CEITEC
Central European Institute of Technology
BRNO | CZECH REPUBLIC

Introduction to Bioinformatics (LF:DSIB01)

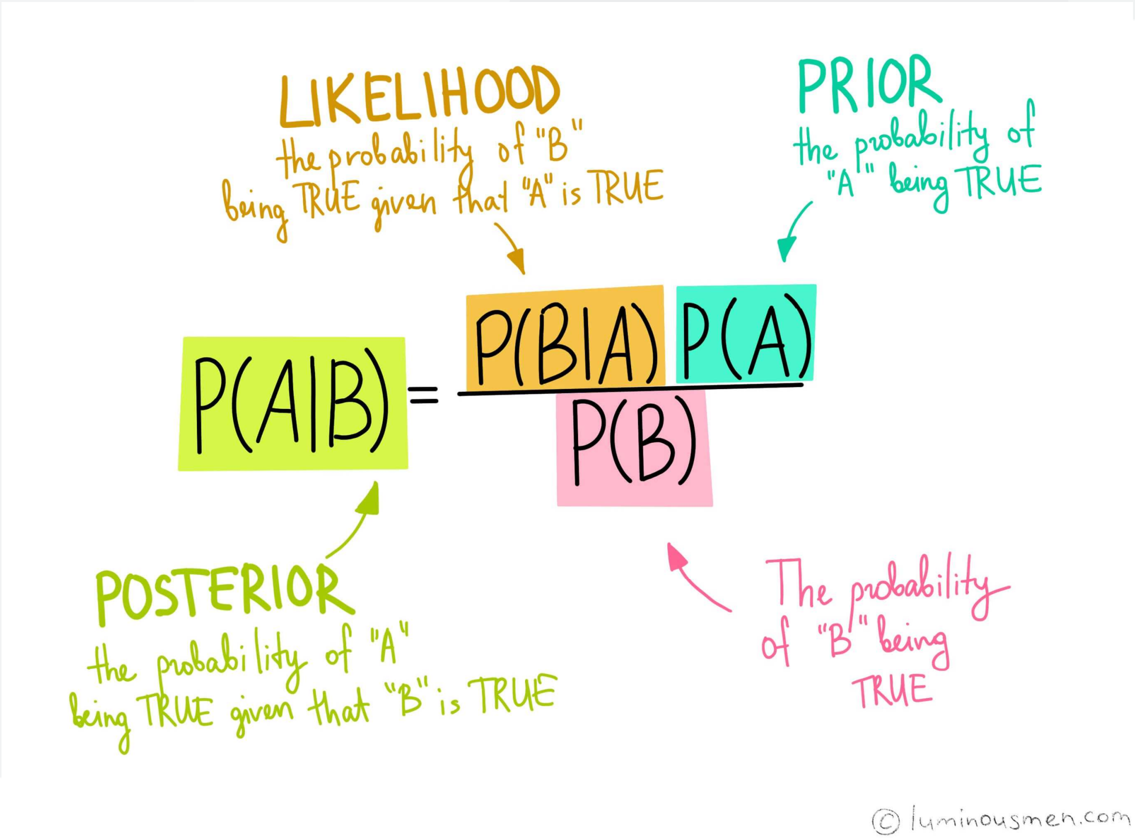
Week 8 : Bayesian modeling

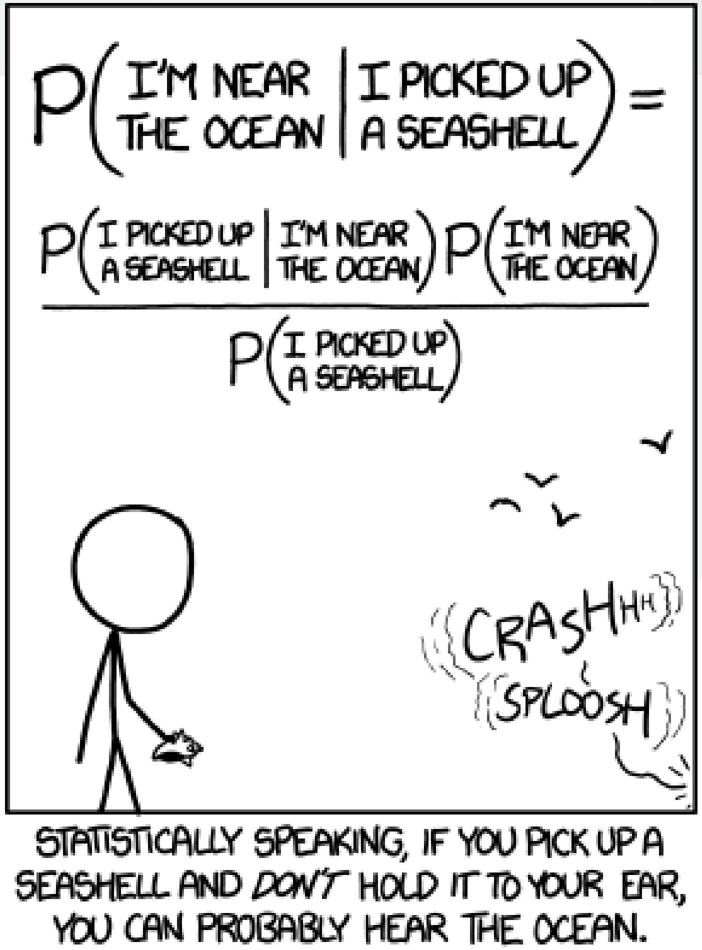
Bayes' theorem

Conditional Probability: Bayes' Theorem

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

Bayes' theorem





Bayes' theorem

Posterior

Likelihood

Prior

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

Normalizing constant

$$P(B) = \sum_{Y} P(B|A)P(A)$$

Bayes' theorem

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}.$$

The normal Bayes' theorem equation, where we chose a case A (called A_i)

