Statistical Data Analysis course

In-depth part : Profile Likelihood Ratio & `MINOS` interval

Prof. Alexis Pompili (University of Bari Aldo Moro)*

* alexis.pompili@ba.infn.it

https://persone.ict.uniba.it/rubrica/alexis.pompili

Connection between Profile Likelihood Ratio & `MINOS` uncertainties

Profile Likelihood

In the next slides this connection will be argued/explained.
Firstly, remember the difference between these two concepts:

- POI(s) = **parameter(s) of interest**: parameter(s) of theoretical model (we assume predicts distribution of observed variables)

- NPs = **nuisance parameters** : additional unknown parameters, appearing together with the POI(s), that represent the effect of the detector response (resolutions, miscalibrations, …), the presence of background, …
Typically they can represent systematic uncertainties & can be usually determined from simulation or data control samples.

Let's assume for simplicity to have a POI $\boldsymbol{\mu}$ and a set of NPs $\vec{\boldsymbol{\theta}}$ (i.e.all parameter are treated as NPs with exception of $\boldsymbol{\mu}$).
The **likelihood function** is written as: $\mathcal{L}(\vec{x}; \boldsymbol{\mu}, \vec{\boldsymbol{\theta}})$. To easy the notation we drop the $\vec{x}$ and write simply $\mathcal{L}(\boldsymbol{\mu}, \vec{\boldsymbol{\theta}})$.

The so-called **profile likelihood** is constructed following this prescription:

- for a *given* value of the POI $\bar{\boldsymbol{\mu}}$ derive the **ML estimates** $\hat{\hat{\vec{\boldsymbol{\theta}}}}(\bar{\boldsymbol{\mu}})$ (it's a *conditional ML estimate*; fit with $\boldsymbol{\mu}$ fixed to a constant value $\bar{\boldsymbol{\mu}}$)

- thus the maximum likelihood for a given value of $\bar{\boldsymbol{\mu}}$ is $\mathcal{L}_{max}(\bar{\boldsymbol{\mu}}, \hat{\hat{\vec{\boldsymbol{\theta}}}}(\bar{\boldsymbol{\mu}}))$;

- recalculating (CPU expensive) for each value of $\boldsymbol{\mu}$ (scan of $\boldsymbol{\mu}$ values) we get a truly function of $\boldsymbol{\mu}$: $\mathcal{L}_{max}(\boldsymbol{\mu}, \hat{\hat{\vec{\boldsymbol{\theta}}}}(\boldsymbol{\mu}))$
which is the **likelihood function maximized w.r.t. all the NPs and** is called **profile likelihood** !

Profile Likelihood Ratio

On the other hand it is always possible to maximize the likelihood getting the best estimates (fit values) of $\boldsymbol{\mu}$ and $\vec{\boldsymbol{\theta}}$ corresponding to the observed data $\vec{\boldsymbol{x}}$ **:** $\hat{\boldsymbol{\mu}}$ and $\hat{\vec{\theta}}$. Thus the maximized likelihood is: $\mathcal{L}_{max}(\hat{\boldsymbol{\mu}}, \hat{\vec{\boldsymbol{\theta}}})$

At this point we can consider the **Profile Likelihood ratio** : $\boldsymbol{\lambda}(\boldsymbol{\mu}) = \dfrac{\mathcal{L}_{max}(\boldsymbol{\mu}, \hat{\hat{\vec{\boldsymbol{\theta}}}}(\boldsymbol{\mu}))}{\mathcal{L}_{max}(\hat{\boldsymbol{\mu}}, \hat{\vec{\boldsymbol{\theta}}})}$ (that does not depend on the NPs $\vec{\boldsymbol{\theta}}$)

This ratio is used in the convenient test statistic $\boldsymbol{t}_{\boldsymbol{\mu}} = -\mathbf{2}\ \boldsymbol{ln}\ \boldsymbol{\lambda}(\boldsymbol{\mu})$. Dropping the obvious “max” index: $\boldsymbol{\lambda}(\boldsymbol{\mu}) = \dfrac{\mathcal{L}(\boldsymbol{\mu}, \hat{\hat{\vec{\boldsymbol{\theta}}}}(\boldsymbol{\mu}))}{\mathcal{L}(\hat{\boldsymbol{\mu}}, \hat{\vec{\boldsymbol{\theta}}})}$

In other words the profile likelihood ratio substitutes the ordinary likelihood ratio, in the test statistics $\boldsymbol{t}_{\boldsymbol{\mu}} = -\mathbf{2}\ \boldsymbol{ln}\ \boldsymbol{\lambda}(\boldsymbol{\mu})$, **when** we have to deal with nuissance parameters:

Likelihood for a given μ

$$\boldsymbol{\lambda}(\boldsymbol{\mu}) = \frac{\mathcal{L}(\boldsymbol{\mu})}{\mathcal{L}(\hat{\boldsymbol{\mu}})} \Rightarrow \boldsymbol{\lambda}(\boldsymbol{\mu}) = \frac{\mathcal{L}(\boldsymbol{\mu}, \hat{\hat{\vec{\boldsymbol{\theta}}}}(\boldsymbol{\mu}))}{\mathcal{L}(\hat{\boldsymbol{\mu}}, \hat{\vec{\boldsymbol{\theta}}})}$$

