Frequentist statistics methods and techniques used in the T2K oscillation results

Denis Carabadjac
dcarabadjac@llr.in2p3.fr

LLR

l'X ÉCOLE POLYTECHNIQUE IP PARIS

LLR group meeting

28.02.2024

université PARIS-SACLAY

cea

- The goal of the talk is to explain in a detailed way which statistics methods and techniques are used in T2K Far Detector Oscillation Analysis. How they are used and why

- Everybody heard about Feldman-Cousins method, Wilks' theorem, Neyman construction but not everybody understands what they really mean: the goal of the talk is to explain fundamentally what mean these concepts and why they are important

- In this lecture I tried to find balance between strict mathematical statements and clear physics interpretations. Hopefully, I could achieve this

- My target was to make it clear for 1^{st} year PhD student

Additional comments:

- Understanding that the topic is hard I will repeat important points many times during the slides
- The best time to ask questions is the end of the subsection because I try to answer main possible questions during the subsection. But if it very unclear interrupt me at any time


STAY STRONG

Part 0:
Neutrino oscillation probability

Electron neutrino appearance in muon neutrino flux

$$
\begin{aligned}
P\left(\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e\right) \approx\; & \sin^2\theta_{23}\frac{\sin^2 2\theta_{13}}{(A-1)^2}\sin^2[(A-1)\Delta_{31}] \\
\overset{-}{(+)}\; & \alpha\frac{J_0 \sin\delta_{CP}}{A(1-A)}\sin\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}] \\
+\; & \alpha\frac{J_0 \cos\delta_{CP}}{A(1-A)}\cos\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}] \\
+\; & \alpha^2\cos^2\theta_{23}\frac{\sin^2 2\theta_{12}}{A^2}\sin^2(A\Delta_{31})
\end{aligned}
$$

- The leading term
- $\sim\theta_{13}$ (RC is important)
- $\sim\sin^2\theta_{23}$(allows to break octant degeneracy)
- $\sim\sin^2\Delta_{31}$ (not sensitive to sign of Δm^2_{31})

$$
\begin{aligned}
\alpha &= \Delta m^2_{21}/\Delta m^2_{31} \\
\Delta_{ij} &= \Delta m^2_{ij}L/4E \\
A &= (-)2\sqrt{2}G_F n_e E/\Delta m^2_{31} \\
J_0 &= \sin 2\theta_{12}\sin 2\theta_{13}\sin 2\theta_{23}\cos\theta_{13}.
\end{aligned}
$$

Part 0: Neutrino oscillation probability

Electron neutrino appearance in muon neutrino flux

$$P\left(\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e\right) \approx \sin^2\theta_{23}\frac{\sin^2 2\theta_{13}}{(A-1)^2}\sin^2[(A-1)\Delta_{31}]$$
$$\overset{-}{(+)}\ \alpha\frac{J_0\sin\delta_{CP}}{A(1-A)}\sin\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}]$$
$$+\ \alpha\frac{J_0\cos\delta_{CP}}{A(1-A)}\cos\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}]$$
$$+\ \alpha^2\cos^2\theta_{23}\frac{\sin^2 2\theta_{12}}{A^2}\sin^2(A\Delta_{31})$$

- CP-odd term (asymmetric for ν and $\bar{\nu}$ oscillations)
- Modulates probability with respect to first leading term
- The term is still "not small" as it is $\sim\sin\Delta_{31}\sim 1$ (30% of first term for $\sin\delta_{CP}$=1)
- $\sim\sin\delta_{CP}$ (sensitive to $\sin\delta_{CP}$)

$$\alpha = \Delta m^2_{21}/\Delta m^2_{31}$$
$$\Delta_{ij} = \Delta m^2_{ij}L/4E$$
$$A = (-)2\sqrt{2}G_F n_e E/\Delta m^2_{31}$$
$$J_0 = \sin 2\theta_{12}\sin 2\theta_{13}\sin 2\theta_{23}\cos\theta_{13}.$$

Part 0: Neutrino oscillation probability

$$P\left(\overset{(-)}{\nu}_\mu \rightarrow \overset{(-)}{\nu}_e\right) \approx \sin^2\theta_{23}\frac{\sin^2 2\theta_{13}}{(A-1)^2}\sin^2[(A-1)\Delta_{31}]$$

$$\overset{-}{(+)}\ \alpha\frac{J_0 \sin\delta_{CP}}{A(1-A)}\sin\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}]$$

$$+\ \alpha\frac{J_0 \cos\delta_{CP}}{A(1-A)}\cos\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}]$$

$$+\ \alpha^2\cos^2\theta_{23}\frac{\sin^2 2\theta_{12}}{A^2}\sin^2(A\Delta_{31})$$

- CP-even term
- Effects precision measurements of δ_{CP} in case it is near to $\pm\frac{\pi}{2}$
- BUT it is smaller that CP-even term as it depends on $\cos\Delta_{31} \sim 0$-> Not so sensitive to $\cos\delta_{CP}$

Part 0: Neutrino oscillation probability

$$
\begin{aligned}
P\left(\overset{(-)}{\nu}_\mu \rightarrow \overset{(-)}{\nu}_e\right) \approx\ & \sin^2\theta_{23}\frac{\sin^2 2\theta_{13}}{(A-1)^2}\sin^2[(A-1)\Delta_{31}] \\
& \overset{-}{(+)}\ \alpha\frac{J_0\sin\delta_{CP}}{A(1-A)}\sin\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}] \\
& +\ \alpha\frac{J_0\cos\delta_{CP}}{A(1-A)}\cos\Delta_{31}\sin(A\Delta_{31})\sin[(1-A)\Delta_{31}] \\
& +\ \alpha^2\cos^2\theta_{23}\frac{\sin^2 2\theta_{12}}{A^2}\sin^2(A\Delta_{31})
\end{aligned}
$$

Conclusions:

1) We have sensitivity to δ_{CP}, more precisely speaking to $\sin\delta_{CP}$ in appearance channel
2) Unknown MO causes degeneracy (but for T2K it is not so dramatic as baseline is not very long (295 km))
3) Sensitivity to MO is small

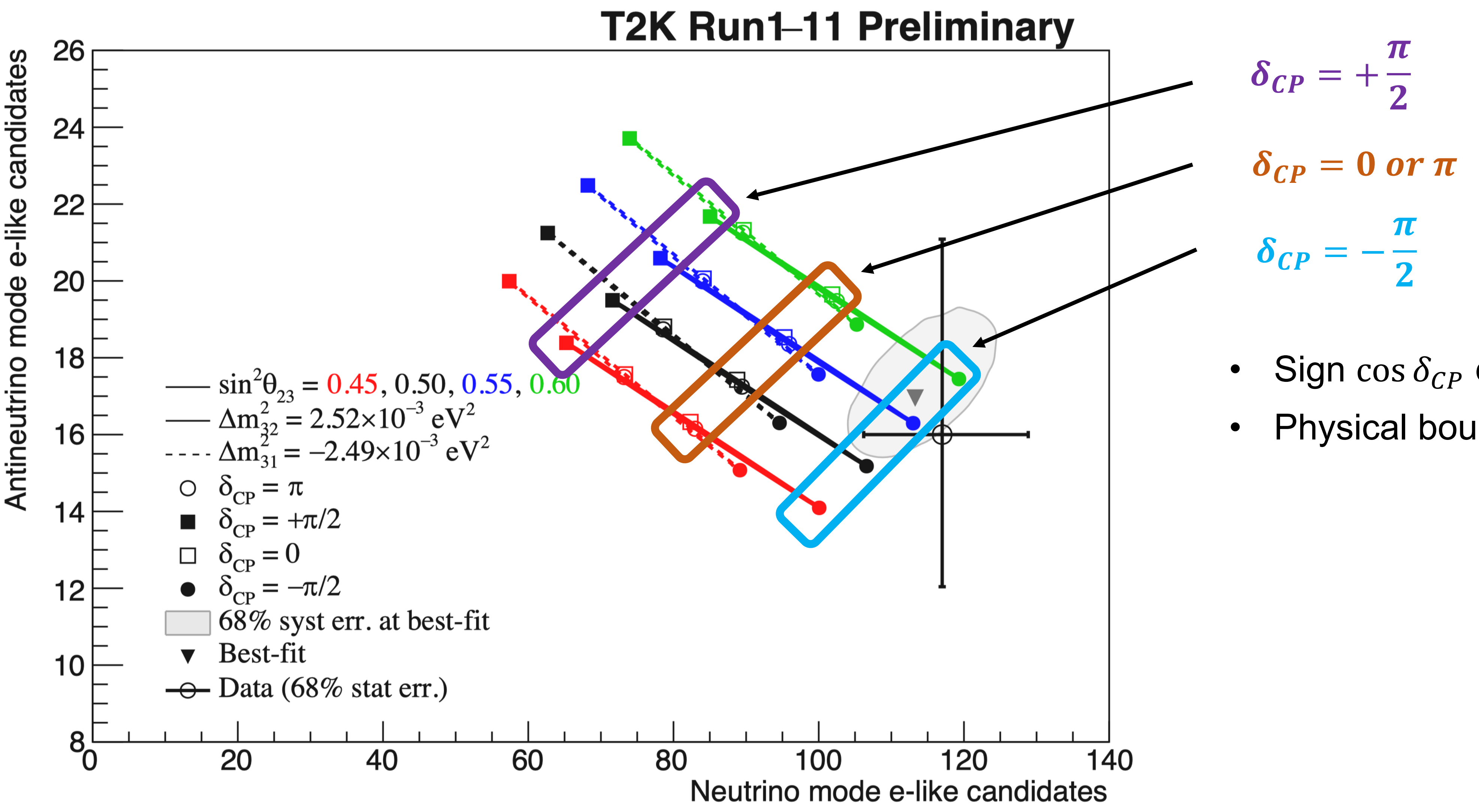
Part 0: Neutrino oscillation probability

Considering also disappearance channel we understand that T2K is sensitive to

δ_{CP}, Δm^2_{32}, θ_{23}, θ_{13}, (MO)

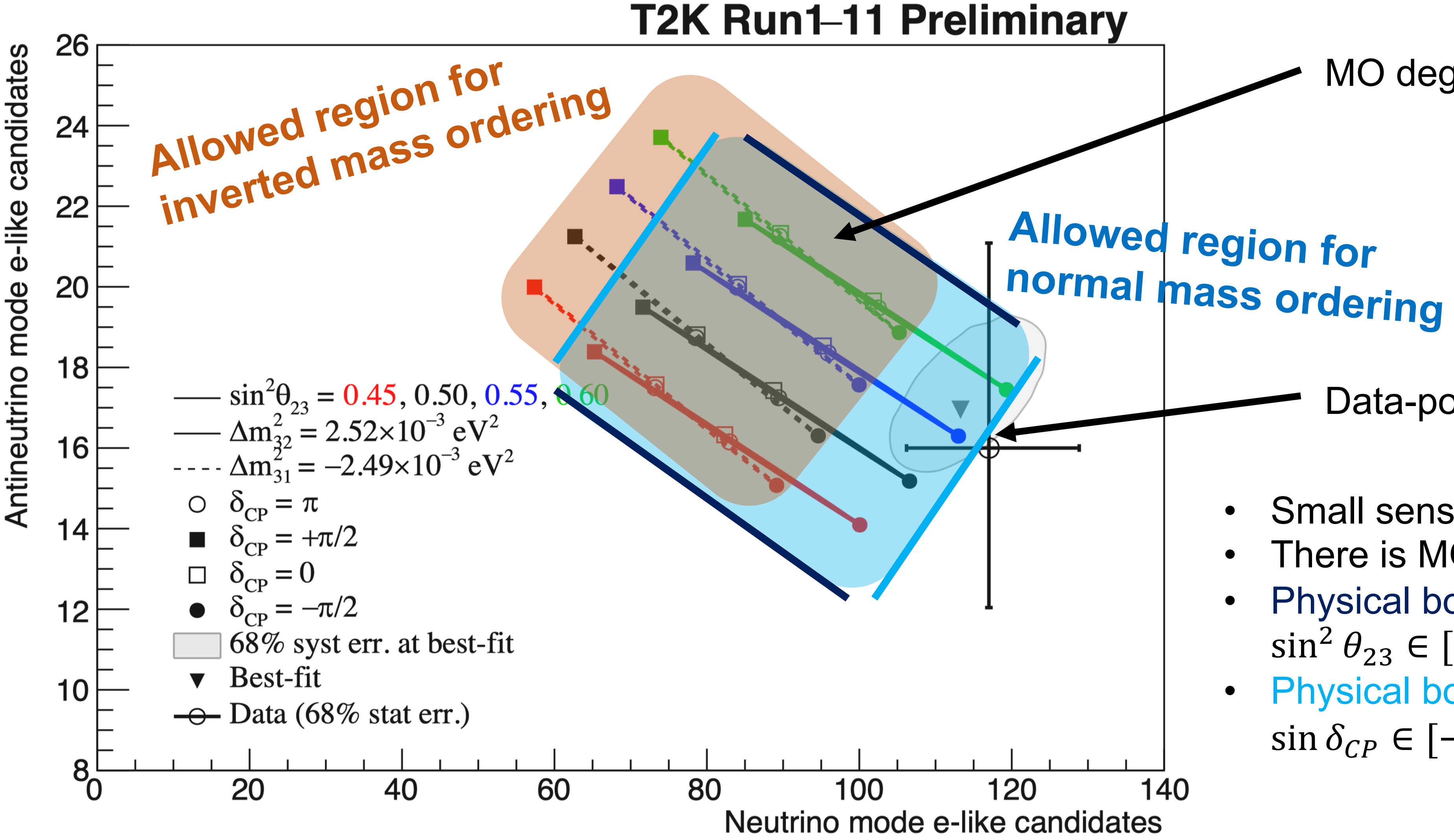
Part I:
Biprobability plots

Link from Oscillation probability to Likelihood fit



- Sign $\cos\delta_{CP}$ degeneracy
- Physical boundaries

Part II: Statistical inference of the parameters



- Small sensitivity to mass ordering
- There is MO-δ_{CP} degeneracy
- Physical boundaries (dark blue) as $\sin^2\theta_{23} \in [0; 1]$
- Physical boundaries (light blue) as $\sin\delta_{CP} \in [-1; 1]$

T2K oscillation analysis

Hadron Production Data

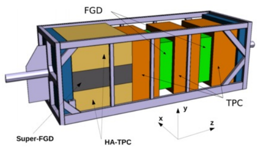
Ingrid/Monitoring Data

1. Flux model Φ

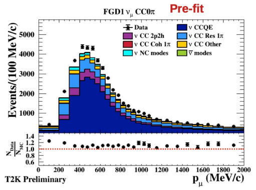
External and Internal cross section Data

2. Cross section model σ

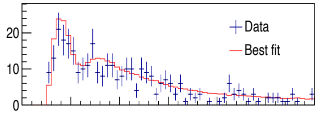
ND280 detector model



ND280 Data

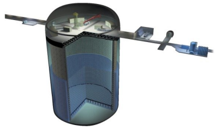


Near detector fit (to constraint systematics)



1 & 2

SuperK Data



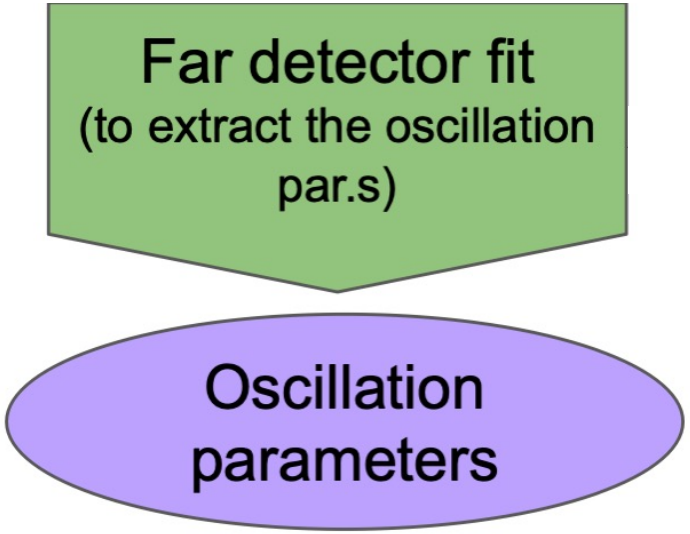
Far detector fit (to extract the oscillation par.s)

SuperK detector model

Oscillation parameters

Complex structure
Challenging task

Focusing on this part



The story starts now…

The story starts now…

Part II: Introduction to the frequentist fit: likelihood definition

Part II: Introduction to the frequentist fit

I. As we perform frequentist fit: need to build the **likelihood** function.

- We don’t want to perform fit event by event as it is CPU consuming → Instead use **binned data**
- Binned data → The **random variables** $\boldsymbol{x_i = N_i^{obs}}$ – are the number of observed events in the bin $\boldsymbol{i}$
  $\boldsymbol{N_i^{obs}} \sim \text{Poisson}(\boldsymbol{N_i^{exp}})$, where $\boldsymbol{N_i^{obs}}$ is the predicted or expected number of events in the bin $\boldsymbol{i}$
- We need to construct **sophisticated model** to make accurate predictions $\boldsymbol{N_i^{exp}}$
  → **physics model:** describes the **neutrino oscillation phenomenon**
  → **systematics model:** describes the relevant effects except neutrino oscillation

$$N_{exp}^{\nu_\alpha}(E_\nu^{true}) = \Phi(E_\nu^{true}) \otimes \sigma(E_\nu^{true}) \otimes \epsilon(E_\nu^{true}) \otimes S(E_\nu^{true}, E_\nu^{reco}) \otimes \boldsymbol{P_{\nu_\mu \to \nu_\alpha}(E_\nu^{true}, \vec{o})}$$

$\alpha = e, \mu$

Neutrino flux | Interection cross-section | Detector efficiency | Energy smearing matrix — Systematics

Oscillation probability — We want to extract this!

Part I: Introduction: definitions and notations

$$-\ln L = \sum_{s,i} \left[N_{s,i}^{exp}(\boldsymbol{o}, \boldsymbol{f}) - N_{s,i}^{obs} + N_{obs} \ln \frac{N_{s,i}^{obs}}{N_{s,i}^{exp}(\boldsymbol{o}, \boldsymbol{f})} \right] + (\boldsymbol{f} - \boldsymbol{f_0})^T V^{-1} (\boldsymbol{f} - \boldsymbol{f_0})$$

Sample likelihood

Penalty term for systematics

$N_{s,i}^{exp}(\boldsymbol{o}, \boldsymbol{f})$ –number of expected events in bin i and sample s

V – prior covariance matrix on systematics

$N_{s,i}^{obs}$ –observed number of events in bin i and sample s

$$L = f\left(\boldsymbol{o}, \boldsymbol{f} \middle| \boldsymbol{N^{obs}}\right), \qquad f: R^{170+5} \rightarrow R$$

- The likelihood function stores all the statistics knowledge of our model and experiment
- We need a sophisticated accurate manipulations of it to **make inference on parameters of interest** (oscillation parameters), **check the validity of our model etc**
- We should develop the techniques for this which will be practically realisable and accurare enough

Part III:

Statistical inference of the oscillation parameters

Part III: Statistical inference of the parameters

The model depends on the large number of parameters:

a) parameters of interest $\boldsymbol{\theta}$ (in our case 4 continuous (δ_{CP}, Δm^2_{32}, θ_{23}, θ_{13}) and 1 discrete - MO)
b) nuisance parameters which come from systematic model $-\boldsymbol{\eta}$ (170).

- How do find the best estimator of the parameters?

- How do we determine the uncertainties of these estimators?

Part III: Statistical inference of the parameters

Goal: Determine the best fit point

$$\boldsymbol{\xi} = (\boldsymbol{o}, \boldsymbol{f})$$

$$L(\boldsymbol{\xi} | \boldsymbol{N^{obs}}) \rightarrow max\colon\ L_{max} = L(\hat{\boldsymbol{\xi}} | \boldsymbol{N^{obs}})$$

$\hat{\boldsymbol{\xi}}$ **– our central value which will be quoted**

This is an easy part. No caveats

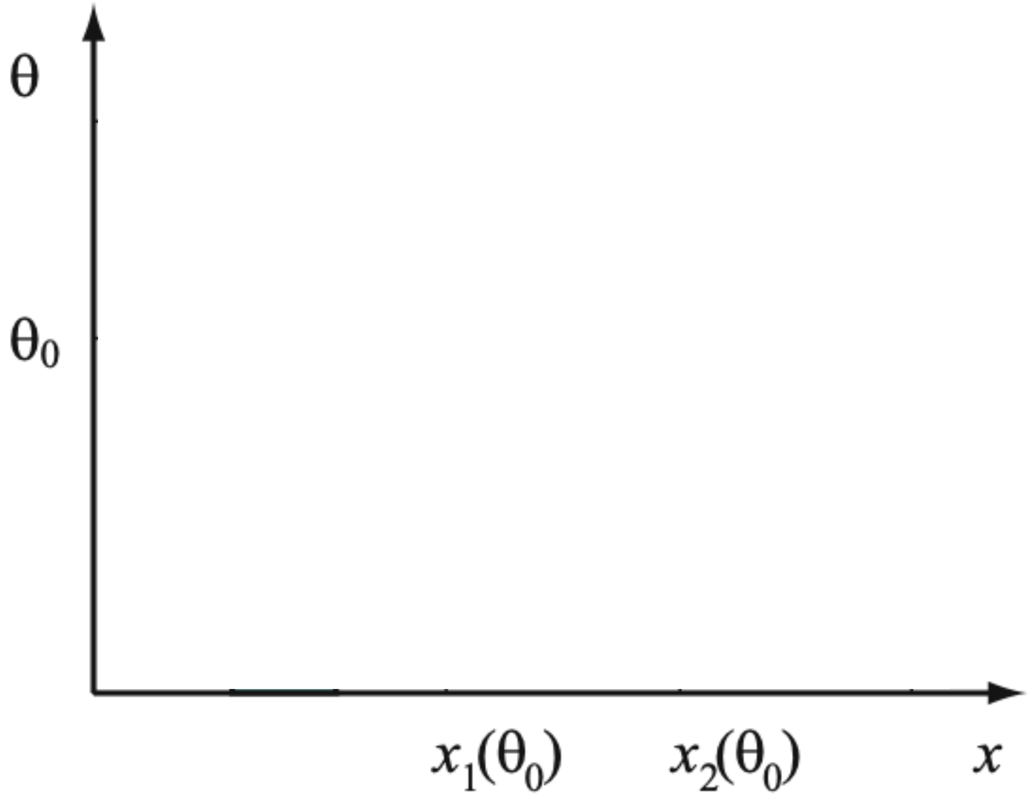
Goal: Determine the error of the best-fit point

Central technique – Neyman belt construction

Very important for understanding of next slides!

