

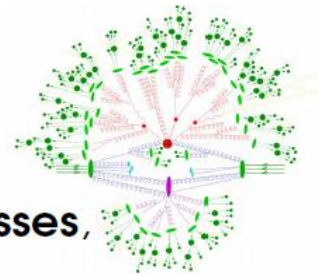
Statistics and Data Analysis (HEP at LHC)

□ Computing statistical results

- Estimating the value of a parameter
- Testing hypotheses
- Discovery
- Limits
- Confidence intervals

Slides extracted from N. Berger lectures at CERN Summer School 2019

How to represent the data



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

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

Model can include multiple **categories**, each with a separate description

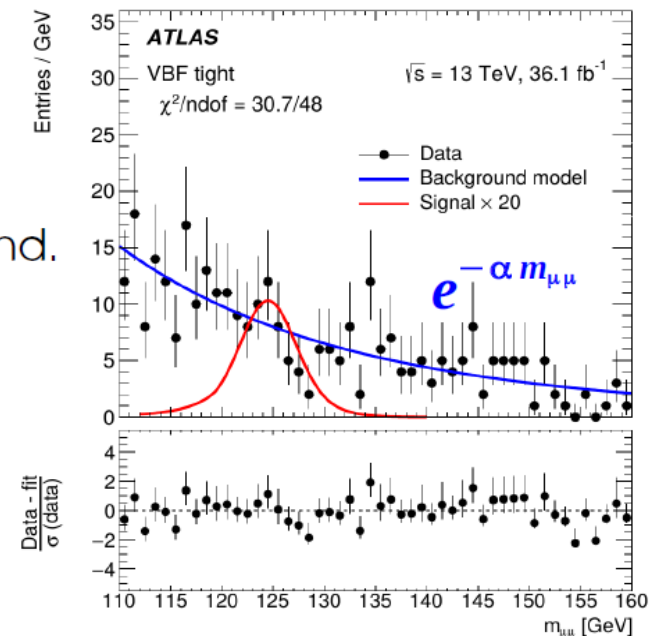
Model parameters

Model typically includes:

- **Parameters of interest** (POIs) : what we want to measure
→ $S, \sigma, m_W, \dots$
- **Nuisance parameters** (NPs) : other parameters needed to define the model
→ **B**
→ For binned data, $f_{\text{sig}_i}, f_{\text{bkg}_i}$
→ For unbinned data, parameters needed to define P_{bkg}
e.g. exponential slope α of $H \rightarrow \mu\mu$ background.

NPs must be either

- **given a value “by hand”** (possibly within systematics) or
- **constrained by the data** (e.g. in sidebands)



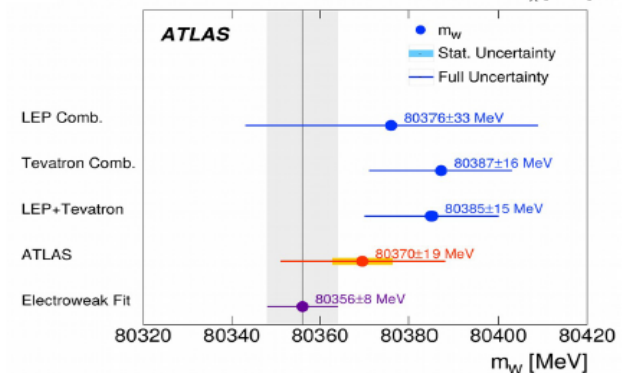
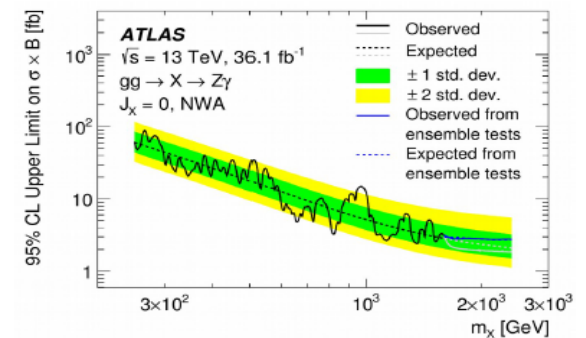
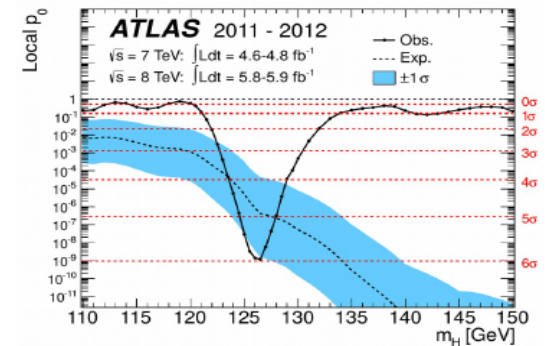
Phys. Rev. Lett. 119 (2017) 051802 4

Statistical computations

Now that we have a model, can use it to compute analysis results:

- **Discovery significance:** we see an excess – is it a (new) signal, or a background fluctuation ?
- **Upper limit on signal yield:** we don't see an excess – if there is a signal present, how small must it be ?
- **Parameter measurement:** what is the allowed range for a model parameter ? (“confidence interval”)

→ The Statistical Model already contains all the needed information – how to use it ?



Using the PDF

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

⇒ Possible outcomes of the experiment, for given parameter values

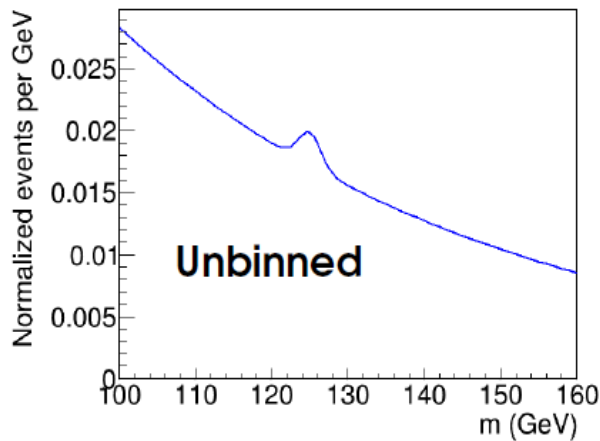
Can draw random events according to PDF : **generate** *pseudo-data*

$$P(\lambda=5)$$

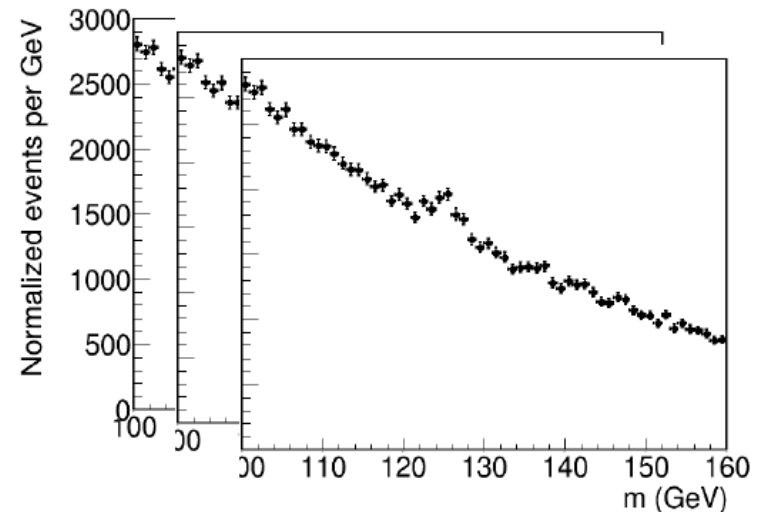


2, 5, 3, 7, 4, 9,

Each entry = separate “experiment”



Generate



Likelihood

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

⇒ Possible outcomes of the experiment, for given parameter values

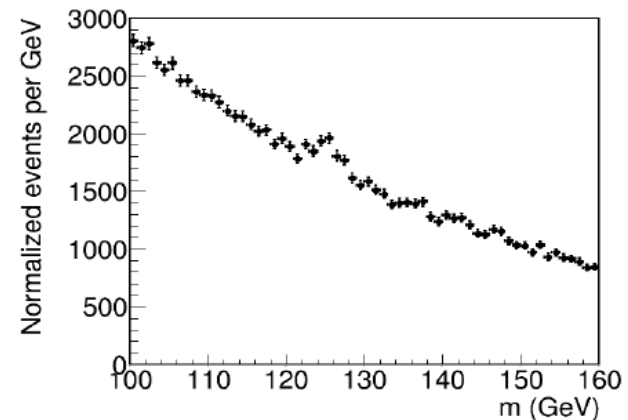
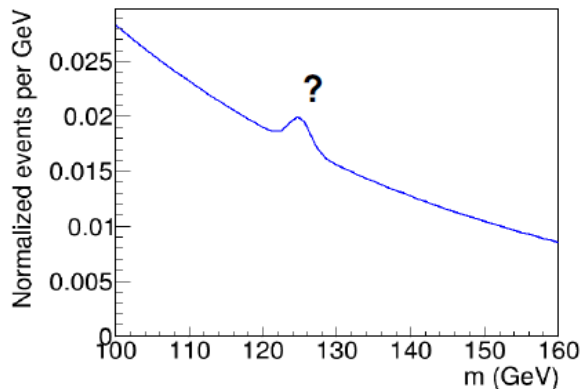
We want the **other** direction: **use data to get information on parameters**

$$P(\lambda = ?)$$

2



Estimate



Likelihood: $L(\text{parameters}) = P(\text{data}; \text{parameters})$

→ same as the PDF, but seen as function of the parameters

Poisson example

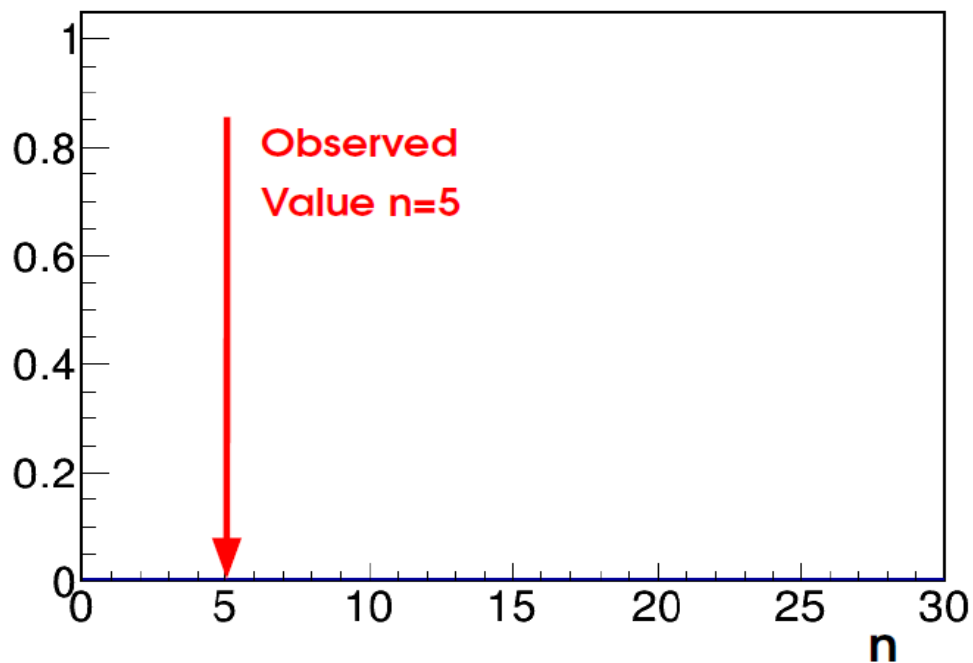
Assume **Poisson distribution** with $\mu = 0$: $P(\mathbf{n}; \mathbf{S}) = e^{-S} \frac{S^n}{n!}$

Say we **observe $n=5$** , want to infer information on the parameter **S**

→ Try different values of S for a fixed data value $n=5$

→ Varying parameter, fixed data: **likelihood**

$$L(\mathbf{S}; \mathbf{n=5}) = e^{-S} \frac{S^5}{5!}$$



Poisson example

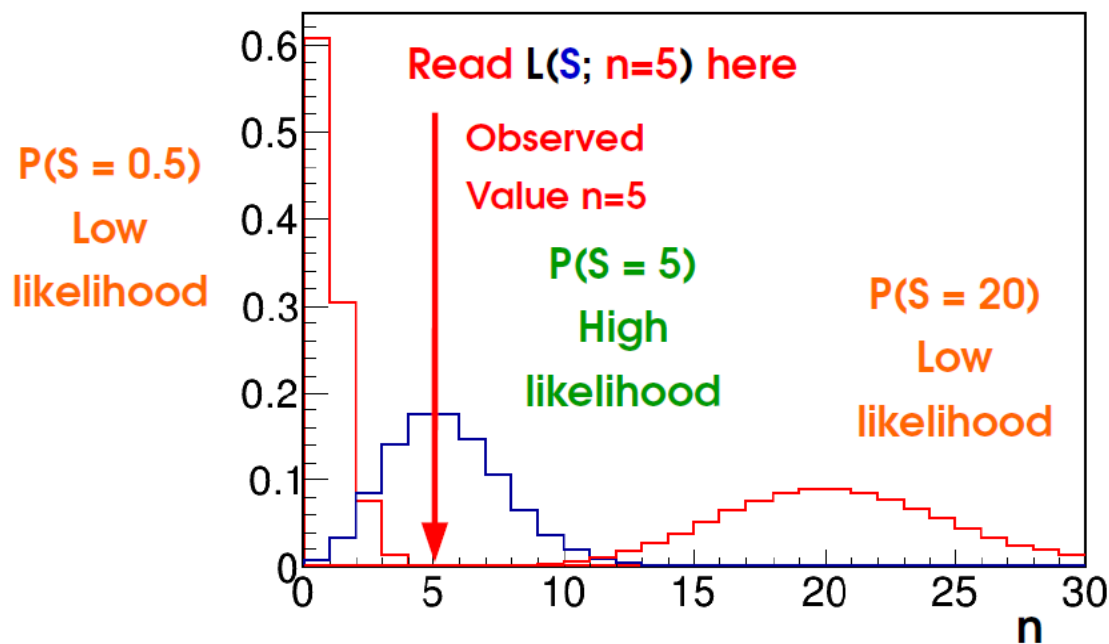
Assume **Poisson distribution** with $B = 0$: $P(\mathbf{n}; \mathbf{S}) = e^{-S} \frac{S^n}{n!}$

Say we **observe** $\mathbf{n=5}$, want to infer information on the parameter $\mathbf{S}$

