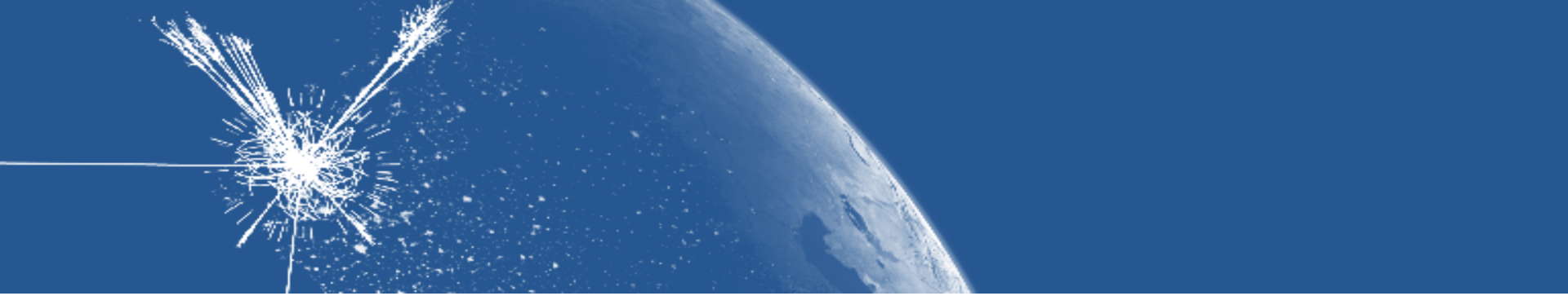
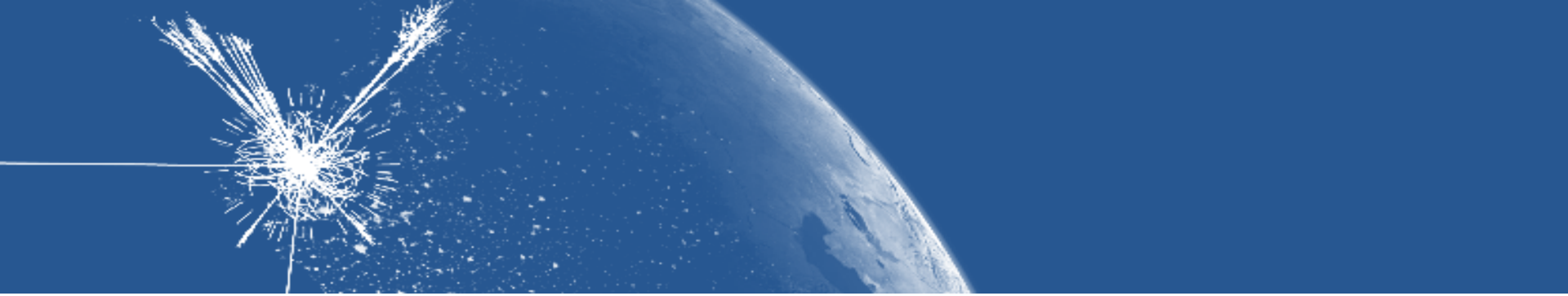


Introduction to probability and statistics (3)

Andreas Hoecker (CERN)

CERN Summer Student Lecture, 17–21 July 2017

If you have questions, please do not hesitate to contact me: **andreas.hoecker@cern.ch**



Outline (4 lectures)

1st lecture:

- Introduction
- Probability

2nd lecture:

- Probability axioms and hypothesis testing
- Parameter estimation
- Confidence levels (some catch up to do…)

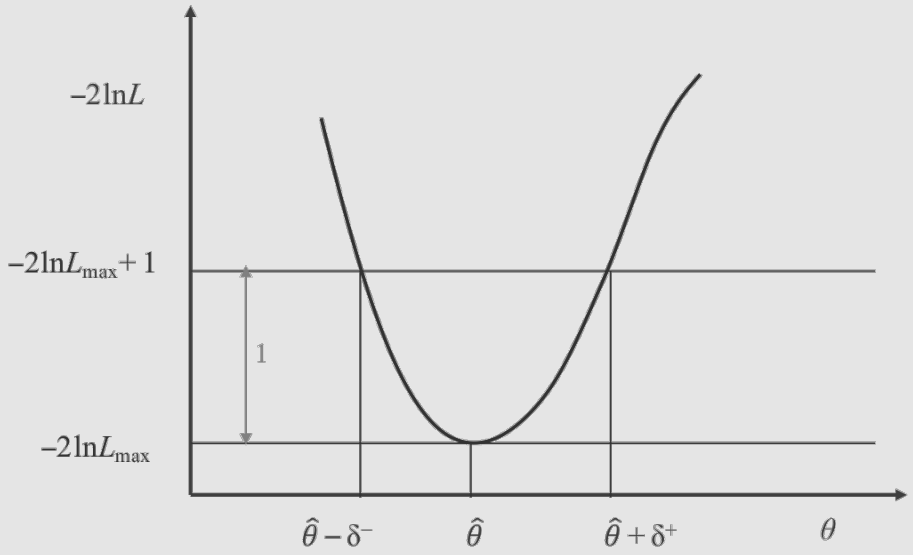
3rd lecture:

- Maximum likelihood fits
- Monte Carlo methods
- Data unfolding

4th lecture:

- Multivariate techniques and machine learning

Catch-up from yesterday

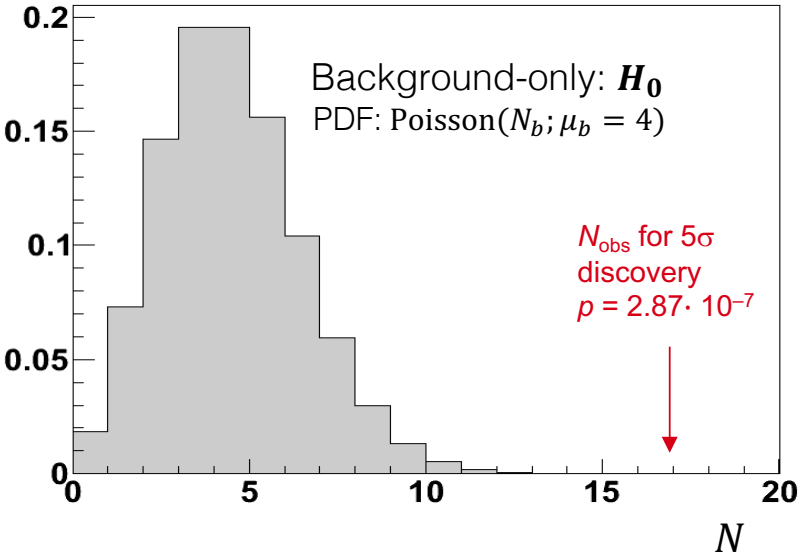


Statistical tests in new particle/physics searches

Discovery test

- Disprove background-only hypothesis $\boldsymbol{H_0}$
- Estimate probability of “upward” (or “signal-like”) fluctuation of background

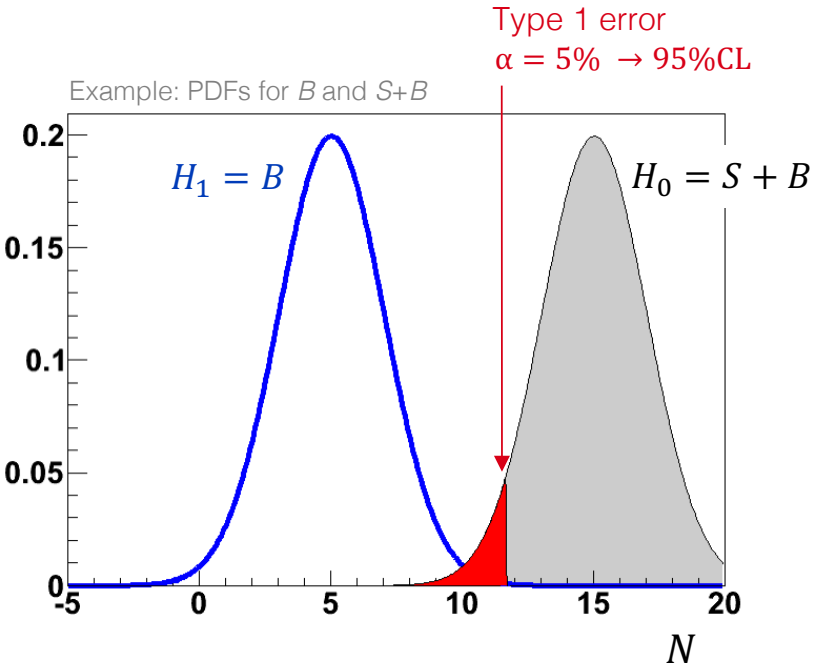
Example: PDF of background-only test statistic



