Gamma-Poisson Inference, Part 2

STAT 251, Unit 5B

Overview

Review

Derivation of the Posterior Distribution in Gamma-Poisson Setting

Examples of Gamma-Poisson Inference

Inference with multiple observations, $y_1, y_2, \ldots, y_n$

Review

Recap from Gamma-Poisson, Part 1

We focused on the Poisson and Gamma families.

- Poisson distribution is a natural candidate for modeling count data with no upper bound (and which do not consist of a specific number of trials).
- Gamma distribution is a natural candidate for the prior distribution of the Poisson's rate parameter.
- Two forms for writing the probability density function of the gamma distribution
  - Gamma(shape=a, rate=b)
  - Gamma(shape=a, scale=β)
- Mean, variance of Poisson and Gamma distributions

In Gamma-Poisson, Part 2

- Analytical derivation of the posterior distribution of θ when $Y|\theta$ has the *Poisson*(θ) distribution and θ has the Gamma(a,b) distribution (the so-called *Gamma-Poisson* setting).
- Practice with posterior inference in the Gamma-Poisson setting.
- Discuss inference if multiple y measurements are made.

Derivation of the Posterior Distribution in Gamma-Poisson Setting

Derivation of Posterior

Assume the so-called *Gamma-Poisson* setting. That is:

- $Y|\theta$ has the *Poisson*(θ) distribution
- θ has the Gamma(shape=a, rate=b) prior distribution
- we observe a single observation, y

In this specific setting, the posterior distribution of $\theta|y$ is the Gamma(shape=a+y, rate=b+1) distribution.

Gamma-Poisson Posterior Derivation, by definition (part 1)

$$\begin{aligned}
\pi(\theta|y) = & \quad \frac{\pi(\theta)f(y|\theta)}{f(y)} \\
= & \quad \frac{\pi(\theta)f(y|\theta)}{\int \pi(\theta)f(y|\theta)d\theta} \\
= & \frac{\frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}\mathbb{1}_{\theta>0}\frac{\theta^y e^{-\theta}}{y!}\mathbb{1}_{y\in\{0,1,2,\ldots\}}}{\int \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}\mathbb{1}_{\theta>0}\frac{\theta^y e^{-\theta}}{y!}\mathbb{1}_{y\in\{0,1,2,\ldots\}}d\theta} \\
= & \frac{\frac{b^a}{\Gamma(a)y!}\theta^{a+y-1}e^{-(b+1)\theta}\mathbb{1}_{\theta>0}}{\frac{b^a}{\Gamma(a)y!}\int \theta^{a+y-1}e^{-(b+1)\theta}\mathbb{1}_{\theta>0}d\theta} \\
= & \frac{\theta^{a+y-1}e^{-(b+1)\theta}\mathbb{1}_{\theta>0}}{\int \theta^{a+y-1}e^{-(b+1)\theta}\mathbb{1}_{\theta>0}d\theta} \\
= & \frac{\theta^{a+y-1}e^{-(b+1)\theta}\mathbb{1}_{\theta>0}}{\int_0^\infty \theta^{a+y-1}e^{-(b+1)\theta}d\theta}
\end{aligned}$$

Continued on next slide:

Gamma-Poisson Posterior Derivation, by definition (part 2)

Continued from previous slide:

$$
\begin{aligned}
\pi(\theta|y) =& \frac{\theta^{a+y-1} e^{-(b+1)\theta} \mathbb{1}_{\theta>0}}{\int_0^\infty \theta^{a+y-1} e^{-(b+1)\theta} d\theta} \\
=& \frac{\frac{(b+1)^{a+y}}{\Gamma(a+y)} \theta^{a+y-1} e^{-(b+1)\theta} \mathbb{1}_{\theta>0}}{\underbrace{\int_0^\infty \frac{(b+1)^{a+y}}{\Gamma(a+y)} \theta^{a+y-1} e^{-(b+1)\theta} d\theta}_{=1}} \\
=& \frac{(b+1)^{a+y}}{\Gamma(a+y)} \theta^{a+y-1} e^{-(b+1)\theta} \mathbb{1}_{\theta>0} \\
=& \text{pdf of the } Gamma(a^\star = a+y, b^\star = b+1) \text{ distribution}
\end{aligned}
$$

The posterior distribution after observing a single y is $\theta|y \sim Gamma(a^\star = a+y, b^\star = b+1)$.

