# On Bayesian Tests

Johannes Siedersleben
Oxford, July 2024

# Recommended Books

## Daniel Kahneman: Thinking, Fast and Slow (2011)

Easy reading, you learn a lot about how humans think. Many examples of Bayesian tests with surprising results.
Kahnemann received the Nobel Prize in 2002.

## Judea Pearl: The Book of Why (2019)

This book is a bit more challenging and explains the theory of causation in some depth, albeit informally. The examples given here are taken from it.
Pearl received the Turing Award in 2011. He invented, among other things, Bayesian Networks.

# Bayes who?

**Thomas Bayes** (1701 – 1761) was an English statistician, philosopher, and Presbyterian minister who is known for formulating a specific case of the theorem that bears his name: Bayes' theorem. Bayes never published what would become his most famous accomplishment; his notes were edited and published posthumously by Richard Price.
One of his major contributions is conditional probabilities.
(From Wikipedia)

# Conditional what?

**P[cancer]** **// the (unconditional) probability of cancer**

That's two numbers adding up to 1

| |
|---|
| P[cancer-no] = 0.9986 |
| P[cancer-yes] = 0.0014 |

Don’t look at the numbers (yet).

**P[cancer|test]** **// the conditional probability of cancer, given a test result (not symmetric!)**

That's four numbers with column sums equal to 1

| | |
|---|---|
| P[cancer-no|test negative] = 0.9995 | P[cancer-no|test positive] = 0.9917 |
| P[cancer-yes|test negative] = 0.0005 | P[cancer-yes|test positive] = 0.0083 |

With conditional probabilities under our belts, let’s move on to Bayesian tests.

# Bayesian Tests

**Given**: a disease (e.g. cancer) and a test. You know somehow

**prior probability = P[cancer]** // proportion of population affected

**Before the roll-out**: You test people with a known diagnosis (thousands, but not a million). This gives you:

likelihood = P[test|cancer] // probability of test positive or negative given the diagnosis (four numbers)

likelihood ratio = likelihood / P[test] (four numbers)

**posterior probability = P[cancer|test] = prior probability * likelihood ratio** (four numbers)

**In the field**: Millions of people with an unknown diagnosis are tested.

**Interpretation**:

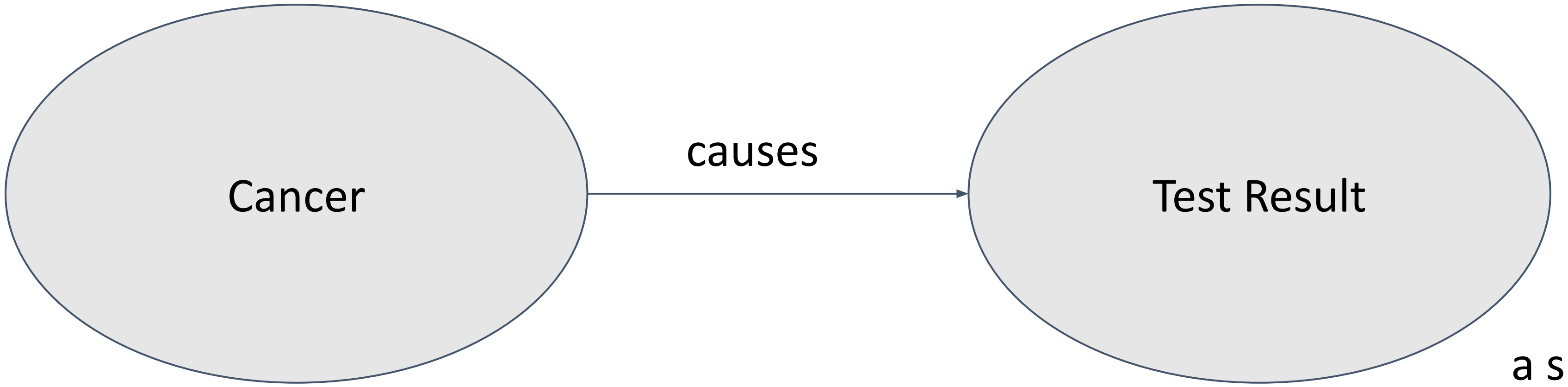
Before the test, the probability of being contracted is the prior probability.
With the information gained from the test, the probability of being contracted is the posterior probability.
The likelihood ratio is a measure of the information gained.

Very nice, but this talk is supposed to be about causation!

