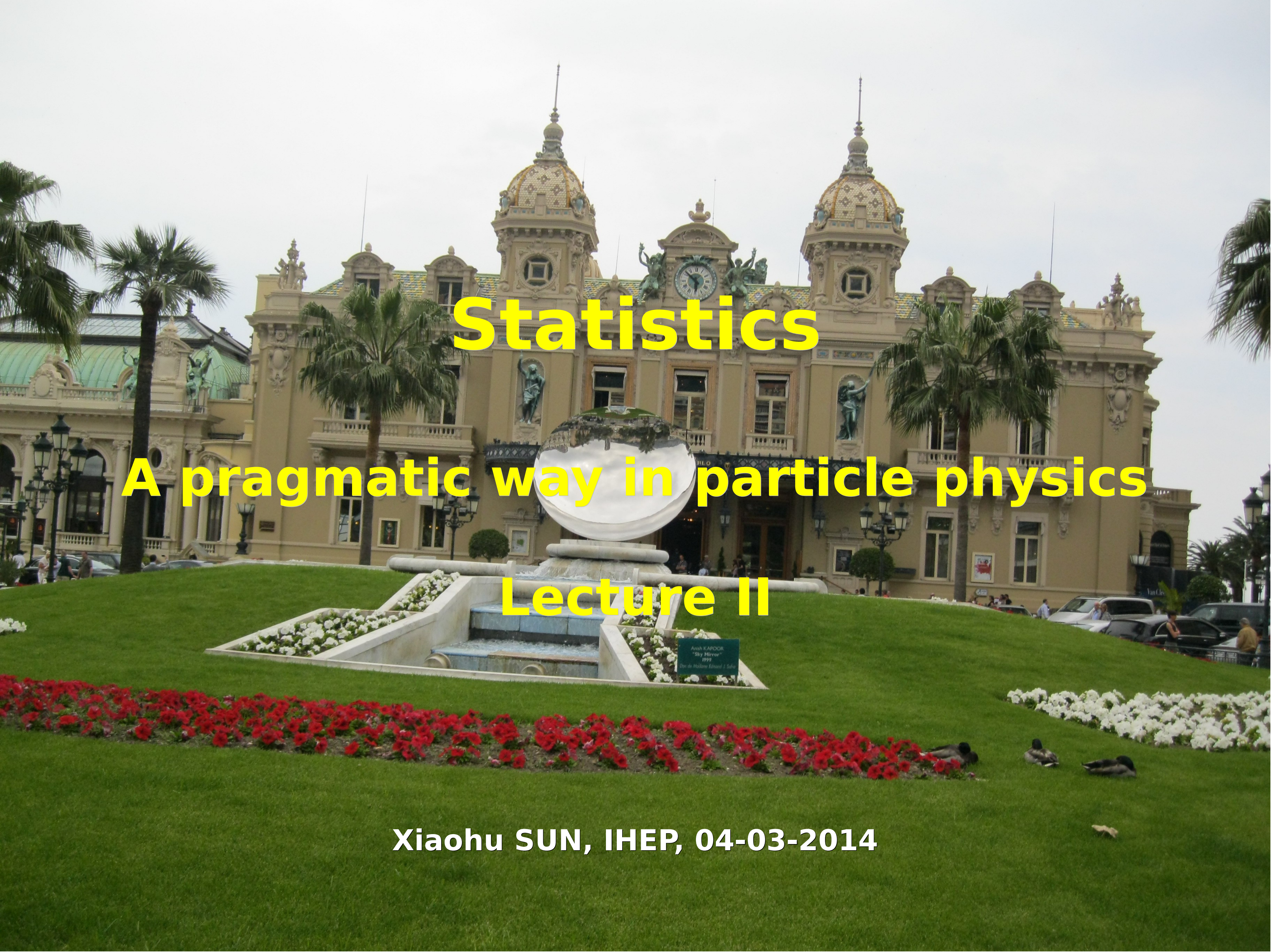


Statistics

A pragmatic way in particle physics

Lecture II

Xiaohu SUN, IHEP, 04-03-2014

What is this lecture about?

