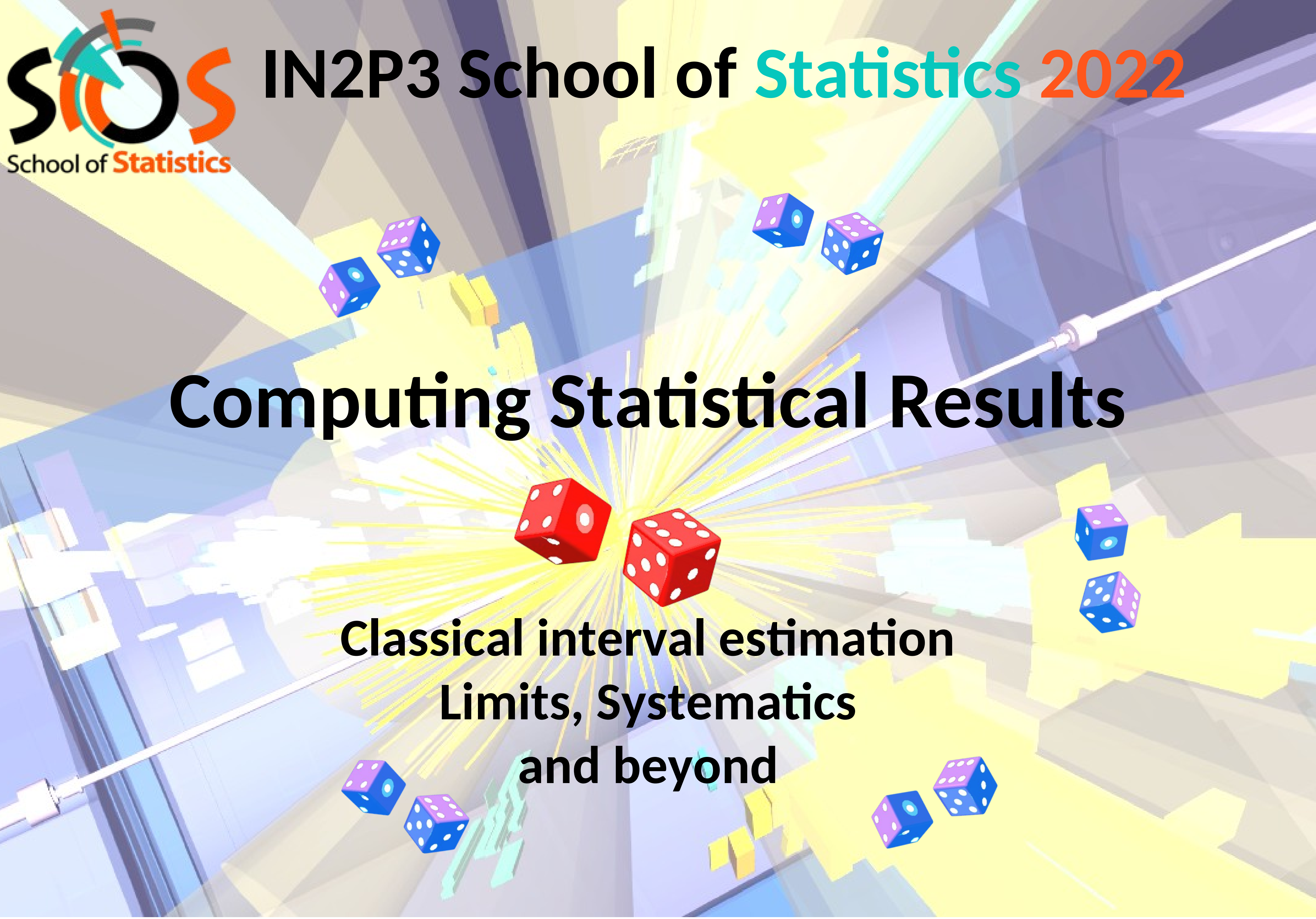
IN2P3 School of Statistics 2022



Computing Statistical Results

Classical interval estimation
Limits, Systematics
and beyond

Nicolas Berger (LAPP Annecy)

Lecture 2

Lecture Plan

Statistics basic concepts (Monday/Tuesday)

- **Basic ingredients** (PDFs, etc.)
- **Parameter estimation** (maximum likelihood, least-squares, ...)
- **Model testing** (χ^2 tests, hypothesis testing, p-values, ...)

These lectures: Computing statistical results

- Statistical modeling
- Review of model testing
- Computing results
  - Discovery
  - Confidence intervals
  - Upper limits
- Systematics and profiling
- Bayesian techniques

See also the Hands-on tutorial yesterday covering both sets of lectures.

Highlights : Hypothesis Tests and Discovery

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

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

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 $$q_0 = -2\log\frac{L(S=0)}{L(\hat{S})} \qquad \hat{S} \geq 0$$

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

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

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

Confidence Intervals

Confidence Intervals

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

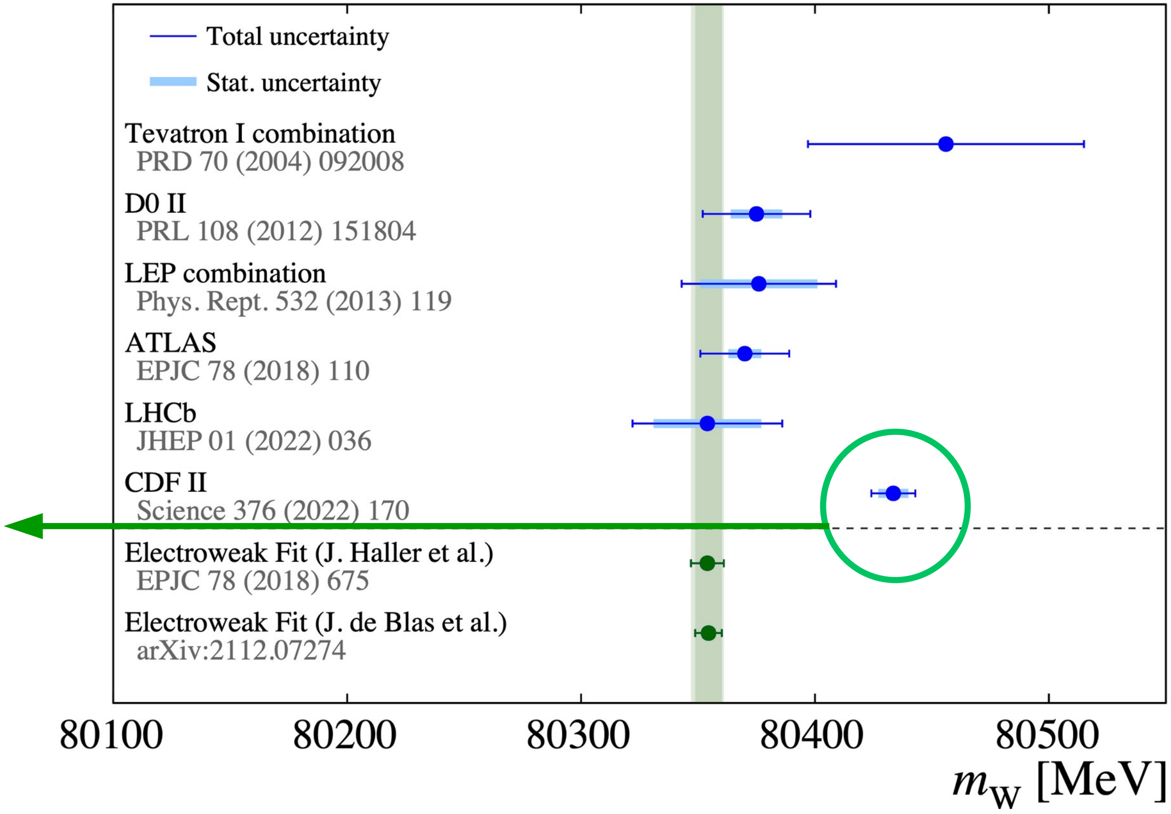
Maximum Likelihood Estimator (MLE) $\hat{\mu}$:

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

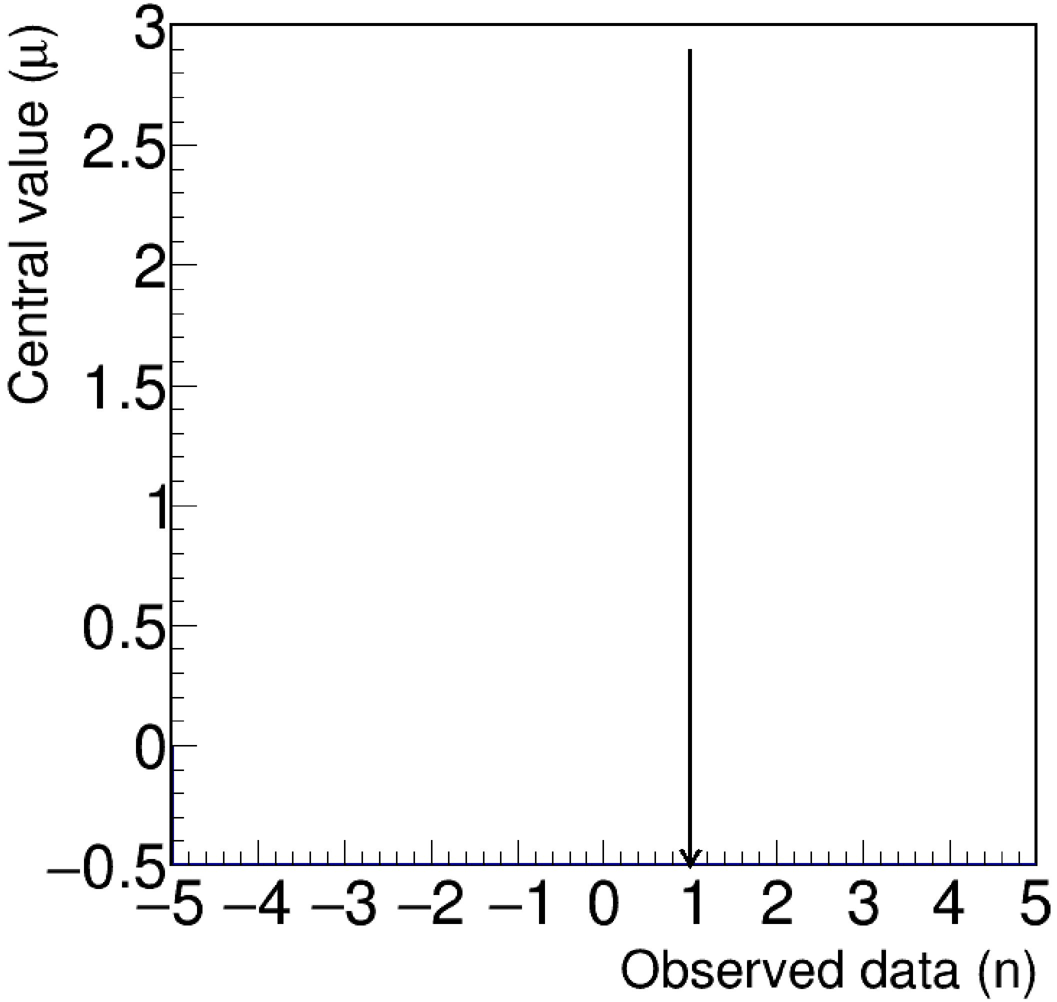
However we also need to estimate the associated uncertainty.

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



Consider a Gaussian likelihood:

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

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

$$\Updownarrow$$

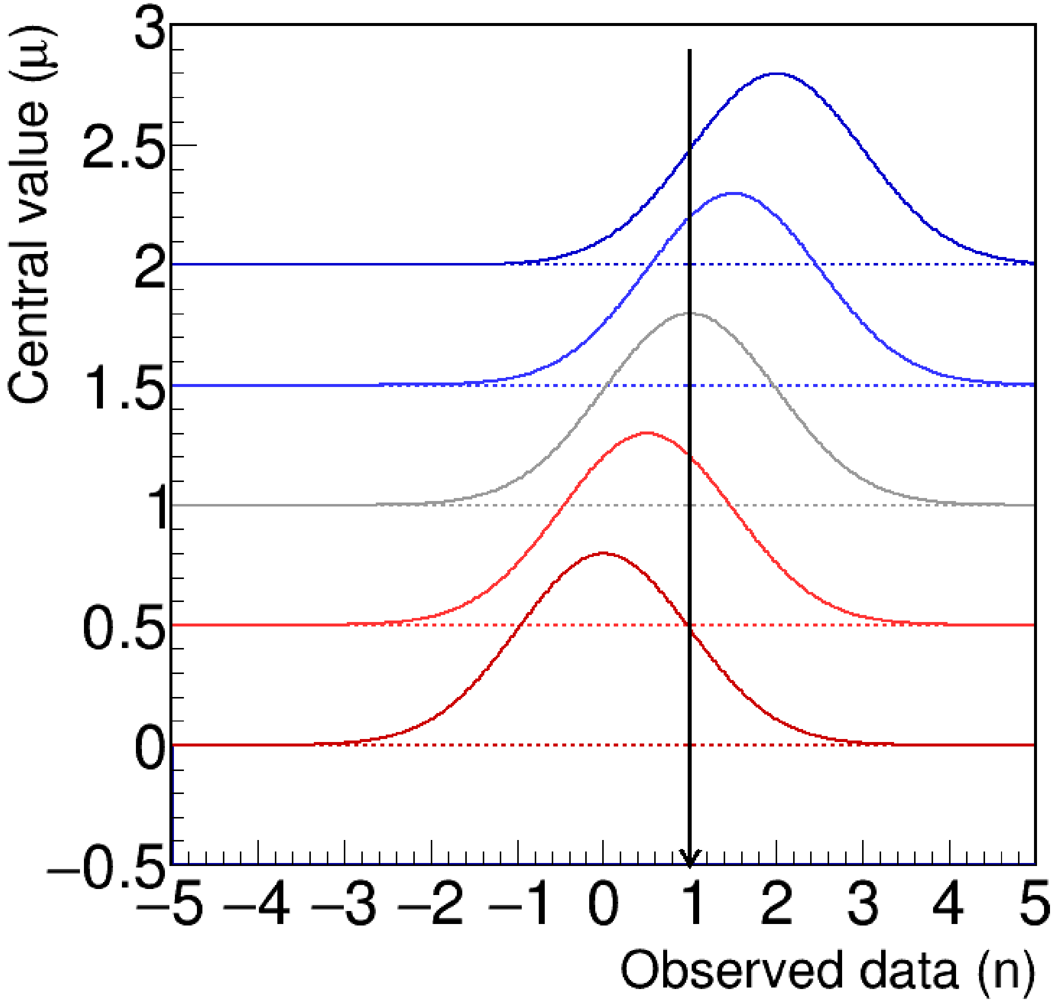
$$P(n - \sigma < \mu < n + \sigma) = 68.3\,\%$$

Still a statement on n!

μ = n ± σ at 68% CL (“1σ”)

The reported interval n ± σ will contain the true value of μ 68.3% of the time

Gaussian confidence intervals



Consider a Gaussian likelihood:

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

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

⇕

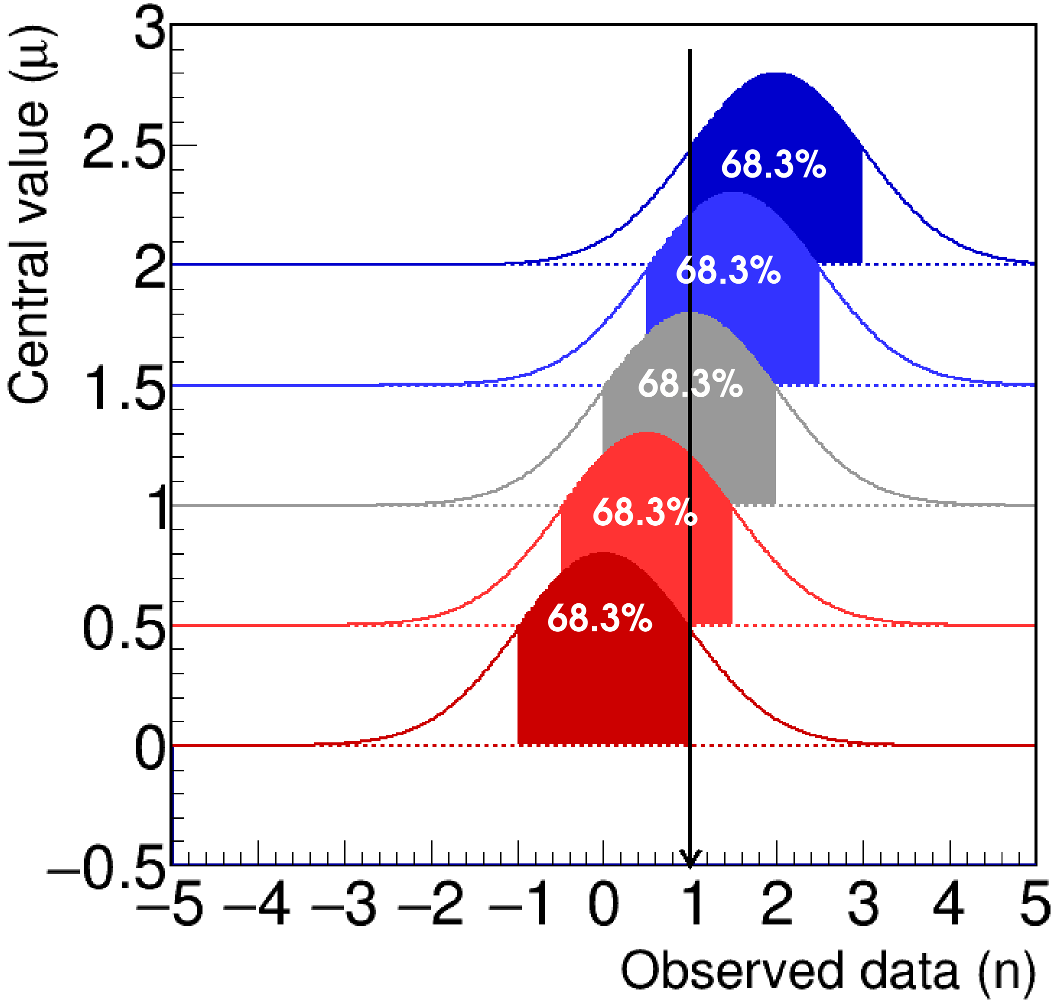
$$P(n - \sigma < \mu < n + \sigma) = 68.3\ \%$$

Still a statement on n!

μ = n ± σ at 68% CL (“1σ”)

The reported interval n ± σ will contain the true value of μ 68.3% of the time

Gaussian confidence intervals



Consider a Gaussian likelihood:

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

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

$$\Updownarrow$$

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

Still a statement on n!

μ = n ± σ at 68% CL (“1σ”)

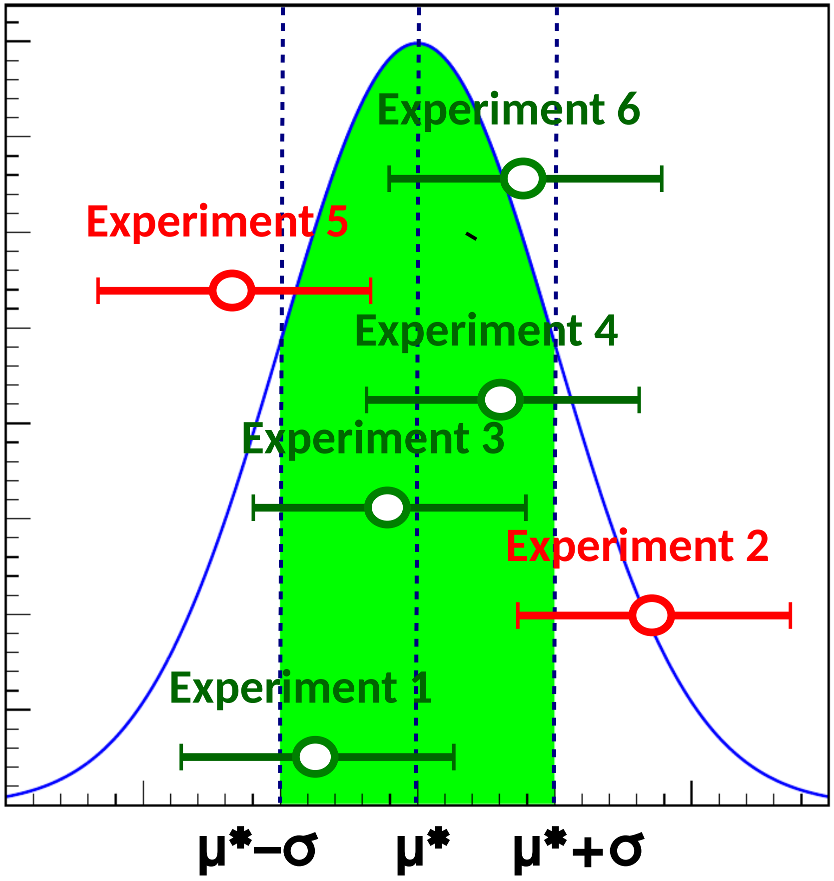
The reported interval n ± σ will contain the true value of μ 68.3% of the time

Gaussian confidence intervals

Frequentist interpretation

If we would repeat the same experiment multiple times, with true value μ*, then 68.3% of the 1σ intervals would contain μ*.

→ Crucially, this works even if we do not know μ* !



For each experiment, get the interval

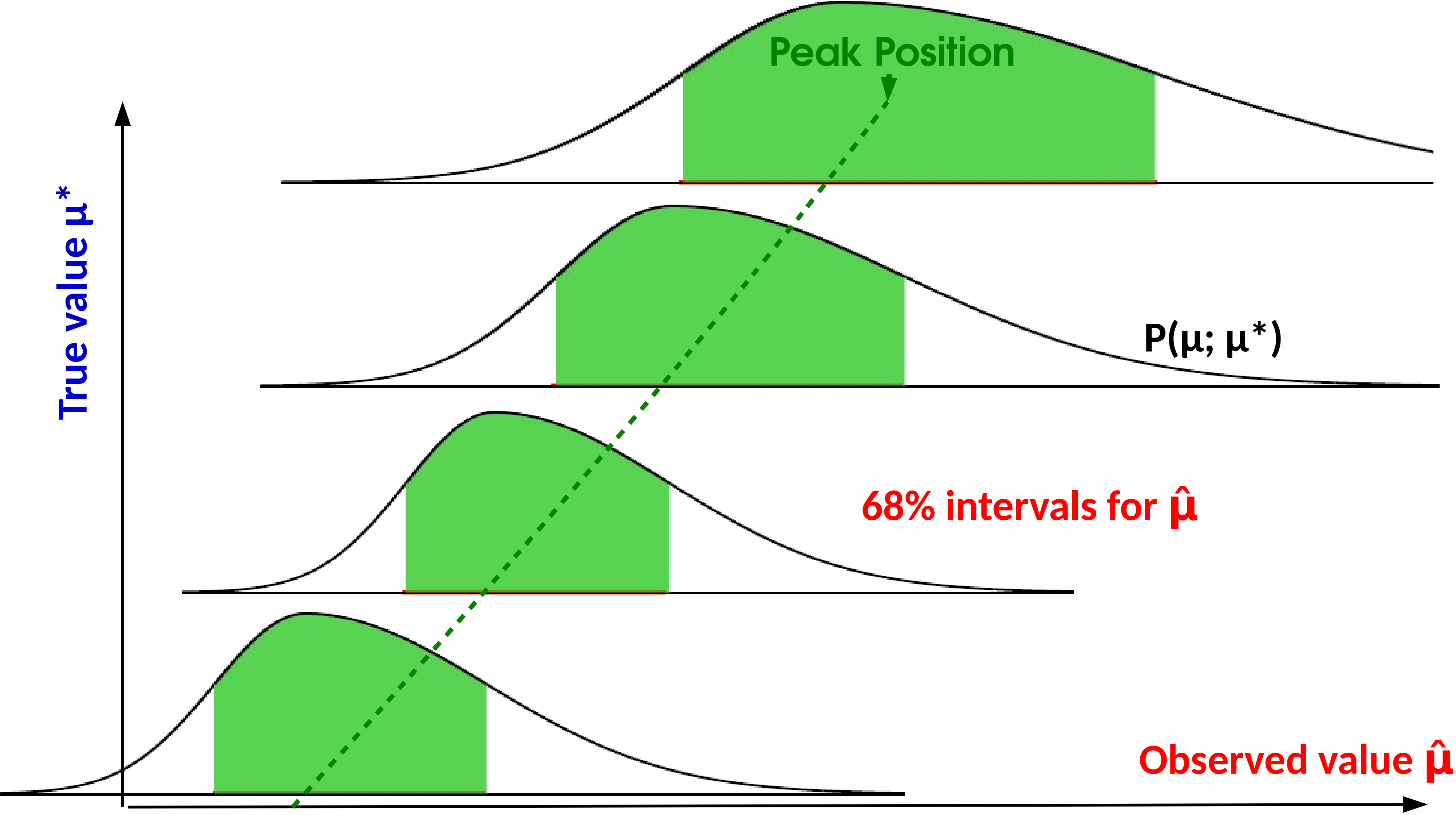
μ = n ± σ at 68% CL ("1σ")

