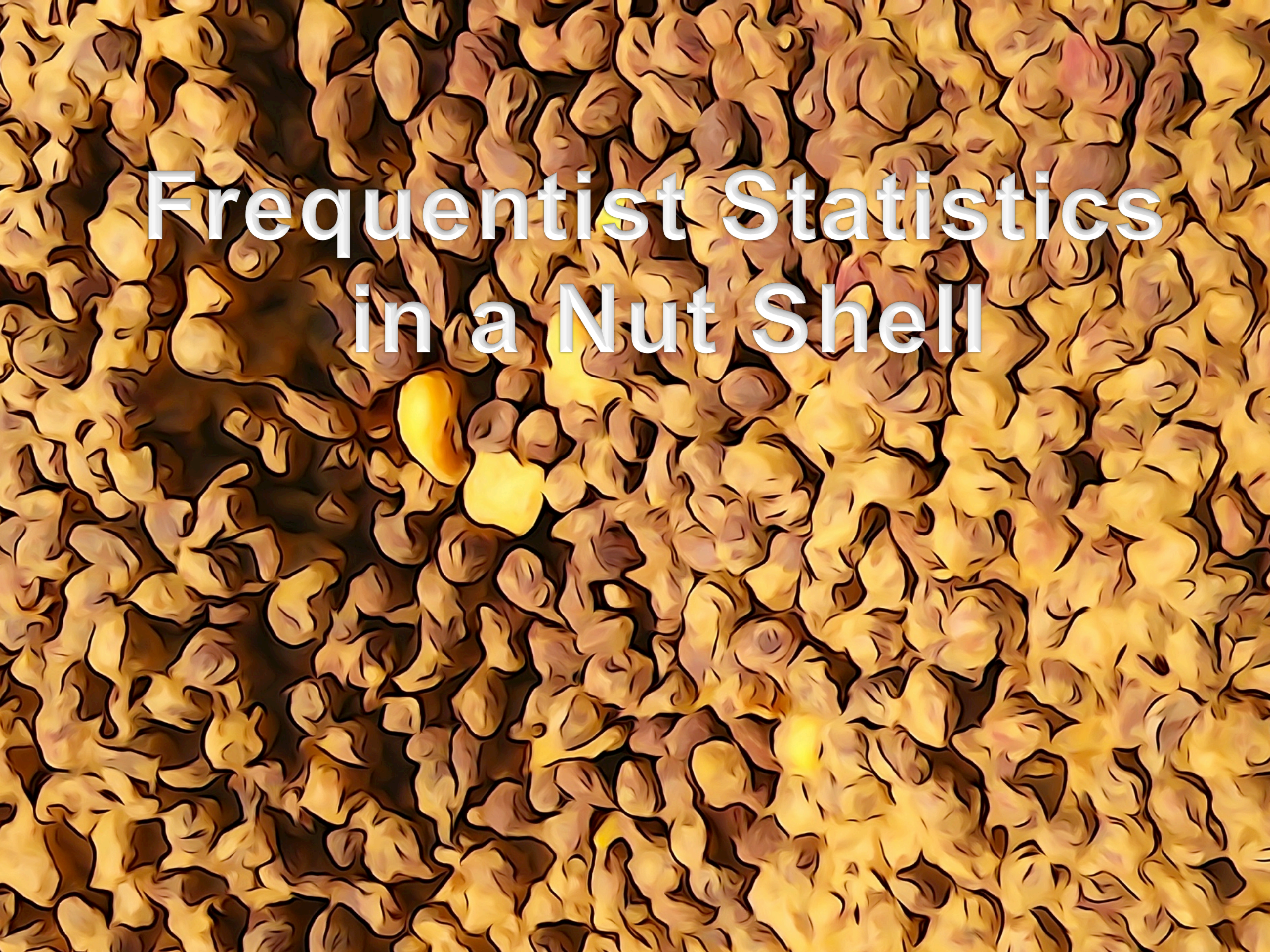
Frequentist Statistics @ LHC

Eilam Gross, Weizmann Institute of Science

This presentation would have not been possible without the tremendous help of the following people throughout many years

Louis Lyons, Alex Read, Glen Cowan ,Kyle Cranmer
Ofer Vitells & Bob Cousins


Frequentist Statistics
in a Nut Shell

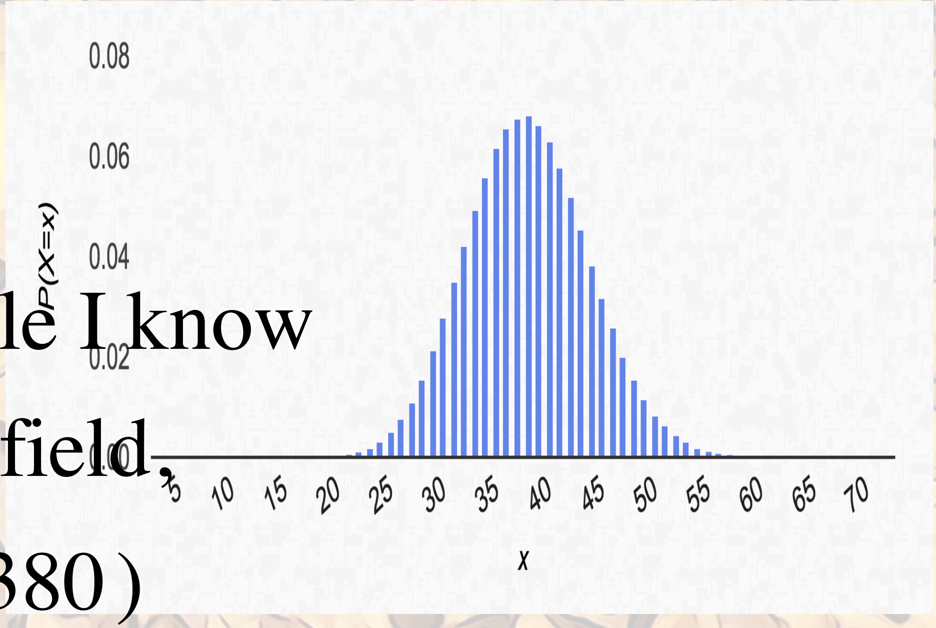
- I came here and realized I know very few people
- Am I too young? NO; Am I THAT old? Hopefully NOT
- Am I part of this community?

$$H_{null} = \text{My Field}$$

$$H_{alt} = \text{Not My Field}$$

test statistic $q = \#$ people I know

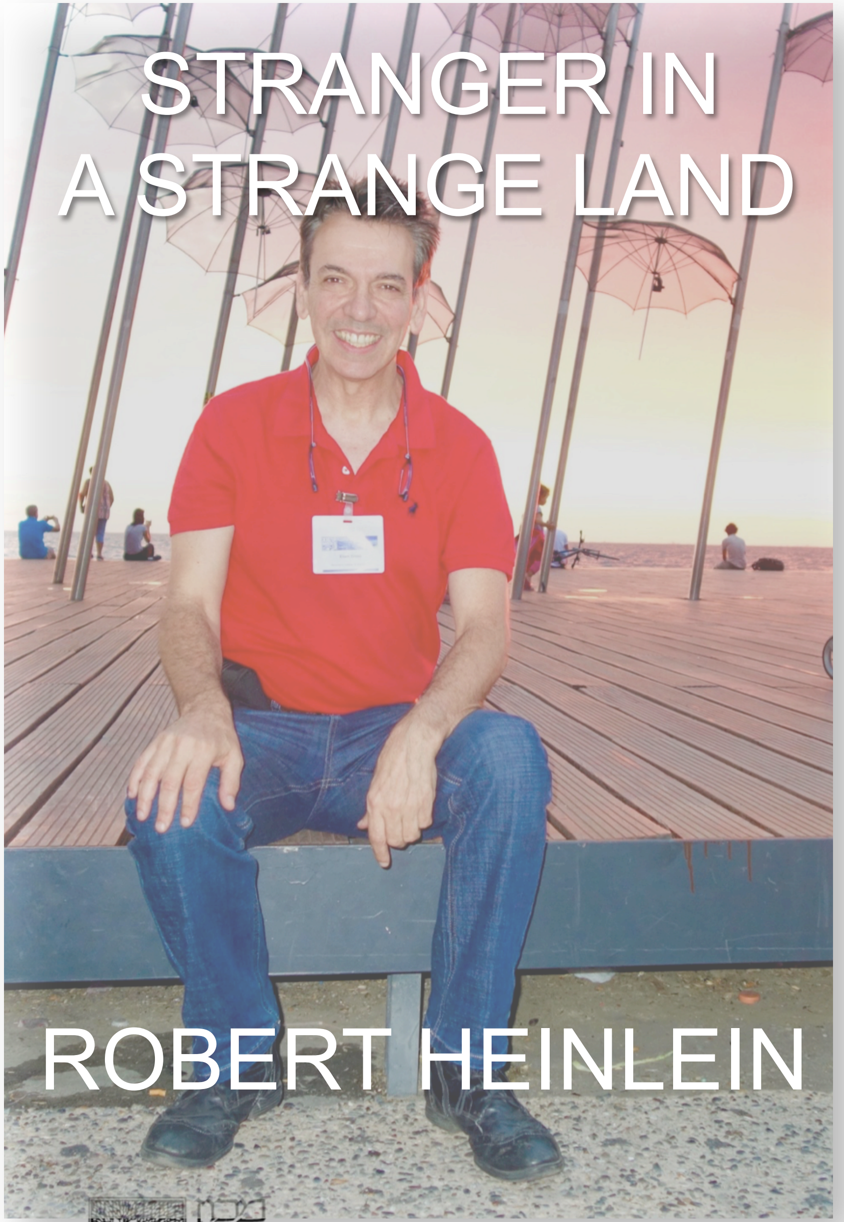
In a typical conf in my field,

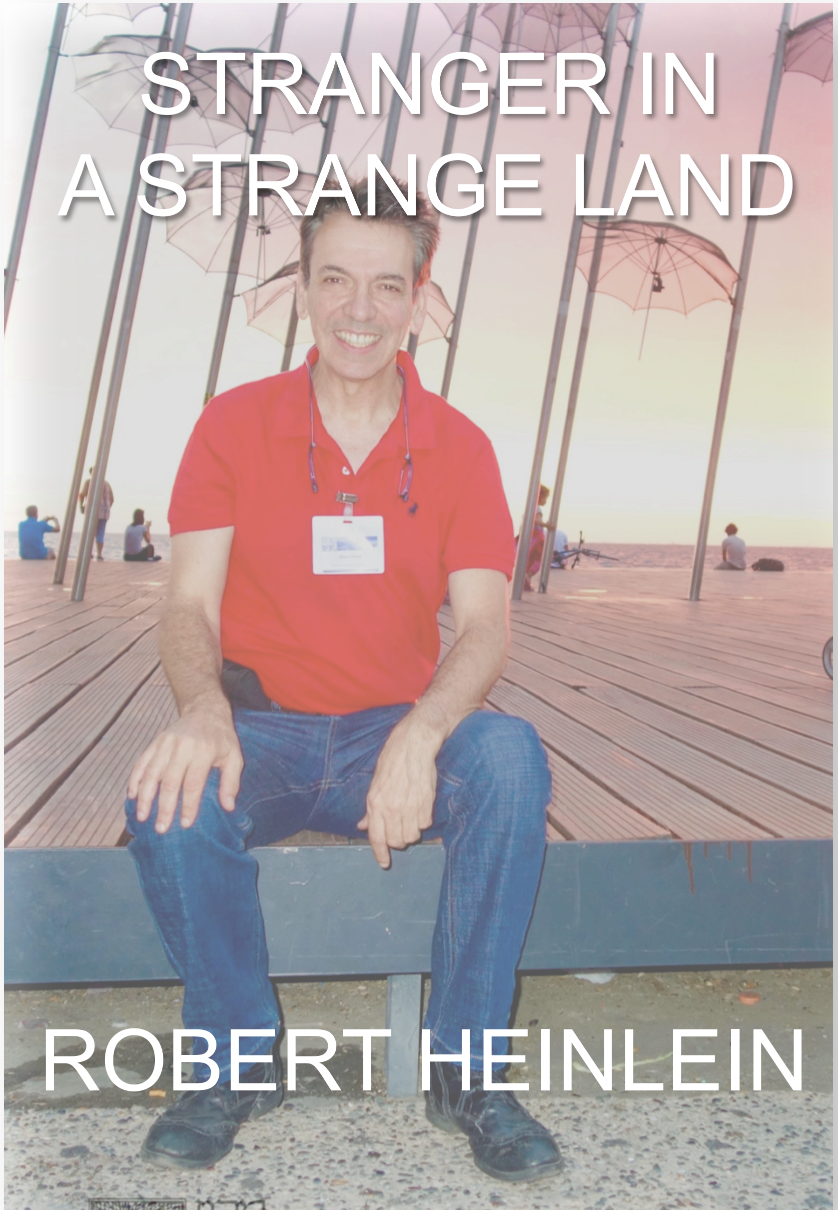


I know $\geq 10\%$ ($n \geq 38|380$)

$$p-value = \Pr ob(n \leq 6|380; p=0.1) \ll 1\% \Rightarrow$$

reject H_{null} @ $> 99\% CL$


STRANGER IN
A STRANGE LAND
ROBERT HEINLEIN



Statistics should not be
a STRANGE land
For no one

Statistics is an international language
A multi-disciplinal language

Instead of an Outline (biased)

- 18 years to Feldman-Cousins (FC)

  Robert D. Cousins and Gary J. Feldman, Phys. Rev. D 57, 3873 (1998).

