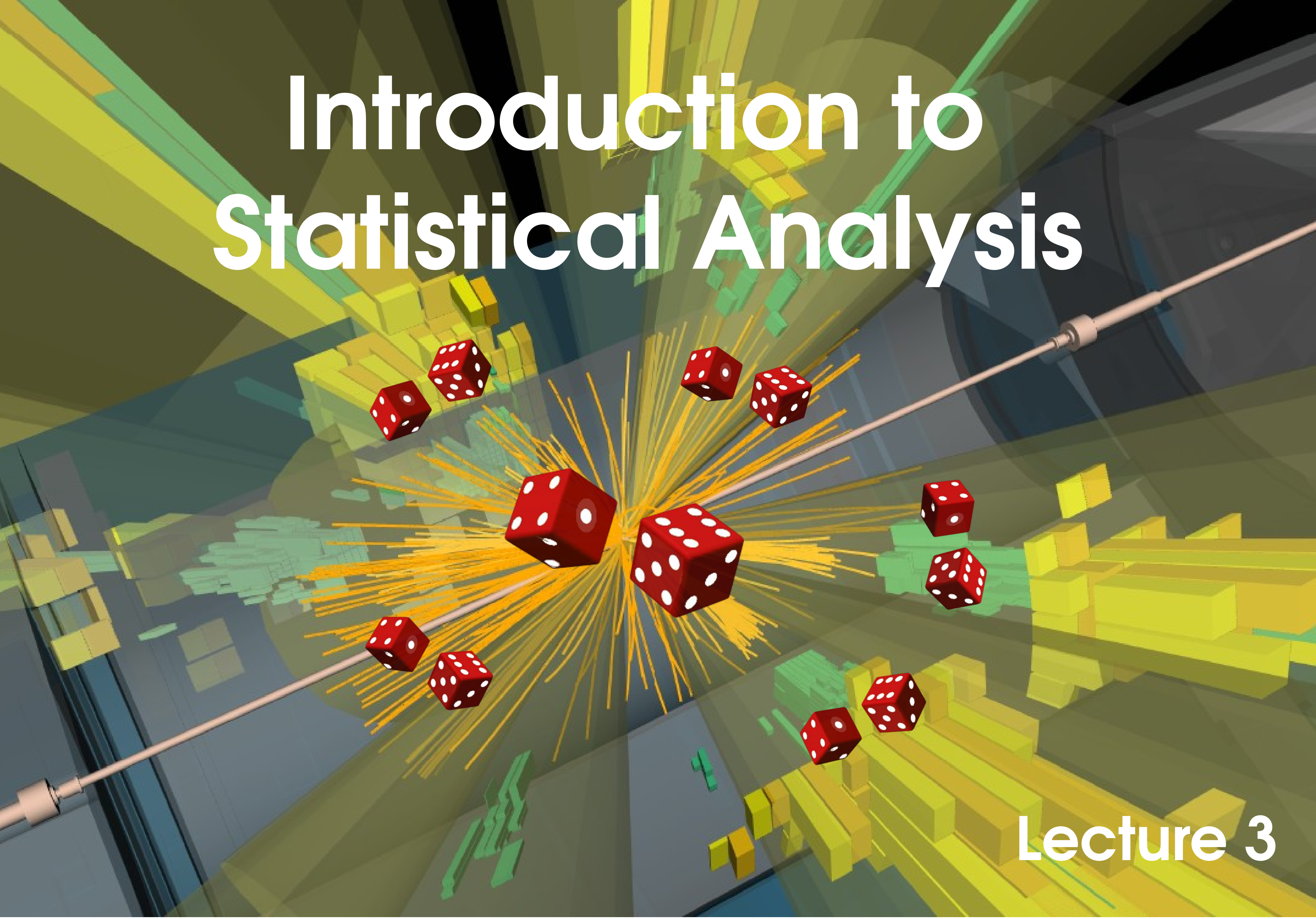
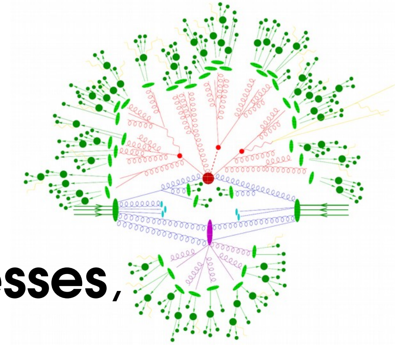
Introduction to Statistical Analysis



Lecture 3

Nicolas Berger (LAPP Annecy)

Reminders From Lecture 1



Physics measurement data are produced through **random processes**,
Need to be described using a statistical model:

Description	Observable	Likelihood
Counting	n	**Poisson** $P(n; S, B) = e^{-(S+B)} \frac{(S+B)^n}{n!}$
Binned shape analysis	n_i, $i=1..N_{bins}$	**Poisson product** $P(n_i; S, B) = \prod_{i=1}^{n_{bins}} e^{-(S f_i^{sig} + B f_i^{bkg})} \frac{(S f_i^{sig} + B f_i^{bkg})^{n_i}}{n_i!}$
Unbinned shape analysis	m_i, $i=1..n_{evts}$	**Extended Unbinned Likelihood** $P(m_i; S, B) = \frac{e^{-(S+B)}}{n_{evts}!} \prod_{i=1}^{n_{evts}} S\, P_{sig}(m_i) + B\, P_{bkg}(m_i)$

Model can include multiple **categories**, each with a separate description
Includes **parameters of interest** (POIs) but also **nuisance parameters** (NPs)

Reminders from Lecture 2: Discovery Significance

Given a statistical model P(data; μ), define likelihood **L(μ) = P(data; μ)**

To estimate a parameter, use the value $\hat{\mu}$ that maximizes L(μ).

To decide between hypotheses H_0 and H_1, use the **likelihood ratio** $$\frac{L(H_0)}{L(H_1)}$$

To test for **discovery**, use $$q_0 = -2\log\frac{L(S=0)}{L(\hat{S})} \quad \hat{S} \geq 0$$

For large enough datasets (n > 5), $$Z = \sqrt{q_0}$$

For a **Gaussian** measurement, $$Z = \frac{\hat{S}}{\sqrt{B}}$$

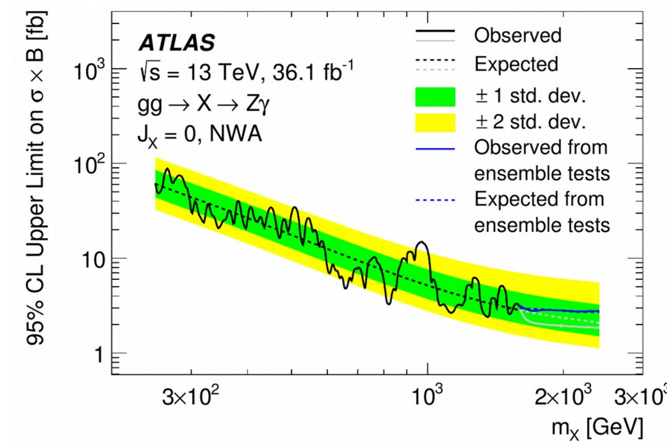
For a **Poisson** measurement, $$Z = \sqrt{2\left[(\hat{S}+B)\log\left(1+\frac{\hat{S}}{B}\right)-\hat{S}\right]}$$

Reminders from Lecture 2: Limits & Intervals

Limits : use LR-based test statistic:

$$q_{S_0} = -2\log\frac{L(S=S_0)}{L(\hat{S})} \qquad \hat{S} \le S_0$$

→ Use **CL$_s$ procedure** to avoid negative limits

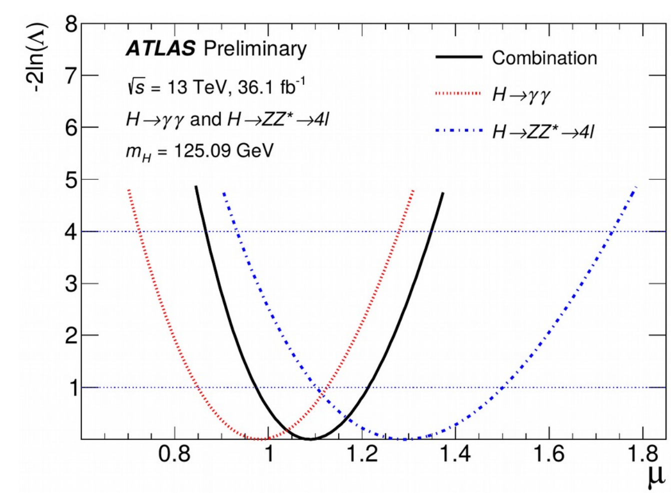


Poisson regime, n=0 : **S_{up} = 3 events**

Confidence intervals: use

$$t_{\mu_0} = -2\log\frac{L(\mu=\mu_0)}{L(\hat{\mu})}$$

→ 1D: crossings with $t_{\mu 0} = Z^2$ for $\pm Z\sigma$ intervals



Gaussian regime: $\mu = \hat{\mu} \pm \sigma_{Gauss}$ for a 1σ interval

Outline

- Expected results and toys
  - Pseudo-experiments and Asimov datasets
  - Dealing with non-asymptotic situations
- Profiling
- Look-Elsewhere Effect
- Bayesian methods
- Presentation of results

Generating Pseudo-data

Model describes the distribution of the observable: **P(data; parameters)**

⇒ Possible outcomes of the experiment, for given parameter values

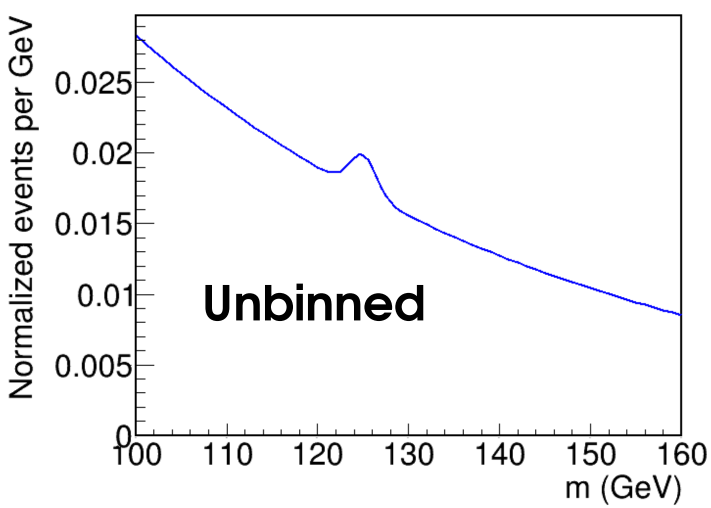
Can draw random events according to PDF : **generate** ***pseudo-data***

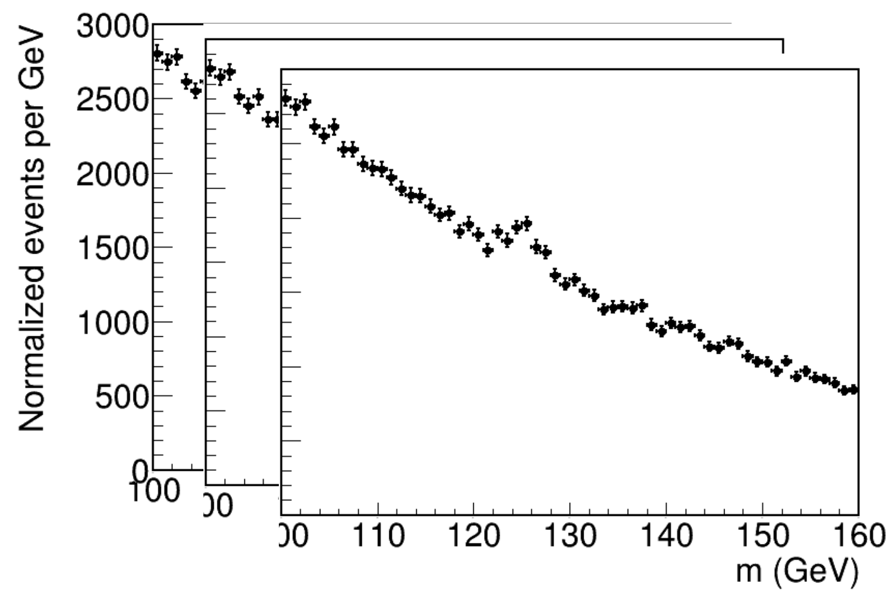
$$P(\lambda = 5)$$

2, 5, 3, 7, 4, 9, ...

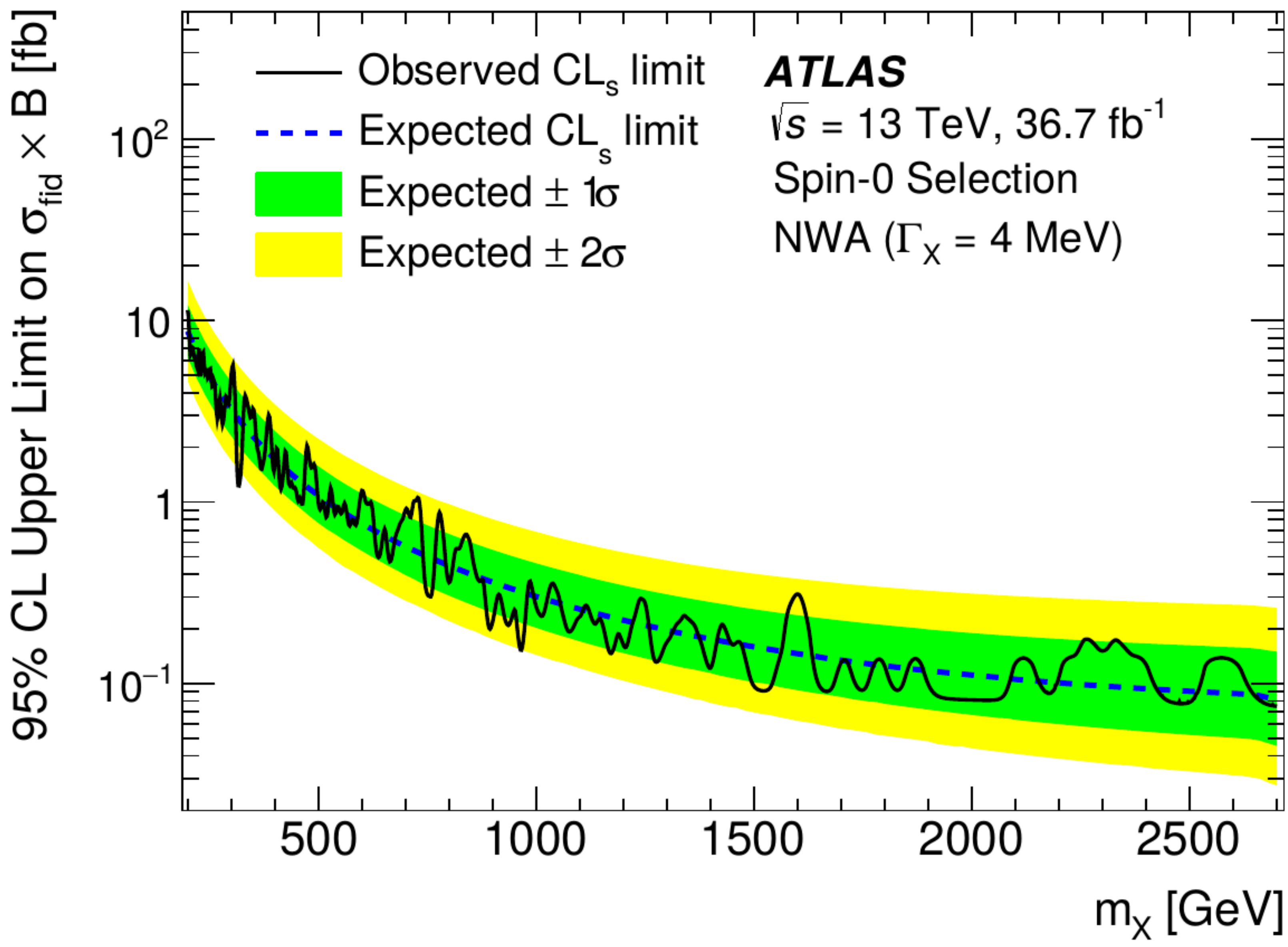
Each entry = separate "experiment"

Generate





Expected Results



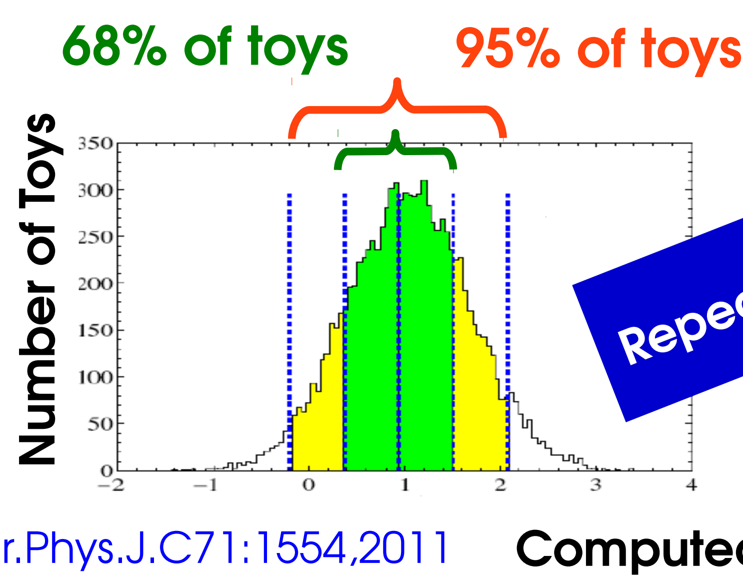
Expected Limits: Toys

***Expected* results**: median outcome under a given hypothesis
→ usually B-only for searches, but other choices possible.

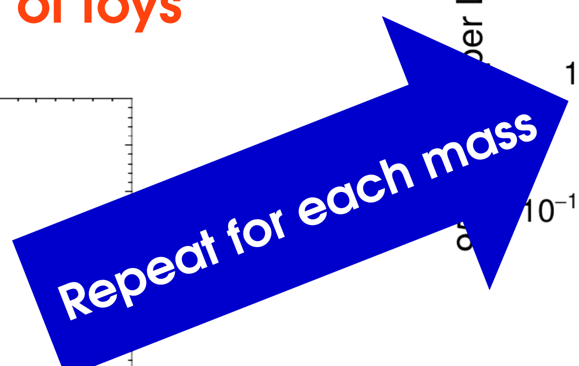
Two main ways to compute:

→ Pseudo-experiments (*toys*):

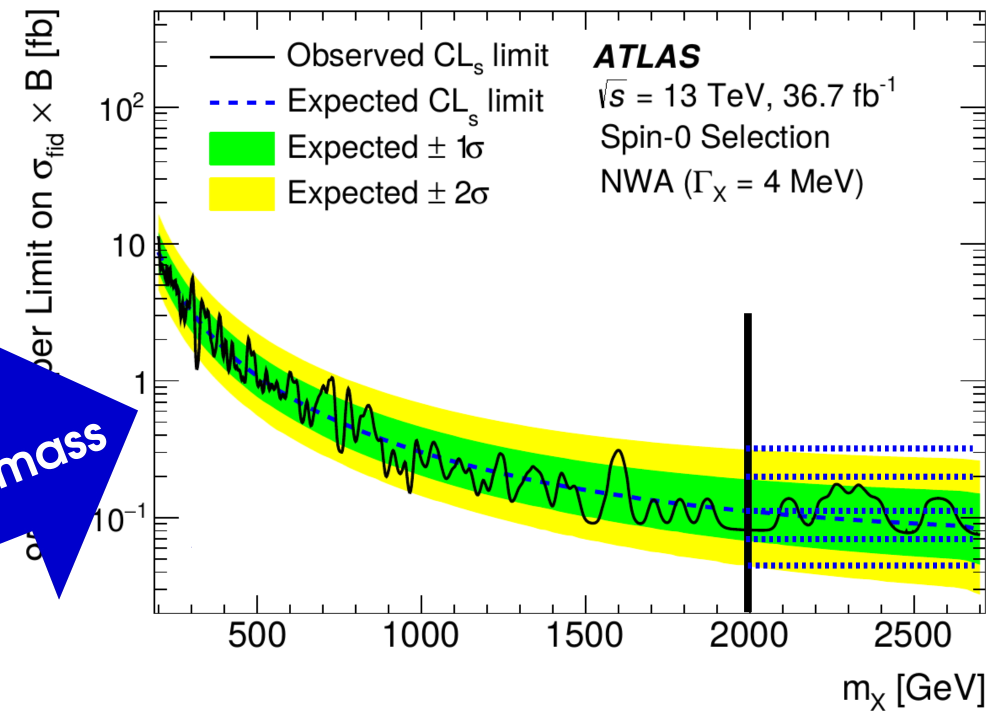
- Generate a pseudo-dataset in B-only hypothesis
- Compute limit
- Repeat and histogram the results
- Central value = median, bands based on quantiles



Eur.Phys.J.C71:1554,2011



Phys. Lett. B 775 (2017) 105



Expected Limits: Asimov Datasets

***Expected* results**: median outcome under a given hypothesis
→ usually B-only by convention, but other choices possible.

