Bayesian Reasoning: Answering the Correct Questions

QTM 110

KEVIN MCALISTER

Administrative Stuff

Problem Set #4 will be posted this evening

- One IV question, one Bayes theorem question
- Push on what we talked about in class to really force understanding!

Due next Thursday evening.

PS3 grades will be released before Monday.

Previously on…

Bayes' Theorem:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

$P(A|B) \neq P(B|A)$!

The conversion depends on the marginal probability that B occurs and the marginal probability that A occurs.

- We're mostly concerned with the marginal probability of B occurring – this is the base rate or the prior

Previously on…

If we ignore P(B), we may fall victim to the **base rate fallacy**:

- Rare events are rare and common events are common
- Ignoring the base rate when presented with new evidence
- Even if we receive evidence that a rare event has occurred, we should be skeptical that it actually happened
- Gotta keep the base rate in mind – don't assume all shy people are librarians!
- Probability and evidence can be wonky. Bayes' theorem gives us a principled way to think through these problems and make conclusions.

Psychics?

I haven't told you this, but I am psychic!

- Really! I am.
- I promise you.
- Let me show you.

Psychics?

https://kmcalist.shinyapps.io/ThePsychicGame/

It’s the same thing…

Different day, same story:

$$\Pr(Innocent \mid Matching) \neq \Pr(Matching \mid Innocent)$$

It’s the same thing…

Different day, same story:

$$\Pr(Innocent \mid Matching) \neq \Pr(Matching \mid Innocent)$$

It doesn’t matter what the question is, the same thing holds!

$$\Pr(Not\ Psychic \mid 10\ Correct\ Guesses) \neq \Pr(10\ Correct\ Guesses \mid Not\ Psychic)$$

We’re answering the wrong question!

This is a generic issue with **p-values** – the probability that we would observe certain data or estimates given that our null hypothesis is true!!!

In general, we care about the probability that the hypothesis is true – not the other way around.

Psychics, maybe?

Let's use Bayes' Theorem.

Priors and Scientific Reasoning

The uncomfortable truth:

- The prior probability is the analyst's **informed belief** that the hypothesis is true.
- Priors are pulled from the world around us and are based on expertise and knowledge about the problem.

Priors and Scientific Reasoning

The uncomfortable truth:

- The prior probability is the analyst's **informed belief** that the hypothesis is true.
- Priors are pulled from the world around us and are based on expertise and knowledge about the problem.

But Kevin (this conversation requires a first name basis because it is so deeply critical to how science works), that seems awfully **subjective**. Couldn’t someone just say that they believe that the effect is real 100% of the time and assign a probability of 1 for the prior?

Priors and Scientific Reasoning

The uncomfortable truth:

- The prior probability is the analyst's **informed belief** that the hypothesis is true.
- Priors are pulled from the world around us and are based on expertise and knowledge about the problem.

But Kevin (this conversation requires a first name basis because it is so deeply critical to how science works), that seems awfully **subjective**. Couldn't someone just say that they believe that the effect is real 100% of the time and assign a probability of 1 for the prior?

- They could! But, this is not good science.
- Prior beliefs are more plentiful and consistent than we may think.

Priors and Scientific Reasoning

A quick exercise. By show of hands:

- How many of you believe that the sun will rise tomorrow?
- How many of you believe that the number of people waiting in line at Starbucks tomorrow morning at 9 AM will be greater than 10?
- How many of you believe that the Carolina Panthers will win the NFC South this season?
- How many of you feel like you could accurately answer a question about this NFL season?
- (If yes) How many of you believe that the Carolina Panthers will win the NFC South this season?
- How many of you believe that having more teacher’s aides in the classroom leads to better student performance on tests?

Priors and Scientific Reasoning

All of the probabilities that we just came up with are pretty reasonable **prior guesses** to the probability that an event will occur.

Good priors come from a mixture of **diverse opinions** and **expertise**. This is called **prior elicitation**.

Important question: As a researcher, why do we collect data related to a specific question?

Priors and Scientific Reasoning

All of the probabilities that we just came up with are pretty reasonable **prior guesses** to the probability that an event will occur.

Good priors come from a mixture of **diverse opinions** and **expertise**. This is called **prior elicitation**.

Important question: As a researcher, why do we collect data related to a specific question?

Answer: Because we believe that an effect is real and we want to prove it! Or be surprised that it isn't there. But, we only collect the data because we believe that there is an effect or a lack of effect. We don't go into problems completely agnostic to the results.

To deny that is to deny that we have developed some level of expertise and that we have some reason to believe that our research is important.

It's why AI will never really be a perfect substitute for humans in research…

Psychics, maybe?

Let's think about priors

Psychics, maybe?

Given my prior, 20 coin flips is not enough to change the probability that she is psychic!

- I think that being a psychic is extremely unlikely – it is much more likely that she just got lucky!
- Or (even more likely) that the coin is rigged or the game is rigged.
- I update my prior belief with the evidence – on its face, it looks like I'm psychic. But, the evidence doesn't really move the needle.

With respect to my prior, I don't want to fall victim to the **base rate fallacy**:

- What I saw happen was rare, but I shouldn't necessarily interpret this as world-changing evidence that an even rarer (almost impossible) thing is true!

Reasonable Evidence

But Kevin (first name again), I believe that your prior is wrong and severely underestimates the probability that someone might be psychic.

Is there any evidence that would sway you?

Reasonable Evidence

But Kevin (first name again), I believe that your prior is wrong and severely underestimates the probability that someone might be psychic.

Is there any evidence that would sway you?

