

---

---

# Introduction to statistics

Tomas Dado (Dortmund)

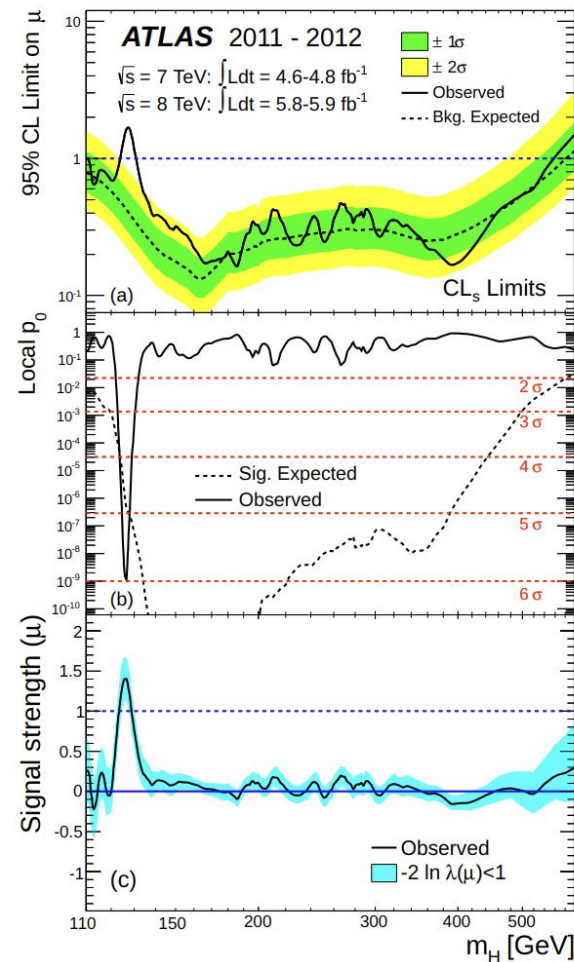
---

---

HASCO, Goettingen, 2023

# Introduction

- Focusing on HEP statistics approaches
- **Quantum mechanics/field theory = statistical theory**
  - Needed for every interpretation
- Here we will go through
  - Basics of statistics
  - Hypotheses testing
  - Discovery and limit setting
  - Parameter estimation
  - Unfolding
- Should be able to understand these plots at the end of this presentation



<https://arxiv.org/abs/1207.7214>

# Useful references

- G. Cowan, *Statistical Data Analysis*, Oxford University Press, 1998
  - Related: Cowan's Academic lectures: [indico link](#)
- F. James, *Statistical methods in experimental physics*, 2nd ed., World Scientific, 2006
- K. Cranmer, *Practical Statistics for the LHC*, <https://arxiv.org/abs/1503.07622>
- Cowan et al, *Asymptotic formulae for likelihood-based tests of new physics*, <https://arxiv.org/abs/1007.1727>
  
- Commonly used model for the binned likelihood fit in HEP: *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, <https://cds.cern.ch/record/1456844>

# Basics

# Frequentist statistics

- Probability = **outcomes of repeatable observations**

$$P(x) = \lim_{n \rightarrow \infty} \frac{\text{number of outcomes of } x}{n}$$

- I.e. we need **repeatable events**
- Does Higgs boson exist? Is the mass of the top quark between 172 and 173 GeV? ...?
  - It is **either true or false** but we do not know which
  - The frequentists tools tell us about outcomes of (hypothetical) **repeated experiments**
- The **preferred theories** (models, hypotheses, ...) are those for which our observations would be considered "usual"

# Bayesian statistics

- Interpretation of probability extended to a degree of belief
  - The degree of belief is updated based on the observations
- Bayes' formula

**Probability  
observing data  $\vec{x}$ ,  
assuming the  
hypothesis  $H$**

**Prior probability  
for hypothesis  $H$**

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H) dH}$$

**Normalisation, i.e. sum of all possible outcomes**

# Bayesian statistics example

- Assume 2% of the population have COVID19 in a given time
- The tests for COVID19 detect the virus in 90% of the cases and give false-positive (show positive result even when there is no COVID19 virus) in 5% of the cases
- The test result is positive, what is the probability that the person has the COVID19 virus?

We can use the Bayes' formula for this

- $P(H) = 0.02$  - this is the prior probability, i.e. before we do the test
- $P(x,H) = 0.9$  - i.e. if the person is positive, what is the likelihood of getting a positive result
- Normalisation =  $0.9 \times 0.02 + 0.05 \times 0.98$  - i.e. has the virus and positive test + does not have virus and has a positive test
- Using the Bayes' formula:

$$\frac{0.9 \times 0.02}{0.9 \times 0.02 + 0.05 \times 0.98} \approx 24\%$$

- How would the probability change if the person would do another test and it came back positive?

# Frequentist vs Bayesian

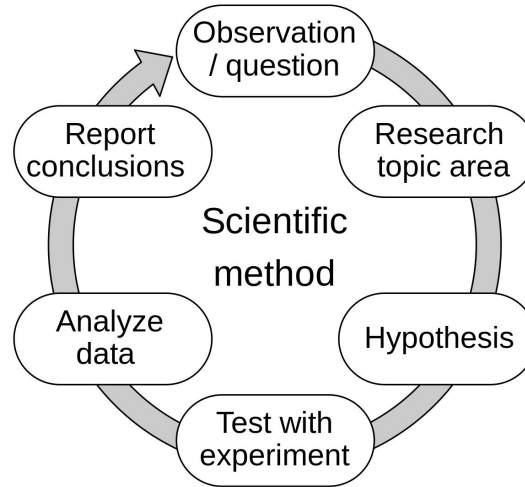
- Frequentist
  - **Limit of a long term frequency**
  - Do not need an infinite sample for the definition to be useful
  - **Sometimes no ensemble exists**
- Bayesian
  - **Probability is a degree of belief**
  - Intrinsically **subjective** (choice of the prior)
    - No golden rule for the **choice of priors**
- “Bayesians address the question everyone is interested in, by using assumptions no-one believes. Frequentists use impeccable logic to deal with an issue of no interest to anyone” - L. Lyons



# Hypothesis testing

# Definitions

- Hypothesis testing is a core of the scientific method

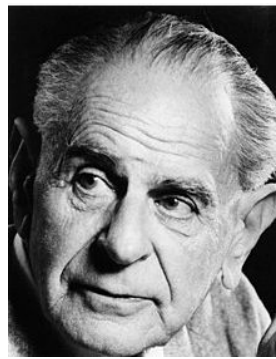


- Hypothesis  $H$**  specifies the **probability for the data**, i.e., the outcome of the observation,  $x$
- Possible values of data ( $x$ ) form the sample space ("data space")
- The **probability for  $x$  given  $H$**  is also called the **likelihood of the hypothesis, written  $L(x|H)$** .
  - E.g. The probability to observe  $N$  number of events with a given selection assuming the validity of the Standard Model

# Hypothesis testing

How to confirm a hypothesis?

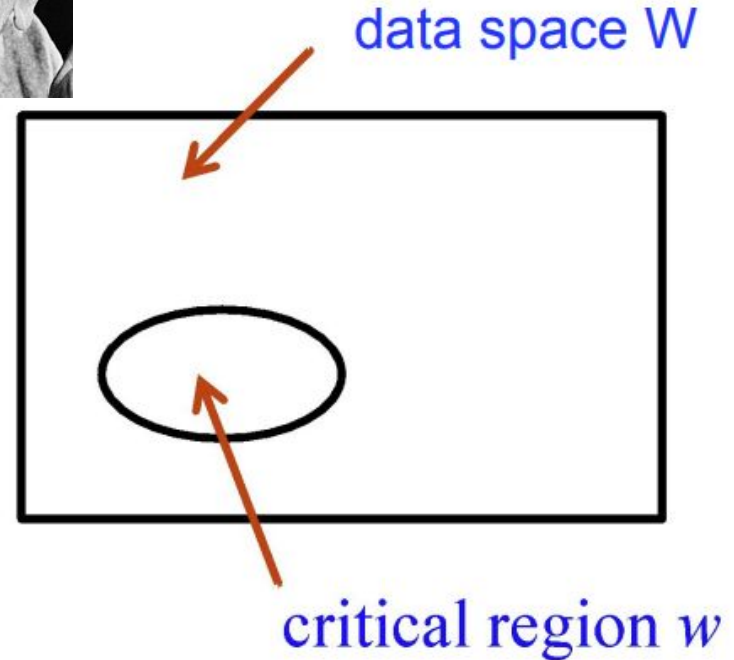
- Karl Popper: **You cannot!**
- But you **can reject a hypothesis!**



- Find a region,  $W$ , of the data space where the is only **small probability  $\alpha$  to observe data  $x$  provided  $H_0$  is true** - this is the “critical region”

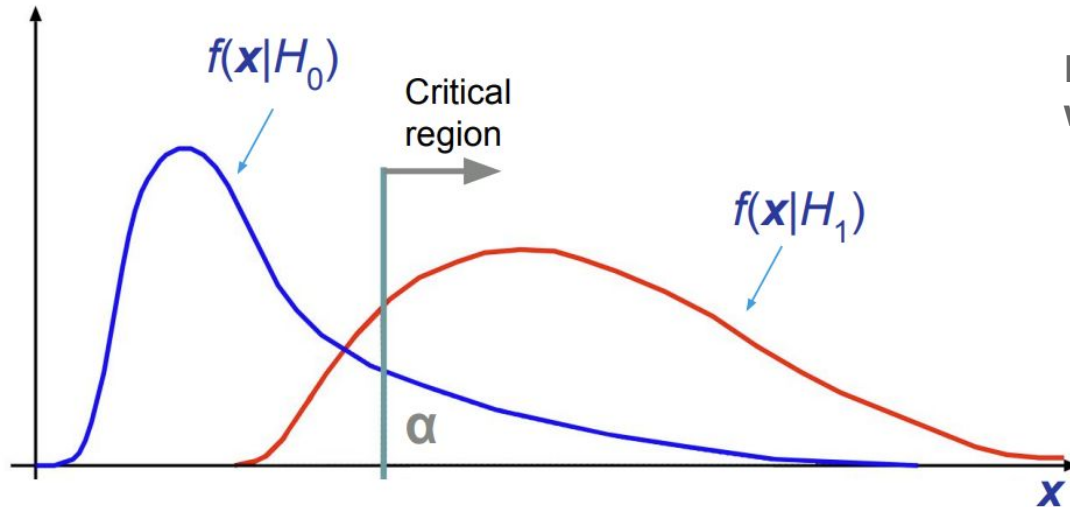
$$P(x \in W | H_0) \leq \alpha$$

- **Reject hypothesis if data is observed in  $W$**
- $\alpha$  is called “size” or “**significance level**” of the test



