Analysis of Microarray Data with Methods from Machine Learning and Network Theory

Summer Lecture 2015

Prof. Dr. A. B. Cremers

Dr. Jörg Zimmermann

Introduction to Inferential Statistics

Goal: Find a Hypothesis which is a good explanation for your data and quantify your uncertainty

There are two basic types of Uncertainty:

Data Uncertainty:	Hypothesis Uncertainty:
Contingency (Event risk)	Plausibility (Model risk)

If data contingency is modelled by probability, we are in the realm of classical statistics, if both contingency of data and plausibility of hypotheses is modelled by probability, it is called Bayesian statistics (there are other approaches to modeling uncertainty, this is an active area of research, especially in Machine Learning and Artificial Intelligence).

Event Algebra

First of all we have to define the domain of discourse, the elementary and complex events we want to speak about:

The set of ementary events: Ω (e.g. {1,2,3,4,5,6} for a dice)

If a set of subsets **A** of Ω has the following properties:

1. If A and B are in **A**, then the union of A and B is in **A**.
2. If A and B are in **A**, then the intersection of A and B is in **A**
3. If A is in **A**, then the complement of A is in **A**

it is called an event algebra.

Probability Axioms

A probability measure is a mapping from an event algebra into the [0,1]-interval of the real numbers, satisfying the following axioms:

1. $P(\Omega) = 1$

2. $P(A \cup B) = P(A) + P(B), \text{ if } A \cap B = \emptyset$

3. $P(AB) = P(A|B) \cdot P(B)$

An equation like "P(A) = p" is read: "The event A has probability p under probability measure P".

The expression "P(A|B)" occurring in the 3. axiom denotes a conditional probability and is read: "The probability of A given B".

Random Variable

Definition: A probability space is a pair (**A**,P), consisting of an event algebra **A** and a probability measure P for this event algebra.

Definition: Let (**A**,P) be a probability space. If Ω = R (real numbers), (**A**,P) defines a continuous random variable.

Definition: Let (**A**,P) be a probability space. If Ω = N (natural numbers) (**A**,P) defines a discrete random variable.

Example: the scan intensities are modeled as continuous random variables.

Random Variables are denoted by upper case letters like X, Y, ...

Expectation, Variance, Correlation

The expectation of a (discrete) random variable is defined as:

$$E(X) = \Sigma_i x_i \cdot p(x_i)$$

(where *i* varies over all elements of Ω)

The variance of a random variable is defined as:

$$Var(X) = E((X - E(X))^2)$$

The mean or standard deviation is defined as:

$$MeanDeviation(X) = \sqrt{Var(X)} = \sigma_X$$

Expectation, Variance, Correlation

The covariance (a measure of relationship of two random variables) is defined as:

$$Cov(X,Y) = E((X - E(X))(Y - E(Y))$$

The correlation is a normed covariance. It assumes always a value in the interval [-1,1]. It is the most common measure used to express the degree of dependency of two random variables:

$$Correlation(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Properties of Correlation:

The correlation between to random variables is a normed covariance. It assumes always a value in the interval [-1,1]:

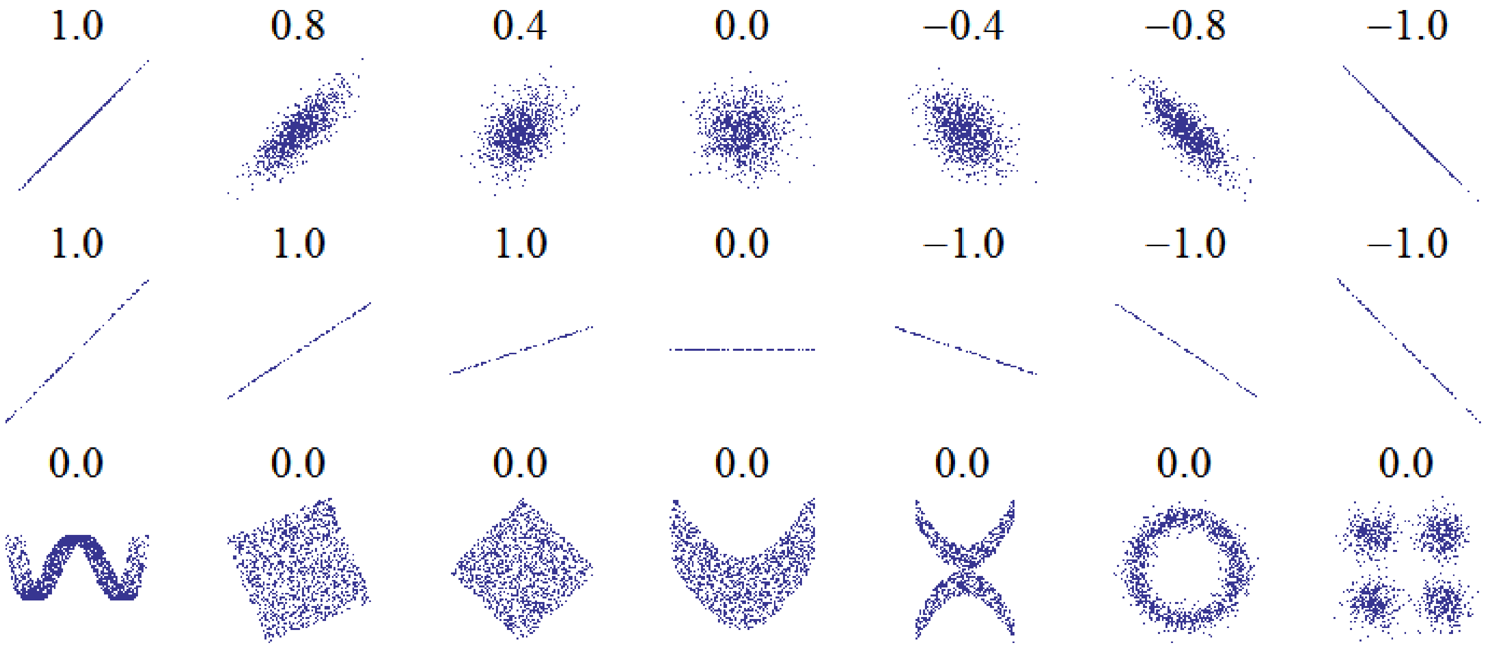
$$Correlation(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

