Stat 136 (Bayesian Statistics)

Lesson 1.4: Bayesian Inference for the Poisson Mean

Review: The Poisson Distribution

Just like the binomial distribution, the Poisson distribution is used to model count data. It is used to model the number of occurrences of rare events which are occurring randomly through time (or space) at a constant rate. For example, the Poisson distribution could be used to model the number of accidents on a highway over a month, or the number of COVID-19 patients arriving at an ER every 1-hour interval.If a random variable Y has a Poisson distribution with mean or *rate parameter* λ, then the PMF of Y is

$$P(y = y) = \frac{\lambda^y e^{-\lambda}}{y!}, y = 0, 1, 2, \ldots$$

If $Y \sim Poisson(\lambda)$ then $E(Y) = V(Y) = \lambda$.

Bayesian Inference for the Poisson Mean Using Discrete Prior

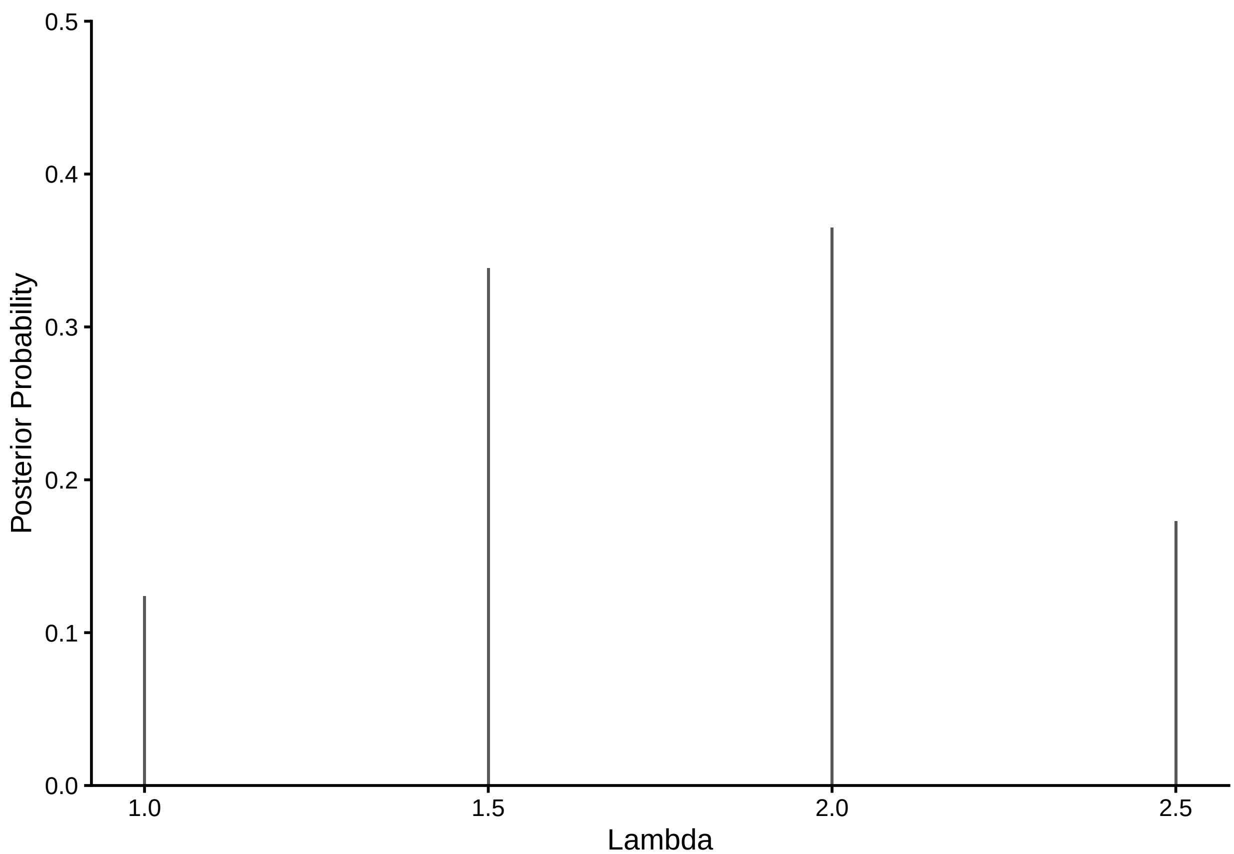
We use a similar approach with Bayesian inference for p using discrete prior. For example, let Y_i be distributed as $Poisson(\lambda)$. Suppose that we believe there are only four possible values for λ: 1.0, 1.5, 2.0, and 2.5. Further, suppose that the two middle values, 1.5 and 2.0, are twice as likely as the two end values 1.0 and 2.5. The observed count is $Y = 2$.

```
l <- c(1.0, 1.5, 2.0, 2.5)
pr <- c(1/6, 1/3, 1/3, 1/6)
lik <- (l^2*exp(-l)/factorial(2))
pl <- pr*lik
post <- pl/sum(pl)
bbox <- as.data.frame(cbind(l,pr,lik,pl,post))
knitr::kable(bbox)
```

