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

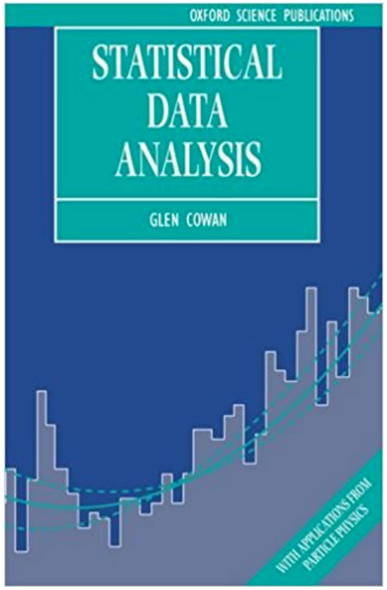
- **Lecture 1:** Basic Concepts
  Histograms, Testing Hypotheses,
  LR as a Test Statistics, p-value, POWER, CLs
  Measurements, Neyman Construction and Feldman Cousins
- **Lecture 2:** Wald Theorem, Asymptotic Formalism, Asimov Data Set, PL, Asimov Significance, Look Elsewhere Effect

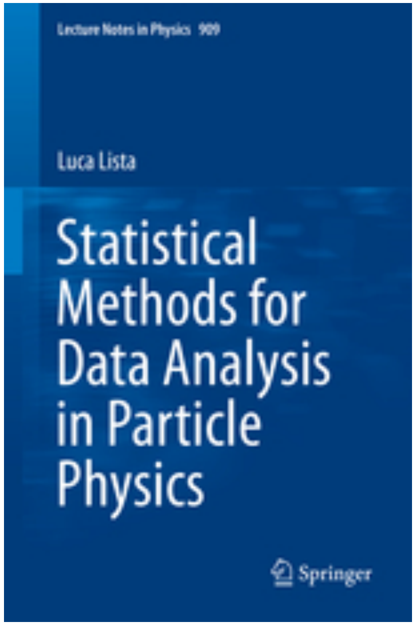
Support Material

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

Histograms

N *collisions*

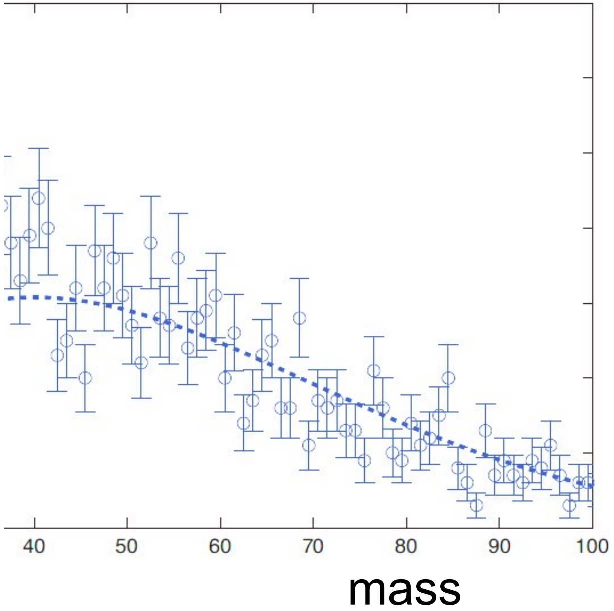
$$p(Higgs\ event) = \frac{\mathcal{L}\sigma(pp \to H)\epsilon_{ff}}{\mathcal{L}\sigma(pp)}$$

Prob *to see* n_H^{obs} *in N collisions is*

$$P(n_H^{obs}) = \binom{N}{n_H^{obs}} p^{n_H^{obs}} (1-p)^{N-n_H^{obs}}$$

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

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



A counting experiment

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

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

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

ALTERNATE

H_0- SM w/o Higgs

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

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

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

$$P(x \mid H) \neq P(H \mid x)$$

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

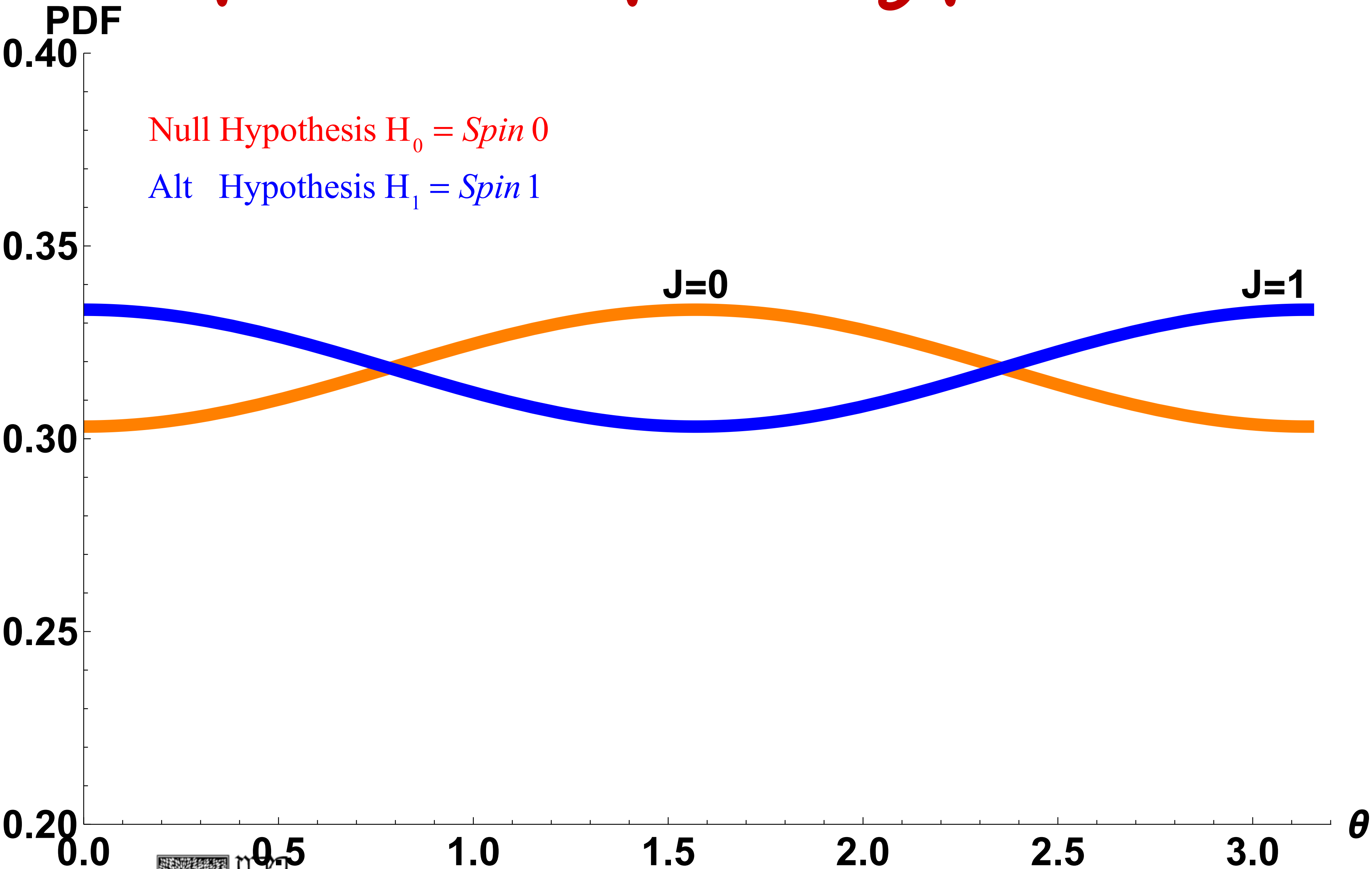
Testing an Hypothesis (wikipedia...)

- 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?