# How to select the critical region?

- Infinitely many critical regions for a given hypothesis
- No **unique way to select it**
- Can define an alternative hypothesis  $H_1$
- Roughly speaking:
  - Choose the critical region so that the **probability of observing data under  $H_0$  is low** and **probability of observing data under  $H_1$  is high**



Rejecting  $H_0$  does not mean “ $H_0$  is wrong and  $H_1$  is right”

- Frequentist - only outcome of repeated experiments
- Bayesian - depends on the priors

# Type-I and type-II errors

- Type-I error (false negative)
  - **Reject hypothesis  $H_0$  if it is true**
  - Maximum probability for this is  $\alpha$

$$P(x \in W \mid H_0) \leq \alpha$$

- Type-II error (false positive)
  - **Accept hypothesis  $H_0$  if it is false and  $H_1$  is true**
  - Occurs with probability  $\beta$

$$P(x \in S - W \mid H_1) = \beta$$

- $1 - \beta$  is called the “power” of the test

True negative



False positive



False negative



True positive



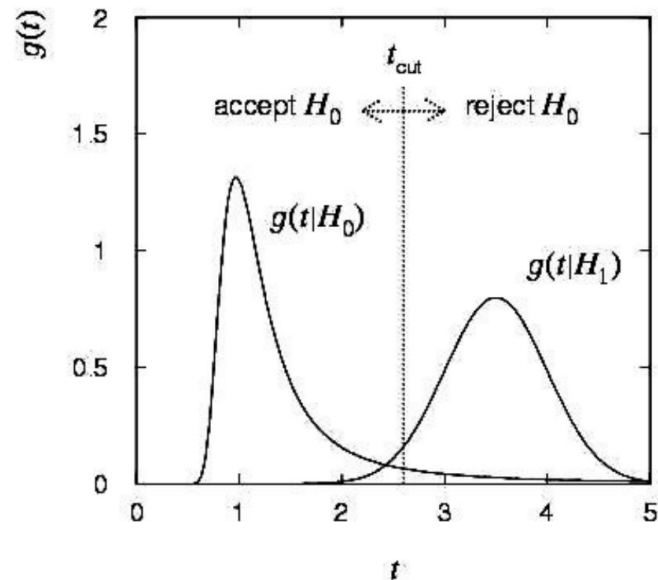
# Test statistics

- Assume that for **each event we have a collection of numbers**
  - Number of jets, leptons, MET value, ..., have multiple bins, ...
  - Data ( $x$ ) will follow some joint PDF for the different observables
  - The critical region is **multidimensional** - cumbersome to work with
- Can define the **boundary** of the critical region using an equation of form

$$t(x_1, \dots, x_n) = t_{\text{cut}}$$

- Where  $t(x_1, \dots, x_n)$  is the **scalar** test statistics

**We have turned an N-dimensional problem to a 1-dimensional one!**



# Optimal choice for the test statistics

- How to choose the test statistics?
- **Neyman-Pearson lemma**: For a test of size  $\alpha$  of the simple hypothesis  $H_0$ , to obtain the highest power with respect to the simple alternative  $H_1$ , choose the critical region  $W$  such that the likelihood ratio satisfies

$$\frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)} \geq k$$

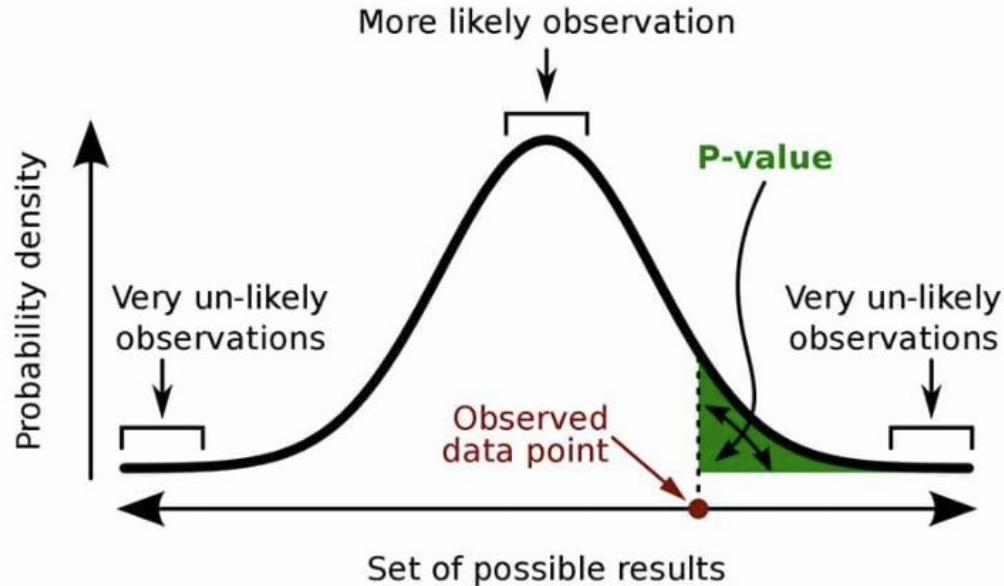
everywhere in  $W$  and is less than  $k$  else -  $k$  is a constant chosen such that the test has size  $\alpha$

- The **optimal scalar test statistics is then**

$$t(\mathbf{x}) = \frac{P(\mathbf{x}|H_1)}{P(\mathbf{x}|H_0)}$$

# p-value

- Level of agreement (compatibility) of data and a given hypothesis (model)  $H$
- p-value  $\rightarrow$  probability, under assumption of  $H$ , to observe data with **equal or lesser compatibility** with  $H$  relative to the data we got
  - **This is NOT a probability that  $H$  is true!**



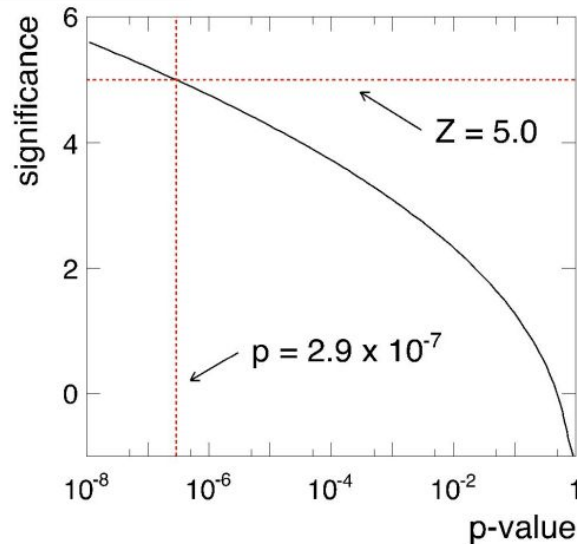
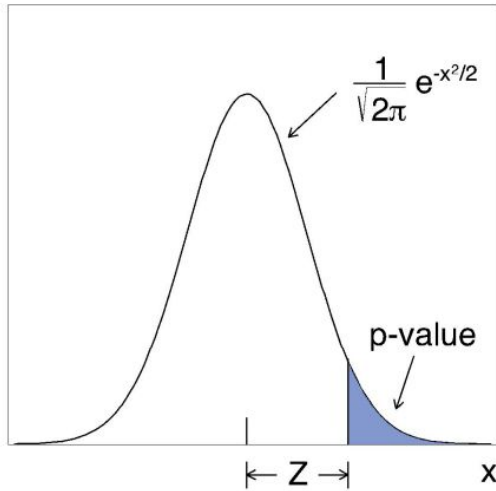


# p-value and significance

- We can define the significance  $Z$  as the **number of standard deviations** ("sigmas") that a Gaussian variable would fluctuate in one direction to give the same p-value

$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z) \longrightarrow Z = \Phi^{-1}(1 - p)$$

Gaussian cumulative function

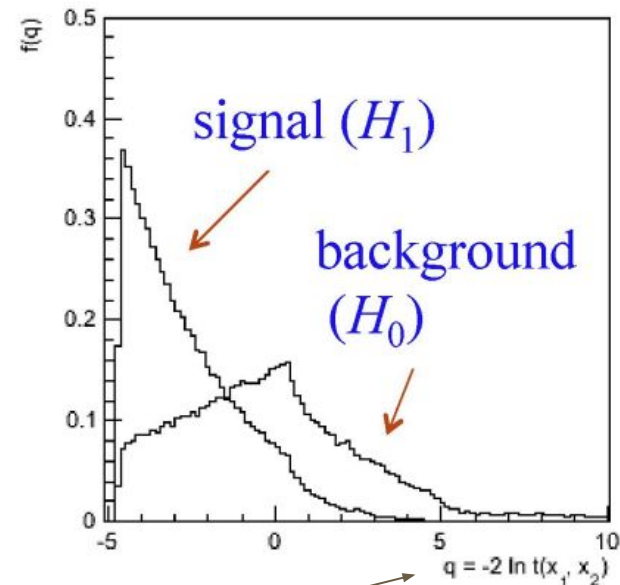


z (one tail)	p-value
1.00	0.16
2.00	0.023
3.00	0.0013
4.00	3.2e-05
5.00	2.9e-07
6.00	9.9e-10

# Discovery and limits

# Discovery in HEP

- We want to discover new physics (BSM)
- Typically
  - Hypothesis  $H_0$ , i.e. the “null hypothesis” is the SM prediction
    - **“Background-only” hypothesis**
  - Alternative hypothesis  $H_1$  is your favourite model
- We know what to do
  - Find the  $P(x, H_0)$  and  $P(x, H_1)$ , i.e. the likelihood
  - **Build the test statistics** using the ratios
  - **Calculate the p-value**
    - **Reject/accept**
- How to get the PDF?
  - Use **MC simulation**
  - Need to get a distribution of the values
    - **Pseudo-experiments/toys!**



