Introduction to Bayesian Statistics

PSCI 1801 · Class Handout

Dr. Marc Trussler

This handout is designed as an in-class supplement to the textbook. The material contained on this handout does not represent the full set of information you need to know for the exams. Exclusively studying off of this handout will not be sufficient for the exams.

The Fundamental Difference

Everything we've done all semester has been frequentist *statistics. There's another way to think about inference:* Bayesian *statistics.*

- **Frequentist question**: If H_0 is true, how likely are these data?
- **Bayesian question**: Given these data, how likely is the hypothesis?

These are not *the same question. A p-value of 0.03 does* not *mean "there's a 3% chance the null is true" — it means "there's a 3% chance of data this extreme if the null were true."*

A Bayesian can say "there's an 85% probability the true effect is positive." A frequentist cannot — even though people say things like this all the time and get it wrong.

You Already Know the Math

Bayes' theorem comes straight from conditional probability we did in the probability chapter:

$$P(A \mid B) = \frac{P(A\&B)}{P(B)}, \qquad P(A\&B) = P(B \mid A) \cdot P(A)$$

Combine them:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Applied to statistical inference (with A = hypothesis, B = data):

$$P(\text{Hyp} \mid \text{Data}) = \frac{P(\text{Data} \mid \text{Hyp}) \cdot P(\text{Hyp})}{P(\text{Data})}$$

The four pieces

- **Posterior** $P(\text{Hyp} \mid \text{Data})$: updated belief after seeing data. This is what we want.
- **Likelihood** $P(\text{Data} \mid \text{Hyp})$: how probable the data are under a hypothesis. Frequentism lives here.
- **Prior** $P(\text{Hyp})$: what we believed *before* seeing the data.
- **Marginal likelihood** $P(\text{Data})$: normalizing constant. Usually not directly interpreted.

The verbal recipe:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Worked Example: Is This Coin Fair?

Flip 10 times, observe 7 heads. What can we conclude?

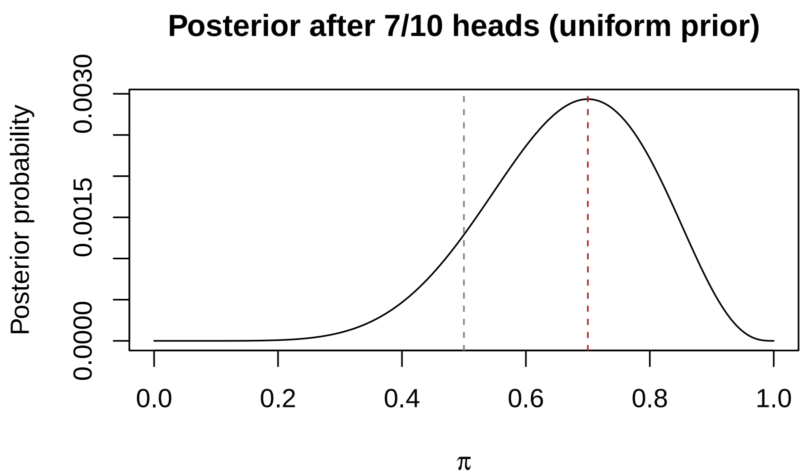
Frequentist answer

```
[1] 0.2265625
```

Fail to reject $H_0 : \pi = 0.5$. Notice what we cannot *say: we cannot say "the coin is probably fair" or attach a probability to fairness.*

Bayesian answer

Consider every *possible value of π from 0 to 1 and compute the posterior.*



Now we can make direct probability statements:

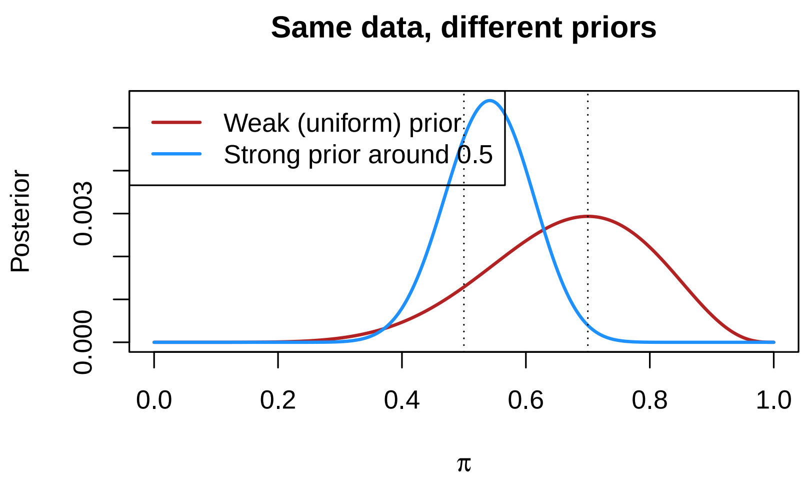
```
[1] 0.8860734
```

"Given our data, there's about an 89% probability the coin is biased towards heads." That's a Bayesian statement.

The Prior Matters — the Controversial Part

Bayesian answers depend on the prior. Two people with different priors see the same data and reach different conclusions.

Example: strong prior belief that the coin is fair (Beta(20, 20)) vs. flat prior:



With more data, the prior matters less — the likelihood dominates. With small n, the prior matters a lot.

When Does the Prior Really Matter?

Two rules of thumb:

1. **Sample size**: with lots of data, priors are washed out. With small n, priors dominate.
2. **Prior strength**: a "vague" prior (broad, flat) doesn't pull much. A "strong" prior (peaked, informative) can override moderate amounts of data.

Frequentist vs Bayesian Correspondences

Frequentist	Bayesian
Parameters are fixed unknowns	Parameters are random variables with distributions
Probability = long-run frequency	Probability = degree of belief
Uses data only	Uses data + prior beliefs
P-values, confidence intervals	Posterior distributions, credible intervals
"95% CI" = 95% of such intervals contain the truth	"95% credible interval" = 95% probability truth is inside

Frequentist	Bayesian
Prior info: not formally incorporated	Prior info: explicitly modeled
Large samples: SEs shrink	Large samples: posterior concentrates (prior irrelevant)
Small samples: wide CIs, low power	Small samples: prior helps (or hurts)

Why Frequentist First

Two reasons this course is frequentist:

1. **Dominant paradigm**: nearly all social science research uses frequentist methods. You need to be able to read this literature.
2. **Simpler machinery**: the frequentist framework has one set of rules (sampling distributions, standard errors, p-values). Bayesian statistics has more machinery (priors, posteriors, MCMC).

Understanding both makes you a better consumer of statistics — regardless of which you use.