October 2001

Statistical Formulas

Probability
Monte Carlo methods
Distributions
Practical problems
Numerical linear algebra
Information and algorithms

Volker Blobel

Institut für Experimentalphysik
Universität Hamburg
2001

List of mathematical symbols

x, y ...	random variable (one-dimensional)
$\boldsymbol{x}$, $\boldsymbol{y}$ ...	random vector (n-dimensional)
μ	expectation value, mean value
$\langle x \rangle$	mean value
μ_x	expectation value, mean value of random variable x
σ	standard deviation (or cross section)
σ_x	standard deviation of random variable x
σ_{xx}	variance = square of standard deviation of random variable x, $\sigma_{xx} = \sigma_x^2$
σ_{xy}	covariance between random variables x and y
$\boldsymbol{V}$	covariance matrix
ρ	correlation coefficient
ρ_{xy}	correlation coefficient between random variables x and y, $\rho_{xy} = \sigma_{xy}/(\sigma_x \sigma_y)$
$N(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$N(0, 1)$	Normalized Gaussian distribution with mean 0 and variance 1
$U(0, 1)$	Uniform distribution between 0 and 1

Contents

1 Probability **5**
1.1 Probability . . . 5
1.2 Permutations . . . 6
1.3 Probability density . . . 6
1.4 Expectation values and moments . . . 7
1.5 Tschebyscheff inequality . . . 8
1.6 Characteristic functions . . . 8
1.7 The central limit theorem . . . 10

2 Monte Carlo methods **11**
2.1 Generation of uniform random numbers . . . 11
2.2 Tests for random number generators . . . 12
2.3 Generators for general random numbers . . . 12
2.4 Monte Carlo integration and numerical integration . . . 13
2.5 Variance reducing methods . . . 15

3 Distributions **17**
3.1 Binomial distribution . . . 17
3.2 Multinomial distribution . . . 18
3.3 Poison distribution . . . 18
3.4 Uniform distribution . . . 19
3.5 One-dimensional normal distribution . . . 19
3.6 Two-dimensional normal distribution . . . 21
3.7 Multivariate normal distribution . . . 22
3.8 Exponential distribution . . . 23
3.9 Gamma distribution . . . 23
3.10 χ^2 distribution . . . 23
3.11 Students t distribution . . . 24
3.12 F distribution . . . 26
3.13 Cauchy distribution . . . 26

4 Practical problems **28**
4.1 Expressing the uncertainty of measurement results . . . 28
4.2 Sample mean and variance . . . 29
4.3 Averages . . . 29
4.4 Method of moments . . . 30

5 Numerical linear algebra **32**
5.1 Vectors and matrices . . . 32
5.2 Symmetric matrices . . . 33
5.3 Exchange algorithm . . . 34
5.4 Triangular matrices . . . 34
5.5 General LR decomposition . . . 35
5.6 Cholesky decomposition . . . 36
5.7 Inversion by partitioning . . . 36
5.8 Diagonalization of symmetric matrices . . . 37
5.9 Singular value decomposition . . . 39

6 Information and algorithms **42**
6.1 Data representation . . . 42
6.2 Information . . . 44
6.3 Analysis of algorithms . . . 44
6.4 Sorting . . . 45
6.5 Searching . . . 46

Glossary **48**

1 Probability

1.1 Probability

Historically the interest in statistical problems started with investigations of games of chance. In games of chance there are often many different outcomes, which a-priori are equally likeli (tossing a coin, throwing a dice); this can be used to make a *combinatorical* definition of probability. Similarly one can give an *empiricial* definition of probability.

Combinatorical definition. If an event can happen in n different ways, which are all equal probable, and if k of the different ways have the probability $\mathcal{A}$, then the probability for event $\mathcal{A}$ is given by

$$P(\mathcal{A}) = \frac{k}{n} \, .$$

Empirical definition of probability. An observation is repeated independently under identical conditions n times; if property $\mathcal{A}$ is observed in k of the n cases, the ratio k/n is the empirical probability for $\mathcal{A}$. The probability $P(\mathcal{A})$ can be defined as the ratio in the limit for n going to infinity:

$$P(\mathcal{A}) = \lim_{n\to\infty} \frac{k}{n} . \qquad (1)$$

According to Richard von Mises[1] this equation can be considered as definition of a random event; for physicists this is often the practical definition.

Both definitions can be critisized: the first one can be considered as a vicious circle, and the second one is based on an impossible experiment.

Formal definition. Elementary events e_i are considered, which *exclude* each other:

$$\begin{aligned} e_i &= \text{elementary event} \\ \Omega &= \text{set of all elementary events} \\ P(e_i) &= \text{probability of event } e_i \end{aligned}$$

The probability for elementary events has to follow the following rules:

1. $P(e_i) \geq 0$ for all i
2. $P(e_i \vee e_j) = P(e_i) + P(e_j) \qquad (\vee = \text{or})$
3. $\sum_\Omega P(e_i) = 1$

Non-elementary events $\mathcal{A}$ and $\mathcal{B}$ are formed from elementary events.

Combination of probability. Given two types of events, $\mathcal{A}$ and $\mathcal{B}$. $P(\mathcal{A})$ is the probability for the occurence of event $\mathcal{A}$ and $P(\mathcal{B})$ is the probability for the occurence of event $\mathcal{B}$. The following equations give the probability for the occurence of $\mathcal{A}$ **or** $\mathcal{B}$ and of $\mathcal{A}$ **and** $\mathcal{B}$.

$$\begin{aligned}
P(\mathcal{A} \text{ or } \mathcal{B}) &= P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \text{ and } \mathcal{B}) && \mathcal{A} \text{ or } \mathcal{B} \\
P(\mathcal{A} \text{ and } \mathcal{B}) &= 0 && \text{if } \mathcal{A} \text{ and } \mathcal{B} \text{ exclude each other} \\
P(\mathcal{A} \text{ or } \mathcal{B}) &= P(\mathcal{A}) + P(\mathcal{B}) && \text{if } \mathcal{A} \text{ and } \mathcal{B} \text{ exclude each other} \\
P(\mathcal{A} \text{ or } \overline{\mathcal{A}}) &= P(\mathcal{A}) + P(\overline{\mathcal{A}}) = 1 && (\overline{\mathcal{A}} = \text{not } \mathcal{A}) \\
P(\mathcal{A} \text{ and } \mathcal{B}) &= P(\mathcal{A}) \cdot\ P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B}) \cdot\ P(\mathcal{A}|\mathcal{B}) && \mathcal{A} \text{ and } \mathcal{B} \\
P(\mathcal{A} \text{ and } \mathcal{B}) &= P(\mathcal{A}) \cdot\ P(\mathcal{B}) && \text{if } \mathcal{A} \text{ and } \mathcal{B} \text{ independent}
\end{aligned}$$

$P(\mathcal{B}|\mathcal{A})$ is the probability for event $\mathcal{B}$, if event $\mathcal{A}$ has happened, and is called *conditional* probability. Bayes' theorem[2] relates the two conditional probabilities $P(\mathcal{A}|\mathcal{B})$ and $P(\mathcal{B}|\mathcal{A})$:

$$\boxed{P(\mathcal{A}|\mathcal{B}) = P(\mathcal{B}|\mathcal{A}) \cdot \frac{P(\mathcal{A})}{P(\mathcal{B})}} \, , \qquad (2)$$

which follows directly from the equation $P(\mathcal{A} \text{ and } \mathcal{B}) = P(\mathcal{A}) \cdot\ P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B}) \cdot\ P(\mathcal{A}|\mathcal{B})$ (see above). There is a special interpretation of the Bayes' theorem, with controversial discussions. Generalized to n classes of events with properties $\mathcal{A}_i$, $(i = 1, 2, \ldots n)$ and $P(\mathcal{B}) = \sum_i P(\mathcal{B}|\mathcal{A}_i) \cdot P(\mathcal{A}_i)$ the theorem reads

$$P(\mathcal{A}_i|\mathcal{B}) = \frac{P(\mathcal{B}|\mathcal{A}_i) P(\mathcal{A}_i)}{\sum_{i=1}^n P(\mathcal{B}|\mathcal{A}_i) \cdot P(\mathcal{A}_i)} \, . \qquad (3)$$

[1] Richard von Mises, Mathematician (1883-1953)

[2] Thomas Bayes, Reverend, 1702-1761

1.2 Permutations

For n different elements there are P_n different posssibilities for an order of the objects, called **permutations**, with

$$\boxed{P_n = 1 \cdot 2 \cdot 3 \cdots n = n!}$$

($0! = 1! = 1$). An example for $n = 3$ and elements `a`, `b` and `c` is $P_3 = 3! = 6$ with the permutations:
`abc bca cab cba bac acb`.

Selection of k elements. The number of possibilites to select k elements out of total n different elements is given by

$$\boxed{P_n^k = \underbrace{n\,(n-1)\,(n-2)\cdots(n-k+1)}_{k \text{ factors}} = \frac{n!}{(n-k)!}} ,$$

if the order of the k elements is relevant. If the order of selection is irrelevant, one has to divide the number P_n^k by $k!$, the number of permutations; the result is called **combination**

$$\boxed{C_n^k = \frac{n\,(n-1)\,(n-2)\cdots(n-k+1)}{1 \cdot 2 \cdot 3 \cdots n} = \frac{P_n^k}{P_k} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k}} .$$

These number are also called Binomial coefficients, they are integers with properties

$$\binom{n}{0} = \binom{n}{n} = 1 \qquad \binom{n}{1} = n \qquad \binom{n}{m} = \binom{n}{n-m} .$$

For numerical calculations the form

$$C_n^k = \left(\frac{n}{k}\right) \cdot \left(\frac{n-1}{k-1}\right) \cdots \left(\frac{n-k+1}{1}\right)$$

is more suitable. The Binomial coefficients appear in the Binomial theorem

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

With $q = 1 - p$ the Bernoulli formula (binomial distribution) is obtained

$$(p+q)^n = (1)^n = \sum_{k=0}^{n} P_n(k) \qquad \text{with} \quad P_n(k) = \binom{n}{k} p^k \,(1-p)^{n-k}$$

The number of possibilities to select, from total n different elements, first k_1 elements, then k_2 elements from the remaining $n - k_1$ elements, then k_3 elements from the remaining $n - k_1 - k_2$ elements etc. is obtained by repeated application of the formula, with $n = k_1 + k_2 + k_3 + \ldots$

$$P_n^{k_1} = \frac{n!}{k_1!\,(n-k_1)!} \cdot \frac{(n-k_1)!}{k_2!\,(n-k_1-k_2)!} \frac{(n-k_1-k_2)!}{k_3!\,(n-k_1-k_2-k_3)!} \cdots = \frac{n!}{\prod_j k_j!}$$

This formula is used in statistical mechanics for the distribution of distinguishable particles.

1.3 Probability density

Random variables are considered. A random variable can take on, for example in repeated identical experiments or measurements, more than one numerical value. Random variables with continous values and with discrete values are distinguished.

Discrete random variables. A discrete variable k can take on a finite or infinite number of values k_i with $i = a, a+1, \ldots b$. The probability for the value k_i is $P(k_i)$ with the normalization

$$\sum_{i=a}^{b} P(k_i) = 1 .$$

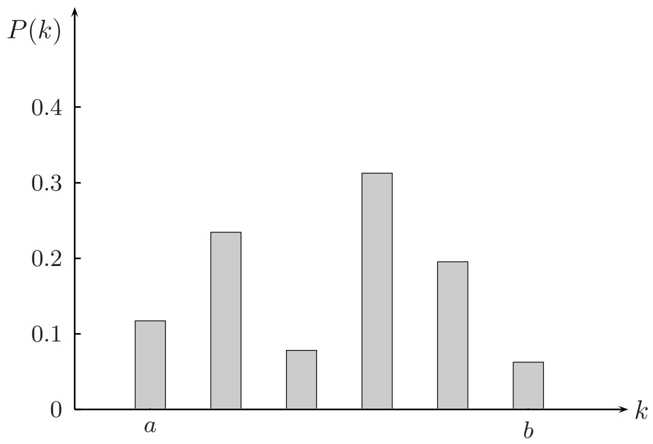

Figure 1: Distribution of a discrete random variable k within the limits a and b.

Continuous random variables. A continuous random variables x can take on values from a continuum. The probability for x to fall in the interval $a \leq x < b$ is

$$P\,(a \leq x < b) = \int_a^b f(x)\,\mathrm{d}x$$

where the function $f(x)$ is called *probability density* of the random variable x; the density is non-negative and is normalized to 1:

$$f(x) \geq 0 \qquad\qquad \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

The *distribution function* $F(x)$ is defined by

$$F(x) = \int_{-\infty}^{x} f(x')\,\mathrm{d}x' \qquad \text{with} \quad F(-\infty) = 0 \quad \text{and} \quad F(+\infty) = 1\ .$$

The distribution function and the probability density function are related also by $f(x) = \mathrm{d}F(x)/\,\mathrm{d}x$.

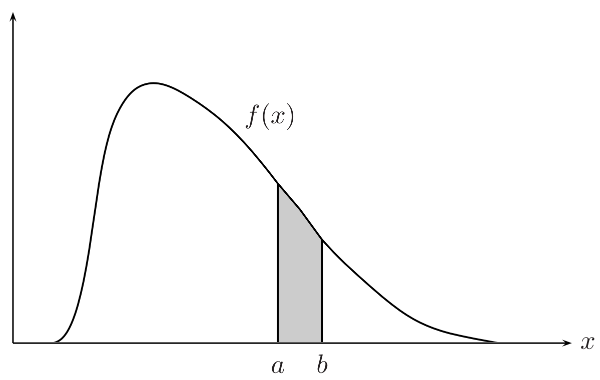

Figure 2: Distribution of a continuous random number x. The grey area corresponds to the probability $P(a \leq x \leq b)$.

1.4 Expectation values and moments

The *expectation value* of the function $h(x)$, where x is a random variable from a density $f(x)$, is defined by

$$E\,[h] = \int_{-\infty}^{+\infty} h(x)\,f(x)\,\mathrm{d}x \qquad\qquad E\,[h] = \sum_{i=a}^{b} h(r_i)\,P(r_i)\ .$$

Important special cases are the expectation value of x or k, or mean value, and the *variance* of x or k:

$$E[x] = \int_{-\infty}^{+\infty} x\,f(x)\,\mathrm{d}x = \langle x\rangle \qquad\qquad E[k] = \sum_{i=a}^{b} k_i P(k_i) = \langle k\rangle \tag{4}$$

$$V[x] = \int_{-\infty}^{+\infty} (x - \langle x\rangle)^2\; f(x)\,\mathrm{d}x \qquad\qquad V[k] = \sum_{i=a}^{b} (k_i - \langle k\rangle)\,P(k_i) \tag{5}$$

The expectation of a random variable x is often called μ, it is the standard parameter describing the "center of mass" of a distribution of a random variable. The variance of a random variable is non-negative; the square root is called the standard deviation σ and is the standard measure of the width of a distribution. The expectation is a linear operator (g and h are functions of x):

$$E\left[a \cdot g(x) + b \cdot h(x)\right] = a \cdot E\left[g(x)\right] + b \cdot E\left[h(x)\right]$$

For the variance operator

$$V\left[a \cdot g(x)\right] = a^2 \cdot V\left[g(x)\right]$$

is valid.

Moments. Algebraic moments are defined by the expectation value of powers of the random variable, central moments are defined with respect to the mean value:

$$\begin{array}{lcl} n\text{-th algebraic moment} & = & \alpha_n = E[x^n] \\ n\text{-th central moment} & = & m_n = E[(x - \langle x \rangle)^n] \end{array}$$

$$m_2 = \sigma^2 = V\left[x\right] = E\left[x^2\right] - \left(E\left[x\right]\right)^2$$

There is a unique relation between the density $f(x)$ and all the moments. Some moments and functions of moments have a special name:

$$\begin{array}{lcl} \text{Mean value = first algebraic moment} & = & E[x] = \mu = \alpha_1 \\ \text{Variance = second central moment} & = & V[x] = m_2 = \sigma^2 = \alpha_2 - \mu^2 \\ \text{Skewness} & = & \beta_1 = m_3^2/m_2^3 \\ \text{Coefficient of skewness} & = & \gamma_1 = \sqrt{\beta_1} = m_3/\sigma^3 \\ \text{Kurtosis} & = & \gamma_2 = m_4/m_2^2 - 3 \end{array}$$

For the normal distribution $\gamma_1 = \gamma_2 = 0$.

1.5 Tschebyscheff inequality

The probability for a deviation of a random varibale from the mean value μ by k times the standard deviation depends on the distribution. An upper limit is given, *independently* of the distribution, by the Tschebyscheff inequality:

$$\boxed{P\left(|x-\mu|^2 \geq k\sigma\right) \leq \frac{1}{k^2}}$$

The statement is rather weak, but general and useful in special cases.

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \langle x \rangle)^2 \cdot f(x)\,\mathrm{d}x = \left[\int_{-\infty}^{\langle x \rangle - k\sigma} + \int_{\langle x \rangle - k\sigma}^{\langle x \rangle + k\sigma} + \int_{\langle x \rangle + k\sigma}^{+\infty}\right] (x - \langle x \rangle)^2 \cdot f(x)\,\mathrm{d}x$$

All three integrals are positive and removing the middle one an inequality is obtained:

$$\sigma^2 \geq \int_{-\infty}^{\langle x \rangle - k\sigma} (x - \langle x \rangle)^2 \cdot f(x)\,\mathrm{d}x + \int_{\langle x \rangle + k\sigma}^{+\infty} (x - \langle x \rangle)^2 \cdot f(x)\,\mathrm{d}x\,.$$

Making use of the inequalities $x < \langle x \rangle - k\sigma$ bzw. $x > \langle x \rangle + k\sigma$ in the two remaining integrals the inequality

$$\sigma^2 \geq k^2\sigma^2 \left[\int_{-\infty}^{\langle x \rangle - k\sigma} f(x)\,\mathrm{d}x + \int_{\langle x \rangle + k\sigma}^{+\infty} f(x)\,\mathrm{d}x\right]\,.$$

is obtained. The expression in parentheses is the probability for $|x - \langle x \rangle| \geq k\sigma$, which proves the Tschebyscheff inequality.

1.6 Characteristic functions

Given a random variable with probability density $f(x)$, the characteristic function $\phi(t)$ is defined as the expectation of e^{itx}

$$\phi(t) = E\left[e^{itx}\right] = \int_{-\infty}^{+\infty} e^{itx} \cdot f(x)\,\mathrm{d}x \tag{6}$$

with the inverse operation

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \cdot \phi(t)\,\mathrm{d}t\,.$$

The characteristic function $\phi(t)$ is essentially the Fourier transform of the probability density $f(x)$. The characteristic function as well as the first and second derivative for the special value $t = 0$ are easily calculated:

$$\begin{aligned}
\phi(0) &= 1 \quad \text{(Normalization)} \\
\left.\frac{d\phi(t)}{dt}\right|_{t=0} &= \left.\int_{-\infty}^{+\infty} ixe^{itx} f(x)\mathrm{d}x\right|_{t=0} = i\langle x\rangle \\
\left.\frac{d^2\phi(t)}{dt^2}\right|_{t=0} &= -\left.\int_{-\infty}^{+\infty} x^2 e^{itx} f(x)\mathrm{d}x\right|_{t=0} = -E\left[x^2\right] = -\left(\sigma^2 + \langle x\rangle^2\right)
\end{aligned}$$

or more general for the n-th derivative:

$$\left.\frac{d^n\phi(t)}{dt^n}\right|_{t=0} = i^n \int_{-\infty}^{+\infty} x^n f(x)\mathrm{d}x = i^n \alpha_n$$

(α_n is the n-th algebraic moment of x). Thus it may be simple to calculate all the moments of a distribution $f(x)$ defined by the characteristic function $\phi(t)$.

The characteristic function for a random variable x can be expanded in a Taylor series by

$$\phi(t) = \sum_{n=0}^{\infty} \left.\frac{\mathrm{d}^n\phi}{\mathrm{d}t^n}\right|_{t=0} \frac{t^n}{n!} = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} E\left[x^n\right] \; . \tag{7}$$

In the special case of a mean value of 0 the first few terms are

$$\phi(t) = 1 - \frac{1}{2!}t^2\sigma^2 - \frac{i}{3!}t^3 E\left[x^3\right] \ldots \tag{8}$$

The characteristic function $\phi_z(t)$ for a *sum* $z = \sum_i x_i$ of n independent random variables $x_1, x_2 \ldots x_n$ with probability densities $f_1(x_1)$, $f_2(x_2) \ldots f_n(x_n)$ with characteristic functions $\phi_1(t)$, $\phi_2(t) \ldots \phi_n(t)$ is given by the *product* of the individual characteristic functions:

$$\boxed{\phi_z(t) = \phi_1(t) \cdot \phi_2(t) \cdots \phi_n(t)} \tag{9}$$

This relation follows simply from the definition of the characteristic function:

$$\begin{aligned}
\phi_z(t) = \int \ldots \int \exp\left(it \sum_{i=1}^{n} x_i\right) f_1(x_1)\, f_2(x_2) \cdots f_n(x_n)\, \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n \\
= \int e^{itx_1} f_1(x_1)\,\mathrm{d}x_1 \cdots \int e^{itx_n} f_n(x_n)\,\mathrm{d}x_n
\end{aligned} \tag{10}$$

The normal distribution. For the normal distribution the density $f(x)$ and the characteristic function $\phi(t)$ are given by

$$f(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) \mathrm{d}x \qquad \phi(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right) \tag{11}$$

The addition theorem (9) can be used to prove, that the sum $z = x + y$ of two random numbers x and y from the normal distribution with mean values μ_x and μ_y and with standard deviations σ_x and σ_y is a random number with mean $\mu_x + \mu_y$ and standard deviation $\sqrt{\sigma_x^2 \sigma_y^2}$, which again follows the normal distribution:

$$\begin{aligned}
\phi_z(t) = \phi_x(t) \cdot \phi_y(t) = \exp\left(i\mu_x t - \frac{1}{2}\sigma_x^2 t^2\right) \cdot \exp\left(i\mu_y t - \frac{1}{2}\sigma_y^2 t^2\right) \\
= \exp\left[i\left(\mu_x + \mu_y\right)t - \frac{1}{2}\left(\sigma_x^2 + \sigma_y^2\right)t^2\right]
\end{aligned} \tag{12}$$

1.7 The central limit theorem

In the central limit theorem n independent random variables x_i with mean values μ_i and variances σ_i^2 are considered. The central limit theorem states:

> In the limit of large n the sum $\sum_i x_i$ becomes a random variable with mean $\sum_i \mu_i$ and with variance $\sum_i \sigma_i^2$, which follows the *normal distribution*, *regardless of the distribution* of the individual variables x_i.

For the proof the case of equal variances σ_i^2 is assumed, with fixed value of n, and the transformed sum

$$z = \sum_{i=1}^{n} \frac{x_i - \mu_i}{\sqrt{n}} \tag{13}$$

is considered; each term in the sum has a mean value of zero and a variance of σ^2/n. The characteristic function of the sum z can, according to equation (8), be expressed in terms of a Taylor expansion:

$$\phi_z(t) = \prod_{i=1}^{n} \phi_i(t) = \prod_{i=1}^{n} \left(1 - \frac{1}{2!}\, t^2 \frac{\sigma^2}{n} - \frac{i}{3!}\, t^3 \frac{E\left[(x_i - \mu_i)^3\right]}{n^{3/2}} + \ldots \right) . \tag{14}$$

Neglecting terms with the third and higher moments in the limit of large n the characteristic function of the sum z is

$$\phi_z(t) = \prod_{i=1}^{n} \phi_i(t) \approx \prod_{i=1}^{n} \left(1 - \frac{1}{2!}\, t^2 \frac{\sigma^2}{n} \right) = \left(1 - \frac{1}{2!}\, t^2 \frac{\sigma^2}{n} \right)^n \tag{15}$$

for which in the limit $n \to \infty$

$$\lim_{n\to\infty} \left(1 - \frac{1}{2!}\, t^2 \frac{\sigma^2}{n} \right)^n = \exp\left(-\frac{1}{2}\, \sigma^2 t^2 \right) \tag{16}$$

follows; this is the characteristic function of the normal distribution with mean zero and variance σ^2.

$$\lim_{n\to\infty} \left(1 + \frac{u}{n} \right)^n = \lim_{n\to\infty} \left(1 + \binom{n}{1} \frac{u}{n} + \binom{n}{2} \left(\frac{u}{n} \right)^2 + \ldots \right) = \lim_{n\to\infty} \left(1 + u + \frac{n-1}{n} u^2 + \ldots \right) = e^u \tag{17}$$

2 Monte Carlo methods

2.1 Generation of uniform random numbers

Basis for Monte Carlo simulations are random numbers, distributed uniformly between the limits 0 and 1. The distribution is called $U(0,1)$ and has the density $g(u)$ and the distribution $G(u)$

$$g(u) = \begin{cases} 1 & \text{if } 0 \leq u < 1 \\ 0 & \text{else} \end{cases} \qquad G(u) = \begin{cases} 0 & \text{if } u < 0 \\ u & \text{if } 0 \leq u < 1 \\ 1 & \text{if } u \leq 1 \end{cases} . \tag{18}$$

A generation of statistically uniform numbers u_i on a digital computer in a sequence $\ldots u_i,\, u_{i+1},\, u_{i+2} \ldots$ of independent values is principally impossible. However generators exist for so-called pseudo random numbers, which pass all tests for statistical independence; they have very long periods and can be used in practice for accurate simulation calculations.
In addition to the uniform random numbers very often random numbers z from the standardized normal distribution $N(0,1)$ with mean 0 and variance 1 are needed. For random numbers the notation

$$\begin{array}{lcll} u_i & \in & U(0,1) & \text{uniform on (0,1)} \\ z_i & \in & N(0,1) & \text{standardized normal distribution} \end{array}$$

is used.
Linear congruential generator. A very simple generator for random numbers u is the general linear congruent generator, based on the recurrence relation

$$I_j = (a \cdot I_{j-1} + c) \quad \mathrm{mod}\ m$$

for the generation of integer random numbers between 0 and $m-1$. This generator needs three parameters, the factor a and the increment b, and m, called modulus. Starting from an initial value I_0, called *seed*, a sequence I_1, $I_2, I_3, \ldots$ is generated. The recurrence will repeat itself after a certain index, which is called the *period*, which is not greater than m. The integer I_j is converted to a real number $u = I_j/m$, which can take values between 0 and 1, ocasionally the value 0.0, but never the exact value 1.0. This range is valid for many generators; the exact value 0.0 is often excluded to avoid arithmetic problems (for example in $\ln u$).
The maximum value m of the period is reached under the following conditions:

1. increment $c \neq 0$;
2. the increment c and the modulus m do not have a common divisor;
3. $(a-1)$ is a multiple of the prime number p for each prime number p, which is a divisor of the modulus m;
4. $(a-1)$ is a multiple of 4, if modulus m is a multiple of 4.

