

# Aplicação do Amostrador de Gibbs à metodologia de captura-recaptura

José H C Monteiro da Silva

**RA 117439**

MI602 - 2s2020

## Introdução

A metodologia de captura-recaptura é uma técnica utilizada inicialmente por Laplace em 1786 para estimar o volume populacional da França e que é amplamente utilizada pela área da Ecologia na estimação de populações de animais (George and Robert 1992; Paula et al. 2009) e também se aplica em estudos epidemiológicos (Chao et al. 2001). Na ecologia, o princípio básico do método consiste na captura e marcação de animais seguida da soltura em múltiplas amostras de armadilhas ao longo do tempo (Paula et al. 2009).

Castledine (1981) revisitou a teoria da metodologia de captura-recaptura de modo a acomodar informações *a priori* sobre as populações e probabilidades de recaptura a partir da construção de estimativas Bayesianas. Tais estimativas foram computadas anos depois utilizando o Amostrador de Gibbs por George and Robert (1992).

O presente exercício se concentra na aplicação de métodos de *Monte Carlo Markov Chain*, no caso o Amostrador de Gibbs, para gerar as estimativas Bayesianas formuladas por Castledine (1981) e George and Robert (1992) para um conjunto de dados de captura-recaptura da

espécie *Panthera onca* na região do Parque Nacional da Serra da Capivara no Estado do Piauí. Esse conjunto de dados foi utilizado para gerar estimativas da população da onça na região por Silveira et al. (2009) utilizando *softwares* específicos e métodos clássicos da área da Ecologia. Dessa maneira, o objetivo do trabalho é gerar estimativas e intervalos de credibilidade para o tamanho populacional e para as probabilidades de captura dado o conjunto de dados amostrados em Silveira et al. (2009), de modo a comparar os resultados e testar a aplicabilidade do método em situações reais de biomas brasileiros.

Esse documento apresenta 3 seções além desta breve introdução. A segunda seção apresenta os dados e os métodos de construção de estimativas. A terceira seção apresenta os resultados. E a quarta seção apresenta uma breve discussão dos resultados obtidos comparando-os com os observados em Silveira et al. (2009).

## Dados e Métodos

### Dados

Os dados para aplicação dos métodos de MCMC para construção de estimativas Bayesianas para o modelo de captura-recaptura foram retirados de Silveira et al. (2009). Esse estudo, baseado na região do Parque Nacional Serra da Capivara do Piauí foi conduzido utilizando de armadilhas por câmeras para identificação e amostragem da população de onças pintadas numa área de  $524 \text{ km}^2$  do parque. A Tabela 1 apresenta os números de indivíduos capturados e recapturados em cada ocasião (armadilha).

Tabela 1: Dados da metodologia de captura-recaptura de onças pintadas coletados por Silveira et al., (2009) na região do Parque Nacional da Serra da Capivara, Piauí.

Ocasão/Armadilha	Capturas (ni)	Recapturas (mi)
1	2	0
2	4	0
3	2	2
4	4	4
5	4	4
6	2	2
7	3	3
8	4	2
9	2	2
10	3	3
11	5	5
12	10	6
13	2	2
14	5	5
Total	52	42

## Base Teórica da Estimativa Bayesiana e Pressupostos para a Metodologia de Captura-Recaptura

Neste trabalho, serão adotados os mesmos pressupostos de George and Robert (1992), com base no trabalho de Castledine (1981). Assim, seja  $N$  a população desconhecida de onças pintadas que se deseja estimar,  $I$  (igual a 14, no caso) o número de ocasiões (armadilhas) de captura e  $p_{ij}$  a probabilidade de captura do indivíduo  $j$  na armadilha  $i$ . Consideram-se ainda os seguintes pressupostos (Castledine 1981; Silveira et al. 2009):

- a população amostrada é constante ao longo do experimento e fechada (ou seja, sem eventos de nascimentos, mortes ou migração dentro da área do estudo);
- inicialmente, nenhum animal está marcado como já capturado;
- em cada uma das 14 armadilhas, todos os animais tem a mesma probabilidade de serem capturados, independente do seu histórico de capturas, ou seja,  $p_{ij} = p_i$ .

