Search and Discovery Statistics in HEP

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, Bob Cousins Glen Cowan ,Kyle Cranmer
Ofer Vitells & Jonathan Shlomi

What can you expect from the Lectures

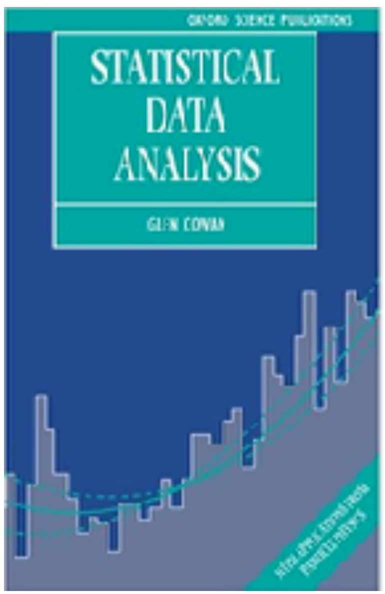
- Lecture 1: Basic Concepts
  Histograms, PDF, Testing Hypotheses,
  LR as a Test Statistics, p-value, POWER, CLs
  Measurements
- Lecture 2: Feldman-Cousins, Wald Theorem,
  Asymptotic Formalism, Asimov Data Set, PL & CLs
- Lecture 3: Asimov Significance
  Look Elsewhere Effect
  1D LEE the non-intuitive thumb rule
  (upcrossings, trial #~Z)
  2D LEE (Euler Characteristic)

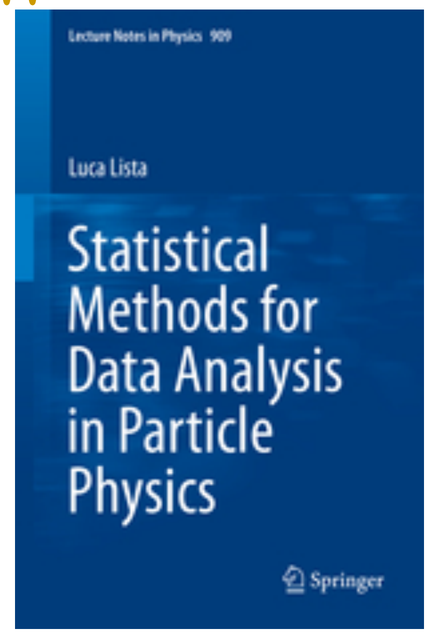
Support Material

G. Cowan, Statistical Data Analysis, Clarendon Press, Oxford, 1998.
PDG

L. Lista Statistical methods for Data Analysis, 2nd Ed. Springer, 2018

G. Cowan PDG
http://pdg.lbl.gov/2017/reviews/rpp2017-rev-statistics.pdf





Preliminaries

In a Nut Shell

The binomial distribution with parameters n and p
is
the discrete probability distribution of the number of successes in a sequence of n independent experiments.
(Wikipedia)

$$P(k : n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

If X ~ B(n, p)

$$E[X] = np$$

$$P(k:n,p)=\binom{n}{k}p^{k}(1-p)^{n-k}$$

The Poisson distribution with parameter λ = np can be used as an approximation to B(n, p) of the binomial distribution if n is sufficiently large and p is sufficiently small.

$$P(k:n,p)\xrightarrow{n\to\infty,np=\lambda} Poiss(k;\lambda)=\frac{\lambda^{k}e^{-k}}{k!}$$

$$If\ X\sim Poiss(k;\lambda)$$

$$E[X]=Var[X]=\lambda$$

From Binomial to Poisson to Gaussian

$$P(k:n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(k:n,p) \xrightarrow{n\to\infty, np=\lambda} Poiss(k;\lambda) = \frac{\lambda^k e^{-k}}{k!}$$

$$\langle k \rangle = \lambda, \ \sigma_k = \sqrt{\lambda}$$

$$k \to \infty \Rightarrow x = k$$

Using Stirling Formula

$$\text{prob(x)=G(x,}\sigma = \sqrt{\lambda}) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\lambda)^2/2\sigma^2}$$

This is a Gaussian, or Normal distribution

with mean and variance of λ

Histograms

$N\ collisions$

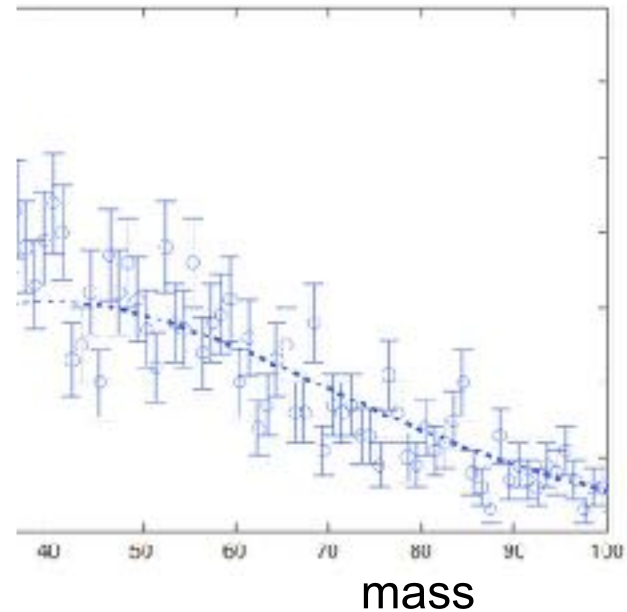
$$p(Higgs\ event) = \frac{\mathcal{L}\sigma(pp \to H)\, A\epsilon_{ff}}{\mathcal{L}\sigma(pp)}$$

$\text{Prob to see } n_H^{obs} \text{ in } N \text{ collisions is}$

$$P(n_H^{obs}) = \begin{pmatrix} N \\ n_H^{obs} \end{pmatrix} p^{n_H^{obs}} (1-p)^{N-n_H^{obs}}$$

$$\ell im_{N\to\infty} P(n_H^{obs}) = Poiss(n_H^{obs}, \lambda) = \frac{e^{-\lambda}\lambda^{n_H^{obs}}}{n_H^{obs}!}$$

$$\lambda = Np = \mathcal{L}\sigma(pp) \cdot \frac{\mathcal{L}\sigma(pp \to H)\, A\epsilon_{ff}}{\mathcal{L}\sigma(pp)} = n_H^{exp}$$

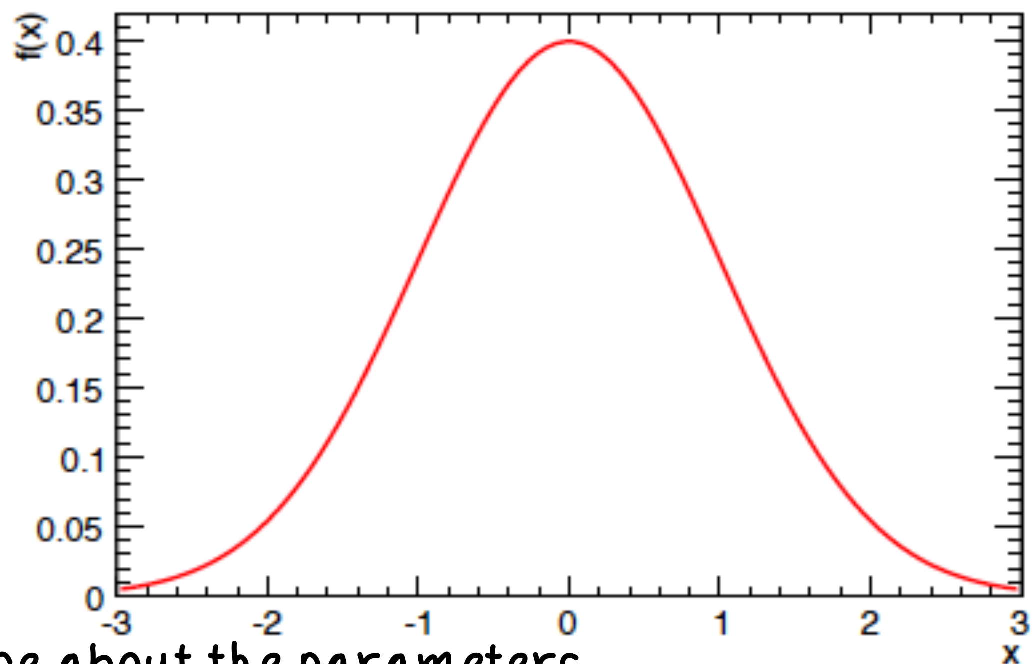


pdf

X is a random variable
Probability Distribution Function
PDF

$$P(x \in [x, x+dx]) = f(x)dx$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$



f(x) is not a probability
f(x)dx is a probability

$G(x|\mu, \sigma)$
Is a parametrized pdf (μ, σ)

We would like to make inference about the parameters

A counting experiment

- The Higgs hypothesis is that of signal $s(m_H)$

$$s(m_H) = L\sigma_{SM} \cdot A \cdot \epsilon$$

For simplicity unless otherwise noted $s(m_H) = L\sigma_{SM}$

- In a counting experiment $n = \mu s(m_H) + b$

$$\mu = \frac{L\sigma_{obs}(m_H)}{L\sigma_{SM}(m_H)} = \frac{\sigma_{obs}(m_H)}{\sigma_{SM}(m_H)}$$

- μ is the strength of the signal (with respect to the expected Standard Model one)

- The hypotheses are therefore denoted by H_μ

- H_1 is the SM with a Higgs, H_0 is the background only model

A Tale of Two Hypotheses

NULL

ALTERNATE

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favor of the alternative hypothesis

A Tale of Two Hypotheses

NULL — H_0 - SM w/o Higgs

ALTERNATE — H_1 - SM with Higgs

- Test the Null hypothesis and try to reject it
- Fail to reject it OR reject it in favor of the alternative hypothesis

A Tale of Two Hypotheses

NULL

H_0 - SM w/o Higgs

ALTERNATE

H_1 - SM with Higgs

- Reject H_0 in favor of H_1 – A DISCOVERY

We quantify rejection by p-value (later)

Swapping Hypotheses → exclusion

NULL

H_0 - SM w/o Higgs

ALTERNATE

H_1 - SM with Higgs

- Reject H_1 in favor of H_0

Excluding $H_1(m_H)$ → Excluding the Higgs with a mass m_H

We quantify rejection by p-value (later)

Likelihood

- Likelihood is the compatibility of the Hypothesis with a given data set.
  But it depends on the data

  Likelihood is <u>not</u> the probability of the hypothesis given the data

$$L(H) = L(H \mid x)$$

$$L(H \mid x) = P(x \mid H)$$

Bayes Theorem

$$P(H \mid x) = \frac{P(x \mid H) \cdot P(H)}{\Sigma_H P(x \mid H) P(H)}$$

$$P(H \mid x) \approx P(x \mid H) \cdot P(H)$$

Prior

Frequentist vs Bayesian

- The Bayesian infers from the data using priors

posterior $P(H\,|\,x) \approx P(x\,|\,H)\cdot P(H)$

- Priors is a science on its own.
  Are they objective? Are they subjective?

- The Frequentist calculates the probability of an hypothesis to be inferred from the data based on a large set of hypothetical experiments
  Ideally, the frequentist does not need priors, or any degree of belief while the Baseian posterior based inference is a "Degree of Belief".

- However, NPs (Systematic) inject a Bayesian flavour to any Frequentist analysis

Likelihood is NOT a PDF

- A Poisson distribution describes a discrete event count n for a real valued Meμ.

$$Pois(n|\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

Say, we observe n_o events

What is the likelihood of μ?

The likelihood of μ is given by

$L(\mu) = Pois(n_o | \mu)$

It is a continues function

of μ but it is NOT a PDF

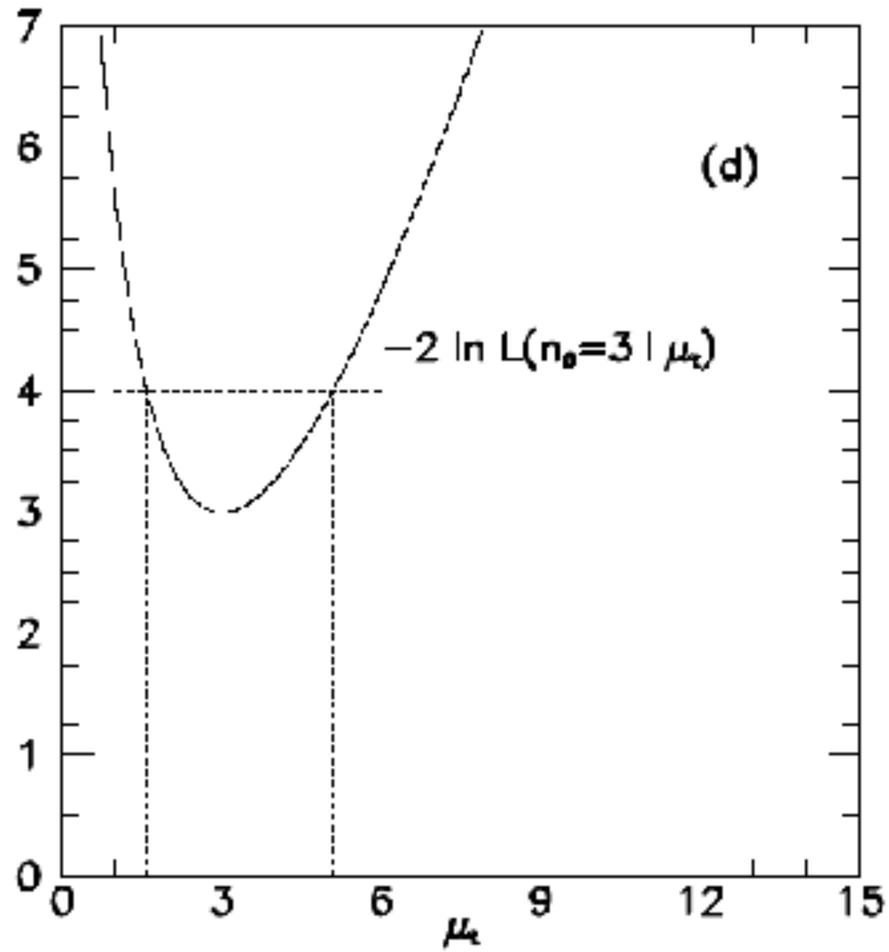


Figure from R. Cousins, Am. J. Phys. 63 398 (1995)

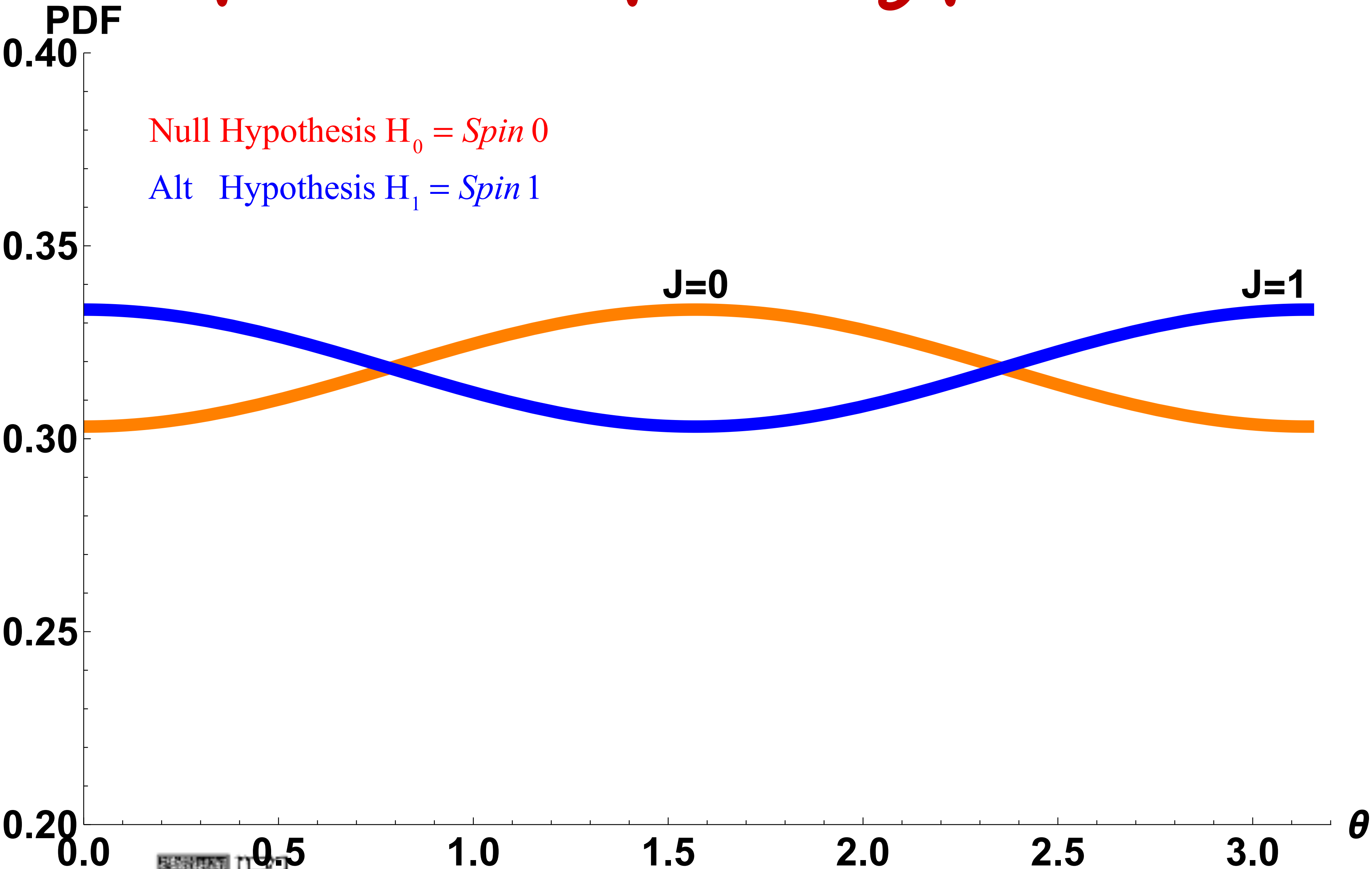
Testing an Hypothesis (wikipedia…)

- The first step in any hypothesis test is to state the relevant null, H_0 and alternative hypotheses, say, H_1
- The next step is to define a test statistic, q, under the null hypothesis
- Compute from the observations the observed value q_{obs} of the test statistic q.
- Decide (based on q_{obs}) to either
  fail to reject the null hypothesis or
  reject it in favor of an alternative hypothesis
- next: How to construct a test statistic, how to decide?