$A_1, \ldots, A_n$

When A has 1 to *n* cases, and we choose some case *i* to find the probability of

The normal Bayes' theorem equation, where we chose a case A (called A_i)

Here we sum over each case, so there is *n* cases and we start at *j*, that is, we start at case 1, then case 2, all the way to case *n*

Bayes' theorem - example

- You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic.
  - Past data tells you that 10% of patients entering your clinic have liver disease. P(A) = 0.10.
  - Five percent of the clinic's patients are alcoholics.
  - Among those patients diagnosed with liver disease, 7% are alcoholics.

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

Bayes' theorem - example

- You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic.
  - Past data tells you that 10% of patients entering your clinic have liver disease. P(A) = 0.10.
  - Five percent of the clinic's patients are alcoholics.
  - Among those patients diagnosed with liver disease, 7% are alcoholics.

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

- P(A|B) = (0.07 * 0.1) / 0.05 = 0.14
- If the patient is an alcoholic, their chances of having liver disease is 0.14 (14%). This is a large increase from the 10% suggested by past data.

Bayes' theorem - example

- What is the probability that a woman has cancer if she has a positive mammogram result?
  - One in 1000 of women have breast cancer.
  - 98 percent of women who have breast cancer test positive on mammograms.
  - 1 percent of women without breast cancer have a positive mammogram.

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

Bayes' theorem - example

- What is the probability that a woman has cancer if she has a positive mammogram result?
  - One in 1000 of women have breast cancer.
  - 98 percent of women who have breast cancer test positive on mammograms.
  - 1 percent of women without breast cancer have a positive mammogram.

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

- P(A)=0.001, P(-A)=0.999, P(B|A)=0.98,P(B|-A)=0.01
- (0.98 * 0.001) / ((0.98 * 0.001) + (0.01 * 0.999)) = 0.0893
- The probability of a woman having cancer, given a positive test result, is ~9%.

Bayes' theorem - applications

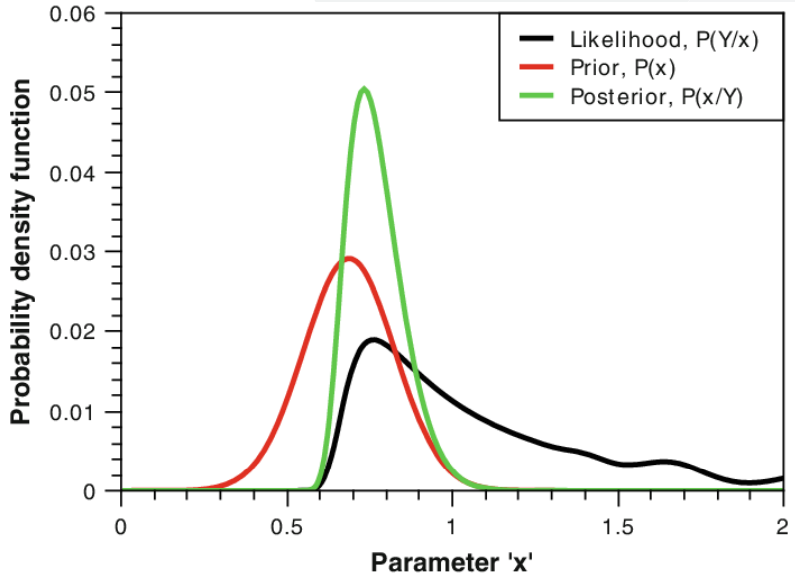
- Bayesian statistics
  - Data modeling
  - Parameter estimation
- Bayesian networks
  - Naïve Bayesian classifier
  - Dynamic Bayesian networks
    - Hidden markov models
- Can be used in many different context
  - Neural networks

Bayesian statistics

- Bayesian statistical methods use Bayes' theorem to compute and update probabilities after obtaining new data.

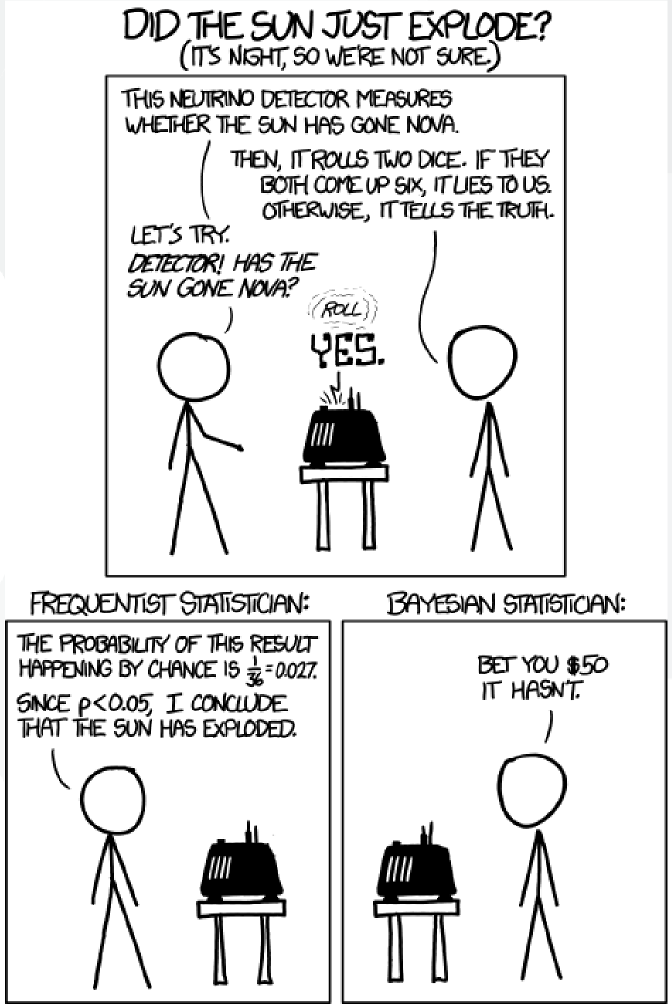
Posterior, Likelihood, Prior, Normalizing constant