Part III: Stat. infer. / Neyman belt construction

- Let's consider a simple model with one observable x and one unknown parameter θ. The distribution of x for a fixed θ is given by $f(x|\theta)$ (likelihood in other words) .

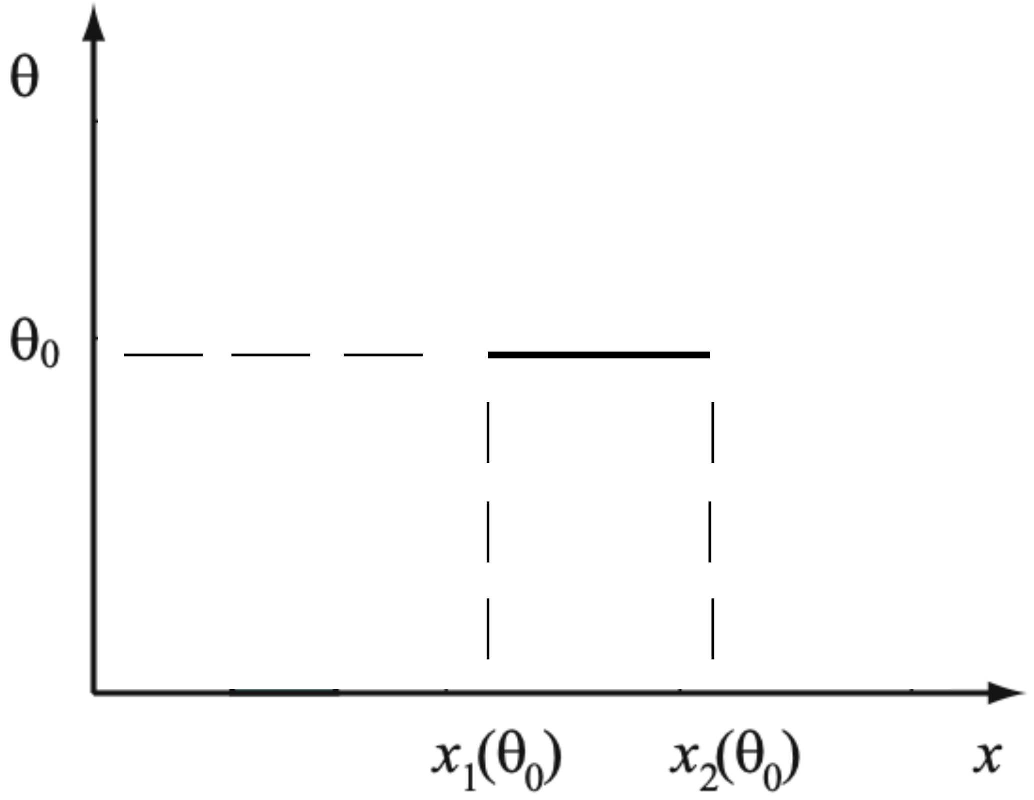


- Let's consider a simple model with one observable x and one unknown parameter θ. The distribution of x for a fixed θ is given by $f(x|\theta)$ (likelihood in other words) .

- Let's fix a true value to θ_0. Find interval $[x_1(\theta_0), x_2(\theta_0)]$ which gives probability to observe x inside it with probability β(**confidence level**):

$$\int_{x_1(\theta_0)}^{x_2(\theta_0)} f(x|\theta_0)\, dx = \beta$$

e.g. $\beta = 68\%(1\sigma), 99.7\%(3\sigma)$



Distr of x and show the deltax region

Part III: Stat. infer. / Neyman belt construction

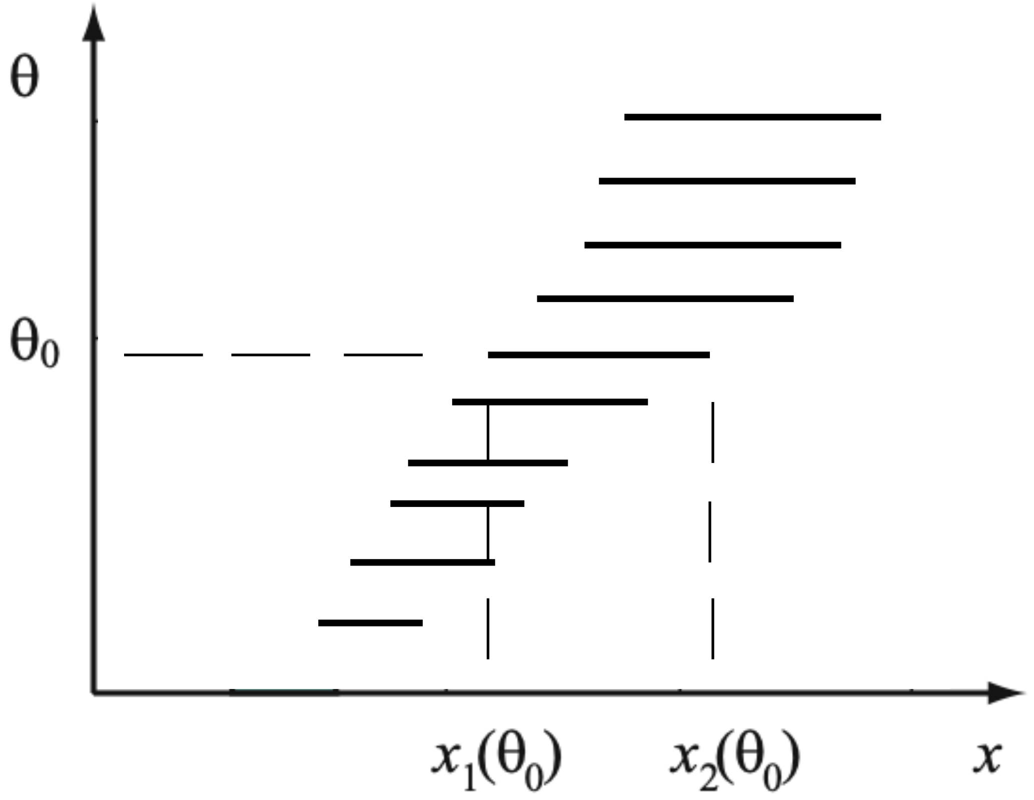
- Let's consider a simple model with one observable x and one unknown parameter θ. The distribution of x for a fixed θ is given by $f(x|\theta)$ (likelihood in other words) .

- Let's fix a true value to θ_0. Find interval $[x_1(\theta_0), x_2(\theta_0)]$ which gives probability to observe x inside it with probability β(**confidence level**):

$$\int_{x_1(\theta_0)}^{x_2(\theta_0)} f(x|\theta_0)\,dx = \beta$$

e.g. $\beta = 68\%(1\sigma)$, $99.7\%(3\sigma)$

- Repeat this for different true values of θ



Part III: Stat. infer. / Neyman belt construction

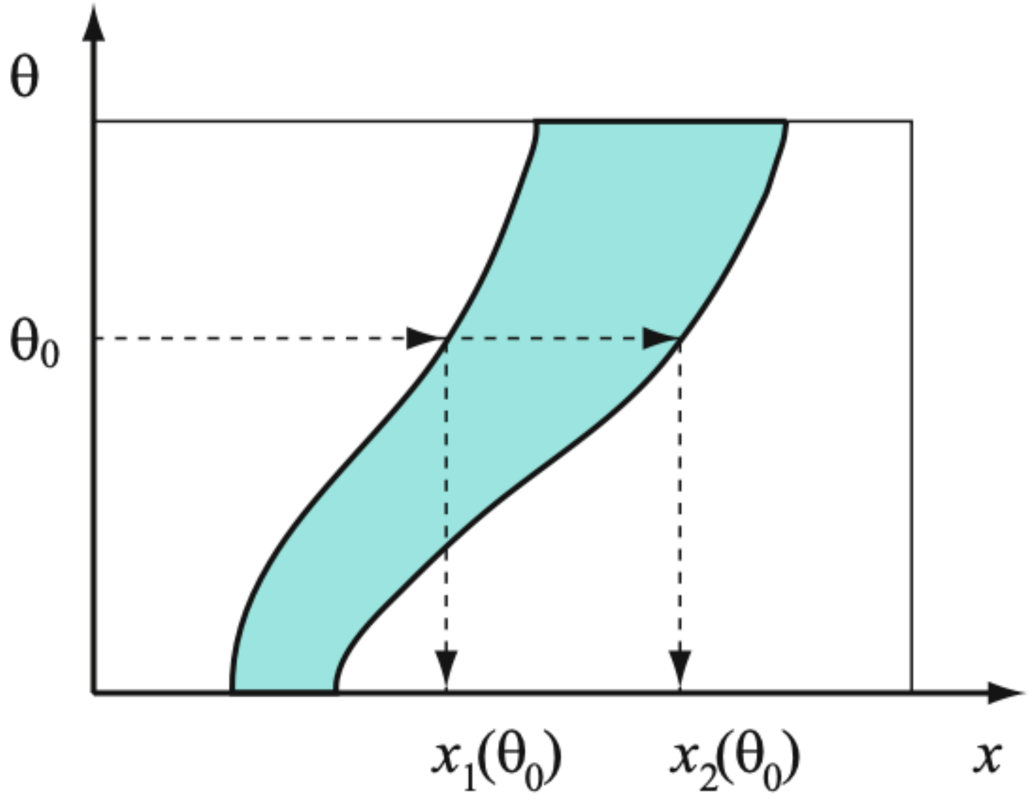
- Let's consider a simple model with one observable x and one unknown parameter θ. The distribution of x for a fixed θ is given by $f(x|\theta)$ (likelihood in other words) .

- Let's fix a true value to θ_0. Find interval $[x_1(\theta_0), x_2(\theta_0)]$ which gives probability to observe x inside it with probability β(**confidence level**):

$$\int_{x_1(\theta_0)}^{x_2(\theta_0)} f(x|\theta_0)\, dx = \beta$$

e.g. $\beta = 68\%(1\sigma)$, $99.7\%(3\sigma)$

- Repeat this for different true values of θ



Neyman belt is contructed

Part III: Stat. infer. / Neyman belt construction

- Let's consider a simple model with one observable x and one unknown parameter θ. The distribution of x for a fixed θ is given by $f(x|\theta)$ (likelihood in other words) .

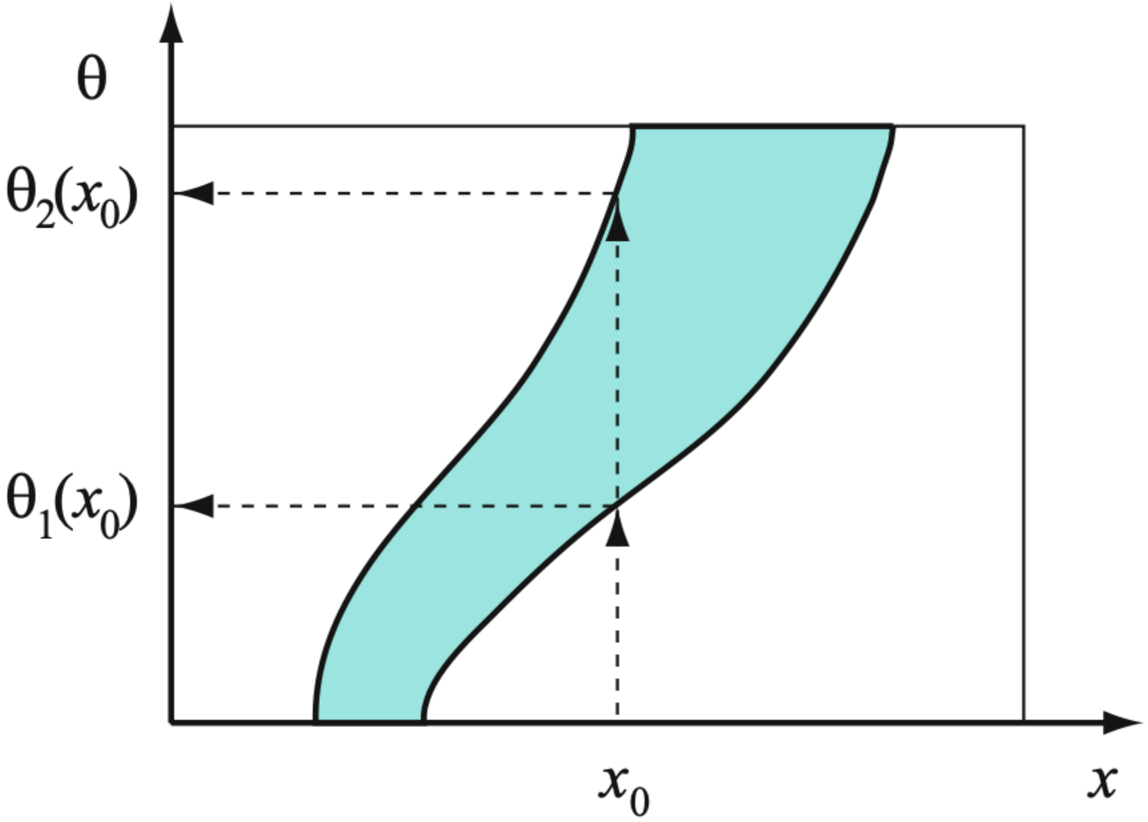
- Let's fix a true value to θ_0. Find interval $[x_1(\theta_0), x_2(\theta_0)]$ which gives probability to observe x inside it with probability β(**confidence level**):

$$\int_{x_1(\theta_0)}^{x_2(\theta_0)} f(x|\theta_0)\, dx = \beta$$

e.g. $\beta = 68\%(1\sigma), 99.7\%(3\sigma)$

- Repeat this for different true values of θ

- Invert the confidence belt: if the data result is x_0, the **confidence interval for θ is $[\theta_1(x_0), \theta_2(x_0)]$** (determined as intersections of vertical line with the belt)

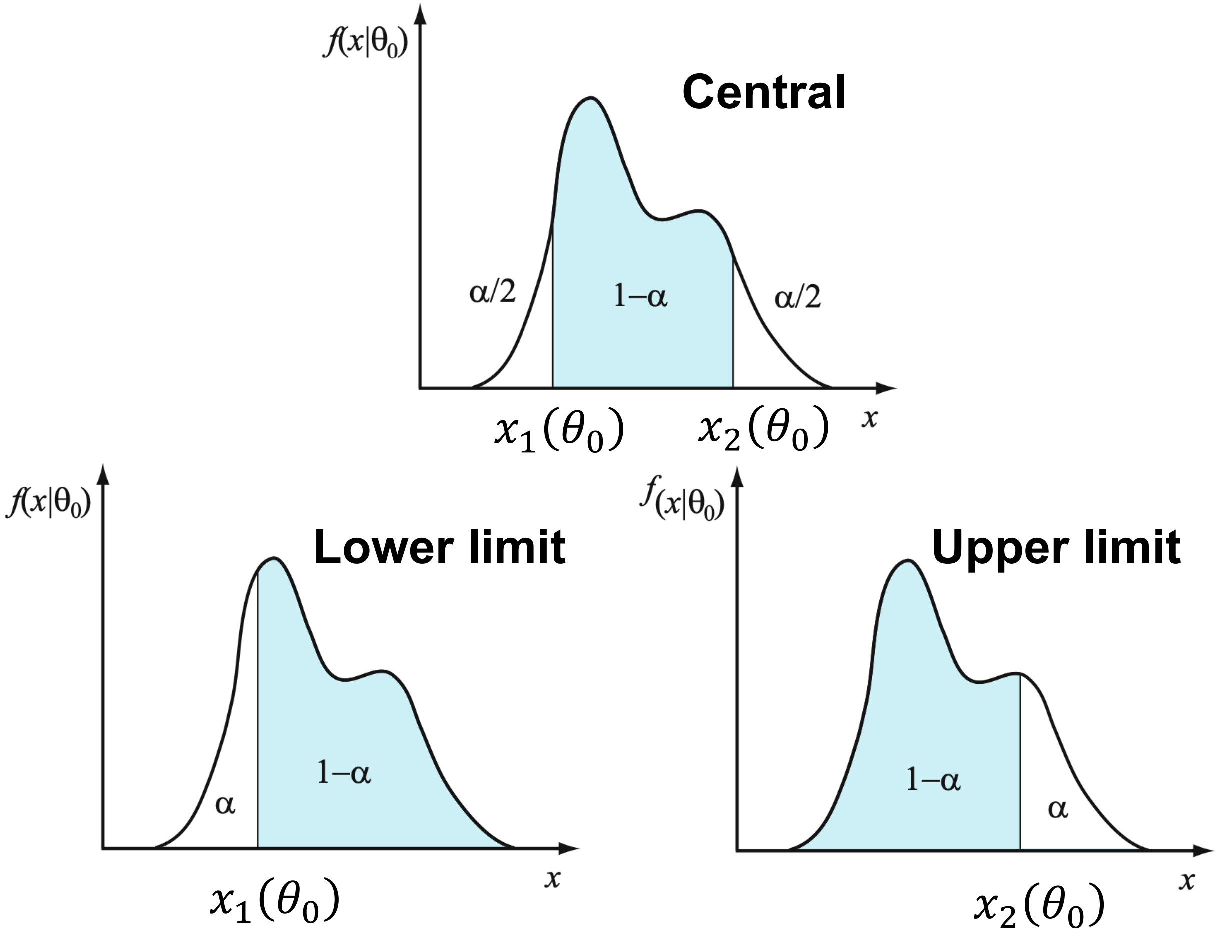


Later I will show some examples, don't worry!

Part III: Stat. infer. / Neyman belt construction

One significant problem of the belt – **the choice of $[x_1(\theta_0), x_2(\theta_0)]$ is arbitrary**

- The choice of the $[x_1(\theta_0), x_2(\theta_0)]$ in called **ordering rule**
- Thus, the confidence interval will depend on the chosen ordering rule
- Also in some cases a chosen ordering rule can give incorrect coverage (e.g. "flip-flopping problem")



Example #1 for better understanding

Finally!

We measure directly a quantity x. Let $x \sim N(\mu, \sigma^2)$, where σ^2 – is known.

$$\text{So, } L(x;\mu) \coloneqq f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

1) Contruct Neyman belt:

For a fixed μ and given β:

$$\int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \beta$$

$$\Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right) = \beta$$

This is equation on x_1, x_2!

