roofit tutorial

references

http://roofit.sourceforge.net/quicktour/first.html

https://www.slideserve.com/denna/statistical-methods-for-data-analysis-parameter-estimates-with-roofit

https://www.slideserve.com/gomer/deviations-from-exponential-decay

RooFit: Your toolkit for data modeling

What is it?

- A powerful toolkit for modeling the expected distribution(s) of events in a physics analysis
- Primarily targeted to high-energy physicists using ROOT
- Originally developed for the BaBar collaboration by Wouter Verkerke and David Kirkby.
- Included with ROOT v5.xx

Documentation:

- http://root.cern.ch/root/Reference.html – for latest class descriptions. RooFit classes start with "Roo".
- http://roofit.sourceforge.net – for documentation and tutorials

Tutorials:

- Dig $ROOTSYS/tutorials/rootfit

Math – The Likelihood estimator

- Definition of Likelihood
  - given **D($\vec{x}$)** and **F($\vec{x}$;$\vec{p}$)**

Functions used in likelihoods must be Probability Density Functions:

$$\int F(\vec{x};\vec{p})d\vec{x} \equiv 1, \quad F(\vec{x};\vec{p}) > 0$$

$$L(\vec{p}) = \prod_i F(\vec{x}_i;\vec{p}), \quad \text{i.e.} \quad L(\vec{p}) = F(x_0;\vec{p}) \cdot F(x_1;\vec{p}) \cdot F(x_2;\vec{p})...$$

  - For convenience the *negative log of the Likelihood* is often used

$$-\ln L(\vec{p}) = -\sum_i \ln F(\vec{x}_i;\vec{p})$$

- Parameters are estimated by maximizing the Likelihood, or equivalently minimizing –log(L)

$$\left.\frac{d\ln L(\vec{p})}{d\vec{p}}\right|_{p_i=\hat{p}_i} = 0$$

Math – Variance on ML parameter estimates

- **Estimator** for the **parameter variance** is

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

  - I.e. variance is estimated from 2nd derivative of –log(L) at minimum
  - Valid if estimator is **efficient** and **unbiased**!

From Rao-Cramer-Frechet inequality

$$V(\hat{p}) \geq \frac{1+\frac{db}{dp}}{\left(\frac{d^2\ln L}{d^2 p}\right)}$$

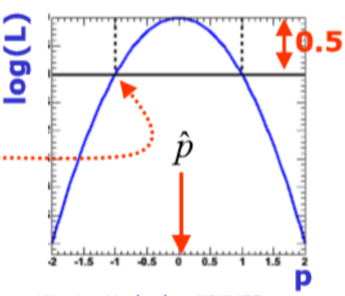
b = bias as function of p, inequality becomes equality in limit of efficient estimator

- **Visual interpretation** of variance estimate
  - Taylor expand –log(L) around minimum

$$\ln L(p) = \ln L(\hat{p}) + \left.\frac{d\ln L}{dp}\right|_{p=\hat{p}}(p-\hat{p}) + \tfrac{1}{2}\left.\frac{d^2\ln L}{d^2 p}\right|_{p=\hat{p}}(p-\hat{p})^2$$

$$= \ln L_{\max} + \left.\frac{d^2\ln L}{d^2 p}\right|_{p=\hat{p}}\frac{(p-\hat{p})^2}{2}$$

$$= \ln L_{\max} + \frac{(p-\hat{p})^2}{2\hat{\sigma}_p^2} \Rightarrow \ln L(p\pm\sigma) = \ln L_{\max} - \frac{1}{2}$$



RooFit core design philosophy

- Mathematical objects are represented as C++ objects

Mathematical concept		RooFit class
variable	x	`RooRealVar`
function	$f(x)$	`RooAbsReal`
PDF	$f(x)$	`RooAbsPdf`
space point	$\vec{x}$	`RooArgSet`
integral	$\int_{x_{\min}}^{x_{\max}} f(x)dx$	`RooRealIntegral`
list of space points		`RooAbsData`

Likelihood function

- Given a sample of N events each with variables $(x_1, \ldots, x_n)$, the likelihood function expresses the probability density of the sample, as a function of the unknown parameters:

$$L = \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m)$$

- Sometimes the used notation for parameters is the same as for conditional probability:

$$f(x_1, \cdots, x_n | \theta_1, \cdots, \theta_m)$$

- If the size N of the sample is also a random variable, the extended likelihood function is also used:

$$L = p(N; \theta_1, \cdots, \theta_m) \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m)$$

