ESERCITAZIONE 12 - A.A.21/22

IN-DEPTH STUDIES about FITTING with ROOFIT

PROFILE LIKELIHOOD & MINOS

A.Pompili

Bibliography

Inspired by part of the theory visualization & exercises by **Wouter Verkerke** :
https://indico.cern.ch/event/72320/contributions/2082589/attachments/1037201/1478048/roofit-intro-roostats-v11a.pdf
https://indico.cern.ch/event/305391/contributions/701304/attachments/580262/798889/Verkerke_Statistics_L2.pdf

See also :
- his slides for the Ferrara School 2009: https://www.nikhef.nl/~verkerke/ferrara)
- his slides for IN2P3 School 2014: https://indico.in2p3.fr/event/9742/contributions/50419/attachments/40828/50594/sos2014_systprof_v38.pdf

Of course a good reference book is : **Luca Lista, Statistical methods for Data Analysis in Particle Physics**, Springer, 2nd Ed.

Retrieve binned data and plot

With reference to the code in the macro **yield.C** …

- Get the histogram of the **J/ψ**(μμ)**ϕ**(KK)**K** invariant mass :

```
TFile f1("DatasetAandB_KaonTrackRefit_Bwin_new_21aug13.root","READ");
TH1D *hist = (TH1D*)f1.Get("myJpsiKKKmass_all");
```

- Declare & initialize the variable to represent the invariant mass and prepare the corresponding RooPlot pointer:

```
RooRealVar y("y","y",5.15,5.45);
RooPlot* yframe = y.frame("");
```

- Import the binned data by creating the RooDataHist object from the histogram and plot it:

```
RooDataHist BmassExt(hist->GetName(),hist->GetTitle(),RooArgSet(y),RooFit::Import(*hist,kFALSE));
BmassExt.plotOn(yframe);
myC->cd();
yframe->Draw();
```

Build the negative log-lokelihood (`nll`)

Build the model: - a gaussian for the signal (2 parameters: mass and width) ;
- a Chebyshev of 2nd (2 parameters:) order for the background.

Based on these two PDFs, build the full PDF to make an **extended fit**:

```
RooRealVar nsig("nsig","sig fraction",500.,0.,5000.);
RooRealVar nbkg("nbkg","bkg fraction",2000.,0.,200000.);
//
RooAddPdf model_extended("model_extended","gauss+cheby EXT",
                          RooArgList(gausse,chebye),RooArgList(nsig,nbkg));
```

Create a function object that represents the negative-log-likelihood (**`nll`**) …
… by using the method RooAbsPdf::**createNLL(**RooAbsData&**)**; the returned object is of type **RooAbsReal***

```
RooAbsReal* nll = model_extended.createNLL(BmassExt);
```

In this way we explicitely constructed the likelihood (function of PDF/data combination)
that can be used as any RooFit funcion object.

Note: likelihood can be created by a calculation that can be parallelized (suppose for instance on 4 cores):

```
RooAbsReal* nll = model_extended.createNLL(BmassExt,NumCPU(4));
```

MINUIT session

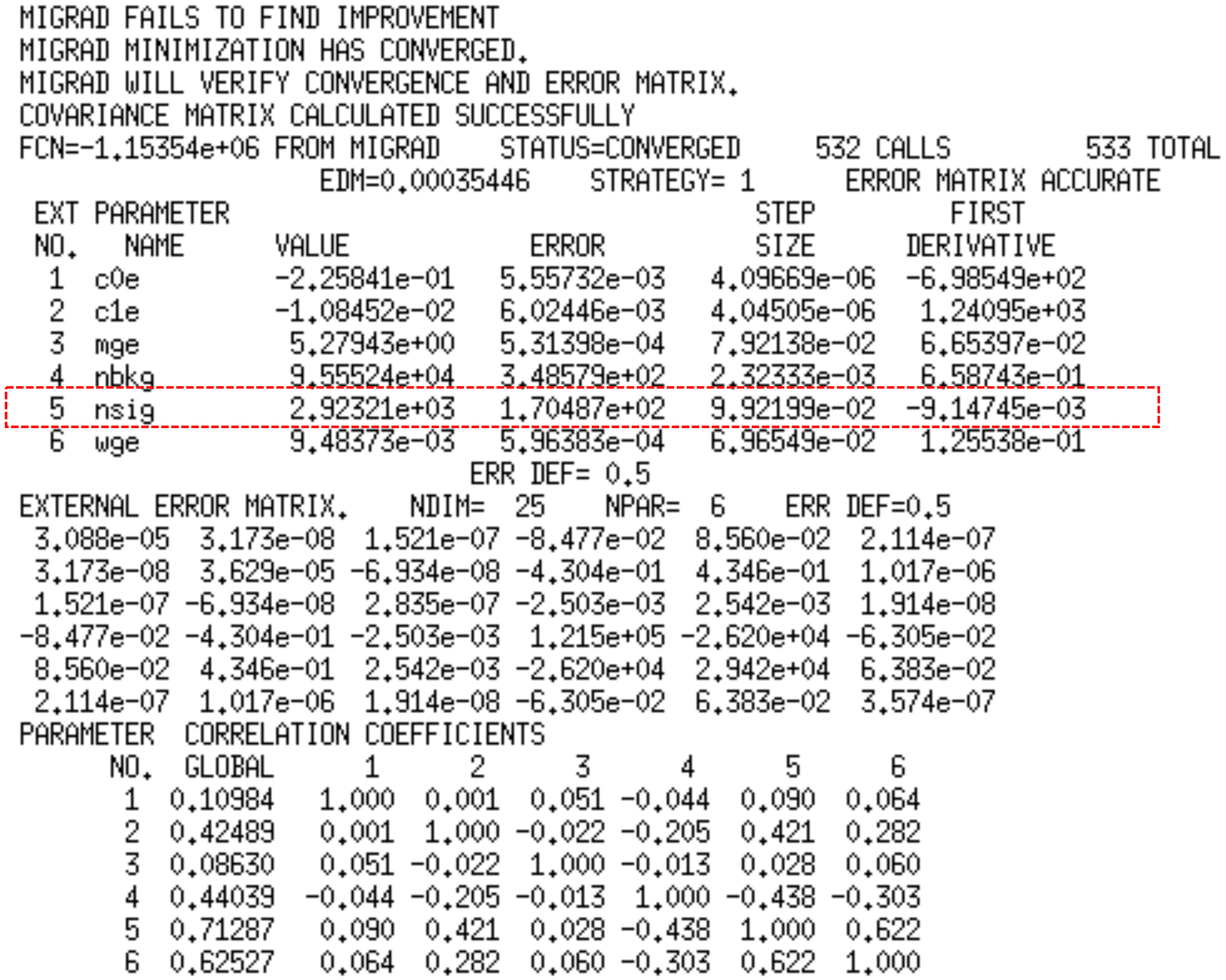
Let us invoke **MINUIT** to perform the binned extended fit.

First we can create a **MINUIT** session:

```
RooMinuit m(*nll);
```

Calling `MIGRAD` we get the central values (best estimates) for the parameters when convergence is reached:

```
m.migrad();
```

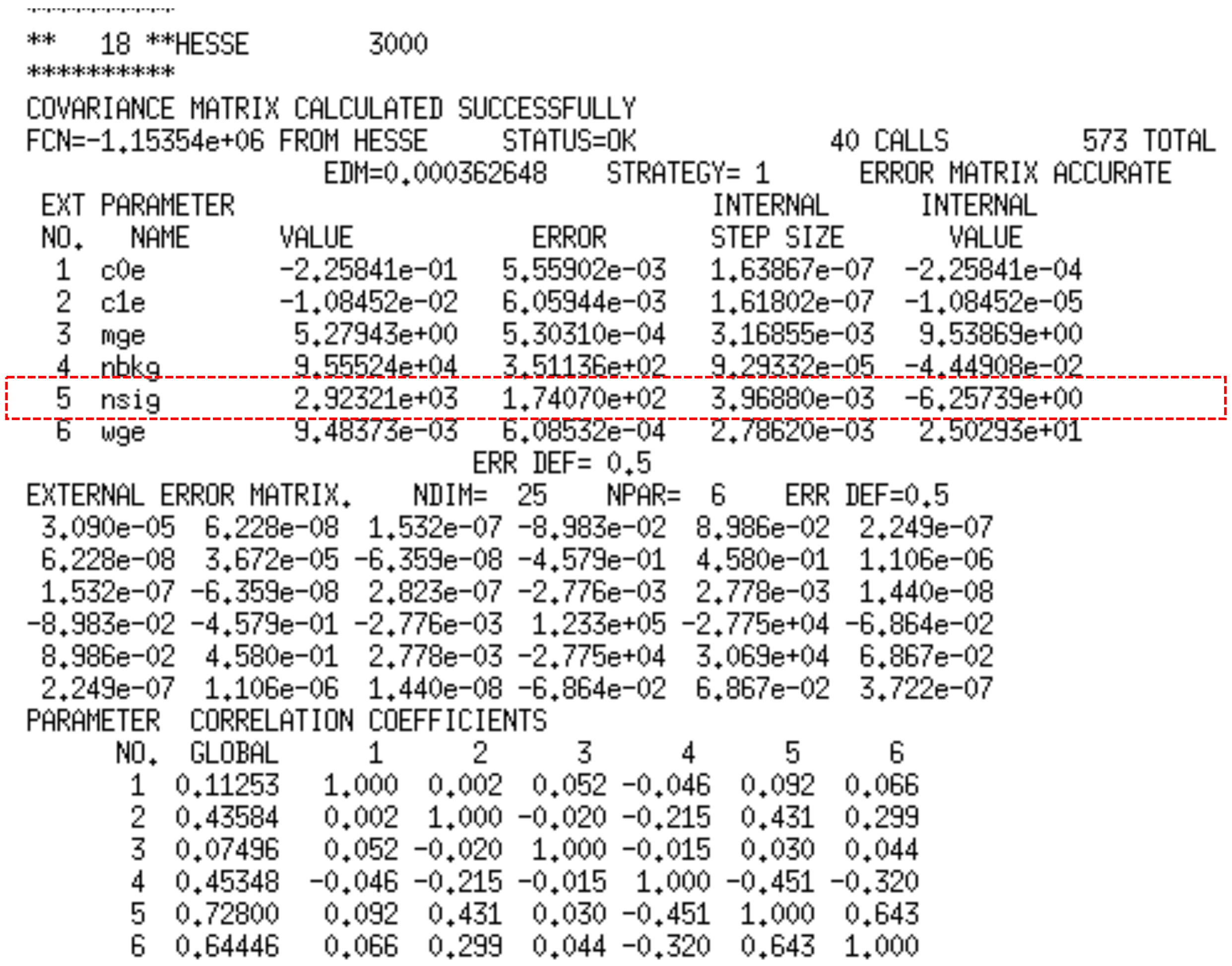


```
MIGRAD FAILS TO FIND IMPROVEMENT
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=-1.15354e+06 FROM MIGRAD    STATUS=CONVERGED     532 CALLS          533 TOTAL
                    EDM=0.00035446    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                   STEP         FIRST
 NO.   NAME      VALUE            ERROR          SIZE      DERIVATIVE
  1  c0e        -2.25841e-01   5.55732e-03   4.09669e-06  -6.98549e+02
  2  c1e        -1.08452e-02   6.02446e-03   4.04505e-06   1.24095e+03
  3  mge         5.27943e+00   5.31398e-04   7.92138e-02   6.65397e-02
  4  nbkg        9.55524e+04   3.48579e+02   2.32333e-03   6.58743e-01
  5  nsig        2.92321e+03   1.70487e+02   9.92199e-02  -9.14745e-03
  6  wge         9.48373e-03   5.96383e-04   6.96549e-02   1.25538e-01
                               ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  6    ERR DEF=0.5
 3.088e-05  3.173e-08  1.521e-07 -8.477e-02  8.560e-02  2.114e-07
 3.173e-08  3.629e-05 -6.934e-08 -4.304e-01  4.346e-01  1.017e-06
 1.521e-07 -6.934e-08  2.835e-07 -2.503e-03  2.542e-03  1.914e-08
-8.477e-02 -4.304e-01 -2.503e-03  1.215e+05 -2.620e+04 -6.305e-02
 8.560e-02  4.346e-01  2.542e-03 -2.620e+04  2.942e+04  6.383e-02
 2.114e-07  1.017e-06  1.914e-08 -6.305e-02  6.383e-02  3.574e-07
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2      3      4      5      6
       1  0.10984   1.000  0.001  0.051 -0.044  0.090  0.064
       2  0.42489   0.001  1.000 -0.022 -0.205  0.421  0.282
       3  0.08630   0.051 -0.022  1.000 -0.013  0.028  0.060
       4  0.44039  -0.044 -0.205 -0.013  1.000 -0.438 -0.303
       5  0.71287   0.090  0.421  0.028 -0.438  1.000  0.622
       6  0.62527   0.064  0.282  0.060 -0.303  0.622  1.000
```