Analytical Derivation, using a shortcut

$$
\begin{aligned}
\pi(\theta|y) =& \frac{\pi(\theta)f(y|\theta)}{f(y)} \\
&\propto \pi(\theta)f(y|\theta) \\
&= \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}\mathbb{1}_{\theta>0}\frac{\theta^y e^{-\theta}}{y!}\mathbb{1}_{y\in\{0,1,2,\ldots\}} \\
&\propto \quad \underbrace{\theta^{a+y-1}e^{-(b+1)\theta}\mathbb{1}_{\theta>0}}_{\text{kernel of Gamma(a+y, b+1) pdf}} \\
&\propto \frac{(b+1)^{a+y}}{\Gamma(a+y)}\theta^{a+y-1}e^{-(b+1)\theta}\mathbb{1}_{\theta>0} \\
&= \text{pdf of } Gamma(a^\star = a+y, b^\star = b+1) \text{ dist.}
\end{aligned}
$$

Although the above equations imply the posterior is *proportional* to the Gamma($a^\star, b^\star$)*distribution*, why are we able to conclude that the posterior (i.e., $\pi(\theta|y)$) *IS* the $Gamma(a^\star, b^\star)$ distribution?

The *kernel* of a pmf/pdf is the portion that depends explicitly on the variable.

Examples of Gamma-Poisson Inference

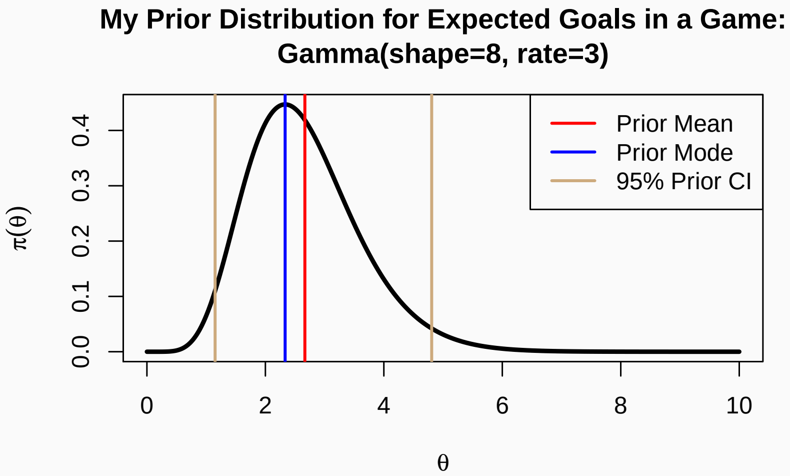
Soccer Goals Scored

Suppose that we are considering the scoring ability of the 2016 BYU women's soccer team. Specifically, we are interested in $\theta =$ expected number of goals scored in a randomly selected game.

I will assume that given θ, Y= number of goals scored in a game has the *Poisson*(θ) distribution.

My prior distribution is the Gamma(shape=8, rate=3) distribution.

Plot of my Prior, with selected Prior Summaries



Prior Mean: $a/b = 8/3$; Prior Mode: $(a\text{-}1)/b = 7/3$; 95% Prior Credible Interval (from 2.5th, 97.5th percentiles): (1.15, 4.81).

Posterior (after 1 game)

In first non-exhibition game, BYU won 2-1 over Washington State.

With this match result, what is my posterior distribution?

Posterior (after 1 game)

In first non-exhibition game, BYU won 2-1 over Washington State.
With this match result, what is my posterior distribution?

