C. Statistics

V. Blobel
Inst. für Experimentalphysik, Hamburg

Sommersemester 2002

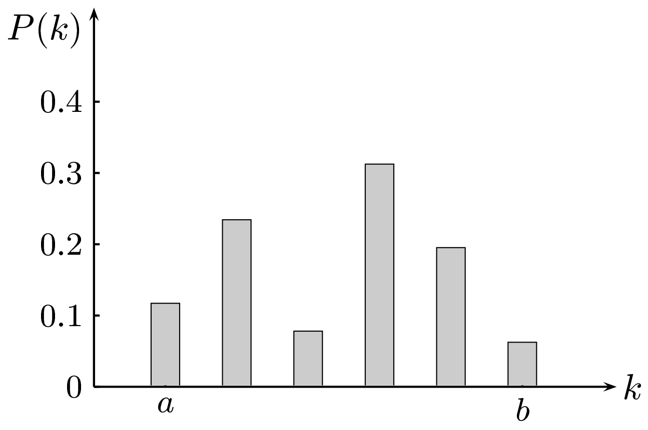
1. Random variables and distributions
2. Standard distributions
3. Central limit theorem
4. Transformations (Error propagation)
5. Multidimensional distributions

C.1 Random variables and distributions

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

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



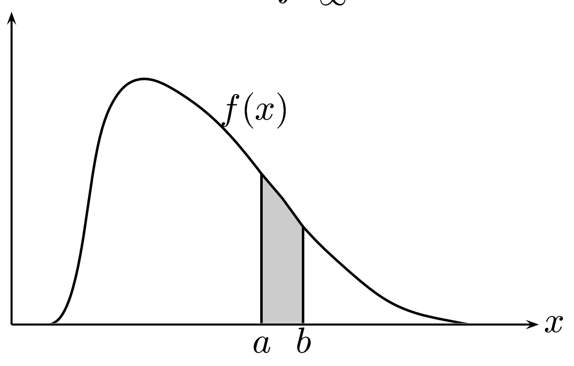
Continuous random variables

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

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

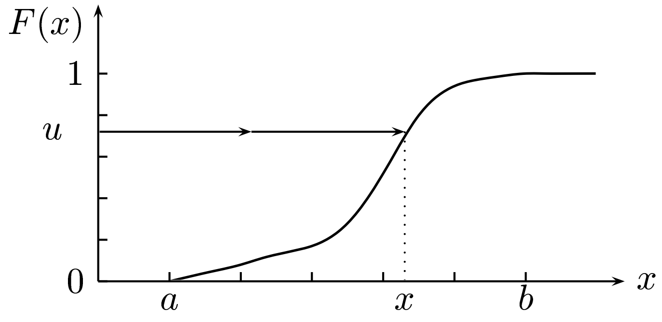


Continuous random variables contnd.

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

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

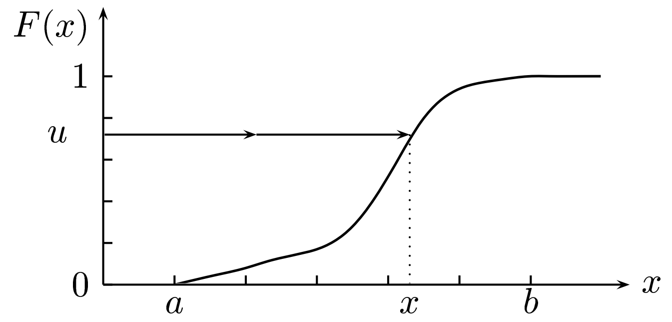
The distribution function and the probability density function are related also by $f(x) = \mathrm{d}F(x)/\,\mathrm{d}x$.



Two applications for relation $x \leftrightarrow u = F(x)$

(1) Assuming random variable $x \in f(x)$ (hypothesis), a value of the random variable x is transformed to u. If the hypothesis is correct, then u follows a uniform distribution (make histogram of u for a test).

