Computational Physics

Ch 05 - Likelihood principle

Korea University
Eunil Won

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

The Maximum Likelihood Estimators

- The maximum likelihood estimators

  Consider a random variable x distributed according to a p.d.f. $f(x;\boldsymbol{\theta})$. Suppose the form of $f(x;\boldsymbol{\theta})$ is known but the value of at least one of $\boldsymbol{\theta}$ is unknown.

  ex) $f(x;\boldsymbol{\theta}) = \theta_0 + \theta_1 x$, and $\{\theta_0, \theta_1\}$ are unknown.

Now, for a set of measurements $x_1,...,x_n$, we can consider the following probability:

$$\text{probability that } x_i \in [x_i, x_i + dx_i] \text{ for all } i = \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta})dx_i$$

If the hypothesized p.d.f. and parameter values are correct, one expects a high probability for the data that were actually measured.

One defines this product as the likelihood:

$$L(x_1, x_2, ..., x_n|\boldsymbol{\theta}) = L(\mathbf{x}|\boldsymbol{\theta}) \equiv \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta})$$

Maximum-Likelihood Principle tells us that we should choose as an estimate of the unknown set parameters $\boldsymbol{\theta}$ that maximizes the likelihood.

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

The Maximum Likelihood Estimators

This means that the estimates $\hat{\boldsymbol{\theta}}$ is such that

$$L(\mathbf{x}|\hat{\boldsymbol{\theta}}) \geq L(\mathbf{x}|\boldsymbol{\theta})$$

for all conceivable values of $\boldsymbol{\theta}$.

This means, in the case of the single parameter:

$$\left.\frac{\partial L(\mathbf{x}|\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial}{\partial \theta}\prod_{i=1}^{n} f(x_i|\theta)\right|_{\theta=\hat{\theta}} = 0, \quad \left.\frac{\partial^2 L(\mathbf{x}|\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial^2}{\partial \theta^2}\prod_{i=1}^{n} f(x_i|\theta)\right|_{\theta=\hat{\theta}} < 0,$$

Since *L* and the logarithm of *L* attain their maxima for the same value of θ, one usually take logarithm of *L*: because the sum is often easier to handle than the product.

$$\left.\frac{\partial}{\partial \theta}\ln L(\mathbf{x}|\theta)\right|_{\theta=\hat{\theta}} = \left.\frac{\partial}{\partial \theta}\sum_{i=1}^{n}\ln f(x_i|\theta)\right|_{\theta=\hat{\theta}} = 0, \quad \left.\frac{\partial^2}{\partial \theta^2}\ln L(\mathbf{x}|\theta)\right|_{\theta=\hat{\theta}} = \left.\frac{\partial^2}{\partial \theta^2}\sum_{i=1}^{n}\ln f(x_i|\theta)\right|_{\theta=\hat{\theta}} < 0$$

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

The Maximum Likelihood Estimators

Example: estimate of mean lifetime

Let us assume that we observe the production and decay of a certain type of particles in an infinite detector. The p.d.f. is then

$$f(t|\tau) = \frac{1}{\tau} e^{-t/\tau}, \qquad 0 \le t \le \infty$$

where τ is the mean lifetime of the particle to be estimated. For n observed t_i, the likelihood function is

$$L(t|\tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau}$$

and the maximum likelihood estimator for τ is found by solving the equation

$$\frac{\partial \ln L(t|\tau)}{\partial \tau} = \frac{\partial}{\partial \tau} \sum_{i=1}^{n} \left(-\ln \tau - \frac{t_i}{\tau} \right) = \sum_{i=1}^{n} \left(-\frac{1}{\tau} + \frac{t_i}{\tau^2} \right) = 0 \quad \therefore \hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i = \bar{t}$$

Hence the maximum likelihood estimate of the parameter τ is equal to the arithmetic mean of the observed flight-times. The solution obtained does correspond to a maximum of the likelihood because

$$\left. \frac{\partial^2 \ln L(t|\tau)}{\partial \tau^2} \right|_{\tau = \hat{\tau}} = -\frac{n}{\hat{\tau}^2} < 0$$

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

The Maximum Likelihood Estimators

Example: measurements with common error

Let $x_1, x_2, .., x_N$ be n independent measurements on the same unknown quantity μ, and assume that the measurement error is σ, common to all observations. We assume that x_i's to constitute a sample of size n drawn from a normal population $N(\mu,\sigma^2)$, where μ is unknown, σ^2 is known.

Let us construct the likelihood first: $L(\mathbf{x};\sigma|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$

Taking a log of it and differentiation w.r.t. μ gives

$$\frac{\partial \ln L}{\partial \mu} = \frac{\partial}{\partial \mu} \sum_{i=1}^{n} \Big(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\big(\frac{x_i-\mu}{\sigma}\big)^2 \Big) = 0$$

and the solution is then

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

so the maximum likelihood estimate of the population mean μ is equal to the sample mean.

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

The Maximum Likelihood Estimators

Example: measurements with different errors (weighted mean)

We assume that the measurements $x_1, x_2, .., x_N$ on the unknown quantity μ have different, but still known errors. If each measurement x_i is normally distributed with measuring error σ_i, the likelihood function is

$$L(\mathbf{x};\boldsymbol{\sigma}|\mu) = L(x_1, ..., x_n; \sigma_1, ..., \sigma_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right)}$$

In this case, the maximum likelihood estimate for μ becomes

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

which is called the *weighted mean* of the observations.