  - Where p is most of the times a Poisson distribution whose average is a function of the unknown parameters

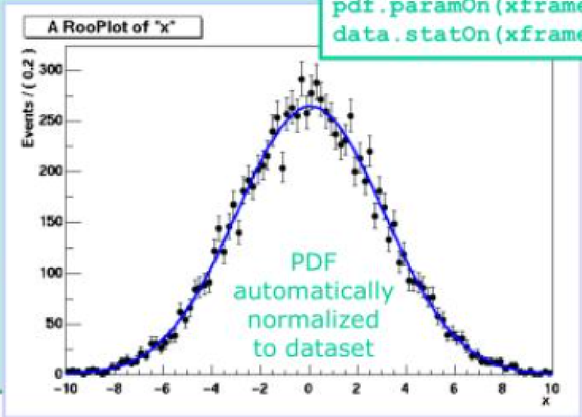
- In many cases it is convenient to use $-\ln L$ or $-2\ln L$: $\prod_i \rightarrow \sum_i$

Example


```
RooRealVar x("x","x",-10,10) ;
RooRealVar mean("mean","mean of gaussian",0,-10,10);
RooRealVar sigma("sigma","width of gaussian",3);

   RooGaussian gauss("gauss","gaussian PDF",x,mean,sigma);

RooDataSet* data = gauss.generate(x,10000);

// ML fit is the default
gauss.fitTo(*data);

mean.Print();
// RooRealVar::mean =
// 0.0172335 +/- 0.0299542
sigma.Print();
// RooRealVar::sigma =
// 2.98094  +/- 0.0217306

RooPlot* xframe = x.frame();
data->plotOn(xframe);
gauss.plotOn(xframe);
xframe->Draw();
```

Further drawing options:

```
pdf.paramOn(xframe,data);
data.statOn(xframe);
```



Extended ML fits

- Specify extended ML fit adding one extra parameter:

```
pdf.fitTo(*data,
  RooFit::Extended(kTRUE));
```

Import external data sets

- Read a ROOT tree:

```
RooRealVar x("x","x",-10,10);
RooRealVar c("c","c",0,30);
RooDataSet data("data","data",inputTree,
                RooArgSet(x,c));
```

  – Automatic removal of entries out of variable range

- Read an ASCII file:

```
RooDataSet* data =
  RooDataSet::read("ascii.file",
                   RooArgList(x,c));
```

  – Une line per entry; variable order given by argument list

Histogram fits

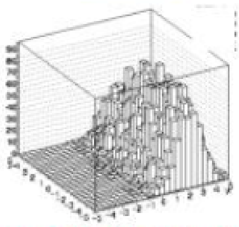
- Use a binned data set:
  – **RooDataHist** instead of **RooDataSet**
- Fit with binned model

Unbinned

x	y	z
1	3	5
2	4	6
1	3	5
2	4	6

RooDataSet

Binned



RooDataHist

RooAbsData

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 **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
```

Value of χ^2 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 minimu


```
**********
**    13 **MIGRAD        1000
**********
(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
```


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                           TOTAL
                     EDM=2.36534e-06    STRAT                    CURATE
 EXT PARAMETER                                 INTERNAL      INTERNAL
 NO.    NAME       VALUE            ERROR       STEP SIZE       VALUE
  1  mean          8.84225e-02   3.23861e-01   7.16689e-05   8.84237e-03
  2  sigma         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
```