Linear congruential generator are rather fast, but have a number of deficiencies. The low-order bits are much less "random" than their high-order bits. More important are sequential correlations: if d subsequent random number are plotted in d dimensional space, they will lie on $(d-1)$-dimensional planes, and there will be at most $m^{1/d}$ such planes, often many fewer than that, depending on the choice of parameters. The table below gives a few set of parameters, which are "good"; they pass the spectral test of D. Knuth for 2 to 6 dimensions and have the maximum period of m. They may be used for simple problems, where certain operations have to be randomized. There are sets of parameters with larger periods, which however may have problems with "integer overflow".

a	205	281	1021	421
c	29573	28411	25673	54773
m	139968	134456	121500	259200

Other random number generators. For serious Monte Carlo simulations a well-tested random-number generator should be used. Lagged-Fibonacci based generators are efficient and have a long period. RANLUX, based on the method of Lüscher, has a solid theoretical basis in chaos theory. Another well-tested generator called `ranshi` is by Gutbrod.
Good random-number generators make use of additional techniques. A large improvement is to combine two generators: an array is filled with random numbers from one generator, and a second generator is used to select random numbers.

Random number generators are often realized as a function like `u=ran()` with no argument. The function call for every rundom number means some overhead (call/return), which is avoided in subprograms which generate a whole array of random numbers.
Usually generators are initialized with a certain number, called *seed*, which determines the sequence of random numbers. In repeated calculations of a problem it is essential to initialize each calculation with a different seed.

2.2 Tests for random number generators

A basic test for uniform random numbers is filling a histogram with a fixed narrow bin width. Statistical fluctuations within each bin around the expected mean value of n should be in agreement with the Poisson statistic with a standard deviation of $\sqrt{n}$.
Sensititive to correlations are plots of random number u_j versus the next number u_{j+1}. Even more sensitive is this plot, if the preceeding or following number is restricted, for example to the range > 0.99.
In the gap test two numbers α and β defining a small gap are used: $0 \leq \alpha < \beta \leq 1$. A large number of random numbers is generated. The probability to have among $k+1$ random numbers the first k numbers outside and the $k+1$.st number inside the interval (α, β), is

$$P_k \; = \; p(1-p)^k \quad \text{with } p = \beta - \alpha \, .$$

according to the Binomial distribution.
Up-down- or sequence test. In this test a large number of $n+1$ random numbers u are generated. Two subsequent random numbers u_j and u_{j+1} are compared: the bit 0 or bit 1 is assigned to the cases $u_j < u_{j+1}$ and $u_j > u_{j+1}$. The number of sequences $N(k)$ of 0's and 1's of length k is determined from a very large sequence of n random numbers; the numbers $N(k)$ satisfy the conditions

$$\sum_{k=1}^{n} k \cdot N(k) = n \qquad \sum_{k=1}^{n} N(k) = \frac{(2n-1)}{3} \, .$$

For truely uncorrelated random numbers u_j the following values are expected:

$$N(1) \; = \; \frac{5n+1}{12} \qquad N(2) = \frac{11n-14}{60} \tag{19}$$

$$N(k) \; = \; \frac{2\left[\left(k^2+3k+1\right)n - \left(k^3+3k^2-k-4\right)\right]}{(k+3)!} \, . \tag{20}$$

Especially the numbers $N(k)$ from larger values of k allow sensitive tests of the random number generator.

2.3 Generators for general random numbers

In this section general methods for the generation of random numbers x from a density function $f(x)$ and (cumulative) distribution $F(x)$ are described.
Inverse transform method for a continuous random variable. If x is chosen with density $f(x)$, then the cumulative probability $F(x)$ is itself a random number which will occur with uniform density $g(u)$ on $[0,1]$:

$$f(x)\mathrm{d}x \; = \; g(u)\,\mathrm{d}u \qquad \int_{-\infty}^{x} f(x')\,\mathrm{d}x' \; = F(x) \; = u.$$

Ignoring the endpoints, the equation $u = F(x)$ is solved for the variable x by

$$x = F^{(-1)}(u)$$

with the inverse function $F^{(-1)}(u)$, provided the inverse function can be found. The method is illustrated in Figure 3.
The transformation can be used for the test of empirical distributions, e.g. χ^2-distributions. If a sample of x-values is assumed to follow a certain density $f(x)$, then the transformation $u = F(x)$ should give a sample of u-values, which is uniform between 0 and 1.
Inverse transform method for a discrete random variable. In a discrete distribution an index k is assigned to the discrete random values with $k = 1, 2, \ldots$. The discrete values x_k occurs with probability $P(x_k)$. To generate a random value x_k a random number $u \in U(0,1)$ is chosen. Then one has to find the index k with

$$P_{k-1} < u \leq P_k = \sum_{i=1}^{k} P(x_i) = P_{k-1} + P(x_k) \; ,$$

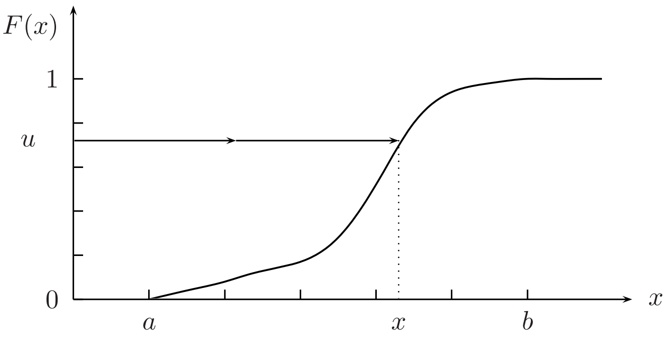

Figure 3: Generation of random numbers from a continuous distribution $f(x)$. The curve representing the cumulative distribution $F(x)$ is shown. A random number $u \in U(0,1)$ is transformed to the random number $x = F^{-1}(u)$.

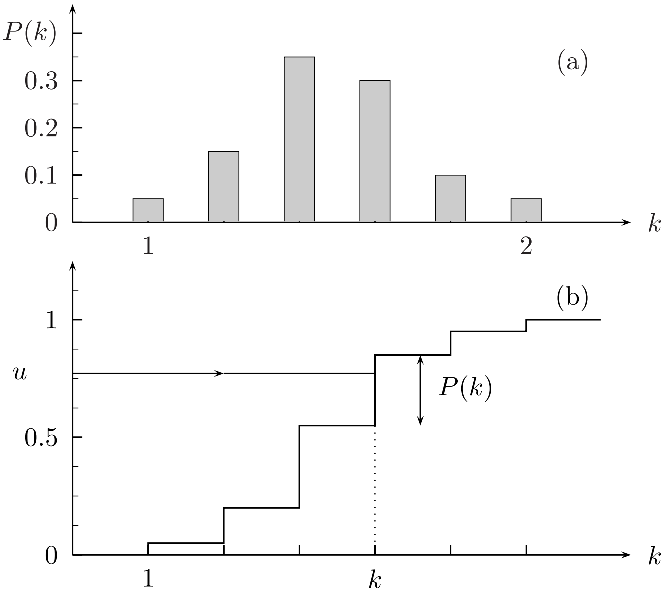

Figure 4: Generation of random numbers of a discrete distribution according to a histogram (a). The cumulative distribution (b) is used to transform a random number $u \in U(0,1)$ to the discrete random number k.

where $P_0 = 0$; for a finite discrete distribution with n possible values $P_n = 1$. The algorithm is illustrated in Figure 4.

Acceptance-rejection method. Obtaining the inverse function $F^{(-1)}(u)$ is often impossible, although the density function can be computed for any possible value of x. The method due to Neumann is based on the relation

$$f(x) \leq C \cdot h(x) \qquad \text{for all } x \ ,$$

where factor C is a (positive) constant and where random values x for the density $h(x)$ are easily generated. The algorithm is as follows:

1. Choose a random number x according to distribution $h(x)$;
2. generate u;
3. if the product $u\,C\,h(x) > f(x)$, reject x and repeat step 1; otherwise accept x.

Because both $f(x)$ and $h(x)$ are normalized, the factor $C > 1$; the acceptance probability in the last step is equal to the ratio of the areas, which is $1/C < 1$. The factor C will be not be much larger than 1, if the shape of $h(x)$ is similar to the shape of $f(x)$; in this case the efficiency of the algorithm will be close to 100 %.

The distribution $h(x)$ may be even a uniform distribution between the limits a and b of x and in this case one has to choose the constant C such that $f(x) \leq C$ for all x. Then the algorithm is:

1. Choose x_i uniformly distributed between a und b: $\quad x = a + u_1 \cdot (b - a)$ with $u_1 \in U(0,1)$;
2. choose another random number $u_2 \in U(0,1)$.
3. if the product $u_2 \cdot C > f(x)$, goto step 1, otherwise accept x.

2.4 Monte Carlo integration and numerical integration

Onedimensional integration. Consider a function $g(x)$ depending on one variable x. The integral of the function $g(x)$ between the limits a and b is given by

$$\mathrm{I} = \int_a^b g(x)\,\mathrm{d}x$$

If the integral can not be evaluated by analytical methods, one can estimate the value by numerical and by Monte Carlo methods.

The simplest *numerical* method is the trapezoidal rule with the formula

$$I = \int_a^b g(x)\,\mathrm{d}x = \frac{b-a}{n}\sum_{j=0}^{n} w_j g(x_j) + R_n \qquad R_n = -\frac{1}{12}\frac{(b-a)^3}{n^2}\tilde{g}''$$

with the notation $x_j = a + j\,(b-a)/n$, and $w_0 = w_n = 1/2$; $w_j = 1$ for $1 \le j \le (n-1)$. The numerical error estimate R_n is proportional to some average value of the second derivative in the interval (a,b) and proportional to $1/n^2$.

In the Monte Calo method in its simplest form the integral I can be expressed by the mean value of $g(x)$, written as $\overline{g}$, in the interval (a,b):

$$I = \int_a^b 1 \cdot g(x)\,\mathrm{d}x = (b-a)\overline{g} = (b-a)\,E[g(x)]\ ,$$

assuming a *random* variable x with an uniform distribution in the x-range. The Monte-Carlo estimate I_{MC} for the integral I is given by

$$I_{\mathrm{MC}} = \frac{b-a}{n}\sum_{i=1}^{n} g(x_i) = \frac{b-a}{n}\sum_{i=1}^{n} g_i\ ,$$

where the x-values x_i are generated from a uniform density in (a,b); if u_i are random numbers from the $U(0,1)$-distribution, then $x_i = a + u_i(b-a)$. The accuracy of the integral I_{MC} can be estimated with standard statistical methods. The variance of I_{MC} is given by

$$V[I_{MC}] = \sigma_I^2 = V\left[\frac{b-a}{n}\sum_{i=1}^{n} g_i\right] = \left(\frac{b-a}{n}\right)^2 V\left[\sum_{i=1}^{n} g_i\right] = \frac{(b-a)^2}{n} V\,[g_i]\ . \tag{21}$$

The standard deviation has, in agreement with the law of large numbers, a dependence $\propto 1/\sqrt{n}$. The variance of I_{MC} is also proportional to the variance $V\,[g_i] = \sigma_g^2$ of the function $g(x)$,

$$\sigma_g^2 = V[g] = E\left[(g(x) - \overline{g})^2\right]$$

where $\overline{g}$ is the mean function value, equal to $I/(b-a)$. The variance of g will be a large factor for a function with large deviation from the mean, which in practice is not easily calculated. The Monte Carlo estimate of this variance is given by

$$V\,[g_i] = E\left[(g_i - \bar{g})^2\right] \approx \left[\frac{\sum g_i^2}{n} - \left(\frac{\sum g_i}{n}\right)^2\right]\ . \tag{22}$$

This gives the final result for the variance of the estimate I_{MC}:

$$V[I_{MC}] = \sigma_I^2 = \frac{(b-a)^2}{n}\left[\frac{\sum g_i^2}{n} - \left(\frac{\sum g_i}{n}\right)^2\right] = \frac{1}{n}\left[(b-a)^2\frac{\sum g_i^2}{n} - I^2\right]\ . \tag{23}$$

A comparison of the two methods shows, that the numerical method even in its simplest form method has a much faster convergence ($1/n^2$ for the trapezoidal rule), and is superior to the Monte Carlo method with a $1/\sqrt{n}$-dependence.

Integration in d dimensions. Often multi-dimensional integrals occur, which have to be evaluated numerically. The one-dimensional integration formulae can be extended to formulae for d-dimensional integration. For simplicity an integral over the d-dimensional unit hypercube $(0,1)^d$ is assumed. The trapezoidal rule in d dimensions, obtained by applying the one-dimensional integration rule with $n+1$ points in each dimension, is

$$I = \int_\Omega g(x_1,\, x_2, \ldots x_d)\,\mathrm{d}\boldsymbol{x} = \frac{1}{n^d} \sum_{j_1=0}^{n} \sum_{j_2=0}^{n} \cdots \sum_{j_d=0}^{n} w_{j_1} w_{j_2} \cdots w_{j_d} \cdot g\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots \frac{j_d}{n}\right) + \mathcal{O}\left(\frac{1}{n^2}\right) \,,$$

where the function $g(x_1,\, x_2, \ldots x_d)$ has to be evaluated $N = (n+1)^d \approx n^d$ times. Although the deviation is still proportional to $1/n^2$, the dependence on the number N of function evaluations is $N^{-2/d}$. The deviation in the Monte Carlo method is $N^{-1/2}$, independent of dimension, and thus becomes superior compared to the trapezoidal rule for $d > 4$. Although a more accurate rule in each dimension improves the dependence, perhaps to $N^{-4/d}$, the method of choice for numerical integration in high dimensions is the Monte Carlo method; it has the additional advantage, that complicated integration boundaries are simpler for the Monte Carlo method than for numerical methods.
However, the convergence of the Monte Carlo estimate I_{MC} by a rate of $1/\sqrt{N}$ is slow; it depends on the variance of the function and techniques to reduce the variance are important.

2.5 Variance reducing methods

Importance sampling. The method corresponds to a change of the variable. The integral of the function $g(x)$ in the intervall (a, b) is written in the form

$$I = \int_a^b g(x)\,\mathrm{d}x = \int_a^b \left[\frac{g(x)}{f(x)}\right] f(x)\,\mathrm{d}x = E\left[\frac{g(x)}{f(x)}\right] \,,$$

with a probability density $f(x)$, normalized in the interval (a, b). Assuming that the density $f(x)$ can be integrated ($F(x)$) and the inverse $F^{(-1)}(u)$ can be found, the expectation value of the function $r(x) = g(x)/f(x)$ for probability density $f(x)$ can easily be determined with Monte Carlo methods.

$$I_{\mathrm{MC}} = \frac{1}{n} \sum_{i=1}^{n} \frac{g(x_i)}{f(x_i)} \quad \text{with} \quad x_i \in (a,\, b) \quad \text{from density } f(x)\,.$$

Importance sampling improves the accuracy of a Monte Carlo integration, if a density $f(x)$ can be found such that the variance of $r(x) = g(x)/f(x)$

$$V\,[r] = E\left[(r(x) - \overline{r})^2\right]$$

is small.

Control variates. In this method for variance reduction a function $h(x)$ is subtracted from the function $g(x)$, which can be integrated analytically:

$$\int_a^b g(x)\,\mathrm{d}x = \int_a^b h(x)\,\mathrm{d}x + \int_a^b \left[g(x) - h(x)\right]\,\mathrm{d}x \,,$$

The first integral is calculated analytically, the second one by Monte Carlo methods. The function $h(x)$ should be chosen to reduce the variance of the second integral.

Stratified sampling. The full integration space is divided into subspaces and a Monte Carlo integration is performed in each subspace; the partial results are added up in the end. The method is based on the property

$$\int_a^b g(x)\,\mathrm{d}x = \int_a^c g(x)\,\mathrm{d}x + \int_c^b g(x)\,\mathrm{d}x \qquad\qquad a < c < b$$

for the Riemann-Integral. For a d-dimensional integral the integration region is divided into k subregions M_j with $j = 1,\, 2, \ldots k$. A subregion has a volume $V(M_j)$ and in each subregion the Monte Carlo integration is done with N_j points. The Monte Carlo estimate of the complete integral is

$$I_{\mathrm{MC}} = \sum_{j=1}^{k} \frac{V(M_j)}{N_j} \sum_{i=1}^{N_j} g(\boldsymbol{x}_{ji}) \,. \tag{24}$$

The variance of the integral is, instead of $\sigma^2(g)/N$, given by the expression

$$V[I_{\mathrm{MC}}] = \sum_{j=1}^{k} \frac{V^2(M_j)}{N_j} \left.\sigma^2(g)\right|_{M_j} , \tag{25}$$

where the $\left.\sigma^2(g)\right|_{M_j}$ are the variances in the subregions. If the subregions and the numbers N_j are carefully chosen, the variance can be reduced compared to the simple Monte Carlo integral. The variances in the subregions should be small, and the number of points in each subregion should be proportional to the product of volume $V(M_j)$ and standard deviation $\left.\sigma(g)\right|_{M_j}$ in each subregion.

In adaptive Monte Carlo methods the subregions are optimized during the calculation, and the number of points N_j in each subregion is optimized.

3 Distributions

3.1 Binomial distribution

If the probability for a certain event is equal to p, the probability to observe in n experiments exactly k of the events, is given by the Binomial distribution:

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots n$$

In a single event there are two possible outcomes, one (often called "success") with probability p and another (often called "failure") with probability $q = 1 - p$. Examples of Binomial distribution are shown in Figure 5. The factor $\binom{n}{k}$ ist the number of possibilities to select k elements out of total n different elements.

$$\text{Expectation value } = E[k] = n\,p \qquad \text{Variance } = V[k] = n\,p(1-p) \tag{26}$$

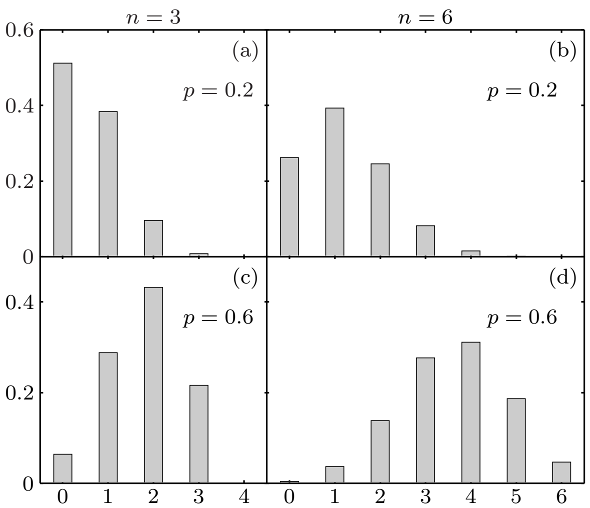

Figure 5: Binomial distributions for $n = 3$ (left) and $n = 6$ (right) with probabilities $p = 0.2$ (top) and $p = 0.6$ (bottom).

Estimate: If for an unknown value of p the values n and k are known from an experiment, the unknown value of the probability p can be estimated by $h = k/n$. An unbiased estimate of the variance of h is given by

$$V(h) = \frac{1}{n(n-1)} N \left(\frac{k}{n}\right)\left(1 - \frac{k}{n}\right)$$

The law of large numbers. The accuracy of an estimate h of the probability in an Binomial process increases with n.

$$V[h] = V[k/n] = \frac{1}{n^2} V[k] = \frac{1}{n^2} n\,p\,(1-p)$$

Because of $p\,(1-p) \leq 1/2$ for any value of p, one can estimate

$$V[h] \leq \frac{1}{4n} \qquad \sigma(h) \propto \frac{1}{\sqrt{n}}$$

The error of the estimate h can be made as small as required by choosing a sufficiently high value of n.

Random number generation. Parameters of the Binomial distribution are n and p. An algorithm for small n is as follows.

1. Set $k = 0$ and $m = 0$;
2. generate a random number u; if $u \leq p$, increase k by one: $k := k + 1$;
3. increase m by one: $m := m + 1$; if $m < n$ goto step 2, otherwise accept k.

The following is an algorithm which is faster for large n and is to be used for the case $p \leq 1/2$.

1. Set $k = 0$ and set $q = 1 - p$;
2. compute $P_k = q^n$ and store in B;
3. generate a random number u;
4. if $u \leq B$ accept k and stop;
5. increment k by 1; if $k = n$, accept k and stop

6. compute next P_k by
$$P_k := P_k \frac{p}{q} \frac{n-k+1}{k} \; ,$$
add to B and go to step 4;

For the case $p > 1/2$ the value of p should be replaced by $1-p$ and the resulting random value k has be replaced by $n-k$.

3.2 Multinomial distribution

The multinomial distribution is the generalization of the binomial distribtion to the case of m different possible outcomes. There are n (non-negative) probabilities p_i with the constraint $\sum_i p_i = 1$. The probability to obtain k_1 events of type 1, k_2 events of type 2, $\ldots k_m$ events of type m for total $n \; (= \sum_i k_i)$ events is given by
$$P(k_1, \ldots, k_m; n. p_1, \ldots, p_m) = \frac{n}{k_1! k_2! \cdots k_m!} p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m}$$
The factor is equal to the number of possibilities to select from total n elements k_1 elements of type 1, k_2 elements of type 2 ... and k_m elements of type m.
$$E[k_i] = n\, p_i \qquad V[\boldsymbol{k}]_{ij} = \begin{cases} -np_i p_j & i \neq j \\ np\,(1-p_i) & i = j \end{cases} \tag{27}$$
Histogram. A histogram with a *fixed* total number n of entries distributed in m bins with k_i entries in the i-th bin is an example of the multinomial distribution; the number of events in any two bins i and j are negatively correlated.

If the observations k_i are independent, then the distribution of entries in bins is the Poisson distribution. The total number $n = \sum_i k_i$ is itself a random variable, correlated with the k_i. If the number n is not treated as a random variable (e.g. if the histogram is normalized), the numbers k_i must be considered as conditional on the observed value of n and the multinomial distribution is applicable. In the case $p_i \ll 1$ (histogram with many bins) the formula for the variance can be appproximated:
$$V[k_i] = n\, p_i\,(1-p_i) \approx n\, p_i \approx k_i \qquad \sigma_i \approx \sqrt{k_i}$$

3.3 Poison distribution

The Poisson distribution is the fundamental distribution for counting processes. It is obtained from the Binomial distribution in the limit $p \to 0$ and $n \to \infty$ with product np fixed, and depends on the single parameter $\mu \; (= np)$.
$$\boxed{P_\mu(k) = \frac{\mu^k}{k!}\, e^{-\mu} \qquad k = 0, 1, \ldots}$$
$$\text{Expectation value } = E[k] = \mu \qquad \text{Variance } = V[k] = \mu \tag{28}$$

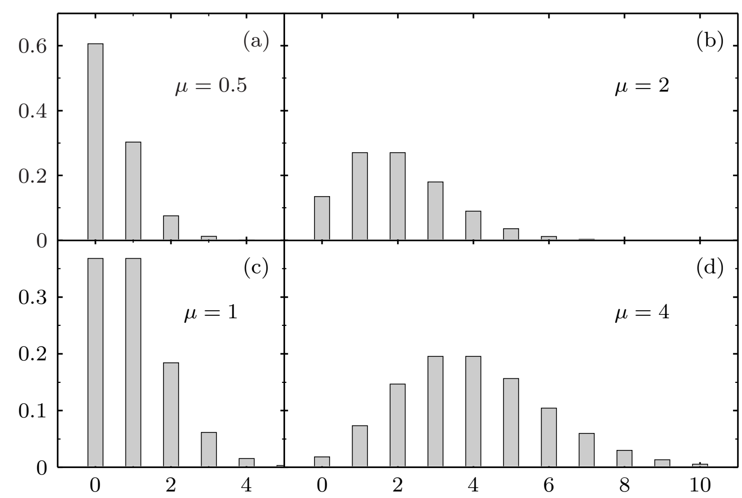

Figure 6: Poisson distributions with mean values $\mu = 0.5$, $\mu = 1$, $\mu = 2$ und $\mu = 4$.

Random number generation. The Poisson distribution is defined by one parameter μ.

1. Set $k := 0$ and $A := 1.0$;

2. generate a random number u and replace A by $u\,A$: $A := u\,A$; if $A \le e^{-\mu}$, accept k and stop
3. increase k by 1: $k := k + 1$ and goto step 2

For larger values of μ like $\mu = 100$ the method makes no sense, because $e^{-\mu}$ may be already outside the floating point range of the computer. Already for $\mu > 10 \ldots 20$ it is better to use the normal approximation of the Poisson distribution and to genererate a random variable z from $N(0, 1)$; the real random number is then converted to the random integer k by $k = max\left(0, \left[\mu + z\sqrt{\mu} + 0.5\right]\right)$.

3.4 Uniform distribution

The density of the uniform distribution $U(a, b)$ is given by

$$\boxed{f(x) = \begin{cases} 1/(b-a) & a \le x < b \\ 0 & \text{outside} \end{cases}}$$

$$\text{Expectation value } = E[x] = (a+b)/2 \qquad \text{Variance } = V[x] = (b-a)^2/12 \tag{29}$$

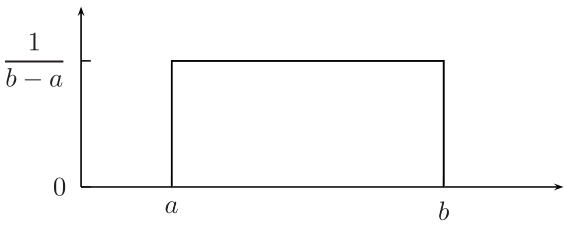

Figure 7: The uniform distribution with constant density between the limits a and b.