Maximum Likelihood for a given μ

Maximum Likelihood

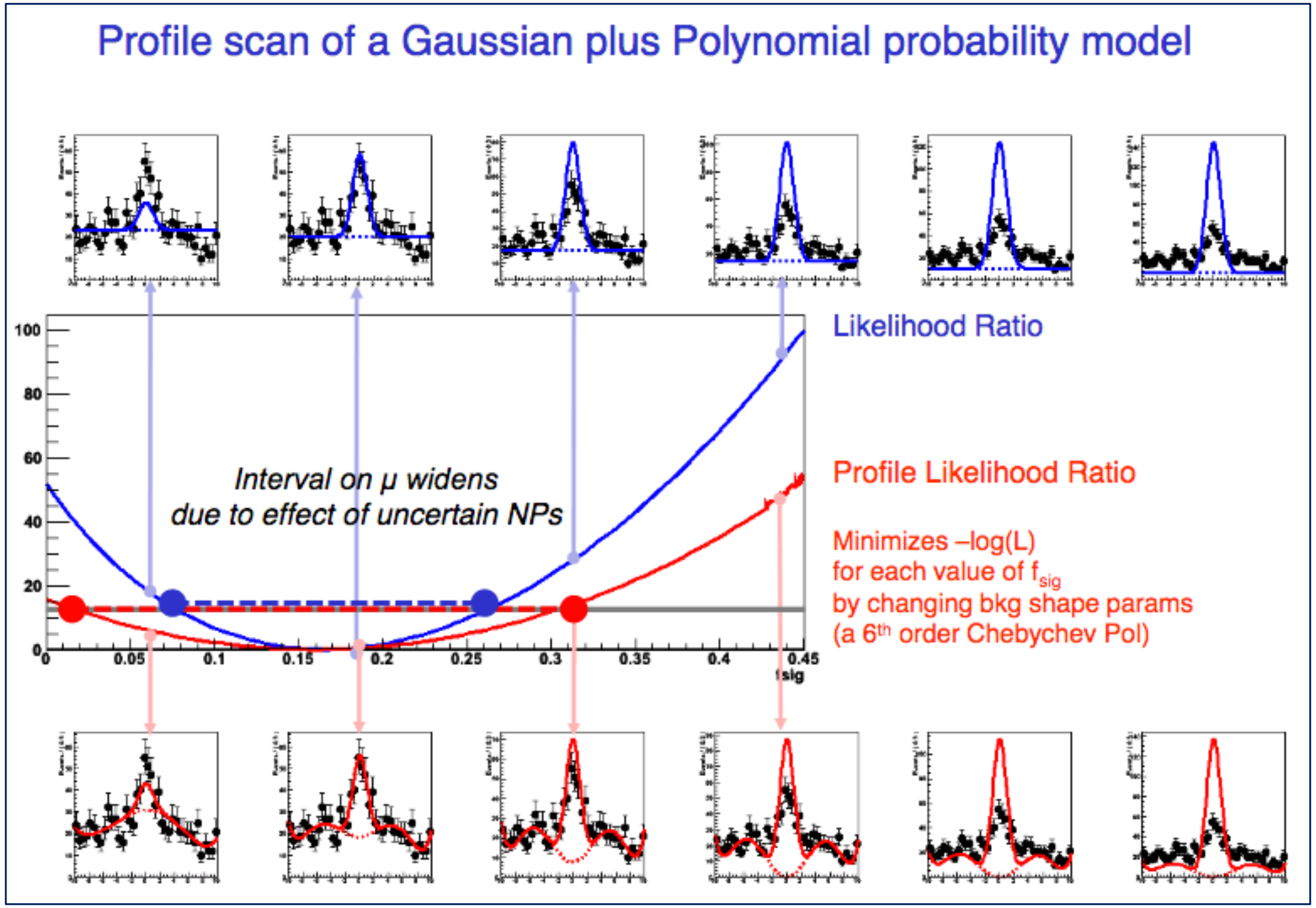
Maximum Likelihood

Comments on the Profile Likelihood approach:

- it is **computationally challenging** because it requires to perform the minimization of the likelihood w.r.t. **all** the nuisance parameters for every point in the profile likelihood curve (see also next slide that illustrates this)
- the minimization can be difficult because of the possibly strong correlation among POIs and NPs or multiple/local minima

How to obtain a Profile Likelihood

For visualization purposes have a look at this figure illustrating the scan of $\boldsymbol{\mu}$ values in order to obtain $\mathcal{L}(\boldsymbol{\mu},\hat{\hat{\vec{\boldsymbol{\theta}}}}(\boldsymbol{\mu}))$:



Profile Likelihood & Contours - I

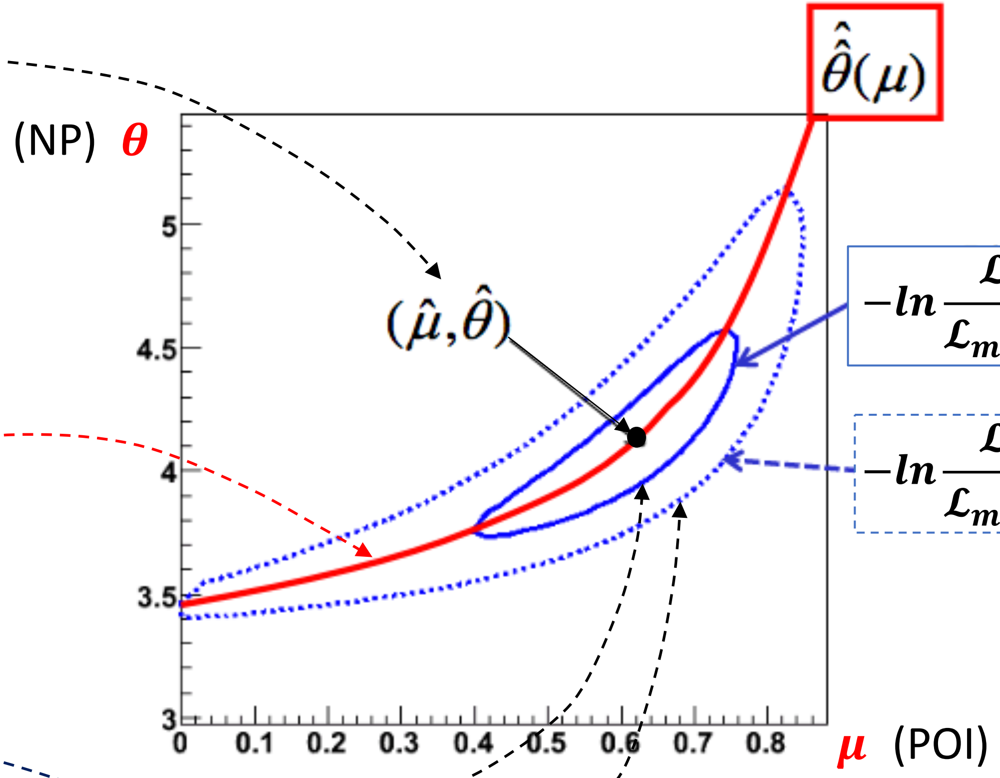
For illustration purposes let us consider one POI ($\boldsymbol{\mu}$) and one NP ($\boldsymbol{\theta}$) in order to visualize the profiling.

Firstly let us design/locate the **(black) point** representing their best estimates that maximize the likelihood: $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\theta}})$

$$(\boldsymbol{\mu},\boldsymbol{\theta}) \equiv (\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\theta}}) \quad \Rightarrow \quad \mathcal{L} \equiv \mathcal{L}_{max}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\theta}})$$

Secondly let's design the **red curve** that represents those points $(\boldsymbol{\mu},\hat{\hat{\boldsymbol{\theta}}})$ for which $\mathcal{L} \equiv \mathcal{L}_{max}(\boldsymbol{\mu},\hat{\hat{\boldsymbol{\theta}}}(\boldsymbol{\mu}))$

…corresponding to a subset of subsequently given/fixed values of $\boldsymbol{\mu}$ which includes also the special value $\bar{\boldsymbol{\mu}} \equiv \hat{\boldsymbol{\mu}}$.



$$-ln\frac{\mathcal{L}(\boldsymbol{\mu},\boldsymbol{\theta})}{\mathcal{L}_{max}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\theta}})} = \frac{1}{2}$$

$$-ln\frac{\mathcal{L}(\boldsymbol{\mu},\boldsymbol{\theta})}{\mathcal{L}_{max}(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\theta}})} = 2$$

Finally we can design the contour curves at 1σ and 2σ with respect to $(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\theta}})$ the maximum (minimum) of the (negative) likelihood. This is discussed in detail in the next slide.

Profile Likelihood & Contours - II

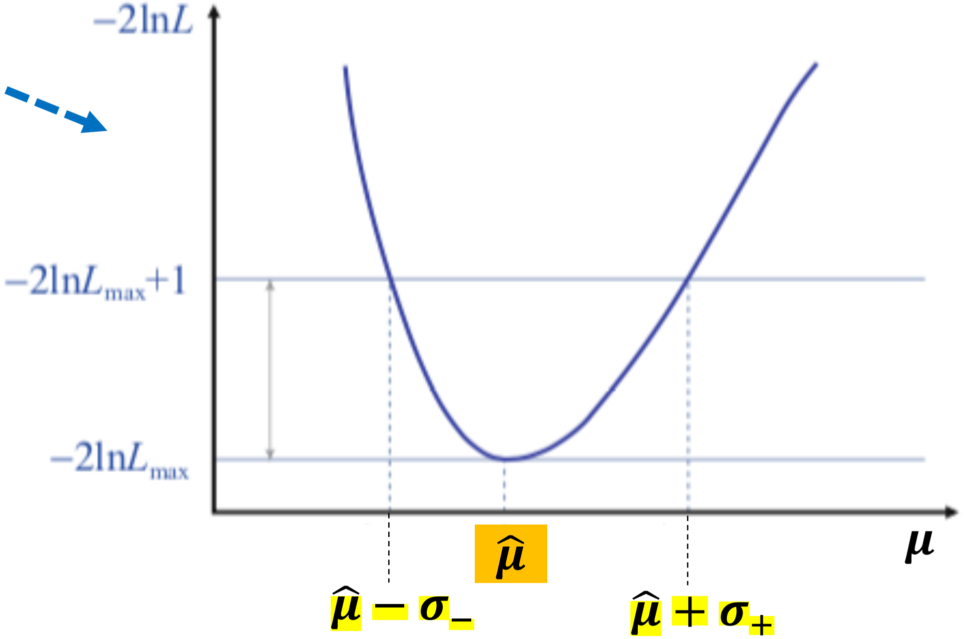
In particular the first contour corresponds to a set of parameters such that: $-2\ln\mathcal{L}(\mu,\theta) = -2\ln\mathcal{L}_{max}(\hat{\mu},\hat{\theta}) + 1$