There is!

- I suppose there is ***some*** chance that psychics exists
- But, my belief is that this is ***really unlikely***

Reasonable Evidence

But Kevin (first name again), I believe that your prior is wrong and severely underestimates the probability that someone might be psychic.

Is there any evidence that would sway you?

There is!

- I suppose there is ***some*** chance that psychics exists
- But, my belief is that this is ***really unlikely***

Given a prior of 1/100,000,000, I would really only need to see 26 in a row to say that I am convinced that it's more likely that I'm psychic than not

- Would that convince you?

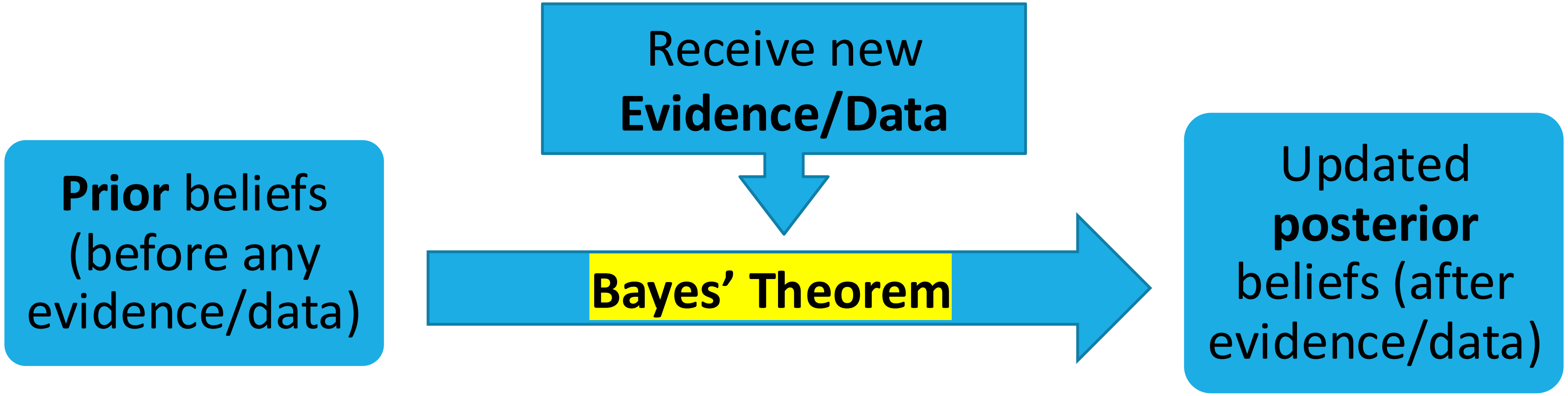
Bayesian Reasoning

This is the beautiful balance of using Bayes' theorem to update beliefs:

- In a world where I receive no new data, I have a prior belief that something will occur
- In the face of new information, my new belief is a combination of my **prior** and the evidence from the new information
- Even if the data and my prior **disagree**, I can cut the difference rationally using Bayes' theorem
- At a certain point, **regardless of my initial beliefs**, there can be too much evidence to continue to hold my prior belief

Bayes’ Theorem, Evidence, and Beliefs

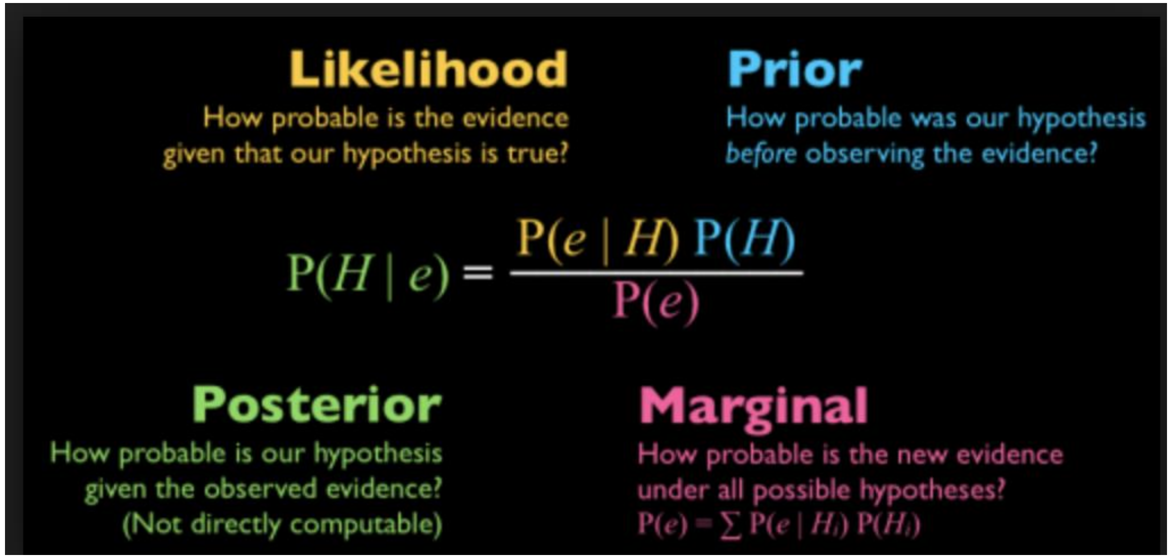
- Translating Evidence to Beliefs, and Updating Those Beliefs
  - How likely is a claim or idea to be true, given some evidence/data
    - In our daily lives
    - About scientific hypotheses



- This process is, broadly, how humans reason about everything!