l	pr	lik	pl	post
1.0	0.1666667	0.1839397	0.0306566	0.1239620
1.5	0.3333333	0.2510214	0.0836738	0.3383404
2.0	0.3333333	0.2706706	0.0902235	0.3648246

l	pr	lik	pl	post
2.5	0.1666667	0.2565156	0.0427526	0.1728729

```
library(tidyverse)
bbox %>%
  ggplot(aes(x = l, y = post)) +
  geom_bar(stat = "identity", width = .004) +
  scale_y_continuous(expand=c(0,0), limit = c(0, 0.5)) +
  labs(x = "Lambda",
       y = "Posterior Probability")+
  theme_classic()
```



Question: What is the mean and variance of the posterior distribution of λ?

```
Mean <-  sum(l*post)
Variance <- sum(l^2*post) - (sum(l*post))^2
print(cbind(Mean, Variance))
```

```
##          Mean  Variance
## [1,] 1.793304 0.2090422
```

Question: What is $P(\lambda < 2)$?

```
bbox %>%
  filter(l<2) %>%
  select(post) %>%
  sum()
```

```
## [1] 0.4623025
```

Bayesian Inference for the Poisson Mean Using Continuous Prior

Just like before we can do this in three steps.

1. Construct a prior expressing an opinion about the location of the rate λ before any data is collected
2. Take the sample of intervals and records the number of arrivals in each interval. From this data, we form the likelihood, the probability of these observations expressed as a function of λ
3. Use the Bayes' rule to compute the posterior – this distribution updates the prior opinion about λ given the information from the data

In addition, we compute the predictive distribution to learn about the number of arrivals in future intervals. The posterior predictive distribution is also useful in checking the appropriateness of our model.

Before we identify prior distributions for making Bayesian inference on λ, first, let us look at the likelihood function of λ:

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \approx \lambda^u e^{-n\lambda}, \text{ where: } u = \sum_{i=1}^{n} y_i$$

This likelihood resembles the kernel of a *Gamma* (α,β) density with $\alpha = \sum_{i=1}^{n} y_i + 1$ and $\beta = n$.

So, we begin by constructing a prior density to express one's opinion about the rate parameter λ. Since the rate is a positive continuous parameter, one needs to construct a prior density that places its support only on positive values. If we have no idea what the value of λ is prior to looking at the data, then we can consider the **positive uniform** prior density given by

$$f(\lambda) = \begin{cases} 1, \text{ if } \lambda > 0 \\ 0, \text{elsewhere} \end{cases}$$

Note that this prior density is improper since its integral over all possible values is infinite. Using this prior density the resulting posterior density will be identical to the likelihood.

