

# UMA REPARAMETRIZAÇÃO DAS PROBABILIDADES DE CAPTURA EM UM MODELO BAYESIANO DE CAPTURA-RECAPTURA

Marcelo de PAULA<sup>1</sup>  
Carlos Alberto Ribeiro DINIZ<sup>2</sup>  
José Galvão LEITE<sup>2</sup>

- RESUMO: A proposta deste trabalho é apresentar uma reparametrização para as probabilidades de captura de um processo de captura-recaptura com o objetivo de verificar possíveis vantagens e desvantagens em relação ao modelo bayesiano proposto por Castledine (1981). Para isso foram consideradas duas diferentes *prioris* de referência para as probabilidades de captura do modelo bayesiano proposto por Castledine a fim de comparar as estimativas *a posteriori* deste com as estimativas *a posteriori* do modelo onde as probabilidades de captura foram reparametrizadas. A-presentamos um estudo de simulação com diferentes probabilidades de capturas com o objetivo de concretizar tal comparação. As estimativas bayesianas do tamanho populacional foram obtidas através de algoritmos de simulação estocástica, via MCMC (Monte Carlo Markov Chain).
- PALAVRAS-CHAVE: Processo de captura-recaptura; população fechada; análise bayesiana; simulação estocástica.

## 1 Introdução

O processo de captura e recaptura, no caso da estimação do tamanho populacional, consiste em selecionar uma amostra de tamanho  $n_1$  de uma população marcando-a e devolvendo-a à população. Após um certo período de tempo seleciona-se uma segunda amostra aleatória de tamanho  $n_2$ , conta-se o número de elementos marcados e marca-se os elementos não marcados, devolvendo-os à população. Após

---

<sup>1</sup>Instituto de Ciências Ambientais e Desenvolvimento Sustentável, Universidade Federal da Bahia – UFBA, CEP: 47805-100, Barreiras, BA, Brasil. E-mail: [marcelop@ufba.br](mailto:marcelop@ufba.br)

<sup>2</sup>Departamento de Estatística, Universidade Federal de São Carlos – UFSCar, CEP: 13565-905, São Carlos, SP, Brasil. E-mail: [dcad@power.ufscar.br](mailto:dcad@power.ufscar.br) / [leite@ufscar.br](mailto:leite@ufscar.br)

um certo período de tempo seleciona-se uma terceira amostra de tamanho  $n_3$ , conta-se o número de elementos marcados, marca-se os elementos não marcados, devolvendo-os à população, e assim por diante. Esse processo é realizado  $s$  vezes ( $s \geq 2$ ).

O interesse em estimar tamanhos de populações surgiu em meados do século XVII, segundo White *et al.* (1982). Historicamente, Laplace (1786) utilizou tal processo para estimar o tamanho da população da França. Em ecologia, o primeiro pesquisador a empregar este método foi o dinamarquês Carl G. J. Petersen (1896), que estudou o fluxo migratório de peixes no mar Báltico.

As técnicas de captura-recaptura podem ser usadas para populações fechadas ou abertas. Uma *população fechada* é aquela em que os efeitos de nascimento, mortalidade e migração não são considerados, isto é, supõem-se que seu tamanho não se altera durante o período de estudo (Comack, 1992). Uma *população aberta* é aquela que durante a realização do experimento se altera em tamanho e em composição por ocorrência de nascimentos, mortes e migrações. Com relação ao estudo de populações abertas, vários outros autores destacaram-se. Entre os mais citados estão Jolly e Seber (1965), Pollock (1991) e Schwarz e Arnason (1996).

Basicamente existem três abordagens para a estimação do tamanho populacional, a partir de amostras obtidas pelo processo de captura-recaptura: a primeira é a abordagem clássica (Otis *et al.* (1978), Engen (1978)), a segunda é a Bayesiana (Smith (1988), Mingoti (2000), Wang (2002), Madigan *et al.* (1997), Castledine (1981), Hunter e Griffiths (1978), Castledine (1981), Smith (1988,1991), George e Robert (1992), Ananda (1997), Yoshida (1996), Yoshida *et al.* (1999), Leite *et al.* (2000), King *et al.* (2001)) e a terceira é a relacionada com a aplicação de modelos log-lineares para tabelas de contingência incompletas (Coull *et al.* (1999), Cormarck (1989), Bishop *et al.* (1975), Abeni (1994), Rivest *et al.* (2004)). Neste trabalho faremos o estudo sob o enfoque bayesiano.

A maior parte das aplicações do processo de captura-recaptura diz respeito à inferência sobre tamanhos de populações animais (Seber (1986, 1992), Pollock (1974, 1991), Pollock *et al.* (1983, 1984)). No entanto, mais recentemente, esta metodologia passou a ser utilizada em estudos sociais e epidemiológicos, oceanografia, dinâmica de frota de veículos, modelagem de demografia de insetos, dentre outras aplicações.