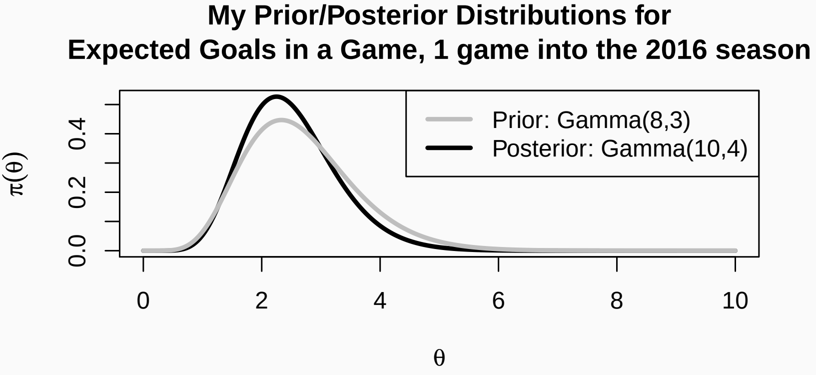
I assumed $Y|\theta \sim \textit{Poisson}(\theta)$ and $\theta \sim \textit{Gamma}(a = 8, b = 3)$; observed $y = 2$.

Posterior (after 1 game)

In first non-exhibition game, BYU won 2-1 over Washington State. With this match result, what is my posterior distribution?

I assumed $Y|\theta \sim Poisson(\theta)$ and $\theta \sim Gamma(a = 8, b = 3)$; observed $y = 2$.

Posterior: $\theta|y \sim Gamma(a^\star = a + y = 10, b^\star = b + 1 = 4)$.



Posterior (after 2nd game)

In the next game, BYU lost 0-1 to Nebraska. With this match result, what is my posterior distribution?

Posterior (after 2nd game)

In the next game, BYU lost 0-1 to Nebraska. With this match result, what is my posterior distribution?

I assumed $Y|\theta \sim Poisson(\theta)$ and my updated prior is $\theta \sim Gamma(a = 10, b = 4)$; observed $y = 0$.

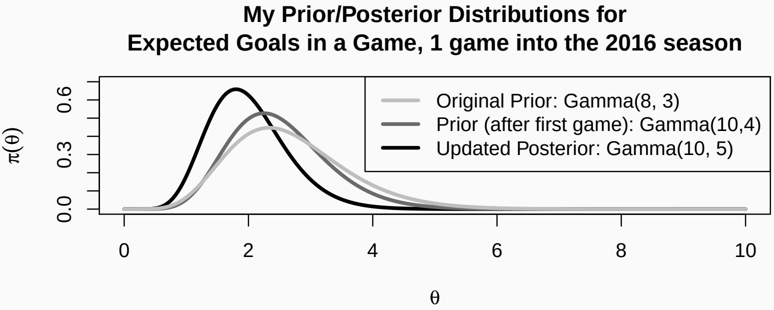
Posterior (after 2nd game)

In the next game, BYU lost 0-1 to Nebraska. With this match result, what is my posterior distribution?

I assumed $Y|\theta \sim Poisson(\theta)$ and my updated prior is $\theta \sim Gamma(a = 10, b = 4)$; observed $y = 0$.

Updated Posterior:

$\theta|y \sim Gamma(a^\star = a + y = 10, b^\star = b + 1 = 5)$.



Prior Elicitation for Gamma-Poisson Setting

In the setting where $Y|\theta \sim Poisson(\theta)$ and $\theta \sim Gamma(shape = a, rate = b)$:

- If we have little idea of what θ might be, it is common to choose both a and b to be very small (e.g., a=1, b=1 or even a=.01, b=.01)
- If we can choose the prior mean, $E(\theta) = \mu$, and the prior standard deviation, $SD(\theta) = \sigma$, then we can solve for a and b using the fact that for a Gamma random variable, the mean is $\mu = \frac{a}{b}$ and the standard deviation is $\sigma = \frac{\sqrt{a}}{b}$.

Another alternative for Prior Elicitation

As with the beta prior distribution, we can choose a and b by considering how our prior beliefs would equate to an equivalent number of observations. To use such a method for the gamma distribution, we interpret b as the total number of prior units "observed", and a as the total number of prior occurrences "observed" in the prior units.

For example, suppose I equate my knowledge about the scoring ability of BYU's 2017 women's soccer team to be about six games' worth of information, with about 17 total goals scored. Then I could use the $gamma(a = 17, b = 6)$ prior distribution for θ.

I find this approach difficult to implement because it is often abstract.

Inference with multiple observations, $y_1, y_2, \ldots, y_n$

What to do with multiple observations?

We have placed greatest emphasis on observing a single random variable y.

Often, multiple observations are made. That is, we have $y_1, y_2, \ldots, y_n$, where n represents the number of observed y's.