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



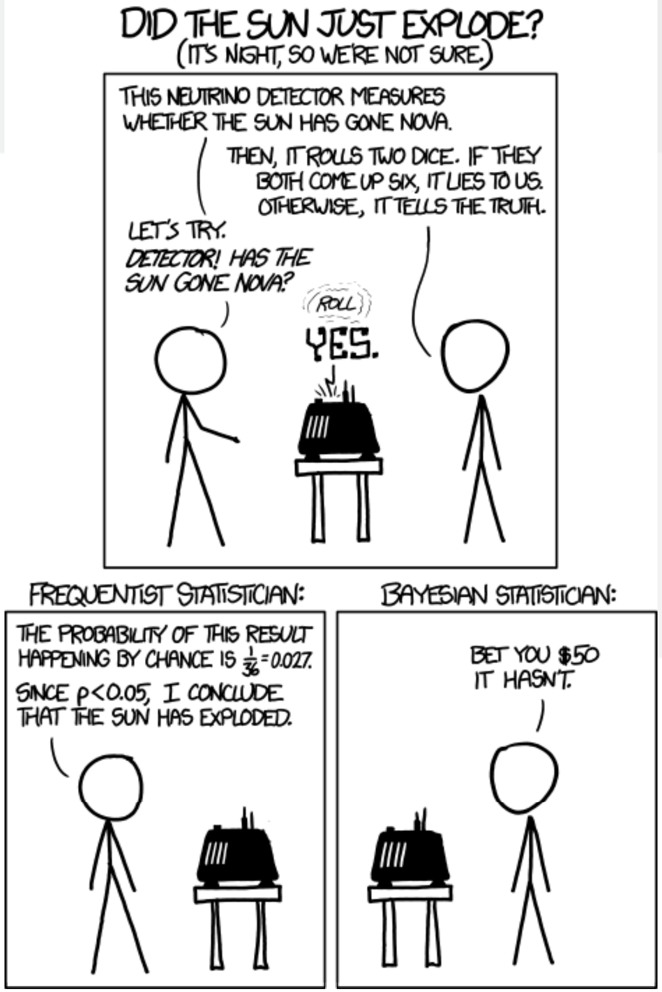
Bayesian vs. frequentists statistics

- Bayesian interpretation of probability where probability expresses a degree of belief in an event
- In the Bayesian view, a probability is assigned to a hypothesis, whereas under frequentist inference, a hypothesis is typically tested without being assigned a probability.
- The frequentist interpretation that views probability as the limit of the relative frequency of an event after many trials



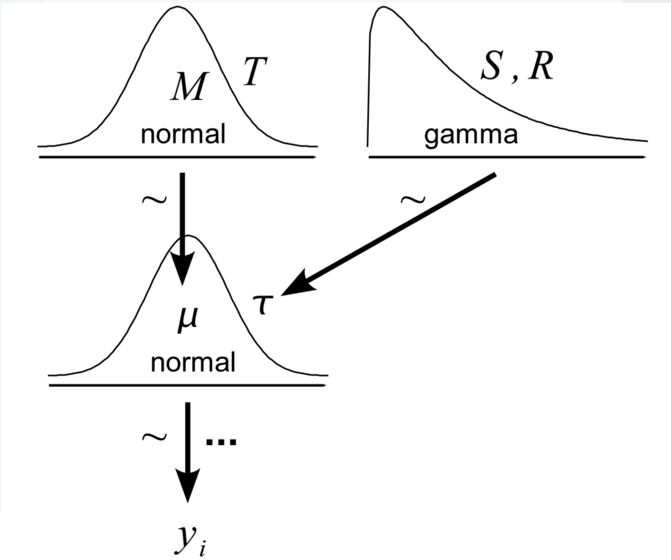
Bayesian vs. frequentists statistics

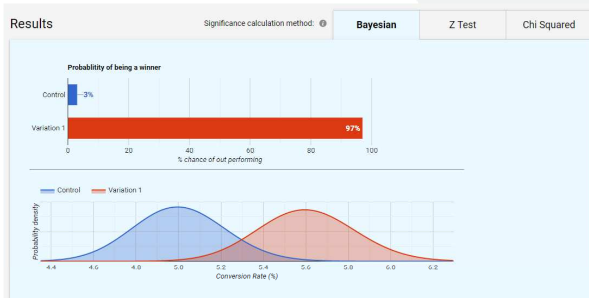
- Two main sticking points
  - Prior believe
  - Small amount of data situations