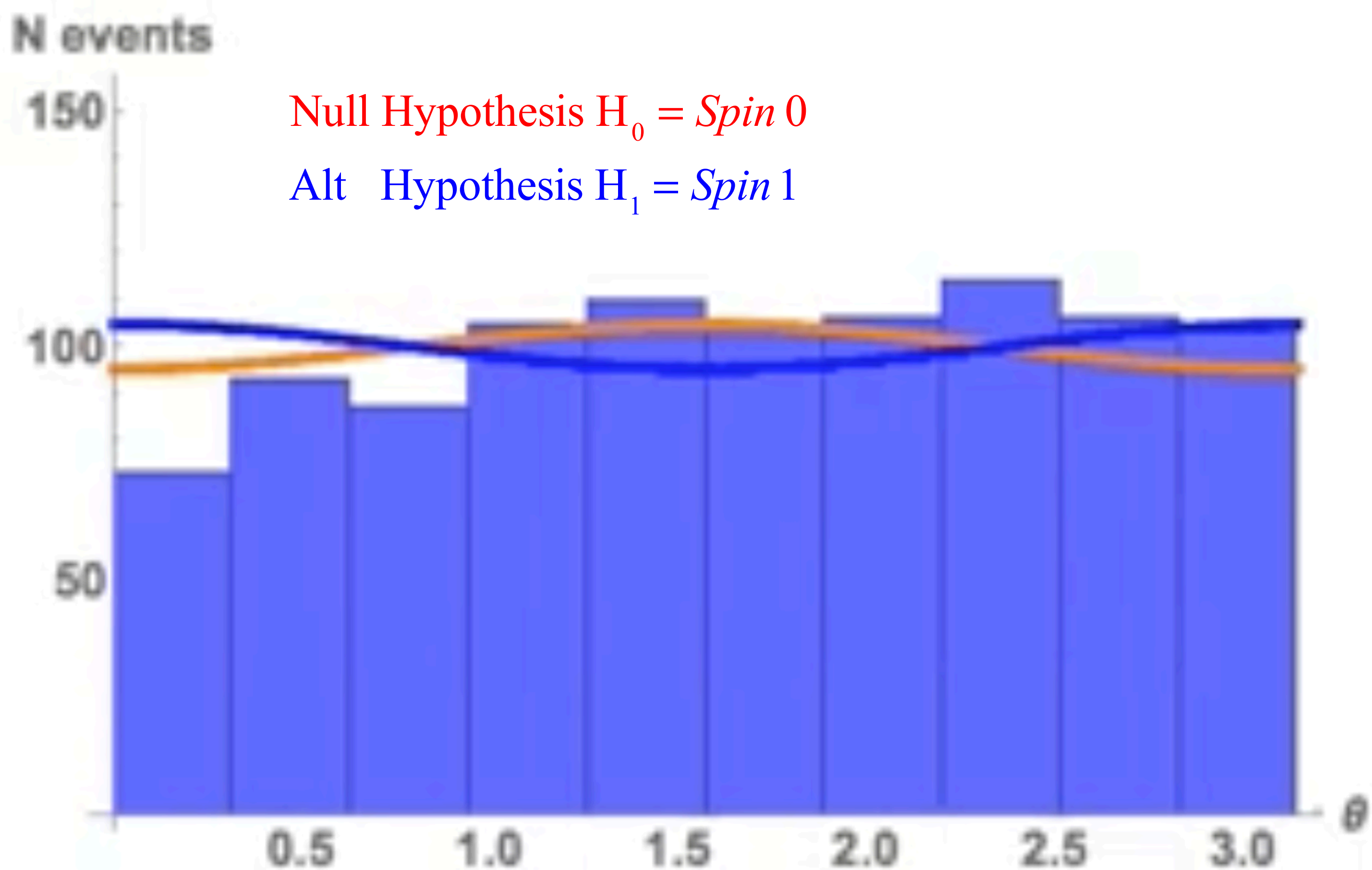
Test statistic and p-value

Case Study 1 : Spin

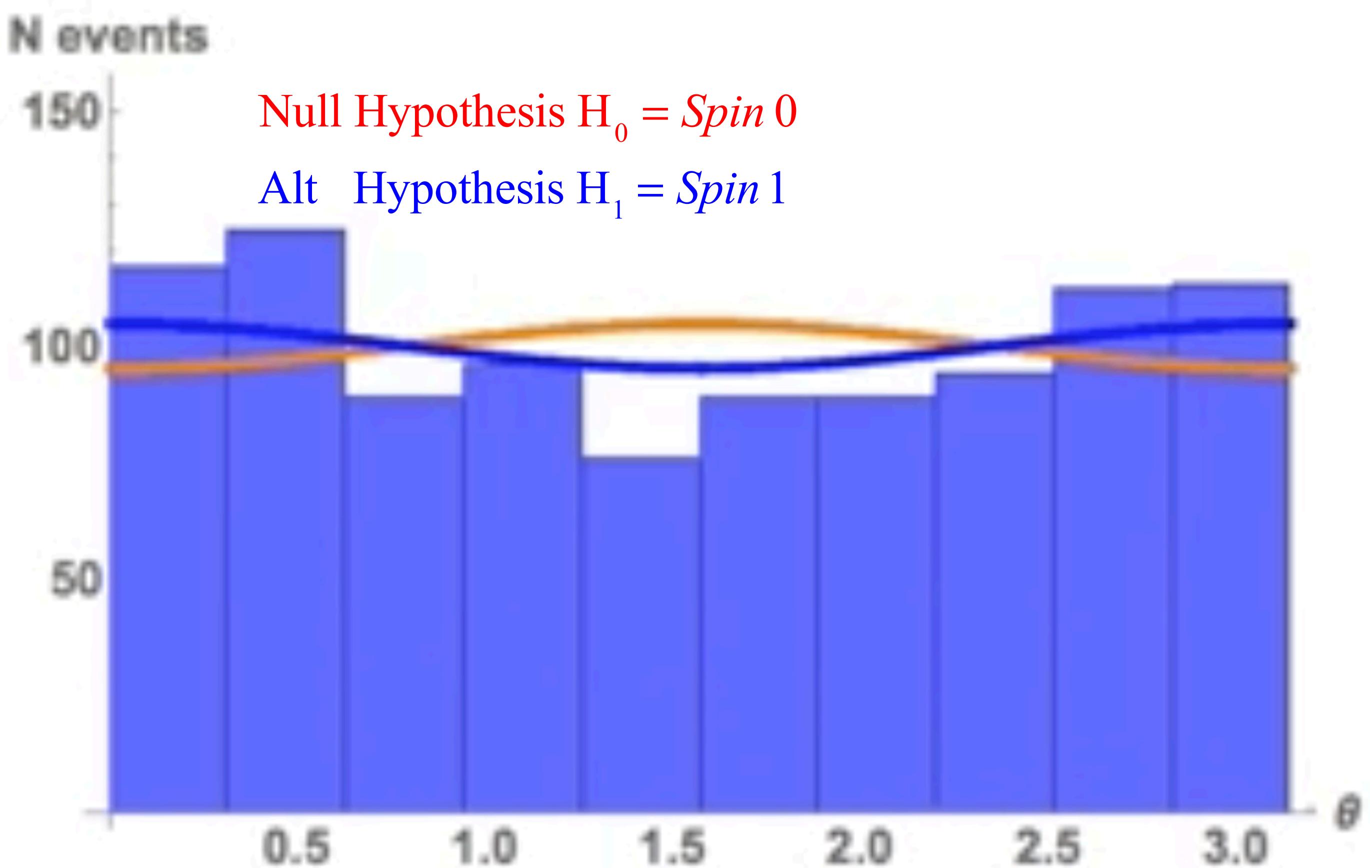
Spin 0 vs Spin 1 Hypotheses



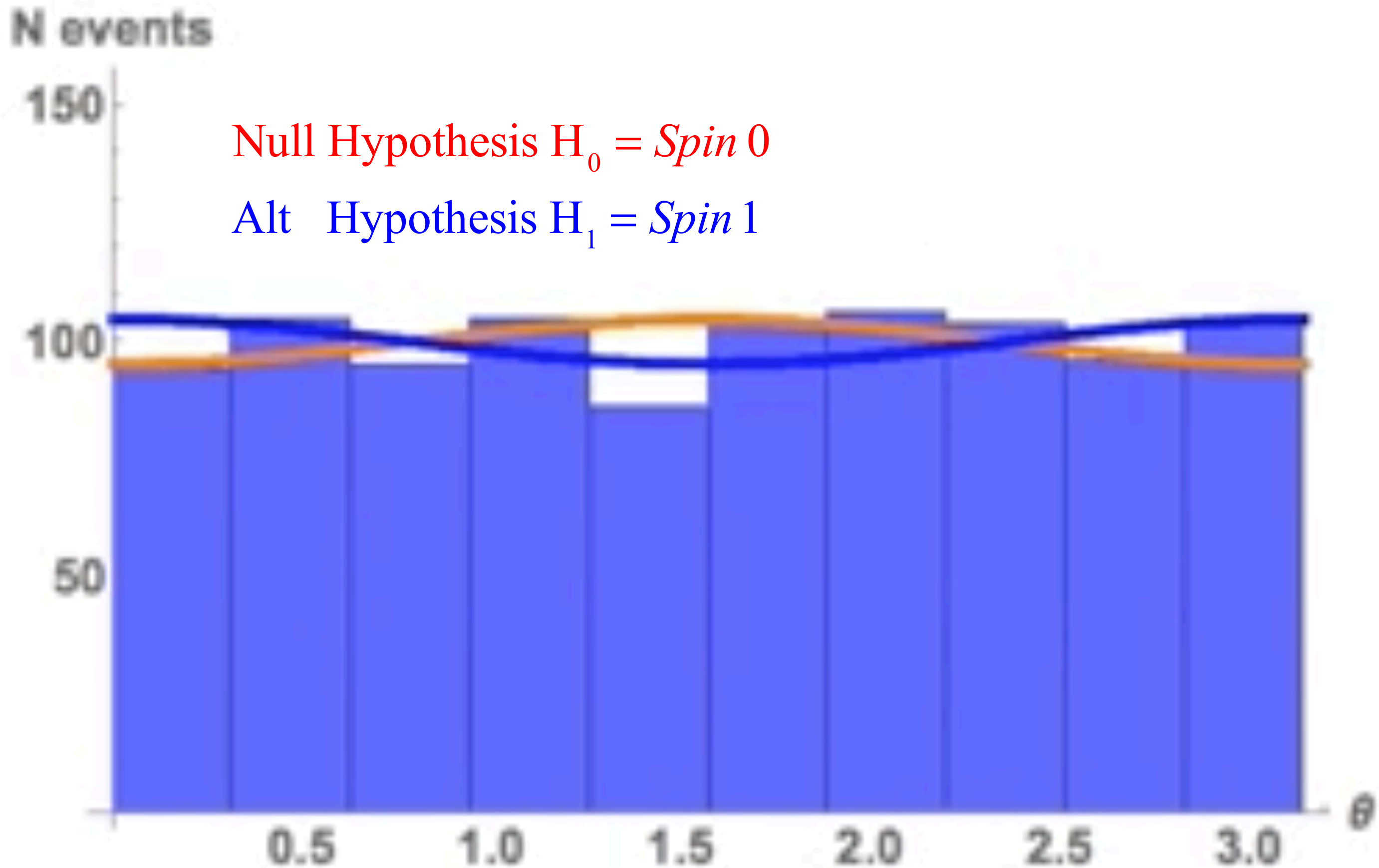
Spin 0 vs Spin 1 Hypotheses



Spin 0 vs Spin 1 Hypotheses



Spin 0 vs Spin 1 Hypotheses



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, $\lambda = \dfrac{L(H_1)}{L(H_0)}$

  which rejects H_0 in favor of H_1,

  is the **most powerful test**

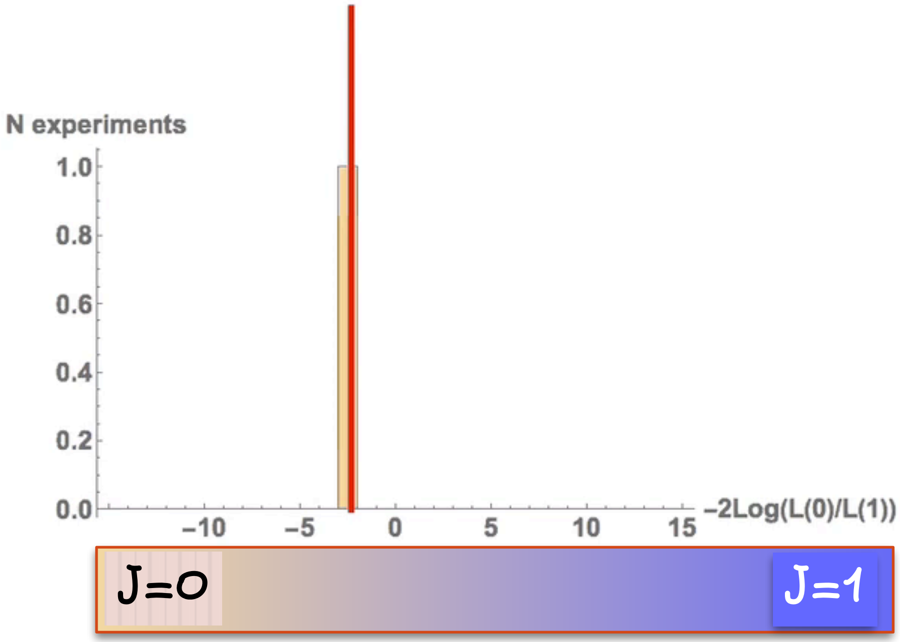
  for a given significance level $\alpha = prob(\lambda \le \eta)$

