Statistics for Machine Learning and Big Data

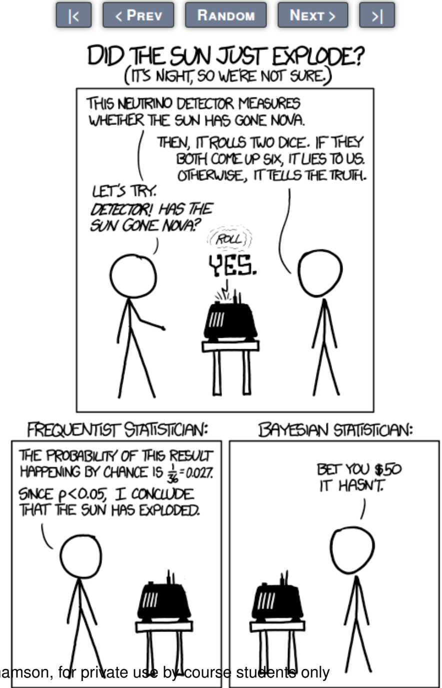
An Introduction

Part 4: Bayesians and frequentists

Jeff Abrahamson

7 April 2015

Frequentists vs. Bayesians



Freqentists vs. Bayesians

https://xkcd.com/1132/

Frequentists

Flipping coins

You have a coin that when flipped ends up head with probability p and ends up tail with probability $1 - p$. (The value of p is unknown.)

Trying to estimate p, you flip the coin **14** times. It ends up heads **10** times.

Then you have to decide on the following event: "In the next two tosses we will get two heads in a row."

Would you bet that the event will happen or that it will not happen?

http://www.behind-the-enemy-lines.com/2008/01/are-you-bayesian-or-frequentist-or.html

Flipping coins, frequentist view

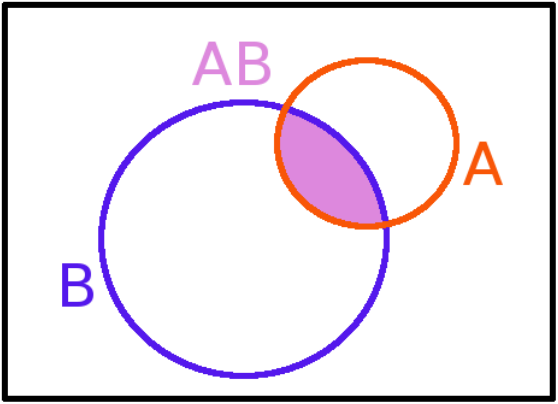
We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

$$
\begin{aligned}
p &= \frac{10}{14} \approx .714 \\
\Pr(HH) &= p^2 \approx .51
\end{aligned}
$$

