Follow up material for
PRACTICAL CLASS 3 & following
for the Course
Laboratorio Analisi Dati
2017/2018
Prof. A.Pompili

Some slides about MINUIT and ML fitting (MIGRAD, HESSE, MINOS)
[borrowed from RooFit tutorials]

Fitting and likelihood minimization

- What happens when you do `pdf->fitTo(*data)`
  - 1) Construct object representing –log of (extended) likelihood
  - 2) Minimize likelihood w.r.t floating parameters using MINUIT

- Can also do these two steps explicitly by hand (*)

```
// Construct function object representing –log(L)
RooAbsReal* nll = pdf.createNLL(data) ;

// Minimize nll w.r.t its parameters
RooMinuit m(*nll) ;
m.migrad() ;
m.hesse() ;
```

(*) You will see it explicitely later in PRACTICAL CLASS 7

Likelihood minimization – class RooMinuit

- Class RooMinuit is an *interface* to the ROOT implementation of the MINUIT minimization and error analysis package.

- RooMinuit takes care of
  - Passing value of miminized RooFit function to MINUIT
  - Propagated changes in parameters both from **`RooRealVar`** to MINUIT and back from MINUIT to **`RooRealVar`**, i.e. it keeps the state of RooFit objects synchronous with the MINUIT internal state
  - Propagate error analysis information back to **`RooRealVar`** parameters objects
  - Exposing high-level MINUIT operations to RooFit uses (MIGRAD,HESSE,MINOS) etc…
  - Making optional snapshots of complete MINUIT information (e.g. convergence state, full error matrix etc)

A brief description of MINUIT functionality

- MIGRAD
  - Find function minimum. Calculates function gradient, follow to (local) minimum, recalculate gradient, iterate until minimum found
    - To see what MIGRAD does, it is very instructive to do RooMinuit::setVerbose(1). It will print a line for each step through parameter space
  - Number of function calls required depends greatly on number of floating parameters, distance from function minimum and shape of function
- HESSE
  - Calculation of error matrix from 2^{nd} derivatives at minimum
  - Gives symmetric error. Valid in assumption that likelihood is (locally parabolic)

$$\hat{\sigma}(p)^2 = \hat{V}(p) = \left(\frac{d^2 \ln L}{d^2 p} \right)^{-1}$$

  - Requires roughly N^2 likelihood evaluations (with N = number of floating parameters)

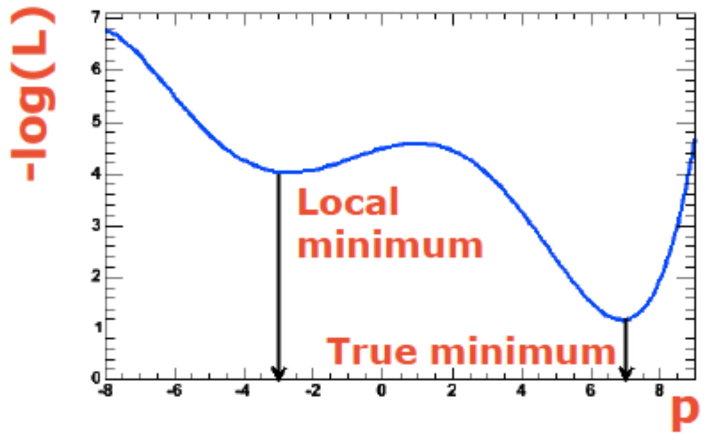
A brief description of MINUIT functionality

- MINOS
  - Calculate errors by explicit finding points (or contour for >1D) where Δ-log(L)=0.5
  - Reported errors can be asymmetric
  - Can be very expensive in with large number of floating parameters

- CONTOUR
  - Find contours of equal Δ-log(L) in two parameters and draw corresponding shape
  - Mostly an interactive analysis tool

Note of MIGRAD function minimization

- For all but the most trivial scenarios it is not possible to automatically find reasonable starting values of parameters
  - So you need to supply 'reasonable' starting values for your parameters



