Data Analysis

Lecture 3: Confidence intervals

In this lecture

- Interval estimation
  - Errors on the fit parameters
- Confidence intervals

Reminder

- Example: histogram fitting

Physicists	Statisticians
1. Determining the "best fit" parameters of a curve	1. Point estimation
2. Determining the errors on the parameters	2. Confidence interval estimation
3. Judging the goodness of a fit	3. Goodness-of-fit (Hypothesis) testing

Adopted from [Baker, Cousins, 1984]

Confidence intervals[7]

For a Gaussian estimator the result of an experiment is usually expressed by

- The parameter's estimated value, plus/minus an estimate of the **standard deviation**, $\hat{\theta} \pm \sigma_{\hat{\theta}}$

If the pdf is not Gaussian, or in the presence of physical boundaries

- One usually quotes instead an **interval**.

The quoted interval or limit should:

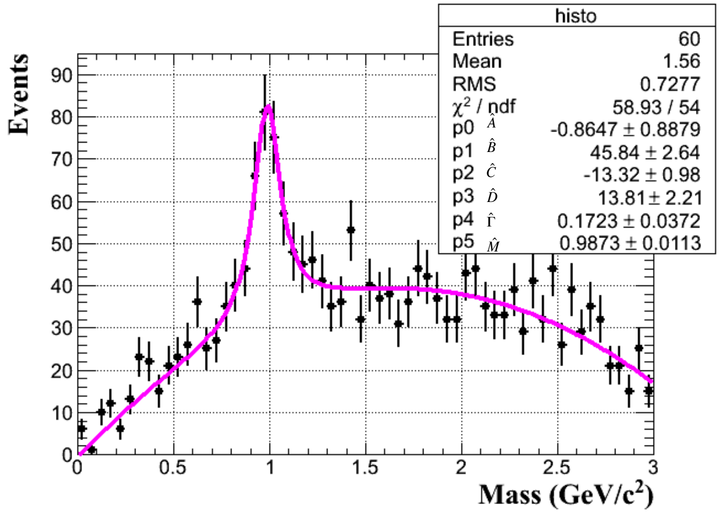
- Objectively communicate the result of the experiment,
- **Communicate incorporated prior beliefs** and relevant assumptions,
- Provide interval that covers the true value of the θ with specified probability,
- Make possible to draw conclusions about the parameter.

These goals are satisfied in **case of large data sample** by $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and in the multi-parameter case by

- The parameter estimates and covariance matrix.

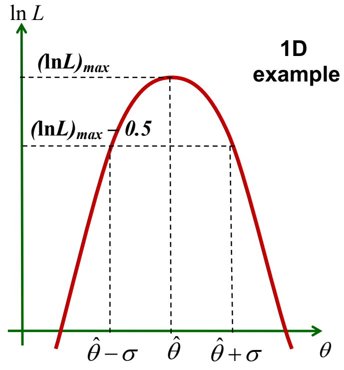
[7] Adapted from Particle Data Group.

Example: results of the fit



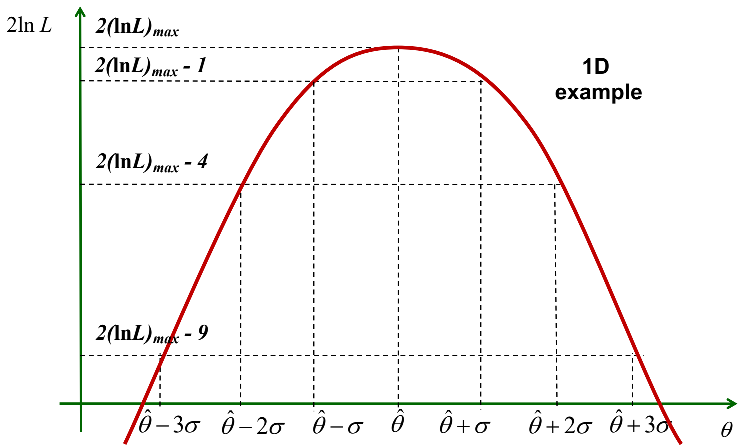
Errors on the ML estimates (1/4)