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

The Maximum Likelihood Estimators

Example: simultaneous estimation of mean and variance

We assume that the measurements $x_1, x_2, .., x_N$ on the unknown quantity μ have unknown error σ that is common for all measurements. Our likelihood becomes

$$L(\mathbf{x}|\mu,\sigma^2) = \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)}$$

So, we require followings simultaneously to get estimates:

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma} = 0 \qquad \frac{\partial \ln L}{\partial \sigma^2} = \sum_{i=1}^{n} \left(-\frac{1}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}(x_i-\mu)^2\right) = 0$$

In this case, we get

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Advanced: the maximum likelihood estimate of σ^2 is biased: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Variance of Maximum-Likelihood Estimators

- General methods for variance estimation

Let us regard the likelihood function $L(\mathbf{x}|\theta) = \Pi\, f(x_i|\theta)$ as the joint p.d.f. of the *n* variables $x_1, x_2, ..., x_n$ for the *k* parameters $\theta_1, \theta_2, ..., \theta_k$. Then, the covariance term may be defined as

$$V_{ij}(\hat{\boldsymbol{\theta}}) = \int (\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) L(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}$$

where the integration is over all x_i's and $\boldsymbol{\theta}=\{\theta_1, \theta_2, ..., \theta_k\}$ represent the true values of the parameters. Therefore, the formula above can be used to find the covariance matrix from the given $f(x|\boldsymbol{\theta})$ alone without having any data available.

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Variance of Maximum-Likelihood Estimators

- Example: Variance of the lifetime estimate

The maximum likelihood estimate for the lifetime was $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$ for the one parameter p.d.f. $f(t|\tau) = \frac{1}{\tau}e^{-t/\tau}, \quad 0 \le t \le \infty$

So, the variance becomes $V(\hat{\tau}) = \int_0^\infty \ldots \int_0^\infty (\hat{\tau} - \tau)^2 \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau} dt_i$

and this can be written as

$$V(\hat{\tau}) = \int \Big(\frac{1}{n}\sum_{k=1}^{n} t_k\Big)\Big(\frac{1}{n}\sum_{j=1}^{n} t_j\Big) \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau} dt_i - 2\tau \int \Big(\frac{1}{n}\sum_{k=1}^{n} t_k\Big) \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau} dt_i + \tau^2 \int \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau} dt_i$$

(the integrations are from zero to infinity for all the *n* variables t_i)

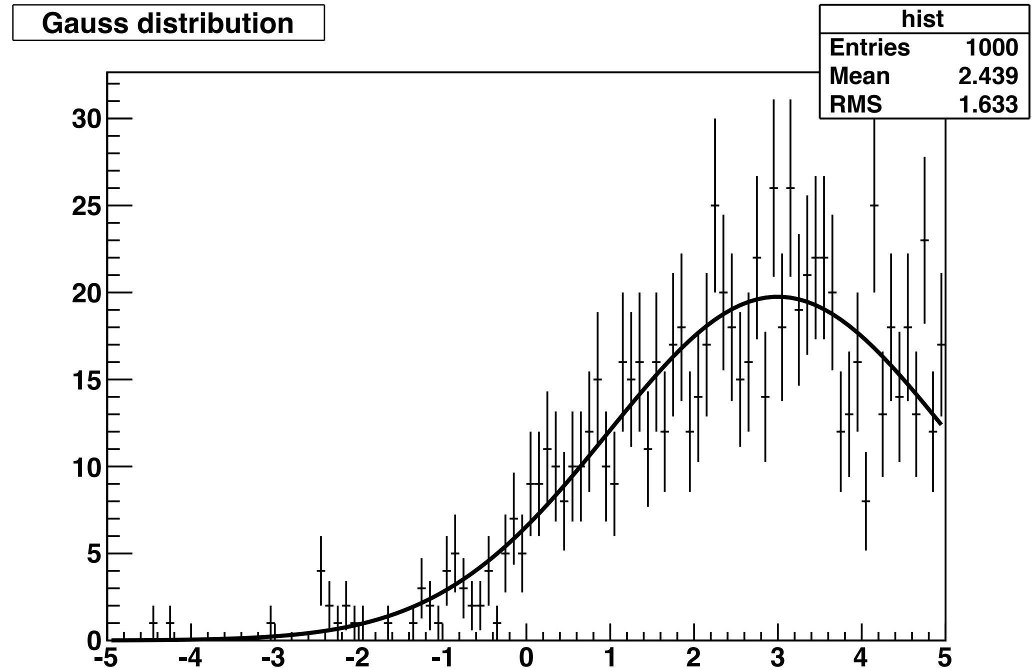
Now, it can be shown as

$$V(\hat{\tau}) = \Big(\frac{2}{n}\tau^2 + \frac{n-1}{n}\tau^2\Big) - (2\tau^2) + \tau^2 = \frac{\tau^2}{n}$$

A precise measurement of the lifetime is possible with a large size data set.

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Likelihood fit with ROOT



- "Unbinned" maximum likelihood fit

1) Generate N random numbers that are distributed according to a gaussian PDF

2) Construct the likelihood function, something similar to

$$L(\mathbf{x}|\mu,\sigma^2) = \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)}$$

3) Since MINUIT minimize a given function, instead of maximizing the likelihood, we minimize -ln(likelihood).

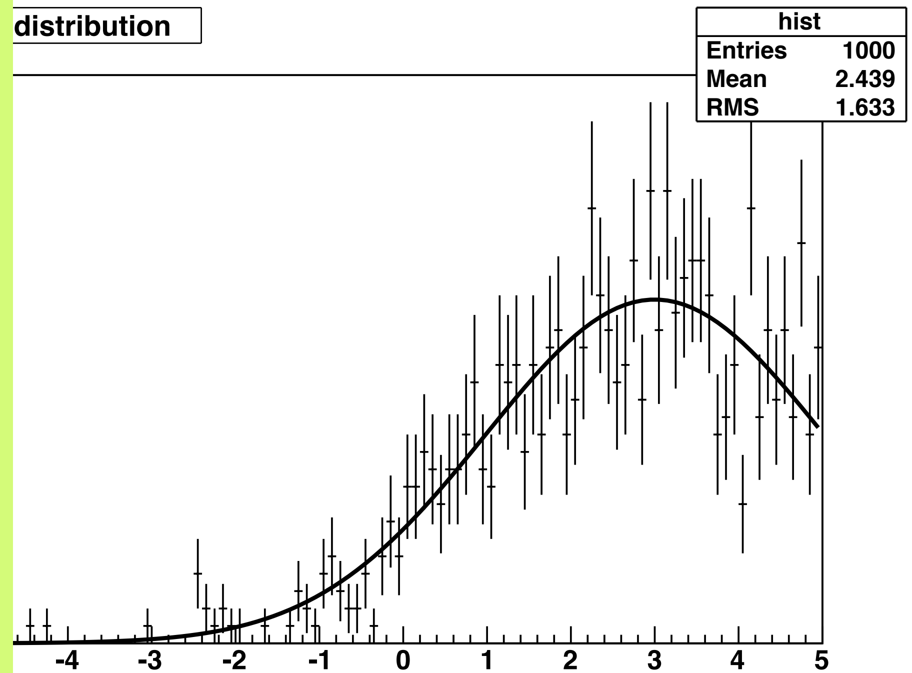
4) We can obtain parameters (central values and one sigma of the gaussian distribution in this case)

