# Statistical analysis of experimental data
# **Least-squares method**

Aleksander Filip Żarnecki

FACULTY OF PHYSICS
UNIVERSITY OF WARSAW

**Lecture 08**

November 21, 2024

## Least-squares method

1. $\chi^2$ distribution
2. Hypothesis Testing
3. Linear Regression
4. Homework

## Maximum Likelihood Method

The product:

$$L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

is called a likelihood function.

The most commonly used approach to parameter estimation is the maximum likelihood approach: as the best estimate of the parameter set $\boldsymbol{\lambda}$ we choose the parameter values for which the likelihood function has a (global) maximum.

Frequently used is also log-likelihood function

$$\ell = \ln L = \sum_{j=1}^{N} \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

we can look for maximum value of $\ell$ or minimum of $-2\,\ell = -2\ln L$

## Multiple parameter estimate

Likelihood function (and log-likelihood) can depend on multiple parameters:

$$\boldsymbol{\lambda} = (\lambda_1 \ldots \lambda_p) \qquad L = \prod_{j=1}^{N} f(\mathbf{x}^{(j)}; \boldsymbol{\lambda}) \qquad \ell = \sum_{j=1}^{N} \ln f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$$

Best estimate of $\boldsymbol{\lambda}$, for given set of experimental results $\mathbf{x}^{(j)}$, corresponds to maximum of the likelihood function, which can be found by solving a system of equations:

$$\left. \frac{\partial \ell}{\partial \lambda_i} \right|_{i=1\ldots p} = 0$$

## The Likelihood Principle

G. Bohm and G. Zech

Given a p.d.f. $f(\mathbf{x}; \boldsymbol{\lambda})$ containing an unknown parameters of interest $\boldsymbol{\lambda}$ and observations $\mathbf{x}^{(j)}$, all information relevant for the estimation of the parameters $\boldsymbol{\lambda}$ is contained in the likelihood function $L(\boldsymbol{\lambda}; \mathbf{x}) = \prod f(\mathbf{x}^{(j)}; \boldsymbol{\lambda})$.

## Parameter covariance matrix

For the considered case of multivariate normal distribution, best parameter estimates $\hat{\boldsymbol{\lambda}}$ are given by the measured variable values $\mathbf{x}$.

Unlike parameters $\boldsymbol{\lambda}$, parameter estimates $\hat{\boldsymbol{\lambda}}$ are random variables (functions of $\mathbf{x}$) and so we can consider covariance matrix for $\hat{\boldsymbol{\lambda}}$:

$$\mathbb{C}_{\mathbf{x}} = \mathbb{C}_{\hat{\boldsymbol{\lambda}}} = \left( - \frac{\partial^2 \ell}{\partial \lambda_i \, \partial \lambda_j} \right)^{-1}$$

Knowing the likelihood function, we can not only estimate parameter values, but also extract uncertainties and correlations of these estimates!

## Recipe for a parameter uncertainty

G. Bohm and G. Zech

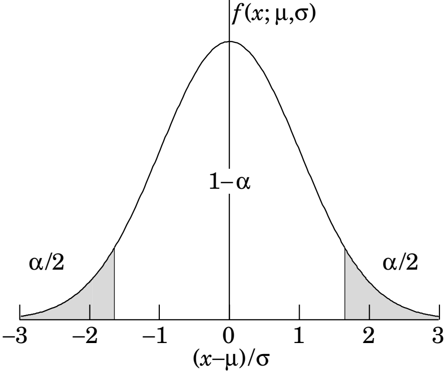
Standard error intervals of the extracted parameter are defined by the decrease of the log-likelihood function by 0.5 for one, by 2 for two and by 4.5 for three standard deviations.

## Normal distribution

Meaning of $\sigma$ is well defined for Gaussian distribution.

Probability for the experimental result to be consistent with the true value within $\pm N\sigma$:

| | | $1-\alpha$ | |
|---|---|---|---|
| $\pm 1\,\sigma$ | $\Rightarrow$ | 68.27 | % |
| $\pm 2\,\sigma$ | $\Rightarrow$ | 95.45 | % |
| $\pm 3\,\sigma$ | $\Rightarrow$ | 99.73 | % |
| $\pm 4\,\sigma$ | $\Rightarrow$ | 99.9937 | % |
| $\pm 5\,\sigma$ | $\Rightarrow$ | 99.999943 | % |




There is a non-zero chance for deviation grater than $5\sigma$, but it is extremely small

## Least-squares method

## Maximum Likelihood Method

see lectures 05 and 06

Let us consider $N$ independent measurements of variable $Y$. Assuming **measurement fluctuations are described by Gaussian pdf**, the likelihood function is:

$$L = \prod_{i=1}^{N} G(y_i;\mu_i,\sigma_i) = \prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma_i^2}\right)$$

Log-likelihood:

assuming $\sigma_i$ are known

$$\ell = -\frac{1}{2}\sum \frac{(y_i-\mu_i)^2}{\sigma_i^2} + \text{const}$$

## Maximum Likelihood Method

see lectures 05 and 06

Let us consider $N$ independent measurements of variable $Y$. Assuming **measurement fluctuations are described by Gaussian pdf**, the likelihood function is:

$$L = \prod_{i=1}^{N} G(y_i; \mu_i, \sigma_i) = \prod_{i=1}^{N} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left( -\frac{1}{2} \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right)$$

Log-likelihood:

assuming $\sigma_i$ are known

$$\ell = -\frac{1}{2} \sum \frac{(y_i - \mu_i)^2}{\sigma_i^2} + \text{const}$$

We can also define

$$\chi^2 = -2\,\ell = -2 \ln L = \sum_{i=1}^{N} \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

Maximum of (log-)likelihood function corresponds to minimum of $\chi^2$ (for Gaussian pdf!)

## Problem

$\chi^2$ calculated for a set of $N$ measurements is a random varible.
It is a function of random variables, results of the measurement
its value changes when we take another set of $N$ measurements.

Can we predict what its probability distribution is?

## Problem

$\chi^2$ calculated for a set of $N$ measurements is a random varible.
It is a function of random variables, results of the measurement
its value changes when we take another set of $N$ measurements.

Can we predict what its probability distribution is?

We will address this problem in two different approaches:

- simple one, based on intuitive case of $N = 2$, extrapolating to other $N$
- more formal one, based on momentum generating functions

## N=2

Let us introduce "shift" variables:

$$z_i = \frac{y_i - \mu_i}{\sigma_i}$$

which are (by construction) described by Gaussian pdf with $\mu = 0,\ \sigma = 1$.

## N=2

Let us introduce "shift" variables:

$$z_i = \frac{y_i - \mu_i}{\sigma_i}$$

which are (by construction) described by Gaussian pdf with $\mu = 0,\ \sigma = 1$.

For $N = 2$ independent variables we can write:

$$f(z_1, z_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(z_1^2 + z_2^2)\right)$$

## N=2

Let us introduce "shift" variables:

$$z_i = \frac{y_i - \mu_i}{\sigma_i}$$

which are (by construction) described by Gaussian pdf with $\mu = 0, \ \sigma = 1$.

For $N = 2$ independent variables we can write:

$$f(z_1, z_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(z_1^2 + z_2^2)\right)$$

and then change variables to polar coordinates see lecture 04

$$f(r_z, \phi_z) = \frac{1}{2\pi} \, r_z \, \exp\left(-\frac{1}{2} r_z^2\right)$$

**N=2**

Integrating over $\phi_z$ and changing variable to $r_z^2$

$$f(r_z^2) \quad = \quad \frac{1}{2} \exp\left(-\frac{1}{2} r_z^2\right)$$

Distribution is exponential, corresponds to decay time distribution for $\tau = 2$...

# $\chi^2$ distribution

## N=2

Integrating over $\phi_z$ and changing variable to $r_z^2 = \chi^2$

$$f(\chi^2) \;=\; \frac{1}{2}\,\exp\left(-\frac{1}{2}\chi^2\right)$$

Distribution is exponential, corresponds to decay time distribution for $\tau = 2$...

## N=2

Integrating over $\phi_z$ and changing variable to $r_z^2$

$$f(\chi^2) \;=\; \frac{1}{2}\,\exp\left(-\frac{1}{2}\chi^2\right)$$

Distribution is exponential, corresponds to decay time distribution for $\tau = 2$...

## Extrapolation to even N case

Sum of $n = N/2$ numbers from exponential distribution, is distributed according to Gamma distribution with $k = n = N/2$, $\lambda = 1/\tau = 1/2$ lecture 03

$$f(\chi^2) \;=\; \frac{1}{\Gamma(k)}\,(\chi^2)^{k-1}\,\lambda^k\,e^{-\lambda\chi^2}$$

# $\chi^2$ distribution

## N=2

Integrating over $\phi_z$ and changing variable to $r_z^2$

$$f(\chi^2) = \frac{1}{2} \exp\left(-\frac{1}{2}\chi^2\right)$$

Distribution is exponential, corresponds to decay time distribution for $\tau = 2$...

## Extrapolation to even N case

Sum of $n = N/2$ numbers from exponential distribution, is distributed according to Gamma distribution with $k = n = N/2$, $\lambda = 1/\tau = 1/2$ lecture 03

$$f(\chi^2) = \frac{1}{\Gamma(\frac{N}{2})} \left(\frac{1}{2}\right)^{\frac{N}{2}} (\chi^2)^{\frac{N}{2}-1} e^{-\chi^2/2}$$

The formula (as one can expect) works also for odd N...

## Moment generating function

Bonamente

One can consider moment generating function (see lecture 04) for distribution of $u = z^2$ for single measurement ($N = 1$):

## Moment generating function

Bonamente

One can consider moment generating function (see lecture 04) for distribution of $u = z^2$ for single measurement $(N = 1)$:

$$M_1(t) = \mathbb{E}(e^{tu}) = \int_{-\infty}^{+\infty} dz \; f(z) \; e^{tz^2}$$

## Moment generating function

Bonamente

One can consider moment generating function (see lecture 04) for distribution of $u = z^2$ for single measurement ($N = 1$):

$$M_1(t) = \mathbb{E}(e^{tu}) = \int_{-\infty}^{+\infty} dz\ f(z)\ e^{tz^2}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dz\ e^{-z^2(\frac{1}{2}-t)}$$

## Moment generating function

Bonamente

One can consider moment generating function (see lecture 04) for distribution of $u = z^2$ for single measurement ($N = 1$):

$$M_1(t) = \mathbb{E}(e^{tu}) = \int_{-\infty}^{+\infty} dz\ f(z)\ e^{tz^2}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dz\ e^{-z^2(\frac{1}{2}-t)}$$

$z'^2 = z^2(1-2t)$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{dz'}{\sqrt{1-2t}}\ e^{-\frac{1}{2}z'^2}$$

## Moment generating function

Bonamente

One can consider moment generating function (see lecture 04) for distribution of $u = z^2$ for single measurement ($N = 1$):

$$M_1(t) = \mathbb{E}(e^{tu}) = \int_{-\infty}^{+\infty} dz\, f(z)\, e^{tz^2}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dz\, e^{-z^2(\frac{1}{2}-t)}$$

$$z'^2 = z^2(1-2t) \qquad = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{dz'}{\sqrt{1-2t}}\, e^{-\frac{1}{2}z'^2}$$

$$= \frac{1}{\sqrt{1-2t}}$$

## Arbitrary N

Bonamente

Considered random variables $z_i$ are independent, and $\chi^2 = \sum z_i^2$. Moment generating function for $\chi^2$ distribution is thus given by:

$$M_N(t) = \prod_{i=1}^{N} M_1(t) = \left( \frac{1}{1-2t} \right)^{N/2}$$

## Arbitrary N

Bonamente

Considered random variables $z_i$ are independent, and $\chi^2 = \sum z_i^2$. Moment generating function for $\chi^2$ distribution is thus given by:

$$M_N(t) \;=\; \prod_{i=1}^{N} M_1(t) \;=\; \left(\frac{1}{1-2t}\right)^{N/2}$$

We can compare it with the moment generating functions for Gamma pdf

$$M_G(t) \;=\; \mathbb{E}(e^{tx}) \;=\; \int_0^{+\infty} dx\, \frac{1}{\Gamma(k)} x^{k-1}\, \lambda^k\, e^{-\lambda x}\, e^{tx}$$

$$=\; \frac{\lambda^k}{\Gamma(k)} \int_0^{+\infty} dx\; x^{k-1}\; e^{-x(\lambda - t)}$$