Reason: There may exist multiple (local) minima in the likelihood or χ^2

  - You may also need to supply 'reasonable' initial step size in parameters. (A step size 10x the range of the above plot is clearly unhelpful)
  - Using RooMinuit, the initial step size is the value of `RooRealVar::getError()`, so you can control this by supplying initial error values

Minuit function MIGRAD

- Purpose: find minimum

```
**********
**   13 **MIGRAD        1000            1
**********
(some output omitted)
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM MIGRAD    STATUS=CONVERGED      31 CALLS         32 TOTAL
                     EDM=2.36773e-06    STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT PARAMETER                                      STEP         FIRST
 NO.   NAME        VALUE            ERROR           SIZE       DERIVATIVE
  1  mean          8.84225e-02   3.23862e-01   3.58344e-04  -2.24755e-02
  2  sigma         3.20763e+00   2.39540e-01   2.78628e-04  -5.34724e-02
                                ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25     NPAR=  2    ERR DEF=0.5
 1.049e-01  3.338e-04
 3.338e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2
       1  0.00430   1.000  0.004
       2  0.00430   0.004  1.000
```

Progress information, watch for errors here

Parameter values and approximate errors reported by MINUIT

Error definition (in this case 0.5 for a likelihood fit)

Minuit function MIGRAD

- Purpose: find minimum

```
**********
**    13 **MIGR
**********
(some output o
MIGRAD MINIMIZ
MIGRAD WILL VERI                              RIX.
COVARIANCE        RIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM MIGRAD    STATUS=CONVERGED       31 CALLS          32 TOTAL
                    EDM=2.36773e-06     STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT PARAMETER                                        STEP         FIRST
 NO.   NAME       VALUE            ERROR              SIZE      DERIVATIVE
  1  mean          8.84225e-02   3.23862e-01   3.58344e-04  -2.24755e-02
  2  sigma         3.20763e+00   2.39540e-01   2.78628e-04  -5.34724e-02
                             ERR DEF= 0.5
EXTERNAL ERROR MATRIX.     NDIM=  25     NPAR=  2     ERR DEF=0.5
 1.049e-01  3.338e-04
 3.338e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL       1       2
       1  0.00430    1.000   0.004
       2  0.00430    0.004   1.000
```

Value of χ2 or likelihood at minimum

(NB: χ^2 values are not **divided by $N_{d.o.f}$)**

Approximate Error matrix And covariance matrix

Minuit function MIGRAD

- Purpose: find minim

```
**********
**    13 **MIGRAD        1000
**********
(some output omitted)
MIGRAD MINIMIZATION HAS CONVERGE
MIGRAD WILL VERIFY CONVERGENCE AND E        ATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM MIGRAD     STATUS=CONVERGED      31 CALLS          32 TOTAL
                     EDM=2.36773e-06     STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT PARAMETER                                       STEP         FIRST
 NO.   NAME       VALUE            ERROR          SIZE       DERIVATIVE
  1   mean         8.84225e-02   3.23862e-01   3.58344e-04  -2.24755e-02
  2   sigma        3.20763e+00   2.39540e-01   2.78628e-04  -5.34724e-02
                              ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  2    ERR DEF=0.5
 1.049e-01  3.338e-04
 3.338e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2
       1  0.00430   1.000  0.004
       2  0.00430   0.004  1.000
```

***Status*:**
Should be 'converged' but can be 'failed'

Estimated Distance to Minimum
should be small $O(10^{-6})$

Error Matrix Quality
should be 'accurate', but can be 'approximate' in case of trouble

Minuit function HESSE

- Purpose: calculate error matrix from $\frac{d^2 L}{dp^2}$

```
**********
**
***
COV                              UCCESSFULLY
FCN                              TUS=OK              10 CALLS          42 TOTAL
                                 4e-06     STRATEGY= 1        ERROR MATRIX ACCURATE
 EX                                                  INTERNAL       INTERNAL
 NO                                   ERROR          STEP SIZE       VALUE
  1                                3.23861e-01    7.16689e-05    8.84237e-03
  2  si               3.20763e+00  2.39539e-01    5.57256e-05    3.26535e-01
                                ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25     NPAR=  2     ERR DEF=0.5
 1.049e-01  2.780e-04
 2.780e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2
       1  0.00358    1.000  0.004
       2  0.00358    0.004  1.000
```