Indeed in the simplest case (only one POI & no NPs) one has:

$$-2\ln\mathcal{L}(\mu) \equiv -2\ln\mathcal{L}_{max}(\hat{\mu}) + 1$$

$$\Longleftrightarrow \quad 2\ln\mathcal{L}(\mu) - 2\ln\mathcal{L}_{max}(\hat{\mu}) = -1$$

$$\Longleftrightarrow \quad 2ln\frac{\mathcal{L}(\mu)}{\mathcal{L}_{max}(\hat{\mu})} = -1 \quad \Longleftrightarrow \quad -ln\frac{\mathcal{L}(\mu)}{\mathcal{L}_{max}(\hat{\mu})} = +\frac{1}{2}$$



Note that in general **the** **uncertainty** (and thus the 1σ interval) **can be** **asymmetric** (as depicted).

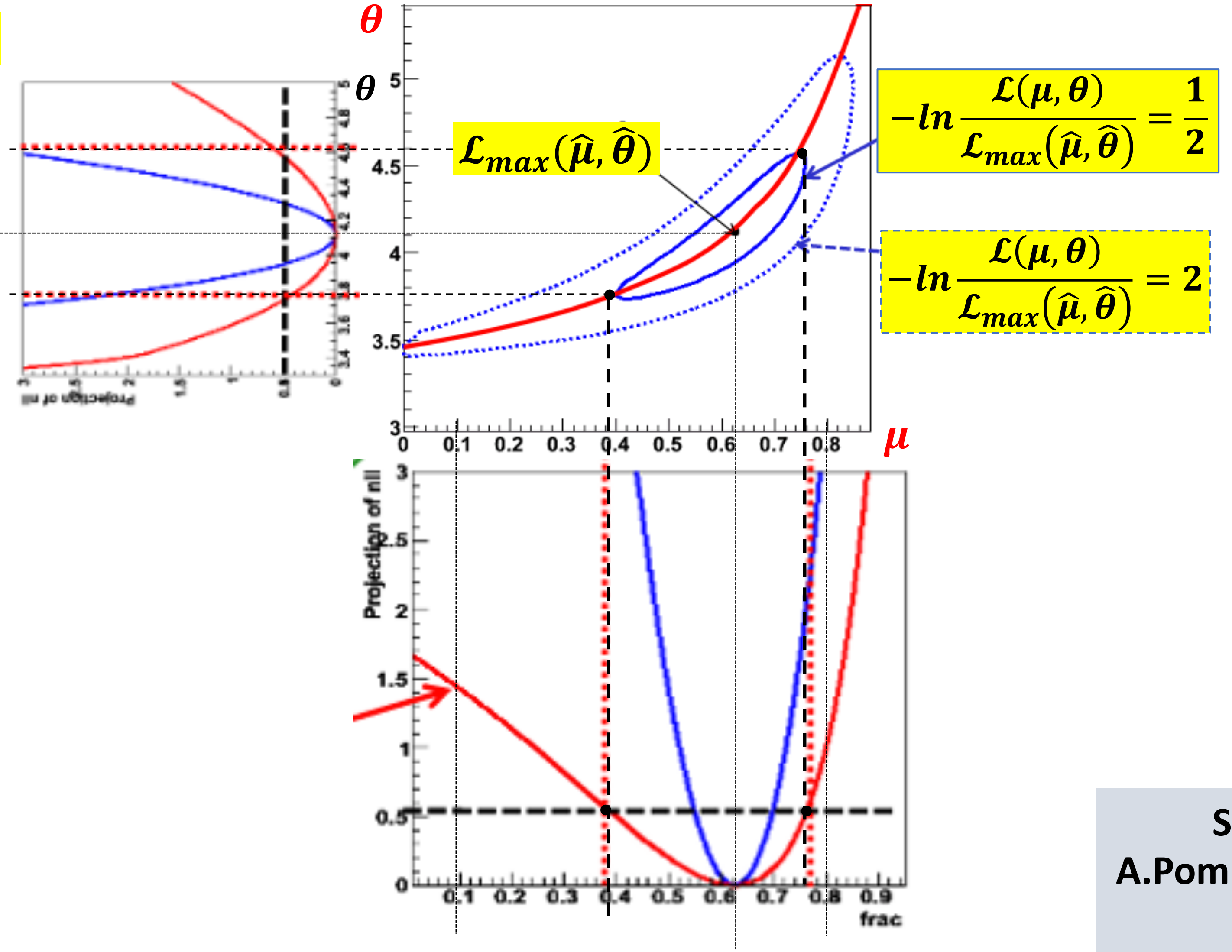
When there is also one NP one gets 2D contours (see next slide) and

$$-ln\frac{\mathcal{L}(\mu)}{\mathcal{L}_{max}(\hat{\mu})} \quad \text{becomes} \quad -ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$$

Profile Likelihood & Contours - III

When there is also one POI and one NP one gets **2D contours**, here designed with the 2 projections:

Clearly **the correct 1σ interval for the POI is given by the projection of the contour** (and not by the - marginalized - likelihood, that is the blue projection, which ignores the effect of the presence of the NP). It can be demonstrated that **this *confidence interval* provides the correct *coverage* in the frequentistic approach**.