## Arbitrary N

Bonamente

Considered random variables $z_i$ are independent, and $\chi^2 = \sum z_i^2$. Moment generating function for $\chi^2$ distribution is thus given by:

$$M_N(t) \;=\; \prod_{i=1}^{N} M_1(t) \;=\; \left(\frac{1}{1-2t}\right)^{N/2}$$

We can compare it with the moment generating functions for Gamma pdf

$$M_G(t) \;=\; \mathbb{E}(e^{tx}) \;=\; \int_0^{+\infty} dx \, \frac{1}{\Gamma(k)} x^{k-1} \, \lambda^k \, e^{-\lambda x} \, e^{tx}$$

$$x' = x(\lambda - t) \qquad = \; \frac{\lambda^k}{\Gamma(k)} \int_0^{+\infty} \frac{dx' \; x'^{k-1}}{(\lambda - t)^k} \, e^{-x'}$$

## Arbitrary N

Bonamente

Considered random variables $z_i$ are independent, and $\chi^2 = \sum z_i^2$. Moment generating function for $\chi^2$ distribution is thus given by:

$$M_N(t) \;=\; \prod_{i=1}^{N} M_1(t) \;=\; \left(\frac{1}{1-2t}\right)^{N/2}$$

We can compare it with the moment generating functions for Gamma pdf

$$M_G(t) \;=\; \mathbb{E}(e^{tx}) \;=\; \int_0^{+\infty} dx\, \frac{1}{\Gamma(k)} x^{k-1}\, \lambda^k\, e^{-\lambda x}\, e^{tx}$$

$$x' = x(\lambda - t) \qquad = \; \frac{\lambda^k}{\Gamma(k)} \int_0^{+\infty} \frac{dx'\; x'^{k-1}}{(\lambda - t)^k}\, e^{-x'} \;=\; \left(\frac{1}{1-\frac{t}{\lambda}}\right)^k$$

## $\chi^2$ distribution

We conclude that distribution of $\chi^2$ is described by Gamma pdf with:

$$k = \frac{N}{2} \quad \text{and} \quad \lambda = \frac{1}{2}$$

## $\chi^2$ distribution

We conclude that distribution of $\chi^2$ is described by Gamma pdf with:

$$k = \frac{N}{2} \quad \text{and} \quad \lambda = \frac{1}{2}$$

Properties of the $\chi^2$ distribution (see lecture 03)

$$\langle \chi^2 \rangle = \frac{k}{\lambda} = N$$

$$\mathbb{V}(\chi^2) = \frac{k}{\lambda^2} = 2N$$

## $\chi^2$ distribution

We conclude that distribution of $\chi^2$ is described by Gamma pdf with:

$$k = \frac{N}{2} \quad \text{and} \quad \lambda = \frac{1}{2}$$

Properties of the $\chi^2$ distribution (see lecture 03)

$$\langle \chi^2 \rangle = \frac{k}{\lambda} = N$$

$$\mathbb{V}(\chi^2) = \frac{k}{\lambda^2} = 2N$$

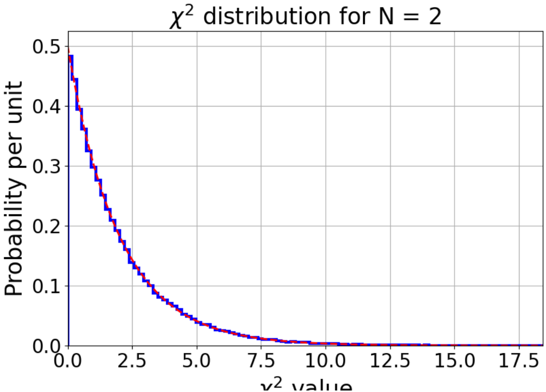
$$\sqrt{\mathbb{V}(\chi^2)} = \sigma_{\chi^2} = \sqrt{2N}$$

For small N, value of $\chi^2$ is a subject to large fluctuations...

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

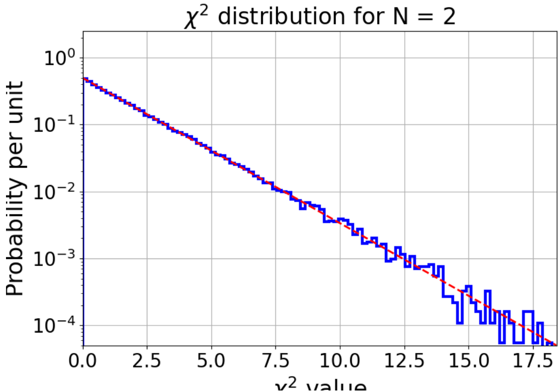



Exponential distribution

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

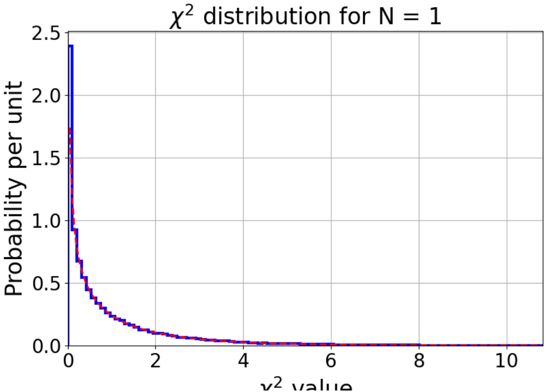



Exponential distribution

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

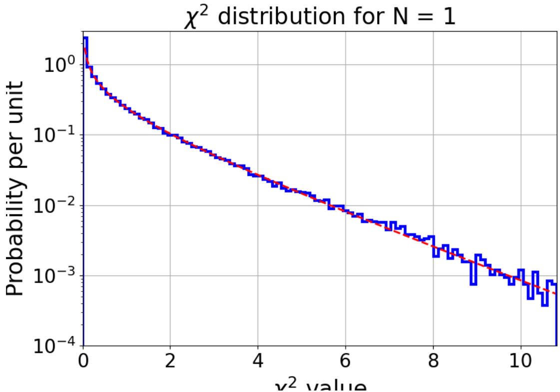



Sharply peaked at zero, but with long tail

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

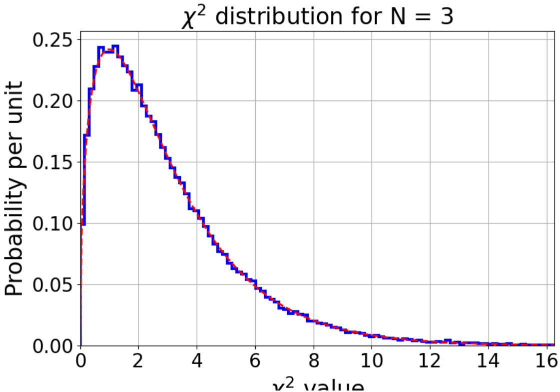



Sharply peaked at zero, but with long tail

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

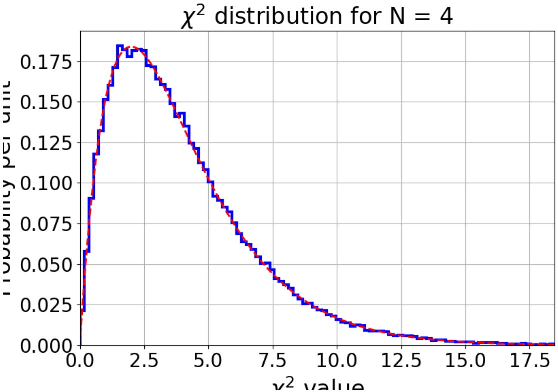



Very asymmetric, most events below average value...

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

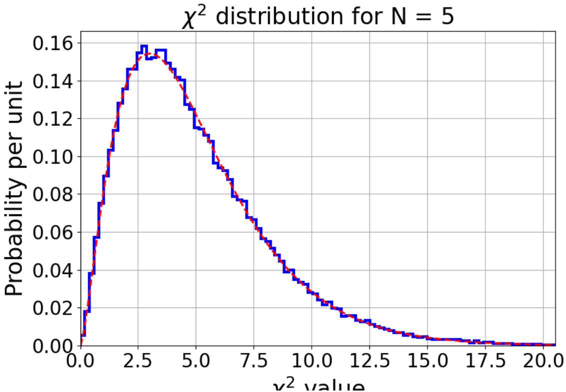



Very asymmetric, most events below average value...

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

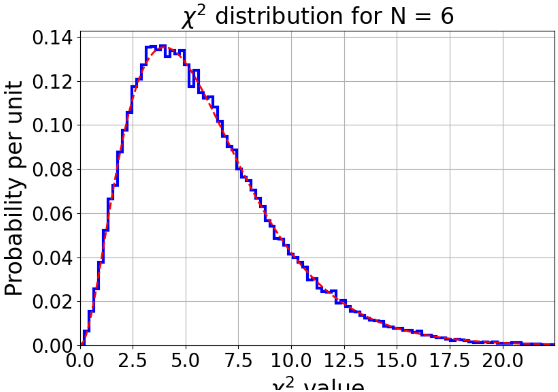



Very asymmetric, most events below average value...

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

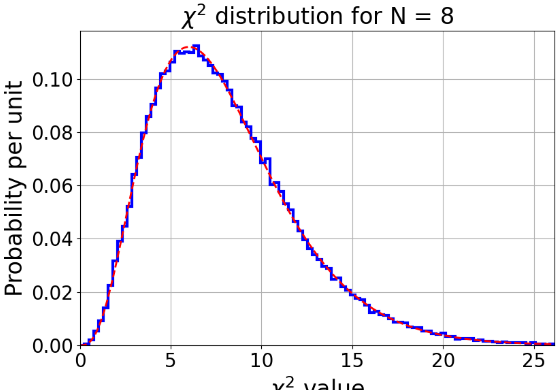



It is interesting to note that maximum position, $\chi^2_{max} = N - 2$ (!!!)

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

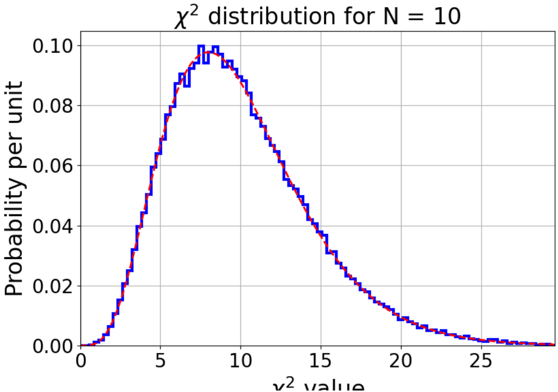



It is interesting to note that maximum position, $\chi^2_{max} = N - 2$ (!!!)