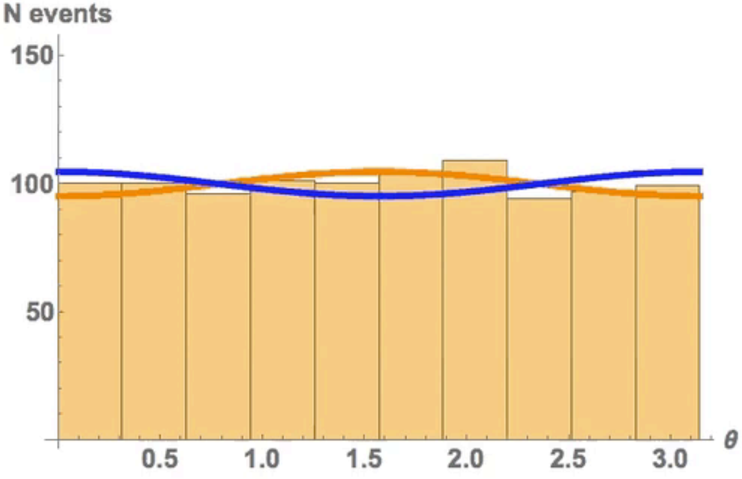
  with a threshold η

Building PDF

Build the pdf of the test statistic

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

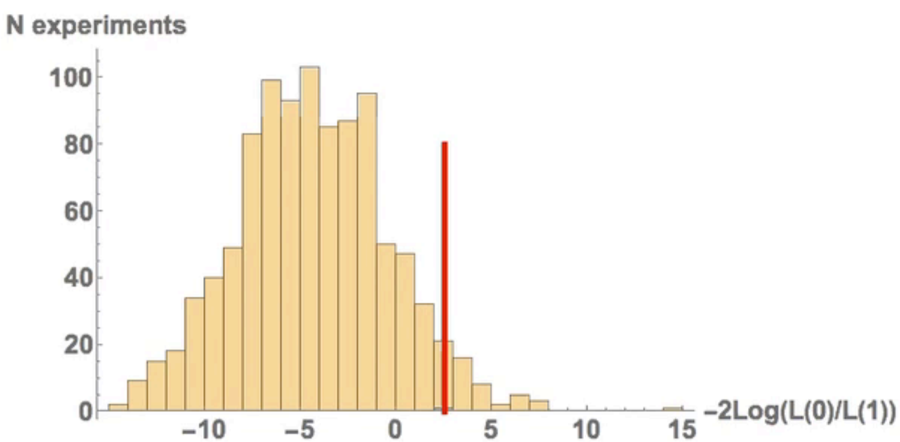




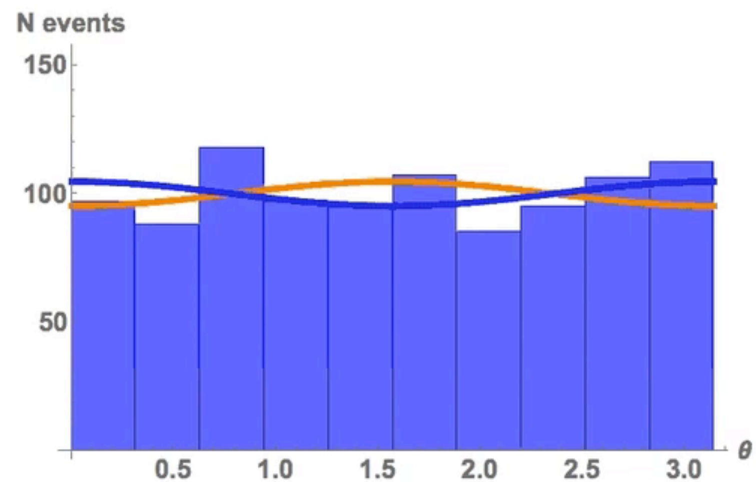
Building PDF

Build the pdf of the test statistic

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



J=0 J=1



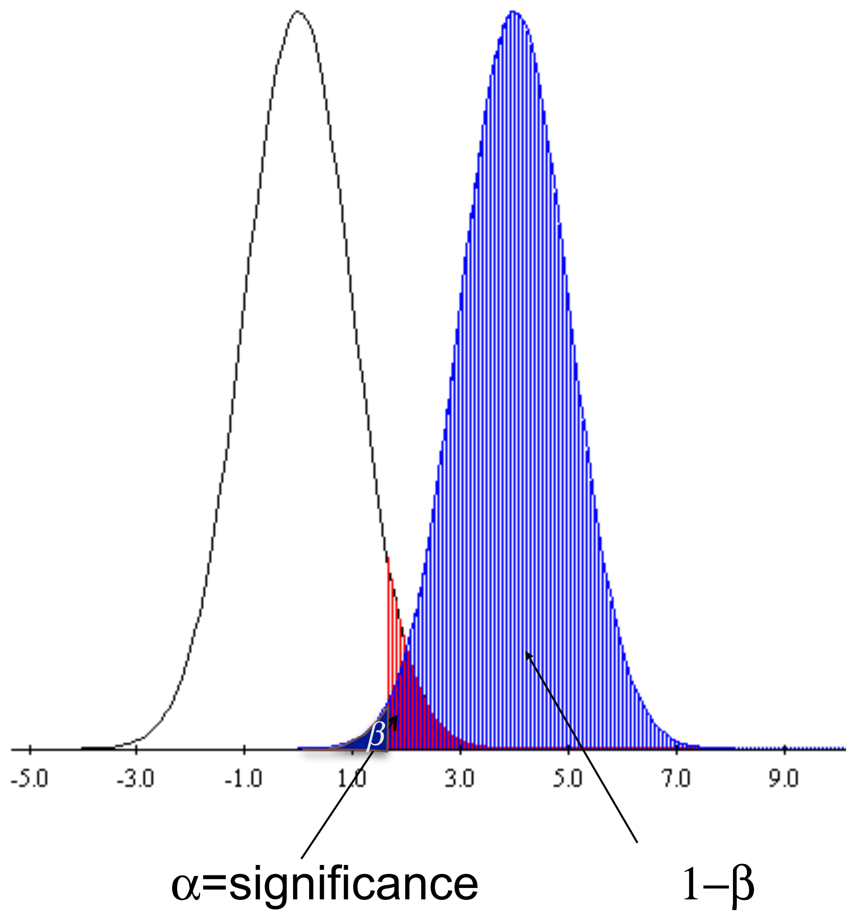
Basic Definitions: type I-II errors

- 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

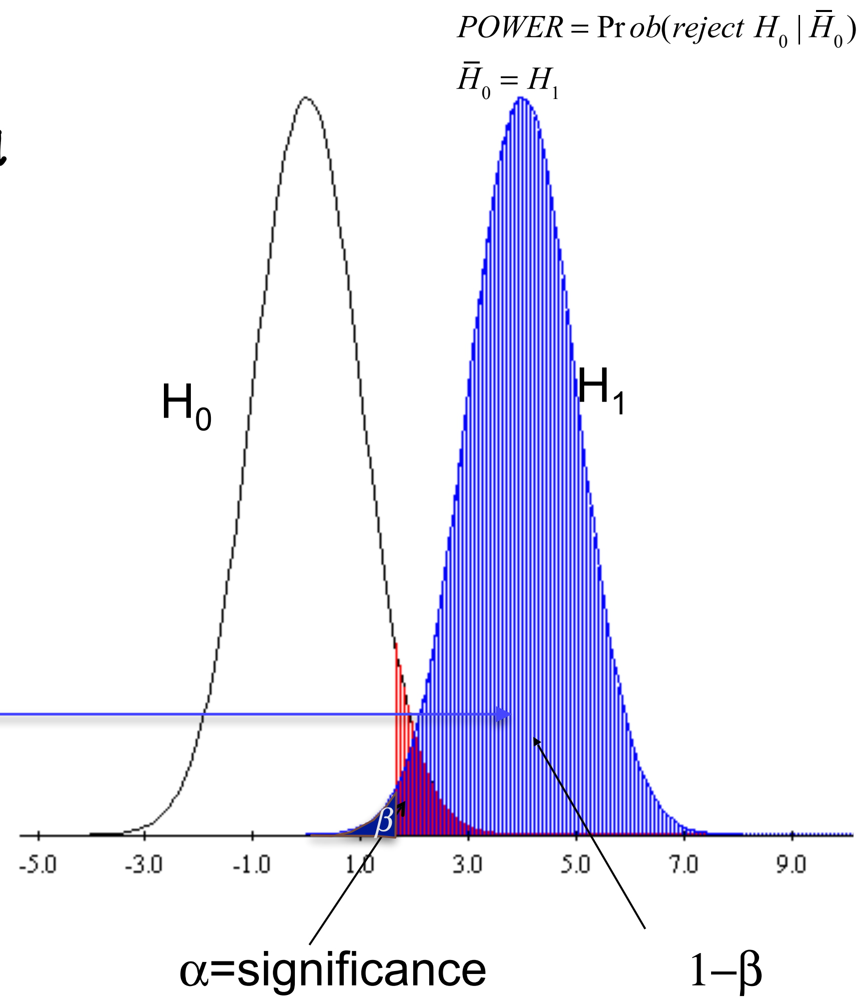
- $\alpha = \Pr ob(reject\ H_0 \mid H_0)$
- The POWER of an hypothesis test is the probability to reject the null hypothesis when it is indeed wrong (the alternate analysis is true)
- $POWER = \Pr ob(reject\ H_0 \mid \bar{H}_0)$

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

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

  $\bar{H}_0 = H_1$

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



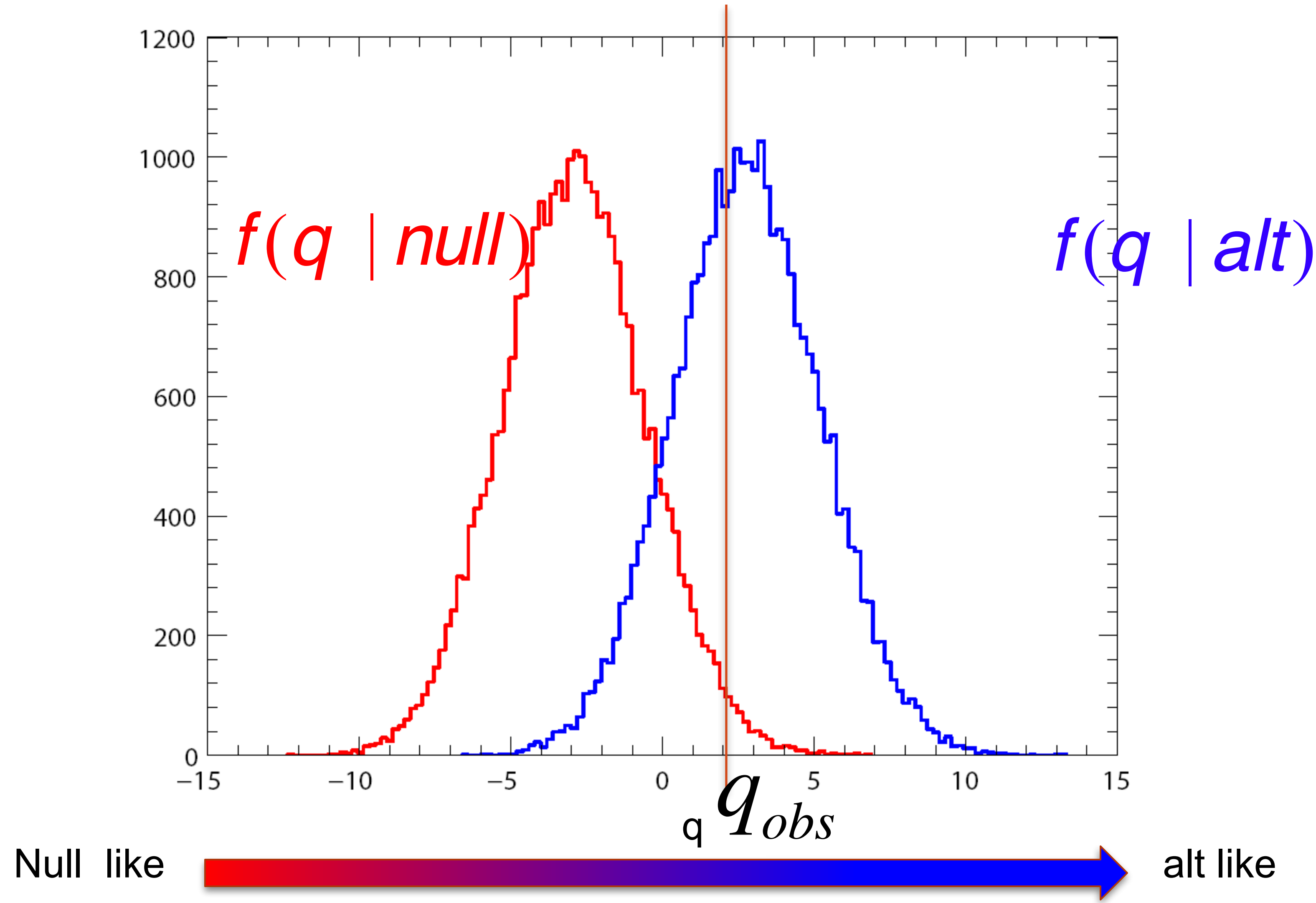
p-Value

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

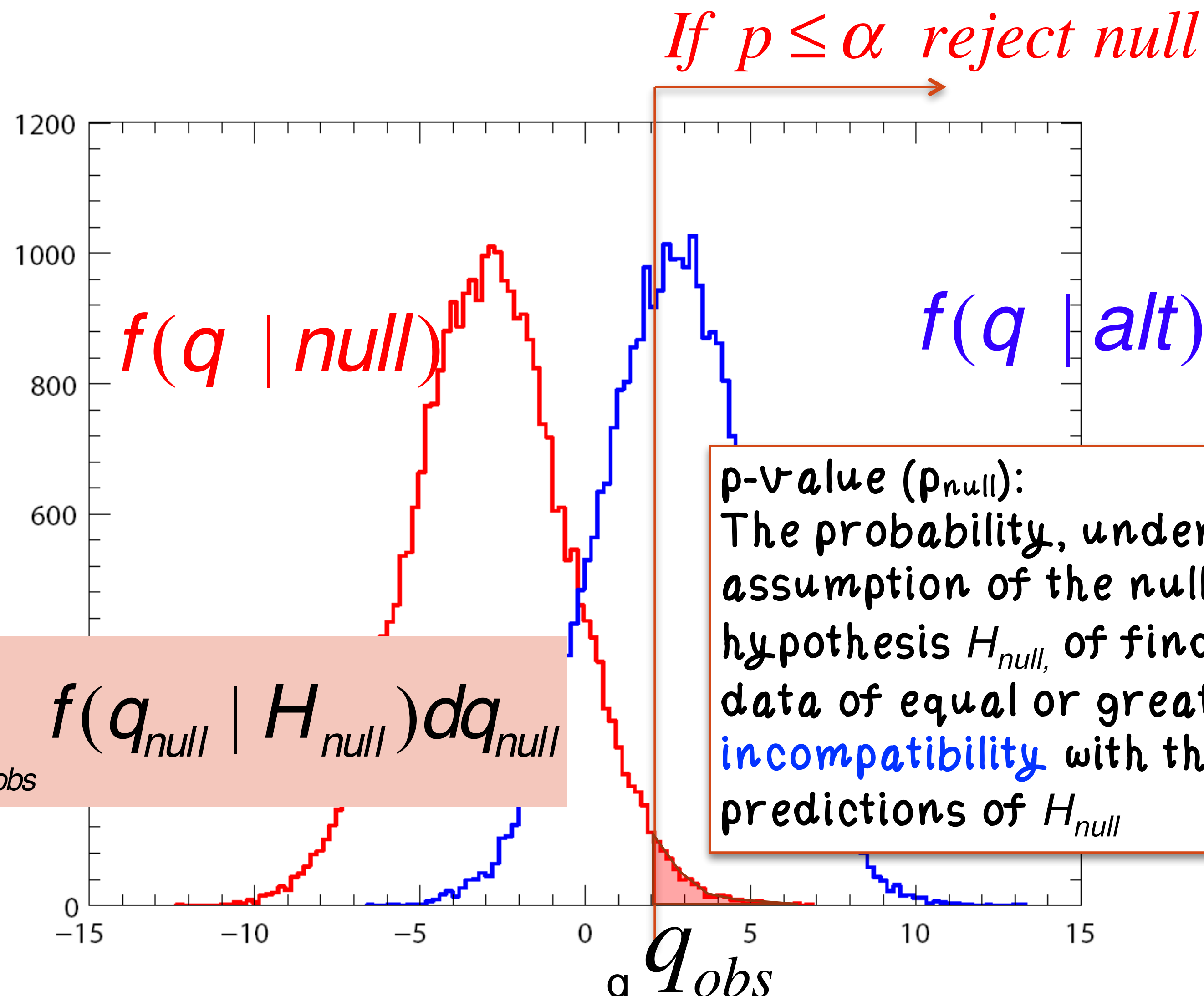
- It is the probability, under assumption of the null hypothesis $H_{null,}$ of finding data of equal or greater incompatibility with the predictions of H_{null}

