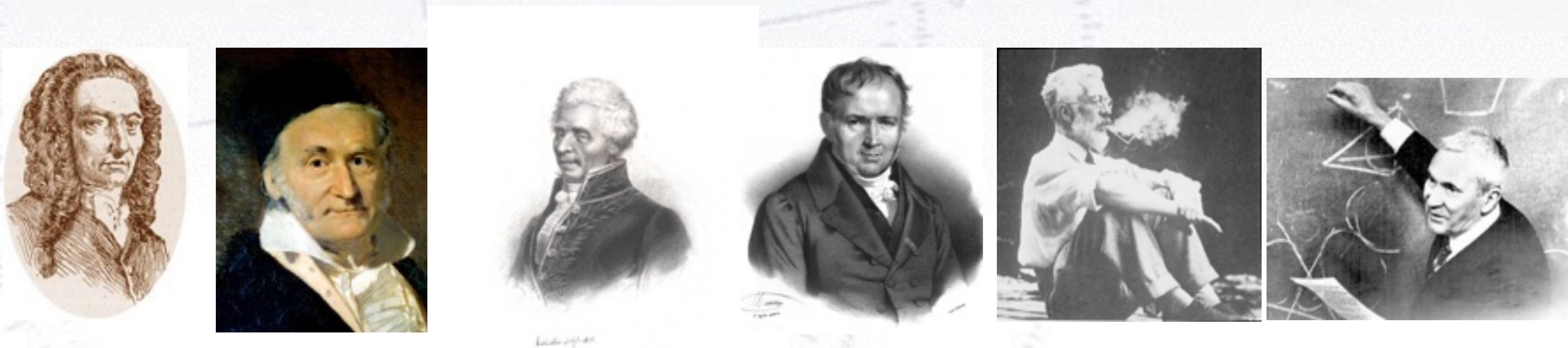
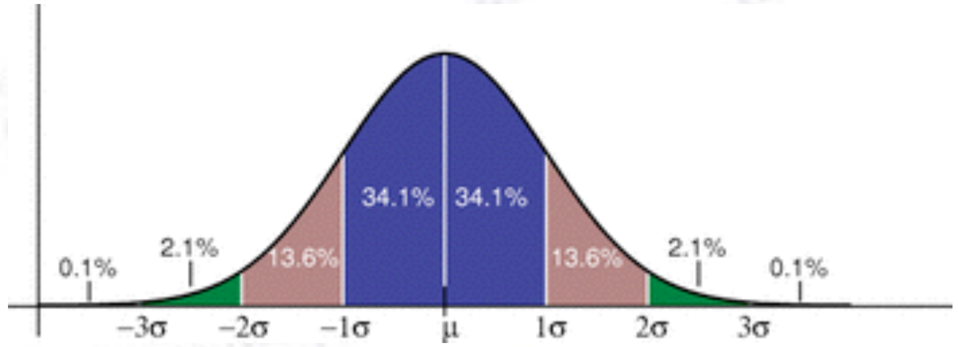
Applied Statistics

Bayes' Theorem



Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Problem

Suppose a drug test can be characterised as follows:

- 99% positive results for users (99% sensitive, i.e. 1% Type I errors).
- 99% negative results for non-users (99% specific, i.e. 1% Type II errors).

If 0.5% of a population is using the drug, and a random person tests positive, what is the chance that he is using the drug?

Problem

Suppose a drug test can be characterised as follows:

