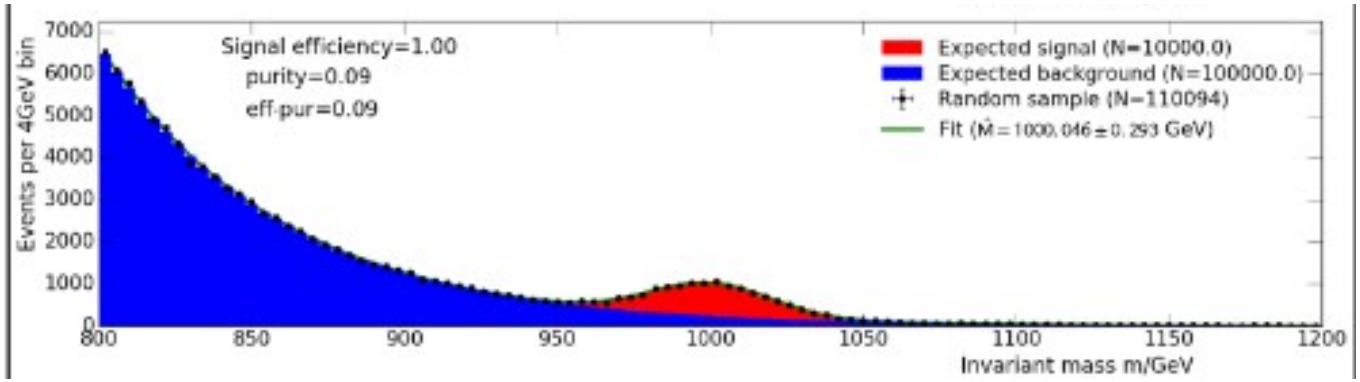
Statistics
(for “daily” data analysis in physics)

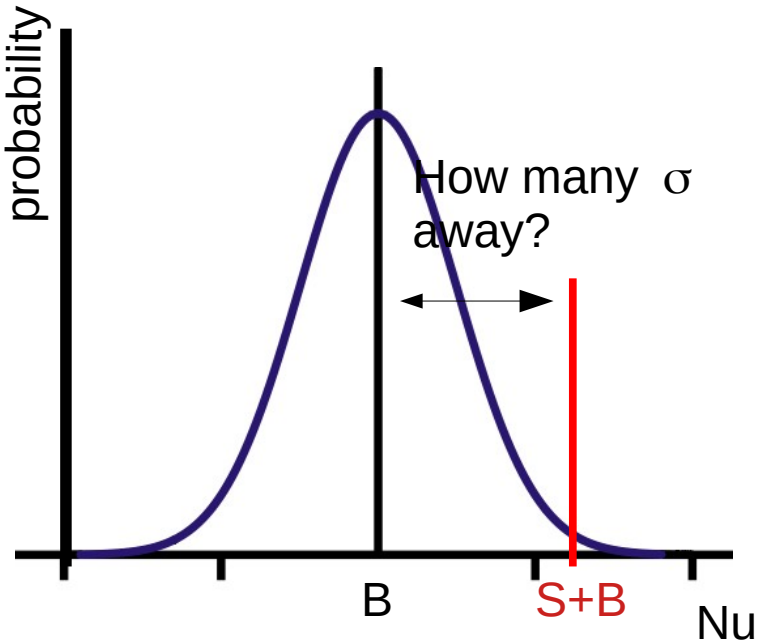
Sara Bolognesi (IRFU, CEA)

Significance



How significant is the presence of signal?

(Metric to optimize analysis cuts to select a subsample with enhanced signal)

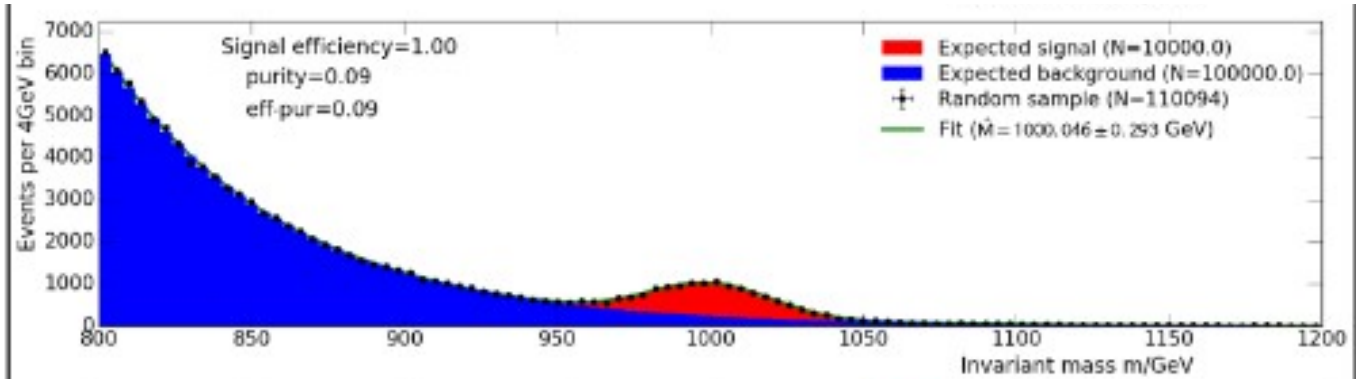


Depend on how large is the (Gaussian) distribution → i.e. uncertainty on B

$\frac{S}{\sqrt{B}}$ If only uncertainty on B is statistical (very small systematics on B)

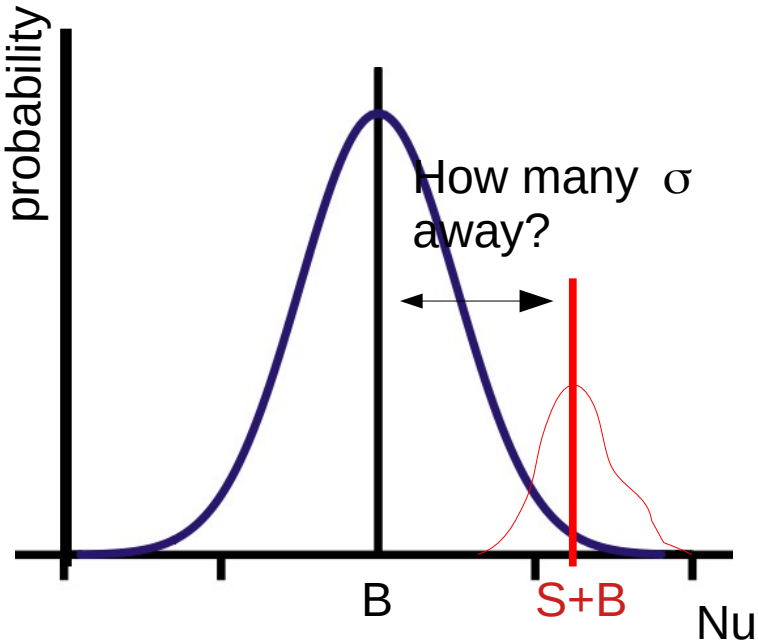
$\frac{S}{\sqrt{\delta B^2 + B}}$ If systematics on B sizable

Significance



How significant is the presence of signal?

(Metric to optimize analysis cuts to select a subsample with enhanced signal)



Depend on how large is the (Gaussian) distribution → i.e. uncertainty on B

$\frac{S}{\sqrt{B}}$ If only uncertainty on B is statistical (very small systematics on B)

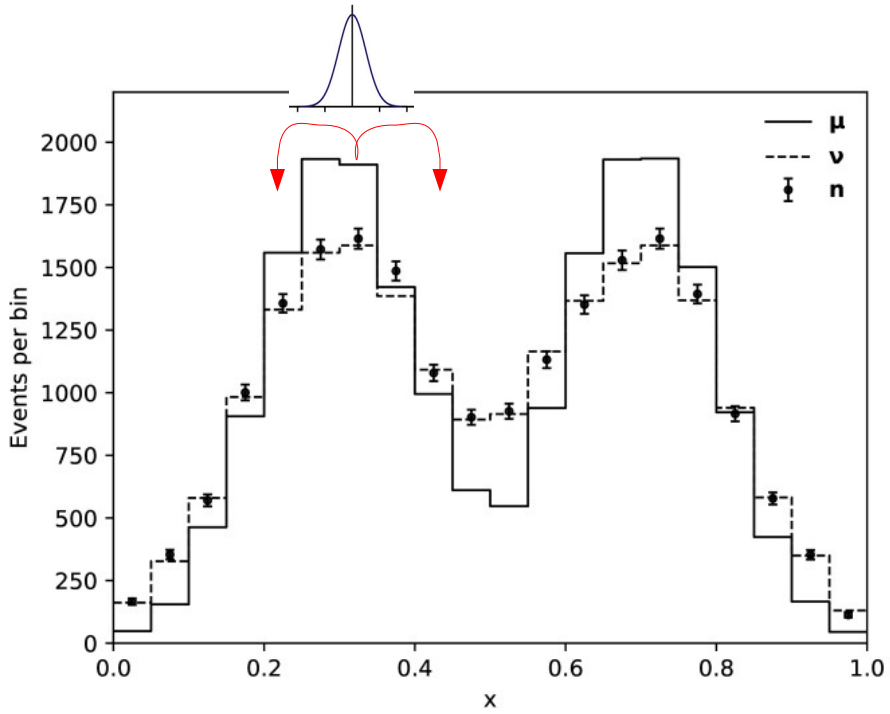
$\frac{S}{\sqrt{\delta B^2 + B}}$ If systematics on B sizable

$\frac{S}{\sqrt{S+B}}$ But also signal has its statistical uncertainty

Actual complete formulation at http://tid.uio.no/epf/seminar/slides/simpleopt.pdf

Resolution

Back to histograms: **observed distribution = true distribution smeared, typically can be described by convolving with a Gaussian.**



true
reco
data

$$N_j^{reco} = N_i^{true} \cdot M_{ij}^{simu}$$

where M_{ji} is a matrix which gives the probability for an event in true bin i to be reconstructed in bin j

Such matrix can be evaluated from MC (typically with cross-check, tuning from control samples)

- How to correct back from observed to "true" → deconvolving detector effects
With histograms is basically an algebric problem

Forward fitting

Yesterday we considered the unfolding (inverting the matrix)
The other possibility is **forward folding**, i.e. describe the true distribution as a function of unknown (to be measured) parameters and performing a fit to find the best values of the parameters which describe the observed data

$$N_i^{true} = N_j^{reco} \cdot M_{ij}^{simu} r_j$$

r_j = renormalize each bin with a semi-free term **with prior value and uncertainty from MC (typically with Gaussian distribution) but to be tuned to data**

Forward fitting

The other possibility is **forward folding**, i.e. describe the true distribution as a function of unknown (to be measured) parameters and performing a fit to find the best values of the parameters which describe the observed data

$$N_i^{true} = N_j^{reco} \cdot M_{ij}^{simu} r_j$$

r_j = renormalize each bin with a semi-free term **with prior value and uncertainty from MC (typically with Gaussian distribution) but to be tuned to data**

- **The fit** is an algorithm that change the MC expectations varying the parameters r_j until it find the 'best match' of MC expectation to data
('best match' = minimum of the likelihood)

Forward fitting

The other possibility is **forward folding**, i.e. describe the true distribution as a function of unknown (to be measured) parameters and performing a fit to find the best values of the parameters which describe the observed data

$$N_i^{true} = N_j^{reco} \cdot M_{ij}^{simu} r_j$$

r_j = renormalize each bin with a semi-free term **with prior value and uncertainty from MC (typically with Gaussian distribution) but to be tuned to data**