Of special importance is the uniform distribution $U(0, 1)$ (limits 0 and 1) with $E[x] = 1/2$ and $V[x] = 1/12$. Computer generated random numbers x from this distribution are used in Monte Carlo calculations.

Generation of a random azimutal angle. A uniformly distributed random azimutal angle is $\phi = 2\pi u$. The sine and the cosine of a uniformly distributed angle ϕ is generated without function computation by the following algorithm:

1. Generate u_1 and u_2 and $v_1 := 2u_1 - 1$, $v_2 := u_2$;
2. calculate $r^2 = v_1^2 + v_2^2$; if $r^2 > 1$, goto step 1;
3. calculate

$$\sin\phi := 2\frac{v_1 v_2}{r^2} \qquad \cos\phi := \frac{v_1^2 - v_2^2}{r^2} \tag{30}$$

Generation of a random direction in 3-dimensions. Isotropy in three dimmensions means a density proportional to the solid angle element $\mathrm{d}\Omega = \mathrm{d}(\cos\theta)\,\mathrm{d}\phi$. The azimutal angle ϕ is uniform (see above) and for the polar angle $\cos\theta := 2u - 1$ (uniform in $\cos\theta$) is calculated using an uniform random number u.

3.5 One-dimensional normal distribution

The density of the normal distribution $N(\mu, \sigma^2)$ is given by

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad x \in (-\infty, \infty)} \,. \tag{31}$$

$$\text{Expectation value} \quad E[x] = \mu \qquad \text{Variance} \quad V[x] = \sigma^2$$

The normal distribution $N(0, 1)$ with $E[x] = 0$ and $V[x] = 1$ is called the *standardized* normal distribution $N(0, 1)$.

The integrated distribution is

$$\Phi(z) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z} e^{-(x-\mu)^2/2\sigma^2}\,\mathrm{d}x = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{z-\mu}{\sqrt{2}\sigma}\right)\right) .$$

The last form gives the formula for the calculation using the so-called error function erf, defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,\mathrm{d}t$$

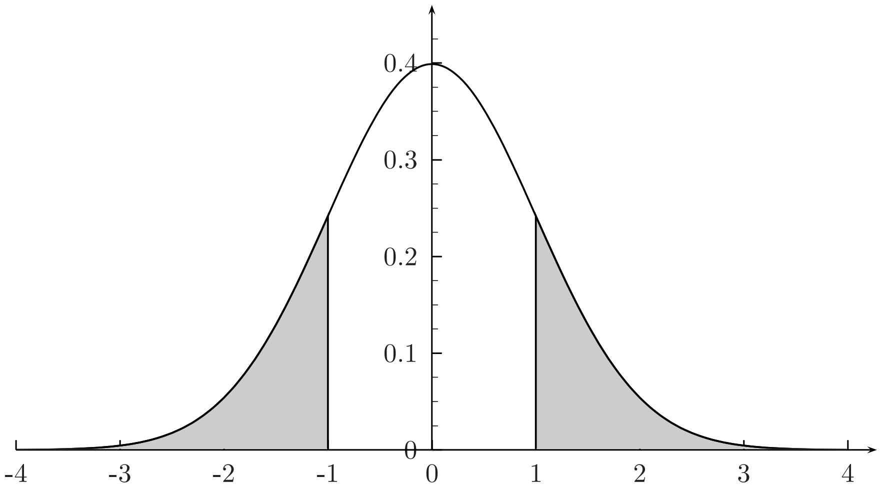

Figure 8: The standardized Gaussian distribution (mean value $\mu = 0$ and standard deviation $\sigma = 1$). The grey area shows the probability for $|x-\mu| \geq 1\sigma$; the probability is 31.74 %.

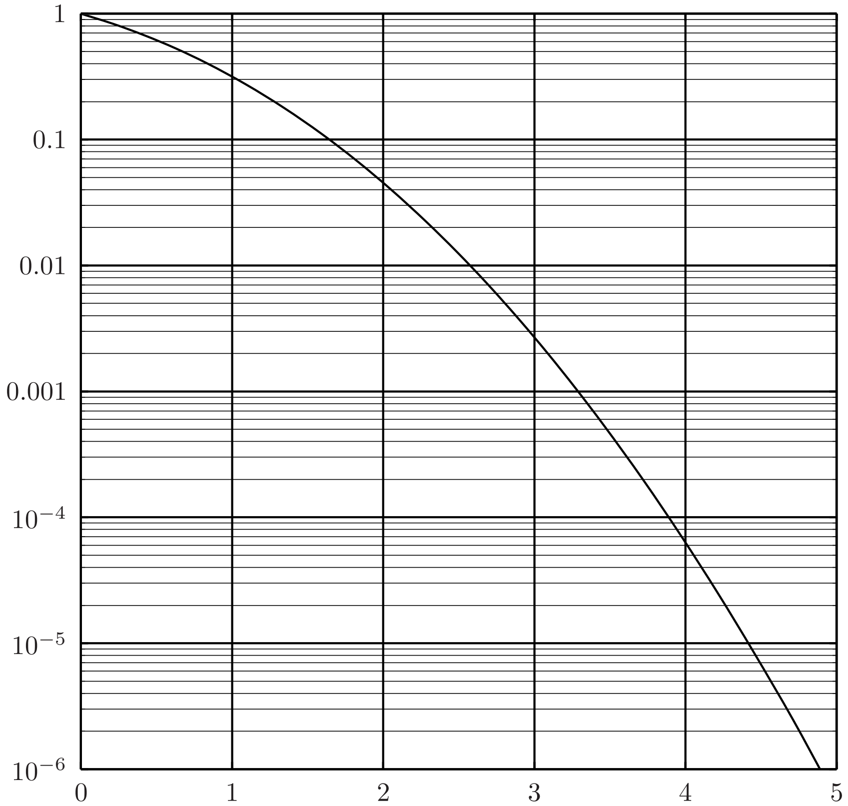

Figure 9: Probability for a deviation of more than k standard deviations for a Gaussian distribution, shown against the number k of standard deviations.

with $\mathrm{erf}(0) = 0$ and $\mathrm{erf}(\infty) = 1$.

Generation of random numbers. A simple approximate method is by adding 12 uniformly distributed random numbers $u_i \in U(0,1)$:

$$z = \sum_{i=1}^{12} u_i - 6$$

The sum z has a mean of 0 and a variance of $12 \cdot 1/12 = 1$.

In the following method two independent random variables from the standardized normal distribution are generated. The algorithm requires on ln- and one sqrt-call for two normal random numbers.

1. Generate two random numbers u_1 and u_2 and construct $v_1 := 2u_1 - 1$ and $v_2 := 2u_2 - 1$.
2. calculate $r^2 = v_1^2 + v_2^2$; if $r^2 > 1$, goto step 1;
3. calculate the two independent random numbers

$$z_1 = v_1 \sqrt{\frac{-2\ln r^2}{r^2}} \qquad\qquad z_2 = v_2 \sqrt{\frac{-2\ln r^2}{r^2}}$$

The random number z is transformed to a random number z' from the normal distribution $N(\mu, \sigma^2)$ with mean μ and standard deviation σ by $z' = \mu + \sigma z$.

3.6 Two-dimensional normal distribution

The density of the two-dimensional normal distribution for random variables x and y is given by

$$f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$
$$\left.\left[\left(\frac{x-\xi}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\xi}{\sigma_x}\right)\left(\frac{y-\eta}{\sigma_y}\right) + \left(\frac{y-\eta}{\sigma_y}\right)^2\right]\right\}\mathrm{d}x\,\mathrm{d}y\ . \tag{32}$$

From this density the two marginal distributions for x and y are obtained by integration over the other variable:

$$f_y(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\frac{1}{2}\left(\frac{x-\xi}{\sigma_x}\right)^2\right\} \tag{33}$$

$$f_x(y) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma_y}\exp\left\{-\frac{1}{2}\left(\frac{y-\eta}{\sigma_y}\right)^2\right\}\ . \tag{34}$$

Mean values and variances of the marginal distributions are ξ, η and ${\sigma_x}^2$, ${\sigma_y}^2$:

$$\langle x\rangle = \xi \qquad\qquad \langle y\rangle = \eta$$
$$V[x] = \sigma_x^2 \qquad\qquad V[y] = \sigma_y^2\ .$$

The correlation coefficient is $\rho_{xy} = \rho$. The conditional densities are

$$f(x|y_0) = \frac{f(x,y_0)}{f_x(y_0)} \tag{35}$$
$$= \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x-\xi}{\sigma_x} - \rho\frac{y_0-\eta}{\sigma_y}\right]^2\right\}$$

$$f(y|x_0) = \frac{f(x_0,y)}{f_y(x_0)} \tag{36}$$
$$= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{y-\eta}{\sigma_y} - \rho\frac{x_0-\xi}{\sigma_x}\right]^2\right\}\ .$$

The *conditional mean values* are

$$E[x|y_0] = \int_{-\infty}^{+\infty} x f(x|y_0)\,\mathrm{d}x = \xi + \rho\sigma_x\left(\frac{y_0-\eta}{\sigma_y}\right) \tag{37}$$

$$E[y|x_0] = \int_{-\infty}^{+\infty} y f(y|x_0)\,\mathrm{d}y = \eta + \rho\sigma_y\left(\frac{x_0-\xi}{\sigma_x}\right)\ . \tag{38}$$

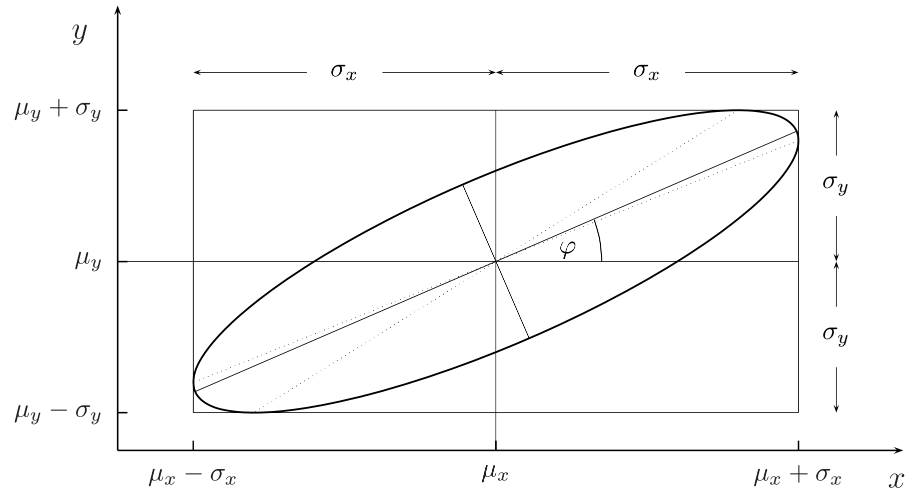

Figure 10: The elliptical 1σ contour of the two-dimensional normal distribution with mean values μ_x, μ_y and standard deviations σ_x, σ_y. The correlation coefficient is $\rho = +0.8$. Shown are the main axes of the ellipse with the angles φ und $\varphi + \pi/2$ given by $\tan 2\varphi = \pm(2\rho\sigma_x\sigma_y)/(\sigma_x^2 - \sigma_y^2)$ The dotted lines are the conditional means $x - \mu_x = (y - \mu_y)\rho\sigma_x/\sigma_y$ and $y - \mu_y = (x - \mu_x)\rho\sigma_y/\sigma_x$.

Random number generation. The generation of a random vector from the two-dimensional normal distribution is based on the Cholesky decomposition $\boldsymbol{V} = \boldsymbol{R}^T\boldsymbol{R}$ of the 2-by-2 covariance matrix $\boldsymbol{V}$

$$R_{11} = \sqrt{V_{11}} = \sigma_x \qquad R_{12} = \frac{V_{11}}{\sqrt{V_{11}}} = \rho\sigma_y \qquad R_{22} = \sqrt{V_{22} - R_{12}^2} = \sigma_y\sqrt{1-\rho^2}$$

as explained in general later. The transformation from two uncorrelated random numbers z_1 and z_2 from the standardized normal distribution to x and y is given by the equations

$$\begin{aligned} x &= \xi + \sigma_x z_1 \\ y &= \eta + \sigma_y \left(\rho z_1 + \sqrt{1-\rho^2} z_2\right) \end{aligned}$$

3.7 Multivariate normal distribution

The n-dimensional normal distribution is the generalization of the onedimensional normal distribution to the case of n dimensions. The random number is a n-vector $\boldsymbol{x}$, the density depends on the n-vector $\boldsymbol{\mu_x}$ of mean values and on the n-by-n covariance matrix $\boldsymbol{V_x}$. The density is given by the formula

$$\boxed{f(\boldsymbol{x}) = \frac{|\boldsymbol{W}|^{1/2}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_x})^T \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{\mu_x})\right)}\,, \tag{39}$$

with $\boldsymbol{W} = \boldsymbol{V}^{-1}$ and $|\boldsymbol{W}| = \det\ \boldsymbol{W} = |\boldsymbol{V}^{-1}|$. The n-dimensional normal distribution is often used as approximation to other n-dimensional distribution; the parameters of the n-dimensional normal distribution are just the first moment (mean $\boldsymbol{\mu}$) and the second central moment (covariance matrix $\boldsymbol{V}$).
Random number generation. A method is based on the representation of the (symmetric) covariance matrix $\boldsymbol{V}$ by the representation $\boldsymbol{V} = \boldsymbol{R}^T\boldsymbol{R}$ with a right triangular matrix $\boldsymbol{R}$, which is obtained by a Cholesky transformation. A random vector $\boldsymbol{z}$, where the elements z_i are independent random numbers from the standardized normal distribution, is transformed by

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{R}^T\boldsymbol{z}$$

to a random vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$.
Cholesky decomposition. The matrix $\boldsymbol{W}$ is transformed to the matrix $\boldsymbol{R}$ in steps. Starting with $\boldsymbol{W}^{(0)} = \boldsymbol{W}$ the transformation is

$$W_{ij}^{(k)} = W_{ij}^{(k-1)} - R_{ki}R_{kj} \qquad i,\, j = k+1,\, k+2,\, \ldots\, n\,. \tag{40}$$

for $k = 1, 2, \ldots (n-1)$. First $k = 1$ and the elements wih $i, j = 2, 3, \ldots n$ are transformed, the $k = 2$ and the elements with $i, j = 3, 4, \ldots n$ are transformed by

$$R_{ii} = \sqrt{W_{ii}^{(i-1)}} \qquad R_{ij} = \frac{W_{ij}^{(i-1)}}{R_{ii}} \quad \text{für} \quad j > i\,. \tag{41}$$

The relation

$$W_{jj} = \sum_{i=1}^{j} R_{ij}^2 \qquad j = 1, 2, \ldots n \tag{42}$$

shows that the elements of the Cholesky factors $\boldsymbol{R}$ are limited by the positive diagonal elements of the matrix $\boldsymbol{W}$.

3.8 Exponential distribution

The density is given by

$$\boxed{f(x) = \lambda e^{-\lambda x}}\,.$$

$$\text{Expectation} \quad E[x] = \frac{1}{\lambda} \qquad \text{Variance} \quad V[x] = \frac{1}{\lambda^2}$$

Consider events occuring randomly in time with a mean of λ events per unit time. According to the Poisson distribution the probability of no ($k = 0$) events in time interval t is $e^{-\lambda t}$.

Random number generation. The inverse transformation method can be used in this case (for $u \in U(0,1)$:

$$x = -\frac{1}{\lambda} \ln u$$

If values x have to be generated between limits a and b, the formula is

$$x = -\frac{1}{\lambda} \ln\left(a + (b-a)u\right)$$

3.9 Gamma distribution

The density of the gamma distribution is given by

$$\boxed{f(x|k) = \frac{x^{k-1} e^{-x}}{\Gamma(k)}}\,.$$

For integer values of k with $k \geq 1$ the density describes the waiting time $t \equiv x$ from the first event to the k-th event in a Poisson process with mean $\mu = 1$. The generalization for other values of μ is

$$\boxed{f(x|k,\mu) = \frac{x^{k-1} \mu^k e^{-\mu x}}{\Gamma(k)}}\,.$$

The parameter k determines the shape of the distribution, while μ is onl;y a scale parameter. For $\mu = 1/2$ und $k = n/2$ the gamma distribution agrees with the χ^2-distribution with n degrees of freedom.

3.10 χ^2 distribution

A sum u of n independently distributed random numbers x_i from the standardized normal distribution $N(0,1)$

$$u = \chi^2 = \sum_{i=1}^{n} x_i^2$$

follows a χ^2 distribution $f_n(u) \equiv f_n(\chi^2)$ with n degrees of freedom:

$$\boxed{f_n(u) \;=\; \frac{\frac{1}{2}\left(\frac{u}{2}\right)^{n/2-1} e^{-u/2}}{\Gamma(n/2)}}\,. \tag{43}$$

$$\langle u \rangle = \langle \chi^2 \rangle = n \qquad V\left[u\right] = V\left[\chi^2\right] = 2n\,. \tag{44}$$

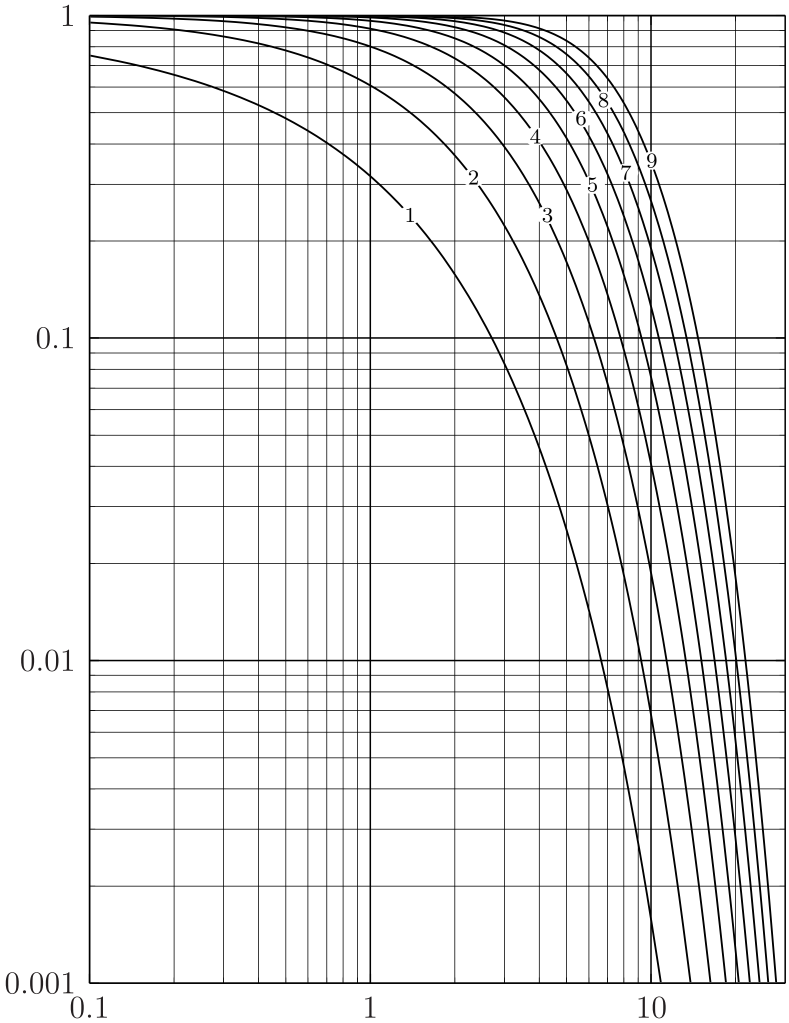

Figure 11: The quantity $1 - P = 1 - F(\chi^2) = 1 - \int_0^{\chi^2} f_n(v)\,\mathrm{d}v$ plotted against χ^2 for n degrees of freedom ($n = 1, 2, \ldots 9$). For a given n the curve shows the probability to get a value at lest as big as χ^2. For example for $n = 9$ a value of $\chi^2 > 10$ will appear in 35% of the cases. Stichproben auftreten.

3.11 Students t distribution

If x and $x_1, x_2, \ldots x_n$ are independent random variables following the standardized normal distribution, the quantity t defined by

$$t = x / \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

is distributed according to a Students t distribution with n degree of freedoms. The density is given by

$$\boxed{f_n(t) \; = \frac{1}{\sqrt{n\pi}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \qquad -\infty < t < +\infty}. \tag{45}$$

The mean is finite for $n > 1$, and the variance is finite for $n > 2$; thus the central limit theorem can not be applied for $n \leq 2$. For $n \geq 3$:

$$\text{Expectation} \quad E[t] = 0 \qquad \text{Variance} \quad V[t] = \frac{N}{n-2}$$

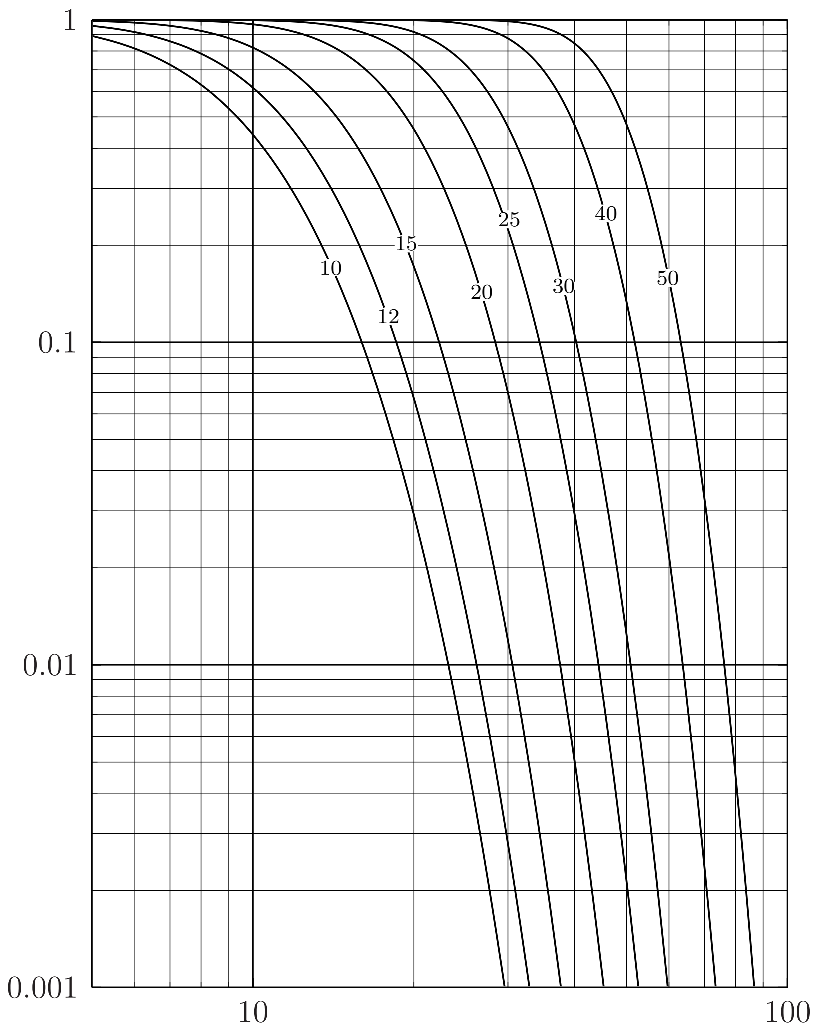

Figure 12: the quantity $1-P$ versus χ^2 forr n degrees of freedom (n = 10, 12, 15, 20, 25, 30, 40 und 50).

Example: The standard application is for normally distributed random variables x_i with unknown mean value μ and unknown variance σ^2. Sample mean and variance are

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

The sample mean $\overline{x}$ has a Gaussian distribution with variance σ^2/n, and the variable $(\overline{x}-\mu)/\sqrt{\sigma^2/n}$ follows the standard Gaussian distribution (mean 0 and variance 1). The variable $(n-1)s^2/\sigma^2$ follows the χ^2-distribution with $n-1$ degrees of freedom, and is independent of the other variable. The ratio of the two variables

$$t = \frac{(\overline{x}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{(n-1)s^2/\sigma^2(n-1)}} = \frac{(\overline{x}-\mu)}{\sqrt{s^2/n}}$$

follows the t-distribution with $n-1$ degrees of freedom; in the ratio above the true variance σ^2 cancels. The value of t can be used to test the probability that the true variable is some particular value μ.

Random number generation. A random number t can be generated from a random number z from the standardized normal distribution and a random number y, distributed according to the gamma distribution with $k = n/2$ and $\lambda = 1$. In the special cases $n = 1$ Students t distribution and the Cauchy distribution are identical and the simpler method shown for the Cauchy distribution can be used.

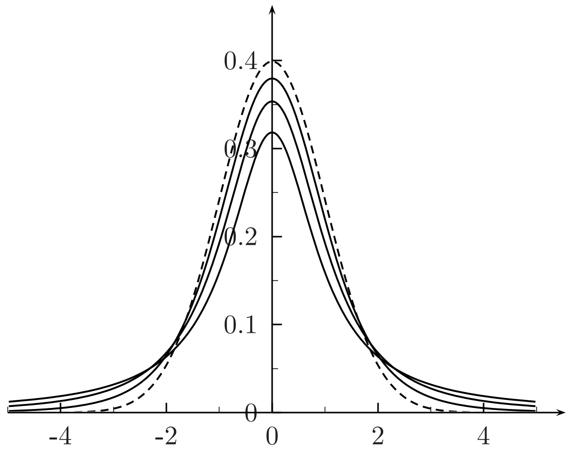

Figure 13: Students t distribution for $n = 1, 2$ und 5 degrees of freedom. For comparison the standarddized Gaussna distribution $N(0, 1)$ is shown, which is approximated by the t distribution with increasing number of degrees of freedom.