Parabolic uncertainties

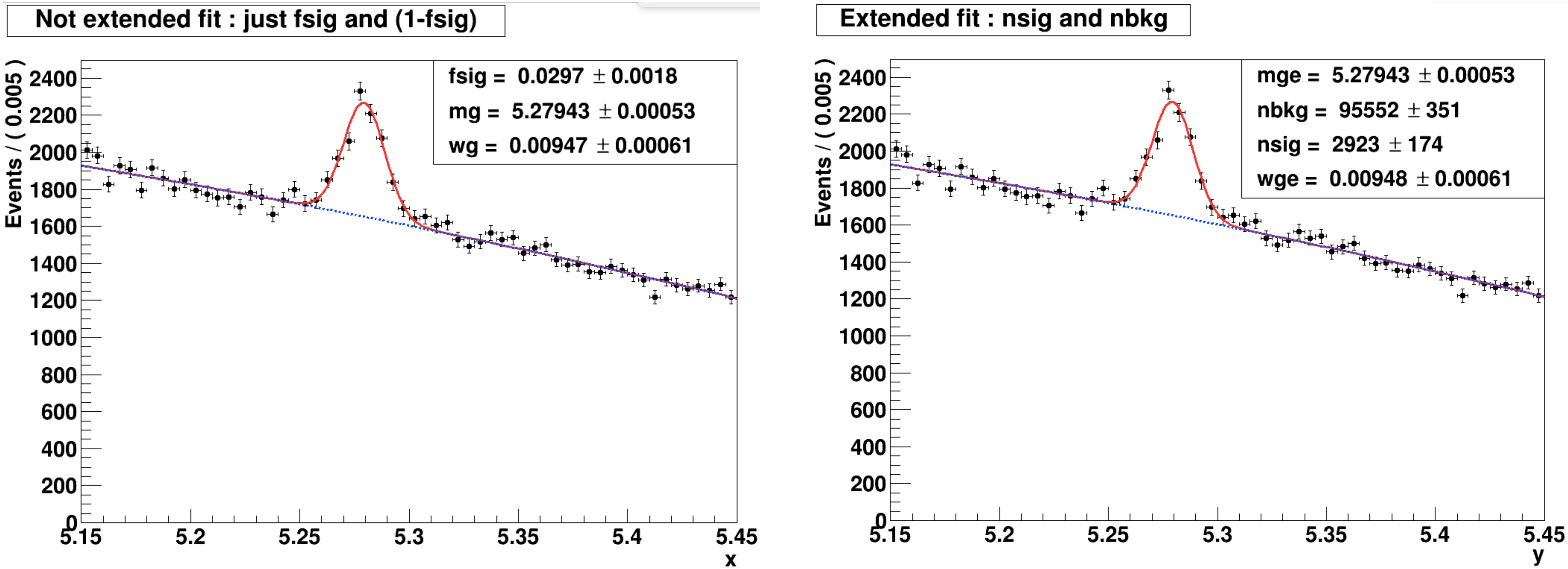
To recalculate the errors and the covariance matrix in an accurate way (still in parabolic assumption) we use `HESSE`, while central values (by `Migrad`) are conserved.

```
m.hesse();
```



```
**********
**   18 **HESSE        3000
**********
 COVARIANCE MATRIX CALCULATED SUCCESSFULLY
 FCN=-1.15354e+06 FROM HESSE     STATUS=OK            40 CALLS         573 TOTAL
                     EDM=0.000362648    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                                INTERNAL      INTERNAL
  NO.   NAME      VALUE            ERROR       STEP SIZE       VALUE
   1  c0e         -2.25841e-01   5.55902e-03   1.63867e-07  -2.25841e-04
   2  c1e         -1.08452e-02   6.05944e-03   1.61802e-07  -1.08452e-05
   3  mge          5.27943e+00   5.30310e-04   3.16855e-03   9.53869e+00
   4  nbkg         9.55524e+04   3.51136e+02   9.29332e-05  -4.44908e-02
   5  nsig         2.92321e+03   1.74070e+02   3.96880e-03  -6.25739e+00
   6  wge          9.48373e-03   6.08532e-04   2.78620e-03   2.50293e+01
                               ERR DEF= 0.5
 EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  6    ERR DEF=0.5
  3.090e-05  6.228e-08  1.532e-07 -8.983e-02  8.986e-02  2.249e-07
  6.228e-08  3.672e-05 -6.359e-08 -4.579e-01  4.580e-01  1.106e-06
  1.532e-07 -6.359e-08  2.823e-07 -2.776e-03  2.778e-03  1.440e-08
 -8.983e-02 -4.579e-01 -2.776e-03  1.233e+05 -2.775e+04 -6.864e-02
  8.986e-02  4.580e-01  2.778e-03 -2.775e+04  3.069e+04  6.867e-02
  2.249e-07  1.106e-06  1.440e-08 -6.864e-02  6.867e-02  3.722e-07
 PARAMETER  CORRELATION COEFFICIENTS
       NO.  GLOBAL      1      2      3      4      5      6
        1  0.11253   1.000  0.002  0.052 -0.046  0.092  0.066
        2  0.43584   0.002  1.000 -0.020 -0.215  0.431  0.299
        3  0.07496   0.052 -0.020  1.000 -0.015  0.030  0.044
        4  0.45348  -0.046 -0.215 -0.015  1.000 -0.451 -0.320
        5  0.72800   0.092  0.431  0.030 -0.451  1.000  0.643
        6  0.64446   0.066  0.299  0.044 -0.320  0.643  1.000
```

Extended vs not-extended fits: a comparison - I

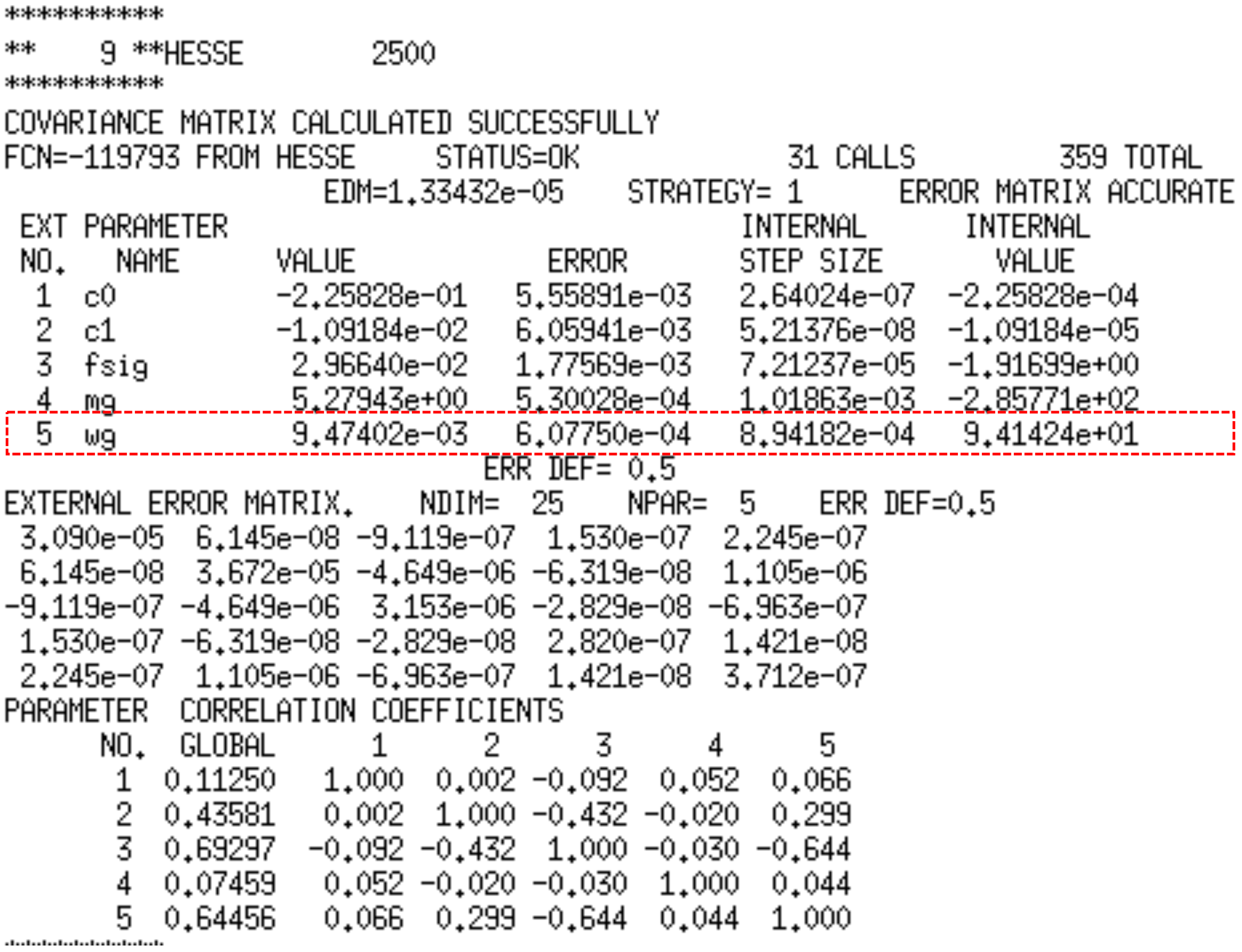


Difference can be hardly appreciated: mass and width are about identical ! (see next slide)