We usually use (-2 times) logarithm of the ratio

# Simple example

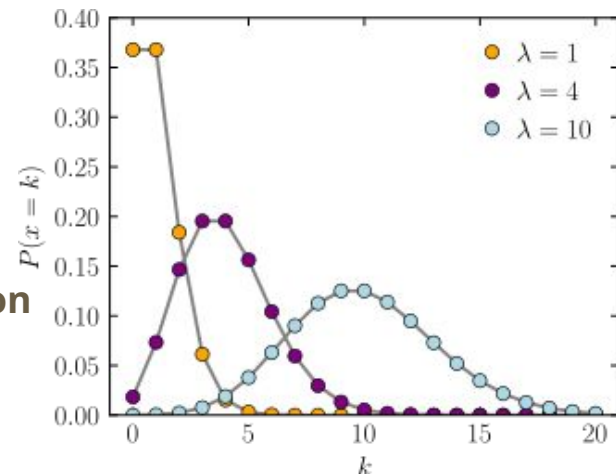
- Suppose we are doing a **counting experiment**
  - Predicted number of background events is ***b***
  - Predicted number of signal events is ***s***
  - Observed number of events will follow **Poisson distribution**

$$P(n|b) = \frac{b^n}{n!} e^{-b}$$

Background only

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

Signal + bkg



- We observe ***n*** instances of ***x***
- Likelihoods for the hypotheses**

- Background only

- Signal + bkg

$$L_b = \frac{b^n}{n!} e^{-b} \prod_{i=1}^n f(\mathbf{x}_i|b)$$

$$L_{s+b} = \frac{(s+b)^n}{n!} e^{-(s+b)} \prod_{i=1}^n (\pi_s f(\mathbf{x}_i|s) + \pi_b f(\mathbf{x}_i|b))$$

(Prior) probabilities for an event to be signal or bkg

# Simple example continued

- Define test statistics (**-2 logarithm of the likelihood ratio**)

Is constant, can be ignored

$$Q = -2 \ln \frac{L_{s+b}}{L_b} = -s + \sum_{i=1}^n \ln \left( 1 + \frac{s}{b} \frac{f(\mathbf{x}_i|s)}{f(\mathbf{x}_i|b)} \right)$$

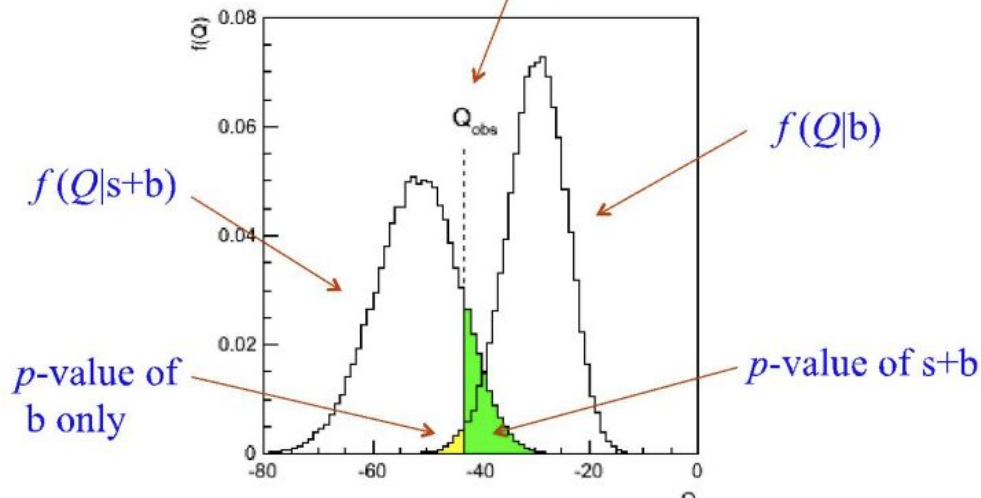
- Let us assume we observe  $Q = Q_{\text{obs}}$

e.g.  $b = 100$ ,  $s = 20$ .

Suppose in real experiment  
 $Q$  is observed here.

- HEP standard**

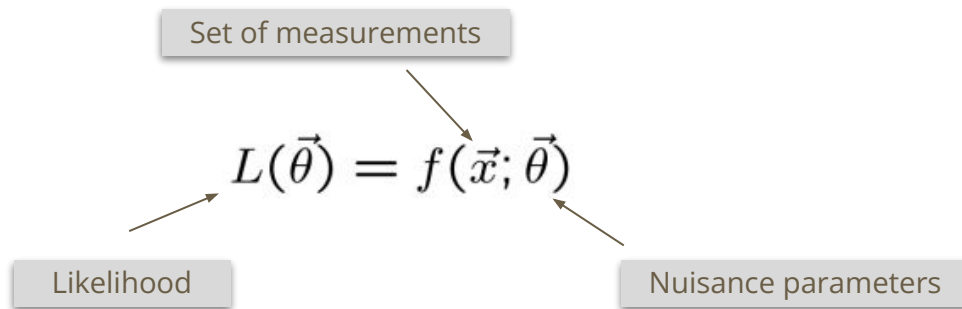
- Claim **discovery at 5 sigma**
- Reject B-only hypothesis  
when p-value is  $< 2.9 \times 10^{-7}$



# Let's add systematics

- So far, only considered statistical uncertainty
- In reality, many **systematic uncertainties affect the predictions**
- Can add the **systematics into the likelihood**
  - Define “*signal strength*”,  $\mu$ , as  $n = \mu \cdot s + b$ 
    - $\mu = 1$  means cross-section as predicted by the model
  - Add “*nuisance parameters*” to the likelihood
    - Parameters that impact the likelihood, but we are not interested in them, e.g. systematic uncertainties
    - Usually, “subsidiary” or “auxiliary” measurements are used to constrain NPs

$$\mu = \frac{\sigma_{obs.}}{\sigma_{pred.}}$$



# Commonly used model

- More and more common approach for including systematics in HEP statistical analysis:
  - include **systematic uncertainties as unknown parameters in the model**
  - **nuisance parameters** modifying expectations in a **parametric** way
  - **nuisance parameters constrained by subsidiary measurements**

- The binned profile-likelihood:

$$L(\vec{n} \mid \vec{\theta}, \vec{k}) = \prod_i P(n_i \mid \underbrace{S_i(\vec{\theta}, \vec{k}) + B_i(\vec{\theta}, \vec{k})}_{\text{prediction in bin } i \text{ (signal+background)}}) \times \prod_j \underbrace{G(\theta_j)}_{\text{constraint term for nuisance parameter } j}$$

data  $\vec{n}$       Poisson  $P(n_i \mid \dots)$       Gaussian (or other pdf...)  $G(\theta_j)$

**constrained parameters:**  
nuisance parameters (**NPs**)  
associated to systematic uncertainties

**unconstrained parameters:**  
parameter of interest (**POI** or “ $\mu$ ”) + unconstrained nuisance parameters (e.g. background normalization parameters)

data events in bin  $i$       prediction in bin  $i$  (signal+background)      constraint term for nuisance parameter  $j$

# Profile-likelihood significance

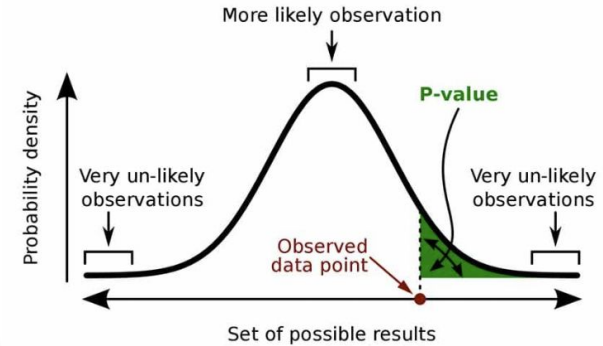
- Define test statistics

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

Maximises  $L$  for a given fixed  $\mu$

Best fit  $\mu$

Likelihood value that maximises the likelihood for all parameters



- Observing new physics  $\Leftrightarrow$  excluding background-only hypothesis  $\Leftrightarrow$  excluding  $\mu = 0$**
- Only consider upward fluctuations

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$


$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0) dq_0$$



# Wald's approximation

- Running the **fit can take a long time**
- We need a PDF for the test statistics  $\Leftrightarrow$  many fits to toy data
  - For 5 sigma discovery we need  $\sim 10^7$  toys!
- Luckily, there is a **powerful approximation** - Wald's approximation
- **For large  $n$ , the likelihood ratio is approximately chi-square distributed!**
  - Does not require the likelihood to be chi-square or gaussian distributed!

$$-2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})$$

$\hat{\mu} \sim \text{Gaussian}(\mu', \sigma)$   sample size

- Under this assumption, the significance is simply

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}$$

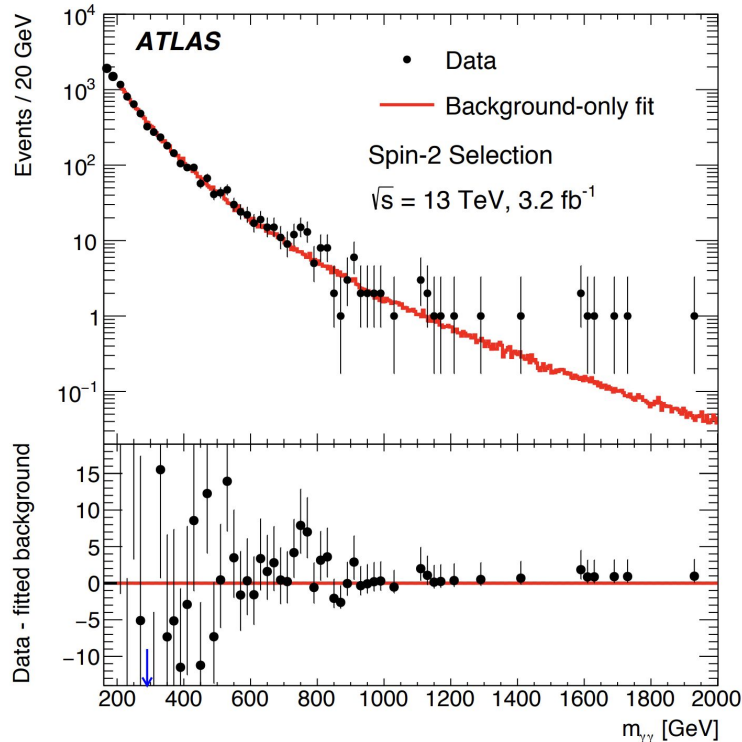
Usually a good approximation  
as long as number events in  
each bin is greater than  $\sim 10$

- I.e. need to **run the fit only twice** - **unconditional** and **with  $\mu$  fixed to 0**
  - Get the  $-2 \ln L$  values for the fits and take the square root of the difference

# Look-elsewhere effect

- What if we are looking for a resonance with an unknown mass and see an excess in some mass?
  - Should we just quote the **significance for that mass point**?

