Bayesian Statistics

The Beta-Binomial Model & Sequential Testing

Alexandra Sarafoglou

June 3, 2026

Building a Bayesian model

- We assigned probabilities using a single number (e.g., $P(B) = 0.40$)
- However, in reality we often cannot be that precise
- Instead of numbers, we express the uncertainty about a parameter using probability distributions

Intended learning outcome

- You will learn how to interpret and tune a continuous Beta prior model to reflect your prior information about a probability π
- Construct the Beta-Binomial model for proportion π

Before we start...

- In Bayesian statistics, we treat unknown quantities as random variables to which we assign prior probabilities
- We can place **prior probabilities** on:
  - **Parameters** (e.g., π, μ, θ)
  - **Hypotheses** (e.g., H_0 vs H_1)

Focus of this course:

- We will primarily focus on placing priors on **parameters**
- This allows us to compute the **posterior distribution**:

$$p(\pi \mid y) \propto p(\pi) \times L(\pi \mid y)$$

Running Example: The Crisis of Confidence in the Social Sciences

A growing awareness (since the early 2010s) that many published findings in psychology and other social sciences fail to replicate.

- Replication attempts often yield weaker or non-significant results. Potential contributors: p-hacking, small sample sizes, publication bias, but also the tendency to publish surprising and counter-intuitive findings.
- It has been shown that social science experts can predict whether a replication will be successful or not.

How do we compare to social science experts? What is the **prediction ability** π in our Bayesian group?

Prior probability model

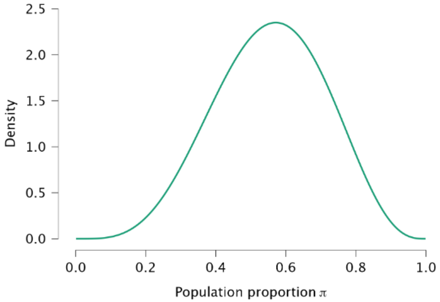
- Based on the last lecture, we could simply take our best guess (or mode, mean, or median if available) to construct our prior probability model that we predict a replication outcome correctly: $P(A) = 0.55$ and $P(A^c) = 0.45$.
- However, this neglects the uncertainty around our beliefs.
- For that reason, we choose a distribution (instead of a single number) to reflect our prior beliefs.

Prior probability model

A valid probability model must:

1. **Account for all possible events:** our ability π to predict replication success can take *any value* between 0 and 1.
2. **Assign probabilities to each event:** We can construct a continuous prior probability model π, which represents our prior beliefs as a curve.
3. **These probabilities sum to one:** For the continuous probability model, the curve does not sum to one, but it must *integrate* to one. Integration is the continuous analogue of summation: instead of adding up probabilities at discrete points, we integrate a probability density function (pdf), $f(\pi)$, over all possible values to make sure the total probability is 1.

Prior probability model



A continuous prior model for π, our prediction ability.

The probability density function $f(\pi)$ answers: What values can π take and which are more plausible than others?

Prior probability model

Continuous probability models

Let π be a continuous random variable with **probability density function** $f(\pi)$. Then $f(\pi)$ has the following properties:

- $f(\pi) \geq 0$
- $\int_{\pi} f(\pi) d\pi = 1$, that is, the area under $f(\pi)$ is 1
- The probability that π falls between two values, a and b, is equal to the area under the curve of the probability density function between those two points. In other words, the area between any two values corresponds to the **probability** of π being in this range.