- 17 years to CLs (CLs)

  A.L. Read, J. Phys. G 28, 2693 (2002

- 6 years to the asympotic formula and the Asimov Data Set (CCGV)

  G. Cowan, K. Cranmer, E. G. et al., "Asymptotic formulae for likelihood-based tests of new physics", Eur. Phys. J. C 71 (2011) 1

- 6 years to the Trial Factors and the LEE asymptotic formulae (LEE)

  E. G. and O. Vitells, "Trial factors for the look elsewhere effect in high energy physics", Eur. Phys. J. C 70 (2010) 525

- 4 years to Higgs Discovery (All the above)

- 1 month to the fall of the 750 (2D LEE)

  O. Vitells and E. G., Estimating the significance of a signal in a multi-dimensional search, 1669 Astropart. Phys. 35 (2011) 230,

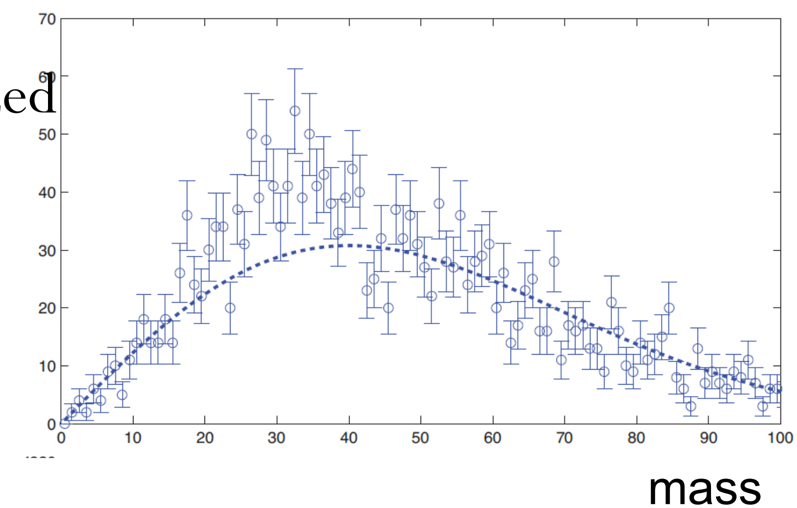
A Tale of Two Hypotheses

What is the statistical challenge?

- The Challenge:
  to tell in the most powerful way, and to the best of our current scientific knowledge, if, in our data, there is new physics (signal), beyond what is already known (background)

- The DATA:
  Billions of Events which could be visualized with histograms

- How compatible is the **data (dots)** with the **SM expectation (dashed)**?



A minimal vocabulary

The background hypothesis, $b(\theta)$

θ represent the Nuisance Parameters

The Signal hypotheis (needs some Model)

$$s(m)+b$$

In a counting experimet $\text{n}=\mu s(m)+b$

Signal Strength $\mu = \dfrac{\sigma_{obs}}{\sigma_{SM}^{exp}}$

$$\mu = 0 \Leftrightarrow H_0$$

$$\mu \neq 0 \Leftrightarrow H_\mu$$

A Frequentist Tale of Two Hypotheses

- The first step in any hypothesis test is to state the relevant **null,** H_{null} and **alternative hypotheses,** say, H_{alt}

$$n = \mu s(m) + b(\theta)$$

$$\mu = 0 \Leftrightarrow H_0$$

$$\mu \neq 0 \Leftrightarrow H_\mu$$

One hypotheis is the *null* hypotheis

Other hypotheis is the *alt*

Test H_{null}, and try to reject it in favour of H_{alt}

fail to reject the null hypothesis or
reject it in favor of an alternative hypothesis (if p-value is small)

A Frequentist Tale of Two Hypotheses

- *The next step is to define a test statistic, q (q_{null}), to test the null hypothesis*

The pdf of q need to be known (toys or asymptotic)

- The two most common test statistics in HEP are the NP and the PL
- The NP is a symmetric test statistic

$$q_{NP} = -2ln\frac{L(H_{\mu})}{L(H_{0})}$$

- The PL is obeying the Wilks' theorem with Nuisance Parameters

$$q_{\mu} = -2ln\frac{L\left(\mu s + b_{\mu}(\hat{\hat{\theta}})\right)}{L\left(\hat{\mu} s + b(\hat{\theta})\right)} \sim \chi^{2}$$

Profile Likelihood (PL) and Wilks' Theorem

$$q_\mu = -2ln \frac{L\left(\mu s + b_\mu(\hat{\hat{\theta}})\right)}{L\left(\hat{\mu} s + b(\hat{\theta})\right)}$$

$$q_\mu = -2ln \frac{\max_\theta L\left(\mu s + b_\mu(\theta)\right)}{\max_{\mu,\theta} L\left(\mu s + b(\theta)\right)}$$

$$q_\mu = q_\mu(\hat{\mu})$$

$\hat{\mu}$ MLE of μ

$\hat{\hat{\theta}}_\mu$ MLE of θ fixing μ

A slight change:
upward fluctuations of the signal do not serve as an evidence against the signal

$$q_\mu = \begin{cases} q_\mu(\hat{\mu}) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

$$f(q_\mu \mid H_\mu) \sim \frac{1}{2}\chi^2$$

- *The pdf of q need to be known (toys or asymptotic)*

q_{null}

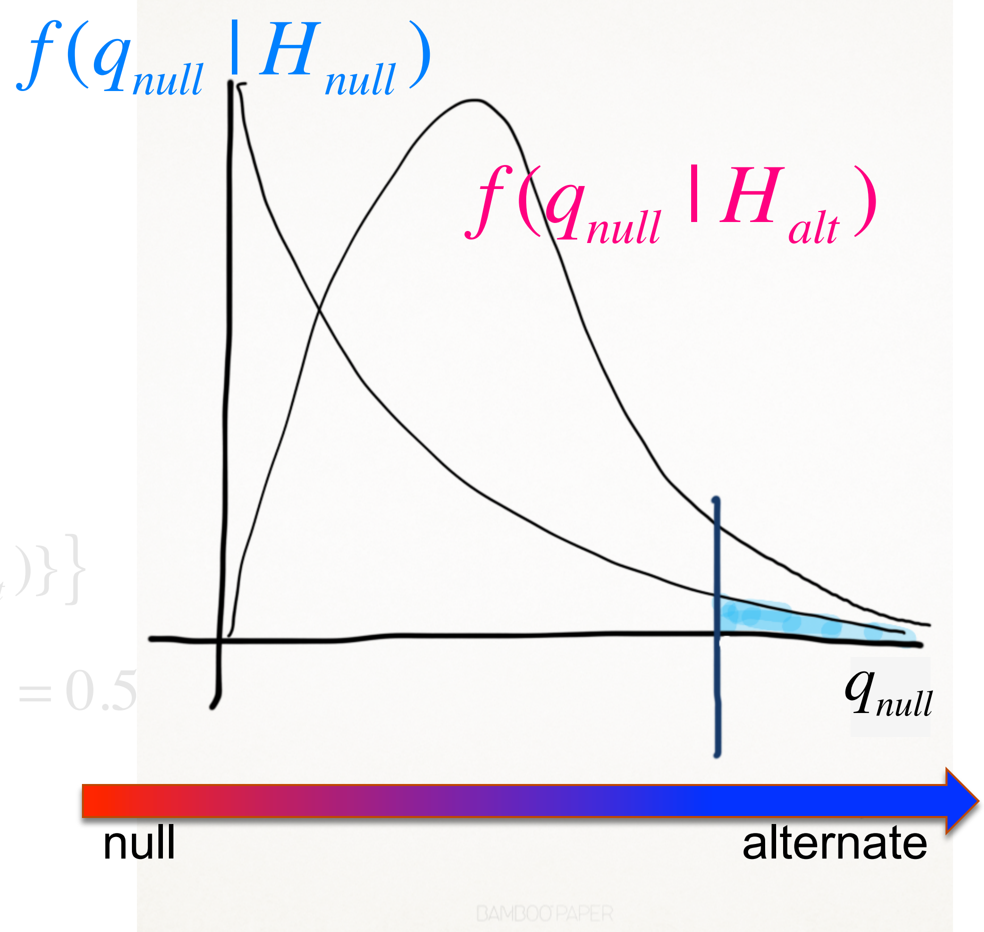
$f(q_{null} \mid H_{null})$

$q_{obs} \equiv q_{null,obs}$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$$

$f(q_{null} \mid H_{alt})$

$$q_A \equiv q_{null,A} = \{q \mid med\{f(q_{null} \mid H_{alt})\}\} \quad \int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null}) dq_{null} = 0.5$$



A Frequentist Tale of Two Hypotheses

- The first step in any hypothesis test is to state the relevant **null,** H_{null} and **alternative hypotheses,** say, H_{alt}

- *The next step is to define a test statistic, q (q_{null}), to test the null hypothesis*

  The pdf of q need to be known (toys or asymptotic)

- **Compute from the observations the observed value q_{obs} *of the test statistic* q.**

- ***Based on q_{obs} find the p-value which is a measure of the compatibility of the data with null hypothesis***

- Decide (based on *the p-value*) to either
  fail to reject the null hypothesis or
  reject it in favor of an alternative hypothesis (if p-value is small)

- Compute from the observations q_{obs} and the ***p-value***

$$q_{null}$$

$$f(q_{null} \mid H_{null})$$

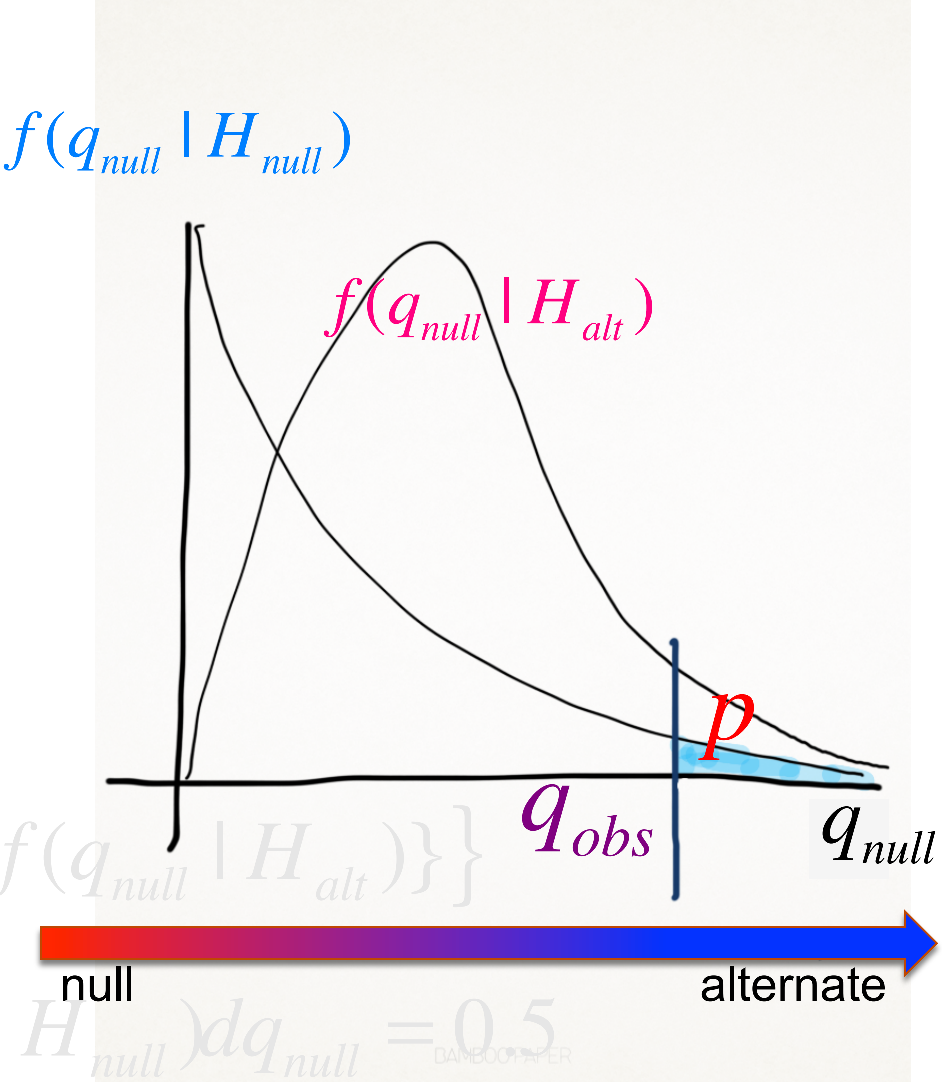
$$q_{obs} \equiv q_{null,obs}$$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$$

$$f(q_{null} \mid H_{alt})$$

$$\{q \mid med\{f(q_{null} \mid H_{alt})\}\}$$

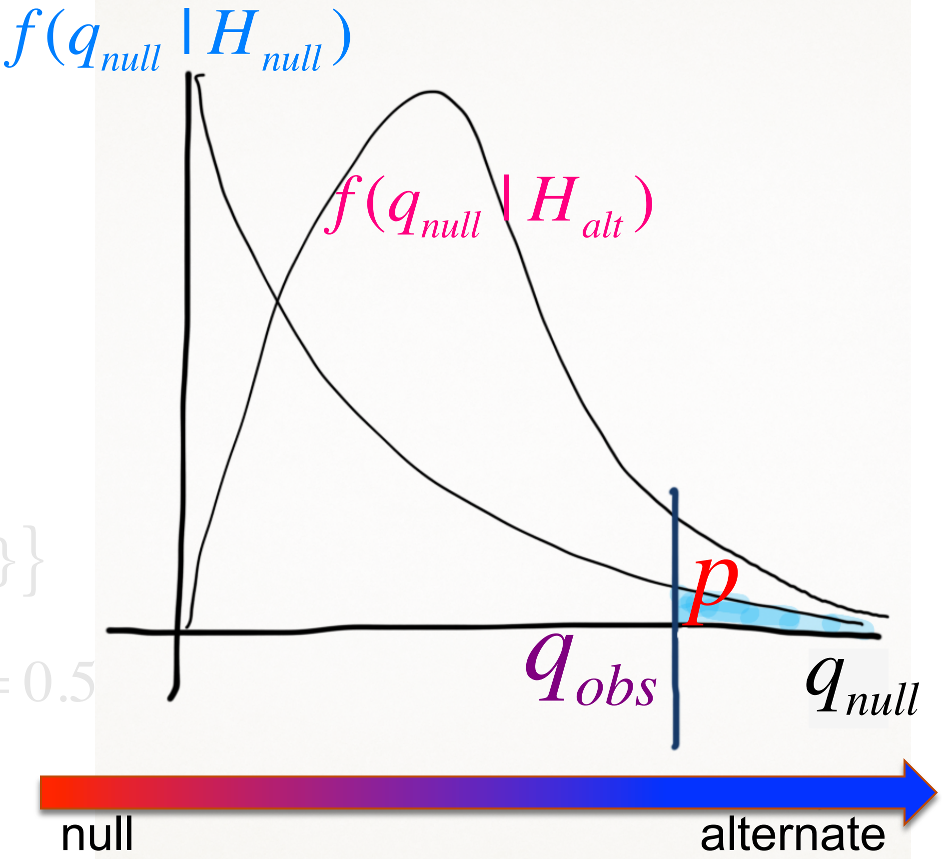
$$q_A \equiv q_{null,A} = \int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null}) dq_{null} = 0.5$$



