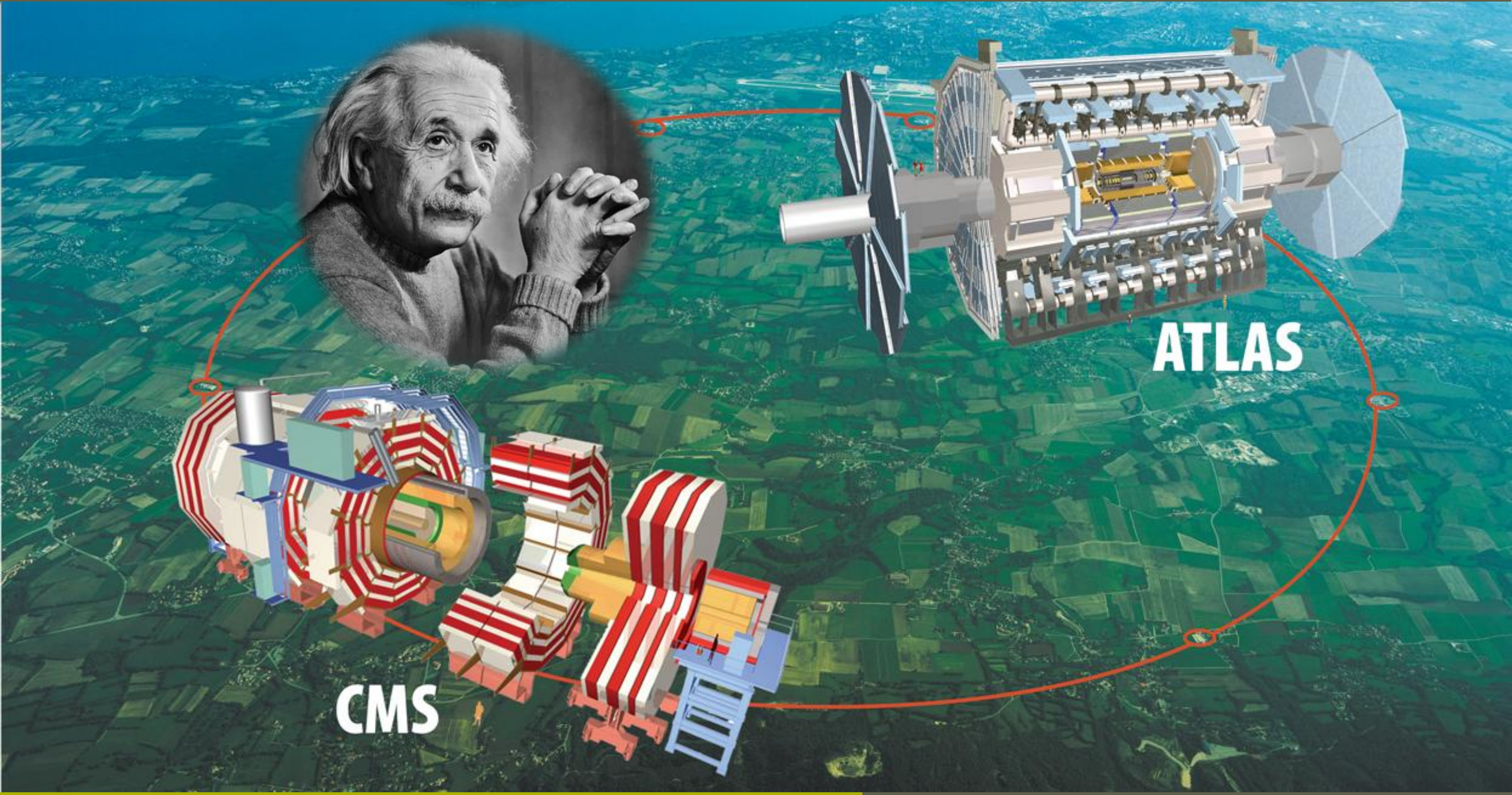
Theory versus Experiment

Prof. Jorgen D'Hondt
Vrije Universiteit Brussel
jodhondt@vub.ac.be

Theory versus Experiment



Dangerous cocktail !!!

The basics in these lectures

Part1 : "**Theory meets experiment**"

① Our QFT description of Nature is a stochastic one
② General stochastic distributions in physics
③ From theoretical to experimental distributions
④ ... and back: unfolding techniques
⑤ Examples from the LHC at CERN

Part 2 : "**Experiment meets theory**"

① Experimental aspects to accumulate experimental data
② Selection of the dedicated signal
③ Performing measurements & parameter estimation
④ Claiming a discovery of new physics or setting limits
⑤ Examples from the LHC at CERN

Lecture 1

"Theory meets experiment"

① Stochastic description of Nature
Stochastic variables, distributions, examples

② General stochastic distributions in physics
Bernouilli, Binomial, Poisson, Gaussian, Central Limit Theorem

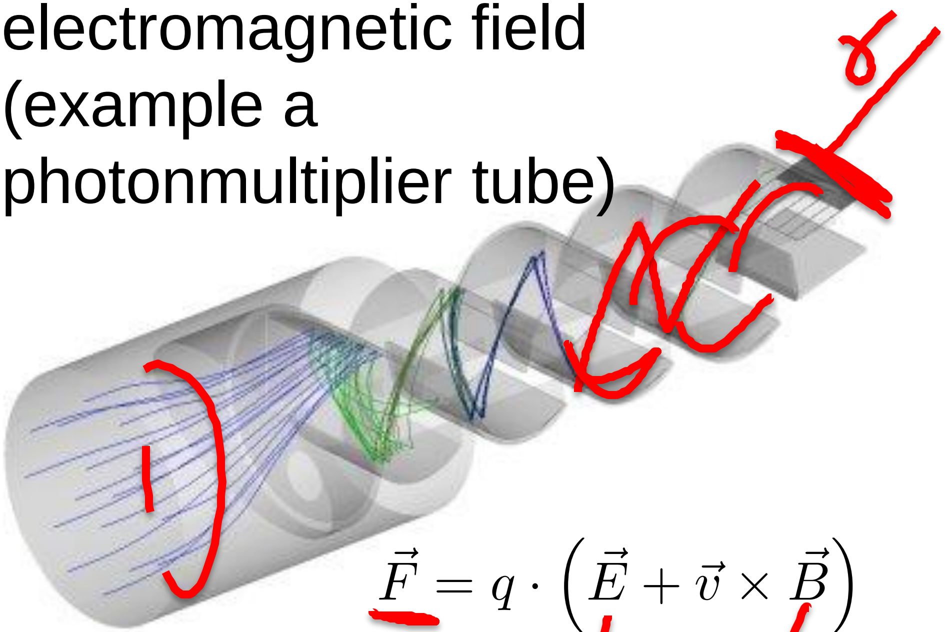
③ From theoretical to experimental distributions
Nuisance of experiment, convolution, Monte Carlo simulation, some reconstruction techniques

④ A basic introduction to unfolding techniques

⑤ An example from the LHC at CERN

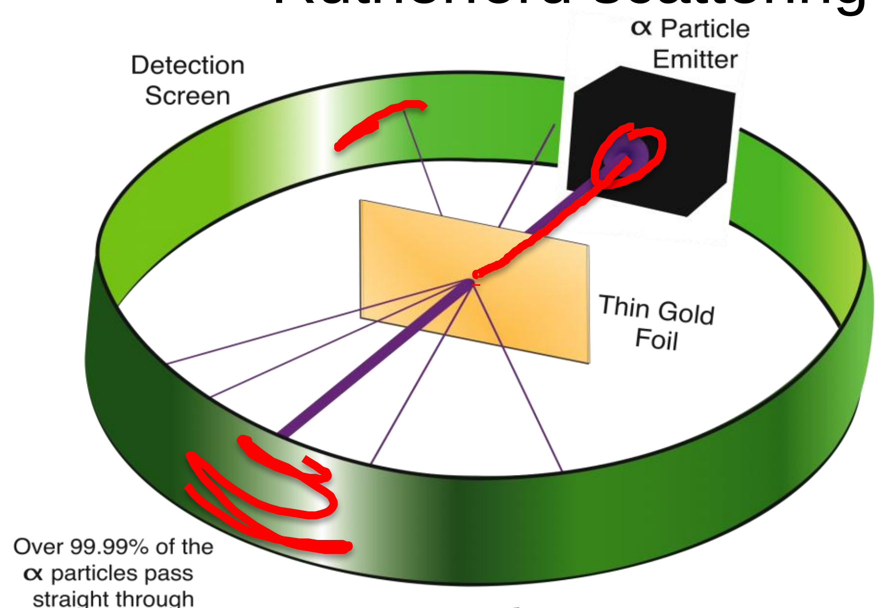
Deterministic vs Stochastic

Particles through an electromagnetic field (example a photonmultiplier tube)



$$\vec{F} = q \cdot \left(\vec{E} + \vec{v} \times \vec{B}\right)$$

Rutherford scattering



$$\frac{d\sigma}{dcos\theta} \sim \frac{1}{(1-cos\theta)^2}$$

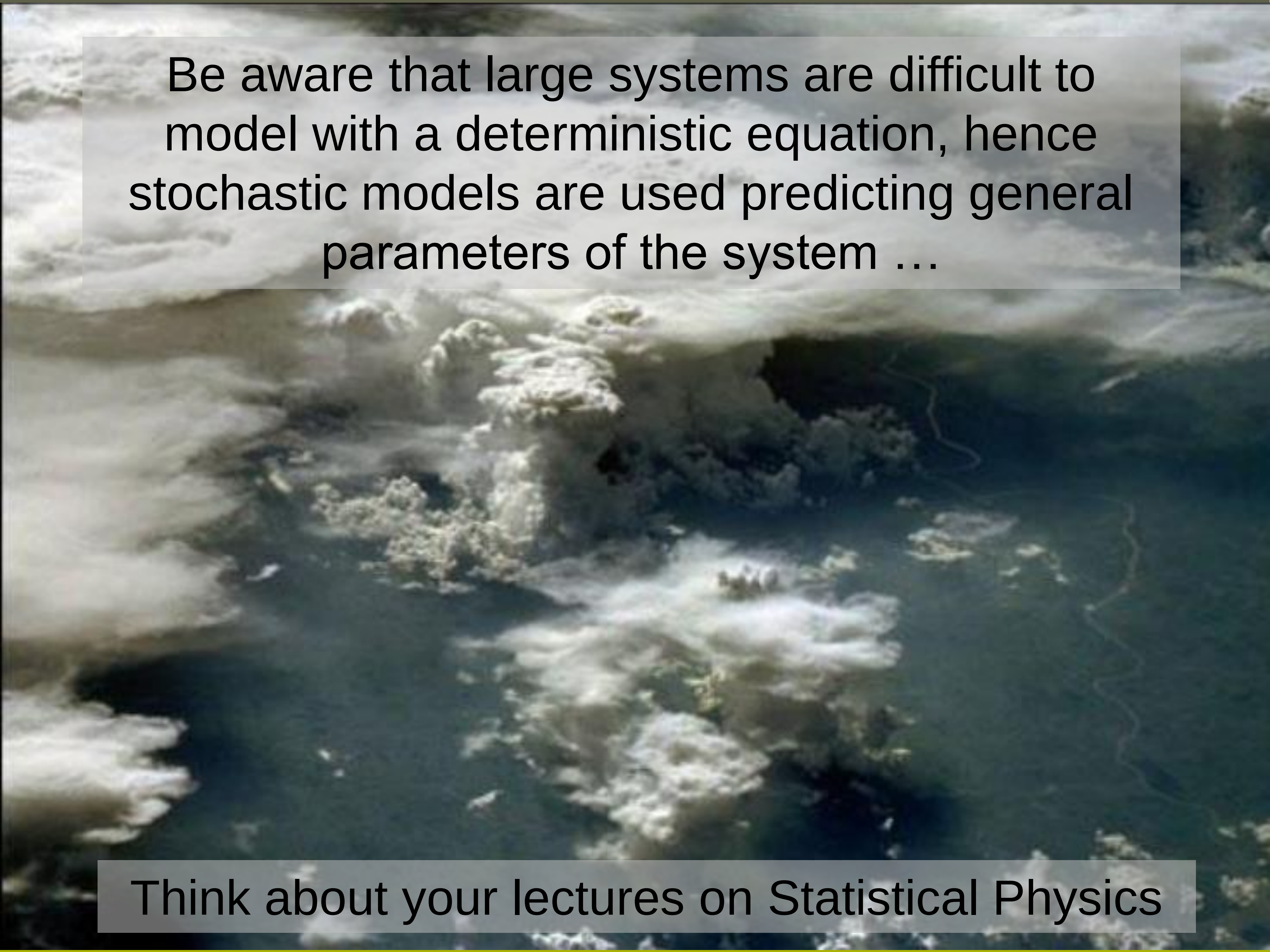
Deterministic

Each "experiment" is predictable (exact)

Stochastic

Set of "experiments" is predictable (distributions)

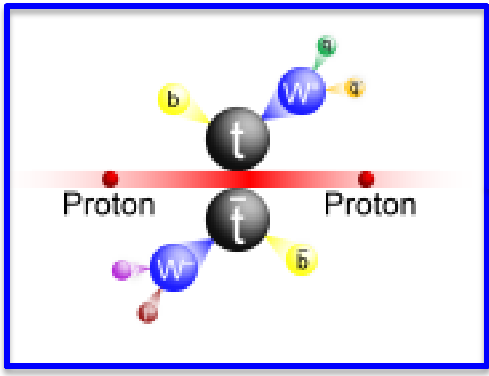
Be aware that large systems are difficult to model with a deterministic equation, hence stochastic models are used predicting general parameters of the system …



Think about your lectures on Statistical Physics

The micro-scale Nature is stochastic!

- Fundamental interactions are described on the level of quantum mechanics which is stochastic. Quantum Field Theory predicts cross-sections and distributions of kinematic variables.



→ Quantum Field Theory →

$$\sigma_{NLO}(pp \to t\bar{t})$$

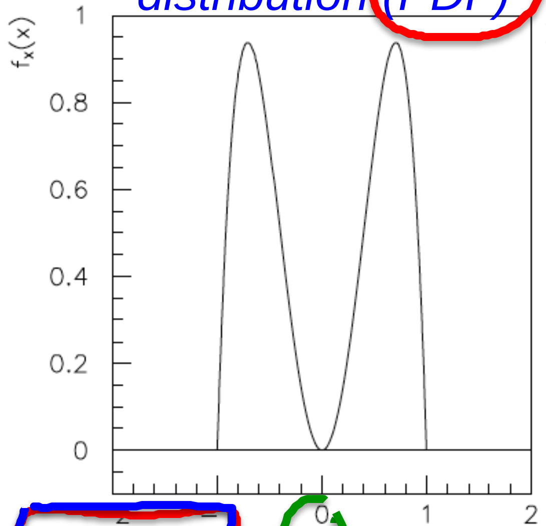
$$\frac{d\sigma_{NLO}(pp \to t\bar{t})}{dm_{t\bar{t}}}$$

- Prediction of the **expected** amount of events observed in collisions and the **expected** kinematic distributions. But not the prediction of the exact result of one single experiment.

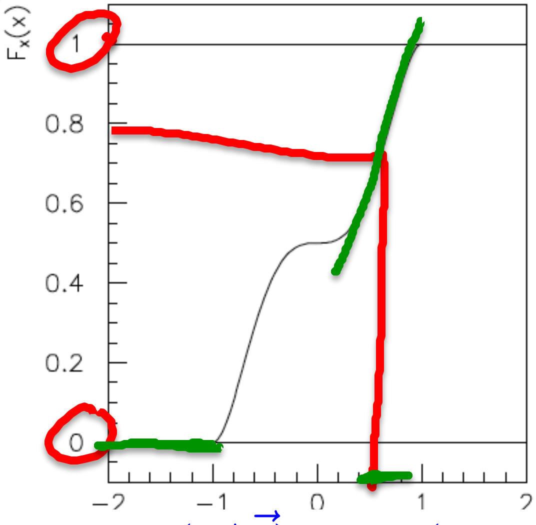
Distributions of stochastic variables

- A stochastic variable X denotes the outcome of one experiment and usually obtains a real value ($\mathbb{R}$) after its measurement.
- When repeated infinite times in the same conditions, a distribution is obtained (can be discrete or continuous).

distribution (PDF)



cumulative distribution



- $f_X(x)$ only obtained with infinite experiments, hence only theoretical
- Describe distribution with its momenta $\mu_k(X)$

$$\mu_k(X) = \int_\Omega x^k f_X(x)dx$$

$$f_X(x|\vec{\theta}) = \frac{d}{dx}F_X(x|\vec{\theta}) \qquad F_X(x|\vec{\theta}) = P(X \le x|\vec{\theta})$$

$$f_X(x) = \sum_{i=0}^{\infty} \left(\frac{\partial f(x)}{\partial x}\right)^i \frac{(x-a)^i}{i!}$$

$x = a$

c_i coefficients (∞)

physics $\Rightarrow$ $c_i = 0$

$\Rightarrow$ finite list c_i

$$f_X(x \mid \{c_i\})$$

theoretical parameters

Distributions of stochastic variables

- The collection of momenta $\{\mu_k(X)|k=\{1\ldots\infty\}\}$ gives a full description of the distribution $f_X(x)$
- Most distributions in physics have only few free momenta (the others are zero or can be expressed in terms of other momenta)
- Definition of expectation value: $E[g(X)] \equiv \int_\Omega g(x) f_X(x) dx$

  Typical examples of theoretical parameters of a distribution:

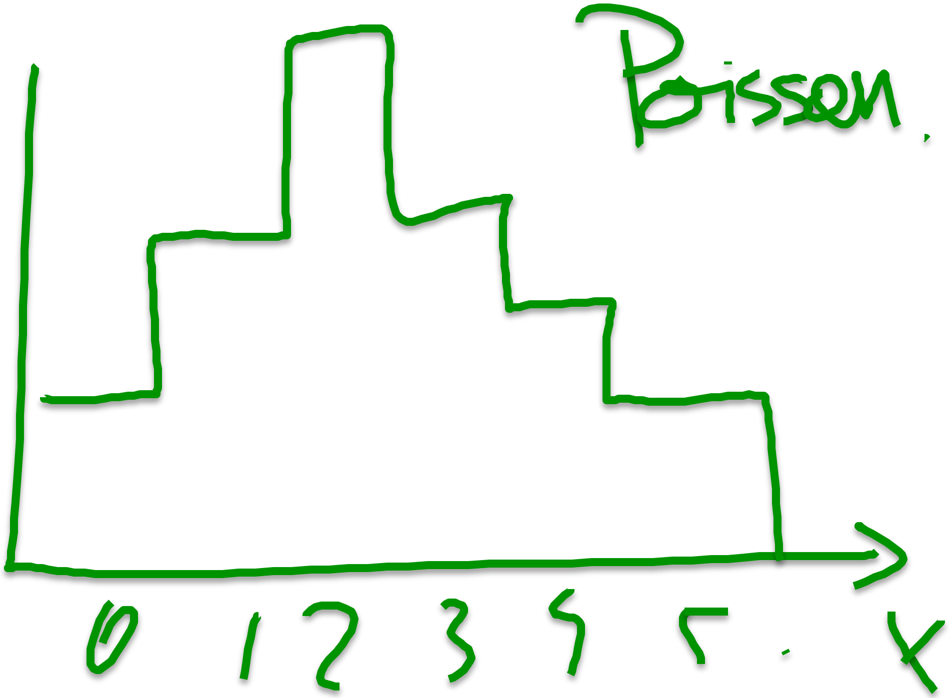
$$\mu_X \equiv \mu_1(X) = E[X^1] = \int_\Omega x^1 f_X(x) dx$$

$$Var[X] \equiv E[X^2] - (E[X])^2 = \int_\Omega (x - E[X])^2 f_X(x) dx$$

$E[(X-E[X])^2]$

  These parameters are unknown and have to be measured

Examples of stochastic variables (X): number of collision events in a dataset of 1/fb, reconstructed value of $m_{t\bar{t}}$ per event, …



X ~ stoch. var

Distributions in physics

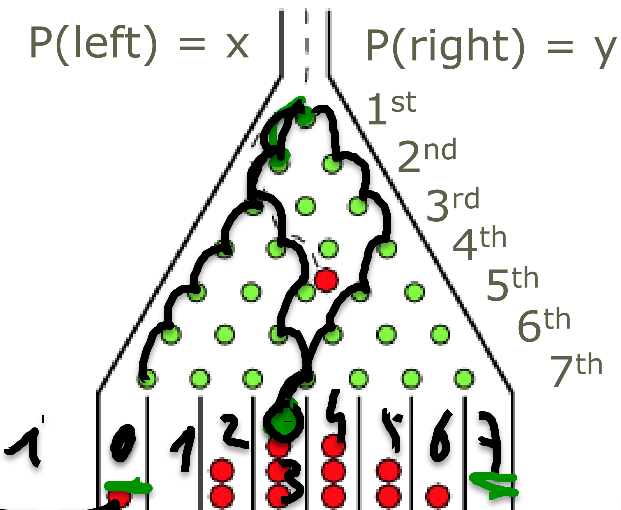
Which distributions are relevant to compare theory with experiment in particle physics?

Binomial distribution

Starts from a very simple Bernouilli distribution: the stochastic variable can only take two values (A or B)

P(A) = 1 – P(B)

Now you do this several times

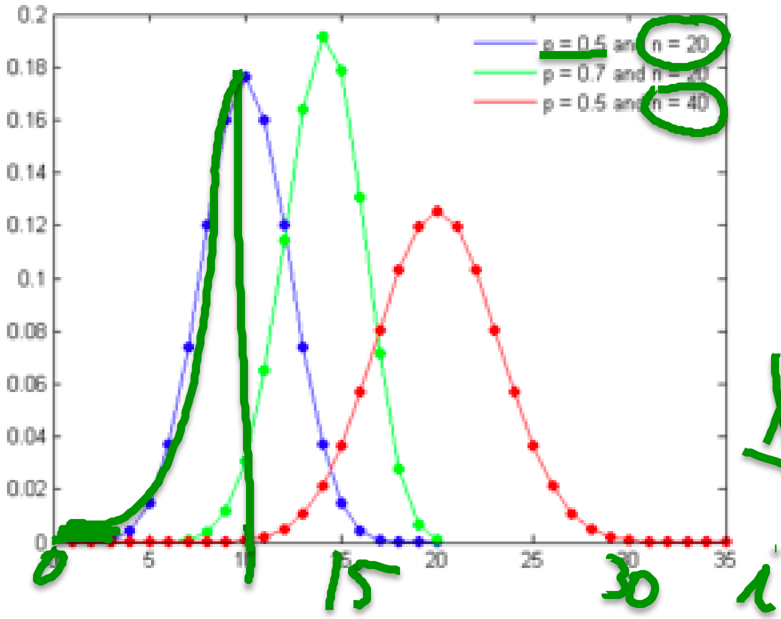


$$(x+y)^N = \sum_{i=0}^{N} \frac{N!}{(N-i)!\,i!} x^i y^{N-i}$$

binomial distribution

Binomial distribution

- One parameter to describe the full distribution P(i): N (number of Bernouilli experiments)



$$(x+y)^N = \sum_{i=0}^{N} \frac{N!}{(N-i)!i!} x^i y^{N-i}$$

$$P_N(i) = \frac{N!}{(N-i)!i!} p^i (1-p)^{N-i}$$

$$\mu_X = E[X] = N.p$$

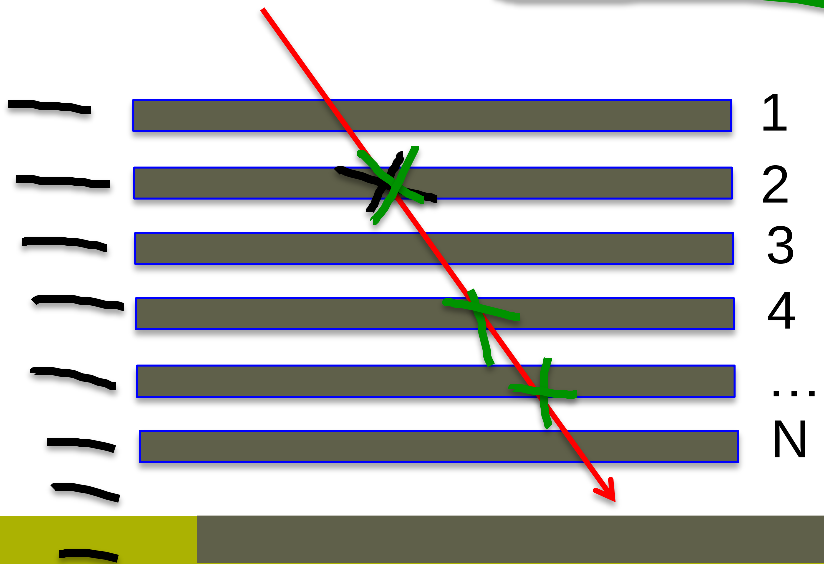
$$Var[X] = N.p.(1-p)$$

p = probablity for result A (eg. ball goes left)

Binomial distribution

- Typical exercise:

- A particle detector has an efficiency of 90% per detector layer. How many detector layers do we need to put on top of each other to get a probability of 99% to detect the particle? Detecting the particle means, that we have to "see" the particle in at least 3 layers.



