ISAPP 2017

Statistical methods for neutrino physics (2/2)

Alessandra Tonazzo
Université Paris-Diderot, Laboratoire APC
tonazzo@in2p3.fr

Yesterday’s hands-on session…

Quick Links: ROOT Homepage | Class Index | Class Hierarchy
Source: header file | source file | viewVC header | viewVC source
Sections: class description | function members | data members | class charts

Search documentation... Search

ROOT » ROOFIT » ROOSTATS » RooStats::ConfidenceBelt

class RooStats::ConfidenceBelt: public TNamed

ConfidenceBelt is a concrete implementation of the ConfInterval interface. It implements simple general purpose interval of arbitrary dimensions and shape. It does not assume the interval is connected. It uses either a RooDataSet (eg. a list of parameter points in the interval) or a RooDataHist (eg. a Histogram-like object for small regions of the parameter space) to store the interval.

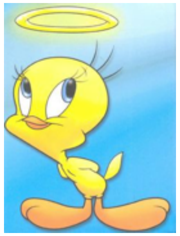
ROOT » ROOFIT » ROOSTATS » RooStats::FeldmanCousins

class RooStats::FeldmanCousins: public RooStats::IntervalCalculator

The FeldmanCousins class (like the Feldman-Cousins technique) is essentially a specific configuration of the more general NeymanConstruction. It is a concrete implementation of the IntervalCalculator interface that, which uses the NeymanConstruction in a particular way. As the name suggests, it returns a ConfidenceInterval. In particular, it produces a RooStats::PointSetInterval, which is a concrete implementation of the ConfInterval interface.

The Neyman Construction is not a uniquely defined statistical technique, it requires that one specify an ordering rule or ordering principle, which is usually incoded by choosing a specific test statistic and limits of integration (corresponding to upper/lower/central limits). As a result, this class must be configured with the corresponding information before it can produce an interval.

In the case of the Feldman-Cousins approach, the ordering principle is the likelihood ratio -- motivated by the Neyman-Pearson lemma. When nuisance parameters are involved, the profile likelihood ratio is the natural generalization. One may either choose to perform the construction over the full space of the nuisance parameters, or restrict the nusiance parameters to their conditional MLE (eg. profiled values).



SUMMARY

- Parameter estimation
  - The Maximum Likelihood Estimator: construction and properties
  - The Least Squares estimator: construction and properties

- Estimation of confidence intervals
  - general case 1D: Neyman belt construction
  - the Feldman-Cousins approach
    - hands-on: the Feldman-Cousins approach

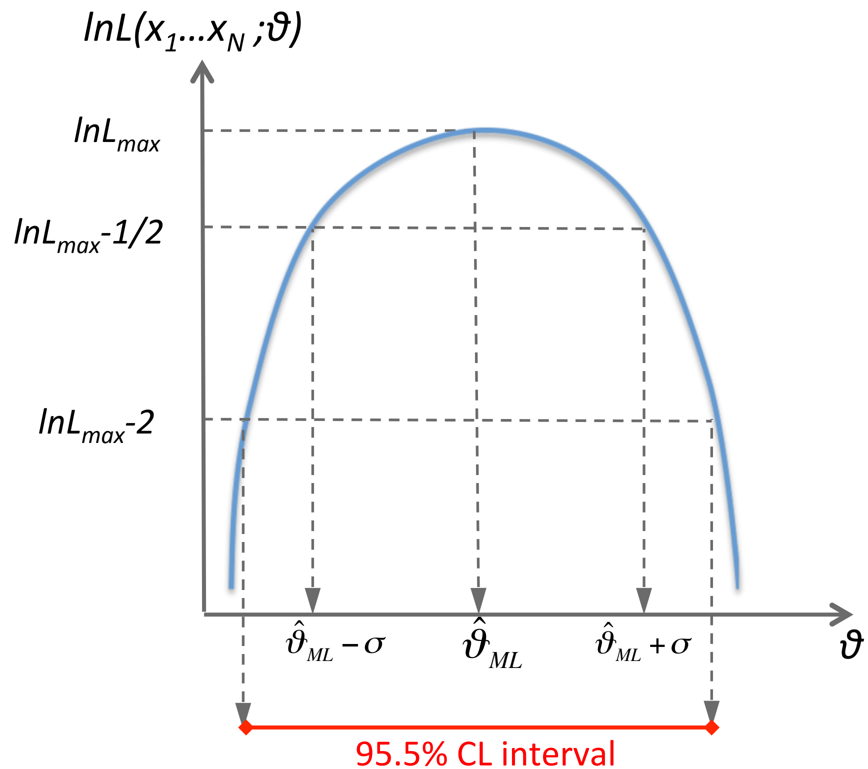
  - use of the Likelihood function
  - case of multiple parameters
  - Bayesian credibility intervals

Interval estimation : use of the Likelihood

Use asymptotic normality of the Likelihood function:

P[$(x-\mu)^2<\sigma$] = 68.3% =>
P[$x-\sigma<\mu<x+\sigma$] = 68.3%

P[$(x-\mu)^2<2\sigma$] = 95.5% =>
P[$x-2\sigma<\mu<x+2\sigma$] = 95.5%



Interval estimation : use of the Likelihood

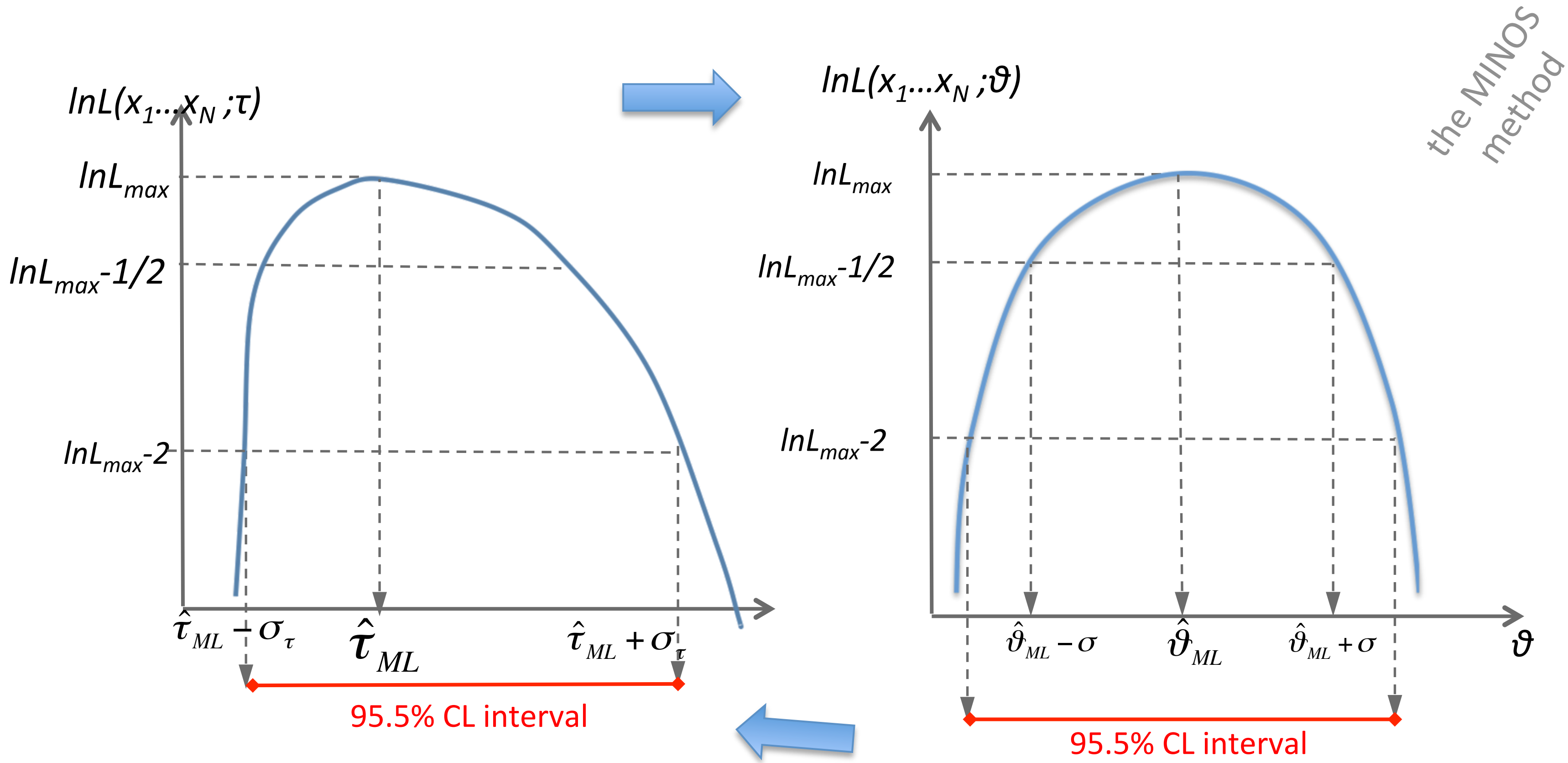
Case of non-Gaussian Likelihood function (non-parabolic *lnL*):

$\Rightarrow$ the <u>invariance</u> property allows to find a transformation that makes the L Gaussian

$\Rightarrow$ the contents of the interval are preserved

$\Rightarrow$ can determine the confidence intervals without actually making the transformation to gaussian

BUT <u>the Confindence Intervals are only approximate</u> for N finite ! USE WITH CAUTION