5) Let us visualize the fit result (maybe tricky)

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Likelihood fit with ROOT

```
#include "TMinuit.h"

const static Int_t   Ngen = 1000;
const static Int_t   Nbin = 100;
const static Float_t Xmax =  5.0;
const static Float_t Xmin = -5.0;
const static Float_t Mu0  = 3.0;
const static Float_t Sig0 = 2.0;

Float_t gdata[Ngen];

Double_t f1(Double_t *x,Double_t *par)
{
  Double_t value = 1.0/(TMath::Sqrt(2.0*3.1415926)*par[1])
                 * TMath::Exp(-0.5*(x[0]-par[0])*(x[0]-par[0])/(par[1]*par[1]));
  return value*Ngen*(Xmax-Xmin)/(Nbin);
}

Double_t func(float x,Double_t *par)
{
  Double_t value = 1.0/(TMath::Sqrt(2.0*3.1415926)*par[1])
                 * TMath::Exp(-0.5*(x-par[0])*(x-par[0])/(par[1]*par[1]));
  return value;
}

void fcn(Int_t &npar, Double_t *gin, Double_t &f, Double_t *par, Int_t iflag)
{
  Int_t i;

  //
  // calculate the -log(likelihood)
  //
  Double_t loglike = 0;
  Double_t sumlike = 0;
  for (i=0;i<Ngen; i++)
  {
    loglike  = -1.0*TMath::Log(func(gdata[i],par));
    sumlike += loglike;
   }
  f = sumlike;
}
```



Note that there is no binning in the calculation

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Likelihood fit with ROOT

```
void lgauss()
{
  gROOT->SetStyle("Plain");
  gROOT->ForceStyle();
  Float_t z1,z2;

  gRandom->SetSeed();
  Int_t dummy;
  hist = new TH1F("hist","Gauss distribution",Nbin,Xmin,Xmax);
  for (Int_t i=0;i<Ngen;i++)
  {
    Float_t u1,u2;
    u1 = gRandom->Rndm(dummy);
    u2 = gRandom->Rndm(dummy);
    z1 = TMath::Sin(2.0*3.1415926*u1)*TMath::Sqrt(-2.0*TMath::Log(u2));
    z2 = TMath::Cos(2.0*3.1415926*u1)*TMath::Sqrt(-2.0*TMath::Log(u2));
    z1 = Mu0 + z1*Sig0;
    gdata[i] = z1;
    hist->Fill(z1);
  }

  TMinuit *gMinuit = new TMinuit(2);    gMinuit->SetFCN(fcn);

  Double_t arglist[10];
  Int_t ierflg = 0;

  arglist[0] = 0.5;
  gMinuit->mnexcm("SET ERR", arglist ,1,ierflg);
  static Double_t vstart[2] = {0.0,  1.0};
  static Double_t step[2]   = {0.1 , 0.1};
  gMinuit->mnparm(0, "mean ", vstart[0], step[0], 0,0,ierflg);
  gMinuit->mnparm(1, "sigma", vstart[1], step[1], 0,0,ierflg);

  arglist[0] = 500;
  arglist[1] = 1.;
  gMinuit->mnexcm("MIGRAD", arglist ,2,ierflg);

  Double_t amin,edm,errdef;
  Int_t nvpar,nparx,icstat;
  gMinuit->mnstat(amin,edm,errdef,nvpar,nparx,icstat);

  Double_t mean, emean, sigma, esigma;
  gMinuit->GetParameter(0, mean, emean);
  gMinuit->GetParameter(1, sigma, esigma);

  printf(" mean  = %5.5f +- %5.5f \n",mean,emean);
  printf(" sigma = %5.5f +- %5.5f \n",sigma,esigma);

  hist->Draw("e");
  TF1 *fnc = new TF1("fnc",f1,-5.0,5.0,2);
  fnc->SetParameters(mean,sigma);
  fnc->Draw("same");
}
```