Bayes’ Theorem, Evidence, and Beliefs

- We start with a prior probability – how likely we think some hypothesis/claim is at baseline
- We add some evidence, that may or may not support our prior
- We divide by the marginal total probability of seeing the evidence at all
  - **Law of Total Probability**
- Combining these with **Bayes’ theorem** gives us a posterior probability for a hypothesis/claim
  - Notice if $\frac{Likelihood}{Marginal} = \frac{P(e|H)}{P(e)} > 1$ our hypothesis becomes more likely, otherwise less likely
  - If the evidence we saw is more likely under our hypothesis than the “average” alternative(s), the probability of our hypothesis rises



Source: https://medium.com/@mark.rethana/bayesian-statistics-and-naive-bayes-classifier-33b735ad7b16

Bayes' Theorem, Evidence, and Beliefs

- Translating Evidence to Beliefs, and Updating Those Beliefs

Prior	Evidence	Posterior
20% chance it is raining	I played disc golf	4% chance it's raining
0.05% chance my neighbor is a terrorist	On security algorithm list that purportedly predicts terrorists	Small chance my neighbor is a terrorist since there are very few terrorists in the world.
<1% chance my dog has run away	Can't find him	He probably hasn't run away (he's probably just chewing on a bone under the bed)
30% chance Dana eats my ice cream	No ice cream in the freezer	Very likely she ate it
1% chance I have Celiac	Positive Celiac test	>1% chance I have Celiac, but depends on strength of evidence from test
5% chance this new cancer drug extends survival	Statistically significant result in well-powered RCT ($p < .05$)	46% chance this new drug extends cancer survival

Bayes’ Theorem, Evidence, and Beliefs

- Translating Evidence to Beliefs, and Updating Those Beliefs

Prior	Evidence	Posterior
20% chance it isn’t raining	I played disc golf	4% chance it’s raining
0.05% chance my neighbor is a terrorist	On security algorithm list that purportedly predicts terrorists	Small chance my neighbor is a terrorist since there are very few terrorists in the world.
<1% chance my dog has run away	Can’t find him	He probably hasn’t run away (he’s probably just chewing on a bone under the bed)
30% chance Dana eats the Ben and Jerry’s I bought at Publix	No ice cream in the freezer	Very likely she ate it
1% chance I have Celiac	Positive Celiac test	>1% chance I have Celiac, but depends on strength of evidence from test
5% chance this new cancer drug extends survival	**Statistically significant result in well-powered RCT ($p < .05$)**	**46% chance this new drug extends cancer survival**

Bayesian Reasoning as a Scientific Device

The p-value:

$$\Pr(Estimate \mid No\ Effect)$$

After seeing the data, what is the probability that we would have gotten our estimate in a world where there is no effect.

For the ATT:

$$\Pr(|ATE| \geq x \mid \ ATT = 0)$$

A p-value tells us the probability that we would have observed the ATE we did if the true data generating process was a null effect.

Scientist: We see that the p-value is sufficiently small, so we should reject the null. Now, we should say that the effect that we see is true? Problem?

Bayesian Reasoning as a Scientific Device

Again, we're answering the wrong question!

$$\underbrace{\Pr(No\ Effect \mid Estimate)}_{\text{What we want to know}} \neq \underbrace{\Pr(Estimate \mid No\ Effect)}_{\text{P-value : What we compute}}$$

The p-value, itself, does not directly tell us the probability that our hypothesis of an effect is true!

It's a flipped conditional, so we need to leverage Bayes' theorem!

Bayesian Reasoning as a Scientific Device

A Scientific Hypothesis:

- The thing we believe to be true or what we want to find evidence for.

Let’s say that our hypothesis is that there is an effect:

- Does going to college have an effect on wealth at 30?
- Does having a teacher’s aid in all K-5 classrooms lead to higher test scores?
- Does surgery on low birth weight newborns increase chances of survival?

We say that there is an effect if we can convincingly show that there is little evidence (or a low posterior probability) for no effect.

Bayesian Reasoning as a Scientific Device

Our null hypothesis is that there is no effect (ATT = 0; β = 0; etc.)

We collect data, analyze the data in an appropriate fashion, compute an estimate of the effect, and compute the **p-value** associated with the estimate:

$$\Pr(Estimate \mid No\ Effect)$$

Remember! Estimate = Estimand + Bias + Noise and we can assume that there's no bias since you took QTM 110 and we talked about appropriate ways to collect and analyze data to eliminate bias!

Bayes' theorem lets us answer the question we really want to answer:

$$\Pr(No\ Effect \mid Estimate) = \frac{\Pr(Estimate \mid No\ Effect)P(No\ Effect)}{\Pr(Estimate \mid No\ Effect)\Pr(No\ Effect) + \Pr(Estimate \mid Effect)P(Effect)}$$

Where:

$$\Pr(No\ Effect) = 1 - \Pr(Effect)$$

Bayesian Reasoning as a Scientific Device

$$\Pr(No\ Effect \mid Estimate) = \frac{\Pr(Estimate \mid No\ Effect) P(No\ Effect)}{\Pr(Estimate \mid No\ Effect)\Pr(No\ Effect) + \Pr(Estimate \mid Effect) P(Effect)}$$

1. $\Pr(Estimate \mid No\ Effect)$ is our significance level – the probability that we would observe what we observed if there was truly no effect.
   - Because statistical hypothesis testing is sharp, this number is our **significance threshold** – the value under which we would say that we have a **statistically significant effect**
   - Most often, this is set to be .05.
   - Our **false positive rate** of our test – the maximum proportion of the time we'd be willing to interpret no effect as an actual effect
2. $\Pr(Estimate \mid Effect)$ is called the **statistical power** of our inferential procedure
   - If the effect is genuine (e.g. there is an effect), how likely is it that we would obtain a statistically significant estimate?
   - We can think of this as our false negative rate – what is the minimum proportion of the time we'd be willing to interpret an actual effect as no effect
   - We can actually estimate this quantity using the data and some assumptions, but beyond the scope of this class.
   - Key point – as N increases, statistical power increases. Why?