Com o objetivo de se obter aproximações dos valores de expressões analíticas complexas, as programações desenvolvidas neste trabalho para o estudo de simulação estocástica via **Monte Carlo Markov Chain (MCMC)**, algoritmo de *Gibbs Sampling* ("Amostrador de Gibbs", ver Geman e Geman (1984) e Gelfand e Smith (1990), Casella *et al.* (1992), Smith *et al.* (1993)) e algoritmo de *Metropolis Hastings* (ver Metropolis *et al.* (1953), Hastings (1970), Chib *et al.* (1995)) foram implementadas no Software **R**, cuja convergência foi monitorada pelo pacote **CODA - Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output** (Best *et al.* 1995).

## 2 Metodologia

Nesta seção apresentamos o modelo estatístico e determinamos a função de verossimilhança para o método de captura-recaptura com  $s$  estágios de marcação ( $s \geq 2$ ) proposto por Castledine (1981). Denotemos por

- $N$  : tamanho desconhecido da população,
- $s$  : número de amostras selecionadas (épocas de captura),  $s \geq 2$ ,
- $p_j$  : probabilidade de qualquer animal ser capturado na  $j$ -ésima amostra,  $j = 1, 2, \dots, s$ ,
- $\mathbf{p} = (p_1, p_2, \dots, p_s)$ ,
- $n_j$  : número de animais capturados na  $j$ -ésima amostra,  $j = 1, 2, \dots, s$ ,
- $m_j$  : número de animais marcados recapturados na  $j$ -ésima amostra,  $j = 1, 2, \dots, s$ .

Vamos supor que as seguintes condições sejam verificadas:

1. a população é fechada;
2. não há animais marcados na população no início do processo, isto é,  $m_1 = 0$ ;
3. os animais comportam-se independentemente uns dos outros;
4. as marcas não afetam a capturabilidade do animal;
5. os animais não perdem suas marcas durante o processo;
6. as épocas de amostragem são independentes.

Neste caso, a função de verossimilhança (ver por exemplo Zacharias (2000) e Paula (2006)) é tal que

$$L(N, \mathbf{p} | D) = P(n_1, m_1; n_2, m_2; \dots; n_s, m_s | \mathbf{p}, N) \propto \binom{N}{r} \prod_{j=1}^s p_j^{n_j} (1 - p_j)^{N - n_j}, N \geq r, \quad (1)$$

onde  $D = (n_1, m_1; n_2, m_2; \dots; n_s, m_s)$  representa os dados,  $r = \sum_{j=1}^s n_j - \sum_{j=1}^s m_j$  corresponde ao número de animais distintos capturados ao longo do processo e  $0 < p_j < 1$ ,  $j = 1, 2, \dots, s$ .

### 2.1 Modelo bayesiano

Nesta seção a inferência sobre  $N$  é analisada sob o enfoque bayesiano, ou seja, utilizamos conhecimento *a priori* sobre o tamanho populacional  $N$ , como do vetor de probabilidades  $\mathbf{p} = (p_1, p_2, \dots, p_s)$ .

Suponhamos que as probabilidades de captura sejam, *a priori*, independentes e identicamente distribuídas, com  $p_j$ ,  $j = 1, 2, \dots, s$ , seguindo distribuição *Beta* ( $\alpha, \beta$ ) com  $\alpha$  e  $\beta$  conhecidos ( $\alpha > 0$ ,  $\beta > 0$ ) e que a distribuição *a priori* de  $N$  seja  $\pi(N)$ ,  $N = 1, 2, \dots$ , com  $N$  e  $\mathbf{p}$  independentes. Então, a distribuição *a priori* conjunta de  $N$  e  $\mathbf{p}$  é igual a

$$\pi(N, \mathbf{p}) = \pi(N) \pi(\mathbf{p}) = \prod_{j=1}^s \frac{1}{B(\alpha, \beta)} p_j^{\alpha-1} (1 - p_j)^{\beta-1} \pi(N), \quad (2)$$

e a distribuição *a posteriori* conjunta de  $N$  e  $\mathbf{p}$  é tal que

$$\pi(N, \mathbf{p} | D) \propto \binom{N}{r} \prod_{j=1}^s p_j^{n_j + \alpha - 1} (1 - p_j)^{N - n_j + \beta - 1} \pi(N), N \geq r. \quad (3)$$

Em muitas situações práticas, podemos não ter informação *a priori* sobre o parâmetro de interesse. Neste caso, devemos considerar uma distribuição *a priori* não-informativa. Consideramos duas diferentes *prioris* não informativas para as probabilidades de captura do modelo bayesiano proposto por Castledine (1981) a fim de comparar as estimativas *a posteriori* deste modelo com as estimativas *a posteriori* do modelo onde a probabilidade de captura foi reparametrizada, a saber:

1. *Priori* não informativa de Jeffreys, isto é, *Beta* ( $\alpha = 0, 5; \beta = 0, 5$ ):

$$\pi(p_j) = \frac{1}{B(\alpha, \beta)} p_j^{-\frac{1}{2}} (1 - p_j)^{-\frac{1}{2}}, j = 1, 2, \dots, s. \quad (4)$$

2. *Priori* não informativa de Novick and Hall's:

$$\pi(p_j) = \frac{1}{p_j(1 - p_j)}, j = 1, 2, \dots, s. \quad (5)$$

Box e Tiao (1973) discutiram as idéias de Jeffreys (1961), sobre a distribuição *a priori* para representar o estado de ausência de informação ou ignorância a respeito do comportamento probabilístico dos parâmetros. O estudo abrangeu os casos uniparamétricos e multiparamétricos. A crítica mais freqüente à análise Bayesiana é que diferentes *prioris* conduzem a diferentes respostas. Contudo, com o interesse de encontrar “objetividade” pode-se usar *prioris* não-informativas.