Φ –CDF of stardard normal distribution

Let's choose **central confidence interval**. Then

$$\Phi\left(\frac{x_2-\mu}{\sigma}\right) = \frac{1-\beta}{2}, \qquad 1 - \Phi\left(\frac{x_2-\mu}{\sigma}\right) = \frac{1-\beta}{2}$$

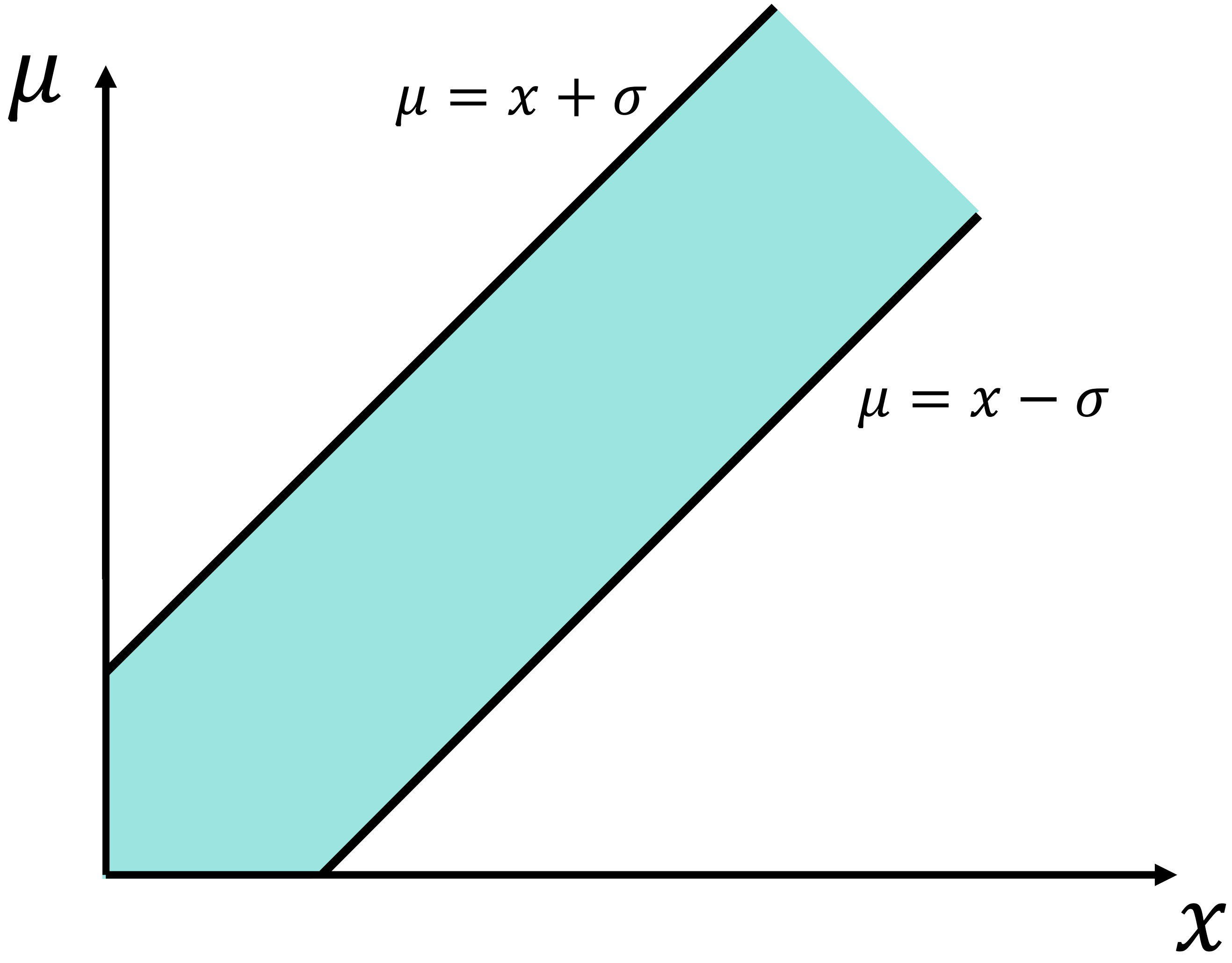
We now need to invert CDF. Let's take 68.27% C.L.. Then

$$x_1 = \mu - \sigma; \qquad x_2 = \mu + \sigma$$

Hooray! Got expected result

Part III: Stat. infer. / Neyman belt construction

$$x_1 = \mu - \sigma; \qquad x_2 = \mu + \sigma$$



Part III: Stat. infer. / Neyman belt construction

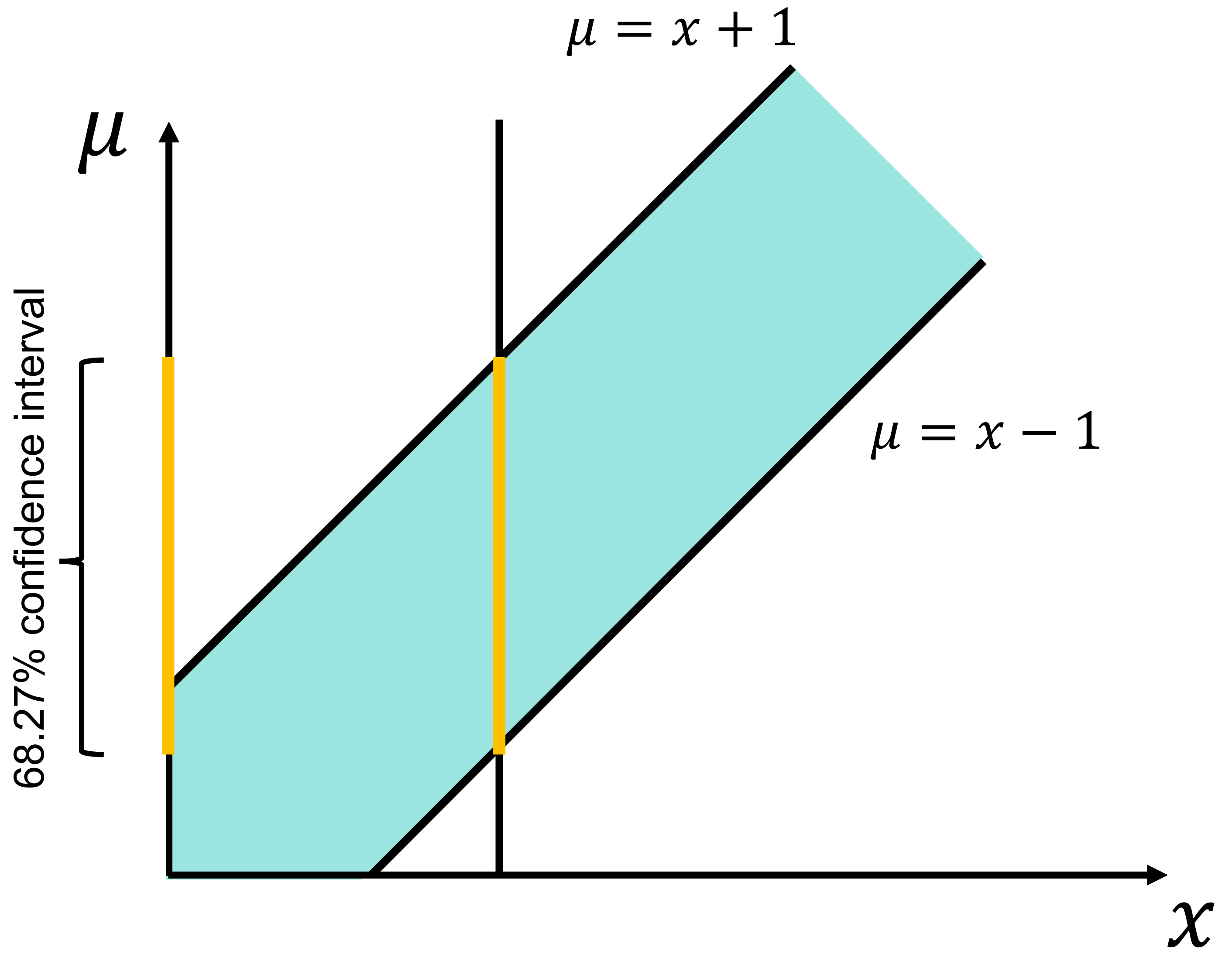
2) Neyman belt inversion

Let's D be our data measurement

Then according to this measurement μ 1σ C.I.:

C.I. = $[D-\sigma;\ D+\sigma]$

- Best-fit point obviosly is at D (can be shown by likelihood maximisation):
$$\hat{\mu} = D$$
- So, $\mu = D^{+\sigma}_{-\sigma}$
- The example is trivial but it illustrates how the Neyman construction works



But! What can be do if we have several measurements (if x is a vector but not a scalar)?

Answer: To use test statistic λ which is scalar

$$\boldsymbol{x} \longrightarrow \lambda(\boldsymbol{x})$$

vector → one number

Use $\lambda(\boldsymbol{x})$ for x-axis in Neyman belt construction

N.B.! We should know pdf for $\lambda(\boldsymbol{x})$, what can be a challenge

$$f(x|\theta) \rightarrow f(\lambda(\boldsymbol{x})|\theta)$$

↑ Usually we do not know analitically this function

Neyman belt construction

Example #2 for better understanding

We have several measurements: $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$. Each measurement $x_i \sim N(\mu, \sigma^2)$, where σ^2 – is known.
So,

$$L(x;\mu) := f(x|\mu) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

- Choose test statistic: **a) Sample mean:** $\lambda(\boldsymbol{x}) = \frac{1}{N}\sum_i x_i = \overline{x}$
- Test statistic distribution is known analitically: $\lambda \sim N(\mu, \frac{\sigma^2}{N})$
- Now we have the same problem as in example 1 for quantity $\lambda \sim N(\mu, \frac{\sigma^2}{N})$.
  So, take example 1, substitute $x \to \lambda = \overline{x}, \sigma \to \frac{\sigma}{\sqrt{N}}$

68.27% C.I. : $[\overline{D} - \frac{\sigma}{\sqrt{N}}, \overline{D} + \frac{\sigma}{\sqrt{N}}]$, where $\overline{D} = \frac{1}{N}\sum_i D_i$, D_i –outcomes of the direct measurements

Neyman belt construction

Example #2 for better understanding

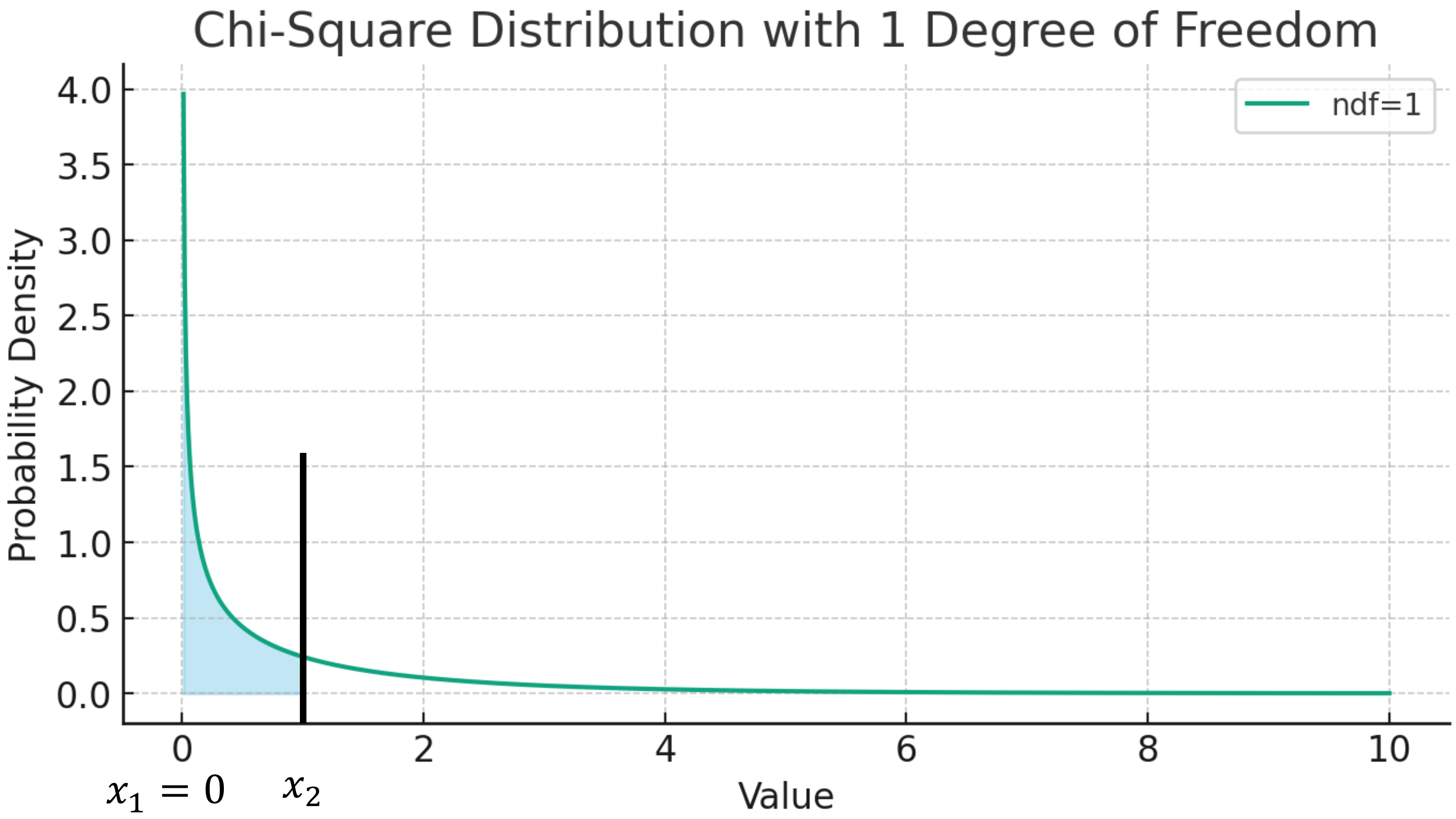
We have several measurements: $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$. Each measurement $x_i \sim N(\mu, \sigma^2)$, where σ^2 – is known.

$$L(x;\mu) \coloneqq f(x|\mu) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Extremely important!

- Choose test statistic: **b) Log-likelihood ratio:** $\lambda(\boldsymbol{x}) = -2\ln\frac{L(x;\mu)}{L(x;\hat{\mu})}$
- It can be shown that $\lambda(\boldsymbol{x}) = \frac{N}{\sigma^2}(\mu - \overline{\boldsymbol{x}})^2 = \frac{(\mu-\overline{\boldsymbol{x}})^2}{\frac{\sigma^2}{N}}$ (exercise for young people ;))
- $\overline{\boldsymbol{x}} \sim N\left(\mu, \frac{\sigma^2}{N}\right) \rightarrow \lambda \sim \chi^2\ (ndof = 1)$ (does not depend to μ value!) ($f(x|\theta) \rightarrow \chi^2$)
- So we know explicitly pdf for λ

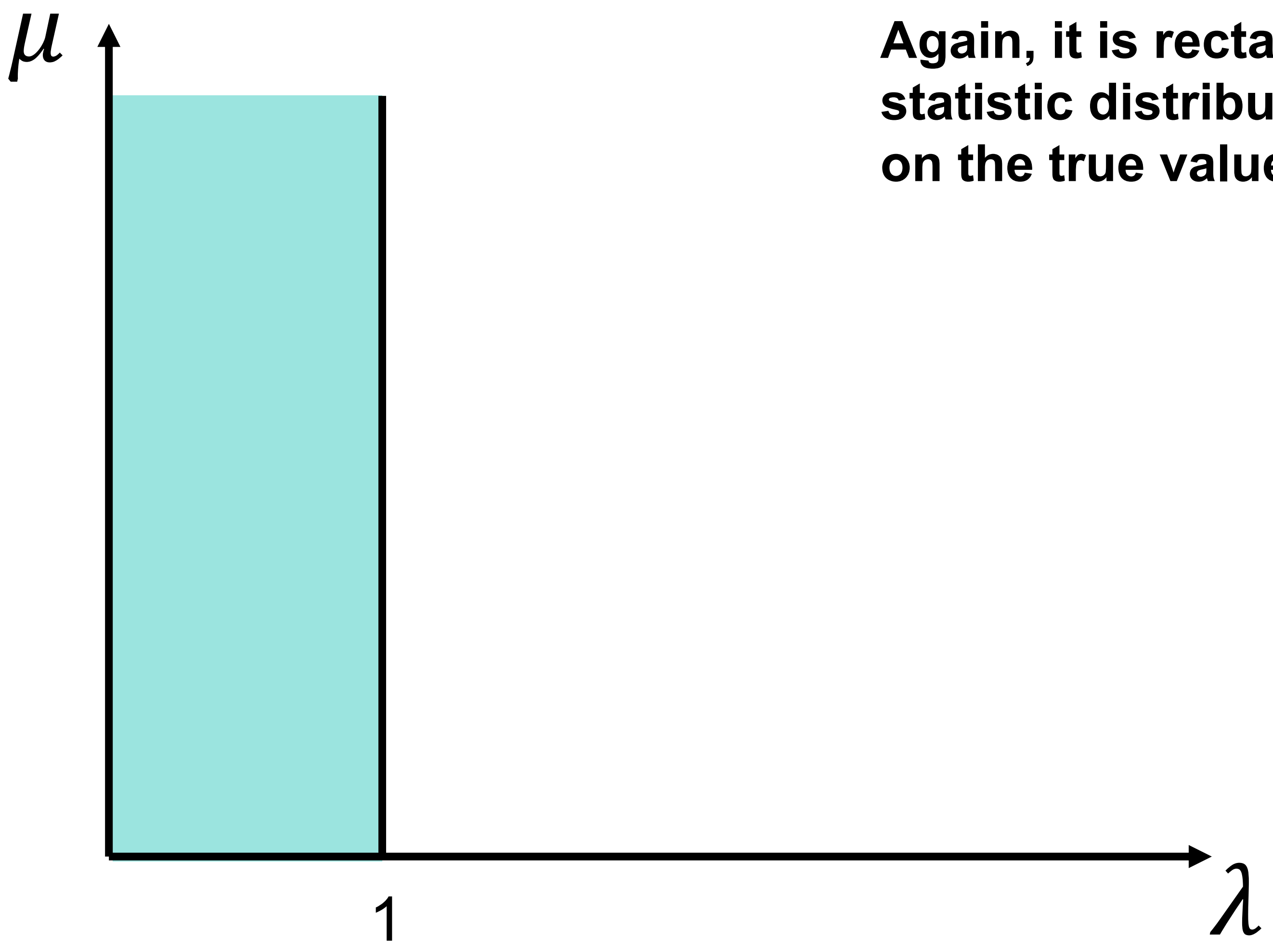
Neyman belt construction



For 68.27% C.L: $x1 = 0;$ $x_2 = 1$

Neyman belt construction

When the test statistic has χ^2 distribution, the Neyman belt is a rectangle



Again, it is rectangle because test statistic distribution does not depend on the true value of parameter μ

Neyman belt construction

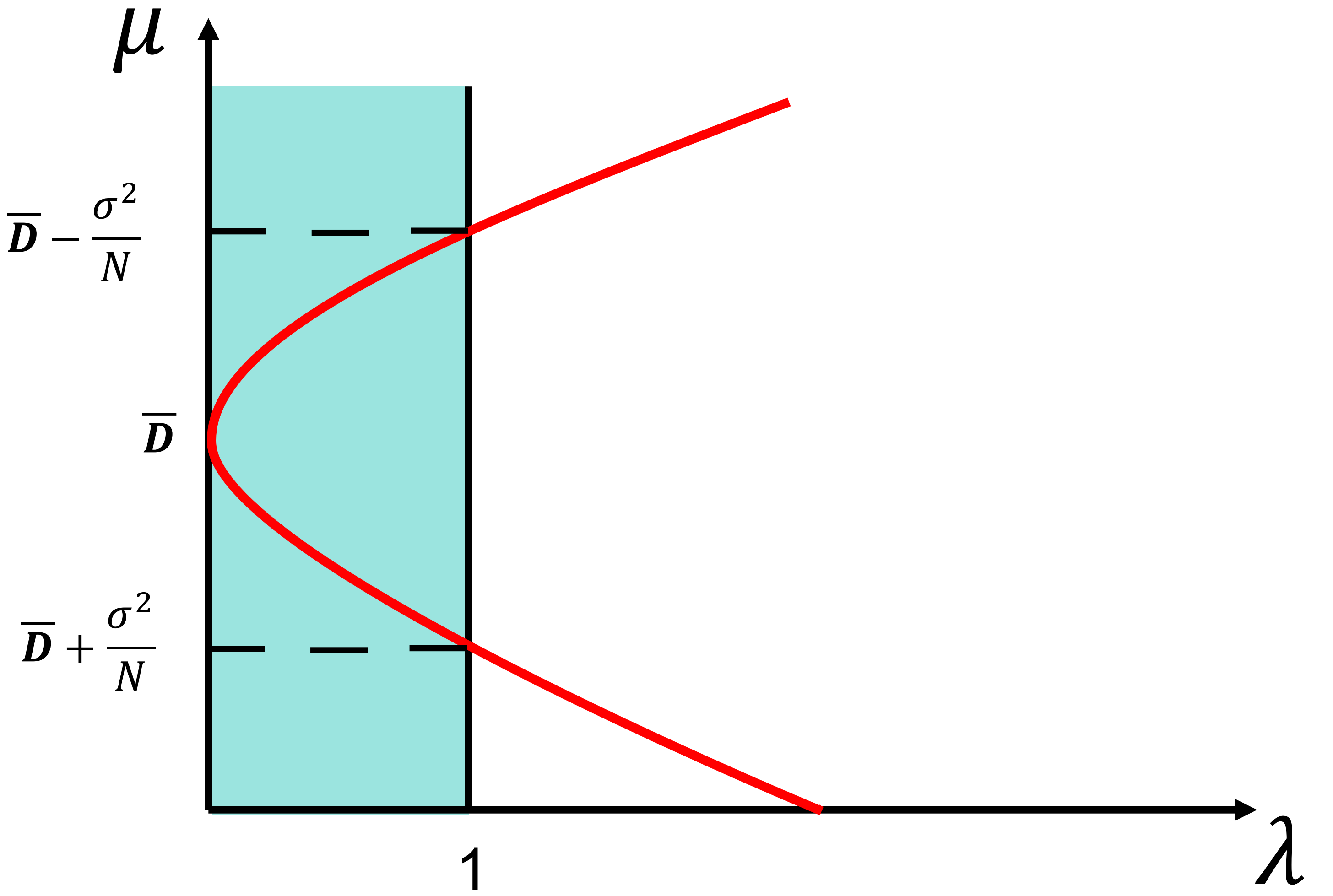
When the test statistic has χ^2 the Neyman belt is a rectangle

If our measurement is $\boldsymbol{D}$, then the test statistic for our measurement:

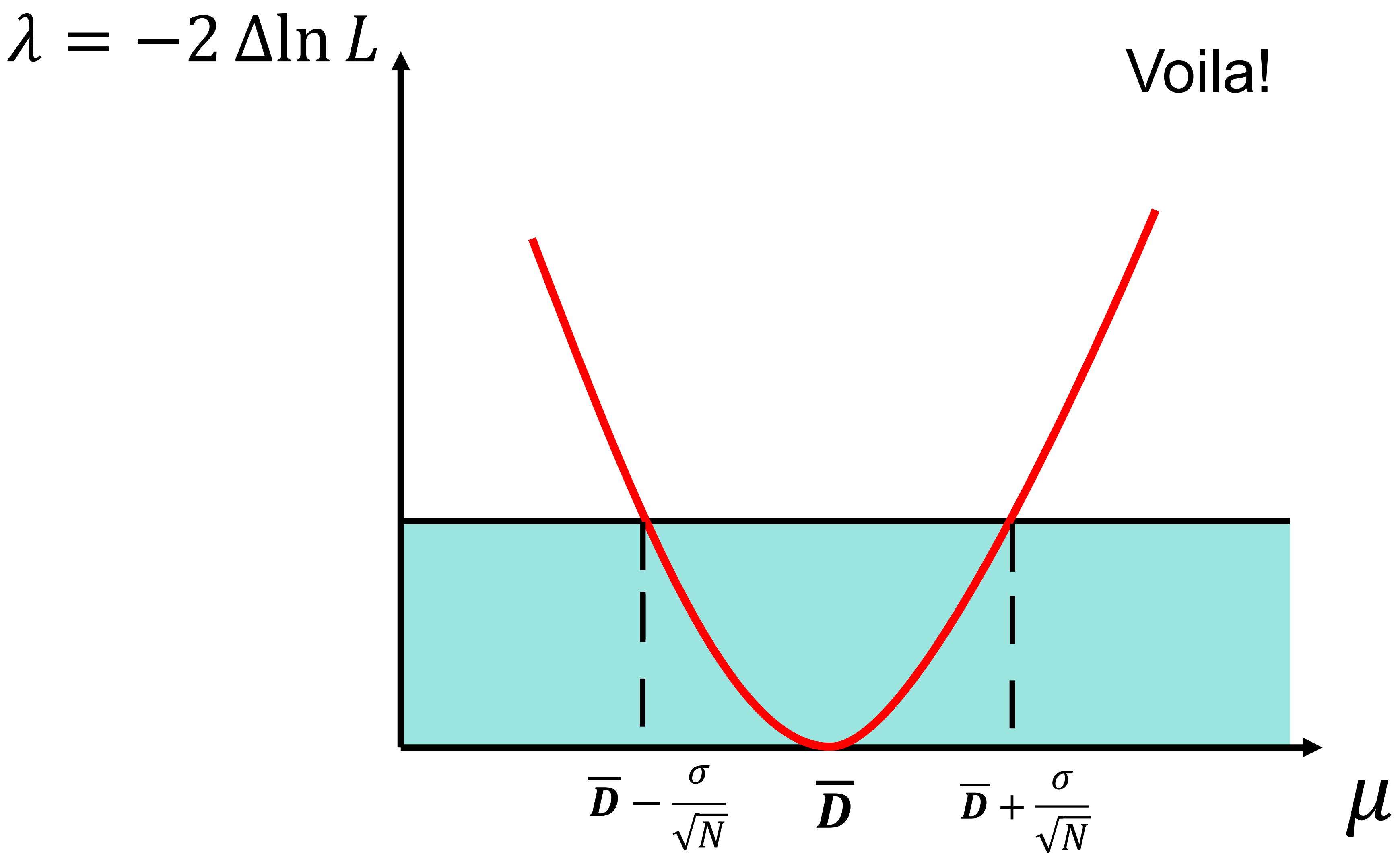
$$\lambda(\boldsymbol{D}) = \frac{N}{\sigma^2}\left(\mu - \overline{\boldsymbol{D}}\right)^2$$

A simple parabola

Let's rotate the plot!



Neyman belt construction



The plot which we are used to see!

Neyman belt construction

- Now we know how to construct the Neyman belt and understand its meaning

But we have not answered some important questions:

- How to choose ordering rule?
- How to choose test statistic?
- How to build distribution of test statistic in less trivial cases?
- How to build the belt for a model with many parameters?

I will not discuss in big details, but I will give some answers

How to build the belt for a model with many parameters?

Answer:

Use profiling or marginalisation to reduce the dimensionality of your likelihood

I do not include here the details. I discussed about these one year ago at the group meeting ;)
Hopefully, you remember everything :D

How to choose the ordering rule?

How to choose the ordering rule?

Good option:

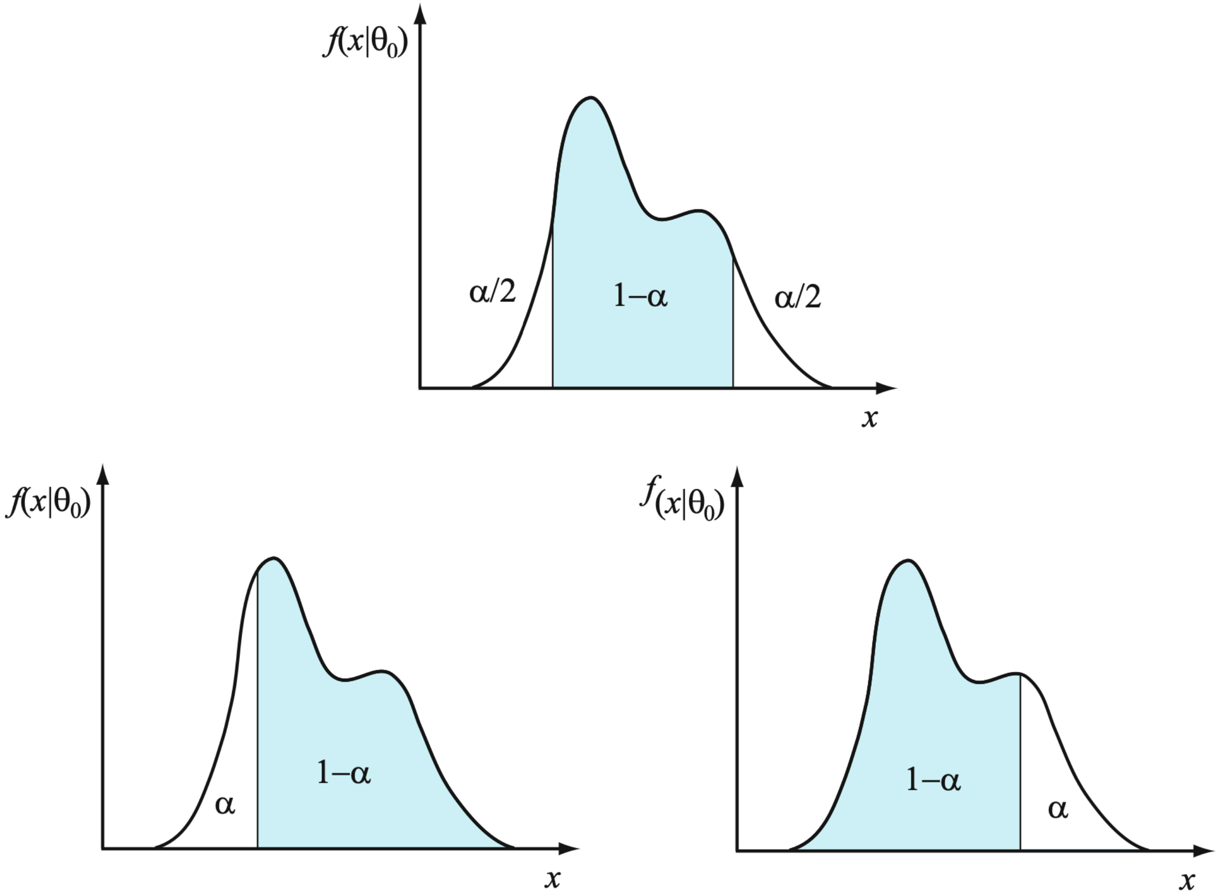
The Unified Feldman-Cousins Approach

(see here)

Neyman belt construction

One significant problem of the belt – **the choice of** $[x_1(\theta_0), x_2(\theta_0)]$ **is arbitrary**

- The choice of the $[x_1(\theta_0), x_2(\theta_0)]$ in called ***ordering rule***
- Thus, the confidence interval will depend on the chosen ordering rule
- Also in some cases a chosen ordering rule can give incorrect coverage (e.g. “flip-flopping problem”)



The Feldman-Cousins Approach

Feldman-Cousins ordering rule is based on ordering of likelihood ratio

It ensures the correct coverage

Use **likelihood ratio:**

$$\lambda(x|\theta_0) = \frac{f(x|\theta_0)}{f\left(x|\hat{\theta}(x)\right)}$$

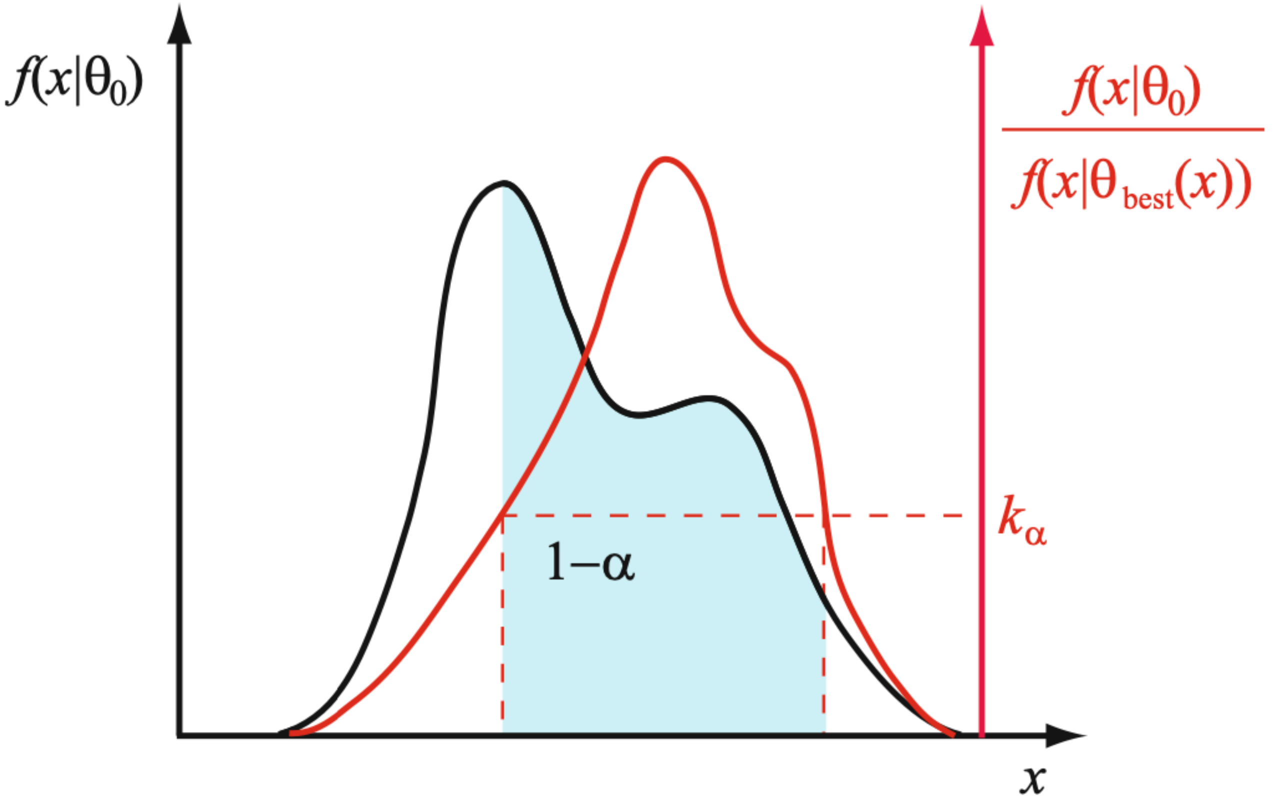
$\hat{\mu}$ –best fit for measurement $\boldsymbol{x}$

The interval $[x_1(\theta_0), x_2(\theta_0)]$ is defined by

$$R_\alpha(\theta_0) = \{x : \lambda(x|\theta_0) > k_\alpha\}$$

where k_α is found from following equation for target coverage $1-\alpha$ (C.L.)

$$\int_{R_\alpha} f(x|\theta_0)\,\mathrm{d}x = 1-\alpha$$



k_α is called critical value corresponding to C.L. $1-\alpha$

How to choose the test statistic?

Good option:

The likelihood ratio

For several reasons:

1) **The Neyman-Pearson lemma:** the likelihood ratio ensures the most optimal hypotheses selection

2) **Wilk's theorem:** Under certain condition the log-likelihood ratio asymptically converges to χ^2 distribution

How to build the distribution of the test statistic for a general case?

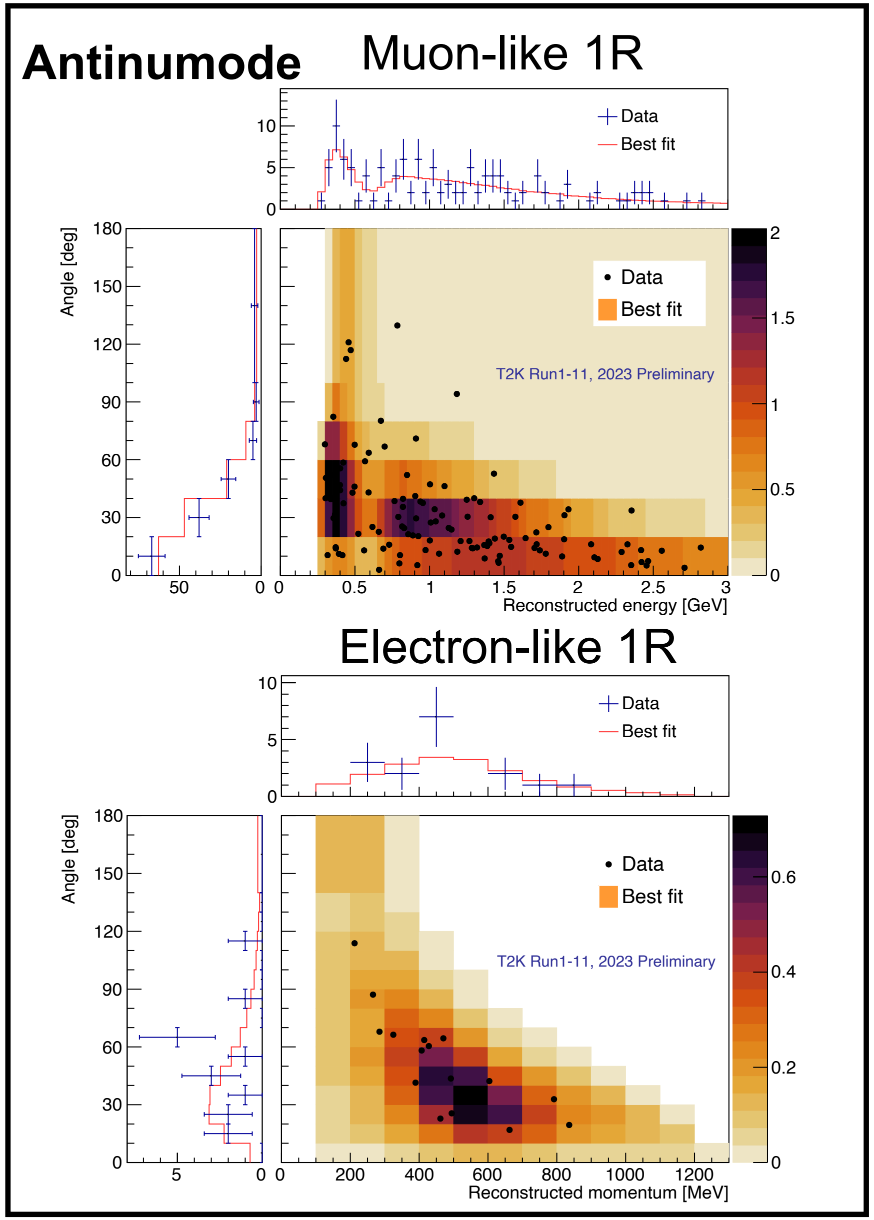
Generate MC simulations

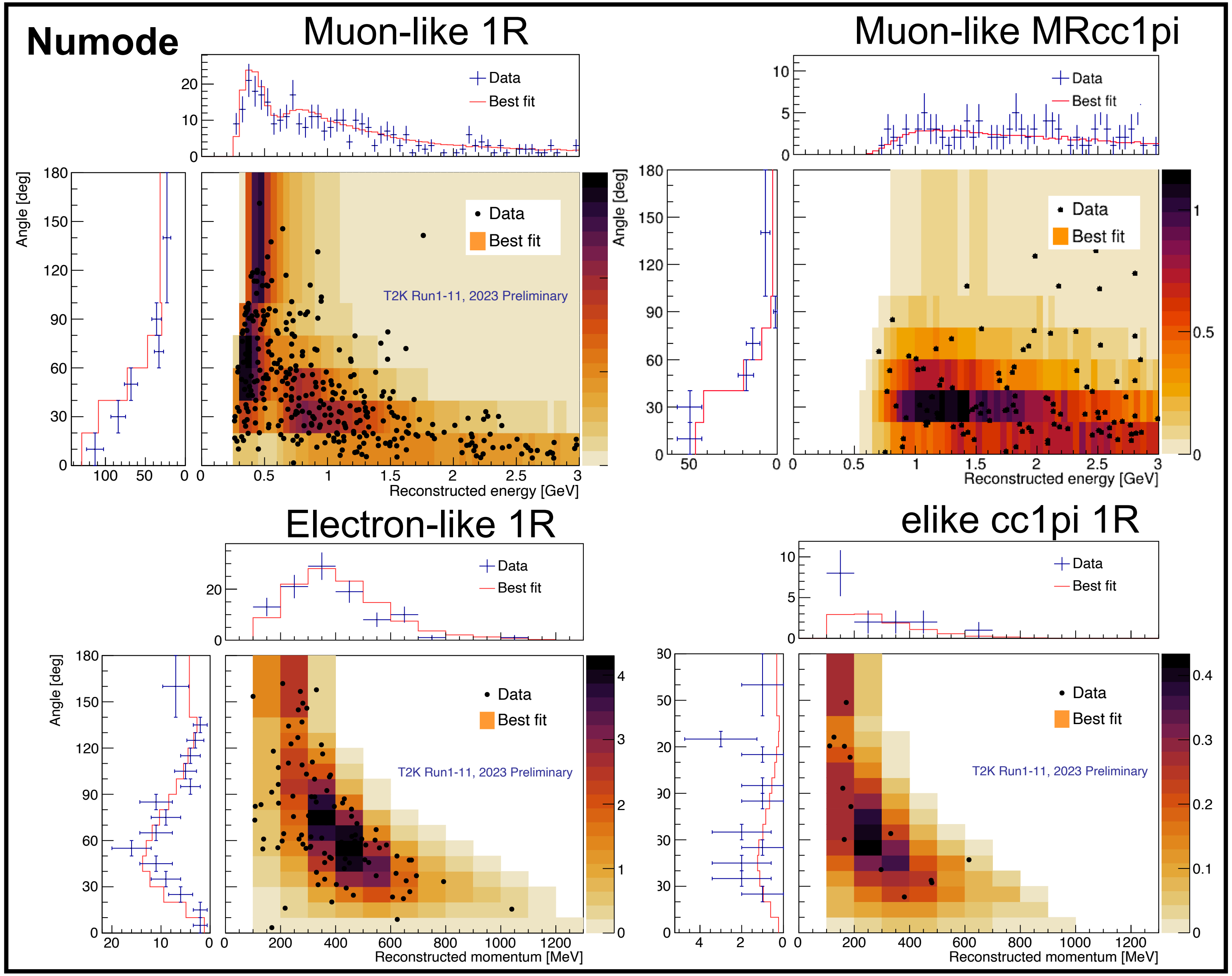
But how? If you have nuisance parameters, then you should decide how to sample them!
There are several options – I will discuss later only what is done in T2K

Part IV:

T2K approach to claim the oscillations results

Part IV: T2K approach to claim the results





Part IV: T2K approach to claim the results

$$-\ln L = \sum_{s,i} \left[N_{s,i}^{exp}(\boldsymbol{o},\boldsymbol{f}) - N_{s,i}^{obs} + N_{obs} \ln \frac{N_{s,i}^{obs}}{N_{s,i}^{exp}(\boldsymbol{o},\boldsymbol{f})} \right] + (\boldsymbol{f} - \boldsymbol{f_0})^T V^{-1} (\boldsymbol{f} - \boldsymbol{f_0})$$

$$L = L(\boldsymbol{N}^{obs}; \delta_{CP}, \Delta m_{32}^2, \theta_{23}, \theta_{13}, MO, \boldsymbol{f})$$

1) $V = \begin{pmatrix} V_{ND} & O \\ O & V_{SK} \end{pmatrix} (170\times170)$, where V_{ND} –comes from BANFF or Gundam fit

V_{ND} – comes from T2K-SK group from atmospheric fit

2) T2K uses reactor constraint (RC) θ_{13} - Gaussian prior

3) Are used 6 SK samples categorised by neutrino mode, neutrino flavour and number of observed rings

Part IV: T2K approach to claim the results

$$\boldsymbol{o}_{bf} = \arg\max_{\boldsymbol{o},\boldsymbol{f}} L(\boldsymbol{N}^{obs}; \boldsymbol{o}, \boldsymbol{f})$$

This is declared as central value

Part IV: T2K approach to claim the results

To estimate the error we find C.I. corresponding to different C.L. T2K uses Neyman belt construction using FC ordering

- Let's build it for a parameter θ which is one the oscillation parameters
- We need to build a scalar test statistic **only as function of true $\boldsymbol{\theta}$** $\lambda(\boldsymbol{N}^{obs}|\theta)$
- For reason mentioned in previous part **we use log-likelihood ratio** as test statistics
- But we cannot directly use our likelihood as it depends on 170 parameters, not only on θ. To reduce the likelihood dimensionality we use marginalisation. So, for **statistic we use log-marginalised likelihood**

Why marginalisation? Why not profiling?