Bayesian Reasoning as a Scientific Device

$$\Pr(No\ Effect \mid Estimate) = \frac{\Pr(Estimate \mid No\ Effect) P(No\ Effect)}{\Pr(Estimate \mid No\ Effect) \Pr(No\ Effect) + \Pr(Estimate \mid Effect) P(Effect)}$$

The first two parts can be computed using statistical inference procedures.

But that leaves us with the third part:

$\Pr(Effect)$ – Our prior belief that there is an effect

- Hopefully, this number is relatively high (or relatively low if we're trying to show no effect)
- Otherwise, why are we looking at it?
- If this number is low (e.g. we're throwing darts and hoping for a significant effect), then this leads to bad science

Bayesian Reasoning as a Scientific Device

$$\Pr(No\ Effect \mid Estimate) = \frac{Significance * (1 - Prior)}{Significance * (1 - Prior) + Power * Prior}$$

Let’s set our significance level to .05 – the industry “standard”

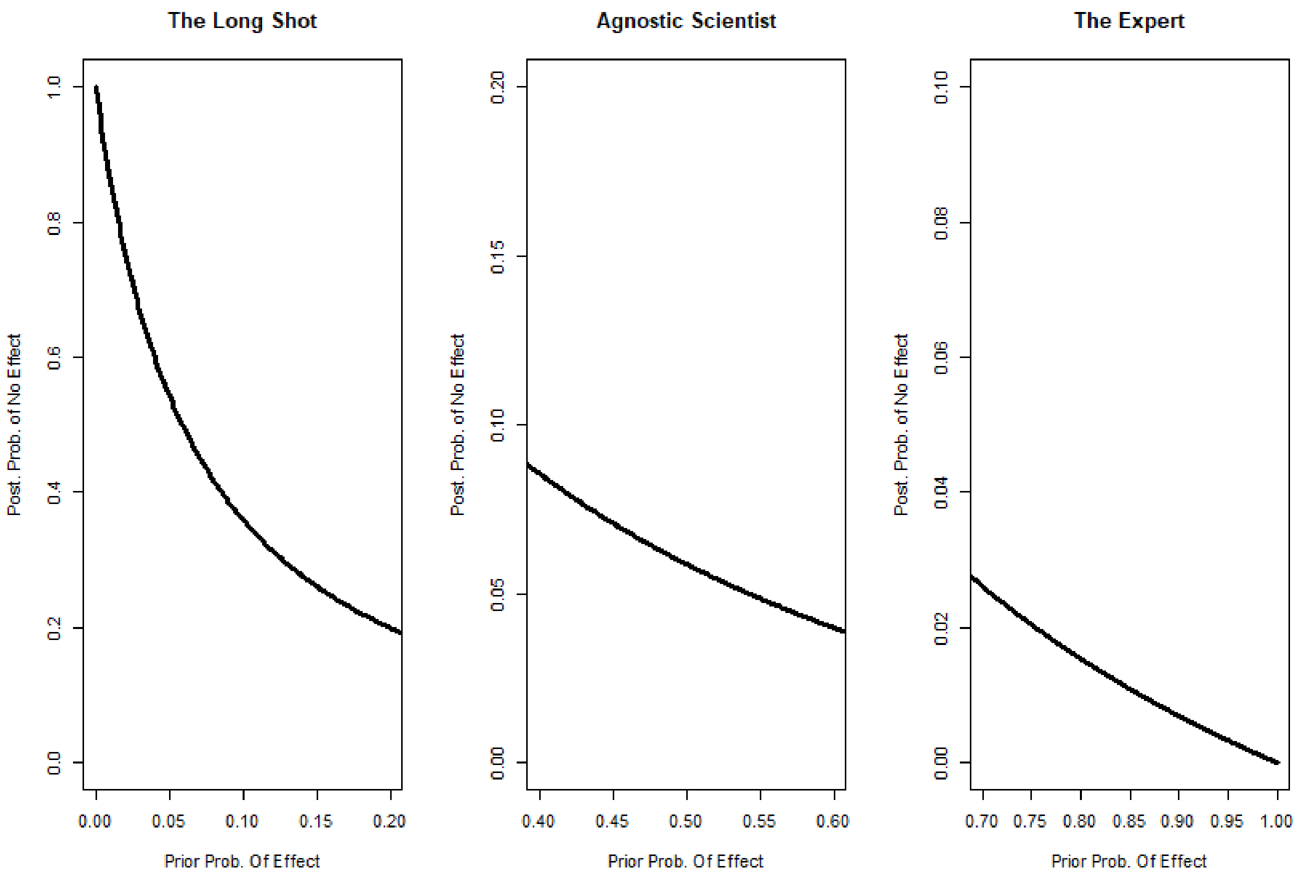
So, we can alter our priors and the power of the test.

- Typically, medical research requires that you are able to sufficiently argue that your statistical power for a test is .8 – the probability that we have an effect and are not able to detect it is less than 20%.
- Larger N means higher power, so let’s assume that we have a large N study and achieve a power of at least .8.