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)

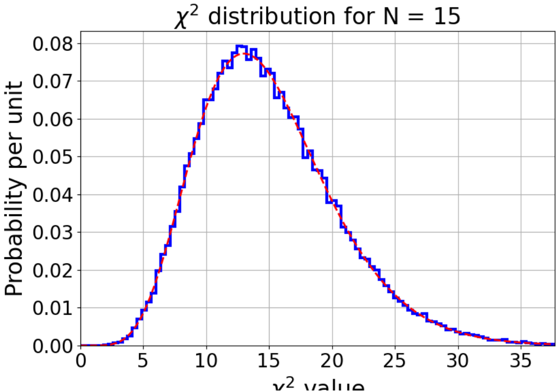



It is interesting to note that maximum position, $\chi^2_{max} = N - 2$ (!!!)

## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

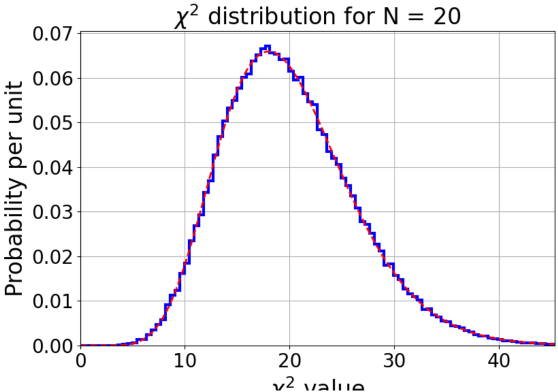
Results of the Monte Carlo sample generation (compared with predictions)



## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

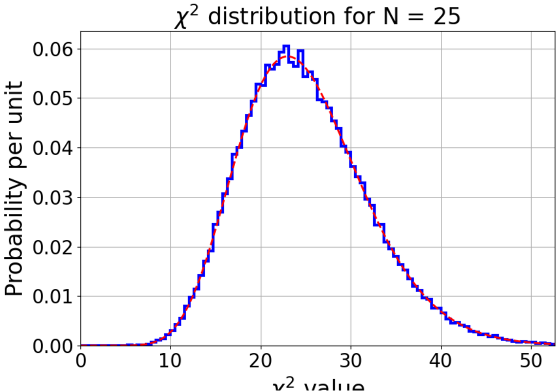
Results of the Monte Carlo sample generation (compared with predictions)



## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

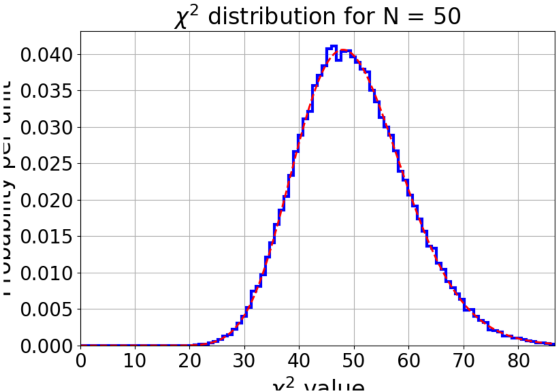
Results of the Monte Carlo sample generation (compared with predictions)



## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

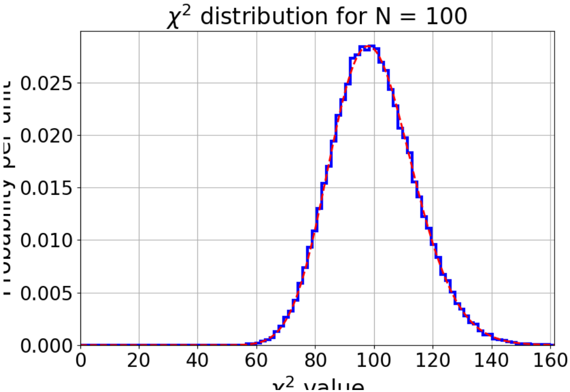
Results of the Monte Carlo sample generation (compared with predictions)



## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

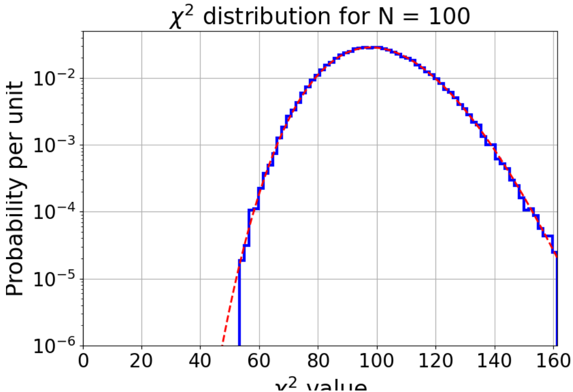
Results of the Monte Carlo sample generation (compared with predictions)



## $\chi^2$ distribution

08_chi2.ipynb Open in Colab

Results of the Monte Carlo sample generation (compared with predictions)




Almost gaussian, but asymmetry in tails remains even for large $N$...

## Reduced $\chi^2$

When discussing consistency of large data samples it is often convenient to use value of "reduced $\chi^2$":

$$\chi^2_{red} = \frac{\chi^2}{N}$$

Distribution of $\chi^2_{red}$ is again described by the Gamma pdf with

$$k = \frac{N}{2} \quad \text{and} \quad \lambda = \frac{N}{2}$$

## Reduced $\chi^2$

When discussing consistency of large data samples it is often convenient to use value of "reduced $\chi^2$":

$$\chi^2_{red} \;=\; \frac{\chi^2}{N}$$

Distribution of $\chi^2_{red}$ is again described by the Gamma pdf with

$$k = \frac{N}{2} \quad \text{and} \quad \lambda = \frac{N}{2}$$

Properties of the distribution:

$$\langle \chi^2_{red} \rangle \;=\; 1 \qquad \mathbb{V}(\chi^2_{red}) \;=\; \sigma^2_{\chi^2_{red}} \;=\; \frac{2}{N} \qquad \chi^2_{red}\Big|_{\mathrm{p=max}} = 1 - \frac{2}{N} \quad (N > 2)$$

## Number of degrees of freedom

So far, we have only considered an ideal case, where both the expected values $\boldsymbol{\mu}$ and measurement uncertainties $\boldsymbol{\sigma}$ are known.

However, it is quite a common situation, when the expected value is extracted from the data:

$$\tilde{\chi}^2 = \sum_{i=1}^{N} \frac{(y_i - \bar{y})^2}{\sigma^2}$$

where we assume uniform uncertainties for simplicity.

What is the expected distribution for $\tilde{\chi}^2$ ?

Mean value corresponds to maximum likelihood $\Rightarrow$ $\tilde{\chi}^2 \leq \chi^2$

## Number of degrees of freedom

We already know (lecture 04) that unbiased variance estimate for N measurements is

$$s^2 = \frac{1}{N-1} \sum_i (y_i - \bar{y})^2$$

so one can conclude:

$$\langle \tilde{\chi}^2 \rangle = N - 1$$

but this does not give us full information about the distribution...

## Number of degrees of freedom

We already know (lecture 04) that unbiased variance estimate for N measurements is

$$s^2 = \frac{1}{N-1} \sum_i (y_i - \bar{y})^2$$

so one can conclude:

$$\langle \tilde{\chi}^2 \rangle = N - 1$$

but this does not give us full information about the distribution...

Simple variable transformation can be used: (Brandt)

$$x_1 = \frac{1}{\sqrt{2}}(y_1 - y_2)$$

$$x_2 = \frac{1}{\sqrt{2 \cdot 3}}(y_1 + y_2 - y_3)$$

$$\ldots$$

## Number of degrees of freedom

We already know (lecture 04) that unbiased variance estimate for N measurements is

$$s^2 = \frac{1}{N-1} \sum_i (y_i - \bar{y})^2$$

so one can conclude:

$$\langle \tilde{\chi}^2 \rangle = N - 1$$

but this does not give us full information about the distribution...

Simple variable transformation can be used: (Brandt)

$$x_k = \frac{1}{\sqrt{k(k+1)}} (y_1 + \ldots + y_k \; - \; y_{k+1}) \qquad k = 1 \ldots N-1$$

$$x_N = \sqrt{N} \cdot \bar{y}$$

One can verify that this is an orthogonal transformation...

## Number of degrees of freedom

If $y_i$ are independent random variables with Gaussian pdf, so are $x_i$. Also:

$$\sum_{i=1}^{N} x_i^2 \; = \; \sum_{i=1}^{N} y_i^2$$

## Number of degrees of freedom

If $y_i$ are independent random variables with Gaussian pdf, so are $x_i$. Also:

$$\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i^2$$

We can now rewrite the formula for $\tilde{\chi}^2$ in the new basis:

$$\sigma^2 \cdot \tilde{\chi}^2 = \sum_{i=1}^{N} (y_i - \bar{y})^2 = \sum_{i=1}^{N} y_i^2 - 2\bar{y} \sum_{i=1}^{N} y_i + N\bar{y}^2$$

$$= \sum_{i=1}^{N} y_i^2 - N\bar{y}^2$$

## Number of degrees of freedom

If $y_i$ are independent random variables with Gaussian pdf, so are $x_i$. Also:

$$\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i^2$$

We can now rewrite the formula for $\tilde{\chi}^2$ in the new basis:

$$\begin{aligned}
\sigma^2 \cdot \tilde{\chi}^2 &= \sum_{i=1}^{N} (y_i - \bar{y})^2 = \sum_{i=1}^{N} y_i^2 - 2\bar{y} \sum_{i=1}^{N} y_i + N\bar{y}^2 \\
&= \sum_{i=1}^{N} y_i^2 - N\bar{y}^2 = \sum_{i=1}^{N} x_i^2 - x_N^2 = \sum_{i=1}^{N-1} x_i^2
\end{aligned}$$

$\Rightarrow$ distribution of $\tilde{\chi}^2$ corresponds to that of $\chi^2$ for $N_{df} = N - 1$ variables...

## Least-squares method

## Data consistency test

Value of $\chi^2$ (or $\chi^2_{red}$) can be used to verify the consistency of the given data set $\mathbf{y}$ (with uncertainties $\boldsymbol{\sigma}$) with the model predictions given by $\boldsymbol{\mu}$

We can try to test the theoretical model, verify our estimates of measurement uncertainties, or check the consistency of the experimental procedure...

If the model does not describe the data, higher $\chi^2$ values are expected.

How to quantify the level of agreement?

## Data consistency test

Value of $\chi^2$ (or $\chi^2_{red}$) can be used to verify the consistency of the given data set $\mathbf{y}$ (with uncertainties $\boldsymbol{\sigma}$) with the model predictions given by $\boldsymbol{\mu}$

We can try to test the theoretical model, verify our estimates of measurement uncertainties, or check the consistency of the experimental procedure...

If the model does not describe the data, higher $\chi^2$ values are expected.

How to quantify the level of agreement?

