Statistical Inference, pt. III

ECON 3640–001

Marcio Santetti

Spring 2022

Motivation

Housekeeping

Notes based on `Johnson et al. (2022)`:

- Chapters 4 (optional) and 5
- Available `here`

Last time...

Last time, we were introduced to **conjugate priors**.

It simply means that, by combining priors and likelihoods from certain families, Bayes' theorem will return a posterior whose distribution is the **same** as the prior.

Consequently, we **know** which distribution the posterior follows and are able to *analytically* calculate it.

Let us study more conjugate families now, namely

1. *Gamma-Poisson*;
2. *Normal-Normal*

The Gamma-Poisson model

The Gamma-Poisson model

Take a look at your email inbox and search for how many **spam messages** you currently have.

Assume that we want to know more about the **rate** with which spam emails come into our inbox for a given number of days.

Does this problem fit into a Beta-Binomial setting?

No!

This rate does not fit solely on the [0,1] interval, just as a proportion.

Furthermore, the number of spam messages is a **count** that can take on any integer value, and is not limited by a number of trials, as with a Binomial experiment.

The Gamma-Poisson model

Our variable of interest is the **rate** with which spam messages come into our inbox over a given number of days.

And the number of spam messages is a **count** random variable.

We are once again dealing with a **discrete random variable**, and this situation fits perfectly well with one discrete distribution we've already studied.

- The **Poisson** distribution.

The Gamma-Poisson model

Let's label the daily **count** of spam messages as Y_i.