Test statistic and p-value

Case Study 1 : Spin

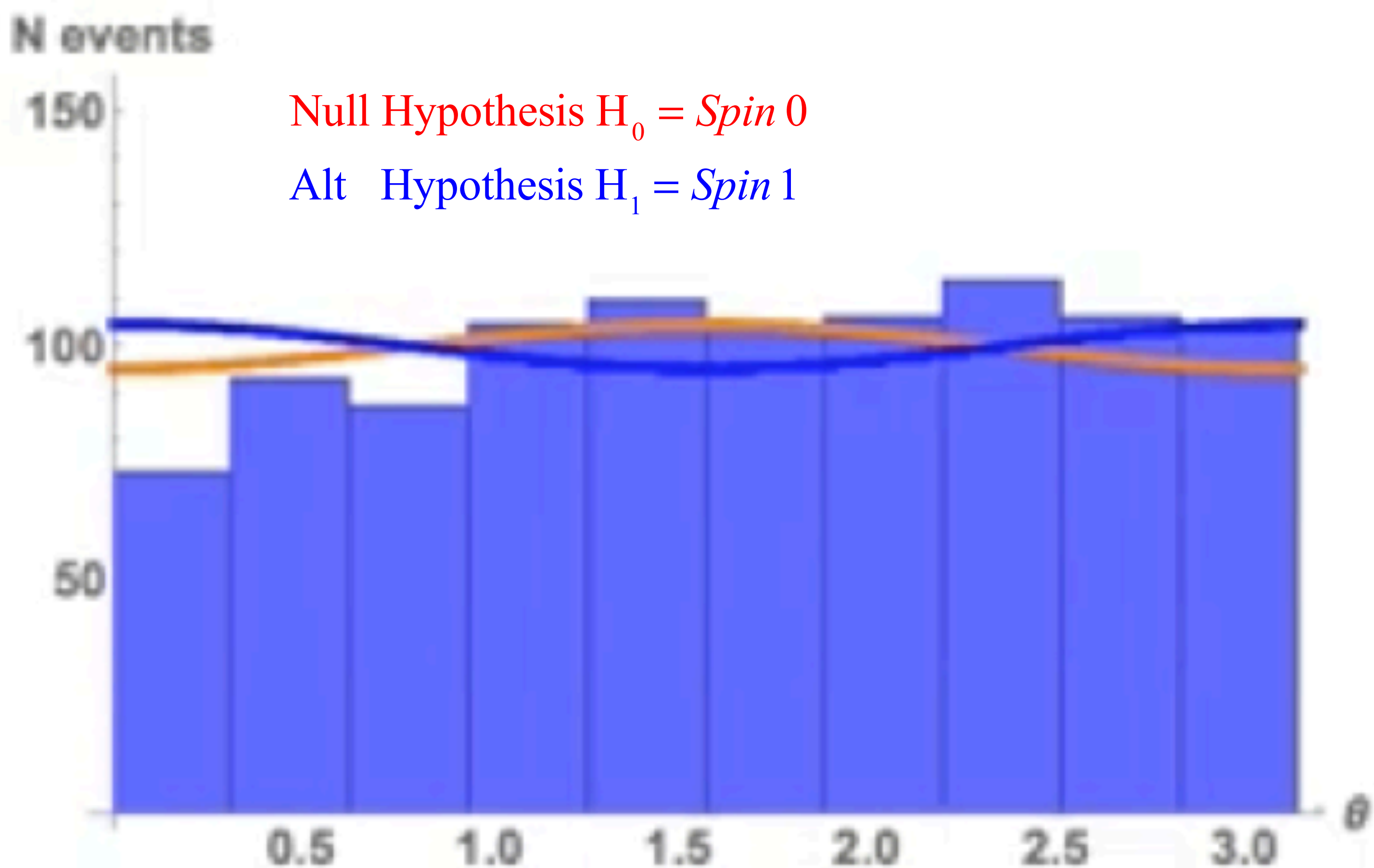
Spin 0 vs Spin 1 Hypotheses



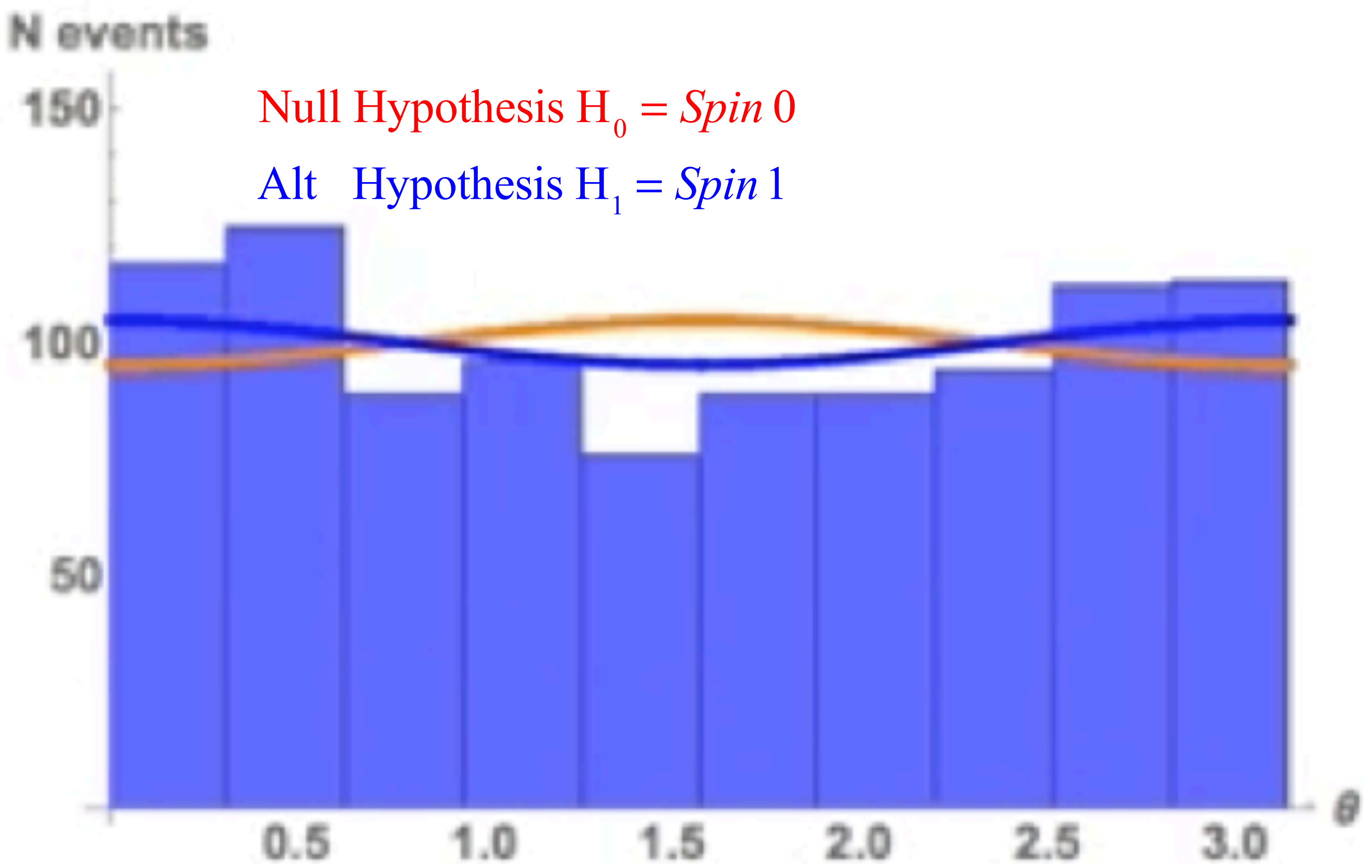
Spin 0 vs Spin 1 Hypotheses

Null Hypothesis $H_0 = \mathit{Spin}\ 0$

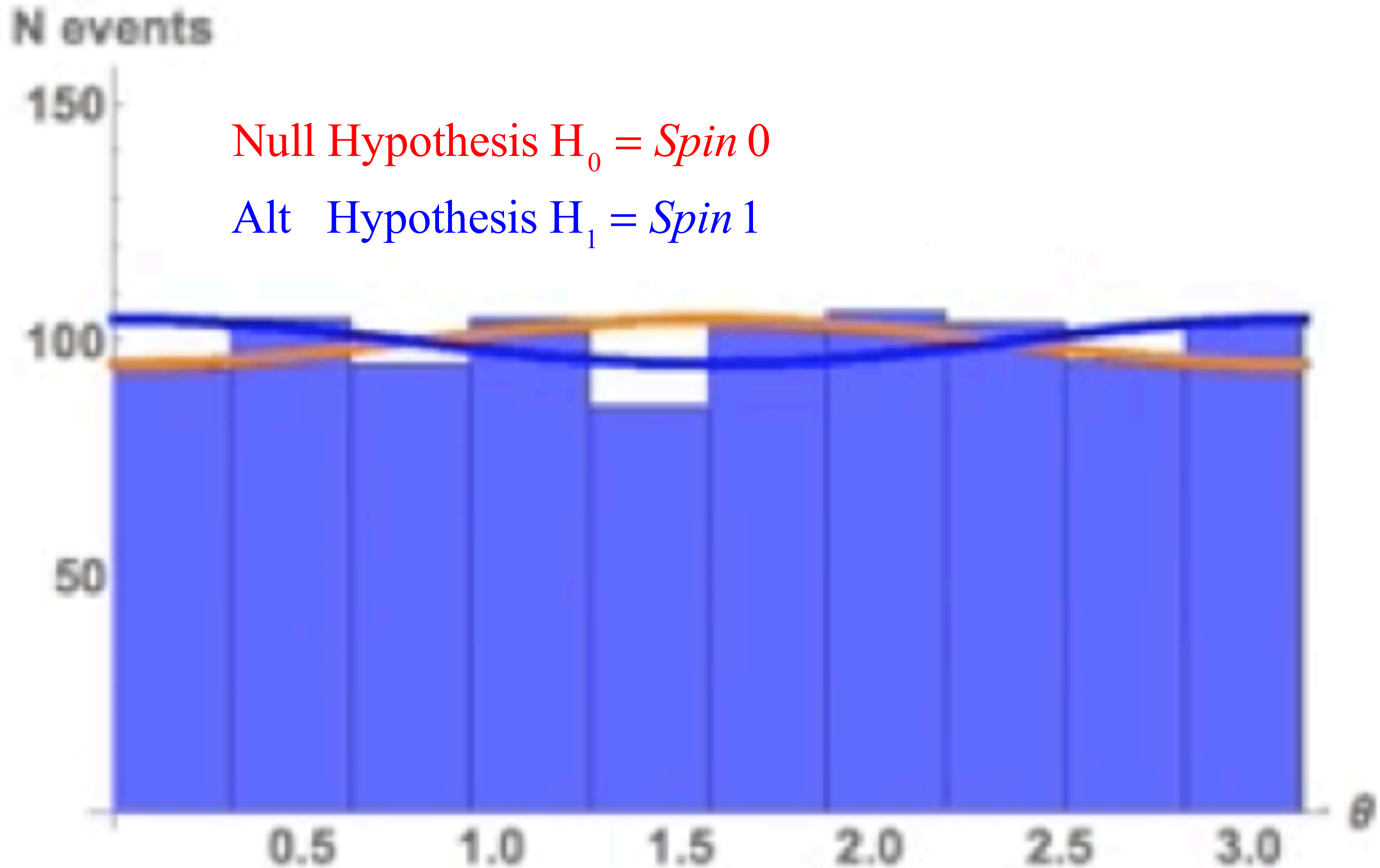
Alt Hypothesis $H_1 = \mathit{Spin}\ 1$



Spin 0 vs Spin 1 Hypotheses



Spin 0 vs Spin 1 Hypotheses



The Neyman-Pearson Lemma

- Define a test statistic $\lambda = \frac{L(H_1)}{L(H_0)}$
- When performing a hypothesis test between two simple hypotheses, H_0 and H_1, the Likelihood Ratio test, $\lambda = \frac{L(H_1)}{L(H_0)}$

  which rejects H_0 in favor of H_1,

  is the **most powerful test**

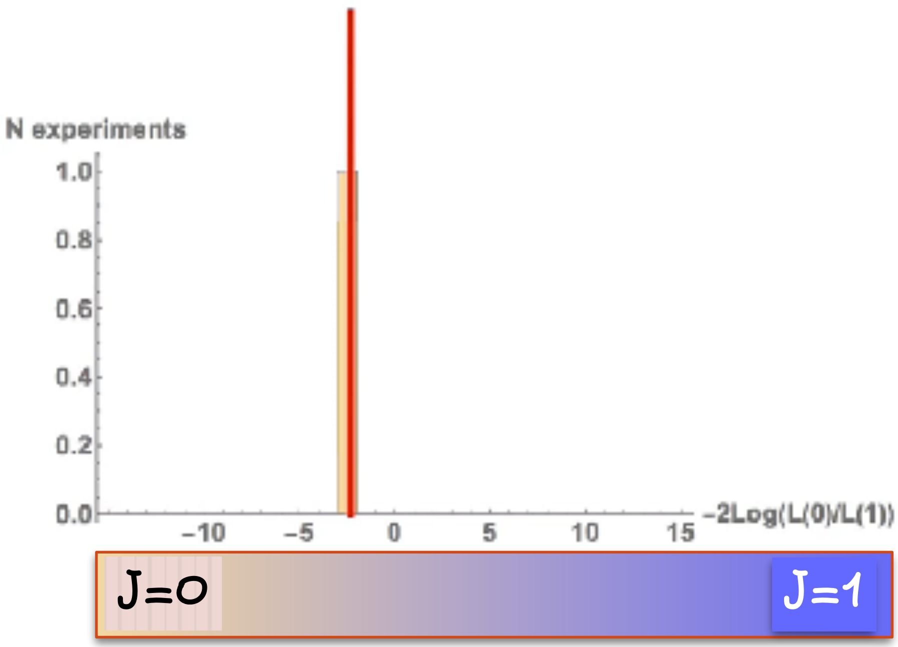
  for a given significance level $\alpha = prob(\lambda \leq \eta)$

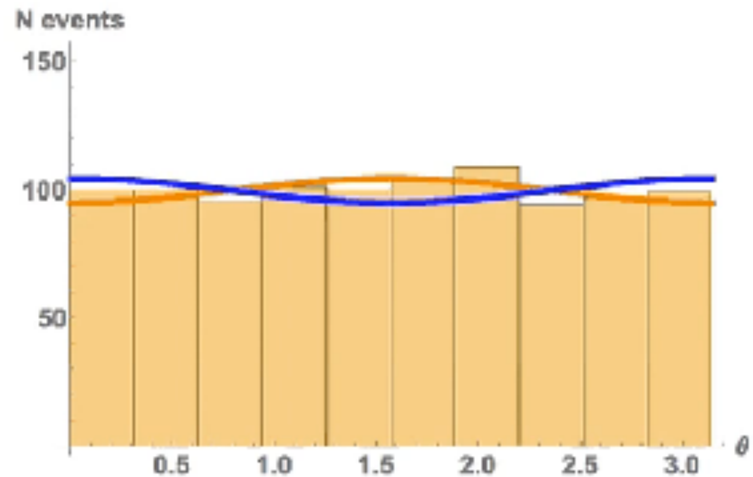
  with a threshold η

Building PDF

Build the pdf of the test statistic

$$q_{NP} = q_{NP}(x) = -2\ln\frac{L(H_0\,|\,x)}{L(H_1\,|\,x)}$$

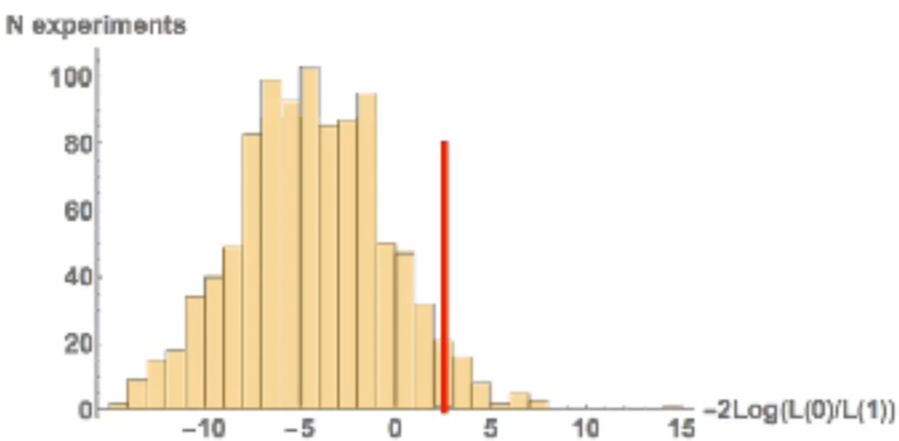


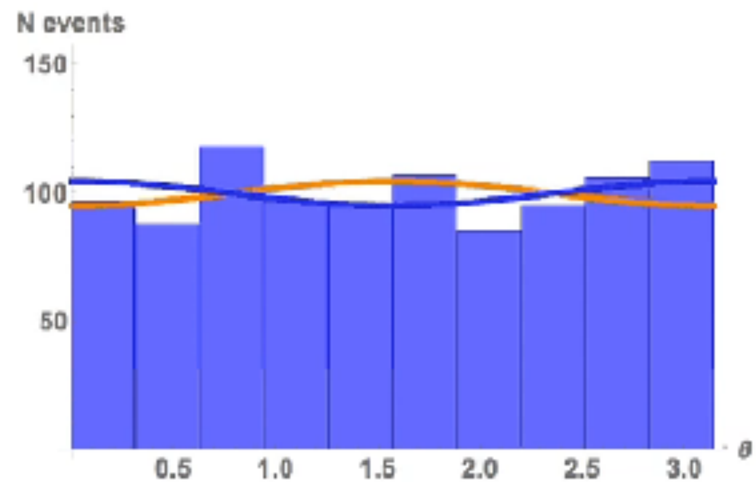


Building PDF

Build the pdf of the test statistic

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





Basic Definitions: type I-II errors

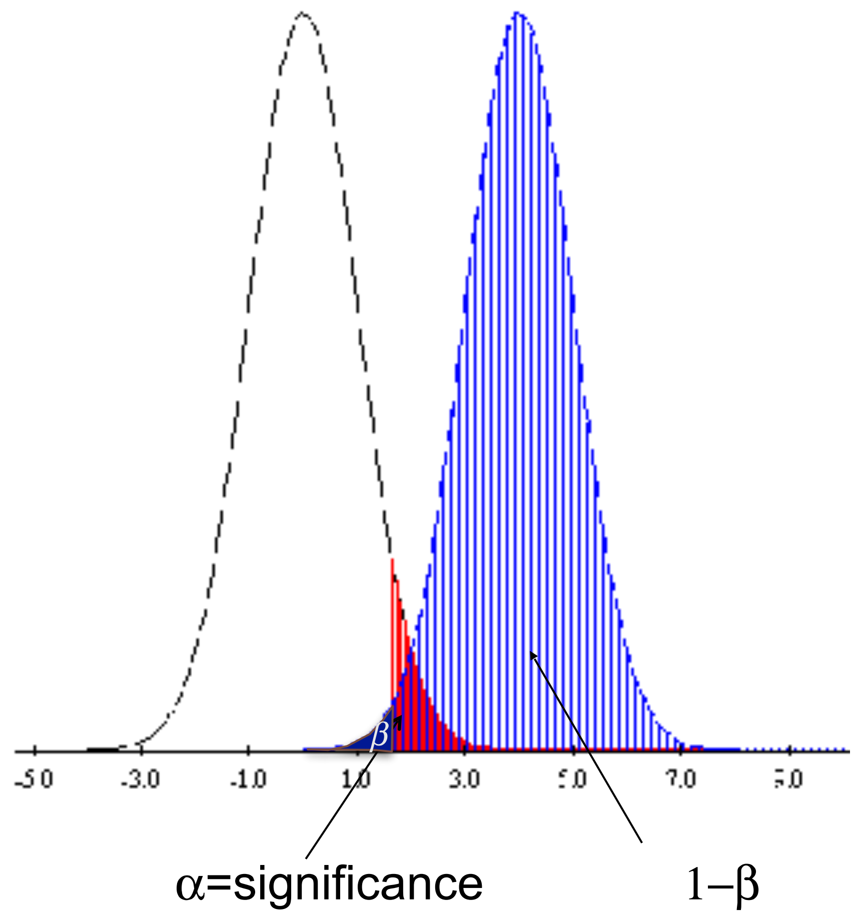
- By defining α you determine your tolerance towards mistakes… (accepted mistakes frequency)
- type-I error: the probability to reject the tested (null) hypothesis (H_o) when it is true
- $\alpha = \Pr ob(reject\ H_0 \mid H_0)$

  $\alpha = typeI\ error$
- Type II: The probability to accept null hypothesis when it is wrong

$\beta = \Pr ob(accept\ H_0 \mid \bar{H}_0)$

$\beta = typeII\ \ error$

- The pdf of q….