(2) Generation of a random variable $x \in f(x)$. Get random value u from the uniform distribution (random number generator), and transform to x by $x = F^{(-1)}(u)$.



Expectation values

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

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

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

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

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

standard deviation $\qquad \sigma = \sqrt{V}$

Expectation and variance

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:

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

For the variance operator

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

is valid.

Algebraic and central moments

Moments:

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

Central moments are defined with respect to the mean value.

Some moments and functions of moments have a special name:

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

Tschebyscheff inequality

The probability for a deviation of a random variable from the mean value μ by k times the standard deviation σ depends on the specific distribution.

An upper limit is given, *independently* of the distribution, by the Tschebyscheff inequality:

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

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

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

Tschebyscheff inequality contnd.

All three integrals are positive; 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

$$\begin{aligned}\sigma^2 &\geq k^2\sigma^2 \left[\int_{-\infty}^{\langle x\rangle-k\sigma} f(x)\,\mathrm{d}x + \int_{\langle x\rangle+k\sigma}^{+\infty} f(x)\,\mathrm{d}x\right] \\ &= P\left(|x-\mu| \geq k\sigma\right)\end{aligned}$$

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

C.2 Standard distributions

Relation between the standard distributions

$$
\begin{tikzcd}
\boxed{\text{Binomial } n, p} \arrow[rr, "p \to 0"', "np = \mu"] \arrow[dr, "n \to \infty"'] & & \boxed{\text{Poisson } \mu} \arrow[dl, "\mu \to \infty"] \\
& \boxed{\text{Normal}} &
\end{tikzcd}
$$

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:

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

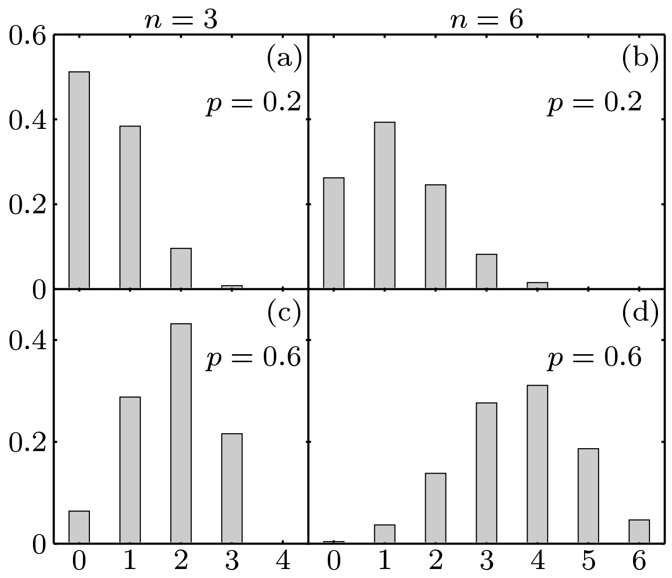
The factor $\binom{n}{k}$ ist the number of possibilities to select k elements out of total n different elements. In a single event there are two possible outcomes, one (often called "success") with probability p and another (often called "failure") with probability $q = 1 - p$.

Expectation value $= E[k] = n\,p$ Variance $= V[k] = n\,p(1-p)$

Binomial theorem $\quad (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = 1$

Binomial distribution contnd.

Expectation value $= E[k] = n\,p$ Variance $= V[k] = n\,p(1-p)$



Law of large numbers

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.