- **The fit** is an algorithm that change the MC expectations varying the parameters r_j until it find the 'best match' of MC expectation to data
('best match' = minimum of the likelihood)

- **Likelihood** ~ function which described how well the data match with my model/expectations. Actually in frequentist terms: how probable is to observe my data, given the model

$L\left(N_j^{data}; N_j^{simu} \cdot f\left(\alpha_k\right)\right)$ built in such a way to be minimal when $N_j^{data} \sim N_i^{simu} \cdot f\left(\alpha_k\right)$

- α_k are parameters describing 'freedom' in the expectation:
  - parameters you want to measure (aka parameters of interest)
  - systematic uncertainties on the model, both the physics model and the detector model (aka nuisances parameters)
- likelihood function written in a statistically correct way to consider statistical uncertainty in data and prior knowledge/uncertainty on nuisances

Systematic uncertainties (aka nuisance parameters)

- The expectations and their dependence on nuisances $N_i^{simu}\, f(\alpha_k)$ typically can be

  - in form of a **full analytical description** (typically unpractical since it is difficult to encode in a single analytical function all the detector effects and their possible variations)

  - in form of a **simulated histogram** which is reproduced with full simulation at each variation of all the parameters (typically unpractical since it is computationally expensive to perform a full simulation for each fit iteration)

  - in form of a simulated histogram with **parametrization of uncertainties in form of reweigthing of the histogram**

Systematic uncertainties (aka nuisance parameters)

- The expectations and their dependence on nuisances $N_i^{simu}\, f(\alpha_k)$ typically are

  - in form of a **full analytical description** (typically unpractical since it is difficult to encode in a single analytical function all the detector effects and their possible variations)

  - in form of a **simulated histogram** which is reproduced with full simulation at each variation of all the parameters (typically unpractical since it is computationally expensive to perform a full simulation for each fit iteration)

  - in form of a simulated histogram with **parametrization of uncertainties in form of reweigthing of the histogram**

- Typically α_k **are not completely free: they are known with a certain precision** from control samples or from simulation

  → included in the likelihood with a 'penalty term' which makes the likelihood large (i.e. makes bad data-MC match) if the nuisance parameters value move away from the 'prior' estimated value

  → you need to decide how well you know this prior value and what is the distribution of its uncertainty

Systematic uncertainties (aka nuisance parameters)

- The expectations and their dependence on nuisances $N_i^{simu}\, f(\alpha_k)$ typically are

  - in form of a **full analytical description** (typically unpractical since it is difficult to encode in a single analytical function all the detector effects and their possible variations)

  - in form of a **simulated histogram** which is reproduced with full simulation at each variation of all the parameters (typically unpractical since it is computationally expensive to perform a full simulation for each fit iteration)

  - in form of a simulated histogram with **parametrization of uncertainties in form of reweigthing of the histogram**

- Typically α_k **are not completely free: they are known with a certain precision** from control samples or from simulation

  → included in the likelihood with a 'penalty term' which makes the likelihood large (i.e. makes bad data-MC match) if the nuisance parameters value move away from the 'prior' estimated value

  → you need to decide how well you know this prior value and what is the distribution of its uncertainty

- Typically α_k prior knoweldge/uncertainty is assumed Gaussian but not always obvious.

  For instance theoretical uncertainty → you can use other distributions (eg flat)

Likelihood

Likelihood ~ function which described how well the data match with my model/expectations.
Actually in frequentist terms: how probable is to observe my data, given the model

$$L\left(N_j^{data}; N_j^{simu}\right) =$$

$$\sum_{j}^{reco\,bins} 2\left(N_j^{simu} - N_j^{data} + N_j^{data} \ln\left(\frac{N_j^{data}}{N_j^{simu}}\right)\right) + \ldots$$

Statistical term: minimum when data ~ simu and written in a statistical correct way for Gaussian (Poisson) uncertainties

All derivations here (*)

Likelihood

Likelihood ~ function which described how well the data match with my model/expectations.
Actually in frequentist terms: how probable is to observe my data, given the model

$$L\left(N_j^{data}; N_j^{simu}\cdot f\left(\alpha_k\right)\right) =$$

$$\sum_{j}^{reco\,bins} 2\left(N_j^{simu}\cdot f\left(\alpha\right) - N_j^{data} + N_j^{data}\ln\left(\frac{N_j^{data}}{N_j^{simu}\cdot f\left(\alpha\right)}\right)\right) +$$

$$\sum_{k,i}\left(\alpha_k - \alpha_k^{prior}\right) M_{ki}\left(\alpha_i - \alpha_i^{prior}\right)$$

"Chi square" multidimensional term considering possible correlations in prior knowledge of nuisances. Large if value of α away from prior (some freedom with σ_α)

1D: $$\chi^2 = \frac{\left(\alpha - \alpha^{prior}\right)}{\sigma_\alpha}$$

The minimization algorithm will change α value until finding the value which make data ~ MC at smallest possible expense of deviation from α prior

Toys and Asimov

You exercise/tune/develop your fit on Monte Carlo samples:

- you produce a sample of simulated events which is your reference MC sample to evaluate N_i^{simu} $f(\alpha_k)$

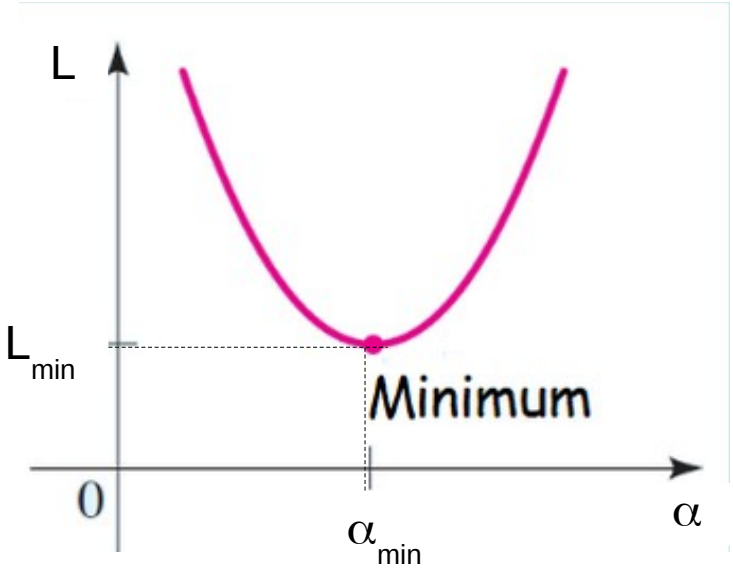
- you produce many other samples of simulated events with small variations: eg, statistical fluctuations as expected in data, small change of systematic value (eg slightly larger detector efficiency, resolution…)
→ You analyse these samples as they were many examples of actual data

Asimov fit = fit of the reference MC sample to itself: both N_i^{simu} $f(\alpha_k)$ and N_i^{simu} $f(\alpha_k)$ from the same MC reference sample
→ the fit must converge to your expectation by definition (basic closure test)
→ you can use it to estimate the **expected sensitivity** (i.e. postfit precision on parameter of interest)

Toys = fit of the reference MC sample to the 'varied' samples of MC
→ all the fits **should** converge
→you can use it to look how your data may look like: in principle data should look like one of those sample → how data fit is similar to them? **(P-value, Confidence level...)**

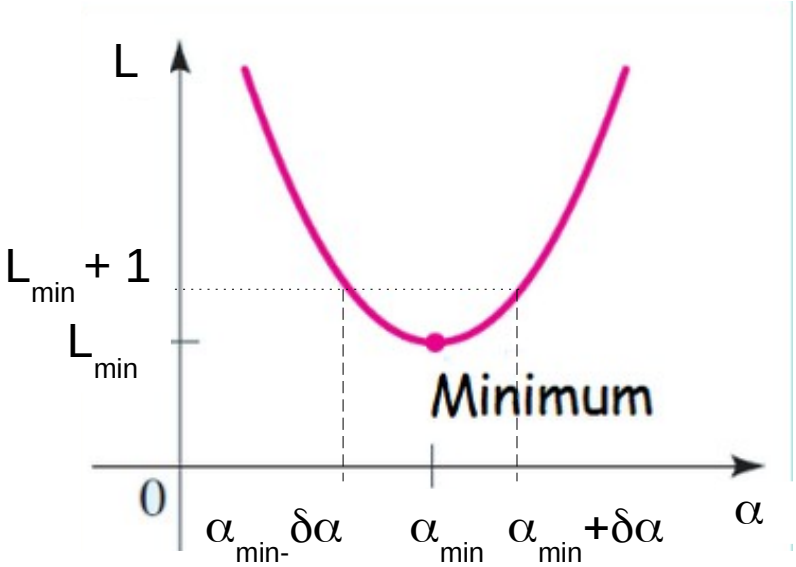
How the fit works: 1D example

MINUIT (or any other algorithm) will find the minimum for you



How the fit works: 1D example

MINUIT (or any other algorithm) will find the minimum for you



How to define "1 sigma" error on α?

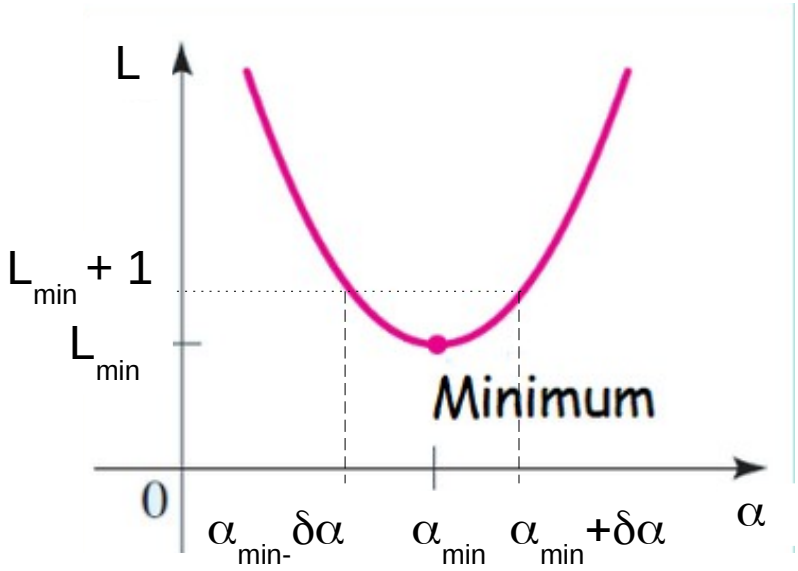
If the likelihood is a $\chi 2$, ie all your uncertainties have a Gaussian distribution then you have the simple $\chi 2$ rules

L_{min} + 1 → α_{min} +/- $\delta\alpha$

Why?
All derivations here (*)

How the fit works: 1D example

MINUIT (or any other algorithm) will find the minimum for you



How to define "1 sigma" error on α?

