Homework 2

Bayesian Data Analysis

Problem 1

Suppose you are interested in estimating the average total snowfall per year μ (in inches) for a large city on the East Coast of the United States.

Before you collect the data, you are 95% confident that the average snowfall could be between 38 and 42 inches, centered around 40 inches per year.

Assume individual yearly snow totals $y_1, \ldots, y_n$ are collected from a population that is assumed to be normally distributed with mean μ and known standard deviation $\sigma = 10$ inches.

Suppose you observe the yearly snowfall totals $38.6, 42.4, 57.5, 40.5, 51.7, 67.1, 33.4, 60.9, 64.1, 40.1, 40.7$, and 6.4 inches.

(1) Propose a conjugate Normal prior $\mu \sim N(\mu_0, \tau_0)$ for this problem.

(2) What would be the posterior distribution in such case? What can you comment about the posterior mean, and the relative contribution of prior and data to the posyerior expectation?

(3) Plot the prior and the posterior distribution in a single figure. What can you comment about the posterior distribution of μ?

(4) Compute the 10% and 90% quantiles of the posterior distribution (i.e. those values a and b such that $P(\mu < a|\text{Data}) = 0.1$ and $P(\mu < b|\text{Data}) = 0.9$, respectively).

(5) For reporting needs, instead of computing the posterior mean of μ you are asked to compute the posterior mean of the following function:

$$g(\mu) = \begin{cases} \log(\mu) & \text{if } \mu > 0 \\ 0 & \text{if } \mu \leq 0 \end{cases}$$

How would you compute it?

Problem 2

The Mayo Clinic conducted a study of $n = 50$ patients followed for one year on a new medication and found that 30 patients experienced no adverse side effects (ASE), 12 experienced one ASE, 6 experienced two ASEs and 2 experienced ten ASEs.

(1) Derive the pasterior of λ for the Poisson/gamma model $Y_1, \ldots, Y_n|\lambda \overset{\text{id}}{\sim}$ Poisson($\lambda\Big)$ and $\lambda \sim \text{Gamma}(a, b)$.

(2) Use the Polsson/gamma model with $a = b = 0.01$ to study the rate of adverse events. Plot the posterior and give the posterior mean and 95% credible interval.

(3) In Bayesian Analysis, it is important to assess the *sensitivity* of the posterior inference to prior specifications. Repeat the analysis with Gamma(0.1,0.1) and Gamma(1,1) priors and discuss the sensitivity of the results to the prior.

(4) Plot the data versus the Poisson($\hat{\lambda}$) PMF, where $\hat{\lambda}$ is the posterior mean of λ from part (2). Does the Poisson likelihood fit the data well?

(5) The current medication is thought to have around one adverse side effect per year. What is the posterior probability that this new medication has a higher side effect rate than the previous medication? Are the results sensitive to the prior?

Problem 3

Over the past 50 years California has experienced an average of $\lambda_0 = 75$ large wildfires per year. For the next 10 years you will record the number of large fires in California and then fit a Poisson/gamma model to these data. Let the rate of large fires in this future period, λ, have prior $\lambda \sim \text{Gamma}(a, b)$. Select a and b so that the prior is somehow uninformative, i.e. has prior variance around 100 and gives prior probability approximately $\text{Prob}\,(\lambda > \lambda_0) = 0.5$. This prior assessment corresponds to an equal probability a priori assigned to the hypothesis that the rate of fires may increase in the next 10 years *vs* that it may decrease.