The reported interval n ± σ will contain the true value of μ 68.3% of the time

Neyman Construction

General case: build 1σ intervals of observed values for each true value

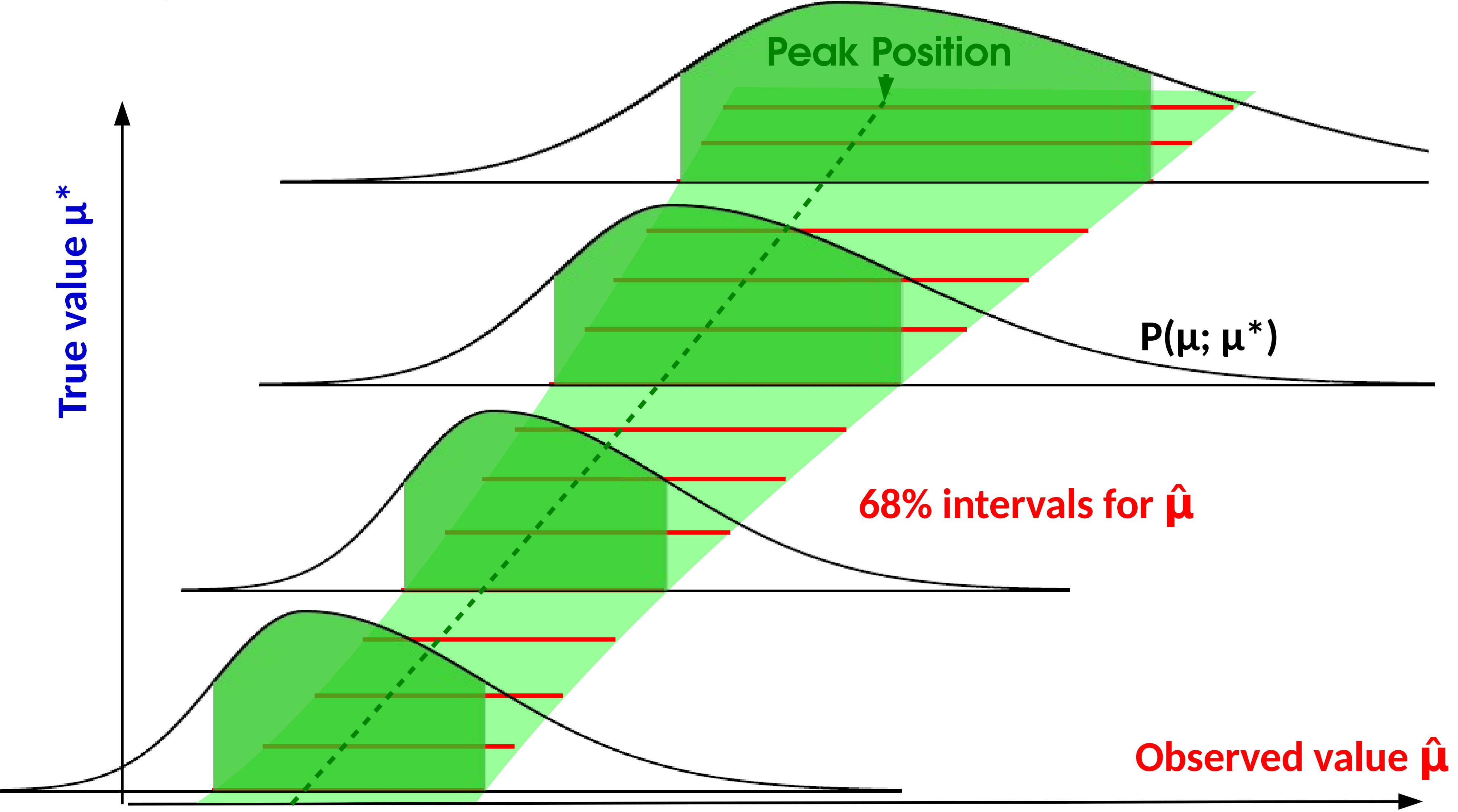
⇒ *Confidence belt*



Neyman Construction

General case: build 1σ intervals of observed values for each true value

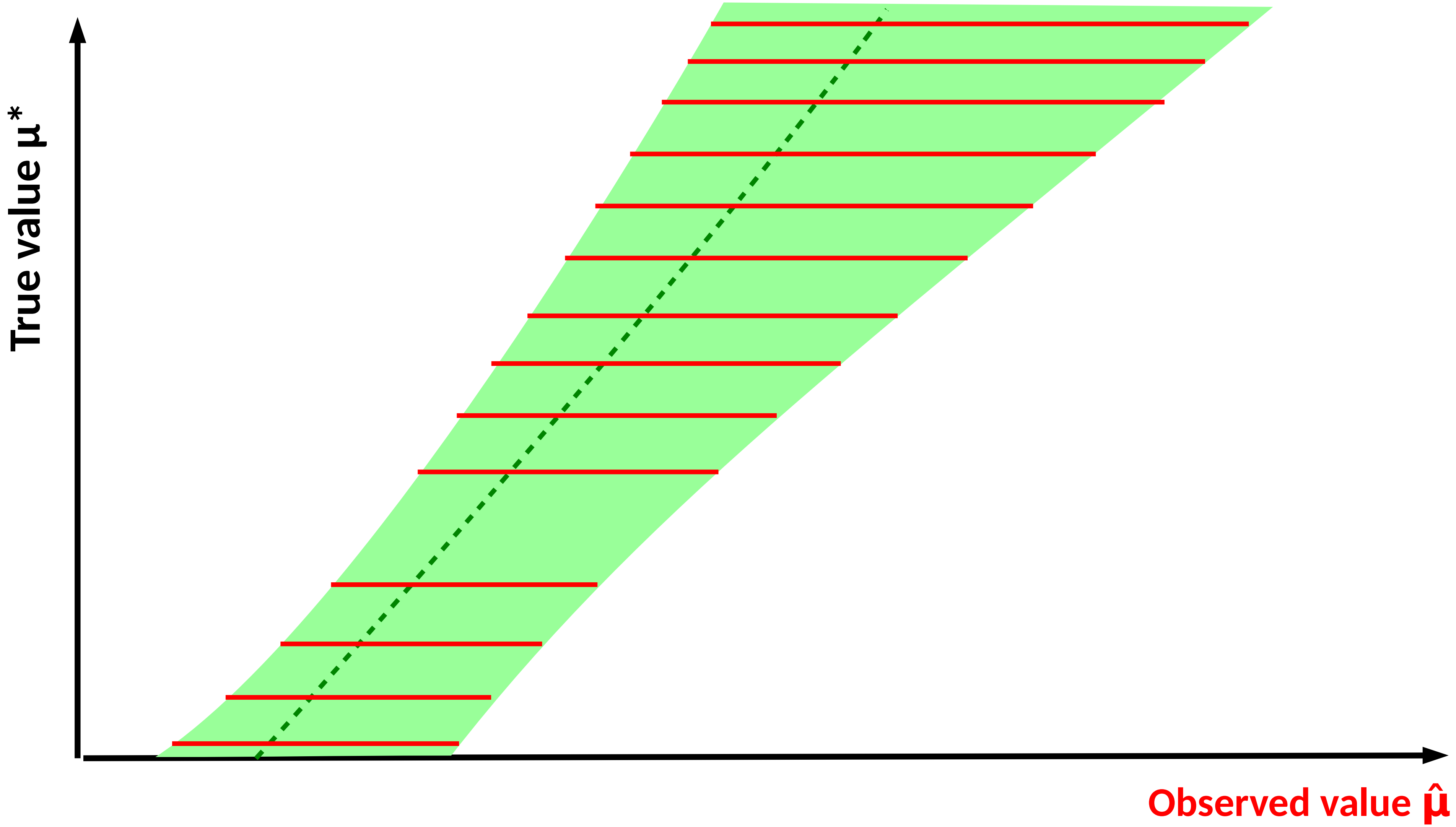
⇒ *Confidence belt*



Inversion using the Confidence Belt

General case: Intersect belt with given $\hat{\mu}$, get $P\left(\hat{\mu} - \sigma_{\mu}^{-} < \mu^{*} < \hat{\mu} + \sigma_{\mu}^{+}\right) = 68\,\%$

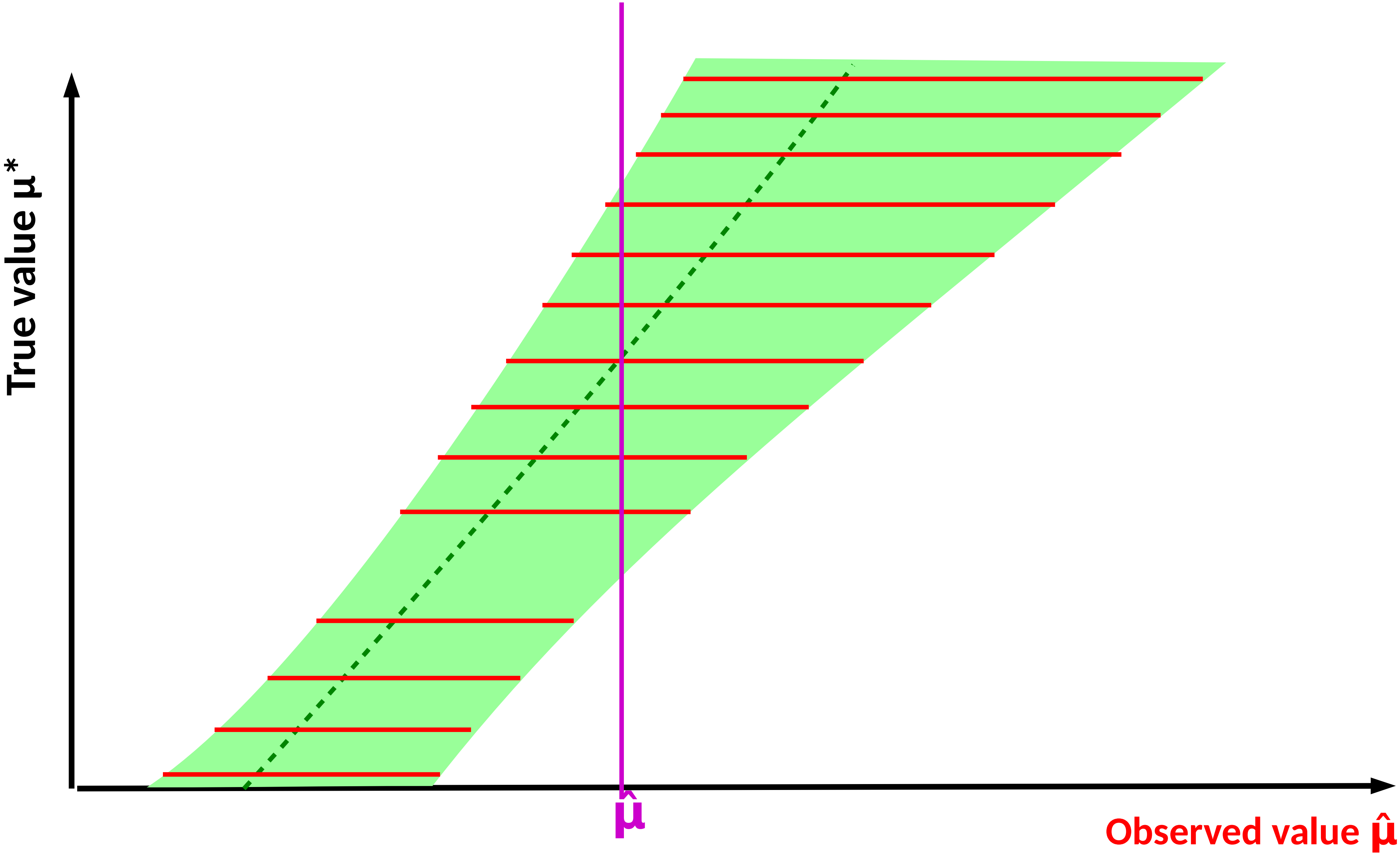
→ Same as before for Gaussian, works also when P(μ^{obs}|μ) varies with μ.



Inversion using the Confidence Belt

General case: Intersect belt with given $\hat{\mu}$, get $P\left(\hat{\mu} - \sigma_{\mu}^{-} < \mu^{*} < \hat{\mu} + \sigma_{\mu}^{+}\right) = 68\,\%$

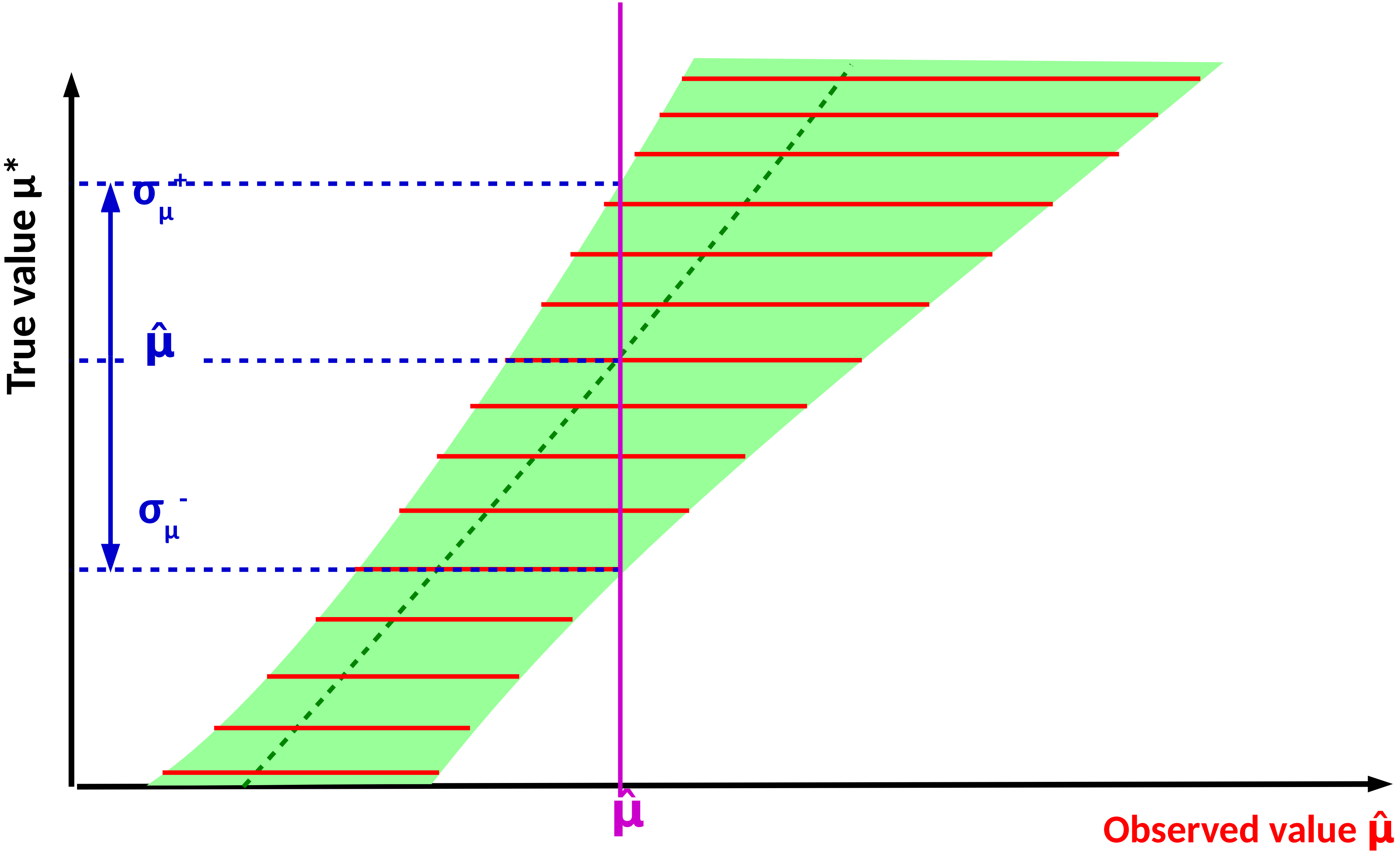
→ Same as before for Gaussian, works also when $P(\mu^{obs}|\mu)$ varies with μ.



Inversion using the Confidence Belt

General case: Intersect belt with given $\hat{\mu}$, get $P(\hat{\mu} - \sigma_{\mu}^{-} < \mu^{*} < \hat{\mu} + \sigma_{\mu}^{+}) = 68\,\%$

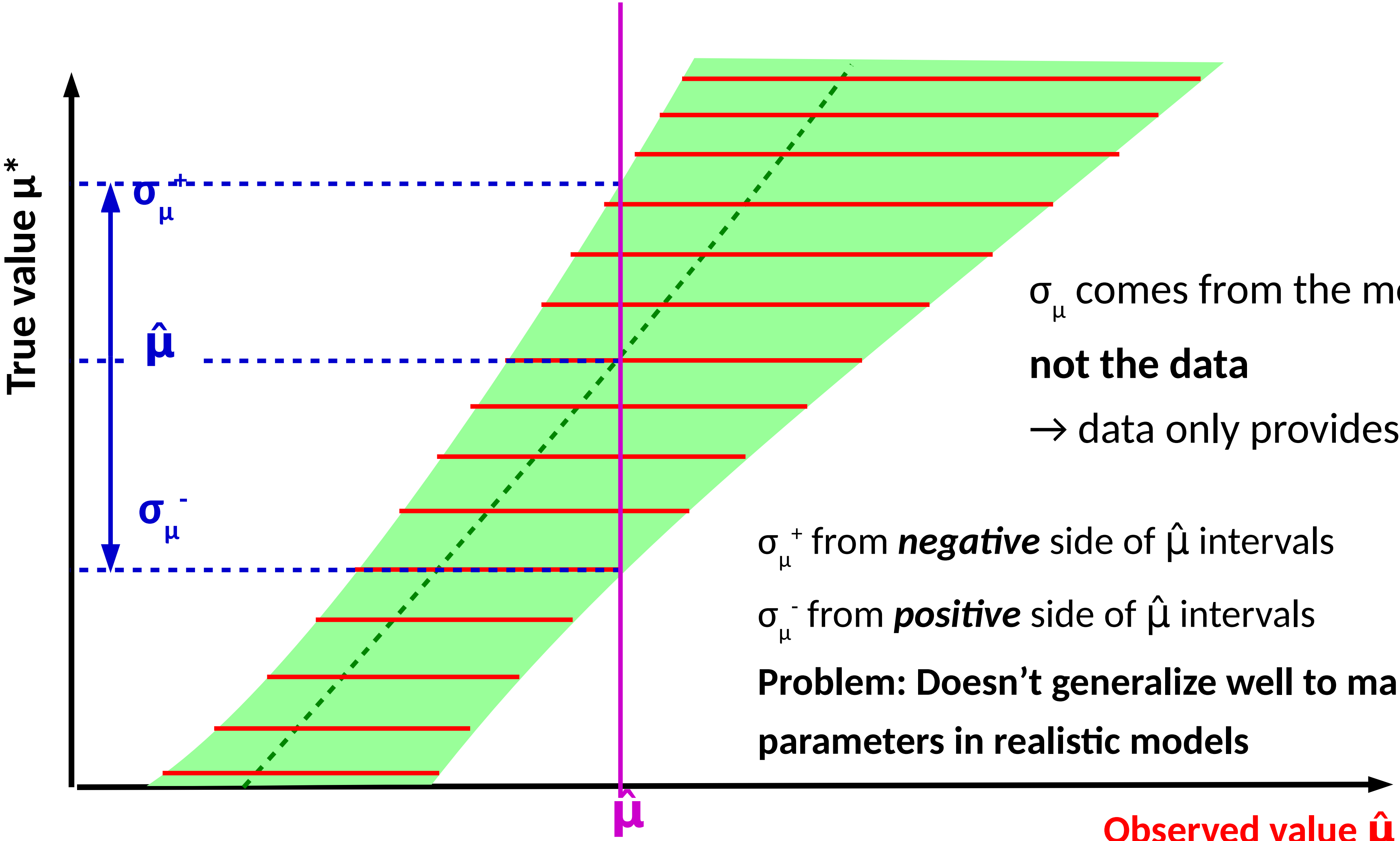
→ Same as before for Gaussian, works also when $P(\mu^{obs}|\mu)$ varies with μ.



Inversion using the Confidence Belt

General case: Intersect belt with given $\hat{\mu}$, get $P\left(\hat{\mu} - \sigma_{\mu}^{-} < \mu^{*} < \hat{\mu} + \sigma_{\mu}^{+}\right) = 68\,\%$

→ Same as before for Gaussian, works also when $P(\mu^{obs}|\mu)$ varies with μ.



σ_μ comes from the model,

not the data

→ data only provides $\hat{\mu}$.

σ_μ^+ from ***negative*** side of $\hat{\mu}$ intervals

σ_μ^- from ***positive*** side of $\hat{\mu}$ intervals

Problem: Doesn't generalize well to many parameters in realistic models

General case: Likelihood Intervals

Confidence intervals from L(μ):

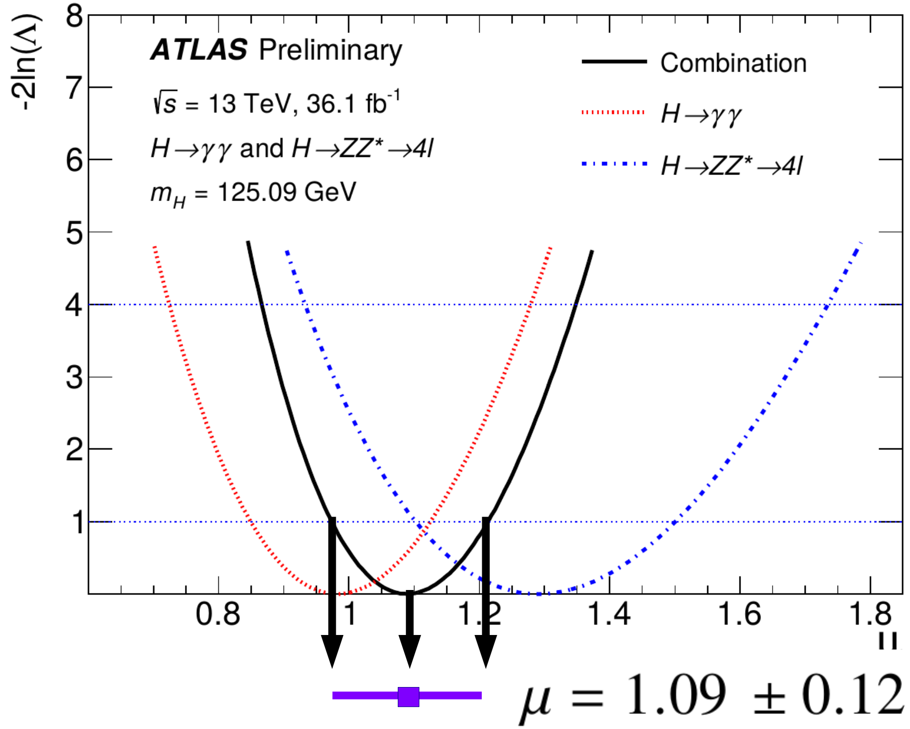
- Test various values μ using the **Profile Likelihood Ratio t(μ)**
- Minimum (=0) for μ=μ̂, rises away from μ̂.
- Good properties thanks to the Neyman-Pearson lemma.

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 μ̂.

ATLAS-CONF-2017-047



$$\mu = 1.09 \pm 0.12$$

Gaussian L(μ):

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

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

- t(μ) is parabolic, distributed as a χ^2
- Minimum occurs at μ = μ̂

$t(\mu_\pm) = 1 \Rightarrow \mu = n \pm \sigma$ 1σ interval!

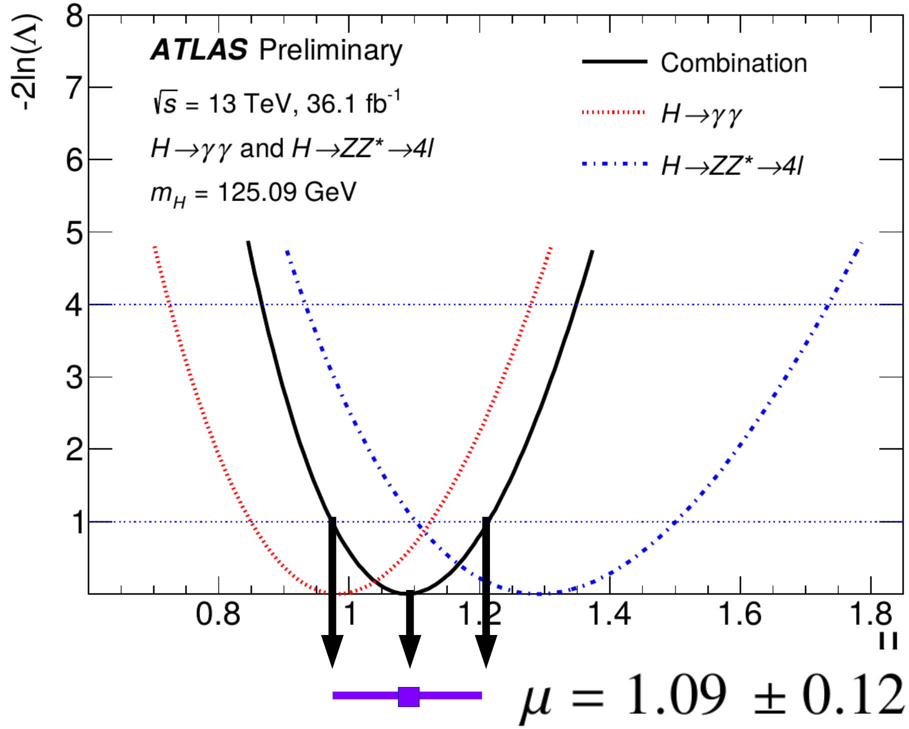
General case: Likelihood Intervals

Confidence intervals from L(μ):

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

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

ATLAS-CONF-2017-047



$\mu = 1.09 \pm 0.12$

General case:

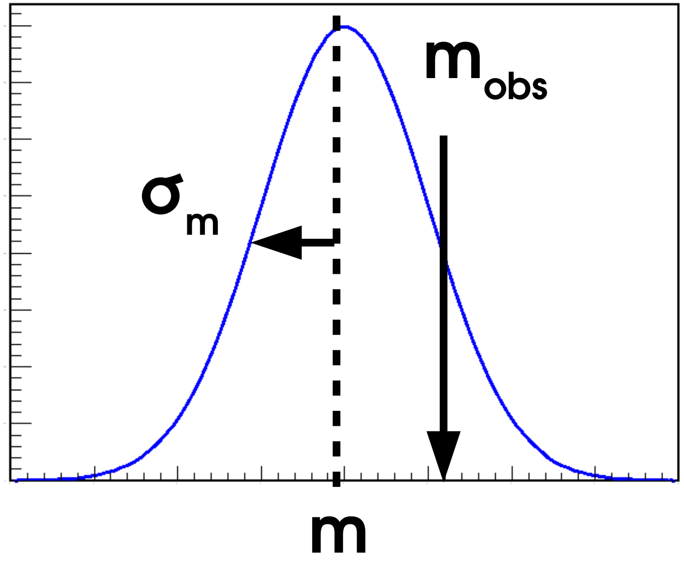
- Generally not a perfect parabola
- Minimum still at **μ = μ̂**

Asymptotic approximation

→ Compute t(μ) using the exact L(μ)

→ **1σ interval given by t(μ) 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}:

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

→ Compute the best-fit value (MLE) $\hat{m}$

→ Compute t(m)

→ Compute the 1σ (68.3% CL) interval on m

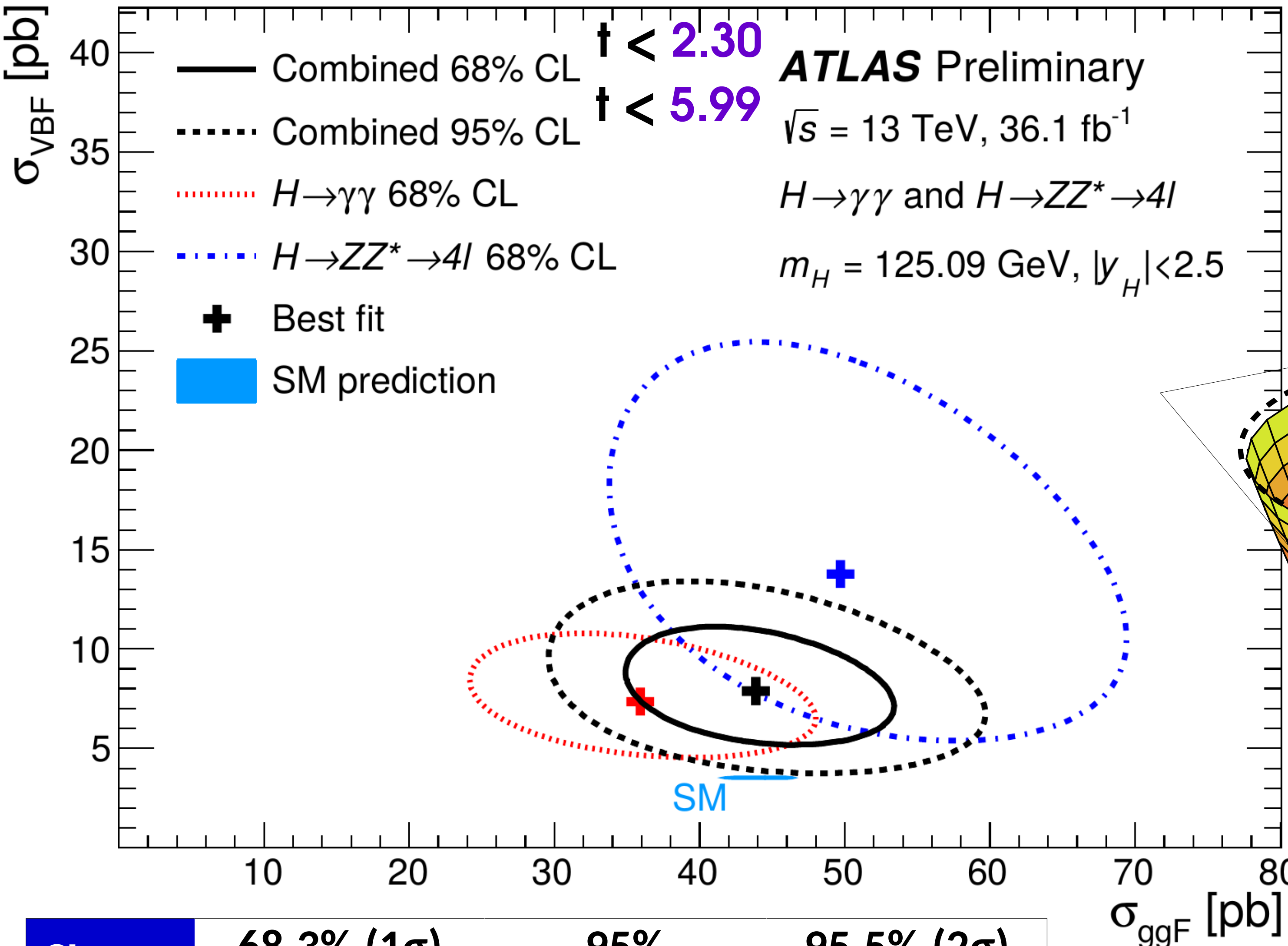
Solution: $m = m_{obs} \pm \sigma_m$

→ As expected!

