Statistical Inference, pt. II

ECON 3640–001

Marcio Santetti

Spring 2022

Motivation

Housekeeping

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

- Chapter 3
- Available `here`

Last time...

Last time, we studied our first **Bayesian** model.

Our purpose was to figure out what is the underlying proportion of Bayesians to frequentists in the Social Sciences.

We started with the **prior** belief that any value of this proportion (θ) is equally likely.

But does it have to be this way?

Updating our prior

Updating our prior

Last time, we defined our prior as

$$\theta \sim \mathrm{Unif}(0, 1)$$

However, assuming that the plausibility of **no** scholars being Bayesians is the *same* as **all** researchers being Bayesians is quite *imprecise* and *uninformative*.

So let us incorporate some **prior** information into our model.

In a recent survey, it was found that 75% of interviewed researchers used more *frequentist* than *Bayesian* methods in their research agendas.

How can we use this **previous knowledge** and interact it with **new data**?

Updating our prior

In order to translate this prior information into a probability distribution, we need some *specific distribution*, lying from 0 to 1, that allows us to move beyond a "*flat*" prior setting.

- Say hello to the **Beta distribution**!

A **Beta** random variable is *continuous*, and lies on the [0,1] interval.

Therefore, it should satisfy our needs of a more **informative** prior to conduct our analysis.

A random variable X follows a *Beta* distribution with **shape** parameters α and β:

$$X \sim \mathrm{Beta}(\alpha, \beta)$$

Updating our prior

A Beta-distributed random variable has a **probability density function** (PDF) as follows:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\; x^{\alpha-1}\,(1-x)^{\beta-1}$$

for $x \in [0, 1]$.

$\Gamma(\cdot)$ is called a *gamma function*.

