Data analysis and Statistics at the LHC

Dr. Nicholas Wardle

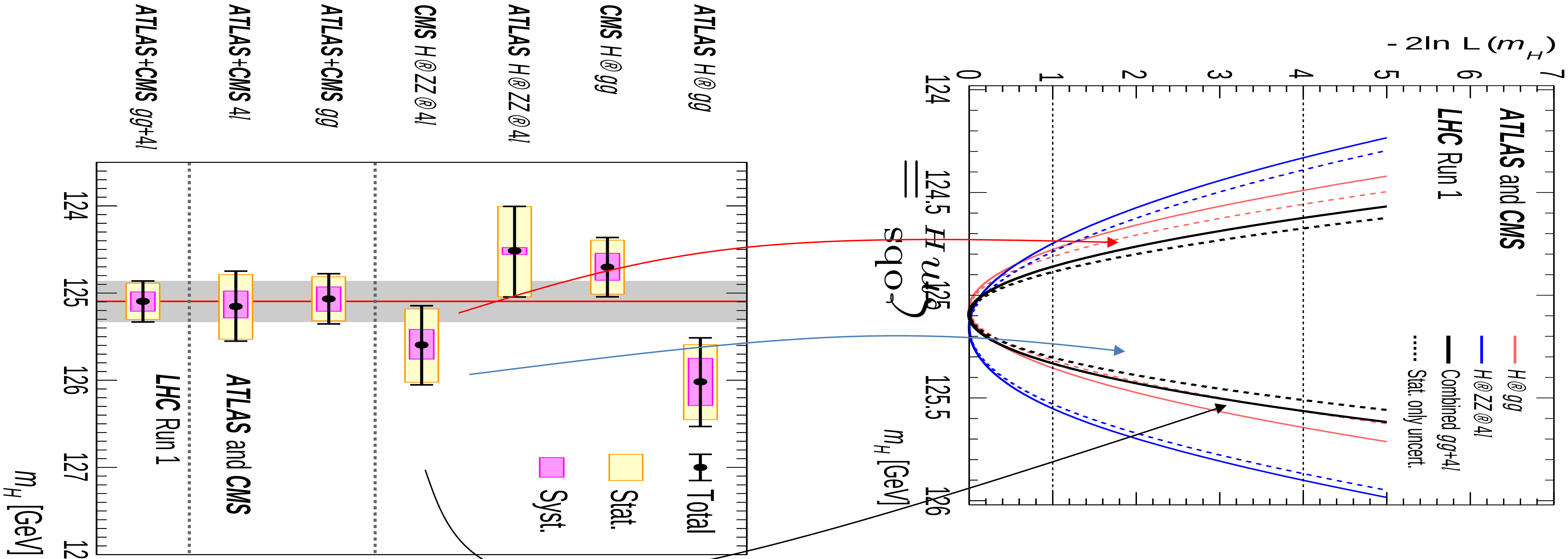
VSOP *Quy Nhon, Vietnam*

15-26 July 2024

Uncertainties @ the LHC Recap

When we see a plot like this…

And we know that the **error bars** are derived from the region for which $\zeta^{\mathrm{obs}}_{m_H} < 1$

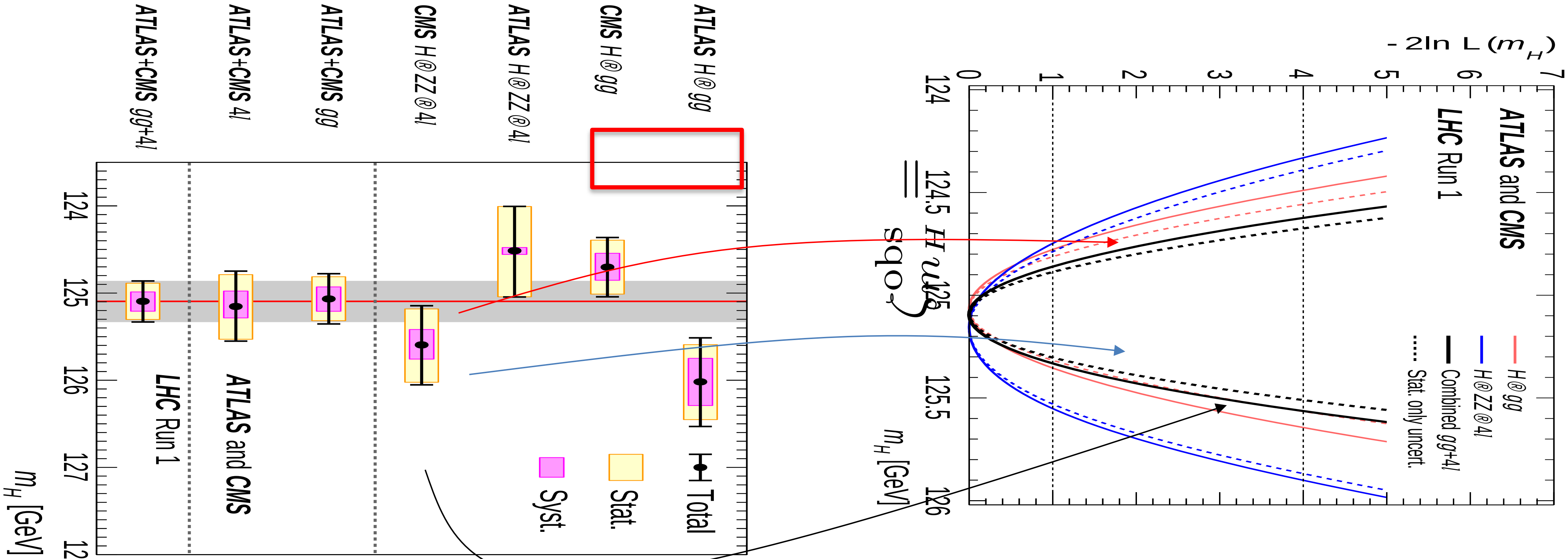


We know the points represent the **Maximum Likelihood estimates**

Uncertainties @ the LHC Recap

When we see a plot like this…

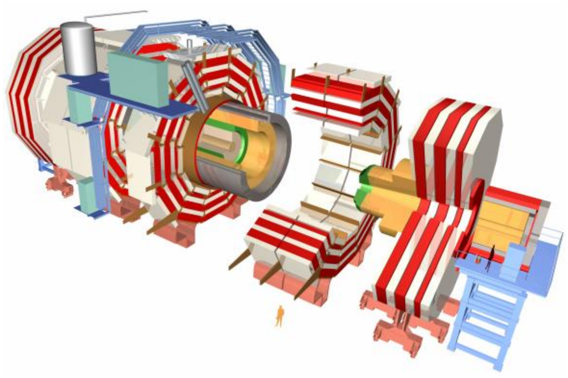
And we know that the **error bars** are derived from the region for which $\zeta^{\mathrm{obs}}_{m_H} < 1$



We know the points represent the **Maximum Likelihood estimates**

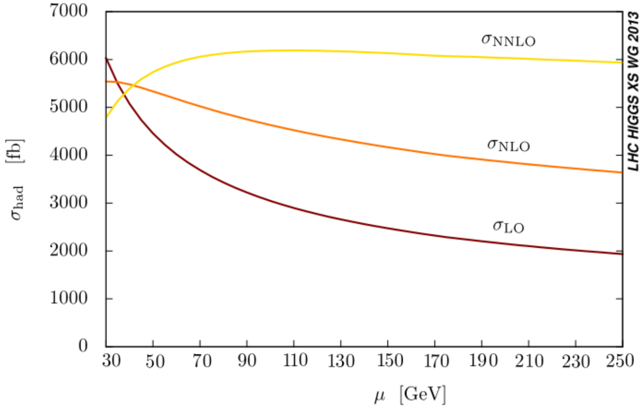
What does **this part** of the error bar mean?

Systematic Uncertainties



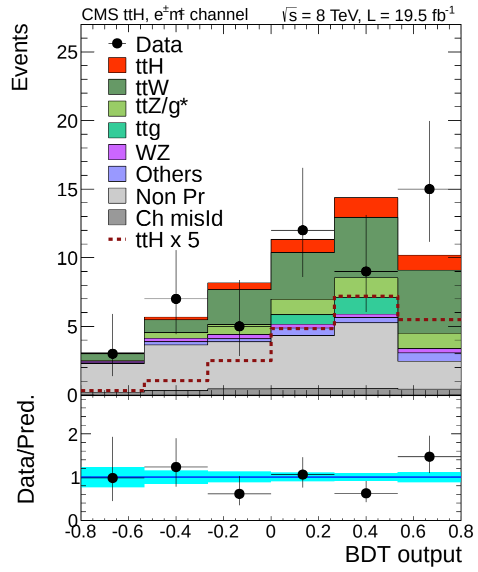
Experimental/Detector systematics:

- Object efficiencies, energy calibrations, luminosity



Signal theory uncertainties:

- Inclusive x-section uncertainties, QCD scale, pdf, UEPS, Branching ratios, jet counting



Background theory uncertainties:

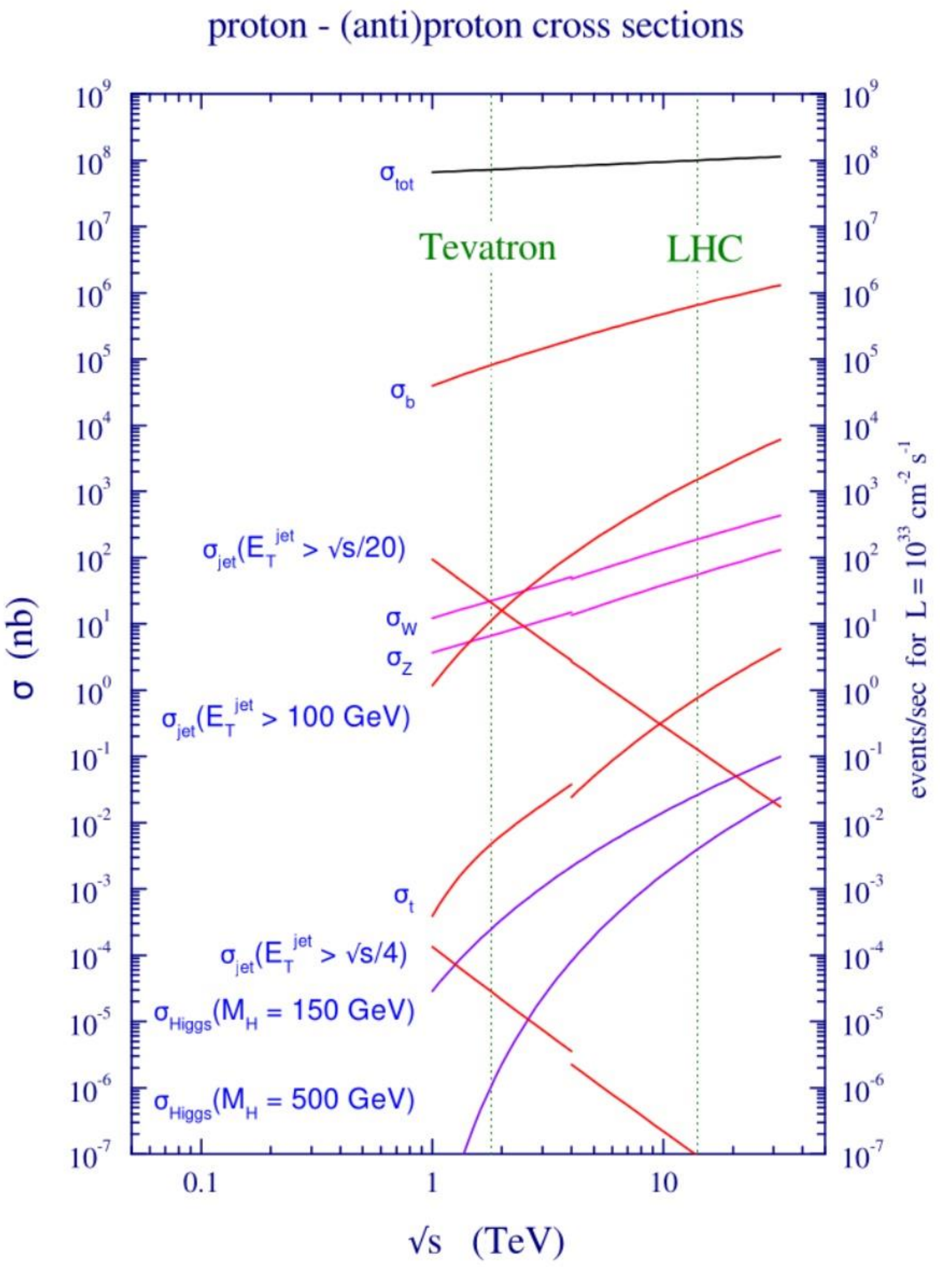
- Often rather different phase-spaces considered for extrapolating from control regions for data-driven estimates
- Limited simulation size to predict p(B)

Cross-section example

Remember our formulae for the number of expected event

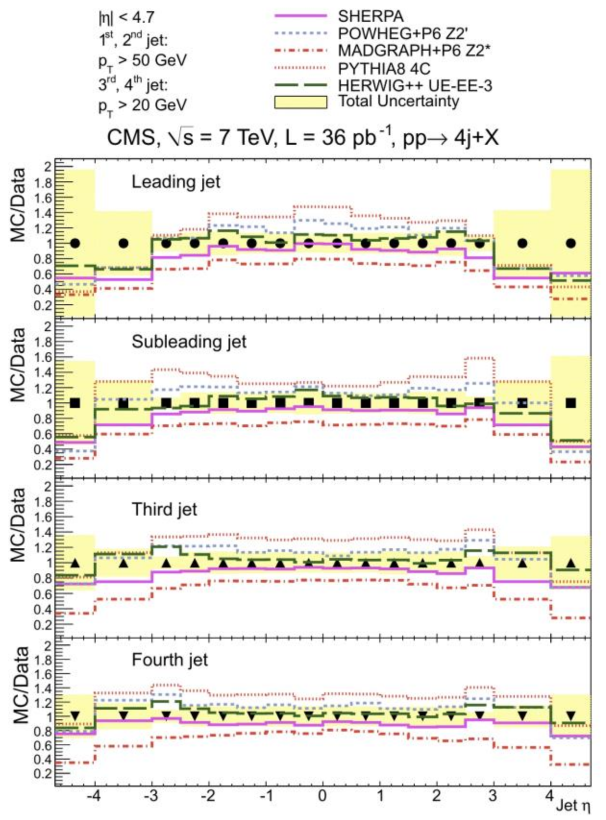
$$N = L\sigma$$

This tells us the total (inclusive) expected rate of events but from lecture 1, we know that we don't keep all of the events → selection!



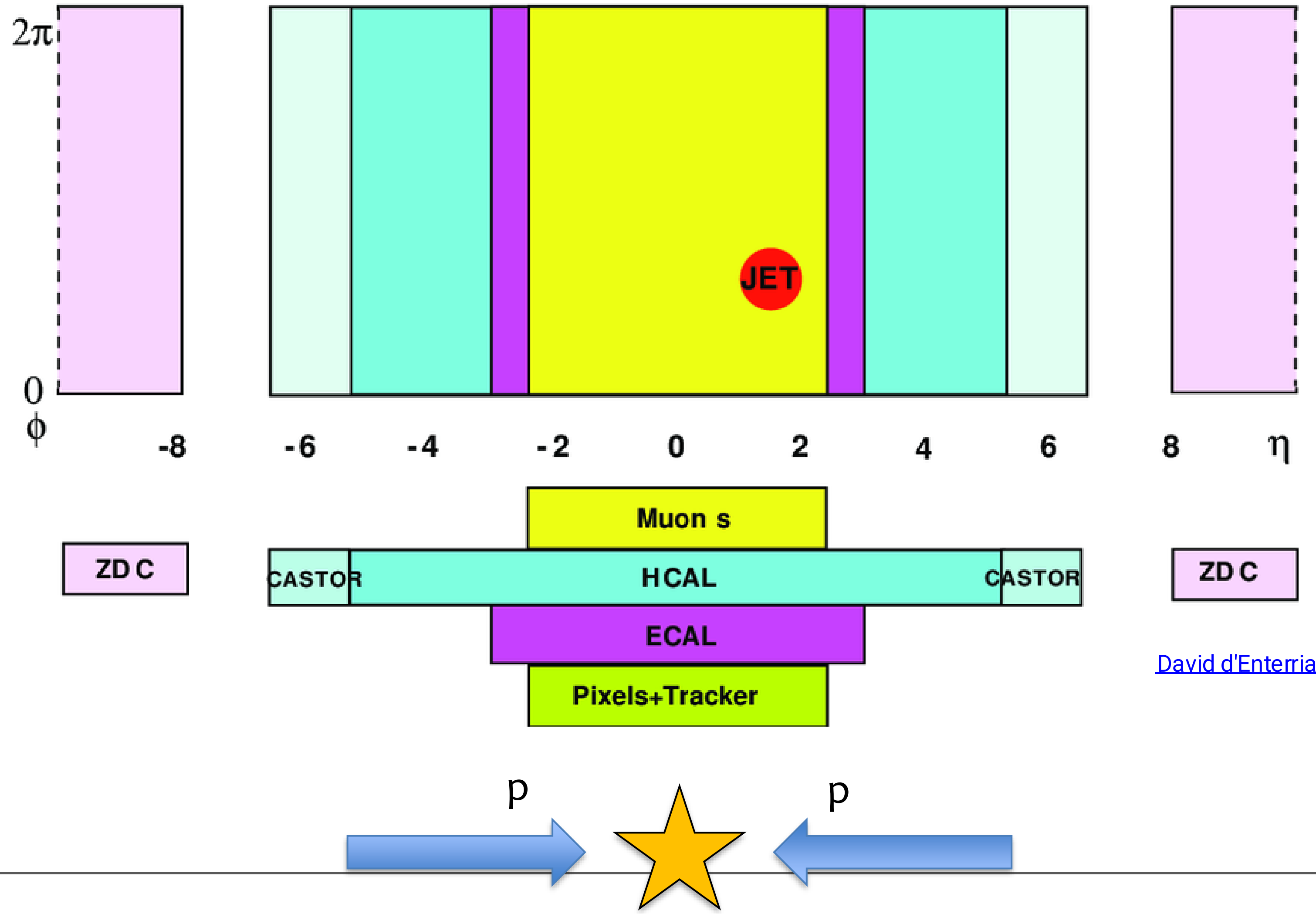
Acceptance

$$N = L\sigma A$$



Event kinematics not known precisely
→ **Acceptance uncertainty**

Need to account for **acceptance of detector** for different physics objects



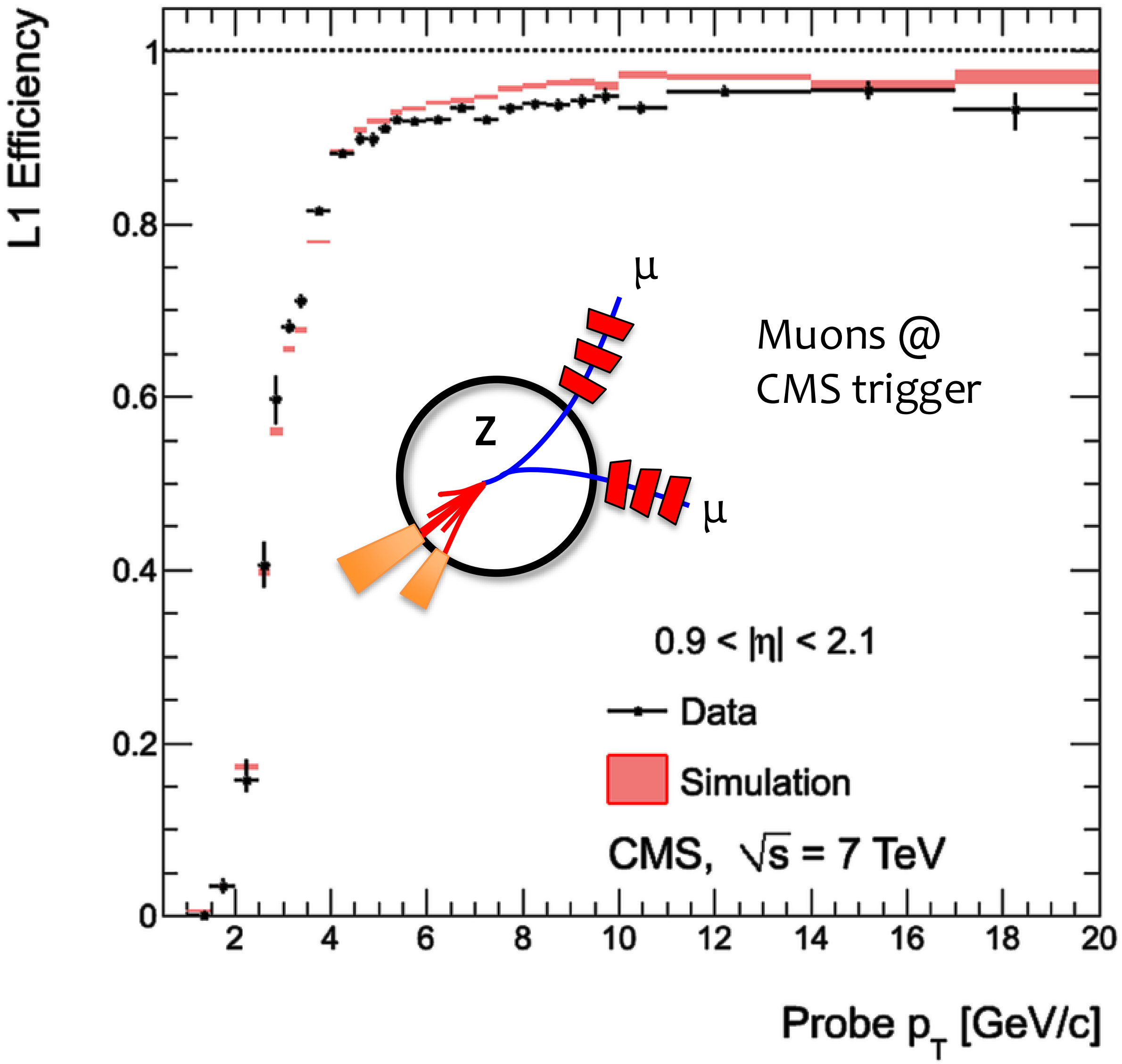
David d'Enterria

Efficiency

$$N = L\sigma A\epsilon$$

Not all particles reconstructed with perfect efficiency (missing hits in tracker, leakage/gaps in calorimeter…)