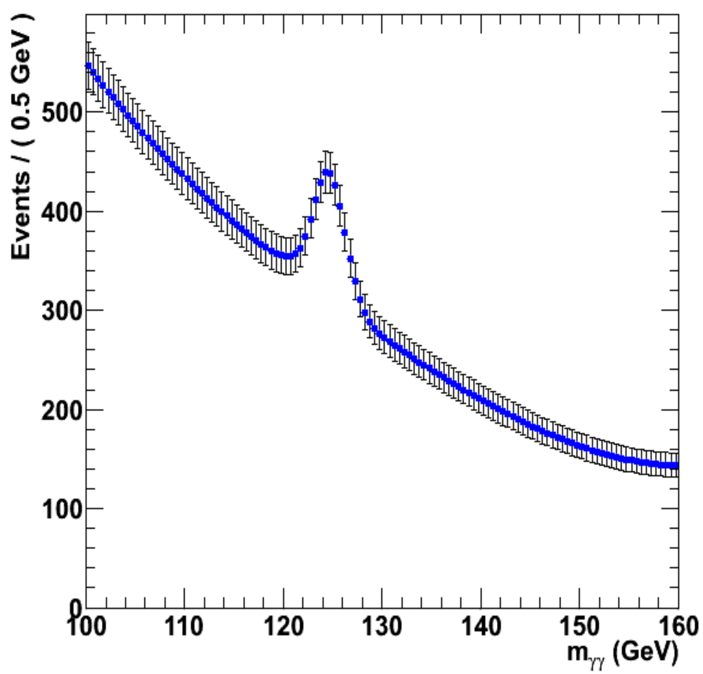
Two main ways to compute:

Strictly speaking, Asimov dataset if
$\hat{X} = X_0$ **for all parameters** X,
where X_0 is the generation value

→ Asimov Datasets

- Generate a "perfect dataset" – *e.g.* for binned data, set bin contents carefully, no fluctuations.
- Gives the median result immediately:
  median(toy results) ↔ result(median dataset)
- Get bands from asymptotic formulas:
  Band width

$$\sigma^2_{S_0,A} = \frac{S_0^2}{q_{S_0}(\text{Asimov})}$$



⊕ **Much faster (1 "toy")**

⊖ **Relies on Gaussian approximation**

Outline

- Expected results and toys
  - Pseudo-experiments and Asimov datasets
  - Dealing with non-asymptotic situations
- Profiling
- Look-Elsewhere Effect
- Bayesian methods
- Presentation of results

Beyond Asymptotics: Toys

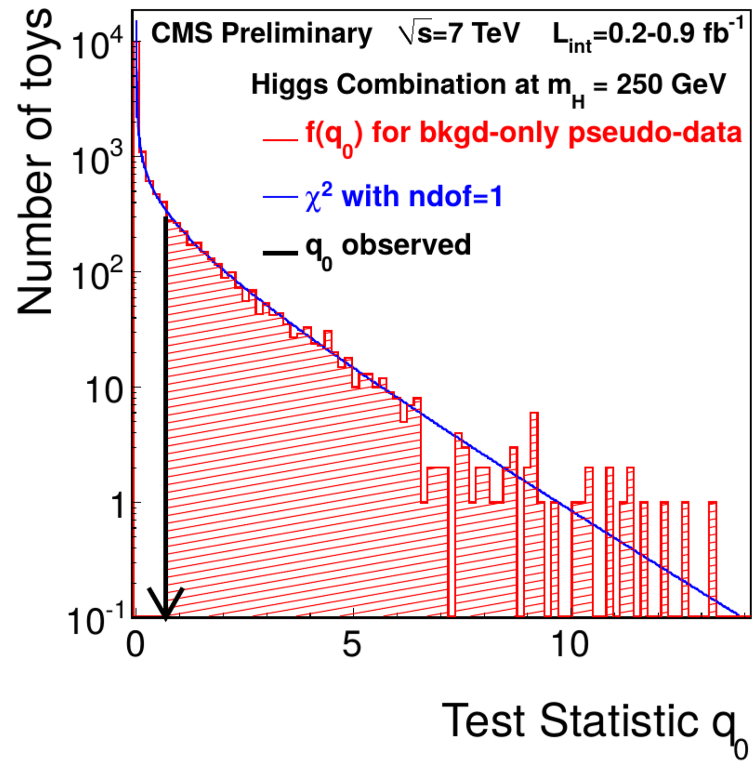
CMS-PAS-HIG-11-022

Asymptotics usually work well, but break down in some cases – e.g. **small event counts.**

Solution: generate ***pseudo data*** (**toys**) using the PDF, under the tested hypothesis

→ Also randomize the observable (θ^{obs}) of each auxiliary experiment: $G(\theta^{obs};\theta,\sigma_{syst})$

→ Samples the true distribution of the PLR

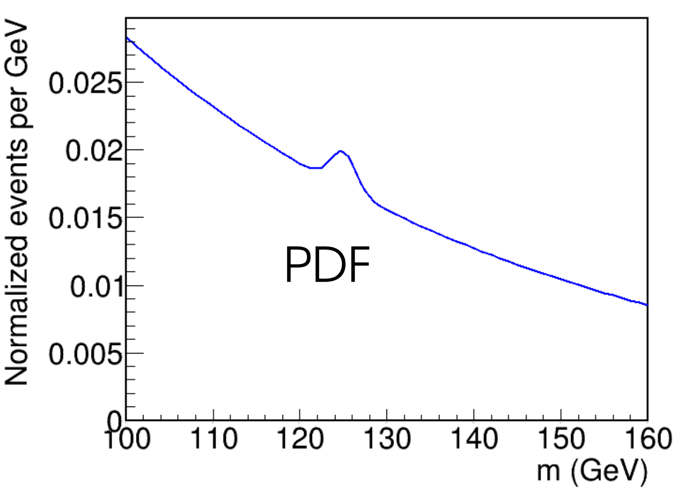


⇒ Integrate above observed PLR to get the p-value

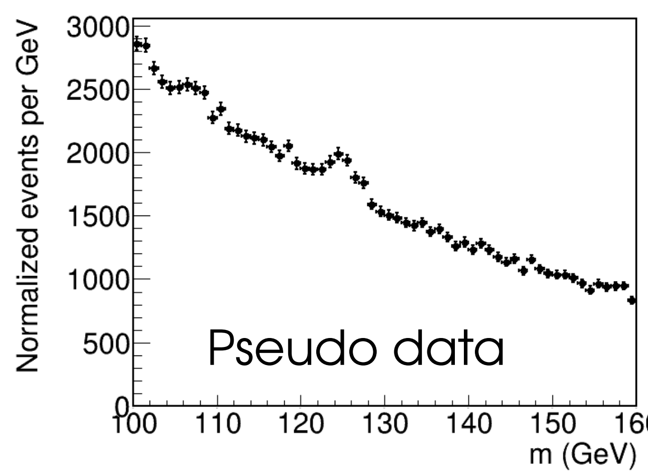
→ Precision limited by number of generated toys,

Small p-values (5σ : p~10^{-7}!) ⇒ **large toy samples**

Repeat N_{toys} times



p(data|x)



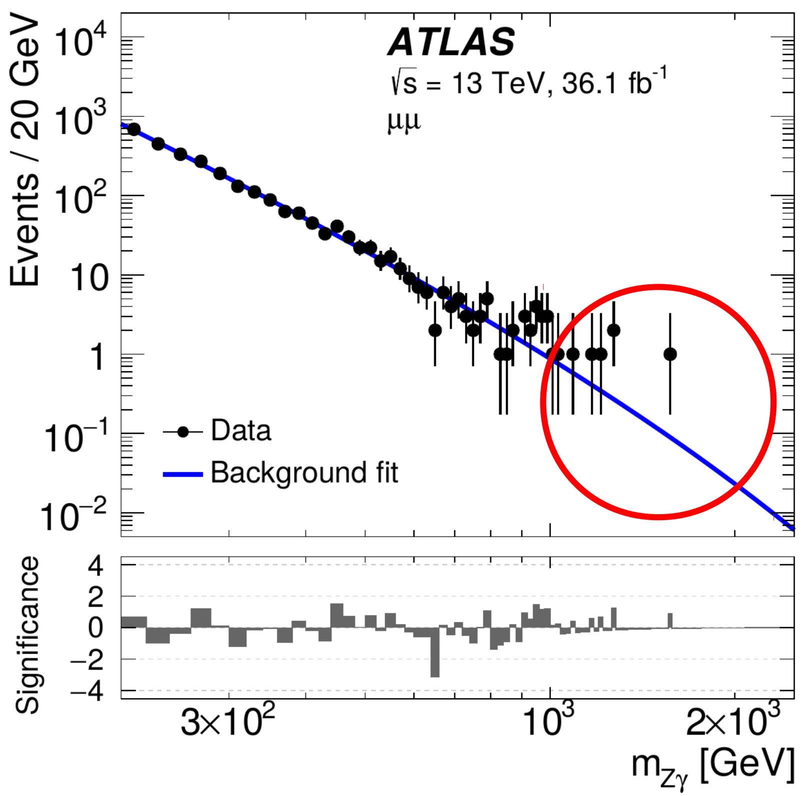
q_0

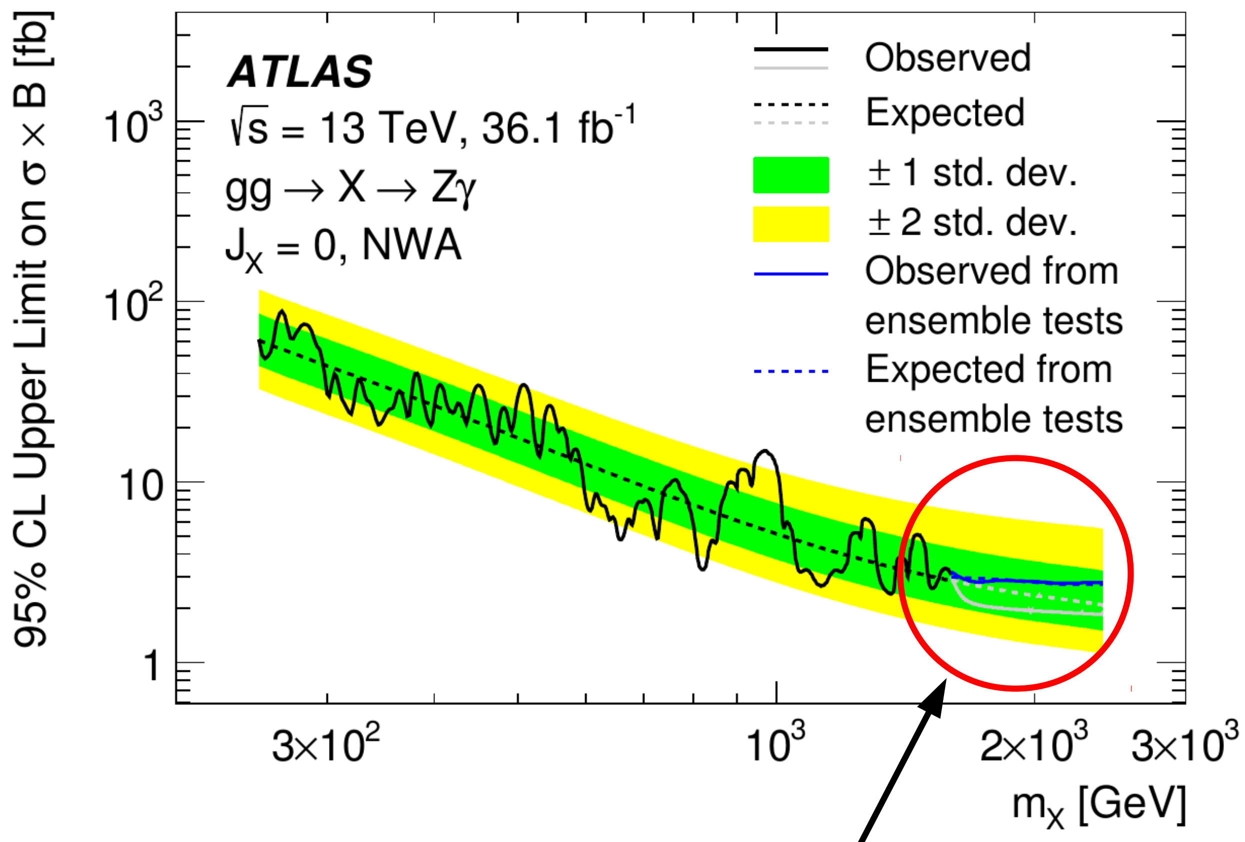
Toys: Example

JHEP 10 (2017) 112

ATLAS $X \to Z\gamma$ Search: covers 200 GeV < m_X < 2.5 TeV

→ for m_X > 1.6 TeV, low event counts ⇒ derive results from toys



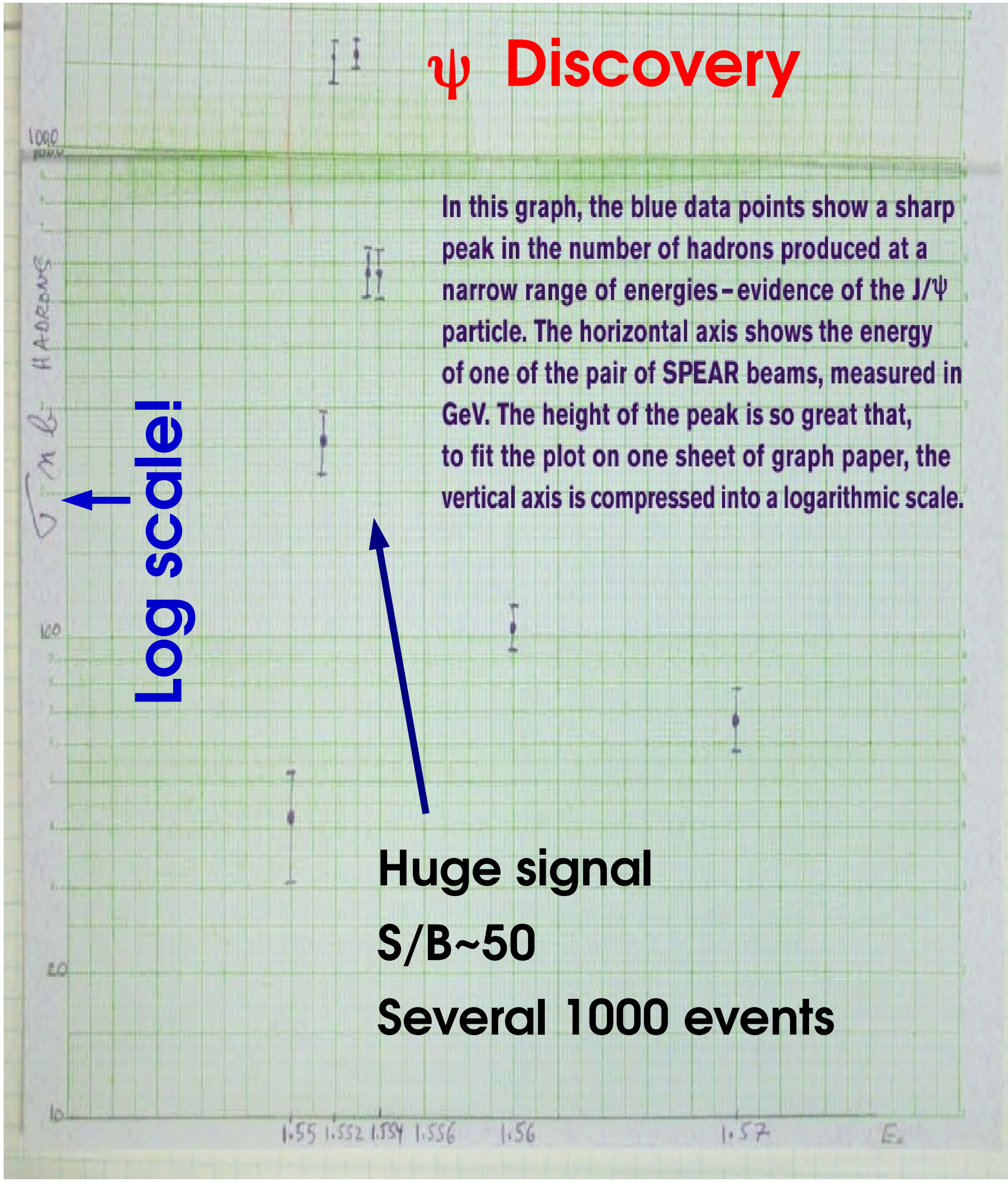


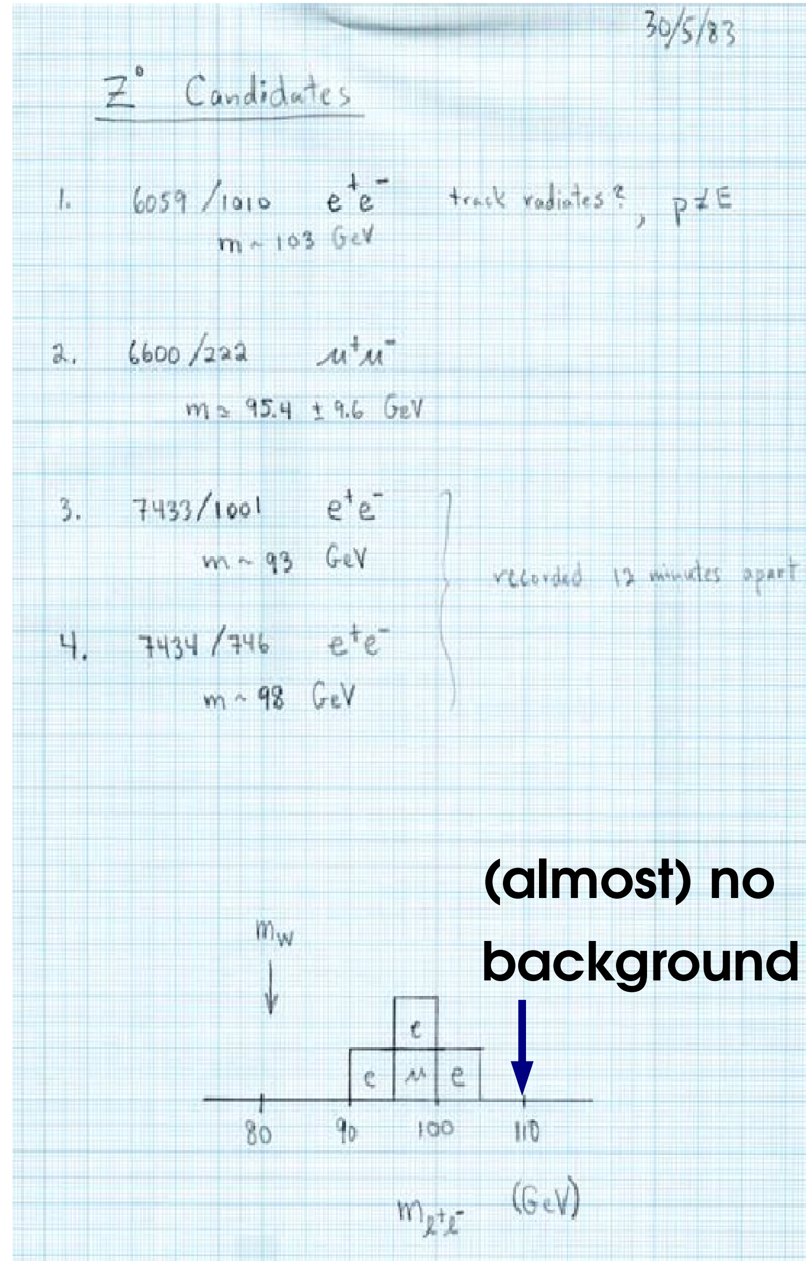
Asymptotic results (in gray) give optimistic result compared to toys (**in blue**)