If *x* is a positive integer, it simplifies to $\Gamma(x) = (x-1)!$

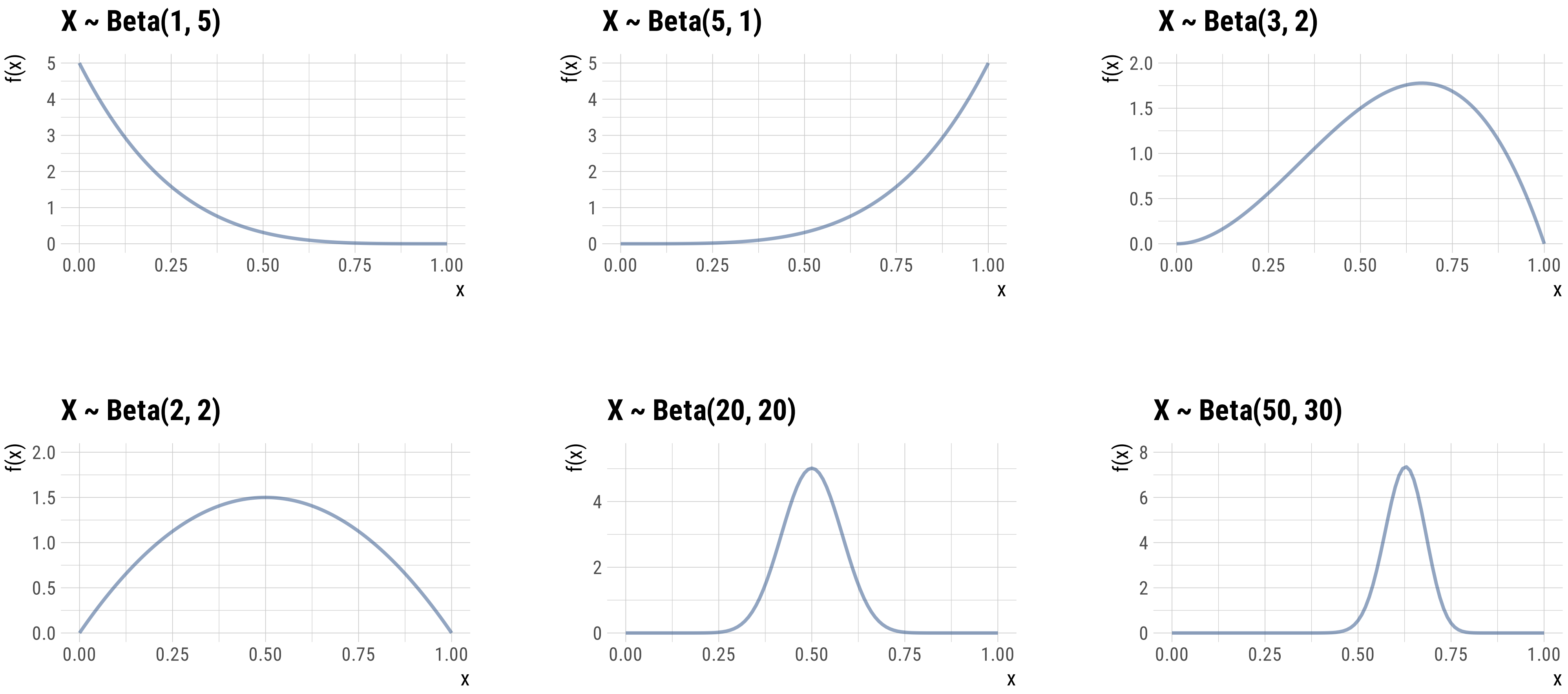
Updating our prior

We will call α and β as the Beta distribution's **hyperparameters**.

> A **hyperparameter** is a parameter used in a prior probability model.

Depending on how one **tunes** these hyperparameters, the Beta distribution's PDF will have different **shapes**.

Updating our prior



Updating our prior

Try out the following code:

```r
library(tidyverse)

data <- tibble(x = c(0,1))

data %>%
  ggplot(aes(x = x)) +
  stat_function(fun = dbeta, args = list(shape1 = 1, shape2 = 1), size = 1) +
  labs(x = "x",
       y = "f(x)")
```

What do you get?

Updating our prior

From these distributions, we can extract some measures of central tendency:

- **Expected Value** (i.e., *mean*);
- **Mode** (most *plausible* value)

For a Beta distribution, these can be calculated as:

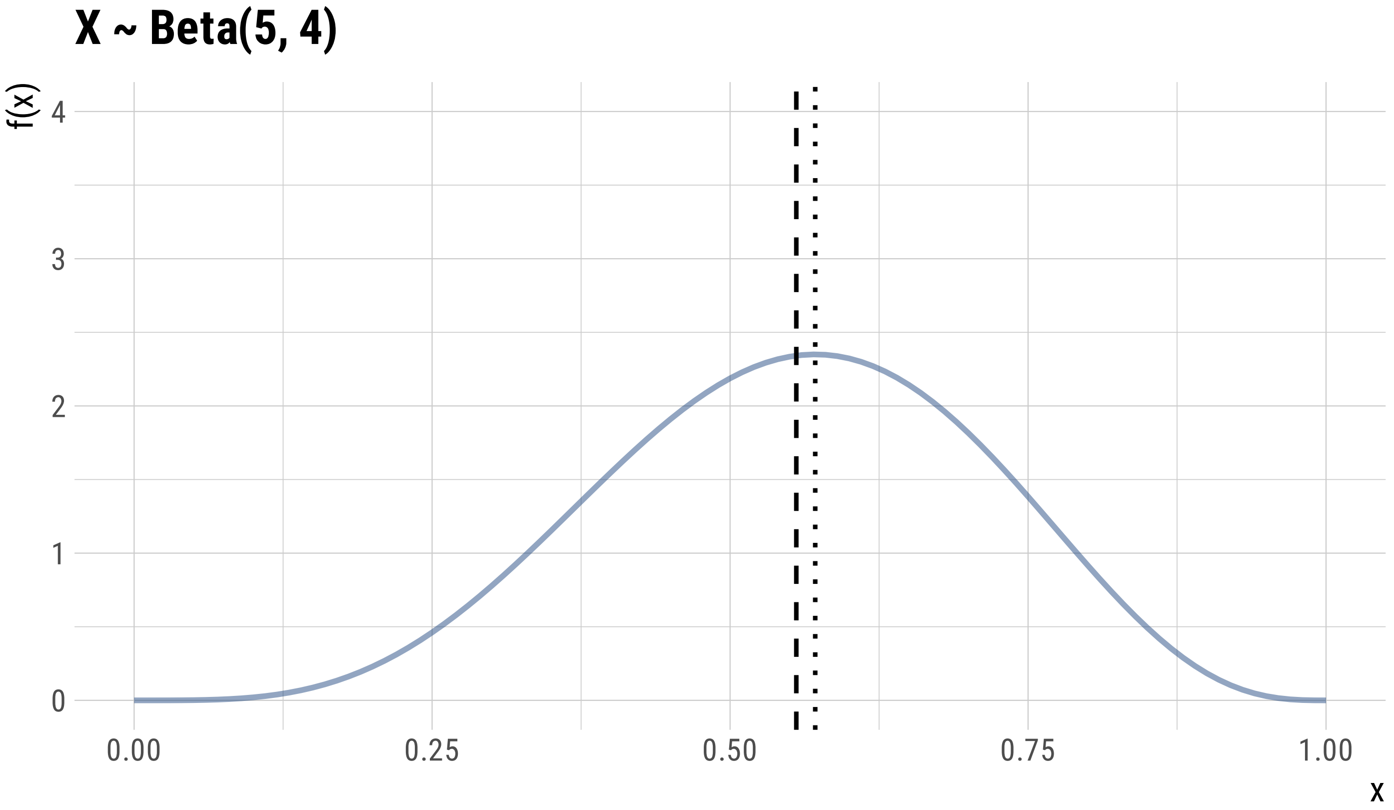
$$E(x) = \frac{\alpha}{\alpha + \beta}$$

