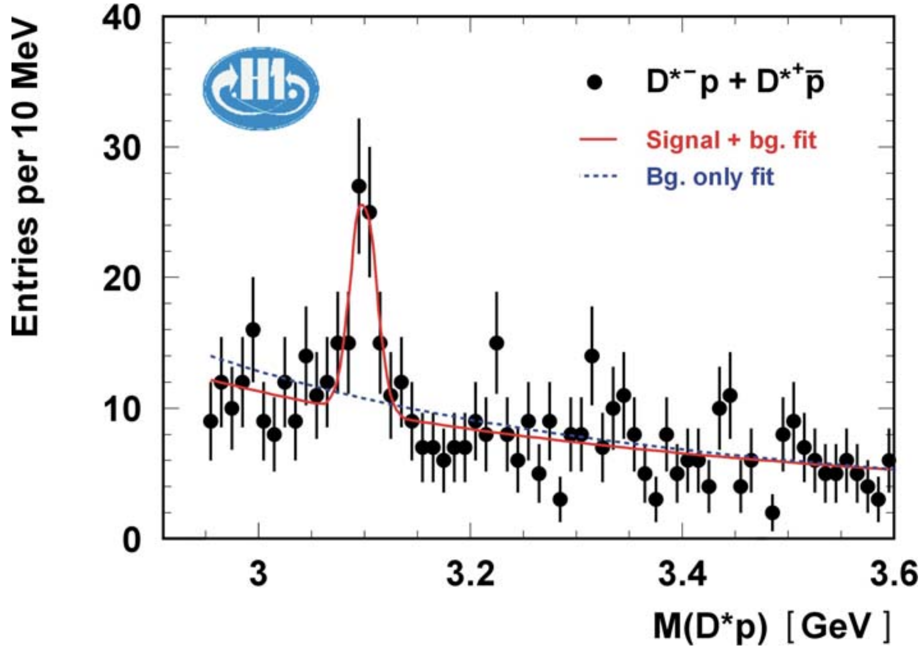
χ^2, Hypothesis Testing and Significance

Last time:
- Estimators and estimation
- (Log) likelihood fitting

Today:
- χ^2 Fitting
- χ^2 Tests
- Confidence and Significance



'We warn the reader that there is no Universal convention for the Term `Confidence Level' (Review of Particle Properties, PDG)

Reminder: Maximum Likelihood

• The **likelihood of a** is just the combined probability to obtain your data points if parameter to be estimated has the given value a:

$$L(x_1, x_2, \ldots x_N; a) = P(x_1; a)P(x_2; a)\ldots.P(x_N; a) = \prod_{i=1}^{N} P(x_i; a)$$

• Max (log) likelihood estimator $\hat{a}$ makes this probability maximal…

• Max (log) likelihood is a nice general prescription for obtaining a `best estimate' of a parameter from any N data points which depend on ("are sensitive to") that parameter.

Important: `weighted means' of multiple measurements x_i of the same quantity with different errors σ_i

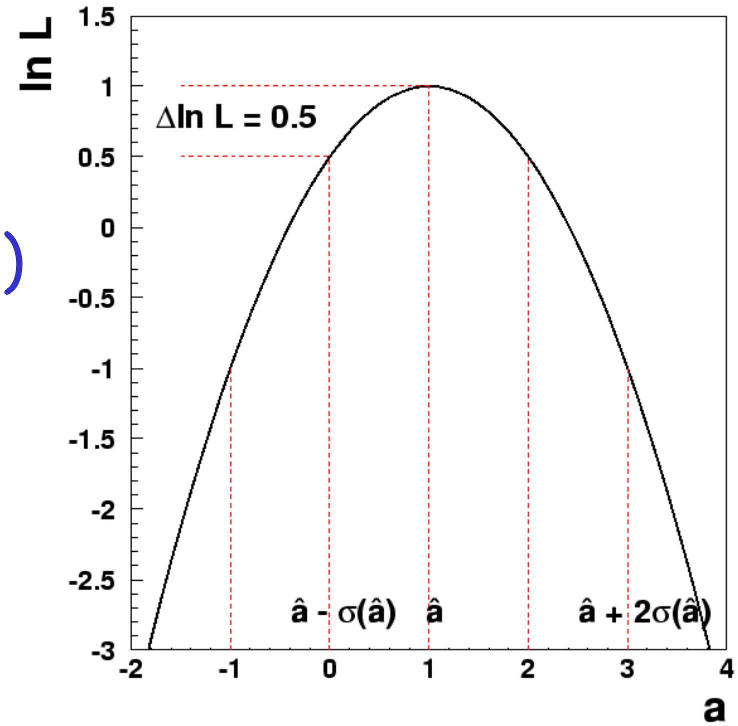
$$\hat{\mu} = \frac{\sum_i (x_i / \sigma_i^2)}{\sum_i (1/\sigma_i^2)}$$