Why is this 0.5 (instead of 1.0 as in the chi-square fits?)

```
Processing lgauss.C...
 **********
 **    1 **SET ERR           1
 **********
 PARAMETER DEFINITIONS:
    NO.   NAME         VALUE      STEP SIZE      LIMITS
     1 mean         0.00000e+00  1.00000e-01     no limits
     2 sigma        1.00000e+00  1.00000e-01     no limits
 **********
 **    2 **MIGRAD        500           1
 **********
 FIRST CALL TO USER FUNCTION AT NEW START POINT, WITH IFLAG=4.
 START MIGRAD MINIMIZATION.  STRATEGY  1.  CONVERGENCE WHEN EDM .LT. 1.00e-03
 FCN=7459.22 FROM MIGRAD    STATUS=INITIATE        8 CALLS           9 TOTAL
                     EDM= unknown      STRATEGY= 1      NO ERROR MATRIX
  EXT PARAMETER               CURRENT GUESS       STEP         FIRST
  NO.   NAME      VALUE            ERROR          SIZE      DERIVATIVE
   1  mean        0.00000e+00   1.00000e-01   1.00000e-01  -3.00051e+03
   2  sigma       1.00000e+00   1.00000e-01   1.00000e-01  -1.20806e+04
 MIGRAD MINIMIZATION HAS CONVERGED.
 MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
 COVARIANCE MATRIX CALCULATED SUCCESSFULLY
 FCN=2121.68 FROM MIGRAD    STATUS=CONVERGED      61 CALLS          62 TOTAL
                     EDM=6.48278e-08    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                                   STEP         FIRST
  NO.   NAME      VALUE            ERROR          SIZE      DERIVATIVE
   1  mean        3.00050e+00   9.03047e-02   2.03457e-03  -1.10811e-03
   2  sigma       2.01927e+00   6.38545e-02   1.44188e-03  -5.41579e-03
 EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  2    ERR DEF=1
  8.155e-03  4.086e-06
  4.086e-06  4.077e-03
 PARAMETER  CORRELATION COEFFICIENTS
       NO.  GLOBAL      1      2
        1  0.00071   1.000  0.001
        2  0.00071   0.001  1.000
 mean  = 3.00050 +- 0.09030
 sigma = 2.01927 +- 0.06385
```

Consistent with the input values within the uncertainty

```
const static Float_t Mu0  = 3.0;
const static Float_t Sig0 = 2.0;
```

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Variance of Maximum-Likelihood Estimators (graphical method)

- A Taylor expansion of the log-likelihood function (one parameter case)

$$\ln L(x|\theta) = \ln L(x|\hat{\theta}) + \underbrace{\Big[\frac{\partial \ln L(x|\theta)}{\partial\theta}\Big]_{\theta=\hat{\theta}}(\theta-\hat{\theta})}_{\text{This term becomes zero.}} + \frac{1}{2!}\Big[\frac{\partial^2 \ln L(x|\theta)}{\partial\theta^2}\Big]_{\theta=\hat{\theta}}(\theta-\hat{\theta})^2 + \ldots$$

In general, we have (without proof)

$$(\hat{V^{-1}})_{ij} = -\frac{\partial^2 \ln L}{\partial\theta_i\partial\theta_j}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

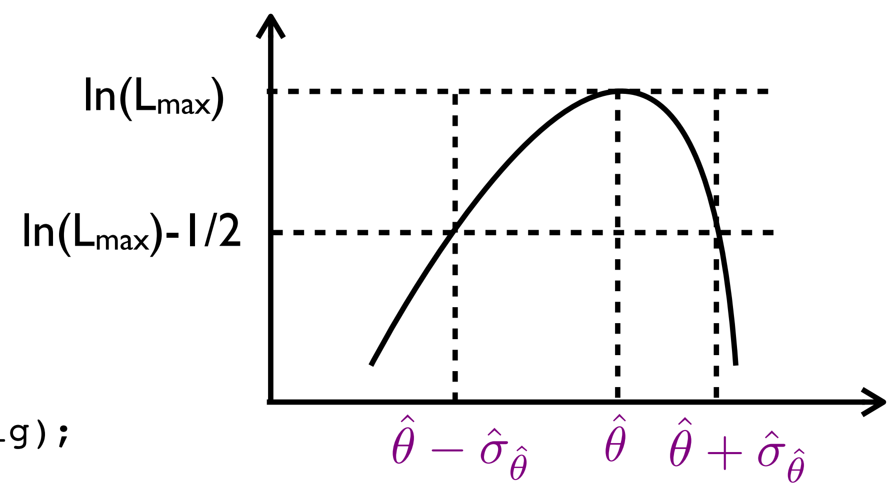
and for the one parameter case, we get

$$\hat{\sigma}^2_{\hat{\theta}} = \Big(-1\Big/\frac{\partial^2 \ln L}{\partial\theta^2}\Big)\Big|_{\theta=\hat{\theta}}$$

$$\ln L(x|\theta) = \ln L_{\max} - \frac{(\theta-\hat{\theta})^2}{2\hat{\sigma}^2_{\hat{\theta}}}$$ *(neglecting higher order terms)*

or

$$\ln L(x|\hat{\theta}\pm\hat{\sigma}_{\hat{\theta}}) = \ln L_{\max} - \frac{1}{2}$$



This is why we had:

```
arglist[0] = 0.5;
gMinuit->mnexcm("SET ERR", arglist,1,ierflg);
```


Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

What about the chi-square ?

- A Taylor expansion of the chi-square function (1st order term vanishes at minimum)

$$\chi^2(\boldsymbol{\theta}) = \chi^2(\hat{\boldsymbol{\theta}}) + \frac{1}{2}\sum_{i,j=1}^{m}\left[\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}(\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)$$

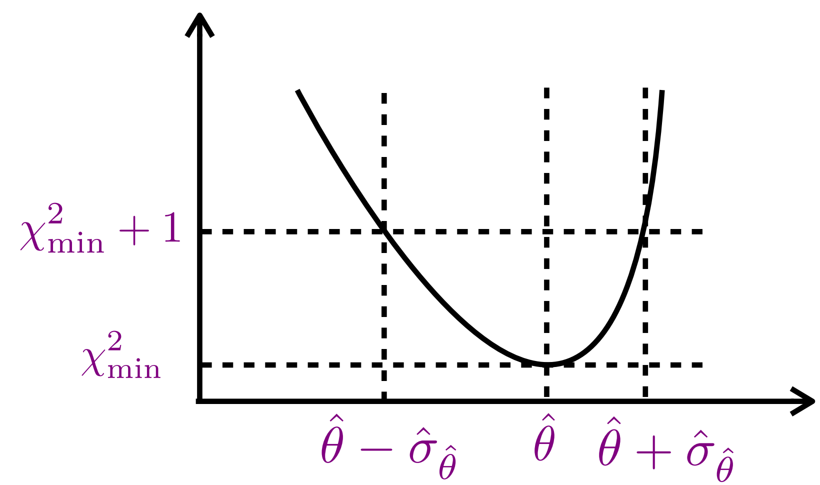
and from previous discussions,

$$U = (A^T V^{-1} A)^{-1}$$

$$(U^{-1})_{ij} = \frac{1}{2}\left[\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

we have

$$\chi^2(\boldsymbol{\theta}) = \chi^2(\hat{\boldsymbol{\theta}}) + 1 = \chi^2_{\min} + 1$$



This is why we had (for chi-square fits):

```
arglist[0] = 0.5;
gMinuit->mnexcm("SET ERR", arglist,1,ierflg);
```


Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Errors can be asymmetric?