- 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

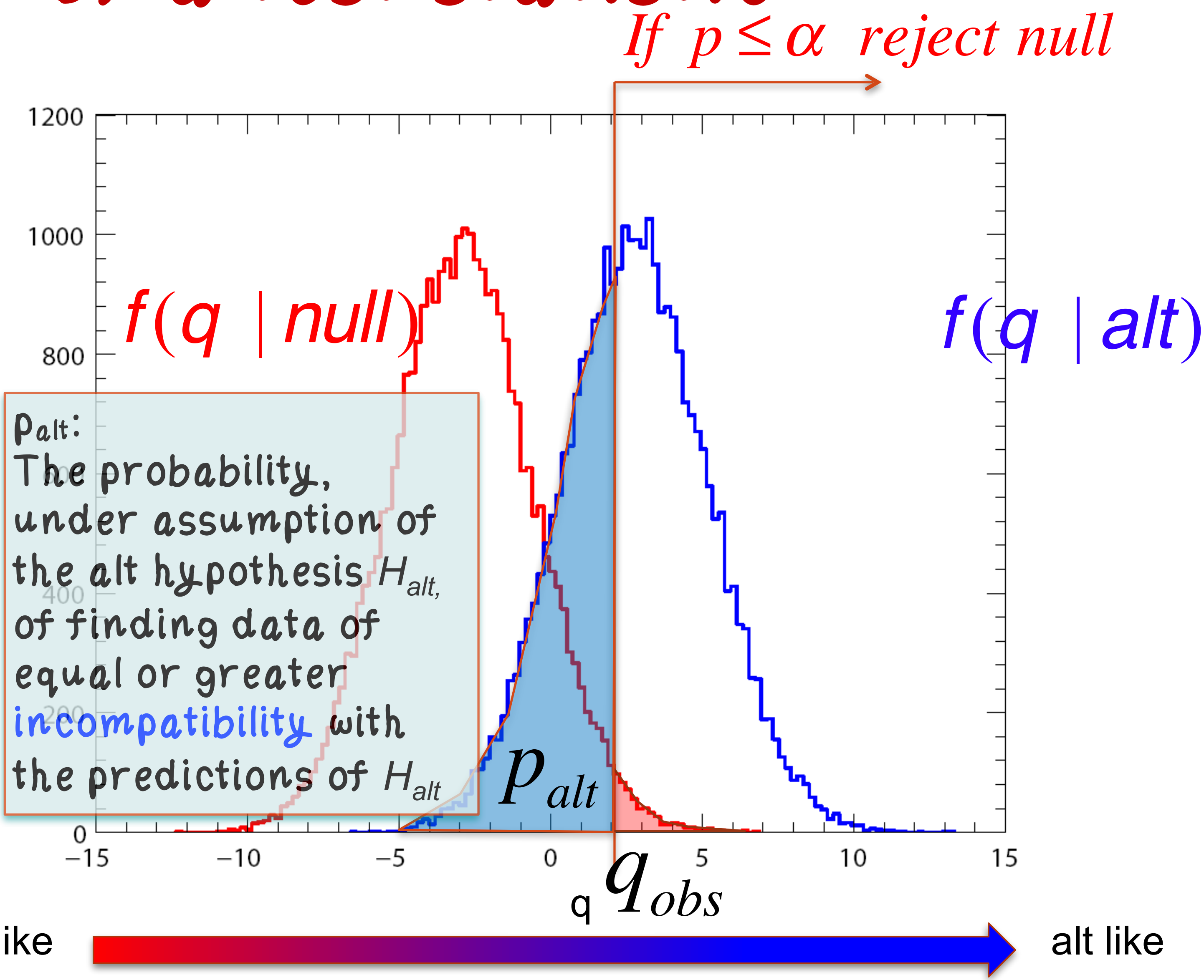


$$p = \int_{q_{obs}}^{\infty} f(q_{null} \mid H_{null})\, dq_{null}$$

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}

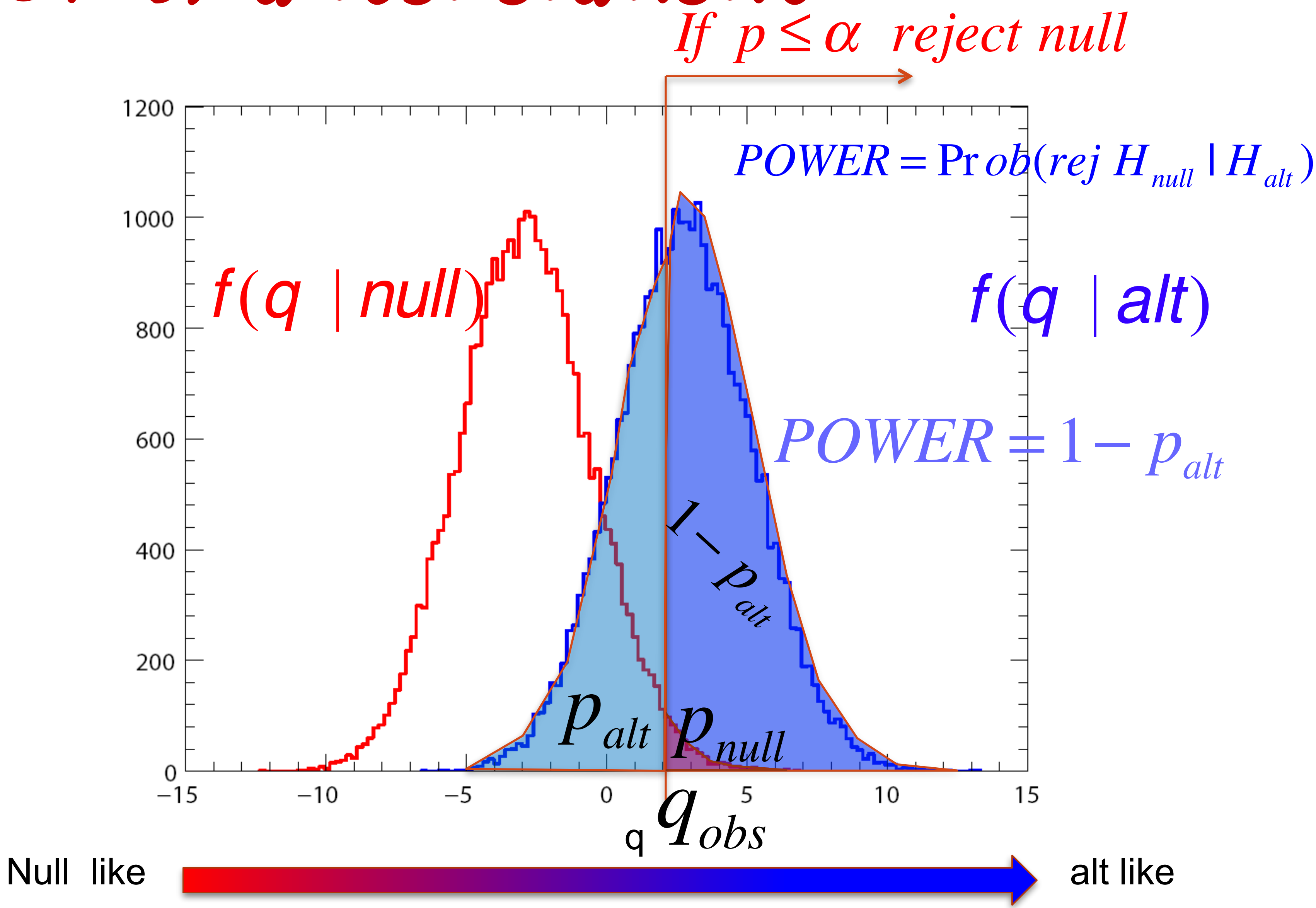
Null like → alt like

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}

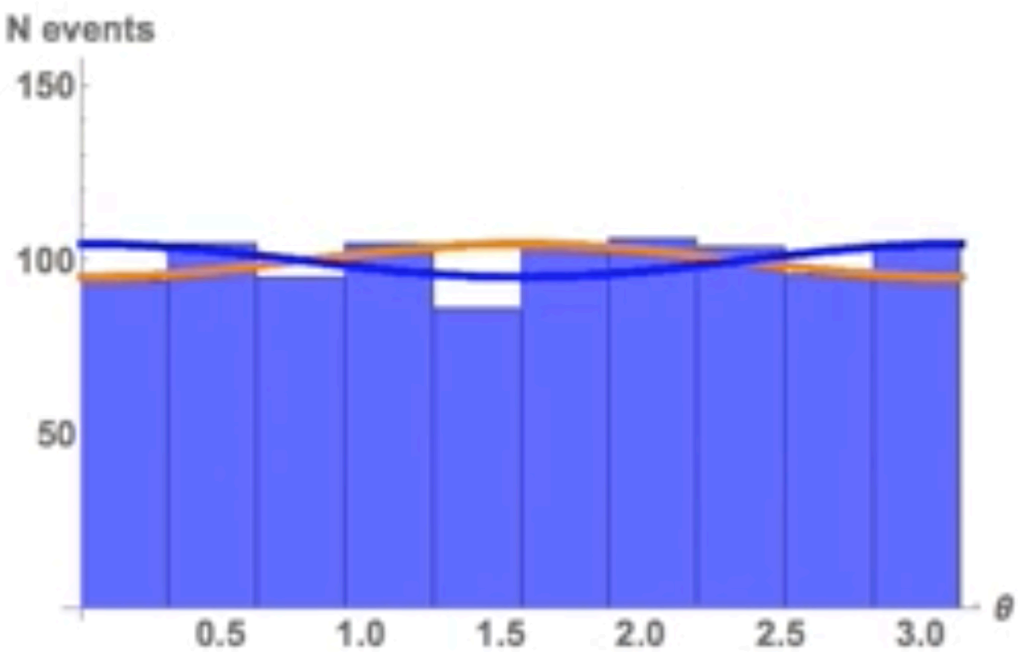
PDF of a test statistic

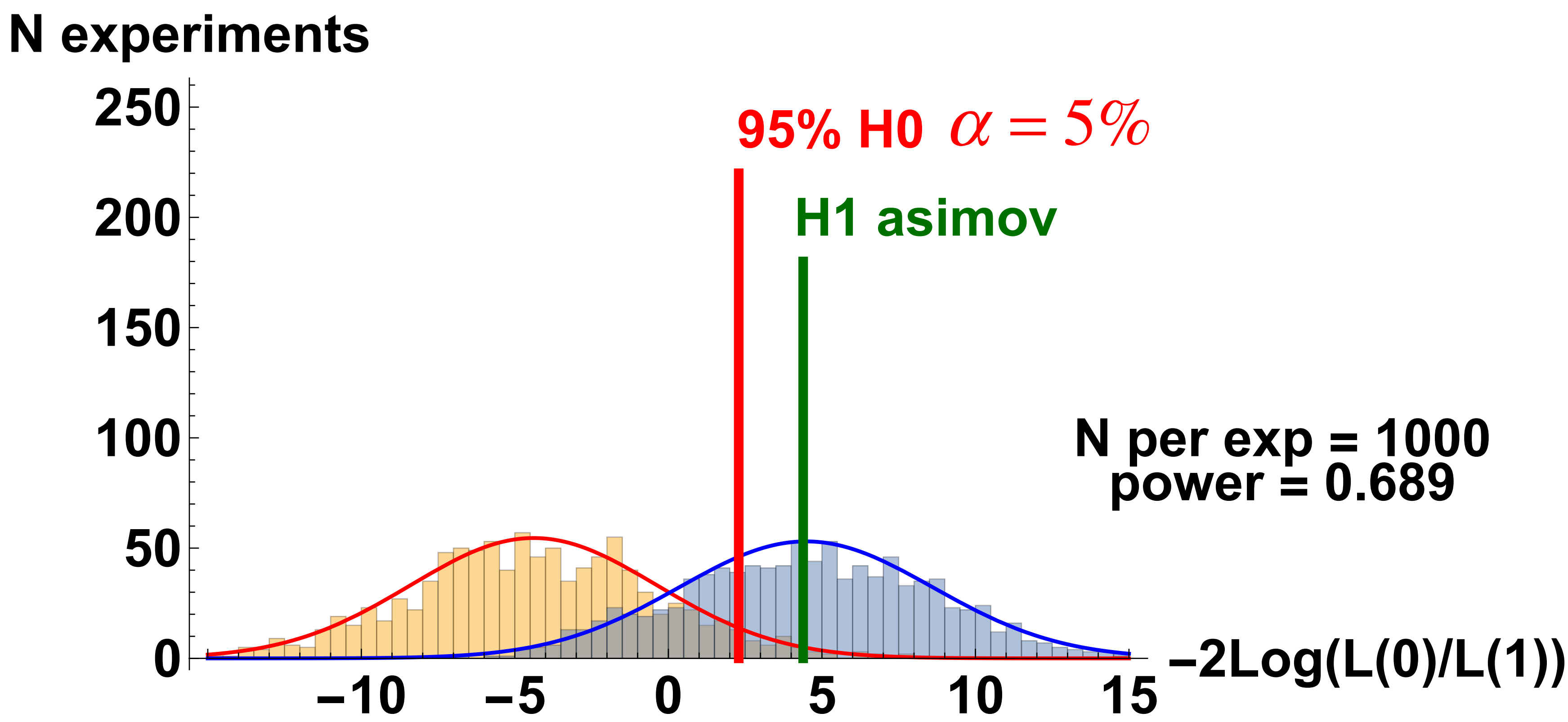


