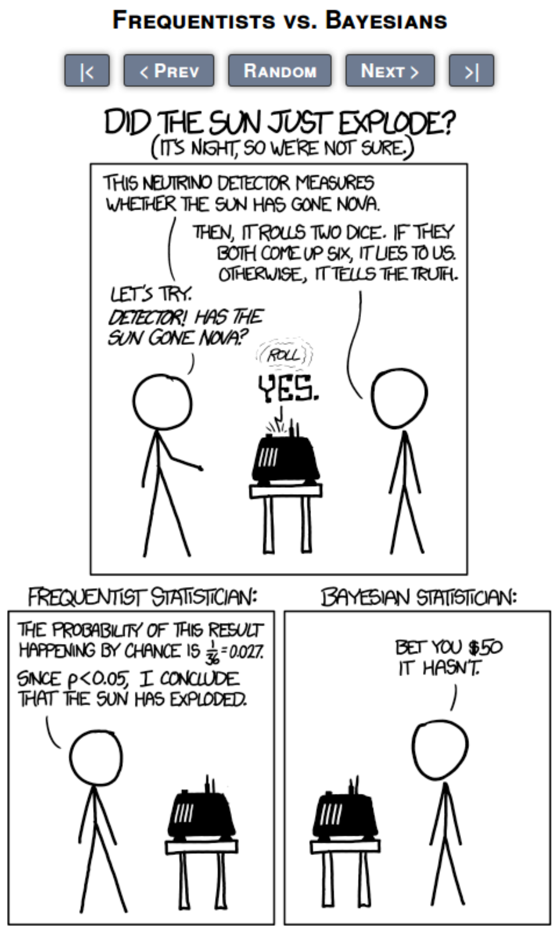


Our goal is to understand this.

Note that betting is not just probability, but also utility. If the sun exploded, money no longer has value, so one should always make this bet.

Freqentists vs. Bayesians

https://xkcd.com/1132/

1 Frequentists

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?

$$p = \frac{10}{14} \approx .714$$

$$\Pr(HH) = p^2 \approx .51$$

We say that this is the best or *maximum likelihood* estimate for p.

This is the probability we learn in school.

(Not to say that frequentists have it easy.)

2 Bayesians

Bayesians

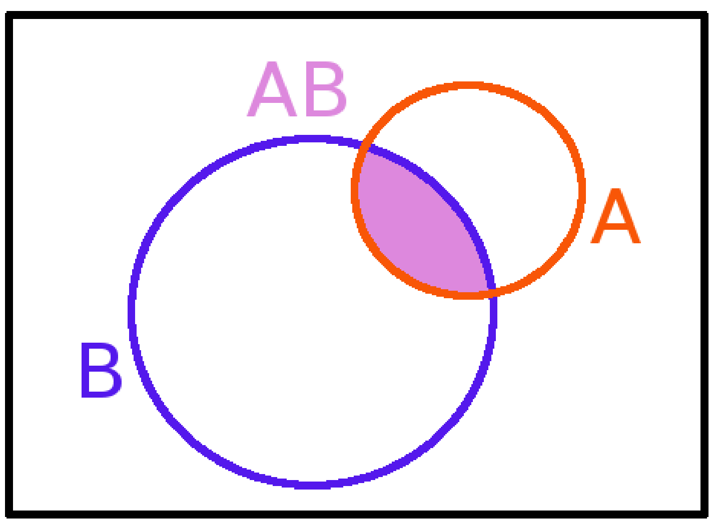
Review: conditional probability

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

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

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

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



Review: Bayes theorem

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

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

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

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

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

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

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

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

Here H is the hypothesis and E is evidence.

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

Prior probability of H

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

Impact of E on H

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

Likelihood of observing E given H

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

Marginal likelihood or *model evidence*

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

Posterior probability of H

Using Bayes theorem

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

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`

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

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

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

Exercise

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

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

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

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

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

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. Bayesian inference became popular with greater computing power.

p is now a distribution instead of a value.

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

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

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

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

Ignoring constants,

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

What is $\Pr(p)$?

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

If the posterior distributions $\Pr(\theta \mid x)$ are in the same family as the prior probability distribution $\Pr(\theta)$, then prior and posterior are then called **conjugate distributions**, and the prior is called a **conjugate prior** for the likelihood function.

The **beta distribution** is a family of continuous probability distributions with two shape parameters, α and β.

$$f(x;\alpha,\beta) = \text{constant} x^{\alpha-1}(1-x)^{\beta-1}$$

Sometimes p and q instead of α and β. Because approaches Bernoulli when $\alpha, \beta \to 0$

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

`ss.beta` exists.

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

For us, that means

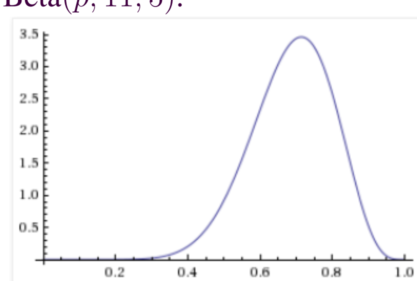
$$\Pr(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$$

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

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

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

We can at least plot $\text{Beta}(p; 11; 5)$.



