

MI602: Métodos Computacionais em Estatística, Atividade I

Guilherme Ludwig
gvludwig@ime.unicamp.br

13 de Outubro de 2022

- A atividade é individual. Referencie o trabalho citado, mas desenvolva as soluções na atividade. Explique resultados com suas próprias palavras.
- A solução deve ser desenvolvida em detalhe. Todas as contas devem estar explicadas e desenvolvidas. Não é preciso fazer operações elementares manualmente mas as contas devem estar claramente indicadas.
- Cada aluno deverá ser responsável por garantir que não há cópia do seu trabalho. Indícios de plágio podem levar à anulação da atividade de todos os envolvidos. Use seu bom senso para guiar as conversas com seus colegas.
- A atividade deve ser entregue através do Moodle, em um único documento em formato PDF.
 - As páginas da sua solução devem estar numeradas, e as linhas também devem estar numeradas. Use o pacote no \LaTeX `\usepackage{lineno}` e `\linenumbers`. Isso é importante pois me ajudará a dar *feedback* à sua solução.
 - Se você usar o Word ou LibreOffice, numere linhas e páginas e imprima o documento em formato PDF antes de enviá-lo.
- Use a linguagem que você preferir, de preferência com o pacote `listings` do \LaTeX para exibir código: R, C, python, julia, ou outra qualquer. Todos os códigos devem estar listados no final da atividade, com instruções de execução (à parte das linguagens listadas).
- Se você quiser, pode usar Rmarkdown, Jupyter notebooks ou o que preferir para mesclar código e texto.
- **Prazo de entrega:** 20/10/2022, às 23:59, via Moodle. Sem tolerância para atraso.

O Instituto Chico Mendes de Conservação da Biodiversidade (icmbio.gov.br) publicou em 2018 o *Livro Vermelho da Fauna Brasileira Ameaçada de Extinção*, em sete volumes (ICMBio, 2018), reunindo o trabalho de 1270 cientistas para a avaliação de 12.254 espécies. Dentre os livros publicados, o *Volume 2: Mamíferos* emprega extensivamente a metodologia de captura-recaptura, ou método de Lincoln-Petersen (Petersen, 1896; Lincoln, 1930), utilizando frequentemente o chamado estimador de Chapman para determinar a população de uma espécie de mamífero (Chapman, 1951). Dentre os diversos animais listados, a onça (*panthera onca*) possui diversos estudos listados. Na página 355 do *Livro Vermelho...*, lê-se que:

“Estudos utilizando armadilhas fotográficas, baseados em modelos de captura-marcação-recaptura, estimam (...) 2.67 ± 1.06 indivíduos/100 km² na caatinga (Silveira et al., 2009),”

entre outros biomas listados na referência. O artigo de Silveira et al. (2009) em particular, lista todos os dados brutos em seu estudo (ao invés de apenas as estatísticas) e por isso vamos nos concentrar nele. O estudo foi conduzido no Parque Nacional da Serra da Capivara (Piauí), e a identificação foi feita através de armadilhas com câmeras, que foram armadas por períodos de 6 dias não-consecutivos. Os dados observados estão listados na Tabela 1.

A metodologia de estimação de abundância foi revisitada quase cem anos depois no contexto Bayesiano por George and Robert (1992). Considere o Exemplo 9.9 de Robert and Casella (2004) e o exercício correspondente da Lista 2: supondo que os ensaios de captura foram independentes (o que é uma suposição questionável, mas

Tabela 1: Histórico de capturas de 12 onças identificadas no Parque Nacional da Serra da Capivara, em 2007 (o indivíduo 4 é um filhote e foi removido). Uma entrada com 1 representa a captura de um indivíduo em uma das sessões de armadilhas de 6 dias. Dados de Silveira et al. (2009).

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

que será feita por simplicidade), e que queremos determinar o tamanho N da população de onças na região do estudo, e suponha que a probabilidade de captura dos animais em um certo dia de experimento/armadilha é constante para todos os animais e as capturas são independentes entre si. Porém, a probabilidade pode mudar entre os dias. Defina n_i como o número de animais capturados na armadilha i ($i = 1, \dots, 14$), m_j como o número de animais capturados na armadilha j que já foram capturados alguma vez anteriormente ($j = 2, \dots, 14$), e

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

Então a verossimilhança é

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

Objetivos

- Assuma a distribuição *a priori* $\pi(N|\cdot) \sim \text{Poisson}(\lambda)$ e $\pi(p_i|\cdot) \sim U(0,1)$, independentes entre si. Encontre as distribuições condicionais completas $f(N|p_1, \dots, p_{14}, \text{dados})$, $f(p_i|N, p_{\{k:k \neq i\}}, \text{dados})$. Implemente o amostrador de Gibbs, usando como hiperparâmetro $\lambda = 50$, e como amostras iniciais as distribuições *a priori*. Justifique que as distribuições *a posteriori* atingiram a estacionariedade.
- Desenvolva a metodologia do item anterior através de um método de *Hamiltonian Monte Carlo*, e compare os resultados do item anterior.
- (pontos extras): Implemente o método de HMC no STAN.
- Produza um relatório descrevendo a implementação, discutindo a metodologia estatística e comparando com os resultados obtidos por (Silveira et al., 2009). O relatório deve ter formato científico e poucas páginas.

Referências

- D. G. Chapman. Some properties of the hypergeometric distribution with applications to zoological sample censuses, 1951.
- E. I. George and C. P. Robert. Capture-recapture estimation via Gibbs sampling. *Biometrika*, 79(4):677–683, 1992.
- ICMBio. Livro vermelho da fauna brasileira ameaçada de extinção, 2018. URL <https://www.icmbio.gov.br/portal/component/content/article/10187>.
- F. C. Lincoln. Calculating waterfowl abundance on the basis of banding returns. *United States Department of Agriculture Circular*, 118, 1930.
- C. G. J. Petersen. The yearly immigration of young plaice in the limfjord from the german sea. *Report of the Danish Biological Station*, 6:1–48, 1896.
- C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.
- L. Silveira, A. T. Jácomo, S. Astete, R. Sollmann, N. M. Tôrres, M. M. Furtado, and J. Marinho-Filho. Density of the near threatened jaguar *Panthera onca* in the caatinga of north-eastern Brazil. *Oryx*, 44(1): 104–109, 2009.