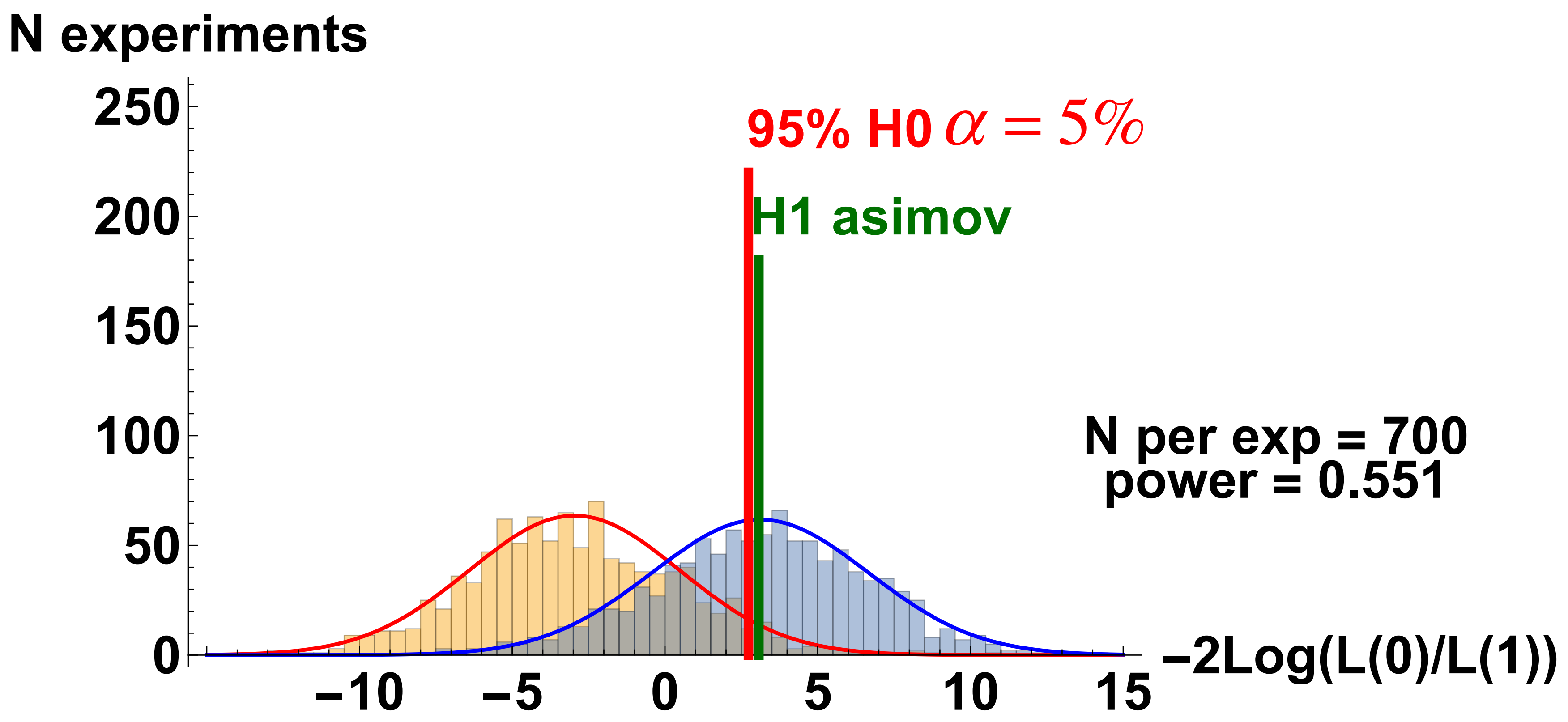
Power and Luminosity

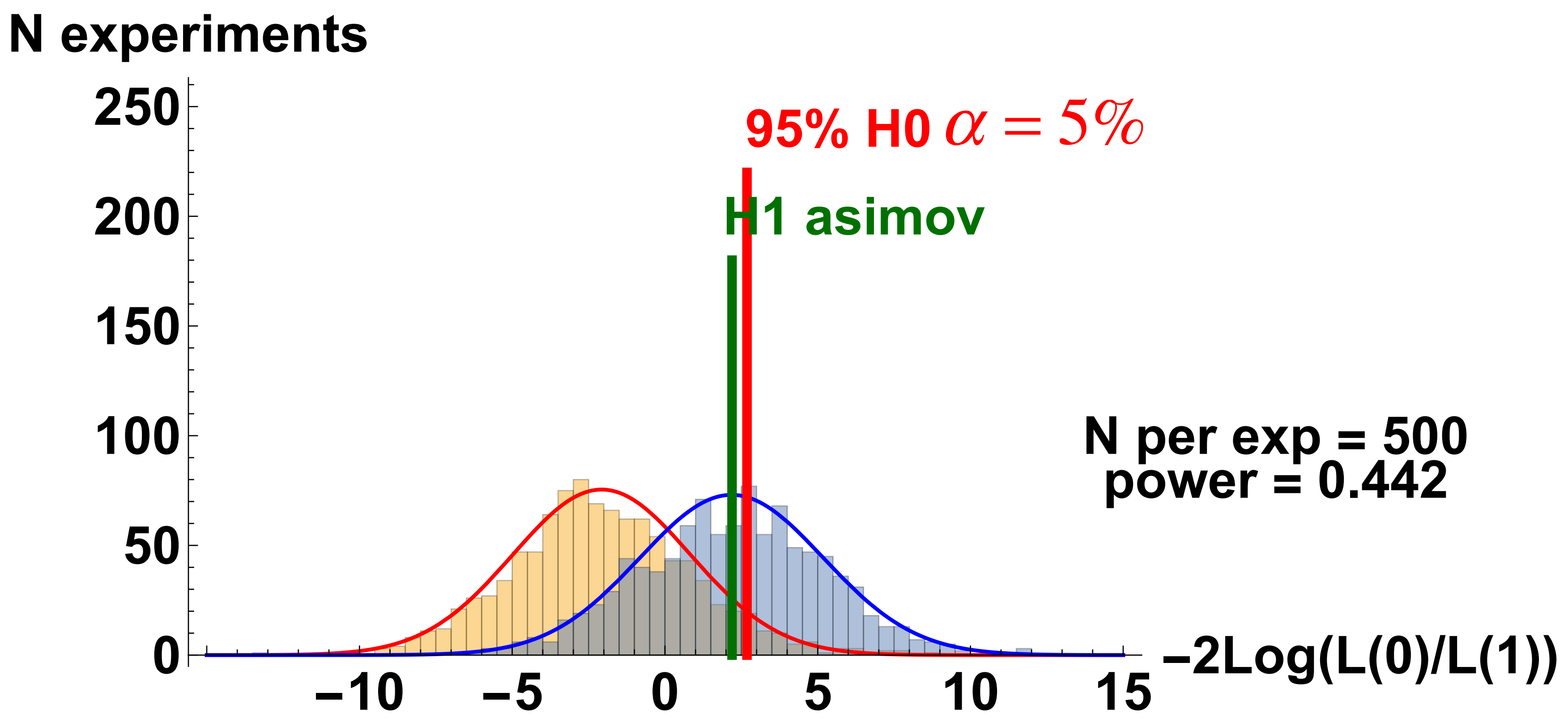
For a given significance the power increases with increased luminosity

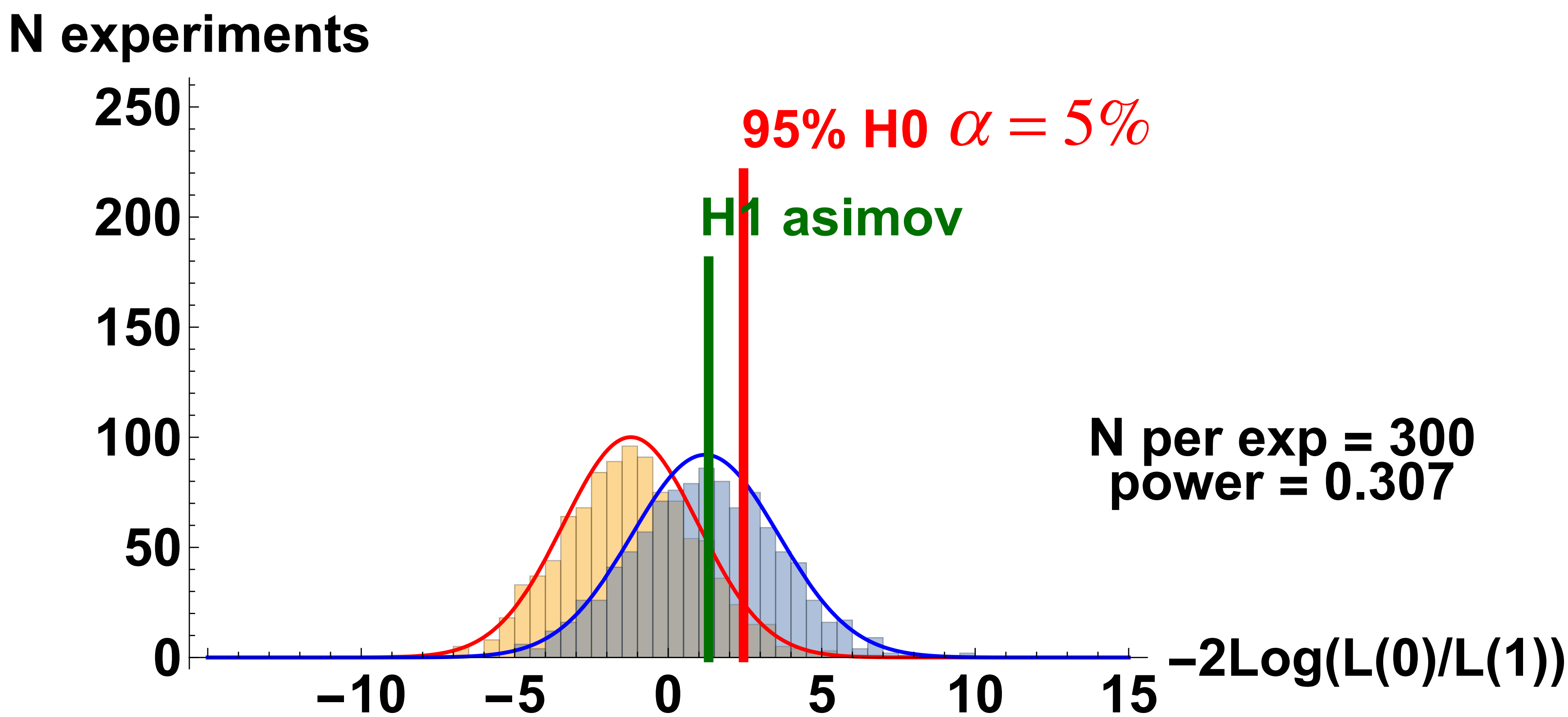
Luminosity ~ Total number of events in an experiment



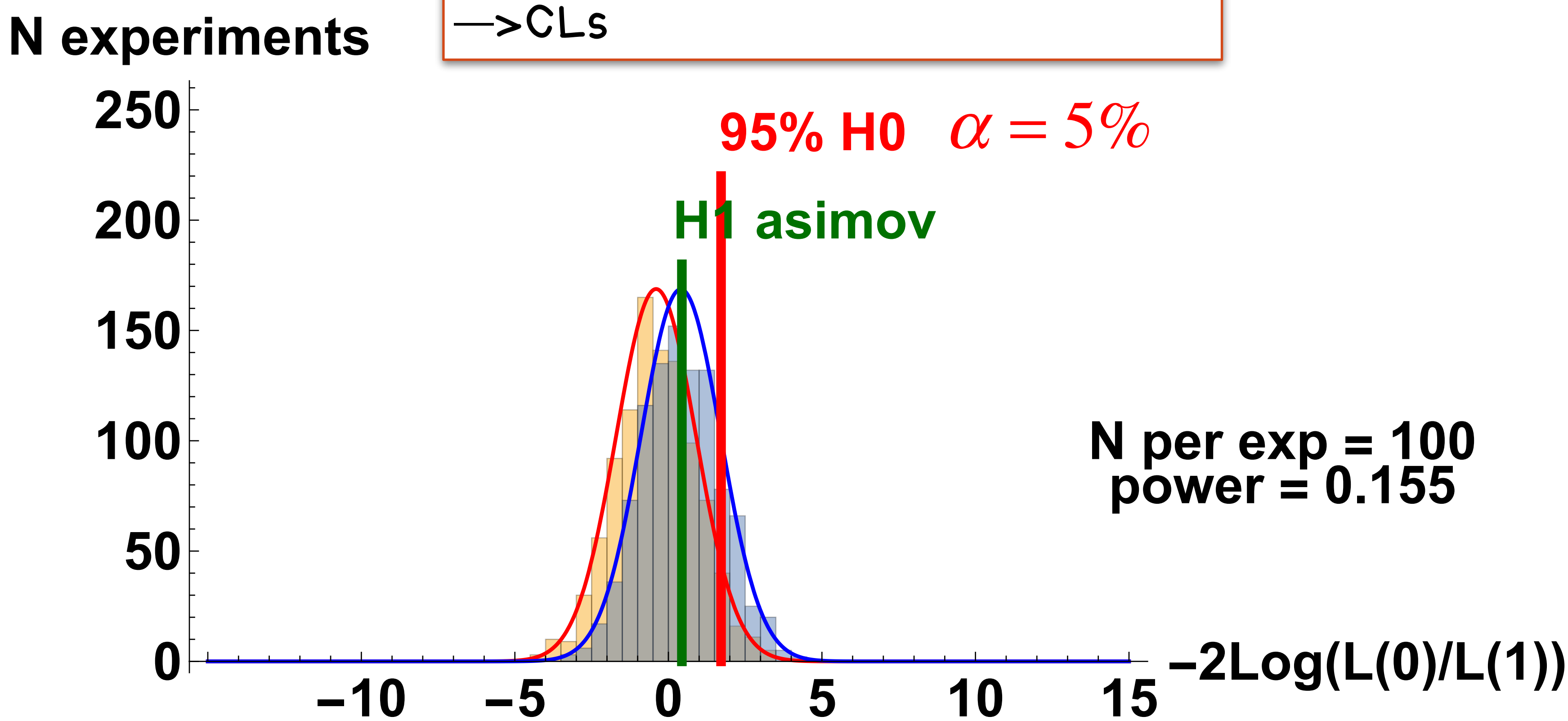

N experiments
250
200
150
100
50
0
95% H0 $\alpha = 5\%$
H1 asimov
N per exp = 1000
power = 0.689
−10
−5
0
5
10
15
−2Log(L(0)/L(1))


