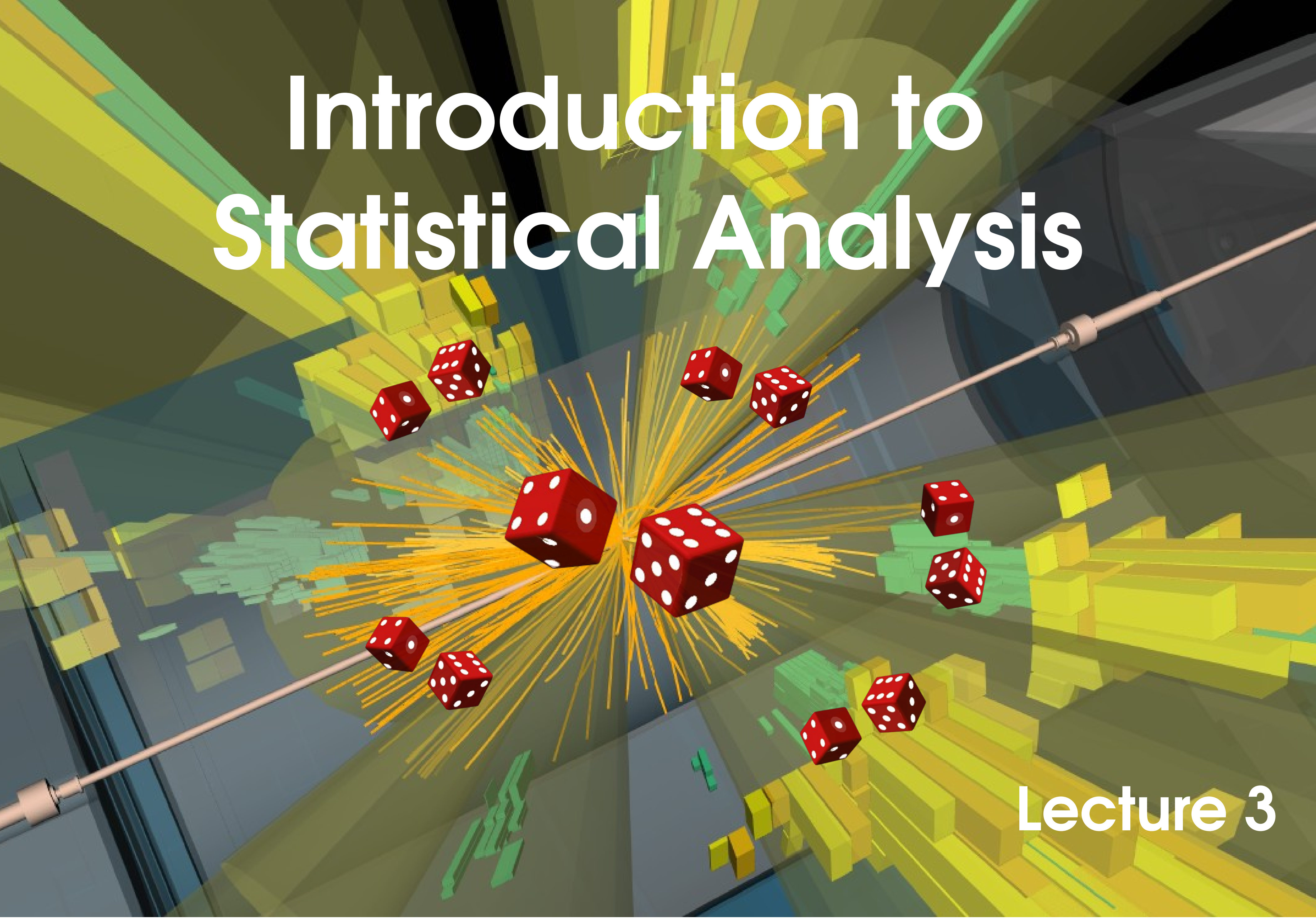
Introduction to Statistical Analysis



Lecture 3

Nicolas Berger (LAPP Annecy)

Course Outline

Previously in this course:

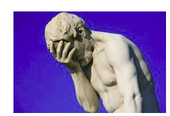
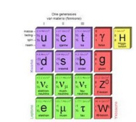
- Statistics basics
- Describing measurements
- Determining the value of a parameter

Today:

- **Computing statistical results:**
  - Discovery testing
  - Confidence intervals
  - Limits
  - Expected limits

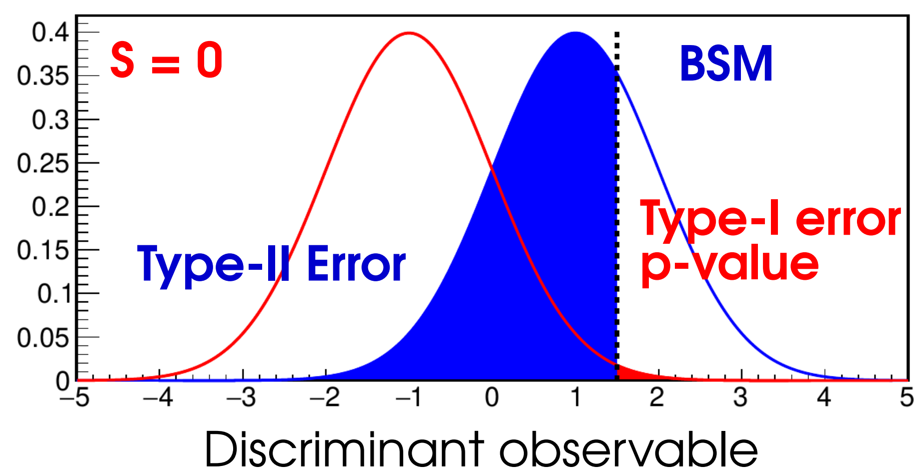
Hypothesis Testing

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 ⇔ Higher Type-II errors and vice versa: cannot have everything!

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



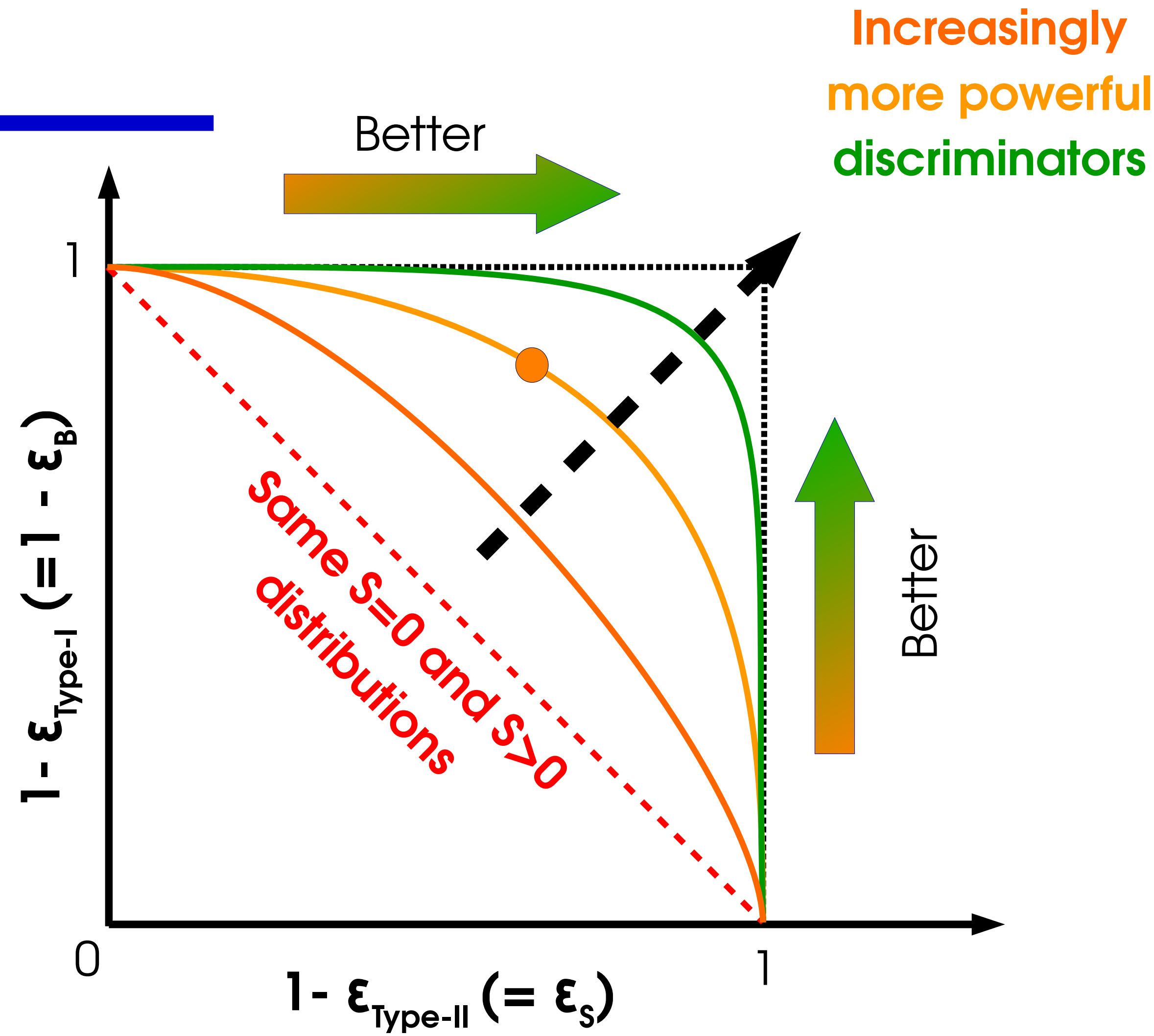
ROC Curves

“Receiver operating characteristic” (ROC) Curve:

→ Plot Type-I vs Type-II rates for different cut values

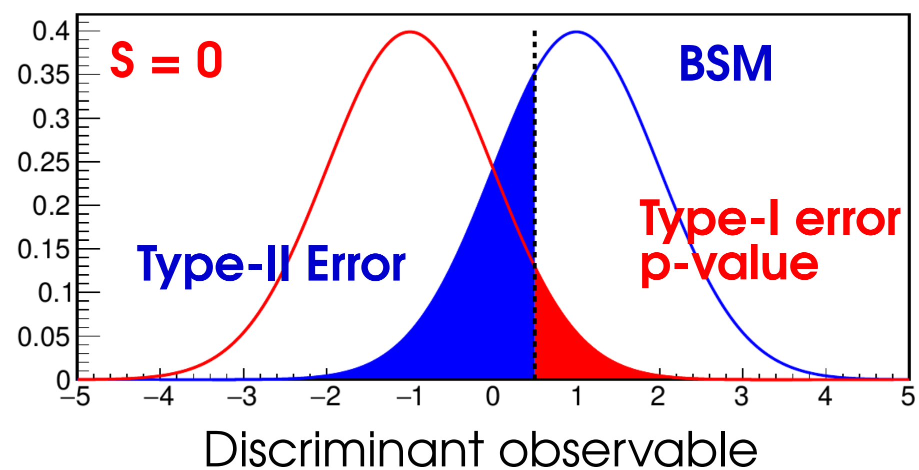
→ All curves monotonically decrease from (0,1) to (1,0)

→ Better discriminators more bent towards (1,1)



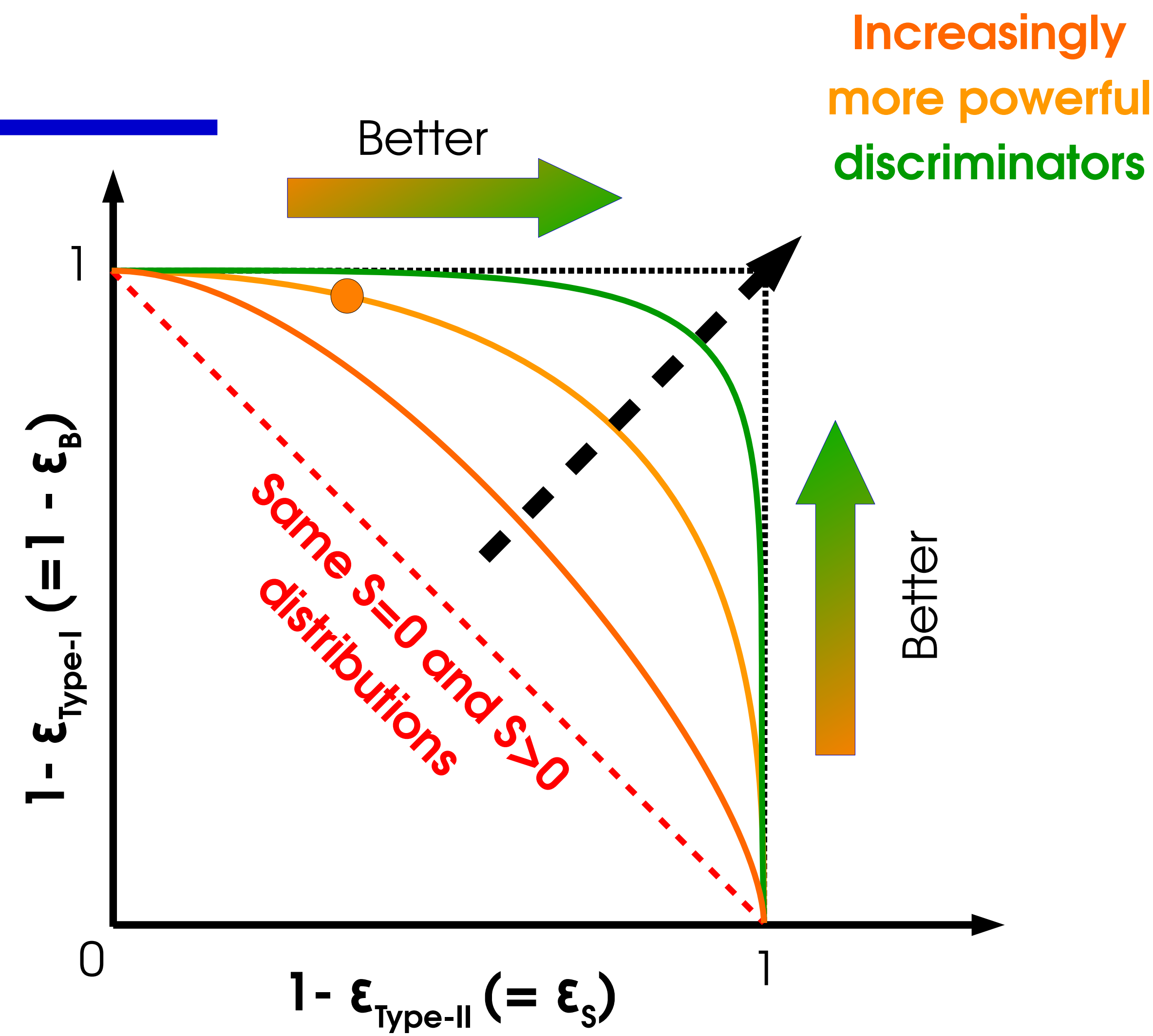
→ **Goal**: test that minimizes Type-II errors **for given level of Type-I error**.

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



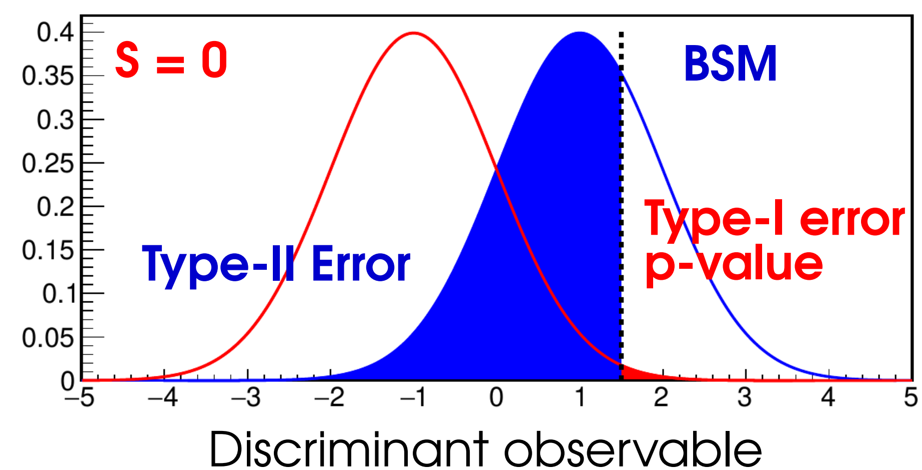
ROC Curves

"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)



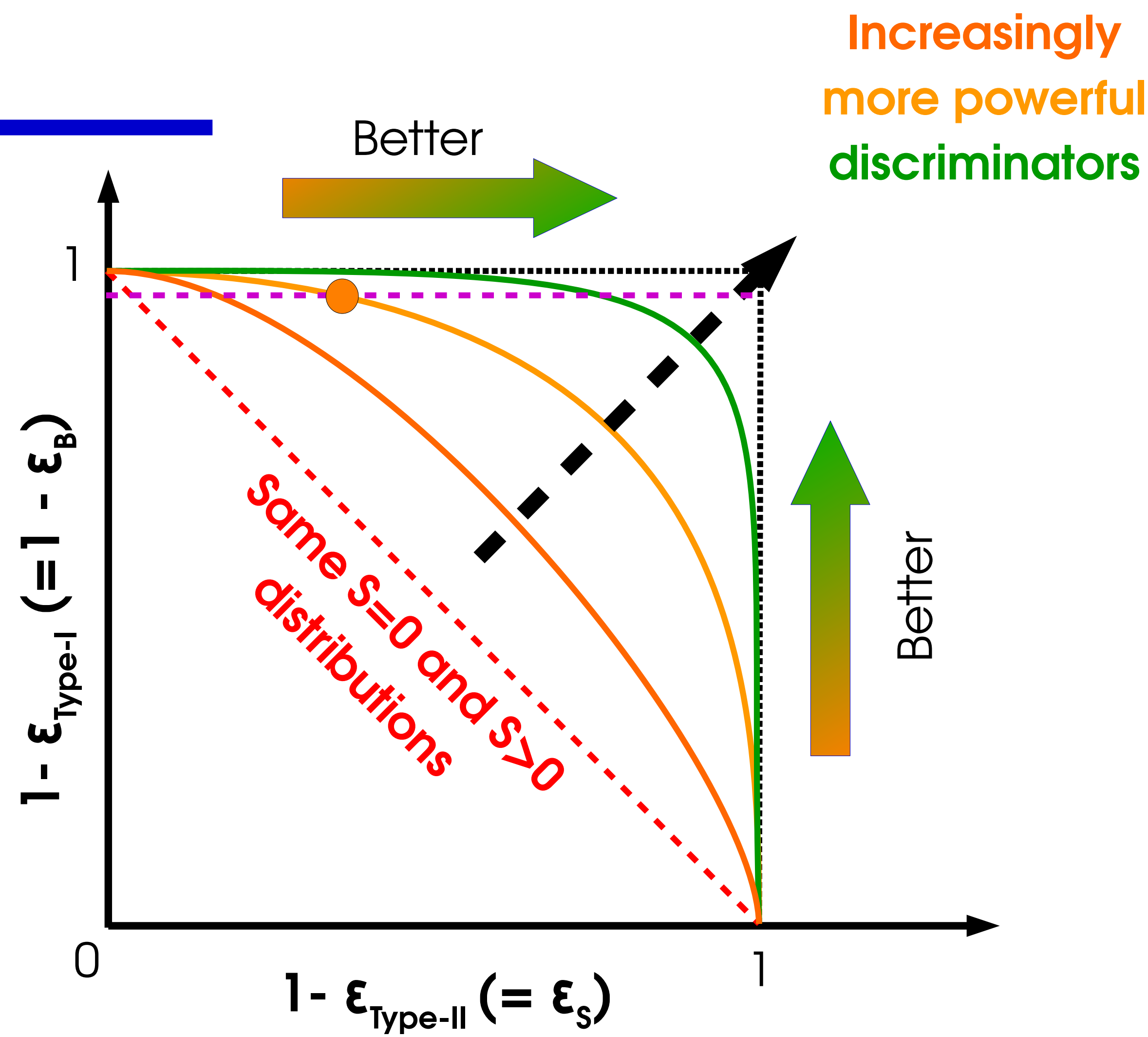
→ **Goal**: test that minimizes Type-II errors **for given level of Type-I error**.

