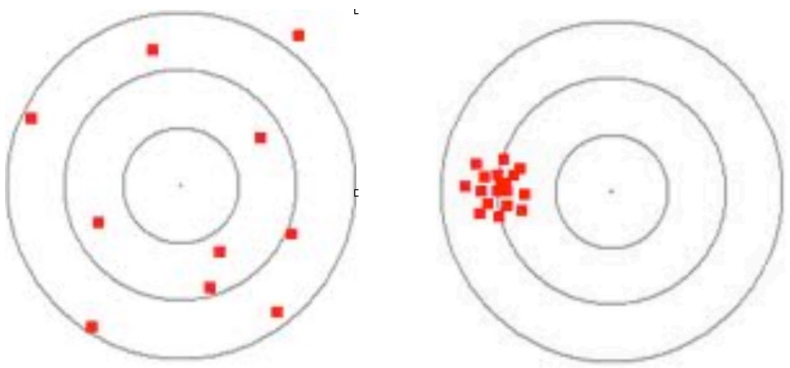
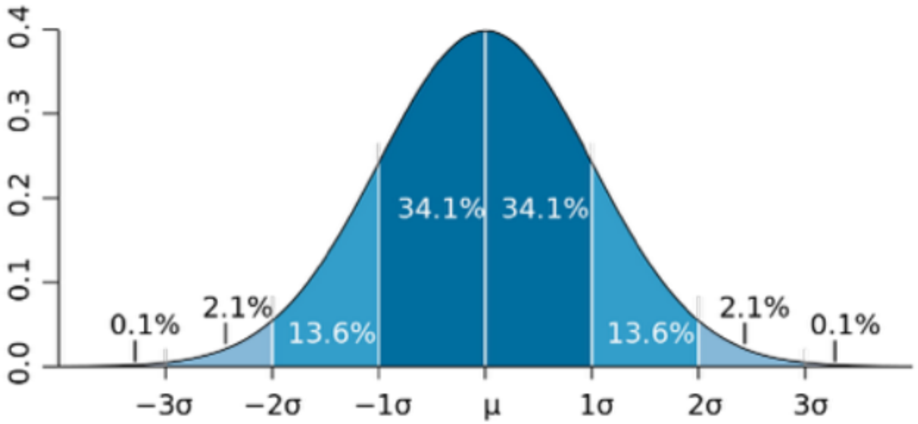
Statistical Methods in Data Analysis

Part 1: Parameter Estimation

Andreas B. Meyer
DESY

18–22 March 2024





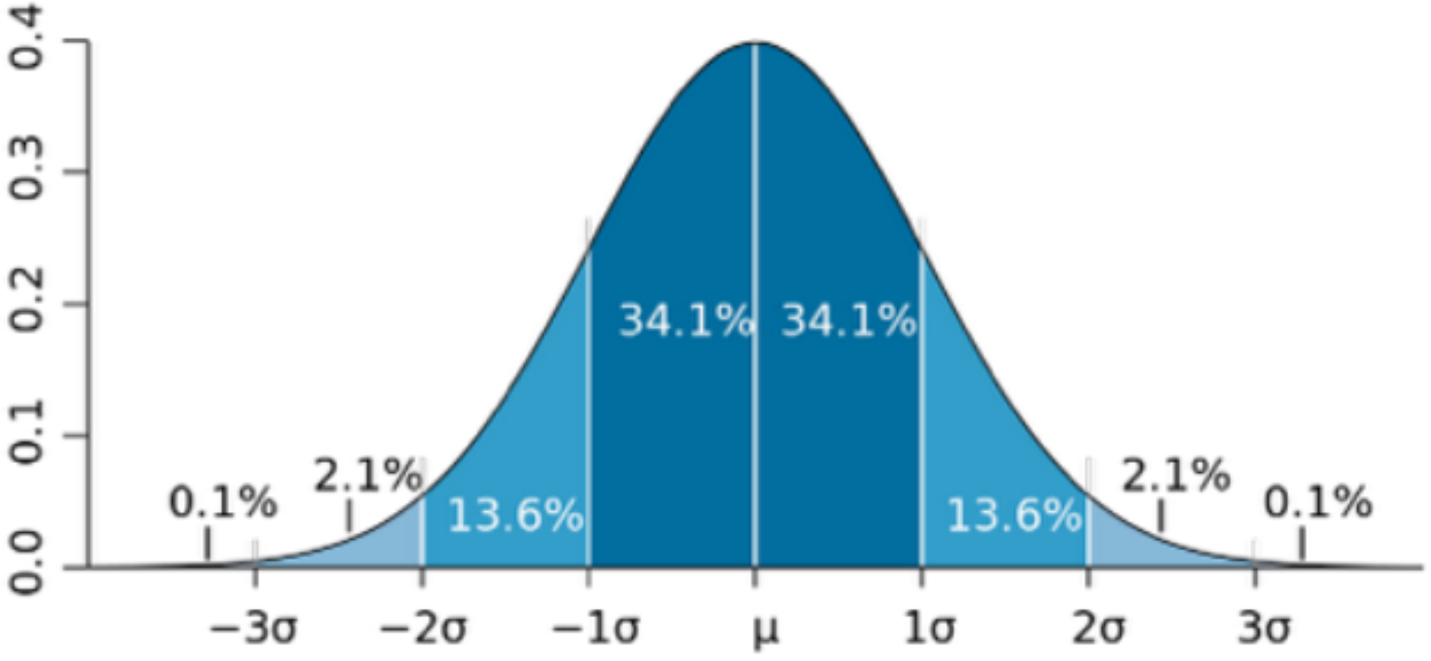
Menu

Parameter Estimation

Today
- Statistical and systematic uncertainties
- Probability
- Parameter estimation

Wednesday
- Hypothesis testing
- Confidence intervals
- Profile likelihood ratio
- Outlook: classification and MVA

Friday
Matthias Komm:
Introduction to machine learning



Quantiles of the Gauss distribution and what they mean

Menu

Confidence Intervals

Tuesday

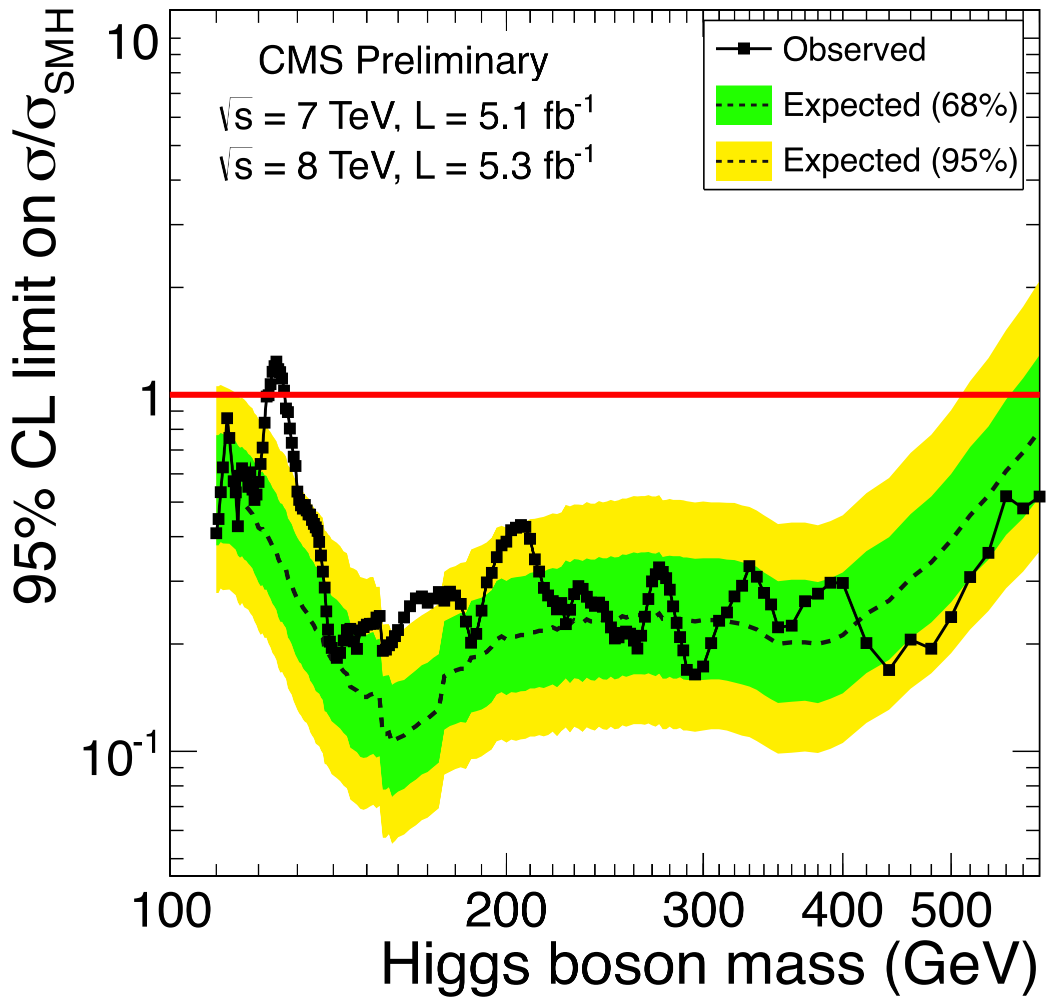
- Statistical and systematic uncertainties
- Probability
- Parameter estimation

Wednesday

- Hypothesis testing
- Confidence intervals
- Profile likelihood ratio
- Outlook: classification and MVA

Friday

Matthias Komm:
Introduction to machine learning



Higgs discovery: What does this figure really show ?

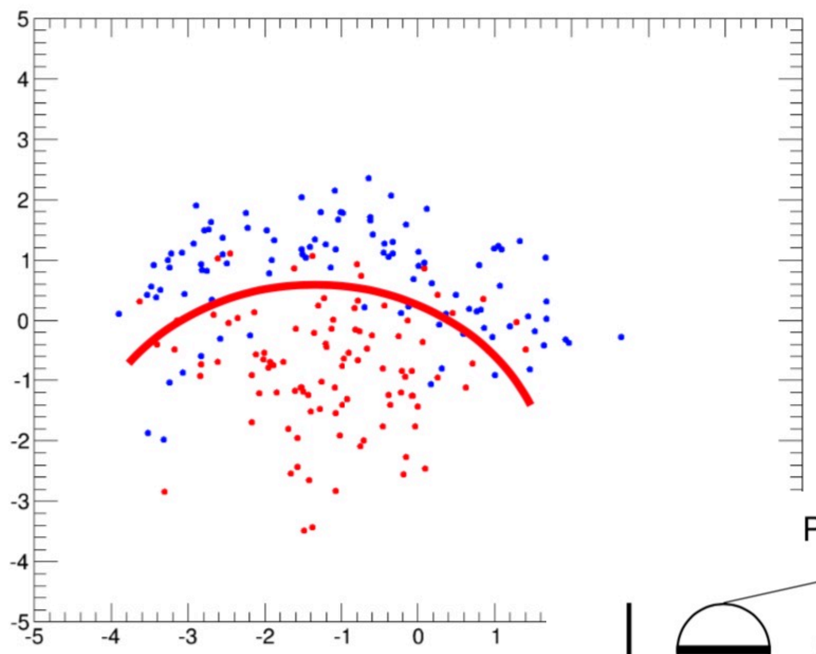
Menu

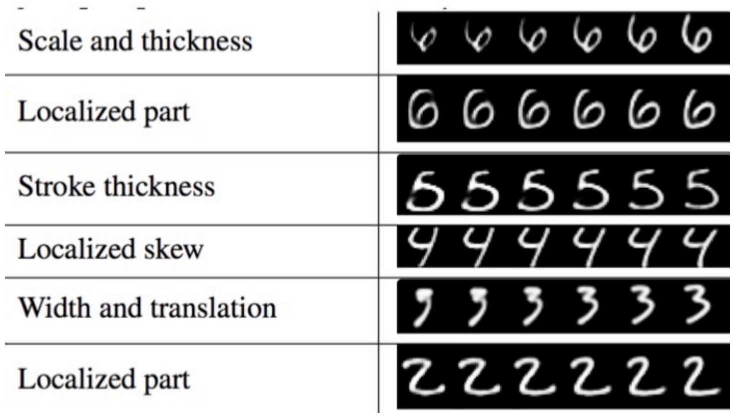
Multivariate Analysis

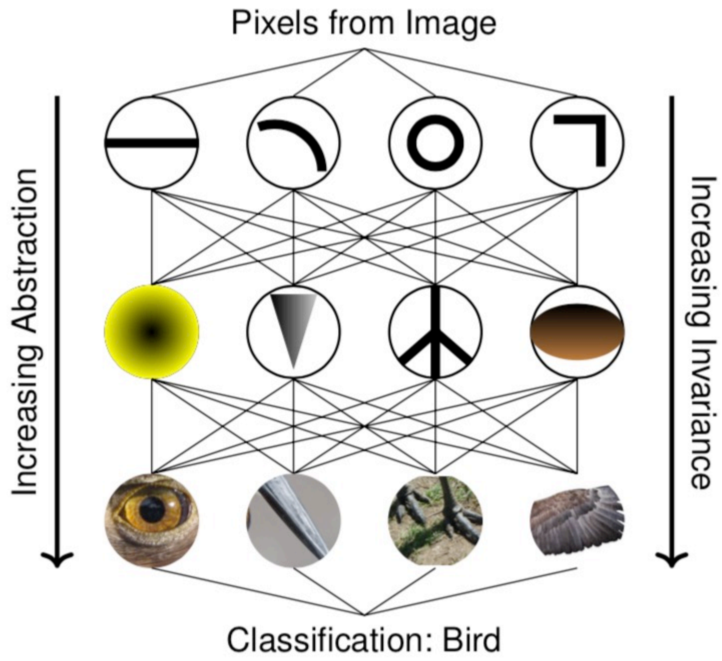
Tuesday
- Statistical and systematic uncertainties
- Probability
- Parameter estimation

Wednesday
- Hypothesis testing
- Confidence intervals
- Profile likelihood ratio
- Outlook: classification and MVA

Friday
Matthias Komm:
Introduction to machine learning



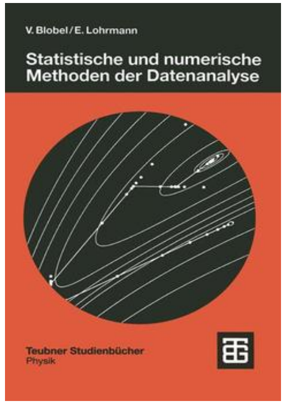


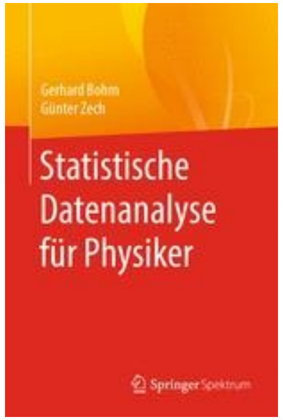


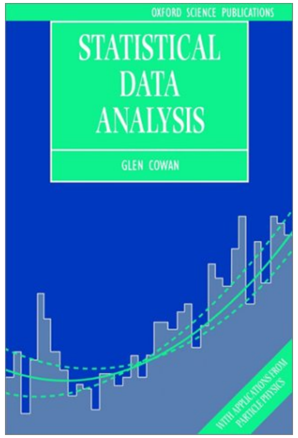
Classification, Multivariate Analysis, Machine-Learning

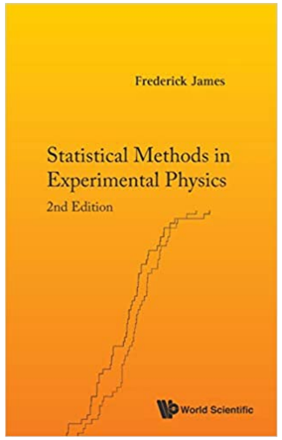
Books and References

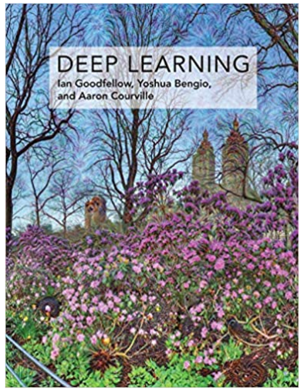
- V. Blobel, E.Lohrmann, **2012**: “Statistische und numerische Methoden der Datenanalyse”, https://www.desy.de/~sschmitt/blobel/ebuch.html
- M. Erdmann, T. Hebbeker, **2013**: Exp.Phys.V: Mod. Meth. der Datenanalyse
- G. Bohm, G.Zech, **2017**: “Introduction to Statistics and Data Analysis for Physicists”, [DESY library]
- G. Cowan, **1998**: “Statistical Data Analysis”
- F. James, 2nd edition, **2006**: “Statistical Methods in Experimental Physics”
- O. Behnke et al, **2013**: “Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods”
- Goodfellow et al, **2016**: “Deep Learning” https://www.deeplearningbook.org/