Basic Definitions: POWER

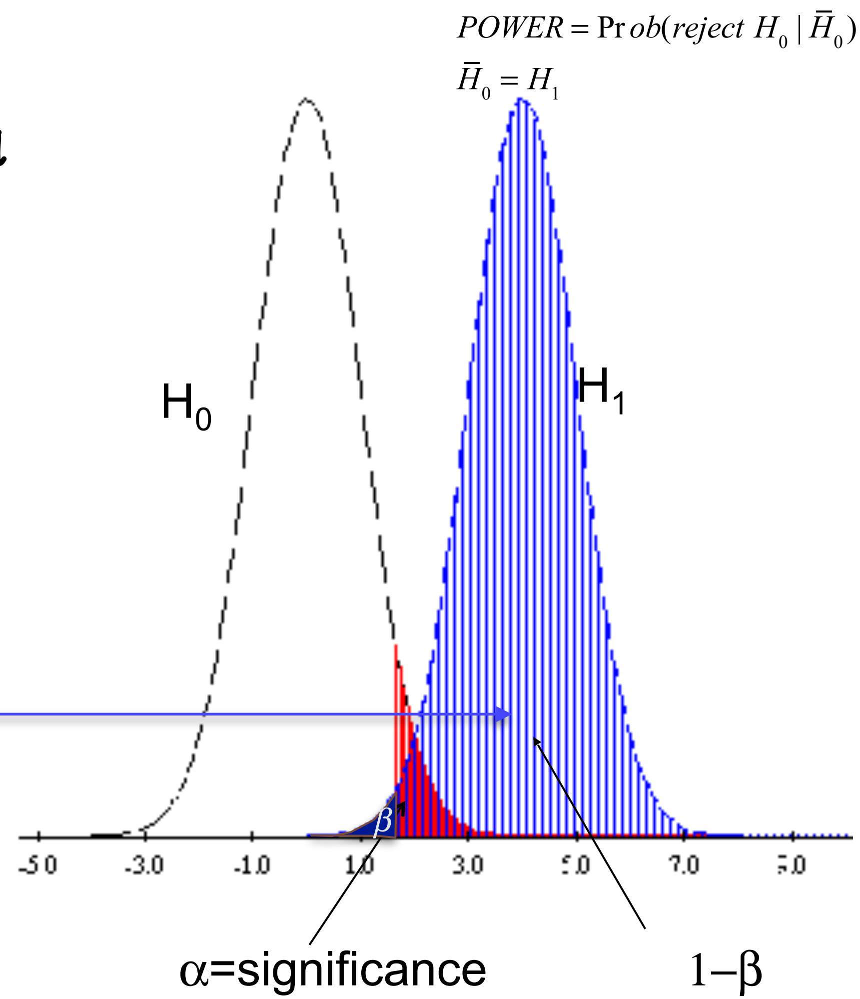
- $\alpha = \Pr ob(reject\ H_0 \mid H_0)$
- The POWER of an hypothesis test is the probability to reject the null hypothesis when it is indeed wrong (the alternate analysis is true)
- $POWER = \Pr ob(reject\ H_0 \mid \bar{H}_0)$

  $\beta = Prob(accept\ H_0 \mid \bar{H}_0)$

  $1-\beta = Prob(reject\ H_0 \mid \bar{H}_0)$

  $\bar{H}_0 = H_1$

  $1-\beta = Prob(reject\ H_0 \mid H_1)$
- The power of a test increases as the rate of type II error decreases



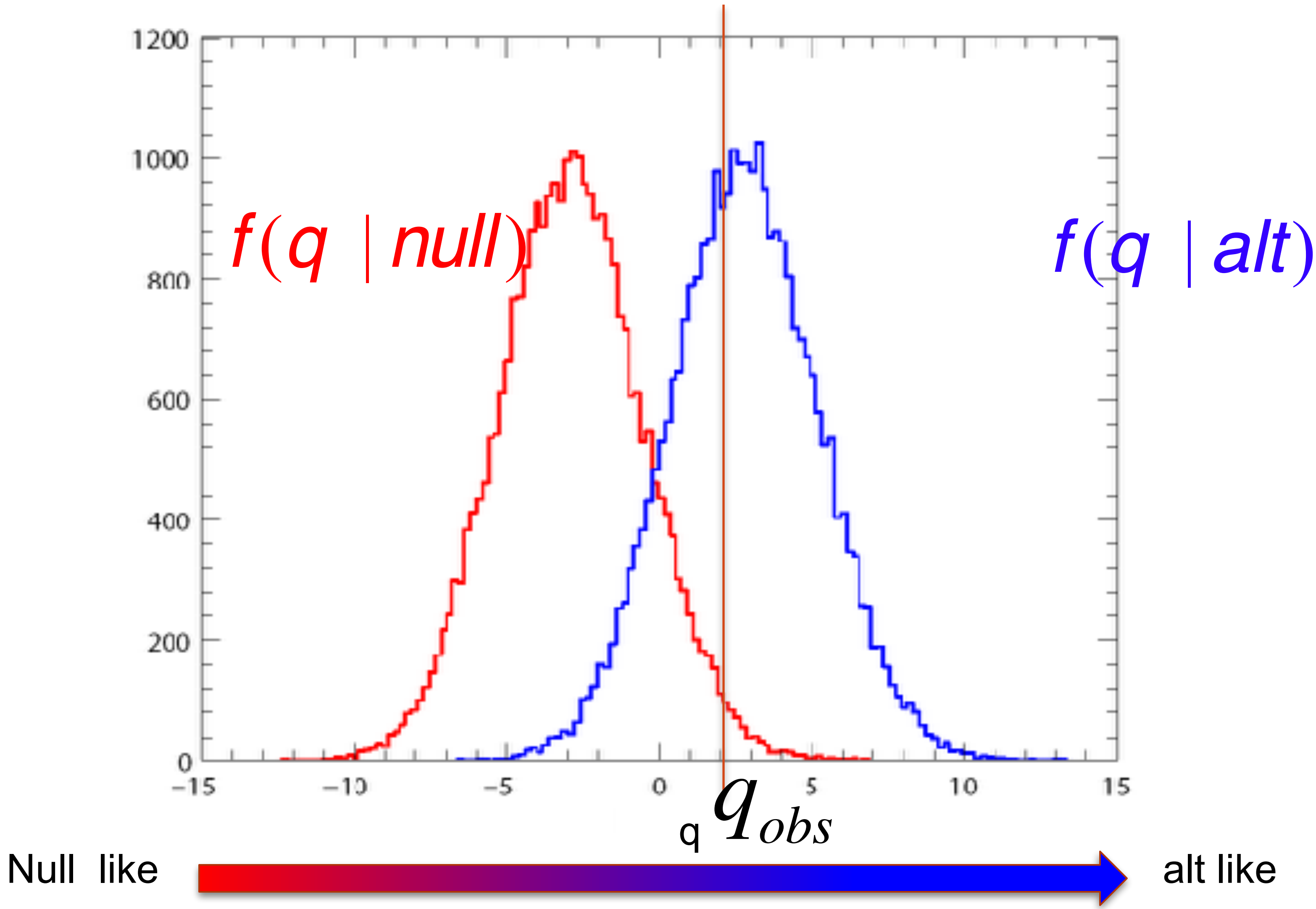
p-Value

- The observed p-value is a measure of the compatibility of the data with the tested hypothesis.

- It is the probability, under assumption of the null hypothesis $H_{null,}$ of finding data of equal or greater incompatibility with the predictions of H_{null}

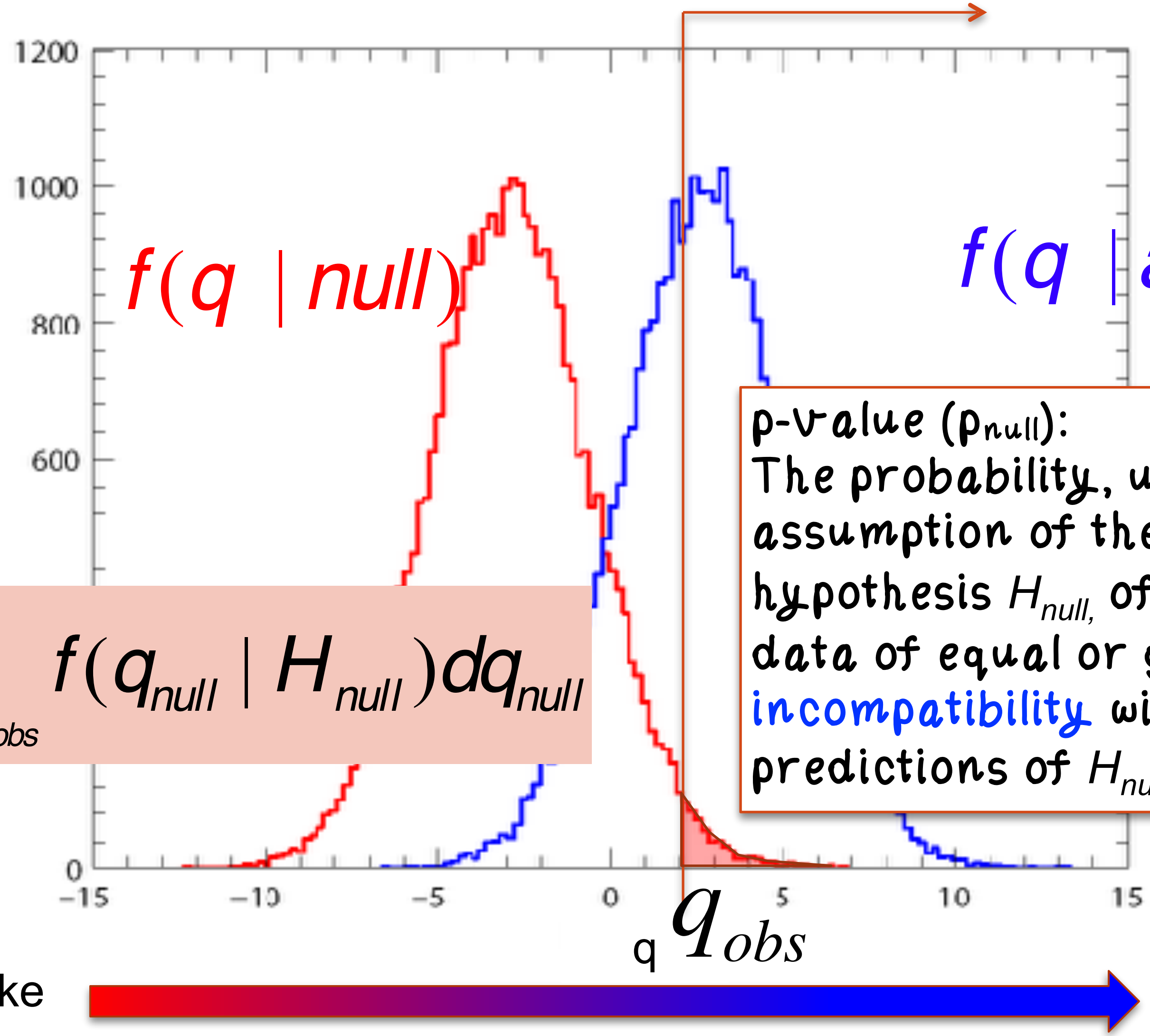
- An important property of a test statistic is that its sampling distribution under the null hypothesis be calculable, either exactly or approximately, which allows p-values to be calculated. (Wiki)

PDF of a test statistic



PDF of a test statistic

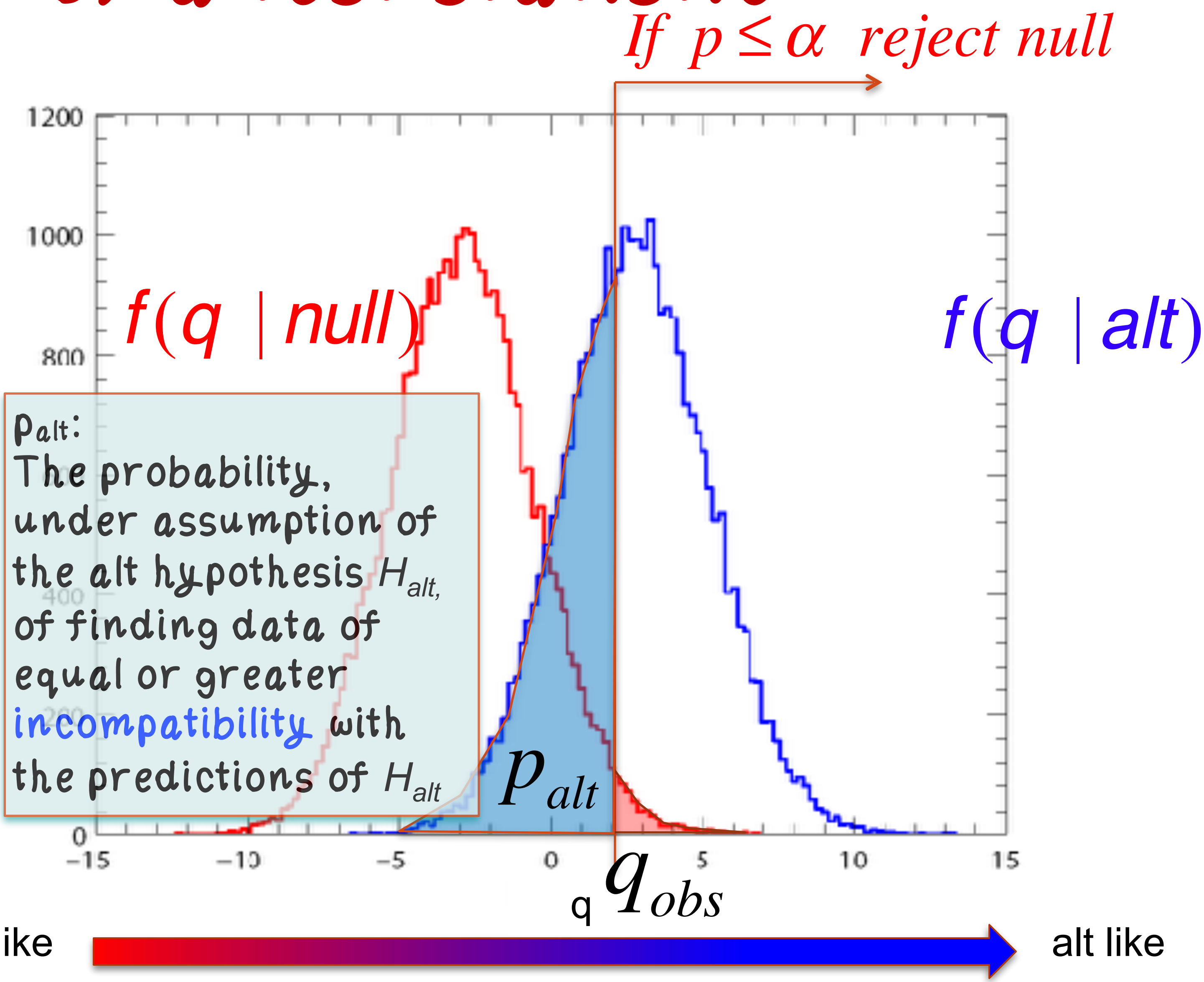
If $p \le \alpha$ *reject null*



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

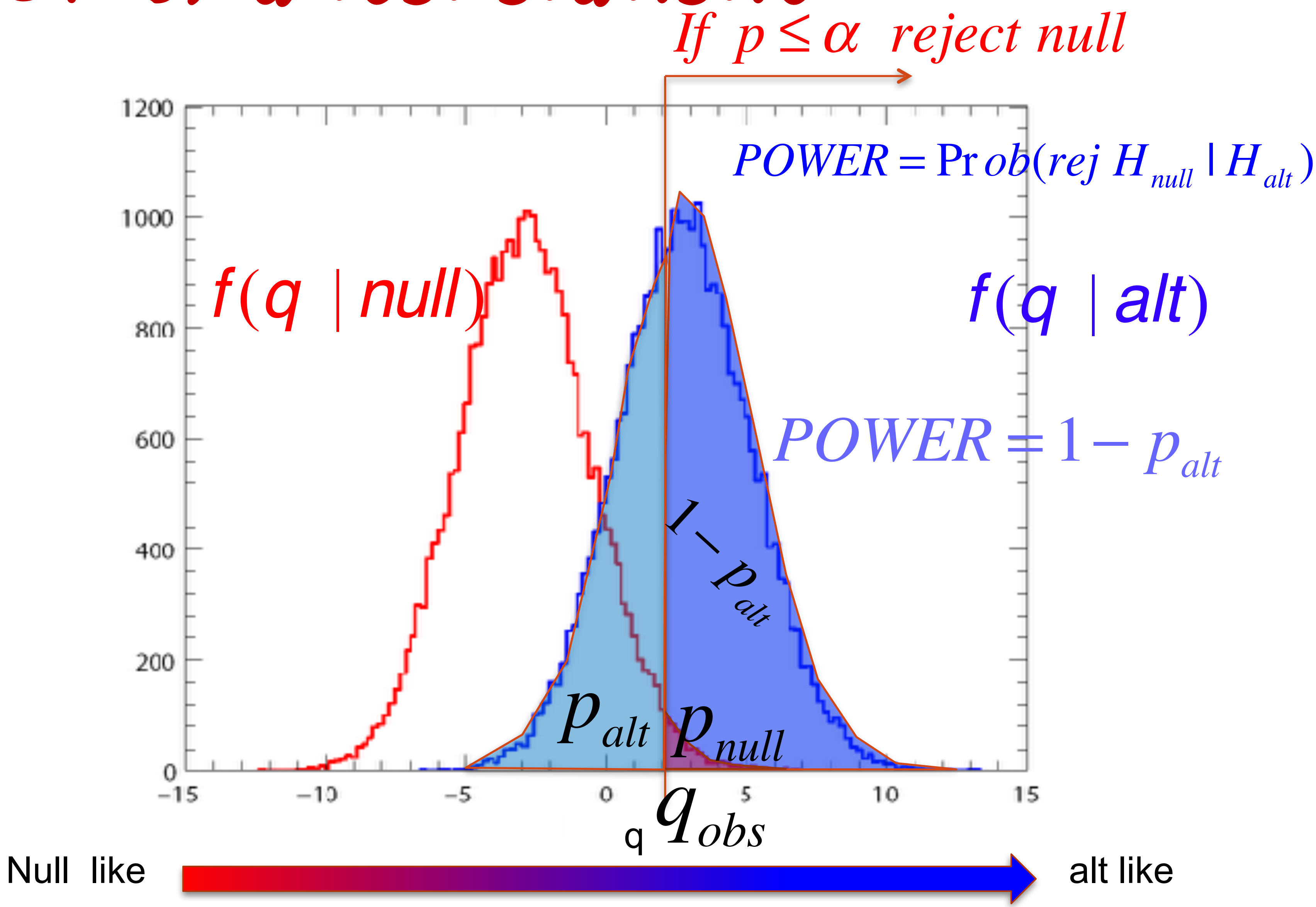
p-value (p_{null}):
The probability, under assumption of the null hypothesis H_{null}, of finding data of equal or greater incompatibility with the predictions of H_{null}

PDF of a test statistic



p_{alt}:
The probability, under assumption of the alt hypothesis H_{alt}, of finding data of equal or greater incompatibility with the predictions of H_{alt}