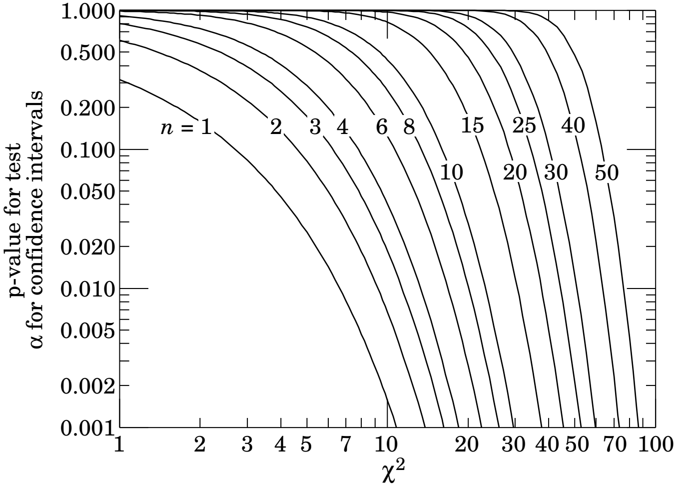
We can calculate the probability of obtaining given value of $\chi^2$ or lower:

$$P(\chi^2) \;=\; \int_0^{\chi^2} d\chi^{2\prime} \, f(\chi^{2\prime})$$

given by the cumulative probability distribution. $1 - P(\chi^2)$ is sometimes referred to as *p*-value

## Data consistency test

Plot of $p$-values
as a function of $\chi^2$
for different $N_{df}$




(PDG)

## Critical $\chi^2$

The other approach is to define, for given probability $CL$ (confidence level) the critical value of $\chi^2$, corresponding the the frequentist upper limit:

$$\int_0^{\chi^2_{crit}} d\chi^2 \, f(\chi^2) \; = \; CL \qquad\qquad \int_{\chi^2_{crit}}^{+\infty} d\chi^2 \, f(\chi^2) \; = \; 1 - CL \; = \; \alpha$$

## Critical $\chi^2$

The other approach is to define, for given probability $CL$ (confidence level) the critical value of $\chi^2$, corresponding the the frequentist upper limit:

$$\int_0^{\chi^2_{crit}} d\chi^2 \, f(\chi^2) \; = \; CL \qquad\qquad \int_{\chi^2_{crit}}^{+\infty} d\chi^2 \, f(\chi^2) \; = \; 1 - CL \; = \; \alpha$$

If the $\chi^2$ value obtained in the actual measurement is higher than the selected $\chi^2_{crit}$, then we should reject the hypothesis of data consistency with the model (can still be due to the data, not the wrong model).

## Critical $\chi^2$

The other approach is to define, for given probability $CL$ (confidence level) the critical value of $\chi^2$, corresponding the the frequentist upper limit:

$$\int_0^{\chi^2_{crit}} d\chi^2 \, f(\chi^2) \; = \; CL \qquad \int_{\chi^2_{crit}}^{+\infty} d\chi^2 \, f(\chi^2) \; = \; 1 - CL \; = \; \alpha$$

If the $\chi^2$ value obtained in the actual measurement is higher than the selected $\chi^2_{crit}$, then we should reject the hypothesis of data consistency with the model (can still be due to the data, not the wrong model).

Very low $P$ values, $P(\chi^2) \ll 1$, are also not expected (not likely)!
If $\chi^2 \ll N$ (except for very small $N$), this usually indicates a problem:

- overestimated uncertainties of measurements (or correlations not properly included)
- hidden correlations between measurements (which we treat as independent variables)

## Critical $\chi^2$

Table of critical $\chi^2$ values (Brand)

$$P = \int_0^{\chi_P^2} f(\chi^2; f)\, d\chi^2$$

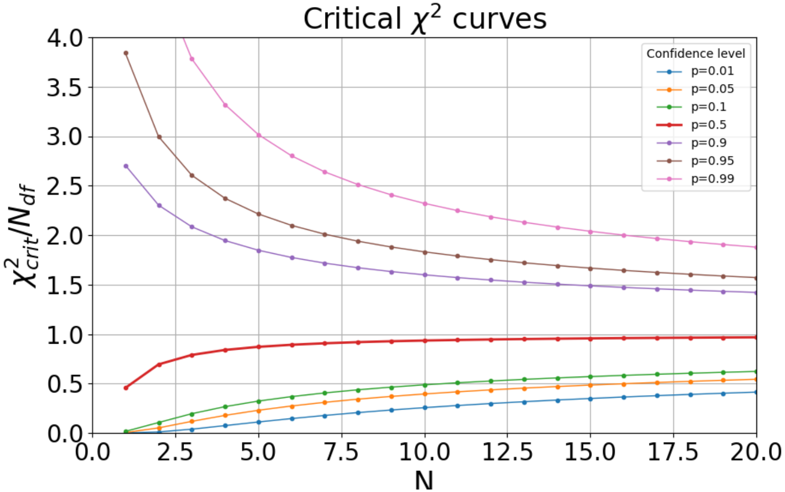
| $f$ | $P$ = 0.900 | 0.950 | 0.990 | 0.995 | 0.999 |
|---|---|---|---|---|---|
| 1 | 2.706 | 3.841 | 6.635 | 7.879 | 10.828 |
| 2 | 4.605 | 5.991 | 9.210 | 10.597 | 13.816 |
| 3 | 6.251 | 7.815 | 11.345 | 12.838 | 16.266 |
| 4 | 7.779 | 9.488 | 13.277 | 14.860 | 18.467 |
| 5 | 9.236 | 11.070 | 15.086 | 16.750 | 20.515 |
| 6 | 10.645 | 12.592 | 16.812 | 18.548 | 22.458 |
| 7 | 12.017 | 14.067 | 18.475 | 20.278 | 24.322 |
| 8 | 13.362 | 15.507 | 20.090 | 21.955 | 26.124 |
| 9 | 14.684 | 16.919 | 21.666 | 23.589 | 27.877 |
| 10 | 15.987 | 18.307 | 23.209 | 25.188 | 29.588 |

## Critical $\chi^2$

08_critical.ipynb Open in Colab

Plot of critical values for reduced $\chi^2$

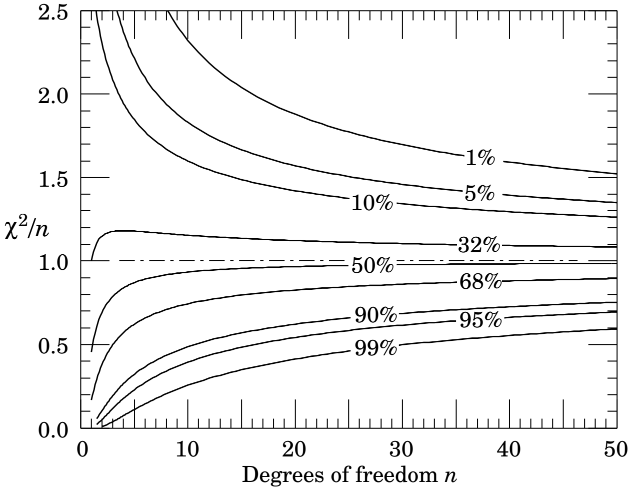
p=0.5 shows the median



## Critical $\chi^2$

Plot of critical values
for reduced $\chi^2$
(indicated is $p = 1 - P$)

(PDG)



## Student's $t$ Distribution

We can verify consistency of the series of measurements **x** with the true value $\mu$ by looking at the shift parameter for the mean

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

where mean value $\bar{x}$ is the best estimate of $\mu$ assuming Gaussian pdf.

## Student's $t$ Distribution

We can verify consistency of the series of measurements **x** with the true value $\mu$ by looking at the shift parameter for the mean

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

where mean value $\bar{x}$ is the best estimate of $\mu$ assuming Gaussian pdf.

But this works only, if we know the uncertainty, $\sigma_{\bar{x}} = \sigma/\sqrt{N}$.
We need to know measurement uncertainties to calculate $\chi^2$ !...

## Student's $t$ Distribution

We can verify consistency of the series of measurements **x** with the true value $\mu$ by looking at the shift parameter for the mean

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

where mean value $\bar{x}$ is the best estimate of $\mu$ assuming Gaussian pdf.

But this works only, if we know the uncertainty, $\sigma_{\bar{x}} = \sigma/\sqrt{N}$.
We need to know measurement uncertainties to calculate $\chi^2$ !...

If the measurement uncertainties are unknown, or not reliable, we can estimate the variance of the sample from the data itself (lecture 04)

$$\hat{s}^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

$\hat{s}^2$ distribution corresponds to $\chi^2$ distribution for $N-1$ degrees of freedom

## Student's $t$ Distribution

Consistency of our measurements **x** with the true value $\mu$ can be now described by

$$t = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{N}}$$

but the distribution of $t$ is no longer Gaussian (due to $\hat{s}$ being a random variable as well).

## Student's $t$ Distribution

Consistency of our measurements **x** with the true value $\mu$ can be now described by

$$t = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{N}}$$

but the distribution of $t$ is no longer Gaussian (due to $\hat{s}$ being a random variable as well).
It can still be calculated analytically:

$$f(t;n) = \frac{1}{\sqrt{n\,\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

where $n$ is the number of degrees of freedom, $n = N - 1$.

## Student's $t$ Distribution

Consistency of our measurements **x** with the true value $\mu$ can be now described by

$$t = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{N}}$$

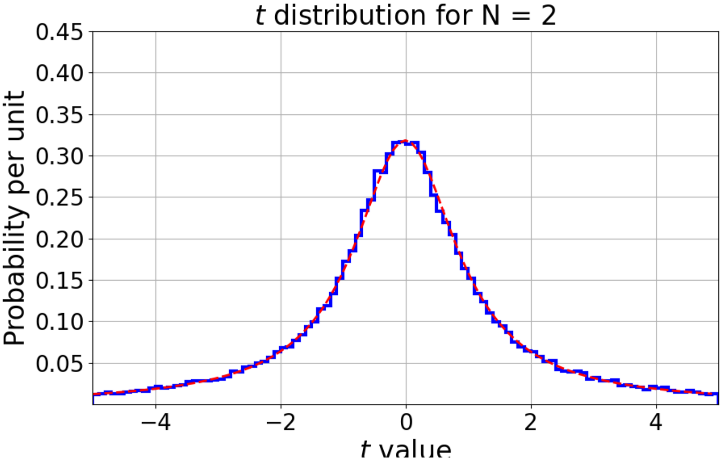
but the distribution of $t$ is no longer Gaussian (due to $\hat{s}$ being a random variable as well).
It can still be calculated analytically:

$$f(t;n) = \frac{1}{\sqrt{n\,\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

where $n$ is the number of degrees of freedom, $n = N - 1$.
Distribution is symmetric and has a mean of zero, but larger tails than the Gaussian distribution, for small $N$ in particular.