A value near 1 is called correlated, a value around 0 is called uncorrelated and a value near -1 is called anti-correlated.

Correlation measures a degree of linear dependence between random variables, not general dependence or independce.

Independent variables (P(X,Y) = P(X)P(Y)) have always correlation of 0, but not vice versa, e.g., if there is a non-linear relationship between X and Y.

Examples of Correlation



Correlation reflects the noisiness and direction of a linear relationship, but not the slope of that relationship, nor nonlinear relationships.

Image Source: Wikipedia

Correlation vs. Causation

It is important to note that correlation between two variables implies by no means causation!
If you measure two variables of a person:

X = Person p has lung cancer

Y = Air Quality of living room of person p

then you will likely find a high correlation between X and Y. But that musn't mean that lung cancer is caused by bad air quality. There could be a third variable, which is hidden and which causes lung cancer and bad air quality.

In this case such a third variable could be “Person is smoking”. Such hidden variables are called confounders and it is an important task of statistical modeling to identify or eliminate such confounding variables.

Law of Large Numbers

How are probabilities of events and observed frequencies connected?

$$P(|\frac{1}{n}(X_1 + ..X_n) - EX| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2 n}$$

This means that if one repeats a stochastic experiment many times, the probability that the average of outcomes deviates significantly from the expected value goes to zero.
The law of large numbers gives a quantitative description of this relationship and is thus a cornerstone of statistics, connecting theoretical probabilities with real observations.

Probability Densities:

In the case of continuous random variables it is convenient to describe a probability measure by a probability density.

A probability density p(x) is a nonnegative function from the real numbers into the real numbers with the property that the whole area below the function graph and the x-axis equals 1.

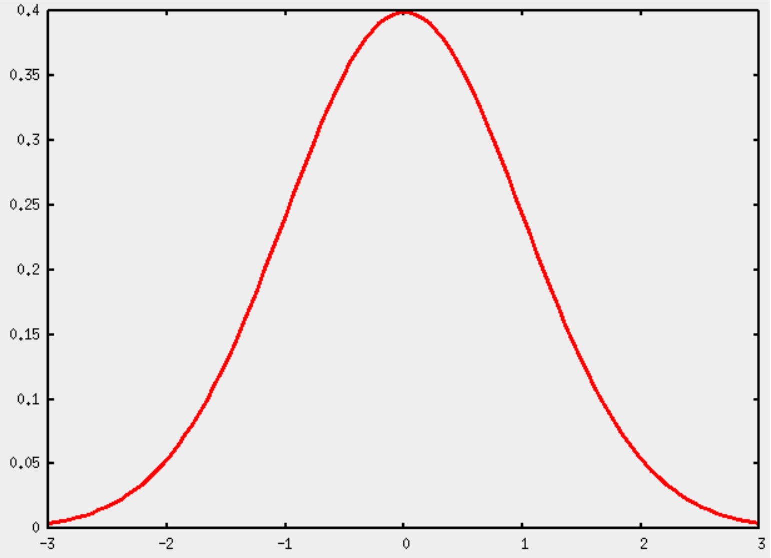
The probability, that an event falls into the interval [a,b] is given by:

$$P([a,b]) = \int_a^b p(x)dx$$

Important distributions

The most common continuous distribution is the Normal- or Gauss-Distribution, given by the following density:

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

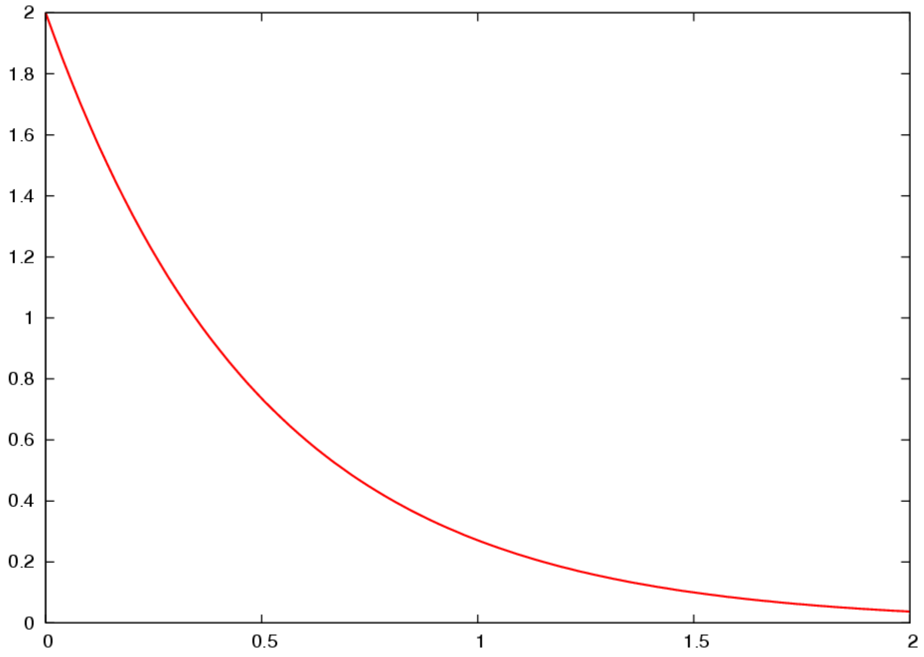


It has a bell-shaped curve. We will later see why it is so common (Central Limit Theorem).

Exponential Distribution

Another common distribution is the exponential distribution. It appears in many real world situations like the waiting times between to random events:

$$f_\lambda(x) = \lambda e^{-\lambda x} \qquad (x \geq 0)$$



Bernoulli-Experiment

The Bernoulli-Experiment is the most simple discrete distribution. It consists of the independent and identical distributed (i.i.d.) repetition of an experiment having only two outcomes: 0 and 1.

$$b_{p,n}(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

x denotes the number of positive outcomes in a Bernoulli-experiment of length *n*. *p* is the probability of a positive outcome and is the only parameter of a Bernoulli-experiment of fixed length *n*.

Point Estimation

Define a family of possible stochastic models as explanations for data parameterized by one (or many) real parameter(s).

Define a function which maps a data sample to the parameter of the model family: this is called an point estimator.

Common approach to construct an estimator: Maximum Likelihood.

Point Estimation

An estimation problem is given by the following data:

1. An event Algebra **A,** called the sample space

2. A family of probability measures $\{P_\theta: \theta \in \Theta\}$, the potential models the observed data

Problem: How to determine a good parameter of the family when you have data from event Algebra A?

Maximum Likelihood Estimator

A common approach to find an estimator is to maximize the Likelihood. The Likelihood function L_x, given observation x, assigns to a parameter θ the probability of x under the probability measure defined by θ:

$$L_x(\theta) = P_\theta(x)$$

So the Likelihoodfunction is just a change in the point of view: now the data x are viewed as fixed, and the parameter θ is varying.

Maximum Likelihood Estimator

If L_x assumes a maximal value for θ*, i.e. if:

$$L_x(\theta^*) = sup\{L_x(\theta) : \theta \in \Theta\}$$

then we call θ* a maximum likelihood estimator of parameter θ.

In many practical situations there is a unique maximum likelihood estimator, and it is usually a good estimator.

But what does “good” mean? We will see criteria for that soon, but first we will have an example of a maximum likelihood estimator.

Example

Maximum Likelihood Estimator for a Bernoulli-Experiment
of fixed length n:

$$L_x(p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Likelihoodfunction

$$logL = ln\binom{n}{x} + x\ ln(p) + (n-x)\ ln(1-p)$$

Loglikelihoodfunction

$$\frac{d}{dp} logL = \frac{x}{p} - \frac{n-x}{1-p}$$

Derivation and solving for zero

The Maximum Likelihood-Estimator for a Bernoulli-Experiment is:

$$\boxed{\hat{p}(x) = \frac{x}{n}}$$

Unbiased Estimator

An Estimator *g* is called unbiased, if the expectation of the estimator (wrt. parameter θ) is the true value of the parameter:

$$E_\theta(g(X)) = \theta$$

If an estimator is unbiased, it will yield goold results on average.

Unbiasedness is a desirable property of an estimator.
The ML-estimator for Bernoulli-experiments, for example,
is unbiased.

Quadratic Error

The mean quadratic error of an estimator is:

$$R(\theta, g) = E_\theta((g(X) - \theta)^2)$$

R depends on θ. In general, one estimator is good for some parameter values and another estimator is good on other parameter values. But if one assumes unbiasedness of estimators, then in many cases there exists an estimator which has uniformly minimum quadratic error.

These estimators, if they exist, are called UMVU-estimators (Uniformly Minimum Variance Unbiased). The ML-estimator for Bernoulli-experiments is an UMVU-estimator.

Discussion

Are there always unbiased estimators? Unfortunately, no.
So we have to seek for other criteria as well.

Are they unique? No, not in general.

Are they optimal? In the case of UMVU-estimators, they can be called optimal in a reasonable sense. But we will see other optimality criteria.