# Gamma-Poisson Inference, Part 1

STAT 251, Unit 5A

# Overview

Review

Poisson Distribution

Gamma Distribution

# Review

# Recap from Previous Lecture Material, part 1

- We combine prior beliefs with information from additional data to obtain posterior beliefs about unknowns (parameters)
- If parameter of interest represents a population proportion, a convenient choice for mathematically representing prior beliefs is to use a $Beta(a, b)$ distribution and a convenient model for information conveyed by the additional data is a binomial likelihood.

In the $Beta(a, b)$ distribution, what interpretation do we give to $a$ and to $b$?

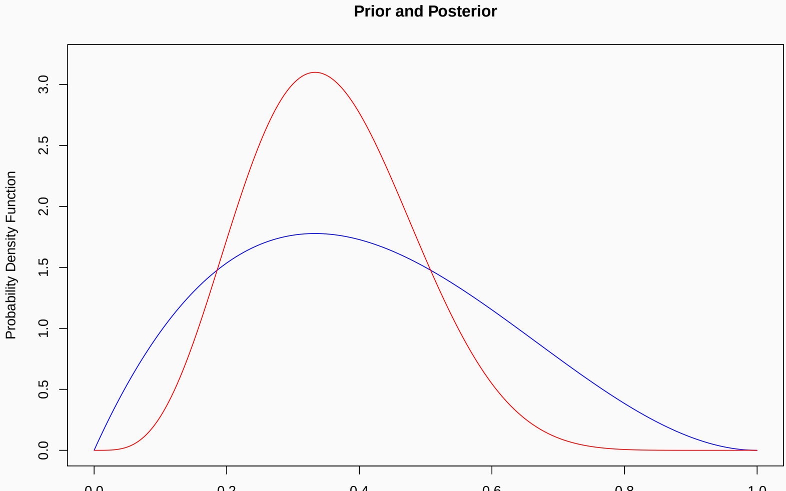
If we observe $y$ successes out of $n$ trials, what would the resulting posterior distribution be?

# Recap from Previous Lecture Material, part 2

- The Monte Carlo method uses random draws from an underlying distribution to estimate expected values associated with the underlying distribution
- This is particularly useful for posterior inference comparing two populations.

# Review Question 1

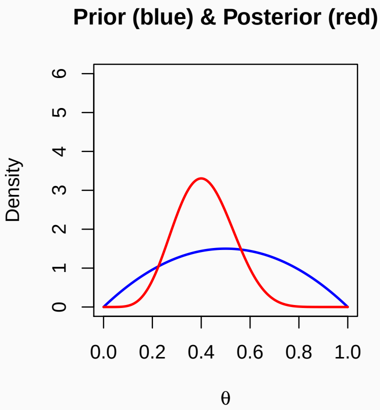
Below, I have included a plot that shows both the prior and the posterior distribution for $\theta$. Which color do you think represents the prior? Why?

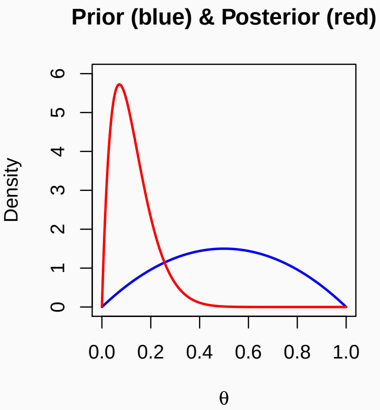


# Review Question 2

In which of the two situations below does the posterior agree less with the corresponding prior distribution?

When the prior distribution does not agree well with the posterior distribution, what does this suggest about the prior beliefs?



# Poisson Distribution

# New Situation

We'll move to a different kind of data. We have looked at binomial data (random variable $y$ represents how many successes we have out of a predetermined number of trials).

What if we have count data where there are not a predetermined number of trials?

One possibility: the Poisson distribution.

# Poisson Distribution

Assumptions for Poisson distribution:

- The random variable $Y$ represents the number of occurrences for some event of interest in a fixed amount of time/space/etc.
- There is a rate parameter, usually called $\lambda$ but which we will denote by $\theta$ to be consistent with established notation. This represents the expected number of occurrences per unit interval (could be a unit of time, or a unit of space, etc.)
- Occurrences are assumed to be independent in time/space/etc.
- Occurrences are assumed to have a constant underlying rate (i.e., $\theta$ does not change over time/space/etc).
- The probability of two (or more) occurrences within a very, very small interval is essentially zero.

