Durham University

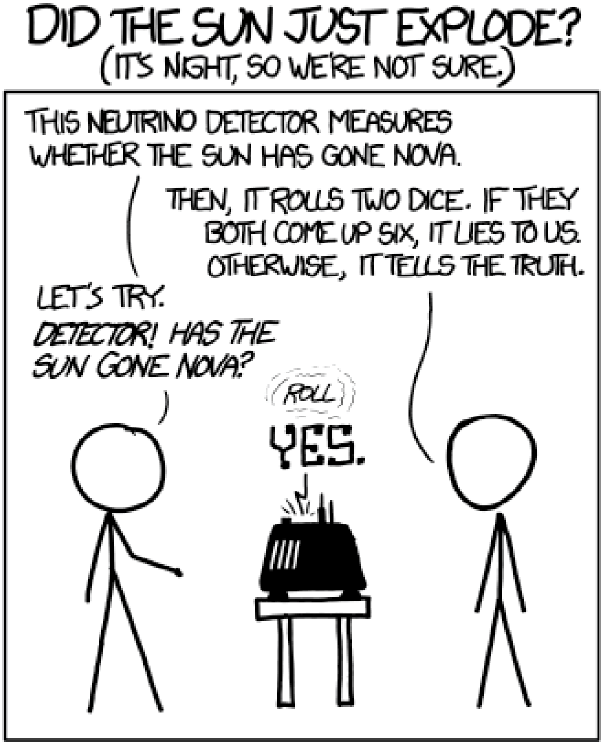
Introduction to Bayesian Statistics

Camila Caiado
Associate Professor in Statistics

Warm-up

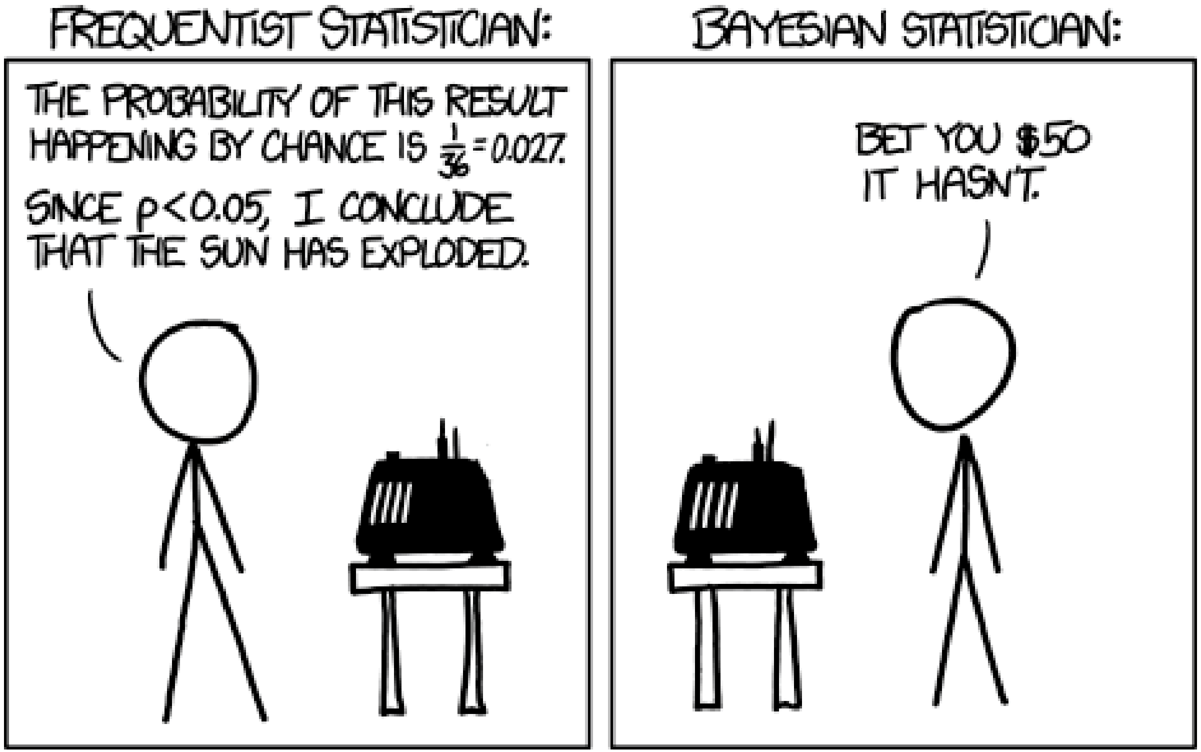
- Think of something you believe to be true.
- You have 1000kr to bet on this truth. If you are right, you receive the full amount you have bet, otherwise you receive nothing. How much are you willing to bet?
- If you didn’t bet the full 1000kr, why?

