STA 365: Applied Bayesian Statistics

Boris Babic
Assistant Professor, University of Toronto

Week 4B: Poisson Models

Posterior Distribution

- Let $y_i|\theta \sim \text{Poisson}(\theta)$ and $\theta|\alpha, \beta \sim \text{G}(\alpha, \beta)$.
- Then the posterior distribution is given by:

$$\begin{aligned}
\pi(\theta|y) &\propto \pi(\theta)\pi(y|\theta) \\
&\propto \left(\theta^{\alpha-1} \exp[-\beta\theta]\right) \times \left(\theta^{\sum y_i} \exp[-n\theta]\right) \\
&\propto \theta^{\alpha+\sum y_i - 1} \exp[-(\beta+n)\theta]
\end{aligned}$$

- This is recognizable as a gamma distribution with parameters $\alpha + \sum y_i, \beta + n$.
- Hence, if $\theta \sim \text{G}(\alpha, \beta)$ and $Y_1, ...Y_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ then

$$\theta|Y_1, ...Y_n \sim \text{G}(\alpha + \sum Y_i, \beta + n)$$

Posterior Inference

- Estimation and prediction proceed in a manner similar to that in the binomial model.
- The posterior mean of θ is a convex combination of the prior mean and the sample mean:

$$\mathrm{E}[\theta|y] = \frac{\alpha + \sum y_i}{\beta + n} = \frac{\beta}{\beta + n}\frac{\alpha}{\beta} + \frac{n}{\beta + n}\frac{\sum y_i}{n}$$

- β is interpreted as the number of prior observations,
- α is interpreted as the sum of counts from b prior observations.
- For large n the information from the data dominates the prior information:

$$n >> \beta \rightarrow \mathrm{E}[\theta|y] \approx \overline{y}, \mathrm{Var}(\theta|y) \approx \overline{y}/n$$

Predictive Distribution

- Predictions about additional data can be obtained as before with the posterior predictive distribution, which is given by

$$
\begin{aligned}
\pi(\widetilde{y}|y) &= \int_0^\infty \pi(\widetilde{y}|\theta, y)\pi(\theta|y)d\theta \\
&= \int \pi(\widetilde{y}|\theta)\pi(\theta|y)d\theta \\
&= \int \widetilde{y} \sim \mathrm{Poisson}(\theta) \times \theta \sim \mathrm{G}(\alpha + n\overline{y}, \beta + n)d\theta \\
&= \frac{(\beta+n)^{\alpha+\sum y_i}}{\Gamma(\widetilde{y}+1)\Gamma(\alpha+\sum y_i)} \int_0^\infty \theta^{\alpha+\sum y_i+\widetilde{y}-1} \exp[-(\beta+n+1)\theta]d\theta
\end{aligned}
$$

- This looks daunting but we can simplify it.
- Using what you know about the Gamma density, see if you can simplify it.

Predictive Distribution

- We know that

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta = 1$$

- Hence,

$$\int_0^\infty \theta^{\alpha-1} e^{-\beta\theta} d\theta = \frac{\Gamma(\alpha)}{\beta^\alpha}$$

- Now if we substitute $\alpha + \sum y_i + \widetilde{y}$ for α and $\beta + n + 1$ for β we get:

$$\int_0^\infty \theta^{\alpha+\sum y_i+\widetilde{y}-1} e^{-(\beta+n+1)\theta} d\theta = \frac{\Gamma(\alpha + \sum y_i + \widetilde{y})}{(\beta+n+1)^{\alpha+\sum y_i+\widetilde{y}}}$$

- We can simplify this into a more recognizable form:

$$\pi(\widetilde{y}|y) = \frac{\Gamma(\alpha + \sum y_i + \widetilde{y})}{\Gamma(\widetilde{y}+1)\Gamma(\alpha + \sum y_i)} \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum y_i} \left(\frac{1}{\beta+n+1}\right)^{\widetilde{y}}$$