$$\hat{\sigma}^2 = 1 \Big/ \sum_i (1/\sigma_i^2)$$

Some Last Notes on Maximum Likelihood

• Central Limit Theorem: like all probability distributions, L becomes Gaussian for large N (so ln L quadratic)

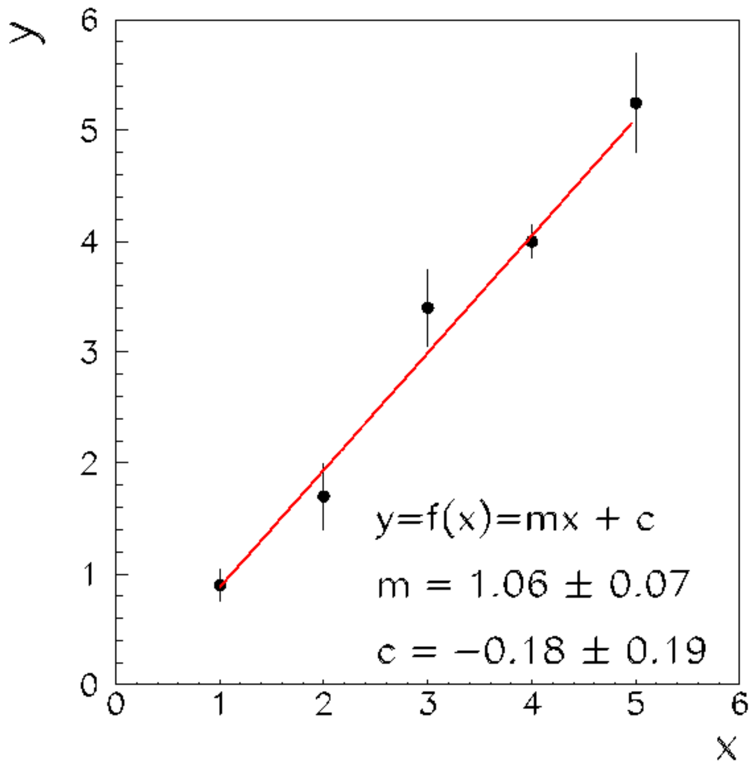
• … 1σ error on $\hat{a}$ given by value of a which reduces ln L by 0.5 (ln L reduces by 2.0 for 2σ etc)



Can use for comparisons:… express $\Delta \ln L_{max}$ in terms of number of σ by which one theory / hypothesis is better than another as a fit to the same data.

• Main drawback:… likelihood doesn't tell us how well theory fits data overall … <u>absolute value of ln L not easily interpreted.</u>

Least Squares and χ^2



Very common situation …

- Measurements y_i (eg differential cross section) with errors σ_i at lots of known points x_i.
- A theory gives $y=f(x;a)$ depending on (unknown) parameter a
- Want to extract a from the data.
- If errors on data points Gaussian:

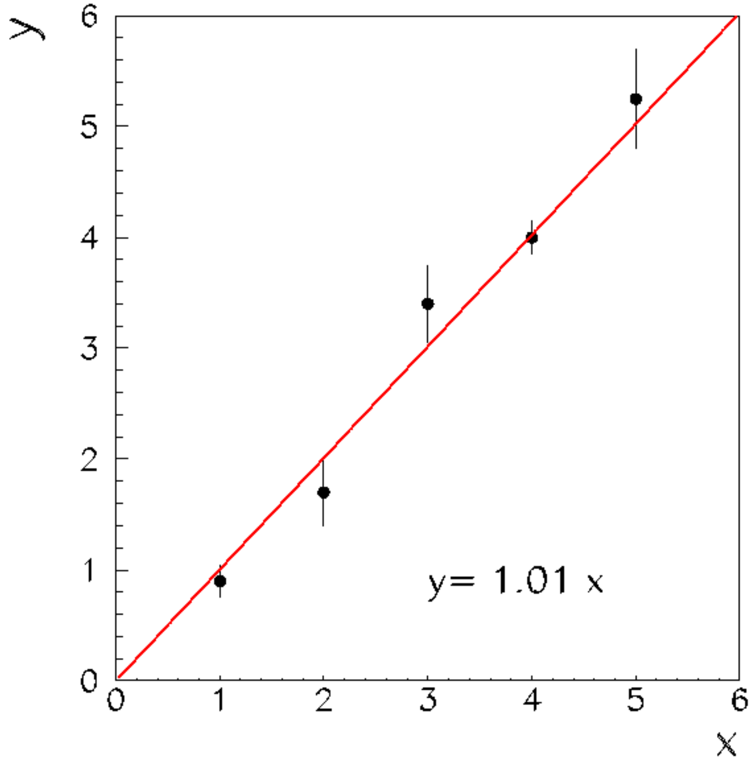
$$P(y_i;a) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-[y_i - f(x_i;a)]^2/2\sigma_i^2}$$

… so log likelihood is $\ln L = -\frac{1}{2}\sum_i \left(\frac{y_i - f(x_i;a)}{\sigma_i} \right)^2 - \sum_i \ln \sigma_i \sqrt{2\pi}$

… and maximising likelihood is equivalent to minimising χ^2 with respect to a …

$$\chi^2 = \sum_i \left(\frac{y_i - f(x_i;a)}{\sigma_i} \right)^2$$

Simplest Example: Straight Line Fit



For simplicity, suppose line must go through origin: y=f(x)=mx

$$\chi^2 = \sum_i \frac{(y_i - mx_i)^2}{\sigma_i^2}$$

Minimise with respect to m....

$$\frac{d\chi^2}{dm} = \sum_i -2x_i \frac{(y_i - mx_i)}{\sigma_i^2} = 0$$

$$m = \frac{\sum_i x_i y_i / \sigma_i^2}{\sum_i x_i^2 / \sigma_i^2}$$

- Simple result because f(y) linear in m (doesn't have to be linear in x) !

- Rapidly gets out of hand if non-linear in m …numerical (iterative) solutions using computers. e.g. MINUIT

Some Remarks on χ^2

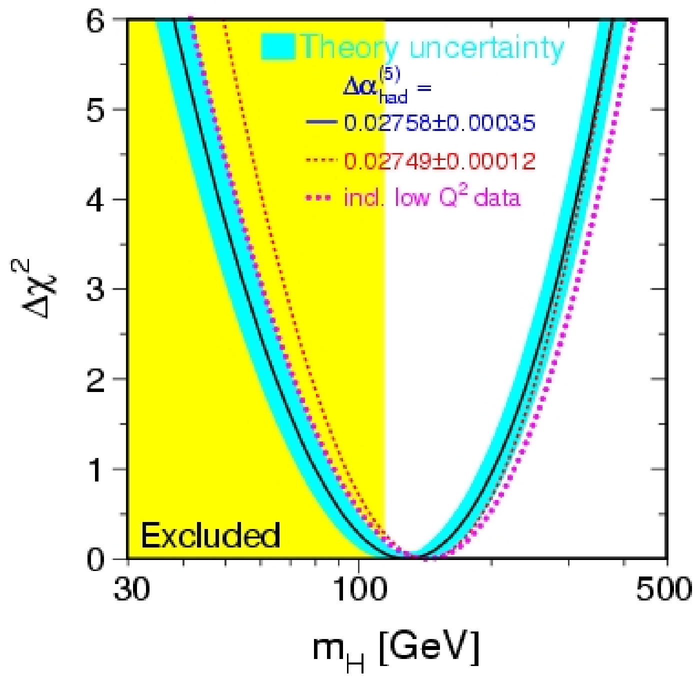
- By definion of $$\chi^2 = \sum_i \frac{(y_i - f(x_i))^2}{\sigma_i^2}$$ expect ~1 per data point.
- More precisely, expect χ^2 ~1 per degree of freedom (dof)

$$N_{dof} = N_{\text{data points}} - N_{\text{fit parameters}}$$

e.g. if we fitted a Gaussian, there were 3 parameters

- χ^2 / dof provides a figure of merit for how well theory describes data

- $\Delta\chi^2=1$ is equivalent to moving 1σ away from the best fit... ($\Delta\chi^2=n^2$ for $n\sigma$)
- e.g. Higgs mass of 250 GeV is about $\Delta\chi^2=4$ (2σ) away from best fit of world data to the Standard Model