Historical Aside

Classic Discoveries (1)

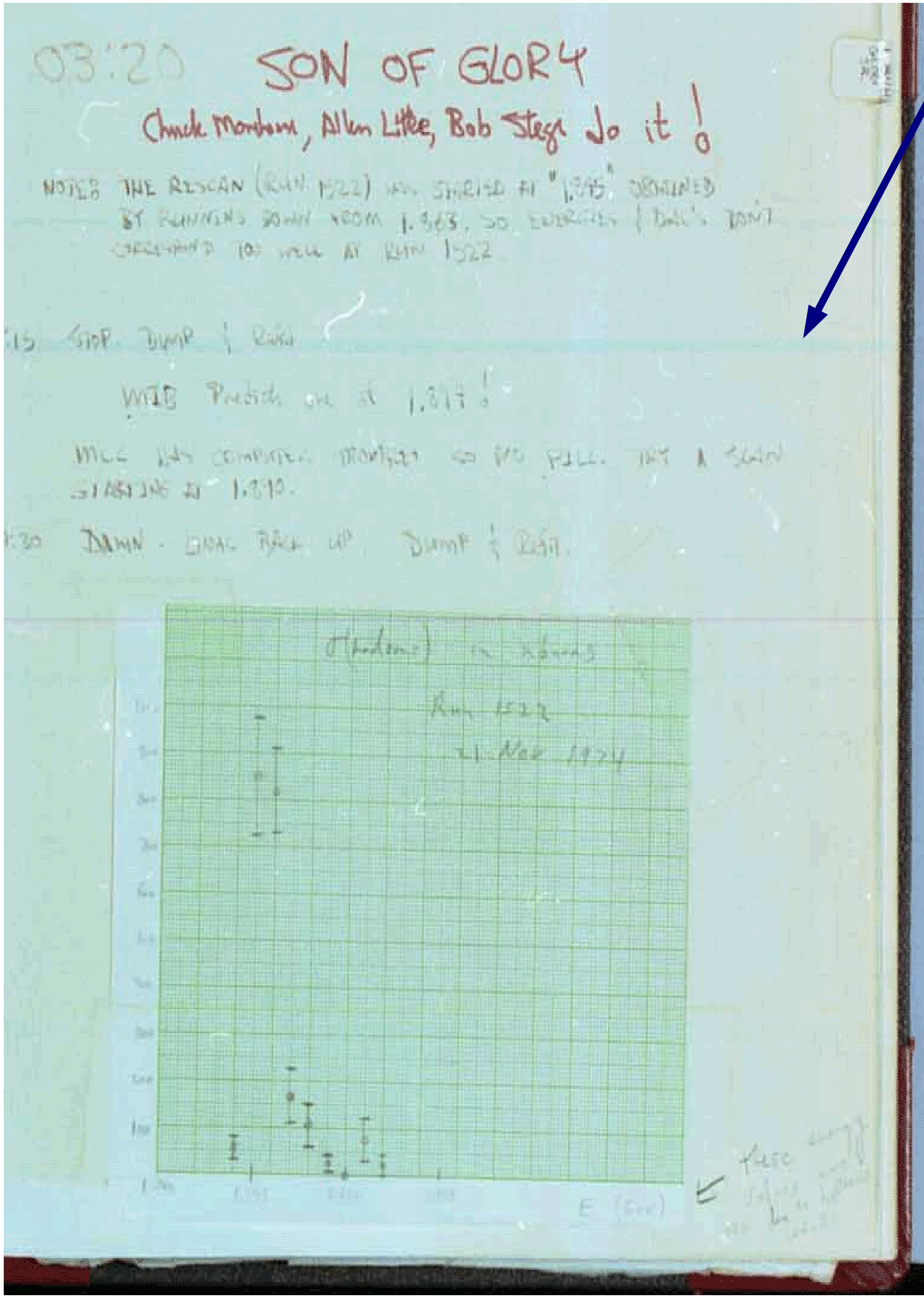
Z^0 Discovery



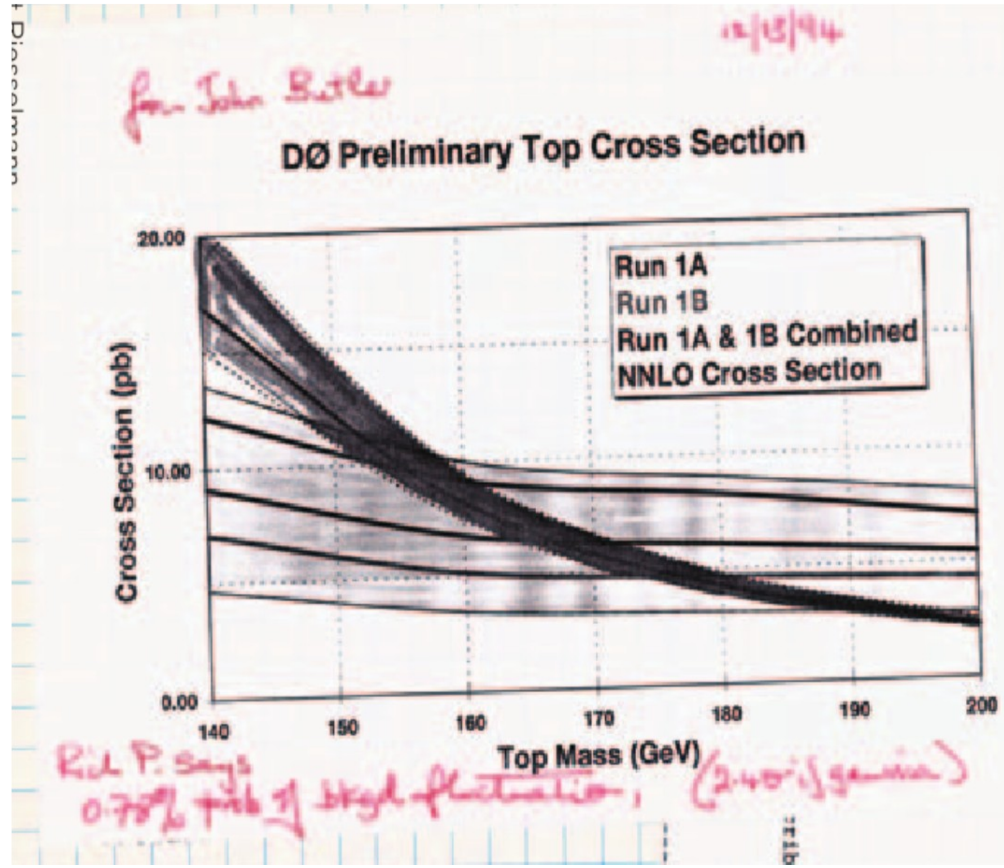


Logbook of J. Rohlf, 1983-05-30

Classic Discoveries (2)



ψ' : discovered in the control room by the (lucky) shifters



First hints of top at D0:
O(10) signal events,
a few bkg events, 2.4σ

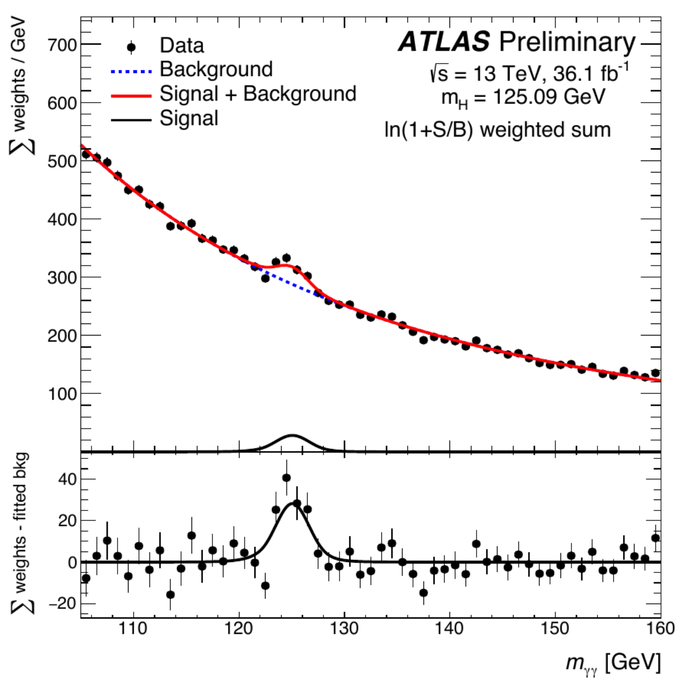
And now ?

Short answer: The high-signal, low-background experiments have been done already (although a surprise would be welcome...)

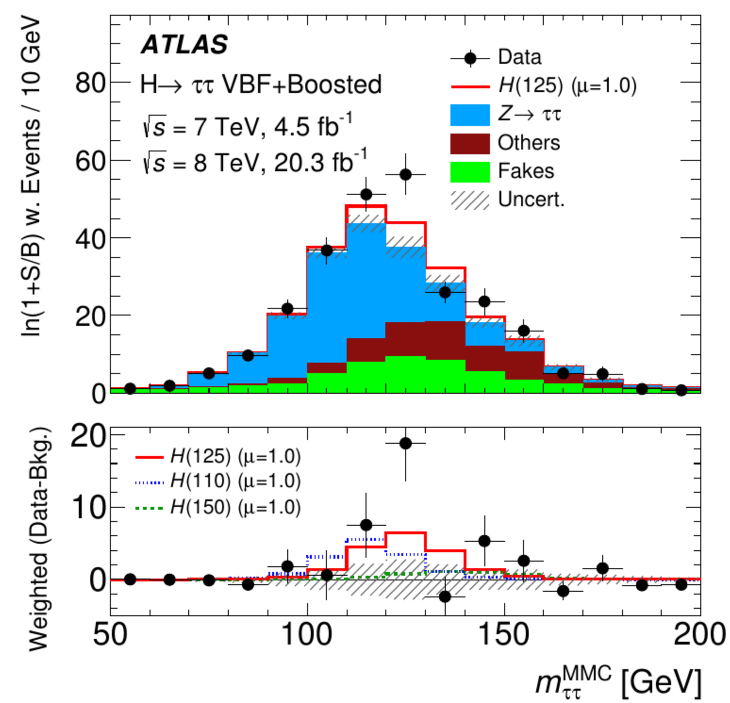
e.g. at LHC:

- **High background levels**, need precise modeling
- **Large systematics,** need to be described accurately
- **Small signals**: need optimal use of available information :
  - **Shape analyses** instead of counting
  - **Categories** to isolated signal-enriched regions

ATLAS-CONF-2017-045



JHEP 12 (2017) 024



Outline

- Expected results and toys
  - Pseudo-experiments and Asimov datasets
  - Dealing with non-asymptotic situations
- **Profiling**
- Look-Elsewhere Effect
- Bayesian methods
- Presentation of results

Reminder: Wilks' Theorem

Cowan, Cranmer, Gross & Vitells
Eur.Phys.J.C71:1554,2011

Consider $t_{S_0} = -2\log\frac{L(S=S_0)}{L(\hat{S})}$

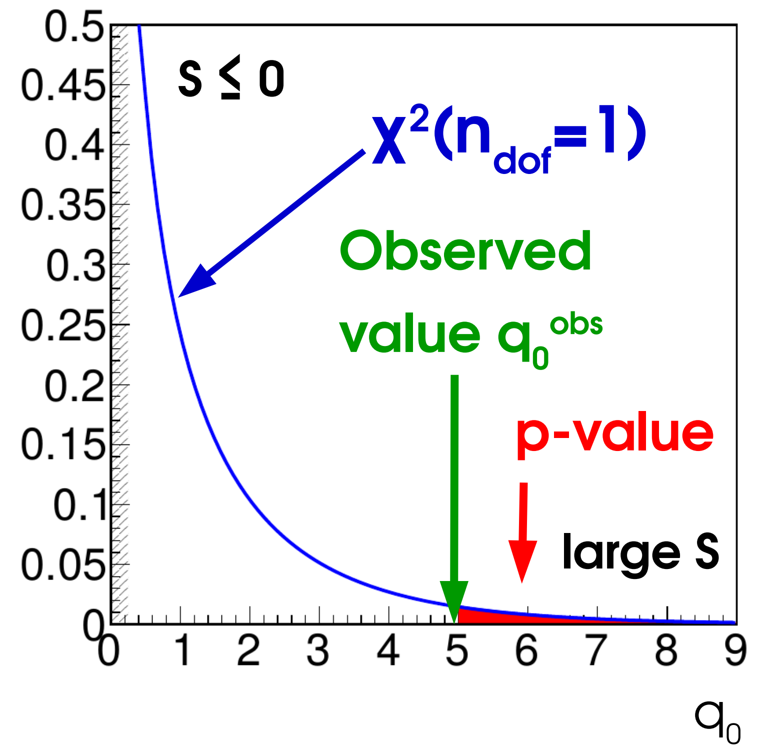
→ Assume **Gaussian regime** (e.g. large n_{evts}, Central-limit theorem) : then:

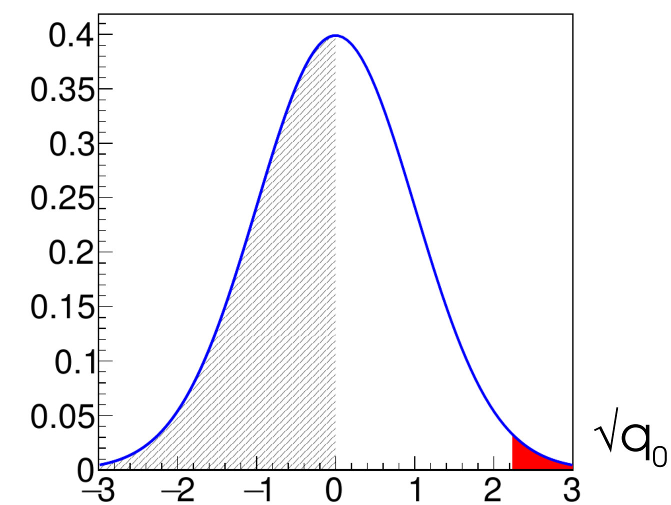
Wilk's Theorem: **t_{S0} is distributed as a χ^2** under $H_{S0}(S=S_0)$:

$$f\left(t_{S_0} \mid S=S_0\right) = f_{\chi^2(n_{dof}=1)}\left(t_{S_0}\right)$$

⇒ The significance is:

$$Z = \sqrt{q_0}$$





Profiling

How to deal with nuisance parameters in likelihood ratios ?

→ **Let the data choose ⇒ use the best-fit values (*Profiling*)**

⇒ **Profile Likelihood Ratio** (PLR)

$$t_{S_0} = -2\log\frac{L(S=S_0, \hat{\hat{\theta}}(S_0))}{L(\hat{S}, \hat{\theta})}$$

$\hat{\hat{\theta}}(S_0)$ best-fit value for $S=S_0$ (conditional MLE)

$\hat{\theta}$ overall best-fit value (unconditional MLE)

***Wilks' Theorem*: same properties as plain likelihood ratio**

$$f\left(t_{S_0} \mid S=S_0\right) = f_{\chi^2(n_{dof}=1)}\left(t_{S_0}\right)$$

also with NPs present

→ Profiling "builds in" the effect of the NPs

⇒ **Can use t_{S0} to compute limits, significance, etc. in the same way as before**

Homework 7: Gaussian Profiling

Counting experiment with background uncertainty: **n = S + B** :

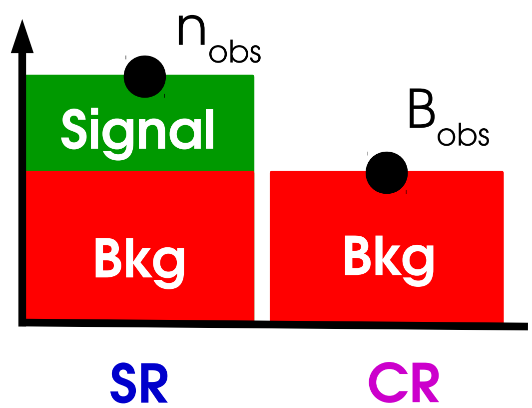
→ **Signal region (SR)**: $n_{obs} \sim G(S + B, \sigma_{stat})$

→ **Control region (CR)**: $B_{obs} \sim G(B, \sigma_{bkg})$

$$L(S,B) = G(n_{obs}; S + B, \sigma_{stat})\, G(B_{obs}; B, \sigma_{bkg})$$

Recall: Signal region only (fixed B):

$$t_S = \left(\frac{S - n_{obs}}{\sigma_{stat}}\right)^2 \qquad S = (n_{obs} - B) \pm \sigma_{stat}$$



→ Compute the best-fit (MLEs) for S and B

→ Show that the conditional MLE for B is

$$\hat{\hat{B}}(S) = B_{obs} + \frac{\sigma_{bkg}^2}{\sigma_{stat}^2 + \sigma_{bkg}^2}(\hat{S} - S)$$

→ Compute the profile likelihood t_S

→ Compute the 1σ confidence interval on S

$$S = (n_{obs} - B_{obs}) \pm \sqrt{\sigma_{stat}^2 + \sigma_{bkg}^2} \qquad \sigma_S = \sqrt{\sigma_{stat}^2 + \sigma_{bkg}^2}$$

Stat uncertainty (on n) and systematic (on B) add in quadrature

Systematics Implementation

Prototype: NP measured in a separate ***auxiliary* experiment**
e.g. luminosity measurement

→ Build the combined likelihood of the main+auxiliary measurements

$$L(\mu, \theta; \mathrm{data}) = L_{\mathrm{main}}(\mu, \theta; \mathrm{main\ data})\ L_{\mathrm{aux}}(\theta; \mathrm{aux.\ data})$$

Independent measurements:
⇒ just a product

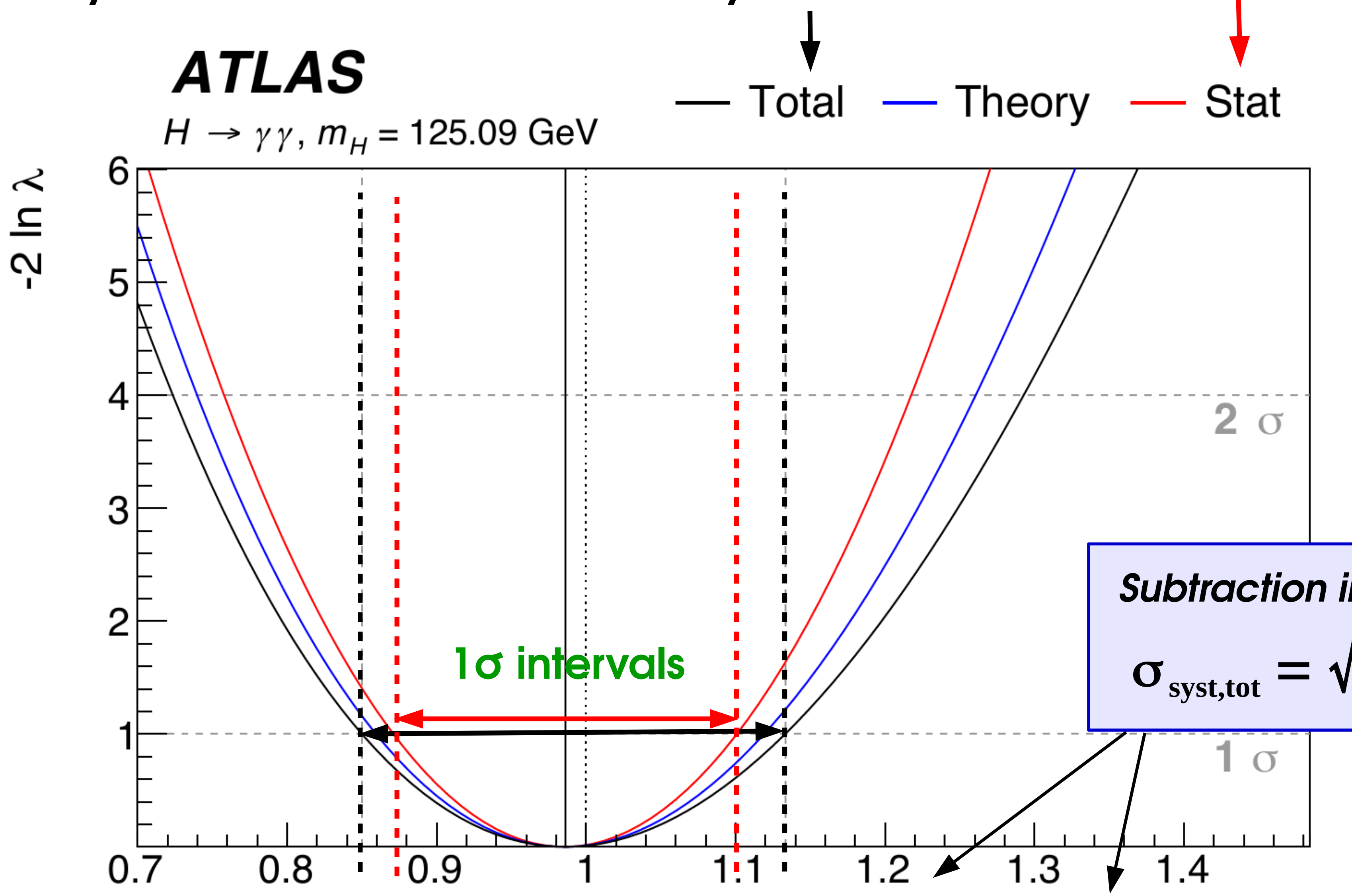
Gaussian form often used by default: $L_{\mathrm{aux}}(\theta; \mathrm{aux.\ data}) = G(\theta^{\mathrm{obs}}; \theta, \sigma_{\mathrm{syst}})$