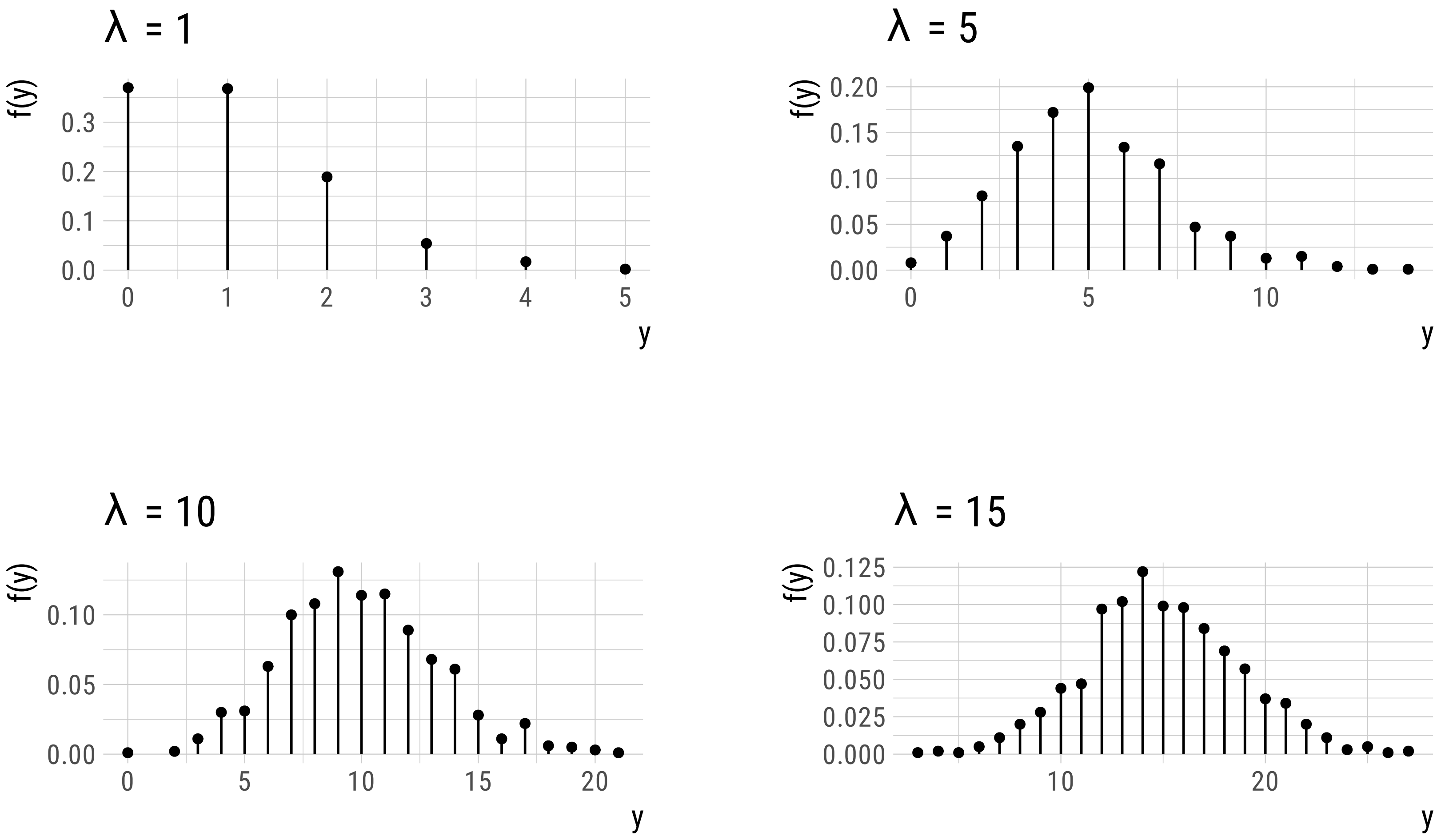
- $Y_i = \{0, 1, 2, 3, 4, \ldots\}$

Recall the **Probability Mass Function** (PMF) of a Poisson-distributed random variable:

$$f(Y \mid \lambda) = \frac{\lambda^Y \, e^{-\lambda}}{Y!} \qquad \text{for } Y \in \{0, 1, 2, 3, \ldots\}$$

Moreover, $E(Y \mid \lambda) = \text{Var}(Y \mid \lambda) = \lambda$.

Depending on the value of λ, the Poisson distribution will have **different shapes**.



λ = 1
f(y)
0.0
0.1
0.2
0.3
0
1
2
3
4
5
y
λ = 5
f(y)
0.00
0.05
0.10
0.15
0.20
0
5
10
y
λ = 10
f(y)
0.00
0.05
0.10
0
5
10
15
20
y
λ = 15
f(y)
0.000
0.025
0.050
0.075
0.100
0.125
10
20
y

The Gamma-Poisson model

Now, assume that $(Y_1, Y_2, Y_3, \ldots, Y_n)$ are the number of spam messages observed on each of the *n* days we are observing these data.

The daily number of spam messages will likely **differ** from day to day.

Therefore, on each day *i*

$$Y_i \mid \lambda \overset{ind}{\sim} \text{Pois}(\lambda)$$

The Gamma-Poisson model

In order to account for **all** individual days, we need to rewrite the Poisson PMF as a **joint** probability mass function:

$$f(\vec{y} \mid \lambda) = \prod_{i=1}^{n} f(y_i \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} \, e^{-\lambda}}{y_i!}$$

The above expression simply follows the product rule for independent events:

- $P(A \cap B) = P(A) \; P(B)$

The Gamma-Poisson model

$$f(\vec{y} \mid \lambda) = \prod_{i=1}^{n} f(y_i \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda_i^{y} \, e^{-\lambda}}{y_i!}$$

can be simplified to

$$f(\vec{y} \mid \lambda) = \frac{\lambda^{\sum y_i} \, e^{-n\lambda}}{\prod_{i=1}^{n} y_i!}$$

The Gamma-Poisson model

Now that the likelihood has been defined, it is time to think about the **prior distribution** for our target parameter, λ.

With λ being a **positive** and **continuous** rate, we can incorporate any **prior** information we have available in order to **tune** our prior's *hyperparameter.*

Luckily, we do have a **conjugate prior** for the Poisson distribution.

This prior is the **Gamma distribution**.