This leaves us with the prior probability of an effect.

Bayesian Reasoning as a Scientific Device

- When we have a strong reason to believe that there is an effect – we have expertise related to the problem – a low p-value points to an effect.
- When we're agnostic – no strong belief one way or the other – it's still pretty low, but not 1-to-1: if we are truly indifferent, the Posterior probability of no effect is still 6%
- But, if we're throwing darts on "long-shot" hypotheses, our p-value tells us very little about whether or not the effect is real – Pr(No Effect | Evidence) is still really high!

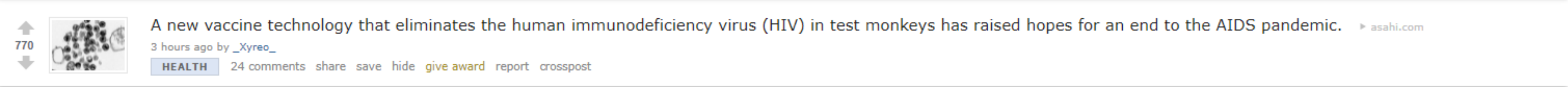


Bayesian Reasoning as a Scientific Device

If we are pursuing a long-shot hypothesis, then a “significant” effect does not mean that we have evidence that our hypothesis is true!

How common is this?

- “Unicorn Drugs” – we have strong theory that certain things should work, but prior evidence has shown that there are no known miracle drugs (low prior that the new one will lead to an effect)

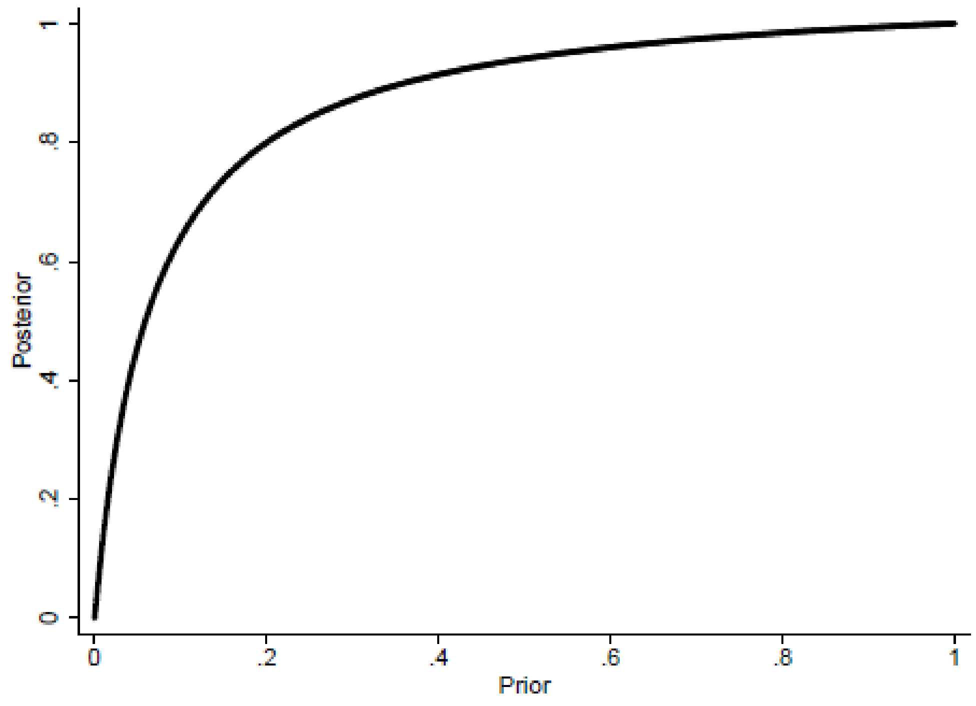


- Too good to be true findings – that sure is cool, but why would that ever happen?

- And many other pieces of “junk science” that we’ll discuss a little later in the semester

The Value of Skeptical Priors

- In 2010, Cornell psychologist Daryl Bem published a paper of an experiment he conducted claiming to have found evidence of extrasensory perception (ESP)
  - College subjects were able to pick which of two curtains an object was behind statistically significantly better than chance would suggest ($p<0.05$), with a (standard) power of 80%
  - How much you believe this should depend *heavily* on your **prior** – which is likely near 0.0!
    - **Priors** can be valuable for cutting through the BS, even in peer-reviewed journals!



The Replication Crisis

If our criteria for accepting a scientific finding as true is that $p < .05$, then we are bound to accept some things as truth when they are, in fact, not. More than 5% of findings will be due to chance!

Without good reason to believe that something is true, the p-value means very little for the efficacy of our hypothesis!

Problem: A large portion of findings cannot be **replicated**

- When separate researchers attempt to replicate findings from a published article, they are often unable to reproduce the finding in a different setting

A scientific world which rewards p-values less than .05 without considering whether or not the finding makes sense will make lots of incorrect conclusions.

We haven't even touched on the problem of **power** – the vast majority of studies in the social sciences and small N medical studies are **underpowered**, meaning that the evidence required to overcome a somewhat skeptical prior is even higher!