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab

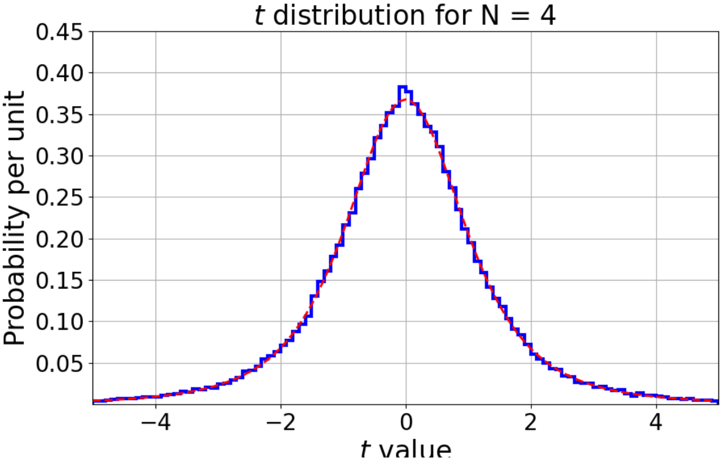



Shape of the $t$ distribution for different numbers of measurements

$$N_{df} = N - 1 = 1$$

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab

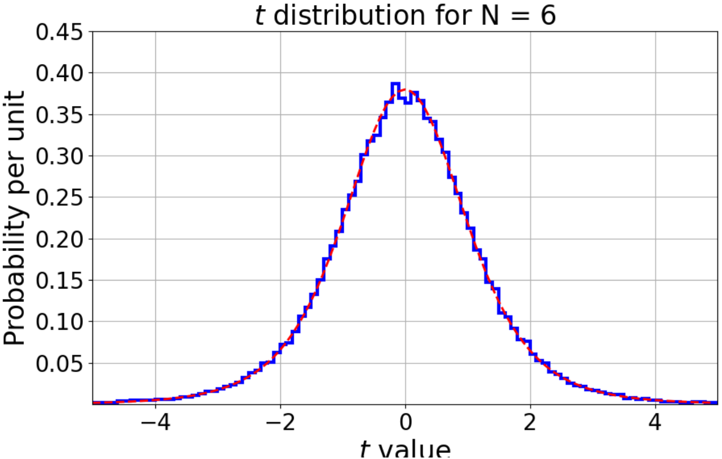



Shape of the $t$ distribution for different numbers of measurements

$$N_{df} = N - 1 = 3$$

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab

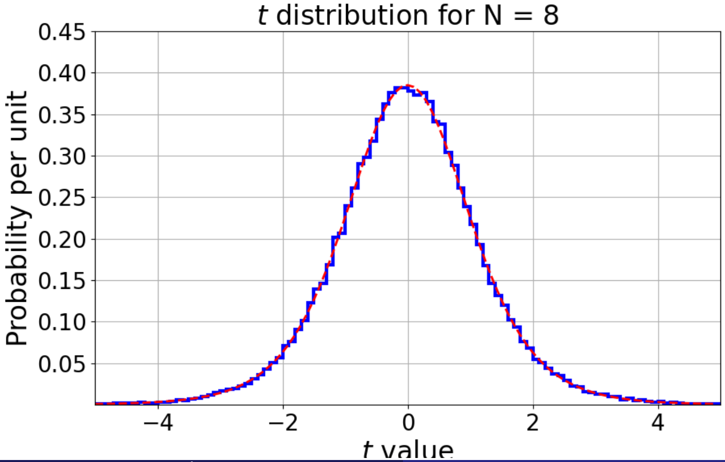



Shape of the $t$ distribution for different numbers of measurements

$$N_{df} = N - 1 = 5$$

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab

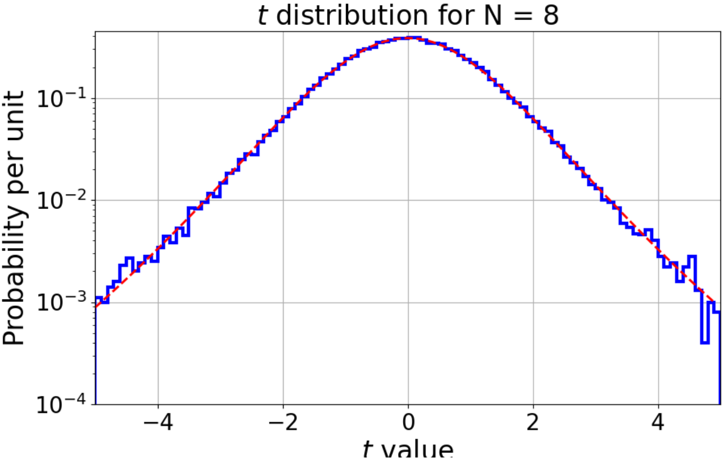



Shape of the $t$ distribution for different numbers of measurements

$N_{df} = N - 1 = 7$

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab



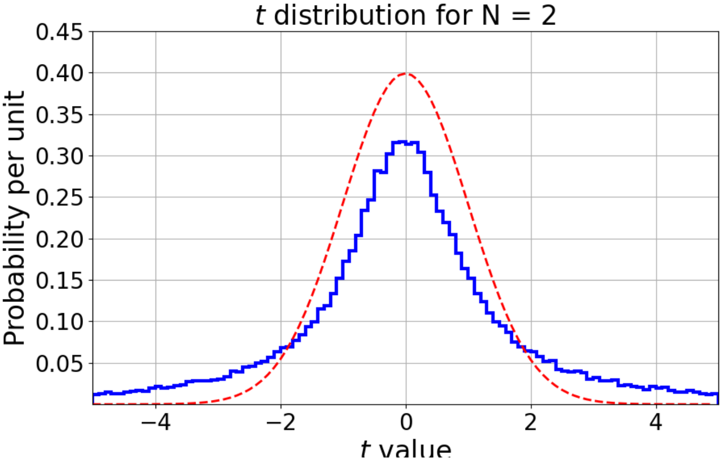

Shape of the $t$ distribution for different numbers of measurements

$N_{df} = N - 1 = 7$

tails are clearly non-Gaussian...

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab

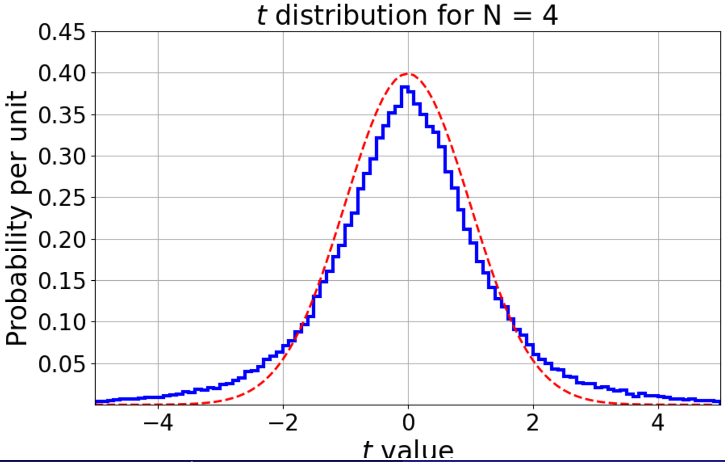



Shape of the $t$ distribution compared with Gaussian distribution

$$N_{df} = N - 1 = 1$$

## Student's $t$ Distribution

08_t-dist.ipynb  Open in Colab

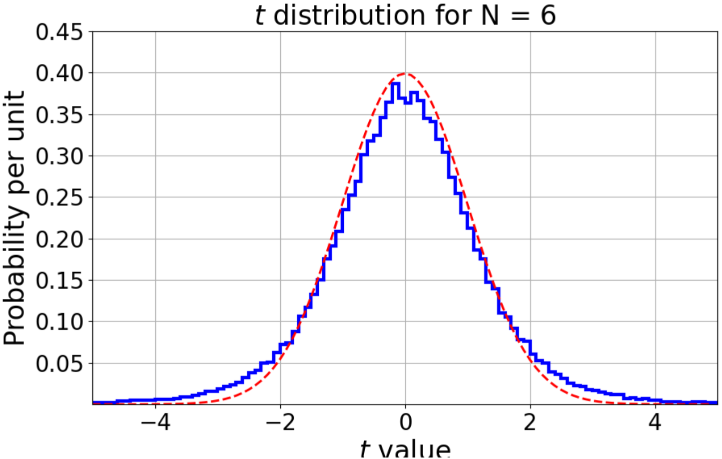



Shape of the $t$ distribution compared with Gaussian distribution

$N_{df} = N - 1 = 3$

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab

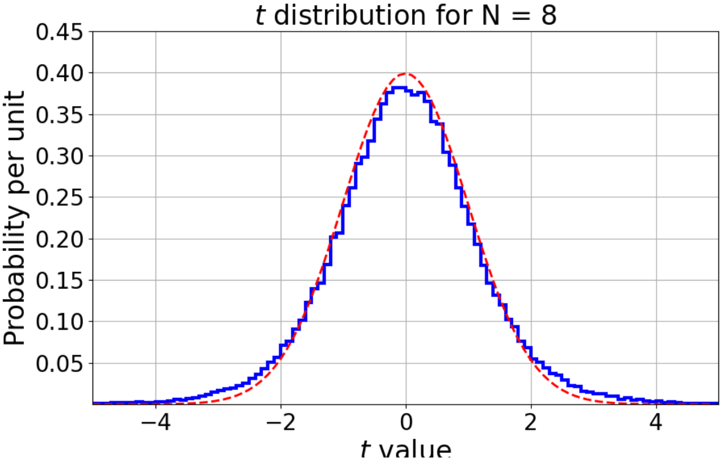



Shape of the $t$ distribution compared with Gaussian distribution

$N_{df} = N - 1 = 5$

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab



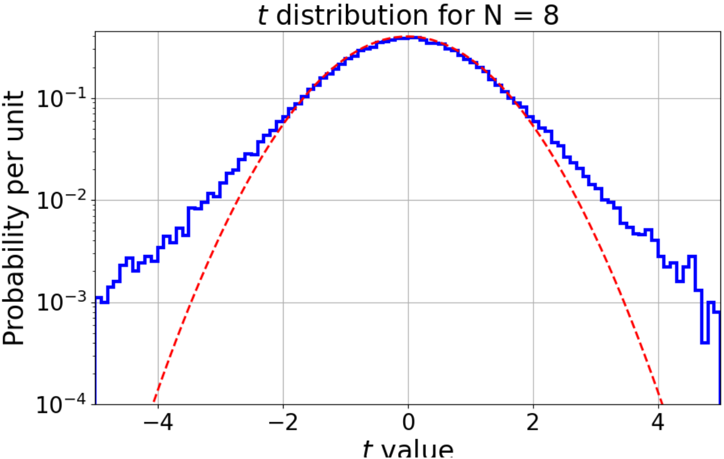

Shape of the $t$ distribution compared with Gaussian distribution

$N_{df} = N - 1 = 7$

## Student's $t$ Distribution

08_t-dist.ipynb

Open in Colab

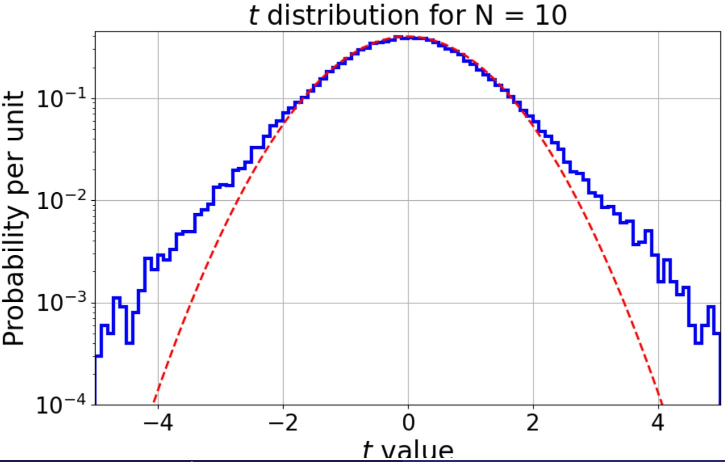



Shape of the $t$ distribution compared with Gaussian distribution

$N_{df} = N - 1 = 7$

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab

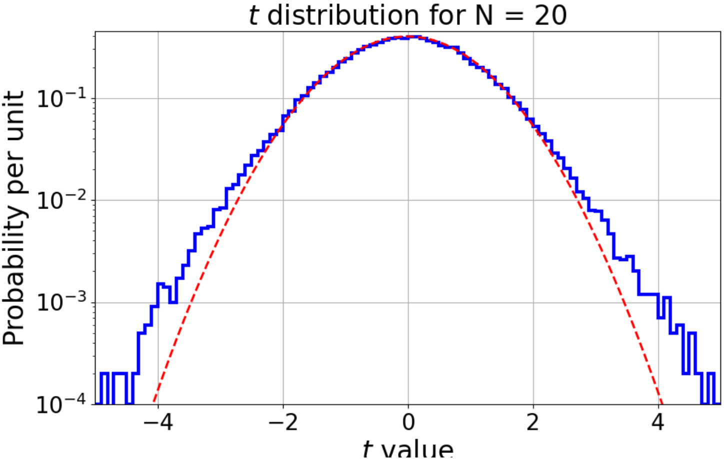



Shape of the $t$ distribution compared with Gaussian distribution

Probability of large fluctuations still significantly enhanced!