Assessing Results: Hypothesis Testing

- Often interested in a question with a yes / no answer ...
  e.g. Does y increase with x?
  Do parameters a and b have the same value?
  Are data well described by a Poisson distribution?
  Are the data well described by the function y=3x?

- Make a **hypothesis** and test it statistically.

- Use χ^2 dist^n or other test to accept / reject hypothesis

... Well defined procedure ...
... Resulting statements always based on probabilities
e.g. probability to get a more extreme result or
`**confidence level**'

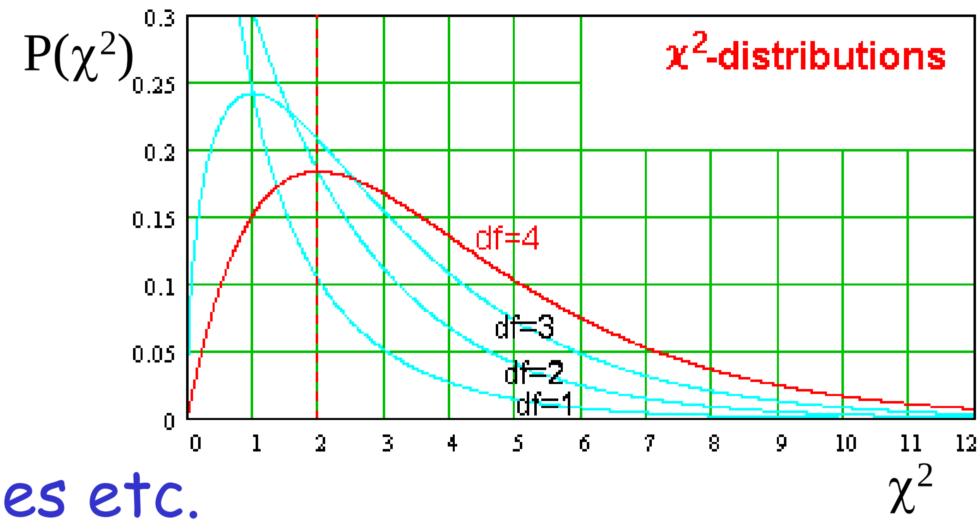
χ^2 Testing

- Again, data points y_i with errors σ_i at known x_i.
- Theorist tells us that the data are described by his new theory, which specifies $y=f(x)$ with no free parameters.

- Construct $\chi^2 = \sum_i \frac{(y_i - f(x_i))^2}{\sigma_i^2}$ 1 degree of freedom per point.

- If χ^2 / dof >> 1, theorist not as brainy as we thought
- If χ^2 / dof << 1, theorist got a preview and fiddled theory?