→ Uncertainty in measurements offline and **online** can lead to **uncertainty in total efficiency** of selection



Nuisance parameters

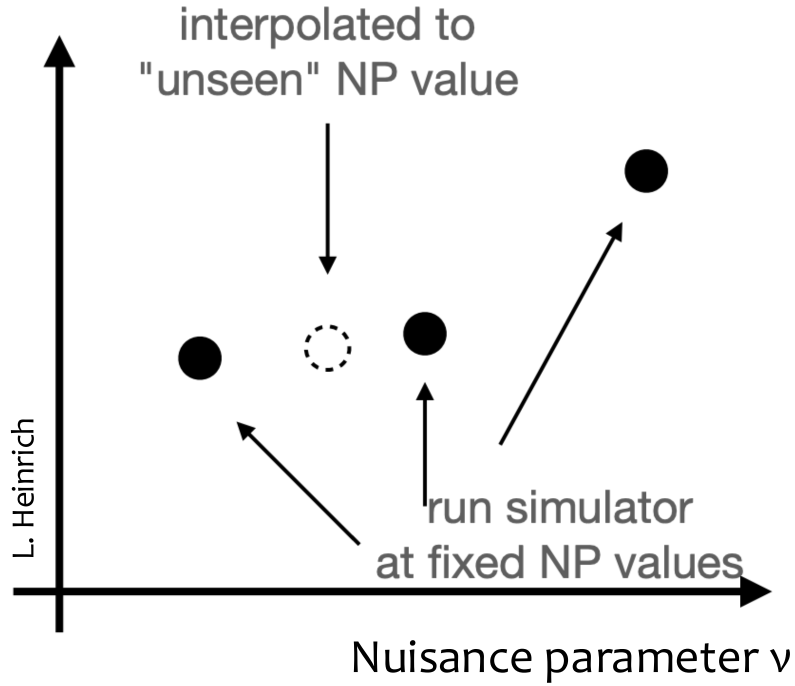
We model the effects of systematic uncertainties through the introduction of nuisance parameters into our model

$$p(X;\theta) \to p(X;\mu,\nu)$$

μ **Parameters of interest:** cross-section, Top mass, …

ν **Nuisance parameters:** Jet energy scale, Luminosity, …

We need to choose a parameterization for the effects of each of our nuisance parameters



Nuisance parameters

We model the effects of systematic uncertainties through the introduction of nuisance parameters into our model

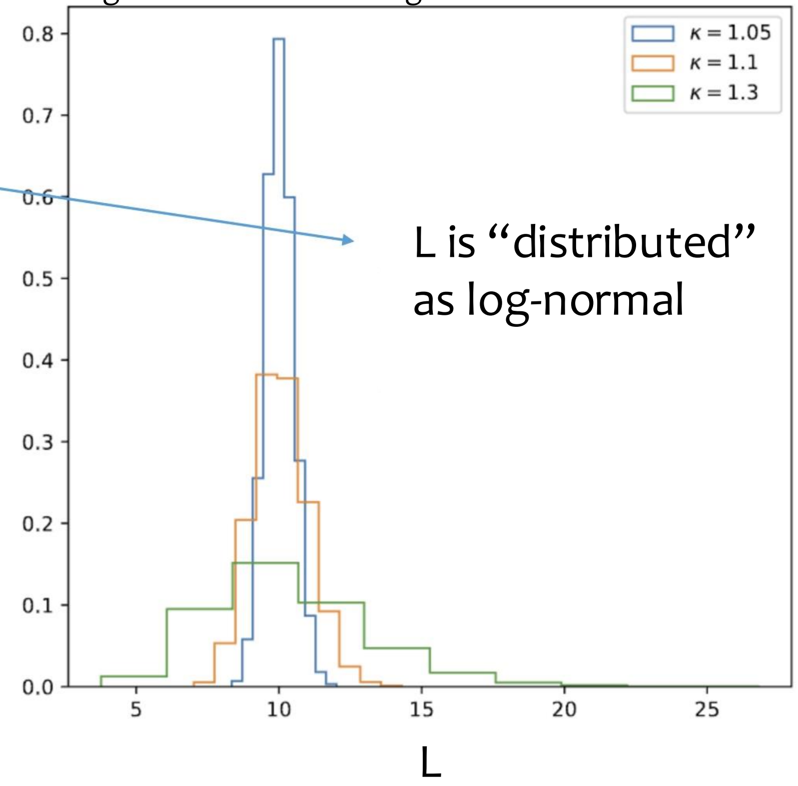
$$p(X;\theta) \to p(X;\mu,\nu)$$

μ **Parameters of interest:** cross-section, Top mass, …

ν **Nuisance parameters:** Jet energy scale, Luminosity, …

$$N = \mu \cdot \sigma L_0(\kappa)^{\nu} A\epsilon$$

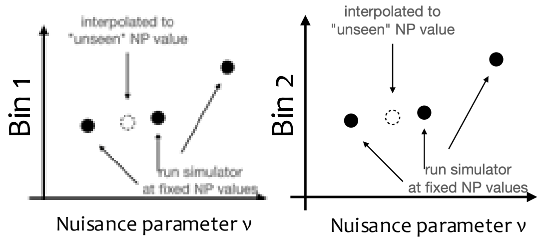
Log-normal cannot be negative



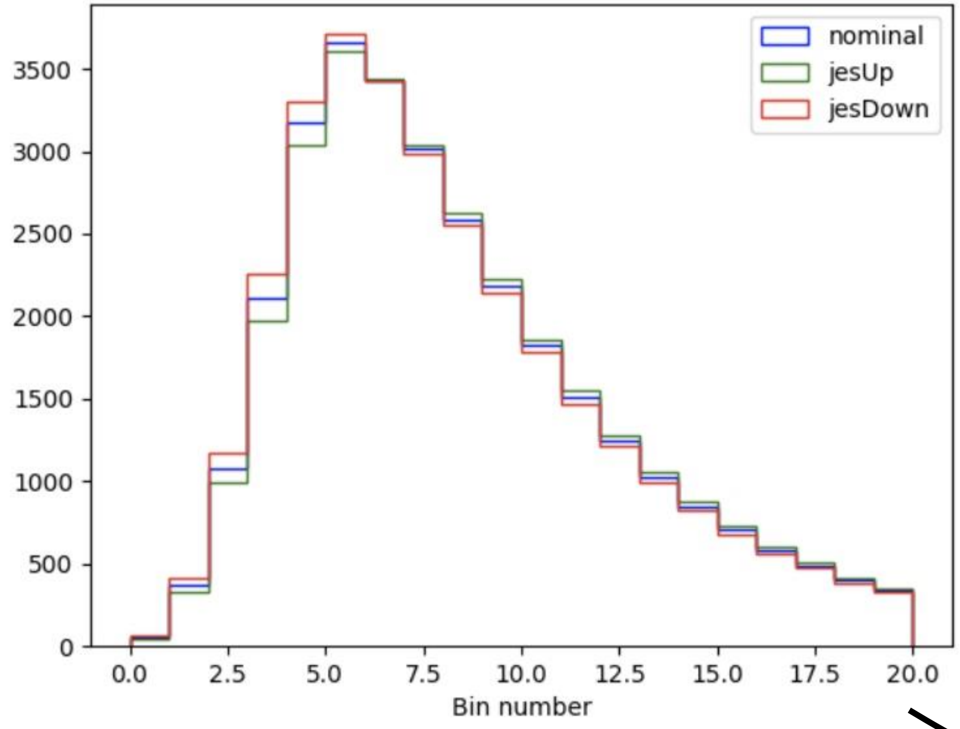
L is “distributed” as log-normal

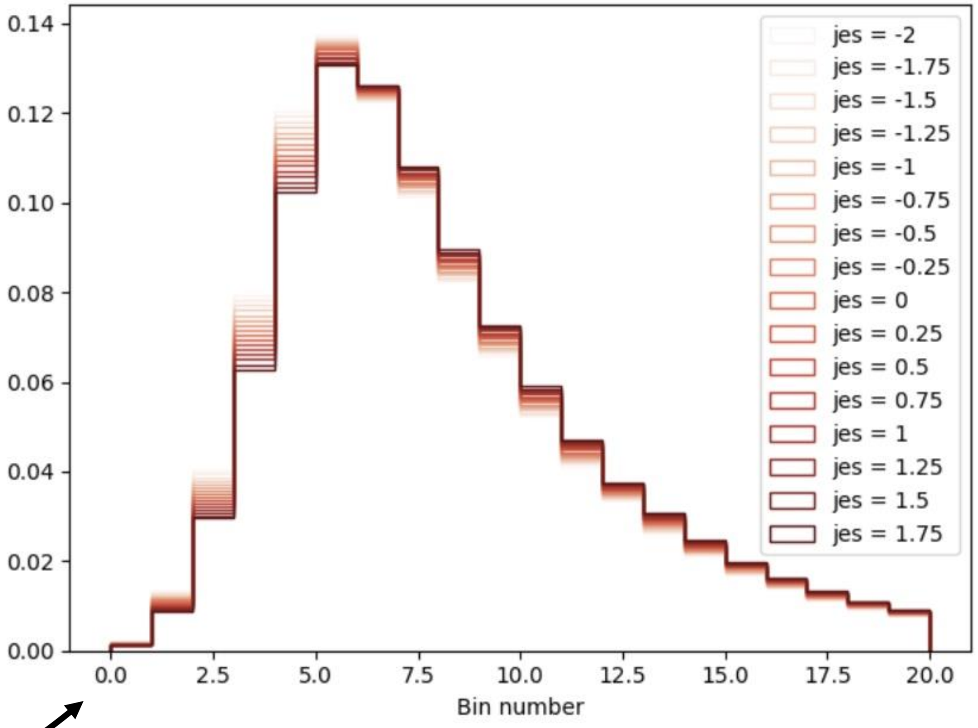
We often use “log-normal” uncertainties to model the effect of each nuisance parameter – eg for luminosity uncertainty

Shape uncertainties



For distributions (shapes) this is more complicated as whole distribution can change as a result of varying nuisance parameters





We use morphing / interpolation from nominal and alternative templates to derive continuous parameterization of shape

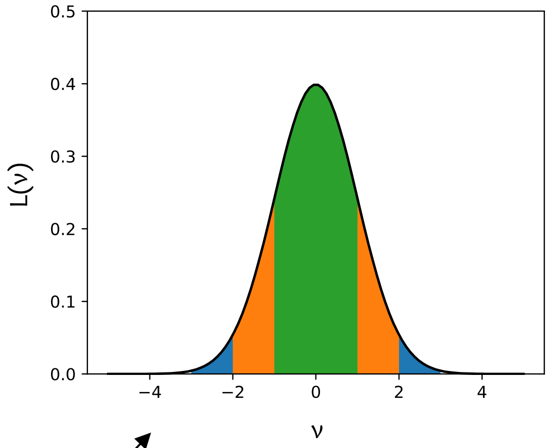
Enhancing the Likelihood function

Nuisance parameters often constrained through measurements
→ calibration measurements for energy scale, tag-and-probe for efficiencies etc …

We can **include these constraints** in the likelihood function!

Eg for a Poisson likelihood,

$$L(\lambda) = e^{-\lambda}\frac{\lambda^k}{k!} \rightarrow e^{-\lambda(r,\nu)}\frac{\lambda(r,\nu)^k}{k!}\boxed{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\nu^2}}$$



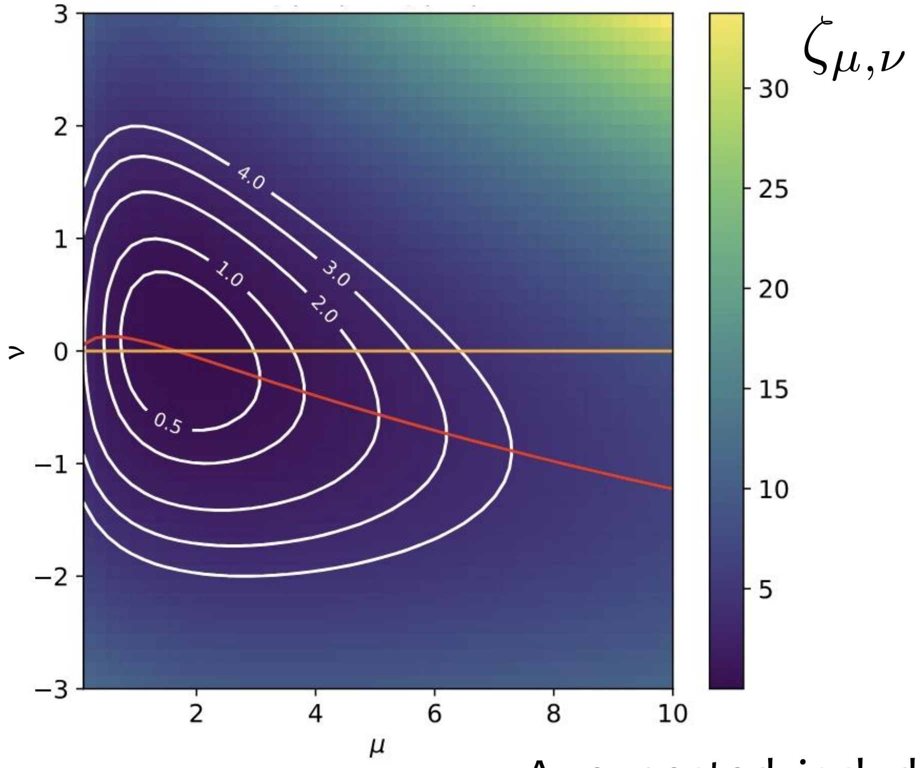
$$\lambda(r,\nu) = r\sigma L_0 (1.3)^{\nu} A\epsilon$$

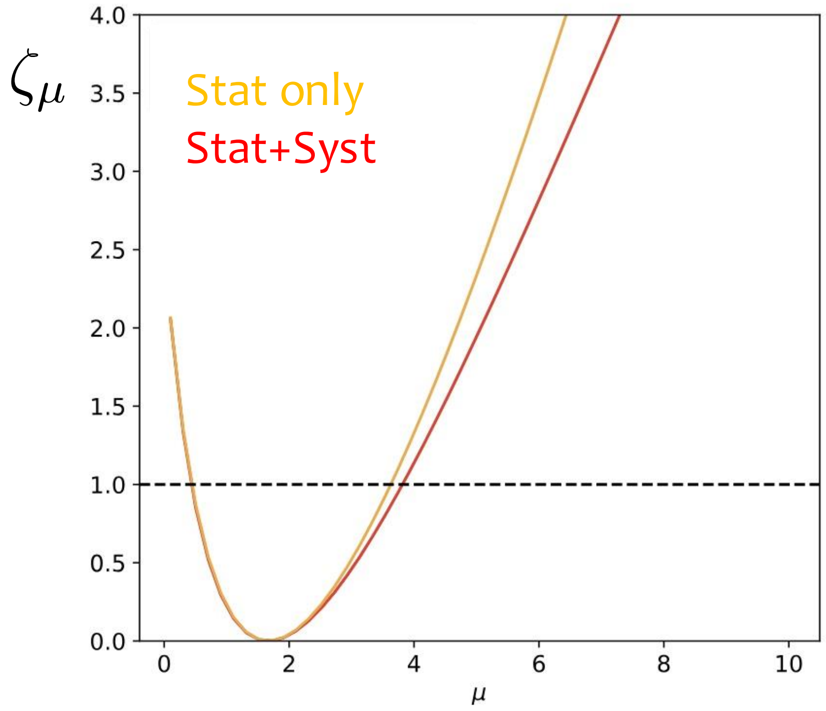
Profiled likelihood

For statistical inference, we replace the likelihood function with the **profiled likelihood function**

$$L(\mu,\nu) \rightarrow L\left(\mu, \hat{\nu}(\mu)\right) \quad \& \quad \zeta_{\mu,\nu} \rightarrow \zeta_{\mu} = -2\log\frac{L\left(\mu,\hat{\nu}(\mu)\right)}{L\left(\hat{\mu},\hat{\nu}(\hat{\mu})\right)}$$

Simple Poisson case





As expected, including systematic uncertainties increases total uncertainty!

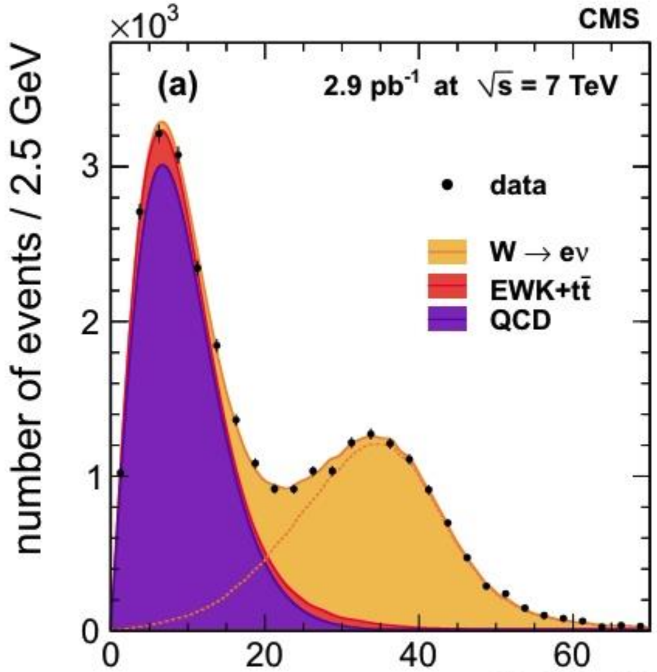
Analysis strategies

When designing an analysis, we try to consider the balance between systematic and statistical effects!
For example, lets consider cross-section measurements of $pp \rightarrow W \rightarrow e\nu$

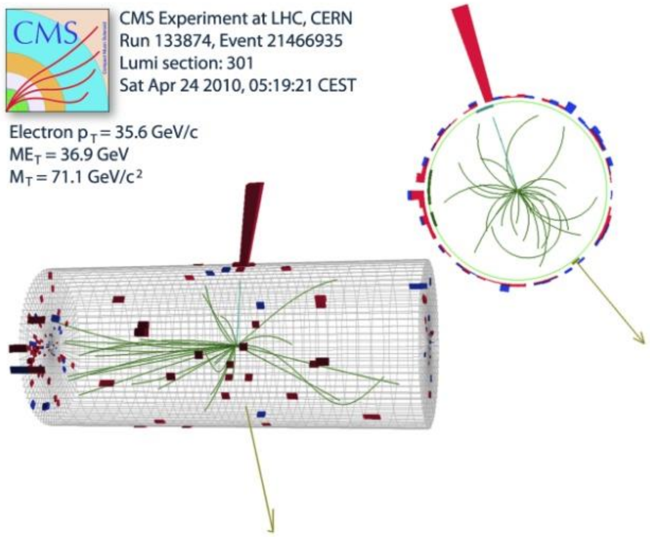
Large background from **QCD multijet events**

Strategy 1: Use simulated events to estimate background contribution

Systematic effects: luminosity, hadronization model, missing momentum model …



Missing transverse momentum



Analysis strategies

When designing an analysis, we try to consider the balance between systematic and statistical effects!
For example, lets consider cross-section measurements of $pp \rightarrow W \rightarrow e\nu$

Large background from **QCD multijet events**

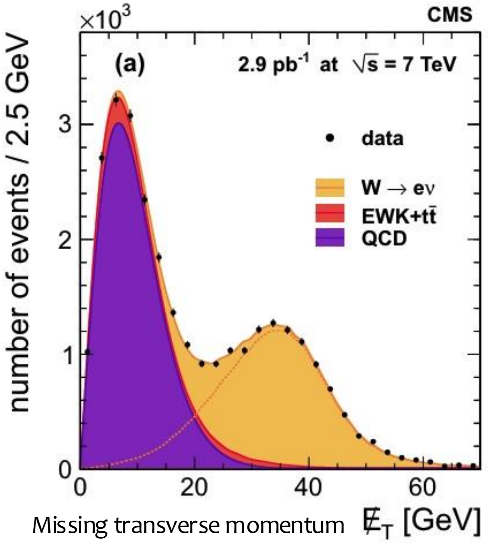
Strategy 1: Use simulated events to estimate background contribution

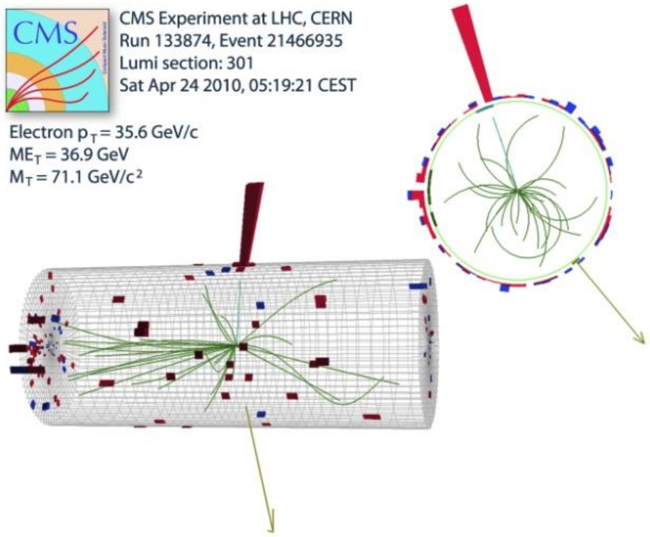
Systematic effects: luminosity, hadronization model, missing momentum model …

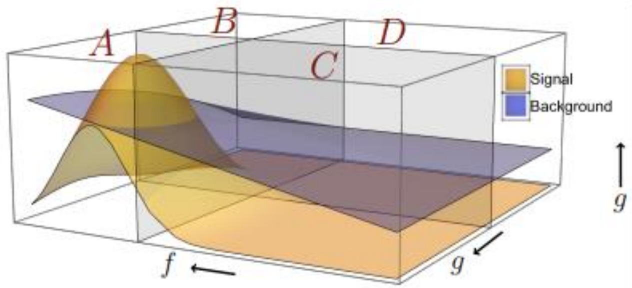
Strategy 2: Use ABCD method to estimate contribution from data

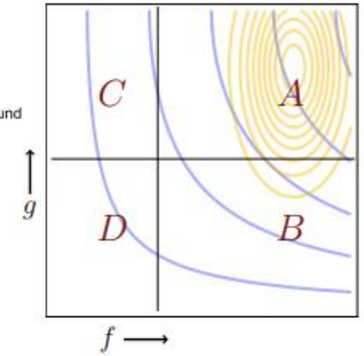
Systematic effects: ~~luminosity~~, ~~hadronization model~~, ~~missing momentum model,~~ limited events in data to estimate, correlation assumptions

Need to study total systematic uncertainties in different scenarios!









