Homework 2: DATA130048

Biostatistics

Due Thursday, April 30th, 2020

1 Problem 1: 20pt

A researcher is trying to estimate the mean number of accidents per week within 100 feet of the Gervais Street/Assembly Street intersection in Columbia. She assumes a Poisson(λ) model for the number of accidents X per week, so that the density function for X given λ is

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!},\ x = 0, 1, 2, \cdots, \lambda \geqslant 0$$

(a) She uses a standard exponential prior distribution for λ (i.e., an exponential with mean 1). Derive her posterior distribution for λ given a random sample $x_1, ..., x_n$ from n weeks.

(b) Using your "expert knowledge" (i.e., any common-sense guess) of the accident rate around this busy intersection (and resisting the urge to use the data given in part (c)!),suggest a different /better prior for λ. Derive the corresponding posterior distribution for λ given a random sample $x_1, ..., x_n$ from n weeks.

(c) If she gathers the following accident counts from 15 randomly selected weeks

$$1\ 0\ 4\ 1\ 4\ 2\ 5\ 3\ 0\ 3\ 1\ 2\ 2\ 4\ 1$$

find the posterior median and a 95% credible interval for λ using the standard exponential prior, along with these data.

(d) Give the posterior median and a 95% credible interval for λ using your own prior,along with the data in part (c).

2 Problem 2: 45pt

The eBay selling prices for auctioned Palm M515 PDAs are assumed to follow a normal distribution with μ and σ^2 unknown. We wish to perform inference on the mean selling price μ.

(a) Suppose we assume an $IG(1100, 250000)$ prior for σ^2 and let the prior for $\mu|\sigma^2$ be

$$p(\mu|\sigma^2) \propto (\sigma^2)^{-\frac{1}{2}} \exp^{-\frac{1}{2\sigma^2/s_0}(\mu-\delta)^2},$$

with $s_0 = 1$ and $\delta = 220$. If our sample data are: $(212, 249, 250, 240, 210, 234, 195, 199, 222, 213, 233, 251)$, then find a point estimate and 95% credible interval for μ. (Note, you can use either the conditional posterior or the marginal posterior of μ to obtain the interval.)

(b) Suppose now that we had assumned the independent improper priors

$$\begin{aligned} p(\mu) =& 1, -\infty < \mu < \infty \\ p(\sigma) =& 1/\sigma, 0 < \sigma < \infty. \end{aligned}$$

Using the same data as in part (a), find a point estimate and 95% credible interval for μ.

(c) How do your substantive conclusions in parts (a) and (b) differ, and how is this related to the different choices of priors?

(d) Now suppose (perhaps unrealistically) that we had known the true population variance was $\sigma^2 = 228$. Assuming a conjugate prior for μ with $\delta = 220$ and $\tau^2 = 25$,find a point estimate and 95% credible interval for the single unknown parameter μ.

(e) How (if at all) does the inference in part (d) differ from the inferences in parts (a) and (b)? Explain your answer intuitively.

3 Problem 3: 21 pt

Let $X_1, ..., X_n$ be i.i.d. data from a normal distribution with known mean 0 and unknown variance θ.

(a) Write the likelihood $L(\theta|\mathbf{x})$.

(b) Derive the Jeffreys prior for θ.

(c) Suppose we observe the 6 data values $x_1 = 2.75, x_2 = 1.78, x_3 = 0.36, x_4 = -1.64, x_5 = 0.17, x_6 = -2.03$. Write the posterior distribution, using your Jeffreys prior from part (b). Do you recognize the form of this posterior? Specify exactly what distribution it is, including the parameter values.

4 Problem 4: 14 pt

Two Bayesian statisticians, Barry and Brianna, are trying to estimate θ, the mean survival time for a population of terminally ill patients who have undergone a certain procedure meant to slow the spread of their disease. They consult with a medical expert, whose best guess of the most likely mean survival time is 400 days. The expert also believes there is a $2/3$ chance that the mean survival time is between 315 and 485 days.

(a) Barry wishes to use a normal prior for θ. Based on the expert opinion, what parameters would be good choices for the parameters of his prior? Explain your reasoning clearly.

(b) Brianna wishes to use a gamma prior for θ. Based on the expert opinion, what parameters would be good choices for the parameters of her prior? Explain your reasoning clearly.