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))

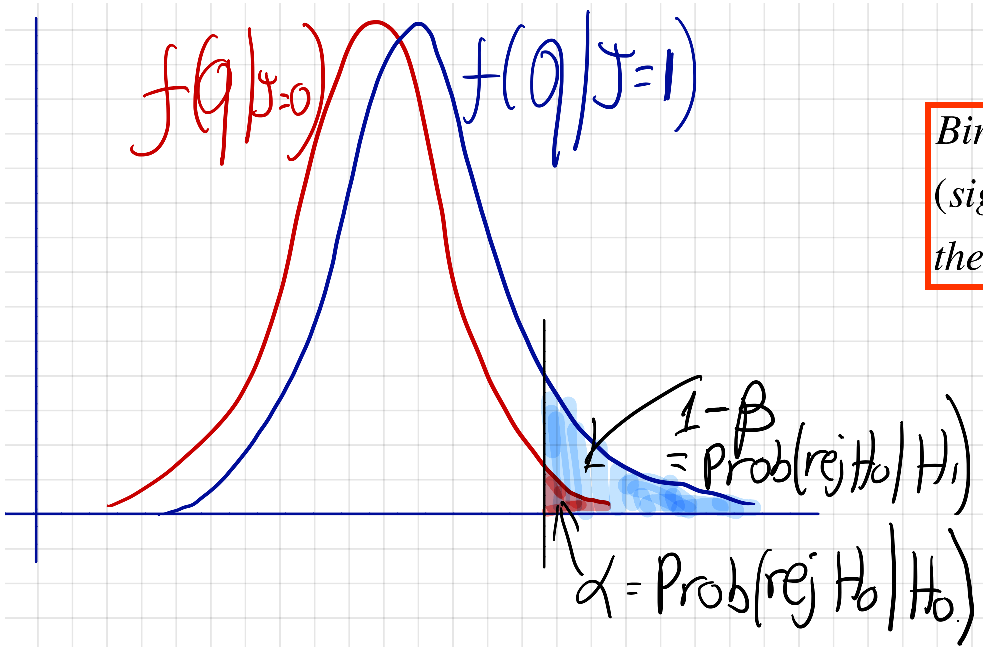
Hard to tell f(q|J=0) from f(q|J=1)
—>CLs



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."



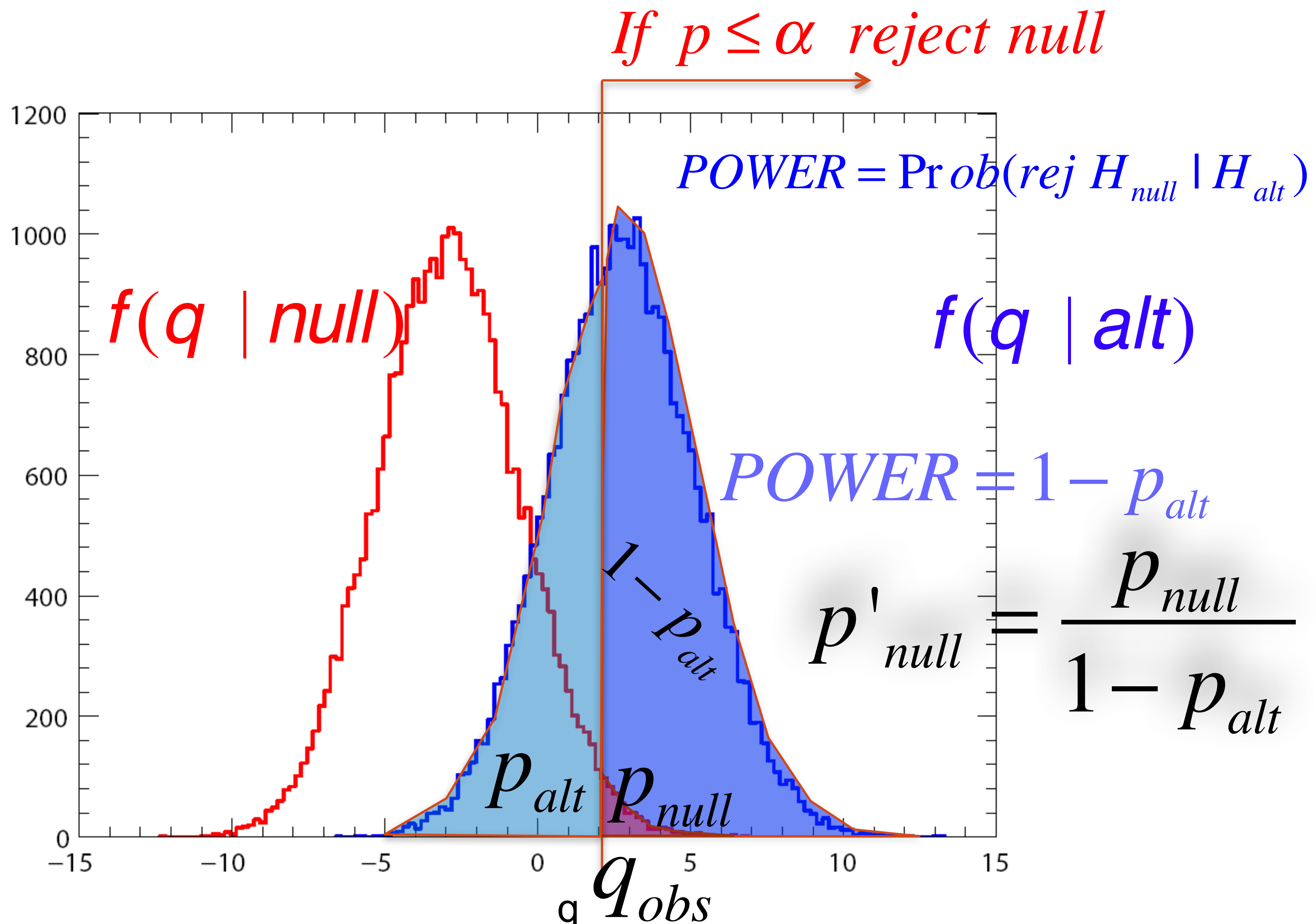
Birnbaum (1962) *suggested that* $\alpha / 1-\beta$ (*significance / power*) *should be used as a measure of the strength of a statistical test, rather than* α *alone*

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

$$p' \equiv CL_s$$

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

CLs

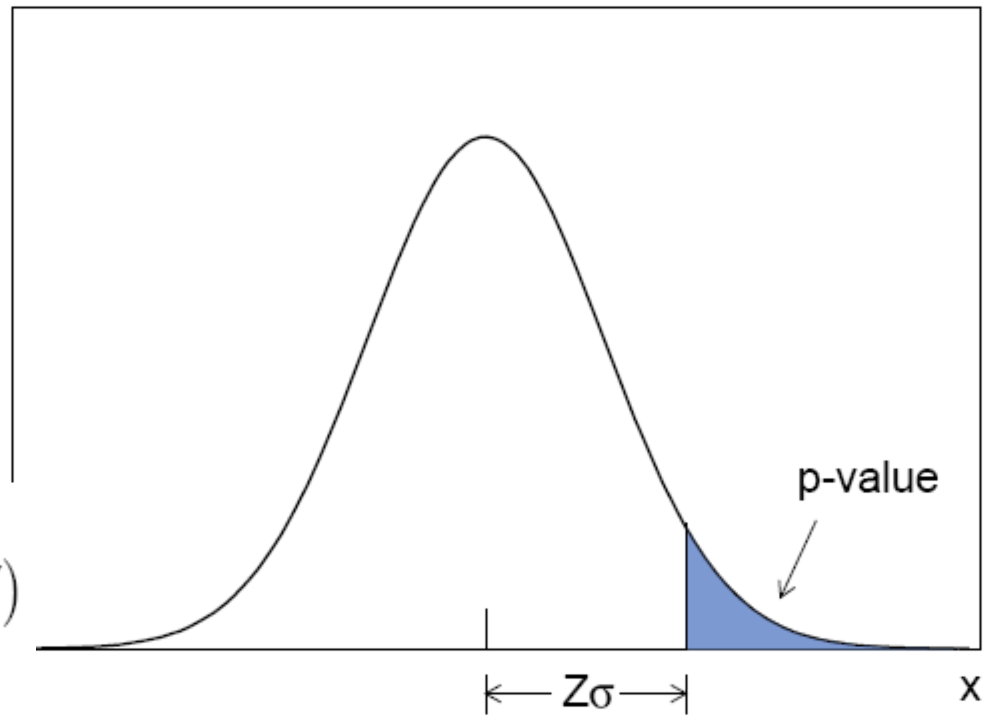


Null like

alt like

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



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

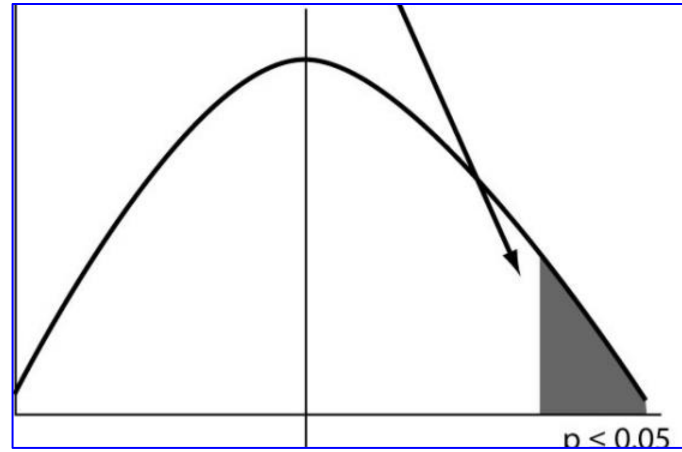
$$Z = \Phi^{-1}(1-p)$$

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

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.

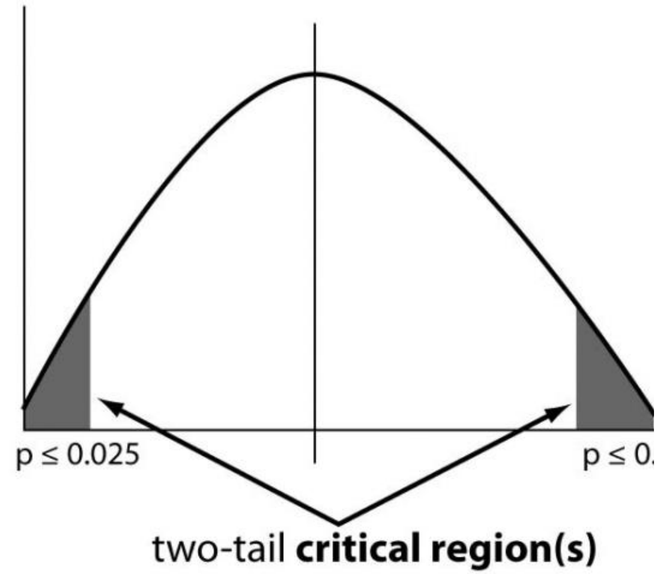


- Downward fluctuations of the background will not serve as an evidence against the background
- Upward fluctuations of the signal will not be considered as an evidence against the signal

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.