→ General method can be applied in the same way to more complex cases

2D Example: Higgs σ_{VBF} vs. σ_{ggF}

ATLAS-CONF-2017-047

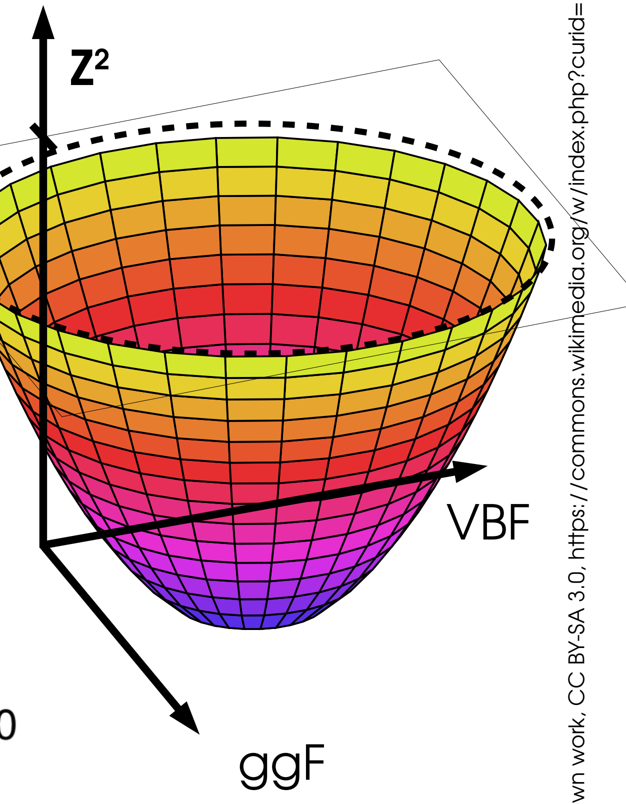


$t < 2.30$

$t < 5.99$

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

$t_{ggF,VBF} \sim \chi^2(N_{dof}=2)$



By Krishnavedala - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=15278826

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

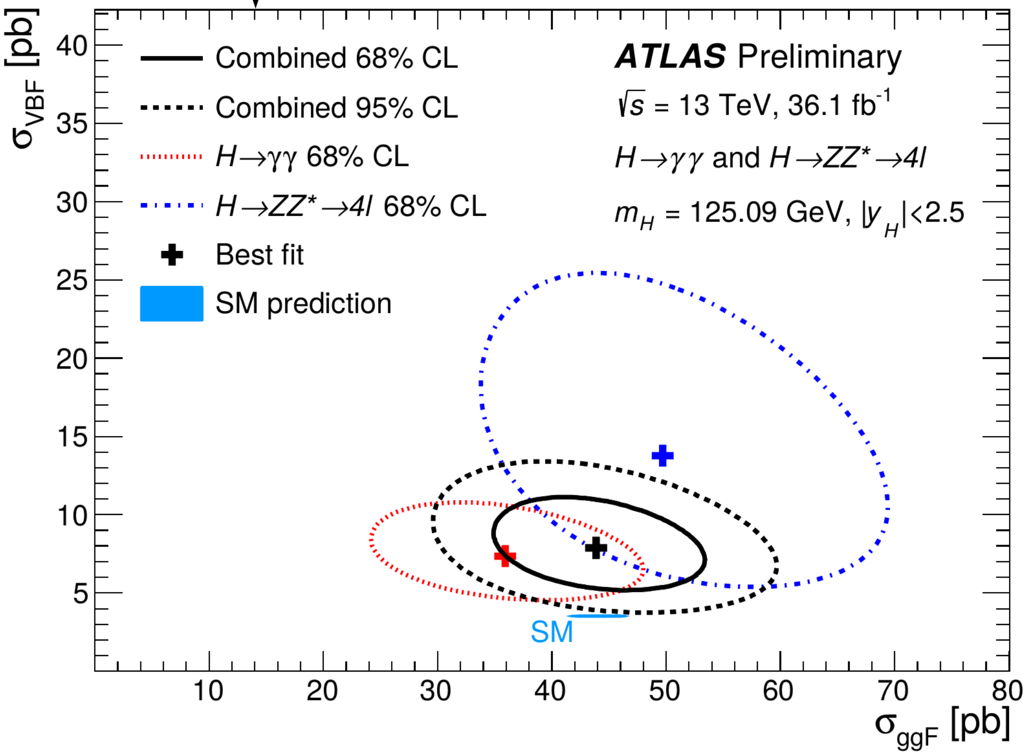
Reparameterization

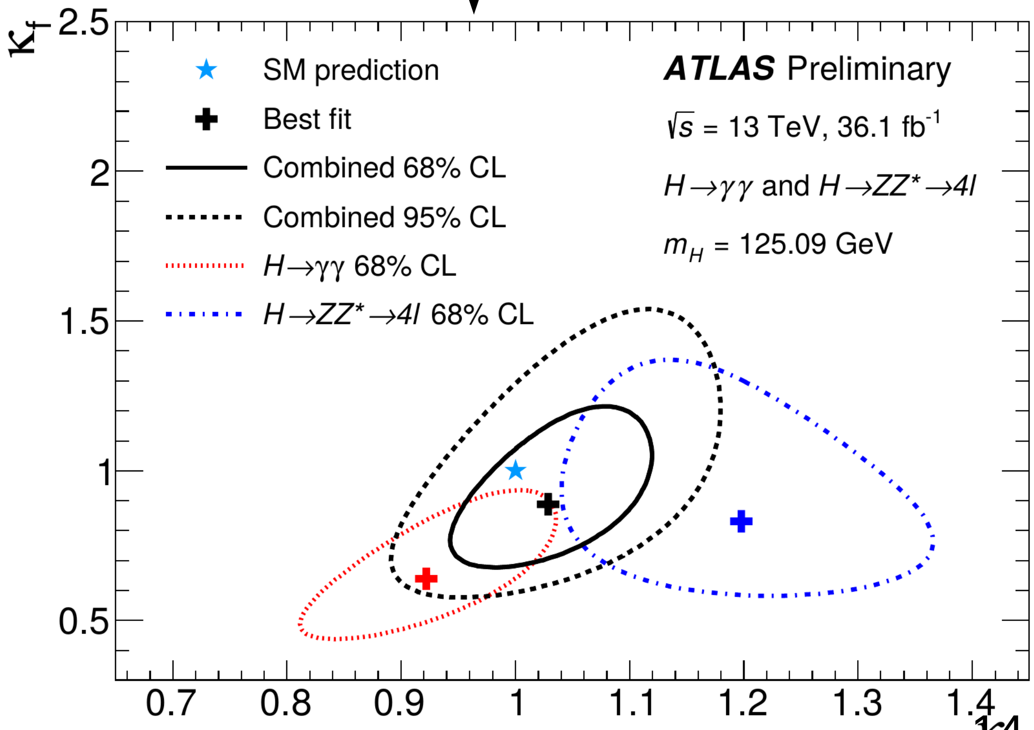
Start with basic measurement in terms of e.g. **($\sigma \times B$)**

→ 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.

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





Upper Limits

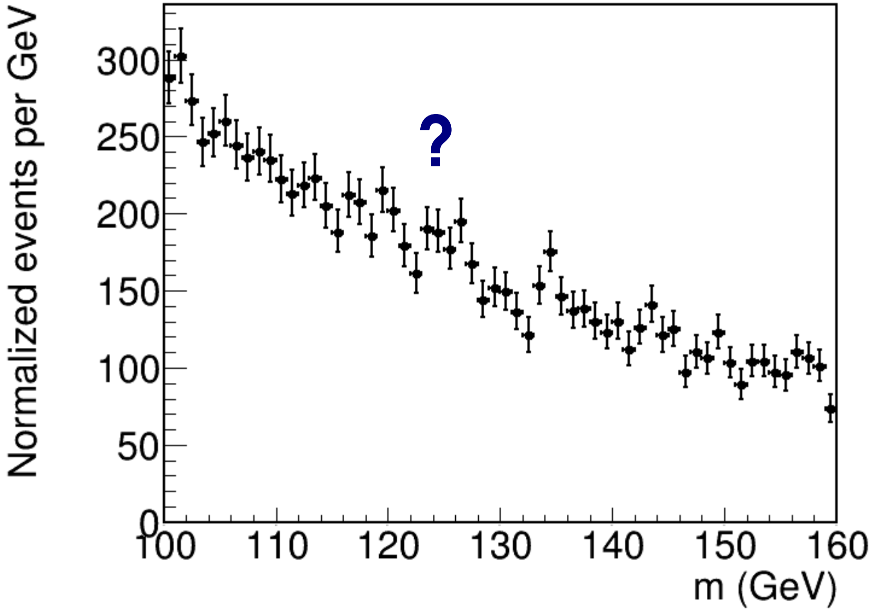
Hypothesis tests for Limits

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”



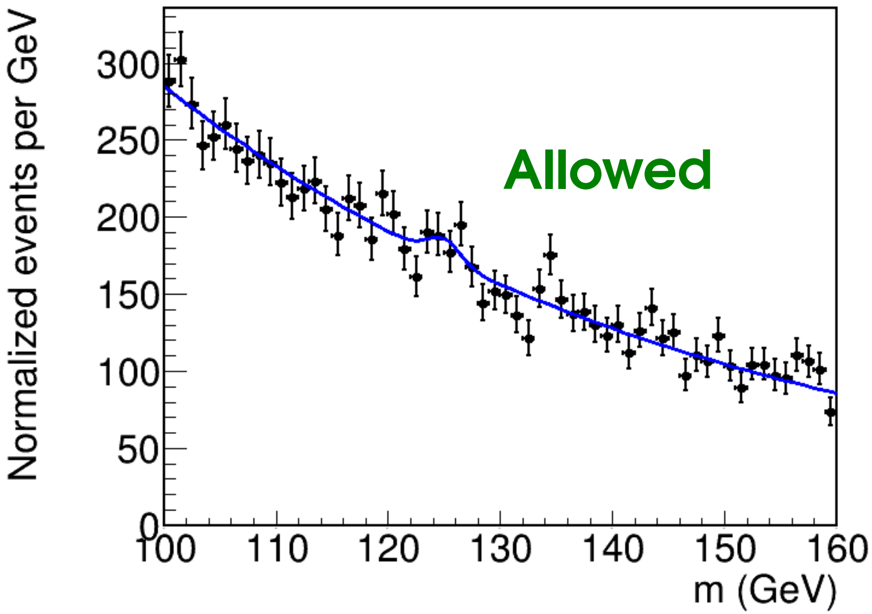
Hypothesis tests for Limits

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”



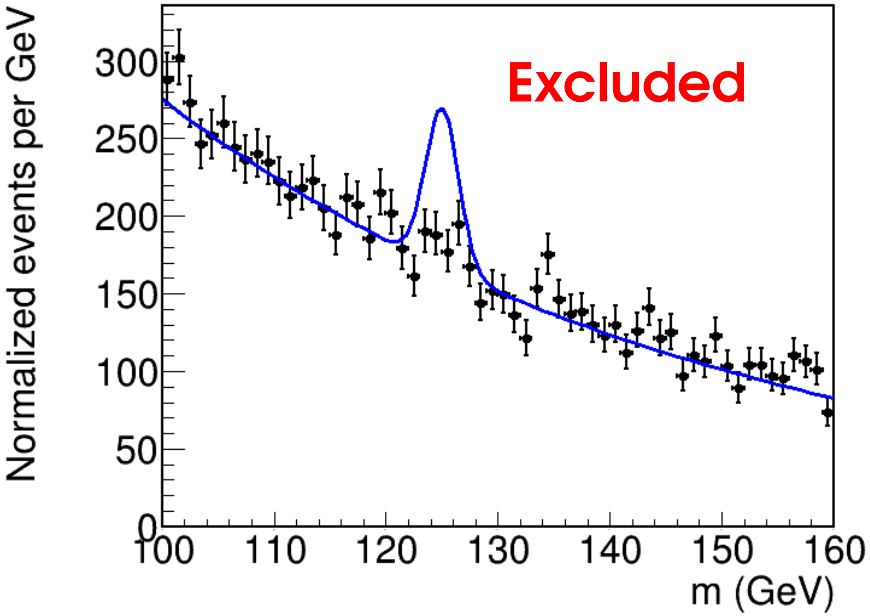
Hypothesis tests for Limits

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”



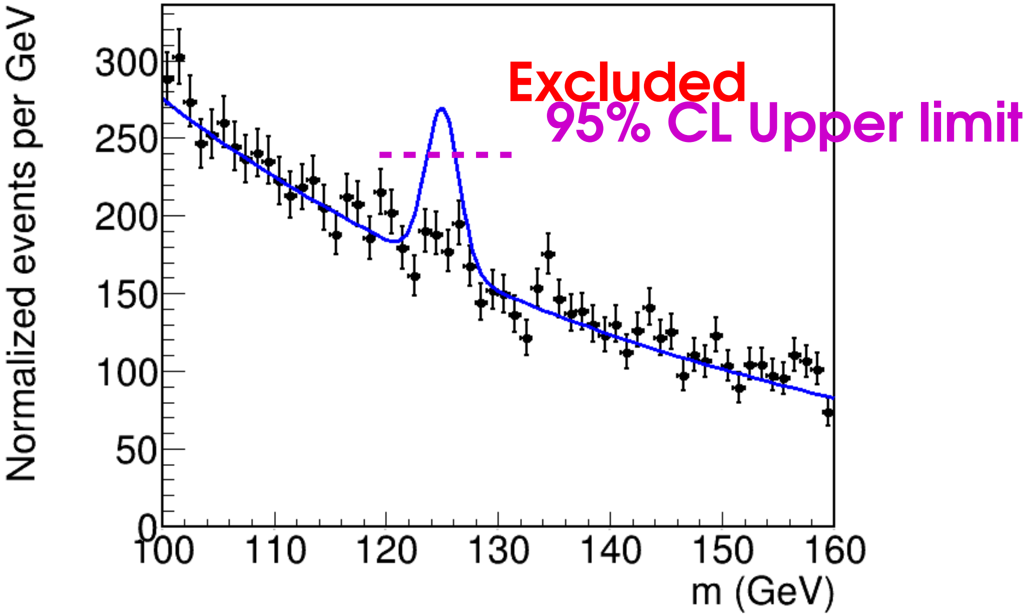
Hypothesis tests for Limits

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**

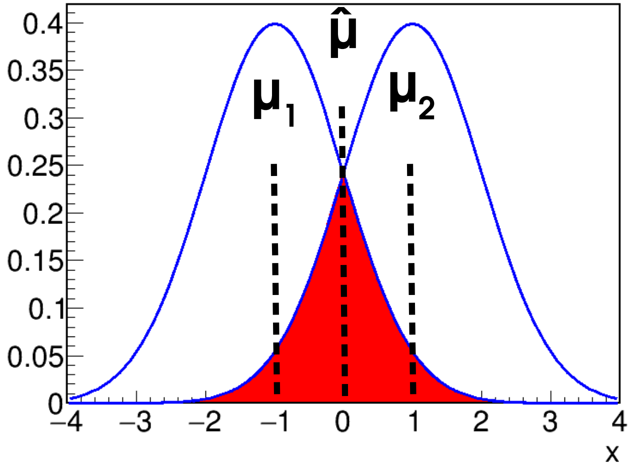
→ Typically report **95% CL** upper limit (p-value = 5%) : “S < S_0 @ 95% CL”



Test Statistics for Limit-Setting

Confidence Interval :

Try to exclude μ values away from μ̂.

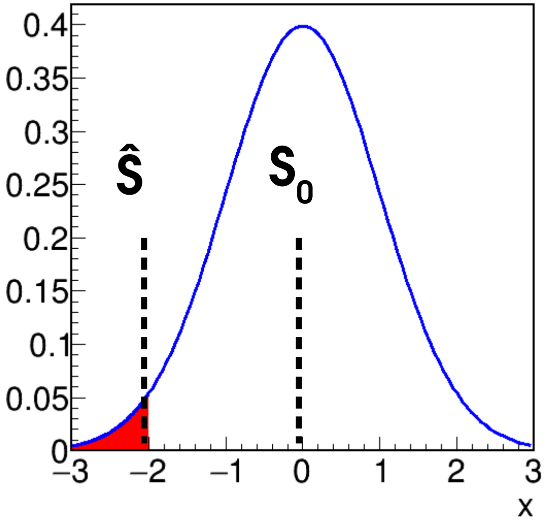


$$t(\mu_0) = -2\log\frac{L(\mu=\mu_0)}{L(\hat{\mu})}$$

“**Two-sided**” test

Limit-setting

Try to exclude values of S that are above Ŝ.



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

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

Discovery was also one-sided, for S>0

Inversion : Getting the limit for a given CL

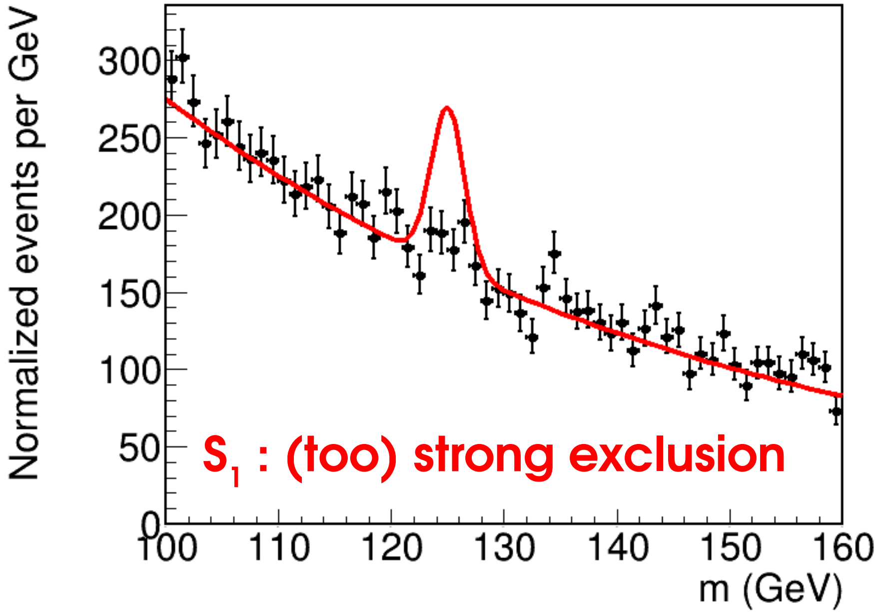
Procedure:

CL	p	Region
90%	10%	$\sqrt{q(S)} > 1.28$
95%	5%	$\sqrt{q(S)} > 1.64$
99%	1%	$\sqrt{q(S)} > 2.33$

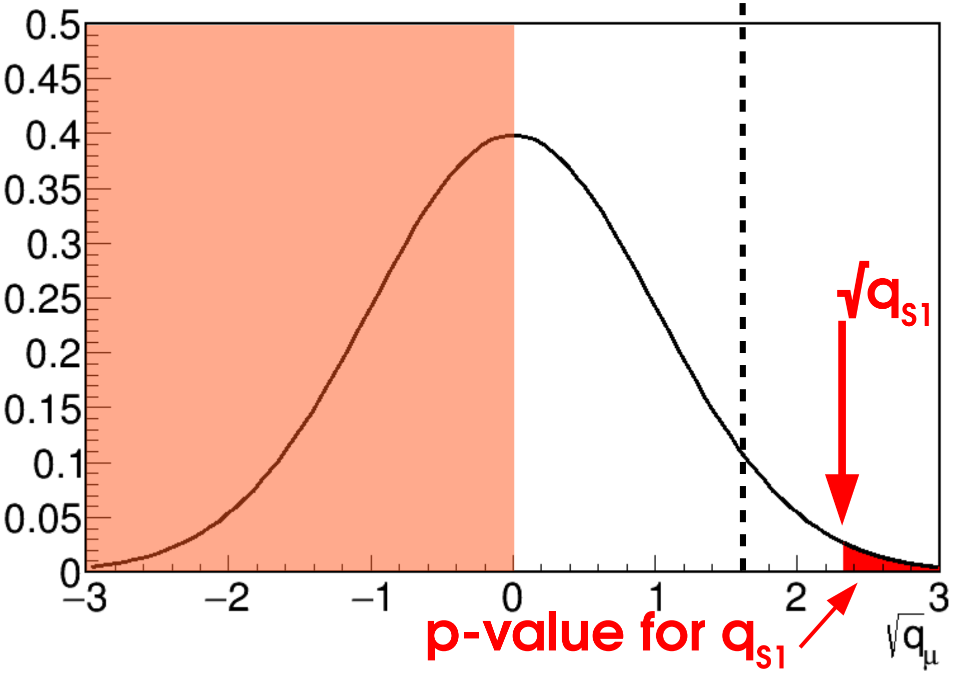
→ Compute $q(S_0)$ for some S_0,
get the exclusion p-value $p(S_0)$.

Asymptotics: $$p(S_0) = 1 - \Phi\left(\sqrt{q(S_0)}\right)$$

→ **Adjust S_0** to get the desired exclusion
Asymptotics: need **$\sqrt{q(S_{95})} = 1.64$** for **95% CL**



$\sqrt{q(S)} = 1.64$
(p = 5%)



Inversion : Getting the limit for a given CL

Procedure:

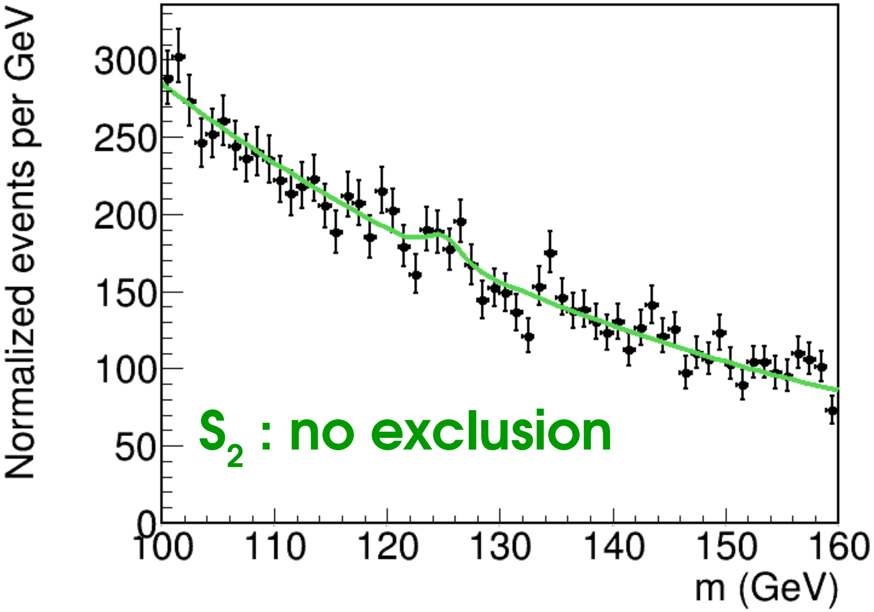
CL	p	Region
90%	10%	$\sqrt{q(S)} > 1.28$
95%	5%	$\sqrt{q(S)} > 1.64$
99%	1%	$\sqrt{q(S)} > 2.33$

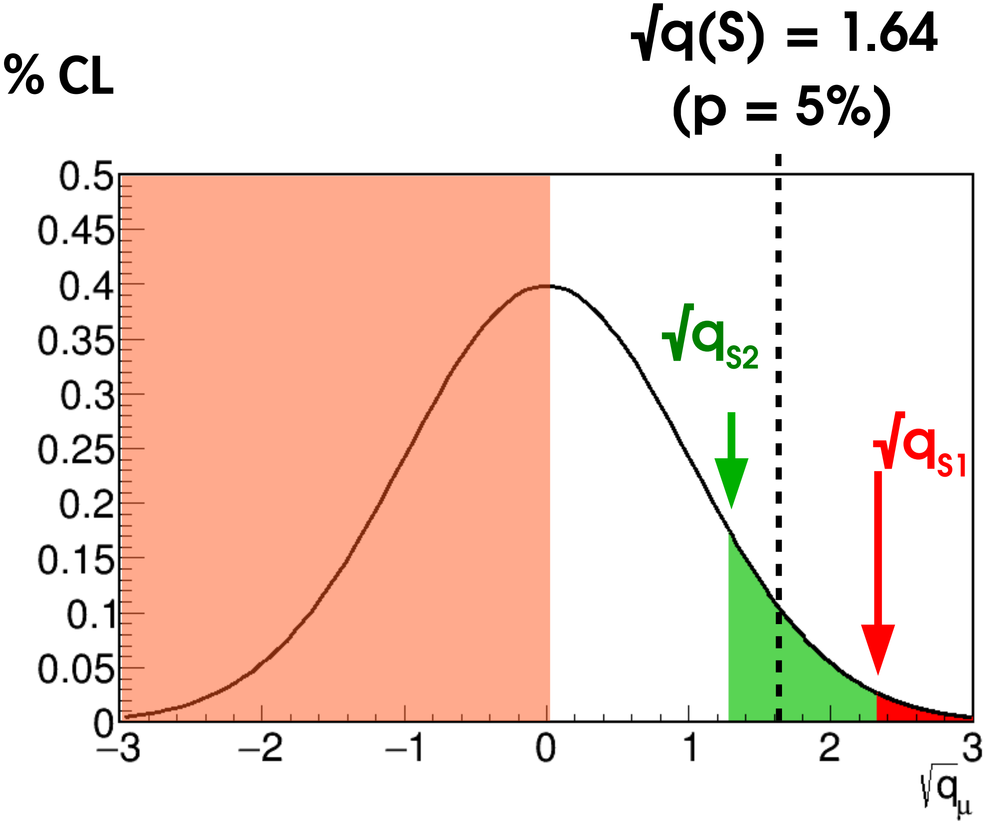
→ Compute $q(S_0)$ for some S_0, get the exclusion p-value $p(S_0)$.

Asymptotics: $$p(S_0) = 1 - \Phi\left(\sqrt{q(S_0)}\right)$$

→ **Adjust S_0** to get the desired exclusion

Asymptotics: need $\sqrt{q(S_{95})} = 1.64$ for **95% CL**





Inversion : Getting the limit for a given CL

Procedure:

CL	p	Region
90%	10%	$\sqrt{q(S)} > 1.28$
95%	5%	$\sqrt{q(S)} > 1.64$
99%	1%	$\sqrt{q(S)} > 2.33$

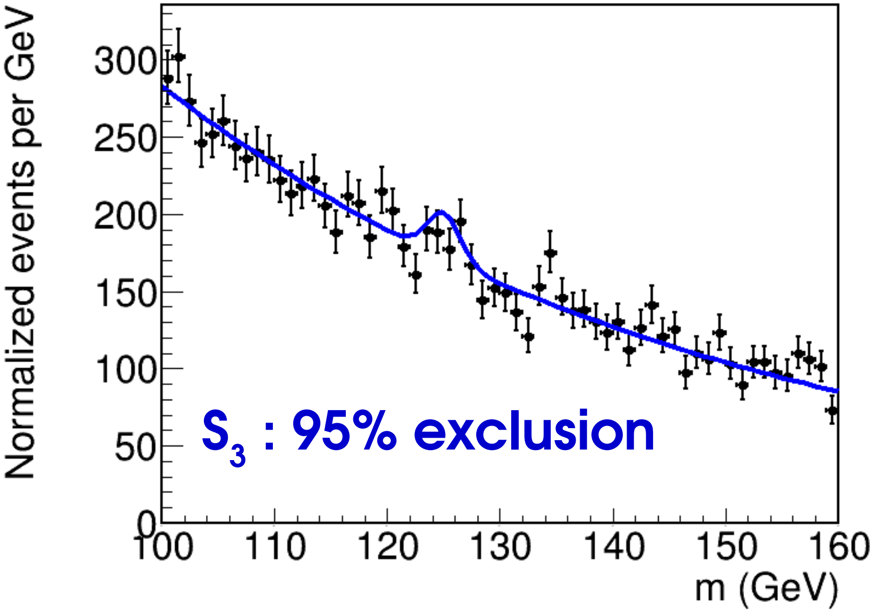
→ Compute $q(S_0)$ for some S_0,
get the **exclusion p-value $p(S_0)$**.