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



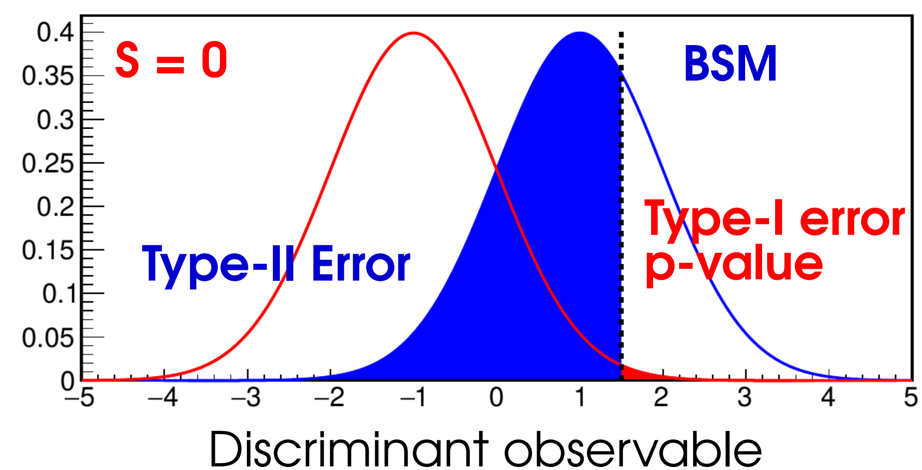
ROC Curves

"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")



Hypothesis Testing with Likelihoods

Neyman-Pearson Lemma

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

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

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

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

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

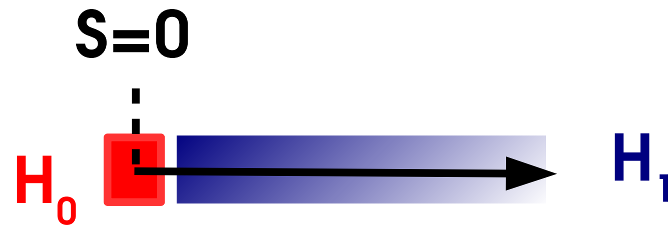
→ **Minimizes Type-II uncertainties** for given level of Type-I uncertainties
→ Always need an **alternate hypothesis** to test against.

→ **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**)

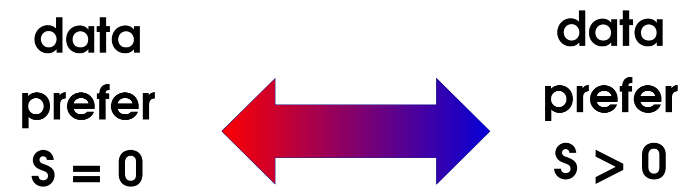
→ For H_1, any S>0 is possible, which to use ? **The one preferred by the data, Ŝ.**

⇒ Use Likelihood ratio: $$\frac{L(S=0)}{L(\hat{S})}$$

→ In fact use the **test statistic** $$q_0 = -2\log\frac{L(S=0)}{L(\hat{S})}$$

Note: for $\hat{S} < 0$, set q_0=0 to reject negative signals ("*one-sided* test statistic")

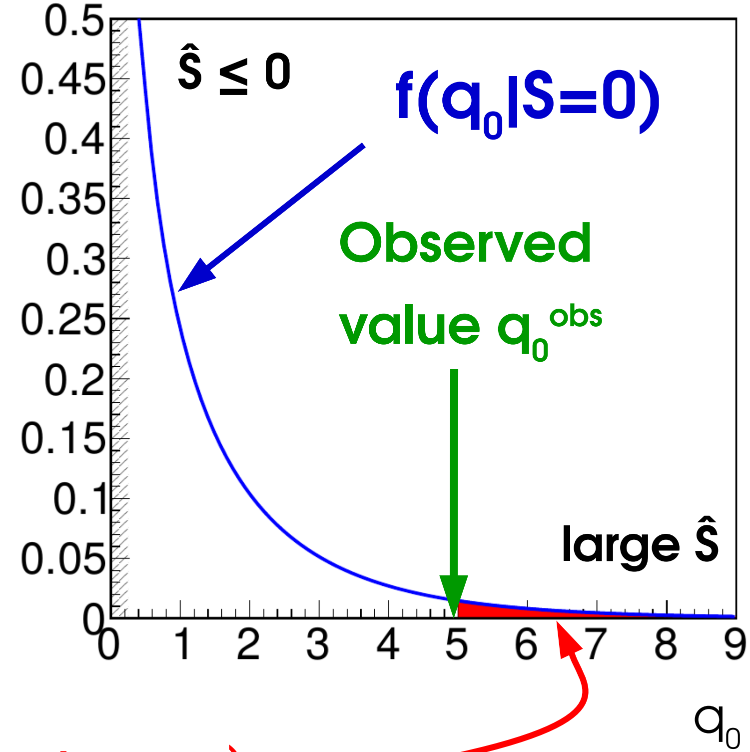
Discovery p-value



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 rejecting H_0)

→ Type-I rate, a.k.a. the ***p-value :*** $$p_0 = \int_{q_0^{obs}}^{\infty} f(q_0|S=0)\, dq_0$$

= *Fraction of outcomes that are*

at least as extreme *(signal-like)* ***as data****, when* ***H_0 is true*** *(no signal).*

Asymptotic distribution of q_0

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

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

Wilk's Theorem (*) : for $S = 0$

q_0 is distributed as χ^2 (n_{par})

⇒ n_{par} = 1 : $\sqrt{q_0}$ is distributed as a Gaussian

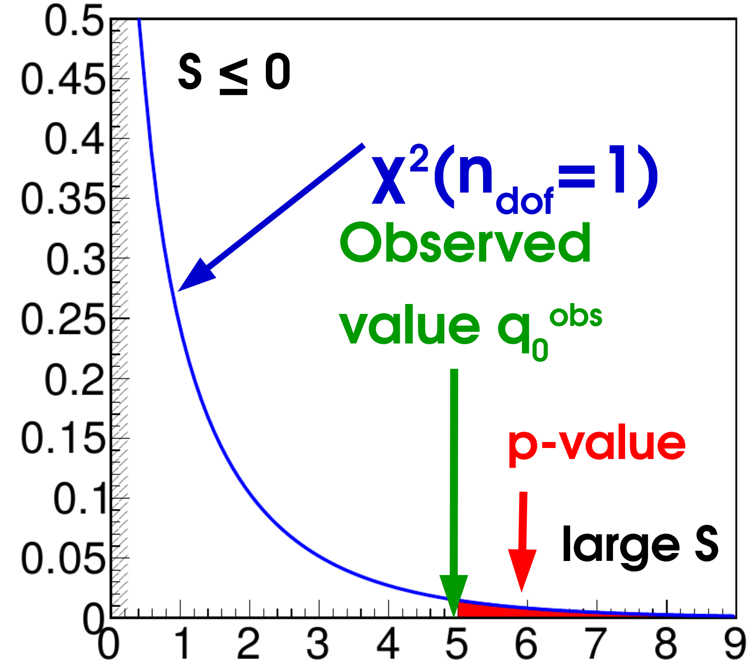
⇒ Can compute p-values from Gaussian quantiles

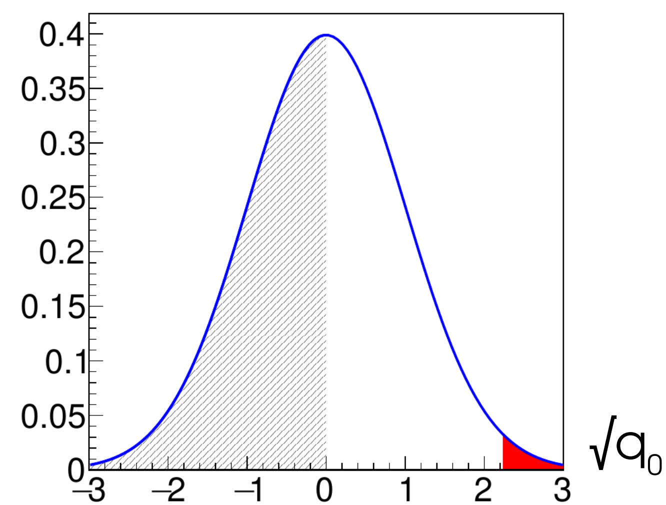
$$p_0 = 1 - \Phi\left(\sqrt{q_0}\right)$$

⇒ Even more simply, the significance is:

$$Z = \sqrt{q_0}$$

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





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

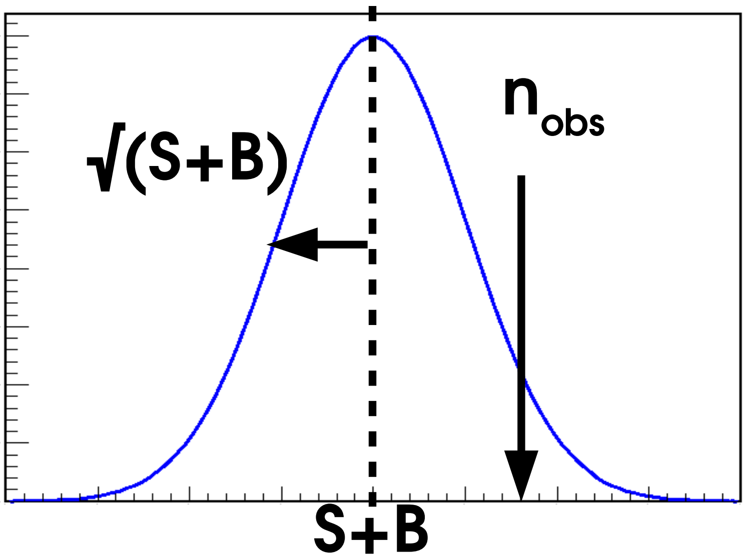
$$L(S) = \exp\left[-\frac{1}{2}\left(\frac{S-\hat{S}}{\sigma}\right)^2\right] \Rightarrow q_0 = \left(\frac{\hat{S}}{\sigma}\right)^2 \Rightarrow \sqrt{q_0} = \frac{\hat{S}}{\sigma} \sim G(0,1) \Rightarrow q_0 \sim \chi^2(n_{dof}=1)$$

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

→ Find the best-fit value (MLE) $\hat{S}$ for the signal

(can use λ = -2 log L instead of L for simplicity)

→ Find the expression of q_0 for $\hat{S} > 0$.

→ Find the expression for the significance

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

Homework 2: Poisson Counting

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

$$L(S;n) = e^{-(S+B)}(S+B)^n$$

(Can remove the n! constant since we're only dealing with L ratios)

→ As before, compute $\hat{S}$, and q_0

→ Compute $Z = \sqrt{q_0}$, assuming asymptotic behavior

Solution:

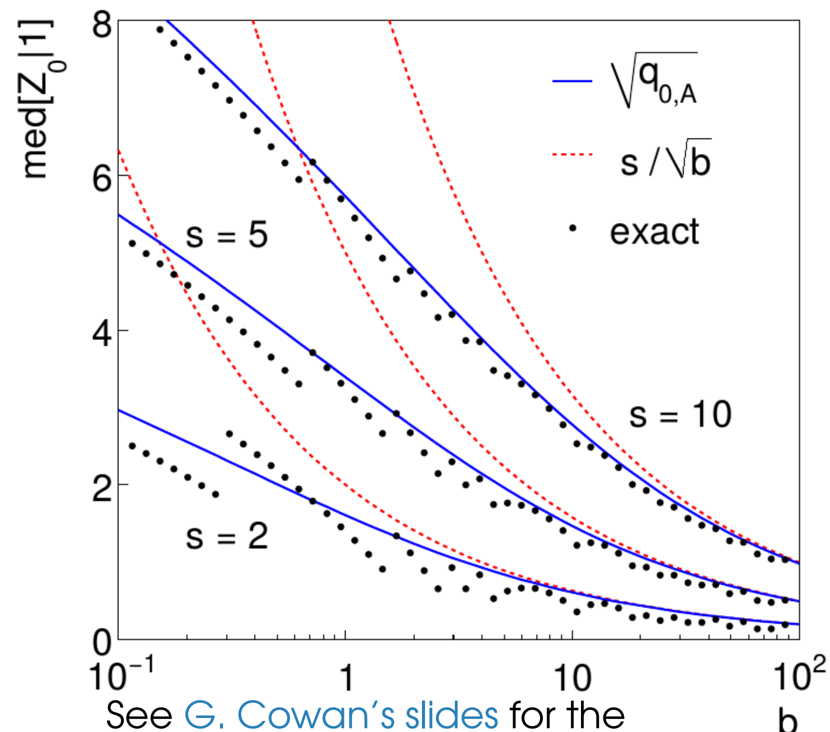
$$Z = \sqrt{2\left[(\hat{S}+B)\log\left(1+\frac{\hat{S}}{B}\right) - \hat{S}\right]}$$

Exact result can be obtained using

pseudo-experiments → close to $\sqrt{q_0}$ result

Asymptotic formulas justified by Gaussian regime, but remain valid even for small values of S+B (down to 5 events!)