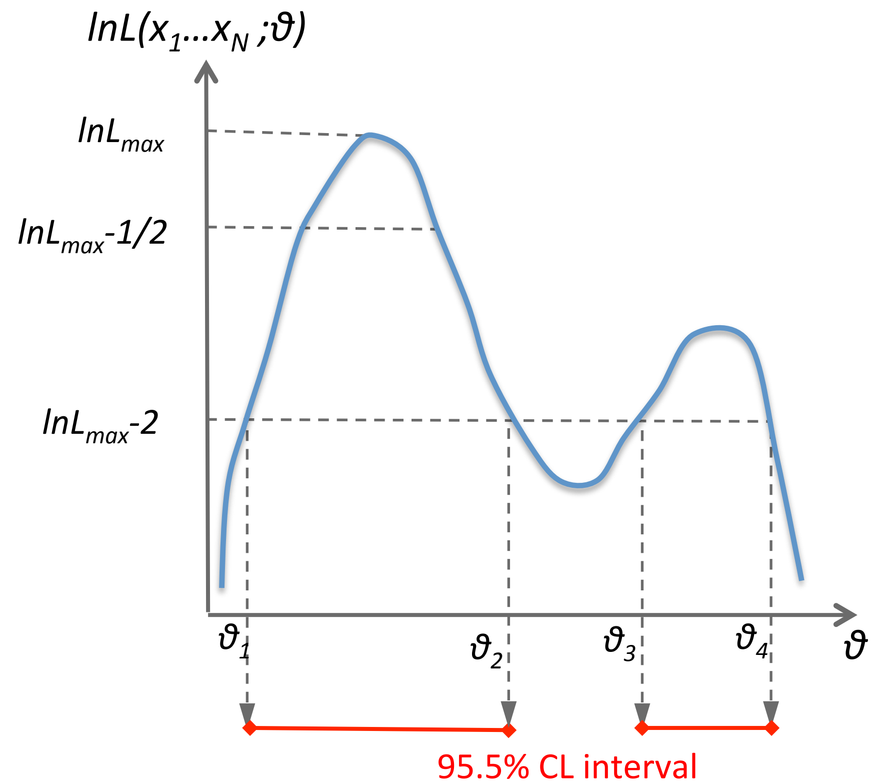
Interval estimation : use of the Likelihood

Case of ill-shaped Likelihood function with multiple maxima
$\Rightarrow$ multi-interval CL

$[\theta_1<\theta<\theta_2] \cup [\theta_3<\theta<\theta_4]$

the interval is approximate

better to show the whole L curve



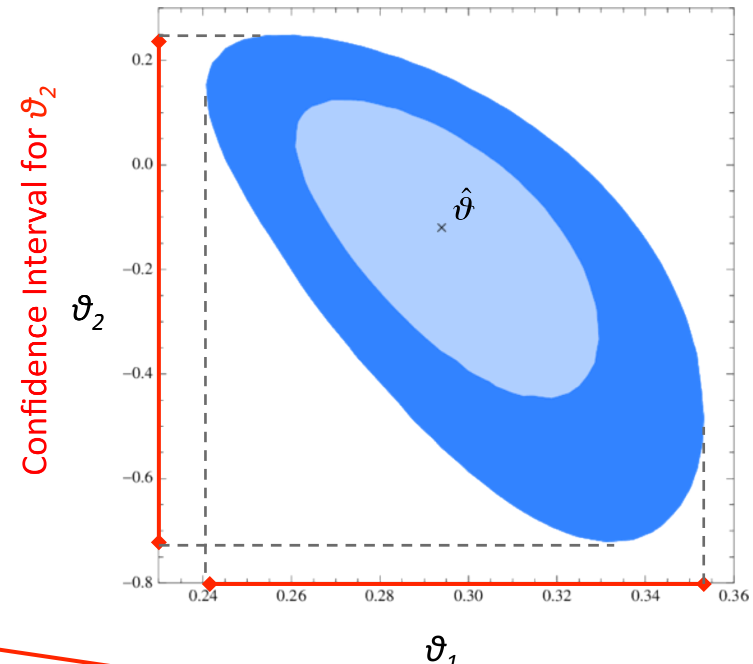
Interval estimation: multi-parameter case

$\hat{\vartheta} = (\hat{\vartheta}_1, \hat{\vartheta}_2)$
simultaneous MLE of
the two parameters

k parameters: the C.I. with probability content *β* is given by the *k*-dimensional confidence regions given by the hyper-surface defined by

$$\ln L(X \mid \vec{\vartheta}) = \ln L_{\max} - \frac{1}{2}\chi^2_\beta(k)$$

(asymptotic χ^2(k) distribution of lnL)



Confidence Interval for ϑ_1

Probabiility content different from that of the 2D-contour !

"Tilt" of the shape (~ellipse) indicate the correlation (negative in this case).
Larger correlation results in larger C.I.

Interval estimation: multi-parameter case

Example: 2D Normal distribution

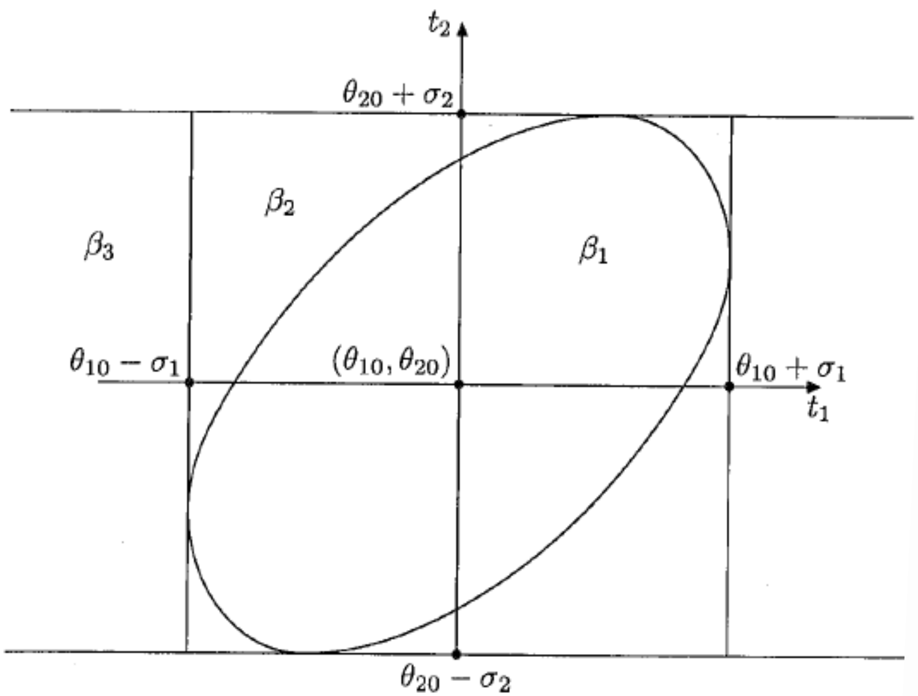


Table 9.1. Probability content of different regions in two variables.

	$K_\beta = 1$	$K_\beta = 2$	$K_\beta = 3$
inner ellipse β_1	0.393	0.865	0.989
square β_2 for $\rho = 0.00$	0.466	0.911	0.995
for $\rho = 0.50$	0.498	0.917	0.995
for $\rho = 0.80$	0.561	0.929	0.996
for $\rho = 0.90$	0.596	0.936	0.996
for $\rho = 0.95$	0.622	0.941	0.996
for $\rho = 1.00$	0.683	0.954	0.997
infinite band β_3	0.683	0.954	0.997

$$\beta_1 = P\left(\vartheta_1 \text{ and } \vartheta_2 \text{ in the ellipse}\right)$$

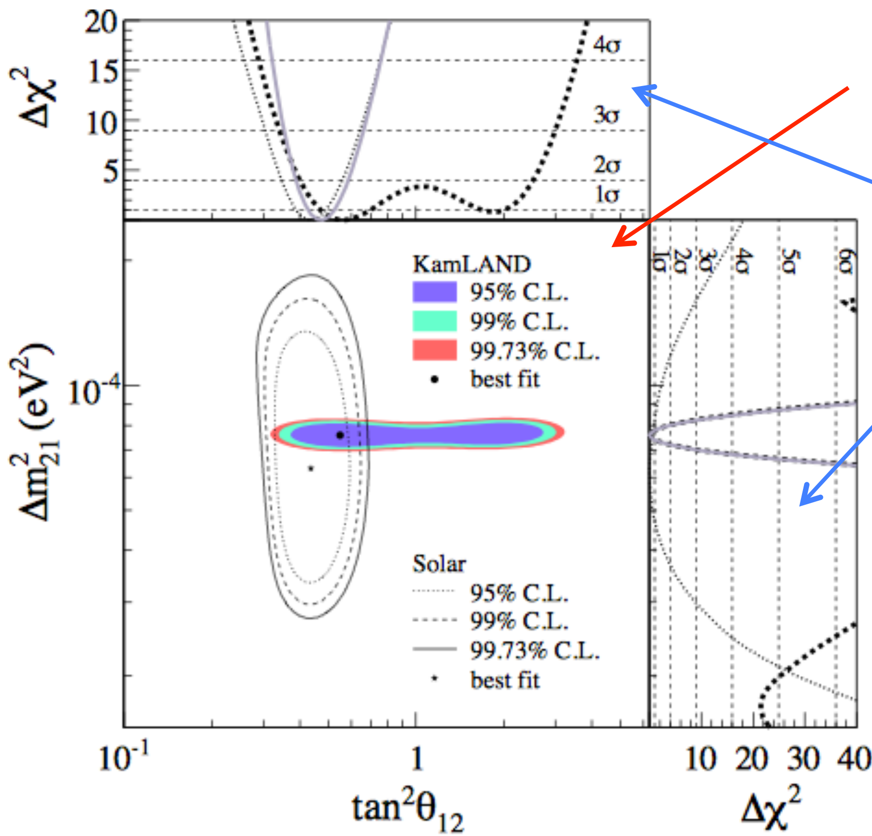
$$\beta_2 = P\left(\left(\vartheta_{10}-\sigma_1 < \vartheta_1 < \vartheta_{10}+\sigma_1\right) \text{ and } \left(\vartheta_{20}-\sigma_2 < \vartheta_2 < \vartheta_{20}+\sigma_2\right)\right)$$

$$\beta_3 = P\left(\vartheta_{20}-\sigma_2 < \vartheta_2 < \vartheta_{20}+\sigma_2\right)$$