Gaussian distribution

- The most important distribution in physics is the Gaussian one which is a rescaling of a binomial stochastic variable X ~ B(n,p)

$$Z = \frac{X - n.p}{\sqrt{n.p.(1-p)}} \xrightarrow{n \to \infty} f_Z(z) = \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2} z^2)$$

N(0,1)

- General Z ~ N(μ,σ²)

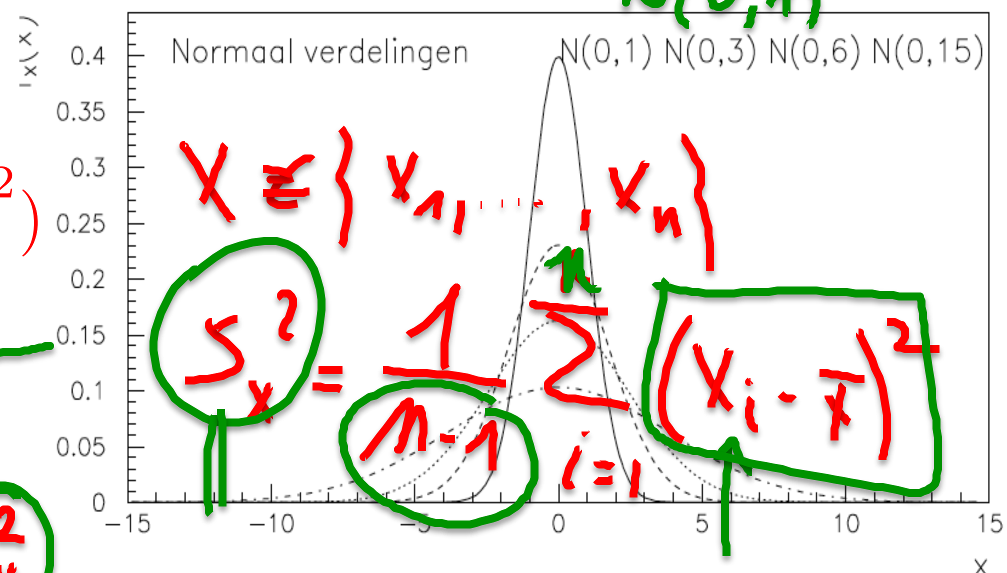
$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2)$$

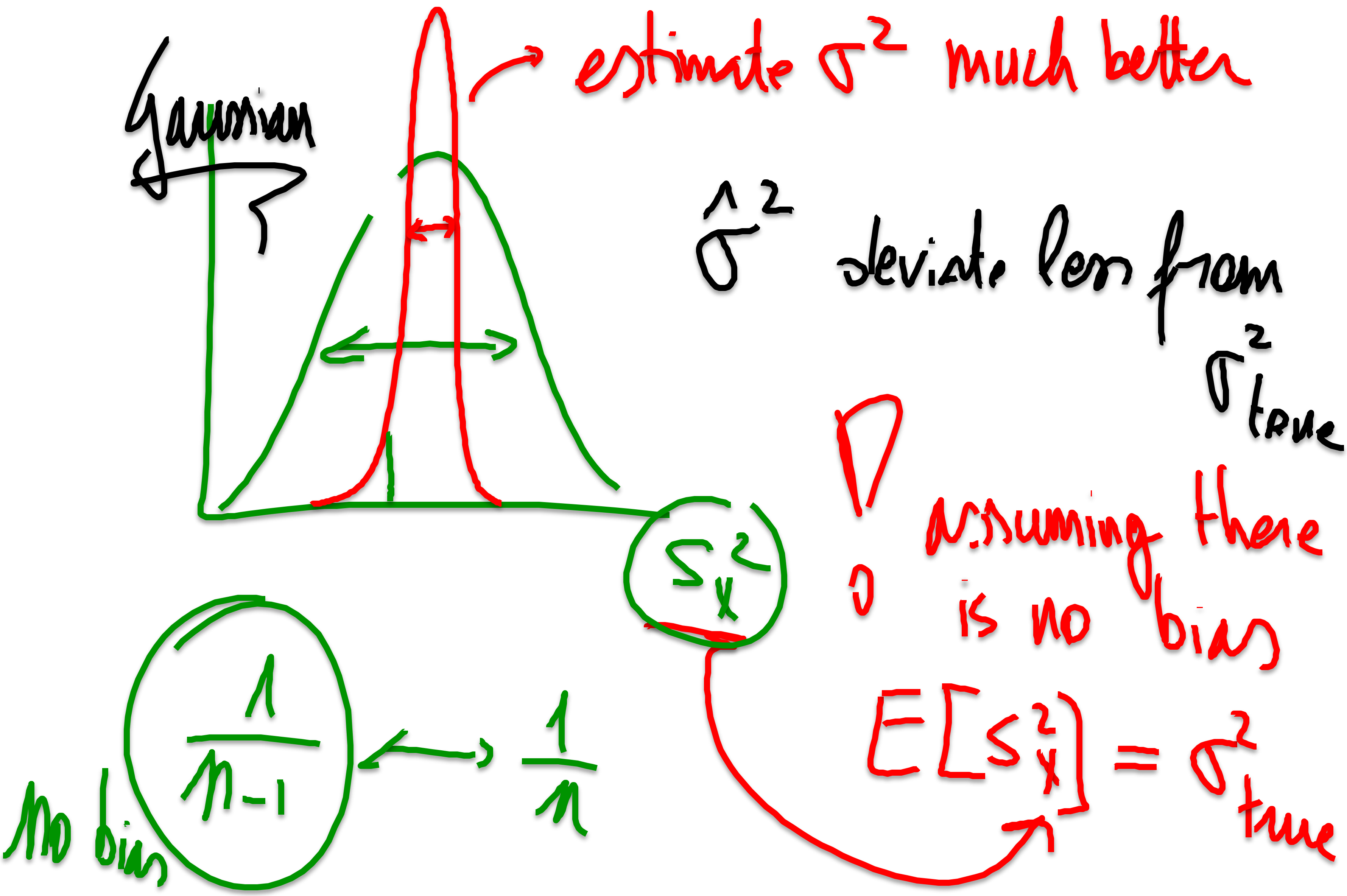
$\mu = E[X] \longrightarrow \hat{\mu} = \bar{x}$

$\sigma^2 = Var[X] \longrightarrow \hat{\sigma}^2 = s_x^2$

$X \in \{x_1, \ldots, x_n\}$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$



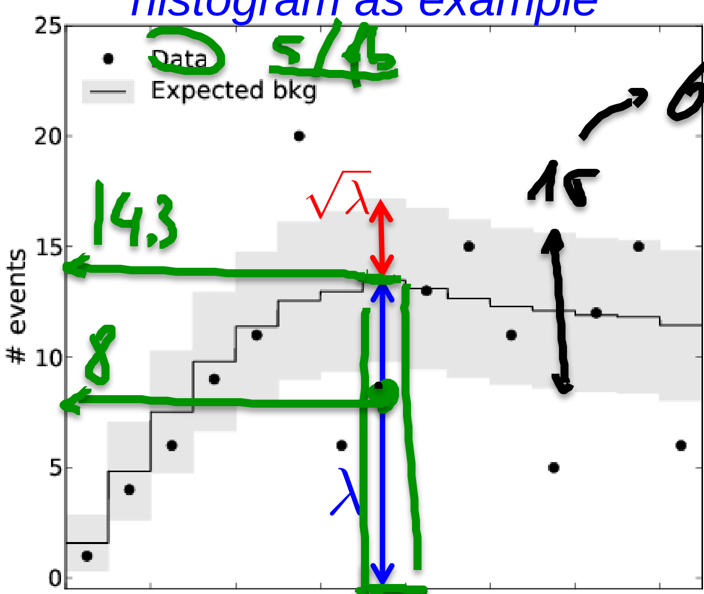

Gaussian
estimate σ^2 much better
$\hat{\sigma}^2$ deviate less from σ^2_{true}
s_x^2
! assuming there is no bias
$E[s_x^2] = \sigma^2_{true}$
$\frac{1}{n-1}$
$\frac{1}{n}$
no bias

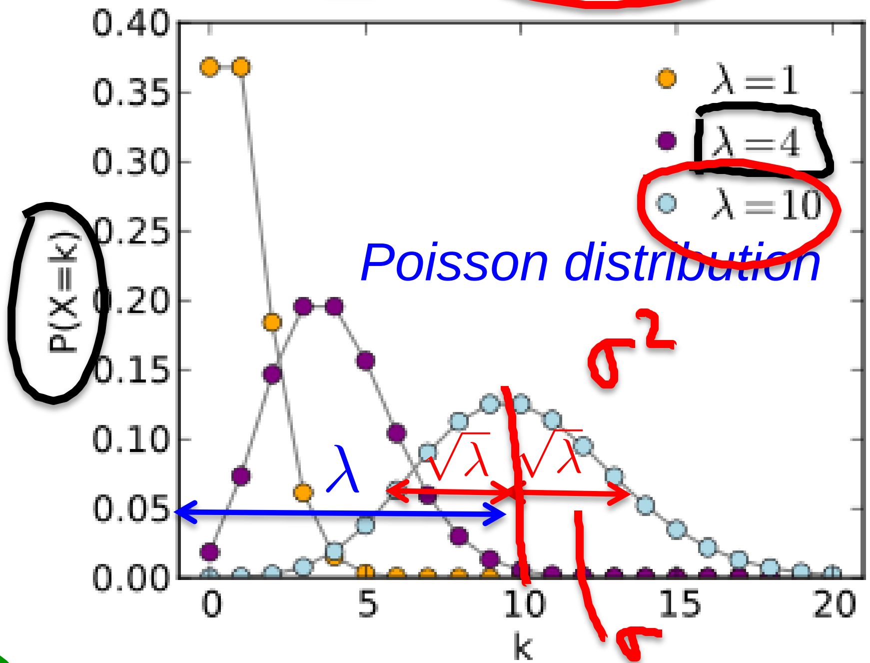
Poisson distribution

Stochastic variable X describes how many events happen in a fixed time interval (eg. how many collisions in 1/fb)

$$P_Y(k) = \frac{\lambda^k}{k!}e^{-\lambda}$$ one parameter $\lambda = E[X] = Var[X]$

histogram as example





Central Limit Theorem

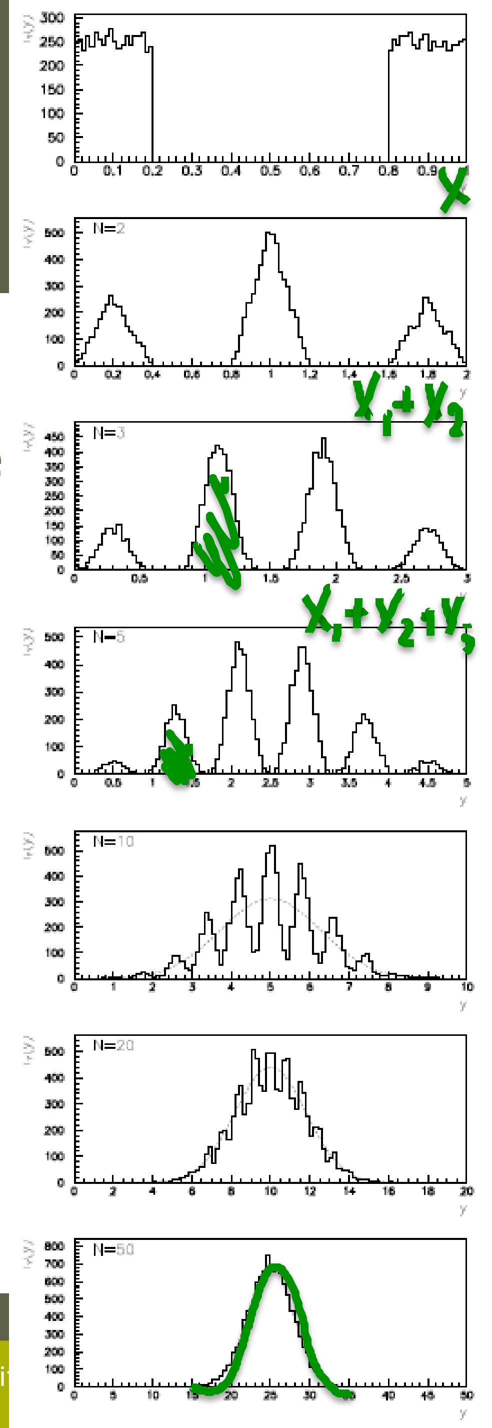
Consider independent stochastics X_i following all a random distribution $f_{X_i}(x_i)$. Each stochastic variable Y being a linear combination of these

$$Y = \sum_{i=1}^{n} c_i X_i$$

Will follow a normal distribution (when the variance Var[Y] is not dominated by one variance Var[X_i]).

$$Y \underset{n \to \infty}{\sim} N\left(\sum_{i=1}^{n} c_i E[X_i], \sum_{i=1}^{n} c_i^2 \sigma_{x_i}^2\right)$$

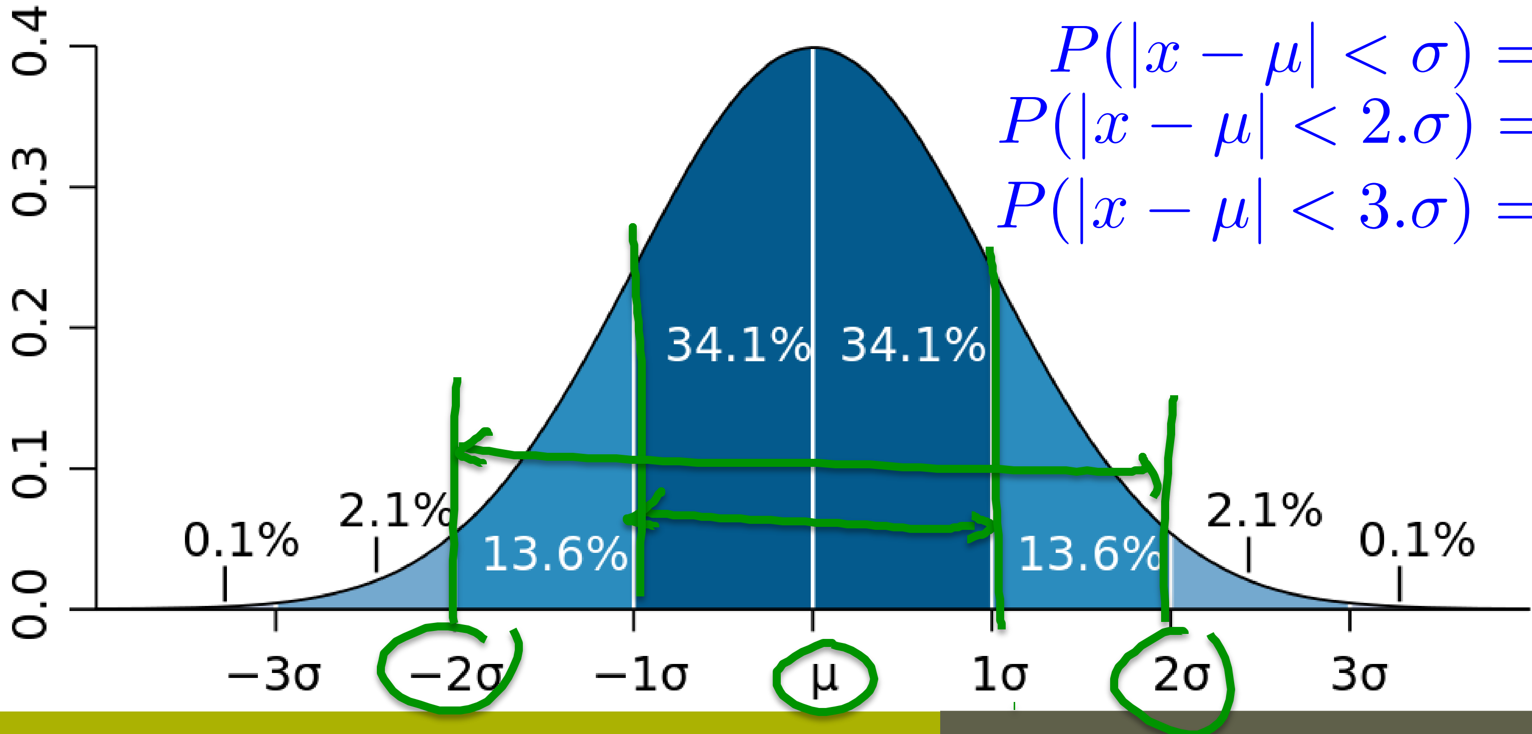
Hence the theory of uncertainties can be developed assuming the measurable stochastic follows a Gaussian distribution.



Gaussian distribution

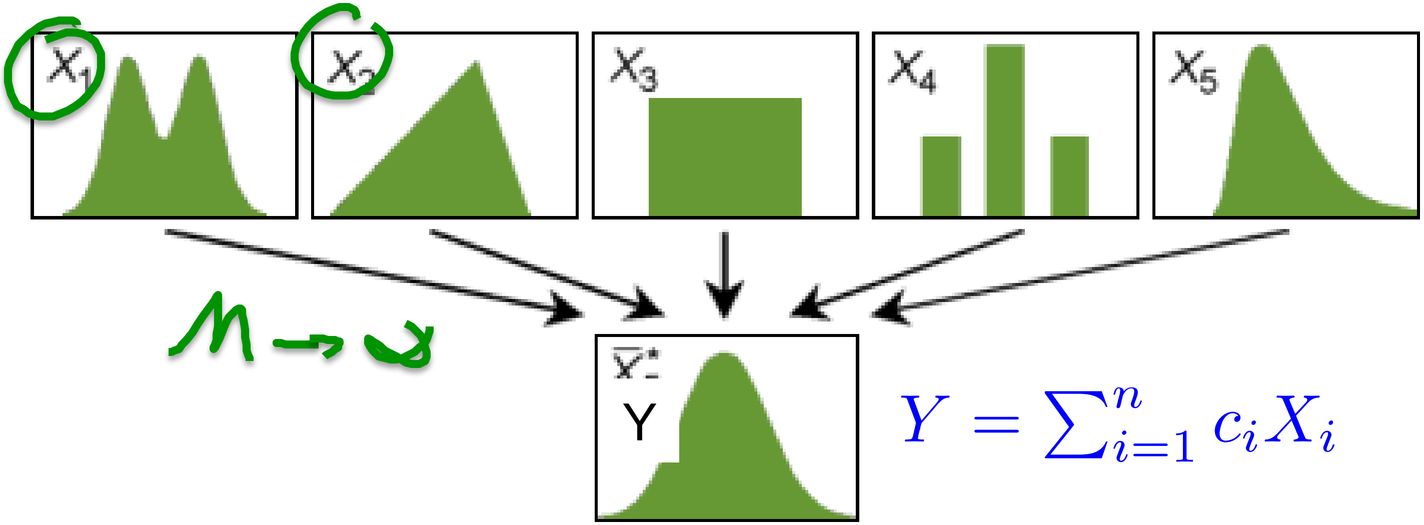
Estimators are usually linear combinations of stochastic variables, hence follow a Gaussian distribution

$$P(|x-\mu| < \sigma) = 68.3\%$$
$$P(|x-\mu| < 2.\sigma) = 95.4\%$$
$$P(|x-\mu| < 3.\sigma) = 99.7\%$$



Central Limit Theorem

The distributions X_i should not have the same PDF:



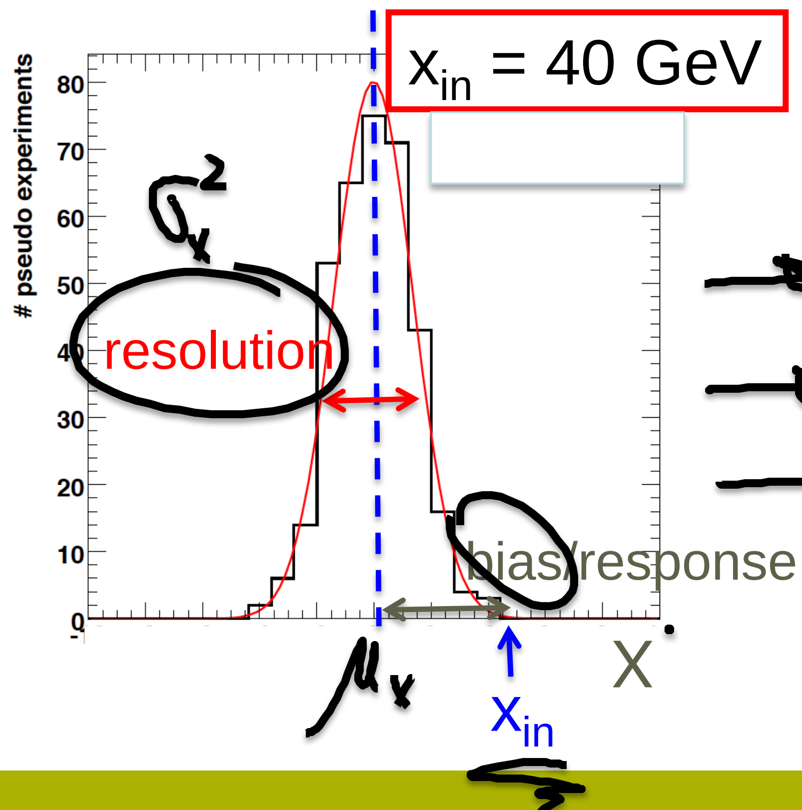
$$Y = \sum_{i=1}^{n} c_i X_i$$

Our particle detectors are typically complex instruments where particles are reconstructed from several "hits" which are combined into estimators for the four-momentum of the particle.

→ Gaussian resolution functions

Response & resolution

If I put in my detector a particle with variable value $X=x_{in}$, I will observe it at another value $X=x'$ according to some distribution.



Stochastic effects from

① the detector granularity

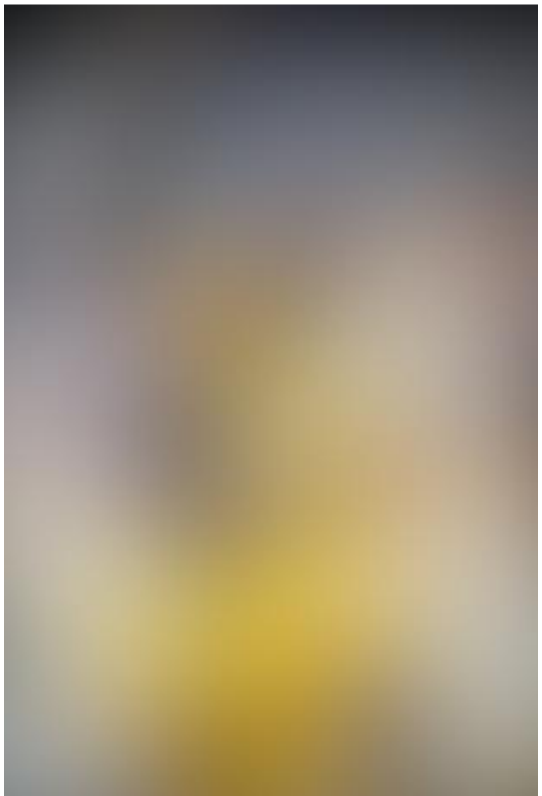
② the detector instrument

③ the physics (eg. jet reconstruction)

Resolution functions

① the detector granularity

Detectors make our life more complicated, but are needed!

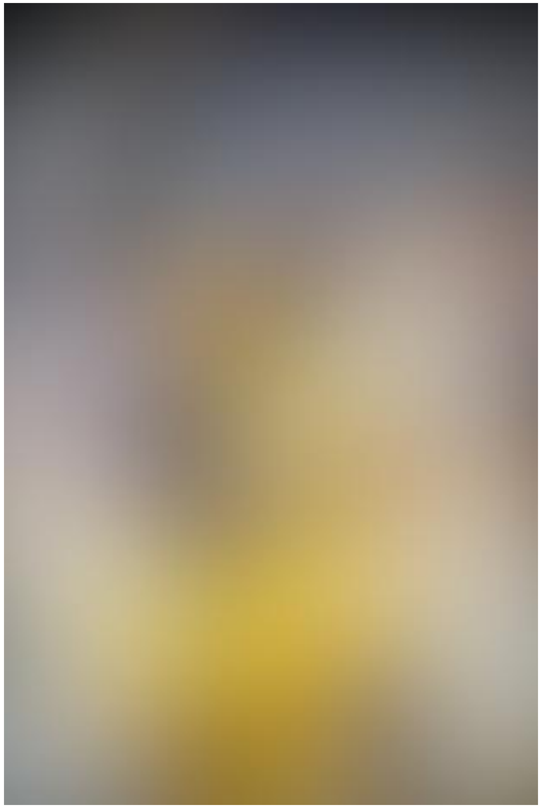


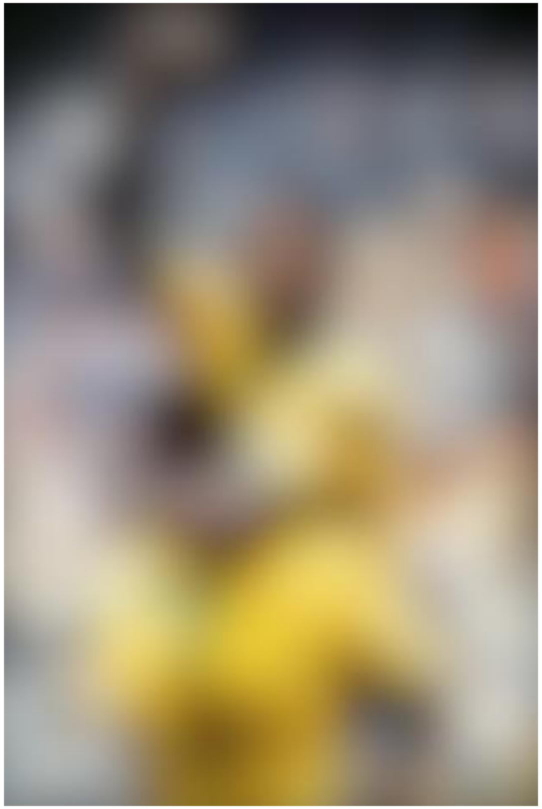
THEORY ?