50 Define-se ainda  $r$  a partir do número de capturas e recapturas:

$$(1) \quad r = \sum_{i=1}^{14} n_i - \sum_{i=1}^{14} m_i = 52 - 40 = 12$$

51 assim, para os pressupostos adotados, a função de verossimilhança  $L(N, p|dados)$  do  
 52 experimento é dada na equação (2), com suporte nos Inteiros não negativos para  $N$  e nos  
 53 Reais no intervalo  $[0,1]$  para  $p_i$ , seguindo George and Robert (1992):

$$(2) \quad L(N, p|dados) \propto \frac{N!}{(N-r)!} \prod_{i=1}^{14} p_i^{n_i} (1-p_i)^{N-n_i}.$$

54 Assumem-se as *prioris* para  $N$  e  $p = \{p_1, p_2, \dots, p_{14}\}$  como sendo  $N \sim Poisson(\lambda)$  com  
 55  $\lambda = 50$  e  $p_i \sim Uniforme(0, 1)$ , respectivamente.

56 Assim, a distribuição *posteriori* é dada por

$$(3) \quad p_i(N, p|dados) \propto L(N, p|dados) * \pi(N) * \prod_{i=1}^{14} \pi(p_i),$$

57 que assume distribuições independentes para as *prioris* e pode ser obtida via Amostrador  
 58 de Gibbs encontrando-se as distribuições condicionais completas para  $f(N|p, dados)$  e  
 59  $f(p_i|N, p_{k:k \neq i}, dados)$ . Aplicando as *prioris* definidas anteriormente na *posteriori*, as  
 60 condicionais assumem seguintes distribuições (Robert and Casella 2004, p. 348):

$$(4) \quad (N-r)|p, dados \sim Poisson(\lambda \prod_{i=1}^{14} (1-p_i)),$$

61 e

$$(5) \quad p_i|N, p_{k:k \neq i}, dados \sim Beta(n_i + 1, N - n_i + 1).$$

## Estimativas via Amostrador de Gibbs

O Amostrador de Gibbs é uma técnica que permite a obtenção de características (média e variância, por exemplo) de uma determinada função de densidade marginal  $f(x)$  dada uma densidade conjunta  $f(x, y_1, y_2, \dots, y_n)$  a partir da integração  $f(x) = \int \dots \int f(x, y_1, \dots, y_n) dy_1 \dots dy_n$  (Casella and George 1992). A técnica de amostragem de Gibbs é utilizada para encontrar as características de  $f(x)$  em casos onde a integração apresentada é de difícil computo.

Assim, pode-se obter uma amostra  $X_1, X_2, \dots, X_m \sim f(x)$  partindo de um tamanho de amostra  $m$  suficientemente grande e de distribuições condicionais  $f(x|y)$  e  $f(y|x)$  conhecidas (Casella and George 1992). A partir das distribuições condicionais, obtém-se uma sequência de Gibbs  $X_0, Y_0, X_1, Y_1, \dots$  calculada iterativamente a partir de valores  $X_0$  e  $Y_0$  especificados.

Seguindo a lógica de George and Robert (1992), pode-se implementar um amostrador de Gibbs para obter as distribuições condicionais para  $N$  e  $p$  do modelo de captura-recaptura  $f(N|p, \text{dados})$  e  $f(p_i|N, p_{k:k \neq i}, \text{dados})$ . Para tal, o seguinte algoritmo é adotado:

1. Gerar estimativas no tempo 0 para  $N$  e  $p$  a partir das *prioris* definidas:  $\hat{N}^0 \sim \text{Poisson}(\lambda)$  e  $\hat{p}_i^0 \sim \text{Uniforme}(0, 1)$ ;
2. Simular  $\hat{N}^t$  a partir da equação (4) e  $\hat{p}_i^{t-1}$ ;
3. Simular  $\hat{p}^t$  a partir da equação (5) e  $\hat{N}^t$ .