- A fresh reminder of why we need statistics in particle physcis
  - It helps to quantify our measurements, especially when the uncertianties are introduced
  - It helps to make decision “find or not a new particle?”
  - It helps to communicate physics results among experimentists as well as theorists

- Previously, I introduced the basics of probability: the definition, E[x], V[x], error propagation and some important distributions
- This time, I discuss
  - The parameter estimation
  - The hypothesis testing

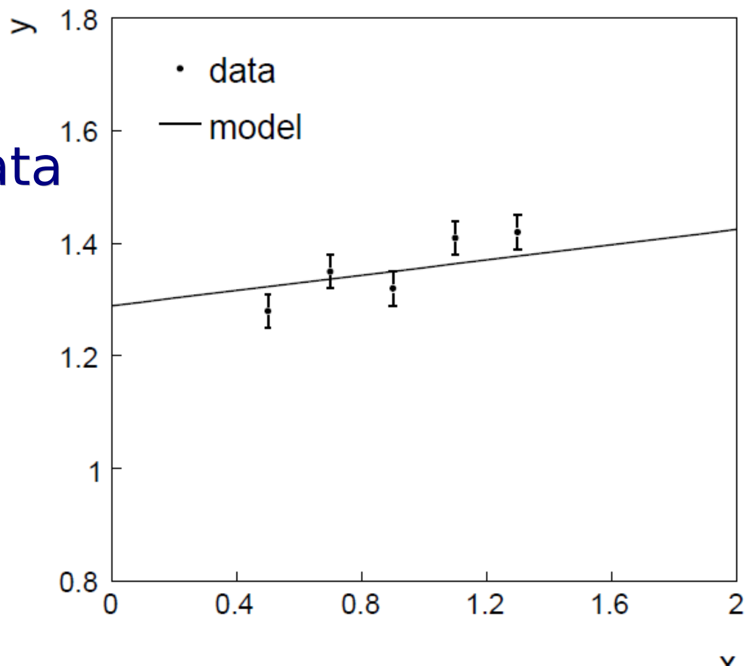
Parameter estimation - intro

- Random variable y follows a Gaussian distribution with a known standard deviation and a mean value as a function of random variable x

$$\mu(x; \theta_0, \theta_1) = \theta_0 + \theta_1 x$$

theoretical model

  - where θ_0 θ_1 are parameters to be fit
  - suppose the real goal is to obtain θ_0 (i.e. POI)
  - then θ_1 is treated as a nuisance parameter
- Data points $y_{1,2,\ldots,n}$, $x_{1,2,\ldots,n}$
- Find the optimal θ_0 θ_1 according to data
- So we need a statistical model
  - to fit theoretical model to data



Parameter estimation - L

- The pdf of y reads,

$$f(y_i;\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i-\mu(x_i;\boldsymbol{\theta}))^2/2\sigma_i^2}$$

- Construct a likelihood function that is a joint pdf of all y_i from data

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i;\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(y_i-\mu(x_i;\boldsymbol{\theta}))^2/2\sigma_i^2}$$

