Statistics in the early days of the Natural Sciences

Elena Sellentin

Sterrewacht
Universiteit Leiden

The actors[1]

- Pierre-Simon Laplace (1749-1827)
- Thomas Bayes (1701-1761)
- Many people who needed money.
- Many people who wanted to be healthy.
- Samuel Hahnemann (1755-1843)

Feel of the time: Marie Antoinette (1755-1793) was alive, bloodletting was a thing.

[1] We apologize to all humans whose story has not been handed down correctly throughout the years.

Ergo: There was no such thing as approved clinical trials.

People had difficulties learning from data:

- Data $\boldsymbol{x} \sim \mathcal{D}(\boldsymbol{x})$
- Hypothesis or theory H
- $\mathcal{L}(\boldsymbol{x}|H)$ ('Claiming the hypothesis is true, this is how often it generates data $\boldsymbol{x}$')
- $\mathcal{P}(H|\boldsymbol{x})$ ('Look, these are my data $\boldsymbol{x}$, how likely is my hypothesis?')

Bayes' Theorem

$$\mathcal{P}(H|\boldsymbol{x}) = \frac{\mathcal{L}(\boldsymbol{x}|H)\pi(H)}{\pi(\boldsymbol{x})} \tag{1}$$

Today: $\pi(H)$ m-a-t-t-e-r-s. (But let's forget that; we're in 1790.)

So we come up with a hypothesis... namely...

H = 'Arsenic[2] is good for your health.'

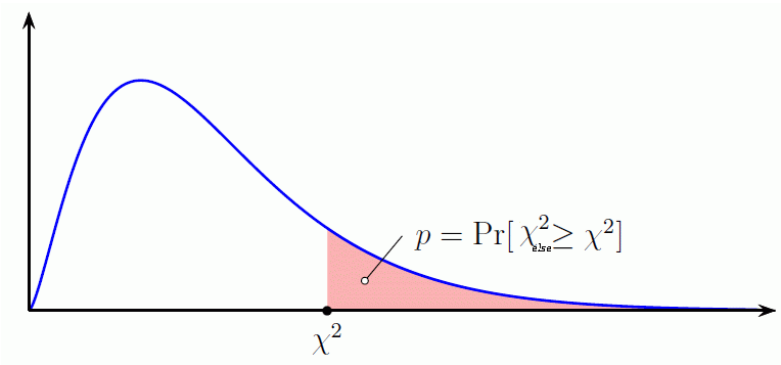
No, it isn't! Arsenic is a deadly poison!
Don't eat it at all!!

[2] Dutch: Arseen. It's a terrible poison!

But we're in 1790...

... so someone conducted a 'study'.

- Person 1: dies.
  Analysis from back then: $\pi(H)$ and $\pi(x)$ don't matter, let's put them to 1. Then $\mathcal{P}(H|\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}|H)$. Since H probably right, x must be from tail.
  Conclusion: well, this happens.
- Person 2: dies.
  Conclusion: well, as above.
- Person 3.... dies.
- Person X.... dies.



Correct conclusion would have been: Arsenic is a poison and kills people.

Ergo: Be a good human and remember Bayes' theorem

$$\mathcal{P}(H|\boldsymbol{x}) = \frac{\mathcal{L}(\boldsymbol{x}|H)\pi(H)}{\pi(\boldsymbol{x})}. \tag{2}$$

And yes, the exam is definitely going to include a question on Bayes' theorem.