Realizando os três passos do algoritmo acima por  $t$  vezes suficientemente grandes, tem-se uma corrente de Markov gerada para as variáveis de interesse que convergem para as distriuições de interesse (Casella and George 1992). Casella and George (1992) sugerem que o monitoramento da convergência e a partir do momento em que a série gerada tiver convergido, fazer a seleção a partir de intervalos de  $k$  passos de modo a evitar problemas de autocorrelação.

## Resultados

### Amostrador de Gibbs

Foram simuladas 100000 realizações do algoritmo apresentado na seção anterior. A Figura 1 apresenta os resultados das 150 primeiras simulações utilizadas para investigar a convergência das séries geradas. Pelo resultado, fica evidente que a partir de  $t = 100$  a convergência das estimativas já é observada para  $N$ .

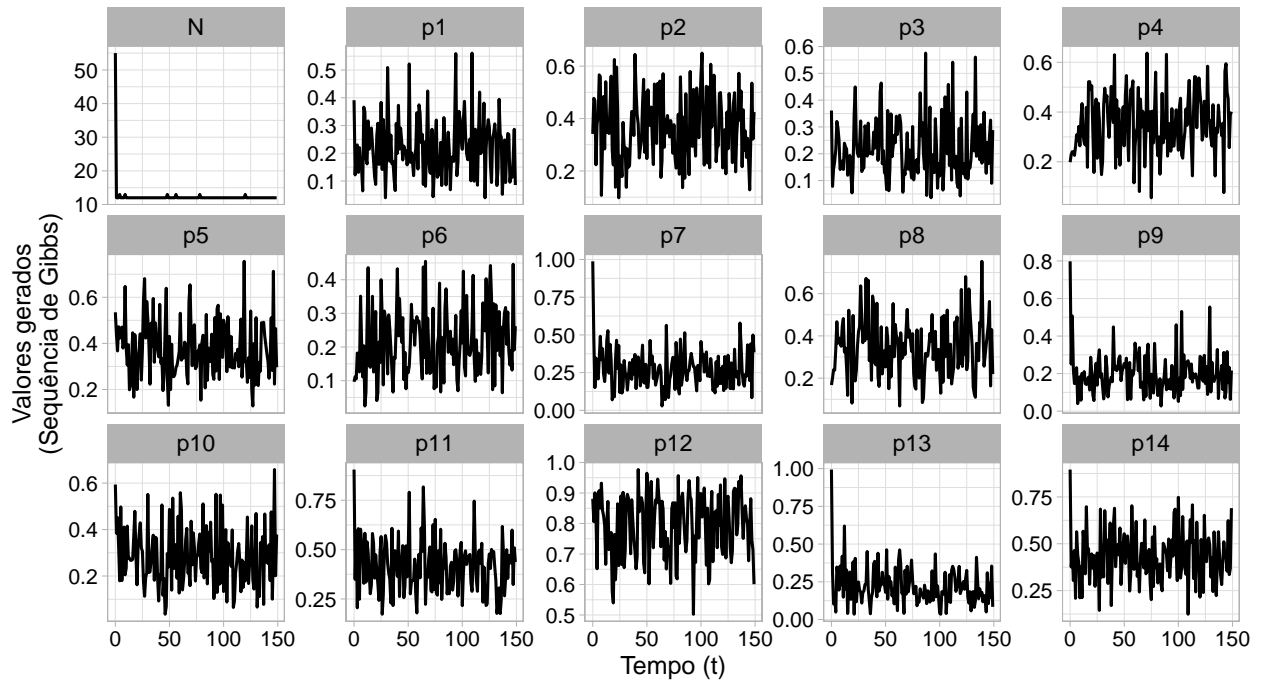


Figura 1: Valores simulados para as estimativas de  $N$  e  $p_i$  para os tempos  $t$  entre 0 e 149

