Bayesian Estimation

Foundations of Data Analysis

February 24, 2022

All models are wrong, but some are useful.

— George Box

Frequentist vs. Bayesian Statistics

Frequentist:

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

Bayesian:

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

Frequentist vs. Bayesian Statistics

Frequentist: θ is a parameter

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

Bayesian: θ 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 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

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

But Bayesian Analysis is *Subjective*, Right?

But Bayesian Analysis is *Subjective*, Right?

- Not necessarily (we'll cover noninformative priors)

But Bayesian Analysis is *Subjective*, Right?

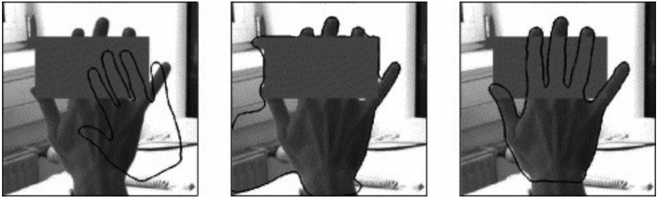
- Not necessarily (we'll cover noninformative priors)
- Frequentist models make assumptions, too!

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.



Cremers, et al., Pattern Recognition, 2003

Deductive Logic

Remember *modus ponens*?

Deductive Logic

Remember *modus ponens*?

$$A \Rightarrow B$$

Deductive Logic

Remember *modus ponens*?

$$A \Rightarrow B$$

A is true

Deductive Logic

Remember *modus ponens*?

$$\frac{\begin{array}{c} A \Rightarrow B \\ A \text{ is true} \end{array}}{B \text{ is true}}$$

Deductive Logic

Remember *modus ponens*?

$$\frac{\begin{array}{c} A \Rightarrow B \\ A \text{ is true} \end{array}}{B \text{ is true}}$$

If it's raining, then the sidewalk is wet.

Deductive Logic

Remember *modus ponens*?

$A \Rightarrow B$	If it's raining, then the sidewalk is wet.
A is true	It's raining.
B is true	

Deductive Logic

Remember *modus ponens*?

$A \Rightarrow B$	If it's raining, then the sidewalk is wet.
A is true	It's raining.
B is true	The sidewalk is wet.

Deductive Logic

How about *modus tollens*?

Deductive Logic

How about *modus tollens*?

$$A \Rightarrow B$$

Deductive Logic

How about *modus tollens*?

$$A \Rightarrow B$$

B is false

Deductive Logic

How about *modus tollens*?

$$A \Rightarrow B$$

$$\frac{B \text{ is false}}{A \text{ is false}}$$

Deductive Logic

How about *modus tollens*?

$$\frac{\begin{array}{c} A \Rightarrow B \\ B \text{ is false} \end{array}}{A \text{ is false}}$$

If it's raining, then the sidewalk is wet.

Deductive Logic

How about *modus tollens*?

$A \Rightarrow B$	If it's raining, then the sidewalk is wet.
B is false	The sidewalk is not wet.
A is false	

Deductive Logic

How about *modus tollens*?

$A \Rightarrow B$	If it's raining, then the sidewalk is wet.
B is false	The sidewalk is not wet.
A is false	It is not raining.

Conditional Probability as Logic

Logic	Probability
A, B are propositions	A, B are events
$A \Rightarrow B$	$P(B \mid A) > P(B)$

Conditional Probability as Logic

Logic	**Probability**
A, B are propositions	A, B are events
$A \Rightarrow B$	$P(B \mid A) > P(B)$

Weak form of *modus ponens*:

Conditional Probability as Logic

Logic	**Probability**
A, B are propositions	A, B are events
$A \Rightarrow B$	$P(B \mid A) > P(B)$

Weak form of *modus ponens*:

If A is true, B becomes more likely.

Conditional Probability as Logic

Logic	Probability
A, B are propositions	A, B are events
$A \Rightarrow B$	$P(B \mid A) > P(B)$

Weak form of *modus ponens*:

If A is true, B becomes more likely.

A is true.

Conditional Probability as Logic

Logic	Probability
A, B are propositions	A, B are events
$A \Rightarrow B$	$P(B \mid A) > P(B)$

Weak form of *modus ponens*:

If A is true, B becomes more likely.

A is true.

B is more likely.

Bayesian Logic

Unlike Boolean logic, we can *flip* the implication!

Bayesian Logic

Unlike Boolean logic, we can *flip* the implication!

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

Bayesian Logic

Unlike Boolean logic, we can *flip* the implication!

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

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

Bayesian Logic

Unlike Boolean logic, we can *flip* the implication!

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

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

$$P(A \mid B) > P(A) \qquad \text{Bayes' Rule}$$

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.

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

$$\frac{\begin{array}{l}\text{If } A \text{ is true, } B \text{ becomes more likely.} \\ B \text{ is true.}\end{array}}{A \text{ is more likely.}}$$

If it's raining, then the sidewalk is more likely to be wet.

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

If A is true, B becomes more likely.

B is true.

A is more likely.

If it's raining, then the sidewalk is more likely to be wet.

The sidewalk is wet.

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

If A is true, B becomes more likely.

B is true.

A is more likely.

If it's raining, then the sidewalk is more likely to be wet.

The sidewalk is wet.

It's more likely to be raining.

Exercise for You

Given that $P(B \mid A) > P(B)$, show that:

1. If $\bar{B}$ happens, A becomes less likely.
   (weak form of *modus tollens*)
2. If $\bar{A}$ happens, B becomes less likely.