# Here comes causation

We have

- a biochemical **theory** of how cancer causes the test result
- **data** which confirms (or possibly contradicts) the theory
- a **method** (Bayesian test) that allows us to quantify the impact.




a small Bayesian network

We are interested in

- **posterior probability = P[cancer|test]**

You'll be surprised!

# Test positive – should I be worried?
# Let's look at the numbers (from J. Pearl, not made up)!

**prior**

| | |
|---|---|
| P[cancer-no] = 0.9986 | |
| P[cancer-yes] = 0.0014 | |

**likelihood**

| | |
|---|---|
| P[test negative\|cancer-no] =0.8800 | P[test positive\|cancer-no] = 0.1200 |
| P[test negative\|cancer-yes] = 0.2857 | P[test positive\|cancer-yes] = 0.7143 |

**posterior**

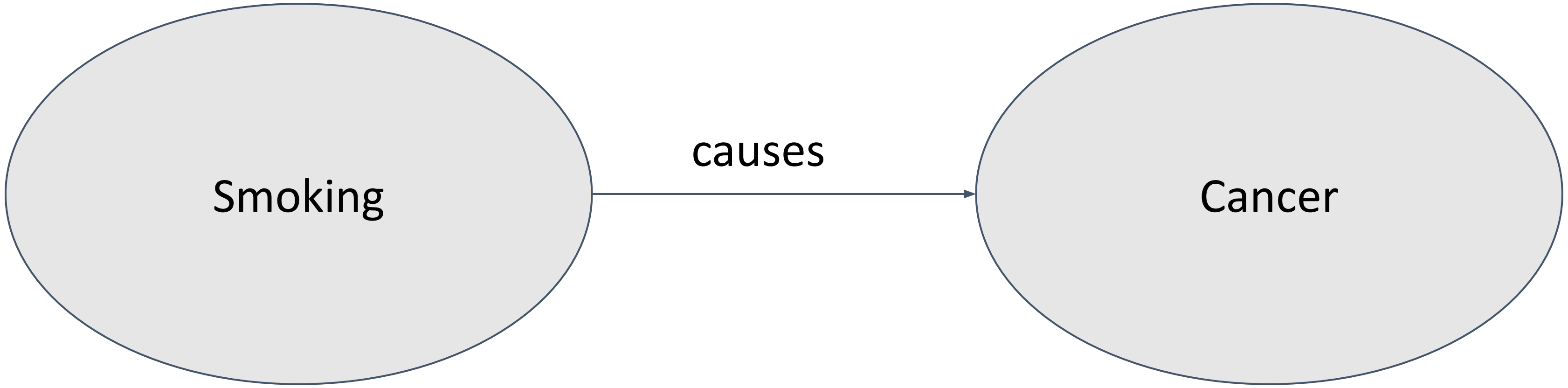
| | |
|---|---|
| P[cancer-no\|test negative] = 0.9986 * 1.00095<br>= 0.9995 | P[cancer-no\|test positive] = 0.9986 * 0.9931<br>= 0.9917 |
| P[cancer-yes\|test negative] = 0.0014 * 0.324971<br>= **0.0005** | P[cancer-yes\|test positive] = 0.0014 * 5.91296<br>= **0.0083** |

Ready for another example?

# Does smoking cause lung cancer?

We have

- a biochemical **theory** of how smoking causes cancer
- **data** which confirms (or possibly contradicts) the theory
- a **method** (Bayesian test) that allows us to quantify the impact.




We are interested in

- **posterior probability = P[cancer|smoking]**



You'll be surprised!

# Let's look at the numbers (they are made up)

**prior**

| | |
|---|---|
| P[cancer-no] = 0.99 | |
| P[cancer-yes] = 0.01 | |

**likelihood**

| | |
|---|---|
| P[smoker-no\|cancer-no] = 0.8081 | P[smoker-yes\|cancer-no] = 0.1919 |
| **P[smoker-no\|cancer-yes] = 0** | **P[smoker-yes\|cancer-yes] = 1** |

**posterior**

| | |
|---|---|
| P[cancer-no\|smoker-no] = 0.99 * 1.0101 = 1 | P[cancer-no\|smoker-yes] = 0.99 * 0.9596 = 0.95 |
| **P[cancer-yes\|smoker-no] = 0.01 * 0 = 0** | **P[cancer-yes\|smoker-yes] = 0.01 * 5 = 0.05** |

Let’s call it a day!

# What can you take home?

- Thomas Bayes, the clergyman
- Conditional probabilities
- The Bayesian approach (swapping what comes before and after the stroke), its versatility and how it's related to causation.
- Surprising numbers
- Two books to read