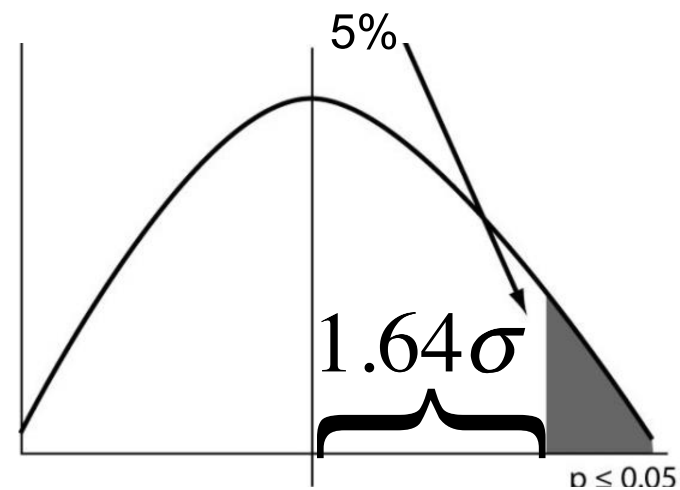
- Here we use a 2-sided p-value

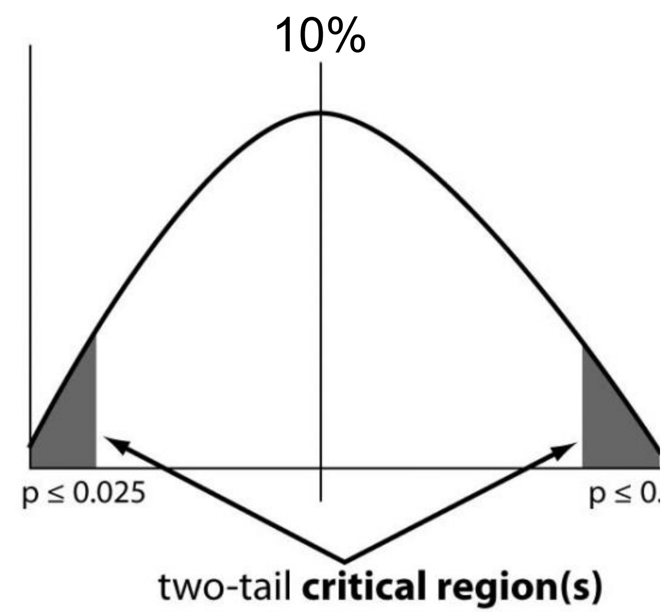


1-sided 2-sided

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)

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

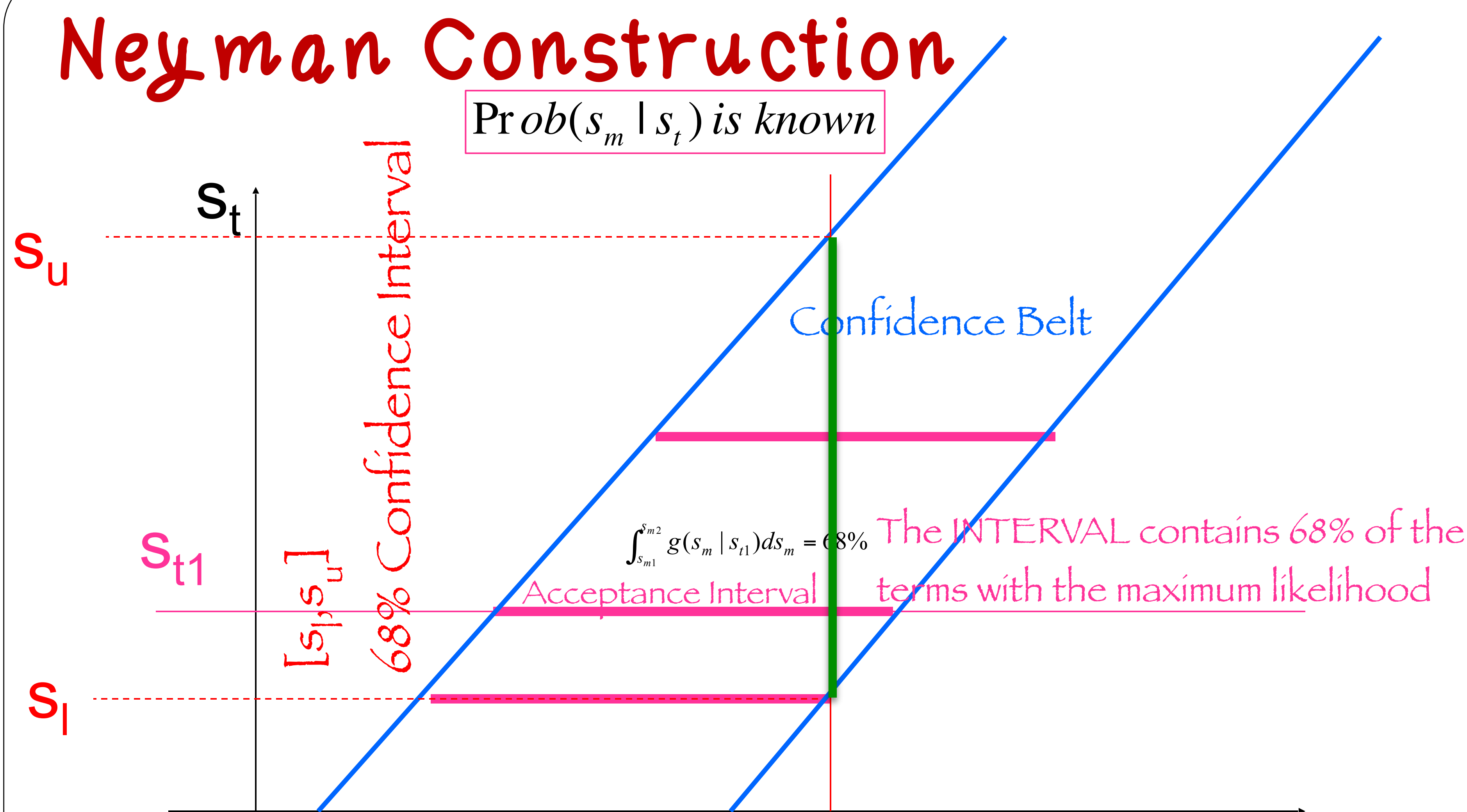
- Given the measurement you deduce somehow (based on your statistical method) that there is a 95% Confidence interval $[0,\mu_{up}]$.
- This means: our interval contains $\mu=0$ (no Higgs)
- We therefore deduce that $\mu<\mu_{up}$ at the 95% Confidence Level (CL)
- μ_{up} is therefore an upper limit on μ
- If $\mu_{up}<1 \rightarrow$
  $\sigma(m_H)<\sigma_{SM}(m_H)\rightarrow$
  a SM Higgs with a mass m_H is excluded at the 95% CL

The Frequentist Game a 'la Neyman

Or

How to ensure a Coverage with Neyman construction

Neyman Construction



$[s_l,s_u]$ 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.

The Feldman Cousins Unified Method

Frequentist Paradise – F&C Unified with Full Coverage

- Let the test statistics be

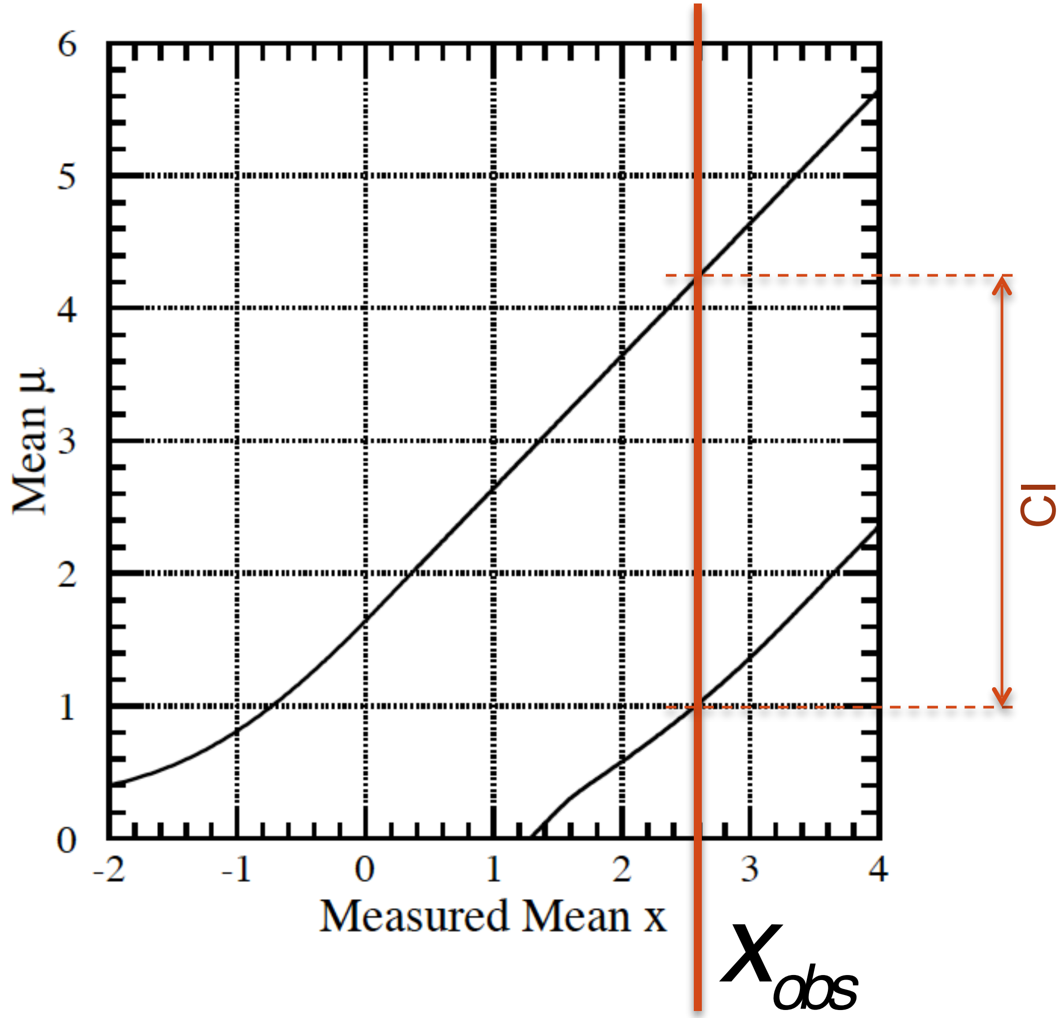
$$q=\begin{cases}-2\ln\dfrac{L(s+b)}{L(\hat{s}+b)} & \hat{s}\geq 0\\ -2\ln\dfrac{L(s+b)}{L(b)} & \hat{s}<0\end{cases}$$

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

- Order by taking the 68% highest q's

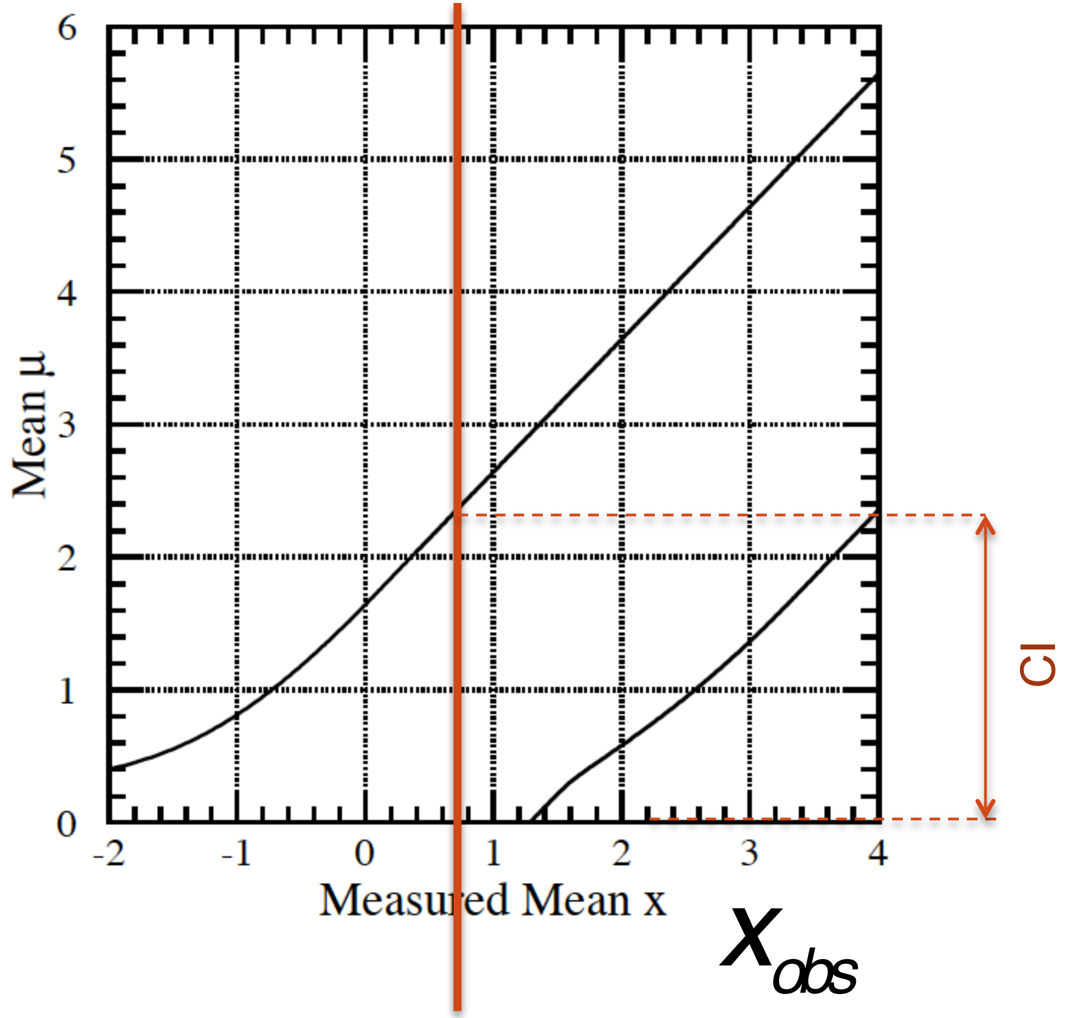
How to tell an Upper limit from a Measurement without Flip Flopping