Extended fit has the advantage to provide as output also the number of B⁺ candidates (`nsig`)

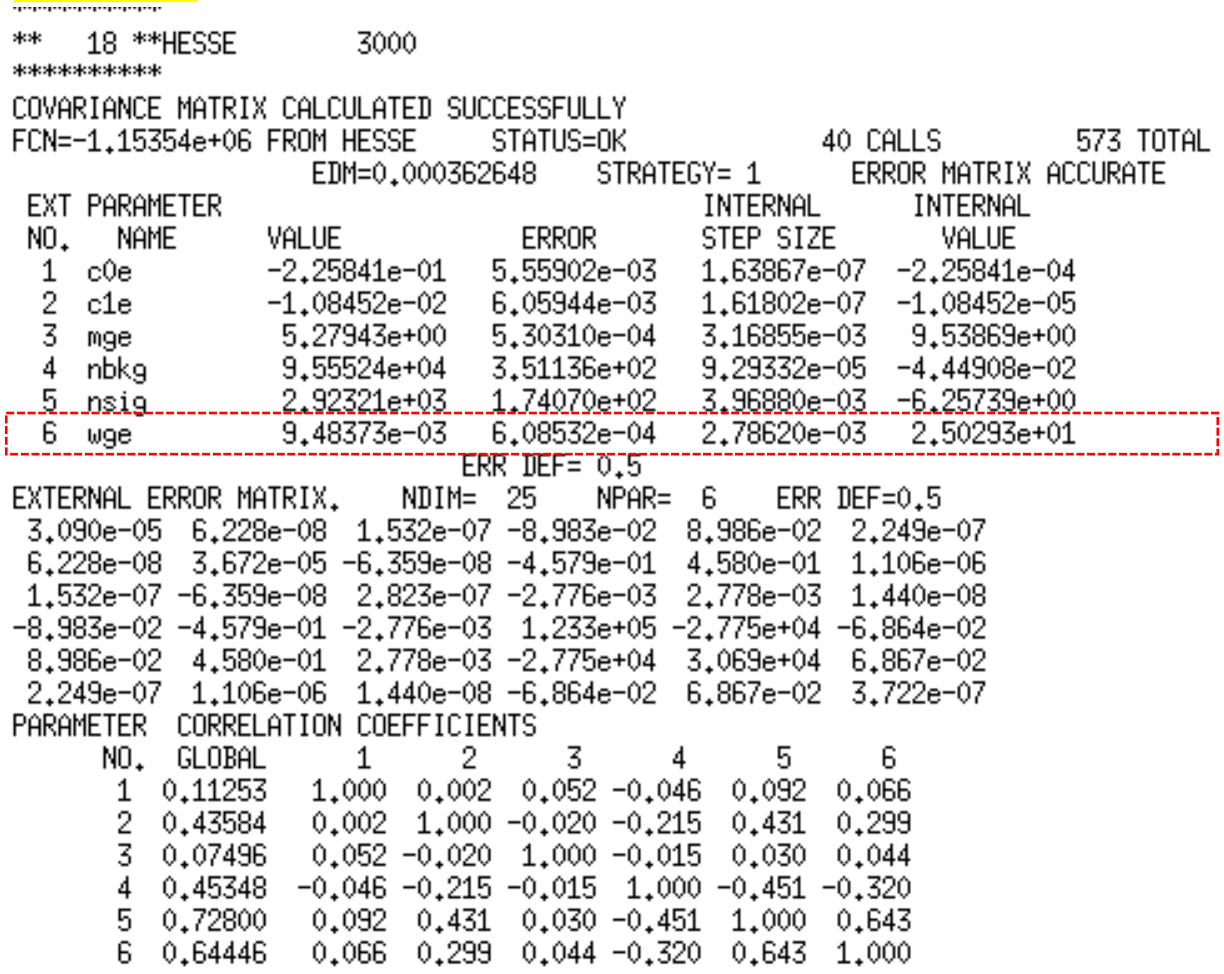
Extended vs not-extended fits: a comparison - II

NOT-extended



```
**********
**    9 **HESSE        2500
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=-119793 FROM HESSE     STATUS=OK             31 CALLS          359 TOTAL
                     EDM=1.33432e-05    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                INTERNAL      INTERNAL
 NO.   NAME      VALUE            ERROR       STEP SIZE       VALUE
  1  c0         -2.25828e-01   5.55891e-03   2.64024e-07  -2.25828e-04
  2  c1         -1.09184e-02   6.05941e-03   5.21376e-08  -1.09184e-05
  3  fsig        2.96640e-02   1.77569e-03   7.21237e-05  -1.91699e+00
  4  mg          5.27943e+00   5.30028e-04   1.01863e-03  -2.85771e+02
  5  wg          9.47402e-03   6.07750e-04   8.94182e-04   9.41424e+01
                               ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  5    ERR DEF=0.5
 3.090e-05  6.145e-08 -9.119e-07  1.530e-07  2.245e-07
 6.145e-08  3.672e-05 -4.649e-06 -6.319e-08  1.105e-06
-9.119e-07 -4.649e-06  3.153e-06 -2.829e-08 -6.963e-07
 1.530e-07 -6.319e-08 -2.829e-08  2.820e-07  1.421e-08
 2.245e-07  1.105e-06 -6.963e-07  1.421e-08  3.712e-07
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2      3      4      5
       1  0.11250   1.000  0.002 -0.092  0.052  0.066
       2  0.43581   0.002  1.000 -0.432 -0.020  0.299
       3  0.69297  -0.092 -0.432  1.000 -0.030 -0.644
       4  0.07459   0.052 -0.020 -0.030  1.000  0.044
       5  0.64456   0.066  0.299 -0.644  0.044  1.000
**********
```

Extended



```
**********
**   18 **HESSE        3000
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=-1.15354e+06 FROM HESSE     STATUS=OK             40 CALLS          573 TOTAL
                     EDM=0.000362648    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                INTERNAL      INTERNAL
 NO.   NAME      VALUE            ERROR       STEP SIZE       VALUE
  1  c0e        -2.25841e-01   5.55902e-03   1.63867e-07  -2.25841e-04
  2  c1e        -1.08452e-02   6.05944e-03   1.61802e-07  -1.08452e-05
  3  mge         5.27943e+00   5.30310e-04   3.16855e-03   9.53869e+00
  4  nbkg        9.55524e+04   3.51136e+02   9.29332e-05  -4.44908e-02
  5  nsig        2.92321e+03   1.74070e+02   3.96880e-03  -6.25739e+00
  6  wge         9.48373e-03   6.08532e-04   2.78620e-03   2.50293e+01
                               ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  6    ERR DEF=0.5
 3.090e-05  6.228e-08  1.532e-07 -8.983e-02  8.986e-02  2.249e-07
 6.228e-08  3.672e-05 -6.359e-08 -4.579e-01  4.580e-01  1.106e-06
 1.532e-07 -6.359e-08  2.823e-07 -2.776e-03  2.778e-03  1.440e-08
-8.983e-02 -4.579e-01 -2.776e-03  1.233e+05 -2.775e+04 -6.864e-02
 8.986e-02  4.580e-01  2.778e-03 -2.775e+04  3.069e+04  6.867e-02
 2.249e-07  1.106e-06  1.440e-08 -6.864e-02  6.867e-02  3.722e-07
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2      3      4      5      6
       1  0.11253   1.000  0.002  0.052 -0.046  0.092  0.066
       2  0.43584   0.002  1.000 -0.020 -0.215  0.431  0.299
       3  0.07496   0.052 -0.020  1.000 -0.015  0.030  0.044
       4  0.45348  -0.046 -0.215 -0.015  1.000 -0.451 -0.320
       5  0.72800   0.092  0.431  0.030 -0.451  1.000  0.643
       6  0.64446   0.066  0.299  0.044 -0.320  0.643  1.000
```

$$m = (5279{,}43 \pm 0.53) MeV$$

$$\sigma = (9{,}4740 \pm 0{,}6078) MeV$$

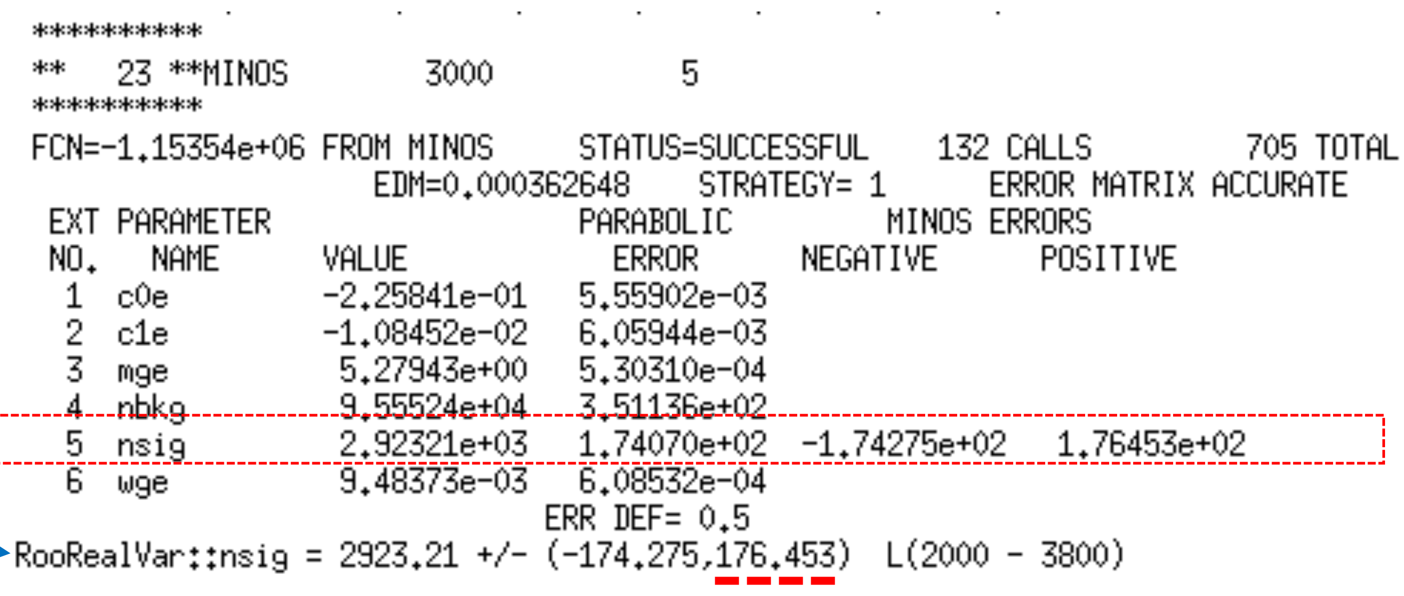
$$m = (5279{,}43 \pm 0.53) MeV$$

$$\sigma = (9{,}4837 \pm 0{,}6085) MeV$$

Asymmetric uncertainties

To get asymmetric error (central values and parabolic are the same) for a specific parameter, like `nsig` :

