela 6 a seguir apresenta as porcentagens de estimativas infinitas para diferentes combinações de valores de N , k e \mathbf{p} .

Tabela 6. Porcentagens de estimativas de máxima verossimilhança perfilada e condicional iguais a $+\infty$.

N	p	k						
		3	4	5	6	7	8	9
50	P_1	48,4	30,1	11,7	4,4	1,1	0,5	0,1
	P_2	13,1	2,2	0	0	0	0	0
	P_3	1,1	0	0	0	0	0	0
	P_4	0,1	0	0	0	0	0	0
100	P_1	25,6	8,3	1,2	0,2	0	0	0
	P_2	1,8	0	0	0	0	0	0
	P_3	0,1	0	0	0	0	0	0
200	P_1	6,3	1,0	0,1	0	0	0	0
400	P_1	0,3	0	0	0	0	0	0

A partir da análise gráfica dos box-plots das estimativas obtidas para todas as combinações do tamanho populacional N e do número de épocas de captura k considerados neste estudo de simulação, observa-se de maneira geral que: (i) o aumento das probabilidades e do número de épocas de captura causam uma diminuição da dispersão das estimativas obtidas; (ii) as estimativas perfilada e condicional tem comportamentos semelhantes entre si, assim como as estimativas integradas uniforme e de Jeffreys; (iii) para probabilidades de captura entre P_1 e P_2 , muitas das estimativas perfilada e condicional foram muito discrepantes, distanciando-se muito do valor verdadeiro de N ; (iv) as estimativas integradas uniforme e de Jeffreys, em geral, subestimam o valor verdadeiro do parâmetro N , sendo esta característica amenizada com o aumento das probabilidades e do número de épocas de captura.

A Figura 8 ilustra o gráfico das estimativas no caso em que $N = 50$ e $k = 3, 6$ e 10 . A partir desta figura nota-se que, para probabilidades de captura em P_1 todos os estimadores subestimam o valor verdadeiro de N . Para as probabilidades de captura em P_3 e P_4 , as estimativas perfilada e condicional tiveram mediana muito próxima ao valor verdadeiro de N , mas apresentaram pontos muito discrepantes acima do terceiro quartil para $k = 3$, enquanto as estimativas das verossimilhanças integradas tiveram mediana abaixo do valor verdadeiro de N para probabilidade de captura em P_1 e P_2 , apresentando poucos pontos discrepantes. Vale ressaltar que para probabilidades de captura em P_1 obteve-se 48,4%, 4,4% de estimativas perfiladas e condicionais iguais a $+\infty$, respectivamente, para $k = 3$ e 6 , as quais não foram consideradas na construção dos box-plots. Para probabilidades de

captura \mathbf{p} em P_3 a mediana das estimativas perfilada e condicional ficaram muito próximas do valor verdadeiro de N , enquanto a mediana das estimativas integradas se mantiveram pouco inferior ao valor verdadeiro de N . O comportamento de todos os estimadores são bastante similares entre si para as probabilidades de captura entre P_4 e valores de $k \geq 6$.

A Figura 9 apresenta os gráficos das estimativas para o caso $N = 800$ e $k = 3, 6$ e 10 . Nota-se que as mesmas conclusões anteriores são válidas, exceto que para probabilidades de captura em P_1 as estimativas de máxima verossimilhança perfilada e condicional tem mediana muito próxima ao valor verdadeiro de N , enquanto o terceiro quartil das estimativas das verossimilhanças integradas estão abaixo do valor verdadeiro de N . Ao comparar os gráficos das Figuras 8 e 9, observamos que a dispersão de todos os estimadores diminui com o aumento de k , em especial quando as probabilidades de captura estão entre em P_3 ou P_4 .

Para os demais valores de tamanhos populacionais N e números de épocas de captura k , observamos um comportamento análogo aos comentados acima.