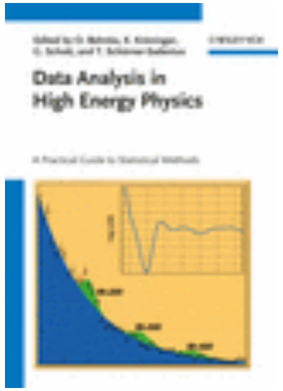












Links, Papers and Sources

Statistical Methods in Data Analysis", Terascale, March 2024: https://www.desy.de/~ameyer/da_desy24

Previous lectures:

- Statistical Methods in Data Analysis", Introduction to the Terascale, March 2023: https://indico.desy.de/event/33888 and https://www.desy.de/~ameyer/da_desy23/
- Statistical Methods in Data Analysis", KSETA lecture, Feb 2022: https://www.desy.de/~ameyer/da_kseta_22/
- "Moderne Methoden der Datenanalyse", Course lecture at KIT, SoSe 2017, slides (in German): https://www.desy.de/~ameyer/kit/da_sose17/index.html **Access to slides and material: (user: Students. pw: only)**

Papers and articles:

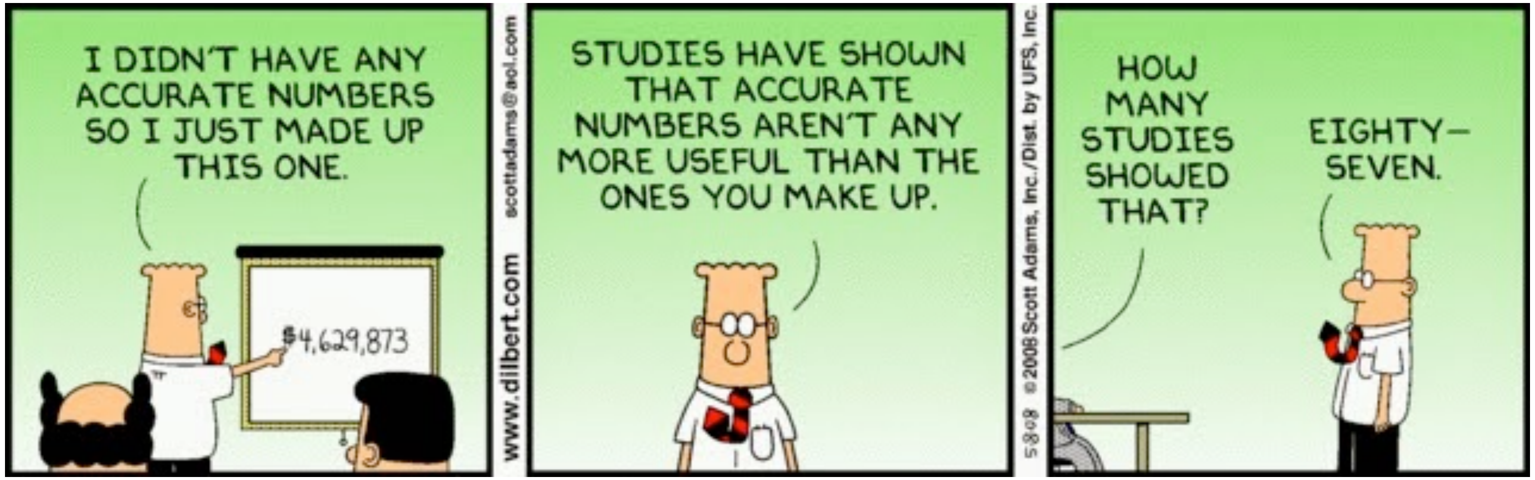
- Robert Cousins: "Why isn't every physicist a Bayesian ?", Am.J.Phys. 65 (1995).
- Robert Cousins: "Lectures on Statistics in Theory: Prelude to Statistics in Practice" [arXiv]
- G.Cowan, Particle Data Group [pdg] 2020, chapter 40 [pdf] or full PDG book for download (80MB) [pdf]
- G.Cowan, K.Cranmer, E.Gross, O.Vitells: "Asymptotic formulae for likelihood-based tests of new physics" [arXiv]
- ATLAS and CMS Collaborations: "Procedure for the LHC Higgs boson search combination" [CDS]
- T.Junk: "Confidence level computation for combining searches with small statistics", NIM, A 434 (1999) 435-443
- A.Read: "Presentation of search results: the CL_s technique", J.Phys.G: 28 (2002)

Many thanks for discussions, material and help go to:

- G. Quast (KIT), R. Wolf (KIT), O. Behnke (DESY), C. Autermann (Aachen)

Introduction

Motivation



Use optimal statistical methods to extract maximal information from the data

- Parameter estimation:
  - What are the most likely "true" values, given the data ?
    n.b. this is a very "Bayesian" question
- Hypothesis testing:
  - Should an email go to the spam folder?
  - Does a person carry a contagious disease?
  - Are the data consistent with the prediction ?
- Confidence intervals:
  - Uncertainty and significance

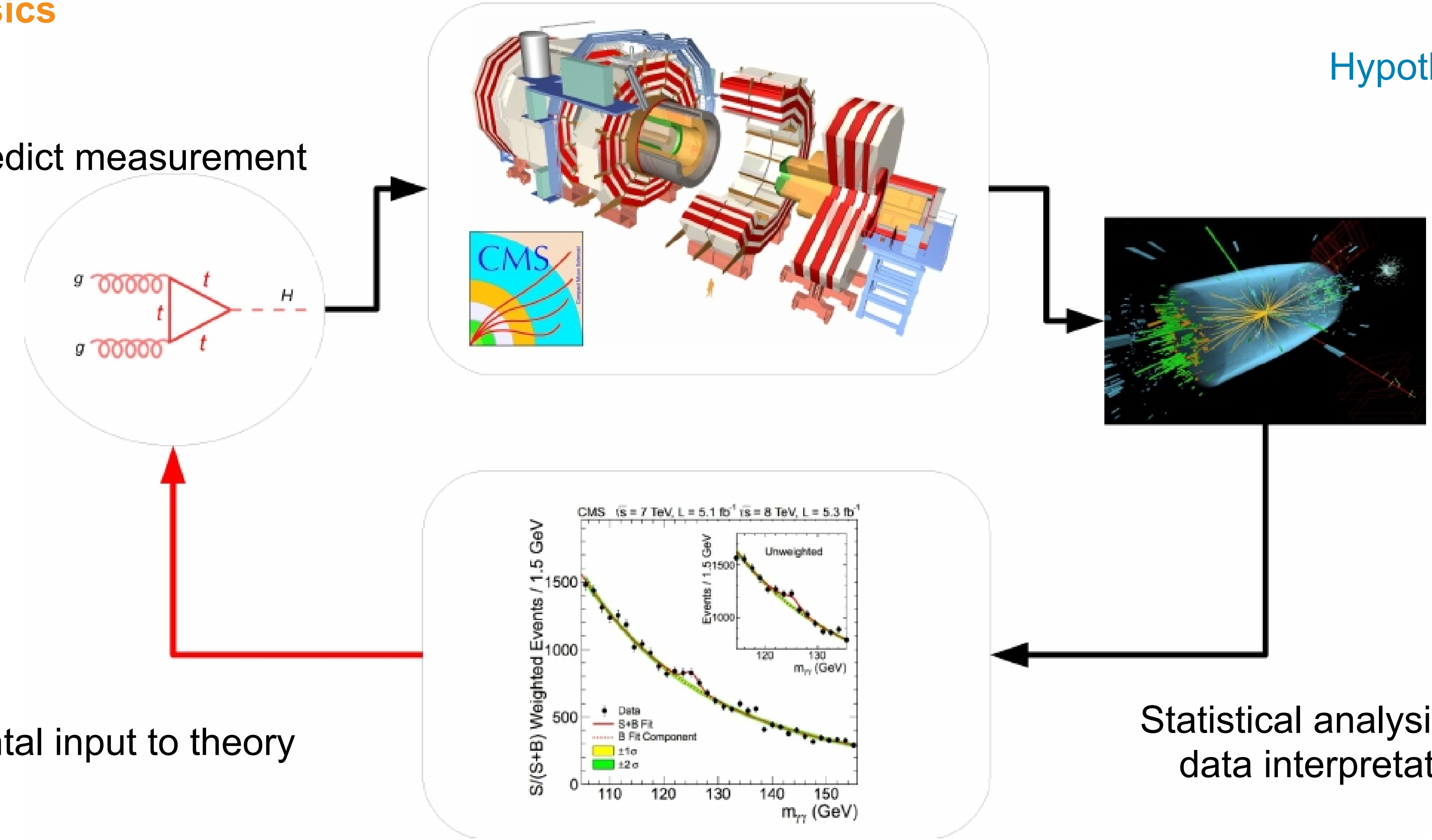
The scientific cycle

Particle physics

Experiment: measure and test theory predictions

Hypothesis tests

Theory: predict measurement



Experimental input to theory

Statistical analysis and data interpretation

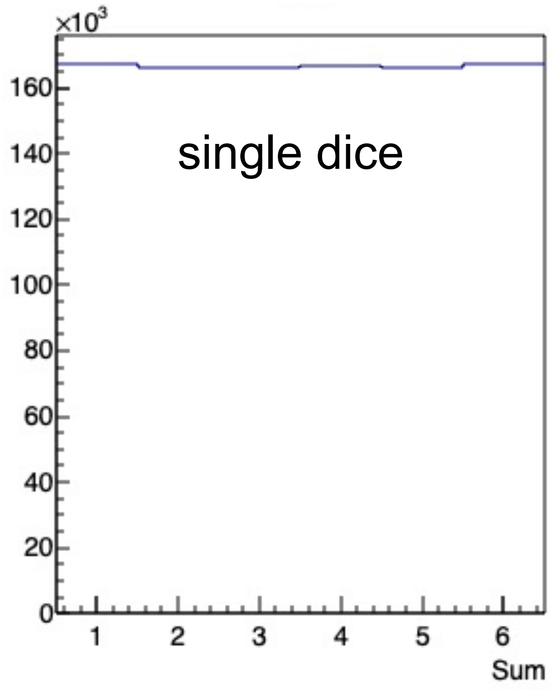
Parameter estimation

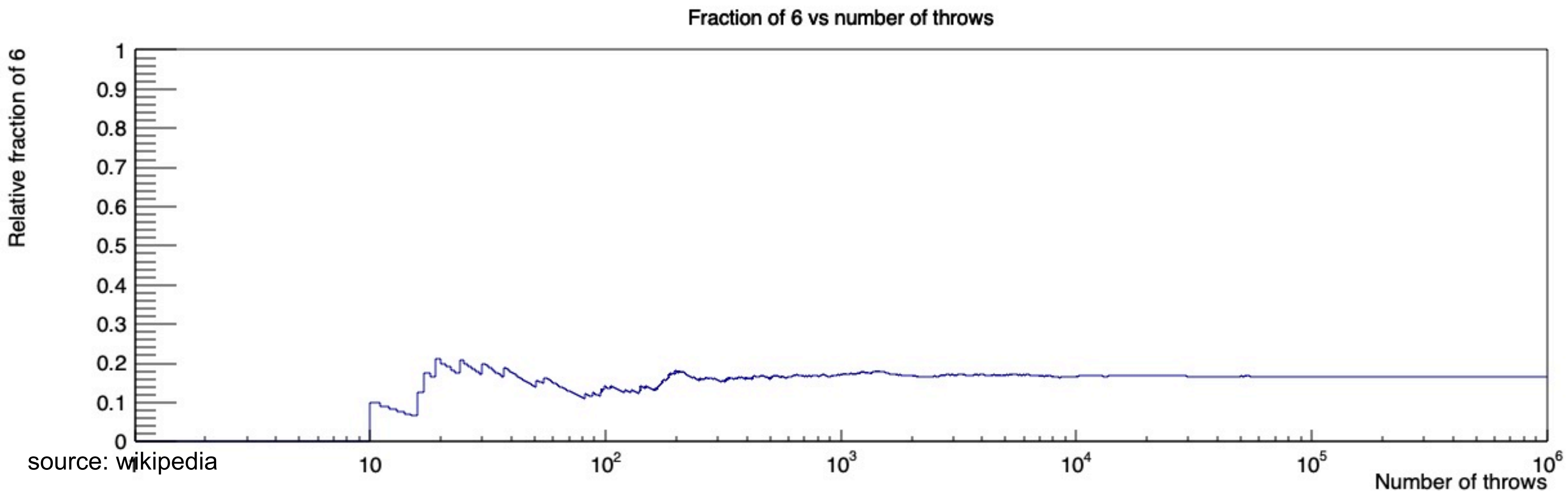
Confidence intervals

The scientific cycle

Example: dice

- Prediction (trivial for ideally cubic dice):
  - Same probability for each number: 1/6 (expectation value)
- Experiment
  - Variance (spread) → statististical uncertainty
  - Bias (distortion) → systematic uncertainty (unless corrected)
- Simulation requires good random number generators !





source: wikipedia

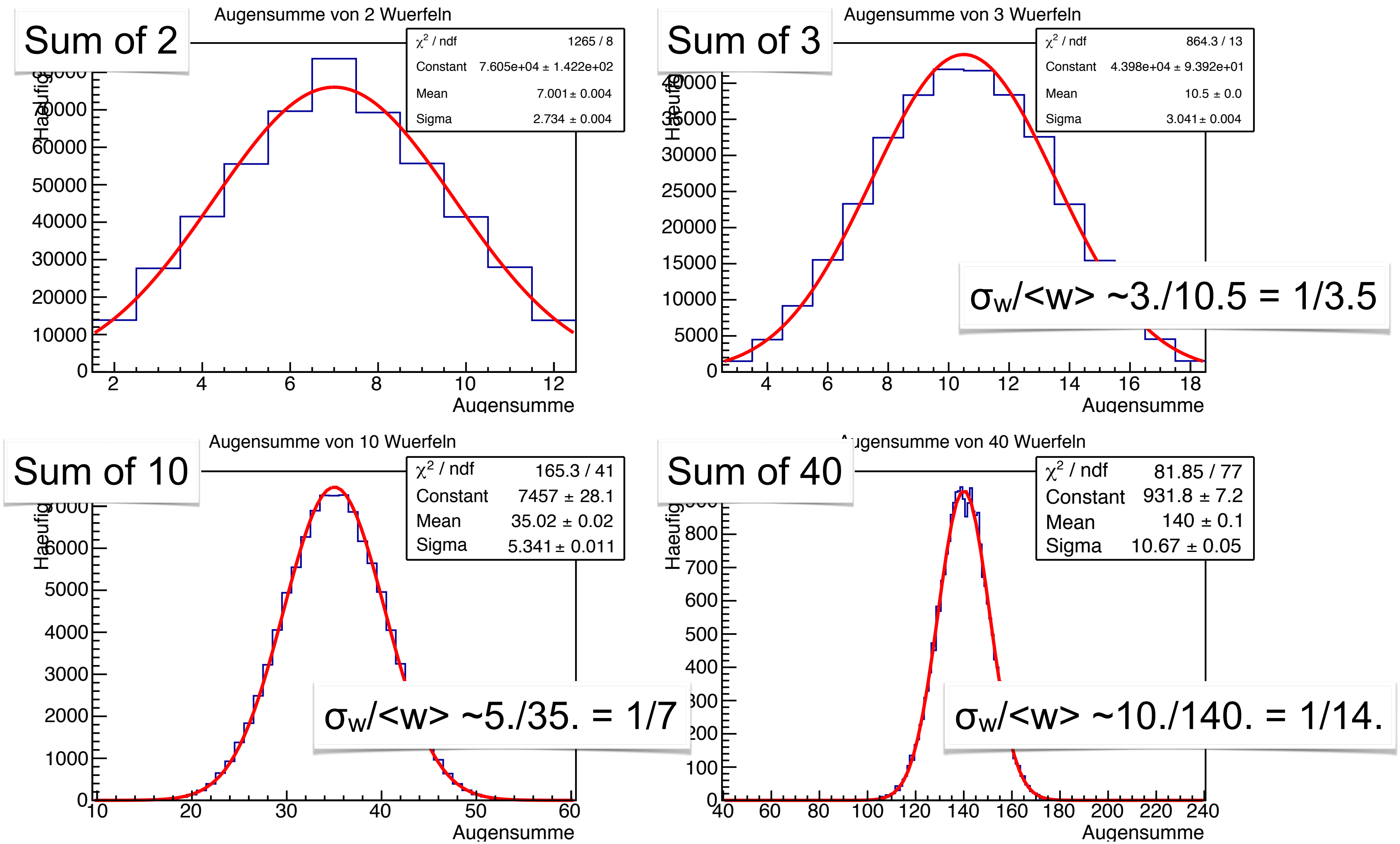
Example: dice.C

Central Limit Theorem

Example: dice

In the limit of large *n*:
the probability distribution of the sum of *n* independent random numbers follows a Gaussian distribution.

