

## Exam 2 (A2)

Class: Bayesian Statistics  
Instructor: Luiz Max Carvalho

02/06/2021

**Turn in date: until 16/06/2021 at 23:59h Brasilia Time.**

- Please read through the whole exam before starting to answer;
- State and prove all non-trivial mathematical results necessary to substantiate your arguments;
- Do not forget to add appropriate scholarly references *at the end* of the document;
- Mathematical expressions also receive punctuation;
- You can write your answer to a question as a point-by-point response or in “essay” form, your call;
- Please hand in a single, **typeset** (  $\LaTeX$  ) PDF file as your final main document. Code appendices are welcome, *in addition* to the main PDF document.
- You may consult any sources, provided you cite **ALL** of your sources (books, papers, blog posts, videos);
- You may use symbolic algebra programs such as Sympy or Wolfram Alpha to help you get through the hairier calculations, provided you cite the tools you have used.
- The exam is worth 100 marks.

## Background

This exam covers applications, namely estimation, prior sensitivity and prediction. You will need a working knowledge of basic computing tools, and knowledge of MCMC is highly valuable. Chapter 6 in [Robert \(2007\)](#) gives an overview of computational techniques for Bayesian statistics.

## Inferring population sizes – theory

Consider the model

$$x_i \sim \text{Binomial}(N, \theta),$$

with **both**  $N$  and  $\theta$  unknown and suppose one observes  $\mathbf{x} = \{x_1, x_2, \dots, x_K\}$ . Here, we will write  $\xi = (N, \theta)$ .

- a) (10 marks) Formulate a hierarchical prior ( $\pi_1$ ) for  $N$ , i.e., elicit  $F$  such that  $N \mid \alpha \sim F(\alpha)$  and  $\alpha \sim \Pi_A$ . Justify your choice;
- b) (5 marks) Using the prior from the previous item, write out the full joint posterior kernel for all unknown quantities in the model,  $p_1(\xi \mid \mathbf{x})$ . *Hint*: do not forget to include the appropriate indicator functions!;
- c) (5 marks) Is your model identifiable?
- d) (5 marks) Exhibit the marginal posterior density for  $N$ ,  $p_1(N \mid \mathbf{x})$ ;
- e) (5 marks) Return to point (a) above and consider an alternative, uninformative prior structure for  $\xi$ ,  $\pi_2$ . Then, derive  $p_2(N \mid \mathbf{x})$ ;
- f) (10 marks) Formulate a third prior structure on  $\xi$ ,  $\pi_3$ , that allows for the closed-form marginalisation over the hyperparameters  $\alpha$  – see (a) – and write out  $p_3(N \mid \mathbf{x})$ ;
- g) (10 marks) Show whether each of the marginal posteriors considered is proper. Then, derive the posterior predictive distribution,  $g_i(\tilde{x} \mid \mathbf{x})$ , for each of the posteriors considered ( $i = 1, 2, 3$ ).
- h) (5 marks) Consider the loss function

$$L(\delta(\mathbf{x}), N) = \left( \frac{\delta(\mathbf{x}) - N}{N} \right)^2. \quad (1)$$

Derive the Bayes estimator under this loss.

## Inferring population sizes – practice

Consider the problem of inferring the population sizes of major herbivores ([Carroll and Lombard, 1985](#)). In the first case, one is interested in estimating the number of impala (*Aepyceros melampus*) herds in the Kruger National Park, in northeastern South Africa. In an initial survey collected the following numbers of herds:  $\mathbf{x}_{\text{impala}} = \{15, 20, 21, 23, 26\}$ . Another scientific question is the number of individual waterbuck (*Kobus ellipsiprymnus*) in the same park. The observed numbers of waterbuck in separate sightings were  $\mathbf{x}_{\text{waterbuck}} = \{53, 57, 66, 67, 72\}$  and may be regarded (for simplicity) as independent and identically distributed.



(a) Impala



(b) Waterbuck

Figure 1: Two antelope species whose population sizes we want to estimate.

- i) (20 marks) For each data set, sketch the marginal posterior distributions  $p_1(N | \mathbf{x})$ ,  $p_2(N | \mathbf{x})$  and  $p_3(N | \mathbf{x})$ . Moreover, under each posterior, provide (i) the Bayes estimator under quadratic loss and under the loss in (1) and (ii) a 95% credibility interval for  $N$ . Discuss the differences and similarities between these distributions and estimates: do the prior modelling choices substantially impact the final inferences? If so, how?
- j) (25 marks) Let  $\bar{x} = K^{-1} \sum_{k=1}^K x_k$  and  $s^2 = K^{-1} \sum_{k=1}^K (x_k - \bar{x})^2$ . For this problem, a sample is said to be *stable* if  $\bar{x}/s^2 \geq (\sqrt{2} + 1)/\sqrt{2}$  and *unstable* otherwise. Devise a simple method of moments estimator (MME) for  $N$ . Then, using a Monte Carlo simulation, compare the MME to the three Bayes estimators under quadratic loss in terms of relative mean squared error. How do the Bayes estimators compare to MME in terms of the stability of the generated samples? *Hint*: You may want to follow the simulation setup of [Carroll and Lombard \(1985\)](#).

## Bibliography

- Carroll, R. J. and Lombard, F. (1985). A note on  $N$  estimators for the binomial distribution. *Journal of the American Statistical Association*, 80(390):423–426.
- Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.