- In all the examples so far we showed, errors were symmetric. For example,

```
 EXT PARAMETER                                   STEP         FIRST
 NO.   NAME       VALUE            ERROR          SIZE      DERIVATIVE
  1  mean          3.00050e+00   9.03047e-02   2.03457e-03  -1.10811e-03
  2  sigma         2.01927e+00   6.38545e-02   1.44188e-03  -5.41579e-03
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  2     ERR DEF=1
 8.155e-03  4.086e-06
 4.086e-06  4.077e-03
PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL      1      2
       1  0.00071   1.000  0.001
       2  0.00071   0.001  1.000
```

but if you have the following extra line in your fitting program:

```
gMinuit->mnexcm("MIGRAD", arglist ,2,ierflg);
gMinuit->mnexcm("MINOS", arglist ,0,ierflg);
```

```
**********
 **    3 **MINOS
 **********
 FCN=2121.68 FROM MINOS     STATUS=SUCCESSFUL     36 CALLS          98 TOTAL
                     EDM=6.48278e-08    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                  PARABOLIC         MINOS ERRORS
  NO.   NAME      VALUE            ERROR      NEGATIVE      POSITIVE
   1  mean         3.00050e+00   9.03047e-02  -9.03457e-02   9.03548e-02
   2  sigma        2.01927e+00   6.38545e-02  -6.21992e-02   6.55896e-02
```


Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

What are `MIGRAD` and `MINOS` anyway?

- These are two different numerical minimization methods implemented in MINUIT package

  - `MIGRAD`

    Davidon-Flettcher-Powell variable-metric algorithm : R. Fletcher and M.J.D. Powell. A rapidly converging descent method for minimization. *Comput. J.*, 6:163, 1963.

    For us, we use this command all the time. It will return symmetric error estimation.

  - `MINOS`

    F. James, and M. Roos, A system for function minimization and analysis of the parameter errors and correlations, *Comput. Phys. Commut*, **10**, 343 (1975).

    For us, we use this command when we need asymmetric error (more precise estimation) and serious fitting purpose

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

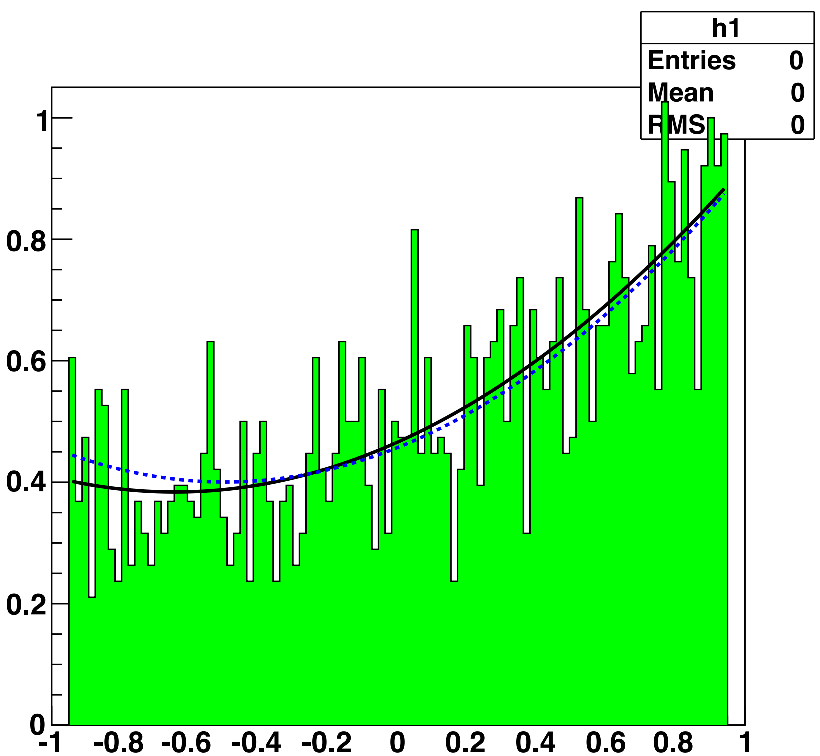
Example of maximum likelihood fit

- A p.d.f. is given as $f(x;\alpha,\beta) = \dfrac{1+\alpha x+\beta x^2}{2+2\beta/3}$ for $-1 \le x \le 1$

If one restricts the range to be

$$f(x;\alpha,\beta) = \frac{1+\alpha x+\beta x^2}{(x_{\max}-x_{\min}+\frac{\alpha}{2}(x_{\max}^2-x_{\min}^2)+\frac{\beta}{3}(x_{\max}^3-x_{\min}^3))}$$

Now, let's generate 2000 random numbers with the above p.d.f. and perform a maximum likelihood fit (unbinned) to it (x_{max} = 0.95, x_{min} = -0.95)



Input : α= 0.5, β= 0.5

Histogram: a binned data of 2000 random numbers generated

Back curve: a maximum likelihood fit result

Dashed blue curve: the original p.d.f. with α= 0.5, β= 0.5

EXT NO.	PARAMETER NAME	VALUE	ERROR	STEP SIZE	FIRST DERIVATIVE
1	alpha	5.49470e-01	5.15671e-02	1.08855e-03	-1.34855e-02
2	beta	4.26312e-01	1.04027e-01	2.19395e-03	6.29645e-03