$$\mathrm{Mode}(x) = \frac{\alpha - 1}{\alpha + \beta - 2}$$

As an example, what are the expected value and mode for a *Beta(5, 4)* random variable?

Updating our prior

Dashed: expected value; dotted: mode



Updating our prior

The **variance** of a Beta-distributed random variable is calculated as follows:

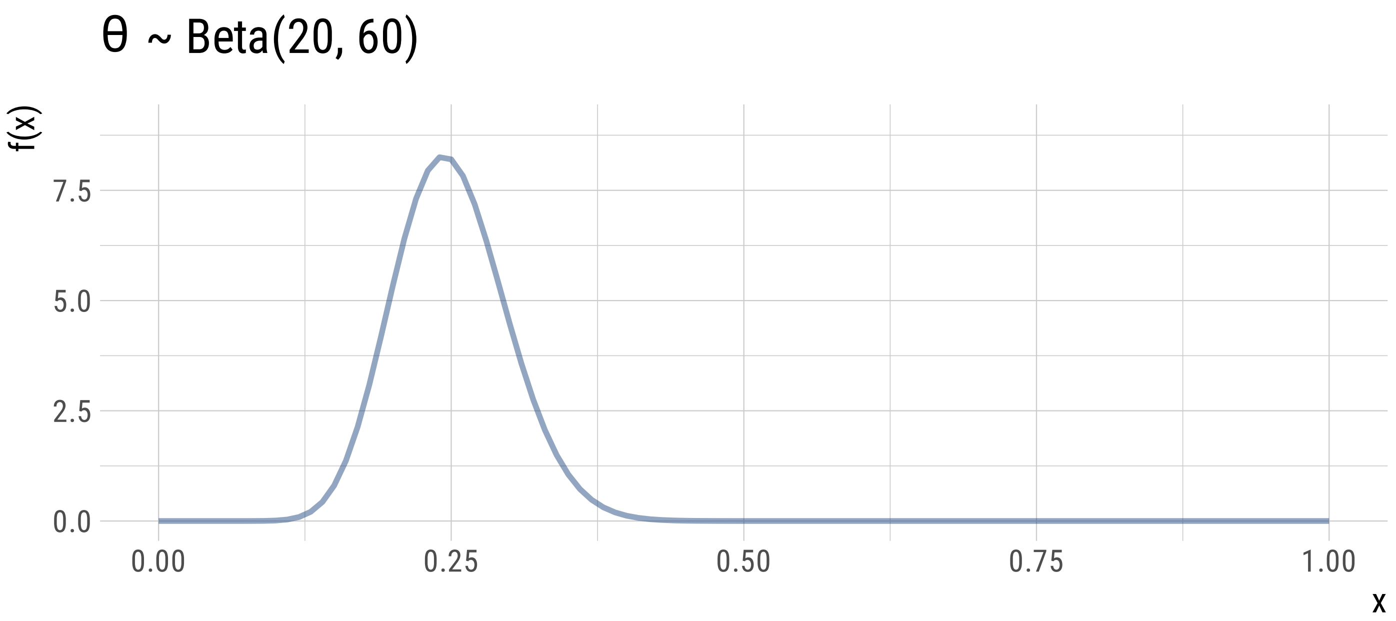
$$\mathrm{Var}(x) = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}$$