→ Try different values of S for a fixed data value $n=5$

→ Varying parameter, fixed data: **likelihood**

$$L(\mathbf{S}; \mathbf{n=5}) = e^{-S} \frac{S^5}{5!}$$



Poisson example

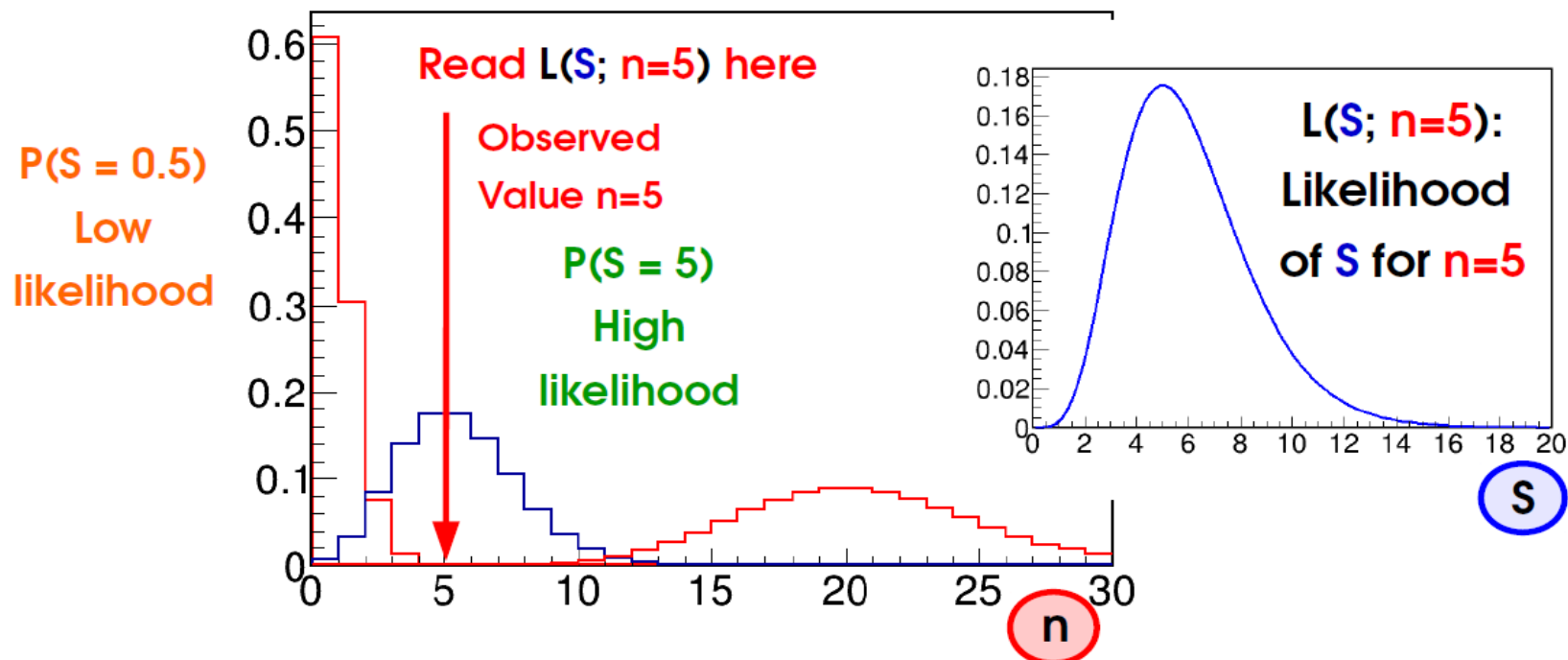
Assume **Poisson distribution** with $B = 0$: $P(\mathbf{n}; \mathbf{S}) = e^{-S} \frac{S^n}{n!}$

Say we **observe** $\mathbf{n=5}$, want to infer information on the parameter $\mathbf{S}$

→ Try different values of S for a fixed data value $n=5$

→ Varying parameter, fixed data: **likelihood**

$$L(\mathbf{S}; \mathbf{n=5}) = e^{-S} \frac{S^5}{5!}$$

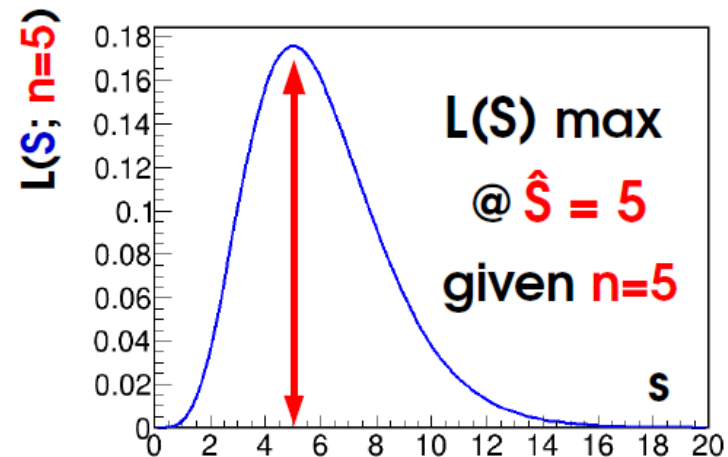
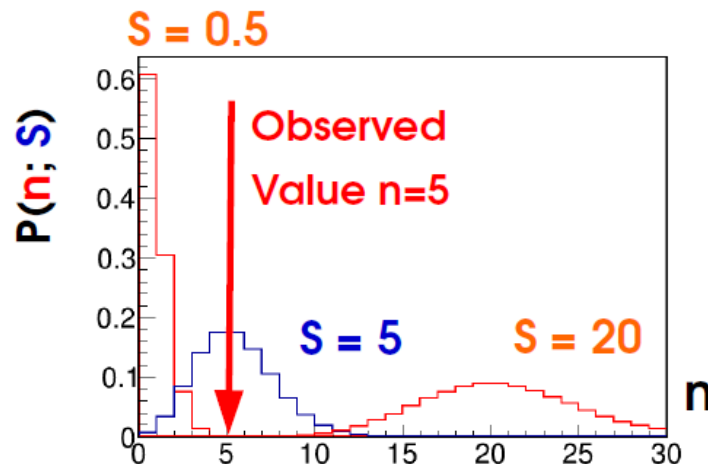


Maximum Likelihood Estimation

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

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

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



MLE: the value of μ for which **this data** was *most likely to occur*

The MLE is a function of the data – itself an **observable**

No guarantee it is the true value (data may be “unlikely”) but sensible estimate

MLEs in shape analyses

Binned shape analysis:

$$L(\mathbf{S}; \mathbf{n}_i) = P(\mathbf{n}_i; \mathbf{S}) = \prod_{i=1}^N \text{Pois}(\mathbf{n}_i; \mathbf{S} f_i + B_i)$$

Maximize global $L(\mathbf{S})$ (each bin may prefer a different $\mathbf{S}$)

In practice easier to minimize

$$\lambda_{\text{Pois}}(\mathbf{S}) = -2 \log L(\mathbf{S}) = -2 \sum_{i=1}^N \log \text{Pois}(\mathbf{n}_i; \mathbf{S} f_i + B_i) \quad \text{Needs a computer...}$$

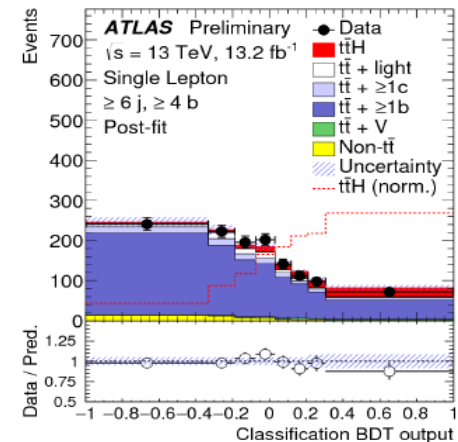
In the Gaussian limit

$$\lambda_{\text{Gaus}}(\mathbf{S}) = \sum_{i=1}^N -2 \log G(\mathbf{n}_i; \mathbf{S} f_i + B_i, \sigma_i) = \sum_{i=1}^N \left(\frac{\mathbf{n}_i - (\mathbf{S} f_i + B_i)}{\sigma_i} \right)^2 \quad \chi^2 \text{ formula!}$$

→ **Gaussian MLE** ($\min \chi^2$ or $\min \lambda_{\text{Gaus}}$) : **Best fit value** in a χ^2 (Least-squares) fit

→ **Poisson MLE** ($\min \lambda_{\text{Pois}}$) : **Best fit value** in a *likelihood* fit (in ROOT, fit option "L")

In RooFit, $\lambda_{\text{Pois}} \Rightarrow \text{RooAbsPdf::fitTo}()$, $\lambda_{\text{Gaus}} \Rightarrow \text{RooAbsPdf::chi2FitTo}()$.

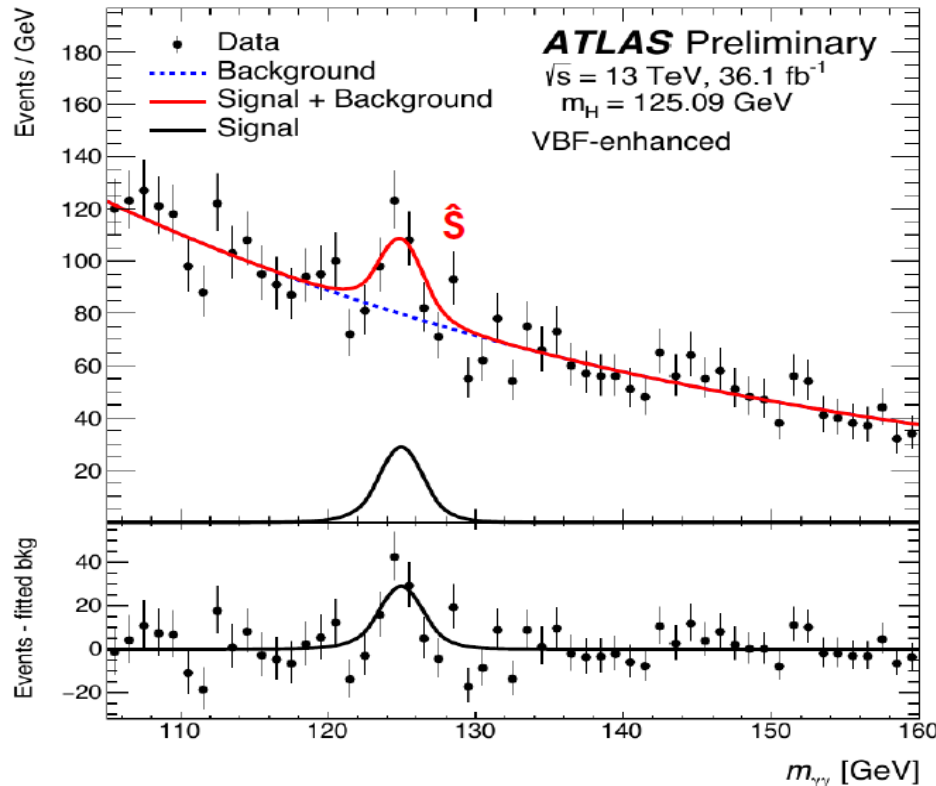


In both cases, MLE $\Leftrightarrow$ Best Fit

MLEs in shape analyses

$H \rightarrow \gamma\gamma$

$$L(\mathbf{S}, \mathbf{B}; \mathbf{m}_i) = e^{-(S+B)} \prod_{i=1}^{n_{\text{evts}}} S P_{\text{sig}}(\mathbf{m}_i) + B P_{\text{bkg}}(\mathbf{m}_i)$$



Estimate the MLE $\hat{S}$ of S ?

→ Perform (likelihood) best-fit of model to data

⇒ fit result for S is the desired $\hat{S}$.

In particle physics, often use the *MINUIT* minimizer within ROOT.

ATLAS-CONF-2017-045

MLE Properties

Asymptotically Gaussian
and unbiased :

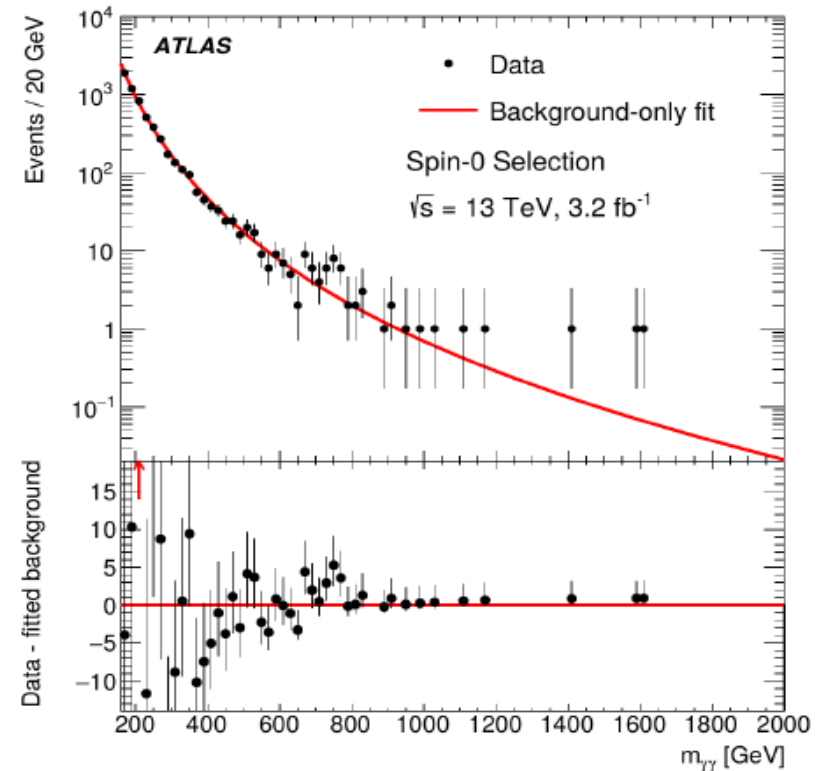
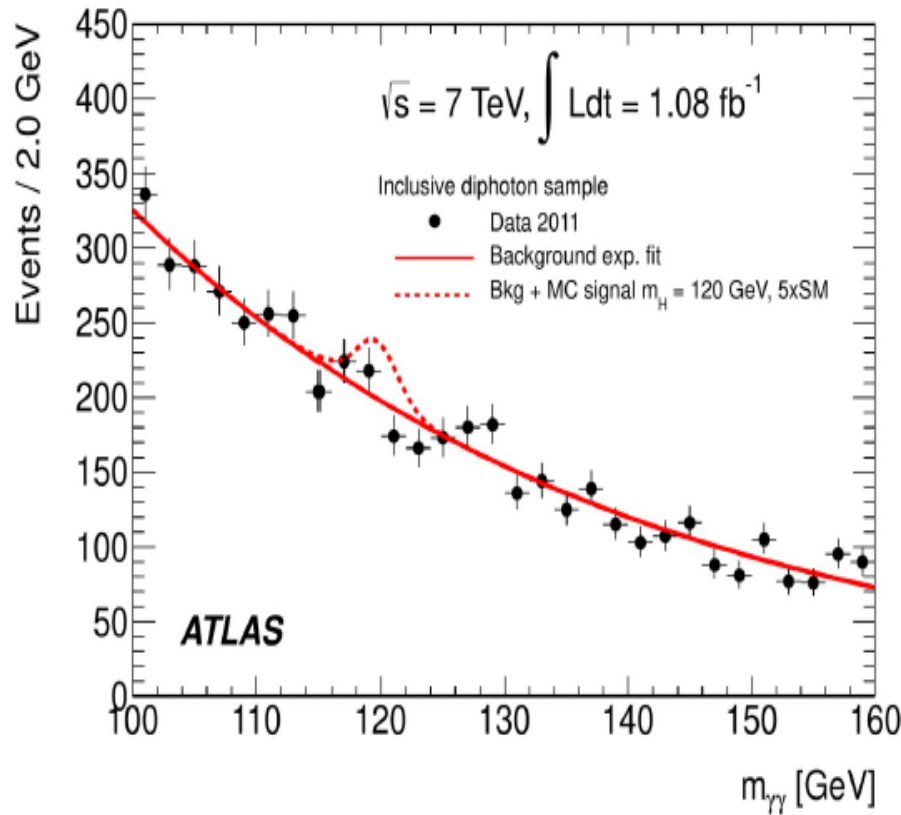
for large enough
datasets

$$P(\hat{\mu}) \propto \exp\left(-\frac{(\hat{\mu} - \mu^*)^2}{2\sigma_{\hat{\mu}}^2}\right) \quad \text{for } n \rightarrow \infty$$

Standard deviation of the distribution of $\hat{\mu}$

- **Asymptotically Efficient** : $\sigma_{\hat{\mu}}$ is the **lowest possible value** (in the limit $n \rightarrow \infty$) among consistent estimators.
→ MLE captures all the available information in the data
- Also **consistent**: $\hat{\mu}$ converges to the true value for large n , $\hat{\mu} \xrightarrow{n \rightarrow \infty} \mu^*$
- **Log-likelihood** : Can also **minimize** $\lambda = -2 \log L$
 - Usually more efficient numerically
 - For Gaussian L , λ is parabolic: $\lambda(\mu) = \left(\frac{\hat{\mu} - \mu}{\sigma_{\mu}}\right)^2$
- Can **drop multiplicative constants in L** (additive constants in λ)