Exclusion limit

- Upper limit on new physics cross section
- Disprove signal + background hypothesis $\boldsymbol{H_0}$
- Estimate probability of downward fluctuation of signal + background: find minimal signal, for which $\boldsymbol{H_0}$ (here: *S*+*B*) can be excluded at specified confidence Level

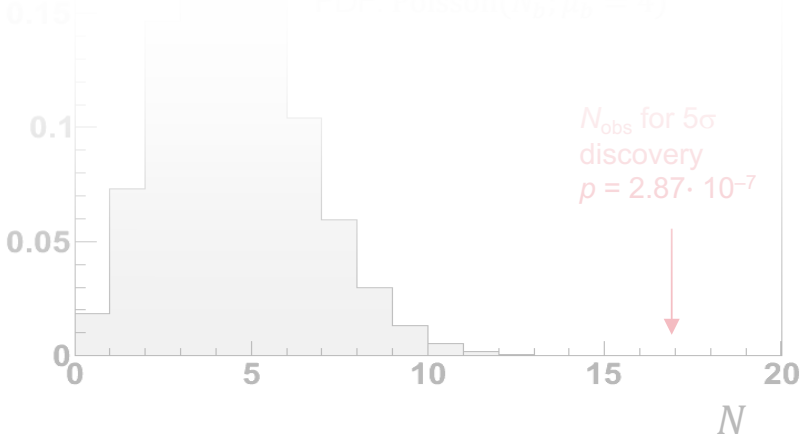
Example: PDFs for *B* and *S*+*B*

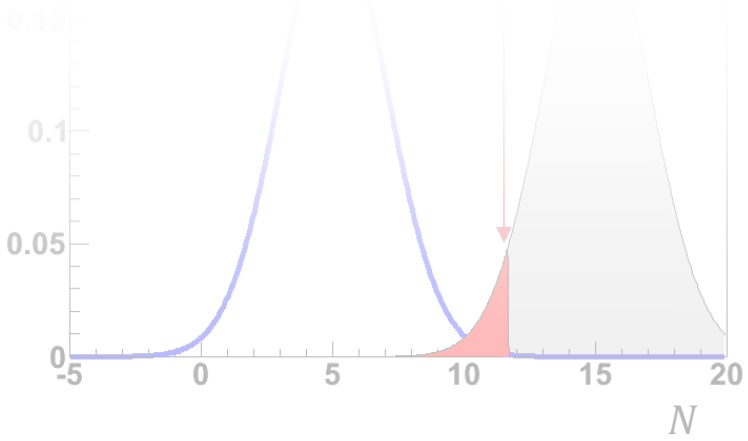


Statistical tests in new particle/physics searches

Realistic discovery and exclusion likelihood tests involve complex fits of several signal and background-normalisation (so-called *control*) regions, signal and background yields, as well as *nuisance parameters* describing systematic uncertainties.

We will come to this, but first need to learn about parameter estimation.

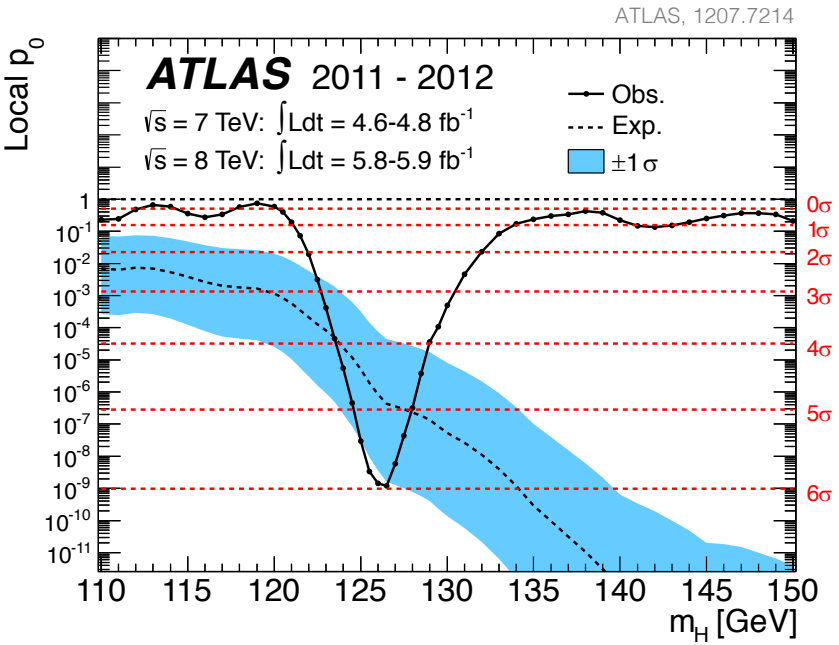




Statistical tests in new particle/physics searches — teaser

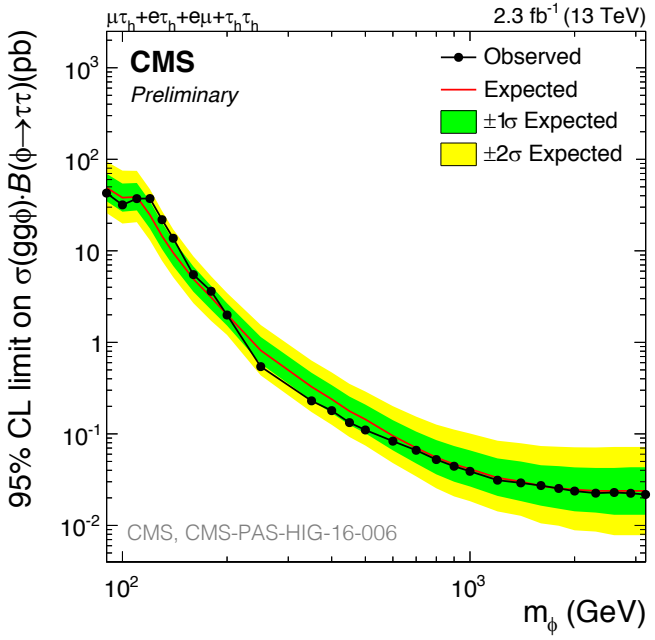
Discovery test — Higgs discovery in 2012

- 5.9σ rejection of background-only hypothesis from statistical combination of dominantly $H \to \gamma\gamma$, ZZ*, WW* decays at m_H = 126 GeV
- No ***trials factor*** (*look-elsewhere-effect*, LEE) taken into account in above number, but would not qualitatively change picture



Exclusion limit

- 13 TeV search for new physics (here: a new heavy Higgs boson) in events with at least two tau leptons
- Figure shows expected and observed 95% confidence level upper limits on cross section times branching fraction



Parameter estimation

An *estimator* is a function of a data sample $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{x_1}, \ldots., \boldsymbol{x_N})$ that estimates the characteristic parameter $\boldsymbol{\theta}$ of a parent distribution.

Examples:

- Mean value estimator: $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$ (one way to define the mean value, there could be others)
- Variance estimator: $\hat{V} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$
- Median estimator …..
- ...but also: CP-asymmetry parameter in *B* meson sample (very complex parameter estimation)

The estimator $\widehat{\boldsymbol{\theta}}$ is a random variable (function of measured data that are random)

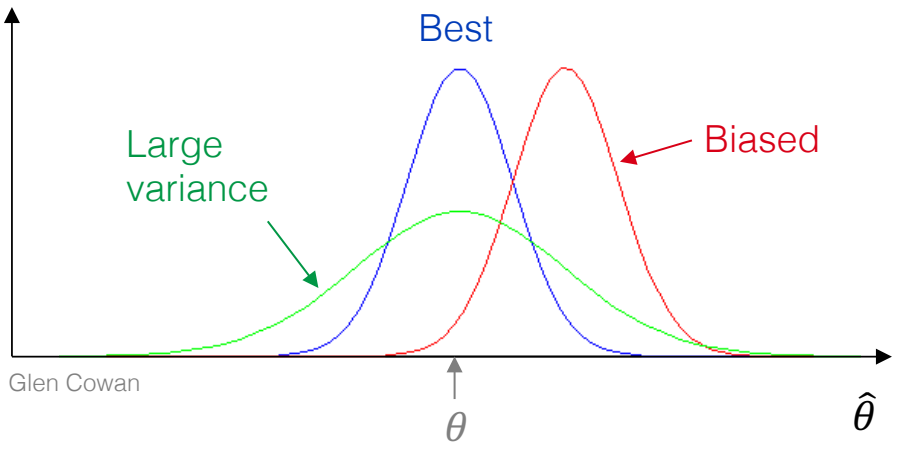
The estimator $\widehat{\boldsymbol{\theta}}$ has itself an expectation value, an expected variance, for given $\boldsymbol{\theta}$:

$\rightarrow \quad E\left[\hat{\theta}(x)\middle|\theta\right] = \int \hat{\theta}\,(x) f(x|\theta) dx$, with $f(x|\theta)$ the distribution (PDF) of the expected data

Parameter estimation

$\hat{\boldsymbol{\theta}}$ is a random variable that follows a PDF. Consider many measurements / experiments:

→ There will be a spread of $\hat{\theta}$ estimates. Different estimators can have different properties:



- Biased or unbiased: if $E\left[\hat{\theta}(x)\middle|\theta\right] = \theta \;\rightarrow$ unbiased
- Small bias and small variance can be "in conflict"
  - asymptotic bias → limit for infinite observations/data samples

Maximum likelihood estimator

Want to estimate (measure !) a parameter $\boldsymbol{\theta}$

Observe $\vec{x}_i = (x_1, \ldots . x_K)_i, i = 1, N$ (ie: K observables per event, and N events)

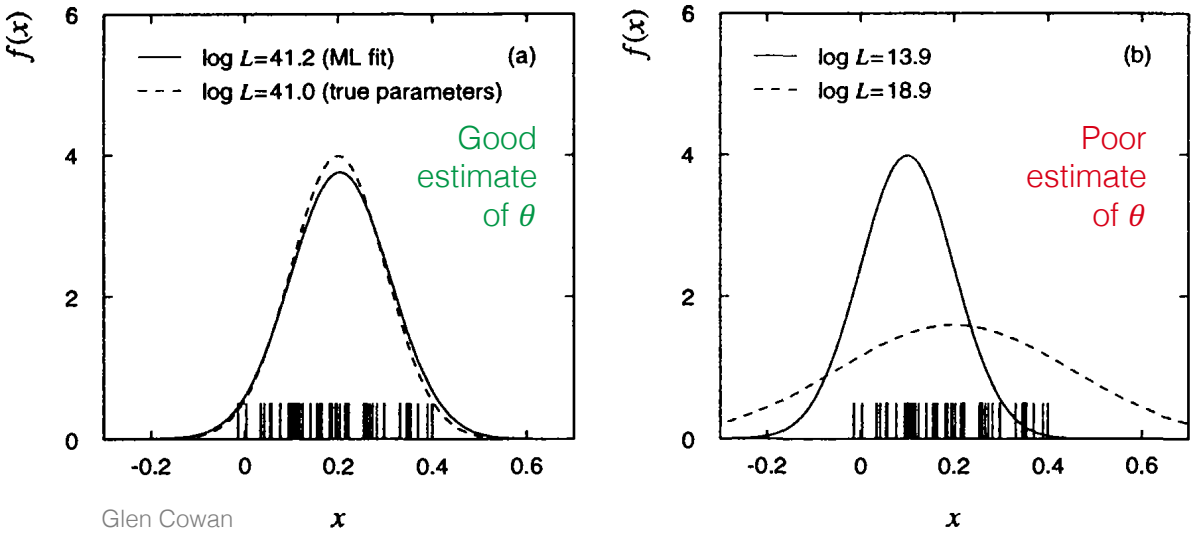
Hypothesis is PDF $p_x(\vec{x}; \theta)$, ie, the distribution of $\vec{x}$ given θ

There are N *independent* events $\rightarrow$ combine their PDFs: $P(\vec{x}_1, .. \vec{x}_N; \theta) = \prod_{i=1}^{N} p_x(\vec{x}_i; \theta)$

For fixed $\vec{x}$ consider $p_x(\vec{x}; \theta)$ as function of θ $\rightarrow$ ***Likelihood $L(\theta)$***

- $L(\theta)$ is at maximum (if unbiased) for $\hat{\theta} = \theta_{\text{true}}$

50 observations of .Gaussian random variable with mean 0.2 and σ=0.1



Task: maximise $L(\theta)$ to derive best estimate for $\hat{\theta}$

In practice, often minimise $-2 \cdot \ln(L(\theta))$ (see later why)

$\rightarrow$ Maximum likelihood fit

Maximum likelihood estimator (continued)

Let's take the Gaussian example from before: $L\,(\mu,\sigma|x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

- Measure N events: $x_1,\ldots,x_N$
- Full likelihood given by: $L(\mu,\sigma|x) = \prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$
- In logarithmic form: $-2\cdot\ln\big(L(\mu,\sigma|x)\big) = \sum_{i=1}^{N}\left(\frac{(x_i-\mu)^2}{\sigma^2}\right) - 2N\cdot\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)$

→ In a full maximum likelihood fit one could now determine $\hat{\mu}$ and $\hat{\sigma}$

→ If one is not interested in fitting σ but just μ, one can omit the (then constant) 2nd term:

$$-2\cdot\Delta\ln\big(L(\mu|x)\big) = \sum_{i=1}^{N}\left(\frac{(x_i-\mu)^2}{\sigma^2}\right)$$ → which is the **"least squares" (χ^2)** expression

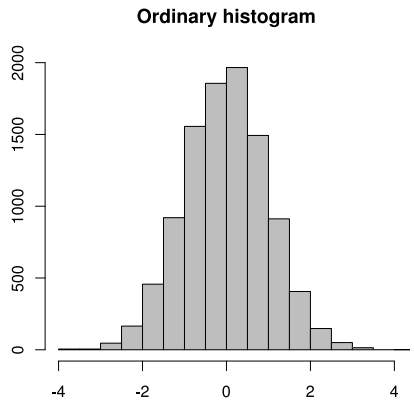
where: $\Delta\ln\big(L(\mu|x)\big) = \ln\big(L(\mu|x)\big) - \text{constant term}$

Maximum likelihood estimator (continued)

So far considered *unbinned* datasets (i.e., likelihood is given by product of PDFs for each event)

One can replace the events by bins of a histogram

- Useful if very large number of events, or PDF has very complex form, or if only broad regions are considered rather than the full shape of a PDF
- Most LHC analyses use binned maximum likelihood fits



Each bin $\boldsymbol{i}$ has $\boldsymbol{N_i}$ events that are Poisson distributed around $\boldsymbol{\mu_i}$

- The prediction of the μ_i can be obtained from Monte Carlo simulation

Likelihood function: $L(\theta) = P\left(N_1, \ldots N_{n_{\text{bins}}}; \theta\right) = \prod_{i=1}^{n_{\text{bins}}} \frac{\mu_i^{N_i}(\theta)}{N_i!} e^{-\mu_i(\theta)}$

…and in log form: $-2 \cdot \ln(L(\theta)) = 2 \sum_{i=1}^{n_{\text{bins}}} \left(\mu_i(\theta) - N_i \ln(\mu_i(\theta)) - \ln(N_i!)\right)$

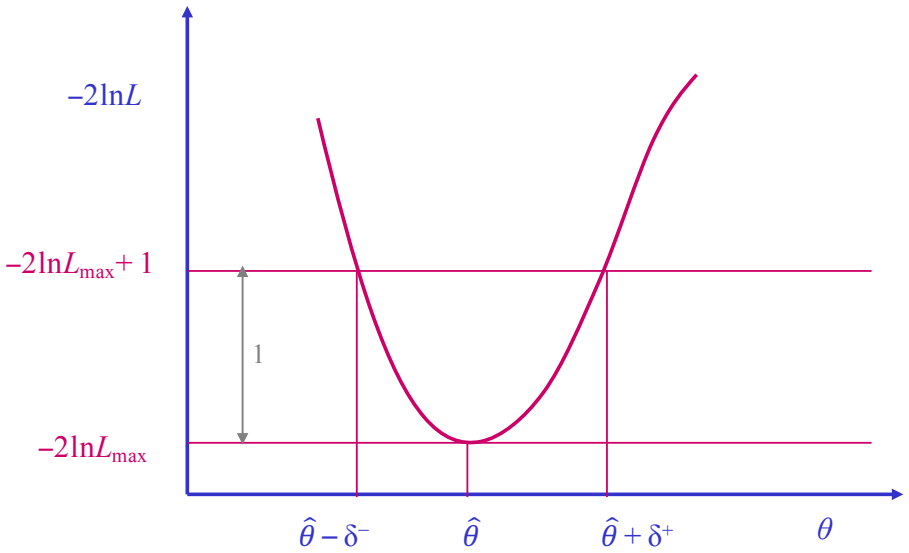
Maximum likelihood estimator (continued)

