

Practical Statistics

For Particle Physicists

Lecture Plan

Statistics basic concepts (Today)

[Basic ingredients (PDFs, etc.)]

Statistical Modeling (PDFs for particle physics measurements)

Parameter estimation (maximum likelihood, least-squares, ...)

Computing statistical results (Today)

Model testing (χ^2 tests, hypothesis testing, p-values, ...)

Discovery testing

Systematics and further topics (Tomorrow)

Confidence intervals

Upper limits

Systematics and profiling

[Bayesian techniques]

Disclaimer: the examples and methods covered in the lectures will be biased towards LHC techniques (generally close to the state of the art anyway)

The class will be based on both lectures and [hands-on tutorials](#)

Highlights : Hypothesis Tests and Discovery

Given a PDF $\mathbf{P}(\text{data}; \boldsymbol{\mu})$, define likelihood $\mathbf{L}(\boldsymbol{\mu}) = \mathbf{P}(\text{data}; \boldsymbol{\mu})$

To estimate a parameter, use the value $\hat{\boldsymbol{\mu}}$ that maximizes $\mathbf{L}(\boldsymbol{\mu}) \rightarrow$ best-fit value

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

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

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

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

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

Confidence Intervals

Confidence Intervals

Last lecture we saw how to estimate (=compute) the value of a parameter

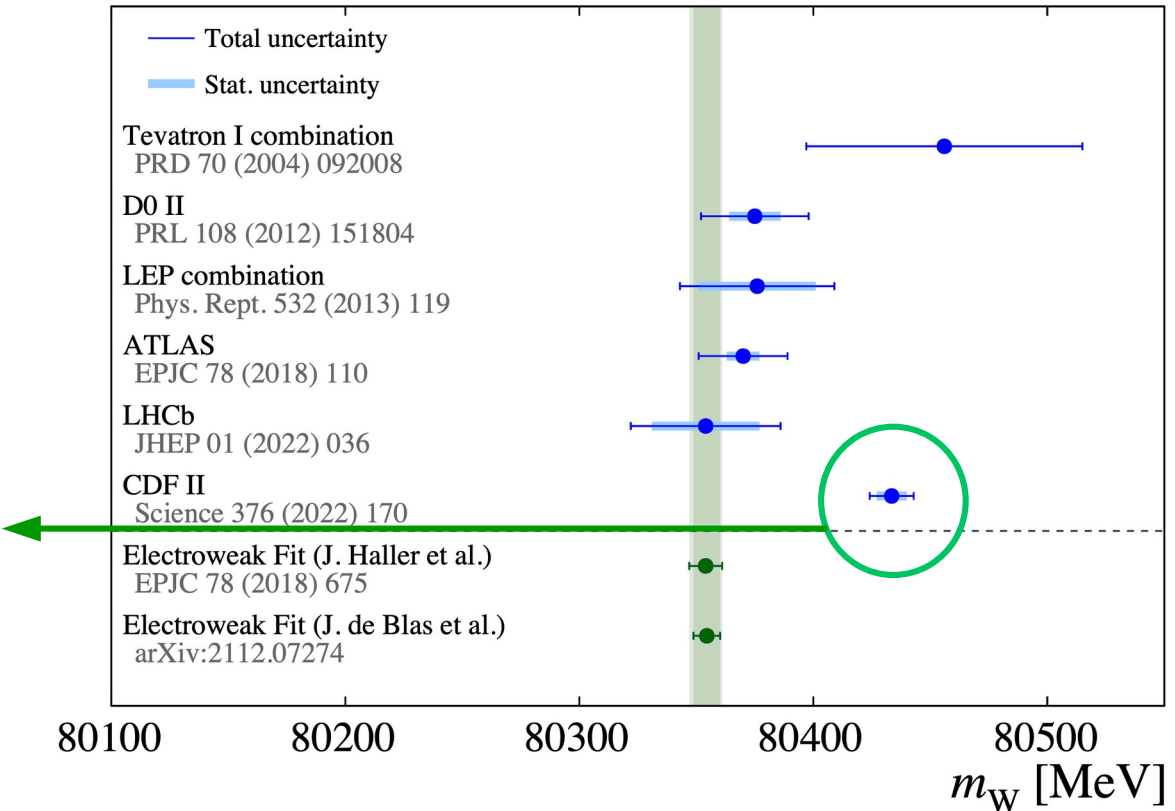
Maximum Likelihood
Estimator (MLE) $\hat{\mu}$:

$$\hat{\mu} = \arg \max L(\mu)$$

However we also need to estimate the associated uncertainty.

What is the meaning of an uncertainty ?

We don't know what the true value is, but **there is a 68% chance that it is within the error bar**

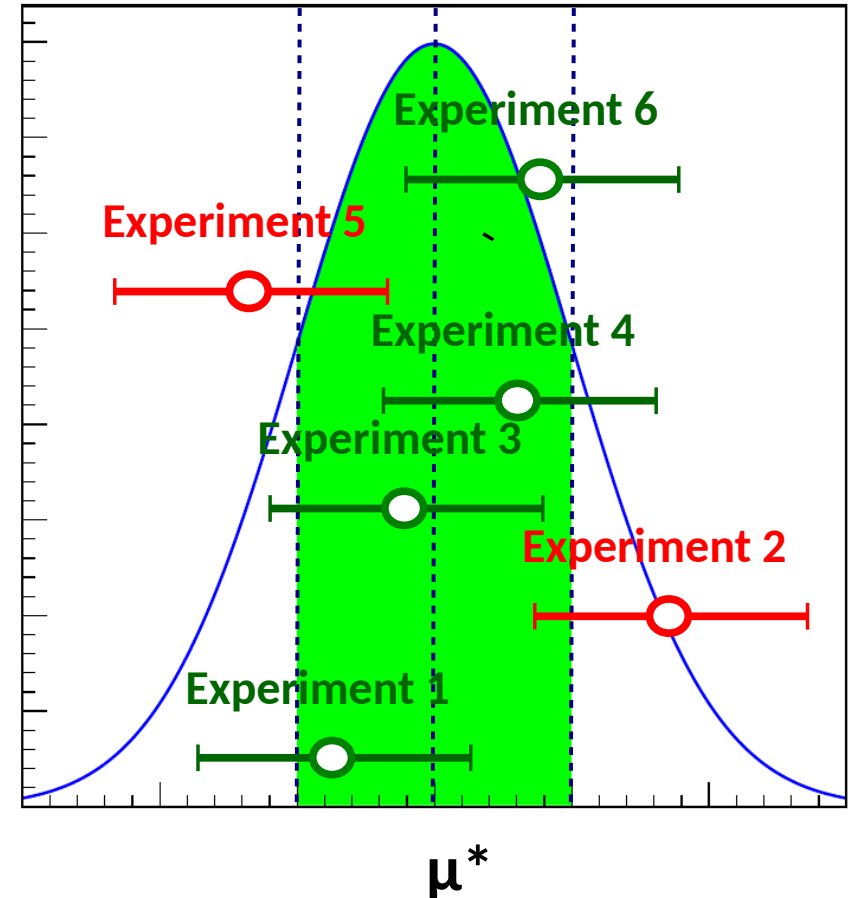


Gaussian confidence intervals

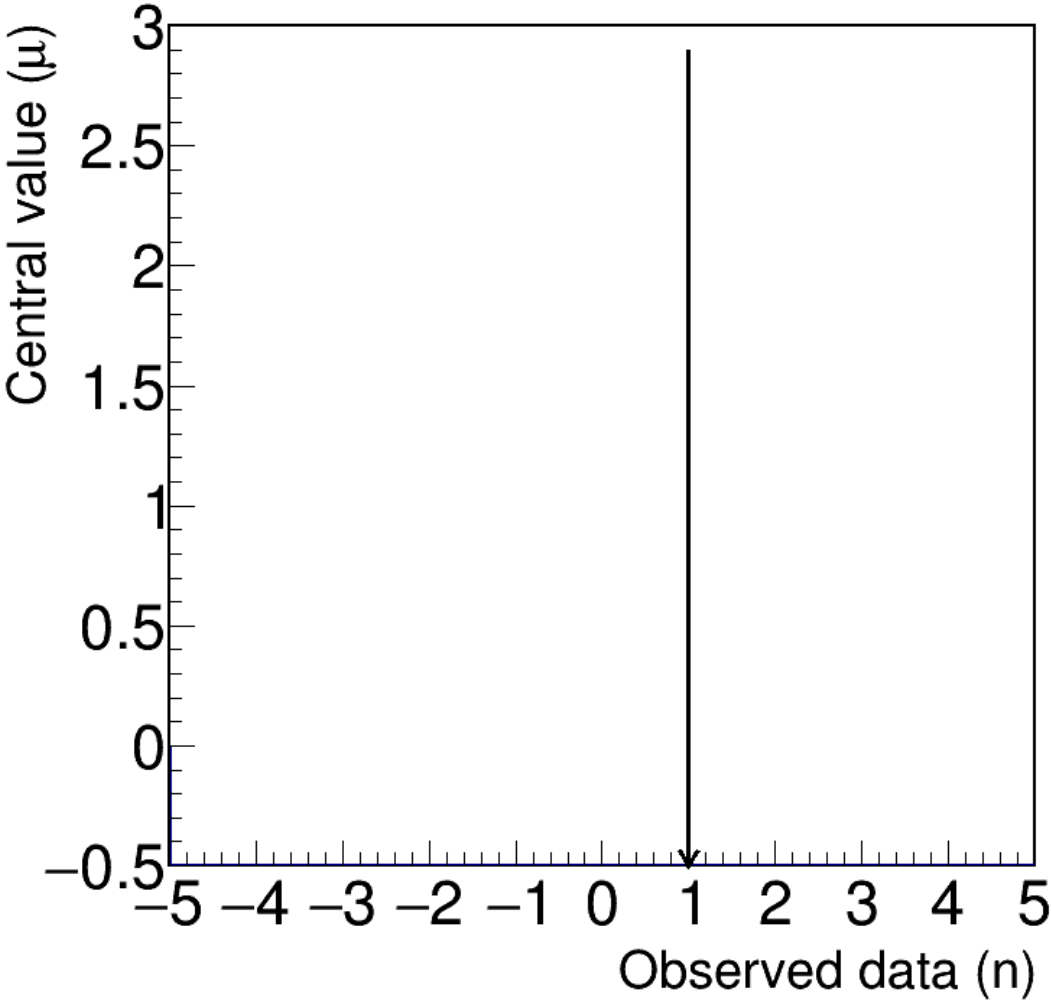
Imagine we repeat the same experiment multiple times, report a 1σ interval in each case.

→ Then 68.3% of these intervals must contain the true value μ^* .

→ **Crucially, must work even if we do not know μ^* !**

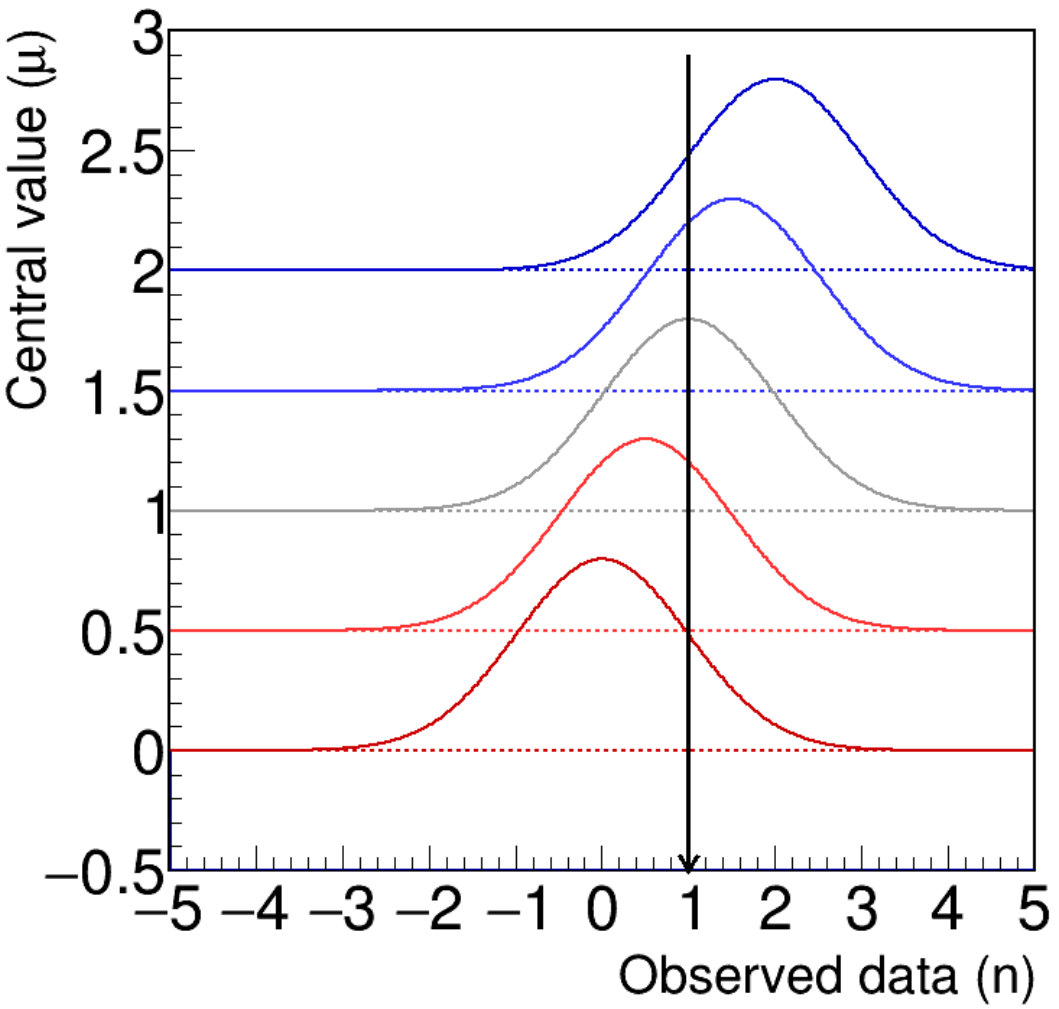


Gaussian confidence intervals



$$P(n; \mu) = \exp\left[-\frac{1}{2}\left(\frac{n - \mu}{\sigma}\right)^2\right]$$

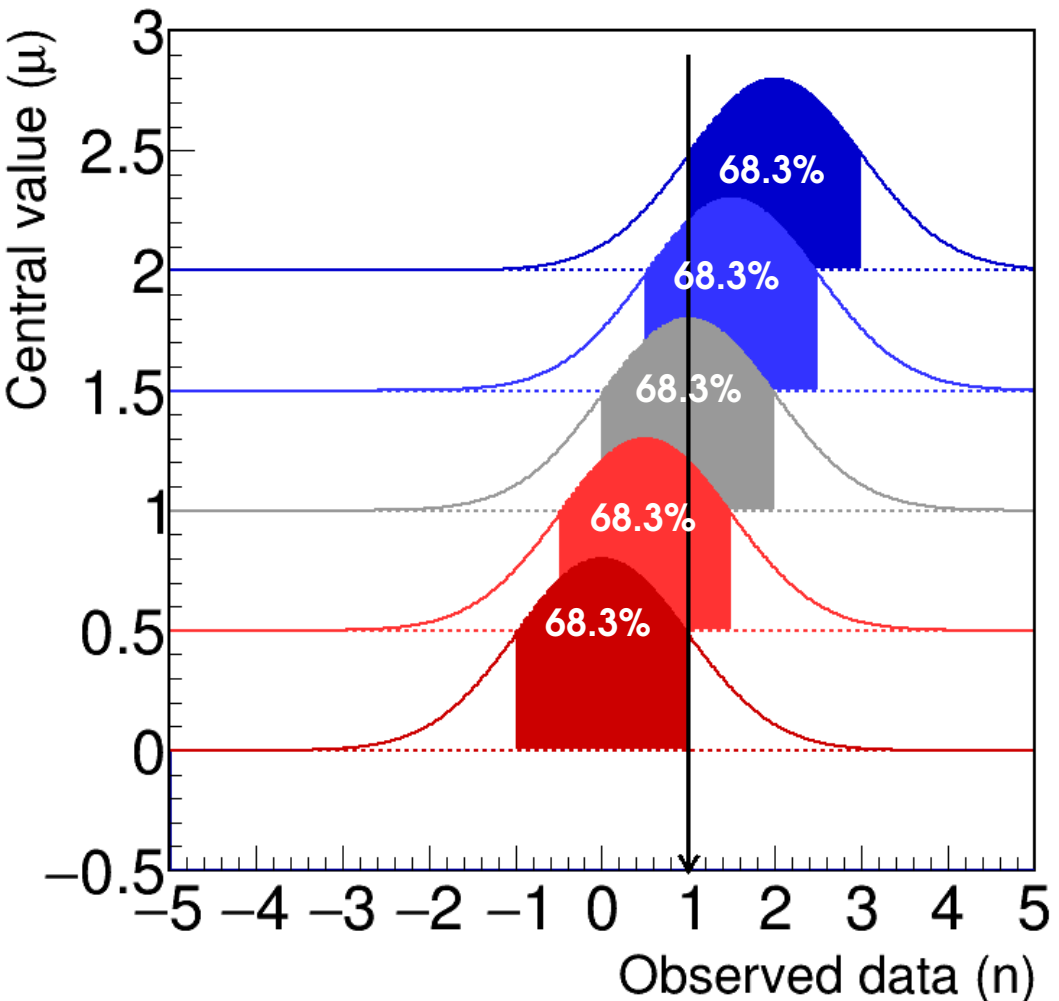
Gaussian confidence intervals



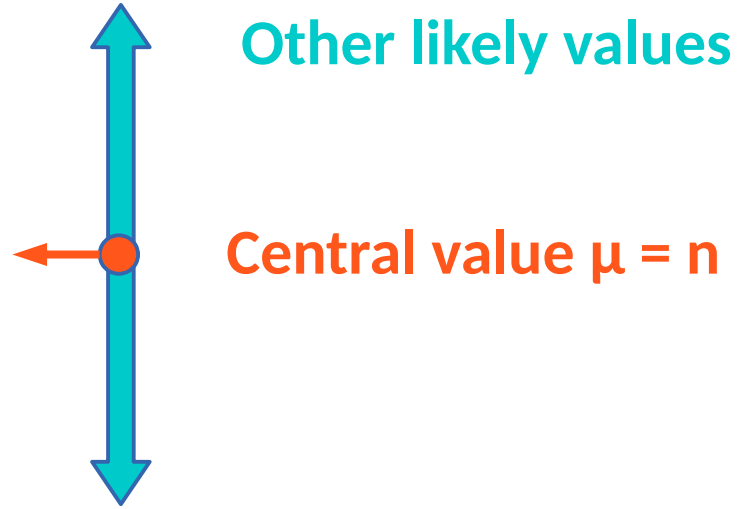
$$P(n; \mu) = \exp\left[-\frac{1}{2}\left(\frac{n - \mu}{\sigma}\right)^2\right]$$

← ● Central value $\mu = n$

Gaussian confidence intervals

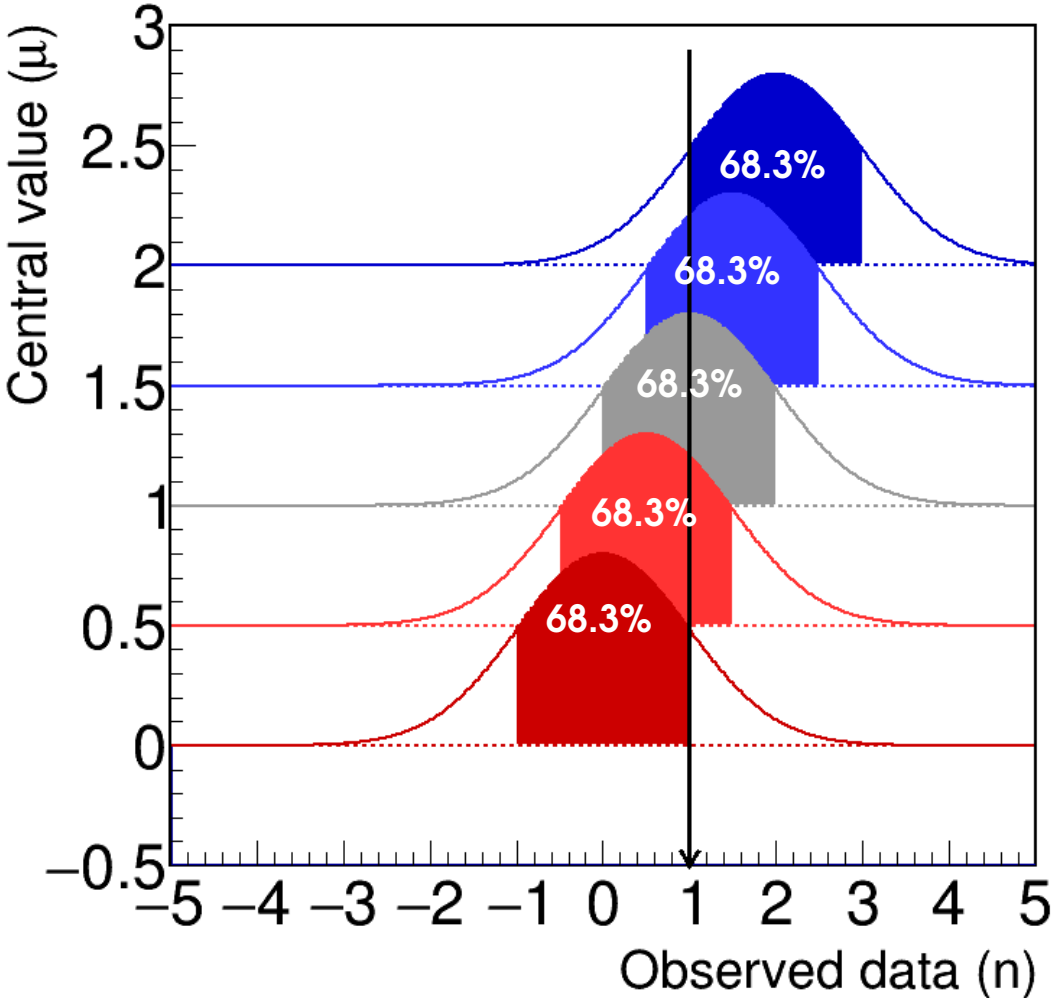


$$P(n; \mu) = \exp\left[-\frac{1}{2}\left(\frac{n-\mu}{\sigma}\right)^2\right]$$



How big an interval do we need around $\mu=n$ to include μ^* 68.3% of the time ?

Gaussian confidence intervals



$$P(n; \mu) = \exp\left[-\frac{1}{2}\left(\frac{n-\mu}{\sigma}\right)^2\right]$$

$$P(\mu^* - \sigma < n < \mu^* + \sigma) = 68.3\%$$



$$P(n - \sigma < \mu^* < n + \sigma) = 68.3\%$$

Still a statement on n !

$\mu = n \pm \sigma$ at 68% CL ("1 σ ")

The reported interval $n \pm \sigma$ will contain the true value of μ 68.3% of the time

Works even if we don't know μ^* !

General case: Likelihood Intervals

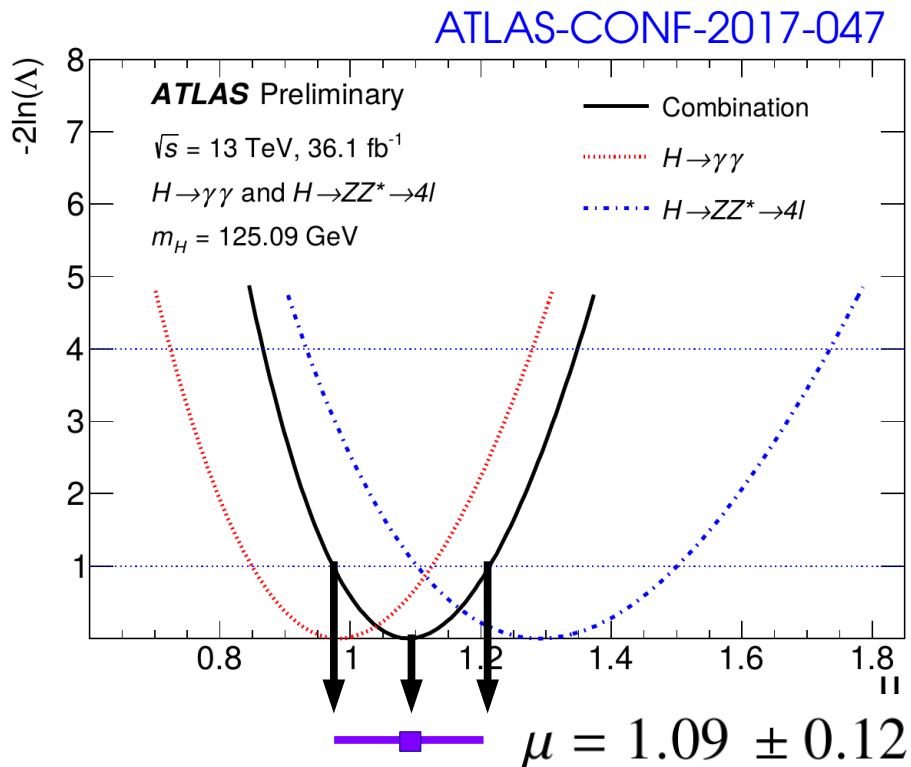
Confidence intervals from $L(\mu)$:

- Test various values μ using the **Profile Likelihood Ratio $t(\mu)$**
- Minimum (=0) for $\mu = \hat{\mu}$, rises away from $\hat{\mu}$.
- Good properties thanks to the Neyman-Pearson lemma.

Probability to observe the data **for a given μ** .

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

Probability to observe the data **for best-fit $\hat{\mu}$** .



Gaussian $L(\mu)$:

$$L(\mu) = \exp \left[-\frac{1}{2} \left(\frac{n - \mu}{\sigma} \right)^2 \right]$$

$$t(\mu) = \left(\frac{n - \mu}{\sigma} \right)^2$$

- $t(\mu)$ is parabolic, distributed as a χ^2
- Minimum occurs at $\mu = \hat{\mu}$

$t(\mu_{\pm}) = 1 \Rightarrow \mu = n \pm \sigma$ **1 σ interval**

General case: Likelihood Intervals

Confidence intervals from $L(\mu)$:

- Test various values μ using the **Profile Likelihood Ratio $t(\mu)$**
- Minimum (=0) for $\mu = \hat{\mu}$, rises away from $\hat{\mu}$.
- Good properties thanks to the Neyman-Pearson lemma.

