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 θ unknown and suppose one observes $\boldsymbol{x} = \{x_1, x_2, \ldots, x_K\}$. Here, we will write $\xi = (N, \theta)$.

a) (10 marks) Formulate a hierarchical prior (π_1) for N, i.e., elicit F such that $N \mid \alpha \sim F(\alpha)$ and $\alpha \sim \Pi_A$. Justify your choice;

Solution. The only restriction prior to the data we have is that $N \in \mathbb{Z}^+$. Therefore we are looking for a distribution in this space. We can follow the assumption made by Raftery (1988) where $N \sim \text{Poisson}(\mu)$. Other approach is to consider $N \sim \text{Geometric}(\nu)$.
The first approach has the advantage in calculations. When

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

and

$$N \sim \text{Poisson}(\mu),$$

we have that $x_i \sim \text{Poisson}(\mu \cdot \theta)$, as proved in Appendix A. And we can convert prior information in x_i in terms of our beliefs about its mean.
The second approach serves as a comparative. The interesting part of this distribution is the compact domain since $\nu \in [0, 1]$. From that, we can build a model that caries the correlation between ν and θ whenever necessary in a direct way: we transform the variables to logistic space and defines a bivariate normal distribution. We would have to define five parameters, but this is a good extension.

Priors to the hyperparameters

Now we shall define the priors to the hyperparameters. Define $\lambda = \mu \cdot \theta$. Given that it is morel likely that previous research made statements about x_i, we use the same idea as Raftery considering the prior over (λ, θ). I will assume they are independent from now on with

$$\lambda \sim \text{Gamma}(\alpha, \beta),$$

and

$$\theta \sim \text{Beta}(a, b).$$

The other approach will have two settings. First I suppose ν and θ independents with

$$\nu \sim \text{Beta}(\alpha_1, \beta_1)$$

and

$$\theta \sim \text{Beta}(\alpha_2, \beta_2).$$

I choose the Beta distribution because it has a flexible shape with a good intuition behind it. Other point is that the Beta distribution forms a conjugate family for the Geometric distribution. Another set up is to consider the correlated case. We do it in the following way:

$$\begin{pmatrix} \text{logit}(\nu) \\ \text{logit}(\theta) \end{pmatrix} \sim \text{Normal}(\eta, \Sigma).$$

This choice is intrinsically linked to the fact the the normal distribution is a good approximation to a series of events, and it has a very good interpretation of the parameters. The problem with this approach is that it is harder to codify prior information. We necessarily need information about N, θ, and how they relate.
From these three approaches, I will call these approaches in the text (1) *Raftery approach*, (2) *Geometric and independent approach*, and (3) *Geometric and correlated approach.*

b) (5 marks) Using the prior from the previous item, write out the full joint posterior kernel for all unknown quantities in the model, $p(\xi \mid \boldsymbol{x})$. *Hint:* do not forget to include the appropriate indicator functions!;

Solution. For the Geometric and correlated approach, it may be impossible to find a simple expression.
Generally speaking, by Bayes' Theorem,

$$\begin{aligned} p(\xi|\boldsymbol{x}) &\propto l(\xi|\boldsymbol{x}) \cdot \pi(\xi) \\ &= \left(\prod_{i=1}^{n} \binom{N}{x_i} \theta^{x_i}(1-\theta)^{N-x_i}\right) \cdot \pi(\xi) \\ &= \left(\prod_{i=1}^{n} \binom{N}{x_i}\right) \theta^{S}(1-\theta)^{nN-S} \cdot \pi(\xi), \end{aligned} \tag{1}$$

where $S = \sum_{i=1}^{n} x_i$. We shall derive for each case the prior $\pi(\xi)$.

(1) **Raftery**:

$$\begin{aligned} \pi(\xi) &= \int_0^\infty \pi(\xi, \lambda)\, d\lambda \\ &= \int_0^\infty \pi(N|\theta, \lambda)\pi(\theta, \lambda)\, d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda/\theta}(\lambda/\theta)^N}{N!} \pi(\lambda)\pi(\theta)\, d\lambda \\ &\propto \int_0^\infty \frac{e^{-\lambda/\theta}(\lambda/\theta)^N}{N!} \lambda^{\alpha-1} e^{-\beta\lambda} \theta^{a-1}(1-\theta)^{b-1} \mathbb{1}(0<\theta<1)\, d\lambda \\ &= \frac{\theta^{a-1-N}(1-\theta)^{b-1}}{N!} \mathbb{1}(0<\theta<1) \int_0^\infty \lambda^{\alpha+N-1} e^{-(\beta+1/\theta)\lambda}\, d\lambda \\ &= \frac{\theta^{a-1-N}(1-\theta)^{b-1}}{N!} \mathbb{1}(0<\theta<1) \cdot \frac{\Gamma(\alpha+N)}{(\beta+1/\theta)^{\alpha+N}}, \end{aligned}$$

since the integrand is the kernel of a gamma distribution. Therefore, rewriting,

$$\pi(\xi) \propto \frac{\Gamma(\alpha+N)\theta^{a-1-N}(1-\theta)^{b-1}}{(\beta+1/\theta)^{\alpha+N} N!} \mathbb{1}(0<\theta<1) \tag{2}$$

and

$$p(\xi|\boldsymbol{x}) \propto \left(\prod_{i=1}^{n} \binom{N}{x_i}\right) \frac{\Gamma(\alpha+N)\theta^{a+S-1-N}(1-\theta)^{b+nN-S-1}}{(\beta+1/\theta)^{\alpha+N} N!} \mathbb{1}(0<\theta<1) \tag{3}$$