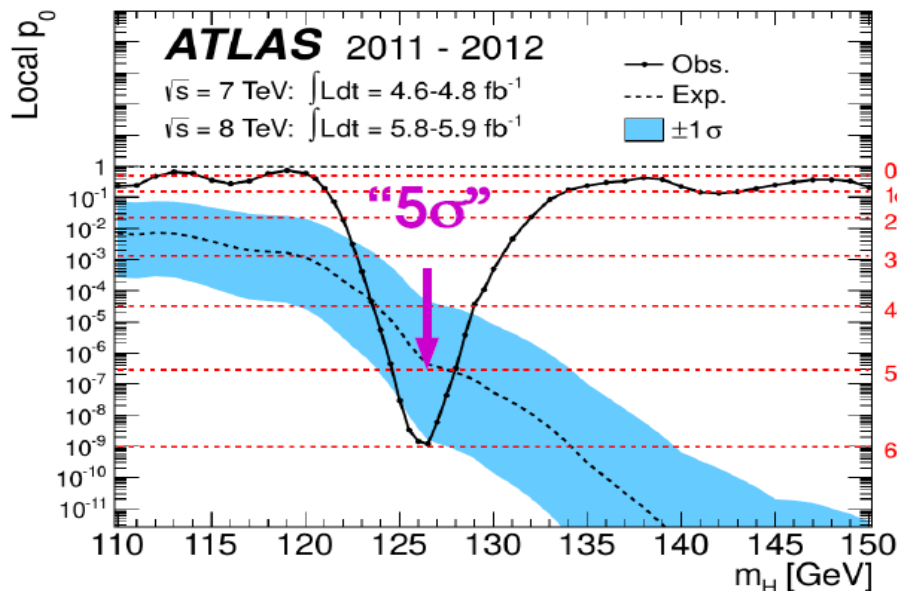
Hypothesis testing and discovery



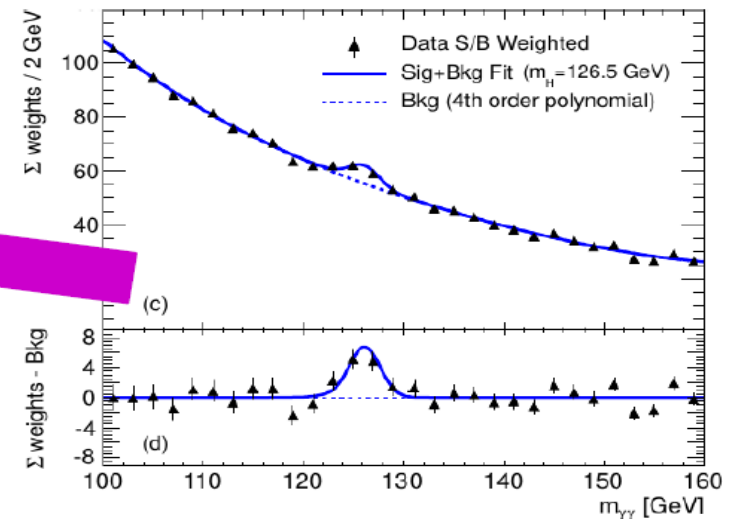
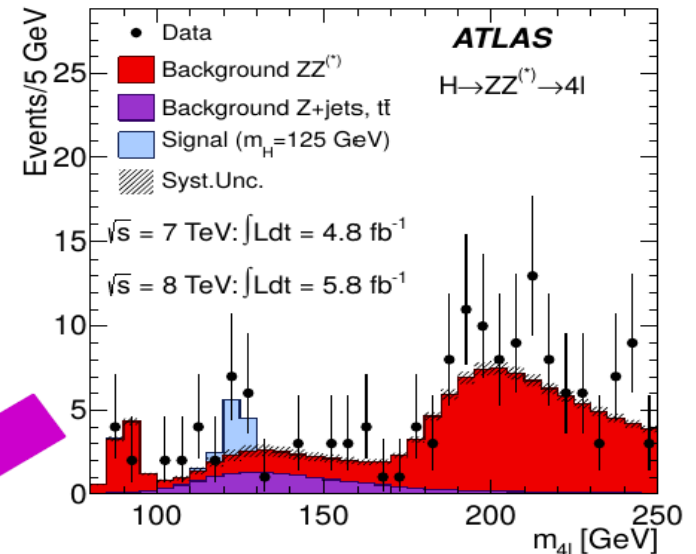
Discovery testing

We see an unexpected feature in our data, is it a signal for new physics or a fluctuation ?

e.g. Higgs discovery : **"We have 5 σ " !**



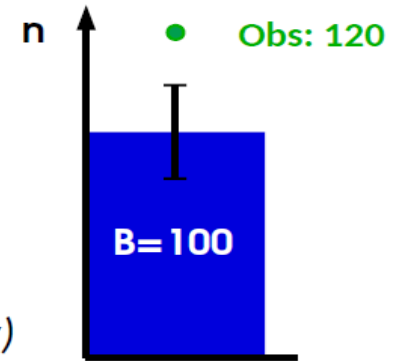
Phys. Lett. B 716 (2012) 1-29



Discovery testing

Say we have a Gaussian measurement with a background $B=100$, and we measure $n=120$

Did we just discover something? *Maybe :-)* (but not very likely)



The measured signal is $S = 20$.

$$S = n_{\text{obs}} - B$$

Uncertainty on B is $\sqrt{B} = 10$

$\Rightarrow$ Significance $Z = 2$

$\Rightarrow$ we are $\sim 2\sigma$ away from $S=0$.

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

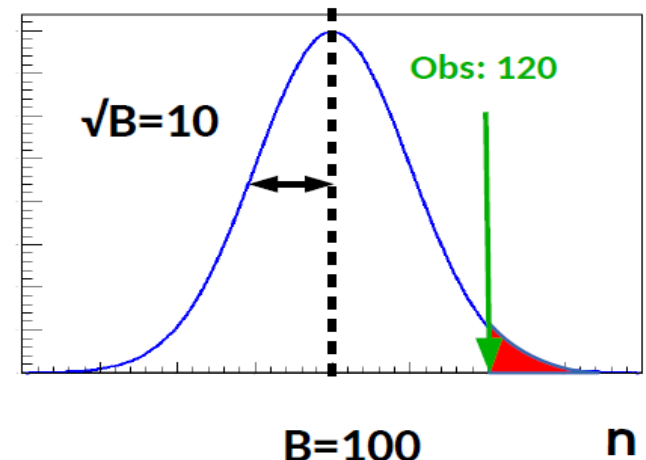
Gaussian quantiles :

$Z = 2$ happens $p_0 \sim 2.3\%$ of the time if $S=0$

P-value:

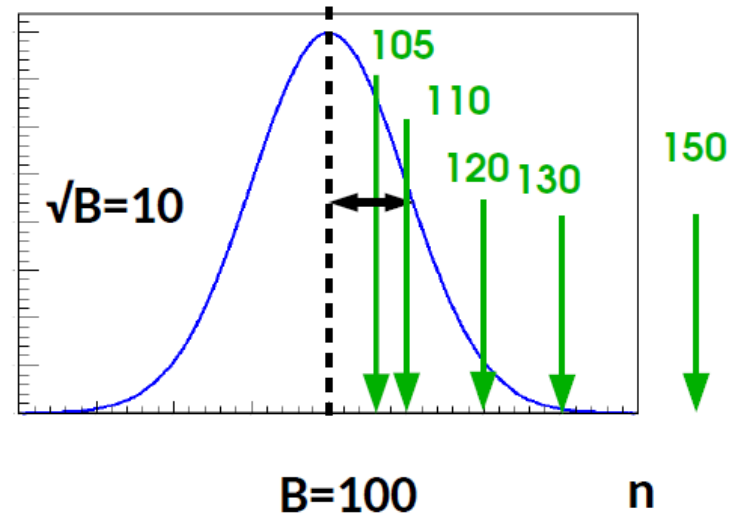
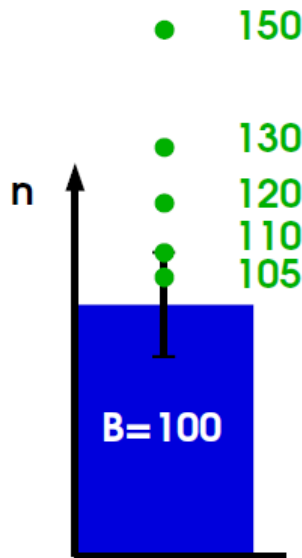
$$p_0 = 1 - \Phi(Z)$$

$\Rightarrow$ Rare, but not exceptional



$$\Phi(Z) = \int_{-\infty}^Z G(u; 0, 1) du$$

Discovery testing



n_{obs}	S	Z	p_0
105	5	0.5σ	31%
110	10	1σ	16%
120	20	2σ	2.3%
130	30	3σ	0.1%
150	50	5σ	$3 \cdot 10^{-7}$

Straightforward in this Gaussian case

Need to be able to do the same in more complex cases:

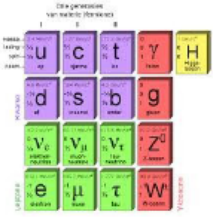
Evidence

Discovery

- Determine S
- Compute Z and p_0

Hypothesis Testing

Hypothesis: assumption on model parameters, say value of S (e.g. $H_0 : S=0$)
 → **Goal** : decide if H_0 is favored or disfavored using a test based on the data

Possible outcomes:	Data disfavors H_0 (Discovery claim)	Data favors H_0 (Nothing found)
H_0 is false (New physics!)	Discovery! 	Missed discovery Type-II error (1 - Power) 
H_0 is true (Nothing new)	False discovery claim Type-I error (→ p-value, significance) 	No new physics, none found 

"... the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only to give the facts a chance of disproving the null hypothesis." – R. A. Fisher

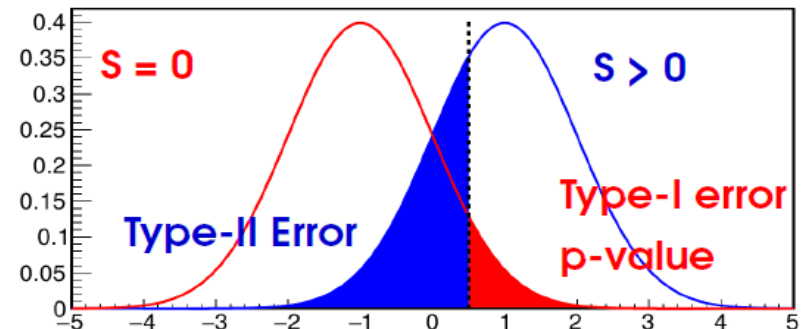
Hypothesis Testing

Hypothesis: assumption on model parameters, say value of S (e.g. $H_0 : S=0$)

	Data disfavors H_0 (Discovery claim)	Data favors H_0 (Nothing found)
H_0 is false (New physics!)	Discovery! 	Type-II error (Missed discovery) 
H_0 is true (Nothing new)	Type-I error (False discovery) 	No new physics, none found 

Lower Type-I errors $\Leftrightarrow$ **Higher Type-II errors** and vice versa: cannot have everything!

→ **Goal:** test that minimizes Type-II errors **for given level of Type-I error.**



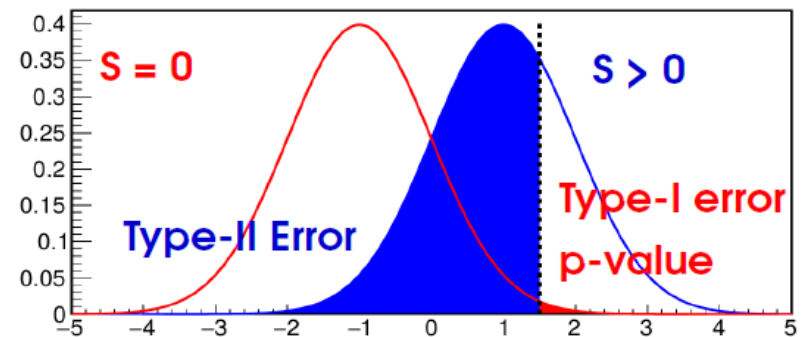
Hypothesis Testing

Hypothesis: assumption on model parameters, say value of S (e.g. $H_0 : S=0$)

	Data disfavors H_0 (Discovery claim)	Data favors H_0 (Nothing found)
H_0 is false (New physics!)	Discovery! 	Type-II error (Missed discovery) 
H_0 is true (Nothing new)	Type-I error (False discovery) 	No new physics, none found 

Lower Type-I errors $\Leftrightarrow$ **Higher Type-II errors** and vice versa: cannot have everything!

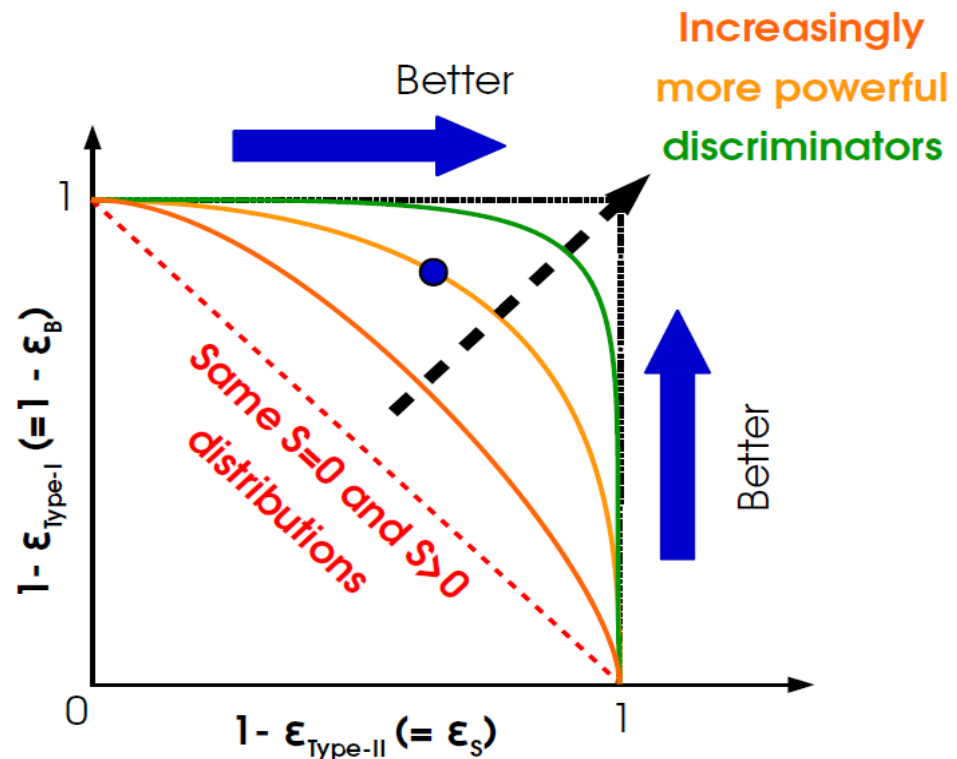
→ **Goal:** test that minimizes Type-II errors **for given level of Type-I error.**



ROC Curves

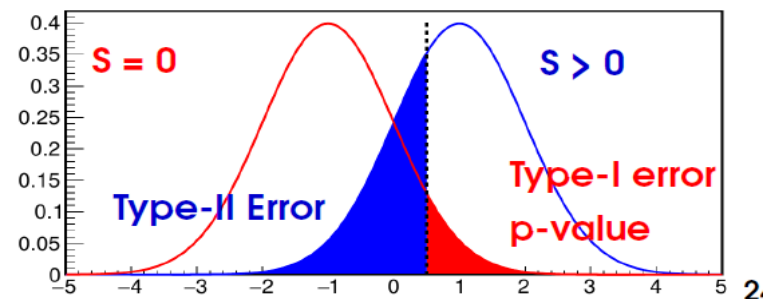
“Receiver operating characteristic” (ROC) Curve:

- Plot Type-I vs Type-II rates for different cut values
- All curves monotonically decrease from (0,1) to (1,0)
- Better discriminators more bent towards (1,1)



→ **Goal:** test that minimizes Type-II errors **for given level of Type-I error.**