Bayesians

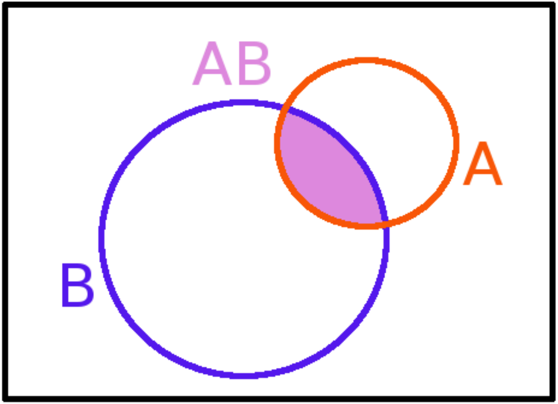
Review: conditional probability

$$\Pr(A \mid B)$$



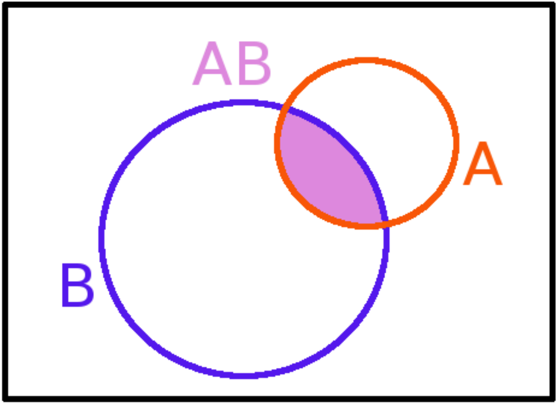
Review: conditional probability

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$



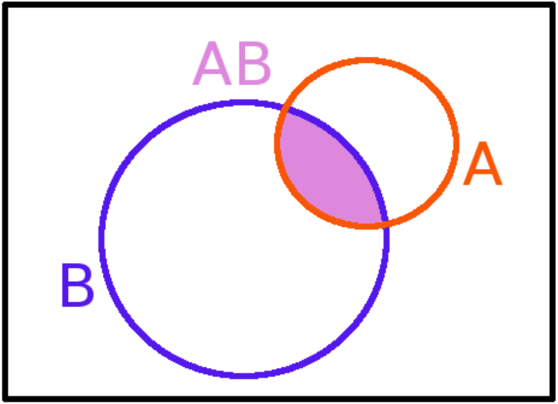
Review: conditional probability

$$\Pr(A \mid B)\Pr(B) \quad = \quad \Pr(A \cap B)$$



Review: conditional probability

$$\begin{aligned} \Pr(A \mid B)\Pr(B) &= \Pr(A \cap B) \\ &= \Pr(B \mid A)\Pr(A) \end{aligned}$$



Review: Bayes theorem

$$
\begin{aligned}
\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\
&= \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}
\end{aligned}
$$

Review: Bayes theorem

$$\begin{aligned}\Pr(A_j \mid B) &= \frac{\Pr(A_j \cap B)}{\Pr(B)} \\ &= \frac{\Pr(B \mid A_j)\Pr(A_j)}{\Pr(B)}\end{aligned}$$

if $A = \cup_i A_i$ covers $A \cup B$ and is a non-overlapping cover of A

Review: Bayes theorem

$$
\begin{aligned}
\Pr(A_j \mid B) &= \frac{\Pr(A_j \cap B)}{\Pr(B)} \\
&= \frac{\Pr(B \mid A_j)\Pr(A_j)}{\Pr(B)} \\
&= \frac{\Pr(B \mid A_j)\Pr(A_j)}{\sum_i \Pr(B \cap A_i)} \\
&= \frac{\Pr(B \mid A_j)\Pr(A_j)}{\sum_i \Pr(B \mid A_i)\Pr(A_i)}
\end{aligned}
$$

if $A = \cup_i A_i$ covers $A \cup B$ and is a non-overlapping cover of A

Review: Bayes theorem

$$
\begin{aligned}
\Pr(A \mid B) &= \frac{\Pr(A \cap B)}{\Pr(B)} \\
&= \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}
\end{aligned}
$$

Review: Bayes theorem

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)}$$

Review: Bayes theorem

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)} = \frac{\Pr(E \mid H)}{\Pr(E)} \cdot \Pr(H)$$

Here H is the hypothesis and E is evidence.

Review: Bayes theorem

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)}{\Pr(E)} \cdot \Pr(H)$$

Prior probability of H

Review: Bayes theorem

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)}{\Pr(E)} \cdot \Pr(H)$$

Impact of E on H

Review: Bayes theorem

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)}{\Pr(E)} \cdot \Pr(H)$$

Likelihood of observing E given H

Review: Bayes theorem

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)}{\Pr(E)} \cdot \Pr(H)$$

Marginal likelihood or *model evidence*

Review: Bayes theorem

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)}{\Pr(E)} \cdot \Pr(H)$$

Posterior probability of H

Using Bayes theorem

If the evidence E doesn't fit H, we reject H.

Using Bayes theorem

If the evidence E doesn't fit H, we reject H.

If the evidence E is extremely unlikely, we also reject H.

Example: a baby!

A friend had a baby, but we don't know more.
A mutual friend says it's a boy and shows us a picture of a baby boy.

http://en.wikipedia.org/wiki/Bayesian_inference

Example: a baby!

