

Estimação Bayesiana da População de Onças

Autor: Lucas Perondi Kist (RA: 236202)

1 Introdução

A estimação da população de animais selvagens é importante para monitorar o risco de extinção de uma espécie, bem como melhorar a compreensão acerca da dinâmica da cadeia alimentar local. Nesse sentido, existem diversas formas de fazer isso, sendo a metodologia de captura-marcação-recaptura e utilização do estimador de Chapman (1951) uma delas.

Nesse sentido, é possível, ainda, utilizar uma abordagem Bayesiana e considerar conhecimentos *a priori* dos parâmetros de interesse, tal qual apresentado em George (1992). Dessa forma, o objetivo deste trabalho é realizar a estimação Bayesiana do tamanho da população de onças (*Panthera onca*) adultas no Parque Nacional da Serra da Capivara, em 2007, utilizando os dados de Silveira et al. (2010), o estimador de Chapman (1951) e o Amostrador de Gibbs.

2 Materiais e Métodos

A seguir, são apresentados em detalhes os dados utilizados para a obtenção dos resultados, além de serem desenvolvidos os cálculos realizados para a obtenção das distribuições condicionais completas, necessárias para a implementação do Amostrador de Gibbs. Ademais, são apresentados alguns detalhes adicionais sobre os estimadores e métodos aplicados.

2.1 Banco de dados

Silveira et al. (2010) realizou um experimento utilizando catorze sessões de armadilhas, cada uma com de seis dias de duração, entre agosto e outubro de 2007 para realizar essa estimação. Os resultados obtidos por ele estão apresentados na Tabela 1, a qual indica que foram observadas doze onças distintas (o indivíduo 4 era um filhote e, por isso, foi removido). Esses dados foram usados na realização deste trabalho para comparar resultados das estimativas obtidas com as de Silveira et al. (2010), o qual reportou um intervalo de confiança (provavelmente de 95%, mas não é especificado no trabalho) para o tamanho populacional como sendo [13,33].

Tabela 1: Resultado das capturas utilizando 14 armadilhas. Foram identificadas doze onças distintas (o indivíduo 4 era um filhote e foi removido). Uma célula igual a 1 indica que esse indivíduo foi capturado por essa armadilha. Dados coletados por Silveira et al. (2010).

Indivíduo	Armadilha													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	0	0	0	0	0	1	0	0	0	1	1	1
2	1	0	0	1	1	0	1	1	0	1	1	1	0	1
3	0	1	1	0	0	0	0	0	0	1	1	1	0	1
5	1	0	0	1	1	0	1	0	1	0	1	0	1	1
6	0	1	1	1	1	1	1	0	1	1	1	1	0	0
7	0	0	0	0	0	0	0	0	0	0	0	1	0	0
8	0	1	0	1	0	0	0	0	0	0	1	1	0	0
9	0	1	0	0	1	1	0	1	0	0	0	1	0	1
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	1	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0
13	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Total	2	4	2	4	4	2	3	4	2	3	5	10	2	5

2.2 Modelo Bayesiano de Captura-Recaptura

Defina $n_i, i = 1, 2, \dots, 14$, o número de animais capturados na armadilha i e seja m_j o número de animais capturados na j -ésima armadilha e que já haviam sido capturados anteriormente, $j = 2, 3, \dots, 14$. Assim, o número de animais distintos capturados em algum momento r é dado por

$$r = \sum_{i=1}^{14} n_i - \sum_{j=2}^{14} m_j.$$

Para determinar o tamanho populacional N , assume-se que os ensaios das capturas são independentes entre si, que as capturas são independentes entre si e que a probabilidade de captura em uma determinada sessão de armadilhas é a mesma para todos os animais. Então, segundo apresentado por Salasara et al. (2016), a verossimilhança é dada por

$$L(N, p_1, p_2, \dots, p_{14} | n_1, \dots, n_{14}, m_2, \dots, m_{14}) \propto \frac{N!}{(N-r)!} \mathbf{1}(N)_{\{0,1,2,3,\dots\}} \prod_{i=1}^{14} [p_i^{n_i} (1-p_i)^{N-n_i} \mathbf{1}(p_i)_{[0,1]}].$$