Se houver uma amostra pequena de dados, é necessário fazer sérias considerações para a informação *a priori*. Quando a amostra é grande, intervalos de confiança clássicos e intervalo de credibilidade Bayesiano serão quase idênticos numericamente. Para mais informações sobre *prioris* não informativas ver Novick and Hall (1965), Jeffreys (1961), Box e Tiao (1973) e Yang et. al (1994, 1995, 1996).

## 2.2 Reparametrização das probabilidades de captura

Nesta seção apresentamos uma reparametrização do vetor de probabilidades de captura  $\mathbf{p} = (p_1, p_2, \dots, p_s)$  considerando uma estrutura hierárquica. Através de um estudo de simulação verificamos o efeito desta reparametrização nas estimativas *a posteriori* dos parâmetros. Reescrevendo as probabilidades de captura como

$$p_j = \frac{\exp(\eta_j)}{1 + \exp(\eta_j)}, j = 1, 2, \dots, s, \quad (6)$$

temos que a função de verossimilhança reparametrizada é tal que

$$L(N, \boldsymbol{\eta} | D) \propto \binom{N}{r} \prod_{j=1}^s \frac{\exp\{n_j \eta_j\}}{[1 + \exp\{\eta_j\}]^N}, \quad N \geq r, \quad -\infty < \eta_j < \infty, \quad (7)$$

onde  $\boldsymbol{\eta} = \{\eta_1, \eta_2, \dots, \eta_s\}$ . De (6) temos a função de ligação

$$\log\left(\frac{p_j}{1-p_j}\right) = \eta_j, \quad -\infty < \eta_j < \infty, \quad j = 1, 2, \dots, s. \quad (8)$$

O uso dessa transformação logito é usual quando desejamos sair do espaço paramétrico de  $p_j$ ,  $0 < p_j < 1$ ,  $j = 1, 2, \dots, s$  e trabalhar no espaço paramétrico de  $\eta_j$ ,  $-\infty < \eta_j < \infty$ ,  $j = 1, 2, \dots, s$ .

### 2.3 Distribuição *a posteriori*

Para a implementação de um modelo bayesiano hierárquico adotamos as seguintes distribuições *a priori*:

$$\begin{aligned} \pi(N) &= N^{-1}; \\ \eta_j &\sim N(\mu, \sigma^2), \quad j = 1, 2, \dots, s; \\ \mu &\sim N(a, b^2); \\ \sigma^2 &\sim IG(c, d); \end{aligned} \quad (9)$$

onde  $a$ ,  $b$ ,  $c$  e  $d$  são hiperparâmetros conhecidos. De (7) e (9) segue que a distribuição *a posteriori* conjunta de  $N$ ,  $\boldsymbol{\eta} = \{\eta_1, \eta_2, \dots, \eta_s\}$ ,  $\mu$ ,  $\sigma^2$  é tal que

$$\begin{aligned} \pi(N, \boldsymbol{\eta}, \mu, \sigma^2 | D) &\propto \binom{N}{r} \prod_{j=1}^s \frac{\exp\{n_j \eta_j\}}{[1 + \exp\{\eta_j\}]^N} \\ &\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\eta_j - \mu)^2\right\} \times \exp\left\{-\frac{1}{2b^2}(\mu - a)^2\right\} \\ &\times (\sigma^2)^{-(c+1)} \exp\left\{\frac{-d}{\sigma^2}\right\} \times \frac{1}{N}. \end{aligned} \quad (10)$$

As distribuições condicionais necessárias para implementar o algoritmo *Gibbs Sampling* (ver por exemplo, Geman e Geman, 1984 e Gelfand e Smith, 1990), conjuntamente com o algoritmo de *Metropolis Hastings*, a partir de (10), são dadas a seguir.

(1) A distribuição condicional de  $N - r$ , dados  $\boldsymbol{\eta}$ ,  $\mu$ ,  $\sigma^2$  e  $D$  é tal que

$$N - r | \boldsymbol{\eta}, \mu, \sigma^2, D \sim BN \left\{ r, 1 - \prod_{j=1}^s \left[ \frac{1}{1 + e^{\eta_j}} \right] \right\}, \quad (11)$$

onde *BN* denota a distribuição binomial negativa;

(2) a distribuição condicional de  $\boldsymbol{\eta}$ , dados  $N$ ,  $\mu$ ,  $\sigma^2$  e  $D$  é tal que

$$\eta_j | N, \mu, \sigma^2, D \propto \exp \left\{ -\frac{1}{2\sigma^2} (\eta_j - \mu)^2 \right\} \times \frac{\exp \{n_j \eta_j\}}{[1 + \exp \{ \eta_j \}]^N}, \quad j = 1, 2, \dots, s; \quad (12)$$

