

# STA 360/602L: MODULE 2.7

## GAMMA-POISSON MODEL I

DR. OLANREWaju MICHAEL AKANDE

# POISSON DISTRIBUTION RECAP

- $Y_1, \dots, Y_n \stackrel{iid}{\sim} \text{Poisson}(\theta)$  denotes that each  $Y_i$  is a **Poisson random variable**.
- The Poisson distribution is commonly used to model count data consisting of the number of events in a given time interval.
- Some examples: # children, # lifetime romantic partners, # songs on iPhone, # tumors on mouse, etc.
- The Poisson distribution is parameterized by  $\theta$  and the pmf is given by

$$\Pr[Y_i = y_i | \theta] = \frac{\theta^{y_i} e^{-\theta}}{y_i!}; \quad y_i = 0, 1, 2, \dots; \quad \theta > 0.$$

where

$$[Y_i] = [Y_i] = \theta.$$

- What is the joint likelihood? What is the best guess (MLE) for the Poisson parameter? What is the sufficient statistic for the Poisson parameter?

# GAMMA DENSITY RECAP

- The **gamma density** will be useful as a prior for parameters that are strictly positive.
- If  $\theta \sim \text{Ga}(a, b)$ , we have the pdf

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}.$$

where  $a$  is known as the **shape parameter** and  $b$ , the **rate parameter**.

- Another parameterization uses the **scale parameter**  $\phi = 1/b$  instead of  $b$ .
- Some properties:
  - $[\theta] = \frac{a}{b}$
  - $[\theta] = \frac{a}{b^2}$
  - $\text{Mode}[\theta] = \frac{a-1}{b}$  for  $a \geq 1$

# GAMMA DENSITY

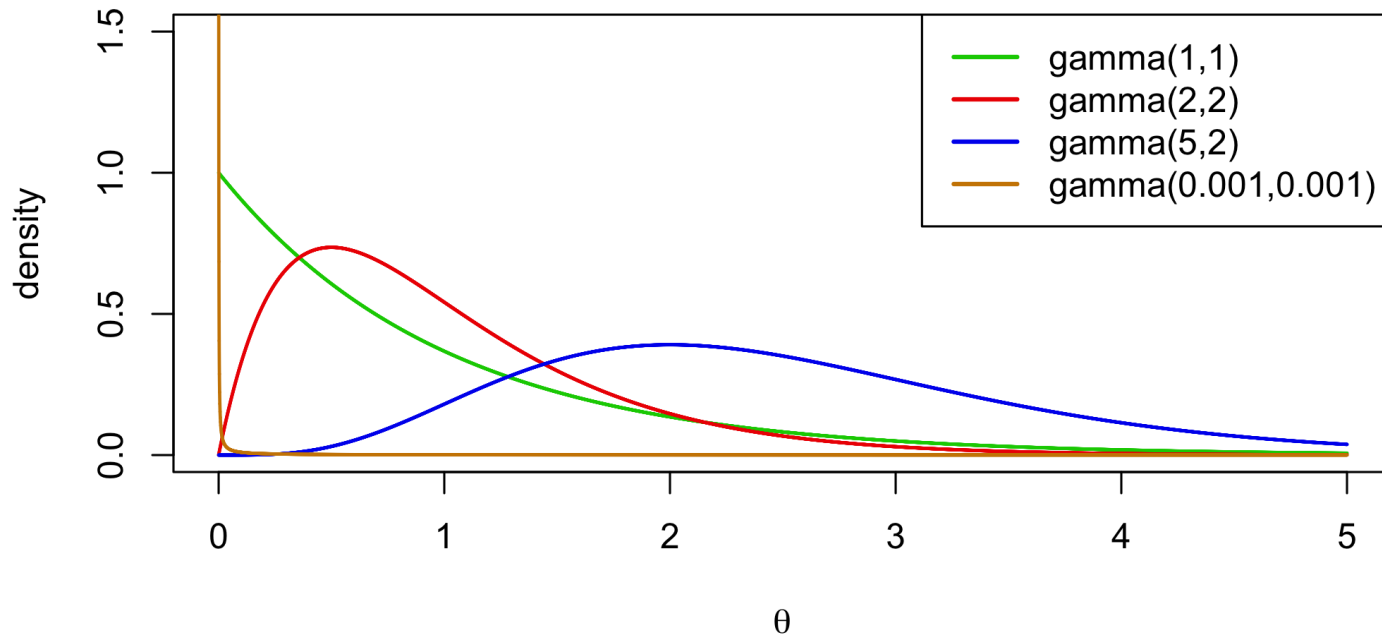
- If our prior guess of the expected count is  $\mu$  & we have a prior "scale"  $\phi$ , we can let

$$[\theta] = \mu = \frac{a}{b}; \quad [\theta] = \mu\phi = \frac{a}{b^2},$$

and solve for  $a, b$ . We can play the same game if we have a prior variance or standard deviation.

