



中国科学技术大学

University of Science and Technology of China

RooFit

王咪

(参考秦龙宇师兄2022年9月PPT)

2023年7月14日



- 拟合初介绍
- 一些例子
- 总结



- 什么是拟合
 - ✓ 我们的数据一般是一些离散的点，我们用比如极大似然值法，得到一个最好的参数化公式去描述数据。
- 为什么拟合
 - ✓ 我们得到的参数化公式往往可能蕴含某种物理规律。比如粒子物理最常见的去拟合不变质量谱得到一个BW函数，就是研究一个我们感兴趣的共振态的最直接有效的方法。



- RooRealVar
- RooAbsPdf
- RooDataSet
- RooFFTConvPDF
- RooAddPdf
- 变量; 参数
- PDF
- 建立在某一变量下的数据的集合
- 卷积
- PDF相加



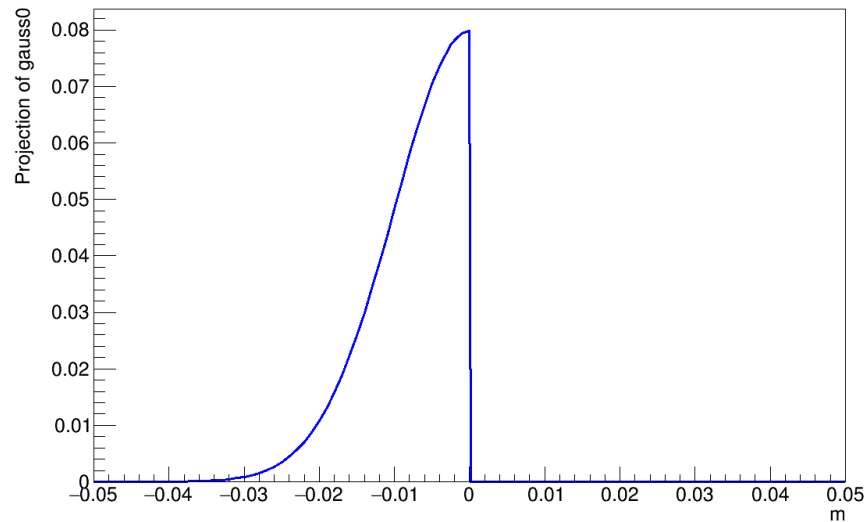
- RooGaussian
- RooExponential
- RooPolynomial
- RooChebychev
- RooBreigWigner
- RooArgusBG
- RooBiFurGauss
- RooCrystalBall
- Gaussian
- Exponential
- Polynomial
- Chebychev polynomial
- Breit-wigner
- Argus
- Bifurcated Gaussian
- Crystal Ball

