Bayesian Inference 101

Arnav Sheth
iCapital Network

November 6, 2023

Summary

- Let's start with a coin-toss; probability of heads assumed to be 50-50:
  - $\mathbb{P}(H) = 0.50$
- But is it *really* 50-50? What if we start with heads up?
- Assume physical symmetry, i.e., we have a fair coin.
  - Let's run an experiment.
  - Flip a coin $n = 10$ times with heads up each time.
- Diaconis, et al (2007) and Bartoš, et al (2023) show that $\mathbb{P}(H) = 0.5\mathbf{1}$!

Motivation

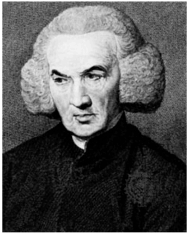
Notation

Probability and Statistics are two sides of the same coin

- Let $X_{1:n} = (X_1, \ldots, X_n)$ be the outcomes of n i.i.d. coin flips.
- Let $\mathbb{P}(H) = \theta$, since we don't know what it is.
- Probability: $\theta \longrightarrow X_{1:n}$.
- Statistics: $X_{1:n} \longrightarrow \theta$.
    - Maybe toss the coin a lot and look at the proportion of heads?

Some History

$\mathbb{P}(\text{this} = \text{Bayes} \mid \text{data}) < 1$



Richard Price

Pierre-Simon Laplace

- Thomas Bayes (1701? - 1761) was a Presbyterian minister.
- Work discovered by Richard Price, published in 1764.
- Laplace developed it further; published in 1774.

Frequentist Statistics

- The basic idea of frequentist statistics is that the world is described by parameters that are:
    - fixed; and
    - unknown.
- That $\exists$ some true parameter value; we are estimating it using the sample we have.
- This true, fixed, unknown parameter describes something about the entire population.

Bayesian Statistics

- Bayesian statistics applies probabilistic methods and reasoning directly to the *parameters*.
- Use the algebra of probabilities to quantify and describe how much we believe various propositions.
- Bayesians apply probabilities directly to their knowledge of the world.
    - Frequentists can only apply probabilities to the act of repeating an experiment.

The Math

- Basic idea: assume a **prior** probability distribution for θ.
  - This represents the *plausibility* of each possible value of θ, BEFORE the data is observed.
- The **posterior** distribution is used to make inferences about θ.
  - This represents the *plausibility* of each possible value of θ, AFTER the data is observed.
- Mathematically, this is expressed:

$$\begin{array}{llll} p(\theta \mid x) = \dfrac{p(x \mid \theta) \times p(\theta)}{p(x)} & \propto & p(x \mid \theta) \times p(\theta) \\ \text{posterior} & \propto & \text{likelihood} \times \text{prior} \end{array}$$

where:

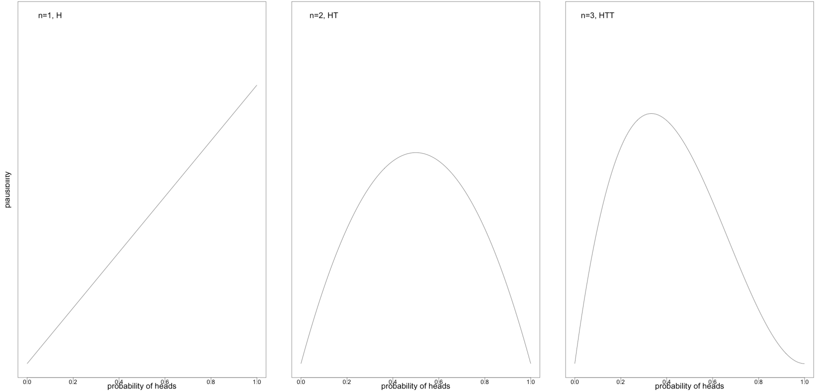
- $x = x_{1:n}$ (for example, the coin flips from earlier);
- $p(\theta \mid x)$ is the posterior;
- $p(x \mid \theta)$ is the likelihood; and
- $p(\theta)$ is the prior.

Advantage: Bayesians?

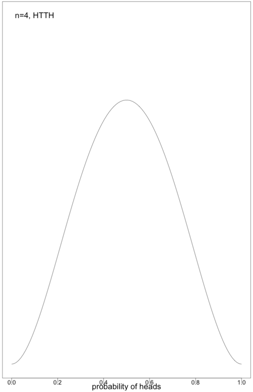
- The results are understandable.
- All parts of the model, including priors and other assumptions, are explicit and open to criticism.
- The model is scrutable.
- Admissibility. (Thanks, Clinton)

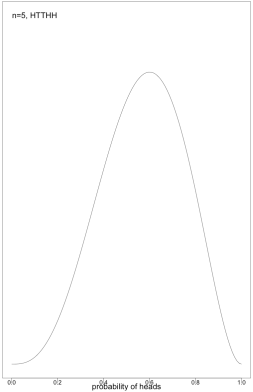
  See, for example: Robert, Christian P. (1994). *The Bayesian Choice*. Springer-Verlag. Also, Wald (1939).

