

Bayesian Estimation

Foundations of Data Analysis

February 6, 2020

All models are wrong, but some are useful.

— George Box

Frequentist vs. Bayesian Statistics

Frequentist:

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

Bayesian:

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

Frequentist vs. Bayesian Statistics

Frequentist: θ is a parameter

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

Bayesian: θ is a random variable

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

Why is Random θ Important?

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

Why is Random θ Important?

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

Why is Random θ Important?

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

Why is Random θ Important?

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

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

But Bayesian Analysis is *Subjective*, Right?

But Bayesian Analysis is *Subjective*, Right?

- ▶ Not necessarily (we'll cover noninformative priors)

But Bayesian Analysis is *Subjective*, Right?

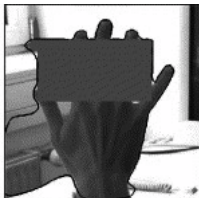
- ▶ Not necessarily (we'll cover noninformative priors)
- ▶ Frequentist models make assumptions, too!

But Bayesian Analysis is *Subjective*, Right?

- ▶ Not necessarily (we'll cover noninformative priors)
- ▶ Frequentist models make assumptions, too!
- ▶ Whether using frequentist or Bayesian models, **always check the assumptions you make.**

But Bayesian Analysis is *Subjective*, Right?

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



MLE of Bernoulli Proportion

$$X_1, \dots, X_n \sim \text{Ber}(\theta)$$

$$L(\theta \mid x_1, \dots, x_n) = \theta^k (1 - \theta)^{n-k}, \quad \text{where } k = \sum_i x_i$$

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

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

Bayesian Inference of a Bernoulli Proportion

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

$$p(\theta) = 1, \quad \text{for } \theta \in [0, 1]$$

Bayesian Inference of a Bernoulli Proportion

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

$$p(\theta) = 1, \quad \text{for } \theta \in [0, 1]$$

Posterior:

$$\begin{aligned} p(\theta \mid x_1, \dots, x_n) &= \frac{p(x_1, \dots, x_n \mid \theta)p(\theta)}{p(x_1, \dots, x_n)} \\ &= \frac{p(x_1, \dots, x_n \mid \theta)}{p(x_1, \dots, x_n)} \end{aligned}$$

Bayesian Inference of a Bernoulli Proportion

Just need the denominator (normalizing constant):

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

Resulting posterior is:

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

Beta Distribution

$\theta \sim \text{Beta}(\alpha, \beta)$ PDF:

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

Beta Distribution

$\theta \sim \text{Beta}(\alpha, \beta)$ PDF:

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

So, posterior of Bernoulli with Uniform prior is

$$\theta \mid x \sim \text{Beta}(k + 1, n - k + 1).$$

Beta Distribution

$\theta \sim \text{Beta}(\alpha, \beta)$ PDF:

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

So, posterior of Bernoulli with Uniform prior is

$$\theta \mid x \sim \text{Beta}(k + 1, n - k + 1).$$

Also notice that $\text{Beta}(1, 1)$ is equivalent to $\text{Unif}(0, 1)$.

Beta Distribution

$\theta \sim \text{Beta}(\alpha, \beta)$ PDF:

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

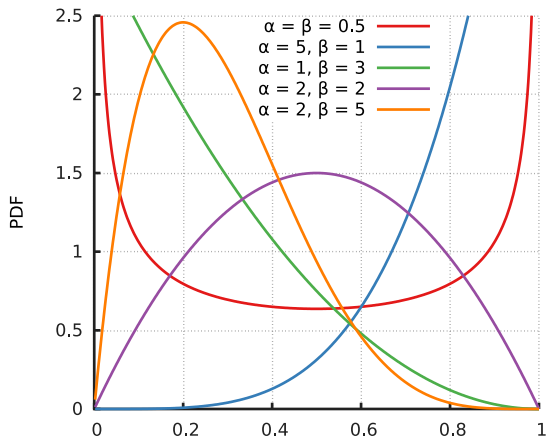
So, posterior of Bernoulli with Uniform prior is

$$\theta \mid x \sim \text{Beta}(k + 1, n - k + 1).$$

Also notice that $\text{Beta}(1, 1)$ is equivalent to $\text{Unif}(0, 1)$.

$$\text{Mode: } \max p(\theta) = \frac{\alpha-1}{\alpha+\beta-2}$$

Beta pdf



See https://en.wikipedia.org/wiki/Beta_distribution

Bernoulli Likelihood with Beta Prior

$$X_1, \dots, X_n \sim \text{Ber}(\theta)$$

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

Bernoulli Likelihood with Beta Prior

$$X_1, \dots, X_n \sim \text{Ber}(\theta)$$

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

Posterior:

$$\begin{aligned} p(\theta \mid x_1, \dots, x_n) &\propto p(x_1, \dots, x_n \mid \theta) p(\theta; \alpha, \beta) \\ &\propto \theta^k (1 - \theta)^{n-k} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &= \theta^{k+\alpha-1} (1 - \theta)^{n-k+\beta-1} \end{aligned}$$