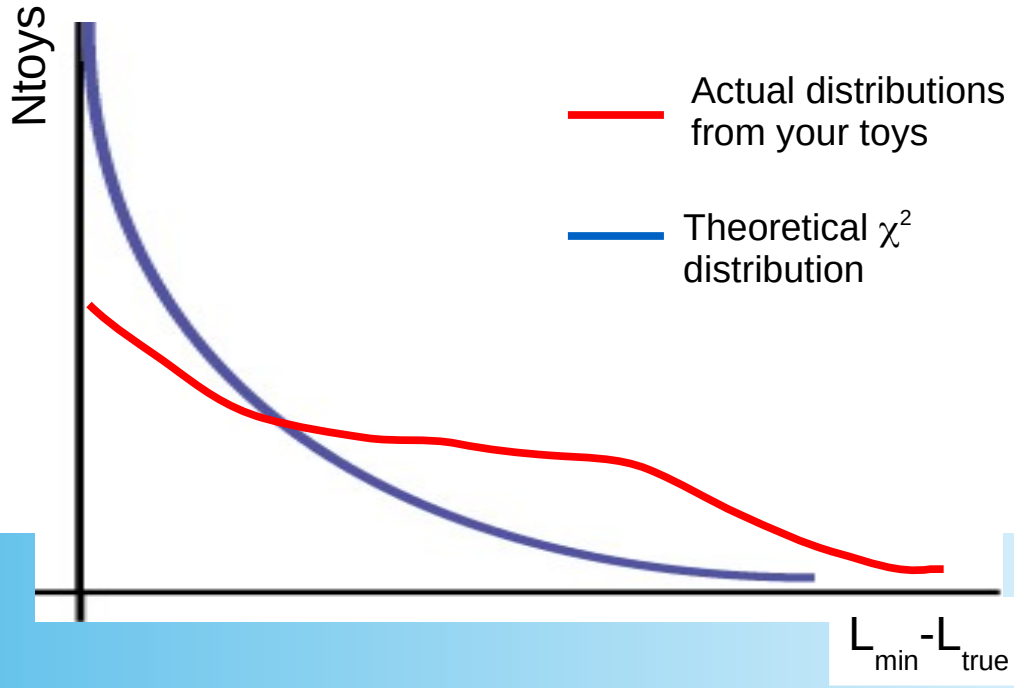
If the likelihood is a χ2, ie all your uncertainties have a Gaussian distribution then you have the simple χ2 rules

L_{min} + 1 → α_{min} +/- $\delta\alpha$

Typically real world is never perfectly Gaussian

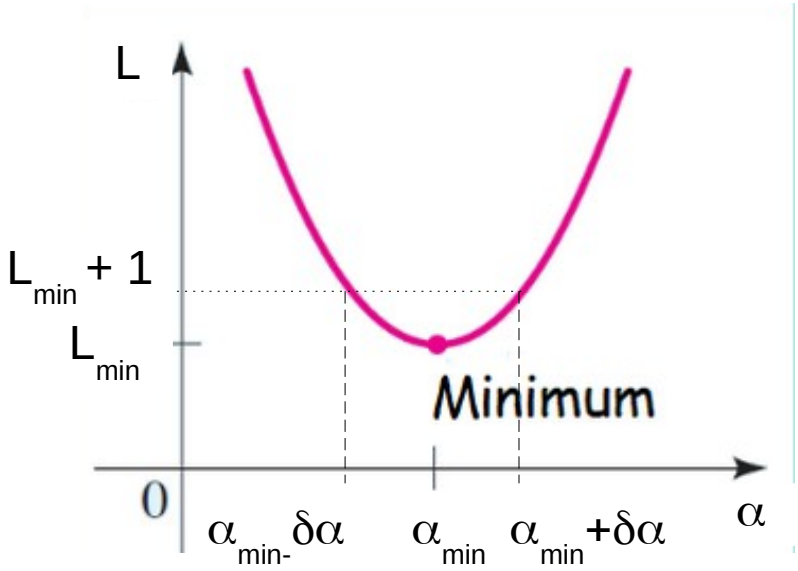
→ toys: run many fits on MC by changing the prior values of your parameters around true values

→ look at distribution of L_{min}-L_{true}



How the fit works: 1D example

MINUIT (or any other algorithm) will find the minimum for you



Typically real world is never perfectly Gaussian

→ toys: run many fits on MC by changing the prior values of your parameters around true values

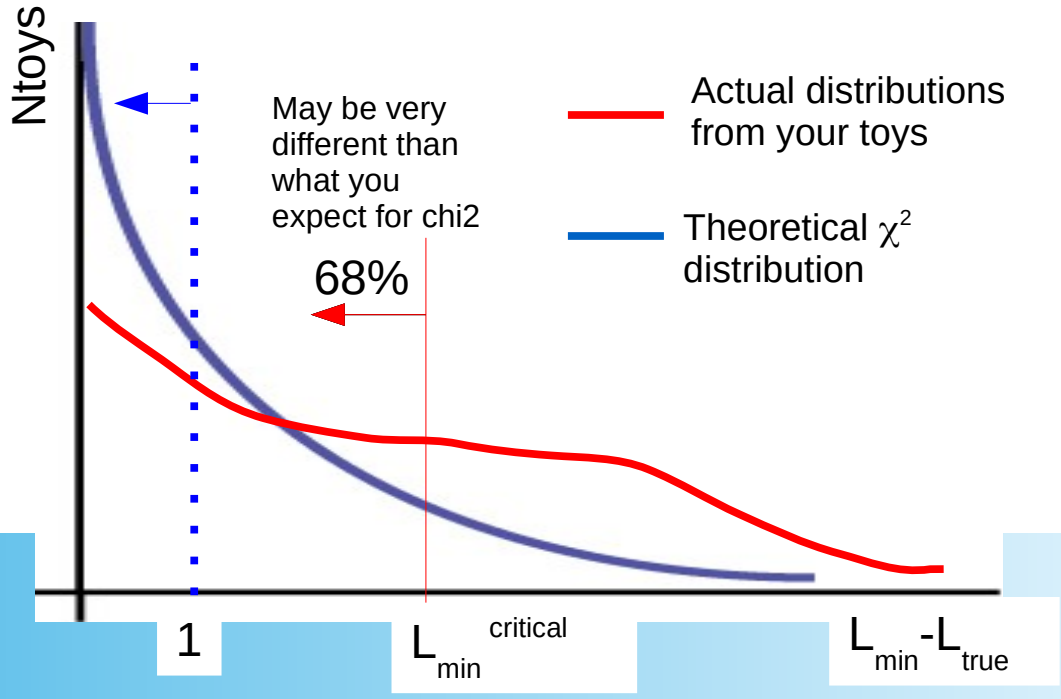
→ look at distribution of L_{min}-L_{true}

e.g. integrate over 68% of your results to know the ΔL~'1σ' error

How to define "1 sigma" error on α?

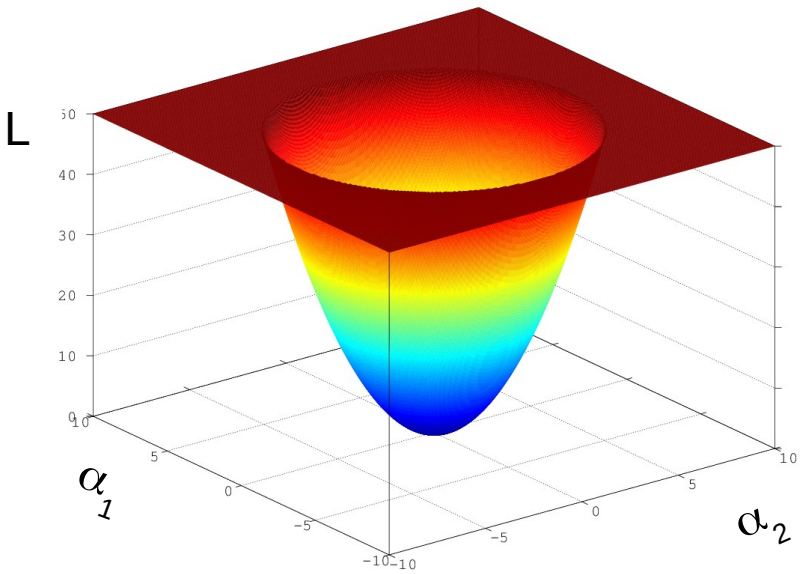
If the likelihood is a $\chi 2$, ie all your uncertainties have a Gaussian distribution then you have the simple $\chi 2$ rules

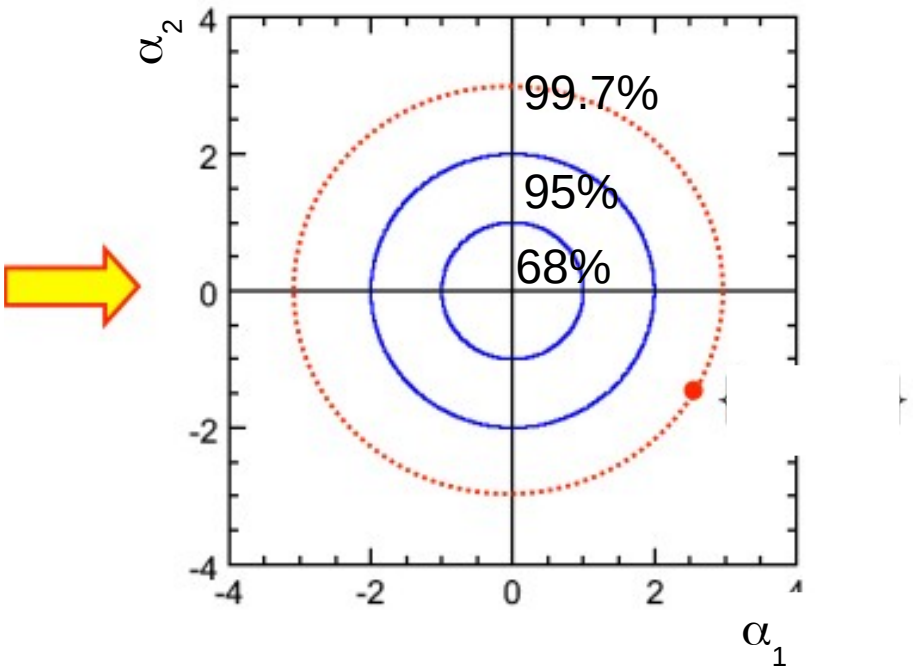
L_{min} + 1 → α_{min} +/- $\delta\alpha$



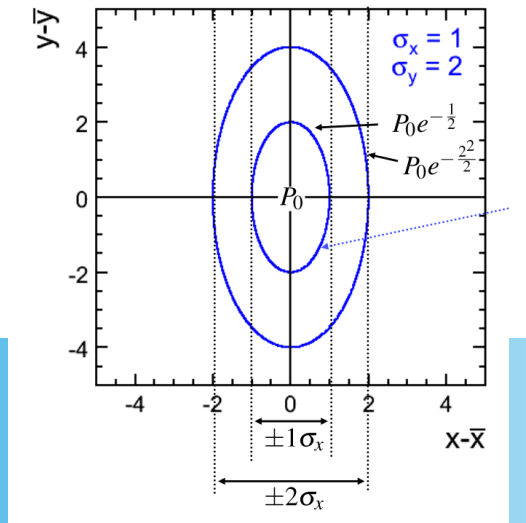
Many dimensions

Typically the likelihood is multidimensional (since you have many unknown parameters α_k)





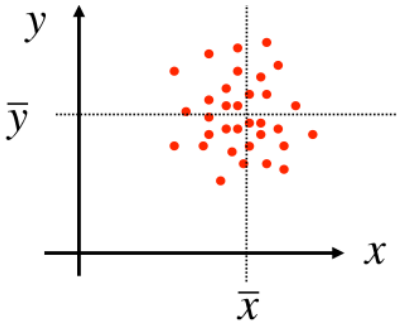
In general if **correlations** are present between parmeters → non-circular projection (eg ellipses)



Linear correlations

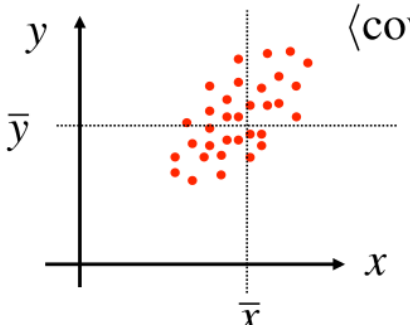
Imagine repeating the measurement of two variables (x,y) many times