Part IV: T2K appr./ Why do we use marginalisation?

Main reason: CPU time

Example. If you are using paralisation with 100 jobs to compute an **Asimov fit:**

Marginal likelihood: you need ~2 minutes
Profiled likelihood: you need ~10*4 minutes

Factor of 4 you gain to cover al local minima: two octants and two sign of $\cos \delta_{CP}$

To compute a toy fit:
Marginal likelihood: you need ~2 minutes
Profiled likelihood: you need ~30*4 minutes

Why do we use marginalisation?

For FC method and p-values calculation you need to fit thousands of toys!

For profiled likelihood it would take ages!

Part IV: T2K approach to claim the results

This will x −axis on the Neyman belt

Test statistic for Neyman belt construction:

$$\lambda(\boldsymbol{N}^{obs}|\theta) = -2\ln\frac{L_{marg}(\boldsymbol{N}^{obs};\theta)}{L_{marg}(\boldsymbol{N}^{obs};\hat{\theta})},$$

$\boldsymbol{\theta}$ – parameter of interest

where $L_{marg}(\boldsymbol{N}^{obs};\theta) := \int L_s(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta})\pi(\boldsymbol{\eta})d\boldsymbol{\eta}$; $\hat{\theta} = \arg\max_\theta L_{marg}(\boldsymbol{N}^{obs};\theta)$

$\boldsymbol{\eta}$ −all params except θ

- <u>Main task here: we need to know the pdf for λ for different true values of θ:</u>

1) **Calculate it by sampling $\boldsymbol{N}^{obs}$**

2) **Use a theorem to know distribution of λ without any calculations**

The uncertainty of the best-fit

1) Calculate λ distribution by sampling $\boldsymbol{N}^{obs}$

$\boldsymbol{\theta}$ – parameter of interest

Sample $\boldsymbol{N}^{obs}$ → Calculate $\lambda(\boldsymbol{N}^{obs}|\theta)$

However, the toys depend on the true value of nuisance parameters!

Sample nuisance parameters distribution → Sample $\boldsymbol{N}^{obs}$ → Calculate $\lambda(\boldsymbol{N}^{obs}|\theta)$

↑

How?!

Part IV: T2K approach to claim the results

Sample nuisance parameters distribution → Sample $\boldsymbol{N}^{obs}$ → Calculate $\lambda(\boldsymbol{N}^{obs}|\theta)$

↑

How?!

There is no any definite answer. Maybe it is the most challenging step in T2K oscillation parameters inference

There are many different methods which can be chosen:

1) A priori estimate
2) Conservative
3) Berger–Boos
4) Highland–Cousins
5) A posteriori Highland–Cousins
6) Profiled method

You can check this paper which shortly descibes these methods

Part IV: T2K approach to claim the results

- T2K **calculates** log-likelihood ratio distribution for $\boldsymbol{\delta_{CP}}$ and $\mathbf{sin^2\theta_{23}}$

- T2K uses Highland–Cousins method to sample nuisance parameters

Highland–Cousins method realisation in T2K to sample nuisance parameters:

1) The posterior distributions of the oscillation parameters are calculated

2) These distributions are used to sample nuisance oscillation parameters

3) Systematic parameters are sampled from prior distributions

Part IV: T2K approach to claim the results

Summary: We sample the toys following Highland-Cousins method and then calculate log-likelihood distribution which are used in Neyman belt construction

Don’t worry I will show an explicit example now!

Part IV: T2K approach to claim the results

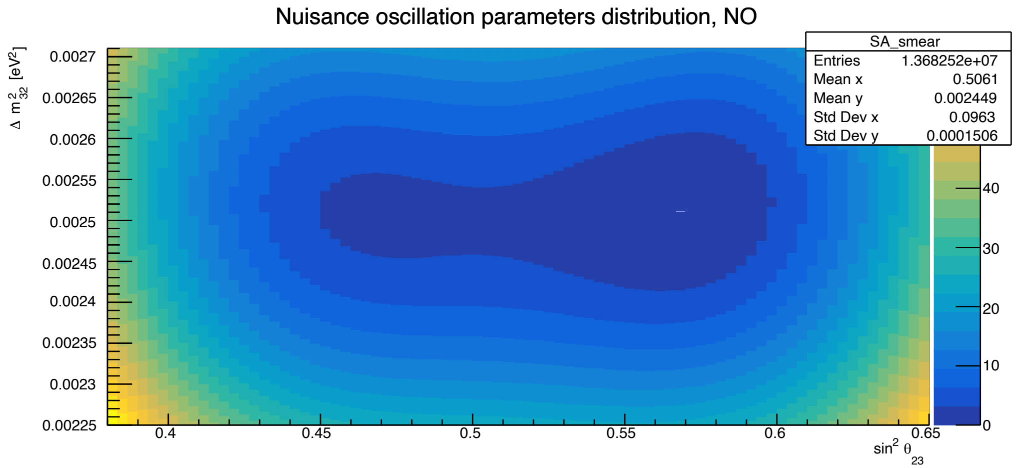
Let's now follow all the mentioned steps for δ_{CP}

Part IV: T2K approach to claim the results

1) We need to generate the toy experiments

For that we need to sample nuisance parameters. For each true value of δ_{CP}:

a) For Δm^2_{32} and $\sin^2\theta_{23}$ generate 2D data-posterior from $\Delta\chi^2$
b) For $\sin^2\theta_{13}$ use RC
c) For all systematics use prior distribution



50k toy experiments

Part IV: T2K approach to claim the results

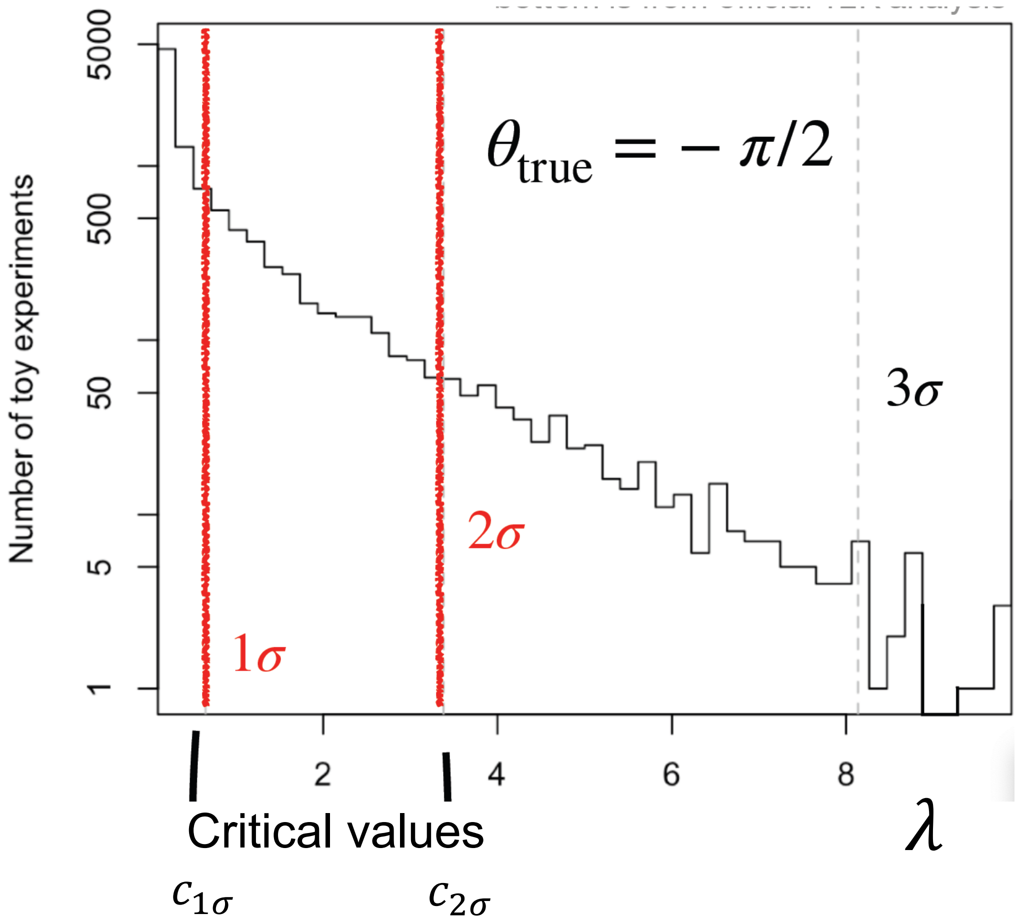
2) We need to fit the toys. For that we marginalise the likelihood at different toys

3) We calculate the test statistics, which is log-likelihood ratio:

$$\lambda\left(\boldsymbol{N}^{obs}\middle|\delta_{CP}\right) = -2\ln\frac{L_{marg}\left(\boldsymbol{N}^{obs};\delta_{CP}\right)}{L_{marg}\left(\boldsymbol{N}^{obs};\widehat{\delta_{CP}}\right)}$$

4) We find critical values cor. to a C.L. α:

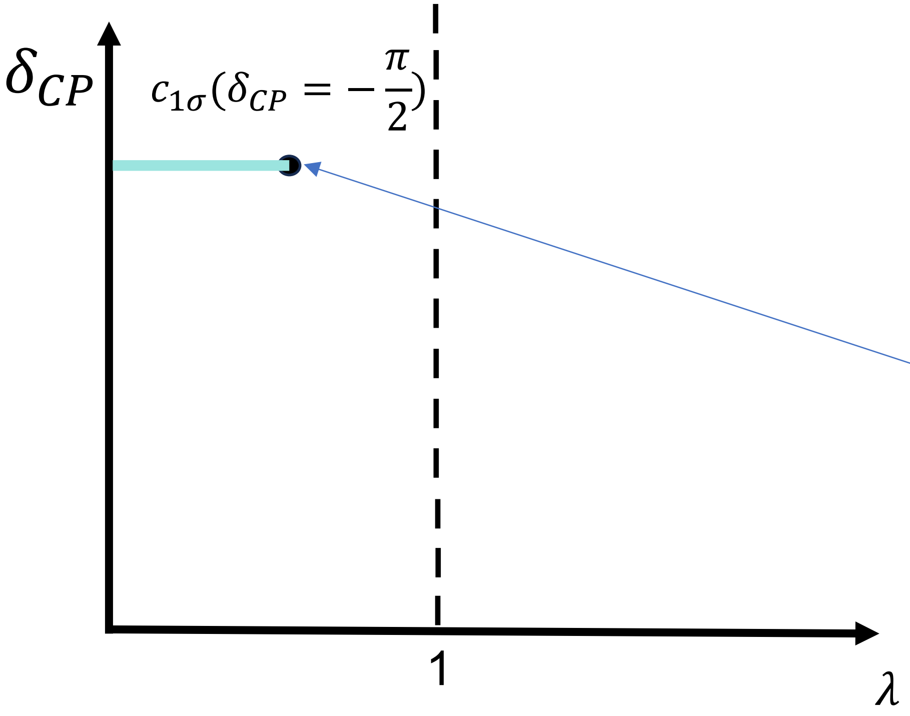
$$\int_0^{c_\alpha(\delta_{CP})} f\left(\lambda\left(\boldsymbol{N}^{obs}\middle|\delta_{CP}\right)\right)d\lambda = \alpha$$



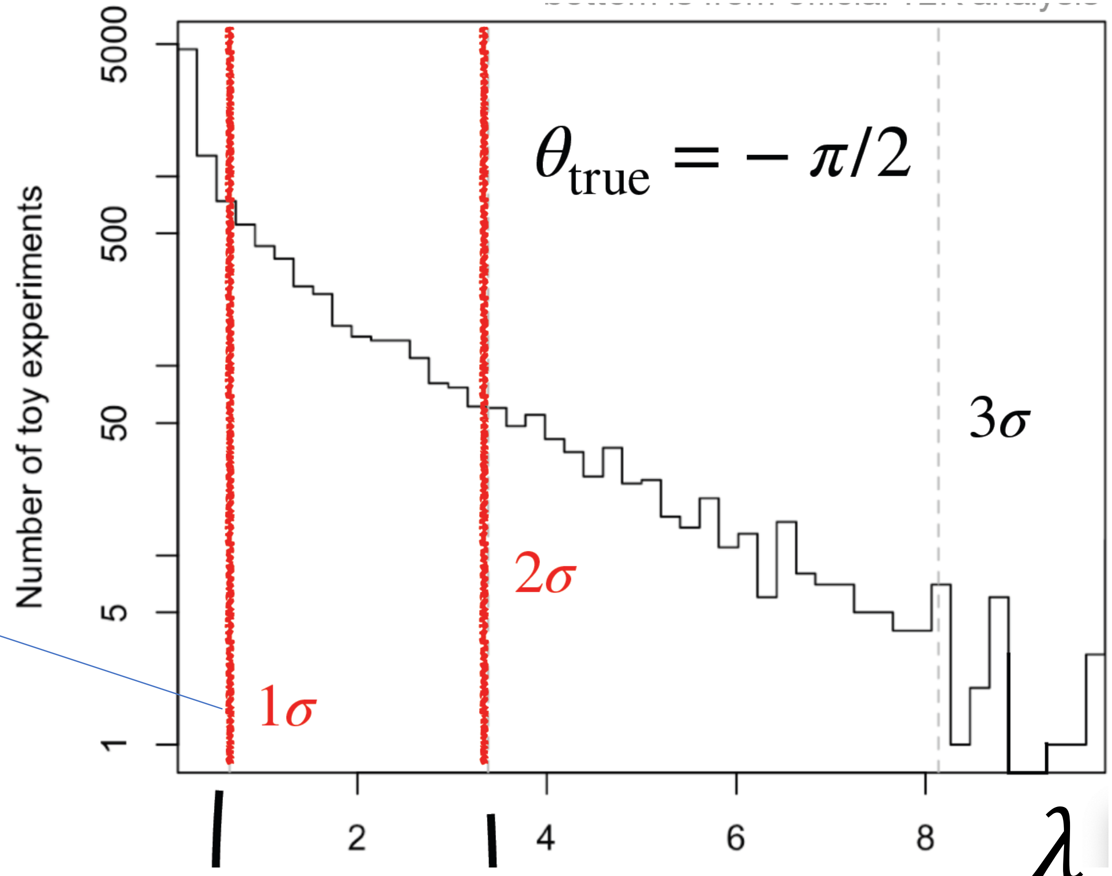
Part IV: T2K approach to claim the results

4) We find critical values cor. to a C.L. α:

$$\int_0^{c_\alpha(\delta_{CP})} f\left(\lambda\left(\boldsymbol{N}^{obs}\middle|\delta_{CP}\right)\right) d\lambda = \alpha$$



Test statistic distribution

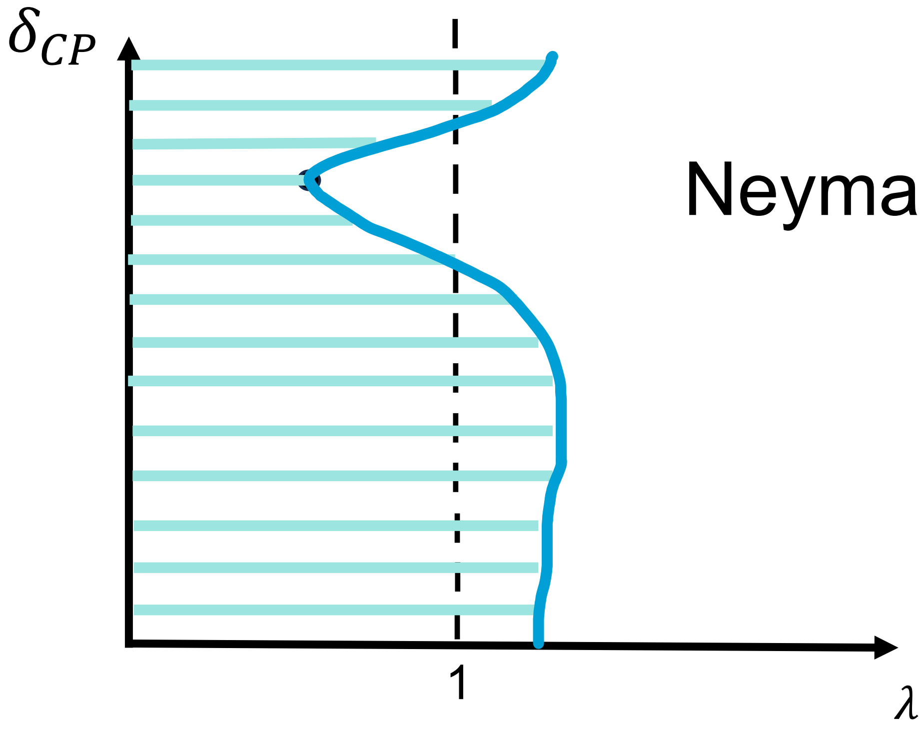


Critical values (which I called x_2)

Part IV: T2K approach to claim the results

4) We find critical values cor. to a C.L. α:

$$\int_0^{c_\alpha(\delta_{CP})} P\left(\lambda\left(\boldsymbol{N}^{obs}\middle|\delta_{CP}\right)\right) d\lambda = \alpha$$



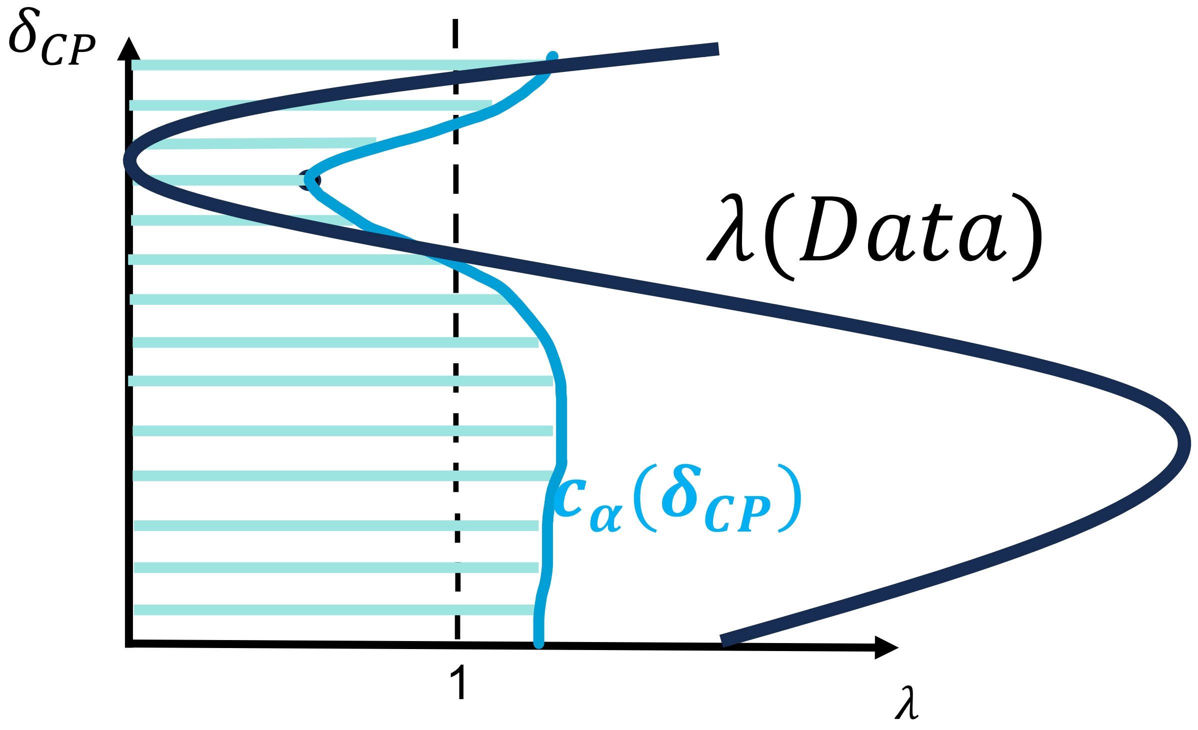
Neyman belt!

- Repeating for different true values of δ_{CP} we contruct the Neyman belt
- We have done full algorithm of Neyman belt contstruction!