(3) a distribuição condicional de  $\mu$ , dados  $N$ ,  $\boldsymbol{\eta}$ ,  $\sigma^2$  e  $D$  é tal que

$$\mu | N, \boldsymbol{\eta}, \sigma^2, D \sim N \left( \frac{b^2 \sum_{j=1}^s \eta_j + \sigma^2 a}{sb^2 + \sigma^2}, \frac{b^2 \sigma^2}{sb^2 + \sigma^2} \right); \quad (13)$$

(4) a distribuição condicional de  $\sigma^2$ , dados  $N$ ,  $\boldsymbol{\eta} = \{\eta_1, \eta_2, \dots, \eta_s\}$ ,  $\mu$  e  $D$  é tal que

$$\sigma^2 | N, \boldsymbol{\eta}, \mu, D \sim IG \left( \frac{s}{2} + c, \sum_{j=1}^s \frac{(\eta_j - \mu)^2}{2} + d \right). \quad (14)$$

onde  $IG$  denota a distribuição gama inversa.

### 3 Resultados e discussão

#### 3.1 Probabilidades de captura iguais

Com o objetivo de comparar as estimativas bayesianas produzidas pelo modelo bayesiano original proposto por Castledine (considerando cada uma das duas *prioris* de referência) e o modelo hierárquico com probabilidades de captura reparametrizadas foi realizado um estudo de simulação com  $s = 6$  épocas de captura, supondo uma população de tamanho  $N = 5000$ . O vetor de probabilidades é dado por  $\mathbf{p} = (p_1, p_2, \dots, p_6)$  com  $p_1 = p_2 = \dots = p_6 = p$ . Foram atribuídas 24 valores a  $p$ . Como  $p = \frac{\exp(\eta)}{1 + \exp(\eta)}$ , temos, para cada valor de  $p$ , um valor de  $\eta$  e um valor simulado da estatística  $r$ , como apresentada na Tabela 1.

Lembremos que a estatística  $r$  é dada por  $r = \sum_{j=1}^s n_j - \sum_{j=1}^s m_j$ , ou seja, é o número de animais distintos capturados durante todo o processo.

A Tabela 2 mostra os resultados *a posteriori* para o tamanho populacional  $N$  segundo os Modelos I ao Modelo III, respectivamente, a saber:

**Modelo I.** Modelo bayesiano proposto por Castledine (1981) com distribuição *a priori* não informativa de Jeffreys,  $\pi(N) \propto N^{-1}$ , para o tamanho populacional  $N$

Tabela 1 - Resultados das simulações com diferentes valores de  $p$  e  $N = 5000$

Prob. Captura	Reparam.	Estatística	Prob. Captura	Reparam.	Estatística
$p_j = p$	$\eta$	$r$	$p_j = p$	$\eta$	$r$
0,005	-5,2933	134	0,40	-0,4055	4713
0,01	-4,5951	275	0,45	-0,2007	4907
0,02	-3,8918	521	0,50	0	4969
0,03	-3,4761	874	0,55	0,2007	4910
0,04	-3,1781	1124	0,60	0,4055	5000
0,05	-2,9444	1336	0,65	0,6190	5000
0,10	-2,1972	2319	0,70	0,8473	5000
0,15	-1,7346	3111	0,75	1,0986	5000
0,20	-1,3863	3626	0,80	1,3863	5000
0,25	-1,0986	4145	0,85	1,7346	5000
0,30	-0,8473	4343	0,90	2,1972	5000
0,35	-0,6190	4556	0,95	2,9444	5000

e um produto de *prioris* não informativas  $Beta(\alpha = 0,5; \beta = 0,5)$  para o vetor de probabilidades  $\mathbf{p} = (p_1, p_2, \dots, p_6)$ .

**Modelo II.** Modelo bayesiano proposto por Castledine (1981) com distribuição *a priori* não informativa de Jeffreys,  $\pi(N) \propto N^{-1}$ , para o tamanho populacional  $N$  e um produto de *prioris* não informativas de Novick and Hall's para o vetor de probabilidades  $\mathbf{p} = (p_1, p_2, \dots, p_6)$ .

**Modelo III.** Modelo de captura-recaptura utilizado por Castledine (1981) com reparametrização das probabilidades de captura considerando uma distribuição *a priori* de Jeffreys,  $\pi(N) \propto N^{-1}$ , para o tamanho populacional  $N$  e uma distribuição *a priori* Normal com média centrada em zero e variância  $\sigma^2 = 1000$  para  $\eta_j, j = 1, 2, \dots, s$ .

Para cada um dos 24 casos considerados na Tabela 1 e considerando os modelos I e II foram geradas duas cadeias de 40000 iterações. Foram descartadas as primeiras 20000 iterações como "burn-in" e foi atribuído um salto de tamanho 20, resultando em uma amostra final de tamanho 1000 da distribuição *a posteriori* conjunta. Para o Modelo III foram geradas duas cadeias de 70000 iterações. Foram descartadas as primeiras 20000 iterações como "burn-in" e foi atribuído um salto de tamanho 50, resultando em uma amostra final de tamanho 1000 da distribuição *a posteriori* conjunta.

A Tabela 2 mostra os resumos *a posteriori* do parâmetro  $N$  (onde  $E(N | D)$  denota a média *a posteriori* e (I C. 95% de  $N$ ) denota o intervalo de credibilidade 95%) dos Modelos I ao Modelo III respectivamente.