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

Because of $p\,(1-p) \leq 1/4$ for any value of p, one can estimate

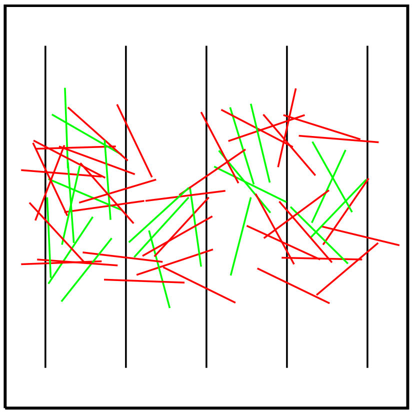
$$V[h] \leq \frac{1}{4}\frac{1}{n} \qquad \sigma(h) \propto \frac{1}{\sqrt{n}}$$

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

Buffons needle – Monte Carlo estimate of π

George L.L. Buffon (1707 – 1788) * Montbard

A needle of length D is falling randomly on a plane with parallel lines of distance D. The probability for crossing a line (red needle) is $(D\cos\varphi)/D = \cos\varphi$ (φ = angle to horizontal direction).



$$P = \langle\cos\varphi\rangle = \frac{\int_0^{\pi/2}\cos\varphi\,\mathrm{d}\varphi}{\int_0^{\pi/2}\mathrm{d}\varphi} = E\left[\frac{k}{n}\right] = \frac{2}{\pi} \approx 0.6366 \quad\rightarrow\quad \pi_{\text{est}} = 2\frac{n}{k}$$

Poisson distribution

Consider Binomial distribution with $n \to \infty$, $p \to 0$, but $n \cdot p \to \mu$ (fixed mean value):

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

Expectation value $E[k] = \mu$ Variance $V[k] = \mu$

For rate of events, which is constant in time t: number of events k follows a Poisson distribution, and the time difference t follows exponential distribution:

$$f(t;\tau)\,\mathrm{d}t = \frac{1}{\tau}e^{-t/\tau}\,\mathrm{d}t$$

Normal distribution

The normal distribution ist the most important continous distribution; almost anything which is a sum of many random contributions, follows (approximately) the normal distribution ($\rightarrow$ central limit theorem).

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

Expectation value $E[x] = \mu$ Variance $V[x] = \sigma^2$

$$\begin{aligned} P(|x-\mu| < 1\cdot\sigma) &= 68.26\% \\ P(|x-\mu| < 2\cdot\sigma) &= 95.45\% \\ P(|x-\mu| < 3\cdot\sigma) &= 99.73\% \end{aligned}$$

The Poisson distribution becomes a normal distribution for large μ.

C.3 Central limit theorem

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 random variables x_i* (σ_i must exist and should be finite).

- Characteristic functions
- Sum of independent random variables
- Normal distribution
- Proof
- Numerical examples

For finite n: theorem is valid, if sum is not dominated by one (few) terms

- velocity component v_x of gas molecules (good example)
- energy loss of charged particles traversing thin gas layer (bad example, Landau distribution)

Characteristic functions

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

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

with the inverse operation

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

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

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

Special case: $t = 0$

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

or more general for the n-th derivative:

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

Thus it may be simple to calculate all the moments of a distribution $f(x)$ defined by the characteristic function $\phi(t)$.
In the special case of a mean value of 0 the first few terms are

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

Sum of independent random variables

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:

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

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

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

Normal distribution

Density $f(x)$ and the characteristic function $\phi(t)$ are:

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

The addition theorem 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:

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

Central limit theorem

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 random variables x_i.*

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

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

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 be expressed in terms of a Taylor expansion:

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

Central limit theorem contnd.

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

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

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

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

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

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

C.4 Transformation (error propagation)

A function of a random variable is itself a random variable.

Suppose x follows a pdf $f_x(x)$ and the transformation

$$y = y(x)$$

of x into a variable y is given.

What is the probability density function $f_y(y)$?

Transformation of a single variable

Consider intervals $(x, x + \mathrm{d}x)$ and $(y, y + \mathrm{d}y)$. For intervals

$$f_x(x)\,|\,\mathrm{d}x| = f_y(y)\,|\,\mathrm{d}y|$$

$$f_y(y) = f_x(x(y)) \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = f_x(x(y)) \left|\frac{1}{\mathrm{d}y/\,\mathrm{d}x}\right|$$

If the inverse of $y(x)$ is not unique, include all branches.