Error matrix (Covariance Matrix) calculated from

$$V_{ij} = \left(\frac{d^2(-\ln L)}{dp_i dp_j} \right)^{-1}$$

Minuit function HESSE

- Purpose: calculate error matrix from $\frac{d^2L}{dp^2}$

```
**********
**   18 **HESSE        1000
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM HESSE     STATUS=OK             10 CALLS          42 TOTAL
                    EDM=2.36534e-06    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                  INTERNAL      INTERNAL
 NO.   NAME      VALUE          ERROR           STEP SIZE       VALUE
  1  mean         8.84225e-02                                  8.84237e-03
  2  sigma        3.20763e+00                                  3.26535e-01

EXTERNAL ERROR MATRIX.    NDIM                                 F=0.5
 1.049e-01  2.780e-04
 2.780e-04  5.739e-02
PARAMETER  CORRELATION COEFFICIENT
      NO.  GLOBAL      1      2
       1  0.00358   1.000  0.004
       2  0.00358   0.004  1.000
```

Correlation matrix ρ_{ij} calculated from

$$V_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Minuit function HESSE

- Purpose: calculate error matrix from $\frac{d^2L}{dp^2}$

```
**********
**    18 **HESSE         1000
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM HESSE      STATUS=OK           10 CALLS         42 TOTAL
                    EDM=2.36534e-06    STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT PARAMETER                                  INTERNAL       INTERNAL
 NO.   NAME       VALUE           ERROR         STEP SIZE       VALUE
  1  mean                                       7.16689e-05    8.84237e-03
  2  sigma                                      5.57256e-05    3.26535e-01

EXTERNAL ERROR                                  2    ERR DEF=0.5
 1.049e-01  2.
 2.780e-04  5.739e-
PARAMETER  CORRELA     N COEFFICIENTS
      NO.  GLOBAL       1      2
       1  0.00358   1.000  0.004
       2  0.00358   0.004  1.000
```

Global correlation vector: correlation of each parameter with *all other* parameters

Minuit function MINOS

- Error analysis through Δnll contour finding

```
**********
**    23 **MINOS        1000
**********
FCN=257.304 FROM MINOS     STATUS=SUCCESSFUL     52 CALLS          94 TOTAL
                    EDM=2.36534e-06    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                  PARABOLIC         MINOS ERRORS
 NO.   NAME      VALUE            ERROR      NEGATIVE      POSITIVE
  1  mean         8.84225e-02   3.23861e-01  -3.24688e-01   3.25391e-01
  2  sigma        3.20763e+00   2.39539e-01  -2.23321e-01   2.58893e-01
                               ERR DEF= 0.5
```

Symmetric error

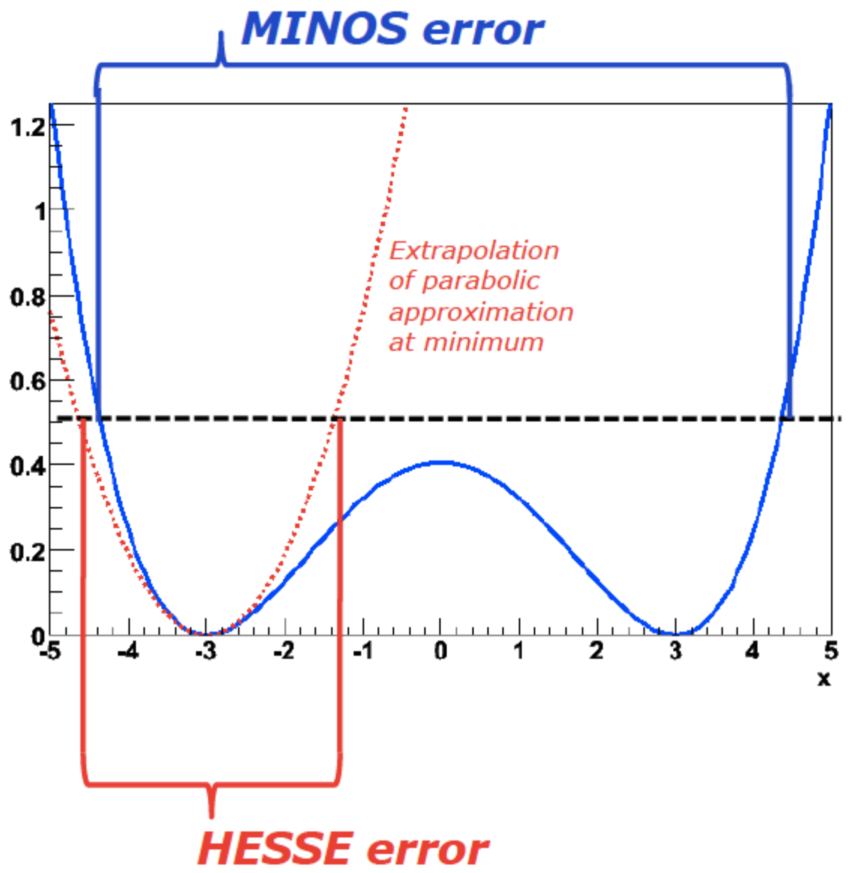
(repeated result from HESSE)