Foram assumidas as distribuições *a priori* $\pi(N|\cdot) \sim \text{Poisson}(\lambda)$ e $\pi(p_i|\cdot) \sim \text{Beta}(a, b)$ independentes entre si, com λ, a e b conhecidos. Note que o tamanho da população N deve ser pelo menos o número de onças distintas capturadas r , então a distribuição *a priori* de N foi truncada, de forma a iniciar em r . Além disso, sendo $\mathbf{x} = (n_1, \dots, n_{14}, m_2, \dots, m_{14})$ segue que a distribuição conjunta *a posteriori* dos parâmetros de interesse é

$$\pi(N, p_1, \dots, p_{14} | \mathbf{x}) \propto \frac{N! e^{-\lambda} \lambda^N}{(N-r)! N!} \mathbf{1}(N)_{\{0,1,\dots\}} \mathbf{1}(N)_{\{r,r+1,\dots\}} \prod_{i=1}^{14} [p_i^{n_i} (1-p_i)^{N-n_i} p_i^{a-1} (1-p_i)^{b-1} \mathbf{1}(p_i)_{[0,1]}],$$

de onde segue que

$$\pi(N, p_1, \dots, p_{14} | \mathbf{x}) \propto \frac{e^{-\lambda} \lambda^N}{(N-r)!} \mathbf{1}(N)_{\{r,r+1,r+2,r+3,\dots\}} \prod_{i=1}^{14} [p_i^{a+n_i-1} (1-p_i)^{b+N-n_i-1} \mathbf{1}(p_i)_{[0,1]}].$$

Assim, as distribuições condicionais completas são dadas por:

$$\pi(N | p_1, \dots, p_{14}, \mathbf{x}) \propto \frac{\lambda^N \prod_{i=1}^{14} (1-p_i)^N}{(N-r)!} \mathbf{1}(N)_{\{r,r+1,r+2,r+3,\dots\}} = \frac{[\lambda \prod_{i=1}^{14} (1-p_i)]^N}{(N-r)!} \mathbf{1}(N-r)_{\{0,1,2,3,\dots\}}$$

$$\pi(p_i | N, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{14}, \mathbf{x}) \propto p_i^{a+n_i-1} (1-p_i)^{b+N-n_i-1} \mathbf{1}(p_i)_{[0,1]} \quad i = 1, 2, \dots, 14.$$

Ou seja,

$$[(N-r) | (p_1, \dots, p_{14}, \mathbf{x})] \sim \text{Poisson} \left(\lambda \prod_{i=1}^{14} (1-p_i) \right)$$

$$[p_i | (N, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{14}, \mathbf{x})] \sim \text{Beta}(a + n_i, b + N - n_i), i = 1, 2, \dots, 14.$$

Em particular, quando $\pi(p_i|\cdot) \sim U(0, 1)$, isto é, $a = b = 1$,

$$[(N-r) | (p_1, \dots, p_{14}, \mathbf{x})] \sim \text{Poisson} \left(\lambda \prod_{i=1}^{14} (1-p_i) \right)$$

$$[p_i | (N, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{14}, \mathbf{x})] \sim \text{Beta}(1 + n_i, 1 + N - n_i), i = 1, 2, \dots, 14.$$

O Amostrador de Gibbs para obter uma amostra de tamanho n da distribuição conjunta *a posteriori* dos parâmetros de interesse, neste caso, é dado por:

1. Escolher chutes iniciais para os parâmetros: $N_1 \sim \text{Poisson}(\lambda)$ e $p_{i1} \sim \text{Beta}(a, b)$;
2. Para $t = 2, 3, \dots, n$:
 - Para $i = 1, 2, \dots, 14$, amostre $P_i^* \sim \text{Beta}(a + n_i, b + N_{t-1} - n_i)$ e faça $p_{it} = P_i^*$;
 - Amostre $N^* \sim \text{Poisson}(\lambda \prod_{i=1}^{14} (1 - p_{it}))$ e faça $N_t = r + N^*$.

