STAT305: Introduction to Inference **Fall 2021-22**

Assignment 3

Due date: see Piazza post.

Instructions

Academic integrity policy: We encourage you to discuss verbally with other students about the assignment. However, you should write your answers by yourself (in particular, without ever putting your eyes on someone else's answer or draft answer). If you use online resources (including asking questions on math exchange or similar), you *have* to cite them. If you have not done se already, read: `http://learningcommons.ubc.ca/guide-to-academic-integrity/` Please also make sure to read the audit process described during lecture 1.

Instructions: read this right away!

1. Due date: refer to the Piazza post announcing the assignment.
2. Justify all your answers. Use precise mathematical notation. Define all symbols you introduce. Marks will be deducted for imprecise, incorrect or insufficient justification.
3. Write your answers in order, and clearly label the question you are answering, for example 1.1 for part 1 of question 1.
4. We only accept submissions via Gradescope. Please make sure at least 24 hours in advance that you are confortable with the submission process, to give us enough time to fix any problems with the submission process.
5. See slide in lecture 1 for late assignment policy.

1 Death by horsekick

Historically, one of the first application of the Poisson approximation (law of rare events), was to study the number of soldiers of the Prussian army killed by accidental horse kick. Each year, the number of soldiers killed by accidental horse kick was recorded:

Year	Number of deaths (y_i)
1881	7
1882	1
1883	3
1884	2
1885	7
1886	6
1887	1
1888	3
1889	2
1890	2
1891	6
1892	4
1893	4
1894	1
1895	6
1896	2

Since the army is large, and the probability of each solider dying from accidental horse kick in one year is low, it is a priori reasonable to invoke the Poisson approximation and to assume a Poisson likelihood model, i.e. to model the number of death Y_i in a given year i using a Poisson distribution with an unknown mean μ_i. Let us assume the army size and horse kick risk is constant, i.e. $\mu_i = \mu$ for all i.

1. Take a Bayesian approach, putting an exponential prior on μ with prior mean 10. Use conjugacy to provide a closed form expression for the posterior distribution on μ given the above observed data.

   [5 mark(s)]

2. Let Y_{1897} denote the number of deaths by accidental horsekicks in 1897 (the year after our dataset ends). Compute $P(Y_{1897} = 0|Y)$, where $Y = (Y_{1881}, Y_{1882}, \ldots, Y_{1896})$, i.e. the probability that there is no accidental death by horsekick in 1897 given the observations. Hint: start with the following identity, a consequence of the law of total expectation:[1]

$$P(Y_{1897} = 0|Y) = \mathbb{E}[P(Y_{1897} = 0|\mu)|Y]$$

   [3 mark(s)]

2 More flexible conjugate prior

Let us extend the example from Worksheet 6.4–6.5 (the HIV example). We assume the same binomial likelihood model, $Y|\Pi \sim \text{Binom}(\Pi, 8197)$, but let us say we cannot decide which of the three priors to use (recall prior 1 is $\text{Beta}(\alpha_1 = 1, \beta_1 = 10)$, prior 2 is $\text{Beta}(\alpha_2 = 1, \beta_2 = 1)$, and prior 3, $\text{Beta}(\alpha_3 = 10, \beta_3 = 1)$).

Instead of sticking to one prior, one possibility is to build a new prior which is a linear combination of the three possible priors:

$$f_\Pi(\pi) = \sum_{i=1}^{3} \lambda_i \text{Beta}(\pi; \alpha_i, \beta_i), \tag{1}$$

where let us say $\lambda_i = 1/3$. In the above equation, $\text{Beta}(\pi; \alpha_i, \beta_i)$ denotes the density of the beta distribution.

[1] The usual law of total expectation gives us $P(A) = \mathbb{E}[P(A|Z)]$, applying this with the conditional probability $P(\cdot|Y)$ and corresponding expectation $\mathbb{E}[\cdot|Y]$ gives the identity.

1. Assuming the prior given by Equation (1), show that the posterior density $f_{\Pi|Y}(\pi|y)$ has the same form as Equation (1) but with updated $\lambda_i, \alpha_i, \beta_i$. In other words, the prior defined in Equation (1) is conjugate to the binomial likelihood. Clearly show the formulae for updating the values of $\lambda_i, \alpha_i, \beta_i$.

[4 mark(s)]

2. Derive the density of the posterior based on the prior in Equation (1) in the HIV example, where $Y = 51$.

   where $Z = 1/3\frac{B(52,8156)}{B(1,10)} + 1/3\frac{B(52,8147)}{B(1,1)} + 1/3\frac{B(61,8147)}{B(10,1)}$.

[1 mark(s)]

3 Bayesian calibration

Let Π denote a parameter of interest, and Y, the observations. Let also $L(y)$ and $R(y)$ denote the left and right end points produced by the 95% credible interval procedure covered in worksheets 6.3–6.5.

1. Compute $P(L(Y) \leq \Pi \leq R(Y))$.

[3 mark(s)]

2. Interpret the above probability in terms of "repeated experiments." Conclude that Bayesian procedures have a certain calibration property.

[2 mark(s)]