# Bayesian Estimation

Foundations of Data Analysis

February 15, 2023

*All models are wrong, but some are useful.*

— George Box

# Likelihood vs. Bayesian Estimation

**Likelihood:**

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i; \theta)$$

**Bayesian:**

$$p(\theta \mid x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n \mid \theta) p(\theta)}{p(x_1, \ldots, x_n)}$$

# Likelihood vs. Bayesian Estimation

**Likelihood:** $\theta$ is a parameter

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i; \theta)$$

**Bayesian:** $\theta$ is a random variable

$$p(\theta \mid x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n \mid \theta) p(\theta)}{p(x_1, \ldots, x_n)}$$

# Why is Random $\theta$ Important?

- The prior, $p(\theta)$, let us use our **beliefs, previous experience, or desires** in the model.

# Why is Random $\theta$ Important?

- The prior, $p(\theta)$, let us use our **beliefs, previous experience, or desires** in the model.
- We can make **probabilistic statements** about $\theta$ (e.g., mean, variance, quantiles, etc.).

# Why is Random $\theta$ Important?

- The prior, $p(\theta)$, let us use our **beliefs, previous experience, or desires** in the model.
- We can make **probabilistic statements** about $\theta$ (e.g., mean, variance, quantiles, etc.).
- If $\theta$ is one of several competing **hypotheses** , we can assign it a probability.

# Why is Random $\theta$ Important?

- The prior, $p(\theta)$, let us use our **beliefs, previous experience, or desires** in the model.
- We can make **probabilistic statements** about $\theta$ (e.g., mean, variance, quantiles, etc.).
- If $\theta$ is one of several competing **hypotheses** , we can assign it a probability.
- We can make **probabilistic predictions** of the next data point, $\hat{x}$, using

$$p(\hat{x} \mid x_1, \ldots, x_n) = \int p(\hat{x} \mid \theta) p(\theta \mid x_1, \ldots, x_n) d\theta$$

# But Bayesian Analysis is *Subjective*, Right?

# But Bayesian Analysis is *Subjective*, Right?

- Not necessarily (we'll cover non-informative priors)

# But Bayesian Analysis is *Subjective*, Right?

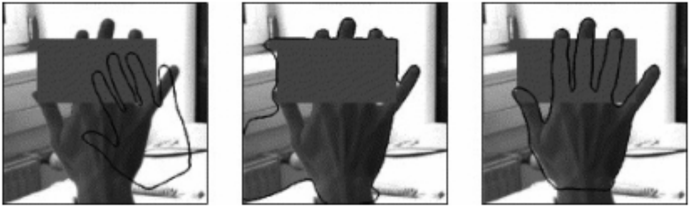
- Not necessarily (we'll cover non-informative priors)
- Frequentist models make assumptions, too!

# But Bayesian Analysis is *Subjective*, Right?

- Not necessarily (we'll cover non-informative priors)
- Frequentist models make assumptions, too!
- Whether using frequentist or Bayesian models, **always c**
  **assumptions you make.**

# But Bayesian Analysis is *Subjective*, Right?

- Not necessarily (we'll cover non-informative priors)
- Frequentist models make assumptions, too!
- Whether using frequentist or Bayesian models, **always c assumptions you make.**
- Sometimes prior knowledge is a good thing.



*Cremers, et al., Pattern Recognition, 2003*

# Bayesian Logic

By Bayesian logic, we can *flip* the implication!

# Bayesian Logic

By Bayesian logic, we can *flip* the implication!

$$P(B \mid A) > P(B) \qquad \text{given}$$

# Bayesian Logic

By Bayesian logic, we can *flip* the implication!

$$P(B|A) > P(B)$$ given

$$\frac{P(A)P(B|A)}{P(B)} > P(A)$$ multiply by $\frac{P(A)}{P(B)}$

# Bayesian Logic

By Bayesian logic, we can *flip* the implication!

$$P(B \mid A) > P(B)$$ given

$$\frac{P(A)P(B \mid A)}{P(B)} > P(A)$$ multiply by $\frac{P(A)}{P(B)}$

$$P(A \mid B) > P(A)$$ Bayes' Rule

# Flipping the implication: $P(B \mid A) > P(B)$

# Flipping the implication: $P(B \mid A) > P(B)$

If $A$ is true, $B$ becomes more likely.