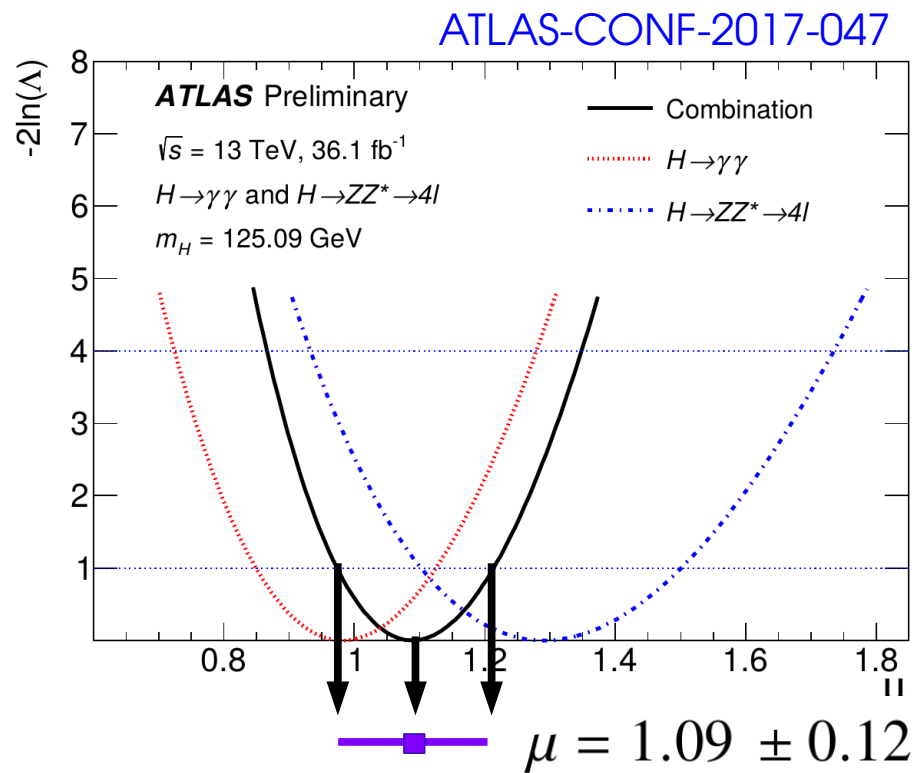
$$t(\mu) = -2 \log \frac{L(\mu)}{L(\hat{\mu})}$$

General case:

- Generally not a perfect parabola
- Minimum still at $\mu = \hat{\mu}$

Asymptotic approximation

- Compute $t(\mu)$ using the exact $L(\mu)$
- **1σ interval given by $t(\mu) = 1$**



Homework 3: Gaussian Case

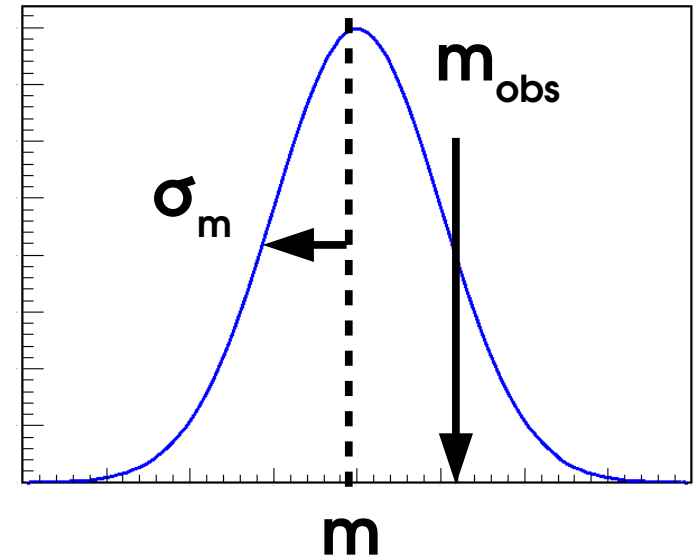
Consider a parameter m (e.g. Higgs boson mass) whose measurement is Gaussian with known width σ_m , and we measure m_{obs} :

$$L(m; m_{\text{obs}}) = e^{-\frac{1}{2} \left(\frac{m - m_{\text{obs}}}{\sigma_m} \right)^2}$$

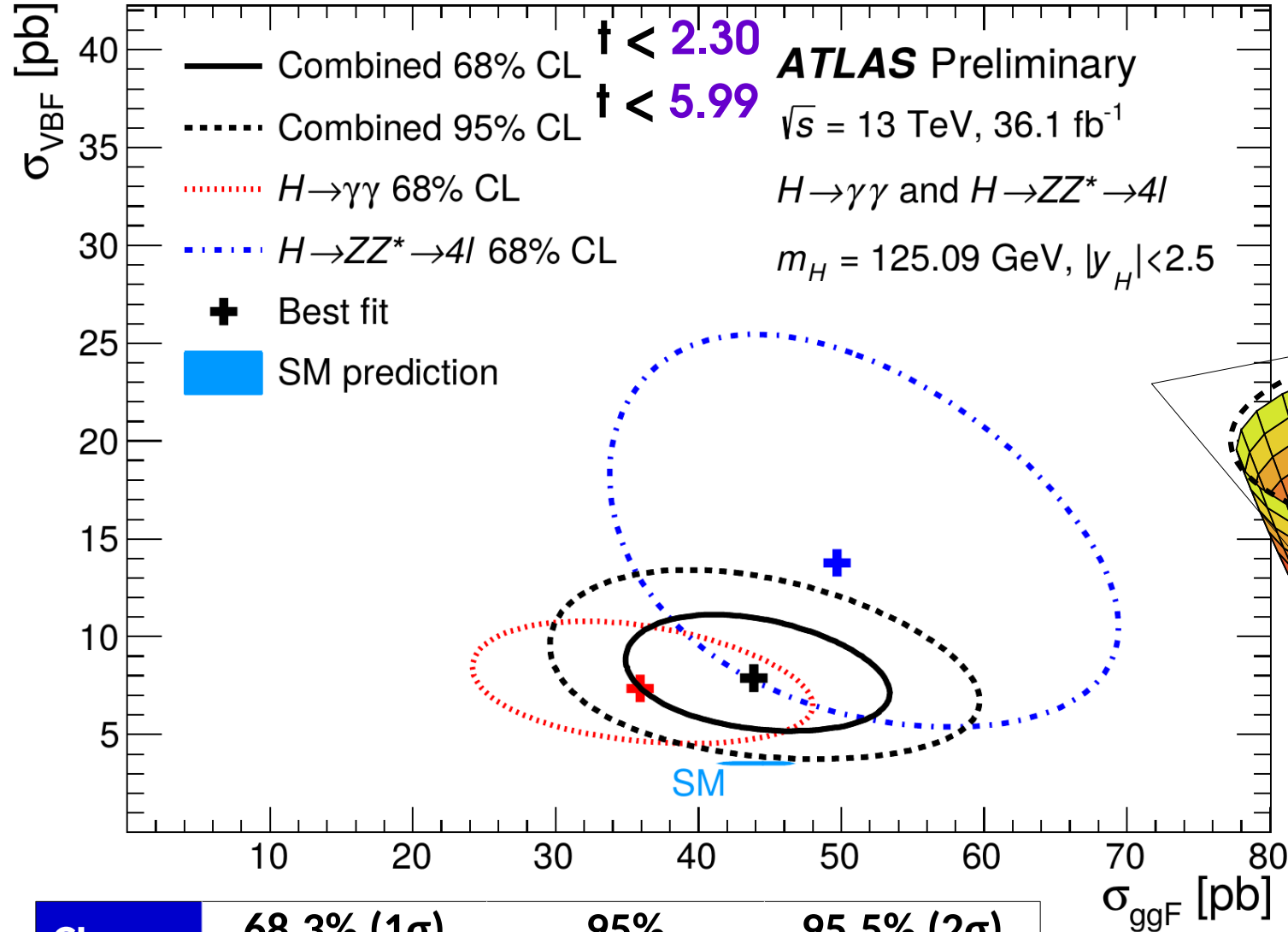
- Compute the best-fit value (MLE) \hat{m}
- Compute $t(m)$
- Compute the 1σ (68.3% CL) interval on m

Solution: $m = m_{\text{obs}} \pm \sigma_m$

- As expected!
- General method can be applied in the same way to more complex cases

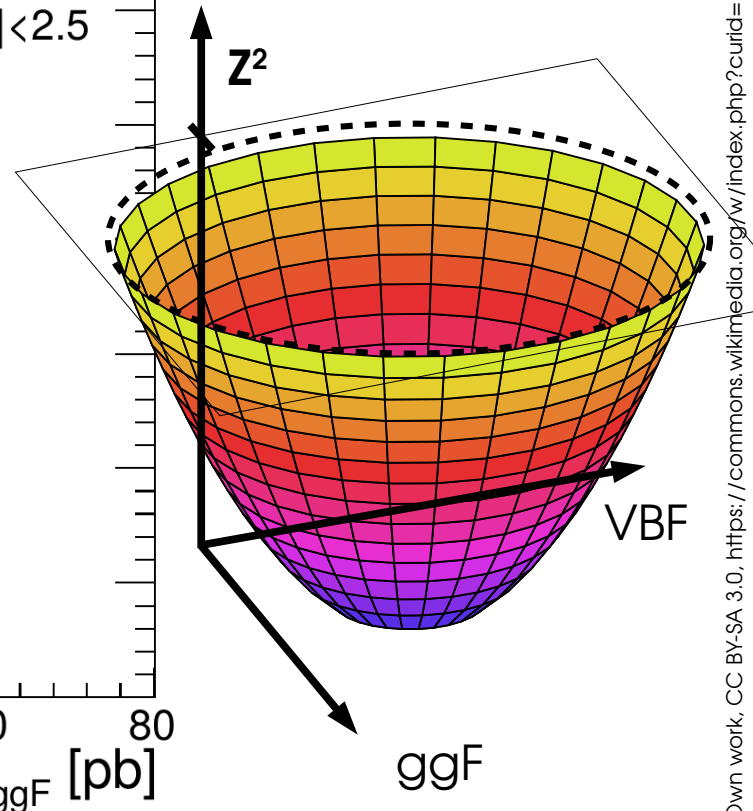


2D Example: Higgs σ_{VBF} vs. σ_{ggF}



$$t = -2 \log \frac{L(X_0, Y_0)}{L(\hat{X}, \hat{Y})}$$

$$\sim \chi^2(N_{\text{dof}}=2)$$



CL	68.3% (1σ)	95%	95.5% (2σ)
1D Z^2	1.00	3.84	4.00
2D Z^2	2.30	5.99	6.18

Gaussian case: elliptic paraboloid surface

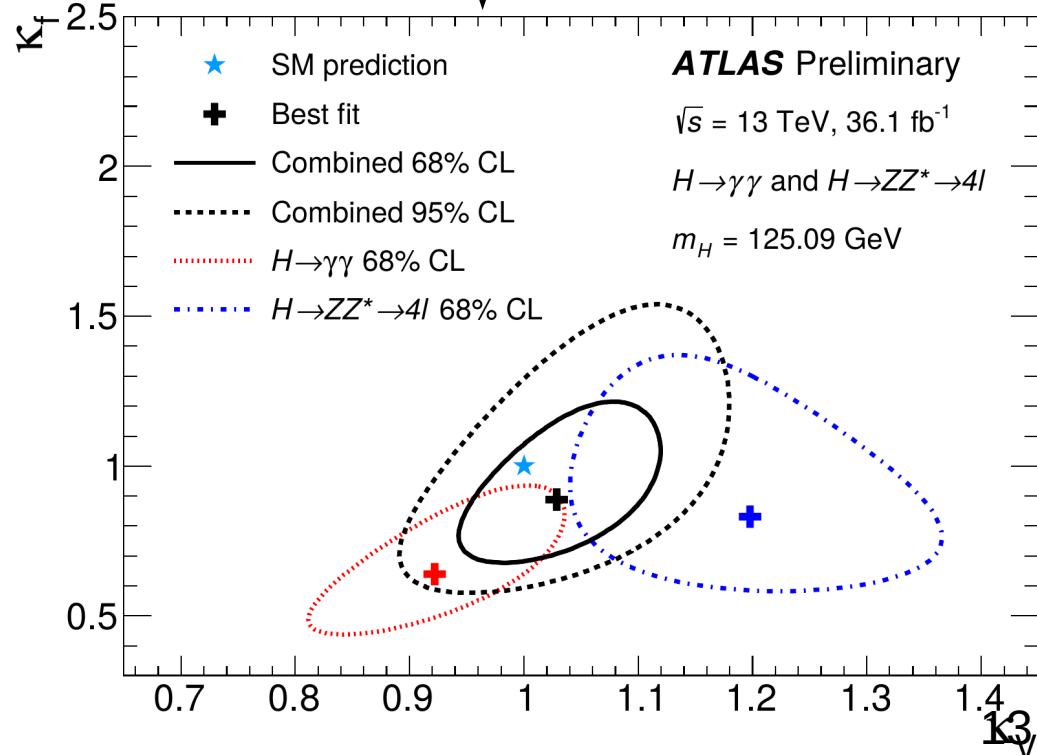
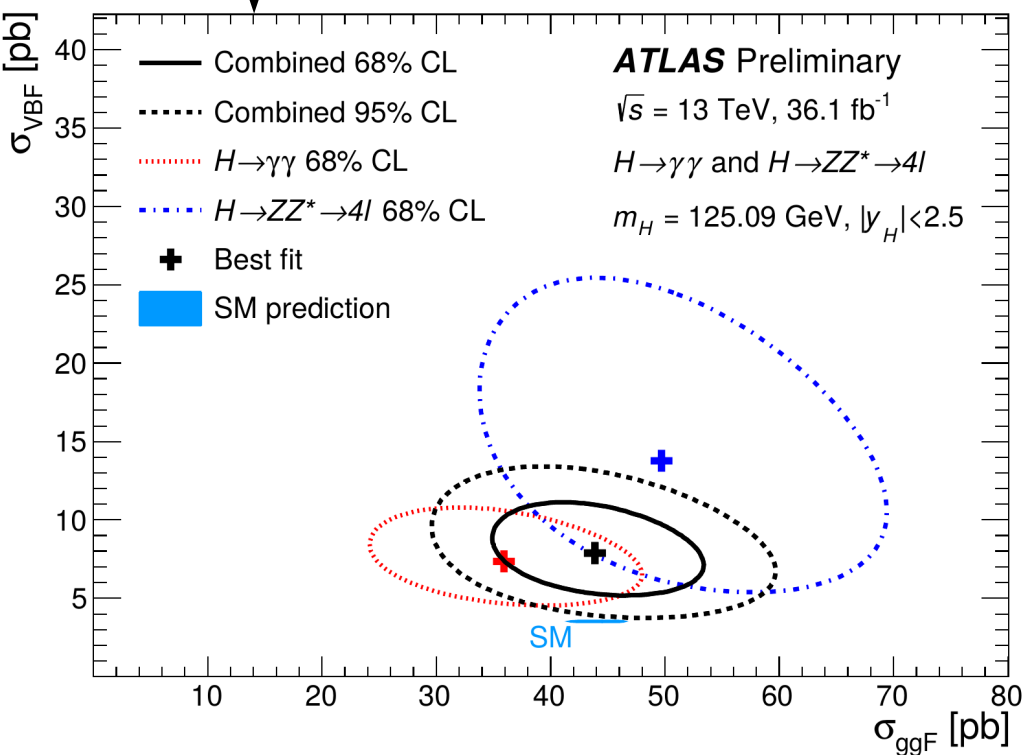
Reparameterization

Start with basic measurement in terms of e.g. $(\sigma \times \mathbf{B})$

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

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

$$L(\sigma_{ggF}, \sigma_{VBF}) \xrightarrow{\substack{\sigma_{ggF} \rightarrow \sigma_{ggF}(\kappa_V, \kappa_F) \\ \sigma_{VBF} \rightarrow \sigma_{VBF}(\kappa_V, \kappa_F)}} L(\sigma_{ggF}(\kappa_V, \kappa_F), \sigma_{VBF}(\kappa_V, \kappa_F)) \equiv L'(\kappa_V, \kappa_F)$$



Upper Limits

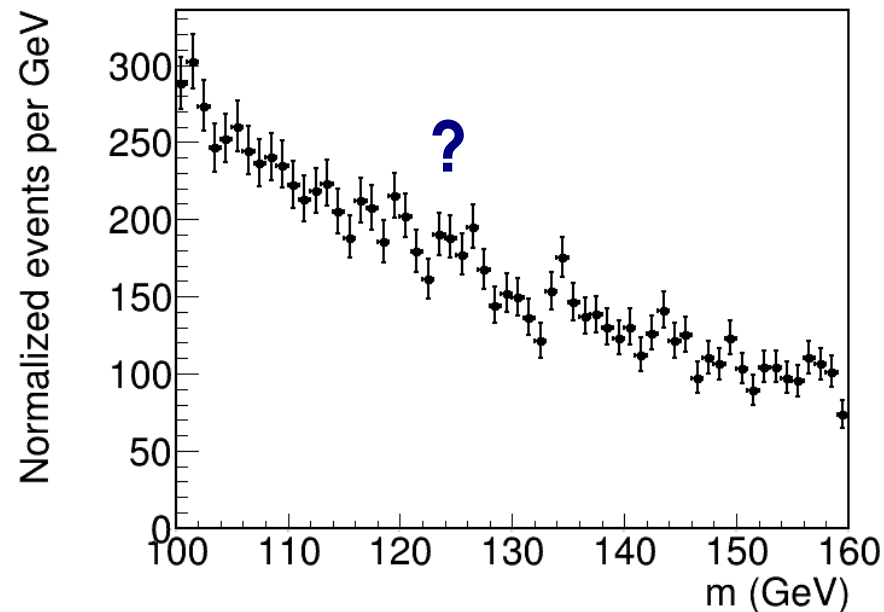
Hypothesis tests for Limits

If no signal in data, testing for discovery not very relevant (report 0.2σ excess ?)

→ More interesting to **exclude large signals**

⇒ **Upper limits on signal yield**

→ Typically report **95% CL** upper limit (p-value = 5%) : “ $S < S_0$ @ 95% CL”



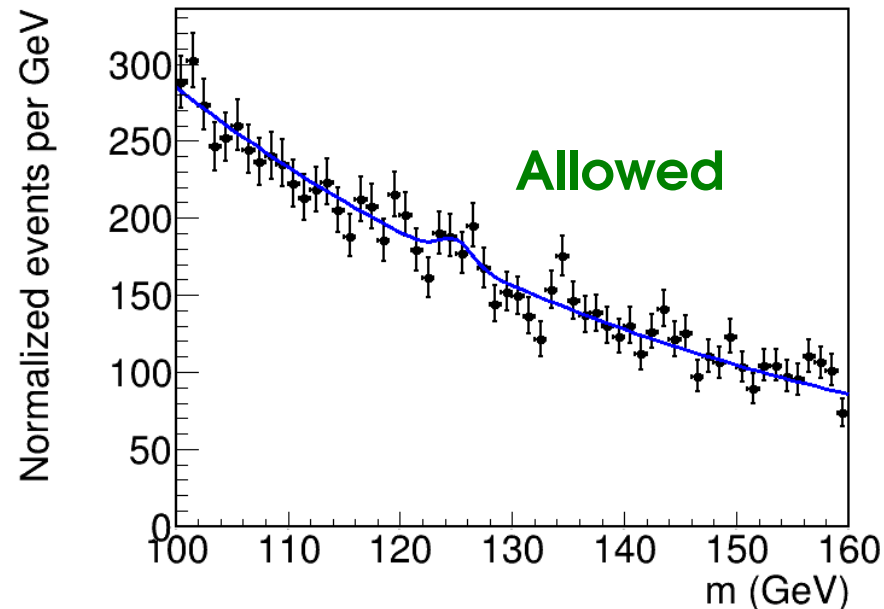
Hypothesis tests for Limits

If no signal in data, testing for discovery not very relevant (report 0.2σ excess ?)

→ More interesting to **exclude large signals**

⇒ **Upper limits on signal yield**

→ Typically report **95% CL** upper limit (p-value = 5%) : “ $S < S_0$ @ 95% CL”



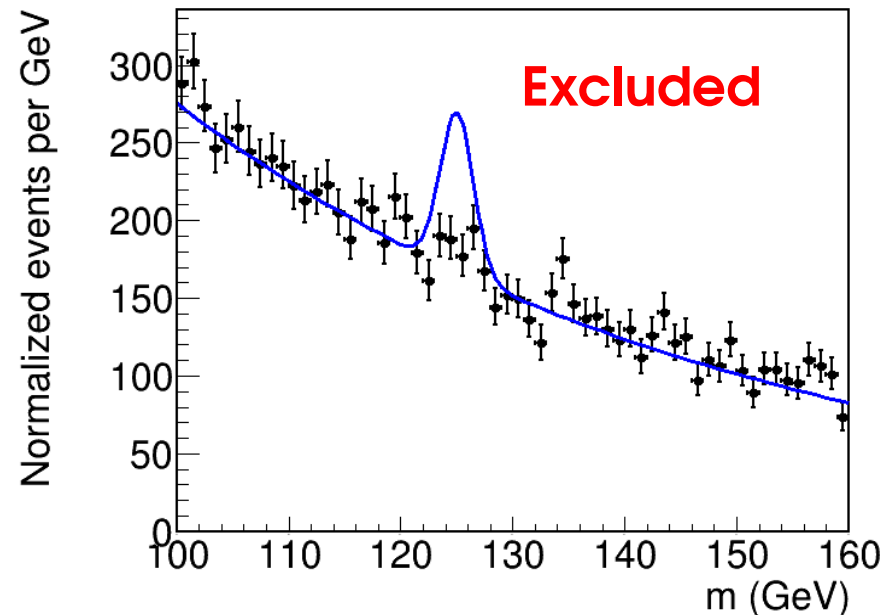
Hypothesis tests for Limits

If no signal in data, testing for discovery not very relevant (report 0.2σ excess ?)

→ More interesting to **exclude large signals**

⇒ **Upper limits on signal yield**

→ Typically report **95% CL** upper limit (p-value = 5%) : “ $S < S_0$ @ 95% CL”



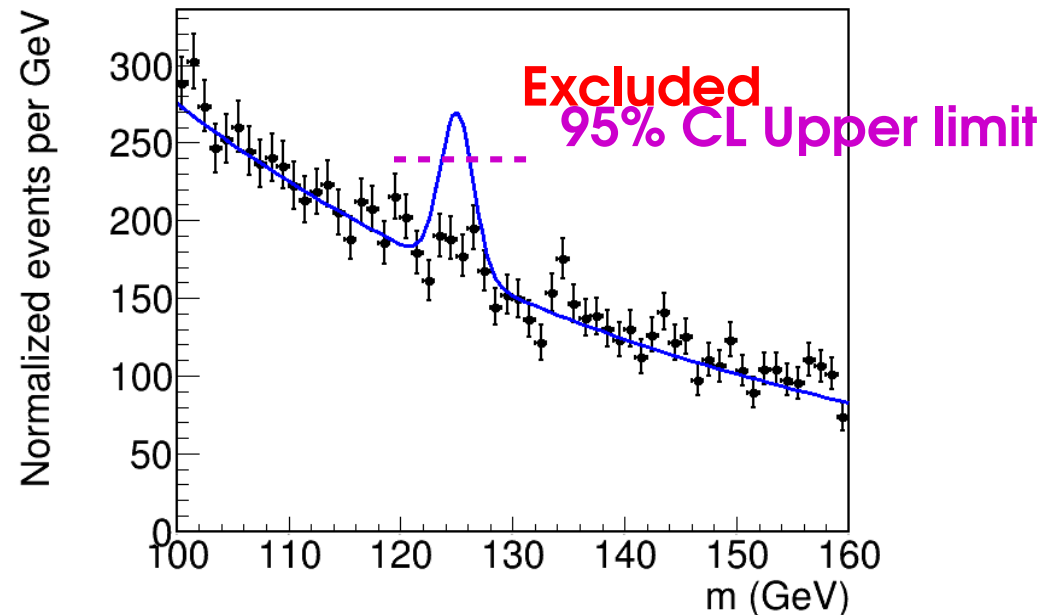
Hypothesis tests for Limits

If no signal in data, testing for discovery not very relevant (report 0.2σ excess ?)