Example: $y = x^2$ $\quad x = \pm\sqrt{y}$ $\quad \mathrm{d}x = \pm\frac{\mathrm{d}y}{2\sqrt{y}}$

Transformed distribution:

transf. distribution $\quad f_y(y) = \dfrac{1}{2\sqrt{y}}\left(f_x(+\sqrt{y}) + f_x(-\sqrt{y})\right)$

Example: Linear transformation $y = ax + b$ $\quad x = (y - b)/a$

transf. distr. $f_y(y) = \dfrac{1}{a} f_x\left(\dfrac{y-b}{a}\right)$ $\quad E[y] = a{\cdot}E[x]{+}b$ $\quad V[y] = a^2 V[x]$

Error propagation

Nonlinear transformation $y = y(x)$ may be complicated.

Simpler: transform mean and variance based on expansion

$$y(x) = y(\langle x \rangle) + (x - \langle x \rangle) \left. \frac{\mathrm{d}y}{\mathrm{d}x} \right|_{x=\langle x \rangle} + \frac{1}{2}(x - \langle x \rangle)^2 \left. \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} \right|_{x=\langle x \rangle} + \ldots$$

Taking terms up to second order with correction term

$$\text{mean} \quad \langle y \rangle \approx y(\langle x \rangle) \qquad + \frac{1}{2}\sigma_x^2 \left. \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} \right|_{x=\langle x \rangle}$$

$$\text{Variance} \qquad V[y] \approx \left(\frac{\mathrm{d}y}{\mathrm{d}x} \right)^2 \cdot V[x] \qquad \text{error propagation}$$

C.5 Multidimensional distributions

Starting with two random variables x and y …

The two-dimensional probability density function $f(x,y)$ is defined by the probability to find (x,y) in intervals $a \leq x < b$ and $c \leq y < d$:

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

Normalisation:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1\,.$$

Marginal probability density function:

$$f_x(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x \qquad f_y(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y\,.$$

They represent the probability density of the random variables x and y, resp., if the other variable is ignored.

Two-dimensional distributions

For a given $x = x_0$ the *Conditional probability density function* is given by

$$f(y|x_0) = \frac{f(x_0, y)}{\int f(x_0, y)\,\mathrm{d}y} = \frac{f(x_0, y)}{f_y(x_0)} .$$

From the marginal and the conditional p.d.f. the joint p.d.f. is obtained:

$$f(x, y) = f(y|x) \cdot f_y(x)$$

For independent variables the joint p.d.f. factorizes:

$$f(x, y) = f_x(y) \cdot f_y(x)$$

Expectation, variance and covariance

$$\begin{aligned} E\,[x] &= \iint x f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int x f_y(x)\,\mathrm{d}x \\ E\,[y] &= \iint y f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int y f_x(y)\,\mathrm{d}y \\ V\,[x] &= \iint (x - \langle x \rangle)^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \sigma_x^2 \\ V\,[y] &= \iint (y - \langle y \rangle)^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \sigma_y^2 \,. \end{aligned}$$

A new quantity:

covariance $\qquad \sigma_{xy} = \mathrm{cov}(x,y) = \displaystyle\iint (x - \langle x \rangle)(y - \langle y \rangle) f(x,y)\,\mathrm{d}x\,\mathrm{d}y \,.$

correlation coefficient $\rho_{xy} \qquad \sigma_{xy} = \rho_{xy}\,\sigma_x \sigma_y \qquad -1 \leq \rho \leq +1$

x and y independent $\rightarrow$ $\rho = 0$ (but $\not\leftarrow$)

$\rho = 0 \rightarrow$ x and y independent or dependent

$\rho = +1$ or $-1 \rightarrow$ x and y completely correlated

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:

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

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

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

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:

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

The *transposed* of matrix $\boldsymbol{A}$ is $\boldsymbol{A}^T$ and has the elements $(\boldsymbol{A}^T)_{ij} = (\boldsymbol{A})_{ji}$. The matrix is called *symmetric* if $\boldsymbol{A} = \boldsymbol{A}^T$.