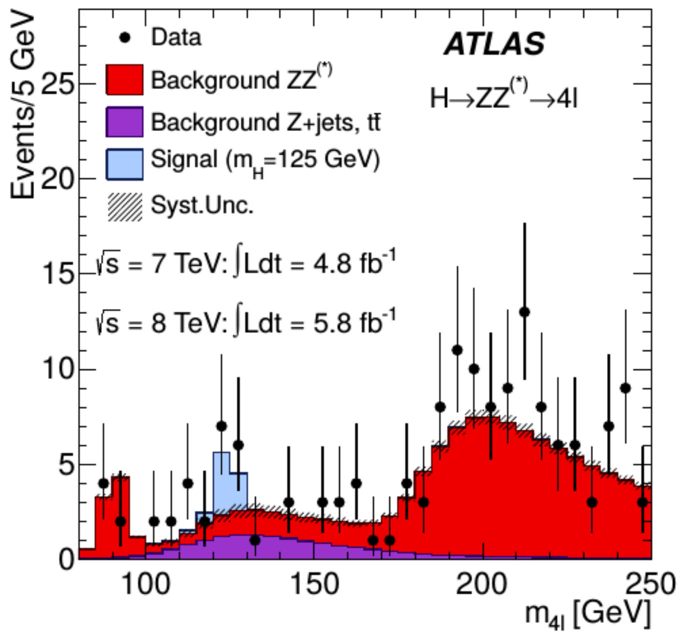
Eur.Phys.J.C71:1554,2011

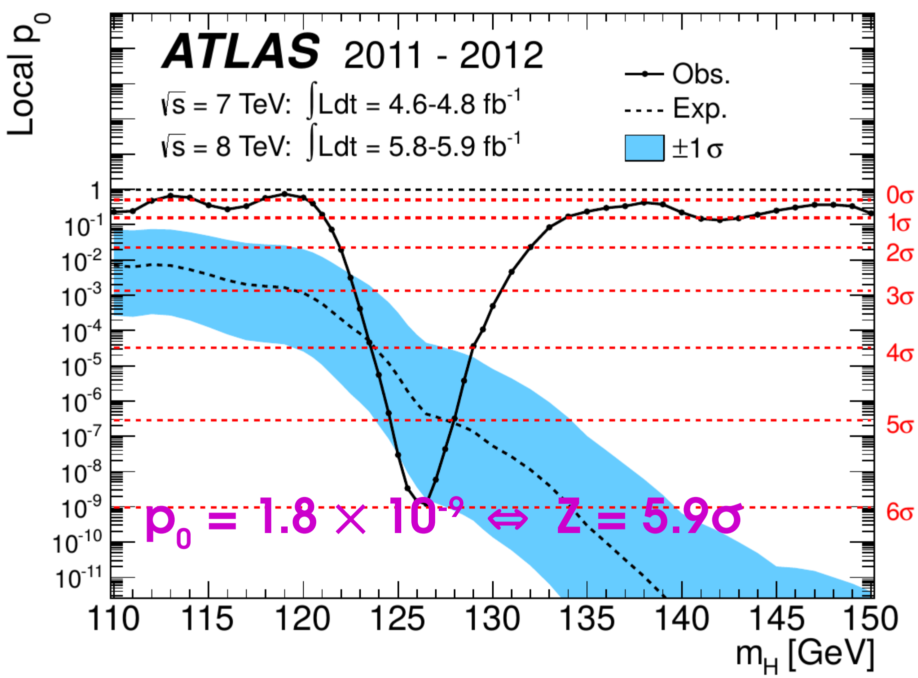


See G. Cowan's slides for the case with B uncertainty

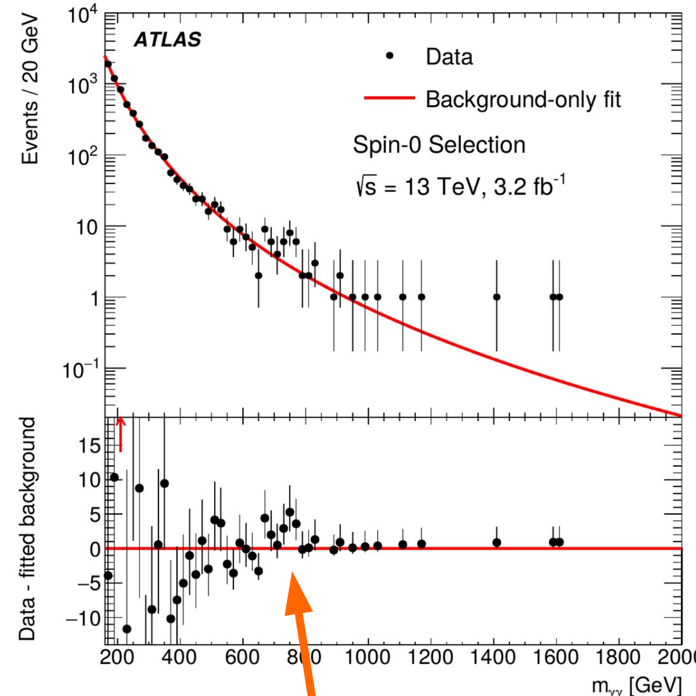
Some Examples

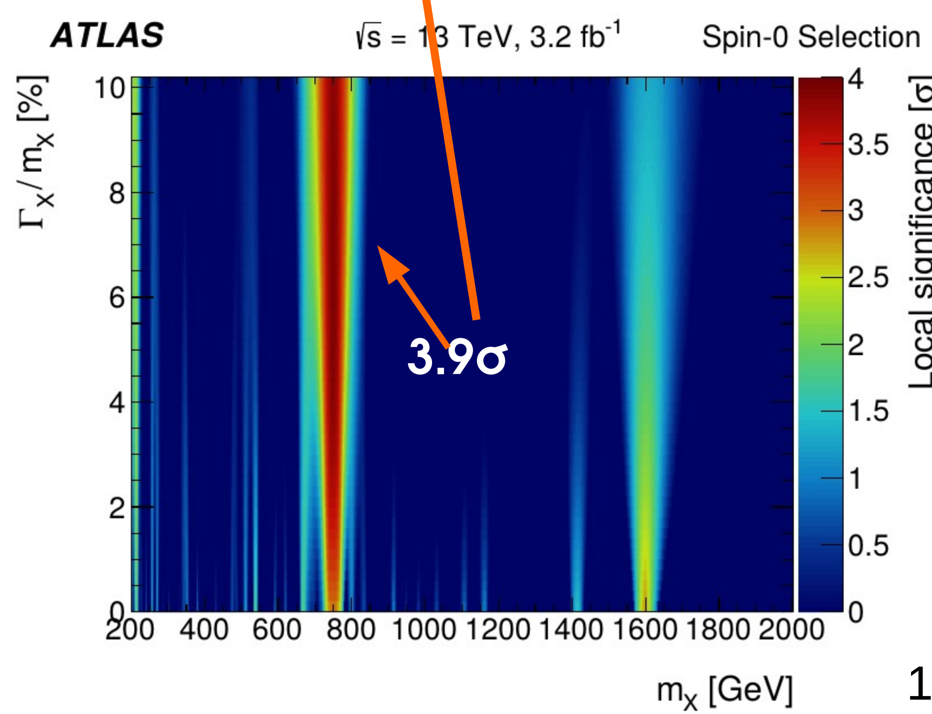
Higgs Discovery: Phys. Lett. B 716 (2012) 1-29





High-mass X→γγ Search: JHEP 09 (2016) 1





Discovery Thresholds

Evidence : **3σ ⇔ p_0 = 0.3% ⇔ 1 chance in 300**

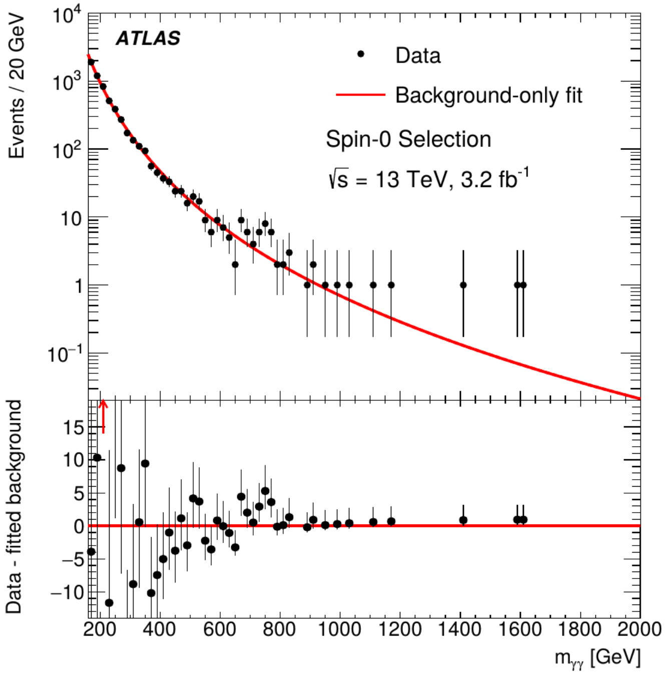
Discovery: **5σ ⇔ p_0 = 3 10-7 ⇔ 1 chance in 3.5M**

Why so high thresholds ? (from Louis Lyons):

- **Look-elsewhere effect** : searches typically cover multiple independent regions ⇒ Higher chance to have a fluctuation "somewhere"

  N_{trials} ~ 1000 : local 5σ ⇔ O(10^{-4}) more reasonable

- **Mismodeled systematics**: factor 2 error in syst-dominated analysis ⇒ factor 2 error on Z…



- **History**: 3σ and 4σ excesses do occur regularly, for the reasons above

Extraordinary claims require extraordinary evidence!

Takeaways

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

To estimate a parameter, use the value $\hat{\mu}$ that maximizes L(μ) → 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 >~ 5), $Z = \sqrt{q_0}$

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

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

Outline

Computing statistical results

Discovery Testing

Confidence intervals

Upper limits on signal yields

Expected limits

Confidence Intervals

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

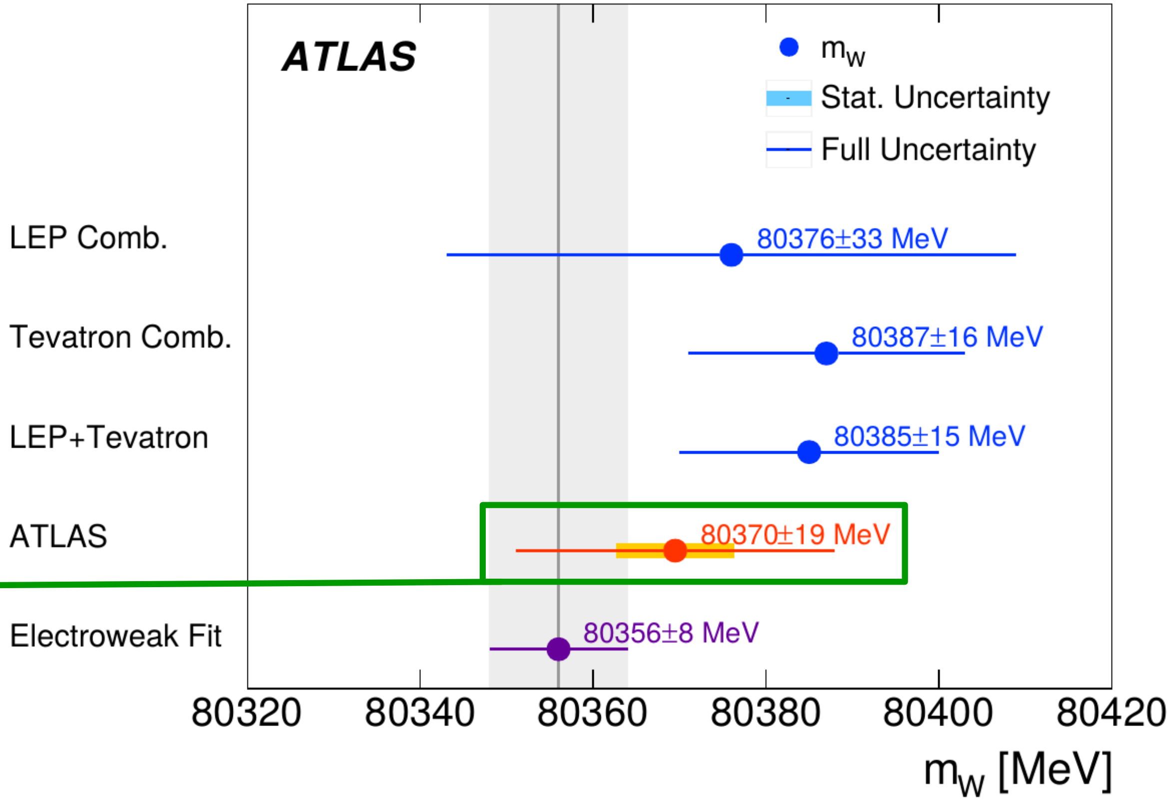
Maximum Likelihood Estimator (MLE) $\hat{\mu}$:

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

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 orange interval**



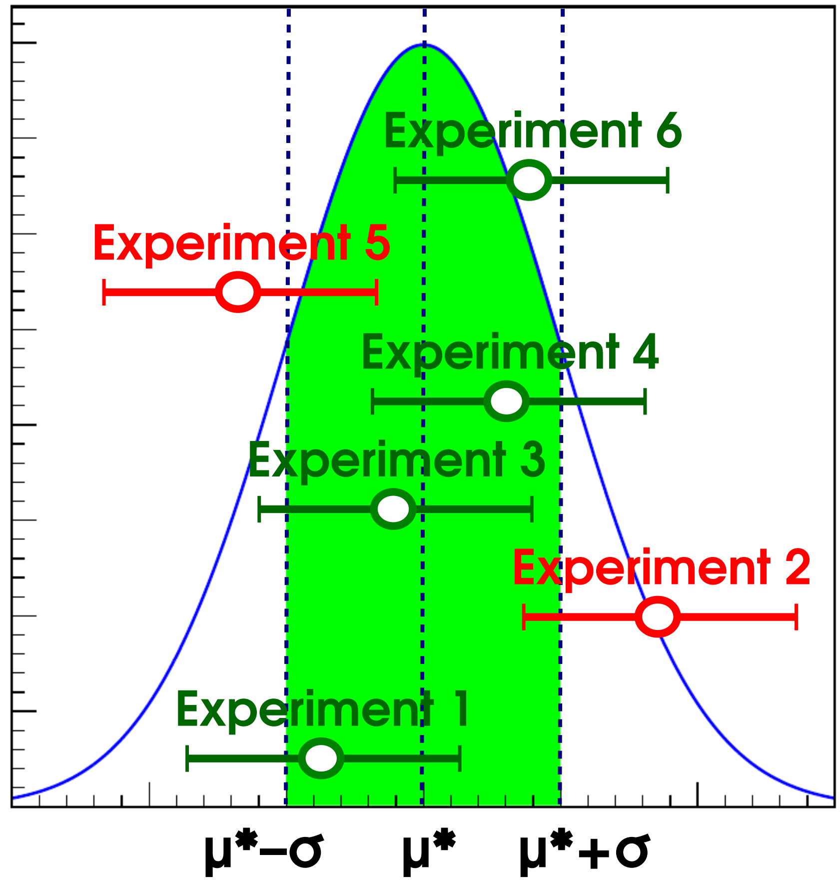
Gaussian Intervals

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

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

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

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



But we can invert the relation:

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

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

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

→ *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

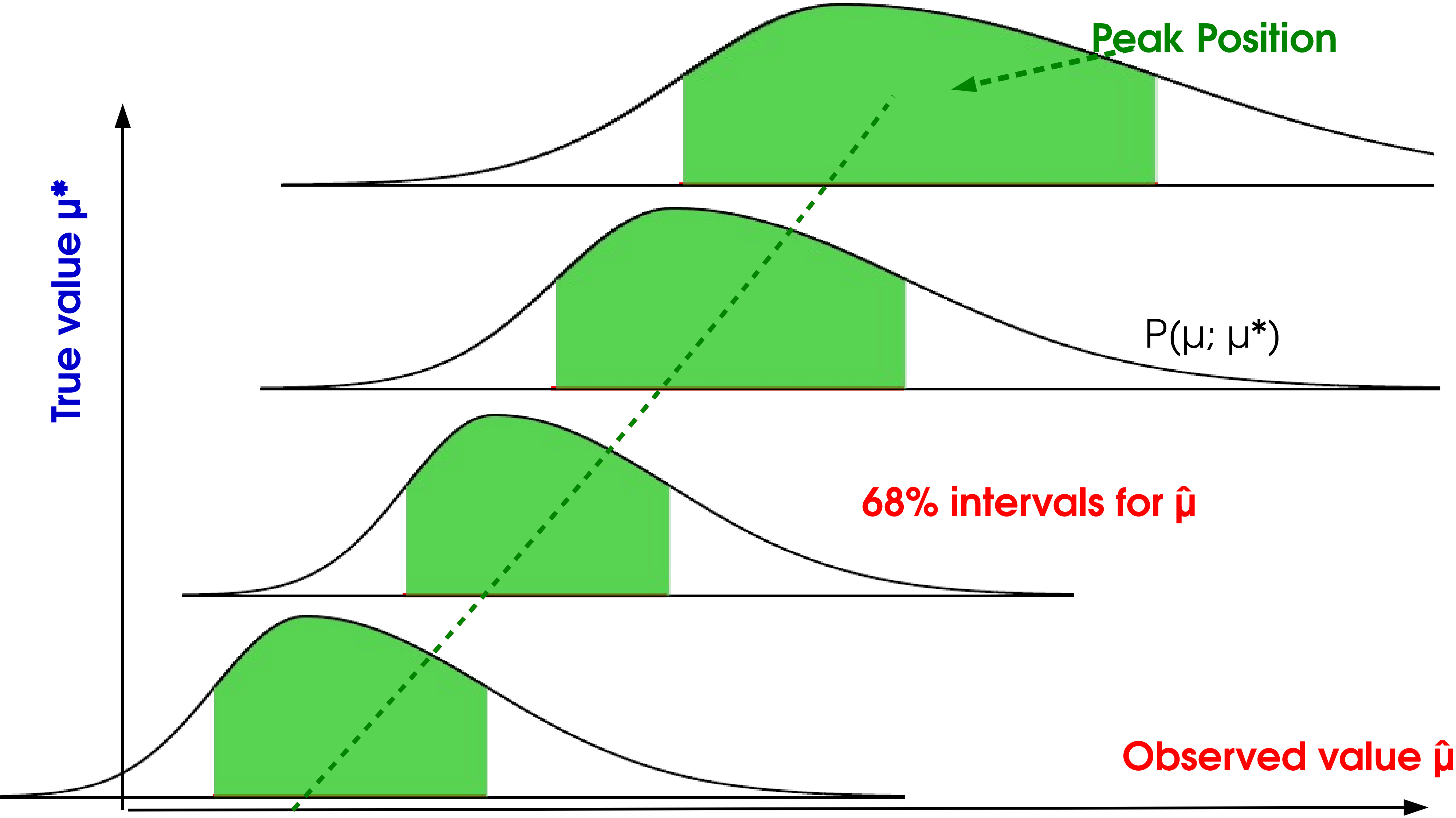
Works in the same way for other interval sizes: **$[\hat{\mu} - Z\sigma, \hat{\mu} + Z\sigma]$** with