$\mathcal{L}_{max}(\hat{\mu},\hat{\theta})$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=\frac{1}{2}$$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=2$$

The overall uncertanty is in general asymmetric!

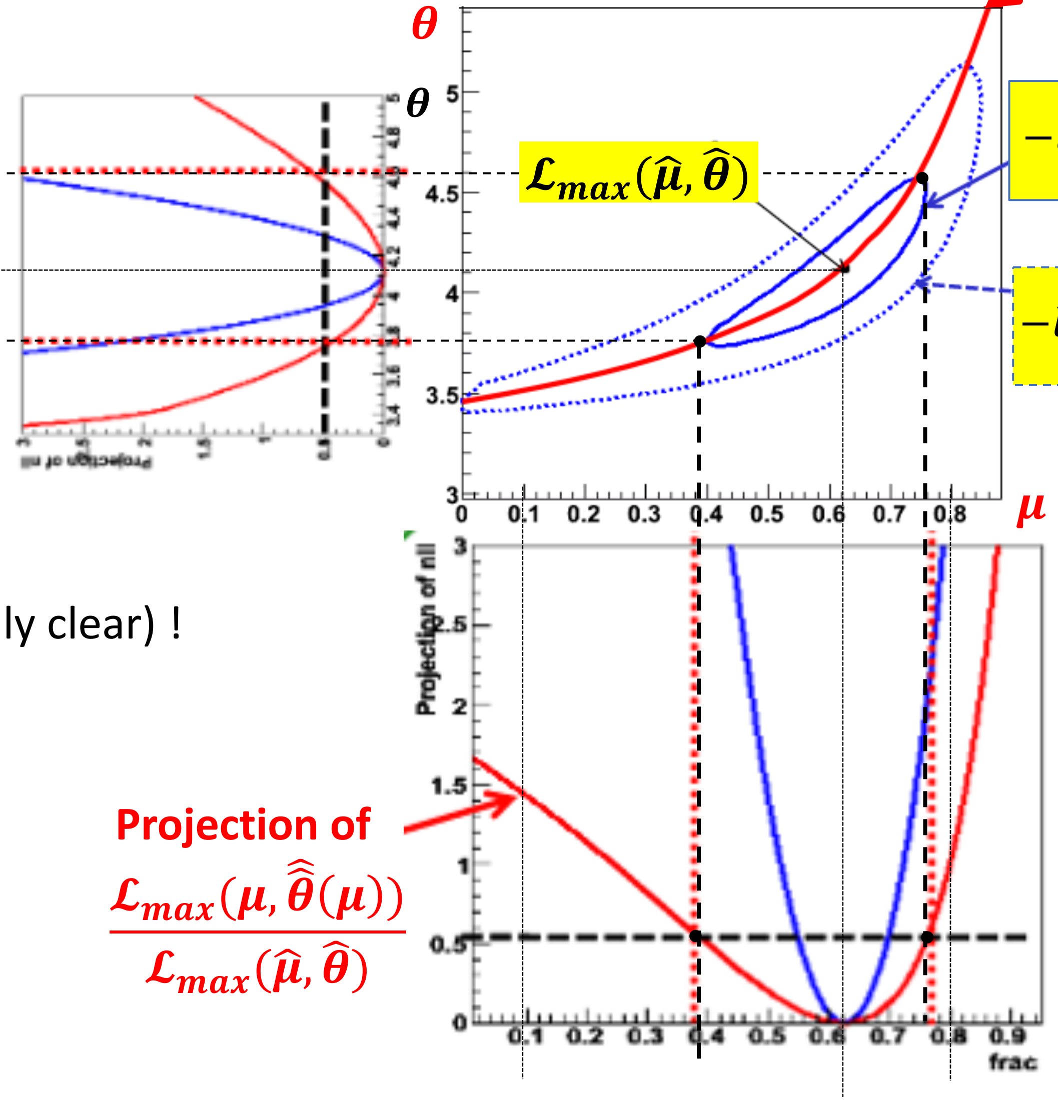
Same confidence interval provided by Profile Likelihood & Contours

When there is also one POI and one NP one gets **2D contours**, here designed with the 2 projections:

Clearly **the correct 1σ interval for the POI is given by the projection of the contour** (and not by the - marginalized - likelihood, that is the blue projection, which ignores the effect of the presence of the NP). It can be demonstrated that **this *confidence interval* provides the correct *coverage* in the frequentistic approach**.

It is also crucial to know that **this interval is the same provided by (the projection of) the Profile Likelihood ratio** based on $\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))$ (as visually clear) ! Indeed the additon of NP(s) broadens the shape of the Profile Likelihood as a function of the POI compared with the case where NP(s) are not added. As a consequence, the uncertainty on the POI increases when NPs - that usually model sources of systematic uncertainties - are included.