(2) **Geometric and independent**:

$$\begin{aligned}
\pi(\xi) &= \int_0^1 \pi(\xi, \nu)\, d\nu \\
&= \int_0^1 \pi(N|\nu, \theta)\pi(\nu)\pi(\theta)\, d\nu \\
&\propto \theta^{\alpha_2 - 1}(1-\theta)^{\beta_2 - 1}\mathbb{1}(0 < \theta < 1) \int_0^1 (1-\nu)^N \nu \cdot \nu^{\alpha_1 - 1}(1-\nu)^{\beta_1 - 1}\, d\nu \\
&= \theta^{\alpha_2 - 1}(1-\theta)^{\beta_2 - 1}\mathbb{1}(0 < \theta < 1) \int_0^1 \nu^{\alpha_1}(1-\nu)^{N+\beta_1 - 1}\, d\nu \\
&= \theta^{\alpha_2 - 1}(1-\theta)^{\beta_2 - 1}\mathbb{1}(0 < \theta < 1) B(\alpha_1 + 1, N + \beta_1),
\end{aligned}$$

since the integrand is the kernel of a Beta distribution. Rewriting,

$$\pi(\xi) \propto B(\alpha_1 + 1, N + \beta_1)\theta^{\alpha_2 - 1}(1-\theta)^{\beta_2 - 1}\mathbb{1}(0 < \theta < 1) \tag{4}$$

and

$$\pi(\xi|\boldsymbol{x}) \propto \left(\prod_{i=1}^n \binom{N}{x_i}\right) B(\alpha_1 + 1, N + \beta_1)\theta^{S + \alpha_2 - 1}(1-\theta)^{nN + \beta_2 - S - 1}\mathbb{1}(0 < \theta < 1) \tag{5}$$

(3) **Geometric and correlated**: This is the harder case. I have to derive $\pi(\nu, \theta)$ from $\pi(\text{logit}(\nu), \text{logit}(\theta))$. Define $f(x_1, x_2) = (\text{logit}^{-1}(x_1), \text{logit}^{-1}(x_2))$. This is an invertible function with $f^{-1}(y_1, y_2) = (\text{logit}(y_1), \text{logit}(y_2))$. By the change of variables,

$$\pi(\nu, \theta) = \pi(f^{-1}(\nu, \theta)) \cdot \left| \det \left[\frac{df^{-1}(z)}{dz}\bigg|_{z=(\nu,\theta)} \right] \right|.$$

Observe that

$$\det \left[\frac{df^{-1}(z)}{dz}\bigg|_{z=(\nu,\theta)} \right] = \det \begin{bmatrix} \frac{d}{d\nu}\text{logit}(\nu) & 0 \\ 0 & \frac{d}{d\theta}\text{logit}(\theta) \end{bmatrix} = \frac{d}{d\nu}\text{logit}(\nu) \cdot \frac{d}{d\theta}\text{logit}(\theta).$$

By the calculation of the Jacobian, we can join everything

$$\begin{aligned}
\pi(\xi) &= \int_0^1 \pi(N|\nu, \theta)\pi(\nu, \theta)\, d\nu \\
&= \int_0^1 (1-\nu)^N \nu \pi(\text{logit}(\nu), \text{logit}(\theta)) |\,\text{logit}'(\nu) \cdot \text{logit}'(\theta)|\, d\nu \\
&= \int_0^1 (1-\nu)^N \nu \pi(\text{logit}(\nu), \text{logit}(\theta)) \frac{1}{\nu(1-\nu)\theta(1-\theta)}\, d\nu \\
&= \frac{1}{\theta(1-\theta)} \int_0^1 (1-\nu)^{N-1} \pi(\text{logit}(\nu), \text{logit}(\theta))\, d\nu.
\end{aligned}$$

Let $z = (\text{logit}(\nu), \text{logit}(\theta))$. Therefore,

$$\pi(\xi) \propto \frac{1}{\theta(1-\theta)} \int_0^1 (1-\nu)^{N-1} \exp\left\{ -\frac{1}{2}(z-\eta)^T \Sigma^{-1}(z-\eta) \right\} d\nu \tag{6}$$

and

$$\pi(\xi|\boldsymbol{x}) \propto \left(\prod_{i=1}^n \binom{N}{x_i}\right) \theta^{S-1}(1-\theta)^{nN - S - 1} \int_0^1 (1-\nu)^{N-1} \exp\left\{ -\frac{1}{2}(z-\eta)^T \Sigma^{-1}(z-\eta) \right\} d\nu. \tag{7}$$

c) (5 marks) Is your model identifiable?

Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

d) (5 marks) Exhibit the marginal posterior density for N, $p_1(N \mid \boldsymbol{x})$;

Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

e) (5 marks) Return to point (a) above and consider an alternative, uninformative prior structure for ξ, π_2. Then, derive $p_2(N \mid \boldsymbol{x})$;

Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

f) (10 marks) Formulate a third prior structure on ξ, π_3, that allows for the closed-form marginalization over the hyperparameters α – see (a) – and write out $p_3(N \mid \boldsymbol{x})$;

Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

g) (10 marks) Show whether each of the marginal posteriors considered is proper. Then, derive the posterior predictive distribution, $g_i(\tilde{x} \mid \boldsymbol{x})$, for each of the posteriors considered ($i = 1, 2, 3$).

Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

h) (5 marks) Consider the loss function

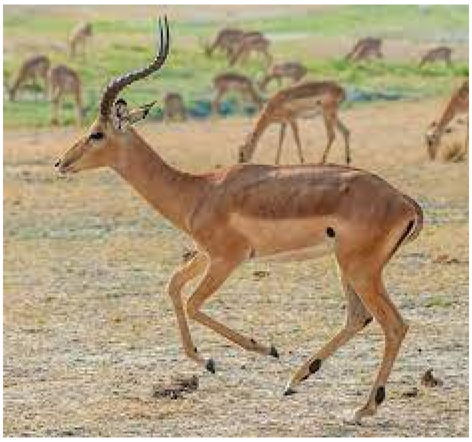
$$L(\delta(\boldsymbol{x}), N) = \left(\frac{\delta(\boldsymbol{x}) - N}{N} \right)^2. \tag{8}$$

Derive the Bayes estimator under this loss.

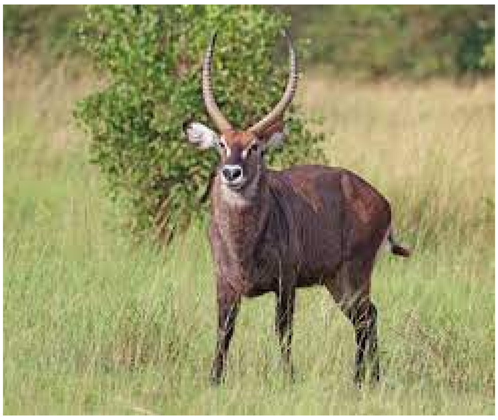
Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

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: $\boldsymbol{x}_{\text{impala}} = \{15, 20, 21, 23, 26\}$. Another scientific question is the number of individual waterbuck (*Kobus ellipsiprymnus*) in the same park. The observed numbers of waterbuck in separate sightings were $\boldsymbol{x}_{\text{waterbuck}} = \{53, 57, 66, 67, 72\}$ 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 $p_1(N \mid \boldsymbol{x})$, $p_2(N \mid \boldsymbol{x})$ and $p_3(N \mid \boldsymbol{x})$. Moreover, under each posterior, provide (i) the Bayes estimator under quadratic loss and under the loss in (8) 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?

Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

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).

Solution. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Appendix

A Binomial and Poisson

Suppose $X|N \sim \text{Binomial}(N, p)$ and $N \sim \text{Poisson}(\mu)$. We shall derive the distribution of X. By the Law of total probability,

$$
\begin{aligned}
\Pr(X = k) &= \sum_{n=0}^{\infty} \Pr(X = k|N = n) \Pr(N = n) \quad [\Pr(X = k|N = n) = 0 \text{ if } k > n] \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{e^{-\mu}\mu^n}{n!} \\
&= \frac{e^{-\mu}p^k}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}}{(n-k)!} \mu^{n-k} \mu^k \\
&= \frac{e^{-\mu}(p\mu)^k}{k!} \sum_{n=k}^{\infty} \frac{((1-p)\mu)^{n-k}}{(n-k)!} \qquad (9) \\
&= \frac{e^{-\mu}(p\mu)^k}{k!} \sum_{m=0}^{\infty} \frac{((1-p)\mu)^m}{m!} \quad [m = n-k] \\
&= \frac{e^{-\mu}(p\mu)^k}{k!} e^{(1-p)\mu} \\
&= \frac{e^{-p\mu}(p\mu)^k}{k!},
\end{aligned}
$$

what implies $X \sim \text{Poisson}(p \cdot \mu)$.

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.

Raftery, A. E. (1988). Inference for the binomial n parameter: A hierarchical bayes approach. *Biometrika*, 75(2):223–228.

Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation.* Springer Science & Business Media.