Minuit function HESSE

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

```
**********
**
***
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

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

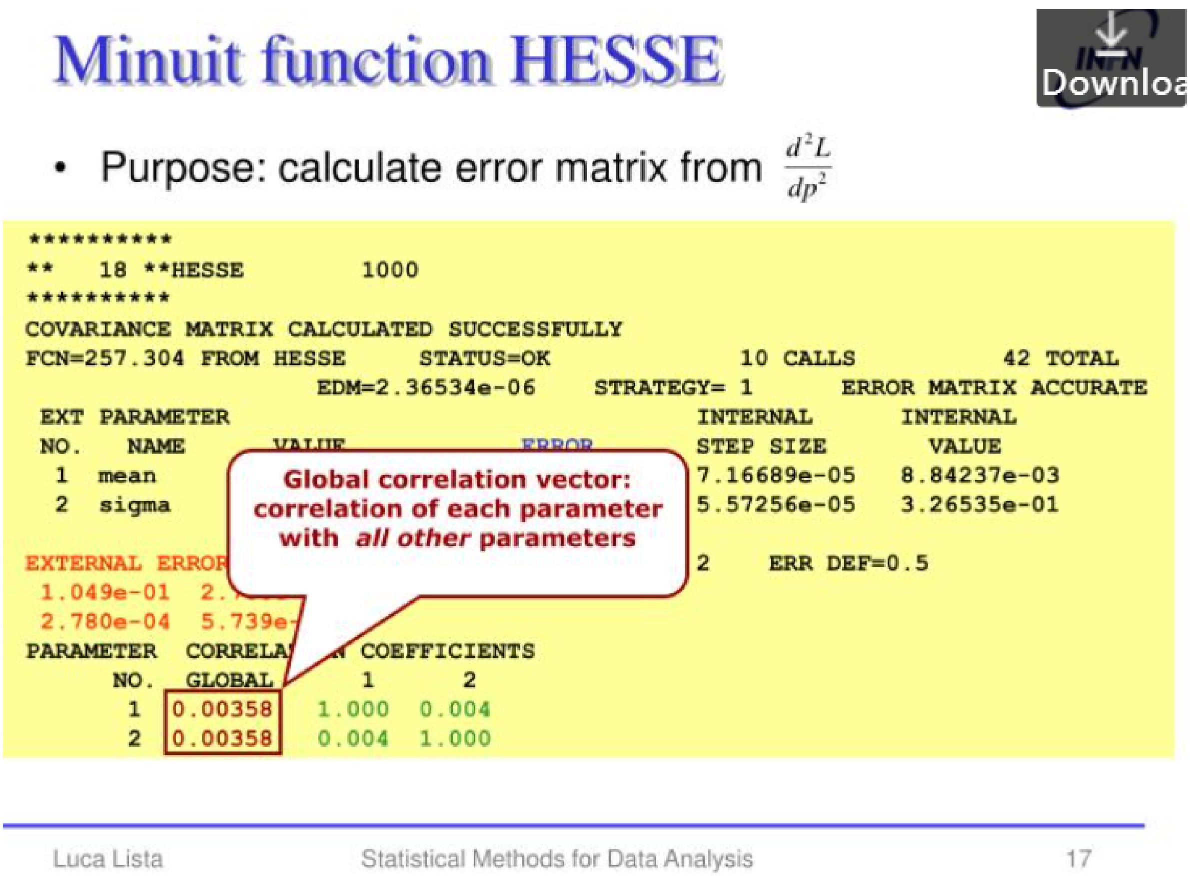
Minuit function HESSE

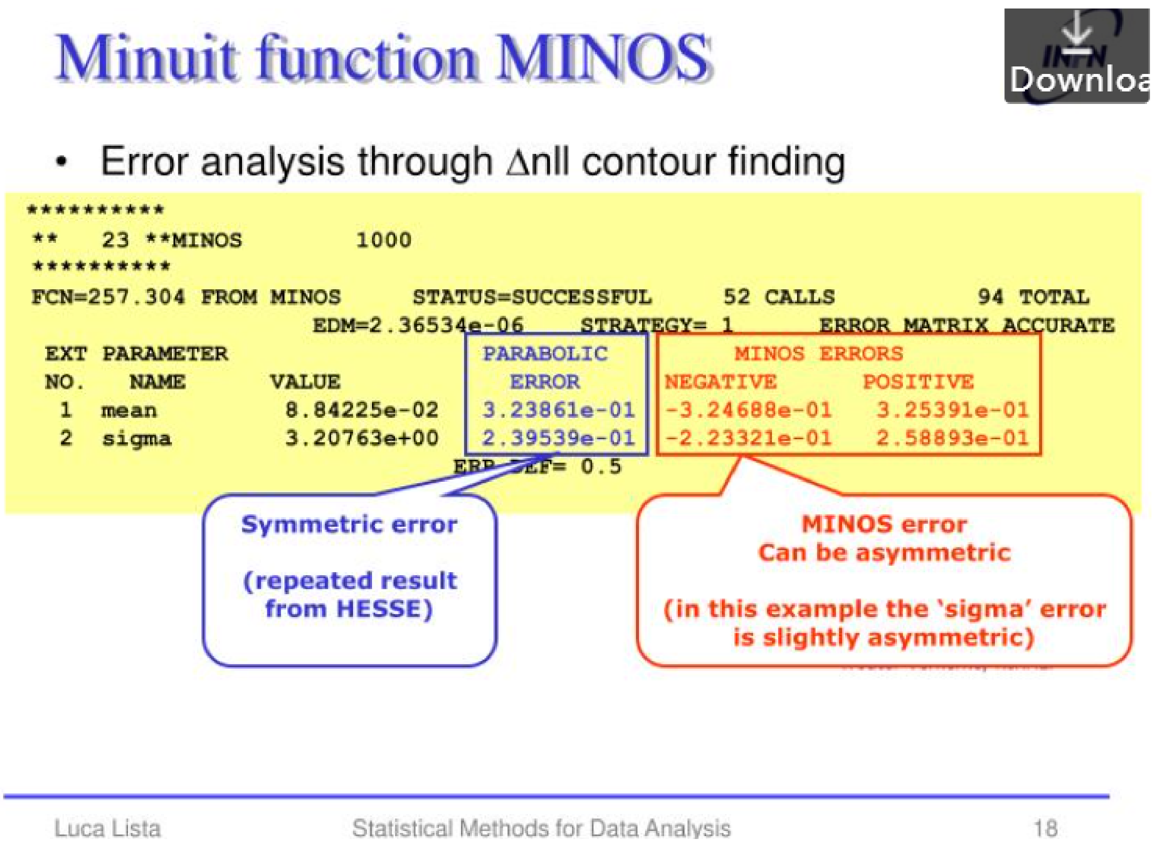
- Purpose: calculate error matrix from $\frac{d^2L}{dp^2}$