Z	1	1.96	2
CL	68.3%	95%	95.5%

Neyman Construction

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

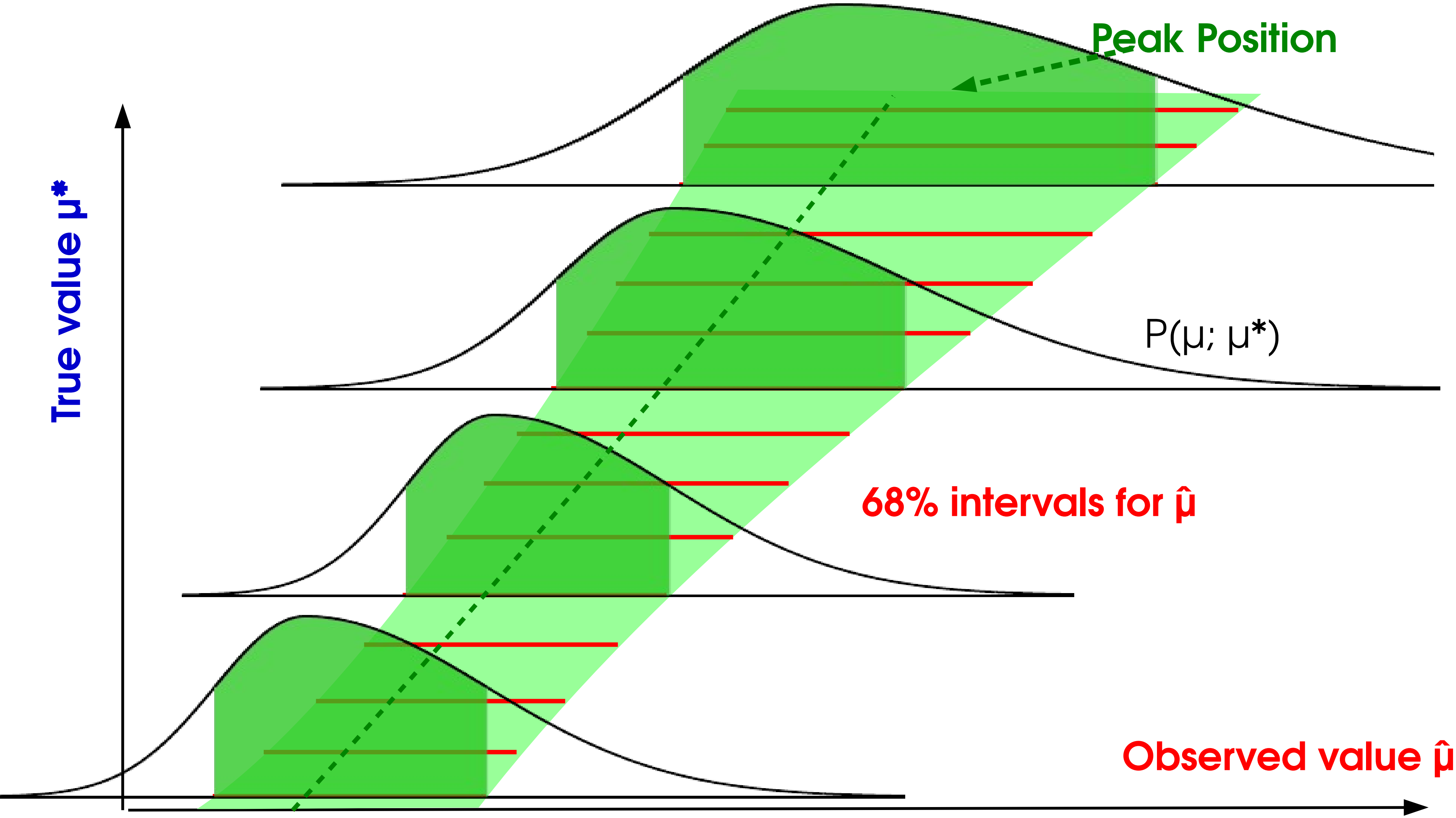
⇒ *Confidence belt*



Neyman Construction

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

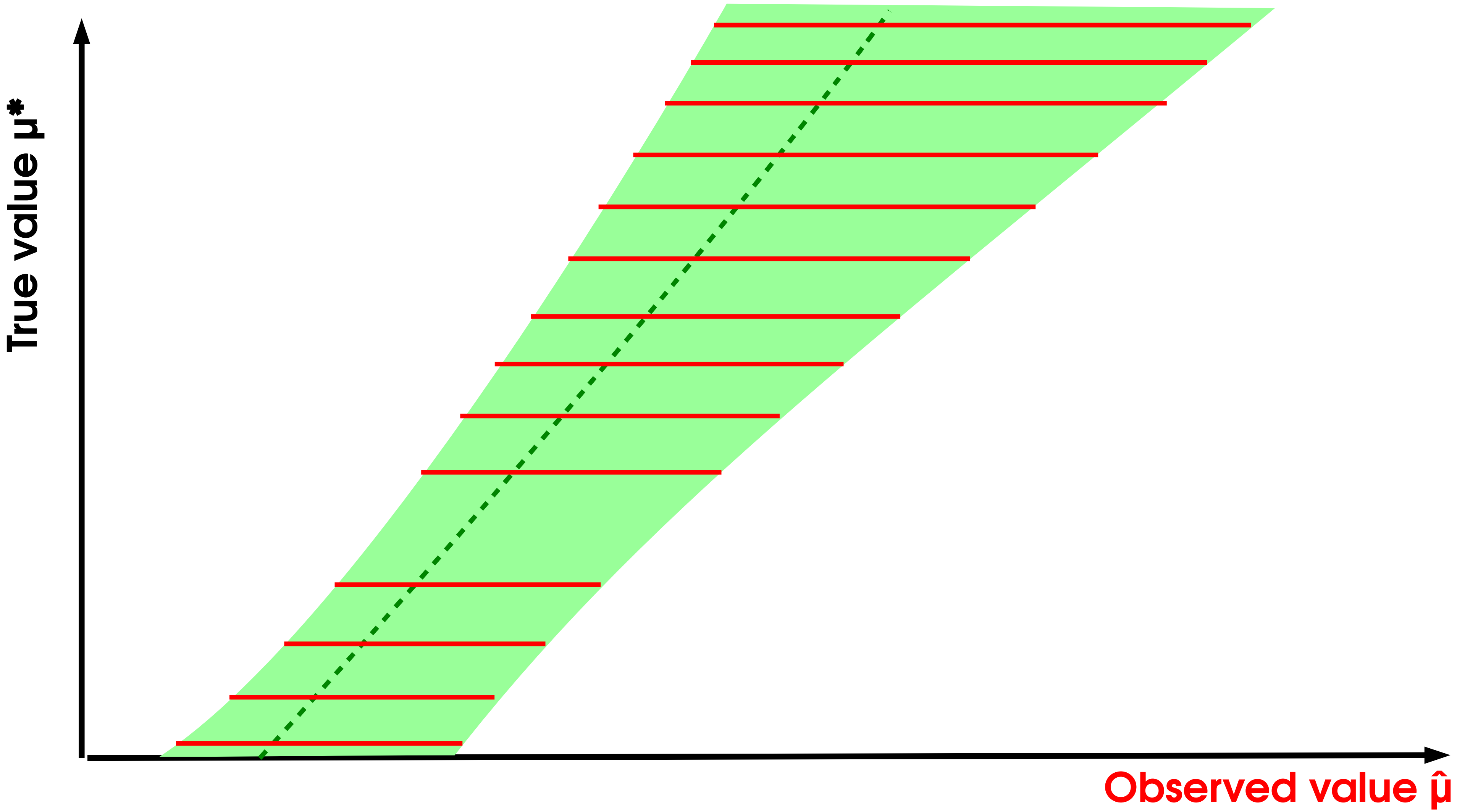
⇒ ***Confidence belt***



Inversion using the Confidence Belt

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

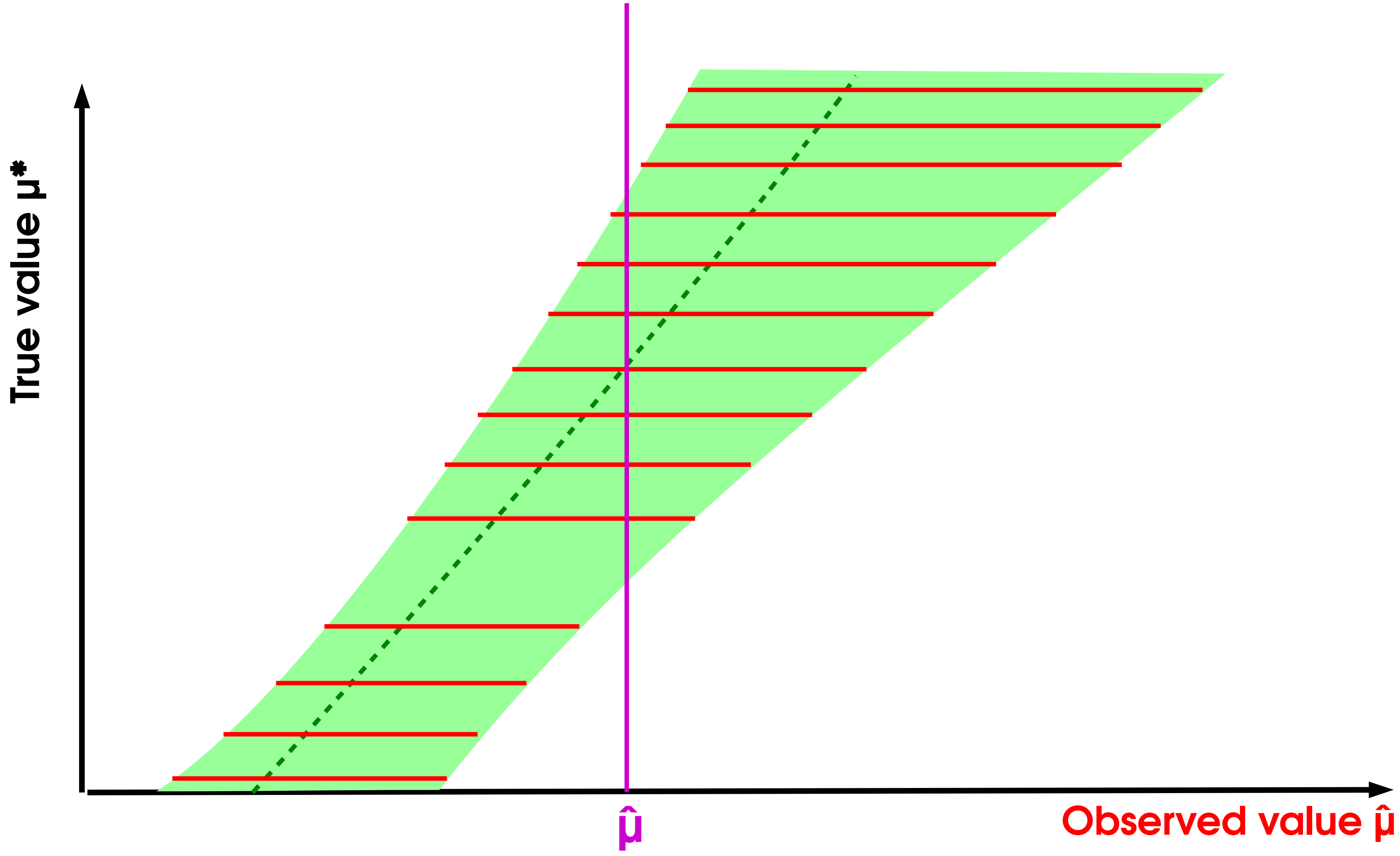
→ Same as before for Gaussian, works also when $P(\mu^{obs}|\mu)$ varies with μ.



Inversion using the Confidence Belt

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

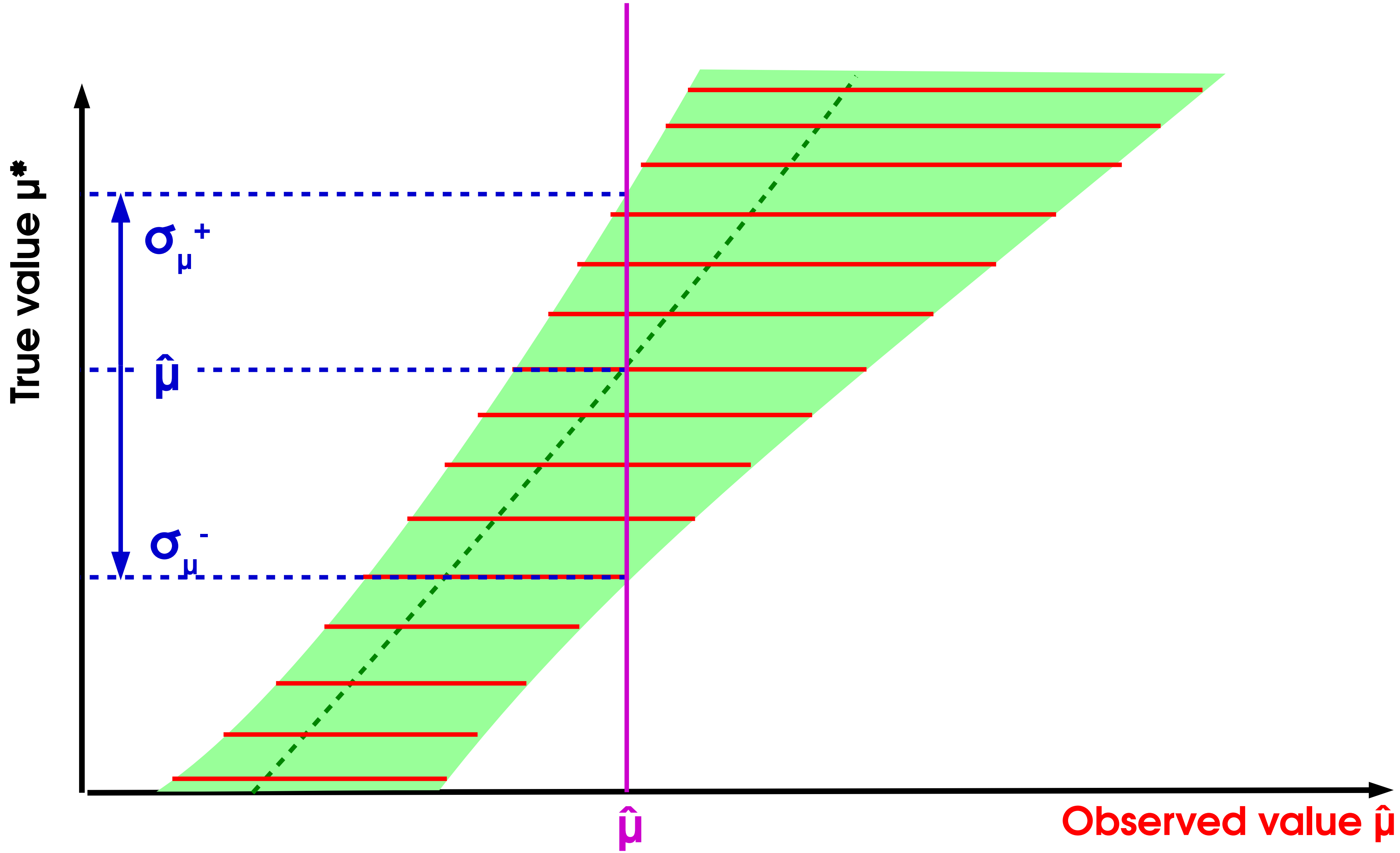
→ Same as before for Gaussian, works also when $P(\mu^{obs}|\mu)$ varies with μ.



Inversion using the Confidence Belt

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

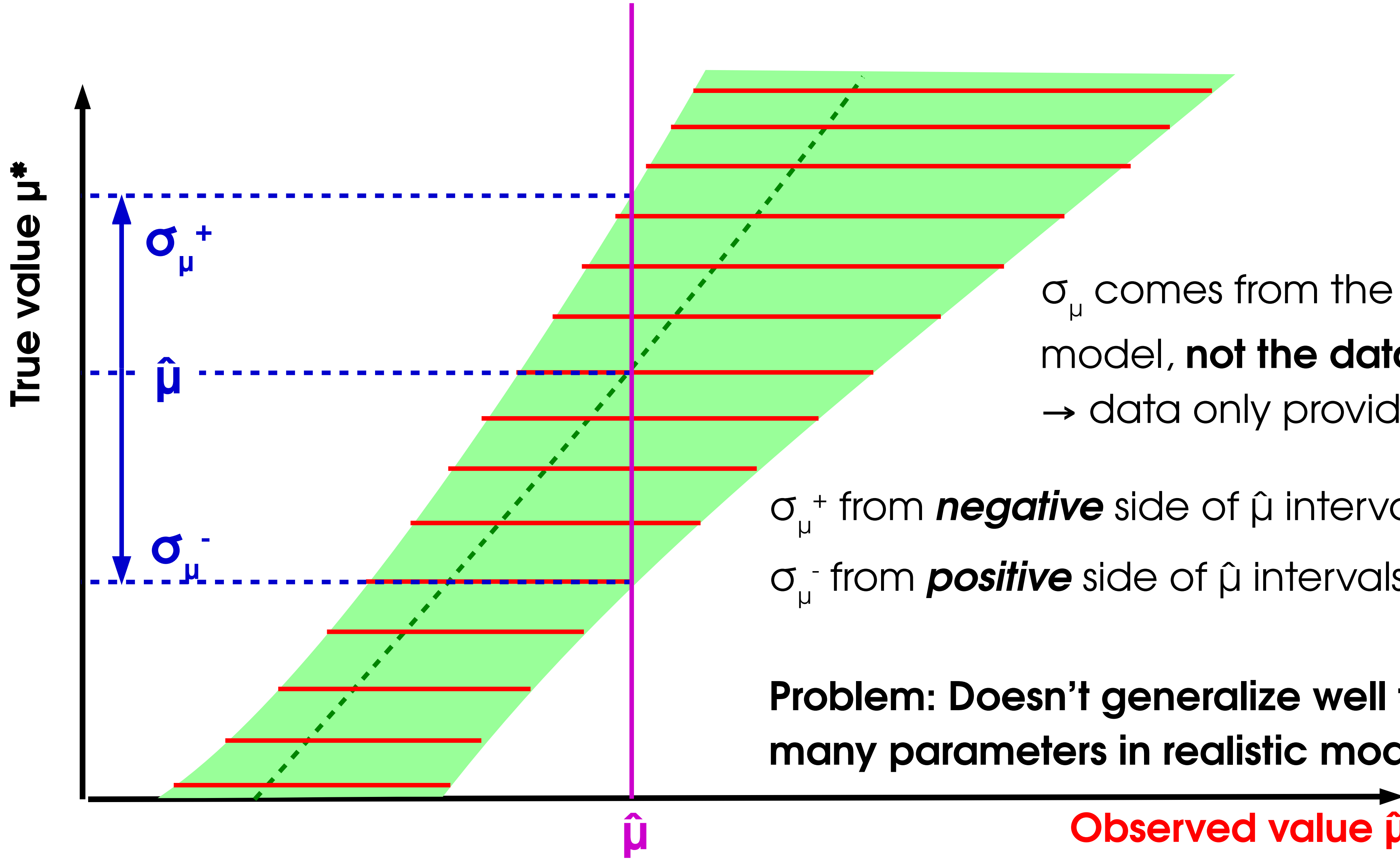
→ Same as before for Gaussian, works also when $P(\mu^{obs}|\mu)$ varies with μ.



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.



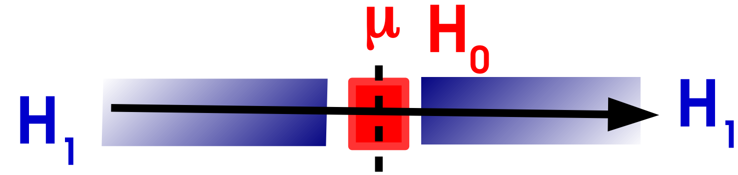
σ_μ comes from the model, **not the data**
→ data only provides $\hat{\mu}$.