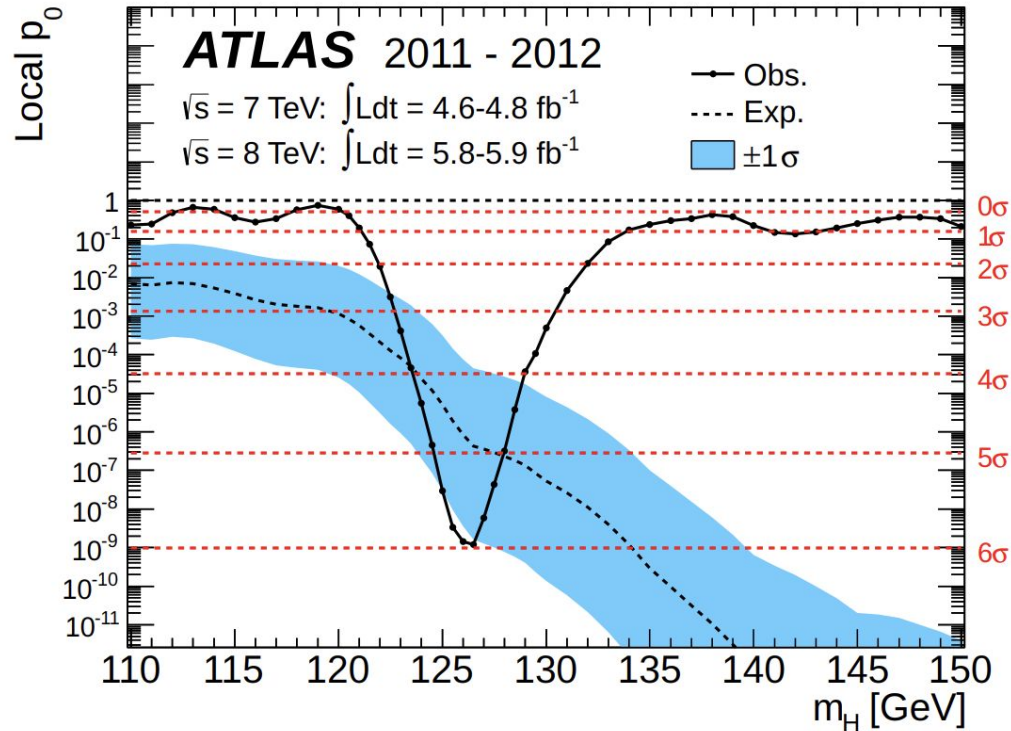
<https://arxiv.org/abs/1606.03833>



- Need to take into account the **"trials"**
  - We are **"testing" multiple bins**
  - We have more options to find an excess
  - **Need to correct for this!**
- Significance for a **fixed mass point**  $\Leftrightarrow$  local significance
- Significance for the **floating mass**  $\Leftrightarrow$  global significance
  - **Global significance  $\leq$  local significance**
- How to relate local significance to the global one?
  - **No simple recipe**
  - Need to **run toys**
    - Usually only 100s, not millions

# Reading significance plots

<https://arxiv.org/abs/1207.7214>



- Dashed curve = "*Expected*" median  $p_0$ 
  - $p_0$  for each mass of the SM Higgs boson - from MC
- Blue band = 1 sigma variations of the  $p_0$  value
- Full line = "*Observed*"  $p_0$  value from real data
- $> 5$  sigma at around  $m_H = 125$  GeV

# Setting limits

- What if we do not see any significant excess?
  - We can **set limits**!
- What values of  $\mu$  can be excluded with the observed data?
  - I.e. the **implied rate for a given  $\mu$  would be very high for the observed data**
  - One-sided test - provide an “upper limit”
- Slightly modify the test statistics used for discovery
  - If  $\mu$  comes out negative (unphysical) we can compare to the closest model with  $\mu = 0$

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\hat{\theta}})} & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\hat{\theta}})}{L(0, \hat{\hat{\theta}})} & \hat{\mu} < 0. \end{cases} \quad \tilde{q}_{\mu} = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

This is the test statistics commonly used (e.g. Higgs combinations)

# Setting limits - continued

- **Settings limits** = finding the highest value of  $\mu$  that results in p-value not smaller than  $y$ 
  - $y$  is **usually chosen as 0.05, i.e. 95% confidence level (CL)**
  - *"What is the largest value of  $\mu$  that is still compatible with the data?"*

The diagram illustrates the formula for the p-value,  $p_\mu = \int_{\tilde{q}_\mu}^{\infty} f(\tilde{q}_\mu | \mu, \hat{\theta}(\mu)) d\tilde{q}_\mu$ , with several annotations in grey boxes:

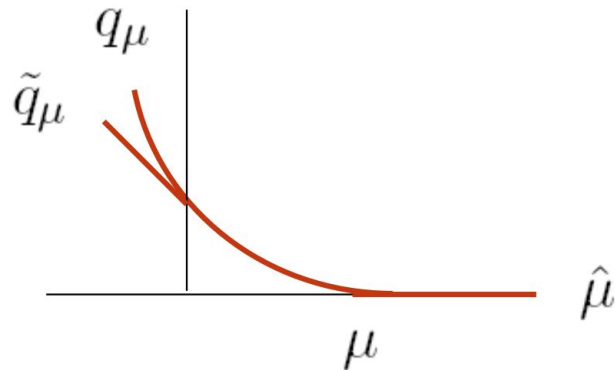
- P-value for a given  $\mu$** : An arrow points from this box to the  $p_\mu$  term in the formula.
- Test statistics**: An arrow points from this box to the  $\tilde{q}_\mu$  variable in the integrand.
- Maximises likelihood for a fixed  $\mu$** : An arrow points from this box to the  $\hat{\theta}(\mu)$  term in the integrand.
- Observed value  $q_\mu$  tilde**: An arrow points from this box to the lower limit  $\tilde{q}_\mu$  of the integral.

- **Need to solve for  $\mu$** 
  - Nasty integral equation
  - Can run **pseudo-experiments** to get the distribution of the test statistics
    - Find  $\mu$  that leads to  $p_\mu = 0.05$

# Asymptotic limit settings

- Can use the Wald's approximation
  - The test statistics approaches chi-square

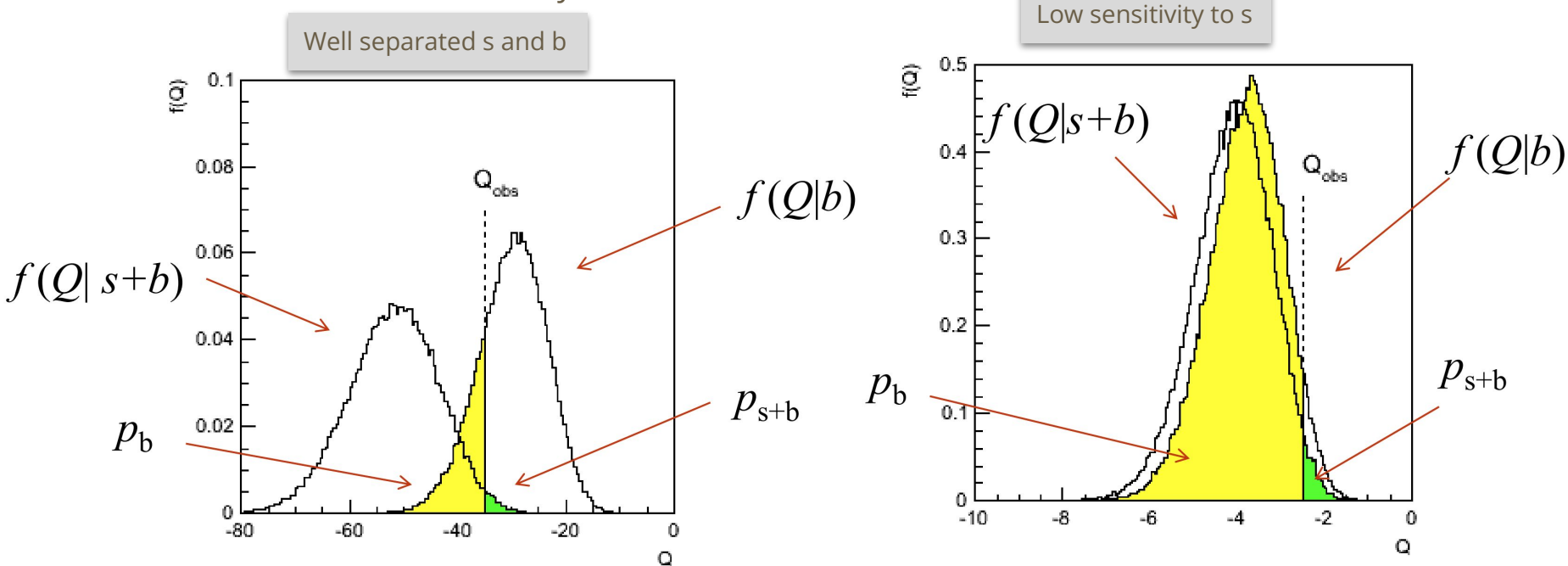
$$q_{\mu} = \begin{cases} \frac{(\mu - \hat{\mu})^2}{\sigma^2} & \hat{\mu} < \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad \tilde{q}_{\mu} = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0 \\ \frac{(\mu - \hat{\mu})^2}{\sigma^2} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$



- Limit estimation in practice (simplified)
  - Get the best fit value of  $\mu$  and its uncertainty (more on this later)
  - Set  $\mu$  to +2 sigma (approximately 95%) - this is a starting point of the iterative estimation
  - Calculate the p-value for this  $\mu$ 
    - If p-value too small, decrease  $\mu$ , if p-value too large increase  $\mu$
    - Repeat!
    - Stop when the p-value is sufficiently close to 0.05
  - Usually requires O(10) fits
- If the asymptotic approximation is not valid, have to use toy experiments

# The CLs issue

- Suppose we **have a low sensitivity to a particular signal**
  - Test statistics for **s+b** is **very similar to background-only**
  - There is non-negligible **probability to exclude s+b even when we have low sensitivity**
    - Can be caused by downward fluctuation



# The CLs procedure

[A. Read et al.](#)

- **Solution** to the issue: do not use only p-value for the s+b but divide by p-value for b-only