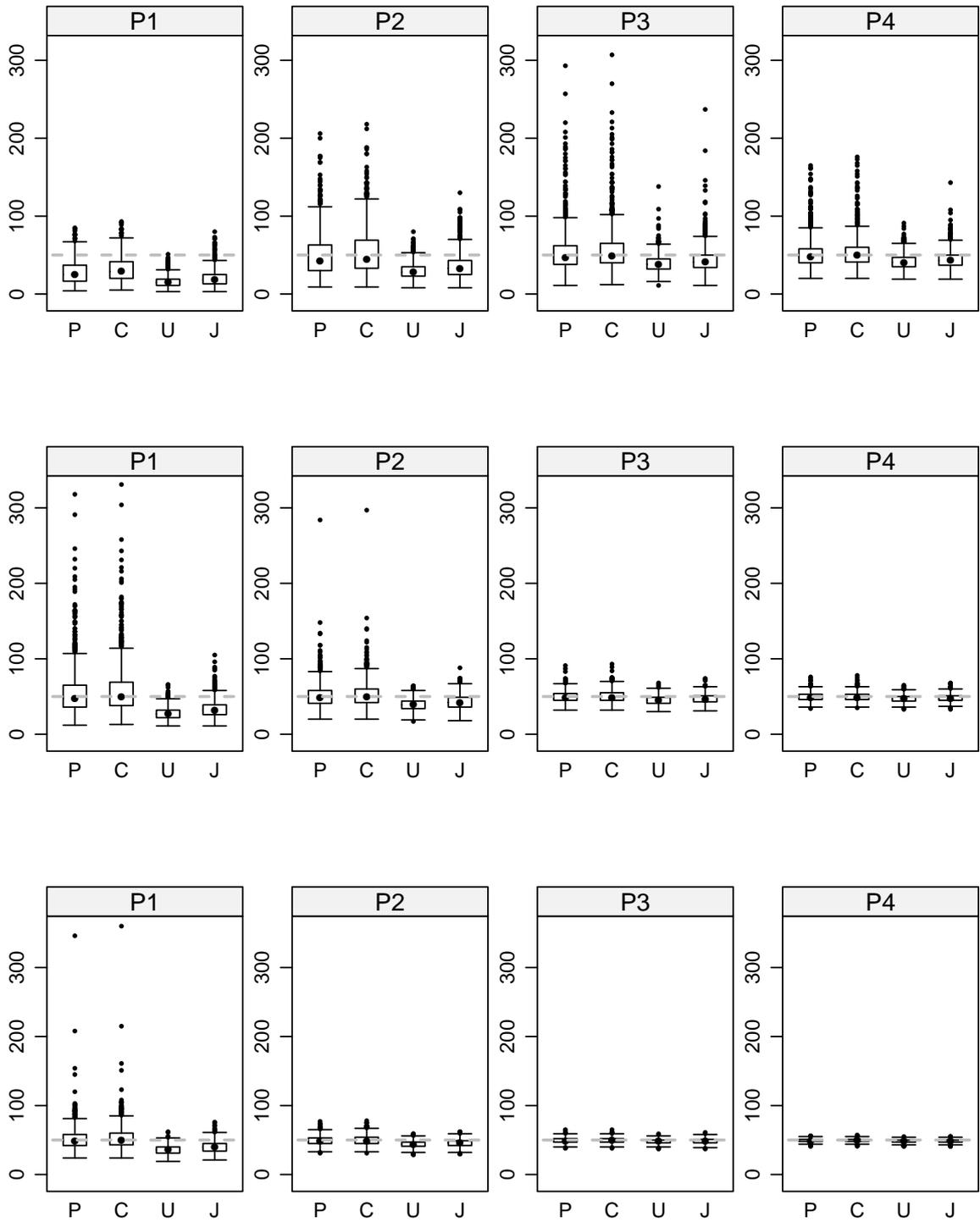


Figura 8. Estimativas de máxima verossimilhança perfilada (P), condicional (C), integrada uniforme (U) e integrada de Jeffreys (J) para amostras simuladas com $N = 50$ (linha tracejada cinza), $k = 3$ (acima), $k = 6$ (meio) e $k = 10$ (abaixo).

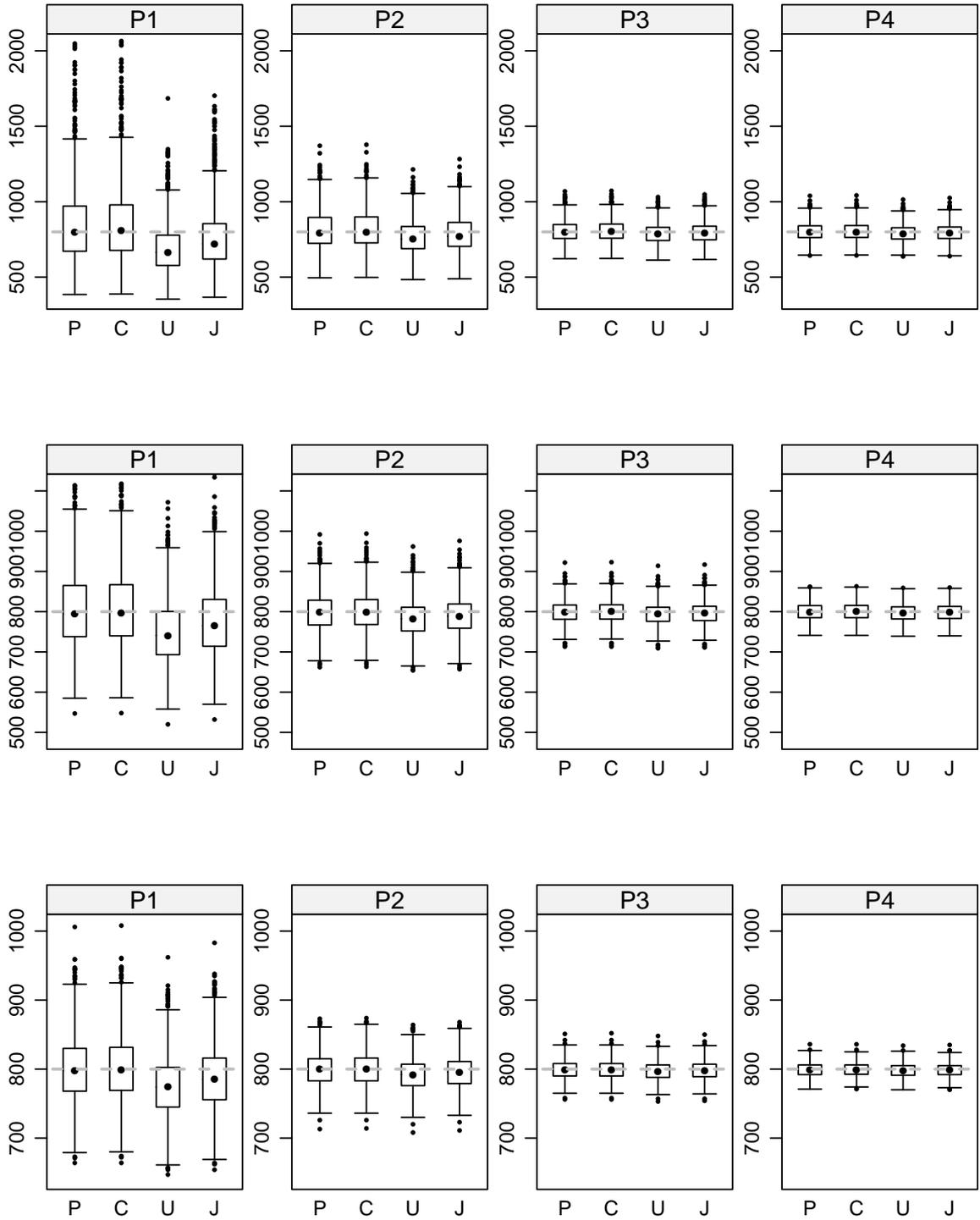


Figura 9. Estimativas de máxima verossimilhança perfilada (P), condicional (C), integrada uniforme (U) e integrada de Jeffreys (J) para amostras simuladas com $N = 800$ (linha tracejada cinza), $k = 3$ (acima), $k = 6$ (meio) e $k = 10$ (abaixo).