N.B.: the sum is a Gaussian distribution even if the original variables are not Gaussian distributed



source: wikipedia

Relative uncertainty of the mean decreases as $1/\sqrt{n}$

Central Limit Theorem

- Be w the sum of n random numbers x_i (with variance σ_i^2)

$$w = \sum_{i=1}^{n} x_i$$

- In the limit of large n, the probability density function (PDF), follows a Gaussian distribution

$$f(w) = \frac{1}{\sqrt{2\pi}\sigma_w} \cdot e^{-\frac{(w-\langle w \rangle)^2}{2\sigma_w^2}}$$

  - with variance V:

$$V[w] = \sigma_w^2 = \sum_{i=1}^{n} \sigma_i^2$$

- In case, $\sigma_i = \sigma$ for all i, then
  - the mean
  - and the standard deviation (see previous page)

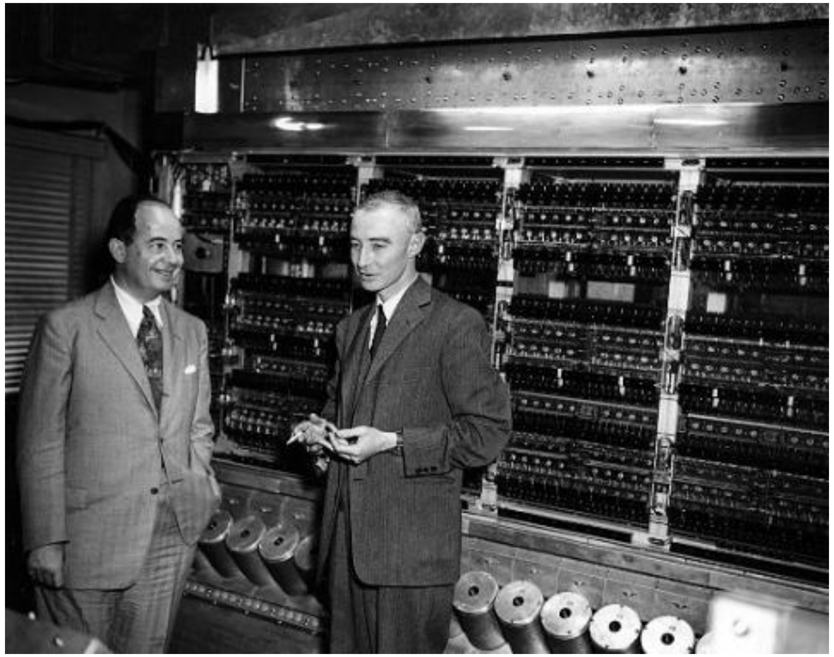
$$\langle w \rangle = n \langle x \rangle$$
$$\sigma_w = \sqrt{n}\,\sigma$$

Erdmann/Hebbeker, p.53

The Monte-Carlo method

Numeric method to solve multi-dimensional integrals

http://lib-www.lanl.gov/la-pubs/00326866.pdf



Stanislaw Ulam, John von Neumann, Manhattan project

- Simple formulation of problems
- Precision only depends on CPU time
  - Statistical uncertainty is proportional to: $1/\sqrt{n}$
  - Other numeric integration methods, e.g. trapezoid integration in *n* intervals and *d* dimensions: $1/n^{2/d}$

MC better for *d* > 4

- In particular physics:
  - Simulation of physics events (generators) and detectors (simulation by Geant)
  - Design and plan experiments, develop analyses, taking into account all resolution and efficiency effects

Convolution (“folding”) of all detector effects

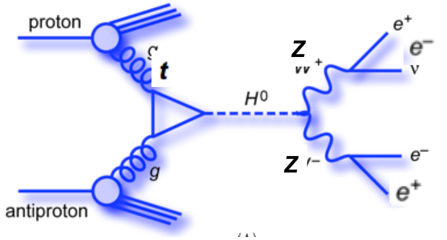
$$f'(x') = \int_{-\infty}^{\infty} t(x, x') f(x) dx$$

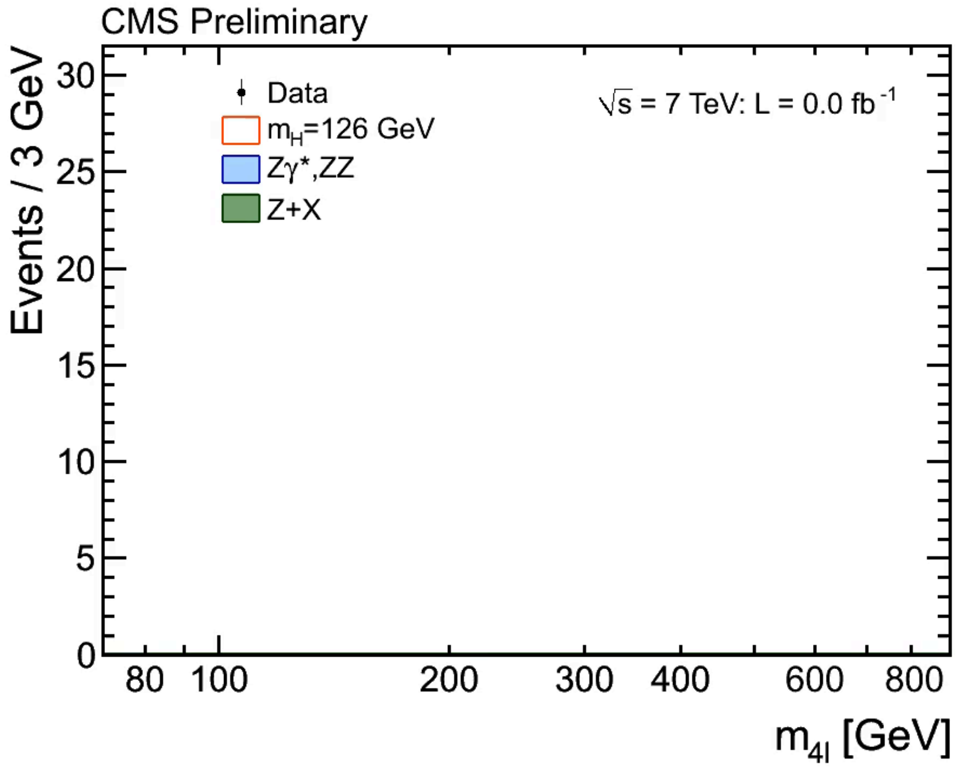
MC method: reformulate series of convolutions by sum of random distributions

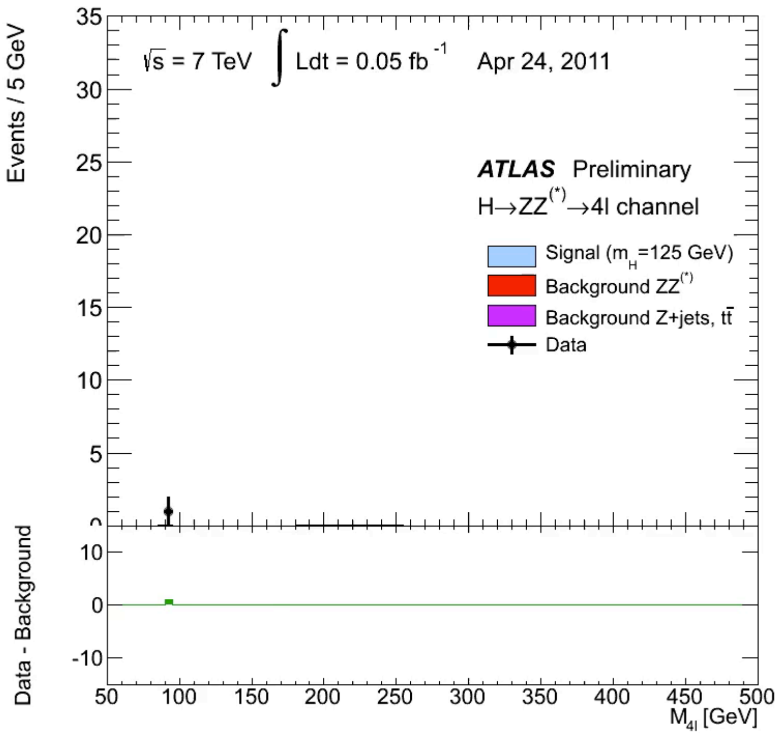
Examples: calc_pi.C animate_pi.py

Higgs boson discovery at the LHC (2012)

Two independent experiments ATLAS and CMS





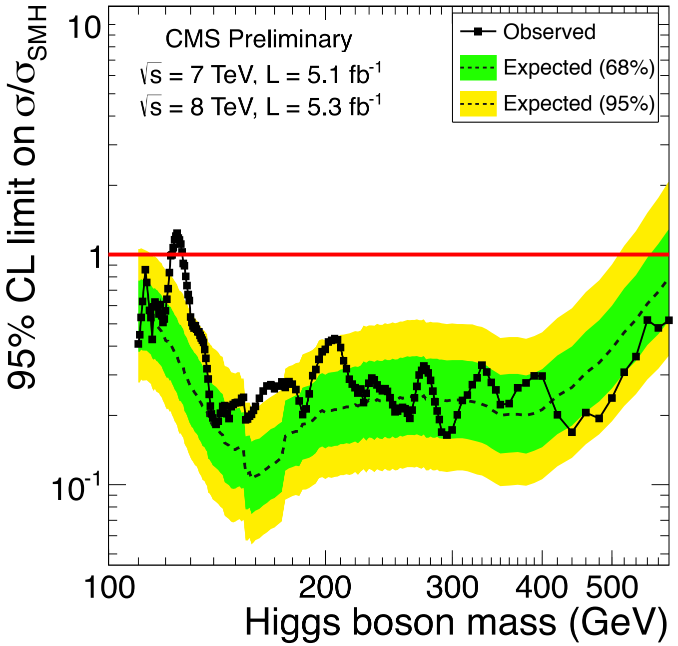


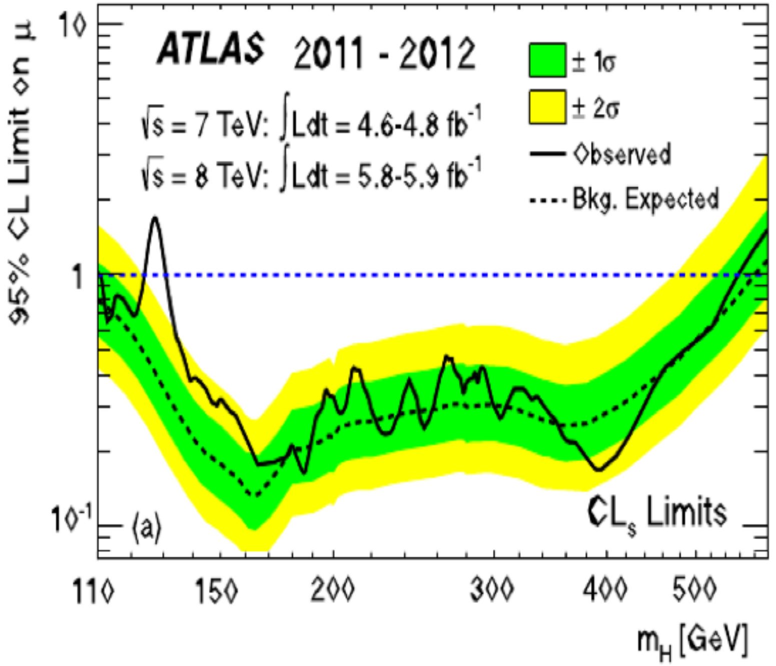
- Statistical distribution of randomly distributed events
  - counting experiments generally follow Poisson distributions
  - gradual appearance of the signal above background

Exclusion limits

“Brazilian-flag” figures

- 4th of July 2012: announcement of the discovery of a new boson
- Exclusion of signals between 131(128) GeV and 523(600) GeV





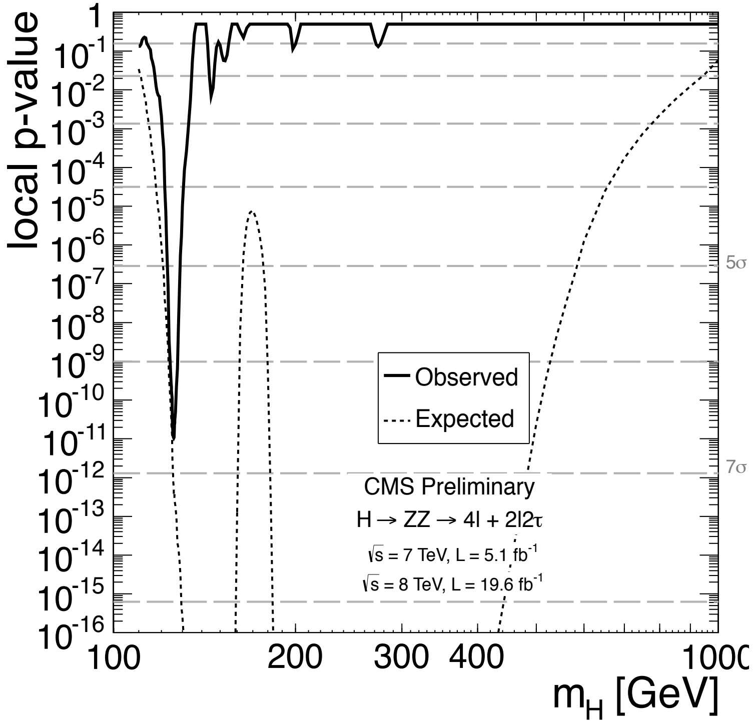
By the end of this lecture: understand what exactly is shown in these figures

Local p-value

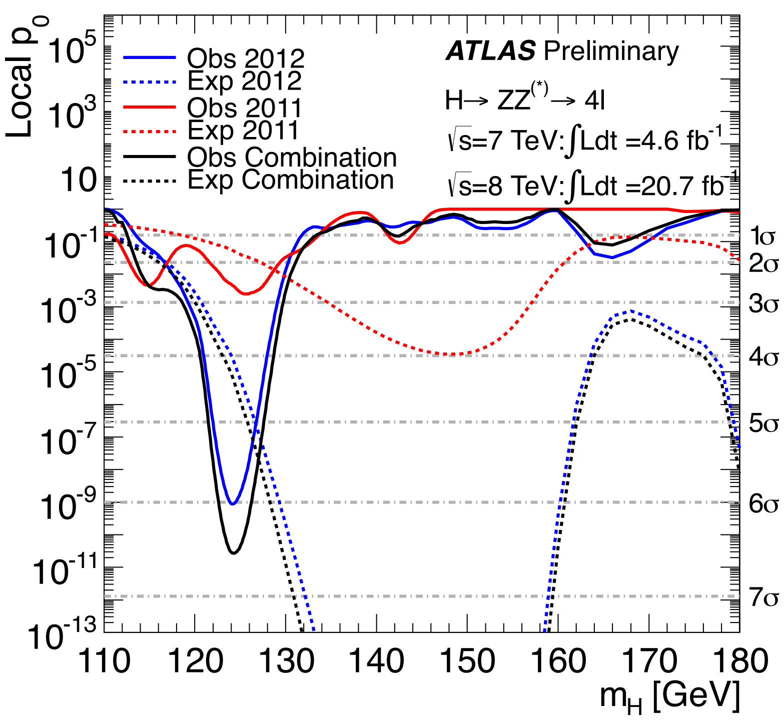
Comparison measurement and theory

- Probability, that the measured distribution is due to a fluctuation of the background: $p(m_H = 125.8\ \text{GeV}) = 10^{-11}$ (corresponds to a significance of ~6.7σ)

CMS



ATLAS



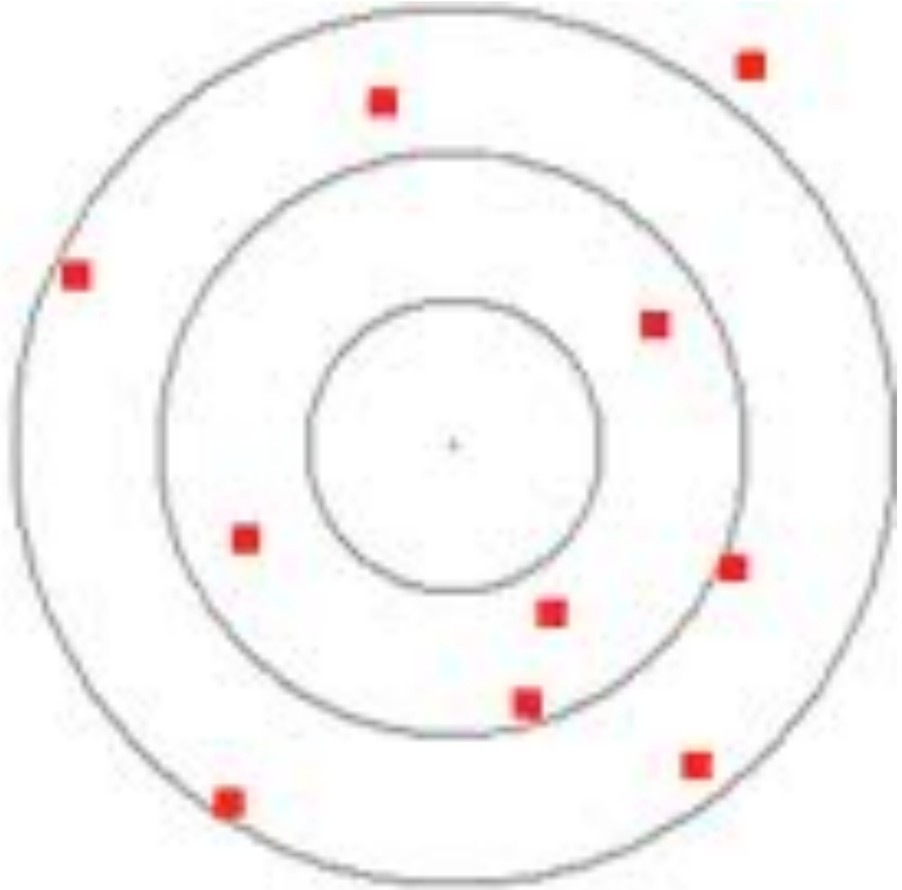
By the end of this lecture: understand what exactly is shown in these figures → Hypothesis Testing