- Decide (based on *the p-value*) to either fail to reject the null hypothesis or reject it in favor of an alternative hypothesis

$$q_{obs} \equiv q_{null,obs}$$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null})\,dq_{null}$$



Discovey

$$H_{null} = BG \quad p_{bg} = 2.9 \cdot 10^{-7} \sim 5\sigma$$

Exclusion

$$H_{null} = s + b \quad p_s = 0.05 = 5\% \sim 2\sigma$$

A Frequentist Tale of Two Hypotheses

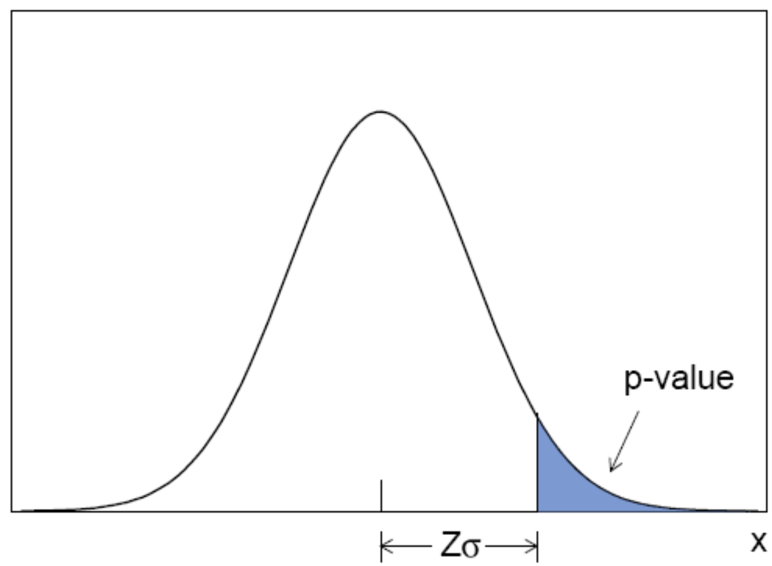
- It is a custom in HEP

Discovey $H_{null} = BG \quad p_{bg} = 2.9 \cdot 10^{-7} \sim 5\sigma$

Exclusion $H_{null} = s + b \quad p_s = 0.05 = 5\% \sim 2\sigma$

$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z)$$

$$Z = \Phi^{-1}(1-p)$$



Sensitivity of an Experiment

- Does your proposed experiment have the sensitivity to discover New Physics?
- Can it discover or exclude a particle with a mass *m*, within some given model? (Model=production and decay)
- Discovery sensitivity:
  - What is the median sensitivity (median p value) by which one would reject the BG hypothesis, given a nominal signal (e.g. $\mu=1$)
- Exclusion sensitivity:
  - What is the median sensitivity (median p value) by which one would reject the signal hypothesis (with signal strength μ), given a nominal BG ($\mu=0$) ?

Sensitivity of an Experiment

- Does your visional experiment have the sensitivity to discover New Physics?
- Can it discover or exclude a particle with a mass m, within some given model? (Model=production and decay)
- Discovery sensitivity:
  - What is the median sensitivity (median p value) by which one would reject the BG hypothesis, given a nominal signal (e.g. $\mu=1$)
- Exclusion sensitivity:
  - What is the median sensitivity (median p value) by which one would reject the signal hypothesis (with signal strength μ), given a nominal BG ($\mu=0$) ?

Sensitivity of an Experiment

- Does your visional experiment have the sensitivity to discover New Physics?
- Can it discover or exclude a particle with a mass m, within some given model? (Model=production and decay)
- Discovery sensitivity:
  - What is the median sensitivity (median p value) by which one would reject the BG (NULL) hypothesis, given a nominal signal (ALT) (e.g. $\mu=1$)
- Exclusion sensitivity:
  - What is the median sensitivity (median p value) by which one would reject the signal (NULL) hypothesis (with signal strength μ), given a nominal BG (ALT) ($\mu=0$) ?

Sensitivity of an Experiment

Testing the null hypothesis

Q: **What is the median sensitivity (median p value) by which one would reject the NULL hypothesis, given a nominal ALT?**

$$q_{null} = q_{null}(DATA)$$

A: nominal alt : Find the typical DATA set, that represents the nominal alt Hypothesis

ASIMOV DATA set satisfies $\int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{alt})dq_{null} = 0.5$

$$q_A \equiv q_{null,A} = \left\{ q_{null} \mid med\{f(q_{null} \mid H_{alt})\} \right\}$$

sensitivity given by p − value with respect to the null Hypothesis

$$p = \int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null})dq_{null} = p_A = med\left[p \middle| alt \right]$$

Q: **What is the median sensitivity (median p value) by which one would reject the NULL hypothesis, given a nominal ALT?**

q_{null}

$f(q_{null} \mid H_{null})$

$q_{obs} \equiv q_{null,obs}$

$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null}) dq_{null}$$

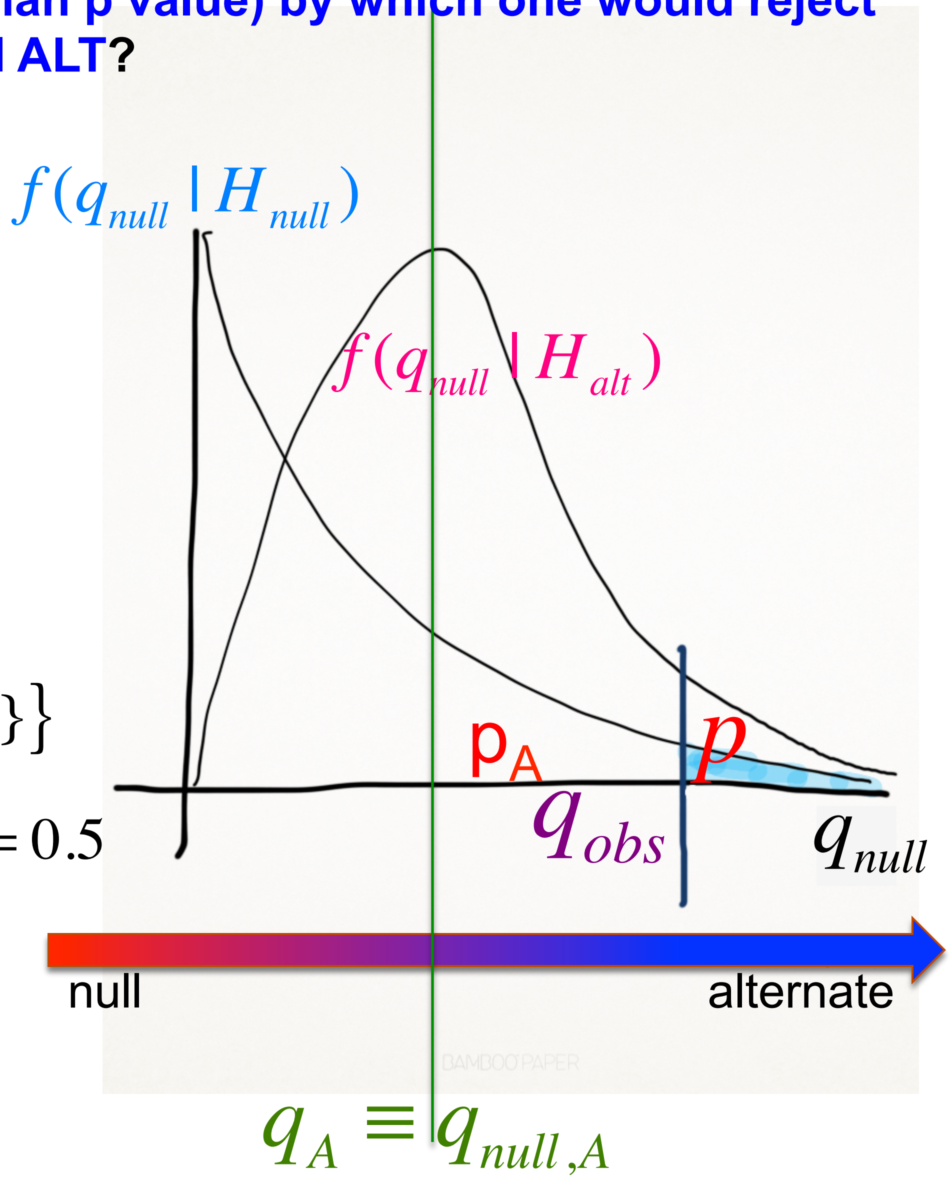
$f(q_{null} \mid H_{alt})$

$$q_A \equiv q_{null,A} = \{q \mid med\{f(q_{null} \mid H_{alt})\}\}$$

$$\int_{q_{null,A}}^{\infty} f(q_{null} \mid H_{null}) dq_{null} = 0.5$$

$$Z_{expected} = \sqrt{q_{null,A}}$$

$$\begin{aligned} \mathrm{med}[Z_0|\mu'] &= \sqrt{q_{0,A}} \\ \mathrm{med}[Z_\mu|0] &= \sqrt{q_{\mu,A}} \end{aligned}$$

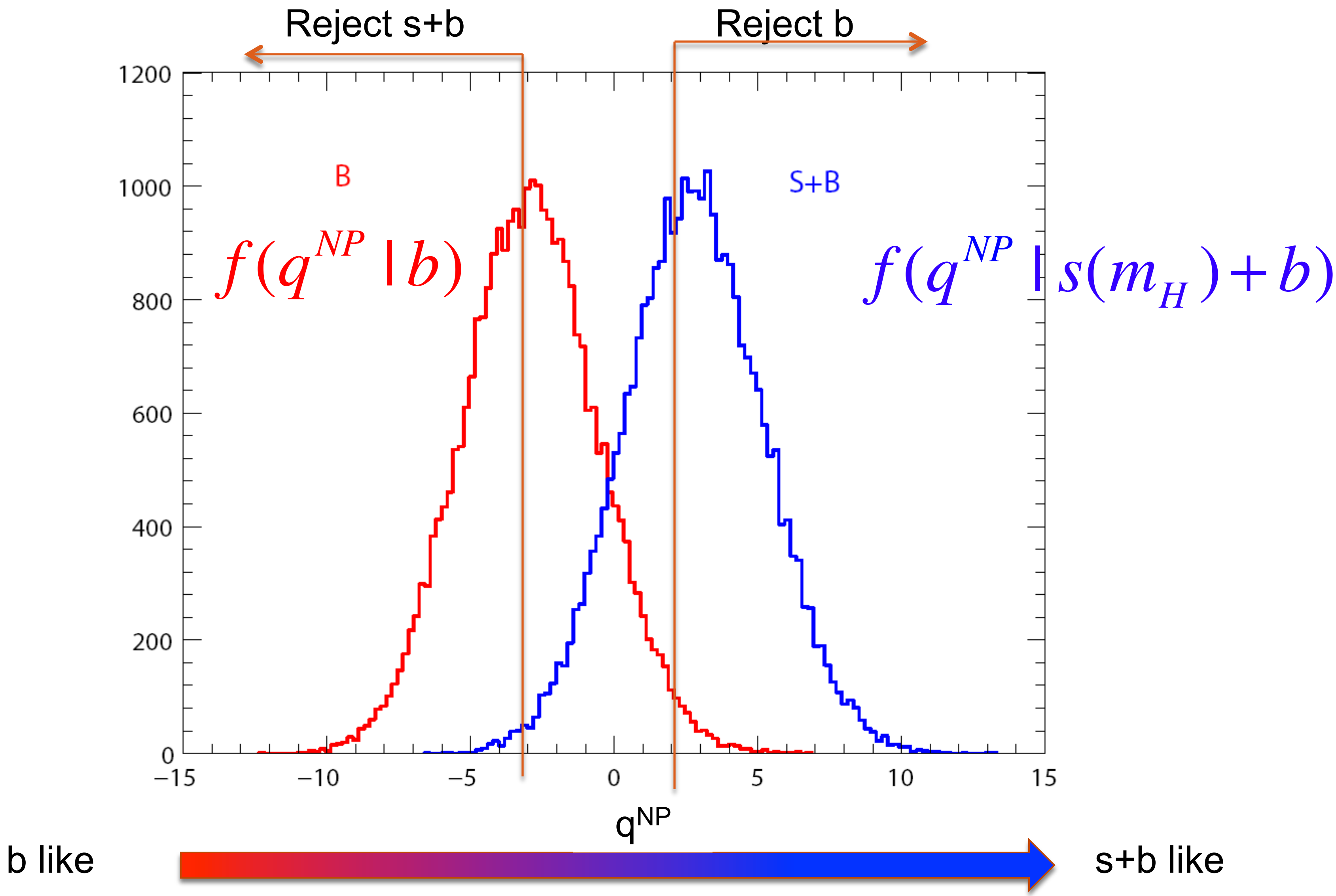


Test Statistics	Purpose	Experession	LR
q_0	discovery of positive signal	$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$	$\lambda(0) = \dfrac{L(0,\hat{\hat{\theta}}_0)}{L(\hat{\mu},\hat{\theta})}$
t_μ	2-sided measurement	$t_\mu = -2\ln\lambda(\mu)$	$\lambda(\mu) = \dfrac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})}$
$\tilde{t}_\mu$	avoid negative signal (FC)	$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu)$	$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})} & \hat{\mu} \geq 0 \\ \dfrac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(0,\hat{\hat{\theta}}_0)} & \hat{\mu} < 0 \end{cases}$
q_μ	exclusion	$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$	
$\tilde{q}_\mu$	exclusion of positive signal	$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$	