Prior probability model

Probability Density Function (PDF)

- Applies to **continuous** random variables (e.g., proportion of success, memory, cognitive ability)
- The probability density function integrates to 1
- It is possible that $f(\pi) > 1$, thus a continuous pdf *cannot* be interpreted as a probability. Rather, $f(\pi)$ can be used to *compare* the plausibility of two different values of π : the greater $f(\pi)$, the more plausible the corresponding value of π.
- Probability is defined over intervals: $P(a \leq \pi \leq b) = \int_a^b f(\pi) d\pi$
- The probability at an exact point is zero: $P(\pi = \pi_0) = 0$

Prior probability model

- Next step is to translate the *picture* of our prior into a formal probability model. For that, we need to specify a formula for the pdf $f(\pi)$.
- There are a variety of "named" models to choose from
- A natural choice for probabilities is the Beta probability model
- A Beta random variable is continuous and restricted to live in the interval $[0, 1]$

The Beta model

Let π be a random variable which can take any value between 0 and 1, that is, $\pi \in [0,1]$. Then the variability in π might be well modeled by a Beta model with **shape hyperparameters** $\alpha > 0$ and $\beta > 0$:[1]

$$\pi \sim \text{Beta}(\alpha, \beta).$$

The Beta model is specified by the following probability density function:

$$f(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1} \text{ for } \pi \in [0,1].$$

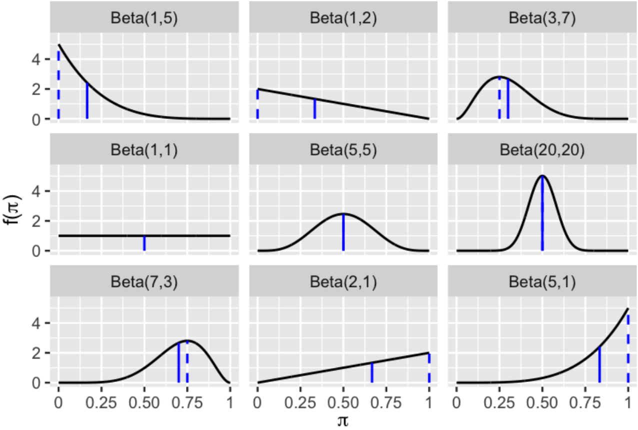
[1] A hyperparameter is a parameter used in a prior model.

The Beta model

We can *tune* the Beta to reflect our prior understanding of π by changing α and β.

- Hyperparameters: $\alpha > 0$ and $\beta > 0$
- When $\alpha, \beta > 1$: in this case α and β can be interpreted as the number of prior successes and failures $+ 1$
- When $\alpha = \beta = 1$: uniform distribution
- When $a < 1$ or $\beta < 1$: U-shape. This models belief in **extreme** values, that is, that the probability is either very close to 0 or very close to 1, but not in between.

The Beta model



The Beta model: R Demo

For a Beta$(5, 4)$ prior on our prediction ability π, we use the `plot_beta()` function to visualize the distribution:

```
library(bayesrules)
plot_beta(5, 4)
# add mean and mode to your plot
plot_beta(alpha = 5, beta = 4,
          mean = TRUE, mode = TRUE)
```

Choosing a suitable prior distribution

Choosing appropriate values for α and β or any other prior hyperparameters is *hard*. When you assign a prior distribution, apply the following principles:

- Principle 1: Priors depend on domain knowledge and theory
  - Theory tells us which values are possible or impossible
  - Theory and past findings can suggest which values are plausible
- Principle 2: You need to be able to justify them to the scientific community
- Principle 3: Priors should be as informative as possible, but no more informative than necessary.

Choosing a suitable prior distribution

When developing a prior, we can use the following strategies:

- Strategy 1: Satisfy constraints
- Strategy 2: Check prior predictions. Simulated data based on the prior should look psychologically plausible
- Strategy 3: When multiple priors are plausible: report sensitivity analyses
- Strategy 4: Use hierarchical extensions
  - For instance, rather than specifying α and β, we estimate them
  - The group-level priors are often easier to justify

Exercise

In pairs:

- Discuss and choose a suitable **prior distribution** for the prediction ability of the class
- Choose values for the Beta distribution parameters that represent your beliefs: α and β
- Plot the resulting Beta distribution using `plot_beta()`
- Compute the mean, mode, and variance of your distribution:
    - Mean: $\mu = \dfrac{\alpha}{\alpha + \beta}$
    - Mode: $\text{Mode}(\pi) = \dfrac{\alpha - 1}{\alpha + \beta - 2}$, when $\alpha, \beta > 1$
    - Variance: $V = \dfrac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