The overall uncertanty is in general asymmetric!



$$\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$$

$$\mathcal{L}_{max}(\hat{\mu},\hat{\theta})$$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=\frac{1}{2}$$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=2$$

Projection of

$$\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$$

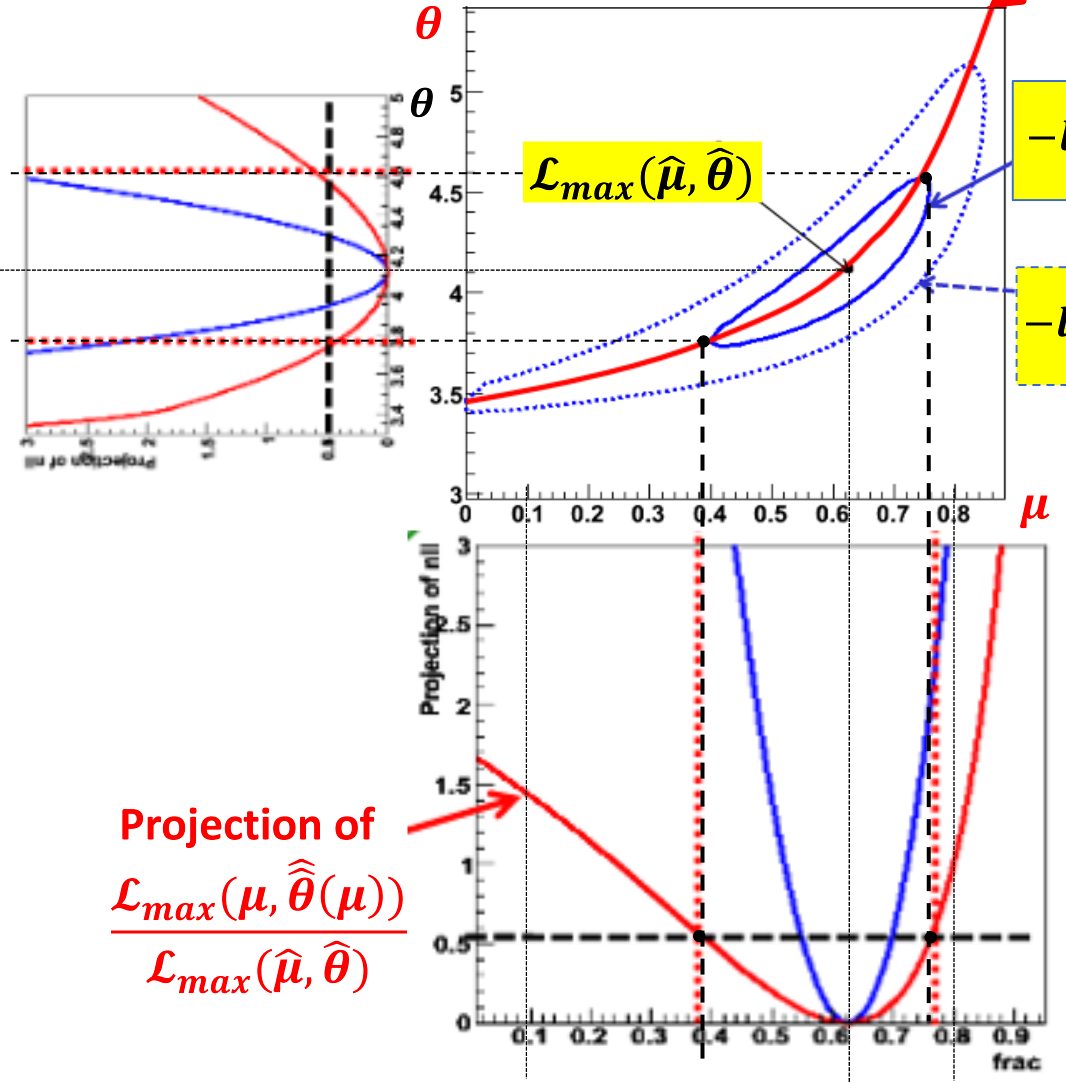
MINOS uncertainties by likelihood scan

When there is also one POI and one NP one gets **2D contours**, here designed with the 2 projections:

$$\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$$

Moreover :

The `MINOS` method (called by `MINUIT`) determines the overall uncertainties (in general asymmetric) **based on the *likelihood scan* namely on the $-2\ln\mathcal{L}(\mu)$ scan used to determine the 1σ contour.**



$\mathcal{L}_{max}(\hat{\mu},\hat{\theta})$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=\frac{1}{2}$$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=2$$