Bayesians vs Frequentists (XKCD)



Durham University

Bayesians vs Frequentists (XKCD)



Durham University

Bayesians vs Frequentists

"Using a spade for some jobs and shovel for others does not require you to sign up to a lifetime of using only Spadian or Shovelist philosophy, or to believing that only spades or only shovels represent the One True Path to garden neatness." – Ken Rice

- A good statistician should behave like a Bayesian, always learning, but only use it when it makes sense.
- There is more than just Bayesian and Frequentist approaches.

Bayes' Theorem

Conditional probability: the probability of A given B, for $P(B) \neq 0$,

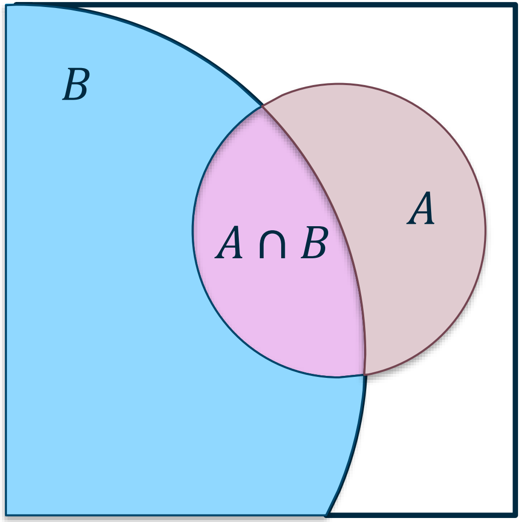
$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Similarly, for $P(A) \neq 0$,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$



Durham University

Medical diagnosis example

Suppose 0.1% of the population have a certain disease. A diagnostic test is carried out: the outcome of the test is either positive or negative. It is known that the test is 90% reliable, that is, a person with the disease will correctly test positive with probability 0.9. A person without the disease will incorrectly test positive with probability 0.1.

A person is tested for the disease. What is the probability that they test positive? What is the probability that they have the disease if they have tested positive?

Medical diagnosis example

D: person has the disease

$$P(D) = 0.001$$
$$P(D^-) = 0.999$$

T: person tests positive

$$P(T|D) = 0.9 \Rightarrow P(T|D)P(D) = P(T \cap D) = 0.0009$$
$$P(T|D^-) = 0.1 \Rightarrow P(T|D^-)P(D^-) = P(T \cap D^-) = 0.0999$$

So, the probability of testing positive is

$$P(T) = P(T \cap D) + P(T \cap D^-) = 0.1008.$$

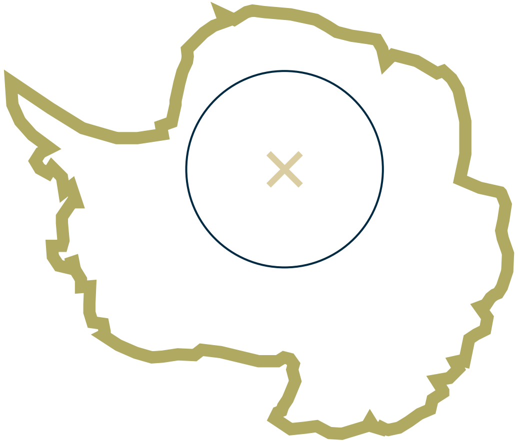
And the probability of having the disease given an positive test:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{0.009}{0.1008} \approx 0.0893.$$

Discussion: What if we test this person again and we observe another positive result?

Frequentist Treasure Hunt

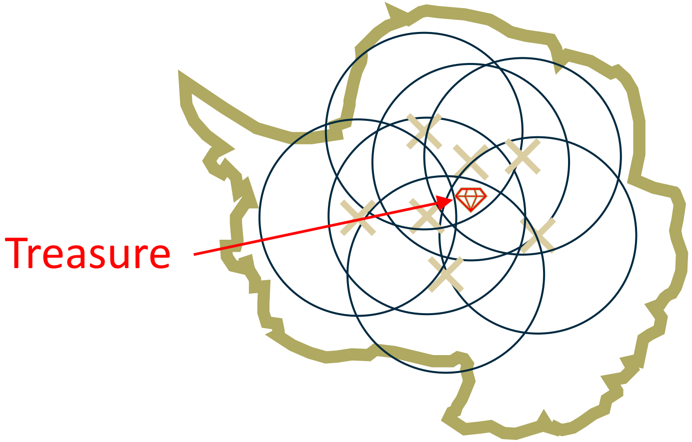
A pirate has hidden a treasure in an island. You have a map but it isn't the most accurate so you and your mates decide to dig holes until you find it. After digging each hole, you draw a 1 metre circle around the hole, you have 95% confidence that the circle contains the treasure.