Exercise

- In step 2 of our Bayesian analysis on our collective prediction ability π, we are ready to collect some data.
- We plan to conduct a new poll of $n \approx 50$ participants to collect Y, the number of accurate predictions
- The results depend upon π: the greater our collective prediction accuracy, the greater Y will tend to be.
- We can assume that (1) each student answers *independently* from each other and (2) the probability that any one can successfully predict the replication outcome is π
- To model the dependence of Y on π, $P(Y \mid \pi)$, we use the Binomial function:

$$Y \mid \pi \sim \text{Binomial}(50, \pi)$$

Exercise

Let's put your prediction ability to the test! Read the study description below. Then, write down your prediction (individually!). **Will this study replicate? (Yes/No)** Will it show a **significant effect in the same direction** as the original finding?

Risen and Gilovich (2008): Tempting fate

Do people belief that tempting fate leads to negative consequences? Participants imagined a scenario in which they would come to a lecture in which the professor picks out one student to answer a difficult question in front of the entire class. In one group, participants imagined that they tempted fate by coming to the lecture unprepared. In the other group, participants imagined that they came to the lecture prepared. Afterwards, participants had to estimate how likely it was that they would get chosen. If participants imagined that they tempted fate, they thought it was more likely that they would get chosen by the professor to answer a difficult question in front of the entire class.

The Binomial data model

- To model the dependence of Y on π, $P(Y \mid \pi)$, we use the Binomial distribution:

$$Y \mid \pi \sim \text{Binomial}(50, \pi)$$

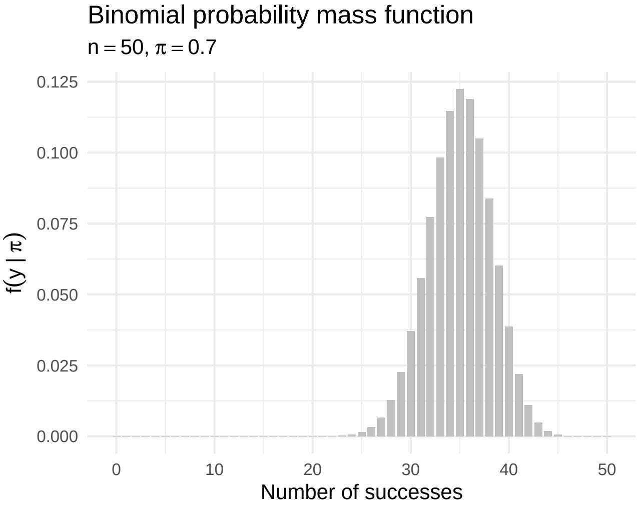
- This means that, for any possible value $y \in \{0, 1, 2, \ldots, 50\}$, the probability of observing exactly y successes is given by the probability mass function (pmf):

$$f(y \mid \pi) = P(Y = y \mid \pi) = \binom{50}{y} \pi^y (1 - \pi)^{(50-y)}$$

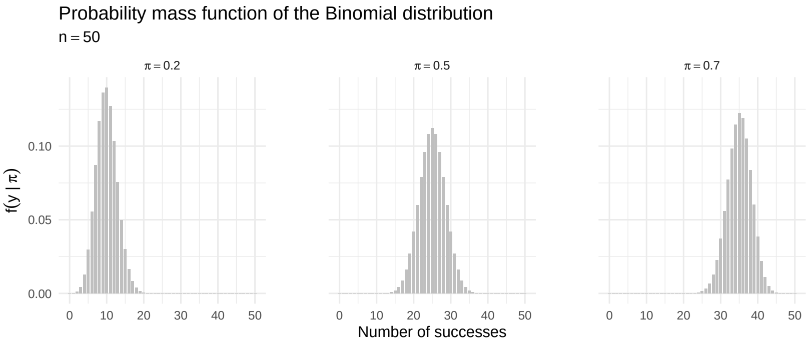
The Binomial data model

$f(y \mid \pi)$ provides answers to a hypothetical question: *if* our prediction accuracy is some given value of π, then how many of the students $Y = y$ might we expect to predict the replication outcome correctly?