Para s -ésima amostra simulada e cada combinação dos valores fixados dos parâme-

tros N , \mathbf{p} e k , as estimativas de máxima verossimilhança perfilada $\widehat{N}^{\mathcal{P}}(s)$, condicional $\widehat{N}^{\mathcal{C}}(s)$, integrada uniforme $\widehat{N}^{\mathcal{U}}(s)$ e integrada de Jeffreys $\widehat{N}^{\mathcal{J}}(s)$ foram determinadas, $s = 1, 2, \dots, 1000$. Para avaliar as propriedades frequentistas dos estimadores, definimos as seguintes medidas descritivas

$$\text{REQM}(\widehat{N}^{\mathcal{M}}) = \sqrt{\frac{\sum_{s=1}^{1000} \left(\widehat{N}^{\mathcal{M}}(s) - N \right)^2}{1000}},$$

$$\text{VR}(\widehat{N}^{\mathcal{M}}) = \frac{1}{1000} \sum_{s=1}^{1000} \frac{\widehat{N}^{\mathcal{M}}(s) - N}{N},$$

onde REQM é a raiz quadrada do erro quadrático médio, VR é o vício relativo das estimativas e \mathcal{M} assume \mathcal{P} (perfilada), \mathcal{C} (condicional), \mathcal{U} (integrada uniforme), \mathcal{J} (integrada de Jeffreys). Para algumas amostras simuladas, as estimativas de máxima verossimilhança perfilada e condicional são infinitas, o que ocorre em geral quando as probabilidades e o número de épocas de captura é pequeno. Nestes casos, definimos REQM e VR como sendo ∞ . As Figuras 10, 11, 12, 13, 14 mostram como estas medidas descritivas se comportam em função de k para os valores de $N = 50, 100, 200, 400, 800$, respectivamente.

A partir da análise destes gráficos de REQM e VR em função de k para diferentes combinações de valores fixados de N e das probabilidades de captura \mathbf{p} , concluímos que: (i) um aumento tanto das probabilidades quanto do número de épocas de captura faz com que haja uma redução de REQM e VR (em valor absoluto); (ii) os estimadores das verossimilhanças integradas tem vício relativo negativo, enquanto os estimadores da perfilada e condicional apresentam pequeno vício positivo; (iii) o vício relativo (em valor absoluto) da integrada de Jeffreys é menor que o da integrada uniforme em todos os cenários considerados; (iv) o desempenho dos estimadores da perfilada e condicional são muito similares entre si, enquanto que o estimador da Integrada Jeffreys tem desempenho igual ou superior ao da Integrada Uniforme; (v) para baixa probabilidade de captura (P_1 ou P_2), quando o número de épocas de captura é pequeno ou moderado ($k \leq 7$) os estimadores da perfilada e condicional assumem maiores valores de REQM (possivelmente ∞) do que os estimadores das integradas, enquanto que se o número de épocas de captura for grande $k \geq 8$, todos os estimadores tem desempenho similar; (vi) para probabilidade de captura alta (P_3 ou P_4), os valores de REQM são muito similares entre todos os estimadores quando $N \geq 200$, enquanto assumem valores menores para os estimadores das integradas do que os da perfilada e condicional quando $N = 50, 100$, notadamente quando $k \leq 5$.

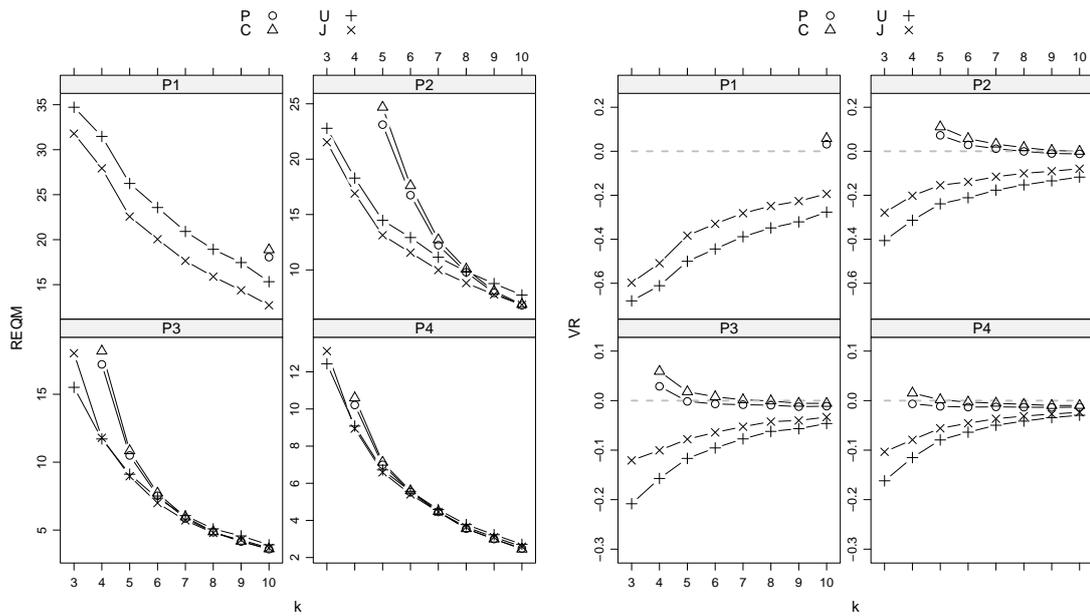


Figura 10. Raiz do erro quadrático médio REQM (esquerda) e vício relativo VR (direita) para as estimativas de máxima verossimilhança perfilada (P), condicional (C), integrada uniforme (U) e integrada de Jeffreys (J), $N = 50$.

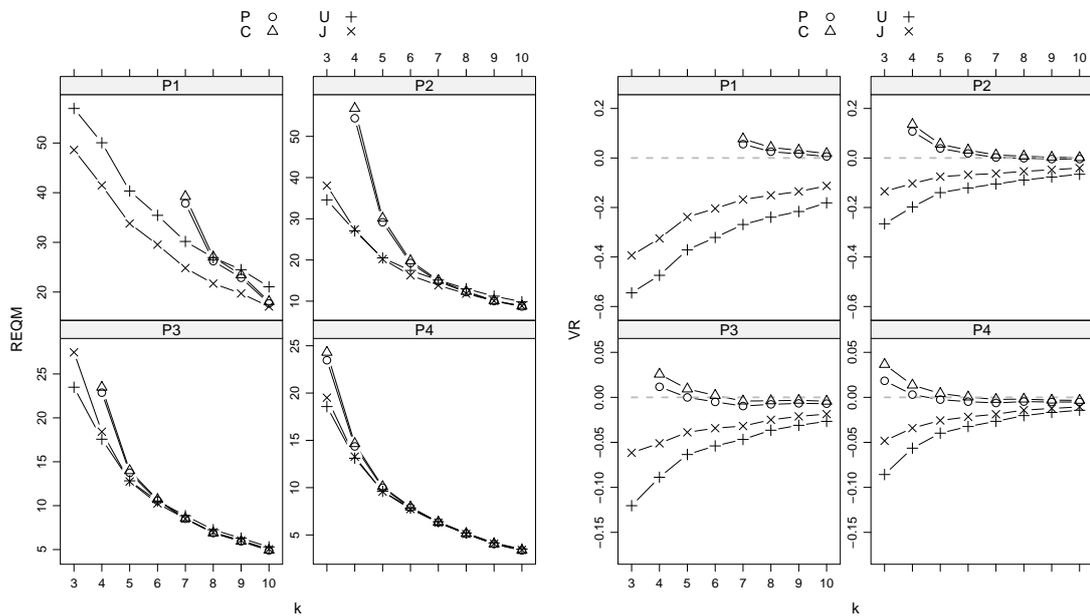


Figura 11. Raiz do erro quadrático médio REQM (esquerda) e vício relativo VR (direita) para as estimativas de máxima verossimilhança perfilada (P), condicional (C), integrada uniforme (U) e integrada de Jeffreys (J), $N = 100$.

