

Stat 343 Bayes Practice with Conjugate Priors

Let's return to a couple of previous examples where we performed maximum likelihood estimation, and see how a Bayesian analysis looks.

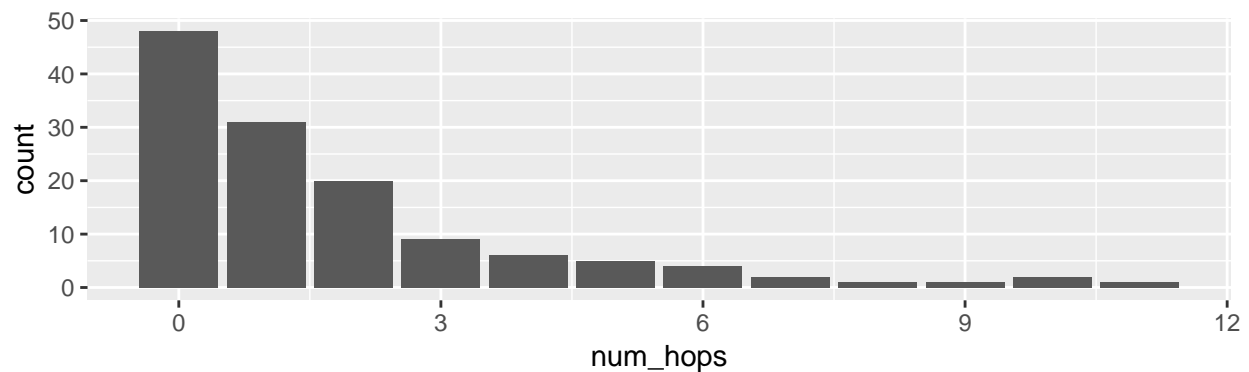
Example 1: Bird Hops

In an ecological study of the feeding behavior of birds, researchers counted the number of hops each of several birds took between flights. The R code below reads in the data and makes an initial plot:

```
bird_hops <- read_csv("http://www.evanlray.com/data/rice/Chapter%208/hops.csv") %>%
  transmute(
    num_hops = num_hops - 1
  )
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
ggplot(data = bird_hops, mapping = aes(x = num_hops)) +
  geom_bar()
```



```
bird_hops %>%
  count(num_hops)
```

```
## # A tibble: 12 x 2
##   num_hops     n
##   <dbl> <int>
## 1         0    48
## 2         1    31
## 3         2    20
## 4         3     9
## 5         4     6
## 6         5     5
## 7         6     4
## 8         7     2
## 9         8     1
## 10        9     1
## 11        10     2
## 12        11     1
```

After subtracting 1 from the originally reported counts (as is done in the code above), the number of hops is 0 if the bird took off directly (0 hops before successful flight), 1 if the bird took one hop before taking off, and so on. Let's model these data with a Geometric(θ) distribution.

Note that there are multiple parameterizations of the geometric distribution in common use. For this problem, please use the parameterization given in the “probability distributions” handout.

You saw in PS3 that the joint pmf of X_1, \dots, X_n is

$$\begin{aligned} f(x_1, \dots, x_n | \theta) &= \prod_{i=1}^n f(x_i | \theta) \\ &= \prod_{i=1}^n (1 - \theta)^{x_i} \theta \\ &= (1 - \theta)^{\sum_{i=1}^n x_i} \theta^n \end{aligned}$$

Show that in a Bayesian analysis, a $\text{Beta}(\alpha, \beta)$ prior distribution is a conjugate prior for θ . What are the parameters of the posterior distribution?

From the probability distributions handout, if $W \sim \text{Beta}(\alpha, \beta)$ then the pdf of W is given by

$$f(w | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1} (1 - w)^{\beta-1},$$

on the support $0 \leq w \leq 1$.

Note: this derivation is extremely similar to the derivation we gave for a Binomial model last class. You should be able to follow along with the steps we took in that derivation quite closely, but you will now need to condition on the full data set x_1, \dots, x_n rather than on a single observed value x as in the Binomial example.

You should carefully state the prior distribution pdf, use Bayes’ Rule to derive the form of the posterior, and “recognize” the posterior as a Beta distribution without explicitly calculating any integrals.