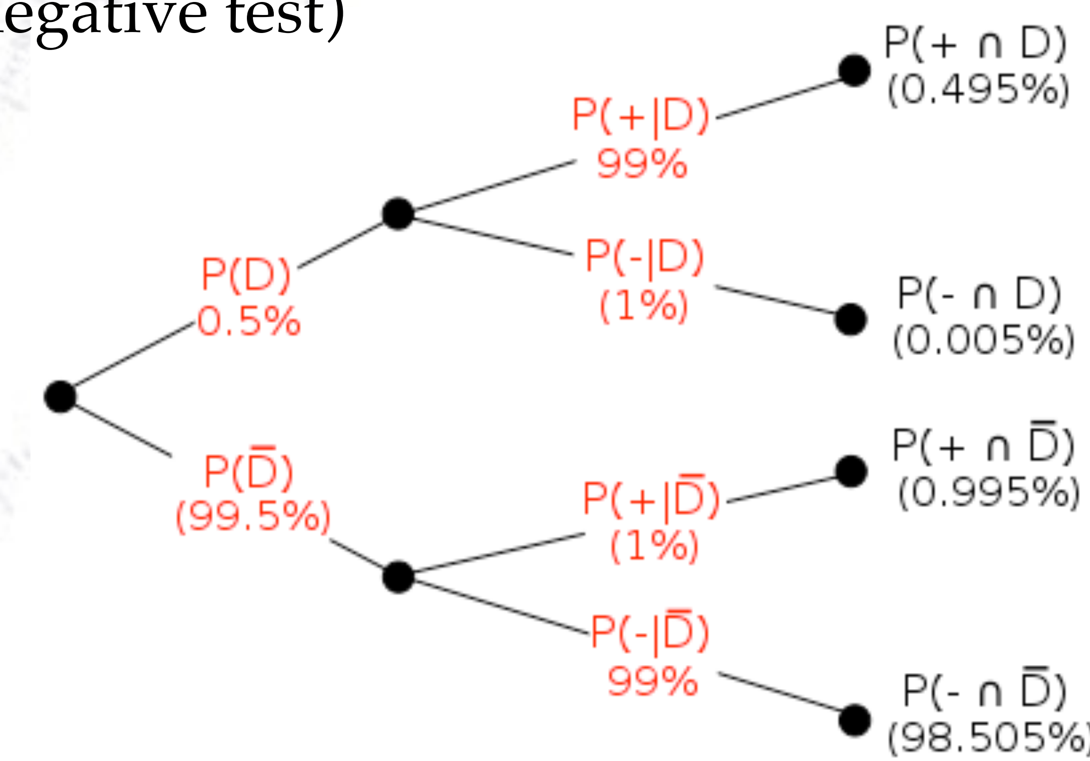
- 99% positive results for users (99% sensitive, i.e. 1% Type I errors).
- 99% negative results for non-users (99% specific, i.e. 1% Type II errors).

If 0.5% of a population is using the drug, and a random person tests positive, what is the chance that he is using the drug?

The answer is 33.2%, i.e. NOT very high! The reason is the **prior probability**. False positives (0.995%) are large compared to true positives (0.495%).

(D = user, $\overline{\text{D}}$ = non=user, + = positive test, - = negative test)

$$
\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\bar{D})P(\bar{D})} \\
&= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\
&= 33.2\%.
\end{aligned}
$$



Different versions...

The "original" version of Bayes' Theorem was stated as follows:

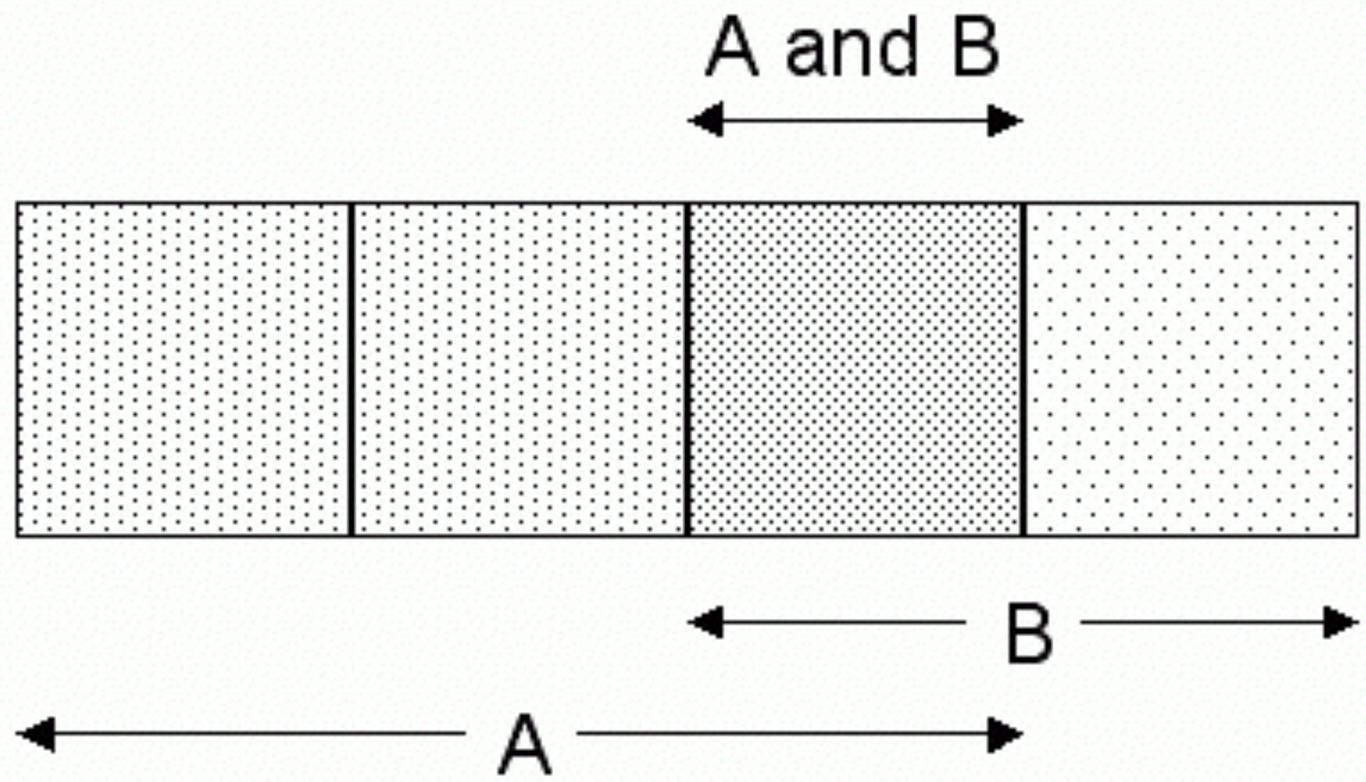
$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}.$$

However, it can be expanded (using the total law of probability) to:

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

It is in this form, that Bayes' Theorem is most often used.

Bayes' Theorem illustrated



P(A) = 3/4

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

P(B) = 2/4

P(A and B) = P(AB) = 1/4

P(A|B) = P(AB) / P(B) = (1/4) / (2/4) = 1/2

P(B|A) = P(AB) / P(A) = (1/4) / (3/4) = 1/3

Overview

★ Apply Bayes' theory to our the measurement of a parameter x

- We determine $P(\text{data}; x)$, i.e. the likelihood function
- We want $P(x; \text{data})$, i.e. the PDF for x in the light of the data
- Bayes' theory gives:

$$P(x; \text{data}) = \frac{P(\text{data}; x) P(x)}{P(\text{data})}$$

$P(\text{data}; x)$ the likelihood function, i.e. what we measure

$P(x; \text{data})$ the posterior PDF for x, i.e. in the light of the data

$P(\text{data})$ { prior probability of the data. Since this doesn't depend on x it is essentially a normalisation constant

$P(x)$ { prior probability of x, i.e. encompassing our knowledge of x before the measurement

★ Bayes' theory tells us how to modify our knowledge of x in the light of new data