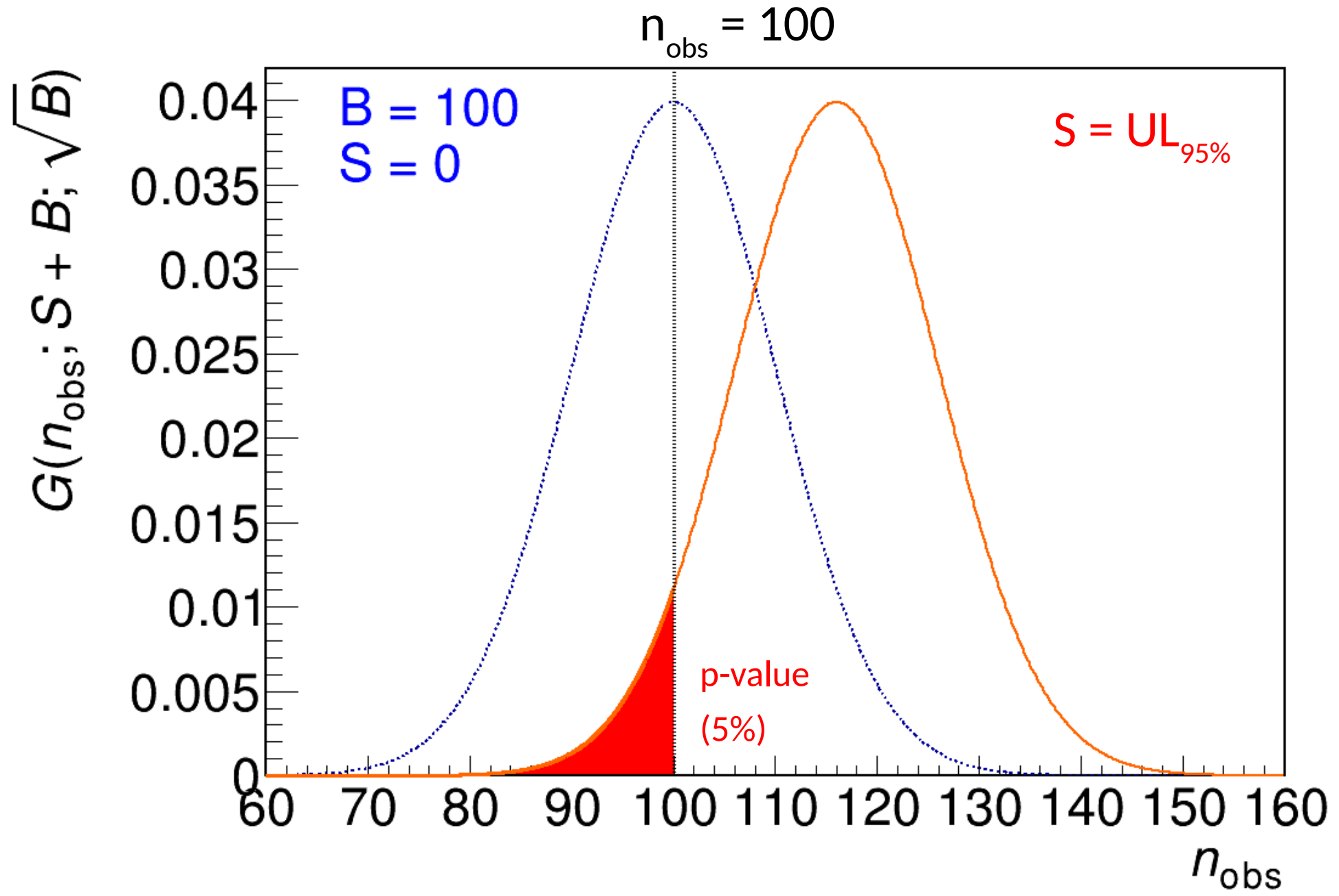
→ More interesting to **exclude large signals**

⇒ **Upper limits on signal yield**

→ Typically report **95% CL** upper limit (p-value = 5%) : “ $S < S_0$ @ 95% CL”

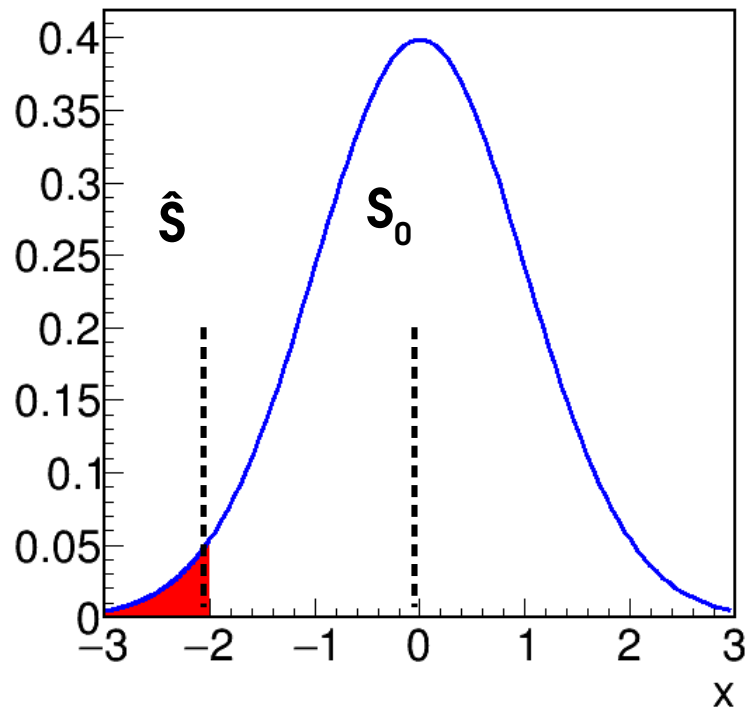


Upper Limit @ 95% CL in Gaussian counting (S=0)



Test Statistics for Limit-Setting

Limit-setting: try to exclude values of S that are above \hat{S} .



$$q(S_0) = \begin{cases} -2 \log \frac{L(S=S_0)}{L(\hat{S})} & S_0 > \hat{S} \\ 0 & S_0 \leq \hat{S} \end{cases}$$

“One-sided” test : only interested in excluding above

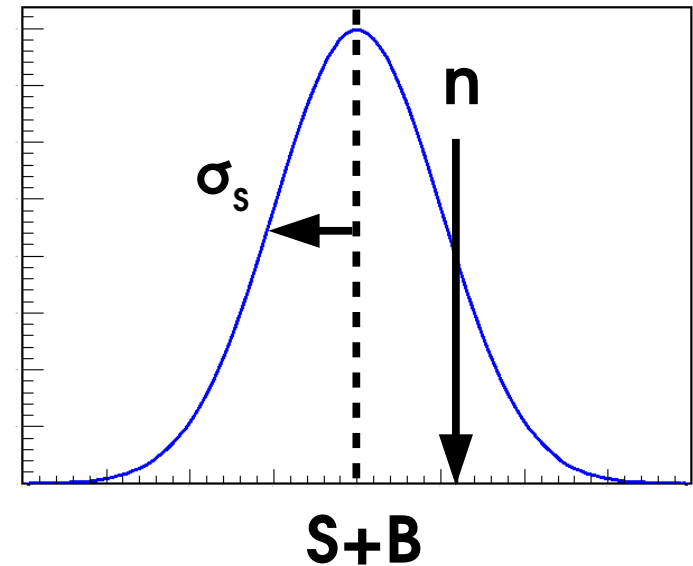
Discovery was also one-sided, for $S > 0$

Homework 4: Gaussian Example

Usual Gaussian counting example with known B:

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

$$\sigma_s \sim \sqrt{B} \text{ for small } S$$



Reminder: Significance: $Z = \hat{S}/\sigma_s$

→ Compute $q(S_0)$

→ Compute the 95% CL upper limit on S , S_{up} , by solving $\sqrt{q_{S_0}} = 1.64$.

Solution: $S_{up} = \hat{S} + 1.64 \sigma_s$ at 95% CL

Upper limits sometimes take negative values (exclude all $S > 0$!)

Known feature – to avoid, usual solution in HEP is to use **CL_s** ”modified p-value”

$$P_{CL_s} = \frac{p(S_0)}{p_B}$$

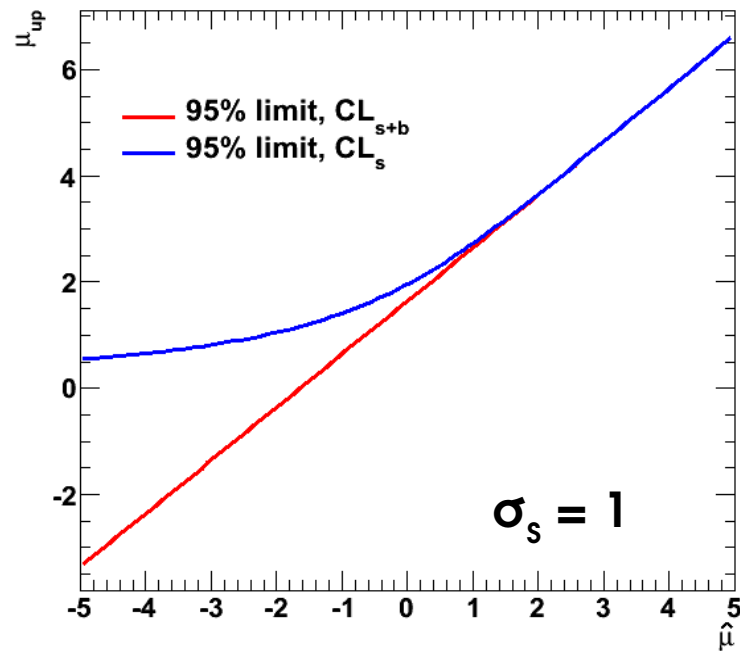
Usual p-value for $S=S_0$

P-value for $S=0$

⇒ Compute exclusion relative to that of $S=0$
 → Somewhat ad-hoc, but good properties...

$\hat{S} \sim 0 \Rightarrow p_B \sim O(1), p_{CL_s} \sim p(S_0)$ no change

$\hat{S} \ll 0 \Rightarrow p_B \ll 1, p_{CL_s} \gg p(S_0)$ no exclusion at $S=0$



Drawback: overcoverage

→ limit is claimed to be 95% CL, but actually >95% CL for small p_B .

Homework 5: CL_s in the Gaussian Case

Usual Gaussian counting example with known B:

$$L(S; n) = e^{-\frac{1}{2} \left(\frac{n - (S+B)}{\sigma_s} \right)^2} \quad \sigma_s \sim \sqrt{B} \text{ for small } S$$

Reminder

CL_{s+b} limit: $S_{up} = \hat{S} + 1.64 \sigma_s$ at 95 % CL

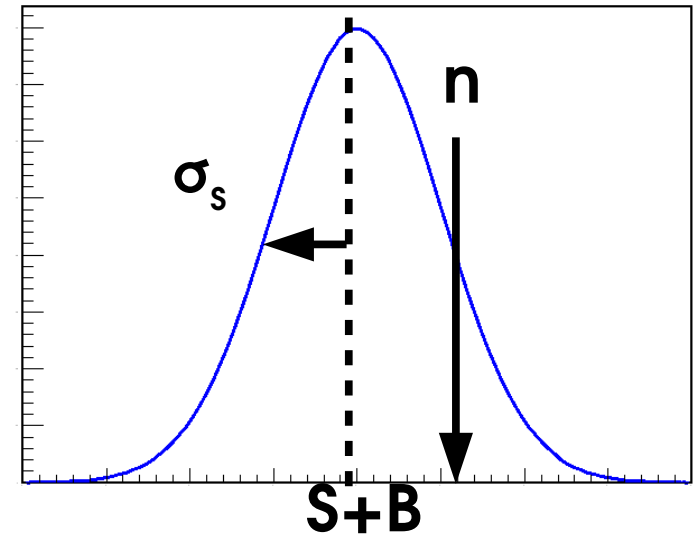
CL_s upper limit :

→ Compute p_{s_0} (same as for CL_{s+b})

→ Compute $1-p_B$ (hard!)

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

for $\hat{S} \sim 0$, $S_{up} = \hat{S} + 1.96 \sigma_s$ at 95 % CL



Homework 6: CL_s Rule of Thumb for $n_{obs}=0$

Same exercise, for the Poisson case with $n_{obs} = 0$. Perform an exact computation of the 95% CL_s upper limit based on the definition of the p-value:

p-value : *sum probabilities of cases at least as extreme as the data*

Hint: for $n_{obs}=0$, there are no “more extreme” cases (cannot have $n < 0$!), so

$p_{S_0} = \text{Poisson}(n=0 \mid S_0+B)$ and $1 - p_B = \text{Poisson}(n=0 \mid B)$

Solution: $S_{up}(n_{obs}=0) = \log(20) = 2.996 \approx 3$

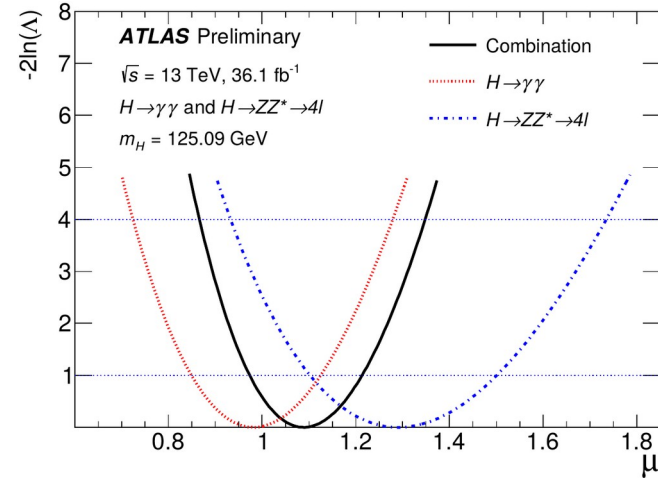
⇒ Rule of thumb: when $n_{obs} = 0$, the 95% CL_s limit is 3 events (for any B)

Highlights: Confidence intervals and Upper Limits

Confidence intervals: use $t(\mu_0) = -2 \log \frac{L(\mu = \mu_0)}{L(\hat{\mu})}$

→ Crossings with $t(\mu_0) = 1$ for 1σ intervals (in 1D)

Gaussian regime: $\mu = \hat{\mu} \pm \sigma_\mu$ at 68.3% CL (1σ interval)

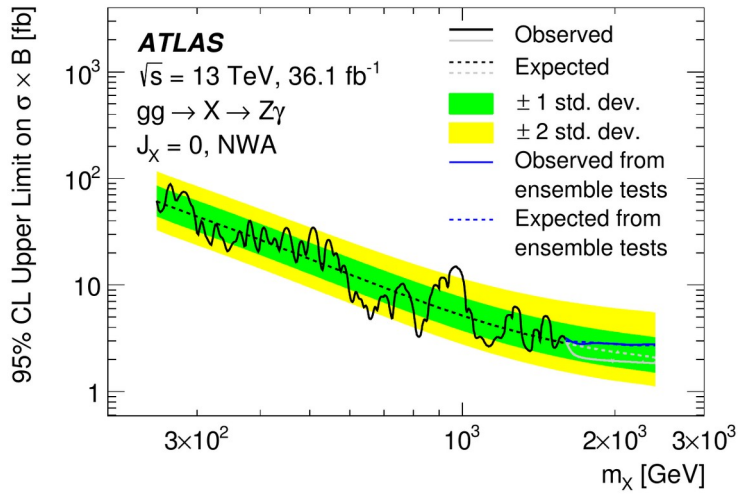


Limits : use LR-based test statistic: $q_{S_0} = -2 \log \frac{L(S=S_0)}{L(\hat{S})} \quad S_0 \geq \hat{S}$

→ Use CL_s procedure to avoid negative limits

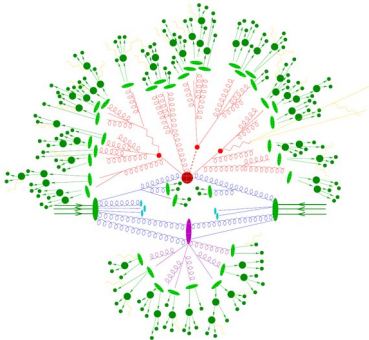
Gaussian regime, $n \sim 0$: $S < \hat{S} + 1.96\sigma$ at 95% CL

Poisson regime, $n=0$: $S < 3$ events at 95% CL



Systematic Errors

Reminder on Statistical Modeling



Random data must be described using a statistical model:

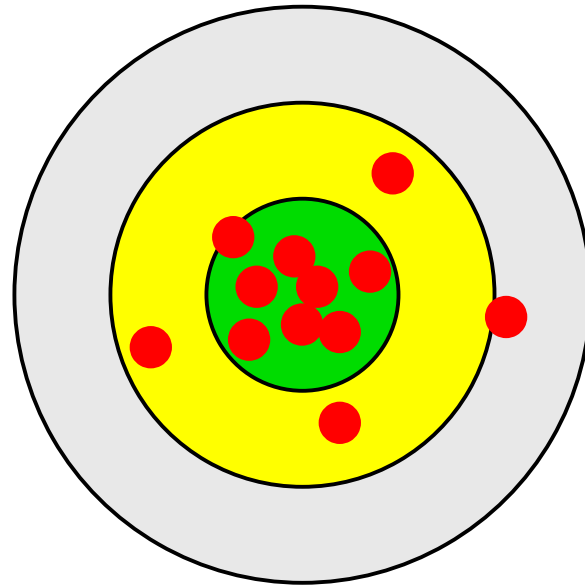
Description	Observable	Likelihood
Counting	n	<p>Poisson</p> $P(n; S, B) = e^{-(S+B)} \frac{(S+B)^n}{n!}$
Binned shape analysis	$n_i, i = 1 \dots N_{bins}$	<p>Poisson product</p> $P(\mathbf{n}_i; S, B) = \prod_{i=1}^{n_{bins}} e^{-(S f_i^{sig} + B f_i^{bkg})} \frac{(S f_i^{sig} + B f_i^{bkg})^{n_i}}{n_i!}$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

Model include

- **Parameters of interest** (POIs) – e.g. S but also
- **Nuisance parameters** (NPs) – e.g. B .

Systematic Errors

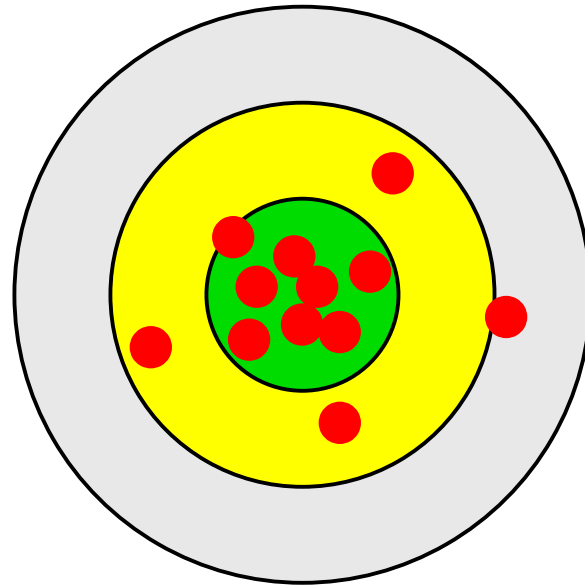
The PDF expresses *uncertainties on the outcome of a measurement*



These uncertainties are called **Statistical Uncertainties**.

Systematic Errors

The PDF expresses *uncertainties on the outcome of a measurement*



These uncertainties are called **Statistical Uncertainties**.

However **the model itself may not be known exactly**

→ We also need **systematic uncertainties**, i.e. *on the form of the PDF itself*.

Systematics

Systematics = what we don't know about the random process.

"Systematic uncertainty is, in any statistical inference procedure, the uncertainty due to the incomplete knowledge of the probability distribution of the observables.

G. Punzi, *What is systematics ?*

How to describe them in practice ?

→ Parameterize using additional nuisance parameters (NPs)

But: if the NPs are completely free, no measurement is possible (e.g. free B ?...)

⇒ Add constraints in the likelihood

$$L(\mu, \theta; \text{data}) = L_{\text{measurement}}(\mu, \theta; \text{data}) C(\theta)$$

Diagram illustrating the components of the likelihood function $L(\mu, \theta; \text{data})$:

- μ (green) is labeled **POI** (Point of Interest).
- θ (red) is labeled **Systematics NP** (Nuisance Parameter).
- $L_{\text{measurement}}(\mu, \theta; \text{data})$ (blue) is labeled **Measurement Likelihood**.
- $C(\theta)$ (purple) is labeled **NP Constraint term**.

$C(\theta)$ represents **external knowledge** about the NPs that we inject into the statistical model – e.g. to say that “ $B \sim 100 \pm 5$ ”

Frequentist Systematics

Prototype: Systematics NP → measured in a separate *auxiliary experiment*
e.g. background levels.

→ Build the **combined PDF** of the main+auxiliary measurements

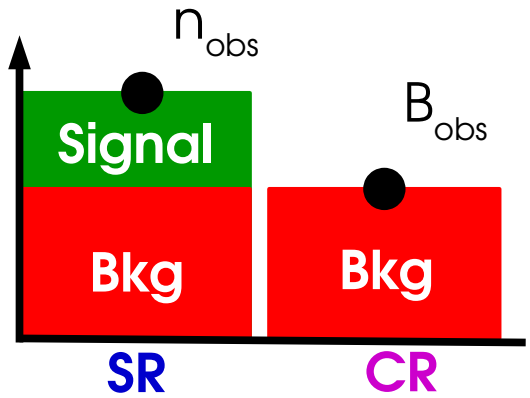
$$P(\mu, \theta; \text{data}) = P_{\text{main}}(\mu, \theta; \text{main data}) P_{\text{aux}}(\theta; \text{aux. data})$$

Independent measurements:
⇒ just a product

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

In the combined likelihood, **systematic NPs are constrained**
⇒ Measured simultaneously with the POIs. in a fit to data.

- Often no clear setup for auxiliary measurements (e.g. theory simulation uncertainties)
- **Define constraints “by hand”** (“pseudo-measurement”)



Profiling Nuisance Parameters

Profiling

How to deal with nuisance parameters in likelihood ratios ?

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

⇒ Profile Likelihood Ratio (PLR)

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

$\hat{\theta}(S_0)$ 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 without NPs

$$f(t_{S_0} | S=S_0) = f_{\chi^2(n_{dof}=1)}(t_{S_0}) \quad \text{also with NPs present}$$

→ Profiling “builds in” the effect of the NPs

⇒ Can use $t(S_0)$ 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)**: $n_{obs} \sim G(S + B, \sigma_{stat})$

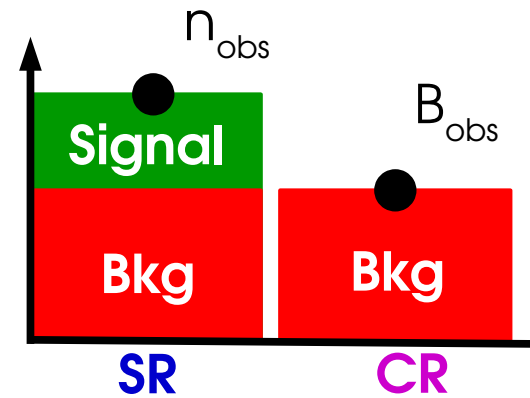
→ **Control region (CR)**: $B_{obs} \sim G(B, \sigma_{bkg})$

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

Recall: Signal region only (fixed B): $t(S) = \left(\frac{S - n_{obs}}{\sigma_{stat}} \right)^2$ $S = (n_{obs} - B) \pm \sigma_{stat}$

- Compute the best-fit (MLEs) for S and B
- Show that the conditional MLE for B is

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



- Compute the profile likelihood $t(S)$
- Compute the 1σ confidence interval on S

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

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

Uncertainty decomposition

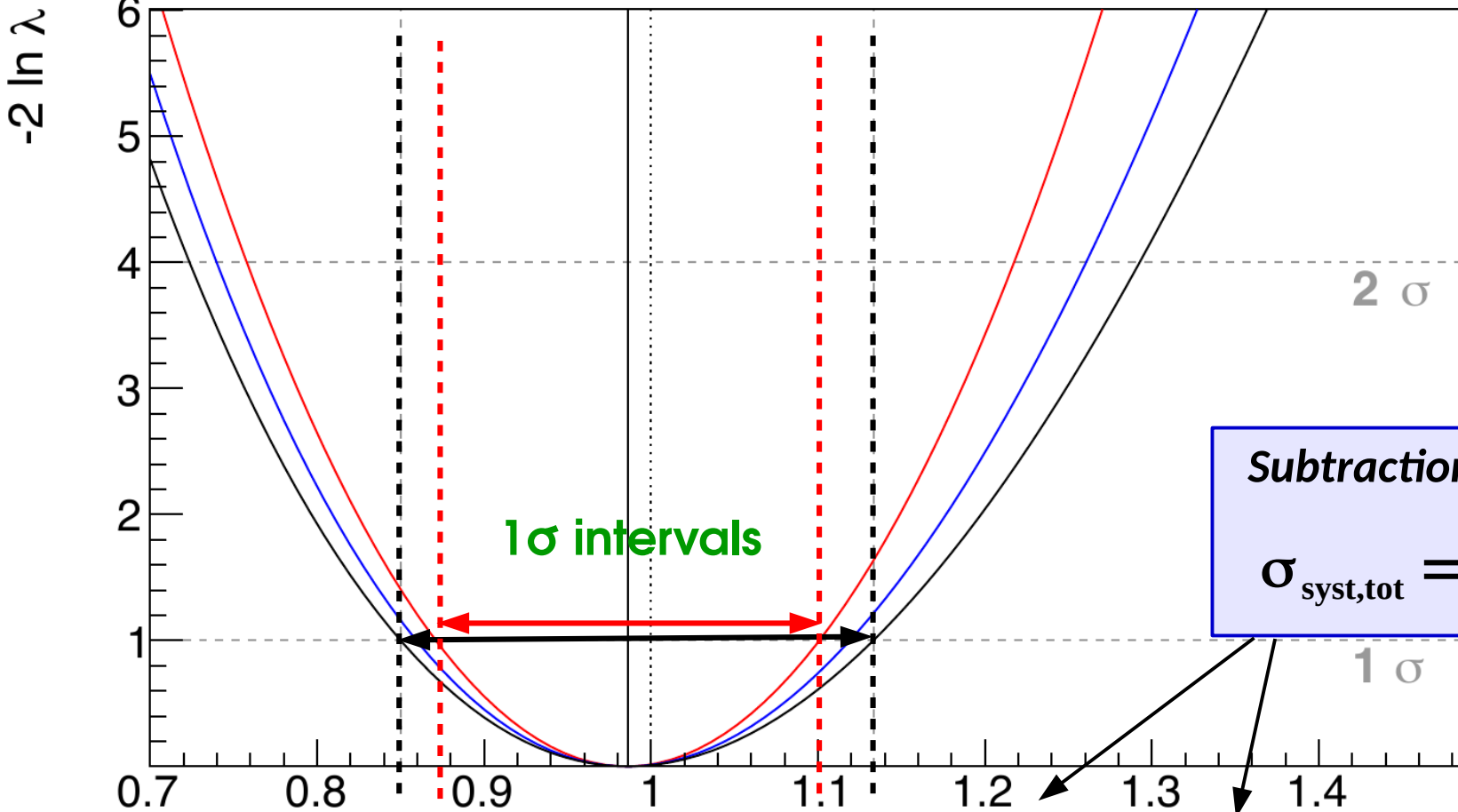
All systematics NPs excluded : statistical uncertainty only

All systematics NPs included: stat+syst uncertainties

— Total
— Theory
— Stat

ATLAS

$H \rightarrow \gamma\gamma, m_H = 125.09 \text{ GeV}$



Subtraction in quadrature

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

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

Pull/Impact plots

Systematics are described by NPs included in the fit. Define **pull** as

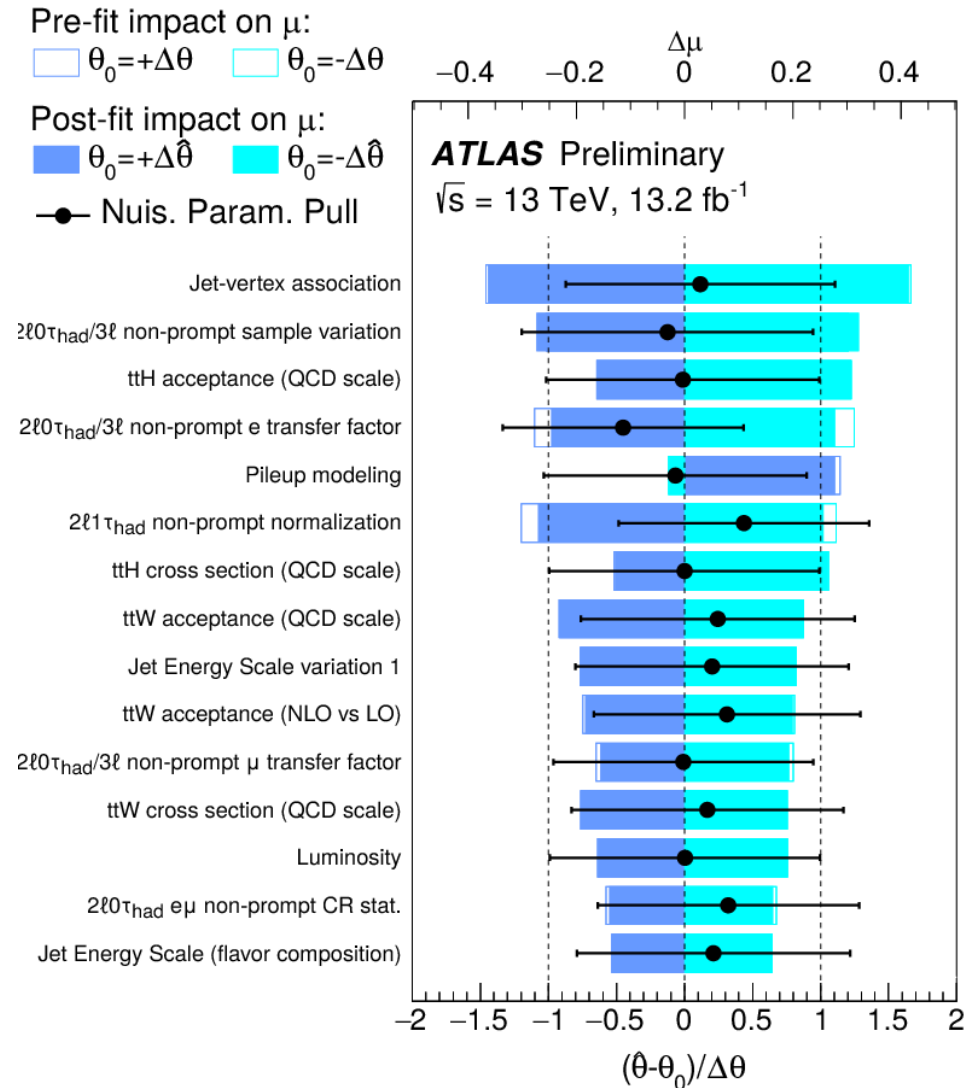
$$(\hat{\theta} - \theta_0) / \sigma_{\theta}$$

Nominally:

- **pull = 0** : i.e. the pre-fit expectation
- **pull uncertainty = 1** : from the Gaussian

However fit results may be different:

- **Central value $\neq 0$** : some data feature differs from MC expectation
 ⇒ Need investigation if large
 - **Uncertainty < 1** : effect is *constrained* by the data ⇒ Needs checking if this legitimate or a modeling issue
- **Impact on result** of $\pm 1\sigma$ shift of NP allows to gauge which NPs matter most .



