Example: Binomial Model + Beta Posterior

Bots on a Social Media Platform (Bayesian inference)

(Lecture slides)

Story / context (why this model appears everywhere)

Problem

A social media company wants to estimate:

$$\theta = \Pr(\text{an account is a bot}).$$

- There are many users (e.g. millions), but checking all accounts is impossible.
- A software engineer takes a **random sample of** $n = 200$ **accounts**.
- Among them, $y = 3$ **are identified as bots**.

Main goal

Use the data $(y = 3, n = 200)$ to infer what values of θ are plausible, and quantify uncertainty.

Modelling assumptions (Binomial setting)

We represent each sampled account by a binary random variable:

$$X_i = \begin{cases} 1, & \text{if account } i \text{ is a bot,} \\ 0, & \text{if account } i \text{ is human.} \end{cases}$$

Assumptions

- **Same bot probability:** $\Pr(X_i = 1) = \theta$ for all i.
- **Independence:** $X_1, \ldots, X_n$ are independent (approximately reasonable if population is large).

Consequence

$$X_i \mid \theta \sim \text{Bernoulli}(\theta), \qquad Y = \sum_{i=1}^{n} X_i \mid \theta \sim \text{Binomial}(n, \theta).$$

Here Y is the number of bots found in n sampled accounts.

Observation in this example

$$n = 200, \qquad y = 3.$$

- We observed only **one number** y (the count of bots).
- But it summarises **200 independent Bernoulli trials**.

Interpretation

This is a “success/failure repeated trials” scenario:

- **Success** = “account is a bot”
- **Number of successes** = y
- **Number of trials** = n

Likelihood: what is $\pi(y \mid \theta)$?

Given θ, the distribution of Y is binomial:

$$Y \mid \theta \sim \text{Binomial}(n, \theta).$$

Binomial likelihood

$$\pi(y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \qquad y = 0, 1, \ldots, n.$$

In our data ($n = 200, y = 3$)

$$\pi(y = 3 \mid \theta) = \binom{200}{3} \theta^3 (1-\theta)^{197}.$$

Bayes' rule: posterior is proportional to prior $\times$ likelihood

Bayes' theorem (density form):

$$\pi(\theta \mid y) = \frac{\pi(\theta)\,\pi(y \mid \theta)}{\int_0^1 \pi(\theta)\,\pi(y \mid \theta)\,d\theta}.$$

Key point (what we often use in algebra)

$$\pi(\theta \mid y) \propto \pi(\theta)\,\pi(y \mid \theta).$$

- The symbol "$\propto$" means: **equal up to a normalising constant**.
- That normalising constant does **not depend on** θ.
- So when identifying the posterior family, we focus on the parts that **depend on** θ.

A simple prior: $\text{Uniform}(0, 1)$

Suppose we have no strong reason to prefer any particular value of θ a priori.

Uniform prior

$$\theta \sim \text{Uniform}(0, 1), \qquad \pi(\theta) = 1, \quad 0 < \theta < 1.$$

- This says: before seeing data, all $\theta \in (0, 1)$ are equally plausible.
- It is a very common "default" prior for a probability parameter.

Posterior with Uniform prior: simplify using proportionality

Start from

$$\pi(\theta \mid y) \propto \pi(\theta)\,\pi(y \mid \theta).$$

Plug in the uniform prior $\pi(\theta) = 1$

$$\pi(\theta \mid y) \propto 1 \cdot \binom{n}{y} \theta^{y} (1-\theta)^{n-y}.$$

Drop constants that do not depend on θ

$$\pi(\theta \mid y) \propto \theta^{y} (1-\theta)^{n-y}.$$

In our data $(n = 200, y = 3)$

$$\pi(\theta \mid y) \propto \theta^{3} (1-\theta)^{197}.$$

Recognising the Beta distribution

A Beta distribution is defined by:

$$\theta \sim \mathrm{Beta}(\alpha, \beta), \qquad \pi(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0 < \theta < 1.$$

- $B(\alpha, \beta)$ is a normalising constant (so density integrates to 1).
- The **shape** comes from:

$$\theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

Matching exponents

If the posterior looks like

$$\theta^{y}(1-\theta)^{n-y},$$

then we match it with Beta form:

$$\alpha - 1 = y, \qquad \beta - 1 = n - y.$$

Posterior result for this dataset

We had:

$$\pi(\theta \mid y) \propto \theta^3 (1-\theta)^{197}.$$

Match to Beta:

$$\theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

Solve for (α, β)

$$\alpha - 1 = 3 \;\Rightarrow\; \alpha = 4, \qquad \beta - 1 = 197 \;\Rightarrow\; \beta = 198.$$