→ Often no clear setup for auxiliary measurements
e.g. theory uncertainties on missing HO terms from scale variations
→ **Implemented in the same way nevertheless** ("pseudo-measurement")

Uncertainty decomposition

All systematics NPs excluded : statistical uncertainty only

All systematics NPs included: stat+syst uncertaintes

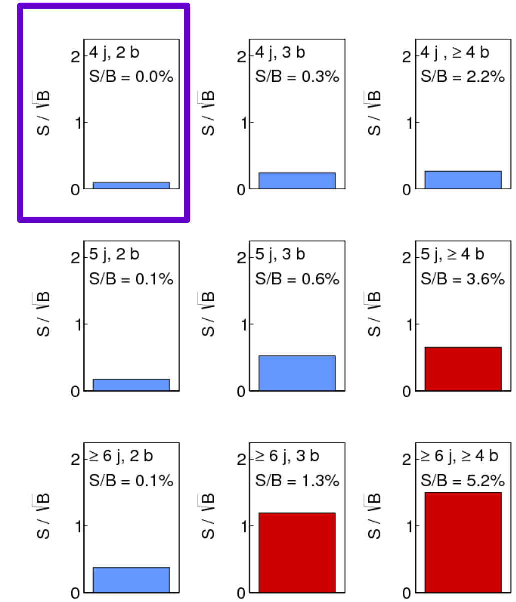


Subtraction in quadrature

$$\sigma_{\text{syst,tot}} = \sqrt{\sigma^2_{\text{total}} - \sigma^2_{\text{stat}}}$$

$$\mu = 0.99 \pm 0.12\,(\text{stat}) \pm 0.06\,(\text{syst}) \pm 0.06\,(\text{theo})$$

Profiling Example: ttH→bb

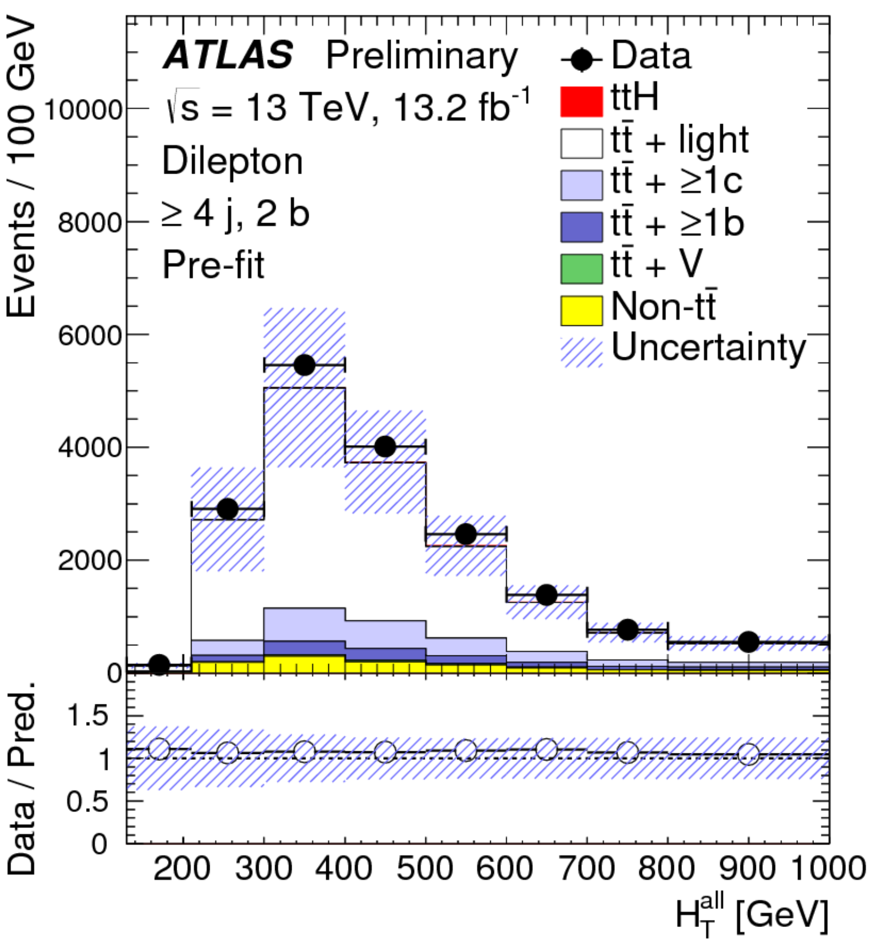


Analysis uses low-S/B categories to constrain backgrounds.

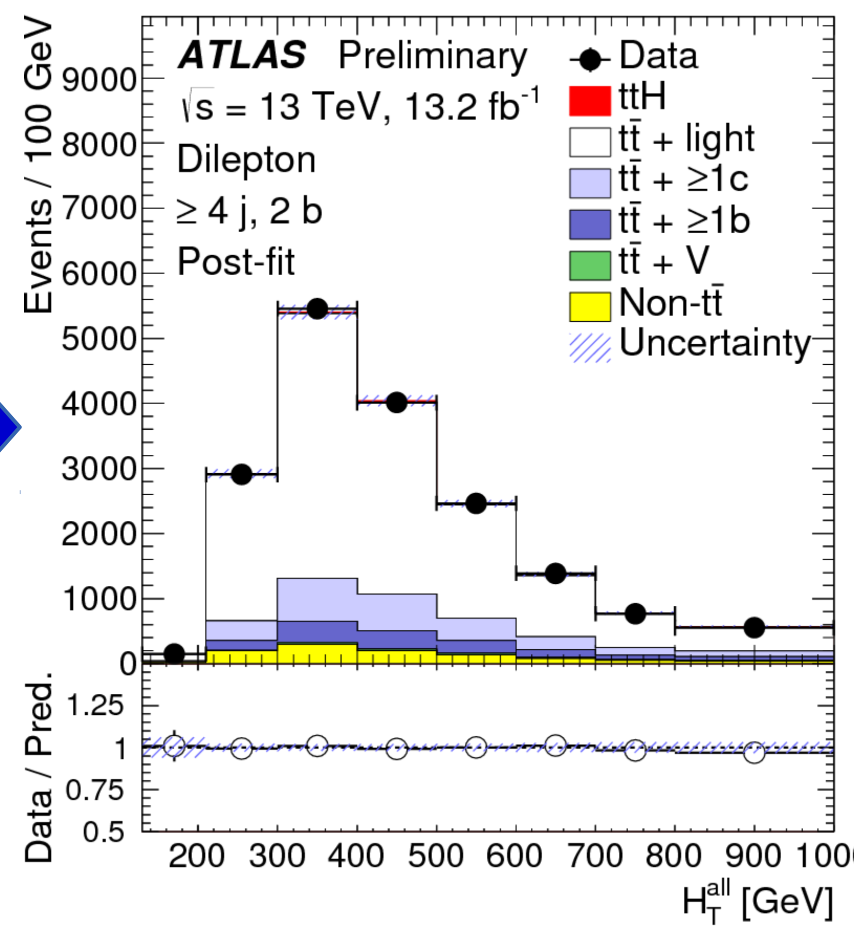
→ **Reduction in large uncertainties on tt bkg**

→ **Propagates to the high-S/B categories** through the statistical modeling

⇒ **Care needed in the propagation** (e.g. different kinematic regimes)



Fit



ATLAS-CONF-2016-08

Profiling Takeaways

When testing a hypothesis, use the best-fit values of the nuisance parameters: ***Profile Likelihood Ratio***.

$$\frac{L(\mu=\mu_0, \hat{\hat{\theta}}_{\mu_0})}{L(\hat{\mu}, \hat{\theta})}$$

Allows to include systematics as uncertainties on nuisance parameters.

Profiling systematics includes their effect into the total uncertainty.

Gaussian:

$$\sigma_{total} = \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$

Guaranteed to work well as long as everything is Gaussian, but typically also robust against non-Gaussian behavior.

Profiling can have unintended effects – need to carefully check behavior

Outline

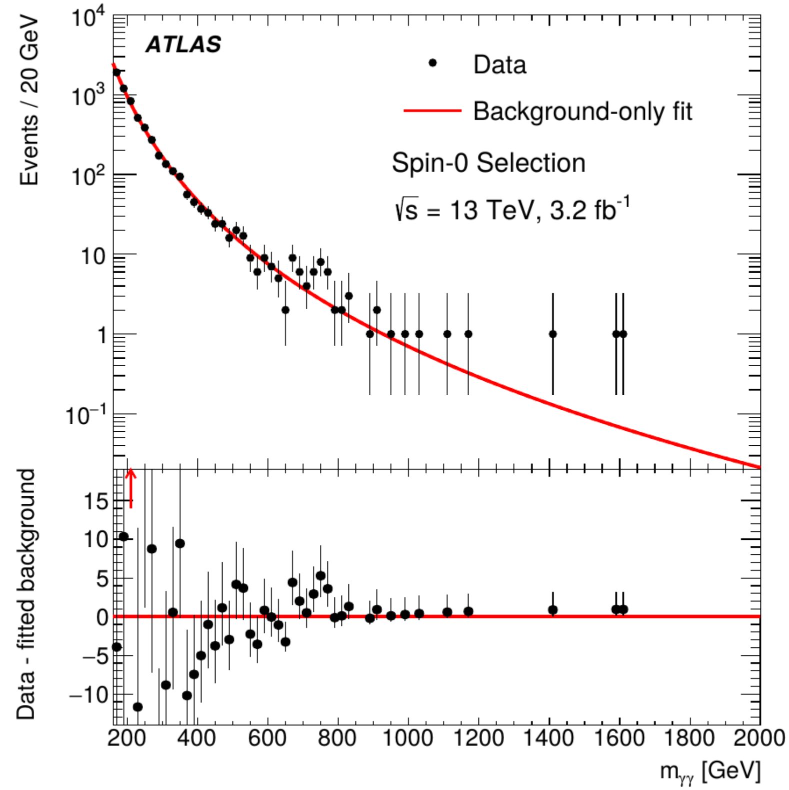
- Expected results and toys
  - Pseudo-experiments and Asimov datasets
  - Dealing with non-asymptotic situations
- Profiling
- **Look-Elsewhere Effect**
- Bayesian methods
- Presentation of results

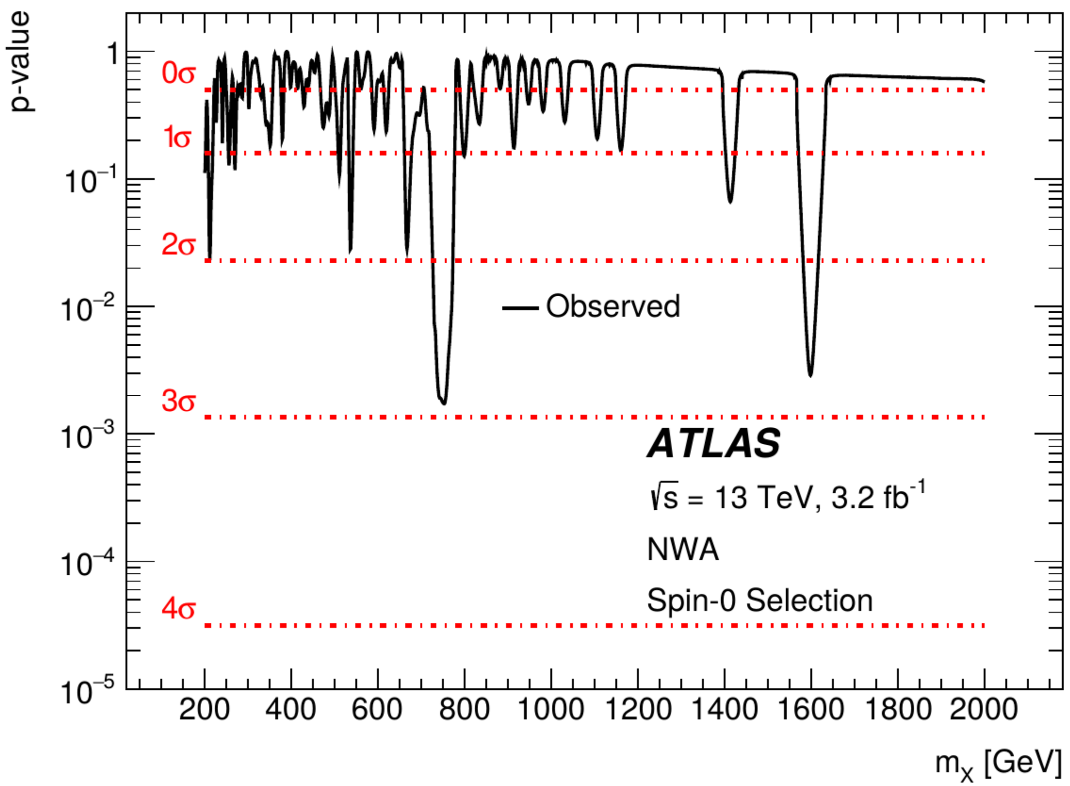
Look-Elsewhere effect

Sometimes, unknown parameters in signal model
e.g. p-values as a function of m_X

⇒ Effectively: **multiple, simultaneous searches**

→ If e.g. small resolution and large scan range, **many independent experiments**





→ More likely to find an excess **anywhere in the range**, rather than in a **predefined** location

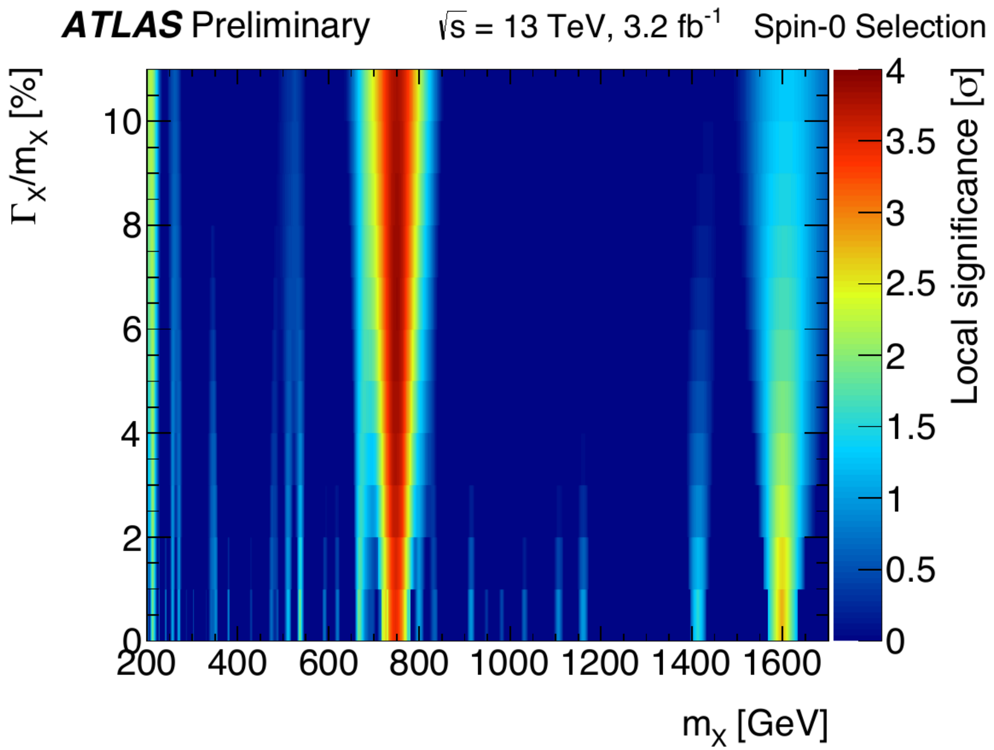
⇒ ***Look-elsewhere effect*** (LEE)

Global Significance

Probability for a fluctuation ***anywhere*** in the range → **Global** p-value.
at a given location → **Local** p-value

For searches over a parameter range, **the global p-value is the relevant one**

→ Accounts for the actual search procedure: look for an excess anywhere in the scanned range



→ Depends on the scanned parameter ranges

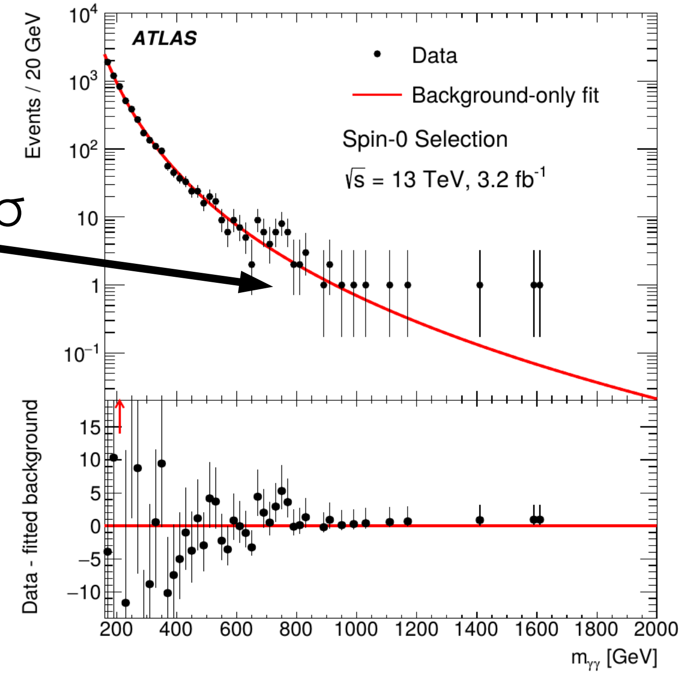
e.g. $X \to \gamma\gamma$:

- $200 < m_X < 2000$ GeV
- $0 < \Gamma_X < 10\%\ m_X$.

→ p_{local} is what comes out of the usual formulas

How to compute p_{global} (or N_{trials}) ?

Global Significance from Toys



Local 3.9σ

Principle: repeat the analysis in toy data:

→ generate pseudo-dataset

→ perform the search, scanning over parameters as in the data

→ report the largest significance found

→ repeat many times

⇒ The frequency at which a given Z_0 is found ***is*** the global p-value

e.g. **X→γγ Search**: **Z_{local} = 3.9σ** (⇒ p_{local} ~ 5 10^{-5}),

→ However we are scanning 200 < m_X < 2000 GeV and 0 < Γ_X < 10% m_X !

→ Toys : find such an excess **2%** of the time somewhere in the range

⇒ p_{global} ~ 2 10^{-2}, **Z_{global} = 2.1σ** Less exciting, and better indication of true Z!