# Discussion

Under which of the following scenarios are the Poisson assumptions reasonable?

- Y=number of goals scored by the BYU women's soccer team in their next game
- Y=number of fatalities on Utah roadways in October
- Y=number of pine trees in a randomly selected 10m$\times$10m plot of the Uinta-Wasatch-Cache National Forest
- Y=count in 1 minute period from Geiger counter at a decommissioned nuclear test site.
- Y=number of free throws I make out of 10 attempts

# Poisson Likelihood

The pmf of the *Poisson*($\theta$) distribution is:

$$f(y|\theta) = \frac{\theta^y e^{-\theta}}{y!} \mathbb{1}_{y \in \{0,1,2,...\}}$$

Again, more commonly this is written as

$$f(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \mathbb{1}_{y \in \{0,1,2,...\}}$$

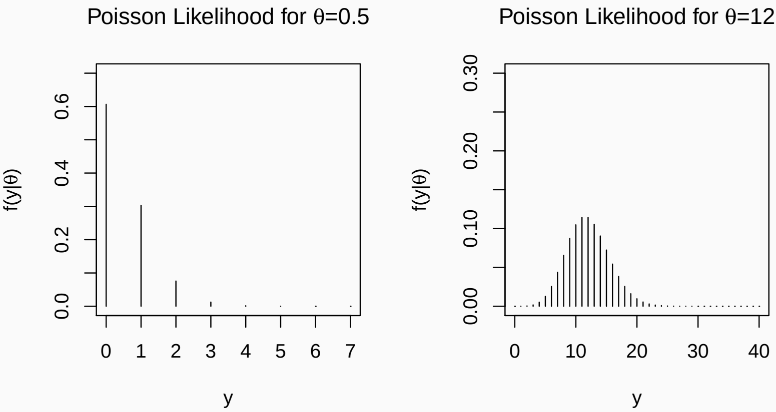
but I want to continue to use $\theta$ to represent parameters.

This can be evaluated in R by using the *dpois* function; use *?dpois* to learn more.

What would the functions *rpois, qpois, and ppois* be used for?

# Two Poissons pmfs, plotted

Here is what the Poisson distribution looks like for two choices of $\theta$.




Your turn: Plot the Poisson distribution if $\theta = 4$.

# Characteristics of Poisson Distribution

If $Y|\theta$ has the Poisson distribution, then $E(Y|\theta)$ (or $\mu_{Y|\theta}$, the mean of Y for a given $\theta$) is

$$
\begin{aligned}
E(Y|\theta) =& \sum_{y=0}^{\infty} y f(y|\theta) = \sum_{y=0}^{\infty} y \frac{\theta^y e^{-\theta}}{y!} \mathbb{1}_{y \in \{0,1,2,\ldots\}} \\
=& \sum_{y=1}^{\infty} y \frac{\theta^y e^{-\theta}}{y!} \mathbb{1}_{y \in \{1,2,\ldots\}} \\
=& \theta \sum_{y=1}^{\infty} \frac{\theta^{y-1} e^{-\theta}}{(y-1)!} \mathbb{1}_{(y-1) \in \{0,1,2,\ldots\}} \\
=& \theta \sum_{y=0}^{\infty} \frac{\theta^{y} e^{-\theta}}{(y)!} \mathbb{1}_{y \in \{0,1,2,\ldots\}} \\
=& \theta
\end{aligned}
$$

# Characteristics of Poisson Distribution (cont.)

If $Y|\theta$ has the Poisson distribution, then $Var(Y|\theta)$ (or $\sigma^2_{Y|\theta}$, the variance of Y for a given $\theta$) is

$$Var(Y|\theta) = \theta$$

Conclusion: If $Y|\theta$ has the Poisson($\theta$) distribution, then both the mean and variance of $Y$ (for a known value of $\theta$) equal $\theta$.

# Prior for the Poisson rate parameter, $\lambda$

Suppose we are assuming a Poisson($\theta$) likelihood.

