STAT 431 — Applied Bayesian Analysis — Spring 2019

Homework 4

Please submit your assignment *on paper*. Make sure your answers are completely justified and clear enough to read! Any computer code and output should be included.

1. From the class survey, $y = 12$ out of $n = 70$ sampled students had pets. R Example 8.1 (`ex8.1.R`, posted under Lecture Materials) illustrates how to approximate the posterior *mean* of the population proportion π 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 *variance* of π.

   (a) [2 pts] Write out all of the mathematical formulas for the integrals you will compute using R.

   (b) [2 pts] Perform the integrations using function `integrate()` in R. (You may either use the exact value for the posterior mean of π, or approximate it using `integrate()`.)

   (c) [2 pts] Now compute the posterior variance analytically (with the help of conjugacy and Table A.2 in Cowles), and compare this answer to your approximation.

2. Official combined city/highway energy consumption ratings (MPGe, miles per gallon gasoline equivalent) are given below for 2019 model *all-electric* vehicles in two categories:[1]

Small Cars	
Hyundai Ioniq Electric	136
Volkswagen e-Golf	119
Honda Clarity EV	114
BMW i3	113
BMW i3s	113
Nissan Leaf	112
Fiat 500e	112
smart EQ fortwo (coupe)	108
smart EQ fortwo (convertible)	102

Sport Utility Vehicles (SUVs)	
Hyundai Kona Electric	120
Tesla Model X 75D	93
Tesla Model X 100D	87
Tesla Model X P100D	85
Jaguar I-Pace	76

   Regard MPGe as independent between vehicles and normally-distributed within category, with both mean *and* variance possibly differing by category. Use "independent" "standard" (product-Jeffreys) priors, as illustrated in R Example 8.3 (`ex8.3.R`, posted under Lecture Materials). Use at least 100000 simulation samples for all of your approximations.

   (a) [1 pt] Compute the *sample* means and *sample* standard deviations for the two categories.

   (b) [2 pts] Compute an approximate 95% equal-tailed credible interval for the difference between the mean for small cars and the mean for SUVs. Do the means appear to differ?

   (c) [1 pt] Approximate the posterior probability that the mean for small cars does *not* exceed the mean for SUVs.

[1] Data from `https://fueleconomy.gov`

(d) [3 pts] Compute the (frequentist) Welch two sample t-test one-sided p-value for testing the null hypothesis that the mean for small cars does *not* exceed the mean for SUVs. (Use R function `t.test(..., ..., alternative=..., var.equal=FALSE)`, making sure to select the correct alternative.) Also compute the usual (frequentist) two sample t-test one-sided p-value that assumes equal variances (`t.test(..., ..., alternative=..., var.equal=TRUE)`). Compare with the Bayesian probability of the previous part.

(e) [2 pts] Compute an approximate 95% equal-tailed credible interval for the *ratio* of the *variance* for small cars to the *variance* for SUVs. Do the variances appear to differ?

3. GRADUATE SECTION ONLY

Consider data that is a single observed value y of (in the parameterization of Cowles)

$$Y \sim \text{Gamma}(\nu + 1, \beta)$$

with unknown ν that is a *non-negative integer*, and unknown $\beta > 0$.

Ordinarily, estimating two parameters with a single datum would be nearly impossible. Fortunately you are a Bayesian, and can use a relatively informative prior.

Under the prior, let ν and β be independent with densities

$$p(\nu) \propto \frac{1}{2^\nu}, \quad \nu = 0, 1, 2, \ldots \qquad\qquad p(\beta) \propto e^{-\beta}, \quad \beta > 0$$

(a) [2 pts] Derive the *full conditional distribution* of β: Name it, and express its parameter(s) in terms of ν and y (in the parameterization of Cowles).

(b) [2 pts] Derive the *full conditional distribution* of ν: Name it, and express its parameter(s) in terms of β and y (in the parameterization of Cowles).

(c) [1 pt] Approximate the value of $\mathrm{P}(\nu = 0 \mid Y = 4)$ to at least two significant digits. [Hint: Use a Gibbs sampler — unless you can figure it out analytically.]