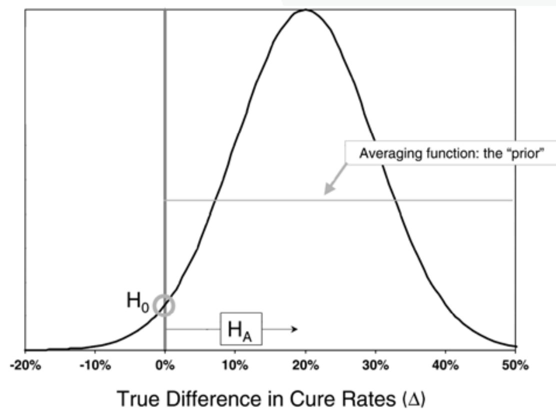


Bayesian statistics – parameter estimation

- Data from some probability distribution function (PDF)
- The goal is to get PDF of the parameter of this function

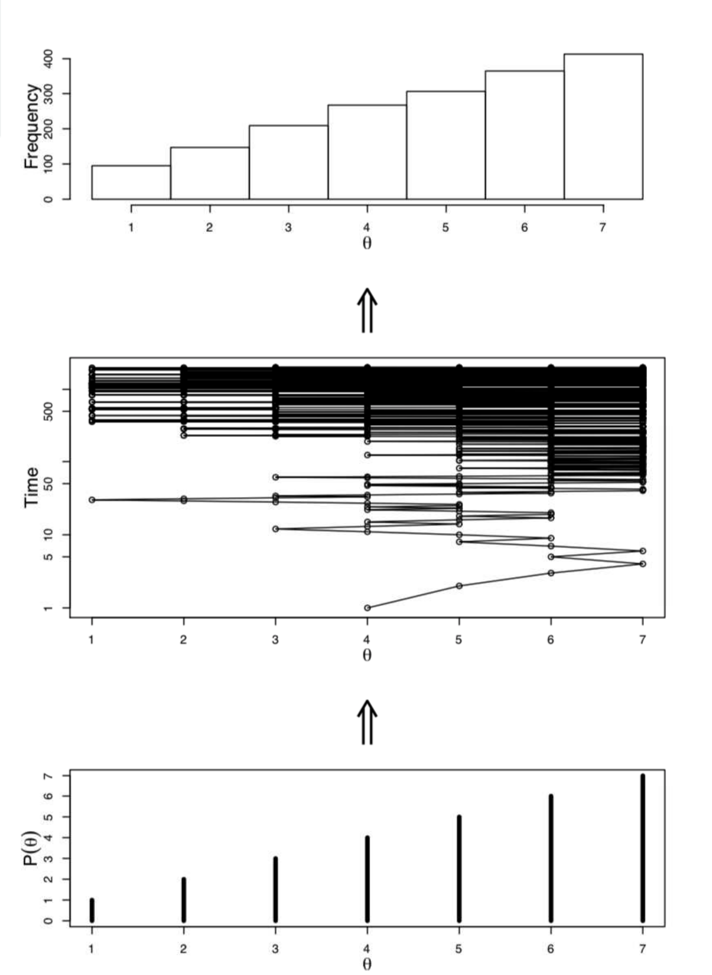






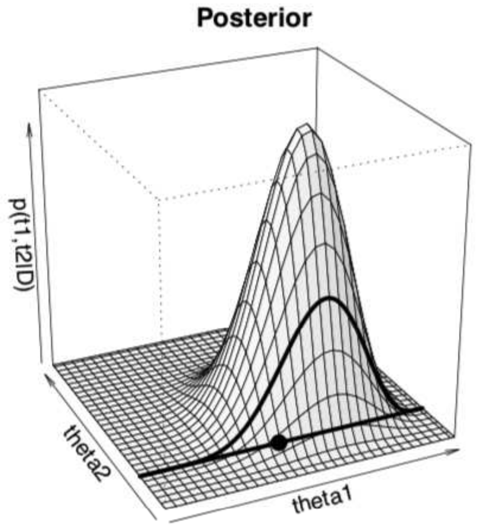
PDF estimation with sampling

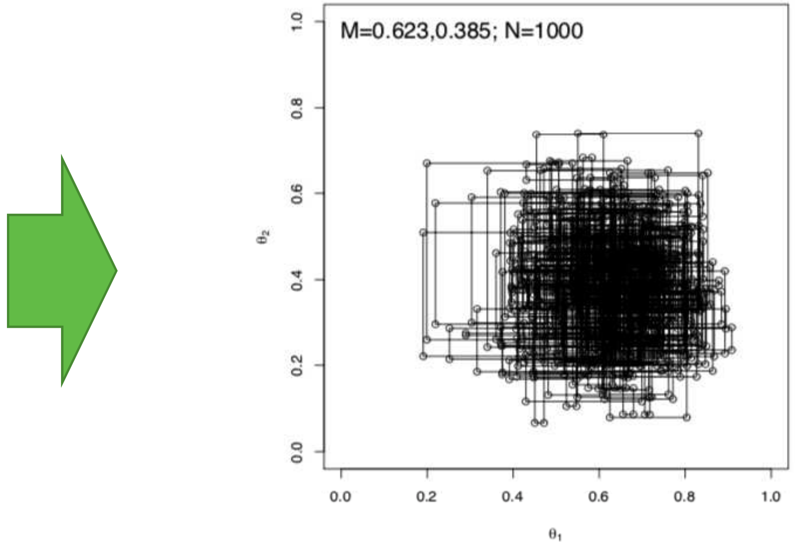
- Impossible (very hard) to compute analytically
- Markov chain Monte Carlo (MCMC)
  - This allowed usage of Bayesian statistics