Pull/Impact plots

Systematics are described by NPs included in the fit. Define **pull** as

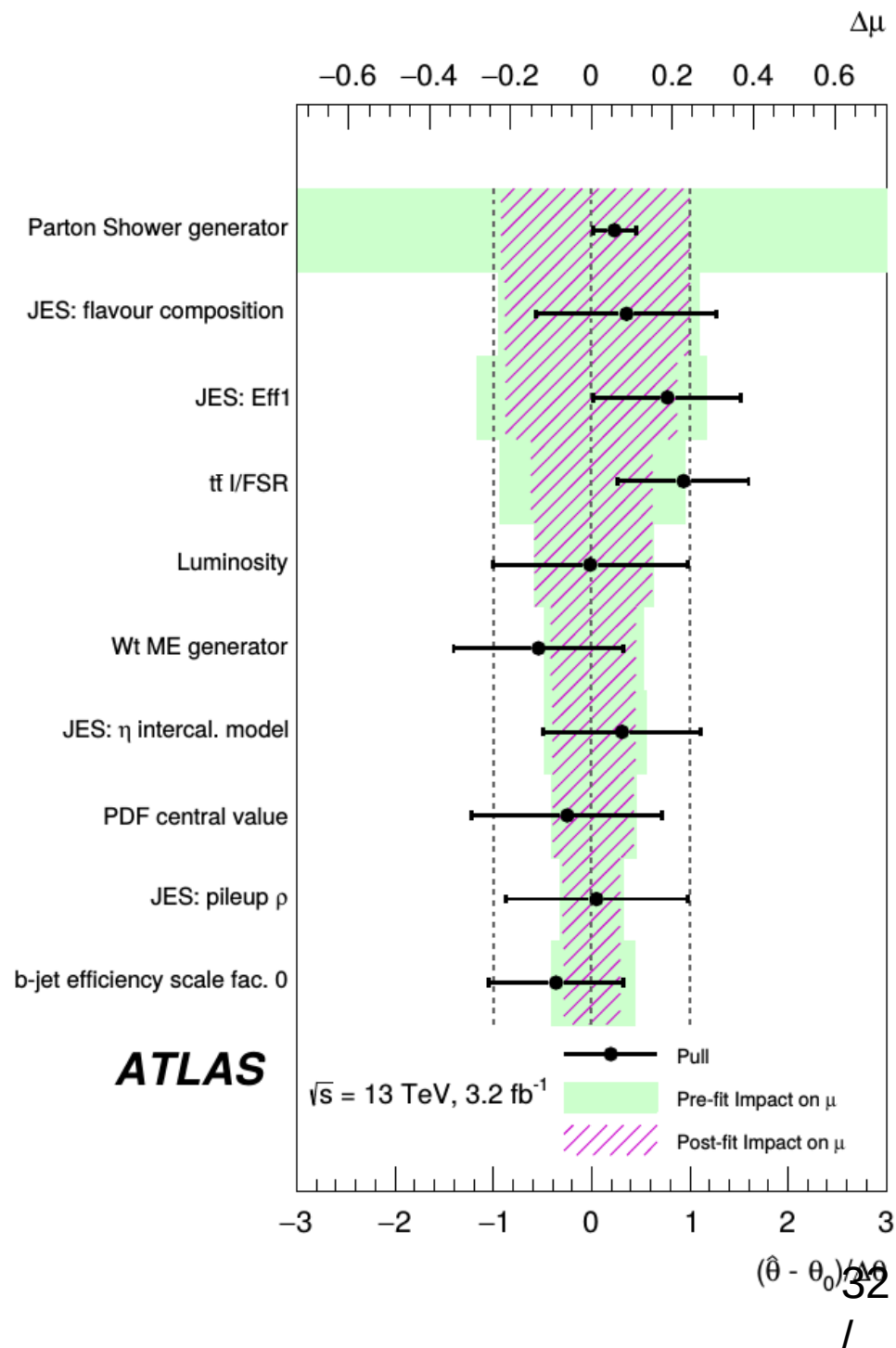
$$(\hat{\theta} - \theta_0) / \sigma_{\theta}$$

Nominally:

- **pull = 0** : i.e. the pre-fit expectation
- **pull uncertainty = 1** : from the Gaussian

However fit results may be different:

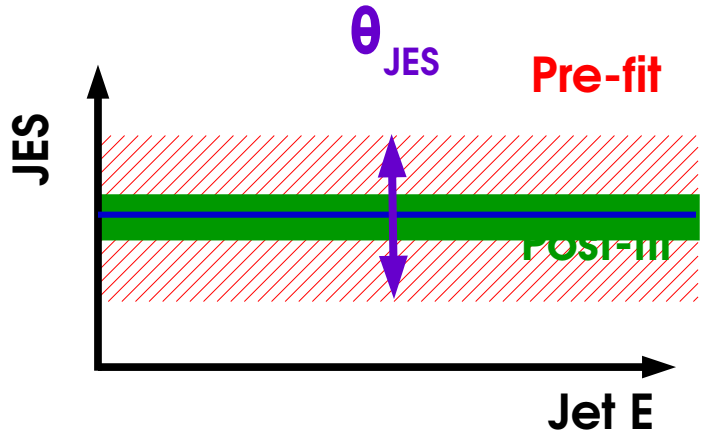
- **Central value $\neq 0$** : some data feature differs from MC expectation
 \Rightarrow Need investigation if large
 - **Uncertainty < 1** : effect is *constrained* by the data \Rightarrow Needs checking if this legitimate or a modeling issue
- \rightarrow **Impact on result** of $\pm 1\sigma$ shift of NP allows to gauge which NPs matter most .



Profiling Issues

Too simple modeling can have unintended effects

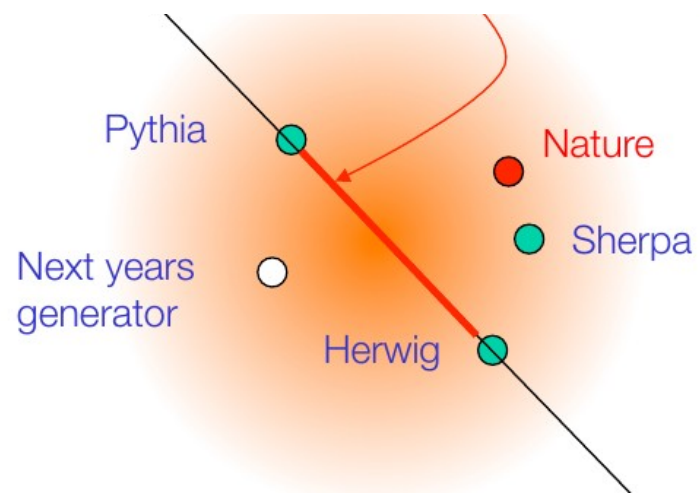
- e.g. single Jet E scale parameter:
- ↳ Low-E jets calibrate high-E jets – intended ?



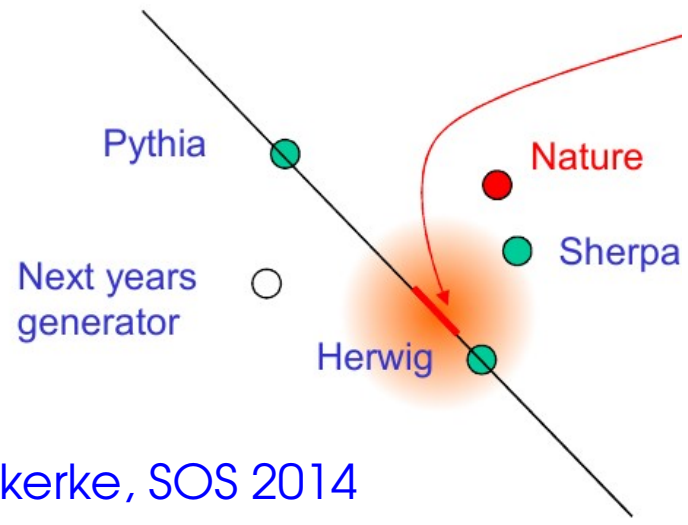
Two-point uncertainties:

→ Interpolation may not cover full configuration space, can lead to too-strong constraints

Pre-fit constraint



Post-fit constraint

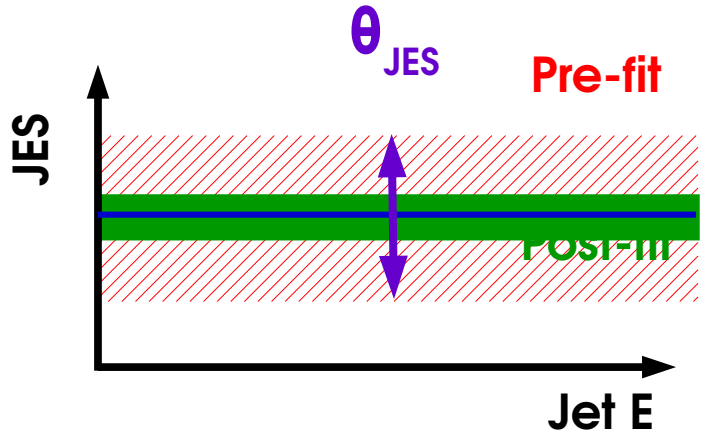


W. Verkerke, SOS 2014

Profiling Issues

Too simple modeling can have unintended effects

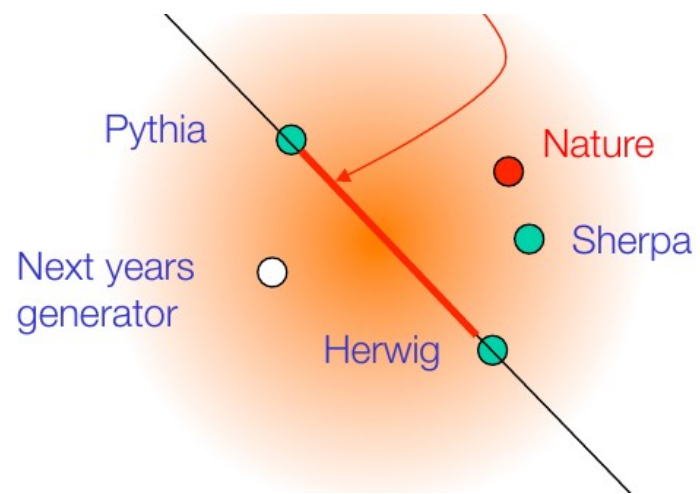
- e.g. single Jet E scale parameter:
- ↳ Low-E jets calibrate high-E jets – intended ?



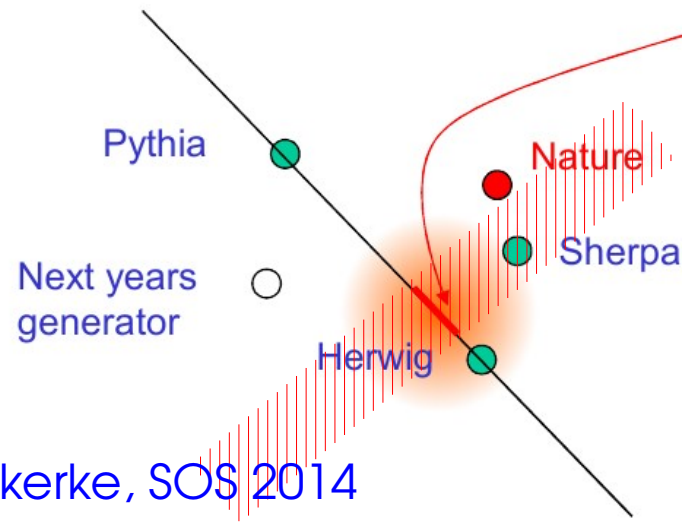
Two-point uncertainties:

- Interpolation may not cover full configuration space, can lead to too-strong constraints

Pre-fit constraint



Post-fit constraint



W. Verkerke, SOS 2014

Profiling Takeaways

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

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

Allows to include systematics as uncertainties on nuisance parameters.

Profiling systematics includes their effect into the total uncertainty.

Gaussian:

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

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**

Other Topics

Generating Pseudo-data

Model describes the distribution of the observable: $P(\text{data}; \text{parameters})$

↳ Possible outcomes of the experiment, for given parameter values

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

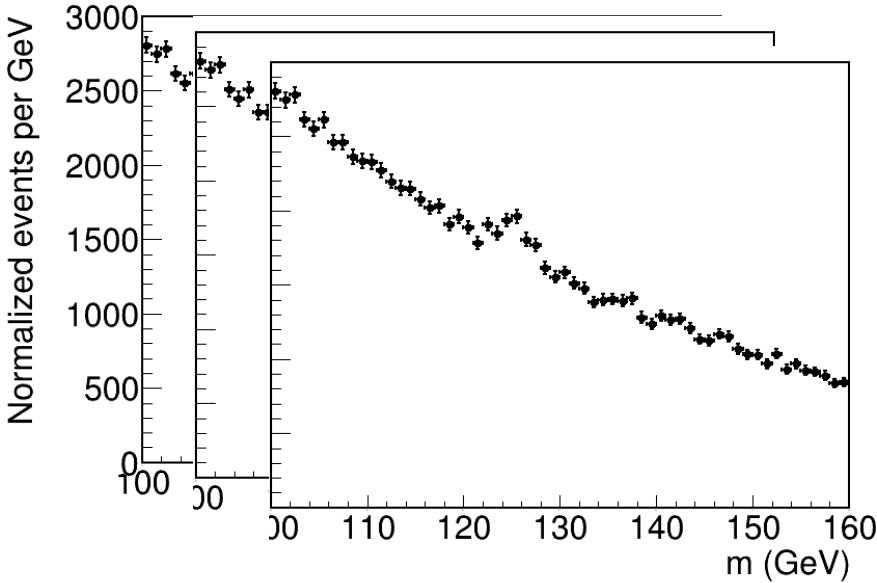
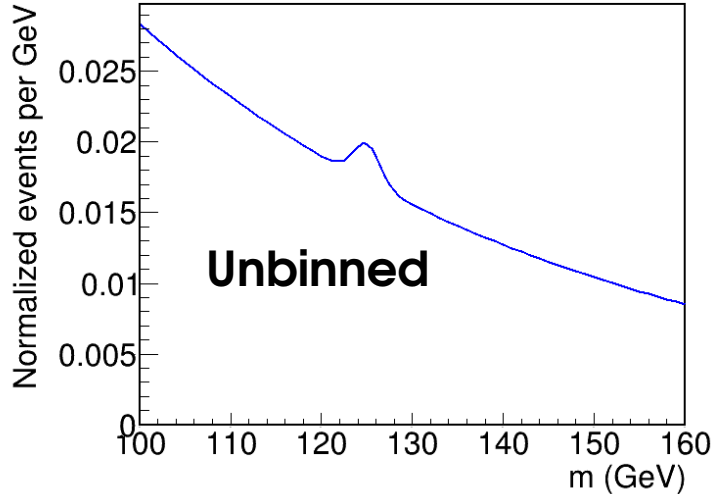
$$P(\lambda=5)$$



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

Each entry = separate "experiment"

Generate



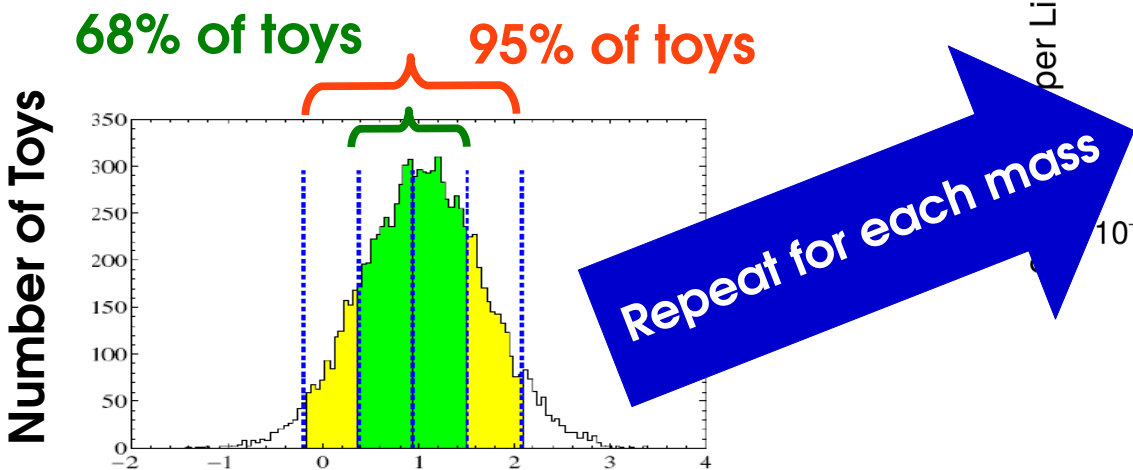
Expected Limits: Toys

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

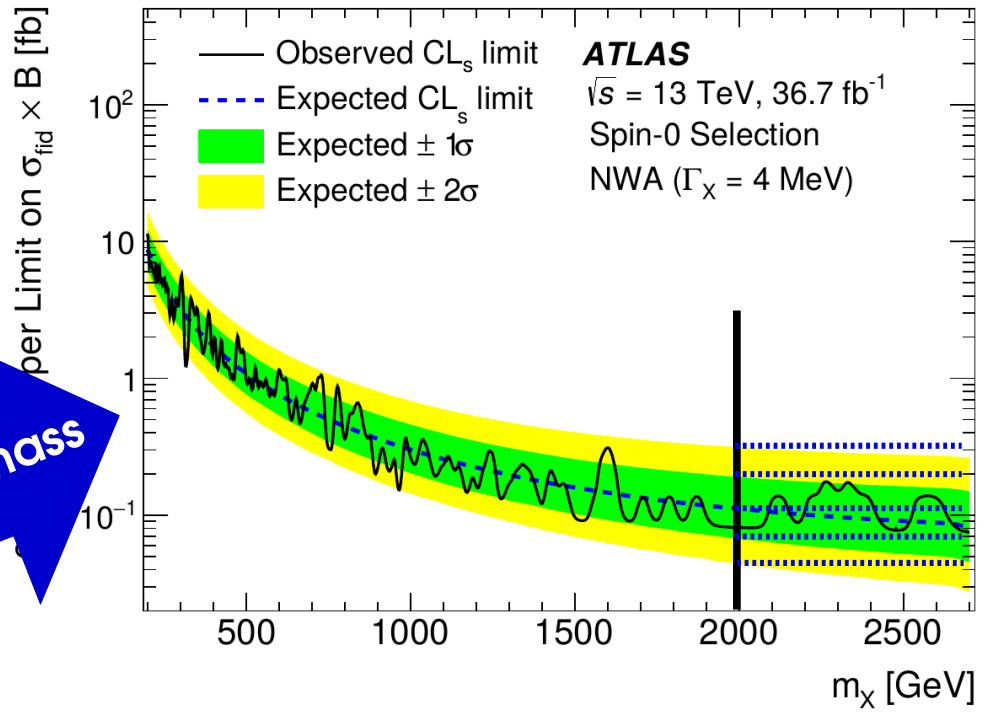
Two main ways to compute:

→ **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



Phys. Lett. B 775 (2017) 105



Expected Limits: Asimov Datasets

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

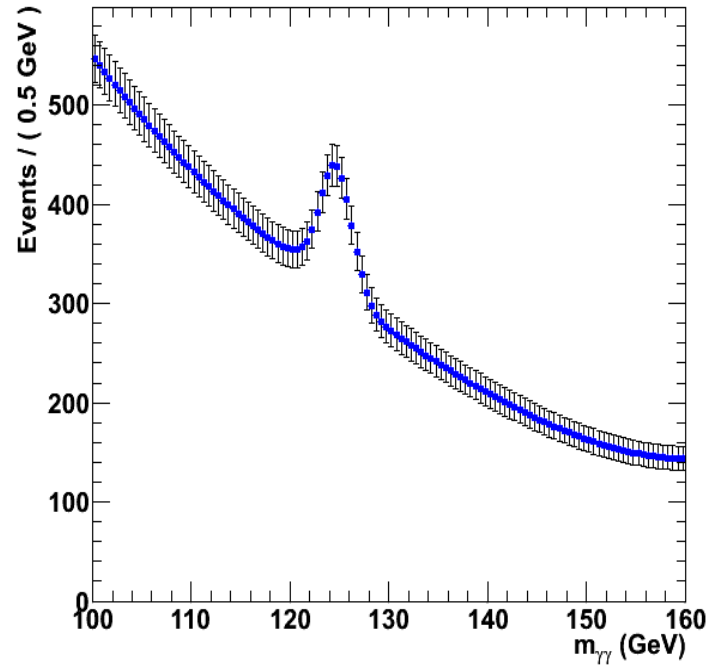
Two main ways to compute:

Strictly speaking, Asimov dataset if
 $\hat{\mathbf{X}} = \mathbf{X}_0$ for all parameters \mathbf{X} ,
 where \mathbf{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:

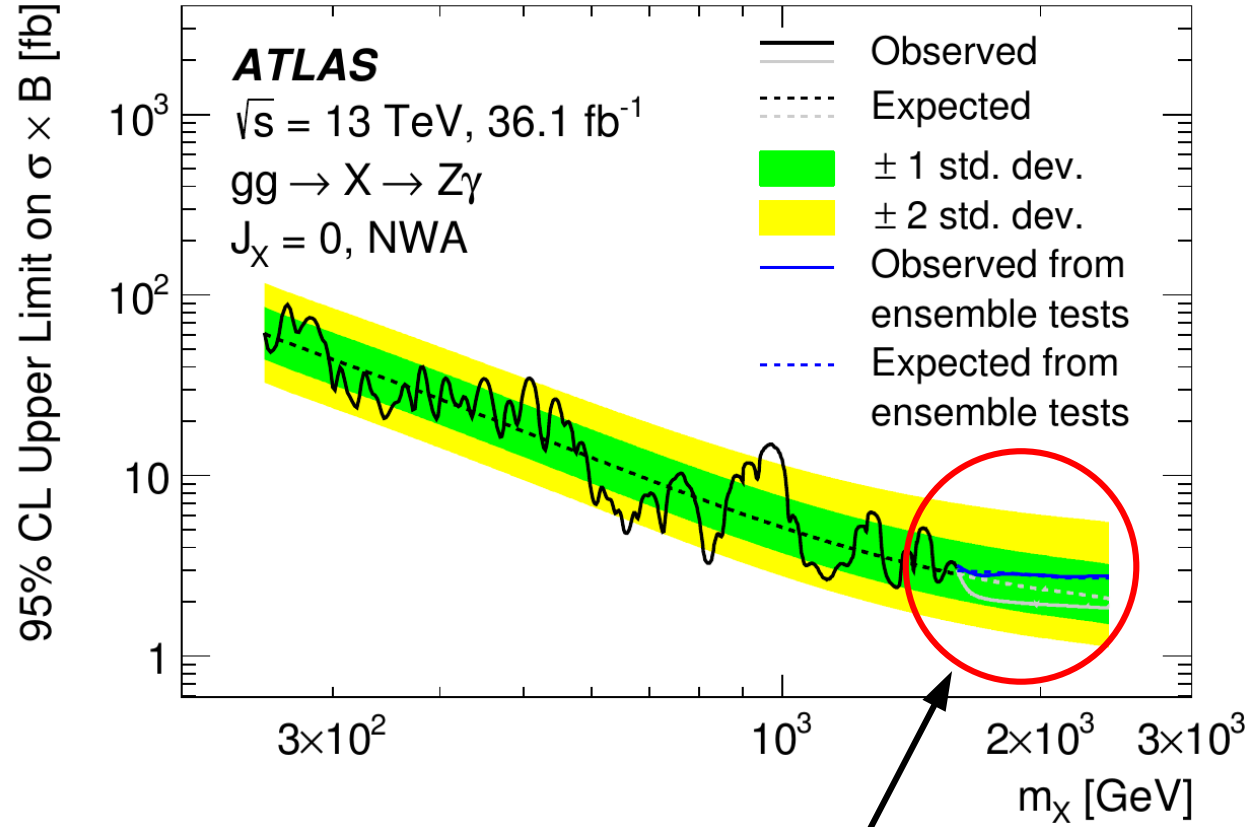
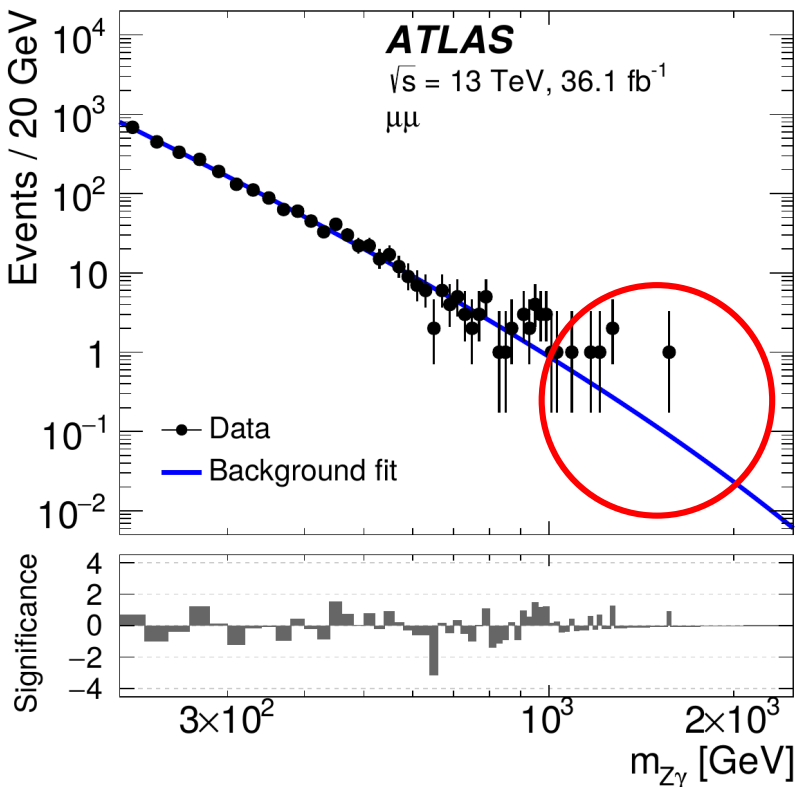
Band width
$$\sigma_{S_0, A}^2 = \frac{S_0^2}{q_{S_0}(\text{Asimov})}$$