This will produce only a slight change in the implementation of our Bayesian procedure. As before, we note that

$posterior = \frac{prior \times likelihood}{\text{marginal likelihood}}$, and $posterior \propto prior \times likelihood$. The change is in how we compute the

likelihood—$f(y_1, y_2, \ldots, y_n|\theta)$—and consequently also the marginal likelihood—$f(y_1, y_2, \ldots, y_n)$.

Determining $f(y_1, y_2, \ldots, y_n|\theta)$: Conditional Independence

In STAT 251, we will assume that observations are *conditionally independent and identically distributed.* What does this mean? First consider conditional independence:

> The random variables $Y_1|\theta, Y_2|\theta, \ldots, Y_n|\theta$ are said to be *conditionally independent* if, conditional on θ, each Y_i is mutually independent of the others.

If $Y_1|\theta, Y_2|\theta, \ldots, Y_n|\theta$ are conditionally independent, then

$$\begin{aligned} f(\{y_1, y_2, \ldots, y_n\}|\theta) =& f_{y_1|\theta}(y_1|\theta) f_{y_2|\theta}(y_2|\theta) \cdots f_{y_n|\theta}(y_n|\theta) \\ =& \prod_{i=1}^{n} f_{y_i|\theta}(y_i|\theta) \end{aligned}$$

The subscripts on the pdf/pmf, (e.g., $f_{y_1|\theta}$ and $f_{y_2|\theta}$) are to allow for the possibility that $Y_i|\theta$ and $Y_j|\theta$ have different marginal distributions.

Determining $f(y_1, y_2, \ldots, y_n|\theta)$: Conditionally Identically Distributed

The random variables $Y_1|\theta, Y_2|\theta, \ldots, Y_n|\theta$ are said to be *conditionally identically distributed* if, conditional on θ, each Y_i has the exact same distribution.

That is, $f_{y_1|\theta}(y|\theta) = f_{y_2|\theta}(y|\theta) = \cdots = f_{y_n|\theta}(y|\theta)$ for all y

Because identically distributed random variables share the same pmf/pdf, then there is no need to have a separate subscript for each pmf/pdf. That is, instead of writing $f_{y_1|\theta}$ or $f_{y_2|\theta}$, we can simply write $f_{y|\theta}$ or even just f.

Determining $f(y_1, y_2, \ldots, y_n|\theta)$: Conditionally Independent and Identically Distributed

$Y_1|\theta, Y_2|\theta, \ldots, Y_n|\theta$ are said to be *conditionally independent and identically distributed* if they are conditionally independent and they are conditionally identically distributed.

> Conditionally independent and identically distributed is abbreviated as conditionally **iid**.

If $Y_1|\theta, Y_2|\theta, \ldots, Y_n|\theta$ are conditionally *iid* then

$$f(\{y_1, y_2, \ldots, y_n\}|\theta) = \prod_{i=1}^{n} f(y_i|\theta).$$

Likelihood of n conditionally *iid* Poisson observations

Suppose that

$$y_i|\theta \overset{iid}{\sim} Poisson(\theta),$$

so that $f(y_i|\theta) = \frac{\exp(-\theta)\theta^{y_i}}{y_i!}$ for all i.

Note the notation above to state that the Y_i's are (conditionally) independent and identically distributed: $\overset{iid}{\sim}$

Then the likelihood is

$$\begin{aligned} f(y_1, y_2, \ldots, y_n|\theta) &= \prod_{i=1}^{n} f(y_i|\theta) \\ &= \prod_{i=1}^{n} \frac{\exp(-\theta)\theta^{y_i}}{y_i!} \\ &= \frac{exp(-n\theta)\theta^{\sum y_i}}{\prod_{i=1}^{n} y_i!} \end{aligned}$$

Technical Note

We often regard the Y_i's as being *conditionally independent*, which means that each $y_i|\theta$ is independent of every other $Y_j|\theta$.

Technical Note: Why would we NOT be able to say that (marginally) Y_i and Y_j are independent? Because both depend on θ, and this induces some dependence between Y_i and Y_j. For example, if θ is large, $Y_i|\theta$ and $Y_j|\theta$ both are likely to be large; if θ is small, they are both likely to be small. Across the distribution for θ, then, there is some positive association between Y_i and Y_j.