- **Define CLs**

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

- **Reject s+b hypothesis if CLs <  $\alpha$**

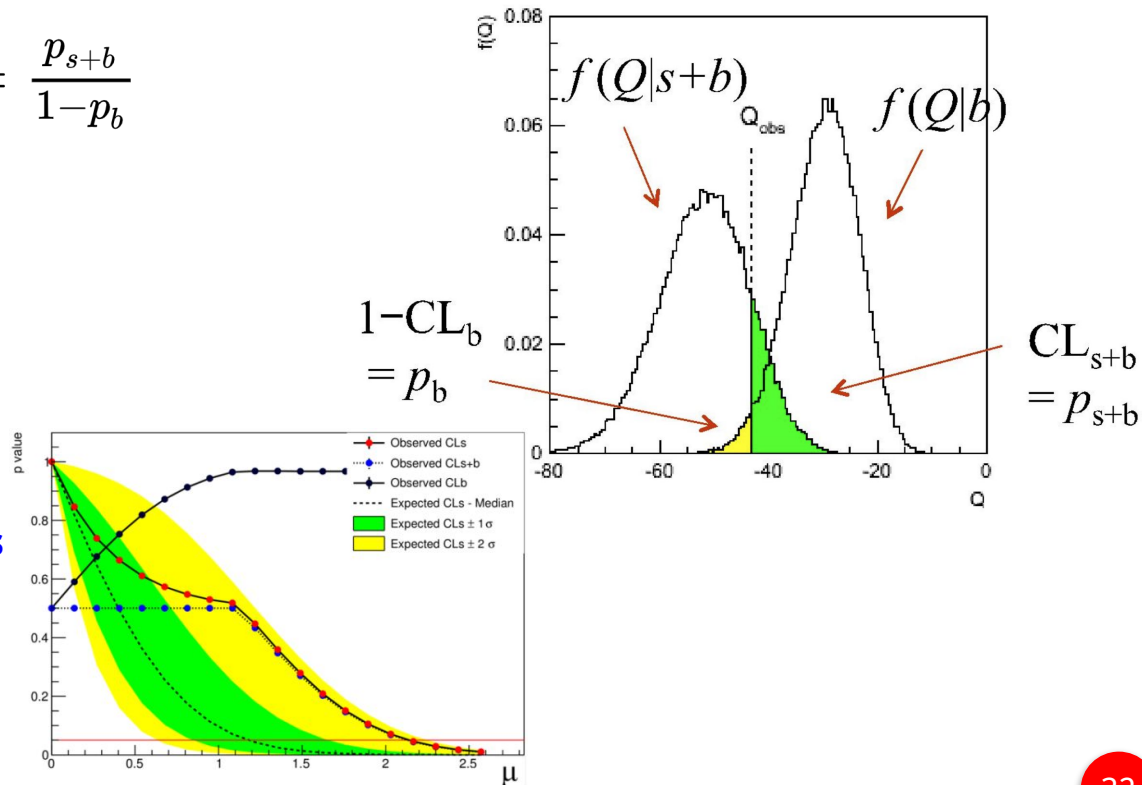
- Reduces “effective p-value”