Part IV: T2K approach to claim the results

4) We find critical values cor. to a C.L. α:

$$\int_0^{c_\alpha(\delta_{CP})} P\left(\lambda\left(\boldsymbol{N}^{obs}\middle|\delta_{CP}\right)\right) d\lambda = \alpha$$

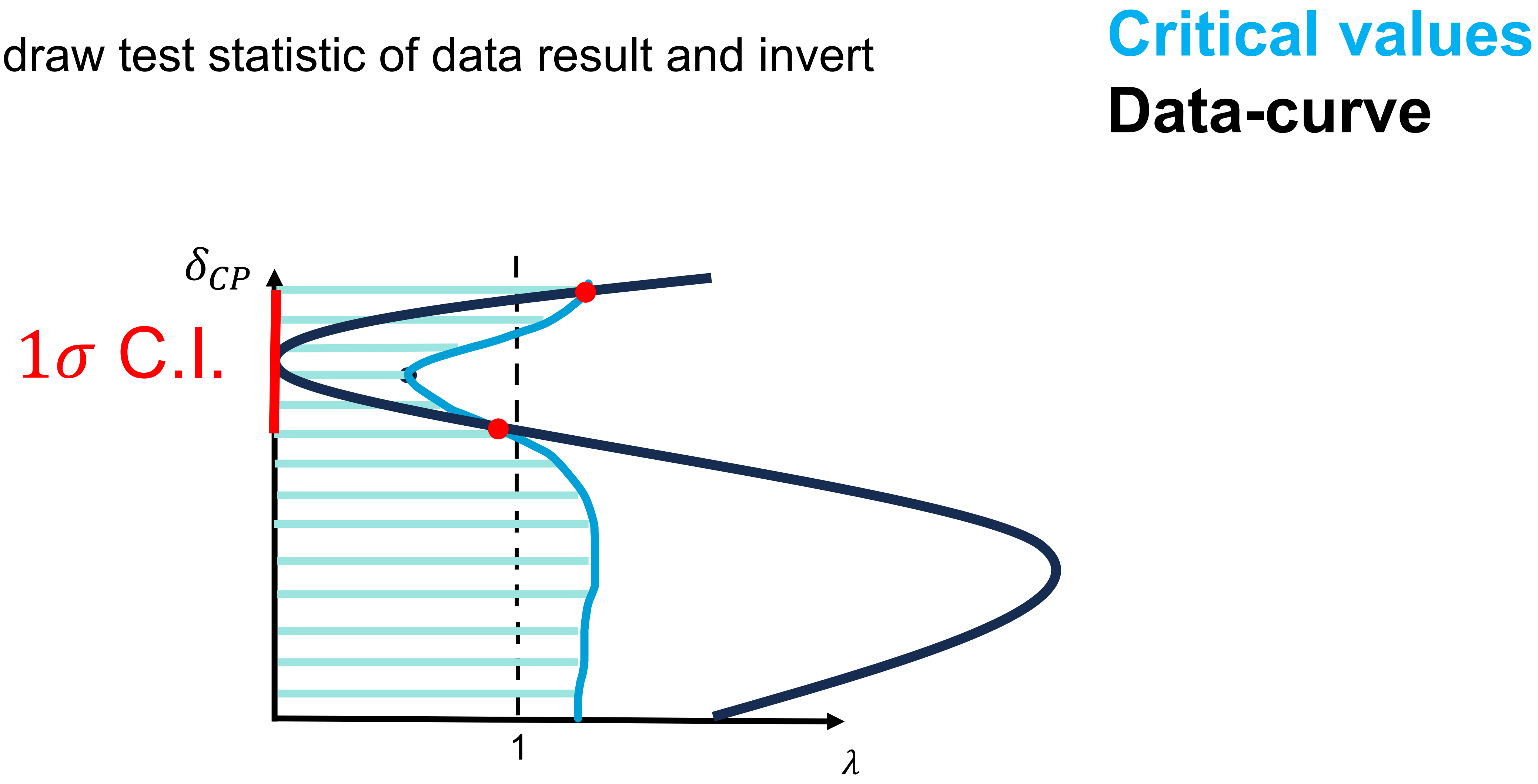


Reminder: $\lambda = -2\Delta \ln L$

- Repeating for different true values of δ_{CP} we contruct the Neyman belt
- We have done full algorithm of Neyman belt contstruction

Part IV: T2K approach to claim the results

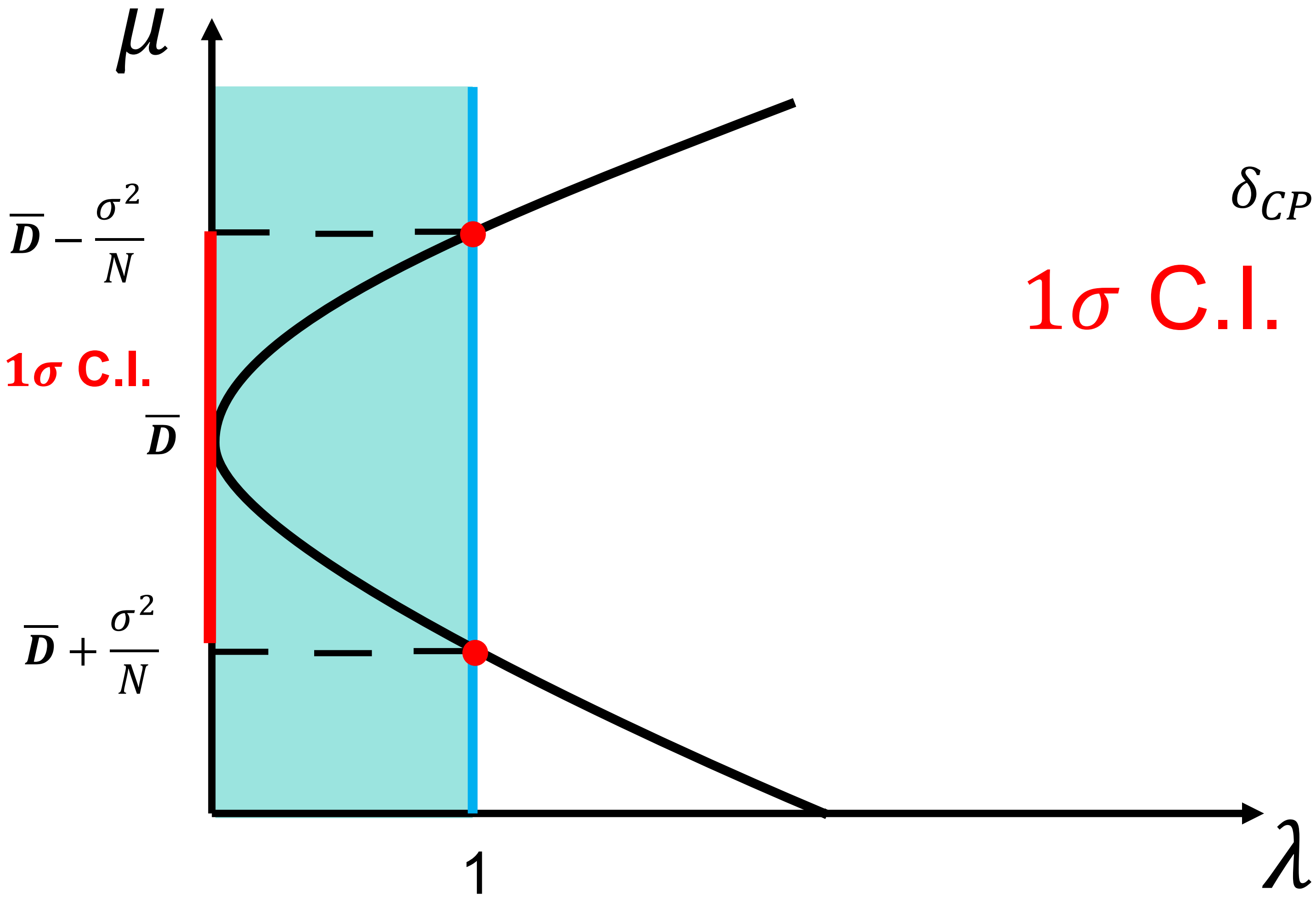
- Last step: draw test statistic of data result and invert the belt.

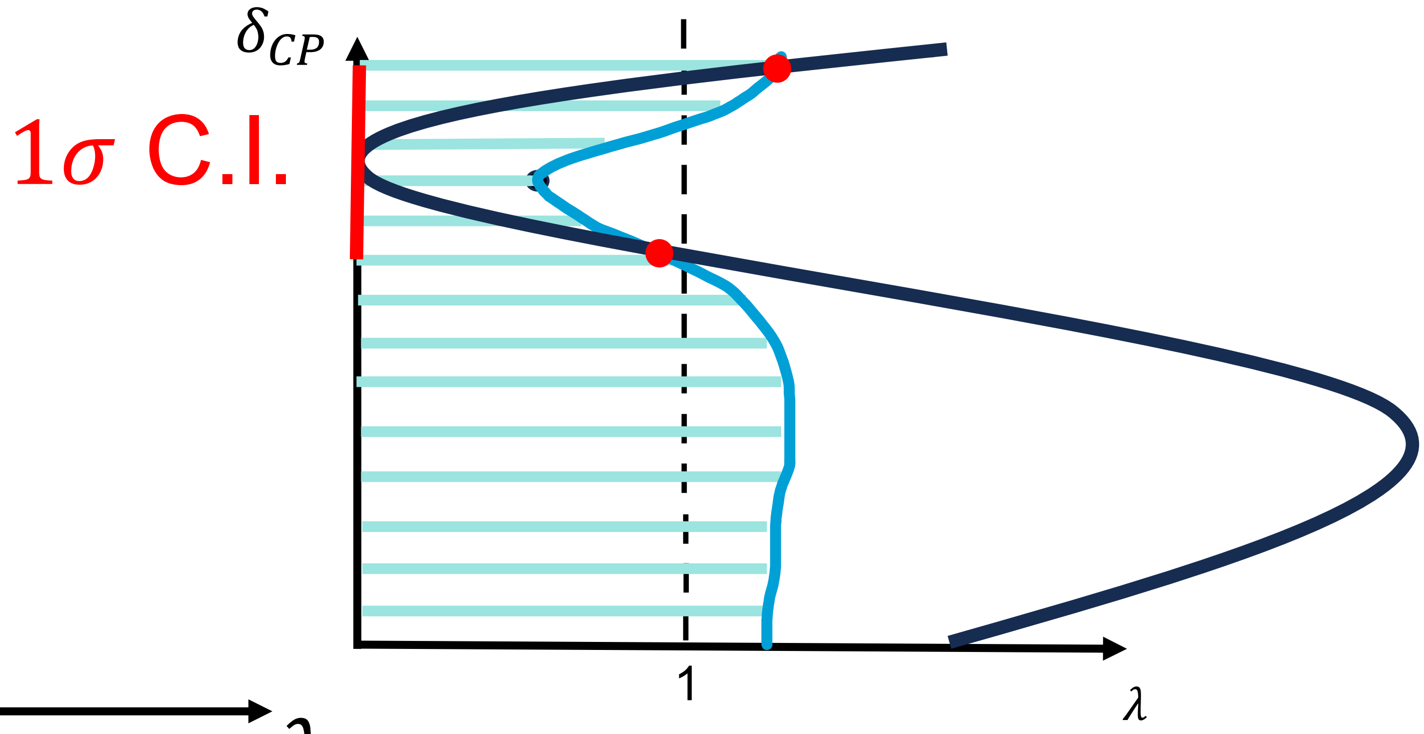


Critical values

Data-curve

Compare with the Gaussian example!





What we have just performed is called in T2K Feldman-Cousins method for δ_{CP}

But why?!

Part IV: T2K approach to claim the results

- Some words about terminology because it can be misleading:
  Instead of “FC method” I think it is better to say “Neyman belt contruction using FC ordering”

- Personally I had the following problem while I was reading the presentations or TN on this topic: I thought that “FC method is a method when you generate the toys and fit them to find C.I.”. This is not correct interpretation!

- According to this interpretation FC method and Neyman belt construction are the same thing – not true at all!

- Feldman-Cousins is an ordering rule! And Feldman-Cousins method is about how to order you toys when you want to extract C.I.!

Part IV: T2K approach to claim the results

- We have clarified the terminology but still it is not clear: at which step we ordered the toys? I just took upper limits of log-likelihood ratio statistic. Is it?

- Let me explicitly show where the ordering was performed

The Feldman-Cousins Approach

Feldman-Cousins ordering rule is based on ordering of likelihood ratio

It ensures the correct coverage

Use **likelihood ratio:**

$$\lambda(x|\theta_0) = \frac{f(x|\theta_0)}{f\left(x|\hat{\theta}(x)\right)}$$

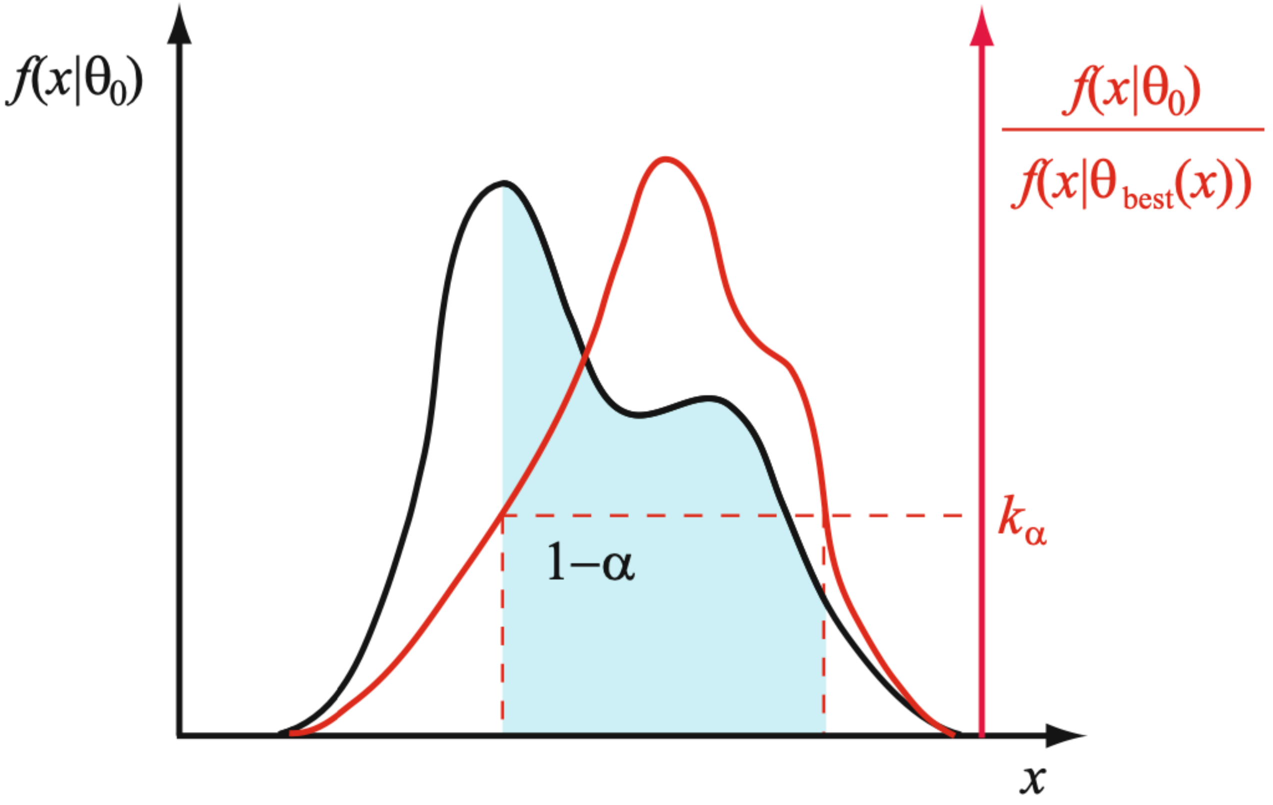
$\hat{\mu}$ –best fit for measurement $\boldsymbol{x}$

The interval $[x_1(\theta_0), x_2(\theta_0)]$ is defined by

$$R_\alpha(\theta_0) = \{x : \lambda(x|\theta_0) > k_\alpha\}$$

where k_α is found from following equation for target coverage $1-\alpha$ (C.L.)

$$\int_{R_\alpha} f(x|\theta_0)\,\mathrm{d}x = 1-\alpha$$



$\boldsymbol{k_\alpha}$ **is called critical value corresponding to C.L.** $\boldsymbol{1-\alpha}$

The Feldman-Cousins Approach

Use **log-likelihood ratio:**

$$LLR\left(\boldsymbol{N}^{obs}|\delta_{CP}\right) = -2\ln\frac{L_{marg}\left(\boldsymbol{N}^{obs};\delta_{CP}\right)}{L_{marg}\left(\boldsymbol{N}^{obs};\widehat{\delta_{CP}}\right)}$$

Using LLR we order the toys:

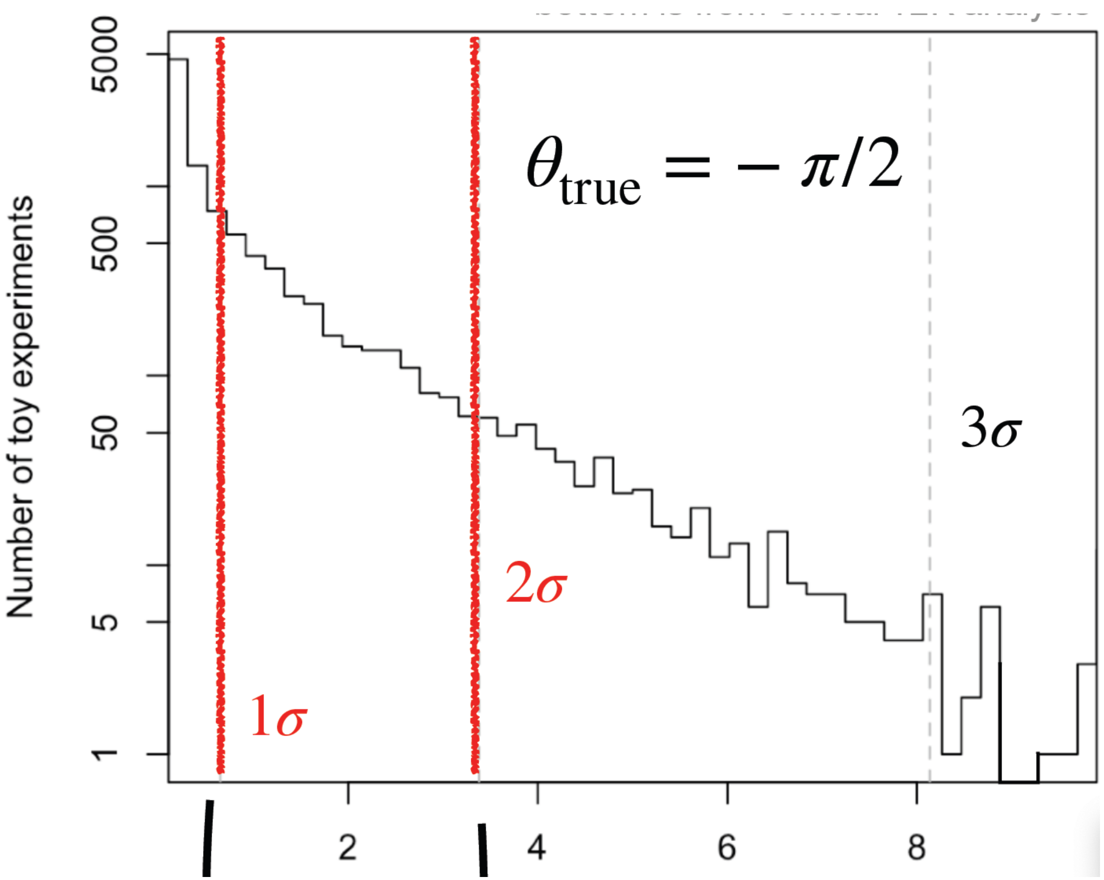
$$R_\alpha(\delta_{CP}) = \{\boldsymbol{N}^{obs}: LLR\left(\boldsymbol{N}^{obs}|\delta_{CP}\right) < c_\alpha\}$$

where c_α is found from following equation for target coverage $1-\alpha$ (C.L.)

$$\int_{R_\alpha} f(\lambda|\delta_{CP})d\lambda = 1-\alpha$$

Then important transition: if $\lambda == LLR$, then we don't need do explicitly this ordering, we just take upper limits:

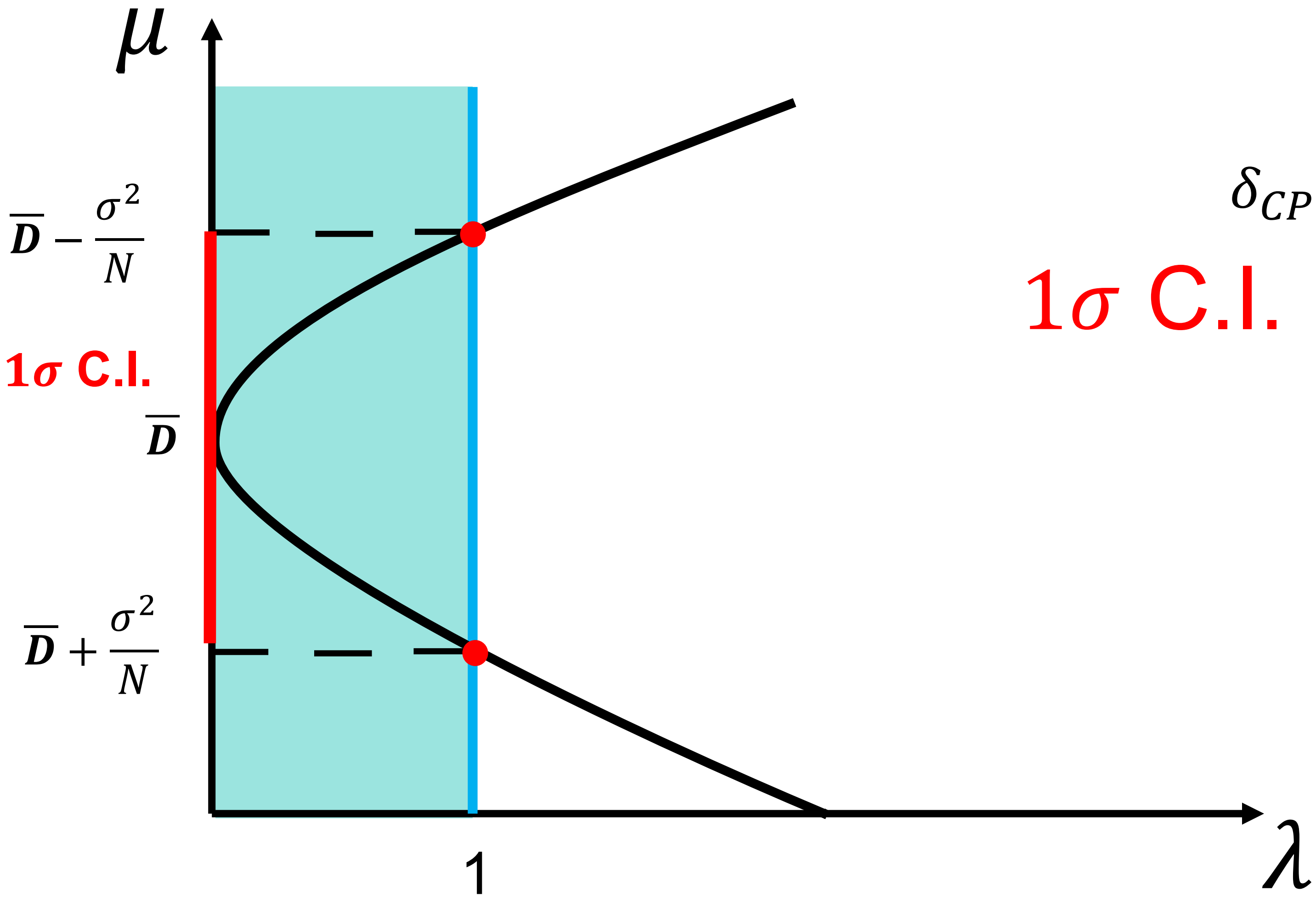
$$\int_0^{c_\alpha} f(LLR|\delta_{CP})d\lambda = 1-\alpha$$

