Unit 6: Bayesian Statistics

1. What is Probability?

5/03/2021

Recap from last time

1. Mixed effects models are a method for statistically modeling the factors that make everyone the same (what we’ve done so far) while accounting for the ways in which everyone is different.

2. Fixed effects are things that you think are universal (or you experimentally manipulated),
   Random effects are things that you think might vary across people. The same factor can contribute to both.

3. This is an active area of research in statistics, and the solutions are less tidy (but also probably less wrong)
   than the models we have used so far

Key ideas

1. What you mean by “probability” has implications for what statistical tools you should use

2. Bayesian probability conceives of probability as *subjective* rather *objective*. That means you can talk about probability of beliefs rather than of data.

3. Bayesian methods are more computationally complex, and have their own issues, but can sometimes be more useful and can often be more intuitive

Rules of probability

For any event *A*, let P(A) be the probability of event *A*

1. $0 \leq P(A) \leq 1$
2. $P(A) + P(\sim A) = 1$
3. Events *A* and *B* are *independent* if P(A + B) = P(A)P(B)

   … etc

But what *is* probability?

Why do we think that if a coin is fair P(heads) = .5?

Classical probability

The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible.

Pierre-Simon Laplace (1812)

P(heads) = .5 because there are two outcomes,
and nothing makes us think they are not equally likely.

So P(heads) = ½ = .5

The problems with classical probability

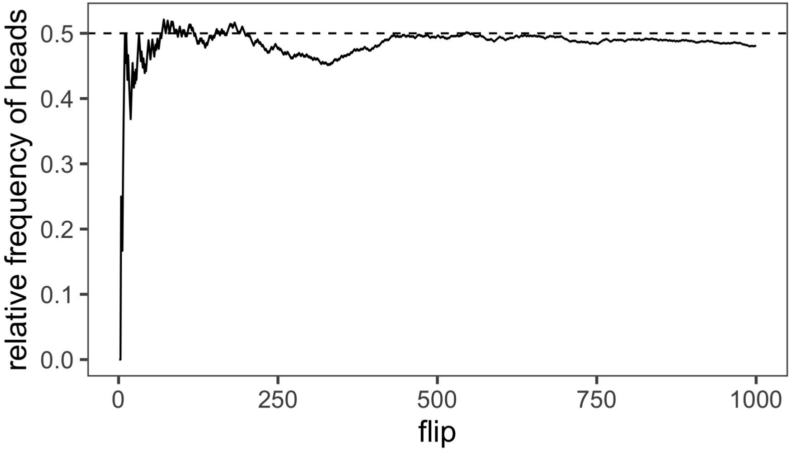
The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible. Pierre-Simon Laplace (1812)

1. *It's circular.* A fair coin is defined a coin that is fair
2. *It's hard to generalize*. In most cases, it's hard to justify this principle of indifference. We'd like to be able to think about cases where we don't know all the possible outcomes, where the possible outcomes aren't equally likely, etc.

   E.g. probability a bus comes on time.

Frequentist probability

The probability of an event is defined by the limit of its *relative frequency* over many trials of an experiment.



P(heads) = .5 because if you flip a coin over and over and over again for long enough, half of the flips will have come up heads.

This is the logic behind the simulations we built and analyzed in Unit 2, and the broader Null Hypothesis Testing framework

The problems with frequentism

The probability of an event is defined by the limit of its *relative frequency* over many trials of an experiment.

But what about events that have never happened before and will never happen again?

E.g. Probability that we will still be wearing masks in the fall

What about things that aren't "events"

E.g. Probability that Germ theory is correct?

Bayesian probability

Probability is *subjective*, it exists only in your mind.

What you mean when you talk about *P(A)* is the strength of your belief that *A* will happen. Think of it as how much you would be willing to bet on A.

Further, your *P(A)* can be different from my *P(A)*.

Reverend Thomas Bayes

P(heads) = .5 because I expect it to come up heads 50% of the time based on my prior belief about the coin and my experience flipping it.

Published posthumously by Price, and generalized into the from we use today by Laplace

But how should you form your beliefs?

In practice, we don't want to say you can have any old belief.
We want to talk about the belief that a **rational** agent should have after observing some data

Likelihood
(What the data say)

Prior Probability
(What you used to believe)

Bayes' Rule:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Posterior Probability
(What you should believe now)

The problem with Bayesianism

Bayes rule gives you a way to compute how much you should believe in some hypothesis (posterior) if you know three things:

1. The likelihood of the data under that hypothesis
2. The prior probability of that hypothesis
3. The probability of the data

Problem: We only know the likelihood (**1**)

Priors are the biggest problem with Bayesianism because priors are *subjective* (i.e. reasonable people can disagree about the right prior).

There are some techniques for dealing with this, but it's a real problem.

Still… *priors matter*!

Why priors matter

Suppose that you wake up with feeling like you have a fever.

$P(fever|cold)$ = .01

$P(fever|covid\text{-}19)$ = .6

(I made these numbers up)

$P(fever|malaria)$ = 1

Which of these ailments do you think you are most likely to have?