The Binomial data model



The Binomial data model



The Binomial$(50, \pi)$ probability mass function pmf, $f(y \mid \pi)$ is plotted for values of $\pi = 0.2$, $\pi = 0.5$, and $\pi = 0.7$.

The Binomial data model

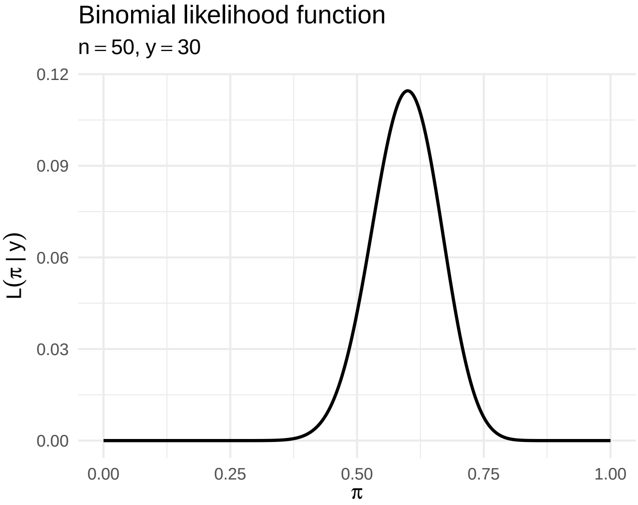
Probability Mass Function (PMF)

- Applies to **discrete** random variables (e.g., number of heads in 10 coin tosses, 10 successful predictions)
- Gives the probability of exact outcomes: $P(Y = y)$
- The pmf sums to 1

The likelihood function

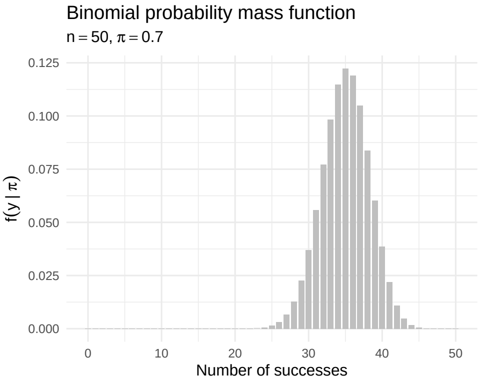
- Let's assume, we would observe that $Y = 30$ of the $n = 50$ students predicted the replication outcome correctly
- Now we can represent the likelihood of the observed polling data, $Y = 30$, at each potential level prediction ability π in $\{0.1, 0.2, \cdots, 0.9\}$.
- We receive a continuous likelihood function $L(\pi \mid y = 30)$ defined for any π between 0 and 1

The likelihood function



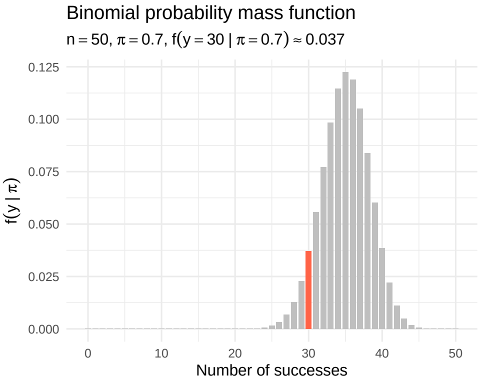
Conditional Probability versus Likelihood

The Binomial data model $P(Y \mid \pi)$ or $f(y \mid \pi)$: "What is the probability of seeing this data, assuming a certain parameter or hypothesis is true?"