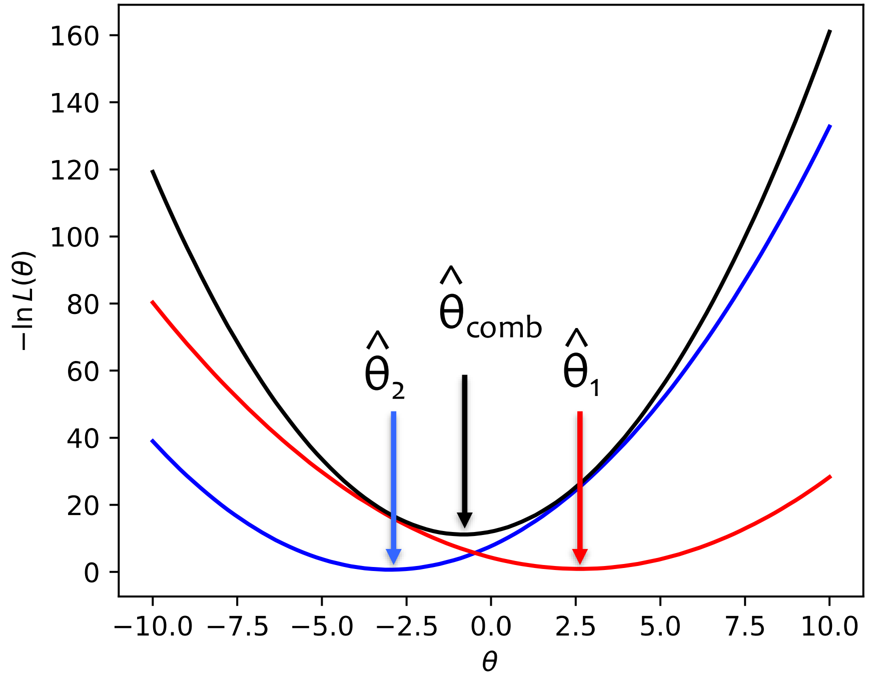
source

Warning about combining likelihoods

Remember that for N independent observations $X = \{X_1, X_2, \ldots, X_N\}$, the likelihood function is,

$$L(\theta) := \prod_{i=1}^{N} f_i(X_i;\theta),$$

This means $-\ln L(\theta) = -\sum_{i=1}^{N} \ln(f_i(X_i;\theta)) = \sum_{i=1}^{N} -\ln(L_i(\theta))$

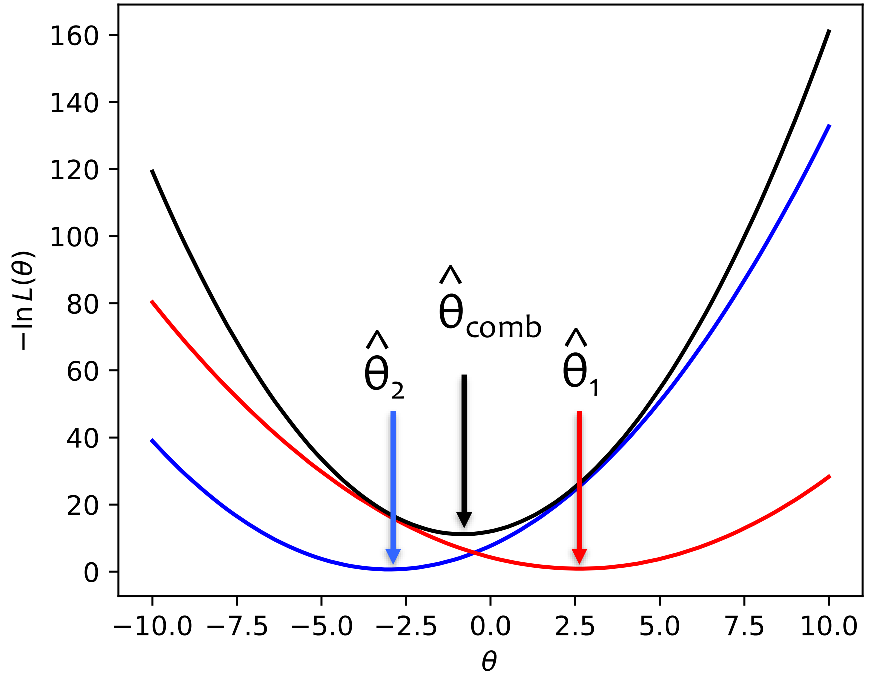


Warning about combining likelihoods

Remember that for N independent observations $X = \{X_1, X_2, \ldots, X_N\}$, the likelihood function is,

$$L(\theta) := \prod_{i=1}^{N} f_i(X_i; \theta),$$

This means $$-\ln L(\theta) = -\sum_{i=1}^{N} \ln(f_i(X_i;\theta)) = \sum_{i=1}^{N} -\ln(L_i(\theta))$$



We can sum the negative log likelihood curves to obtained the **combined negative log likelihood** → measurements can be easily combined

$$-\ln L_{\mathrm{comb}}(\theta) = -\ln L_1(\theta) - \ln L_2(\theta)$$

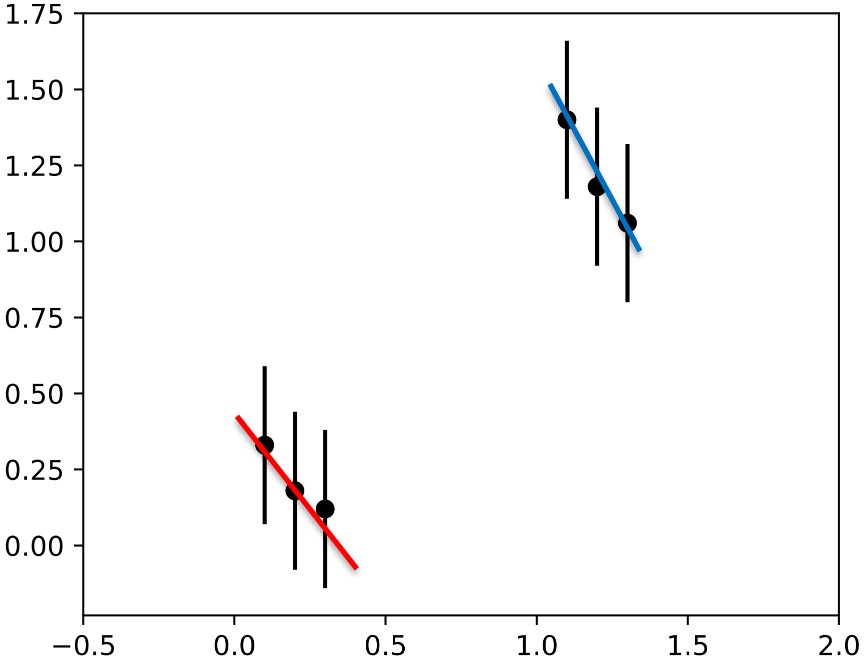
Warning about combining likelihoods

This is **not true** for **profiled likelihoods**!

$$-\ln L_{\mathrm{comb}}(\mu,\hat{\nu}(\mu)) \neq -\ln L_{1}(\mu,\hat{\nu}(\mu)) - \ln L_{2}(\mu,\hat{\nu}(\mu))$$

Imaging fitting a straight line to the points

$$y = mx + c$$



Warning about combining likelihoods

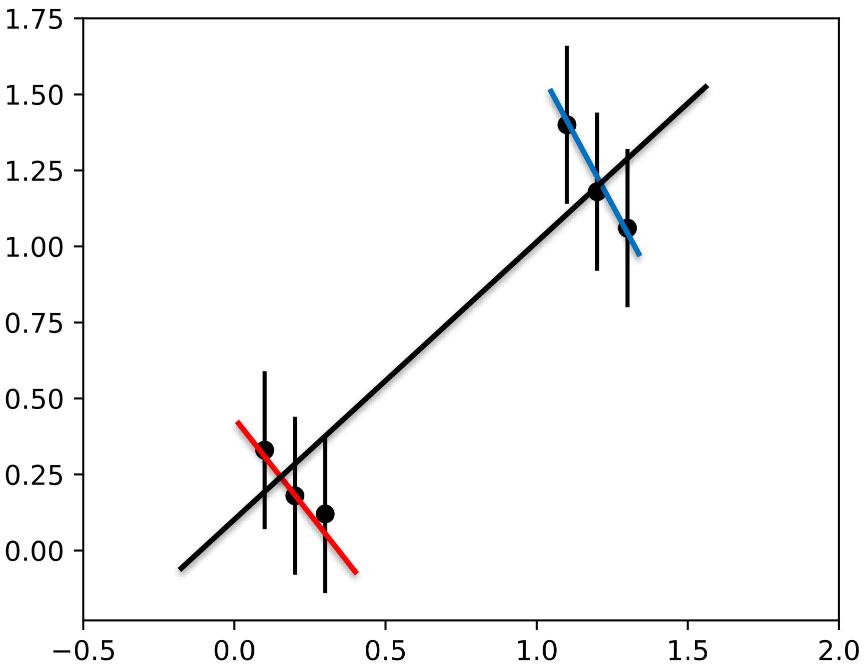
This is **not true** for **profiled likelihoods**!

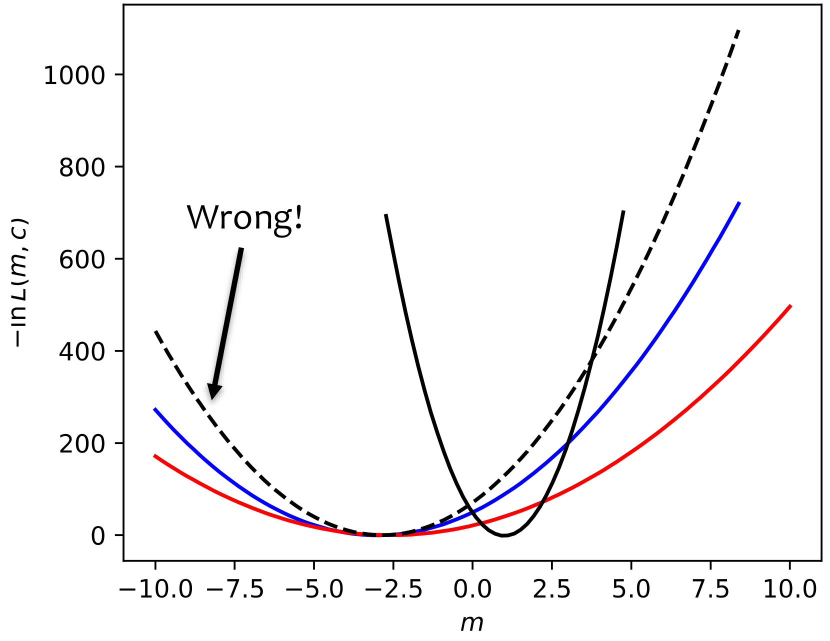
$$-\ln L_{\text{comb}}(\mu, \hat{\nu}(\mu)) \neq -\ln L_1(\mu, \hat{\nu}(\mu)) - \ln L_2(\mu, \hat{\nu}(\mu))$$

Imaging fitting a straight line to the points

$$y = mx + c$$

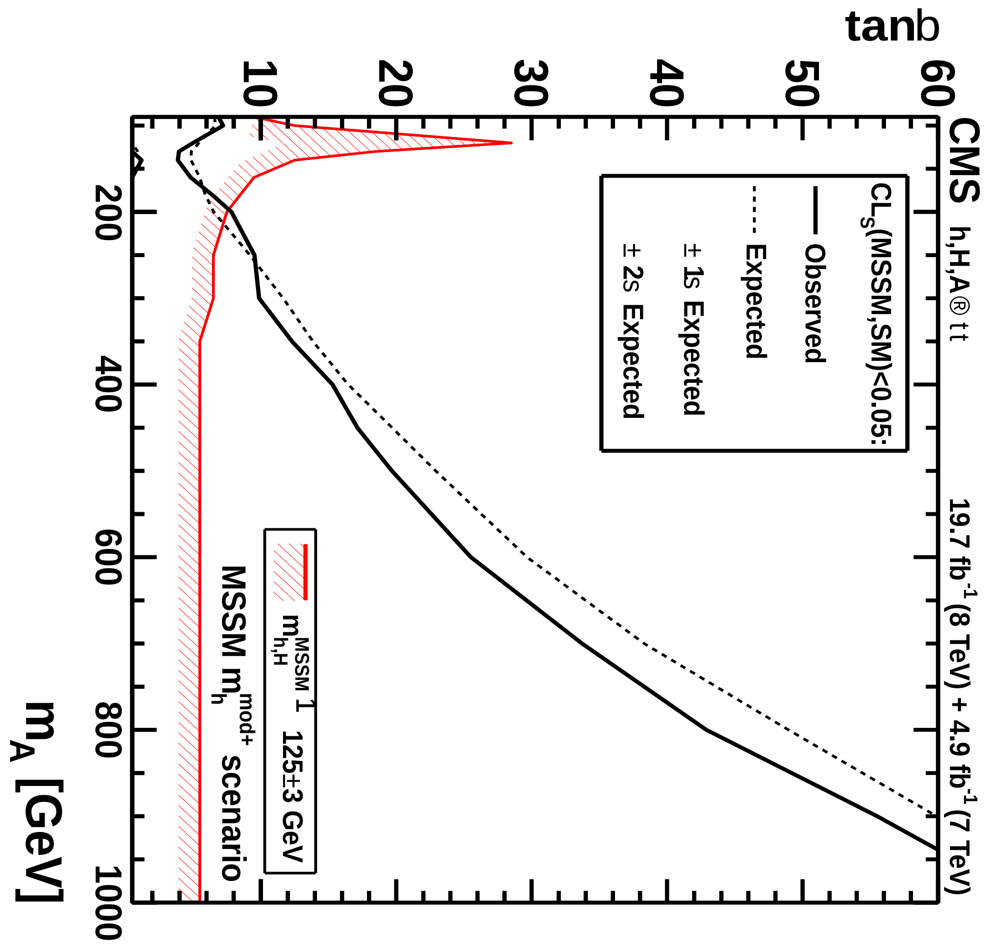
The profiled likelihoods as a function of the slope **do not sum to give the correct combined profiled likelihood!**





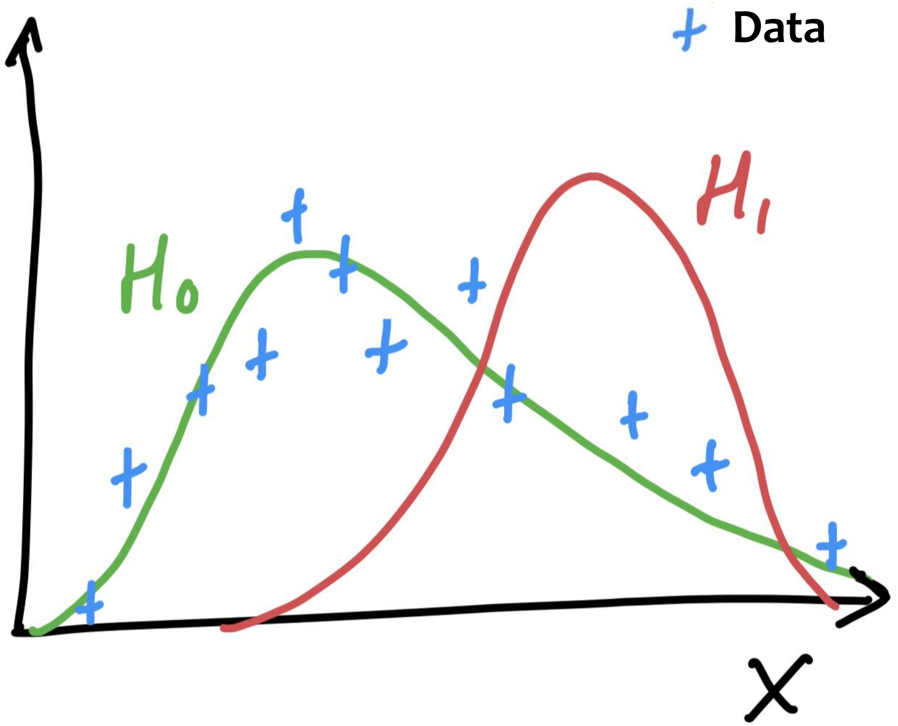
Hypothesis testing

Now we know how measurements are made, what about results like this one?



Hypothesis testing

Suppose that we have to choose between two hypotheses labelled H0 and H1 . We typically distinguish the two as;



H_0 := the ***null*** hypothesis
H_1 := the ***alternate*** hypothesis

Example : H_0 = Standard Model, H_1 = Supersymmetry

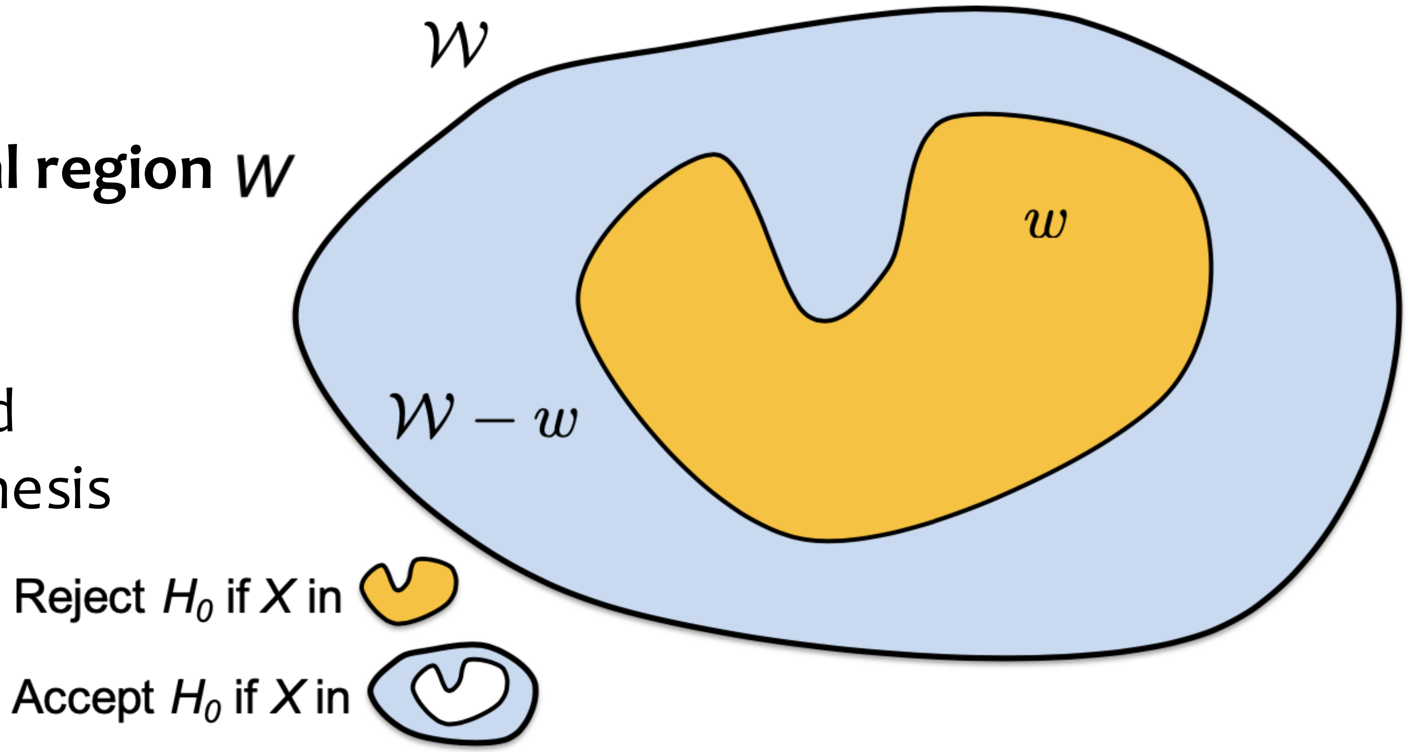
X is a function of the experimental observations which is supposed to summarize the observations – this is known as a ***test statistic***.