Interval estimation: multi-parameter case

An example from neutrino physics [KamLAND, Phys.Rev.Lett. 100 (2008) 221803]

Δχ² ~ -2lnL



Multi-parameter

Profile

Note: different confidence intervals/regions in the two cases !

FIG. 2: Allowed region for neutrino oscillation parameters from KamLAND and solar neutrino experiments. The side-panels show the $\Delta\chi^2$-profiles for KamLAND (dashed) and solar experiments (dotted) individually, as well as the combination of the two (solid).

Bayesian credibility intervals

Base on the probability contents of intervals for the posterior distribution *$\pi(\vartheta|X)$* (distribution of degrees of belief), based on the prior distribution *$\pi(\vartheta)$*

$$\pi\left(\vartheta \mid \vec{X}\right) = \frac{\prod_{i=1}^{N} f\left(X_i,\vartheta\right)\pi\left(\vartheta\right)}{\int \prod_{i=1}^{N} f\left(X_i,\vartheta\right)\pi\left(\vartheta\right)d\vartheta}$$

Define an interval $[\vartheta'_L,\vartheta'_U]$ such that $\int_{\vartheta'_L}^{\vartheta'_U} \pi\left(\vartheta \mid \vec{X}\right)d\vartheta = \beta$

$\Rightarrow$ degree-of-belief $(\vartheta'_L < \vartheta < \vartheta'_U) = \beta$

$\Rightarrow$ $[\vartheta'_L,\vartheta'_U]$ is the credibility interval with probability content β for the parameter ϑ

Often coincides with the frequentist interval, as a uniform prior is used… but not always

SUMMARY of TOPIC n.1

- Parameter estimation
  - The Maximum Likelihood Estimator: construction and properties
  - The Least Squares estimator: construction and properties

- Estimation of confidence intervals
  - general case 1D: Neyman belt construction
  - the Feldman-Cousins approach
    - hands-on: the Feldman-Cousins approach

  - use of the Likelihood function
  - case of multiple parameters
  - Bayesian credibility intervals

HYPOTHESIS TESTING

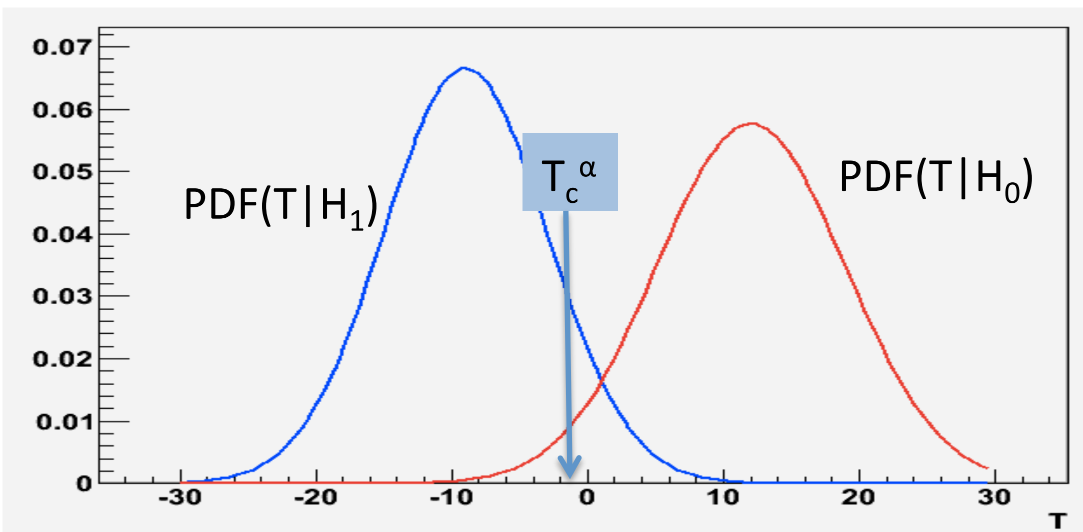
Test of simple hypotheses

Aim: test a "Null Hypothesis" H_0 against an "alternative hypothesis" H_1.

Both hypotheses are "simple" = completely specified

Method ("frequentist"):

1) define a "test statistic" T, function of the data
2) construct the PDF of T under under each hypothesis
3) define a "critical region" Ω_C such that T in Ω_C suggests H_0 is true



"critical region" $T>T_c^\alpha$

4) Evaluate the probability to give the wrong answer

Test of simple hypotheses

4) Evaluate the probability to give the wrong answer,

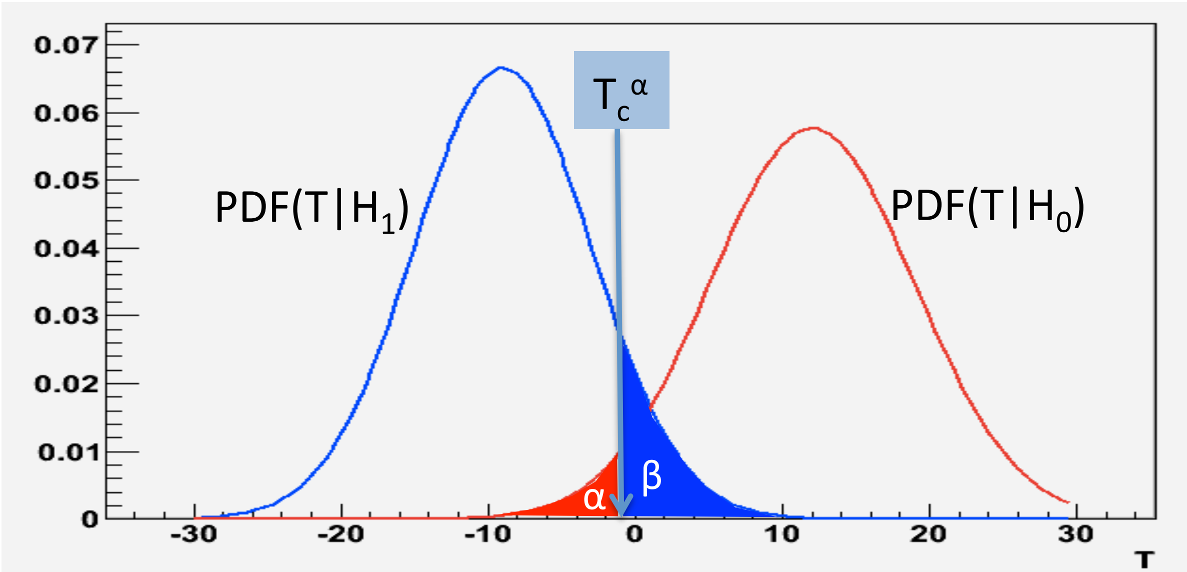
which you can do in two ways:

- reject H_0 when it's true : Error of type I or "loss": α
- accept H_0 when H_1 is true: Error of type II or "contamination": β

<u>Definitions</u>: Confidence Level CL = 1-α $= \int_{T_C}^{+\infty} PDF(T \mid H_0)\, dT$

Power p= 1-β $= \int_{-\infty}^{T_C} PDF(T \mid H_1)\, dT$

Note: "3σ" means $\int_{T_C}^{+\infty} PDF(T \mid H_0)\, dT = 99.73\%$ if 1-sided integrals are used



"critical region" $T > T_c^\alpha$

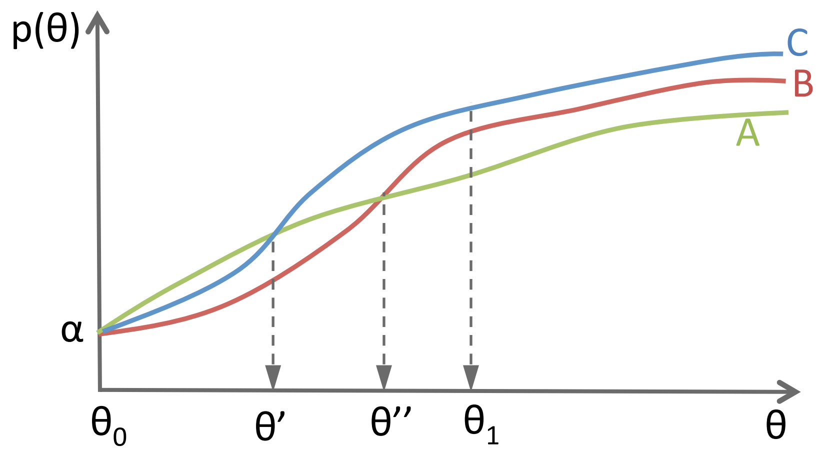
reject H_0 at CL (1-α) if $T < T_C^\alpha$