Pelos resultados apresentados na Tabela 2, com relação ao modelo I, observa-se que as estimativas *a posteriori* dos parâmetros  $N$  são satisfatórias quando  $p_j = p \geq 0,02, j = 1, 2, \dots, 6$ . Nas situações onde  $p_j = p < 0,02$ , o tamanho populacional foi em média subestimado, como é o caso de  $p_j = 0,005$  e  $p_j = 0,01$ ,

Tabela 2 - Estimativas *a posteriori* de  $N$  referentes aos Modelos I a III

Prob.	$r$	Modelo I		Modelo II		Modelo III	
		$E(N D)$	IC.95%	$E(N D)$	IC.95%	$E(N D)$	IC.95%
0,005	134	2608	(914; 7109)	7825	(1973; 13948)	136121	(2009; 478402)
0,01	275	3633	(1993; 6696)	5347	(2532; 11404)	5405	(2544; 11604)
0,02	521	4999	(3452; 7421)	5663	(3781; 8711)	5768	(3757; 8727)
0,03	874	4872	(3598; 5318)	4501	(3686; 5490)	4508	(3713; 5506)
0,04	1124	4542	(3910; 5274)	4621	(3972; 5376)	4611	(3996; 5386)
0,05	1336	4735	(4164; 5375)	4790	(4224; 5437)	4787	(4225; 5457)
0,10	2319	4873	(4600; 5174)	4888	(4609; 5180)	4886	(4608; 5212)
0,15	3111	4954	(4780; 5140)	4960	(4794; 5144)	4957	(4782; 5132)
0,20	3626	4832	(4724; 5152)	4833	(4725; 5109)	4835	(4728; 4947)
0,25	4145	5069	(4984; 5157)	5070	(4985; 5157)	5069	(4983; 5151)
0,30	4343	4894	(4835; 5053)	4903	(4837; 5125)	4894	(4833; 4954)
0,35	4556	4909	(4865; 5054)	4915	(4867; 5112)	4910	(4868; 4954)
0,40	4713	4933	(4901; 5066)	4934	(4900; 5088)	4933	(4902; 4968)
0,45	4907	5054	(4961; 5081)	5054	(5029; 5081)	5054	(5028; 5080)
0,50	4969	5053	(4968; 5073)	5053	(5034; 5073)	5053	(5034; 5072)
0,55	4910	4997	(4989; 5018)	4988	(4936; 5038)	4948	(4936; 4961)
0,60	4998	5006	(5003; 5012)	5012	(5008; 5032)	5022	(5012; 5032)
0,65	4999	5003	(5002; 5003)	5008	(5004; 5017)	5010	(5004; 5017)
0,70	5000	5002	(5001; 5003)	5002	(5000; 5003)	5004	(5000; 5004)
0,75	5000	5001	(5001; 5002)	5001	(5000; 5002)	5001	(5000; 5004)
0,80	5000	5000	(5000; 5001)	5000	(5000; 5002)	5000	(5000; 5002)
0,85	5000	5000	(5000; 5001)	5000	(5000; 5001)	5000	(5000; 5001)
0,90	5000	5000	(5000; 5000)	5000	(5000; 5000)	5000	(5000; 5000)
0,95	5000	5000	(5000; 5000)	5000	(5000; 5000)	5000	(5000; 5000)

$j = 1, 2, \dots, 6$ , onde a amplitude dos intervalos de credibilidade foram maiores que os demais casos. Isso ocorre devido ao fato da probabilidade de captura ser pequena resultando numa estatística  $r$  pequena, ou seja, há falta de dados. Os estudos de simulação evidenciaram que quando a estatística  $r$  é maior que a metade do tamanho populacional  $N$  ( $r \geq \frac{N}{2}$ ), as estimativas *a posteriori* dos parâmetros são satisfatórias.

Com relação ao modelo II observa-se que as estimativas *a posteriori* dos parâmetros  $N$  foram superestimadas nos casos de probabilidade de captura pequena, isto é, onde a estatística  $r$  é pequena e a amplitude dos intervalos de credibilidade foram maiores que os demais casos.

Considerando o modelo III as estimativas *a posteriori* para  $N$  são satisfatórias quando  $p_j = p \geq 0,15$ ,  $j = 1, 2, \dots, 6$ . Pelos resultados apresentados na Tabela 2 podemos concluir que os Modelos I e II são, no mínimo, equivalentes ao Modelo III. Existem intervalos de credibilidade que não contém o verdadeiro valor de  $N$  nos resultados referentes ao modelo III. Ou seja, a reparametrização não é mais eficiente do que a utilização pura e simples do parâmetro.

A Figura (1) mostra graficamente o comportamento das estimativas do tamanho populacional  $N$  segundo cada modelo estudado. Observe que, para valores de  $p > 0,6$ , os comportamentos das estimativas de  $N$  em ambos os modelos são equivalentes.