PDF estimation with sampling

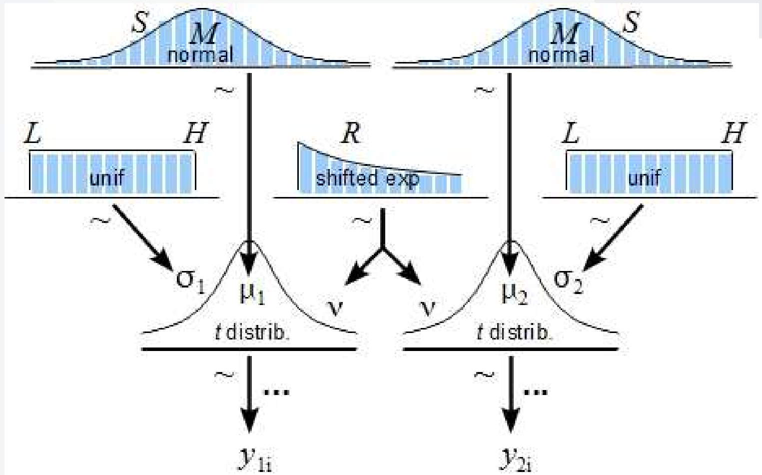
- Impossible (very hard) to compute analytically
- Markov chain Monte Carlo (MCMC)
  - This allowed usage of Bayesian statistics