Posterior distribution

$$\theta \mid y = 3 \sim \mathrm{Beta}(4, 198).$$

Uniform$(0, 1)$ is a special case of Beta

Recall:

$$\theta \sim \mathrm{Beta}(\alpha, \beta) \quad \Longrightarrow \quad \pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

Set $\alpha = \beta = 1$

$$\pi(\theta) \propto \theta^0(1-\theta)^0 = 1, \qquad 0 < \theta < 1.$$

Conclusion

$$\mathrm{Uniform}(0, 1) \equiv \mathrm{Beta}(1, 1).$$

- This is useful because it puts our "uninformative" prior inside the Beta family.
- Once we use Beta priors, the algebra becomes very clean.

Conjugacy: Beta prior + Binomial likelihood

Now suppose a more general prior:

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

Prior density (up to constants)

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}.$$

Likelihood:

$$\pi(y \mid \theta) \propto \theta^{y}(1-\theta)^{n-y}.$$

Posterior

$$\pi(\theta \mid y) \propto \theta^{\alpha-1+y}(1-\theta)^{\beta-1+n-y}.$$

Therefore

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

Why we call this a "conjugate prior"

Definition idea (informal)

A prior is **conjugate** to a likelihood if the posterior stays in the same distribution family.

- Here:

$$\text{Prior: } \mathrm{Beta}(\alpha, \beta) \quad \Rightarrow \quad \text{Posterior: } \mathrm{Beta}(\alpha + y, \beta + n - y).$$

- So Beta is conjugate to Binomial.

What changes after seeing data?

Only the **parameters** update:

$$(\alpha, \beta) \quad \longrightarrow \quad (\alpha + y,\ \beta + n - y).$$

The **family does not change**.

Posterior interpretation: “pseudo-counts” intuition

With $\theta \sim \mathrm{Beta}(\alpha, \beta)$:

Interpretation

$$\alpha - 1 \approx \text{prior bot-count}, \qquad \beta - 1 \approx \text{prior human-count}.$$

After observing y bots and $n - y$ humans:

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

Meaning

Data simply **adds counts**:

$$(\text{bot count}) + y, \qquad (\text{human count}) + (n - y).$$

- Uniform prior $\mathrm{Beta}(1, 1)$ has “zero” pseudo-counts.
- So with $y = 3$, $n = 200$:

$$\mathrm{Beta}(1, 1) \to \mathrm{Beta}(4, 198).$$

Point estimates from the posterior (optional but useful)

For $\theta \sim \mathrm{Beta}(\alpha, \beta)$:

Posterior mean

$$\mathbb{E}[\theta \mid y] = \frac{\alpha}{\alpha + \beta}.$$

Posterior mode (MAP), if $\alpha, \beta > 1$

$$\theta_{\mathrm{MAP}} = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

For our posterior $\mathrm{Beta}(4, 198)$

$$\mathbb{E}[\theta \mid y] = \frac{4}{4 + 198} = \frac{4}{202} \approx 0.0198, \qquad \theta_{\mathrm{MAP}} = \frac{3}{200} = 0.015.$$

- Posterior mean is slightly larger than 3/200 due to the prior adding 1 to both counts.

Summary: what did we learn from this example?

The modelling pipeline

1. Identify repeated success/failure trials $\Rightarrow$ Binomial model.
2. Write likelihood:
$$\pi(y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$
3. Choose prior for θ (Uniform or Beta).
4. Apply Bayes:
$$\pi(\theta \mid y) \propto \pi(\theta)\pi(y \mid \theta).$$
5. Recognise posterior as Beta:
$$\theta \mid y \sim \mathrm{Beta}(\alpha + y, \beta + n - y).$$

For $n = 200$, $y = 3$ with uniform prior

$$\theta \mid y \sim \mathrm{Beta}(4, 198).$$

What this sets up for later lectures

- **Conjugacy** gives closed-form posteriors (fast inference).
- Beta–Binomial is a prototype for many Bayesian models:
  - Dirichlet–Multinomial (categorical outcomes)
  - Normal–Normal (Gaussian mean with known variance)
  - Gamma–Poisson (count data)
- We will later discuss:
  - prior choice (informative vs weakly informative),
  - uncertainty summaries (credible intervals),
  - prediction for new data.

Big message

Choosing a good prior family can make Bayesian updating **algebraic, interpretable, and scalable**.