STAT 339: Statistical Theory

Bayesian Parameter Estimation

Anthony Scotina

Bayes' Rule (for events)

Theorem (Bayes' Rule)

For events A and B,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A)},$$

where by the *Law of Total Probability*,

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c).$$

Real or Fake News?

From *Bayes Rules!* by Johnson, Ott, and Dogucu:

> Is the article fake or not?!

In a sample of $n = 150$ articles posted on Facebook and fact-checked…

- **60%** are real: $P(B) = 0.6$
- **40%** are fake: $P(B^c) = 0.4$

These are **prior probabilities**. They suggest that, *assuming the sample is representative*, incoming articles are most likely real.

- **2.22%** of *real news* titles (2 of 90) used an exclamation point: $P(A \mid B) = 0.0222$
- **26.67%** of *fake news* titles (16 of 60) used an exclamation point: $P(A \mid B^c) = 0.2667$

The **data** suggest that exclamation points are more consistent with *fake news* titles.

Real or Fake News?

Using **Bayes' Rule**, we can calculate the **posterior probability** of whether an article is real:

$$
\begin{aligned}
P(B \mid A) &= \frac{P(A \mid B)P(B)}{P(A)} \\
&= \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)} \\
&= \frac{0.0222 \times 0.6}{0.0222 \times 0.6 + 0.2667 \times 0.4} \\
&= 0.111
\end{aligned}
$$

Thus, after *balancing* our **prior** information and the information present in the **data**, we have developed a **posterior** understanding of whether an article is real.

- Equivalently, there is a ~**89%** *posterior probability* that an article is *fake*, given the presence of an exclamation point in the title.

Bayesian Estimation

We will *still* refer to θ as the **target parameter** of interest.

- But in *Bayesian* parameter estimation, we treat θ as a **random variable**, rather than a *fixed* value. That is, θ is still unknown, but it can *vary* or *fluctuate* over time.

We can still use *Bayes' Rule* to evaluate distributions of θ, *given* the observed data:

$$f(\theta \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \theta) f(\theta)}{f(\mathbf{y})},$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ and $f(\mathbf{y}) = \int_{-\infty}^{\infty} f(\mathbf{y} \mid \theta) f(\theta)\, d\theta$.

- $f(\theta)$ is the **prior distribution** PDF of a parameter *before observing any data.*

- $f(\mathbf{y} \mid \theta)$ is the **likelihood function**, which gives the relative likelihood of observing data $\mathbf{y}$ under different values of θ.

- $f(\theta \mid \mathbf{y})$ is the **posterior distribution** PDF of the parameter, *given the observed data.*

Animal Crossing!

Suppose a group of college students are interested in starting an *Animal Crossing* club.

- In order to estimate demand, the students want to estimate θ, the **proportion of students who play Animal Crossing**.

Based on anecdotal evidence, the students think that θ could reasonably range from 0.1 to 0.25.

- Though *in reality*, θ could be **any** value between 0 and 1.

> How might we model our **prior** understanding of the parameter, θ?

Prior Distribution

If we treat θ as *random*, then the distribution that one assigns to θ *before observing any data* is called the **prior distribution**.

- In the Animal Crossing example, because θ can be *any number between 0 and 1*, what might be a suitable prior distribution? 🤔

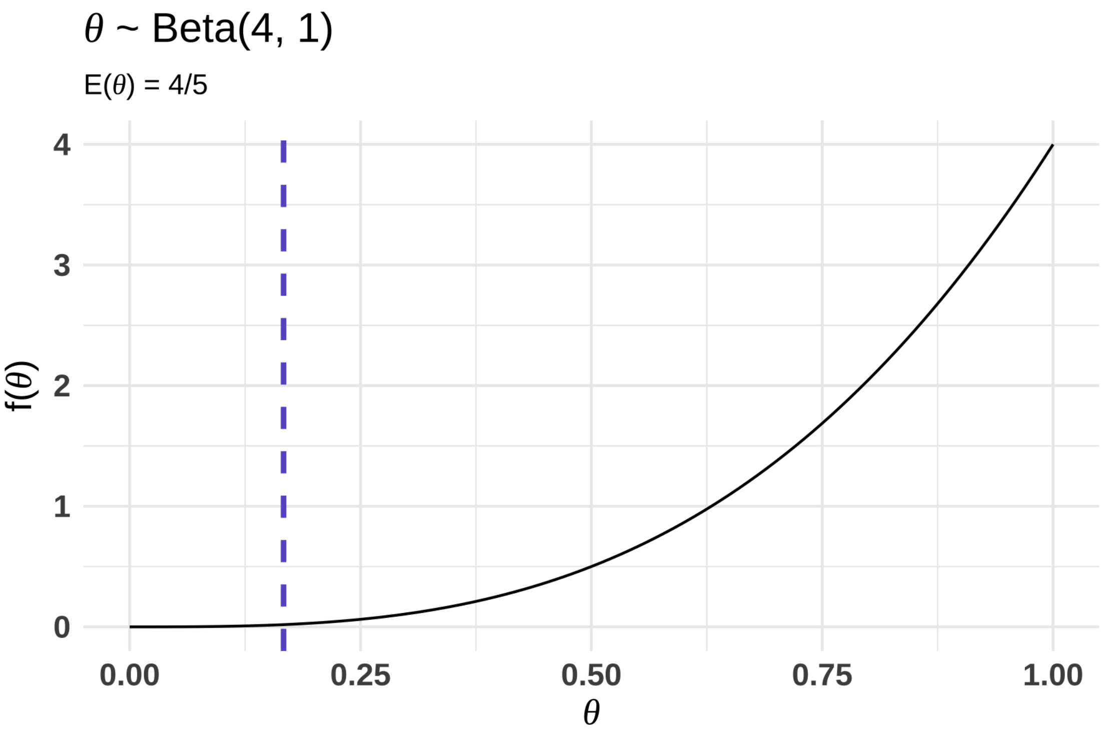
It is reasonable to use a **Beta prior** here. That is:

- $\theta \sim Beta(\alpha, \beta)$
  - In a *prior model*, α and β are called **hyperparameters**.
- $f(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0 \leq \theta \leq 1$
  - This PDF can tell us which values of θ are *more plausible* than others.

Assuming the α and β *hyperparameters* are *fixed* values, we can **tune** them to reflect our *prior understanding* about Animal Crossing popularity among students.

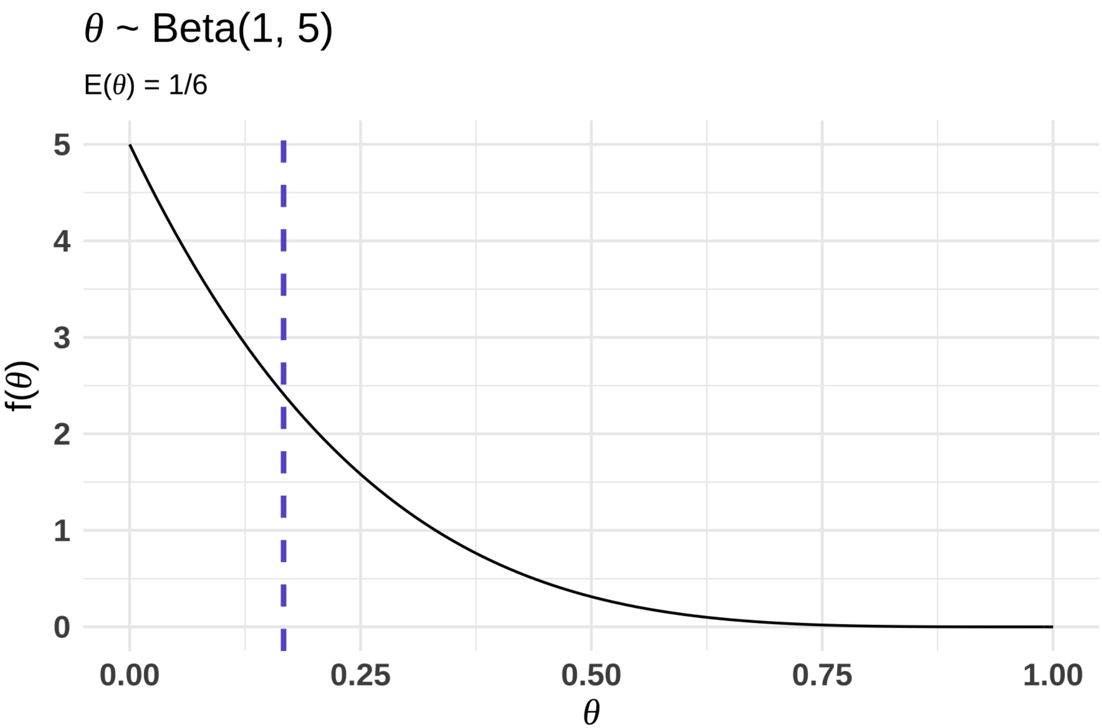
Different Beta Priors

> Based on anecdotal evidence, the students think that θ could reasonably range from 0.1 to 0.25.