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* matrix $\boldsymbol{1}$ 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.

Proof of inequality for correlation coefficient

Correlation coefficient ρ defined by $\sigma_{xy} = \rho\sigma_x\sigma_y$:

$$|\rho_{xy}| = \frac{|\sigma_{xy}|}{\sqrt{\sigma_x^2\sigma_y^2}} \leq 1$$

Consider expectation value

$$E\left[(a(x - \langle x \rangle) + b(y - \langle y \rangle))^2\right] = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_{xy} \geq 0$$

...for any values of a and b.

Written with matrices: quadratic form

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_{xy} \geq 0$$

$$\begin{vmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{vmatrix} = \sigma_x^2\sigma_y^2 - \sigma_{xy}^2 \geq 0$$

which proves the inequality.

n-dimensional random variable

Consider random vector $\boldsymbol{x}$ with components x_i, $(i = 1, 2, \ldots n)$.

Probability density function:

$$f(x_1, x_2, \ldots x_n) = f(\boldsymbol{x})$$

Expectation values with $\langle \boldsymbol{x} \rangle = E[\boldsymbol{x}]$:

$$E[x_i] = \langle x_i \rangle = \int x_i f(x_1, x_2, \ldots, x_n) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n$$
$$= \int x_i \cdot f(\boldsymbol{x}) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \; .$$

Variance

Generalization of the variance:

$$\text{covariance matrix} \quad \boldsymbol{V} = \boldsymbol{V}[\boldsymbol{x}] = E\left[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])^T\right]$$

with diagonal and not-diagonal elements:

$$V_{ii} = \text{var}(x_i) = \int (x_i - \langle x_i \rangle)^2 \cdot f(\boldsymbol{x}) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n$$

$$V_{ij} = \text{cov}(x_i, x_j) = \int (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \cdot f(\boldsymbol{x}) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \, .$$

$$\begin{pmatrix} V_{11} & V_{12} & \cdots & V_{1n} \\ V_{21} & V_{22} & \cdots & V_{2n} \\ \cdots & \cdots & & \\ V_{n1} & V_{n2} & \cdots & V_{nn} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \cdots & \cdots & & \\ \rho_{1n}\sigma_n\sigma_1 & \rho_{2n}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{pmatrix}$$

Multivariate normal distribution

Generalization of the one-dimensional normal distribution to the case of n dimensions.
The random variable is a n-vector $\boldsymbol{x}$, the density depends on the n-vector μ_x of mean values and on the n-by-n covariance matrix V_x.

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

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 distributions; the parameters of the n-dimensional normal distribution are just

- the first moment (mean $\boldsymbol{\mu}$), and
- the second central moment (covariance matrix $\boldsymbol{V}$).

Two-dimensional normal distribution

Covariance matrix $\boldsymbol{V}$ in terms of x and y:

$$\boldsymbol{V} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_y\sigma_x & \sigma_y^2 \end{pmatrix}$$

Inverse covariance matrix = weight matrix:

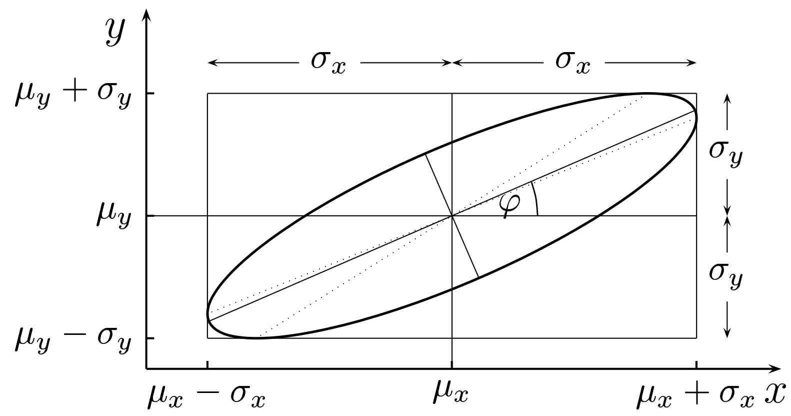
$$\boldsymbol{W} = \boldsymbol{V}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1/\sigma_x^2 & -\rho/(\sigma_x\sigma_y) \\ -\rho/(\sigma_y\sigma_2) & 1/\sigma_y^2 \end{pmatrix}$$