What about the prior distribution for $\theta$ (the rate parameter)?

A common (and convenient) choice is to choose a prior from the *gamma* family of distributions.

A gamma distribution assumes positive-valued variables and has no upper bound.

Good news! Our rate parameter is positive-valued and need not have an upper bound.

# Gamma Distribution

# Gamma Distribution–two forms

There are two common forms in which the Gamma distribution is expressed. To help us keep track of which form is which, I will refer to one as the Gamma(a, b) distribution, and to the other as the Gamma(a, $\beta$) distribution.

$$\text{What I'll call Gamma(a, b):} \quad \pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{1}_{\theta>0}$$

$$\text{What I'll call } \mathit{Gamma}(a, \beta): \quad \pi(\theta) = \frac{1}{\beta^a \Gamma(a)} \theta^{a-1} e^{-\theta/\beta} \mathbb{1}_{\theta>0}$$

(This notation—(a,b) vs. (a, $\beta$)—is NOT a standard convention, but I'll use it to help us keep track of which form is which.)

# Correspondence between the Two Forms

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{1}_{\theta>0}$$

$$\pi(\theta) = \frac{1}{\beta^a \Gamma(a)} \theta^{a-1} e^{-\theta/\beta} \mathbb{1}_{\theta>0}$$

Notice how they relate: By substituting $\beta = 1/b$, the second form would become the first form.

Also, a, b, and $\beta$ must be positive, but are otherwise unrestricted!

# How is the value of $a$ interpreted in what I refer to as the Gamma($a$, $b$) distribution?

In each of the forms, the role of $a$ is the same: it is a *shape* parameter. To see this, I'll plot the function with $b = 1$, which is the same as the function with $\beta = 1$ in the second form.

Aside: The exponential distribution is a special case of the gamma distribution (with $a = 1$).

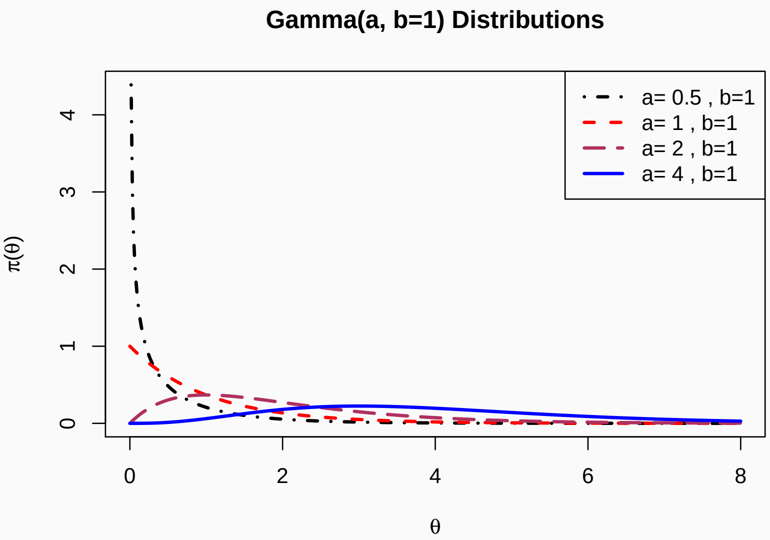
What R function will we need to plot the gamma distribution?

```r
?dgamma
```

# R Code to Plot Gamma(a, b=1) pdfs

```
thetas <- seq(0, 8, length=501)
plot(thetas, dgamma(thetas, .5, 1), col="black",
   lwd=2.5, main="Gamma(a, b=1) Distributions",
   xlab=expression(theta), lty=4, type="l",
   ylab=expression(paste(pi, "(", theta, ")", sep="")))
lines(thetas, dgamma(thetas,1,1), col="red",
      lwd=2.5, lty=2)
lines(thetas, dgamma(thetas,2,1), col="maroon",
      lwd=2,lty=5)
lines(thetas, dgamma(thetas,4,1), col="blue",
      lwd=2.5,lty=1)
legend("topright", paste("a=", c(.5,1,2,4),", b=1"),
       col=c("black","red","maroon","blue"),
       lwd=2.5, lty=c(4,2,5,1))
```

## Notice how the shape of the density function changes with $a$



## Interpretations of $b$ (in first gamma form) and $\beta$ (in second gamma form)

*a* is a shape parameter.