- if the two measurement are independent → uncorrelated



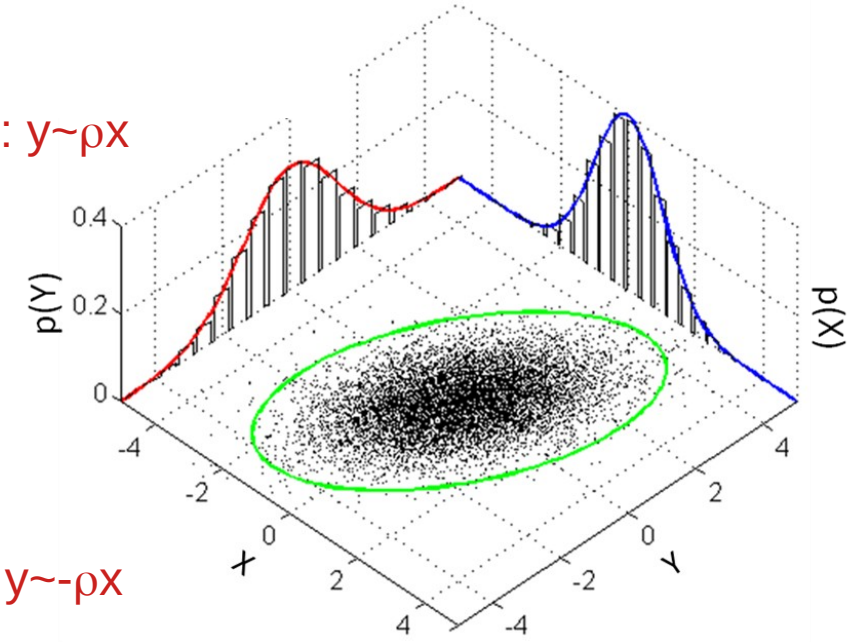
$$\langle \mathrm{cov}(x,y) \rangle = 0$$

- if the two measurement are positively correlated: y~ρx

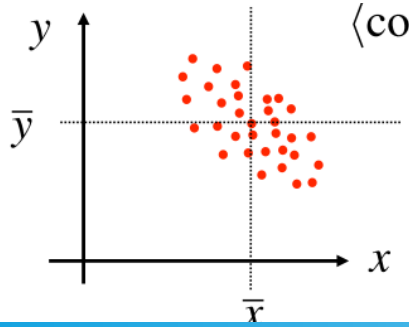


$$\langle \mathrm{cov}(x,y) \rangle > 0$$

(e.g. two xsec measurement at same experiment share same uncertainty on L: luminosity)



- if the two measurement are negatively correlated y~-ρx

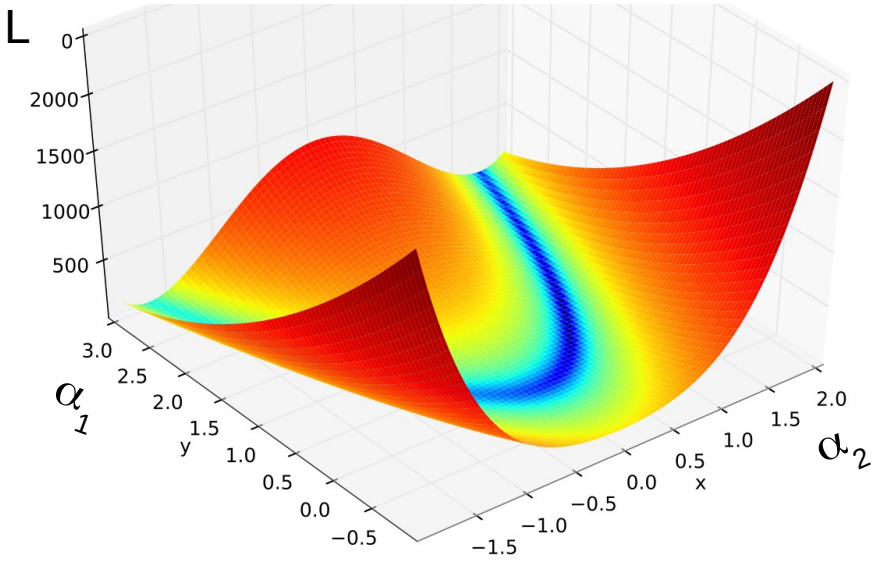


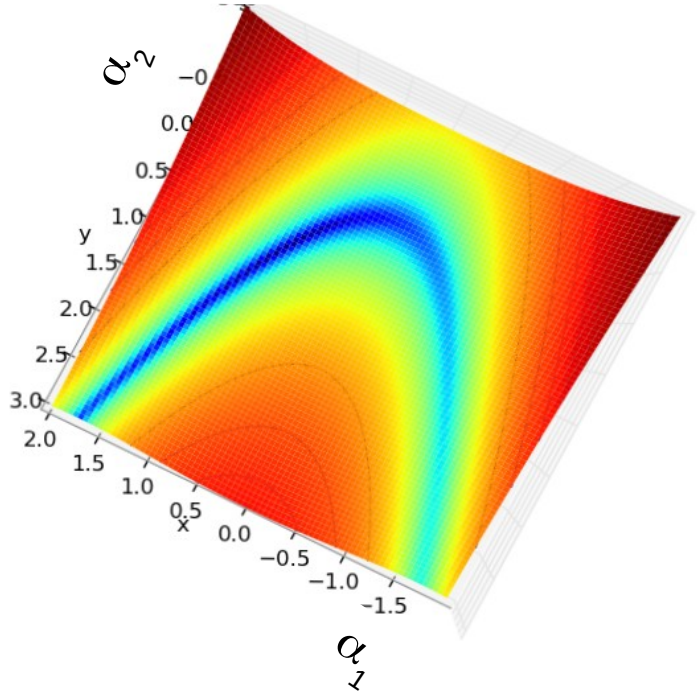
$$\langle \mathrm{cov}(x,y) \rangle < 0$$

(e.g. rate of $\nu_e \rightarrow e$ and ν_μ → μ are linked by μ-e mis-identification)

Relations

Typically the likelihood is multidimensional and you (almost certainly) have relations between various free parameters





The projection of “1 sigma” contour on a set of measured variables may not be circular
Eg here relation such that $\alpha_2 \sim \alpha_1^2$

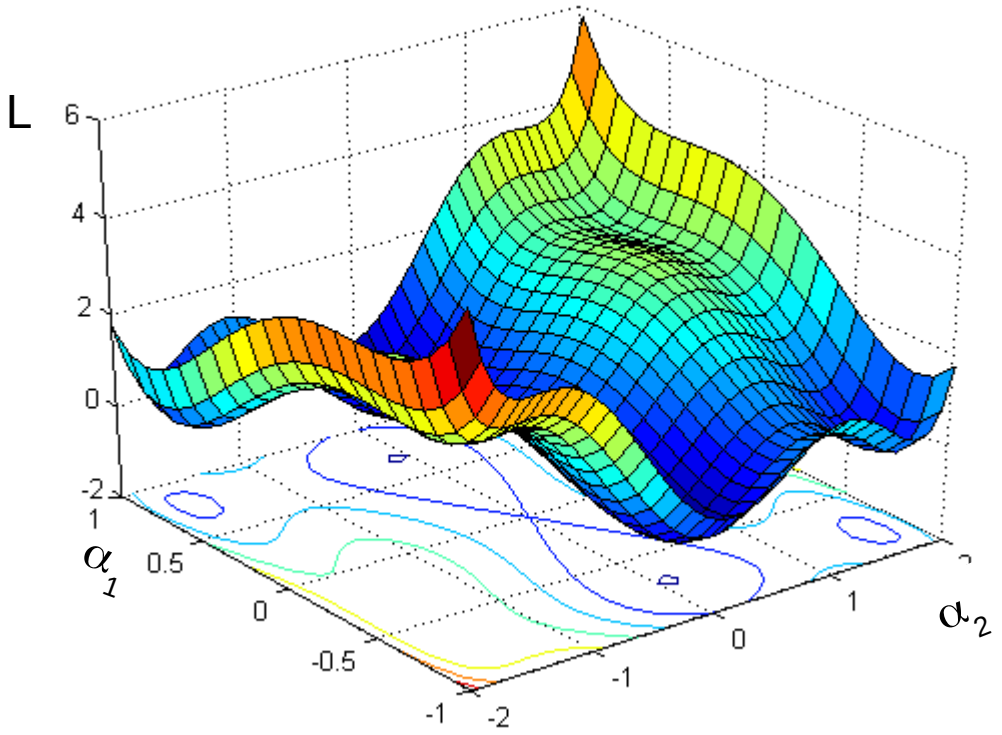
(By the way: if $x \sim y^2$ then mathematically their correlation = 0 !)

Nice example at
http://www.statisticalengineering.com/correlation.htm

Multiple minima

The likelihood may have multiple minima which appears as ‘islands’ in the projected contours

→ careful to explore the likelihood in all the domain where minima could be present (otherwise you may overestimate the power of your experimental data)



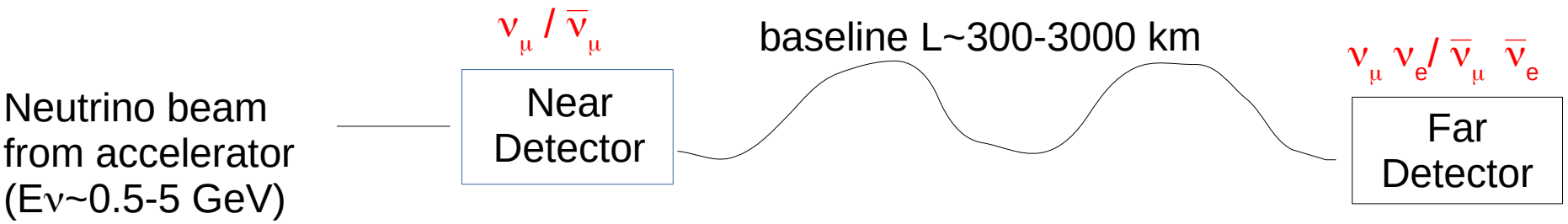
BREAK !