Após verificação do critério da convergência, olha-se a autocorrelação da sequência de Gibbs gerada. Pela Figura 2, percebe-se que para valores relativamente baixos de  $k$  (3 e 5, por exemplo) a autocorrelação entre os valores da série é estatisticamente não significativa a 95% de confiança. Assim, pode-se selecionar a partir do elemento  $t = 100$  valores espaçados a cada 10 passos ( $t = \{100, 110, 120, \dots, 100000\}$ ) de modo a obter uma sequência final com estimativas dos parâmetros  $N$  e  $p$ .

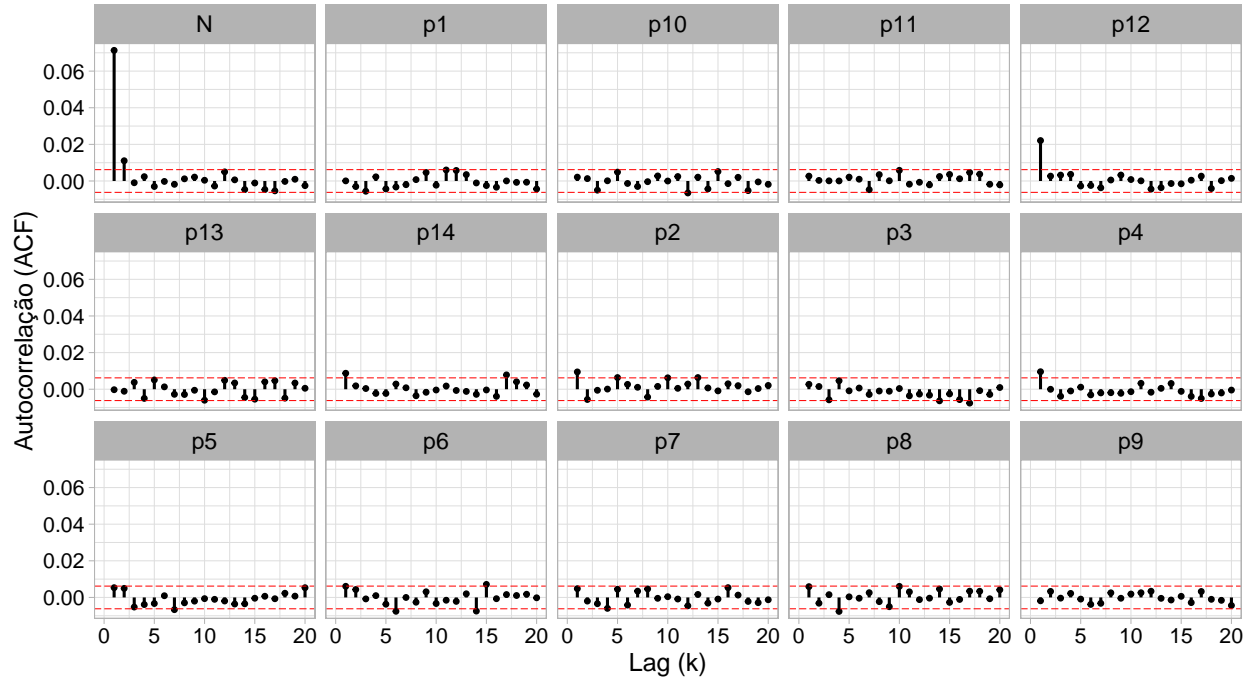


Figura 2: Resultados para estimativas de autocorrelação das sequências de Gibbs geradas para N e  $\pi$ . As linha vermelhas indicam o intervalo de confiança de 0.95 para a ausência de correlação para um determinado lag k.

97 Na Tabela 2, apresentam-se algumas estatísticas para as distribuições geradas. Vê-se que a  
 98 estimativa da população de onça pintada na região do estudo encontra-se entre 12 e 13 com  
 99 95% de credibilidade. Silveira et al. (2009) estimou uma população de 12 onças, valor dentro  
 100 da estimativa observada via Amostrador de Gibbs.