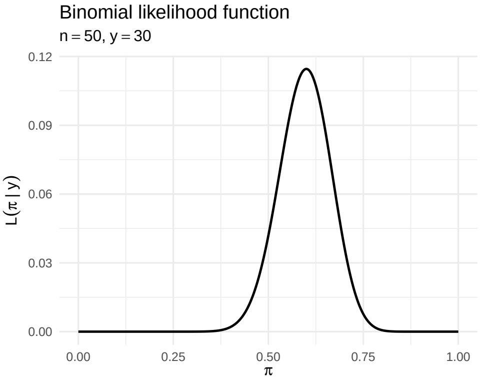
Conditional Probability versus Likelihood

The Binomial data model $P(Y \mid \pi)$ or $f(y \mid \pi)$: "What is the probability of seeing this data, assuming a certain parameter or hypothesis is true?"



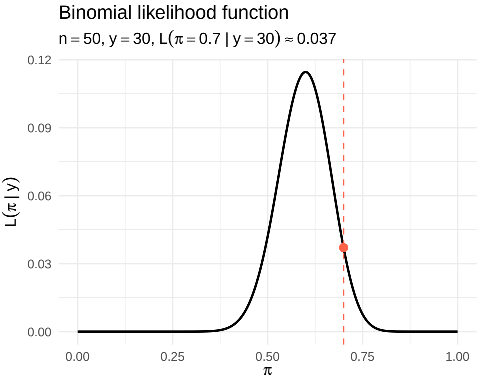
Conditional Probability versus Likelihood

The likelihood function $L(\pi \mid Y)$ is used when the **data is fixed** and you are treating the parameter as variable: "Given this observed data, how likely is each possible value of π?"



Conditional Probability versus Likelihood

The likelihood function $L(\pi \mid Y)$ is used when the **data is fixed** and you are treating the parameter as variable: "Given this observed data, how likely is each possible value of π?"



The Binomial data model

- The Binomial pmf assuming $\pi = 0.7$ and $n = 50$ observations:

$$f(y \mid \pi = 0.7) = \binom{50}{y} 0.7^y (1 - 0.7)^{50-y} \text{ for } y \in [0, 50]$$

- The Binomial likelihood function of π follows from plugging $y = 30$ into the Binomial pmf and letting π free to vary:

$$L(\pi \mid y = 30) = \binom{50}{30} \pi^{30} (1 - \pi)^{50-30} \text{ for } \pi \in [0, 1]$$

- In both cases we get:

$$f(y = 30 \mid \pi = 0.7) = L(\pi = 0.7 \mid y = 30) = \binom{50}{30} 0.7^{30} (1 - 0.7)^{50-30} \approx 0.037$$

Exercise

Original Finding:

- Participants who imagined tempting fate (coming unprepared) felt **more likely** to be selected by the professor.

Replication Outcome:

The study did replicate. The replication attempt did **find a significant effect** in the same direction.

Exercise

You have prior beliefs. You have data. Now update them!

1. Plot the **prior**, **likelihood** (scaled), and **posterior** distributions using the `plot_beta_binomial()` function
2. Compute the mean, mode, and variance of your posterior distribution using the `summarize_beta_binomial()` function.

Discuss: How did the data shift your beliefs?

The Beta posterior model

Conjugate prior

We say that a prior is **conjugate** to a likelihood if, after combining it with the data, the resulting posterior distribution belongs to the same type or family of distributions as the prior.

The Beta posterior model

- The posterior distribution reflects a compromise between prior information and the data.
- Like the prior, the posterior model of π is continuous and lives on [0, 1].
- Like the prior, the posterior model appears to be a Beta(α, β) model where the shape parameters have been *updated* to combine information from the prior and data.