The Neyman-Pearson Test Statistic

- NP **test statistic** $$\boxed{q = -2\ln\frac{L(H_0)}{L(H_1)}}$$
- $$n = \mu s + b$$
$$H_0;\ \hat{\mu} = 0, \langle n_{obs} \rangle = b$$
$$H_1;\ \hat{\mu} = 1, \langle n_{obs} \rangle = s + b$$
$$L\left(H_0\right) = prob(x \mid H_0) = prob(x \mid b)$$
$$L(H_1) = prob(x \mid H_1) = prob(x \mid s + b)$$
$$L(H_\mu) = prob(x \mid H_\mu) = L(\hat{\mu})$$

PDF of NP test statistic

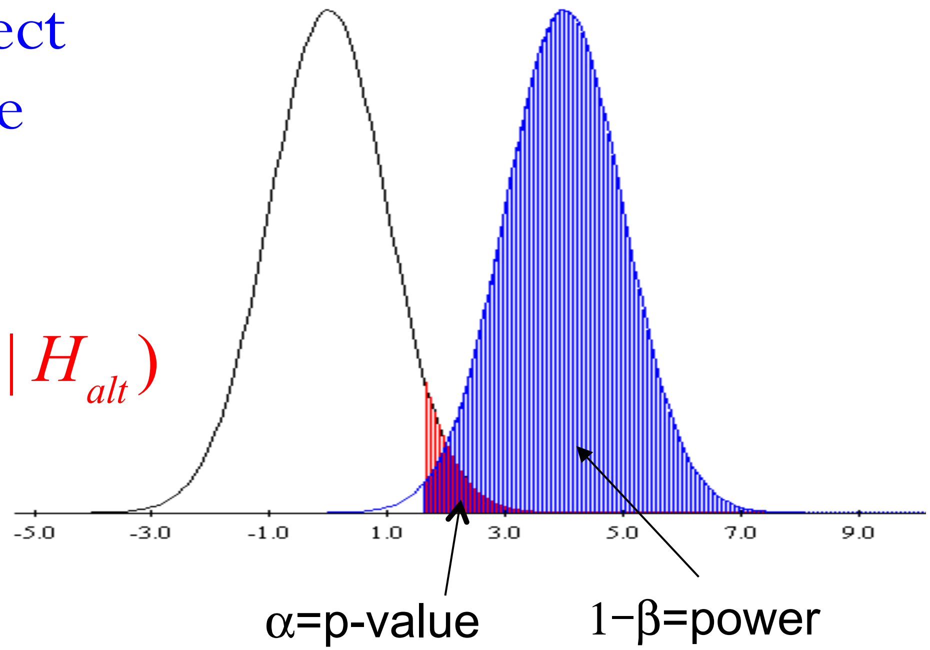


Basic Definitions: POWER

- The POWER of an hypothesis test is the probability to reject the null hypothesis when the alternate analysis is true

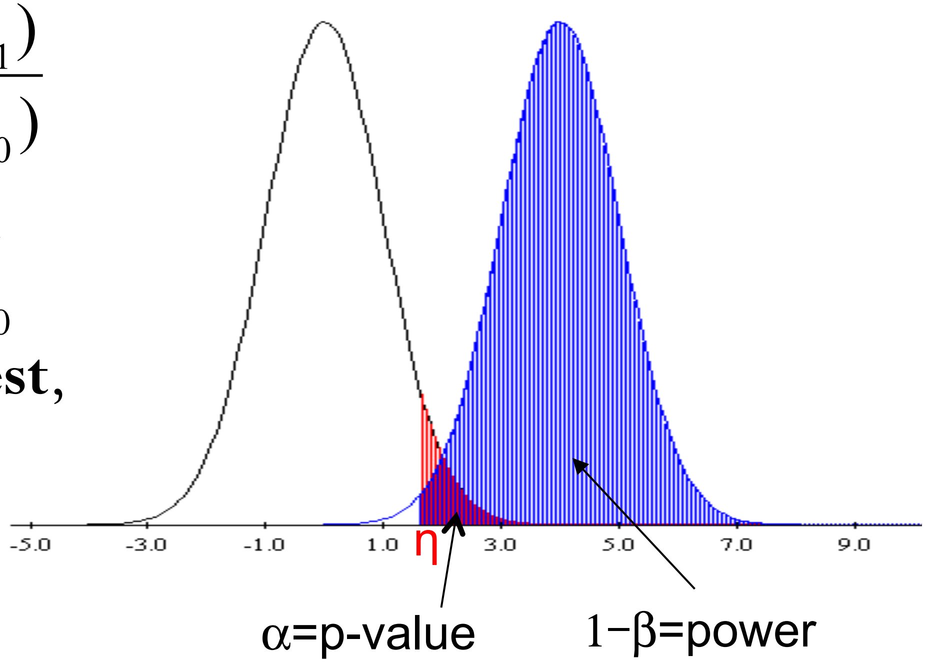
$$POWER = \mathrm{Prob}(reject\ H_{null} \mid H_{alt})$$

- $$POWER = 1 - p_{alt}$$



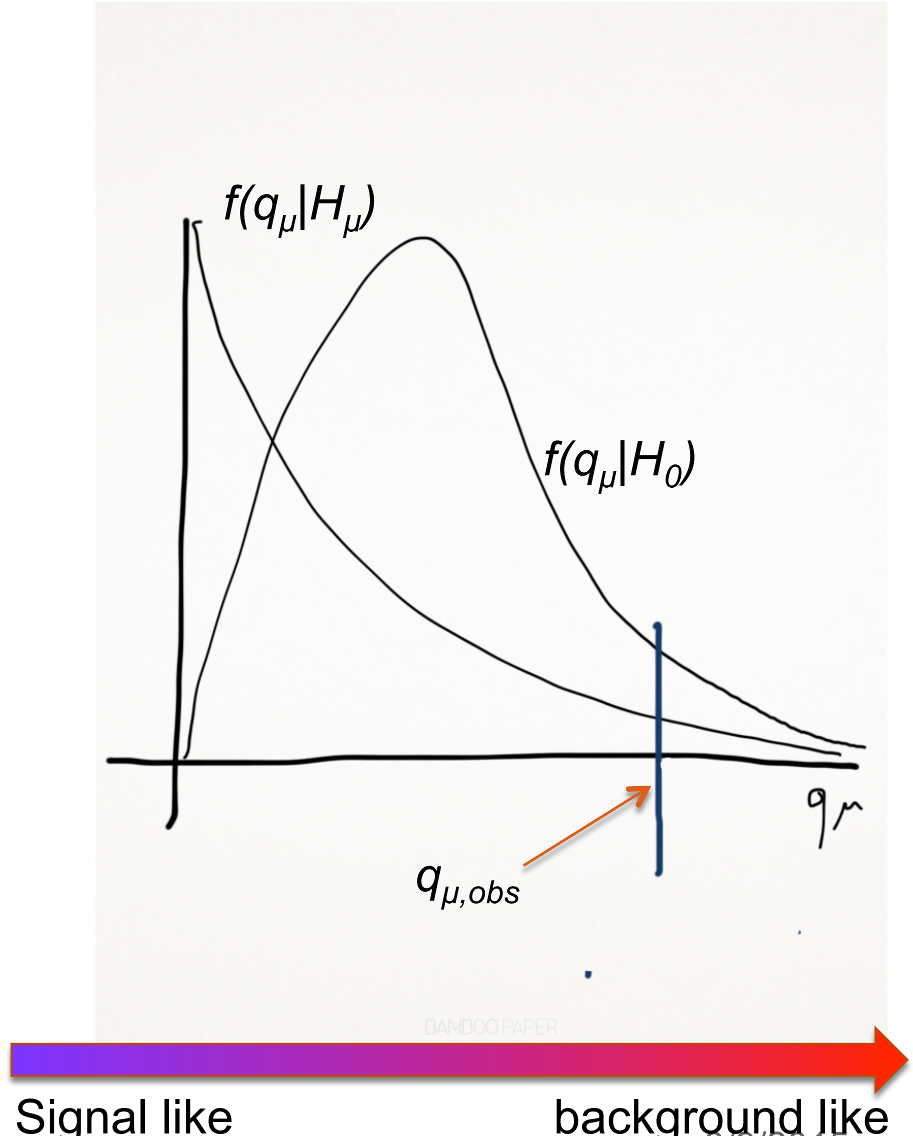
The Neyman-Pearson Lemma

- Define a **test statistic** $\lambda = \dfrac{L(H_1)}{L(H_0)}$
- When performing a hypothesis test between two simple hypotheses, H_0 and H_1, **the Likelihood Ratio test,** which rejects H_0 in favor of H_1, **is the most powerful test** of size α for a threshold η
- To find out which of two inference methods is better, find the one with the highest power with a given p-value.

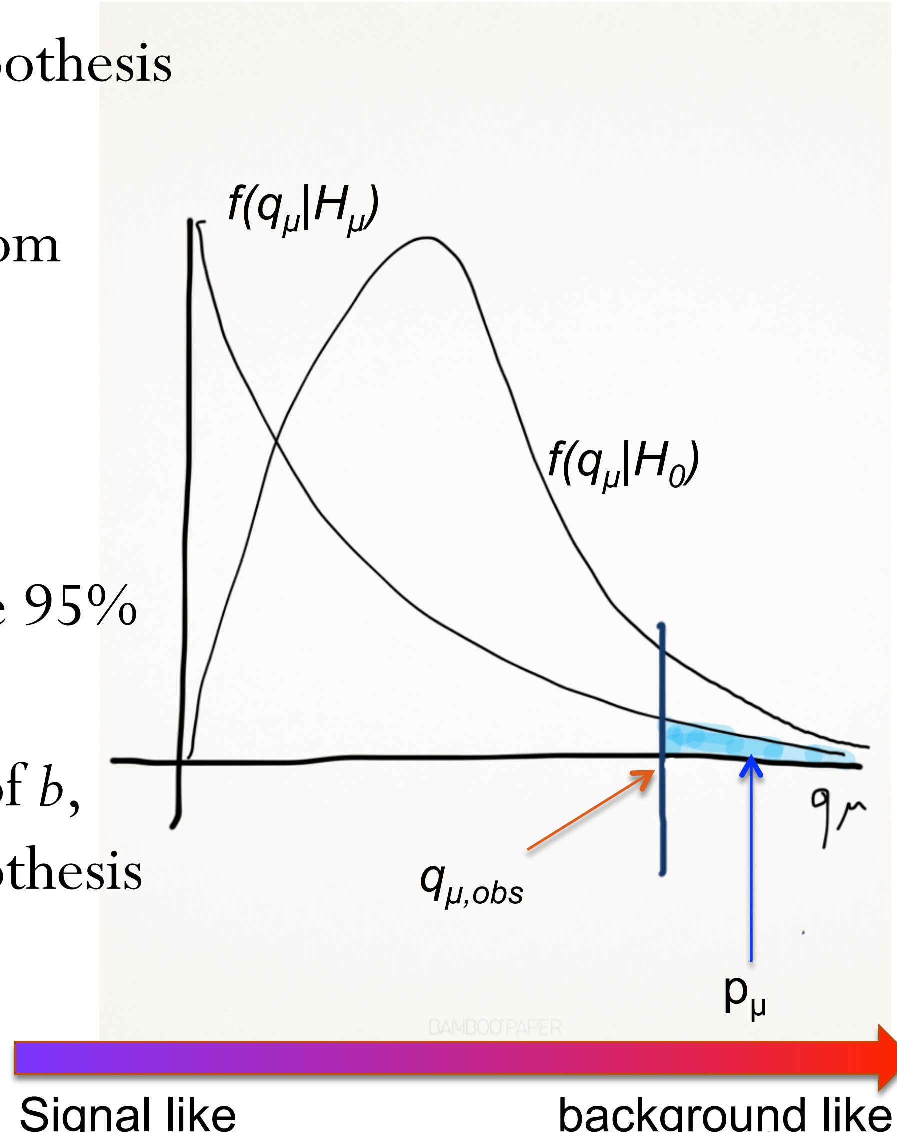


Exclusion and CLs

- We test hypothesis H_μ
- $n = \mu s(m) + b$
- We calculate the PL (profile likelihood) ratio with the one observed data
- $q_{\mu,obs}$