The Gamma-Poisson model

If λ is a continuous RV, taking on any positive value ($\lambda > 0$), its variability may be represented by a **Gamma distribution** with **shape and rate** hyperparameters s and r, respectively:

$$\lambda \sim \mathrm{Gamma}(s, r) \quad \text{with } \; s, r > 0$$

The **PDF** of a Gamma distribution is represented by

$$f(\lambda) = \frac{r^s}{\Gamma(s)} \lambda^{s-1} e^{-r\lambda} \quad \text{for } \; \lambda > 0$$

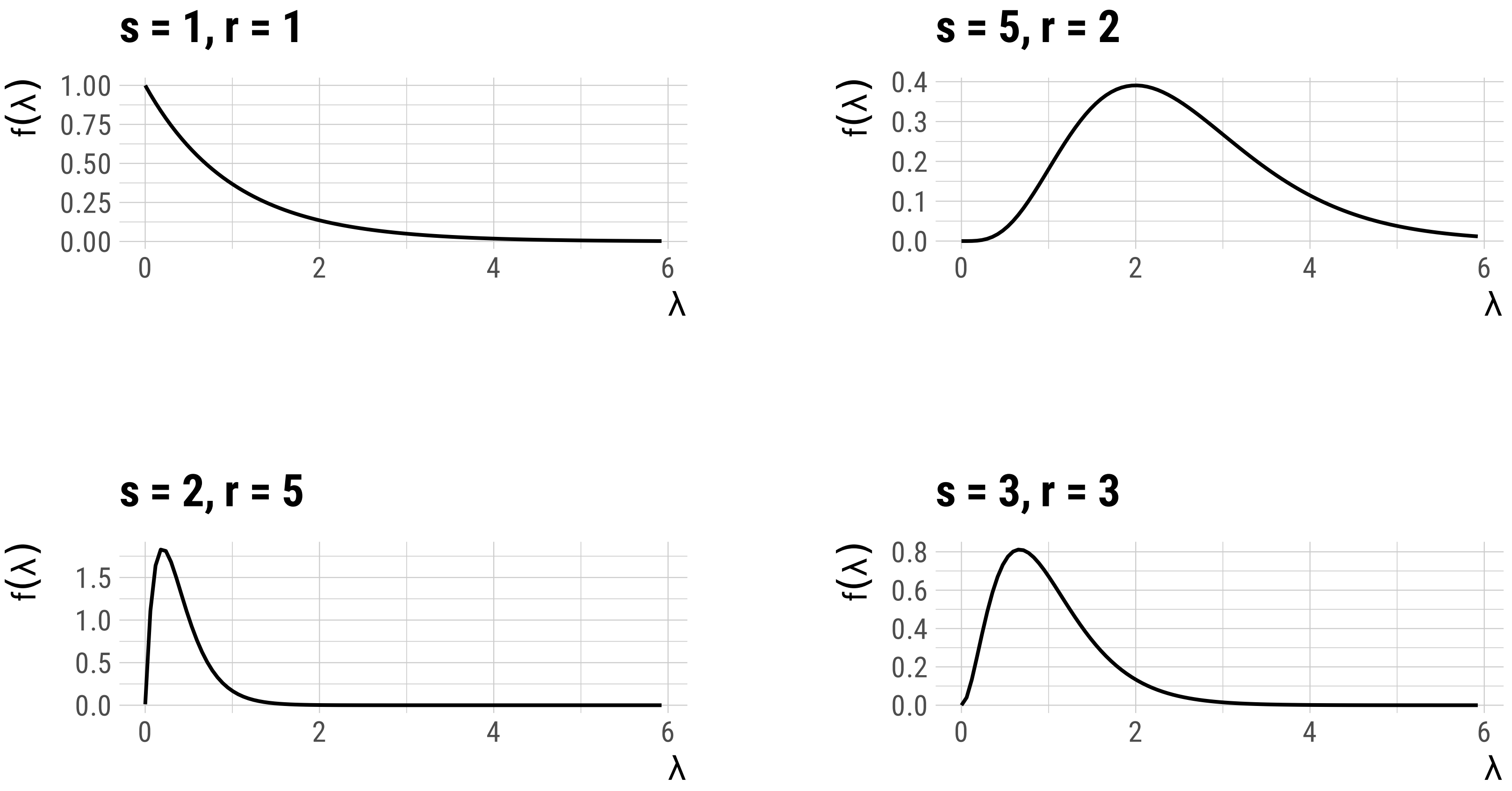
The Gamma-Poisson model

The expected value, mode, and variance for the Gamma distribution are given by:

- **Expected Value**: $E(\lambda) = \dfrac{s}{r}$;
- **Mode**: $\mathrm{Mode}(\lambda) = \dfrac{s-1}{r}$;
- **Variance**: $\mathrm{Var}(\lambda) = \dfrac{s}{r^2}$.

When the shape (s) hyperparameter of a Gamma distribution equals 1, λ follows an **Exponential distribution**:

$$\lambda \sim \mathrm{Exp}(r)$$


s = 1, r = 1
f(λ)
1.00
0.75
0.50
0.25
0.00
0
2
4
6
λ
s = 5, r = 2
f(λ)
0.4
0.3
0.2
0.1
0.0
0
2
4
6
λ
s = 2, r = 5
f(λ)
1.5
1.0
0.5
0.0
0
2
4
6
λ
s = 3, r = 3
f(λ)
0.8
0.6
0.4
0.2
0.0
0
2
4
6
λ

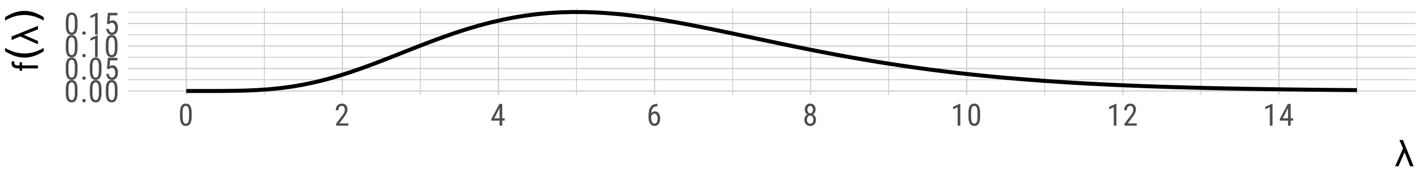
The Gamma-Poisson model

Now, let us **tune** our prior distribution's hyperparameters according to any prior knowledge we have on the problem we are facing.

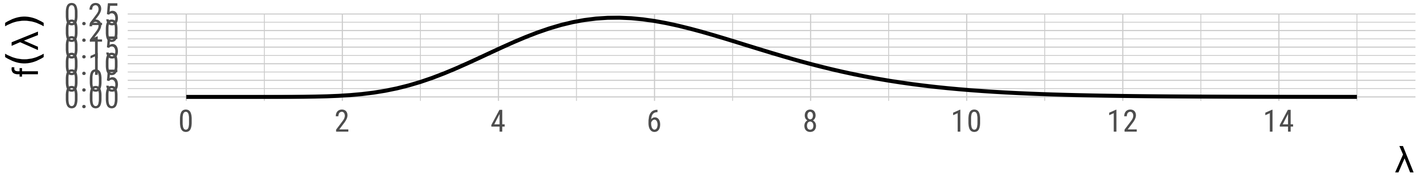
Suppose that, in the past, you've noticed that about 6 spam emails would come each day, varying between 2 and 10.

How do we translate this information into a probability distribution?

Gamma(6, 1)



Gamma(12, 2)



The Gamma-Poisson model

With prior and likelihood defined, we can move on the the **posterior estimation**.

- **Prior**: $\lambda \sim \mathrm{Gamma}(6, 1)$
- **Likelihood**: $Y_i \,|\lambda \;\sim \mathrm{Poisson}(\lambda)$

By **conjugacy**, the posterior will be

$$\lambda \mid \vec{y} \sim \mathrm{Gamma}\big(s + \sum y_i, r + n\big)$$

where n is the number of data points (in our case, days) used in our analysis.