- A measurement (2 sided)



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

- An upper limit (1 sided)



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)

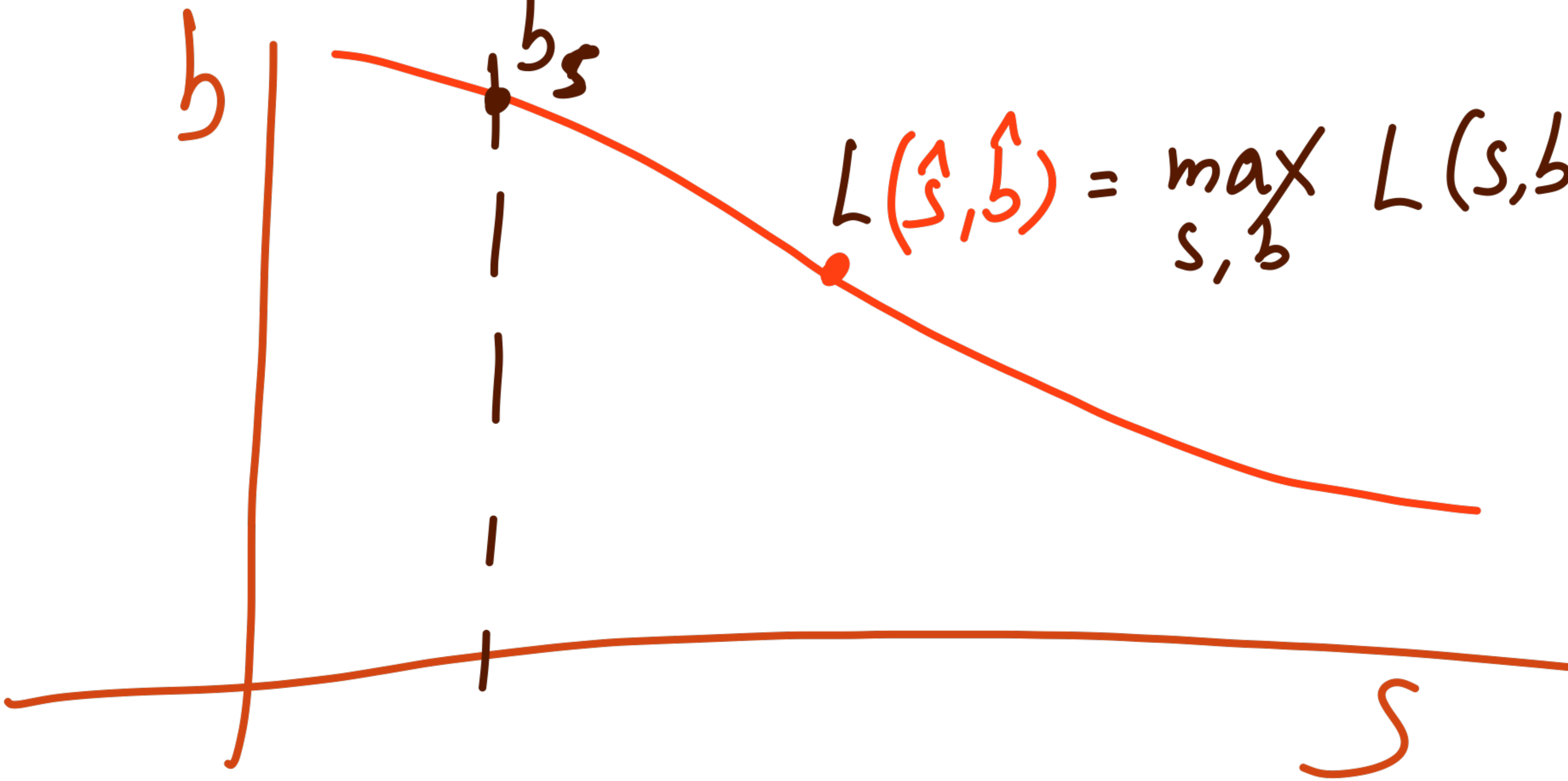
- Hybrid: One can also use a frequentist test statistics (PL) while treating the NPs via marginalization (Hybrid, Cousins & Highland way)

- Marginalization (Integrating))
  - Integrate the Likelihood, L, over possible values of nuisance parameters (weighted by their prior belief functions -- Gaussian,gamma, others...)
  - $$L(\mu) = \int L(\mu,\theta)\pi(\theta)d\theta$$

Profile Likelihood

$$q = \frac{L(s, \hat{\hat{b}})}{L(\hat{s}, \hat{b})}$$

$$L(s, \hat{\hat{b}}_s) = \max_{b} L(s,b)$$



$$L(\hat{s}, \hat{b}) = \max_{s,b} L(s,b)$$

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

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

$$\boxed{\hat{\hat{\theta}}_s \text{ is the MLE of } \theta \text{ fixing } s}$$

Pulls and Ranking of NPs

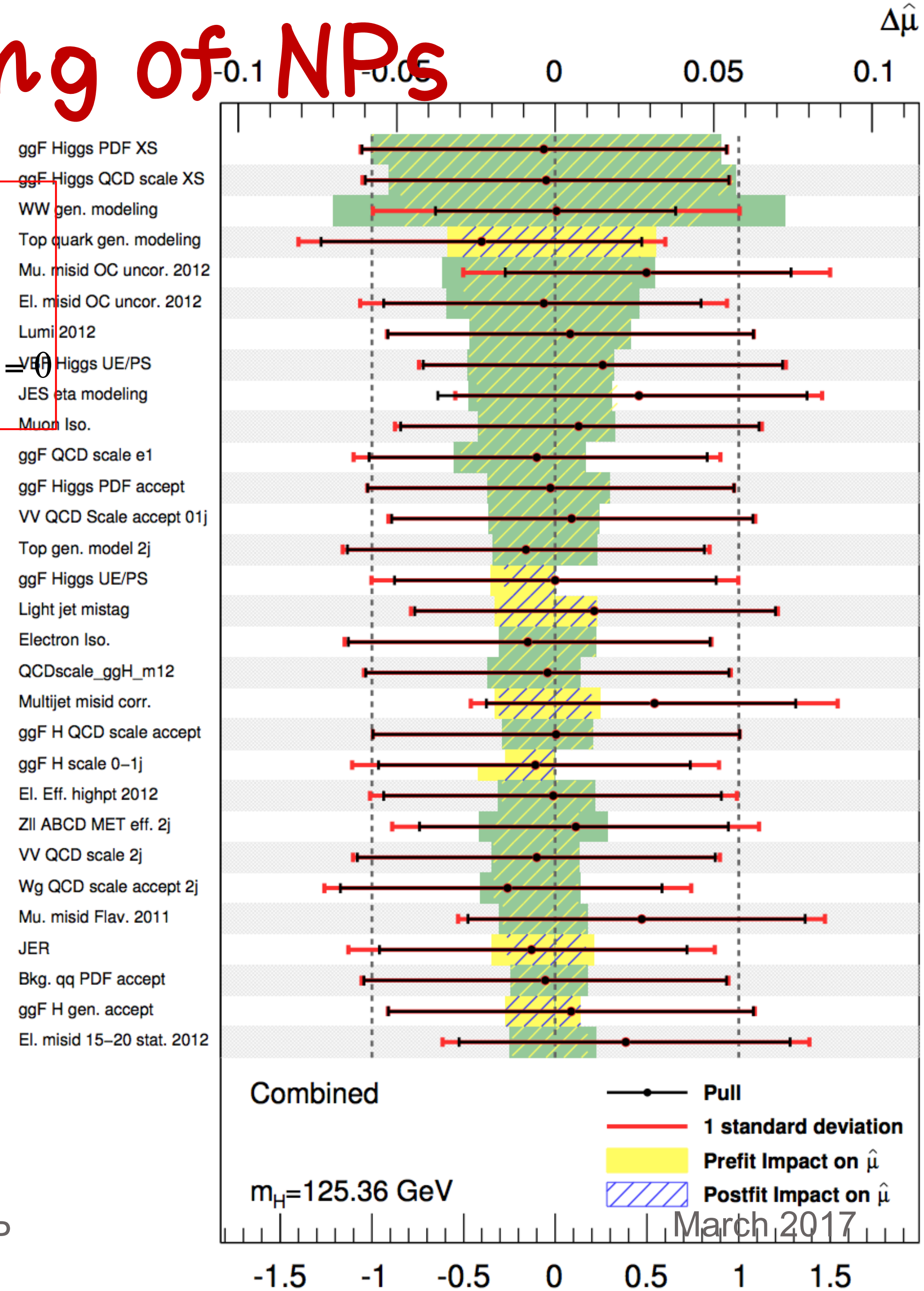
The pull of θ_i is given by $\dfrac{\hat{\theta}_i - \theta_{0,i}}{\sigma_0}$

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

Ranking θ_i by its effect on the p.o.i.

$$\Delta\mu^{\pm} = \hat{\hat{\mu}}_{\hat{\theta}_i \pm \sigma^{\pm}_{\theta_i}} - \hat{\mu}$$

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



Test Spin 0 parity

$$H_0 = 0^+$$

$$H_1 = 0^-$$

$$q^{NP} = -2\ln\frac{L(H_0)}{L(H_1)}$$

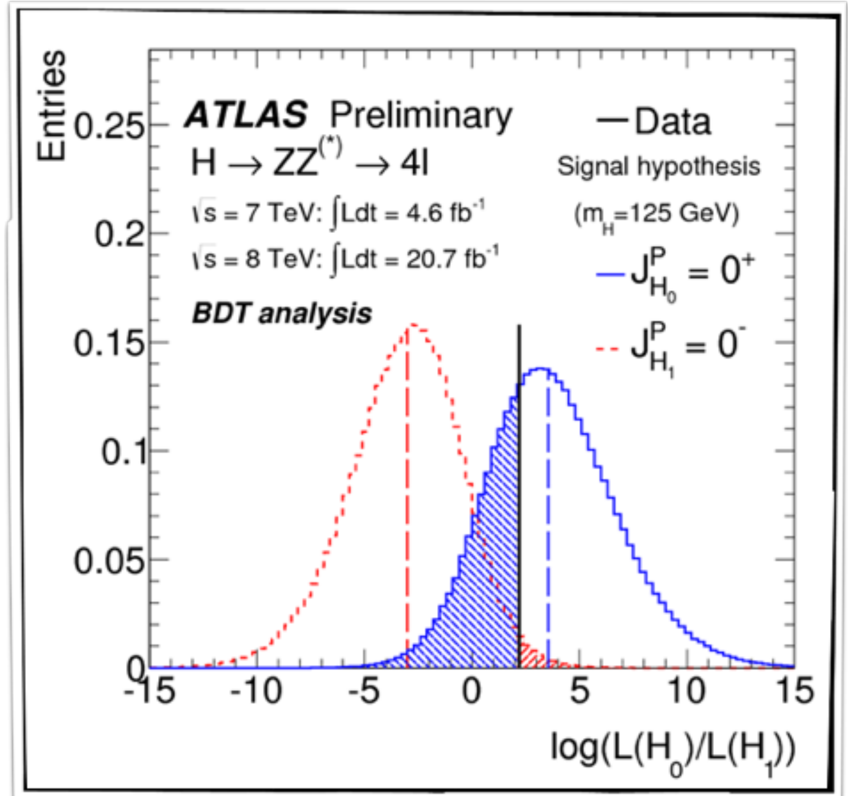
$$p_{H_1}(\exp \mid H_0) = 0.37\%,$$

$$p_{H_1}(obs) = 1.5\%$$

$$p_{H_0}(obs) = 31\%$$

$$p_{H_1}{}^{CL_s}(obs) = 2.2\%$$

$$p_{H_1}^{CL_s} = \frac{p_{H_1}}{1-p_{H_0}} = \frac{1.5\%}{1-0.31} = 2.2\%$$



Which means
$J^p=0^-$ is excluded at the
97.8% CL in favour of $J^p=0^+$

H_1 like

H_0 like