Uncertainties

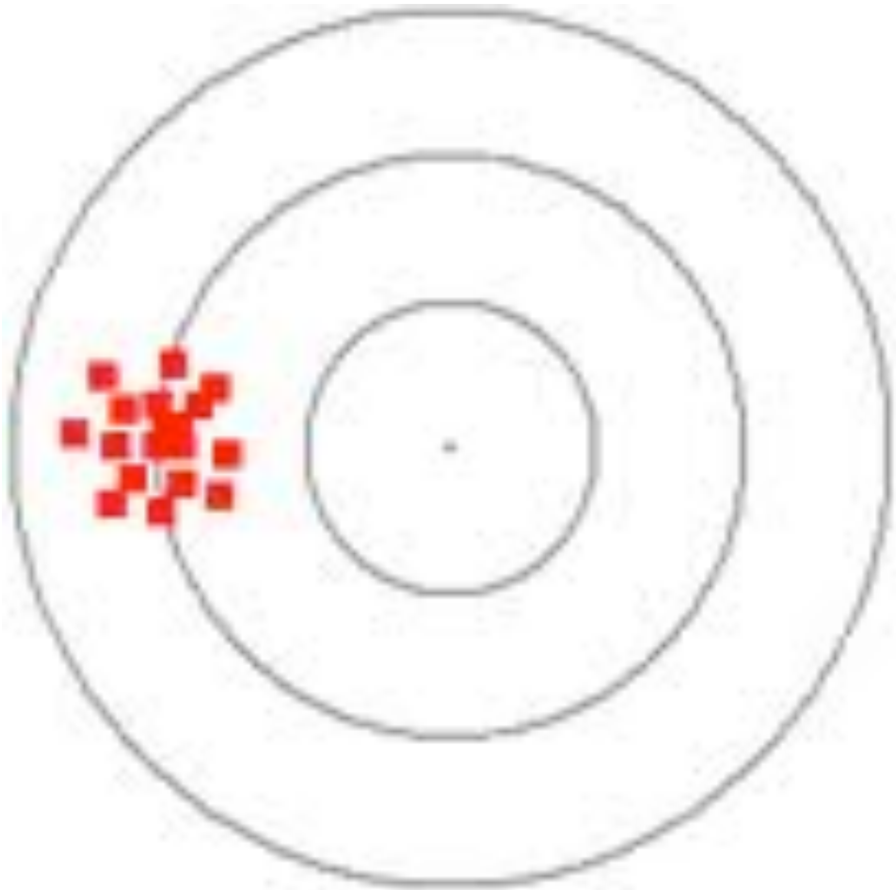
Error and uncertainty

- Strictly speaking, a <u>correct</u> measurement <u>can not have errors</u>
- Any measurement has uncertainties. In practice, the term “error” is often used
- Uncertainties can be statistical or systematic

Statistical and systematic uncertainties



Spread



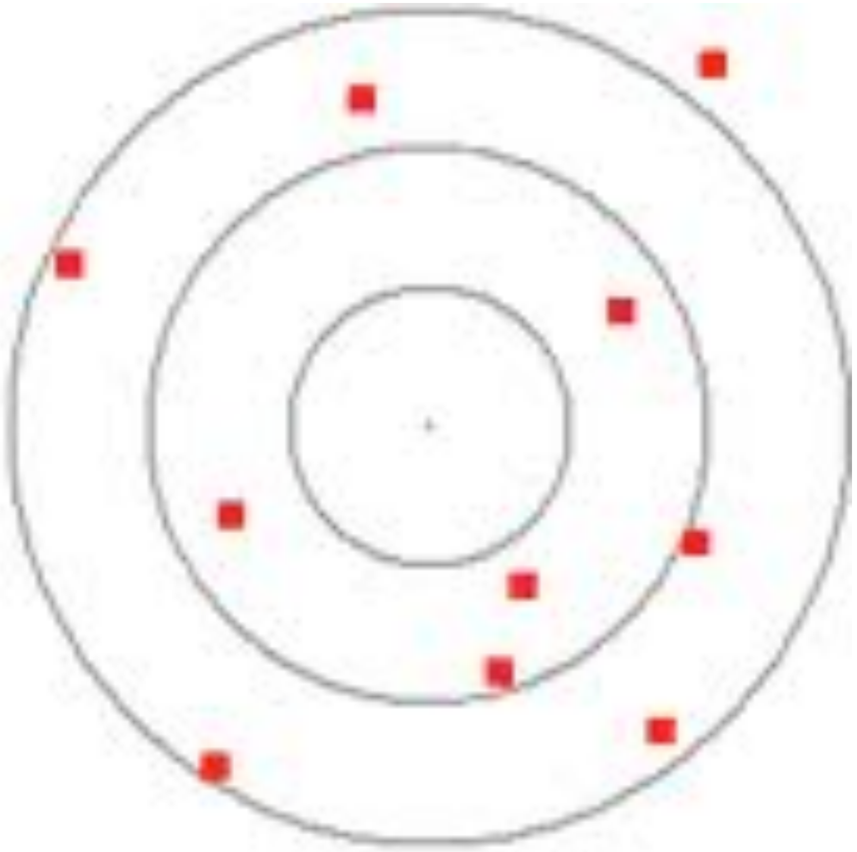
Bias

Systematic uncertainty: uncertainty in the size of the bias

Statistical uncertainties

- Spread of a single measurement for reasons that are practically (e.g. cube) and/or principally (QM) untraceable
  - => Variance: distribution around mean
- Repeated measurements are independent (uncorrelated)
- **Statistical uncertainties are theoretically well understood**
  - Error propagation
  - Correlations

For lack of time I will not say much about these aspects

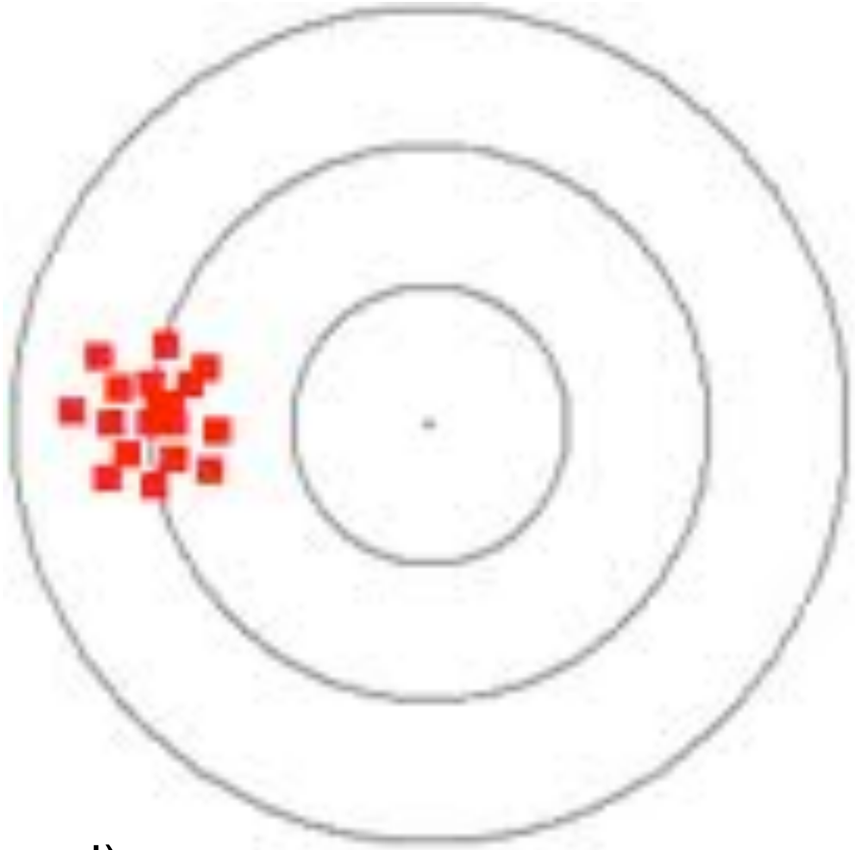


Systematic uncertainties

- Bias (distortion) of measurement

- Examples:
  - Reading errors
  - Noise
  - Miscalibration
  - Band-wagon effects,
    i.e. prejudice about result

"I already wrote the paper. That's why it's so hard to get the right data."



- Systematic uncertainties are (in principle) traceable
- Repeated measurements are usually correlated
  (unless underlying assumptions or analysis approach are changed)
- In practice, no general method for quantification

Systematic uncertainties

An attempt to categorise

- **Explicit**: purely experimental uncertainties
  - Often due to statistical uncertainties of ancillary measurements, e.g. calibrations
  - More data would help, but are often not available
- **Implicit**: due to assumptions of the measurement setup or analysis
  1. Continuous parameters, e.g. QCD scale uncertainties
  2. Discrete values only: no unambiguous interpolation between assumptions

- Systematic uncertainties also follow the central limit theorem,
  - The <u>shape</u> of a single uncertainty <u>is usually irrelevant</u> for the total uncertainty of the measurement (CLT)

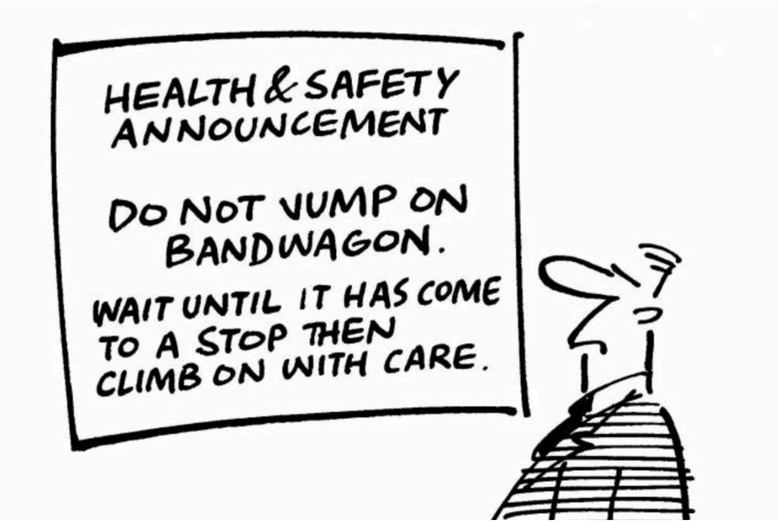
- “Unknown unknowns”: it is never excluded that some sources of uncertainty remain undetected
  - Known uncertainties are often deliberately overestimated (“conservative”)

Accurate estimate of systematics requires care and courage

“Blind analysis”

Good scientific practice

- Do not be guided by prejudice
  - Never delete the primary data, all results are to be doubted
  - Take all effects into account
  - Publish all results. About ~5% of the results should be off by more than 2σ!



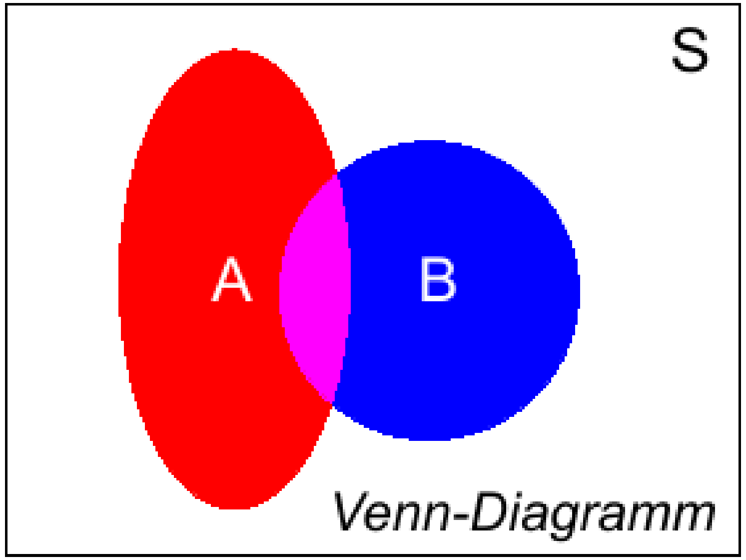
- Blinding policy
  - Produce control distributions for data and MC w/o looking at the final result.
  - Implement the actual measurement using MC “pseudo” data, and optimise the analysis
  - Document the analysis and present the result for pre-approval
  - Unblind during peer-review

- N.B.: worse than blind ⟶ unblind ⟶ publish is:
  - Blind analysis ⟶ surprise@unblinding ⟶ study surprise until surprise is gone ⟶ publish unsurprising result
  - Find middle ground, try to avoid bias, but be pragmatic

Probability

Probability

- S = full set of all events
- A, B subsets of events in S
- P(A) = probability that A takes place

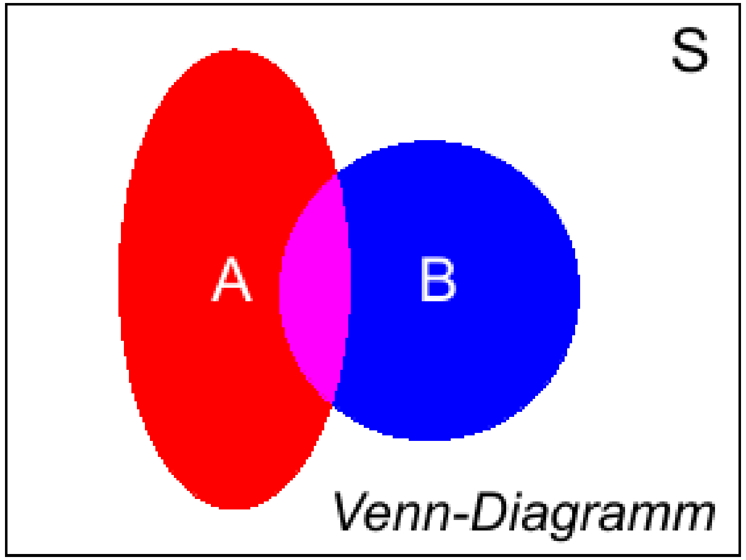


- Kolmogorov-Axioms (1931):
  - 1. $P(A) \geq 0$ positive
  - 2. If A and B disjoint, i.e. if $A \cap B = 0$ ("and") additive

    $P(A \cup B) = P(A) + P(B)$ ("or") "A or B"
  - 3. $P(S) = 1$ normalised

Cowan, p.2

Probability



- Derived properties: $$0 \leq P(A) \leq 1$$ $$P(A \cup \overline{A}) = 1$$ $$P(\overline{A}) = 1 - P(A)$$

- Moreover: “A or B”, for overlapping subsets:

$$P(A \overset{\text{“or”}}{\cup} B) = P(A) + P(B) - P(A \overset{\text{“and”}}{\cap} B)$$