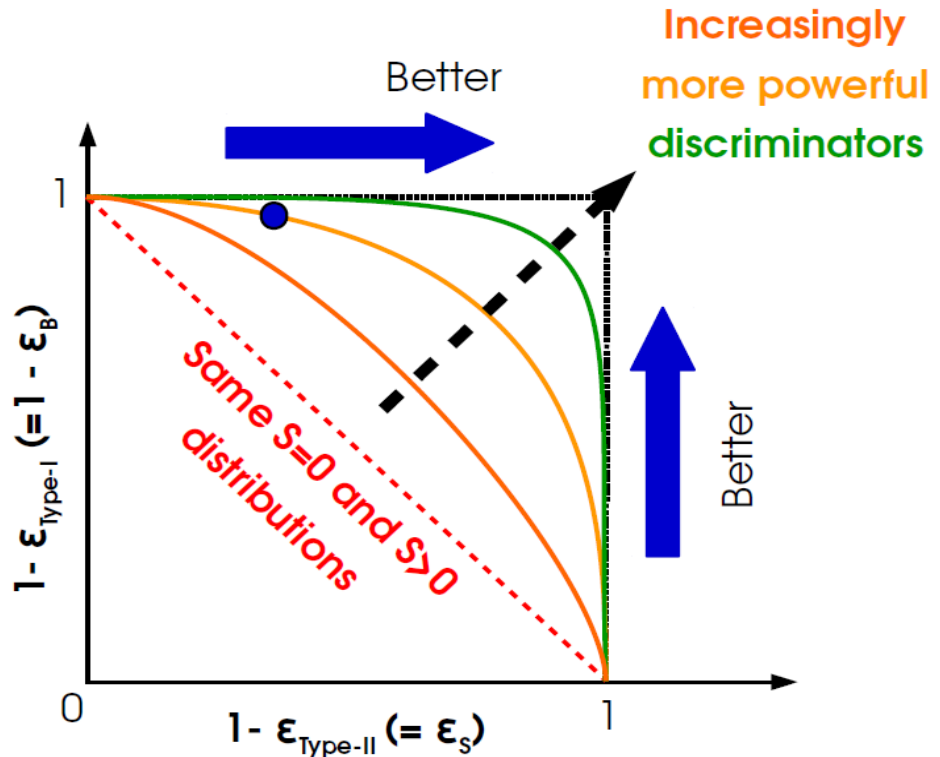
→ Usually set predefined level of **acceptable Type-I error** (e.g. “ 5σ ”)



ROC Curves

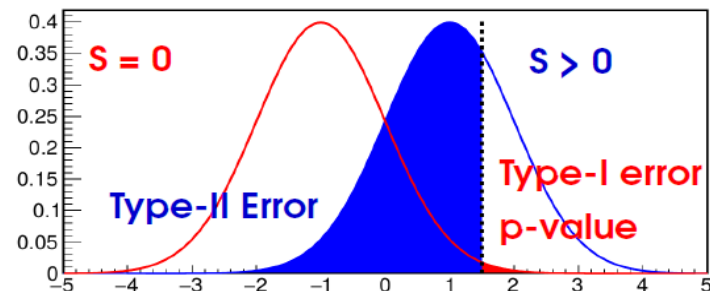
“Receiver operating characteristic” (ROC) Curve:

- Plot Type-I vs Type-II rates for different cut values
- All curves monotonically decrease from (0,1) to (1,0)
- Better discriminators more bent towards (1,1)



→ **Goal:** test that minimizes Type-II errors **for given level of Type-I error.**

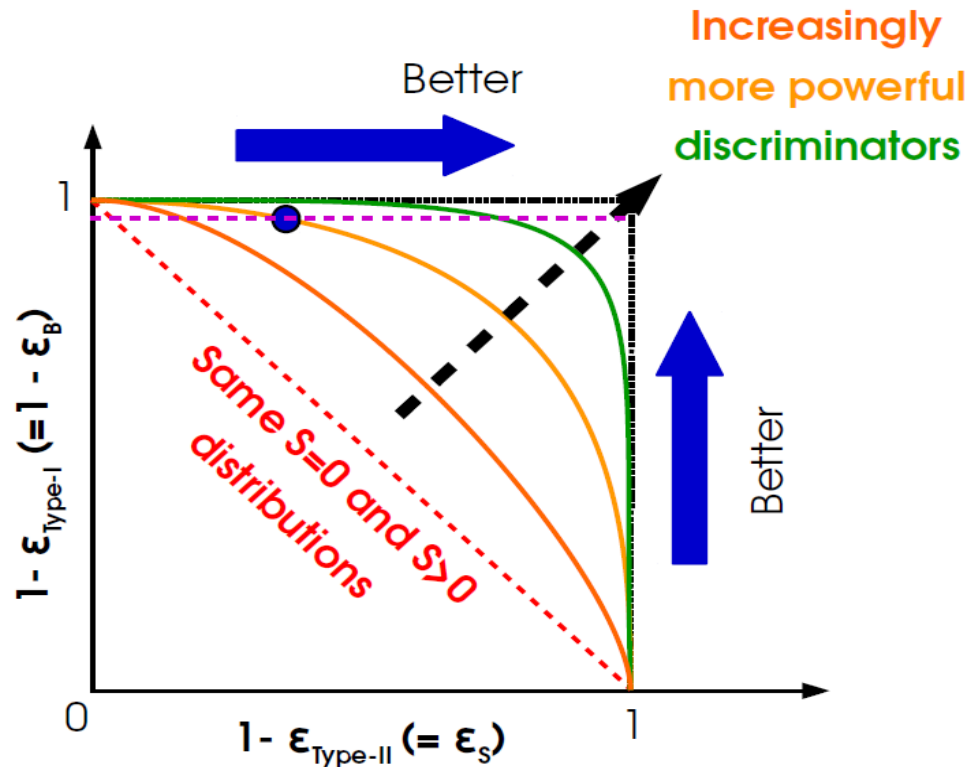
→ Usually set predefined level of **acceptable Type-I error** (e.g. “ 5σ ”)



ROC Curves

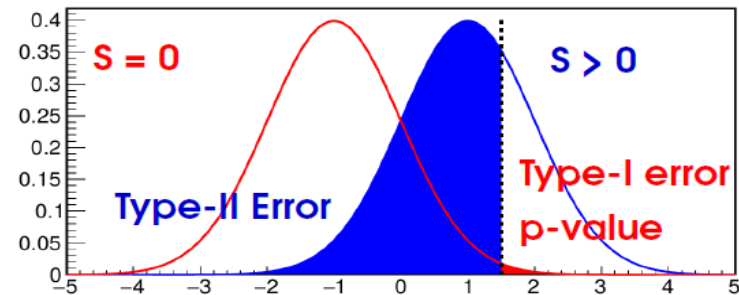
“Receiver operating characteristic” (ROC) Curve:

- Plot Type-I vs Type-II rates for different cut values
- All curves monotonically decrease from (0,1) to (1,0)
- Better discriminators more bent towards (1,1)

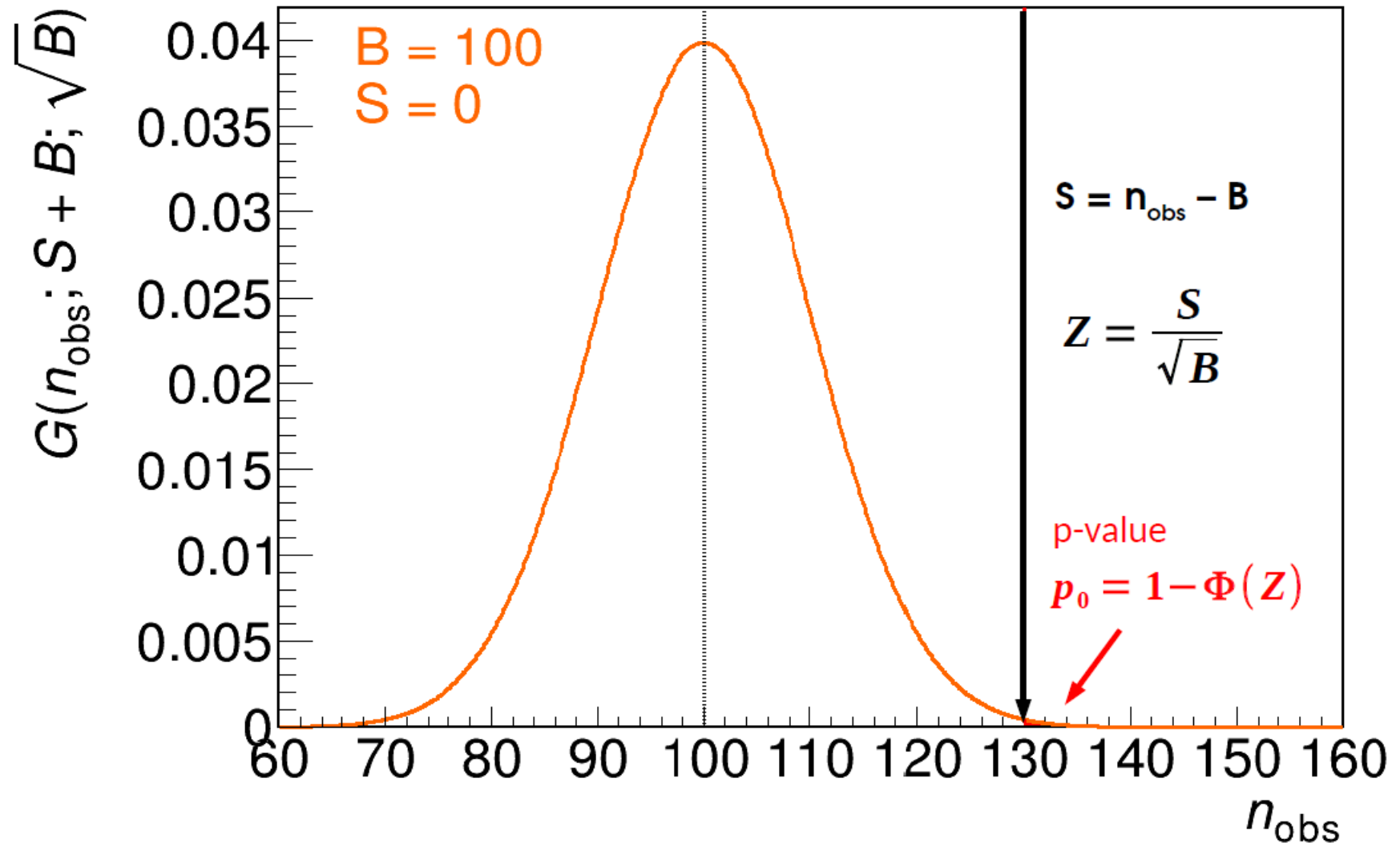


→ **Goal:** test that minimizes Type-II errors **for given level of Type-I error.**

→ Usually set predefined level of **acceptable Type-I error** (e.g. “ 5σ ”)



Discovery testing in Gaussian counting



Hypothesis testing with Likelihoods

Neyman-Pearson Lemma

When comparing two hypotheses H_0 and H_1 , the optimal discriminator is the **Likelihood ratio** (LR)

$$\frac{L(H_1; \text{data})}{L(H_0; \text{data})}$$

e.g.
$$\frac{L(S = 5; \text{data})}{L(S = 0; \text{data})}$$

As for MLE, choose the hypothesis that is more likely **given the data we have**.

- **Minimizes Type-II uncertainties** for given level of Type-I uncertainties
- Always need an **alternate hypothesis** to test against.

Caveat: Strictly true only for *simple hypotheses* (no free parameters)

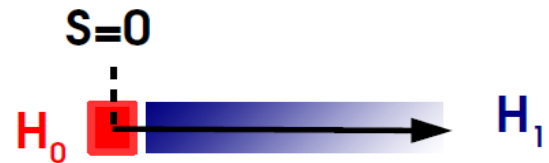
- **In the following:** all tests based on LR, will focus on p-values (Type-I errors), trusting that Type-II errors are anyway as small as they can be...

Discovery: Test Statistic

Cowan, Cranmer, Gross & Vitells,
[Eur.Phys.J.C71:1554,2011](#)

Discovery :

- H_0 : **background only** ($S = 0$) against
- H_1 : presence of a signal ($S > 0$)



→ For H_1 , any $S > 0$ is possible, which to use ? **The one preferred by the data, $\hat{S}$.**

⇒ Use LR $\frac{L(S=0)}{L(\hat{S})}$

→ In fact use the **test statistic**

$$q_0 = \begin{cases} -2 \log \frac{L(S=0)}{L(\hat{S})} & \hat{S} \geq 0 \\ 0 & \hat{S} < 0 \end{cases}$$

→ Set $q_0=0$ for $\hat{S} < 0$, same as for $\hat{S}=0$: negative signal is same as no signal

→ *one-sided* test statistic

Discovery: p-value

Large values of $-2 \log \frac{L(S=0)}{L(\hat{S})}$ if:

⇒ observed $\hat{S}$ is far from 0

⇒ $H_0(S=0)$ **disfavored** compared to $H_1(S \neq 0)$.

How large q_0 before we can exclude H_0 ?

(and **claim a discovery!**)

→ Need small Type-I rate (falsely accepting H_0)

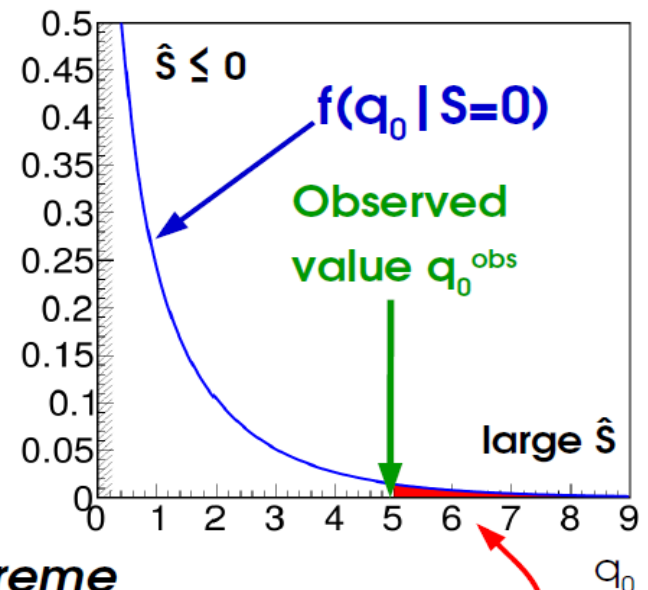
→ Type-I rate also known as the **p-value** p_0 :

*Fraction of outcomes that are **at least as extreme** (signal-like) **as data**, when H_0 is true (no signal present).*

→ Compute from the distribution $f(q_0 | S=0)$: $p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0 | S=0) dq_0$

→ Smaller p-value ⇒ Stronger case for discovery

data prefer $S = 0$ $\longleftrightarrow$ data prefer $S > 0$



Asymptotic distribution of q_0

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

→ Assume **Gaussian regime for $\hat{S}$** (e.g. large n_{evts} , Central-limit theorem)

⇒ **q_0 is distributed as a χ^2** under $H_0(S=0)$, for $\hat{S} \geq 0$: **Wilk's Theorem** (*)

$$f(q_0 | H_0, \hat{S} \geq 0) = f_{\chi^2(n_{\text{dof}}=1)}(q_0)$$

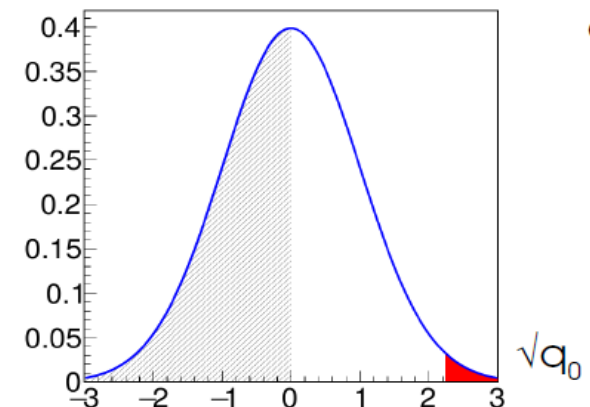
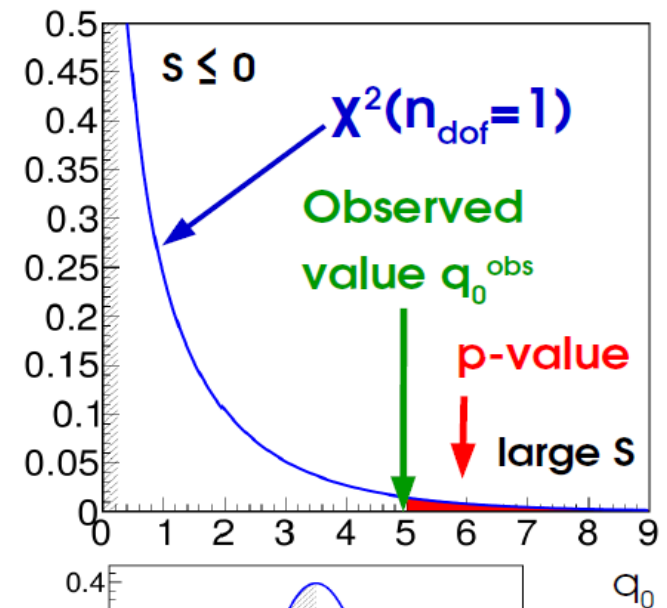
⇒ Can compute p-values from Gaussian quantiles

$$p_0 = 1 - \Phi(\sqrt{q_0}) \quad \text{By definition, } q_0 \sim \chi^2 \Rightarrow \sqrt{q_0} \sim G(0,1)$$