If we use the prior $\Theta \sim \text{Beta}(\alpha, \beta)$, then Θ has pdf

$$f_{\Theta}(\theta | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

The posterior distribution for θ has pdf

$$\begin{aligned} f_{\Theta | X_1, \dots, X_n}(\theta | x_1, \dots, x_n, \alpha, \beta) &= \frac{f_{\Theta, X_1, \dots, X_n}(\theta, x_1, \dots, x_n, \alpha, \beta)}{f_{X_1, \dots, X_n}(x_1, \dots, x_n | \alpha, \beta)} \\ &= \frac{f_{\Theta}(\theta | \alpha, \beta) f_{X_1, \dots, X_n | \Theta}(x_1, \dots, x_n)}{\int_0^1 f_{\Theta}(\theta | \alpha, \beta) f_{X_1, \dots, X_n | \Theta}(x_1, \dots, x_n) d\theta} \\ &= \frac{\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} (1 - \theta)^{\sum_{i=1}^n x_i} \theta^n}{\int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} (1 - \theta)^{\sum_{i=1}^n x_i} \theta^n d\theta} \\ &\propto \theta^{\alpha + n - 1} (1 - \theta)^{\beta + \sum_{i=1}^n x_i - 1} \end{aligned}$$

As a function of θ , this is proportional to the pdf of a $\text{Beta}(\alpha + n, \beta + \sum_{i=1}^n x_i)$ distribution. Therefore, the posterior distribution of Θ is

$$\Theta | \alpha, \beta, x_1, \dots, x_n \sim \text{Beta}(\alpha + n, \beta + \sum_{i=1}^n x_i)$$

Example 2: Saplings

This example comes from “Introduction to Statistical Thought,” by Michael Lavine. Lavine writes:

Tree populations move by dispersing their seeds. Seeds become seedlings, seedlings become saplings, and saplings become adults which eventually produce more seeds. Over time, whole populations may migrate in response to climate change. One instance occurred at the end of the Ice Age when species that had been sequestered in the south were free to move north. Another instance may be occurring today in response to global warming. One critical feature of the migration is its speed. Some of the factors determining the speed are the typical distances of long range seed dispersal, the proportion of seeds that germinate and emerge from the forest floor to become seedlings, and the proportion of seedlings that survive each year. To learn about emergence and survival, ecologists return annually to forest quadrats (square meter sites) to count seedlings that have emerged since the previous year. One such study was reported in Lavine et al. [2002].

In each year from 1991 to 1997, the ecologists recorded the number of “old” seedlings in each of 60 quadrats, where an old seedling is at least 1 year old (they can identify old seedlings by whether they have a bud mark). These values are recorded in the columns of the data frame named like `91_old`. For the years 1992 through 1997, they also recorded the number of “new” seedlings, i.e. the number of seedlings that were less than 1 year old. The number of new seedlings and the number of old seedlings in each quadrat don’t add up like you might expect. This might be due to a variety of factors, like seedlings dying or errors in data collection.

```
library(tidyverse)

seedlings <- read_table("http://www.evanlray.com/data/lavine_intro_stat_thought/seedlings.txt") %>%
  select(quadrat = Block,
         old_1991 = `91`,
         old_1992 = `92-t`,
         old_1993 = `93-t`,
         old_1994 = `94-t`,
         old_1995 = `95-t`,
         old_1996 = `96-t`,
         old_1997 = `97-t`,
         new_1992 = `92-1`,
         new_1993 = `93-1`,
         new_1994 = `94-1`,
         new_1995 = `95-1`,
         new_1996 = `96-1`,
         new_1997 = `97-1`
  )

head(seedlings)

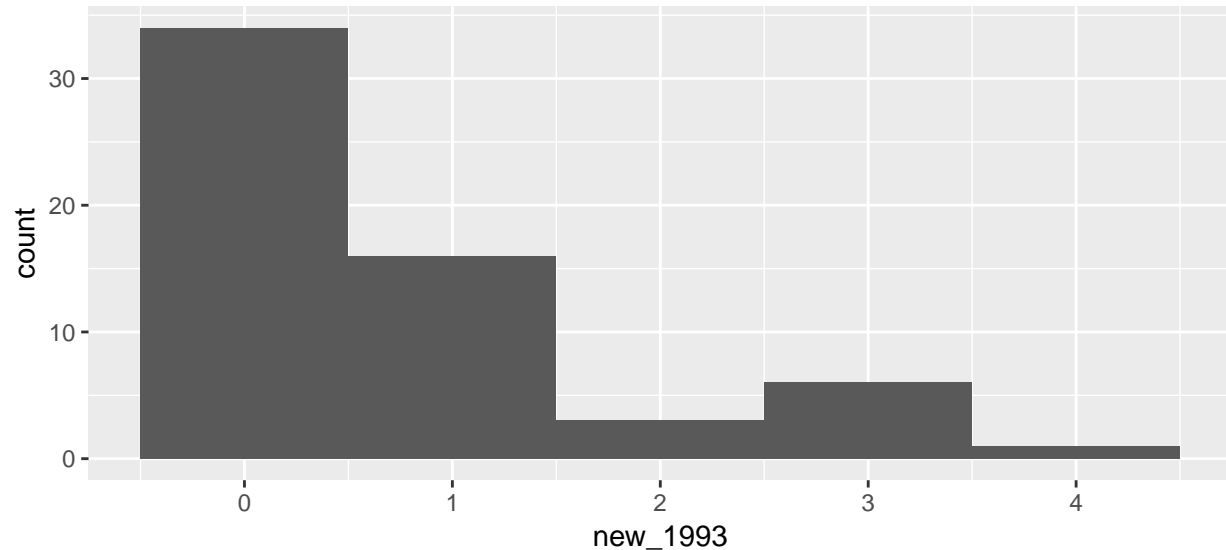
## # A tibble: 6 x 14
##   quadrat old_1991 old_1992 old_1993 old_1994 old_1995 old_1996 old_1997
##   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     1         1         1         1         1         2         1         1
## 2     2         2         1         1         1         0         1         0
## 3     3         3         1         1         0         0         1         0
## 4     4         4         0         0         0         0         0         0
## 5     5         5         1         1         2         1         1         1
## 6     6         6         1         0         1         0         0         0
## # ... with 6 more variables: new_1992 <dbl>, new_1993 <dbl>,
## #   new_1994 <dbl>, new_1995 <dbl>, new_1996 <dbl>, new_1997 <dbl>
```

```
dim(seedlings)
```

```
## [1] 60 14
```

