STAT 431 — Applied Bayesian Analysis — Course Notes

Conjugate Priors for Count Data Models

Fall 2022

Conjugacy

Consider a model for data $\boldsymbol{Y}$ with densities

$$f(\boldsymbol{y} \mid \boldsymbol{\theta})$$

parameterized by $\boldsymbol{\theta}$. (For the observed data $\boldsymbol{y}$, this is the likelihood function of parameter $\boldsymbol{\theta}$.)

Let $\mathcal{F}$ be the family of data distributions having these densities.

A family $\mathcal{P}$ of distributions for $\boldsymbol{\theta}$ is **conjugate** for $\mathcal{F}$ if

using a prior from $\mathcal{P}$ always produces a posterior from $\mathcal{P}$

(no matter what the observed data values are).

Count Data

Data representing counts take values in $\{0, 1, 2, \ldots\}$.

- If a count Y has a fixed and known maximum possible value n, a commonly-used model is

$$Y \mid \theta \sim \text{Binomial}(n, \theta)$$

  e.g. when, in a sample of size n, Y individuals have a certain characteristic.

- If a count Y has no obvious maximum, a commonly-used model is

$$Y \mid \theta \sim \text{Poisson}(N\theta)$$

  where N is a known scaling factor ("exposure") for rate θ.

Beta-Binomial Conjugacy

For fixed n, the family $\mathcal{F}$ of binomial data distributions

$$Y \mid \theta \quad \sim \quad \mathrm{Binomial}(n, \theta)$$

has the $\mathrm{Beta}(\alpha, \beta)$ family as a conjugate family $\mathcal{P}$ of priors:

As we have seen,

$$\theta \quad \sim \quad \mathrm{Beta}(\alpha, \beta)$$

implies

$$\theta \mid Y = y \quad \sim \quad \mathrm{Beta}(y + \alpha, n - y + \beta)$$

for any $y \in \{0, 1, \ldots, n\}$

prior $\mathrm{Beta}(\alpha, \beta)$ leads to

posterior $\mathrm{Beta}(y + \alpha, n - y + \beta)$

- adding 1 to α is equivalent to adding an extra "success" to the data
- adding 1 to β is equivalent to adding an extra "failure" to the data

$$\begin{aligned} \alpha &= \text{"equivalent prior number of successes"} \\ \beta &= \text{"equivalent prior number of failures"} \end{aligned}$$

(Can we take $\alpha = \beta = 0$? Later.)

Poisson Rate Model

Let

$$Y = \text{number of discrete "incidents" of some kind observed over a known time period of length } N$$

For random incidents that take place independently of each other and at a steady rate θ per unit time,

$$Y \mid \theta \sim \text{Poisson}(N\theta)$$

Want inference about rate θ.

Example: Gravitational Waves

Mergers of black holes and/or neutron stars generate a distinctive gravity wave signal that may be detected by sensitive instruments (LIGO/Virgo).

Let

$$
\begin{aligned}
Y &= \text{number of "confident" detections observed} \\
& \text{ in } N \text{ years} \\
&\sim \mathrm{Poisson}(N\theta) \qquad\qquad (\theta = \text{rate per year})
\end{aligned}
$$

A recent observational run (O3b) lasted about 147 days, or $N = 0.402$ years, and observed $y = 35$.

Naively, we might estimate the rate per year θ to be the number of detections over the period of observation in years:

$$\hat{\theta} = \frac{y}{N} = \frac{35}{0.402} \approx 87$$

Call this the sample rate.

Poisson-Gamma Conjugacy

If

$$Y \mid \theta \;\sim\; \mathrm{Poisson}(N\theta) \qquad \text{(known } N > 0\text{)}$$

then

$$f(y \mid \theta) \;=\; \frac{(N\theta)^y \, \exp(-N\theta)}{y!} \qquad y \;=\; 0, 1, 2, \ldots$$

so the likelihood is

$$\propto \;\; \theta^y \, \exp(-N\theta) \qquad \theta > 0$$

Note: Resembles the kernel of a gamma density (in θ).

Try a gamma prior:

$$\theta \;\sim\; \text{Gamma}(\alpha, \beta)$$

$$\pi(\theta) \;=\; \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)}\,\theta^{\alpha-1}\exp(-\beta\theta) & \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\pi(\theta) \;\propto\; \theta^{\alpha-1}\,\exp(-\beta\theta) \qquad\qquad \theta > 0$$

Applying Bayes' rule,

$$\begin{aligned} p(\theta \mid y) \quad &\propto \quad \theta^{y} \exp(-N\theta) \;\cdot\; \theta^{\alpha-1} \exp(-\beta\theta) \\ &\propto \quad \theta^{y+\alpha-1} \exp\big(-(N+\beta)\theta\big) \qquad\qquad \theta > 0 \end{aligned}$$

which is the kernel of a $\mathrm{Gamma}(y+\alpha, N+\beta)$ density:

$$\theta \mid Y = y \quad \sim \quad \mathrm{Gamma}(y+\alpha, N+\beta)$$

Interpret:

α = "equivalent prior number of incidents"

β = "equivalent prior time period"

Remark: Since $\mathrm{Gamma}(\alpha, \beta)$ has mean α/β,

$$\begin{aligned}
\mathrm{E}(\theta \mid y) &= \frac{y+\alpha}{N+\beta} \\
&= \frac{\beta}{N+\beta} \cdot \frac{\alpha}{\beta} \quad + \quad \frac{N}{N+\beta} \cdot \frac{y}{N} \\
&= (1-w_N) \cdot \mathrm{E}(\theta) \quad + \quad w_N \cdot \hat{\theta}
\end{aligned}$$

where

$$w_N = \frac{N}{N+\beta} \qquad\qquad \hat{\theta} = \frac{y}{N}$$

So the posterior mean is a weighted average of the prior mean and the sample rate.

R Example 2.1:

Poisson Rate — Posterior Inference

Remark: N could represent something other than time.

For example:

- Y is the number of bird nests found on an island, and N is the area of that island.
- Y is the number of people in a country who have a rare disease, and N is the population of that country.

In general, we regard N as an **exposure** variable, perhaps representing an amount of "observational effort."