Homework 1

Your Name Here

2026-01-07

Due January 25 end of day. The solution file should include a `.qmd` file rendered to `pdf` and any additional files needed.

Theory problems:

1. Exponential model with conjugate prior distribution:

1a Show that if $y|\theta$ is exponentially distributed with rate θ, then the gamma prior distribution is conjugate for inferences about θ given an iid sample of y values.

1b Show that the equivalent prior specification for the mean, $\phi = 1/\theta$, is inverse-gamma. (That is, derive the latter density function.)

1c The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ. Suppose the prior distribution for θ is a gamma distribution with coefficient of variation 0.5. (The *coefficient of variation* is defined as the standard deviation divided by the mean.) A random sample of light bulbs is to be tested and the lifetime of each obtained. If the coefficient of variation of the distribution of θ is to be reduced to 0.1, how many light bulbs need to be tested?

2. A Mixture Prior for Heart Transplant Surgeries

A hospital in the United States wants to evaluate their success rate of heart transplant surgeries. We observe the number of deaths, y, in a number of heart transplant surgeries. Let $y \sim \text{Pois}(\nu\lambda)$ where λ is the rate of deaths/patient and ν is the exposure (total number of heart transplant patients). When measuring rare events with low rates, maximum likelihood estimation can be notoriously bad and we have good prior knowledge to incorporate. To construct your prior distribution you talk to two experts. The first expert thinks that $p_1(\lambda)$ with a gamma$(3, 2000)$ density is a reasonable prior. The second expert thinks that $p_2(\lambda)$

with a gamma$(7, 1000)$ density is a reasonable prior distribution. You decide that each expert is equally credible so you combine their prior distributions into a mixture prior with equal weights: $p(\lambda) = 0.5 * p_1(\lambda) + 0.5 * p_2(\lambda)$

2a. What does each expert think the mean rate is, *a priori*? Which expert is more confident about the value of λ a priori (i.e. before seeing any data)?

2bWrite the posterior distribution up to a proportionality constant then solve for the proportionality constant (be careful about what constants a proportionality constant in this problem!). Solve for $K = \int L(\lambda; y)p(\lambda)d\lambda$, the normalizing constant for the density. This constant will involve Γ functions. Compute this posterior density and clearly express the density as a mixture of two gamma distributions.

2c. Suppose the hospital has $y = 8$ deaths with an exposure of $\nu = 1767$ surgeries performed. Plot this *normalized* posterior density and add a vertical line at the MLE. Add the prior density to the plot for comparison, in a different color.

Computing problem / Applied problem

3. Visualizing Election Results

This exercise has three challenges: first, manipulating the data in order to get the totals by state; second, replicating the Bayesian calculations to estimate the parameters of the prior distribution and completing the Bayesian analysis by state; third, making the graphs.

The file `pew_data.csv` consists of data from Pew Research Center polls taken during the 2008 election campaign. Each row of the dataset includes results on an indivual from a particular U.S. state. You can read these data into R using the tidy function `read_csv`.

For each state, estimate the percentage of the (adult) population who label themselves as "very liberal" using the ideology column, `ideo`. For your estimates, replicate the procedure for the Poisson-Gamma model that was used in BDA Section 2.7 to estimate cancer rates, that is, treat the number of very liberal respondents as Poisson distributed with an exposure related to the fraction of surveyed individuals (by state). You don't need to make maps; it will be enough to make scatterplots. Graph raw data estimates of Obama vote share vs observed proportion liberal, then graph Obama vote share (x-axis) vs posterior mean estimate of the propoprtion liberal (y-axis). For both of these plots, the limits of the x and y axes should be the same for boht plots.

Then make two more graphs: the number of respondents on the x-axis versus the raw proportion very liberal and another plot with the posterior mean inferred proportion very liberal. Again make sure the x and y limits are the same in these two plots.

Put them all in a single figure using `patchwork` package in R,). Obama's vote share can be found in `2008ElectionResult.csv`.

Explicitly comment on how the posterior means compare to the raw proportions as a function of vote share and population.

Important: you will need to drop Alaska and Hawaii from the computations when computing α and β for the procedure proposed in 2.7 to work (what happens when you leave in the observation for Hawaii? Why is it statistically reasonable to drop these observations?). You should still estimate θ_j for Alaska and Hawaii (as well as all other states).