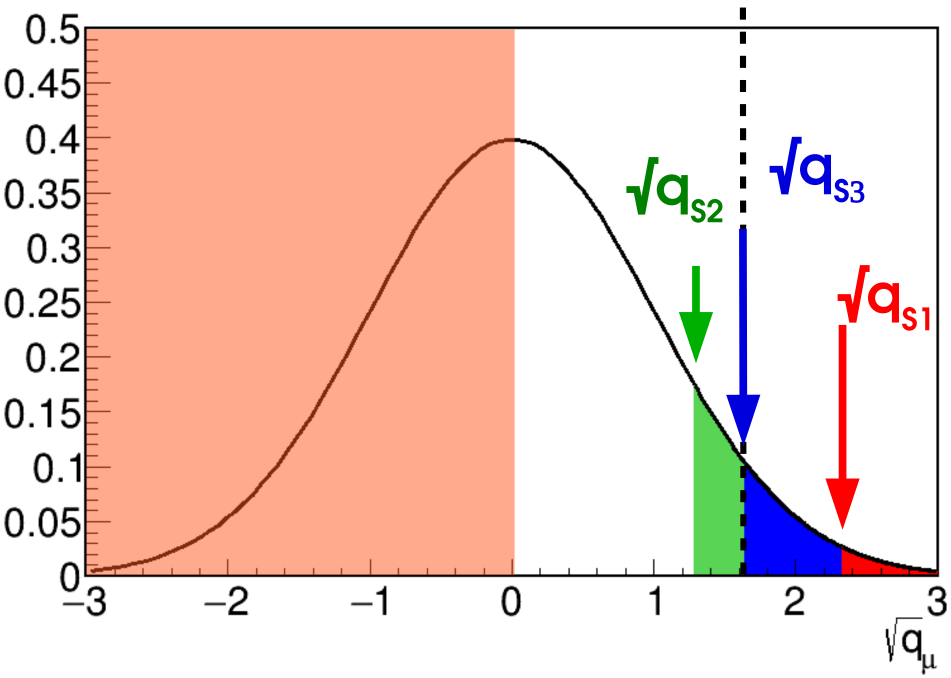
Asymptotics:

$$p(S_0) = 1 - \Phi\left(\sqrt{q(S_0)}\right)$$

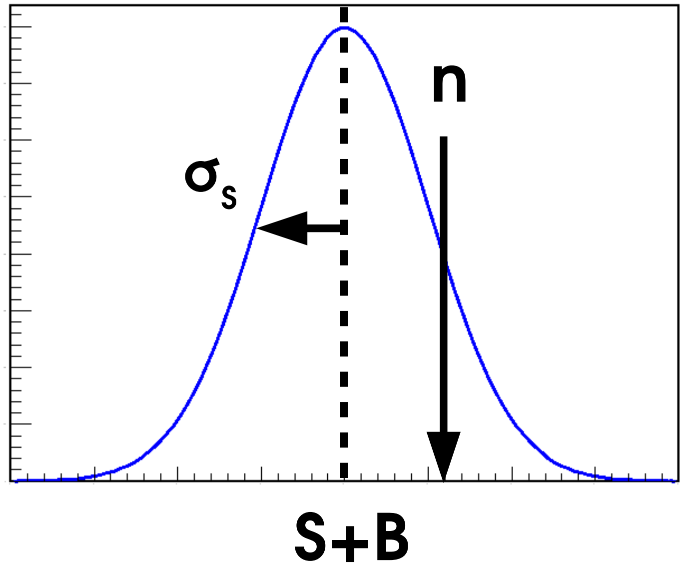
→ **Adjust S_0** to get the desired exclusion
Asymptotics: need **$\sqrt{q(S_{95})} = 1.64$** for **95% CL**

$\sqrt{q(S)} = 1.64$
($p = 5\%$)





Homework 4: Gaussian Example



Usual Gaussian counting example with known B:

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

σ_S ~ √B 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 $\sqrt{q_{S0}} = 1.64$.

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

CL_s

A. Read, J.Phys. G28 (2002) 2693-2704

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

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

Usual p-value for $S=S_0$

P-value for S=0

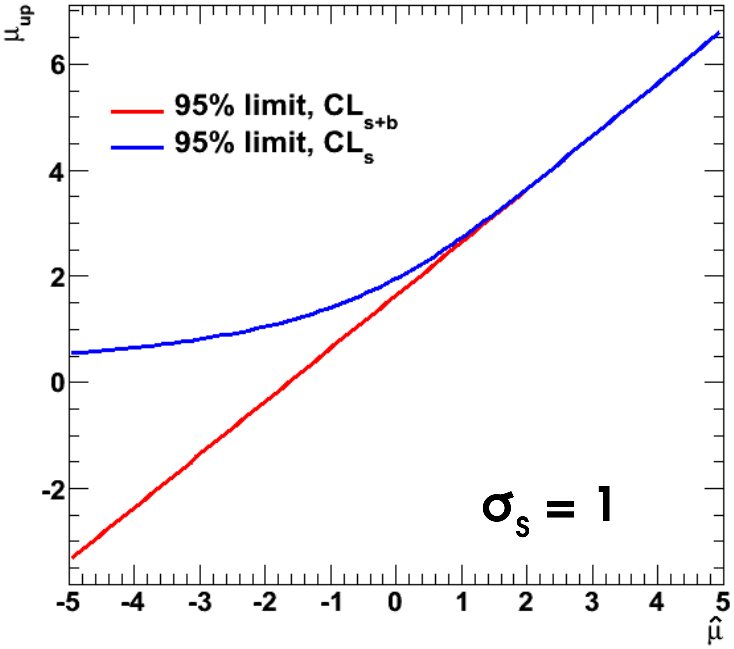
Known feature – to avoid, usual solution in HEP is to use **CL_s** ”modified p-value”

⇒ Compute exclusion relative to that of S=0

→ Somewhat ad-hoc, but good properties...

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

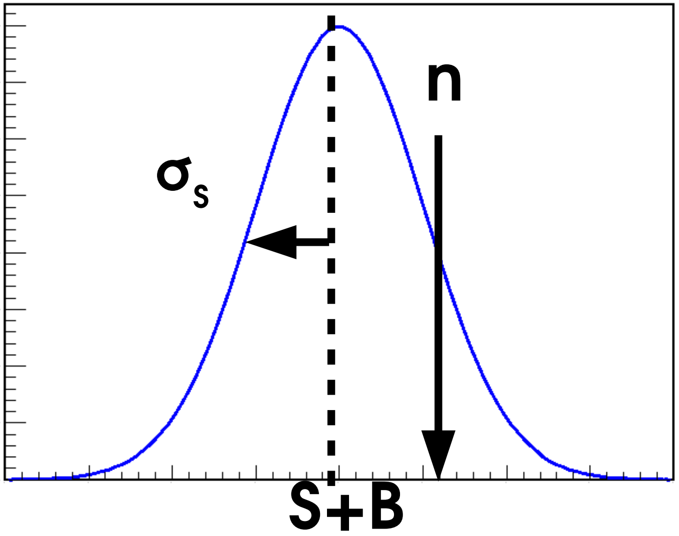
$\hat{S} \ll 0 \Rightarrow p_B \ll 1$, $p_{CLs} \gg p(S_0)$ **no exclusion at S=0**



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:

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

$\sigma_S \sim \sqrt{B}$ for small S

Reminder

CL$_{s+b}$ limit: $\mathbf{S_{up} = \hat{S} + 1.64\,\sigma_S\ \text{at}\ 95\,\%\ \text{CL}}$

CL$_s$ upper limit :

→ Compute p_{S0} (same as for CLs+b)

→ Compute 1-p_B (hard!)

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

for Ŝ ~ 0, $$S_{up} = \hat{S} + 1.96\ \sigma_S\ \text{at}\ 95\,\%\ \text{CL}$$

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% CLs 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

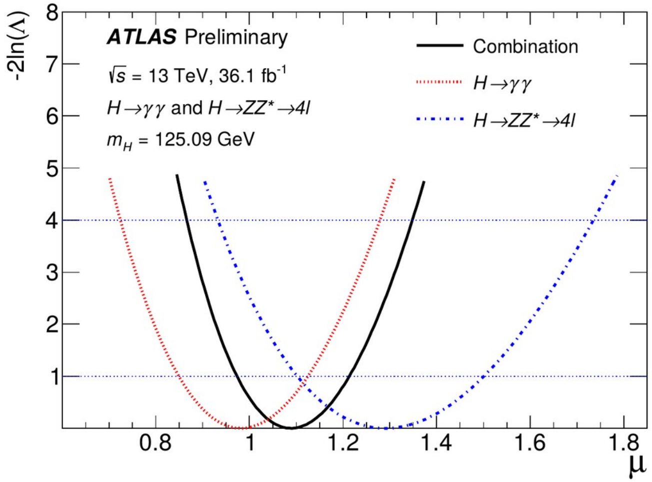
p_{S0} = Poisson(n=0 | S_0+B) and 1 - p_B = Poisson(n=0 | B)

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

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

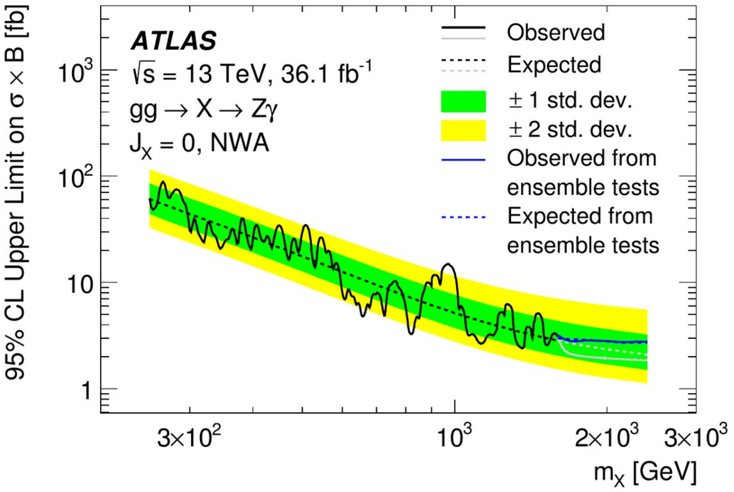


→ Crossings with $t(\mu_0) = 1$ for 1σ intervals (in 1D)

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

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

→ Use **CL_s procedure** to avoid negative limits

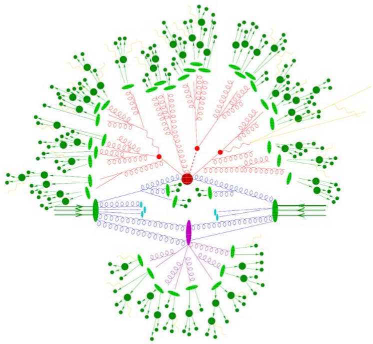


Gaussian regime, n~0: **$S < \hat{S} + 1.96\sigma$ at 95% CL**

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:

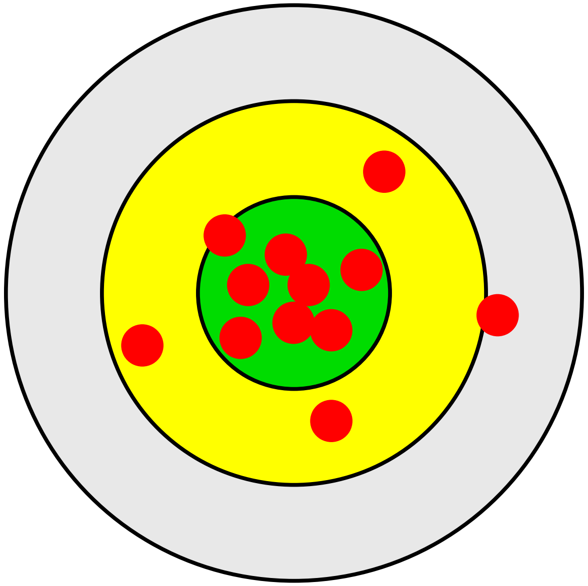
Description	Observable	Likelihood
Counting	**n**	**Poisson** $P(\boldsymbol{n};\boldsymbol{S},\boldsymbol{B})=e^{-(\boldsymbol{S}+\boldsymbol{B})}\frac{(\boldsymbol{S}+\boldsymbol{B})^{\boldsymbol{n}}}{\boldsymbol{n}!}$
Binned shape analysis	n_i, i = 1 .. N_{bins}	**Poisson product** $P(\boldsymbol{n_i};\boldsymbol{S},\boldsymbol{B})=\prod_{i=1}^{n_{\text{bins}}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

Model include

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

Systematic Errors

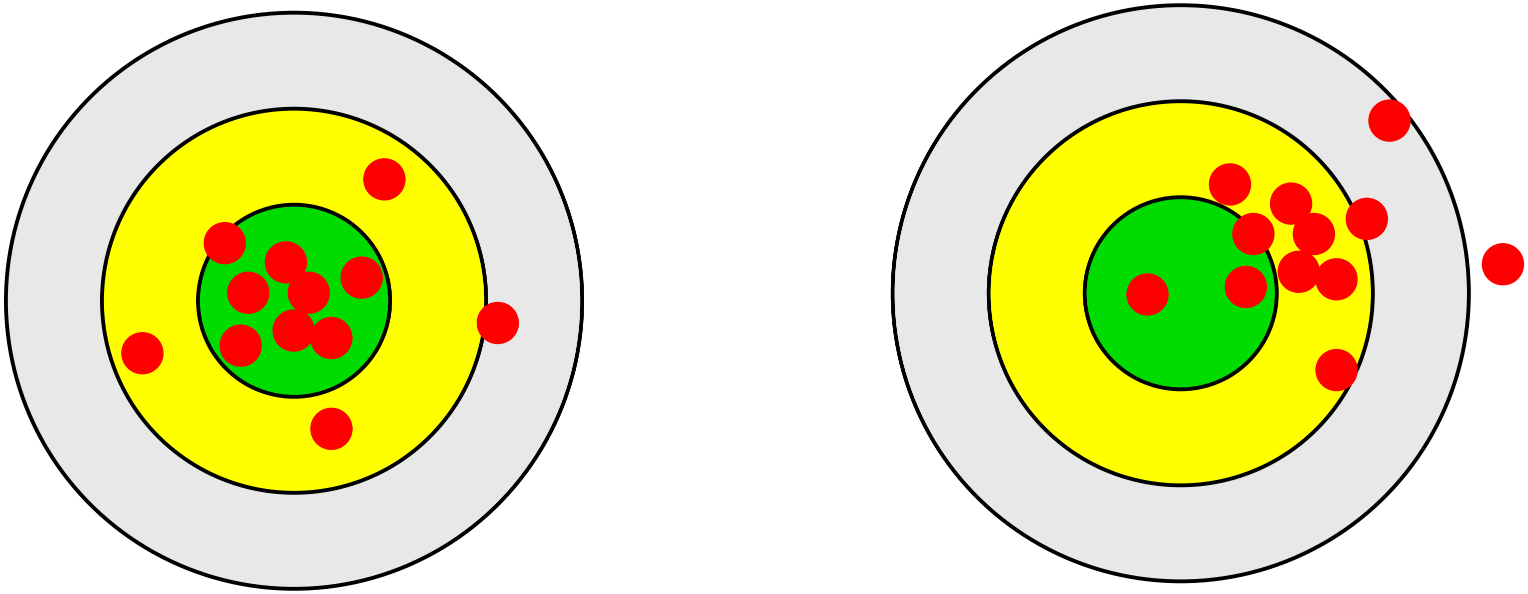
The statistical model (PDF) is a way to express **uncertainty** on the outcome of an experiment. e.g. 2D Gaussian :



These uncertainties are also called **Statistical Uncertainties** – they are the ones encoded in the model PDF.

Systematic Errors

The statistical model (PDF) is a way to express **uncertainty** on the outcome of an experiment. e.g. 2D Gaussian :



These uncertainties are also called **Statistical Uncertainties** – they are the ones encoded in the model PDF.

However **the model itself may be wrong** : this is a *systematic error*
→ To account for them, need a set of **Systematic uncertainties,** i.e. uncertainties *on the form of the PDF itself*.

Systematics

> "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 ?***

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

How to describe them in practice ?

Þ 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

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

POI

Systematics NP

Measurement Likelihood

NP Constraint term

C(θ) represents **external knowledge** about the NPs that we inject into the statistical model – e.g. to say that “**B ~ 100 ± 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**

⇒ Can be 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*)**

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

⇒ **Profile Likelihood Ratio (PLR)**

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

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

Wilks' Theorem* : *same properties as plain likelihood ratio without NPs

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

also with NPs present

→ Profiling "builds in" the effect of the NPs

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

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

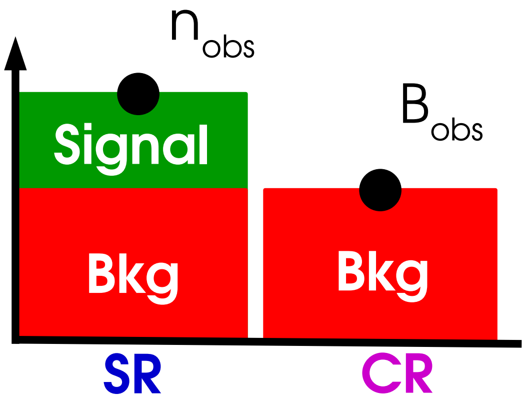
$$L(S, B) = G(n_{obs}; S + B, \sigma_{stat})\, G(B_{obs}; B, \sigma_{bkg})$$

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

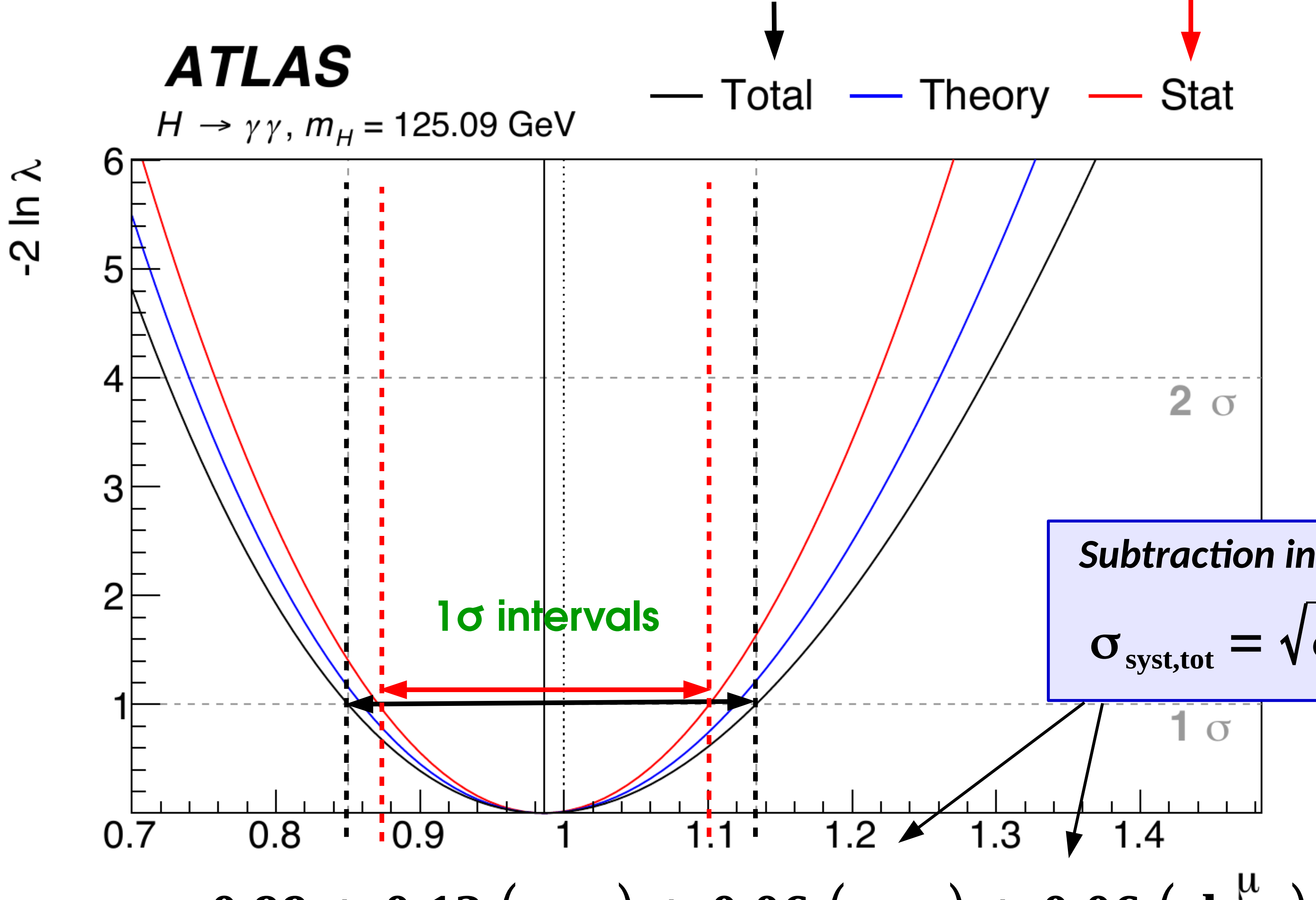
Answer: $S = (n_{obs} - B_{obs}) \pm \sqrt{\sigma_{stat}^2 + \sigma_{bkg}^2} \qquad \sigma_S = \sqrt{\sigma_{stat}^2 + \sigma_{bkg}^2}$

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

Uncertainty decomposition

All systematics NPs excluded : statistical uncertainty only

All systematics NPs included: stat+syst uncertainties



Subtraction in quadrature

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

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

Pull/Impact plots

ATLAS-CONF-2016-058

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

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

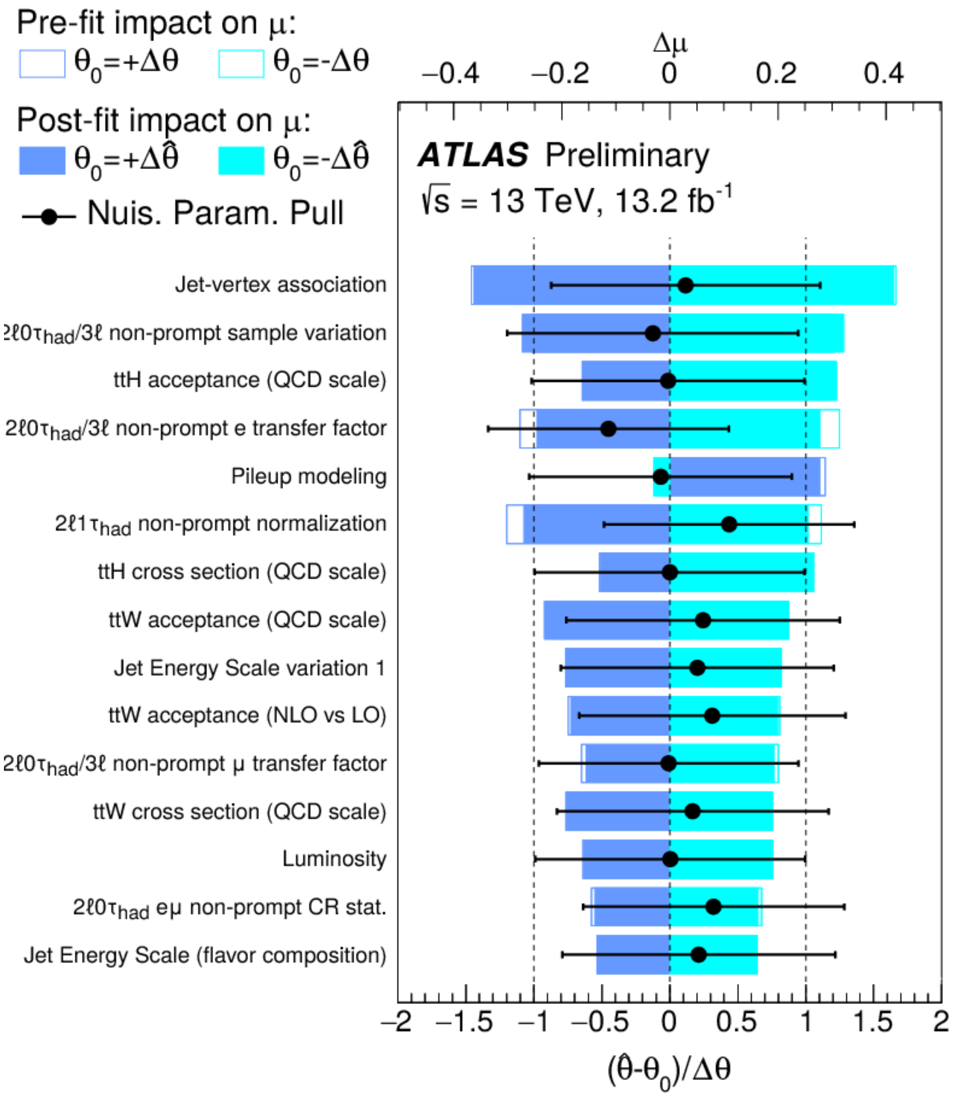
Nominally:

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

However fit results may be different:

- **Central value ≠ 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 ±1σ 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

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

Nominally:

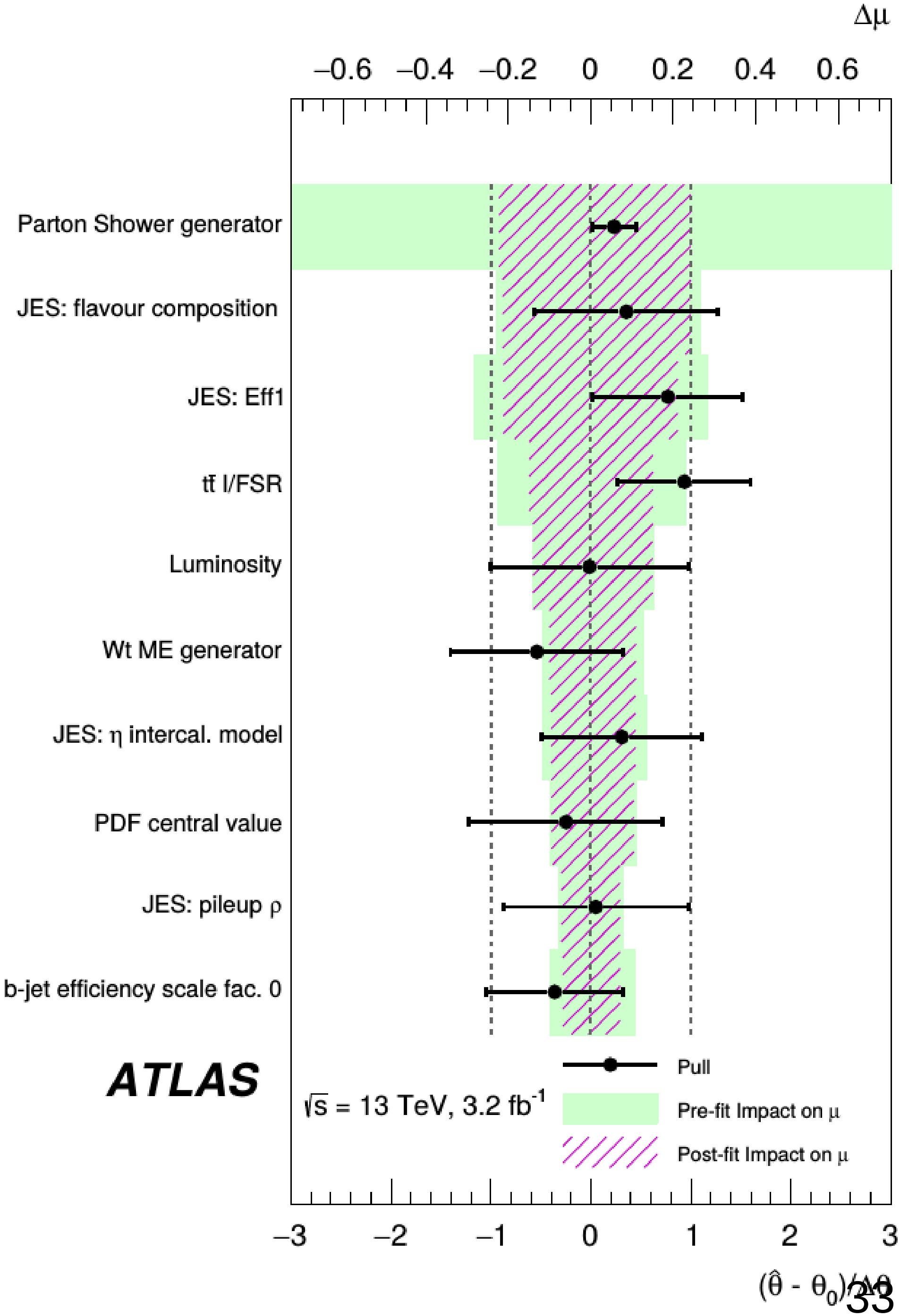
- **pull = 0** : i.e. the pre-fit expectation
- **pull uncertainty = 1** : from the Gaussian

However fit results may be different:

- **Central value ≠ 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 ±1σ shift of NP allows to gauge which NPs matter most .