A implementação foi realizada da seguinte forma:

```
gibbs_cap_recap <- function(n, a, b, lambda, r, vetor_ns){
  aa <- matrix(nrow = n, ncol = 15) #amostra da posteriori
  colnames(aa) <- c(paste0("p_", 1:14), "N")
  while(T){ N1 <- rpois(1, lambda); if(N1 >= r) break}
  aa[1,] <- c(rbeta(14, a, b), N1) #chutes iniciais
  for(linha in 2:n){ #vou amostrar p_1, p_2, ..., p_14, N
    pis <- purrr::map(1:14, function(i) rbeta(1, a + vetor_ns[i],
                                              b + aa[linha-1, 15] - vetor_ns[i]))
    pis <- purrr::list_c(pis)
    N_i <- r + rpois(1, lambda*prod(1-pis))
    aa[linha, ] <- c(pis, N_i)
  }; return(aa)}
library(purrr); v = c(2, 4, 2, 4, 4, 2, 3, 4, 2, 3, 5, 10, 2, 5)
```

3 Resultados

A seguir podem ser encontrados os resultados da aplicação dos métodos de estimação populacional, realizados utilizando $\lambda = 30$. Em relação aos dados obtidos, nota-se que $r = 12$ onças distintas capturadas. Além disso, a armadilha 12 foi a que mais capturou indivíduos distintos (10), o número de indivíduos capturados mais frequente foi 2 (5 vezes) e a média foi de 3,7 onças diferentes por armadilha. Para cada valor de (a, b) , foi fixada a semente indicada e os códigos foram rodados sequencialmente.

3.1 Prioris uniformes para as probabilidades de captura

Inicialmente, foi obtida uma amostra de tamanho 1000 utilizando três vetores de partidas diferentes para verificar se houve convergência da série e determinar o *burn-in* a ser utilizado. A Figura 1 ilustra esses resultados para o caso com menos capturas (p_1), mais (p_{12}), intermediário (p_{11}) e de N . A partir de sua análise, conclui-se que as cadeias estão misturadas em todos os casos e que foi atingida a estacionariedade. Assim, utilizou-se um *burn-in* de 1000.

```
set.seed(236202)
aas_unif <- map(1:3, function(x) gibbs_cap_recap(1000, 1, 1, 30, 12, v))
```

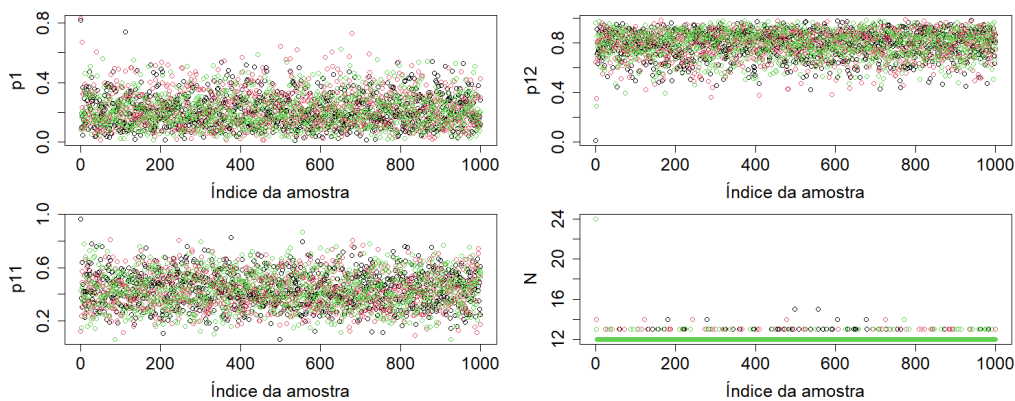


Figura 1: Gráficos sequenciais de algumas distribuições *a posteriori* selecionadas: p_1, p_{12}, p_{11} e N . Foi utilizada distribuição *a priori* $U(0, 1)$ para p_1, \dots, p_{14} .

Já a Figura 2 contém as autocorrelações (ACF) de uma amostra de tamanho 10000 (após remover as 1000 primeiras) para as mesmas distribuições *a posteriori*. Nota-se uma correlação significativa a um nível de 5% no *lag* 1 de N . Por isso, tomou-se $d = 2$ e foi verificado que essa correlação foi eliminada.

```
aa_unif <- gibbs_cap_recap(11000, 1, 1, 30, 12, v)[1001:11000,]
```