3.12 F distribution

If x_1, x_2 are independent random variables distributed as $\chi^2_{\nu_1}$, $\chi^2_{\nu_2}$ respectively, then the distribution of the ratio $(x_1^2\nu_2)/(x_2^2\nu_1)$ is the F distribution with ν_1, ν_2 degrees of freedom. The F distribution is applicable to the distribution of rations of independent estimates of variance. The distribution is given by

$$\boxed{f(F) = \left(\frac{n_1}{n_2}\right)^{n_1/2} \cdot \frac{\Gamma((n_1+n_2)/2)}{\Gamma(n_1/2)\Gamma(n_2/2)} \cdot F^{(n_1-2)/2}\left(1+\frac{n_1}{n_2}F\right)^{-(n_1+n_2)/2}} .$$

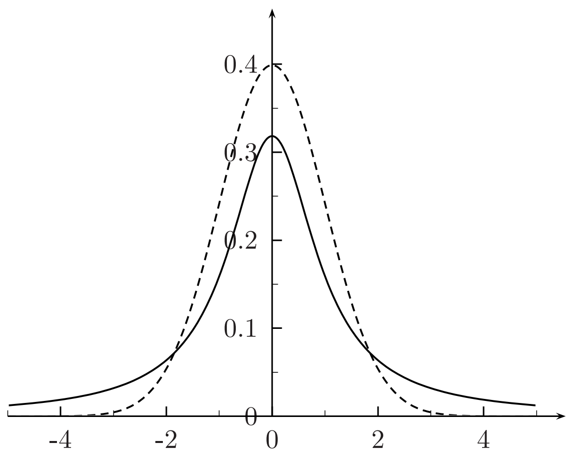

Figure 14: The Cauchy distribution $f(x)$ with FWHM = 2, which is identical to the t distribution with one degree of freedom. The normal distribution shown by a dashed curve approaches zero much smaller for large $|x|$.

3.13 Cauchy distribution

The Cauchy distribution has the density

$$\boxed{f(x) = \frac{1}{\pi}\frac{1}{1+x^2}} .$$

The expecation value is undefined and all other moments diverge. The center value is zero and the full width at half maximum (FWHM) is equal to 2. The distribution

$$f(x) = \frac{1}{\pi}\frac{\Gamma/2}{(x-x_0)^2+\Gamma^2/4} .$$

is called the Breit-Wigner distribution. It corresponds to the Cauchy distribution with central value x_0 and with full width at half maximum of Γ.

Random number generation. The inverse transformation method yields the formula

$$x = \frac{1}{\tan(\pi u)} \qquad (46)$$

for a random Cauchy distributed variable x from a uniform random number u. A method which avoids a function call is the following:

1. Generate two random numbers u_1 and u_2 and construct $v_1 := 2u_1 - 1$ and $v_2 := 2u_2 - 1$.
2. calculate $r^2 = v_1^2 + v_2^2$; if $r^2 > 1$, goto step 1;
3. calculate $x = v_1/v_2$

4 Practical problems

4.1 Expressing the uncertainty of measurement results

In general, the result of a measurement is only an approximation or estimate of the value of the specific quantity subject to measurement, that is, the **measurand**, and the result is complete only when accompanied by a quantitative statement of its uncertainty. The problem is discussed in the paper National Institute of Standards and Technology (NIST), USA: Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, Technical Note 1297, by B.N. Taylor and C.E. Kuyatt; the text below gives a summary of this paper

The uncertainty of the result of a measurement generally consists of several components which may be grouped into two categories according to the method used to estimate their numerical values:

A. those which are evaluated by statistical methods,

B. those which are evaluated by other means.

There is not always a simple correspondence between the classification into categories A and B and the commonly used classification of uncertainty components as "random" and "systematic". The nature of an uncertainty component is conditioned by the use made of the corresponding quantity, that is, on how that quantity appears in the mathematical model that describes the measurement process. When the corresponding quantity is used in a different way, a "random" component may become a "systematic" component and vice versa. Thus the terms "random uncertainty" and "systematic uncertainty" can be misleading when generally applied, and alternative nomenclature that might be used is

"component of uncertainty arising from a random effect,"

"component of uncertainty arising from a systematic effect,"

where a random effect is one that gives rise to a possible random error in the *current measurement process* and a systematic effect is one that gives rise to a possible systematic error in the current measurement process. In principle, an uncertainty component arising from a systematic effect may be in some cases be evaluated by method A while in other cases by method B, as may be an uncertainty component arising from a random effect.

Each component of uncertainty that contributes to the uncertainty of a measurement result is represented by an estimated standard deviation, termed **standard uncertainty** with suggested symbol u_i, and equal to the positive square root of the estimated variance u_i^2.

(A) The evaluation of uncertainty by the statistical analysis of series of observations is termed **Type A evaluation (of uncertainty)**. An uncertainty component in category A is represented by a statistically estimated standard deviation s_i (positive square root of the statistically estimated variance s_i^2), and the associated number of degrees of freedom ν_i.

The evaluation may be based on any valid statistical method for treating data. Examples are calculating the standard deviation of the mean of a series of independent observations or using the method of least squares to fit a curve to data in order to estimate the parameters of the curve and their standard deviations.

(B) In a similar manner, an uncertainty component in category B is represented by a quantity u_j, which may be considered an approximation to the corresponding standard deviation; it is equal to the positive square root of u_j^2, which may be considered an approximation to the corresponding variance and which is obtained from an assumed probability distribution based on all the available information (**Type B evaluation (of uncertainty)**).

A Type B evaluation of standard uncertainty is usually based on scientific judgment using all the relevant infprmation available, which may include previous measurement data, experience with the behavior and property of relevant materials and instruments, calibration data.

The **combined standard uncertainty** of a measurement result, suggested symbol u_c, is taken to represent the estimated standard deviation of the result. It is obtained by combining the individual standard uncertainties u_i (and covariances as appropriate), wether arising from a Type A evaluation or Type B evaluation, using the usual method for combining standard deviations. This method is the *law of propagation of uncertainty* and in common parlance the "root-sum-of-squares" (square root of the sum-of-the-squares) or "RSS" method.

Terminology

Accuracy of measurement = closeness of the agreement between the result of a measurement and the value of the measurand. "Accuracy" is a qualitative concept; it should not be used quantitatively (associate numbers with it). Numbers should be associated with measures of uncertainty, e.g. "the standard uncertainty is 2 μC", but not "the accuracy is 2 μC". The word inaccuracy should not be used. The term **precision** should not be used for "accuracy". The phrase "the value of the measurand" is use instead of the phrase "true value of the measurand" (or simply "true value"). The "true value of a measurand" is simply the value of the measurand.
repeatability (of results of measurements) = closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement (repeatibility conditions, include same measurement procedure, same observer, same measurement instrument, used under the same conditions, same location, and repetition over a short period of time).
reproducibility (of results of measurements) = closeness of the agreement of the same measurand carried out under changed conditions of measurement. The statement of reproducibility requires specification of the conditions changed, which may include principle of measurement, Method of measurement, observer, measurement instrument, reference standard, location, conditions of use, time.
Both the repeatability and the reproducibility can be expressed quantitatively in terms of the dispersion chracteristics of the result.
error (of measurement) = result of a measurement minus the value of the measurand; in general this is unknown.
random error = result of a measurement minus the mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions; it can only be estimated, because only a finite number of measurements can be made.
systematic error = mean that would result from a infinite number of measurements of the same measurand caried out under repeatability conditions minus the value of the measurand.
correction = value added algebraically to the uncorrected result of a measurement to compensate for systematic error (= negative of the estimated systematic error);
correction factor = numerical factor by which the uncorrected result of a measurement is multiplied to compensate for systematic error.

4.2 Sample mean and variance

For a sample of x values $x_1, \ldots, x_i, \ldots, x_n$ the sample mean and variance are given by

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad\qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 .$$

The sample variance can also be calculated by the formula

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) .$$

In practice both results can be obtained in one loop. In order to reduce round-off errors, one should use an estimate x_e for the mean in the formulas

$$S_x = \sum_{i=1}^{n}(x_i - x_e) \qquad\qquad S_{xx} = \sum_{i=1}^{n}(x_i - x_e)^2$$

with the result

$$\overline{x} = x_e + \frac{1}{n}S_x \qquad\qquad s^2 = \frac{1}{n-1}\left(S_{xx} - \frac{1}{n}S_x^2\right)$$

4.3 Averages

Averages from independent measurements x_i with standard deviations σ_i are formed by weighted means, with weight factors $w_i = 1/\sigma_1^2$:

$$\overline{x} \pm \delta\overline{x} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i} \pm \left(\sum_{i=1}^{N} w_i\right)^{-1/2} \tag{47}$$

The measurements may be incompatible or inconsistent; for a test wether the data are consistent with each other the test quantity

$$\chi^2 = \sum_{i=1}^{N} w_i \left(\overline{x} - x_i\right)^2 \tag{48}$$

should be calculated. This quantity should follow the χ^2 distribution with $(N-1)$ degrees of freedom, with expectation value $(N-1)$, if the measurements are Gaussian. The startegy of the PDG is the following: depending on the value of χ^2 the following decisions can be taken

$\chi^2/(N-1) \leq 1$ the result can be accepted;

$\chi^2/(N-1) \gg 1$ the data contradict each other. Either averaging should not be done, or the calculated average may be quoted, but with a conservative estimate of the uncertainty, taking into account the problems;

$\chi^2/(N-1) \gtrsim 1$ (not much greater than 1) averaging can still be done, but the uncertainties δx_i appear to be underestimated on at least one of the measurements. One can assume that all are underestimted by the factor

$$S = \left[\chi^2/(N-1)\right]^{1/2} \tag{49}$$

this factor can be used as a scale factor used to increase the uncertainty $\delta\overline{x}$ calculated in equation (47). This is equivalent to scaling up *all* input errors δx_i before taking the average.

If all measurements have the same error δx_i, then the error of the average is $\delta\overline{x} = \delta x_i/\sqrt{N}$. Averaging of measurements with very different uncertainties δx_i can present special problems. In this case one can calculate from δ the unscaled uncertainty $\delta\overline{x}$ the value

$$\delta_0 = 3\sqrt{N}\delta\overline{x} \tag{50}$$

One strategy is to evaluate the scale factor S only from the measurements with $\delta x_i < \delta_0$. The use of a scale factor S has no influence on the central averaged value $\overline{x}$.

4.4 Method of moments

The method of moments is a general and simple method of parameter estimation. It is optimal only in certain cases, but it is often useful to obtain a rough estimate a parameter. The estimate can be used as initial parameter value in more effective, but also more complicated method.

Given a sample x_i, $i = 1, 2, \ldots n$, assumed to come from a distribution with density $f(x;a)$, where x is the random variable and a is a parameter, to be estimated from the data.

The theoretical expectation of a function $h(x)$ of the random variable, given the density $f(x;a)$, is itself a function $g(a)$ of the parameter a:

$$E\left[h(x)\right] = \int_{-\infty}^{+\infty} h(x)\, f(x,a)\, \mathrm{d}x = g(a)\ . \tag{51}$$

On the other hand, the average value of the function $h(x)$ for the data sample is

$$\overline{h(x)} = \frac{1}{n}\sum_{i=1}^{n} h(x_i)\ . \tag{52}$$

The principle of the method of moments is to to equate the theoretical expectation value and the empirical average of the function $h(x)$, to obtain the parameter estimate $\widehat{a}$:

$$g(\widehat{a}) = \overline{h(x)} \qquad\qquad \widehat{a} = g^{-1}\left(\overline{h(x)}\right) \tag{53}$$

which requires the function g^{-1} inverse to $g(a)$. According to the central limit theorem the mean value $\overline{h(x)}$ is asymptotically normal distributed with variance

$$V\left[\overline{h(x)}\right] = \frac{1}{n} V\left[h(x)\right]$$

Using standard error propagation the variance of the estimate $\widehat{a}$ is given by

$$V\left[\widehat{a}\right] = \frac{1}{n}\left(\frac{\partial g^{-1}}{\partial a}\right)^2 V\left[h(x)\right] \tag{54}$$

The approximate normality of the estimate is often destroyed by the inverse function $g^{-1}\left(\overline{h(x)}\right)$. The estimate has a potential bias $b_n(\widehat{a}) \propto 1/n$.

Example: The angular distribution $f(x) = 1/2\,(1+ax)$ with $x \equiv \cos\vartheta$ is assumed. In order to get an estimate $\widehat{a}$ the function $h(x) \equiv x$ is used.

$$\begin{aligned} E\left[x\right] &= \int_{-1}^{+1} x \frac{1}{2}\left(1+ax\right)\,\mathrm{d}x = \frac{1}{3}a \qquad\qquad \Rightarrow \quad \widehat{a} = 3\widehat{a} \\ V\left[x\right] &= E\left[\left(x - \frac{1}{3}a\right)^2\right] = \int_{-1}^{+1}\left(x-\frac{1}{3}a\right)^2 \frac{1}{2}\left(1+ax\right)\,\mathrm{d}x = \frac{1}{3} - \frac{1}{9}a^2 \end{aligned}$$

For the variance of the average $\overline{x}$ and the variance of the estimate $\widehat{a}$ follows:

$$V\left[\overline{x}\right] = \frac{1}{n}\left(\frac{1}{3} - \frac{1}{9}a^2\right) \qquad\qquad V\left[\widehat{a}\right] = \frac{1}{n}\left(3 - a^2\right) \tag{55}$$

5 Numerical linear algebra

5.1 Vectors and matrices

Vectors. A n-vector $\boldsymbol{x}$ has n Elements $x_1, x_2, \ldots x_n$ in a column; the transposed vector $\boldsymbol{x}^T$ is a row vector:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \boldsymbol{x}^T = (x_1, x_2, \ldots x_n) \; . \tag{56}$$

The scalar product of two vectors $\boldsymbol{x}$ und $\boldsymbol{y}$ is defined by

$$\boldsymbol{x}^T \boldsymbol{y} = \sum_{i=1}^{n} x_i \, y_i = x_1 \, y_1 + x_2 \, y_2 + \ldots + x_n \, y_n \; . \tag{57}$$

with $\boldsymbol{x}^T \boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{x}$. The norm of a vector is defined by

$$|\boldsymbol{x}| = \left(\boldsymbol{x}^T \boldsymbol{x}\right)^{1/2} \; .$$

Matrices. A $m \times n$ Matrix $\boldsymbol{A}$ is a scheme of m rows and n columns; the element A_{ij} is the element in row i and column j:

$$\boldsymbol{A} = (A_{ij}) = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ . & . & \ldots & . \\ A_{m1} & A_{m2} & \ldots & A_{mn} \end{pmatrix} \; . \tag{58}$$

The transposed of matrix $\boldsymbol{A}$ is $\boldsymbol{A}^T$ and has the elements $(\boldsymbol{A}^T)_{ij} = (\boldsymbol{A})_{ji}$. The elements A_{ii} of a square matrix are called diagonal elements of matrix $\boldsymbol{A}$; the sum of the diagonal elements is called trace of the matrix: $\text{trace}(\boldsymbol{A}) = \sum_i A_{ii}$. The unit matrx $\boldsymbol{I}$ has all diagonal elements equal to 1, all other elements are 0. Addition and subtraction of vectors and matrices is done elementwise; the operands have to agree in their dimensions.
The product of the $m \times n$ matrix $\boldsymbol{A}$ and the $n \times k$ matrix $\boldsymbol{B}$ is a $m \times k$ matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \boldsymbol{A}\,\boldsymbol{B} \qquad C_{ij} = \sum_{k=1}^{n} A_{ik} \, B_{kj} \; , \tag{59}$$

in general $\boldsymbol{AB} \neq \boldsymbol{BA}$. The transposed of a product is $(\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$.

Determinants. The determinant of a $n \times n$ matrix is defined by

$$\det \boldsymbol{A} = \sum_{\text{perm}} (-1)^k \, A_{1\alpha} A_{2\beta} \ldots A_{n\omega} , \tag{60}$$

where the indices $\alpha, \beta, \ldots \omega$ are all $n!$ permutations of $1, 2, \ldots n$; the sign $(-1)^k$ of a term is determined by the number of inversions in the permutation. The determinant is computed by the recursion for column j and for row i, respectively

$$\det \boldsymbol{A} = \sum_{i=1}^{n} (-1)^{i+j} \, A_{ij} \det \boldsymbol{A}^{(ij)} \qquad \det \boldsymbol{A} = \sum_{j=1}^{n} (-1)^{i+j} \, A_{ij} \det \boldsymbol{A}^{(ij)} \; , \tag{61}$$

where $\boldsymbol{A}^{(ij)}$ is the matrix obtained by removing row i and column j (Cramers rule). The equations

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A}) \cdot \det(\boldsymbol{B}) \qquad \det\left(\boldsymbol{A}^T\right) = \det(\boldsymbol{A}) \; . \tag{62}$$

are valid.
Linear transformation. A *linear* transformation is written in matrix form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \qquad y_i = \sum_{k=1}^{n} A_{ik} \, x_k \; . \tag{63}$$

The inverse transformation, if possible (same dimension of x and y) and if existing, is written in the form $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$, with the inverse matrix $\boldsymbol{A}^{-1}$:

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I} \; .$$

The transformation is called *regular* or *non-singular* if the inverse $\boldsymbol{A}^{-1}$ exists; in this case $\det(\boldsymbol{A}) \neq 0$.
A set of linear equations in matrix notation is given by

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \qquad \sum_{k=1}^{n} A_{ik}\, x_k = b_i \qquad i = 1, 2, \ldots m \; . \tag{64}$$

For $n = m$ there is a unique solution for $\boldsymbol{x}$, given $\boldsymbol{b}$, unless matrix $\boldsymbol{A}$ is singular.
For a 2×2 matrix $\boldsymbol{A}$ the determinant is $\det(\boldsymbol{A}) = A_{11} \cdot A_{22} - A_{12} \cdot A_{21}$. The inverse exists for $\det \boldsymbol{A} \neq 0$:

$$\boldsymbol{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \qquad \boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}} \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix} \; ,$$

The solution of the matrix equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is

$$\begin{aligned} x_1 &= (A_{22}b_1 - A_{12}b_2) \,/\, \det(\boldsymbol{A}) \\ x_2 &= (A_{11}b_2 - A_{21}b_1) \,/\, \det(\boldsymbol{A}) \, . \end{aligned}$$

Band matrices. In a band matrix all elements A_{ij} are 0 for all i, j with $|i - j| > m$, i.e. only the elements on the diagonal and in a band around the diagonal are nonzero. The value m is called band width. For $m = 0$ the matrix is diagonal and for $m = 1$ *tridiagonal*. The inverse of a band matrix is a general matrix, i.e. the band structure is lost by inversion.

Eigenvectors and eigenvalues. A non-vanishing vektor $\boldsymbol{x}$ satisfying equation

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \qquad \text{oder} \qquad (\boldsymbol{A} - \lambda \boldsymbol{I})\, \boldsymbol{x} = 0 \tag{65}$$

is called *eigenvector* of the square matrix $\boldsymbol{A}$: the scalar λ is called *eigenvalue*. The eigenvalues are determined by the condition $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$. The characteristic polynomial of order n in λ is $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = p(\lambda) = (\lambda)^n + (\lambda)^{n-1} \operatorname{trace}(\boldsymbol{A}) + \ldots + \det(\boldsymbol{A})$, it has n zeros. The n zeros are the n eigenvalues $\lambda_1, \lambda_2 \ldots \lambda_n$, which in general are complex (for symmetric matrices they are real, see below).
The transformation of a matrix $\boldsymbol{A}$ to matrix $\boldsymbol{B}$ by

$$\boldsymbol{B} = \boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{C} \tag{66}$$

with a regular matrix $\boldsymbol{C}$ is called similarity transformation. The eigenvalues (and the characteristic polynomial) are not changed by a similarity transformation:

$$\begin{aligned} \det(\boldsymbol{B} - \lambda \boldsymbol{I}) &= \det\left(\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{C} - \lambda \boldsymbol{I}\right) = \det\left(\boldsymbol{C}^{-1}(\boldsymbol{A} - \lambda \boldsymbol{I})\, \boldsymbol{C}\right) \\ &= \det \boldsymbol{C}^{-1} \cdot \det(\boldsymbol{A} - \lambda \boldsymbol{I}) \cdot \det \boldsymbol{C} = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) \; . \end{aligned}$$

A matrix can be transformed to a diagonal matrix $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \ldots \lambda_n)$

5.2 Symmetric matrices

A square matrix $\boldsymbol{V}$ is called symmetric if $\boldsymbol{V}^T = \boldsymbol{V}$, i.e. $V_{ij} = V_{ji}$. Many matrices in statistical applications are symmetric. The inverse of a non-singular symmetric matrix is again symmetric. All eigenvalues of a real symmetric matrix are *real*. They are all positive for a positive definite symmetric matrix. The eigenvectors form an orthogonal system.
For a general $n \times m$ matrix $\boldsymbol{A}$ the product $\boldsymbol{W} = \boldsymbol{A}^T\boldsymbol{A}$ is a symmetric $n \times n$ matrix. The quadratic form

$$Q(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{W}\, \boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^T (\boldsymbol{A}\boldsymbol{x}) \; , \tag{67}$$

is a non-negative scalar. The matrix $\boldsymbol{W}$ is called *positive definite*, if for the quadratic form $Q(\boldsymbol{x}) > 0$ for all (non-vanishing) vectors $\boldsymbol{x}$.
The product $\boldsymbol{V}_2 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{A}^T$ with the symmetric matrix $\boldsymbol{V}_1$ of order n and a general $m \times n$ matrix $\boldsymbol{A}$ is a symmetric matrix of order m (error propagation).
The elements A_{ij} of a symmetric matrix can be mapped to a one-dimensional table with index

$$q = \begin{cases} i + j(j-1)/2 & \text{für} \quad i \leq j \\ j + i(i-1)/2 & \text{für} \quad i \geq j \end{cases} \tag{68}$$

oder $q = \min(i,j) + [\max(i,j)\,(\max(i,j) - 1)]/2$ for $q = 1 \ldots (n^2 + n)/2$.

5.3 Exchange algorithm

The exchange algorithm allows the solution of linear systems of equations. Given for example the system

$$\begin{aligned} y_1 &= A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + A_{14}x_4 \\ y_2 &= A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + A_{24}x_4 \\ y_3 &= A_{31}x_1 + A_{32}x_2 + A_{33}x_3 + A_{34}x_4 \; . \end{aligned} \tag{69}$$

of equations for unknowns x_1, x_2, x_3 and x_4. In an exchange step the role of one of the unknowns and one of the y elements is exchanged; in the case of elements x_4 and y_2 the second equation becomes

$$x_3 = -\frac{A_{21}}{A_{23}}x_1 - \frac{A_{22}}{A_{23}}x_2 + \frac{1}{A_{23}}y_2 - \frac{A_{24}}{A_{23}}x_4 \tag{70}$$

($A_{23} \neq 0$ required). The variable x_3 can be expressed in x_1, x_2, x_4 and y in the scheme:

$$\begin{array}{r} \\ y_1 = \\ y_2 = \\ y_3 = \end{array} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ \hline A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & \underline{A_{23}} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{array} \Longrightarrow \begin{array}{r} \\ y_1 = \\ x_3 = \\ y_3 = \end{array} \begin{array}{cccc} x_1 & x_2 & y_2 & x_4 \\ \hline \alpha_{11} & \alpha_{12} & \dfrac{A_{13}}{A_{23}} & \alpha_{14} \\ -\dfrac{A_{21}}{A_{23}} & -\dfrac{A_{22}}{A_{23}} & \dfrac{1}{A_{23}} & -\dfrac{A_{24}}{A_{23}} \\ \alpha_{31} & \alpha_{32} & \dfrac{A_{33}}{A_{23}} & \alpha_{34} \end{array}$$

$$\begin{aligned} \alpha_{11} &= A_{11} - \frac{A_{21}A_{13}}{A_{23}} & \alpha_{12} &= A_{12} - \frac{A_{22}A_{13}}{A_{23}} & \alpha_{14} &= A_{14} - \frac{A_{24}A_{13}}{A_{23}} \\ \alpha_{31} &= A_{31} - \frac{A_{21}A_{33}}{A_{23}} & \alpha_{32} &= A_{32} - \frac{A_{22}A_{33}}{A_{23}} & \alpha_{34} &= A_{34} - \frac{A_{24}A_{33}}{A_{23}} \; . \end{aligned}$$

The underlined element (A_{23} in the example) is called *pivot element*; in the exchange step it is replaced by its inverse. The other elements in the column and row of the pivot element are divided by the pivot element, with sign change in the pivot row. All other elements are modified by the product of coresponding elements in the row and column of the pivot element, divided by the pivot element. (rectangular rule). The rules for an exchange step are:

1. Elements outside of the pivot row and column are transformed by the rectangular rule;
2. elements of the pivot column are divided by the pivot element;
3. elements of the pivot row are divided by the pivot element, their sign is reversed;
4. the pivot element is replaced by its inverse.

One exchange step requires $n \times m$ operations in the case of m linear functions and n variables. Exchange steps are *reversible*. The pivot element has to be non-zero. A small value of the pivot element may result in large rounding errors.

Computation of the inverse matrix. In a matrix equation $\boldsymbol{Ax} = \boldsymbol{y}$ for n linear equations and n variables n exchange steps are necessary to exchange all x variables by all y variables. The matrix is inverted in this process, which requires n^3 operations. In order to reduce rounding errors the largest of all remaining elements should be used as the pivot element in each step. The matrix is singular if all candidates for the pivot element are zero.

Symmetric matrices. For symmetric matrices the exchange algorithm can be modified such that all intermediate matrices are symmetric too, with a reduction of the number of operations to about 1/2. All pivot elements have to be selected on the diagonal. The signs of all elements of the pivot row including the pivot element have to be reversed (rules 3 and 4). At the end the signs of all elements of the matrix have to be reversed.