13 TeV single-t XS (arXiv:1612.07231)



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:

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

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

Bayesian Analysis

Bayesian methods

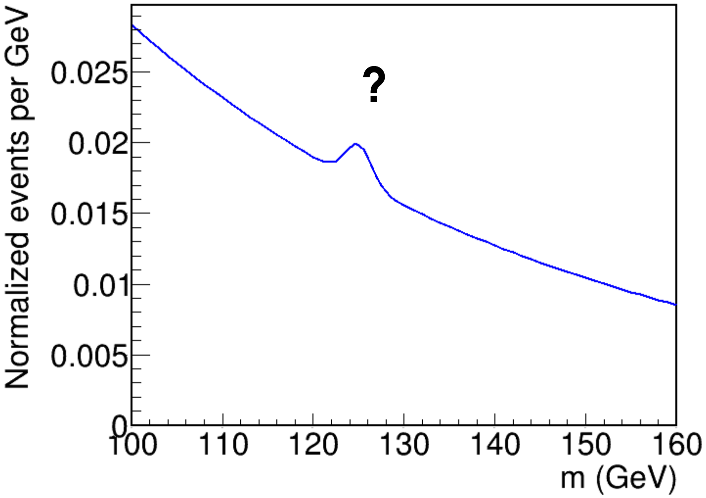
Remember the problem from yesterday:

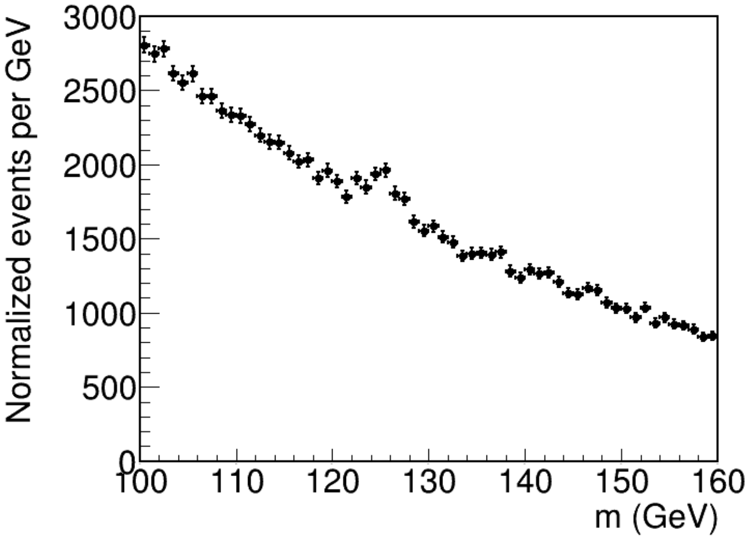
- PDFs give possible outcomes for known parameters
- We already know the outcome, and want information on the parameters

$$P(\lambda = ?)$$

2

Estimate





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

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

Prior PDF on μ: represents our knowledge before the measurement

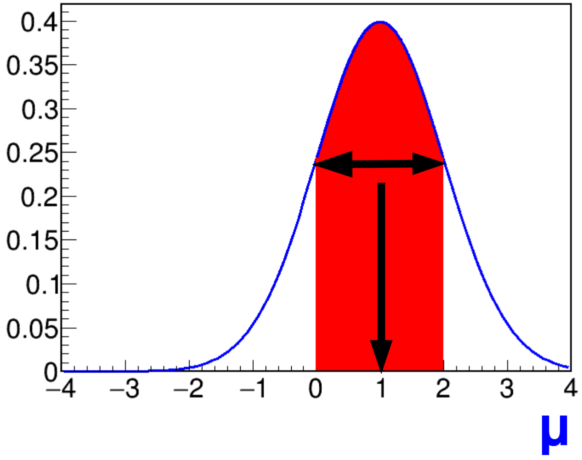
Posterior PDF: represents our total knowledge from prior + measurement

Measurement PDF, same as for the frequentist P(n;μ)

Normalization factor: adjusted so P(μ|n) is normalized to 1)

Immediately useful to get intervals on μ:

- Peak of P(μ|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(μ, θ | n)**

Define a prior **π(θ)** for each nuisance parameter.

⇒ Obtain **P(μ|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(μ) to compute intervals and limits as before.

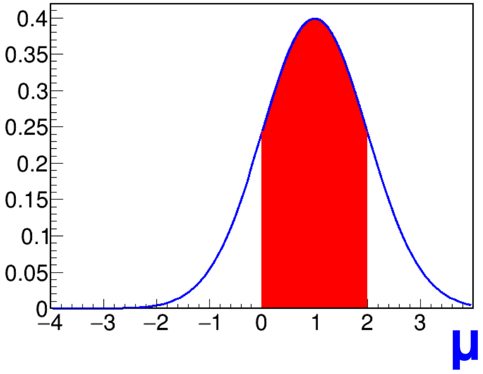
68.3% CL interval:

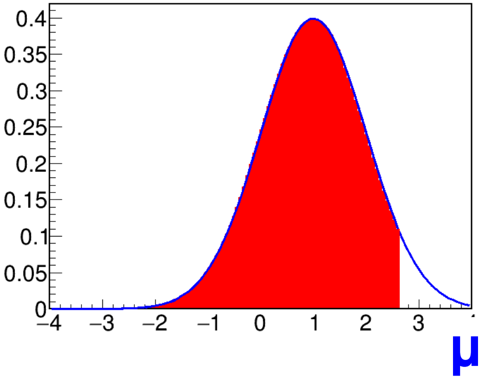
$$\int_A^B P(\mu|n)\, d\mu = 68.3\,\%$$

Here CL means “Credibility Level”)

95% CL upper limit

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





Bayesian vs. frequentist

> *"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

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!

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 flat prior over S>0.
- Find the 95% CL limit by solving $\int_{S_{up}}^{\infty} P(S|n)\,dS = 5\,\%$

Solution:

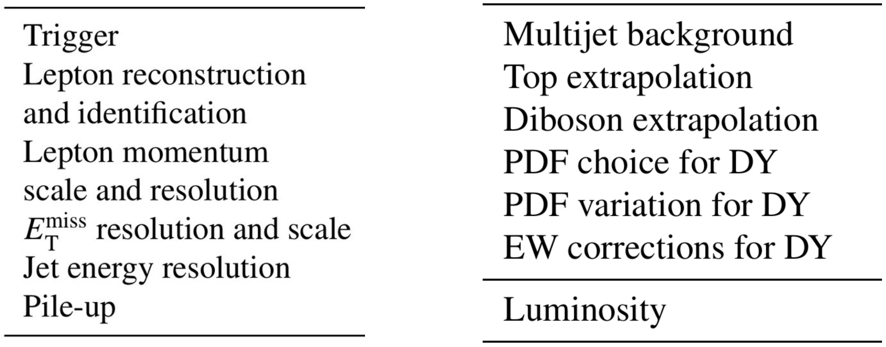
In both cases

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

Example: W'→lν Search

arXiv:1706.04786

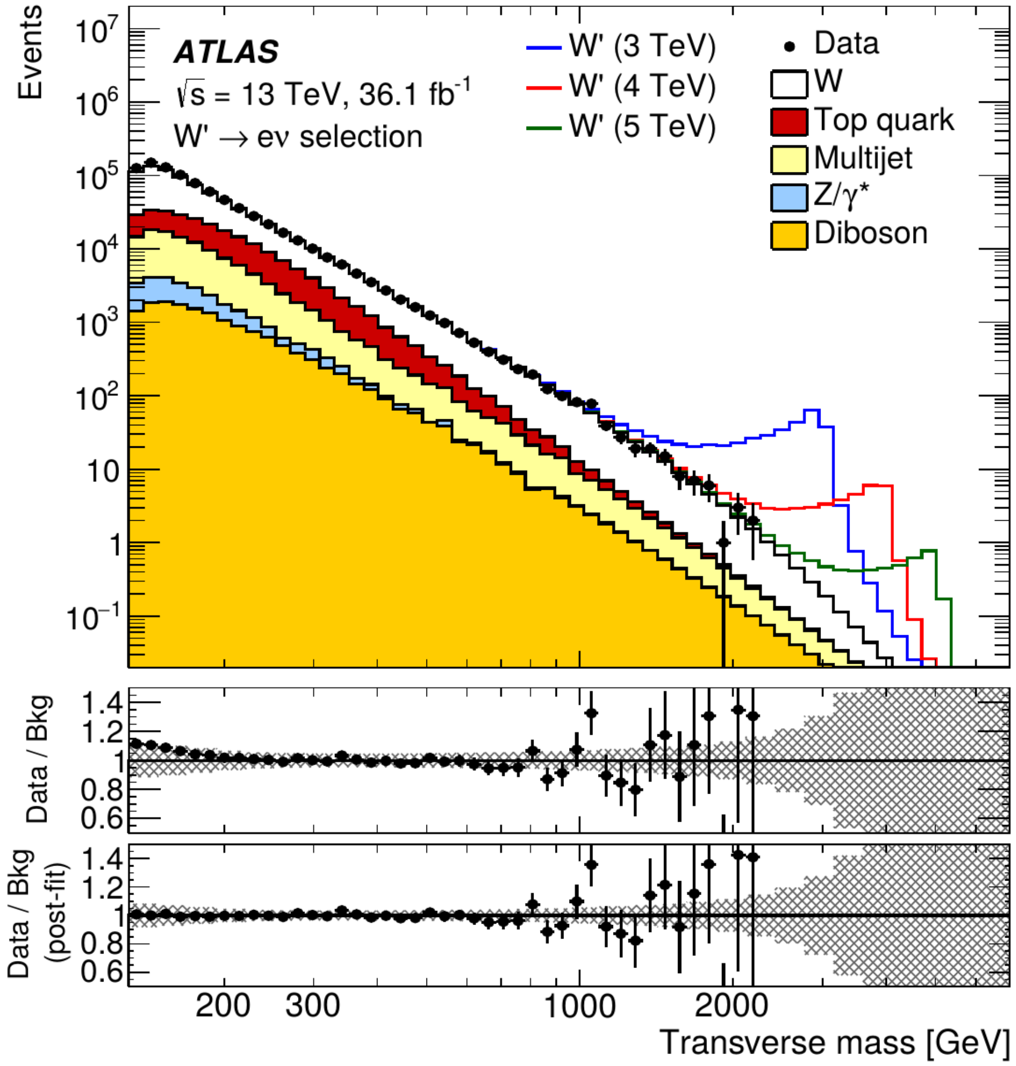
- **POI**: **W' σ × B** → use **flat prior over [0, +inf[**.
- **NPs**: syst on **signal ε** (6 NPs), **bkg** (6), **lumi** (1) → integrate over Gaussian priors

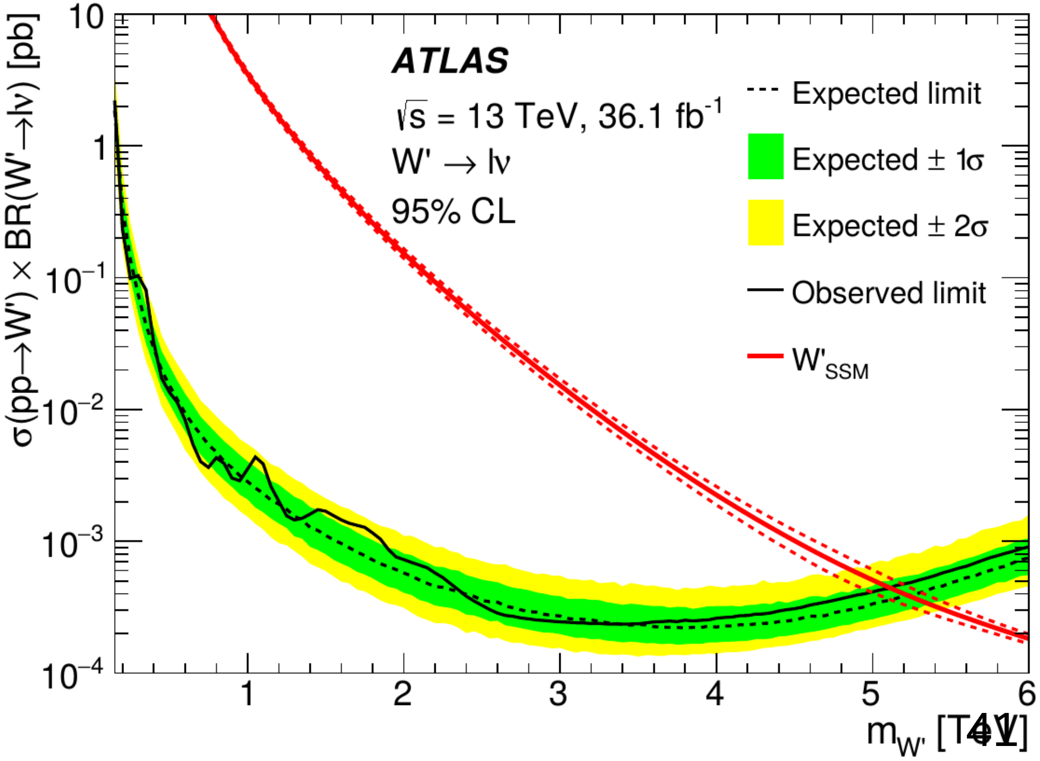


Trigger
Lepton reconstruction and identification
Lepton momentum scale and resolution
$E_{\mathrm{T}}^{\mathrm{miss}}$ resolution and scale
Jet energy resolution
Pile-up

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

Luminosity



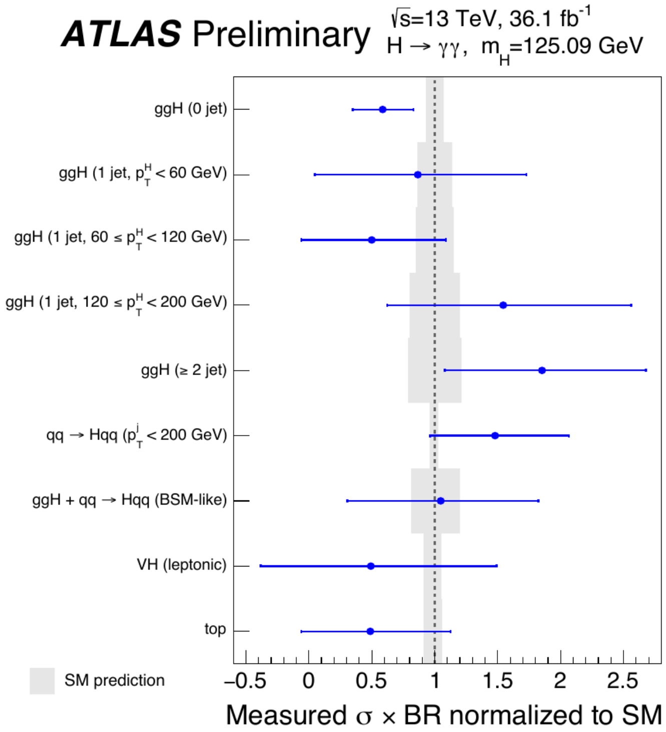


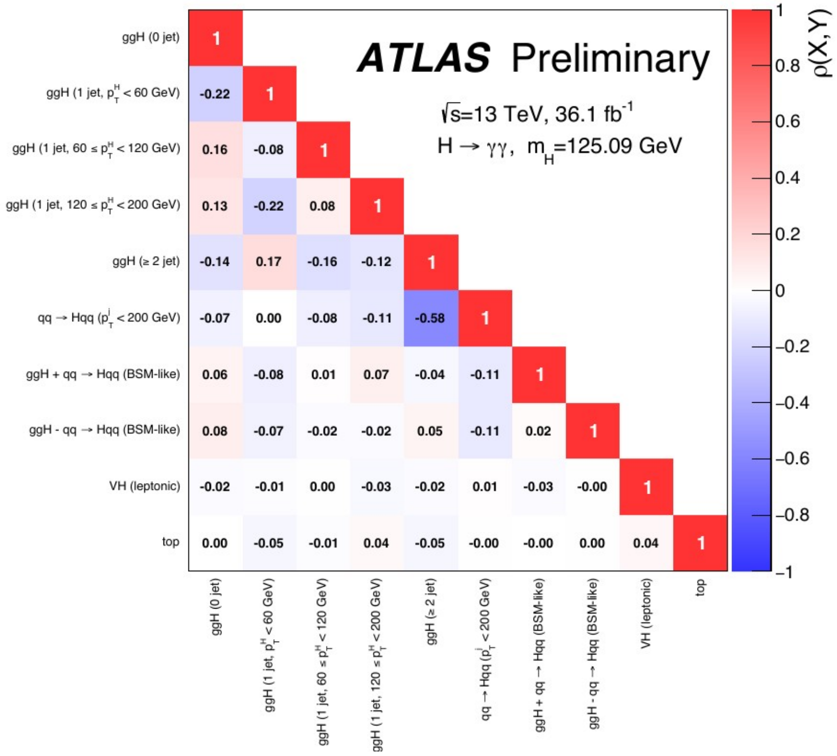
Presentation of Results

→ 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.





	ggH (0 jet)	ggH (1 jet, p_T^H < 60 GeV)	ggH (1 jet, 60 ≤ p_T^H < 120 GeV)	ggH (1 jet, 120 ≤ p_T^H < 200 GeV)	ggH (≥ 2 jet)	qq → Hqq (p_T^j < 200 GeV)	ggH + qq → Hqq (BSM-like)	ggH - qq → Hqq (BSM-like)	VH (leptonic)	top
ggH (0 jet)	1									
ggH (1 jet, p_T^H < 60 GeV)	-0.22	1								
ggH (1 jet, 60 ≤ p_T^H < 120 GeV)	0.16	-0.08	1							
ggH (1 jet, 120 ≤ p_T^H < 200 GeV)	0.13	-0.22	0.08	1						
ggH (≥ 2 jet)	-0.14	0.17	-0.16	-0.12	1					
qq → Hqq (p_T^j < 200 GeV)	-0.07	0.00	-0.08	-0.11	-0.58	1				
ggH + qq → Hqq (BSM-like)	0.06	-0.08	0.01	0.07	-0.04	-0.11	1			
ggH - qq → Hqq (BSM-like)	0.08	-0.07	-0.02	-0.02	0.05	-0.11	0.02	1		
VH (leptonic)	-0.02	-0.01	0.00	-0.03	-0.02	0.01	-0.03	-0.00	1	
top	0.00	-0.05	-0.01	0.04	-0.05	-0.00	-0.00	0.00	0.04	1

Non-Gaussian case: not so simple, but can publish full likelihood (e.g. here)

Generating Pseudo-data

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

Þ Possible outcomes of the experiment, for given parameter values

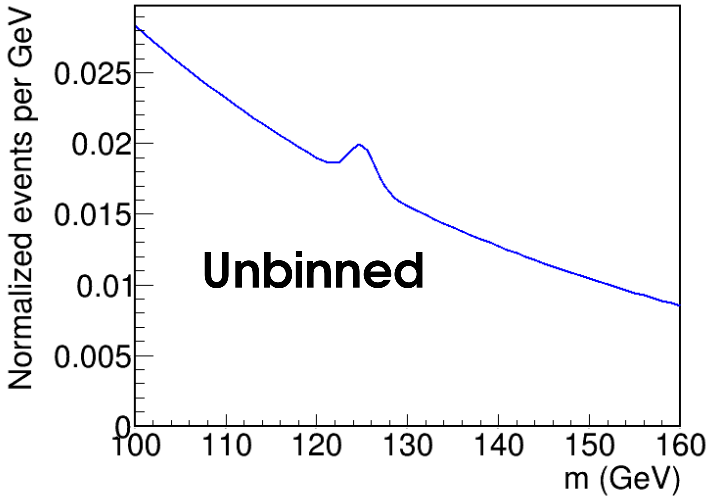
Can draw random events according to PDF : **generate *pseudo-data***

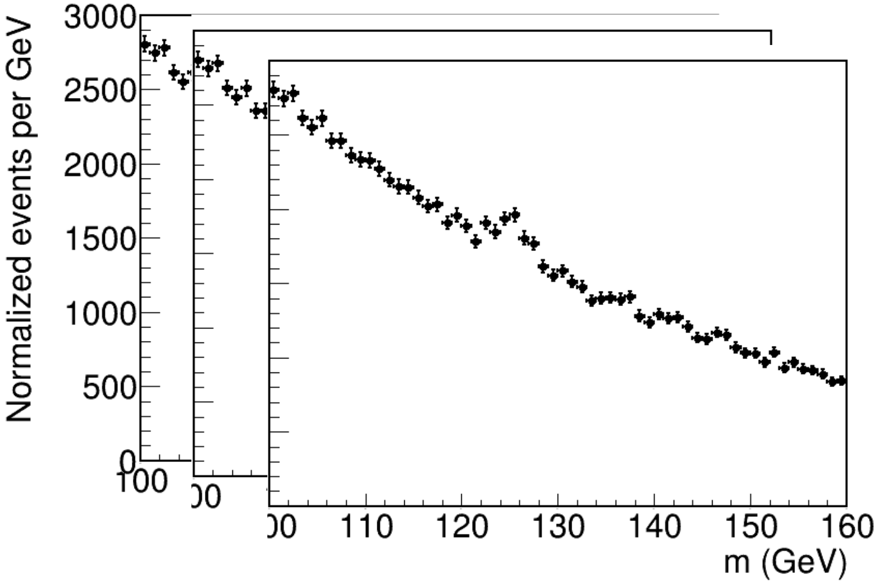
$$P(\lambda=5)$$

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

Each entry = separate "experiment"

Generate





Expected Limits: Toys

***Expected* results**: median outcome under a given hypothesis

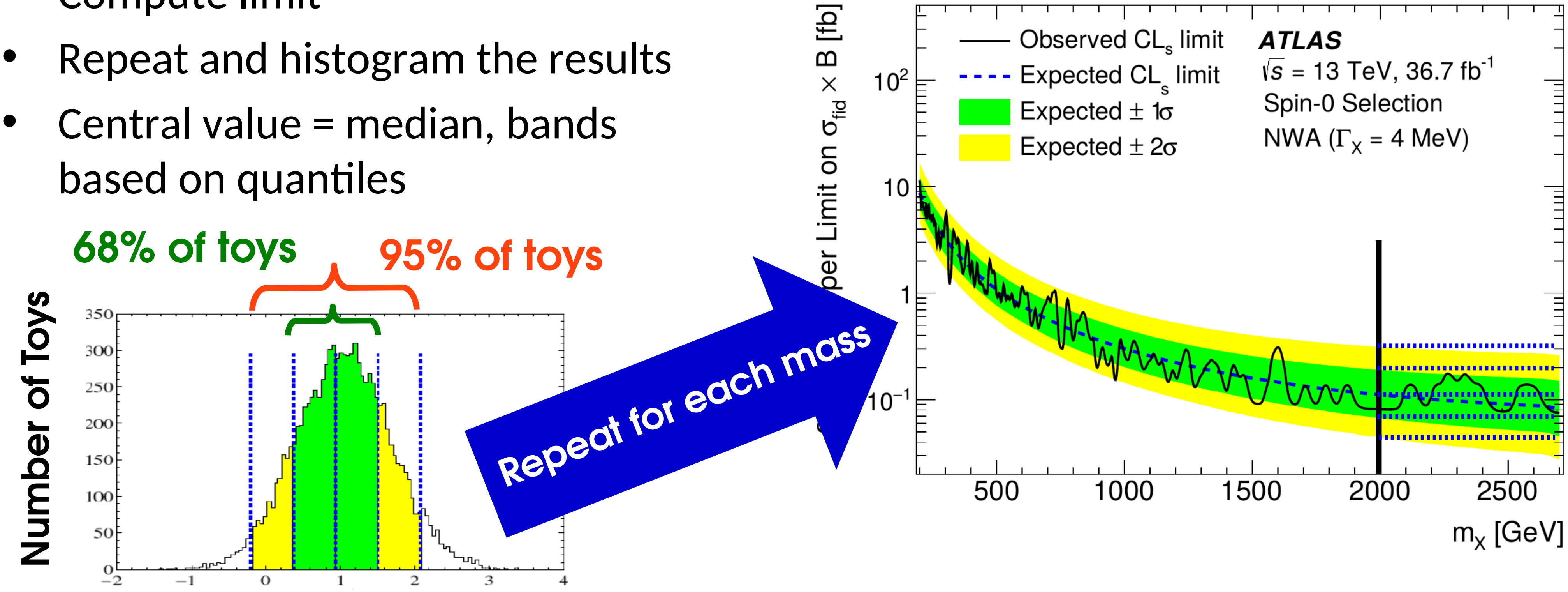
→ usually B-only for searches, but other choices possible.

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



Eur.Phys.J.C71:1554,2011

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

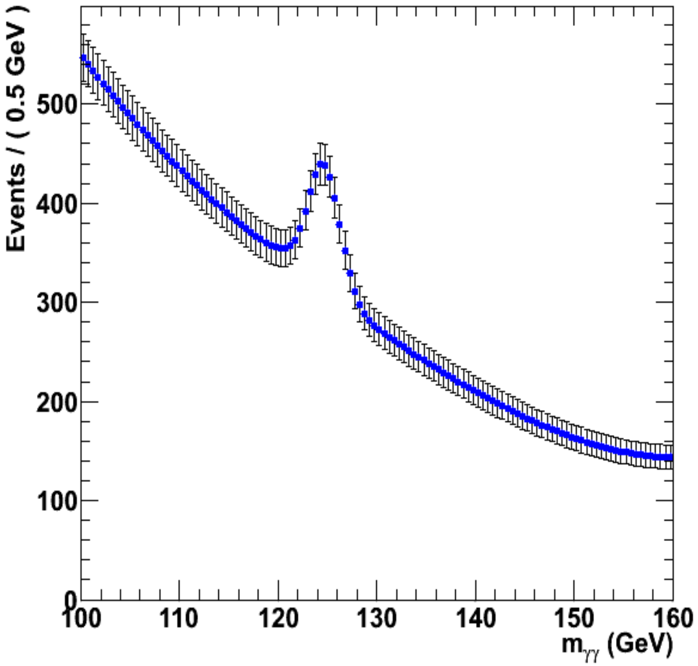
$\hat{X} = X_0$ for all parameters X,

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

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



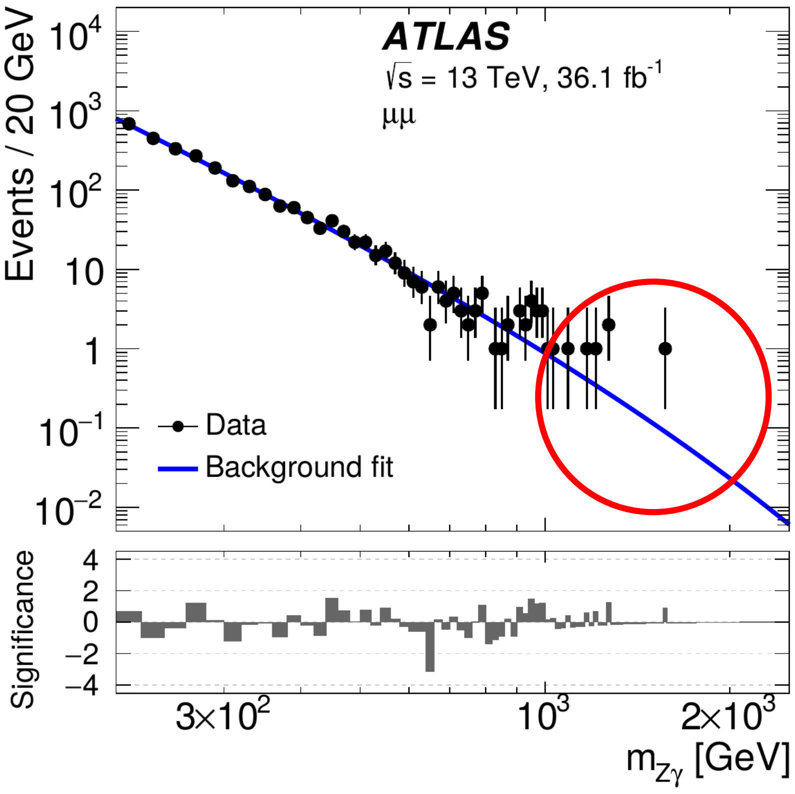
⊕ Much faster (1 “toy”)