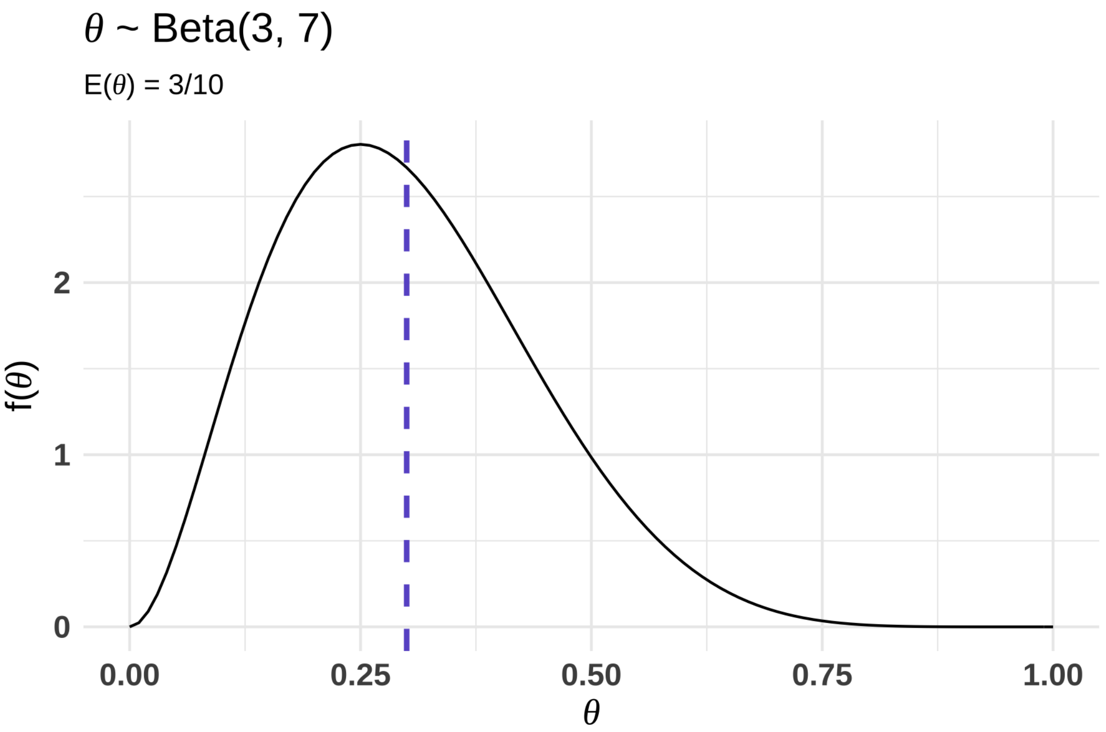
Different Beta Priors

> Based on anecdotal evidence, the students think that θ could reasonably range from 0.1 to 0.25.



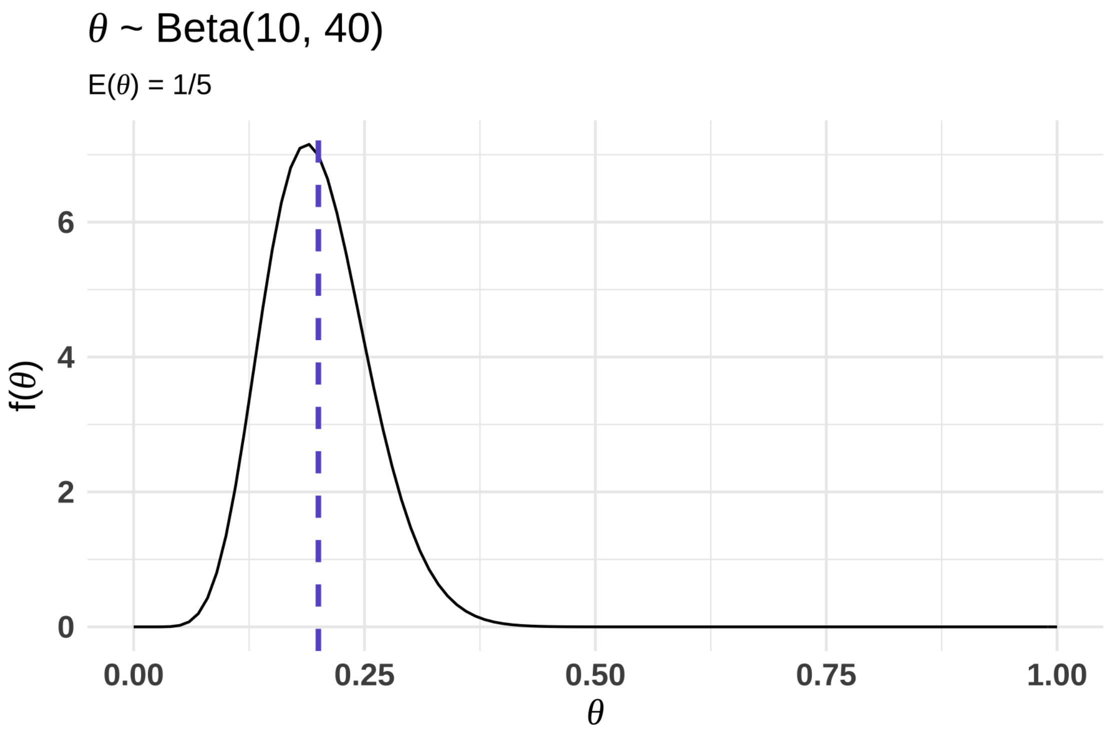
Different Beta Priors

> Based on anecdotal evidence, the students think that θ could reasonably range from 0.1 to 0.25.



Different Beta Priors

> Based on anecdotal evidence, the students think that θ could reasonably range from 0.1 to 0.25.



Prior Distribution

Let's work with the $Beta(10, 40)$ prior. That is:

- $\theta \sim Beta(10, 40)$
  - $E(\theta) = 0.2$
  - $Var(\theta) = 0.003 \implies SD(\theta) = 0.056$
- $f(\theta) = \frac{\Gamma(50)}{\Gamma(10)\Gamma(40)}\theta^9(1-\theta)^{39}, \quad 0 \leq \theta \leq 1$

🚨 This distribution represents our *prior assumptions* about the possible proportion of students who play Animal Crossing.

- It tends to deviate by ~6% from the prior mean of 20%.

The Data Model

In the next step of our Bayesian analysis, we're ready to collect some **data**!

- We'll take a *random sample* of $n = 30$ students, and let Y denote the number that play Animal Crossing.
- **Note**: The data, Y, depend on θ; the greater the *actual* proportion of students who play Animal Crossing, the greater Y will be.

Assumptions:

- Students are sampled *independently* from one another.
- The *probability* that any student plays Animal Crossing is fixed at θ.