5.4 Triangular matrices

Definition and properties. Elements of right triangular matrices $\boldsymbol{R}$ below the diagonal, and of left triangular matrices $\boldsymbol{L}$ above the diagonal are zero:

$$R_{ij} = 0 \quad \text{for} \quad i > j \qquad\qquad \boldsymbol{R} = \begin{pmatrix} R_{11} & R_{12} & R_{13} & \dots & R_{1n} \\ & R_{22} & R_{23} & \dots & R_{2n} \\ & & R_{33} & \dots & R_{3n} \\ & & & & \vdots \\ & & & & R_{nn} \end{pmatrix} .$$

$$L_{ij} = 0 \quad \text{for} \quad i < j \qquad\qquad \boldsymbol{L} = \begin{pmatrix} L_{11} & & & & \\ L_{21} & L_{22} & & & \\ L_{31} & L_{32} & L_{33} & & \\ \vdots & & & & \\ L_{n1} & L_{n2} & L_{n3} & \dots & L_{nn} \end{pmatrix} .$$

The solution of systems of equations ist straightforward, if the matrix is a right or left triangular matrix. A general matrix non-singular matrix $\boldsymbol{A}$ can be represented (or decomposed) as product $\boldsymbol{LR}$ of a left and a right triangular matrix: $\boldsymbol{A} = \boldsymbol{LR}$. The general matrix equation $\boldsymbol{Ax} = \boldsymbol{b}$ with decomposition $\boldsymbol{A} = \boldsymbol{LR}$,

$$\boldsymbol{Ax} = \boldsymbol{L}\,(\boldsymbol{Rx}) = \boldsymbol{b} \; , \tag{71}$$

is solved in two steps, using an auxiliary vector $\boldsymbol{y}$, defined by $\boldsymbol{Rx} = \boldsymbol{y}$:

$$\begin{array}{rrcll} & \boldsymbol{L\,y} & = & \boldsymbol{b} & \qquad\text{forward solution,} \\ \text{followed by} & \boldsymbol{R\,x} & = & \boldsymbol{y} & \qquad\text{backward solution.} \end{array}$$

There are different decompositon methods. The Cholesky method can be used for symmetric positive definite matrices and Crouts method can be used in the general case. The inverse of a triangular matrix is again triangular.

Solution of matrix equations with triangular matrices. For a linear system $\boldsymbol{Lx} = \boldsymbol{b}$ the general formula is

$$x_i = \frac{b_i - \sum_{j=1}^{i-1} L_{ij} x_j}{L_{ii}} \qquad i = 1,\, 2,\, \dots n \qquad \text{for} \quad L_{ii} \neq 0 \tag{72}$$

(forward solution, starting with x_1). The general formula for a right triangular matrix is:

$$x_i = \frac{b_i - \sum_{j=i+1}^{n} R_{ij} x_j}{R_{ii}} \qquad i = n,\, \dots 2,\, 1 \qquad \text{for} \quad R_{ii} \neq 0 \tag{73}$$

(backward solution, starting with x_n).

Inversion of a triangular matrix. For any non-singular left triangular matrix $\boldsymbol{L}$ there is an inverse $\boldsymbol{M}$, which is left triangular too. The elements of the inverse matrix have to satisfy the conditions

$$\sum_{j=1}^{n} L_{ij} M_{jk} = \begin{cases} 1 & \text{for} \quad i = k \\ 0 & \text{for} \quad i \neq k \end{cases} .$$

The general formula for inversion is

$$M_{ik} = -\frac{\sum_{j=k}^{i-1} L_{ij} M_{jk}}{L_{ii}} \quad k < i \qquad\qquad M_{ii} = \frac{1}{L_{ii}} \qquad i = 1,\, 2, \dots\; n \; . \tag{74}$$

5.5 General LR decomposition

Crouts method is a decomposition of the $n \times n$ matrix $\boldsymbol{A}$ into $\boldsymbol{A} = \boldsymbol{LR}$, where $\boldsymbol{L}$ is a left triangular matrix with diagonal elements 1 and $\boldsymbol{R}$ is a right triangular matrix:

$$\boldsymbol{L} = \begin{pmatrix} 1 & & & & \\ L_{21} & 1 & & & \\ L_{31} & L_{32} & 1 & & \\ \dots & & & & \\ \dots & & & & \\ L_{n1} & L_{n2} & L_{n3} & \dots & 1 \end{pmatrix} \qquad \boldsymbol{R} = \begin{pmatrix} R_{11} & R_{12} & R_{13} & \dots & R_{1n} \\ & R_{22} & R_{23} & \dots & R_{2n} \\ & & R_{33} & \dots & R_{3n} \\ & & & & \vdots \\ & & & & R_{nn} \end{pmatrix} . \tag{75}$$

The general formulae for $k = 1,\, 2, \dots n$ follow from the comparison of the elements of the product matrix $\boldsymbol{LR}$ with the elements of matrix $\boldsymbol{A}$ for the rows,

$$R_{ik} = A_{ik} - \sum_{j=1}^{i-1} L_{ij} R_{jk} \qquad\qquad i = 1, 2, \dots k$$

and for the columns

$$L_{ik} = \frac{1}{R_{kk}} \left[A_{ik} - \sum_{j=1}^{k-1} L_{ij} R_{jk} \right] \qquad\qquad i = k,\, k+1, \dots n \; .$$

Matrices $\boldsymbol{L}$ and $\boldsymbol{R}$ both can be stored in the array of matrix $\boldsymbol{A}$.

5.6 Cholesky decomposition

Symmetric decomposition of a symmetric matrix. A positive definite symmetric matrix $\boldsymbol{W}$ can be expressed in the form

$$\boldsymbol{W} = \boldsymbol{R}^T \boldsymbol{R} \tag{76}$$

with a right triangular matrix $\boldsymbol{R}$. The elements of $\boldsymbol{R}$ can be determined by a comparison of the elements of product $\boldsymbol{R}^T \boldsymbol{R}$ with the elements of $\boldsymbol{W}$. The matrix $\boldsymbol{W}$ is transformed in steps into matrix $\boldsymbol{R}$. Starting with $\boldsymbol{W}^{(0)} = \boldsymbol{W}$ and for $k = 1, 2, \ldots (n-1)$

$$W_{ij}^{(k)} = W_{ij}^{(k-1)} - R_{ki} R_{kj} \qquad i, j = k+1, k+2, \ldots n\,. \tag{77}$$

First $k = 1$ and elememts with $i, j = 2, 3, \ldots n$ are transformed, then $k = 2$ and elements with $i, j = 3, 4, \ldots n$ are transformed etc. according to

$$R_{ii} = \sqrt{W_{ii}^{(i-1)}} \qquad\qquad R_{ij} = \frac{W_{ij}^{(i-1)}}{R_{ii}} \quad \text{für} \quad j > i\,. \tag{78}$$

Computation of the inverse matrix. The inverse matrix $\boldsymbol{V} = \boldsymbol{W}^{-1}$ is calculated from $\boldsymbol{R}$:

$$\boldsymbol{W}^{-1} = \left(\boldsymbol{R}^T \boldsymbol{R}\right)^{-1} = \boldsymbol{R}^{-1} \left(\boldsymbol{R}^T\right)^{-1} = \boldsymbol{R}^{-1} \left(\boldsymbol{R}^{-1}\right)^T \;, \tag{79}$$

which requires the inverse (triangular) matrix $\boldsymbol{S} = \boldsymbol{R}^{-1}$. The elements of the inverse matrix $\boldsymbol{V} = \boldsymbol{W}^{-1} = \boldsymbol{S}\boldsymbol{S}^T$ are

$$V_{ik} = \sum_{j=k}^{n} S_{ij} S_{kj} \quad \text{for} \quad i \leq k \qquad\qquad V_{kk} = \sum_{j=k}^{n} S_{kj}^2 \qquad\quad k = 1, 2, \ldots n\;. \tag{80}$$

The Cholesky decomposition is especially useful for band matrices, because the band structure is preserved in the decomposition.

5.7 Inversion by partitioning

Sometimes a matrix $\boldsymbol{A}$ to be inverted has a special structure, which allows a significant reduction of the computational effort in the inversion by partitioning. Examples are matrices with a diagonal submatrix or with an existing inverse submatrix.
A square $n \times n$ matrix $\boldsymbol{A}$ can be partitioned in submatrices:

$$\boldsymbol{A} = \left(\begin{array}{c|c} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \hline \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{array}\right) \;, \tag{81}$$

where submatrix $\boldsymbol{A}_{11}$ is a $p \times p$ square matrix and submatrix $\boldsymbol{A}_{22}$ is a $q \times q$ square matrix with $p + q = n$ (values $p = 1$ or $q = 1$ are possible). The inverse matrix $\boldsymbol{B} = \boldsymbol{A}^{-1}$ can be partitioned in the same way:

$$\boldsymbol{B} = \left(\begin{array}{c|c} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \hline \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{array}\right) \;. \tag{82}$$

Matrix product $\boldsymbol{AB}$ (mit $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$) is given by the following matrix equations;

$$\begin{aligned} \boldsymbol{A}_{11}\,\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\,\boldsymbol{B}_{21} &= \boldsymbol{I} \\ \boldsymbol{A}_{11}\,\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\,\boldsymbol{B}_{22} &= \boldsymbol{0} \\ \boldsymbol{A}_{21}\,\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\,\boldsymbol{B}_{21} &= \boldsymbol{0} \\ \boldsymbol{A}_{21}\,\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\,\boldsymbol{B}_{22} &= \boldsymbol{I} \end{aligned} \tag{83}$$

The submatrices of the inverse $\boldsymbol{B}$ are determined by two equivalent set of equations:

$$\begin{aligned} \boldsymbol{B}_{22} &= \left(\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right)^{-1} \\ \boldsymbol{B}_{21} &= -\boldsymbol{B}_{22}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} \\ \boldsymbol{B}_{12} &= -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{22} \\ \boldsymbol{B}_{11} &= \boldsymbol{A}_{11}^{-1} - \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}_{21}\,. \end{aligned} \tag{84}$$

Here the inverse $p \times p$ matrix $\boldsymbol{A}_{11}^{-1}$ is used, while the original matrix $\boldsymbol{A}_{11}$ is not explicitly required. One matrix inversion is required for the $q \times q$ matrix $\boldsymbol{B}_{22}$. By exchange of indices 1 and 2 a second set of equations is obtained:

$$\begin{aligned} \boldsymbol{B}_{11} &= \left(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\right)^{-1} \\ \boldsymbol{B}_{12} &= -\boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ \boldsymbol{B}_{21} &= -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{11} \\ \boldsymbol{B}_{22} &= \boldsymbol{A}_{22}^{-1} - \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{B}_{12} \, . \end{aligned} \tag{85}$$

In the case of *symmetric* matrices $\boldsymbol{A}$ und $\boldsymbol{B}$ the calculation can be further simplified because of

$$\boldsymbol{A}_{12} = (\boldsymbol{A}_{21})^T \qquad \text{und} \qquad \boldsymbol{B}_{12} = (\boldsymbol{B}_{21})^T \; .$$

Inversion by bordering. Partitioning in the case $q = 1$ is called bordering. The $n \times n$ matrix $\boldsymbol{A}_n$ has the form

$$\boldsymbol{A}_n = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2,n-1} & a_{2n} \\ \dots & & & & \dots \\ a_{n-1,1} & a_{n-1,2} & \dots & a_{n-1,n-1} & a_{n-1,n} \\ \hline a_{n1} & a_{n2} & \dots & a_{n,n-1} & a_{nn} \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{A}_{n-1} & \boldsymbol{u} \\ \hline \boldsymbol{v}^T & a_{nn} \end{array}\right)$$

with two $(n-1)$ vectors $\boldsymbol{u}$ und $\boldsymbol{v}$ and a scalar a_{nn}. Inversion is simple if the matrix $\boldsymbol{A}_{n-1}^{-1}$ is already known or easily calculable. One has to calculate

$$\alpha_n = a_{nn} - \boldsymbol{v}^T \boldsymbol{A}_{n-1}^{-1} \boldsymbol{u} \; ,$$

where α_n corresponds to matrix $\boldsymbol{B}_{22}^{-1}$ (see equation (85)). The other submatrices are obtained by matrix products:

$$\boldsymbol{A}_n^{-1} = \left(\begin{array}{c|c} \boldsymbol{A}_{n-1}^{-1} + \dfrac{\boldsymbol{A}_{n-1}^{-1}\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{A}_{n-1}^{-1}}{\alpha_n} & -\dfrac{\boldsymbol{A}_{n-1}^{-1}\boldsymbol{u}}{\alpha_n} \\ \hline -\dfrac{\boldsymbol{v}^T\boldsymbol{A}_{n-1}^{-1}}{\alpha_n} & \dfrac{1}{\alpha_n} \end{array}\right) . \tag{86}$$

The calculation is further simplified in the case of a symmetric matrix (with $\boldsymbol{u} = \boldsymbol{v}$).
A method for matrix inversion could be based on repeated application of the bordering method. However compared with other methods rounding errors are larger and the application of the method should be restricted to special cases.

5.8 Diagonalization of symmetric matrices

The eigenvalues of real symmetric matrices are real. Eingenvalues and eigenvectors can be determined by a transformation with orthogonal matrices. A matrix $\boldsymbol{U}$ is called *orthogonal*, if

$$\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{D} \; , \qquad \boldsymbol{D} = \text{diagonal matrix,}$$

(coulumn vectors of $\boldsymbol{U}$ orthogonal). The matrix $\boldsymbol{U}$ is called *orthonormal*, if the diagonal matrix $\boldsymbol{D}$ is a unit matrix with $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}$. If the columns vectors of a square matrix are orthogonal, then the column vectors are orthogonal and normalized too.
For any real symmetric matrix $\boldsymbol{V}$ a orthonormal matrix $\boldsymbol{U}$ can be constructed, which transforms the matrix $\boldsymbol{V}$ to a diagonal matrix:

$$\boldsymbol{U}^T\boldsymbol{V}\boldsymbol{U} = \boldsymbol{D} \quad \text{und} \quad \boldsymbol{V} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \tag{87}$$

with square matrices $\boldsymbol{V}$, $\boldsymbol{U}$ and $\boldsymbol{D}$.

$$\boxed{\boldsymbol{V}} = \boxed{\boldsymbol{U}} \times \boxed{\boldsymbol{D}} \times \boxed{\boldsymbol{U}^T}$$

The diagonal elements $D_{11},\, D_{22} \ldots D_{nn}$ of $\boldsymbol{D}$ are the eigenvalues of $\boldsymbol{V}$, the column vectors of the orthogonal matrix $\boldsymbol{U}$ are the corresponding eigenvectors.
The quadratic form $Q = \boldsymbol{x}^T \boldsymbol{V} \boldsymbol{x}$ of the transformed vector $\boldsymbol{y} = \boldsymbol{U}^T \boldsymbol{x}$ can be written in the form

$$Q = \boldsymbol{x}^T \boldsymbol{V}\, \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \boldsymbol{x} = \boldsymbol{y}^T \boldsymbol{D}\boldsymbol{y} = \sum_{i=1}^{n} D_{ii} y_i^2 \;, \tag{88}$$

and for all non-vanishing $\boldsymbol{y}$ the value of the quadratic form $Q > 0$, if for all diagonal elements $D_{ii} > 0$; in this case the matrix $\boldsymbol{V}$ is positive definite.
Rotation matrix. The simplest non-trivial orthogonal matrices are square matrices with elements $U_{ii} = 1$ for $i \neq p,\, q$ and with

$$\begin{array}{ll} U_{pp} = \cos\varphi & \qquad U_{pq} = \sin\varphi \\ U_{qp} = -\sin\varphi & \qquad U_{qq} = \cos\varphi \;, \end{array}$$

all remaining elements are 0

$$\boldsymbol{U} = \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \cos\varphi & \ldots & \sin\varphi & & & \\ & & & \vdots & & \vdots & & & \\ & & & -\sin\varphi & \ldots & \cos\varphi & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix} \tag{89}$$

with $\boldsymbol{U}^{-1} = \boldsymbol{U}^T$. The result of a transformation

$$\boldsymbol{V}' = \boldsymbol{U}^T \boldsymbol{V} \boldsymbol{U} \tag{90}$$

is as follows. The result for the special elements with column and row indices p and q are:

$$\begin{aligned} V'_{pp} &= V_{pp}\cos^2\varphi - 2V_{pq}\cos\varphi\sin\varphi + V_{qq}\sin^2\varphi \\ V'_{qq} &= V_{pp}\sin^2\varphi + 2V_{pq}\cos\varphi\sin\varphi + V_{qq}\cos^2\varphi \\ V'_{pq} &= (V_{pp} - V_{qq})\cos\varphi\sin\varphi + V_{pq}(\cos^2\varphi - \sin^2\varphi) = V'_{qp}. \end{aligned} \tag{91}$$

The remaining elements in rows p and q for $j = 1,\ 2, \ldots\ n$ are:

$$\begin{aligned} V'_{pj} &= V_{pj}\cos\varphi - V_{qj}\sin\varphi \\ V'_{qj} &= V_{pj}\sin\varphi + V_{qj}\cos\varphi \,, \end{aligned} \tag{92}$$

and the remaining elements in columns p und q für $i = 1,\ 2, \ldots\ n$ are:

$$\begin{aligned} V'_{ip} &= V_{ip}\cos\varphi - V_{iq}\sin\varphi \\ V'_{iq} &= V_{ip}\sin\varphi + V_{iq}\cos\varphi \,. \end{aligned} \tag{93}$$

All other elements are unchanged ($V'_{ij} = V_{ij}$ für $i,\, j \neq p,\, q$). This transformation is called *Jacobi rotation* or $(p,\, q)$ rotation.
Determination of rotation angle φ. The rotation angle φ can be determined such that element V_{pq} is transformed to 0:

$$V'_{pq} = V'_{qp} = (V_{pp} - V_{qq})\cos\varphi\sin\varphi + V_{pq}(\cos^2\varphi - \sin^2\varphi) = 0$$

The angle φ is not uniquely determined by this condition. In the algorithm below, which has small rounding errors, the angle φ is restricted to the range $-\pi/4 < \varphi \leq +\pi/4$.

$$\begin{aligned}
\theta \equiv \operatorname{cotan}(2\varphi) &= \frac{V_{qq} - V_{pp}}{2V_{pq}} \\
\tan\varphi &= \begin{cases} \dfrac{1}{\theta + \operatorname{sign}(\theta)\sqrt{1+\theta^2}} & \text{for} \quad \theta \neq 0 \\ 1 & \text{for} \quad \theta = 0 \end{cases} \\
\cos\varphi &= \frac{1}{\sqrt{1+\tan^2\varphi}} \qquad\qquad \sin\varphi = \cos\varphi \cdot \tan\varphi
\end{aligned}$$

The relation $V'_{pp} + V'_{qq} = V_{pp} + V_{qq}$ is valid for the diagonal elements before and after the transformation.
Jacobi method. The matrix $\boldsymbol{U}$ which transforms a real symmetric matrix $\boldsymbol{V}$ to a diagonal matrix can be determined iteratively as the product of many rotation matrices. The method converges after some initial iterations quadratically. One rotation requires $4n$ operations and in general $3n^2$ to $5m^2$ rotations are necessary. For larger matrices the Householder method is more efficient.
Simultaneous diagonalization of two symmetric matrices. There are problems with a quadratic expression of the form

$$Q = \boldsymbol{x}^T \boldsymbol{V}_a \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{V}_b \boldsymbol{x} \tag{94}$$

with a vector $\boldsymbol{x}$ and two symmetric matrices $\boldsymbol{V}_a$ and $\boldsymbol{V}_b$, where a transformation matrix $\boldsymbol{A}$ has to diagonalize both symmetric matrices to diagonal form. The result is a quadratic expression

$$Q = \boldsymbol{y}^T \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{S} \boldsymbol{y} \tag{95}$$

with a transformed vector $\boldsymbol{y}$, with a diagonal matrix $\boldsymbol{S}$. The matrix $\boldsymbol{V}_a$, assumed to be positive definite, has a decomposition

$$\boldsymbol{V}_a = \boldsymbol{U}_a \boldsymbol{D} \boldsymbol{U}_a^T \, ,$$

with the orthogonal matrix $\boldsymbol{U}_a$. Matrix $\boldsymbol{D}$ is diagonal ($\boldsymbol{D} = \operatorname{diag}(D_{11}, D_{22}, \ldots D_{nn})$). Another diagonal matrices $\boldsymbol{D}^{1/2} = \operatorname{diag}(D_{11}^{1/2}, D_{22}^{1/2}, \ldots D_{nn}^{1/2})$ and $\boldsymbol{D}^{-1/2}$ are defined; the latter has the inverse square roots of the diagonal elements of $\boldsymbol{D}$ on the diagonal. The *symmetric* transformation matrix $\boldsymbol{M}$ is defined by

$$\boldsymbol{M} = \boldsymbol{U_a} \boldsymbol{D}^{-1/2} \boldsymbol{U}_a^T \; .$$

The quadratic form

$$\begin{aligned}
Q = \boldsymbol{x}^T \boldsymbol{V}_a \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{V}_b \boldsymbol{x} &= \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{V}_a \boldsymbol{M} \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{V}_b \boldsymbol{M} \boldsymbol{y} \\
&= \boldsymbol{y}^T \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{V}_b \boldsymbol{M} \boldsymbol{y} \; .
\end{aligned}$$

is obtained by the transformation $\boldsymbol{x} = \boldsymbol{M}\boldsymbol{y}$. The second (symmetric) matrix $\boldsymbol{M}\boldsymbol{V}_b\boldsymbol{M}$ has the decomposition

$$\boldsymbol{M} \boldsymbol{V}_b \boldsymbol{M} = \boldsymbol{U}_b \boldsymbol{S} \boldsymbol{U}_b^T \; .$$

The final transformation matrix $\boldsymbol{A}$ is

$$\boldsymbol{A} = \boldsymbol{M} \boldsymbol{U}_b = \boldsymbol{U_a} \boldsymbol{D}^{-1/2} \boldsymbol{U}_a^T \boldsymbol{U}_b \; ; \tag{96}$$

the transformation of $\boldsymbol{V}_a$ and $\boldsymbol{V}_b$ by matrix $\boldsymbol{A}$ has the result

$$\boldsymbol{A}^T \boldsymbol{V}_a \boldsymbol{A} = \boldsymbol{I} \qquad\qquad \boldsymbol{A}^T \boldsymbol{V}_b \boldsymbol{A} = \boldsymbol{S} \; .$$

The quadratic form of equation (95) is obtained by the transformation $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y}$.

5.9 Singular value decomposition

In this section the problem of solving the matrix equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ for an n-Vektor $\boldsymbol{x}$ of unknowns with a $m \times n$ Matrix $\boldsymbol{A}$ is considered, especially the cases of overdetermined and of ill-conditioned sytems are considered.
Overdetermined system of equations. In general there is no solution vector $\boldsymbol{x}$ in the case $m > n$; There is no vector $\boldsymbol{x}$, where the vector of residua $\boldsymbol{r}$, defined by

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \, , \tag{97}$$

is zero. The standard method for a solution, which at least is in the neighbourhood of an exact solution, is the method of least squares: the solution is determined by the requirement of a minimum value of the square of the vector of residua:

$$\begin{aligned} \boldsymbol{r}^T\boldsymbol{r} &= (\boldsymbol{Ax}-\boldsymbol{b})^T(\boldsymbol{Ax}-\boldsymbol{b}) \\ &= \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{b} = \quad \text{Minimum} . \end{aligned}$$

The minimum is determined, by setting the derivative of the above expression to zero:

$$\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{A}^T\boldsymbol{b} = 0 \quad \text{oder} \quad \boldsymbol{A}^T(\boldsymbol{Ax}-\boldsymbol{b}) = \boldsymbol{A}^T\boldsymbol{r} = 0 .$$

These equation are called *normal equations*; the vector $\boldsymbol{r}$ of residua is *normal* with respect to the column vectors of matrix $\boldsymbol{A}$. In general the system has a unique solution

$$\boldsymbol{x} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

as shown below. First the symmetric matrix $\boldsymbol{W} = \boldsymbol{A}^T\boldsymbol{A}$ and the vector $\boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{b}$ is formed;

$$\boxed{\boldsymbol{W}} = \boxed{\quad\boldsymbol{A}^T\quad} \times \boxed{\boldsymbol{A}} \qquad \boxed{\boldsymbol{c}} = \boxed{\quad\boldsymbol{A}^T\quad} \times \boxed{\boldsymbol{b}} \; .$$

The solution vector $\boldsymbol{x}$ is determined by the solution of as matrix equation $\boldsymbol{Wx} = \boldsymbol{c}$:

$$\boxed{\boldsymbol{x}} = \boxed{\boldsymbol{W}^{-1}} \times \boxed{\boldsymbol{c}} \; .$$

In order to discuss potential problems the solution is writen in a different form. Matrix $\boldsymbol{W}$ can be written as product

$$\boldsymbol{W} = \boldsymbol{UDU}^T \; . \tag{98}$$

Diagonalization with matrices $\boldsymbol{W}$, $\boldsymbol{D}$ and $\boldsymbol{U}$ can be expressed by

$$\boxed{\boldsymbol{W}} = \boxed{\boldsymbol{U}} \times \boxed{\boldsymbol{D}} \times \boxed{\boldsymbol{U}^T}$$

Matrix $\boldsymbol{U}$ is orthonormal and matrix D is diagonal with the eigenvalues λ of matrix W on the diagonal:

$$\boldsymbol{D} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} . \tag{99}$$

The inverse matrix $\boldsymbol{W}^{-1}$ is the product

$$\boldsymbol{W}^{-1} = \boldsymbol{UD}^{-1}\boldsymbol{U}^T ,$$

where matrix $\boldsymbol{D}^{-1}$ has the inverse eigenvalues $1/\lambda_i$, $i = 1, 2, \ldots n$ on its diagonal. The solution can be written in the form

$$\boldsymbol{x} = \boldsymbol{UD}^{-1}\boldsymbol{U}^T\boldsymbol{c} .$$

For a *singular* matrix $\boldsymbol{W}$ there is at least one of the eigenvalues zero and the solution above is impossible (the eigenvalues can not be negative, by construction of the matrix $\boldsymbol{W}$). The columns of matrix $\boldsymbol{U}$ (these are the eigenvectors of matrix $\boldsymbol{W}$), which correspond to the *zero* eigenvalues, form an orthogonal basis of the nullspace of $\boldsymbol{W}$. An arbitrary vector of the null space can be added to the solution $\boldsymbol{x}$. In order to avoid this kind of arbitrariness it is better to define a solution with *no* contribution from the null space. This can be done by defining, instead of matrix $\boldsymbol{D}^{-1}$, a modified matrix $\boldsymbol{D}^{*-1}$, where the diagonal elements for zero eigenvalues of $\boldsymbol{D}$ are set to zero (thus replacing $1/0 = \infty$ by 0). The modified solution

$$\boldsymbol{x} = \boldsymbol{UD}^{*-1}\boldsymbol{U}^T\boldsymbol{c}$$

contains *no* contribution from the null space. It can be shown, that this is the solution with smallest length of $|\boldsymbol{x}|$.
Sometimes a system of equation is not singular, but ill conditioned: some of the eigenvalues are very small and the inversion is inaccurate. This property is visible in the condition number of the matrix, the ratio of the largest to the smallest eigenvalue: $\kappa = \lambda_{\max}/\lambda_{\min}$. In such a case a solution with modified matrix Matrix $\boldsymbol{D}^{*-1}$ is preferred, where the small eigenvalues are replaced by zero.
Singular value decomposition. This decomposition is rather general and is applied, if the matrix is singular or almost singular. A $m \times n$ Matrix $\boldsymbol{A}$ can be expressed in the form

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T , \tag{100}$$

where matrices $\boldsymbol{U}$ and $\boldsymbol{A}$ have the same dimension, while matrices $\boldsymbol{S}$ und $\boldsymbol{V}$ are quadratic $n \times n$ matrices. The matrices $\boldsymbol{U}$ und $\boldsymbol{V}$ should have properties

$$\boldsymbol{U}^T\boldsymbol{U} = \mathbf{1} \qquad \boldsymbol{V}^T\boldsymbol{V} = \mathbf{1}$$
$$\boldsymbol{V}\boldsymbol{V}^T = \mathbf{1}$$

i.e. matrix $\boldsymbol{U}$ is a column-orthogonal $m \times n$ matrix and matrix $\boldsymbol{V}$ is an orthogonal $n \times n$ matrix. Matrix S is diagonal and has positive or zero elements, called *singular values*. The decomposition of matrix $\boldsymbol{A}$ is as follows:

$$\boxed{\boldsymbol{A}} = \boxed{\boldsymbol{U}} \times \boxed{\boldsymbol{S}} \times \boxed{\boldsymbol{V}^T}$$

This representation is possible for $m < n$ too: in this case at least $n - m$ of the singular values are zero. Only the case $m \geq n$ is considered in the following:
In the case of full rank all singular values are non-zero. For $m = n$ the solution

$$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{S}^{-1}\left(\boldsymbol{U}^T\boldsymbol{b}\right) \tag{101}$$

can be identified immediately as the solution of the matrix equation, and in the case $m > n$ the solution is the least squares solution. In the general case the matrix operations are:

$$\boxed{\boldsymbol{x}} = \boxed{\boldsymbol{V}} \times \boxed{\boldsymbol{S}^{-1}} \times \boxed{\quad \boldsymbol{U}^T \quad} \times \boxed{\boldsymbol{b}}$$

The comparison of matrix product $\boldsymbol{W} = \boldsymbol{A}^T\boldsymbol{A}$ with $\boldsymbol{A}$ represented in the form of equation (100)

$$\boldsymbol{W} = \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{S}\boldsymbol{V}^T \tag{102}$$

with equation (98) shows, that $\boldsymbol{D} = \boldsymbol{S}\boldsymbol{S}$, i.e. the eigenvalues λ of matrix $\boldsymbol{W}$ are identical to the squares of the singular values.
As above a modified matrix $\boldsymbol{S}^{*-1}$ is defined, where the diagonal elements for small singular values S_{jj} are set to zero. The solution vector is the least squares solution with minimum length

$$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{S}^{*-1}\left(\boldsymbol{U}^T\boldsymbol{b}\right) . \tag{103}$$

Although mathematically equivalent to the eigenvalue method described above the singular value decomposition is numerically more accurate.