Type-I/Type-II Errors

Suppose then that we have our chosen test statistic $X \in \mathcal{W}$

We divide this region $\mathcal{W}$ into a **critical region** w and a **region of acceptance** $\mathcal{W} - w$

Observations of X falling into w would lead us to believe that our null hypothesis is **not true.**

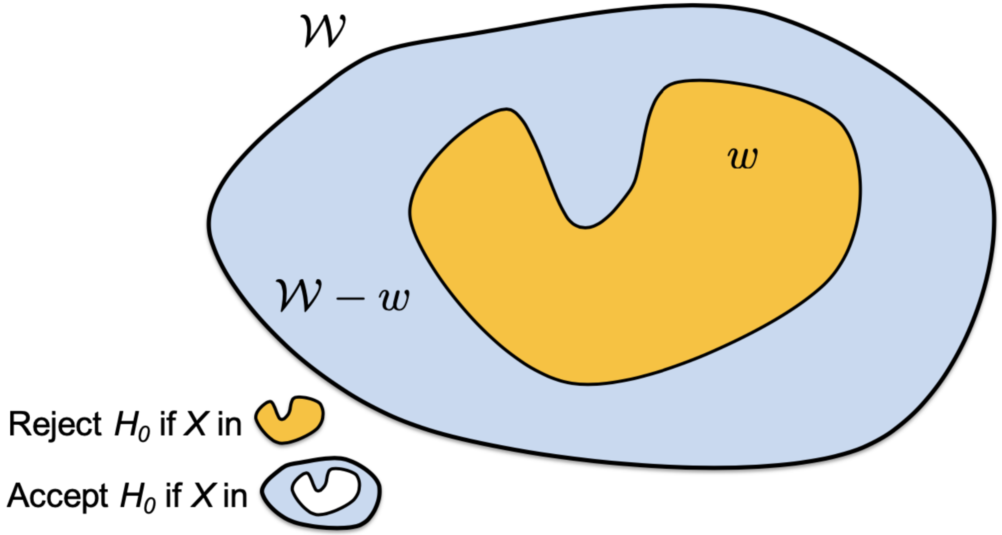


Defining a test of H_0 , given we've decided on our test statistic, then becomes choosing a critical region w

Type-I/Type-II Errors

Type-I Error : In practise, we often tune the critical region so as to obtain a particular probability (known as the size of the test) α that X falls into the critical region when H_0 is true (we usually say "under H_0")

$$P(X \in w | H_0) = \alpha$$

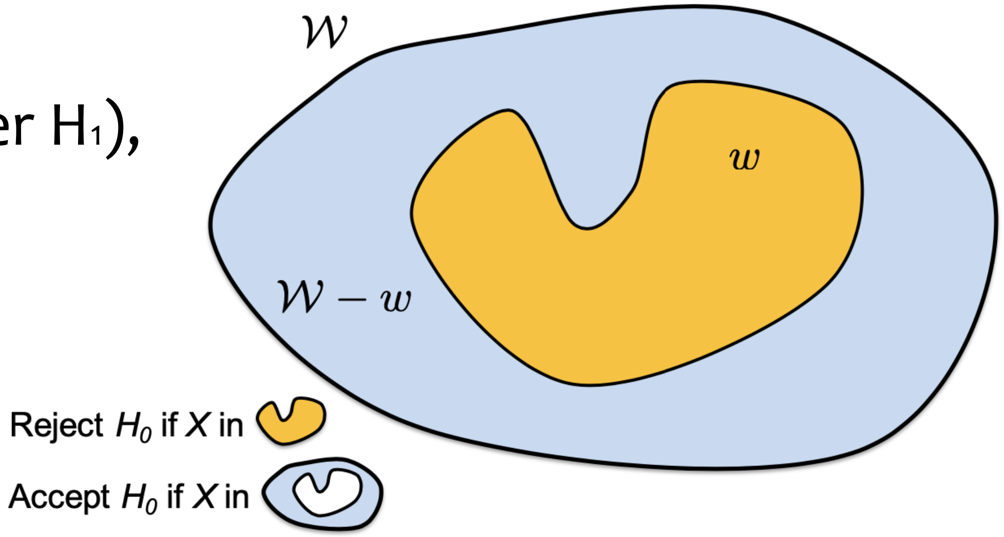


You can see then that α is exactly **the probability to reject the null hypothesis if the null hypothesis is true**

Type-I/Type-II Errors

Type-II Error: Of course, we also want to know how useful a test is at discriminating against the alternate hypothesis. This is known as the ***power of the test***, and is defined as the probability of X falling into the critical region if H_1 is true (under H_1),

$$P(X \in w | H_1) = 1 - \beta$$

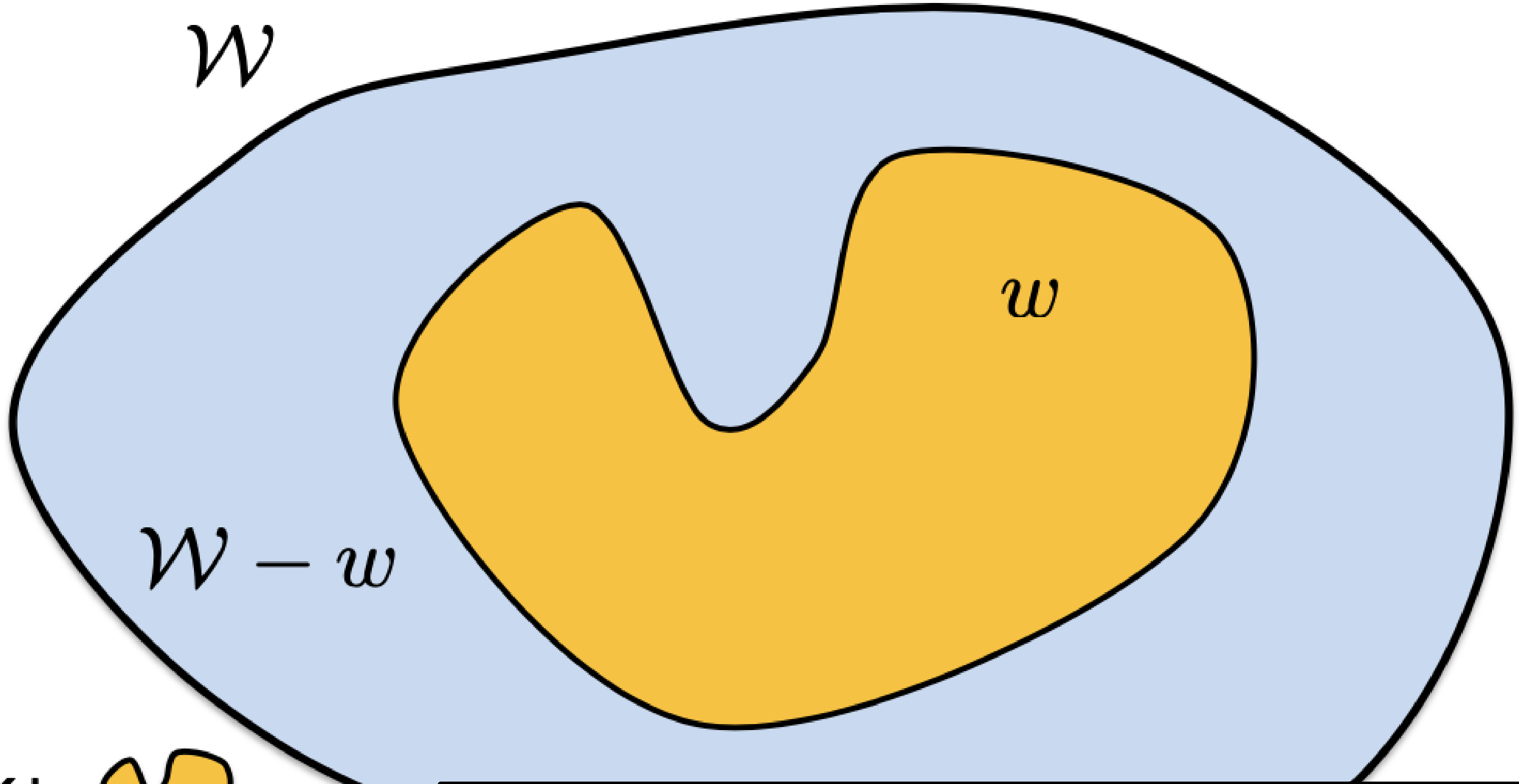


Clearly this is related to the probability that X falls into the acceptance region via

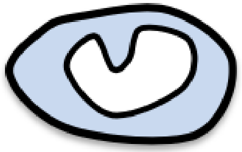
$$P(X \in \mathcal{W} - w | H_1) = 1 - P(X \in w | H_1) = \beta.$$

Then **β** is the probability that we would **accept the null hypothesis when the alternative hypothesis is true**

Type-I/Type-II Errors



Reject H_0 if X in 

Accept H_0 if X in 

$P(X \in w | H_0) = \alpha$ → How likely to *falsely reject* H_0

$P(X \in \mathcal{W} - w | H_1) = \beta$ → How likely to *falsely accept* H_0

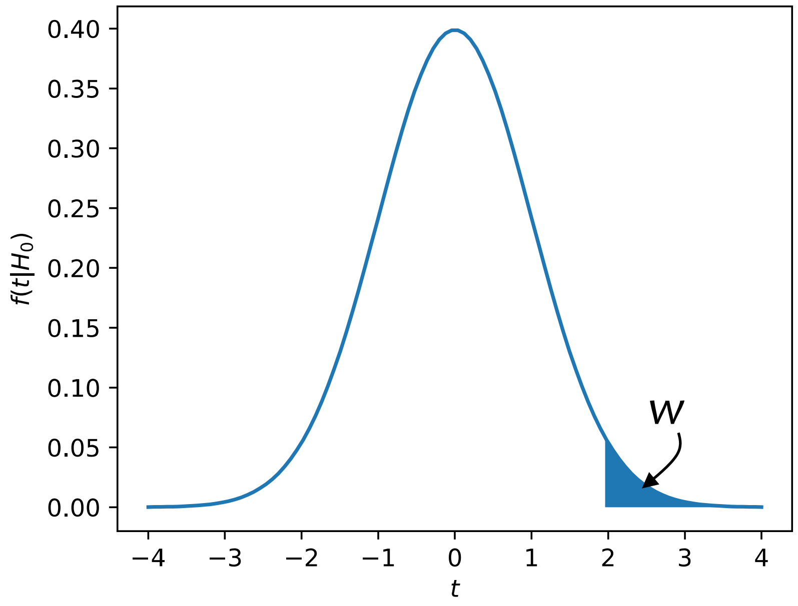
Hypothesis tests

There are a huge number of hypothesis tests on the market for use in various problems, however they generally follow the same routine;

1. Define a test (summary) statistic $t \in \mathbb{R}$ that summarizes the observations and has some separation between H_0 and H_1
2. Define a critical region w such that

$$\int_w f(t|H_0) = \alpha$$

   where α is a predefined value between 0 and 1.



Hypothesis tests

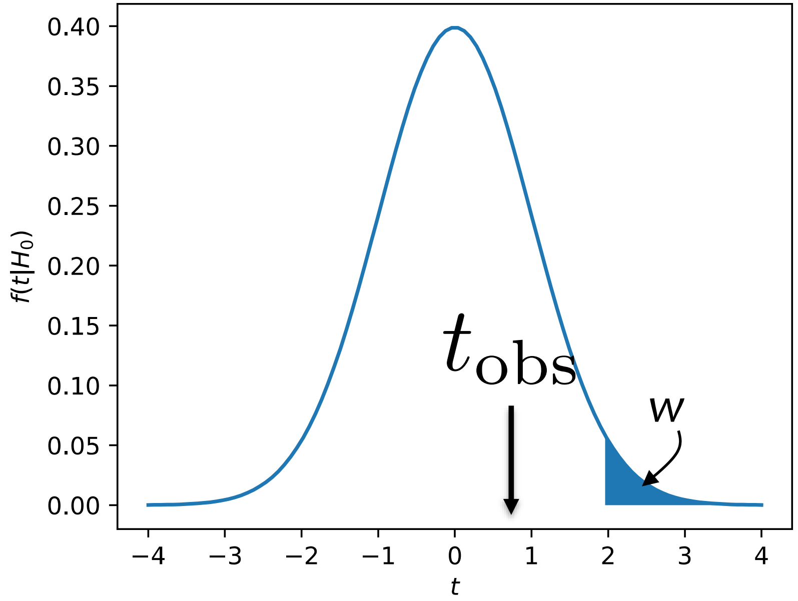
There are a huge number of hypothesis tests on the market for use in various problems, however they generally follow the same routine;

1. Define a test (summary) statistic $t \in \mathbb{R}$ that summarizes the observations and has some separation between H_0 and H_1
2. Define a critical region *w* such that

$$\int_w f(t|H_0) = \alpha$$

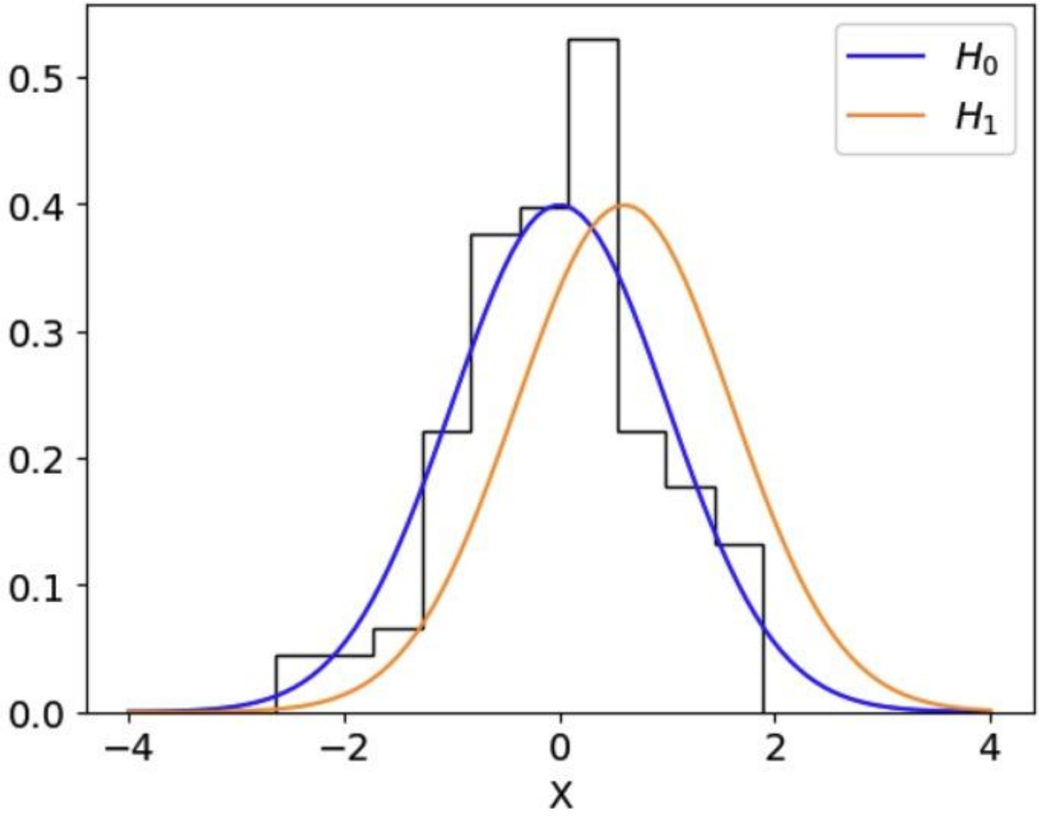
   where α is a predefined value between 0 and 1.

3. Determine the value of *t* in the observed data, t_{obs}
4. Reject H_0 if $t_{\mathrm{obs}} \in w$



Example – Student's t-test

The Student's t-test is a simple hypothesis test based on the expectation values of distributions



$$t = \sqrt{\frac{N}{\bar{V}}}\,(\bar{X} - \mu)$$

Small and large values of t indicate that the data has a significantly different expectation compared to H_0

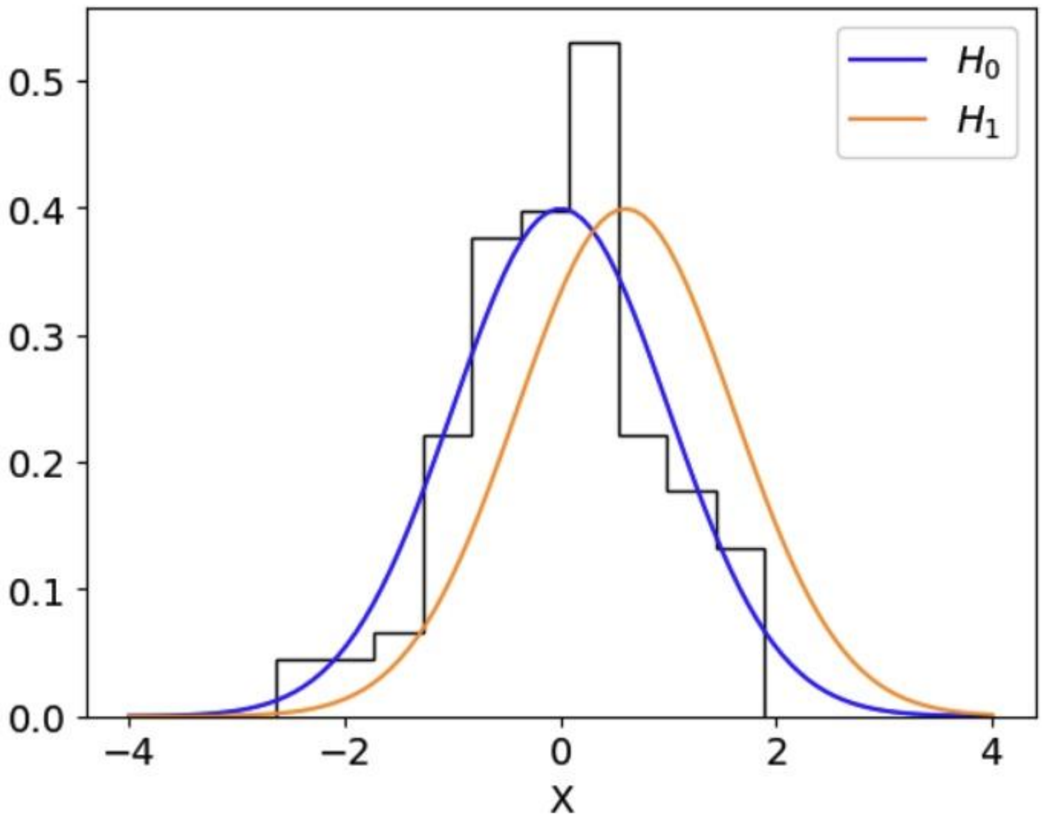
Example – Student's t-test

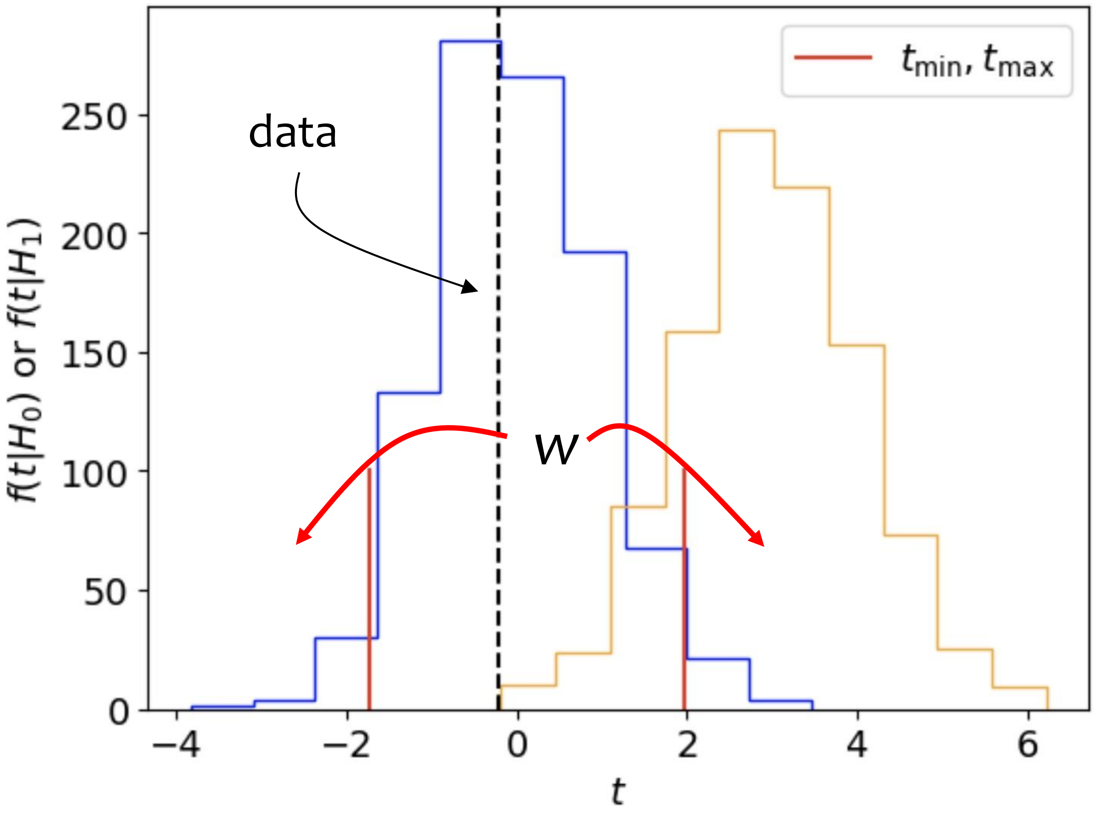
Choosing α = 0.05, we can define t_{min} and t_{max} such that

$$t = \sqrt{\frac{N}{\bar{V}}}\,(\bar{X} - \mu)$$

μ=0

$$\int_{-\infty}^{t_{min}} f(t|H_0)dt = \int_{t_{max}}^{+\infty} f(t|H_0)dt = 0.025$$





Power will depend on H_1

The most powerful test

For a fixed value of α can we find the test that gives the best power $(1\text{-}\beta)$?

