Bayes Theorem
Bayes’ theorem was presented in "An Essay towards solving a Problem in the Doctrine of Chances" which was read to the Royal Society in 1763 already after Bayes' death.
Bayes’ theorem (simple)

\[ P(A \cap B) = P(A|B)P(B) = P(B \cap A) = P(B|A)P(A) \]

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

• In Science we often want to know:
  “How much faith should I put into hypothesis, given the data?”
  or \( P(H|D) \) (see also the inductive definition of probability)
  
• What we usually can calculate if the hypothesis/model is OK:
  “Assuming that this hypothesis is true, what is the probability of the observed data?” or \( P(D|H) \)

• Bayes’ theorem can help: \( P(H|D) = P(D|H) \cdot P(H)/P(D) \)

• The problem is \( P(H) \) (so-called prior) is often not known
Bayes’ theorem (continued)

Works best with exhaustive and mutually-exclusive hypotheses: $H_1, H_2, \ldots, H_n$ such that $H_1 \cup H_2 \cup H_3 \ldots \cup H_n = S$ and $H_i \cap H_j = \emptyset$ for $i \neq j$

$$P(H_k | D) = P(D | H_k) \cdot P(H_k) / P(D)$$

where:

$$P(D) = P(D | H_1) \cdot P(H_1) + P(D | H_2) \cdot P(H_2) + \ldots + P(D | H_n) \cdot P(H_n)$$
An **awesome new test** has been invented for an early detection of cancer. The probability that it correctly identifies someone with cancer as positive is 95%, and the probability that it correctly identifies someone without cancer as negative is 99%. The *incidence* of this type of cancer in the general population is $10^{-4}$. A random person in the population takes the test, and the result is positive.

**What is the probability that he/she has cancer?**

A. 99%
B. 95%
C. 30%
D. 1%

Get your i-clickers
An **awesome new test** has been invented for an early detection of cancer. The probability that it **correctly identifies someone with cancer as positive** is 95%, and the probability that it **correctly identifies someone without cancer as negative** is 99%. The **incidence** of this type of cancer in the general population is $10^{-4}$. A random person in the population takes the test, and the result is positive.

**What is the probability that he/she has cancer?**

A. 99%
B. 95%
C. 30%
D. 1%

Get your i-clickers
How come?
I thought it was a great test..

• Let $C$ – be the event that the patient has cancer; $C'$ – patient is cancer free
• $Y/N$ – events that test is Positive/Negative ($N=Y'$)
• We know: $P(C)=10^{-4}$, $P(Y|C)=0.95$, $P(N|C')=0.99$
• We need to find $P(C|Y)$
• Bayes to the rescue: $P(C|Y)=P(Y|C)*P(C)/P(Y)$
• What on Earth is $P(Y)$ ???
What on Earth is $P(Y)$ ???

- Likelihood that a random patient would test $Y$:
  
  \[
P(Y) = P(Y \cap C) + P(Y \cap C') = P(Y|C)P(C) + P(Y|C')P(C') = 0.95 \times 10^{-4} + (1-0.99)(1-10^{-4}) \approx 0.01\]

- Hence $P(C|Y) = P(Y|C) \times P(C)/P(Y)$
  
  \[
  \approx 10^{-4}/0.01 = 0.01 = 1\%
  \]

- But we would like it to be 100%, please!!!

- Until the false positive discovery rate $1-P(N|C')$ does not fall below the general population prevalence the result will never be close to 100%
What if I am already 50% sure (based on other tests) that a patient has cancer?

- That changes everything!
- Now $P(C) = P(C') = 0.5$
- $P(C|Y) = \frac{P(Y|C) * P(C)}{P(Y|C) * P(C) + P(Y|C') * P(C')} = \frac{0.95 * 0.5}{0.95 * 0.5 + (1-0.99) * 0.5} = 0.99$
- Now the doctor can be almost 100% sure.
- The importance of prior:
  - If prior belief that one has cancer is $10^{-4}$ – test is useless
  - If prior belief is at least 1% - the test is useful
Sensitivity/specificity of the standard test for prostate cancer: \text{PSA level} > 4.0\text{ng/mL}

- **Sensitivity of the test is 71.9%**
  - fraction of cancer patients who will test positive
  - False negative rate is 28.1%

- **Specificity of the test is 90%**
  - fraction of healthy patients who will test negative
  - False positive rate is 10%

Prostate cancer is the most common type of cancer found in males. It is checked by PSA test that is notoriously unreliable. The probability that a noncancerous man will have an elevated PSA level $>4.0$ ng/mL is approximately 0.1, with this probability increasing to approximately 0.719 if the man does have prostate cancer. If, based on other factors, a physician is 50 percent certain that a male has prostate cancer, what is the conditional probability that he has the cancer given that the test indicates an elevated PSA level?

A. 99.9%
B. 95%
C. 88%
D. 55%
All this trouble for a lousy 38% gain in confidence? I don’t believe you!

• Let $C$ – be the event that the patient has cancer; $C'$ – patient is cancer free, $E$ – events that the PSA test was elevated

• We know **doctor’s prior belief**: $P(C)=0.5$

• We know test stats: $P(E|C)=0.719$, $P(E|C')=0.1$

• We need to find $P(C|E)=P(E|C)*P(C)/P(E)$

• $P(E)=P(E|C)*P(C)+P(E|C')*P(C')=0.719*0.5+0.1*0.5=0.41$

• $P(C|E)=0.5*0.719/0.41=0.88$ or 88%