Maximum likelihood estimator is typically unbiased only in limit $N \to \infty$

If likelihood function is Gaussian (often the case for large N by virtue of central limit theorem):

→ Estimate 1σ confidence interval for θ ("parameter uncertainty") by finding intersections $\mathbf{-2 \cdot \Delta \ln(L) = 1}$ around minimum

→ Resulting uncertainty on θ may be asymmetric



If (very) non-Gaussian:

→ revert typically to (classical) *Neyman confidence intervals* (→ see later)

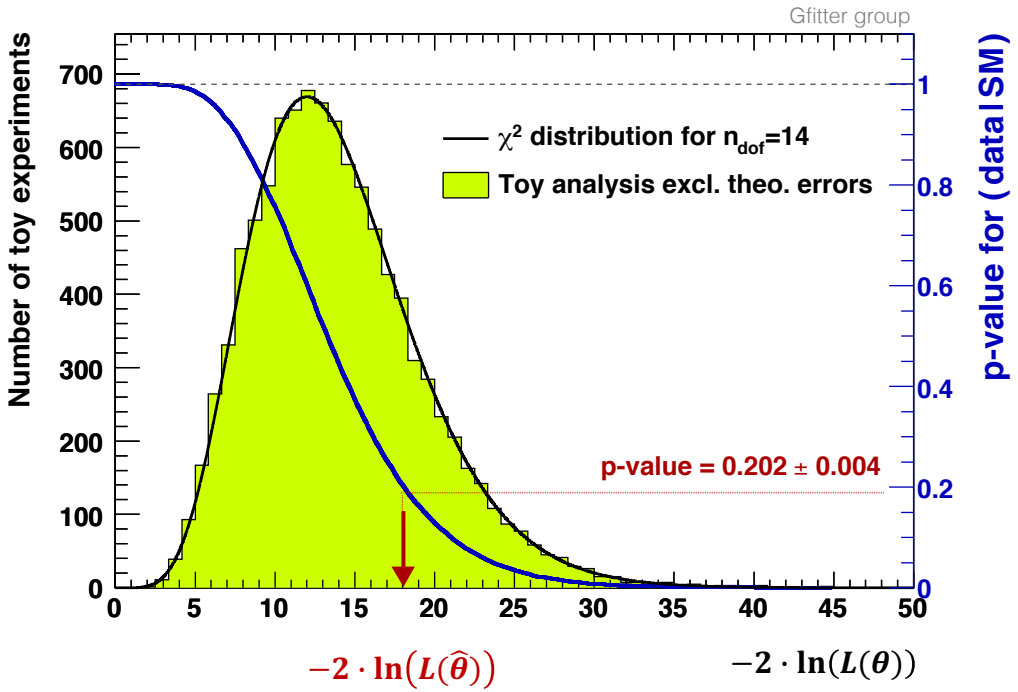
Goodness-of-Fit (GoF)

Maximum likelihood estimator determines the best parameter $\widehat{\boldsymbol{\theta}}$

But: does the model with the best $\hat{\theta}$ fit the data well ?

The value of $-2 \cdot \ln\left(L(\hat{\theta})\right)$ at minimum does not mean much → needs *calibration*

→ Determine the expected distribution of $-2 \cdot \ln\left(L(\hat{\theta})\right)$ using pseudo Monte Carlo events, and compare measured value to expected ones



Goodness-of-Fit (continued)

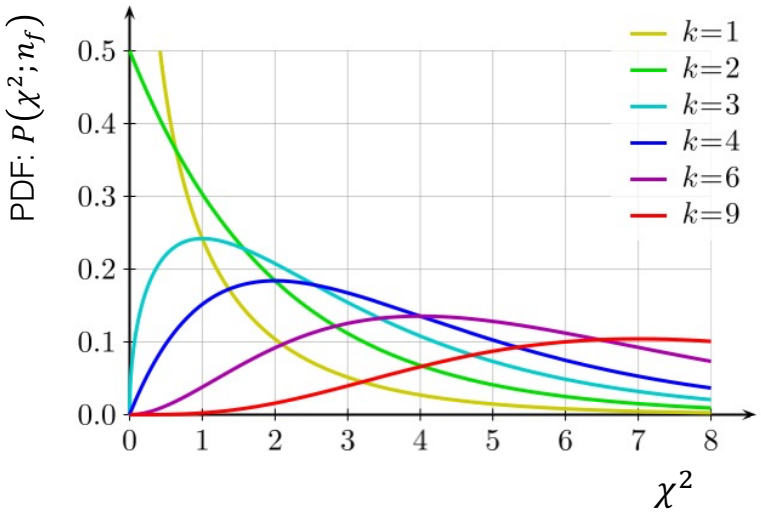
A Goodness-of-fit test is more straightforward with $\boldsymbol{\chi^2}$ estimator

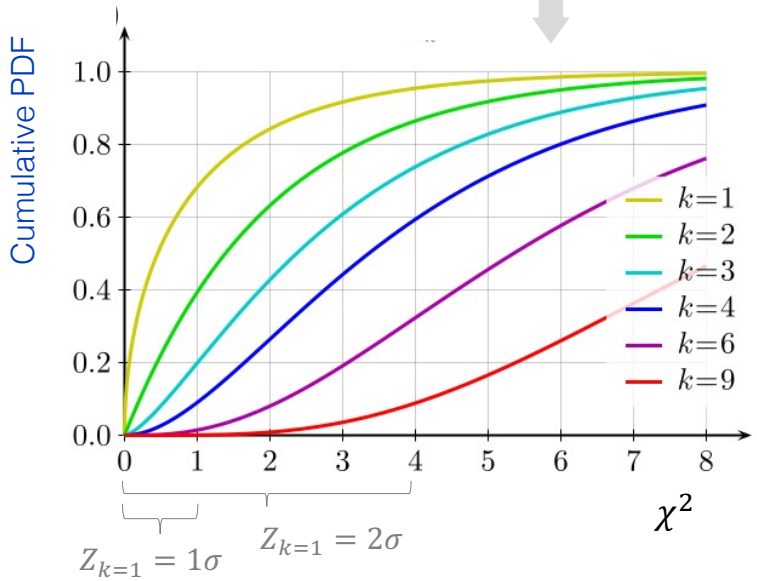
Let's use the binned example again. The task is to minimise versus θ:

$$\chi^2_{\min}(\hat{\theta}) = \min_\theta \left\{ \chi^2(\theta) = \sum_{i=1}^{n_{\text{bins}}} \left(\frac{(N_i - \mu_i(\theta))^2}{\sigma_i^2} \right) \right\}$$

χ^2 has known properties: $E[\chi^2] = n_{\text{d.o.f}} = k$ (= number of degrees of freedom)

Cumulative PDF: probability to find $\chi^2 > \chi^2_{\min}$: $\mathbf{TMath::Prob}(\chi^2_{\min}, k)$