And let's just jump to the solution,

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

The equivalent of a confidence interval is a credible interval.

Summary

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

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

Both viewpoints have their place. The topic extends well beyond this course. Knowing when to use Bayesian vs frequentist methods is an advanced topic.

Humans might be more Bayesian than frequentist when not being scientists.

A Bayesian, strictly speaking, it's incorrect to say "I predict that there's a 30% chance of P", but rather "Based on the current state of my knowledge, I am 30% certain that P will occur."

Bayesian methods are certainly far more elegant theoretically, but generally far, far more difficult to apply in practice, due to computational issues and the necessity of identifying an appropriate prior (a fundamentally non-mathematical problem, and often intractable). Also, scientists find (or think they find) p-values easier to understand than posterior distributions. This is why the medical literature (for instance) is full of frequentism.

http://scienceblogs.com/goodmath/2008/04/07/schools-of-thought-in-probabil/ \#comment-15371

Beware bogus science by using bogus priors! **Example:** Physics seems to be finely tuned to support human life, so God must have created us. But we have no information on the prior (what the universe would be like with different constants).

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 $13H + 7T$?

$$\Pr(12H + 8T) = \binom{20}{12}\left(\frac{1}{2}\right)^{20} \approx .12$$

```
scipy.misc.comb(20, 12) * .5**20
```

$$\Pr(11H + 9T) + \Pr(10H + 10T) + \Pr(9H + 11T) \approx .50$$

```
sum([scipy.misc.comb(20, i) * .5**20 for i in [9, 10,
11]])
```

So 50% chance of being "closer to the truth", but still 50% chance of being this far away

$$\Pr(13H + 7T) \approx .074$$

So now 74% chance of being closer to the expected population mean.

Suppose $p = \frac{13}{20}$. Then

$$\Pr(13H + 7T) = \binom{20}{13} p^{13}(1-p)^7 \approx .18$$

which is about $2.5\times$ better odds.

What's the probability that the coin is not fair?

Frequentists can't help here. Bayesians see an experiment and tests the hypothesis: "What is the probability that the coin is fair given the experimental results?"

$$\Pr(\text{fair} \mid 13H{+}7T) = \frac{\Pr(13H + 7T \mid \text{fair} \Pr(\text{fair}}{\Pr(13H + 7T)} = \frac{\Pr(13H + 7T \mid \text{fair} \Pr(\text{fair}}{\sum_i \Pr(13H + 7T \mid i) \Pr(i)}$$

where we are summing over all alternative realities i.

Added complexity, not clear we've helped. But we have made clear what assumptions we'd need.

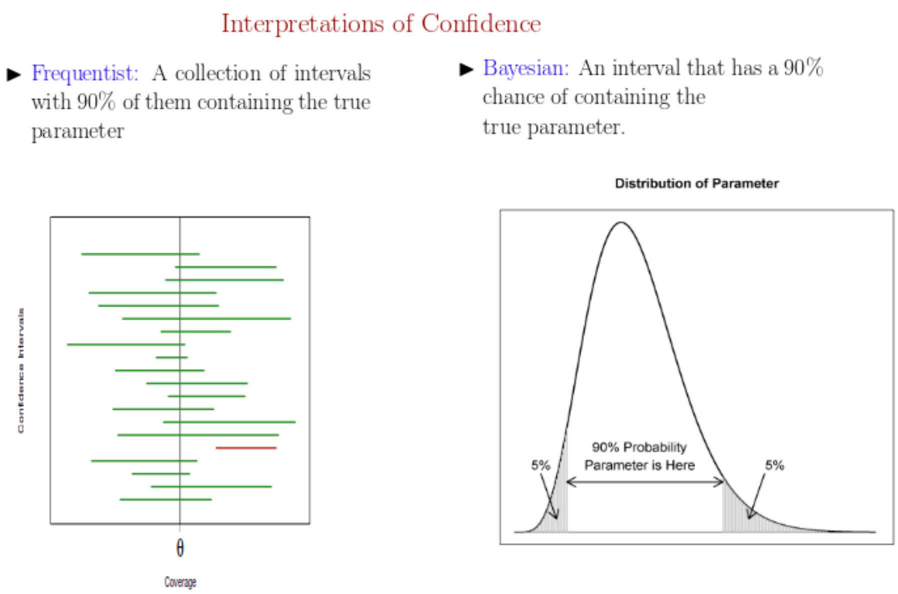
Frequentist: Probability of observing a particular event given the the truth. This is the basis of significance testing.

Bayesian: Probability of the truth given an event.

Frequentists compute precisely but maybe not what we want. Bayesians compute often approximately but maybe what we want.

http://www.met.reading.ac.uk/~sws97mha/Publications/Bayesvsfreq.pdf

Confidence interval vs credible interval



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

3 Break

Break

Questions?

purple.com/talk-feedback