Probably covid-19, because *P(covid-19) >> P(malaria).*

But note you probably don't have a cold because $P(fever|cold)$ is very low

The problem with Bayesianism

Bayes rule gives you a way to compute how much you should believe in some hypothesis (posterior) if you know three things:

1. The likelihood of the data under that hypothesis
2. The prior probability of that hypothesis
3. The probability of the data

Problem: We only know the likelihood (**1**)

You can't compute the probability of the data, but often you don't actually care about the posterior probability of the hypothesis H_1.

You only care whether it is more probable or less probable than some alternative hypothesis H_2

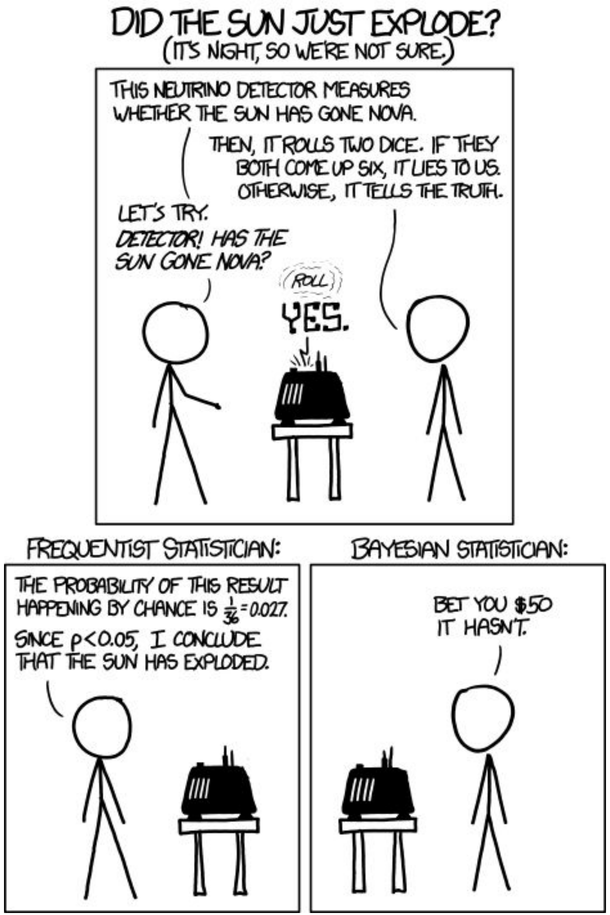
The relative probability of two hypotheses

$$\frac{\mathrm{P}(H_1|D)}{\mathrm{P}(H_2|D)} = \frac{\dfrac{\mathrm{P}(D|H_1)P(H_1)}{P(D)}}{\dfrac{\mathrm{P}(D|H_2)P(H_2)}{P(D)}}$$

$$= \frac{\mathrm{P}(D|H_1)P(H_1)\cancel{P(D)}}{\mathrm{P}(D|H_2)P(H_2)\cancel{P(D)}}$$

$$= \frac{\mathrm{P}(D|H_1)P(H_1)}{\mathrm{P}(D|H_2)P(H_2)}$$

Often you actually want to compare hypotheses



Null hypothesis testing draws inferences by rejecting the Null (i.e. finding that you observed data that is unlikely under the null)

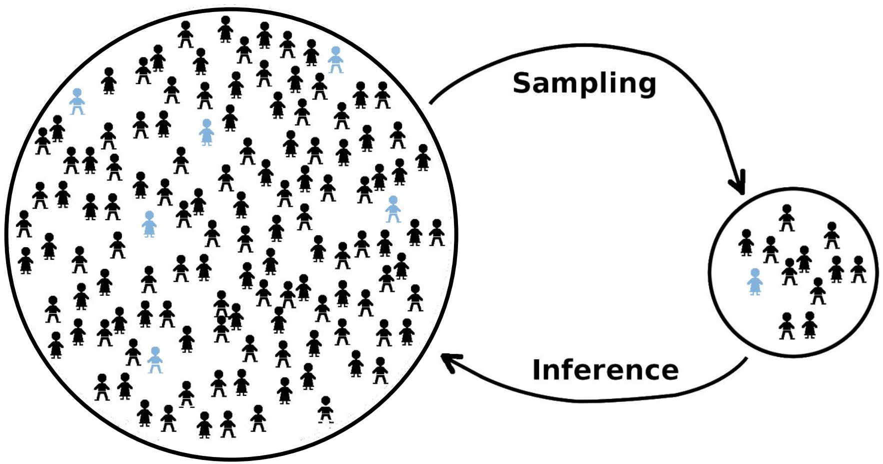
But sometimes the data are just unlikely!

Sometimes the data are even more unlikely under a reasonable alternative hypothesis.