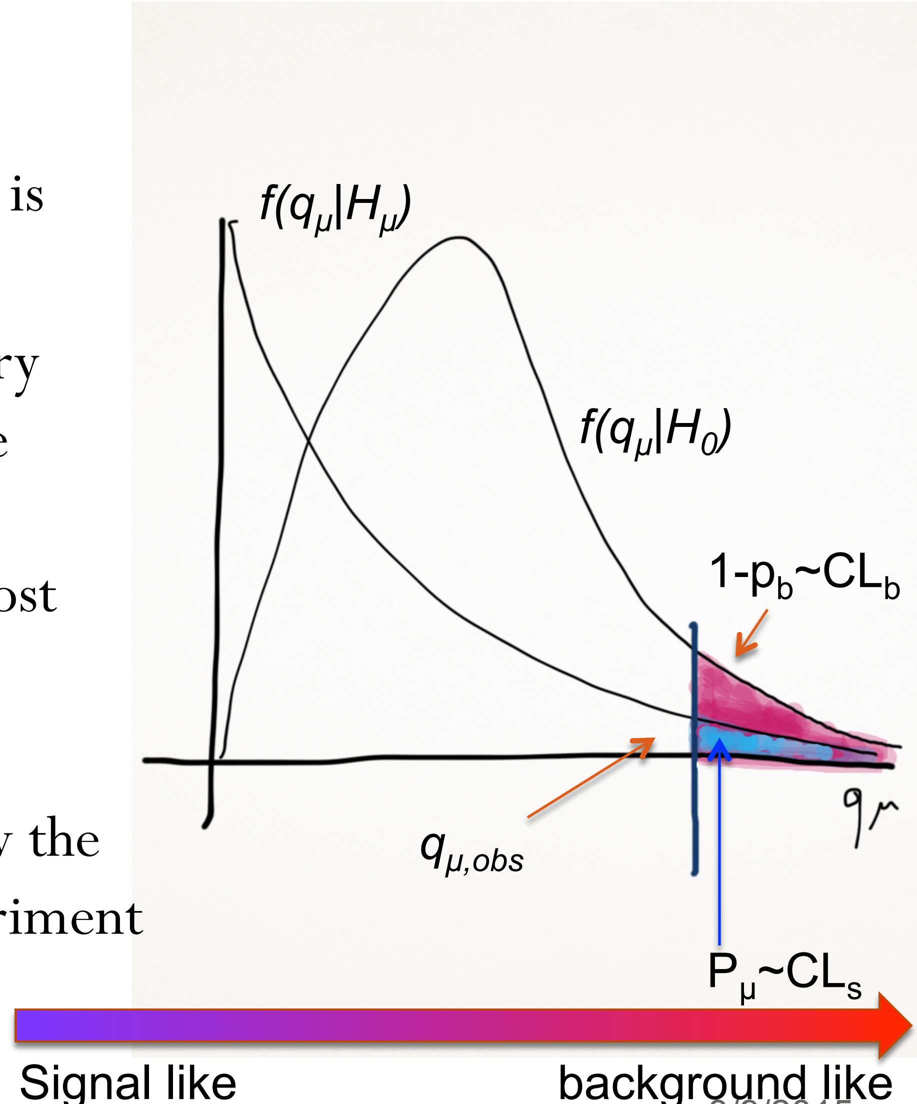


- Find the *p-value* of the signal hypothesis H_μ
- But one cannot separate s(m) from *s(m)+b*
- $$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid H_\mu)\,dq_\mu$$
- In principle if p_μ<5%, H_μ hypothesis is excluded at the 95% CL
- Could a downward fluctuation of *b*, drives a rejection of *s(m)+b* hypothesis (when *s* is small)?



CLs

- A complication arises when $\mu s+b \sim b$
- In these cases the Power of the test is very low (red area).
- When the signal cross section is very small the $s(m_H)+b$ hypothesis can be rejected but at the same time the background-only hypothesis is almost rejected as well due to downward fluctuations of the background
- These downward fluctuations allow the exclusion of a signal, that the experiment is not sensitive to



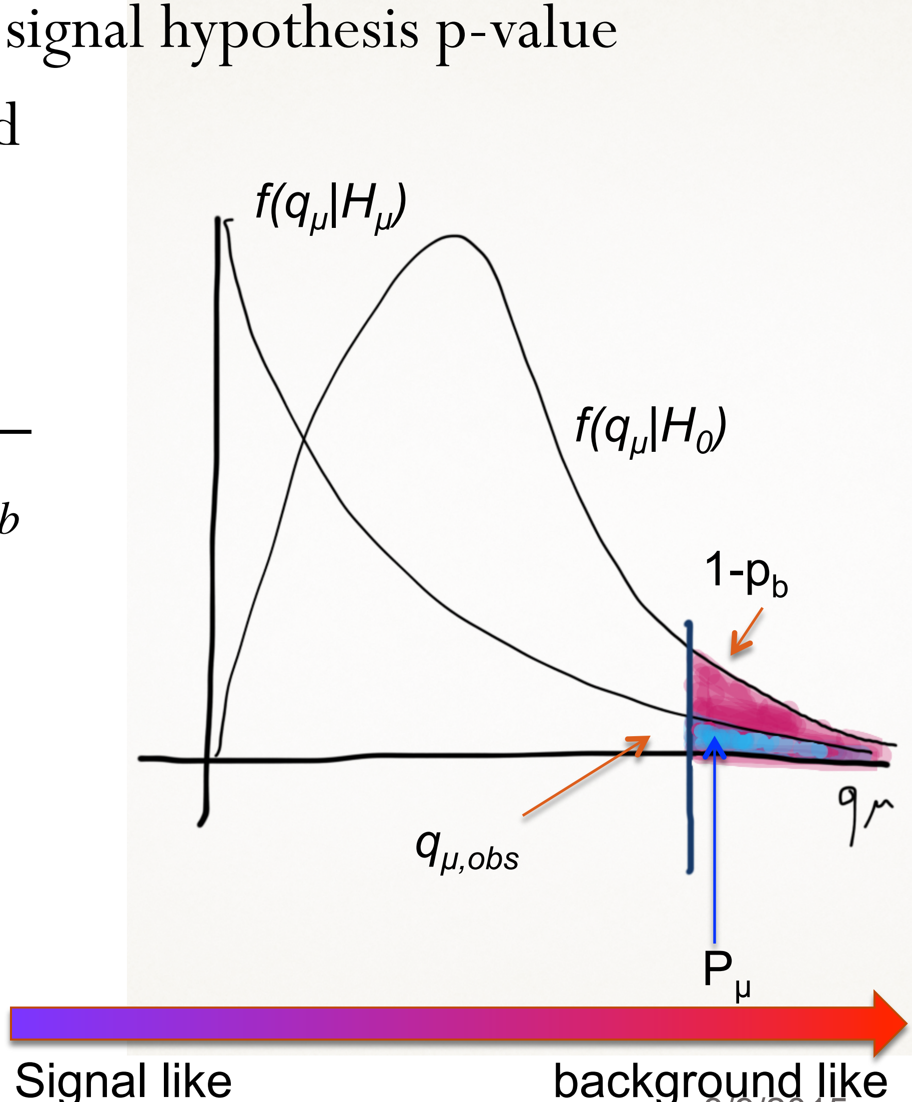
The Modified CLs with the PL test statistic

- The CLs method means that the signal hypothesis p-value p_μ is normalized by the power and is modified to

$$p_\mu \rightarrow p'_\mu = \frac{p_{null}}{1 - p_{alt}} = \frac{p_\mu}{1 - p_b}$$

$$p_\mu = \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|\mu)\,d\tilde{q}_\mu$$

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



Upper Limit

- Upper fluctuations of the signal do not serve as an evidence against the signal, use

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases} \qquad \lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}_\mu)}{L(\hat{\mu}, \hat{\theta})}$$

- Given a DATA set, For each hypothesized mass, m, scan μ and find

$$\mu \in [0, \mu_{up}] \Rightarrow$$

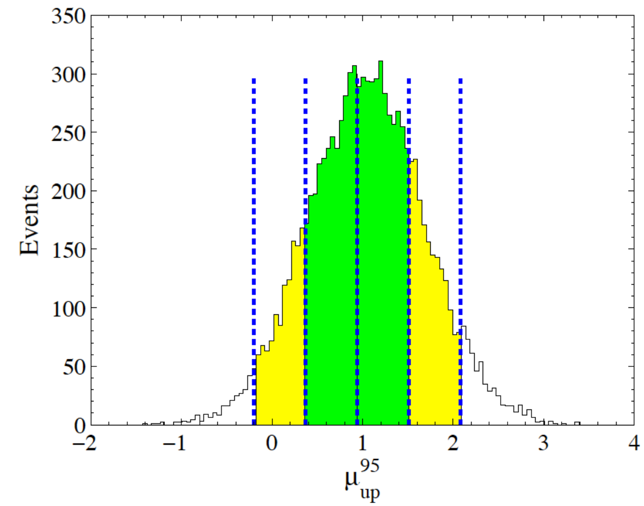
$$\forall \mu > \mu_{up} \Rightarrow p'_\mu = \frac{p_\mu}{1 - p_b} \le 5\% \Rightarrow$$

$$\mu > \mu_{up} \ excluded \ @ \ge 95\%CL$$

$$\Rightarrow \mu \in [0, \mu_{up}], 95\% \text{ Confidence Interval } \Rightarrow$$

$$\mu \le \mu_{up} \ @ \ge 95\%CL$$

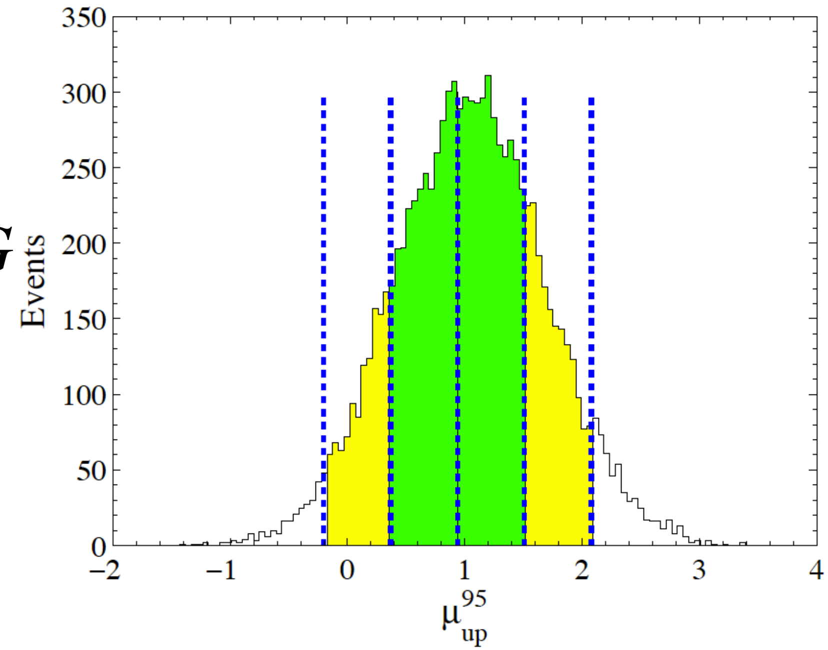
- If we generate toy BG only experiments, and for each finds μ_{up}, we find uncertainty bands on the upper limit



Sensitivity

- The sensitivity of an experiment to exclude a particle with a mass m is the median upper limit
- $\mu_{up}^{med} = med\{\mu_{up} \mid H_{alt}\};\ H_{alt} = BG$
- The 68% (green) and 95% (yellow) are the 1 and 2 σ bands
- The median and the bands can be derived with the Asimov background only dataset $n=b$

Distribution of the upper limit with background only experiments

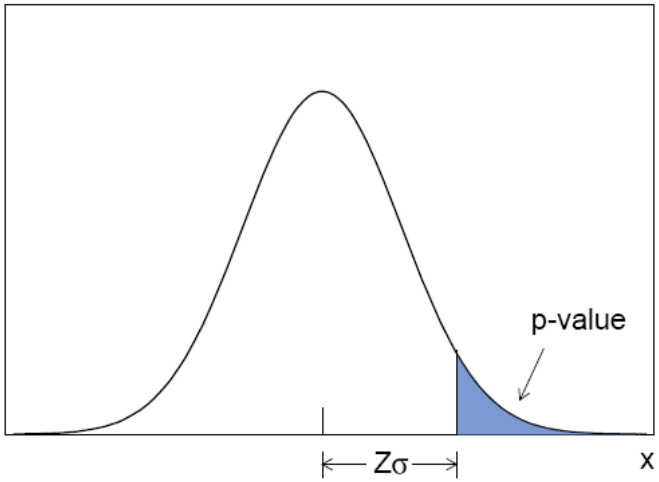


The Asimov data set is n=b
-> median upper limit

CCGV Useful Formulae

The power of asymototic: No need to scan

$$p'_{\mu_{95}} = \frac{1-\Phi(\sqrt{q_{\mu_{95}}})}{\Phi(\sqrt{q_{\mu_{95},A}} - \sqrt{q_{\mu_{95}}})} = 0.05$$



Φ is the cumulative distribution of the standard (zero mean, unit variance) Gaussian.