statistical model

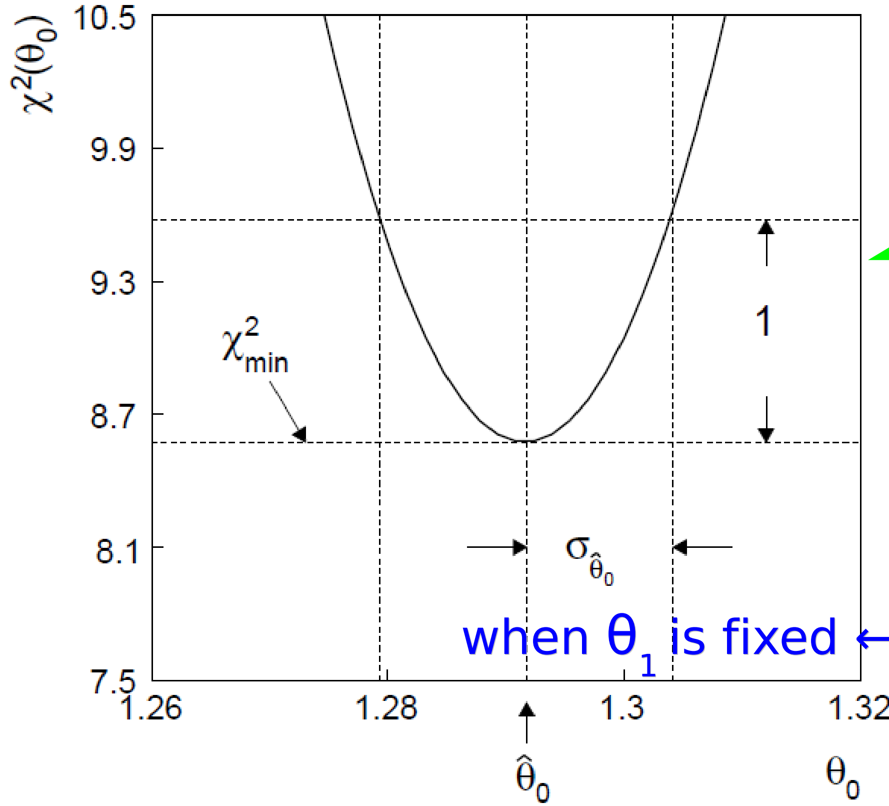
- L contains our theoretical model with to-be-fit parameters as well as the data from experiments
- L get maximized when the theoretical model describes well the data, in which case the parameters are optimal; otherwise, L value decreases
- The optimal (fitted) parameter is usually called estimator $\hat{\boldsymbol{\theta}}$
- Fit → maximization/minimization

Parameter estimation – NLL (LS)

- Maximize LL → Minimize NLL (negative LL)

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{(y_i - \mu(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2} = -2\ln L(\boldsymbol{\theta}) + C$$

- The left equation is actually the method of least square
- In our case LS coincide with NLL (condition:Gaussian distirbution)