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

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

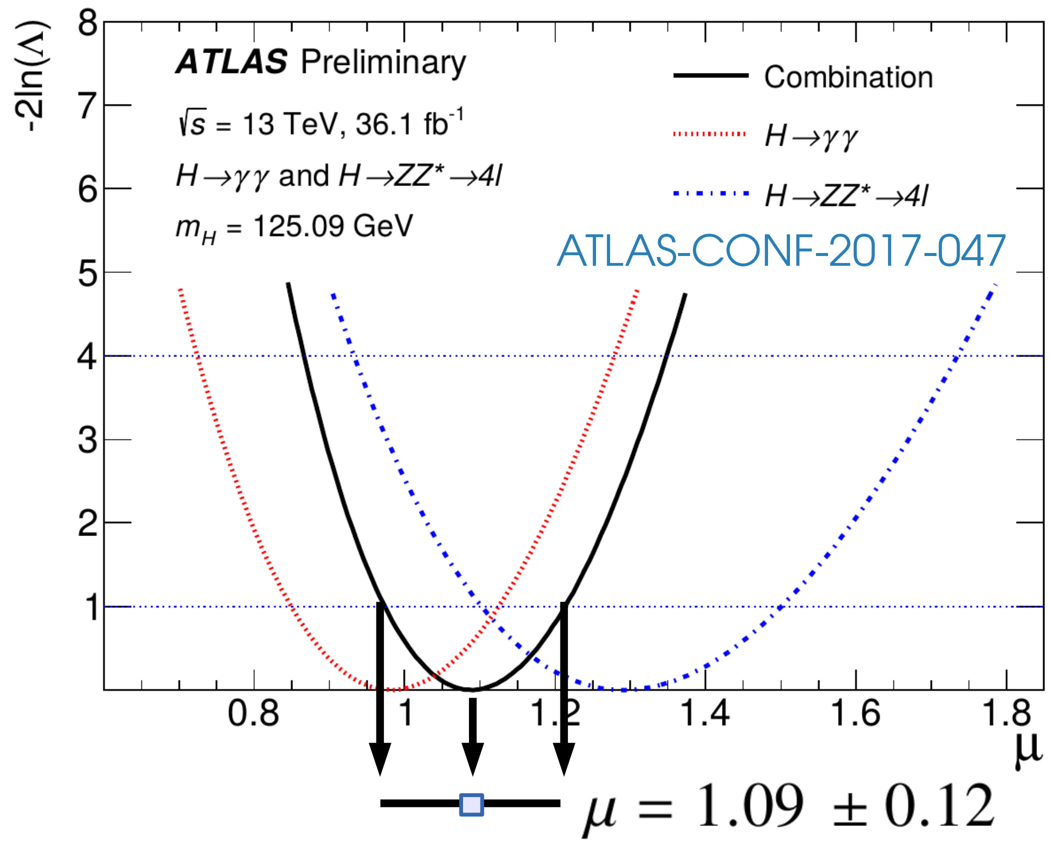
Problem: Doesn't generalize well to many parameters in realistic models

Likelihood Intervals



Confidence intervals from L:

- Test $H(\mu_0)$ against alternative using $$t_{\mu_0} = -2\log\frac{L(\mu=\mu_0)}{L(\hat{\mu})}$$
- Two-sided test since true value can be higher or lower than observed



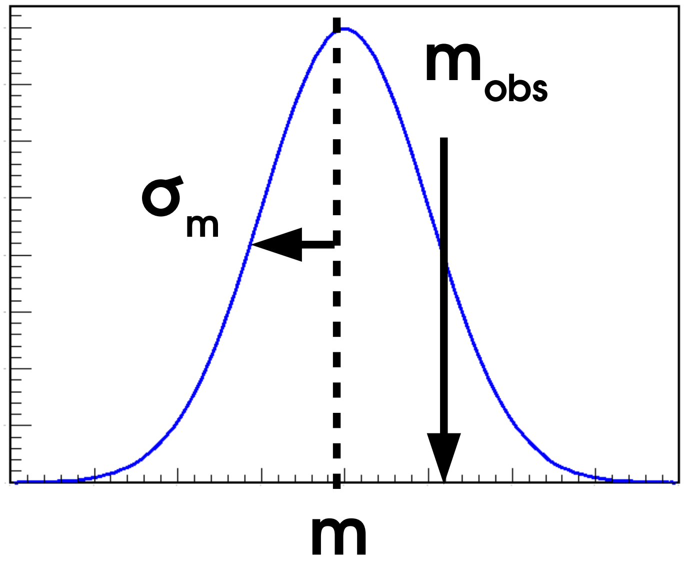
Gaussian L:

- $t_{\mu_0} = \left(\frac{\hat{\mu}-\mu_0}{\sigma_\mu}\right)^2$: 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

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 1-σ (Z=1, ~68% CL) interval on m

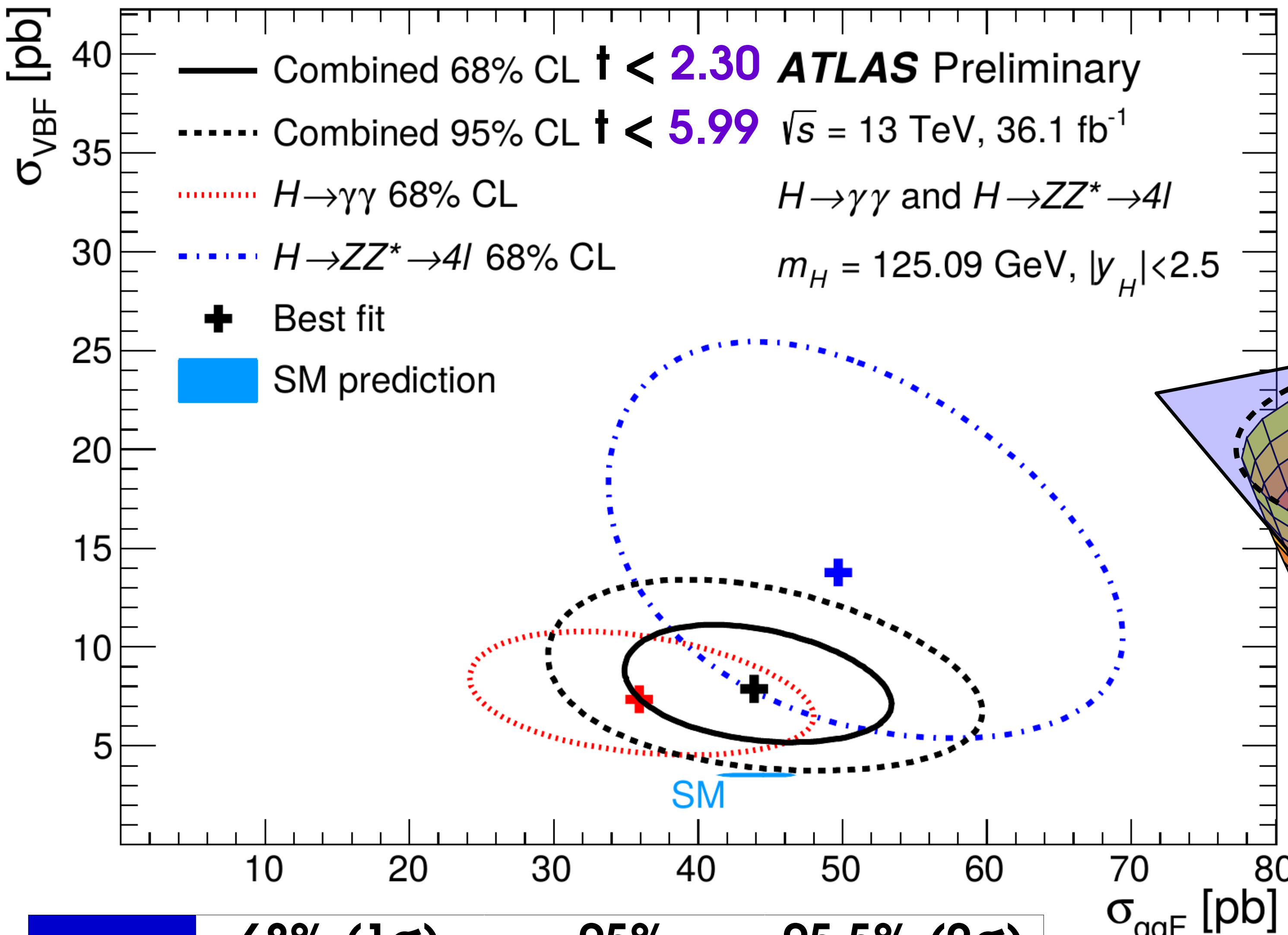
Solution: $m = m_{obs} \pm \sigma_m$

→ Not really a surprise – the method works as expected on this simple case

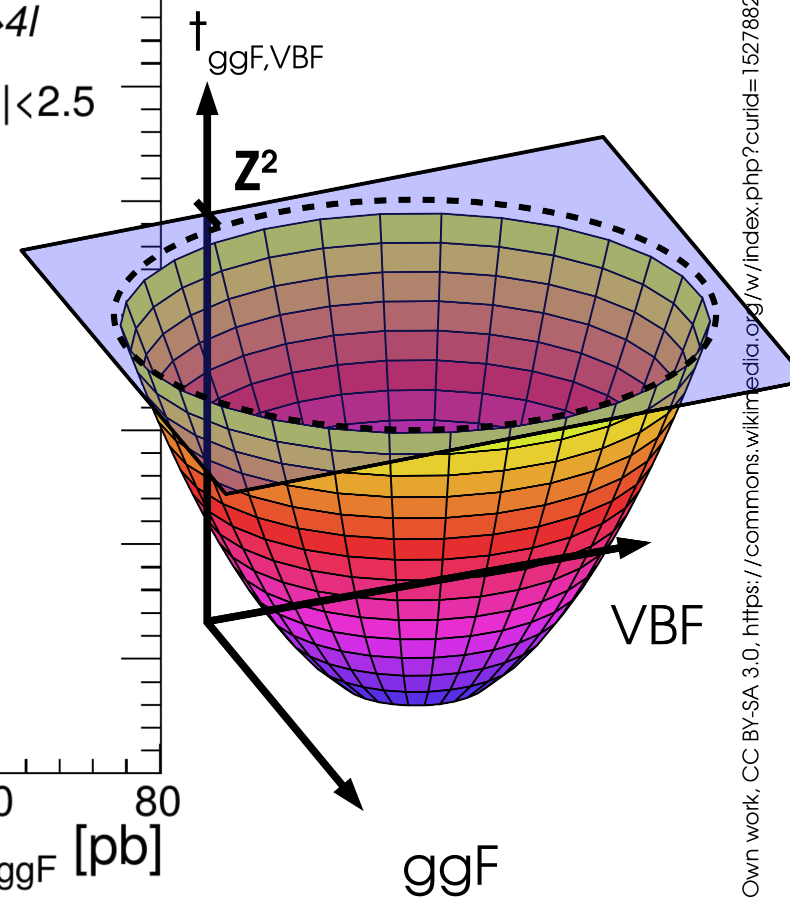
→ General method can be applied in the same way to more complex cases

2D Example: Higgs σ_{VBF} vs. σ_{ggF}

ATLAS-CONF-2017-047



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



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

Gaussian case: elliptic paraboloid surface

Outline

Computing statistical results

- Discovery Testing
- Confidence intervals
- Upper limits on signal yields
- Expected Limits

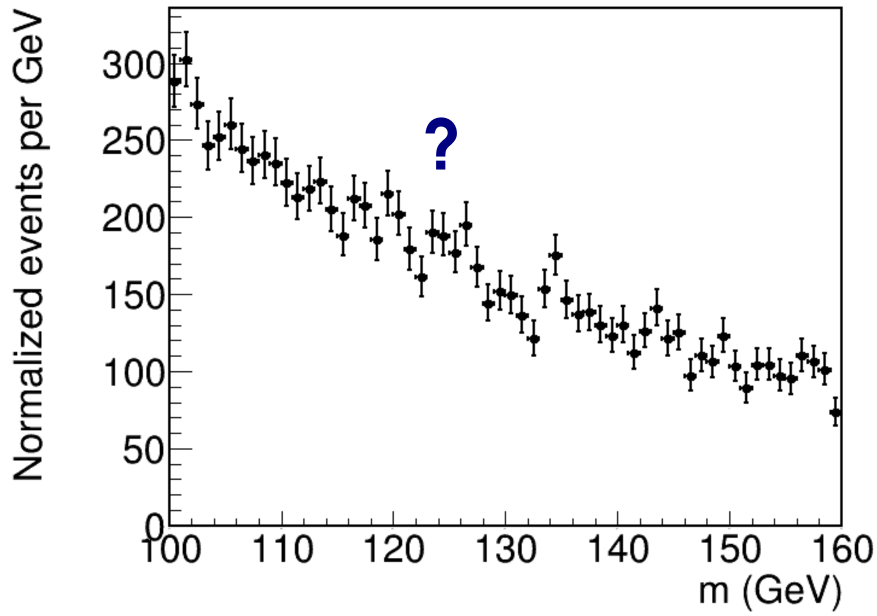
Hypothesis tests for Limits

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"



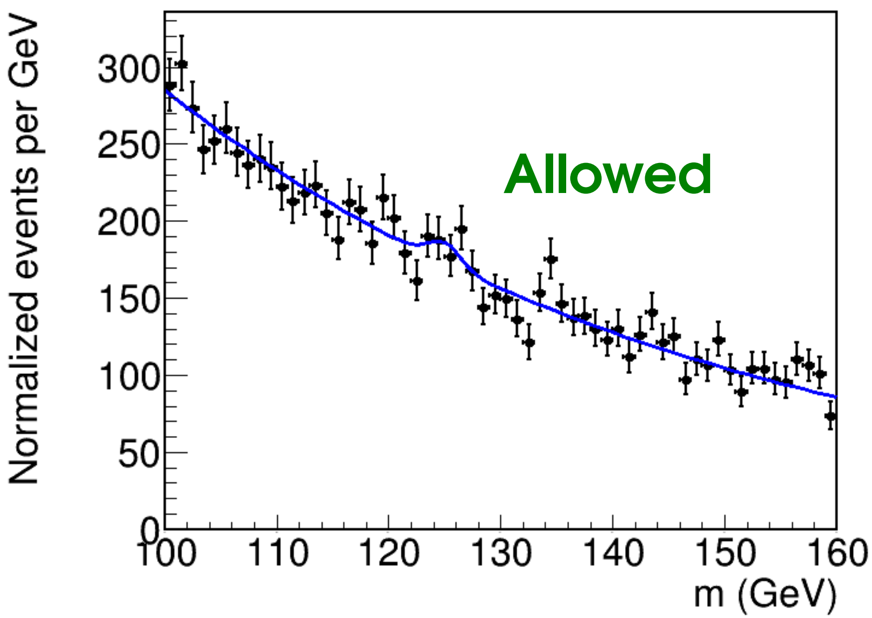
Hypothesis tests for Limits

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"



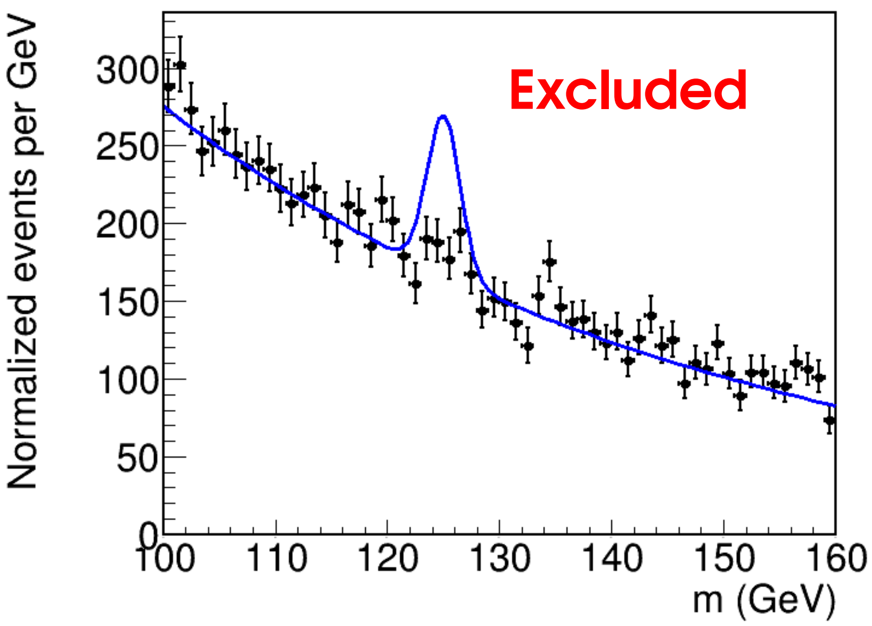
Hypothesis tests for Limits

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"



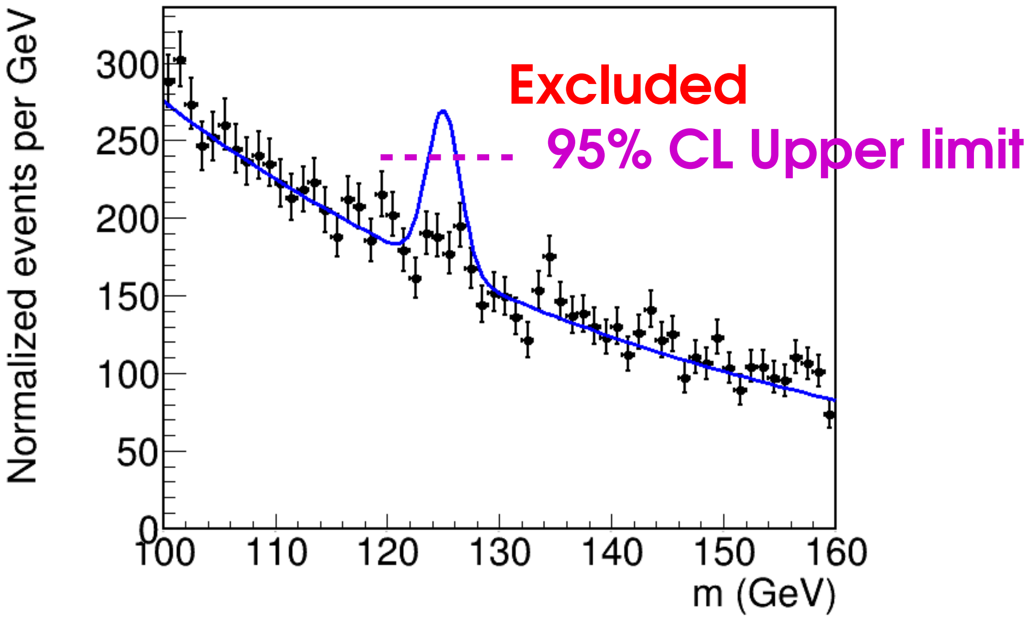
Hypothesis tests for Limits

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**

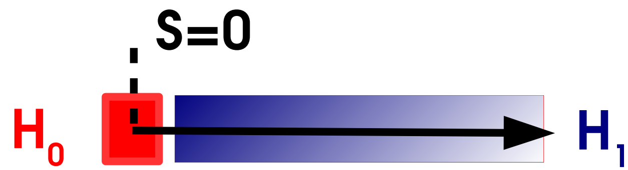
→ Typically report **95% CL** upper limit (p-value = 5%) : "S < S_0 @ 95% CL"



Test 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
← Likelihood of H_1

$(\hat{S} > 0)$

Limit-setting

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

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

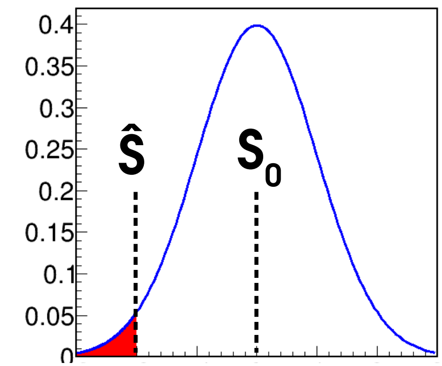
Compare
← Likelihood of H_0
← Likelihood of H_1

$(\hat{S} < S_0)$

Same as q_0 :

→ large values ⇒ good rejection of H_0.