Practical example 1

Neutrino oscillation:

fit to likelihood of near and far detector data to extract best value and uncertainty of parameters of interest which dictate the oscillation probability

Practical example: ν oscillation



Oscillation probability estimated by comparing ν_μ and ν_e rate between near and far detectors:

$$P(\nu_\alpha \rightarrow \nu_\beta) = \underbrace{\sin^2(2\theta)}_{\text{amplitude}} \sin^2\left(\underbrace{1.27\frac{\Delta m^2_{ji}[\text{eV}^2]L[\text{km}]}{E_\nu[\text{GeV}]}}_{\text{frequency}}\right)$$

(simplified 2-flavors approximation)

Parameters of interest: mixing angle θ, mass difference Δm^2 between neutrino mass-eigenstates

Practical example: ν oscillation

$$N^{FD}_{\nu_{\alpha'}} \approx P_{\nu_\alpha \rightarrow \nu_{\alpha'}} \times N^{ND}_{\nu_\alpha}$$

Number of neutrinos at the Far Detector (FD) of a given flavour α' (α=e,μ,τ)

Number of neutrinos at the Near detector (ND) of a given flavour α' (α=e,μ,τ)

The **oscillation probability** $\nu_\alpha \rightarrow \nu_{\alpha'}$ which you want to estimate: it depends on the parameters you want to measure (mixing angle θ, mass difference Δm^2)

ν oscillation: near detector

Predictions depend on number of produced neutrinos and their probability to interact.

$$N_{\nu_\alpha}^{ND}(E_\nu) = \phi(E_\nu) \times \sigma(E_\nu)\, dE_\nu$$

flux= number of neutrinos produced by the accelerator

cross-section = probability of interaction of the neutrinos in the material of the detector

ν oscillation: near detector

Predictions depend on number of produced neutrinos and their probability to interact.

$$N_{\nu_\alpha}^{ND}(E_\nu) = \phi(E_\nu) \times \sigma(E_\nu)\, dE_\nu$$

flux= number of neutrinos produced by the accelerator

cross-section = probability of interaction of the neutrinos in the material of the detector

Detector effects:

$$\epsilon = \frac{N_{\nu_\alpha}^{signal-measured}}{N_{\nu_\alpha}^{signal}}$$

efficiency corrects for events which escape the detection
(threshold, acceptance, containment...)

$$p = \frac{N_{\nu_\alpha}^{measured} - N^{background}}{N_{\nu_\alpha}^{measured}} = \frac{N_{\nu_\alpha}^{signal-measured}}{N_{\nu_\alpha}^{measured}}$$

purity corrects for background
(events wrongly identified as ν_α)

ν oscillation: near detector fit

$$N_{\nu_\alpha}^{ND}(E_\nu) \approx \phi_{\nu_\alpha}^{ND}(E_\nu) \times \sigma_{\nu_\alpha}^{ND}(E_\nu) \times \frac{1}{\epsilon^{ND}} \times p^{ND}$$

Model implemented in **MC simulation predicts expectations for flux and cross-section and detector effects:** uncertainties described by nuisance parameters constrained in a fit to near detector data

$$N_j^{data} \sim N_i^{simu} \cdot f(b_k, x_i, d_j)$$

Likelihood fit to find nuisance values which best match the data:

$$-2\log L_{ND} = 2\sum_{i=0}^{N_{bins}} \left(N_i^{pred}(\vec{b},\vec{x},\vec{d}) - N_i^{obs} + N_i^{obs} \log \frac{N_i^{obs}}{N_i^{pred}(\vec{b},\vec{x},\vec{d})} \right)$$

$$+ \sum_{i=0}^{N_b}\sum_{j=0}^{N_b} \Delta\vec{b}_i \left(V_b^{-1}\right)_{ij} \Delta\vec{b}_j^T$$

flux (b)

$$+ \sum_{i=0}^{N_x}\sum_{j=0}^{N_x} \Delta\vec{x}_i \left(V_x^{-1}\right)_{ij} \Delta\vec{x}_j^T$$

cross-section (x)

$$+ \sum_{i=0}^{N_d}\sum_{j=0}^{N_d} \Delta\vec{d}_i \left(V_d^{-1}\right)_{ij} \Delta\vec{d}_j^T$$

detector (d)

Nuisances (b, x, d) are not completely free: prior knowledge from simulation and control-samples in “penalty terms”

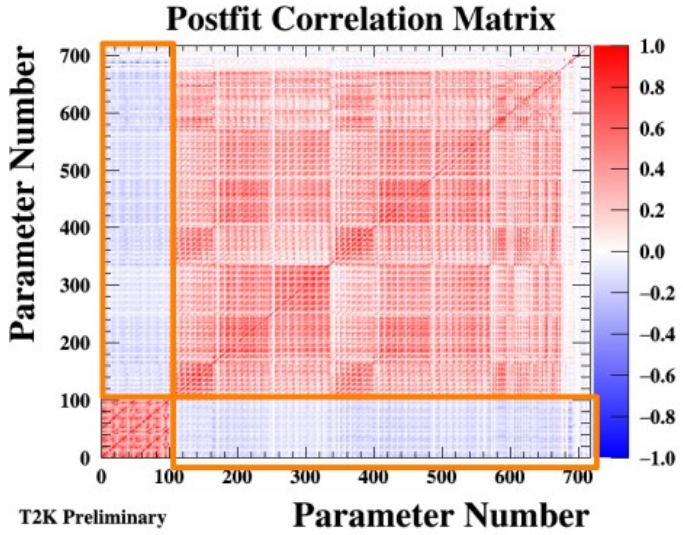
ν oscillation: near detector fit

$$N_{\nu_\alpha}^{ND}(E_\nu) \approx \phi_{\nu_\alpha}^{ND}(E_\nu) \times \sigma_{\nu_\alpha}^{ND}(E_\nu) \times \frac{1}{\epsilon^{ND}} \times p^{ND}$$

Model implemented in **MC simulation predicts expectations for flux and cross-section and detector effects:** uncertainties described by nuisance parameters constrained in a fit to near detector data

$$N_j^{data} \sim N_i^{simu} \cdot f(\alpha_k) \quad \Rightarrow \quad \alpha_k \pm \delta\alpha_k \rightarrow \alpha_k M_{ik} \alpha_i$$

Actually the **flux and xsec uncertainties are strongly anticorrelated**



Number of events in data ~ flux times xsec → if you increase xsec then you need to decrease xsec and viceversa...

ν oscillation: far detector fit

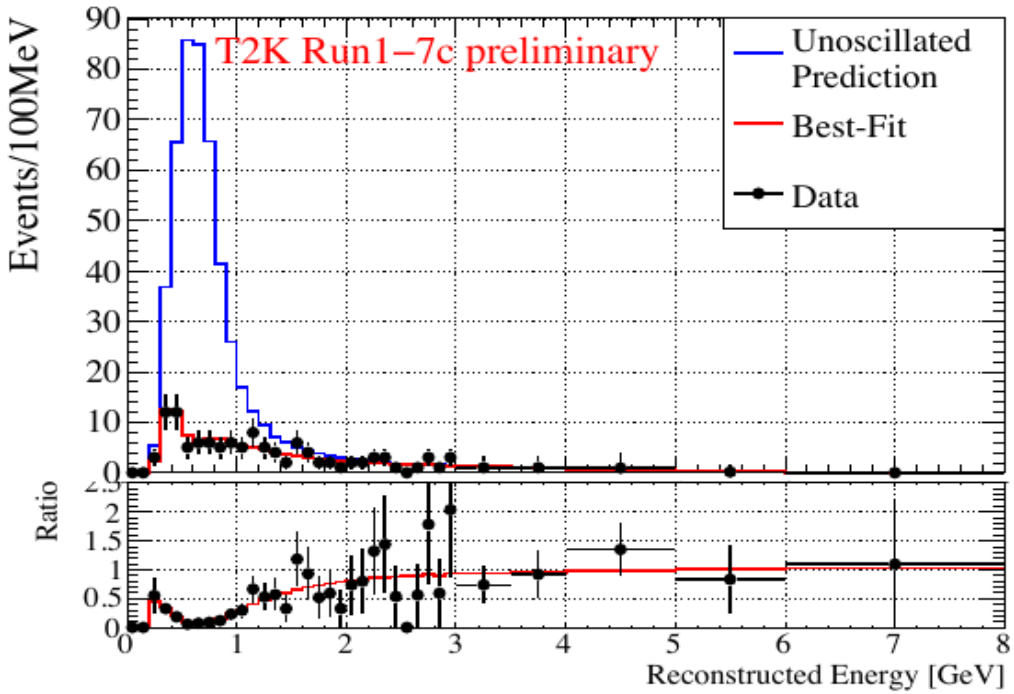
$$N^{FD}_{\nu_{\alpha'}}(E_\nu) \approx P_{\nu_\alpha \to \nu_{\alpha'}}(E_\nu) \times \phi^{FD}_{\nu_{\alpha'}}(E_\nu) \times \sigma^{FD}_{\nu_{\alpha'}}(E_\nu) \times \frac{1}{\epsilon^{FD}} \times p^{FD}$$

$$N_j^{data} \sim N_i^{simu} \cdot f(\alpha_k, \beta)$$

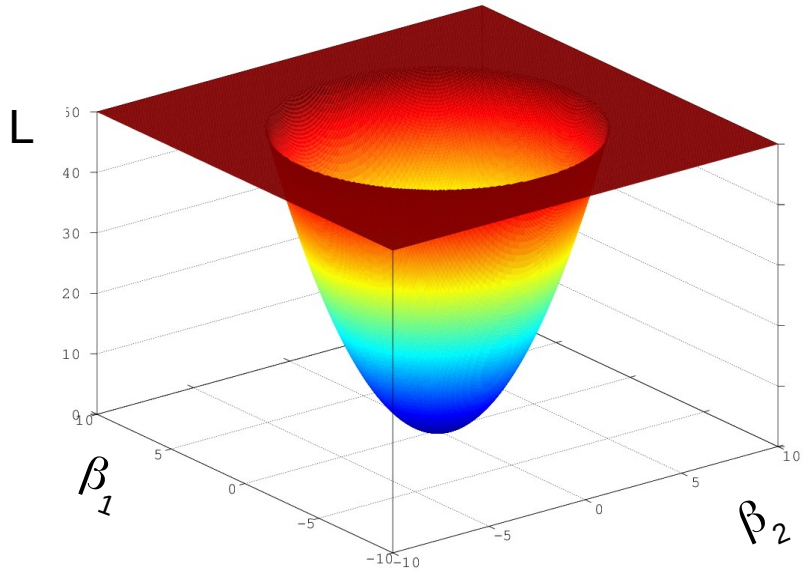
where α **are nuisances of flux and xsec strongly constrained by ND + nuisances on detector systematics (efficiency and purity)**

β = oscillation parameters. Described by standard oscillation formulas (PMNS)

$$P(\nu_\alpha \to \nu_\beta) = \sin^2(2\theta)\sin^2\left(1.27\frac{\Delta m^2_{ji}[\mathrm{eV}^2]L[\mathrm{km}]}{E_\nu[\mathrm{GeV}]}\right)$$