⇒ Even more simply, the significance is:

$$Z = \sqrt{q_0}$$

Typically works well already for event counts of $O(5)$ and above ⇒ Widely applicable



(*) 1-line "proof" : asymptotically L and S are Gaussian, so

$$L(S) = \exp\left[-\frac{1}{2}\left(\frac{S-\hat{S}}{\sigma}\right)^2\right] \Rightarrow q_0 = \left(\frac{\hat{S}}{\sigma}\right)^2 \Rightarrow \sqrt{q_0} = \frac{\hat{S}}{\sigma} \sim G(0,1) \Rightarrow q_0 \sim \chi^2(n_{\text{dof}}=1)$$

Discovery Thresholds

Evidence : $3\sigma \Leftrightarrow p_0 = 0.3\% \Leftrightarrow 1 \text{ chance in } 300$

Discovery: $5\sigma \Leftrightarrow p_0 = 3 \cdot 10^{-7} \Leftrightarrow 1 \text{ chance in } 3.5\text{M}$

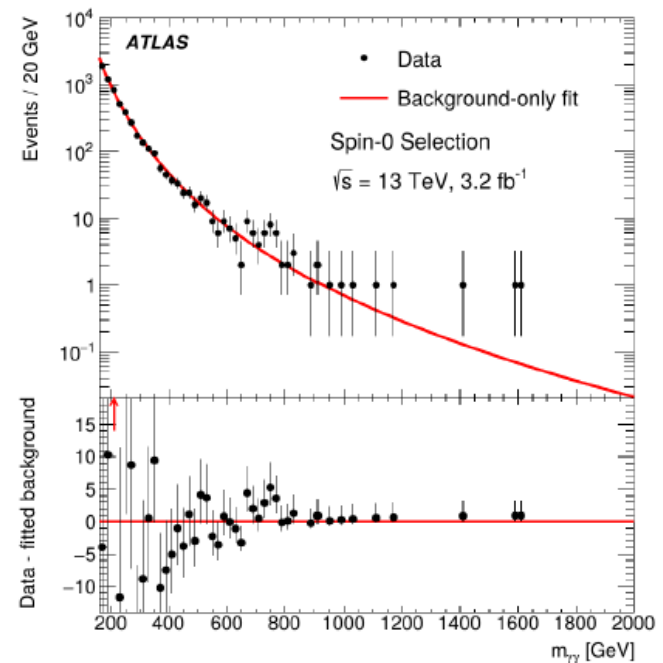
Why so high thresholds ? (from Louis Lyons):

- **Look-elsewhere effect**: searches typically cover multiple independent regions $\Rightarrow$ Higher chance to have a fluctuation “somewhere”

$N_{\text{trials}} \sim 1000$: local $5\sigma \Leftrightarrow O(10^{-4})$ more reasonable

- **Mismodeled systematics**: factor 2 error in syst-dominated analysis $\Rightarrow$ factor 2 error on Z...
- **History**: 3σ and 4σ excesses do occur regularly, for the reasons above

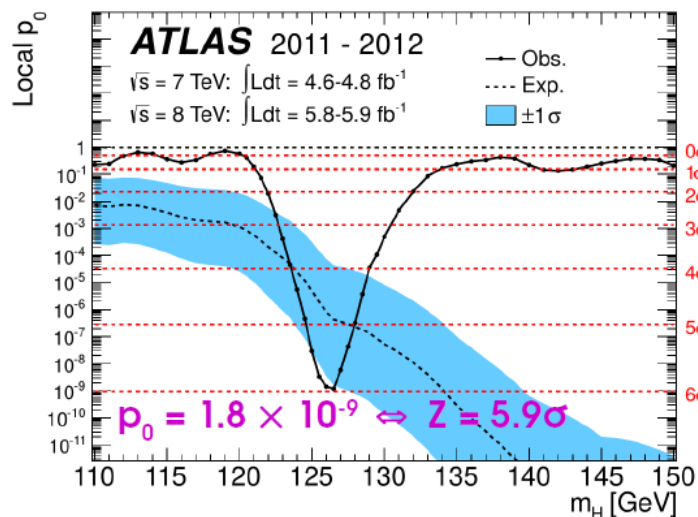
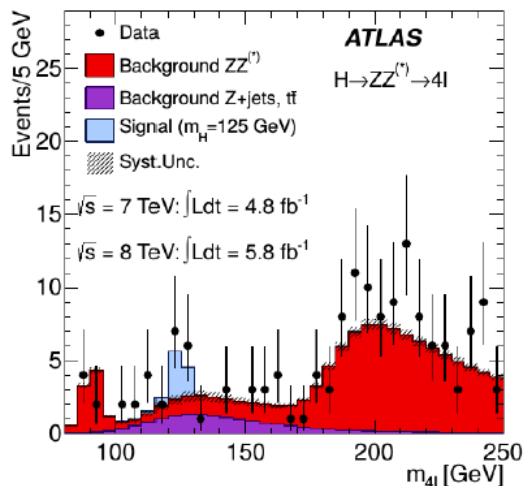
Extraordinary claims require extraordinary evidence!



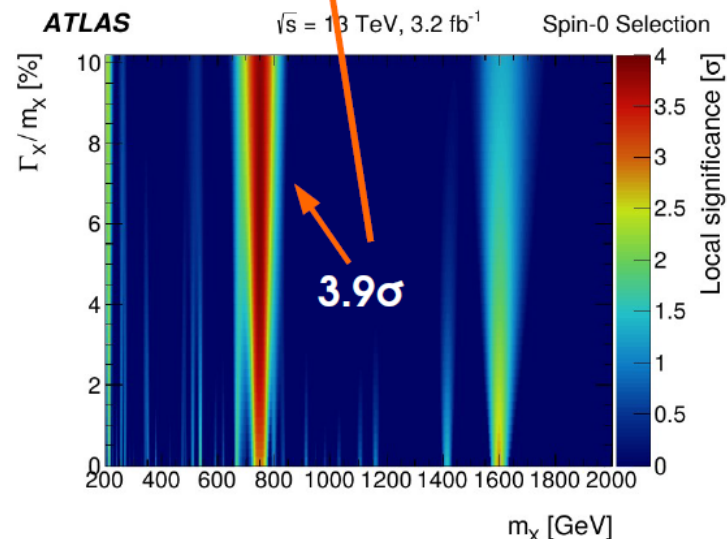
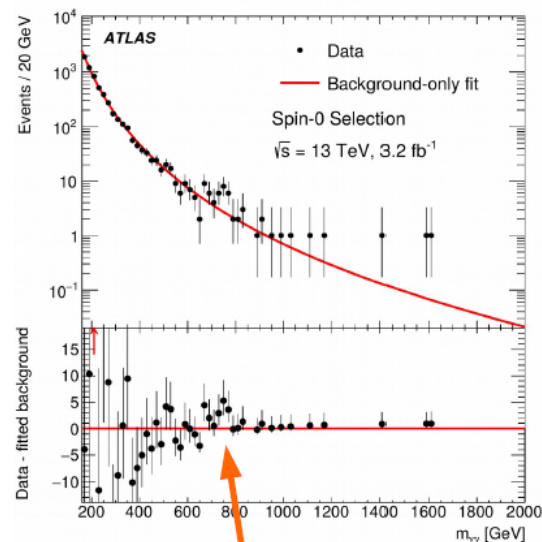
Some examples

Some Examples

Higgs Discovery: [Phys. Lett. B 716 \(2012\) 1-29](#)



High-mass $X \rightarrow \gamma\gamma$ Search: [JHEP 09 \(2016\) 1](#)



Highlights: Hypothesis Tests and Discovery

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

To estimate a parameter, use the value $\hat{\mu}$ that maximizes $L(\mu) \rightarrow$ 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 $q_0 = -2 \log \frac{L(S=0)}{L(\hat{S})} \quad \hat{S} \geq 0$

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

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

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

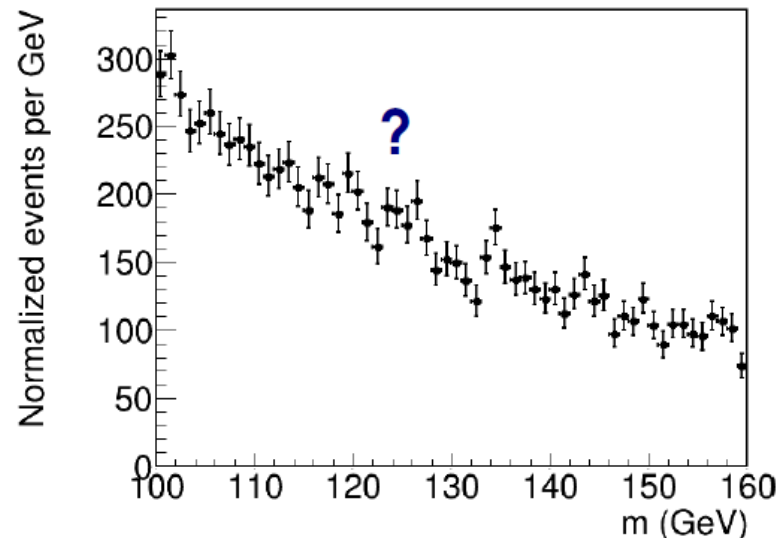
Hypothesis tests for Limits

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"



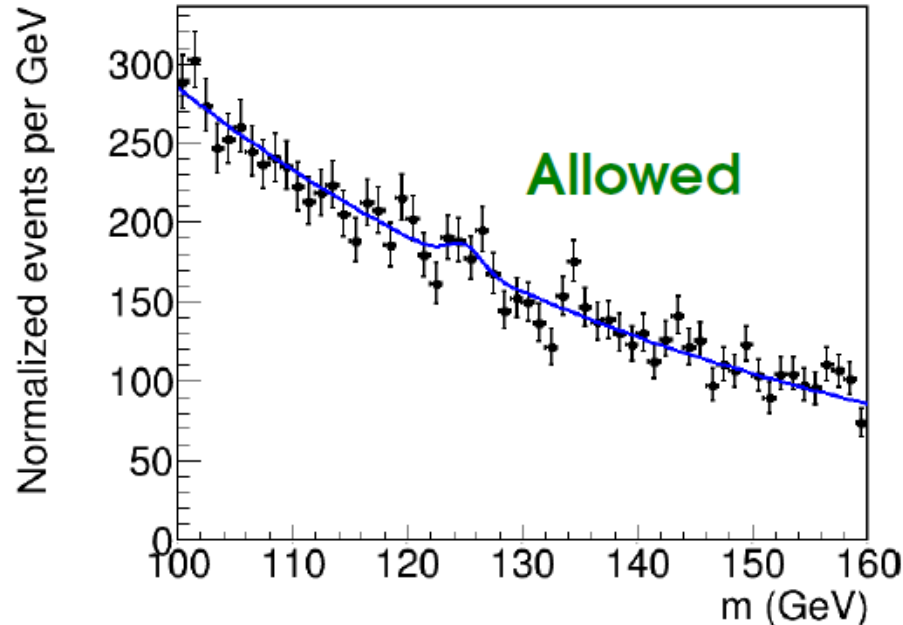
Hypothesis tests for Limits

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"



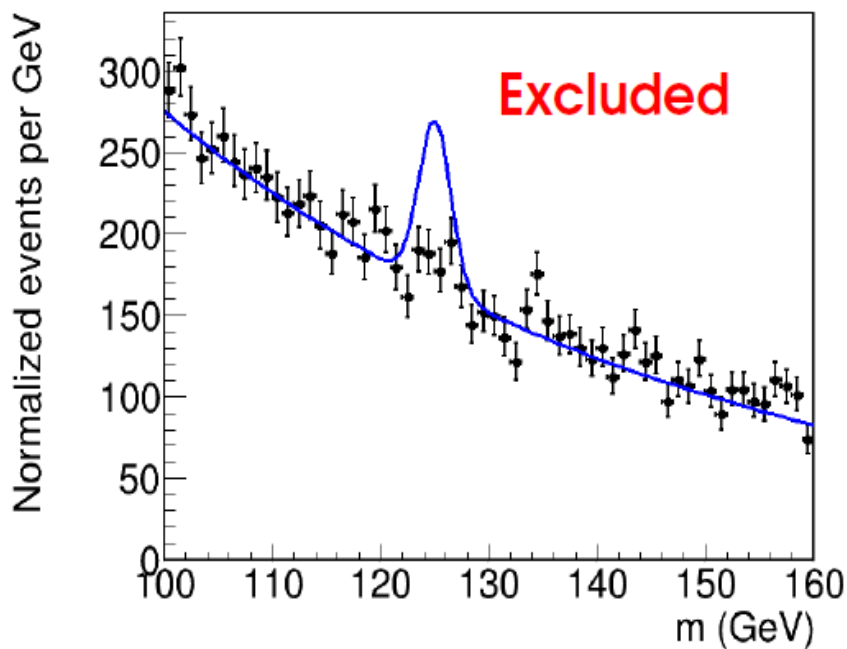
Hypothesis tests for Limits

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”



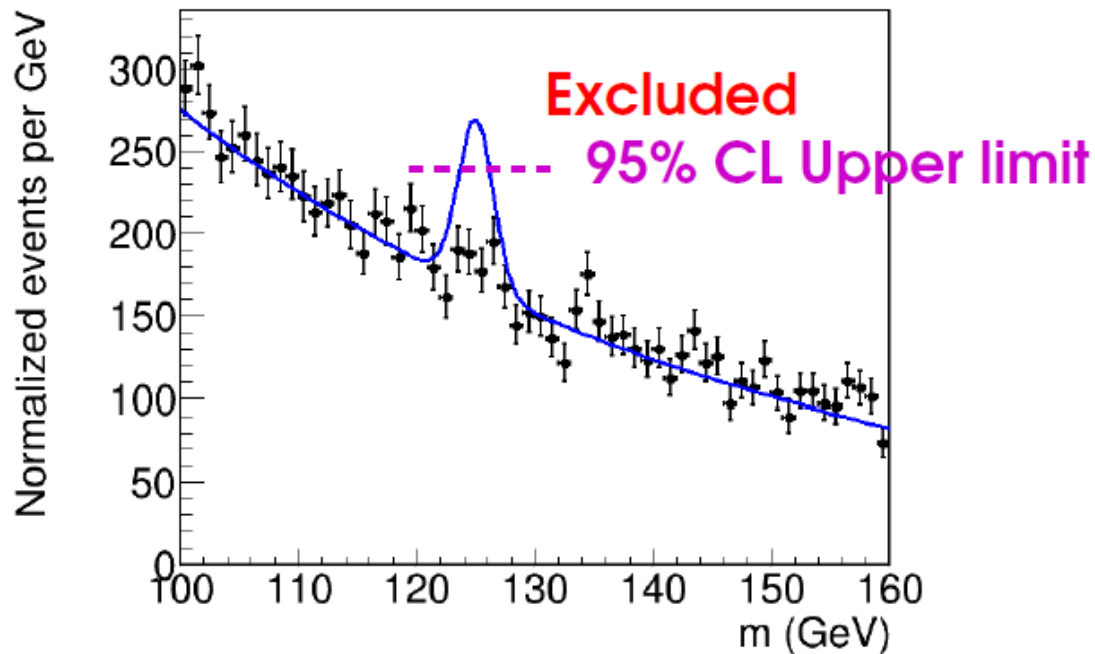
Hypothesis tests for Limits

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**

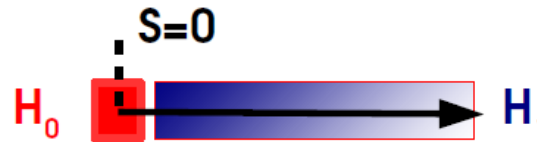
→ Typically report **95% CL** upper limit (p-value = 5%) : “ $S < S_0$ @ 95% CL”



Test Statistic for Limit-Setting

Discovery :

- $H_0 : S = 0$
- $H_1 : S > 0$



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

Compare

← Likelihood of H_0 ($\hat{S} > 0$)

← Likelihood of H_1

Limit-setting

- $H_0 : S = S_0$
- $H_1 : S < S_0$



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

Compare

← Likelihood of H_0 ($\hat{S} < S_0$)

← Likelihood of H_1

Same as q_0 :

→ large values $\Rightarrow$ good rejection of H_0 .