Theory and Expertise

We want to be good scientists and actually assess the probability that our hypotheses are true in the face of data:

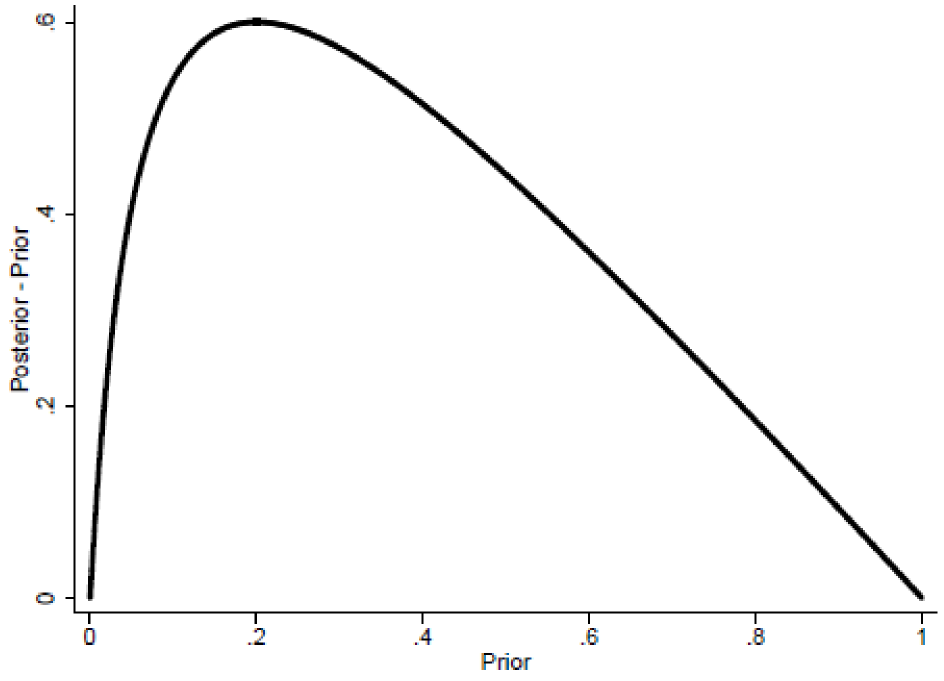
- That requires **priors** to translate what we can compute to what we want to know

How do we **credibly** say that our prior probability of an effect is high?

- Theory – the literature review portion of a paper fixes your question within existing research. These pieces are intended to **raise** the credibility of your question and the credibility that your hypothesis is true!
- Expertise – If a lot of people with a lot of subject specific knowledge believe that something is true, then it's probably true! Wisdom of crowds.
- Prior research – there may be other research that partially addresses your question. Combine those results to create a meaningful prior.

Different People, Different Priors

- It's also fine (to a point/within reason) for different people to react differently to the same evidence. The same piece of evidence (say, about global warming) could affect different people in different ways:
  - If you really don't believe in global warming, it won't change your mind
  - If you really *do* believe in global warming, it won't change your mind
  - If you're closer to the middle, you may shift your beliefs dramatically
- All these *responses* are "rational" under Bayes' theorem – though we can always fight over whether the priors are!



An existential crisis…

So, what are we doing then?

- Why do science at all if we need to have a good strong prior belief that the thing we're looking at has an effect?
- What about finding fundamental truths of the universe that no one else has discovered?
- There's always going to be someone out there who can reasonably say that our **prior** is wrong.

As data analysts, aren't we supposed to let the data speak for itself without imposing our own prejudices?

This is a tough one. Personally, I am not too offended by the notion of priors.

- Unfortunately, there's no way around priors if we really want to assess the probability that our hypothesis is true!

But, I think we can all agree that p-values alone are not sufficient for stating the probability that a finding is true!

An existential crisis…

A more comforting thought:

- Let our priors determine **what** questions we should ask. Only ask questions where we have a plausible belief that there is an effect.
  - Seriously, farts and cancer? Who thought that was a good idea?
- Once we have decided to seek an important answer, analyze the data in a principled and fair way
- Approach the problem with the same rigor that we outlined in the first 10 weeks of the class
  - Is the chosen estimator unbiased for the quantity in question?
  - Is our data good enough to say that our estimate is both internally and externally valid?
  - Can we address this question in a way that minimizes harm?

An existential crisis…

A more comforting thought:

- Let our priors determine **what** questions we should ask. Only ask questions where we have a plausible belief that there is an effect.
  - Seriously, farts and cancer? Who thought that was a good idea?
- Once we have decided to seek an important answer, analyze the data in a principled and fair way
- Approach the problem with the same rigor that we outlined in the first 10 weeks of the class
  - Is the chosen estimator unbiased for the quantity in question?
  - Is our data good enough to say that our estimate is both internally and externally valid?
  - Can we address this question in a way that minimizes harm?

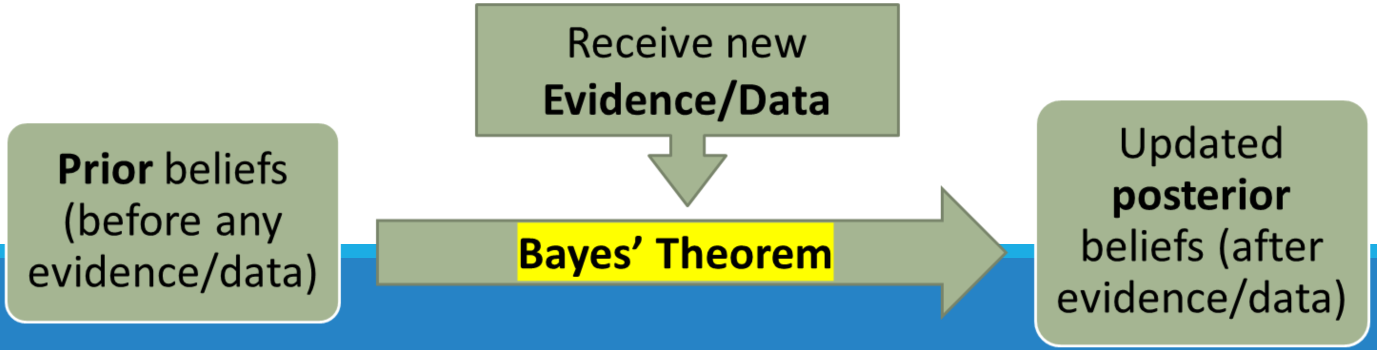
Good analysts don't ignore their priors.