Bernoulli Likelihood with Beta Prior

$$X_1, \dots, X_n \sim \text{Ber}(\theta)$$

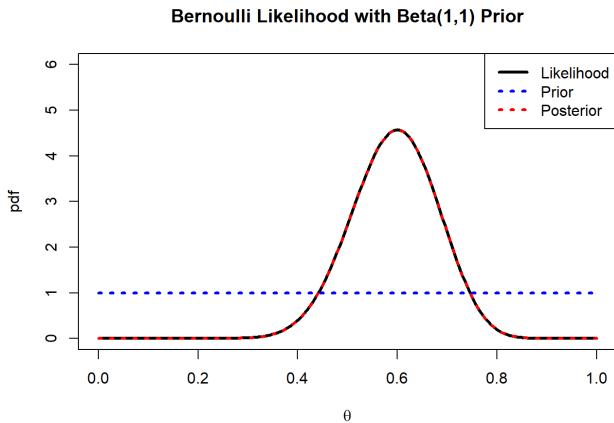
$$\theta \sim \text{Beta}(\alpha, \beta)$$

Posterior:

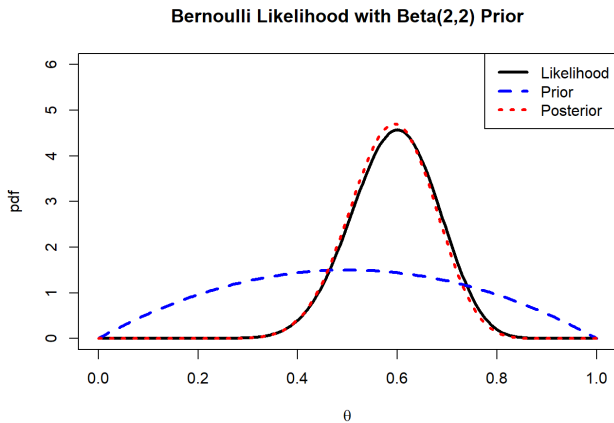
$$\begin{aligned} p(\theta \mid x_1, \dots, x_n) &\propto p(x_1, \dots, x_n \mid \theta) p(\theta; \alpha, \beta) \\ &\propto \theta^k (1 - \theta)^{n-k} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &= \theta^{k+\alpha-1} (1 - \theta)^{n-k+\beta-1} \end{aligned}$$

So, posterior dist. of θ is $\text{Beta}(k + \alpha, n - k + \beta)$.

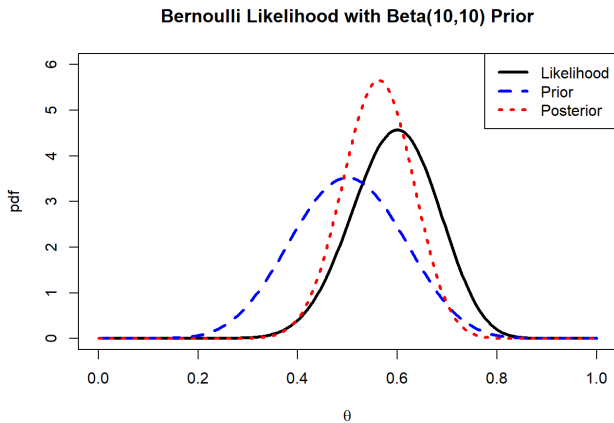
Example



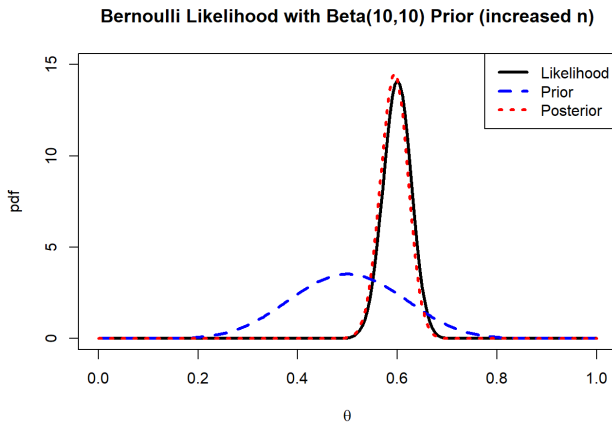
Example



Example



Example



Laplace's Analysis of Birth Rates

Mémoire sur les probabilités (1778)

<http://cerebro.xu.edu/math/Sources/Laplace/>

Problem: Boys were born at a consistently, but only slightly, higher rate than girls in Paris. Was this a real effect or just due to chance?

Laplace's Analysis of Birth Rates

Mémoire sur les probabilités (1778)

<http://cerebro.xu.edu/math/Sources/Laplace/>

Problem: Boys were born at a consistently, but only slightly, higher rate than girls in Paris. Was this a real effect or just due to chance?

Boys: $k = 251527$ # Girls: $n - k = 241945$

Laplace's Analysis of Birth Rates

Mémoire sur les probabilités (1778)

<http://cerebro.xu.edu/math/Sources/Laplace/>

Problem: Boys were born at a consistently, but only slightly, higher rate than girls in Paris. Was this a real effect or just due to chance?

Boys: $k = 251527$ # Girls: $n - k = 241945$

Solution: Model the proportion of boys as the posterior:
 $\theta \mid k \sim \text{Beta}(251528, 241946)$. Then,

$$P(\theta \leq 0.5 \mid k) = F_{\theta \mid k}(0.5) = 1.15 \times 10^{-42}$$