Asymptotic case: p-value $\quad p_{S_0} = 1 - \Phi\left(\sqrt{q_{S_0}}\right)$



Inversion : Getting the limit for a given CL

Procedure:

→ Compute q_{S0} for some S_0, get the **exclusion p-value p_{S0}**.

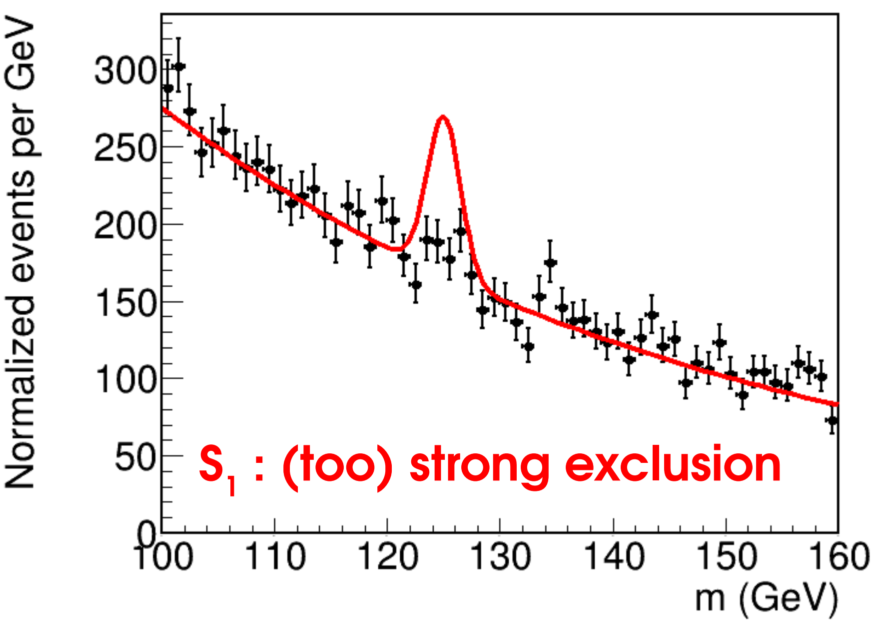
→ Adjust S_0 until 95% CL exclusion (p_{S0} = 5%) is reached

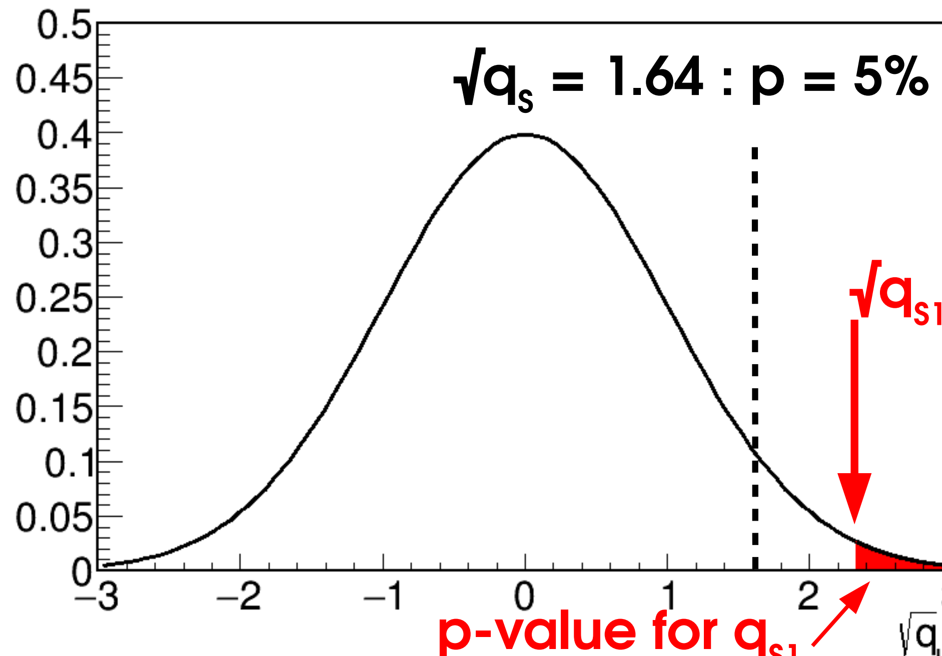
Asymptotic case: need $\sqrt{q_{S0}}$ = 1.64

Asymptotics

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

CL	Region
90%	$\sqrt{q_S} > 1.28$
95%	$\sqrt{q_S} > 1.64$
99%	$\sqrt{q_S} > 2.33$





Inversion : Getting the limit for a given CL

Procedure:

→ Compute q_{S0} for some S_0, get the **exclusion p-value p_{S0}**.

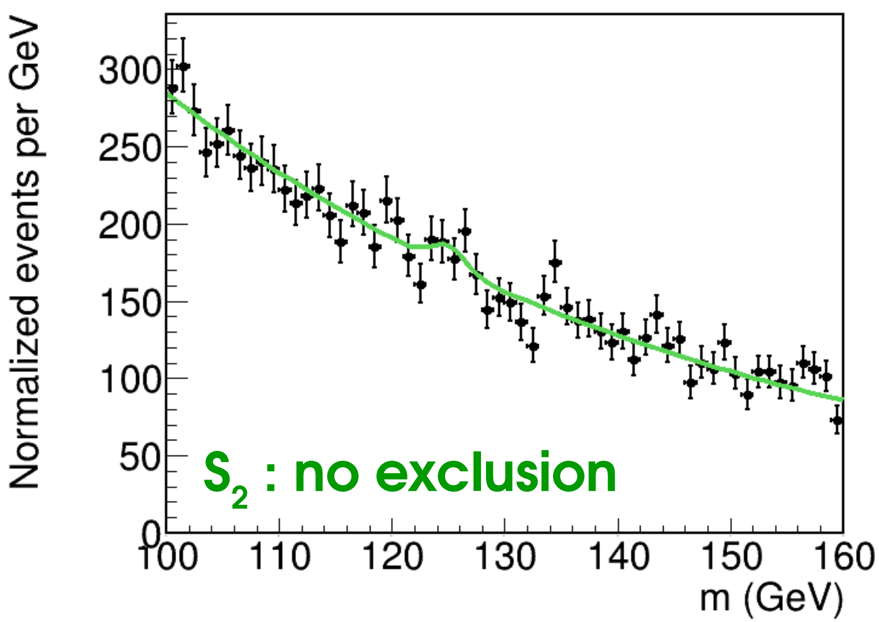
→ Adjust S_0 until 95% CL exclusion (p_{S0} = 5%) is reached

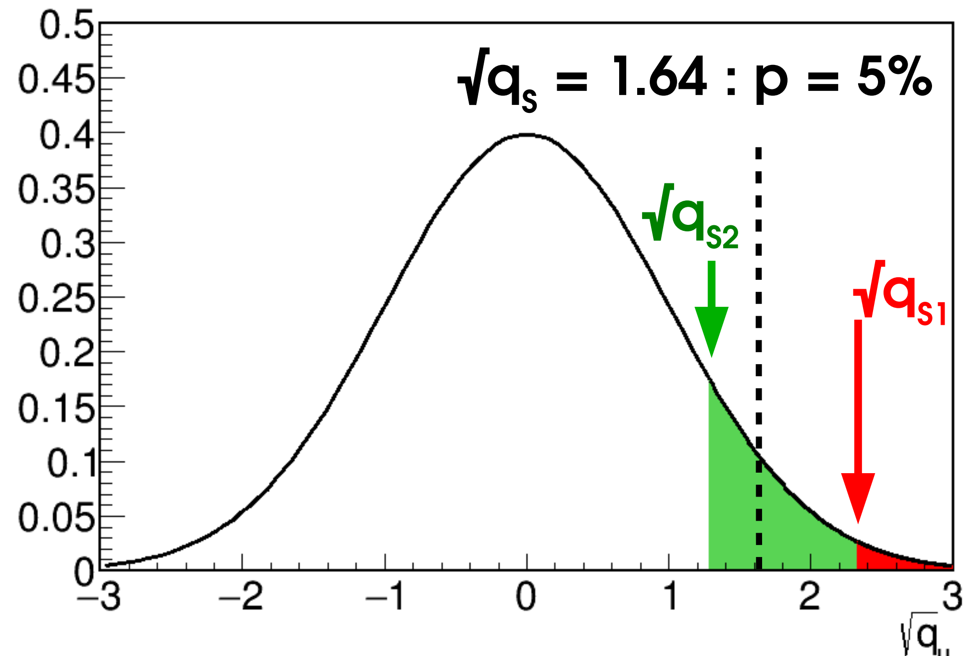
Asymptotic case: need $\sqrt{q_{S0}}$ = 1.64

Asymptotics

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

CL	Region
90%	$\sqrt{q_S} > 1.28$
95%	$\sqrt{q_S} > 1.64$
99%	$\sqrt{q_S} > 2.33$





Inversion : Getting the limit for a given CL

Procedure:

→ Compute q_{S0} for some S_0, get the **exclusion p-value p_{S0}**.

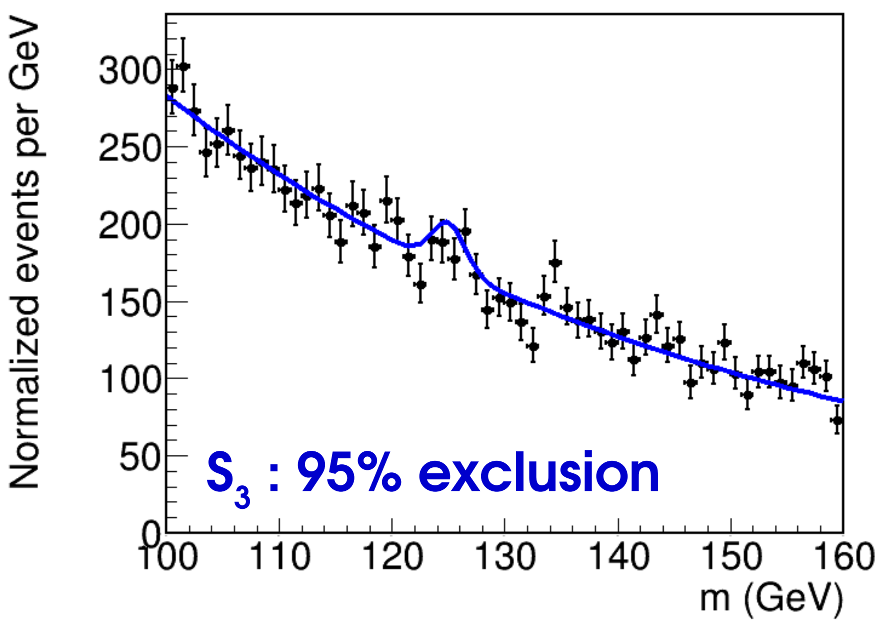
→ Adjust S_0 until 95% CL exclusion (p_{S0} = 5%) is reached

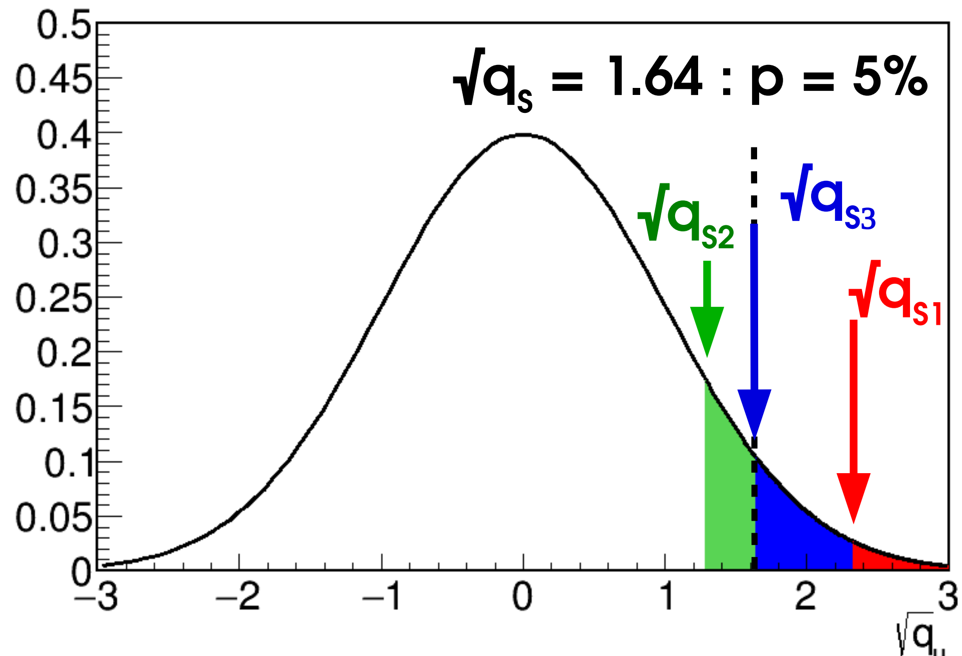
Asymptotic case: need $\sqrt{q_{S0}} = 1.64$

Asymptotics

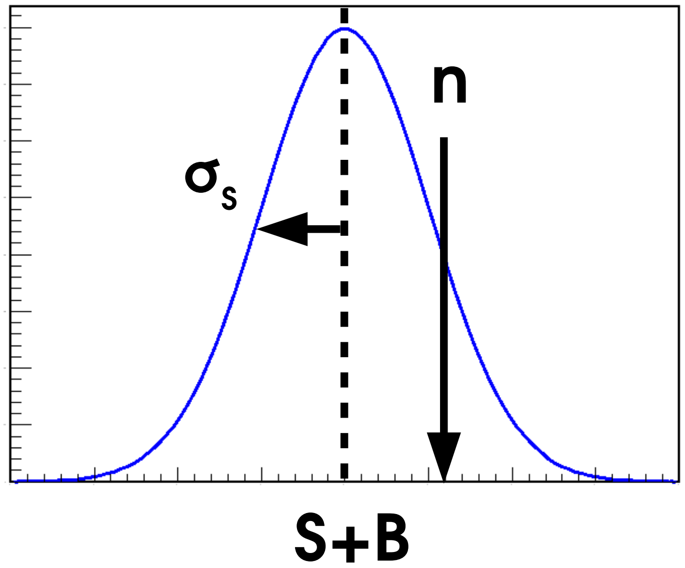
$$\sqrt{q_{S_0}} = \Phi^{-1}(1 - p_0)$$

CL	Region
90%	$\sqrt{q_S} > 1.28$
95%	$\sqrt{q_S} > 1.64$
99%	$\sqrt{q_S} > 2.33$





Homework 4: Gaussian Example



Usual Gaussian counting example with known B:

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

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

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

→ Compute q_{S0}

→ 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

Upper Limit Pathologies

Upper limit: $S_{up} \sim \hat{S} + 1.64\,\sigma_S$.

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

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

How can this be ?

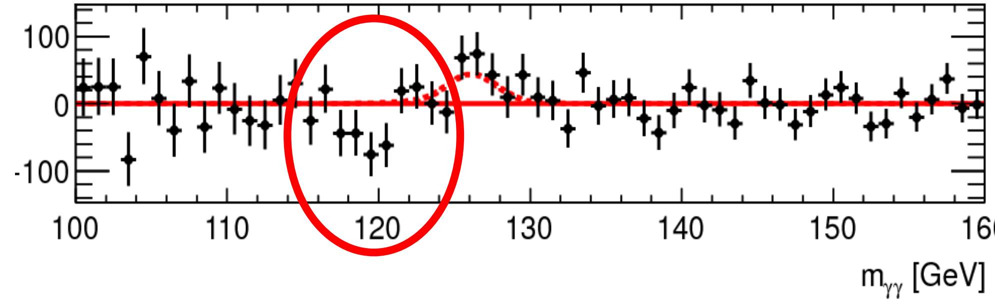
→ **Background modeling issue ?...** Or:

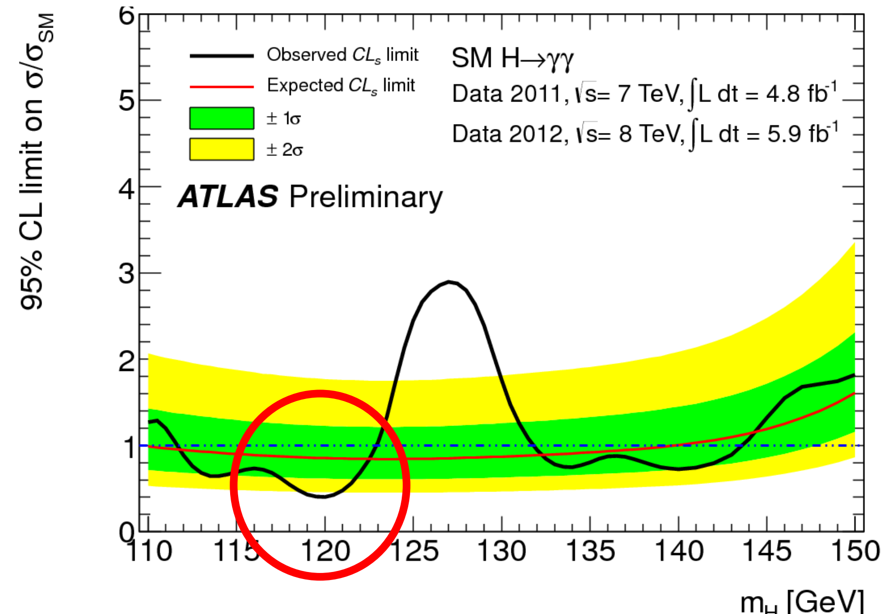
→ This is a **95%** limit ⇒ **5% of the time, the limit wrongly excludes the true value,** *e.g.* S*=0.