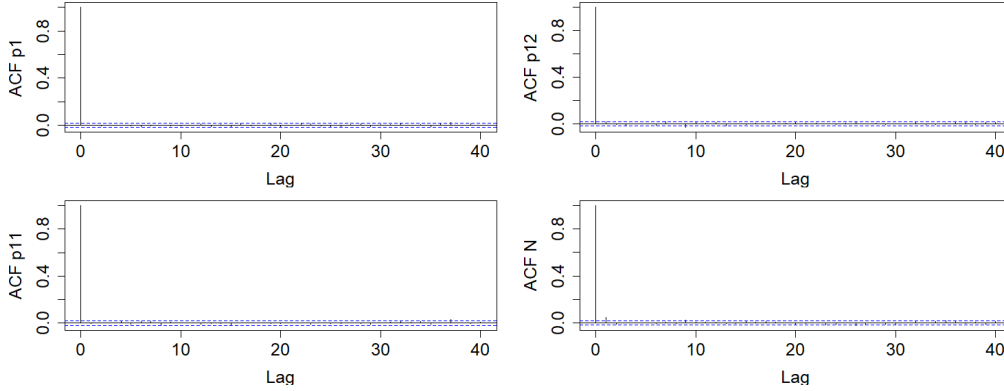


Figura 2: Gráficos de ACF de algumas distribuições *a posteriori* selecionadas: p_1, p_{12}, p_{11} e N . Foi utilizada distribuição *a priori* $U(0, 1)$ para p_1, \dots, p_{14} .

Por fim, foi obtida uma amostra aleatória de tamanho 5000 das distribuições *a posteriori* de todos os parâmetros de interesse utilizando um *burn-in* de 1000 e $d = 2$. Os histogramas de $p_1, p_{12}, p_{11}, p_{12}$ e N estão apresentadas na Figura 3, assim como a distribuição conjunta de N e p_1 e N e p_{12} . A partir dela percebe-se que a distribuição de N está bastante concentrada em 12, com a probabilidade de $N = 12$ igual a 94,4%. Em relação às probabilidades de captura, é visível a influência da *priori* utilizada, que deslocou a distribuição na direção de uma proporção 0,5. Já nas distribuições conjuntas, nota-se uma correlação negativa entre N e a probabilidade de captura.

```
aa_unif_f <- gibbs_cap_recap(11000, 1, 1, 30, 12, v)[seq(1001,11000,by=2),]
```

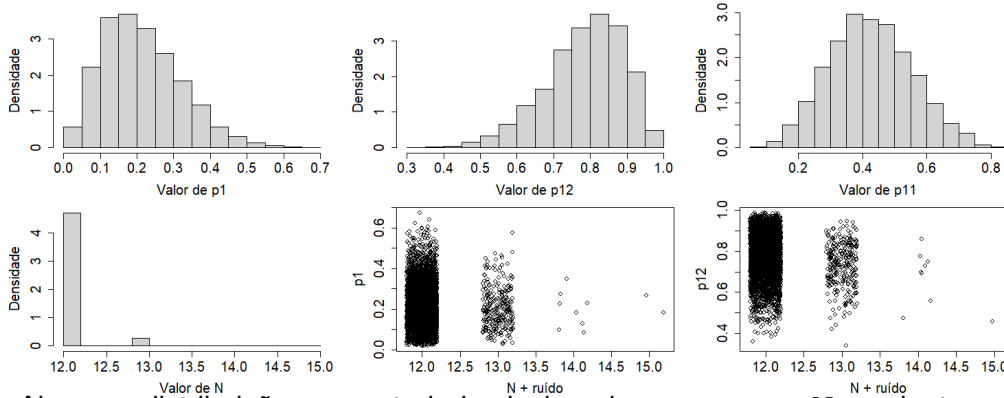


Figura 3: Algumas distribuições *a posteriori* selecionadas p_1, p_{12}, p_{11}, N , conjunta entre N e p_1 e entre N e p_{12} (nesses casos, foi adicionado um ruído aos valores de N para melhorar a visualização). Foi utilizada distribuição *a priori* $U(0, 1)$ para p_1, \dots, p_{14} .

3.2 *Prioris* $Beta(a, b)$ para as probabilidades de captura

Ao levar em consideração que as onças são animais ariscos, as distribuições *a priori* foram modificadas para que a probabilidade média de uma onça ser capturada seja de cerca de 7% e que seja pouco provável que ela seja maior do que 20%. Assim, foram utilizados $a = 1$ e $b = 13$, com média 0,071 e probabilidade de ultrapassar 20% igual a 0,055.