Bayes' theory is the formal basis of Statistical Inference

From: M. A. Thomson

Example of priors influence

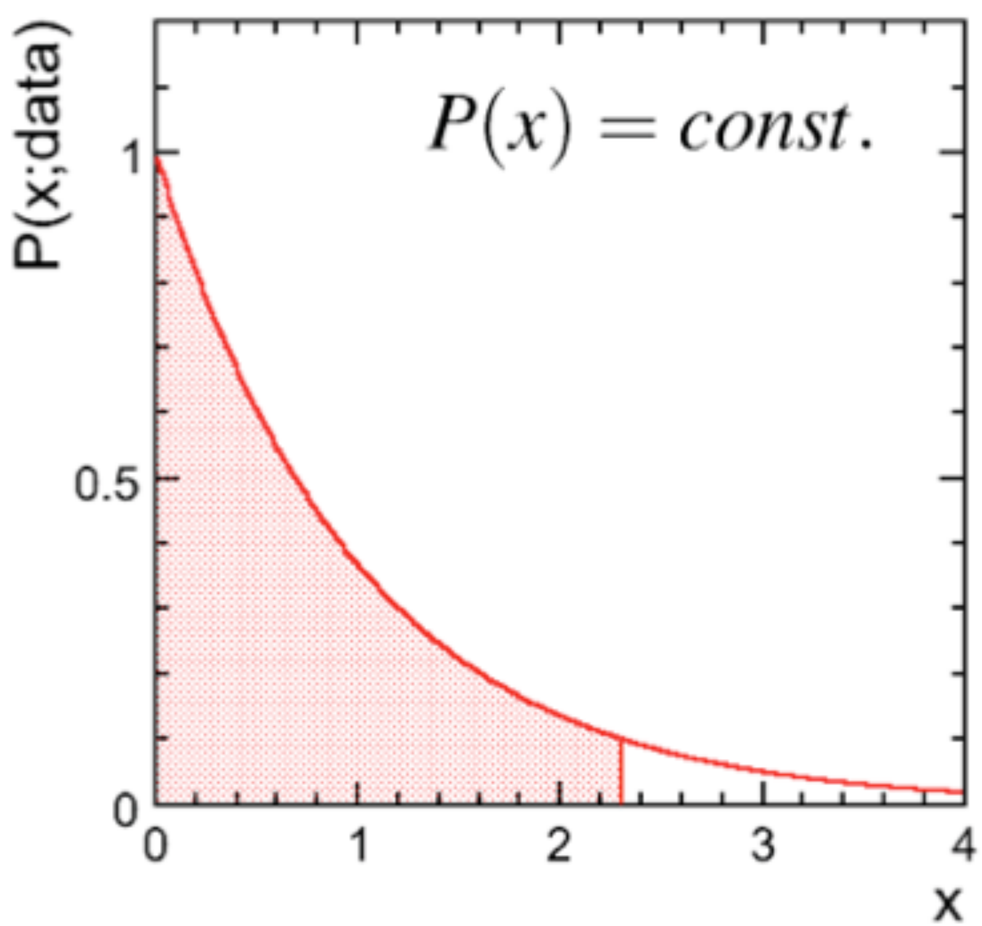
★ See no events…

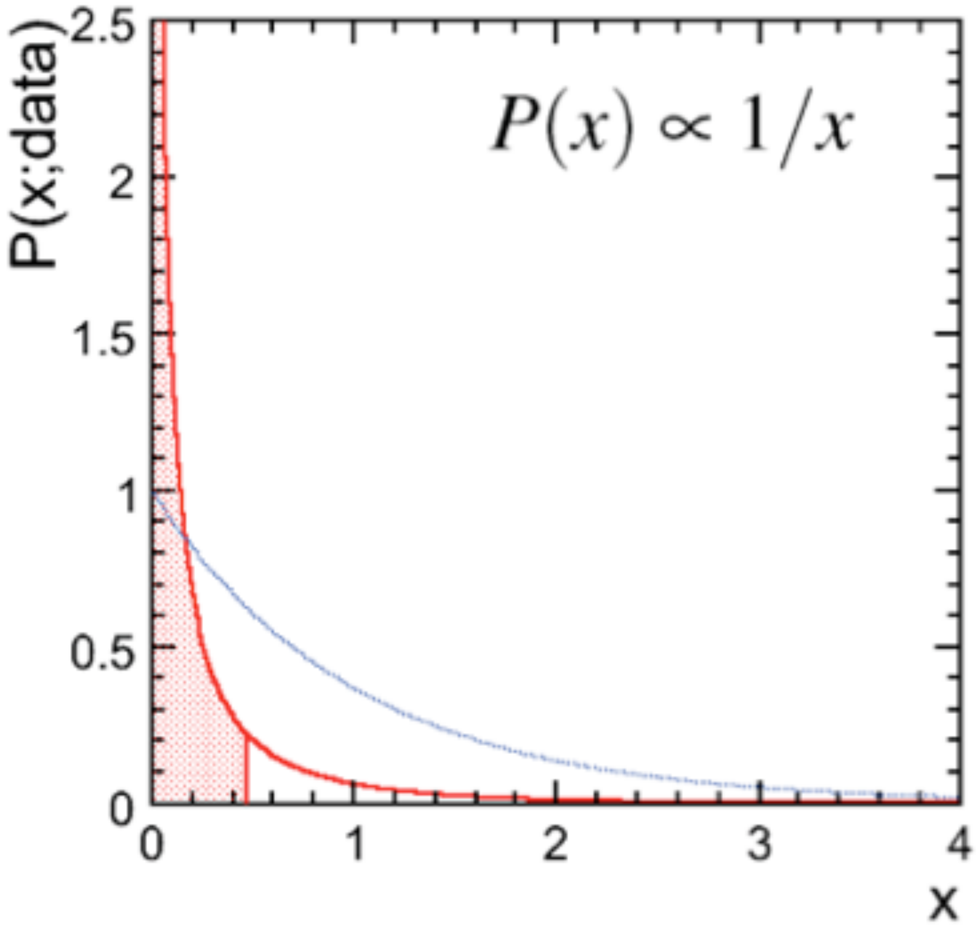
$$P(\text{data};x) = P(0;x) = e^{-x}$$

Poisson prob. for observing 0

Prior flat prior in x : $P(x) = const.$

Prior flat prior in lnx : $P(\ln x) = const.$





★ The Conclusions are very different. Compare regions containing **90 %** of probability

$x < 2.3$

$x < 0.46$

- In this case, the choice of prior is important

From: M. A. Thomson

Frequentist problems...

One of the reasons for Bayesian statistics, is the following problem for frequentist!

Imagine that you observe **5 events in data**, when expecting **0.9 background events**.
Then you can say with 68% confidence, that the signal is in the range **[2.8, 8.4]**.

But what if the expectation was **10.9 background events**?
Then you would say with 68% confidence, that the signal is in the range **[-8.1, -2.5]**.

Frequentist problems...

One of the reasons for Bayesian statistics, is the following problem for frequentist!

Imagine that you observe **5 events in data**, when expecting **0.9 background events**.
Then you can say with 68% confidence, that the signal is in the range **[2.8, 8.4]**.

But what if the expectation was **10.9 background events**?
Then you would say with 68% confidence, that the signal is in the range **[-8.1, -2.5]**.

While this is technically correct...

it is completely stupid!

We of course knew ahead of time, that the signal is either zero or positive.

The possible solution is to include the prior information, that the number is positive.
But exactly what prior to use? That is the problem.

Interpretations

One way Bayes' Theorem is often used in normal thinking is:

$$P(\text{theory}|\text{data}) \propto P(\text{data}|\text{theory}) \cdot P(\text{theory}).$$

Here, P(data) has been omitted (doesn't depend on parameters, so normalisation).

The trouble is, that it is hard to define P(theory) = a "degree of belief" in a theory.

Perhaps Glen Cowan sums it up best (chapter 1):

> Bayesian statistics provides no fundamental rule for assigning the prior probability to a theory, but once this has been done, it says how one's degree of belief should change in the light of experimental data.

"When the facts change, I change my opinion. What do you do, sir?"

[John Maynard Keynes]