Projection of

$$\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$$

MINOS uncertainties by likelihood scan

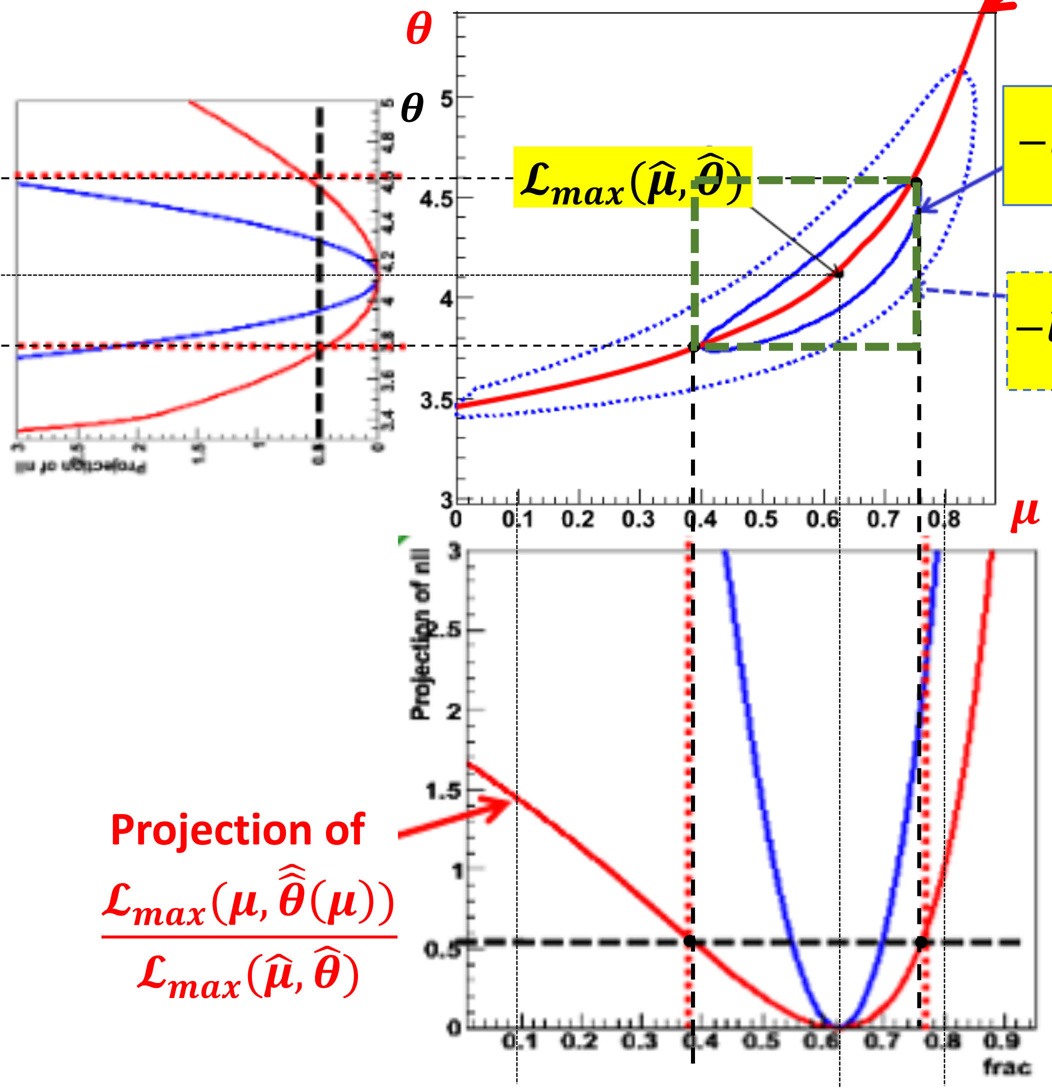
When there is also one POI and one NP one gets **2D contours**, here designed with the 2 projections:

Moreover :

The MINOS method (called by MINUIT) determines the overall uncertainties (in general asymmetric) **based on the *likelihood scan* namely on the $-2\ln\mathcal{L}(\mu)$ scan used to determine the 1σ contour.**

The MINOS errors can be visualized with the size of the green bounding box around the contour given by $-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=\frac{1}{2}$ **!**

$$\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$$



$\mathcal{L}_{max}(\hat{\mu},\hat{\theta})$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=\frac{1}{2}$$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=2$$