Analogamente ao caso anterior, foi obtida uma amostra de tamanho 1000 utilizando três vetores de valores iniciais distintos. Novamente, realizou-se um estudo de convergência para o caso com menos capturas (p_1), mais (p_{12}) e outro intermediário (p_{11}), além de N . A Figura 4 apresenta esses

gráficos sequenciais, de onde é possível observar que um *burn-in* de tamanho 1000 é suficiente e que foi atingida a estacionariedade

```
set.seed(236203)
aas_b_ab <- map(1:3, function(x) gibbs_cap_recap(1000, 1, 13, 30, 12, v))
```

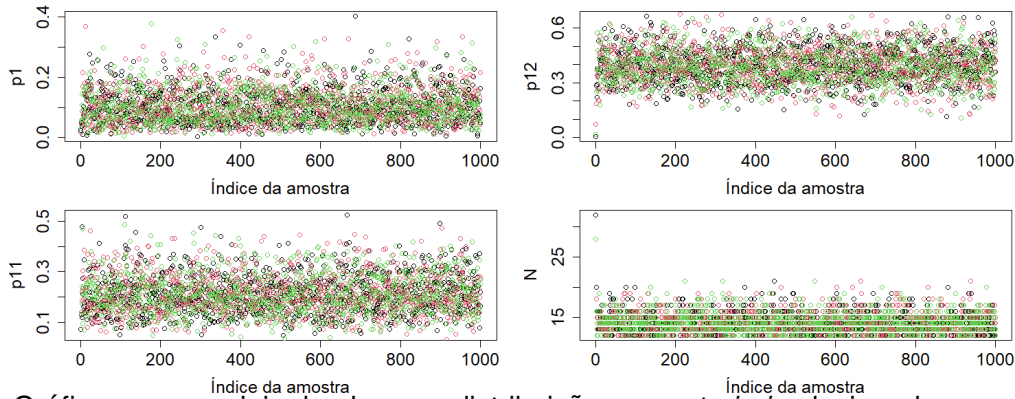


Figura 4: Gráficos sequenciais de algumas distribuições *a posteriori* selecionadas: p_1, p_{12}, p_{11} e N . Foi utilizada distribuição *a priori* $Beta(1, 13)$ para p_1, \dots, p_{14} .

Na sequência, foi obtida uma amostra de tamanho 10000 (após descartar as 1000 iniciais) para analisar o comportamento do ACF, o qual está ilustrado na Figura 5. Nota-se que há 3 *lags* significativos na distribuição de N . Após tomar $d = 3$, percebeu-se que todas as correlações passaram a ser não significativas a 5%.

```
aa_b_ab <- gibbs_cap_recap(11000, 1, 13, 30, 12, v)[1001:11000,]
```

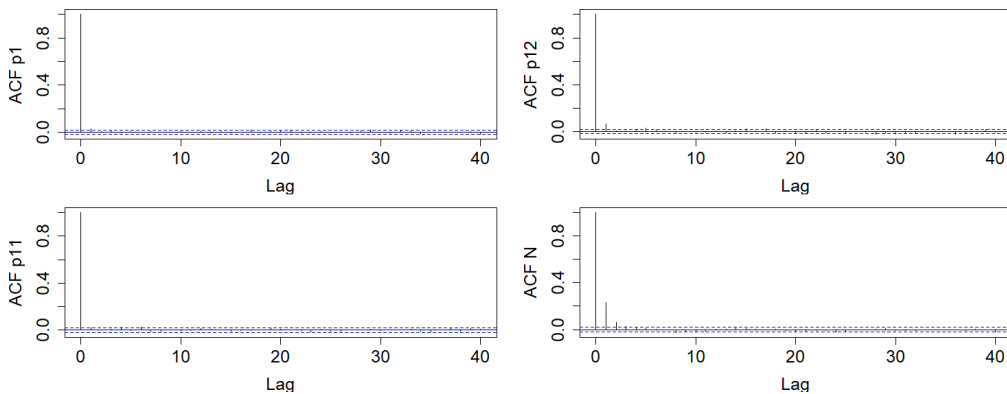


Figura 5: Gráficos de ACF de algumas distribuições *a posteriori* selecionadas: p_1, p_{12}, p_{11} e N . Foi utilizada distribuição *a priori* $Beta(1, 13)$ para p_1, \dots, p_{14} .

Definidos o *burn-in* e o valor de d , foi obtida uma amostra aleatória de tamanho 5000 da distribuição conjunta *a posteriori*. A Figura 6 apresenta histogramas das mesmas distribuições *a posteriori*, além da distribuição conjunta entre N e p_1 e N e p_{12} , com um ruído tendo sido adicionado aos valores de N para melhorar a visualização. Ao analisá-la, percebe-se que os histogramas das probabilidades de captura foram deslocados para valores menores, quando comparados com a Figura 3, devido à *priori* utilizada.