- If low sensitivity

- Ratio of p-values

- Not liked by statisticians

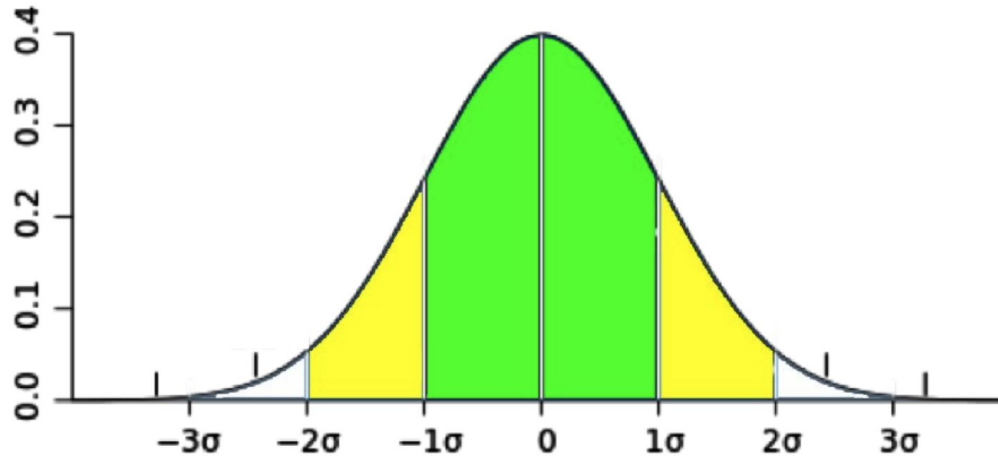
- **Used in almost all HEP searches**

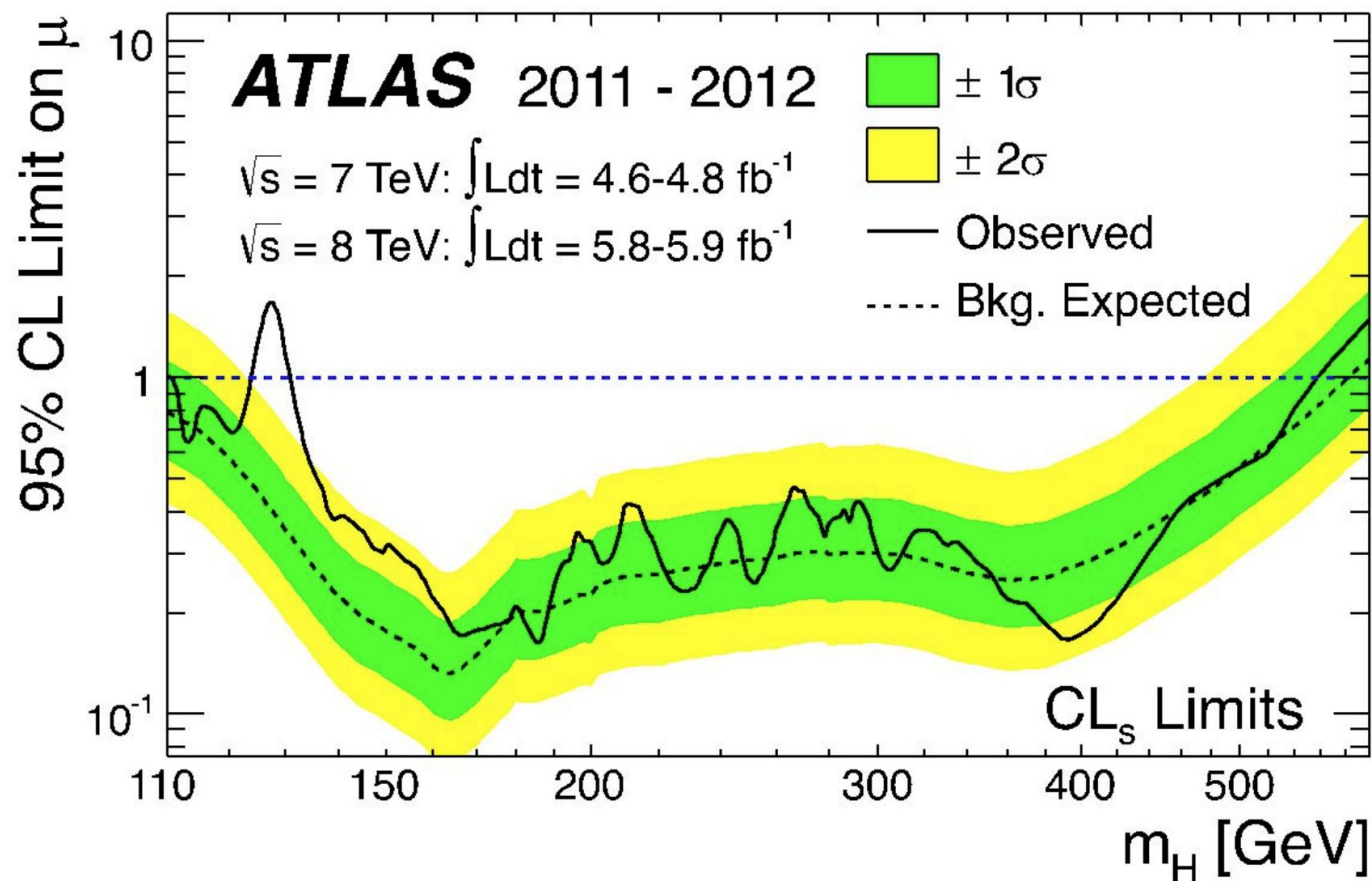




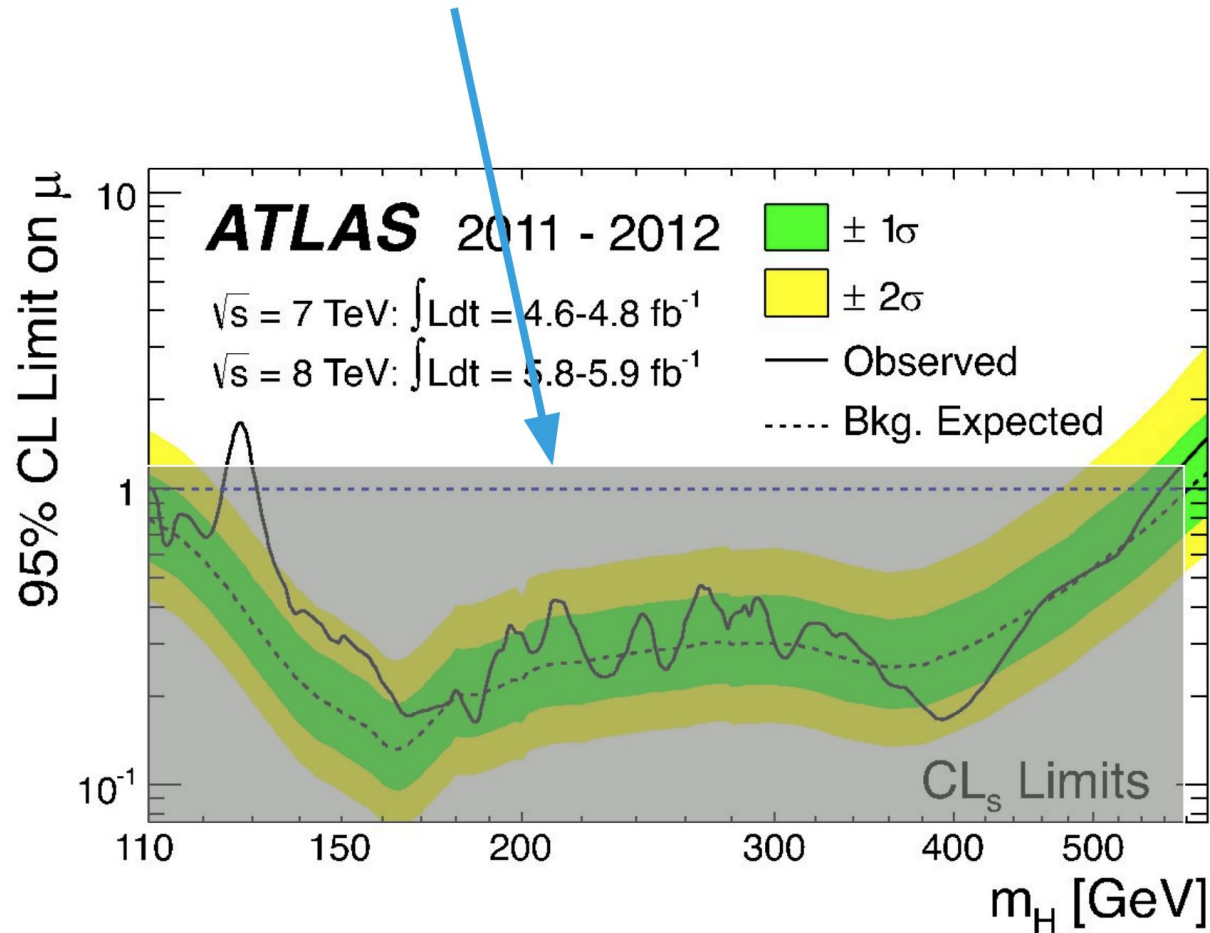
# Expected limits

- **Expected limits** can be calculated using the MC prediction
  - Assume background only, what would be the limit on  $\mu$  in case data = MC?
  - Can do it for several models, e.g. different masses of the Higgs boson
- Frequentist approach
  - Distribution of the p-value  $\Leftrightarrow$  distribution of the 95% CL limits
  - Can quote **median expected limit** and  **$\pm 1(2)$  sigma variations**

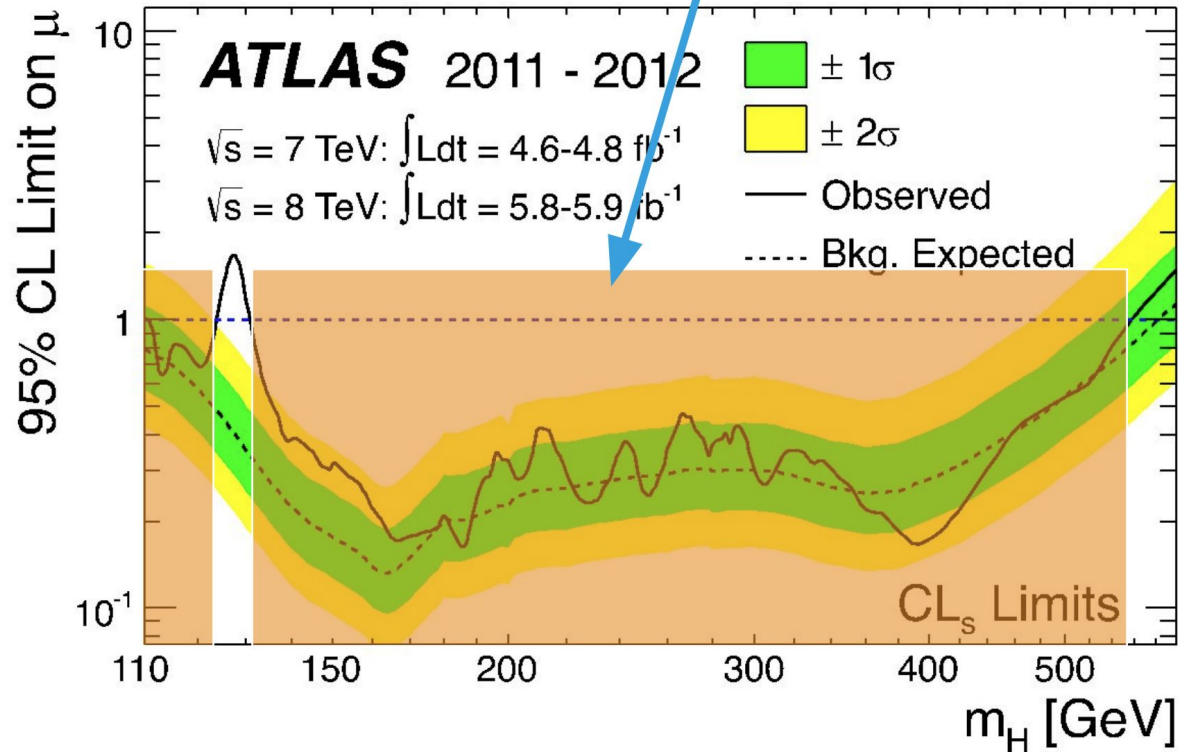




## Expected excluded mass range



## Observed excluded mass range



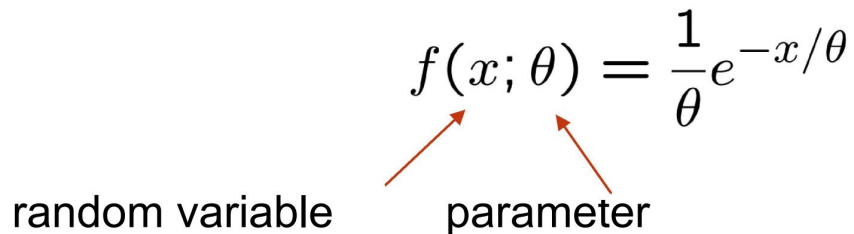
# Parameter estimation

# Estimators

- Often *not searching for a new process*
  - E.g. Measuring top-quark mass, CKM matrix elements, ...
- How to get the parameters from the model with their uncertainties?
- We need the PDF of the estimation
- **Parameters** are **constants of the estimator** that characterise the shape

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}$$

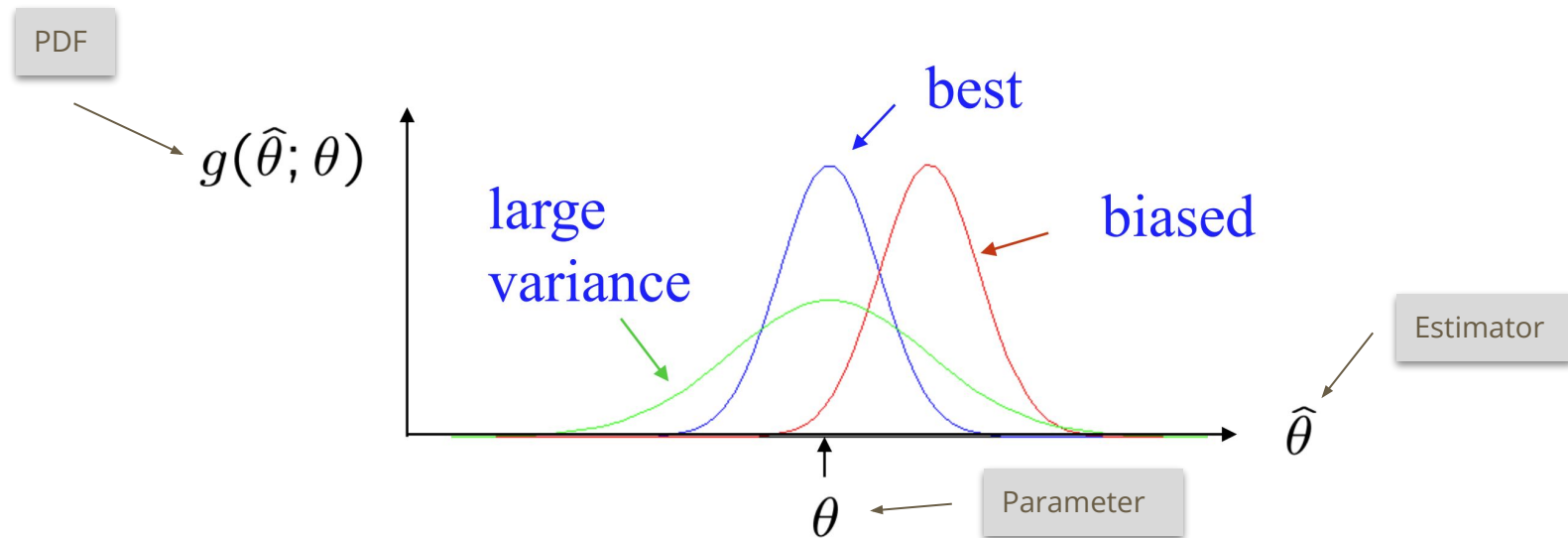
random variable      parameter



- We want to **find some function of data to estimate the parameter(s)**:  $\hat{\theta}(\vec{x})$ 
  - Estimator written with a hat

# Estimators continued

- Repeating the measurement -> get PDF



- We want unbiased estimator (bias = 0) with small variance (small statistical uncertainty)
  - **Generally: conflicting requirements**