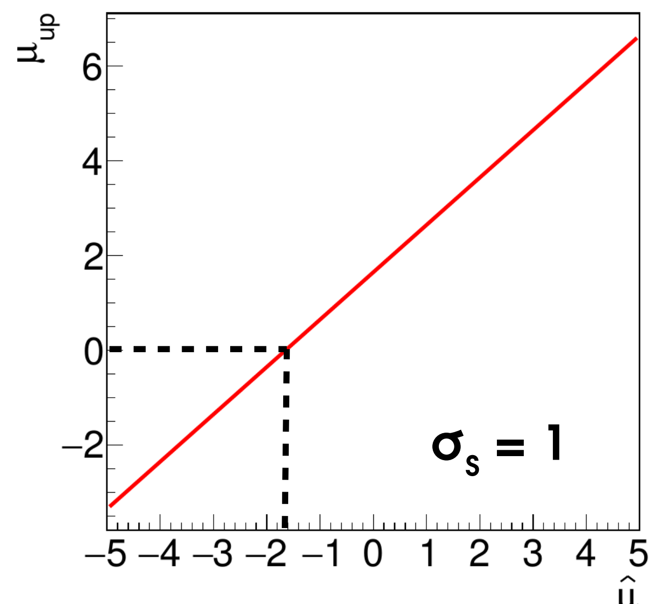
Options

→ **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

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

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

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

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

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

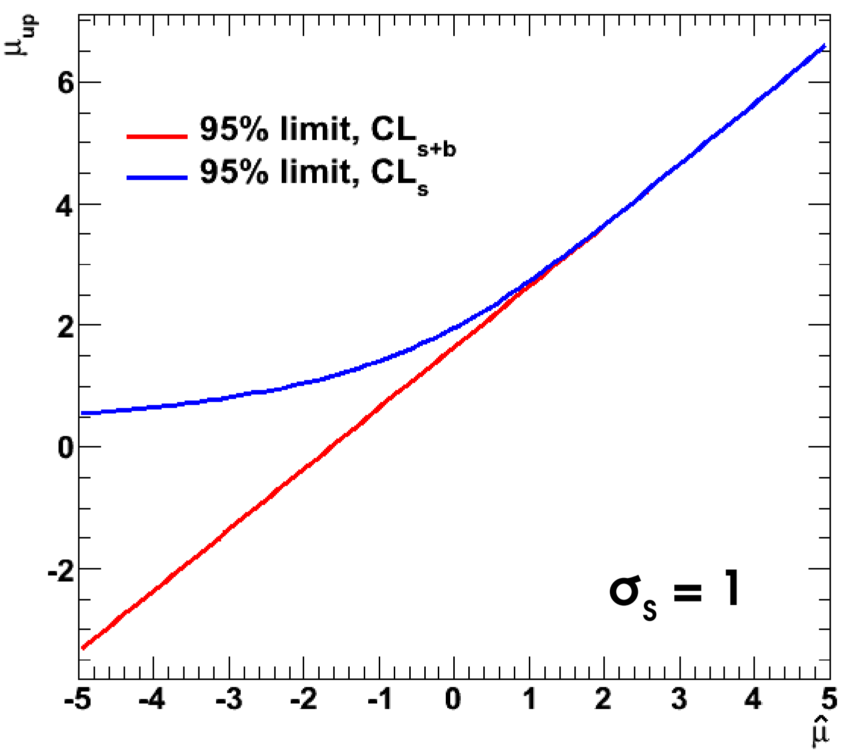
$\hat{S}$ compatible with 0 : $p_B \sim O(1)$

p_{CLs} **~ p_{S0} ~ 5%, no change**.

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

$p_{CLs} \sim p_{S0}/(1-p_B) \gg 5\%$

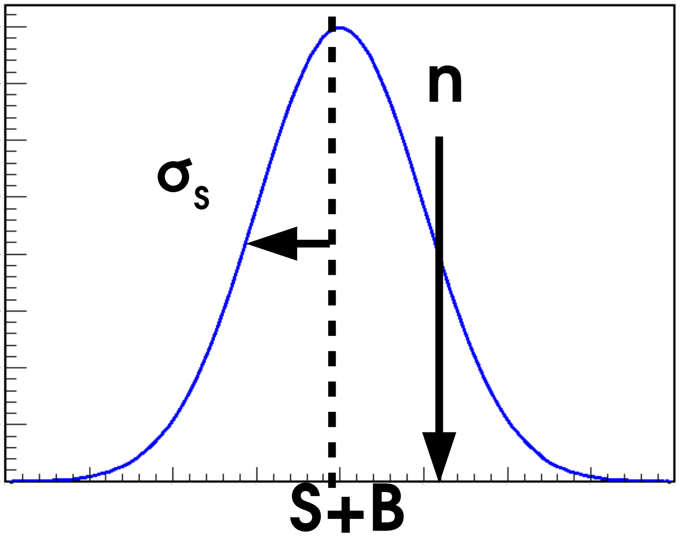
→ lower exclusion ⇒ higher limit, usually >0 as desired



Drawback: ***overcoverage***

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

Homework 5: CL_s : 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: $S_{up} = \hat{S} + 1.64\,\sigma_S$ at 95 % 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 $\hat{S} \sim 0$, $S_{up} = \hat{S} + 1.96\,\sigma_S$ at 95 % CL

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

Same exercise, for the Poisson case with $n_{obs} = 0$. Perform an exact computation of the 95% 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)

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

Solution:

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

Outline

Computing statistical results

Confidence intervals

Upper limits on signal yields

Expected Limits

Generating Pseudo-data

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

⇒ Possible outcomes of the experiment, for given parameter values

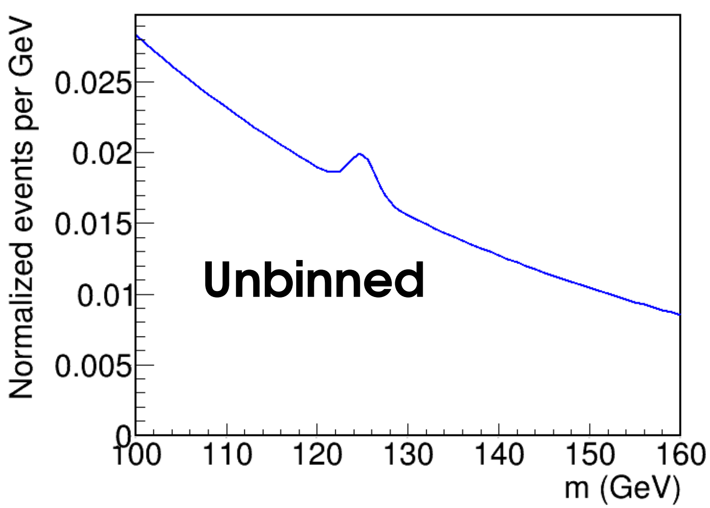
Can draw random events according to PDF : **generate** ***pseudo-data***

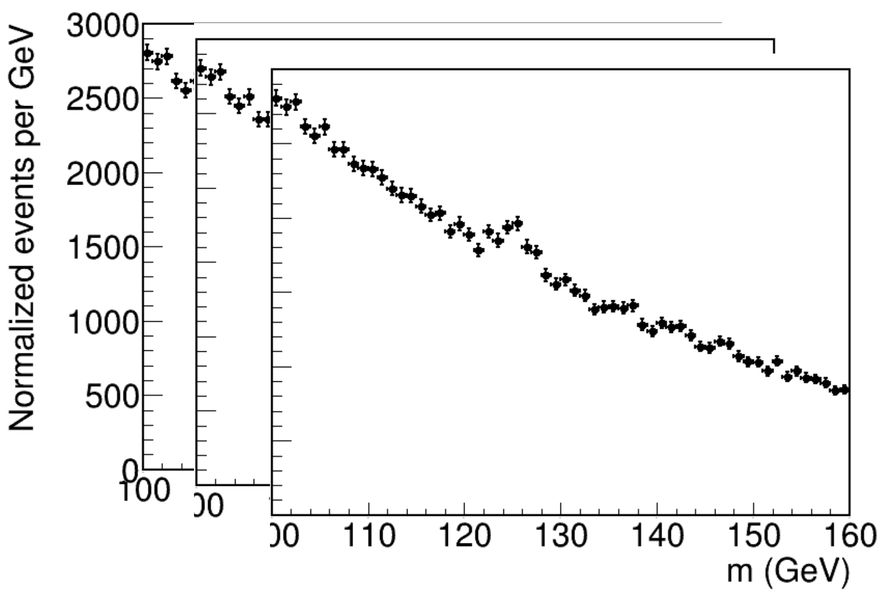
$$P(\lambda=5)$$

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

Each entry = separate "experiment"

Generate





Expected Limits: Toys

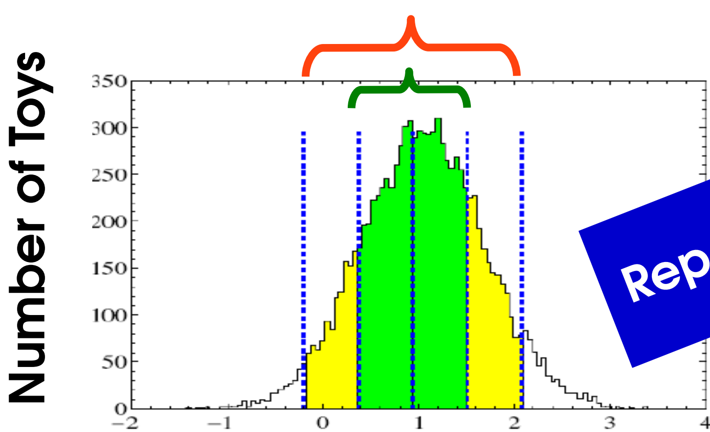
***Expected* results**: median outcome under a given hypothesis
→ usually B-only for searches, but other choices possible.

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

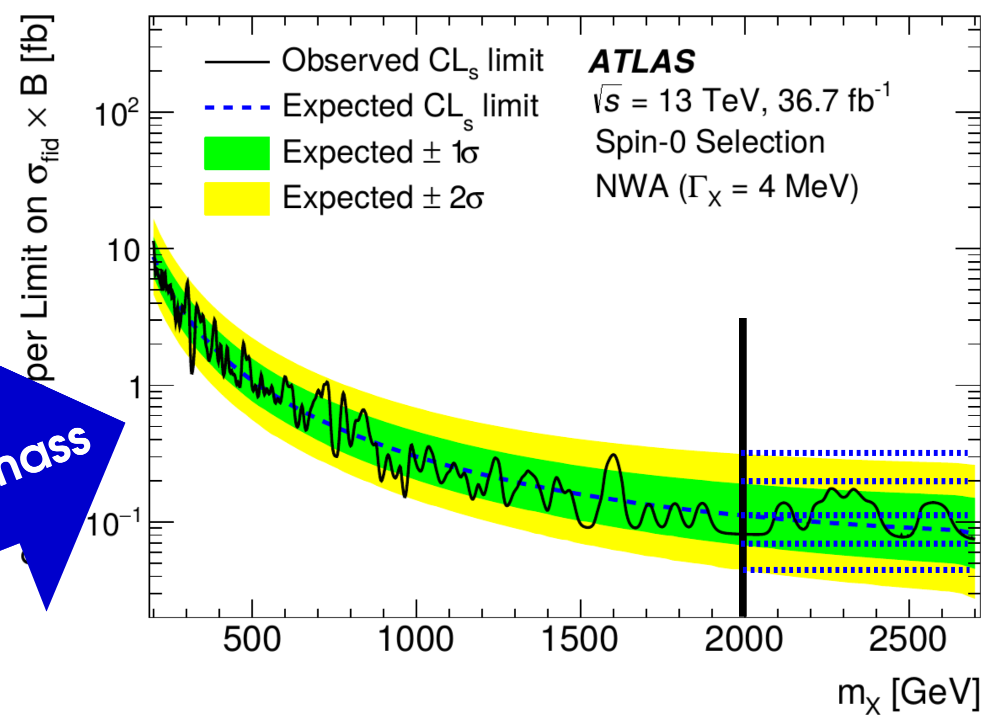
68% of toys 95% of toys



Eur.Phys.J.C71:1554,2011 **Computed limit**

Repeat for each mass

Phys. Lett. B 775 (2017) 105



Expected Limits: Asimov Datasets

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

Two main ways to compute:

Strictly speaking, Asimov dataset if

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

where X_0 is the generation value

→ ***Asimov Datasets***

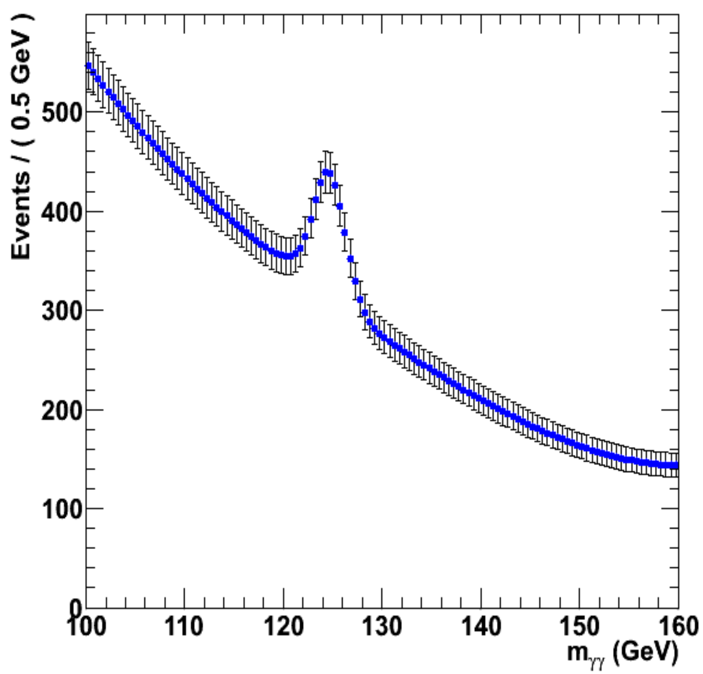
- Generate a "perfect dataset" – *e.g.* for binned data, set bin contents carefully, no fluctuations.
- Gives the median result immediately:

  median(toy results) ↔ result(median dataset)
- Get bands from asymptotic formulas:
  Band width