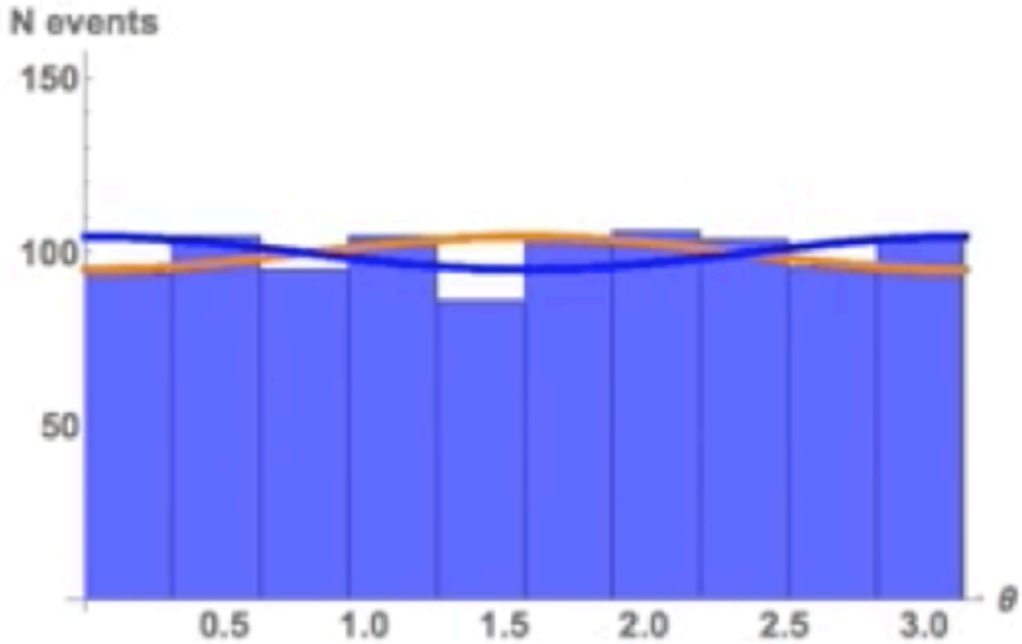
PDF of a test statistic

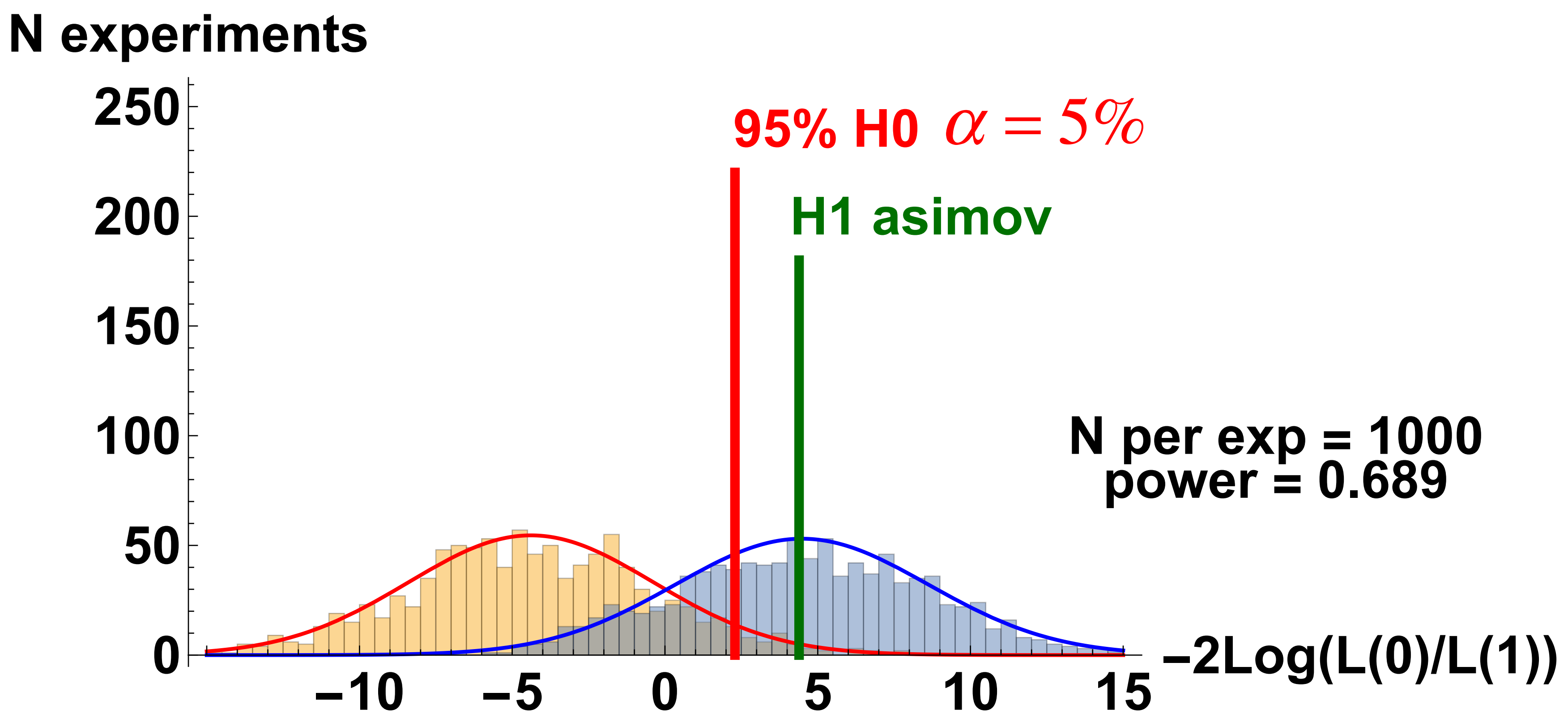


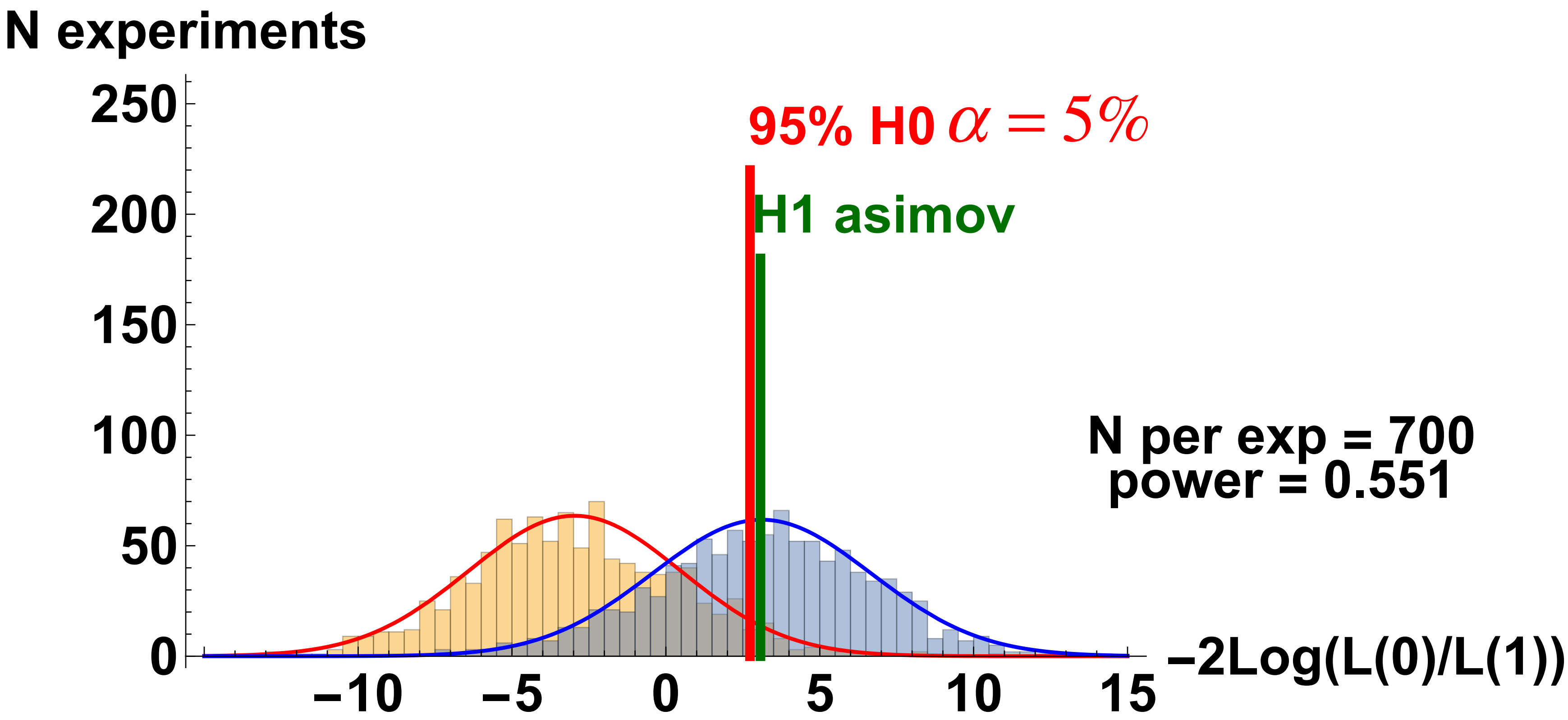
Power and Luminosity

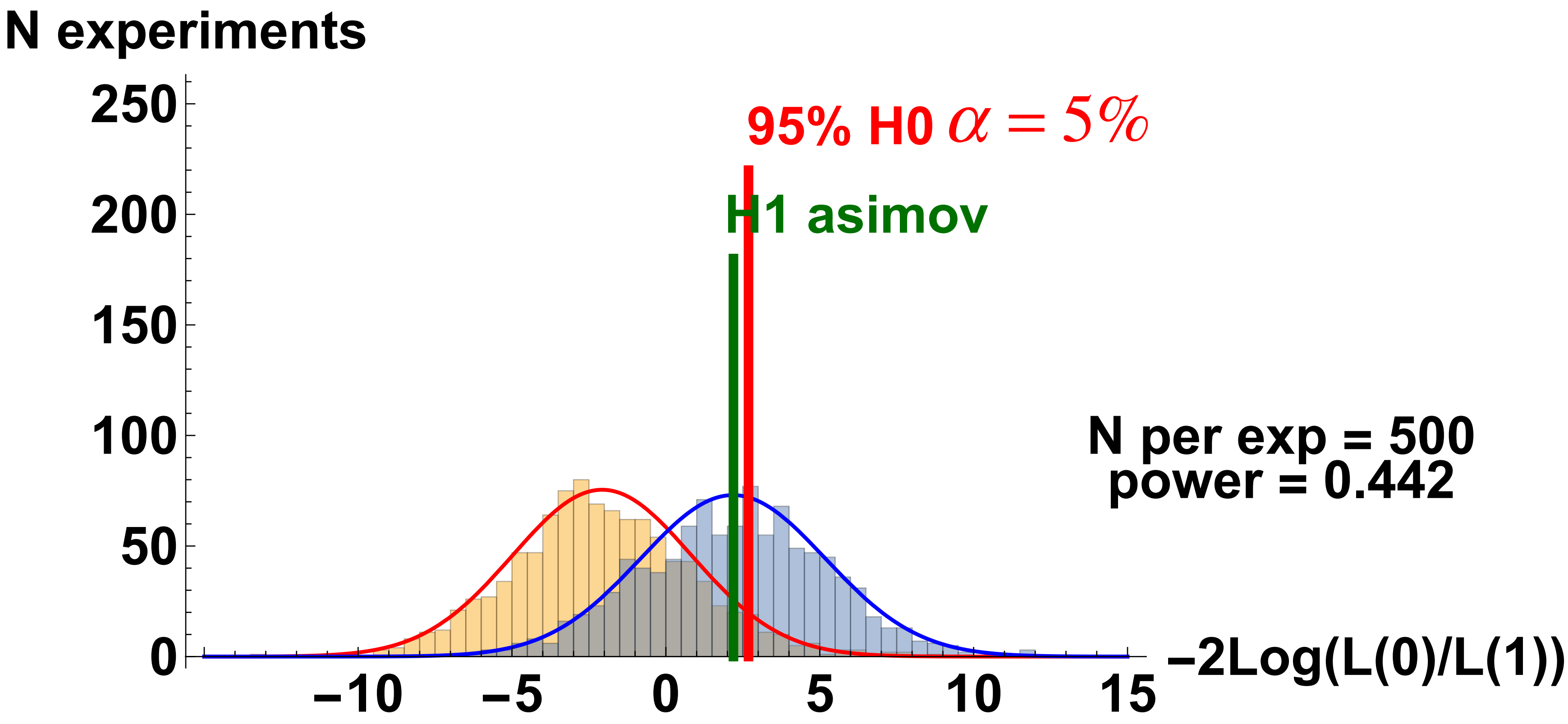
For a given significance the power increases with increased luminosity

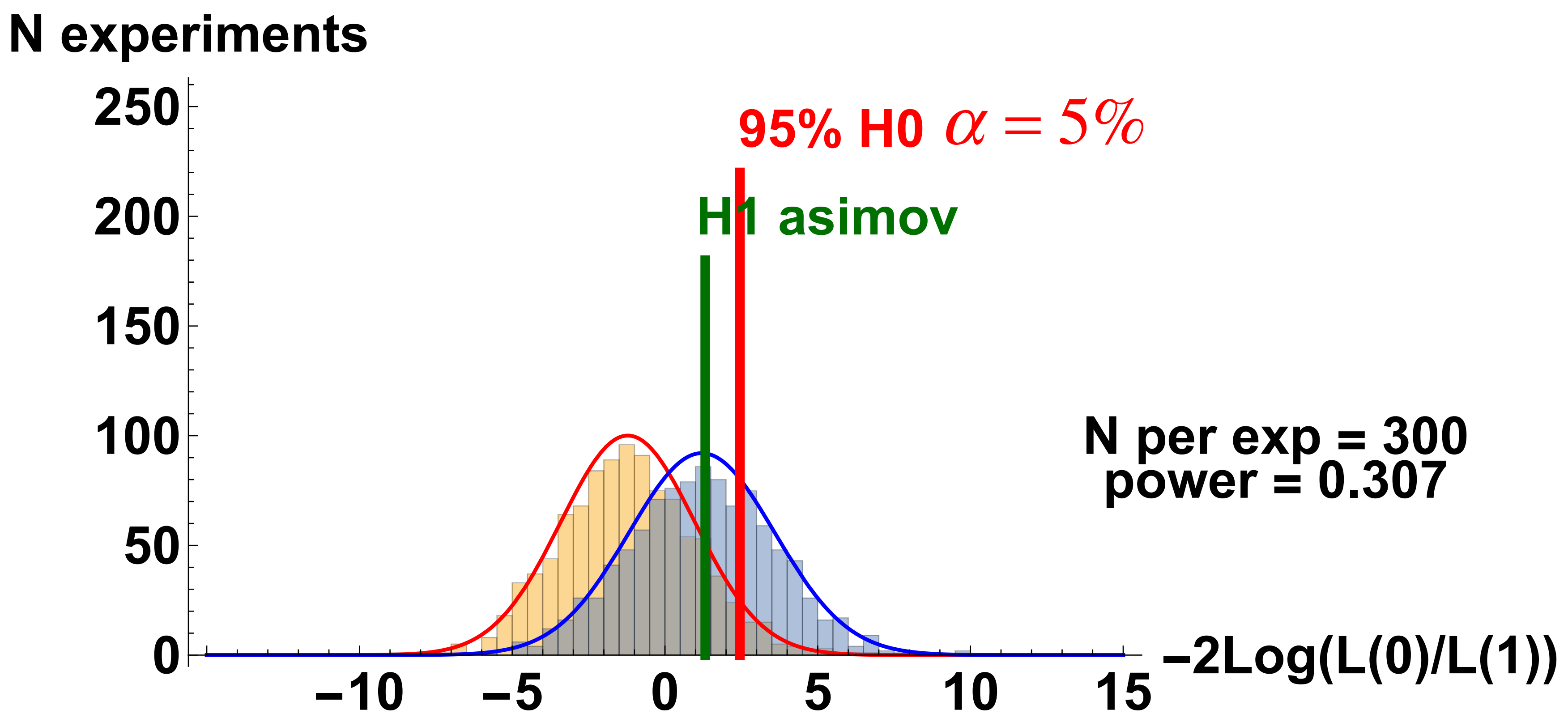
Luminosity ~ Total number of events in an experiment



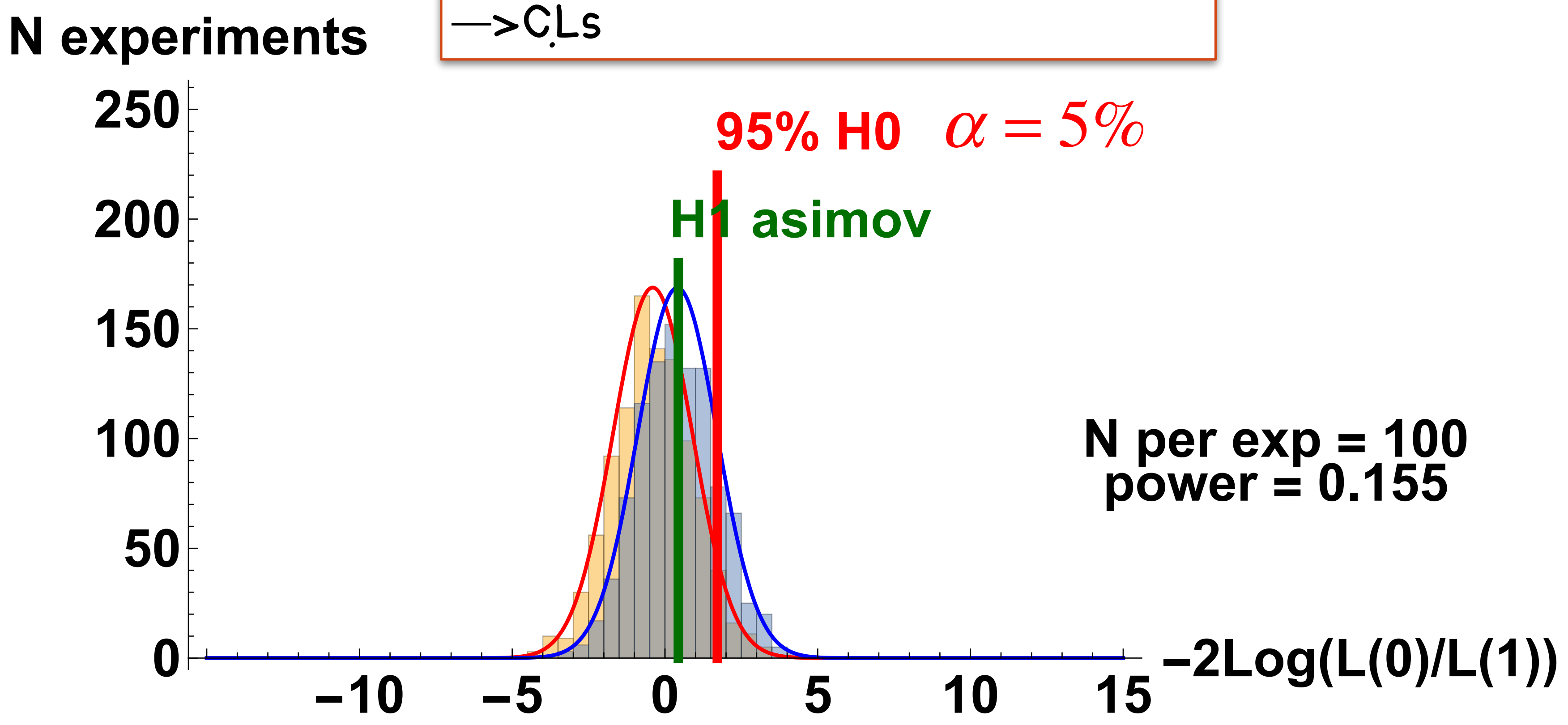

N experiments
250
200
150
100
50
0
95% H0 $\alpha = 5\%$
H1 asimov
N per exp = 1000
power = 0.689
−10
−5
0
5
10
15
−2Log(L(0)/L(1))


N experiments
250
200
150
100
50
0
95% H0 $\alpha = 5\%$
H1 asimov
N per exp = 700
power = 0.551
−10
−5
0
5
10
15
−2Log(L(0)/L(1))


N experiments
250
200
150
100
50
0
95% H0 $\alpha = 5\%$
H1 asimov
N per exp = 500
power = 0.442
−10
−5
0
5
10
15
−2Log(L(0)/L(1))


N experiments
250
200
150
100
50
0
95% H0 $\alpha = 5\%$
H1 asimov
N per exp = 300
power = 0.307
−10
−5
0
5
10
15
−2Log(L(0)/L(1))

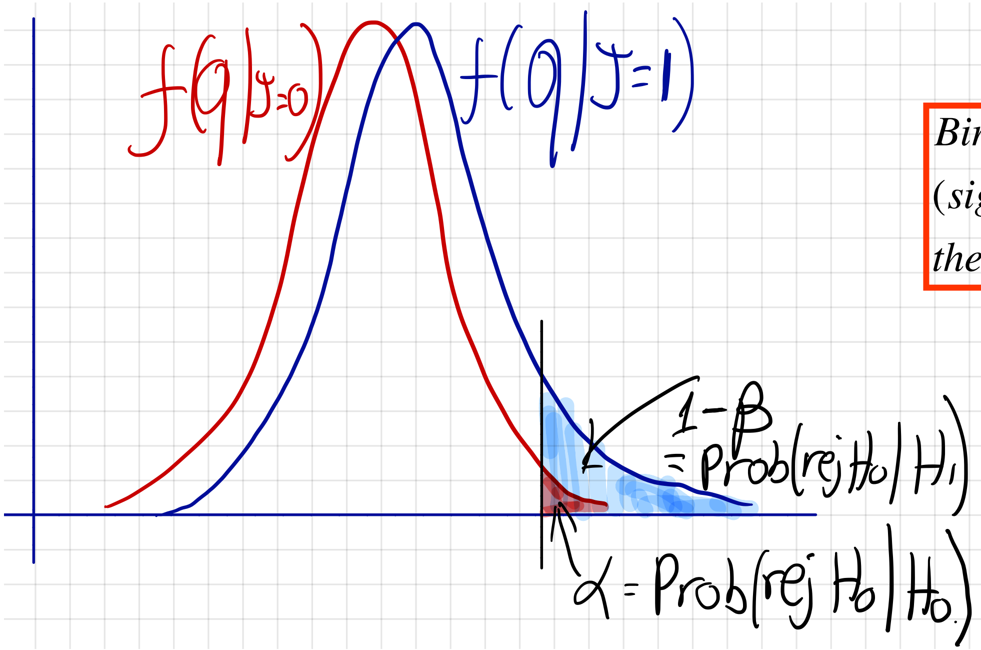
Hard to tell f(q|J=0) from f(q|J=1)
—>CLs



CLs

Birnbaum (1977)

"A concept of statistical evidence is not plausible unless it finds 'strong evidence for H_1 as against H_0' with small probability (α) when H_0 is true, and with much larger probability ($1-\beta$) when H_1 is true."