⊕ **Exact treatment**

⊖ **CPU-intensive** especially for large Z (need ~O(100)/p_{global} toys)

Trials Factor

Trials factor N = # of independent searches:

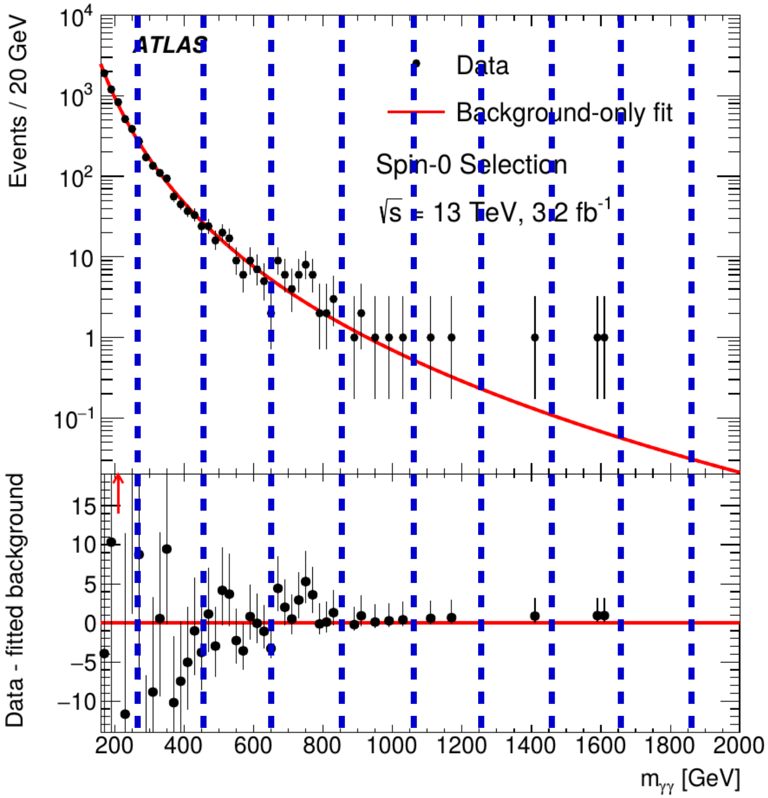
Global p-value → $p_{global} = 1 - \left(1 - p_{local}\right)^N \approx N\ p_{local}$ ← **Local** p-value

↑ **Trials factor**

Naively, one could expect

$$N_{\text{trials}} \overset{??}{=} N_{\text{indep}} = \frac{\text{scan range}}{\text{peak width}}$$

However this is usually **wrong** !



Trials Factor from Asymptotics

Gross & Vitells, EPJ.C70:525-530,2010

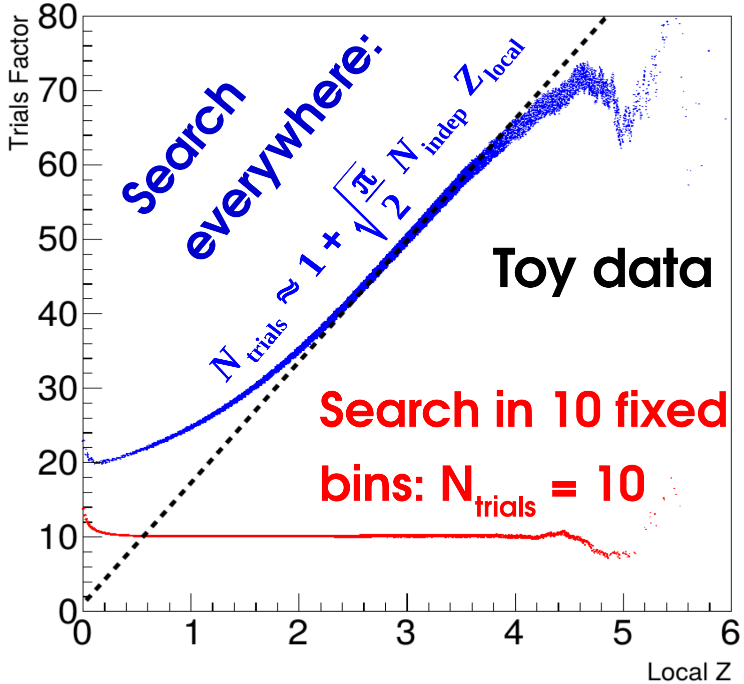
Asymptotic limit : trials factor (1 POI) is $$N_{trials} = 1 + \sqrt{\frac{\pi}{2}}\, N_{indep}\, Z_{local}$$

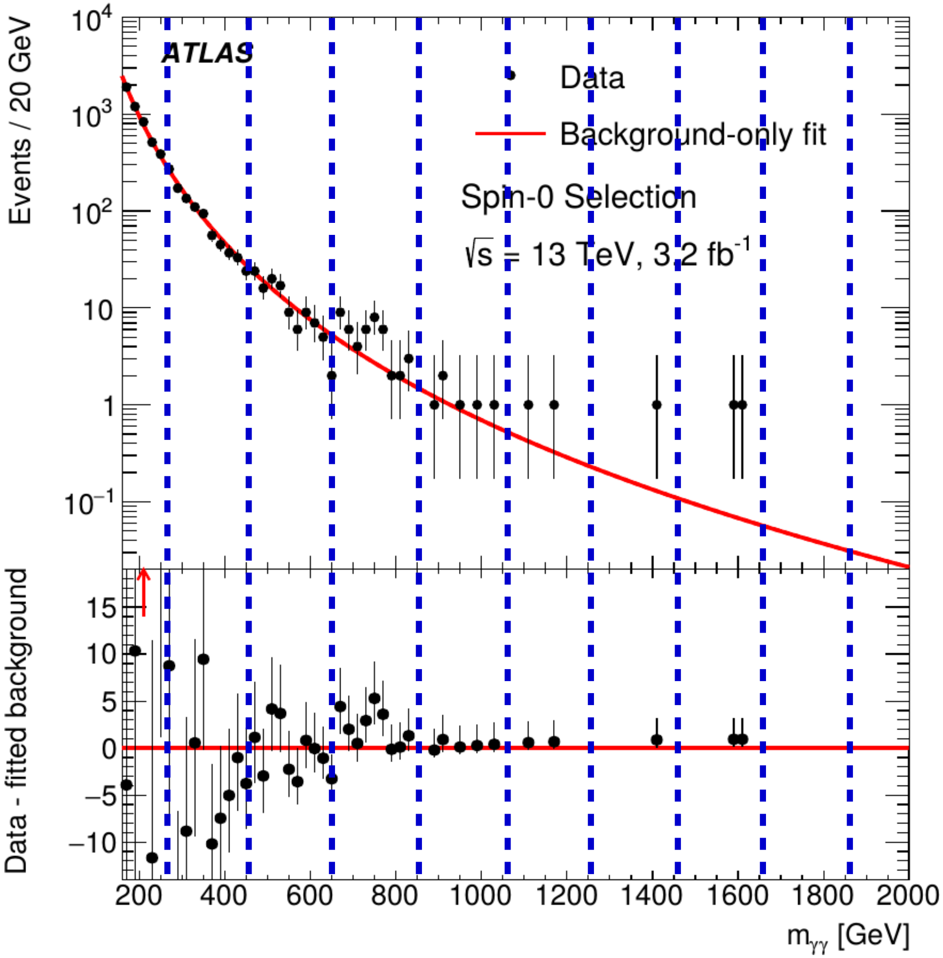
$$N_{indep} = \frac{\text{scan range}}{\text{peak width}}$$

→ Trials factor is **not just N_{indep}**, also depends on Z_{local} !

Why ? Slicing range into N_{indep} regions misses peaks sitting on **edges between regions**

⇒ true N_{trials} is **> N_{indep}**!





Trials Factor from Asymptotics

Gross & Vitells, EPJ.C70:525-530,2010

Asymptotic limit : trials factor (1 POI) is

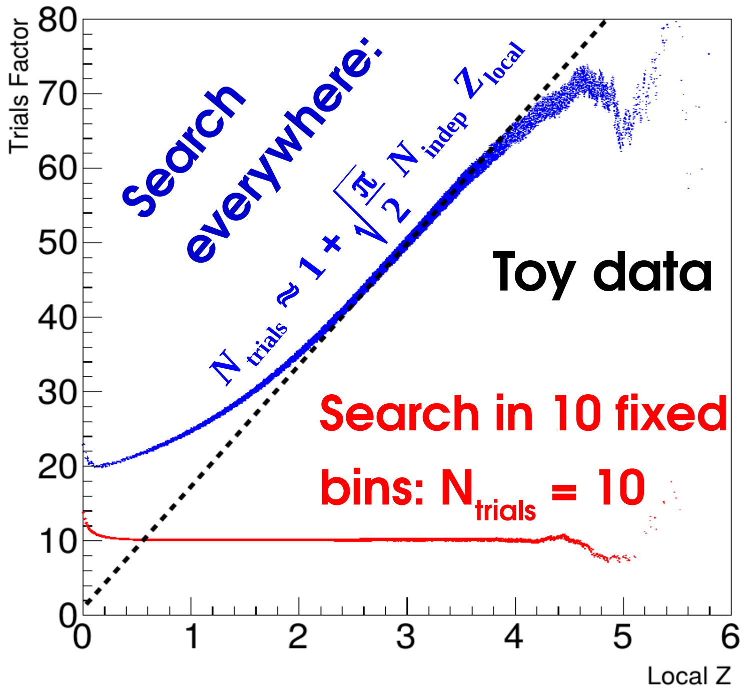
$$N_{\mathbf{trials}} = 1 + \sqrt{\frac{\pi}{2}}\, N_{\mathbf{indep}}\, Z_{\mathbf{local}}$$

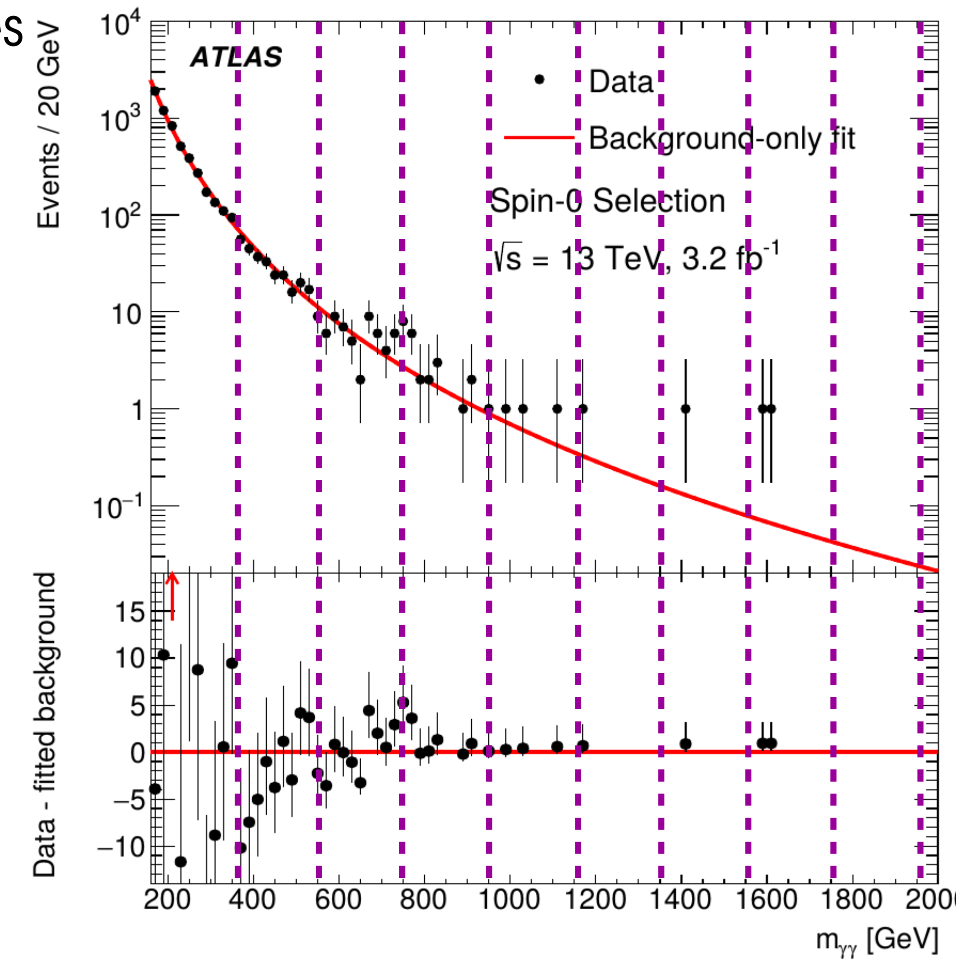
$$N_{\mathbf{indep}} = \frac{\textbf{scan range}}{\textbf{peak width}}$$

→ Trials factor is **not just N_{indep}**, also depends on Z_{local} !

Why ? Slicing range into N_{indep} regions misses peaks sitting on **edges between regions**

⇒ true N_{trials} is **> N_{indep}**!

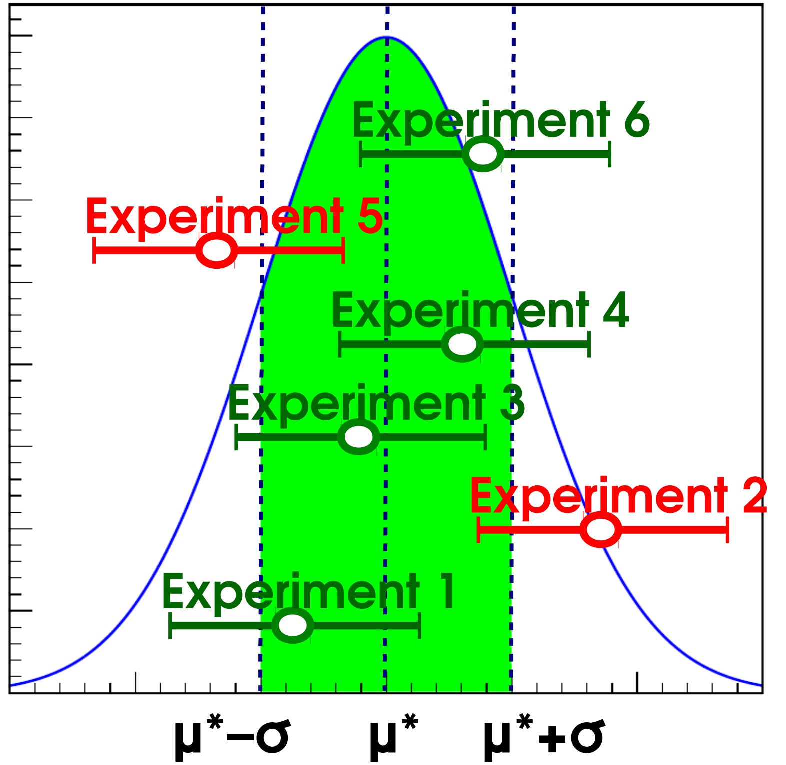




Outline

- Expected results and toys
  - Pseudo-experiments and Asimov datasets
  - Dealing with non-asymptotic situations
- Profiling
- Look-Elsewhere Effect
- **Bayesian methods**
- Presentation of results

Frequentist vs. Bayesian



All methods described so far are **frequentist**

- Measurement outcomes are random
- Parameters value are **fixed but unknown**

Must be careful about meaning:

→ **"5σ Higgs discovery"**

- → if there is really no Higgs, such fluctuations are observed in only one in 3 million experiments : **P(data | no Higgs) is small**

This is not the crucial question! What we would really like to know is

What is the probability that the excess we see is a fluctuation

→ we want **P(no Higgs |data)** – but all we have is **P(data | no Higgs)**

However **P(no Higgs |data)** is not well-defined in the frequentist framework

Frequentist vs. Bayesian

Can use **Bayes' theorem** to address this:

$$P(\text{no Higgs}|\text{data}) = \frac{P(\text{data}|\text{no Higgs})}{P(\text{data})} P(\text{no Higgs})$$

same as in the frequentist formalism (=likelihood)

Prior Probability

irrelevant normalization factor

Can compute P(no Higgs | data), **if we provide P(no Higgs)**

→ An hypothesis ("no Higgs") is now considered something random

- Is the presence of the Higgs in a experiment randomly chosen ?
- In fact, different definition of p: ***degree of belief***, not from frequencies.
- P(no Higgs) ***Prior degree of belief*** – critical ingredient in the computation

Compared to frequentist PLR:

⊕ **answers the "right" question**

⊖ **answer depends on the prior**

⊕ **In practice, frequentist and Bayesian methods usually give similar results**

"Bayesians address the questions everyone is interested in by using assumptions that no one believes. Frequentist use impeccable logic to deal with an issue that is of no interest to anyone." - Louis Lyons

Bayesian methods

Probability distribution (= likelihood) :

→ Same as frequentist case, but treat systematics by **integrating over priors**, instead of profiling:

→ Integrate out θ to get P(μ) : $P(\mu) = \int P(\mu, \theta)\, C(\theta)\, d\theta$

→ Use probability distribution P(μ) directly for limits & intervals

e.g. define 68% CL ("Credibility Level") interval (A, B) by: $\int_A^B P(\mu)\, d\mu = 68\,\%$

⊖ No simple way to test for discovery

⊖ Integration over NPs can be CPU-intensive (but can use MCMC methods)

Priors : most analyses use flat priors in the analysis variable(s)

⇒ **Parameterization-dependent**: if flat in σ×B , them not flat in couplings….

→ Can use the Jeffreys' or reference priors, but difficult in practice

Homework 8: Bayesian methods and CL_s

Gaussian counting problem with systematic on background: $n = S + B + \sigma_{syst}\theta$

$$P(n; S, \theta) = G(n; S+B+\sigma_{syst}\theta, \sigma_{stat})\, G(\theta_{obs}=0; \theta, 1)$$

→ **What is the 95% CL upper limit on S, given a measurement n_{obs} ?**

1. CLs computation:

- Use the result of Homework 7 to compute the PLR for S
- Use the result of Homework 6 to compute the CLs upper limit

2. Bayesian computation:

- Integrate P(n; S, θ) over θ to get the marginalized P(n | S)
- Use Bayes' theorem to compute P(S | n) ∝ P(n | S) P(S), with P(S) a constant prior over S>0.
- Find the 95% CL limit by solving $\int_{S_{up}}^{\infty} P(S|n)\, dS = 5\,\%$

Solution:

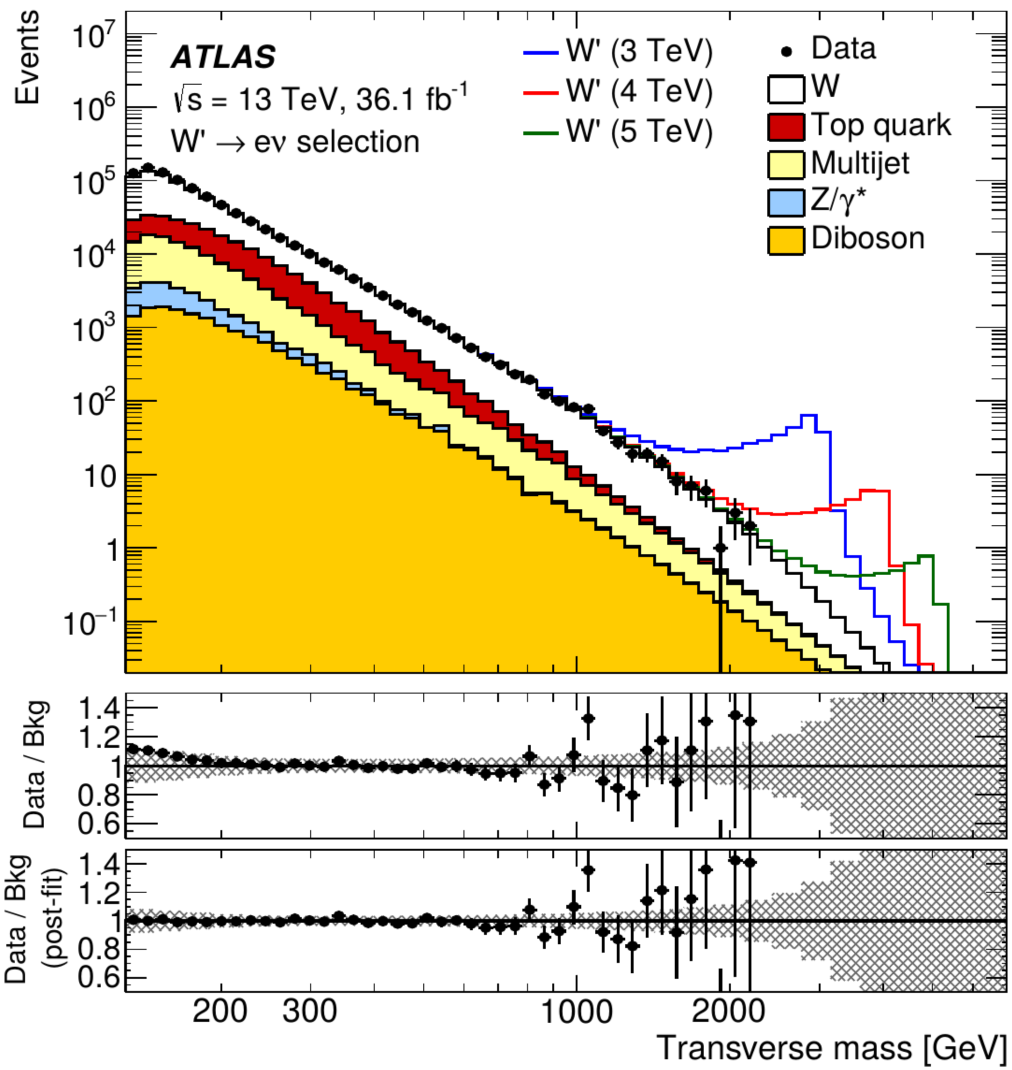
In both cases

$$S_{up}^{CL_s} = n - B + \left[\Phi^{-1}\left(1 - 0.05\, \Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)\right)\right]\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$