2D likelihood over 2 parameters of interest



ν oscillation: far detector fit

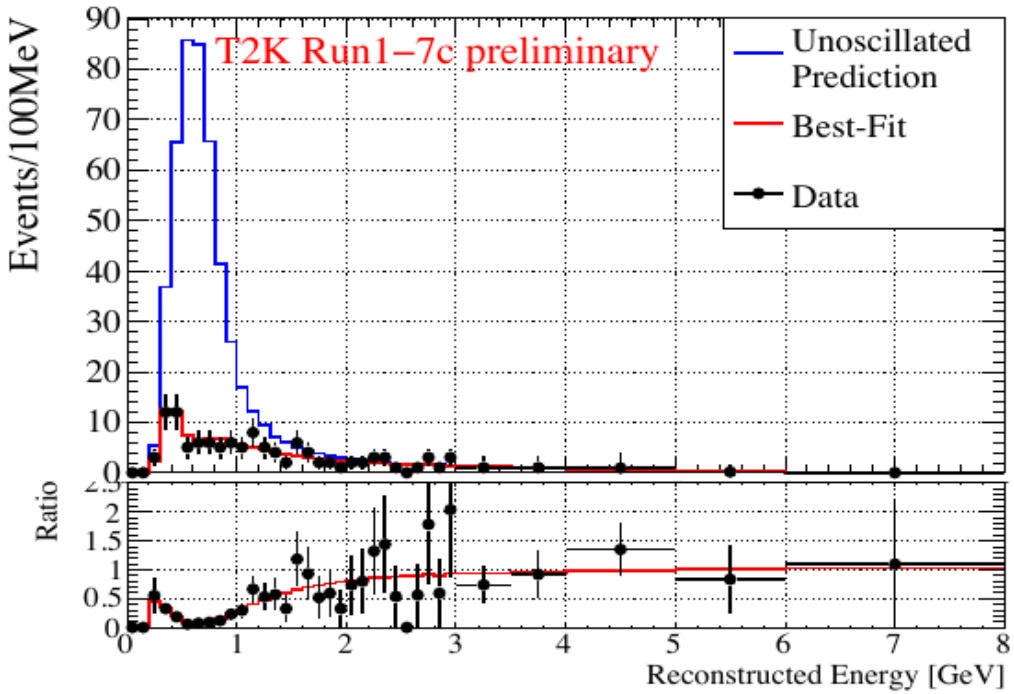
$$N^{FD}_{\nu_{\alpha'}}(E_\nu) \approx P_{\nu_\alpha \to \nu_{\alpha'}}(E_\nu) \times \phi^{FD}_{\nu_{\alpha'}}(E_\nu) \times \sigma^{FD}_{\nu_{\alpha'}}(E_\nu) \times \frac{1}{\epsilon^{FD}} \times p^{FD}$$

$$N_j^{data} \sim N_i^{simu} \cdot f(\alpha_k, \beta)$$

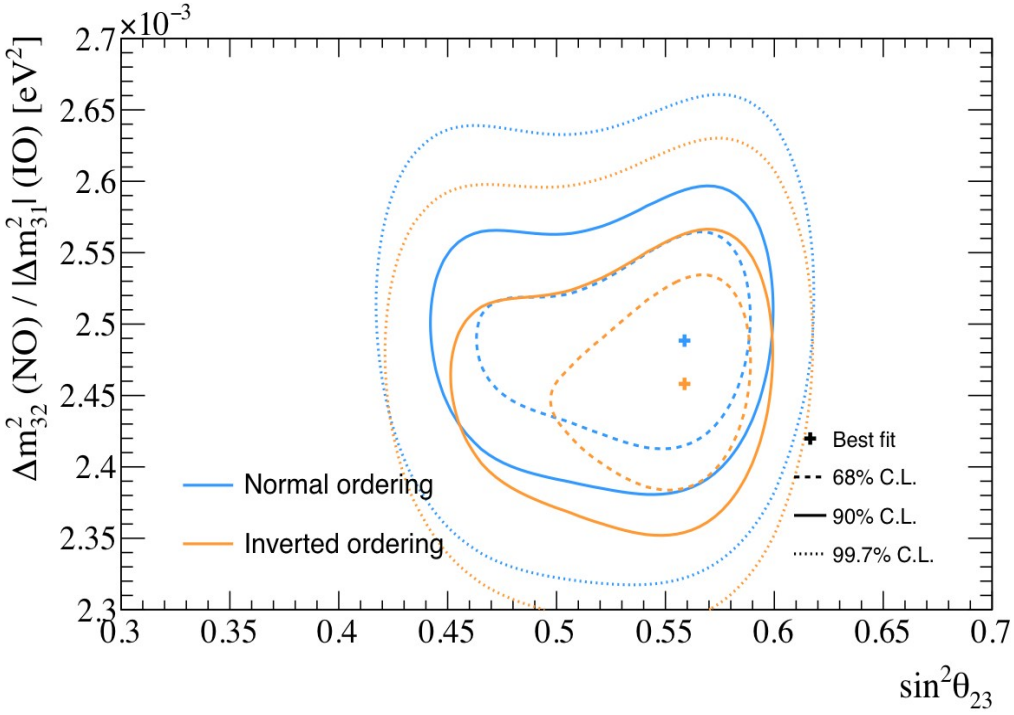
where α **are nuisances of flux and xsec strongly constrained by ND + nuisances on detector systematics (efficiency and purity)**

β = oscillation parameters. Described by standard oscillation formulas (PMNS)

$$P(\nu_\alpha \to \nu_\beta) = \sin^2(2\theta)\sin^2\left(1.27\frac{\Delta m^2_{ji}[\mathrm{eV}^2]L[\mathrm{km}]}{E_\nu[\mathrm{GeV}]}\right)$$



Projection of 2D likelihood over 2 parameters of interest



Where did the nuisances α go?

The likelihood depends both on the parameters of interest you want to measure (PMNS parameters: β) and the nuisances parameters describing just systematics effects (flux, xsec, detector: α)

How to "project" the likelihood on β? Profiling or marginalizing on nuisance α

- When we minimize a likelihood, we can just add our nuisance parameters to the list of things to minimize

- Find a global minimum across all parameters

- Look at the variation of the parameter of interest at the best estimate of the nuisance parameters

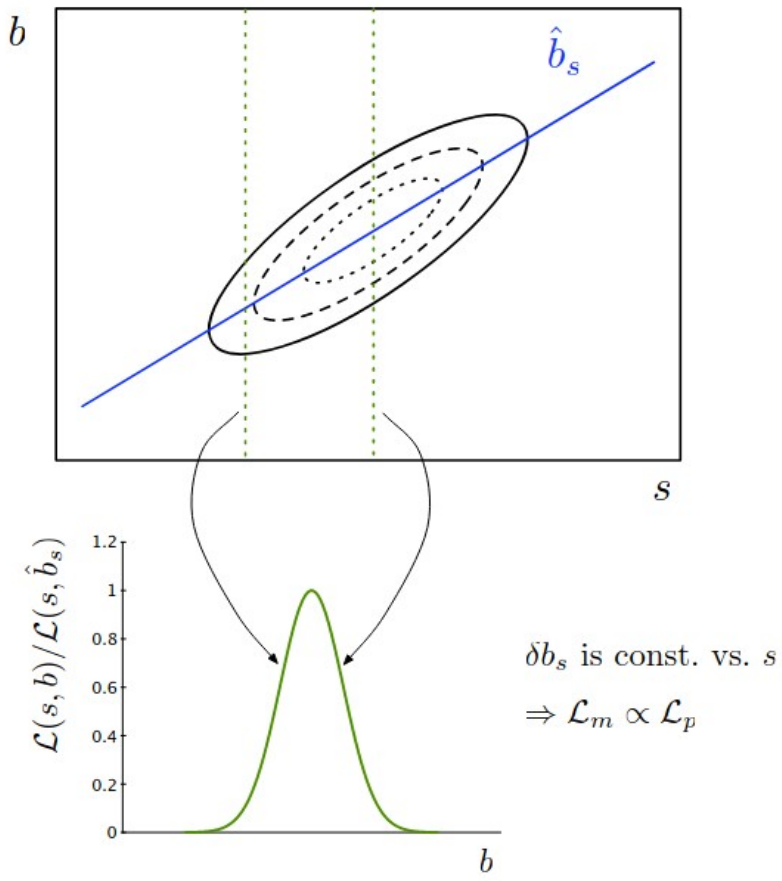
$$L(\beta)\cdot d\,\alpha \sim L(\alpha_{min},\beta)$$

Alternatively (Bayesian) we integrate (or marginalize) over the nuisance parameters

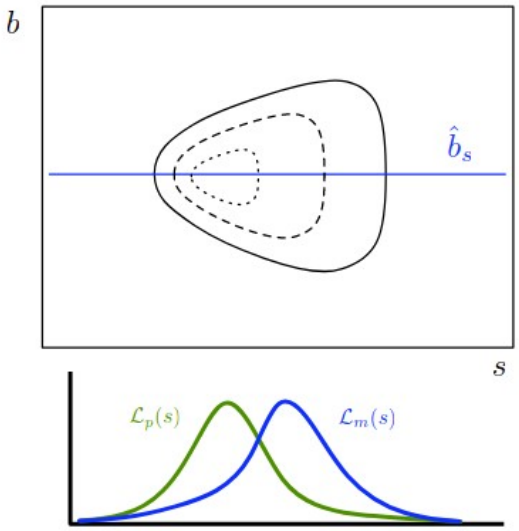
$$L(\beta)=\int L(\alpha,\beta)\,d\,\alpha$$

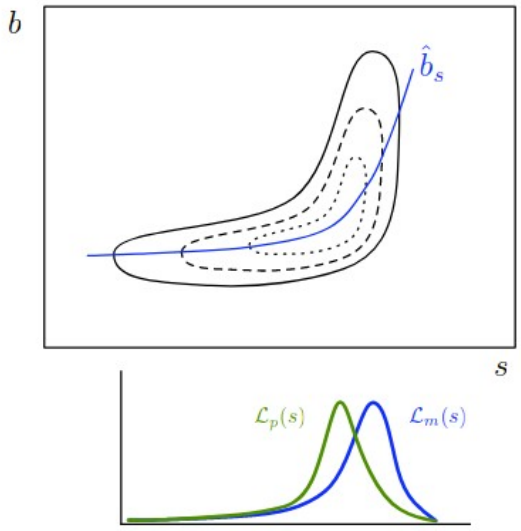
Profiling vs marginalization?

Profiling ~ marginalization, if error on β ~ constant over α nuisances



If error on POI β changes with α values and/or non linear correlation then **results can be widely different!**





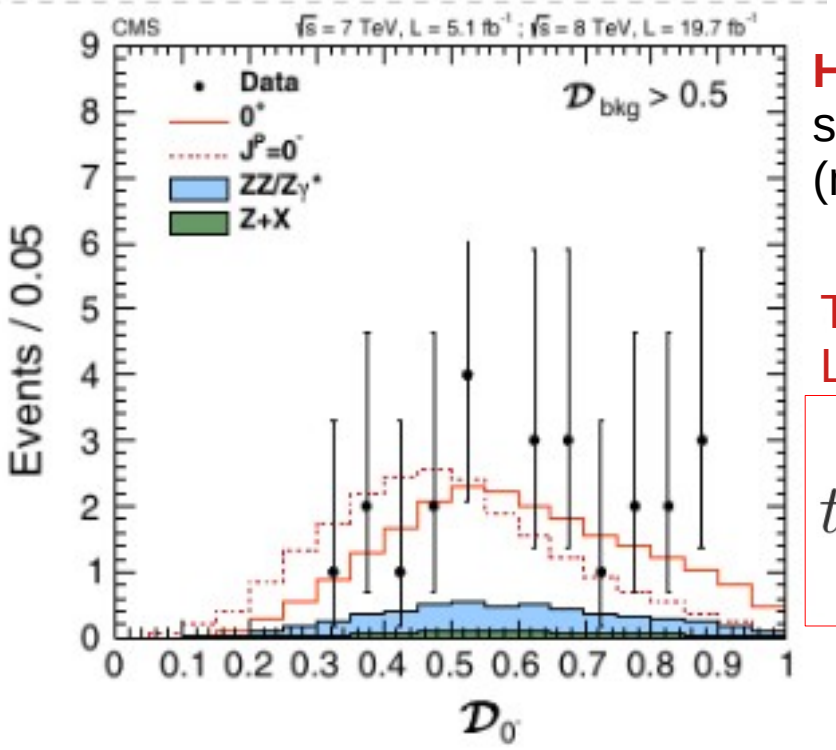
Practical example 2

Higgs spin-parity:

fit to likelihood of newly discovered “Higgs-like” resonance decay kinematics to do hypothesis testing on spin/parity of the resonance

Practical example: Higgs spin-parity

Collapse all the information on the kinematics of the final state Higgs decay in a single discriminant (you do not need to know how is built, consider it as an observable)