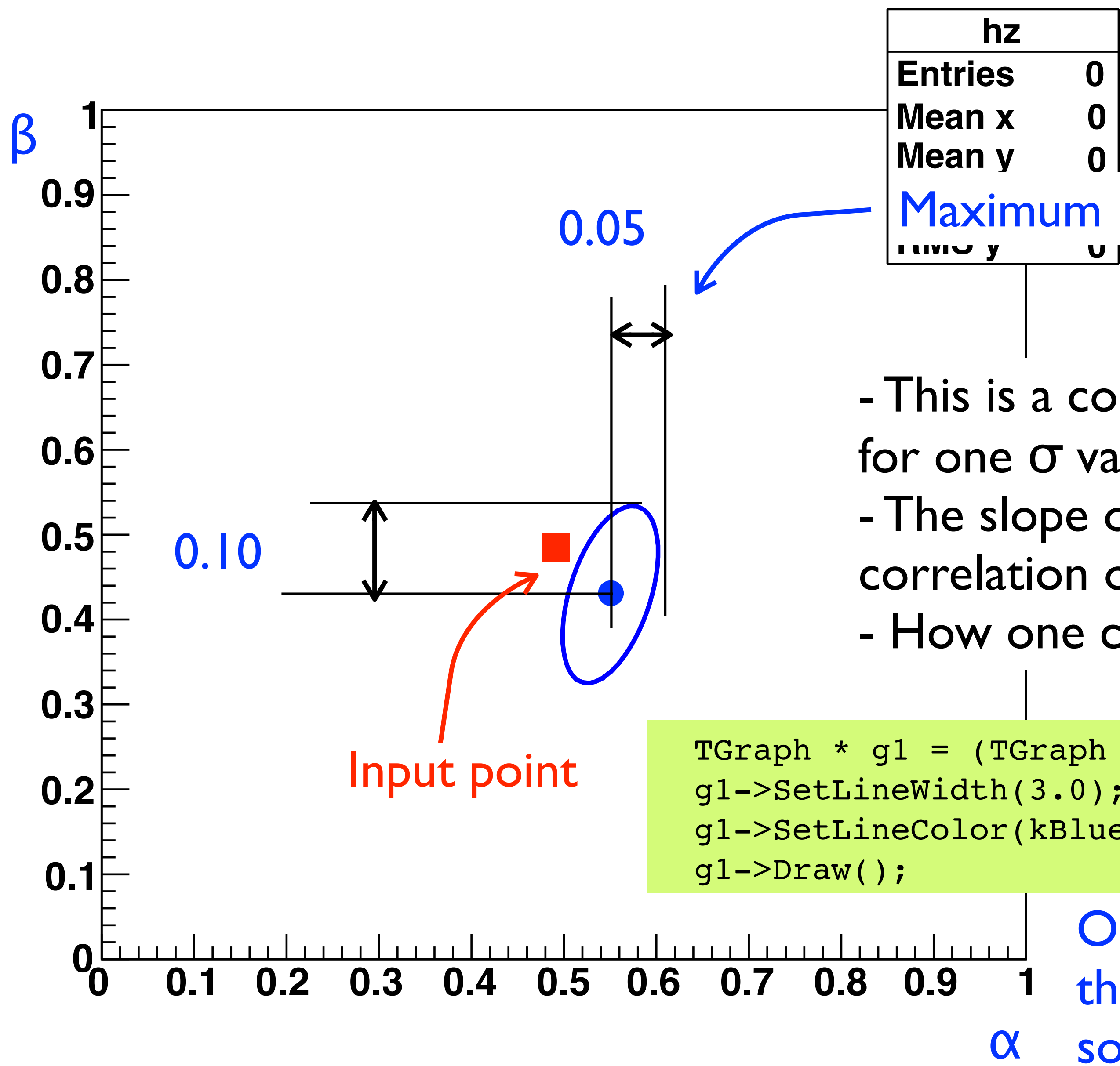
ERR DEF= 0.5

Fit result : α= 0.55±0.05, β= 0.43±0.10

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Example of maximum likelihood fit

Fit result : α= 0.55±0.05, β= 0.43±0.10



- This is a contour curve of fit parameters, for one σ values from the fit center
- The slope of the ellipse represent the correlation of two parameters
- How one can make this in ROOT?

```
TGraph * g1 = (TGraph *) gMinuit->Contour(100,0,1);
g1->SetLineWidth(3.0);
g1->SetLineColor(kBlue);
g1->Draw();
```

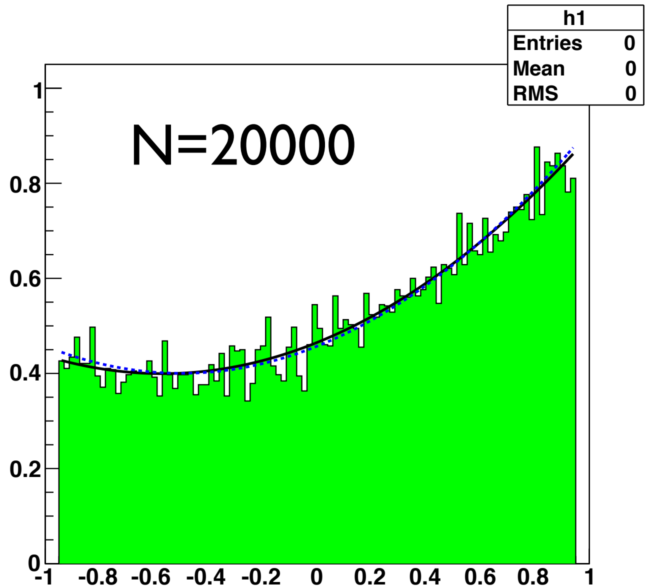
Our fit is about one σ away from the input point: we were somewhat unlucky...

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Example of maximum likelihood fit

- Is the fit result α= 0.55±0.05, β= 0.43±0.10 not good?