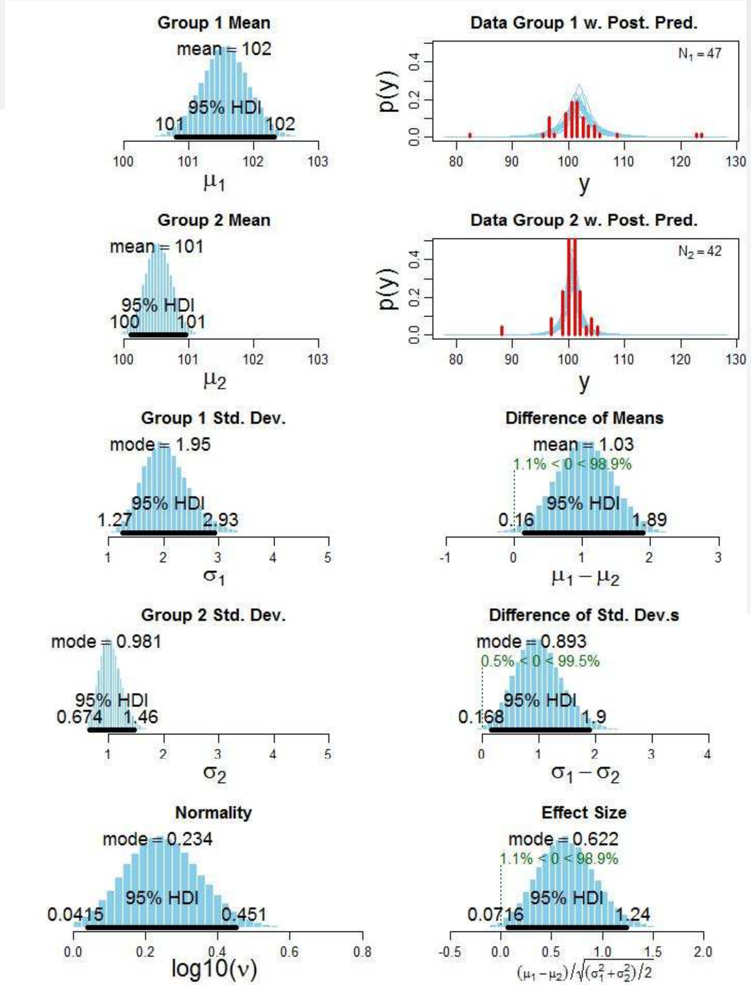




Bayesian statistics - example

- Student t-test

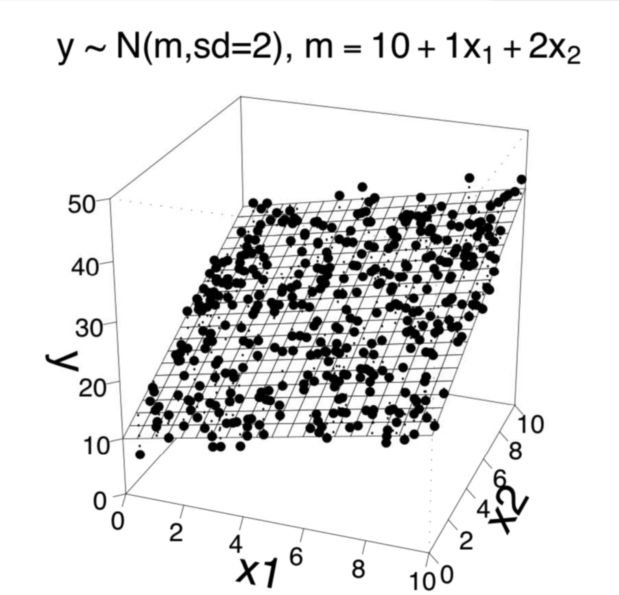


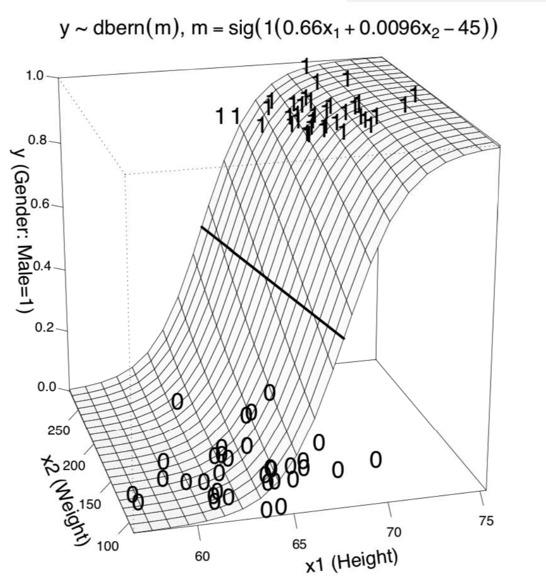


Bayesian statistics – Generalized linear model

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$y = \mathrm{sig}(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$





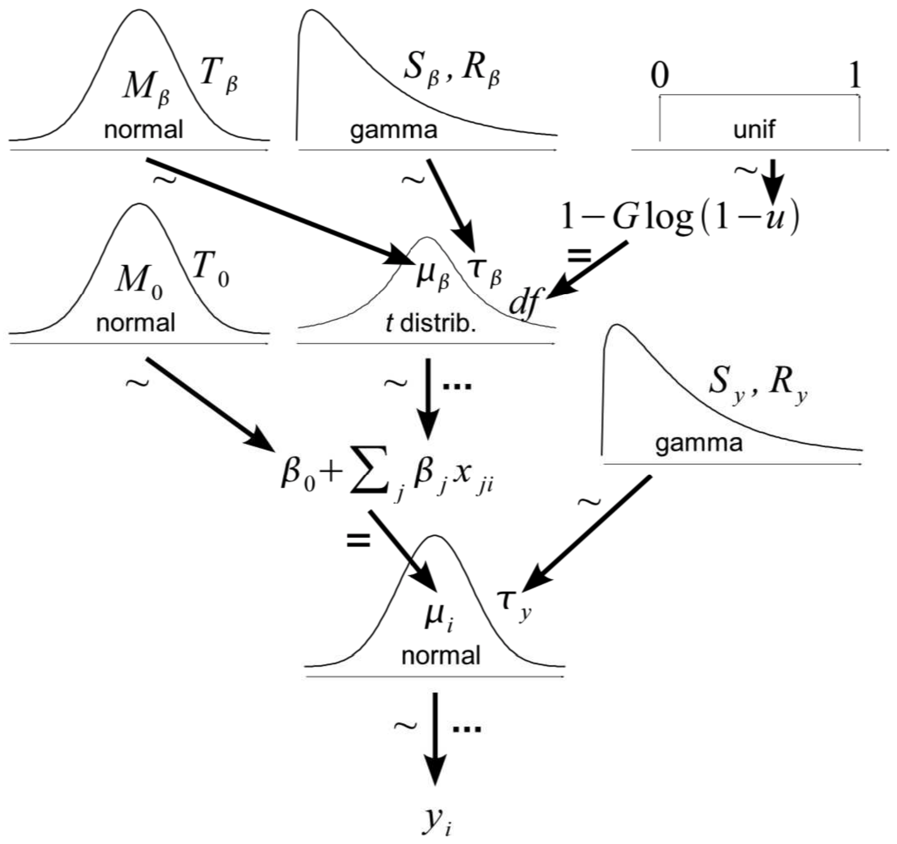
Bayesian statistics – Generalized linear model

y Scale Type	Link Function	pdf
metric	identity	normal
dichotomous	logistic	Bernoulli
ordinal	thresholded cumulative normal	categorical
count	exponential	Poisson

Bayesian statistics – Generalized linear model

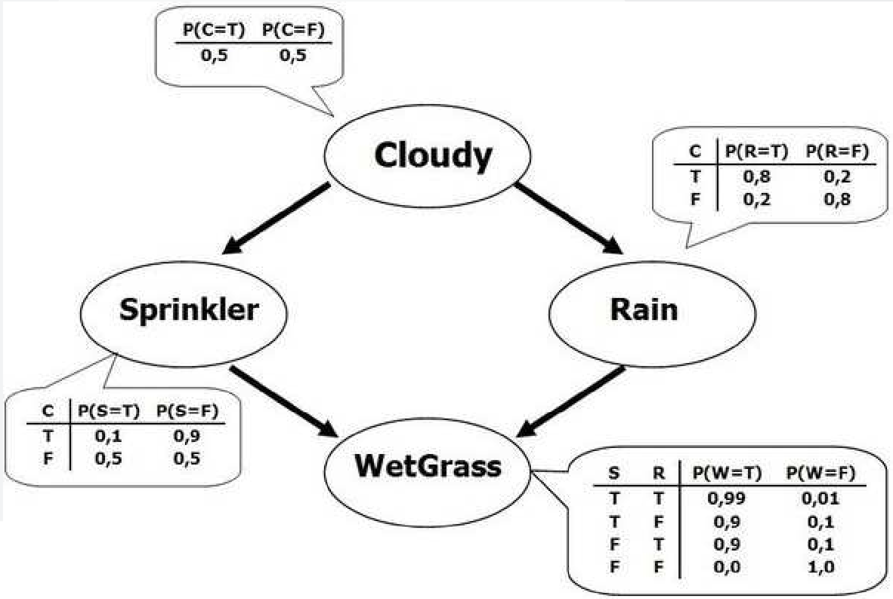
Response variable type	Explanatory variable type	Example test type
Categorical	Categorical	Fisher test
Categorical (two groups)	Continuous	t-test
Categorical (multiple groups)	Continuous	ANOVA
Continuous	Continuous	Linear regression
Continuous	Categorical (two groups)	Logistic regression