$\Rightarrow$ Can compute p-value from q_{S_0} .

Inversion: Getting the limit for a given CL

Procedure:

→ Compute q_{s_0} for some S_0 , get the **exclusion p-value** p_{s_0} .

Asymptotic case: can use $p_{s_0} = 1 - \Phi(\sqrt{q_{s_0}})$

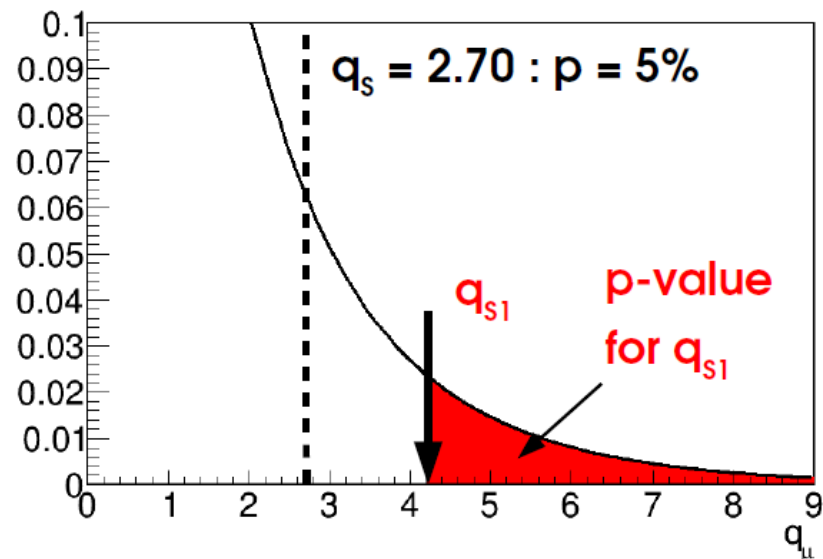
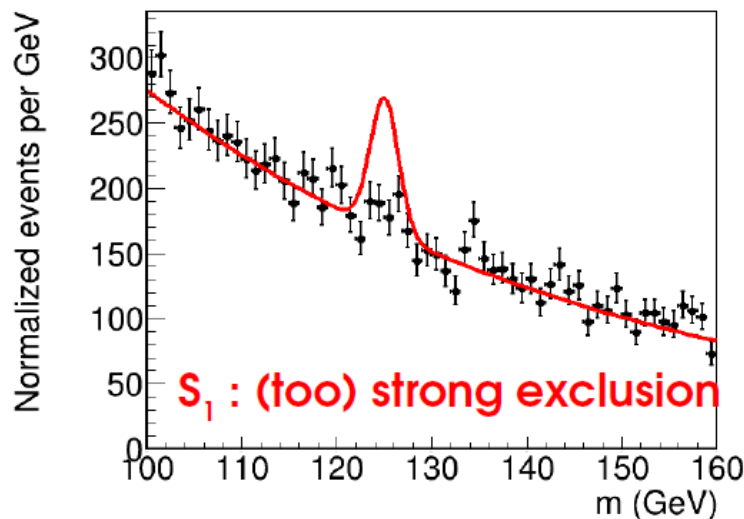
→ Adjust S_0 until 95% CL exclusion ($p_{s_0} = 5\%$) is reached

Asymptotic case: need $q_{s_0} = 2.70$

Asymptotics

$$\sqrt{q_{s_0}} = \Phi^{-1}(1 - p_0)$$

CL	Region
90%	$q_s > 1.64$
95%	$q_s > 2.70$
99%	$q_s > 5.41$



Inversion: Getting the limit for a given CL

Procedure:

→ Compute q_{s_0} for some S_0 , get the **exclusion p-value** p_{s_0} .

Asymptotic case: can use $p_{s_0} = 1 - \Phi(\sqrt{q_{s_0}})$

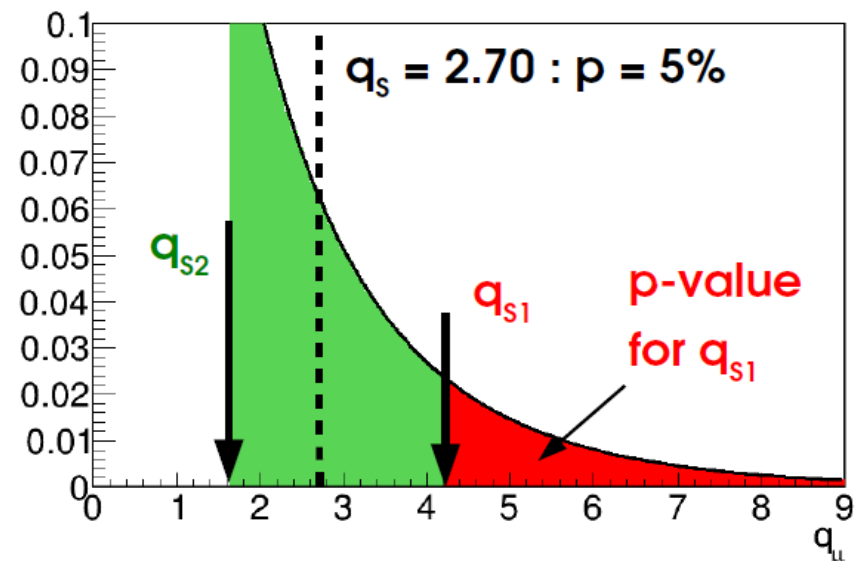
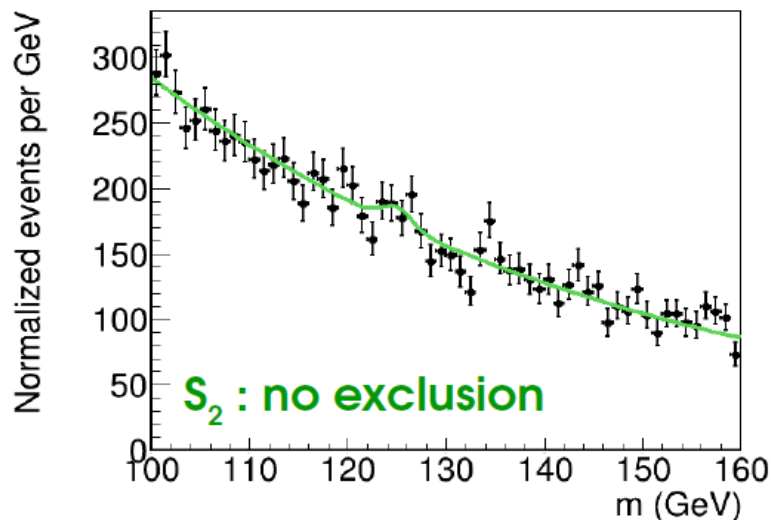
→ Adjust S_0 until 95% CL exclusion ($p_{s_0} = 5\%$) is reached

Asymptotic case: need $q_{s_0} = 2.70$

Asymptotics

$$\sqrt{q_{s_0}} = \Phi^{-1}(1 - p_0)$$

CL	Region
90%	$q_s > 1.64$
95%	$q_s > 2.70$
99%	$q_s > 5.41$



Inversion: Getting the limit for a given CL

Procedure:

→ Compute q_{S_0} for some S_0 , get the **exclusion p-value** p_{S_0} .

Asymptotic case: can use $p_{S_0} = 1 - \Phi(\sqrt{q_{S_0}})$

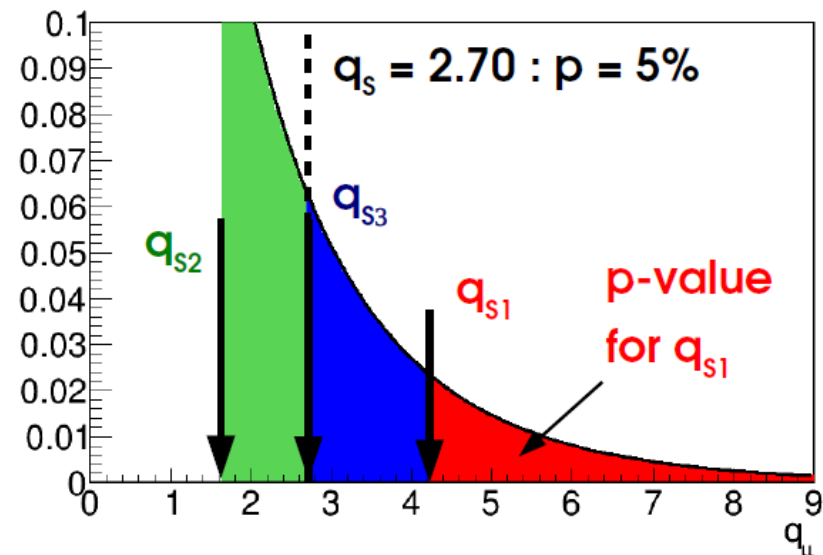
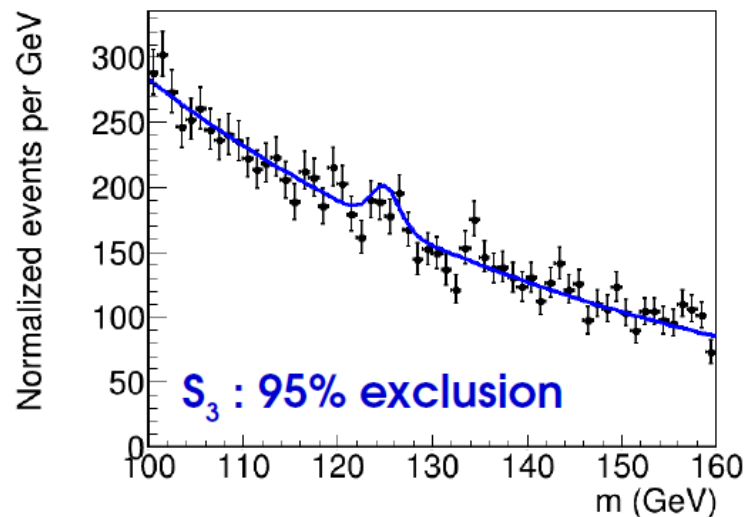
→ **Adjust S_0 until 95% CL exclusion ($p_{S_0} = 5\%$) is reached**

Asymptotic case: need $q_{S_0} = 2.70$

Asymptotics

$$\sqrt{q_{S_0}} = \Phi^{-1}(1 - p_0)$$

CL	Region
90%	$q_s > 1.64$
95%	$q_s > 2.70$
99%	$q_s > 5.41$



Upper Limit Pathologies

Upper limit: $S_{\text{up}} \sim \hat{S} + 1.64 \sigma_s$.

Problem: for negative $\hat{S}$, get **very** good observed limit.

→ For $\hat{S}$ sufficiently negative, even $S_{\text{up}} < 0$!

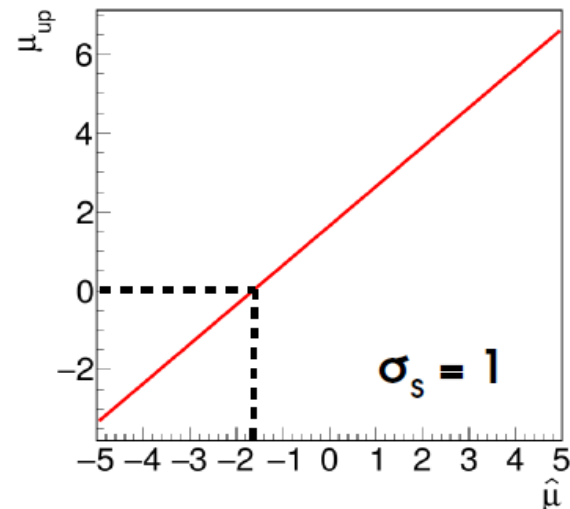
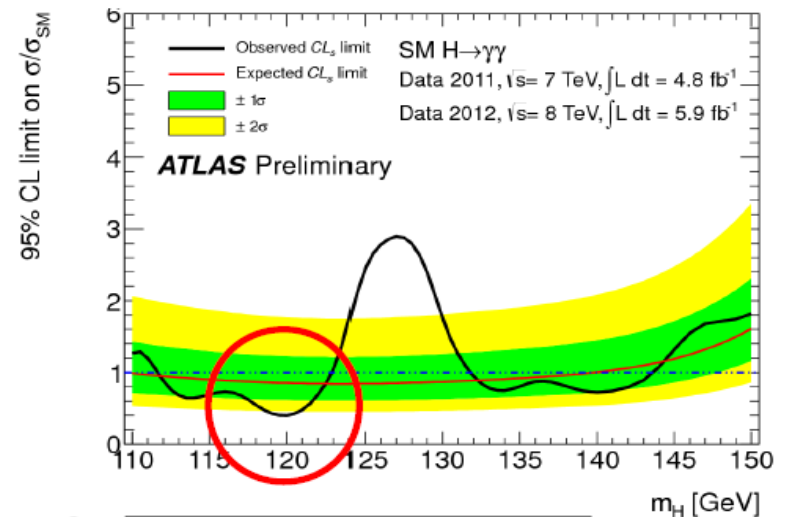
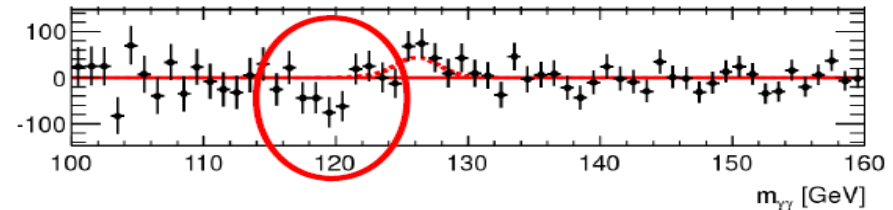
How can this be ?

→ **Background modeling issue ?...** Or:

→ This is a **95%** limit $\Rightarrow$ **5%** of the time, the limit wrongly excludes the true value, e.g. $S^*=0$.

Options

- **live with it:** sometimes report limit < 0
- **Special procedure to avoid these cases,** since if we assume S must be > 0 , we know a priori this is just a fluctuation.



CL_s

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

Usual solution in HEP : **CL_s**.

→ Compute modified p-value

⇒ **Rescale** exclusion at S_0 by exclusion at $S=0$.

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

$\hat{S}$ compatible with 0 : $p_B \sim O(1)$

$p_{CL_s} \sim p_{S_0} \sim 5\%$, no change.