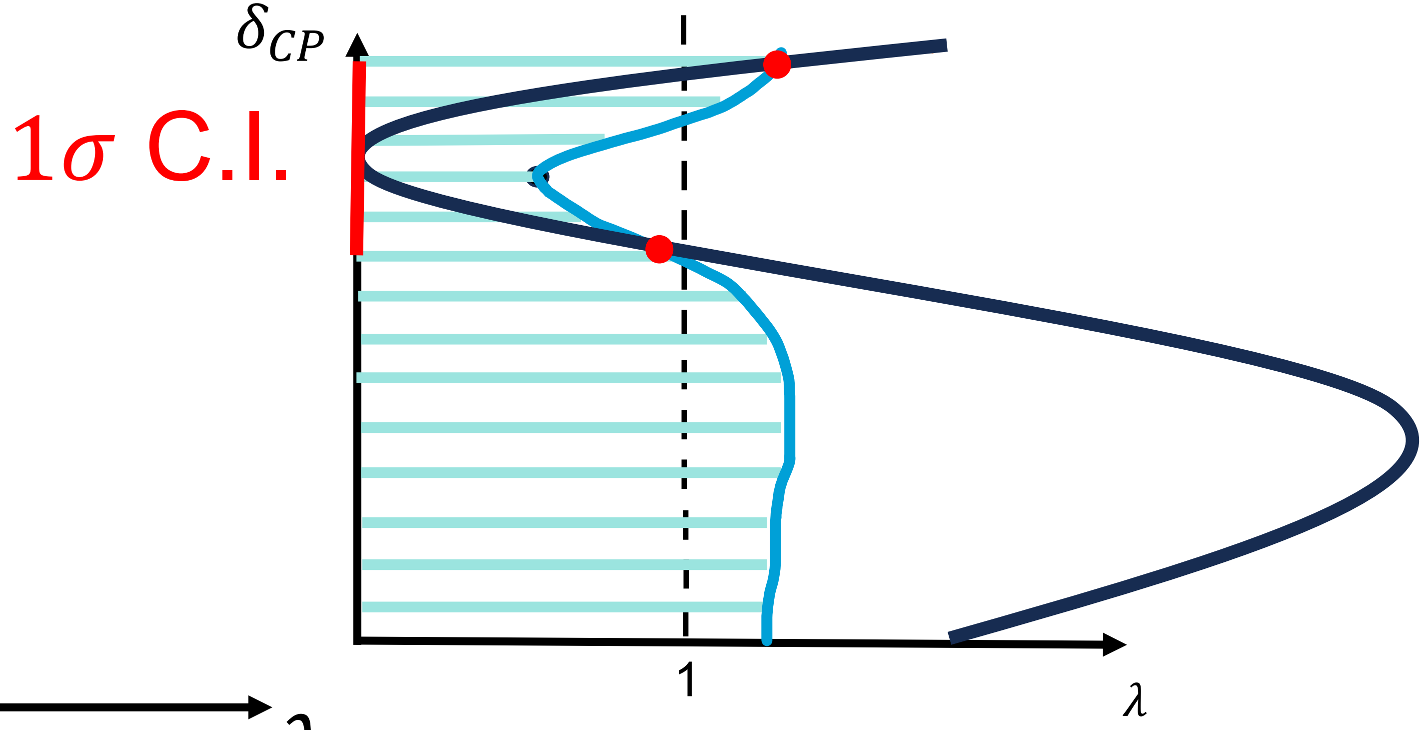


But! It is not necessary that $\lambda = \text{LLR}$, it can be different! But the ordering is always based on LLR – and this is essence of Feldman-Cousins!

Compare with the Gaussian example!

Critical values
Data-curve

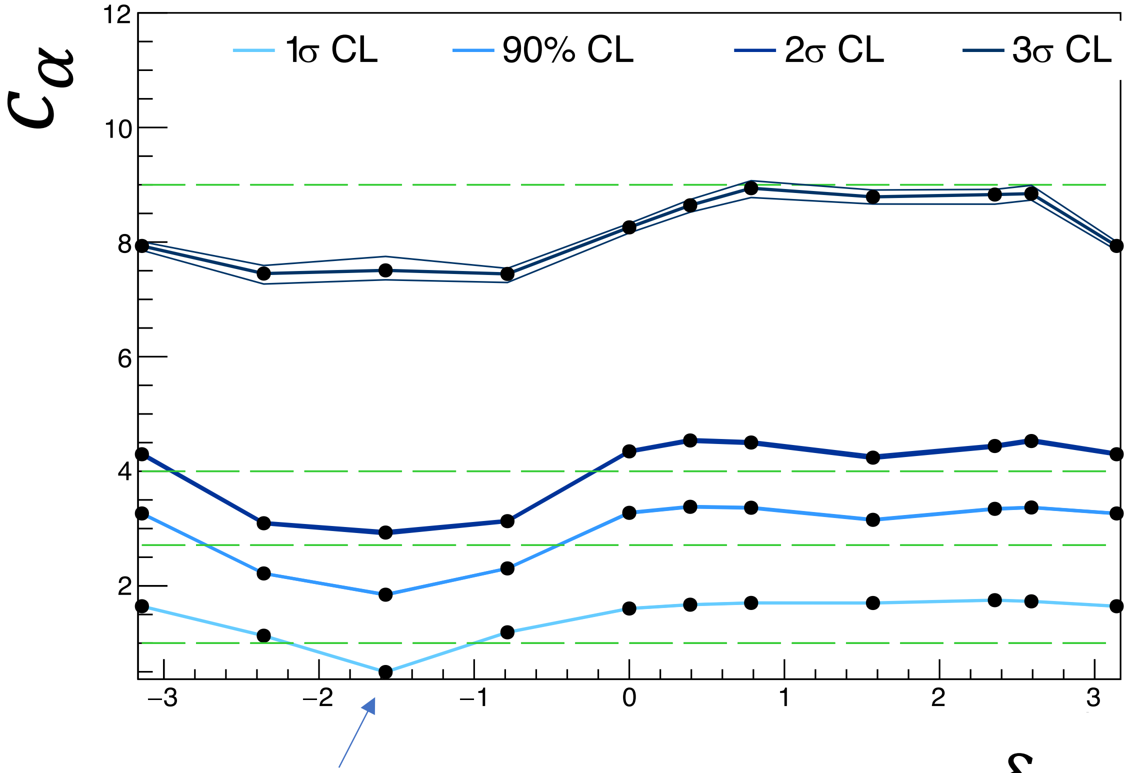




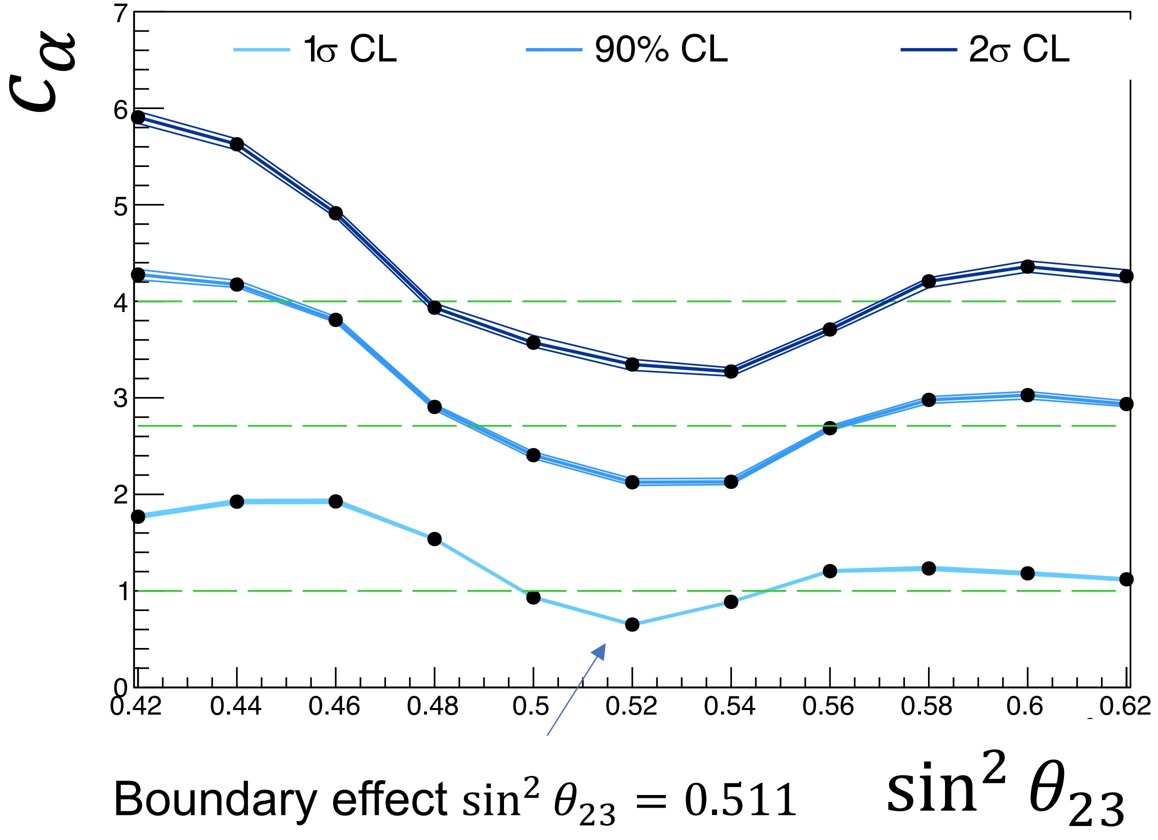
P-theta result (Frequentist)

Critical values for FC method

NH



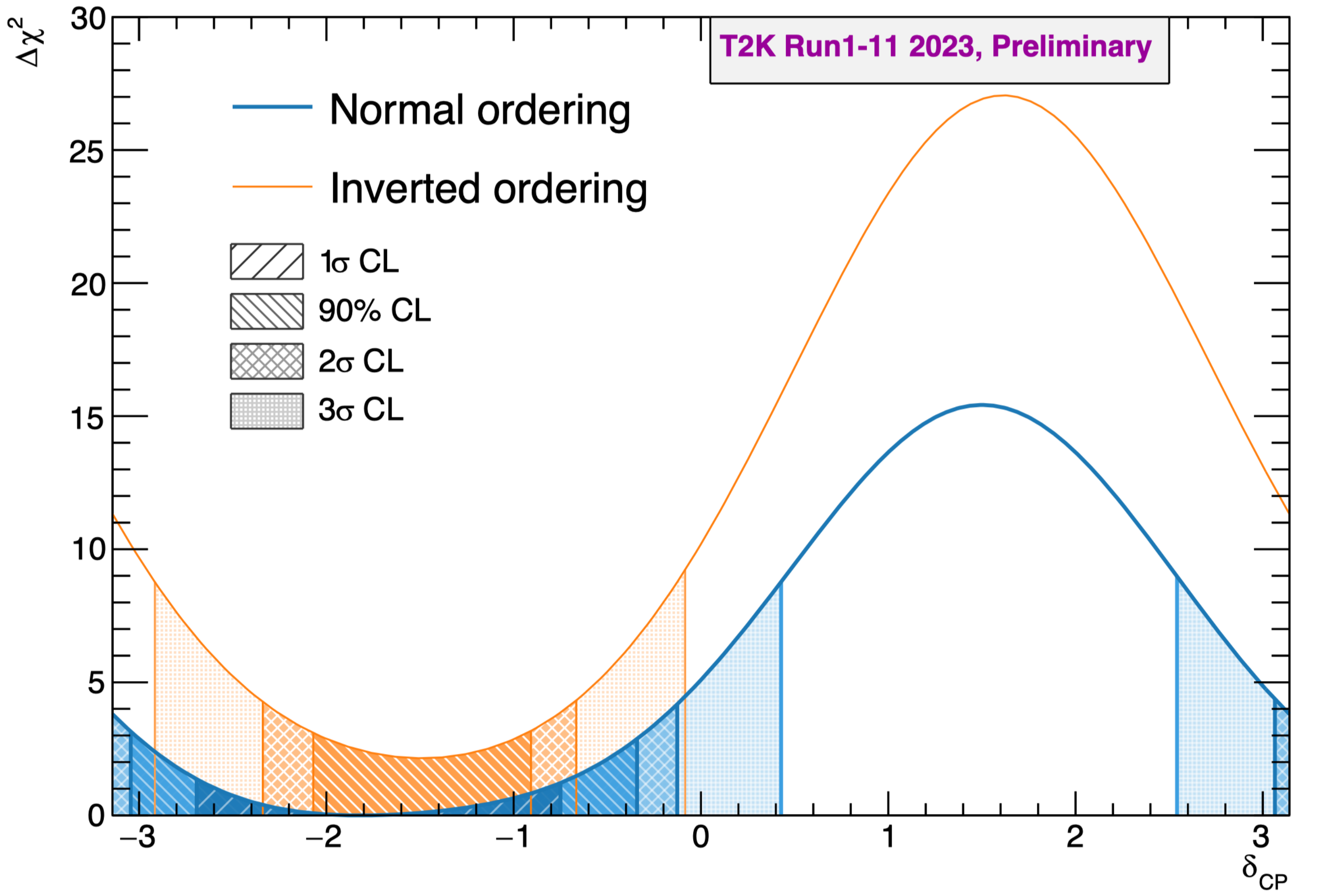
Boundary effect $\sin\delta_{CP} = -\frac{\pi}{2}$



Boundary effect $\sin^2\theta_{23} = 0.511$

Part IV: T2K approach to claim the results

Drawing $\lambda(Data|\delta_{CP})$ and rotating the plot we get the result



Repeating the same procedure for two different fixed mass orderings

Part IV: T2K approach to claim the results

Test statistic for Neyman belt construction:

$$\lambda\left(\boldsymbol{N}^{obs}\middle|\theta\right) = -2\ln\frac{L_{marg}\left(\boldsymbol{N}^{obs};\theta\right)}{L_{marg}\left(\boldsymbol{N}^{obs};\hat{\theta}\right)},$$

where $L_{marg}\left(\boldsymbol{N}^{obs};\theta\right) := \int L_s\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)\pi(\boldsymbol{\eta})d\boldsymbol{\eta};\quad \hat{\theta} = \arg\max_{\theta} L_{marg}\left(\boldsymbol{N}^{obs};\theta\right)$

- We need to know the pdf for λ for different true values of θ:

1) Calculate it by sampling $\boldsymbol{N}^{obs}$

2) Use a theorem to find out distribtuion of λ without any calculations

Part IV: T2K approach to claim the results

2) Use a theorem to find out distribution of λ without any calculations

Wilks' theorem

Wilks' theorem

Let's assume that we have two hypotheses: H_0 and H_1.
$H_0 \Leftrightarrow \xi \in \Xi_0$ and $H_1 \Leftrightarrow \xi \in \Xi_1$. Let's also define statistic λ:

$$\lambda\left(\boldsymbol{N}^{obs}\middle|\xi\right) = -2\ln\frac{\max\limits_{\xi\in\Xi_1} L\left(\boldsymbol{N}^{obs};\xi\right)}{\max\limits_{\xi\in\Xi_0} L\left(\boldsymbol{N}^{obs};\xi\right)}$$

If:

1) the maximum likelihood estimators of the parameters have ellipsoidal distributions (no physical boundaries, no degeneracies etc)
2) $\Xi_1 \subset \Xi_0$(nested hypotheses)
3) Data sample size $N \rightarrow \infty$

Then :

$$\boldsymbol{\lambda \sim \chi^2 \text{ with } k \text{ ndof},}$$

where $k = \dim(\Xi_0) - \dim(\Xi_1)$

Wilks' theorem

$$\boldsymbol{\xi} = (\theta, \boldsymbol{\eta})$$

$$\lambda\left(\boldsymbol{N}^{obs}\middle|\xi\right) = -2\ln\frac{\max_{\xi\in\Xi_1} L\left(\boldsymbol{N}^{obs};\xi\right)}{\max_{\xi\in\Xi_0} L\left(\boldsymbol{N}^{obs};\xi\right)}$$

$\boldsymbol{\lambda}\sim\boldsymbol{\chi}^2$ with $\boldsymbol{k}$ ndof, $\quad k = \dim(\Xi_0) - \dim(\Xi_1)$

- How is it related to our case? What is Ξ_0 and Ξ_1?

- Answer: in our case we should take $\Xi_1 = \{Space\ of\ nuisance\ parameters\}$
  $\Xi_0 = \{Full\ parameter\ space\}$

So,

$$\lambda\left(\boldsymbol{N}^{obs}\middle|\theta\right) = -2\ln\frac{\max_{\boldsymbol{\eta}} L\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)}{\max_{\theta,\boldsymbol{\eta}} L\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)}$$

Wilks' theorem

$$\lambda\left(\boldsymbol{N}^{obs}\middle|\theta\right) = -2\ln\frac{\max\limits_{\boldsymbol{\eta}} L\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)}{\max\limits_{\theta,\boldsymbol{\eta}} L(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta})}$$

We got log-profiled likelihood ratio!
So, if Wilks' theorem conditions are satisfied then the log-profiled likelihood asymptotically converges to χ^2

Wilks' theorem

$$\lambda_p\left(\boldsymbol{N}^{obs}\middle|\theta\right) = -2\ln\frac{\max\limits_{\boldsymbol{\eta}} L\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)}{\max\limits_{\theta,\boldsymbol{\eta}} L\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)} \sim \boldsymbol{\chi}^{2}$$

So, if Wilks' theorem conditions are satisfied we can avoid multiple toys fit
We don't need to **calculate** test statistic distribution – **it is known** – it is χ^2

- Denis, wait but Wilks' theorem works for **profiled log-likelihood ratio**, but as you mentioned T2K uses **marginalised log-likelihood ratio.** How can we use then the theorem?
- **Answer: There is second theorem: under exactly the same contitions of Wilks' theorem $\lambda_m \sim \lambda_p$ for $N \to \infty$**

$$\lambda_m\left(\boldsymbol{N}^{obs}\middle|\theta\right) = -2\ln\frac{L_{marg}\left(\boldsymbol{N}^{obs};\theta\right)}{L_{marg}\left(\boldsymbol{N}^{obs};\hat{\theta}\right)} \sim -2\ln\frac{\max\limits_{\boldsymbol{\eta}} L\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)}{\max\limits_{\theta,\boldsymbol{\eta}} L\left(\boldsymbol{N}^{obs};\theta,\boldsymbol{\eta}\right)} = \lambda_p\left(\boldsymbol{N}^{obs}\middle|\theta\right)$$

Wilks' theorem

Thus, under Wilks' theorem conditions:

$$\lambda_m\left(\boldsymbol{N}^{obs}\middle|\theta\right) = -2\ln\frac{L_{marg}\left(\boldsymbol{N}^{obs};\theta\right)}{L_{marg}\left(\boldsymbol{N}^{obs};\hat{\theta}\right)} \sim \boldsymbol{\chi}^2$$

- **$\boldsymbol{\delta_{CP}}, \boldsymbol{\sin^2\theta_{23}},$ MH do not satisfy the Wilks' theorem conditions that's we explicitly calculate λ_m distribution**

- Why?