Probability density function:

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

Two-dimensional normal distribution - contour

Elliptical 1σ contour (38 % probability content) with mean values μ_x, μ_y, standard deviations σ_x, σ_y and $\rho = +0.8$.



Main axes of the ellipse with the angles φ und $\varphi + \pi/2$ given by $\tan 2\varphi = \pm(2\rho\sigma_x\sigma_y)/(\sigma_x^2 - \sigma_y^2)$.
Dotted lines = conditional means $x - \mu_x = (y - \mu_y)\rho\sigma_x/\sigma_y$ and $y - \mu_y = (x - \mu_x)\rho\sigma_y/\sigma_x$.

Linear transformation $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$

n-by-m matrix $\boldsymbol{B}$ with elements $B_{ik} = \partial y_i / \partial x_k$ transforms n-vector $\boldsymbol{x}$ with mean $\boldsymbol{\mu_x}$ and covariance matrix $\boldsymbol{V}(\boldsymbol{x})$ to m-vector $\boldsymbol{y}$

$$\langle \boldsymbol{y} \rangle = \boldsymbol{\mu_y} \quad = \quad E\left[\boldsymbol{y}\right] = E\left[\boldsymbol{B}\boldsymbol{x}\right] = \boldsymbol{B}E\left[\boldsymbol{x}\right] = \boldsymbol{B}\boldsymbol{\mu_x}$$

(expectation is a linear operator)

Covariance matrix: ”error” propagation

$$\begin{aligned}\boldsymbol{V}\left[\boldsymbol{y}\right] = E\left[(\boldsymbol{y}-\boldsymbol{\mu_y})(\boldsymbol{y}-\boldsymbol{\mu_y})^T)\right] &= E\left[(\boldsymbol{B}\boldsymbol{x}-\boldsymbol{B}\boldsymbol{\mu_x})(\boldsymbol{B}\boldsymbol{x}-\boldsymbol{B}\boldsymbol{\mu_x})^T\right] \\ &= E\left[\boldsymbol{B}(\boldsymbol{x}-\boldsymbol{\mu}_x)(\boldsymbol{x}-\boldsymbol{\mu}_x)^T\boldsymbol{B}^T\right] \\ &= \boldsymbol{B}\cdot E\left[(\boldsymbol{x}-\boldsymbol{\mu}_x)(\boldsymbol{x}-\boldsymbol{\mu}_x)^T\right]\boldsymbol{B}^T = \boldsymbol{B}\boldsymbol{V}\left[\boldsymbol{x}\right]\boldsymbol{B}^T\end{aligned}$$

General transformation $\boldsymbol{x} \to \boldsymbol{y}$

nonlinear transformation $\quad y_i = y_i(x_1, x_2, \dots x_n) = y_i(\boldsymbol{x})$

In principle p.d.f. $g(\boldsymbol{y})$ can be determined from $f(\boldsymbol{x})$ (assuming same dimensions):

$$f(\boldsymbol{x})\,\mathrm{d}x_1\,\mathrm{d}x_2\dots\mathrm{d}x_n = g(\boldsymbol{y})\,\mathrm{d}y_1\,\mathrm{d}y_2\dots\mathrm{d}y_n\ ,$$

where $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are multi-dimensional p.d.f's.