The most powerful test

For a fixed value of α can we find the test that gives the best power $(1-\beta)$?
Yes → Neyman-Pearson lemma

Let's think about some observed data **X** and suppose it has a probability distribution function $f(\mathbf{X};\theta)$,
where θ is used to represent our hypotheses:

$\theta = \theta_0$ represents the null hypothesis H_0

$\theta = \theta_1$ represents the alternate hypothesis H_1

What I mean is $f(X|H(\theta)) = f(X;\theta)$

The most powerful test

For a fixed value of α can we find the test that gives the best power (1-β) ?
Yes → Neyman-Pearson lemma

Let's think about some observed data **X** and suppose it has a probability distribution function f(**X**;θ), where θ is used to represent our hypotheses:

$\theta = \theta_0$ represents the null hypothesis H_0

$\theta = \theta_1$ represents the alternate hypothesis H_1

What I mean is $f(X|H(\theta)) = f(X;\theta)$

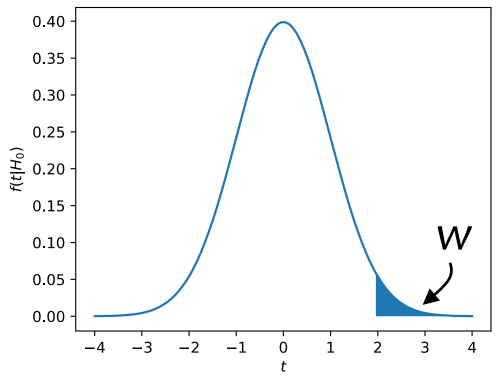
For a specific size of test α we choose w such that,

$$\int_w f(\mathbf{X};\theta_0)d\mathbf{X} = \alpha$$

and we want to find the region w which maximises 1 – β

$$1-\beta \ = \ \int_w f(\mathbf{X};\theta_1)d\mathbf{X}$$

The most powerful test



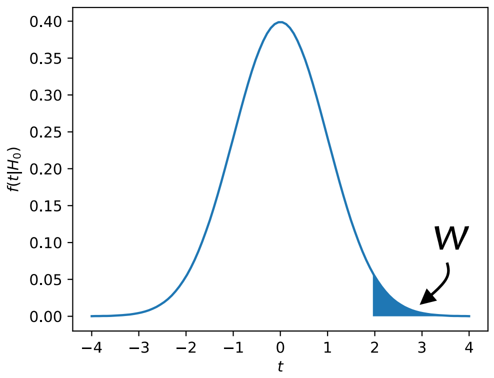
$$1-\beta = \int_w f(\mathbf{X};\theta_1)\,d\mathbf{X}$$

$$= \int_w \frac{f(\mathbf{X};\theta_1)}{f(\mathbf{X};\theta_0)} f(\mathbf{X};\theta_0)\,d\mathbf{X}$$

Expectation value in a restricted space *w*

$$= E\left[\frac{f(\mathbf{X};\theta_1)}{f(\mathbf{X};\theta_0)}\middle|\theta=\theta_0\right]_w$$

The most powerful test



Define $$\Lambda = \frac{f(\mathbf{X};\theta_1)}{f(\mathbf{X};\theta_0)} = \frac{L(\theta_1)}{L(\theta_0)}$$

$$1-\beta = \int_w f(\mathbf{X};\theta_1)d\mathbf{X}$$

$$= \int_w \frac{f(\mathbf{X};\theta_1)}{f(\mathbf{X};\theta_0)} f(\mathbf{X};\theta_0)d\mathbf{X}$$

Expectation value in a restricted space *w*

$$= E\left[\frac{f(\mathbf{X};\theta_1)}{f(\mathbf{X};\theta_0)}\middle|\theta=\theta_0\right]_w$$

This quantity is the **ratio of the likelihood function**, evaluated under the two hypotheses

1-β will be maximal when *w* is chosen to contain the largest values of Λ

→ The best critical region is the set of points for which $\Lambda \geq c_\alpha \in \mathbb{R}$, where c_α satisfies

$$\int_w f(\mathbf{X};\theta_0)d\mathbf{X} = \alpha$$

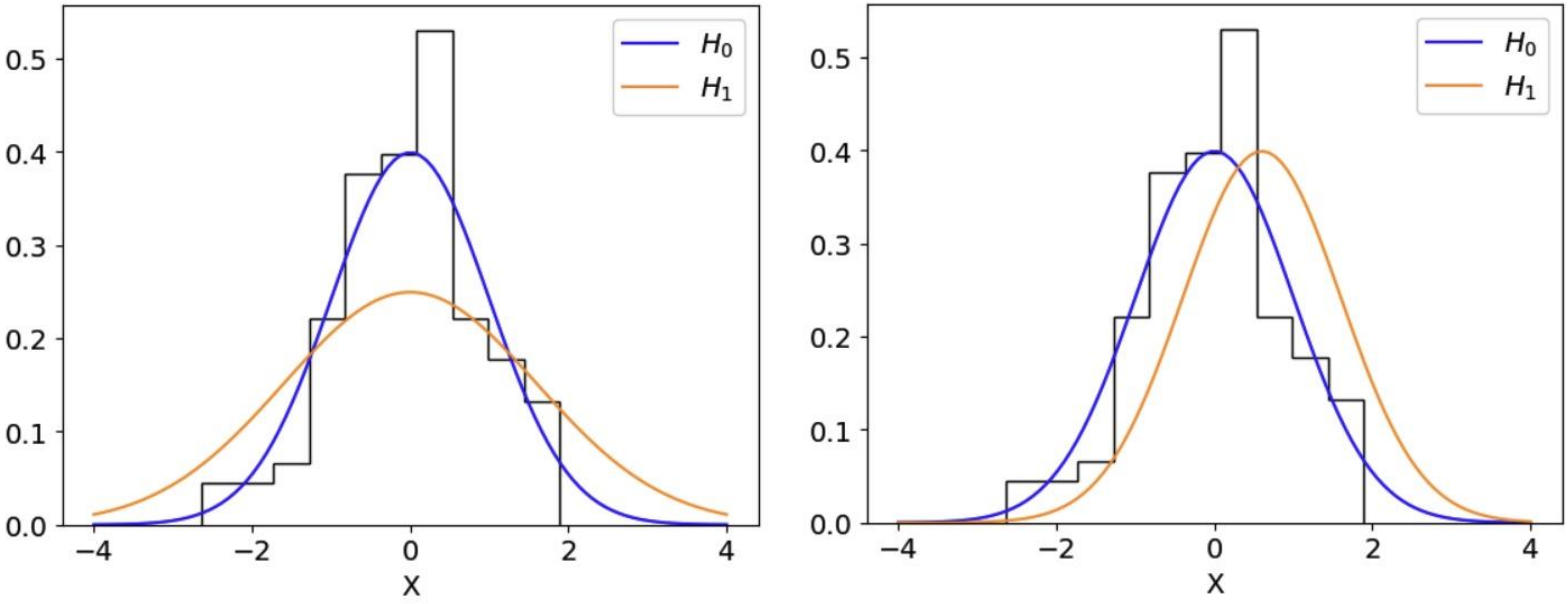
The likelihood ratio is the most powerful test

(fact used very often in Machine Learning!)

→ If $\Lambda > c_\alpha$, we would choose H_1, while $\Lambda \leq c_\alpha$ leads us to choose H_0!

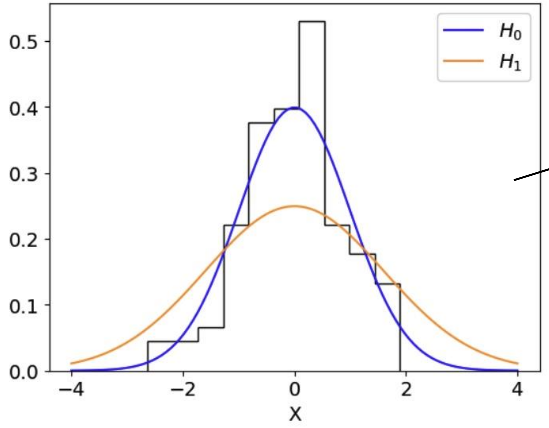
Example: Gaussian distributions

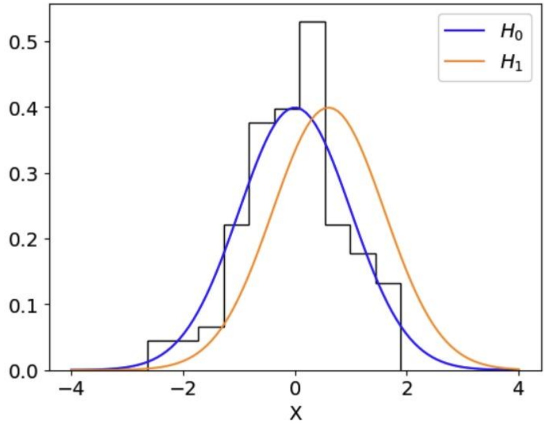
H_0 and H_1 are both Gaussian distributions with the same mean (μ) but different width (σ) or vice-versa

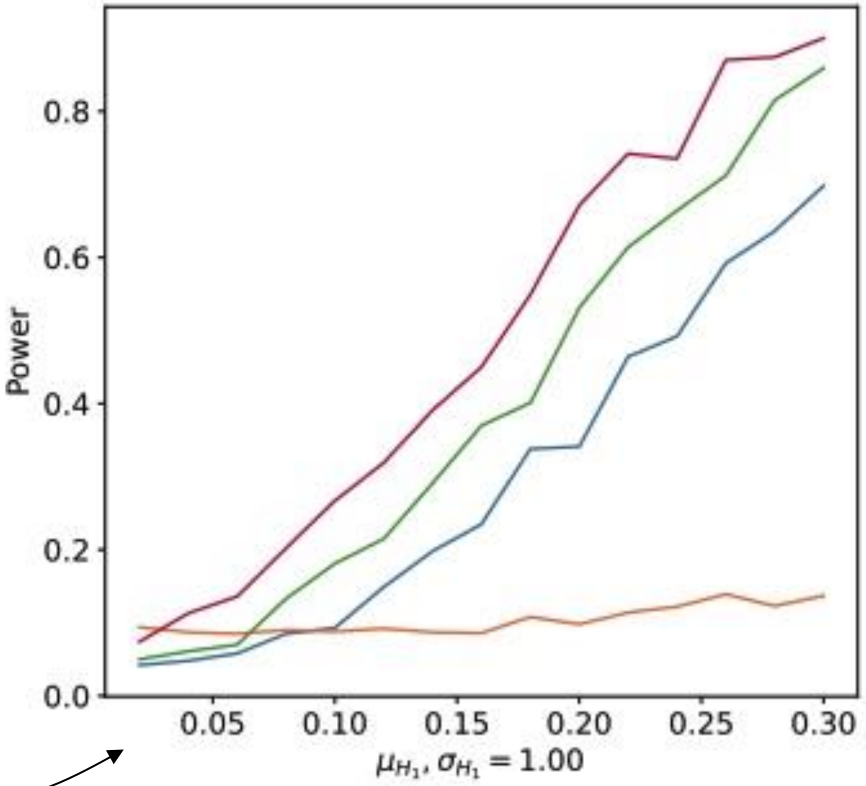


Example: Gaussian distributions

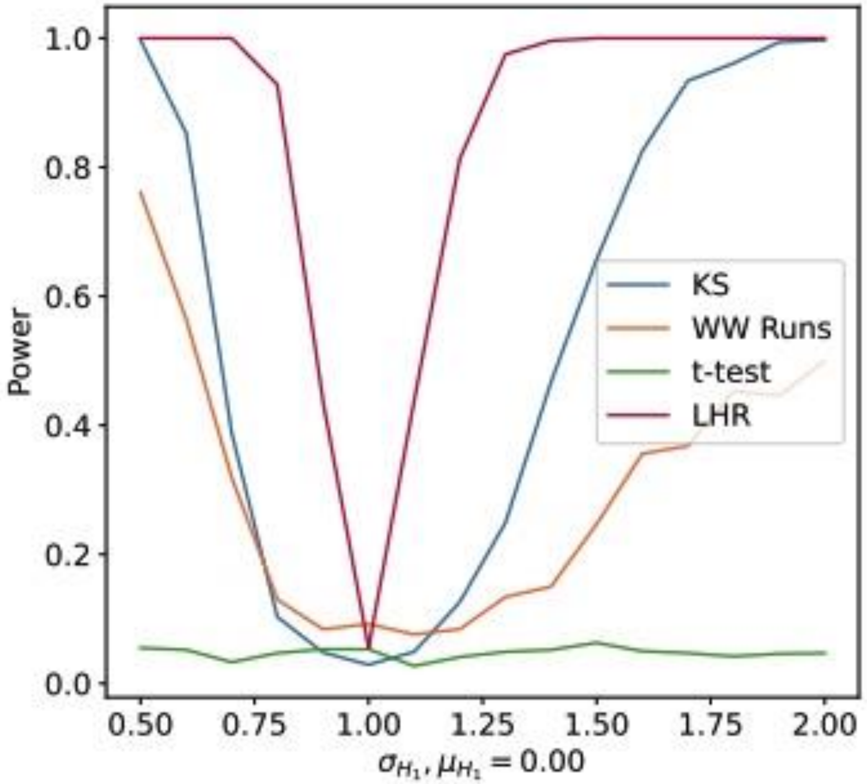
H_0 and H_1 are both Gaussian distributions with the same mean (μ) but different width (σ)







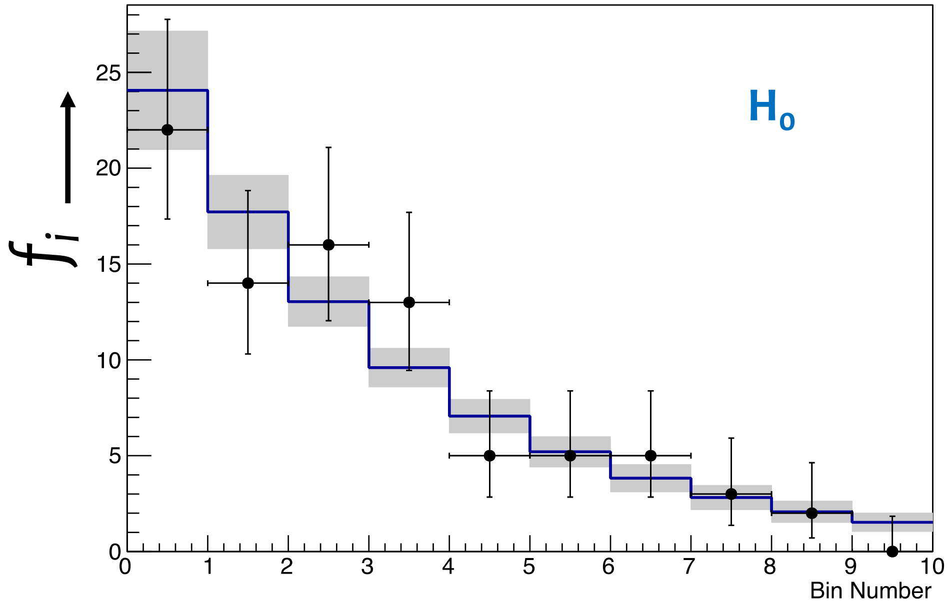
α=0.05



Profiled likelihood ratio based test

Compatibility of data and prediction (H_0) in distributions

$$x = -2\ln\frac{L(H_0)}{L(H_1)}$$



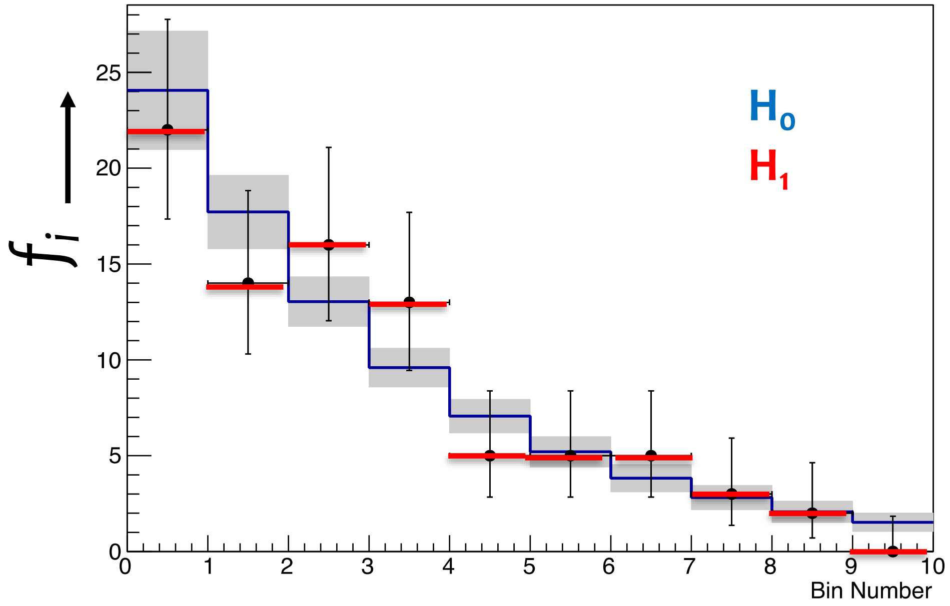
Profiled likelihood ratio based test

Compatibility of data and prediction (H_0) in distributions

Make use of the **saturated likelihood as alternative H_1**

→ Best possible fit to data

$$x = 2\sum_i f_i(\hat{\boldsymbol{\theta}}) - d_i + d_i \ln \frac{d_i}{f_i(\hat{\boldsymbol{\theta}})}$$



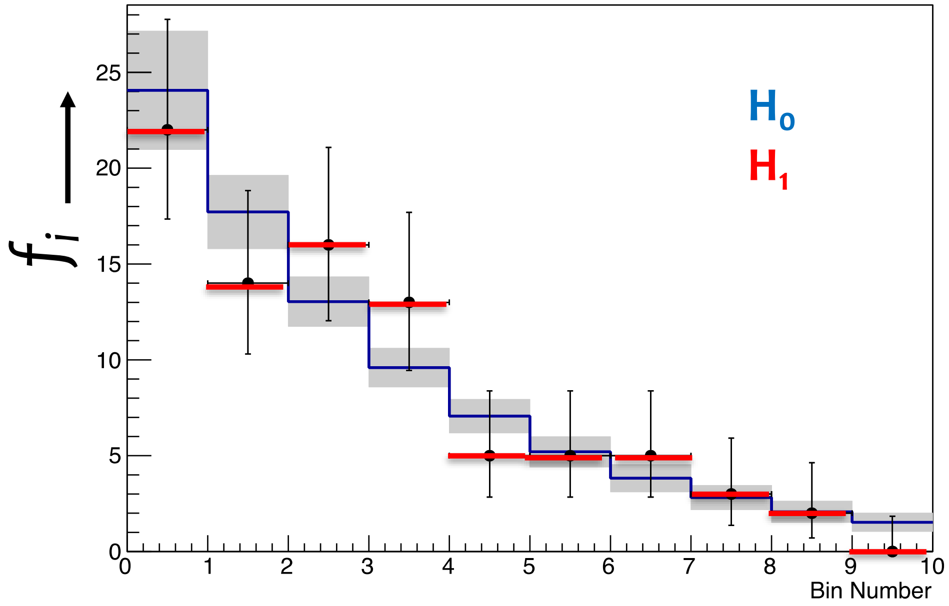
Profiled likelihood ratio based test

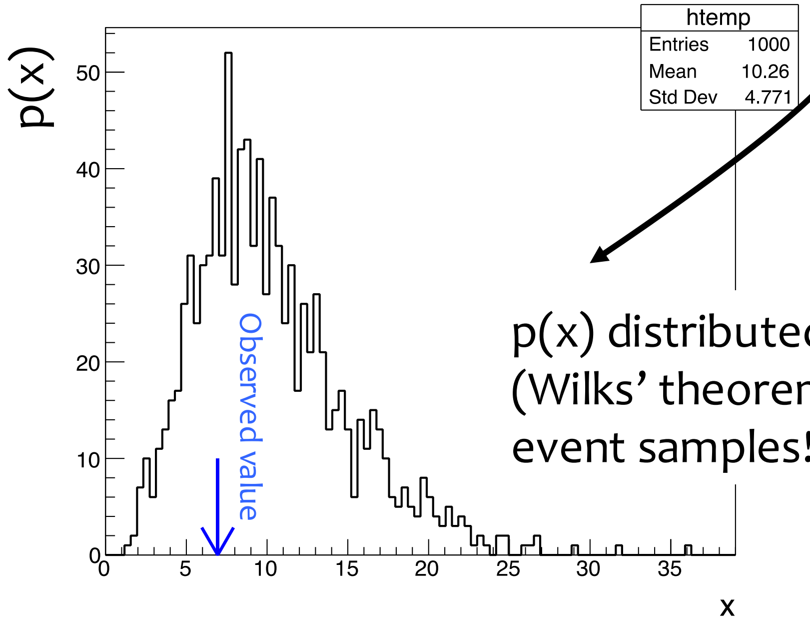
Compatibility of data and prediction (H_0) in distributions