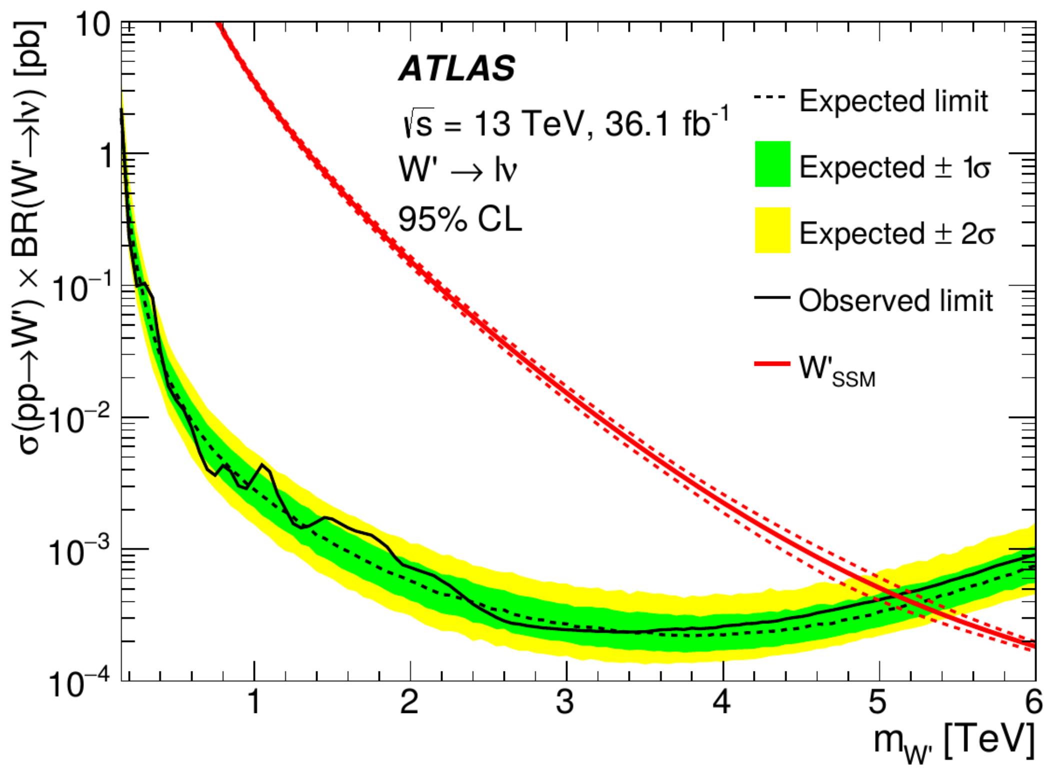
Example: W’→lν Search

arXiv:1706.04786

- **POI**: **W’ σ×B** → use **flat prior over** [0, +∞[.
- **NPs**: syst on **signal ε** (6 NPs), **bkg** (6), **lumi** (1) → integrate over Gaussian priors



Trigger	Multijet background
Lepton reconstruction and identification	Top extrapolation
Lepton momentum scale and resolution	Diboson extrapolation
E_T^{miss} resolution and scale	PDF choice for DY
Jet energy resolution	PDF variation for DY
Pile-up	EW corrections for DY
	Luminosity



Why 5σ ?

One-sided discovery: **5σ ⇔ p_0 = 3 10^{-7} ⇔ 1 chance in 3.5M**

→ Overly conservative ?

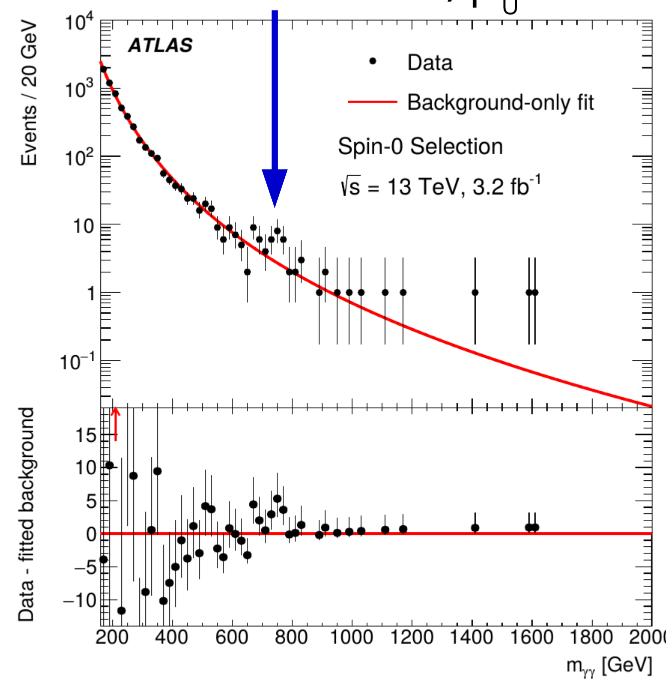
→ Do we even control such small probabilities ?

Reasons for sticking with 5σ (from Louis Lyons):

- **LEE** : searches typically cover multiple independent regions
  ⇒ Global p-value is the relevant one
  N_{trials} ~ 1000 : local 5σ ⇔ O(10^{-4}) more reasonable
- **Mismodeled systematics**: factor 2 error in syst-dominated analysis ⇒ factor 2 error on Z…
- **History**: 3σ and 4σ excesses do occur regularly, for the reasons above
- **"Subconscious Bayes Factor"** : p-value should be at least as small as the subjective p(S):

Local 3.9σ, p_0 = 5E-5

Global 2.1σ, p_0 = 2E-2



Extraordinary claims require extraodinary evidence

⇒ Stay with 5σ...

Outline

- Expected results and toys
  - Pseudo-experiments and Asimov datasets
  - Dealing with non-asymptotic situations
- Profiling
- Look-Elsewhere Effect
- Bayesian methods
- **Presentation of results**

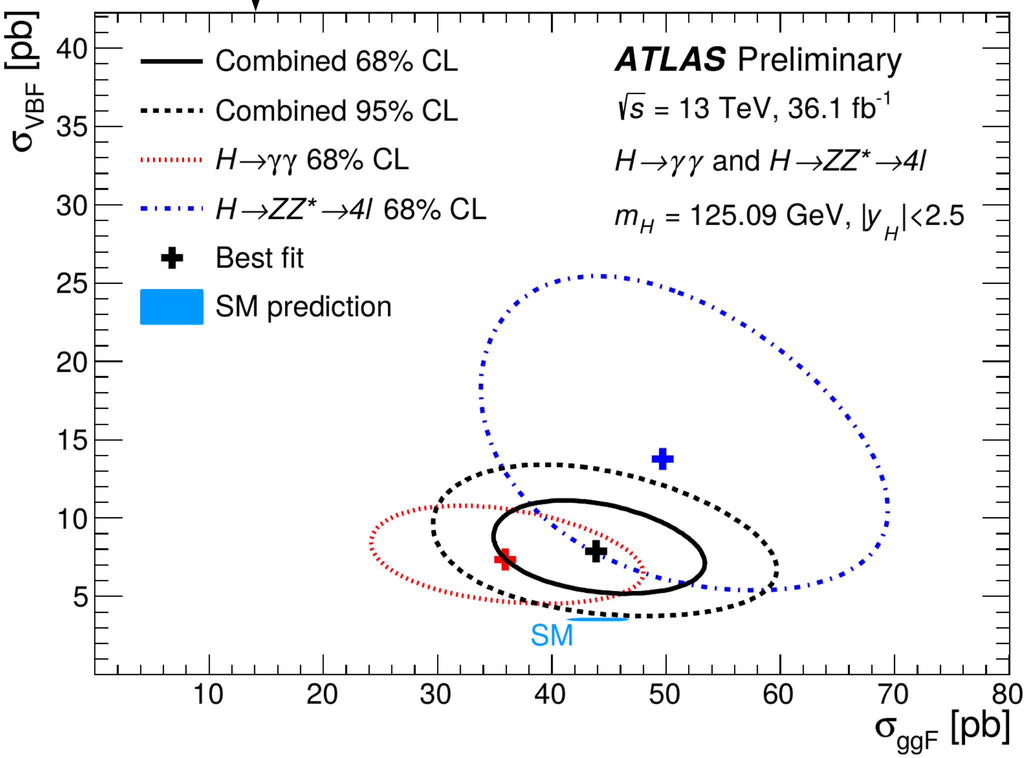
Reparameterization

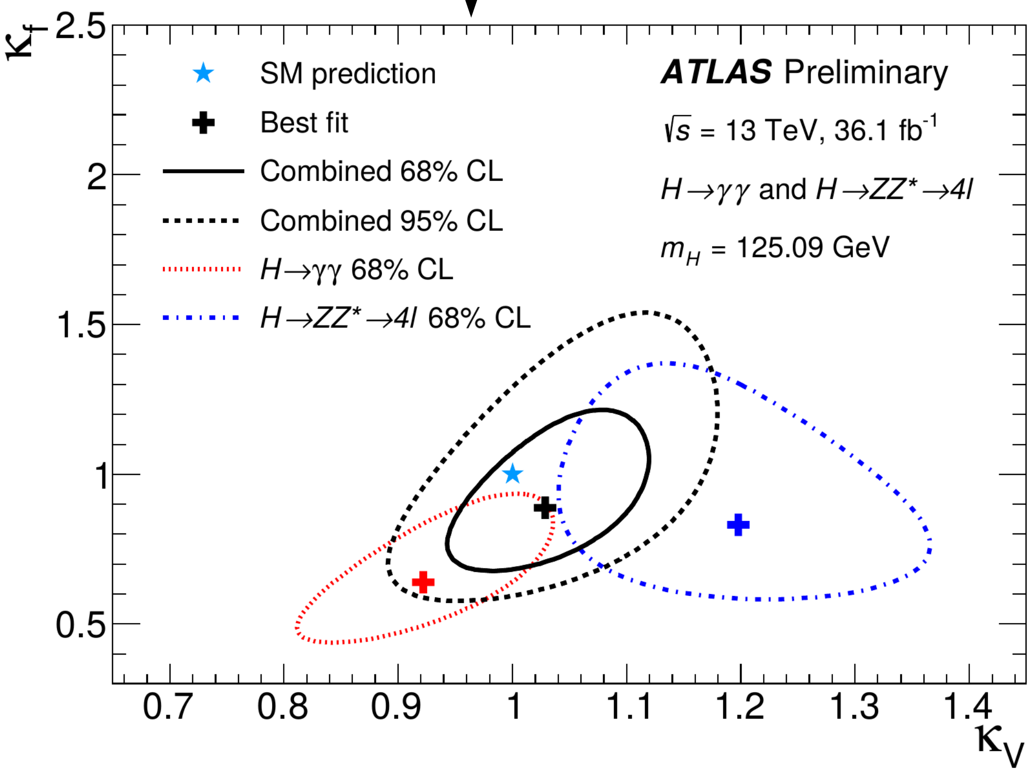
Start with basic measurement in terms of e.g. **σ×B**

→ How to measure derived quantities (couplings, parameters in some theory model, etc.) ? **→ just reparameterize the likelihood:**

e.g. Higgs couplings: σ_{ggF}, σ_{VBF} sensitive to Higgs coupling modifiers κ_V, κ_F.

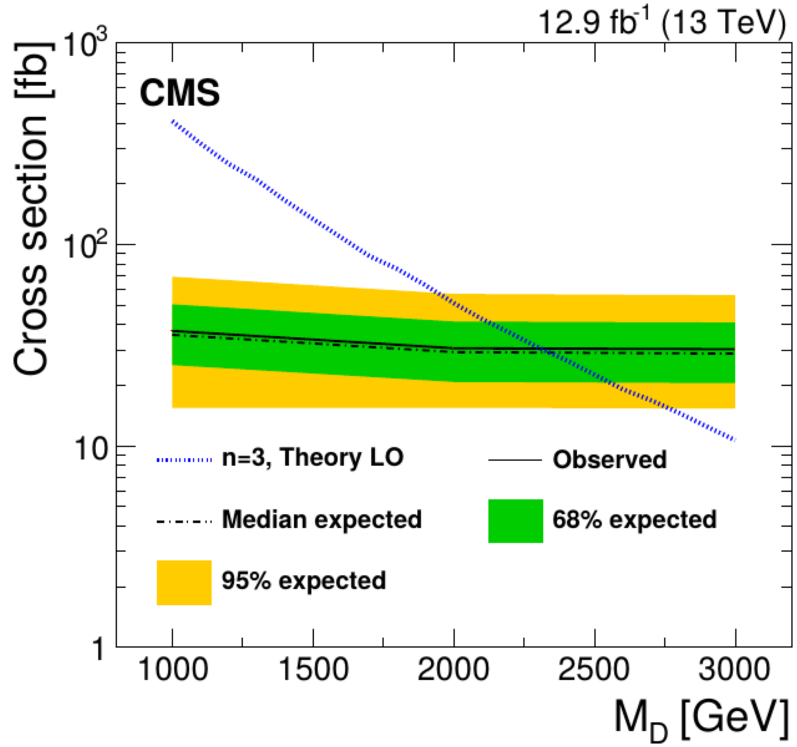
$$L(\sigma_{ggF}, \sigma_{VBF}) \xrightarrow[\sigma_{VBF} \to \sigma_{VBF}(\kappa_V, \kappa_F)]{\sigma_{ggF} \to \sigma_{ggF}(\kappa_V, \kappa_F)} L(\sigma_{ggF}(\kappa_V, \kappa_F), \sigma_{VBF}(\kappa_V, \kappa_F)) \equiv L'(\kappa_V, \kappa_F)$$

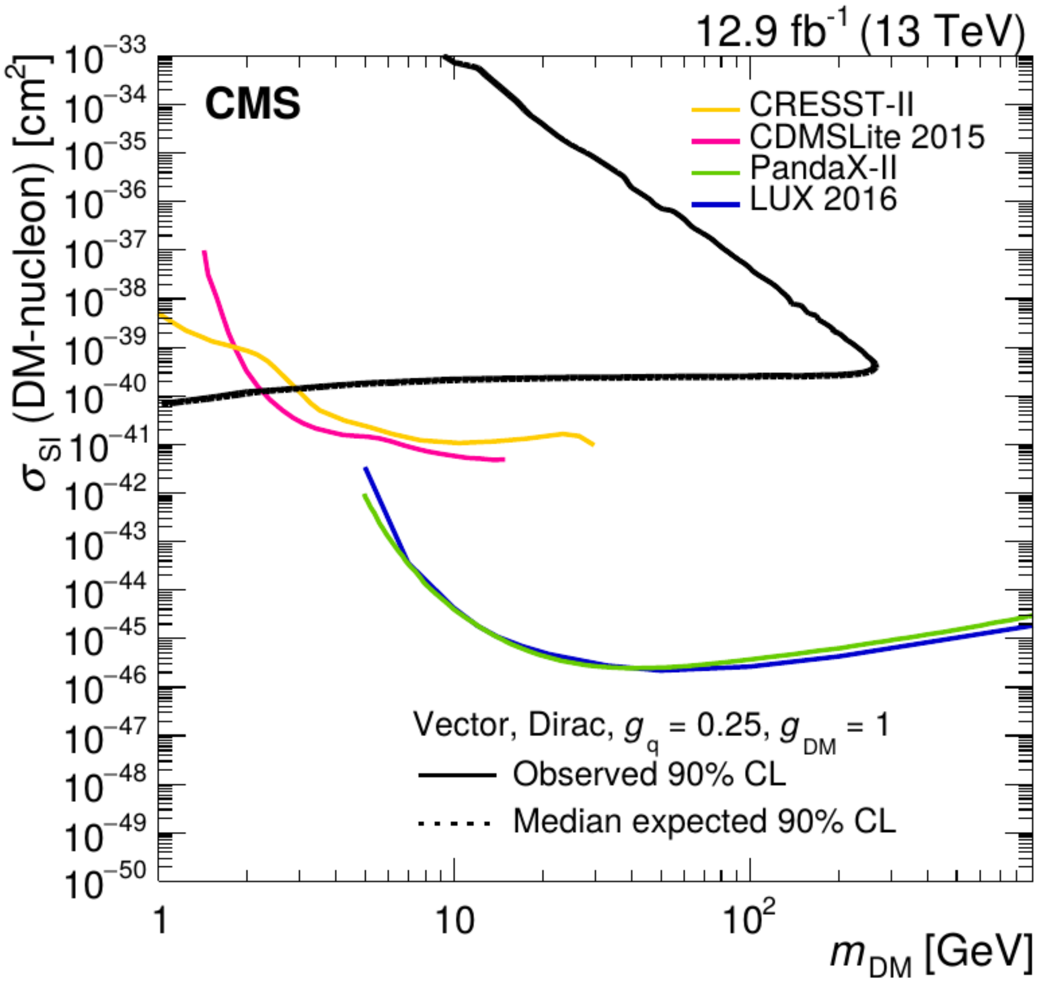


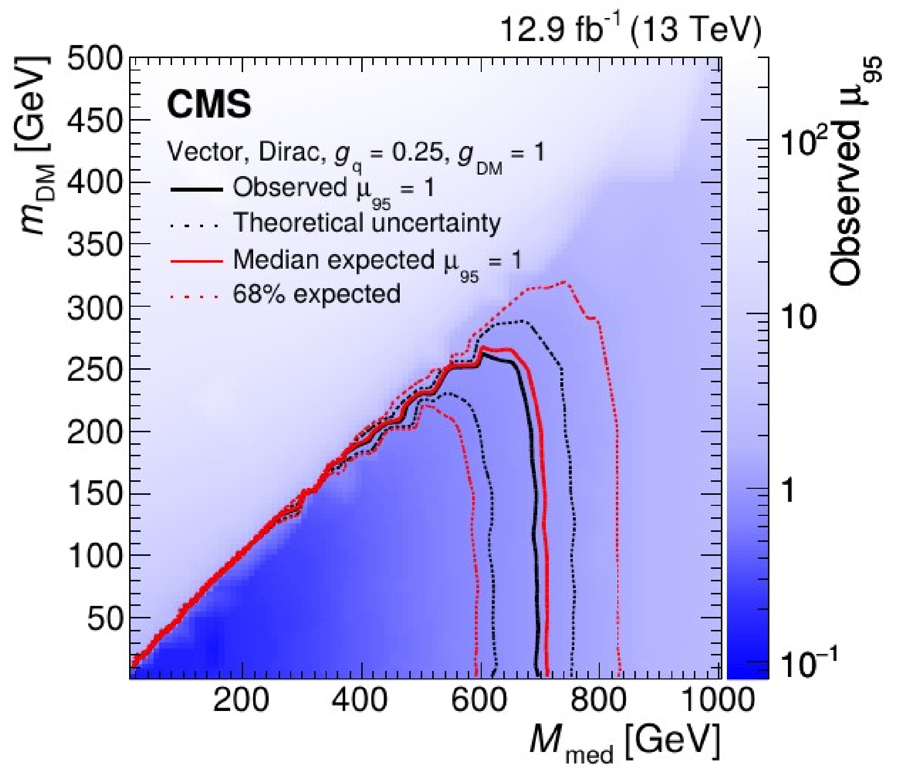


Reparameterization: Limits

CMS Run 2 Monophoton Search: measured N_s in a counting experiment reparameterized according to various DM models