- More properties:
  - If  $\theta_1, \dots, \theta_p \stackrel{\text{ind}}{\sim} \text{Ga}(a_i, b)$ , then  $\sum_i \theta_i \sim \text{Ga}(\sum_i a_i, b)$ .
  - If  $\theta \sim \text{Ga}(a, b)$ , then for any  $c > 0$ ,  $c\theta \sim \text{Ga}(a, b/c)$ .
  - If  $\theta \sim \text{Ga}(a, b)$ , then  $1/\theta$  has an **Inverse-Gamma distribution**. We'll take advantage of these soon!

# EXAMPLE GAMMA DISTRIBUTIONS



*R has the option to specify either the rate or scale parameter so always make sure to specify correctly when using "dgamma", "rgamma", etc!.*

# GAMMA-POISSON

Generally, it turns out that

Poisson data:

$$p(y_i|\theta) : y_1, \dots, y_n \stackrel{iid}{\sim} \text{Poisson}(\theta)$$

+ Gamma Prior:

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} = \text{Ga}(a, b)$$

⇒ Gamma posterior:

$$\pi(\theta|\{y_i\}) : \theta|\{y_i\} \sim \text{Ga}(a + \sum y_i, b + n).$$

That is, updating a gamma prior with a Poisson likelihood leads to a gamma posterior – we once again have conjugacy.

Can we derive the posterior distribution and its parameters? Let's do some work on the board.

# GAMMA-POISSON

- With  $\pi(\theta|\{y_i\}) = \text{Ga}(a + y_i, b + n)$ , we can think of
  - $b$  as the "number prior of observations" from some past data, and
  - $a$  as the "sum of the counts from the  $b$  prior observations".
- Using the properties of the gamma distribution, we have
  - $E[\theta|\{y_i\}] = \frac{a + y_i}{b + n}$
  - $\text{Var}[\theta|\{y_i\}] = \frac{a + y_i}{(b + n)^2}$
- So, as we did with the beta-binomial, we can once again write the posterior expectation as a weighted average of prior and data.

$$E(\theta|\{y_i\}) = \frac{a + y_i}{b + n} = \frac{b}{b + n} \times \text{prior mean} + \frac{n}{b + n} \times \text{MLE}.$$

- Again, as we get more and more data, the majority of our information about  $\theta$  comes from the data as opposed to the prior.

# HOFF EXAMPLE: BIRTH RATES

- Survey data on educational attainment and number of children of 155 forty-year-old women during the 1990's.
- These women were in their 20s during the 1970s, a period of historically low fertility rates in the US.
- **Goal:** compare birth rate  $\theta_1$  for women with bachelor's degrees to the rate  $\theta_2$  for women without.
- **Data:**
  - 111 women without a bachelor's degree had 217 children: ( $\bar{y}_1 = 1.95$ )
  - 44 women with bachelor's degrees had 66 children: ( $\bar{y}_2 = 1.50$ )
- Based on the data alone, looks like  $\theta_1$  should be greater than  $\theta_2$ . But...how sure are we?
- **Priors:**  $\theta_1, \theta_2 \sim \text{Ga}(2, 1)$  (not much prior information; equivalent to 1 prior woman with 2 children). Posterior means will be close to the MLEs.