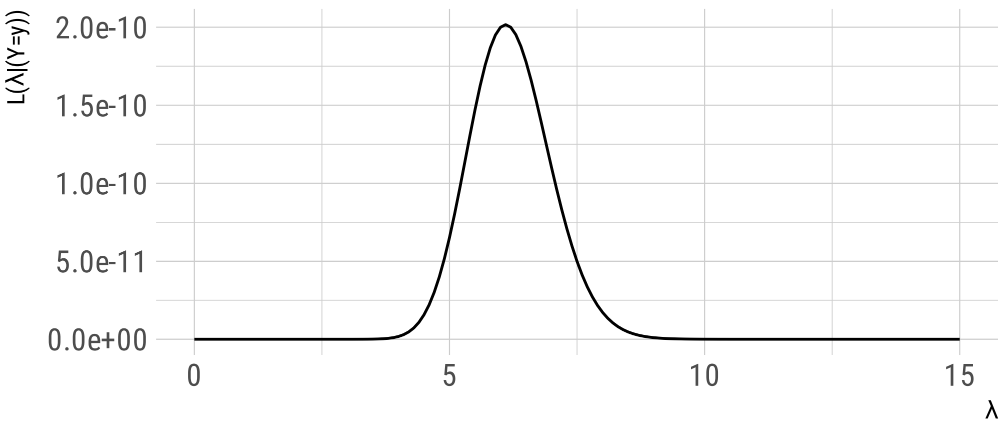
The Gamma-Poisson model

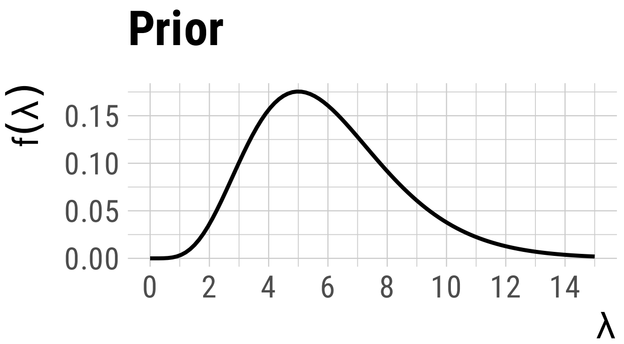
Suppose that we observe **new data** for 10 days.

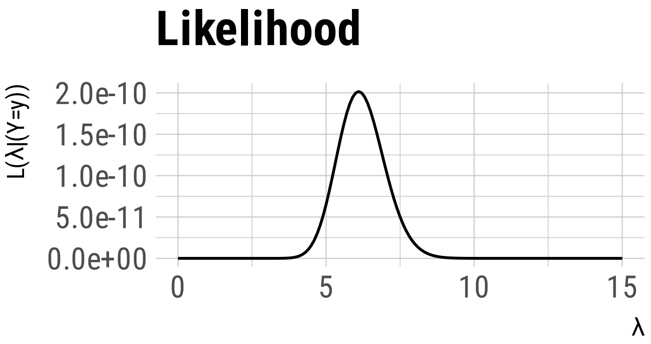
And these are the number of spam messages received each day:

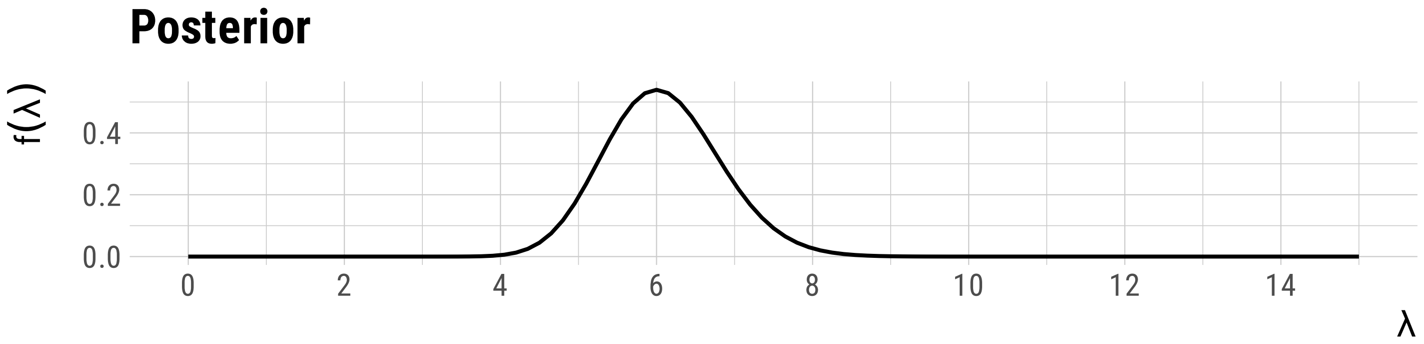
$$\vec{y} = \{6, 10, 3, 5, 7, 6, 6, 10, 3, 5\}$$

```
library(bayesrules)
plot_poisson_likelihood(y = c(6, 10, 3, 5, 7, 6, 6, 10, 3, 5), lambda_upper_bound = 15)
```