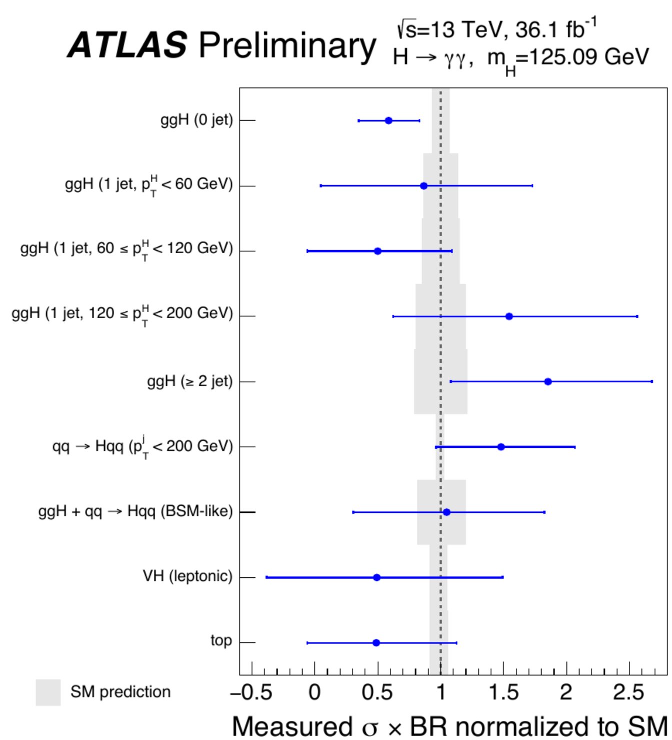


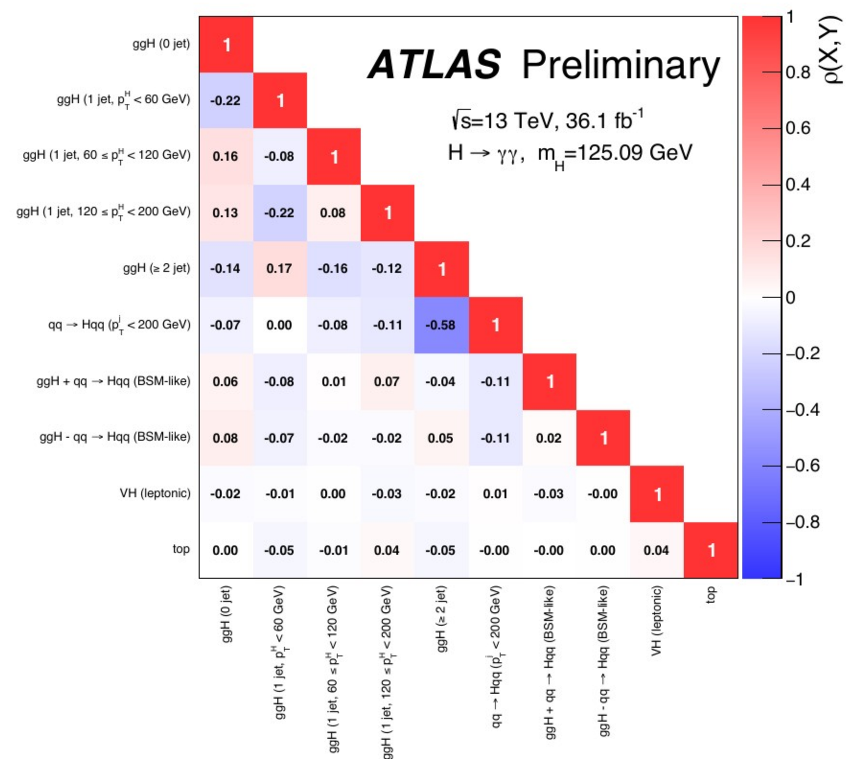
Presentation of Results

→ Cannot test every model : need to make enough information public so that others (theorists) are able to do it independently

⇒ **Gaussian case**: sufficient to provide measurements + covariance matrix

→ For example using the HEPData repository.





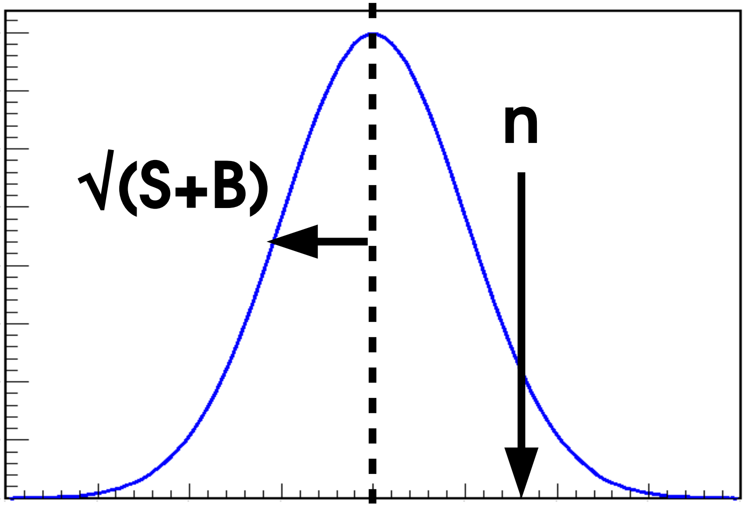
Non-Gaussian case: no simple method

Conclusion

- Significant evolution in the statistical methods used in HEP
- Variety of methods, adapted to various situations and target results
- Allow to
  - model the statistical process with high precision in difficult situations (large systematics, small signals)
  - make optimal use of available information
- Implemented in standard RooFit/RooStat toolkits within the ROOT framework, as well as other tools (BAT)
- Still many open questions and areas that could use improvement
  → e.g. how to present results with all available information

Homework solutions for Lecture 2

Homework 1: Gaussian Counting



Count number of events n in data

→ assume n large enough so process is Gaussian

→ assume B is known, measure S

Likelihood :

$$L(S;n) = e^{-\frac{1}{2}\left(\frac{n-(S+B)}{\sqrt{S+B}}\right)^2}$$

$$\lambda(S;n) = \left(\frac{n-(S+B)}{\sqrt{S+B}}\right)^2$$

MLE for S : $\hat{S} = n - B$

Test statistic: assume $\hat{S} > 0$,

$$q_0 = -2\log\frac{L(S=0)}{L(\hat{S})} = \lambda(S=0) - \lambda(\hat{S}) = \left(\frac{n-B}{\sqrt{B}}\right)^2 = \left(\frac{\hat{S}}{\sqrt{B}}\right)^2$$

Finally:

$$Z = \sqrt{q_0} = \frac{\hat{S}}{\sqrt{B}}$$

Known formula!

→ Strictly speaking only valid in Gaussian regime

Homework 2: Poisson Counting

Same problem but now ***not*** assuming Gaussian behavior:

$$L(S;n) = e^{-(S+B)}(S+B)^n \qquad \lambda(S;n) = 2(S+B) - 2n\log(S+B)$$

MLE: **$\hat{S} = n - B$**, same as Gaussian

Test statistic (for $\hat{S} > 0$):

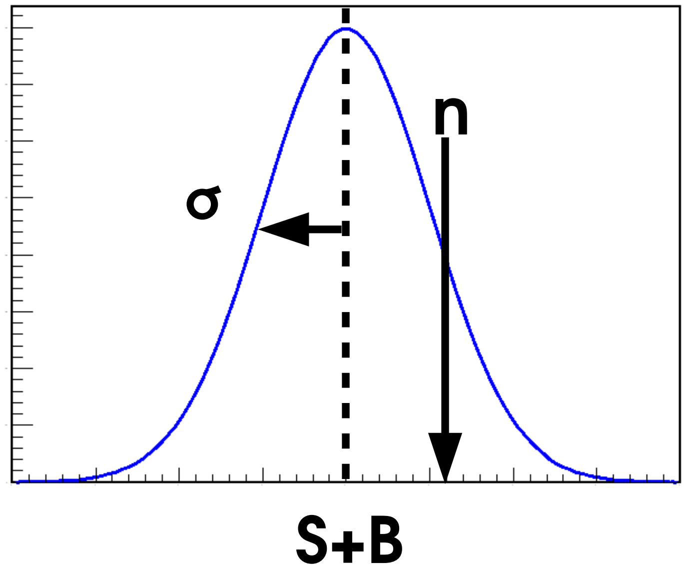
$$q_0 = \lambda(S=0) - \lambda(\hat{S}) = -2\hat{S} - 2(\hat{S}+B)\log\frac{B}{\hat{S}+B}$$

Assuming asymptotic distribution for q_0,

$$Z = \sqrt{2\left[(\hat{S}+B)\log\left(1+\frac{\hat{S}}{B}\right) - \hat{S}\right]}$$

See G. Cowan's slides for case with B uncertainty

Homework 3: Gaussian CL_{s+b}



Usual Gaussian counting example with known B:

$$\lambda(S) = \left(\frac{n-(S+B)}{\sigma_S}\right)^2$$

Reminder:

Best fit signal : $\hat{S} = n - B$

Significance: $Z = \hat{S}/\sqrt{B}$

Compute the 95% CL upper limit on S:

$$q_{S_0} = -2\log\frac{L(S=S_0)}{L(\hat{S})} = \lambda(S_0) - \lambda(\hat{S}) = \left(\frac{n-(S_0+B)}{\sigma_S}\right)^2 = \left(\frac{S_0-\hat{S}}{\sigma_S}\right)^2 \quad \text{for } S_0 > \hat{S}$$

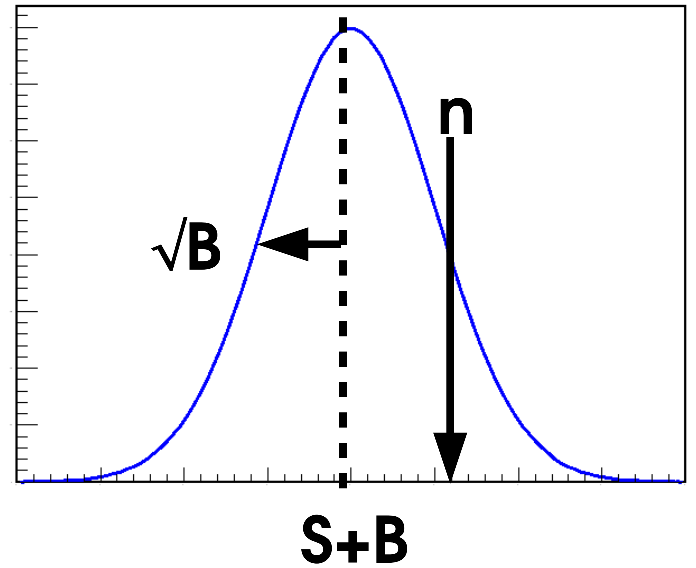
so $q_{S_0} = 2.70$ for $S_0 = \hat{S} + \sqrt{2.70}\,\sigma_S$

And finally $S_{up} = \hat{S} + 1.64\,\sigma_S$ at 95 % CL

Homework 4 : Gaussian CL_s

Usual Gaussian counting example with known B:

$$\lambda(S) = \left(\frac{n-(S+B)}{\sigma_S}\right)^2$$



Reminder

Best fit signal : $\hat{S}$ **= n - B**

CL_{s+b} limit:

$$S_{up} = \hat{S} + 1.64\,\sigma_S \text{ at } 95\,\% \text{ CL}$$

CL_s upper limit : still have $q_{S_0} = \left(\frac{S_0-\hat{S}}{\sigma_S}\right)^2$ (for $S_0 > \hat{S}$)

so need to solve

$$p_{CL_s} = \frac{p_{S_0}}{1-p_B} = \frac{1-\Phi(\sqrt{q_{S_0}})}{1-\Phi(\sqrt{q_{S_0}} - S_0/\sigma_S)} = 5\,\%$$

for $\hat{S} = 0$,

$$S_{up} = \hat{S} + \left[\Phi^{-1}\left(1 - 0.05\,\Phi\left(\hat{S}/\sigma_S\right)\right)\right]\sigma_S \text{ at } 95\,\% \text{ CL}$$

$\hat{S} \sim G(S, \sigma_S)$ so

Under $H_0(S = S_0)$:

$$\sqrt{q_{S_0}} \sim G(0,1)$$

$$p_{S_0} = 1-\Phi(\sqrt{q_{S_0}})$$

Under $H_0(S = 0)$:

$$\sqrt{q_{S_0}} \sim G(S_0/\sigma_S, 1)$$

$$p_B = \Phi(\sqrt{q_{S_0}} - S_0/\sigma_S)$$

Homework 5: Poisson CL_s

Same exercise, for the Poisson case

Exact computation : sum probabilities of cases "at least as extreme as data" (n)

$$p_{S_0}(n) = \sum_0^n e^{-(S_0+B)} \frac{(S_0+B)^k}{k!}$$ and one should solve $p_{CL_s} = \frac{p_{S_{up}}(n)}{p_0(n)} = 5\%$ for S_{up}

For n = 0: $$p_{CL_s} = \frac{p_{S_{up}}(0)}{p_0(0)} = e^{-S_{up}} = 5\% \Rightarrow S_{up} = \log(20) = 2.996 \approx 3$$

⇒ Rule of thumb: when n_{obs}=0, the 95% CL_s limit is 3 events (for any B)

Asymptotics: as before, $$q_{S_0} = \lambda(S_0) - \lambda(\hat{S}) = 2(S_0 + B - n) - 2n \log \frac{S_0+B}{n}$$

For n = 0, $$q_{S_0}(n=0) = 2(S_0+B)$$

$$p_{CL_s} = \frac{p_{S_0}}{p_0} = \frac{1-\Phi\left(\sqrt{q_{S_0}(n=0)}\right)}{1-\Phi\left(\sqrt{q_{S_0}(n=0)} - \sqrt{q_{S_0}(n=B)}\right)} = 5\%$$

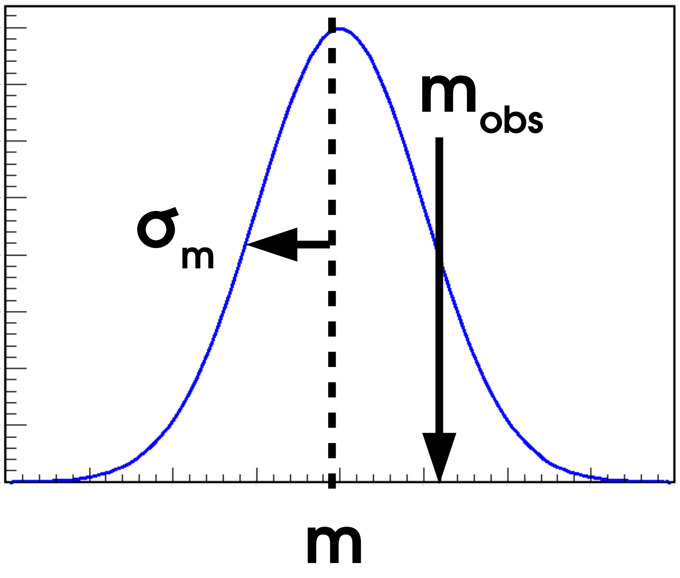
⇒ S_{up} ~ 2, exact value depends on B

⇒ Asymptotics not valid in this case (n=0) – need to use exact results, or toys

Homework 6: Gaussian Intervals

Consider a parameter m (e.g. Higgs boson mass) whose measurement is Gaussian with known width σ_m, and we measure m_{obs}:

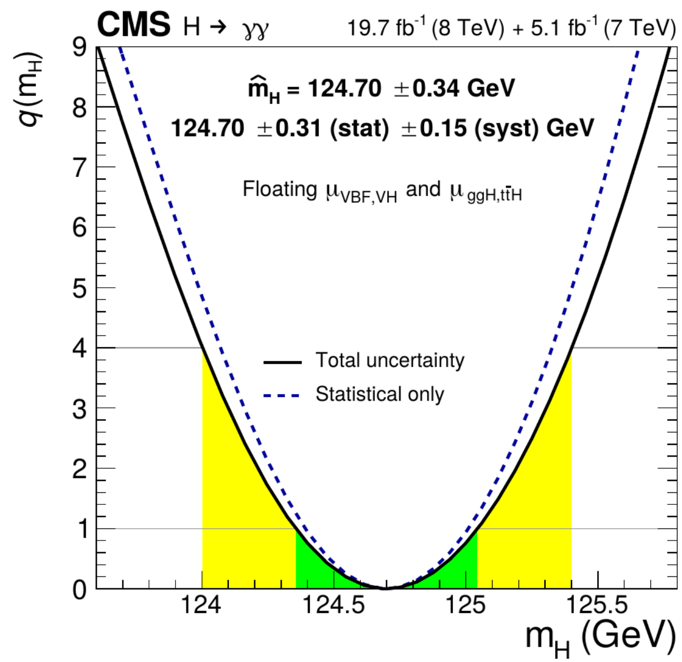
$$\lambda(m; m_{obs}) = \left(\frac{m - m_{obs}}{\sigma_m}\right)^2$$



→ Best-fit value (MLE): $\hat{m} = m_{obs}$.

→ Test statistic : $t_m = \left(\frac{m - m_{obs}}{\sigma_m}\right)^2$

→ 1σ Interval $m = m_{obs} \pm \sigma_m$



JHEP 11 (2017) 047

Homework solutions for Lecture 3

Homework 7: Gaussian Profiling

Counting experiment with background uncertainty: **$n = S + \theta$** :

→ Signal region: **$n \sim G(S + \theta, \sigma_{stat})$**

→ Control region: **$\theta^{obs} \sim G(\theta, \sigma_{syst})$**

$$L(S,\theta) = G(n; S+\theta, \sigma_{stat})\, G(\theta^{obs}; \theta, \sigma_{syst})$$

Then: $$\lambda(S,\theta) = \left(\frac{n-(S+\theta)}{\sigma_{stat}}\right)^2 + \left(\frac{\theta^{obs}-\theta}{\sigma_{syst}}\right)^2$$

MLEs: $$\hat{S} = n - \theta^{obs}$$ $$\hat{\theta} = \theta^{obs}$$

Conditional MLE: $$\hat{\hat{\theta}}(S) = \theta^{obs} + \frac{\sigma_{syst}^2}{\sigma_{stat}^2 + \sigma_{syst}^2}(\hat{S} - S)$$

For S = Ŝ, matches MLE as it should

PLR: $$t_S = -2\log\frac{L(S,\hat{\hat{\theta}}(S))}{L(\hat{S},\hat{\theta})} = \lambda(S,\hat{\hat{\theta}}(S)) - \lambda(\hat{S},\hat{\theta}) = \frac{(S-\hat{S})^2}{\sigma_{stat}^2 + \sigma_{syst}^2}$$

1σ interval $$S = \hat{S} \pm \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2} \qquad \sigma_S = \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$

Stat uncertainty (on n) and systematic (on θ) add in quadrature

Homework 8: CL_s computation

Gaussian counting with systematic on background: $n = S + B + \sigma_{syst}\theta$

$$L(n; S, \theta) = G(n; S + B + \sigma_{syst}\theta, \sigma_{stat})\, G(\theta_{obs} = 0; \theta, 1)$$

MLE: $\hat{S} = n - B$

Conditional MLE: $\hat{\hat{\theta}}(\mu) = \frac{\sigma_{syst}}{\sigma_{stat}^2 + \sigma_{syst}^2}(n - S - B)$

PLR : $\lambda(\mu) = \left(\frac{S + B - n}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)^2$

This boils down to the Gaussian case of HW 6, so the CL_s limit is

$$CL_s: \quad S_{up}^{CL_s} = n - B + \left[\Phi^{-1}\left(1 - 0.05\, \Phi\left(\frac{n - B}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)\right)\right]\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$

Homework 8: Bayesian computation

Gaussian counting with systematic on background: $n = S + B + \sigma_{syst}\theta$

$$P(n \mid S, \theta) = G(n; S + B + \sigma_{syst}\theta, \sigma_{stat})\, G(\theta \mid 0, 1)$$

Bayesian: G(θ) is actually a ***prior*** on θ ⇒ perform integral (***marginalization***)

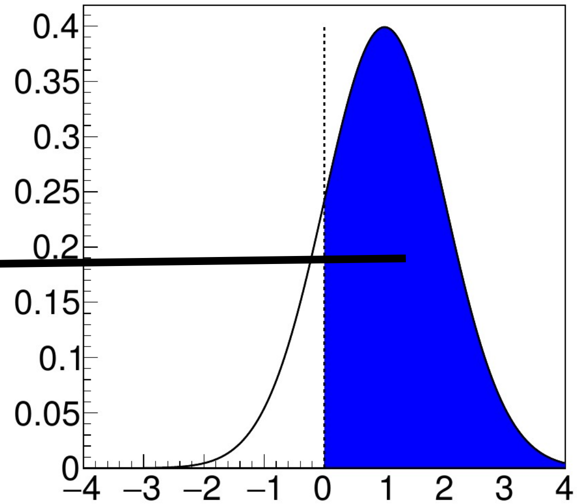
$$P(n \mid S) = G\left(S;\, n - B, \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}\right)$$

same effect as profiling!