- How to obtain errors on the parameters estimated by the ML?
- Option 1: **Matrix inversion**
  - Covariance matrix is minus the inverse of the matrix of second derivatives $\mathrm{cov}^{-1}(\theta_i,\theta_j) = -\left.\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right|_{\theta=\hat{\theta}}$
  - Done with MINUIT/HESSE in ROOT
- Option 2: **Log – likelihood curve**
  - In the large N limits the likelihood function is Gaussian and the log-likelihood is parabola
  - By definition $(\ln L)_{max} = \ln L(\hat{\theta})$
  - $\pm 1\sigma$ limits on θ are those values of θ for which $\ln L$ falls by 0.5 from its maximum value L_{max}
  - For $\pm 2\sigma$ ($\pm 3\sigma$) limits $\ln L$ falls by 2 (4.5)
  - Done with MINUIT/MINOS in ROOT



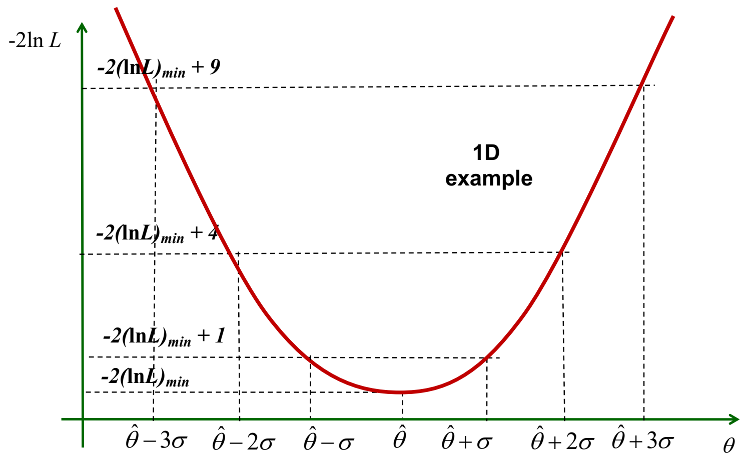
Errors on the ML estimates (2/4)

- The same, but now maximizing $2\ln L$



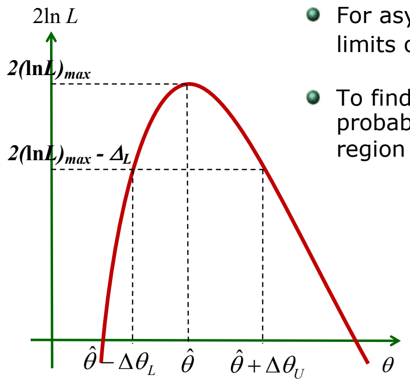
Maximisation → Minimisation

- In reality one is minimizing $-2\ln L$



Errors on the ML estimates (3/4)

- **Asymmetric example**
  - For finite samples and/or non-linear problems lnL is not necessarily parabolic nor symmetric
  - Confidence intervals can still be extracted from the lnL curve



1D example

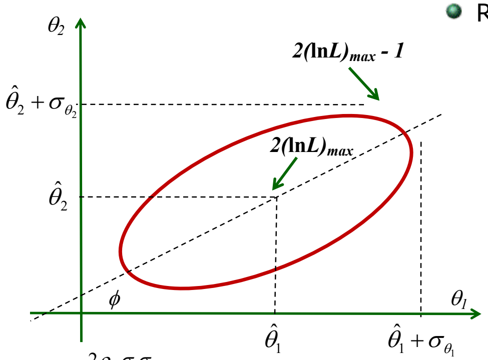
- For asymmetric lnL curve **upper** and **lower** limits on θ are not the same $\theta = \hat{\theta}^{+\Delta\theta_U}_{-\Delta\theta_L}$
- To find upper and lower limits with a certain probability content (β) of the confidence region → use Δ_L from the table:

Δ_L	β (%)
1	68.27
4	95.45
9	99.73

- ROOT uses Minuit/MINOS to extract limits (errors) in this way

Errors on the ML estimates (4/4)

- 2D example: **Standard error ellipse**
  - For more information see f.g. PDG
- This is so called the **plane tangent method**