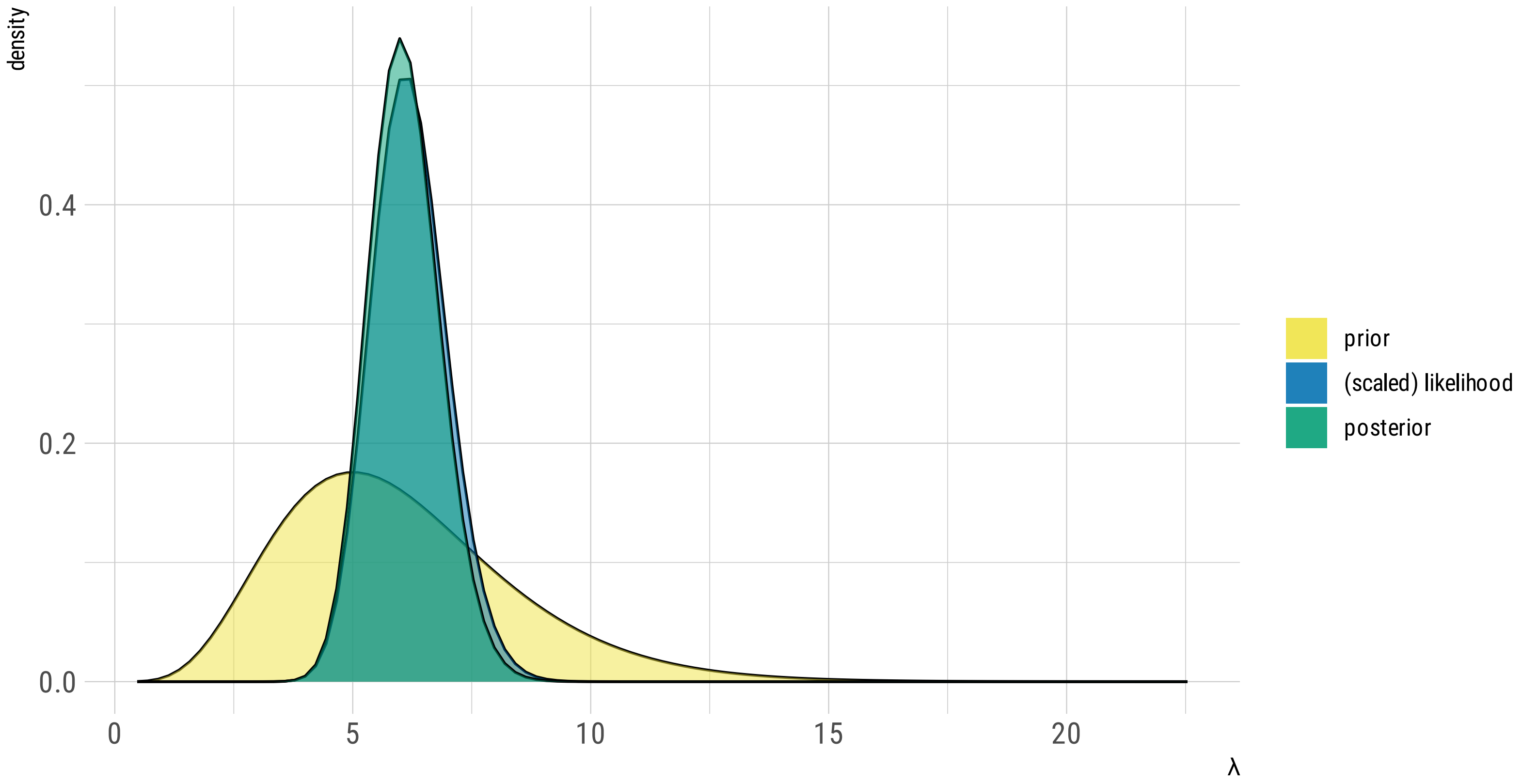

Prior
f(λ)
0.15
0.10
0.05
0.00
0 2 4 6 8 10 12 14
λ


Likelihood
L(λ|(Y=y))
2.0e-10
1.5e-10
1.0e-10
5.0e-11
0.0e+00
0 5 10 15
λ


Posterior
f(λ)
0.4
0.2
0.0
0 2 4 6 8 10 12 14
λ

The Gamma-Poisson model

```r
library(bayesrules)
plot_gamma_poisson(shape = 6, rate = 1, sum_y = 61, n = 10)
```



The Gamma-Poisson model

```r
library(bayesrules)
summarize_gamma_poisson(shape = 6, rate = 1, sum_y = 61, n = 10)
```

```
#>       model shape rate     mean mode      var       sd
#> 1     prior     6    1 6.000000    5 6.000000 2.449490
#> 2 posterior    67   11 6.090909    6 0.553719 0.744123
```

The Normal-Normal model

The Normal-Normal model

The last conjugate family we will cover in detail is the **Normal-Normal** model.

If a random variable Y is continuous and can take on any value between $-\infty$ and $+\infty$, its variability can be modeled through a **Normal** distribution with **mean** μ and **standard deviation** σ:

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

Recall the Normal **PDF**:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \; ; \; -\infty < x < \infty$$

The Normal-Normal model

Some useful summary statistics:

- **Expected Value and Mode**: $E(Y) = \text{Mode}(Y) = \mu$;
- **Variance**: $\text{Var}(Y) = \sigma^2$.
- **Standard deviation**: $\text{SD}(Y) = \sigma$.

When a random variable follows a probability distribution, we can state that roughly **95%** of its values fall within **2 standard deviations** of its mean, μ:

$$\mu \ \pm \ 2\sigma$$

The Normal-Normal model

For a likelihood function

$$Y_i \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$$

For a sample size of n, the **joint PDF** for all individuals is

$$f(\vec{y} \mid \mu) = \prod_{i=1}^{n} f(y_i \mid \mu) \propto \exp\left[- \frac{(\bar{y} - \mu)^2}{2\sigma^2/n} \right] \quad \text{for} \;\; \mu \in (-\infty, +\infty)$$

The Normal-Normal model

When our parameter of interest is μ, the averag value of a variable we are curious about, we can define its prior model as

$$\mu \sim \mathcal{N}(\theta, \tau^2)$$

A posterior model for μ with both prior and likelihood following Normal distributions will be given by

$$\mu \mid \vec{y} \sim \mathcal{N}\left[\theta \frac{\sigma^2}{n\tau^2 + \sigma^2} + \bar{y}\frac{n\tau^2}{n\tau^2 + \sigma^2},\ \frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2}\right]$$

The Normal-Normal model

Carefully read section **5.5** from the `Bayes Rules!` book.

Next time: Approximating the posterior