But what is the interpretation of *b*?

What is the interpretation of $\beta$ (for those instances where you may encounter the second form)?

We'll consult the R documentation to learn more about the two forms:

```
?dgamma
```

Note that you can specify the *rate* for the gamma distribution or you can specify the *scale*.

# Rate and Scale parameters of the Gamma Distribution

In the first version of the gamma distribution, $b$ is the rate parameter.

In the second version, $\beta$ is the scale parameter.

Why are they called *rate* and *scale* parameters?

# Interpretation of the Gamma's rate Parameter

Sometimes the gamma distribution is applied when modeling the wait time until an event happens.

Suppose we had a random variable, $Y$, to represent the wait time (in minutes).

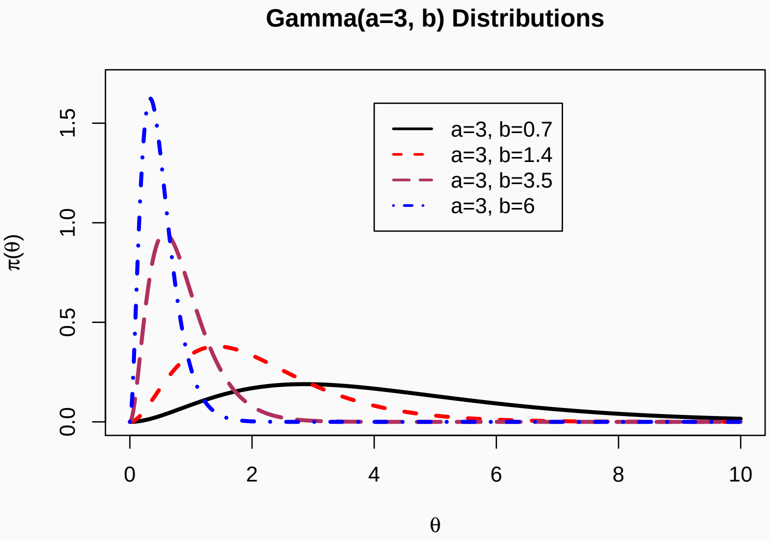
If the *rate* of occurrences is high (e.g., tendency for many occurrences per minute) then what would happen to the wait time? It would tend to be small.

If the rate of occurrences decreases, what would happen to the wait time? It would tend to be larger.

# R Code to plot Gamma(a=3, b) Densities

```
thetas <- seq(0, 10, length=501)
plot(thetas, dgamma(thetas, 3, rate=.7), col="black",
   lwd=3, main="Gamma(a=3, b) Distributions",
   xlab=expression(theta), lty=1, type="l",
   ylim=c(0,1.7),
   ylab=expression(paste(pi, "(", theta, ")", sep="")))
lines(thetas, dgamma(thetas,3, rate=1.4), col="red",
      lwd=3, lty=2)
lines(thetas, dgamma(thetas,3, rate=3.5), col="maroon",
      lwd=3,lty=5)
lines(thetas, dgamma(thetas,3,rate=6), col="blue",
      lwd=3,lty=4)
legend(4, 1.6, paste("a=3, b=", c(.7,1.4,3.5,6), sep=""),
       col=c("black","red","maroon","blue"),
       lwd=2, lty=c(1,2,5,4))
```

# Notice how the mean changes with b (the rate)



## Interpretation of the Gamma's scale parameter

Earlier I noted that $\beta = 1/b$ makes the two forms for the gamma distribution identical.