Birnbaum (1962) *suggested that* $\alpha / 1-\beta$ *(significance / power) should be used as a measure of the strength of a statistical test, rather than* α *alone*

$$p = 5\% \rightarrow p' = 5\% / 0.155 = 32\%$$

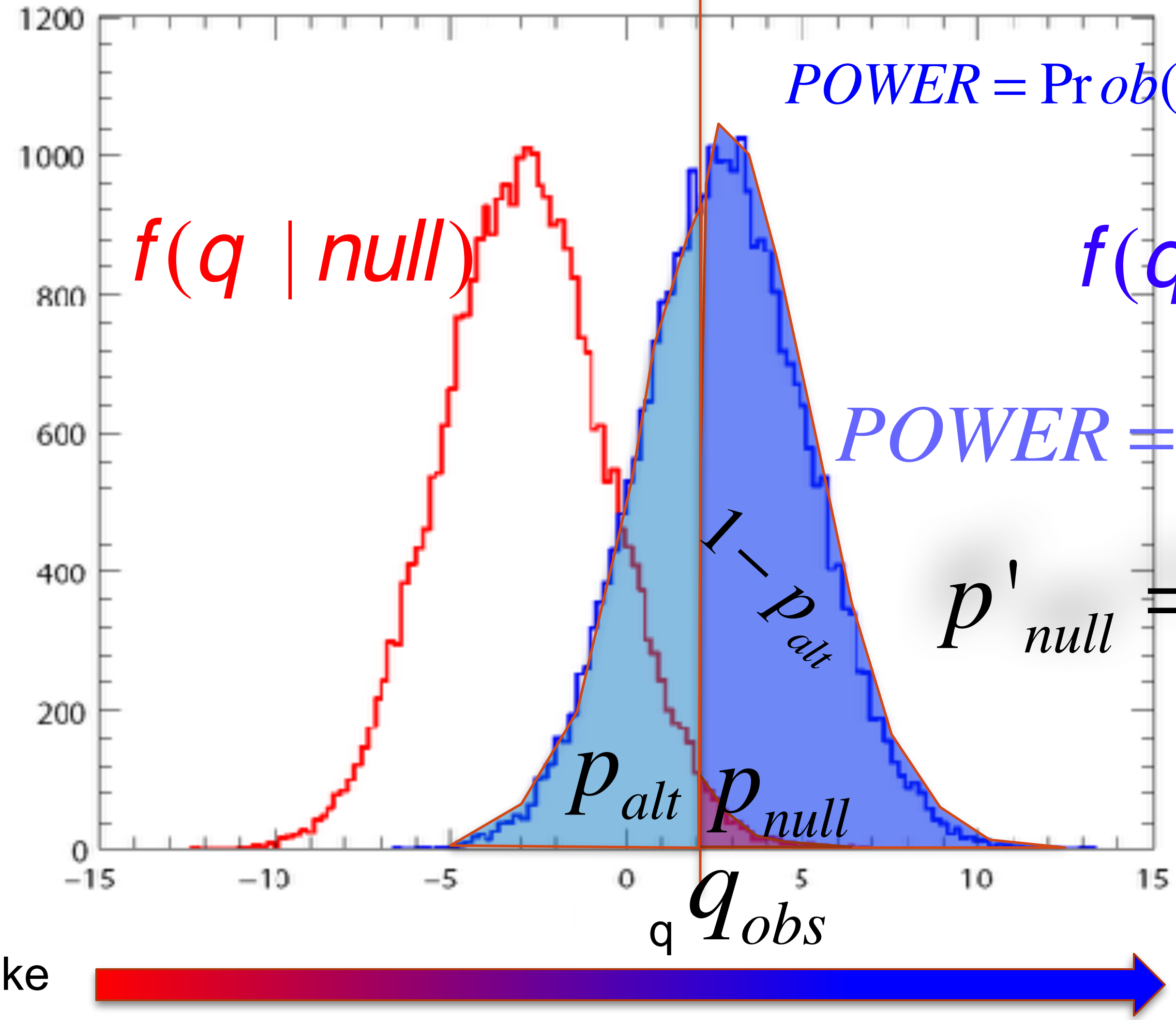
$$p' \equiv CL_s$$

$$p'_\mu = \frac{p_\mu}{1-p_0}$$

CLs

$If\ \ p \le \alpha\ \ reject\ null$

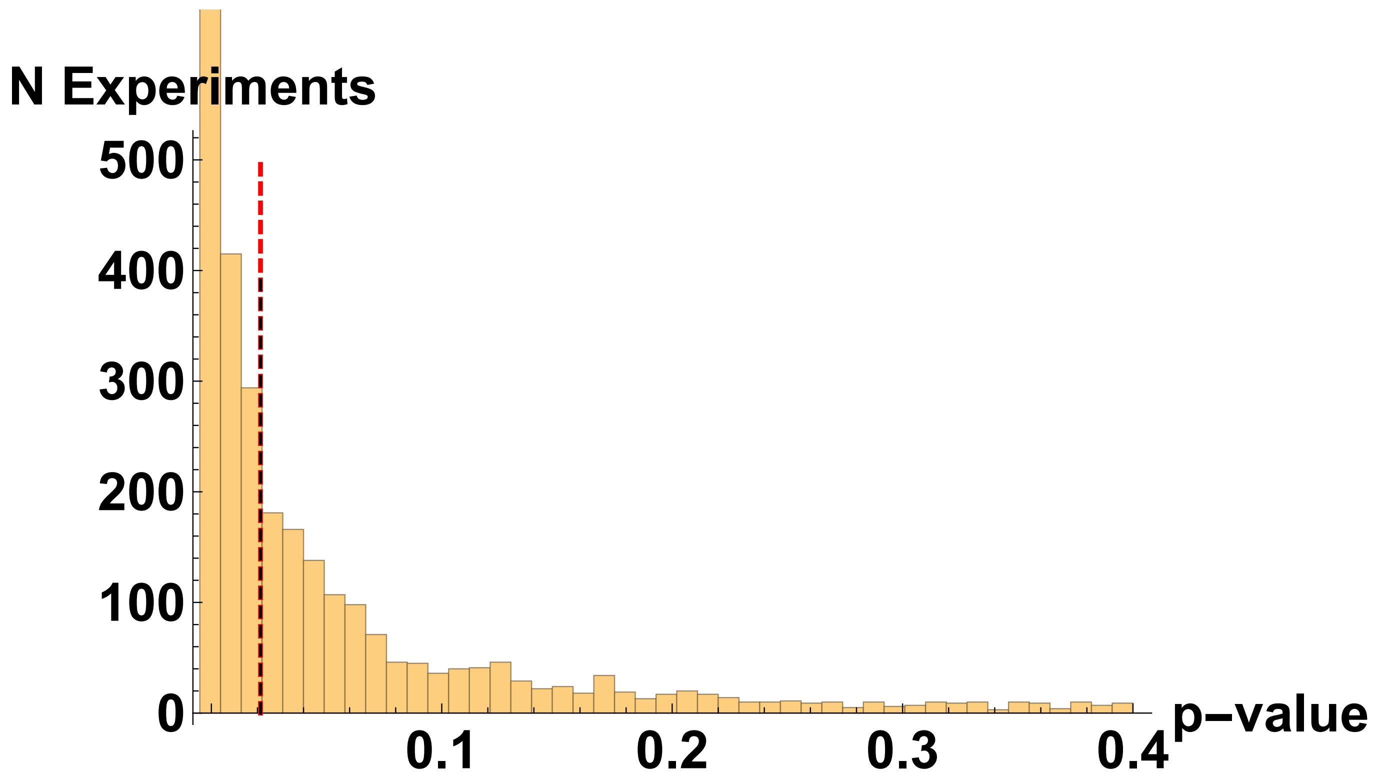
$POWER = \Pr ob(rej\ H_{null} \mid H_{alt})$



$POWER = 1 - p_{alt}$

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

Distribution of p-value under H1



Distribution of p-value under H0

$f(x)\ PDF$

$$cumulative\ F(x) = \int_{-\infty}^{x} f(x')dx'$$

$let\ \ y = F(x)$

$PDF\ of\ \ y$

$$\frac{dP}{dy} = \frac{dP}{dx}\frac{dx}{dy} = f(x)/(dF/dx) = f(x)/f(x) = 1$$

$F(x)$ *distributes uniform between* 0 *and* 1

$p = 1 - F(x)$ *distributes uniform between* 0 *and* 1

Distribution of p-value under H0

$f(x)$ *PDF*

cumulative $F(x) = \int_{-\infty}^{x} f(x')dx'$

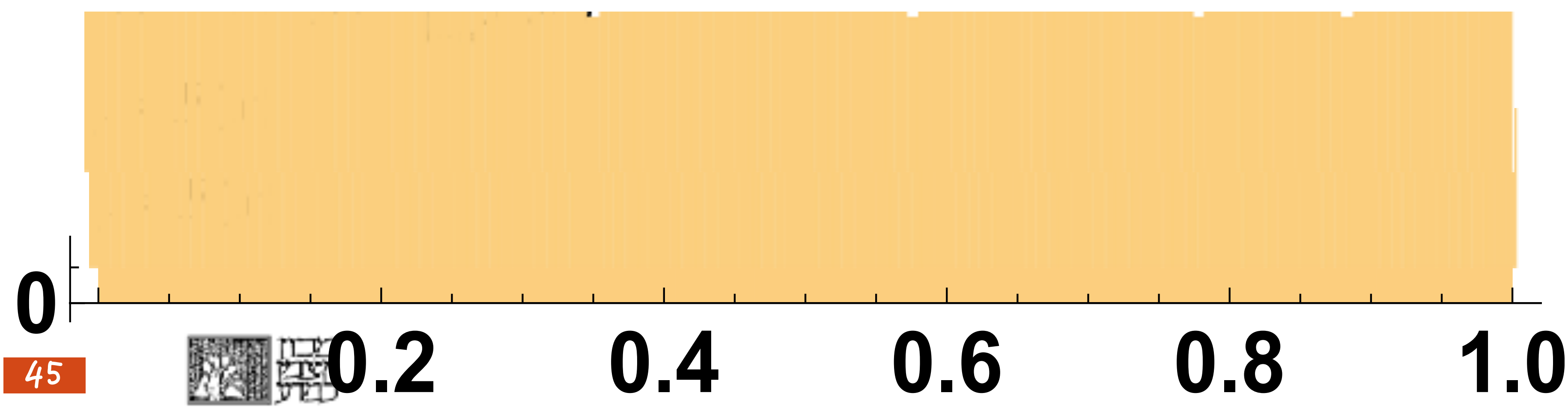
let $y = F(x)$

PDF of y

$$\frac{dP}{dy} = \frac{dP}{dx}\frac{dx}{dy} = f(x)/(dF/dx) = f(x)/f(x) = 1$$

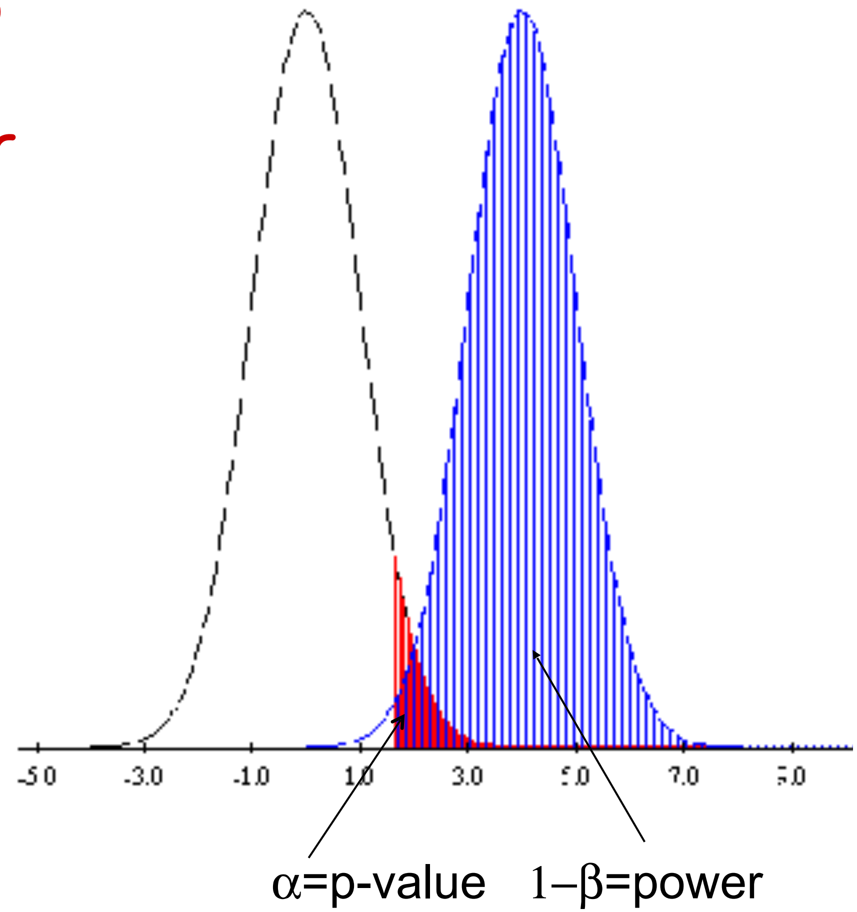
$F(x)$ *distributes uniform between* 0 *and* 1

$p = 1 - F(x)$ *distributes uniform between* 0 *and* 1



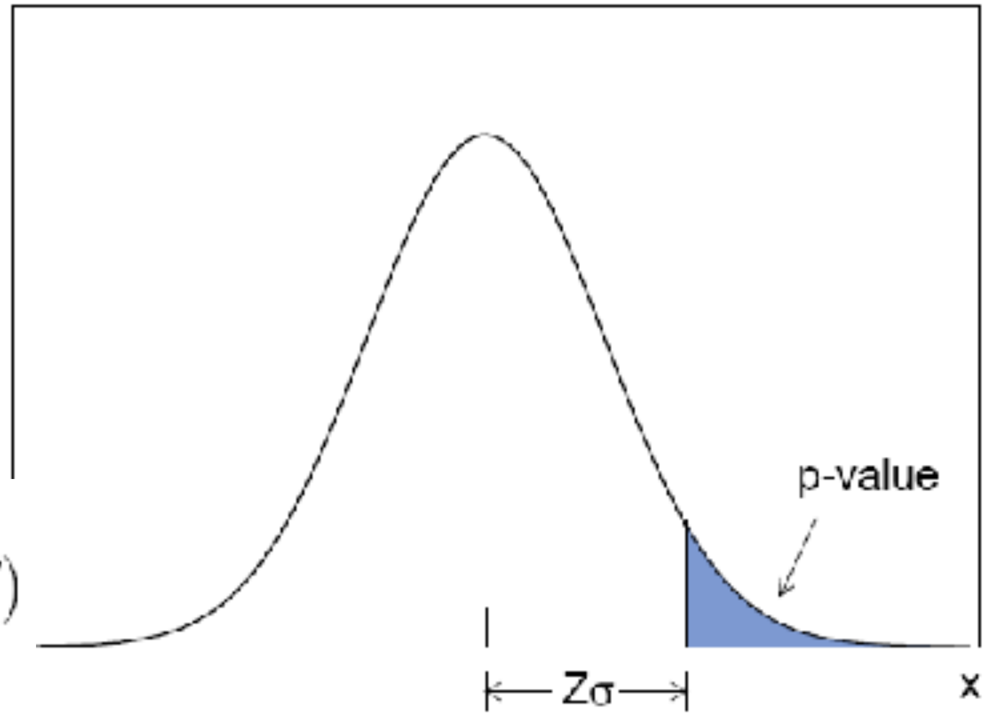
Which Statistical Method is Better

- To find out which of two methods is better plot the p-value vs the power for each analysis method
- Given the p-value, the one with the higher power is better
- p-value~significance



From p-values to Gaussian significance

It is a custom to express the p-value as the significance associated to it, had the pdf were Gaussians



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

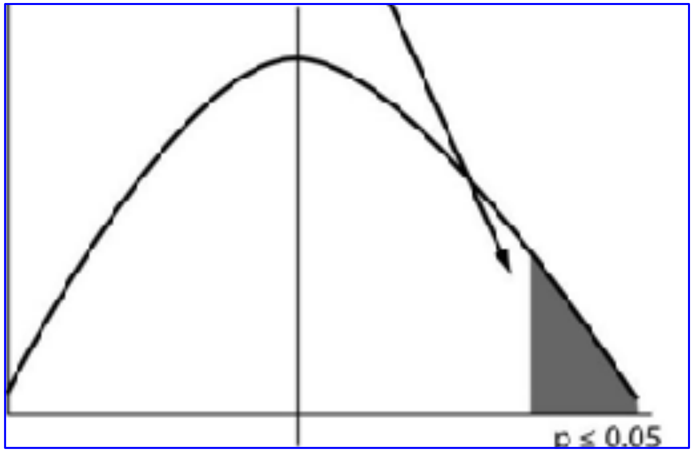
$$Z = \Phi^{-1}(1 - p)$$

A significance of $Z = 5$ corresponds to $p = 2.87 \times 10^{-7}$

Beware of 1 vs 2-sided definitions!

1-Sided p-value

- When trying to reject an hypothesis while performing searches, one usually considers only one-sided tail probabilities.

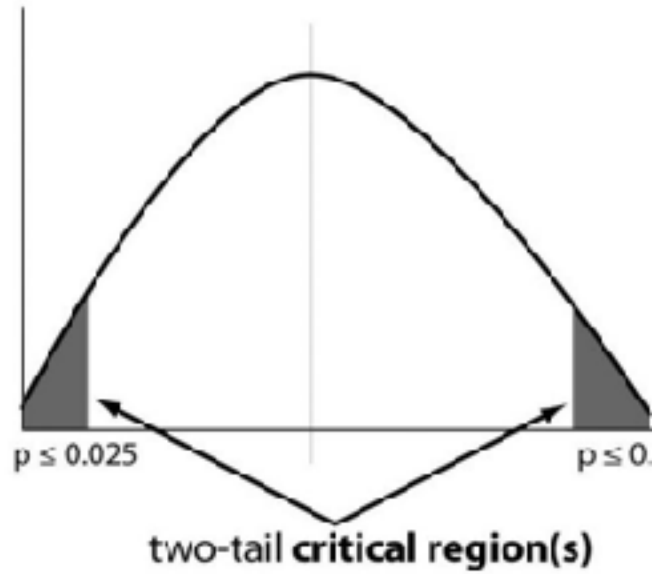


- Downward fluctuations of the background will not serve as an evidence against the background
- Upward fluctuations of the signal will not be considered as an evidence against the signal

2-Sided p-value

- When performing a measurement (t_μ), any deviation above or below the expected null is drawing our attention and might serve an indication of some anomaly or new physics.

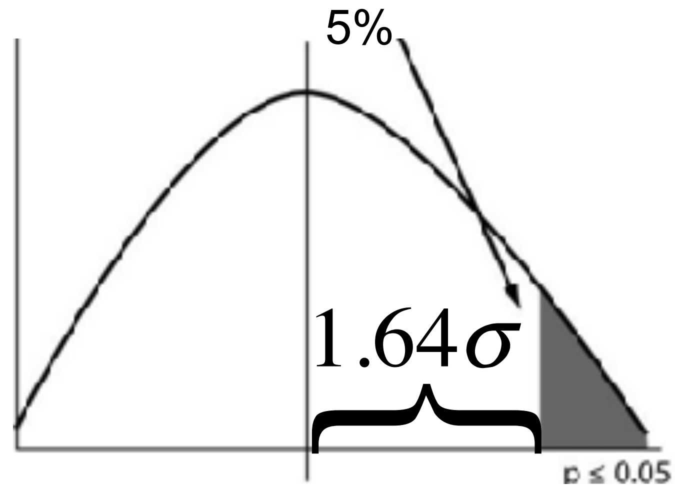
- Here we use a 2-sided p-value

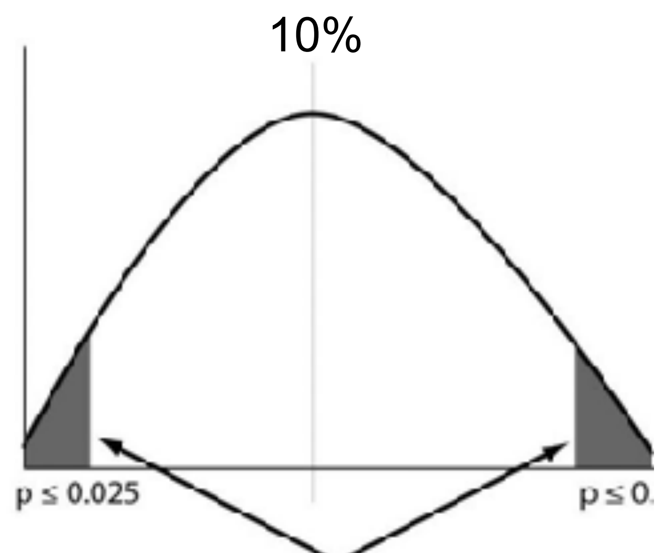


1-sided 2-sided

To determine a 1 sided 95% CL,
we sometimes need to set the critical region to 10% 2 sided

2-sided 5% is 1.95 σ
2-sided 10% is 1.64 σ





p-value – testing the null hypothesis

When testing the b hypotheis (null=b), it is custom to set

$\alpha = 2.9\ 10^{-7}$

→ if $p_b < 2.9\ 10^{-7}$ the b hypothesis is rejected

→ Discovery

When testing the s+b hypothesis (null=s+b), set α =5%

if $p_{s+b} < 5\%$ the signal hypothesis is rejected at the 95% Confidence Level (CL)

→ Exclusion

Confidence Interval and Confidence Level (CL)

CL & CI

$$measurement\ \hat{\mu} = 1.1 \pm 0.3$$

$$L(\mu) = G(\mu; \hat{\mu}, \sigma_{\hat{\mu}})$$

$$\Rightarrow CI\ of\ \mu = [0.8, 1.4]\ at\ 68\%\ CL$$

- 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
- 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 $p(\mu_{meas}|\mu)$
- based on your statistical method you deduce 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.