Predictive Distribution

- This is a negative binomial distribution with parameters $(\alpha + \sum y_i, \beta + n)$, and the posterior mean and variance are given by:

$$\mathrm{E}[\widetilde{Y}|y] = \frac{\alpha + \sum y_i}{\beta + n} = \mathrm{E}[\theta|y]$$

$$\mathrm{Var}(\widetilde{Y}|y) = \frac{\alpha + \sum y_i}{\beta + n} \frac{\beta + n + 1}{\beta + n} = \mathrm{Var}(\theta|y) \times (\beta + n + 1)$$

Example: Birth Rates (Hoff)

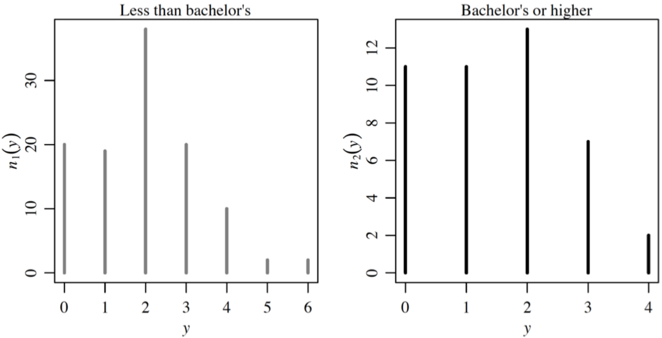
- Over the course of the 1990s the General Social Survey gathered data on the educational attainment and number of children of 155 women who were 40 years of age at the time of their participation in the survey.
- These women were in their 20s during the 1970s, a period of historically low fertility rates in the United States.
- In this example we will examine the difference in the numbers of children between the women without college degrees (group 1) and those with college degrees (group 2).
- Suppose the number of women without college degrees is n_1 and the number of women with college degrees is n_2. For $k = 1, 2$, and $j = 1, ..., n_k$, let $y_{k,j}$ be the number of children of woman j in group k.
- We assume that

  $$y_{k,j}|\theta_k \overset{\text{iid}}{\sim} \text{Poisson}(\theta_k)$$

  And write

  $$y_k = (y_{k,1}, ..., y_{k,n_k})$$

Example: Birth Rates (Hoff)



Posterior Inference

- The group sums and means are as follows:
  - Less than bachelor's: $n_1 = 111, \sum_{j=1}^{n_1} y_{1,j} = 217, \overline{y}_1 = 1.95$
  - Bachelor's or higher: $n_2 = 44, \sum_{j=1}^{n_2} y_{2,j} = 66, \overline{y}_2 = 1.50$
- Suppose we assign the prior: $\theta_1, \theta_2 \overset{\text{iid}}{\sim} \text{G}(2, 1)$.
- What are the posterior distributions, $\theta_1 | y_1$ and $\theta_2 | y_2$?

$$\theta_1 | y_1 \sim \text{G}(219, 112), \qquad \theta_2 | y_2 \sim \text{G}(68, 45)$$

- What are the posterior predictive distributions $\widetilde{y}_{1,*} | y_1, \widetilde{y}_{2,*} | y_2$?

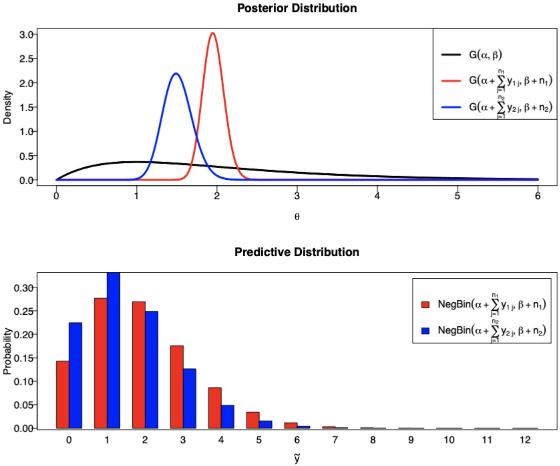
$$\widetilde{y}_{1,*} | y_1 \sim \text{NegBin}(219, 112), \qquad \widetilde{y}_{2,*} | y_2 \sim \text{NegBin}(68, 45)$$

- Conditional on data, are θ_1 and θ_2 independent?