$\tan 2\phi = \dfrac{2\rho_{12}\sigma_1\sigma_2}{\sigma_2^2 - \sigma_1^2}$

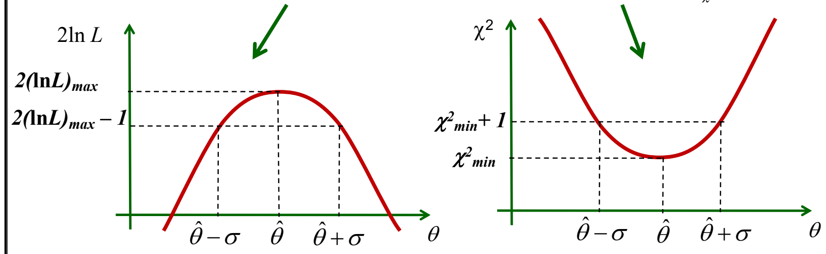
ρ_{12} is the correlation coeficient

- ROOT uses **Minuit/MINOS**
  - Works well also with non-regular iso-probability curves
  - Upper and lower limits for parameter θ_i are those values of θ_i for which
    $\max_{\theta_j, j\neq i}[2\ln L] = 2(\ln L)_{\max} - \Delta_L$
    with Δ_L from the table on the slide before
  - This is OK when interested in errors for only **one** parameter, regardless all others
  - Case of **simultaneous errors** estimate for more parameters → later in this lecture

Chi-square: Finding errors

- **Errors** (or limits) on parameters are found in the equivalent way as for the ML method
  - Matrix inversion
  - Shape of χ^2 arround it's minimum value

$$\text{Prob}(2\ln L) \geq 2\ln L_{\max} - \Delta_L \quad \Leftrightarrow \quad \text{Prob}(\chi^2) \leq \chi^2_{\min} + \Delta_{\chi^2}$$



Multiparameters errors

- When interested in simultaneous error estimation on more than one parameter, then the probability content (coverage probability) of the constant *-2lnL* or *χ2* contours is much smaller then in 1D case
- Example (recall 2D Gaussians probabilities):

	$\Delta_L / \Delta_{\chi^2}$	P_{1D}	P_{2D}
1σ	1	0.68	0.39
2σ	4	0.96	0.86

- Therefore, to increase the coverage probability we have to increase Δ_L or Δ_{χ^2} → see the values in the table (from PDG)

Table 32.2: $\Delta\chi^2$ or $2\Delta \ln L$ corresponding to a coverage probability $1-\alpha$ in the large data sample limit, for joint estimation of m parameters.

$(1-\alpha)$ (%)	$m=1$	$m=2$	$m=3$
68.27	1.00	2.30	3.53
90.	2.71	4.61	6.25
95.	3.84	5.99	7.82
95.45	4.00	6.18	8.03
99.	6.63	9.21	11.34
99.73	9.00	11.83	14.16

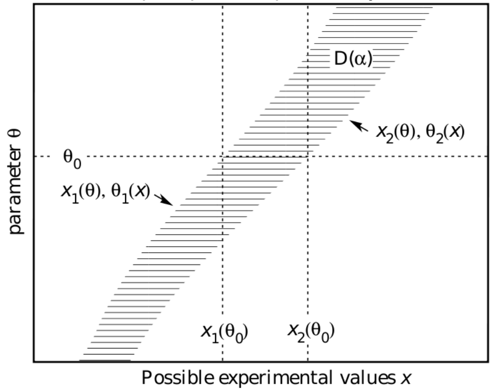
ROOT Tminuit::Contour draws contours of constant -2lnL or χ2 with a given probability coverage

Neyman Confidence Interval (1/2)

Using frequentist approach Neyman defines confidence interval to unknown parameter θ:

$$P(x_1 < x < x_2; \theta) = 1 - \alpha = \int_{x_1}^{x_2} f(x;\theta)dx,$$

where 1-α is pre-specified probability and x is measurement.