Confidence Interval & Coverage

- You claim, $CI_\mu=[\mu_1,\mu_2]$ at the 95% CL
  i.e. In an ensemble of experiments CL (95%) of the obtained confidence intervals will contain the true value of μ.

  - If your statement is accurate, you have full coverage
  - If the true CL is>95%, your interval has an over coverage
  - If the true CL is <95%, your interval has an undercoverage

Upper Limit

- Given the measurement you deduce somehow (based on your statistical method) that there is a 95% Confidence interval $[0,\mu_{up}]$.

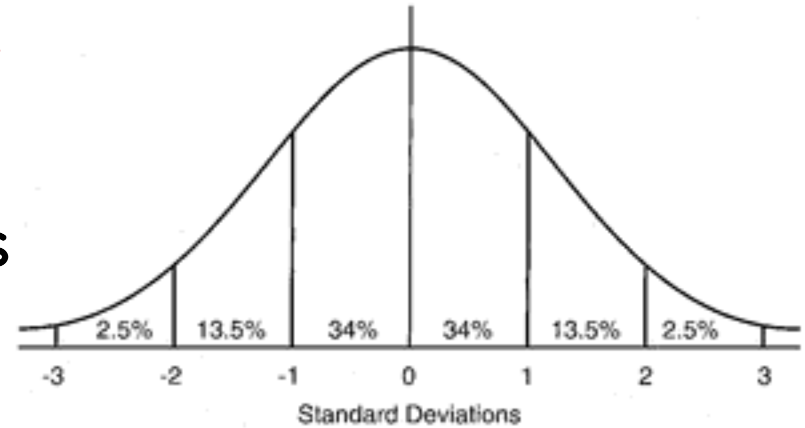
- This means: our interval contains $\mu=0$ (no Higgs)

- We therefore deduce that $\mu<\mu_{up}$ at the 95% Confidence Level (CL)

- μ_{up} is therefore an upper limit on μ

- If $\mu_{up}<1$ →
  $\sigma(m_H)<\sigma_{SM}(m_H)$→
  a SM Higgs with a mass m_H is excluded at the 95% CL

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

$$\hat{X} \pm \sigma \qquad i.e.\ x_{true} \in \left[\hat{x} - \sigma_{\hat{x}}, \hat{x} + \sigma_{\hat{x}}\right] @\, 68\%\ CL$$

Side Note:
A CI is an interval in the true parameters phase-space

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

- One can guarantee a coverage with the Neyman Construction (1937)

Neyman, J. (1937) "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability" Philosophical Transactions of the Royal Society of London A, 236, 333-380.

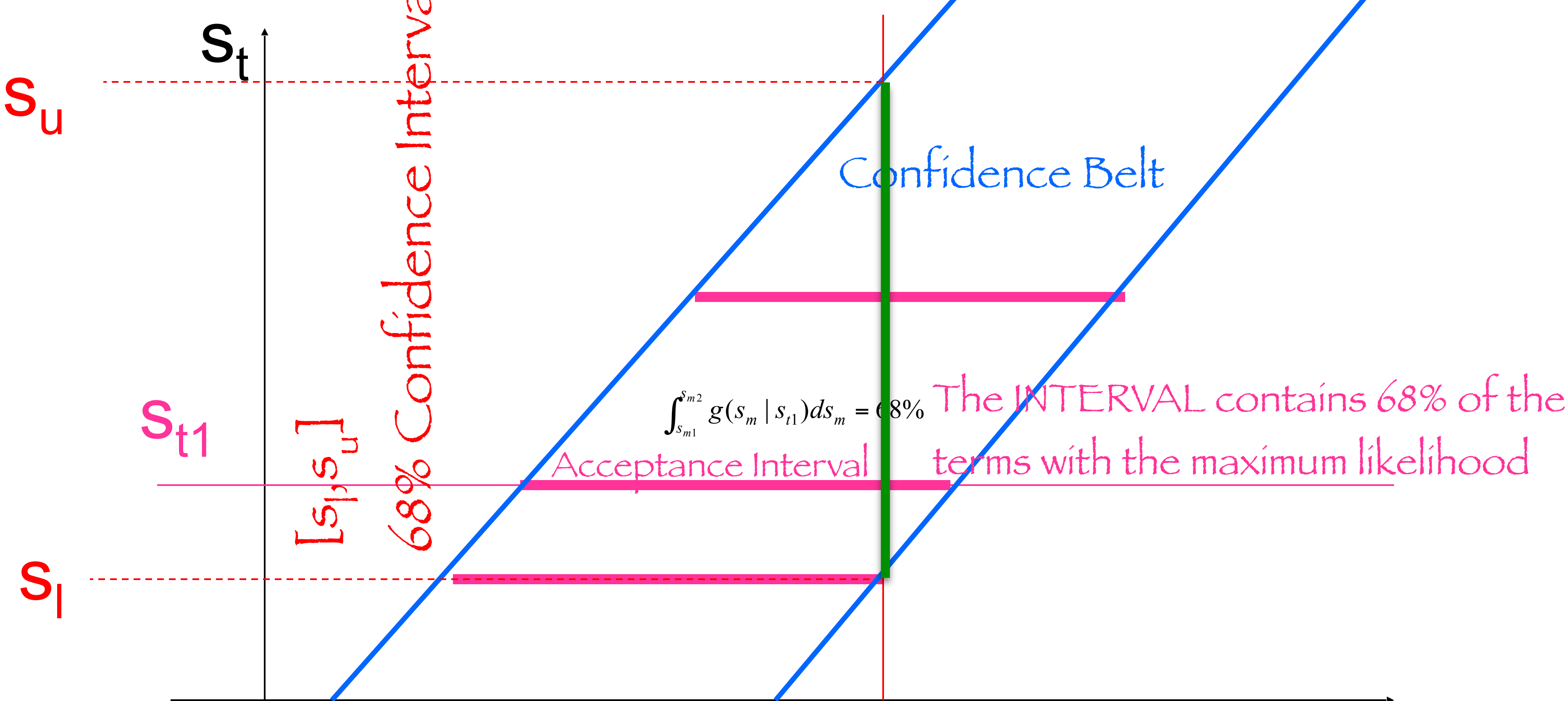
The Frequentist Game a 'la Neyman

Or

How to ensure a Coverage with Neyman construction

Neyman Construction

$Prob(s_m \mid s_t)$ *is known*



$[s_l, s_u]$ 68% Confidence Interval

In **68%** of the experiments the derived **C.I. contains** the **unknown true value of s**

- With Neyman Construction we guarantee a coverage via construction, i.e. for any value of the unknown true s, the Construction Confidence Interval will cover s with the correct rate.

Neyman Construction

$\theta \equiv s_{true} \qquad x \equiv s_{measured} \qquad pdf\ \ f(x|\theta)\ is\ known$

for each prospective θ *generate* x

construct an interval in DATA phase – space

$$Interval = \int_{x_l}^{x_h} f(x|\theta)dx = 68\%$$

repeat for each θ

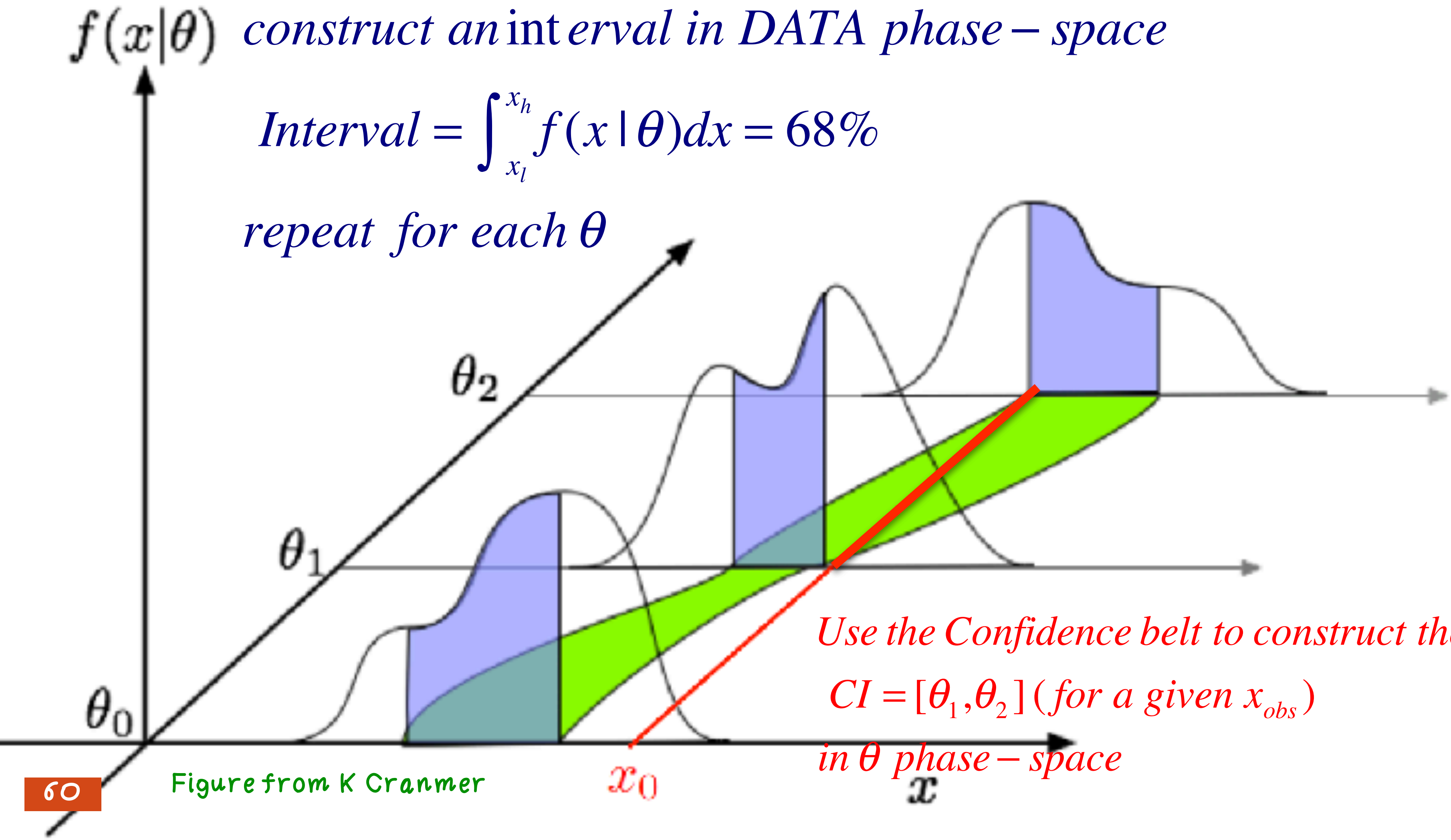


Figure from K Cranmer

Nuisance Parameters

or Systematics

Nuisance Parameters (Systematics)

- There are two kinds of parameters:
  - Parameters of interest (signal strength… cross section… μ)
  - Nuisance parameters (background (b), signal efficiency, resolution, energy scale,…)

- The nuisance parameters carry systematic uncertainties

- There are two related issues:
  - Classifying and estimating the systematic uncertainties
  - Implementing them in the analysis

- The physicist must make the difference between cross checks and identifying the sources of the systematic uncertainty.
  - Shifting cuts around and measure the effect on the observable… Very often the observed variation is dominated by the statistical uncertainty in the measurement.

Implementation of Nuisance Parameters

- Implement by marginalizing (Bayesian) or profiling (Frequentist)
- Hybrid: One can also use a frequentist test statistics (PL) while treating the NPs via marginalization (Hybrid, Cousins & Highland way)
- Marginalization (Integrating))
  - Integrate the Likelihood, L, over possible values of nuisance parameters (weighted by their prior belief functions -- Gaussian,gamma, others...)
  - $$L(\mu)=\int L(\mu,\theta)\pi(\theta)d\theta$$

The Hybrid Cousins-Highland Marginalization

Cousins & Highland

$$q = \frac{L(s+b(\theta))}{L(b(\theta))} \Rightarrow \frac{\int L\big(s+b(\theta)\big)\pi(\theta)d\theta}{\int L\big(b(\theta)\big)\pi(\theta)d\theta}$$

Profiling the NPs

$$q = \frac{L(s+b(\theta))}{L(b(\theta))} \Rightarrow \frac{L(s+b(\hat{\hat{\theta}}_s))}{L(b(\hat{\hat{\theta}}_b))}$$

$\hat{\hat{\theta}}_s$ *is the MLE of* θ *fixing s*

Nuisance Parameters and Subsidiary Measurements

- Usually the nuisance parameters are auxiliary parameters and their values are constrained by auxiliary measurements
- Example

$$n \sim \mu s(m_H) + b \qquad \langle n \rangle = \mu s + b$$

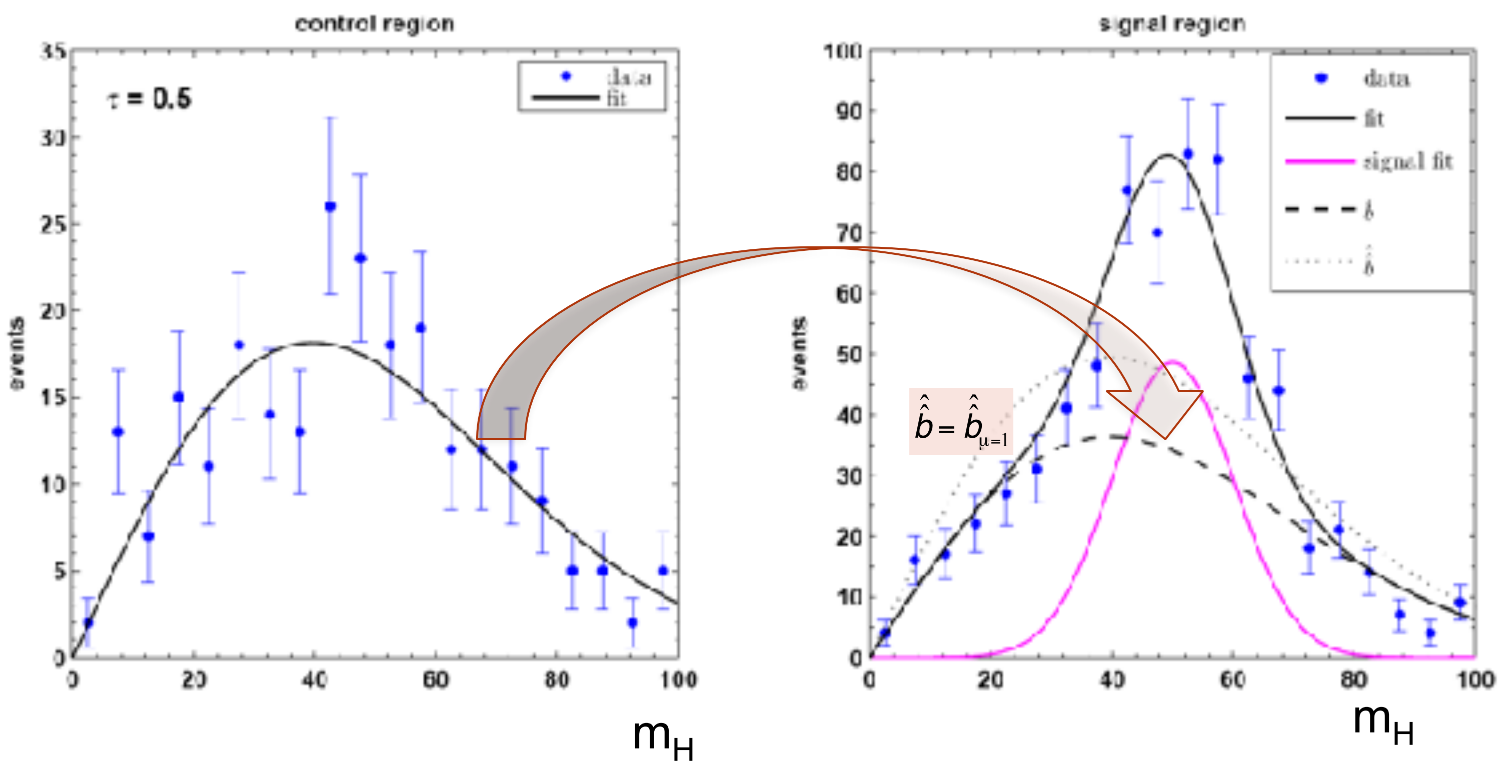
$$m = \tau b$$

$$L\left(\mu \cdot s + b(\theta)\right) = Poisson\left(n; \mu \cdot s + b(\theta)\right) \cdot Poisson\left(m; \tau b(\theta)\right)$$

Mass shape as a discriminator

$$n \sim \mu s(m_H) + b \qquad m \sim \tau b$$

$$L\left(\mu \cdot s + b(\theta)\right) = \prod_{i=1}^{nbins} Poisson\left(n_i; \mu \cdot s_i + b_i(\theta)\right) \cdot Poisson\left(m_i; \tau b_i(\theta)\right)$$



Wilks Theorem

S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60-2.

Profile Likelihood with Nuisance Parameters

$$q_\mu = -2\ln\frac{L(\mu s + \hat{\hat{b}}_\mu)}{L(\hat{\mu}s + \hat{b})}$$

$$q_\mu = -2\ln\frac{\max_b L(\mu s + b)}{\max_{\mu,b} L(\mu s + b)}$$

$$q_\mu = q_\mu(\hat{\mu}) = -2\ln\frac{L(\mu s + \hat{\hat{b}}_\mu)}{L(\hat{\mu}s + \hat{b})}$$

$\hat{\mu}$ MLE of μ

$\hat{b}$ MLE of b

$\hat{\hat{b}}_\mu$ MLE of b fixing μ

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

A toy case with 1-3 poi

$$n = \mu\epsilon A s + b$$

$$L = L(\mu,\epsilon,A,b)$$

3 cases studied

$1poi: \mu$ while ϵ, A, b *profiled*

$2poi: \mu,\epsilon$ *profile* A and b

3poi: μ,ϵ,A profile b

$$1poi:\ t_{\mu} = \frac{L(\mu,\hat{\hat{\epsilon}},\hat{\hat{A}},\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

$$2poi:\ t_{\mu,\epsilon} = \frac{L(\mu,\epsilon,\hat{\hat{A}},\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