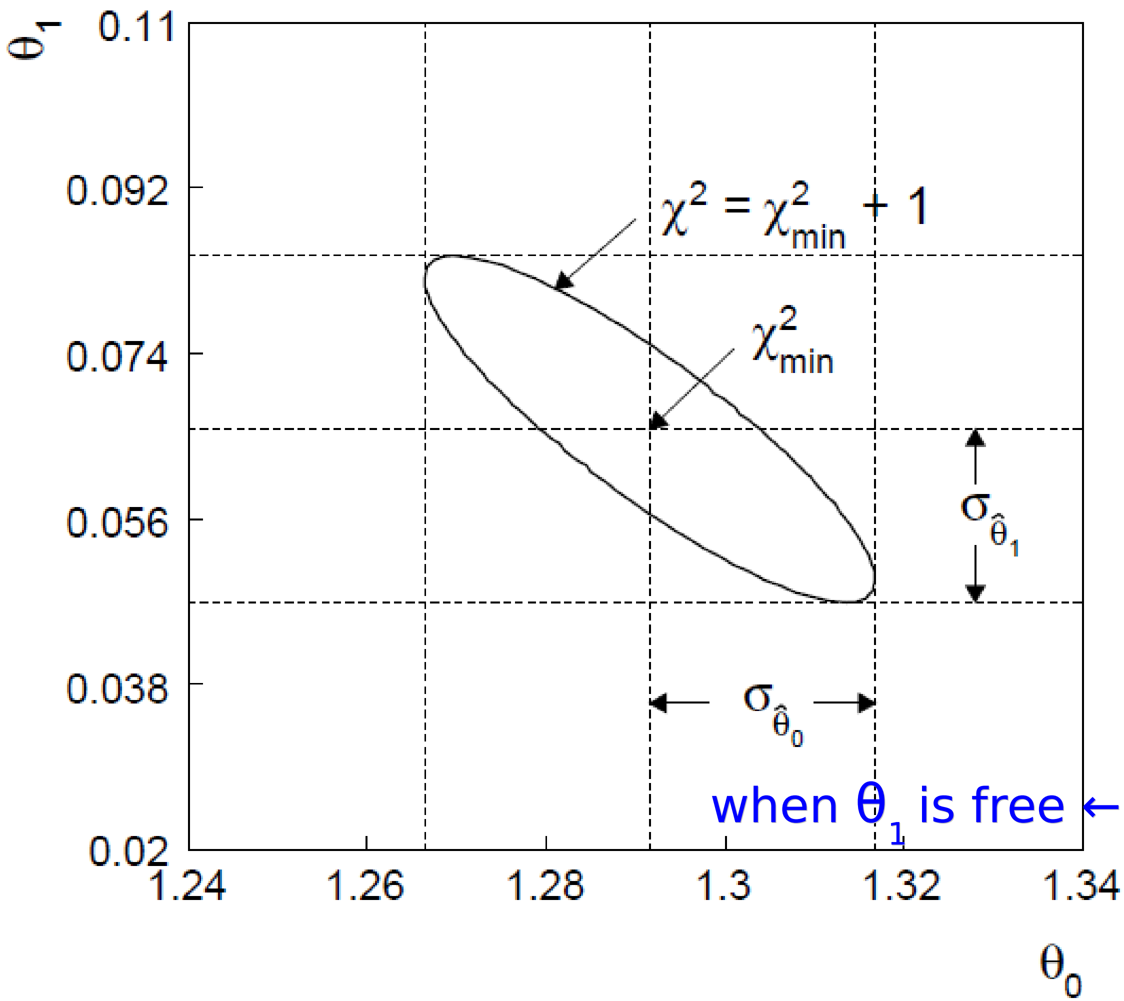


To find the error
LS/NLL: +1.0
LL: -0.5

when θ_1 is fixed ← NO nuisance parameter

Parameter estimation – θ_1 is free

- By introducing one nuisance parameter, the error of POI is getting larger as a price