Back to the soccer example

Let's consider the likelihood function for the number of goals scored in the first game, y_1, then for the second game, y_2, etc.

We will assume the number of goals scored in each game are *conditionally iid*.

More precisely, we assume $Y_i|\theta \overset{iid}{\sim} Poisson(\theta)$.

Suppose (temporarily–just for the next four slides) that we knew $\theta = 3.1$.

What is the likelihood of the soccer team scoring 2 goals in a game?

Back to the soccer example

Let's consider the likelihood function for the number of goals scored in the first game, y_1, then for the second game, y_2, etc.

We will assume the number of goals scored in each game are *conditionally iid*.

More precisely, we assume $Y_i|\theta \overset{iid}{\sim} \textit{Poisson}(\theta)$.

Suppose (temporarily–just for the next four slides) that we knew $\theta = 3.1$.

What is the likelihood of the soccer team scoring 2 goals in a game? $f(2|\theta = 3.1) = \frac{\exp(-3.1)3.1^2}{2!}$

```
dpois(2, 3.1)
```

```
## [1] 0.2164614
```

Soccer Example, continued

What is the likelihood of the soccer team scoring 0 goals in a game?

```
dpois(0, 3.1)

## [1] 0.0450492
```

What is the likelihood of the soccer team scoring 2 goals in the first game **and** 0 goals in the second?

Soccer Example, continued

What is the likelihood of the soccer team scoring 0 goals in a game?

```
dpois(0, 3.1)


## [1] 0.0450492
```

What is the likelihood of the soccer team scoring 2 goals in the first game **and** 0 goals in the second?

Because we assumed the goals scored are conditionally iid,

$f(y_1 = 2, y_2 = 0|\theta = 3.1) = f(y_1 = 2|\theta = 3.1)f(y_2 = 0|\theta = 3.1)$

```
dpois(2, 3.1)* dpois(0, 3.1)


## [1] 0.009751414
```

R trick to compute likelihoods when assuming data are conditionally iid

The R functions dbinom, dpois, and dnorm can accept multiple values for the first argument. When a vector of values are passed as the first argument, it evaluates the probability/density at each value. Compare with the previous two slides:

```
dpois( c( 0, 2),  3.1)

## [1] 0.0450492 0.2164614
```

To get the likelihood for conditionally iid data, you can feed the vector of data values into the function and then take the product using the *prod()* function. Compare with the previous slide:

```
prod( dpois( c(0,2),  3.1) )

## [1] 0.009751414
```

Compute the likelihood function for all 19 regular season games

Here is a vector of the goals scored:

```
goals <- c(2, 0, 3, 5, 2, 4, 3, 3, 7, 1,
           0, 4, 1, 0, 3, 2, 4, 4, 6)
```

Evaluate the likelihood of observing these data if θ were known to equal 3.1

Evaluate the likelihood of observing these data if θ were known to equal 2.5

Evaluate the likelihood of observing these data if θ were known to equal $\bar{y} = 54/19$; this is the maximum likelihood estimate of θ

How do we interpret these likelihood values? And why are they so small?

Posterior Inference after observing $y_1, y_2, \ldots, y_n$

If $Y_i|\theta \overset{iid}{\sim} \textit{Poisson}(\theta)$, and if $\theta \sim \textit{Gamma}(a, b)$, then $\theta|\{y_1, y_2, \ldots, y_n\} \sim \textit{Gamma}(a + \sum y_i, b + n)$

Derivation: on board

Soccer Goal Example, finalized

Recall that we have a vector for the goals scored in each game. A quick method to determine n is to find out the *length* (Number of elements) in the data vector *goals*.

```
length(goals)

## [1] 19
```

A quick method to determine the total number of occurrences of interest (in this case, total goals scored), is to use the *sum* function.

```
sum(goals)

## [1] 54
```

Posterior for $\theta \equiv$expected goals scored in a game, after observing all 19 Regular Season match scores

The prior was the Gamma(a=8, b=3) distribution, and thus the posterior (after observing that 54 goals were scored in 19 games) is the Gamma($a^\star = 8 + 54 = 62, b^\star = 3 + 19 = 22$) distribution.