Let's try fits with more random numbers (N=20000)



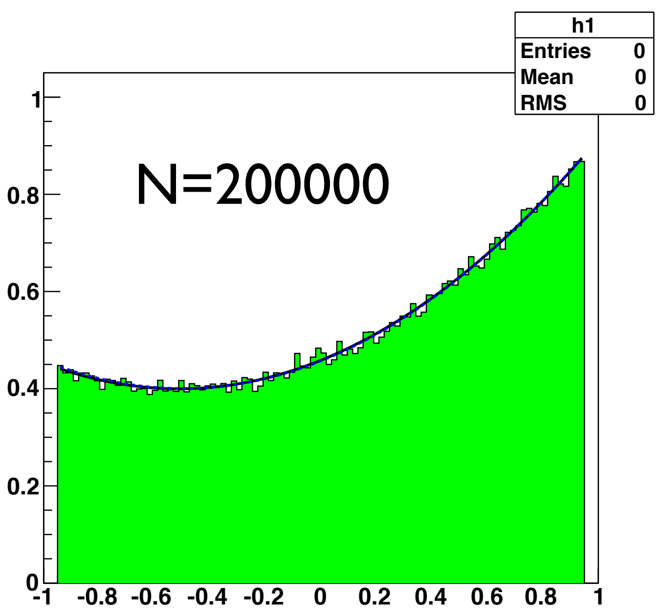
EXT NO.	PARAMETER NAME	VALUE	ERROR	STEP SIZE	FIRST DERIVATIVE
1	alpha	4.96123e-01	1.61868e-02	1.10832e-03	-1.56625e-01
2	beta	4.35307e-01	3.37413e-02	2.30833e-03	6.62779e-02

ERR DEF= 0.5

Fit result : α= 0.50±0.02, β= 0.44±0.03

OK, now it is much better. Two curves are almost identical to each other (can you tell?)

If I do the fit with N=200000 (It took 23 seconds with my laptop)



EXT NO.	PARAMETER NAME	VALUE	ERROR	STEP SIZE	FIRST DERIVATIVE
1	alpha	4.98558e-01	5.24949e-03	1.14557e-03	-1.49391e+00
2	beta	5.25891e-01	1.12444e-02	2.45057e-03	2.69860e-01

ERR DEF= 0.5

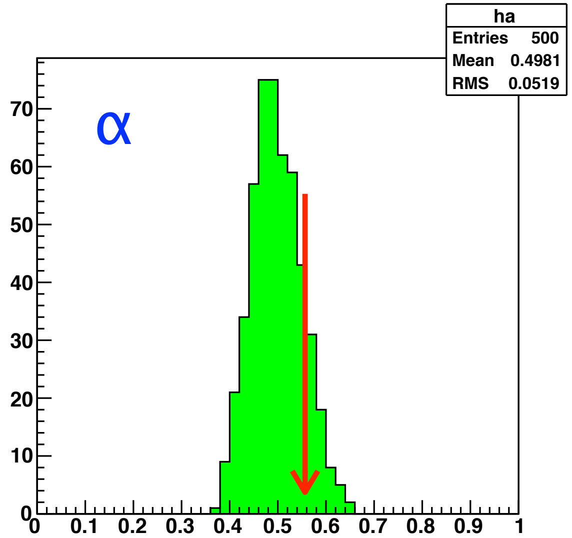
Fit result : α= 0.50±0.005, β= 0.53±0.01

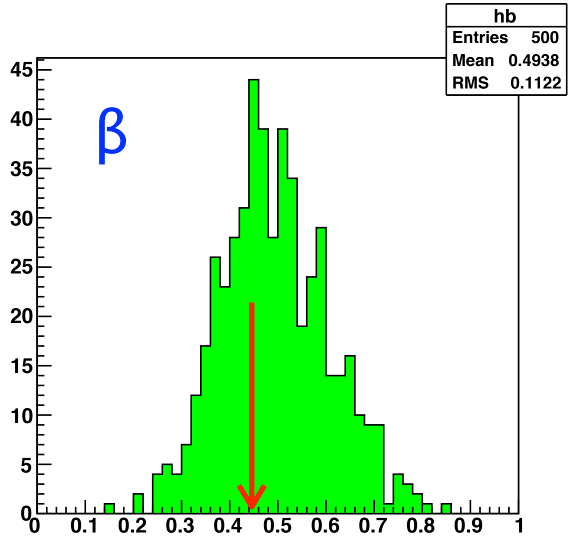
Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Example of maximum likelihood fit

- Back to the question: is the fit result α= 0.55±0.05, β= 0.43±0.10 not good?

  Let's repeat the likelihood fits many times with newly generated set of random numbers each time: this is called pseudo experiment (or ensemble test).





500 independent likelihood fits

- Distributions of α and β are centered at input values (α=0.5, β=0.5)

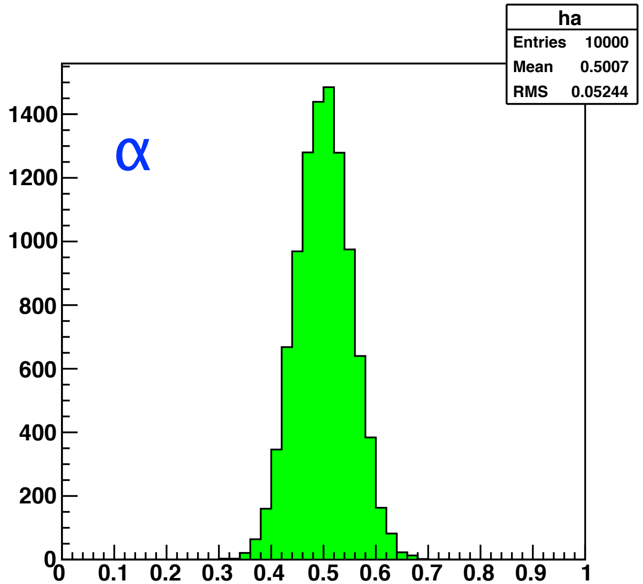
- Our first fit result α= 0.55±0.05, β= 0.43±0.10 is shown as arrows