Além disso, N passou a assumir valores maiores, refletindo em um intervalo de credibilidade de 95,3% (para ser comparável com o anterior) sendo $[12, 17]$. Já em relação às distribuições conjuntas, novamente pode ser vista uma correlação negativa entre a probabilidade de captura e o tamanho da população.

```
aa_b_ab_f <- gibbs_cap_recap(16000, 1, 13, 30, 12, v)[seq(1001, 16000, by=3),]
```

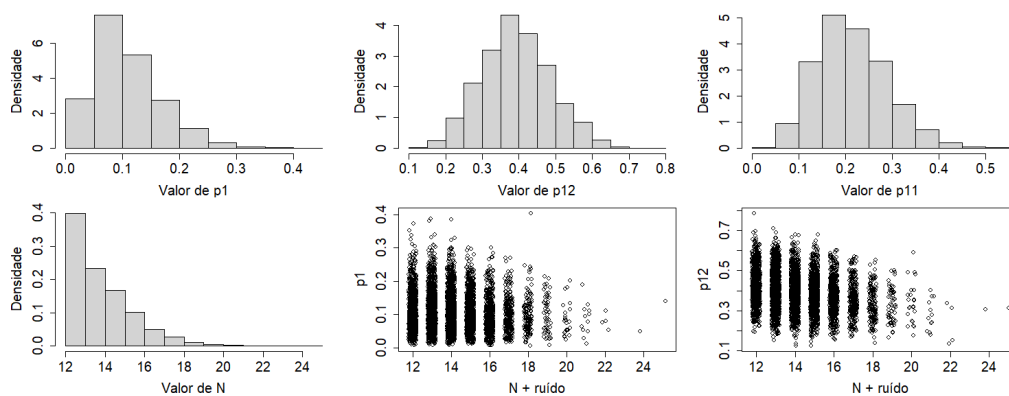


Figura 6: Algumas distribuições *a posteriori* selecionadas p_1, p_{12}, p_{11}, N , conjunta entre N e p_1 e entre N e p_{12} (nesses casos, foi adicionado um ruído aos valores de N para melhorar a visualização. Foi utilizada distribuição *a priori* $Beta(1, 13)$ para p_1, \dots, p_{14} .

4 Conclusão

Com base no exposto, conclui-se que é possível utilizar métodos Bayesianos para estimar uma população de onças a partir de uma metodologia de captura-recaptura e do estimador de Chapman (1951). Além disso, foi possível verificar o impacto da escolha dos hiperparâmetros a e b nas distribuições *a posteriori* dos parâmetros. Em particular, considerar que onças são animais ariscos (aumentar o valor de b de 1 para 13) deslocou a massa das probabilidades de captura para valores menores e aumentou a estimativa da população N .

Em relação aos intervalos de credibilidade calculados para o tamanho populacional - $\{12\}$ e $[12, 17]$ para distribuições *a priori* das probabilidades de captura $U(0, 1)$ e $Beta(1, 13)$, respectivamente -, são menores do que os apresentados por Silveira et al. (2010) - $[13, 33]$. Vale destacar que o nível de confiança não foi reportado (assumiu-se como sendo 95%), mas, em ambos os casos, não foram observados valores de N maiores do que 30. Por fim, percebeu-se uma correlação negativa entre as distribuições *a posteriori* de N e p_1 e N e p_{12} .

Referências

- Chapman, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological sample censuses. Berkeley, Calif: University of California Publications in Statistics, 195(1).
- George, E. I. (1992). Capture—recapture estimation via gibbs sampling. *Biometrika*, 79(4):677–683.
- Salasara, L. E. B., Leite, J. G., e Louzada, F. (2016). Likelihood-based inference for population size in a capture—recapture experiment with varying probabilities from occasion to occasion. *Brazilian Journal of Probability and Statistics*, pages 47–69.
- Silveira, L., JAcomo, A. T., Astete, S., Sollmann, R., Tôrres, N. M., Furtado, M. M., e Marinho-Filho, J. (2010). Density of the near threatened jaguar panthera onca in the caatinga of north-eastern brazil. *Oryx*, 44(1):104–109.