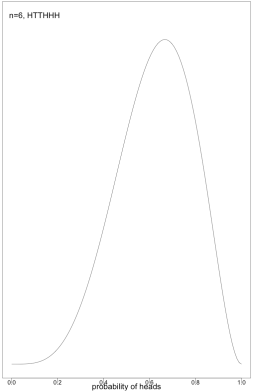
Ten Coin Tosses



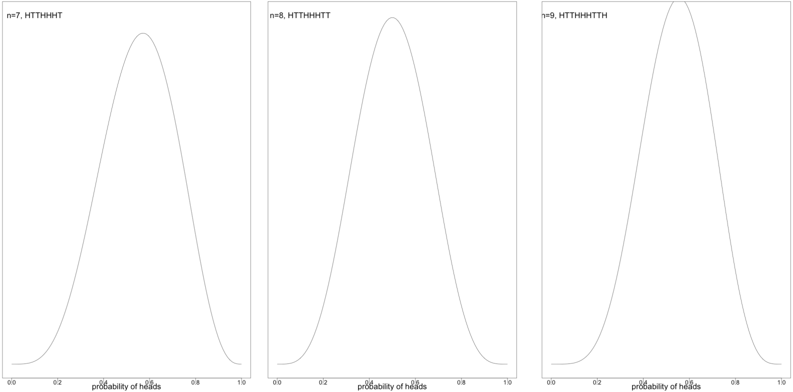
Ten Coin Tosses



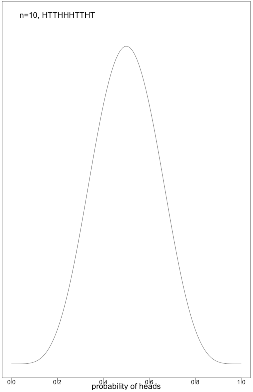




Ten Coin Tosses

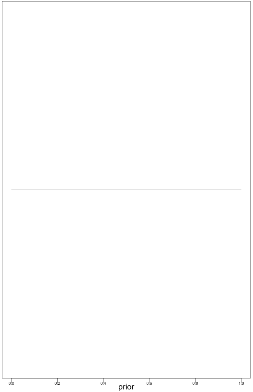


Ten Coin Tosses

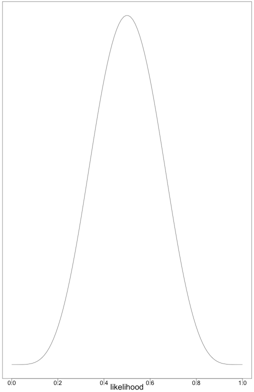


Ten Coin Tosses

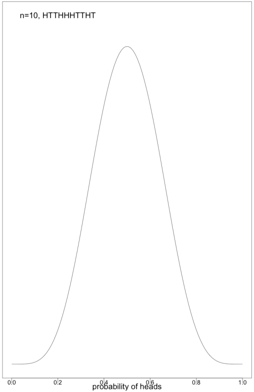
Some Other Priors



$\times$

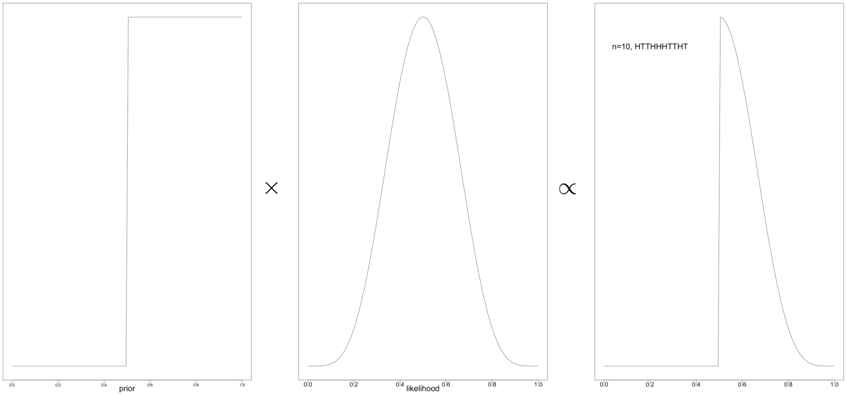


$\propto$



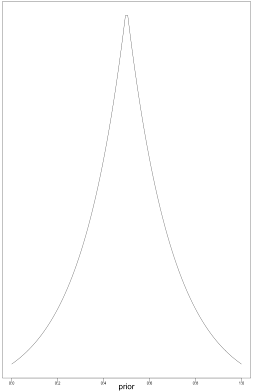
Ten Coin Tosses

Some Other Priors

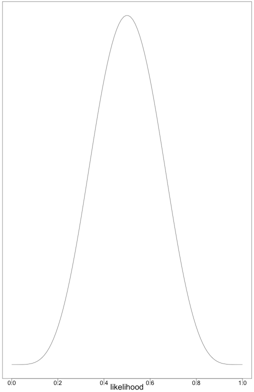


Ten Coin Tosses

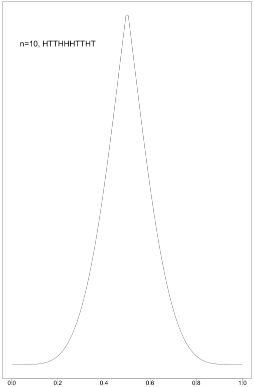
Some Other Priors



$\times$

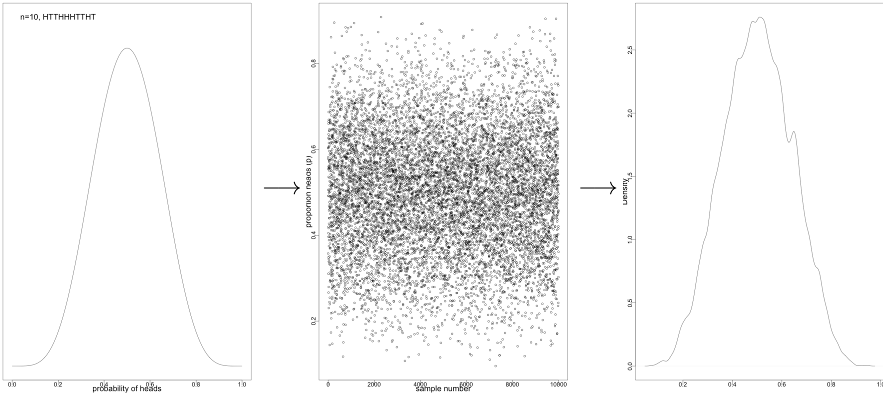


$\propto$



Ten Coin Tosses

Sampling From Posterior



- There is a blood test that correctly detects vampirism 95% of the time.
    - A “95% success rate”
- One percent of the time it incorrectly diagnoses normal people.
- Vampires are only 0.1% of the population.
- Say you test positive for vampirism:

What’s the probability that you **really** are a vampire?

We are told the following:

- $\mathbb{P}$ (positive test result | vampire) = 0.95
- $\mathbb{P}$ (positive test result | mortal) = 0.01
- $\mathbb{P}$ (vampire) = 0.001

We can now use Bayes' theorem to invert the probability to get $\mathbb{P}$(vampire | positive):

$$\mathbb{P}(\text{vampire} \mid \text{positive}) = \frac{\mathbb{P}(\text{positive} \mid \text{vampire})\mathbb{P}(\text{vampire})}{\mathbb{P}(\text{positive})}$$