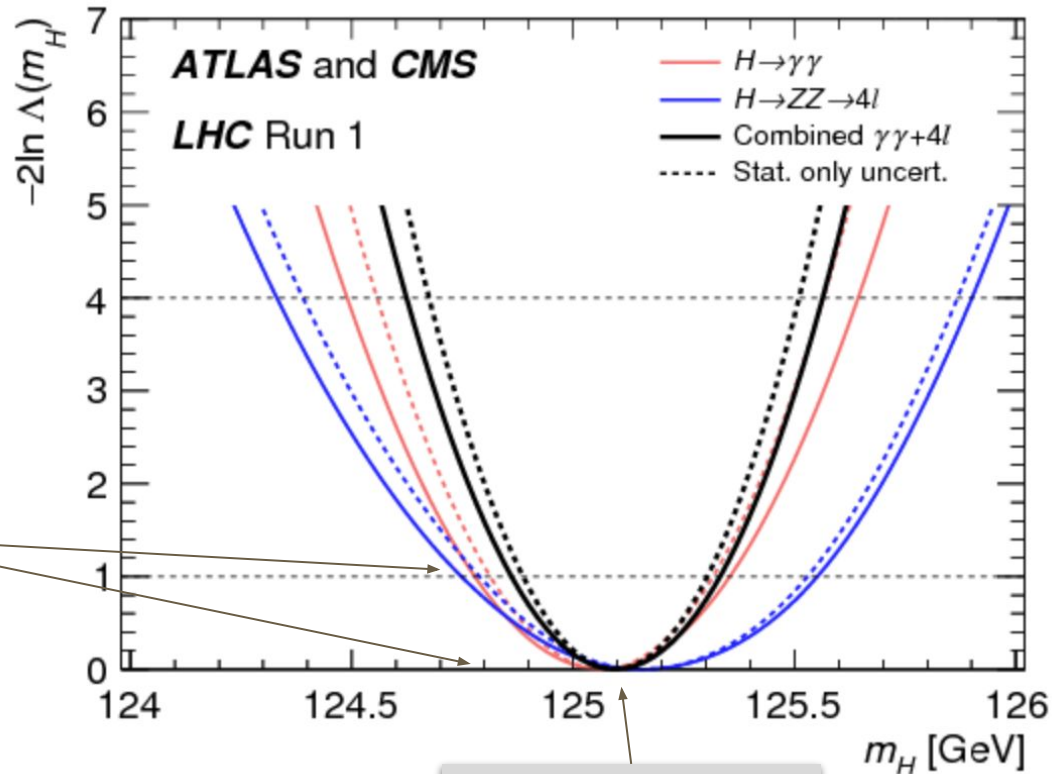
# Maximum-likelihood estimate

- **Maximum-likelihood estimate**  $\Leftrightarrow$  values of parameters that maximize the likelihood
  - Usually: use negative log likelihood
  - Frequentists statistics: **Minimise the NLL (i.e “fit”)**
    - Use minimiser tools, e.g. [Minuit](#)
  - Bayesian statistics: **Sample posterior likelihood**, using Markov-chain Monte Carlo (MCMC)
- If the hypothesized  $\theta$  is close to the true value, then we expect a high probability to get data like that which we actually found
- **ML estimators are not guaranteed** to have any ‘optimal’ properties
  - **In practice they’re very good**
- **Uncertainty of the parameter?**
  - **Value of  $\theta$  where the negative log likelihood shifts by one half** (1 sigma = 0.5, 2 sigma = 2, 3 sigma = 4.5, ...)
    - Motivated by the Normal distribution where shift of 0.5 happens at exactly 1 sigma

*MINUIT*



# Example: Higgs mass measurement - <https://arxiv.org/abs/1503.07589>



Likelihood scan wider when systematic uncertainties are added (next slide)

# Adding systematic uncertainties

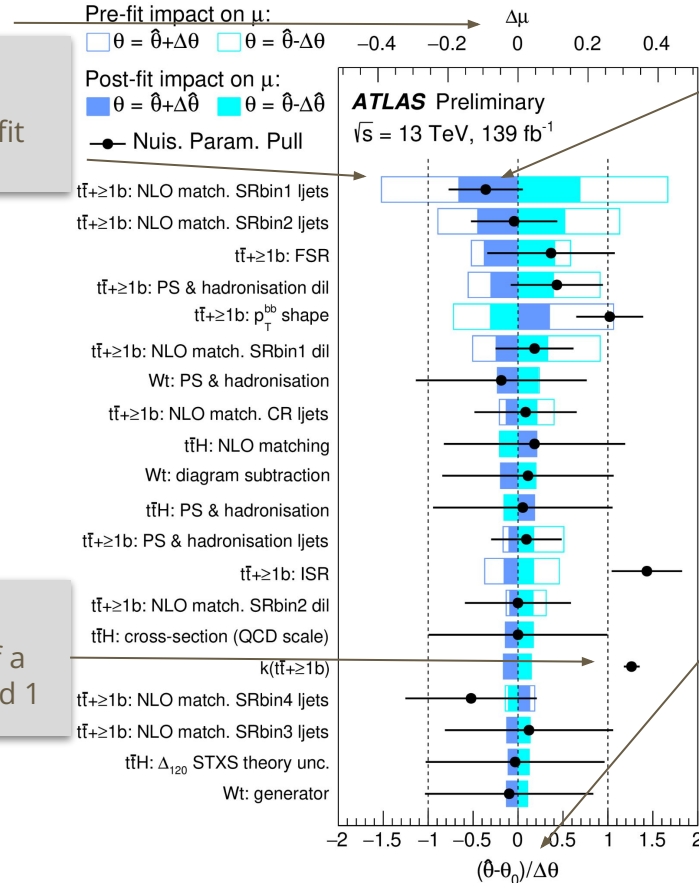
- **Nuisance parameters** (systematic uncertainties) can be added to the likelihood
  - Recall the common model  $L(\vec{n} | \theta, k) = \prod_i P(n_i | S_i(\theta, k) + B_i(\theta, k)) \times \prod_j G(\theta_j)$
- **Maximum-likelihood**  $\Leftrightarrow$  **also the NPs get their best fit value and an uncertainty**
  - **Covariance matrix** of all parameters (including NPs)
    - Can also get **correlations of the parameters** ("post-fit")
  - Lot of physics in these values!
- The uncertainty (likelihood shifts by one half) includes stat+syst
  - **How to get an impact of individual sources of the uncertainties?**
  - Fix a given NP value to  $\pm 1$  sigma, repeat the minimisation and check impact on the parameter of interest
    - Repeat for all NPs
  - Stat-only uncertainty can be obtained by fixing all NPs to their fitted values and repeating the fit and getting the uncertainty on the POI

# Reading pull/ranking plots ATLAS-CONF-2020-058

Impact of a given NP on the POI (ttH signal strength here). Full boxes  $\Leftrightarrow$  post-fit impact, empty boxes  $\Leftrightarrow$  pre-fit impact

NPs “ranked” by their impact on the POI

Some parameters do not have a Gaussian term (e.g. normalisation of a given background)  $\Leftrightarrow$  centred around 1



Central value and uncertainty of a Nuisance parameter indicated with the black point and error bar

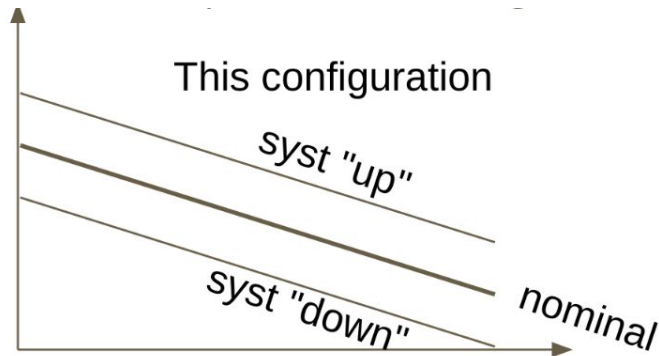
- Is the central value postfit different than 0 (“pull”)?
- Is the post-fit uncertainty smaller than prefit (“constraints”)?

In the model, most of the NPs have a Gaussian term in the likelihood  $\Leftrightarrow$  can talk about “sigmas”.

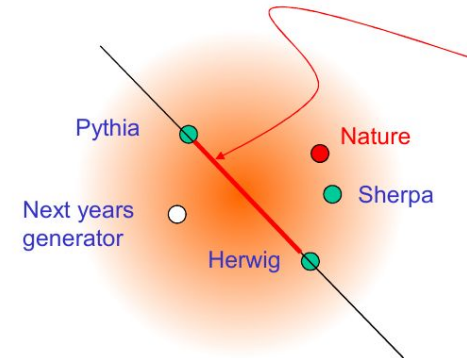
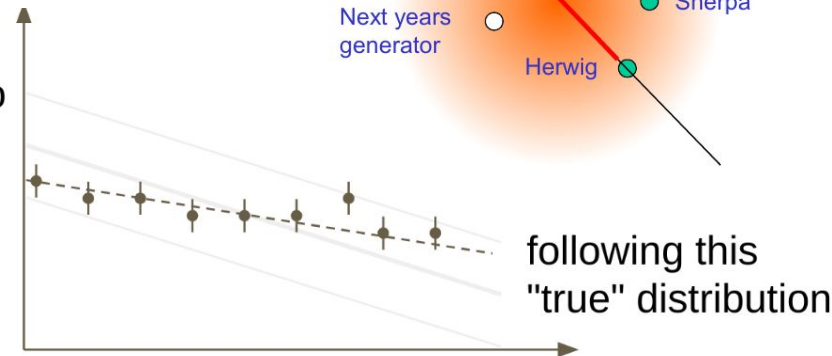
# Dangers of constraining systematic uncertainties

- Post-fit uncertainty smaller than prefit  $\Leftrightarrow$  **constraint**
  - Reduces total uncertainty - good!
  - **Is it reliable?**
    - Should the measurements have power to constrain a given uncertainty?
    - Is the measurements "better" than dedicated calibrations?
    - Are the variation granular enough?
- Usually: pass nominal and  $\pm 1$  sigma variations
  - Interpolation/extrapolation to get **continuous** impact

2-point variations especially problematic!



will not be able to fit these points

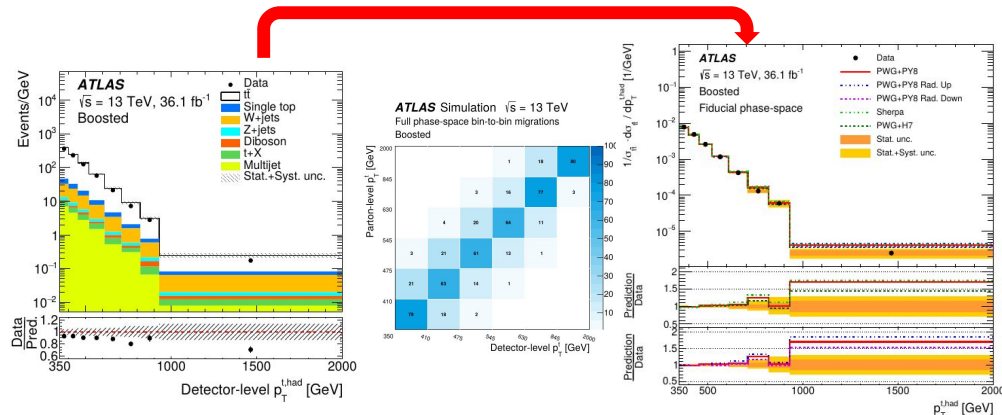


# Unfolding

Slides from: Michele Pinamonti

# What is *unfolding* about?

- **Unfolding** is:
  - removal of detector resolution effects from observed distribution, to extract (our best-guess of) underlying true distribution
  - i.e. extraction of a **differential cross-section**
- Can be done to extract:
  - **total-phase-space** or **fiducial-phase-space** cross-sections
  - cross-sections vs. variable defined at **particle-level** or at **parton-level**
- The unfolding problem can be essentially reduced to a **response-matrix-inversion** problem