`m.minos(nsig);`

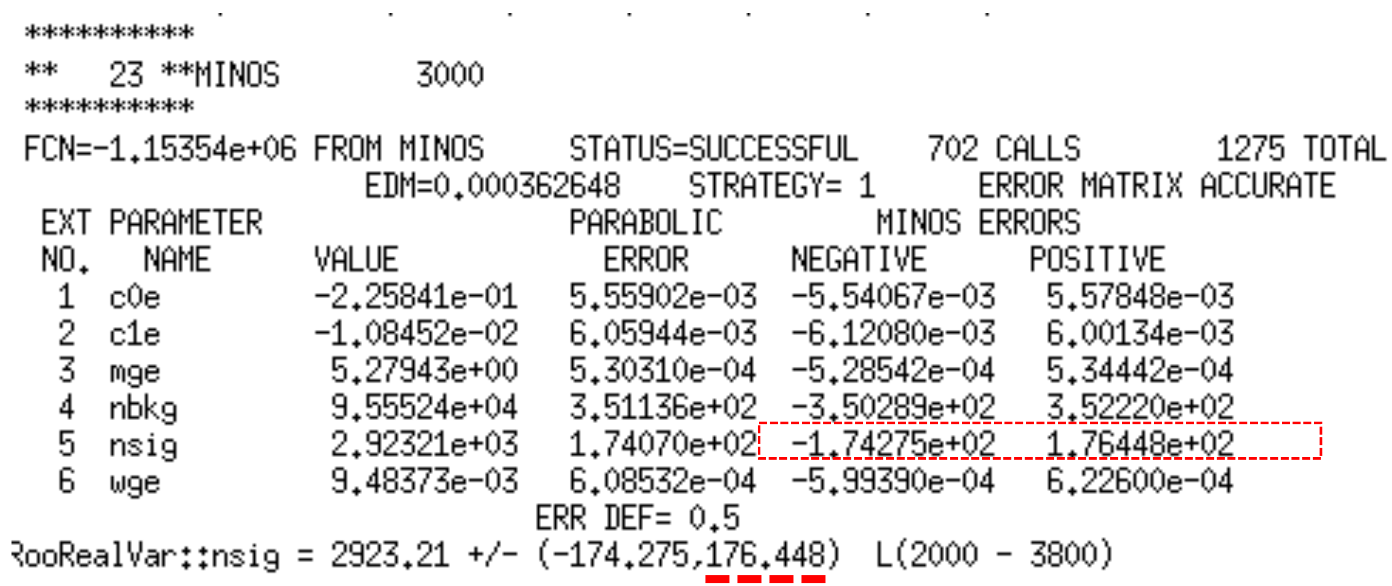


```
**********
**   23 **MINOS        3000          5
**********
FCN=-1.15354e+06 FROM MINOS     STATUS=SUCCESSFUL    132 CALLS         705 TOTAL
                    EDM=0.000362648    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                       PARABOLIC         MINOS ERRORS
  NO.   NAME      VALUE              ERROR       NEGATIVE      POSITIVE
   1  c0e        -2.25841e-01   5.55902e-03
   2  c1e        -1.08452e-02   6.05944e-03
   3  mge         5.27943e+00   5.30310e-04
   4  nbkg        9.55524e+04   3.51136e+02
   5  nsig        2.92321e+03   1.74070e+02  -1.74275e+02   1.76453e+02
   6  wge         9.48373e-03   6.08532e-04
                               ERR DEF= 0.5
RooRealVar::nsig = 2923.21 +/- (-174.275,176.453)  L(2000 - 3800)
```

To additionally print the result just do: `nsig.Print();`

To get asymmetric error for all the parameters :

`m.minos();`



```
**********
**   23 **MINOS        3000
**********
FCN=-1.15354e+06 FROM MINOS     STATUS=SUCCESSFUL    702 CALLS        1275 TOTAL
                    EDM=0.000362648    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                       PARABOLIC         MINOS ERRORS
  NO.   NAME      VALUE              ERROR       NEGATIVE      POSITIVE
   1  c0e        -2.25841e-01   5.55902e-03  -5.54067e-03   5.57848e-03
   2  c1e        -1.08452e-02   6.05944e-03  -6.12080e-03   6.00134e-03
   3  mge         5.27943e+00   5.30310e-04  -5.28542e-04   5.34442e-04
   4  nbkg        9.55524e+04   3.51136e+02  -3.50289e+02   3.52220e+02
   5  nsig        2.92321e+03   1.74070e+02  -1.74275e+02   1.76448e+02
   6  wge         9.48373e-03   6.08532e-04  -5.99390e-04   6.22600e-04
                               ERR DEF= 0.5
RooRealVar::nsig = 2923.21 +/- (-174.275,176.448)  L(2000 - 3800)
```

Note: asymmetric errors can slightly change if you execute `MINOS` for 1 or all parameters

(in this case only … upper uncertaity changes)

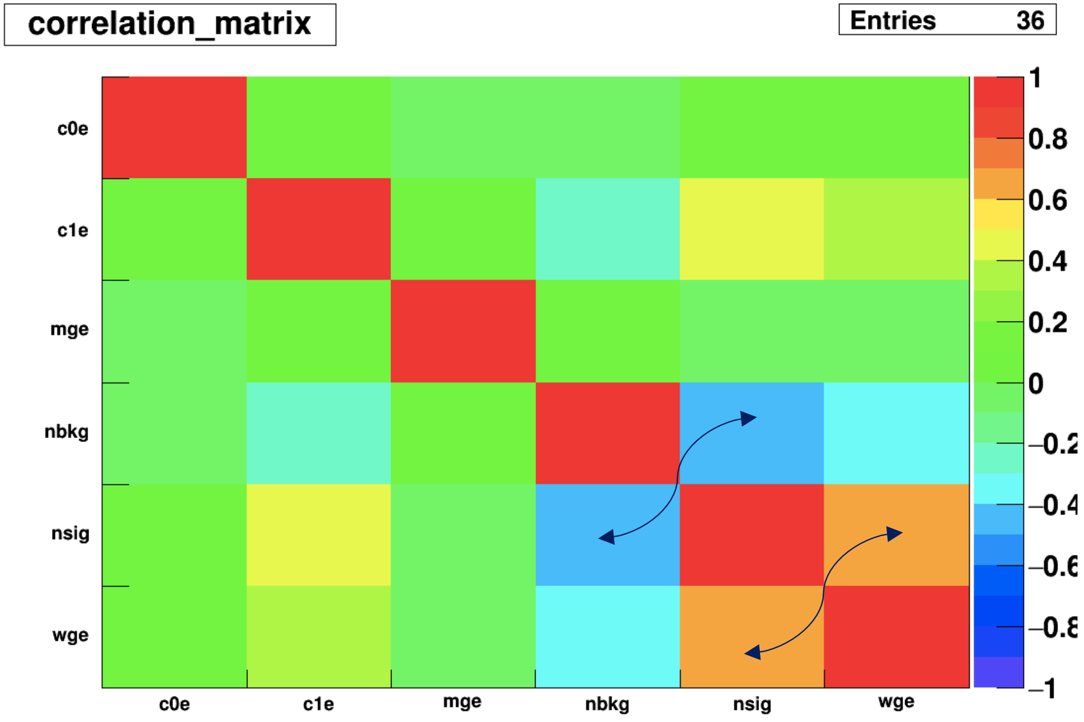
Correlation Matrix

It is possible to save the status of the fit, including the information about the convariance matrix:

```
RooFitResult* fitres = m.save();
```

It is possible to visualize the correlation matrix:

```
gStyle->SetPalette(1); //- for better color choice
fitres->correlationHist->Draw(“colz”);
```



Note:

- anticorrelation between `nsig` and `nbkg` as expected
- correlation between `nsig` and width of Gaussian `wge`

Note: in general, if correlations are very strong (i.e. > 0.9) the model may become unstable and it may be worthwhile to fix one of the parameters in the fit.

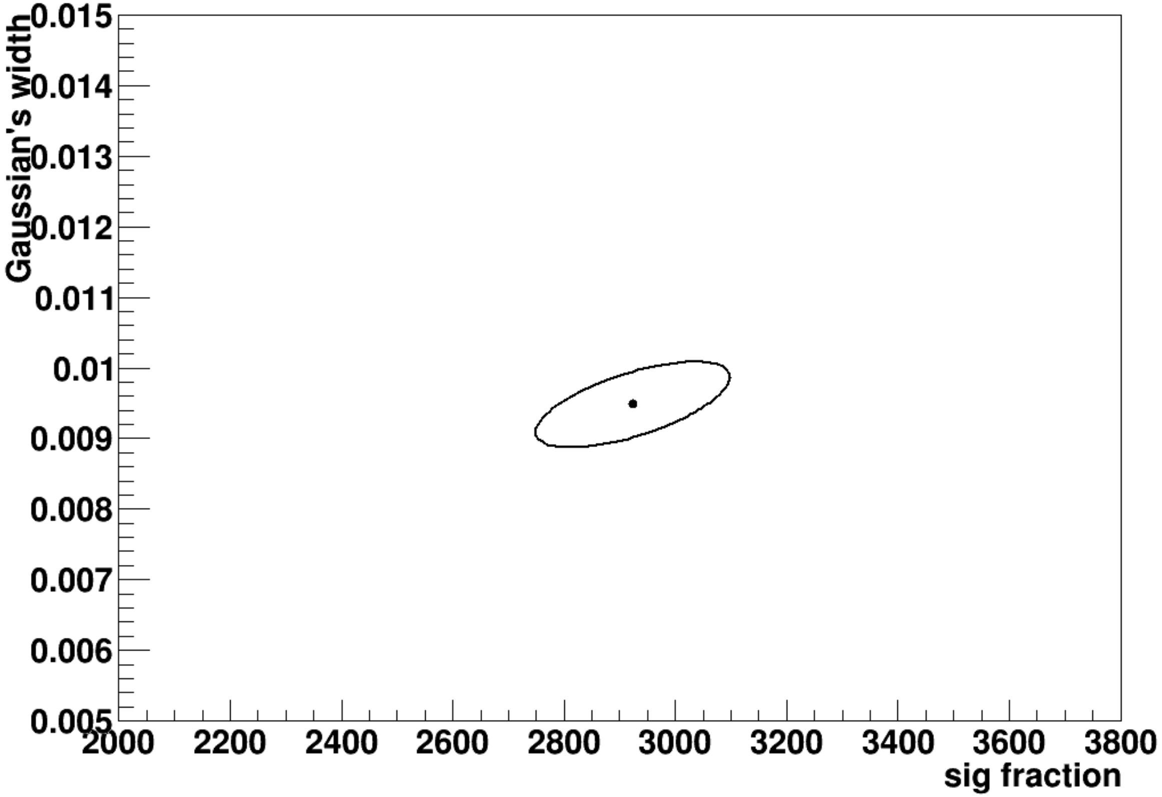
If the strong correlation is between two nuisance parameters, this is not a problem. Instead, when a parameter-of-interest is correlated with a nuisance one, it must be avoided to fix the nuisance parameter because the risk is to strongly under-estimate the uncertainty on the physical parameter!