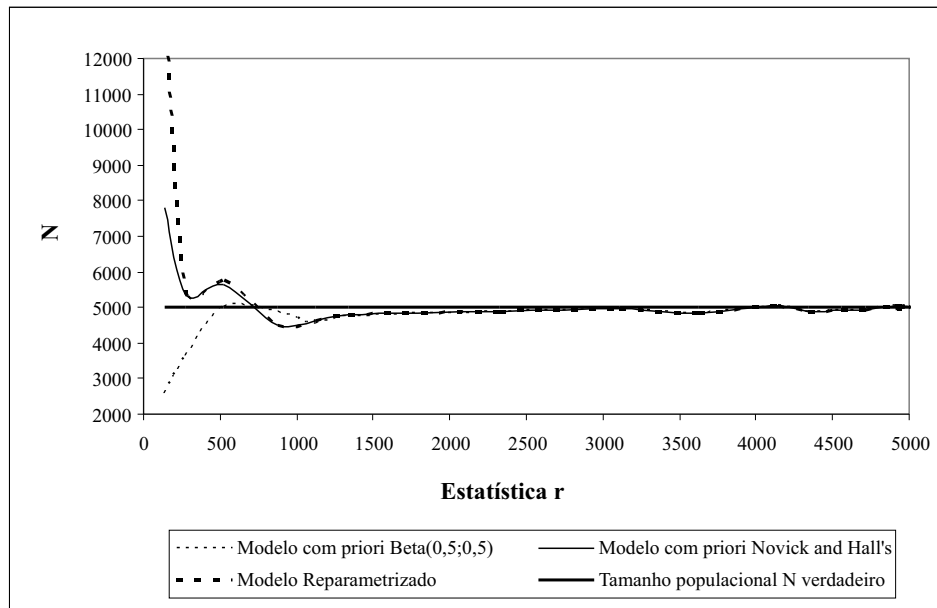


Figura 1 - Comportamento das estimativas de  $N$  segundo os modelos I, II e III.

### 3.2 Probabilidades de captura diferentes

Para completar o estudo sobre os modelos, comparamos os mesmos considerando probabilidades de captura diferentes para cada época por meio de diferentes cenários, conforme mostra a Tabela 3.

Tabela 3 - Diferentes cenários

$N = 5000, s = 6.$						
	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$
Cenário 1	0,05	0,10	0,15	0,20	0,25	0,30
Cenário 2	0,30	0,25	0,20	0,15	0,10	0,05
Cenário 3	0,05	0,10	0,15	0,15	0,10	0,05
Cenário 4	0,25	0,10	0,02	0,02	0,10	0,25
Cenário 5	0,01	0,02	0,03	0,04	0,05	0,06
Cenário 6	0,06	0,05	0,04	0,03	0,02	0,01
Cenário 7	0,35	0,12	0,03	0,03	0,12	0,35
Cenário 8	0,01	0,08	0,22	0,22	0,08	0,01

**Cenário 1:** As probabilidades de captura aumentam a cada época, produzindo uma estatística  $r$  grande.

**Cenário 2:** As probabilidades de captura diminuem a cada época, produzindo uma estatística  $r$  grande.

**Cenário 3:** As probabilidades de captura aumentam até a terceira época e em seguida diminuem até a última, produzindo uma estatística  $r$  moderada.

**Cenário 4:** As probabilidades de captura diminuem até a terceira época e em seguida aumentam até a última, produzindo uma estatística  $r$  moderada.

**Cenário 5:** As probabilidades de captura aumentam a cada época, produzindo uma estatística  $r$  pequena.

**Cenário 6:** As probabilidades de captura diminuem a cada época, produzindo uma estatística  $r$  pequena.

**Cenário 7:** As probabilidades de captura diminuem até a terceira época e em seguida aumentam até a última, produzindo uma estatística  $r$  grande.

**Cenário 8:** As probabilidades de captura aumentam até a terceira época e em seguida diminuem até a última, produzindo uma estatística  $r$  moderada.

A Tabela 4 mostra as estimativas *a posteriori* do tamanho populacional  $N$  considerando cada Modelo para cada cenário.

Tabela 4 - Estimativas *a posteriori* do tamanho populacional  $N$

	$r$	Modelo I		Modelo II		Modelo III	
		$E(N D)$	$I.C.(95\%)$ de $N$	$E(N D)$	$I.C.(95\%)$ de $N$	$E(N D)$	$I.C.(95\%)$ de $N$
Cenário 1	3444	5006	(4544; 5009)	4896	(4802; 5034)	4867	(4731; 5000)
Cenário 2	3422	4936	(4765; 5011)	4987	(4837; 5018)	4866	(4735; 5009)
Cenário 3	2358	4904	(4605; 5173)	4867	(4721; 5145)	4887	(4610; 5181)
Cenário 4	2774	4979	(4504; 5014)	5054	(4669; 5017)	4799	(4596; 5028)
Cenário 5	949	5063	(4127; 5872)	4898	(4326; 6211)	5188	(4205; 6366)
Cenário 6	938	5028	(3542; 5170)	4943	(3804; 5392)	4882	(3702; 5416)
Cenário 7	3395	4974	(4675; 5139)	4954	(4543; 5061)	4805	(4667; 5049)
Cenário 8	2523	5018	(4927; 5522)	5029	(4945; 5550)	5241	(4950; 5575)

Observe que os Modelos I, II e III começam a produzir resultados similares à medida que a estatística  $r$  aumenta. Novamente os estudos de simulação evidenciaram que quando a estatística  $r$  é maior que a metade do tamanho populacional  $N$  ( $r \geq \frac{N}{2}$ ), as estimativas *a posteriori* dos parâmetros são satisfatórias para ambos os modelos estudados. Neste caso podemos também concluir que os modelos são equivalentes.