The Beta posterior model

Thus, for a Beta prior and Binomial data, the posterior is defined as follows:

$$\pi \mid Y = y \sim \text{Beta}(\alpha + y, \beta + n - y)$$

The Beta posterior model

As with our previous Bayesian models, the posterior pdf of π strikes a balance between the prior pdf ($f(\pi)$) and the likelihood function $L(\pi \mid y)$ via Bayes' Rule:

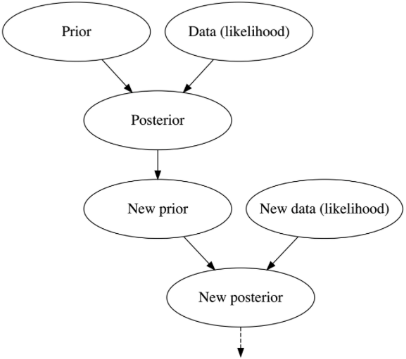
$$f(\pi \mid y) = \frac{f(\pi)L(\pi \mid y)}{f(y)},$$

where $f(y)$ is the marginal probability or *normalizing constant*, that is, a constant which scales the posterior pdf $f(\pi \mid y = 30)$ to integrate to 1.

The Beta posterior model

Just as we continually update our understanding of the world as we live life, our posterior belief continues to evolve as new data come in.
For instance, our understanding of students' prediction ability changes each time we ask another student for their answer.

The Beta posterior model



At each step, the current posterior summarizes everything we have learned so far.

The Beta posterior model

So Bayesian learning is cumulative: each new batch of data updates the current belief.

Suppose we begin with

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

After observing an initial sample with y_1 successes out of n_1 trials, we obtain

$$\pi \mid y_1 \sim \mathrm{Beta}(\alpha + y_1,\ \beta + n_1 - y_1).$$

If we then observe a second sample with y_2 successes out of n_2 trials, we update again:

$$\pi \mid y_1, y_2 \sim \mathrm{Beta}(\alpha + y_1 + y_2,\ \beta + (n_1 - y_1) + (n_2 - y_2)).$$

The Beta posterior model

This Beta-Binomial model has vast applications! It can be used in any situation where the probability of success π is between 0 and 1, with any choice of Beta prior, and any data Y representing the number of successes in n fixed, independent trials. This is a very convenient model because the prior and posterior have the same form.

However, Bayesian models are not always like that.

We will now look at a case where:

- Binomial likelihood
- A non-Beta prior
- The model is non-conjugate

Rouder (2007): Subliminal Priming

Researchers want to know whether a participant can truly detect a briefly presented prime, or whether they are performing at chance. For each participant, we observe:

- A number of correct responses out of a fixed number of trials
- So it is natural to model the data as *number correct out of N trials with success probability p*
- That is just a Binomial likelihood
- However: Rouder makes important theoretical assumption, which cannot be modeled using a Beta prior
- Population is a *mix* of people (1) who always perform at chance level and (2) whose ability to detect a prime is above chance

Rouder (2007): Subliminal Priming

Instead of placing a Beta prior directly on π, Rouder introduces a latent ability x for each participant.
Instead of a Beta prior, a Normal prior is placed on this latent ability x. The probability of a correct response π is then linked to x through a nonlinear transformation. The rule is as follows:

$$\pi_i = \begin{cases} .5, & x_i \leq 0, \\ \Phi(x_i), & x_i > 0, \end{cases}$$

where Φ is the standard normal cumulative distribution function which transforms numbers on the real line to the probability scale.
The model is still Bayesian, but the prior is now placed on x rather than directly on π. This means we no longer have the familiar Beta-Binomial conjugate setup.