- Definition of conditional probability:

$$P(A|B) = \frac{P(A \overset{\text{“and”}}{\cap} B)}{P(B)}$$

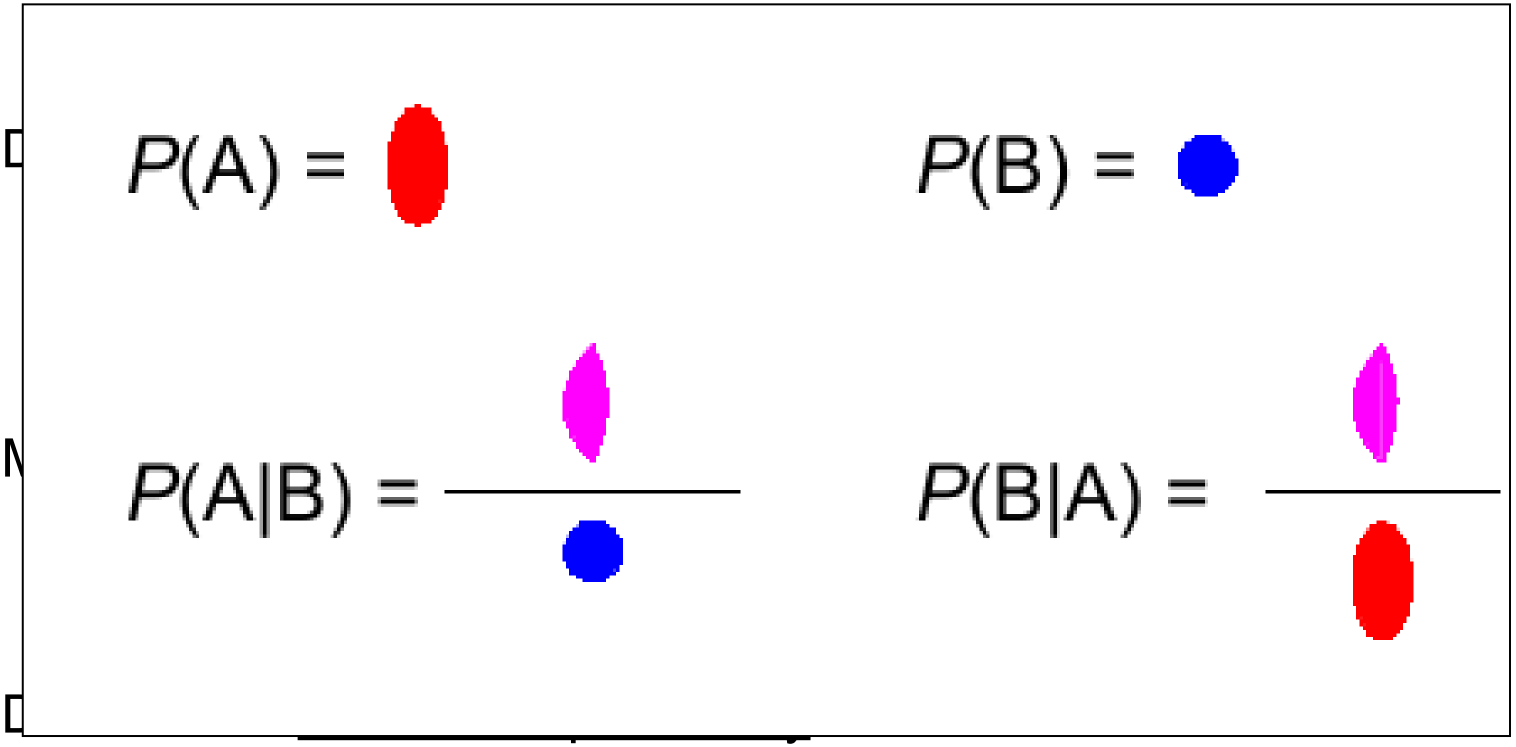
“Probability of *A* under the condition that B”
Or in short: “probability of *A* given *B*”

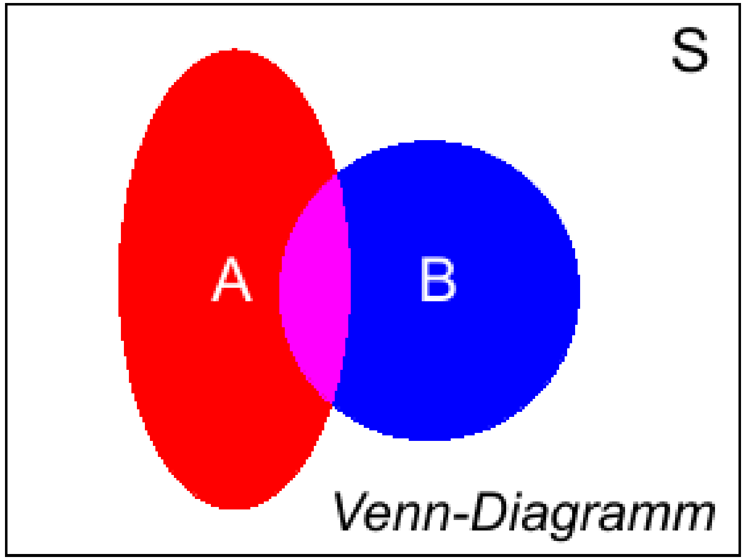
- A and B are said to be “independent” if:

$$P(A \overset{\text{“and”}}{\cap} B) = P(A) \cdot P(B) \quad \Longleftrightarrow \quad P(A|B) = P(A)$$

Cowan, p.2

Probability





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

"Probability of *A* under the condition that B"
Or in short: "probability of *A* given *B*"

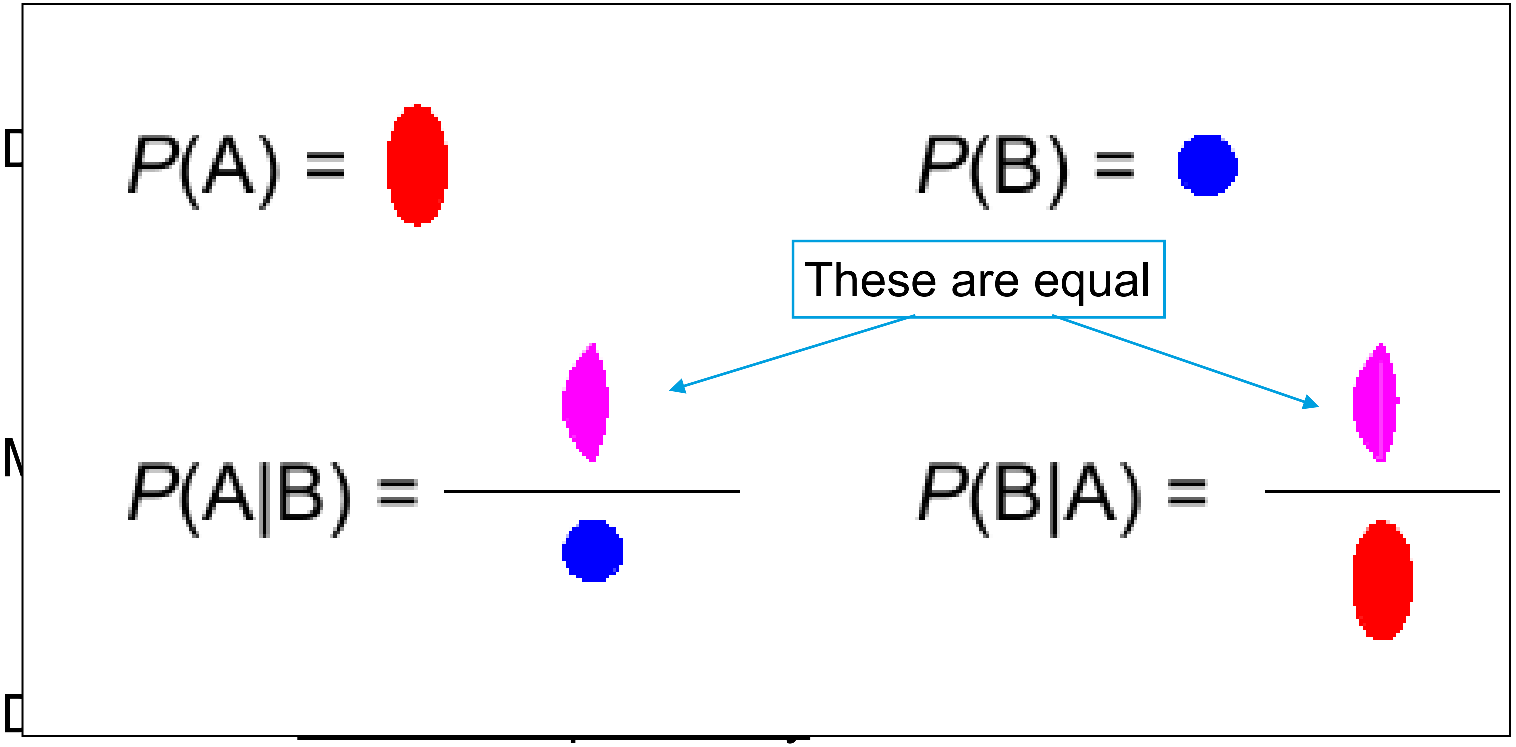
- A and B are said to be "independent" if:

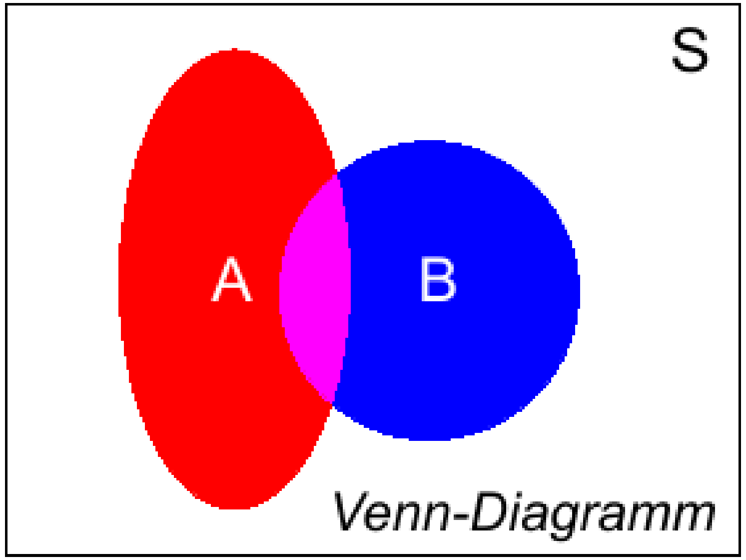
"and"

$$P(A \cap B) = P(A) \cdot P(B) \quad \Leftrightarrow \quad P(A|B) = P(A)$$

Cowan, p.2

Probability





- D
- N
- D

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

"Probability of *A* under the condition that B"
Or in short: "probability of *A* given *B*"

- A and B are said to be "independent" if:

"and"

$$P(A \cap B) = P(A) \cdot P(B) \quad <=> \quad P(A|B) = P(A)$$

Cowan, p.2

Bayes' Theorem

Application in measurements

Thomas Bayes, 1763

"Likelihood" "Prior"

"Posterior"

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

"Evidence"

Probability that theory "A" is correct, given data "B" have been measured

Conditional probability to measure data "B" assuming that theory "A" is correct

Quantitative relation between correctness of a theory ↔ and observation of actual data

Bayes' Theorem

Example: Covid test

- Assume Covid infection rate has a prior probability of 1 permille (0.001)

  P(A) = 0.001 → *P(not A)* = 0.999 "Prior"

- Reliable Covid test (numbers invented)
  - If the patient is infected, the test delivers a correct result in 90% of the cases ("correct positive")

    P(+|A) = 0.90 → *P(- |A)* = 0.10 ("false negative")
  - In case the patient is not infected, the test delivers a correct result in 99% of the cases ("correct negative")

    P(- |not A) = 0.99 → *P(+|not A)* = 0.01 ("false positive")

- If the test is positive, what is the probability *P(A|+)* that the patient is really infected ? "Posterior"

$$P(A|+) = \frac{P(+|A)P(A)}{P(+)} = \frac{P(+|A)P(A)}{P(+|A)P(A) + P(+|\overline{A})P(\overline{A})}$$

$$P(A|+) = \frac{0.90 \times 0.001}{0.90 \times 0.001 + 0.01 \times 0.999} \approx 8.2\%$$

In practice, more tests with symptoms → larger prior

Basic probability distributions

The binomial distribution

Probability to find k events
in the first k trials
and not in the last n-k

- Be p the probability to observe a certain event
- What is the probability, to see k such events in n trials (e.g. find a "6" in 2 out of 3 dice)

$$P(k;n) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

- Binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$: number of combinations to select k out of n elements.

- The probability to find 2 times a "6" in 3 attempts (*p=1/6, n=3, k=2*) is:

$$\frac{3!}{2!(3-2)!} \cdot \left(\frac{1}{6}\right)^2 \left(1-\frac{1}{6}\right)^{3-2} = 0.069$$

animate_Binomial.py
`(pip3 install scipy)`

Basic probability distributions

The Poisson distribution

- Binomial distribution in the limit $n \rightarrow \infty$, $p \rightarrow 0$ and $np = \mu < \infty$ fixed, i.e.
  - $p = \mu/n$ for n large, i.e. $p \ll 1$
  - Only one parameter μ !

$$P(k) = e^{-\mu} \cdot \frac{\mu^k}{k!}$$

- Examples for Poisson distributions:
  - Probability that 4 (=k) out of 100 chips (=n) are produced with a defect, if the error rate is 2% (=p).

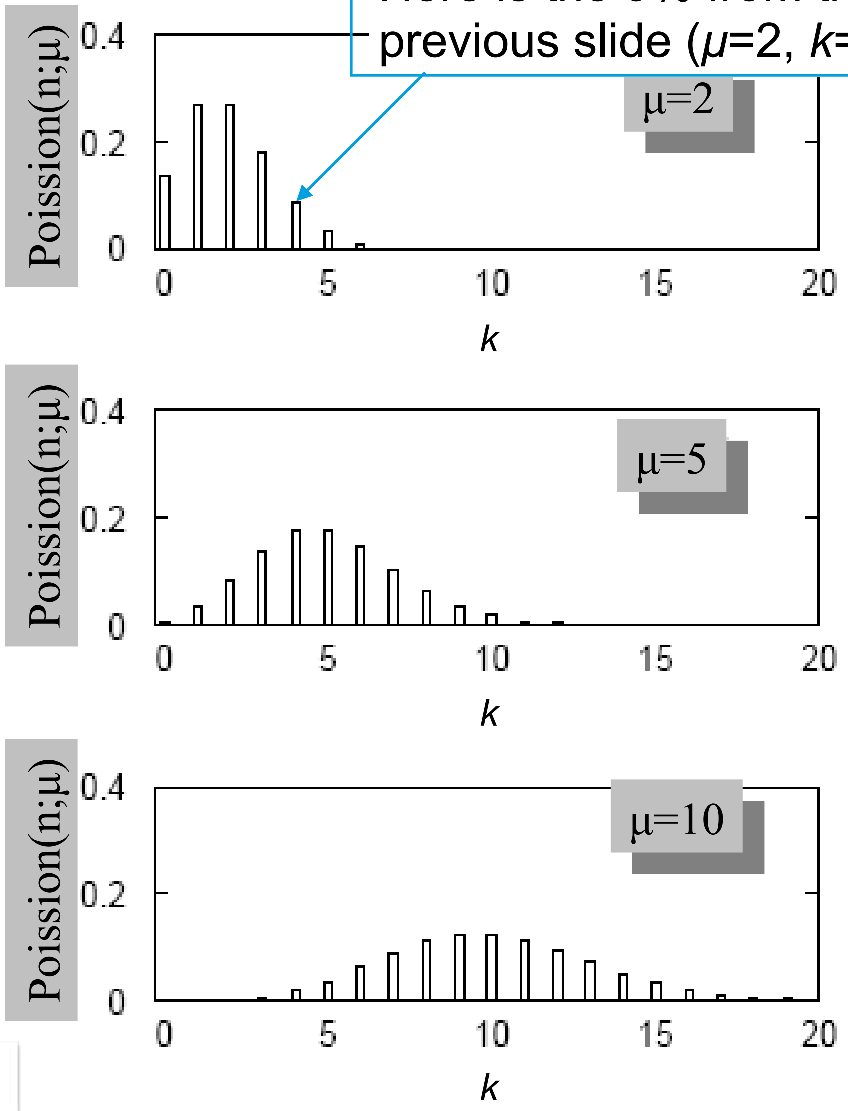
$$\mu = n \cdot p = 100 \cdot 0.02 = 2 \qquad P(4) = e^{-2} \cdot \frac{2^4}{4!} \qquad \sim 9\%$$

  - At fixed event rate, the number of events observed in a time interval t (the distribution of differences in time follows an exponential)
  - The number of entries in a histogram

Erdmann/Hebbeker p.37 / Blobel/Lohrmann Example 4.11

Poisson distribution

Here is the 9% from the previous slide (μ=2, k=4)



Cowan, Fig 2.3

$$P(k) = e^{-\mu} \cdot \frac{\mu^k}{k!}$$

Expectation value: $\mathrm{E}[k] = \mu$

Variance: $\mathrm{V}[k] = \mu$

Standard deviation: $\sigma = \sqrt{\mu}$

Statistical uncertainty is often estimated as $\sqrt{k}$.
Assumption $k \approx \mu$. Not correct, as k can fluctuate from μ

Example: animate_Gaussfunctions.py

	$1-\alpha$	α	$\alpha/2$
1σ	0.683	0.317	0.158
1.65σ	0.90	0.10	0.05
1.96σ	0.95	0.05	0.025
2σ	0.9545	0.0455	0.0228
3σ	0.9973	0.0027	0.0013
5σ			3×10^{-7}