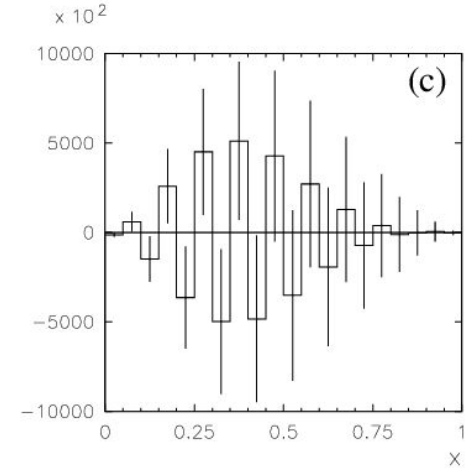
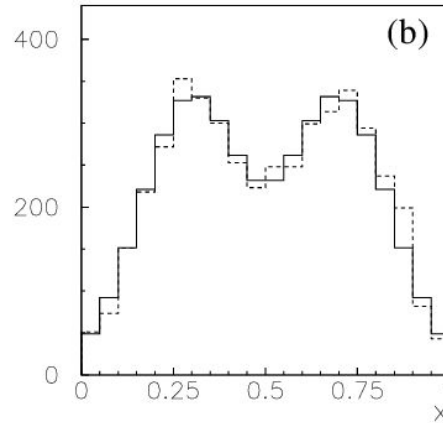
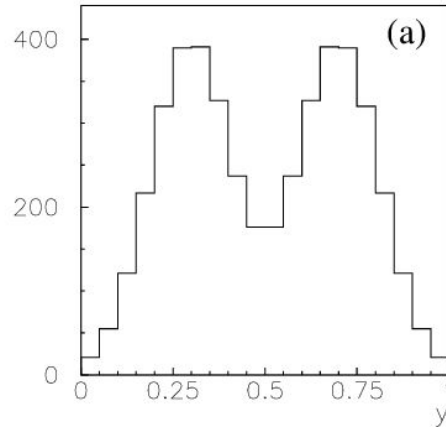


# The regularization concept

*“To regularize or not to regularize?  
This is the question...”*



- Most delicate point is the so-called **regularization**:
  - introduced to avoid **amplification of statistical fluctuations** in unfolded data (**oscillations**), happening when just **inverting** response matrix



- Regularization techniques always imply some level of **assumptions**  $\Rightarrow$  inevitable **bias**
  - Variance-bias optimisation

# Tikhonov regularisation

- Recall the unfolding problem  $A\vec{x} = \vec{b}$
- This can be reformulated as a **minimisation** problem (chi-square):  $\chi^2 = (A\vec{x} - \vec{b})^T (A\vec{x} - \vec{b}) = \min$ 
  - Can minimise to find the best fit for  $\vec{x}$
  - Can **impose some additional constraint** (will bias the result!)

$$L(\vec{x}) \equiv \chi^2(\vec{x}) + \Phi(\vec{x}) \rightarrow \min$$

- **Common choice** for the constraint: **second discrete derivative (Tikhonov)**

$$\Phi(\vec{x}) = \tau \sum_i (x_{i-1} - 2x_i + x_{i+1})^2$$

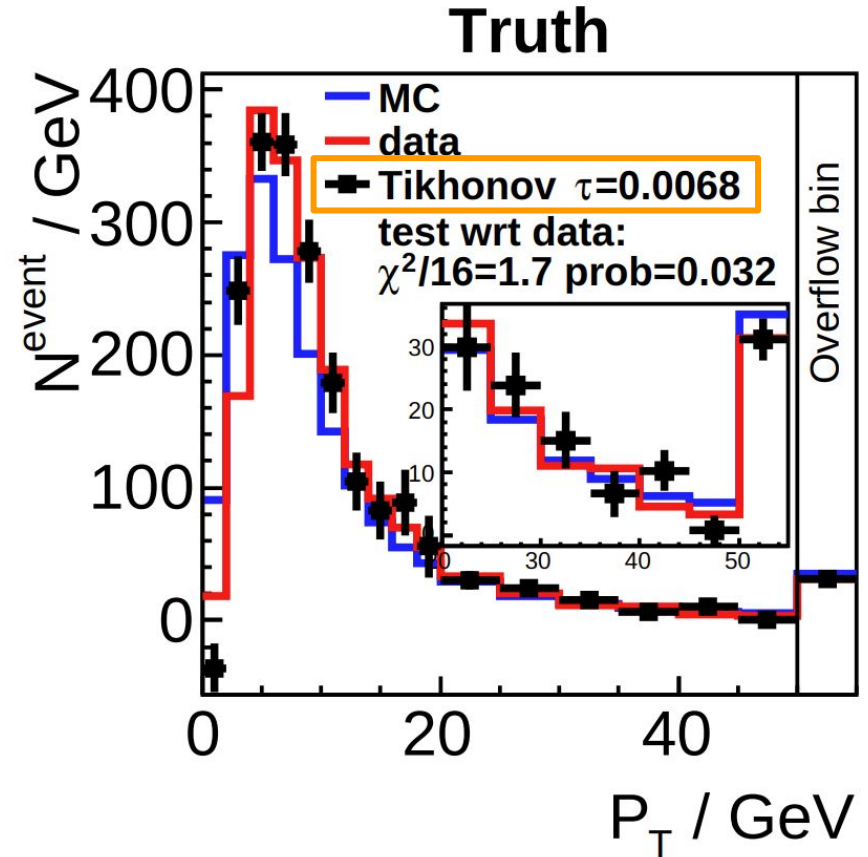
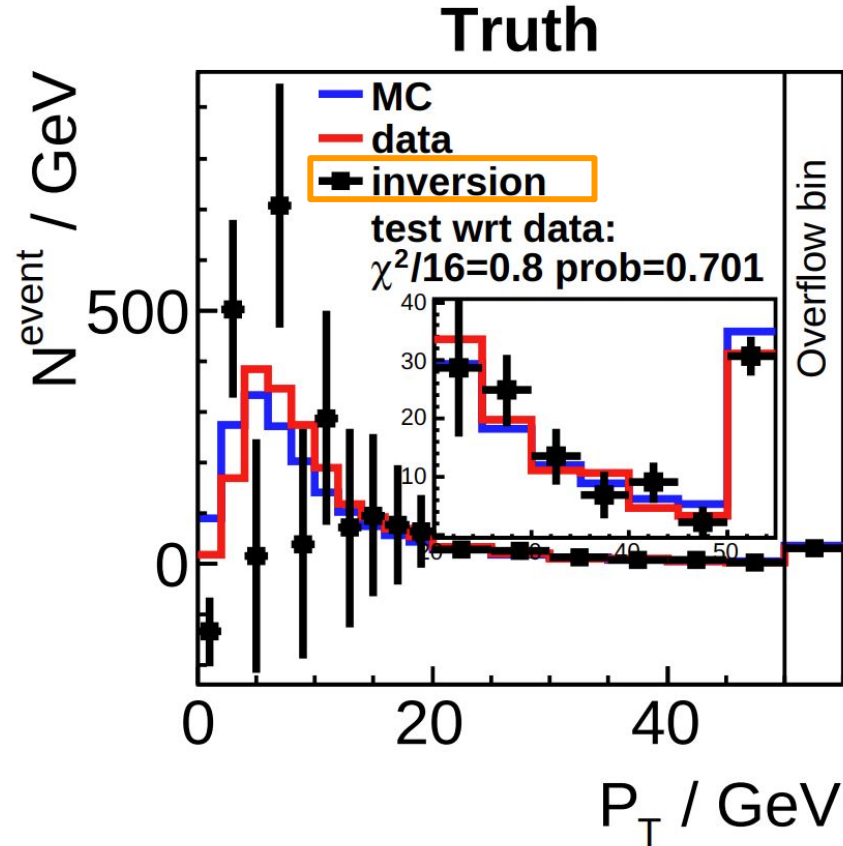
- Choice of  $\tau \Leftrightarrow$  strength of the regularisation
- Different choices of  $\Phi(\vec{x})$  possible - e.g. **SVD**  $\longrightarrow A = U S V^T$ 
  - See e.g. <https://arxiv.org/abs/hep-ph/9509307>



# Impact of regularisation

Taken from:

<https://arxiv.org/abs/1611.01927>



# Iterative Bayesian Unfolding (IBU)

- Frequently used in high-signal measurements
- Uses Bayes theorem iteratively:

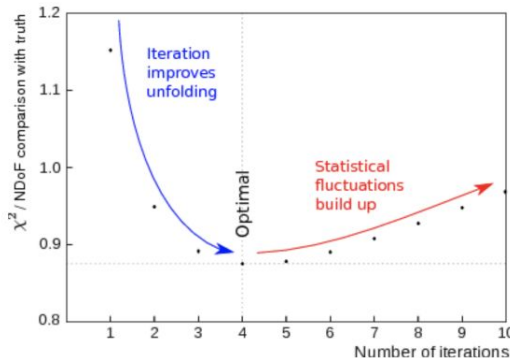
$$\underbrace{p(T|D, \mathcal{M})}_{\text{posterior}} \propto \underbrace{\mathcal{L}(D|T, \mathcal{M})}_{\text{likelihood}} \cdot \underbrace{\pi(T)}_{\text{prior}}$$

*true distribution* points to  $p(T|D, \mathcal{M})$   
*data ("reco") distribution* points to  $\mathcal{L}(D|T, \mathcal{M})$   
*response matrix* points to  $\mathcal{L}(D|T, \mathcal{M})$

- prior based on theoretical prediction in first iteration
- following iterations use result of previous ones as prior



$$\begin{aligned} p_1(T|D) &\propto \mathcal{L} \cdot \pi(T) \\ p_2(T|D) &\propto \mathcal{L} \cdot p_1(T|D) \\ p_3(T|D) &\propto \mathcal{L} \cdot p_2(T|D) \\ &\dots \end{aligned}$$



## Regularization:

- achieved by stopping after a few iterations  
( $N_{\text{iter}} \rightarrow \infty \Rightarrow$  unregularized unfolding, i.e. matrix inversion)
- finding optimal stopping point is an important feature of using IBU

# Thank you for your attention

## Questions?

“If your experiment needs a statistician, you need a better experiment.”

— Ernest Rutherford