# Bayesian Coin Toss Inference

**Advanced Statistical Inference**

Simone Rossi

## Coin toss: A class exercise

**Setup**: I toss the coin $n$ times and observe $y$ heads

**Question**: Is the coin fair? What is the probability of heads?

Steps:

1. Define a model for the number of heads
2. Choose a prior distribution
3. Compute the posterior distribution
4. Make predictions

## Model

**Assumptions**:

1. The probability of heads is the same for all tosses.
2. Each coin toss is independent of the others.

We model the number of heads $y$ with a **binomial distribution** and probability $\theta$:

$$p(y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

where $\theta$ is the probability of heads, $n$ is the number of tosses, $y$ is the number of heads, and

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

is the binomial coefficient.

## Prior

We need a prior distribution for $\theta$.

How to choose it? Recall what $\theta$ represents: the probability of heads.

1. It must be between 0 and 1.
2. It must be continuous.
3. We want a posterior we can compute analytically (conjugacy with binomial).

---

**Beta distribution**:

$$p(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

where

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- Parameters: $\alpha$ and $\beta$
- $\alpha = \beta = 1$: uniform prior (no preference)
- $\alpha = \beta > 1$: preference for fairness, with increasing certainty as values grow
- $\alpha > \beta$: prior preference for heads
- $\alpha < \beta$: prior preference for tails

## Prior intuition: pseudo-counts

The beta prior can be interpreted as prior observations:

- $\alpha - 1$: prior pseudo-heads
- $\beta - 1$: prior pseudo-tails
- $\alpha + \beta$: prior strength

Interpretation:

- Small $\alpha + \beta$: weak prior influence
- Large $\alpha + \beta$: strong prior influence
- Posterior update behaves like adding data to pseudo-counts

### Posterior

By Bayes' rule:

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

Because beta is conjugate to binomial, the posterior is also beta:

$$p(\theta \mid y) = \frac{1}{B(\alpha', \beta')}\theta^{\alpha'-1}(1-\theta)^{\beta'-1}$$

So we only need $\alpha'$ and $\beta'$.

### Posterior: parameter update

From conjugacy:

$$p(\theta \mid y) = \frac{1}{B(\alpha', \beta')}\theta^{\alpha'-1}(1-\theta)^{\beta'-1}$$

From Bayes' rule:

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

Matching powers gives:

$$\alpha' = \alpha + y, \qquad \beta' = \beta + n - y$$

## Posterior summaries

If $\theta \mid y \sim \text{Beta}(\alpha', \beta')$, then

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

$$\text{Var}(\theta \mid y) = \frac{\alpha'\beta'}{(\alpha' + \beta')^2(\alpha' + \beta' + 1)}$$

If $\alpha', \beta' > 1$, the MAP is

$$hetasc_{\text{MAP}} = \frac{\alpha' - 1}{\alpha' + \beta' - 2}$$

As $n$ grows, posterior mean and MAP become close.

## Concentrating the posterior

Assume the coin is fair, $\widehat{\theta} = 0.5$.

1. As we toss more times, the posterior concentrates around 0.5, regardless of prior.
2. The prior still controls **how fast** concentration happens.
3. The posterior becomes increasingly close to a normal distribution.

## Prediction for the next toss

Use the same likelihood as before, but for one trial ($n = 1$):

$$p(y_{\text{new}} \mid \theta, n = 1) = \binom{1}{y_{\text{new}}} \theta^{y_{\text{new}}}(1 - \theta)^{1 - y_{\text{new}}}$$

So for a head ($y_{\text{new}} = 1$):

$$p(y_{\text{new}} = 1 \mid \theta, n = 1) = \binom{1}{1} \theta^1 (1 - \theta)^0 = \theta$$

Since $\theta$ is unknown, we integrate over its posterior:

$$p(y_{\text{new}} = 1 \mid y) = \int p(y_{\text{new}} = 1 \mid \theta, n = 1) p(\theta \mid y) \, \mathrm{d}\theta = \int \theta \, p(\theta \mid y) \, \mathrm{d}\theta = \mathbb{E}[\theta \mid y]$$

With beta-binomial conjugacy:

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

## Is the coin fair? Bayesian decision view

Useful diagnostics:

- 95% credible interval for $\theta$
- Posterior probability $p(\theta > 0.5 \mid y)$
- Sensitivity to prior choices $(\alpha, \beta)$

If credible interval is tight around 0.5, data supports fairness.