Visualization of correlated errors - I

It is also possible to visualize errors & correlation matrix elements:

```
RooPlot* paramFrame = new RooPlot(nsig,wge);
fitres->plotOn(paramFrame,nsig,wge);
paramFrame->Draw();
```

where this example shows a certain level of correlation.



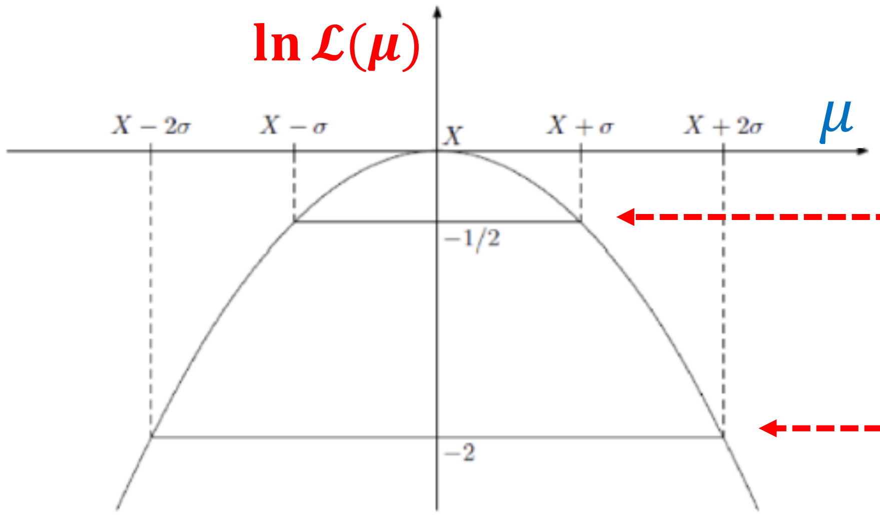
```
m.contour(nsig,wge);
```

It starts the `MNCONTOUR` calculation of 50 point on two contours (for `ERRDEF`=0.5 and 2.0). Each point identified by a pair of values of parameters 5(`nsig`) and 6(`wge`) is a minimum with respect to the remaining (four, in this case) variable parameters. Values are printed on the screen @ execution time. The contour with `ERRDEF`=0.5 is the one obtained before/above.

Visualization of correlated errors - II

But why `ERRDEF`=0.5 and 2.0 are considered?

Well, do not forget that a PDF can be converted into a Likelihood function $\mathcal{L}$ by "exchanging" the vector of observations $\vec{x}$ with the vector of parameters $\vec{\theta}$!
For only one parameter, say μ , the likelihood is a function of it, namely $\mathcal{L}(\mu)$, and $\ln \mathcal{L}(\mu)$ **is a parabola!**



`ERRDEF`=0.5 → 1σ

`ERRDEF`=2.0 → 2σ

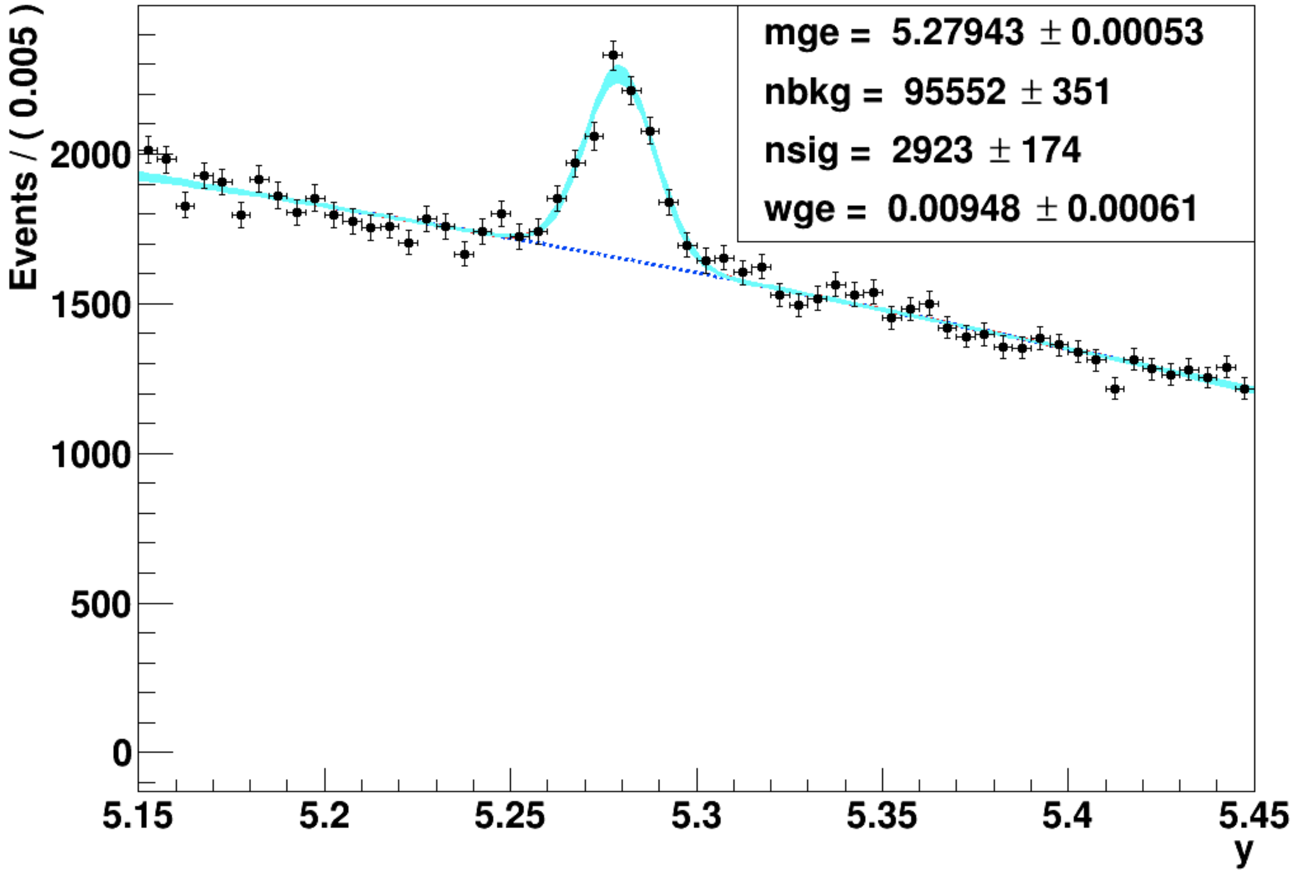
Note : if you put the "-" in front of it, thus getting the neg-log-likelihood, $-\ln \mathcal{L}(\mu)$, the parabola changes sign and "points" upwards instead of downwards.

Extension: Now suppose we've 2 parameters of interest; in this case you can imagine a paraboloid instead of a parabola with different aperture when projecting in 1-dim. The "multivariate" uncertainty is then represented by an elliptic contour.

Visualization of the fit uncertainty

It is possible to propagate the errors (stored in the covariance matrix of a fit result) to a PDF projection:

```
model_extended.plotOn(yframe, VisualizeError(*fitres));
yframe->Draw();
```



To get the points' errors over the cyan shadowed region describing the uncertainty we need to add the following two lines (to get the "trick" done):

```
BmassExt.plotOn(yframe);
yframe->Draw("Esame");
```

Visualization of the fit log-likelihood function and of the Profile Likelihood ratio

We can obtain the best estimate for `nsig` & the **`MINOS`** uncertainty **corresponding** to the interval provided by the PL ratio :

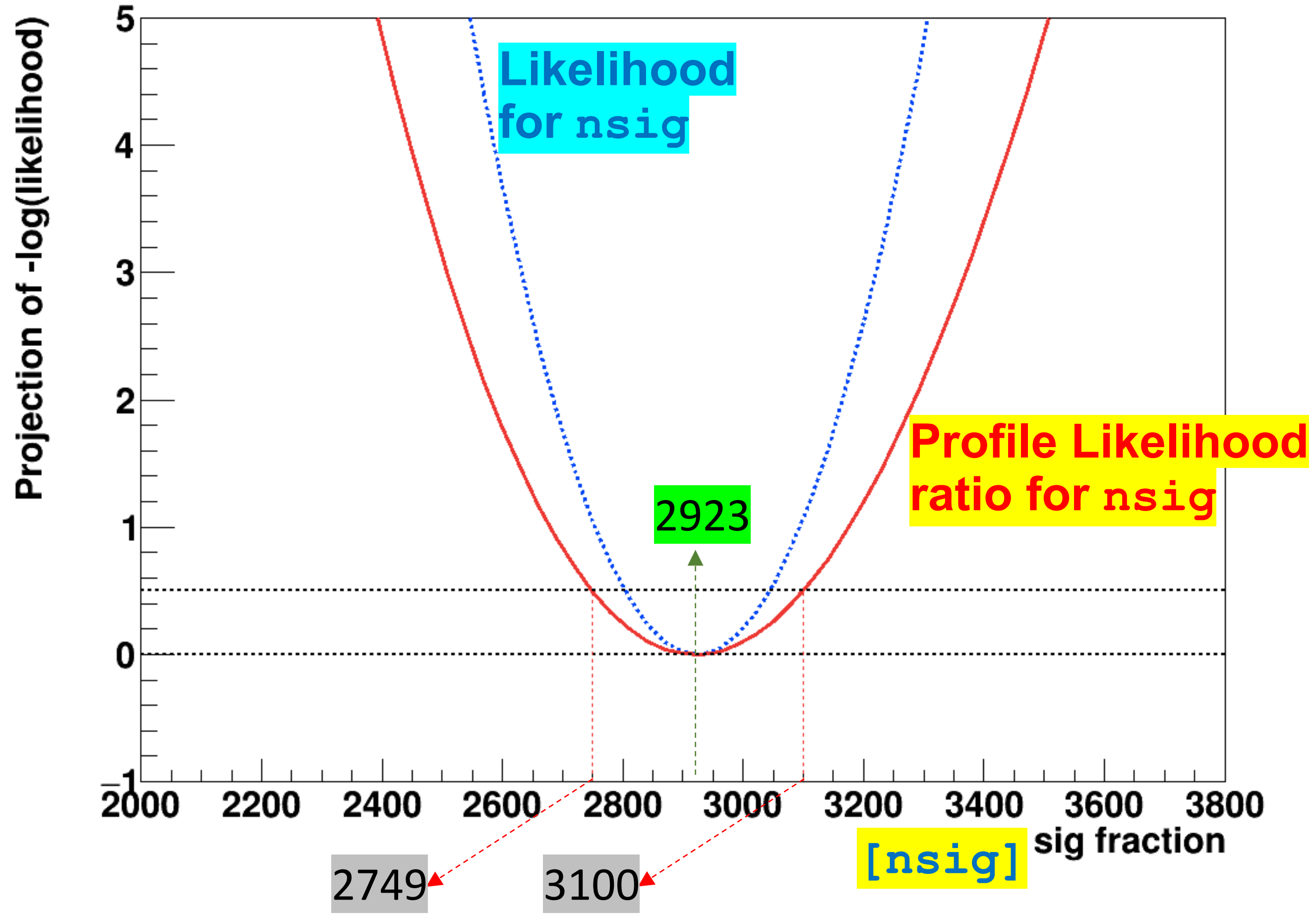
```
// plot the likelihood as a function of the parameter of interest (here nsig):
RooPlot* nsig_frame = nsig.frame(RooFit::Bins(60),RooFit::Range(2000,3800));
nll->plotOn(nsig_frame,RooFit::ShiftToZero(),RooFit::LineStyle(kDashed),LineColor(kBlue));
//
// make the Profile Likelihood ratio (that can be represented as a regular RooFit function)
RooAbsReal* pll_nsig = nll->createProfile(nsig);
pll_nsig->plotOn(nsig_frame,RooFit::ShiftToZero(),LineColor(kRed));
nsig_frame->SetMinimum(-1);
nsig_frame->SetMaximum(5);
nsig_frame->Draw();
```