Resolution functions

① the detector granularity

Detectors make our life more complicated, but are needed!

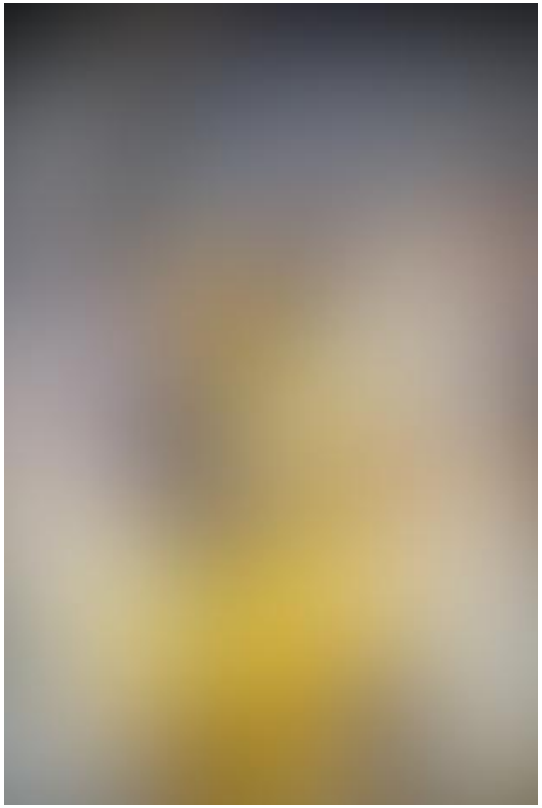


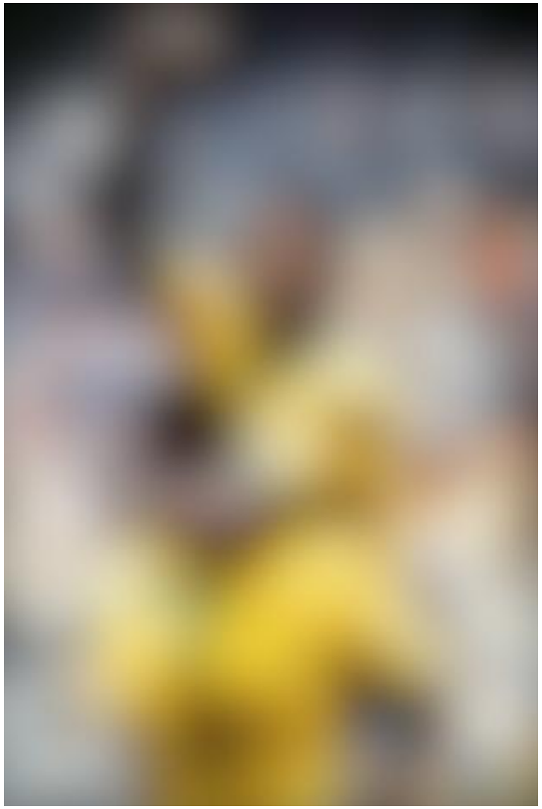


Resolution functions

① the detector granularity

Detectors make our life more complicated, but are needed!





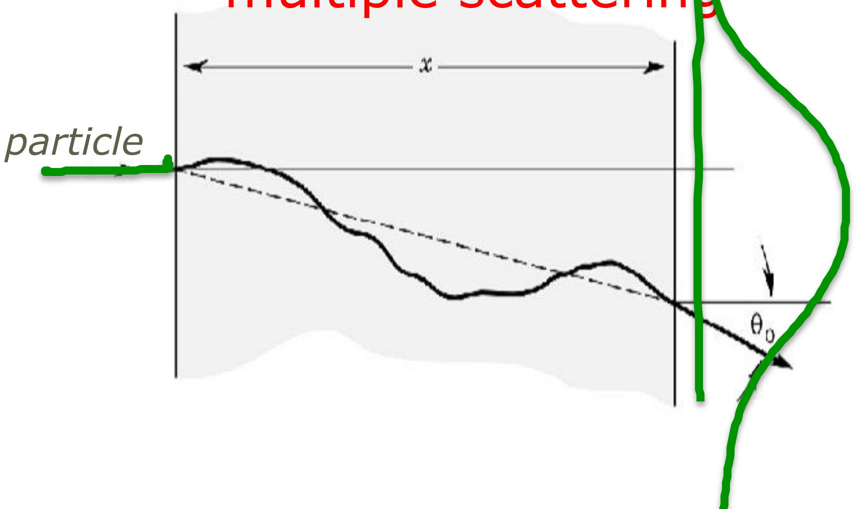
The more granular the detector, the more expensive the detector!

Resolution functions

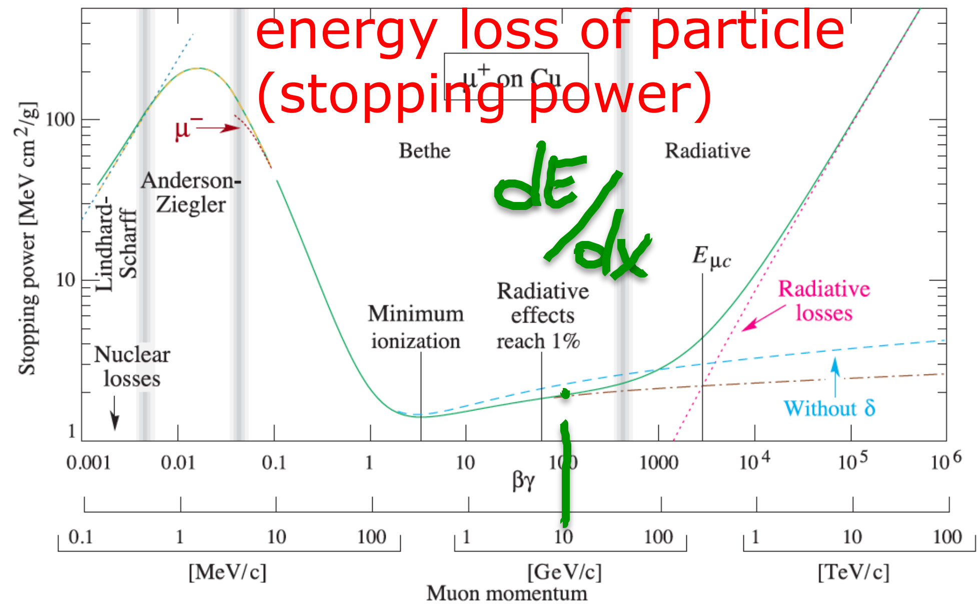
② the detector instrument

Physics itself can smear the "true" direction or energy of a particle

multiple scattering



energy loss of particle (stopping power)

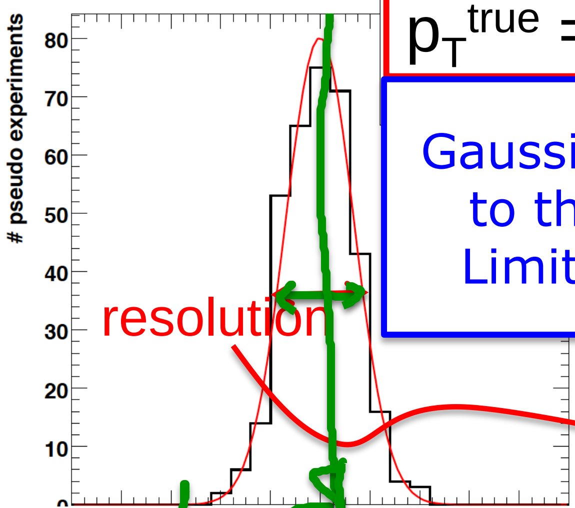


The incoming direction or energy is altered in a stochastic way and the <u>expected behavior</u> can be described by models put in the simulation of collision events at particle detectors.

Resolution functions

granularity + instrument + physics

The resolution function describes the stochastic distribution of the observable variable X (eg. p_T) for fixed settings before the measurement.



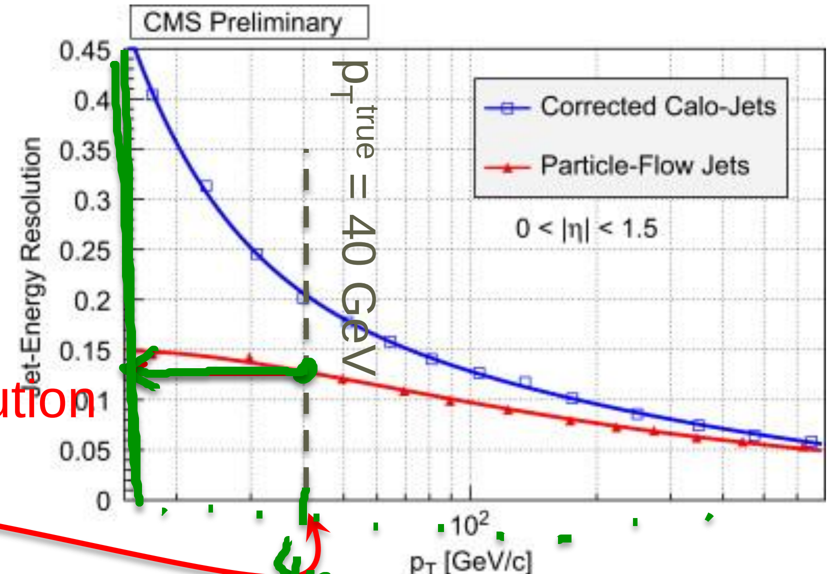
$p_T^{true} = 40$ GeV

Gaussian-like due to the Central Limit Theorem

resolution

p_T observed

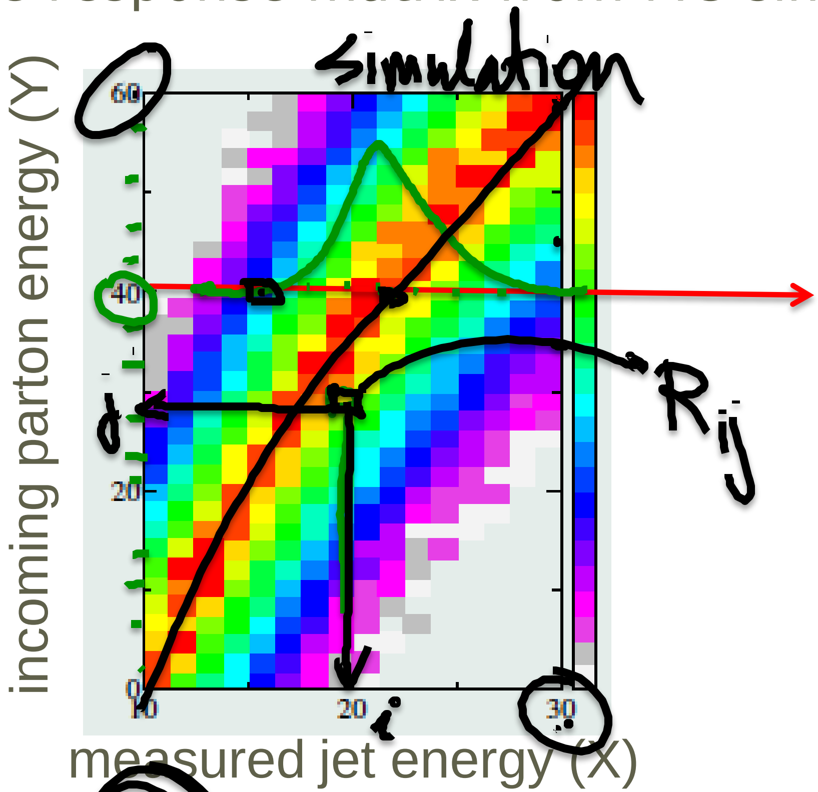
12% resolution

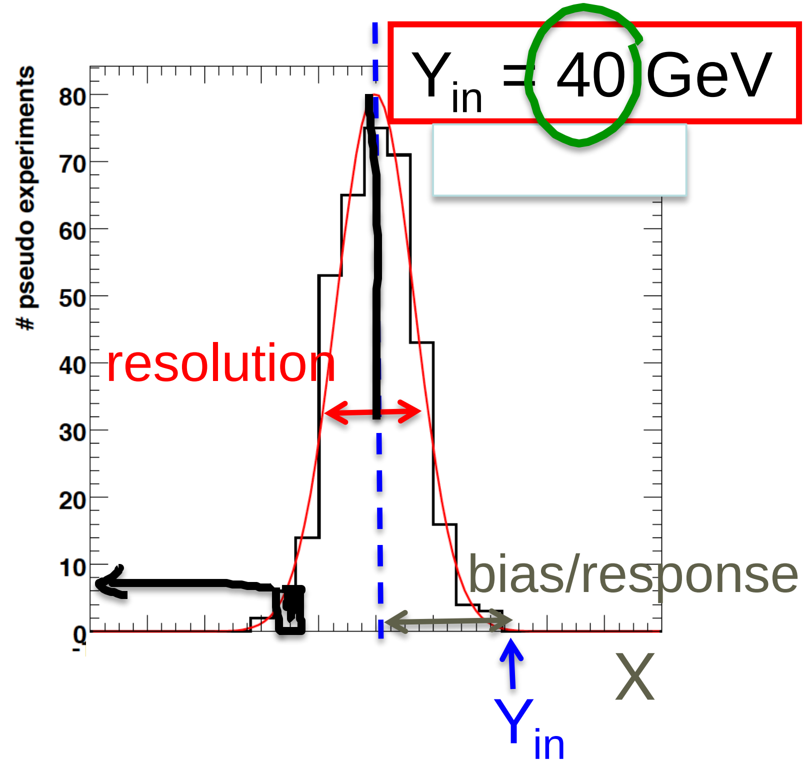


- One can reflect both the variance (resolution) and the bias (response) of the estimator into a response matrix.

Response matrix

The response matrix from MC simulation





matrix R_{ij} = Prob(obs in bin i | true value in bin j)

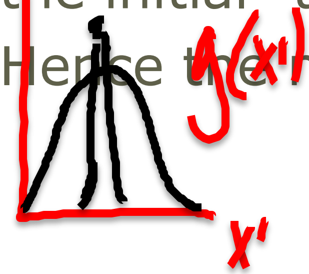
From TH to EXP distributions

Most of the resolution functions R(X,X') are Gaussian

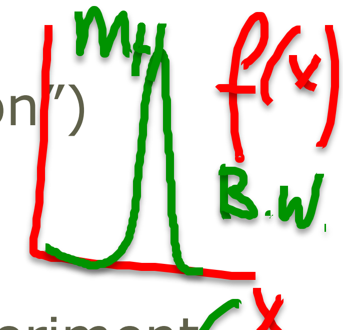
$$R(X,X') = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2}\left(\frac{x-x'}{\sigma}\right)^2\right)$$

and reflect the stochastic aspect of the measurement of X' given the initial "true" value X.

Hence the measured distribution becomes ("convolution")



$$g(x') = \int R(x,x') f_X(x) dx$$

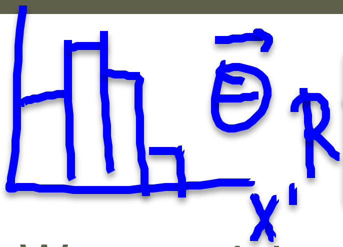


It is the distribution g(x') which is sampled by the experiment.

Including physics ($\vec{\theta}$) and detector "nuisance" ($\vec{\alpha}$) parameters:

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

From TH to EXP distributions



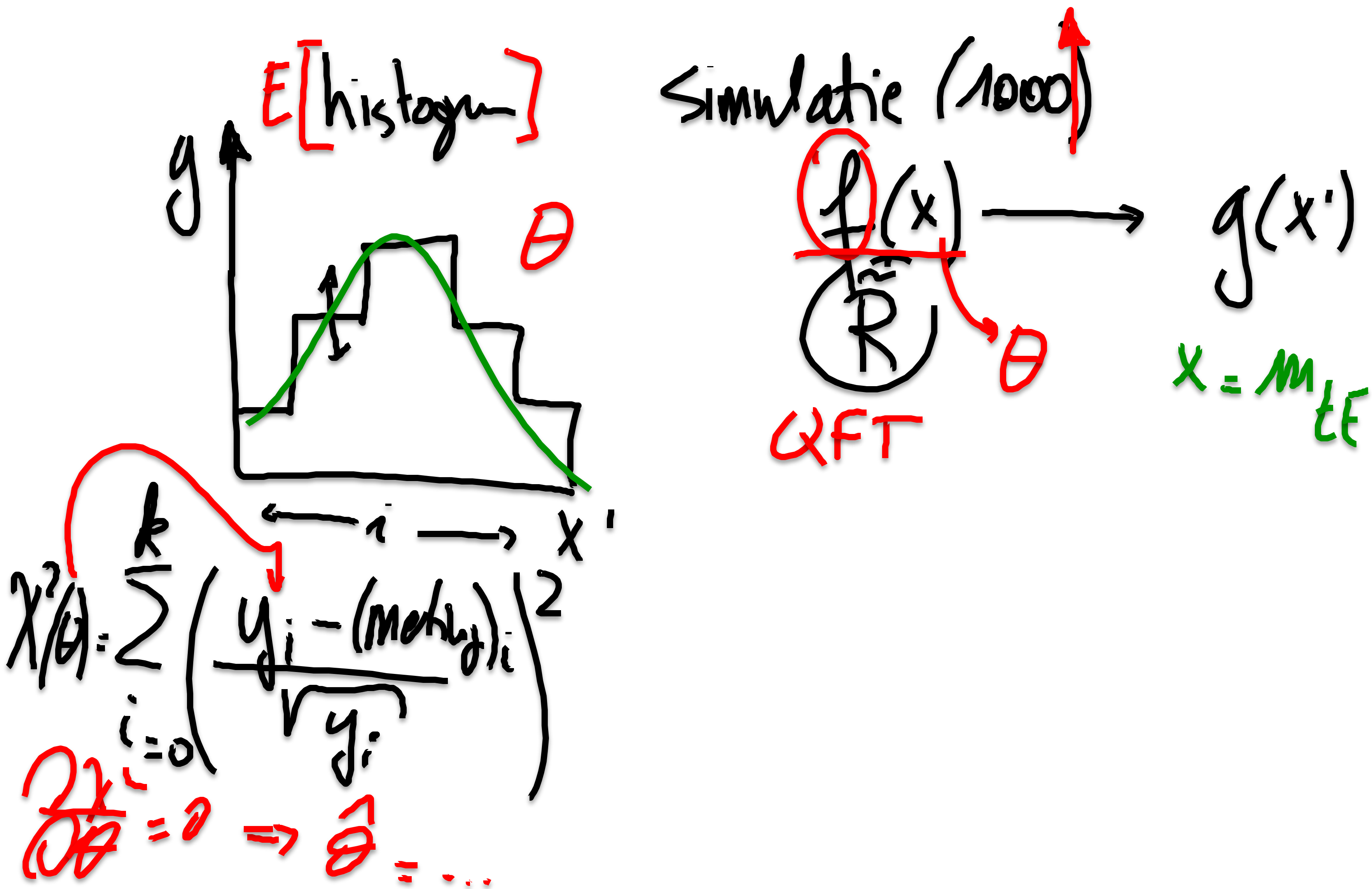
$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

We want to know something about the physics model $f_X(x|\theta)$ using a finite amount of measurements of stochastic variable X'.

The data of the measurements will be collected into a histogram with a finite amount of bins. In general two options:

① With simulation which incorporates the stochastic behavior of the response matrix R, we can predict the expectation value in each bin (= the bin content). And compare the data histogram with these expectations which depend on the parameter θ.

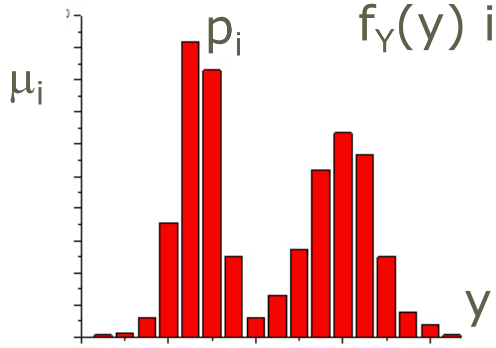
② When the response matrix R is well-known we can try to inverse the problem and estimate directly $f_X(x|\theta)$.