Bayesian statistics – hierarchical models

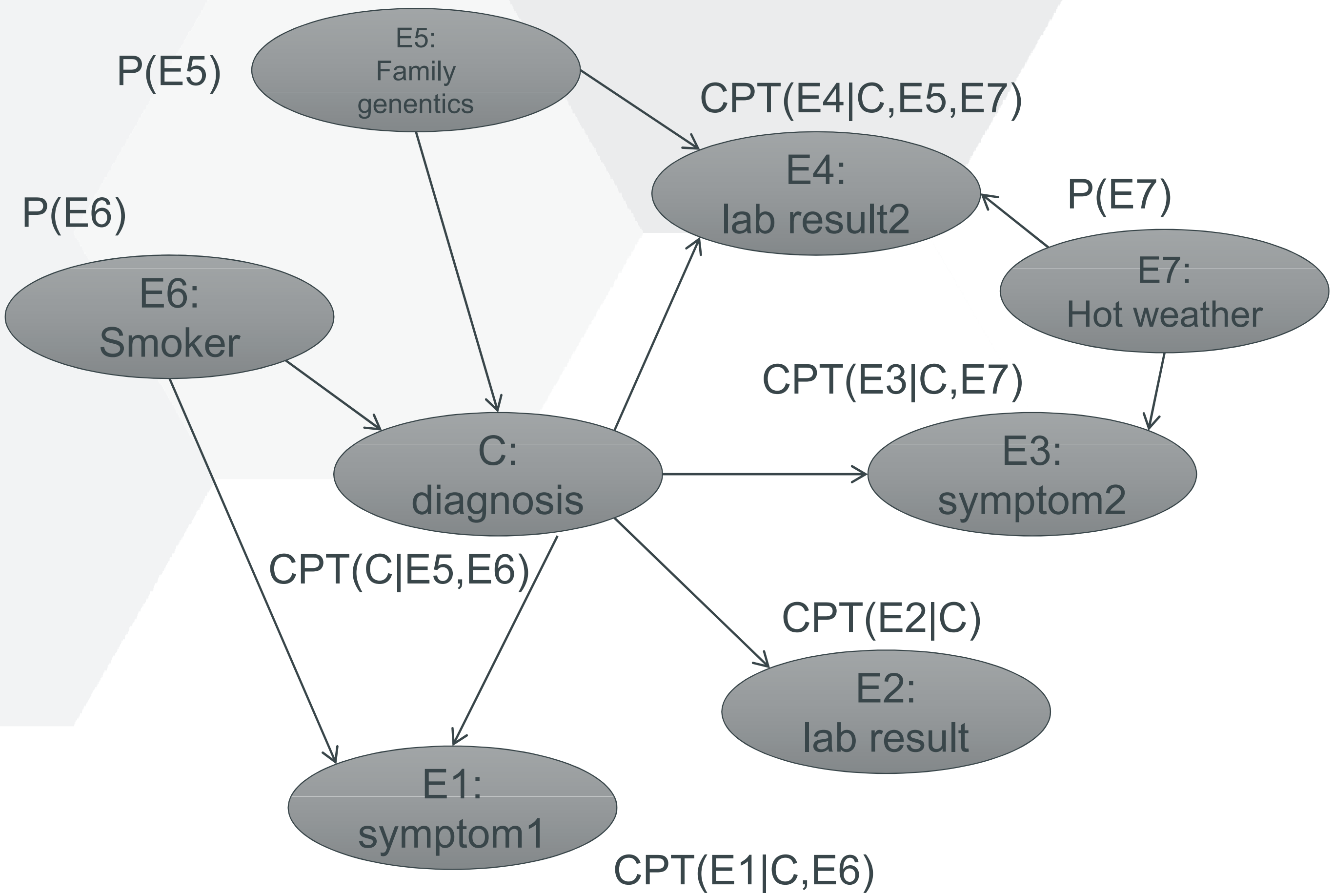


Bayesian networks

- Diracted acyclic probabilistic graph
- Represents a set of variables and their conditional dependencies

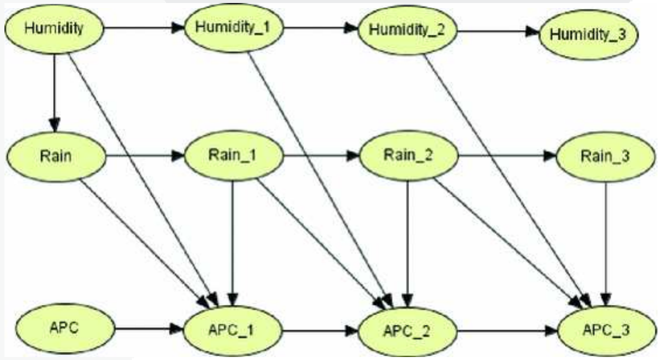


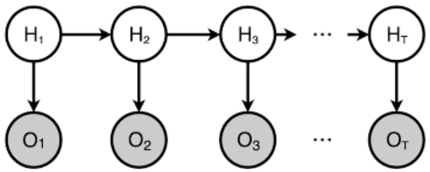
Bayesian networks - example



Dynamic Bayesian networks

- Hidden markov models (hmm) are a (simple) special case of DBN





Naive Bayesian Classifier

The normal Bayes' theorem equation, where we chose a case A (called A_i)