Instead, they are as transparent as possible about the extent to which their conclusions come from the data vs. the priors.

Summary

- Conditional probability: Probability of event A given event B = $\mathrm{P(A|B)} = \frac{P(A\ and\ B)}{P(B)}$
- $\mathrm{P(A|B)} \neq P(B|A)$, but…
- Bayes' theorem allows us to flip conditional probabilities: $P(A|B) = \frac{P(B\mid A)*P(A)}{P(B)}$
  - A more general application of this is to get the probability of some hypothesis/claim given some evidence
  - Bayesian inference allows us to evaluate and update beliefs in a principled, quantitative manner based on evidence (which is, implicitly, how humans reason – though we're bad at it):

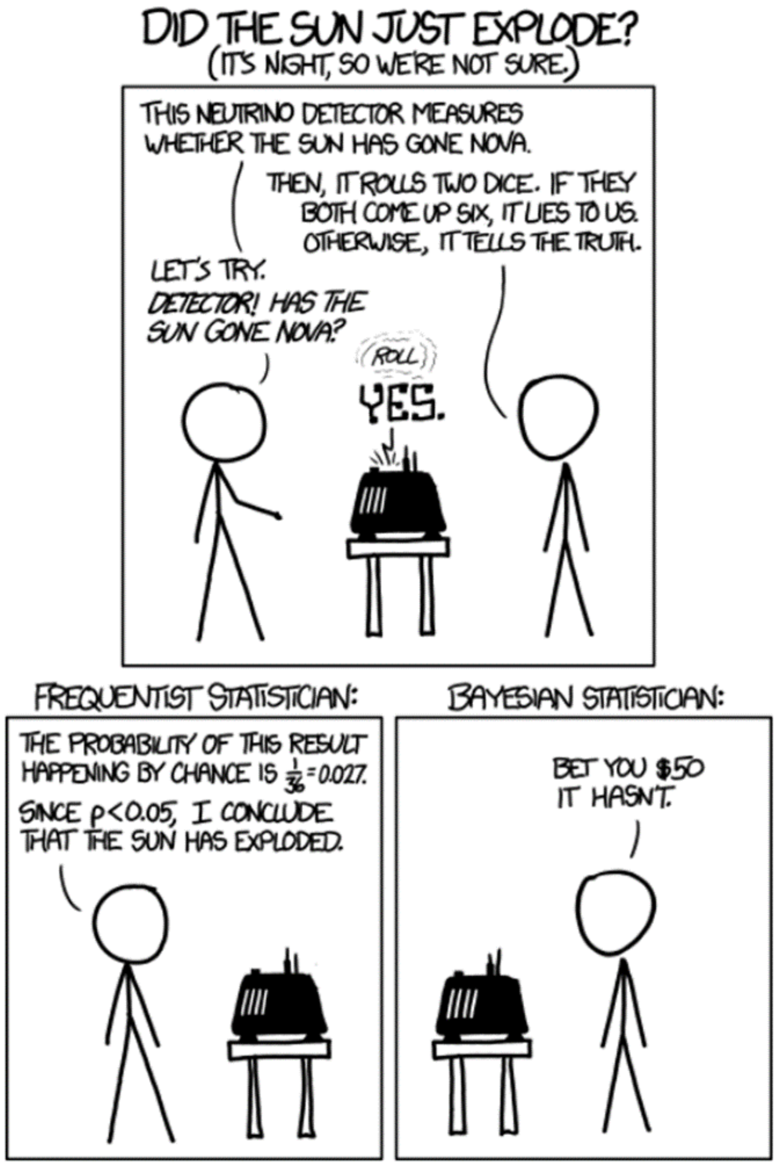
$$P(Hypothesis|Evidence) = \frac{P(Evidence\mid Hypothesis) * P(Hypothesis)}{P(Evidence)}$$



Summary

- Key message:

  - Explicitly state and remember your question
    - With questions of probability, goal is very often some **posterior** belief in something given some data/evidence
  - Make sure you're answering *that question* and not something different
  - P-values provide the opposite conditional of what we actually want to know.
  - Don't ignore the prior/base rate when calculating the posterior probability of something


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
ROLL
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

How Bayes’ Theorem Works – Non-Binary Hypothesis

- Nothing stops you from expanding **Bayes’ theorem** to 3 possible hypotheses

  - Example: is a roulette wheel fair, biased towards red, or biased towards black?
    - You set a **prior** for each possibility; they all must sum to 1
      - In our previous example P(Psychic) + P(Not Psychic) = 1, so this isn’t new
    - The marginal $P(e)$ becomes trickier to calculate because you have to calculate and sum the probability of the evidence (e.g. 5 straight red spins) under all *three* scenarios, rather than just two

Source: https://customtablefelt.com/shop/professional-roulette-wheel/

How Bayes' Theorem Works – Non-Binary Hypothesis

- Nothing stops you from expanding **Bayes' theorem** to 3 possible hypotheses
  - Example: is a roulette wheel fair (F), biased towards red (R), or biased towards black (B)?
    - $P(R|5\,reds) =$

$$\frac{P(5\,reds|R) * P(R)}{P(5\,reds)} =$$

$$\frac{P(5\,reds|R) * P(R)}{P(5\,reds|R) * P(R) + P(5\,reds|B) * P(B) + P(5\,reds|F) * P(F)}$$

- NOTE: How likely you think the wheel is to be biased after 5 reds depends, in part, on *how skeptical of the wheel you were to begin with* (i.e. your **prior,** $P(R)$)!