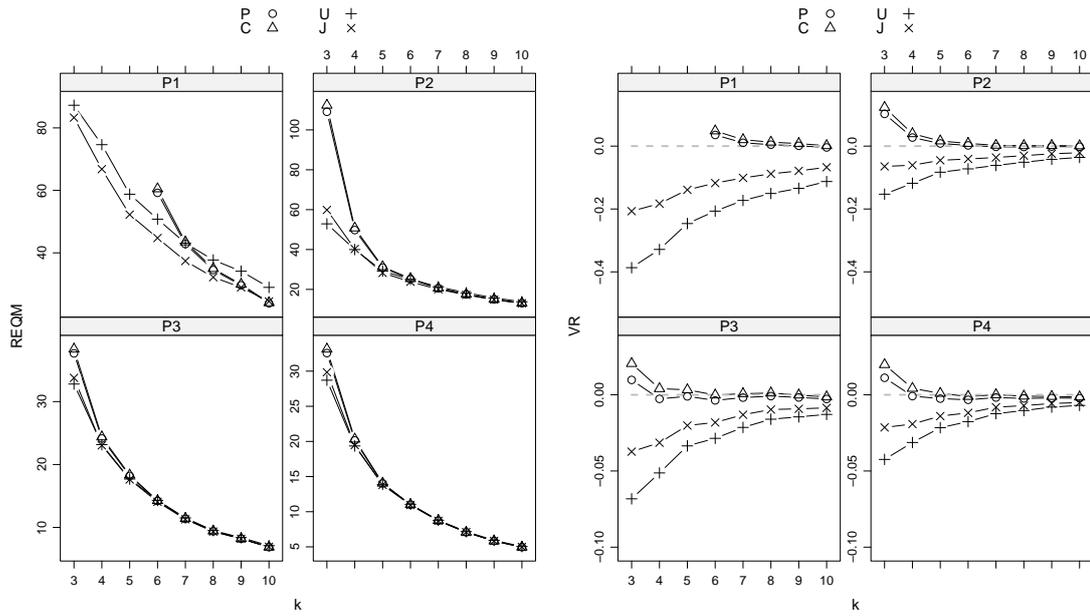


Figura 12. Raiz do erro quadrático médio REQM (esquerda) e vício relativo VR (direita) para as estimativas de máxima verossimilhança perfilada (P), condicional (C), integrada uniforme (U) e integrada de Jeffreys (J), $N = 200$.

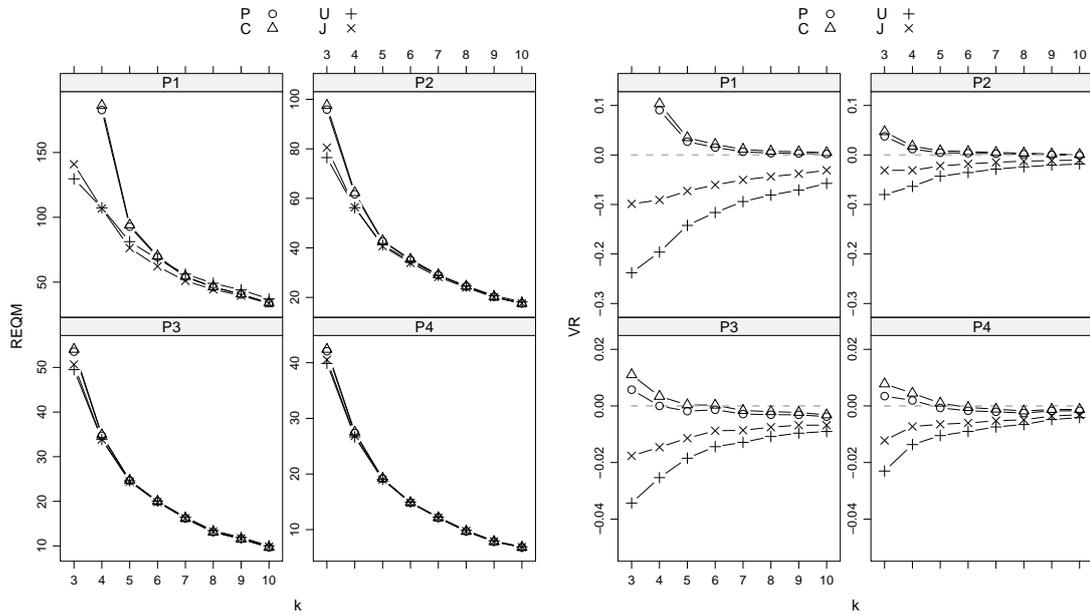


Figura 13. Raiz do erro quadrático médio REQM (esquerda) e vício relativo VR (direita) para as estimativas de máxima verossimilhança perfilada (P), condicional (C), integrada uniforme (U) e integrada de Jeffreys (J), $N = 400$.