# Flipping the implication: $P(B \mid A) > P(B)$

If $A$ is true, $B$ becomes more likely.

$B$ is true.

# Flipping the implication: $P(B \mid A) > P(B)$

If $A$ is true, $B$ becomes more likely.

$B$ is true.

---

$A$ is more likely.

# Cold Example

| | | $C$ = 0 | $C$ = 1 | |
|---|---|---|---|---|
| $R$ | 0 | 0.50 | 0.05 | 0.55 |
| | 1 | 0.20 | 0.25 | 0.45 |
| | | 0.7 | 0.3 | |

Remember:

$$P(C) = 0.3$$

$$P(C \mid R) = 0.56$$

# Cold Example

| | | C = 0 | C = 1 | |
|---|---|---|---|---|
| R | 0 | 0.50 | 0.05 | 0.55 |
| | 1 | 0.20 | 0.25 | 0.45 |
| | | 0.7 | 0.3 | |

Remember:

$$P(C) = 0.3$$

$$P(C \mid R) = 0.56$$

What if I didn't give you the full table, but just:

$$P(R \mid C) = 0.83 > P(R) = 0.45$$

What can you say about the increase $P(C \mid R) > P(C)$?

# Cold Example

Notice, having a cold *increases* my chance for a runny nose by the factor,

$$\frac{P(R|C)}{P(R)} = \frac{0.83}{0.45} = 1.85$$

# Cold Example

Notice, having a cold *increases* my chance for a runny nose by the factor,

$$\frac{P(R \mid C)}{P(R)} = \frac{0.83}{0.45} = 1.85$$

How does such a ratio increase if I flip the conditional?

$$\frac{P(C \mid R)}{P(C)} = \frac{P(C \cap R)}{P(R)P(C)} = \frac{P(R \mid C)}{P(R)} = 1.85$$

# MLE of Bernoulli Proportion

$$X \sim \mathrm{Ber}(\theta)$$

$$L(\theta \,|\, x_1, \ldots, x_n) = \theta^k (1-\theta)^{n-k}, \text{where } \mathrm{k} = \sum_{\mathrm{i}} \mathrm{x_i}$$

# MLE of Bernoulli Proportion

$$X \sim \mathrm{Ber}(\theta)$$

$$L(\theta \,|\, x_1, \ldots, x_n) = \theta^k (1-\theta)^{n-k}, \text{where } \mathrm{k} = \sum_{\mathrm{i}} \mathrm{x_i}$$

$$\frac{dL}{d\theta} = k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1}$$

$$= (k(1-\theta) - (n-k)\theta)\theta^{k-1}(1-\theta)^{n-k-1}$$

$$= (k - n\theta)\theta^{k-1}(1-\theta)^{n-k-1}$$

$$\frac{dL}{d\theta}(\hat{\theta}) = 0 \;\Rightarrow\; \hat{\theta} = \frac{k}{n}$$

# Bayesian Inference of a Bernoulli Proportion

Let's give $\theta$ a uniform prior: $\theta \sim \mathrm{Unif}(0,1)$

Posterior:

$$p(\theta \,|\, x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n \,|\, \theta) p(\theta)}{p(x_1, \ldots, x_n)}$$

$$= \frac{p(x_1, \ldots, x_n \,|\, \theta)}{p(x_1, \ldots, x_n)}$$

# Bayesian Inference of a Bernoulli Proportion

Just need the denominator (normalizing constant):

