STAT 431 — Applied Bayesian Analysis — Fall 2022

Homework 4

Due: October 21, 11:59 PM (US Central)

In your submission, please include any computer code and output that you used to answer the following problems. Any numerical values should be accurate to at least three significant digits, unless otherwise stated.

1. Recall the Jevons's Coins weight comparison:

   - For $n_1 = 24$ coins minted before 1830: $\overline{y}_1 = 7.8730$, $s_1 = 0.05353$
   - For $n_2 = 123$ newer coins (1860's): $\overline{y}_2 = 7.9725$, $s_2 = 0.01409$

   R Example 3.3 (`ex3.3.R`, posted in the Classroom Lecture Materials module on Canvas) illustrates a Bayesian analysis of the mean difference $(\mu_1 - \mu_2)$, using Monte Carlo simulation. Using the *same simulation*, you will analyze the variance ratio σ_1^2/σ_2^2.

   (a) [1 pt] Estimate the posterior mean of σ_1^2/σ_2^2.

   (b) [2 pts] Estimate the Monte Carlo error of your estimated posterior mean.

   (c) [1 pt] Estimate the posterior *standard deviation* of σ_1^2/σ_2^2.

   (d) [2 pts] Form an approximate 95% equal-tailed credible interval for σ_1^2/σ_2^2.

   (e) [2 pts] There is a frequentist $(1-\alpha)100\%$ confidence interval for σ_1^2/σ_2^2:

   $$\left[\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} \frac{s_1^2}{s_2^2}, \quad F_{\alpha/2,n_2-1,n_1-1} \frac{s_1^2}{s_2^2} \right]$$

   where $F_{\alpha/2,\nu_1,\nu_2}$ is the $\alpha/2$ *upper* quantile of the F distribution with ν_1 numerator and ν_2 denominator degrees of freedom.[1] Note carefully that the degrees of freedom are $(n_1 - 1, n_2 - 1)$ in the expression for the lower endpoint, but $(n_2 - 1, n_1 - 1)$ in the expression for the upper endpoint.

   Compute the 95% confidence interval and compare it with your credible interval from the previous part.

2. If a probability model for random data Y has density $f(y \mid \theta)$, where θ is the unknown parameter, then the *deviance function* of θ, for observed data y, is

   $$D(y \mid \theta) \quad = \quad -2\ln\big(f(y \mid \theta)\big)$$

   that is, twice the negative log likelihood. Its *posterior expected value*

   $$\mathrm{E}\big(D(y \mid \theta) \mid y\big)$$

   plays a role in Bayesian model selection (see BSM, Sec. 5.5).

   From a past class survey, $y = 12$ out of $n = 70$ sampled students had pets. R Example 3.1 (`ex3.1.R`, posted under Classroom Lecture Materials) illustrates how to approximate the

[1] See, for example, Sec. 6.4 of R.V. Hogg & E.A. Tanis (2010) *Probability and Statistical Inference*, 8th ed., Pearson.

posterior mean of the population proportion θ of people like us who have pets. It assumes a binomial model and Jeffreys' prior. Using the same binomial model and Jeffreys' prior, you will approximate the *posterior expected deviance*, i.e., the posterior expected value of the deviance.

(a) [2 pts] Show that, for this situation, the deviance function is

$$D(y = 12 \mid \theta) \quad = \quad -2\left(\ln\binom{70}{12} \; + \; 12\ln\theta \; + \; 58\ln(1-\theta)\right)$$

(b) [2 pts] Write out the mathematical formulas for all of the integrals you will approximate in the next part.

(c) [3 pts] Approximate the integrals using R function `integrate`, then use the results to approximate the posterior expected deviance.

3. GRADUATE SECTION ONLY

Let $N \geq 0$ be the (unknown) number of birds of a certain species that live in a forest. An ecological expedition into the forest sights any particular bird with (unknown) probability θ ($0 < \theta < 1$), independently between birds. Thus, if the expedition sights a total of Y different birds of that species,

$$Y \mid N = n, \theta \quad \sim \quad \text{Binomial}(n, \theta)$$

(Note that we always have $N \geq Y$.)

Ordinarily, estimating two parameters with a single datum would be nearly impossible. Fortunately, a Bayesian can compensate by using a relatively informative prior.

Under the prior, let N and θ be independent with densities

$$\pi(n) \;\propto\; (0.9)^n, \quad n = 0, 1, 2, \ldots \qquad\qquad \pi(\theta) \;\propto\; 1, \quad 0 < \theta < 1$$

where $\pi(n)$ is a PMF for N and $\pi(\theta)$ a PDF for θ.

(a) [2 pts] Identify the (posterior) *full conditional distribution* of θ: Name it, and (in the parameterization of BSM) express its constant(s) in terms of n (the value of N) and y (the value of Y).

(b) [2 pts] Let $R = N - Y$. Show that, conditional on $Y = y$ and θ, the distribution of R is negative binomial, and (in the parameterization of BSM) express its two constants in terms of y and θ.

(c) [1 pt] Suppose $y = 20$ birds were sighted. To at least two significant digits, approximate the posterior probability that at least half of the birds in the forest were sighted, i.e., $\text{Prob}(N \leq 2Y \mid Y = 20)$. [Hint: Use a Gibbs sampler, noting that you can sample R to get a draw for $N = R + 20$. R function `rnbinom` will be helpful, but be careful about the order of its arguments.]