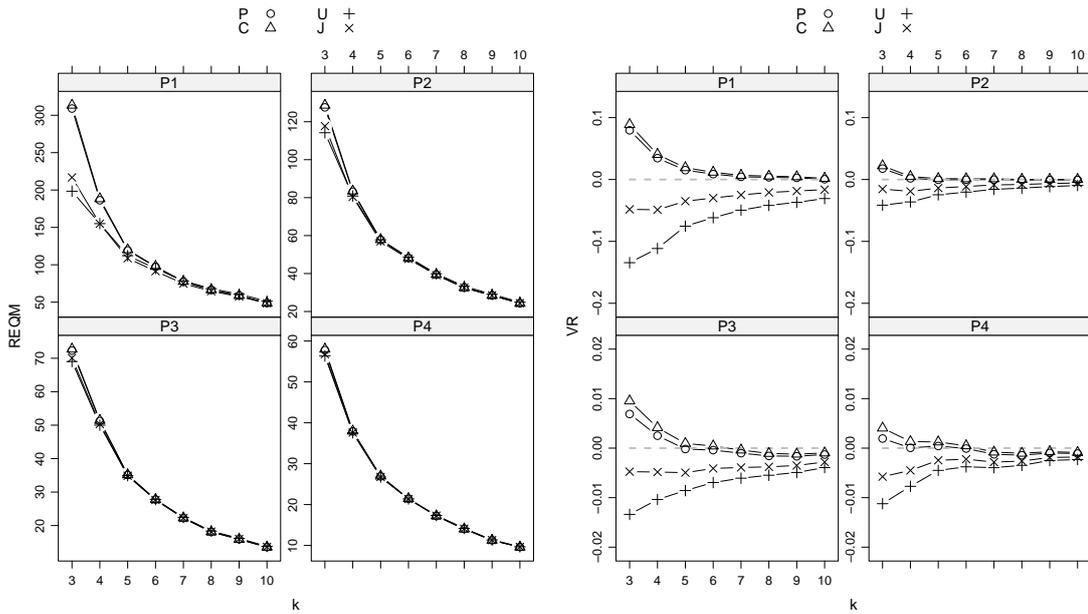


Figura 14. Raiz do erro quadrático médio REQM (esquerda) e vício relativo VR (direita) para as estimativas de máxima verossimilhança perfilada (P), condicional (C), integrada uniforme (U) e integrada de Jeffreys (J), $N = 800$.

4.2 Estimação Intervalar

Para s -ésima amostra simulada e cada configuração dos parâmetros, os intervalos de confiança de 95% $[N_L^{\mathcal{P}}(s), N_U^{\mathcal{P}}(s)]$ (perfilada), $[N_L^{\mathcal{C}}(s), N_U^{\mathcal{C}}(s)]$ (condicional), $[N_L^{\mathcal{U}}(s), N_U^{\mathcal{U}}(s)]$ (integrada uniforme) e $[N_L^{\mathcal{J}}(s), N_U^{\mathcal{J}}(s)]$ (integrada de Jeffreys) e $[N_L^{\mathcal{W}}(s), N_U^{\mathcal{W}}(s)]$ (Wald) foram determinados, $s = 1, 2, \dots, S = 1.000$. Para cada método intervalar, definimos as seguintes medidas descritivas

$$PC = \frac{1}{1000} \sum_{k=1}^S \mathbb{1}_{[N_L^{\mathcal{M}}(s), N_U^{\mathcal{M}}(s)]}(N),$$

$$AR = \frac{1}{1000} \sum_{k=1}^S \frac{N_U^{\mathcal{M}}(s) - N_L^{\mathcal{M}}(s)}{N},$$

com $\mathbb{1}_A(\cdot)$ denotando a função indicadora do conjunto A , PC a proporção de cobertura e AR a amplitude média relativa das estimativas intervalares, \mathcal{M} assumindo \mathcal{P} (perfilada), \mathcal{C} (condicional), \mathcal{U} (integrada uniforme), \mathcal{J} (integrada de Jeffreys) e \mathcal{W} (Wald). Como para algumas das amostras simuladas a estimativa de máxima verossimilhança perfilada e condicional é igual a $+\infty$ (vide Tabela 6), o intervalo de confiança associado é da forma $[N_1, \infty)$, ou seja, tem amplitude infinita. Para as configurações de parâmetros em que se obteve pelo menos um intervalo desta forma, definimos a Amplitude Média relativa

(AR) como $+\infty$. As Figuras 15, 16, 17, 18, 19 mostram como estas medidas descritivas se comportam em função de k para os valores de $N = 50, 100, 200, 400, 800$, respectivamente.

A partir da análise dos gráficos de PC e AR em função de k para todos os tamanhos populacionais N considerados, concluímos que: (i) o aumento do número de elementos N implica em valores de PC mais próximos de 0,95 e em menores valores de AR; (ii) o aumento das probabilidades de captura \mathbf{p} implica em proporções de cobertura mais próximas do coeficiente de confiança nominal (0,95) e em menores valores de AR; (iii) o aumento de k mantendo N e \mathbf{p} fixos não causa mudança significativa em PC; (iv) os intervalos da perfilada e da condicional tem desempenho bastante similar em todos os cenários considerados; (v) os intervalos podem ser colocados em ordem crescente de acordo com AR como: integrada uniforme, integrada de Jeffreys, Wald, perfilada e condicional; (vi) o intervalo da integrada uniforme tem desempenho muito abaixo dos demais com respeito a PC, particularmente para probabilidades de captura em P_1 e P_2 ; (vii) para $N \geq 400$, todos os intervalos tem desempenho similar quando as probabilidades de captura estão em P_3 e P_4 , o intervalo de Jeffreys tem desempenho superior quando as probabilidades de captura estão em P_1 e P_2 ; (viii) para $N = 50, 100, 200$, o intervalo da perfilada, condicional e Wald tem PC levemente mais próxima de 0,95 do que o intervalo de Jeffreys, embora assumam maiores valores de AR, possivelmente assumindo o valor ∞ (ou não sendo definido no caso do intervalo Wald) quando \mathbf{p} está em P_1 e P_2 .