## Student's $t$ Distribution

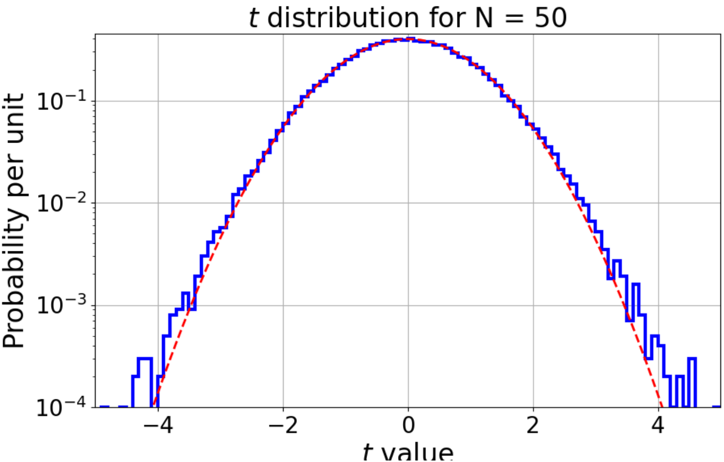
08_t-dist.ipynb Open in Colab




Shape of the $t$ distribution compared with Gaussian distribution

## Student's $t$ Distribution

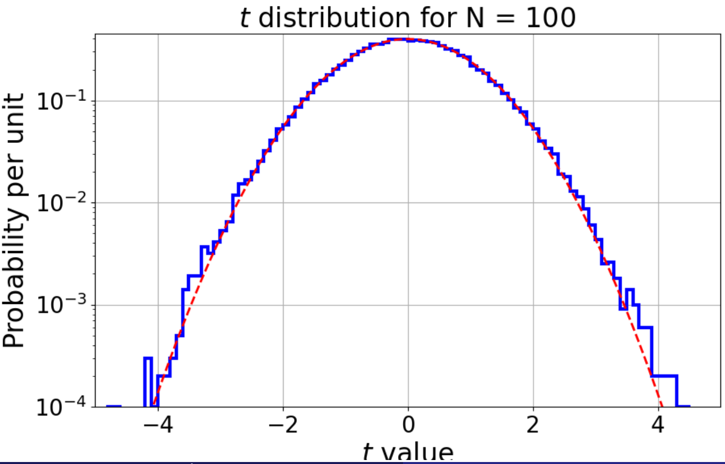
08_t-dist.ipynb Open in Colab




Shape of the $t$ distribution compared with Gaussian distribution

## Student's $t$ Distribution

08_t-dist.ipynb Open in Colab




Shape of the $t$ distribution compared with Gaussian distribution

Converges to the Gaussian distribution for large $N$

## Student's $t$ Distribution (Brandt)

"Critical values" of $t$ for small numbers of degrees of freedom $f$

$$P = \int_{-\infty}^{t_P} f(t;f)\,\mathrm{d}t$$

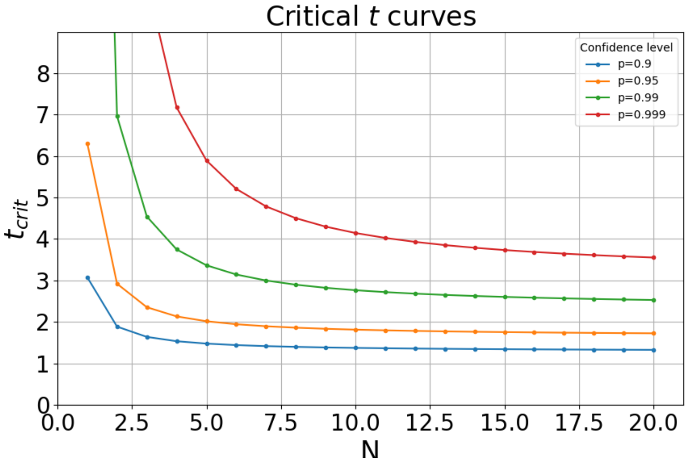
| $f$ | $P$ = 0.9000 | 0.9500 | 0.9750 | 0.9900 | 0.9950 | 0.9990 | 0.9995 |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.309 | 636.619 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 | 31.599 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.215 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |

## Student's $t$ Distribution

08_t-limit.ipynb Open in Colab

Plot of critical values for $t$

Large deviations much more probable for small $N$



## Least-squares method

## General case

We introduced $\chi^2$ in a very general form:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

where different $\mu_i$ and $\sigma_i$ are possible for each of N measurement $y_i$

## General case

We introduced $\chi^2$ in a very general form:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

where different $\mu_i$ and $\sigma_i$ are possible for each of N measurement $y_i$

It is quite often the case that values of $\mu_i$ depend on some controlled variables $x_i$ and a smaller set of model parameters:

$$\mu_i = \mu(x_i; \mathbf{a})$$

we can then use the least-squares method to extract the best estimates of parameters **a** from the collected set of data points $(x_i, y_i)$

We can look for minimum of $\chi^2$ using different numerical algorithms...

## Linear case

The case which is particularly interesting is when the dependence is linear in parameters (!):

$$\mu(x; \mathbf{a}) \; = \; \sum_{k=1}^{M} a_k \, f_k(x)$$

where $f_k(x)$ is a set of functions with arbitrary analytical form.

## Linear case

The case which is particularly interesting is when the dependence is linear in parameters (!):

$$\mu(x; \mathbf{a}) = \sum_{k=1}^{M} a_k \, f_k(x)$$

where $f_k(x)$ is a set of functions with arbitrary analytical form.

One of the examples is the polynomial series:

$$f_k(x) = x^k \qquad \Rightarrow \quad \mu(x; \mathbf{a}) = \sum_{k=1}^{M} a_k \, x^{k-1}$$

but any set of functions can be used, if they are not linearly dependent.
Set of functions ortogonal for a given set of points $x_i$ should work best...

## Parameter fit

Bonamente

Least-squares method is the special case of the maximum likelihood approach, when we can assume Gaussian pdf for measurements $y_i$.

As the best estimate of the parameter set **a** we choose the parameter values which correspond to the (global) $\chi^2$ minimum ($\Rightarrow$ maximum of log-likelihood)

## Parameter fit

Bonamente

Least-squares method is the special case of the maximum likelihood approach, when we can assume Gaussian pdf for measurements $y_i$.

As the best estimate of the parameter set **a** we choose the parameter values which correspond to the (global) $\chi^2$ minimum ($\Rightarrow$ maximum of log-likelihood)

To look for $\chi^2$ maximum, we consider partial derivatives:

$$\frac{\partial \chi^2}{\partial a_l} = \frac{\partial}{\partial a_l} \sum_{i=1}^{N} \left( \frac{y_i - \sum_{k=1}^{M} a_k \, f_k(x_i)}{\sigma_i} \right)^2 = 0$$

## Parameter fit

Bonamente

Least-squares method is the special case of the maximum likelihood approach, when we can assume Gaussian pdf for measurements $y_i$.

As the best estimate of the parameter set **a** we choose the parameter values which correspond to the (global) $\chi^2$ minimum (⇒ maximum of log-likelihood)

To look for $\chi^2$ maximum, we consider partial derivatives:

$$\frac{\partial \chi^2}{\partial a_l} = \frac{\partial}{\partial a_l} \sum_{i=1}^{N} \left( \frac{y_i - \sum_{k=1}^{M} a_k \, f_k(x_i)}{\sigma_i} \right)^2$$

$$= -2 \sum_{i=1}^{N} \left( \frac{y_i - \sum_{k=1}^{M} a_k \, f_k(x_i)}{\sigma_i^2} \right) f_l(x_i)$$

## Parameter fit

Bonamente

We obtain a set of M equations for M parameters $a_i$:

$$\sum_{i=1}^{N} \frac{f_l(x_i)}{\sigma_i^2} \left( y_i - \sum_{k=1}^{M} a_k \; f_k(x_i) \right) \quad = \quad 0 \qquad\qquad l = 1 \ldots M$$

## Parameter fit

Bonamente

We obtain a set of M equations for M parameters $a_i$:

$$\sum_{i=1}^{N} \frac{f_l(x_i)}{\sigma_i^2} \left( y_i - \sum_{k=1}^{M} a_k \, f_k(x_i) \right) \;=\; 0 \qquad l = 1 \ldots M$$

which can be rewritten as:

$$\sum_{k=1}^{M} \left( \sum_{i=1}^{N} \frac{f_l(x_i) \, f_k(x_i)}{\sigma_i^2} \right) a_k \;=\; \sum_{i=1}^{N} \frac{f_l(x_i) \, y_i}{\sigma_i^2}$$

## Parameter fit

Bonamente

We obtain a set of M equations for M parameters $a_i$:

$$\sum_{i=1}^{N} \frac{f_l(x_i)}{\sigma_i^2} \left( y_i - \sum_{k=1}^{M} a_k \, f_k(x_i) \right) \;=\; 0 \qquad l = 1 \dots M$$

which can be rewritten as:

$$\sum_{k=1}^{M} \left( \sum_{i=1}^{N} \frac{f_l(x_i) \; f_k(x_i)}{\sigma_i^2} \right) a_k \;=\; \sum_{i=1}^{N} \frac{f_l(x_i) \; y_i}{\sigma_i^2}$$

or in the matrix form:

$$\mathbb{A} \cdot \mathbf{a} \;=\; \mathbf{b}$$

where $\mathbb{A}_{lk} \;=\; \sum_{i=1}^{N} \frac{f_l(x_i) \; f_k(x_i)}{\sigma_i^2}$ and $b_l \;=\; \sum_{i=1}^{N} \frac{f_l(x_i) \; y_i}{\sigma_i^2}$

## Parameter fit

Solution of this set of equations can be obtained by inverting matrix $\mathbb{A}$

$$\mathbf{a} \;=\; \mathbb{A}^{-1} \cdot \mathbf{b}$$

## Parameter fit

Solution of this set of equations can be obtained by inverting matrix $\mathbb{A}$

$$\mathbf{a} \;=\; \mathbb{A}^{-1} \cdot \mathbf{b}$$

This also gives us the estimate of parameter covariance matrix:

$$\mathbb{C}_{\mathbf{a}} \;=\; \left( - \frac{\partial^2 \ell}{\partial a_l \, \partial a_k} \right)^{-1} = \left( \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_l \, \partial a_k} \right)^{-1} = \mathbb{A}^{-1}$$

## Parameter fit

Solution of this set of equations can be obtained by inverting matrix $\mathbb{A}$

$$\mathbf{a} \;=\; \mathbb{A}^{-1} \cdot \mathbf{b}$$

This also gives us the estimate of parameter covariance matrix:

$$\mathbb{C}_{\mathbf{a}} \;=\; \left( - \frac{\partial^2 \ell}{\partial a_l \, \partial a_k} \right)^{-1} = \left( \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_l \, \partial a_k} \right)^{-1} = \mathbb{A}^{-1}$$

One can write

$$\mathbb{C}_{\mathbf{a}} \;=\; \left( \sum_{i=1}^{N} \frac{f_l(x_i)\; f_k(x_i)}{\sigma_i^2} \right)^{-1}$$

Expected uncertainties of the extracted parameter values depend on the choice of measurement points $x_i$ but, surprisingly, do not depend on the actual results $y_i$

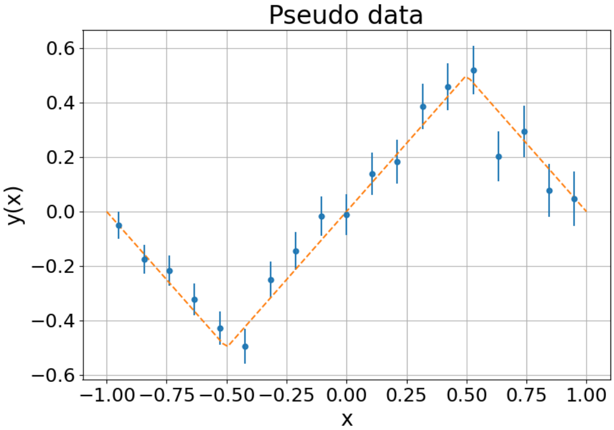
$\Rightarrow$ very useful when planning the experiment...