E[histogram]
Simulatie (1000)
f(x)
g(x')
R
QFT
θ
x = m_tE
y
x'
i
k
i=0
2

Unfolding techniques

("inverse problem")

Estimating probability distributions $f_Y(y)$ with an additional stochastic effect on Y from the measurement process (effect of R), in case when no parametric form is available



$f_Y(y)$ is an unknown PDF → histogram with M bins

μ_i expectation value

$$\vec{\mu} = (\mu_1, \ldots, \mu_M)$$

response function R(x|y) or response matrix R_{ij}

R_{ij} = Prob(obs in bin i | true value in bin j)

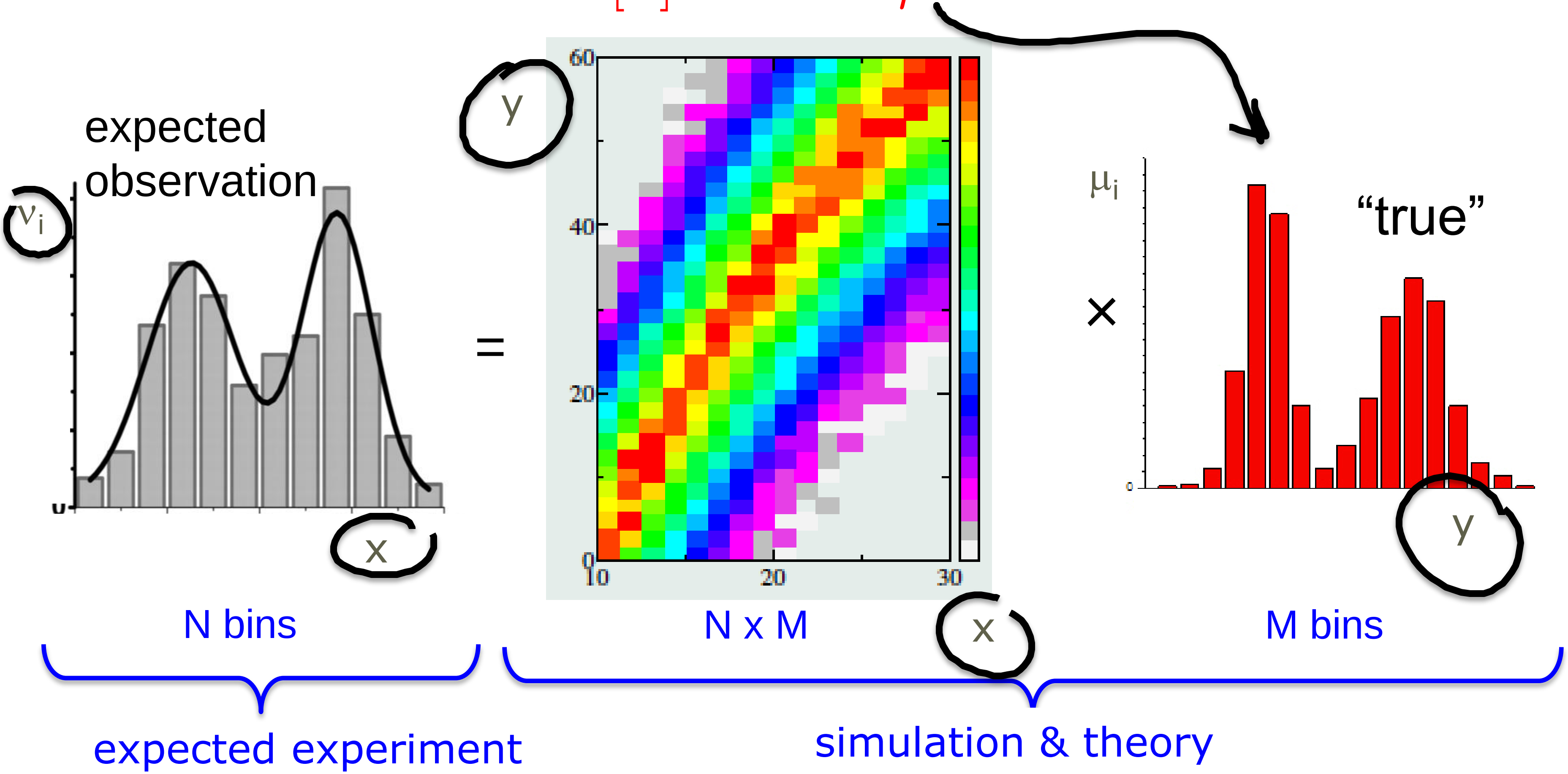
"functional" $$f_{meas}(x) = \int R(x|y).f_{true}(y)dy$$

"histograms" $$E[n_i] \equiv \nu_i = \sum_{j=1}^{M} R_{ij}\mu_j \rightarrow E[\vec{n}] = \vec{\nu} = \tilde{R}\vec{\mu}$$

$\sum_{j=1}^{N} R_{ij}$ = Prob(obs anywhere | true value in bin j) = ε_j *(efficiency)*

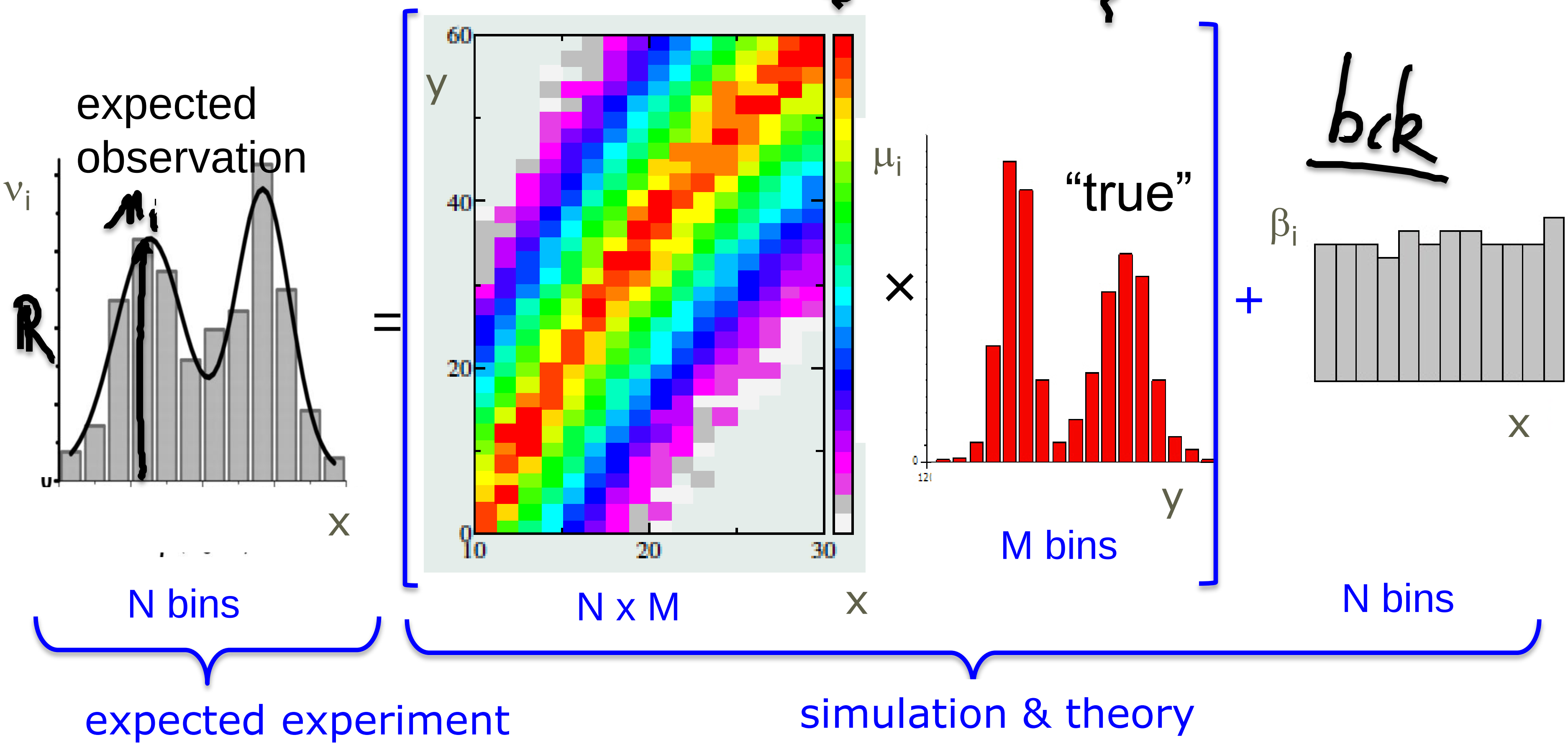
Unfolding techniques

With pictures… $\rightarrow E[\vec{n}] = \vec{\nu} = \tilde{R}\vec{\mu}$



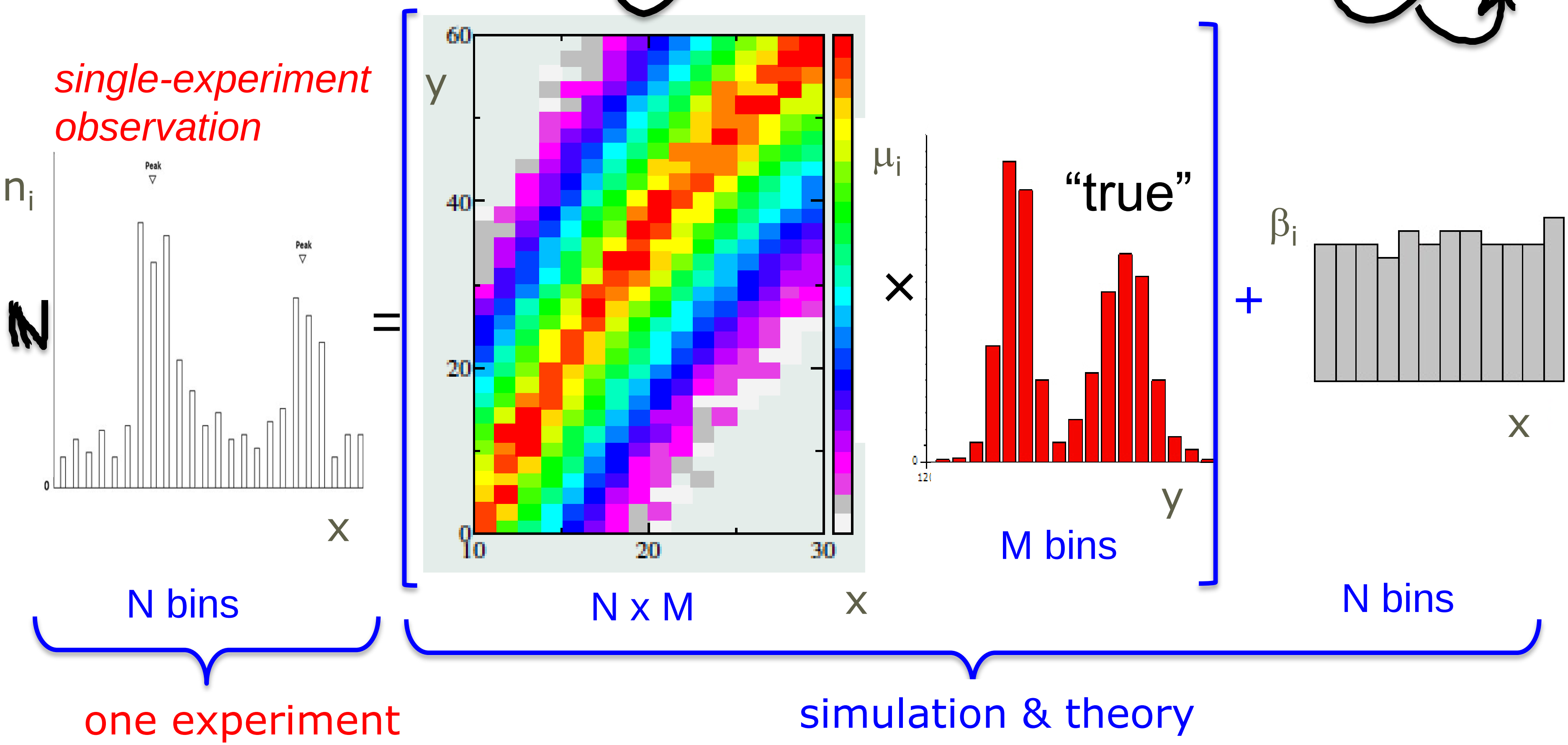
Unfolding techniques

Take care of the background $E[\vec{n}] = \vec{\nu} = \tilde{R}\vec{\mu} + \vec{\beta}$



Unfolding techniques

With pictures… $E[\vec{n}] = \vec{\nu} = \tilde{R}\vec{\mu} + \vec{\beta}$ → now estimate $\vec{\nu} \to \hat{\vec{\nu}}$



Unfolding techniques

- Now we should invert the equation to estimate $\vec{\mu} \rightarrow \hat{\vec{\mu}}$

$$E[\vec{n}] = \vec{\nu} = \tilde{R}\vec{\mu} + \vec{\beta} \quad \rightarrow \quad \vec{\mu} = \tilde{R}^{-1} \cdot (\vec{\nu} - \vec{\beta})$$

- Suppose every n_i has a Poisson distribution, then the maximum likelihood estimator of $\vec{\nu}$ is simply the observed $\vec{n}$

$$P(n_i|\lambda = \nu_i) = \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$$

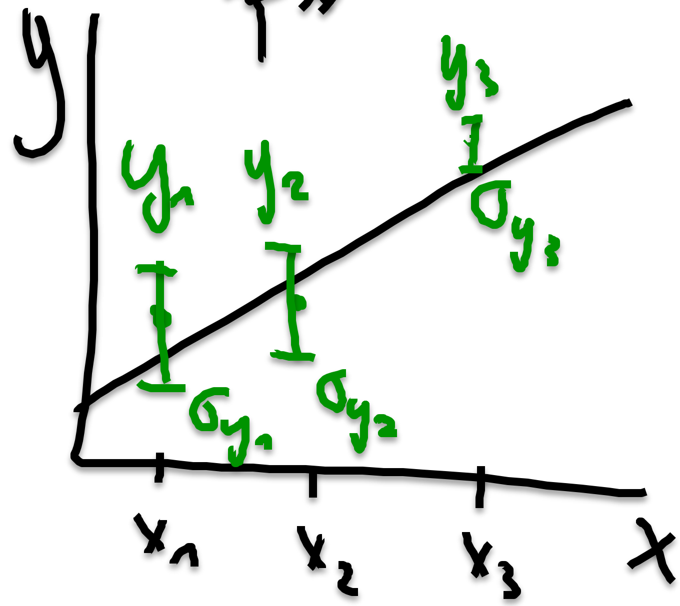
$$\left(\frac{\partial \mathcal{L}(\nu_i|n_i)}{\partial \nu_i} \right)_{\nu_i = \hat{\nu}_i} = 0 \quad \rightarrow \quad \hat{\nu}_i = n_i$$

- Hence we have $\hat{\vec{\mu}} = \tilde{R}^{-1} \cdot (\vec{n} - \vec{\beta})$
- One can show that this estimator is the most efficient one (i.e. with the smallest variance) among all unbiased estimators
- But…

$$y = ax + b$$

$$\chi^2 = \sum_{i=1}^{k}\left(\frac{y_i - (a x_i + b)}{\sigma_{y_i}}\right)^2$$

$$\chi^2(a,b) \longrightarrow \text{minimize}$$



$$y \to N \to f(x \mid \mu, \sigma^2) \quad \text{with } \mu \to E[X],\ \sigma^2 \to Var[X]$$

$$\ln \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

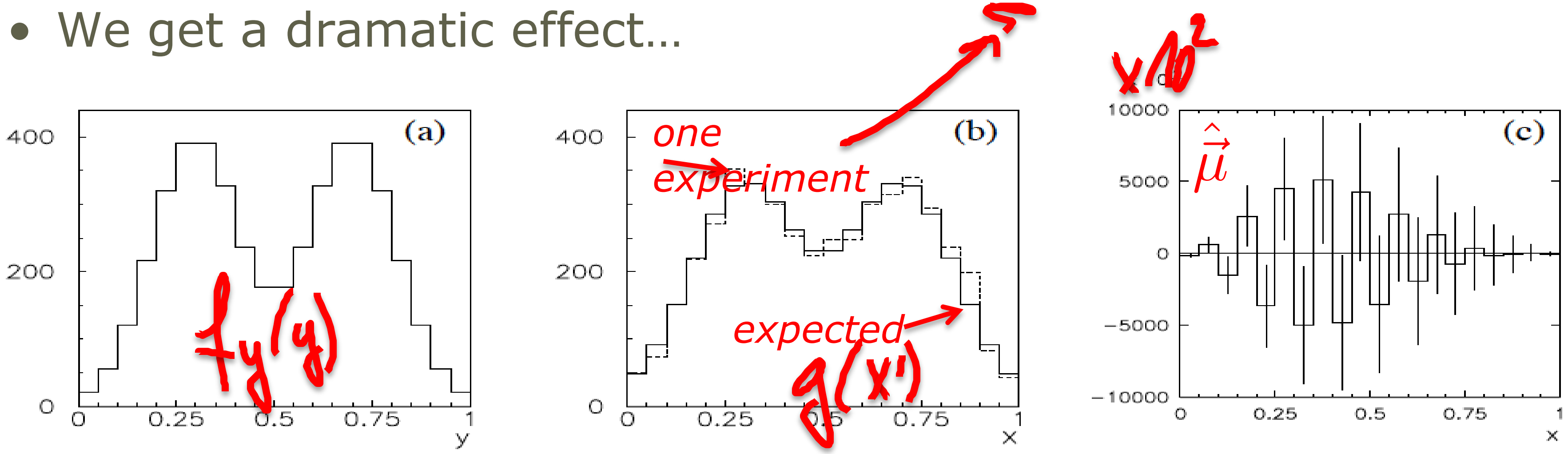
$$= c^{te} + \sum_{i=1}^{k} \left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \to \min \sum_{i=1}^{k} \left(\frac{x_i-\mu}{\sigma}\right)^2 = \chi^2$$

$$P(\vec{x} \mid \mu, \sigma^2) = \prod_{i=1}^{k} f(x_i \mid \mu, \sigma^2) = \mathcal{L}(\vec{x} \mid \mu, \sigma^2)$$

maximize $\Rightarrow \ln \mathcal{L}(\vec{x} \mid \mu, \sigma^2)$

Unfolding techniques

- Using the simple equation $\hat{\vec{\mu}} = \tilde{R}^{-1} \cdot (\vec{n} - \vec{\beta})$
- We get a dramatic effect…



"true" distribution

observed distribution

- Huge fluctuations because of the fluctuations of $\vec{n}$ around $\vec{\nu}$
- Introduce some bias (systematic uncertainty) to reduce the variance (statistical uncertainty) → regularized unfolding

Regularized Unfolding

Still use the same master equation $\hat{\vec{\mu}} = \tilde{R}^{-1} \cdot (\hat{\vec{\nu}} - \vec{\beta})$

But adapt the Likelihood in the Maximum Likelihood method

$$ln\mathcal{L}(\vec{\nu}|\vec{n}) \longrightarrow ln\mathcal{L}'(\vec{\nu}|\vec{n}) = \alpha ln\mathcal{L}(\vec{\nu}|\vec{n}) + S(\vec{\nu})$$

with α is the regularization parameter and S(μ) the regularization function. Hence $\hat{\vec{\nu}} \neq \hat{\vec{\nu}}_{MaxLik} = \vec{n}$, but the solution of

$$\left(\frac{\partial ln\mathcal{L}'(\nu_i|n_i)}{\partial \nu_i}\right)_{\nu_i=\hat{\nu}_i} = 0 \quad \longrightarrow \quad \hat{\nu}_i = \ldots$$

Different options for the regularization parameter and function.

Examples in eg. arXiv:hep-ex/0208022 and arXiv:1104.2962

Regularized Unfolding

$$ln\mathcal{L}(\vec{\nu}|\vec{n}) \longrightarrow ln\mathcal{L}'(\vec{\nu}|\vec{n}) = \alpha ln\mathcal{L}(\vec{\nu}|\vec{n}) + S(\vec{\nu})$$

The regularization function needs to smoothen the likelihood. Therefore a good choice could be the mean square of the second derivative.

$$S(\vec{\nu}) \sim \sum_{i=1}^{N-2} (-\nu_i + 2 \cdot \nu_{i+1} - \nu_{i+2})^2$$

This is related to the amount of curvature in the expectation value of $\vec{\nu}$ versus the measurant X.

The value for the regularization parameter is a trade-off between variance and bias. Hence it can be chosen to minimize both simultaneous.

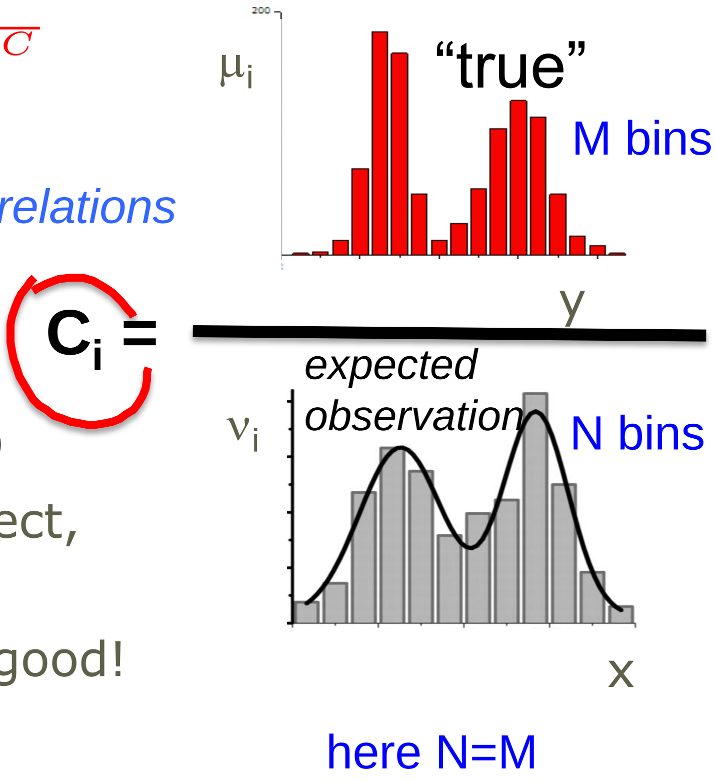
$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (Var[\hat{\mu}_i] + \hat{b}_i^2) \quad \text{with} \quad b_i = E[\hat{\mu}_i] - \mu_i$$

Alternative to unfolding

- Using simulations a simple method exist with multiplicative correction factors.

$$\hat{\mu}_i = C_i (n_i - \beta_i) \qquad C_i = \frac{\mu_i^{MC}}{\nu_i^{MC}}$$

simple method does not take into account all correlations



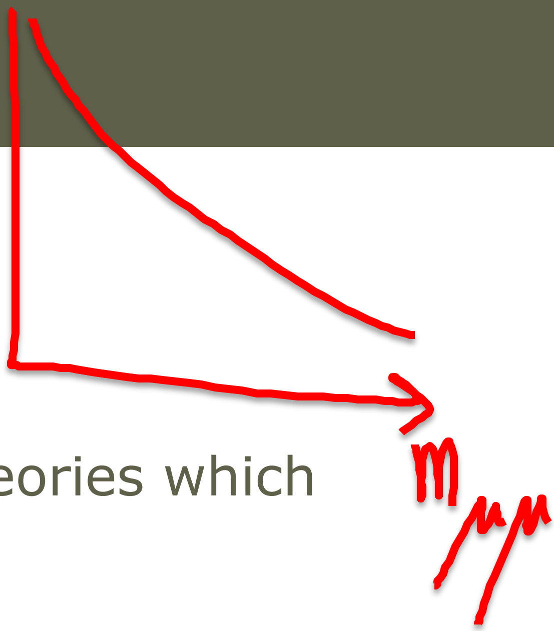
- The bias of the estimator is

$$b_i = \left(\frac{\mu_i^{MC}}{\nu_i^{MC}} - \frac{\mu_i}{\nu_i^{sig}} \right) \nu_i^{sig} \qquad (\nu_i^{sig} = \nu_i - \beta_i)$$

Only no bias if the MC simulated model is correct, and the estimators are pulled towards the MC model used for the correction factors C_i … not good!

→ simply compare $\vec{n}$ directly with $\vec{\nu}$

Why unfolding ?



① Comparison between experiments

① Comparison with new theories (or other theories which experimentalists have not considered)

Unfolding in ROOT: RooUnfold (arXiv:1105.1160)

Examples from the LHC

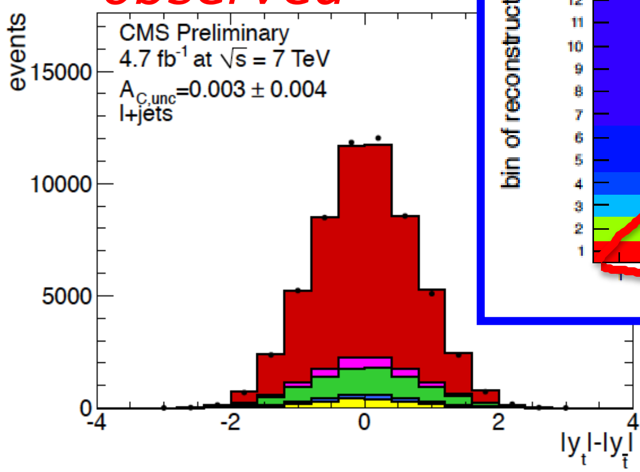
arXiv:1207.0065

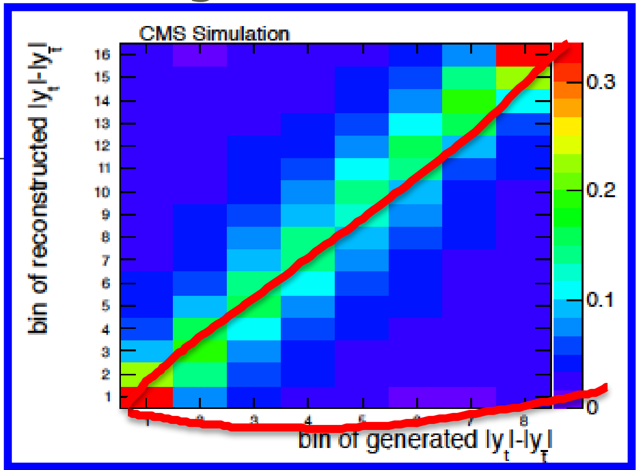
The charge asymmetry A_C in top quark pair events:

$$A_C = \frac{N^+ - N^-}{N^+ + N^-}$$

where $N^{+/-}$ reflects the amount of events for which $\Delta|y| = |y_t| - |y_{\bar{t}}|$ is either positive or negative.

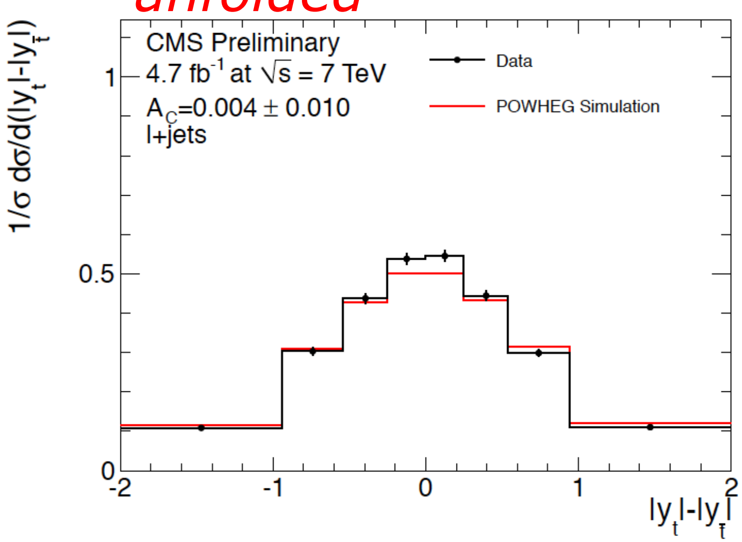
observed





response matrix

unfolded



Addendum: Rivet

http://rivet.hepforge.org/

"Rivet is a library and set of programs which produce simulated distributions which can be directly compared to measured data for MC validation and tuning studies. It can also be used without reference data to compare two or more generators to each other for regression testing or tune comparison. The Rivet library contains the tools needed to calculate physical observables from HepMC files or objects, a large set of important experimental analyses, and histogramming tools for data comparison and presentation."

Summary of part 1

Summary of part 1

Theory needs to meet experiment!