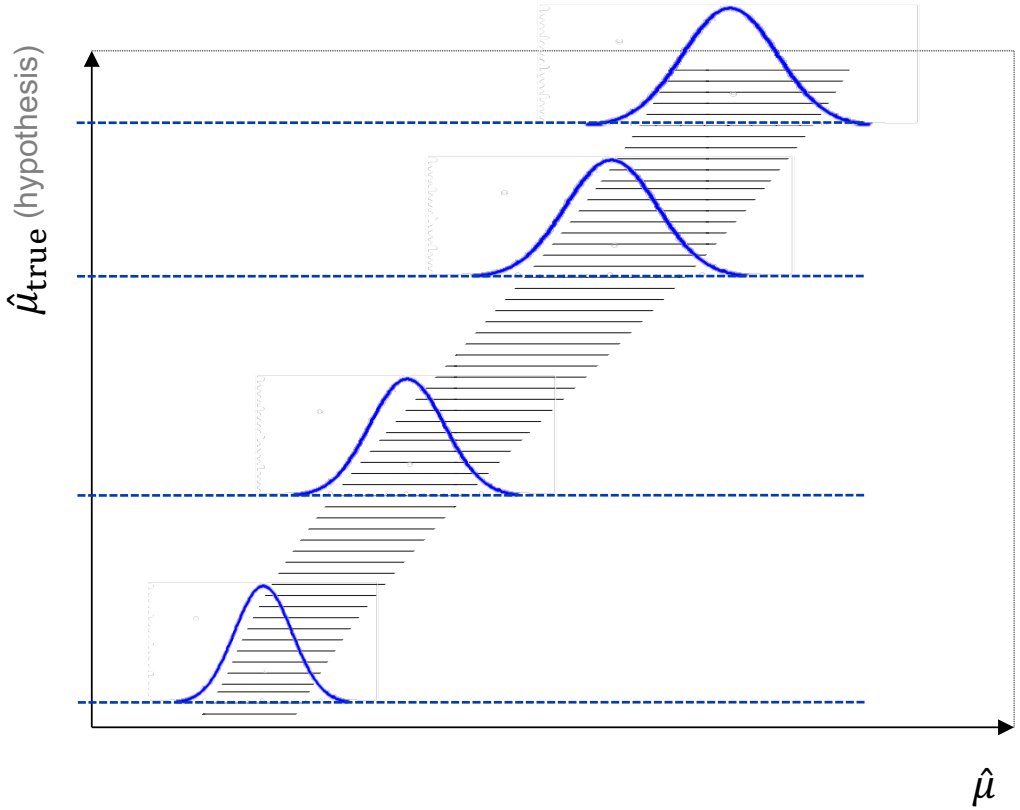


Figures from:
https://en.wikipedia.org/wiki/Chi-squared_distribution

Classical confidence level

Neyman *confidence belt* for confidence level (CL) $\boldsymbol{\alpha}$ (e.g. 95%)

Statement about probability to cover true value $\hat{\boldsymbol{\mu}}_{\mathbf{true}}$ of parameter $\hat{\boldsymbol{\mu}}$ fit to data

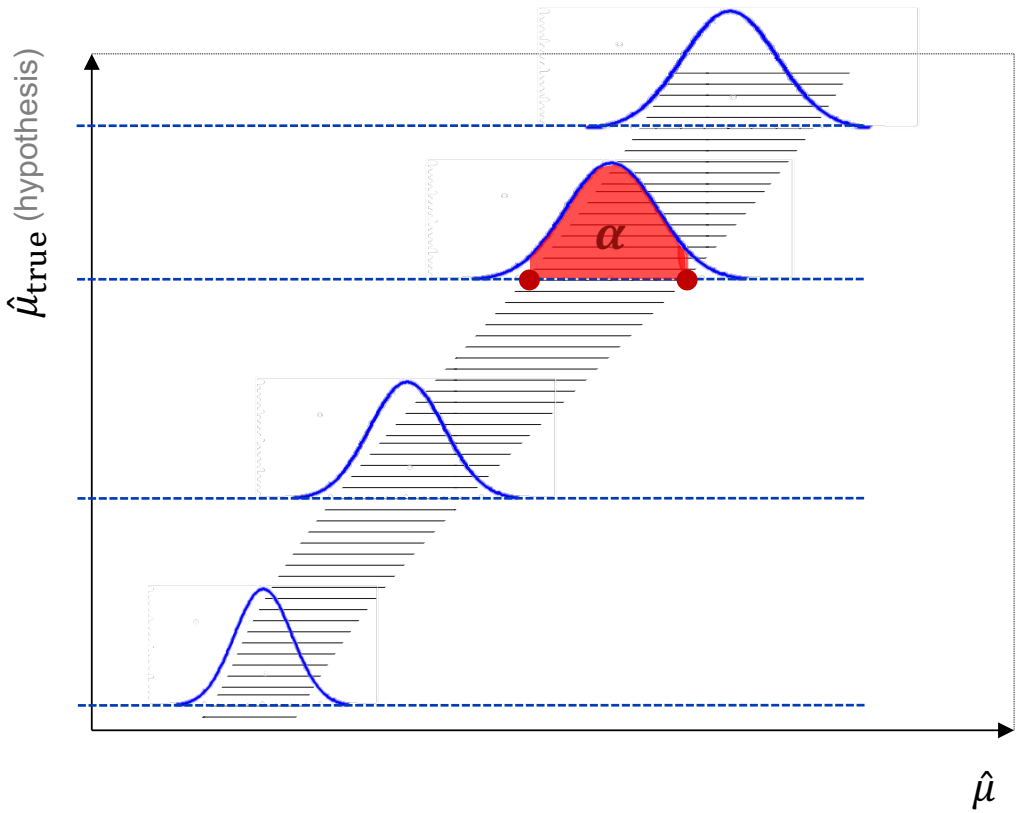


- Each hypothesis $\hat{\mu}_{\mathbf{true}}$ has a PDF of how the measured values $\hat{\mu}_{\mathbf{obs}}$ will be distributed

Classical confidence level

Neyman *confidence belt* for confidence level (CL) $\boldsymbol{\alpha}$ (e.g. 95%)

Statement about probability to cover true value $\hat{\boldsymbol{\mu}}_{\mathbf{true}}$ of parameter $\hat{\boldsymbol{\mu}}$ fit to data

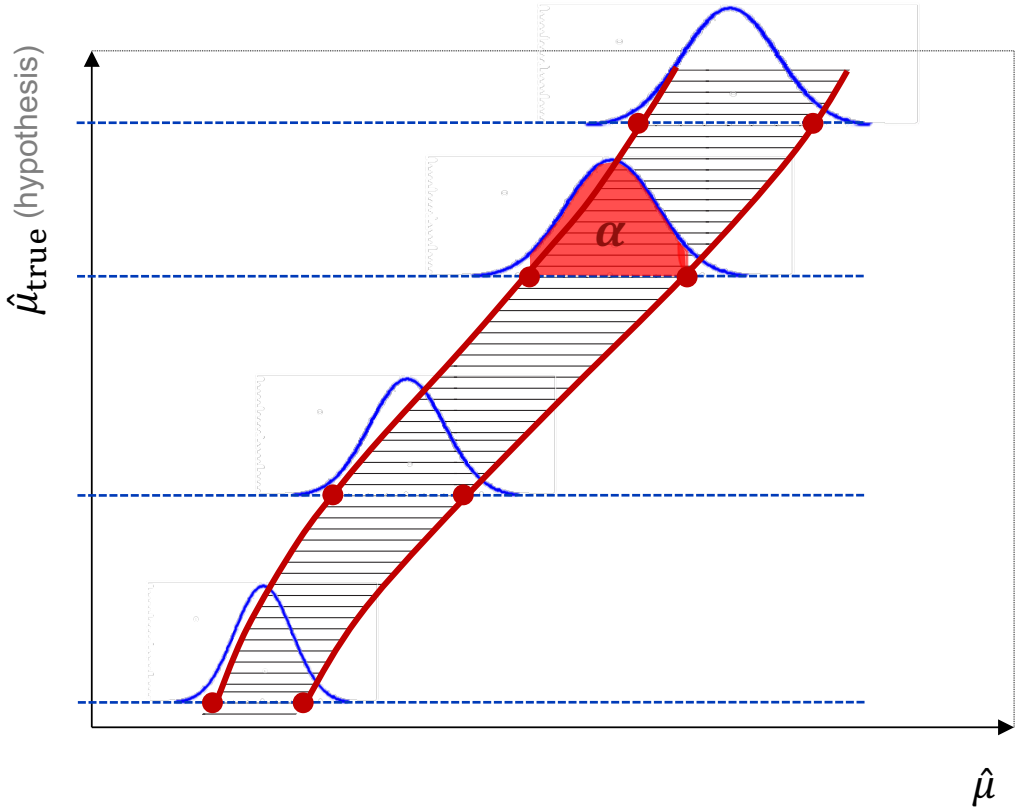


- Each hypothesis $\hat{\mu}_{\mathbf{true}}$ has a PDF of how the measured values $\hat{\mu}_{\mathbf{obs}}$ will be distributed
- Determine the (central) intervals ("acceptance region") in these PDFs such that they contain $\boldsymbol{\alpha}$

Classical confidence level

Neyman *confidence belt* for confidence level (CL) $\boldsymbol{\alpha}$ (e.g. 95%)

Statement about probability to cover true value $\hat{\boldsymbol{\mu}}_{\mathbf{true}}$ of parameter $\hat{\boldsymbol{\mu}}$ fit to data