Make use of the **saturated likelihood as alternative H_1**
→ Best possible fit to data

$$x = 2\sum_i f_i(\hat{\boldsymbol{\theta}}) - d_i + d_i \ln \frac{d_i}{f_i(\hat{\boldsymbol{\theta}})}$$

Generate pseudo-data to calculate p-value



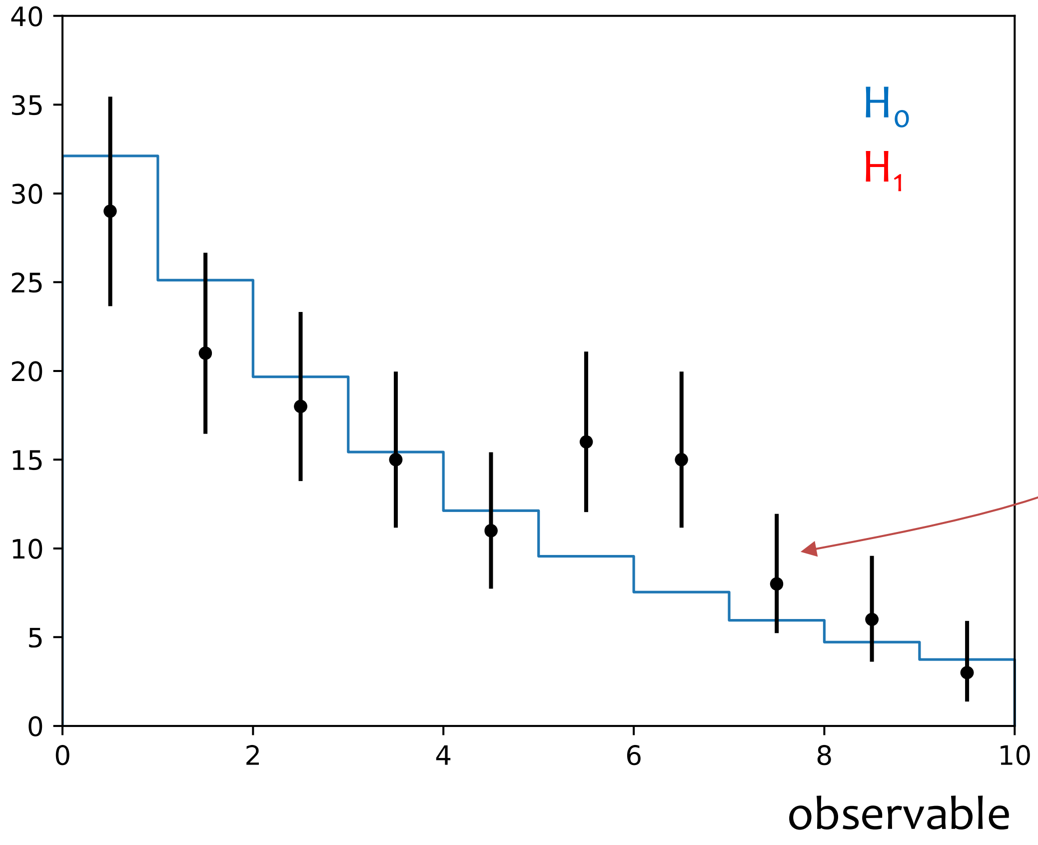


p(x) distributed as $\chi^2(n)$ (Wilks' theorem) for large event samples!

Searches for new physics

Let's imagine we are searching for a new particle X:

- Our null hypothesis H_0 is the standard model - sometimes called "background only"
- The alternate hypothesis H_1 is the standard model + the new particle - or signal + background



Cross-section parameterized with signal strength

$$\sigma A\epsilon L = \mu\sigma(pp \to X)A'\epsilon' L$$

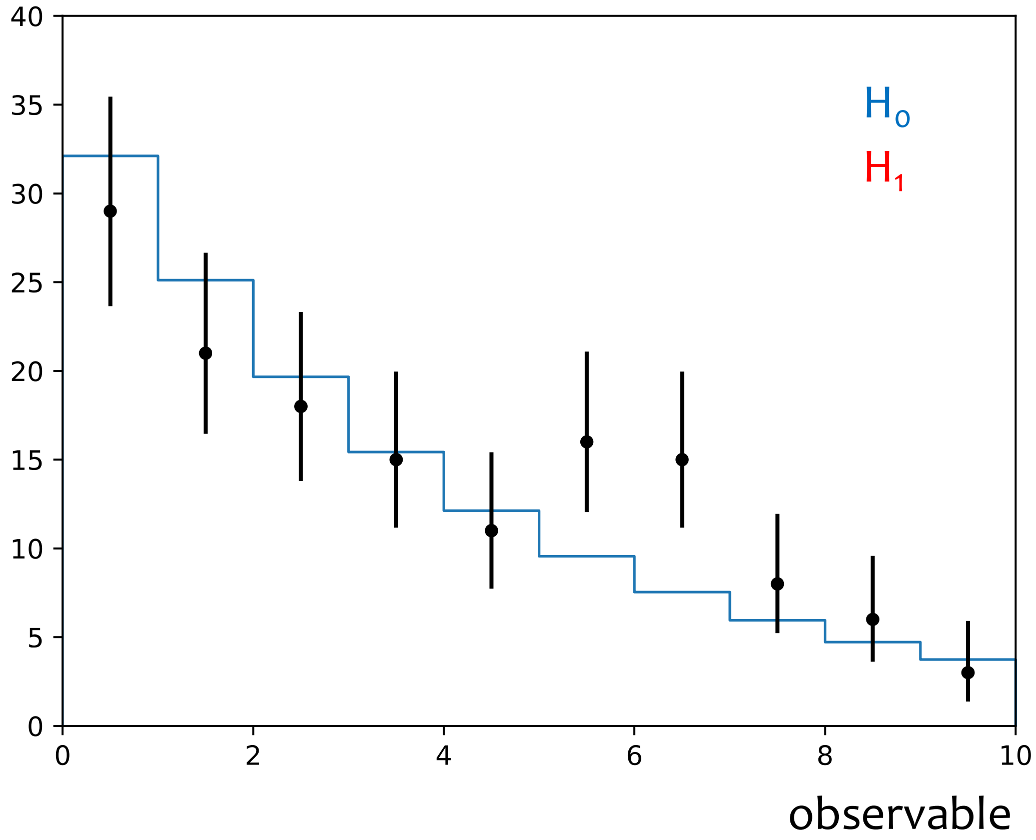
If we assume $A'\epsilon' = A\epsilon$, then we can think of

$$\mu = \frac{\sigma}{\sigma(pp \to X)}$$

Searches for new physics

Let's imagine we are searching for a new particle X:

- Our null hypothesis H_0 is the standard model - sometimes called "background only"
- The alternate hypothesis H_1 is the standard model + the new particle - or signal + background



The test-statistic we use

$$q_0 = \begin{cases} -2\ln\left(\frac{\mathcal{L}(0,\hat{\hat{\nu}}(0))}{\mathcal{L}(\hat{\mu},\hat{\nu})}\right) & \text{if } \hat{\mu} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

Where μ is the signal strength (μ=0 → background only)

Searches for new physics

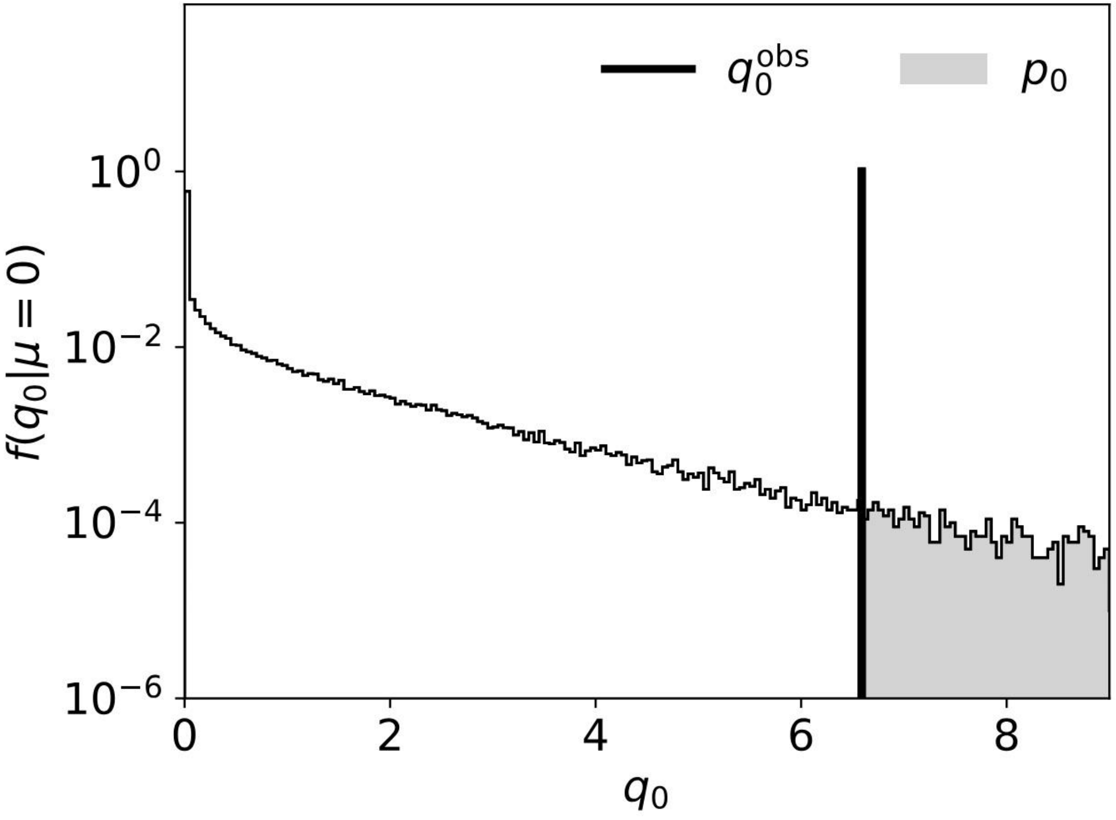
Let's imagine we are searching for a new particle X:

- Our null hypothesis H_0 is the standard model - sometimes called "background only"
- The alternate hypothesis H_1 is the standard model + the new particle - or signal + background

$$q_0 = \begin{cases} -2\ln\left(\frac{\mathcal{L}(0,\hat{\hat{\vec{\nu}}}(0))}{\mathcal{L}(\hat{\mu},\hat{\vec{\nu}})}\right) & \text{if } \hat{\mu} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

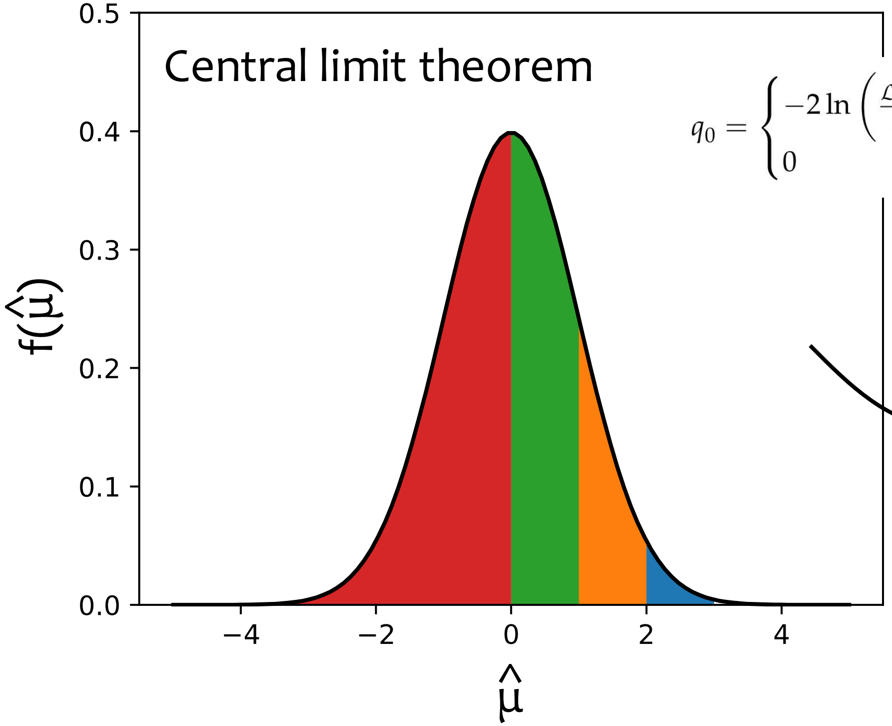
We calculate a p-value using the distribution under the null

$$p_0 = \int_{q_0^{\mathrm{obs}}}^{\infty} f(q_0|0)\,\mathrm{d}q_0.$$



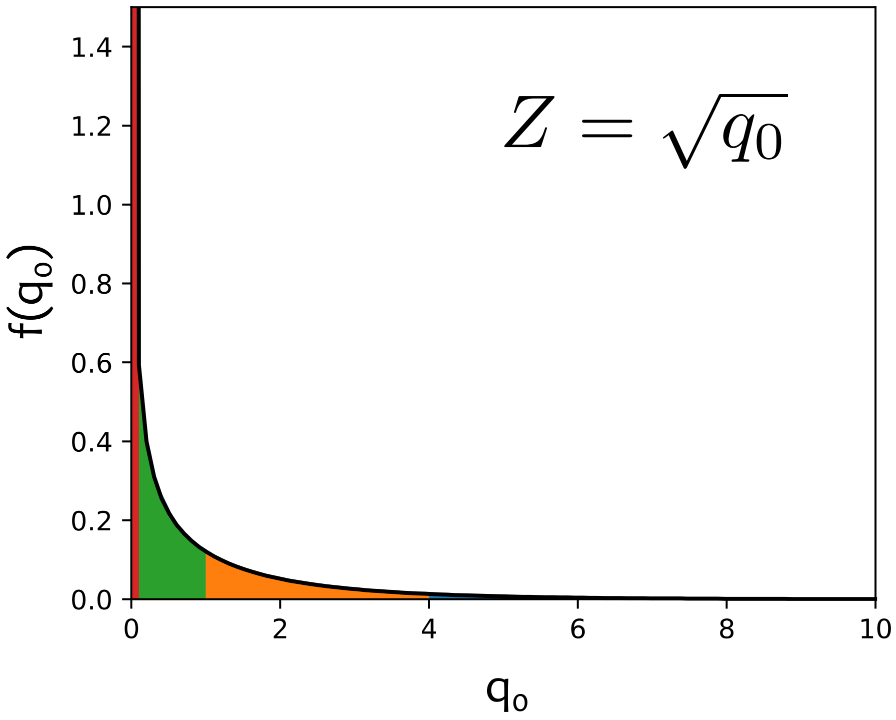
Searches for new physics

In the large statistics limit, the distribution of the test statistic is known (see [Eur.Phys.J.C71:1554,2011])



$$q_0 = \begin{cases} -2\ln\left(\frac{\mathcal{L}(0,\hat{\hat{\nu}}(0))}{\mathcal{L}(\hat{\mu},\hat{\nu})}\right) & \text{if } \hat{\mu} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

We convert p-values into significances (Z-score) through simple formula

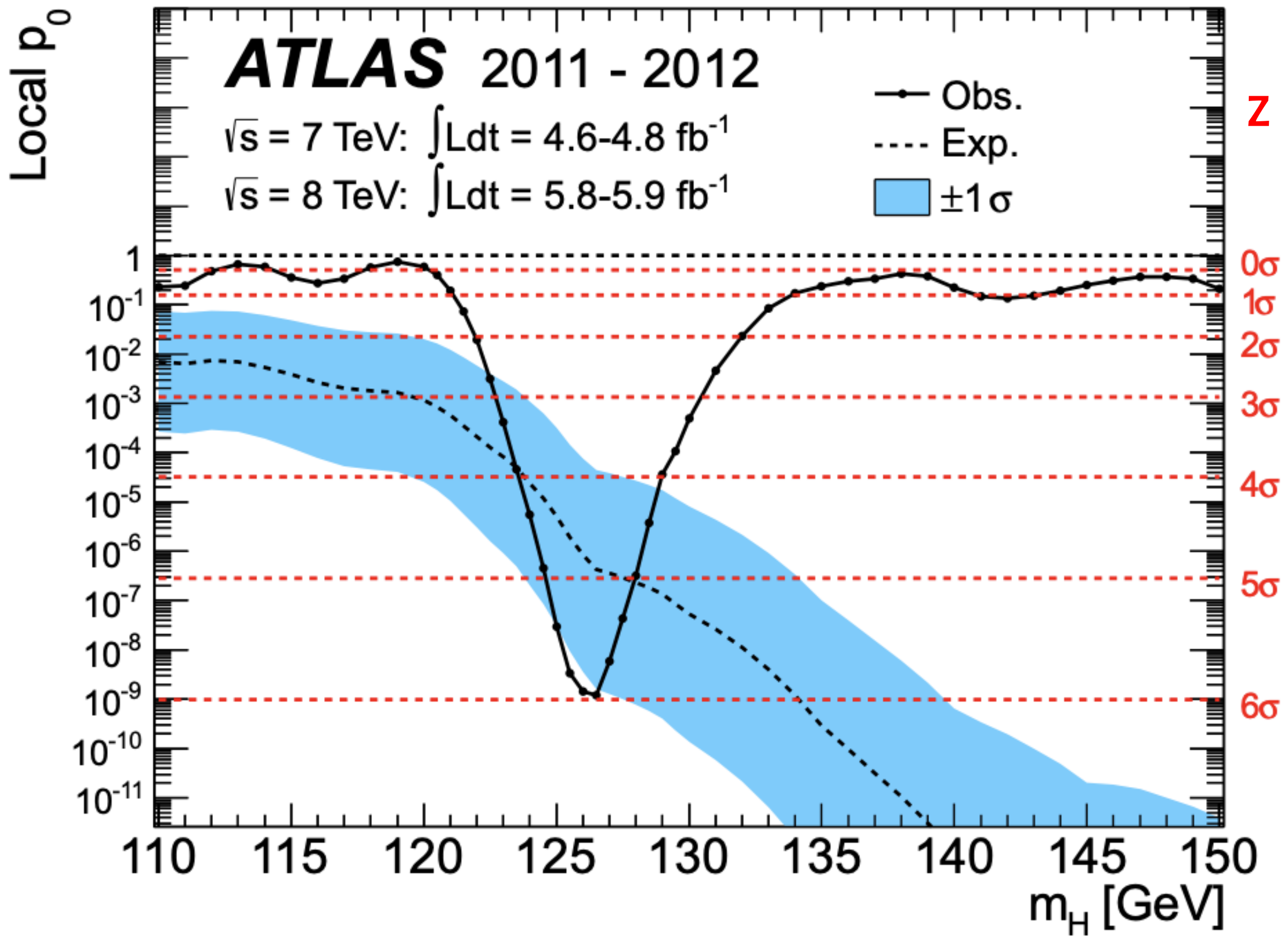


Searches for new physics

If p_0 very small or Z large we reject H_0 → discovery of new physics!