6 Information and algorithms

6.1 Data representation

Data representation in a computer is defined by the hardware. In general eight bits are combined into one byte. There are two conventions used: systems that place the most significant byte at the lowest address in memory (*Big-Endian*), for example IBM 370 series, and systems that place the least significant byte at the lowest address (*Little-Endian*), for example Intel 8086 series. Regardless of the byte ordering the least significant bit is generally referred as bit 0. The most significant bit is bit 7, bit 15 or bit 31, depending on wether the reference is to a byte, a half-word (two bytes), or a word (four bytes).

Integer representation. Integer quantities are usually represented as signed, two's complement numbers of two or four bytes, respectively. The most significant bit of this number represents the sign. Examples for two-byte integers are shown below.

	bits		bits
+1	0000 0000 0000 0001	-1	1111 1111 1111 1111
+2	0000 0000 0000 0010	-2	1111 1111 1111 1110
+3	0000 0000 0000 0011	-3	1111 1111 1111 1101

In the one's complement of a number each bit is reversed (0 replaced by 1 and 1 replaced by 0). The two's complement of a number is the one's complement plus 1. Thus the -1 is represented in a word with all bits equal to 1. An integer quantity can represent integer values between -32768 and $+32767$ (two bytes) and between $-2\,147\,483\,648$ and $+2\,147\,483\,647$ (four bytes). The addition of two large positive numbers may have a "result" which is negative!

Floating point representation. The IEEE floating-point representation (IEEE 754) has become a standard, which is particular widespread in the UNIX world. In this representation a single-precision floating-point number is a 32-bit word quantity, divided into three fields according to the table below.

31	30 23	22 0
sign	biased exponent	fractional part
1 bit	8 bits	23 bits

Real numbers are represented by a fraction and a binary exponent. The value of a single-precision floating-point number is

$$(-1)^{\text{sign}} \times (1 + \text{fraction}) \times 2^{(\text{biased exponent}-127)} \; . \tag{104}$$

The sign is 0 for a positive number, and 1 for a negative number. The binary exponent may have values between -126 and $+127$. The *biased* exponent is the number's binary exponent plus a *bias* of 127; the biased exponent is always a positive integer between 1 and 254. The exponent values -127 and $+128$ are reserved for exception handling.
The *fraction* represent the fractional part of the mantissa. The representation is *normalized*: the start of the integer part is *always* 1 and is not represented. Thus 23 bits are used for the fraction, but the precision corresponds to 24 bits.
Real numbers can have values in the range -3.7×10^{38} to $+3.7 \times 10^{38}$. The minimum magnitude of a floating-point number is 1.2×10^{-38}. Zero is represented by 0s in all fields. Thus real numbers represented by floating-point numbers have a discrete structure with a gap for small numbers except zero.

Double-precision floating-point numbers have 64 bits, with 11 bits for the biased exponent (bias of 1023) and 52 bits for the fractional part. Values in the range -1.7×10^{308} to $+1.7 \times 10^{308}$ can be represented and the minimum number is 2.3×10^{-308}.

Floating point error conditions. In IEEE 754 positive or negative overflow is represented by positive or negative infinity, printed *Inf* and *−Inf*; an indeterminant or impossible result like arccos(1.2) is represented by "Not a number", printed *NaN*, Details of exception handling depend upon the system's implementation. When infinities and NANs appear in computations, they behave in the expected way; for example any operation on a NaN produces another NaN.

Characters. Characters, i.e. the characters of the alphabet in lower and upper case and special characters like $ and & are usually stored in a (8 bit) byte. The Table 1 shows the ASCII code; for the characters shown only 7 bits are necessary.

int	ch	int	ch	int	ch	int	ch	int	ch	int	ch	int	ch	int	ch
0		1		2		3		4		5		6		7	
8		9		10		11		12		13		14		15	
16		17		18		19		20		21		22		23	
24		25		26		27		28		29		30		31	
32	blk	33	!	34	"	35	#	36	$	37	%	38	&	39	’
40	(	41	)	42	*	43	+	44	,	45	-	46	.	47	/
48	0	49	1	50	2	51	3	52	4	53	5	54	6	55	7
56	8	57	9	58	:	59	;	60	¡	61	=	62	¿	63	?
64	@	65	A	66	B	67	C	68	D	69	E	70	F	71	G
72	H	73	I	74	J	75	K	76	L	77	M	78	N	79	O
80	P	81	Q	82	R	83	S	84	T	85	U	86	V	87	W
88	X	89	Y	90	Z	91	[	92	\	93	]	94	ˆ	95	_
96		97	a	98	b	99	c	100	d	101	e	102	f	103	g
104	h	105	i	106	j	107	k	108	l	109	m	110	n	111	o
112	p	113	q	114	r	115	s	116	t	117	u	118	v	119	w
120	x	121	y	122	z	123	{	124	|	125	}	126	˜	127	

Table 1: The ASCII character table. The code of blank is 32.

Logical operations. The basic logical operators operating on single bits are AND, OR and NOT. The unary operator NOT inverts the bit: NOT(0)=1 and NOT(1)=0. The definition of the operators AND and OR for single bits is shown below. Additional logical operators can be constructed form the basic operators. For example EOR (exclusive or) shown in the table too can be expressed by EOR(a, b) = OR(a, b)AND(NOT(AND(a, b))).

AND	0	1
0	0	0
1	0	1

OR	0	1
0	0	1
1	1	1

EOR	0	1
0	0	1
1	1	0

Logical functions. Under Linux the following functions are available in Fortran. Bits in a word are numbered $k = 0, 1, \ldots 31$. Functions like NOT, AND and OR act simultaneously on all bits of a word.

```
L  = NOT(M)         ! bit-wise NOT of M

L  = IOR(M1,M2)     ! logical bit-wise OR between M1 and M2
L  = IAND(M1,M2)    ! logical bit-wise AND between M1 and M2
L  = IEOR(M1,M2)    ! bit-wise exclusive OR between M1 and M2

L  = ISHFT(M,K)     ! shift of M logically left (K>0) or right (K<0) |K| bits

LL = BTEST(M,K)     ! logical test of bit K in integer M,
                    !    result .TRUE. or .FALSE. (type logical)

L  = IBCLR(M,K)     ! clear bit K in M
L  = IBSET(M,K)     ! set bit K in M
```

Arrays. Arrays with n dimensions are declared by $A(k_1, k_2, \ldots k_n)$, where the indices start either with 0 or 1, and increase in steps of 1. More general the index range may be i_j bis k_j, with

$$\text{declaration} \qquad A(i_1 : k_1, i_2 : k_2, \ldots i_n : k_n) \quad \text{mit} \quad i_j \le \text{Index}_j \le k_j \quad \text{für} \quad j = 1, 2, \ldots n\,.$$

The total number of elements in the declaration of A is $M = \prod_{j=1}^{n}(k_j - i_j + 1)$. There are two conventions for the linear index. The *lexical ordering* is used languages like C and Pascal; the element $A_{j_1\, j_2\, \ldots j_n}$ is store as the q-th element of the array with

$$q = 1 + \sum_{s=1}^{n-1} \left((j_s - i_s) \cdot \prod_{l=s+1}^{n} (k_l - i_l + 1) \right) + (j_n - i_n) \qquad \text{(row-wise)}\,. \tag{105}$$

Inverse lexical ordering is used in Fortran:

$$q = 1 + (j_1 - i_1) + \sum_{s=2}^{n} \left((j_s - i_s) \prod_{l=1}^{s-1} (k_l - i_l + 1) \right) \qquad \text{(column-wise)} . \tag{106}$$

6.2 Information

Consider the case of N symbols, which are *equal probable*. In order to select a particular symbol the set of N symbols is divided into two sets of each $N/2$ symbols: one unit of information is needed to decide which part contains the particular symbol. This process is repeated until the symbol is isolated. In total $\log_2 N = \lg N$ steps are necessary. The probability of a particular symbol is $p_N = 1/N$ and the information to select a symbol is

$$H = \lg N = -\lg p_N \tag{107}$$

Suppose an alphabet with N symbols or characters, which occur with probabilities p_i for $i = 1, 2, \ldots N$ with $sum_i p_i = 1$. The fundamental theorem of information theory says that strings that are independent random sequences of characters require, on average, at least

$$H = -\sum_{i=1}^{N} p_i \lg p_i \tag{108}$$

bits per character. H is called the entropy of the probability distribution and is the lower bound for the number of bits per character. Compared to the case of equal probabilities $p_i = 1/N$, where $H = -\sum_i (1/N) \lg(1/N) = \lg N$, any other set of probabilities has a smaller entropy. The entropy bound given by equation (108) would be achieved if each character i is encoded with a code of length $-\lg p_i$ bits.

A special case is the alphabet of binary symbols 0 and 1, where the probabilities $P(0) = p$ and $P(1) = 1-p$ are assumed. The mean information per bit in a stream of binary symbols the entropy is

$$H_2 = -p \lg p - (1-p) \lg(1-p) = S(p) \tag{109}$$

The function $S(p)$ is symmetric around its maximum at $p = 1/2$ (with $S(1/2) = 1$) and is zero for $p = 0$ and $p = 1$. If symbols are coded by binary numbers, the coding is efficient if the bits 0 and 1 occur in about the same number.

Huffmann coding. The entropy bound given by equation (108) would be achieved if each character i is encoded with a code of length $-\lg p_i$ bits (which is usually not an integer). A general method to determine a code which leads to a bit string of minimal length is called Huffmann encoding, where all probabilities are approximated by integer powers of 1/2.
The Huffmann encoding is as follows. Initially the list of all symbols is sorted according to decreasing probability. Then the following steps are repeatedly applied:

1. The last symbol gets a 0, the previous a 1;
2. the last two symbols are joined and their probability is added;
3. they are moved to the correct place.

Example: The case of five different symbols is shown below. The entropy bound according to equation (108) is $H = 1.4579$, while is would be $H = \lg 5 = 2.322$ in the case of equal probability.

char	p_i	code
E	0.42	0
O	0.30	11
A	0.12	100
I	0.09	1011
U	0.07	1010

Huffman coding (table) results in 2.02 bits on average. The string 1011111010 for example corresponds to characters IOU. Huffman code has the prefix property: no binary symbol is the begin of another symbol.

6.3 Analysis of algorithms

Complex problems are divided into smaller problems, which hopefully can be solved by standard algorithms. Important standard problems are *sorting* of objects according to the value of a key and *searching*

function	sorting	searching
$f(n) = \text{const.}$		hash search
$f(n) = \text{const.} \cdot \log n$		binary search
$f(n) = \text{const.} \cdot n$		sequential search
$f(n) = \text{const.} \cdot n \log n$	quicksort, heapsort	
$f(n) = \text{const.} \cdot n^2$	quicksort with sorted input	
$f(n) = \text{const.} \cdot 2^n$		

Table 2: Typical leading dependences on the number n of objects in algorithms. Examples of algorithms for sorting and searching are shown.

in a list. Algorithms usually apply to a set of objects of a certain size n. The dependence $f(n)$ of the computer time or the number of operations on the number n of objects for larger values of n, the *leading* dependence, is often an essential property of an algorithm. Examples for the typical dependences of algorithms are shown in Table 2.
Sometimes the function $f(n)$ depends critically on numerical values or certain conditions. Then the behaviour $A(n)$ in the average case or the worst case behaviour $W(n)$ is of interest. A sorting algorithm may have a special behaviour in the case of already sorted or partially sorted input.

6.4 Sorting

Sorting, i.e. rearranging the elements of a list into order, is a standard task. The order is determined by a key t_i, assigned to each element. Keys can be compared to determine which is larger or wether they are equal; the normal case is sorting into nondecreasing order:

$$t_i \leq t_{i+1} \qquad \text{nondecreasing order} \tag{110}$$

A sorting algorithm may require additional space; if the amount of extra space is constant the algorithm is called to work *in place*. Only internal sort is discussed (no external sort on files).
Sorting is related to many other algorithms like searching and data compression. To find the median of a set to data can also be done by sorting.

Selection sort. The algorithm is straightforward. The smallest element is searched for, and it is exchanged with the element in the first place; then the second smallest key is searched for and exchanged with the element in the second place, and so on. The number of comparisons is increasing with the number n as

$$A(n) = W(n) = (n-1) + (n-2) + \ldots + 2 + 1 = \frac{n(n-1)}{2} \approx \frac{n^2}{2}$$

The number of exchange operations is n at the maximum, and thus depends linearly on n.

Insertion sort. The algorithm is rather simple and natural, it is a in-place sort. Assume that the first elements are already sorted. One step of the algorithm is inserting the next element x in its proper place, by moving each key in the sorted part by one place until a key smaller or equal to x is found.
For the analysis the number of key comparisons is counted. The worst case $W(n)$ is when the keys are in reverse order.

$$W(n) = \frac{n(n-1)}{2} \propto n^2 \qquad\qquad A(n) \approx \frac{n^2}{2} \propto n^2$$

Shell sort. The algorithm (by Donald Shell) is easy to program, and it is rather fast. In the algorithm an increment m is defined. For an increment m there are m sublists, formed by the elements with index increment m; each sublist is sorted by insertion. Complete sorting is reached by a sequence of increments, ending with $m = 1$. Empirically increments $\ldots, 1093$, 364, 121, 40, 13, 4, 1 have been found to give a good performance; these increments follow, starting with $m_1 = 1$, from the relation

$$m_{s+1} = 3m_s + 1 \tag{111}$$

and the start is $s = t$, where t is the smallest index with $m_{t+2} \geq n$. The analysis of the algorithm is difficult; empirically the number of comparisons is $\propto n^{3/2}$.

Quicksort. The algorithm (by C.A.R. Hoare) is very fast for large n and can be defined recursively. One key, called the pivot, is selected from the keys. The other keys are divided (split) into two sets: one set is formed by all keys less than the pivot, and the second set is formed by all keys greater than or equal to the pivot. The same procedure is then applied recursively to the subsets. The algorithm has for large n an average behaviour $A(n) = \text{const} n \ln n$. The behaviour depends on the split: two subsets of about equal size are optimal. In the worst case Quicksort behaves poorly ($\propto n^2$); unfortunately sorted input is among the worst cases. Usually extra strategies are applied to avoid the worst-case behaviour. Quicksort is not an in-place sort; a stack depending on n is used for the book keeping on the sublists, with a worst case $\propto n$.

Heapsort. The algorithm (by J. Williams) has a behaviour of $A(n) = \text{const.} \cdot n \ln n$ and the worst case is only 20 % slower than the average. It is an in-place sort, and is usually the recommended algorithm for the general case.
The algorithm is based on a special data structure called *heap*. This is a set of n numbers t_i for $i = 1, 2, \ldots n$ that follow the relation

$$t_{\lfloor j/2 \rfloor} \geq t_j \qquad \text{for} \quad 1 \geq \lfloor j/2 \rfloor < j \leq n \tag{112}$$

($j/2$ is rounded to the next lower integer in the division $\lfloor j/2 \rfloor$). The element t_1 has thus the largest value. There are two stages in heapsort. First a heap data structure is constructed. Then sorting is done from the high-end, removing always the largest key from the heap, and re-establishing the heap again.

Merging sorted lists and merge sort. Two already sorted lists A and B with n and m elements can be combined to a sorted list C with $n + m$ elements efficiently by the following algorithm.

```
indexA=1;   indexB=1;   indexC=1;
while indexA =< n and indexB =< m do
   if A[indexA] < B[indexB]
      then   C[indexC] = A(indexA]; indexA = indexA + 1
      else   C[indexC] = B(indexA]; indexB = indexB + 1
   end if
   indexC = indexC + 1
end while
if indexB > m
   then  move A[indexA]...A[n]  to  C[indexC]...C[n+m]
   else  move B[indexB]...B[m]  to  C[indexC]...C[n+m]
end if
```

In the worst case the algorithm requires only $n + m - 1$ comparisons. The algorithm can be extended to do complete sorting by splitting a large list into small lists, which are sorted and then combined by merging. The worst-case behaviour is $W(n) = n \lg n$; a disadvantage of merge sort is the extra space required by the algorithm.

6.5 Searching

Sequential search. The simple sequential search in a list for a given key requires a time proportional to the number of elements. Some small improvements are possible to improve the search; for example often used keys can be placed early in the list.

Binary search. For a sorted list the binary search can be used, which requires never more the $1 + \lceil \lg n \rceil$ comparisons. In order to search for a given key x this key is compared to the element of the table with index $(n + 1)/2$. If $x < t_{(n+1)/2}$, further search is in this part of the table, otherwise in the second part of the table. This method is repeated for the reduced lists, dividing the length of the list each time by two. Binary search is illustrated in Figure 15.

Hash search. A fast search method with constant search time can be used, if the keys are integers in the range 0 to $(p - 1)$. In this case a table of length p can be used, where the key is directly taken as index. However large values of p require a lot of space.

Hashing is a generalization of the simple adressing method. A so-called hash-function is defined, which transforms the value of the key to an integer $0 \ldots (p - 1)$, which then is used a index for the table. The aim is to find a function with a small number of collisions: a collision is the case, where different keys are transformed to the same index. Collisions have to be handled by an additional algorithm, which for

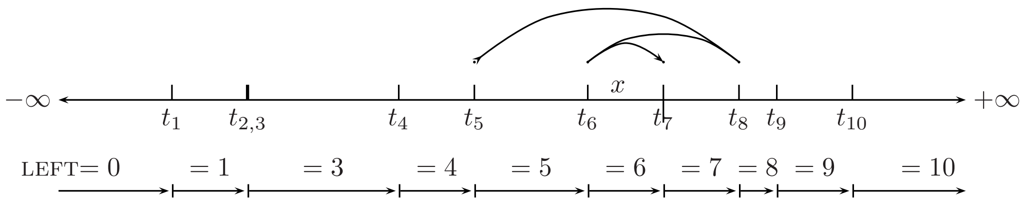

Figure 15: Binary search for the interval with $t_\ell \leq x < t_{\ell+1}$.

example may use linked lists. Essentially hashing divides a long list into p short lists, which may be handled sequentially. A large value of p with corresponding space reduces the search time, and usually a constant time can be reached.

For integer keys m the hash function can be realized by the modulo function $m \bmod p$, resulting in a number $j = 0 \ldots (p-1)$. A prime number p is usually preferred.
Often keys are alphanumerical strings, possible quite long, which can in principle be represented by an integer $\ldots + 256^2 \cdot \mathrm{ch}(3) + 256^1 \cdot \mathrm{ch}(2) + 256^0 \cdot \mathrm{ch}(1)$, but the integer may be too large for normal arithmetic, if the number k of characters `ch(.)` is too large. Taking advantage of the properties of the modulo function the following code can be used to avoid very large integers:

```
j = ch(1); for i=2 to k do j = j*256+ch(i) mod p
```

Glossary

(a) Probability and probability distributions

Bayesian probability. In Bayesian or subjective probability the elements of the sample space correspond to hypotheses. $\mathcal{H}$ is a hypothesis or proposition, which is either true or not true, and the probability $P(\mathcal{H})$ is interpreted as a measure of the *degree of belief*, that $\mathcal{H}$ is true. Bayesian probability is the basis of → Bayesian statistics (see also → Probability).

Bayesian statistics. Basis of the Bayesian statistics is the → Bayesian probability or subjective probability. In the Bayesian statistics Bayes' theorem takes the form of a proportionality

$$P(\mathcal{H}|\text{data}) \propto P(\text{data}|\mathcal{H}) \cdot P(\mathcal{H}) \ ,$$

where $P(\mathcal{H})$ represents the prior probability of hypothesis $\mathcal{H}$ being true, and $P(\text{data}|\mathcal{H})$ is the likelihood of observing the data under the assumption of the hypothesis $\mathcal{H}$. The posterior probability of the hypothesis $\mathcal{H}$ after having measured the data is $P(\mathcal{H}|\text{data})$.

There is no fundamental rule how to assign the prior probabilities to a hypothesis. Once an assignment has been made, the Bayesian statistics gives the change of the degree of belief after having measured the data (see also → Classical statistics, page 50).

Bayes theorem. Bayes' theorem[3] relates the → conditional probabilities $P(\mathcal{A}|\mathcal{B})$ and $P(\mathcal{B}|\mathcal{A})$. The theorem is derived from the basic probability formula

$$P(\mathcal{A} \text{ and } \mathcal{B}) = P(\mathcal{A}) \cdot P(\mathcal{B}) = P(\mathcal{A}) \cdot P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B}) \cdot P(\mathcal{A}|\mathcal{B})$$

and has the following form:

$$\text{Bayes' theorem} \qquad P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{B}|\mathcal{A}) \cdot P(\mathcal{A})}{P(\mathcal{B})}$$

For a sample space with disjoint subsets $\mathcal{A}_1, \ldots, \mathcal{A}_i, \ldots, \mathcal{A}_n$, i.e. $P(\mathcal{A}_i) \cdot P(\mathcal{A}_j) = 0$ for $i \neq j$, the total probability for $\mathcal{B}$ is

$$P(\mathcal{B}) = \sum_i P(\mathcal{B}|\mathcal{A}_i) \cdot P(\mathcal{A}_i)$$

and the Bayes theorem is written in the form

$$P(\mathcal{A}_i|\mathcal{B}) = \frac{P(\mathcal{B}|\mathcal{A}_i) \cdot P(\mathcal{A}_i)}{\sum_i P(\mathcal{B}|\mathcal{A}_i) \cdot P(\mathcal{A}_i)} \ .$$

Beta distribution. The Beta distribution depends on the positive integers n and m and is nonzero over the interval $0 < x < 1$ of the random variable x:

$$f(x) = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} x^{m-1} (1-x)^{n-1} \ .$$

Bias. The difference between the expectation value of an estimator $\widehat{a}$ (→ estimator, page 51) of a physical parameter a and the true value a_{true} is called bias:

$$B_n(\widehat{a}) = E[\widehat{a}] - a_{\text{true}} \ .$$

The bias depends on the size n of the sample, used in the estimator. The estimator is called *unbiased* if $E[\widehat{a}] = a_{\text{true}}$ for *all* n. The estimator is called *consistent*, if the bias $B_n(\widehat{a})$ vanishes asymptotically (for $n \to \infty$).

[3]Thomas Bayes, British reverend, 1702 – 1761

Binomial distribution. Consider an experiment which yields the event A ("success") with probability p (and the event $\overline{A}$ ("failure") with probability $q = 1 - p$). The Binomial distribution gives the probability of k successes (or events) in n trials

$$\boxed{P(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0,\, 1,\, 2 \ldots\, n} \,.$$

The probability that the event accurs in the first k trials is $p^k(1-p)^{n-k}$; this ordering is only one out of $\binom{n}{r} = n!/\left(k!(n-k)!\right)$ ways with the same number k of events in n trials. The normalization follows from the Binomial theorem (→ Binomial theorem, page 60). The Binomial distribution is determined by the two parameters n and p, the random variable r has the mean value np and a variance of $\sigma^2 = np(1-p)$.

The maximum likelihood estimator of the parameter p and its variance is, for a sample of n events with k successes, given by

$$\widehat{p} = \frac{k}{n} \qquad\qquad \sigma^2(\widehat{p}) = \frac{\widehat{p}(1-\widehat{p})}{n}$$

The latter estimate is biased for finite n; an unbiased estimate of the variance is $\sigma^2(\widehat{p}) = \widehat{p}(1 - \widehat{p})/(n-1)$.