① Nature at the quantum level is stochastic … deal with it!

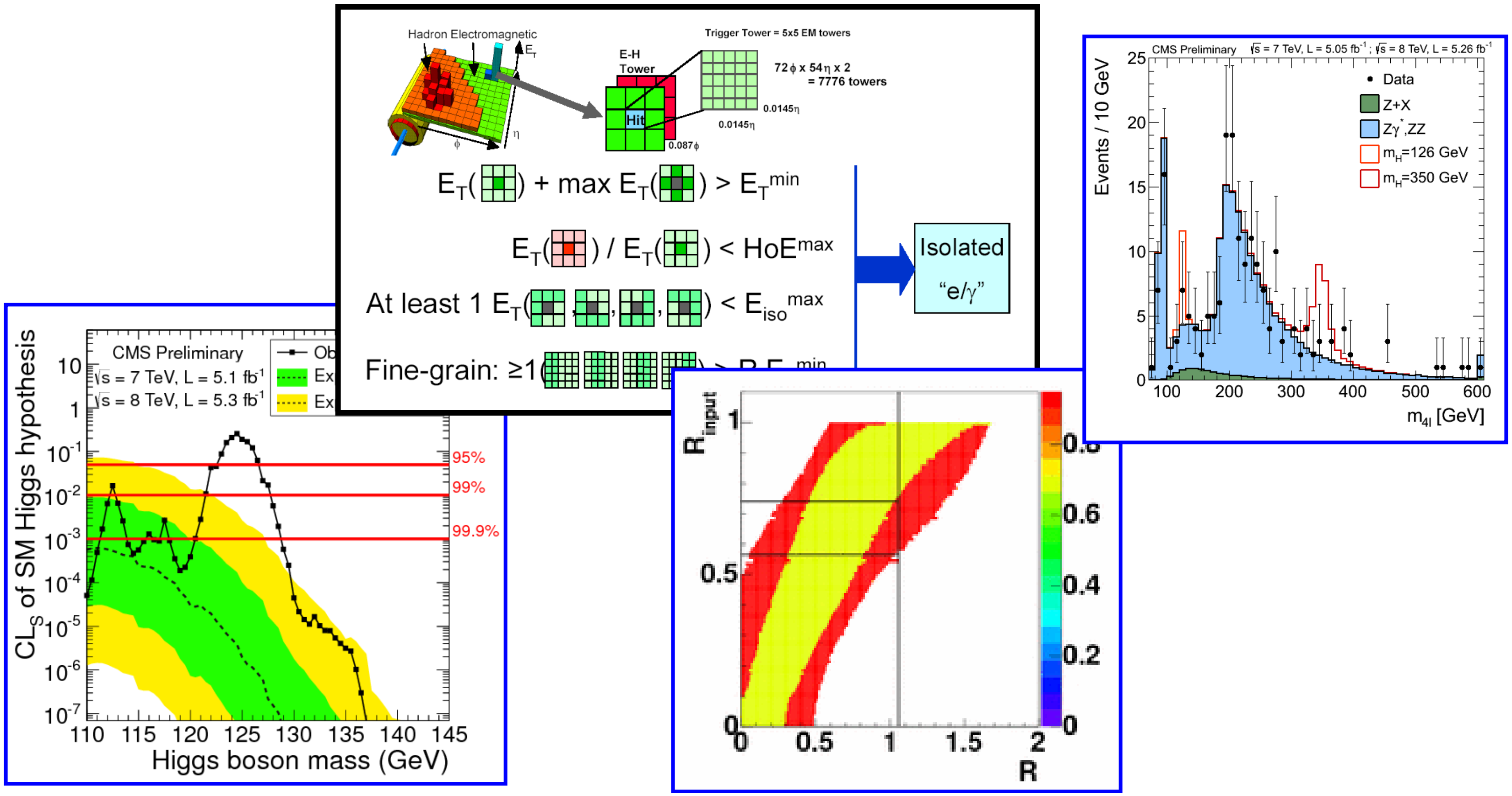
② Fundamental physics is reflected in distributions $f_X(x|\theta)$

③ Detector nuisance effects: $g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$

④ The distributions $g(x'|\vec{\alpha},\theta)$ are measurable

⑤ Unfolding techniques can give you $f_X(x|\theta)$ back…

Lecture 2 – pre-view



Lecture 2

The basics in these lectures

Part1 : "**Theory meets experiment**"

① Our QFT description of Nature is a stochastic one
② General stochastic distributions in physics
③ From theoretical to experimental distributions
④ ... and back: unfolding techniques
⑤ Examples from the LHC at CERN

Part 2 : "**Experiment meets theory**"

① Experimental aspects to accumulate experimental data
② Selection of the dedicated signal
③ Performing measurements & parameter estimation
④ Claiming a discovery of new physics or setting limits
⑤ Examples from the LHC at CERN

"Experiment meets theory"

physics/0311105

① Collecting experimental data

From detectors, over triggers to collision data at the LHC

② Selection of dedicated signals

Hypothesis testing, efficiency versus purity, use of Monte Carlo simulation, optimal event selection, significance

③ Performing measurements

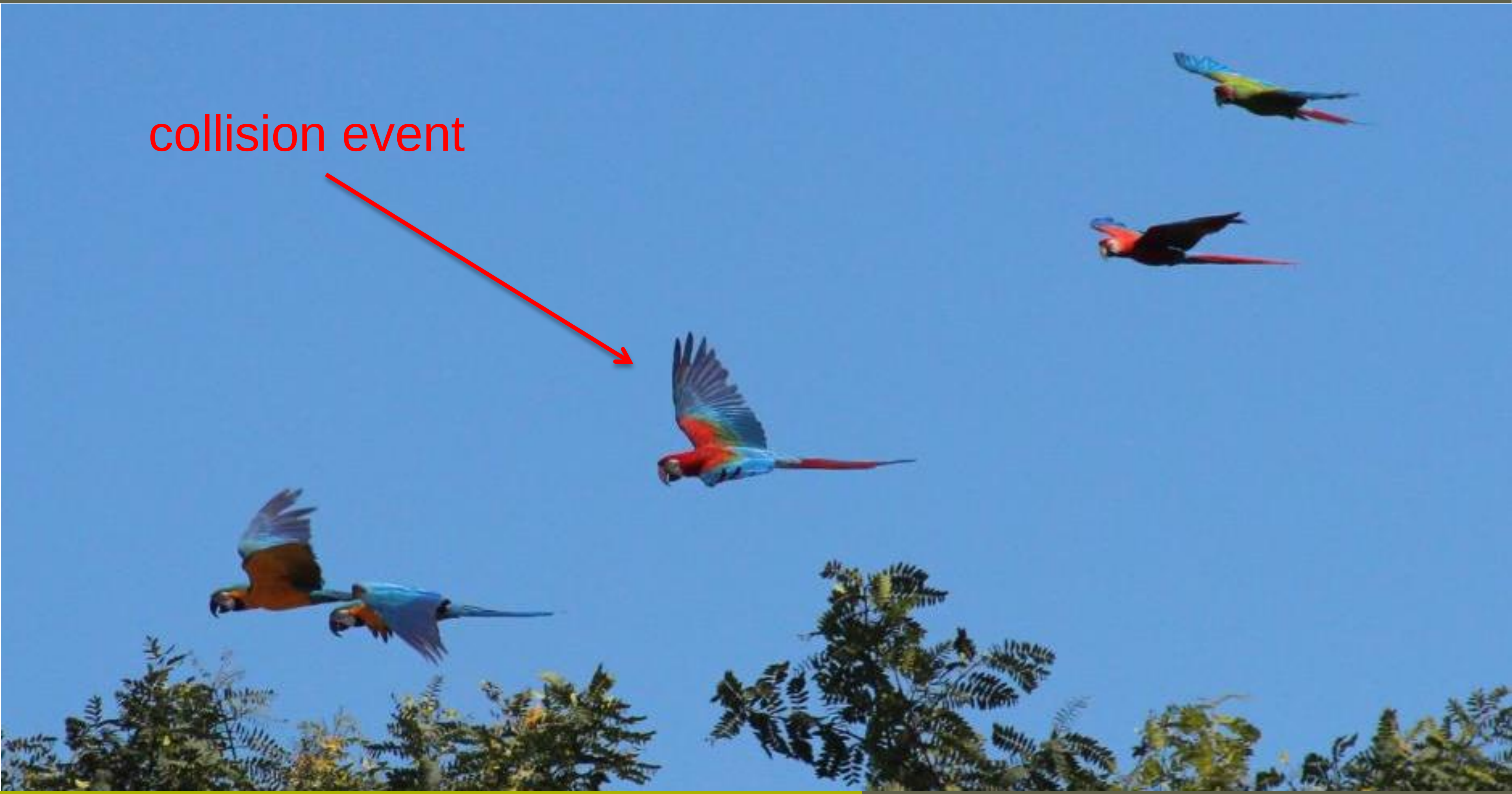
Concepts of parameter estimation, counting experiments versus fitting distributions, least-square and maximum likelihood method, Neyman confidence belts (extension to Feldman-Cousins method)

④ Claiming discovery or setting limits

Systematic uncertainties and significance, the CLs method

The Trigger… with pictures

The Trigger… with pictures



The Trigger… with pictures

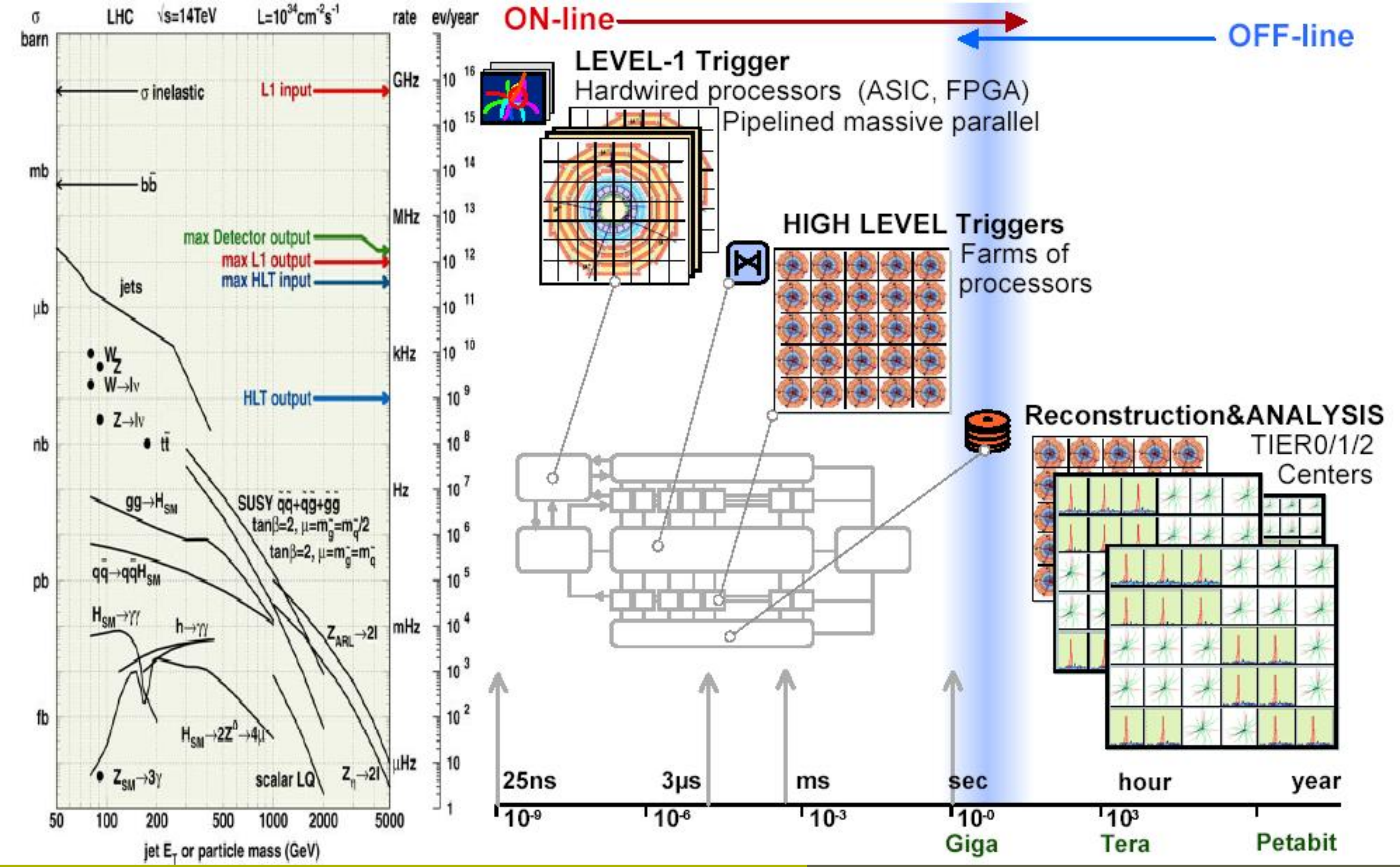
The Trigger… with pictures

The Trigger… with pictures

The Trigger… with pictures

sometimes bad luck…
too much interesting collisions

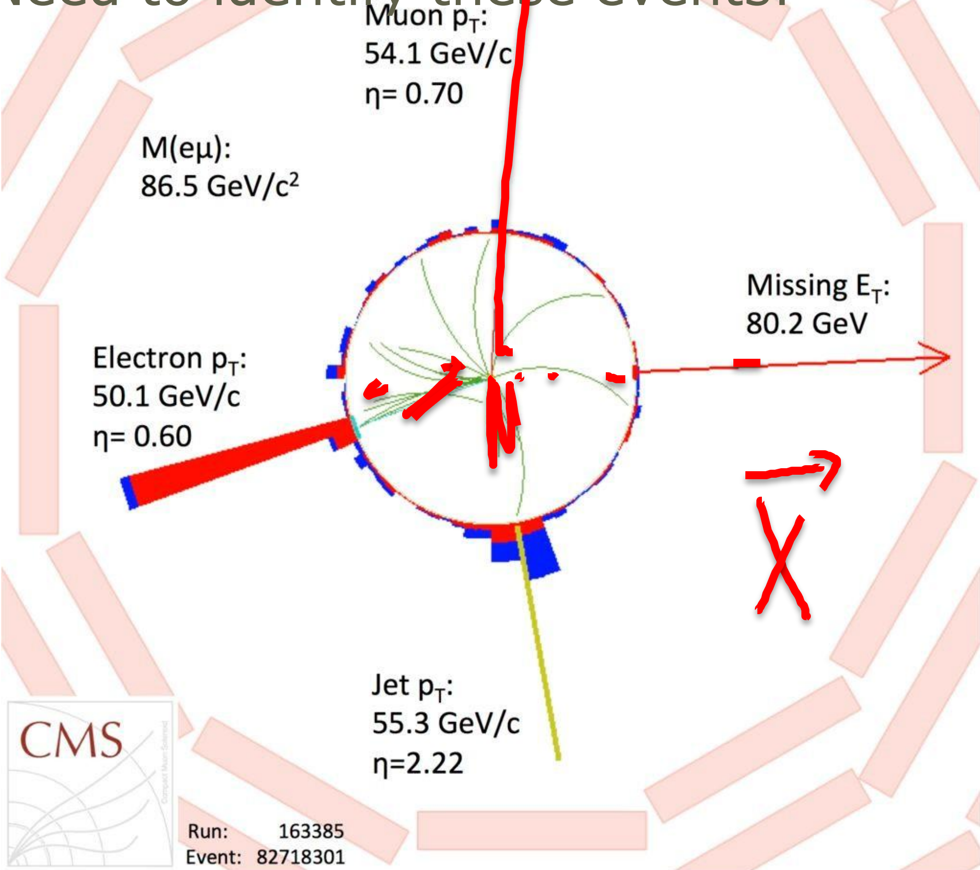
The Trigger: general layout

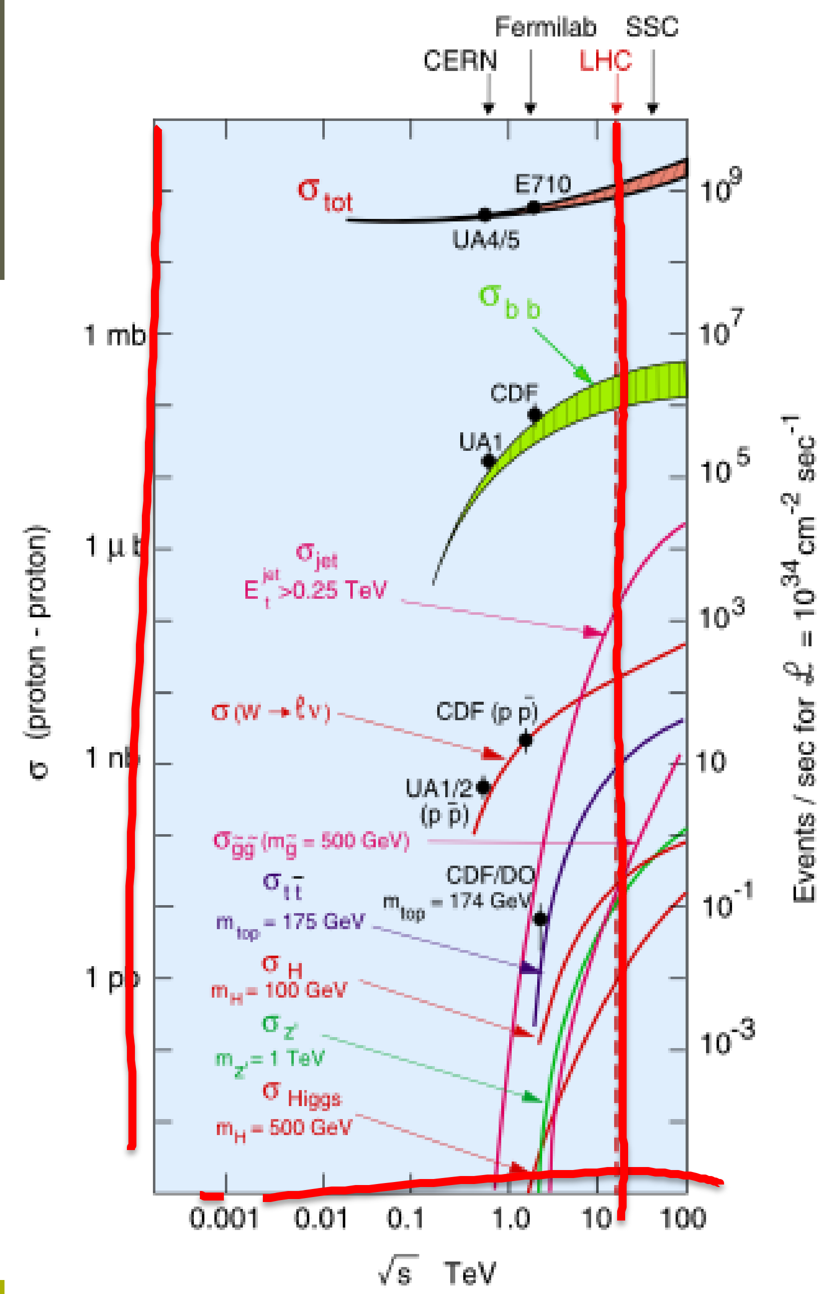


Signal selection

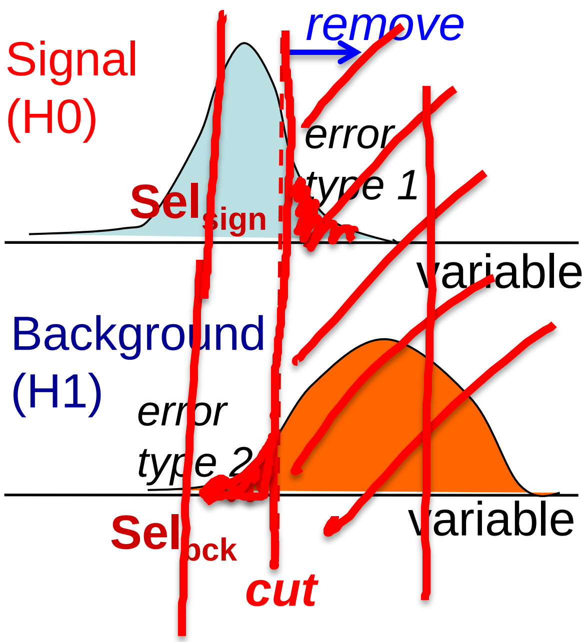
Typically we want to study rare events

Need to identify these events!





Hypothesis testing



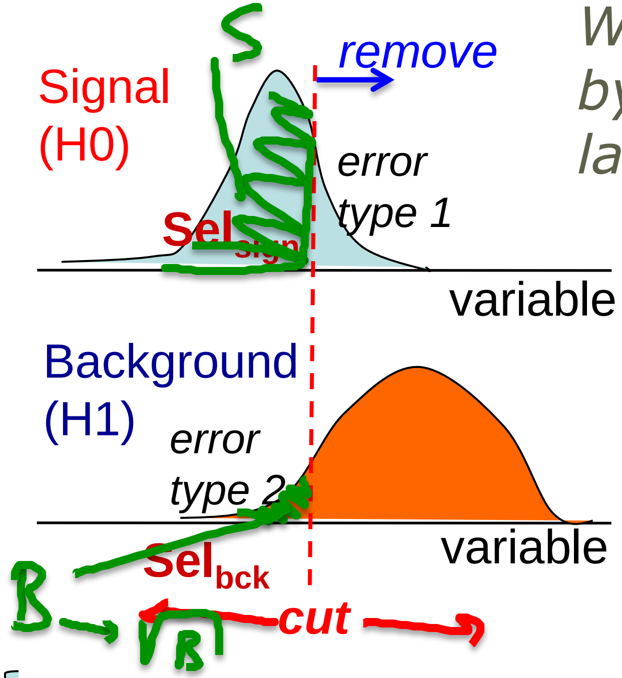
{ "Error of type-1": rejecting a signal event
{ "Error of type-2": accepting a background event

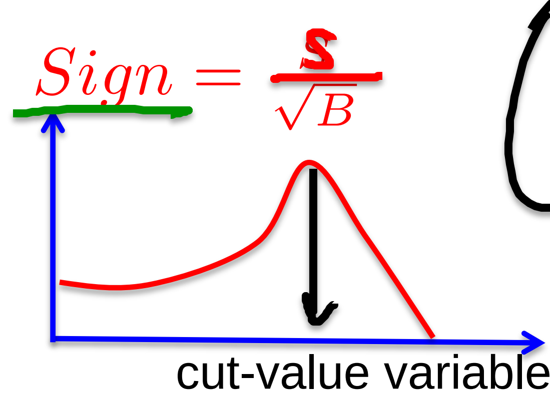
or

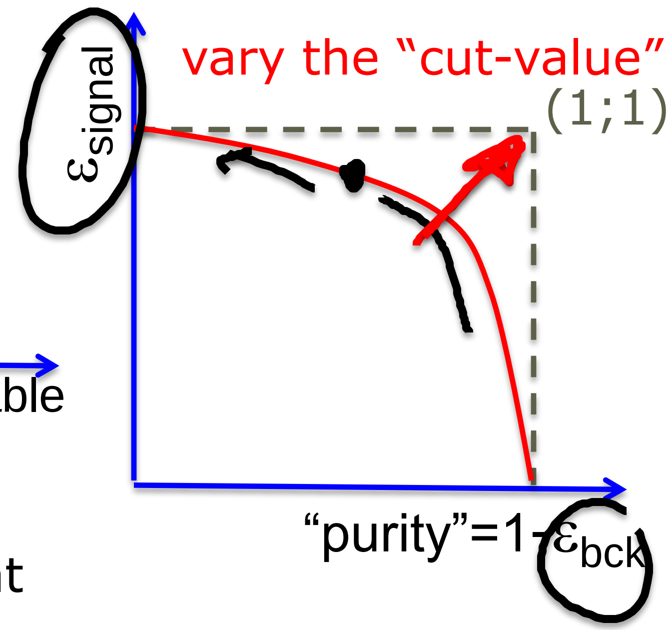
{ ε_{signal} = signal efficiency = # Sel$_{sign}$ / total # signal events
{ ε_{bck} = background efficiency = # Sel$_{bck}$ / total # background events

Hypothesis testing

Where to cut on this variable? For a discovery by counting the number of events, need the largest significance







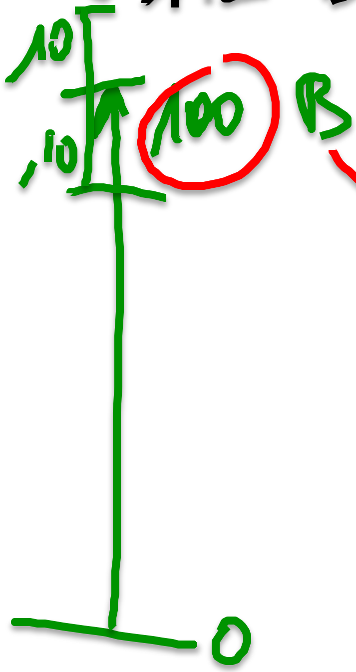
"Error of type-1": rejecting a signal event
"Error of type-2": accepting a background event

or

ε_{signal} = signal efficiency = # Sel_{sign} / total # signal events
ε_{bck} = background efficiency = # Sel_{bck} / total # background events