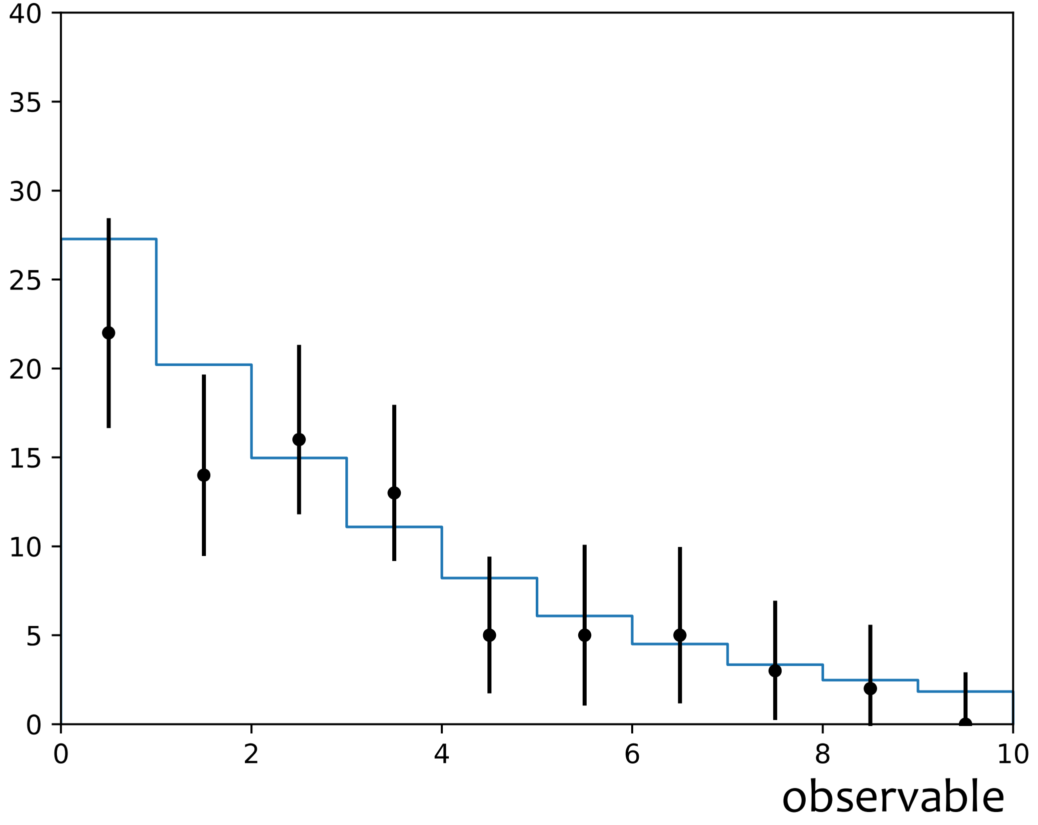
Phys. Lett. B 716 (2012)

This is exactly how we discovered the Higgs boson in 2012!



Upper limits

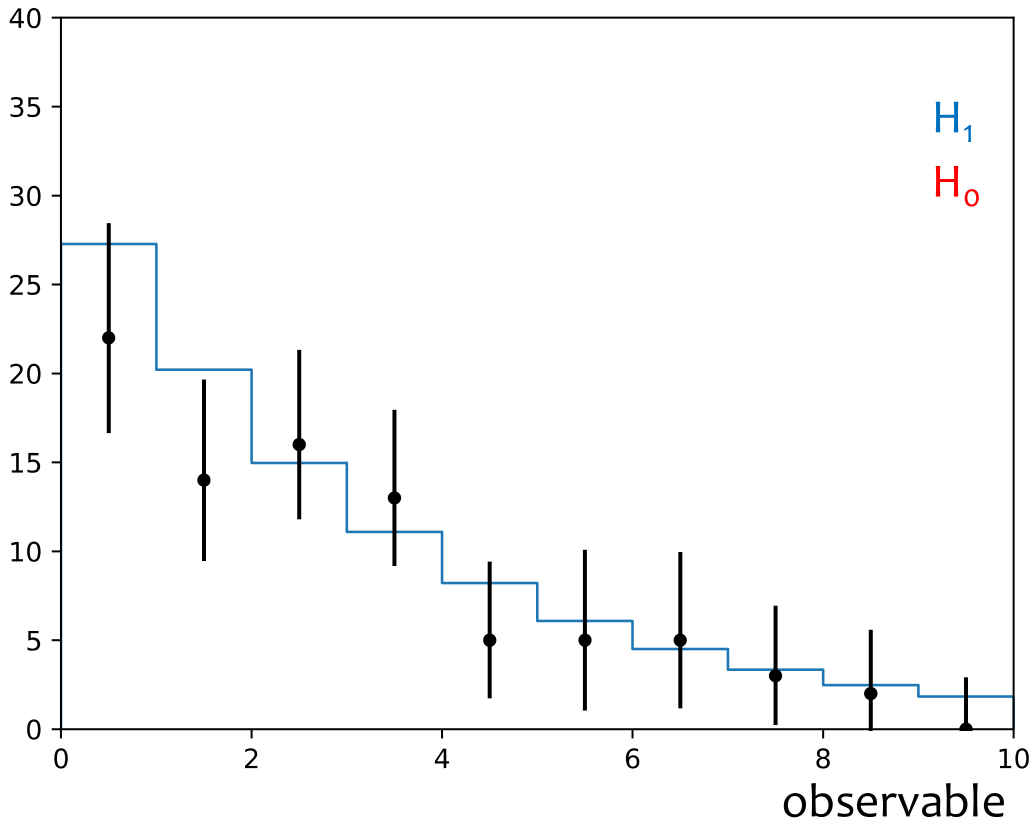
What about if we don't see any excess in the data?



Upper limits

What about if we don't see any excess in the data? → We invert the hypotheses

- Our null hypothesis H_0 is the standard model + new particle - sometimes called "background only"
- The alternate hypothesis H_1 is the standard model



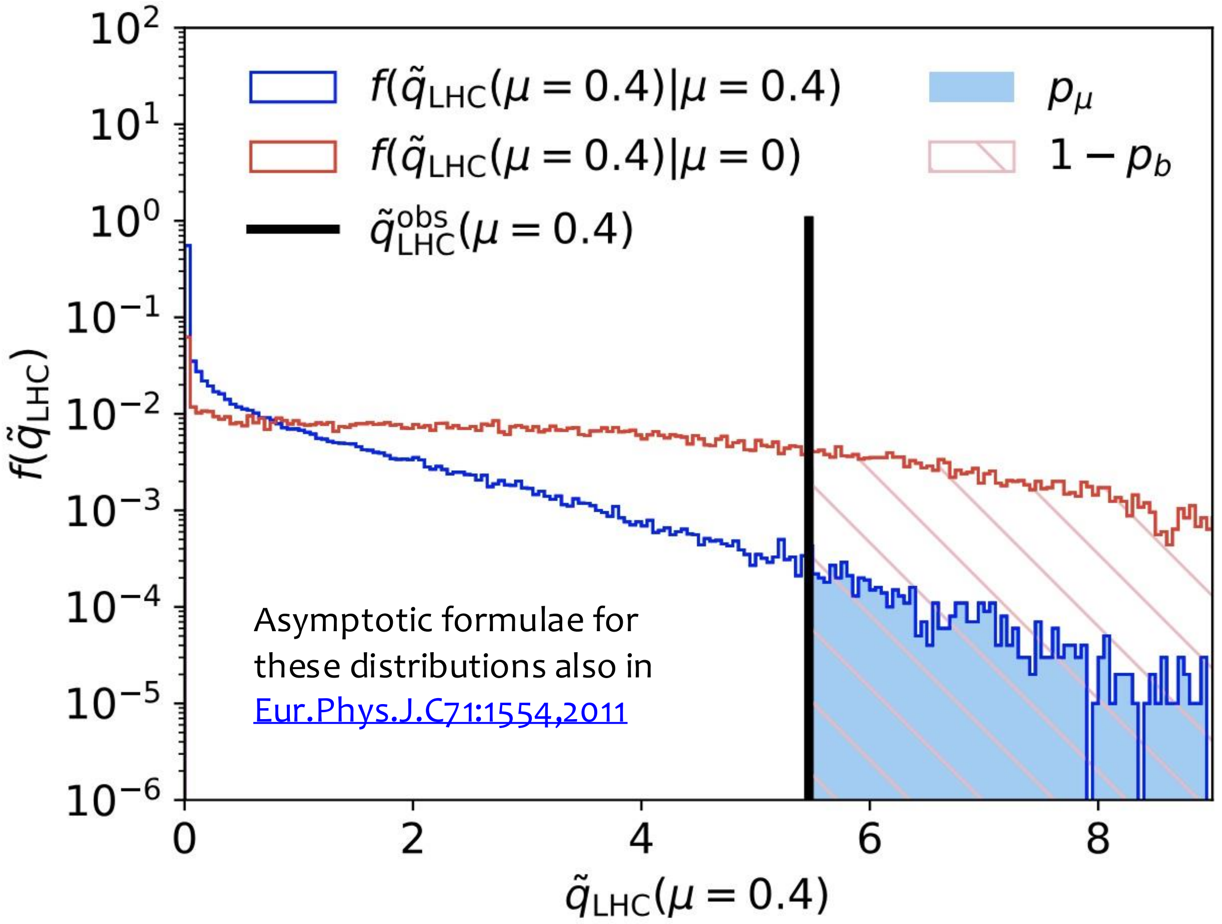
Test-statistic for upper limits at the LHC

$$\tilde{q}_{\mathrm{LHC}}(\mu) = \begin{cases} -2\ln\left(\dfrac{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(\hat{\mu},\hat{\vec{\nu}})}\right) & \text{if } 0 \leq \hat{\mu} \leq \mu, \\ -2\ln\left(\dfrac{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(0,\hat{\hat{\vec{\nu}}}(0))}\right) & \text{if } \hat{\mu} < 0, \\ 0 & \text{if } \hat{\mu} > \mu, \end{cases}$$

Test-statistic is a function of the signal strength: **$H_0 \rightarrow H(\mu)$**

Upper limits - CLs

We need to be careful in this case to avoid excluding a signal when the data also doesn't agree well with the background hypothesis → replace p-value with ratio of p-values



Asymptotic formulae for these distributions also in [Eur.Phys.J.C71:1554,2011]

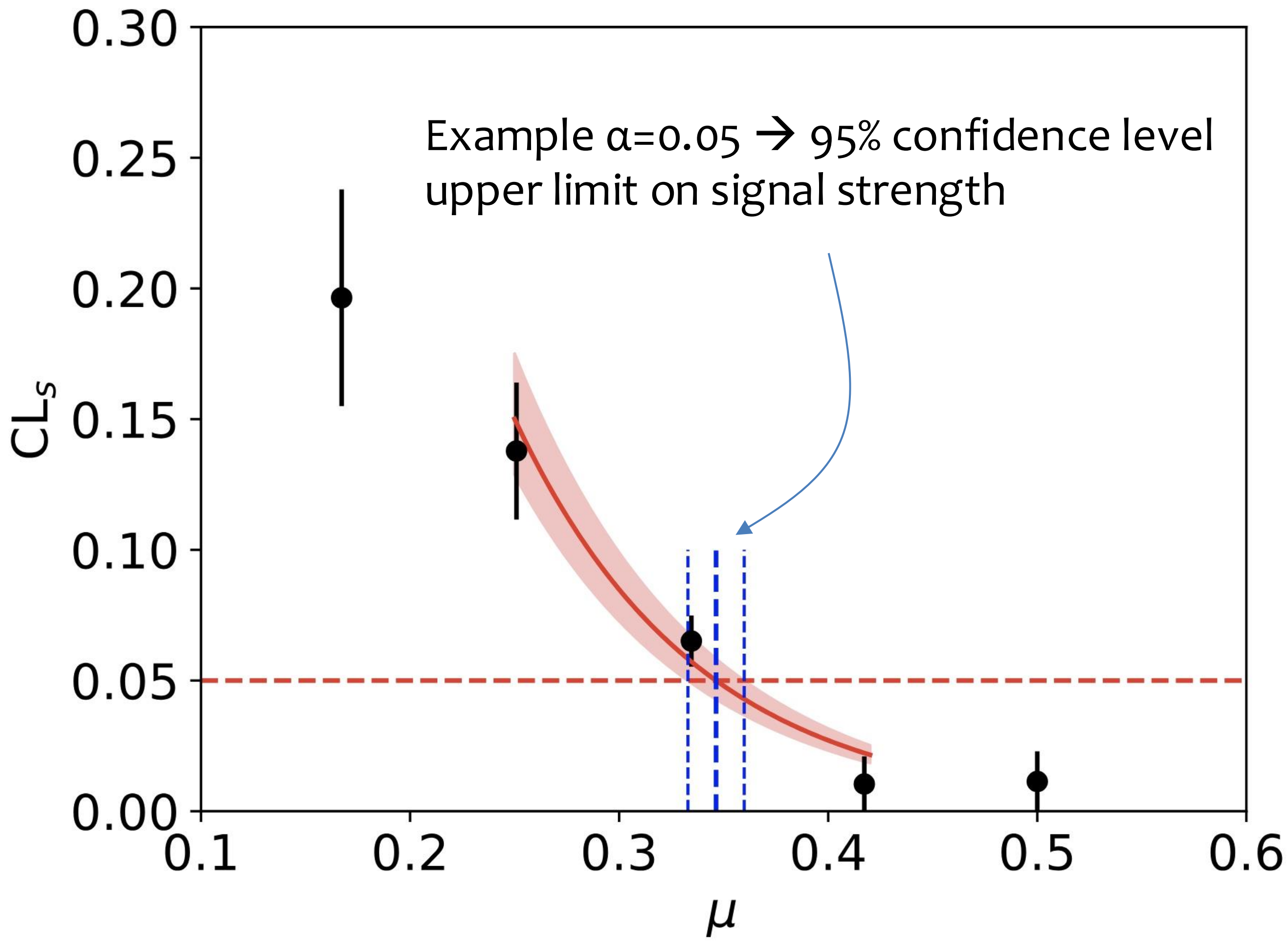
$$p_\mu = \int_{\tilde{q}(\mu)^{\mathrm{obs}}_{\mathrm{LHC}}}^{+\infty} f(\tilde{q}(\mu)_{\mathrm{LHC}}|\mu)$$

$$1 - p_b = \int_{\tilde{q}(\mu)^{\mathrm{obs}}_{\mathrm{LHC}}}^{+\infty} f(\tilde{q}(\mu)_{\mathrm{LHC}}|0)$$

$$CL_s = \frac{p_\mu}{1-p_b}$$

Upper limits - CLs

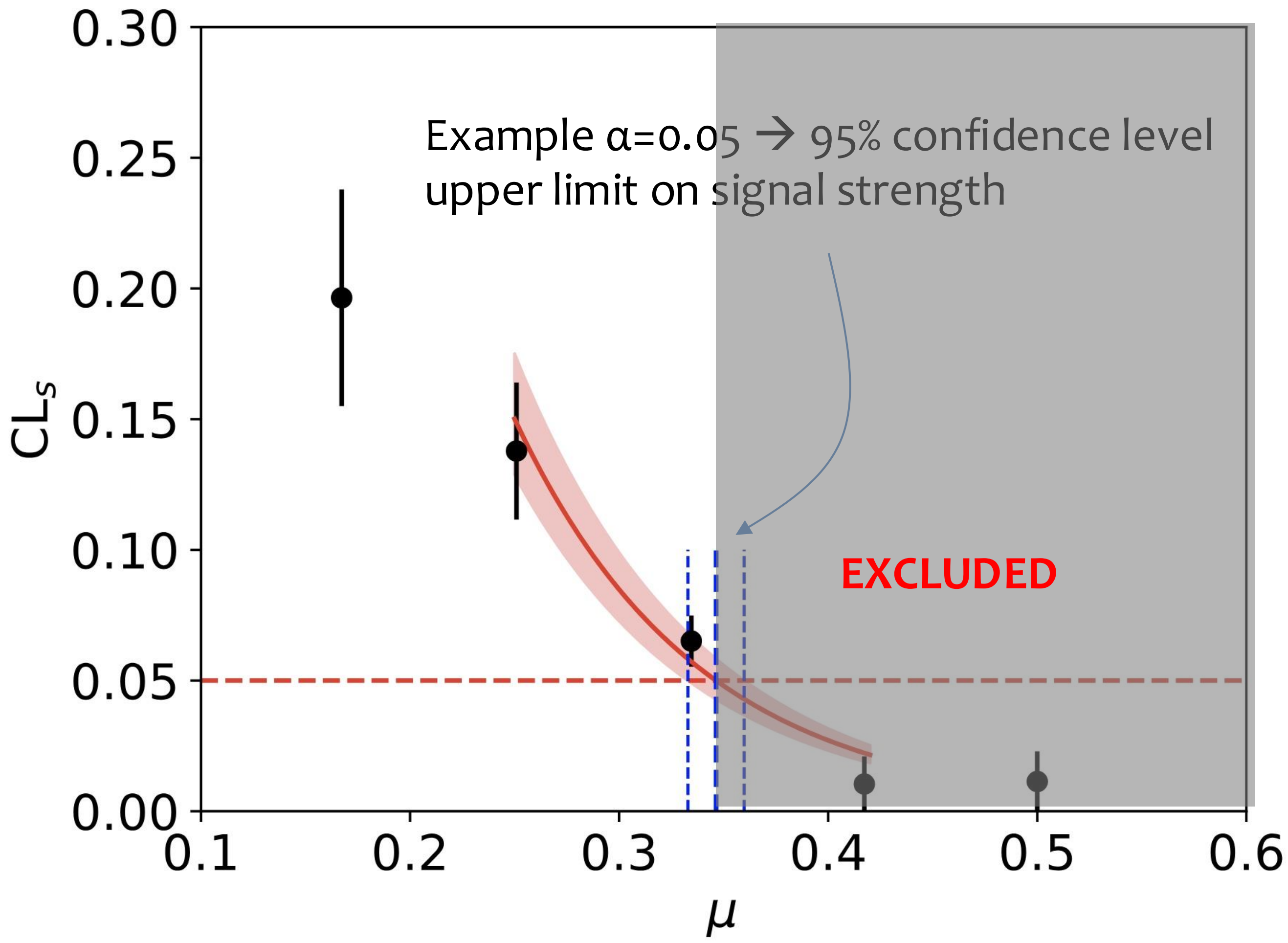
Derive upper limits on μ by scanning for $CL_s = \alpha$



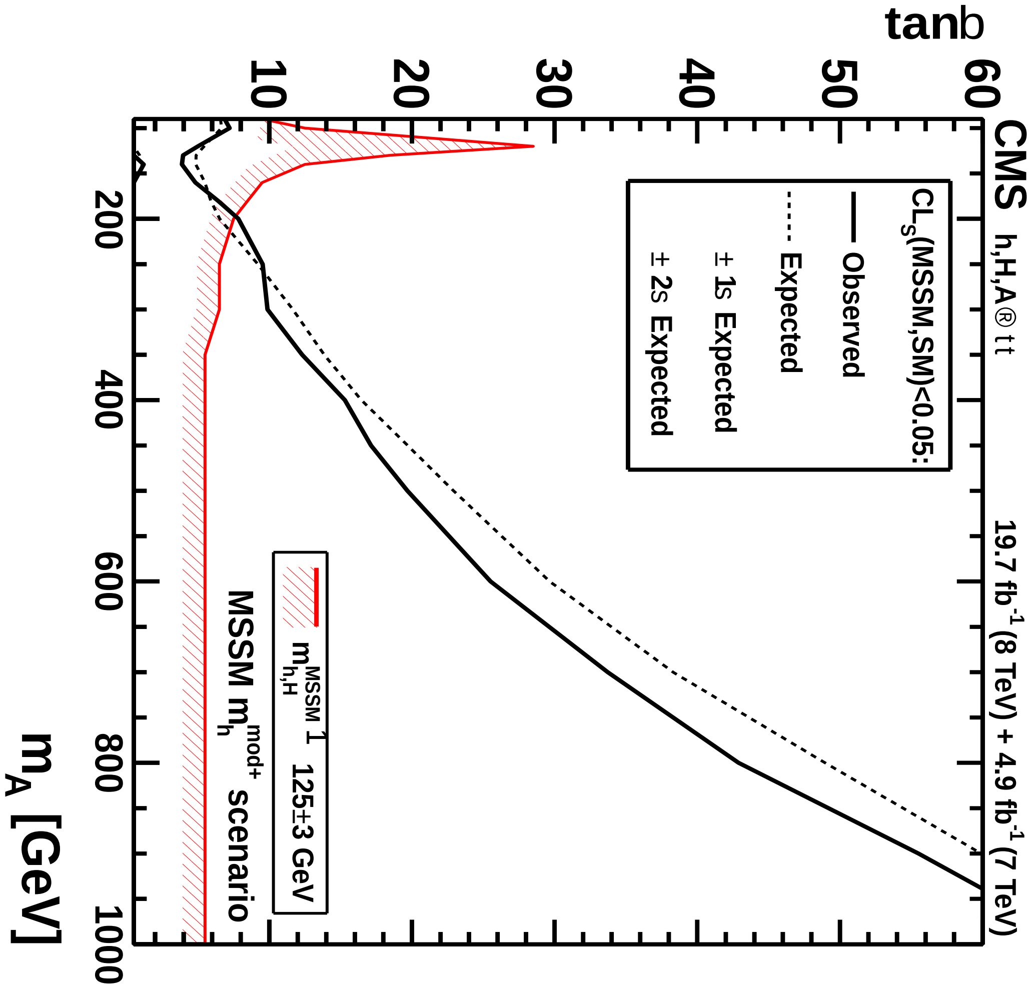
Upper limits - CLs

Derive upper limits on μ by scanning for $CL_s = \alpha$

Upper limit on μ tells us the ***smallest*** amount of signal that can be excluded
→ every value larger is excluded at 95% CL
→ every H(μ) with μ>~0.35 is excluded