⊕ Much faster (1 “toy”)

ATLAS $X \rightarrow Z\gamma$ Search: covers $200 \text{ GeV} < m_X < 2.5 \text{ TeV}$

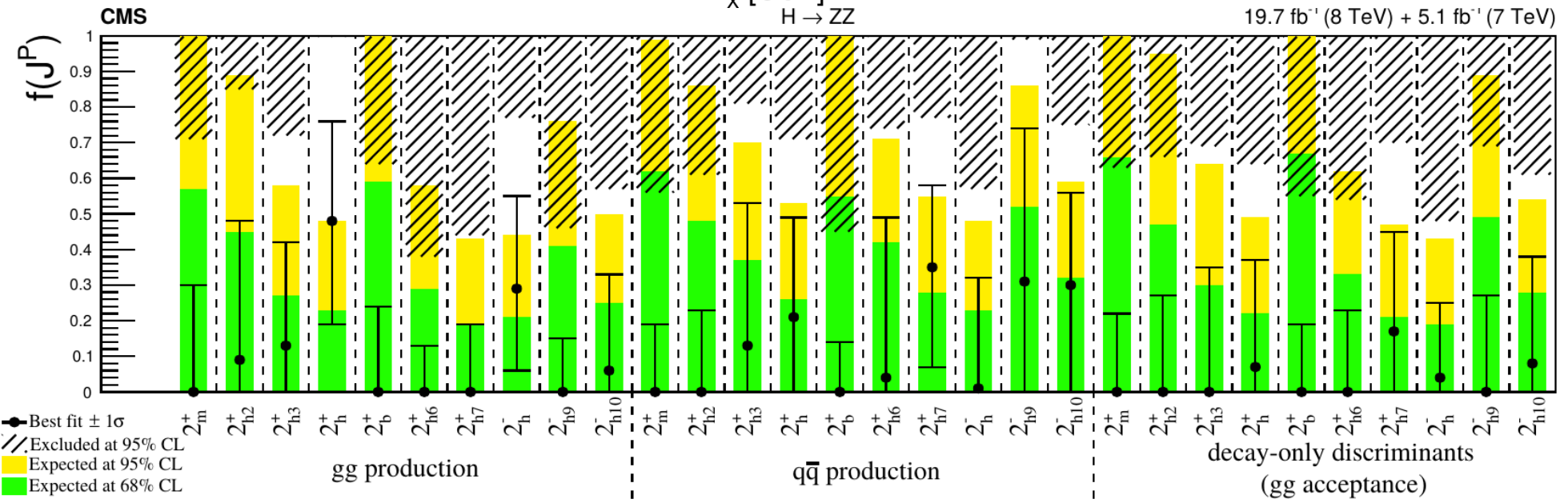
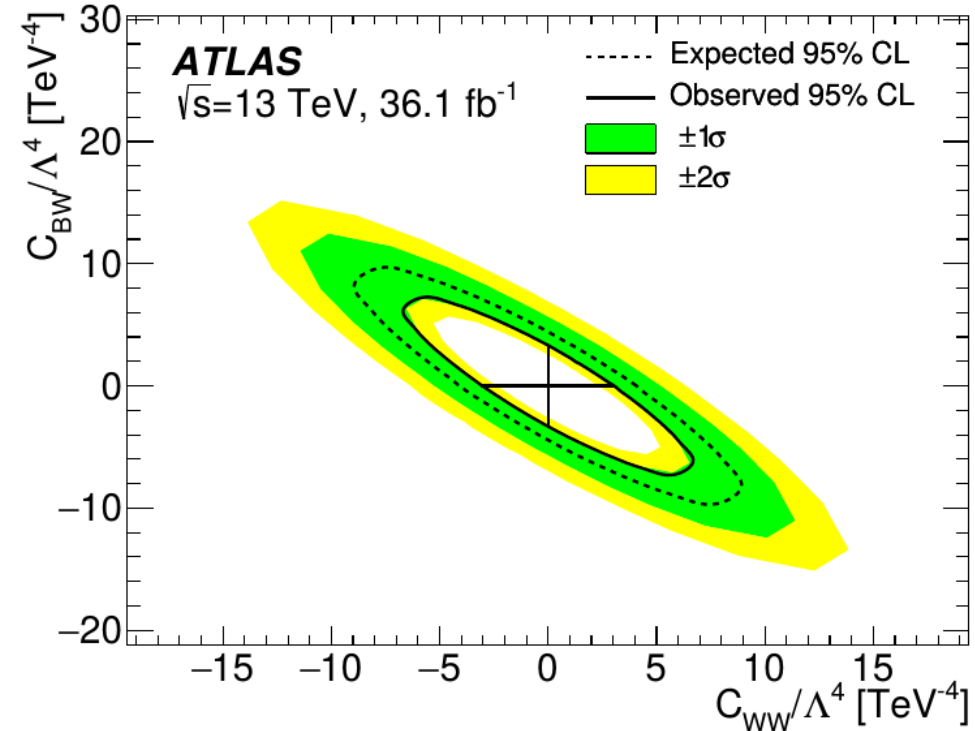
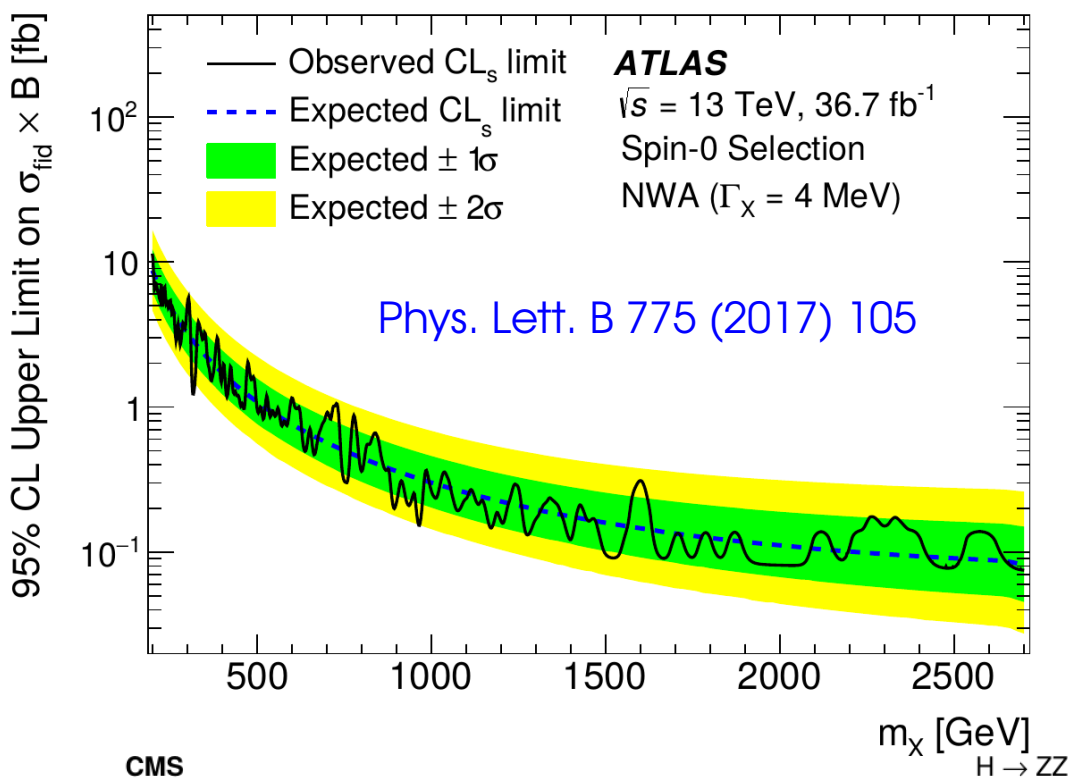
\rightarrow for $m_X > 1.6 \text{ TeV}$, low event counts \Rightarrow derive results from toys



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

Upper Limit Examples

ATLAS 2015-2016 4l aTGC Search



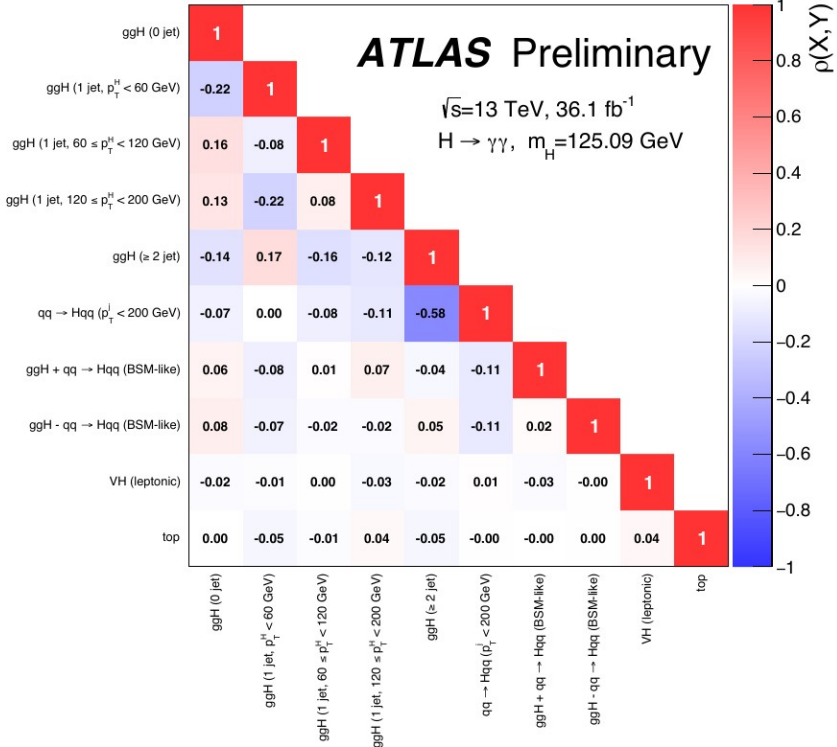
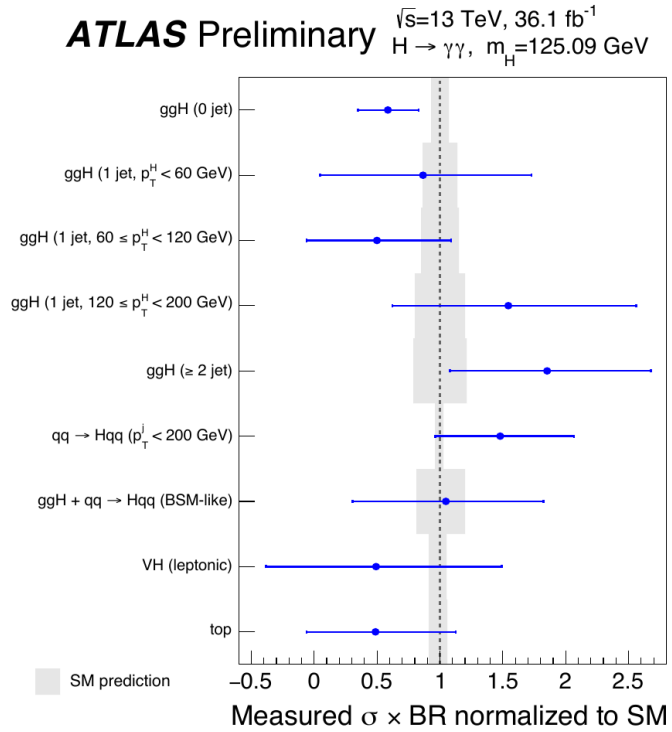
Phys. Rev. D 92 (2015) 012004

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: not so simple, but can publish full likelihood (e.g. [here](#))

Bayesian Analysis

Bayesian methods

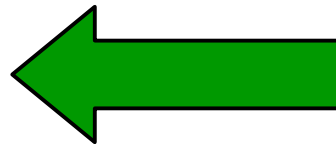
Remember the problem from yesterday:

- PDFs give possible outcomes for known parameters
- We already know the outcome, and want information on the parameters

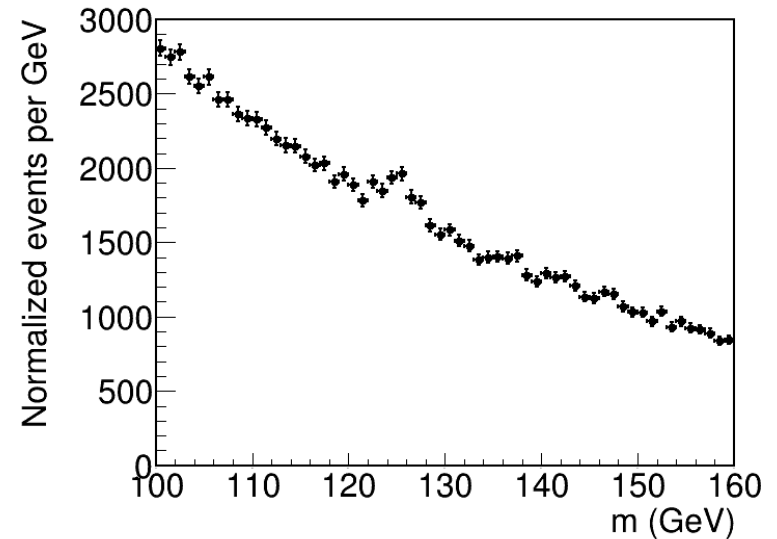
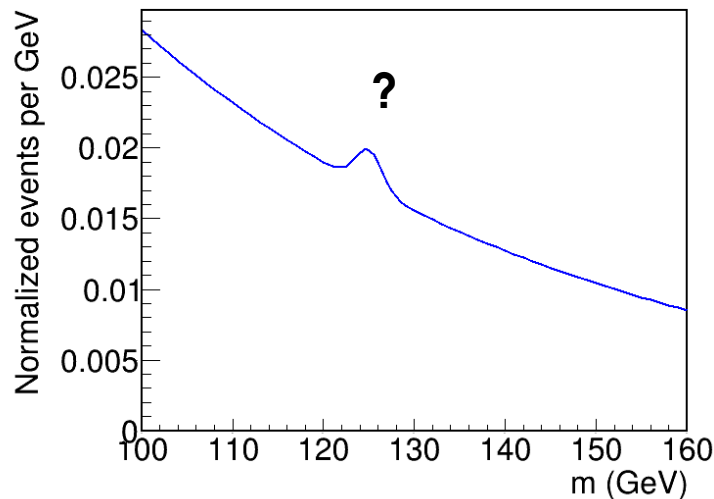
$$P(\lambda = ?)$$



Estimate



2



Solution: maximum likelihood estimation of the parameters, given the data

This is a (good) solution (“classical/frequentist”) but there is another way.

Bayesian methods

Bayesian methods: promote parameters (POIs and NPs) to random variables
→ Represent our best knowledge of their value, not the true values.

Can use **Bayes' Theorem** to obtain a PDF for the parameters

Bayes' Theorem

$$P(\mu | n) = P(n | \mu) \frac{P(\mu)}{P(n)}$$

Posterior PDF: represents our total knowledge from prior + measurement

Measurement PDF, same as for the frequentist $P(n; \mu)$

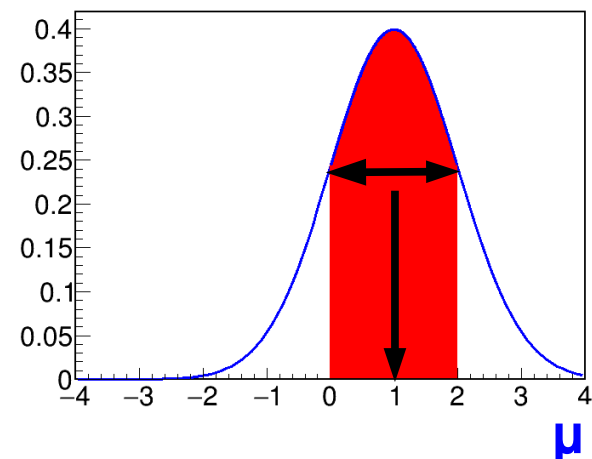
Prior PDF on μ : represents our knowledge before the measurement

Normalization factor: adjusted so $P(\mu | n)$ is normalized to 1)

Immediately useful to get intervals on μ :

- Peak of $P(\mu | n)$ gives the central value : *Maximum a posteriori* (MAP).
- 68.3% interquantile gives the 1σ interval

Problem: what to use for the prior ?...



Bayesian methods

Systematics and nuisance parameters:

Each NP is considered a random variable: Bayes theorem gives $P(\mu, \theta | n)$

Define a prior $\pi(\theta)$ for each nuisance parameter.

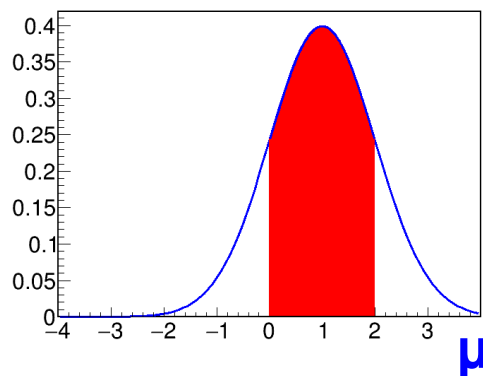
⇒ Obtain $P(\mu | n)$ for μ alone by integrating out the θ :

$$P(\mu | n) = \int P(\mu, \theta | n) C(\theta) d\theta$$

Use probability distribution $P(\mu)$ to compute intervals and limits as before.

68.3% CL interval:

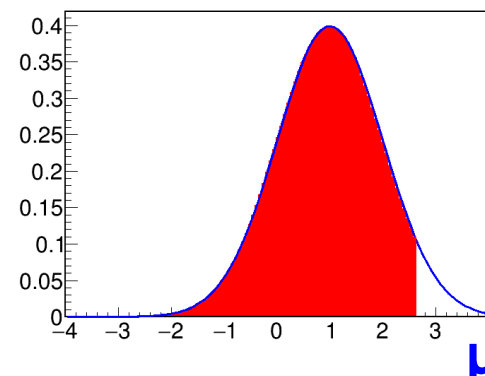
$$\int_A^B P(\mu | n) d\mu = 68.3 \%$$



Here CL means
“Credibility Level”)

95% CL upper limit

$$\int_{-\infty}^L P(\mu | n) d\mu = 95 \%$$



Bayesian vs. frequentist

Many points of commonality

Bayesian analysis typically

- ⊕ Conceptually simpler – frequentist results often difficult to interpret
- ⊖ No simple way to test for discovery
- ⊕ Hybrid methods sometimes used (frequentist discovery + Bayesian systs)

- ⊖ No support for NPs constrained in data
- ⊖ Integration over NPs can be CPU-intensive (but can use MCMC methods)
- ⊕ Minimization over many NPs also not a simple problem for frequentist case...

- ⊖ Need to specify priors, which often contains some arbitrariness – e.g. a prior flat in one parameterization is usually not flat in another.
- ⊕ Can use Jeffreys' or reference priors to avoid this, although difficult in practice.

- ⊕ Frequentist and Bayesian results often agree, so not a big issue in practice!

“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.”

– Louis Lyons

Homework 8: Bayesian methods and CL_s

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

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

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

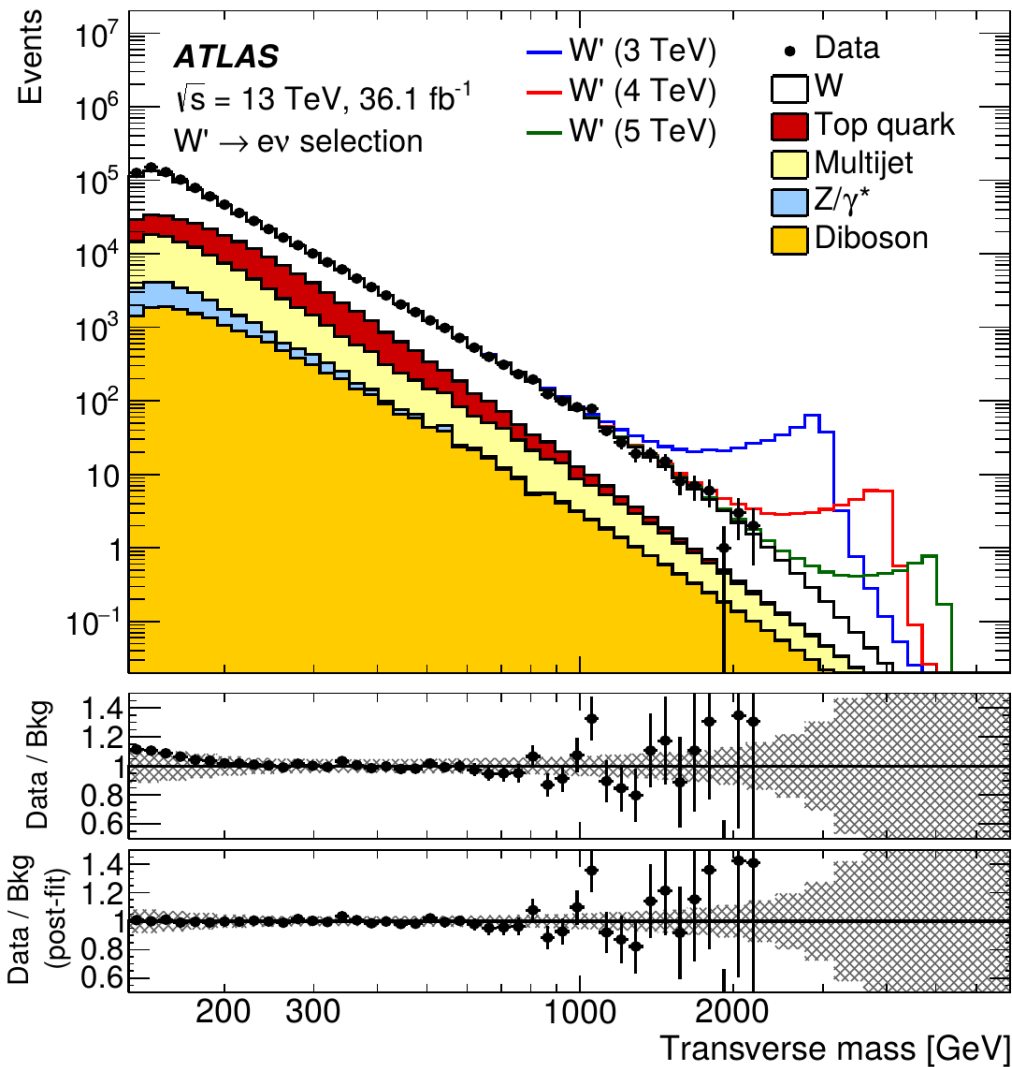
Solution:

In both cases

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

Example: $W' \rightarrow l\nu$ Search

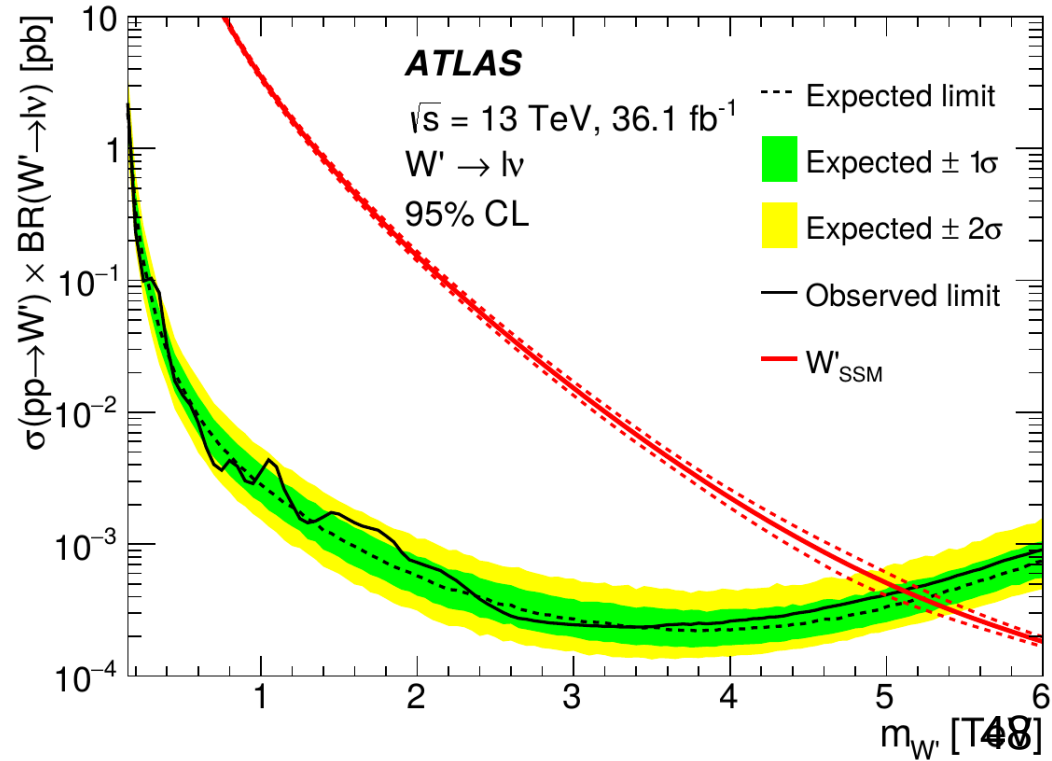
- **POI:** $W' \sigma \times B \rightarrow$ use flat prior over $[0, +\text{inf}]$.
- **NPs:** syst on **signal ϵ** (6 NPs), **bkg** (6), **lumi** (1) \rightarrow integrate over Gaussian priors



Trigger
 Lepton reconstruction and identification
 Lepton momentum scale and resolution
 E_T^{miss} resolution and scale
 Jet energy resolution
 Pile-up

Multijet background
 Top extrapolation
 Diboson extrapolation
 PDF choice for DY
 PDF variation for DY
 EW corrections for DY