$q_{\mu_{95},A}$ Is evaluated with the Asimov data set (background only)

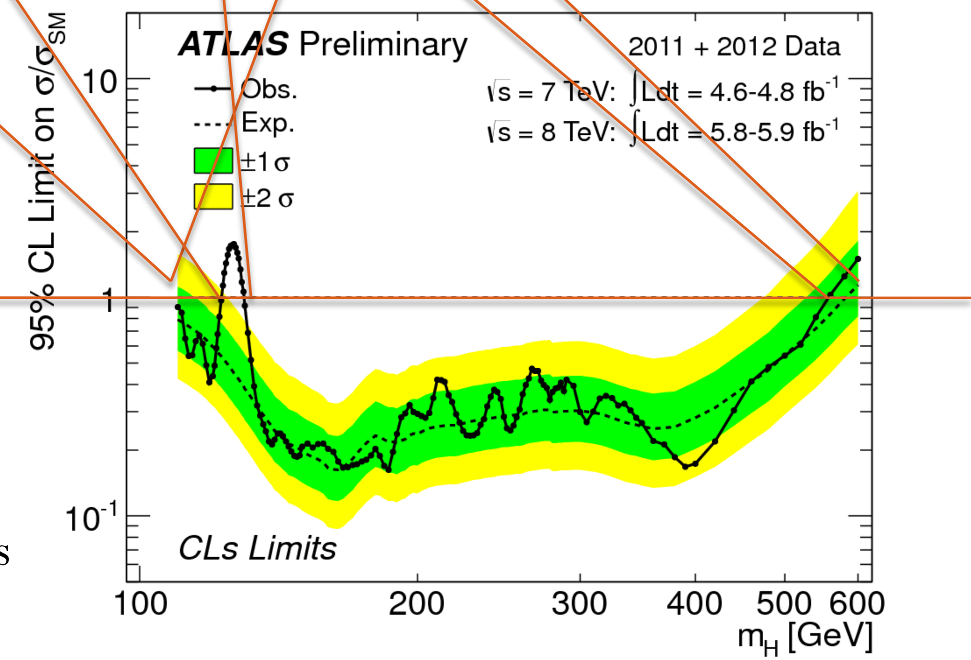
$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 1 - \Phi(Z)$$

$$Z = \Phi^{-1}(1-p)$$

Understanding the Brazil Plot

The expected 95% CL exclusion region covers the m_H range from 110 GeV to 582 GeV. The observed 95% CL exclusion regions are from 110 GeV to 122.6 GeV and 129.7 GeV to 558 GeV. The addition of

- $\mu_{up} = \sigma(m_H)/\sigma_{SM}(m_H) < 1 \rightarrow$
  $\sigma(m_H) < \sigma_{SM}(m_H) \rightarrow$ SM m_H excluded



- The line μ up=1 corresponds to CLs=5% (p'$_s$=5%)
- If μ up<1 the exclusion of a SM Higgs is deeper → p' s<5%,
  p' s=CLs → CL=1-p' s>95%

Asymptotic Distributions (CCGV)

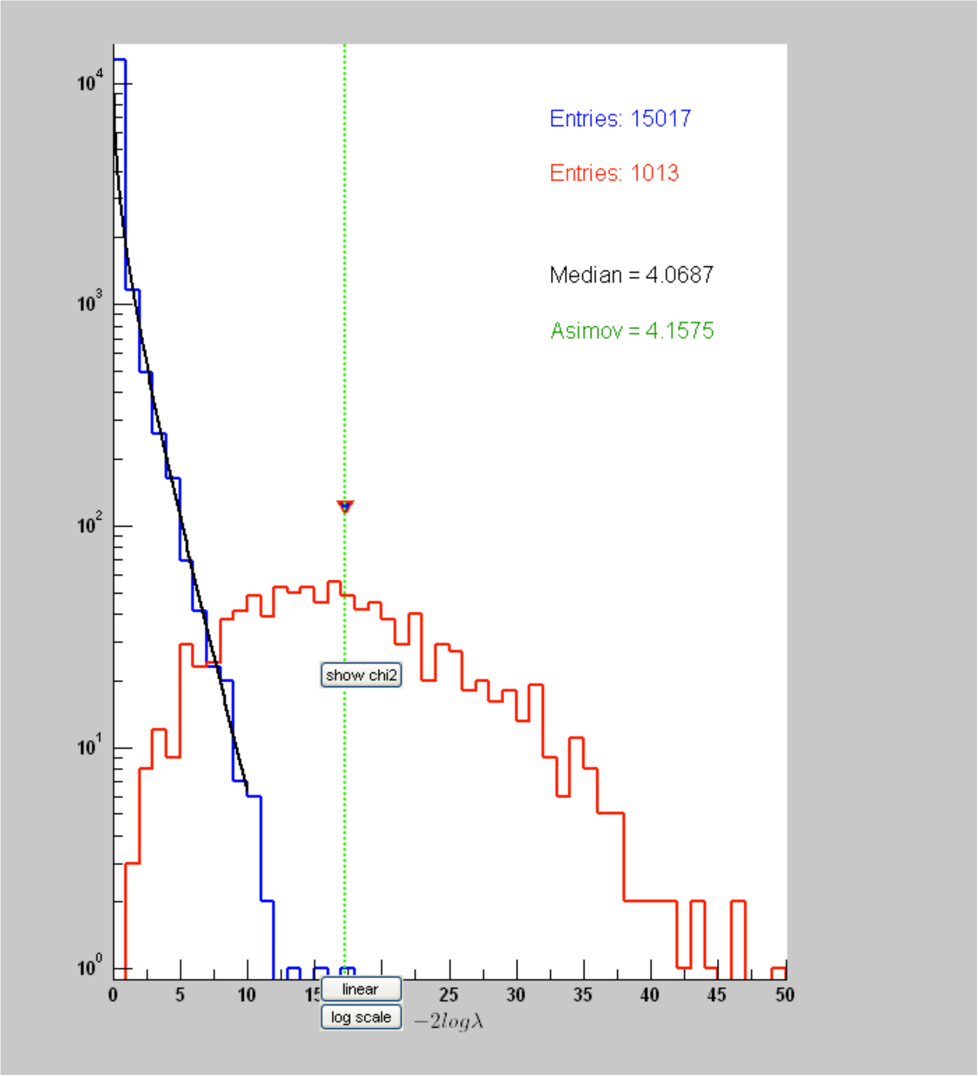
Tossing Monte Carlos to get the test statistic distribution functions (PDF) is sometimes beyond the experiment technical capability.

Knowing both PDF

$$f(q_{null} \mid H_{null})$$

$$f(q_{null} \mid H_{alternate})$$

enables calculating both the observed and expected significance (or exclusion) without a single toy….



Wald Theorem

- Consider a test of the strength parameter μ , which here can either be zero (for discovery) or nonzero (for an upper limit), and suppose the data are distributed according to a strength parameter μ'
- The desired distribution $f(q_\mu | \mu')$ can be found using a result due to Wald [1946], who showed that for the case of a single parameter of interest,

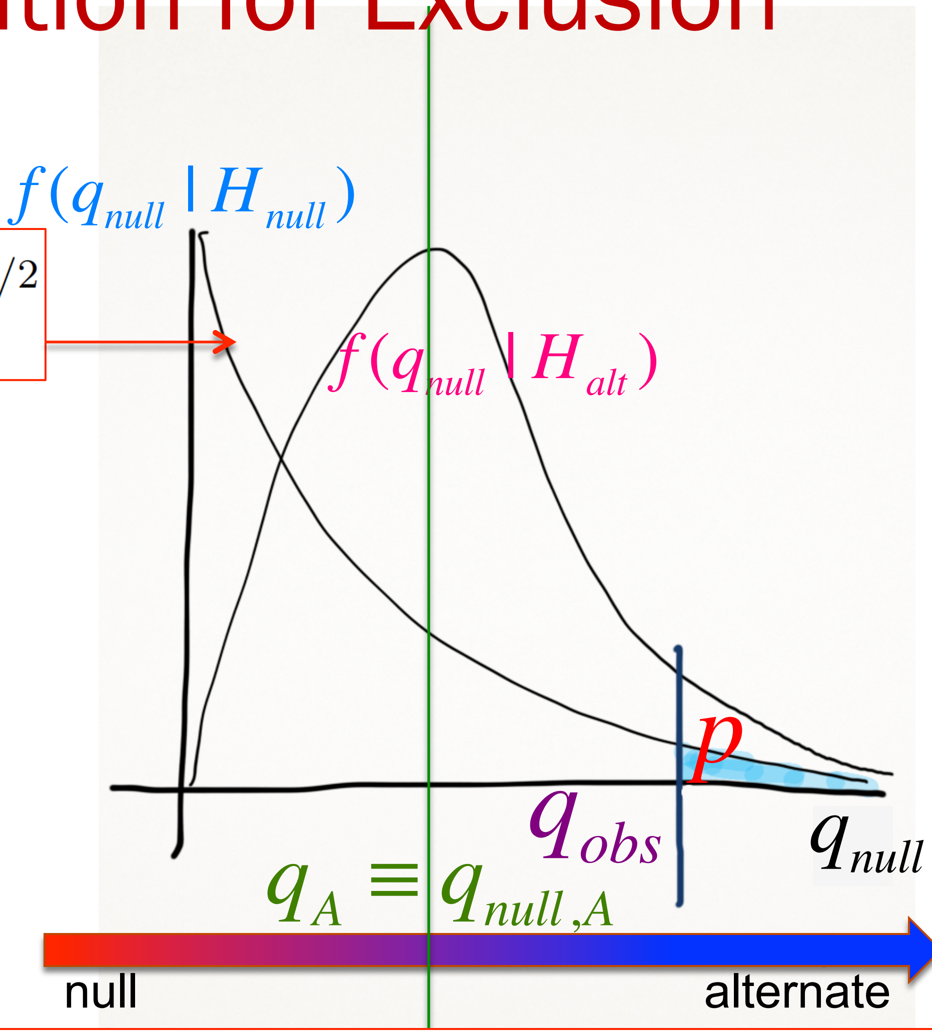
$$\boxed{-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})}$$

$$\boxed{\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})}}$$

$$\boxed{\langle\hat{\mu}\rangle = \mu'}$$

Asymptotic Distribution for Exclusion (CCGV)

$$f(q_\mu|\mu) = \frac{1}{2}\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-q_\mu/2}$$



1 sided CI

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}\exp\left[-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{\mu-\mu'}{\sigma}\right)^2\right]$$

DISCOVERY

Reminder: The test statistic

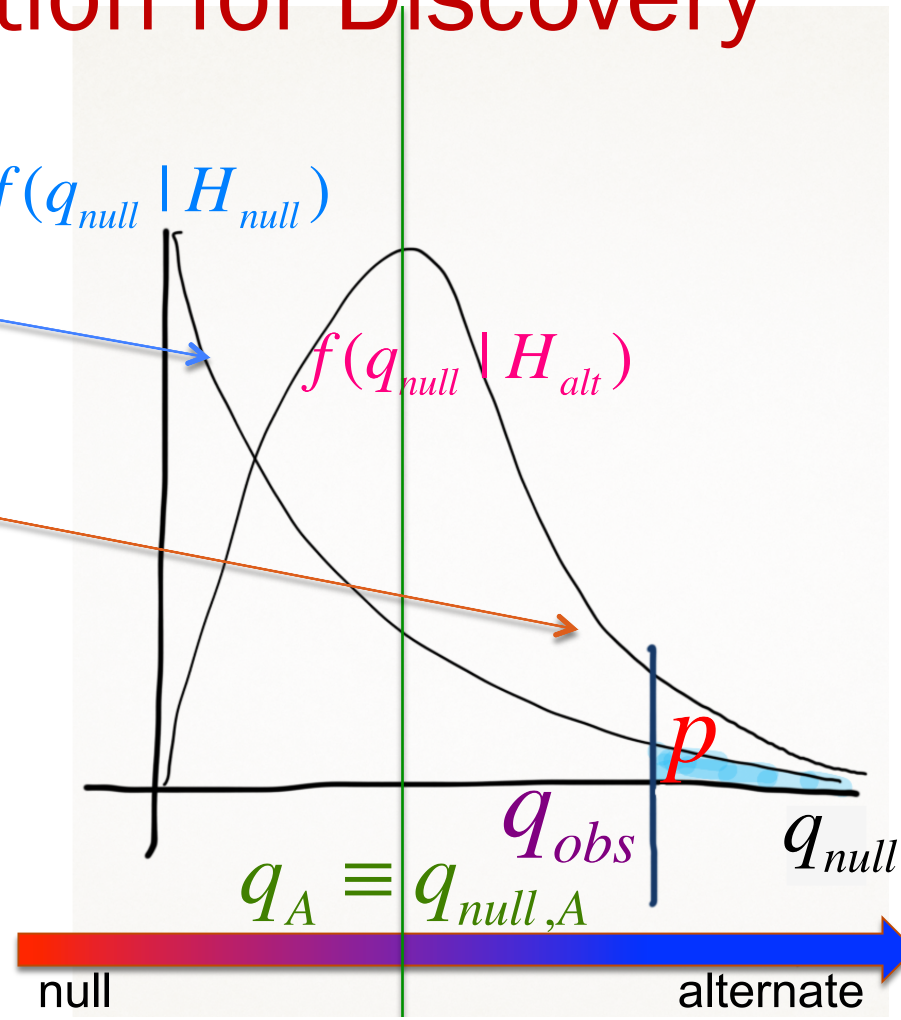
$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0\,, \\ 0 & \hat{\mu} < 0\,, \end{cases}$$

- Downward fluctuations of the background do not serve as an evidence against the background