MINOS error Can be asymmetric

(in this example the 'sigma' error is slightly asymmetric)

Illustration of difference between HESSE and MINOS errors

- 'Pathological' example likelihood with multiple minima and non-parabolic behavior



Practical estimation – Fit converge problems

- Sometimes fits don't converge because, e.g.
  - MIGRAD unable to find minimum
  - HESSE finds negative second derivatives (which would imply negative errors)

- Reason is usually numerical precision and stability problems, but
  - The underlying cause of fit stability problems is usually by **highly correlated parameters** in fit

- HESSE correlation matrix in primary investigative tool

PARAMETER NO.	CORRELATION GLOBAL	COEFFICIENTS 1	2
1	0.99835	1.000	0.998
2	0.99835	0.998	1.000

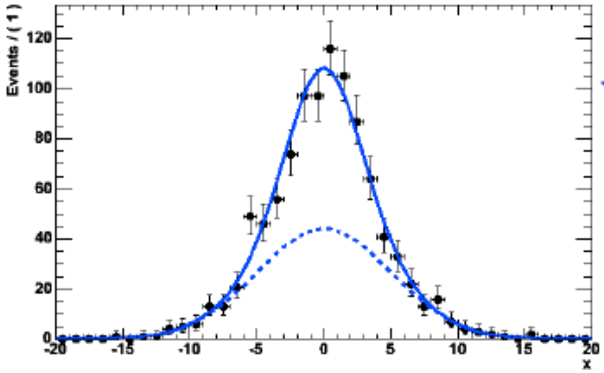
Signs of trouble...

  - In limit of 100% correlation, the usual point solution becomes a line solution (or surface solution) in parameter space. Minimization problem is no longer well defined

Mitigating fit stability problems

- Strategy I – More orthogonal choice of parameters
  - Example: fitting sum of 2 Gaussians of similar width

$$F(x; f, m, s_1, s_2) = fG_1(x; s_1, m) + (1-f)G_2(x; s_2, m)$$



HESSE correlation matrix

PARAMETER NO.	CORRELATION COEFFICIENTS GLOBAL	[f]	[m]	[s1]	[s2]
[f]	0.96973	1.000	-0.135	0.918	0.915
[m]	0.14407	-0.135	1.000	-0.144	-0.114
[s1]	0.92762	0.918	-0.144	1.000	0.786
[s2]	0.92486	0.915	-0.114	0.786	1.000

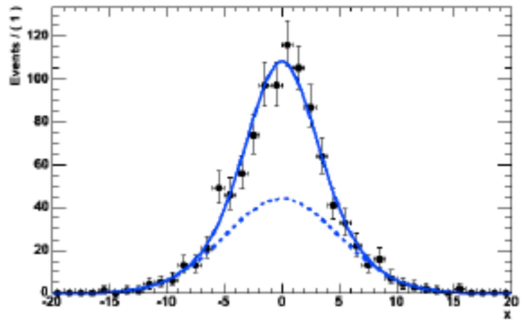
Widths s_1,s_2 strongly correlated fraction f