Hyperparameter tuning

Hyperparameter tuning

Now, it is time to **update** our previous "*flat*" prior, in order to incorporate previous knowledge about our proportion of interest.

Practically, what we need is to properly **tune** the α and β hyperparameters.



The likelihood function

The likelihood function

As our prior is set up, it is time to incorporate **new data** into our analysis.

Assume we interview 100 scholars.

We can assume that each individual we interview will answer the question *independently* of what other people will say.

Thus, we may define our **Binomial** likelihood function as

$$P(y|\theta) \sim \mathrm{Binomial}(100, \theta)$$

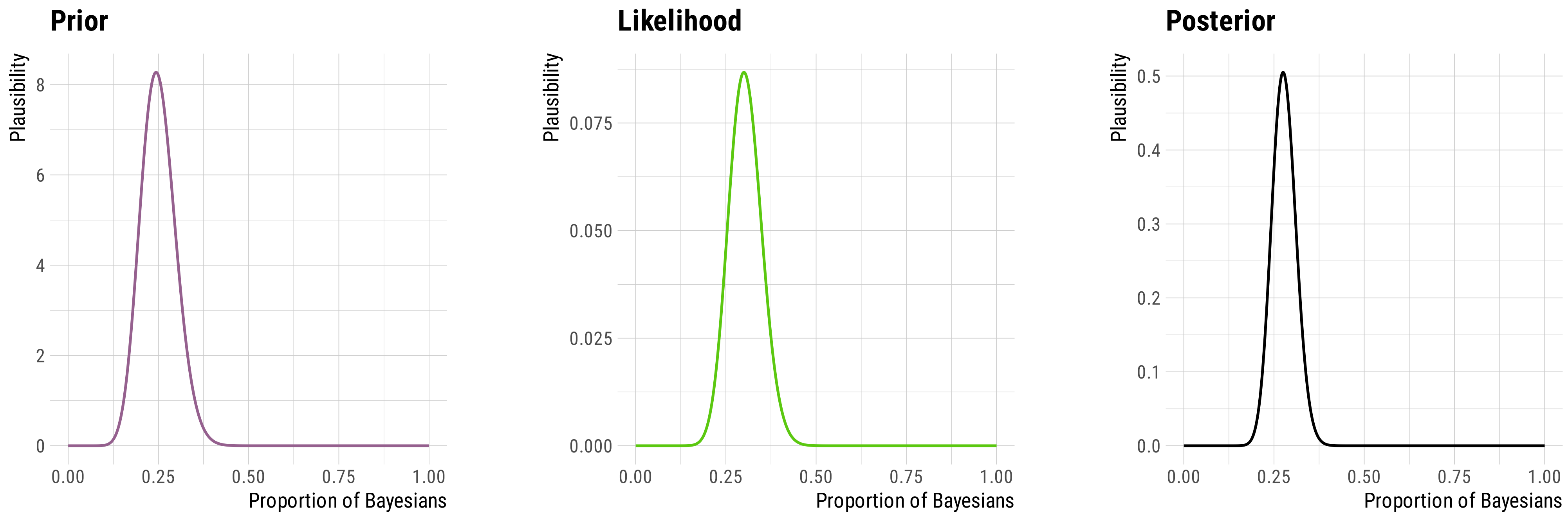
The likelihood function

By the end of this new survey, we find out that 30 people answered "Bayesian."