Far-negative $\hat{S}$: $1 - p_B \ll 1$

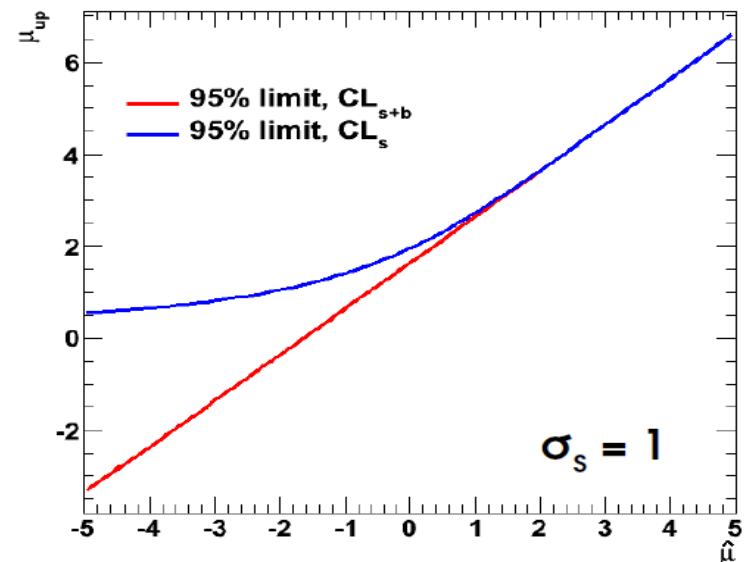
$p_{CL_s} \sim p_{S_0}/(1-p_B) \gg 5\%$

→ lower exclusion ⇒ higher limit,
usually >0 as desired

$$p_{CL_s} = \frac{p_{S_0}}{1 - p_B}$$

The usual p-value under $H(S=S_0)$ (=5%)

The p-value computed under $H(S=0)$

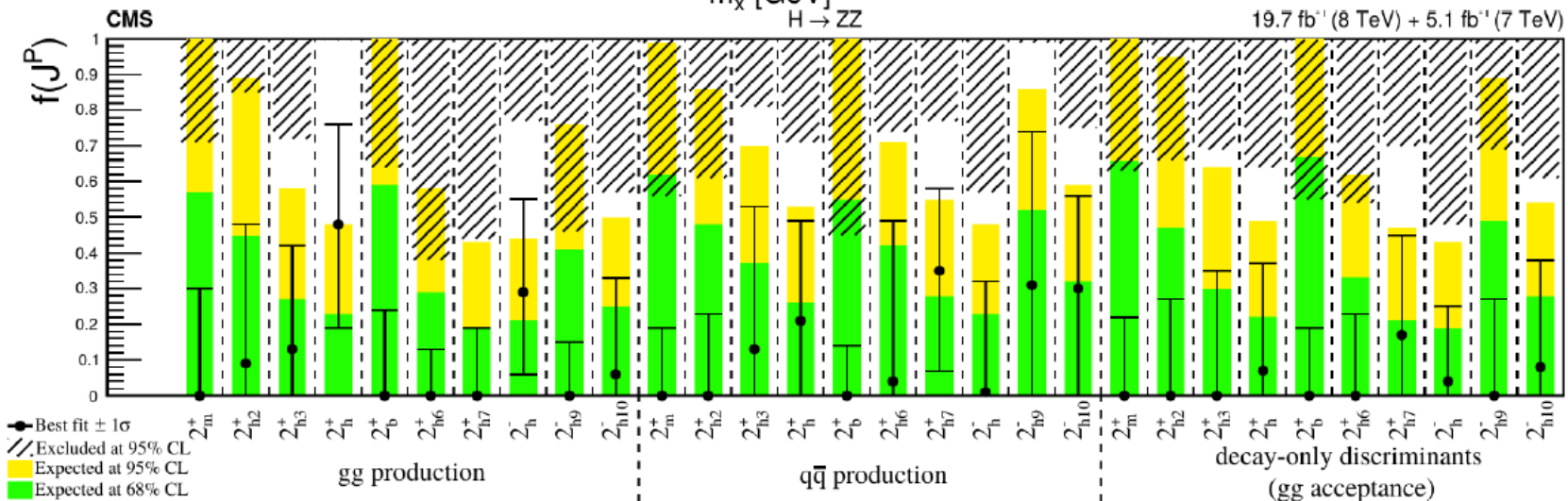
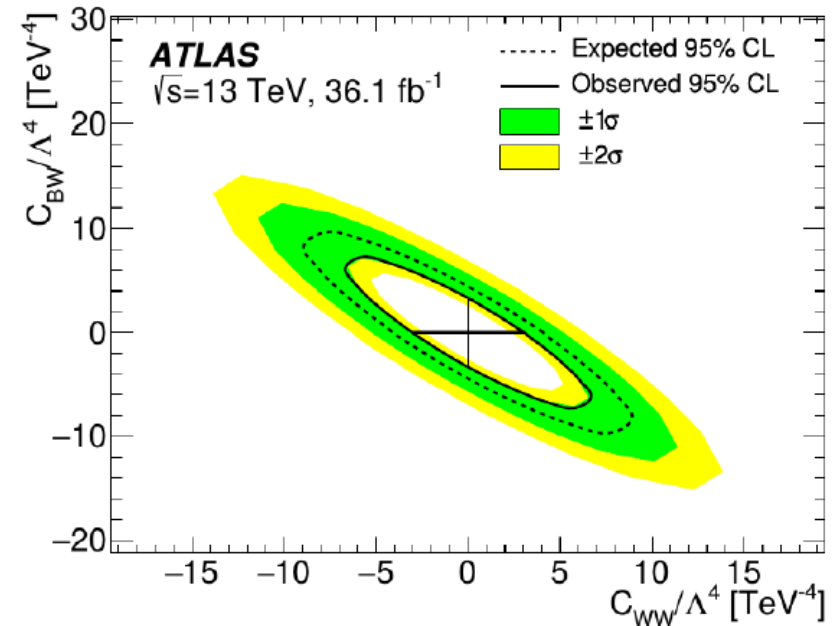
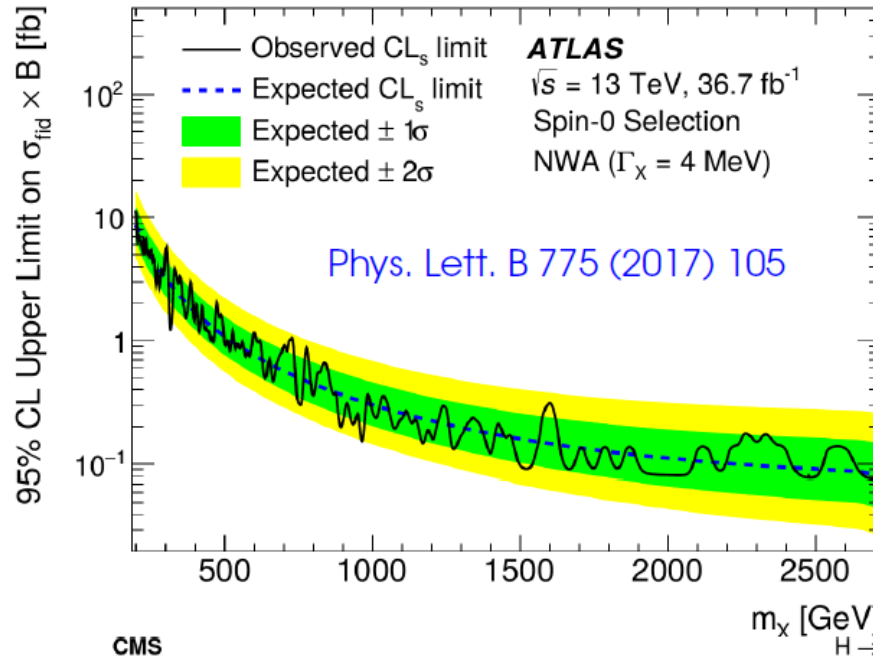


Drawback: overcoverage

→ limit is claimed to be 95% CL, but actually $>95\%$ CL for small $1-p_B$.

Upper Limit Examples

ATLAS 2015-2016 4l α TGC Search



Phys. Rev. D 92 (2015) 012004

Gaussian Intervals

If $\hat{\mu} \sim G(\mu^*, \sigma)$, known quantiles :

$$P(\mu^* - \sigma < \hat{\mu} < \mu^* + \sigma) = 68 \%$$

This is a probability for $\hat{\mu}$, not μ !

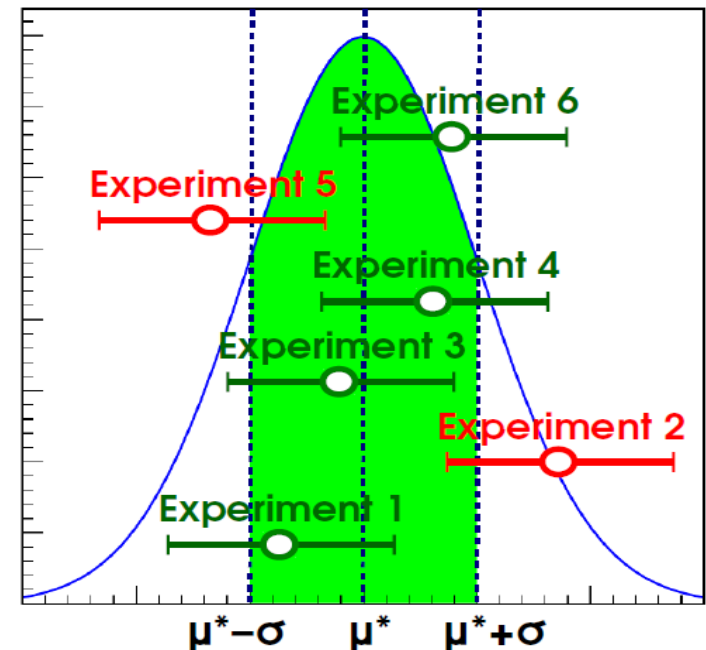
→ μ^* is a **fixed number**, **not a random variable**

But we can invert the relation:

$$P(\mu^* - \sigma < \hat{\mu} < \mu^* + \sigma) = 68 \%$$

$$\Rightarrow P(|\hat{\mu} - \mu^*| < \sigma) = 68 \%$$

$$\Rightarrow P(\hat{\mu} - \sigma < \mu^* < \hat{\mu} + \sigma) = 68 \%$$



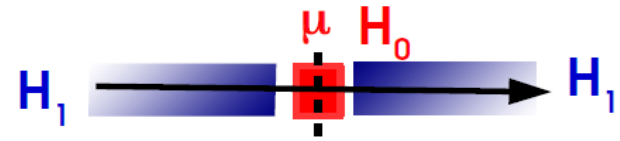
→ This gives the desired statement on μ^* : *if we repeat the experiment many times, $[\hat{\mu} - \sigma, \hat{\mu} + \sigma]$ will contain the true value 68.3% of the time: $\mu^* = \hat{\mu} \pm \sigma$*

This is a statement on the interval $[\hat{\mu} - \sigma, \hat{\mu} + \sigma]$ obtained for each experiment

Works in the same way for other interval sizes: $[\hat{\mu} - Z\sigma, \hat{\mu} + Z\sigma]$ with

Z	1	1.96	2
CL	0.683	0.95	0.955

Likelihood Intervals



Confidence intervals from L:

- Test $H(\mu_0)$ against alternative using
- Two-sided test since true value can be higher or lower than observed

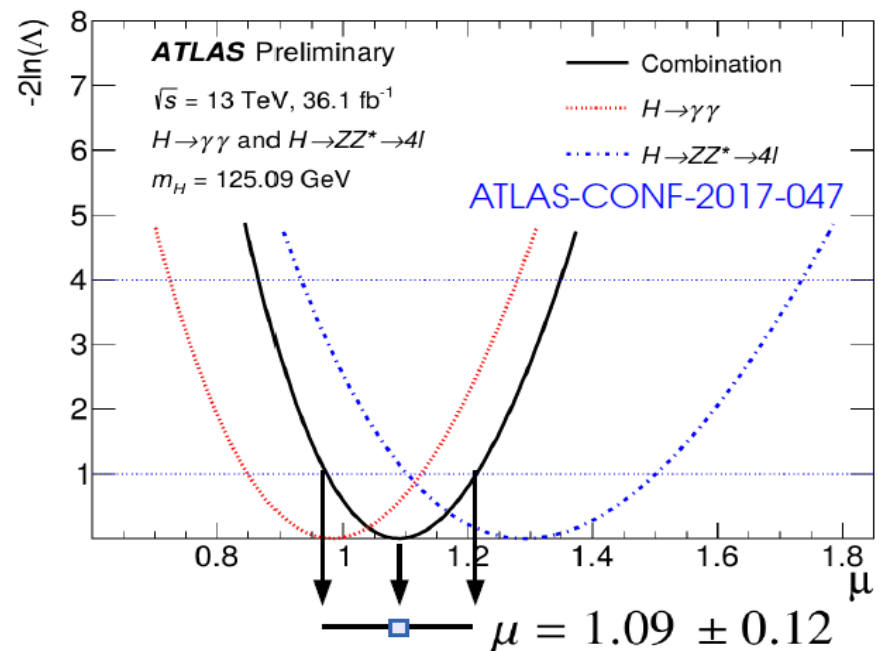
$$t_{\mu_0} = -2 \log \frac{L(\mu = \mu_0)}{L(\hat{\mu})}$$

Gaussian L:

- $t_{\mu_0} = \left(\frac{\hat{\mu} - \mu_0}{\sigma_{\mu}} \right)^2$: parabolic in μ_0 .
- Minimum occurs at $\mu = \hat{\mu}$
- Crossings with $t_{\mu} = 1$ give the 1σ interval

General case:

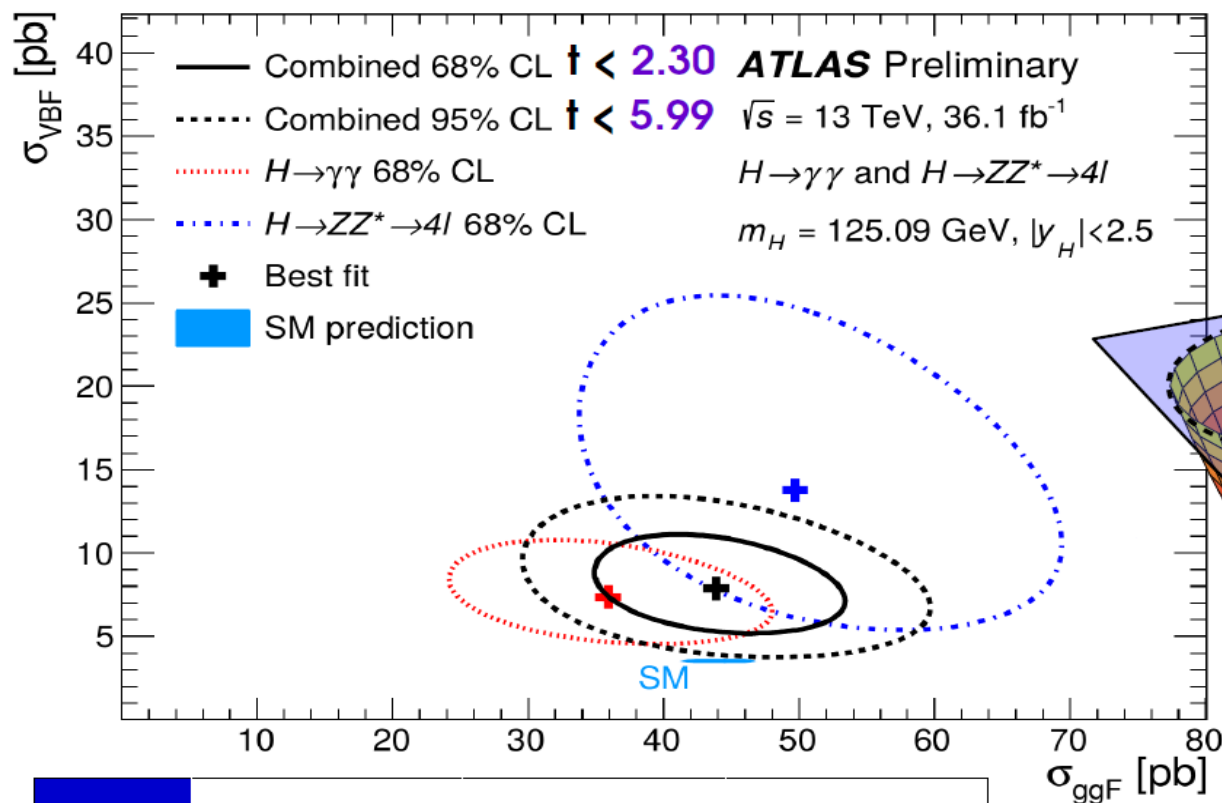
- Generally not a perfect parabola
- Minimum still occurs at $\mu = \hat{\mu}$
- Still define 1σ interval from the $t_{\mu} = \pm 1$ crossings