• 150 DATA → 5 σ

• 112 DATA → 1.2 σ



$$\sigma_B^{TH} = B \text{ pb} \pm \Delta B \text{ pb}$$

Significance

The significance is influenced by systematic uncertainties

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

used to obtain the simulated distributions of variable x'

The imprecision of the model of R and f_X has to be taken into account when you estimate the background (and signal) rate after the event selection cuts → for example an effect of $\Delta_{syst}B$

The significance is the amount of "standard deviations" the signal excess (=S) is above the background level (=B).

$$Sign = \frac{S}{\sqrt{Var[B]}} = \frac{S}{\sqrt{B}} \longrightarrow \frac{S}{\sqrt{Var_{stat}[B]+Var_{syst}[B]}} = \frac{S}{\sqrt{B+(\Delta_{syst}B)^2}}$$

Poisson distribution

remember: $Var[X] = \sigma_X^2$

Significance

$$Sign = \frac{S}{\sqrt{Var[B]}} = \frac{S}{\sqrt{B}} \longrightarrow \frac{S}{\sqrt{Var_{stat}[B]+Var_{syst}[B]}} = \frac{S}{\sqrt{B+(\Delta_{syst}B)^2}}$$

B = 100 → $\sqrt{B}$ = 10

S = 30

Δ_{syst}B = 0

→ Sign = 3 "sigma"

B = 100

S = 30

Δ_{syst}B = **10**

(hence 10% uncertainty on Bck rate)

→ Sign = 2.1 "sigma"

Significance

$$Sign = \frac{S}{\sqrt{Var[B]}} = \frac{S}{\sqrt{B}} \longrightarrow \frac{S}{\sqrt{Var_{stat}[B]+Var_{syst}[B]}} = \frac{S}{\sqrt{B+(\Delta_{syst}B)^2}}$$

- B = 100
- S = 30
- Δ_{syst}B = 0

→ Sign = 3 “sigma”

- B = 100
- S = 30
- Δ_{syst}B = **20**

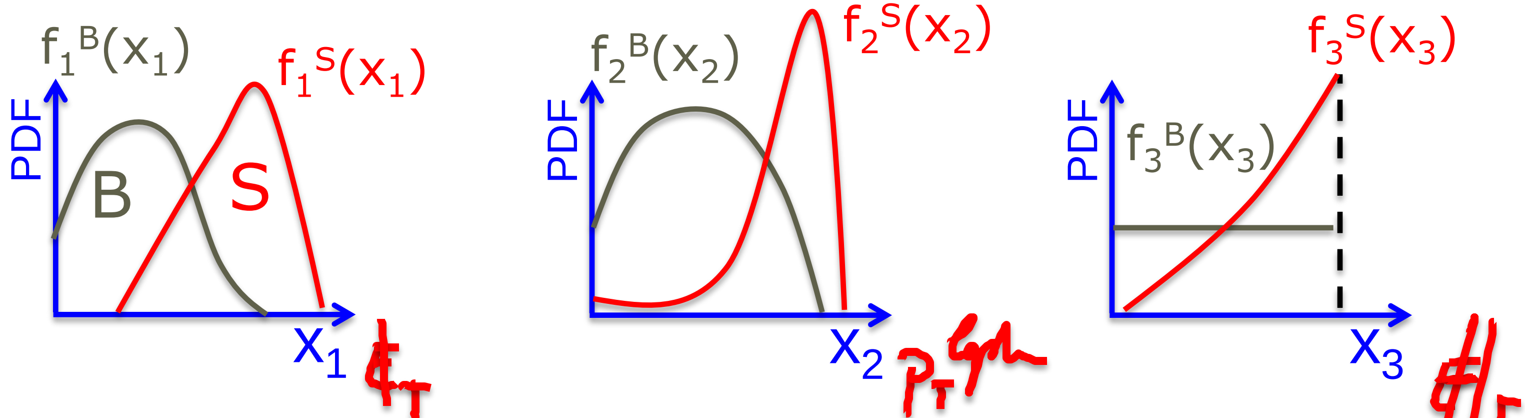
(hence 20% uncertainty on Bck rate)

→ Sign = 1.3 “sigma”

Optimal selection

Multi-Variate Analysis → combine variables $\vec{x}$ into a single variable

(optimal variable is the Likelihood Ratio variable from the optimal Neyman-Pearson hypothesis test)

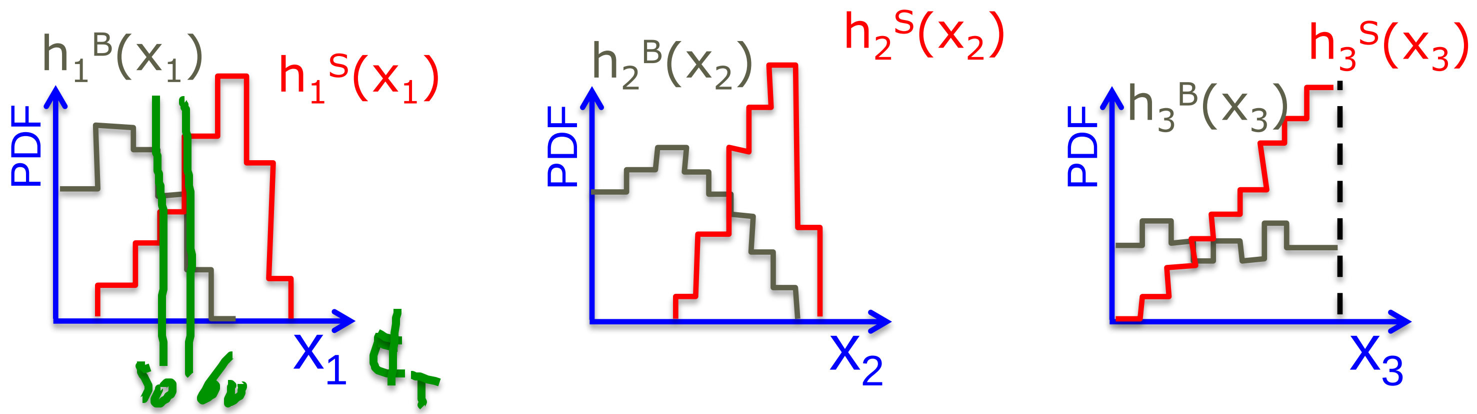


① functional form $f_i(x_i)$ not known because from ∞ events
use Monte Carlo simulation to approximate this

Optimal selection

Multi-Variate Analysis → combine variables $\vec{x}$ into a single variable

(optimal variable is the Likelihood Ratio variable from the optimal Neyman-Pearson hypothesis test)



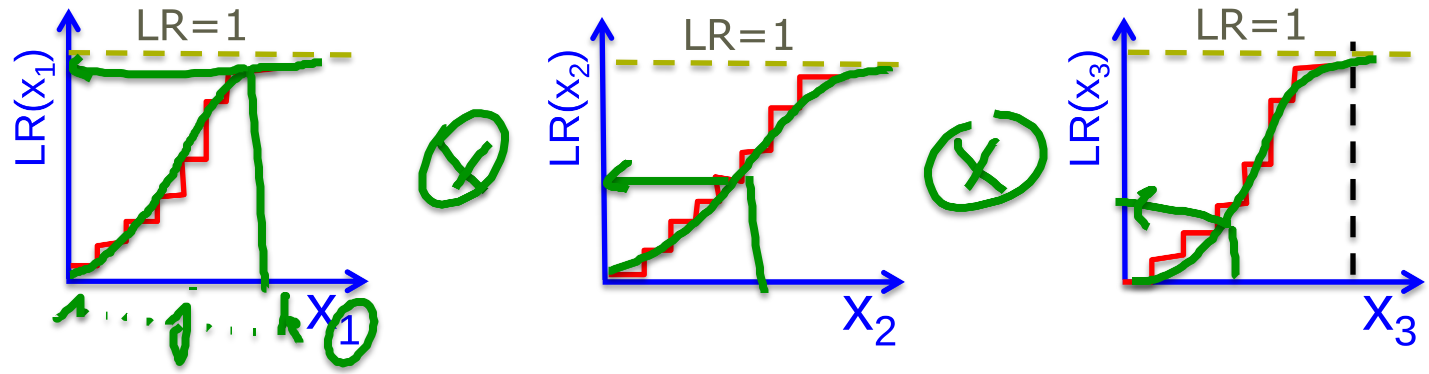
② histograms $h_i(x_i)$ from finite amount of events (normalized & use lots of simulation to get a fine binning)

Optimal selection

Multi-Variate Analysis → combine variables $\vec{x}$ into a single variable

(optimal variable is the Likelihood Ratio variable from the optimal Neyman-Pearson hypothesis test)

$$bin\ j \quad \rightarrow \quad \mathcal{LR}_i(x_{i,j}) = \frac{h_i^S(x_{i,j})}{h_i^S(x_{i,j}) + h_i^B(x_{i,j})} \quad \longrightarrow \quad fit \quad \longrightarrow \quad \mathcal{LR}_i(x_i)$$



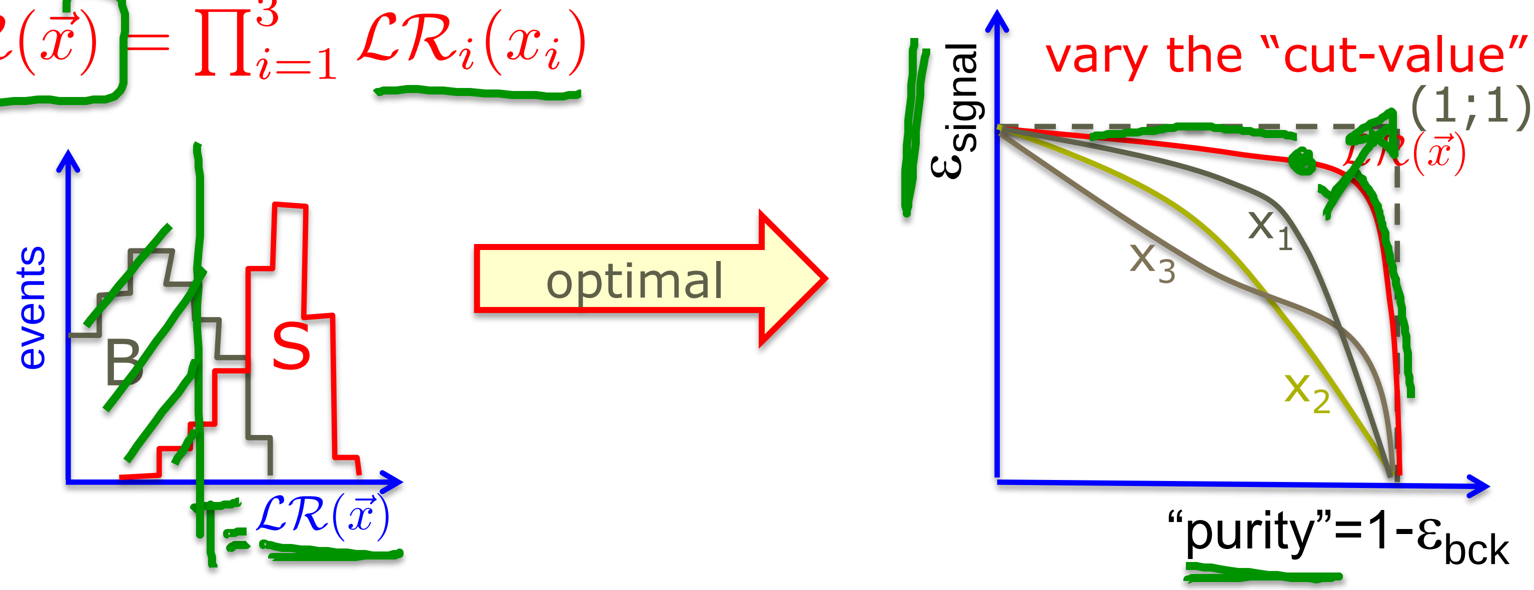
③ use histograms to calculate bin-per-bin a Likelihood Ratio

④ fit these tendencies with an adequate function $\mathcal{LR}_i(x_i)$

Optimal selection

Multi-Variate Analysis → combine variables $\vec{x}$ into a single variable

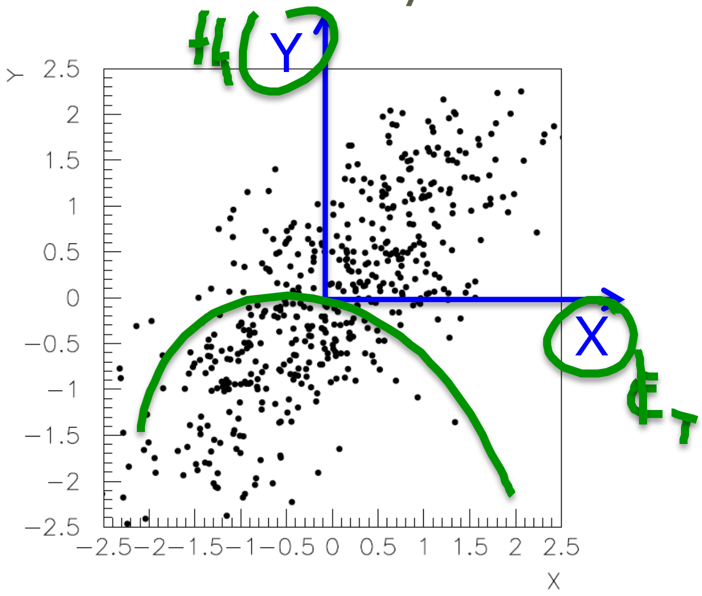
$$\mathcal{LR}(\vec{x}) = \prod_{i=1}^{3} \mathcal{LR}_i(x_i)$$

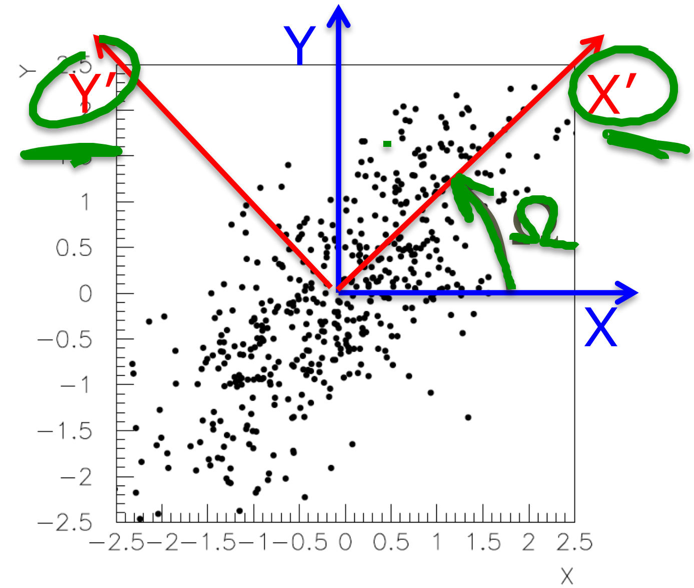


⑤ combined into one variable per event $\mathcal{LR}(\vec{x})$
most optimal variable if the variables x_i are not correlated

Optimal selection

Multi-Variate Analysis → combine variables $\vec{x}$ into a single variable





when linearly correlated, first rotate the variable space

the set of new variables $\vec{X}' = \tilde{R}\,\vec{X}$ can be used into a LR method

(other MVA methods: Neural Networks, Boosted Decision Trees, Fisher Discriminant, … MVA in ROOT: http://tmva.sourceforge.net/)

Now we have the selected data

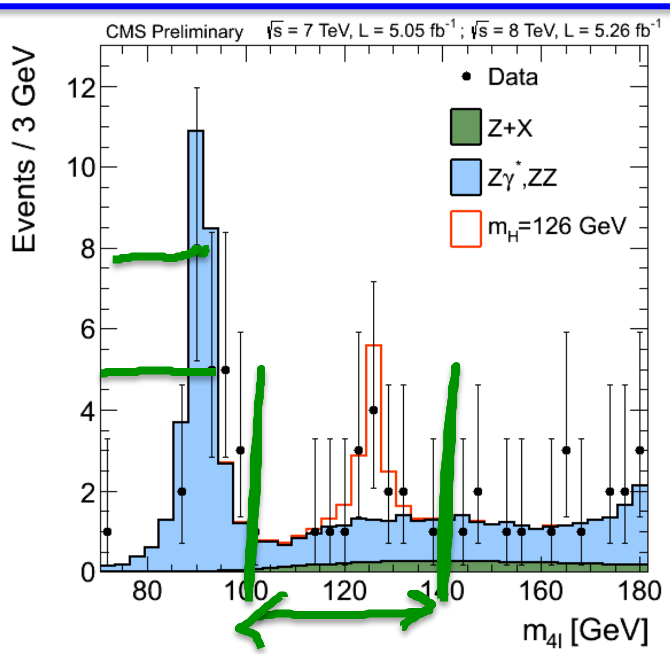
N_{DATA} events

$\varepsilon_{signal}^{MC}$ ε_{bck}^{MC}

(*typically parts from data itself*)

σ_{signal} σ_{bck}

L_{int}



counting experiments

study the shape

(in general you estimate parameters of the signal)

Parameter estimation: the basics

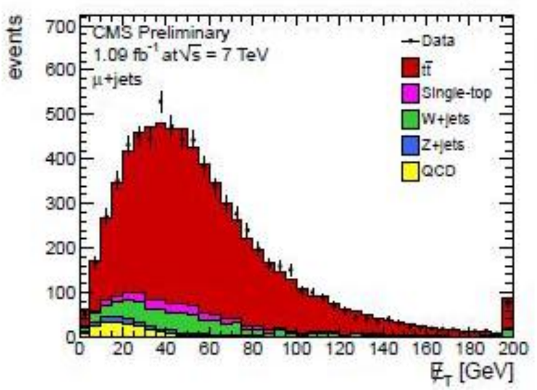
$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

could be one number per experiment,
or a set of numbers per experiment

one of these
is to be estimated

① number of selected events in data

② distribution of an observable variable per event

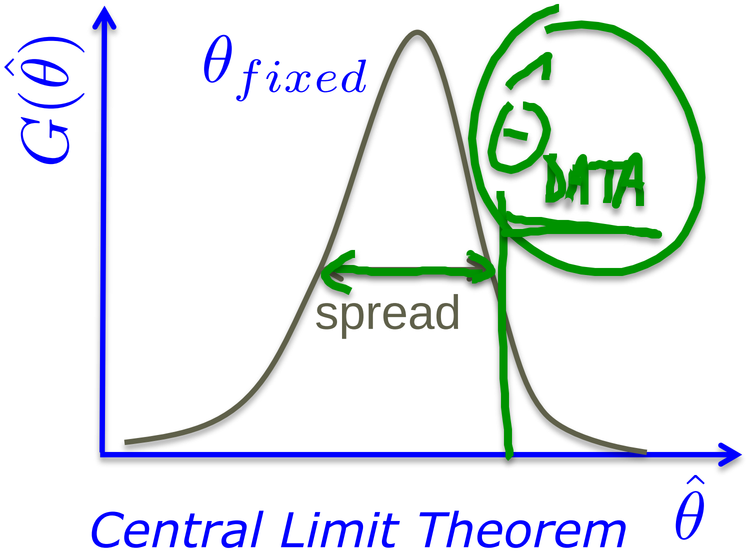


construct an estimator which depends on the observed value(s) of x' … hence the estimator itself is a stochastic variable

Parameter estimation: the basics

The estimator of θ is function of the measurements {x'}: $\hat{\theta} = F(\{x'\})$

It is a stochastic variable which follows a distribution : $G(\hat{\theta})$



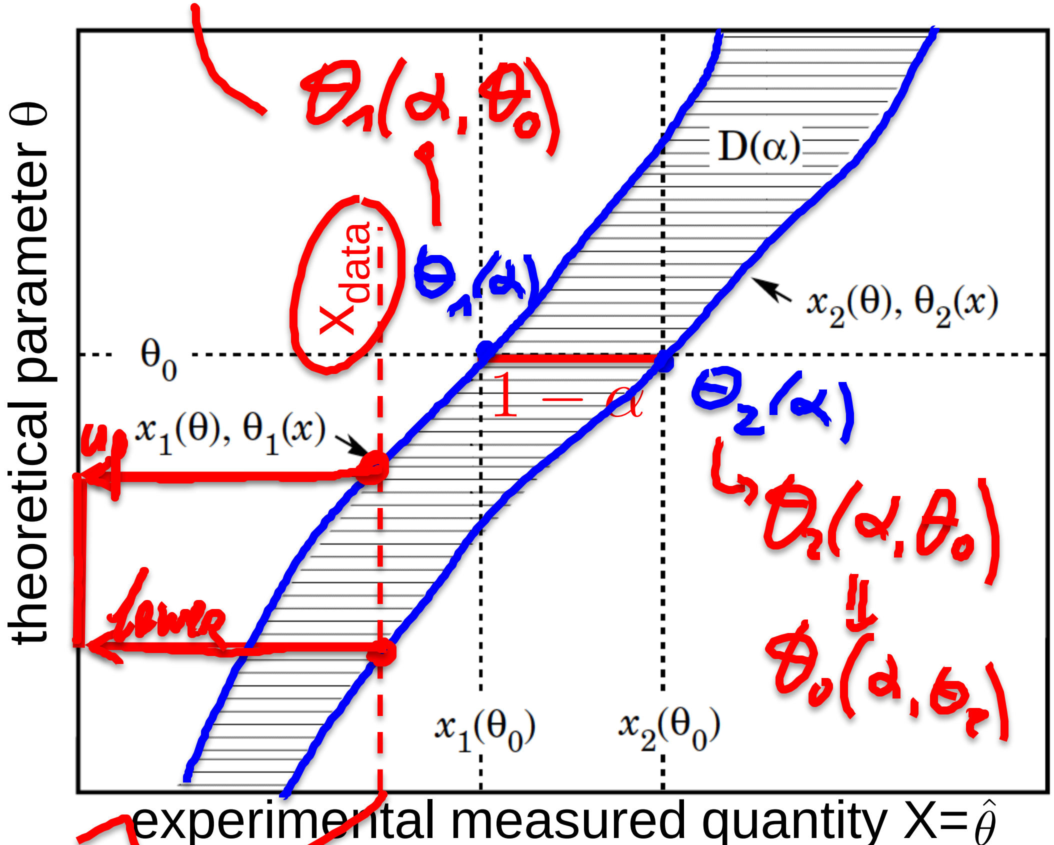
Central Limit Theorem
Gaussian distribution

The distribution $G(\hat{\theta})$ depends on the value of the physics parameter θ itself.

The uncertainty on the estimator is related to the variance of the distribution $G(\hat{\theta})$.

But how to quote an uncertainty on the physical parameter θ itself?

Confidence belts (Neyman's method)



For each θ_0, determine the distribution $G(\hat{\theta}|\theta_0)$.

Take the central 1-α percent probability.

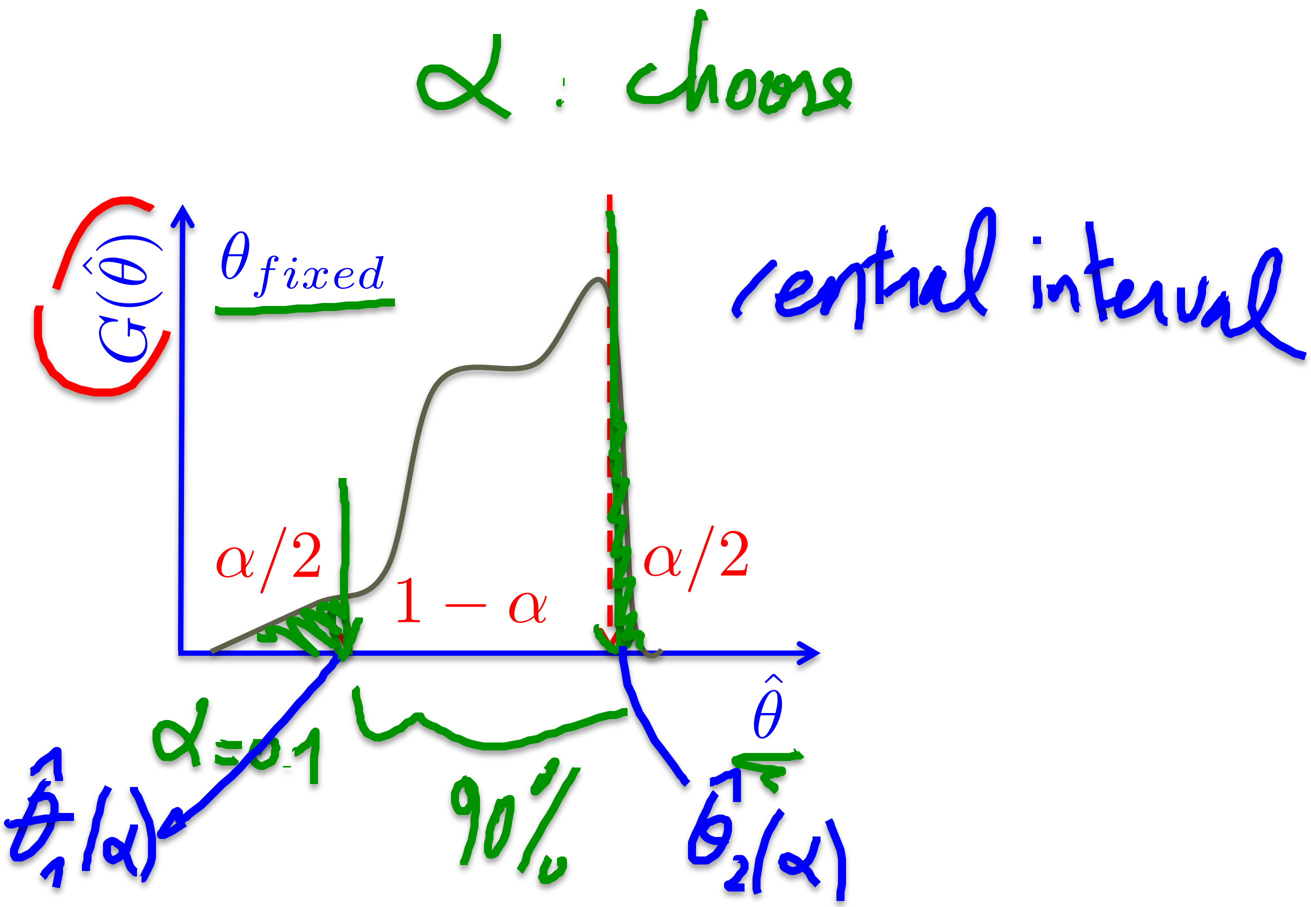
Defines $x_1(\theta_0)$ and $x_2(\theta_0)$.