Let us **update** our prior!

The posterior

The posterior

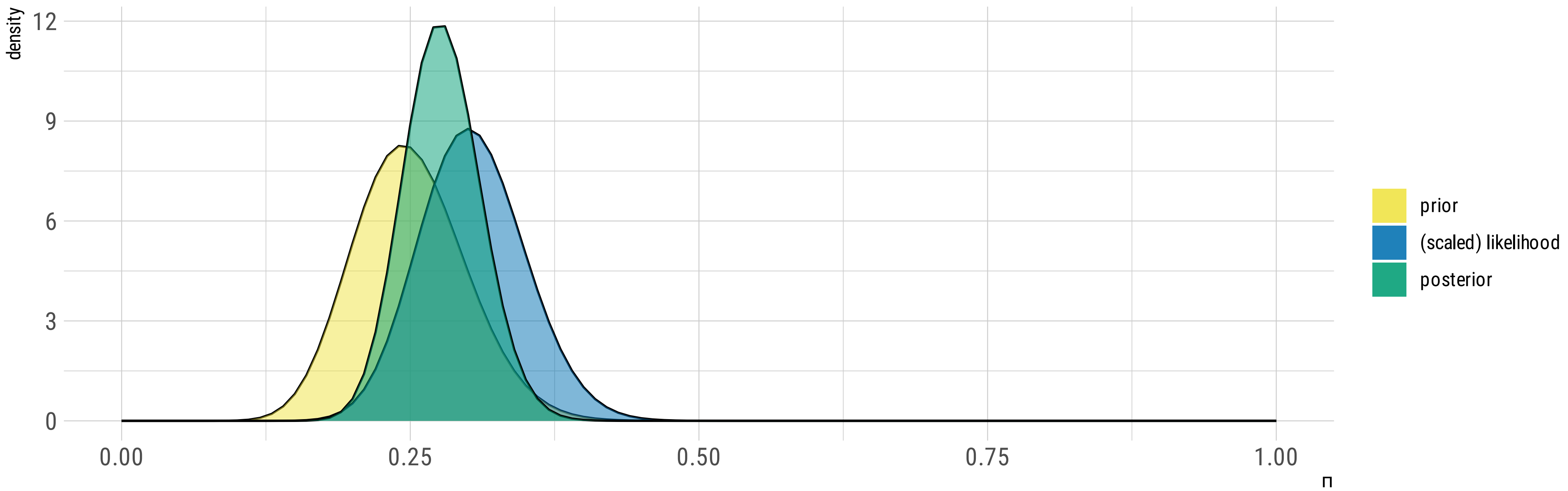


The posterior

All in one picture…

```
library(bayesrules)

plot_beta_binomial(alpha = 20, beta = 60, y = 30, n = 100)
```



The posterior

We are ready for some (*posterior*) **inference**:

```r
library(bayesrules)

summarize_beta_binomial(alpha = 20, beta = 60, y = 30, n = 100)
```

```
#>       model alpha beta      mean      mode         var         sd
#> 1     prior    20   60 0.2500000 0.2435897 0.002314815 0.04811252
#> 2 posterior    50  130 0.2777778 0.2752809 0.001108383 0.03329238
```

The Beta-Binomial model

The Beta-Binomial model

This example illustrated the **Beta-Binomial model**.

In other words, we have **combined** a *Beta prior probability model* with a *Binomial likelihood function*.

When combining these two, our posterior will follow a **Beta distribution** as follows:

$$\theta \mid y \sim \mathrm{Beta}(\alpha + y, \beta + n - y)$$

where y is the number of successes, and n is the number of trials from the binomial experiment.

The Beta-Binomial model

The Beta-Binomial model is an example for **conjugate priors**.

It simply means that, when combining a specific prior with a specific likelihood function, the posterior will follow the **same** distribution as the prior's.

> $P(\theta)$ is a conjugate prior for $P(y \mid \theta)$ if the posterior, $P(\theta \mid y) \propto P(\theta)\, P(y \mid \theta)$ is from the same model family as the prior.

Next time: More conjugate families