Projection of $\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$

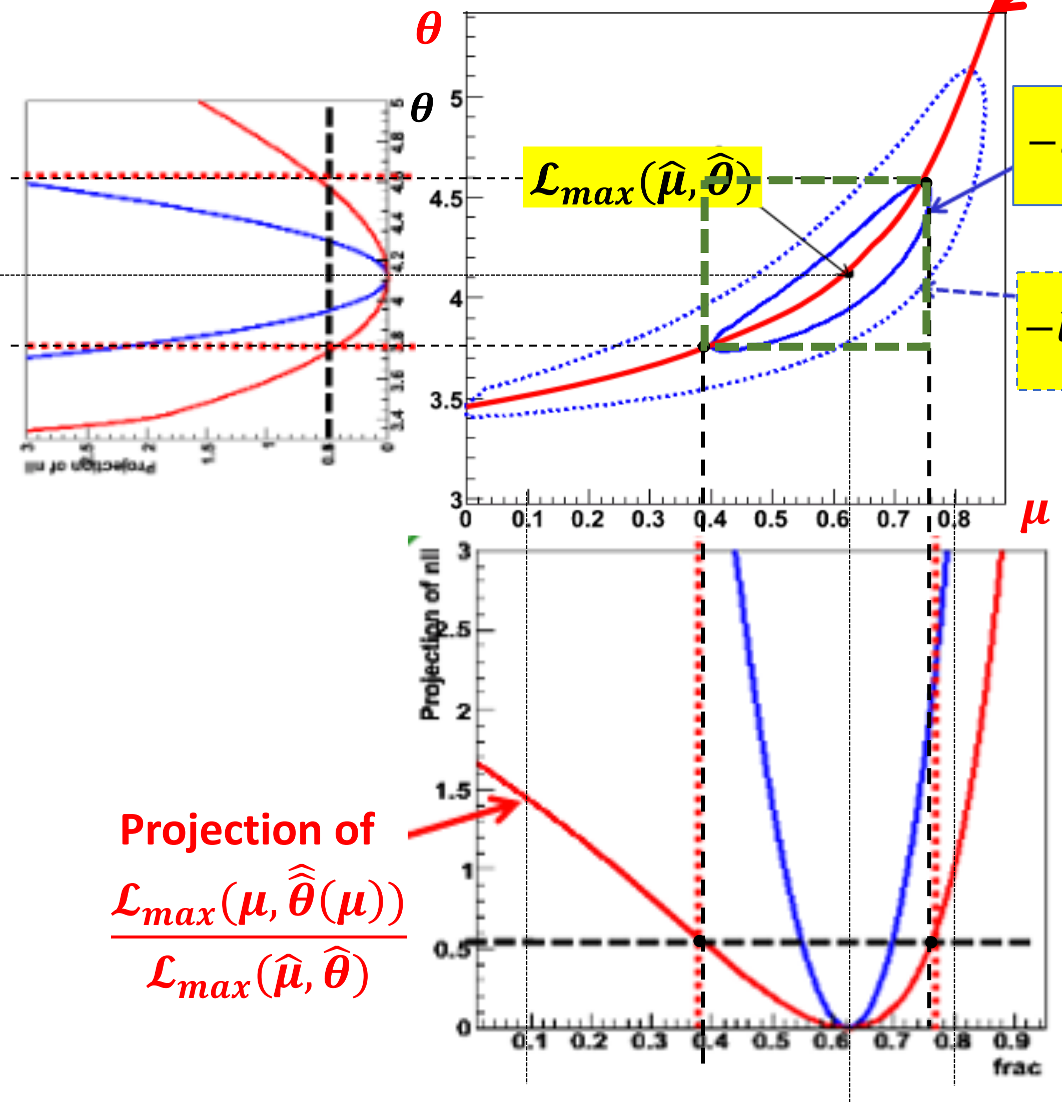
MINOS uncertainties by likelihood scan

When there is also one POI and one NP one gets **2D contours**, here designed with the 2 projections:

Moreover :

The MINOS method (called by MINUIT) determines the overall uncertainties (in general asymmetric) **based on the *likelihood scan* namely on the $-2\ln\mathcal{L}(\mu)$ scan used to determine the 1σ contour.**
The MINOS errors can be visualized with the size of the green bounding box around the contour given by $-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=\frac{1}{2}$!

In the gaussian case/regime $-2\ln\mathcal{L}(\mu)$ can have a parabolic shape and the uncertainty of the POI is symmetric, or "close to symmetric" if parabolic approximation is good and contour is an ellipsis. However - in general - the *coverage* in usually improved performing the likelihood scan instead of the parabolic approximation (given by HESSE).



$$\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$$

$\mathcal{L}_{max}(\hat{\mu},\hat{\theta})$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=\frac{1}{2}$$

$$-ln\frac{\mathcal{L}(\mu,\theta)}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}=2$$

Projection of $\frac{\mathcal{L}_{max}(\mu,\hat{\hat{\theta}}(\mu))}{\mathcal{L}_{max}(\hat{\mu},\hat{\theta})}$

Correspondence between MINOS uncertainties & Profile Likelihood intervals

Summarizing : the `MINOS` algorithm (*) provides the same (asymmetric) uncertainties given by the Profile Likelihood ratio

For both … the resulting confidence interval is satisfactorily "covered".

(*) The `MINOS` algorithm will be studied next year in the Laboratory Course (*Statistical Data Analysis Lab.*)

Let us remind that in the frequentist approach:

For a large fraction of repeated experiments - usually 68.27% - the unknown true value of μ is contained in the confidence interval $[\hat{\mu} - \sigma, \hat{\mu} + \sigma]$. The fraction is meant in the limit of infinitely large number of repetitions of the experiment , and $\hat{\mu}$ & σ may vary from on experiment to the other, being the result of a measurement in each experiment.

- Coverage: property of the estimated interval to contain the true value in 68.27% of the experiments.
- Confidence level : the reference *probability level* usually taken as 68.27%.
Interval estimates that have a larger (or smaller) probability of containing the true value, *compared to the desired confidence level*, are said to overcover (or undercover).

It is important to know that the resulting confidence interval from the Profile Likelihood construction will have exact coverage for the points $(\mu, \hat{\hat{\vec{\theta}}}(\mu))$; elewhere it might be over- or under- covering.

We conclude stating: in the asymptotic regime (very large number of experiments) the `MINOS` algorithm (in `ROOT`) provides the (asymmetric) uncertainties used in the definition of the frequentist confidence intervals !

Frequentist confidence intervals when NP are present

Exact confidence intervals are difficult when nuisance parameters are present:

- intervals should cover for any value of NPs (technically difficult)
- typically there can be a significant over-coverage

The approach to use the Profile Likelihood ratio guarantees the coverage at the measured values of NPs (only !)

- technically replace Likelihood ratio with Profile Likelihood ratio
- computationally more intensive but still very tractable

Asymptotically confidence intervals costructed with Profile Likelihood ratio correspond to `MINOS` likelihood ratios intervals

- as the distribution of the Profile Likelihood becomes asymptotically independent of θ the coverage for all values of θ is restored !