A reasonable model for the data, Y, *conditional on* θ, is:

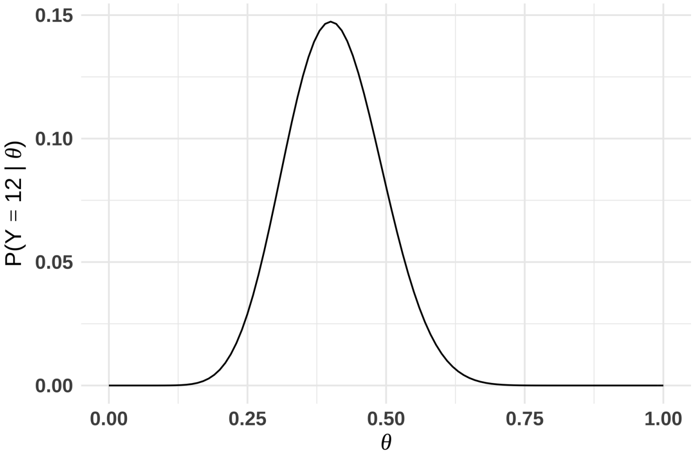
- $Y \mid \theta \sim Binomial(30, \theta)$
- $f(y \mid \theta) = P(Y = y \mid \theta) = \binom{30}{y} \theta^y (1-\theta)^{30-y}, \quad y = 0, 1, \ldots, 30$

The Data Model

The **likelihood**, $f(y \mid \theta)$, provides the probability of obtaining certain values of Y, *if the proportion of students who play Animal Crossing were some given value of* θ.

- **If** θ is low, then Y is more likely to be low.
- **If** θ is high, then Y is more likely to be high.

Suppose, *in reality*, we observe that $Y = 12$. That is, in our sample of 30 randomly selected students, 40% play Animal Crossing!

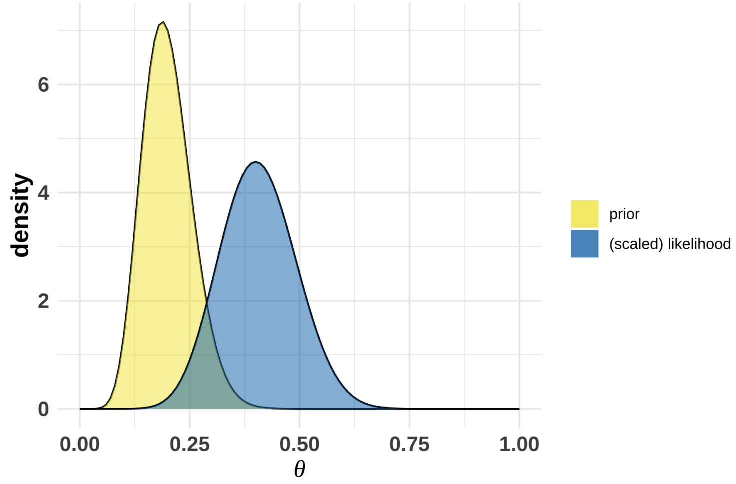


- $P(Y = 12 \mid \theta = 0.4) = \binom{30}{12} 0.4^{12}(1 - 0.4)^{18} \approx 0.147$

Summary (so far)

Let's recap what we have so far:

- $\theta \sim Beta(10, 40)$ is our **prior distribution** for θ
- $Y \mid \theta \sim Binomial(30, \theta)$ is the distribution for our **data**, Y, given θ



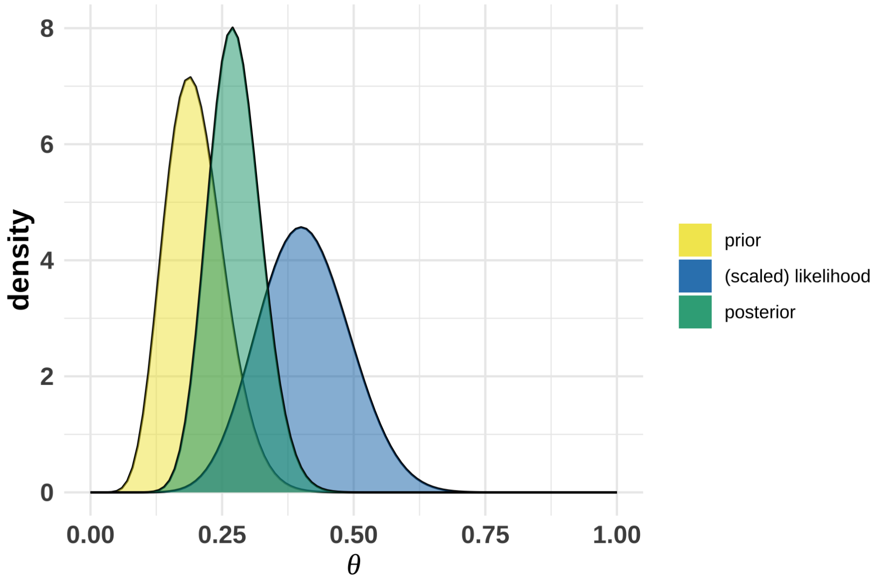
The **prior** and **data** aren't in *perfect* agreement!

- The prior generally assumes fewer students play Animal Crossing than the data suggest.
  - **That doesn't make the prior wrong!!!**

Posterior Distribution

The **prior** and **data** are both valuable to Bayesians.

- The **posterior** distribution will *combine* information from the prior and data.



It turns out...

- $\theta \mid Y \sim Beta(22, 58)$ is the **posterior distribution** of θ, *given Y*.
  - This is the *updated* distribution of θ that combines information from the prior and data.

Deriving the Posterior

We have:

- $\theta \sim Beta(10, 40)$
- $Y \mid \theta \sim Binomial(30, \theta)$

We can derive the **posterior distribution** using *Bayes' Rule...*

$$\begin{aligned}
f(\theta \mid y) &= \frac{f(y \mid \theta) f(\theta)}{f(y)} \\
&= \frac{\left(\binom{30}{12}\theta^{12}(1-\theta)^{18}\right)\left(\frac{\Gamma(50)}{\Gamma(10)\Gamma(40)}\theta^{9}(1-\theta)^{39}\right)}{f(y)} \\
&\propto \theta^{21}(1-\theta)^{57}
\end{aligned}$$

This is the **kernel** of a $Beta(22, 58)$ distribution!

- The remaining "stuff" that *doesn't depend on* θ is lumped into a **normalizing constant** so that $f(\theta \mid y)$ integrates to 1.

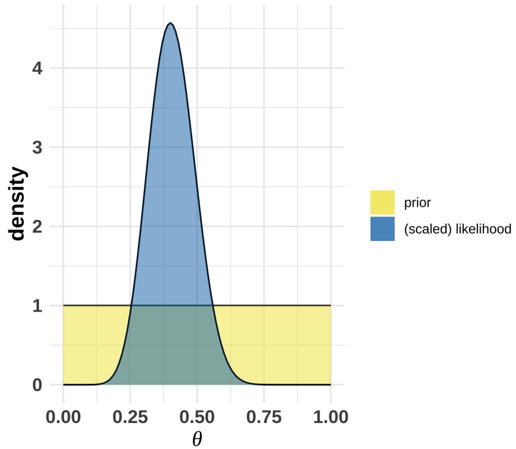
Flat (Uniform) Prior

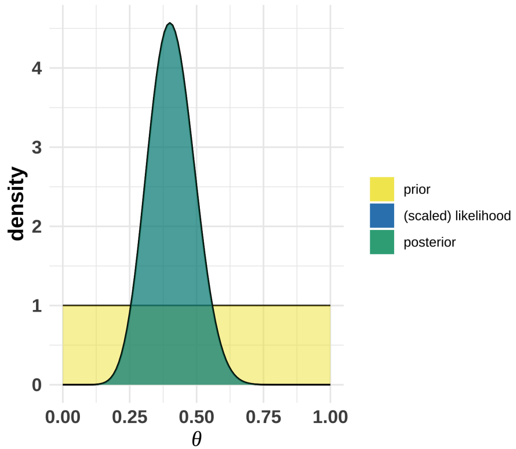
Suppose we had **absolutely no idea** how many students played Animal Crossing.

- It wouldn't really make sense to assign any particular $Beta(\alpha, \beta)$ distribution. How would we even know what to choose for α and β?