Breit Wigner distribution. The distribution, also called Lorentz distribution, is the Fourier transform of the → exponential distribution with the relation $\tau\Gamma = 1$ between the mean value τ of the exponential distributin and the width parameter Γ of the Breit Wigner distribution

$$\boxed{p(x;\, x_0, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(x-x_0)^2 + \Gamma^2/4}} \,.$$

The parameter x_0 is the mode of the distribution and Γ is the full width at half maximum (FWHM) (see also → Cauchy distribution)

Cauchy distribution. The Cauchy distribution is symmetric around $x = 0$ and is defined by

$$\boxed{p(x) = \frac{1}{\pi} \frac{1}{1+x^2}} \,.$$

Mean, variance and higher moments of the Cauchy distribution are undefined, because the integrals diverge; the Cauchy distribution is the example of a density with long tails. Mode and the median are zero, the full width at half maximum (FWHM) is two. A more general form of the Cauchy distribution is the → Breit Wigner distribution, page 49.

If a variable φ has an uniform distribution between $-\pi/2$ and $+\pi/2$, then the variable $x = \tan\varphi$ follows the Cauchy distribution. If y and z follow (independent) standard Gaussian distributions $N(0,1)$, then the ratio $x = y/z$ follows the Cauchy distribution.

χ^2 distribution. If $x_1, \ldots, x_i, \ldots, x_n$ are independently distributed according to the standardized Gaussian distribution $N(0,1)$, then

$$u = \sum_{i=1}^{n} x_i^2$$

follows a χ^2 distribution with n degrees of freedom. The density is given by

$$f_n(u)\, \mathrm{d}u = \frac{\frac{1}{2}\left(\frac{u}{2}\right)^{n/2-1} e^{-u/2}}{\Gamma(n/2)}\, \mathrm{d}u \,.$$

Expectation value and variance are:

$$E[u] = n \qquad\qquad V[u] = 2n$$

The χ^2 distribution approaches the Gaussian distribution with mean n and variance $2n$ in the limit $n \to \infty$. The transformed variable

$$t = \sqrt{2u} - \sqrt{2n-1}$$

approaches the standard Gaussian distribution faster, and this variable can be used in comparing χ^2-distributed values with different values of n.

If $x_1, \ldots, x_i, \ldots, x_n$ are *independently* Gaussian distributed with means μ_i and variances σ_i^2, then the quantity

$$u = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

follows a χ^2 distribution with n degrees of freedom. The generalization to correlated Gaussian variables with a covariance matrix $\boldsymbol{V}$ is written in matrix notation as quadratic form

$$u = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{V}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \ .$$

Classical statistics. ($\to$ Bayes statistics, page 48)

Basic tasks of classical statistics are to estimate probabilities, given a finite set of measured data, and to test wether probabilities, predicted by a certain model or theory, are compatible with the measured data. It is based on the interpretation of probability as a relative frequency.

Conditional distribution. Given a joint probability function $p(x_1, x_2)$ of random variables x_1, x_2, the conditional probability density function of x_2 is the probability density function of x_2 for a given fixed value of x_1. The variable x_1 is treated as a constant. The conditional probability density function for the random varibale x_2, given x_1, written in the form $q(x_2|x_1)$, is defined as

$$q(x_2|x_1) = \frac{p(x_1, x_2)}{\int p(x_1, x_2)\, \mathrm{d}x_2}$$

The density is renormalized such that its area is unity when integrated over x_2. The generalization to the case of a joint probability function of more variables is straightforward.

Conditional probability. Consider a sample space with subsets $\mathcal{A}$ and $\mathcal{B}$. The conditional probability of $\mathcal{B}$, given $\mathcal{A}$, is defined as

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

Confidence interval. The statistical error of an estimate $\widehat{a}$ can be expressed by a confidence interval for a given $\to$ confidence level with probabilities α and β. Suppose that the estimate $\widehat{a}$ has been obtained from data by an estimator $\widehat{a}(x_1, \ldots, x_i, \ldots, x_n$, and that the probability density function $g(\widehat{a}; a)$, depending on the true value a of the parameter, is known. The limits of the confidence interval are called a_{lower} and a_{upper} and can be determined by solving equations

$$\alpha = \int_{\widehat{a}}^{+\infty} g(\widehat{a}; a_{\text{lower}})\, \mathrm{d}\widehat{a} \qquad\qquad \beta = \int_{-\infty}^{\widehat{a}} g(\widehat{a}; a_{\text{upper}})\, \mathrm{d}\widehat{a}$$

for a_{lower} and a_{upper} (one of the limits is the observed value $\widehat{a}$). The confidence interval is called *central* for $\alpha = \beta = 1/2\,\gamma$.

One-sided confidence intervals represent limits. The value a_{lower} is a lower limit on the parameter a such that $a_{\text{lower}} \leq a$ with probability $1 - \alpha$, and the value a_{upper} is an upper limit on the parameter a such that $a \leq a_{\text{upper}}$ with probability $1 - \beta$.

Confidence level. The confidence level is a probability associated with the comparison of observed values with a given probability density function. The confidence level corresponds to the integral over the probability density function from the observed value to (minus or plus) infinity ($\to$ Confidence interval, page 50, and $\to$ Goodness-of-fit, page 55).

Correlation coefficient. The correlation coefficient ρ_{jk} between two (correlated) random variables x_j and x_k is defined by

$$\rho_{jk} = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}} = \frac{\mathrm{cov}\,(x_j, x_k)}{\sigma_j \sigma_k} \qquad\qquad -1 \leq \rho_{jk} \leq +1$$

If $|\rho_{jk}| = 1$, the variables x_j and x_k are linearly dependent; in this case the covariance matrix is singular (→ Variance, page 59, → Covariance, page 51, and → Covariance matrix, page 51).

Covariance. The Covariance $\mathrm{cov}(x_j, x_k)$ between two random variables x_j and x_k is equal to the element V_{jk} of the covariance matrix, and can be written in terms of expectation values:

$$\mathrm{cov}(x_j, x_k) = E\left[(x_j - \langle x_j \rangle)\,(x_k - \langle x_k \rangle)\right]$$

For independent variables (→ Independence, page 55) the covariance vanishes, but the converse is not true.

Covariance matrix. For a n-dimensional random variable $\boldsymbol{x}$ the symmetric $n \times n$ covariance matrix is defined by

$$\boldsymbol{V}\left[\boldsymbol{x}\right] = \int_\Omega \left(\boldsymbol{x} - \langle \boldsymbol{x} \rangle\right)^T \left(\boldsymbol{x} - \langle \boldsymbol{x} \rangle\right) p(\boldsymbol{x})\, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n$$

where $p(\boldsymbol{x})$ is the probability density of the random variable $\boldsymbol{x}$, with elements

$$\begin{aligned} V_{jj} = \sigma_j^2 \quad &= \quad \int_\Omega \left(x_j - \langle x_j \rangle\right)^2 p(\boldsymbol{x})\, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \\ V_{jk} = \sigma_{jk} \quad &= \quad \int_\Omega \left(x_j - \langle x_j \rangle\right)\left(x_k - \langle x_k \rangle\right) p(\boldsymbol{x})\, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \; . \end{aligned}$$

The elements σ_j^2 are called variances (→ Variance, page 59), and the elements σ_{jk} are called covariances (→ Covariance, page 51); the covariance can be expressed by $\sigma_{jk} = \rho_{jk}\sigma_j\sigma_k$ in terms of the correlation coefficient ρ_{jk} (→ correlation coefficient, page 51) with the range $-1 < \rho_{jk} < +1$.

Error propagation. Given a random varibale $\boldsymbol{x}$ with covariance matrix $V_{\boldsymbol{x}}$. For a variable $\boldsymbol{y}$, related to $\boldsymbol{x}$ by a linear transformation $\boldsymbol{y} = B\boldsymbol{x}$, the covariance matrix $V_{\boldsymbol{y}}$ is given by

$$\boldsymbol{V_y} = \boldsymbol{A V_x A}^T \; .$$

In case of a nonlinera transformation ann approximation for $V_{\boldsymbol{y}}$ is calculated by the same formula whre the elements B_{ik} are the derivatives $\partial y_i / \partial x_k$, calculated at the position of mean value.

Estimator. An estimator for a parameter a is written as $\widehat{a}$, a function of the data $x_1, \ldots, x_i, \ldots, x_n$, and is itself a random variable.

An estimator $\widehat{a}(x_1, \ldots, x_i, \ldots, x_n)$ is called *consistent*, if it converges to the true value in the limit of large n (asymptotic limit). A consistent estimator is asymptotically unbiased.

The mean squared error of an estimate is

$$V\left[\widehat{a}\right] = E\left[\left(\widehat{a} - a_{\text{true}}\right)^2\right] = E\left[\left(\widehat{a} - E\left[\widehat{a}\right]\right)^2\right] + \left(E\left[\left(\widehat{a} - a_{\text{true}}\right]\right)\right)^2 = V\left[\widehat{a}\right] + B^2$$

where B ist the bias (→ Bias).

An estimator is called *efficient*, if its variance attains the minimum variance bound given by the → Cramer-Rao inequality.

Expectation value. The expectation value of the continuous random variable x (or mean value, population mean) with probability density function $p(x)$, and a discrete random variable with probabilities $P(x)$, respectively, is defined by

$$E\left[x\right] = \int_\Omega x\, p(x)\, \mathrm{d}x \qquad\qquad E\left[x\right] = \sum_i x_i\, P(x_i) \; .$$

The expectation value for a function $g(x)$ of a random variable is defined by

$$E\left[g(x)\right] = \int_\Omega g(x)\, p(x)\, \mathrm{d}x \qquad E\left[g(x)\right] = \sum_i g(x_i)\, P(x_i)\ .$$

Expectation values are also denoted by $\langle x\rangle$ and $\langle g(x)\rangle$, respectively.

The expectation value is a linear operator, for example

$$E\left[a\cdot g(x) + b\cdot h(x)\right] = a\cdot E\left[g(x)\right] + b\cdot E\left[h(x)\right]\ .$$

Exponential distribution. The exponential distribution of the continuous random variable x is defined by

$$p(x;\,\tau) = \frac{1}{\tau}e^{-x/\tau} \qquad 0\le x<\infty$$

and depends on a single parameter τ, which is equal to the mean and the standard deviation of the random variable:

$$E\left[x\right] = \tau \qquad \sigma = \sqrt{V\left[x\right]} = \tau$$

An example for a random varibale with an exponential distribution is the decay time of an unstable particle, where the parameter τ is the mean decay time. The transformed parameter $\lambda = 1/\tau$ is the decay probability.

The exponential distribution can also describe attenuation phenomena, where the parameter τ has the meaning of an attenuation length.

F distribution. Let s_1^2 and s_2^2 be estimates of the two variances σ_1^2 and σ_2^2 with n_1 and n_2 degrees of freedom. In this case the ratio $n_1 s_1^2/\sigma_1^2$ is distributed as χ^2 with n_1 degrees of freedom and $n_2 s_1^2/\sigma_2^2$ is distributed as χ^2 with n_2 degrees of freedom. The random variable F, defined by

$$F = \frac{s_1^2}{\sigma_1^2}\,\frac{\sigma_2^2}{s_2^2}$$

follows the F distributon with (n_1, n_2) degrees of freedom. The density is given by

$$p(F;\,n_1,n_2) = \left(\frac{n_1}{n_2}\right)^{n_1/2}\cdot\frac{\Gamma\left((n_1+n_2)/2\right)}{\Gamma\left(n_1/2\right)\Gamma\left(n_2/2\right)}\cdot F^{(n_1-2)/2}\left(1+\frac{n_1}{n_2}F\right)^{-(n_1+n_2)/2}$$

Gamma distribution. The Gamma distribution is given by the expression:

$$\boxed{f(x|k) = \frac{x^{k-1}e^{-x}}{\Gamma(k)}}\ .$$

For an integer value of k with $k\ge 1$ (order k) the Gamma distribution gives the waiting time to the k-th event in a Poisson random process of unit mean ($\lambda = 1$); however the parameter k is not required to be an integer. The generalization for other values of the parameter λ is:

$$\boxed{f(x|k,\lambda) = \frac{x^{k-1}\lambda^k e^{-\lambda x}}{\Gamma(k)}}\ .$$

The parameter k affects the *shape* of the distribution, while the parameter λ is merely a scale parameter. For $\lambda = 1/2$ and $k = n/2$ the Gamma distribution becomes the χ^2-distribution.

Gaussian distribution. The Gaussian (or normal) distribution of the continuous random variable x is defined by

$$\boxed{f(x;\,\mu,\sigma^2)\,dx = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad x\in(-\infty,+\infty)}$$

and has two parameters, the mean μ and the variance σ^2; it is also denoted $N(\mu,\sigma^2)$.

The distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called standard Gaussian, and denoted $N(0,1)$. The cumulative distribution of the Gaussian can be expressed by the Gaussian error function $\mathrm{erf}(x)$, defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,\mathrm{d}t$$

with $\mathrm{erf}(0) = 0$ and $\mathrm{erf}(\infty) = 1$. The cumulative Gaussian distribution is

$$\Phi(x) = \int_{-\infty}^{x} f(x;0,1)\,\mathrm{d}x = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right) .$$

Some values of the cumulative Gaussian distribution for values $x = 0 \ldots 3.49$ are shown below in a table.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Gaussian distribution, bivariate. The twodimensional of bivariate Gaussian density for two random variables x_1 and x_2 is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

Marginal distributions. From this two-dimensional density the marginal distribution for x_1

$$p(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2)\,\mathrm{d}x_2 = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}$$

and the marginal distribution for x_2

$$p(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2)\,\mathrm{d}x_1 = \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}$$

follow by integration over the other variable. The mean values of the two marginal distributions are μ_1 and μ_2, and the standard deviations are σ_1 und σ_2.

Conditional distributions. The conditional distributions depend on the correlation coefficient, defined by

$$\rho = \frac{E\left[(x_1-\mu_1)(x_2-\mu_2)\right]}{\sigma_1\sigma_2} \qquad\qquad -1 < \rho < +1$$

The conditional distribution of the random variable x_1 for a given value of the second variable x_2

$$p(x_1|x_2) = \frac{f(x_1,x_2)}{p(x_2)} = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_1-\mu_1}{\sigma_1} - \rho\frac{x_2-\mu_2}{\sigma_2}\right]^2\right\}$$

and the conditional distribution of the random variable x_2 for a given value of the first variable x_1

$$p(x_2|x_1) = \frac{f(x_1,x_2)}{p(x_1)} = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_2-\mu_2}{\sigma_2} - \rho\frac{x_1-\mu_1}{\sigma_1}\right]^2\right\}$$

follow. The mean values of the conditional distributions

$$\begin{aligned} E\left[x_1|x_2\right] &= \int_{-\infty}^{+\infty} x_1\, p(x_1|x_2)\,\mathrm{d}x_1 = \mu_1 + \rho\sigma_1\left(\frac{x_2-\mu_2}{\sigma_2}\right) \\ E\left[x_2|x_1\right] &= \int_{-\infty}^{+\infty} x_2\, p(x_2|x_1)\,\mathrm{d}x_1 = \mu_2 + \rho\sigma_2\left(\frac{x_1-\mu_1}{\sigma_1}\right) \end{aligned}$$

depend linearly on the correlation coefficient ρ. The variances of the conditional distributions are equal respectively to $\sigma_1^2(1-\rho^2)$ and $\sigma_2^2(1-\rho^2)$. The conditional means in the x_1, x_2 plane are straight lines with the slopes (dx_2/dx_1) of respectively $1/\rho(\sigma_2/\sigma_1)$ and $\rho(\sigma_2/\sigma_1)$. For values of $|\rho|$ close to one the two slopes are almos identical.

Lines of constant probability density. Lines of constant probability density of the two-dimensional distribution correspond to constant values of the exponent and are are ellipes. If the constant is choosen to be one, the condition is

$$\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\frac{(x_2-\mu_2)^2}{\sigma_2^2} = 1-\rho^2$$

The probability to observe a point (x_1, x_2) inside this error ellipse is $1 - 1/\sqrt{e} = 0.3935$. To find a point for a given probability P (x_1, x_2) inside an ellipse the corresponding value of the χ^2-distribution with two degrees of freedom is $\chi^2_P = 2\ln(1-P)$. The axes of the ellipse are inclined to the x_1-axis at the angles φ, defines by $\tan 2\varphi = \pm 2\rho\sigma_1\sigma_2/(\sigma_1^2-\sigma_2^2)$. The relation

$$\rho = \pm(2-\varepsilon^2)^{-1}\sqrt{\varepsilon^4 - (1-\varepsilon^2)(\sigma_1/\sigma_2 - \sigma_2/\sigma_1)^2}$$

is valid for the excentricity ε of the ellipse.

Gaussian distribution, multivariate. The Gaussian distribution of a n-dimensional random variable $\boldsymbol{x}$ with elements $x_1, \ldots, x_i, \ldots, x_n$ is defined by

$$\boxed{f(\boldsymbol{x};\, \boldsymbol{\mu}, \boldsymbol{V}) = \frac{1}{(2\pi)^{n/2}\left|\boldsymbol{V}\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{V}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\}}$$

and depends on the parameters $\boldsymbol{\mu}$ and $\boldsymbol{V}$.
The n-dimensional vector $\boldsymbol{\mu}$ with elements $\mu_1, \ldots, \mu_i, \ldots, \mu_n$ is the vector of mean values, and the symmetric $n \times n$-matrix $\boldsymbol{V}$ is the covariance matrix:

$$\boldsymbol{\mu} = E\left[\boldsymbol{x}\right] \qquad\qquad \boldsymbol{V} = V\left[\boldsymbol{x}\right] = E\left[\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T\right]$$

The n-dimensional Gaussian distribution is the model for the distribution of correlated variables, since its parameters are just the mean values, the variances and the covariances.

Global correlation coefficient. The (unsigned) largest correlation coefficient between the random variable x_j and a linear combination of all other variables is called global correlation coefficient and is given by

$$\rho_j = \sqrt{1 - \left[(\boldsymbol{V})_{jj} \cdot \left(\boldsymbol{V}^{-1}\right)_{jj}\right]^{-1}} \qquad 0 \leq \rho_j \leq 1 \, .$$

It measures the total amount of correlation between the variable x_j and all other variables. If $\rho_j = 0$, the variable x_j is uncorrelated with all other variables, and in the limit $\rho_j \to 1$, the variable x_j is completely correlated with at least one linear combination of the other variables.

Goodness-of-fit test. For a goodness-of-fit test, testing a given hypothesis, a test statistic is defined, which is a function of the data, measuring the distance between the data and a given hypothesis. The distance is converted to a probability P of obtaining data with a larger value of this test statistic than the observed value, assuming the hypothesis is true. The value of P is called the → confidence level of the fit. A small value of P indicates a low level of confidence in the hypothesis. The most common test for goodness-of-fit is the χ^2-test.

Independence. Consider a sample space with subset $\mathcal{A}$ and $\mathcal{B}$. The two subsets are called independent if

$$P(\mathcal{A} \text{ and } \mathcal{B}) = P(\mathcal{A}) \cdot P(\mathcal{B}) \, .$$

Information. The amount of *information* about a parameter a given by a sample $x_1, \ldots, x_i, \ldots, x_n$ of measured values from a distribution $p(x|a)$ is given by the information I defined by the expection value

$$I(a) = E\left[\left(\frac{\partial}{\partial a} \ln \prod_{i=1}^{n} p(x_i|a)\right)^2\right] \, .$$

The inverse of the information is the minimal variance of any estimate $\widehat{a} = \widehat{a}(x_1, \ldots, x_i, \ldots, x_n)$.

Jacobian peak. A peak in a probability density can be caused by the variation of the Jacobi determinant. An example is a cross section which, as a function of tranverse momentum squared, falls off exponentially:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}p_T^2} = \frac{1}{2p_T}\frac{\mathrm{d}\sigma}{\mathrm{d}p_T} = b\, e^{-ap_T} \, .$$

As a function of the transverse momentum p_T the distribution is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}p_T} = 2b\, p_T\, e^{-ap_T} \, ,$$

which has a peak at $p_T = 1/a$.

Marginal distribution. Given a joint probability function $p(x_1, x_2)$ of random variables x_1, x_2, the marginal probability density of the random variable x_1 is the density regardless of the value of x_2. The marginal probability density of x_1 is obtained by integrating the $p(x_1, x_2)$ over x_2:

$$p_{x_1}(x_1) = \int_{-\infty}^{+\infty} p(x_1, x_2)\, \mathrm{d}x_2$$

and corresponds to a projection of $p(x_1, x_2)$ onto the x_1-axis.

Median. The median $x_{1/2}$ of a probability density function is the → quantile of order 50 %. The probabilities for x to be observed greater of less than $x_{1/2}$ are equal.

For a random sample $x_1, \ldots, x_i, \ldots, x_n$ a practical method for the determination of the median is to sort the x-values and then to define

$$\text{median} = \begin{cases} x_{(n+1)/2} & \text{if } n \text{ is odd} \\ \left(x_{n/2} + x_{(1+n/2)}\right)/2 & \text{if } n \text{ is even} \end{cases}$$

The median can be used as a measure of the location of the random variable, which is rather insensitive to long tails of the distribution.

Mode. The mode of a probability density function is the value of the random variable x at which the probability density function is a maximum. A probability density function may have local maxima.

Moment. The (central) moment of order m, or the moment of order m abut the mean $E[x] = \langle x \rangle$ of a random variable x is defined by the expectation value

$$\mu = E\left[(x - \langle x \rangle)^m\right]$$

In the case of a multidimensional random variable $\boldsymbol{x} = (x_1, x_2, \ldots)$ the moment of order $(m, n, \ldots)$ about the mean $\langle \boldsymbol{x} \rangle$ is defined by the expectation value

$$\mu = E\left[(x_1 - \langle x_1 \rangle)^m (x_2 - \langle x_2 \rangle)^n \ldots\right]$$

Multinomial distribution. The multinomial distribution is a generalization of the binomial distribution with n (more than two) classes into which an event can fall. If p_i is the probability of an individual event belonging to the i-th class, the probability of exacly k_i events belonging to class i for each i is given by the multinomial distribution

$$P(k_1, k_2, \ldots k_n) = \frac{N!}{p_1! p_2! \cdots p_n!} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n}$$

with the normalization $\sum_i p_i = 1$ and with $\sum_i k_i = N$. Expectation value and variance of the k_i are $E[k_i] = Np_i$ and $V[k_i] = Np_i(1 - p_i)$; the correlation between k_i and k_j is $\rho_{ij} = -\sqrt{p_i p_j / (1 - p_i)(1 - p_j)}$.

A common example for the multinomial distribution is a histogram containing a fixed number N of independent events distributed into n bins. There is a correlation between bin contents because the sum is fixed to N. For $p_i \ll 1$ the variance $Np_i(1 - p_i)$ of k_i can be approximated by the Poisson variance Np_i. If the total number N of independent events is itself a Poisson random number, then all k_i follow individual Poisson distributions and the correlation vanishes.

Point Binomial distribution. Consider a process with two possible results in a trial, called "success" and "failure" (for example "head" or "tail", when a coin is tossed). The random variable x can take on two different values $x = 1$ and 0, associated to the two outcomes "success" and "failure". The probability can be written in the form

$$\boxed{P = p^x (1 - p)^{(1-x)} \qquad x = 0,\ 1}$$

called point binomial, which depends on one parameter p, with $0 \leq p \leq 1$. The probability for $x = 1$ is p, and the probability for $x = 0$ is $(1 - p)$.

Poisson distribution. The Poisson distribution gives the probability of exactly k events, if the number N of trials is very large and the probability p of the event to occur in a trial is very small, with a finite average value of $\mu = Np$. The Poisson distribution can be derived as the limiting form of the Binomial distribution (→ Binomial distribution, page 49) and has only one parameter μ, which is equal to the mean value *and* equal to the variance.

$$\boxed{P_\mu(k) = \frac{\mu^k e^{-\mu}}{k!} \qquad k = 0,\ 1,\ 2,\ \ldots} \ .$$

It is easy to see that the Poisson distribution is correctly normalized:

$$\sum_{k=0}^{\infty} P_\mu(k) = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{-\mu} e^{+\mu} = 1$$

Poisson distribution as limiting form of the Binomial distribution. From the Binomial distribution the probability for k successes in N trials, where p is the probability for a success in 1 trial, is given by the formula

$$P(k) = \binom{N}{k} p^k (1 - p)^{N-k} \ .$$

with an average (finite) number of $\mu = Np$. Expressing p by the formula μ/N the formula becomes

$$P(k) = \frac{N(N-1)\cdots(N-k+1)}{k!}\left(\frac{\mu}{N}\right)^k\left(1-\frac{\mu}{N}\right)^{N-k}$$
$$= \frac{N(N-1)\cdots(N-k+1)}{N^k}\frac{\mu^r}{k!}\left(1-\frac{\mu}{N}\right)^{N-k}$$

and by reordering the formula becomes

$$P(k) = \frac{\mu^k}{k!}\left[1\cdot\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{k-1}{N}\right)\frac{(1-\mu/N)^N}{(1-\mu/N)^k}\right] .$$

Now the number N is increased and the probability p is decreased, keeping the average $\mu = Np$ at a fixed value. In the limit $N \to \infty$ for one factor in the parenthesis $\lim_{N\to\infty}(1-\mu/N)^N = e^{-\mu}$ and the other factors become 1, thus in this limit

$$\lim_{N\to\infty} P(k) = \frac{\mu^k}{k!}e^{-\mu}$$

the Poisson probability is obtained.

Probability. In the analysis of experimental data probability is interpreted as a limiting relative frequency. Elements of the sample space are the possible outcomes of a measurement and an elementary outcome or event $\mathcal{A}$ corresponds to the occurance of this single outcome. The probability $P(\mathcal{A})$ is the fraction of times event $\mathcal{A}$ occurs in repeated measurements. The assigned probabilities can be tested in experiments repeated under identical conditions.

The concept of probability as a relative frequency can not be applied to unique phanomena, where repeated tests are impossible. In these cases the probability expresses the *degree of belief* about the phenomena and is subjective ($\to$ Bayesian probability).