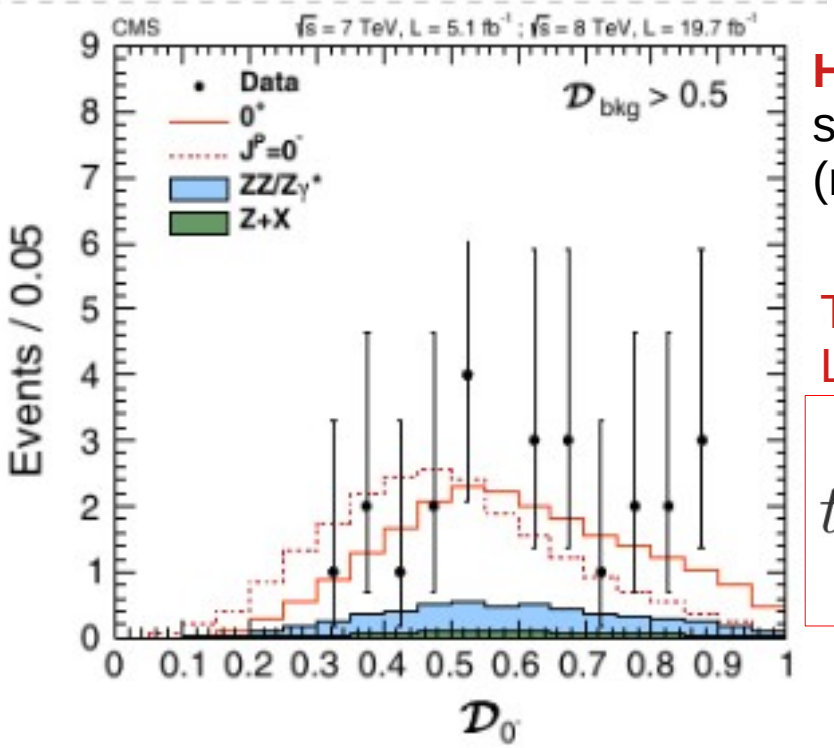
Hypothesis testing: is your observed data indicating a scalar Higgs (as in Standard Model) or a pseudoscalar (negative parity)?

The best "test" statistics is given (Neyman-Person Lemma) by

$$t(x) = \frac{P(x|H_1)}{P(x|H_0)}$$

Practical example: Higgs spin-parity

Collapse all the information on the kinematics of the final state Higgs decay in a single discriminant (you do not need to know how is built, consider it as an observable)



Hypothesis testing: is your observed data indicating a scalar Higgs (as in Standard Model) or a pseudoscalar (negative parity)?

The best "test" statistics is given (Neyman-Person Lemma) by

$$t(x) = \frac{P(x|H_1)}{P(x|H_0)} \qquad q = -2\ln\left(L^{alt}/L^{SM}\right)$$

P(x|H1, H0) = probability of observing data given alternative hypothesis 1 (alternative = 0-) or given hypothesis 0 (baseline = Standard Model)

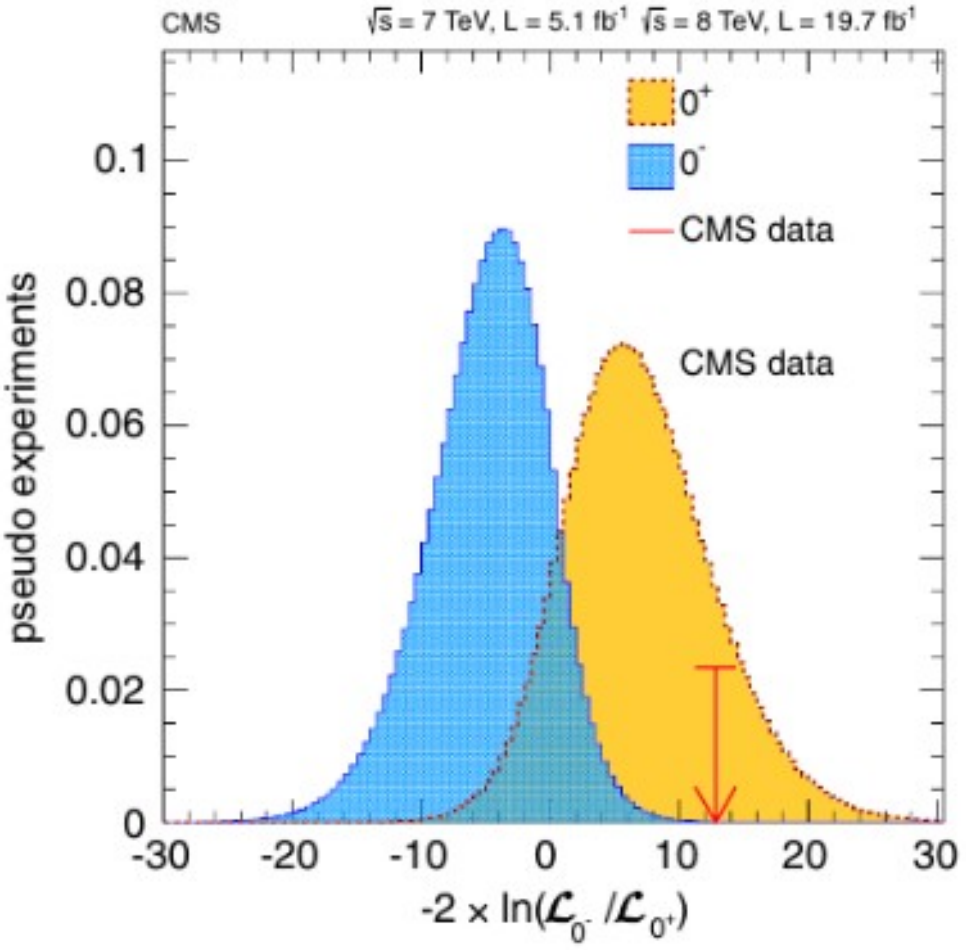
= these are the likelihood !

$$L^{SM}\left(N_j^{data}; N_j^{simu\,SM}\cdot f\left(\alpha_k\right)\right)$$

$$L^{alt}\left(N_j^{data}; N_j^{simu\,alt}\cdot f\left(\alpha_k\right)\right)$$

Practical example: Higgs spin-parity

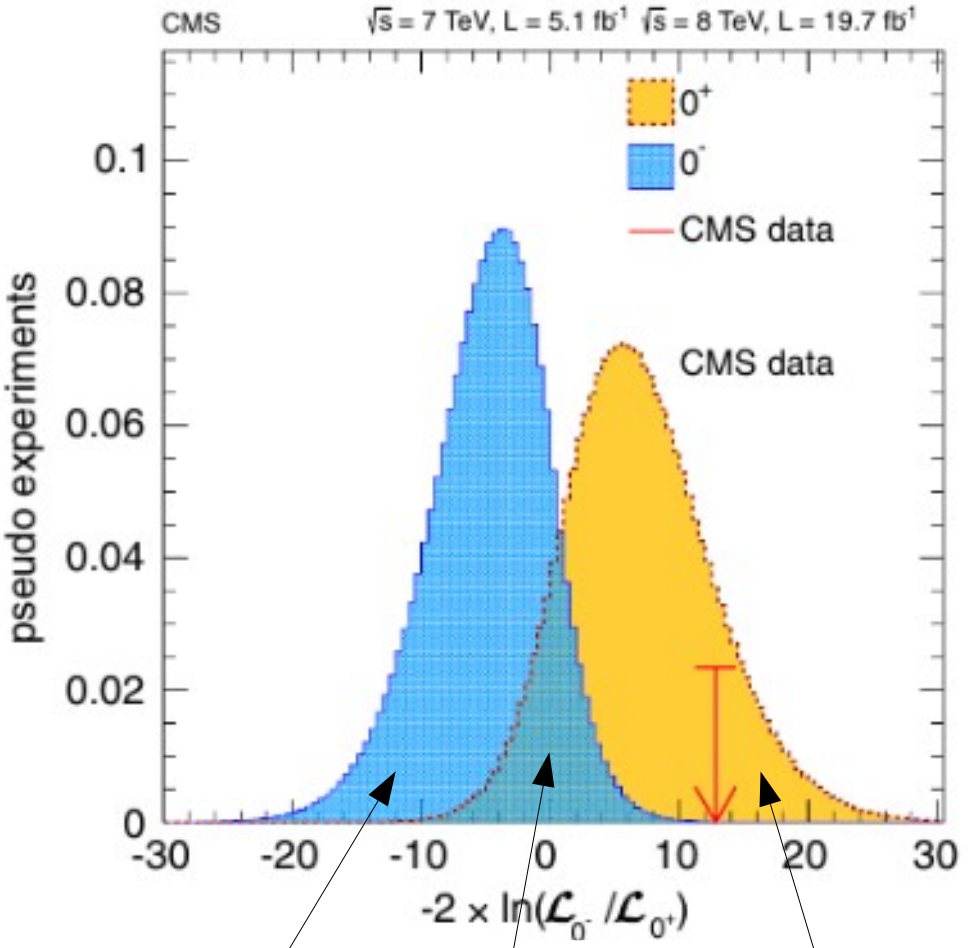
Expected distribution of test statistics q over many toys in Monte Carlo varying the systematic uncertainties (nuisances)



$$q = -2\ln\left(L^{alt}/L^{SM}\right)$$

Practical example: Higgs spin-parity

Expected distribution of test statistics q over many toys in Monte Carlo varying the systematic uncertainties (nuisances)



$$q = -2\ln\left(L^{alt}/L^{SM}\right)$$

Value of test statistics from a likelihood fit to data is SM-like. By how much?

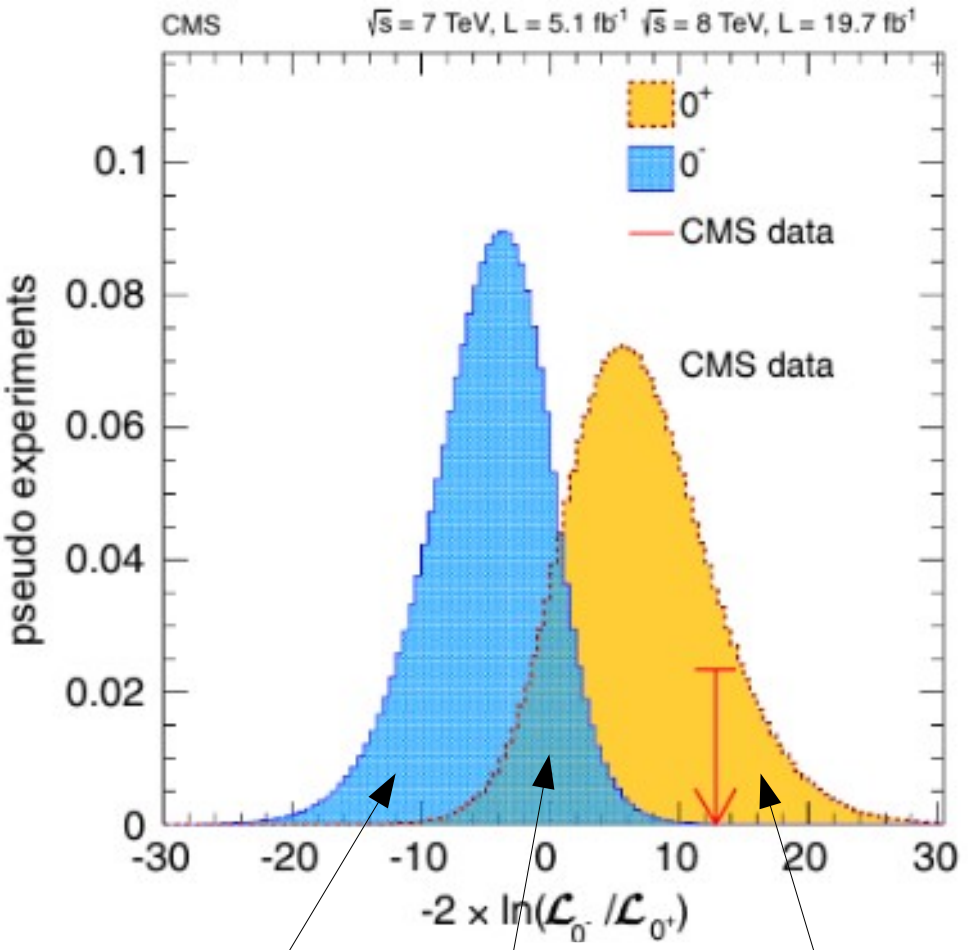
Region where data are BSM-like (0-)

