Prior information in the Poisson-gamma model

Math 315, Adam Loy

Working toward a data augmentation prior

Consider the model Poisson-gamma model for a rate, where $Y|\lambda \sim Poisson(N\lambda)$ and $\lambda \sim Gamma(a, b)$. In a previous class, we showed that the posterior is $\lambda|Y \sim \text{Gamma}(a + y, b + N)$. The posterior mean is then given by

$$\widehat{\lambda} = E(\lambda|Y) = \frac{a+Y}{b+N} = (1-w)\frac{a}{b} + w\frac{Y}{N}$$

Use this fact to answer the following questions.

a. Find an expression for w.

$w = \frac{N}{N+b}$

b. Describe Y/N and a/b in words.

Y/N = sample rate

a/b = prior (mean) rate

c. If we view Gamma(a, b) as a data augmentation prior, what do a and b represent?

a = prior # events

b = prior observation time/space

d. When (in terms of N, a, and b) is $\widehat{\lambda}$ close to Y/N? How does this help you set a "weak" prior?

weak prior

if a, b are small (say .01)

or N, Y are large

e. When (in terms of N, a, and b) is $\widehat{\lambda}$ shrunk towards a/b?

N, Y are small (little data)

Using domain expertise

Suppose that causes of death in 2018 are reviewed in detail for a city with a population of 200,000. It is found that 3 people died of asthma, giving a crude estimated asthma mortality rate in the city of 1.5 cases per 100,000 people per year.

You propose the Poisson model $Y|\lambda \sim \text{Poisson}(2\lambda)$, where λ denotes the true underlying long-term asthma mortality rate in the city, measured in cases per 100,000 people per year.

Reviews of asthma mortality rates around the world suggest that mortality rates about 1.5 per 100,000 people are rare in Western countries, with typical asthma mortality rates around 0.6 per 100,000.

a. Using the above information from the domain area, set up two equations that could be used to specify the hyperparameters of $\lambda \sim \text{Gamma}(a, b)$.

Answers will vary. One possibility is

$E(\lambda) = \frac{a}{b} = 0.6$ [median or mode could also be used!]

$\int_0^{1.5} \pi(\lambda)\, d\lambda = 0.975$ [quantile chosen is clearly subjective!]

[Some might use mean ± 2SD idea]

b. Use R to solve this two equation, two unknown problem. What gamma prior distribution did you select?

c. If you are using a domain expert to elicit prior information, one way to check whether your choice is reasonable is to simulate observations (i.e. $y^{(i)}$s) from the prior predictive distribution. To do this in R, run the following code chunk.

```
S <- 1e4 # no. simulations
a <- ___ # specify your choice for a
b <- ___ # specify your choice for b
prior_lambdas <- rgamma(S, a, b)
prior_ys <- rpois(S, 2 * prior_lambdas)
hist(prior_ys, nclass = 50, main = "Prior predictive distribution", xlab = "Y")
```