$$\text{p.d.f. transformation:} \qquad g(\boldsymbol{y}) = f(\boldsymbol{x}) \cdot J\left(\frac{x}{y}\right)\ ,$$

where the elements x_k have to be expressed by the elements y_i.
$J\left(x/y\right)$ is Jacobi determinant:

$$J\left(\frac{x}{y}\right) = \left|\frac{\partial x_1 \partial x_2 \cdots \partial x_n}{\partial y_1 \partial y_2 \cdots \partial y_n}\right| \qquad \text{or} \qquad J\left(\frac{y}{x}\right) = \left|\frac{\partial y_1 \partial y_2 \cdots \partial y_n}{\partial x_1 \partial x_2 \cdots \partial x_n}\right|$$

…this can be very complicated or impossible.

Nonlinear transformation of covariance matrix

The approximate calculation of the covariance matrix $\boldsymbol{V}[\boldsymbol{y}]$ from the covariance matrix $\boldsymbol{V}[\boldsymbol{x}]$ uses the constant matrix $\boldsymbol{B}$ of derivatives:

$$\boldsymbol{B} = \begin{pmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 & \dots & \partial y_1/\partial x_n \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 & \dots & \partial y_2/\partial x_n \\ \dots & & & \\ \dots & & & \\ \partial y_m/\partial x_1 & \partial y_m/\partial x_2 & \dots & \partial y_m/\partial x_n \end{pmatrix},$$

(each element calculated at the mean value $\boldsymbol{\mu}_y$).

General law of approximate "error" propagation:

$$\boxed{\boldsymbol{V}[\boldsymbol{y}] = \boldsymbol{B}\boldsymbol{V}[\boldsymbol{x}]\boldsymbol{B}^T}$$

Special cases: independent random variables

If y has the form

$$y = x_1 + x_2 - x_3 \cdots \qquad \text{(sums, differences)},$$

then the result of error propagation is (exactly)

$$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \sigma_{x_3}^2 + \cdots$$

if $\sigma_{x_1}, \sigma_{x_2}, \sigma_{x_3} \dots$ are the standard deviations of the independent random variables x_1, x_1, $x_3 \dots$; *the squares of the individual standard deviations are added*.

If y has the form

$$y = \frac{x_1 \cdot x_2 \cdots}{x_3 \cdots} \qquad \text{(product, quotient)},$$

then the result of error propagation is (approximately)

$$\left(\frac{\sigma_y}{\mu_y}\right)^2 = \left(\frac{\sigma_{x_1}}{\mu_{x_1}}\right)^2 + \left(\frac{\sigma_{x_2}}{\mu_{x_2}}\right)^2 + \left(\frac{\sigma_{x_3}}{\mu_{x_3}}\right)^2 + \cdots$$

The squares of the individual relative standard deviations are added.

χ^2 *distribution*

For independent normal random variables $x_i \in N(0,1)$ (i.e. $\mu = 0$ and $\sigma = 1$) for $i = 1, 2, \ldots n$ the sum

$$u = \sum_{i=1}^{n} x_i^2 \qquad \text{(mean value} = n\text{)}$$

follows a χ^2 distribution $f_n(u)$ with n degrees of freedom.

$$\boxed{f_n(u) \; = \frac{\frac{1}{2}\left(\frac{u}{2}\right)^{n/2-1} e^{-u/2}}{\Gamma(n/2)}} \qquad \text{density}$$

Important for statistical tests (”goodness of fit”).

Random variables are (linearly) transformed before to random variables with $\mu = 0$ and $\sigma = 1$.