We *could* choose to assign a *uniform*, or **flat prior** to θ (which is technically a Beta(1, 1)). That is, let's assume the following hierarchy:

- $\theta \sim Uniform(0, 1) \implies f(\theta) = 1, \quad 0 \leq \theta \leq 1$
- $Y \mid \theta \sim Binomial(30, \theta)$

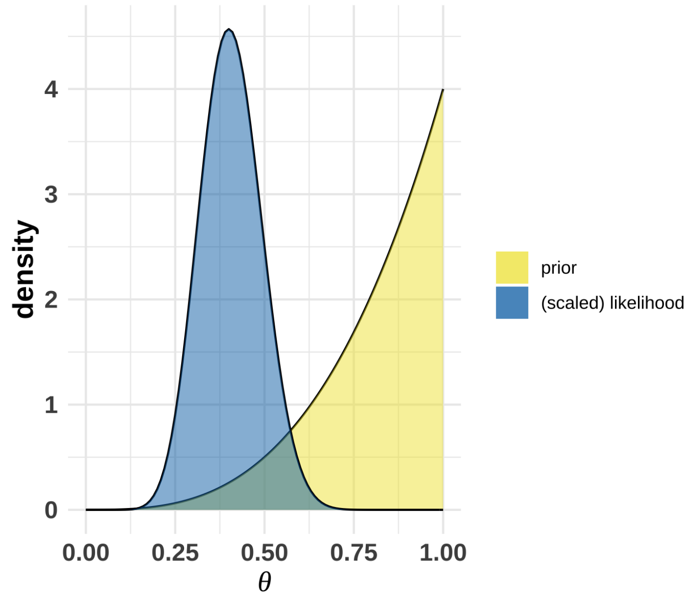


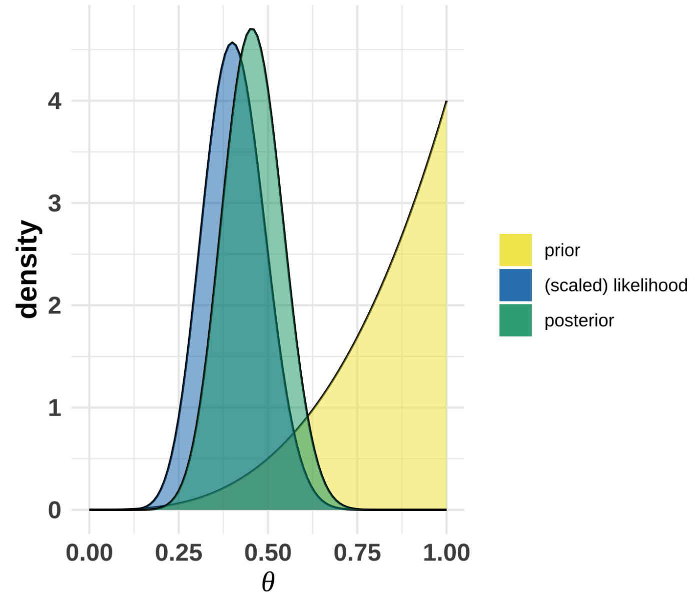


Other Beta Priors

Maybe instead of a *uniform prior*, we assign a different prior with *more variability* but *higher mean*:

- $\theta \sim Beta(1, 4)$
- $Y \mid \theta \sim Binomial(30, \theta)$





The Beta Binomial Model

We just worked with the **beta-binomial** Bayesian model! In general...

- **Prior**: $\theta \sim Beta(\alpha, \beta)$
- **Likelihood**: $Y \mid \theta \sim Binomial(n, \theta)$
- **Posterior**: $\theta \mid Y \sim Beta(\alpha + y, \beta + n - y)$

This model is very useful when:

- The *parameter* of interest, θ, is a number between 0 and 1.
- The data, Y, represents the number of "successes" in *n* independent **Bernoulli** trials.

Sequential Observations

Suppose that, in the previous example, we didn't observe all $n = 30$ observations at once.

- Rather, we observed 10 observations *each day, for three days*.

We still assume the following:

- **Prior**: $\theta \sim Beta(10, 40)$
- **Likelihood**: $Y \mid \theta \sim Binomial(n, \theta)$

But *now*, we observe the following data over three days:

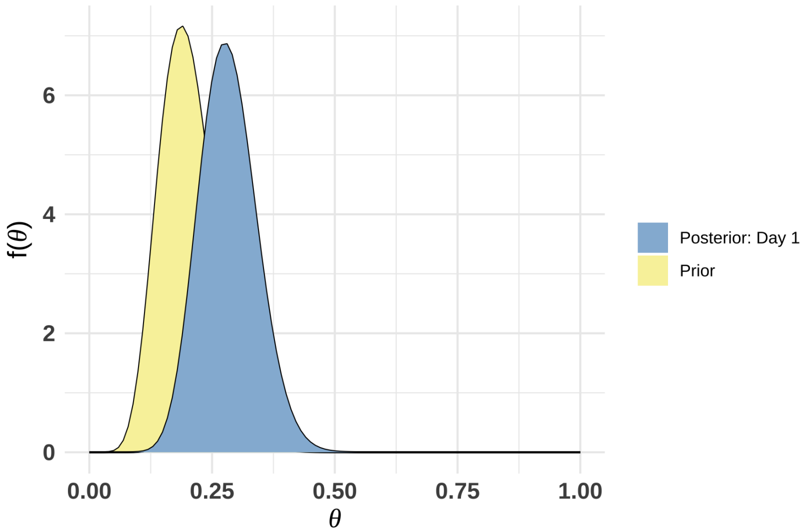
- **Day 1**: $n = 10, Y = 7$ play Animal Crossing
- **Day 2**: $n = 5, Y = 1$ play Animal Crossing
- **Day 3**: $n = 15, Y = 4$ play Animal Crossing

Each day, our understanding of θ evolves, *conditional on the previous day(s)*!

Sequential Observations

Using the general **Beta-Binomial** model from a previous slide, we can obtain the **posterior** for $\theta \mid Y$ after **Day 1**:

- **Prior**: $\theta \sim Beta(10, 40)$
- **Likelihood**: $Y \mid \theta \sim Binomial(10, \theta)$; we observe $Y = 7$ students who play AC
- **Posterior**: $\theta \mid Y \sim Beta(10 + 7, 40 + 10 - 7)$

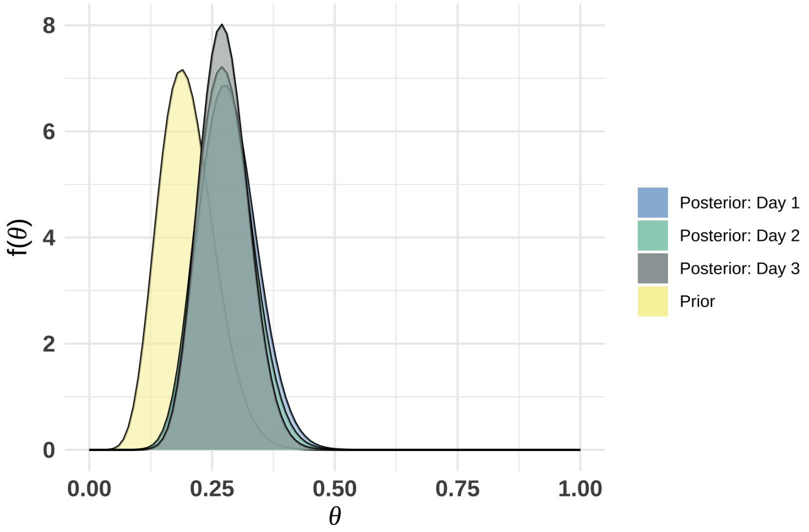


Sequential Observations

Each day, we *update* the posterior by essentially treating the *previous posterior* as a *new prior.*

In other words,

$$f(\theta \mid y_2) = \frac{f(y_2 \mid \theta) f(\theta \mid y_1)}{f(y_2)} = \frac{f(y_2 \mid \theta) \left[\frac{f(y_1|\theta) f(\theta)}{f(y_1)} \right]}{f(y_2)}.$$



Example

Suppose we want to estimate the lifetime (in hours), θ, of a certain electrical component.

Consider the following:

- **Prior**: $\theta \sim Gamma(\alpha, \beta)$, where

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

- **Likelihood**: $Y_1, Y_2, \dots, Y_n \mid \theta \sim Exponential(\theta)$, where

$$f(y_i \mid \theta) = \theta e^{-\beta\theta}$$

Let's derive the **posterior distribution**, $\theta \mid \mathbf{Y}$.