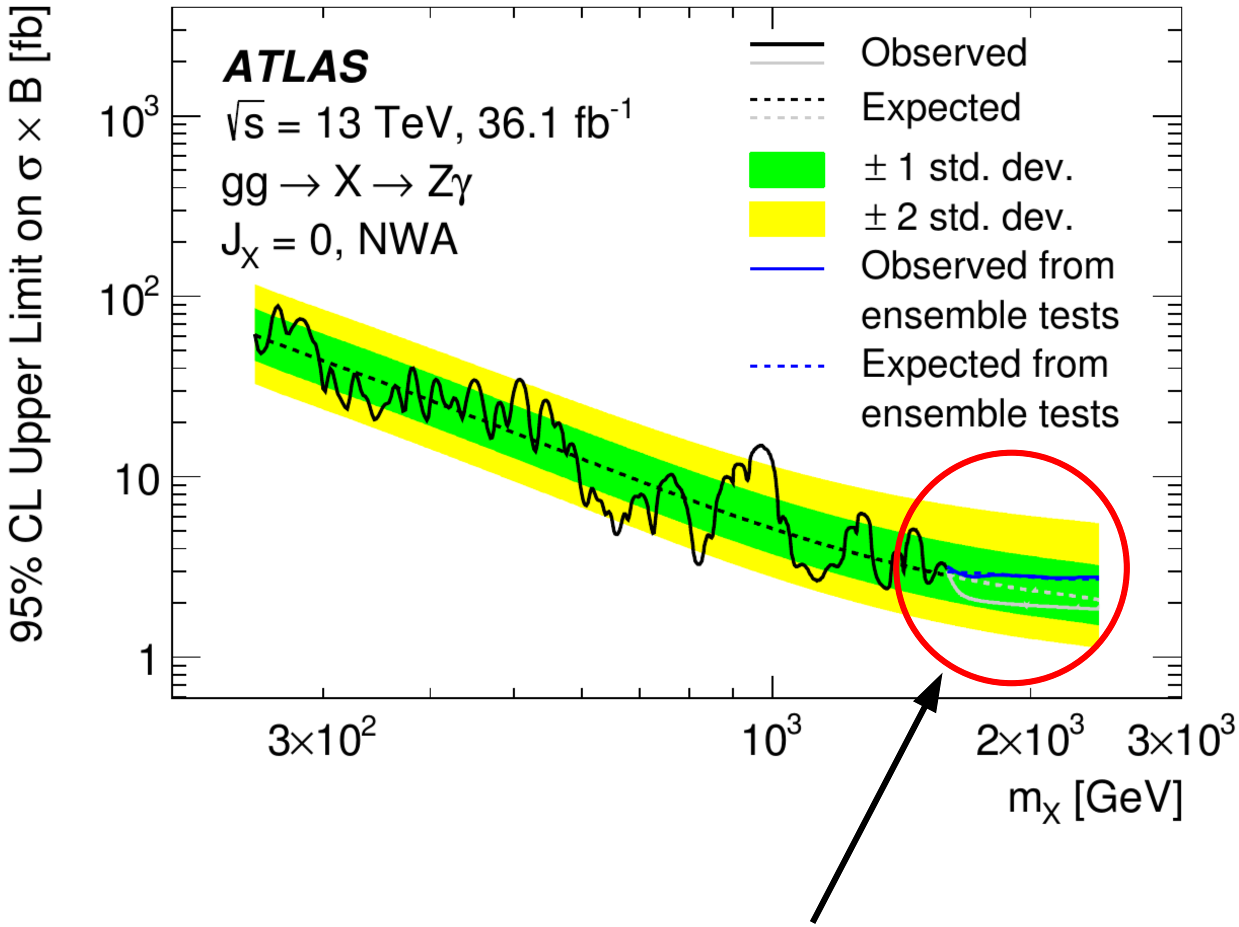
Toys: Example

JHEP 10 (2017) 112

ATLAS X→Zγ Search: covers 200 GeV < m_X < 2.5 TeV

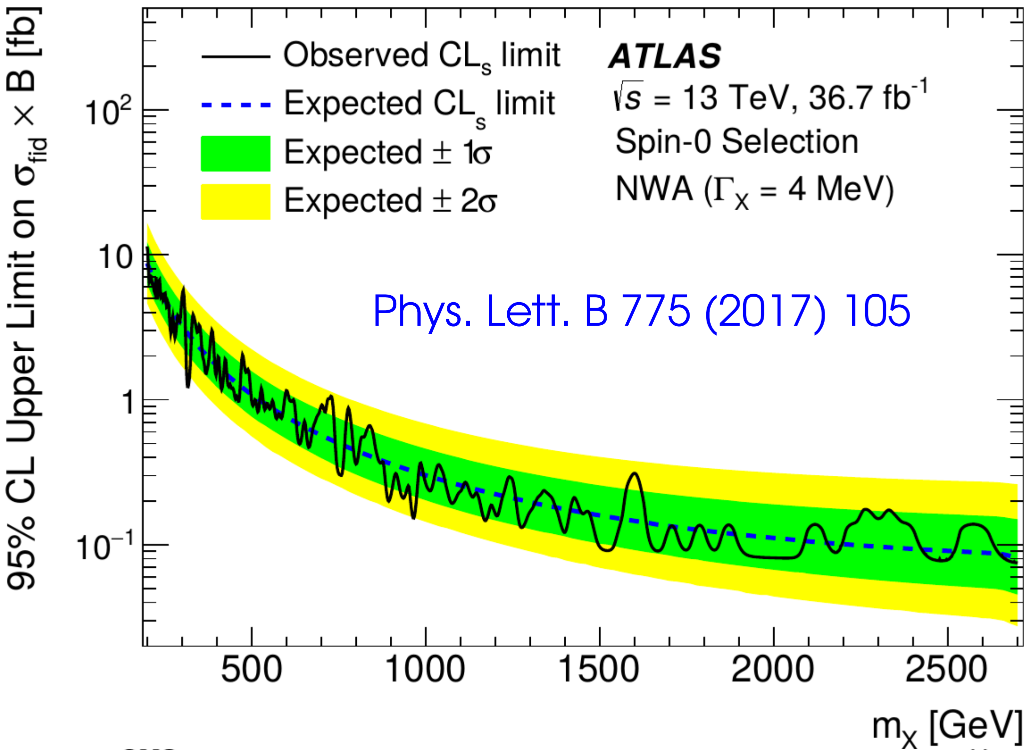
→ for m_X > 1.6 TeV, low event counts ⇒ derive results from toys





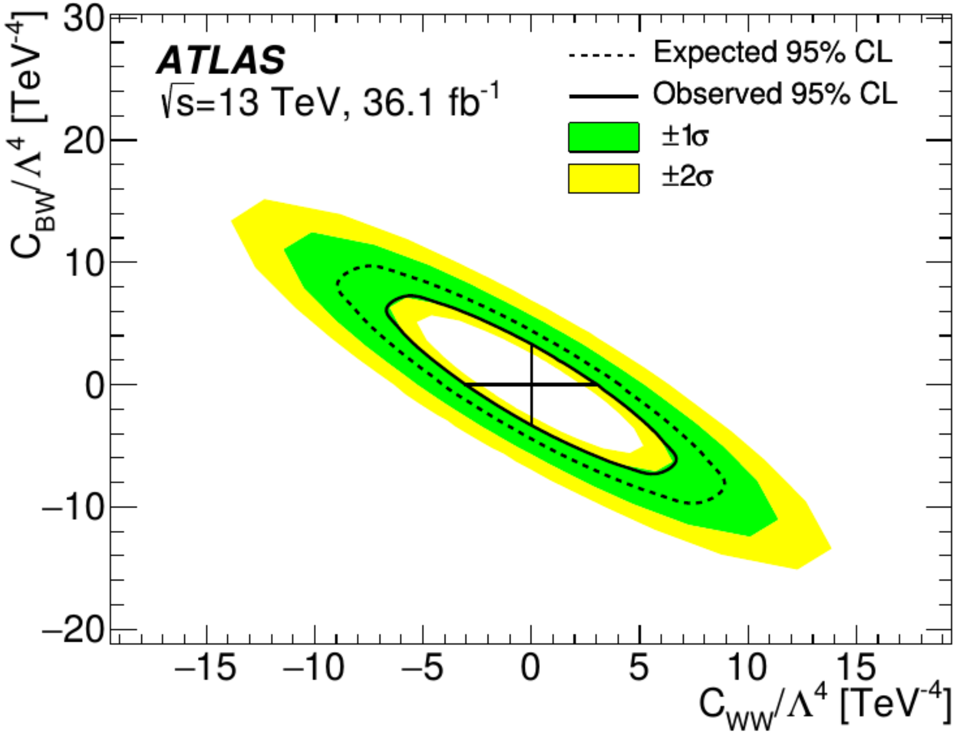
Asimov results (**in gray**) give optimistic result compared to toys (**in blue**)

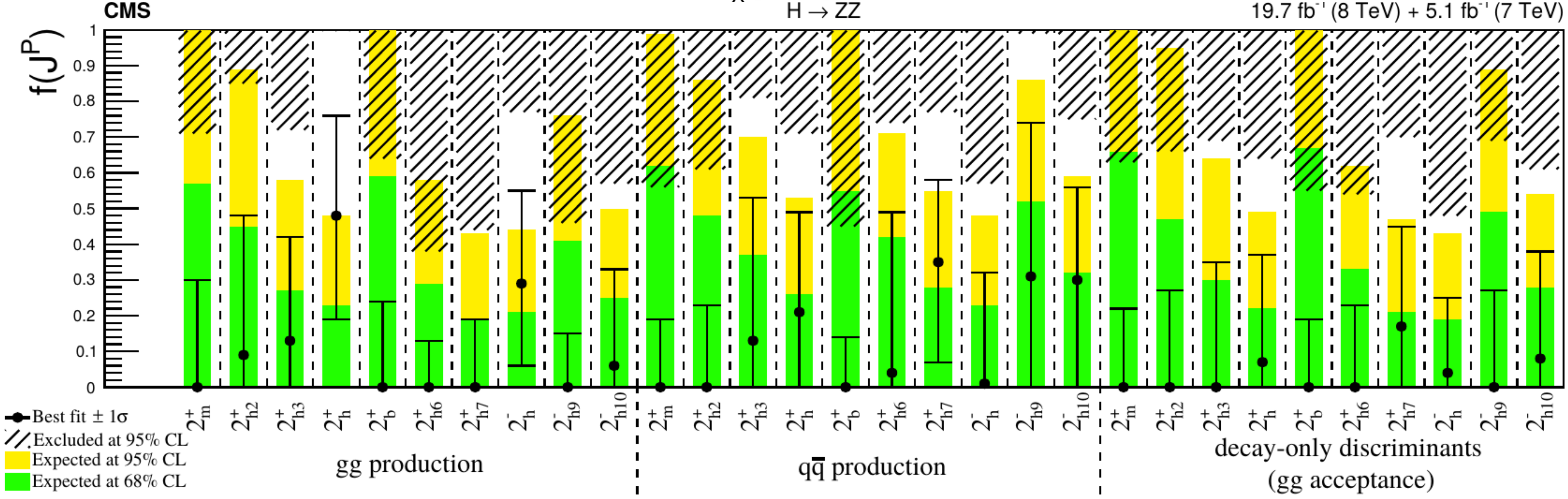
Upper Limit Examples



Phys. Lett. B 775 (2017) 105

ATLAS 2015-2016 4l aTGC Search

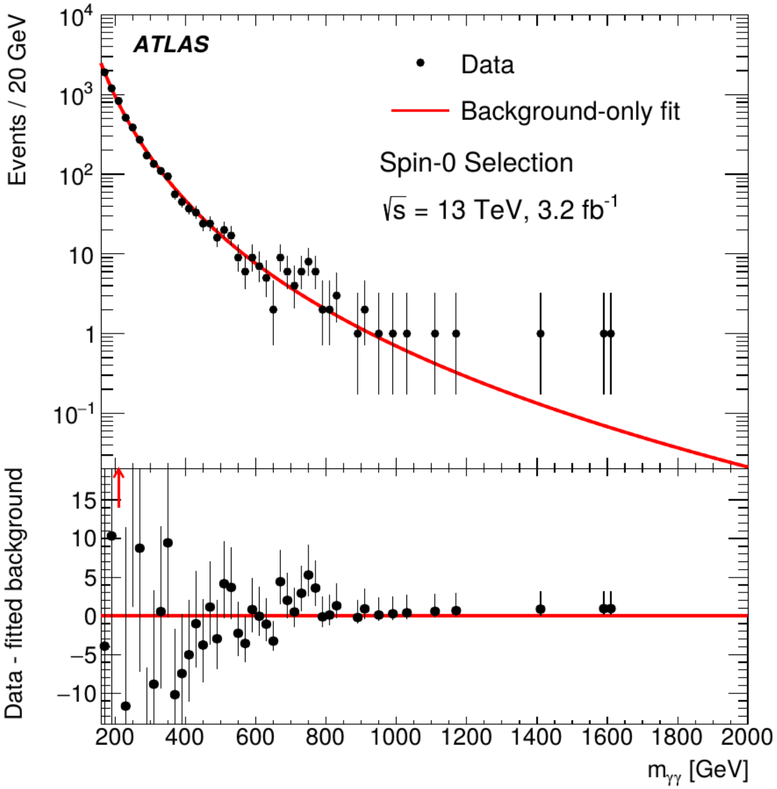




Phys. Rev. D 92 (2015) 012004

Look-Elsewhere Effect

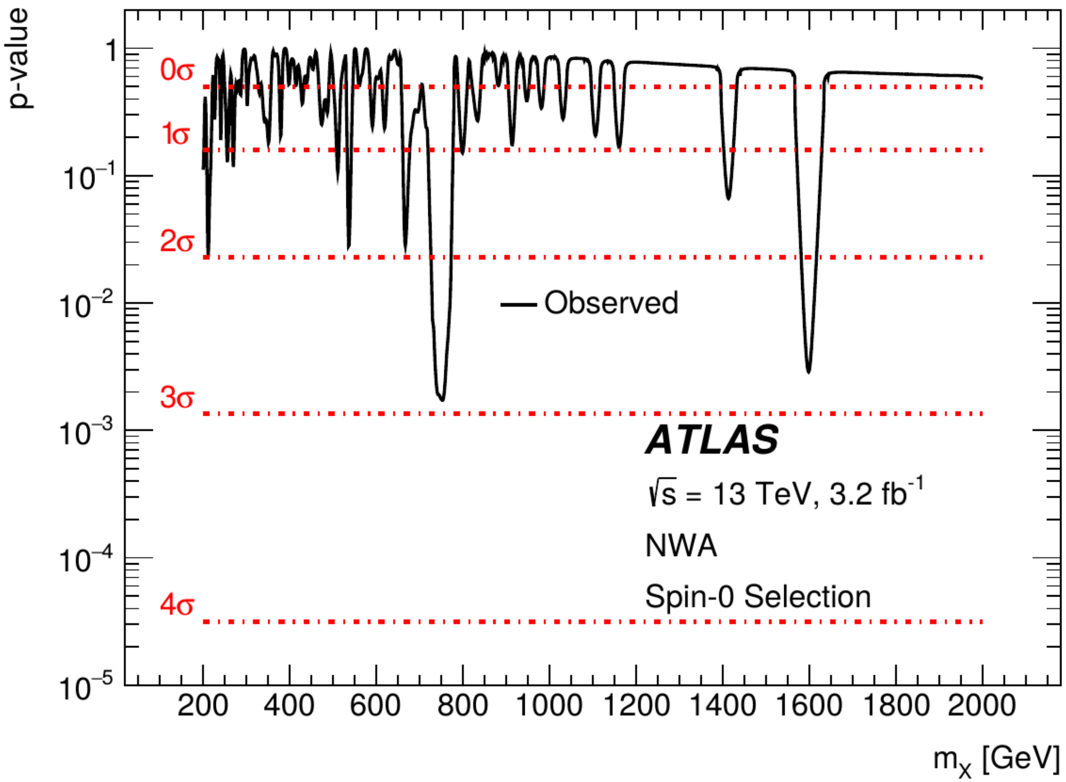
Look-Elsewhere effect



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**

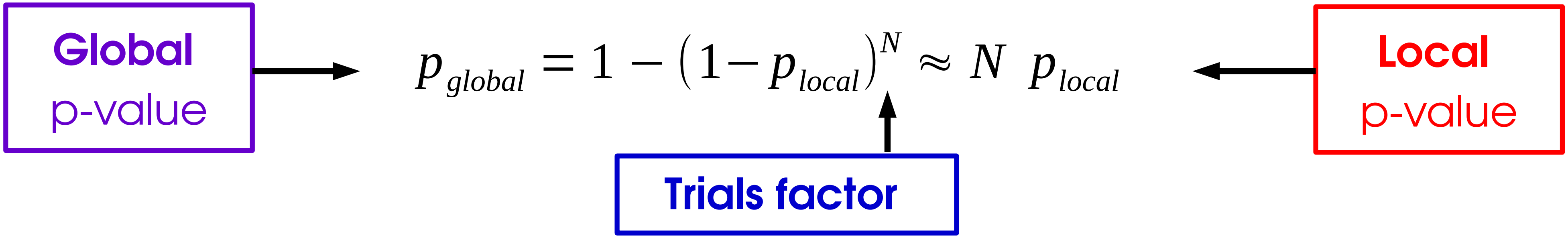


→ More likely to find an excess **anywhere in the range**, rather than in a **predefined** location

⇒ ***Look-elsewhere effect*** (LEE)

Global Significance

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



Global p-value

$$p_{global} = 1-(1-p_{local})^N \approx N\ p_{local}$$

Local p-value

Trials factor

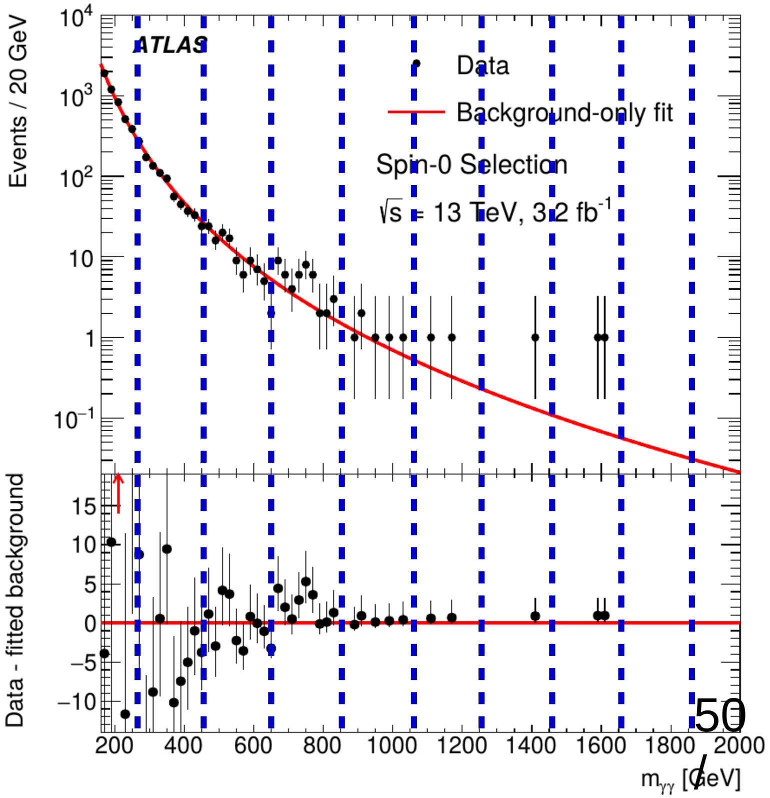
→ $\mathbf{p_{global} > p_{local}}$ ⇒ $\mathbf{Z_{global} < Z_{local}}$: global fluctuation more likely ⇒ less significant

⚠

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

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

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

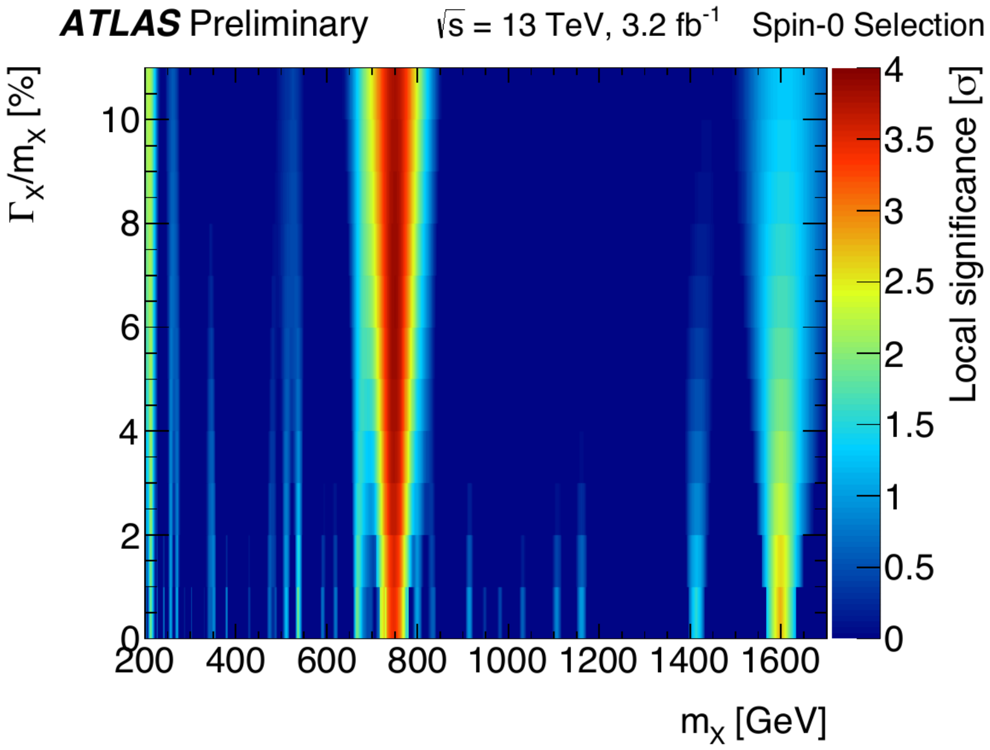


Global Significance

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. $X \to \gamma\gamma$:

- $200 < m_X < 2000$ 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}) ?

Trials Factor

Trials factor N = # of independent searches:

Global p-value

$$p_{global} = 1 - \left(1 - p_{local}\right)^{N} \approx N\ p_{local}$$

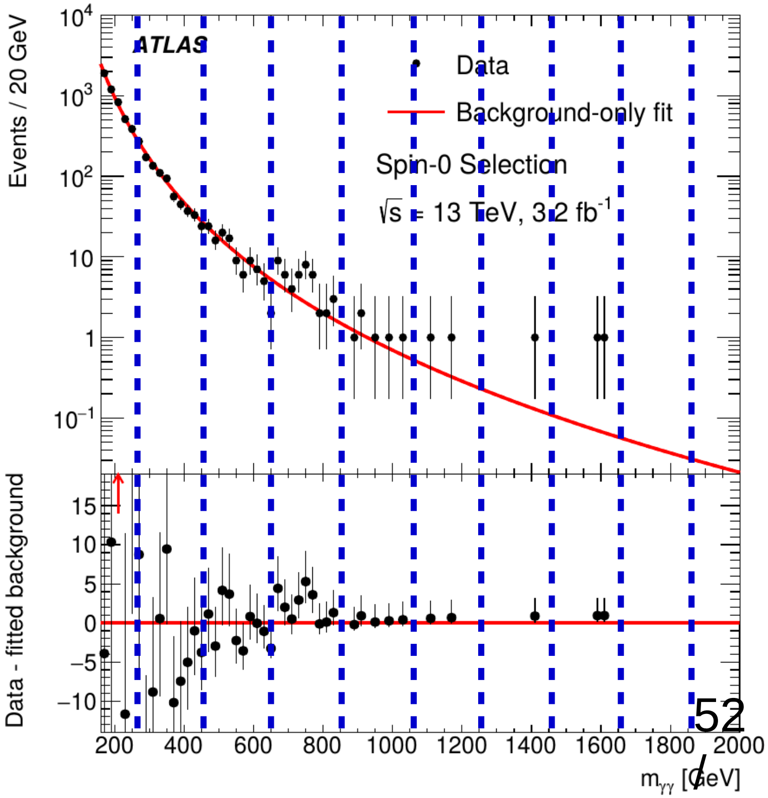
Local p-value

Trials factor

Naively, one could expect

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

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



Trials Factor for continuous variables

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

Asymptotic limit : trials factor (1 POI) is

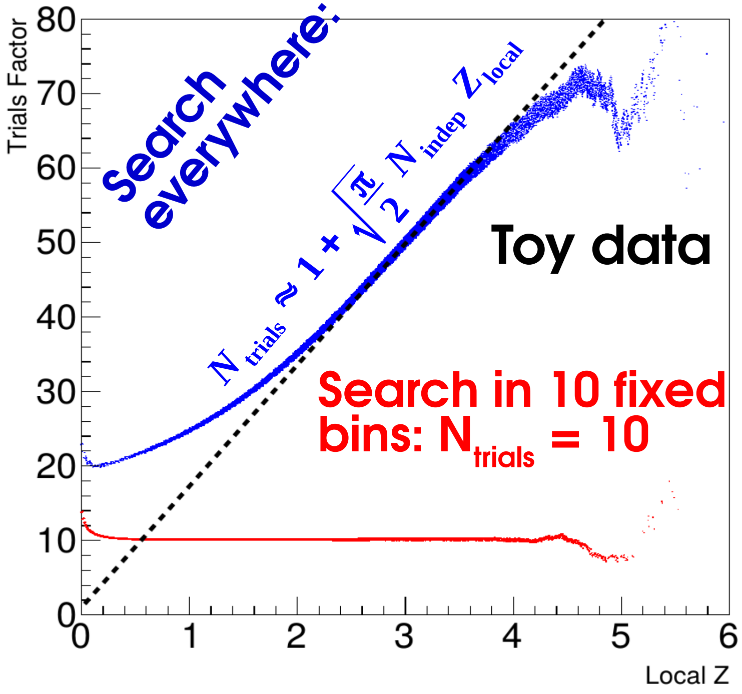
$$N_{\text{trials}} = 1 + \sqrt{\frac{\pi}{2}}\, N_{\text{indep}}\, Z_{\text{local}}$$

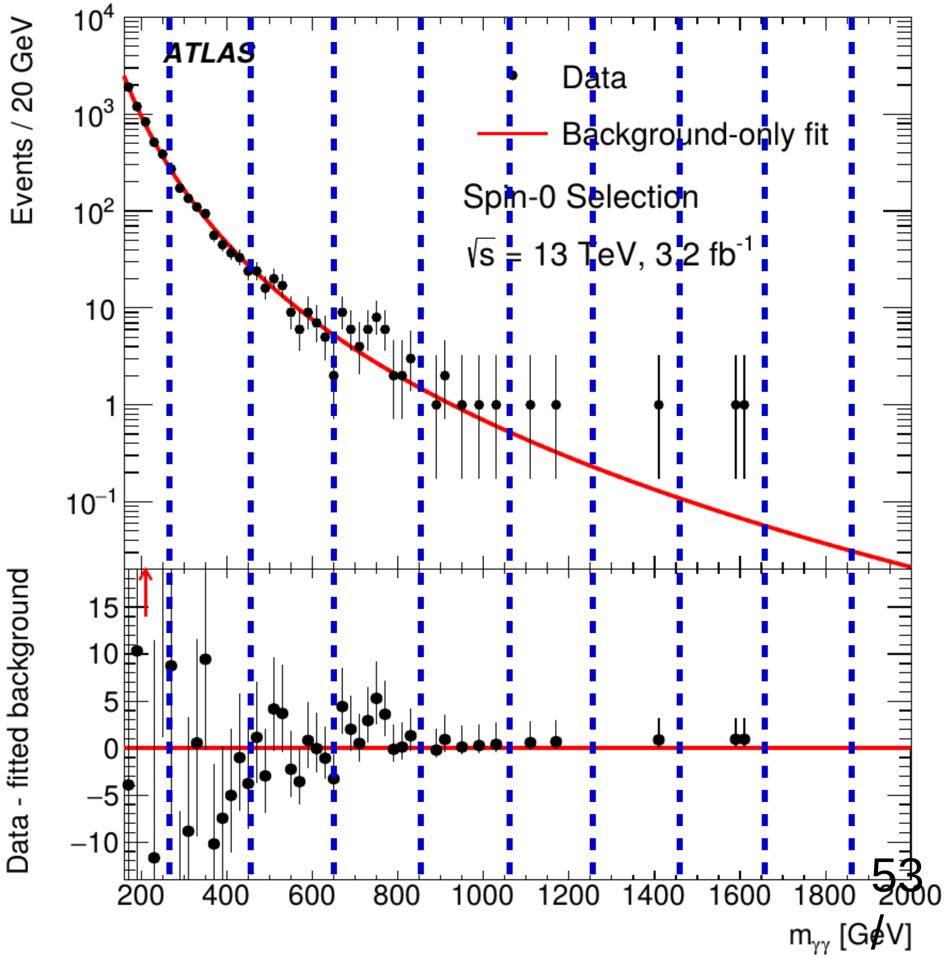
$$N_{\text{indep}} = \frac{\text{scan range}}{\text{peak width}}$$

→ 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 for continuous variables

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

Asymptotic limit : trials factor (1 POI) is

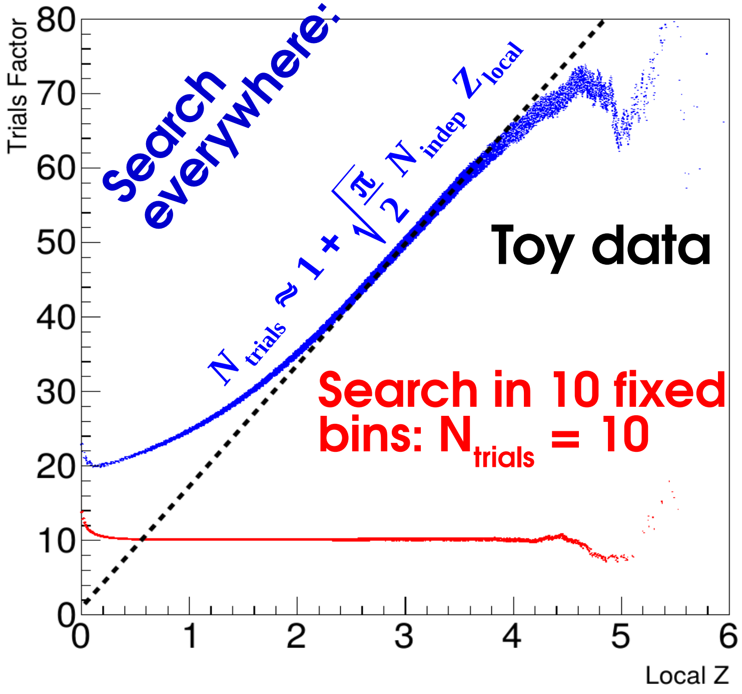
$$N_{\text{trials}} = 1 + \sqrt{\frac{\pi}{2}}\, \underset{\uparrow}{N_{\text{indep}}}\, Z_{\text{local}}$$