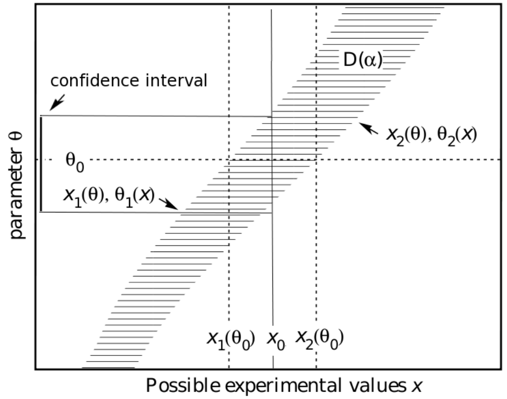


Union of $[x_1,x_2]$ segments for all values of θ is know as the **confidence belt**, D(α).

(Image courtesy of Particle Data Group.)

Neyman Confidence Interval (2/2)

Determining **confidence interval for** θ:



- Draw a vertical line through a measurement x_0.
- Inspect where the confidence belt is intercepted by the line.

Bayesian Confidence Intervals

In Bayesian statistics, all knowledge about parameter $\boldsymbol{\theta}$ is summarized by the posterior pdf $p(\boldsymbol{\theta}|\boldsymbol{x})$,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{x}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')d\boldsymbol{\theta}'}$$

which gives the degree of belief for $\boldsymbol{\theta}$ to have values in a certain region given the data $\boldsymbol{x}$.

- $\pi(\boldsymbol{\theta})$ is the prior pdf for $\boldsymbol{\theta}$, reflecting experimenter's subjective degree of belief about $\boldsymbol{\theta}$ before the measurement.
- $L(\boldsymbol{x}|\boldsymbol{\theta})$ is the likelihood function, i.e. the joint pdf for the data given a certain value of $\boldsymbol{\theta}$.
  - $L(\boldsymbol{x}|\boldsymbol{\theta})$ **should be published whenever possible**, to enable readers to calculate their own posterior pdf.
- The denominator simply normalizes the posterior pdf to unity.

Example: Non-negative constraint of a Poisson variable

Consider Poisson variable n which counts known background event with mean b, and unknown signal events with mean s constrained to be non-negative using the prior pdf

$$\pi(s) = \begin{cases} 0 & \text{if } s \leq 0 \\ 1 & \text{if } s > 0. \end{cases}$$

The likelihood function for Poisson distributed n is

$$L(n|s) = \frac{(s+b)^n}{n!}e^{-(s+b)}.$$

An upper limit s_+ at **credibility level** 1-α can be obtained by requiring

$$1 - \alpha = \int_{-\infty}^{s_+} p(s|n)ds = \frac{\int_{-\infty}^{s_+} L(n|s)\pi(s)ds}{\int_{-\infty}^{\infty} L(n|s)\pi(s)ds}$$

If $b = 0$ the equation reduces to the quantile of the χ^2 distribution

$$s_+ = \frac{1}{2}F_{\chi^2}^{-1}(1-\alpha; n_d),$$

where $n_d = 2(n+1)$ is the number of degrees of freedom.
(*e.g.* `0.5*TMath::ChisquareQuantile(0.95,2*(10+1))` ≈ 17.0.)

Uncertainty in physics

The sources of uncertainty in measurement[9]:

- **Incomplete definition** of the measurand; or its imperfect realization
- **Non-representative sampling**
- inadequate knowledge of the effects of environmental conditions; or imperfect measurements of these conditions
- **Personal bias** in reading instruments
- **Finite instrument resolution**
- Inexact values of measurement standards and reference materials
- **Inexact values of constants** and other parameters obtained from external sources and used in the data-reduction algorithm
- **Approximations and assumptions** incorporated in the measurement procedure
- **Variations of repeated observations** of the measurand under apparently identical conditions

[9] Adapted from the The International Organization for Standardization (ISO) Guide to the Expression of Uncertainty in Measurement.

Optimal presentation of search results

Optimal presentation of search results has some desired properties[10]:

- **Uncertainties due to systematic effects should be included in a clear and consistent way.**
  - Often it is useful to quote the statistical and systematical error separately, e.g. $\sigma = 45 \pm 4 \pm 1\ mb$.
- The result should summarize completely the experiment; so that no extra information should be required for further analysis.
- Results should be easily turned into probabilistic statements.
- Analysis should be transparent, and result should be stated in such a way that it cannot be misleading. The presentation of the result should not depend on the particular application.
- **If possible full pdf-distributions and even data sets can be attached into analysis results**.
- In **unified approach to data analysis**, the transitions between exclusion, observation, discovery, and measurement are kept as small as possible.