A friend had a baby, but we don't know more.
A mutual friend says it's a boy and shows us a picture of a baby boy.

E: picture of a baby boy
H: the baby is a boy

$$\begin{aligned}
\Pr(H \mid E) &= \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)} \\
&= \frac{(1)(\frac{1}{2})}{\frac{1}{2}} \\
&= 1
\end{aligned}$$

Example: a baby!

A friend had a baby, but we don't know more.
A mutual friend says it's a dog and shows us a picture to prove it.

Example: a baby!

A friend had a baby, but we don't know more.
A mutual friend says it's a dog and shows us a picture to prove it.

E: picture of a puppy
H: the baby is a boy

$$\begin{aligned}
\Pr(H \mid E) &= \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)} \\
&= \frac{(1)(0)}{\epsilon} \\
&= 0
\end{aligned}$$

Example: medical testing

A disease affects 0.1% of the population.
The test is not perfect:

	+	−
Sick	.95	.05
Healthy	.01	.99

Example: medical testing

A disease affects 0.1% of the population.
The test is not perfect:

	+	−
Sick	.95	.05
Healthy	.01	.99

Exercise

Example: medical testing

A disease affects 0.1% of the population.
The test is not perfect:

	$+$	$-$
Sick	.95	.05
Healthy	.01	.99

$\Pr(S \mid -)$	.005%
$\Pr(S \mid +)$	8.6%
$\Pr(H \mid -)$	99.9%
$\Pr(H \mid +)$	91.3%

Example: medical testing

A disease affects 0.1% of the population.
The test is not perfect:

	$+$	$-$
Sick	.95	.05
Healthy	.01	.99

$$\begin{aligned}
\Pr(S \mid -) &= \frac{\Pr(- \mid S)\Pr(S)}{\Pr(-)} \\
&= \frac{\Pr(- \mid S)\Pr(S)}{\Pr(- \mid H)\Pr(H) + \Pr(- \mid S)\Pr(S)} \\
&= \frac{(.05)(.001)}{(.95)(.999) + (.05)(.001)} = .005\%
\end{aligned}$$

Example: medical testing

A disease affects 0.1% of the population.
The test is not perfect:

	+	$-$
Sick	.95	.05
Healthy	.01	.99

$$
\begin{aligned}
\Pr(S \mid +) &= \frac{\Pr(+ \mid S)\Pr(S)}{\Pr(+)} \\
&= \frac{\Pr(+ \mid S)\Pr(S)}{\Pr(+ \mid H)\Pr(H) + \Pr(+ \mid S)\Pr(S)} \\
&= \frac{(.95)(.001)}{(.01)(.999) + (.95)(.001)} = 8.6\%
\end{aligned}
$$

Example: medical testing

A disease affects 0.1% of the population.
The test is not perfect:

	$+$	$-$
Sick	.95	.05
Healthy	.01	.99

$$
\begin{aligned}
\Pr(H \mid -) &= \frac{\Pr(- \mid H)\Pr(H)}{\Pr(-)} \\
&= \frac{\Pr(- \mid H)\Pr(H)}{\Pr(- \mid H)\Pr(H) + \Pr(- \mid S)\Pr(S)} \\
&= \frac{(.99)(.999)}{(.99)(.999) + (.05)(.001)} = 99.9\%
\end{aligned}
$$

Example: medical testing

A disease affects 0.1% of the population.
The test is not perfect:

	+	$-$
Sick	.95	.05
Healthy	.01	.99

$$
\begin{aligned}
\Pr(H \mid +) &= \frac{\Pr(+ \mid H)\Pr(H)}{\Pr(+)} \\
&= \frac{\Pr(+ \mid H)\Pr(H)}{\Pr(+ \mid H)\Pr(H) + \Pr(+ \mid S)\Pr(S)} \\
&= \frac{(.01)(.999)}{(.01)(.999) + (.95)(.001)} = 91.3\%
\end{aligned}
$$

Flipping coins

You have a coin that when flipped ends up head with probability p and ends up tail with probability $1 - p$. (The value of p is unknown.)