Correlations between data points

Case of two *correlated* data points:

$$\text{Covariance matrix} \quad \boldsymbol{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\begin{aligned} \sigma_{1,2} &= \text{standard deviations of the variables} \\ \rho &= \text{correlation coefficient} \qquad -1 < \rho < +1 \end{aligned}$$

The χ^2 value for a deviation $\boldsymbol{\Delta}$ (= 2-vector of deviations) of a theoretical prediction from the data points is given by:

$$\boxed{\chi^2 = \boldsymbol{\Delta}^T \left(\boldsymbol{V}^{-1}\right) \boldsymbol{\Delta}}$$

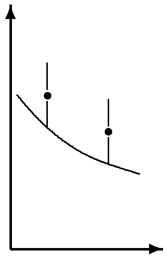
with the inverse of the covariance matrix

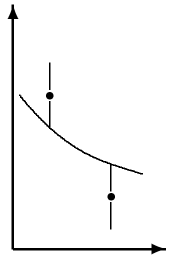
$$\boldsymbol{V}^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1/\sigma_1^2 & -\rho/\sigma_1\sigma_2 \\ -\rho/\sigma_1\sigma_2 & 1/\sigma_2^2 \end{pmatrix}$$

Numerical Example for deviations

$\sigma_1 = \sigma_2 = \sigma$ and two deviations $\Delta = \pm 1\,\sigma$

$$\chi^2 = \begin{cases} 2/(1+\rho) & \text{same sign of } \Delta \\ 2/(1-\rho) & \text{opposite sign of } \Delta \end{cases}$$





Probability from χ^2

	same sign	opposite sign
$\rho = 0$	37 %	37 %
$\rho = +0.9$	59 %	0.0005 %
$\rho = -0.9$	0.0005 %	59 %

Conclusion: Interpretation of *common* deviations of > 1 data point in case of *correlated* ($\rho \neq 0$) data points is difficult.

Numerical Example - averaging of two data points

$\sigma_1 = \sigma_2 = \sigma$ and correlation coefficient $\rho = -0.9,\ 0,\ +0.9$

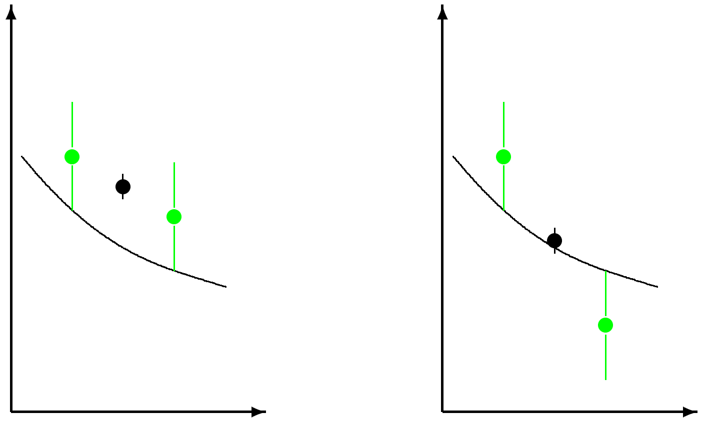
Standard deviation of average is $\sqrt{\dfrac{1+\rho}{2}}\sigma$ (by error propagation)

Numerical values:

	$\sqrt{\frac{1+\rho}{2}}\,\sigma$		comment
$\rho = 0$	0.707	σ	$1/\sqrt{2}$
$\rho = +0.9$	0.975	σ	$\approx$ same precison as single point
$\rho = -0.9$	0.224	σ	much more accurate

Numerical Example - averaging of two data points contnd.

Averaging in case of negatively correlated data points:



Conclusion:

- Averaging of data points with large negative correlations results in very accurate values.
- Averaging of data points with large positive correlation reduces the number of data points without increasing their accuracy.

Multinomial distribution

The multinomial distribution is the generalization of the binomial distribution to the case of m different possible outcomes.

There are m (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.

$$E[k_i] = n\, p_i \qquad V[\boldsymbol{k}]_{ij} = \begin{cases} -np_i p_j & i \neq j \\ np_i\,(1 - p_i) & i = j \end{cases}$$

(Note: correlation is always negative; important for case of small number m.)

Application: 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 approximated ($\rightarrow$ Poisson distribution):

$$V[k_i] = n\,p_i\,(1 - p_i) \approx n\,p_i \approx k_i \qquad \sigma_i \approx \sqrt{k_i}$$