From `MIGRAD`: 2923.2

From `MINOS`: $(2923.2) - 174.3 + 176.4$

(slightly asymmetric)

Overall interval: $\cong [2749, 3100]$



Connection between `MINOS` uncertainties & Profile Likelihood

In the next slides this connection will be investigated & explained.

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: $\boldsymbol{\mathcal{L}}(\vec{\boldsymbol{x}};\boldsymbol{\mu},\vec{\boldsymbol{\theta}})$. To easy the notation we drop the $\vec{\boldsymbol{x}}$ and write simply $\boldsymbol{\mathcal{L}}(\boldsymbol{\mu},\vec{\boldsymbol{\theta}})$.

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 $\boldsymbol{\mathcal{L}}_{max}(\bar{\boldsymbol{\mu}},\hat{\hat{\vec{\boldsymbol{\theta}}}}(\bar{\boldsymbol{\mu}}))$;

- recalculating (CPU expensive) for each value of $\boldsymbol{\mu}$ (scan of $\boldsymbol{\mu}$ values) we get a truly function of $\boldsymbol{\mu}$: $\boldsymbol{\mathcal{L}}_{max}(\boldsymbol{\mu},\hat{\hat{\vec{\boldsymbol{\theta}}}}(\boldsymbol{\mu}))$
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 $t_{\mu} = -2 \, ln \, \lambda(\mu)$. 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 $t_{\mu} = -2 \, ln \, \lambda(\mu)$, **when** we have to deal with nuissance parameters:

Likelihood for a given μ

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

Maximum Likelihood

Maximum Likelihood

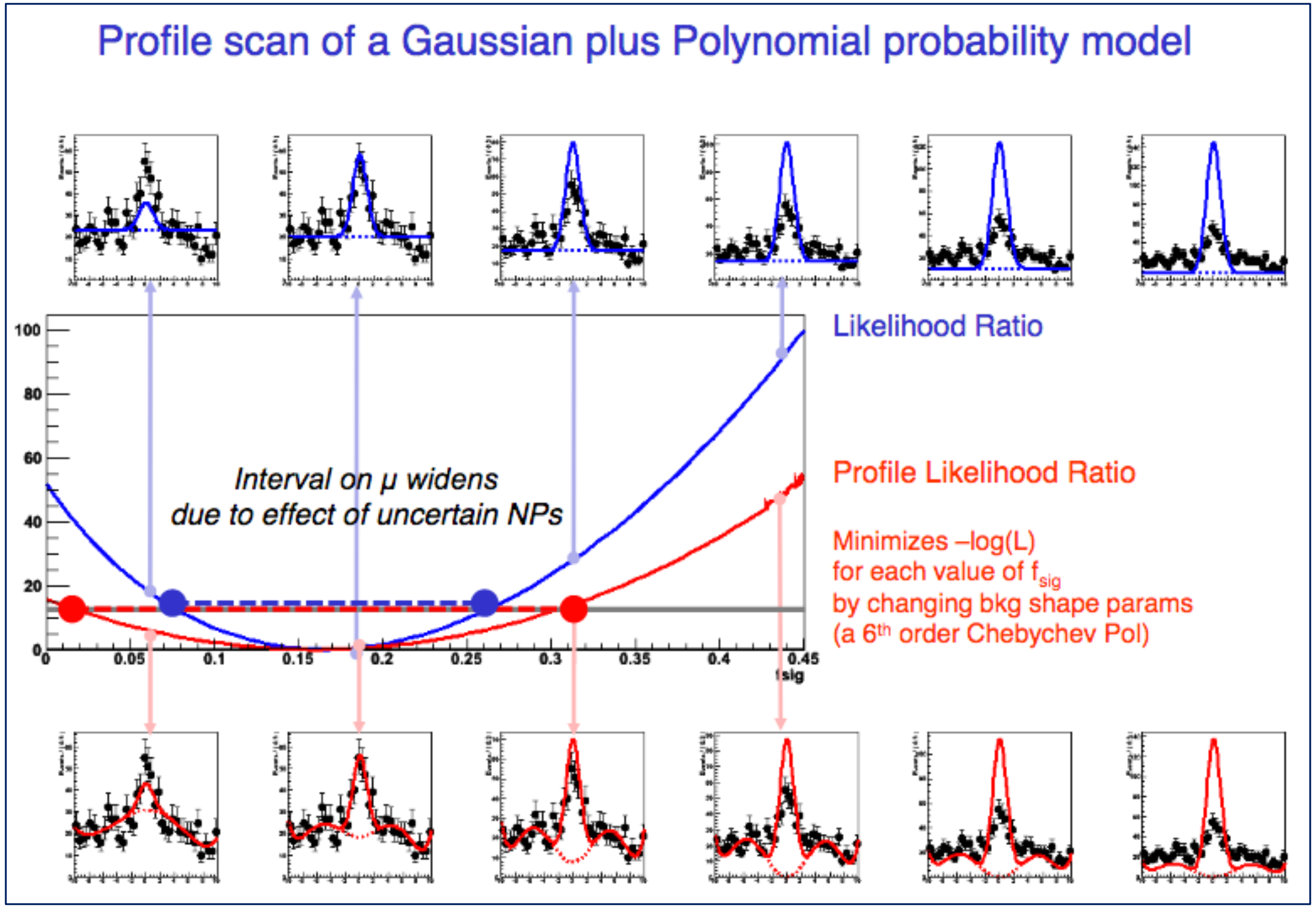
Maximum Likelihood for a given μ

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 ar this figure illustrating the scan of $\boldsymbol{\mu}$ values in order to obtain $\mathcal{L}(\mu,\hat{\hat{\vec{\theta}}}(\mu))$:



Profile scan of a Gaussian plus Polynomial probability model

Likelihood Ratio

Profile Likelihood Ratio

Minimizes –log(L)
for each value of f_{sig}
by changing bkg shape params
(a 6th order Chebychev Pol)

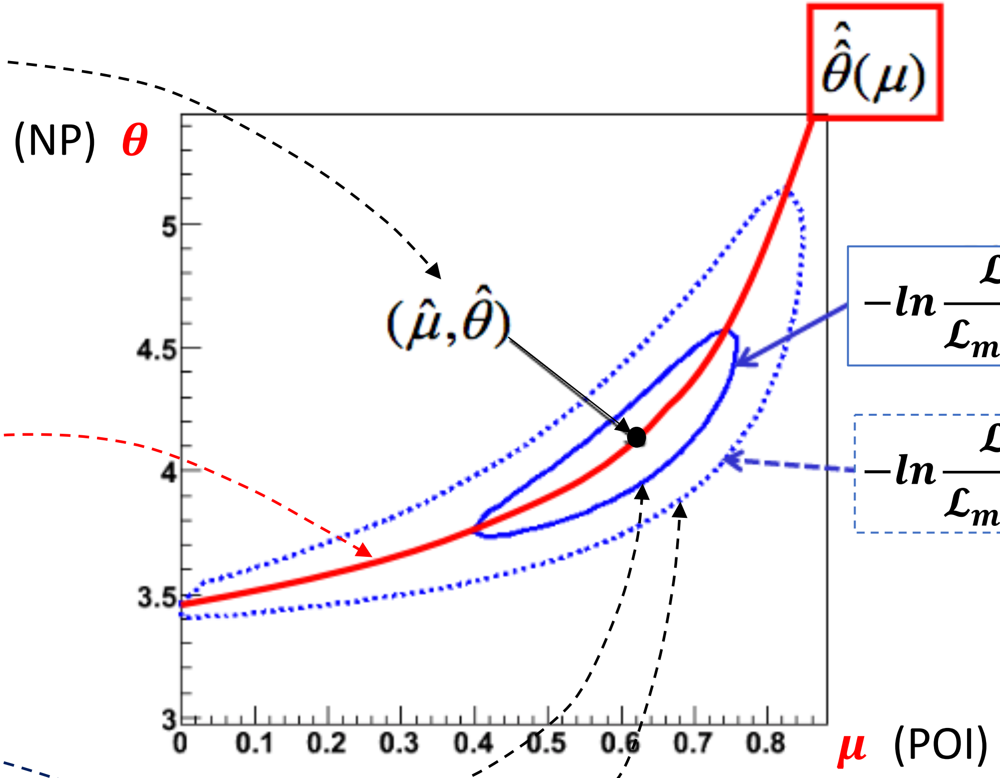
Interval on μ widens
due to effect of uncertain NPs

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: $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})$