Randall Monroe, XKCD
https://xkcd.com/1132/

Frequentism in practice

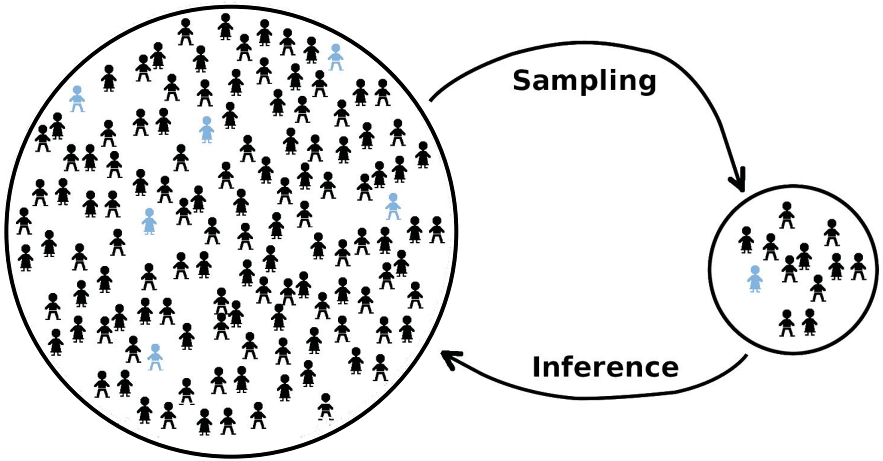
We assume that there is an *unknown* but *fixed* underlying parameter, θ, for a population (i.e., the proportion of people who are left handed). Random variation (environmental factors, measurement errors, ...) means that each observation does not result in the true value.



Slide adapted from Claus Ekstrøm

The meta-experiment

Think of the data you have in hand as one realization of all possible datasets that you could have seen if you had run the experiment over and over again.

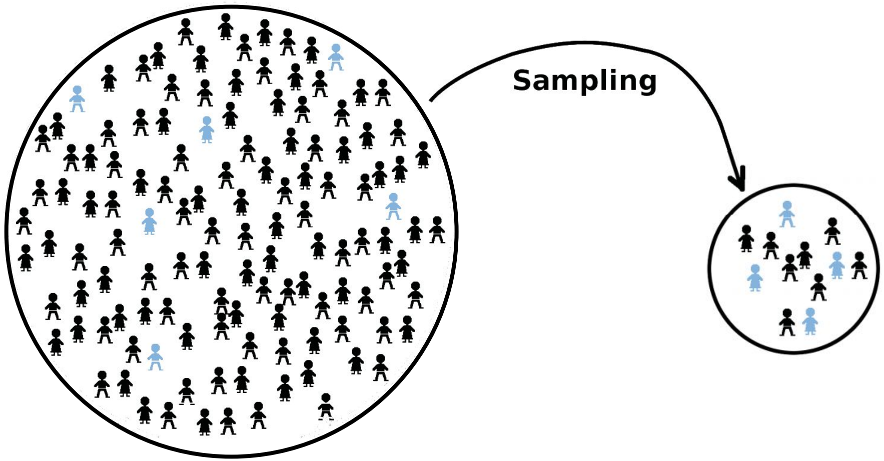


1/12 ~ .083

Slide adapted from Claus Ekstrøm

The meta-experiment

Think of the data you have in hand as one realization of all possible datasets that you could have seen if you had run the experiment over and over again.



1/12 ~ .083

4/12 ~ .333

Slide adapted from Claus Ekstrøm

Frequentist Confidence intervals

In frequentism, the population mean is real, but unknown, and unknowable, and can only be *estimated* from the data.

Knowing the distribution for the sample mean, you constructs a **confidence interval**, centered at the sample mean.

- Either the true mean is in the interval or it is not. Can't say there's a 95% probability (long-run fraction having this characteristic) that the true mean is in this interval, because it's either already in or not.
- Reason: true mean is fixed value, which doesn't have a distribution.
- The sample mean does have a distribution! That's why you say things like "95% of similar intervals would contain the true mean, if each interval were constructed from a different random sample."

Bayesian Credible Intervals

Bayesians have an altogether different world-view.

They say that only the data are real. The population mean is an abstraction, and as such you should believe some values more than others based on the data and your prior beliefs.

The Bayesian constructs a **credible interval**, centered near the sample mean, but tempered by “prior” beliefs concerning the mean.

Now the Bayesian can say what the frequentist cannot: “There is a 95% probability (degree of believability) that this interval contains the mean.”

Frequentism vs Bayesian

In frequentism, probabilities are *objective*. They are properties of the world defined by the long-run outcomes of random process.

The *parameters* we want to estimate have some true exact value, and we can try to estimate them by talking about how future samples from the random process would look. $\mathrm{P}(D|H)$

In Bayesianism, probabilities are *subjective*. They are properties of the mind of the experimenter.

What are estimating the parameters of hypotheses and not the world. We can talk about how much or how little certainty we have about the truth of our hypotheses. $\mathrm{P}(H|D)$

Key ideas

1. What you mean by “probability” has implications for what statistical tools you should use

2. Bayesian probability conceives of probability as *subjective* rather *objective*. That means you can talk about probability of beliefs rather than of data.

3. Bayesian methods are more computationally complex, and have their own issues, but can sometimes be more useful and can often be more intuitive