

12.2 The Base-Rate Fallacy

WORKSHEET

The question

The diagnostic test AL has a probability 0.9 of giving a positive result when applied to a person suffering from the rare disease mathematitis. It also has a probability $1/11$ of giving a false positive result when applied to a non-sufferer. It is known that only 1% of the population suffer from the disease. Given that the test AL is positive when applied to Frankie, who is chosen at random from the population, what is the probability that Frankie is a sufferer?

In an attempt to identify sufferers more accurately, a second diagnostic test STEP is given to those for whom the test AL gave a positive result. The probability of STEP giving a positive result on a sufferer is 0.9, and the probability that it gives a false positive result on a non-sufferer is p . Half of those for whom AL was positive and on whom STEP then also gives a positive result are sufferers. Find p .

Hints

Try the question on your own first. If you get stuck, read the hints *one at a time*, and go back to the question after each one. Hints marked *Pause and try* are the key steps.



Ask yourself Imagine 1000 people. How many have the disease? How many of them test positive? How many of the healthy people test positive?

STAGE 1 Bayes' theorem



Ask yourself Write D for “Frankie has the disease” and $+$ for “AL is positive”. You know $P(+ | D)$ and want $P(D | +)$. How are they related?

STAGE 2 A second test



Ask yourself The second test is given only to those who tested positive on AL. What is the proportion of sufferers in that group? Use it as the new “base rate”.



Pause and try With $p = 0.09$, STEP’s false-positive rate is about the same as AL’s, yet a second positive result raises the probability from $\frac{1}{11}$ to $\frac{1}{2}$. Why does the same quality of test have a much bigger effect the second time?

Try this next

- 1 A body in the fridge.** (From STEP II, 1987.) Two friends, who tell the truth independently with probabilities p and q , both tell me there is a body in my fridge. If I had thought this equally likely as not, explain why my new estimate should be $\frac{pq}{1-p-q+2pq}$. There is indeed a body. It could be my enemy A , my enemy B , or both, which I judged equally likely; both friends now say A is in the fridge. What is my new estimate that both are there?
- 2 How rare?** For a test with sensitivity 0.9 and false-positive rate $\frac{1}{11}$, how common must a disease be for a positive result to mean a better than even chance of having it?
- 3 Repeat the test.** If AL itself is repeated, and the results are independent given the person’s true state, what is the probability of disease after two positive AL results? After three?
- 4 Screening.** A screening programme tests a million people for a condition with prevalence 1 in 10 000, using a test with sensitivity 99% and false-positive rate 1%. How many people will be told they may have the condition, and how many of

them actually do?

Further reading

- ▶ **Gerd Gigerenzer, *Reckoning with Risk*** (Penguin). Natural frequencies, medical tests and the courtroom.
- ▶ **David Spiegelhalter, *The Art of Statistics*** (Pelican). Includes medical screening and the Sally Clark case.
- ▶ **Sharon Bertsch McGrayne, *The Theory That Would Not Die*** (Yale University Press). The history of Bayes' theorem, including Turing's use of it.

The full guided solution, and what the question leads to at university, is in Section 12.2:
https://keithmcnulty.github.io/step_book_mcnulty/ch12/base-rate.html