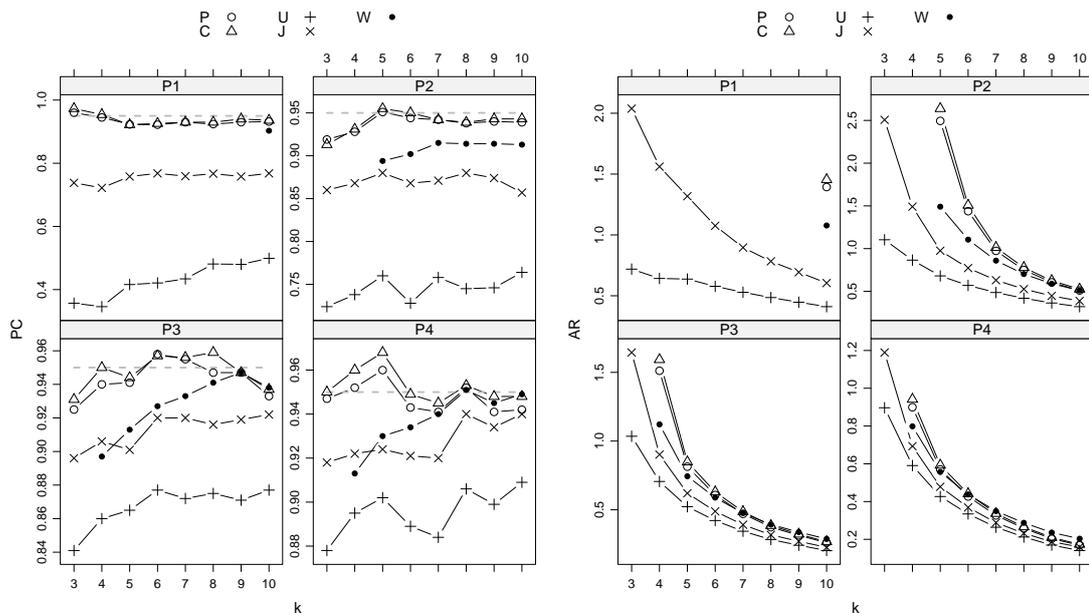


Figura 15. Proporção de cobertura PC (esquerda) e amplitude relativa (direita) dos intervalos da perfilada (P), condicional (C), integrada uniforme (U), integrada de Jeffreys (J) e Wald (W) para $N = 50$.

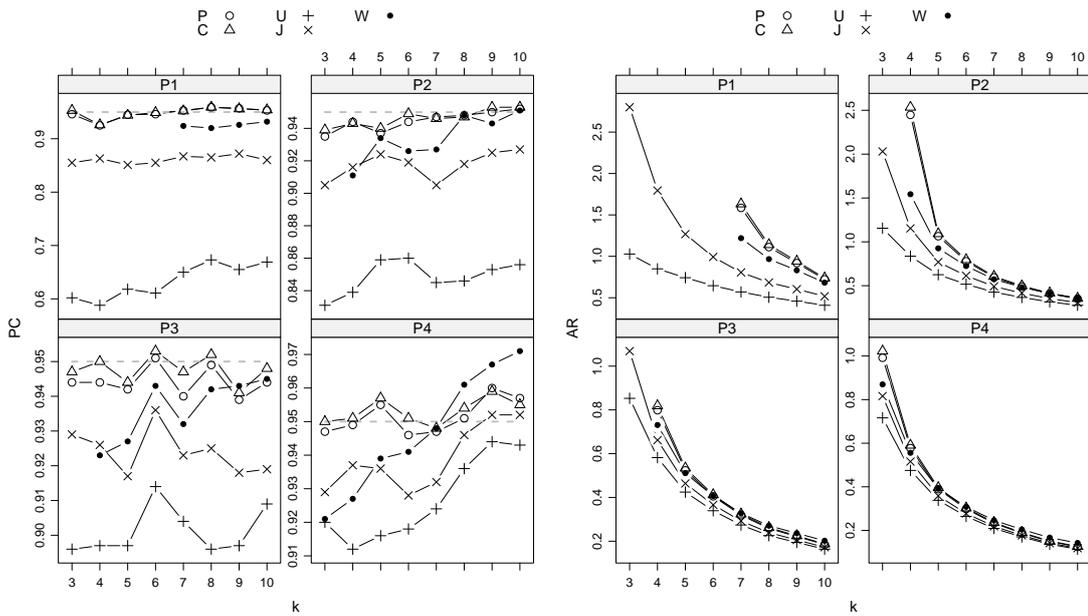


Figura 16. Proporção de cobertura PC (esquerda) e amplitude relativa (direita) dos intervalos da perfilada (P), condicional (C), integrada uniforme (U), integrada de Jeffreys (J) e Wald (W) para $N = 100$.

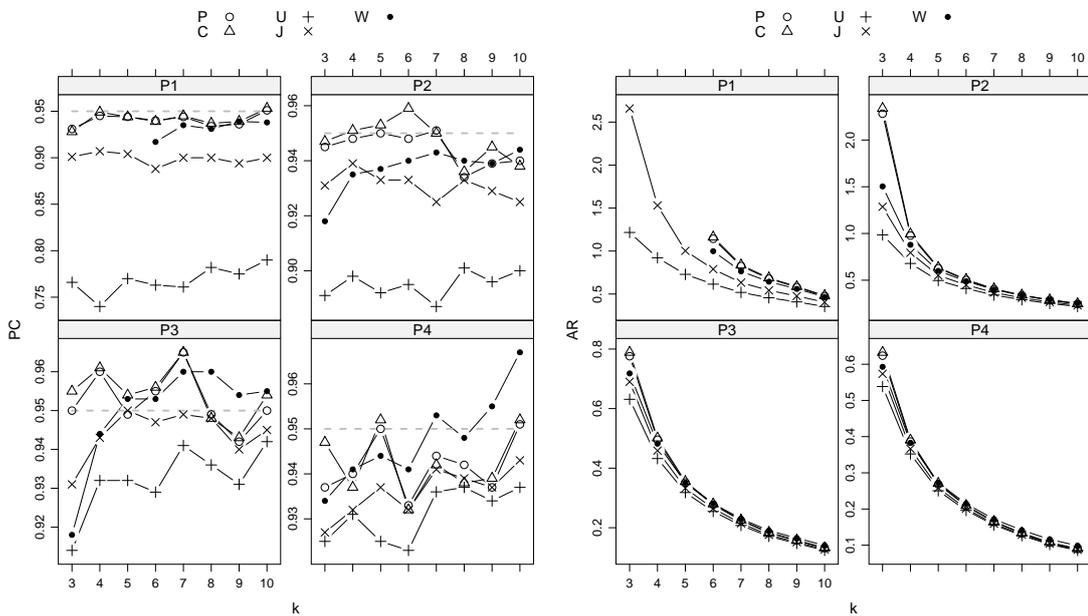


Figura 17. Proporção de cobertura PC (esquerda) e amplitude relativa (direita) dos intervalos da perfilada (P), condicional (C), integrada uniforme (U), integrada de Jeffreys (J) e Wald (W) para $N = 200$.