  Yes
- Conditional on data, are $\widetilde{y}_{1,*}$ and $\widetilde{y}_{2,*}$ independent?

  Yes

Example: Birth Rates (Hoff)



Example: Birth Rates (Hoff)

- What are the posterior mean, mode, and 95% credible intervals for θ_1 and θ_2?

Example: Birth Rates (Hoff)

```
a<-2; b<-1 #prior parameters
n1<-111 ; sy1 <-217 # data in group 1
n2<-44 ; sy2<-66 # data in group 2

(a+sy1)/(b+n1) # posterior mean
[1] 1.955357

(a + sy1 -1)/(b+n1) # posterior mode
[1] 1.946429

qgamma( c(.025 ,.975) ,a+sy1 ,b+n1) # posterior 95 percent CI
[1] 1.704943 2.222679

# repeat for second group:

(a+sy2)/(b+n2)
[1] 1.511111

(a + sy2 - 1)/(b + n2)
[1] 1.488889

qgamma( c(.025 ,.975) ,a+sy2 ,b+n2)
[1] 1.173437 1.890836
```

Exponential Family

- Binomial, Poisson and Normal models are all in the exponential family.
- A one-parameter exponential family model is any model whose densities can be expressed as

$$\pi(y|\varphi) = h(y)c(\varphi)\exp[\varphi t(y)],$$

  where φ is the unknown parameter and $t(y)$ is the sufficient statistic.
- The conjugate prior distributions can be expressed in terms of their exponential family representation as:

$$\pi(\varphi|n_0, t_0) = k(n_0, t_0)c(\varphi)^{n_0}\exp[n_0 t_0 \varphi]$$

- Interpretations:
  - n_0: prior sampl size (measure of how informative the prior is)
  - t_0: prior expected value of $t(Y)$.

Binomial Model Representation

- We have:

$$\pi(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-1}$$

- Let

$$\varphi = \log \frac{\theta}{1-\theta}$$

- Then

$$\pi(y|\varphi) = \binom{n}{y} (1+\exp[\varphi])^{-n} \exp[\varphi y]$$

- The conjugate prior for φ is:

$$\pi(\varphi|n_0, t_0) \propto (1+\exp[\varphi])^{-n_0} \exp[n_0 t_0 \varphi]$$

  where t_0 represents the prior expectation of $t(y) = y$.

- What is the prior distribution in terms of θ?

$$\pi(\theta|n_0, t_0) \propto \theta^{n_0 t_0 - 1} (1-\theta)^{n_0(1-t_0)-1}$$

- What is the posterior distribution in terms of θ?

$$\pi(\theta|n_0, t_0, y) \propto \theta^{n_0 t_0 + n\overline{y} - 1} (1-\theta)^{n_0(1-t_0)+n(1-\overline{y})-1}$$

Poisson Model Representation

- The $\mathrm{Poisson}(\theta)$ model can be shown to be an exponential family model with

$$\begin{aligned} t(y) &= y \\ \varphi &= \log(\theta) \\ c(\varphi) &= \exp\left[\,\exp[-\varphi]\right] \end{aligned}$$

- The conjugate prior for φ is given by

$$\pi(\varphi|n_0, t_0) = \exp\left[n_0 \exp[-\varphi]\right] \exp[n_0 t_0 \varphi]$$

- What is the prior distribution in terms of θ?

$$\pi(\theta|n_0, t_0) = \theta^{n_0 t_0 - 1} \exp[-n_0 \theta]$$