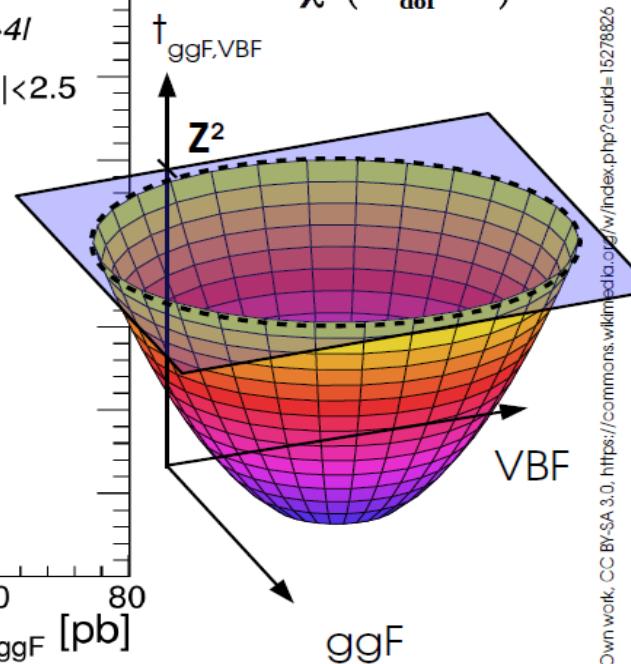
2D examples

2D Example: Higgs σ_{VBF} vs. σ_{ggF}

ATLAS-CONF-2017-047



$$t = -2 \log \frac{L(X_0, Y_0)}{L(\hat{X}, \hat{Y})} \sim \chi^2(N_{\text{dof}}=2)$$



Gaussian case: elliptic paraboloid surface

CL	68% (1σ)	95%	95.5% (2σ)
1D Z^2	1	3.84	4
2D Z^2	2.30	5.99	6.18

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

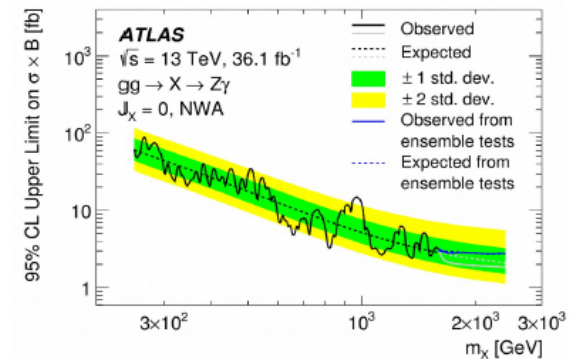
Takeaways

Limits : use LR-based test statistic:

$$q_{s_0} = -2 \log \frac{L(S=S_0)}{L(\hat{S})} \quad \hat{S} \leq S_0$$

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

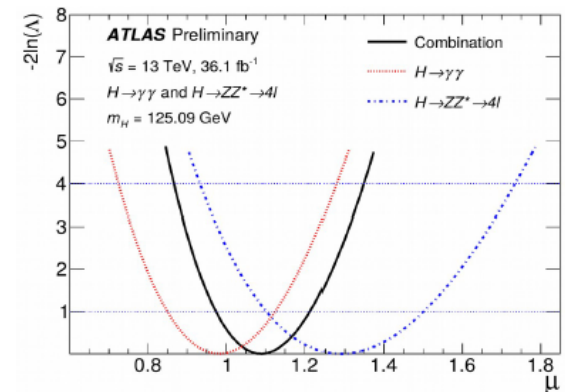
Poisson regime, $n=0$: **S_{up} = 3 events**



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

→ Crossings with $t_{\mu_0} = Z^2$ for $\pm Z\sigma$ intervals (in 1D)

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

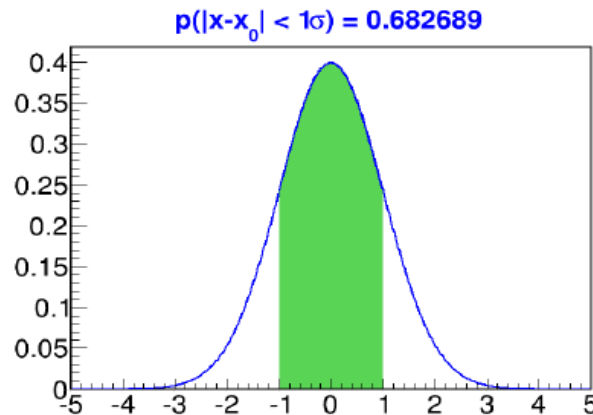


Extra slides

Discovery significance

Interesting p-values are quite small
 ⇒ express in terms of Gaussian quantiles

→ **Significance Z**



$$p_0 = 1 - \int_{-Z}^{+Z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$= 1 - 2 \Phi(Z)$$

$$\Phi(Z) = \int_{-\infty}^Z G(u; 0, 1) du$$

Z	p-value
1	0.32
2	0.045
3	0.003
5	6×10^{-7}

In ROOT:

$p_0 \rightarrow Z \ (\Phi)$: `ROOT::Math::gaussian_quantile_c`

$Z \rightarrow p_0 \ (\Phi^{-1})$: `ROOT::Math::gaussian_cdf_c`

⇒ How small is small enough ?

→ Conventionally, discovery for $p_0 = 6 \cdot 10^{-7} \Leftrightarrow Z = 5\sigma$

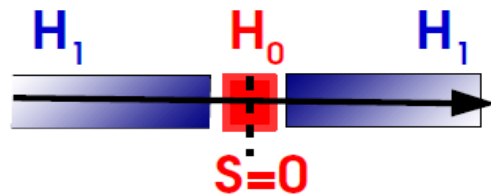
One-sided vs. Two-sided

If $\hat{S} < 0$, is it a *discovery* ? (does reject the $S=0$ hypothesis...)

Usual assumption : only $\hat{S} > 0$ is a *bona fide* signal

$\Rightarrow$ Change statistic so that $\hat{S} < 0 \Rightarrow t_0 = 0$ (perfect agreement with H_0 , as for $\hat{S} = 0$)

Two-sided

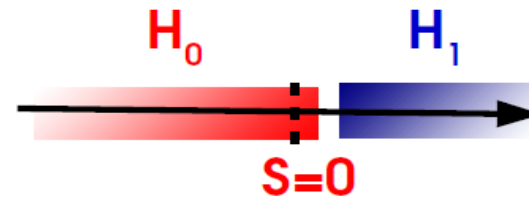


$$t_0 = -2 \log \frac{L(S=0)}{L(\hat{S})}$$

$$Z = \Phi^{-1}\left(1 - \frac{p_0}{2}\right) = \sqrt{t_0}$$

By convention, factor 2
in p-values for a given Z

One-sided



$$q_0 = \begin{cases} -2 \log \frac{L(S=0)}{L(\hat{S})} & \hat{S} \geq 0 \\ 0 & \hat{S} < 0 \end{cases}$$

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

$\Rightarrow$ Same Z in both cases
for a given signal S

p_0	Z	p_0
0.32	1	0.16
0.003	3	0.0015
6×10^{-7}	5	3×10^{-7}

One-sided Asymptotics

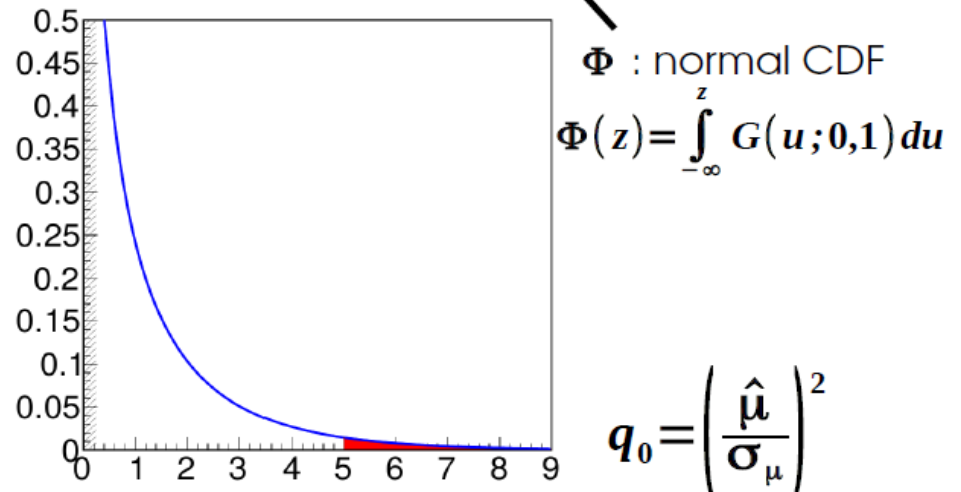
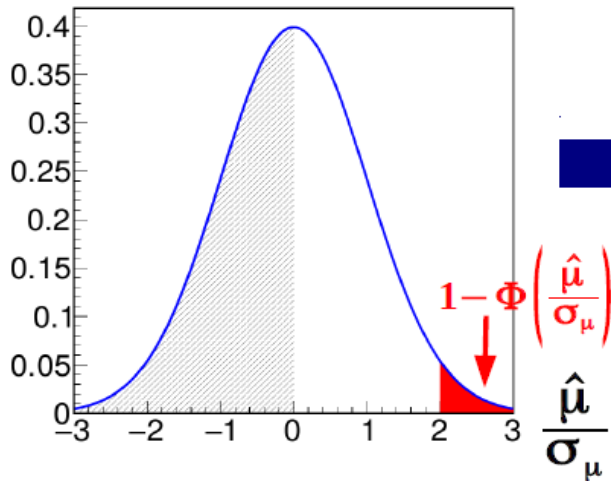
→ One-sided test:



$$q_0 = \begin{cases} -2 \log \frac{L(S=0)}{L(\hat{S})} & \hat{S} \geq 0 \\ 0 & \hat{S} < 0 \end{cases}$$

Asymptotics: “half- χ^2 ” distribution: $f(q_0 | S=0) = \frac{1}{2} \delta(q_0) + \frac{1}{2} f_{\chi^2(n_{dof}=1)}(q_0)$

Discovery p-value: $p_0 = 1 - \Phi(\sqrt{q_0})$ Significance: $Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}$

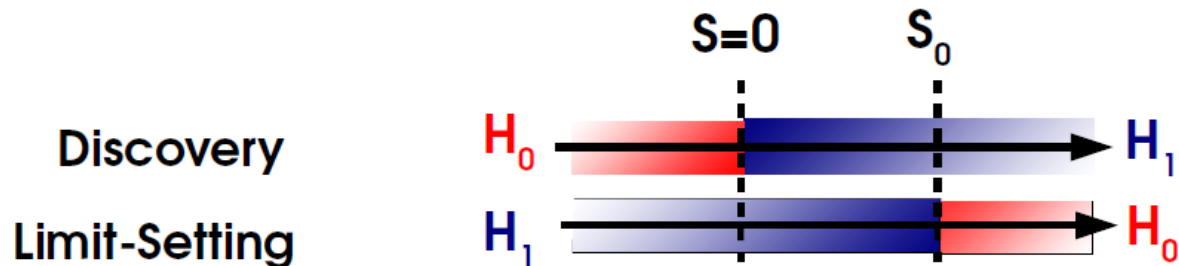


One-sided Test Statistic

For upper limits, alternate is $H_1 : S < \mu_0$:

→ If **large** signal observed ($\hat{S} \gg S_0$), does not favor H_1 over H_0

→ Only consider $\hat{S} < S_0$ for H_1 , and include $\hat{S} \geq S_0$ in H_0 .



⇒ Set $q_{s_0} = 0$ for $\hat{S} > S_0$ – only small signals ($\hat{S} < S_0$) help lower the limit.

→ Also treat separately the case $S < 0$ to avoid technical issues in $-2\log L$ fits.

Asymptotics:

$q_{s_0} \sim \text{"}\frac{1}{2}\chi^2\text{"}$ under $H_0(S=S_0)$, same as q_0 , except for special treatment of $\hat{S} < 0$.

$$\tilde{q}_{s_0} = \begin{cases} 0 & \hat{S} \geq S_0 \\ -2 \log \frac{L(S=S_0)}{L(\hat{S})} & 0 \leq \hat{S} \leq S_0 \\ -2 \log \frac{L(S=S_0)}{L(S=0)} & \hat{S} < 0 \end{cases}$$

$$p_0 = 1 - \Phi\left(\sqrt{q_{s_0}}\right)$$

Cowan, Cranmer, Gross & Vitells, [Eur.Phys.J.C71:1554,2011](#)

CL_s: Gaussian Bands

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

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

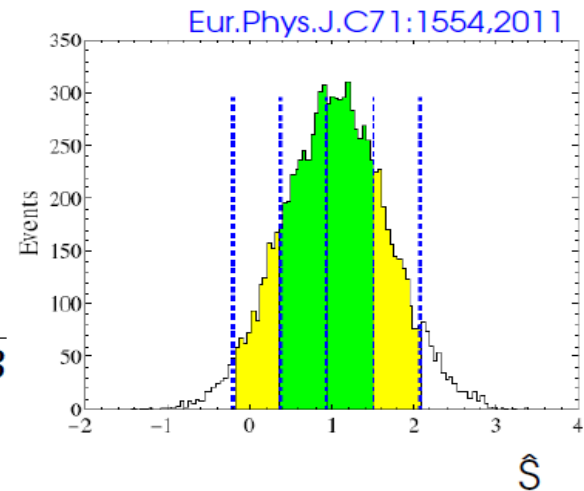
Compute expected bands for S=0:

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

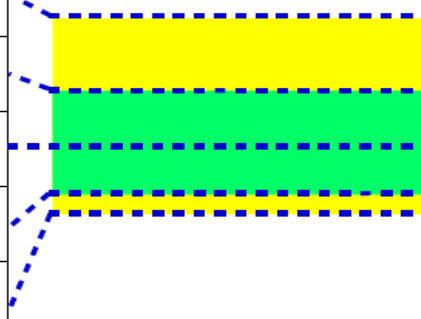
$$S_{\text{up,exp}}^0 = 1.96 \sigma_s$$

→ **$\pm n \sigma$ bands**:

$$S_{\text{up,exp}}^{\pm n} = \left(\pm n + \left[1 - \Phi^{-1} \left(0.05 \Phi(\mp n) \right) \right] \right) \sigma_s$$



n	$S_{\text{exp}}^{\pm n} / \sqrt{B}$
+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 $\sigma_{S,A}^2 = \frac{S^2}{q_s(\text{Asimov})}$
depends on S, for non-Gaussian cases, different values for each band...

Comparison with LEP/Tevatron definitions

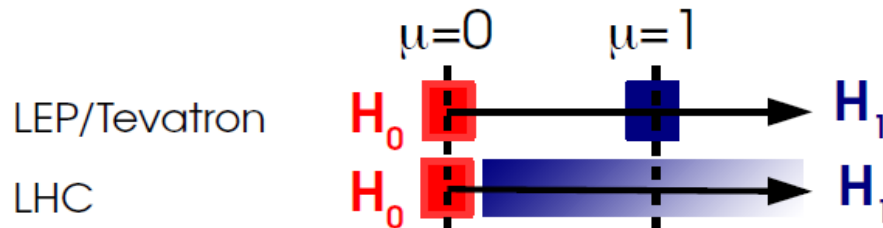
Likelihood ratios are not a new idea:

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

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

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

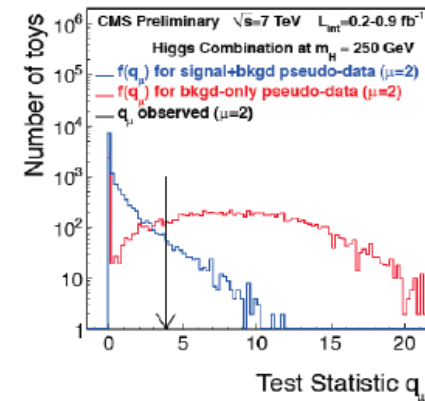
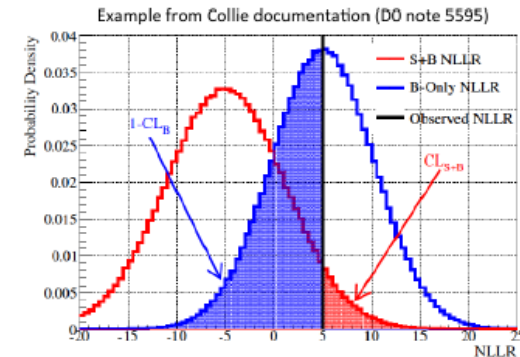
Both compare to $\mu=1$ instead of best-fit $\hat{\mu}$



→ 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



Spin/Parity measurements

Phys. Rev. D 92 (2015) 012004