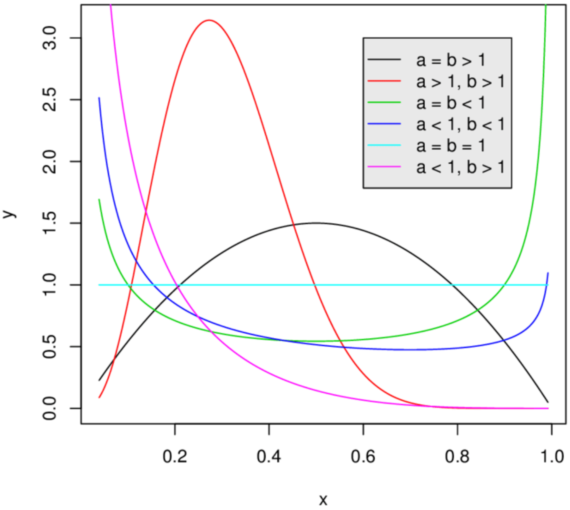
$$\begin{aligned} p(x_1, \ldots, x_n) &= \int_0^1 p(x_1, \ldots, x_n \,|\, \theta) p(\theta) d\theta \\ &= \int_0^1 \theta^k (1 - \theta)^{n-k} d\theta \\ &= \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \end{aligned}$$

Resulting posterior is:

Beta distribution

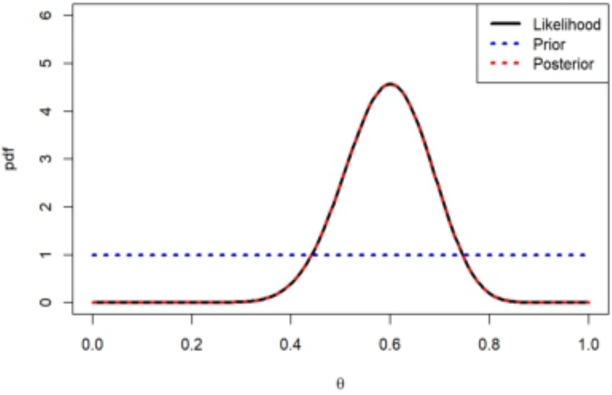
$$p(\theta \,|\, x_1, \ldots, x_n) = \boxed{\frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} \theta^k (1-\theta)^{n-k}}$$

# Shape of Beta Distribution



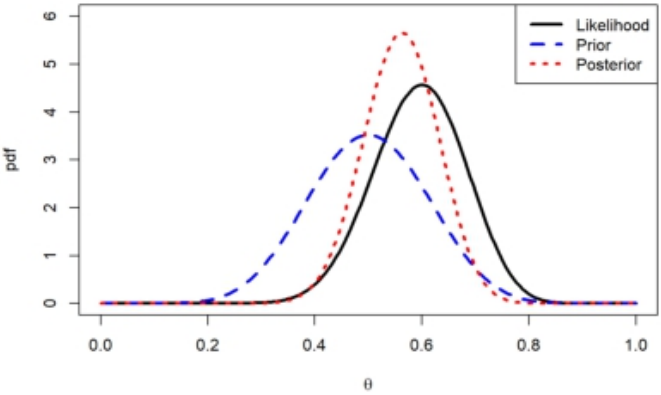
# Example

Bernoulli likelihood with Beta(1,1) prior

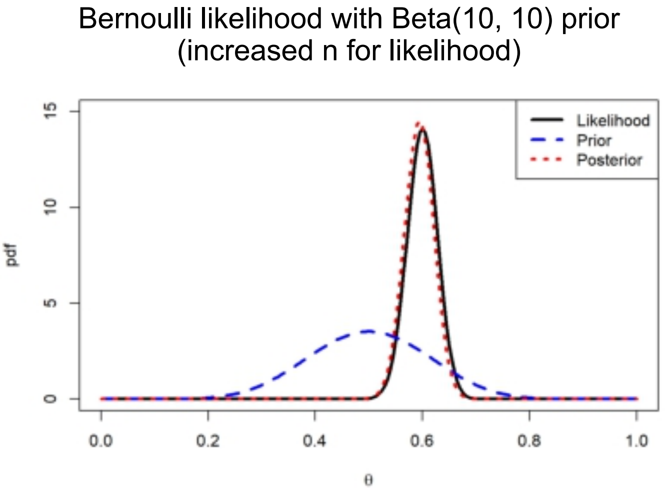


# Example

Bernoulli likelihood with Beta(2, 2) prior



# Example



Bernoulli likelihood with Beta(10, 10) prior
(increased n for likelihood)
Likelihood
Prior
Posterior
pdf
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
θ