Measurements: 2-sided interval: *p*-value = α

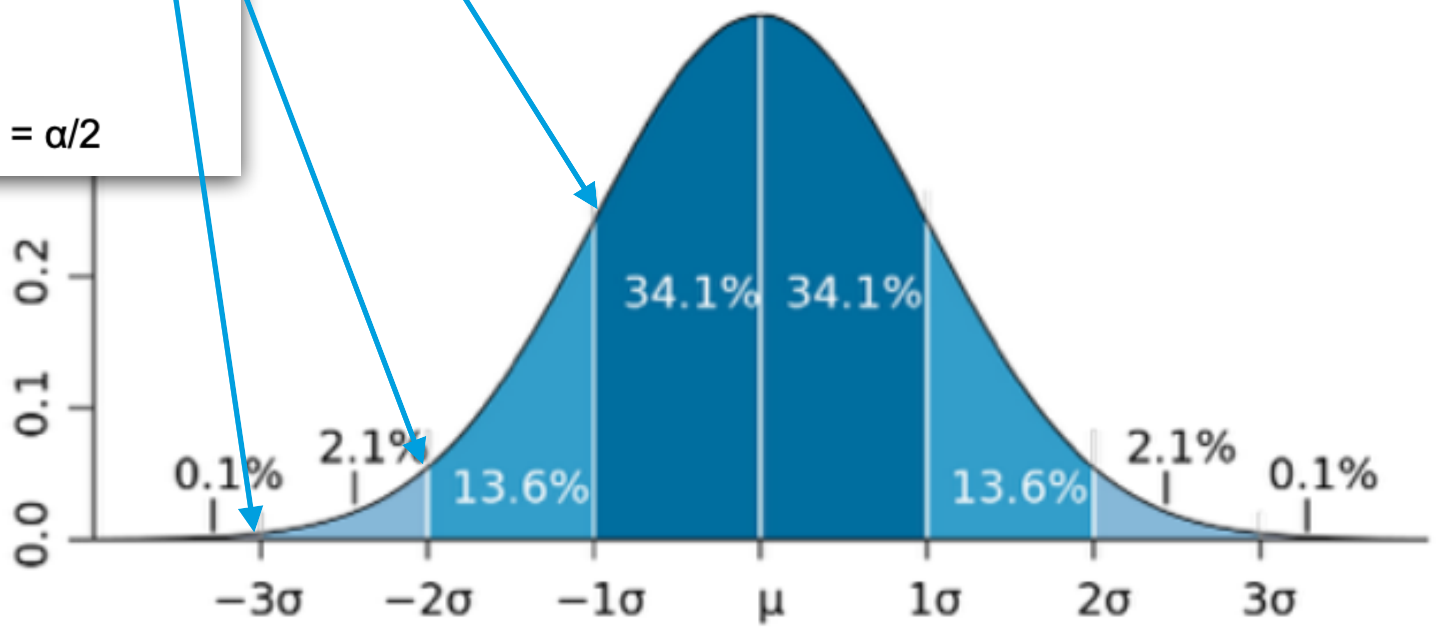
Exclusion/discovery: 1-sided interval: *p*-value = α/2

$-\frac{(x-\mu)^2}{2\sigma^2}$

Towards large *n and μ*:
binomial and Poisson-distributions
approach the Gauss-distribution

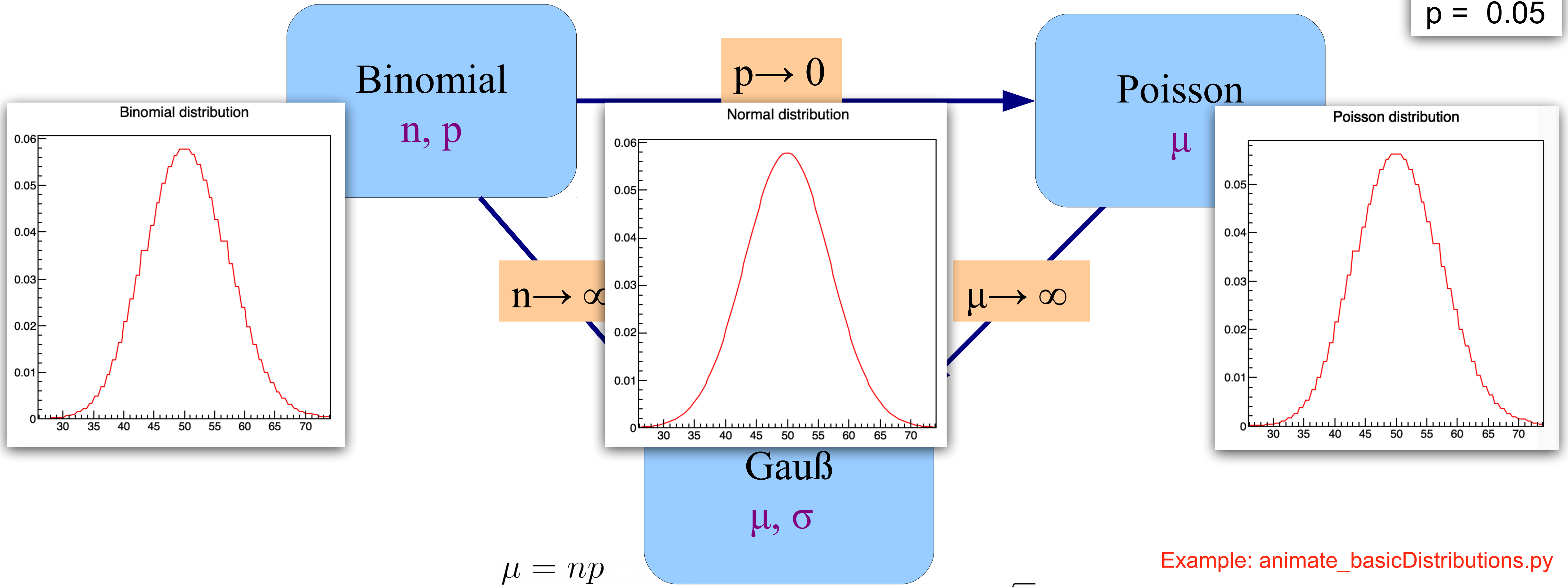
- Quantiles of the Gauss-distribution
  (a.k.a. confidence intervals)

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



Gauss - Poisson - Binomial

n = 1000
p = 0.05



$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

$$\sigma = \sqrt{\mu}$$

Example: animate_basicDistributions.py

See also: Blobel/Lohrmann Example 4.16

Parameter Estimation

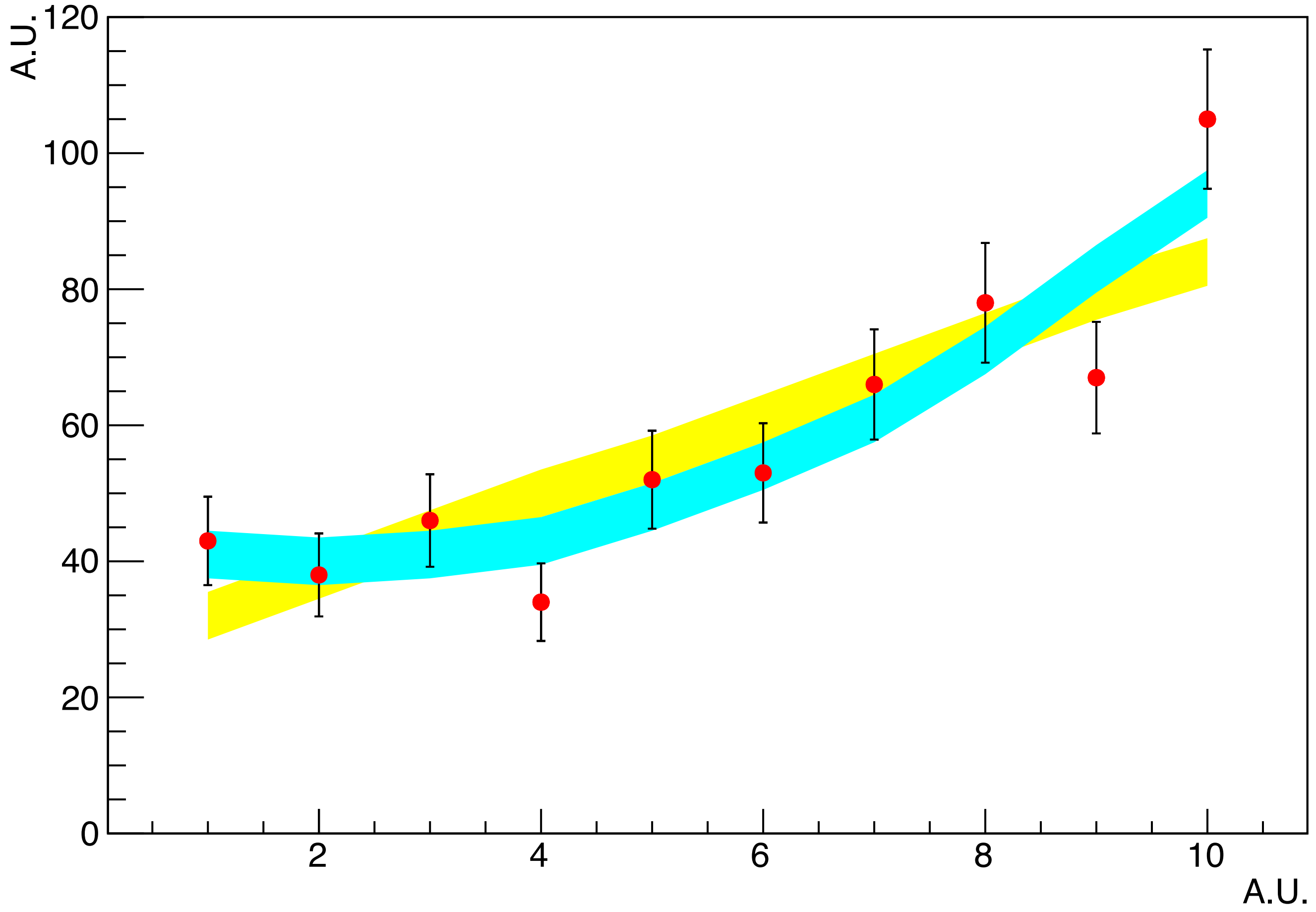
Parameter estimation

- Probability density distributions (PDF) are defined by “truth” parameters a
- Measurements of distributions consist of random samples, i.e. n elements $x_1, x_2 \dots x_n$
- The random sample contains information to determine the estimator $\hat{a}$
- The estimator $\hat{a}$ is also a random number
- In the limit of many identical experiments, the distribution of $\hat{a}$ should approach the PDF of a

- In a measurement, the estimator $\hat{a}$ should be:
  - consistent: $\lim_{n\to\infty} \hat{a} = a$
  - unbiased: $E(\hat{a}) = a$ also for $n < \infty$ (!)
  - efficient: $V(\hat{a})$ as small as possible
  - robust: stable against wrong data

Goodness of fit

Example: χ^2 distribution



- Which theory gives a better description of the data ?

Goodness of fit

Example: χ^2 distribution

- How well does a prediction fit the data ?
- Weighted sum of squares of the difference between Gaussian-distributed data y_i with uncertainties σ_i, and theory $f(x_i)$:

$$S = \chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i, \{p\})}{\sigma_i} \right)^2$$

- The estimator S follows a χ^2-distribution (with $n = N - k$ degrees of freedom):

$$f_n(\chi^2) = \frac{\frac{1}{2}\left(\frac{\chi^2}{2}\right)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}}{\Gamma\left(\frac{n}{2}\right)}$$

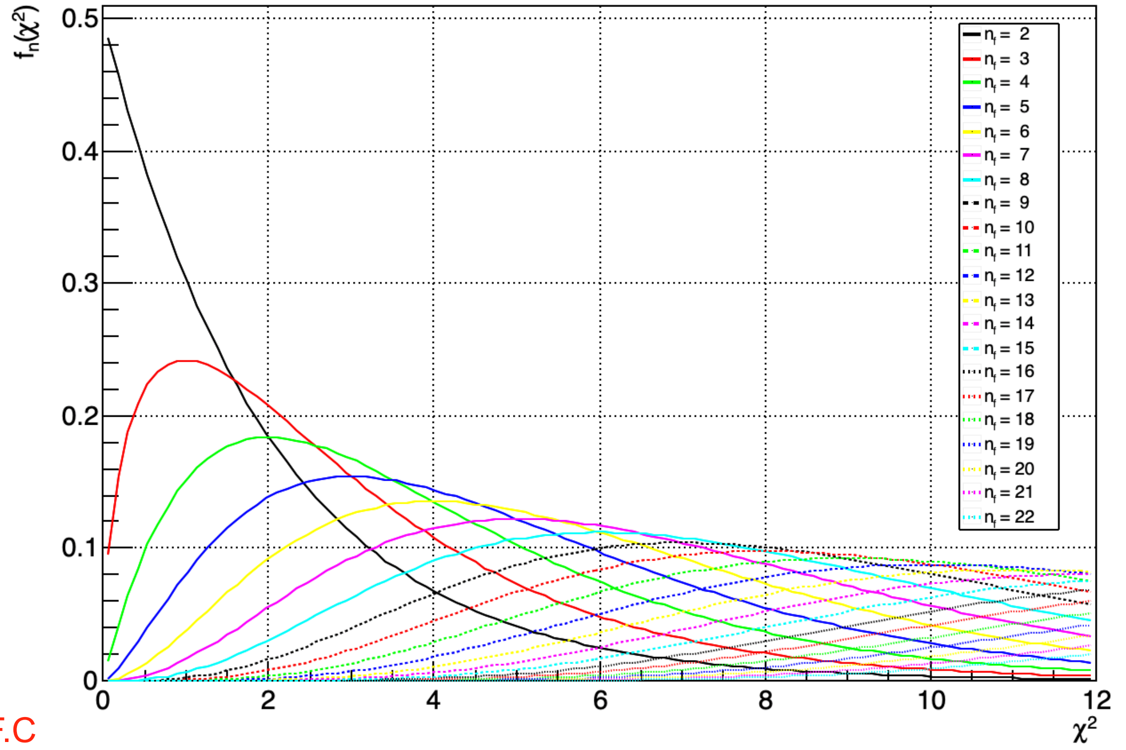


Figure: chi2PDF.C

- Expectation value: $\langle \chi^2 \rangle = n$ $\Rightarrow$ $\langle \chi^2/n \rangle = 1$
- Variance: $V[\chi^2] = 2n$

Many other goodness-of-fit tests, e.g. Kolmogorov Smirnov

χ^2-Probability

- Probability to find a bigger value of χ^2 than the one actually measured:

$$\chi^2_{\text{Prob}} = \int_{\chi^2}^{\infty} f_n(v)dv$$

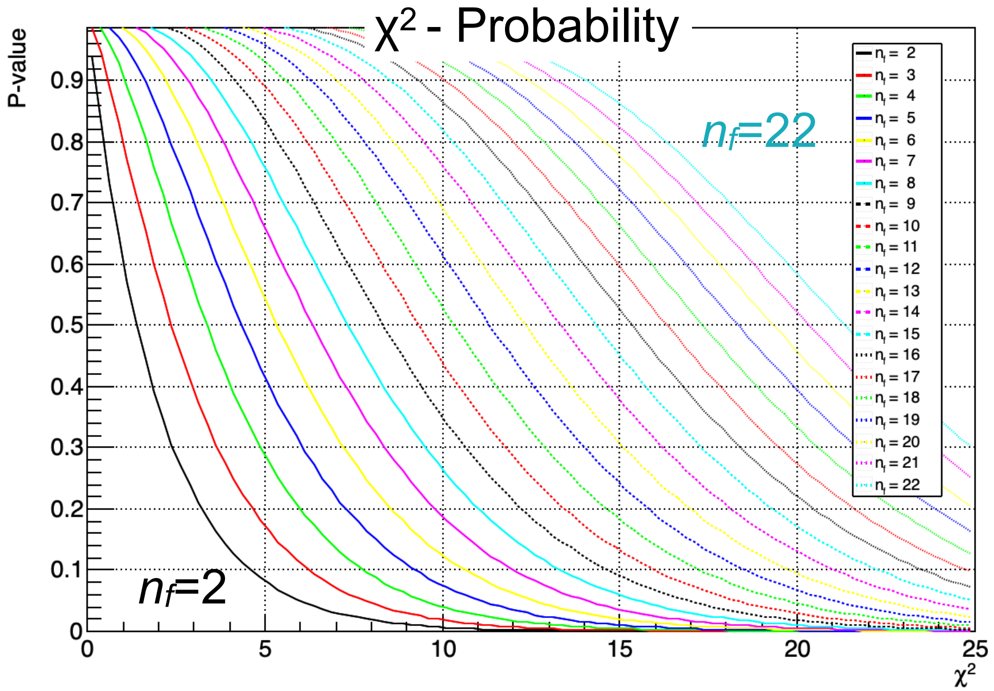
$$= 1 - \int_0^{\chi^2} f_n(v)dv$$

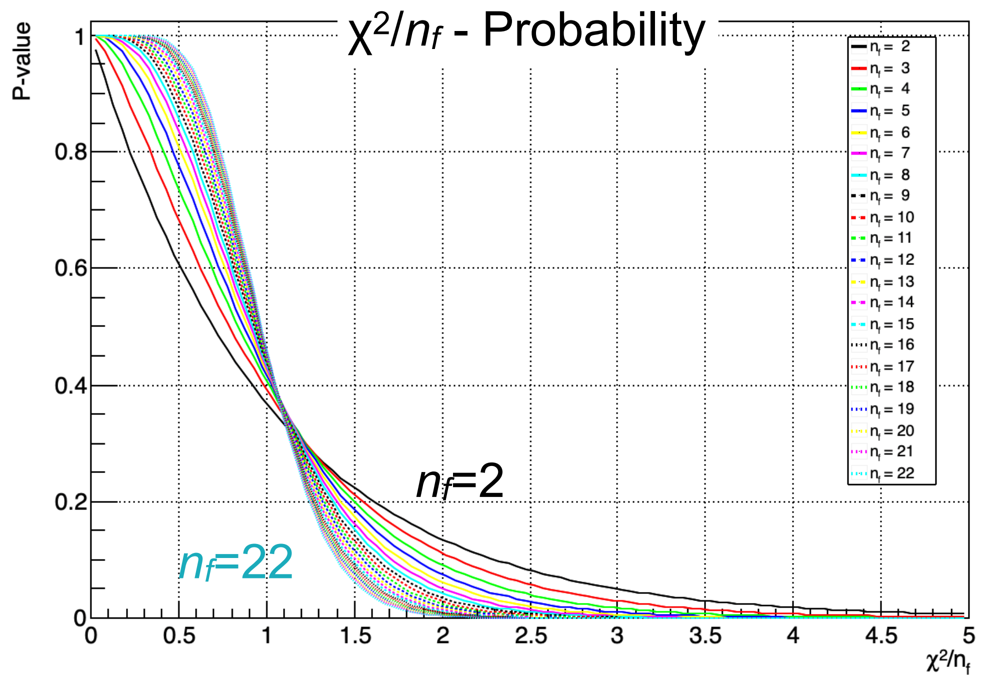
  - Useful to quantify agreement

- Probability of χ^2/n_f > 1 is ~40% (largely independent of n_f)

