Conditional Probability:

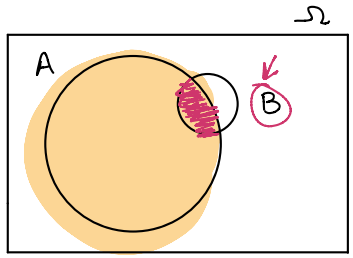
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

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

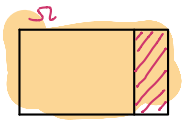
Interpretation

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

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

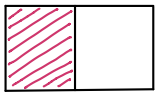


$P(B|A)$



≈ 0.2

$P(A|B)$



≈ 0.5

Problem:

- A disease is prevalent in 0.2% of a population.
- We have a test that, given to a sick person, gives a +ve result 85% of the time.
- Of all the people ever tested, 8% were positive.

Q: If Nazo is tested and test comes back positive, what are the chances that she actually has the disease?

☐ 85% ☐ 77% ☐ 21% ☑ 2%

(*) Event "Disease" is difficult to measure directly!

Event "Pos" is relatively easier to measure

$P(\text{Disease}) = 0.002$ (0.2%)

$P(\text{Pos} \mid \text{Disease}) = 0.85$

$P(\text{Pos}) = 0.08$

$P(\text{Disease} \mid \text{Pos}) = ?$

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

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

$P(A \cap B) = P(A|B) * P(B)$

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

$$P(\underset{B}{\text{Disease}} \mid \underset{A}{\text{Pos}}) = \frac{P(\underset{A}{\text{Pos}} \mid \underset{B}{\text{Disease}})\,P(\underset{B}{\text{Disease}})}{P(\underset{A}{\text{Pos}})}$$

$$= \frac{(0.85)(0.002)}{0.08}$$

$$= 0.021$$

2.1%

Posterior — likelihood — prior — before

$$P(\text{Disease} \mid \text{Pos}) = \frac{P(\text{Pos} \mid \text{Disease})\ P(\text{Disease})}{P(\text{Pos})}$$

after — normalizing factor

$P(\text{Disease}) = 0.002$ — Prior belief

$P(\text{Pos} \mid \text{Disease}) = 0.85$ — Result of experiment

$P(\text{Pos}) = 0.08$

$P(\text{Disease} \mid \text{Pos}) = 0.21$ — Updated belief

⊛ You start off with some belief and update it based on some experiment!

This is the "Bayes' Rule" of inference.

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

Classical Statistics →

→ Bayesian " →

Applying Bayes' Rule to Spam Detection.

- "You have inherited a million dollars." (Spam)
- "There will be a meeting at noon." (not-spam)
- Assumption: We have a dataset of spam emails. (UCI dataset)
- Need to find whether a piece of text is spam.

- Let's first consider a single word.

$$P(\text{Spam} \mid w) = \frac{P(w \mid \text{spam})\, P(\text{spam})}{P(w)}$$

- "Given that this word appears, how likely is it that the message is spam?"
- "how frequently does this word appear in spam?" — from dataset.
- "How much spam is there in the world?"
- "How frequent is this word?"

$$P(\text{spam}) = \frac{\#\text{ of spam messages}}{\#\text{ of all messages}} \quad \frac{100}{500}$$

$$P(w \mid \text{spam}) = \frac{\#\text{ of times this word appears in spam}}{\#\text{ of spam messages}}$$

$$P(w) = \frac{\#\text{ of times this word appears}}{\#\text{ of total messages}}$$

- Now, do this for all words
- $P(\text{spam} \mid \text{words}) = P(\text{spam} \mid w_1) * P(\text{spam} \mid w_2) * \dots P(\text{spam} \mid w_n)$

"independence"

$$P(\text{spam} \mid \text{words}) = \prod_{i=1}^{|\text{words}|} P(\text{spam} \mid w_i)$$

→ words — independent

"Naive Bayes Model"

$\prod_{i=1}^{n}$ $\sum_{i=1}^{n}$