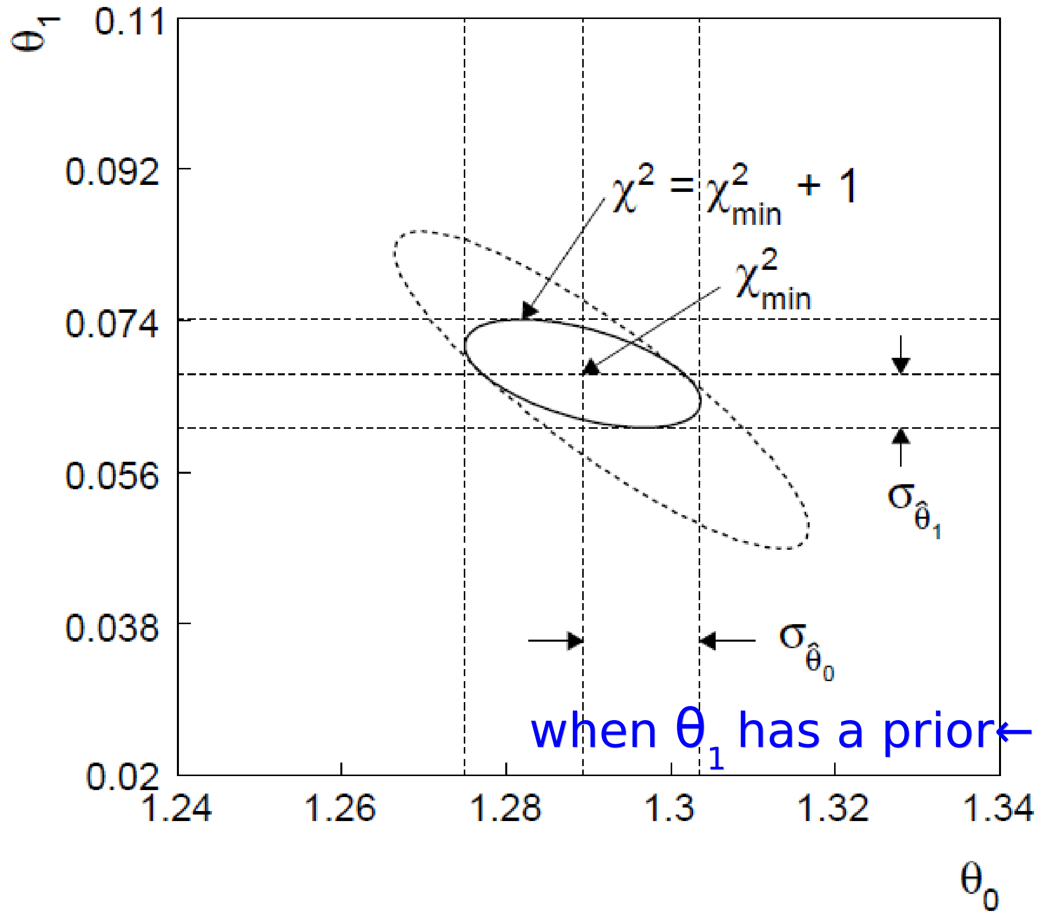


when θ_1 is free ← nuisance parameter

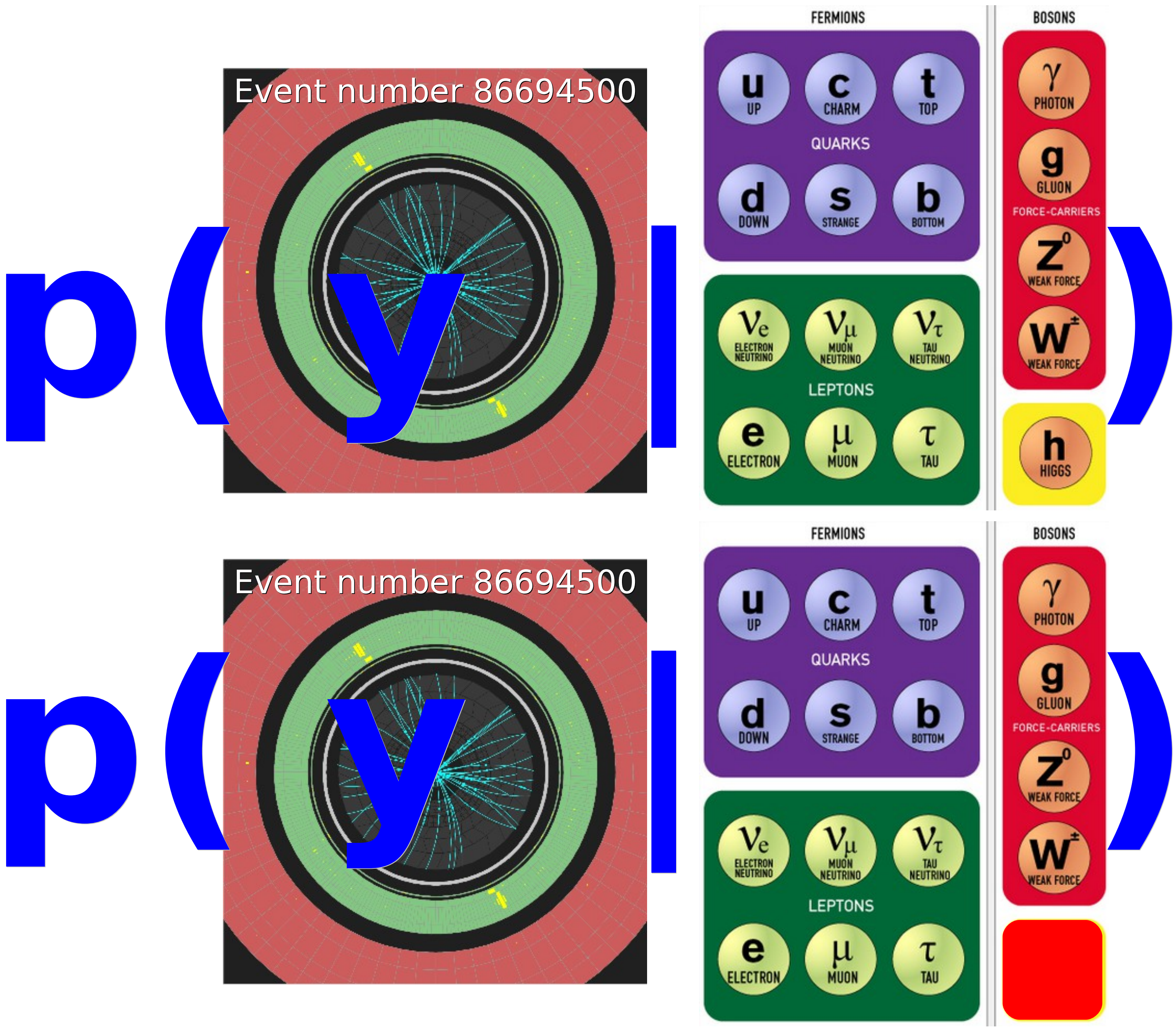
Parameter estimation – θ_1 has a prior

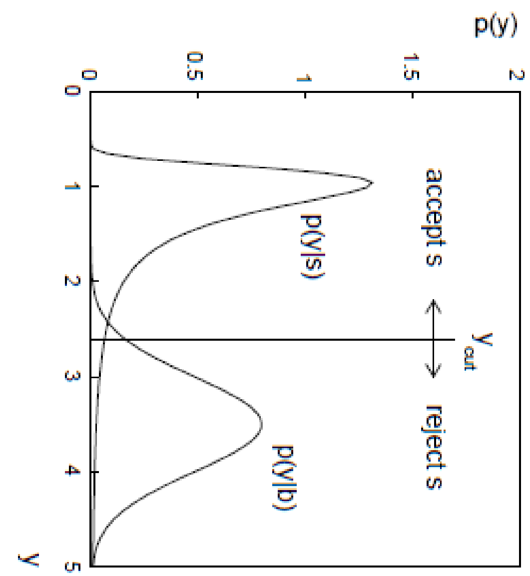
- To constain the error of POI, one can introduce a prior of nuisance parameter (a knowladge on how nui behaves)

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{(y_i - \mu(x_i;\boldsymbol{\theta}))^2}{\sigma_i^2} + \frac{(\theta_1 - t_1)^2}{\sigma_{t_1}^2}$$