```
**********
**    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

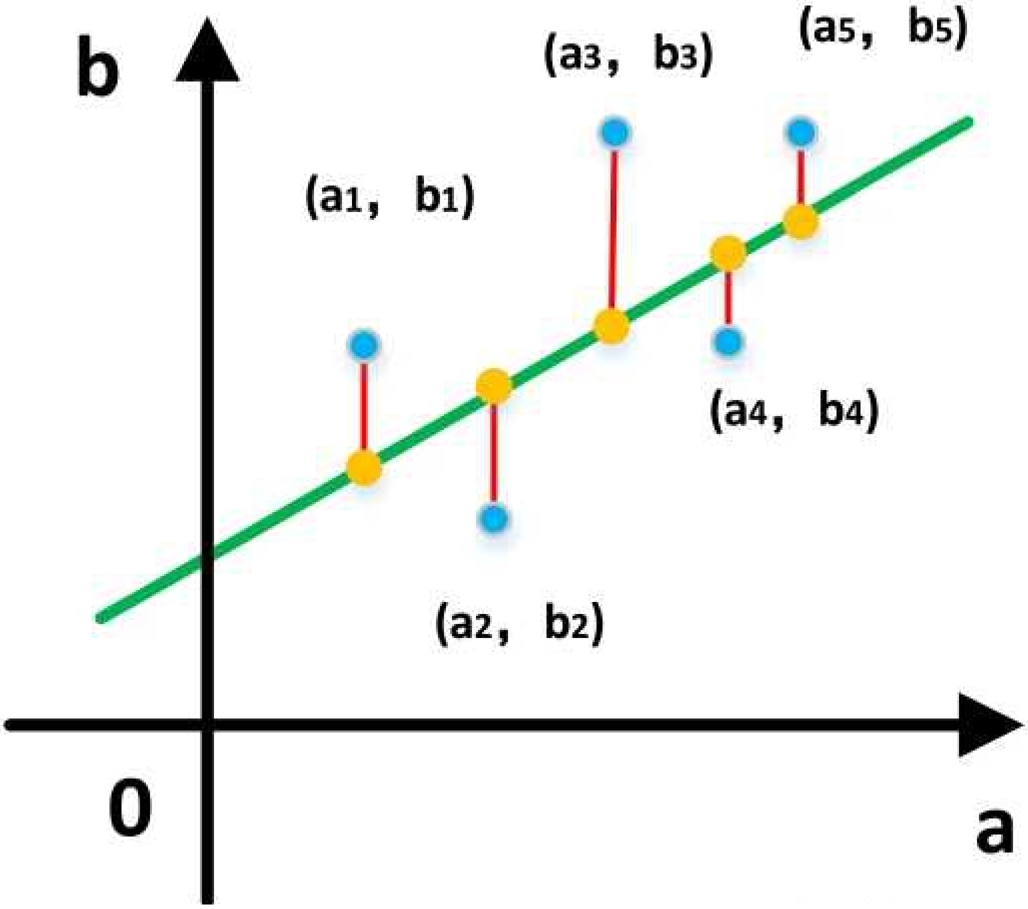
$$V_{ij} = \sigma_i \sigma_j \rho_{ij}$$