Comparison of tests

More general case: "composite hypotheses", with dependence on a parameter θ

H_0 is $\theta=\theta_0$, H_1 is $\theta=\theta_1$

Define the <u>power function</u> as $p(\theta)=1-\beta(\theta)$. By construction, $p(\theta_0)=1-\beta(\theta_0)=\alpha$



Test B is better (more powerful) than test A for $\theta>\theta''$

Test C is the best (most powerful) of the three for $\theta>\theta'$ (in particular for $\theta=\theta_1$)

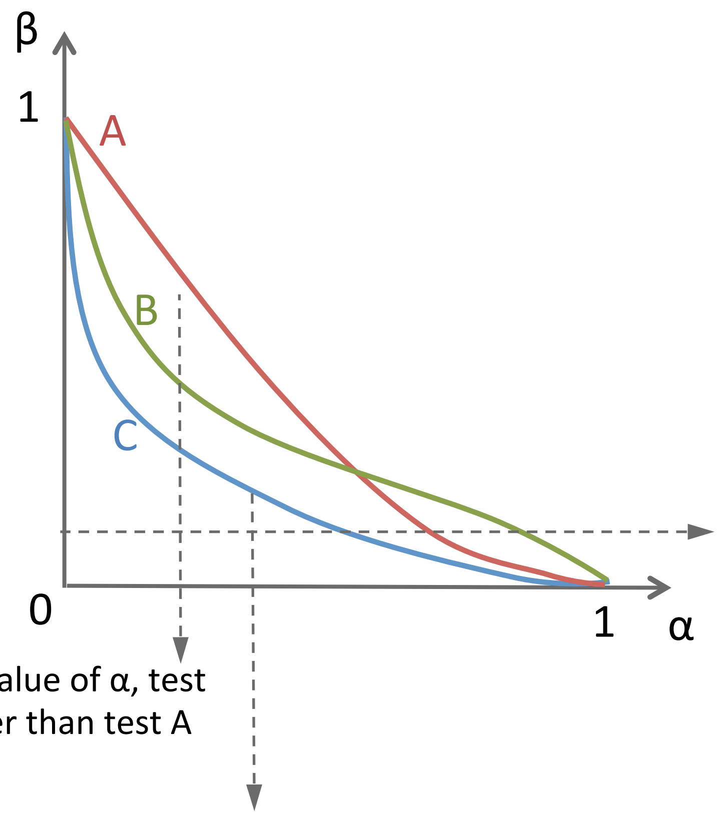
A test at least as powerful as any other at a given θ is "the most powerful test" for that θ

A test which is most powerful for all values of θ is <u>Uniformly Most Powerful (UMP)</u>

Important criteria are also: robustness, consistency, unbiasedness

Comparison of tests

α-β symmetric comparison :



for this value of β, test A is better than test B

for this value of α, test B is better than test A

Test C is “better” than A and B for any value of α and for any value of β

For simple hypotheses, a UMP test exists: the Neyman-Pearson test

The Neyman-Pearson test

finding the most powerful test ⇔ finding the best critical region

Given a signficance α, find the region ω_α maximising P=1-β

$$1-\beta = \int_{\omega_\alpha} PDF(X \mid H_1)\,dX = \int_{\omega_\alpha} \frac{PDF(X \mid H_1)}{PDF(X \mid H_0)} PDF(X \mid H_0)\,dX = E_{\omega_\alpha}\left[\left.\frac{PDF(X \mid H_1)}{PDF(X \mid H_0)}\right|_{H_0} \right]$$

this will be maximal only if ω_α contains the largest values of the ratio of PDFs

$$\boxed{l(X \mid H_0, H_1) = \frac{PDF(X \mid H_1)}{PDF(X \mid H_0)} > c_\alpha}$$

⇒ if $l(X|H_0.H_1)>c_\alpha$, choose H_1
⇒ if $l(X|H_0.H_1)<c_\alpha$, choose H_0

$l(X|H_0.H_1)$ is the ratio of likelihoods for the two hypotheses.
It is the best test if H_0 and H_1 are completely specified

HYPOTHESIS TESTING
The case of neutrino Mass Hierarchy in future experiments

Some references :

[Quian] X.Qian et al., Phys.Rev. D86 (2012) 113011, arXiv:1210.3651
[CEZ] E.Ciuffoli, J.Evslin and X.Zhang, arXiv:1305.5150
[BCHS] M.Blennow, P.Coloma, P.Huber and T.Schwetz, arXiv:1311.1822
[Blen] M.Blennow, arXiv:1311.3183
[Ciuf] E.Ciuffoli, ArVix:1704.08043

The test statistic for MH

H_0=IH , H_1=NH

If the two hypotheses are “simple”, the uniformly most powerful test is provided by the Likelihood Ratio (Neyman-Pearson lemma)

- In general, the LR is a good choice
- For the MH problem, it is equivalent to a difference of χ^2’s

  (in the gaussian limit)

$$-\ln L_{NH} = -\ln\left(\prod_{i\in bins} L\left(n_i \mid \mu_i^{NH}\right)\right) \approx -\ln\left(\prod_{i\in bins} \exp\left[-\frac{\left(n_i-\mu_i^{NH}\right)^2}{2\mu_i^{NH}}\right]\right) = \sum_{i\in bins}\left[\frac{\left(n_i-\mu_i^{NH}\right)^2}{2\mu_i^{NH}}\right] = \chi^2_{NH}$$

idem for IH

$$\ln\frac{L_{NH}}{L_{IH}} = -\chi^2_{NH} + \chi^2_{IH}$$

- Definition of the test statistic for MH : **$T = \chi^2_{IH}-\chi^2_{NH}$** (“$\Delta\chi^2$”)

PDFs of the test statistic for MH

If the hypotheses are "nested" (eg: H_0 is {θ in ω, subset ofΩ} and H_1 is {θ in Ω}), Wilk's theorem states that

$PDF(T|H_0) = X^2(1dof)$ => "Nσ" CL is ensured at $T_c=N^2$

But, for the MH case, the hypotheses are NOT nested ! **H_0=IH , H_1=NH**

⇒ $PDF(T|H_0)$ is not $X^2(1dof)$ and the T_c values must be computed based on the correct distribution [CEZ,Qian]

- Ideally: get PDFs from toy-MC simulations
- A good approximation for the MH case [CEZ,Qian,Ciuf]:

$$\boxed{\text{Gaussians with } \sigma=2\sqrt{\mu}}$$

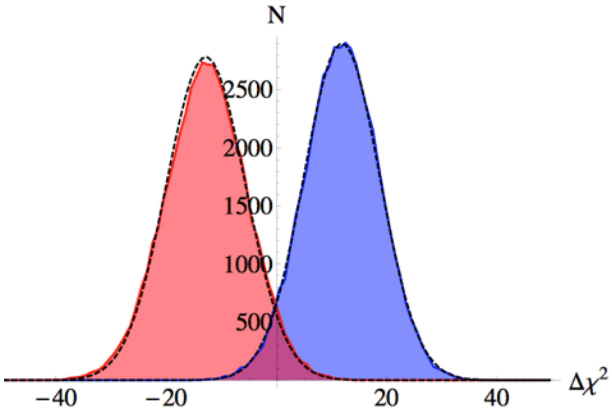
Only under certain conditions :

(1) expected number of events per bin is an approximately linear function of the parameters

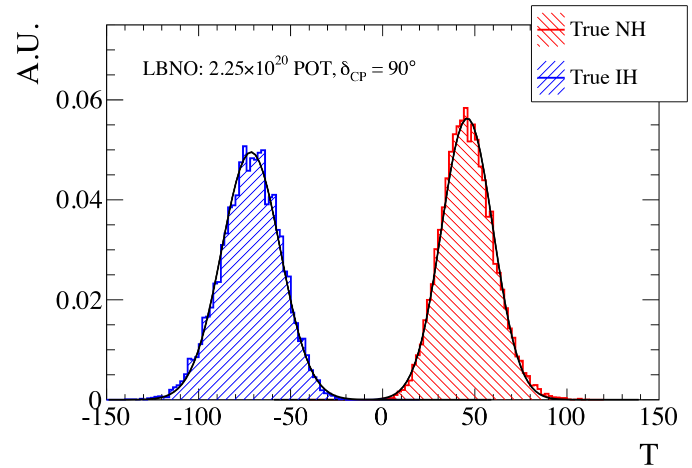
(2) the hyper-planes defined by (1) for the two hypotheses are parallel