## Linear fit example

08_fit1.ipynb Open in Colab

Fitting Fourier series to example data set

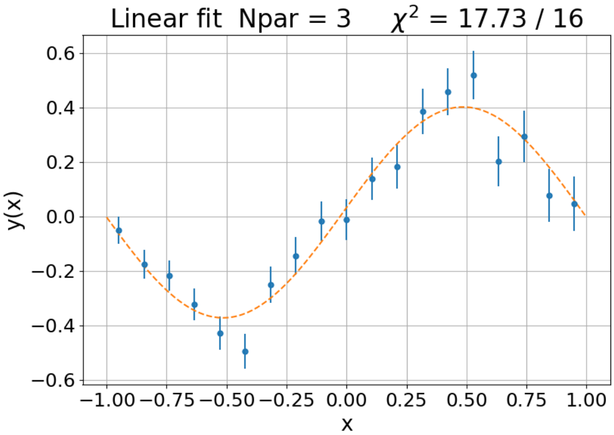
example model ⇒



## Linear fit example

08_fit1.ipynb Open in Colab

Fitting Fourier series to example data set

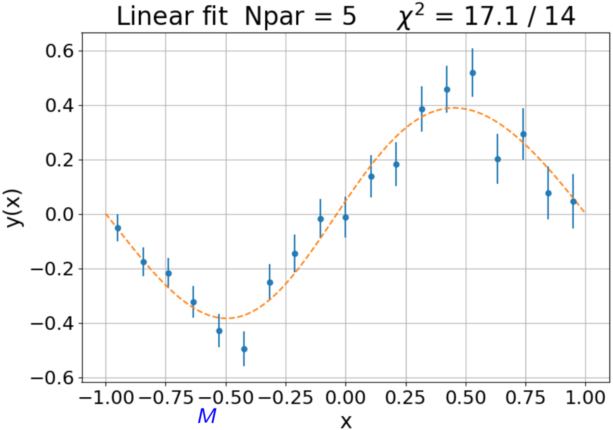



$$y(x) = a_0 + a_1 \sin(x) + a_2 \cos(x)$$

## Linear fit example

08_fit1.ipynb Open in Colab

Fitting Fourier series to example data set



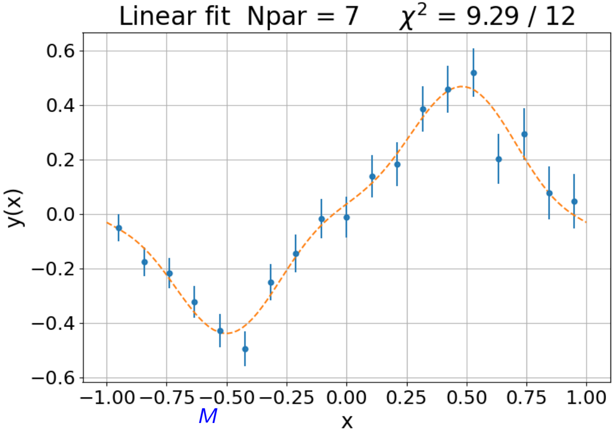

$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin(nx) + a_{2n} \cos(nx) \qquad M = 2$$

## Linear fit example

08_fit1.ipynb Open in Colab

Fitting Fourier series to example data set

Probably optimal choice



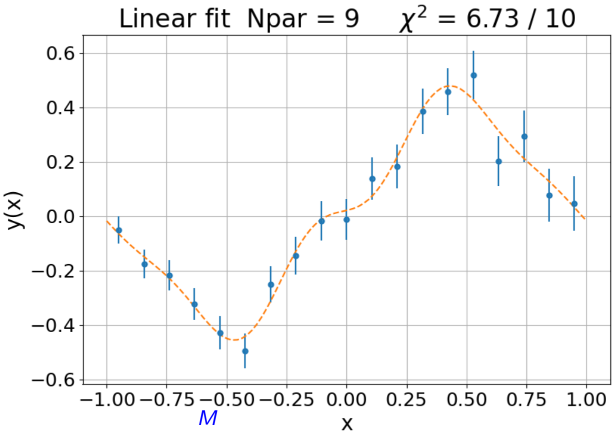

$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin(nx) + a_{2n} \cos(nx) \qquad M = 3$$

## Linear fit example

08_fit1.ipynb Open in Colab

Fitting Fourier series to example data set

"Overtraining ?"




$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin(nx) + a_{2n} \cos(nx) \qquad M = 4$$

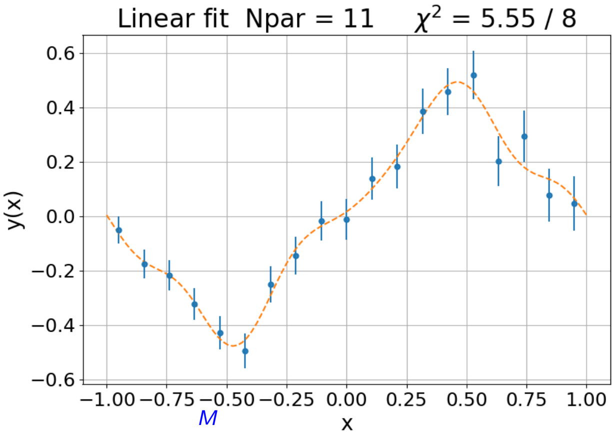
## Linear fit example

08_fit1.ipynb

Open in Colab

Fitting Fourier series to example data set

"Overtraining ?"



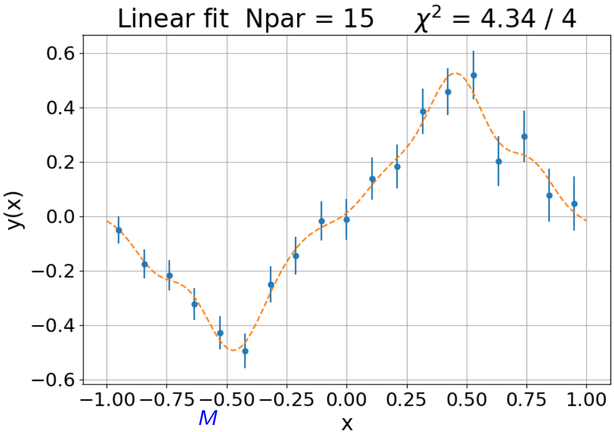

$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin(nx) + a_{2n} \cos(nx) \qquad M = 5$$

## Linear fit example

08_fit1.ipynb Open in Colab

Fitting Fourier series to example data set

"Overtraining ?"




$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin(nx) + a_{2n} \cos(nx) \qquad M = 6$$

## Best fit choice?

If the functional dependence is not predicted by any theory/model, we can try to fit a polynomial or function series. When should we stop?

How can we recognize that we have too many parameters?
Beside looking at the $P$ value corresponding to the obtained $\chi^2$

## Best fit choice?

If the functional dependence is not predicted by any theory/model, we can try to fit a polynomial or function series. When should we stop?

How can we recognize that we have too many parameters?
Beside looking at the $P$ value corresponding to the obtained $\chi^2$

- Adding new parameter results in only moderate $\chi^2$ decrease, $\mathcal{O}(1)$

## Best fit choice?

If the functional dependence is not predicted by any theory/model, we can try to fit a polynomial or function series. When should we stop?

How can we recognize that we have too many parameters?
Beside looking at the $P$ value corresponding to the obtained $\chi^2$

- Adding new parameter results in only moderate $\chi^2$ decrease, $\mathcal{O}(1)$
- Parameters become highly correlated

## Best fit choice?

If the functional dependence is not predicted by any theory/model, we can try to fit a polynomial or function series. When should we stop?

How can we recognize that we have too many parameters?
Beside looking at the $P$ value corresponding to the obtained $\chi^2$

- Adding new parameter results in only moderate $\chi^2$ decrease, $\mathcal{O}(1)$
- Parameters become highly correlated
- Values and errors of the individual parameters increase
  differences of large contributions

## Best fit choice?

If the functional dependence is not predicted by any theory/model, we can try to fit a polynomial or function series. When should we stop?

How can we recognize that we have too many parameters?
Beside looking at the $P$ value corresponding to the obtained $\chi^2$

- Adding new parameter results in only moderate $\chi^2$ decrease, $\mathcal{O}(1)$
- Parameters become highly correlated
- Values and errors of the individual parameters increase
  differences of large contributions
- Additional parameters are consistent with zero

## Best fit choice?

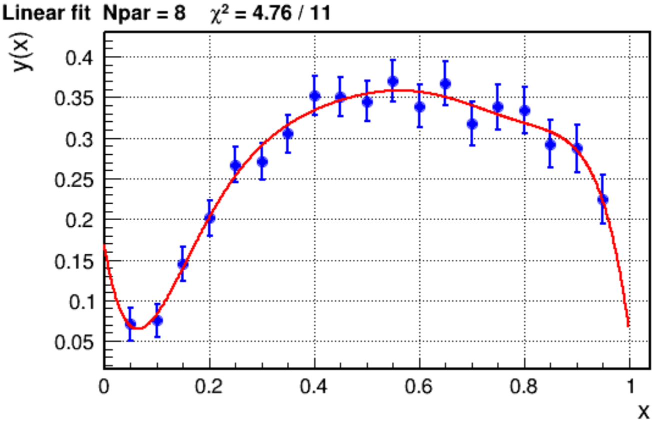
If the functional dependence is not predicted by any theory/model, we can try to fit a polynomial or function series. When should we stop?

How can we recognize that we have too many parameters?
Beside looking at the $P$ value corresponding to the obtained $\chi^2$

- Adding new parameter results in only moderate $\chi^2$ decrease, $\mathcal{O}(1)$
- Parameters become highly correlated
- Values and errors of the individual parameters increase
  differences of large contributions
- Additional parameters are consistent with zero
- Fit starts to follow fluctuations of the measurement results

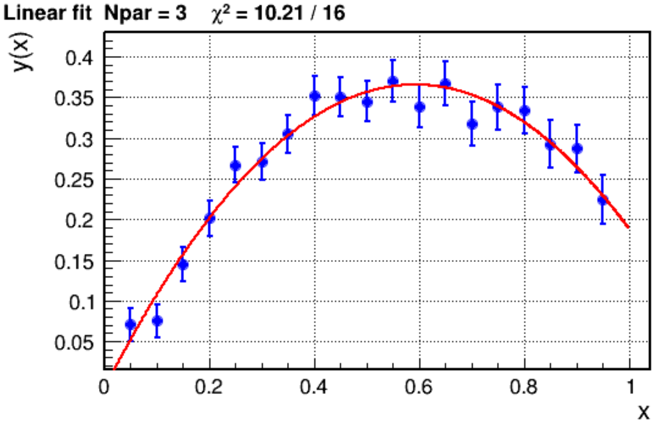
## Best fit choice?

Example of fit with too high polynomial order



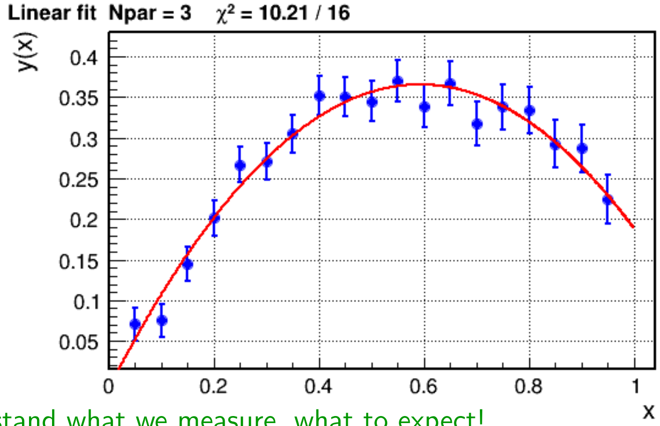
## Best fit choice?