# HOFF EXAMPLE: BIRTH RATES

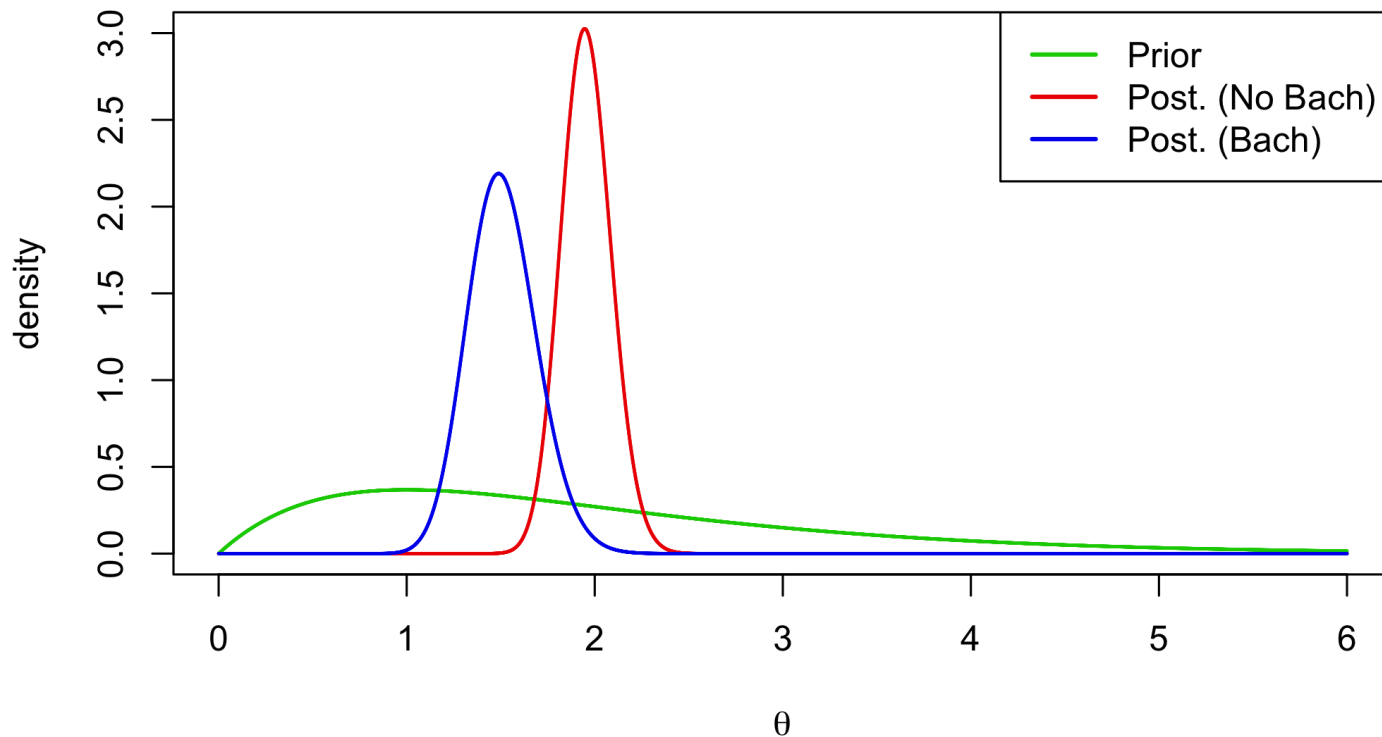
- Then,
  - $\theta_1 | \{n_1 = 111, y_{i,1} = 217\} \sim \text{Ga}(2 + 217, 1 + 111) = \text{Ga}(219, 112).$
  - $\theta_2 | \{n_2 = 44, y_{i,2} = 66\} \sim \text{Ga}(2 + 66, 1 + 44) = \text{Ga}(68, 45).$
- Use R to calculate posterior means and 95% CIs for  $\theta_1$  and  $\theta_2$ .

```
a=2; b=1; #prior
n1=111; sumy1=217; n2=44; sumy2=66 #data
(a+sumy1)/(b+n1); (a+sumy2)/(b+n2); #post means
qgamma(c(0.025, 0.975), a+sumy1, b+n1) #95% ci 1
qgamma(c(0.025, 0.975), a+sumy2, b+n2) #95% ci 2
```

- Posterior means:  $[\theta_1 | \{y_{i,1}\}] = 1.955$  and  $[\theta_2 | \{y_{i,2}\}] = 1.511.$
- 95% credible intervals
  - $\theta_1$ : [1.71, 2.22].
  - $\theta_2$ : [1.17, 1.89].

# HOFF EXAMPLE: BIRTH RATES

Prior and posteriors:



# HOFF EXAMPLE: BIRTH RATES

- Posteriors indicate considerable evidence birth rates are higher among women without bachelor's degrees.
- Confirms what we observed.
- Using sampling we can quickly calculate  $\Pr(\theta_1 > \theta_2 | \text{data})$ .

```
mean(rgamma(10000,219,112)>rgamma(10000,68,45))
```

We have  $\Pr(\theta_1 > \theta_2 | \text{data}) = 0.97$ .

- Why/how does it work?
- Monte Carlo approximation coming soon!
- Clearly, that probability will change with different priors.

# WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!