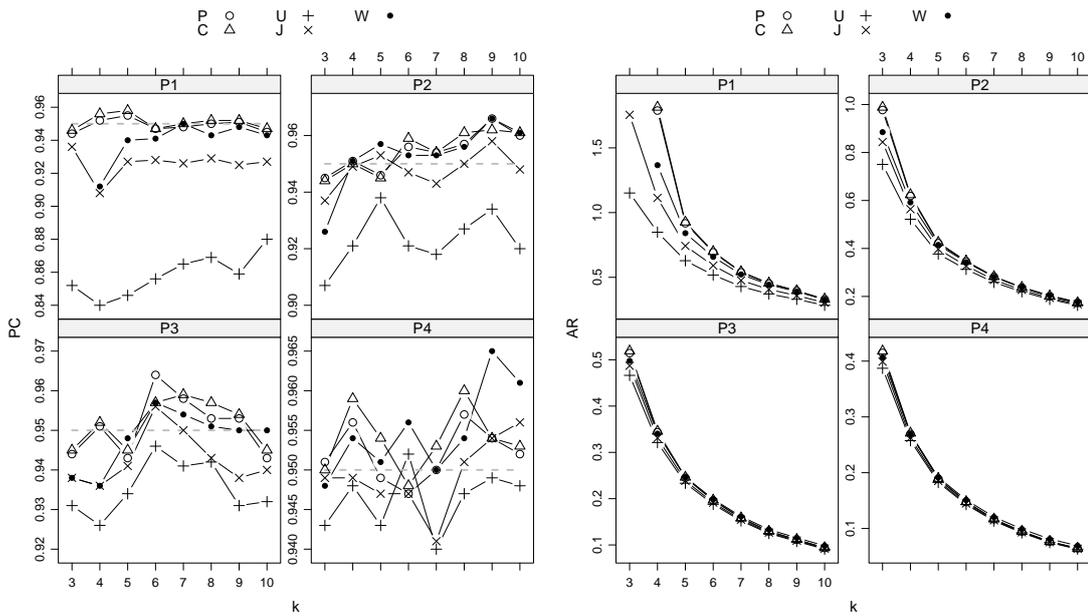


Figura 18. Proporção de cobertura PC (esquerda) e amplitude relativa (direita) dos intervalos da perfilada (P), condicional (C), integrada uniforme (U), integrada de Jeffreys (J) e Wald (W) para $N = 400$.

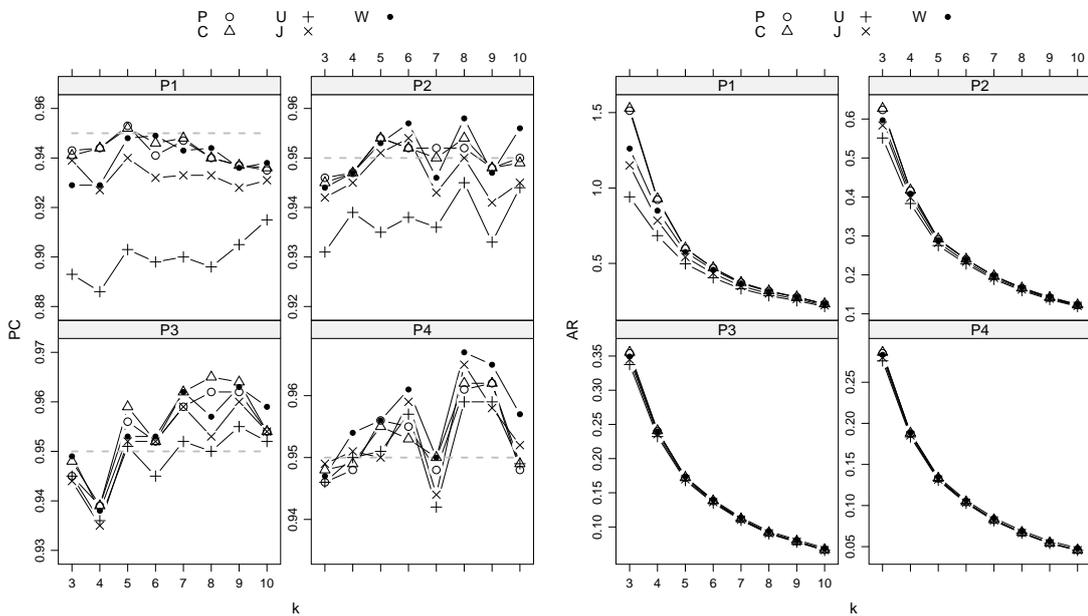


Figura 19. Proporção de cobertura PC (esquerda) e amplitude relativa (direita) dos intervalos da perfilada (P), condicional (C), integrada uniforme (U), integrada de Jeffreys (J) e Wald (W) para $N = 800$.

Ao considerar uma combinação dos critérios de maior proximidade das proporções de cobertura do coeficiente de confiança nominal e de menor amplitude média, podemos dizer que: (i) o intervalo da Integrada Uniforme tem o pior desempenho dentre todos os

métodos considerados por ter proporção de cobertura muito abaixo dos demais métodos; (ii) o intervalo da integrada de Jeffreys teve o melhor desempenho quando $N \leq 200$, particularmente para probabilidades de captura em P_1 e P_2 ; (iii) Para $N \geq 400$, os intervalos da perfilada, conditional, Jeffreys e Wald tiveram desempenho similar.

5 *Considerações Finais*

Neste trabalho consideramos métodos de eliminação de parâmetros perturbadores para um modelo de captura-recaptura, amplamente discutido na literatura, para uma população fechada. Diferentes funções de verossimilhança exclusivamente dependentes do parâmetro de interesse N , tamanho populacional, foram obtidas. Para cada uma das funções de verossimilhança, o comportamento da função foi caracterizado e assim uma expressão para a estimativa de máxima verossimilhança foi determinada. Mostrou-se que as estimativas de máxima verossimilhança perfilada e condicional podem assumir o valor infinito e que para as estimativas integradas uniforme e de Jeffreys isto nunca ocorre. As funções de verossimilhança perfilada e condicional podem assumir dois pontos de máximo, o que implica em duas possíveis estimativas de máxima verossimilhança. No caso das funções de verossimilhança integrada uniforme e de Jeffreys, mostrou-se que existe um único ponto de máximo.

A partir das funções de verossimilhança estudadas, podemos definir intervalos de confiança que somente se baseiam na função de verossimilhança. Este método de construção de intervalos tem apelo intuitivo, pois inclui os pontos mais plausíveis de acordo com a função de verossimilhança. Além disso, o intervalo de confiança do tipo Wald, que é comumente usado na prática, pode ter limite inferior menor do que o número de indivíduos distintos capturados e até pode assumir um valor negativo. Esta abordagem para construção de intervalos de confiança baseia-se na suposição de normalidade assintótica do estimador de máxima verossimilhança, no entanto na maioria dos casos a função de verossimilhança é assimétrica.