## Conclusões

À luz dos resultados apresentados na Tabela 2, considerando probabilidades de captura iguais para as épocas, e os resultados apresentados na Tabela 4, considerando o caso de probabilidades de captura diferentes para cada época, a medida que a estatística  $r$  aumenta, não há diferenças entre os Modelos

considerados. Os estudos de simulação evidenciaram que quando a estatística  $r$  é maior que a metade do tamanho populacional  $N$  ( $r \geq \frac{N}{2}$ ), as estimativas *a posteriori* dos parâmetros são satisfatórias para ambos os modelos estudados. Desta forma a reparametrização da probabilidade de captura não se mostrou uma alternativa eficiente para o modelo de captura-recaptura.

PAULA, M.; DINIZ, C. A. R.; LEITE, J. G. Transformation to the capture-recapture probabilities from Bayesian model. *Rev. Bras. Biom.*, São Paulo, v.26, n.2, p.115-128, 2008.

- **ABSTRACT:** *The proposal this paper is to present a transformation to the capture-recapture probability from capture-recapture process with objective that to verify possible the ins and outs in respect to bayesian model considered by Castledine (1981). For this on considered two different non informative priors for the capture probabilities from bayesian model considered by Castledine in order to compare the them posterior estimates with the posterior estimates that model where the capture probabilities was transformed. We present a simulation study considering different capture probabilities with objective to found this comparison. The population size posteriori bayesian estimates was made through several studies of stochastic simulation through MCMC (Monte Carlo Markov Chain).*
- **KEYWORDS:** *Capture-recapture process; closed population; bayesian analysis; stochastic simulation.*

## Referências

- ABENI, D. D.; BRANCATO, G.; PERUCCI, C. A. Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection. *Epidemiology*, Baltimore, v.5, p.410-414, 1994.
- ANANDA, M. M. A. Bayesian methods for mark-resighting surveys. *Commun. Statist. Theory Methods*, New York, v.26, p.685-697, 1997.
- BEST, N.; COWLES, M. K.; VINES, K. *CODA - Convergence diagnosis and output analysis software for Gibbs sampling output. Version 0.4*. Cambridge: Biostatistics Unit MRC, 1995. (Technical report).
- BISHOP, Y. M. M.; FIENBERG, S. E.; HOLLAND, P. W. *Discrete multivariate analysis: theory and practice*. Cambridge: MIT Press, 1975. 557p.
- BOX, G. E. P.; TIAO, G. C. *Bayesian inference in statistical analysis*. New York: John Wiley, 1973. 588p.
- CASELLA, G.; GEORGE, E. I. Explaining the Gibbs sampler. *Am. Stat.*, Washington, v.46, p.167-74, 1992.
- CASTLEDINE, B. A. Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika*, London, v.68, p.197-210, 1981.

- CHAPMAN, D. G. Some properties of the hypergeometric distribution with applications to zoological sample censuses. *Univ. Calif. Publ. Stat.*, Berkley, v.1 p.131-60, 1951.
- CHIB, S.; GREEMBERG, E. Understanding the metropolis-hastings algorithm. *Am. Stat.*, Washington, v.49, p.327-35, 1995.
- COMACK, D. G. Internal estimation for mark-recapture studies of closed populations. *Biometrics*, Washington, v.48, p.567-76, 1992.
- CORMACK, R. M. Loglinear models for capture-recapture. *Biometrics*, Washington, v.45, p.395-413, 1989.
- COULL, B. A.; AGRESTI, A. The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, v.55, Washington, p.294-301, 1999.
- ENGEN, S. *Stochastic abundance models*. New York: Chapman and Hall, 1978. 389p.
- GAMERMAN, D. *Silmulação estocástica via cadeias de Markov*. Rio de Janeiro: IME-UFRJ, 1996. 196p.
- GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, New York, v.85, p.398-409, 1990.
- GEMAN, S.; GEMAN, D. Stochastic relaxation Gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, New York, v.6, p.721-41, 1984.
- GELMAN, A.; RUBIN, D. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, Hayward, v.7, p.457-511, 1992a.
- GEORGE, E. I.; ROBERT, C. P. Capture-recapture estimation via Gibbs sampling. *Biometrika*, London, v.79, n.4, p.677-83, 1992.
- GERKING, S. D. Vital Statistics of the fish population of Gordy Lake, Indiana. *Trans. Am. Fisheries Soc.*, Bethesda, v.82, p.48-67, 1953.
- HASTINGS, W. K. Monte Carlo Sampling methods using Markov chains and their applications. *Biometrika*, London, v.57, p.97-109, 1970.
- HUNTER, A. J.; GRIFFITHS, H. J. Bayesian approach to estimation of insect population size. *Technometrics*, Washington, v.20, n.3, p.231-234, 1978.
- JEFFREYS, H. *Theory of probability*. London: Oxford University Press, 1961.
- JOLLY, G. M. Explicit estimates from capture-recapture data with both death and immigration - stochastic model. *Biometrika*, London, v.52, p.225-247, 1965.
- KING, R.; BROOKS, S. P. On the Bayesian analysis of population size. *Biometrika*, London, v.88, p.317-336, 2001.
- LAPLACE, P. S. Sur les naissances, les mariages et les morts. In: *Histoire de l'académie royale des sciences*. Paris, 1783. 693p.