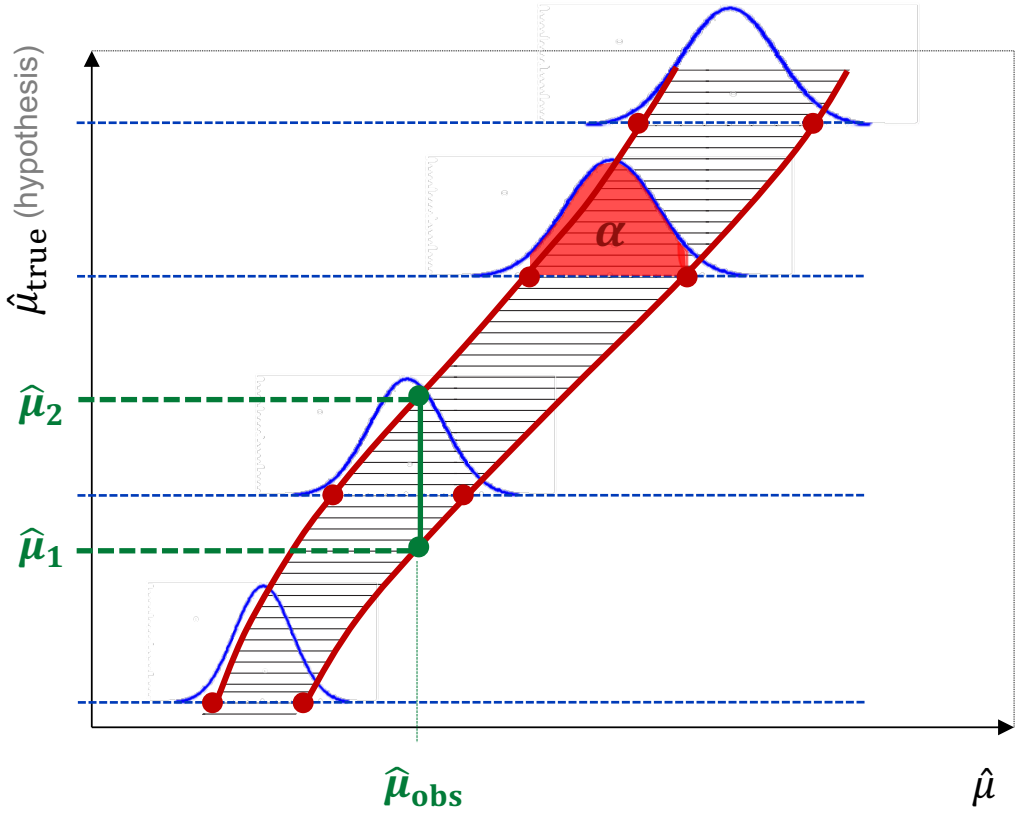


- Each hypothesis $\hat{\mu}_{\mathrm{true}}$ has a PDF of how the measured values $\hat{\mu}_{\mathrm{obs}}$ will be distributed
- Determine the (central) intervals ("acceptance region") in these PDFs such that they contain $\boldsymbol{\alpha}$
- Do this for all $\hat{\mu}_{\mathrm{true}}$ hypotheses
- Connect all the red dots: **confidence belt**

Classical confidence level

Neyman *confidence belt* for confidence level (CL) $\boldsymbol{\alpha}$ (e.g. 95%)

Statement about probability to cover true value $\hat{\boldsymbol{\mu}}_{\mathbf{true}}$ of parameter $\hat{\boldsymbol{\mu}}$ fit to data



- Each hypothesis $\hat{\mu}_{\mathrm{true}}$ has a PDF of how the measured values $\hat{\mu}_{\mathrm{obs}}$ will be distributed
- Determine the (central) intervals ("acceptance region") in these PDFs such that they contain $\boldsymbol{\alpha}$
- Do this for all $\hat{\mu}_{\mathrm{true}}$ hypotheses
- Connect all the red dots: **confidence belt**
- Measure $\hat{\boldsymbol{\mu}}_{\mathbf{obs}}$

→ Confidence interval $[\,\hat{\mu}_1, \hat{\mu}_2\,]$ given by **vertical** line intersecting the belt

→ α = 95% of the intervals $[\,\hat{\mu}_1, \hat{\mu}_2\,]$ contain $\hat{\mu}_{\mathrm{true}}$

Combining confidence intervals

The construction of Neyman intervals may involve large resources if done with pseudo Monte Carlo experiments. In many cases, experiments take "Gaussian" short cut, assuming that the $\mathbf{PDF}(\hat{\mu}_{\mathbf{true}})$ is Gaussian and does not depend on $\hat{\mu}_{\mathbf{true}}$ (see previous slides)

In Gaussian case, measurements can be combined by multiplying their likelihood functions

Otherwise: it is important to combine individual measurements, not the confidence intervals: construct confidence belt of combined measurement

The following "Gaussian shortcut" will be wrong in that case:

SME coefficient determined in [8] and $(CL)_{\bar{\nu}}$ the 99.7% C.L. upper limit determined here. We combine the two limits as

$$1/(CL)^2 = 1/(CL)_{\nu}^2 + 1/(CL)_{\bar{\nu}}^2,$$

where (CL) is the combined 99.7% C.L. upper limit. The most sensitive upper limits we have determined with the MINOS neutrino and antineutrino data are given in Table IV. As discussed, the way we determine the upper lim-

arXiv:1201.2631v2

In a perfectly Gaussian and uncorrelated case, this simple formula is correct

Combining confidence intervals

The construction of Neyman intervals may involve large resources if done with pseudo Monte Carlo experiments. In many cases, experiments take “Gaussian” short cut, assuming that the $\mathrm{PDF}(\hat{\mu}_{\mathrm{true}})$ is Gaussian and does not depend on $\hat{\mu}_{\mathrm{true}}$ (see previous slides)

In Gaussian case, measurements can be combined by multiplying their likelihood functions

Otherwise: it is important to combine individual measurements, not the confidence intervals: construct confidence belt of combined measurement

The following “Gaussian shortcut” will be wrong in that case:

SME coefficient determined in [8] and $(CL)_{\bar{\nu}}$ the 99.7% C.L. upper limit determined here. We combine the two limits as

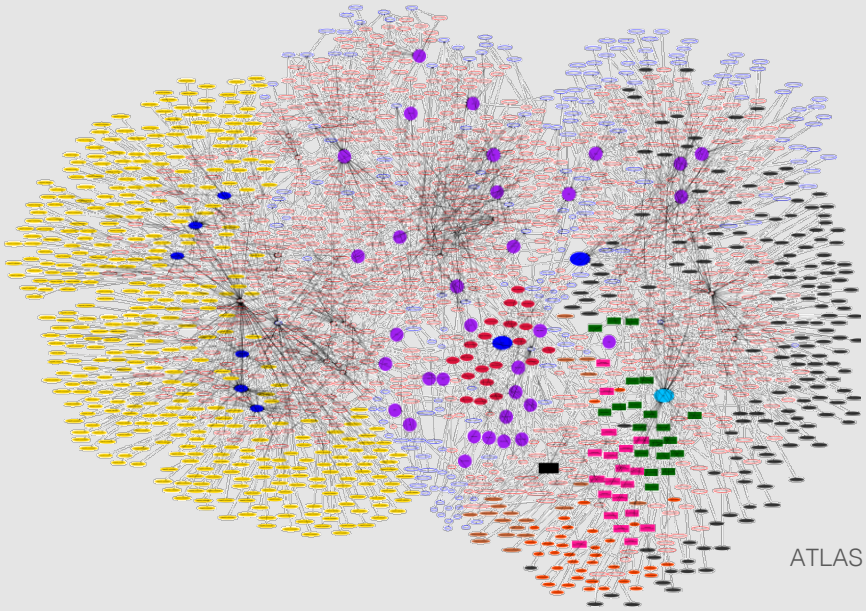
$$1/(CL)^2 = 1/(CL)_{\nu}^2 + 1/(CL)_{\bar{\nu}}^2,$$

where (CL) is the combined 99.7% C.L. upper limit. The most sensitive upper limits we have determined with the MINOS neutrino and antineutrino data are given in Table IV. As discussed, the way we determine the upper lim-

arXiv:1201.2631v2

In a perfectly Gaussian and uncorrelated case, this simple formula is correct

Maximum likelihood fits



ATLAS $H \to \gamma\gamma$ likelihood model used in fit

Likelihood functions

The *likelihood function* for a simple counting experiment is given by the Poisson PDFs:

$$L(\text{data}(N_{\text{obs}})|\mu) = \frac{(\mu S + B)^{N_{\text{obs}}}}{N_{\text{obs}}!} \cdot e^{-(\mu S+B)} \text{ ,}$$

where:

N_{obs}	observed number of events
S	expected number of signal events
B	expected number of background events
μ	“signal strength” modifier

In an unbinned case, the relevant likelihood function for N_{events} events reads:

$$L(\text{data}|\mu) = e^{-(\mu S+B)} \cdot \prod_{i=1}^{N_{\text{events}}} \left(\mu S \cdot p_s(x_i) + B \cdot p_b(x_i)\right)$$

where $p_s(x_i)$ and $p_b(x_i)$ are the values of the signal and background PDFs for the variable x_i