Final Bayesian Logic Rules

Given that $P(B \mid A) > P(B)$, analagous to $A \Rightarrow B$, we have four rules:

1. If A, then B is more likely (weak *modus ponens*)
2. If $\bar{B}$, then A is less likely (weak *modus tollens*)
3. If B, then A is more likely (no logical equivalent)
4. If $\bar{A}$, then B is less likely (no logical equivalent)

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:

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

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,

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

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?

$$\frac{P(C \mid R)}{P(C)}$$

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?

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

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?

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

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?

$$\begin{aligned}\frac{P(C \mid R)}{P(C)} &= \frac{P(C \cap R)}{P(R)P(C)} \\ &= \frac{P(R \mid C)}{P(R)} \\ &= 1.85\end{aligned}$$

MLE of Bernoulli Proportion

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

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

MLE of Bernoulli Proportion

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

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

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

MLE of Bernoulli Proportion

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

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

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

MLE of Bernoulli Proportion

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

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

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

MLE of Bernoulli Proportion

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

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

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

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

Bayesian Inference of a Bernoulli Proportion

Let's give θ 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):

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

Resulting posterior is:

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

Beta Distribution

$X \sim \mathrm{Beta}(\alpha, \beta)$ PDF:

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

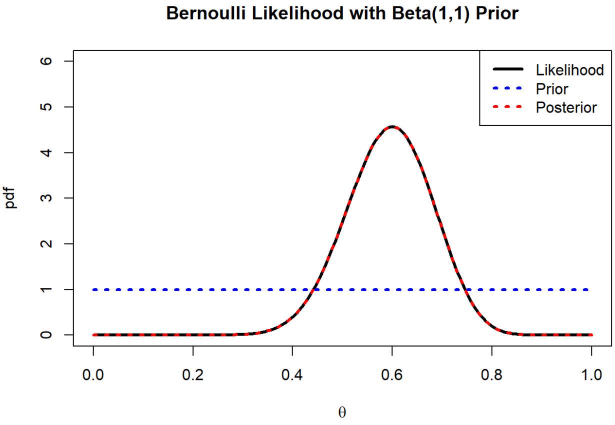
Beta Distribution

$X \sim \mathrm{Beta}(\alpha, \beta)$ PDF:

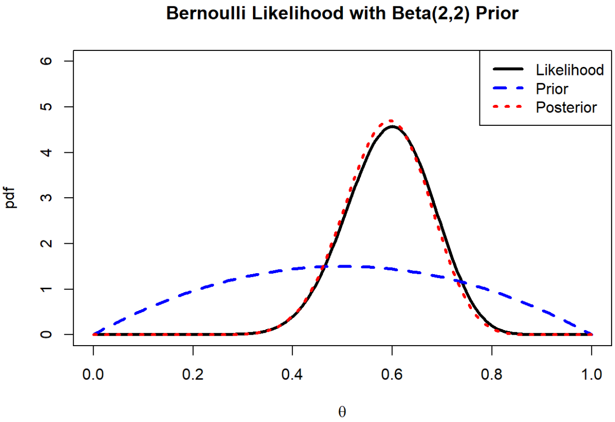
$$p(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1}(1 - x)^{\beta - 1}$$

So, posterior of Bernoulli with Uniform prior is $\theta \sim \mathrm{Beta}(k + 1, n - k + 1)$.

Example

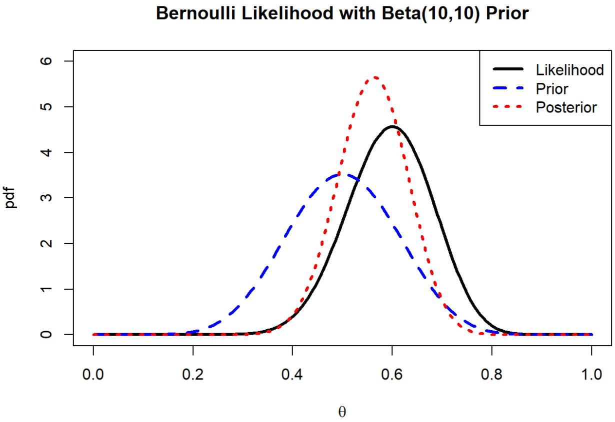

Bernoulli Likelihood with Beta(1,1) Prior
pdf
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood
Prior
Posterior

Example

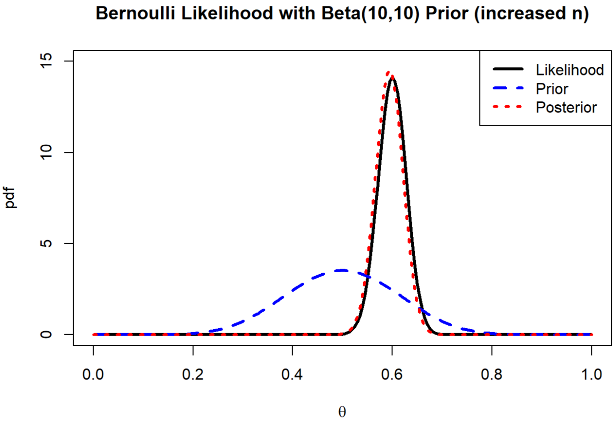


Bernoulli Likelihood with Beta(2,2) Prior
Likelihood
Prior
Posterior
pdf
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
θ

Example


Bernoulli Likelihood with Beta(10,10) Prior
pdf
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
θ
Likelihood
Prior
Posterior

Example


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

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: $\theta \,|\, k \sim \mathrm{Beta}(251528, 241946)$. Then,

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