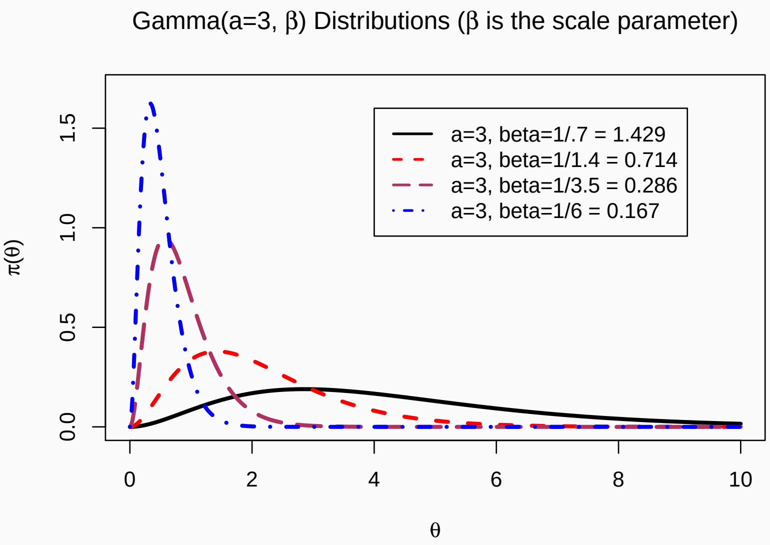
$b$ is the *rate* for the gamma distribution.

$\beta$ is 1/rate, which is referred to as the *scale*.

Larger values of $\beta$ *scale* the distribution so that it tends to produce larger values when sampled.

```r
plot(thetas, dgamma(thetas, 3, scale=1/.7), col="black",
   lwd=3, main=expression(paste("Gamma(a=3, ",beta,
     ") Distributions (",beta," is the scale parameter)",
    sep="")), xlab=expression(theta), lty=1, type="l",
   ylim=c(0,1.7),
   ylab=expression(paste(pi, "(", theta, ")", sep="")))
lines(thetas, dgamma(thetas,3, scale=1/1.4), col="red",
      lwd=3, lty=2)
lines(thetas, dgamma(thetas,3, scale=1/3.5),
      col="maroon", lwd=3,lty=5)
lines(thetas, dgamma(thetas,3,scale=1/6), col="blue",
      lwd=3,lty=4)
legend(4, 1.6, paste("a=3, beta=",
  c("1/.7","1/1.4","1/3.5","1/6")," = ",
  round(1/c(.7,1.4, 3.5, 6),3), sep=""),
       col=c("black","red","maroon","blue"),
       lwd=2, lty=c(1,2,5,4))
```

# Notice how the mean changes with $\beta$ (the scale)



# Numerical summaries of the gamma distribution

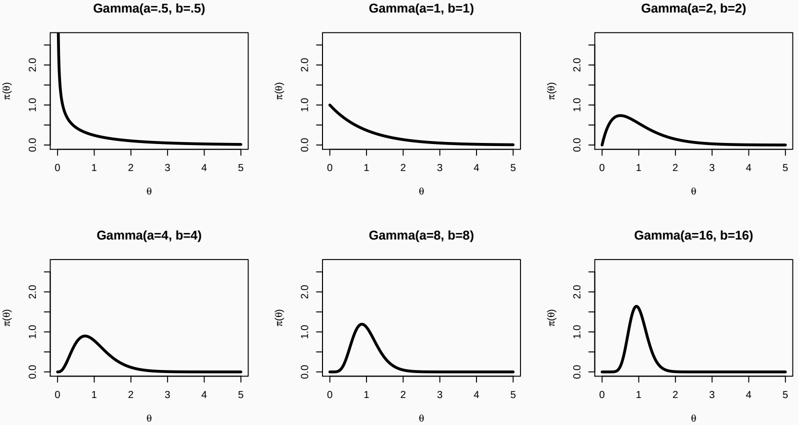
If $\theta \sim Gamma(a, b)$, then:

- The expected value (mean) of $\theta$ is $E(\theta) = a/b$.
- The variance of $\theta$ is $Var(\theta) = a/b^2$.
- The mode is also a function of $a$ and $b$, but it is left as a homework problem to derive the expression.

If expressing as $\theta \sim Gamma(a, scale = \beta)$, what are the mean and variance of $\theta$ (as functions of $a$ and $\beta$)?

# Approximate normality of Gamma(a,b) as $a$ increases

What happens to the gamma distribution as the shape parameter $a$ increases? Without loss of generality, I'll use $b = a$ in all of the following graphs so that the mean is 1 regardless of the value of a.



# In Gamma-Poisson, Part 2

- Analytical derivation of the posterior distribution of $\theta$ when $Y|\theta$ has the *Poisson*($\theta$) distribution and $\theta$ 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.