1) Cyclic nature of δ_{CP} introduces effective boundaries ($|\sin\delta_{CP}| < 1$)
2) Other boundaries: $\boldsymbol{\sin^2 2\theta_{23} < 1}$
3) Degeneracies: sign($\cos\delta_{CP}$) vs MH, octant degeneracy etc.
4) Small number of events in appearance channel
5) MH do not form nested hypotheses: $IH \notin NH$

Wilks' theorem

Thus, under Wilks' theorem conditions:

$$\lambda_m(\boldsymbol{N}^{obs}|\theta) = -2\ln\frac{L_{marg}(\boldsymbol{N}^{obs};\theta)}{L_{marg}(\boldsymbol{N}^{obs};\hat{\theta})} \sim \boldsymbol{\chi}^2$$

- $\delta_{CP}, \sin^2\theta_{23}, \mathrm{MH}$ do not satisfy the Wilks' theorem
- **Only for Δm^2_{32} the Wilks' theorem can be applied**

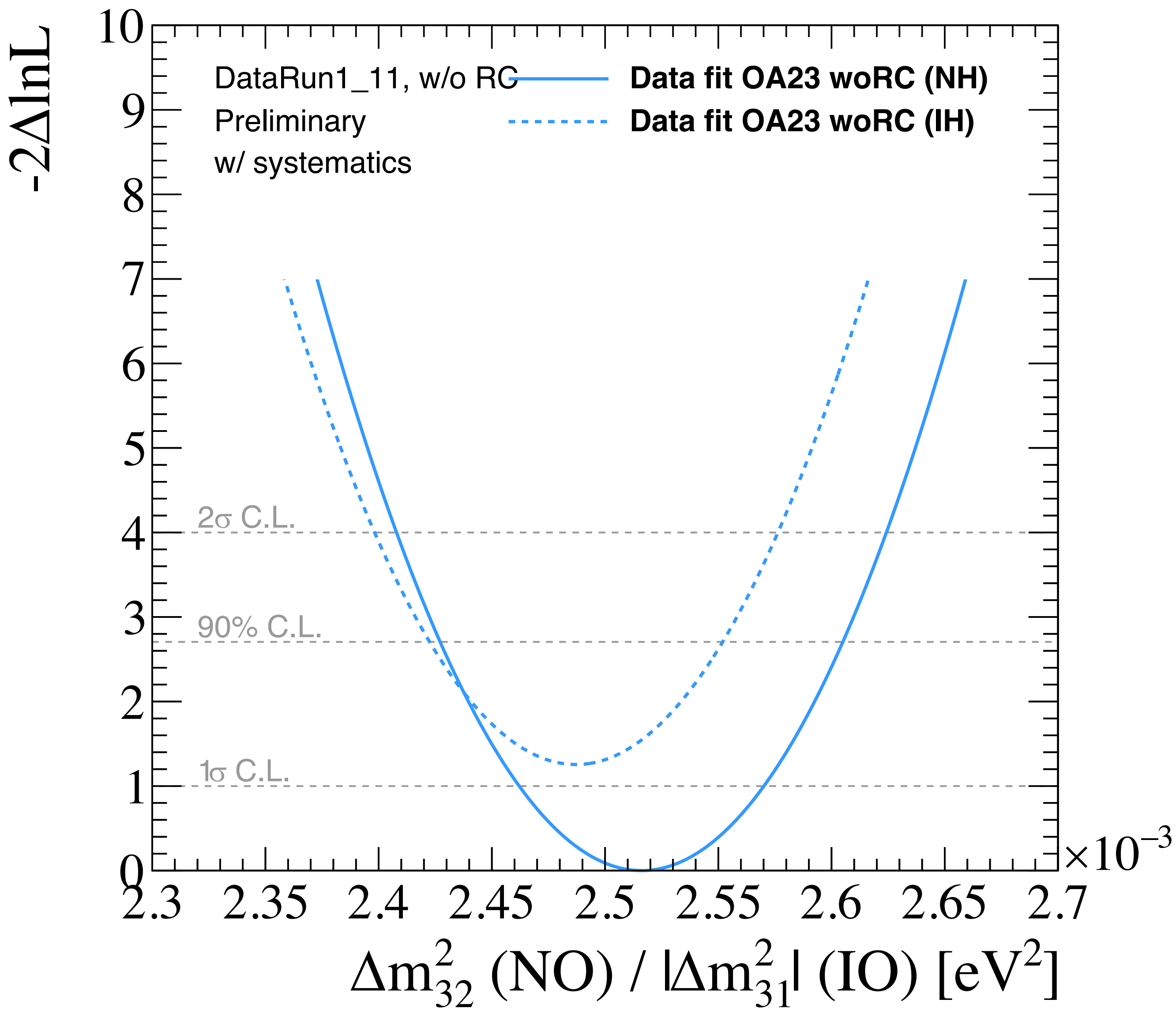
Thus, for δ_{CP} and $\sin^2\theta_{23}$ T2K calculates the critical values, performs explicit Neyman contruction

For Δm^2_{32} we apply Wilks' theorem according to which the Neyman belt will be a rectangle with width $\sqrt{z}$, where z is number of sigmas of corresponding C.L. (remember our example)

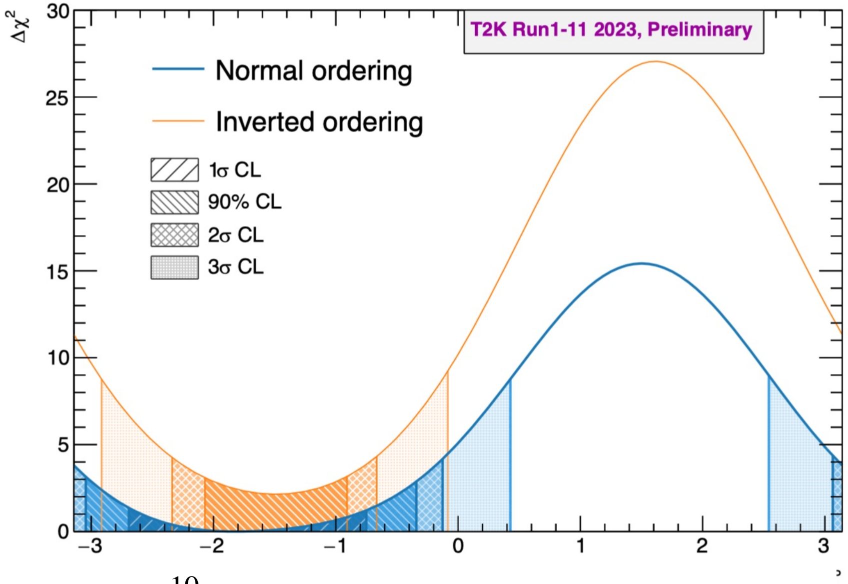
Part IV: T2K approach to claim the results

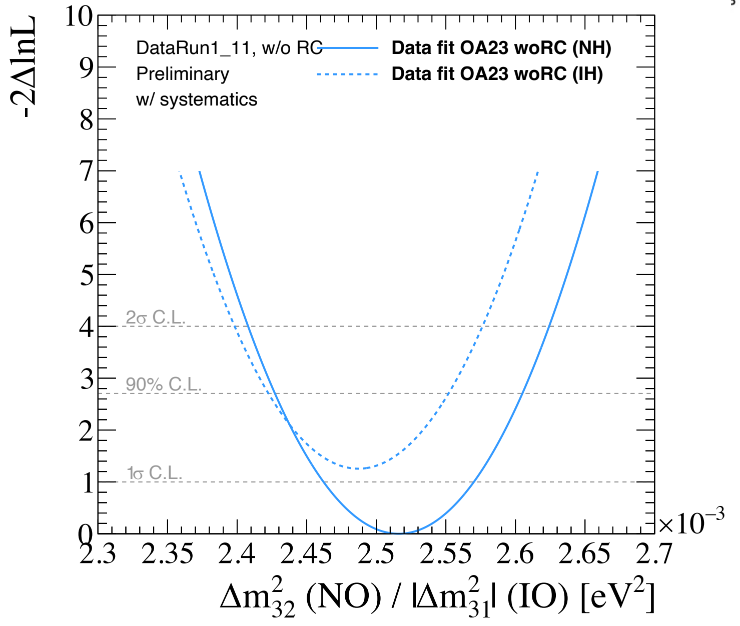
Δm^2_{32}

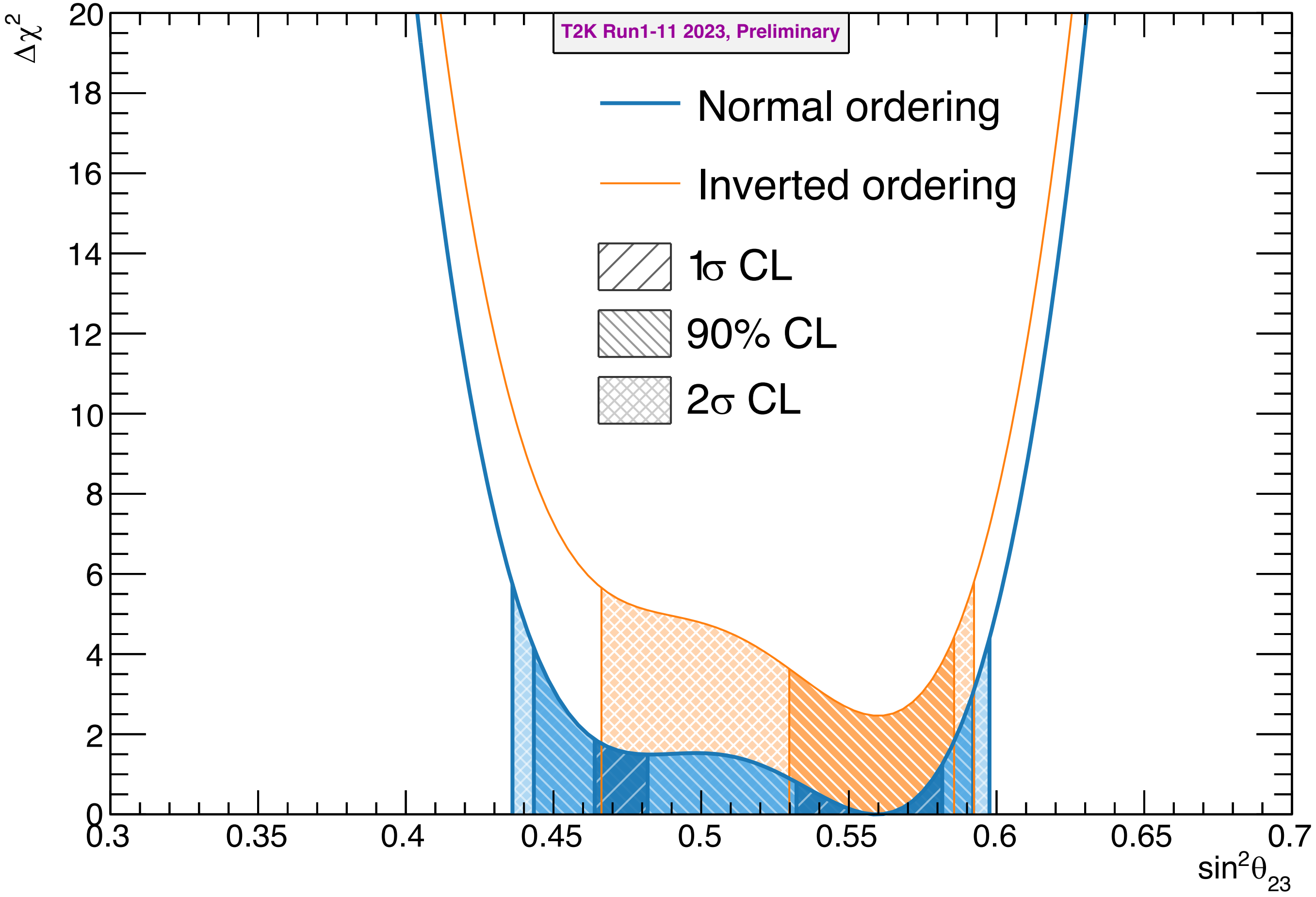
- Critical value are constant, not dependent on δ_{CP}
- Just take $\Delta\chi^2 = 1, 4, 9$ for C.L. 1σ, 2σ and 3σ corr.
- So, repeating for Δm^2_{32} the procedure is the same, we just a priori know the critical values



Part IV: T2K approach to claim the results







What about mass ordering?

To test MH hypothesis T2K calculates p-values

Mass ordering studies

- The **p-value** of a hypothesis test is a measure of the strength of evidence against the null hypothesis.
- It represents the probability of obtaining a test statistic as extreme, or more extreme than, the one observed in the data, assuming that the null hypothesis is true.
- In hypothesis testing, typically, a smaller p-value suggests stronger evidence against the null hypothesis, leading to its rejection in favor of an alternative hypothesis.

For mass ordering tests:

- Test statistic: $\boldsymbol{\lambda(H) := \chi^2_{NH} - \chi^2_{IH}}$, **where** $\boldsymbol{H \in \{\text{True NH}, \text{True IH}\}}$

$$p_0(\mathrm{NH}) = \int_{\lambda_D}^{+\infty} P(\lambda(NH))d\lambda -$$

we take **right tail** because for NH high λ are more extreme

$$p_0(\mathrm{IH}) = \int_{-\infty}^{\lambda_D} P(\lambda(IH))d\lambda -$$

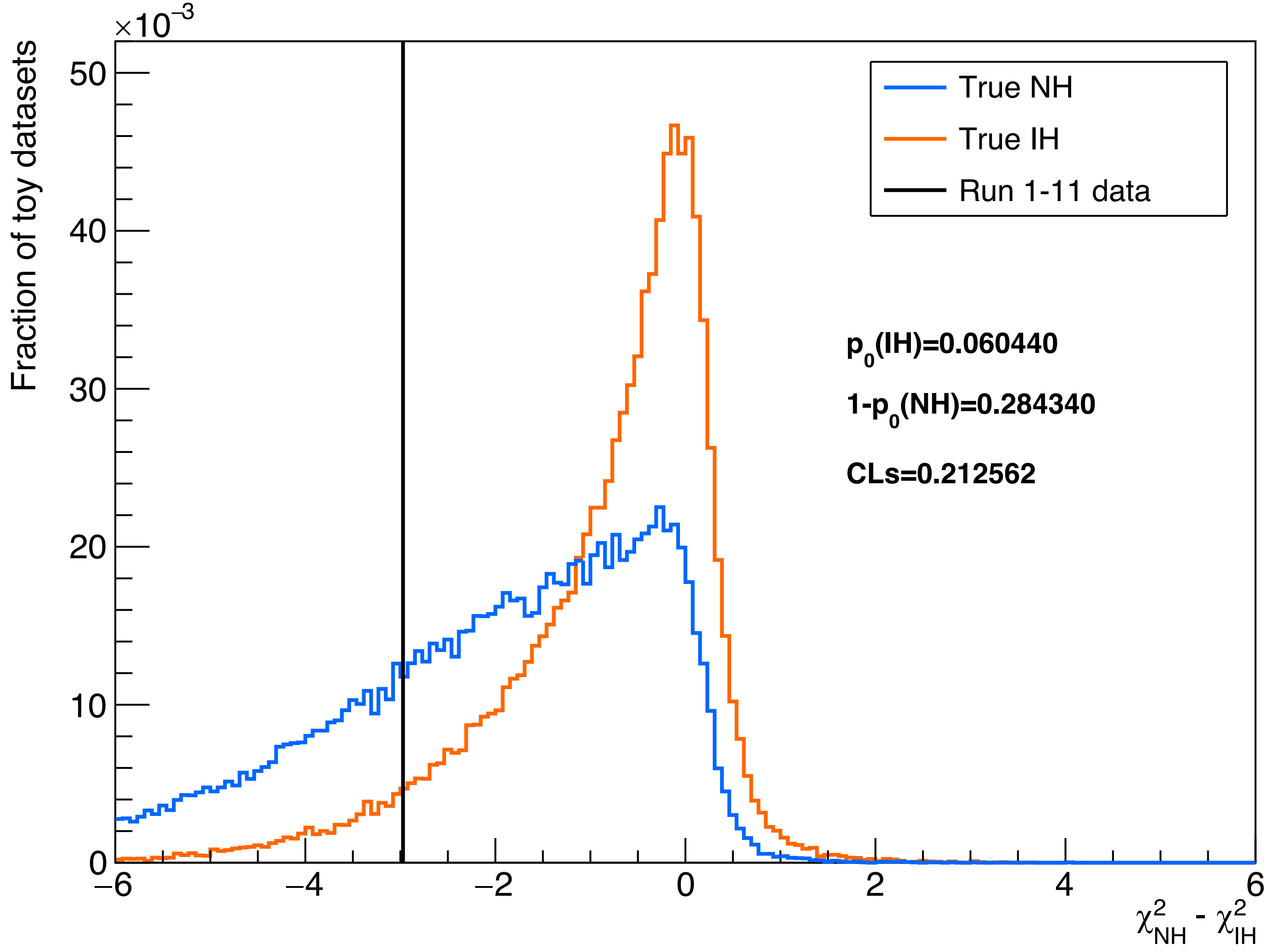
we take **left tail** because for IH low λ are more extreme

Mass ordering studies

- Again the question: how to generate the toys? How to sample nuisance parameters?
- In T2K the sampling in done using two ways: for fixed δ_{CP} and not fixed δ_{CP}
- In case of fixed δ_{CP} we use the same toys used for FC Neyman construction for δ_{CP}
- In case of not-fixed δ_{CP}:
  1) syst. parameters thrown according prior
  2) δ_{CP} and $\sin^2\theta_{23}$ sampled from 2D posterior
  3) Δm^2_{32} from 1D posterior

Mass ordering studies

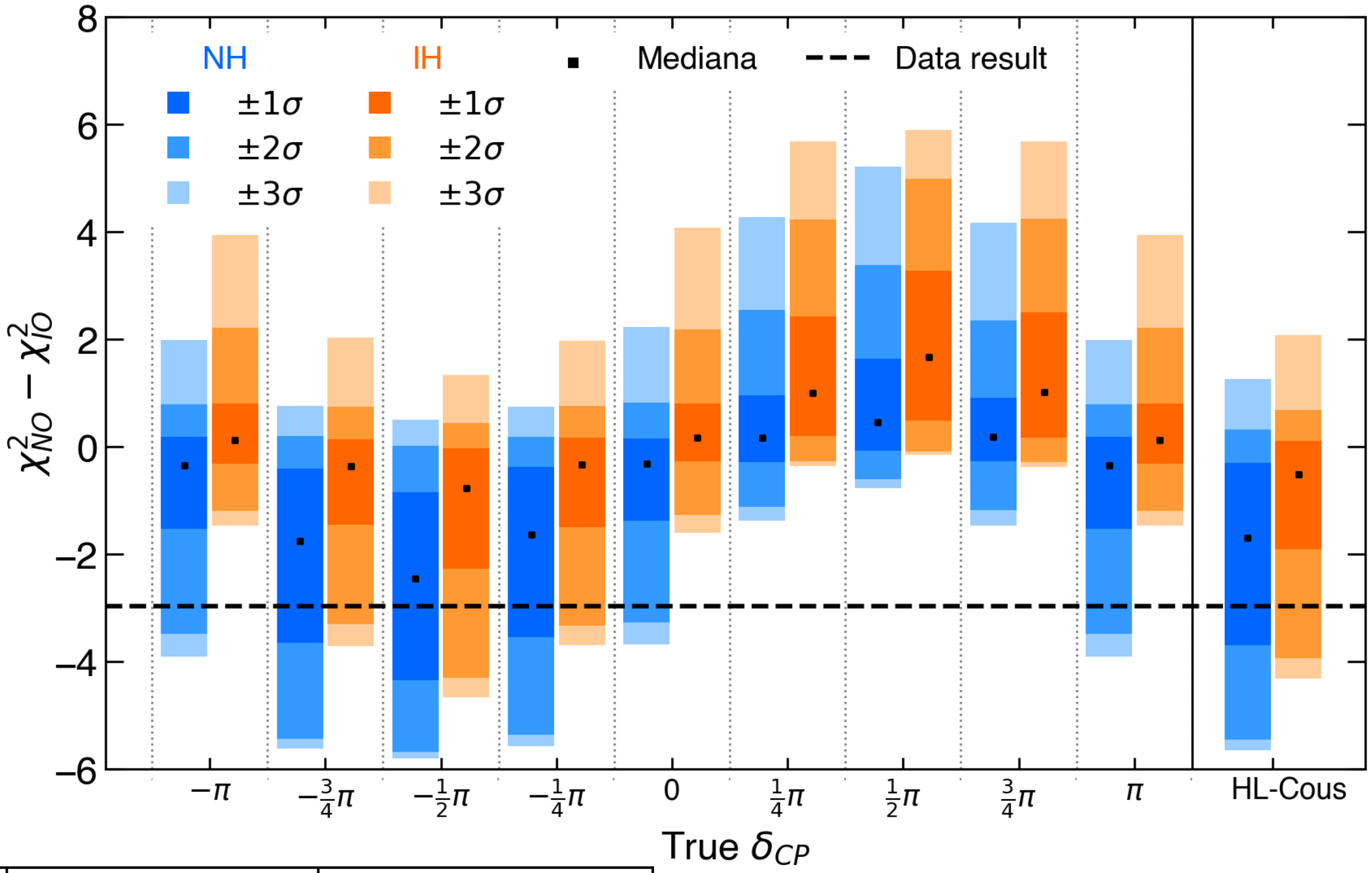
- Test statistic: $\lambda(H) := \chi^2_{NH} - \chi^2_{IH}$, **where** $H \in \{NH, IH\}$



- Potential p-value problem: p_0 can be low for both hypothesis simultaneously
- To deal with this problem sometimes renormalised p-value is used called CLs:

$$CLs := \frac{p_0(\mathrm{IH})}{1 - p_0(\mathrm{NH})}$$

Mass ordering studies



	Fixed $\delta_{CP} = -\frac{\pi}{2}$	Highland-Cousins
$p-\text{value}(IH)$	1.69σ	1.88σ
$CLs(IH)$	1.26σ	1.25σ

Final words

- I didn’t covered many other aspects, such as another possible ways to report CP-conservation p-values, GoF p-values, octant CLs etc

- I hope that this lecture gave a general picture on main statistics methods in T2K

Summary

Now we finally can have a lunch!