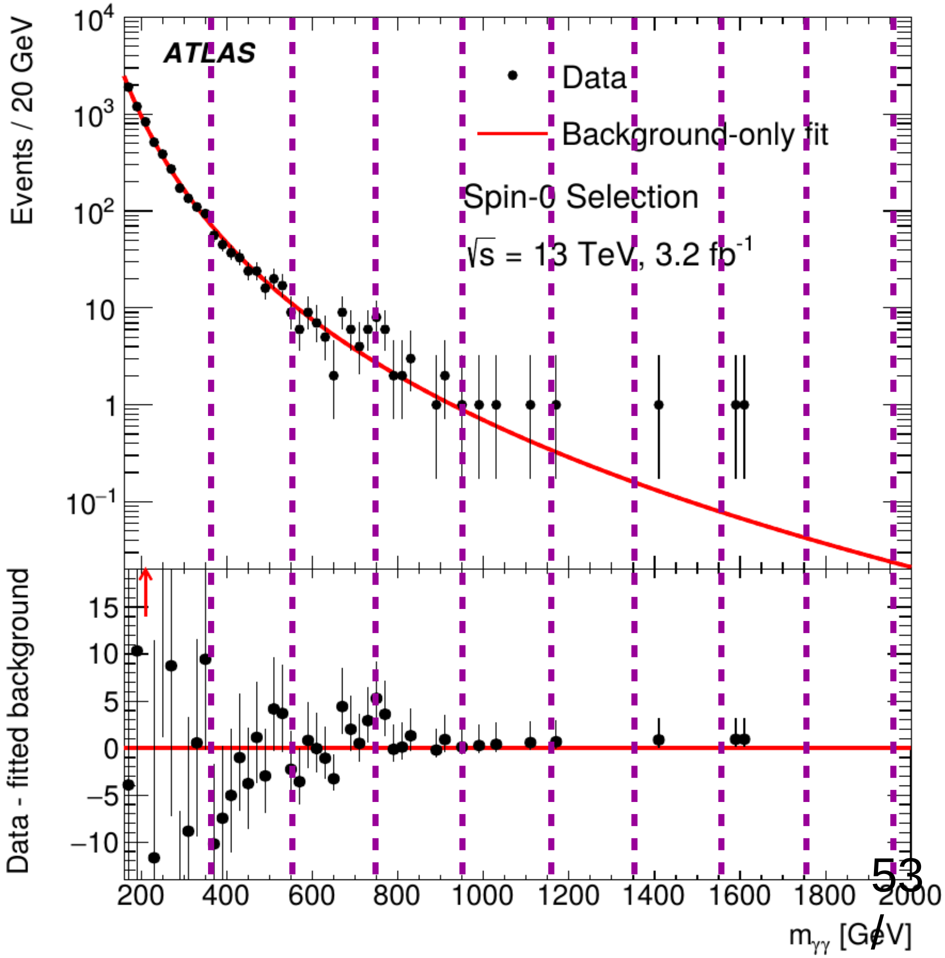
$$N_{\text{indep}} = \frac{\text{scan range}}{\text{peak width}}$$

→ 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**

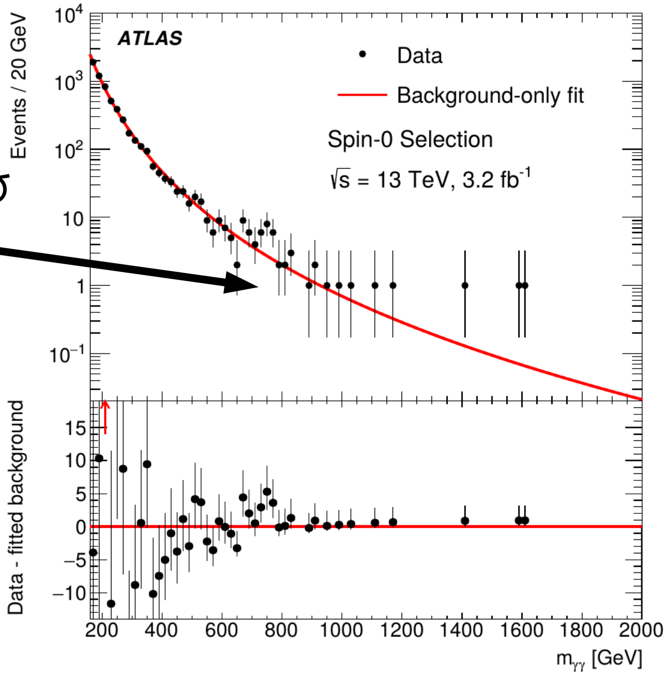
⇒ true N_{trials} is **> N_{indep}**!





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)

Conclusion

- Significant evolution in the statistical methods used in HEP

- Variety of methods, adapted to various situations and target results

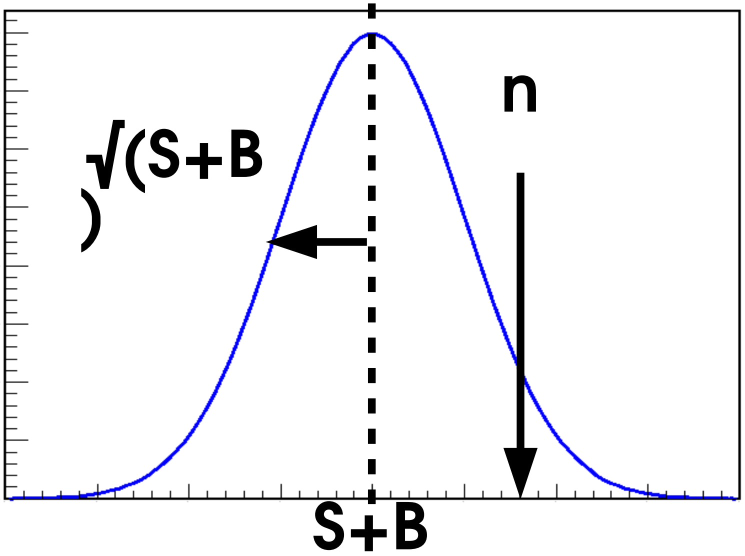
- Allow to
  - model the statistical process with high precision in difficult situations (large systematics, small signals)
  - make optimal use of available information

- 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

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 : **Ŝ = n – B**

Test statistic: assume Ŝ > 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: **Ŝ = n – B**, same as Gaussian

Test statistic (for Ŝ > 0):

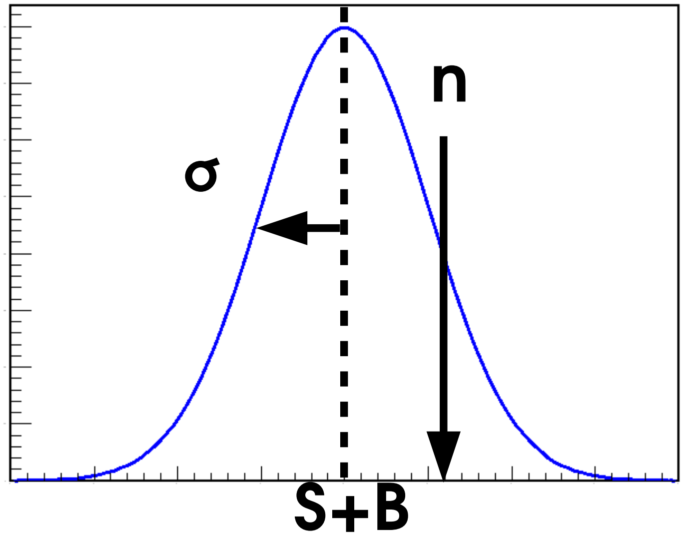
$$q_0 = \lambda(S=0) - \lambda(\hat{S}) = -2\hat{S} - 2(\hat{S}+B)\,\log\frac{B}{\hat{S}+B}$$

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:

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

for $S_0 > \hat{S}$

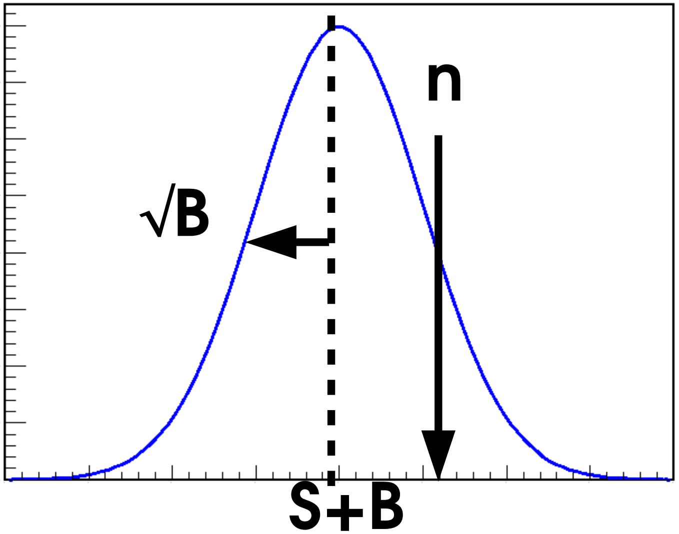
so $q_{S_0} = 2.70$ for $S_0 = \hat{S} + \sqrt{2.70}\,\sigma_S$

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 : **Ŝ = n - B**

CL_{s+b} limit: $S_{\mathrm{up}} = \hat{S} + 1.64\,\sigma_S$ 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

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

for Ŝ = 0,

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

Ŝ ~ G(S, σ_S) so

Under $H_0(S = S_0)$:

$$\sqrt{q_{S_0}} \sim G(0,1)$$
$$p_{S_0} = 1-\Phi\left(\sqrt{q_{S_0}}\right)$$

Under $H_0(S = 0)$:

$$\sqrt{q_{S_0}} \sim G(S_0/\sigma_S, 1)$$
$$p_B = \Phi\left(\sqrt{q_{S_0}} - S_0/\sigma_S\right)$$

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

⇒ 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, $$q_{S_0}(n=0) = 2(S_0+B)$$

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

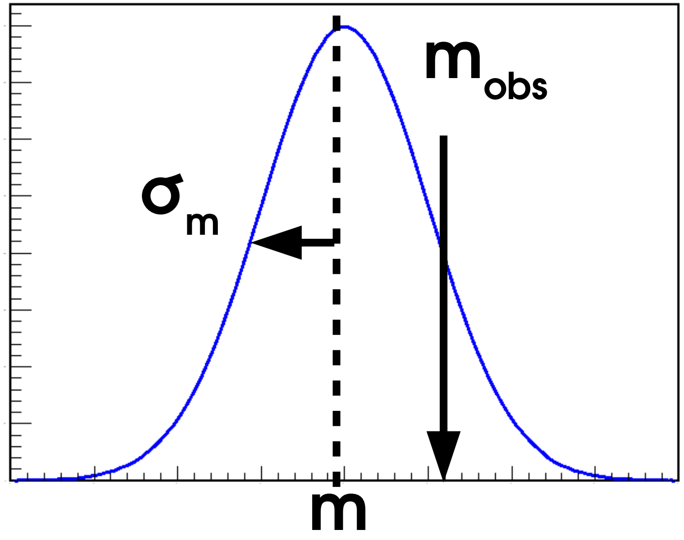
⇒ S$_{up}$ ~ 2, exact value depends on B

⇒ 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}:

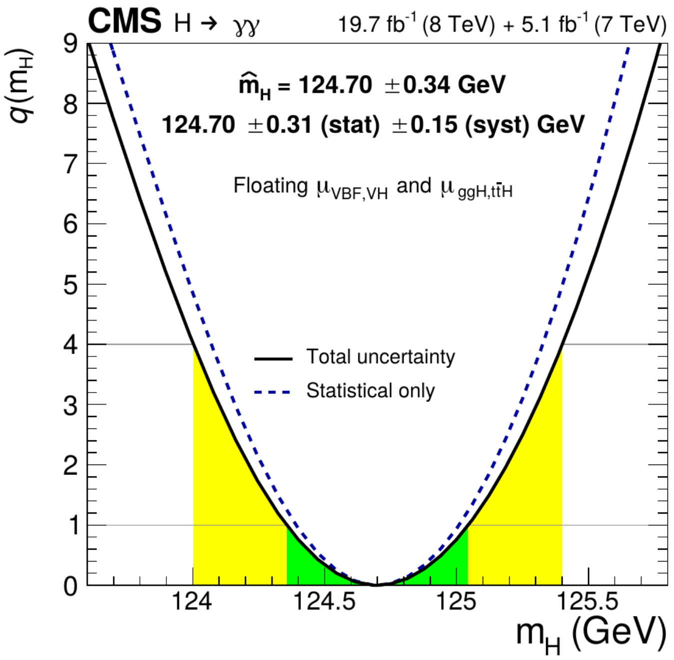
$$\lambda(m;m_{\mathbf{obs}}) = \left(\frac{m-m_{\mathbf{obs}}}{\sigma_m}\right)^2$$



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

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

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



JHEP 11 (2017) 047

Homework 7: Gaussian Profiling

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

→ 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 = Ŝ, matches MLE as it should

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

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

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

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

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

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

This boils down to the Gaussian case of HW 6, so the CL$_s$ limit is

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

Homework 8: Bayesian computation

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

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

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

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

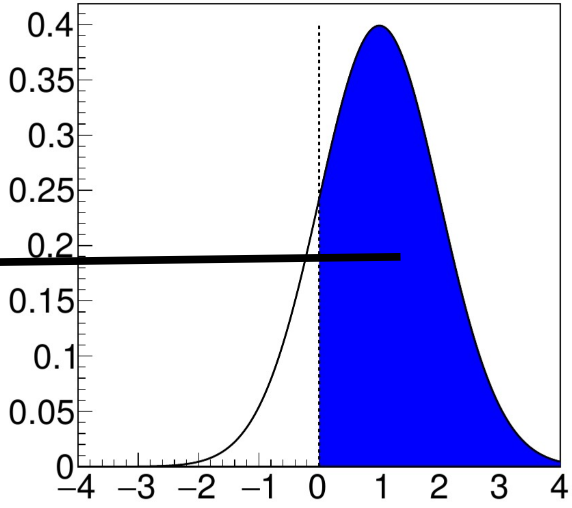
same effect as profiling!

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

⇒ Truncate Gaussian at S=0:

$$P(S\,|\,n) = P(n\,|\,S)\; P(S)$$

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



Bayesian Limit:

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

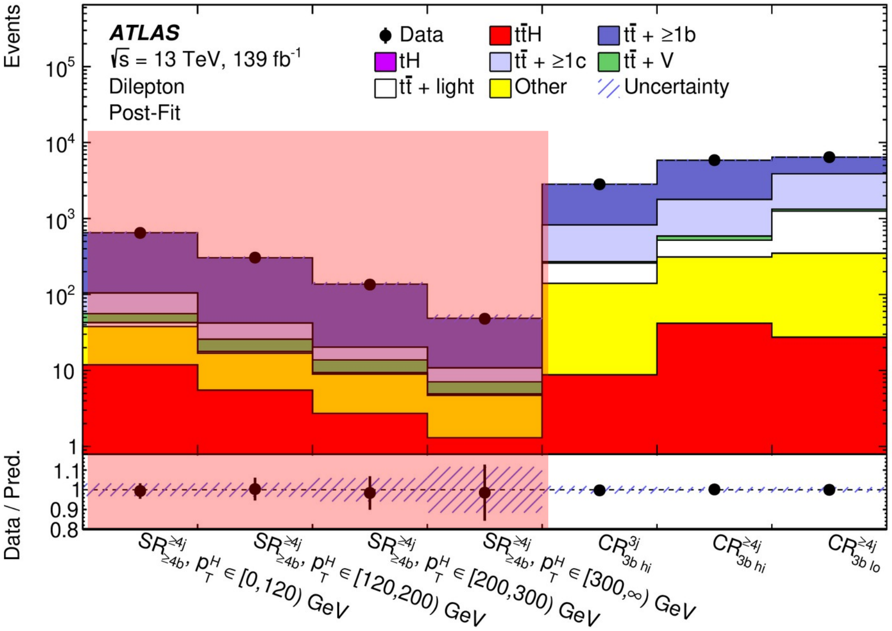
$$S_{up}^{Bayes} = n-B + \left[\Phi^{-1}\left(1 - 0.05\,\Phi\left(\frac{n-B}{\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}}\right)\right)\right]\sqrt{\sigma_{stat}^2+\sigma_{syst}^2}$$

same result as CL_s!

Extra Slides

Categories

arXiv:2111.06712

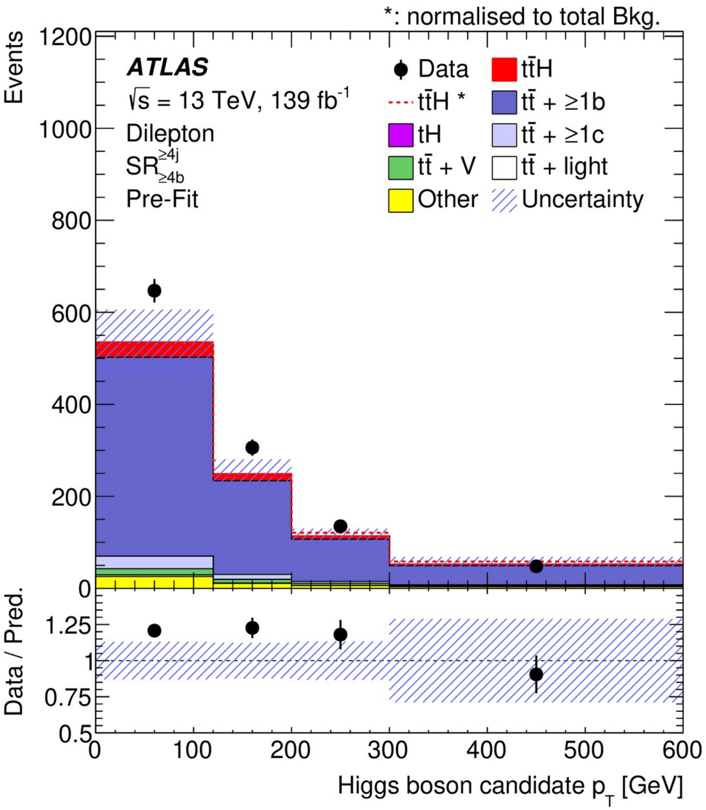


Multiple analysis regions often used.

→ Exploit better sensitivity in some regions

Here (ttH, H→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!

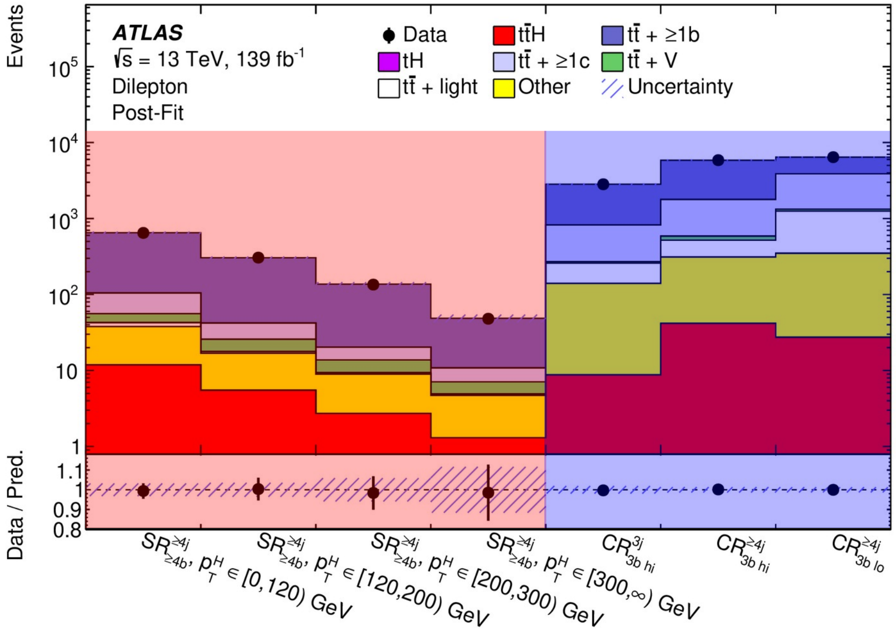
Categories

arXiv:2111.06712

Multiple analysis regions often used.

→ Exploit better sensitivity in some regions

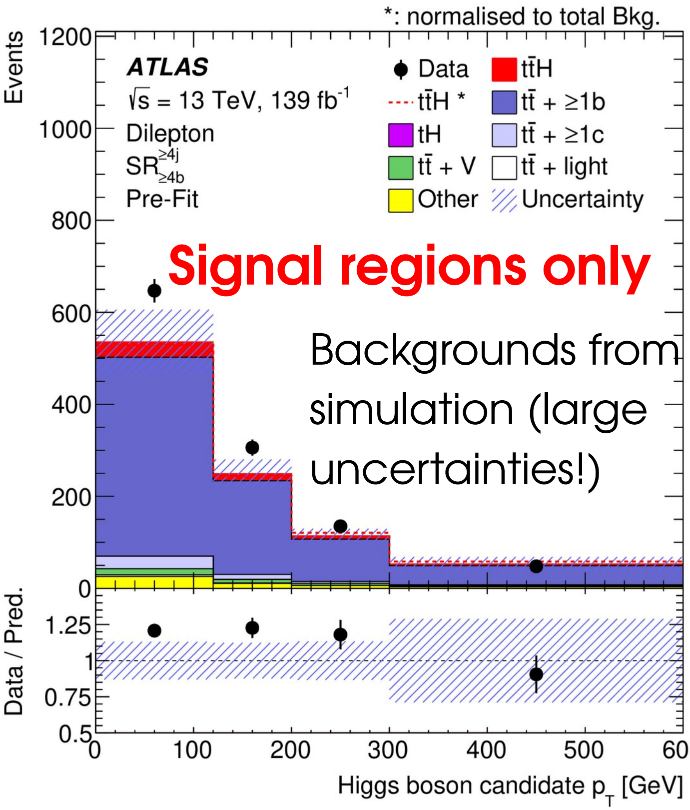
→ Constrain NPs: ***Control regions*** for bkgs



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

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

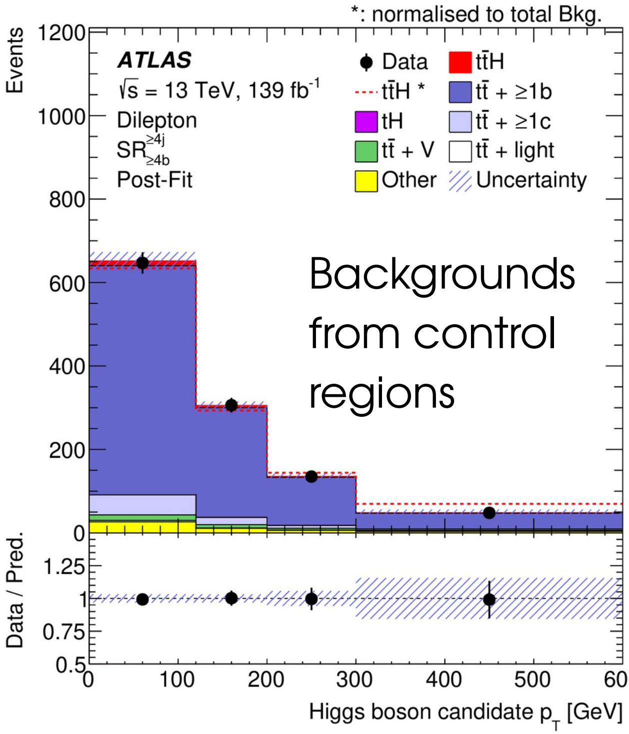
→ **3** Background *Control Regions* (**CR**)



Signal regions only

Backgrounds from simulation (large uncertainties!)

Include Background CRs

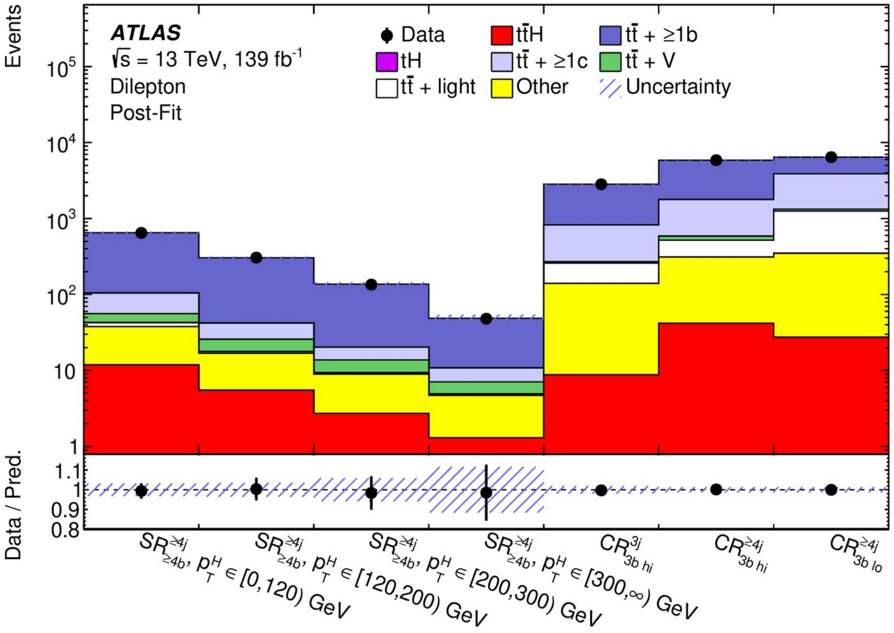


Backgrounds from control regions

Signal + Bkg regions

Categories

arXiv:2111.06712



Multiple analysis regions often used.

→ Exploit better sensitivity in some regions

→ Constrain NPs: ***Control regions*** for bkgs

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

↓

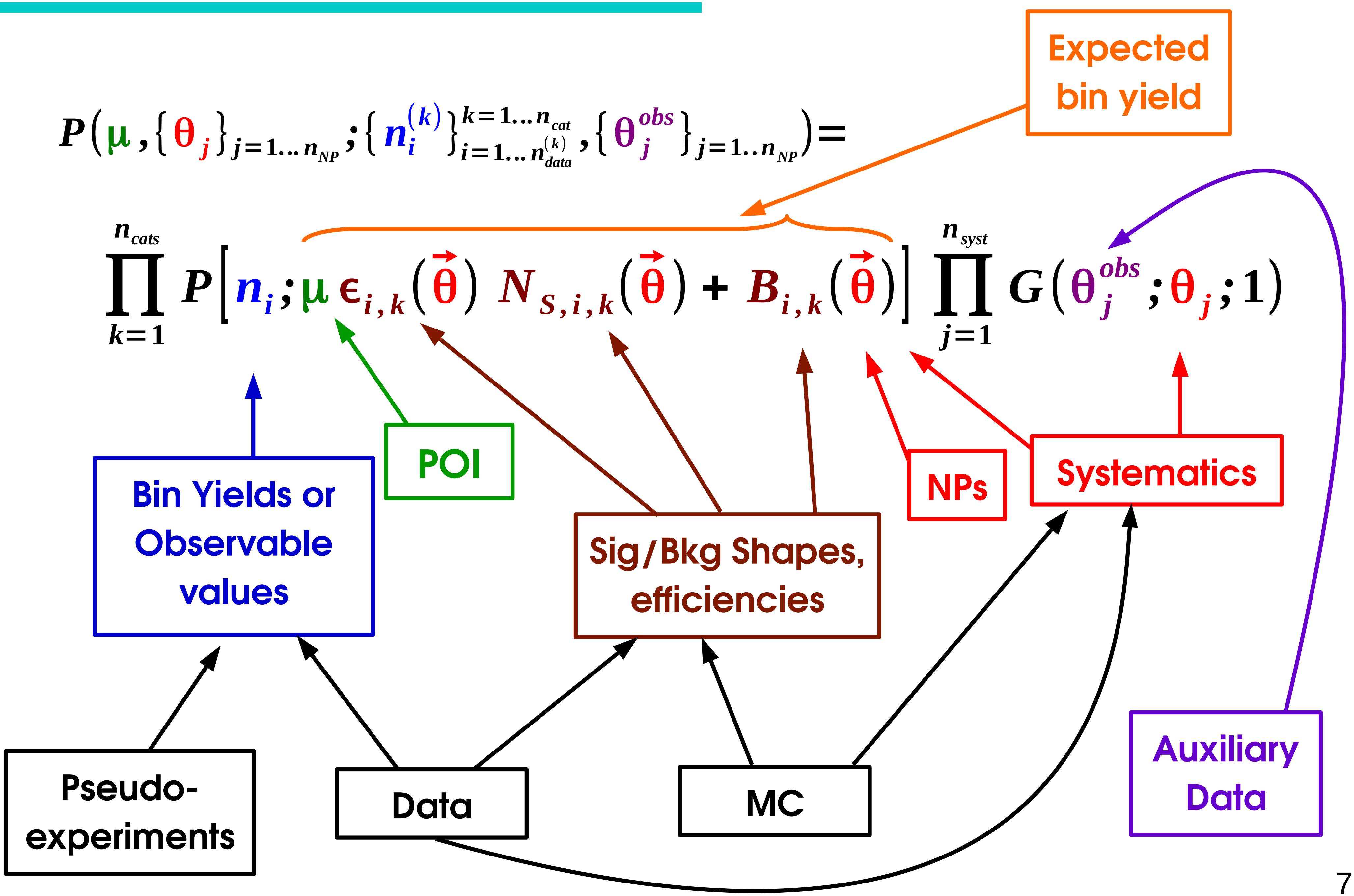
$$P\left(S, B; \{n_i^{(k)}\}^{k=1\ldots n_{\text{cats}}}_{i=1\ldots n_{\text{evts}}^{(k)}}\right) = \prod_{k=1}^{n_{\text{cats}}} P_k\left(S, B; \{n_i^{(k)}\}_{i=1\ldots n_{\text{evts}}^{(k)}}\right)$$

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



$$P\left(\mu, \{\theta_j\}_{j=1\ldots n_{NP}}; \{n_i^{(k)}\}_{i=1\ldots n_{data}^{(k)}}^{k=1\ldots n_{cat}}, \{\theta_j^{obs}\}_{j=1..n_{NP}}\right) =$$

$$\prod_{k=1}^{n_{cats}} P\left[n_i; \overbrace{\mu\, \epsilon_{i,k}(\vec{\theta})\; N_{s,i,k}(\vec{\theta}) + B_{i,k}(\vec{\theta})}^{\text{Expected bin yield}}\right] \prod_{j=1}^{n_{syst}} G(\theta_j^{obs}; \theta_j; 1)$$

Expected bin yield

POI

NPs

Systematics

Bin Yields or Observable values

Sig/Bkg Shapes, efficiencies

Pseudo-experiments

Data

MC

Auxiliary Data

× number of categories!