PDFs of the test statistic for MH

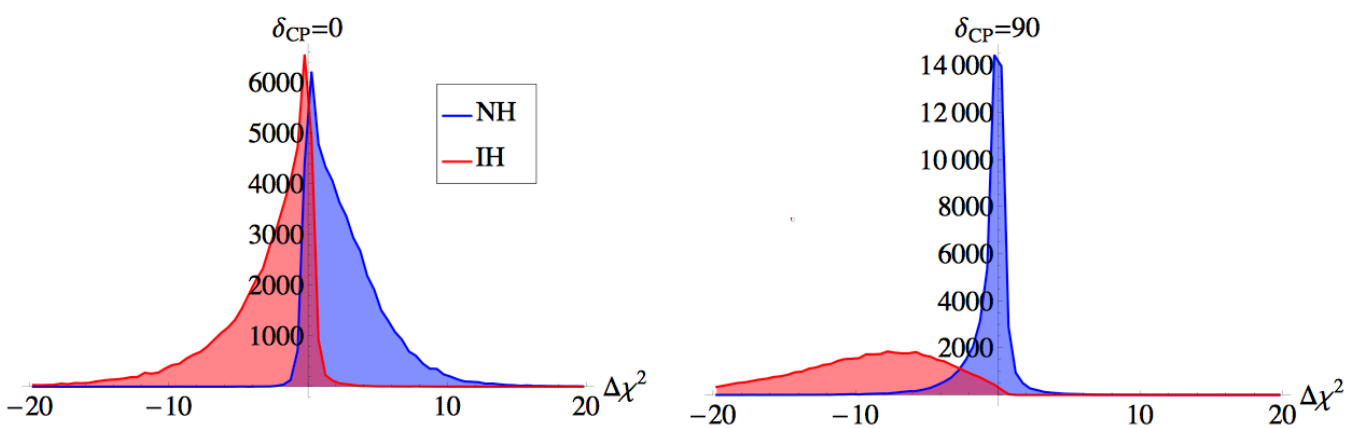
Case of reactor neutrino experiments



Case of Very Long BaseLine experiments (LBNO)



Case of accelerator experiments (NOvA)



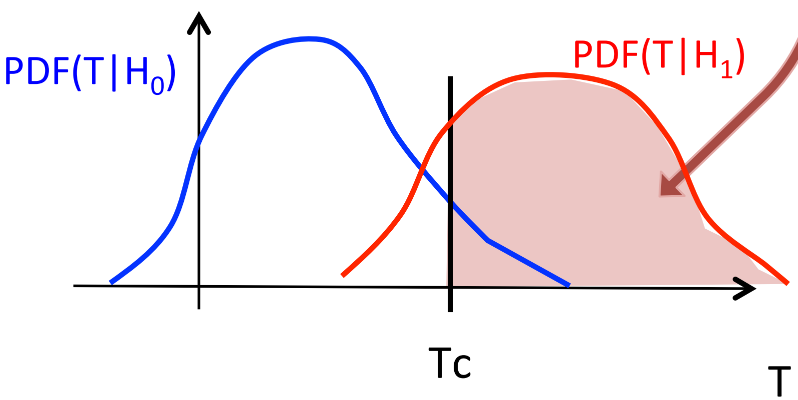
Quantifying the sensitivity: Frequentist approach

Common approaches for a future experiment

1) for a given CL, consider the possible fluctuations and quote Power

$$P = \int_{Tc}^{+\infty} PDF(T \mid H_1)\, dT$$

~ probability to do "at least as good"



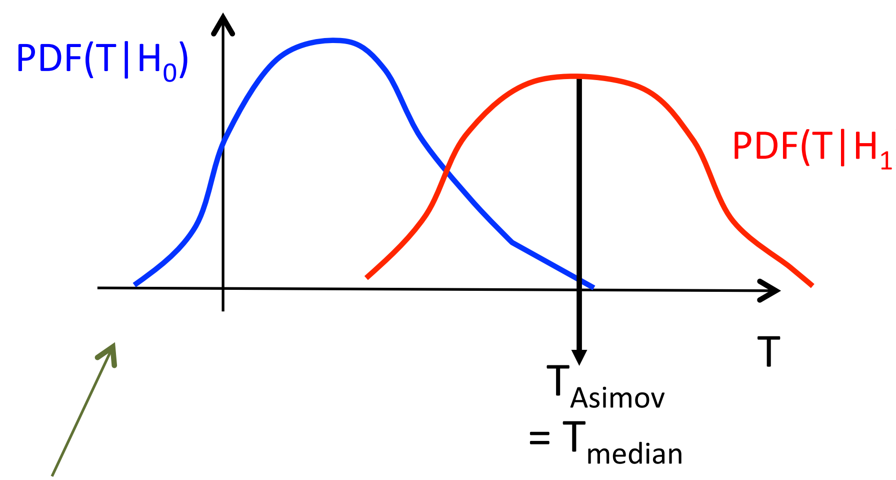
2) Quote the CL with a given outcome

2a) "typical"or "Asimov"* experiment:

$P(T|H_1)$=max

2b) "median" experiment: P=0.5

coinciding if PDF symmetric



More common

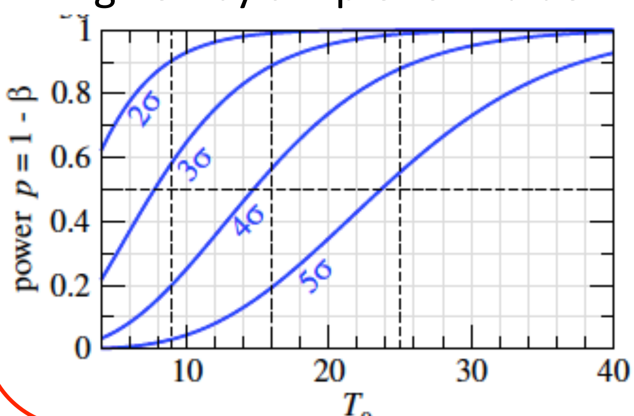
* «Franchise», a story by Isaac Asimov

Quantifying the sensitivity: Frequentist approach

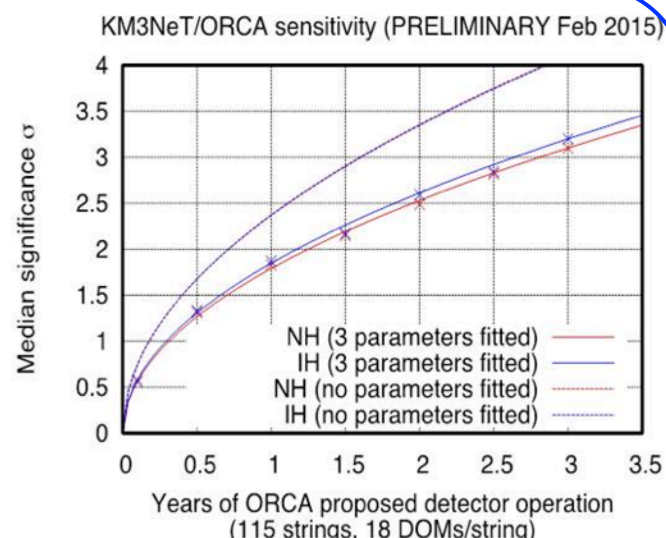
$$PDF(T|IH)=N(-T_0^{IH},2\sqrt{T_0^{IH}})$$
$$PDF(T|NH)=N(T_0^{NH},2\sqrt{T_0^{NH}})$$

typically, T_0^{NH} and T_0^{IH} increase (in absolute value) with exposure, so you get this kind of plots

→ The Power for fixed CL is given by simple formulae



→ The "median" or "typical" sensitivity increases with $\sqrt{\mu}$



More common

→ Hands-on session later today

Quantifying the sensitivity: Bayesian approach

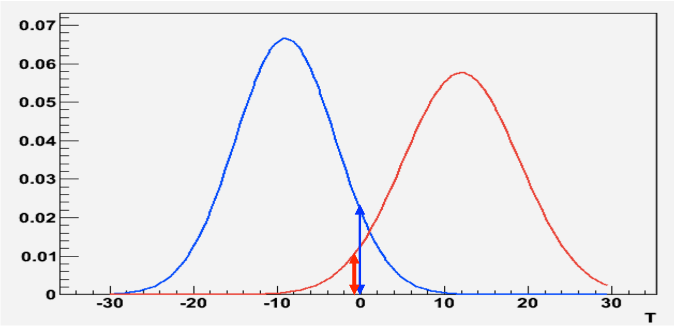
In the "frequentist approach" discussed so far, two sets of information must be provided (for NH and for IH).

With a Bayesian approach [Blen], a single set of values can contain all the information on the test :

- You need a "prior" P on each hypothesis, providing a relative normalization between the two PDFs
- It can be proven that P(NH)=P(IH)=0.5 is the most conservative choice

1) Define a threshold *t*
2) Compute T_c^{NH} such that $$\frac{PDF(T \mid NH)}{PDF(T \mid NH) + PDF(T \mid IH)} > t$$ for $T > T_C^{NH}$

   and T_C^{IH} such that $$\frac{PDF(T \mid IH)}{PDF(T \mid NH) + PDF(T \mid IH)} > t$$ for $T < T_C^{IH}$



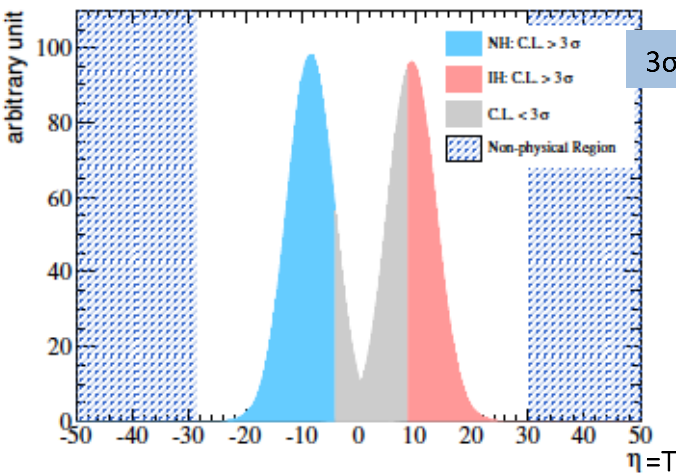
NOTE !!! we are looking the "Odds" or "ratio of posterior probabilities" :
IF YOUR EXPERIMENT GIVES A RESULT T, THE PROBABILITY THAT NATURE IS ACTUALLY NH(IH) is >CL . A question that cannot be answered in the frequentist approach !