Mitigating fit stability problems

- Different parameterization:

$$fG_1(x;s_1,m_1)+(1-f)G_2(x;\underline{s_1\cdot s_2},m_2)$$

PARAMETER NO.	CORRELATION COEFFICIENTS GLOBAL	[f]	[m]	[s1]	[s2]
[f]	0.96951	1.000	-0.134	0.917	-0.681
[m]	0.14312	-0.134	1.000	-0.143	0.127
[s1]	0.98879	0.917	-0.143	1.000	-0.895
[s2]	0.96156	0.681	0.127	-0.895	1.000



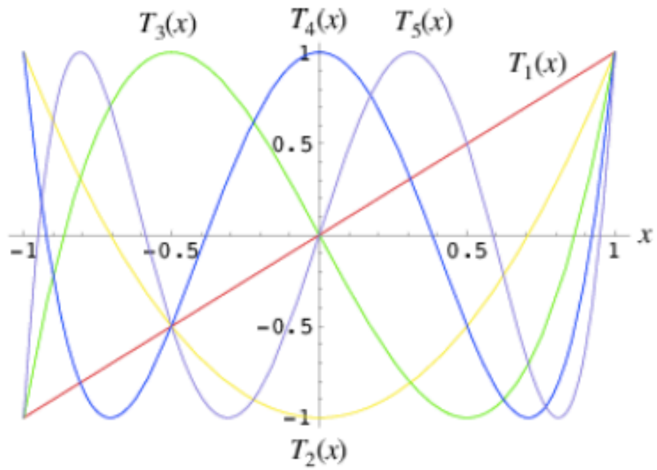
- Correlation of width s2 and fraction f reduced from 0.92 to 0.68
- Choice of parameterization matters!

- **Strategy II – Fix all but one of the correlated parameters**
  - If floating parameters are highly correlated, some of them may be redundant and not contribute to additional degrees of freedom in your model

Mitigating fit stability problems -- Polynomials

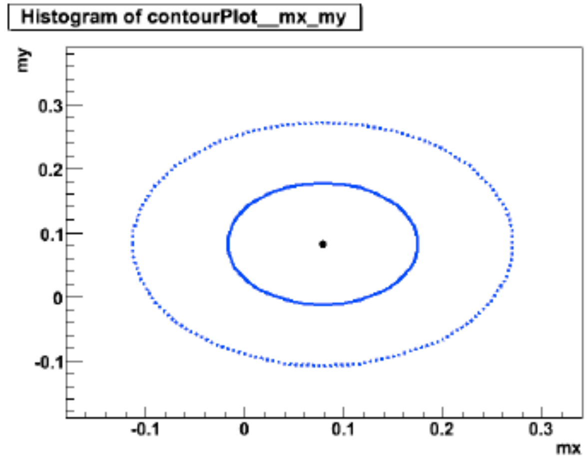
- **Warning:** Regular parameterization of polynomials $a_0+a_1x+a_2x^2+a_3x^3$ nearly always results in strong correlations between the coefficients a_i.
  - *Fit stability problems, inability to find right solution common at higher orders*
- **Solution:** Use existing parameterizations of polynomials that have (mostly) uncorrelated variables
  - ***Example: Chebychev polynomials***

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_2(x) &= 2x^2 - 1 \\
T_3(x) &= 4x^3 - 3x \\
T_4(x) &= 8x^4 - 8x^2 + 1 \\
T_5(x) &= 16x^5 - 20x^3 + 5x \\
T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1.
\end{aligned}
$$



Minuit CONTOUR tool also useful to examine 'bad' correlations

- Example of 1,2 sigma contour of two uncorrelated variables
  - Elliptical shape. In this example parameters are uncorrelation



- Example of 1,2 sigma contour of two variables with problematic correlation
  - Pdf = f·G1(x,0,3)+(1-f)·G2(x,0,s) with s=4 in data