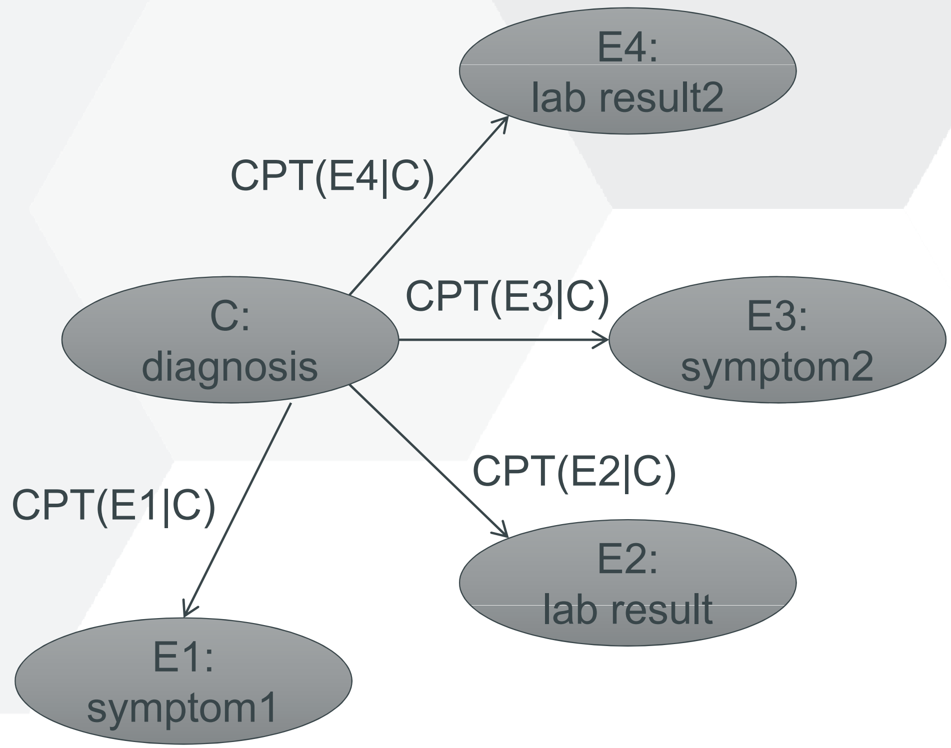
$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}.$$

$A_1, \ldots, A_n$
When A has 1 to *n* cases, and we choose some case *i* to find the probability of

The normal Bayes' theorem equation, where we chose a case A (called A_i)

Here we sum over each case, so there is *n* cases and we start at *j*, that is, we start at case 1, then case 2, all the way to case *n*

Naive Bayesian Classifier – simple example



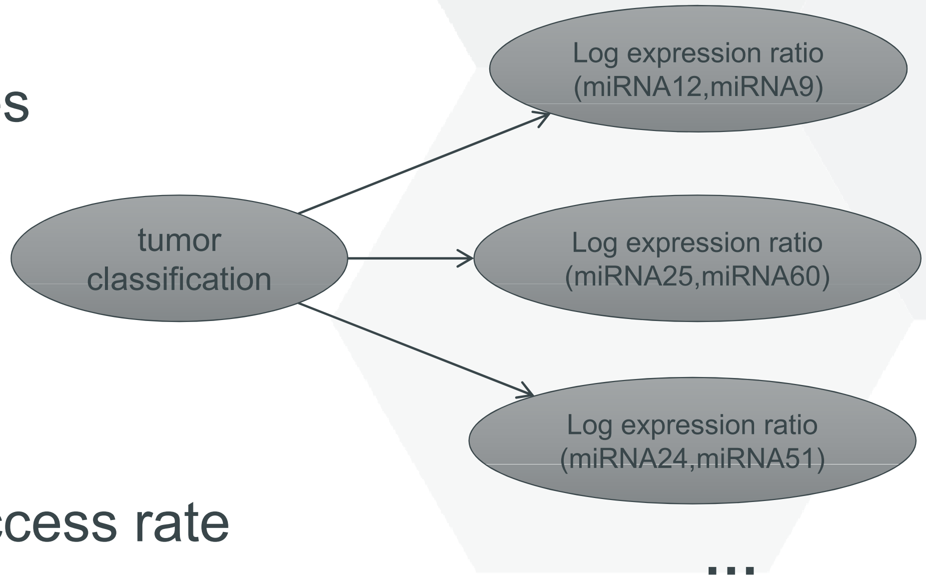
Naive Bayes classifiers - properties

- highly scalable, requiring number of parameters linear in the number of variables
  - curse of dimensionality
- training can be done by evaluating a closed-form expression which takes linear time

- Assumption of independence
  - Simple model
- Successfully being used
- Best if you have poor understanding of the problem.

Naive Bayesian Classifier - example

- Pancreatic tumor classification based on miRNA levels
- miRNA sequencing from plasma samples
  - ~300 miRNAs
- Select ~20 miRNAs to classify tumor types



- Bayesian model - 24 miRNAs -> 85% success rate

CEITEC

@CEITEC_Brno

Thank you for your attention!
60 minutes lunch break.

www.ceitec.eu