The Bayesian approach for MH

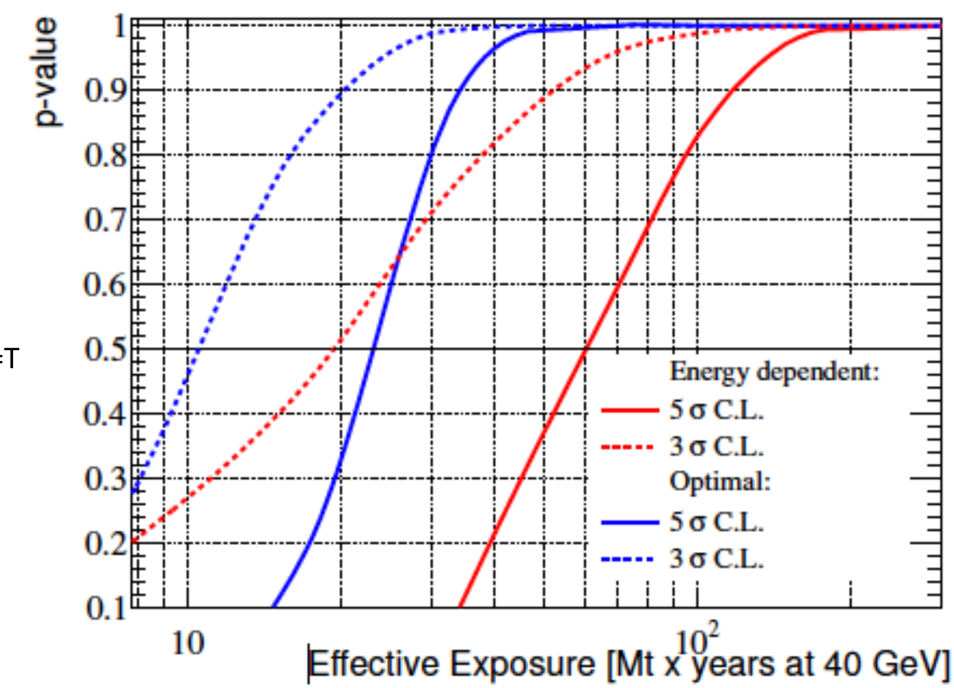
3) Define the "p-value" as $\int_{-\infty}^{T_C^{IH}} PDF(T \mid IH)\,dT + \int_{T_C^{NH}}^{+\infty} PDF(T \mid NH)\,dT$

For example, in

D.Franco et al, **JHEP 1304 (2013) 008** ArXiv 1301.4332 (ORCA/PINGU)



3σ CL means t=99.73%



- T_C' are set by the ratio of the heights of the two gaussians at a given value of T
- p-value is
  0.5* (blue area + red area)

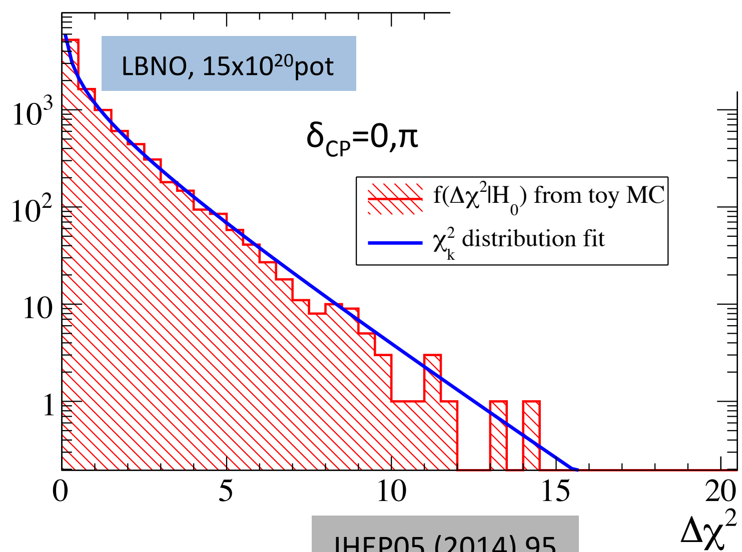
The case of δ_{CP}

Test statistic $\Delta\chi^2 = \min(\Delta\chi^2_{\delta CP=0}, \Delta\chi^2_{\delta CP=\pi})$

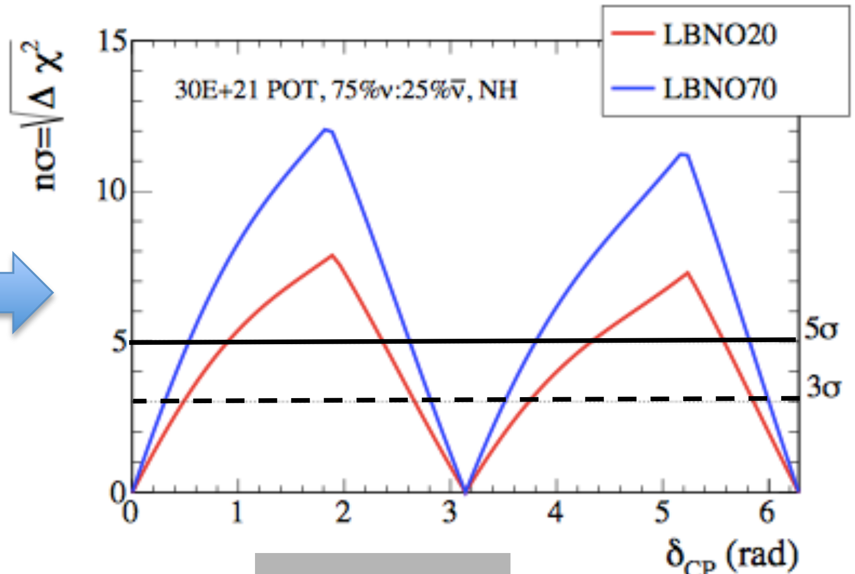
This is a case of "nested hypotheses": $H_0 = \{\delta_{CP}=0 \text{ or } \pi\}$, $H_1 = \{0<\delta_{CP}<2\pi\}$

$\Rightarrow$ Wilk's theorem $PDF(\Delta\chi^2|H_0)=X^2(1dof)$ => $T_C=(\#\text{ of }\sigma)^2$

independent of exposure (unlike for MH)



JHEP05 (2014) 95



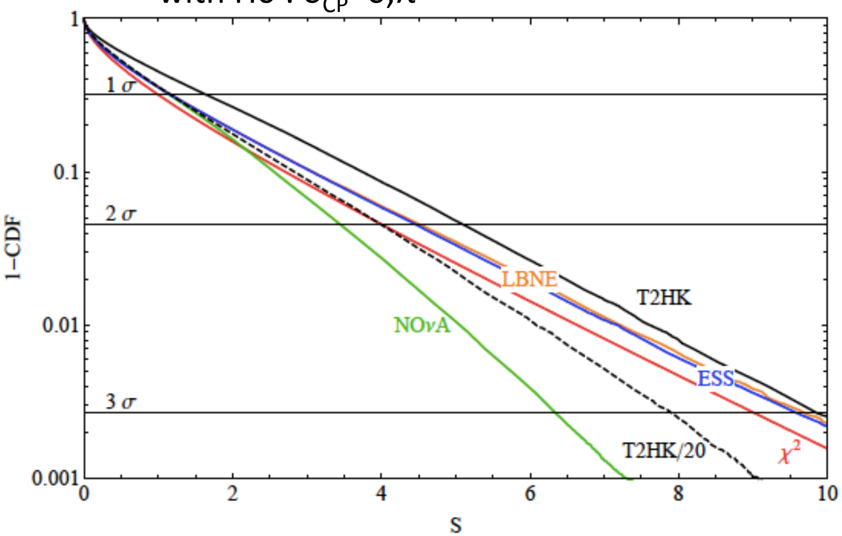
ArXiv:1412.0593

The case of δ_{CP}

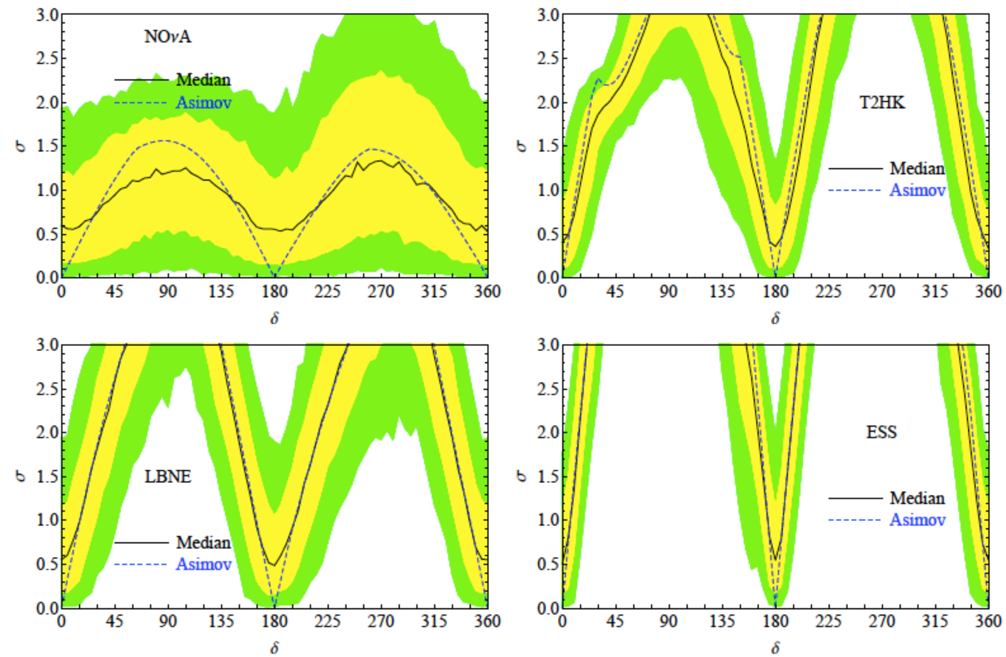
However, Wilk's theorem does not always hold when the experiment has limited sensitivity to δ_{CP} => PDF(T|H_0) is not $\chi^2(1)$