Luminosity



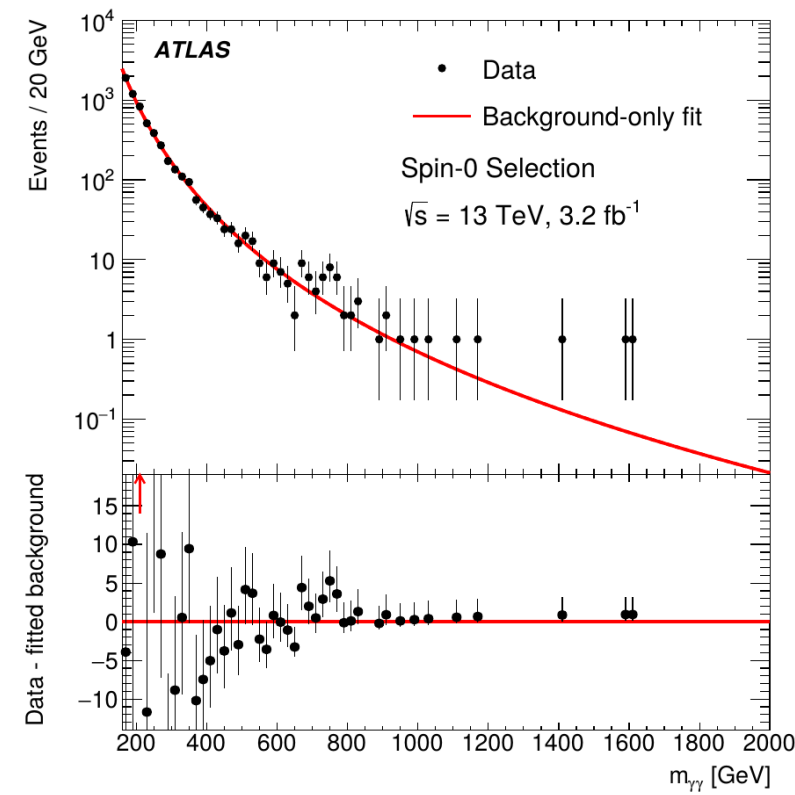
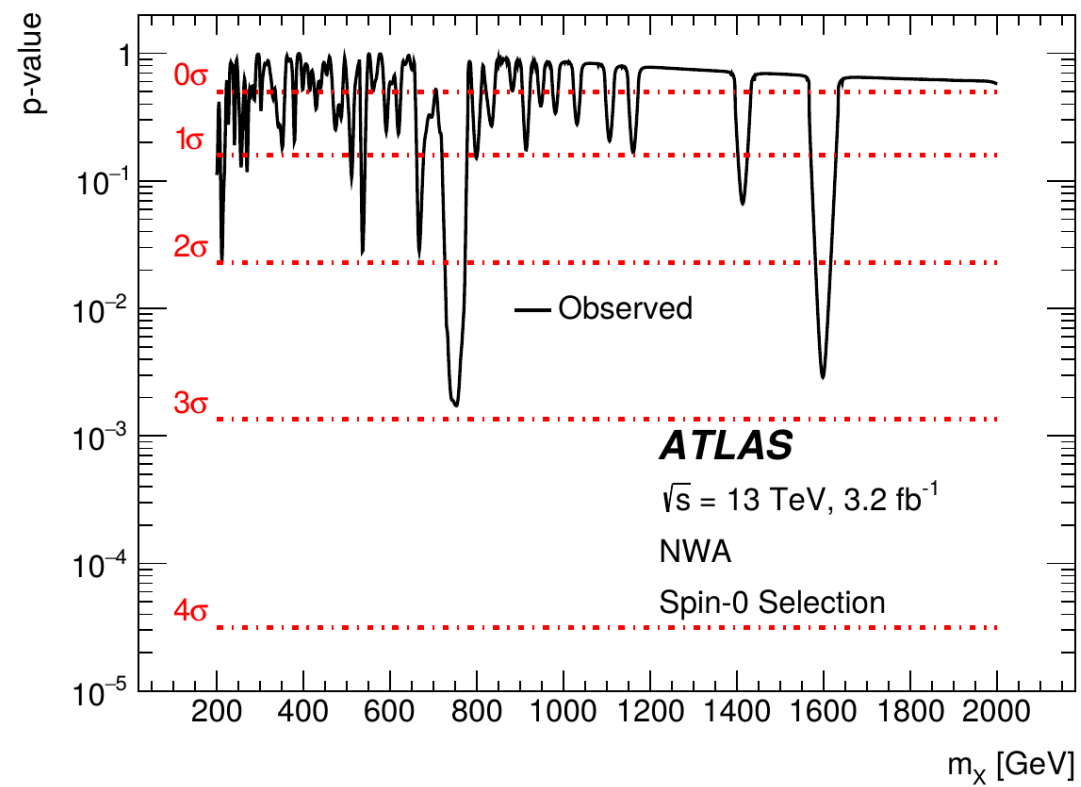
Look-Elsewhere Effect

Look-Elsewhere effect

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

⇒ Effectively: **multiple, simultaneous searches**

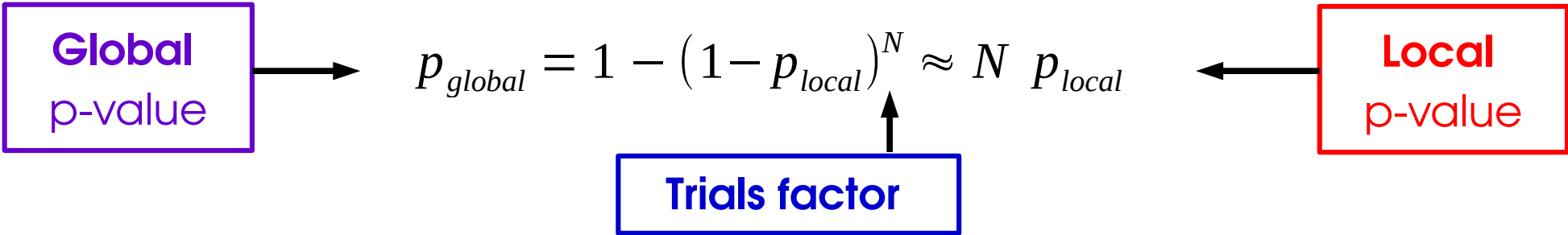
→ If e.g. small resolution and large scan range,
many independent experiments



→ 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



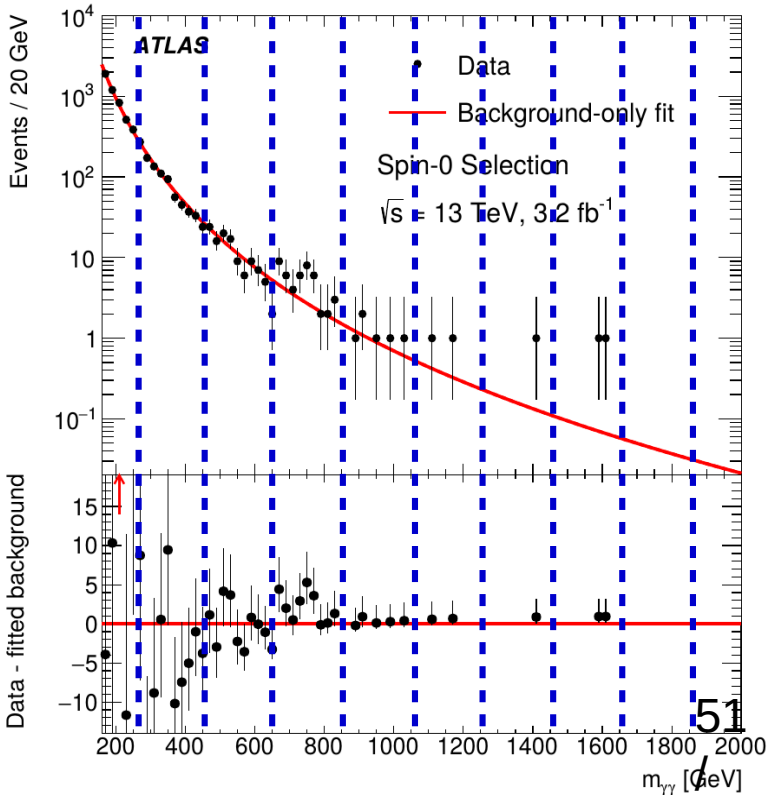
→ $p_{global} > p_{local} \Rightarrow z_{global} < z_{local}$: global fluctuation more likely ⇒ less significant



Trials factor : **naively** = # of independent intervals:

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

However this is usually **wrong** – more on this later



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

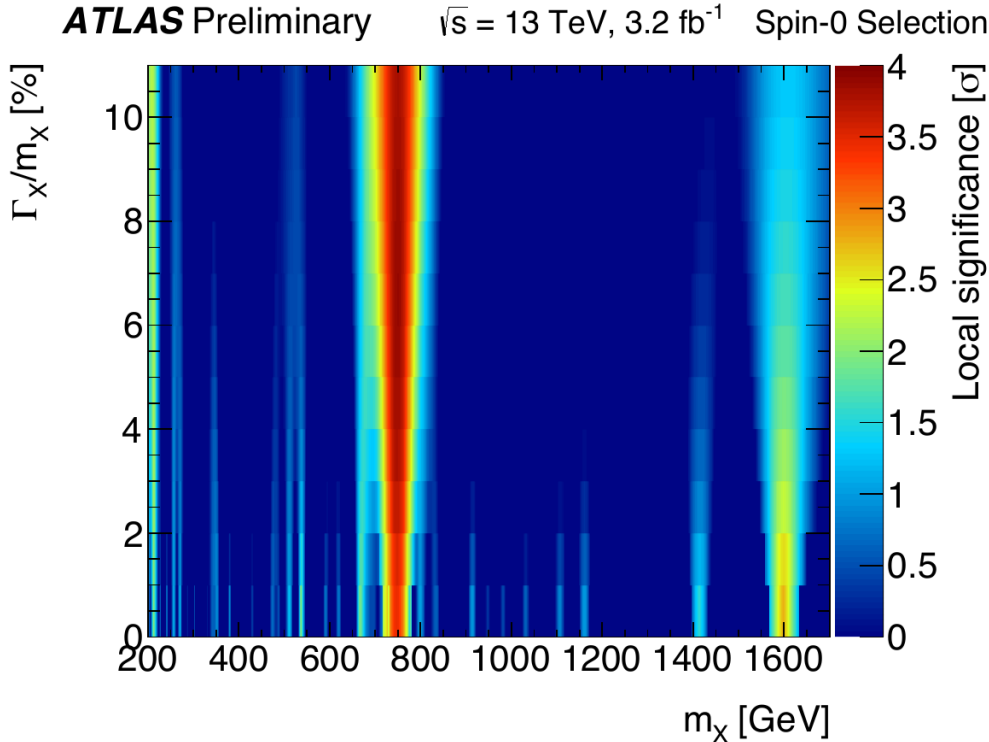
→ Depends on the scanned parameter ranges

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

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

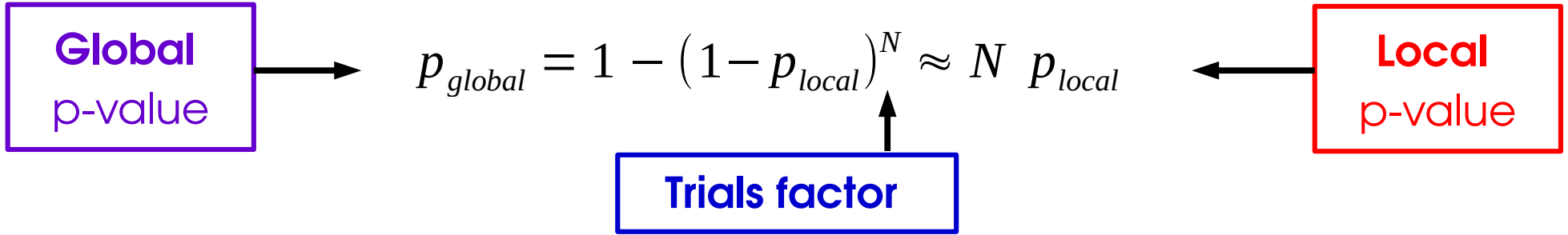
→ p_{local} is what comes out of the usual formulas

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



Trials Factor

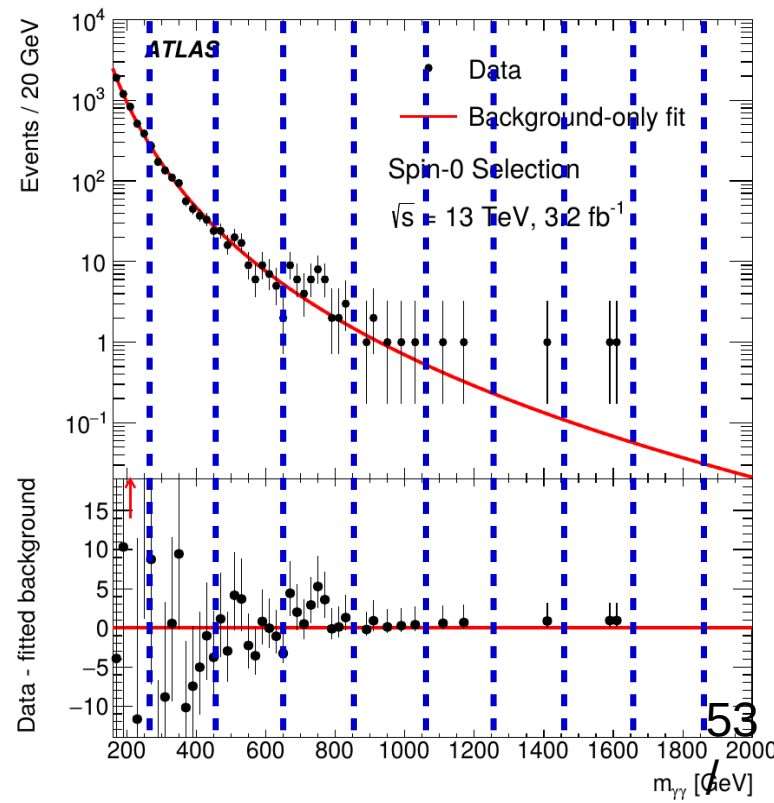
Trials factor $N = \#$ of independent searches:



Naively, one could expect

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

However this is only correct for a discrete Number of experiments (i.e. 10 different regions)



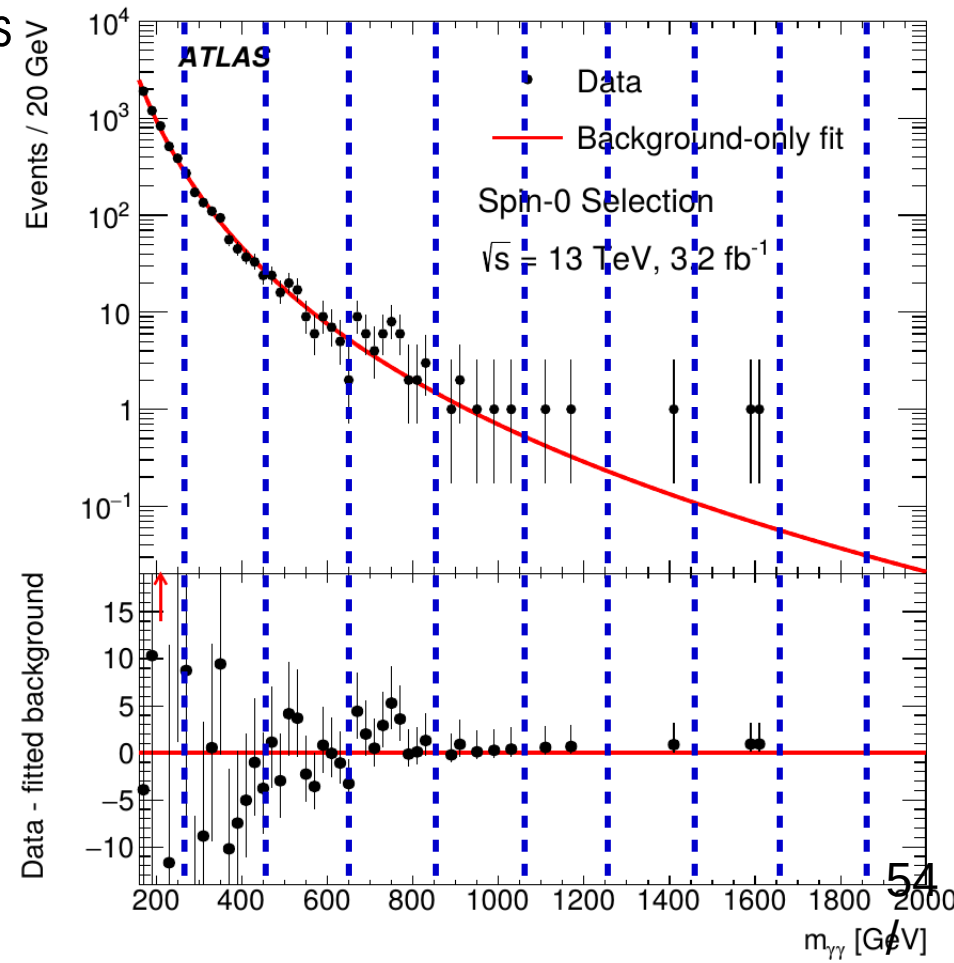
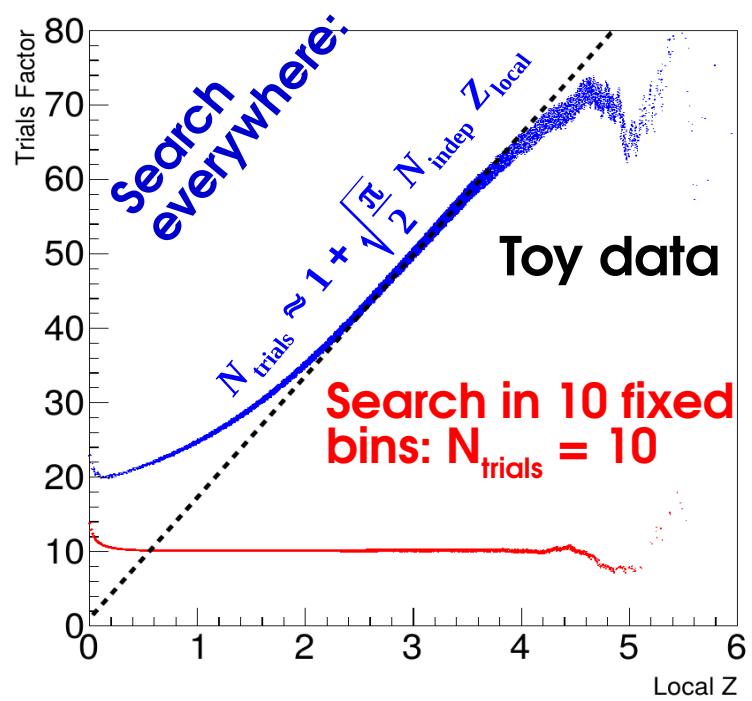
Trials Factor for continuous variables

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

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

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

Why ? Slicing range into N_{indep} regions misses peaks sitting on **edges between regions**
 ⇒ true N_{trials} is **>** N_{indep} !



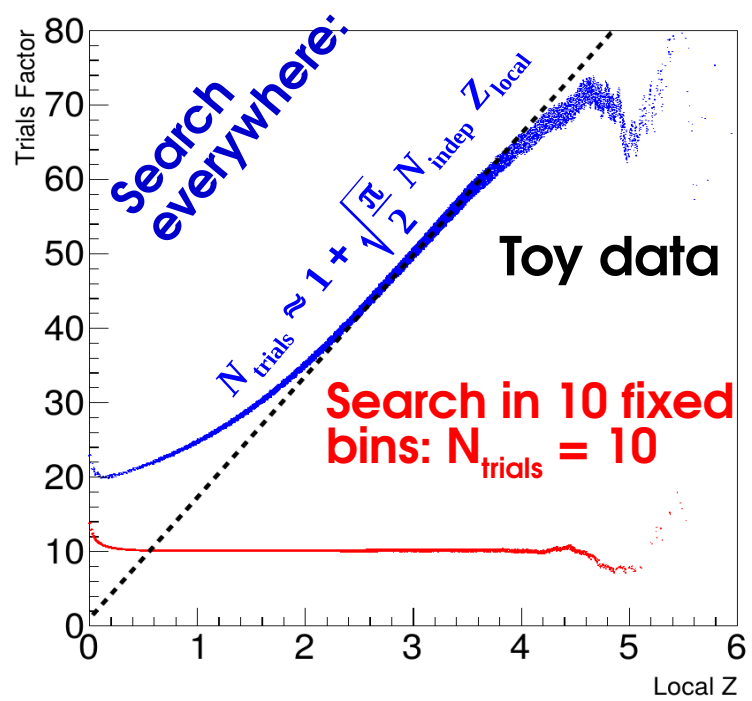
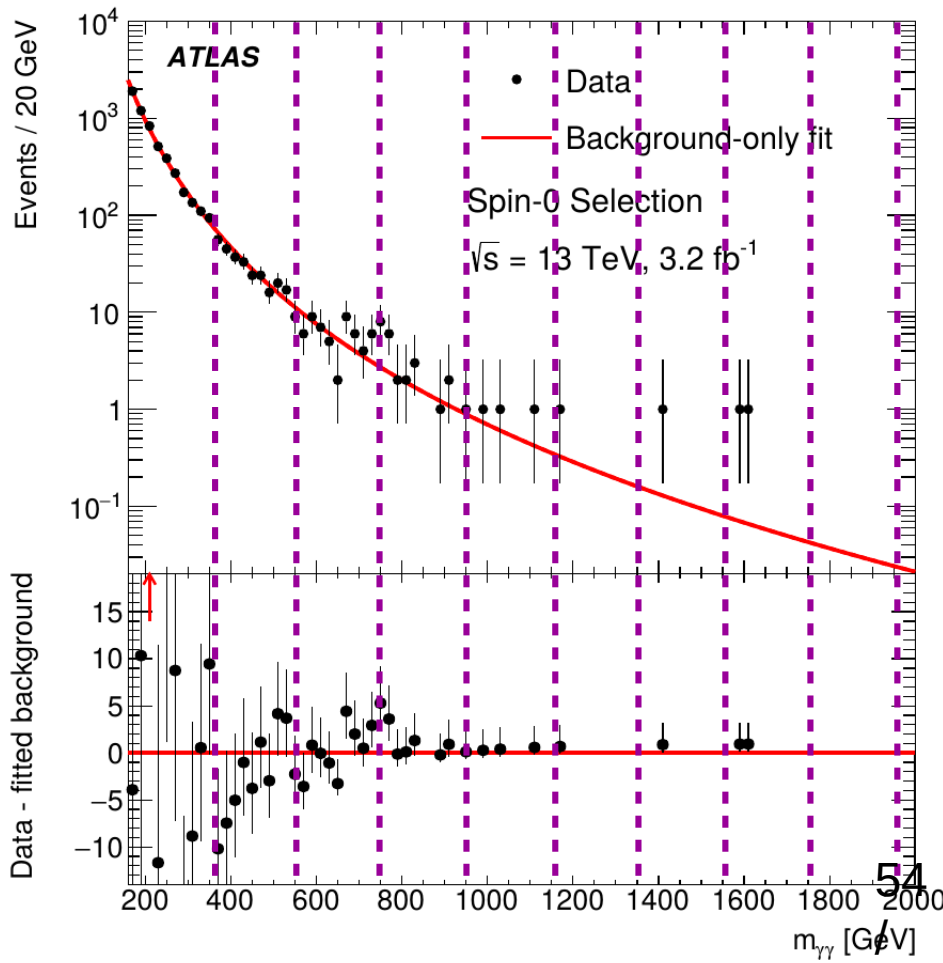
Trials Factor for continuous variables

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

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

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

Why ? Slicing range into N_{indep} regions misses peaks sitting on **edges between regions**
 ⇒ true N_{trials} is **>** N_{indep} !

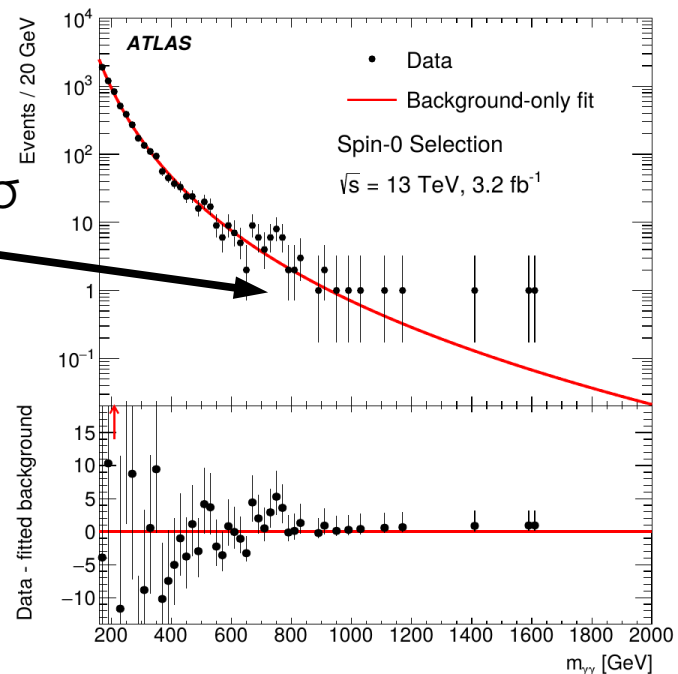


Global Significance from Toys

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

Local 3.9σ



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

e.g. **X → γγ Search:** $Z_{\text{local}} = 3.9\sigma$ ($\Rightarrow p_{\text{local}} \sim 5 \cdot 10^{-5}$),

→ However we are scanning $200 < m_X < 2000 \text{ GeV}$ and $0 < \Gamma_X < 10\% m_X$!

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

⇒ $p_{\text{global}} \sim 2 \cdot 10^{-2}$, $Z_{\text{global}} = 2.1\sigma$ Less exciting, and better indication of true Z!