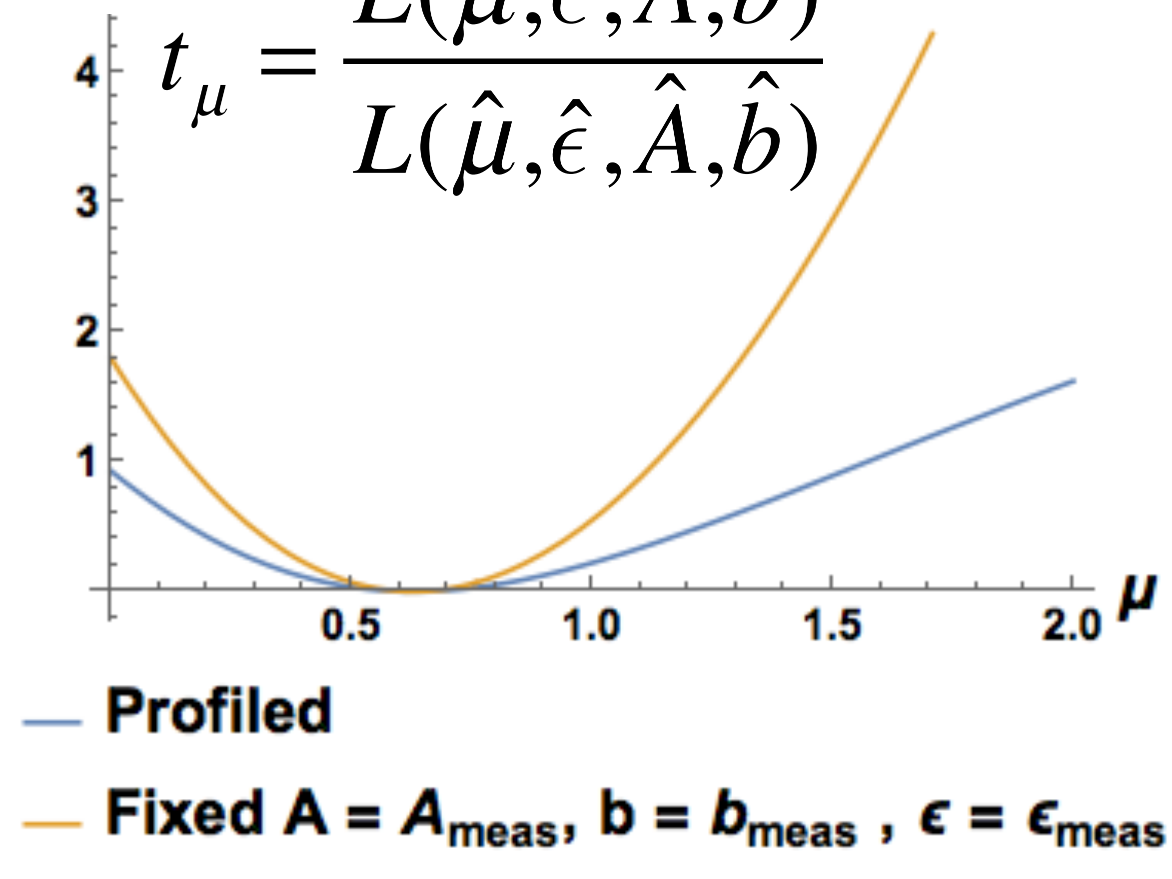
$$3poi:\ t_{\mu,\epsilon A} = \frac{L(\mu,\epsilon,A,\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

$$L(\mu,\epsilon.A) = Poiss(n \mid \mu\epsilon A + b) G(A_{meas} \mid A,\sigma_A) G(\epsilon_{meas} \mid \epsilon,\sigma_\epsilon) G(b_{meas} \mid b,\sigma_b)$$

$$L(\mu,\varepsilon,A) = \frac{(\mu\varepsilon As + b)^n}{n!} e^{-(\mu\varepsilon As+b)} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2} \frac{1}{\sigma_b\sqrt{2\pi}} e^{-(b_{meas}-b)^2/2\sigma_b^2} \frac{1}{\sigma_A\sqrt{2\pi}} e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

Profile Likelihood for Measurement

$$t_\mu = \frac{L(\mu,\hat{\hat{\epsilon}},\hat{\hat{A}},\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

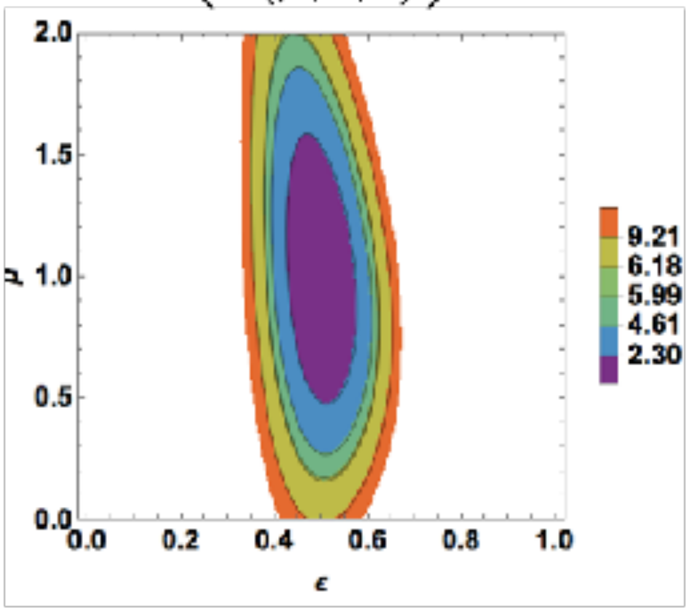


A toy case with 2 poi

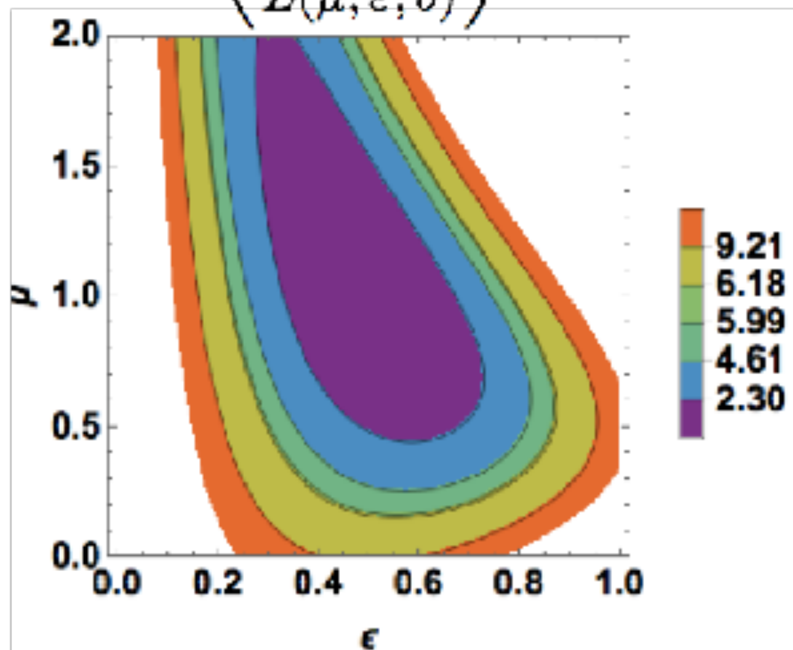
$n = \mu \epsilon A s + b$

$L = L(\mu, \epsilon, A, b)$

$$-2log\left(\frac{L(\mu,\varepsilon,\hat{\hat{b}})}{L(\hat{\mu},\hat{\varepsilon},\hat{b})}\right)$$



$$-2log\left(\frac{L(\mu,\varepsilon,\hat{\hat{b}})}{L(\hat{\mu},\hat{\varepsilon},\hat{b})}\right)$$



background = 100
signal = 90
ε = 0.5
σ_ε **= 0.05**
σ_b = 20

background = 100
signal = 90
ε = 0.5
σ_ε **= 0.15**
σ_b = 10

Profile Likelihood

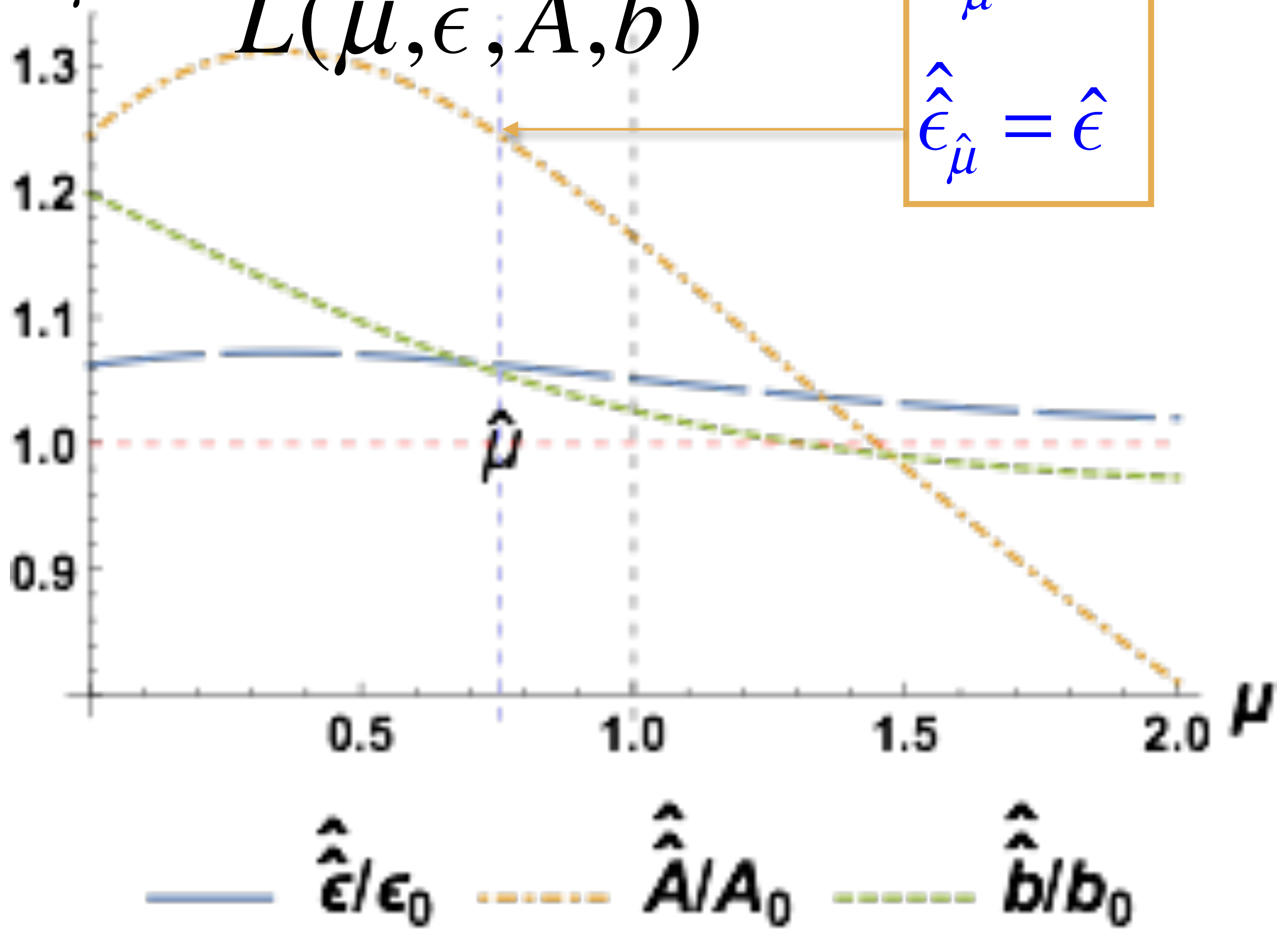
background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε = 0.05
σ_b = 10
σ_A = 0.2

n_{meas} =	137
b_{meas} =	105.533
ϵ_{meas} =	0.531025
A_{meas} =	0.870554
μ_{meas} =	0.756304

$$t_\mu = \frac{L(\mu,\hat{\hat{\epsilon}},\hat{\hat{A}},\hat{\hat{b}})}{L(\hat{\mu},\hat{\epsilon},\hat{A},\hat{b})}$$

$$\hat{\hat{\theta}}_{\hat{\mu}} = \hat{\theta}$$

$$\hat{\hat{\epsilon}}_{\hat{\mu}} = \hat{\epsilon}$$



Wilks theorem in the presence of NPs

- Given n parameters of interest and any number of NPs, then

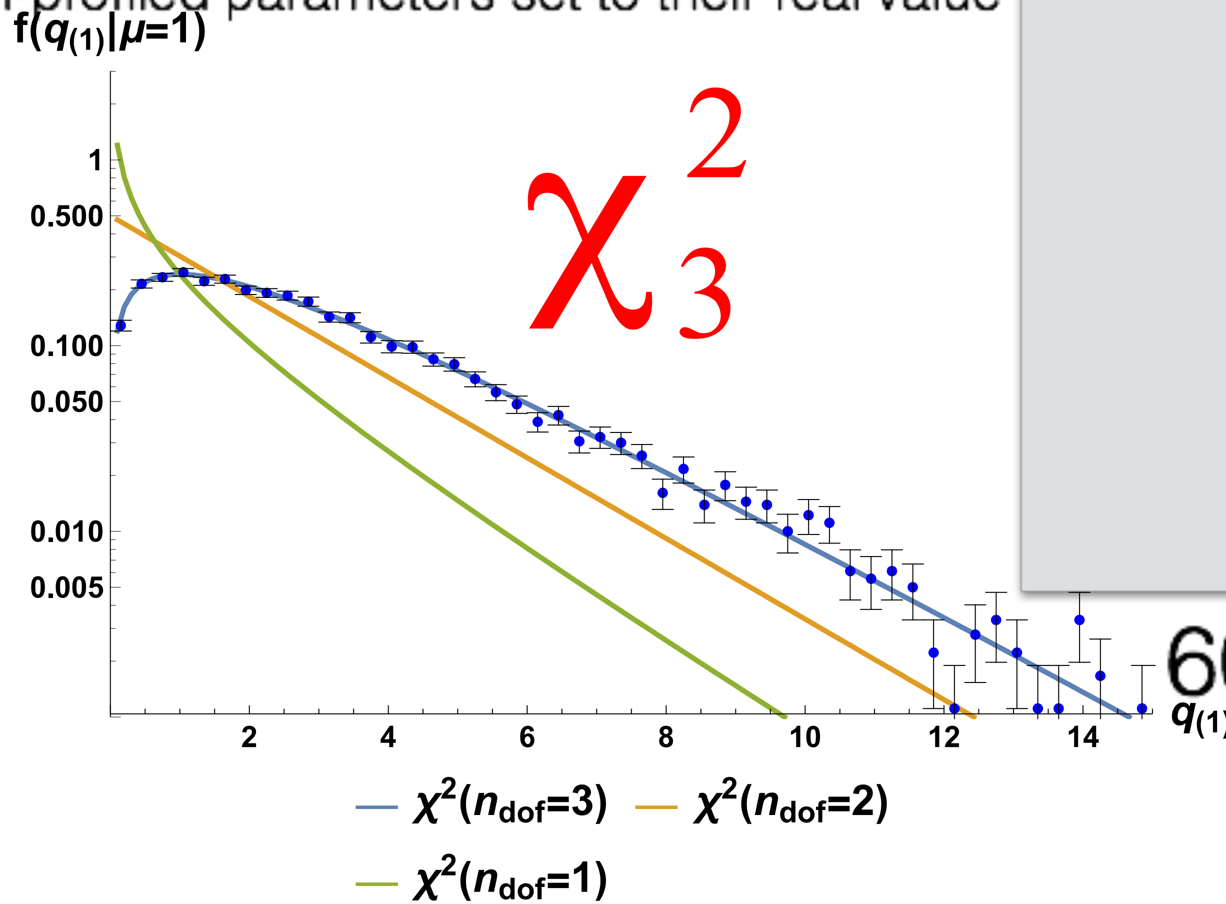
$$\lambda(\alpha_i) = \frac{L(\alpha_i, \hat{\hat{\theta}}_j)}{L(\hat{\alpha}_i, \hat{\theta}_j)}$$

$$q(\alpha_i) \equiv -2 \log \lambda(\alpha_i) \sim \chi_n^2$$

A toy case with 3 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

three parameters of interest (profiling only b)
non-profiled parameters set to their real value



χ^2_3

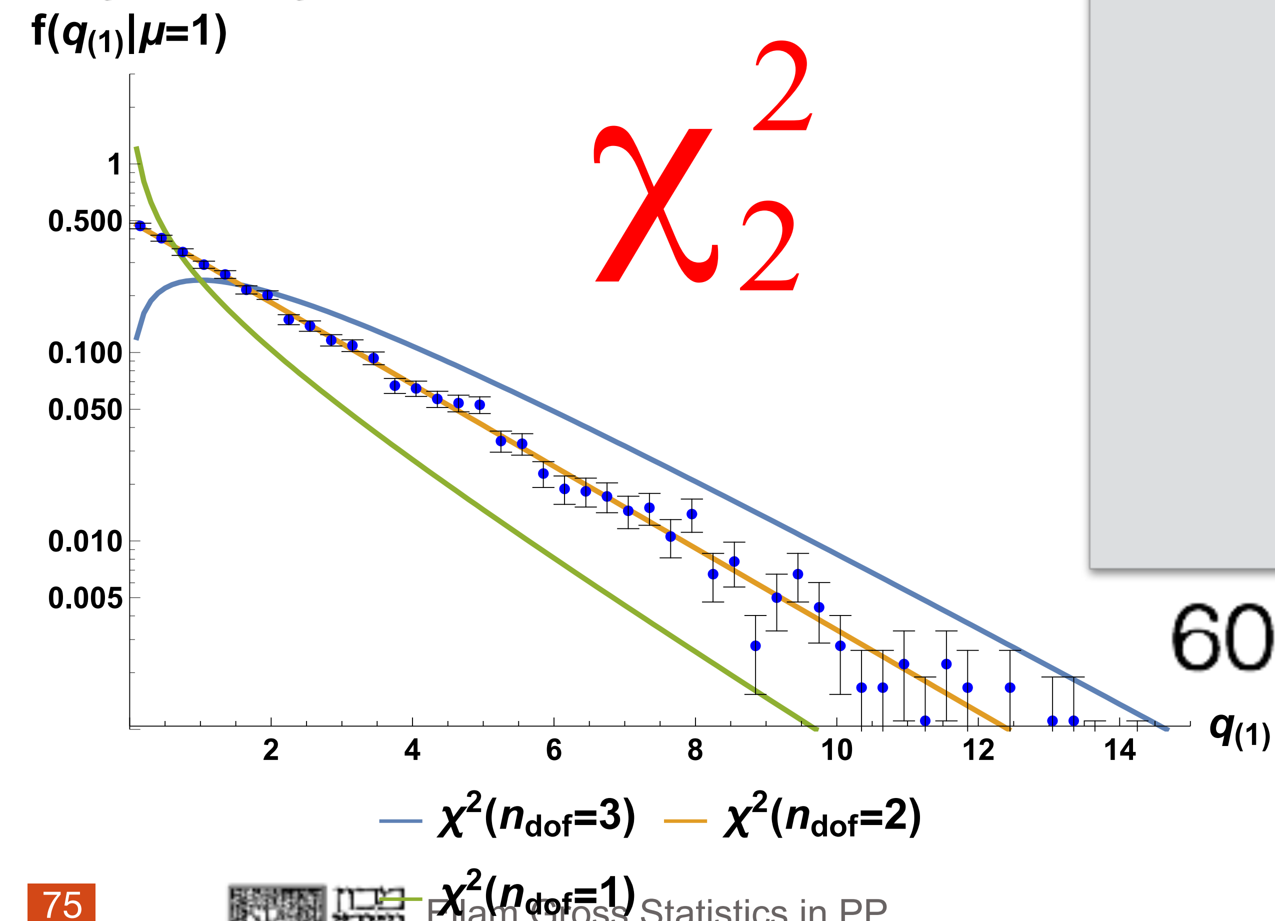
background = 100
signal = 90
ε = 0.5
A=0.7
σ_ε **= 0.05**
σ_b = 10
σ_A **= 0.2**

6000 events

A toy case with 2 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

two parameters of interest (profiling A and b)
non-profiled parameters set to their real value



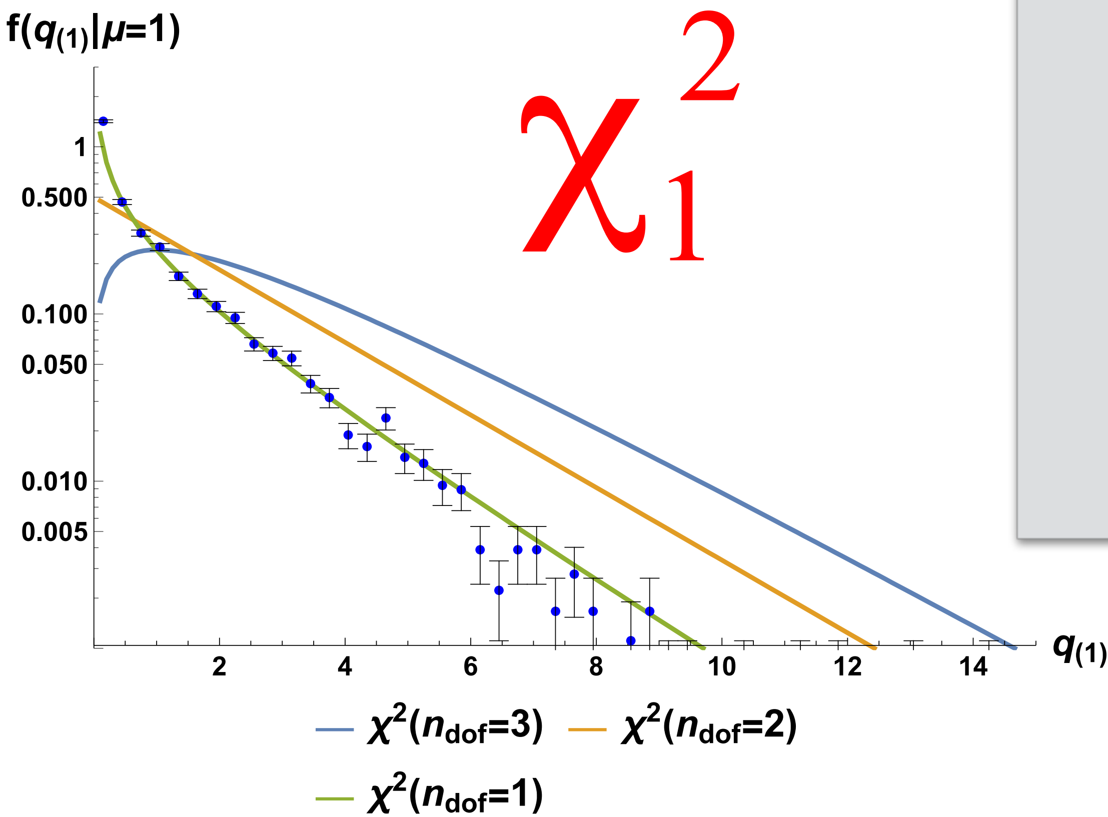
background = 100
signal = 90
ε = 0.5
A=0.7
= 0.05
= 10
= 0.2