[10] Adapted from F. James, *Workshop on Confidence Limits*, CERN-2000-005, 2000.

Example: Propagation of errors/uncertainty

Your 16 GB MicroSD has package size (HxWxD) (mm) 15 x 11 x 1.
Manufacturer claims 0.1 mm deviation from the nominal dimensions.

Estimate you cards volume V = H W D and its uncertainty?

V = H W D = 15 mm 11 mm 1 mm = 165 mm^3.
Assuming uncorrelated variables we calculate

$$(\frac{\Delta V}{V})^2 = (\frac{\Delta H}{H})^2 + (\frac{\Delta W}{W})^2 + (\frac{\Delta D}{D})^2$$

$(\frac{\Delta V}{165})^2 = (\frac{0.1}{15})^2 + (\frac{0.1}{11})^2 + (\frac{\Delta 0.1}{1})^2 \approx 0.01.$

$V \pm \Delta V = 165 \pm 17\ mm^3$.

Note: Commonly the error on a quantity, here ΔV, is given as the standard deviation, σ. If the statistical probability distribution of the variable is assumed to be a normal distribution, there is a 68% probability that the true value of volume lies in the region $V \pm \Delta V$.

Example: MC estimation of errors

ADVANCED

If Taylor series aproximation (1) fails, errors can be estimated using Monte-Carlo simulation. A real life example from neutrino physics[11], where Data Handling

department at CERN had to estimate value of

$$R = \frac{a}{\frac{d}{ke}(b-c) - 2(1-\frac{k^2 d}{ke})a}$$

$a = 3.84 \pm 1.33$
$b = 74 \pm 4$
$c = 9.5 \pm 3$
$d = 0.112 \pm 0.009$
$e = 0.320 \pm 0.002$
$k = 0.89$

Using (1) one finds $\hat{R} = 0.191 \pm 0.073$. What is the probability P that R is different from theoretical value $R_{th} = 0.42$? If gaussian form is assumed

$$P = \int_{3.14}^{+\infty} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = 8 \times 10^{-4}$$

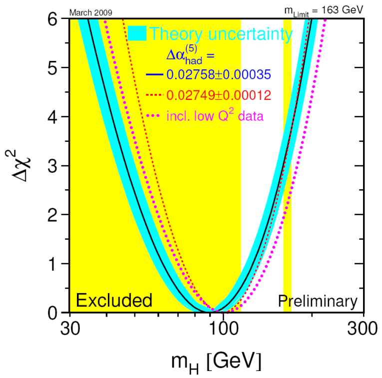
`P=ROOT::Math::gaussian_cdf_c(3.14)` as introduced in `ProbFuncMathCore.h`

$(R_{th} - \hat{R})/\sigma_R = (0.42 - 0.191)/0.073 = \mathbf{3.14}$

Monte-Carlo simulation gives a quite different result.

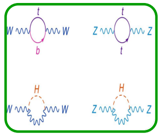
[11] From B. Escoubès, *Probabilités et statistiques à l'usage des physiciens*, Ellipses, 1998.

Example
Higgs boson mass costraints from Electroweak precision tests

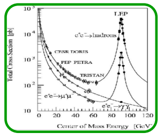


Method

Step 1 – Very precise measurements of SM
- Measure SM parameters extremely well
- **α, M_Z, G_F**
- μ lifetime, $(g-2)_e$, LEP …



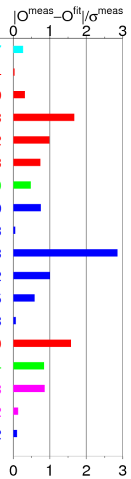
Step 2 – Predictions (assuming Higgs boson)
- Calculate quantum corrections to other observables
  - m_W, A_{LR}, $\sin^2\theta_W$ …
- Depending on **α, M_Z, G_F**, but also on **m_t, m_H …**



Step 3 – Precise electroweak measurements
- Measure very precisely observables from Step 2
- @ SLC, LEP, Tevatron …