=> need to get PDFs from (toy)-MC

1-cumulative of PDF(T|H_0) from MC
with H0 : δ_{CP}=0,π



Blennow et al., ArXiv:1407.3274



Median sensitivity

- - - - using √T

—— using PDF(T|H_1) from MC

GOODNESS-OF-FIT

Goodness-of-fit

"Are my observations compatible with hypothesis H_0 ?

The alternative hypothesis cannot be formulated ! Therefore

- the Likelihood ratio is useless
- the risk of 2nd kind β is unknown
- the power of the test is undefined
- impossible to tell whether a test is better than another

A new type of problem : "Goodness-of-fit":

compare the experimental data with their p.d.f. under the null hypothesis H_0

If H_0 was true and one repeated the experiment many times, one would obtain data more far away from H_0 than the observed values with probability P

- small P = "bad fit"
- large P = "good fit"

➔ The "P-value" is the figure of merit of the goodness-of

Goodness-of-fit test

To construct a goodness-of-fit test, we need

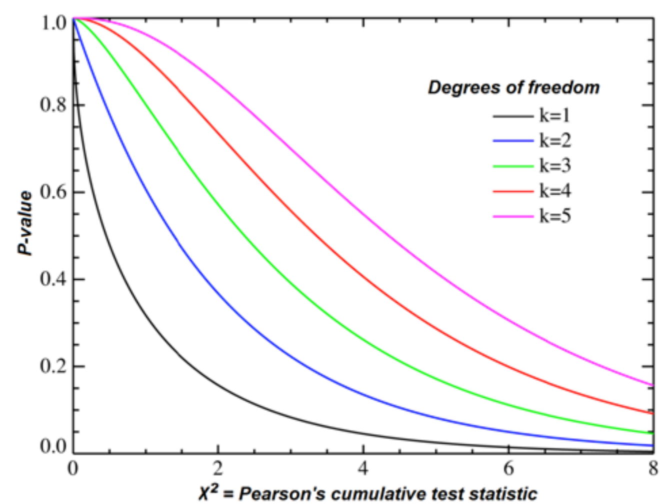
- a test statistic T: a function of H_0 and of the data X, which measures how far away the data is from the hypothesis
- a way to calculate the probability of exceeding the observed value of T, if H_0 were true: a map from values of T to the P-value

$$\mathrm{Prob}(T>T(X)|H_0)=P$$

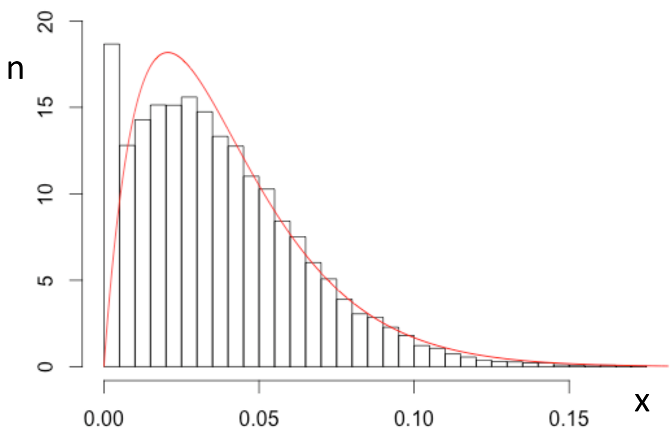
In practice: look for a test statistic T such that its distribution is independent of H_0 ("distribution-free" test) => P does not depend on the details of H_0 => the mapping from T to P is easy to find (tables etc.)

An example of such T: the χ^2 distribution

$$P=\int_{T(X)}^{\infty}\chi_k^2(t)\,dt$$



Pearson's χ^2 test for histograms



k bins
N = total number of events, fixed
$n_1 \ldots n_k$ events per bin: $n_1+\ldots+n_k = N$
V : covariance matrix of the observations (kxk)

Model (H_0): probability per bin = p_i , i=1…k

Consider $\left(\vec{n}-N\vec{p}\right)^T V^{-1}\left(\vec{n}-N\vec{p}\right)$

normalization => rank(V) = (k-1) : take (k-1) terms and sub-matrix W ((k-1)x(k-1));
it can be shown that

$$N\left(W^{-1}\right)_{ij} = \frac{1}{p_i}\delta_{ij} + p_k$$

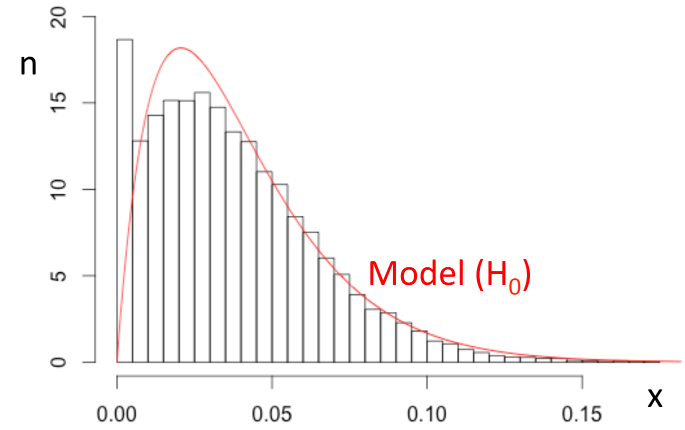
Now consider the statistic $T = \left(\vec{n}-N\vec{p}\right)^T W^{-1}\left(\vec{n}-N\vec{p}\right)$ (only k-1 components of the vectors)
=> T is asymptotically distributed as χ^2(k-1)

Make the expression symmetric in the bins: with some manipulations, you will get

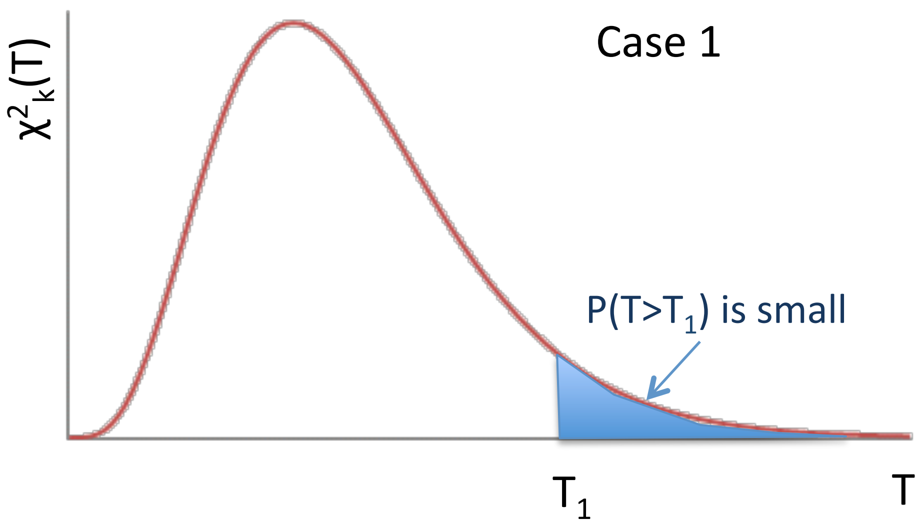
$$\boxed{T = \frac{1}{N}\sum_{i=1}^{k}\frac{\left(n_i - Np_i\right)^2}{p_i} = \frac{1}{N}\sum_{i=1}^{k}\frac{n_i^2}{p_i} - N}$$

close to χ^2(k-1) if Np_i>5

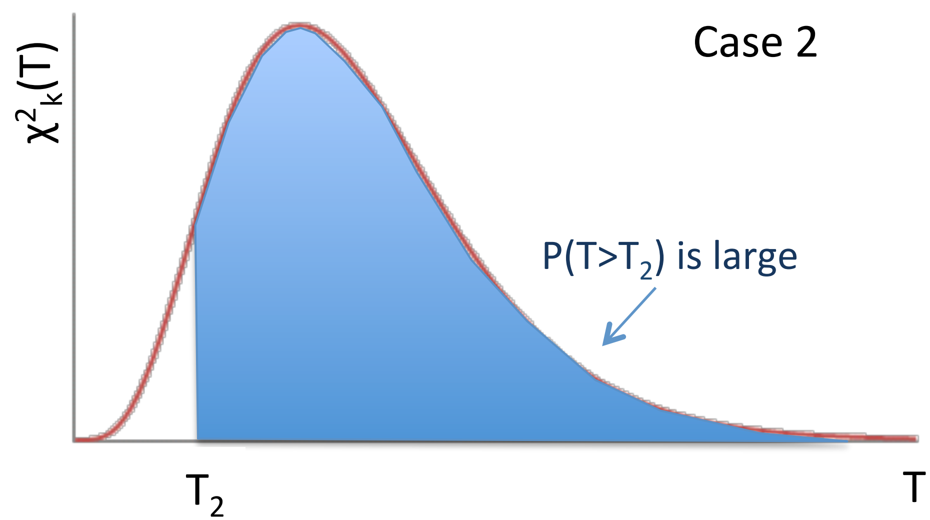
Pearson's χ^2 test for histograms



$$T = \frac{1}{N}\sum_{i=1}^{k}\frac{(n_i - Np_i)^2}{p_i} = \frac{1}{N}\sum_{i=1}^{k}\frac{n_i^2}{p_i} - N$$