- LEITE, J. G.; RODRIGUES, J.; MILAN, L. A. A Bayesian analysis for estimating the number of species in a population using nonhomogeneous Poisson process. *Stat. Probabil. Lett.*, Amsterdam, v.48, p.153-161, 2000.
- MADIGAN, D.; YORK, J. C. Bayesian methods for estimation of the size of a close population. *Biometrika*, London, v.84, p.19-31, 1997.
- METROPOLIS, N. et al. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, New York, v.21, p.1087-1091, 1953.
- MINGOTI, S. A. A stepwise Bayesian estimator for the total number of distinct species in finite populations: sampling by elements. *J. Appl. Stat.*, Abingdon, v.275, p.651-670, 2000.
- NOVICK, W. R.; HALL, W. J. A Bayesian indifference procedure. *J. Am. Stat. Assoc.*, New York, v.60, p.1104-1117. 1965.
- OTIS D. L. et al. Statistical inference from capture data on closed animal populations. *It Wildlife Monographs*, v.62, p.1-135, 1978.
- PAULA, M. *Um enfoque Bayesiano do modelo de captura-recaptura na presença de covariáveis.*, 2006. Dissertação (Mestrado em Estatística), Universidade Federal de São Carlos, São Carlos, 2006.
- PETERSEN, C. G. J. The yearly immigration of young plaice into Limfjord from the German sea, Rept. *Dan. Biol. Stn.*, v.6, p.1-48, 1896.
- POLLOCK, K. H. *The assumption of equal catchability of animals in tag-recapture experiments.* 1974. Ph.D. (Ph.D em Estatística), Cornell University, Ithaca, New York, 1974.
- POLLOCK, K. H. Modeling capture-recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: Past, present, and future. *J. Am. Stat. Assoc.*, New York, v.86, p.225-238, 1991.
- POLLOCK, K. H.; OTTO. Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics*, Washington, v.39, p.1035-1049, 1983.
- POLLOCK, K. H.; HINES, J. E.; NICHOLS, J. D. The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics*, Washington, v.40, p.329-340, 1984.
- RICKER, W. E. Handbook of Computations for Biological Statistics of Fish populations. *Bull. Fish. Board Can.*, Ottawa, v.119, p.1-300, 1958.
- RICKER, W. E. *Computation and interpretation of biological statistics of fish population.* Ottawa: Department of the Environment, Fisheries and Marine Service, 1975. 382p.
- RIVEST, L. P.; DAIGLE, G. Loglinear models for the robust design in mark-recapture experiments. *Biometrics*, Washington, v.60, p.100-107, 2004.

- SCHWARZ, C. J.; ARNASON, A.N. A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, Washington, v.52, p.860-73, 1996.
- SEBER, G. A. F. A review of estimating animal abundance. *Biometrics*, Washington, v.42, p.267-92, 1986.
- SEBER, G. A. F. A review of estimating animal abundance II. *Int. Stat. Rev.*, Sidney, v.602, p.129-166, 1992.
- SMITH, P. J. Bayesian methods for multiple capture-recapture surveys. *Biometrics*, Washington, v.44, p.1177-189, 1988.
- SMITH, P. J. Bayesian analysis for a multiple capture-recapture model. *Biometrika*, v.78, London, p.399-408, 1991.
- WANG, X. *Bayesian analysis of capture-recapture models*. 2002. Ph.D. (Doutorado em Estatística). University of Missouri, Columbia. 2002.
- WHITE, G. C. et al. *Capture-recapture and removal methods for sampling closed populations*. Los Alamos: National Laboratory, New Mexico, 1982.
- YANG, R.; BERGER, J. O. Estimation of a covariance matrix using the reference prior. *Ann. Stat.*, Hayward, v.22, n.3, p.1195-1211. 1994.
- YANG, R.; CHEN, M. H. Bayesian analysis for random coefficient regression models using noninformative priors. *J. Multiv. Anal.*, v.55, n.2, p.283-311. 1995.
- YANG, R.; PYNE, D. *Bayesian analysis with mixed model in unbalanced case*. Ph.D. Dissertation. Purdue University.
- YOSHIDA, O. S. *Inferência Bayesiana do número de espécies de uma população*. 1996. Dissertação (Mestrado em Estatística), Instituto de Matemática e estatística, Universidade de São Paulo, São Paulo, 1996.
- YOSHIDA, O. S.; LEITE, J. G.; BOLFARINE, H. Stochastic monotonicity properties of Bayes estimation of the population size for capture-recapture data. *Stat. Probabil. Lett.*, Amsterdam, v.42, p.257-66, 1999.
- ZACHARIAS, H. P. *Aplicação do algoritmo Gibbs sampling no processo de captura-recaptura*. 2000. Dissertação (Mestrado em Estatística), Universidade Federal de São Carlos, São Carlos, 2000.

Received in 06.11.2007.

Approved after revised in 13.07.2007.