Upper limits - CLs

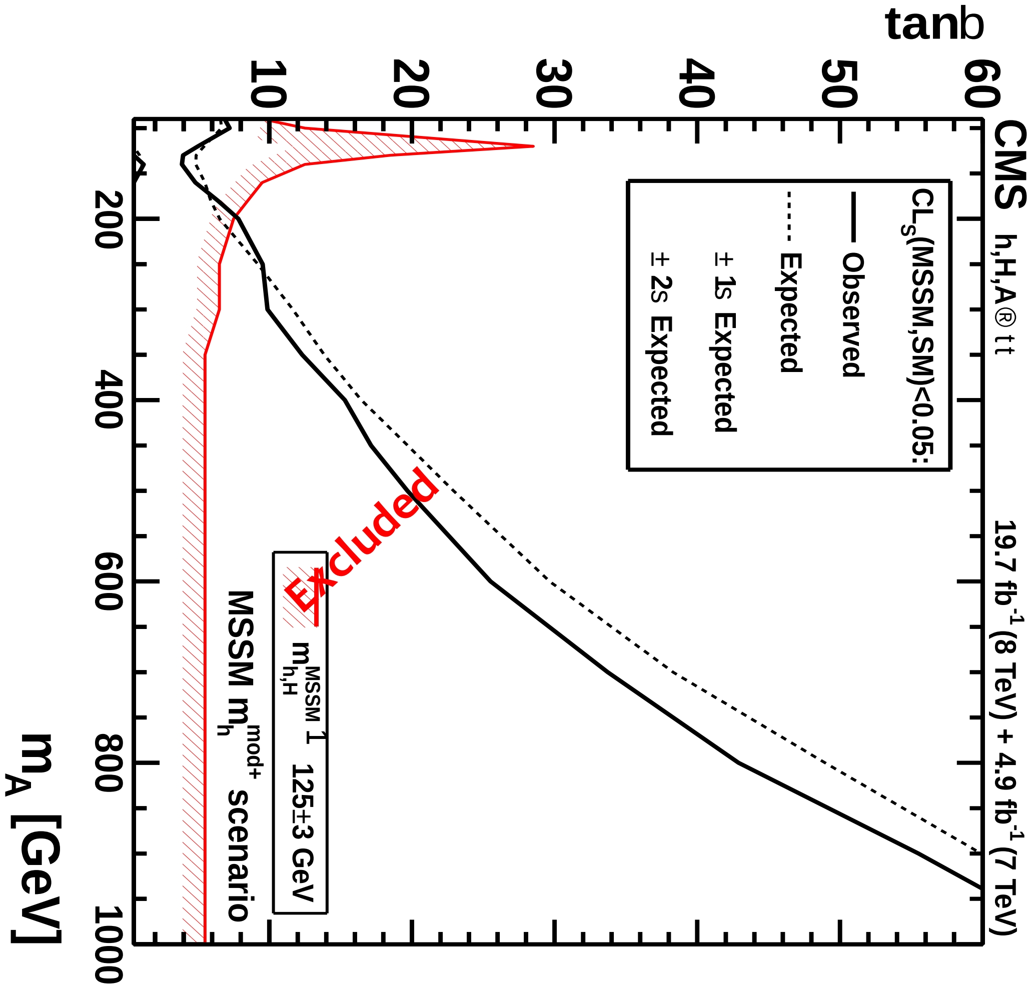


Remember that $$\mu = \frac{\sigma}{\sigma(pp \to X)}$$

So whenever $\mu < 1$ we exclude the signal model (at 95% CL)

Many BSM theories will include parameters that must be specified to predict $\sigma(pp \to X)$

Upper limits - CLs



Remember that $$\mu = \frac{\sigma}{\sigma(pp \rightarrow X)}$$

So whenever $\mu < 1$ we exclude the signal model (at 95% CL)

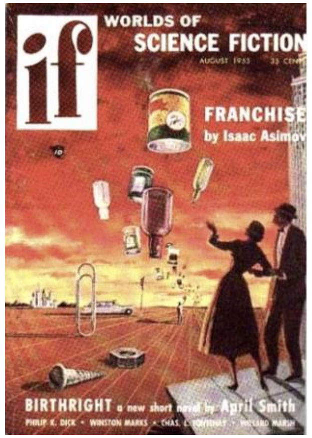
Many BSM theories will include parameters that must be specified to predict $\sigma(pp \rightarrow X)$

→ Scan over parameters and shade region for which $\mu < 1$

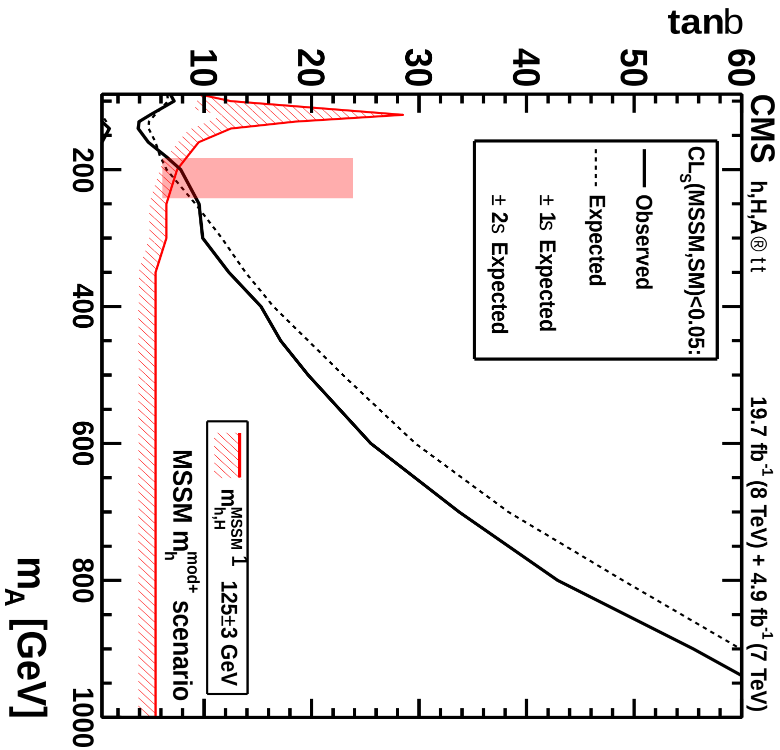
→ excluded region of the BSM theory!

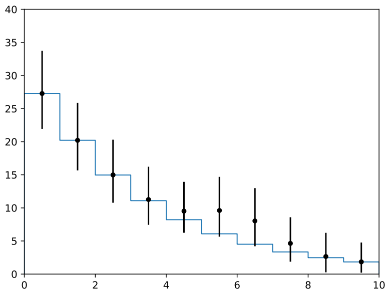
Expected results

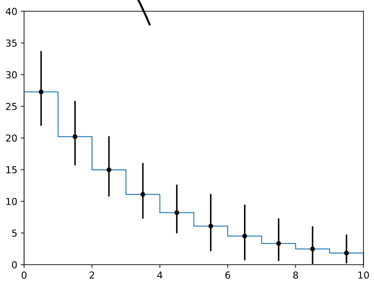
Significance and upper limits are also random variables!
→ If we want to know how sensitive our analysis is, we can calculate expected results

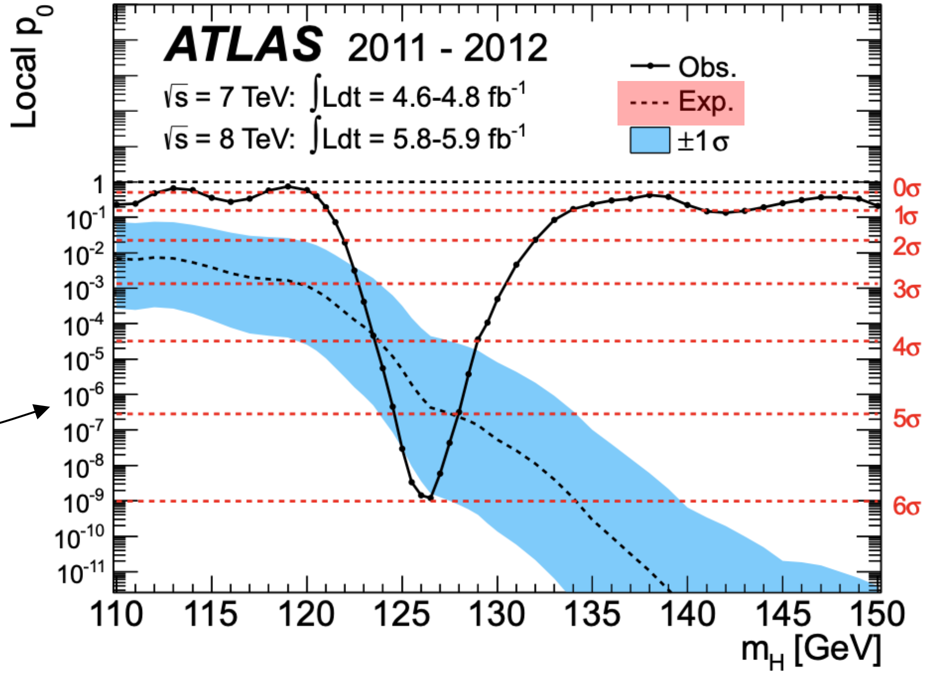


Replace data with expectation
→ **Asimov datasets**









End of lectures

We have covered the basic data analysis and statistics methods used mostly at the LHC, however, there are many more techniques that have been / are used depending on the analysis that we don't have time to cover

Here are some further reading links for LHC statistics in case you are interested

- F. James, "*Statistical Methods in Experimental Physics*", ISBN: 978-9-812-70527-3 (2006).
- G. Cowan, "*Statistical Data Analysis*", ISBN: 978-0-198-50155-8 (1998).
- G. Cowan, "Statistics" (section 39) in "*Review of particle physics*", Chin. Phys. C 40, 100001 (2016).
- L. Lista, "*Statistical Methods for Data Analysis in Particle Physics*", ISBN 978-3-319-20176-4, (2015).
- A. Stuart, K. Ord, S. Arnold, "*Kendall's Advanced theory of Statistics*", Vol 2A: Classical inference and the linear model, ISBN: 978-0-470-68924-0 (2010).
- L. Lyons, N. Wardle, "*Statistical issues in searches for new phenomena in High Energy Physics*", Journal of Physics G: Nuclear and Particle Physics, Volume 45, Number 3.
- O. Behnke, K. Kroninger, G. Schott, T. Schorner-Sadenius, "*Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*", ISBN: 978-3-527-41058-3 (2013).
- K. Cramner, "*Practical Statistics for the LHC*", Proceedings, 2011 European School of High-Energy Physics, (2011).

You can also find some more interactive practical statistics examples in my GitHub area :
https://github.com/nucleosynthesis/IntroductionToStatistics?tab=readme-ov-file

launch binder

Now it's your turn

In tomorrow's exercise, you will include control regions and systematic uncertainties in your statistical analysis and see how this degrades the sensitivity of the measurement

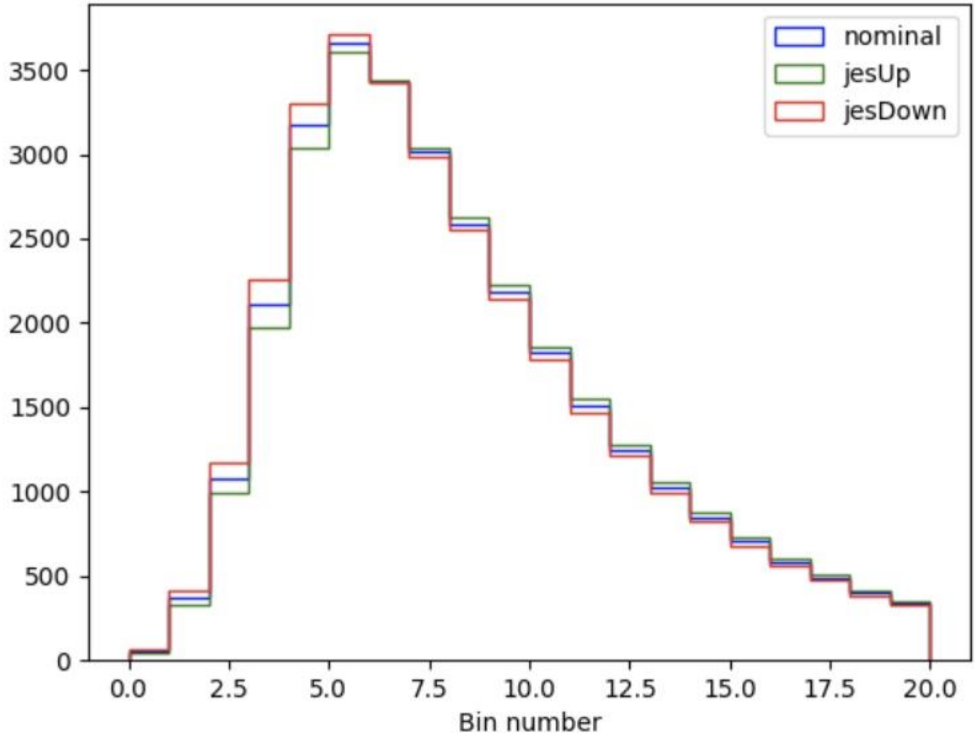
Exercise 3 - Control Regions and Systematic Uncertainties

Launch the `cms_combine` container by typing the following into a terminal on your laptop (or by clicking the play button next to the cms_combine container in the Docker desktop application and using the terminal there).

Bash

```
docker start -i cms_combine
```

In today's exercise, we are going to use our 4j0b control region that we populated at the end of exercise 2 to constrain our `wjets` process in our 4j2b signal region. Don't worry if you didn't manage to process the samples to create the histograms for the 4j0b region. I have put a `.csv` file `exercise2solutions/allregions_mbjj.csv` that has both the signal region and control region histograms for you. You'll also find the datacard for the signal region in the same folder: `signalregion_mbjj.txt`.



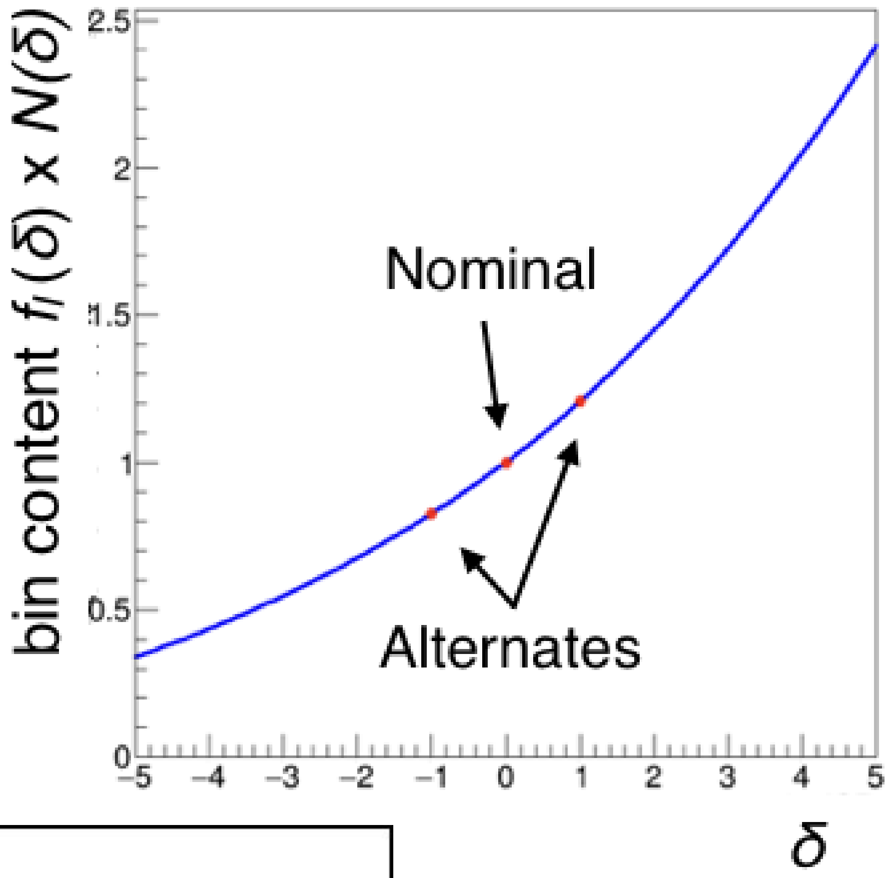
Don't worry if you didn't complete the previous exercise, all of the solutions can be found in `ttbarAnalysis/exercise2solutions`

(Extra Slide) Interpolation example

The effects of correlated systematic uncertainties on n_I are modelled using quadratic(linear) interpo(extrapo)lation function – simplified example of interpolation

$$\boxed{f_I(\boldsymbol{\delta}) = f_I^0 \cdot \frac{1}{F(\boldsymbol{\delta})} \prod_j p_{Ij}(\delta_j)}$$

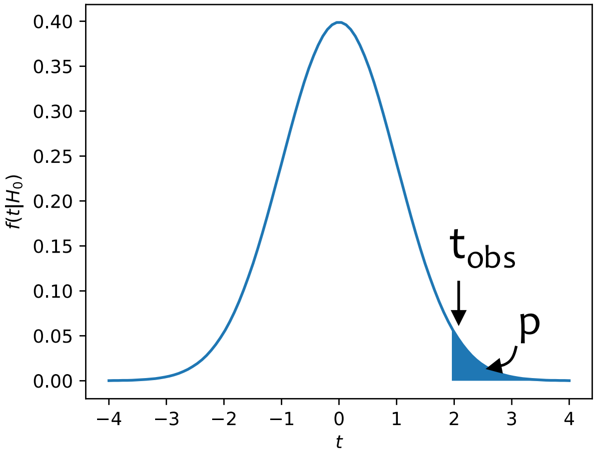
$F(\boldsymbol{\delta}) = \sum_I f_I(\boldsymbol{\delta})$



$$p_{Ij}(\delta_j) = \begin{cases} \frac{1}{2}\delta_j(\delta_j - 1)\kappa_{Ij}^{-} - (\delta_j - 1)(\delta_j + 1) + \frac{1}{2}\delta_j(\delta_j + 1)\kappa_{Ij}^{+} & \text{for } |\delta_j| < 1 \\ \left[\frac{1}{2}(3\kappa_{Ij}^{+} + \kappa_{Ij}^{-}) - 2\right]\delta_j - \frac{1}{2}(\kappa_{Ij}^{+} + \kappa_{Ij}^{-}) + 2 & \text{for } \delta_j > 1 \\ \left[2 - \frac{1}{2}(3\kappa_{Ij}^{-} + \kappa_{Ij}^{+})\right]\delta_j - \frac{1}{2}(\kappa_{Ij}^{+} + \kappa_{Ij}^{-}) + 2 & \text{for } \delta_j < -1 \end{cases}$$

(Extra Slide) p_0 distribution

The p-value is
- A random variable that depends on the observed data (it's a post-observation quantity)
- Distributed uniformly between 0 and 1 under the null hypothesis



$$p = P(t > t_{obs}|H_0) = 1 - P(t < t_{obs}|H_0) = 1 - F(t)$$

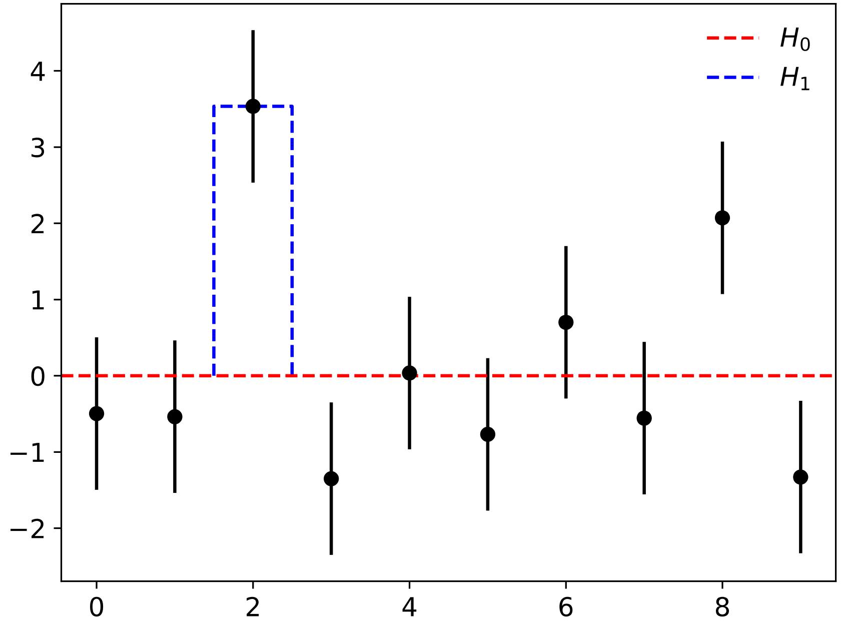
$$1 - F(t) = P(t > t_{obs}|H_0) = P(F(t) > F(t_{obs})|H_0) = 1 - P(F(t) < F(t_{obs})|H_0)$$

Since F(.) is monotonic and increasing

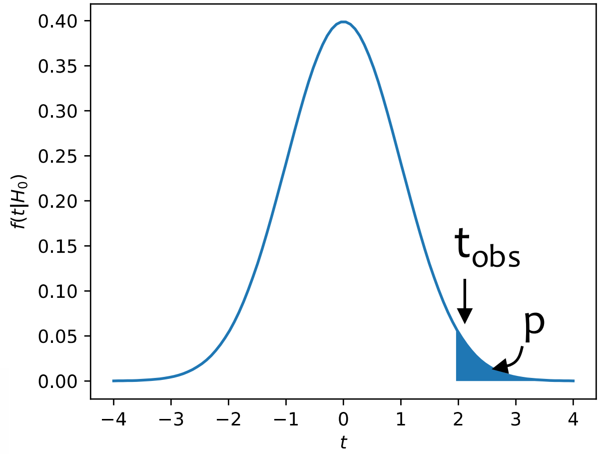
$$F(t) = P(F(t) < F(t_{obs})|H_0)$$

Which is true for any t_{obs}

→ F(t) is uniform
→ p is uniform



(Extra Slide) p_0 distribution



p-value is flat under H_0