$$\begin{aligned}\mathbb{P}(\text{positive}) &= \mathbb{P}(\text{positive} \mid \text{vampire})\mathbb{P}(\text{vampire}) + \\ &\quad \mathbb{P}(\text{positive} \mid \text{mortal})\,(1 - \mathbb{P}(\text{vampire})) \\ &= (0.95)(0.001) + (0.01)(0.999) = 0.0868\end{aligned}$$

Or an 8.7% chance that the suspect is actually a vampire.

Not convinced? Ok let's try it a different way. Here is what we are given:

1. In a population of 100,000 people, 100 of them are vampires.
2. Of the 100 who are vampires, 95 of them will test positive for vampirism.
3. Of the 99,900 mortals, 999 will test positive for vampirism.

So if we test all 100,000 people, what proportion of those who test positive for vampirism are actually vampires?

- There are $95 + 999 = 1094$ people who test positive.
- Of the 1094, 95 are really vampires.
- So: $\mathbb{P}(\text{vampire} \mid \text{positive}) = \frac{95}{1094} \approx 0.087!$

Implication

It's important to see here that a test that a "95% success rate" ends up being right only $\approx$ 9% of the time!

Ten Coin Tosses

Technical Details

- The uninformed prior for the coin tosses are distributed uniformly between 0 and 1:

$$p(H) \sim U(0,1)$$

- The binomial likelihood is given by:

$$\mathcal{L}\left(H \mid N, p\right) = \frac{N!}{H!(N-H)!} p^H (1-p)^{N-H}$$

- where:
  - N is the number of tosses, which in our final case was 10;
  - H is the observed count of heads, which in our final case was 5;
  - p is the probability of heads as defined above.
    - The likelihood generates the data for all values of $p \in 0:1$.