Source: https://customtablefelt.com/shop/professional-roulette-wheel/

How Bayes' Theorem Works – Non-Binary Hypothesis

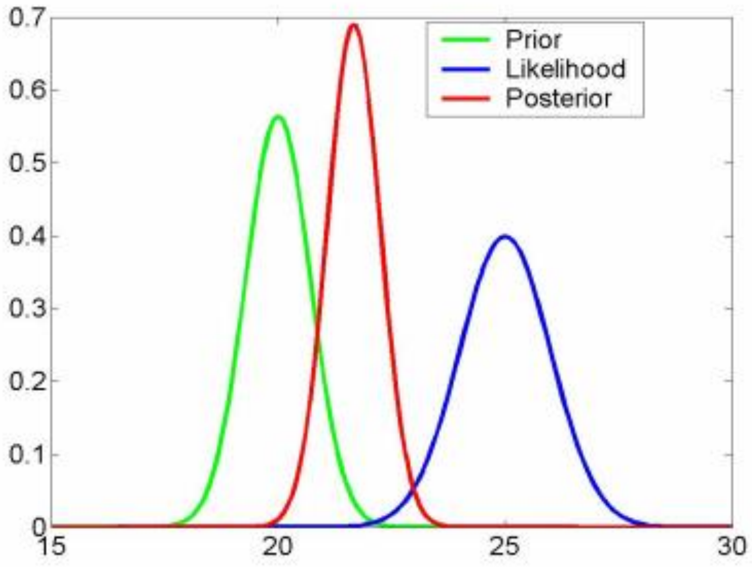
- Nothing stops you from expanding **Bayes' theorem** to an infinite number of hypotheses

  - This comes up when you're trying to estimate the size of an effect in a study – such as "how much does this new policy increase annual GDP?"
    - H1: -$1 billion
    - H2: -$1.00005 billion
    - H3: +$1,000
    - H4: +$1,001
    - H5: +$537,000
    - …

  - You want to know how likely *all* these possible effect sizes are – some will be very likely, some less so, some nearly impossible

How Bayes' Theorem Works – Non-Binary Hypothesis

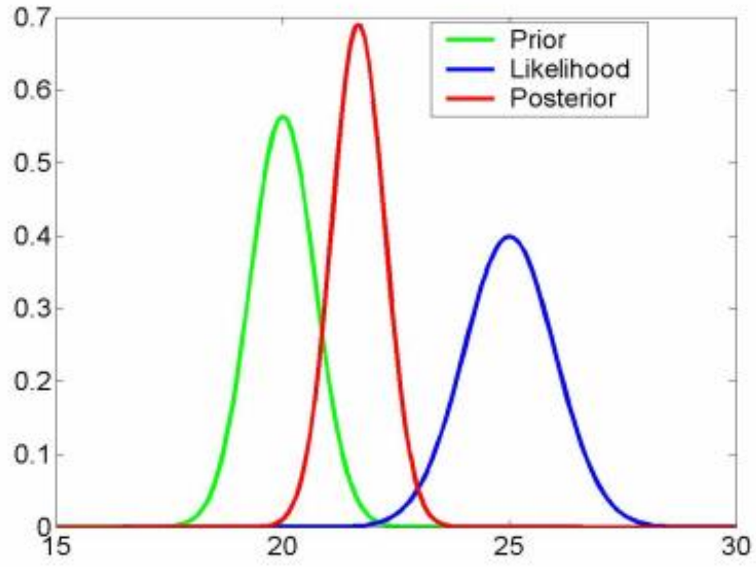
- Nothing stops you from expanding **Bayes' theorem** to an infinite number of hypotheses
  - This comes up when you're trying to estimate the size of an effect in a study – such as "how much does this new policy increase annual GDP?"
    - You start with a **prior *distribution*** for how likely you think each value is – some will be quite likely, some extremely unlikely
      - Probabilities for all values must sum (actually, integrate – area under the curve) to 1.0
    - You also get a **likelihood** from the data that tells you the probability of seeing the data if each possible value of the effect were true
      - Some values will be supported (high $P(e|H)$), some not
    - You combine these using **Bayes' theorem** to get your **posterior distribution** of probabilities for possible values of the effect – pulling **prior** towards **likelihood**



Source: https://www.fil.ion.ucl.ac.uk/spm/course/slides10-vancouver/08_Bayes.pdf

How Bayes' Theorem Works – Non-Binary Hypothesis

- Nothing stops you from expanding **Bayes' theorem** to an infinite number of hypotheses
  - How do you combine the prior and likelihood using **Bayes' theorem**?
    - Essentially the same way, except you have to use integral calculus to get, for example, P(e)
      - And the likelihood and prior aren't single probabilities/numbers, but probability *distributions* that can themselves be multiplied together
    - More details are way beyond the scope of this course, but it's fundamentally the same process – combining prior probability estimates for each possible value with data to create updated posterior probabilities of each possible value



Source: https://www.fil.ion.ucl.ac.uk/spm/course/slides10-vancouver/08_Bayes.pdf