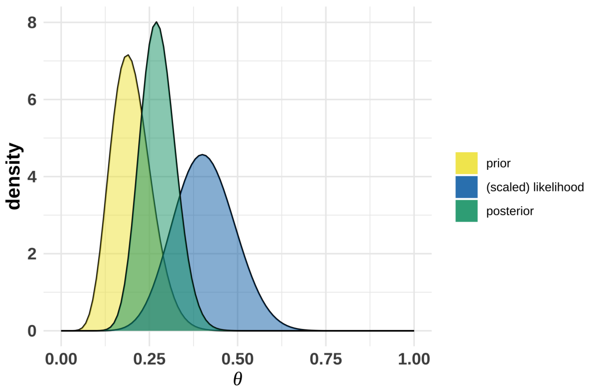
Conjugate Priors

Revisiting the Beta-Binomial

- **Prior**: $\theta \sim Beta(\alpha, \beta)$
- **Likelihood (Data)**: $Y \mid \theta \sim Binomial(n, \theta)$
- **Posterior**: $\theta \mid Y \sim Beta(\alpha + y, \beta + n - y)$

What's so great about this?!

- It's fairly **simple** to compute and work with.
- *Interpretability*
  - Posterior distribution can be understood as a *balance* between the **data** and **prior** models.



Conjugate Families

The beta-binomial Bayesian model is also a **conjugate family**.

Suppose that...

- The *prior model* for θ has PDF $f(\theta)$
- The *data model* for $Y_1, \dots, Y_n$ conditional on θ has likelihood function $f(\mathbf{y} \mid \theta)$.

If the resulting posterior model with PDF $f(\theta \mid \mathbf{y}) \propto f(\mathbf{y} \mid \theta) f(\theta)$ is of the *same model family as the prior*, then the prior is a **conjugate prior**.

We've already seen some examples!

- **Prior**: *beta*; **Data**: *binomial*; **Posterior**: *beta*
- **Prior**: *gamma*; **Data**: *exponential*; **Posterior**: *gamma*

These are wayyyyy simpler to work with than *non-conjugate priors*! For example...

A Non-Conjugate Prior

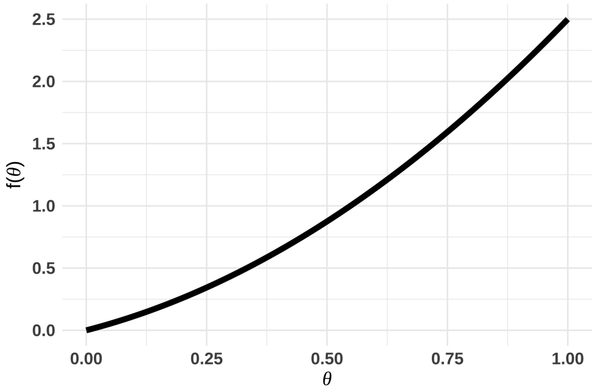
Suppose we still want to estimate the **proportion**, θ, of college students who play Animal Crossing.

- We still model the data Y, conditional on θ, as $Y \mid \theta \sim Binomial(n, \theta)$.

However, instead of $\theta \sim Beta$, we choose a different probability distribution:

$$f(\theta) = (3/2)\theta^2 + \theta, \quad 0 \leq \theta \leq 1$$

.



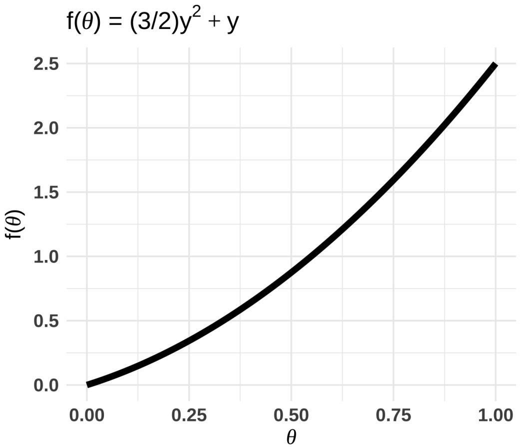
- We can use Bayes' Rule to derive the posterior distribution, but it's *not fun*!

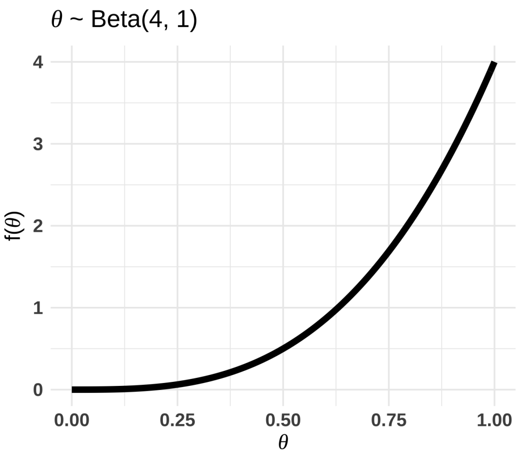
Non-Conjugate Priors

Recap:

- The calculation of the posterior was *not easy*!
  - As such, it's more challenging to develop an understanding of the posterior as a *balance* between the *prior* and *data* models.
- Because the posterior PDF is *messy*, it's more challenging to derive a **posterior mean**, mode, etc. (more on this in a bit!)

We *could* just use a **conjugate** beta prior!





Back to Piéchart Emporium

Piéchart
Emporium

Goal 👇👇👇

> Model rate λ, the *typical* number of customers at Pié Emporium on *weekday afternoons.*

Prior to collecting *data*, I'm guessing that the rate λ could be anywhere between 3 to 9 customers. Because it's *the place to be.* 😎

- To learn more, I record the number of weekday afternoon customers on each of n days, $Y_1, Y_2, \ldots, Y_n$.

> Why *shouldn't* we model the **data** with a *binomial distribution*?.
>
> Why *shouldn't* we use a **beta prior** for λ?

Potential Data Models

Each data point, Y_i, is a **count** representing the number of customers *observed* on a given weekday afternoon.

- Y_i can range from **0** to something *very large*.

The **Poisson** distribution is useful for modeling the data, Y_i, *conditional on* λ:

- $Y_i \mid \lambda \sim Poisson(\lambda)$
- $f(y_i \mid \lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad y = 0, 1, 2, \ldots$

Potential Priors

The rate parameter, λ, represents the typical number of customers on a weekday afternoon.

- λ is *positive* and *continuous.*

There are a few distributions that satisfy this property (i.e., *continuous* with support **> 0**).

- But let's try to choose a useful *conjugate* prior to use with the Poisson data model.

A **Gamma** prior for λ would work here! (trust me)

- $\lambda \sim Gamma(\alpha, \beta)$
- $f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}, \quad \lambda > 0$

Gamma-Poisson Conjugacy (aka "The Logo")

Let $\lambda > 0$ be an unknown *rate* parameter and $(Y_1, Y_2, \ldots, Y_n)$ be iid Poisson(λ) observations. In other words:

- $Y_i \mid \lambda \sim iid\ Poisson(\lambda)$
- $\lambda \sim Gamma(\alpha, \beta)$

Upon observing the **data** $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, the **posterior distribution** for λ also follows a Gamma distribution with *updated parameters*:

- $\lambda \mid \mathbf{y} \sim Gamma(\alpha + \sum_{i=1}^{n} y_i, \beta + n)$

Tuning the Prior

While we originally derived the Gamma-Poisson conjugacy in general terms, let's tune our Gamma prior to reflect our prior beliefs about weekday afternoon customers:

> I'm guessing that the rate λ could be anywhere between 3 to 9 customers.