Note: χ^2/n_f > 1 is more acceptable if n_f is small

Figures: chi2Prob.C



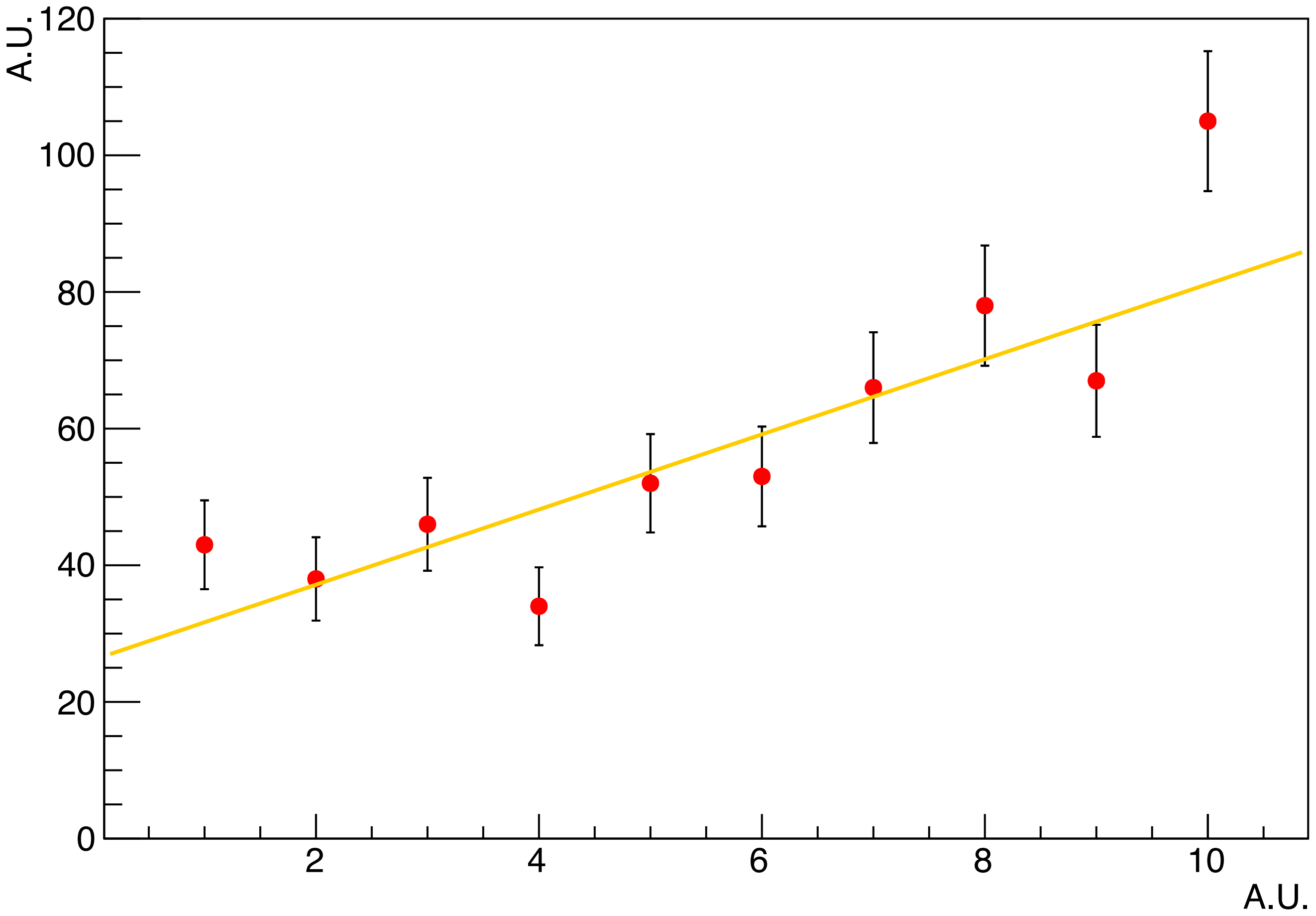


Least Squares Method

- Fit of a function to the data: linear function ax+b (2 free parameters)
- Minimize sum of squares:

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

- $\chi^2 / n_{dof} = 17.6 / 8$ (p-value: 2.4%)



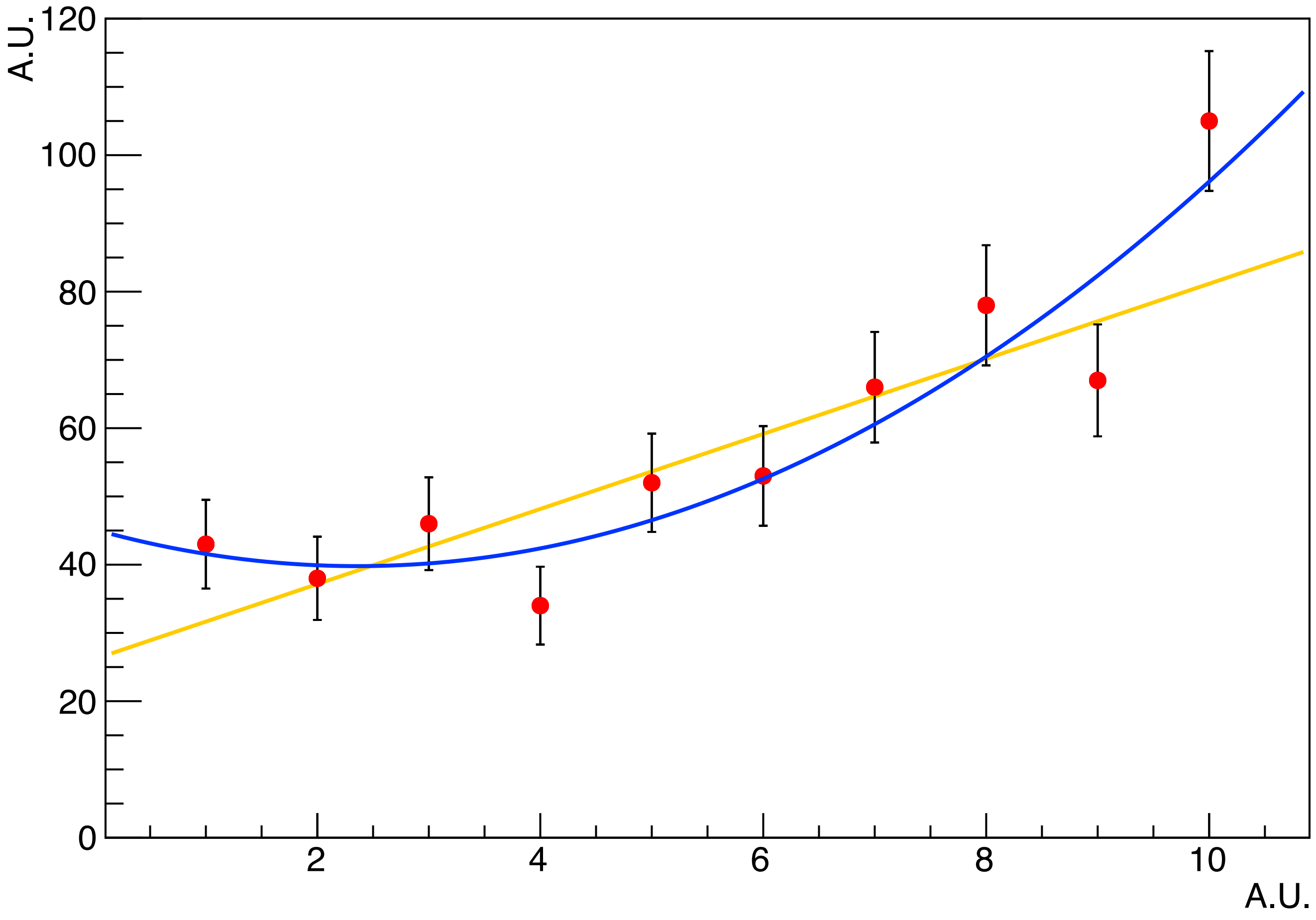
See also Cowan section 7.3

Least Squares Method

- Fit of a function to the data: quadratic function ax^2+bx+c (3 free parameters)
- Minimize sum of squares

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

- $\chi^2 / n_{dof} = 9.1 / 7$ (p-value: 24%)



See also Cowan section 7.3

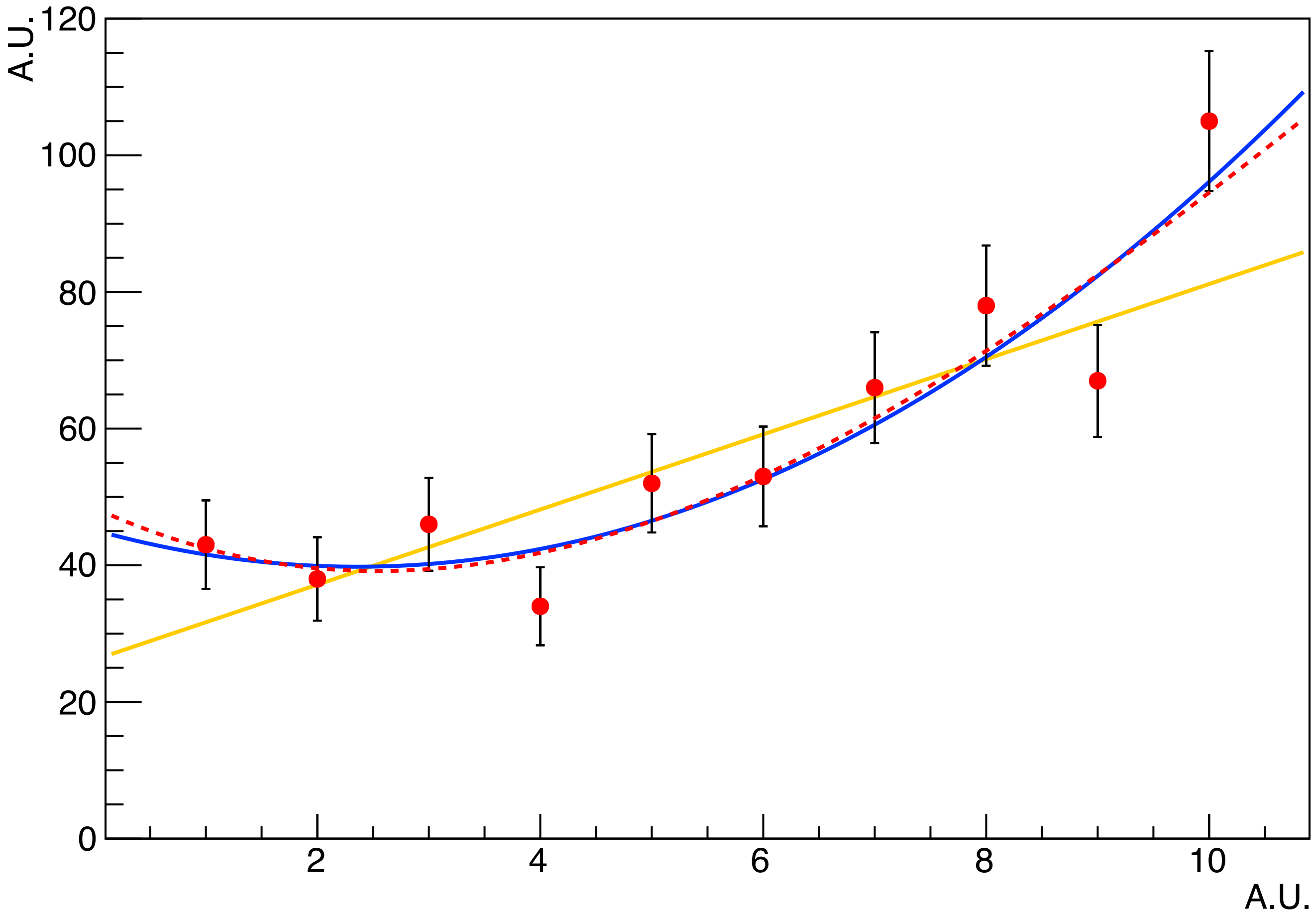
Better p-value than with 2 parameters

Least Squares Method

- Fit of a function to the data: quadratic function ax^3+bx^2+cx+d (4 free parameters)
- Minimize sum of squares

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

- $\chi^2 / n_{dof} = 8.9 / 6$ (*p*-value: 17%)



See also Cowan section 7.3

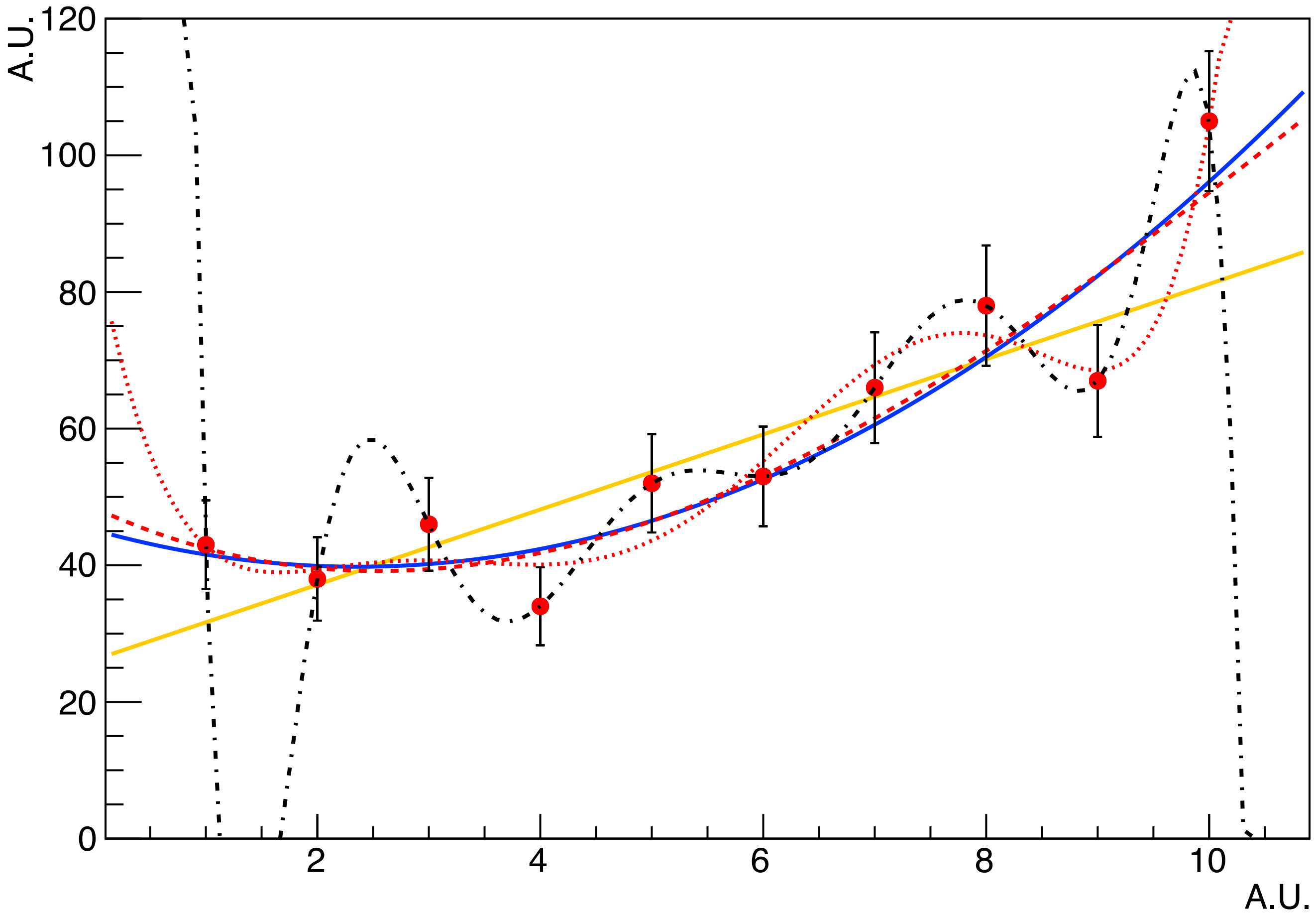
Rule of thumb:
Do not use more parameters than needed => 4 parameters already too many