"Bad fit" : reject H_0



"Good fit" : accept H_0

Tests free of binning

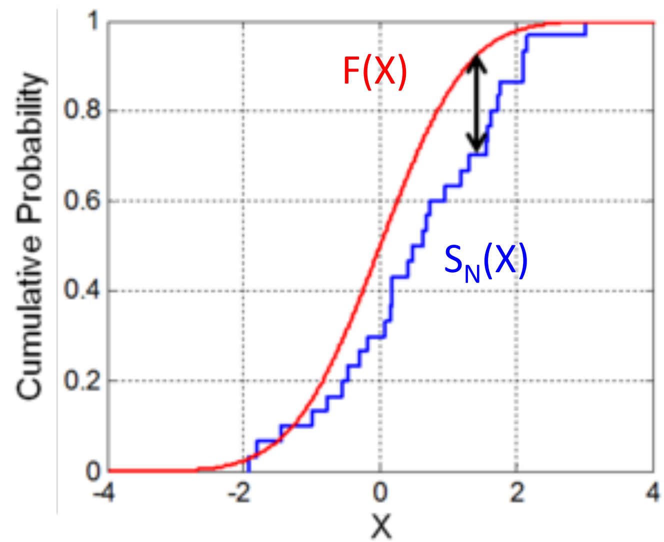
Based on the “distance” between the cumulative distribution of the hypothesis **F(X)** and the empirical distribution function of the data **S_N(X)**

- <u>Smirnov-Cramer-vonMises test</u>

$$W = \int_{-\infty}^{\infty} \left[S_N(X) - F(X) \right]^2 f(X)\,dX$$

=> binomial distribution

with <W>=1/6N and V(W)=(4N-3)/180N^3



- <u>Kolmogorov (-Smirnov) test</u>

$$D_N = \max_{X} \left| S_N(X) - F(X) \right|$$

=> the limiting distribution can be computed

$$\lim_{N \to \infty} P\left(\sqrt{N} D_N > z\right) = 2 \sum_{r=1}^{\infty} (-1)^{r-1} \exp\left(-2r^2 z^2\right)$$

- <u>Anderson-Darling test</u>

$$A = -N - \sum_{i=1}^{N} \frac{2i-1}{N} \left[\ln\left(F(X_i)\right) + \ln\left(1 - F(X_{N+1-i})\right) \right]$$

all available in Root TH1

Sensitivity and Discovery potential

Both are often quoted for future neutrino experiment

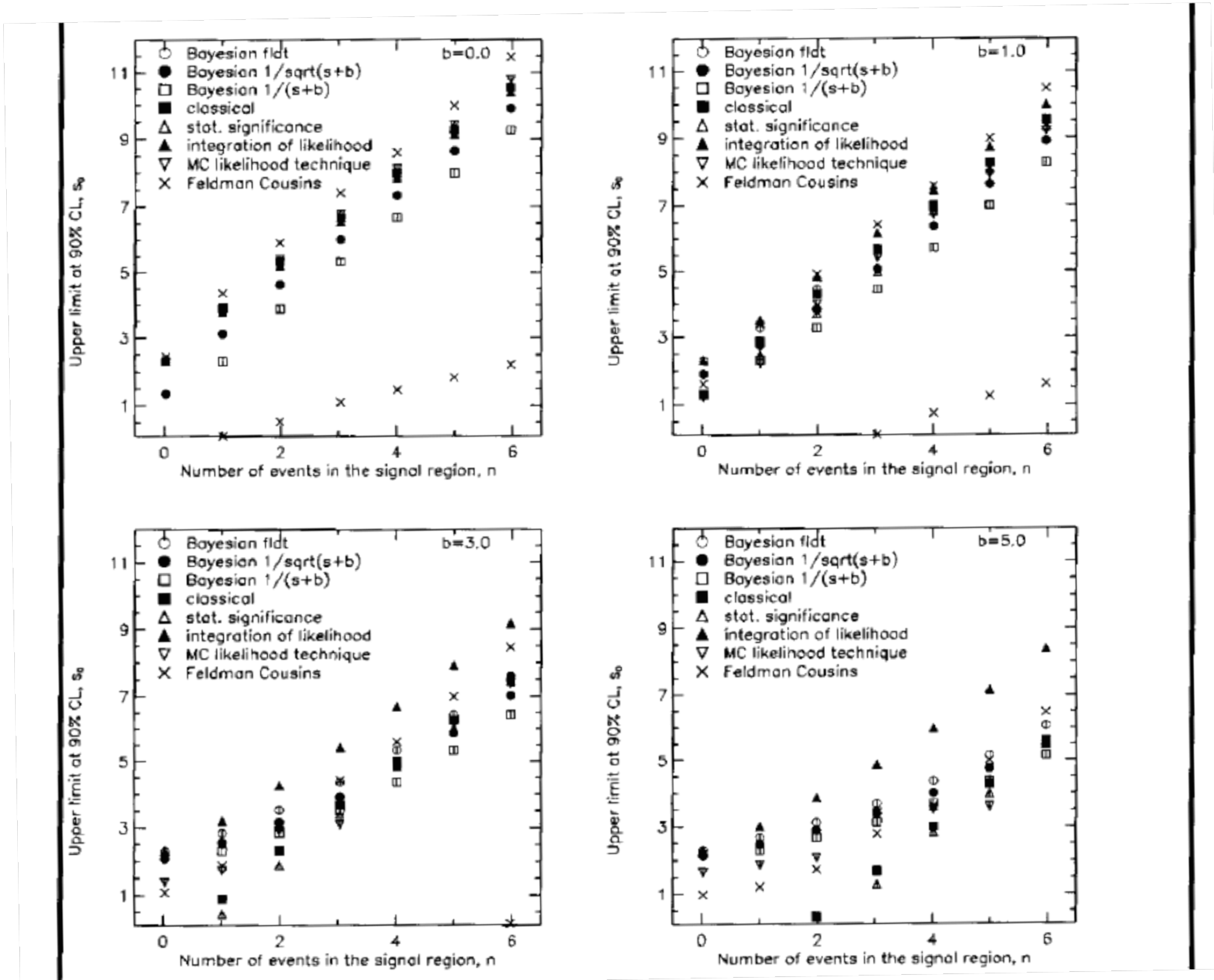
- e.g.: CUORE for ββ-0ν decay of Te, Double-Chooz (in the past) for $\theta_{13}>0$, KM3NeT for point-like ν sources. PROSPECT for sterile neutrinos…

- Sensitivity = fluctuation of the “null hypothesis” (background only) up to a fixed “significance”
  - a goodness-of-fit problem ! significance = 1 – P-value, set the discovery potential at T(P)

- Discovery potential = the smallest signal required to obtain an observation at a given significance level (e.g. 5σ or 3σ) with 50% probability
  - a test of hypotheses ! H_0=background only, H_1=bkg+signal, require α=3σ or 5σ for P=0.5

Different methods give different results…

Example: experiment with mean background b, observed n events: quote upper limit on signal



I. Narsky
PhyStat05

The conclusion

- Be explicit on the statistical method you are using to quote your results (or expected results) !

- … which implies you have to know exactly what you are doing…

SUMMARY of TOPIC n.2

- Test of hypotheses
  - classical approach
  - Neyman-Pearson's likelihood ratio
  - the case of neutrino Mass Hierarchy and the case of δ_{CP}
- Goodness-of-fit
  - chi2 test statistic
  - tests free of binning

Why ?

4 Physics Topics

Within the realm of neutrino physics, subjects for which statistical issues seem particularly relevant and which produced interesting discussions included:

- Fitting parameters for 3 neutrino oscillation situations
- Searching for sterile neutrinos
- Determining the neutrino mass hierarchy
- Determining the CP phase
- Searching for rare processes, e.g. ultra high energy cosmic neutrinos, neutrinoless double beta decay*, supernovae neutrinos, etc.
- Neutrino cross-sections
- Reconstruction and classification issues, e.g. for rings in Cerenkov detectors

L. Lyons, arXiv:1705.01874 [hep-ex]

PHYSTAT-ν Workshop Series :

- May 30 - June 1, 2016, IPMU, Japan
- September 19-21, 2016, Fermilab, USA