6000 events

A toy case with 1 poi

$$L(\mu,\varepsilon,A)=\frac{(\mu\varepsilon As+b)^n}{n!}e^{-(\mu\varepsilon As+b)}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}e^{-(\varepsilon_{meas}-\varepsilon)^2/2\sigma_\varepsilon^2}\frac{1}{\sigma_b\sqrt{2\pi}}e^{-(b_{meas}-b)^2/2\sigma_b^2}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-(A_{meas}-A)^2/2\sigma_A^2}$$

one parameter of interest (profiling ε A and b)
non-profiled parameters set to their real value



χ_1^2

background = 100
signal = 90
$\varepsilon = 0.5$
A=0.7
σ_ε **= 0.05**
$\sigma_b = 10$
σ_A **= 0.2**

6000 events

Pulls and Ranking of NPs

The pull of θ_i is given by $\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}$

without constraint $\sigma\left(\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}\right) = 1 \quad \left\langle \dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0} \right\rangle = 0$

It's a good habit to look at the pulls of the NPs and make sure that Nothing irregular is seen

In particular one would like to guarantee that the fits do not over constrain a NP in a non sensible way

Random Data Set

n_{meas} = 132

b_{meas} = 103.208

ϵ_{meas} = 0.465459

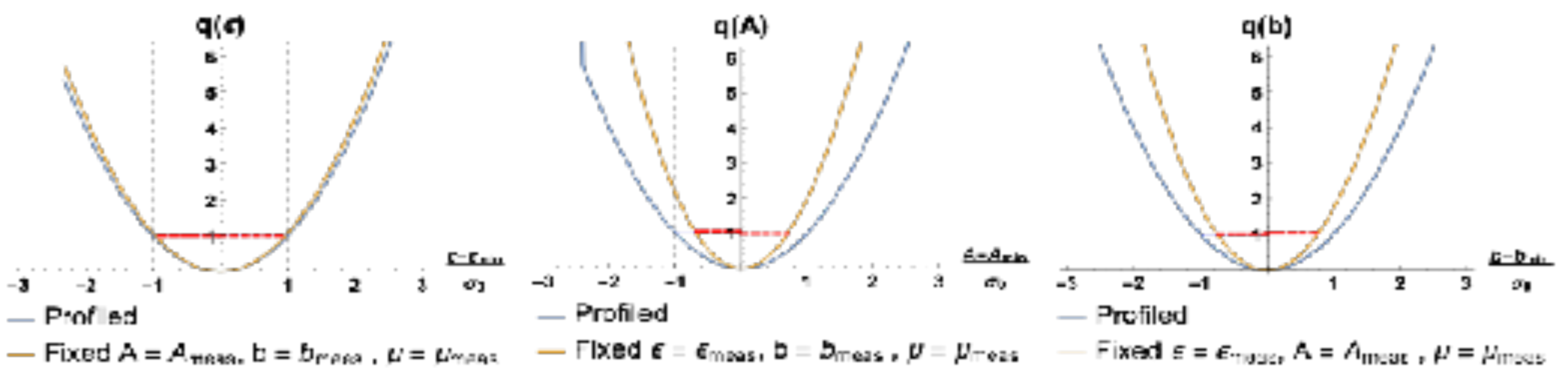
A_{meas} = 0.487107

μ_{meas} = 1.41099

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ_ε = 0.05
σ_b = 10
σ_A = 0.2

To get the pulls:
–scan $q(\epsilon)$
–Find $\hat{\epsilon}$
–Find σ_ϵ^+ and σ_ϵ^- i.e. the poitive and negative error bar substituting $q(\epsilon)=1$



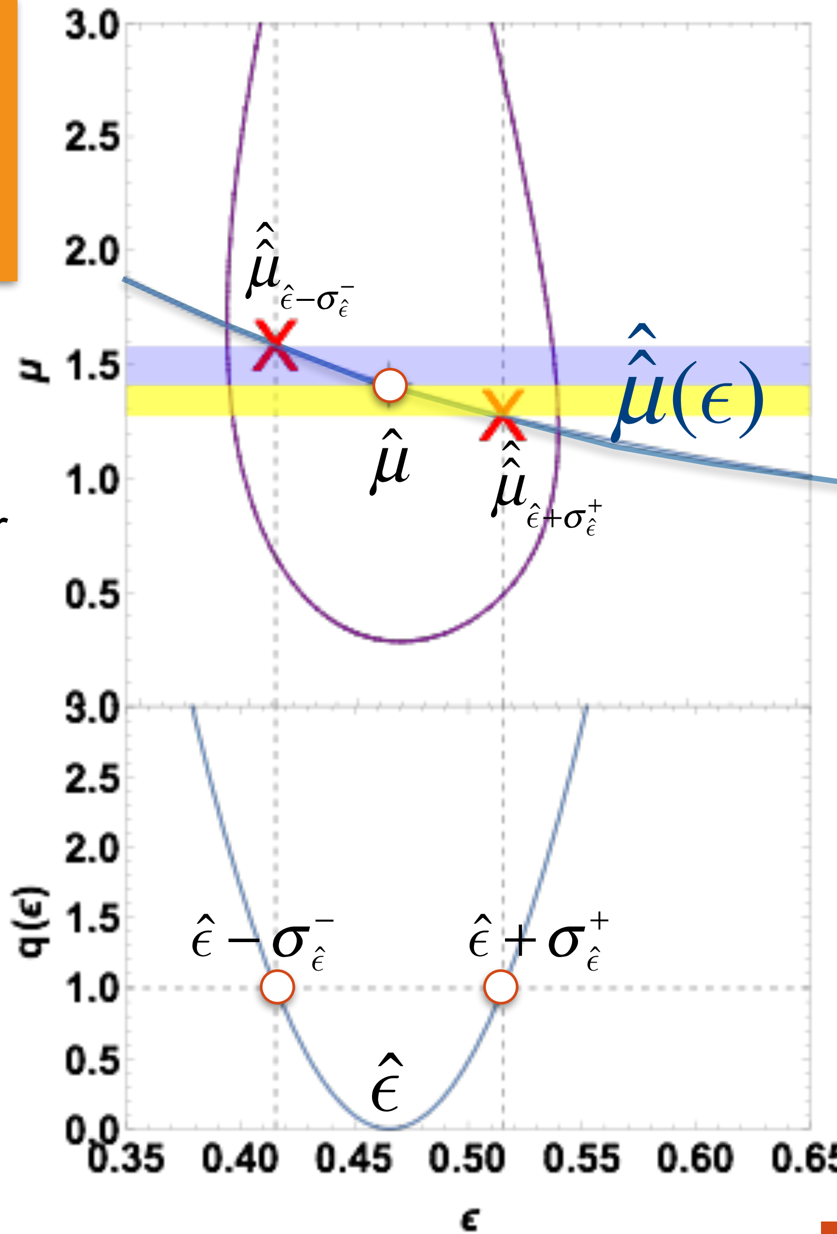
With the random data sets we find perfect pulls for the profiled scans
But not for the fix scans!

Random Data Set: Find the Impact of NP

n_{meas} = 132
b_{meas} = 103.208
ϵ_{meas} = 0.465459
A_{meas} = 0.487107
μ_{meas} = 1.41099

reminder:
$b_0 = 100$
$\varepsilon_0 = 0.5$
$A_0 = 0.7$
$\mu_0 = 1$
$n_0 = 131.5$
signal = 90

$\sigma_\varepsilon = 0.05$
$\sigma_b = 10$
$\sigma_A = 0.2$



To get the impact of a Nuisance Parameter in order to rank them:

Say we want the impact of ϵ

– Scan q(ϵ), profiling all other NPs

– Find $\hat{\epsilon}$

– (note that $\hat{\mu}_{\hat{\epsilon}} = \hat{\mu}$)

– Find $\hat{\mu}_{\hat{\epsilon}\pm\sigma_{\epsilon}^{\pm}} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_{\epsilon}^{\pm}}$

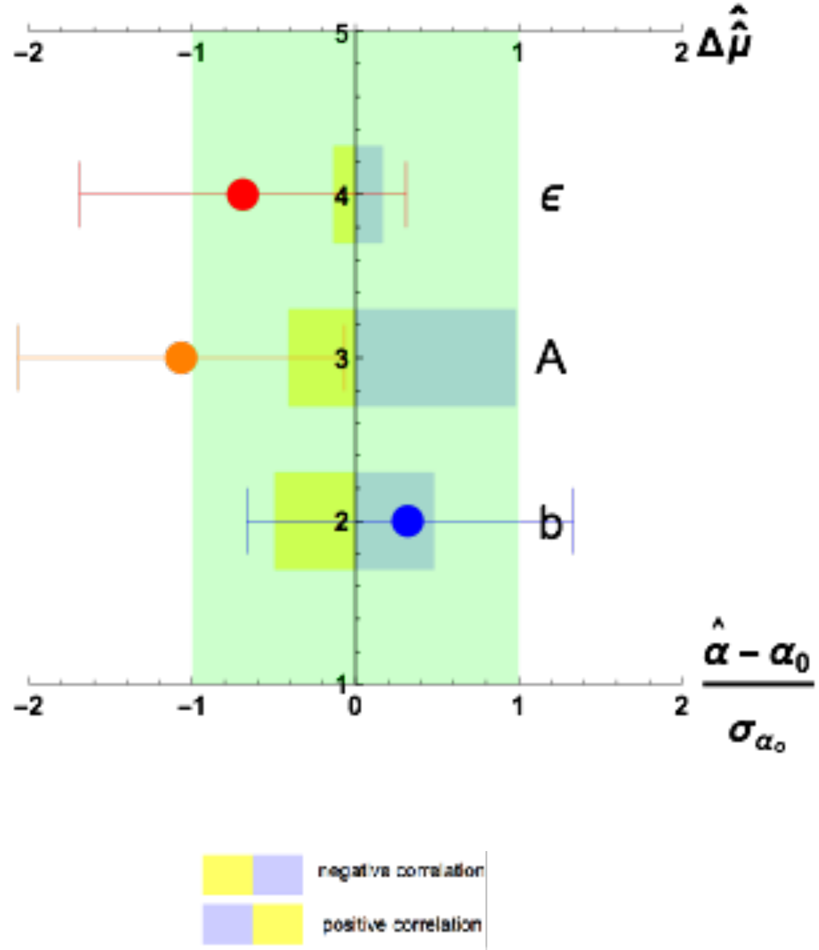
– The impact is given by $\Delta\mu^{\pm} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_{\epsilon}^{\pm}} - \hat{\mu}$

Random Data Set: SUMMARY of Pulls and Impact

n_{meas} = 132
b_{meas} = 103.208
ϵ_{meas} = 0.465459
A_{meas} = 0.487107
μ_{meas} = 1.41099

reminder:
b_0 = 100
ε_0 = 0.5
A_0 =0.7
μ_0 = 1
n_0=131.5
signal =90

σ_ϵ = 0.05
σ_b = 10
σ_A = 0.2

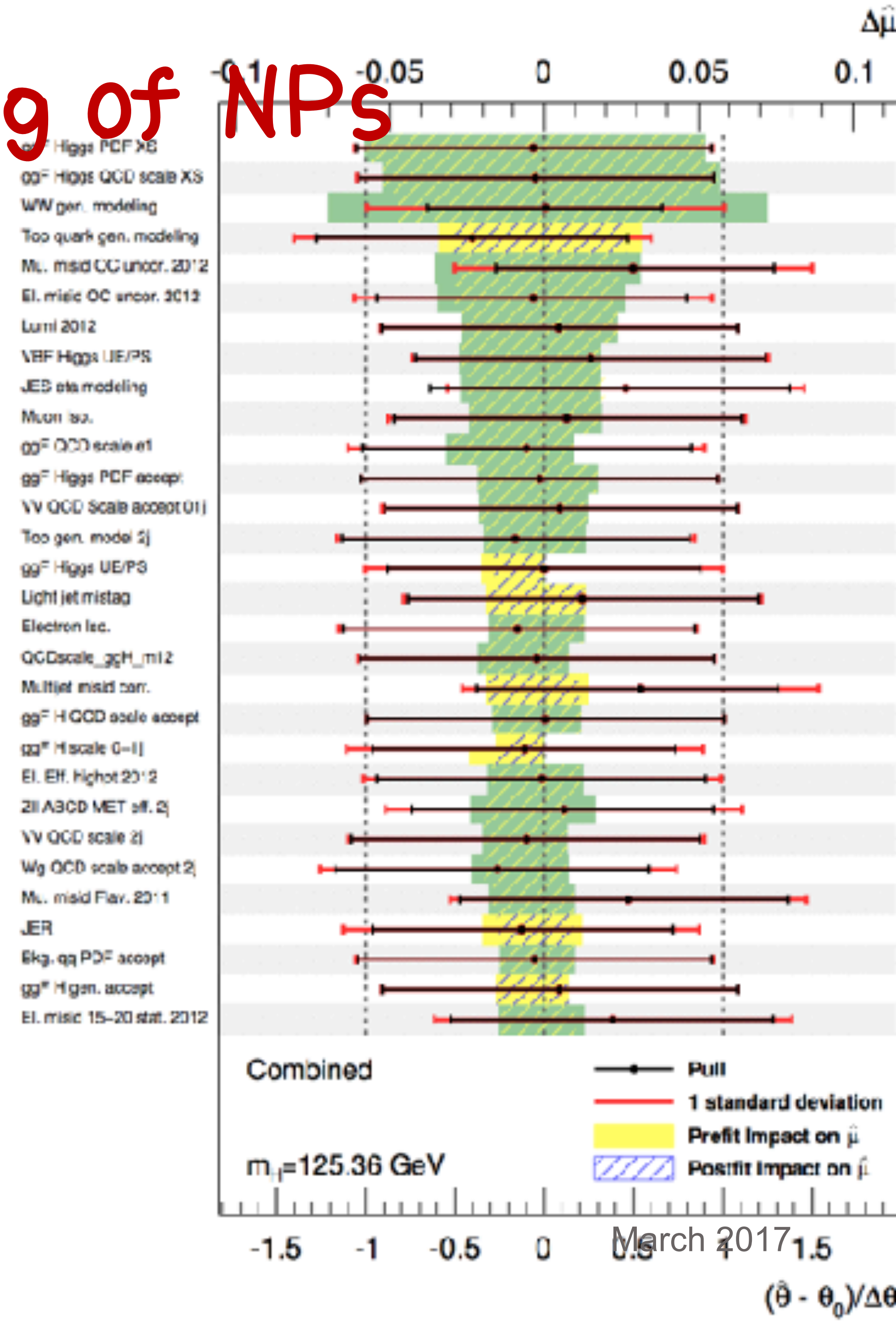


Pulls and Ranking of NPs

Ranking θ_i by its effect in the NP

$$\Delta\mu^{\pm} = \hat{\hat{\mu}}_{\hat{\epsilon}\pm\sigma_{\epsilon}^{\pm}} - \hat{\mu}$$

By ranking we can tell which NPs are the important ones and which can be pruned



If time permits:
The Feldman Cousins Unified Method

The Flip Flop Way of an Experiment

- The most intuitive way to analyze the results of an experiment would be

  If the significance based on q_{obs}, is less than 3 sigma, derive an upper limit (just looking at tables), if the result is >5 sigma derive a discovery central confidence interval for the measured parameter (cross section, mass….)

- **This Flip Flopping policy leads to undercoverage:**
  Is that really a problem for Physicists?
  Some physicists say, for each experiment quote always two results, an upper limit, and a (central?) discovery confidence interval

- Many LHC analyses report both ways.

Frequentist Paradise – F&C Unified with Full Coverage

- Frequentist Paradise is certainly made up of an interpretation by constructing a confidence interval in brute force ensuring a coverage!
- This is the Neyman confidence interval adopted by F&C....

- The motivation:
  - Ensures Coverage
  - Avoid Flip-Flopping – an ordering rule determines the nature of the interval
    (1-sided or 2-sided depending on your observed data)
  - Ensures Physical Intervals

- Let the test statistics be

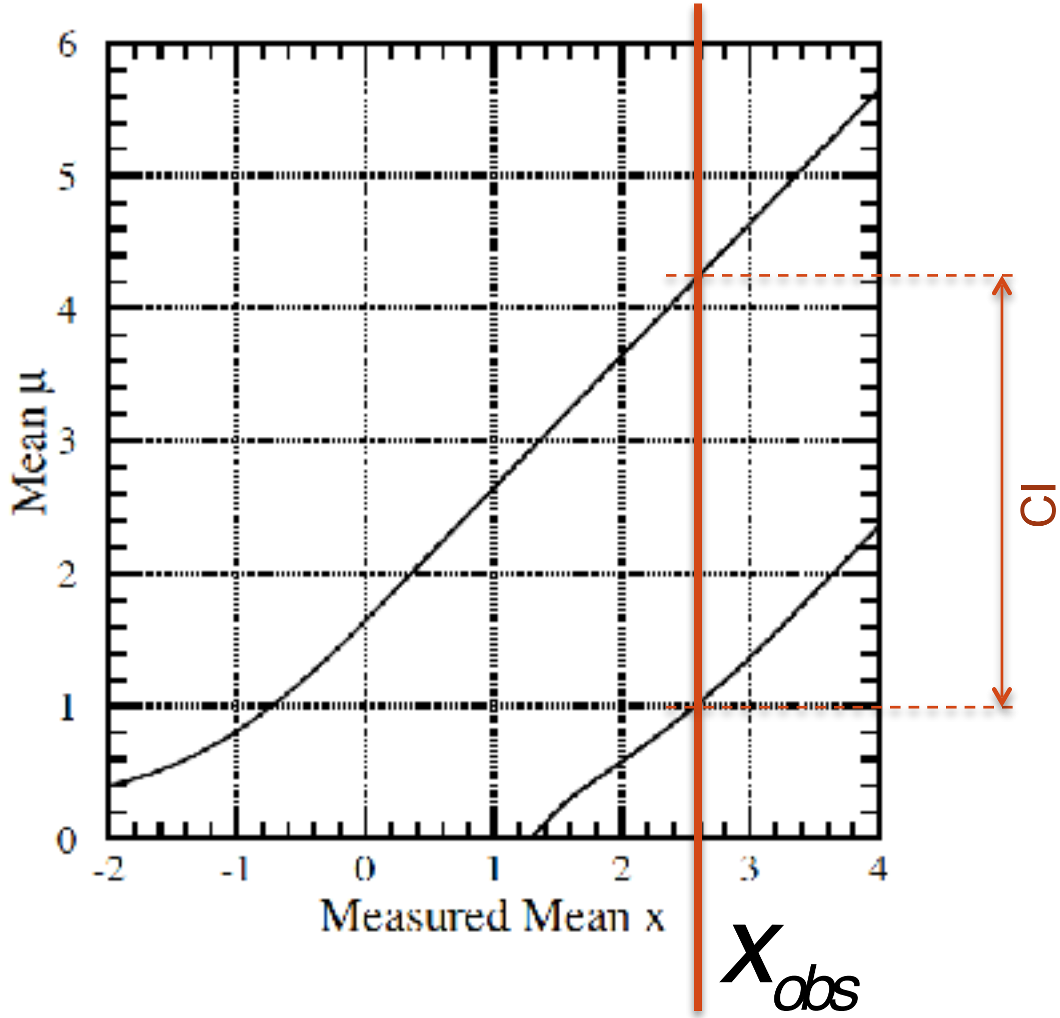
$$q=\begin{cases} -2\ln\dfrac{L(s+b)}{L(\hat{s}+b)} & \hat{s}\geq 0 \\ -2\ln\dfrac{L(s+b)}{L(b)} & \hat{s}<0 \end{cases}$$

  where $\hat{s}$ is the physically allowed mean s that maximizes $L(\hat{s}+b)$
  (protect a downward fluctuation of the background, $n_{obs}>b$; $\hat{s}>0$)

- Order by taking the 68% highest q's

How to tell an Upper limit from a Measurement without Flip Flopping

- A measurement (2 sided)



How to tell an Upper limit from a Measurement without Flip Flopping

- An upper limit (1 sided)