If $\lambda \sim Gamma(\alpha, \beta)$, then:

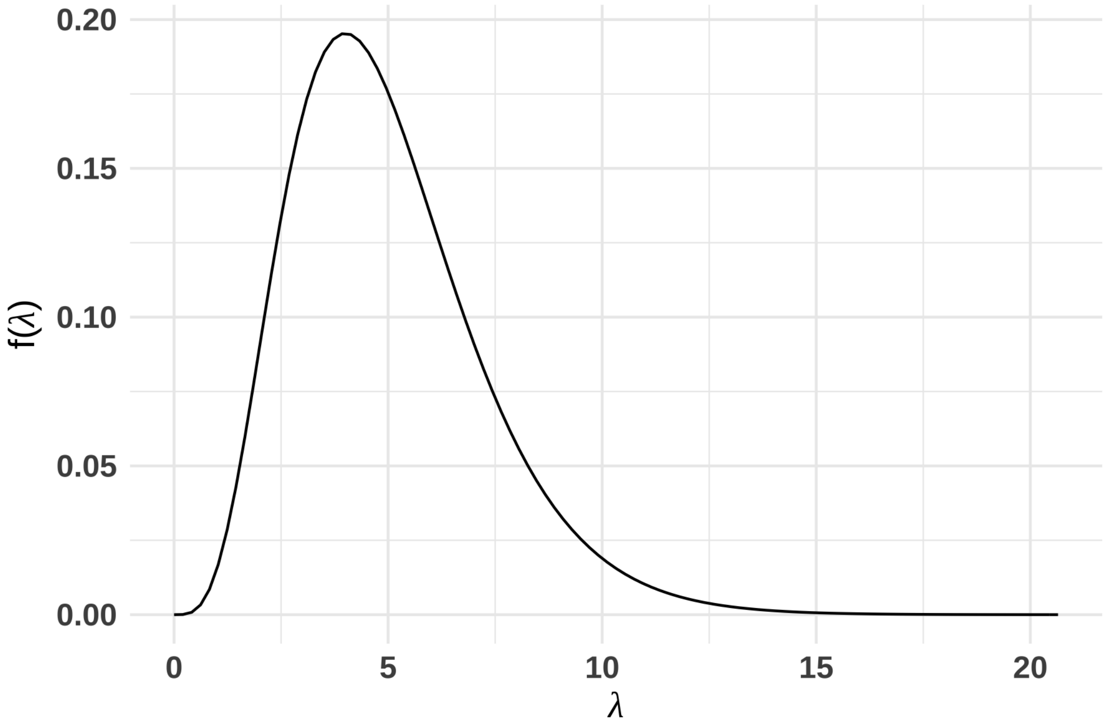
- $E(\lambda) = \alpha\beta$
- $Var(\lambda) = \alpha\beta^2$

Let's try to choose α and β such that $E(\lambda) \approx 5$ and $Var(\lambda) \approx 3$

Tuning the Prior

A Gamma(5, 1) prior could also work:

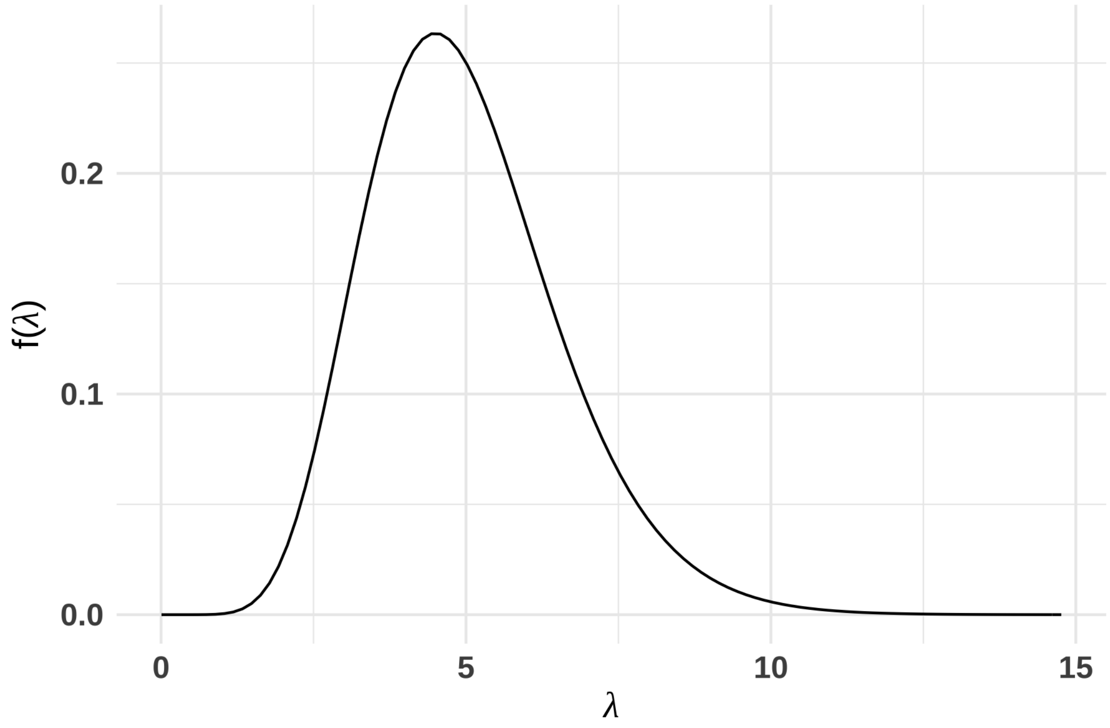
```
bayesrules::plot_gamma(shape = 5, rate = 1)
```



Tuning the Prior

A Gamma(10, 2) prior (where 2 is the *rate* parameter*) could also work:

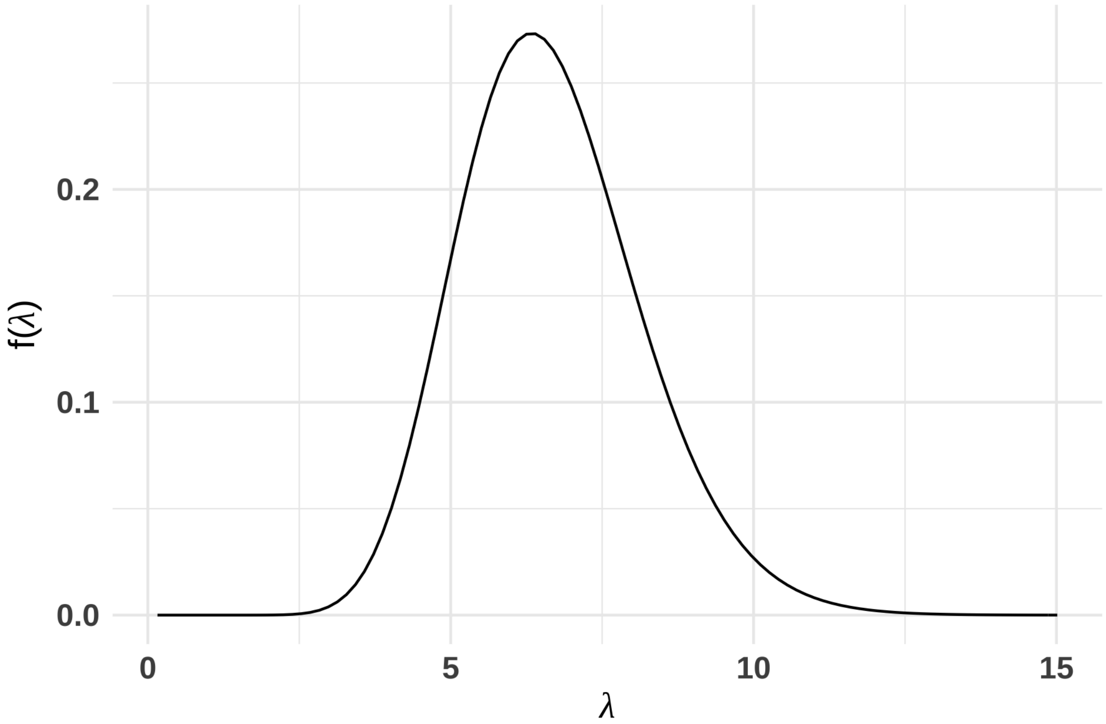
```
bayesrules::plot_gamma(shape = 10, rate = 2)
```



Tuning the Prior

Maybe a Gamma(20, 3) prior?

```
bayesrules::plot_gamma(shape = 20, rate = 3)
```



Onto the DATA

Let's stick with $\lambda \sim Gamma(20, 33)$.