Trying to estimate p, you flip the coin **14** times. It ends up headss **10** times.

Then you have to decide on the following event: "In the next two tosses we will get two heads in a row."

Would you bet that the event will happen or that it will not happen?

http://www.behind-the-enemy-lines.com/2008/01/are-you-bayesian-or-frequentist-or.html

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

This is going to hurt.

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

p is now a distribution instead of a value.

Instead of ML estimate of p, we consider p as a random variable.

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

$$\Pr(p \mid \text{data}) = \frac{\Pr(\text{data} \mid p)\,\Pr(p)}{\Pr(\text{data})}$$

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

$$\Pr(p \mid \text{data}) = \frac{\Pr(\text{data} \mid p) \Pr(p)}{\Pr(\text{data})}$$

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

$$\Pr(\text{data} \mid p) = \binom{14}{10} p^{10}(1-p)^4$$

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

Ignoring constants,

$$\Pr(p \mid \text{data}) \propto p^{10}(1-p)^4 \Pr(p)$$

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

What is $\Pr(p)$?

Convenient to use the Beta distribution (a.k.a. the conjugate prior to the binomial distribution).

$$\mathrm{Beta}(p; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$$

For us, that means

$$\Pr(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$$

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

$$
\begin{aligned}
\Pr(p \mid \text{data}) &\propto p^{10}(1-p)^4 p^{a-1}(1-p)^{b-1} \\
&\propto p^{10+a-1}(1-p)^{4+b-1} \\
&\propto \text{Beta}(p; a+10, b+4)
\end{aligned}
$$

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

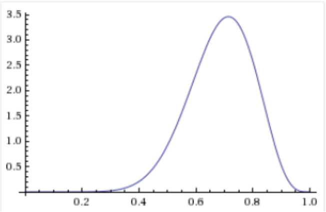
If we assume we know nothing about p, then let's say the prior is uniform, which is $a = b = 1$.

$$\Pr(p \mid \text{data}) \propto \text{Beta}(p; a + 10, b + 4) = \text{Beta}(p; 11; 5)$$

Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

We can at least plot Beta(p; 11; 5).



Flipping coins, Bayesian view

We saw 10 heads in 14 tosses. What's the probability of the next two tosses are both heads?

And let's just jump to the solution,

$$\begin{aligned}\Pr(HH \mid \text{data}) &= \int_0^1 \Pr(HH \mid p)\Pr(p \mid \text{data})\,\mathrm{d}p \\ &= \frac{B(13,5)}{B(11,5)} \approx .485\end{aligned}$$

Summary

The frequentist bets on HH, $\Pr(HH) \approx .51$.

The Bayesian bets against HH, $\Pr(HH) \approx .49$.

Fair coins

Suppose we flip 20 coins and get $12H + 8T$.

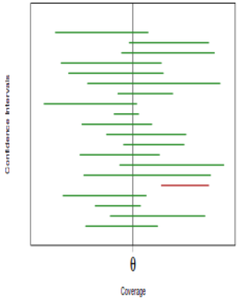
Exercise.

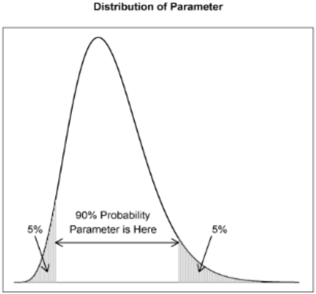
- What is the probability of being closer to the "truth" assuming the coin is fair?
- What if we got $13H + 7T$?

Confidence interval vs credible interval

Interpretations of Confidence

- Frequentist: A collection of intervals with 90% of them containing the true parameter
- Bayesian: An interval that has a 90% chance of containing the true parameter.





http://www.stat.ufl.edu/archived/casella/Talks/BayesRefresher.pdf

Break

Questions?

`purple.com/talk-feedback`