Os métodos propostos foram ilustrados para um conjunto de dados reais e as propriedades frequentistas dos estimadores foram avaliadas a partir de um estudo de simulação. Os estimadores pontual e intervalar da função de verossimilhança integrada de Jeffreys mostraram-se adequados em comparação com os demais na situação em que há pouca informação nos dados, isto é, quando tem-se poucos elementos na população, poucas épocas de captura e baixa probabilidade de captura.

Acreditamos que este trabalho é um primeiro passo no estudo do comportamento das funções de verossimilhança livres de parâmetros perturbadores para modelos de captura-recaptura. Um passo adiante, seria estudar o comportamento de funções de verossimilhanças integradas em que a função de densidade de probabilidade do parâmetro perturbador seja mais geral e que, possivelmente, dependa do número de ocasiões de captura. Uma outra contribuição seria dada pelo estudo do valor exato do vício dos estimadores e de métodos para correção deste vício. Por outro lado, o estudo da distribuição assintótica dos estimadores e da estatística da razão de verossimilhança seria importante para esta classe de distribuições.

Além do modelo de captura-recaptura estudado neste trabalho, os métodos de eliminação de parâmetros perturbadores podem ser aplicados a outros modelos discutidos na literatura (OTIS et al., 1978). Uma possibilidade interessante seria aplicar estes métodos para experimentos em que o número de ocasiões de captura não é fixado de antemão, mas dependa dos dados, uma vez que o aumento do número de ocasiões determina um melhor desempenho dos estimadores estudados.

Referências

- AMSTRUP, S. C.; MCDONALD, T. L.; MANLY, B. F. J. (Ed.). *Handbook of Capture-Recapture Analysis*. London: John Wiley, 2003.
- BASU, D. On the elimination of nuisance parameters. *Journal of the American Statistical Association*, v. 72, p. 355–366, 1977.
- BASU, S.; EBRAHIMI, N. Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika*, v. 88, p. 269–279, 2001.
- BERGER, J. O.; LISEO, B.; WOLPERT, L. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, v. 14, p. 1–28, 1999.
- BOENDER, C. G. E.; RINOOY KAN, A. H. G. A multinomial bayesian approach to the estimation of population and vocabulary size. *Biometrika*, v. 74, p. 849–856, 1987.
- CASTLEDINE, B. J. A bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika*, v. 67, p. 197–210, 1981.
- CHAO, A. An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, v. 6, p. 158–175, 2001.
- CHAO, A. et al. The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, v. 20, p. 3123–3157, 2001.
- CHAPMAN, D. G. The estimation of biological populations. *The Annals of Mathematical Statistics*, v. 25, n. 1, p. 1–15, 1954.
- CORMACK, R. M. The statistics of capture-recapture methods. *Oceanography and Marine Biology*, v. 6, p. 455–506, 1968.
- COX, D. R. Partial likelihood. *Biometrika*, v. 62, p. 269–276, 1975.
- DARROCH, J. N. The multiple-recapture census: I. Estimation of a closed population. *Biometrika*, v. 45, p. 343–359, 1958.
- DARROCH, J. N. The multiple-recapture census: II. Estimation when there is immigration or death. *Biometrika*, v. 46, p. 336–351, 1959.
- GEORGE, E. I.; ROBERT, C. P. Capture-recapture estimation via gibbs sampling. *Biometrika*, v. 79, p. 677–683, 1992.
- GERKING, S. D. Vital statistics of the fish population of gordy lake, indiana. *Transactions of the American Fisheries Society*, v. 82, p. 48–67, 1953.

- JOLLY, G. M. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, v. 52, p. 225–247, 1965.
- LAPLACE, P. S. Sur les naissances, les mariages, et les morts. In: _____. *Historie de L'Academie Royale des Sciences*. Paris: [s.n.], 1783.
- LEE, A. J. Effect of list errors on the estimation of population size. *Biometrics*, v. 58, p. 185–191, 2002.
- LEE, A. J. et al. Capture-recapture, epidemiology, and list mismatches: several lists. *Biometrics*, v. 57, p. 707–713, 2001.
- LINCOLN, F. C. *Calculating Waterfowl Abundance on the Basis of Banding Returns*. [S.l.], 1930. v. 118, 1–4 p.
- NAYAK, T. K. Estimating population size by recapture sampling. *Biometrika*, v. 75, p. 113–4120, 1988.
- OTIS, D. L. et al. Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, v. 62, p. 3–135, 1978.
- PETERSEN, C. G. *The yearly immigration of young plaice into the Limfjord from the German Sea*. [S.l.], 1896.
- POLLOCK, K. H. 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*, v. 86, p. 225–238, 1991.
- POLLOCK, K. H. Capture-recapture models. *Journal of the American Statistical Association*, v. 95, p. 293–296, 2000.
- RICKER, W. E. *Computation and Interpretation of Biological Statistics of Fish Populations*. [S.l.], 1975.
- SCHWARZ, C. J.; SEBER, G. A. F. Estimating animal abundance: Review III. *Statistical Science*, v. 14, p. 427–456, 1999.
- SEBER, G. A. F. A note on multiple recapture census. *Biometrika*, v. 52, p. 249–259, 1965.
- SEBER, G. A. F. *The estimation of animal abundance and related parameters*. London: John Wiley, 1982.
- SEBER, G. A. F. A review of estimating animal abundance. *Biometrics*, v. 42, p. 267–292, 1986.
- SEBER, G. A. F. A review of estimating animal abundance II. *International Statistical Review*, v. 60, p. 129–166, 1992.
- SEBER, G. A. F.; HUAKAU, J. T.; SIMMONS, D. Capture-recapture, epidemiology and lists mismatches: Two lists. *Biometrics*, v. 56, p. 1227–1232, 2000.
- SMITH, P. J. Bayesian methods for capture-recapture surveys. *Biometrics*, v. 44, p. 1177–1189, 1988.

SMITH, P. J. Bayesian analyses for a multiple capture-recapture model. *Biometrika*, v. 78, p. 399–407, 1991.

THISTED, R.; EFRON, B. Did shakespeare write a newly-discovery poem? *Biometrika*, v. 74, p. 445–455, 1987.

WHITE, G. C.; GARROT, R. A. *Analysis of Wildlife Radio-Tracking Data*. New York: Academic Press, 1990.