Least Squares Method

- Fit of data with 9 parameters:
- Minimize sum of squares

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

- $\chi^2 / n_{dof} = 0 / 0$



See also Cowan section 7.3

With enough parameters "can fit an elephant”

Parameter correlation

"Covariance"

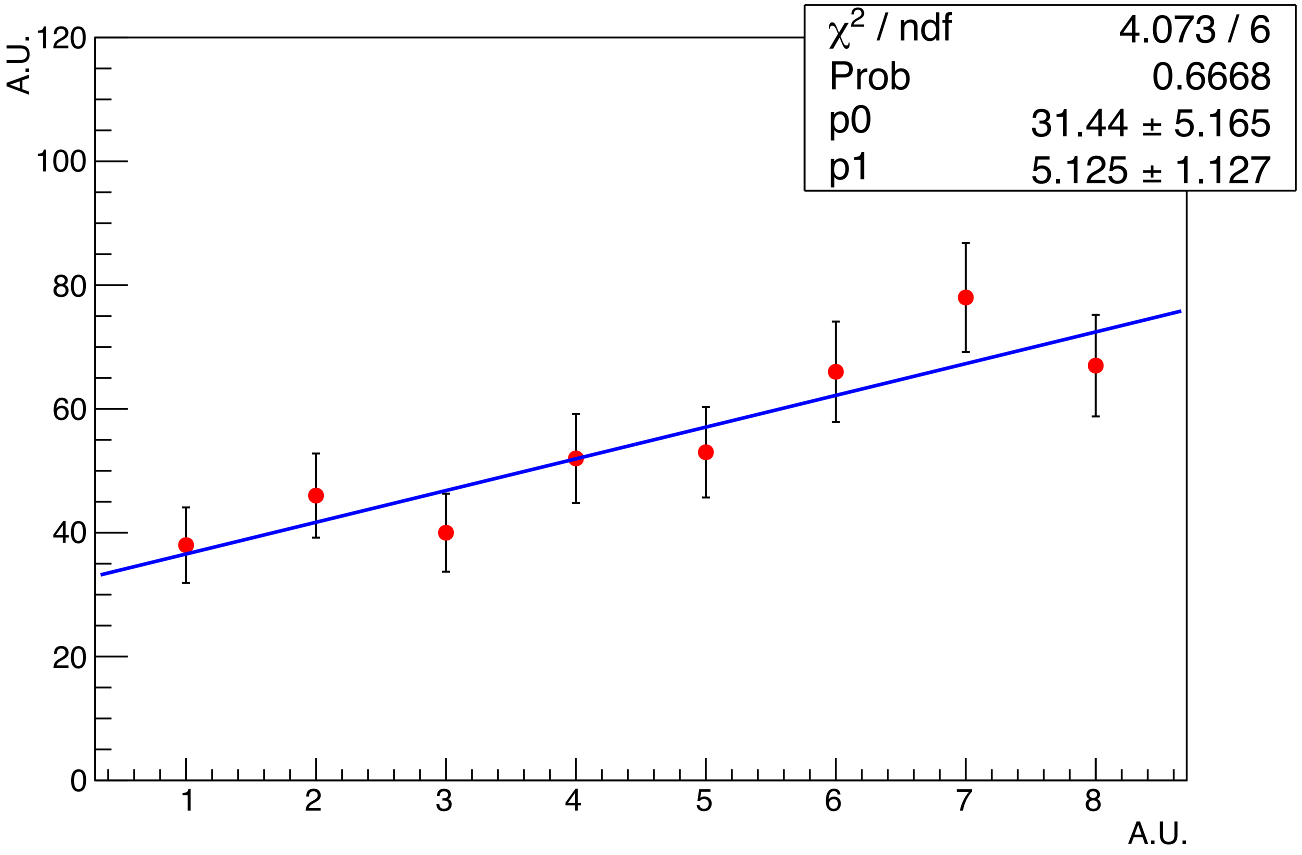
$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

"Correlation"

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ where } \rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

- Ansatz: $f(x) = p_0 + p_1 x$ (2 parameters)

- Result of the fit:
  - Parameters and χ^2: see caption
  - Covariance: $\begin{pmatrix} 26.7 & -5.07 \\ -5.07 & 1.27 \end{pmatrix}$
  - Correlation: $\begin{pmatrix} 1 & -0.86 \\ -0.86 & 1 \end{pmatrix}$



See also Cowan section 7.3

Parameters p_0 und p_1 are strongly anti-correlated

=> bigger p_0 requires smaller p_1, such that *f(x)* can still go through the data points

Parameter correlation

"Covariance"

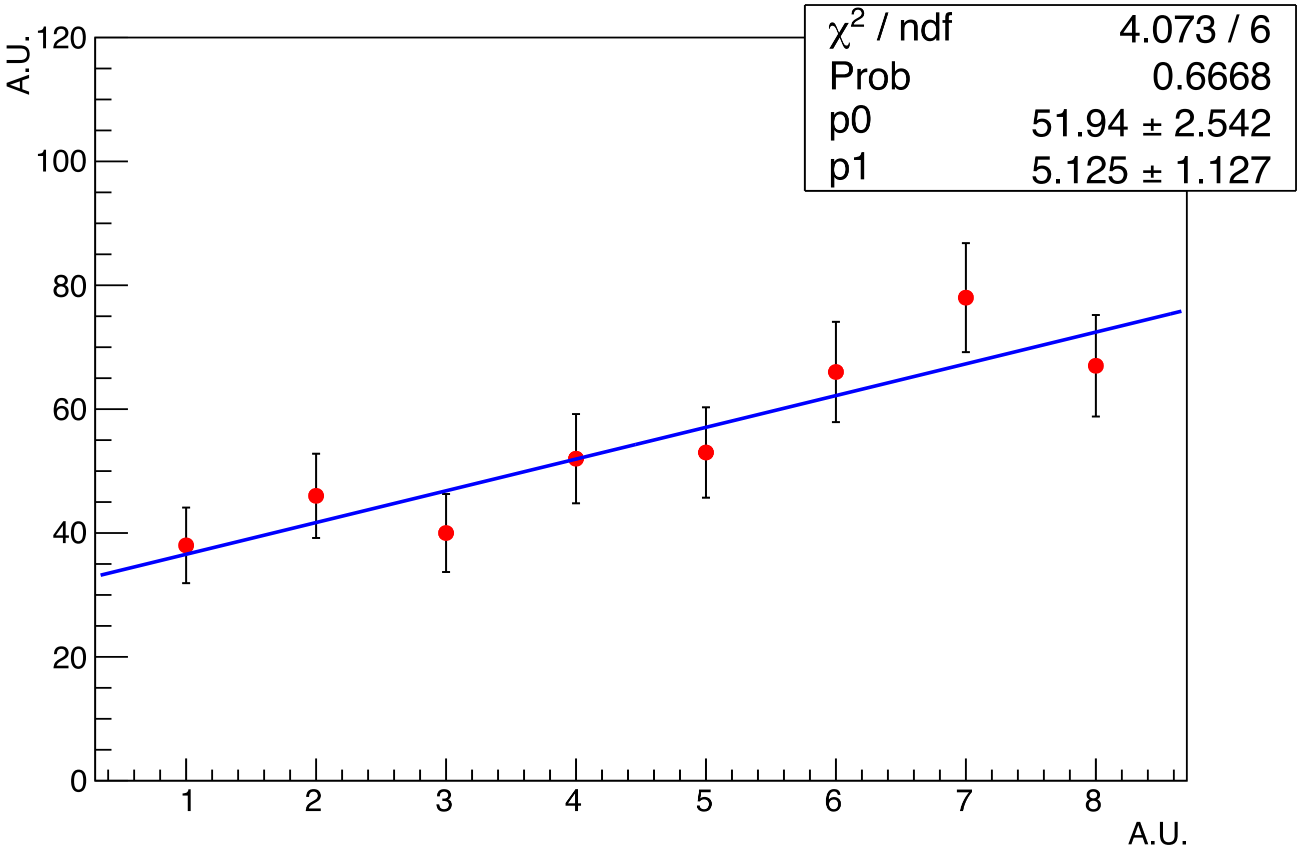
$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

"Correlation"

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ where } \rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

- Ansatz: $f(x) = p_0 + p_1\,(x-4)$
  (2 parameters)

- Result of the fit:
  - Parameters and χ^2: see caption
  - Covariance: $\begin{pmatrix} 6.4 & 0.013 \\ 0.013 & 1.27 \end{pmatrix}$
  - Correlation: $\begin{pmatrix} 1 & 0.004 \\ 0.004 & 1 \end{pmatrix}$



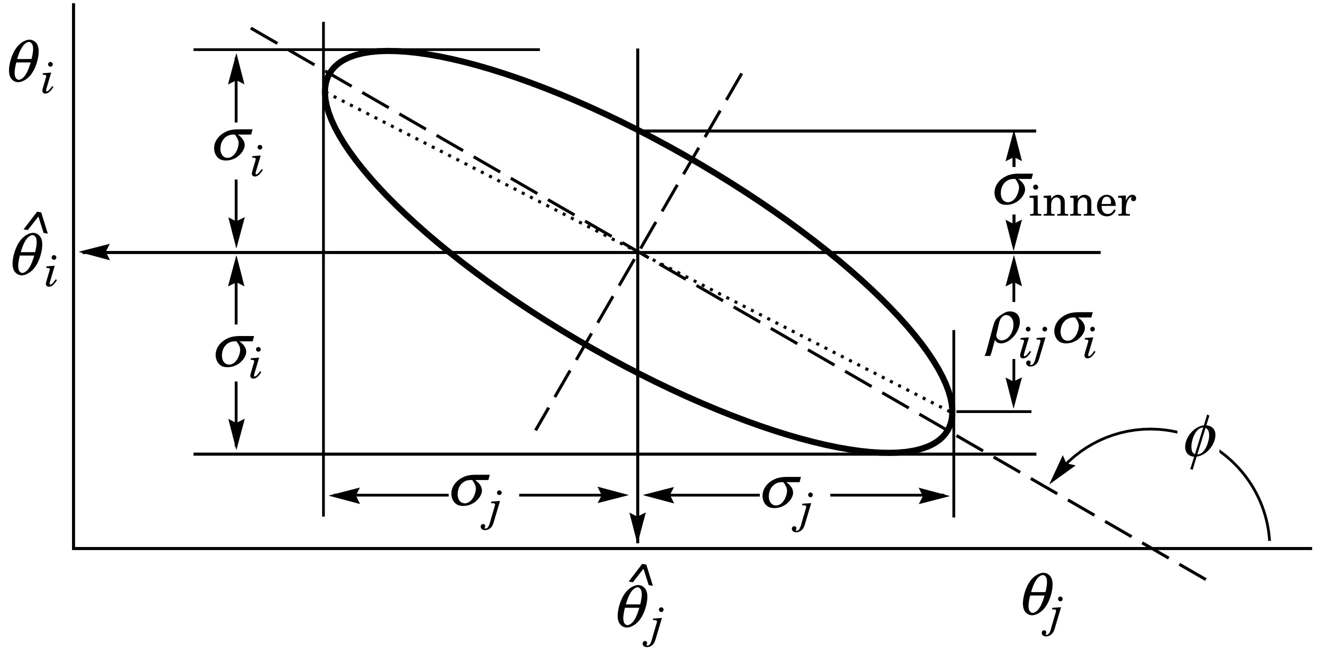
See also Cowan section 7.3

Decorrelation often possible by appropriate transformation of coordinates (e.g. principle component analysis)

Parameter correlation

"Covariance" $\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$ "Correlation" $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ where $\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$

- 1σ contour, corresponding to $\Delta\chi^2 = 1$, for the measured (estimated) parameters θ_i and θ_j



Cowan PDG, Fig 40.5

Fixing one parameter (here θ_j) leads to reduction of the uncertainty of parameter θ_i

$$\sigma_{\text{inner}} = \sqrt{1 - \rho_{ij}^2} \cdot \sigma_i$$

Summary

- The scientific cycle:
  - Theoretical predictions are tested by experiment
  - Experimentally determined model parameters are input to theoretical predictions (e.g. Standard Model)

- Measurements are random samples drawn from a true distribution described by a PDF.
  - Statistical uncertainties (variance -> spread): well understood
  - Systematic uncertainties (bias -> distortion): require care and courage

- Probability
  - Bayes' Theorem
  - Binomial-, Poisson- and Gaussian distributions

- Parameter estimation
  - χ^2 - function, goodness of fit and decorrelation

Menu

Today

- Statistical and systematic uncertainties
- Probability
- Parameter estimation

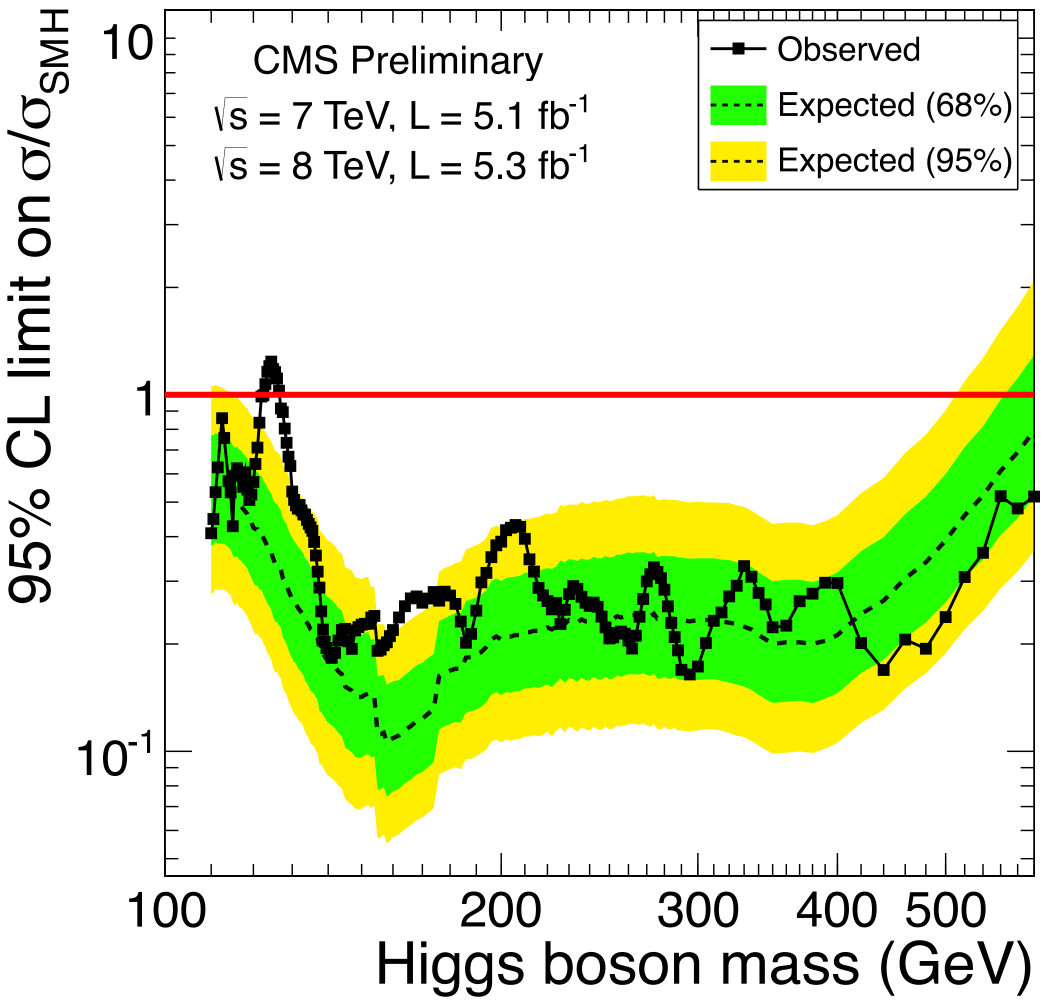
Tomorrow

- Hypothesis testing
- Confidence intervals
- Profile likelihood ratio
- Outlook: classification and MVA

Friday

Matthias Komm:

Introduction to machine learning



Higgs discovery: What does this figure really show ?