Rouder (2007): Subliminal Priming

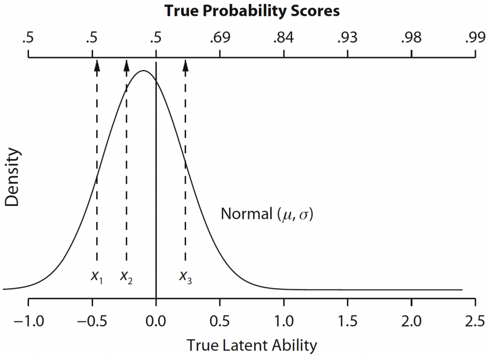


Figure 2. The mass at chance model. The top and bottom axes depict participants' latent ability (x_i) and true probability (p_i), respectively. The curve shows the population distribution of latent abilities. Sixty-two percent of the area under this curve is to the left of zero, indicating that 62% of the population is at chance. The three vertical lines denote latent abilities and corresponding probabilities for three hypothetical participants.

Rouder (2007): Subliminal Priming

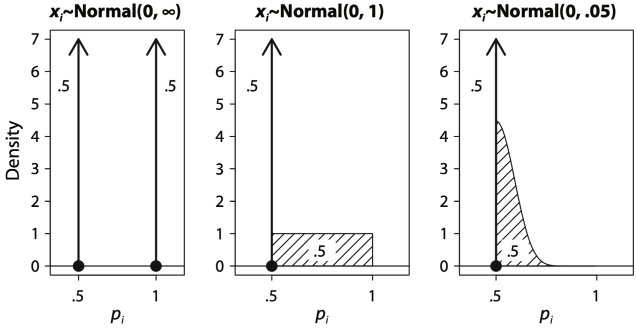


Figure 3. Relationship between priors on x_i and those on p_i. The lines extending upward indicate that half of the mass is concentrated at a point.

Rouder (2007): Subliminal Priming

- **Bayes' rule still works:** Even if the posterior is not a named, closed-form distribution like a Beta, it is still well defined: posterior $\propto$ likelihood x prior
- This is true regardless of the forms of the prior and likelihood
- Bayes' rule gives us the posterior, but it does not promise a neat mathematical expression

Summary: Non-conjugate cases

Bayes' rule tells us how to compute the posterior

$$P(\pi \mid Y) = \frac{P(Y \mid \pi)P(\pi)}{P(Y)}$$

But in many real-world cases:

- The likelihood is complicated or non-linear
- The prior is not conjugate
- The normalizing constant involves a hard integral:

$$P(Y) = \int P(Y \mid \pi)P(\pi)d\pi$$

This means we cannot write down the posterior in a nice closed form.

Summary: Non-conjugate cases

Closed-form posteriors are nice, not necessary.

Challenge	**Bayesian Response**
No closed form	Sampling-based methods
Harder to interpret	Posterior summaries
No conjugacy	Flexibility to formalize beliefs & theory

Summary

We build the Beta-Binomial model for π, an unknown **proportion** that can take any value between 0 and 1:

$$\pi \sim \text{Beta}(\alpha, \beta)$$
$$Y \mid \pi \sim \text{Binomial}(n, \pi)$$
$$\pi|(Y = y) \sim \text{Beta}(\alpha + y, \beta + n - y)$$

Summary

Prior model:

The Beta **prior model** for π can be tuned to reflect the relative prior plausibility of each $\pi \in [0, 1]$:

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

Summary

Likelihood function:

Upon observing data $Y = y$, where $y \in \{0, 1, \cdots, n\}$, the **likelihood function** of π, compares the compatibility of the data with different π:

$$L(\pi \mid y) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \text{ for } \pi \in [0, 1]$$

Summary

Posterior model:

Via Bayes' Rule, the **conjugate** Beta prior combined with the Binomial data model produce a Beta posterior model for π. The updated Beta posterior parameters $(\alpha + y, \beta + n - y)$ reflect the influence of the prior (via α and β) and the observed data (via y and n).

Ungraded Homework Assignment (Exam Prep)

In the book:

- Read Chapter 3.5 “Simulating the Beta-Binomial” and work through the example until you can solve it on your own.
- Exercise 3.3, 3.4, 3.6, 3.12 (solve a. using a simulation), 3.15
- Exercise 4.2, 4.5, 4.6, 4.19