⊕ **Exact treatment**

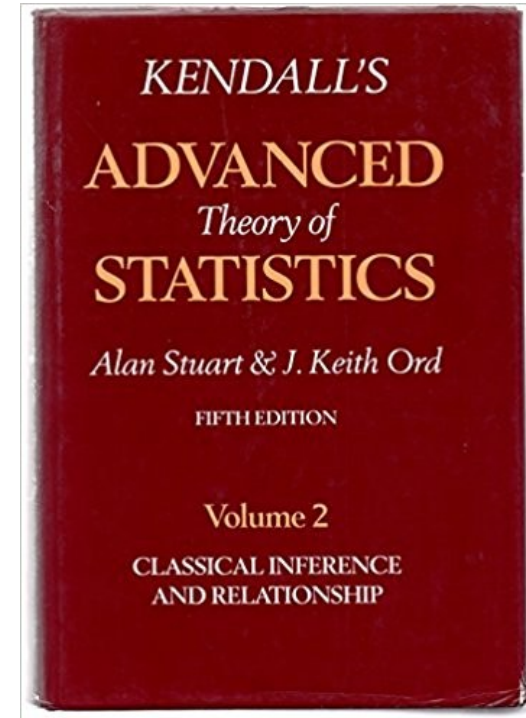
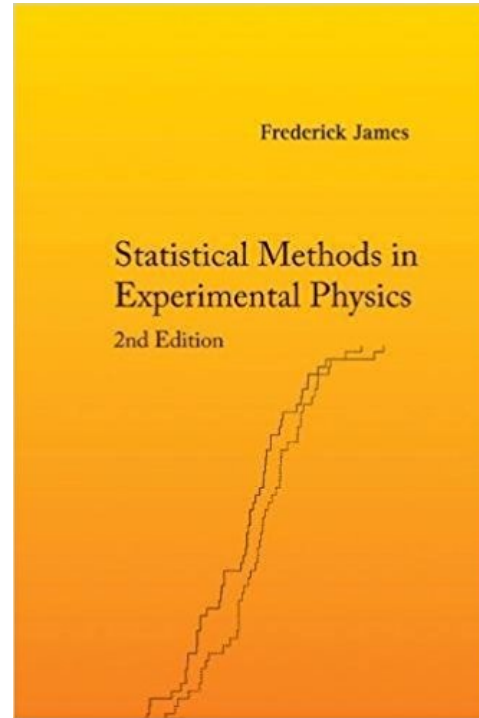
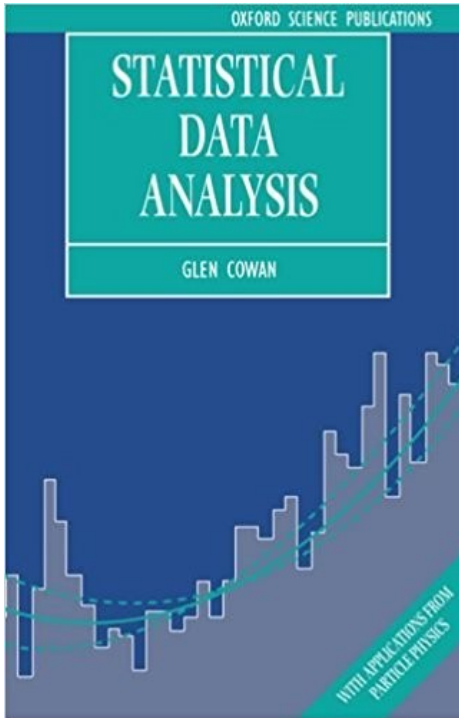
⊖ **CPU-intensive** especially for large Z (need $\sim O(100)/p_{\text{global}}$ toys)

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

Books and Courses

See A. Höcker's [CERN Summer student program statistics course](#)



Some other courses available online:

Glen Cowan's [Cours d'Hiver](#) and [2010 CERN Academic Training lectures](#)

Kyle Cranmer's [CERN Academic Training lectures](#)

Louis Lyons' and Lorenzo Moneta's [CERN Academic Training Lectures](#)

Hands-on session

The hands-on session will be based on **jupyter notebooks** built using the **numpy/scipy/pyplot** stack.

If you have a computer with you, **please install anaconda** as this provides a consistent installation of python, JupyterLab, etc.

→ Alternatively, you can also install **JupyterLab** as a standalone package.

→ Another solution is to run the notebooks on the public jupyter servers at **mybinder.org**. This will probably be slower but avoids a local install.



Lecture 1	Lecture Notes	notebook [solutions]	binder [solutions]
Lecture 2	Lecture Notes	notebook [solutions]	binder [solutions]
Lecture 3	Lecture Notes	notebook [solutions]	binder [solutions]

The **warmup** item includes material that will not be covered in detail in the class, as well as an introduction to the notebooks. **Please have a look before the beginning of the classes** if you are unfamiliar with any of this.

Homework Solutions

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

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

Likelihood :

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

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

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

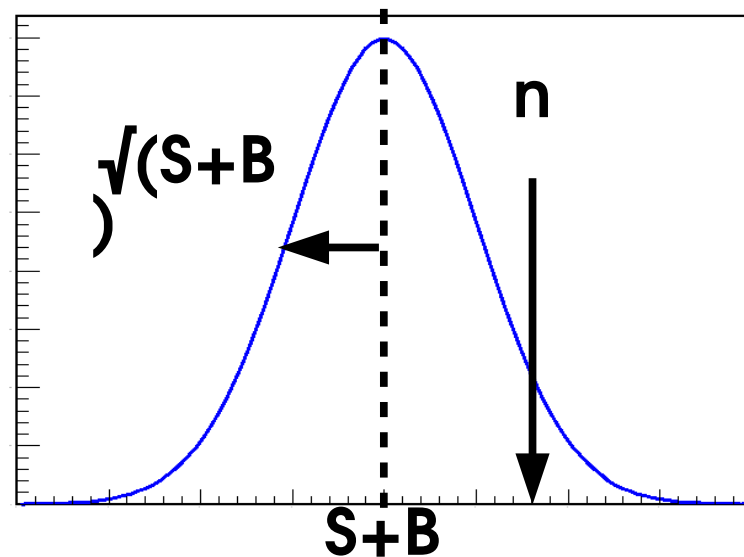
$$q_0 = -2 \log \frac{L(S=0)}{L(\hat{S})} = \lambda(S=0) - \lambda(\hat{S}) = \left(\frac{n-B}{\sqrt{B}} \right)^2 = \left(\frac{\hat{S}}{\sqrt{B}} \right)^2$$

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:

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

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 ,

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

Homework 4 : Gaussian CL_{s+b}

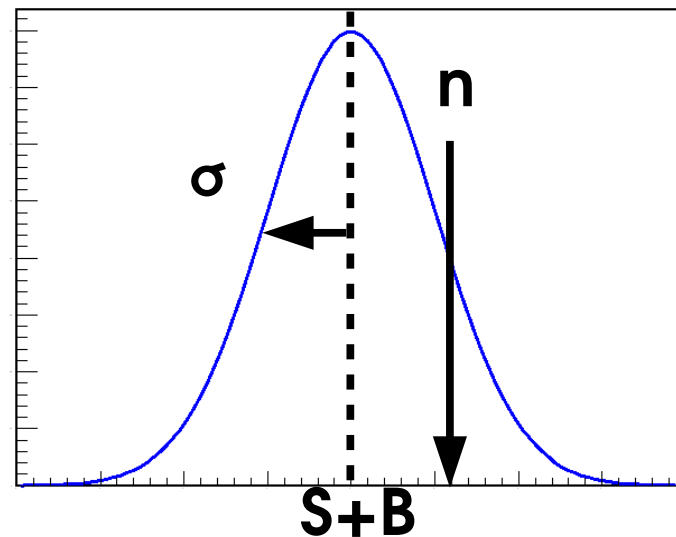
Usual Gaussian counting example with known B:

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

Reminder:

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

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



Compute the 95% CL upper limit on S:

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

$$\text{so } q_{S_0} = 2.70 \quad \text{for } S_0 = \hat{S} + \sqrt{2.70} \sigma_s$$

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

Homework 5 : 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: $S_{\text{up}} = \hat{S} + 1.64 \sigma_S$ at 95 % CL

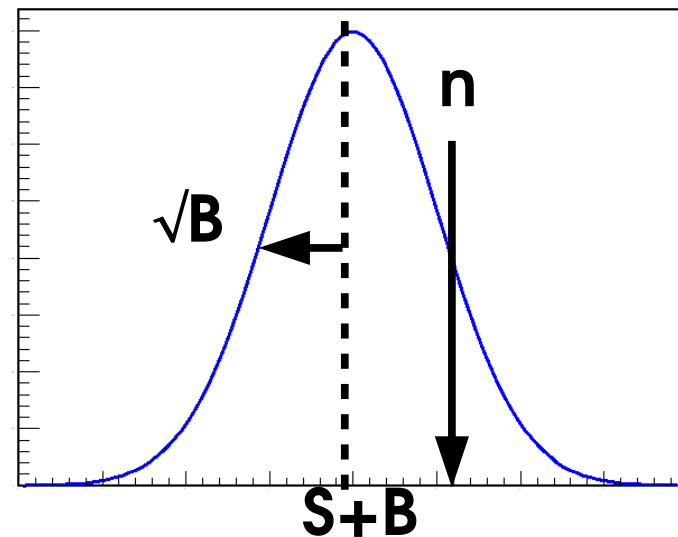
CL_s upper limit : still have $q_{S_0} = \left(\frac{S_0 - \hat{S}}{\sigma_S} \right)^2$ (for $S_0 > \hat{S}$)

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 $\hat{S} = 0$,

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



$\hat{S} \sim G(S, \sigma_S)$ 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 6 : Poisson CL_s

Same exercise, for the Poisson case

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

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

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

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

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

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

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

⇒ $S_{up} \sim 2$, exact value depends on B

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

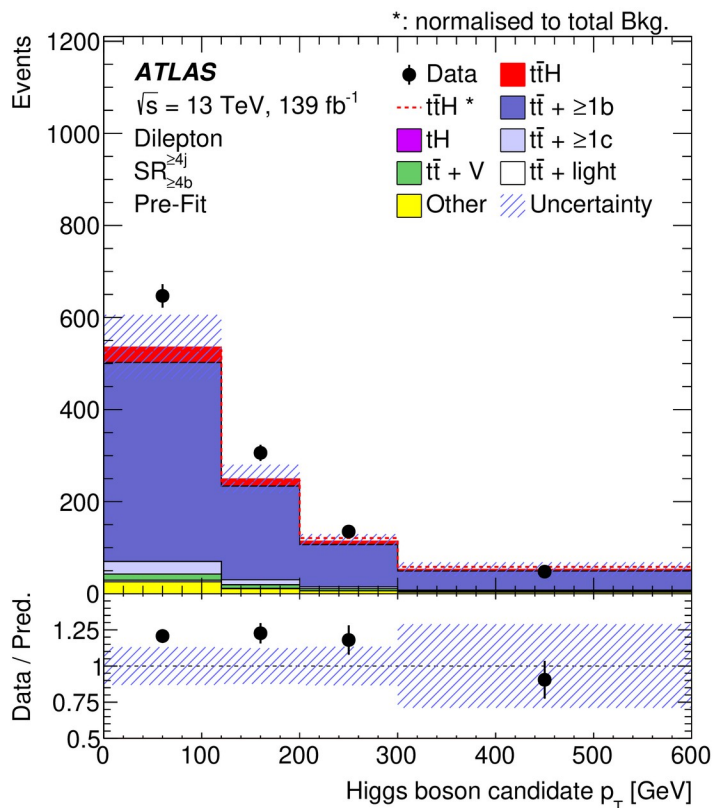
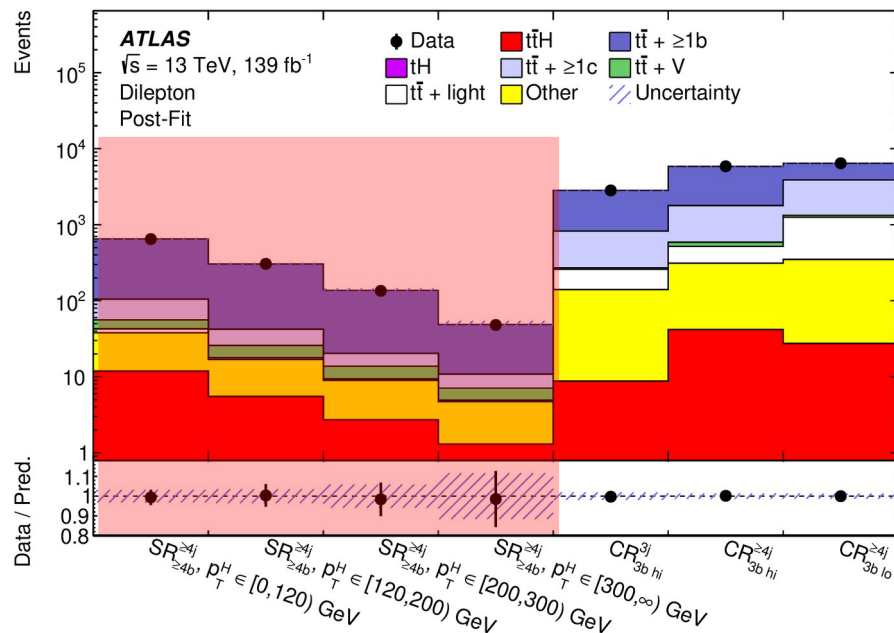
Extra Slides

Multiple analysis regions often used.

→ Exploit better sensitivity in some regions

Here ($t\bar{t}H$, $H \rightarrow bb$ analysis) **7** regions:

→ **4** Signal Regions (**SR**) split in p_T (Higgs)



Better sensitivity at high p_T

→ lower B backgrounds, higher S/B

Backgrounds levels from simulation here

→ Large systematic uncertainties!

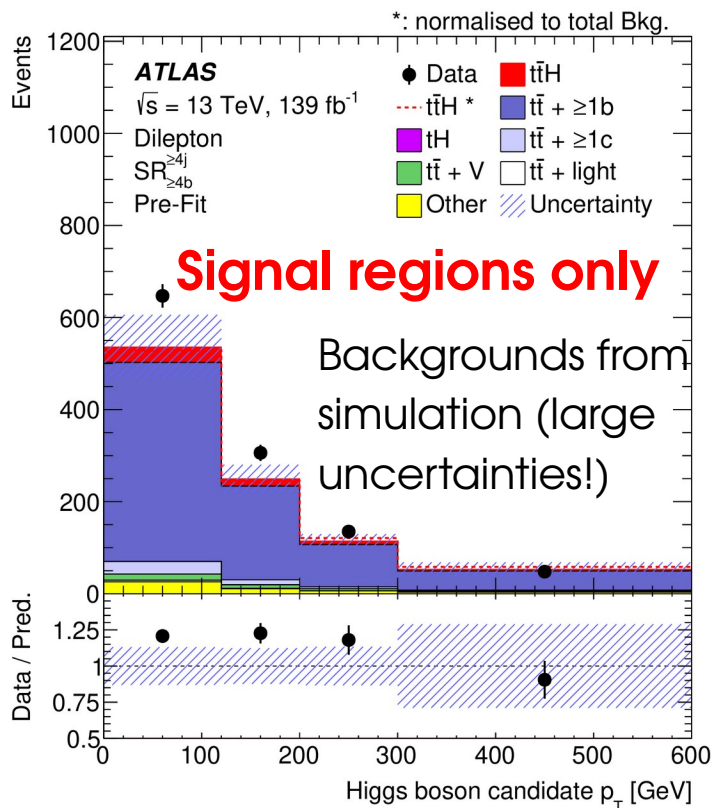
Categories

Multiple analysis regions often used.

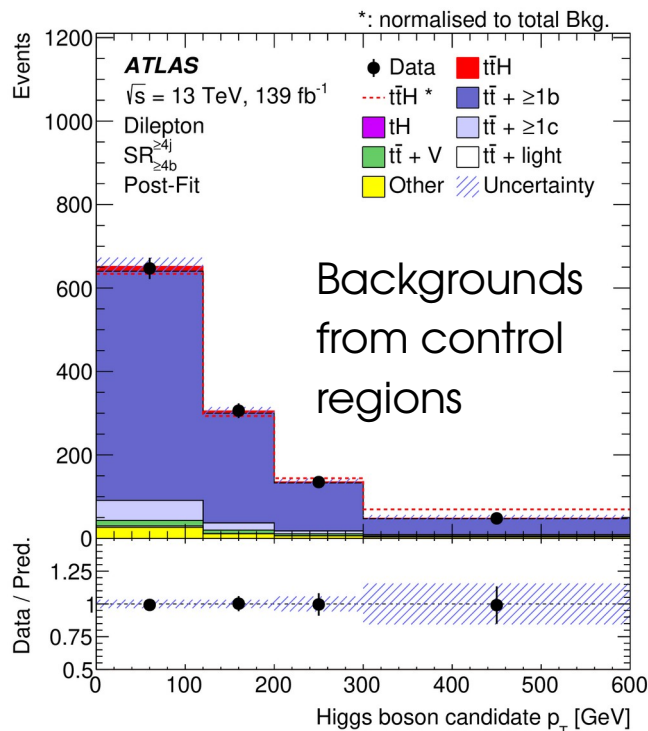
- Exploit better sensitivity in some regions
- Constrain NPs: **Control regions** for bkg

Here (ttH, H→bb analysis) **7** regions:

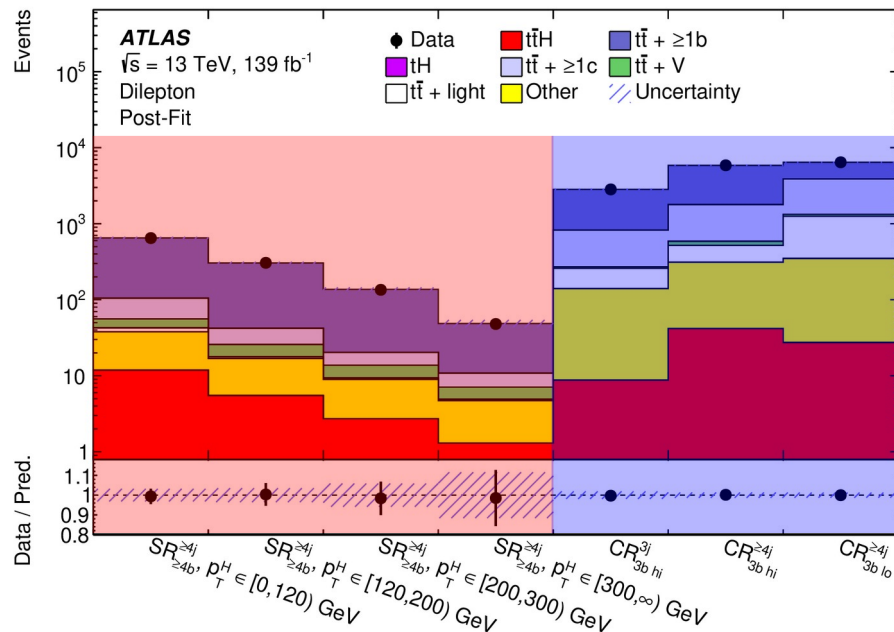
- **4 Signal Regions (SR)** split in p_T (Higgs)
- **3 Background Control Regions (CR)**



Include Background CRs



Signal + Bkg regions



Multiple analysis regions often used.

- Exploit better sensitivity in some regions
- Constrain NPs: **Control regions** for bkg

Here (ttH, H→bb analysis) **7** regions:

- **4 Signal Regions (SR)** split in p_T(Higgs)
- **3 Background Control Regions (CR)**

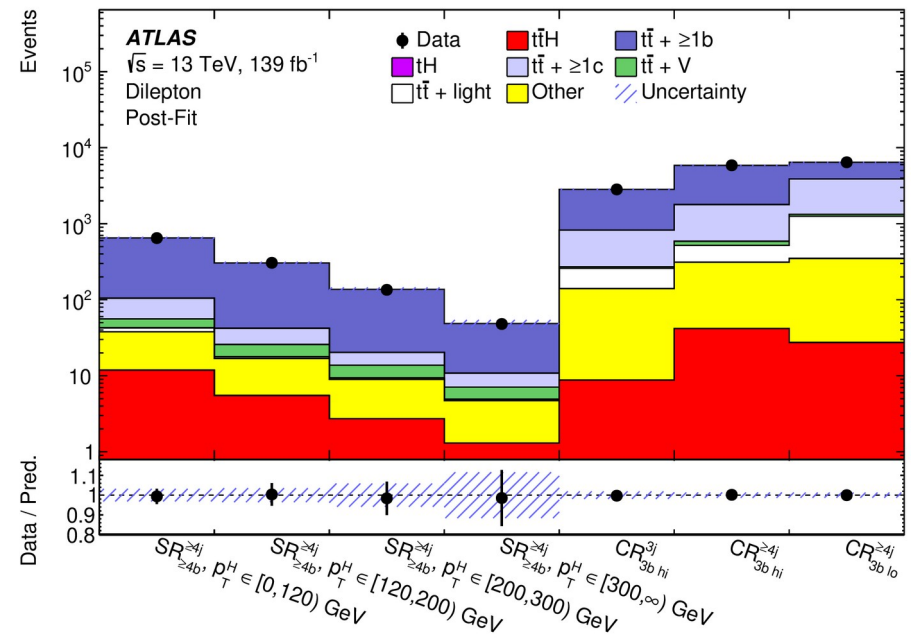
⇒ **Combined PDF** :

PDF for category k

$$P(S, B; \{n_i^{(k)}\}_{i=1 \dots n_{evts}^{(k)}}^{k=1 \dots n_{cats}}) = \prod_{k=1}^{n_{cats}} P_k(S, B; \{n_i^{(k)}\}_{i=1 \dots n_{evts}^{(k)}})$$

No overlaps between categories ⇒ No statistical correlations
 ⇒ can simply take product of individual PDFs.

Multiple categories allows to **constrain nuisance parameters** (e.g. **B**)

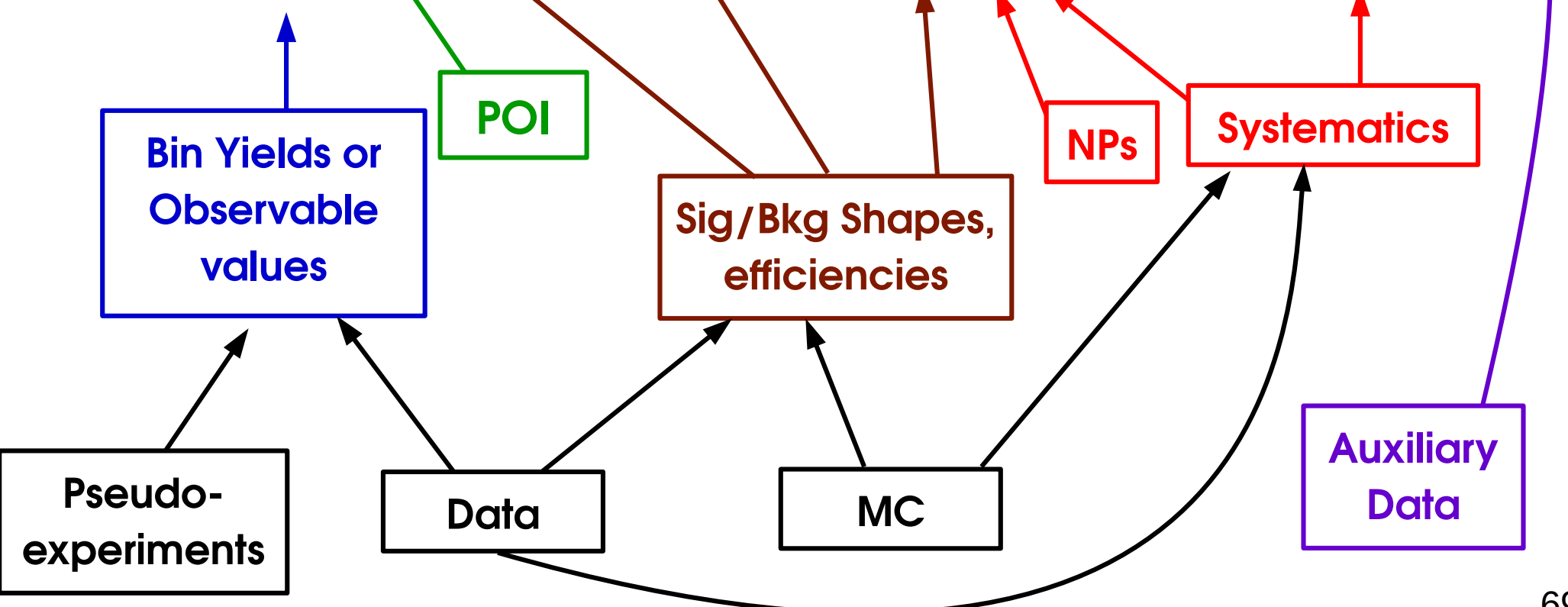


Counting model, the full version

$$P(\boldsymbol{\mu}, \{\boldsymbol{\theta}_j\}_{j=1 \dots n_{NP}}; \{n_i^{(k)}\}_{i=1 \dots n_{data}^{(k)}}, \{\boldsymbol{\theta}_j^{obs}\}_{j=1 \dots n_{NP}}) =$$

Expected bin yield

$$\prod_{k=1}^{n_{cats}} P[n_i; \boldsymbol{\mu} \epsilon_{i,k}(\vec{\theta}) N_{S,i,k}(\vec{\theta}) + B_{i,k}(\vec{\theta})] \prod_{j=1}^{n_{syst}} G(\boldsymbol{\theta}_j^{obs}; \boldsymbol{\theta}_j; 1)$$

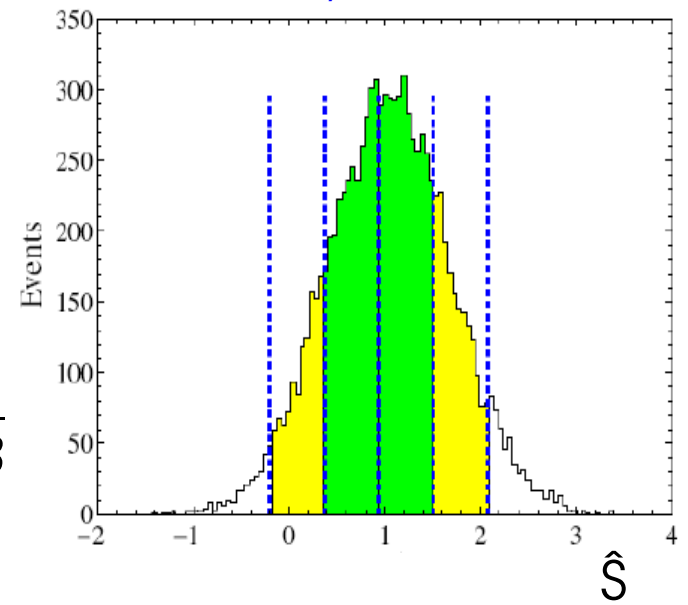


x number of categories!

CL_s : Gaussian Bands

Usual Gaussian counting example with known B:
 95% CL_s upper limit on S:

$$S_{\text{up}} = \hat{S} + \left[\Phi^{-1} \left(1 - 0.05 \Phi \left(\hat{S} / \sigma_S \right) \right) \right] \sigma_S \quad \text{with} \quad \sigma_S = \sqrt{B}$$



Compute expected bands for S=0:

→ **Asimov dataset** $\Leftrightarrow \hat{S} = 0$: $S_{\text{up,exp}}^0 = 1.96 \sigma_S$

→ **$\pm n \sigma$ bands**: $S_{\text{up,exp}}^{\pm n} = \left(\pm n + \left[1 - \Phi^{-1} \left(0.05 \Phi(\mp n) \right) \right] \right) \sigma_S$

- CLs :**
- Positive bands somewhat reduced,
 - Negative ones more so

n	$S_{\text{exp}}^{\pm n} / \sqrt{B}$
+2	3.66
+1	2.72
0	1.96
-1	1.41
-2	1.05

Band width from $\sigma_{S,A}^2 = \frac{S^2}{q_S(\text{Asimov})}$ depends on S, for non-Gaussian cases, different values for each band...