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



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

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

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

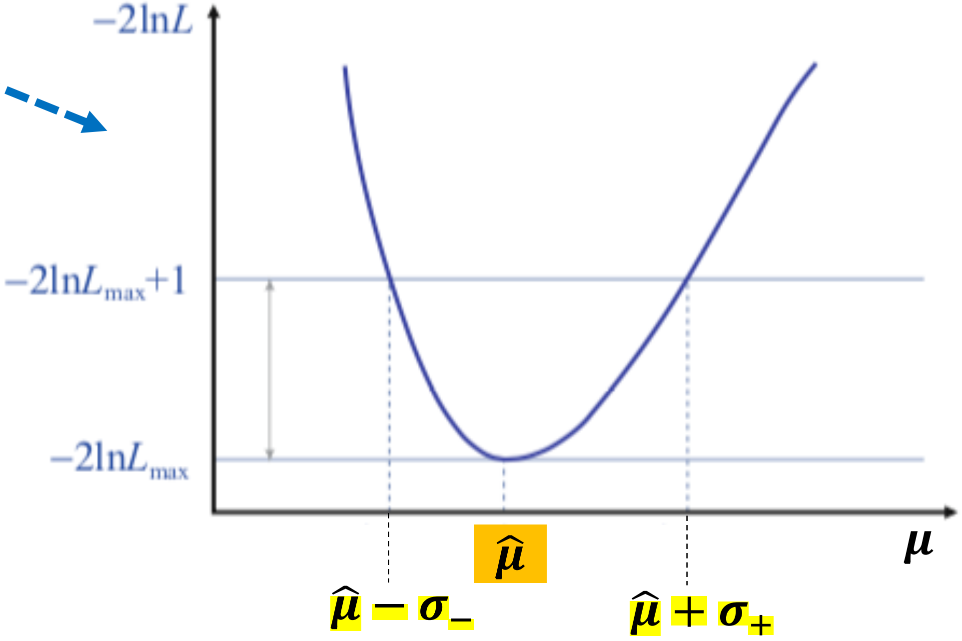
…corresponding to a subset of subsequently given/fixed values of $\boldsymbol{\mu}$ which includes also the special value $\overline{\boldsymbol{\mu}} \equiv \widehat{\boldsymbol{\mu}}$.

Finally we can design the contour curves at 1σ and 2σ with respect to $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})$ 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).

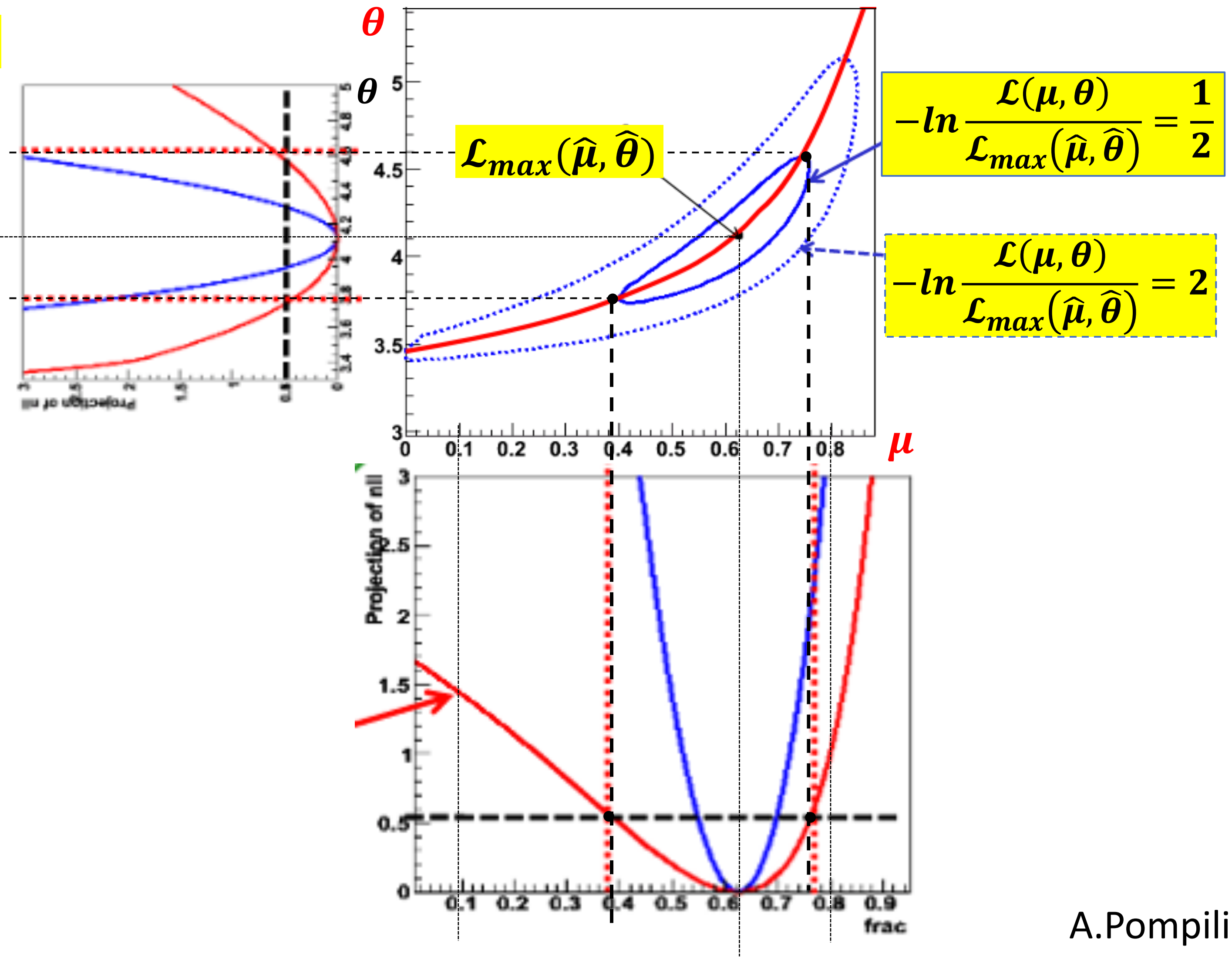
When there is also one NP one gets 2D contours (see next slide) and

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

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**.



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

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

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

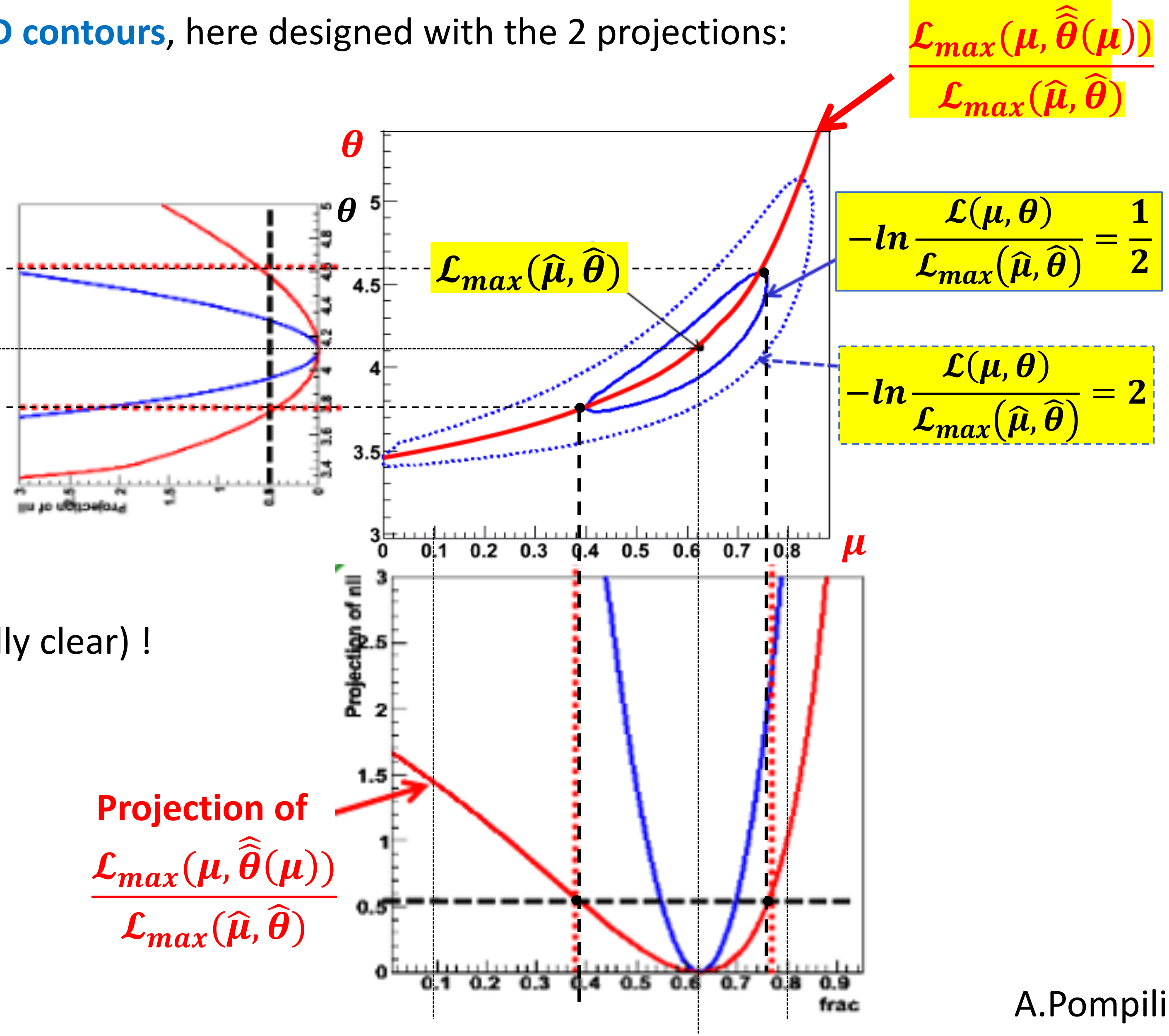
The overall uncertanty is in general asymmetric!