For today, let's analyze the number of new seedlings observed in each of the 60 quadrats during 1993. Here's a plot:

```
ggplot(data = seedlings, mapping = aes(x = new_1993)) +  
  geom_histogram(binwidth = 1)
```



Define the random variables X_i , $i = 1, \dots, 60$ to represent the number of new seedlings observed in each quadrat i in 1993.

We will use the model $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$. Our goal is to estimate the parameter λ , which describes the rate at which seedlings emerge in a quadrat.

We have previously seen that with this model, the pmf for a single X_i is

$$f(x_i|\lambda) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}.$$

We also saw that the joint pmf for X_1, \dots, X_n is

$$\begin{aligned} f(x_1, \dots, x_n|\lambda) &= \prod_{i=1}^n f(x_i|\lambda) \\ &= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\ &= e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!} \end{aligned}$$

Show that in a Bayesian analysis, a $\text{Gamma}(\alpha, \beta)$ prior distribution is a conjugate prior for the parameter λ in the Poisson model. What are the parameters of the posterior distribution?

From the probability distributions handout, if $W \sim \text{Gamma}(\alpha, \beta)$ then the pdf of W is $f(w|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} e^{-\beta w}$

You should carefully state the prior distribution pdf, use Bayes' Rule to derive the form of the posterior, and "recognize" the posterior as a Gamma distribution without explicitly calculating any integrals.

If we use a $\text{Gamma}(\alpha, \beta)$ prior for Λ , the pdf of Λ is

$$f_\Lambda(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

The posterior distribution for Λ is

$$\begin{aligned} f_{\Lambda|X_1, \dots, X_n}(\lambda|x_1, \dots, x_n, \alpha, \beta) &= \frac{f_{\Lambda, X_1, \dots, X_n}(\lambda, x_1, \dots, x_n|\alpha, \beta)}{f_{X_1, \dots, X_n}(x_1, \dots, x_n|\alpha, \beta)} \\ &= \frac{f_\Lambda(\lambda|\alpha, \beta) f_{X_1, \dots, X_n|\Lambda}(x_1, \dots, x_n|\alpha, \beta)}{\int f_\Lambda(\lambda|\alpha, \beta) f_{X_1, \dots, X_n|\Lambda}(x_1, \dots, x_n|\alpha, \beta) d\lambda} \\ &= \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!}}{\int \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!} d\lambda} \\ &\propto \lambda^{\alpha-1} e^{-\beta\lambda} e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \\ &= \lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-(\beta+n)\lambda} \end{aligned}$$

As a function of λ , this is proportional to the pdf of a $\text{Gamma}(\alpha + \sum_{i=1}^n x_i, \beta + n)$ distribution. Therefore, the posterior distribution of Λ is

$$\Lambda|\alpha, \beta, x_1, \dots, x_n \sim \text{Gamma}(\alpha + \sum_{i=1}^n x_i, \beta + n)$$