Likelihood functions with nuisance parameters

If the background prediction is subject to an uncertainty, one adds a *nuisance parameter* $\boldsymbol{\theta}$:

$$L(N_{\text{obs}}, \mu, \theta) = \frac{(\mu S + \theta B)^{N_{\text{obs}}}}{N_{\text{obs}}!} e^{-(\mu S + \theta B)} \cdot \text{Gauss}(\theta - 1, \sigma_\theta)$$

which is (in this example) constrained to $\boldsymbol{\theta} = \mathbf{1}$ within $\boldsymbol{\sigma_\theta}$ by a Gaussian PDF

The profile likelihood function is maximised with respect to both μ and θ

In realistic use cases, $L(N_{\text{obs}}, \mu, \theta)$ can be more complex:

- Both signal and background predictions are subject to multiple uncertainties parametrised by a set of m nuisance parameters $\theta = \{\theta_1, \ldots, \theta_m\}$
- There are several distinct signal and background contributions
- Several signal and background *control regions* are simultaneously fit
- The parameter of interests may not only be event abundances, but also signal properties
- The likelihood may be split into categories with different subpopulations of events with common and non-common parameters

One-sided test statistics

To compare the compatibility of the data with the background-only and signal+background hypotheses, where the signal is allowed to be scaled by some factor μ, we construct the following test statistic based on the profile likelihood ratio:

$$\tilde{q}_\mu = -2 \cdot \ln \frac{L\left(\text{data}|\mu, \hat{\theta}_\mu\right)}{L\left(\text{data}|\hat{\mu}, \hat{\theta}\right)}, \qquad 0 < \hat{\mu} < \mu$$

(Condition enforces one-sided confidence intervals for discovery and upper limit tests)

where nominator and denominator are independently maximised.

$\hat{\theta}_\mu$ is the *conditional* maximum given the signal strength modifier value μ

$\hat{\mu}, \hat{\theta}$ are the values corresponding to the global maximum of the likelihood

Remarks:

- Large $\tilde{q}_\mu$ values correspond to disagreement between data and hypothesis μ.
- $\tilde{q}_\mu$ behaves as χ^2 for large data samples and Gaussian θ parameters
- Note that the denominator in $\tilde{q}_\mu$ is independent of μ and only a normalisation term

Frequentist limit setting procedure

See: ATLAS & CMS https://cds.cern.ch/record/1375842

1. Construct likelihood function $L(\mu, \theta)$

2. Construct test statistics $\tilde{q}_\mu$

3. Perform fits on data and determine observed $\tilde{q}_{\mu,\mathrm{obs}}$ and $\hat{\theta}_{\mu,\mathrm{obs}}$ for hypothesis μ

4. Generate pseudo Monte Carlo events to construct the PDF $p_\mu(\tilde{q}_\mu|\mu, \hat{\theta}_{\mu,\mathrm{obs}})$ of $\tilde{q}_\mu$
   (for hypothesis μ, and where $\hat{\theta}_{\mu,\mathrm{obs}}$ is the set of conditional nuisance parameters found in fit to data).
   The nuisance parameters are fixed to $\hat{\theta}_{\mu,\mathrm{obs}}$ for the MC generation, but allowed to float in the fits.
   In the asymptotic limit, $p_\mu(\tilde{q}_\mu|\mu, \theta)$ is independent of θ.

5. Determine the observed p-value for hypothesis μ: $P(\mu) = \int_{\tilde{q}_{\mu,\mathrm{obs}}}^{\infty} p_\mu(\tilde{q}_\mu|\mu, \hat{\theta}_{\mu,\mathrm{obs}})\, d\tilde{q}_\mu$

6. Perform “discovery” test by computing $P(\mu = 0)$

7. Find the 95% upper bound $\mu = \mu_{95,\mathrm{obs}}$ for which: $P(\mu) = 0.05$

In case of complex fits the pseudo-MC procedure can be very CPU intensive. For sufficiently large number of expected and observed events one therefore usually employs asymptotic formulas that are based on the identification: $\tilde{q}_\mu \approx \chi^2$

Cowan, Cranmer, Gross, Vitells: https://arxiv.org/abs/1007.1727

Frequentist limit setting procedure (continued)

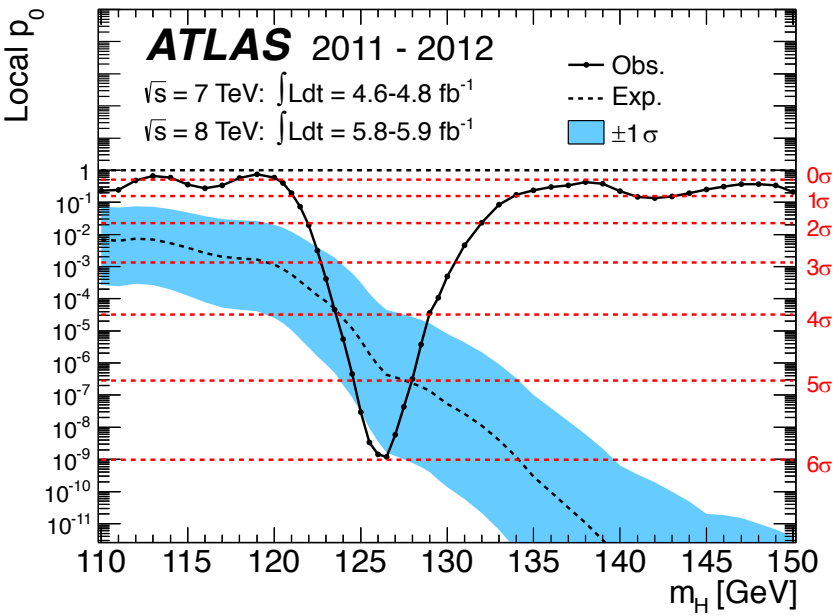
A. Read, https://cds.cern.ch/record/451614

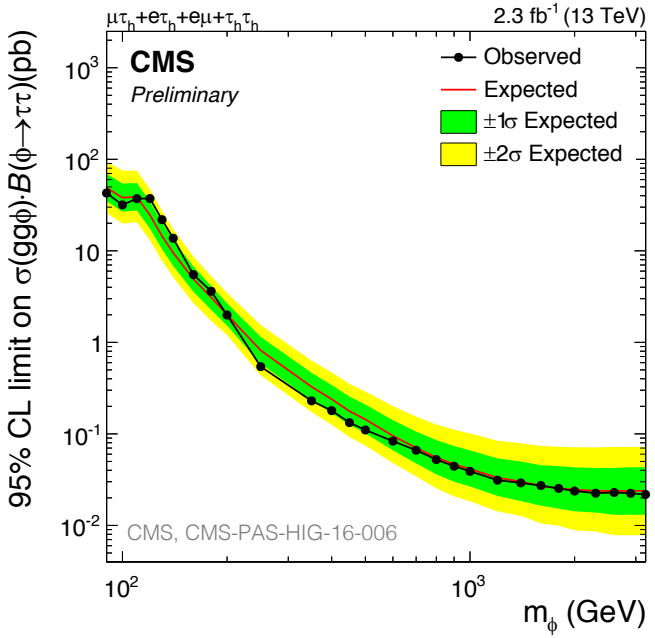
To be more conservative (to avoid that downward fluctuations of background contribute to the p-value), the LHC experiments compute upper limits using: $P_{\mathrm{CL}_s}(\mu) = P(\mu)/P(0) = 0.05$

- CL_s usually *over-covers*, ie, less than 5% of repeated experiments would lie outside the given bound
- A property of CL_s is that in case of $N_{\mathrm{obs}} = 0$, the resulting 95% CL upper limit is $\mu_{95,\mathrm{obs}} S \cong 3$, independent of the background expectation and the nuisance parameters

Let's get back to our earlier discovery and limit plots:

ATLAS, 1207.7214





Frequentist limit setting procedure (continued)

The underlying fits are often complex. On the right a graph of *only* the $H \to \gamma\gamma$ likelihood model:

The ATLAS & CMS Run-1 Higgs coupling combination analysis comprises a total of 4200 nuisance parameters ! (Of which a large fraction is of statistical nature)