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

⊕ **Much faster (1 "toy")**

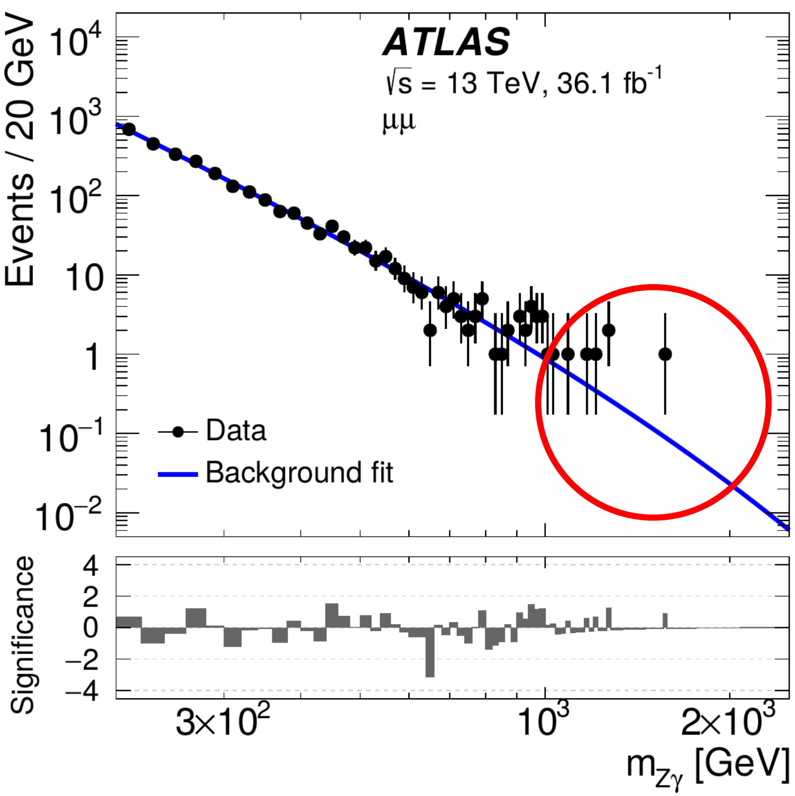
⊖ **Relies on Gaussian approximation**

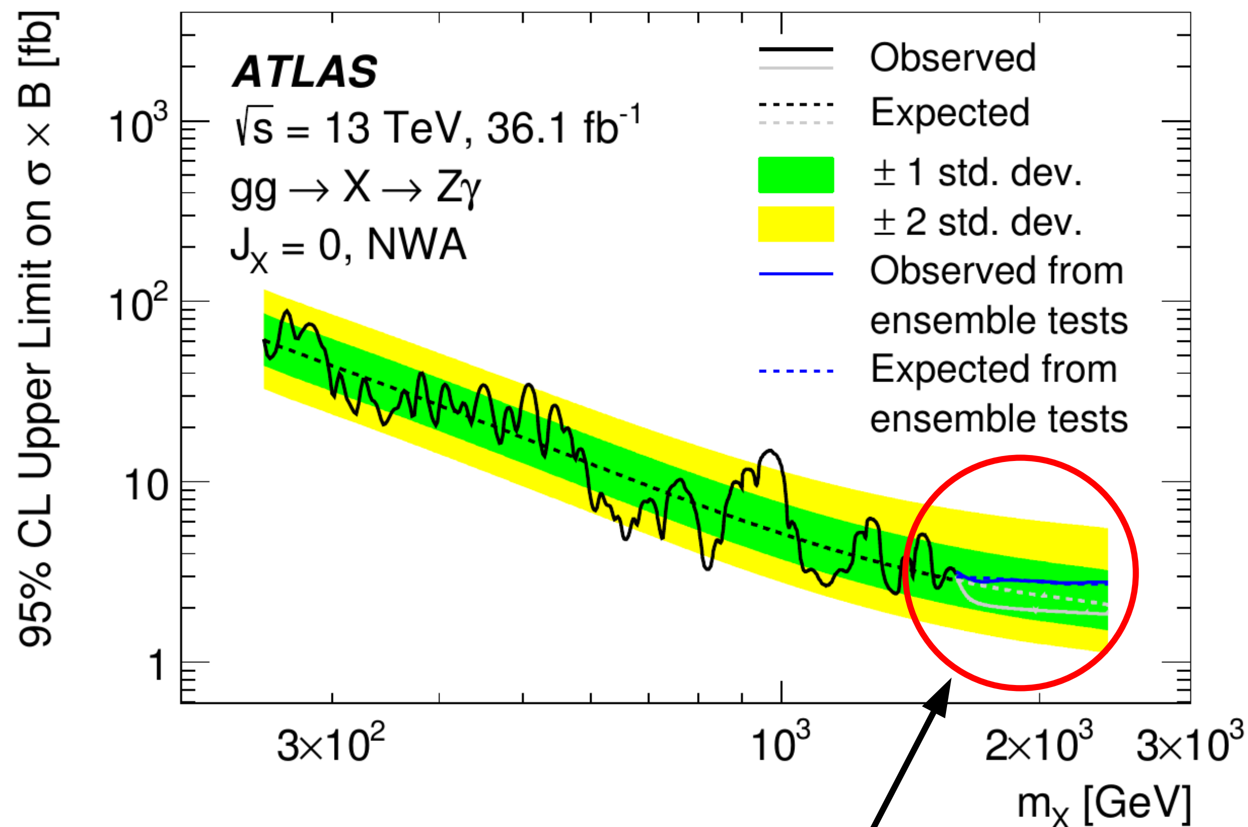


Toys: Example

JHEP 10 (2017) 112

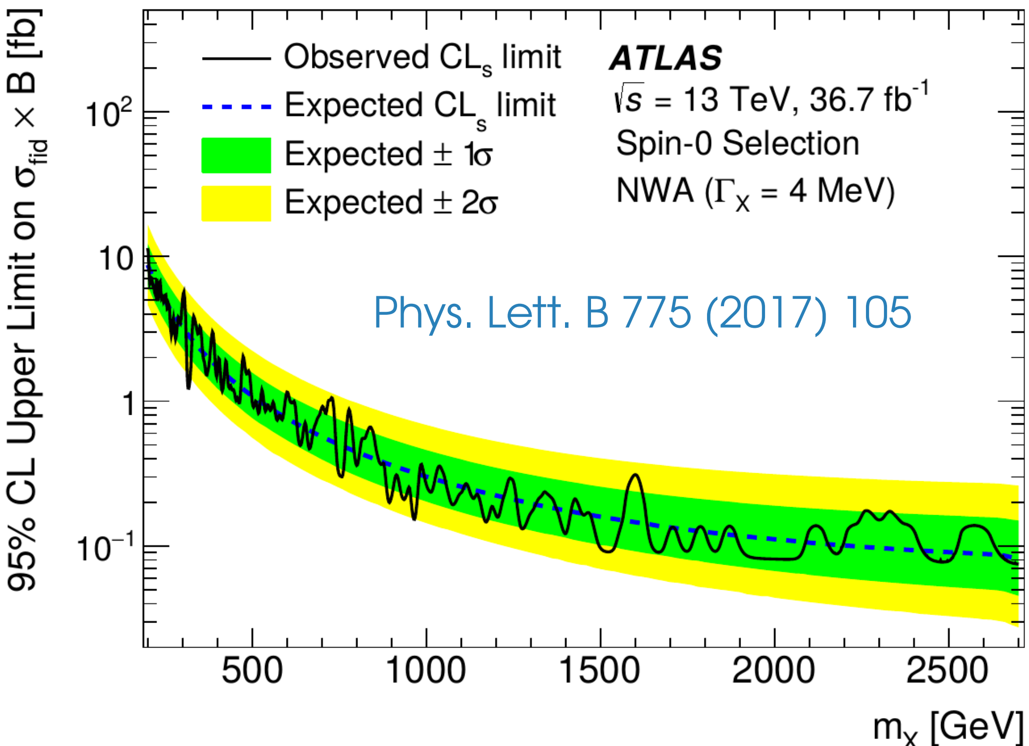
ATLAS X→Zγ Search: covers 200 GeV < m_X < 2.5 TeV
→ for m_X > 1.6 TeV, low event counts ⇒ derive results from toys



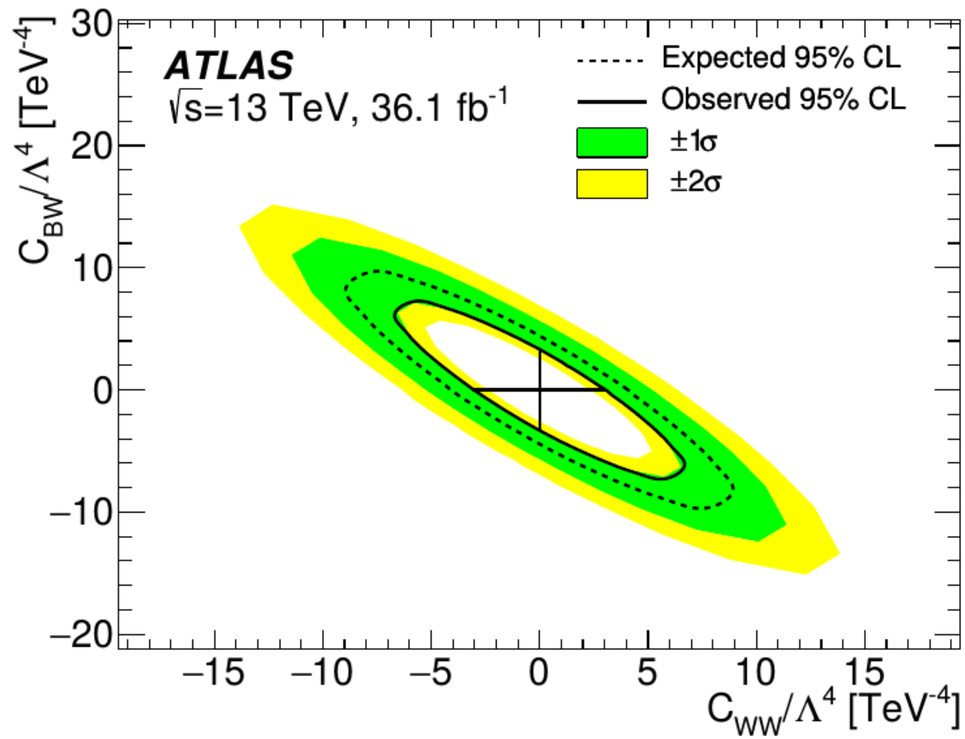


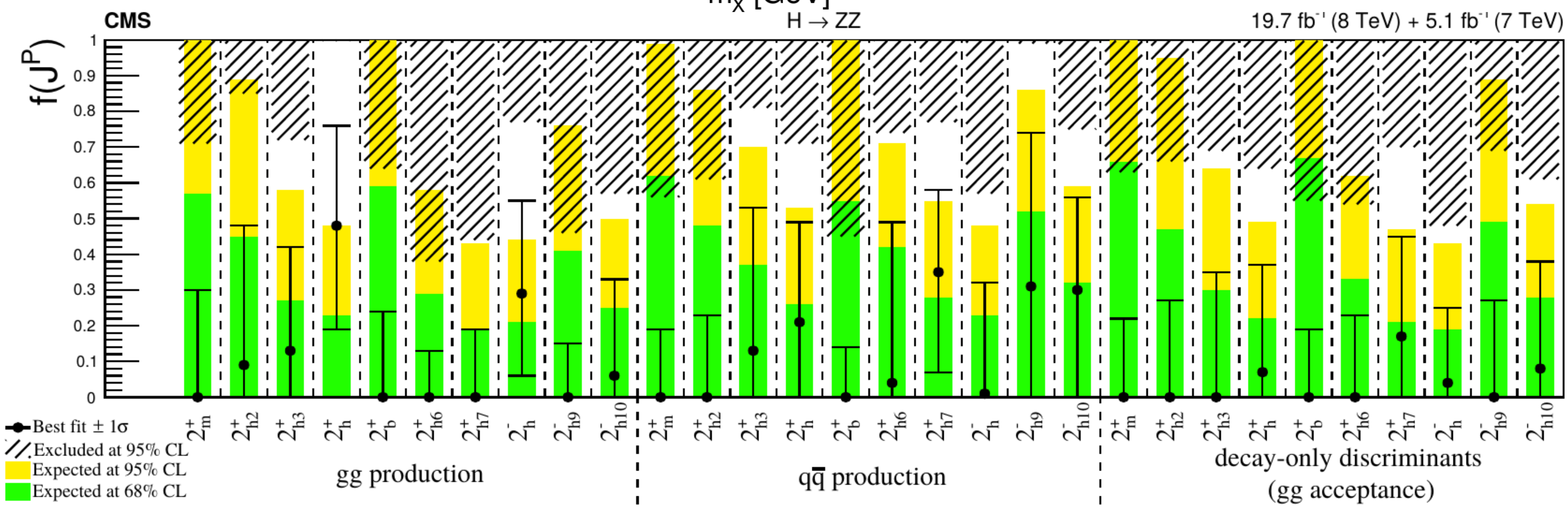
Asimov results (**in gray**) give optimistic result compared to toys (**in blue**)

Upper Limit Examples



ATLAS 2015-2016 4l aTGC Search

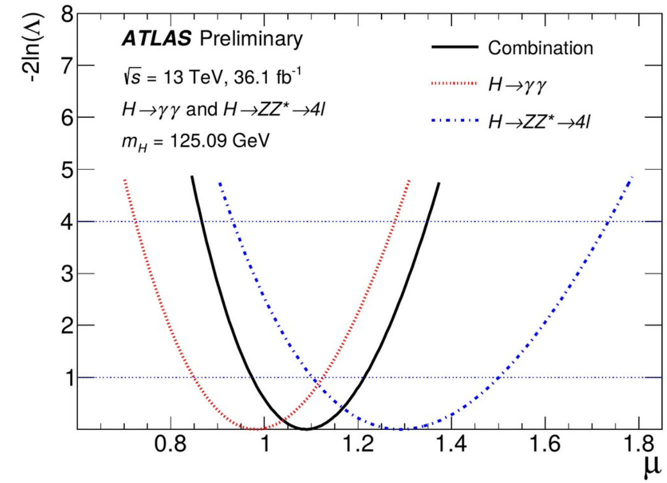




Phys. Rev. D 92 (2015) 012004

Takeaways

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



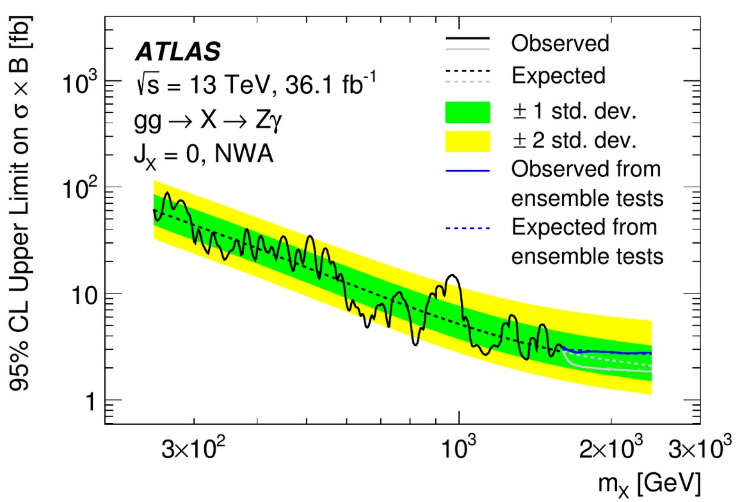
→ Crossings with $t_{\mu 0} = Z^2$ for $\pm Z\sigma$ intervals (in 1D)

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

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

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

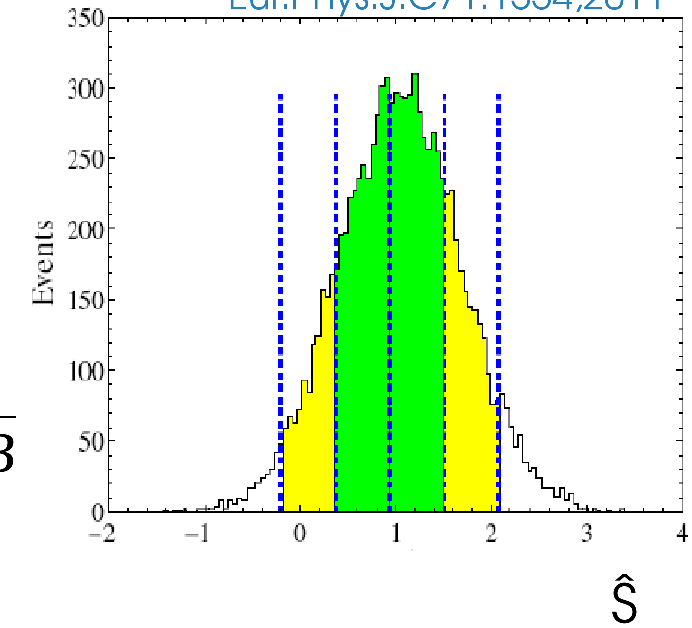
Poisson regime, n=0 : **S_{up} = 3 events**



Extra Slides

CL_s : Gaussian Bands

Eur.Phys.J.C71:1554,2011



Usual Gaussian counting example with known B:

95% CL_s upper limit on S:

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

with

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

Compute expected bands for S=0:

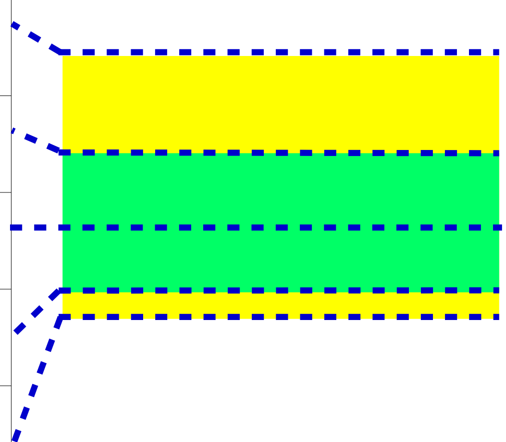
→ **Asimov dataset ⇔ $\hat{S} = 0$** :

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

→ **± nσ bands**:

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

n	$S_{exp}^{\pm n}$ / $\sqrt{B}$
+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 $\sigma^2_{S,A} = \dfrac{S^2}{q_S(\mathrm{Asimov})}$ depends on S, for non-Gaussian cases,different values for each band...

Comparison with LEP/TeVatron definitions

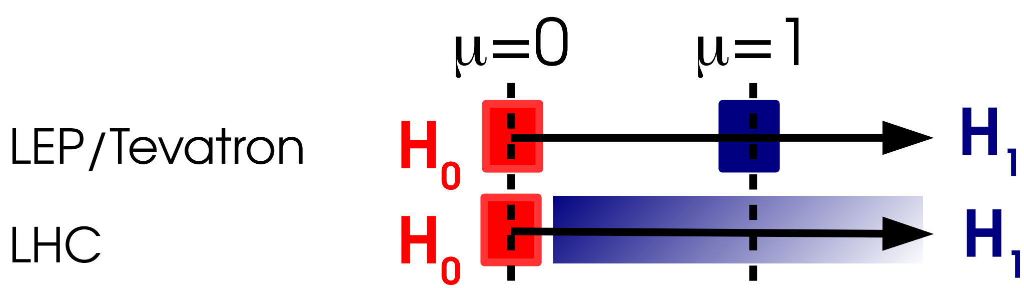
Likelihood ratios are not a new idea:

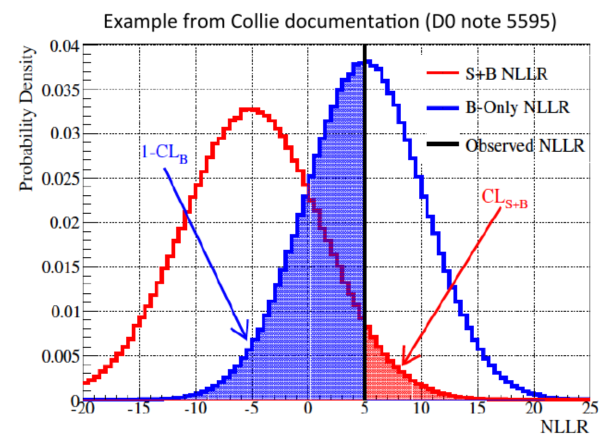
- **LEP**: Simple LR with NPs from MC
  - Compare μ=0 and μ=1
- **Tevatron**: PLR with profiled NPs

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

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

Both compare to **μ=1** instead of best-fit **μ̂**

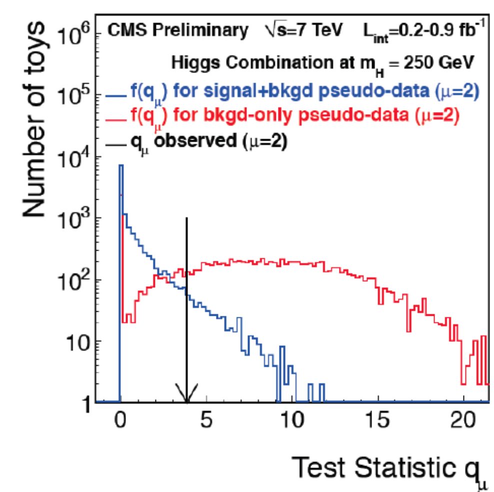




→ Asymptotically:

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

→ Still use TeVatron-style for discrete cases



Andrey Korytov, EPS 2011