CL_s : Gaussian Bands

Usual Gaussian counting example with known B:

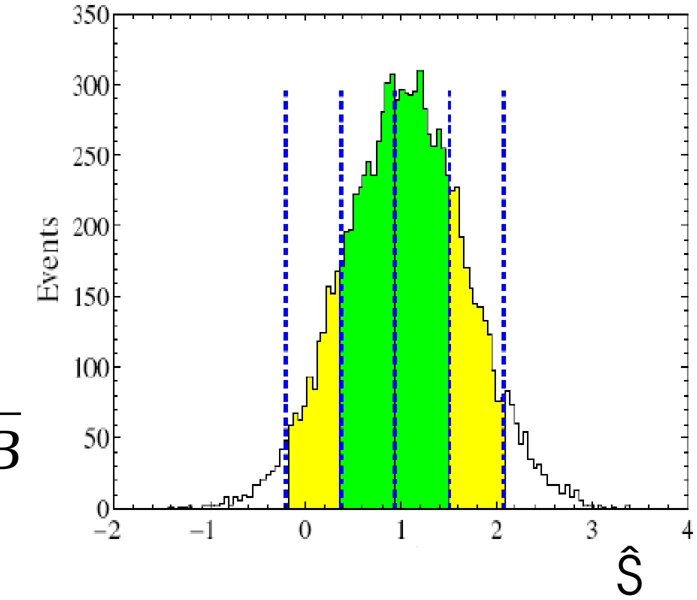
95% CL_s upper limit on S:

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

with

$$\sigma_S = \sqrt{B}$$

Eur.Phys.J.C71:1554,2011



Compute expected bands for S=0:

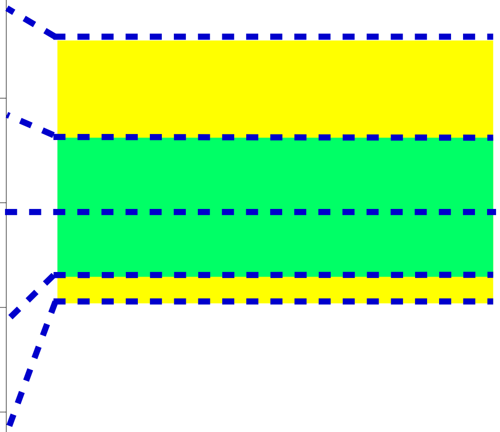
→ **Asimov dataset ⇔ $\hat{S} = 0$** :

$$S^{0}_{up,exp} = 1.96\,\sigma_S$$

→ **± nσ bands**:

$$S^{\pm n}_{up,exp} = \left(\pm n + \left[1 - \Phi^{-1}\left(0.05\,\Phi\left(\mp n\right)\right)\right]\right)\sigma_S$$

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



CLs :

- Positive bands somewhat reduced,
- Negative ones more so

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

Comparison with LEP/TeVatron definitions

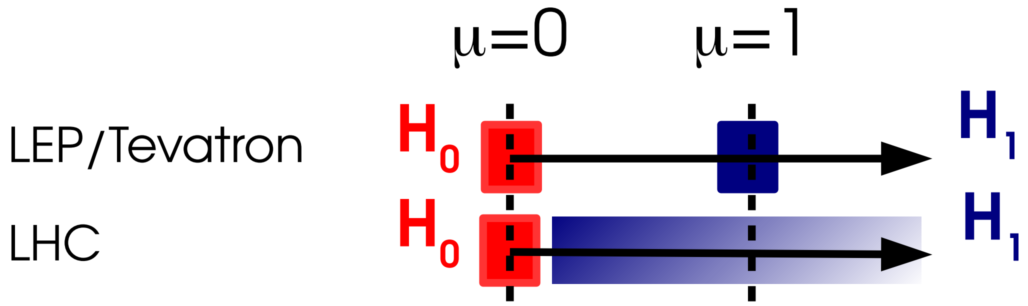
Likelihood ratios are not a new idea:

- **LEP**: Simple LR with NPs from MC
  - Compare μ=0 and μ=1
- **Tevatron**: PLR with profiled NPs

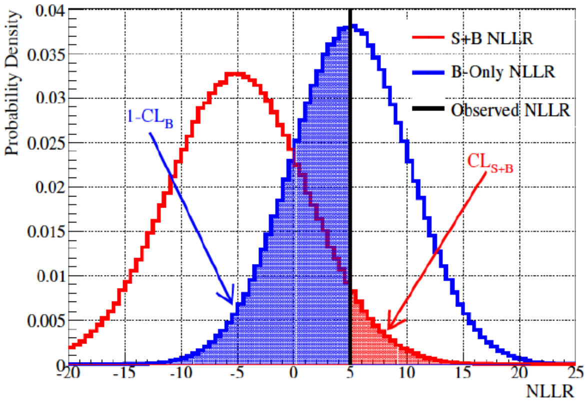
$$q_{LEP} = -2\log\frac{L(\mu=0,\tilde{\theta})}{L(\mu=1,\tilde{\theta})}$$

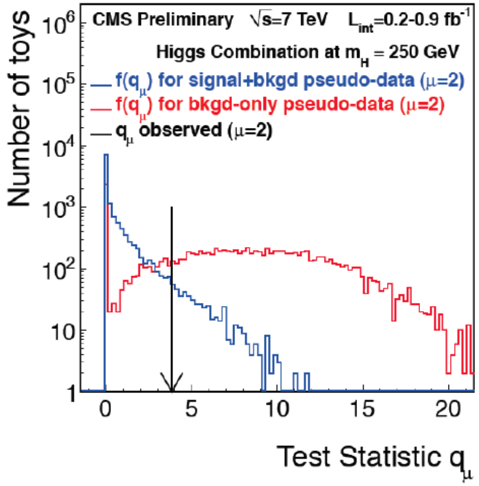
$$q_{Tevatron} = -2\log\frac{L(\mu=0,\hat{\hat{\theta}}_0)}{L(\mu=1,\hat{\hat{\theta}}_1)}$$

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



Example from Collie documentation (D0 note 5595)





→ Asymptotically:

- **LEP/Tevaton**: q linear in μ ⇒ **~Gaussian**
- **LHC**: q quadratic in μ ⇒ **~χ2**

→ Still use TeVatron-style for discrete cases

Andrey Korytov, EPS 2011

Wilks’ Theorem

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

To test the $S=S_0$ hypothesis, consider

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

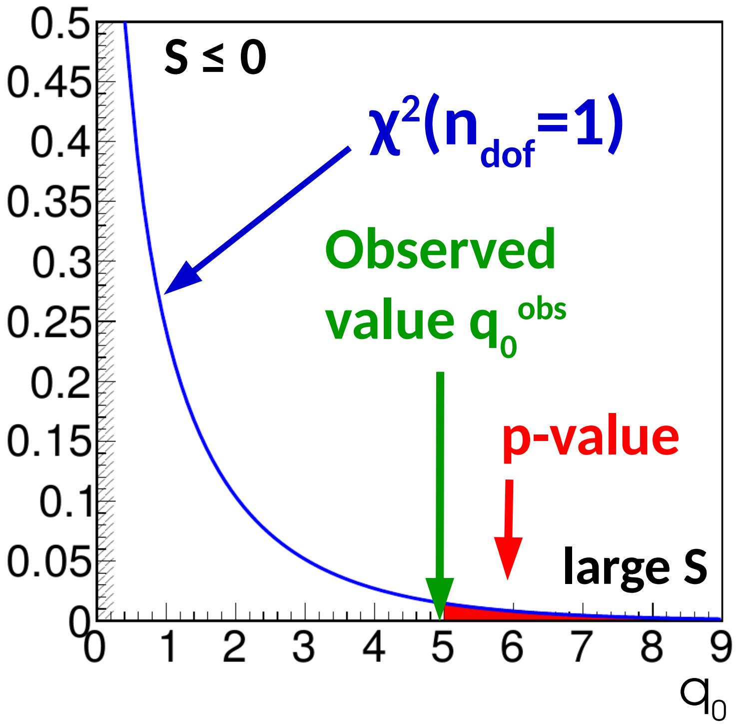
→ Assume **Gaussian regime** (e.g. large n_{evts}, Central-limit theorem) : then:

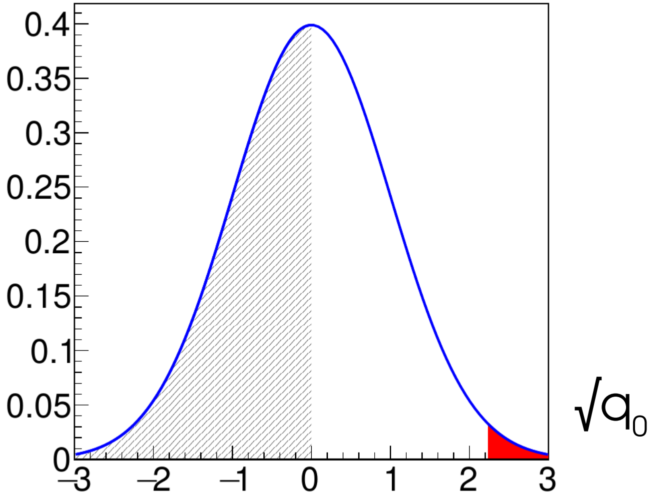
Wilk’s Theorem: **$t(S_0)$ is distributed as a χ^2**

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

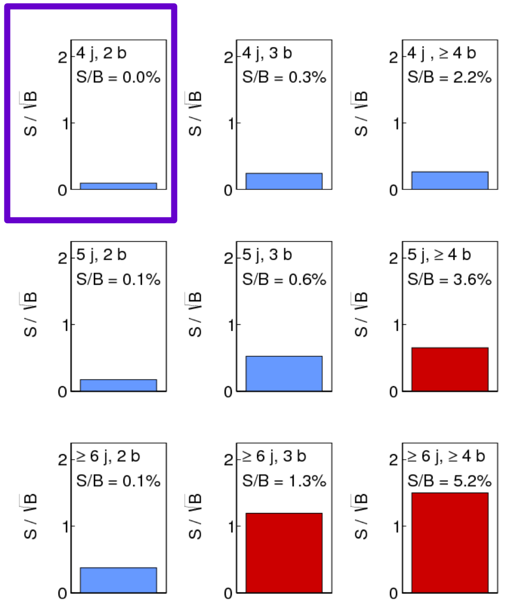
⇒ In particular, the significance is:

$$Z = \sqrt{q_0}$$





Profiling Example: ttH→bb

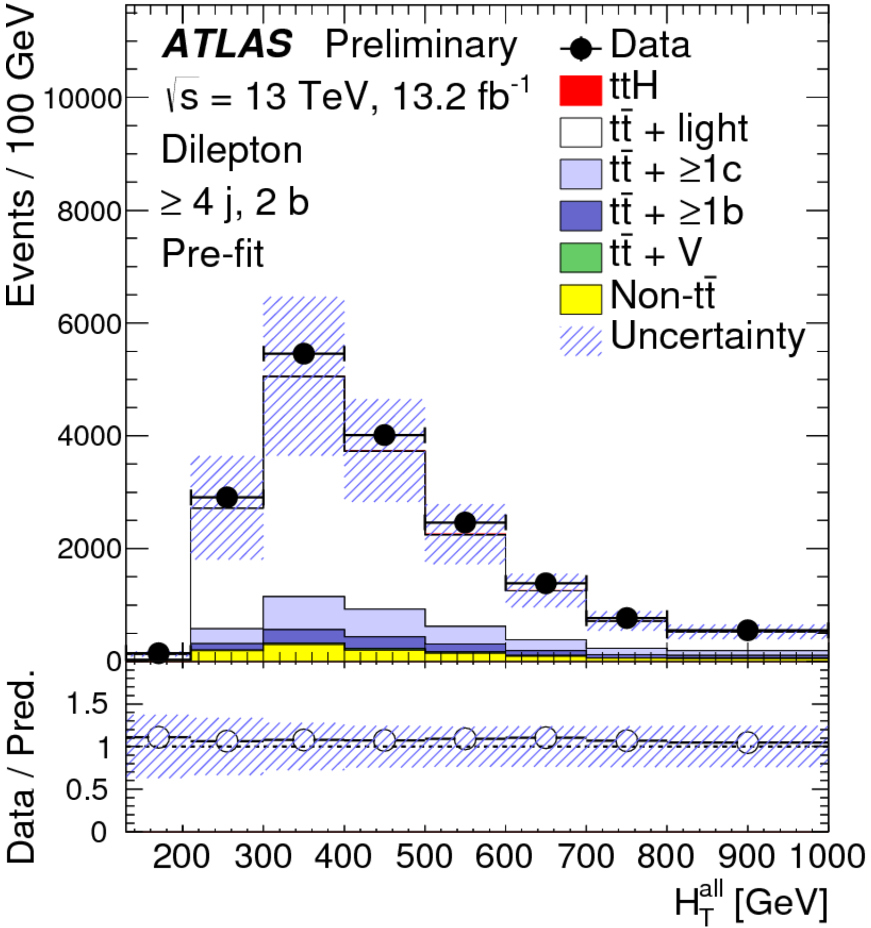


Analysis uses low-S/B categories to constrain backgrounds.

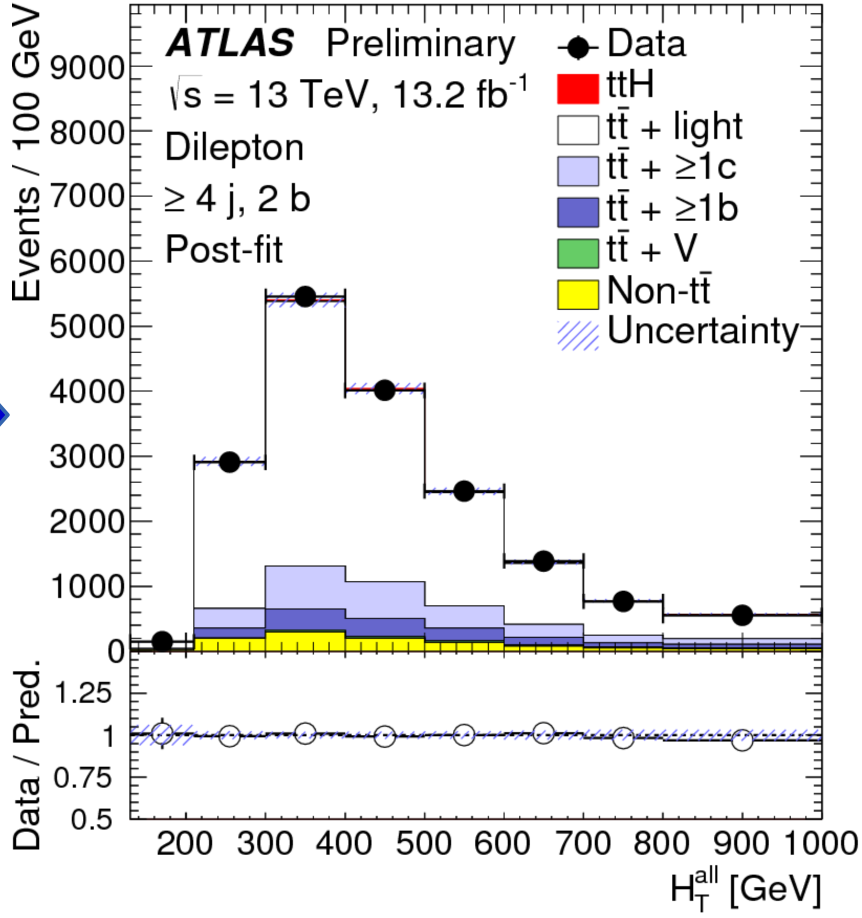
→ **Reduction in large uncertainties on tt bkg**

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

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



Fit



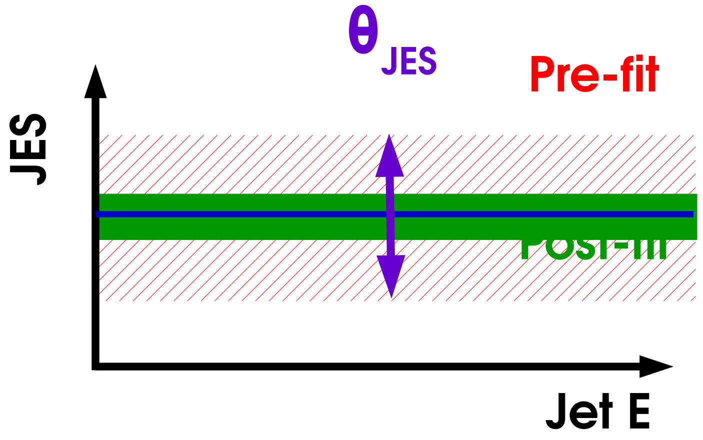
ATLAS-CONF-2016-08

Profiling Issues

Too simple modeling can have unintended effects

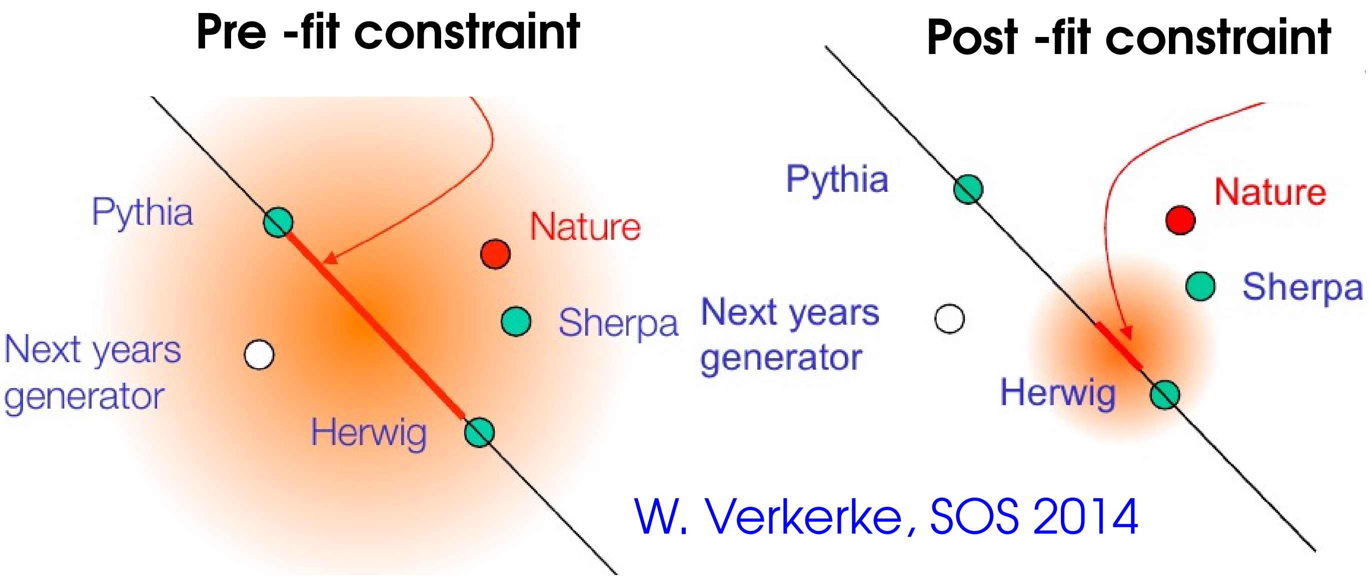
→ e.g. single Jet E scale parameter:

⇒ Low-E jets calibrate high-E jets – intended ?



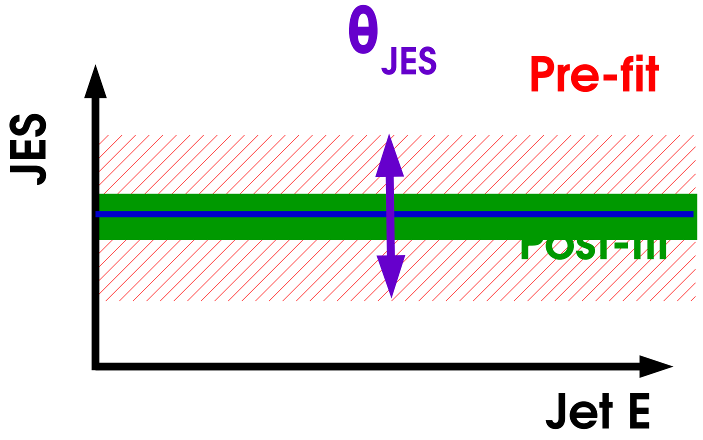
***Two-point* uncertainties**:

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



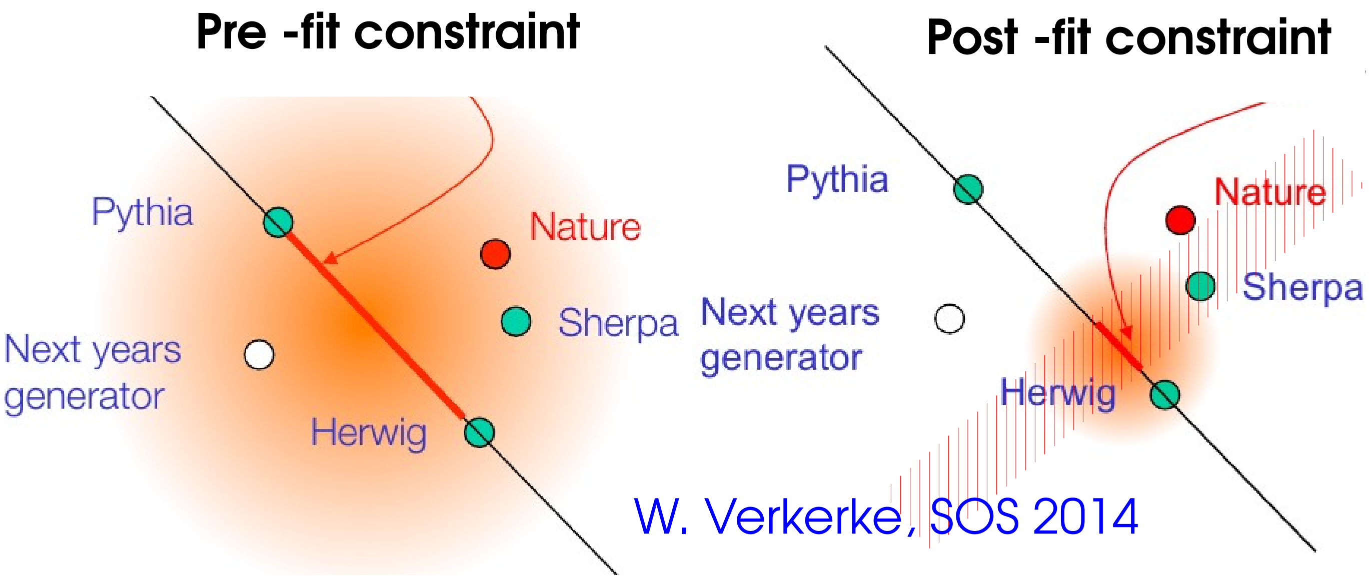
W. Verkerke, SOS 2014

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



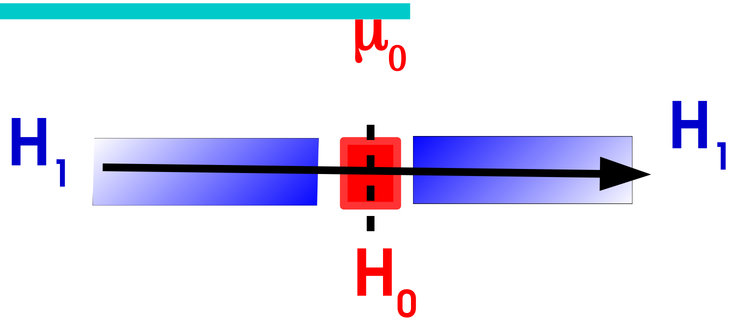
W. Verkerke, SOS 2014

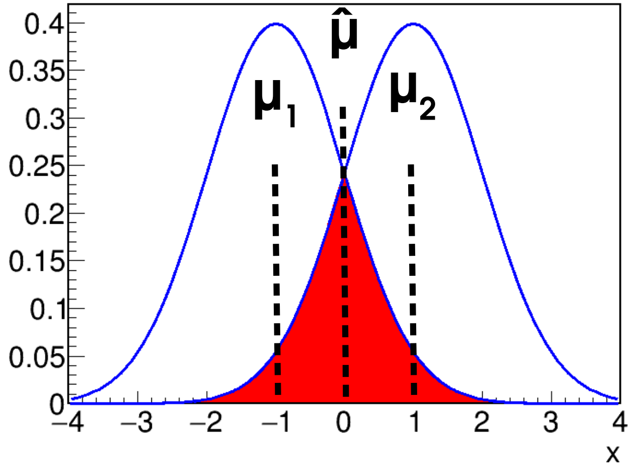
Test Statistics for Limit-Setting

Interval :

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$





Try to exclude μ values away from μ̂.

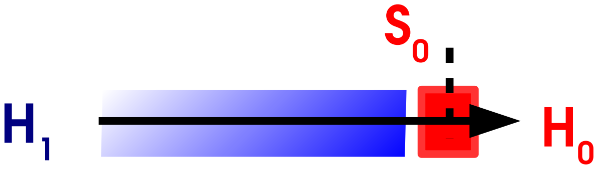
$$t(\mu_0) = -2\log\frac{L(\mu=\mu_0)}{L(\hat{\mu})}$$

“Two-sided” test

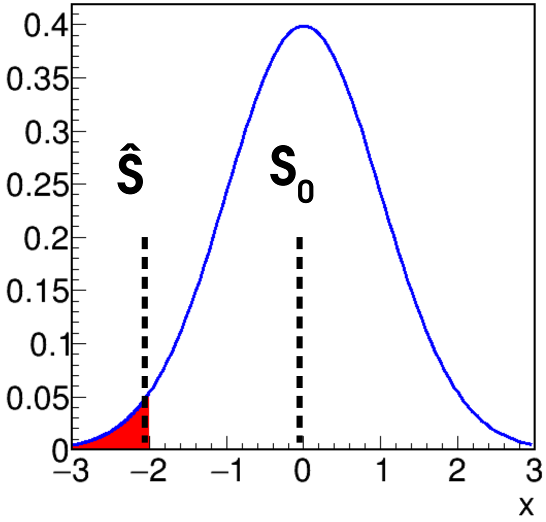
Limit-setting

$H_0 : S = S_0$

$H_1 : S < S_0$



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



Try to exclude values of S that are above Ŝ.

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

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