Now suppose we record the number of customers for *five* weekday afternoons:

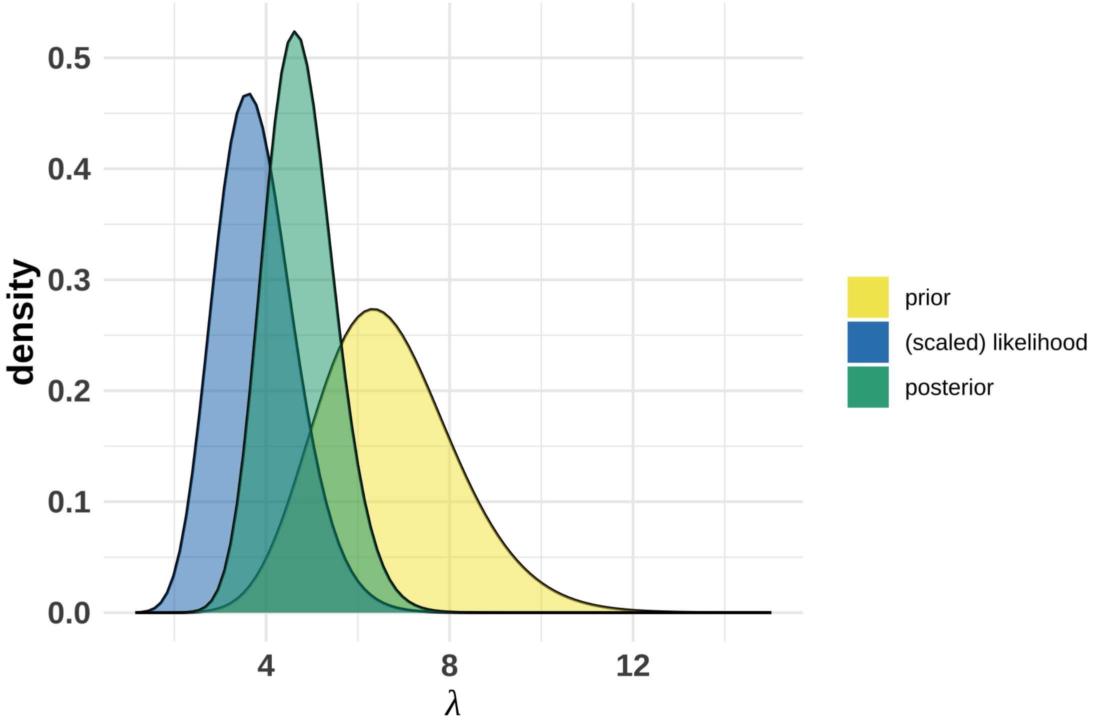
- $(Y_1 = 2, Y_2 = 5, Y_3 = 4, Y_4 = 2, Y_5 = 5)$
- In other words: $n = 5$ and $\sum_{i=1}^{5} y_i = 18$

That means $\lambda \mid \mathbf{y} \sim Gamma(20 + 18, 3 + 5)$!

Gamma-Poisson Conjugacy (aka "The Logo")

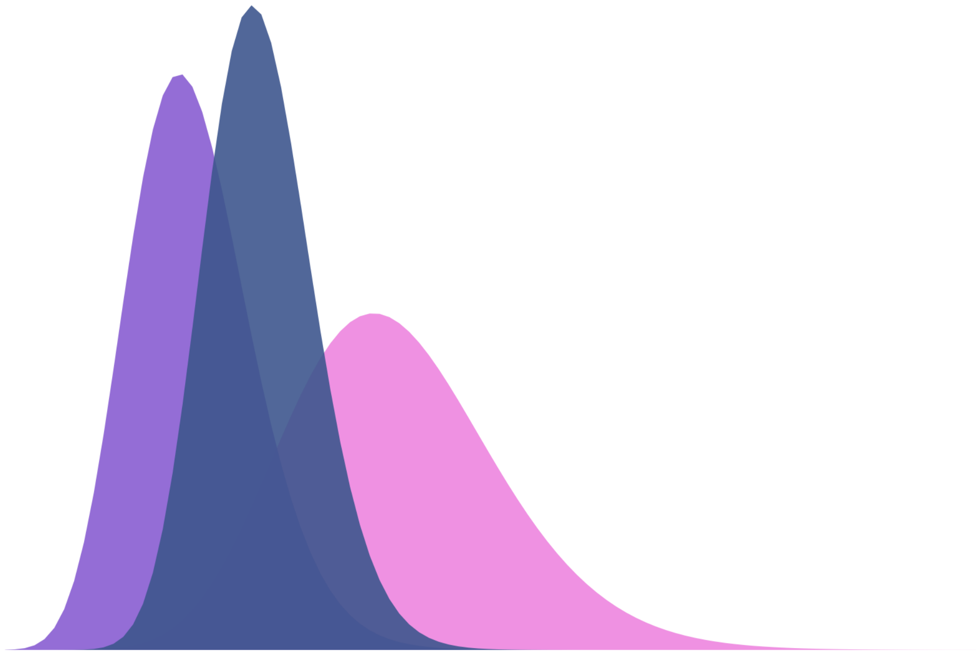
$\lambda \mid \mathbf{y} \sim Gamma(20 + 18, 3 + 5)$

```
bayesrules::plot_gamma_poisson(shape = 20, rate = 3, sum_y = 18, n = 5)
```



Gamma-Poisson Conjugacy (aka "The Logo")

$\lambda \mid \mathbf{y} \sim Gamma(20 + 18, 3 + 5)$



Critiques of Conjugate Familes

Conjugate families can be very **convenient** to work with, but they are not without their limitations!

- Sometimes a *conjugate prior* is simply not as appropriate as a non-conjugate prior.
    - Maybe the best fit for our prior understanding *isn't* a Gamma or Beta model...
- We cannot always choose a **flat prior** in a conjugate family.
    - Because Uniform(0, 1) = Beta(1, 1), a *uniform prior* is conjugate if the data are modeled with a *binomial* distribution.
    - But a Uniform(0, 1) isn't conjugate if the data are modeled with a Poisson distribution!

💡 One potential workaround could be to just choose a non-uniform prior with **high variance**...

Improper Priors

If we can't use a **flat prior** in a conjugate family, we could also use an **improper prior**.

- An *improper prior distribution* (like flat priors) capture the idea that the *data are worth more than our prior understanding.*

An **improper prior** has a PDF that *does not* integrate to 1. In other words, we are using an improper prior for θ if

$$\int_\theta p(\theta)\, d\theta = \infty.$$

- Usually we can obtain an improper prior by replacing a *conjugate prior's* hyperparameter(s) with 0.
- Beta(0, 0), Gamma(0, 0), Normal(μ, $\sigma^2 = \infty$)

Improper Gamma Prior

Suppose in our Piéchart Emporium customer example, we obtain customer counts on $n = 150$ days, where $\sum_{i=1}^{150} y_i = 1100$.

- If we model $Y_1, \ldots, Y_{150} \mid \lambda$ using a Poisson(λ) distribution, the Gamma(α, β) is a *conjugate prior*.
- We could *also* use an **improper** Gamma(0, 0) prior, with "pdf" $f(\lambda) = \lambda^{-1}$

We can apply Bayes' Rule and obtain:

$$\begin{aligned} f(\lambda \mid \mathbf{y}) &\propto f(\mathbf{y} \mid \lambda) f(\lambda) \\ &= Gamma(n, \sum_i y_i) \end{aligned}$$

Bayes Estimators

Estimating θ

Recall: Rather than treat the parameter θ as a *fixed value*, a *Bayesian framework* assumes that θ is a *random variable* with a probability distribution.

- How can we *estimate* θ in a Bayesian framework?