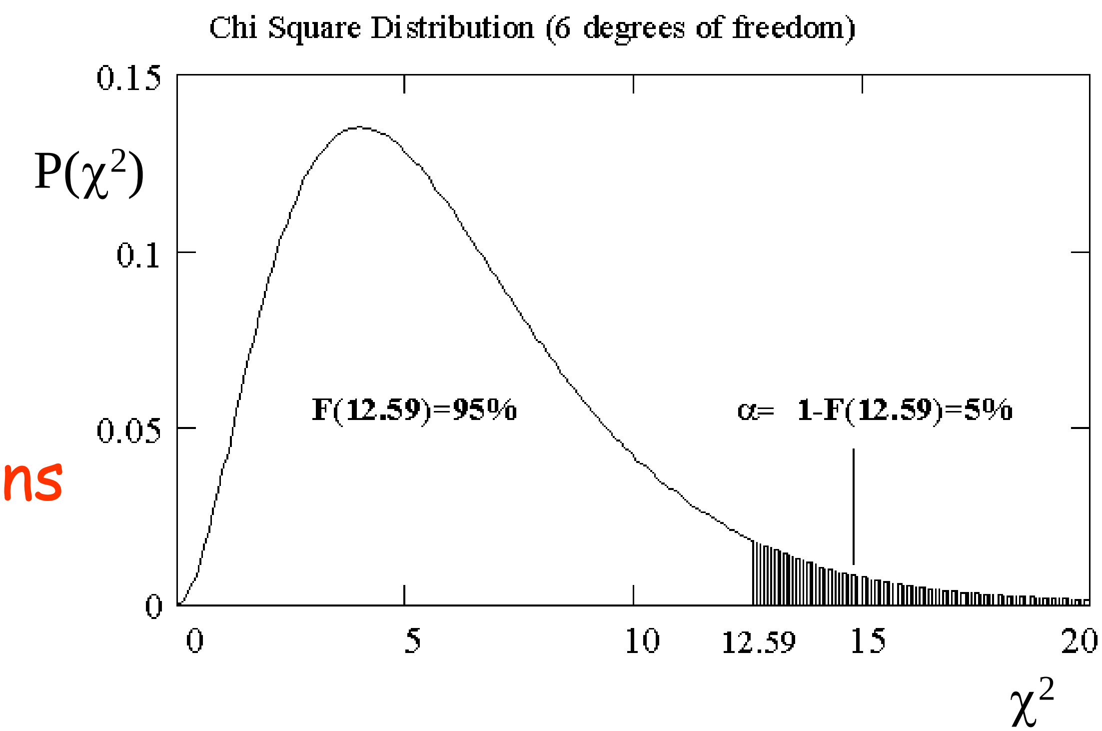
In between, need to use χ^2 / dof distribution to quantify probability.



… in practice, χ^2 / dof looked up in tables / CERNLIB routines etc.

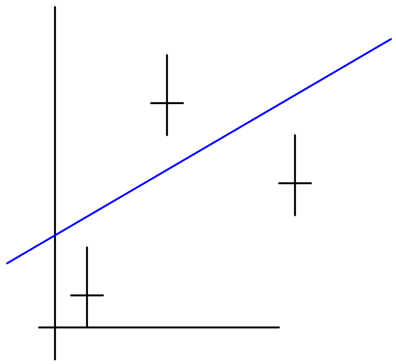
Illustration of χ^2 Testing

- Often 95% (or 90%) Confidence Level quoted
- i.e. if theory is correct, 95% of experiments will yield $\chi^2 < \chi^2$ (95% CL)
- For 6 data points, happens at χ^2 (95% CL) = 12.59

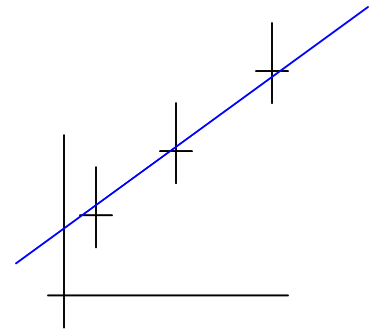


- If we get result of χ^2 = 16.8, reject hypothesis that "theory is correct" at the chosen 95% Confidence Level.
- Quote probability $P(\chi^2 > 16.8) = 0.01$ … i.e. if the theory is true, only 1% of experiments would give a more extreme χ^2

χ^2 Results and their Interpretation



$\chi^2 / \text{dof} > 1$



$\chi^2 / \text{dof} < 1$

Large χ^2 may come from …

1. Bad Measurements
2. Bad Theory
3. Underestimated errors
4. Bad luck
5. Ignoring systematic errors

Small χ^2 may comes from …

1. Overestimated errors
2. Good luck
3. Ignoring correlations in systematic errors

Testing for New Physics: The Null Hypothesis

- In last example, rejected hypothesis `theory is correct'.

- Can never meaningfully accept a hypothesis, only reject one.

- e.g. Suppose experiment tests for extra-sensory perception … a candidate scores 99/100 when random result is 10/100.

- Possible statement: "Result is consistent with the existence of ESP"

…. So what? It's also consistent with the existence of a Higgs boson with a mass 121 GeV, but it doesn't prove it!

- Better statement: "The result is inconsistent with the null hypothesis that there is no such thing as ESP at the 99.99…% Confidence Level"