Example of fit with proper polynomial order



## Best fit choice?

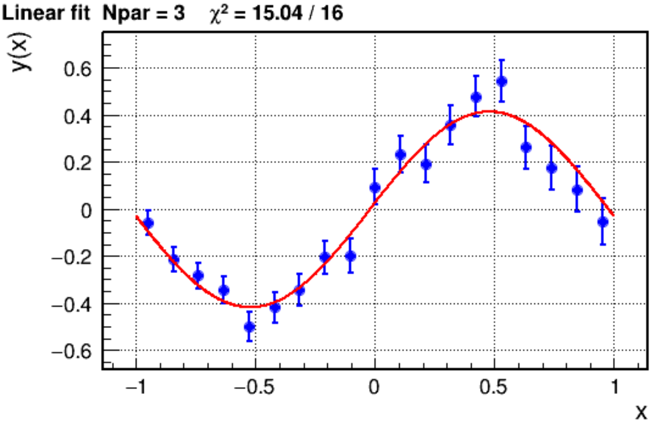
Example of fit with




We need to understand what we measure, what to expect!

## Learning on errors

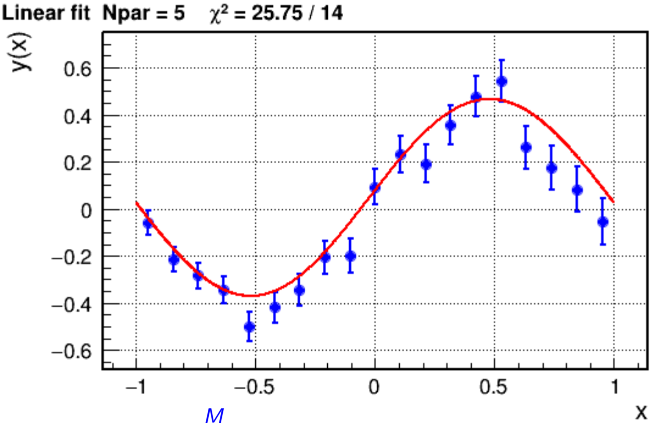
When "wrong" set of functions (highly correlated) is selected...




$$y(x) = a_0 + a_1 \sin(x) + a_2 \cos(x)$$

## Learning on errors

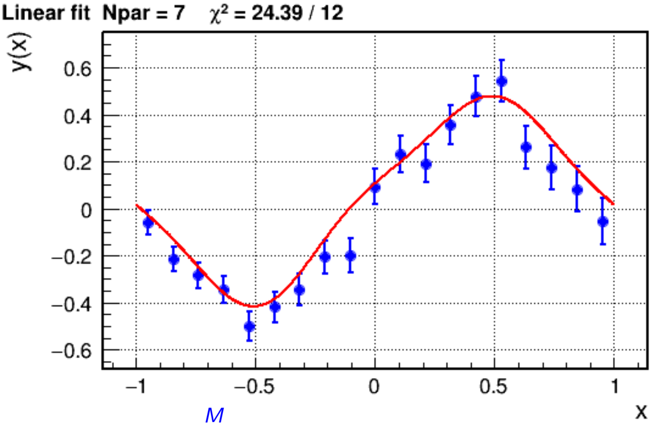
When "wrong" set of functions (highly correlated) is selected...




$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin^n(x) + a_{2n} \cos^n(x) \qquad M = 2$$

## Learning on errors

When "wrong" set of functions (highly correlated) is selected...

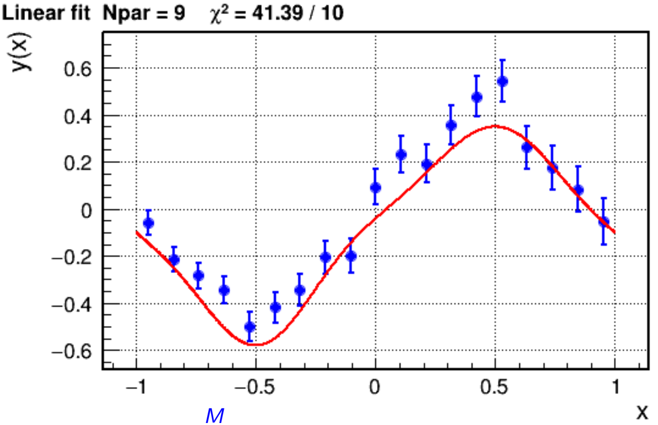



$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin^n(x) + a_{2n} \cos^n(x) \qquad M = 3$$

## Learning on errors

When "wrong" set of functions (highly correlated) is selected...

Poor numerical precision due to high correlations between parameters.



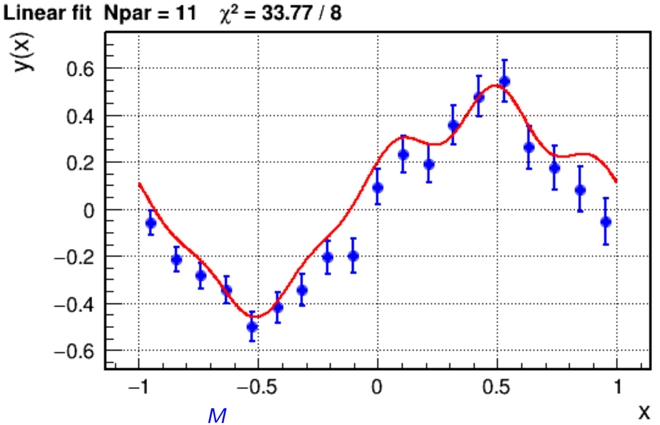

$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1} \sin^n(x) + a_{2n} \cos^n(x) \qquad M = 4$$

## Learning on errors

When "wrong" set of functions (highly correlated) is selected...

Poor numerical precision due to high correlations between parameters.

$\sin^2(x) + \cos^2(x) = 1$




$$y(x) = a_0 + \sum_{n=1}^{M} a_{2n-1}\sin^n(x) + a_{2n}\cos^n(x) \qquad M = 5$$

## Polynomial fit example

For clarity of notation, it is convenient to change parameter numbering to $k, l = 0 \ldots M$ (for polynomial fit of order $M$, $M+1$ parameters).

$$\mathbb{A}_{l\,k} = \sum_{i=1}^{N} \frac{x_i^{(l+k)}}{\sigma_i^2} \quad \text{and} \quad b_l = \sum_{i=1}^{N} \frac{x_i^l \; y_i}{\sigma_i^2}$$

## Polynomial fit example

For clarity of notation, it is convenient to change parameter numbering to $k,\ l = 0 \ldots M$ (for polynomial fit of order $M$, $M+1$ parameters).

$$\mathbb{A}_{l\,k} \; = \; \sum_{i=1}^{N} \frac{x_i^{(l+k)}}{\sigma_i^2} \quad \text{and} \quad b_l \; = \; \sum_{i=1}^{N} \frac{x_i^l \; y_i}{\sigma_i^2}$$

For uniform uncertainties it is then:

$$\mathbb{A} \; = \; \frac{1}{\sigma^2} \sum_{i=1}^{N} \begin{pmatrix} 1 & x_i & \ldots & x_i^M \\ x_i & x_i^2 & \ldots & x_i^{M+1} \\ \vdots & & & \vdots \\ x_i^M & x_i^{M+1} & \ldots & x_i^{2M} \end{pmatrix} \qquad \mathbf{b} \; = \; \frac{1}{\sigma^2} \sum_{i=1}^{N} \begin{pmatrix} y_i \\ y_i \; x_i \\ \vdots \\ y_i \; x_i^M \end{pmatrix}$$

quite simple to implement...

## Uncertainty estimate

The $\chi^2$ value at the minimum can be then calculated as:

$$\tilde{\chi}^2 \;=\; (\mathbf{y} - \mu(\mathbf{x};\mathbf{a}))^{\mathsf{T}} \,\mathbb{A}\, (\mathbf{y} - \mu(\mathbf{x};\mathbf{a}))$$

Its distribution should correspond to the $\chi^2$ distribution for $N_{df} = N - M$

## Uncertainty estimate

The $\chi^2$ value at the minimum can be then calculated as:

$$\tilde{\chi}^2 = (\mathbf{y} - \mu(\mathbf{x};\mathbf{a}))^\mathsf{T} \, \mathbb{A} \, (\mathbf{y} - \mu(\mathbf{x};\mathbf{a}))$$

Its distribution should correspond to the $\chi^2$ distribution for $N_{df} = N - M$

If $\sigma$ is the same for all measurements, the extracted parameter values do not dependent on it!

## Uncertainty estimate

The $\chi^2$ value at the minimum can be then calculated as:

$$\tilde{\chi}^2 = (\mathbf{y} - \mu(\mathbf{x};\mathbf{a}))^\intercal \, \mathbb{A} \, (\mathbf{y} - \mu(\mathbf{x};\mathbf{a}))$$

Its distribution should correspond to the $\chi^2$ distribution for $N_{df} = N - M$

If $\sigma$ is the same for all measurements, the extracted parameter values do not dependent on it!
We can use the calculated value of $\tilde{\chi}^2$ to validate the model (test model hypothesis), but also to "correct" our uncertainties, if we consider them unreliable (or they are unknown).
Resulting variance estimate:

$$\tilde{\sigma}^2 = \sigma^2 \cdot \frac{\tilde{\chi}^2}{N - M}$$

This is useful in particular when $\tilde{\chi}^2 \ll N_{df}$ (overestimated $\sigma$)
For $\tilde{\chi}^2 \gg N_{df}$ we need to consider the possibility that our model is wrong...

## Multiple independent variables

The described approach works also for multi-dimensional dependencies!

For example, we can consider polynomial of order $M$ in two coordinates:

$$\mu(x, z; \mathbf{a}) = \sum_{k=0}^{M} \sum_{l=0}^{M} a_{kl} \, x^k \, z^l$$

## Multiple independent variables

The described approach works also for multi-dimensional dependencies!
For example, we can consider polynomial of order $M$ in two coordinates:

$$\mu(x, z; \mathbf{a}) \; = \; \sum_{k=0}^{M} \sum_{l=0}^{M} a_{kl} \, x^k \, z^l$$

All we need to do is to order the pairs of indexes, so that vector **a** is properly defined.
Example for $M = 1$ (2-D plane): $\mathbf{a} = (a_{00}, a_{10}, a_{01})$

$$\mathbb{A} \; = \; \frac{1}{\sigma^2} \sum_{i=1}^{N} \begin{pmatrix} 1 & x & z \\ x & x^2 & x\,z \\ z & x\,z & z^2 \end{pmatrix} \qquad \mathbf{b} \; = \; \frac{1}{\sigma^2} \sum_{i=1}^{N} \begin{pmatrix} y \\ y\,x \\ y\,z \end{pmatrix}$$

where measurement indexes $i = 1 \ldots N$ were skipped for variables $x_i$, $y_i$ and $z_i$

## Least-squares method

## Homework

Solutions to be uploaded by December 4.

Download the set of data from the lecture home page.
Text file with three columns: $x_i$, $y_i$, $\sigma_{y_i}$

Use linear regression method to fit polynomial dependence to the data.

Calculate p-value for the 3rd order polynomial fit.

Find the order of polynomial, which is adequate for the description of the data and give arguments for your choice.