Tabela 2: Estatísticas a posteriori da estimativa Bayesiana via Amostrador de Gibbs para metodologia de captura-recaptura de onças pintadas coletados por Silveira et al., (2009) na região do Parque Nacional da Serra da Capivara, Piauí.

Variável	Percentil 0.025	Percentil 0.975	Esperança	Mediana
N	12.000	13.000	12.101	12.000
p1	0.049	0.447	0.213	0.199
p2	0.143	0.611	0.357	0.350
p3	0.052	0.453	0.212	0.197
p4	0.139	0.611	0.354	0.345
p5	0.137	0.611	0.352	0.345
p6	0.048	0.447	0.211	0.196
p7	0.089	0.534	0.282	0.273
p8	0.140	0.612	0.355	0.347
p9	0.052	0.452	0.213	0.197
p10	0.093	0.530	0.286	0.276
p11	0.195	0.677	0.424	0.421
p12	0.543	0.946	0.780	0.793
p13	0.050	0.456	0.213	0.199
p14	0.190	0.677	0.425	0.423

## Discussão

Este trabalho apresentou a aplicação do Amostrador de Gibbs para estimar a população de onças pintadas no Parque Nacional da Serra da Capivara a partir da metodologia de captura-recaptura (Silveira et al. 2009) utilizando da teoria Bayesiana (Castledine 1981; George and Robert 1992).

Silveira et al. (2009) obteve um resultado de volume populacional de onças na região igual a 12, enquanto que os resultados do trabalho apontam para uma população entre 12 e 13 indivíduos com 95% de credibilidade. Portanto, os resultados aqui observados para as *prioris* escolhidas e pressupostos adotados foram similares ao resultado obtido pelos modelos clássicos da área apresentados em Silveira et al. (2009).



## Referências Bibliográficas

- Casella, G., and George, E. I. (1992), “Explaining the gibbs samples,” *The American Statistician*, American Statistical Association, 46, 167–174.
- Castledine, B. J. (1981), “A bayesian analysis of multiple-recapture sampling for a closed population,” *Biometrika*, Oxford University Press, 67, 197–210.
- Chao, A., Tsay, P. K., Lin, S.-H., Shau, W.-Y., and CHao, D.-Y. (2001), “The applications of capture-recapture models to epidemiological data,” *Statistics in Medicine*, John Wiley & Sons, 20, 3123–3157.
- George, E. I., and Robert, C. P. (1992), “Capture-recapture estimation via gibbs sampling,” *Biometrika*, Oxford University Press, 79, 677–683.
- Paula, M., Diniz, C A R, and Leite, J. G. (2009), “Modelo bayesiano hierárquico de captura-recaptura com distribuição poisson-gama,” *Trends in Applied and Computational Mathematics*, Sociedade Brasileira de Matemática Aplicada e Computacional, 10, 191–202.
- Robert, C. P., and Casella, G. (2004), *Monte carlo statistical methods*, Springer.
- Silveira, L., Jácomo, A. T., Astete, S., Sollmann, R., Tôrres, N. M., Furtado, M. M., and Marinho-Filho, J. and (2009), “Density of the near threatened jaguar panthera onca in the caatinga of north-eastern brazil,” *Oryx*, Cambridge University Press, 44, 104–109. <https://doi.org/10.1017/S0030605309990433>.

## Agradecimentos

Agradeço à colega Marília Gabriela Rocha pelo suporte durante a realização dessa atividade e, principalmente, pela paciência com as dúvidas elementares.

## <sup>132</sup> **Material para Replicação**

<sup>133</sup> Os códigos utilizados no presente trabalho encontram-se disponíveis no repositório do **github**:

<sup>134</sup> [https://github.com/josehcms/MI602\\_2s2020](https://github.com/josehcms/MI602_2s2020).