Test statistic – mapping into 1-D

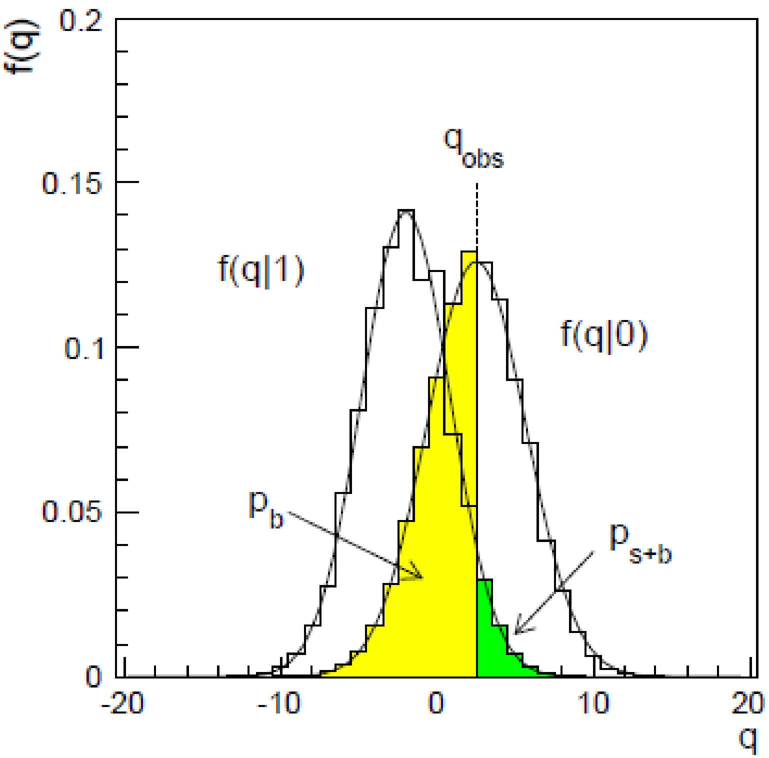




Data is mapped into a 1-d variable (**test statistics**)
with two pdfs for S+B and B-only hypotheses

Test statistic

- Pdfs f(q|1) f(q|0) correspond to S+B and B hypo
- I leave the discussion on how to construct q in the future



q_{obs} is the test statistics calculated from the data

The p-value of S+B is the prob to find q greater than or equal to q_{obs} under the assumption of S+B

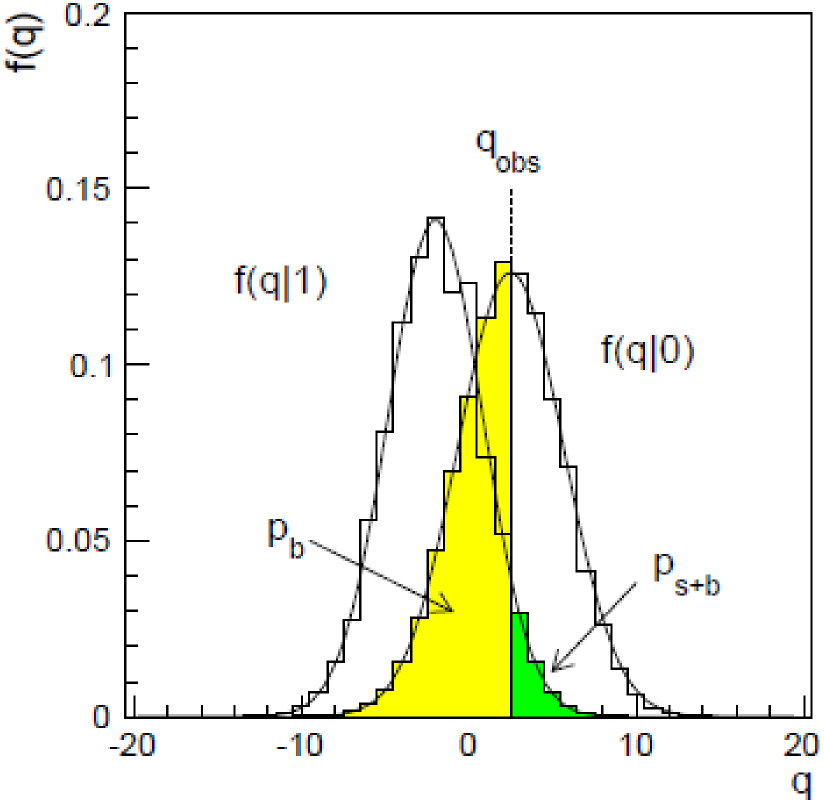
$$p_{s+b} = P(q \geq q_{\rm obs}|s+b) = \int_{q_{\rm obs}}^{\infty} f(q|s+b)\, dq$$