These can be transformed into $\theta_1(x,\alpha)$ and $\theta_2(x,\alpha)$.

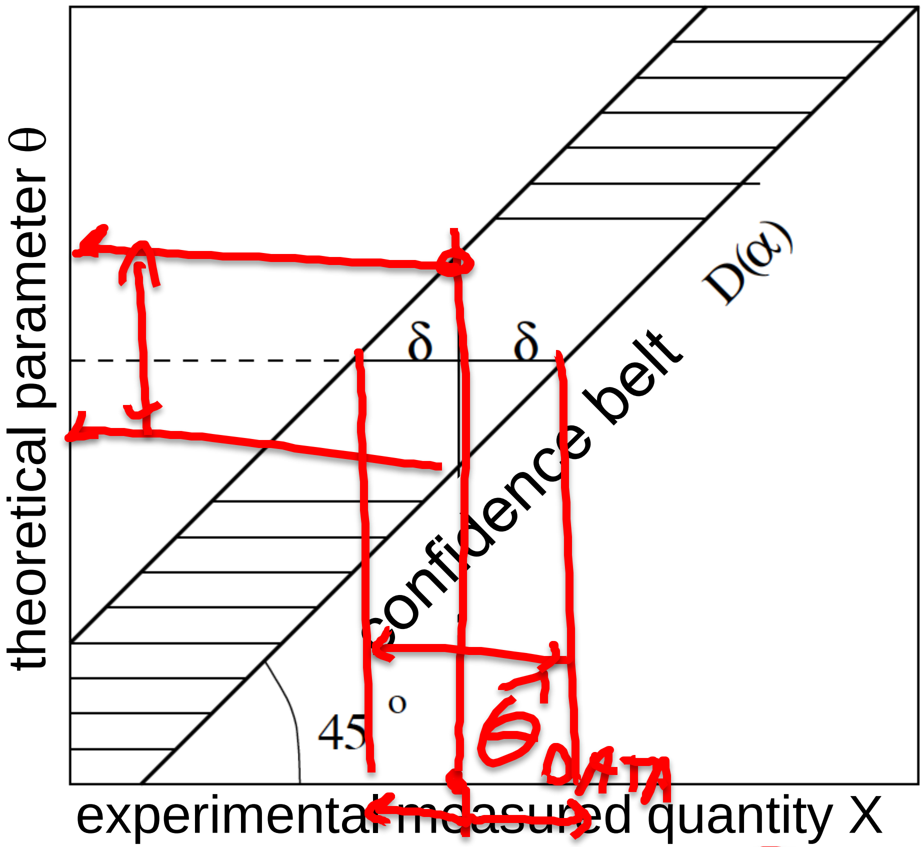
Now the data value x_{data} provides the 1-α confidence interval for the true value of θ, namely $[\theta_2,\theta_1]$.

$$1-\alpha = P(\theta_2(x_{data},\alpha) < \theta_{true} < \theta_1(x_{data},\alpha))$$

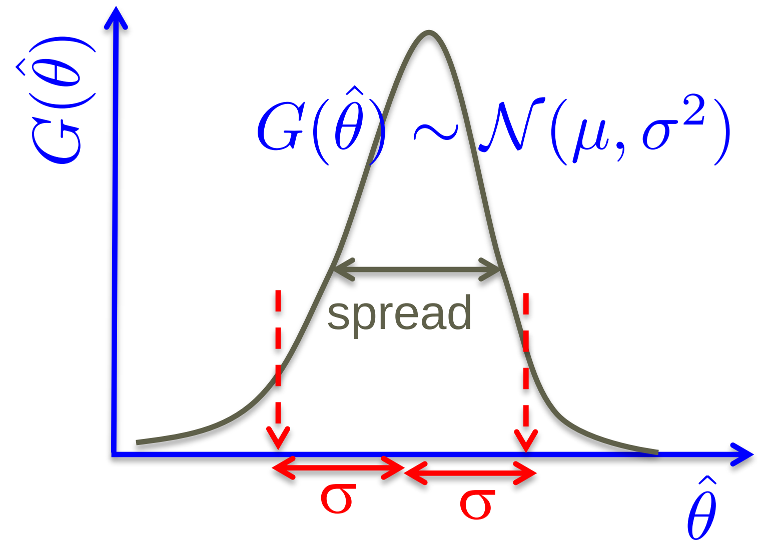
1-α interval


α : choose
central interval
θ_{fixed}
$G(\hat{\theta})$
$\alpha/2$
$1-\alpha$
$\alpha/2$
$\hat{\theta}$
α=0.1
90%
$\hat{\theta}_1(\alpha)$
$\hat{\theta}_2(\alpha)$

Confidence belts (Neyman's method)



Simplified with Gaussian distributions of the estimator.

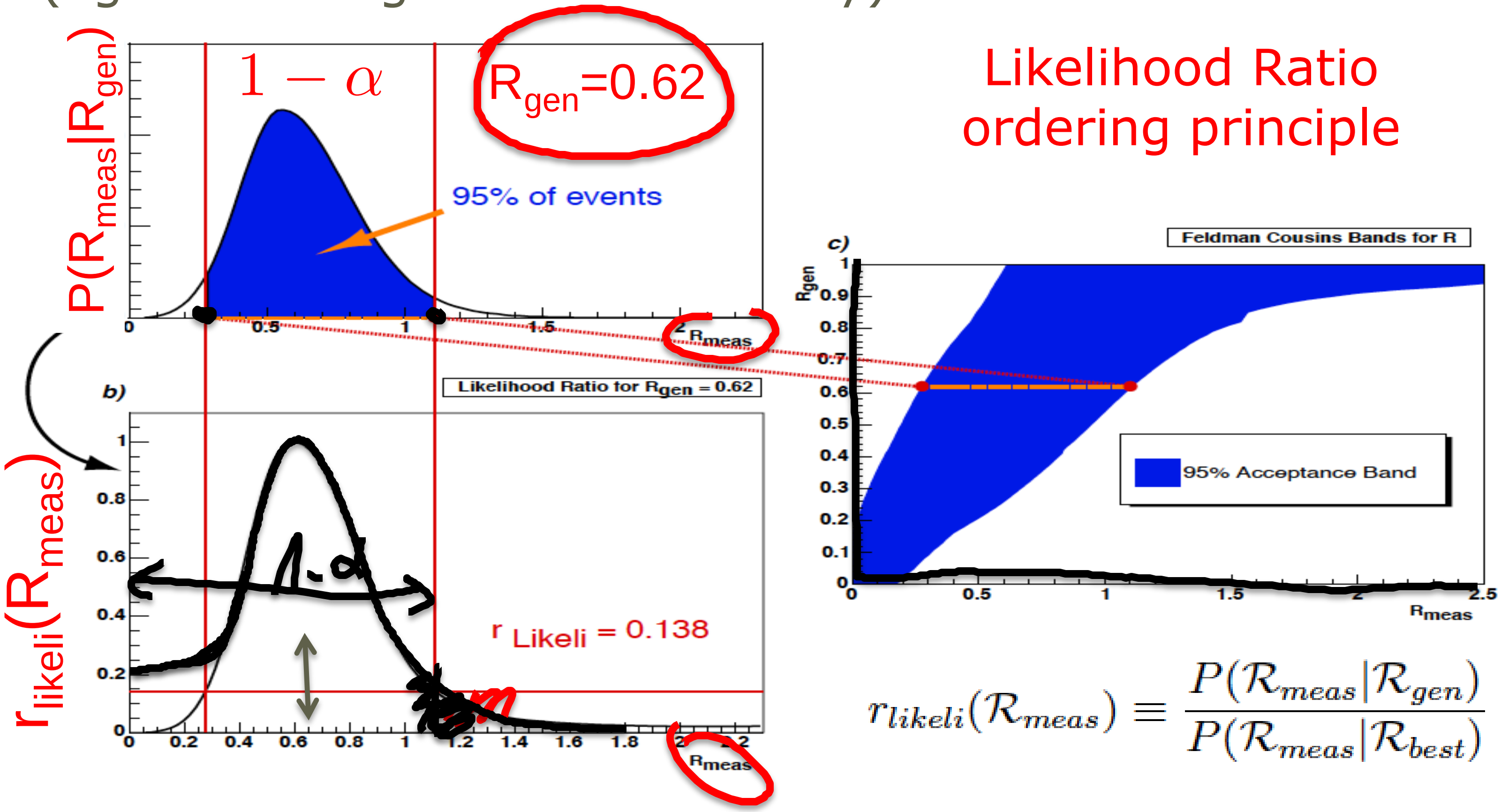


The confidence interval becomes symmetric.

$$P(x_{data} - \sigma < \theta_{true} < x_{data} + \sigma) = 68.3\%$$

Feldman-Cousins

What to do when your measurement hints for a “non-physical” signal (eg. Branching Ratio above unity): Feldman-Cousins

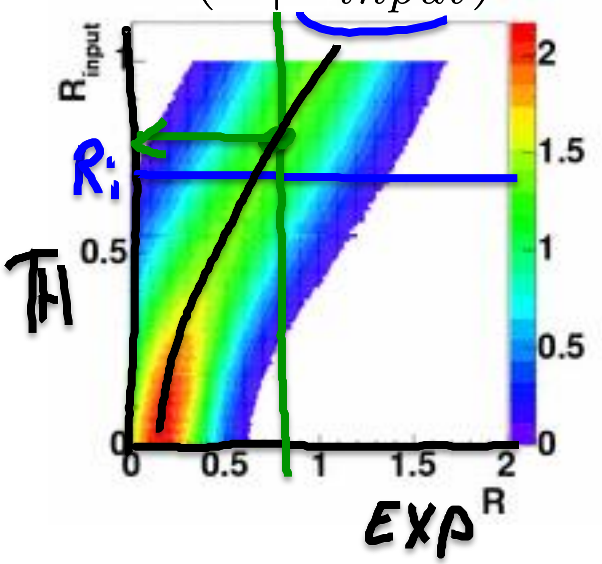


Likelihood Ratio ordering principle

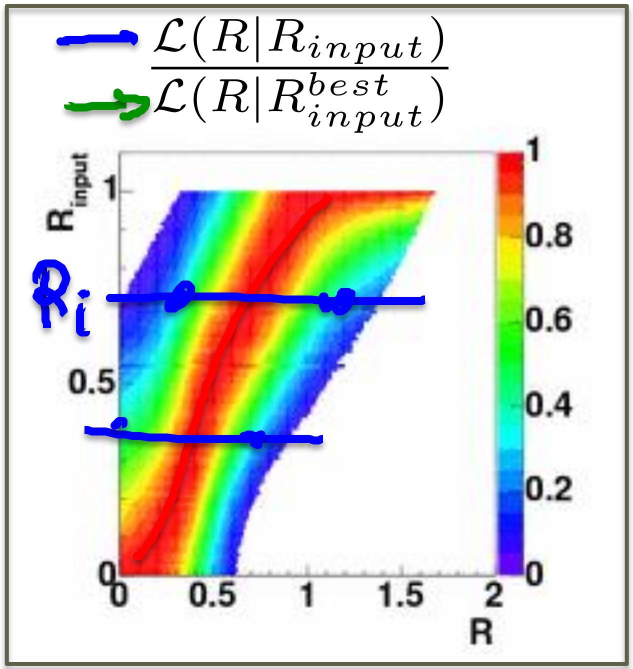
$$r_{likeli}(\mathcal{R}_{meas}) \equiv \frac{P(\mathcal{R}_{meas}|\mathcal{R}_{gen})}{P(\mathcal{R}_{meas}|\mathcal{R}_{best})}$$

Feldman-Cousins

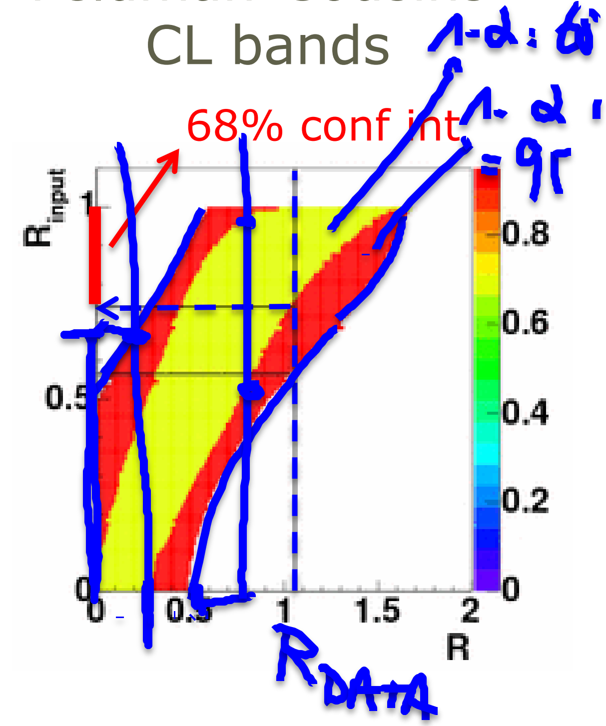
$\mathcal{L}(R|R_{input})$



Used to define the two-sided interval



Feldman-Cousins CL bands

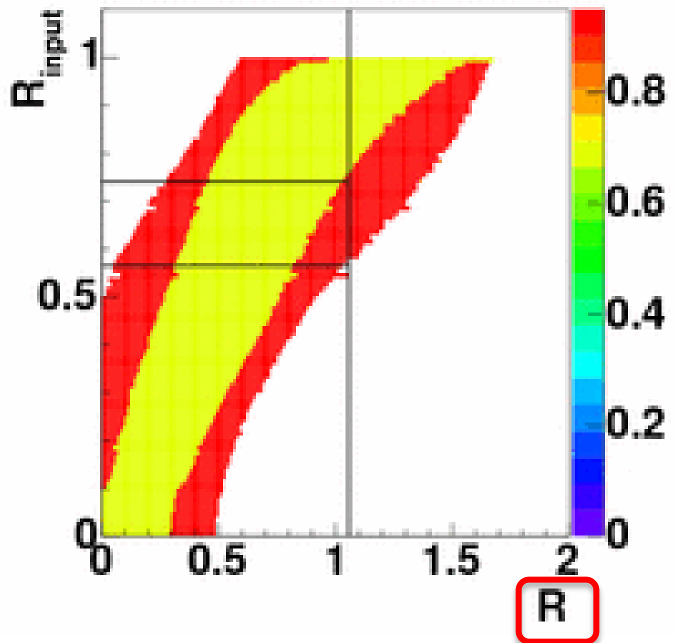


where R_{input}^{best} is the most probable value of R_{input} for a given R

(automatic switching between central and one-sided confidence regions)

Which estimator ?

$$1 - \alpha = P(R_2(R_{data}, \alpha) < R_{true} < R_1(R_{data}, \alpha))$$



this could be any estimator
(something which depends on the observed data)

Counting experiments

expected observable distribution

fixed integrated luminosity

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

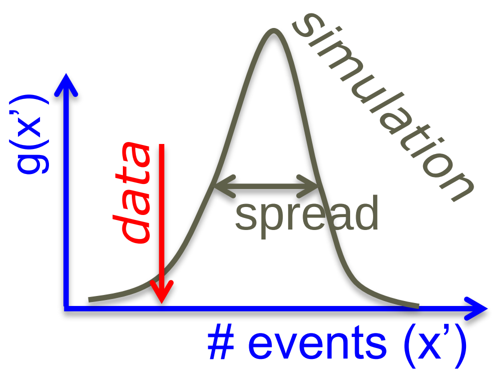
observable stochastic (number of events selected)

effect of experiment

Poisson distribution

number of events

cross section



many pseudo-experiments

$$\hat{\sigma}_{signal} = \frac{(N_{DATA} - N_{bck})}{\epsilon_{trigger} \cdot \epsilon_{signal} \cdot \mathcal{L}_{int}}$$

statistical uncertainty on estimator related to the spread of the distribution of x'

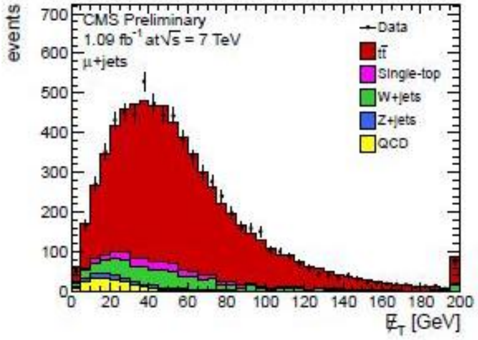
(systematics due to detector nuisance α , the model of R and theory nuisance θ)

Fitting distributions

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

If the information on the physics parameter θ of interest is hidden in a distribution g(x') rather than in an event rate, then we need to compare the expected distribution with the observed one.

binned data
(least-square method)



$\vec{n} = \{n_i\}$

i = 1 … #bins

event-per-event
(maximum likelihood method)

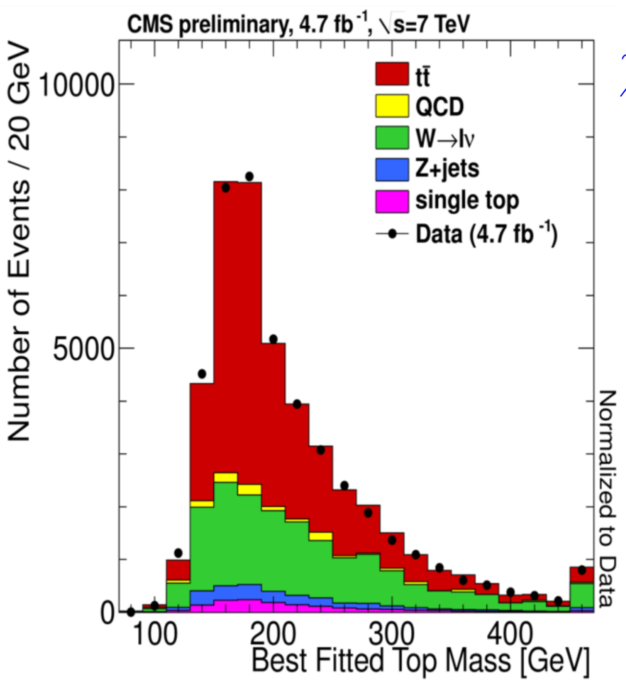
$\vec{x}' = \{x'_i\}$

i = 1 … #events

Least-square method

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

- Expectation (binned) : $g(x'|\vec{\alpha},\vec{\theta},m_t) \longrightarrow h_i(\vec{\alpha},\vec{\theta},m_t)$
- Observed : n_i



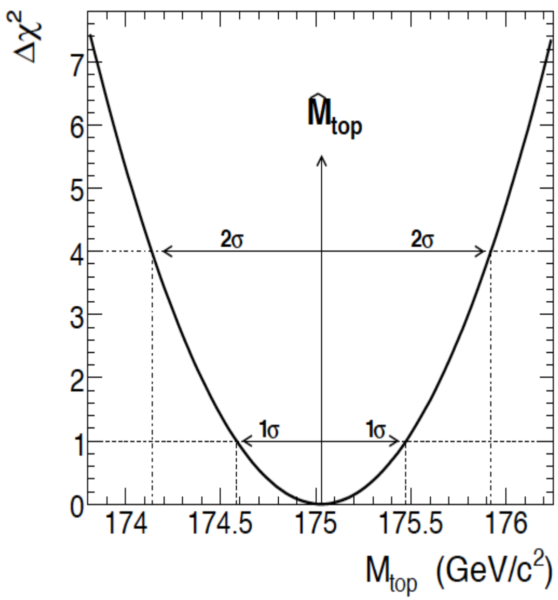
$$\chi^2 = \sum_{i=1}^{n} \left(\frac{n_i - h_i(\vec{\alpha},\vec{\theta},m_t)}{\sqrt{h_i(\vec{\alpha},\vec{\theta},m_t)}} \right)^2$$

↓

minimize $\chi^2(m_t)$

↓

$$\hat{m}_{t,LS}$$



Maximum Likelihood method

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

Calculate for each event a likelihood $g(x'|\vec{\alpha},\vec{\theta},m_t) \longrightarrow \mathcal{L}(m_t|x',\vec{\alpha},\vec{\theta})$

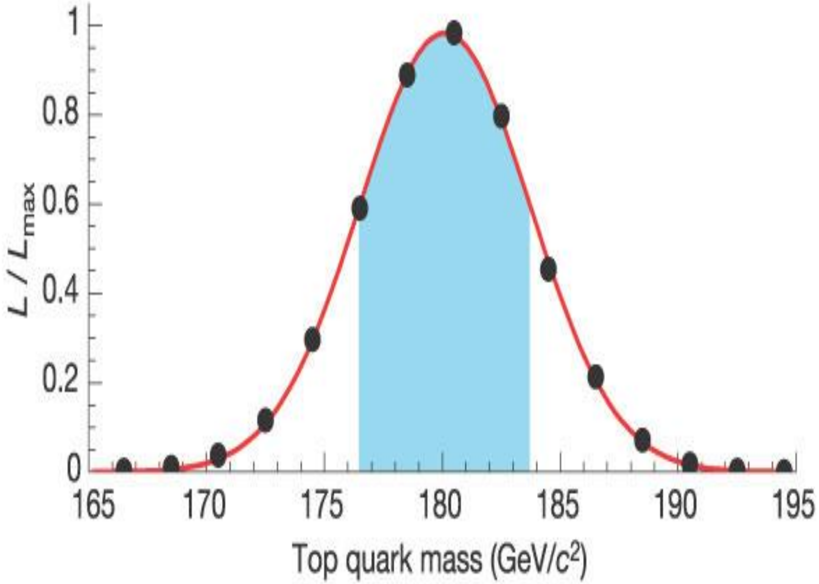
Put them together $\mathcal{L}(m_t|\vec{x},\vec{\alpha},\vec{\theta}) = \prod_{i=1}^{N} \mathcal{L}(m_t|x'_i,\vec{\alpha},\vec{\theta})$

Take the maximum of the likelihood
or from the log-likelihood:

$$ln\mathcal{L}(m_t|\vec{x},\vec{\alpha},\vec{\theta}) = \sum_{i=1}^{N} ln\mathcal{L}(m_t|x'_i,\vec{\alpha},\vec{\theta})$$

$$\longrightarrow \left(\frac{\partial ln\mathcal{L}(m_t|\vec{x},\vec{\alpha},\vec{\theta})}{\partial m_t}\right)_{m_t=\hat{m}_{t,ML}} = 0$$

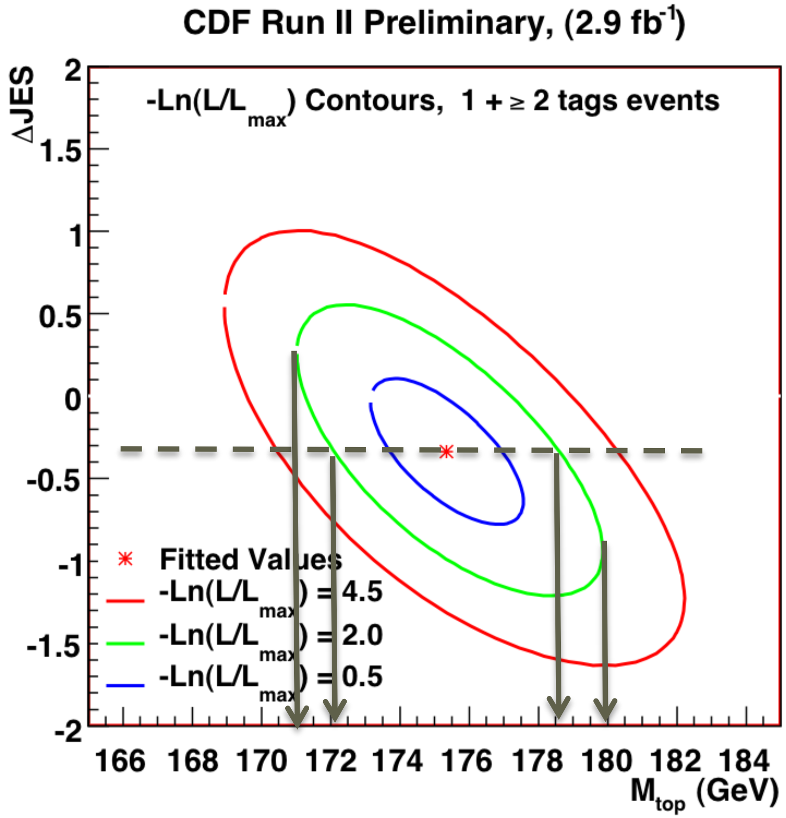
$$\longrightarrow \quad \hat{m}_{t,ML} = \ldots$$



Systematic uncertainties

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta})dx$$

There are uncertainties in the physics description f_X as well as in the experimental modelling R. Be clever, for the dominant systematic uncertainties try to fit the nuisance parameter α together with the physics parameter θ of interest.