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

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

A.Pompili-18

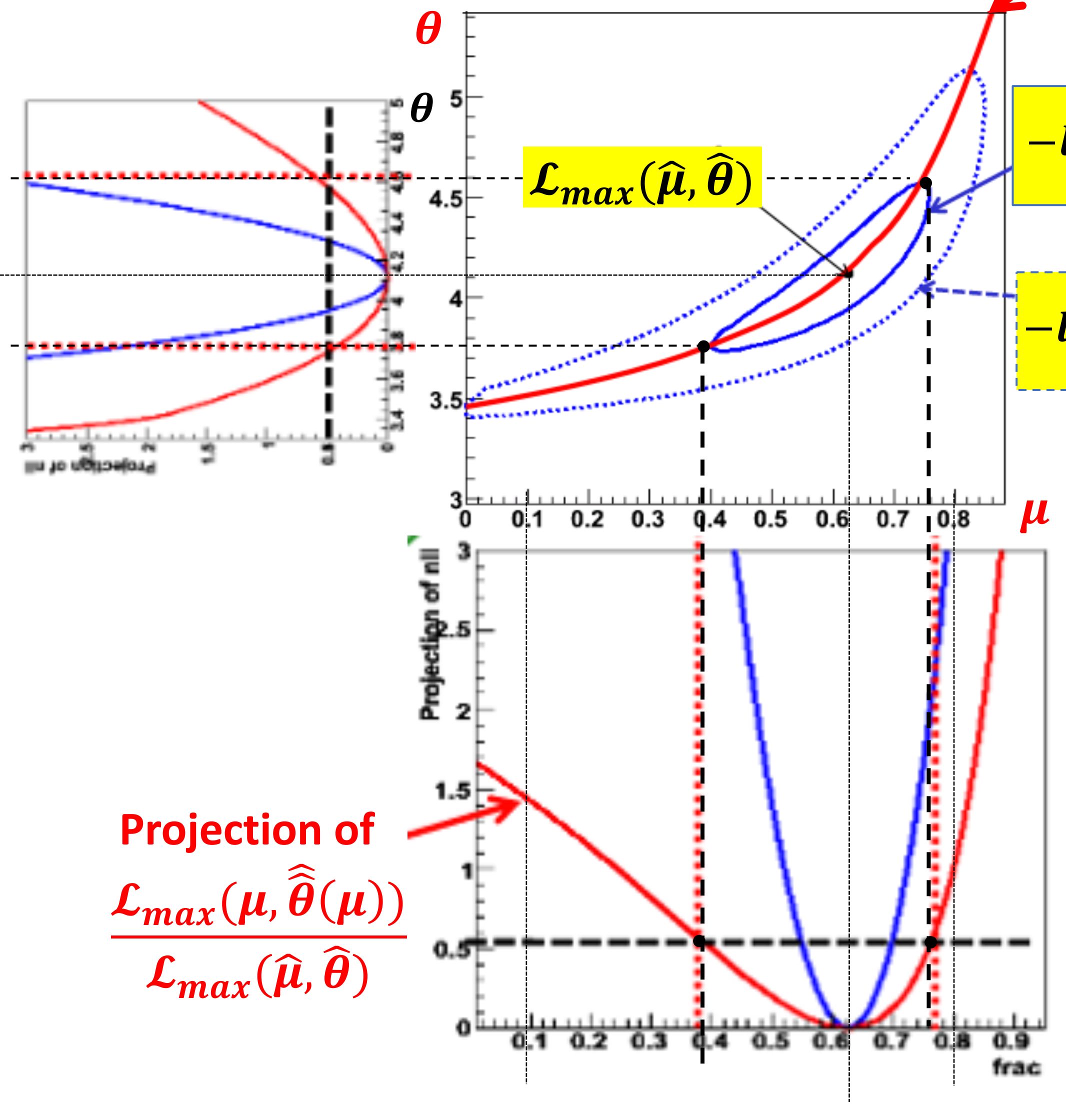
MINOS uncertainties by likelihood scan

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.**

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



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

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

Projection of

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

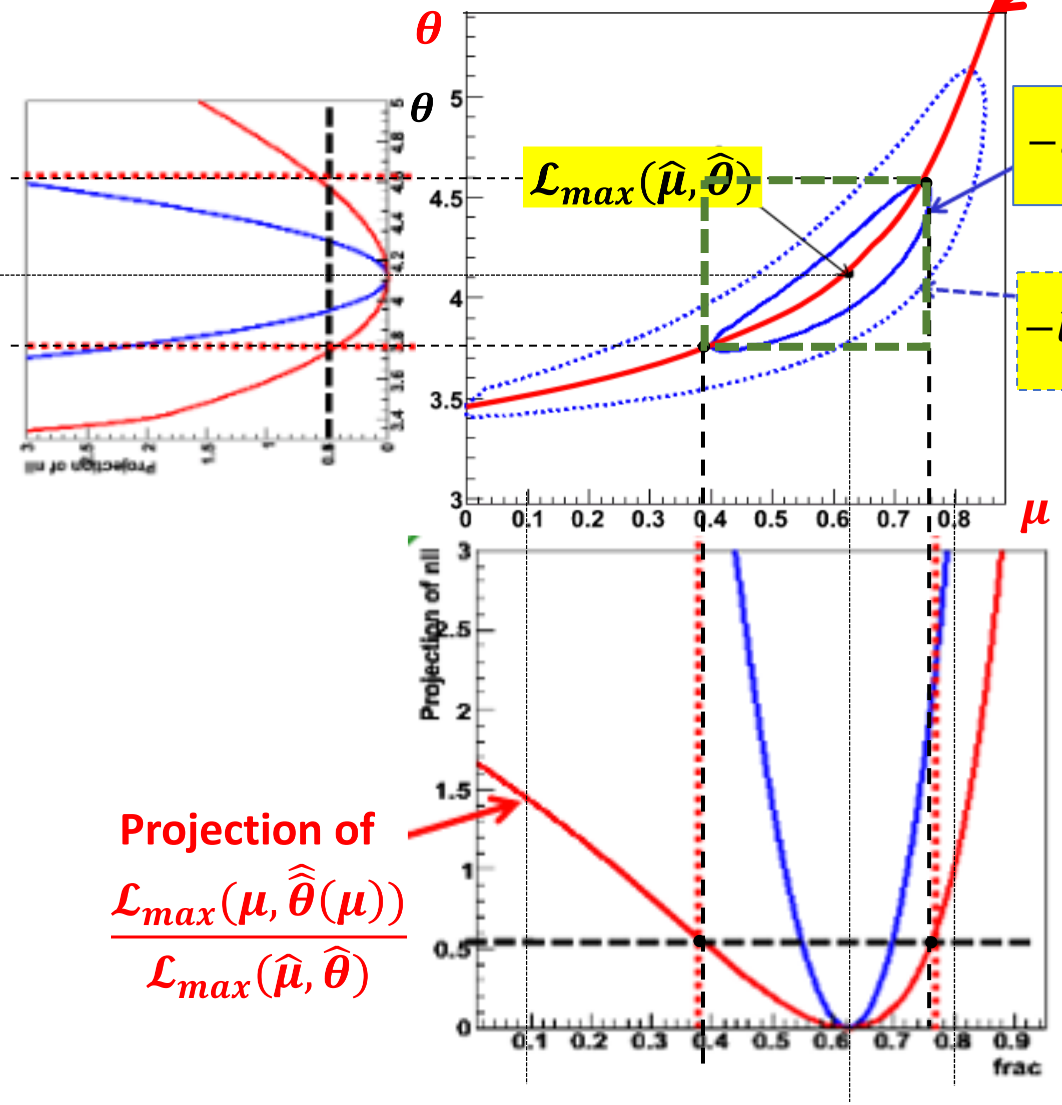
MINOS uncertainties by likelihood scan

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



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

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

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

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

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

MINOS uncertainties by likelihood scan

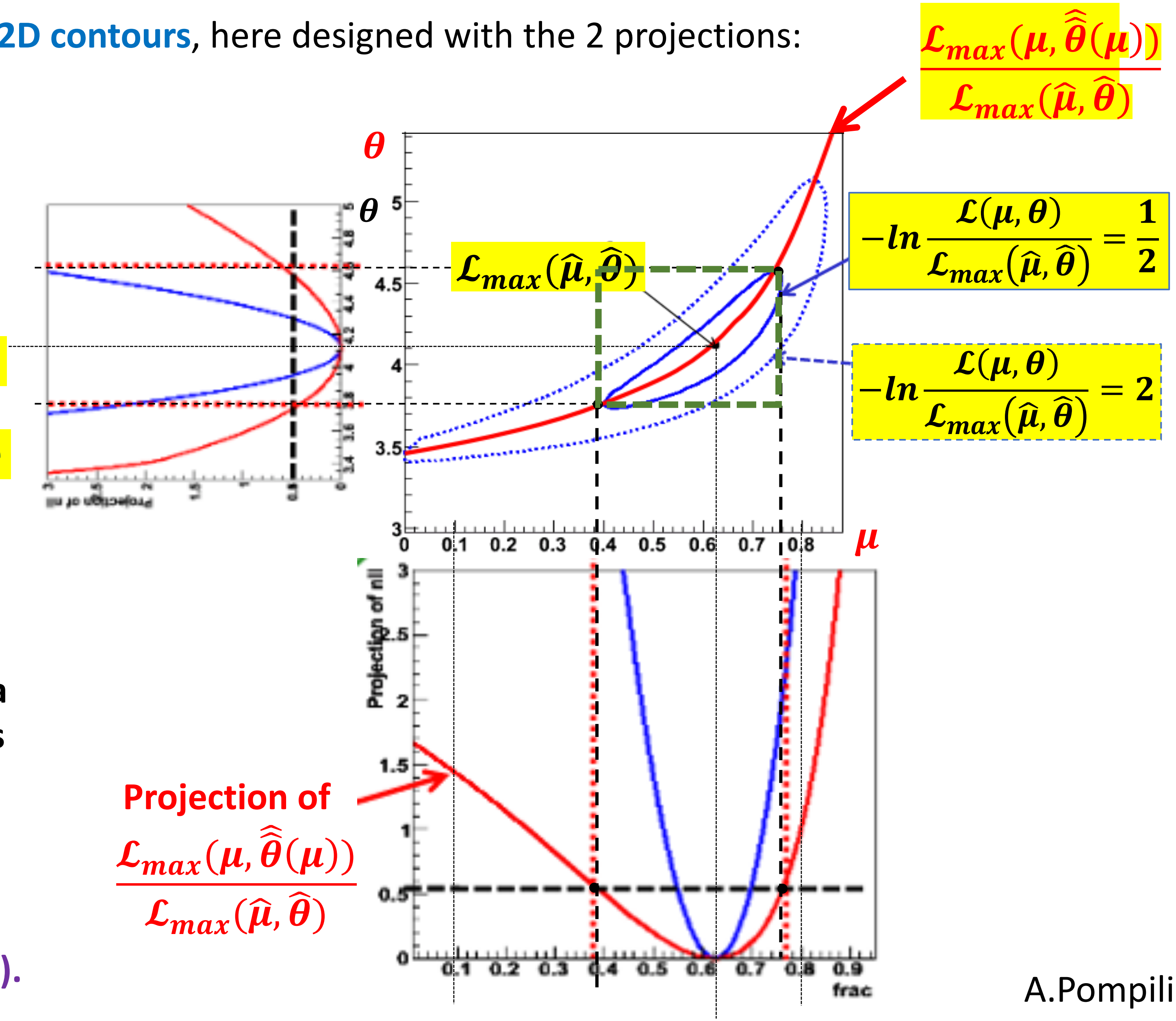
When there is also one POI and one NP one gets **2D contours**, here designed with the 2 projections:

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".

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 that: in the asymptotic regime (very large number of experiments) the MINOS algorithm 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 !