Results from step 2 and 3

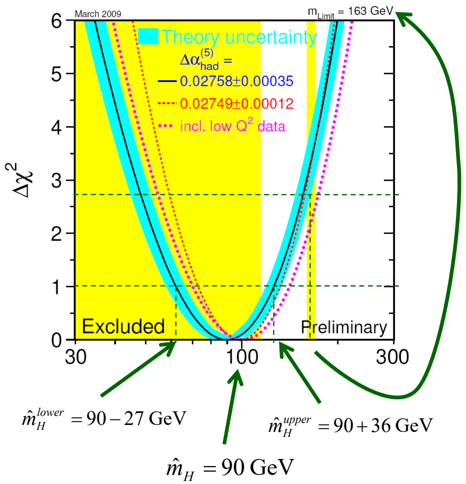
	Measurement	Fit
$\Delta\alpha_{had}^{(5)}(m_Z)$	0.02758 ± 0.00035	0.02767
m_Z [GeV]	91.1875 ± 0.0021	91.1874
Γ_Z [GeV]	2.4952 ± 0.0023	2.4959
σ_{had}^0 [nb]	41.540 ± 0.037	41.478
R_l	20.767 ± 0.025	20.742
$A_{fb}^{0,l}$	0.01714 ± 0.00095	0.01643
$A_l(P_\tau)$	0.1465 ± 0.0032	0.1480
R_b	0.21629 ± 0.00066	0.21579
R_c	0.1721 ± 0.0030	0.1723
$A_{fb}^{0,b}$	0.0992 ± 0.0016	0.1038
$A_{fb}^{0,c}$	0.0707 ± 0.0035	0.0742
A_b	0.923 ± 0.020	0.935
A_c	0.670 ± 0.027	0.668
A_l(SLD)	0.1513 ± 0.0021	0.1480
$\sin^2\theta_{eff}^{lept}(Q_{fb})$	0.2324 ± 0.0012	0.2314
m_W [GeV]	80.399 ± 0.025	80.378
Γ_W [GeV]	2.098 ± 0.048	2.092
m_t [GeV]	173.1 ± 1.3	173.2



March 2009

Adopted from http://lepewwg.web.cern.ch/LEPEWWG/

The best fit



- From the LEP Electroweak Working group:
  - *"The preferred value for its mass, corresponding to the minimum of the curve, is at 90 GeV, with an experimental uncertainty of +36 and -27 GeV (at 68 percent confidence level derived from Delta chi2 = 1 for the black line, thus not taking the theoretical uncertainty shown as the blue band into account)."*
  - *"The precision electroweak measurements tell us that the mass of the Standard-Model Higgs boson is lower than about 163 GeV (one-sided 95 percent confidence level upper limit derived from Delta chi2 = 2.7 for the blue band, thus including both the experimental and the theoretical uncertainty)."*

Mass of a new boson

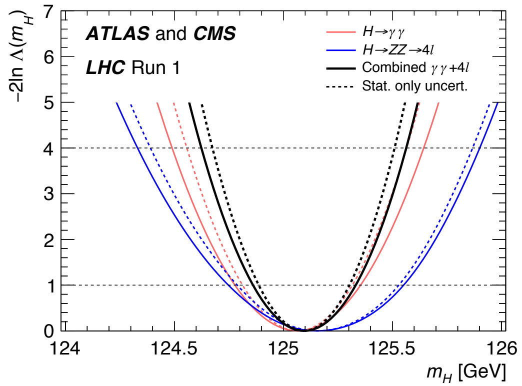


Figure 1: Scans of twice the negative log-likelihood ratio $-2\ln\Lambda(m_H)$ as functions of the Higgs boson mass m_H for the ATLAS and CMS combination of the $H \to \gamma\gamma$ (red), $H \to ZZ \to 4\ell$ (blue), and combined (black) channels. The dashed curves show the results accounting for statistical uncertainties only, with all nuisance parameters associated with systematic uncertainties fixed to their best-fit values. The 1 and 2 standard deviation limits are indicated by the intersections of the horizontal lines at 1 and 4, respectively, with the log-likelihood scan curves.

$$m_H = 125.09 \pm 0.24\ \text{GeV}$$
$$= 125.09 \pm 0.21\ (\text{stat.}) \pm 0.11\ (\text{syst.})\ \text{GeV}.$$