ATLAS & CMS http://arxiv.org/abs/1606.02266

The tool of choice to perform such complex likelihood fits is **RooFit** (contained in ROOT)

https://root.cern.ch/roofit

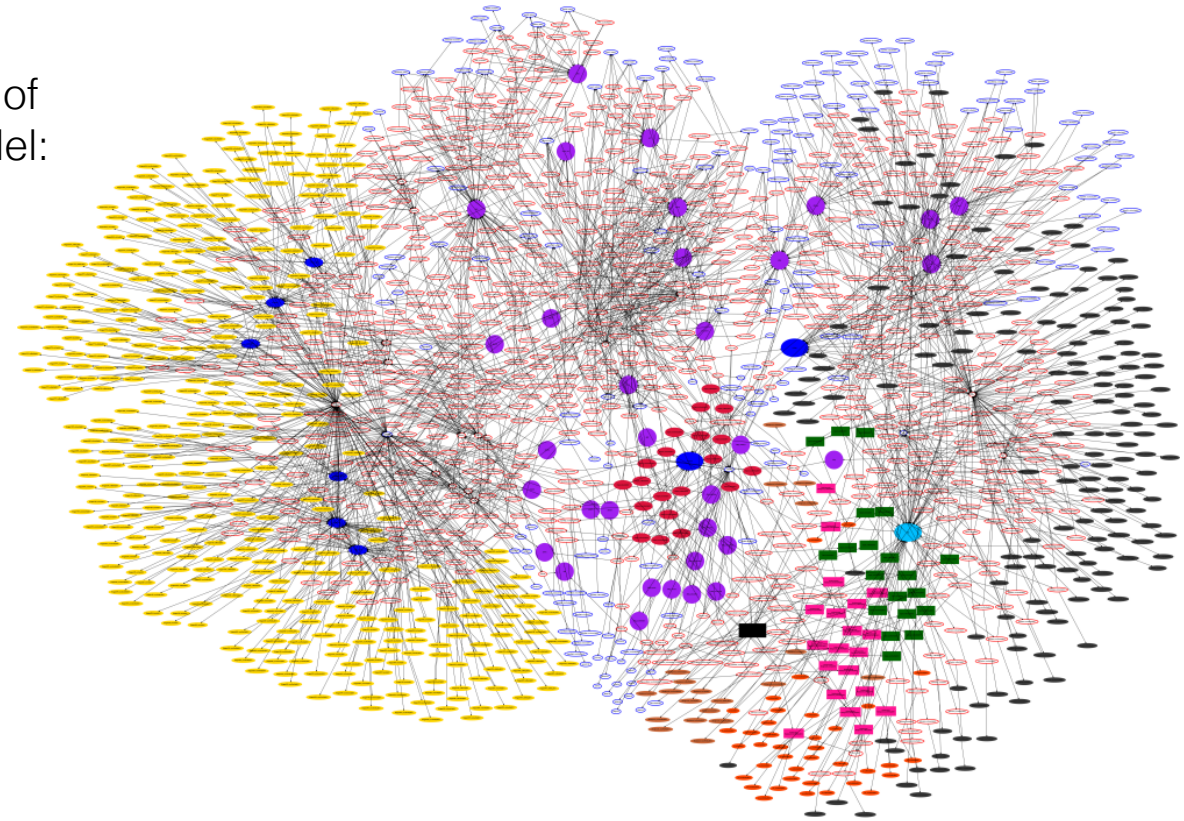


Figure caption: Each node represents either a numerical value, an expression or a PDF. The black box is the top- level PDF, the green boxes are the signal PDFs for each category, the pink boxes are the background PDFs. The bottom part of the graph describes the background: the brown ellipses are the background normalisation parameters, while the orange ellipses are the shape parameters. The dark red ellipses are the signal normalization expressions, and the blue ellipse in the center represents the μ parameter. The left part of the graph is devoted to the parameterization of SM signal yields: the gold ellipses are the coefficients of the parameterization, while the blue ellipses are per-mode μ parameters. The right side of the plot describes the signal shape: the dark gray boxes are the signal shape parameters, the blue ellipse represents m_H, and the cyan ellipse is $m_{\gamma\gamma}$. Finally, the purple ellipses represent the nuisance parameters associated with systematic uncertainties, the white boxes with blue outlines are the parameters describing the uncertainties. The red-lined boxes are expressions that bind the model together.

Nicolas Berger, Habilitation thesis, 2016

Why 5σ for a discovery ?

See also G. Cowan, https://arxiv.org/abs/1307.2487

As we have discussed yesterday, it is common practice in particle physics to regard an observed signal a “discovery” when its significance exceeds $Z = 5$, corresponding to a one-sided p-value of the background-only hypothesis of $2.9 \cdot 10^{-7}$

This is in contrast to many other fields (e.g., medicine, psychology) where a p-value of 5% (Z = 1.64) may be considered significant

Discoveries of new particles have been relatively frequent during the last ~20 years in the low-energy hadron spectra, but are very rare at high energy

Why 5σ for a discovery ?

See also G. Cowan, https://arxiv.org/abs/1307.2487

As we have discussed yesterday, it is common practice in particle physics to regard an observed signal a “discovery” when its significance exceeds $Z = 5$, corresponding to a one-sided p-value of the background-only hypothesis of $2.9 \cdot 10^{-7}$

This is in contrast to many other fields (e.g., medicine, psychology) where a p-value of 5% (Z = 1.64) may be considered significant

Discoveries of new particles have been relatively frequent during the last ~20 years in the low-energy hadron spectra, but are very rare at high energy

Certainly, from Bayesian reasoning: “*extraordinary claims require extraordinary evidence*”

A discovery (beyond the SM) will be a game changer that we do not want to have to unsay

Another reason for the high Z is the influence of non-statistical systematic uncertainties in some of our particle searches, which alter the properties of the p-value found

Finally, and importantly, the large look-elsewhere-effect (LEE) is a source of fluctuations. While it can be accounted for in a given analysis, the LEE is a global phenomenon that affects the entirety of the searches: the probability of seeing a fluctuation with local $Z = 5$ *anywhere* is much larger than $2.9 \cdot 10^{-7}$!

Why 5σ for a discovery ?

See also G. Cowan, https://arxiv.org/abs/1307.2487

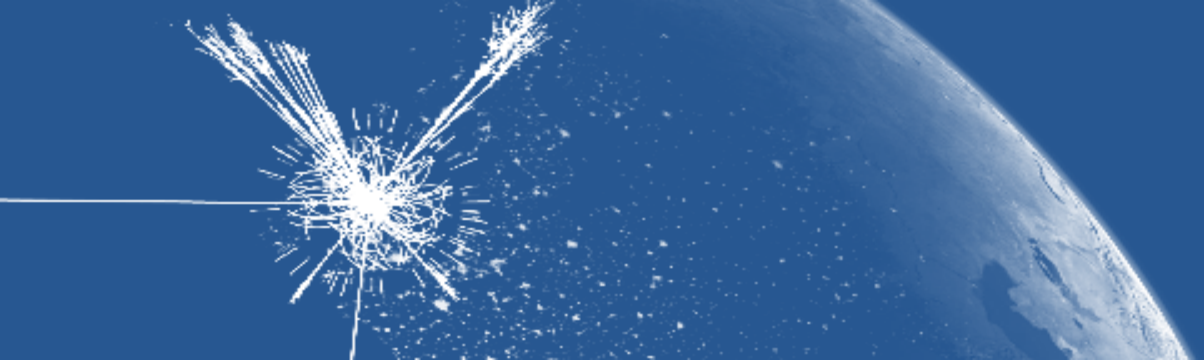
As we have discussed yesterday, it is common practice in particle physics to regard an observed signal a “discovery” when its significance exceeds $Z = 5$, corresponding to a one-sided p-value of the background-only hypothesis of $2.9 \cdot 10^{-7}$

This is in contrast to many other fields (e.g. medicine, psychology) where a p-value of 5%

Note: a discovery requires more than a “5σ” value. It needs the judgement of the scientist that the question asked and the experimental setup used are meaningful, that systematic uncertainties are under control, and that the analysis and interpretation were performed in an unbiased manner.

Another reason for the high Z is the influence of non-statistical systematic uncertainties in some of our particle searches, which alter the properties of the p-value found

Finally, and importantly, the large look-elsewhere-effect (LEE) is a source of fluctuations. While it can be accounted for in a given analysis, the LEE is a global phenomenon that affects the entirety of the searches: the probability of seeing a fluctuation with local $Z = 5$ *anywhere* is much larger than $2.9 \cdot 10^{-7}$!



Summary for today

Parameter estimation with the maximum likelihood technique, goodness-of-fit, and the derivation of a classical Neyman confidence belt were discussed

Maximum likelihood fits are powerful optimisation tools that allow for any required complexity

Next: Monte Carlo techniques and data unfolding