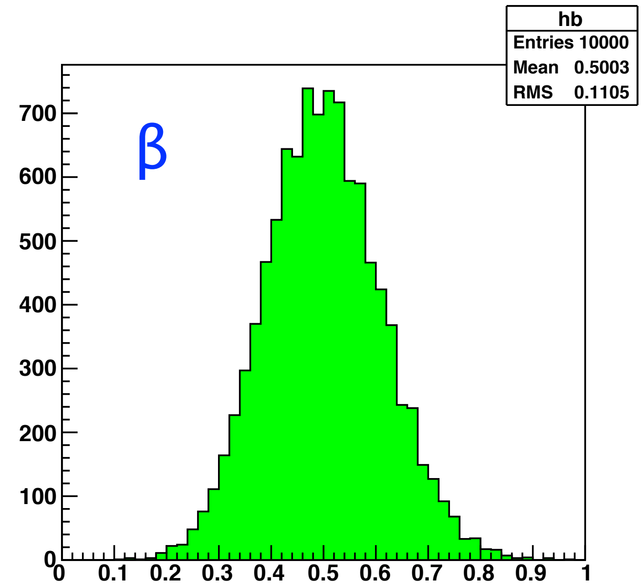
: so, we were somewhat "unlucky" at the first fit, but the distribution from many pseudo experiment looks right.

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

Example of maximum likelihood fit

- You think distributions of α and β are not smooth? Let's try 10000 ensemble tests





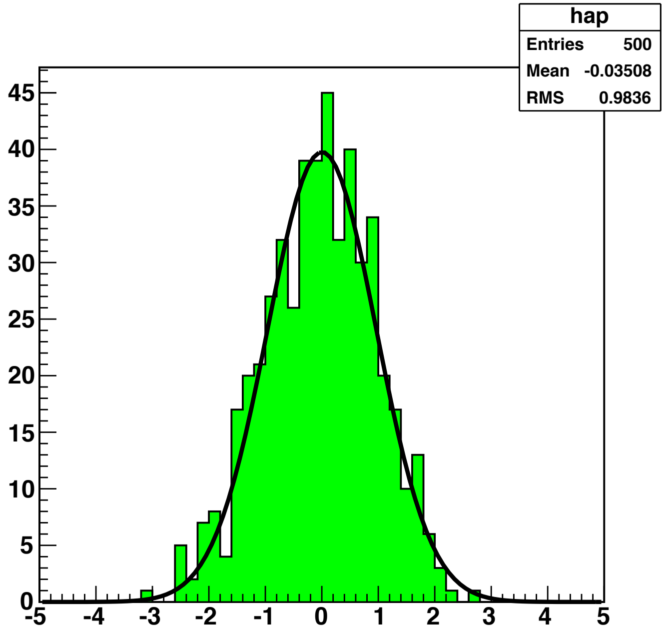
10000 independent likelihood fits
(It took 1.5 hours with my laptop)

If everything is right (errors are assigned right and fits are all done properly) two distributions should look gaussian. We will not prove it (too advanced)

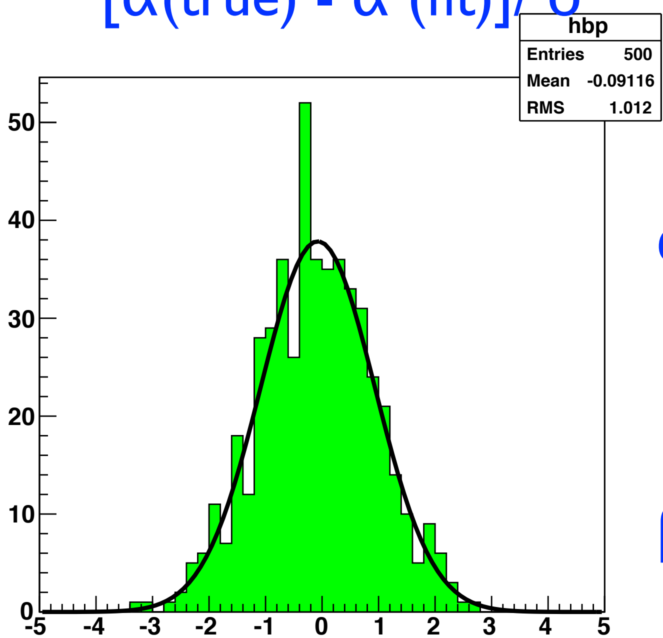
It looks gaussian to me anyway

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)

The pull distribution



[α(true) - α (fit)]/ σ



[β(true) - β(fit)]/ σ

- One can validate the fit procedure by checking the pull distribution defined as

$$\mathrm{pull} = \frac{\theta_{\mathrm{input}} - \theta_{\mathrm{fit}}}{\sigma_{\mathrm{fit}}}$$

If everything is right (errors are assigned right and fits are all done properly) pull distributions must be gaussian distributions with mean=0, sigma=1.

α

NO.	NAME	VALUE	ERROR	SIZE	DERIVATIVE
1	Constant	3.97693e+01	2.28396e+00	4.47228e-03	-9.53974e-07
2	Mean	1.88098e-03	4.59380e-02	1.11360e-04	1.86270e-04
3	Sigma	9.61759e-01	3.52218e-02	2.34404e-05	6.96852e-05

pull of α : mean = 0.00±0.05 sigma = 0.96±0.04

β

NO.	NAME	VALUE	ERROR	SIZE	DERIVATIVE
1	Constant	3.78635e+01	2.21027e+00	4.13630e-03	2.09199e-05
2	Mean	-7.60167e-02	4.71340e-02	1.12357e-04	-8.53878e-04
3	Sigma	1.00969e+00	3.79280e-02	2.33290e-05	3.29532e-03

pull of β : mean = 0.08±0.05 sigma = 1.01±0.04

Korea U/Dept. Physics, Prof. Eunil Won (All rights are reserved)