Frequentist Treasure Hunt

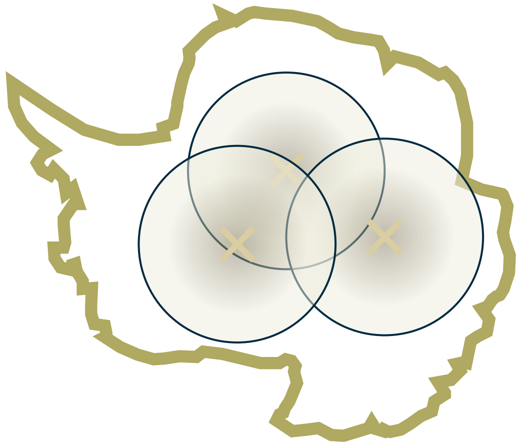
Repeat many times, your circles would contain the treasure 95% of the time.

Say you choose one more location to dig. How do you decide if that's a bad choice? What if the pirate was with you and shouted if you are close or not each time?



Bayesian Treasure Hunt

Repeat as many times as you can afford. Sequentially update your prior based on your previous observations. Use Bayes' theorem to update your beliefs about the treasure location.



If the pirate is with you, you might want to include in your model whether you believe what they are saying or not!

Bayesian Inference

Prior distribution: this distribution describes what you know about the parameter you want to estimate but excludes information in the data.

Likelihood: includes modelling assumptions and asserts how likely the observed data are given your parameter.

Posterior distribution: update what we know about the parameter by combining the prior and the data.

$$\textbf{Posterior} \propto \textbf{Likelihood} \times \textbf{Prior}$$

Update your model as more data becomes available and you will eventually be able to bet on the location of the treasure.

How do you choose a prior?

- The prior should reflect your current knowledge of the problem. If you don’t know what you are doing, call the experts. That is called ***elicitation***. If you are the expert and you are unsure, the prior should reflect your uncertainty.
- The prior contains information external to the current study. If you have a pilot study, that can be used to inform your prior. If you don’t have pilot data, can you talk to experts in the field and try to quantify their knowledge and uncertainty? Design of experiments exist in the Bayesian world too!

Types of Priors

Subjective: this prior expresses the expert's personal uncertainty. E.g. using my personal experience, I think there is a 60% chance of rain today.

Objective and informative: information or data might exist to help formulate a prior. Historical data and past experiments can help build this type of prior. E.g. Yesterday rained and I have seen the weather forecast indicating rain, I think now that there is a 80% chance of rain today.

Today's posterior is tomorrow's prior."

Noninformative: this type of prior expresses ignorance. I have just arrived in this country and have no idea what the weather is like so the chance of rain must be 50-50. There is a lot of debate here around *improper* priors and how informative an noninformative prior can be.

Garbage in, garbage out

- Like with any statistical tool, a Bayesian model is just as good as your assumptions and your data.
- If your data is very informative, your likelihood will overshadow your prior and whether you have succeeded in your elicitation is irrelevant.
- Being uncertain doesn't imply ignorance. You (or your expert) might actually think that extreme values or cases are just as likely to start with. E.g. for Covid it is unknown what the rate of asymptomatic infections is so it isn't invalid to think that the rate can be between 50% and 80%.

Durham University

If you like a good argument

- Bayesian epistemology and Bayesian dogmatism
  - Bayesianism assumes rationality but humans aren’t always rational.
- Subjective vs Objective Bayes
- Noninformative priors are informative.
- What do you do if your experts think they know more than they actually do?

References and resources

- Gelman et. al, *Bayesian Data Analysis*
- Carlin and Louis, *Bayesian Methods for Data Analysis*
- Wakefield, *Bayesian and Frequentist Regression Methods*
- De Finetti, *Theory of Probability: A Critical Introductory Treatment*
- De Finetti, *Probability, induction and statistics: The art of guessing*
- Open University: https://www.open.edu/openlearn/science-maths-technology/bayesian-statistics/content-section-0?active-tab=description-tab