In the **Animal Crossing** example, we used the following model:

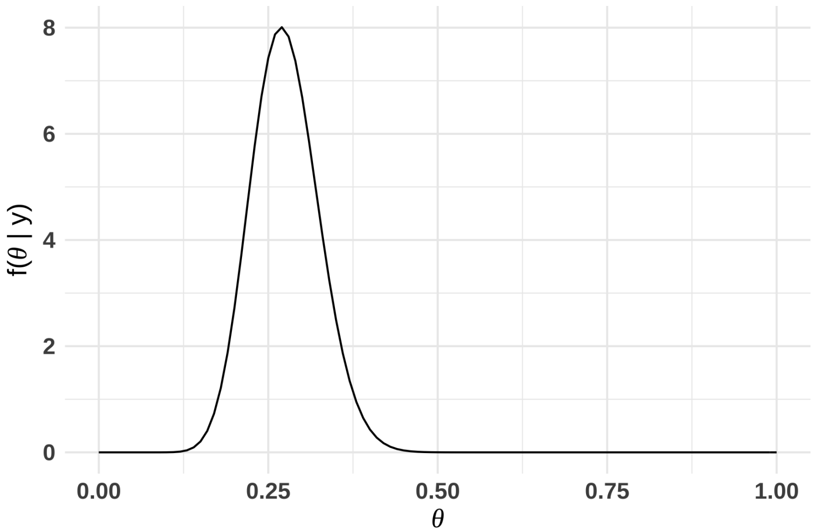
- $\theta \sim Beta(10, 40)$
  - This distribution represents our *prior assumptions* about the possible proportion of students who play Animal Crossing.
- $Y \mid \theta \sim Binomial(30, \theta)$
  - This is the distribution for our **data** (the number of students in our sample who play AC), Y, given θ.
  - In our sample of $n = 30, Y = 12$ students played AC.

Estimating θ

- $\theta \sim Beta(10, 40)$
- $Y \mid \theta \sim Binomial(30, \theta)$

Because this is a **conjugate family**, we derived the following posterior distribution for $\theta \mid Y$:

$$\theta \mid Y \sim Beta(10 + 12, 40 + 30 - 12)$$



- What metric(s) can we use to *summarize* $\theta \mid Y$?

Bayes Estimator

While there are many different types of **Bayes estimators** for θ, we will use the *posterior expected value*:

> Let $Y_1, Y_2, \ldots, Y_n$ be a random sample with likelihood function $f(\mathbf{y} \mid \theta)$, and let θ have prior density $f(\theta)$. The **posterior Bayes estimator** for θ is given by
>
> $$\hat{\theta}_B = E(\theta \mid \mathbf{Y})$$

Example

In the Animal Crossing Example, our posterior distribution for $\theta \mid Y$ was

$$\theta \mid Y \sim Beta(22, 58).$$

- Therefore, $\hat{\theta}_B = E(\theta \mid Y) = 22/(22+58) = 0.275.$

Bayes Estimator for Beta-Binomial

In general, the Beta-Binomial model consists of the following:

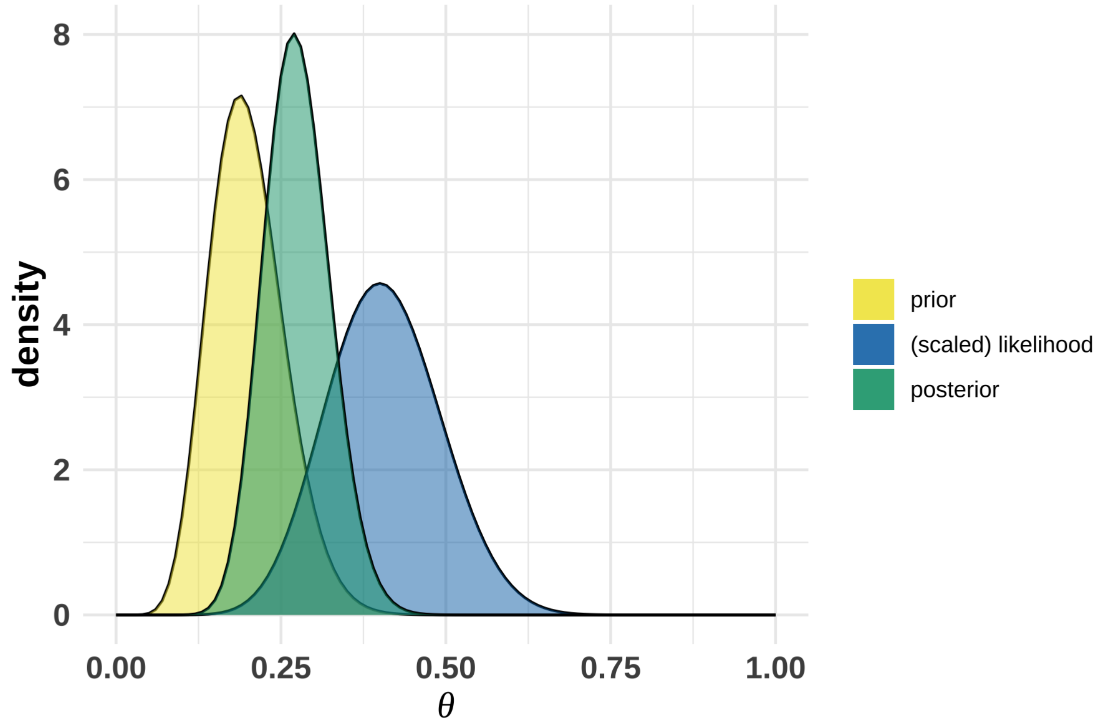
- $\theta \sim Beta(\alpha, \beta)$
- $Y \mid \theta \sim Binomial(n, \theta)$
- $\theta \mid Y \sim Beta(\alpha + y, \beta + n - y)$

Therefore the **Bayes estimator** (posterior expected value), $\hat{\theta}_B$ is:

$$\hat{\theta}_B = E(\theta \mid Y = y) = \frac{\alpha + y}{\alpha + \beta + n}$$

Posterior as a Balance

A great thing about Bayesian estimation (especially when working with *conjugate families*) is that we can think of the *posterior* distribution as a **balance** between the *data* and *prior*.



Posterior as a Balance

A great thing about Bayesian estimation (especially when working with *conjugate families*) is that we can think of the *posterior* distribution as a **balance** between the *data* and *prior*.

But let's see what is going on with the expected value of θ...

- $\theta \sim Beta(\alpha, \beta) \implies E(\theta) = \frac{\alpha}{\alpha+\beta}$
- $\theta \mid Y \sim Beta(\alpha + y, \beta + n - y) \implies E(\theta \mid Y) = \frac{\alpha+y}{\alpha+\beta+n}$

The posterior mean is actually a **weighted average** between the prior and data!

In the Animal Crossing example...

- $\theta \sim Beta(10, 40) \implies E(\theta) = 0.2$
- $Y = 12$ out of $n = 30$ (40% of sample plays Animal Crossing)
- $\theta \mid Y \sim Beta(22, 58) \implies E(\theta \mid Y) = 0.275$

Sensitivity of Estimators

How sensitive are our results to *different priors*?

- Either way, we observe $Y = 12$ Animal Crossing players out of 30, but let's play with different priors.

Functions of θ

We can also derive Bayes estimators for *functions of* θ.

Example: Using the Beta-Binomial conjugate family, find the Bayes estimator for $\theta(1-\theta)$.

- **Note**: $\theta(1-\theta)$ is the *variance* of a Bernoulli RV with "success" probability θ.

In general we can calculate

$$\theta\widehat{(1-\theta)}_B = E(\theta(1-\theta) \mid Y)$$

using the fact that $\theta \mid Y \sim Beta(\alpha + y, \beta + n - y)$.

Posterior Median

While the posterior mean generally provides a solid summary metric for $\theta \mid Y$, other Bayes estimators exist!

- For example, we could calculate the **posterior median**.

The *posterior median* isn't as straightforward to calculate as the *posterior mean*, but we could *estimate* it **via simulation**.

- If $\theta \mid Y \sim Beta(22, 58)$, we can estimate the posterior median with R:

```r
median(
  rbeta(n = 10000, shape1 = 22, shape2 = 58)
)
```

```
## [1] 0.2737252
```

- Or we could just find it **exactly**:

```r
qbeta(0.5, shape1 = 22, shape2 = 58)
```

```
## [1] 0.2731171
```

Gamma-Poisson Bayes Estimator

The **Gamma-Poisson** conjugate family:

- $\theta \sim Gamma(\alpha, \beta)$
  - Using the alternate version of the Gamma PDF where $E(\theta) = \alpha / \beta$
- $\mathbf{Y} \mid \theta \sim Poisson(\theta)$
- $\theta \mid \mathbf{Y} \sim Gamma(\alpha + \sum_i y_i, \beta + n)$

> What is the **Bayes estimator** for θ? 🤔