Region where you cannot decide

Region where data are SM-like

Practical example: Higgs spin-parity

Expected distribution of test statistics q over many toys in Monte Carlo varying the systematic uncertainties (nuisances)



$$q = -2\ln\left(L^{alt}/L^{SM}\right)$$

Value of test statistics from a likelihood fit to data is SM-like. By how much?

$$\frac{P\left(q \geqslant q_{obs}; H_{alt}\right)}{P\left(q \geqslant q_{obs}; H_{SM}\right)} < \alpha$$

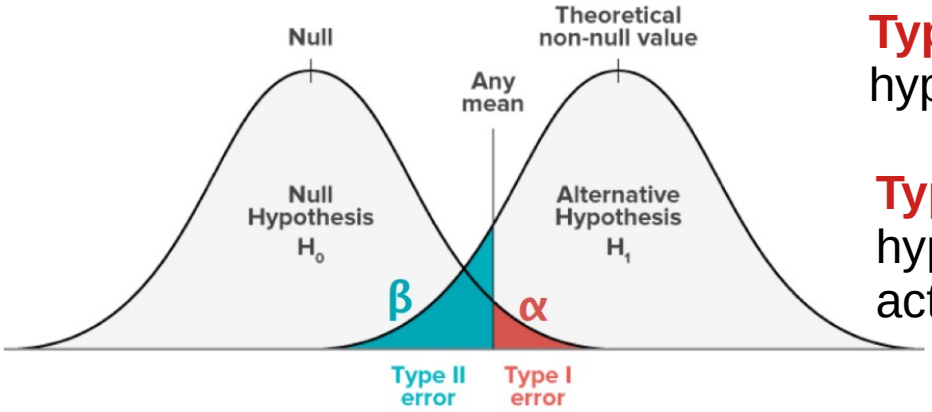
alternative signal hypotheses is excluded or not with a given confidence level (1 − α).
→ alternative H: 0- excluded at 99.9%

Region where data are BSM-like (0-)

Region where you cannot decide

Region where data are SM-like

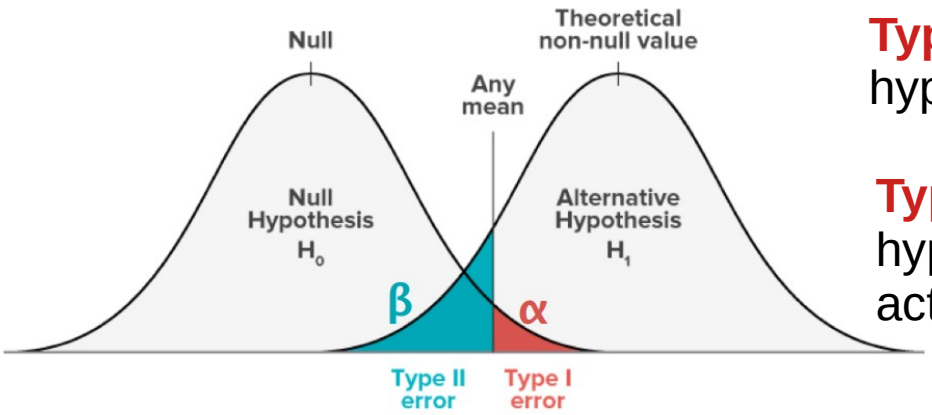
Hypothesis testing



Type I error: reject baseline (null) hypothesis when it is true (α)

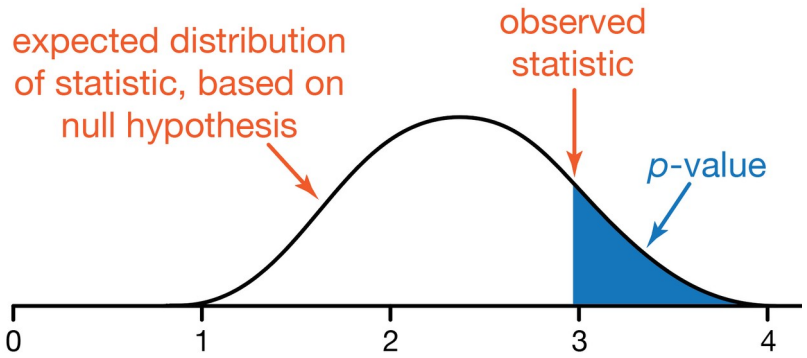
Type II error: fail to reject baseline (null) hypothesis when the alternative hypothesis is actually true (β)

Hypothesis testing



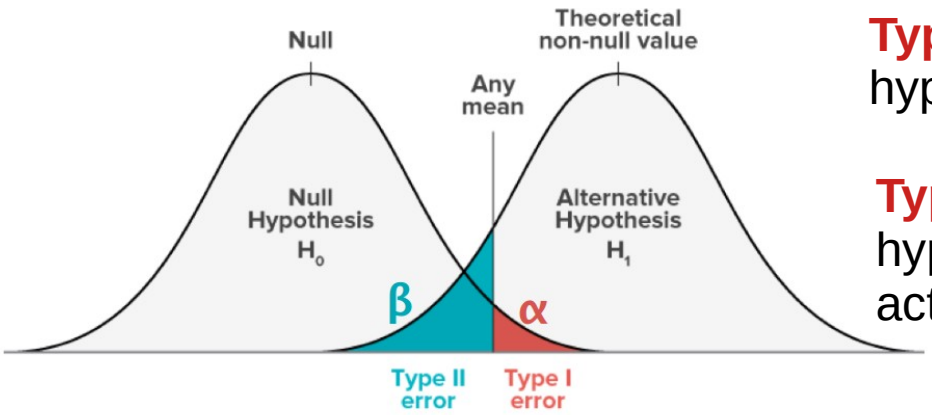
Type I error: reject baseline (null) hypothesis when it is true (α)

Type II error: fail to reject baseline (null) hypothesis when the alternative hypothesis is actually true (β)



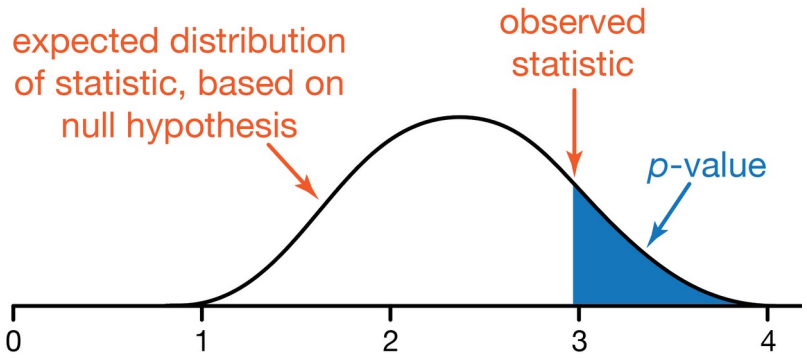
P-value: probability, assuming H, to observe data with equal or lesser compatibility with H relative to the data we got.

Hypothesis testing

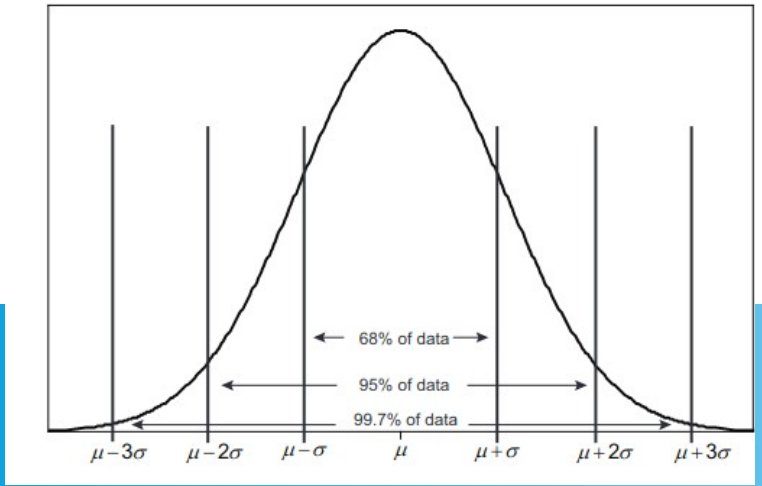


Type I error: reject baseline (null) hypothesis when it is true (α)

Type II error: fail to reject baseline (null) hypothesis when the alternative hypothesis is actually true (β)



P-value: probability, assuming H, to observe data with equal or lesser compatibility with H relative to the data we got.



We often use the **Gaussian distribution as an intuitive "metric"**

Even if our test statistics does not have a Gaussian distribution,

often we translate the "confidence level" or a "p-value" as the number of standard deviation that a Gaussian variable would fluctuate in one direction to give the same p-value
(5σ discovery ~ 3×10^{-7} probability)

Frequentist vs Bayesian

The likelihood is the probability of observing data, given a certain hypothesis

$$L\left(N_{data}; N_{simu} \cdot f\left(\alpha\right)\right) = P\left(data | H\left(\alpha\right)\right)$$

often used as probability of hypothesis given data but it is not correct. **What we would like is the posterior PDF of H(α) = probability of H (or a value) given the data**

Frequentist vs Bayesian

The likelihood is the probability of observing data, given a certain hypothesis

$$L\left(N_{data}; N_{simu}\cdot f\left(\alpha\right)\right)=P\left(data|H\left(\alpha\right)\right)$$

often used as probability of hypothesis given data but it is not correct. **What we would like is the posterior PDF of H(α) = probability of H (or a value) given the data**

$$P\left(H\left(\alpha\right)|data\right)=\frac{P\left(data|H\left(\alpha\right)\right)\cdot P\left(H\left(\alpha\right)\right)}{P\left(data\right)}$$

This is our frequentist likelihood

prior probability of the data: since this doesn t depend on α it is essentially a normalisation constant

prior probability of α , i.e. encompassing our knowledge of α before the measurement

Frequentist vs Bayesian

The likelihood is the probability of observing data, given a certain hypothesis

$$L\left(N_{data}; N_{simu}\cdot f\left(\alpha\right)\right)=P\left(data|H\left(\alpha\right)\right)$$

often used as probability of hypothesis given data but it is not correct. **What we would like is the posterior PDF of H(α) = probability of H (or a value) given the data**

$$P\left(H\left(\alpha\right)|data\right)=\frac{P\left(data|H\left(\alpha\right)\right)\cdot P\left(H\left(\alpha\right)\right)}{P\left(data\right)}$$

This is our frequentist likelihood

prior probability of the data: since this doesn't depend on α it is essentially a normalisation constant

prior probability of α , i.e. encompassing our knowledge of α before the measurement

There is some arbitrariness on how to chose the functional form the prior P(H(α))

A good experiment (with large sensitivity to H(α)) does not depend on the choice of the prior … but then it means that you can choose a flat prior P(H(α))~constant so:

$$P\left(H\left(\alpha\right)|data\right)\propto P\left(data|H\left(\alpha\right)\right)$$