Comparison with LEP/TeVatron definitions

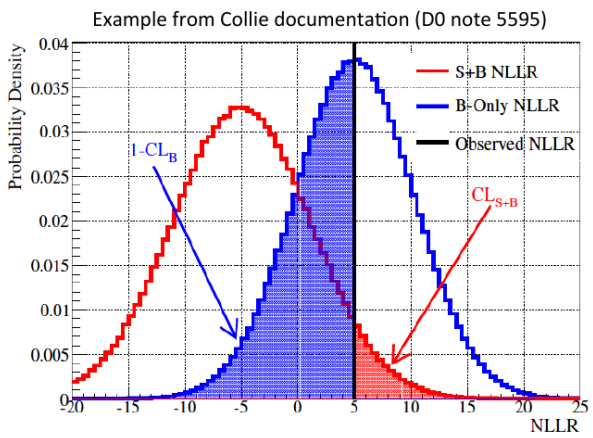
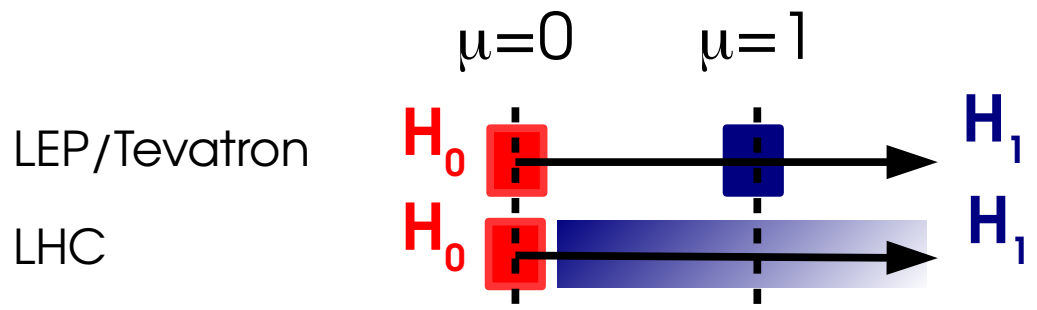
Likelihood ratios are not a new idea:

- **LEP**: Simple LR with NPs from MC
 - Compare $\mu=0$ and $\mu=1$
- **TeVatron**: PLR with profiled NPs

$$q_{LEP} = -2 \log \frac{L(\mu=0, \tilde{\theta})}{L(\mu=1, \tilde{\theta})}$$

$$q_{TeVatron} = -2 \log \frac{L(\mu=0, \hat{\theta}_0)}{L(\mu=1, \hat{\theta}_1)}$$

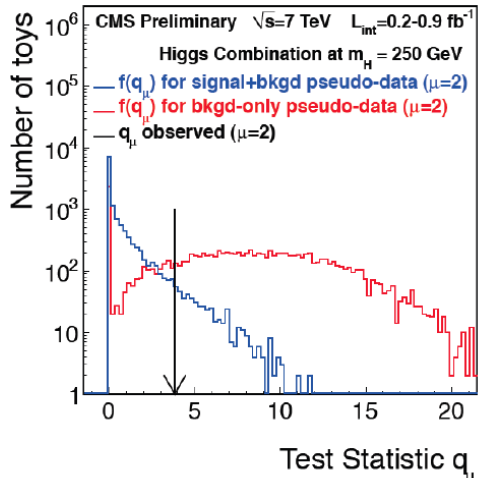
Both compare to $\mu=1$ instead of best-fit $\hat{\mu}$



→ Asymptotically:

- **LEP/TeVatron**: q linear in $\mu \Rightarrow \sim \text{Gaussian}$
- **LHC**: q quadratic in $\mu \Rightarrow \sim \chi^2$

→ Still use TeVatron-style for discrete cases



Wilks' Theorem

To test the $S=S_0$ hypothesis, consider

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

→ Assume **Gaussian regime** (e.g. large n_{evts} ,

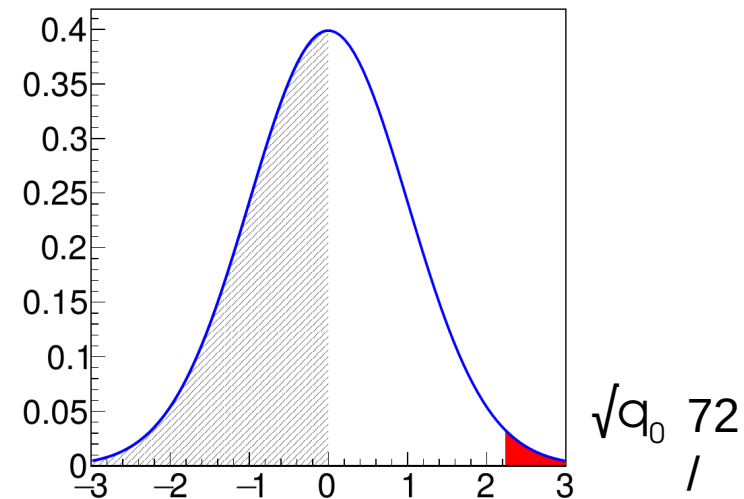
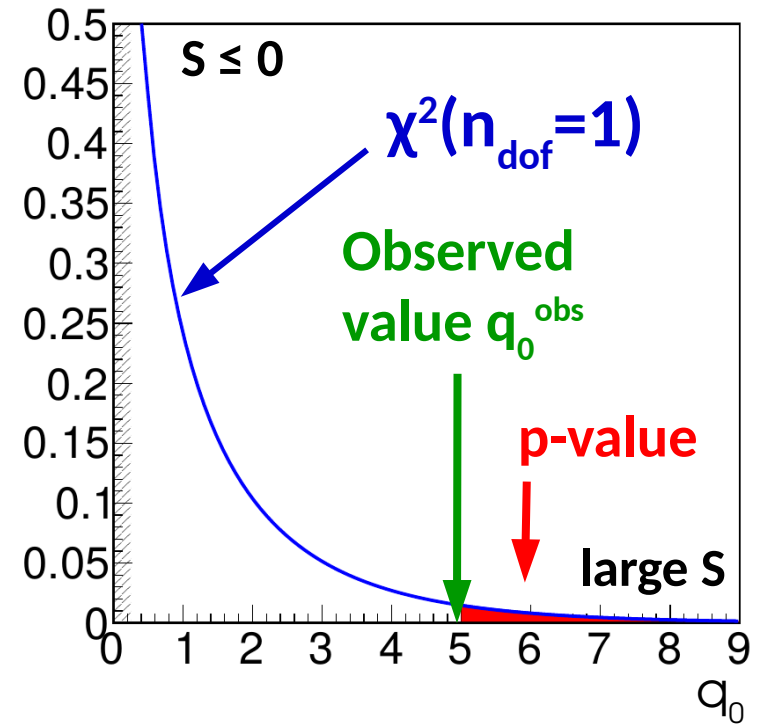
Central-limit theorem) : then:

Wilk's Theorem: $t(S_0)$ is distributed as a χ^2

under $S=S_0$: $f(t_{S_0} | S=S_0) = f_{\chi^2(n_{\text{dof}}=1)}(t_{S_0})$

⇒ In particular, the significance is:

$$Z = \sqrt{q_0}$$



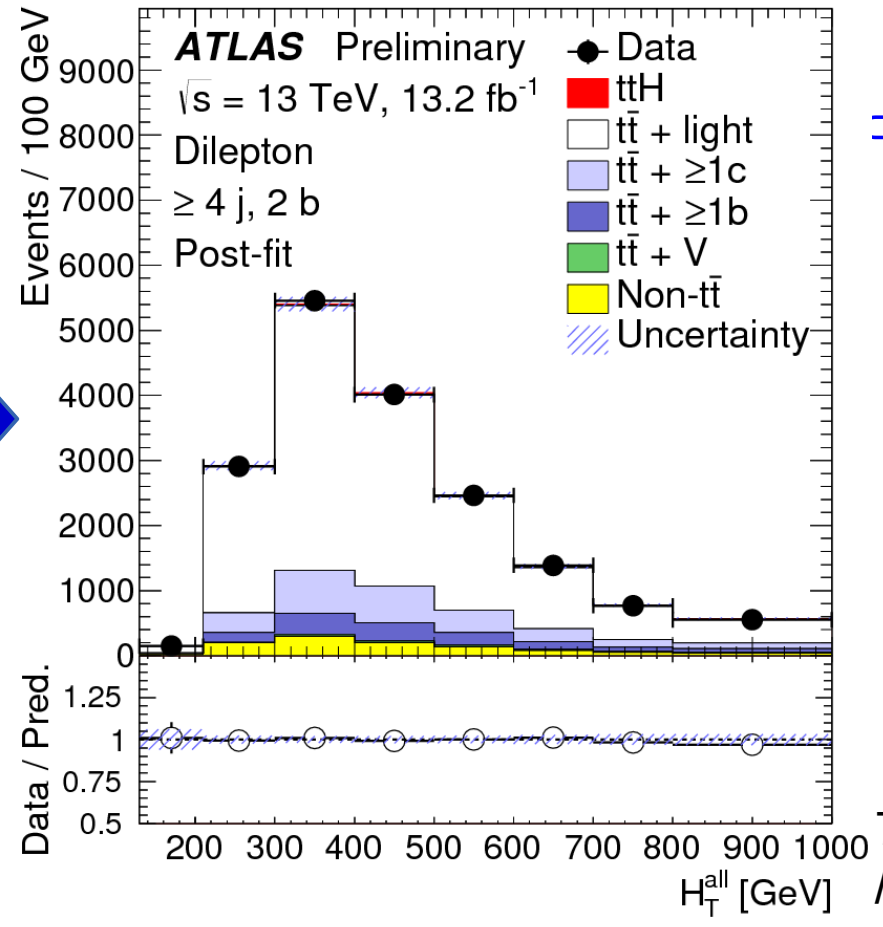
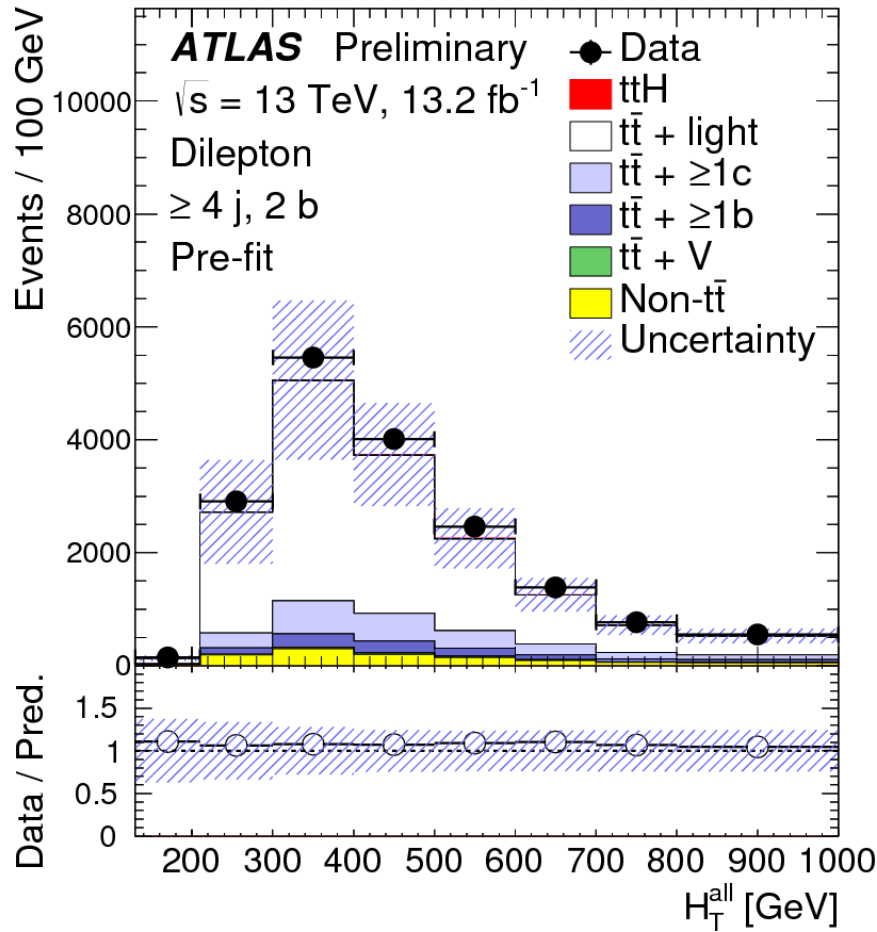
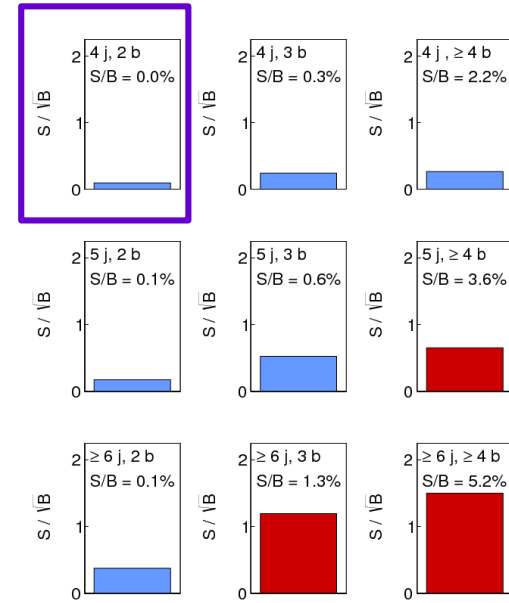
Profiling Example: $t\bar{t}H \rightarrow b\bar{b}$

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

→ **Reduction in large uncertainties on $t\bar{t}$ bkg**

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

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



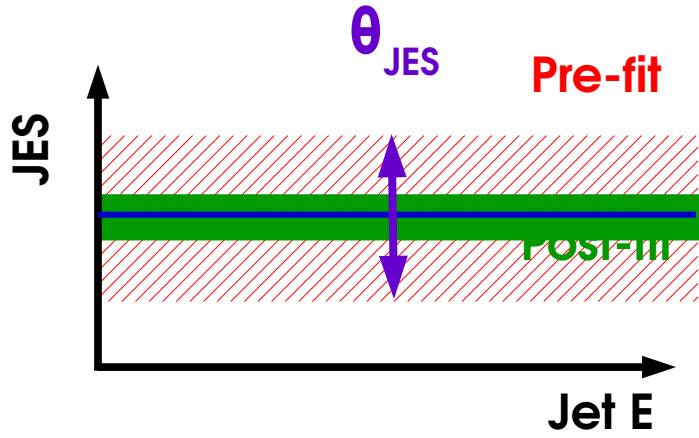
ATLAS-CONF-2016-08

73 /

Profiling Issues

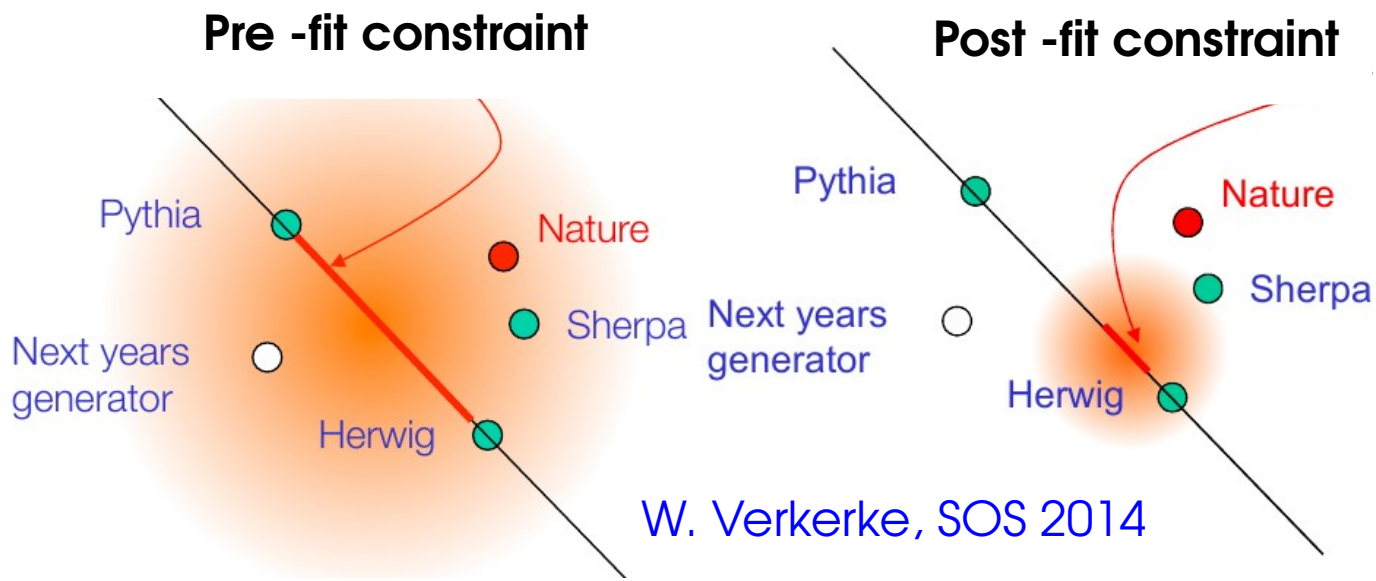
Too simple modeling can have unintended effects

- e.g. single Jet E scale parameter:
- ⇒ Low-E jets calibrate high-E jets – intended ?



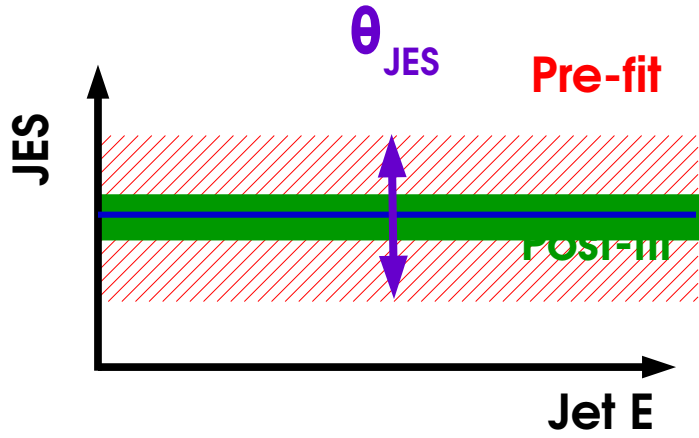
Two-point uncertainties:

→ Interpolation may not cover full configuration space, can lead to too-strong constraints

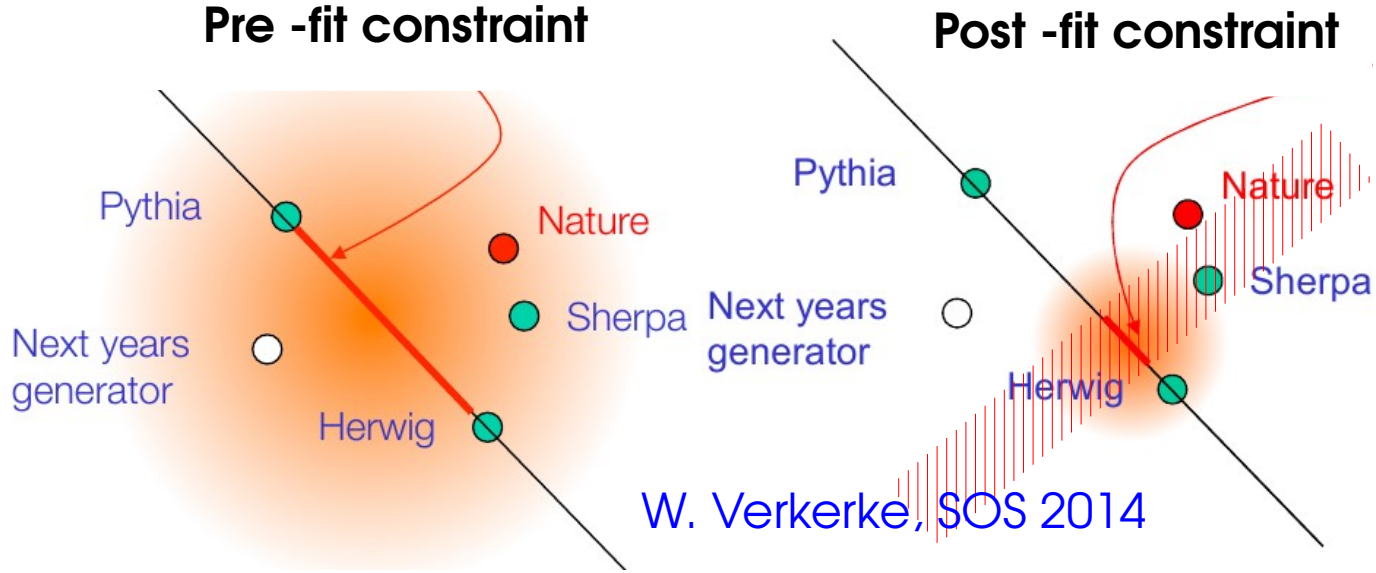


Profiling Issues

Too simple modeling can have unintended effects
→ e.g. single Jet E scale parameter:
⇒ Low-E jets calibrate high-E jets – intended ?



Two-point uncertainties:
→ Interpolation may not cover full configuration space, can lead to too-strong constraints

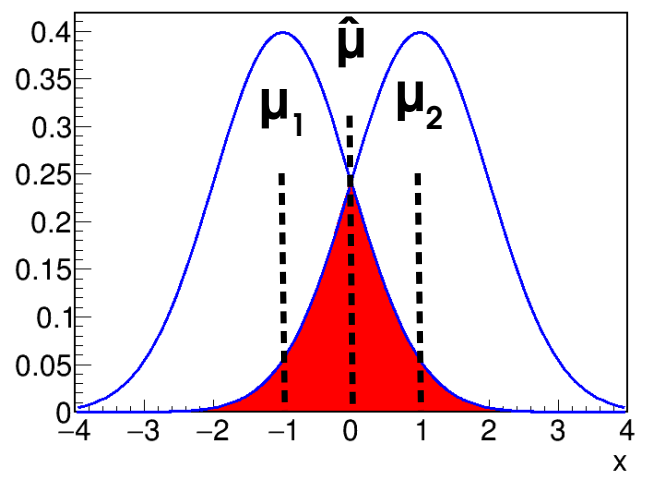
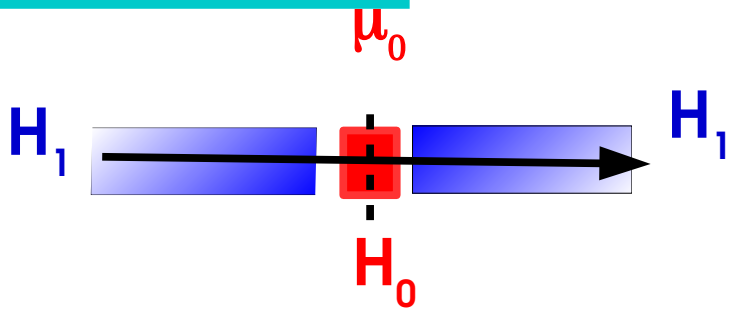


Test Statistics for Limit-Setting

Interval :

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$



“Two-sided” test

Try to exclude μ values away from $\hat{\mu}$.

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

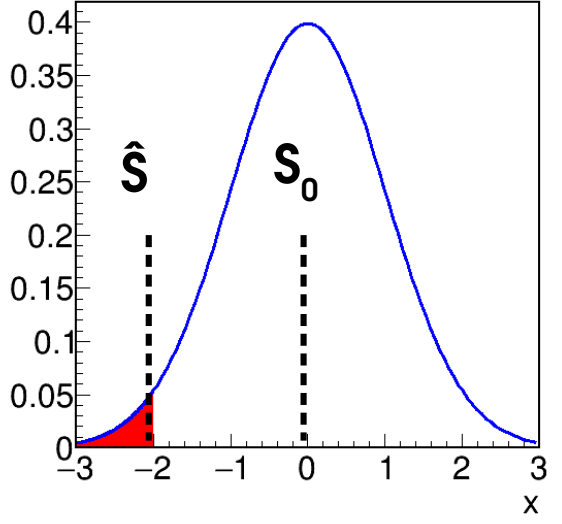
Limit-setting

$H_0 : S = S_0$

$H_1 : S < S_0$



$$q(S_0) = \begin{cases} -2 \log \frac{L(S = S_0)}{L(\hat{S})} & S_0 > \hat{S} \\ 0 & S_0 \leq \hat{S} \end{cases}$$



Discovery is also one-sided, for $S > 0$!

Try to exclude values of S that are above \hat{S} .

⇒ “One-sided” test : only interested in excluding above

Homework solutions

Homework 7: Gaussian Profiling

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

→ Signal region: $n \sim G(S + \theta, \sigma_{\text{stat}})$
 → Control region: $\theta^{\text{obs}} \sim G(\theta, \sigma_{\text{syst}})$

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

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

For $S = \hat{S}$, matches MLE as it should

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

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

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

Stat uncertainty (on n) and systematic (on θ) add in quadrature 77

Homework 8: CL_s computation

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

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

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

$$\text{Conditional MLE: } \hat{\theta}(\mu) = \frac{\sigma_{\text{syst}}}{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2} (n - S - B) \quad \left. \vphantom{\hat{\theta}(\mu)} \right\} \text{PLR: } \lambda(\mu) = \left(\frac{S + B - n}{\sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}} \right)^2$$

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

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

Homework 8: Bayesian computation

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

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

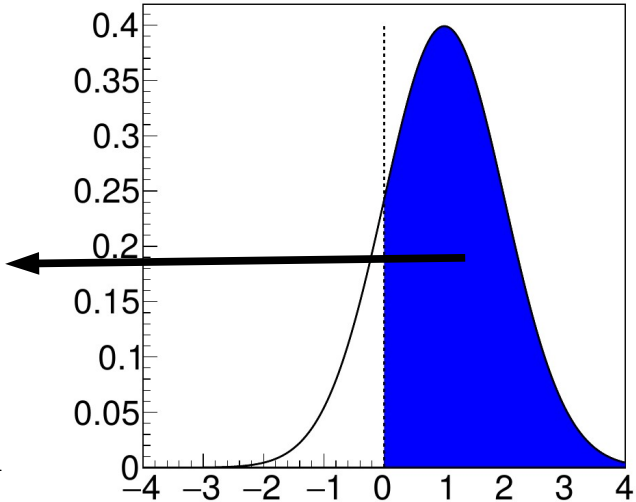
Bayesian: $G(\theta)$ is actually a **prior** on $\theta \Rightarrow$ perform integral (**marginalization**)

$$P(n | S) = G(S; n - B, \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}) \quad \text{same effect as profiling!}$$

Need $P(S | n) \Rightarrow$ a prior for S - take flat PDF over $S > 0$
 \Rightarrow Truncate Gaussian at $S=0$:

$$P(S | n) = P(n | S) P(S)$$

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



Bayesian Limit:

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

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

same result as CL_s !