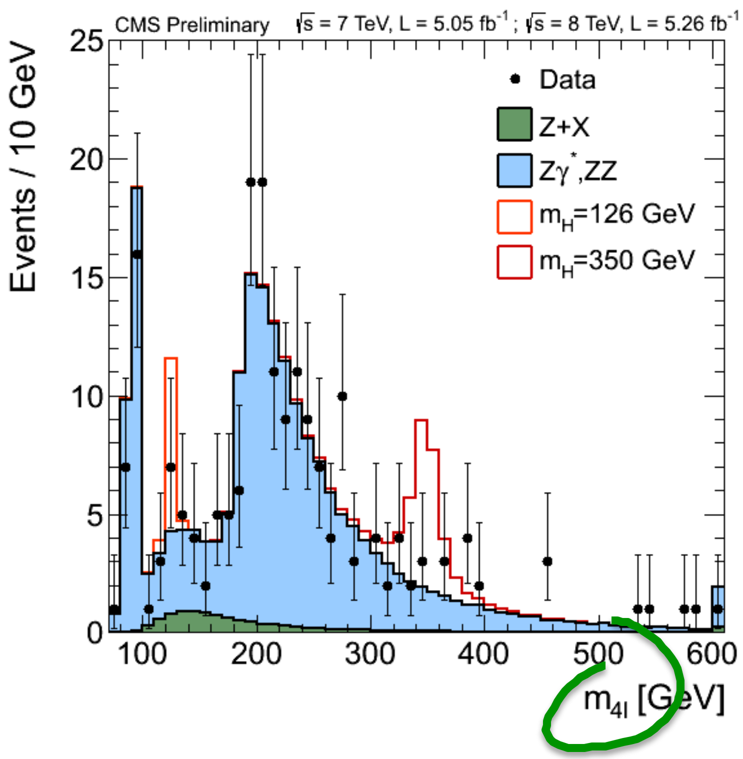


This results in a reduced total uncertainty on the physics parameter of interest.

Example: top mass measurement and the jet energy scale

Confronting data with theory

Is my fancy model of new physics predicting a new particle present in the collision data or not ?



Monte Carlo simulation — *theory* — *experiment*

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

$$g(m'_{4l}|\vec{\alpha},\vec{\theta},m_H) = \int R(m_{4l},m'_{4l}|\vec{\alpha}) f_X(x|\vec{\theta},m_H) dm_{4l}$$

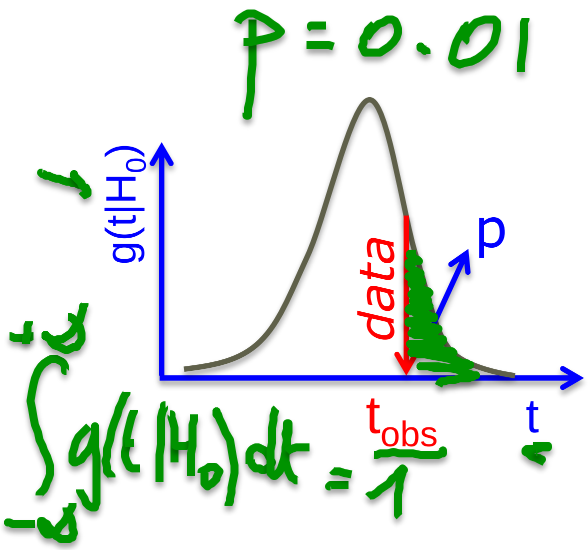
For each value of m_H we need to compare the model with data

$$g(m'_{4l}|\vec{\alpha},\vec{\theta},m_H) \longleftrightarrow \mathrm{data} = \{\mathrm{m}'_{4l,i}\}$$

p-values

Quantify the level of agreement between the model (=hypothesis H_0) and the data without an explicit alternative hypothesis.

Collect the data $\{x_i\}$ into one test statistics T (another stochastic variable), which has a distribution $g(t|H_0)$. If T is defined such that large values of t reflect a worse agreement with data, then the p-value is defined as: $p = \int_{t_{obs}}^{\infty} g(t|H_0)\ dt$

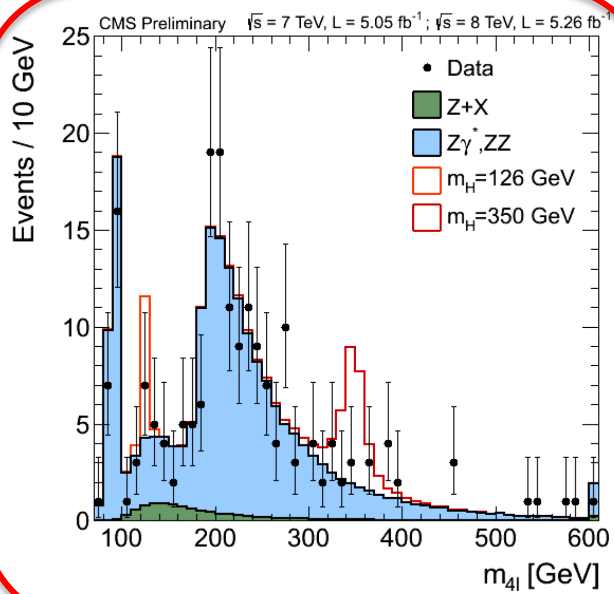


The p-value takes values between 0 and 1, and reflects the probability that a new experiment (in the same conditions) would results in a worse agreement with the hypothesis H_0, given the hypothesis H_0 is correct.

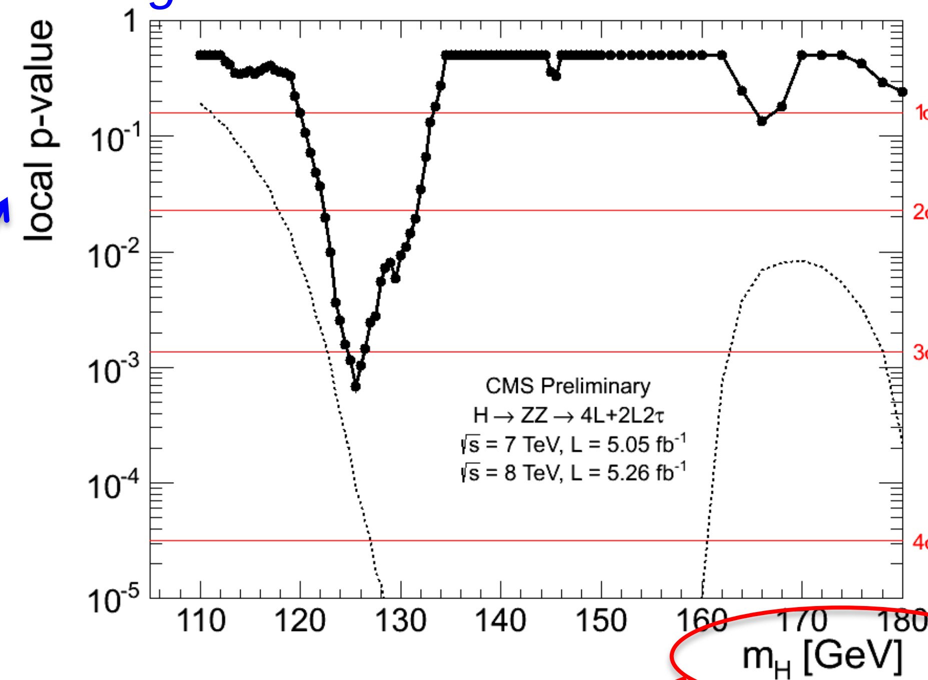
Eg.: $t = \chi^2 = \sum_{i=1}^{n_{bins}} \left(\frac{n_i - h_i(m_H, \vec{\alpha}, \vec{\theta})}{\sqrt{h_i(m_H, \vec{\alpha}, \vec{\theta})}} \right)^2$

p-values (eg. H→4l)

$$g(m'_{4l}|\vec{\alpha},\vec{\theta},m_H) \longleftrightarrow \mathrm{data} = \{\mathrm{m}'_{4\mathrm{l},\mathrm{i}}\}$$



goodness-of-fit of model to data



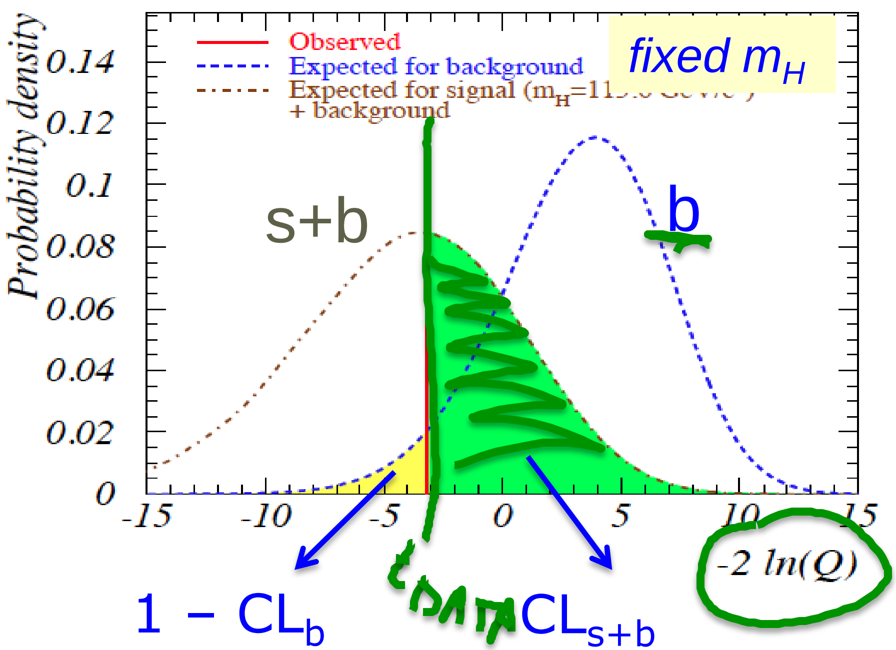
value for test statistics t

fixes the model

The CL_s method

We have the freedom to choose a test statistics to calculate the p-values from the data $\vec{n}$, an example is based on the likelihood ratio

$$Q(\vec{n}) = \frac{\mathcal{L}(\vec{n}|s+b)}{\mathcal{L}(\vec{n}|b)} \longrightarrow t = -2lnQ(\vec{n})$$



distributions from MC simulation

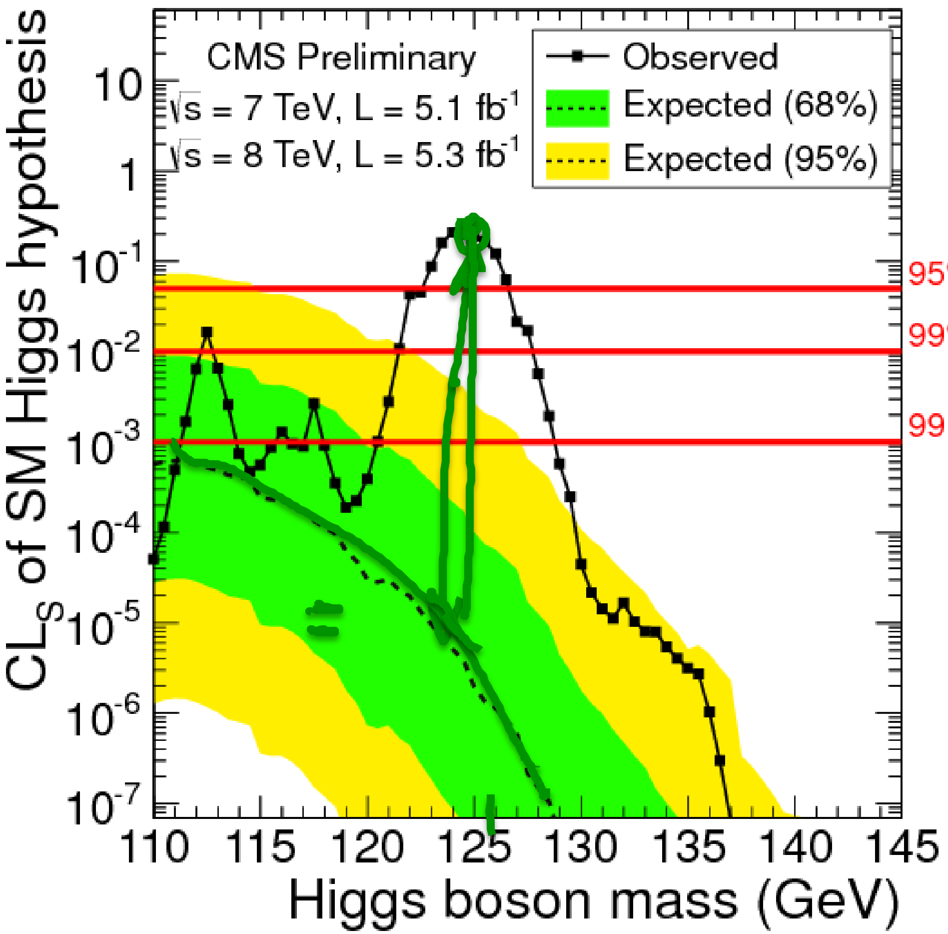
The distribution of t can be obtained using "b-only" simulation or using "s+b" simulation.

This is the optimal test statistics if one uses the value of CL_S to characterize the signal confidence.

$$CL_s = \frac{CL_{s+b}}{1-CL_b}$$

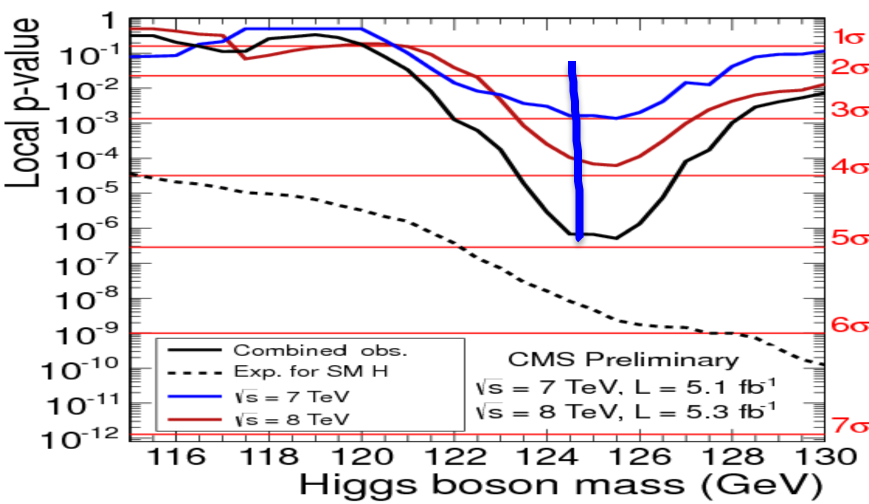
For each m_H a value of CL_s using the data $\vec{n}$ is obtained.

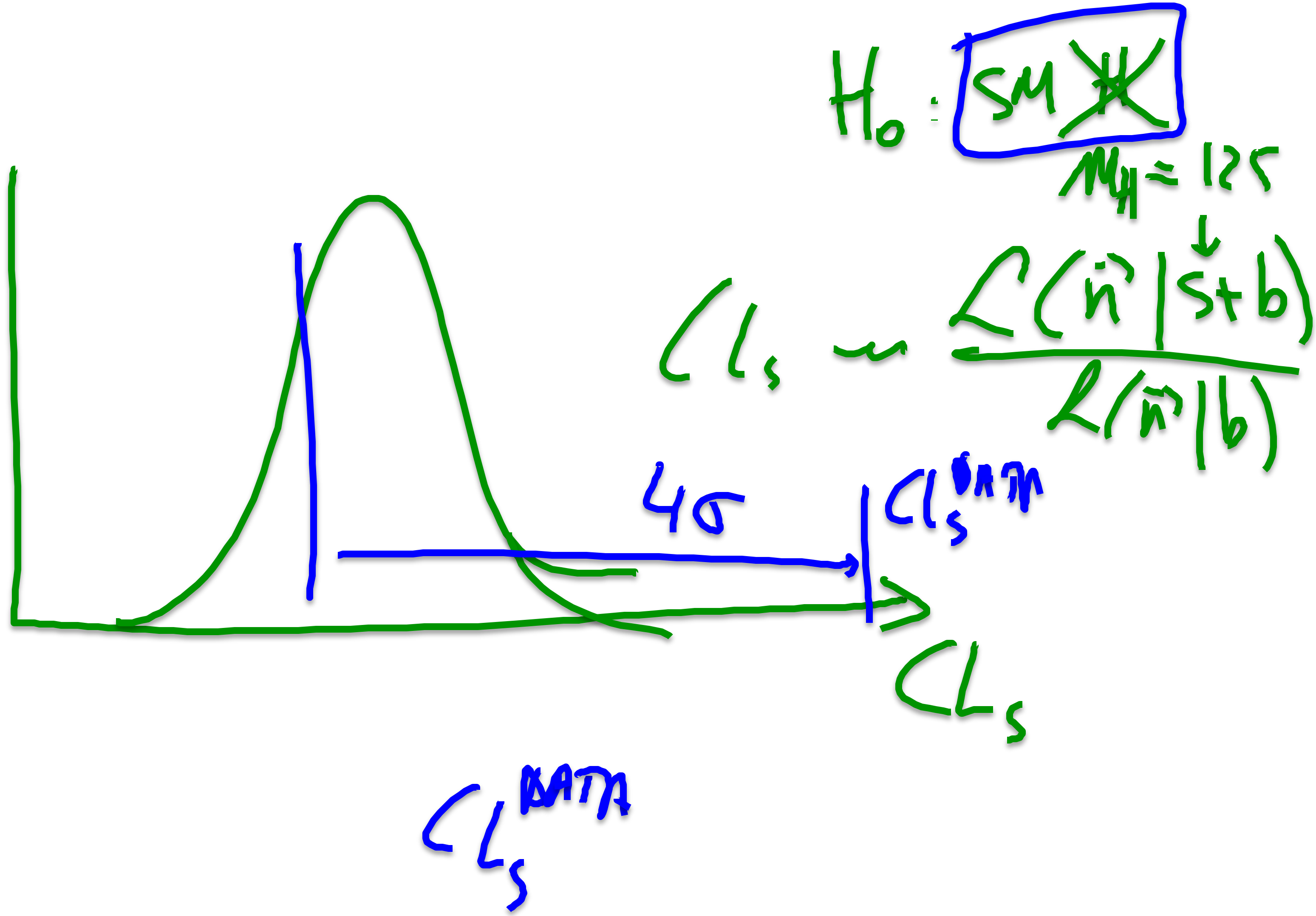
Setting limits (Higgs search)



$$CL_s = \frac{CL_{s+b}}{1-CL_b} = \frac{\text{"}p-value\ \ s+b\text{"}}{1-\text{"}p-value\ \ b\text{"}}$$

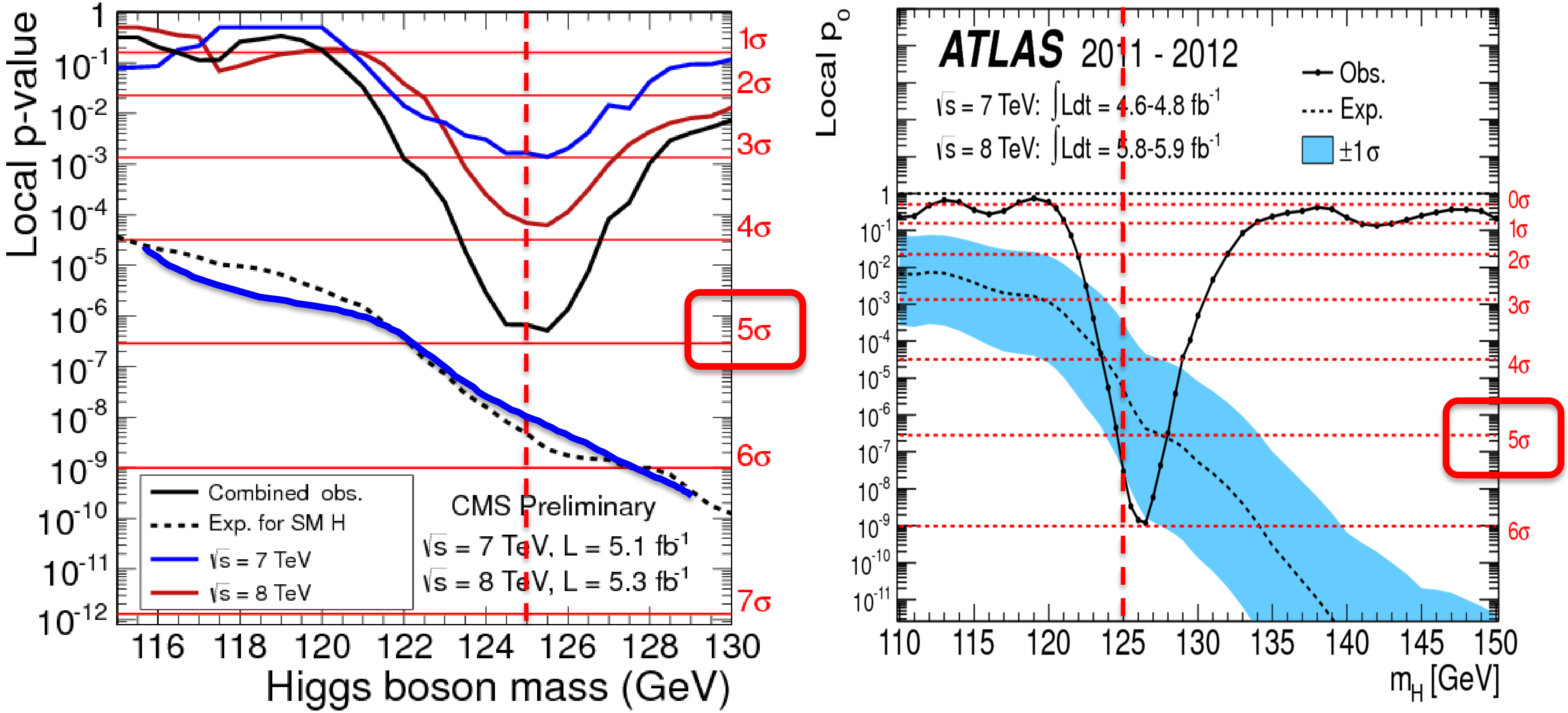
The distribution of CL_s is obtained from simulation, and the green (yellow) band reflects for each value of m_H the expected 1σ (2σ) interval if the data would be background only.




H0 : SM
MH = 125
CLs ~ L(n|s+b) / L(n|b)
4σ
CLs DATA
CLs
CLs DATA

Claiming discovery (Higgs boson)

5σ = 1 in 3,5 million



… for most other analyses however the result is a limit …

Summary of part 2

Summary of part 2

Experiment meets theory!

① Detectors make our life complicated… deal with it!

② The event selection (incl trigger) is essential in physics analyses

③ Different methods to estimate parameters… learn the differences!

$$g(x' | \vec{\alpha}, \vec{\theta}, m_H) \longleftrightarrow \mathrm{data} = \{\mathrm{x}'_\mathrm{i}\}$$

④ When confronting data with theory limits are set or discoveries claimed

⑤ Experimentalists have a hard (but fruitful) life…

Big summary…

How to test your theory?

$$g(x'|\vec{\alpha},\vec{\theta}) = \int R(x,x'|\vec{\alpha}) f_X(x|\vec{\theta}) dx$$

① Test it yourself on existing results

(*part 1 of these lectures*)

② Make it available to experimentalists to be included in their analyses

(*part 2 of these lectures*)