Similarly

$$p_b = P(q \leq q_{\rm obs}|b) = \int_{-\infty}^{q_{\rm obs}} f(q|b)\, dq$$

Test statistic → exclusion or upper limit

- Calculate the test statistic of S+B hypo based on its p-value
- The signal model is regarded as excluded at a confidence level of 1-alpha = 95% if one finds



$$p_{s+b} < \alpha$$

$$1 - \alpha = 95\%$$

$$\text{confidence level CL} = 1 - \alpha$$

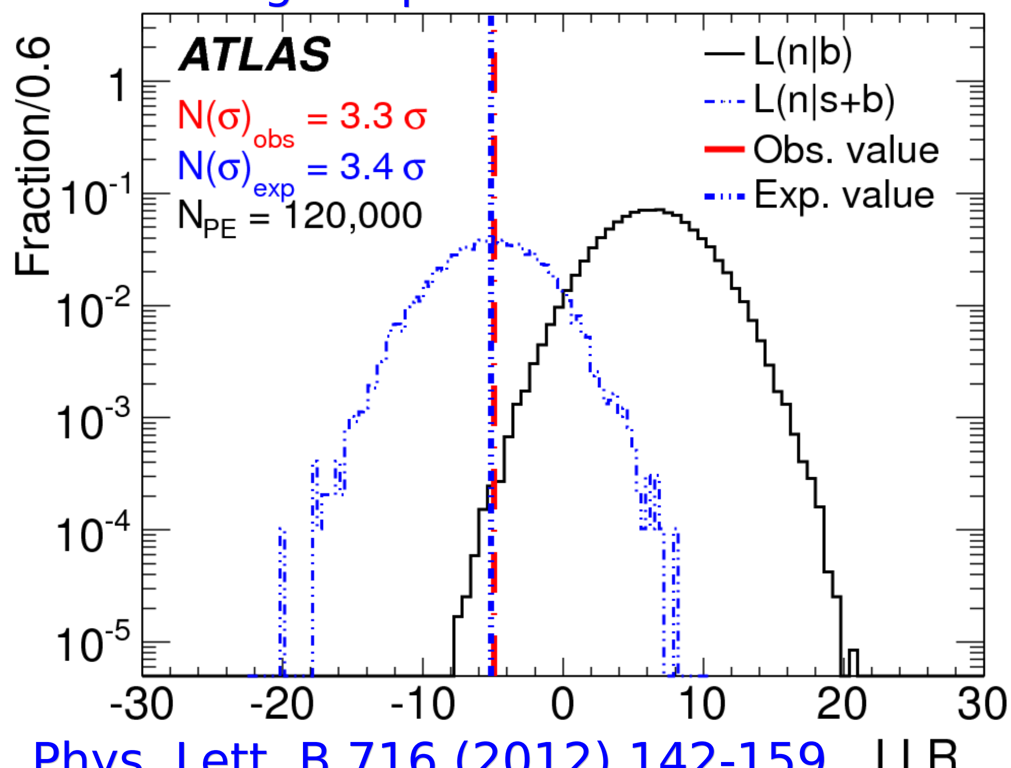
NOTE: in another word, the signal rate s under which the 5% p_{s+b} reaches, can be regarded as a upper limit s_{up}, for which signal is not excluded.

So the interval $[0,s_{up}]$ covers s with a probability of at least 95%

Test statistic → discovery

- Calculate the test statistic of B hypo based on its p-value: p_b
- Convert p-value into standard deviation (XX sigma)
- 3 sigma → evidence; 5 sigma → discovery
- The background-only model is regarded as rejected if 5 sigma

Single top Wt channel evidence



Phys. Lett. B 716 (2012) 142-159

Summary

- In this lecture, the basics of probability is introduced
  - The parameter estimation
  - The L/NLL, the least square LS
  - The test statistic
  - The exclusion, upper limit, discovery
- In the following lectures, more interesting topics on statistics will be covered
  - How ATLAS usually defines the statistical model
  - How ATLAS usually defines the test statistic
  - Confidence level (optional)


Stay tuned
for the next episode
P.O.I.