Need P(S | n) ⇒ a prior for S – take flat PDF over S > 0

⇒ Truncate Gaussian at S=0: $P(S \mid n) = P(n \mid S)\, P(S)$

$$P(S \mid n) = G\left(S; n - B, \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}\right)\left[\Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)\right]^{-1}$$



Bayesian Limit:

$$\int_{S_{up}}^{\infty} P(S \mid n)\, dS = 5\,\% = \left[1 - \Phi\left(\frac{S_{up} - (n-B)}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)\right]\left[\Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)\right]^{-1}$$

$$S_{up}^{Bayes} = n - B + \left[\Phi^{-1}\left(1 - 0.05\, \Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}}\right)\right)\right]\sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$

same result as CL_s!

Extra Slides

Illustrative Example

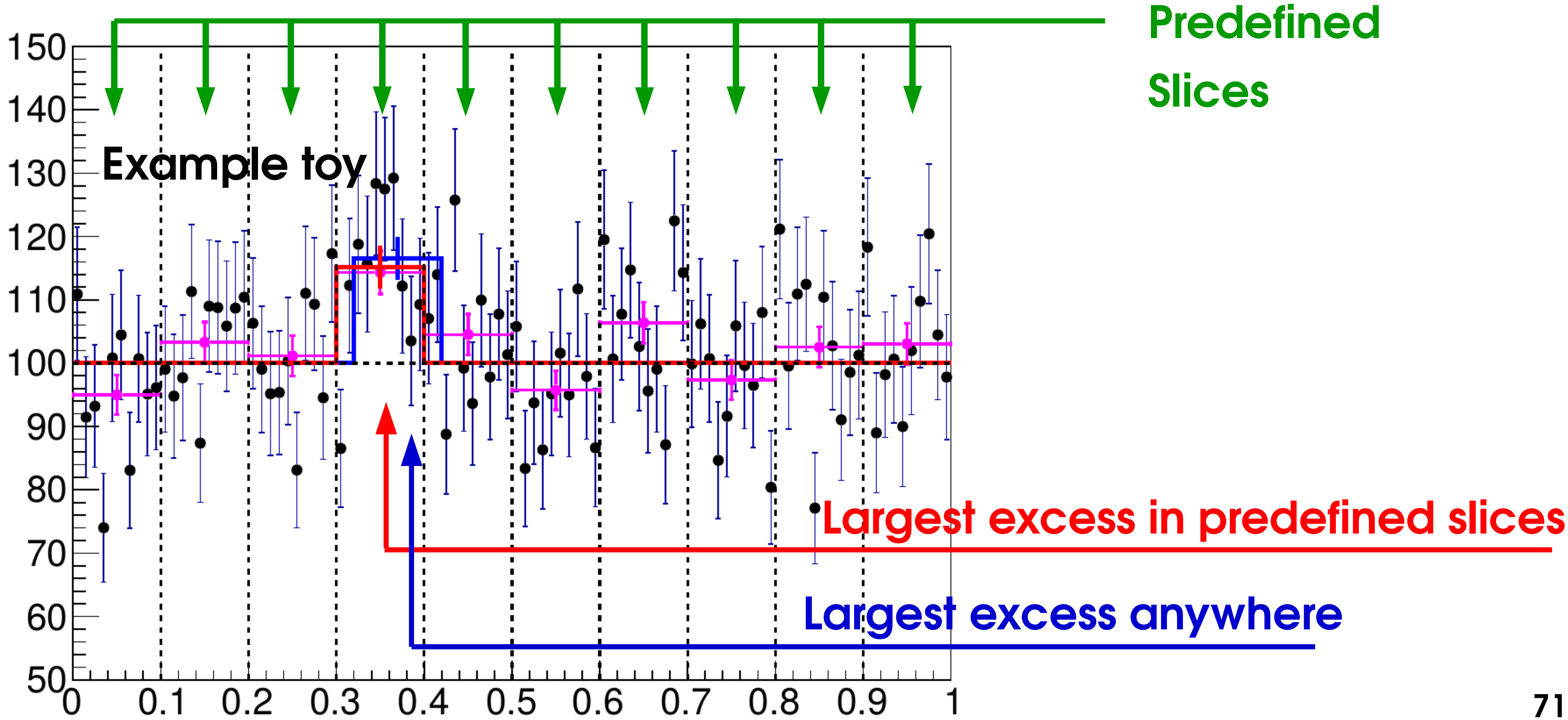
Test on a simple example: generate toys with

→ flat background (100 events/bin)

→ count events in a fixed-size sliding window, look for excesses

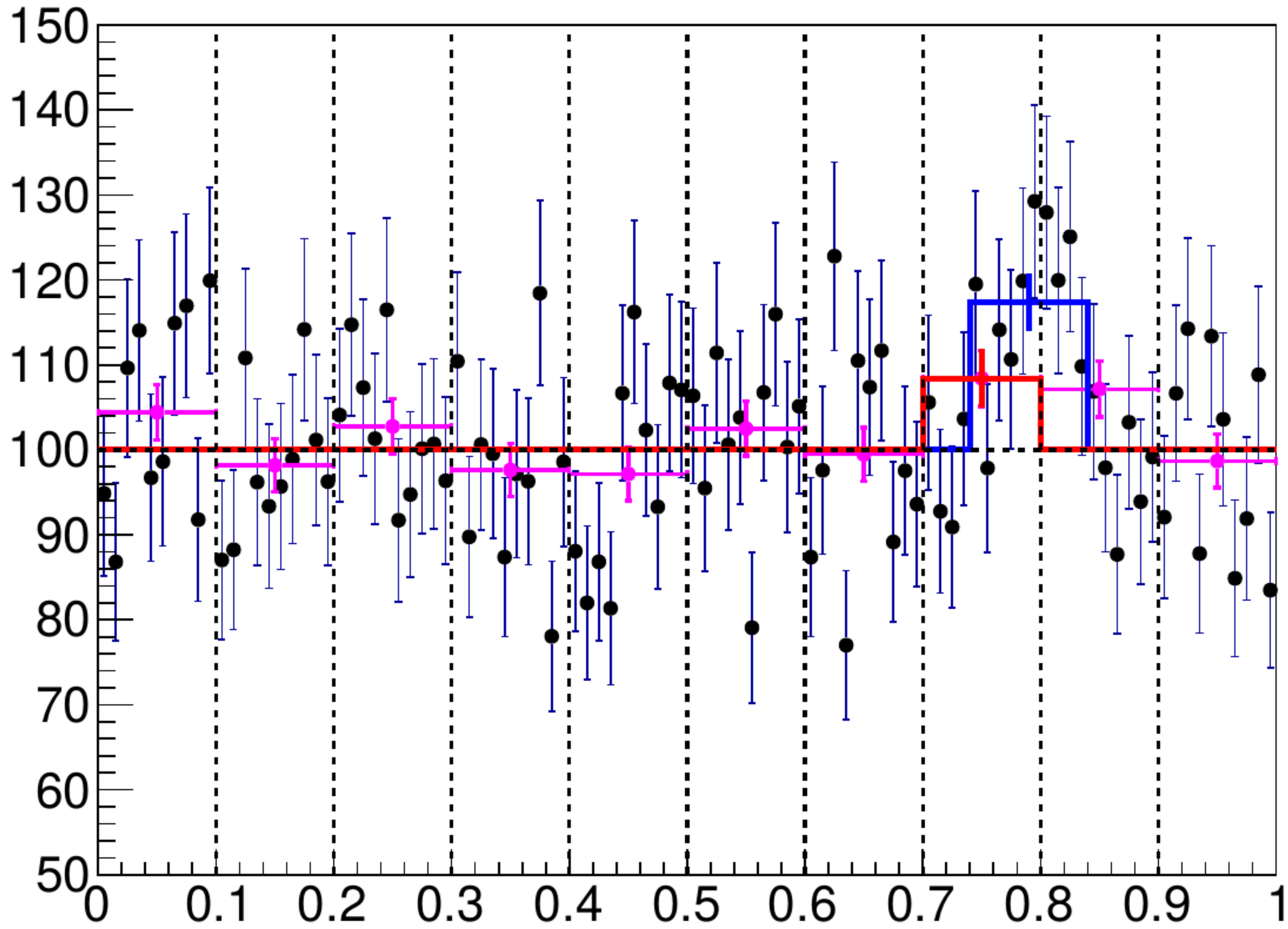
Two configurations:

1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum



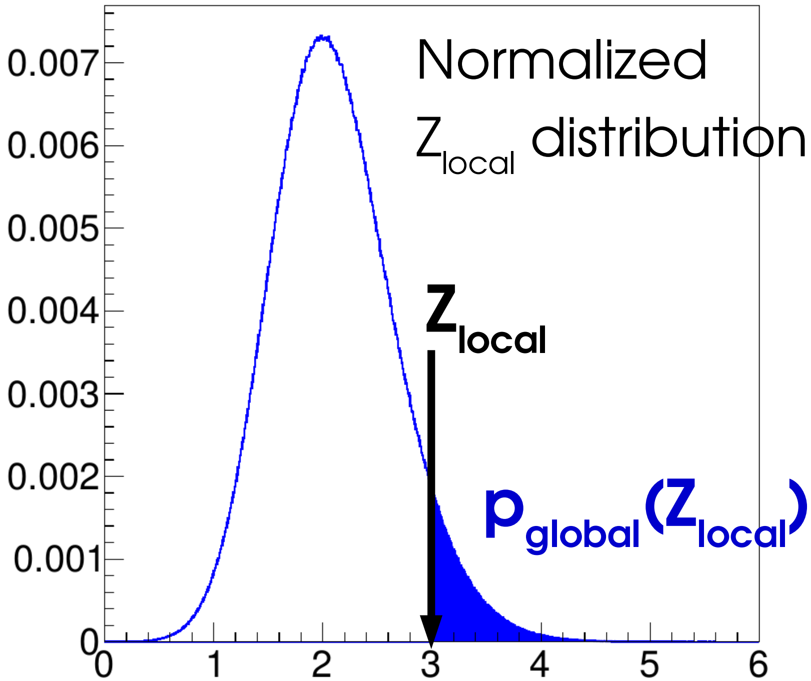
Illustrative Example (2)

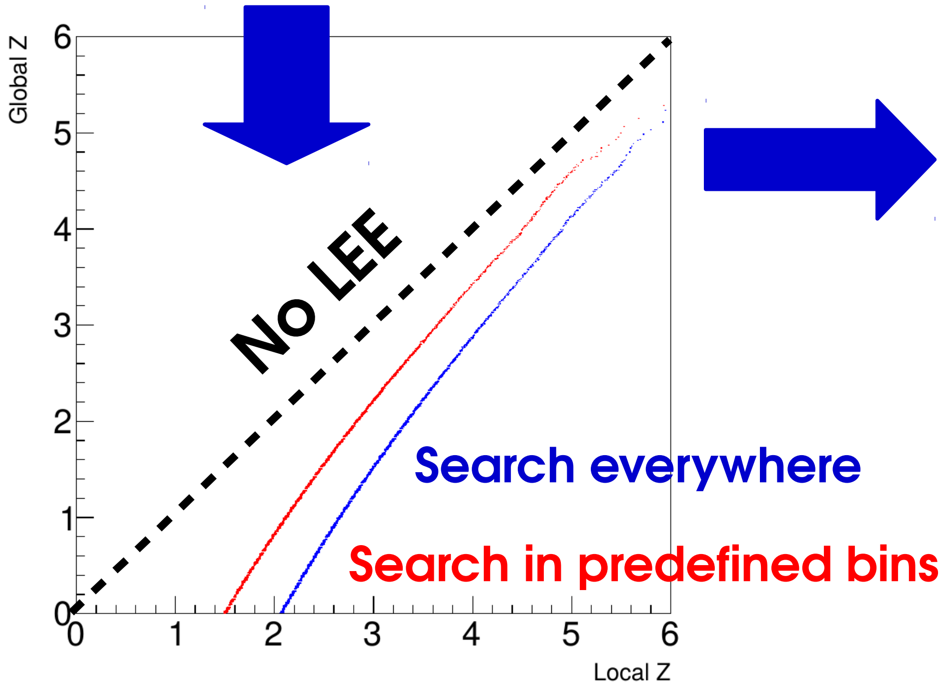
Very different results if the excess is **near a boundary :**

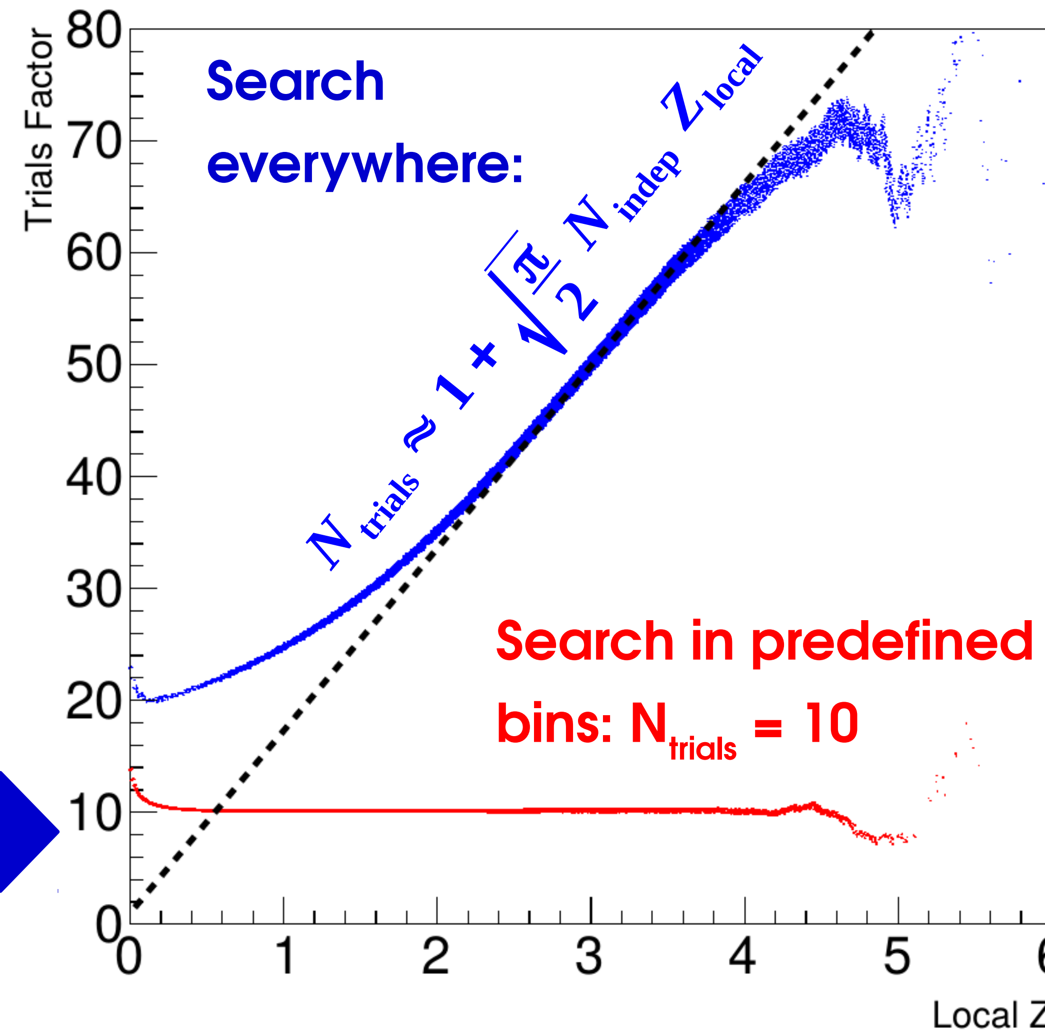


1. Look only in 10 slices of the full spectrum
2. Look in any window of same size as above, anywhere in the spectrum

Illustrative Example (3)







Searching everywhere gives the extra Z_{local} dependence

Z_{Global} Asymptotics Extrapolation

Asymptotic trials factor (1 POI):

$$N_{trials} = 1 + \sqrt{\frac{\pi}{2}}\, N_{indep}\, Z_{local}$$

How to get $\mathbf{N_{indep}}$? Usually work with a slightly different formula:

$$N_{trials} = 1 + \frac{1}{p_{local}} \langle N_{up}(Z_{test}) \rangle e^{\frac{Z_{test}^2 - Z_{local}^2}{2}}$$

Number of excesses with $Z > Z_{test}$

⇒ **calibrate for small Z_{test}, apply result to higher Z_{local}.**

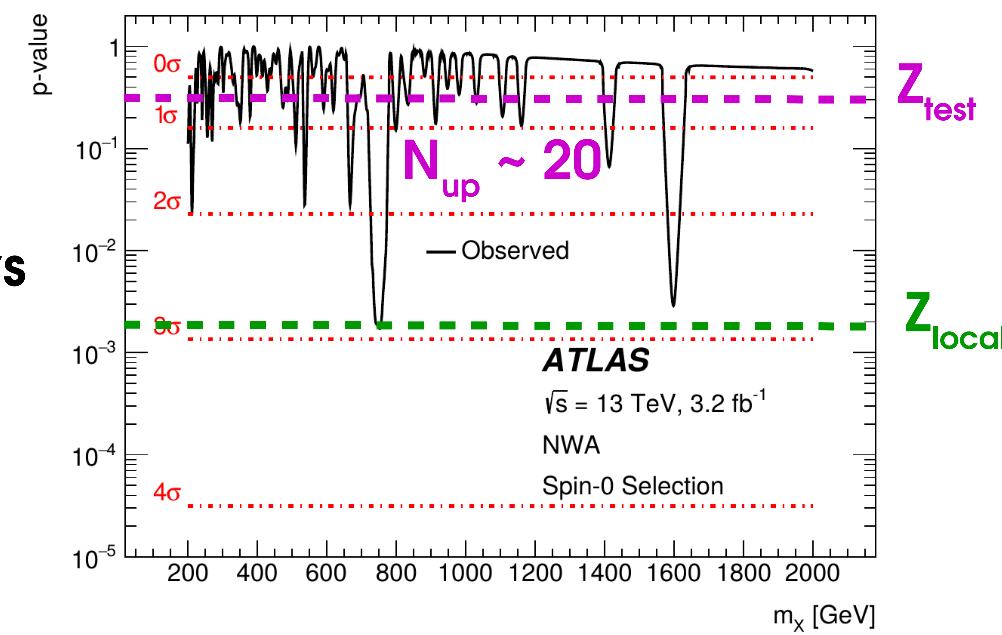
Can choose arbitrarily small Z_{test}

⇒ many excesses

⇒ can measure N_{up} in data (1 "toy")

Can also measure $\langle N_{up} \rangle$ in multiple toys

if large stat uncertainty from
too few excesses

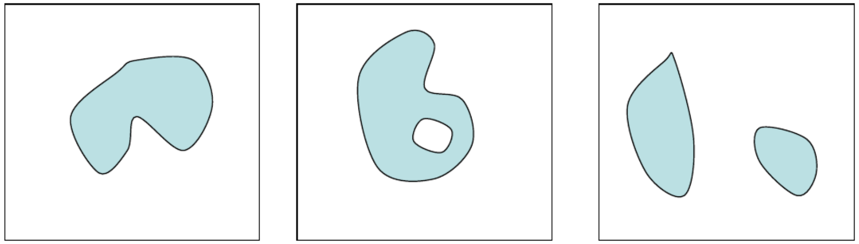


In 2D

O. Vitells and E. Gross, Astropart. Phys. 35 (2011) 230

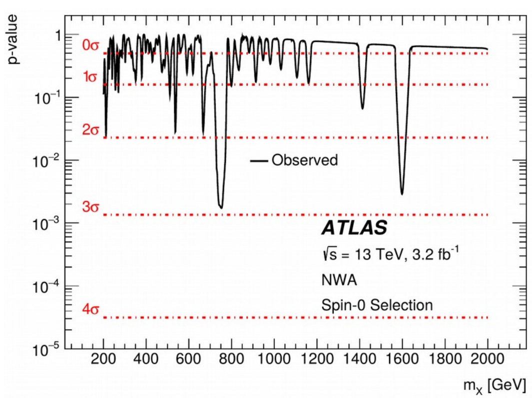
Generalization to 2D scans: consider sections at a fixed Z_{test}, compute its ***Euler characteristic*** φ, and use

$$p_{\text{global}} \approx E[\phi(A_u)] = p_{\text{local}} + e^{-u/2}(N_1 + \sqrt{u}N_2)$$



$\phi=1$ $\phi=0$ $\phi=2$

→ Generalizes 1D bump counting



Now need to determine 2 constants N_1 and N_2, from Euler φ measurements at 2 different Z_{test} values.