Another possible prior for λ is the **Jeffrey's** prior given by

$$f(\lambda) = \begin{cases} \frac{1}{\sqrt{\lambda}}, & \text{if } \lambda > 0 \\ 0, \text{elsewhere} \end{cases}$$

This also an improper prior, but informative since it gives more weight to small values of λ. Using this prior density, the posterior density becomes

$$f(\lambda|y) \approx \lambda^{u-\frac{1}{2}} e^{-n\lambda}, \text{ where: } u = \sum_{i=1}^{n} y_i$$

Observe that this likelihood resembles the kernel of a *Gamma* (α,β) density with $\alpha = \sum_{i=1}^{n} y_i + \frac{1}{2}$ and $\beta = n$. This shows that the convenient choice (**conjugate**) of prior distributions for Poisson sampling is the Gamma distribution with PDF.

$$f(\lambda|\alpha,\beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, & \text{if } \lambda > 0, \text{ and } \alpha, \beta > 0 \\ 0, & \text{elsewhere} \end{cases}$$

The Gamma density is a continuous density where the support is on positive values. It depends on two parameters, a positive shape parameter α and a positive rate parameter β. The Gamma density is a flexible family of distributions that can reflect many different types of prior beliefs about the location of the parameter λ.

One chooses values of the shape α and the rate β so that the Gamma density matches one's prior information about the location of λ. In R, the function *dgamma()* gives the density, *pgamma()* gives the distribution function, and *qgamma()* gives the quantile function of the Gamma distribution. These functions are helpful in graphing the prior and choosing values of the shape and rate parameters that match prior statements about Gamma percentiles and probabilities.

After plugging in the prior density and the likelihood into the Bayes' formula, it can be shown (verify!) that the posterior distribution for λ is

$$f(\lambda|y) \approx \lambda^{\sum_{i=1}^{n} y_i + \alpha - 1} e^{-\lambda(\beta+n)}$$

Once again the posterior resembles the kernel of the Gamma(α', β') density, where:

$$\alpha' = \alpha + \sum_{i=1}^{n} y_i$$

and

$$\beta' = \beta + n$$

Question: What would be the form of the *Gamma(a, b)* prior? That is, what should α and β be?

One way to answer these questions is to plot your prior to check if looks reasonably close to your prior belief. Or, summarize your prior knowledge in terms of the mean m and standard deviation s (or variance s^2).

Recall that the mean and variance of Gamma(α, β) are $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$, respectively. We then solve the following system of equations for α and β:

$$\overline{y} = \frac{\alpha}{\beta}$$

and

$$s^2 = \frac{\alpha}{\beta^2}$$

An Example

The weekly number of traffic accidents on a highway has the Poisson(λ) distribution. Three students are going to count the number of traffic accidents for each of the next eight weeks. They are going to analyze the data in a Bayesian manner, so they each need a prior distribution. The number of accidents on the highway over the next 8 weeks are: 3, 2, 0, 8, 2, 4, 6, 1.

Students A and B tried the positive uniform prior and the Jeffrey's prior, respectively. While, Student C believes that, on average, λ will be close to 2.5 with standard deviation of 1.

Solving the system of equations outlined earlier the Student C comes up with a Gamma(6.25,2.5) prior.

The prior distributions of the 3 students are summarized below.

Student	Prior	Prior density
A	Positive uniform	1
B	Jeffrey's	$\frac{1}{\sqrt{\lambda}}$
C	$\overline{Y} = 2.5$; $s = 1$	Gamma(6.25, 2.5)

Based on the data we have $\sum_{i=1}^{8} = 26$, thus, the likelihoood is

$$L(\lambda|y) \approx \lambda^{26} e^{-8\lambda}$$

Thus, after using the Bayes formula, the posterior distributions are:

Student	Prior	Prior density	Posterior
A	Positive uniform	1	Gamma(27, 8)
B	Jeffrey's	$\frac{1}{\sqrt{\lambda}}$	Gamma(26.5, 8)
C	$\overline{Y} = 2.5$; $s = 1$	Gamma(6.25, 2.5)	Gamma(32.25, 10.5)

1. **Summary statistics**

Posterior	Mean	Median	SD
Gamma(27, 8)	3.375	3.333	0.6495
Gamma(26.5, 8)	3.313	3.271	0.6435
Gamma(32.25, 10.5)	3.071	3.040	0.5408

The means and standard deviations were computed using the formulas $\frac{\alpha'}{\beta'}$ and $\frac{\alpha'}{\beta'^2}$, respectively. While, the medians were computed using the *qgamma()* function

```
mdA <- qgamma(0.5,27,8)
mdB <- qgamma(0.5,26.5,8)
mdC <- qgamma(0.5,32.25,10.5)
print(cbind(mdA,mdB,mdC))
```

```
##           mdA      mdB      mdC
## [1,] 3.333426 3.270928 3.039742
```

We can also simulate observations from each posterior distribution using the *rgamma()* function and use these simulated observations to compute summary statistics.

```
g1 <- rgamma(10000,27,8)
g2 <- rgamma(10000,26.5,8)
g3 <- rgamma(10000,32.25,10.5)
sim.gamma <- as.data.frame(cbind(g1, g2, g3))
sim.gamma %>%
  summarise_all(list(m=mean, sd=sd))
```

```
##       g1_m     g2_m    g3_m     g1_sd     g2_sd     g3_sd
## 1 3.361198 3.308366 3.07201 0.6496109 0.6467711 0.5493279
```

2. **Credible intervals**

Using the *qgamma()* function we can obtain credible intervals for λ based on its posterior distribution. For example, the 95% credible interval for λ based on *Gamma(27,8)* is given by

```r
ci1L <- qgamma(0.025,27,8)
ci1U <- qgamma(0.975,27,8)
print(cbind(ci1L,ci1U))
```

```
##          ci1L     ci1U
## [1,] 2.224146 4.762003
```

The 95% credible intervals for λ for each of the posterior distributions are given in the following table:

Posterior	95% Credible Interval
Gamma(27, 8)	(2.224, 4.762)
Gamma(26.5, 8)	(2.174, 4.688)
Gamma(32.25, 10.5)	(2.104, 4.219)

3. **Test of hypothesis**

In testing one-side hypothesis, we compute the probability of observing the event in the null hypothesis based on the posterior distribution.

- *If the posterior probability of the null hypothesis is less than α , then we reject the null hypothesis at the α level of significance*

In testing two-sided hypothesis, we construct the (1-α)% credible interval

- *If the value of λ under the null hypothesis lies outside the credible interval, reject H_0; else, we do not reject the null hypothesis and conclude λ remains a credible value*

Suppose we want to test $H_0 : \lambda \leq 3$ versus $H_1 : \lambda > 3$

```r
p1 <- pgamma(3,27,8)
p2 <- pgamma(3,26.5,8)
p3 <- pgamma(3,32.25,10.5)
print(cbind(p1,p2,p3))
```

```
##             p1        p2        p3
## [1,] 0.2961806 0.3312435 0.4704052
```

Posterior	$P(\lambda \leq 3)$
Gamma(27, 8)	0.2962
Gamma(26.5, 8)	0.3312
Gamma(32.25, 10.5)	0.4704

In all cases, H_0 is not rejected.

4. **Prediction**

Suppose we wanted to predict the number of car accidents. We simulate λ's from the posterior distribution and use these as input to the Poisson model and simulate Poisson distributed observations . Then, we compute the mean of the predictive distribution.

Suppose we predict an observation from the posterior distribution of Student C

```
lambda <- rgamma(10000,32.25,10.5)
pois <- rpois(10000,lambda)
print(mean(pois))
```

```
## [1] 3.0975
```

Therefore, We expect about 3 accidents to occur.