Asymptotic Distribution for Discovery CCGV

$$f(q_0 \mid 0) \sim \frac{1}{2}\chi^2$$

$$f(q_0 \mid \mu') \sim ?$$



1 sided CI

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]$$

$$\Phi(Z) = 1 - \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

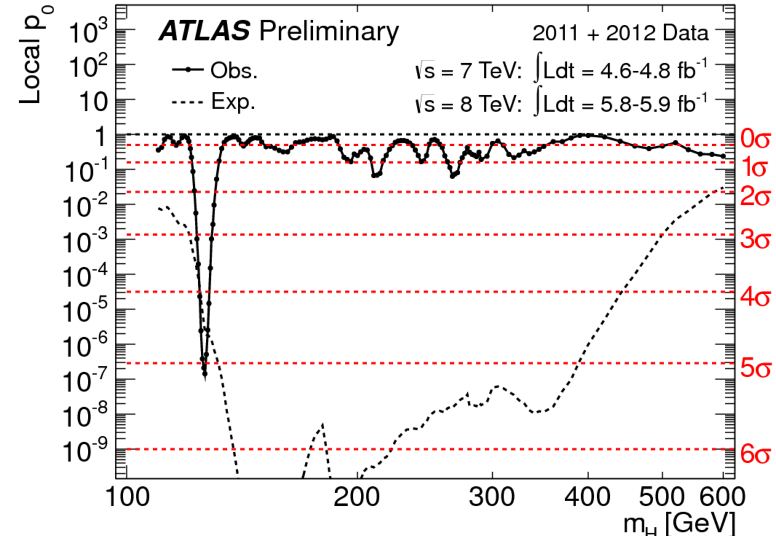
Significance - p_0

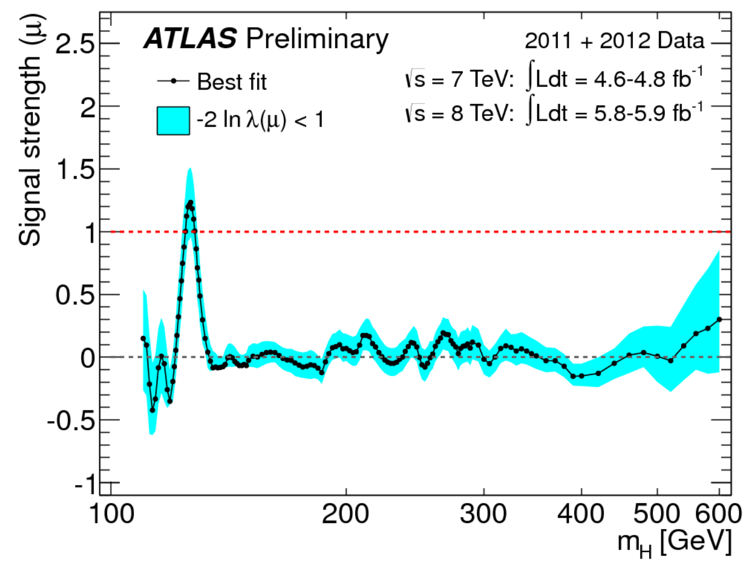
$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\,dq_0 \qquad Z = \Phi^{-1}(1-p)$$

Expected p0 is obtained with the Asimov (*s(m)*+*b*) DATA set

$$\mu(m) = \frac{\sigma(m)}{\sigma_{SM}(m)}$$

$$\hat{\mu}(m) = \text{MLE of } \mu$$





The New $s/\sqrt{b}$

The new $s/\sqrt{b}$

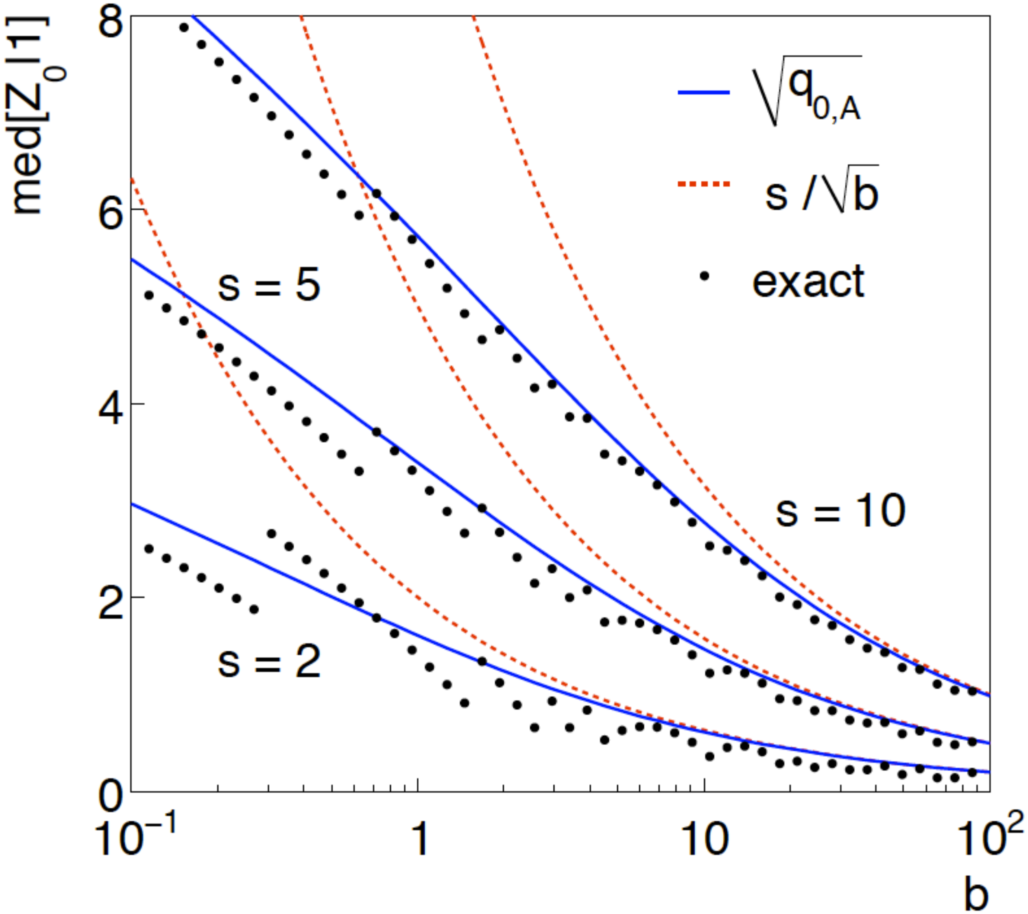
$$Z_A = \sqrt{q_{0,A}}$$

$$\mathrm{med}[Z_0|1] = \sqrt{q_{0,\mathrm{A}}} = \sqrt{2\left((s+b)\ln(1+s/b) - s\right)}$$

$$Z_A = \sqrt{q_{0,A}} \xrightarrow{s/b \ll 1} \frac{s}{\sqrt{b}} + O(s/b)$$

The New $s/\sqrt{b}$

$s/\sqrt{b}$?



The new $s/\sqrt{b}$

$$\mathrm{med}[Z_0|1] = \sqrt{q_{0,\mathrm{A}}} = \sqrt{2\left((s+b)\ln(1+s/b) - s\right)}$$

Taking Background Systematics into Account

- The NEW $$Z \approx s / \sqrt{b + \sigma_b^2} \xrightarrow{L \to \infty} \frac{s}{\sigma_b}$$
  $$\sigma_b = \Delta \cdot b$$

$$\boxed{Z_{\mathrm{A}} = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}}$$

- Expanding the Asimov formula in powers of s/b and $\sigma_b{}^2/b$ gives

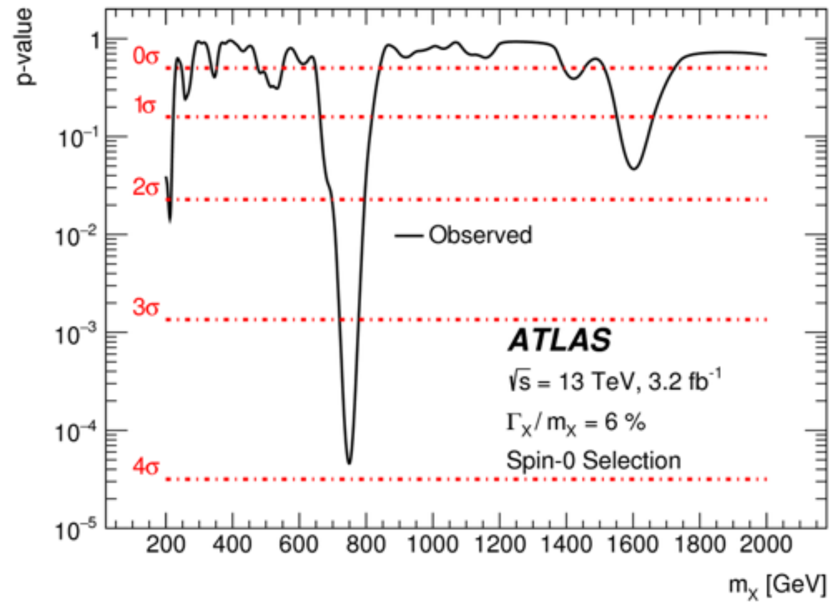
$$Z_{\mathrm{A}} = \frac{s}{\sqrt{b + \sigma_b^2}} \left(1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b) \right)$$

Last , yEt , not lEast

Local vs Global Significance-
The Look Elsewhere Effect

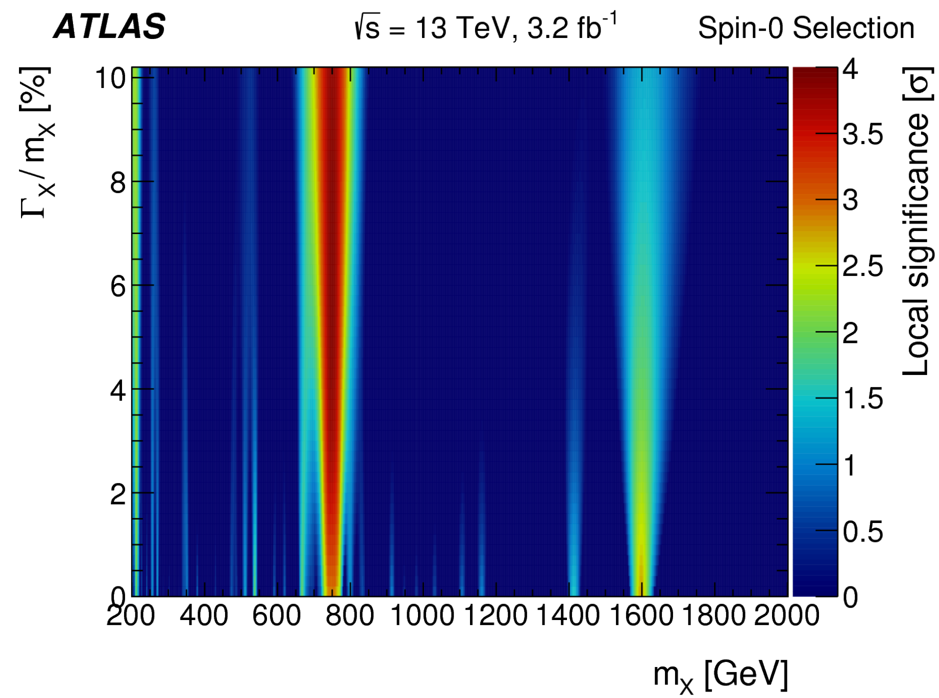
Local vs Global Significance

- There is a clear ~4 Sigma "signal" here
  - Local significance: The probability to observe a fluctuation at a GIVEN mass, m_X
  - Global Significance: The probability to observe a fluctuation of >4 sigma, at SOME mass, m_x in the search range



LEE in Multidimensions

- There is a clear ~4 Sigma "signal" here
  - Local significance: The probability to observe a fluctuation at a GIVEN mass, m_X
  - Global Significance: The probability to observe a fluctuation of >4 sigma, at SOME mass, m_x in the search range, and with SOME width



Look Elsewhere Effect (LEE) in multidimensions

More details
Statistics Session, this afternoon

CONCLUSIONS

CONCLUSIONS

- The Higgs search and discovery (since LEP days in the 1990s) has advanced the Frequentist Statistical methods beyond the foreseen horizon (at the time)
- Asymptotic methods exist to avoid tedious time consuming. sometimes impossible toys generations
- Better understanding of local vs global significance has been achieved (see talk this afternoon in statistics session)
- Methods were used to search, set limits, discover, measure and finally combine ATLAS and CMS reS. Gadatsch sults with ~4000 Nuisance Parameters (see talk by S. Gadatsch on Friday statistics session)
- The ROOSTATS framework allows to do it all fast, relatively smooth and simple

http://root.cern.ch/root/html/tutorials/roostats/index.html

Statistics Session, This afternoon and Friday afternoon

13:00 - 14:30

- Combination measurements and the BLU...
- Confidence intervals for the ratio of two quantities
- Look Elsewhere Effect in 2D
- A method for the construction ...
- Empty

16:30 - 16:50

- Deep Learning and Bayesian Methods
- Bayesian n parametric modelling o...
- Unfolding techniques in Particle Physics
- Experience with using ...
- Classifiers for centrality determination...