Quantile. The quantile of order α is defined as the value x_α of the random variable x such that

$$\alpha = F(x_\alpha) = \int_{-\infty}^{x_\alpha} p(x)\,\mathrm{d}x$$

with $0 \le \alpha \le 1$.

Random variable. A variable that takes on certain values with given probabilities is a random variable. A random variable is called continuous, if the value can take on value from a continuum (real numbers), and the distribution are described by a probability density function $p(x)$. A random variable k is called discrete, if it can take on a finite or infinite number of discrete values, which can be numbered $k_1, k_2, \ldots k_i, \ldots$; the probabilities are given by values P_i.

Sample. A sample $x_1, \ldots, x_i, \ldots, x_n$ of size n from a probability density function $p(x)$ or from a discrete distribution.

Sample mean and variance. For a sample $x_1, \ldots, x_i, \ldots, x_n$ of size n the arithmetic mean of the elements is called sample mean and denoted by $\overline{x}$; it is an estimate of the expectation value of x. The sample mean $\overline{x}$ and the sample variance s^2 of a sample of n values $x_1, \ldots, x_i, \ldots, x_n$ is defined and calculated by the formulas

$$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i = x_{\text{est}} + \frac{1}{n}S_x$$
$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2 = \frac{1}{n-1}\left\{S_{xx} - \frac{1}{n}S_x^2\right\}$$

where S_x and S_{xx} are the sums

$$S_x = \sum_{i=1}^n (x_i - x_{\text{est}}) \qquad S_{xx} = \sum_{i=1}^n (x_i - x_{\text{est}})^2 .$$

The quantity x_{est} is a guess for the sample mean $\overline{x}$, for example $x_{\text{est}} = x_1$, introduced to avoid roundoff-errors.

The sample mean $\overline{x}$ is a consistent estimator for the population mean μ, if the variance of x exists (weak law of large numbers). Sample mean is an unbiased estimate of the population mean μ;

$$E\left[\overline{x}\right] = E\left[\frac{1}{n}\sum_i x_i\right] = \frac{1}{n}\sum_i E\left[x_i\right] = \frac{1}{n}\sum_i \mu = \mu$$

The sample variance, containing the factor $1/(n-1)$, is an unbiased estimate of the population variance σ^2.

For a sample of multidimensional random variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_n$ the generalization of the n sample means is straightforward. The quantities

$$\widehat{V}_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\left(\boldsymbol{x}_i\right)_j - \left(\overline{\boldsymbol{x}}\right)_j\right)\left(\left(\boldsymbol{x}_i\right)_k - \left(\overline{\boldsymbol{x}}\right)_k\right)\right]$$

are unbiased estimates for the covariance between elements j and k of the random variables $\boldsymbol{x}$.

Standard deviation. The standard deviation of a random variable with a probability density function $p(x)$, symbol σ, is defined as the square root of the variance:

$$\sigma = \sqrt{V\left[x\right]}$$

The actual probability content P corresponding to a deviation of k times the standard deviation from the mean of the distribution,

$$P = P\left(\langle x\rangle - k\sigma < x < \langle x\rangle + k\sigma\right)$$

depends on the specific probability density function $p(x)$. The Tschebyscheff inequality gives a (weak) limit on this probability, which is valid for any distribution (→ Tschebyscheff inequality, page 61). The following values are valid for the Gaussian density:

k	P in %
1	31.73
2	4.55
3	0.27
4	6.3×10^{-3}
5	5.7×10^{-5}
6	2.0×10^{-7}

k	P in %
0.67	50
1.28	20
1.64	10
1.96	5
2.58	1
3.29	0.1
3.89	0.01

Student's distribution. The t distribution is used for statistical tests of sample means. The distribution of the random variable t for n degrees of freedom is given by

$$p_n\left(t\right) = \frac{1}{\sqrt{n\pi}}\frac{\Gamma\left(\left(n+1\right)/2\right)}{\Gamma\left(n/2\right)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \qquad -\infty < t < +\infty$$

Student's distribution is used for a test wether an observed mean $\overline{x}$ from a sample of size n is consistent with a theoretical value μ, with a s determined from $s^2 = \sum_i (x_i - \overline{x})^2/(n-1)$. The correct statistics for this test is

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

which follows a Student's distribution with $(n-1)$ degrees of freedom.

Transformation of random variable.

Single variable. Given the probability density $p(x)$ of a random variable x. If the variable x is transformed to a new random variable $y = y(x)$, the probability density $q(y)$ of y is given by

$$q(y) = p\left(x\left(y\right)\right)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|$$

where $x = x(y)$ is the function inverse to $y = y(x)$. The shape of the density is changed by a transformation, unless the transformation is linear. The transformation has to be done for each branch for a transformation, that does not have a unique inverse.

Several variables. A linear transformation ist given by $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$, where the n-vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu_x}$ and covariance matrix $\boldsymbol{V_x}$ is transformed to the m-vector $\boldsymbol{y}$ by the $m \times n$ matrix $\boldsymbol{B}$. The dimensions n and m may be different; the matrix elements correspond to $B_{ik} = \partial y_i/\partial x_k$. The vector of mean values $\boldsymbol{\mu_y}$ and the covariance matrix $\boldsymbol{V_y}$ are given by

$$\langle \boldsymbol{y} \rangle = \boldsymbol{\mu_y} = \boldsymbol{B}\boldsymbol{\mu_x} \qquad \boldsymbol{V_y} = \boldsymbol{B}\boldsymbol{V_x}\boldsymbol{B}^T$$

The transformed probability density in the special case $n = m$ is given by

$$q(\boldsymbol{y}) = p\left(\boldsymbol{x}(\boldsymbol{y})\right) \det |\boldsymbol{B}^{-1}|$$

with the inverse transformation $\boldsymbol{x} = \boldsymbol{x}\left(\boldsymbol{y}\right) = \boldsymbol{B}^{-1}\boldsymbol{y}$.

In the general (nonlinear) case the transformation is defined by equations of the form

$$y_i = y_i(x_1,\, x_2 \ldots x_n) = y_i(\boldsymbol{x})$$

In the case $n = m$ the probility density of $\boldsymbol{y}$ is given by

$$q(\boldsymbol{y}) = p(\boldsymbol{x}) \cdot J\left(\frac{x}{y}\right) \qquad J\left(\frac{x}{y}\right) = \left| \frac{\partial x_1 \partial x_2 \ldots \partial x_n}{\partial y_1 \partial y_2 \ldots \partial y_n} \right|$$

Only an approximate transformation of the covariance matrix is possible in the general case of a nonlinear transforamtion of the n-vector $\boldsymbol{x}$ to the m-vector $\boldsymbol{y}$ by the formula (→ Error propagation)

$$\boldsymbol{V_y} = \boldsymbol{B}\boldsymbol{V_x}\boldsymbol{B}^T$$

where the elements $B_{ik} = \partial y_i/\partial x_k$ are calculated at the position of the mean $\boldsymbol{\mu_y}$.

Uniform distribution. The uniform distribution is constant over a finite range and zero outside:

$$\boxed{p(x;\, a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{elsewhere .} \end{cases}}$$

The distribution with limits a and b is also denoted $U(a, b)$. Mean and variance are

$$\mu = E\left[x\right] = \frac{a+b}{2} \qquad \sigma^2 = V\left[x\right] = \frac{(b-a)^2}{12}$$

The uniform distribution $U(0, 1)$ with limits 0 and 1, denoted $U(0, 1)$, has mean 1/2 and standard deviation $\sqrt{1/12} = 0.2887$. Variables from $U(0, 1)$ are generated by standard → random number generators.

Variance. The variance of a continuous random variable x is defined by

$$V\left[x\right] = E\left[\left(x - \langle x \rangle\right)^2\right] = \int_\Omega \left(x - \langle x \rangle\right)^2 p(x)\, \mathrm{d}x$$

where $p(x)$ is the probability density of the random variable x. It is the second moment about the mean value $\langle x \rangle = E\left[x\right]$ and equal to the square of the standard deviation σ (→ standard deviation, page 58).

$$V\left[a \cdot x\right] = a^2 \cdot V\left[x\right] \qquad V\left[x\right] = E\left[x^2\right] - \langle r \rangle^2$$

For a discrete random variable k the definition of the variance is

$$V\left[k\right] = \sum_i \left(k_i - \langle k \rangle\right)^2 P(k_i)$$

with the probability $P(k)$

For the variance of a finite sample → Sample mean and variance, page 57. A matrix called covariance matrix is defined for a multivariate random variable (→ Covariance matrix, page 51).

(b) Theorems

Binomial theorem. For p with $0 < p < 1$ and $q = 1 - p$ the following expression is called the Binomial theorem:

$$(p+q)^n = 1 = \sum_{k=0}^{n} \binom{k}{n} p^k \cdot (1-p)^{n-k}$$

Central limit theorem. *The density function of the sum $w = \sum_{i=1}^{n} x_i$ of a sample of n independent random values x_i of any distribution with mean $\bar{x}$ and variance σ^2 approaches in the limit $n \to \infty$ a Gaussian density function with mean $n \cdot \bar{x}$ and Variance $n \cdot \sigma$.*

Cramer-Rao inequality. This inequality gives a lower limit for the variance of any estimator $\widehat{a}$, which has a bias $B(a)$:

$$V(\widehat{a}) \geq \frac{(1 + (\mathrm{d}B/\mathrm{d}a))^2}{I(a)}$$

The variance of an unbiased estimate $\widehat{a}$ is bounded below by the inverse of the information $I(a)$ contained in the data sample (→ Information, page 55).

Characteristic function. For a real random variable x with density function $p(x)$ the *characteristic function* $\phi(t)$ is defined as expectation value of e^{itx}:

$$\phi(t) = E\left[e^{itx}\right] = \int_{-\infty}^{+\infty} e^{itx} \cdot p(x)\,\mathrm{d}x$$

The inverse operation is:

$$p(x) = \int_{-\infty}^{+\infty} e^{itx} \cdot \phi(t)\,\mathrm{d}t\ .$$

The values of the characteristic function and its first and second derivative at the special value $t = 0$ are easily calculated:

$$\phi(0) = 1 \tag{113}$$

$$\frac{d\phi(0)}{dt} = \int_{-\infty}^{+\infty} ixe^{itx} P(x)\mathrm{d}x \bigg|_{t=0} = i\,\langle x \rangle \tag{114}$$

$$\frac{d^2\phi(0)}{dt^2} = \int_{-\infty}^{+\infty} -x^2 e^{itx} P(x)\mathrm{d}x \bigg|_{t=0} = -E\left\{x^2\right\} = -\left(\sigma^2 + \langle x \rangle^2\right) \tag{115}$$

For the characteristic function with respect to the mean value $\langle x \rangle = E[x]$, i.e. for the expectation value $E\left[\exp(it(x - \langle x \rangle))\right]$ the terms containing $\langle x \rangle$ are removed from the last two expressions.

The characteristic function for the Gaussian density. The characteristic function of the standard Gaussian density is defined by

$$\phi_{\mathrm{G}}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2\sigma^2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-((x-it\sigma^2)^2-(it\sigma^2)^2)/2\sigma^2}\,\mathrm{d}x$$

$$\frac{1}{\sqrt{2\pi}} e^{-t^2\sigma^2/2} \int_{-\infty}^{+\infty} e^{-(x-it\sigma^2)^2/2\sigma^2}\,\mathrm{d}x$$

The last integral turns out to be a constant:

$$\int_{-\infty}^{+\infty} e^{-(x-it\sigma^2)^2/2\sigma^2}\,\mathrm{d}x = \sigma^2 \int_{r=-\infty-it}^{r=+\infty-it} e^{-r^2/2}\,\mathrm{d}r$$

by the transformation $r = (x - it\sigma^2)/\sigma^2$ with $\mathrm{d}x = \sigma^2\,\mathrm{d}r$; the integral vanishes on a closed path, because no singularities are included (residue theorem), or symbolically

$$\int_{-L}^{+L} e^{-r^2/2}\,\mathrm{d}r + \int_{-L}^{+L-it} e^{-r^2/2}\,\mathrm{d}r + \int_{L-it}^{-L-it} e^{-r^2/2}\,\mathrm{d}r + \int_{-L-it}^{-L} e^{-r^2/2}\,\mathrm{d}r = 0$$

The 2. and 4. integral vanish for $L \to \infty$, because the integrand vanishes. It follows:

$$\int_{-\infty-it}^{+\infty-it} e^{-r^2/2}\,\mathrm{d}r = \int_{-\infty}^{+\infty} e^{-r^2/2}\,\mathrm{d}r = \text{constant}$$

The characteristic function

$$\phi_{\mathrm{G}}(t) = N_0 \cdot e^{-t^2\sigma^2/2}$$

follows for the Gaussian density.

Gauß-Markov theorem. The Gauß-Markov theorem makes a statement about the variance of $\to$ estimators, which are *unbiased* and are *linear functions* of the data. The linear least squares estimator has the smallest possible variance among all estimators with these properties, if the data are unbiased, regardless of the distribution of the data; i.e. the distribution does not need to be Gaussian.

Law of large numbers. In N statistical experiments the number n_j of an event j is counted. The ratio $h_j = n_j/N$ is a random variable, which follows the Binomial distribution. The expectation value of the random variable h_j, $E(h_j)$, is the probability p_j for the event j. In real life the probabilities p_j of the different events are unknown; they have to be determined (estmated) from experiments. The question ist: How accurate is the estimate h_j for p_j? Because the random variable $h(j)$ follows a Binomial distribution, the variance of h_j can be calculated:

$$\sigma^2(h_j) = \sigma^2(n_j/N) = \frac{1}{N^2} \cdot \sigma^2(n_j) = \frac{1}{N^2} \cdot Np_j(1-p_j)$$

Since the product $p_j(1-p_j)$ is limited (maximum value is 1/4), the inequality

$$\sigma^2(h_j) \le 1/N$$

follows, known as the **Law of large numbers**.

Tschebyschew inequality. *The probability for $|x-\bar{x}| \ge k\sigma$ is less or equal to $1/k^2$ for a random variable x with any density function $p(x)$.*

The inequality gives an upper limit on the probability of exceeding a given number of standard deviations, which does not depend on the shape of the density, provided the variance is known.

Proof: The proof starts from the general definition of the variance σ^2.

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\bar{x})^2 \cdot p(x)\,\mathrm{d}x =$$

$$\int_{-\infty}^{\bar{x}-k\sigma} (x-\bar{x})^2 \cdot p(x)\,\mathrm{d}x + \int_{\bar{x}-k\sigma}^{\bar{x}+k\sigma} (x-\bar{x})^2 \cdot p(x)\,\mathrm{d}x + \int_{\bar{x}+k\sigma}^{+\infty} (x-\bar{x})^2 \cdot p(x)\,\mathrm{d}x$$

The integrands in all three integrals are positive. By removing the middle integral an inequality is defined:

$$\sigma^2 \ge \int_{-\infty}^{\bar{x}-k\sigma} (x-\bar{x})^2 \cdot p(x)\,\mathrm{d}x + \int_{\bar{x}+k\sigma}^{+\infty} (x-\bar{x})^2 \cdot p(x)\,\mathrm{d}x$$

In the first and last integral the inequalities $x > \bar{x}+k\sigma$ and $x < \bar{x}-k\sigma$, respectively, are valid, thus

$$\sigma^2 \ge k^2\sigma^2 \left[\int_{-\infty}^{\bar{x}-k\sigma} p(x)\,\mathrm{d}x + \int_{\bar{x}+k\sigma}^{+\infty} p(x)\,\mathrm{d}x \right]$$

The expression in parentheses is the probability for $|x-\bar{x}| \ge k\sigma$, which proves the Tschebyscheff inequality.

The inequality is valid under rather general conditions and is rather weak. It is useful for theoretical considerations and for those practical cases where the knowledge about the distribution is very limited.

(c) Miscellaneous

Binomial coefficient. The expression

$$\binom{r}{n} = \frac{n!}{k!(n-k)!} = \binom{k}{n-k} \qquad 0 \le k \le n$$

is called Binomial coefficient. This number is identical to the number of combinations C_k^n without repetition.

Gamma function. Euler's Gamma function $\Gamma(x)$ is related to the factorial by $\Gamma(x) = (x-1)!$ for integer, positive x. The function is defined by

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} \, \mathrm{d}t \qquad x > 0$$

Properties are: $\Gamma(x) = 1/x \cdot \Gamma(x+1) = (x-1)\Gamma(x-1)$, and $\Gamma(x)\Gamma(1-x) = \pi/\sin(\pi x)$. Some values are:

x	=	$-1/2$	$1/2$	1	$3/2$	2
$\Gamma(x)$	=	$-2\sqrt{\pi}$	$\sqrt{\pi}$	1	$\sqrt{\pi}/2$	1

An asymptotic formula for large arguments x is Stirling's formula

$$\Gamma(x) \approx e^{-x} x^x \sqrt{2\pi/x} \left(1 + \frac{1}{12x} + \frac{1}{288x^2} + \dots \right)$$

which can also be used to approximate the factorial $x!$ or the logarithm of the factorial (for integer x):

$$x! \approx \left(\frac{x}{e}\right)^x \sqrt{2\pi/x} \left(1 + \frac{1}{12x} + \frac{1}{288x^2} + \dots \right) \qquad \ln x! \approx \left(x + \frac{1}{2}\right) \ln n - n + \ln \sqrt{2\pi}$$

Hessian. The symmetric $n \times n$ of second partial derivatives of a scalar function $F(\boldsymbol{x})$ of an n-vector $\boldsymbol{x}$ is called Hessian matrix $\boldsymbol{H}$:

$$\boldsymbol{H} = \begin{pmatrix} \partial^2 F/\partial x_1^2 & \partial^2 F/\partial x_1 \partial x_2 & \dots & \partial^2 F/\partial x_1 \partial x_n \\ \partial^2 F/\partial x_2 \partial x_1 & \partial^2 F/\partial x_2^2 & \dots & \partial^2 F/\partial x_2 \partial x_n \\ \dots & & & \dots \\ \partial^2 F/\partial x_n \partial x_1 & \partial^2 F/\partial x_n \partial x_2 & \dots & \partial^2 F/\partial x_n^2 \end{pmatrix}$$

Integrals. Some useful integrals occuring frequently in statistical calculations are given below.

$$\int_{-\infty}^{+\infty} e^{-ax^2} \, \mathrm{d}x = \sqrt{\frac{\pi}{a}} \qquad \int_0^\infty x\, e^{-ax^2} \, \mathrm{d}x = \frac{1}{2a}$$

$$\int_{-\infty}^{+\infty} x^2\, e^{-ax^2} \, \mathrm{d}x = \frac{1}{2a}\sqrt{\frac{\pi}{a}} \qquad \int_0^\infty x^3\, e^{-ax^2} \, \mathrm{d}x = \frac{1}{2a^2}$$

$$\int_{-\infty}^{+\infty} x^{2n}\, e^{-ax^2} \, \mathrm{d}x = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{(2a)^n}\sqrt{\frac{\pi}{a}} \qquad \int_0^\infty x^{2n+1}\, e^{-ax^2} \, \mathrm{d}x = \frac{n!}{2a^{n+1}}$$

Jacobi matrix. Given m functions $f_1(x_1, \dots x_n), \dots f_m(x_1, \dots x_n)$ of n variables $x_1, \dots x_n$, the Jacobi matrix is the matrix of first derivatives of the functions f_i woth respect to the variables x_j:

$$\boldsymbol{J} = \frac{\partial(f_1, \dots f_m)}{\partial(x_1, \dots x_n)} = \begin{pmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \dots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \dots & \partial f_2/\partial x_n \\ \dots & & & \dots \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \dots & \partial f_m/x_n \end{pmatrix}$$

or short $\partial \boldsymbol{f}/\partial \boldsymbol{x}$.

The chain rule is valid: if $\boldsymbol{x} = \boldsymbol{x}(u_1, \dots u_p)$, then the derivative $\partial \boldsymbol{f}/\partial \boldsymbol{u}$ is given by the matrix product $(\partial \boldsymbol{f}/\partial \boldsymbol{x})\,(\partial \boldsymbol{x}/\partial \boldsymbol{u})$, i.e.

$$\frac{\partial f_i}{\partial u_k} = \sum_{j=1}^{n} \frac{\partial u_i}{\partial x_j} \frac{\partial x_j}{\partial u_k} \,.$$

For a coordinate transformation $\boldsymbol{u} \to \boldsymbol{u} = \boldsymbol{u}(x_1, \dots x_n)$ with $p = n$ the product of $\mathrm{d}\boldsymbol{u}/\mathrm{d}\boldsymbol{x}$ and $\mathrm{d}\boldsymbol{u}/\mathrm{d}\boldsymbol{x}$ is the unit matrix, i.e. $\mathrm{d}\boldsymbol{u}/\mathrm{d}\boldsymbol{x} = (\mathrm{d}\boldsymbol{u}/\mathrm{d}\boldsymbol{x})^{-1}$.

Jacobian. The Jacobian or Jacobi determinant is the factor in the transformation of a function of n variables.

$$p\,(\boldsymbol{x})\;\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_n = p\,(\boldsymbol{x}\,(\boldsymbol{u}))\,|\det(\mathrm{d}\boldsymbol{x}/\mathrm{d}\boldsymbol{u})\,|\,\mathrm{d}u_1\mathrm{d}u_2\dots\mathrm{d}u_p$$

where $(\mathrm{d}\boldsymbol{x}/\mathrm{d}\boldsymbol{u}) = (\mathrm{d}\boldsymbol{u}/\mathrm{d}\boldsymbol{x})^{-1}$ is the → Jacobi matrix. The factor

$$|\det(\mathrm{d}\boldsymbol{x}/\mathrm{d}\boldsymbol{u})\,| = |\det(\mathrm{d}\boldsymbol{u}/\mathrm{d}\boldsymbol{x})\,|^{-1}$$

is the absolut value of the Jacobi determinant.

(→ Transformation of variables).

Minimization. The determination of the minimum of a function $F(\boldsymbol{a}$ is called minimization. Minimization or maximization problems arise in optimization problems (e.g. maximum likelihood method), where an objective function $F(\boldsymbol{a})$ depending on a parameter vector $\boldsymbol{a}$ is defined; the minimum or the maximum corresponds to the optimal solution of the optimization problem.

A local minimum of a function $F(\boldsymbol{a})$, depending on n variables $a_1, \dots, a_i, \dots, a_n$, can be determined by solving simultaneously the n equations

$$\frac{\partial F}{\partial a_1} = 0 \quad \dots \quad \frac{\partial F}{\partial a_i} = 0 \quad \dots \quad \frac{\partial F}{\partial a_n} = 0$$

which correspond to a vanishing gradient $\boldsymbol{\nabla} F$ and in general, define a minimum, a maximum or a saddle point. Methods for the determination are search methods or iterative methods; none of the many methods is universically applicable. There are simple methods like the simplex method of Nelder and Mead which do not use derivatives and which are useful especially in regions far from the minimum. The iterative Newton-Raphson method requires derivatives and is based on the Taylor expansion of the obejctive function at some value $\boldsymbol{a}_t$:

$$F(\boldsymbol{a}) = F(\boldsymbol{a}_t) + (\boldsymbol{a} - \boldsymbol{a}_t)^T\,\boldsymbol{g} + \frac{1}{2}\,(\boldsymbol{a} - \boldsymbol{a}_t)^T\,\boldsymbol{H}\,(\boldsymbol{a} - \boldsymbol{a}_t) + \dots$$

where $\boldsymbol{g}$ is the gradient with elements $g_j = \partial F/\partial a_j$ and $\boldsymbol{H}$ is the Hessian with elements $H_{jk} = \partial^2 F/\partial a_j \partial a_k$ at the value $\boldsymbol{x} = \boldsymbol{x}_t$. Ignoring the higher-order terms a step $\boldsymbol{\Delta a}$ can be calculated from the condition $\boldsymbol{\nabla} F = 0$;

$$\boldsymbol{\Delta a} = \boldsymbol{a} - \boldsymbol{a}_t = -\boldsymbol{H}^{-1}\boldsymbol{g} \,.$$

which in general gives an improvement towards the solution. In a robust method a so-called line search is made for the minimum of the function $f(\alpha) = F(\boldsymbol{a}_t + \alpha \cdot \boldsymbol{\Delta a})$. The convergence towards the solution is usually very fast near the minimum; monitoring of the convergence can be done with the quantity

$$d = -\frac{1}{2}\boldsymbol{g}^T\boldsymbol{\Delta a} = \frac{1}{2}\boldsymbol{g}^T\boldsymbol{H}\boldsymbol{g}$$

which corresponds to the expected function decrease. The method requires the computation of the derivatives, which can be done analytically or by numerical methods.

A special case of minimization is the maximum likelihood method, where the function $F(\boldsymbol{a}) = -\ln \mathcal{L}(\boldsymbol{a})$ is approximately quadratic and the inverse Hessian $\boldsymbol{H}$ can be considered as a good approximation of the covariance matrix of parameters.

Standard minimization problems are impractical in case of very large number of parameters, or if the number of possible discrete solutions is given by permutations. Different methods are available in those cases like *simulated annealing* and *genetic algorithms*.

Monte Carlo method. A statistical process can be simulated by mathematical methods, where all random variables of the statistical process are artificially generated. Standard methods are used to generated random variables u from a uniform distribution $U(0,1)$ ($\rightarrow$ random number generator), and these random variables are transformed to the random variables x of the process. This method is called Monte Carlo method.

The Monte Carlo method allows to simulate experiments and to study the statistical properties of the relevant variables, e.g. the distribution of an $\rightarrow$ estimator.

Optimization. Optimization according to a certain criterium is usually expressed in terms of an objective function, which depends on parameters. The optimality condition corresponds to the minimum or maximum of the objective function.

Permutation. The number of permutations on n distinguishable objects is equal to $P_n = n!$ (with $0! = 1! = 1$). For example for the three objects a, b and c there are $3! = 6$ permutations abc, bca, cab, cba, bac and acb.

The number of combinations of n distinguishable objects to r objects is equal to the $\rightarrow$ Binomial coefficient $C_n^r = \binom{r}{n}$.

Schwarz inequality.

$$\int f^2 \, \mathrm{d}x \int g^2 \, \mathrm{d}x \geq \left(\int f g \, \mathrm{d}x \right)^2$$

and for vectors

$$|\boldsymbol{x}| \; |\boldsymbol{y}| \geq |\boldsymbol{x} \cdot \boldsymbol{y}|$$