最小二乘法和最大似然法

最小二乘法



代表：hist->Fit(f1);

最大似然法

Likelihood function

- Given a sample of N events each with variables $(x_1, \ldots, x_n)$, the likelihood function expresses the probability density of the sample, as a function of the unknown parameters:

$$L = \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m)$$

- Sometimes the used notation for parameters is the same as for conditional probability:

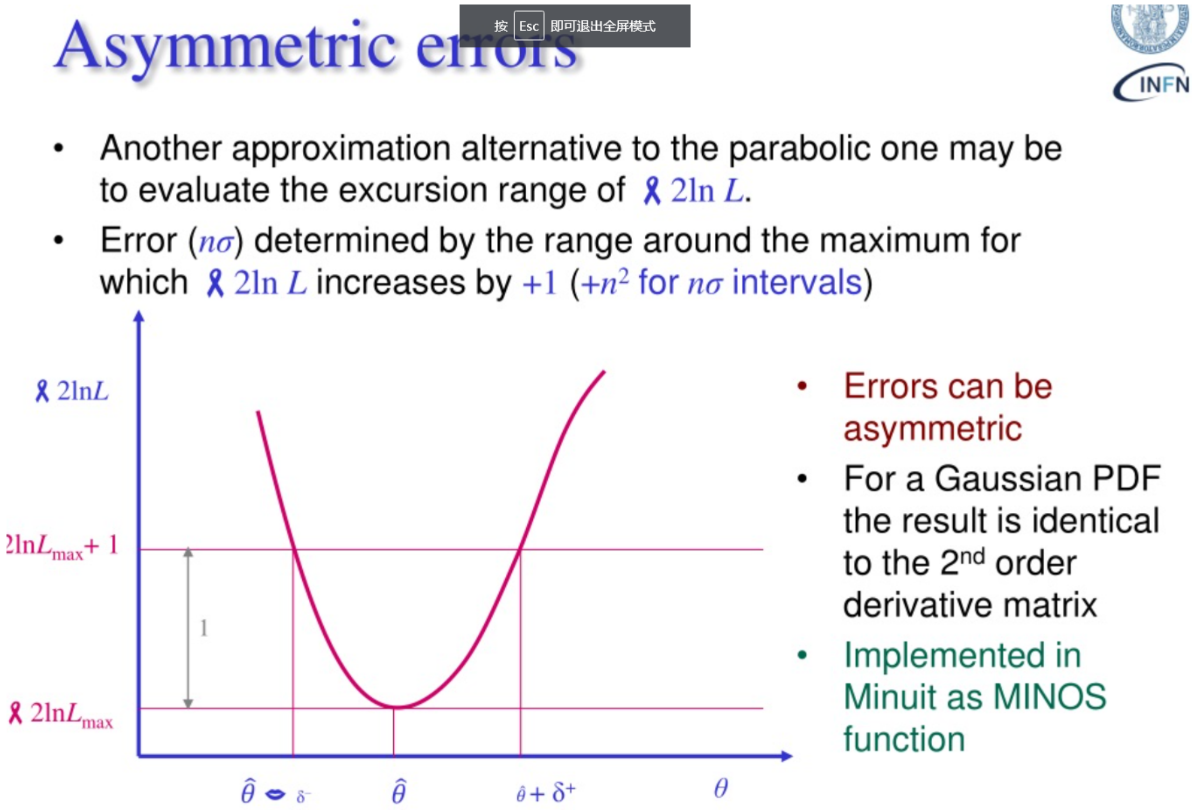
$$f(x_1, \cdots, x_n | \theta_1, \cdots, \theta_m)$$

- If the size N of the sample is also a random variable, the extended likelihood function is also used:

$$L = p(N; \theta_1, \cdots, \theta_m) \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m)$$

  - Where p is most of the times a Poisson distribution whose average is a function of the unknown parameters

- In many cases it is convenient to use $-\ln L$ or $-2\ln L$: $\prod_i \rightarrow \sum_i$



roofit是一个容易进行最大似然拟合的框架

roofit也可以进行最小二乘拟合（rf602_chi2fit.C）

```

// Construct a chi^2 of the data and the model.
// When a p.d.f. is used in a chi^2 fit, the probability density scaled
// by the number of events in the dataset to obtain the fit function
// If model is an extended p.d.f, the expected number events is used
// instead of the observed number of events.
model.chi2FitTo(*dh) ;
return;
```

root 的Fit是常用的最小二乘拟合手段，

但是调用minuit方法，也可以进行最大似然拟合