13:30 - 15:00

- *Kai Yi* Statistical and other experie...
- Statistical combination of experimental results in...
- QCD multijet background modelling by ...
- The Thermal Model of Heavy Ion Collisions

17:00 - 17:30

- Statistical significance ...
- The inverse bagging algo...
- The Matrix Element Method in the...
- Machine Learning Software Development...
- Wrapping up

BACKUP

Significance with BG systematics

Taking Background Systematics into Account

- The intuitive explanation of $s/\sqrt{b}$ is that it compares the signal, s, to the standard deviation of n assuming no signal, $\sqrt{b}$.
- Now suppose the value of b is uncertain, characterized by a standard deviation σ_b.
- A reasonable guess is to replace $\sqrt{b}$ by the quadratic sum of $\sqrt{b}$ and σ_b, i.e.,

$$b \pm \Delta \cdot b \Rightarrow \sigma_b = \sqrt{\left(\sqrt{b}\right)^2 + (\Delta \cdot b)^2} = \sqrt{b + \Delta^2 b^2}$$

$$s/\sqrt{b} \Rightarrow s/\sqrt{b(1 + b\Delta^2)} \xrightarrow{L \to \infty} \frac{s/b}{\Delta}$$

$$\frac{s/b}{\Delta} \geq 5 \to s/b \geq 0.5 \text{ for } \Delta \sim 10\%$$

If s/b<0.5 we will never be able to make a discovery

But even that formula can be omproved using the Asimov formalism

Significance with systematics

- We find (G. Cowan)

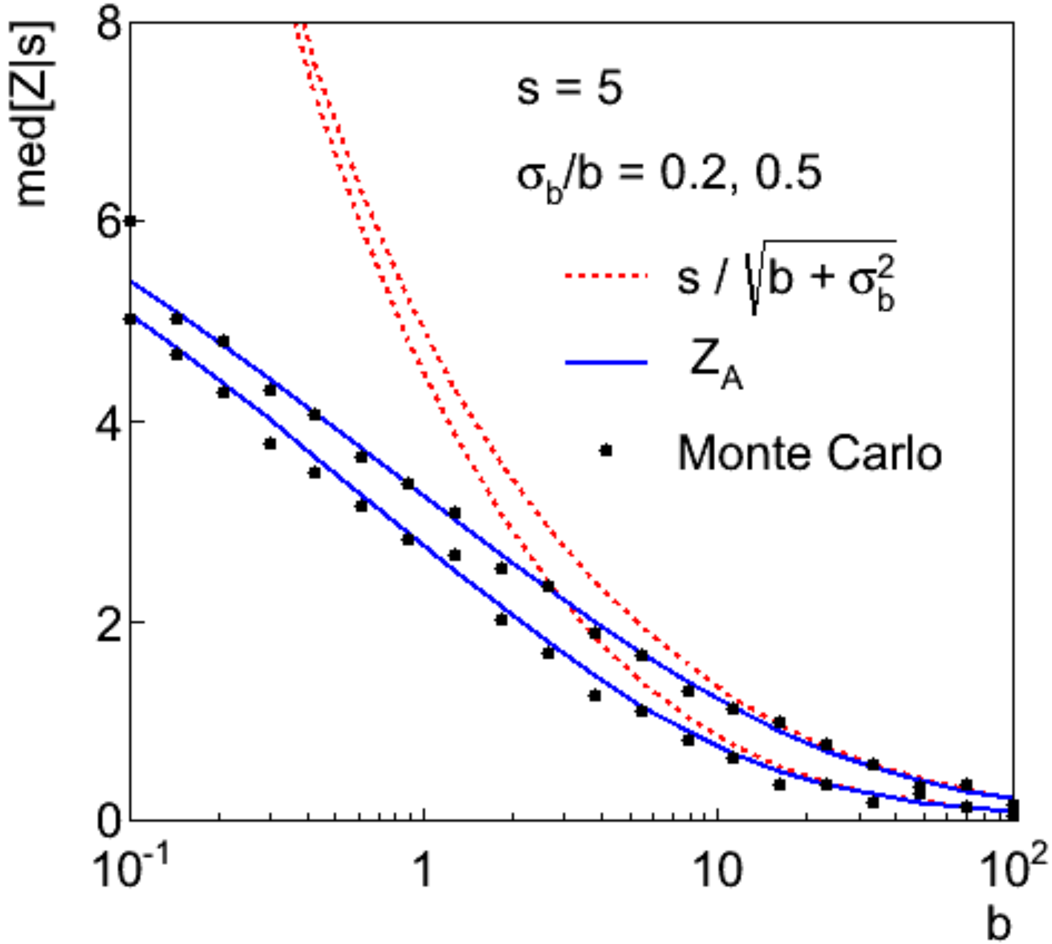
$$Z_{\mathrm{A}} = \left[2\left((s+b)\ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)\right]^{1/2}$$

Expanding the Asimov formula in powers of s/b and $\sigma_b{}^2/b$ gives

$$Z_{\mathrm{A}} = \frac{s}{\sqrt{b+\sigma_b^2}}\left(1+\mathcal{O}(s/b)+\mathcal{O}(\sigma_b^2/b)\right)$$

- So the "intuitive" formula can be justified as a limiting case of the significance from the profile likelihood ratio test evaluated with the Asimov data set.

Significance with systematics



Asymptotic, how to determine Sigma and the Bands

How to determine σ

- To estimate the uncertainty σ there are a few possibilities
  - Given the asymptotic formulae, fit the distribution of

$$f(q_{null} \mid H_{alt}) = f(q_\mu \mid \mu')$$ and extract σ

  - Implement the Wald formula to the Asimov data set and find

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{q_{\mu,A}}$$

where μ is the tested (null) hypothesis and μ ' is the alt hypothesis. For discovery, μ =0 while for exclusion μ '=0.

CCGV Useful Formulae – The Bands

$$\mu_{up}^{med} = \sigma\Phi^{-1}(1-0.5\alpha) = \sigma\Phi^{-1}(0.975)$$

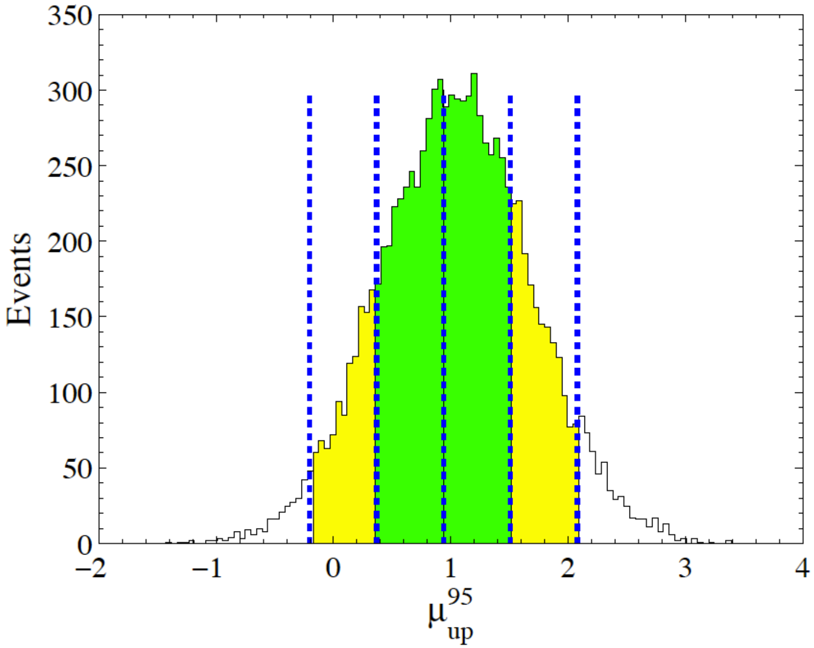
$$\sigma_{\hat{\mu}}^2 = Var[\hat{\mu}]$$

$$\mu_{up+N} = N\sigma_0 + \sigma_{\mu_{up+N}}(\Phi^{-1}(1-\alpha\Phi(N))$$

$$\alpha = 0.05$$

$$\sigma_{\mu_{up+N}}^2 = \frac{\mu_{up+N}^2}{q_{\mu_{up+N},A}}$$

Distribution of the upper limit with background only experiments



The Asimov data set is n=b
-> median upper limit

Power os an inference test

Basic Definitions: POWER

- $p_{null} = \Pr ob(reject\ H_{null} \mid H_{null}) \sim typeI$
- $p_{alt} = \Pr ob(reject\ H_{alt} \mid H_{alt}) = \Pr ob(accept\ \ H_0 \mid H_{alt}) \sim typeII$
- The POWER of an hypothesis test is the probability to reject the null hypothesis when the alternate analysis is true!
- $$POWER = \mathrm{Prob}(reject\ H_{null} \mid H_{alt})$$
  $$p_{alt} = \Pr ob(reject\ H_{alt} \mid H_{alt}) \Rightarrow$$
  $$1 - p_{alt} = \Pr ob(reject\ H_{null} \mid H_{alt}) \Rightarrow$$
  $$POWER = 1 - p_{alt}$$
- The power of a test increases as the rate of type II error decreases

Confidence Interval and Confidence Level (CL)

CL & CI - Wikipedia

$$\mu = 1.1 \pm 0.3$$

$$\mu = [0.8, 1.4] \ @ \ 68\% \ CL$$

CI=[0.8,1.4]

what does it mean?

- A **confidence interval** (**CI**) is a particular kind of interval estimate of a population parameter.
- Instead of estimating the parameter by a single value, an interval likely to include the parameter is given.
- How likely the interval is to contain the parameter is determined by the **confidence level** or confidence coefficient.
- Increasing the desired confidence level will widen the confidence interval.

Confidence Interval & Coverage

- Say you have a measurement μ_{meas} of μ with μ_{true} being the unknown true value of μ
- Assume you know the probability distribution function $p(\mu_{meas} | \mu)$
- Given the measurement you deduce somehow (based on your statistical model) that there is a 95% Confidence interval $[\mu_1, \mu_2]$.

  (it is 95% likely that the μ_{true} is in the quoted interval)

The correct statement:

- In an ensemble of experiments 95% of the obtained confidence intervals will contain the true value of μ.

Upper limit

- Given the measurement you deduce somehow (based on your statistical model) that there is a 95% Confidence interval $[0, \mu_{up}]$.
- This means: In an ensemble of experiments 95% of the obtained confidence intervals will contain the true value of μ, including $\mu=0$ (no Higgs)
- We therefore deduce that $\mu < \mu_{up}$ at the 95% Confidence Level (CL)
- μ_{up} is therefore an upper limit on μ
- If $\mu_{up}<1 \rightarrow$
  $\sigma(m_H)<\sigma_{SM}(m_H) \rightarrow$
  a SM Higgs with a mass m_H is excluded at the 95% CL

Confidence Interval & Coverage

- Confidence Level: A CL of (e.g.) 95% means that in an ensemble of experiments, each producing a confidence interval, 95% of the confidence intervals will contain the true value of μ
- Normally, we make one experiment and try to estimate from this one experiment the confidence interval at a specified CL
- If in an ensemble of (MC) experiments our estimated Confidence Interval fail to contain the true value of μ 95% of the cases (for every possible μ) we claim that our method undercover
- If in an ensemble of (MC) experiments our estimated Confidence Interval contains the true value of μ more than 95% of the cases (for every possible μ) we claim that our method overcover (being conservative)
- If in an ensemble of (MC) experiments the true value of μ is covered within the estimated confidence interval , we claim a coverage

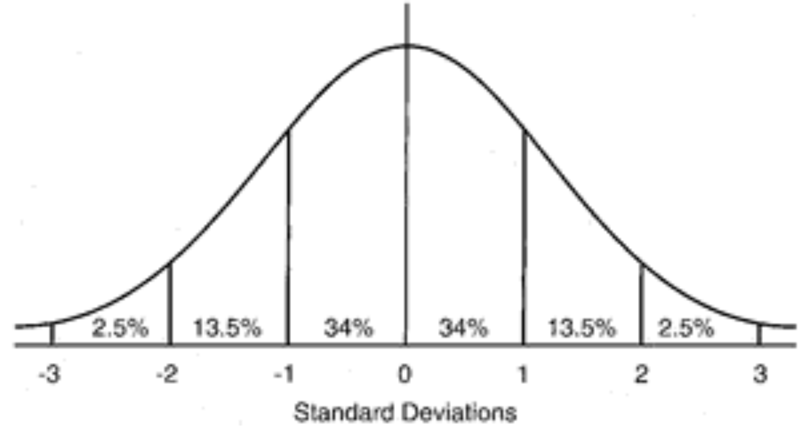
How to deduce a CI?

- One can show that if the data is distributed normal around the average i.e. P(data | μ)=normal

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

then one can construct a 68% CI around the estimator of μ to be

$$\hat{x} \pm \sigma$$



However, not all distributions are normal, many distributions are even unknown and coverage might be a real issue

How to deduce a CI?

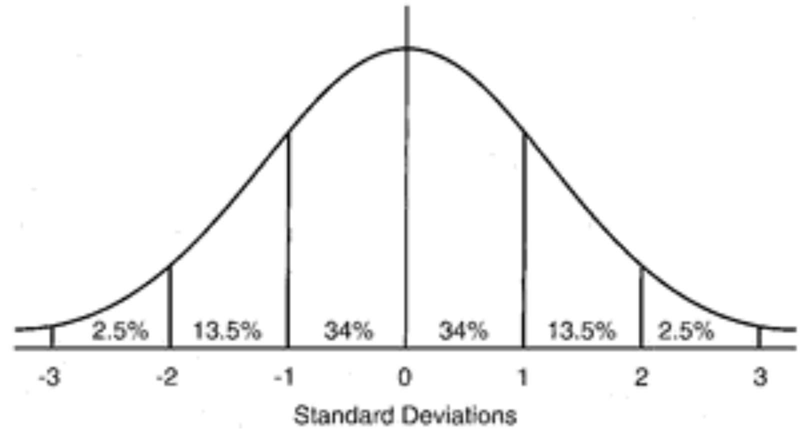
- One can show that if the data is distributed normal around the average i.e. P(data | μ)=normal

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

then one can construct a 68% CI around the estimator of μ to be

$$\hat{x} \pm \sigma$$

However, not all distributions are normal, many distributions are even unknown and coverage might be a real issue



- One may construct many 68% intervals.... $CI = [\mu_L, \mu_U]$

$$\int_{\mu_L}^{\mu_U} f(x \mid \hat{x})dx = 68\%$$

- Which one has a full coverage?
- How can we guarantee a coverage

- The QUESTION is NOT how to construct a CI, it is
- HOW TO CONSTRUCT A CI WHICH HAS A COVERAGE @ THE 68% CL