这些都算作前面提到的RooAbsPdf！！

常用的PDF类型



中国科学技术大学
University of Science and Technology of China

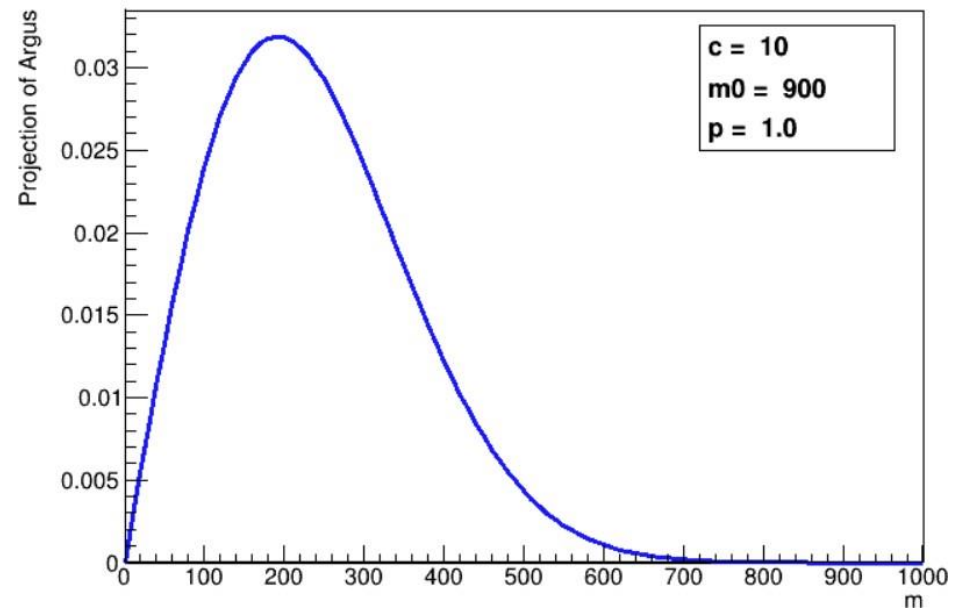


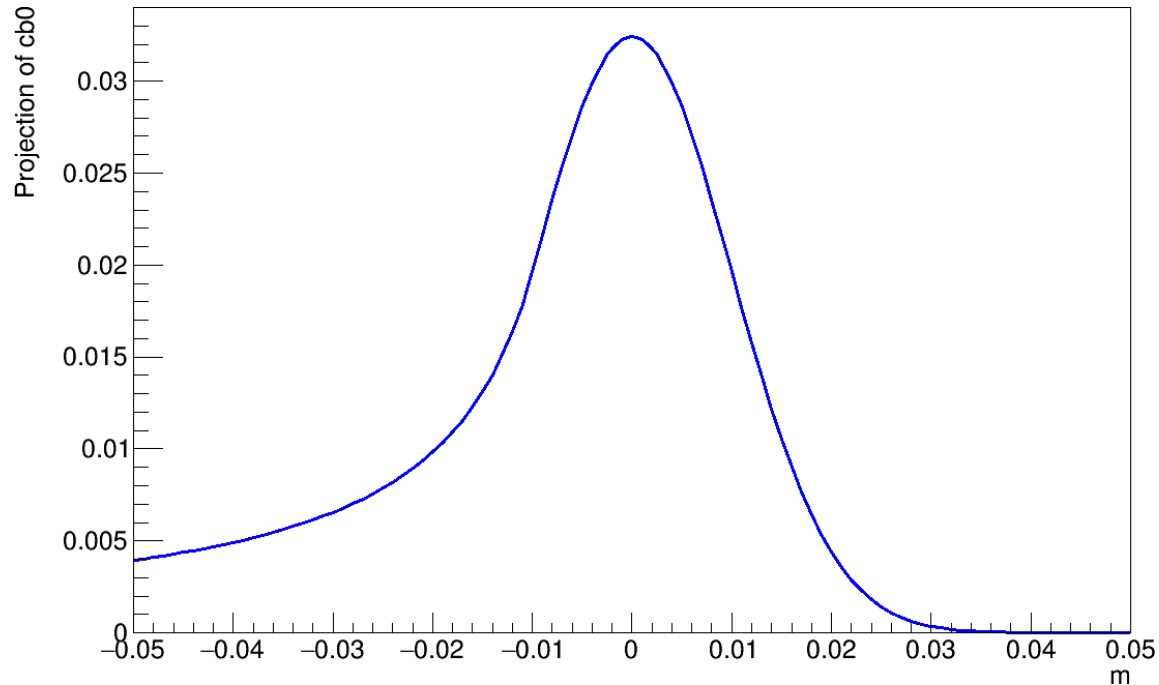
Bifurcated Gaussian

Argus

$$\text{Argus}(m, m_0, c, p) = \mathcal{N} \cdot m \cdot \left[1 - \left(\frac{m}{m_0} \right)^2 \right]^p \cdot \exp \left[c \cdot \left(1 - \left(\frac{m}{m_0} \right)^2 \right) \right]$$

A RooPlot of "m"





Crystal Ball

$$f(m; m_0, \sigma, \alpha_L, n_L, \alpha_R, n_R) = \begin{cases} A_L \cdot (B_L - \frac{m-m_0}{\sigma_L})^{-n_L}, & \text{for } \frac{m-m_0}{\sigma_L} < -\alpha_L \\ \exp\left(-\frac{1}{2} \cdot \left[\frac{m-m_0}{\sigma_L}\right]^2\right), & \text{for } \frac{m-m_0}{\sigma_L} \leq 0 \\ \exp\left(-\frac{1}{2} \cdot \left[\frac{m-m_0}{\sigma_R}\right]^2\right), & \text{for } \frac{m-m_0}{\sigma_R} \leq \alpha_R \\ A_R \cdot (B_R + \frac{m-m_0}{\sigma_R})^{-n_R}, & \text{otherwise,} \end{cases}$$

$$A_i = \left(\frac{n_i}{|\alpha_i|}\right)^{n_i} \cdot \exp\left(-\frac{|\alpha_i|^2}{2}\right)$$

$$B_i = \frac{n_i}{|\alpha_i|} - |\alpha_i|$$



- 最终我们要进行的操作是：把一个用参数(RooRealVar)描述的PDF(RooAbsPdf)(比如一个用mean值和sigma值描述的gaussian分布)，去拟合到一个数据集(RooDataSet)上面，让程序去根据自己的算法，找到这些参数最好的值，来让这个PDF最精确地表述这个数据集。

复制一下先！



中国科学技术大学
University of Science and Technology of China

- 先进到自己的主目录下随便创个文件夹，进去之后cp:
- `cp -r /besfs5/groups/jpsi/jpsigroup/user/qinlongyu/shareble/roofit ./`
- `cd roofit`
- `source ROOT6_24.csh`
 - ✓ 对于某些复杂的拟合，root版本的差别对拟合结果会有挺大的影响(建议用root6)
- `vi gauss1.cxx`

一个简单的例子



- vi gauss1.cxx

```
14 void gauss1()
15 {
16     //define my variable
17     RooRealVar x("x","x",-10,10) ;
18
19     //define a gaussian distribution
20     RooRealVar mean("mean","mean of gaussian",1,-10,10) ;
21     RooRealVar sigma("sigma","width of gaussian",1,0.1,10) ;
22     RooGaussian gauss("gauss","gaussian PDF",x,mean,sigma) ;
23
24     // Generate a dataset of 1000 events in x from gauss
25     RooDataSet* data = gauss.generate(x,10000) ;
26
27     //define another gaussian distribution
28     RooRealVar mean1("mean1","mean1 of gaussian",2,-10,10) ;
29     RooRealVar sigma1("sigma1","width1 of gaussian",2,0.1,10) ;
30     RooGaussian gauss1("gauss1","gaussian PDF1",x,mean1,sigma1) ;
31
32
33
34
35
36     // Draw all frames on a canvas
37     RooPlot* xframe = x.frame() ;
38     RooPlot* xframe2 = x.frame() ;
39     TCanvas* c = new TCanvas("c","c",800,400) ;
40
41     c->Divide(2) ;
42     c->cd(1) ;
43     data->plotOn(xframe) ;
44     gauss1.plotOn(xframe) ;
45     xframe->Draw() ;
46
47     c->cd(2) ;
48     //this line means fit pdf to dataset
49     gauss1.fitTo(*data) ;
50     data->plotOn(xframe2) ;
51     gauss1.plotOn(xframe2) ;
52     xframe2->Draw() ;
53
54 }
```

先定义一个高斯pdf (gauss)

用这个高斯pdf产生了一个dataset

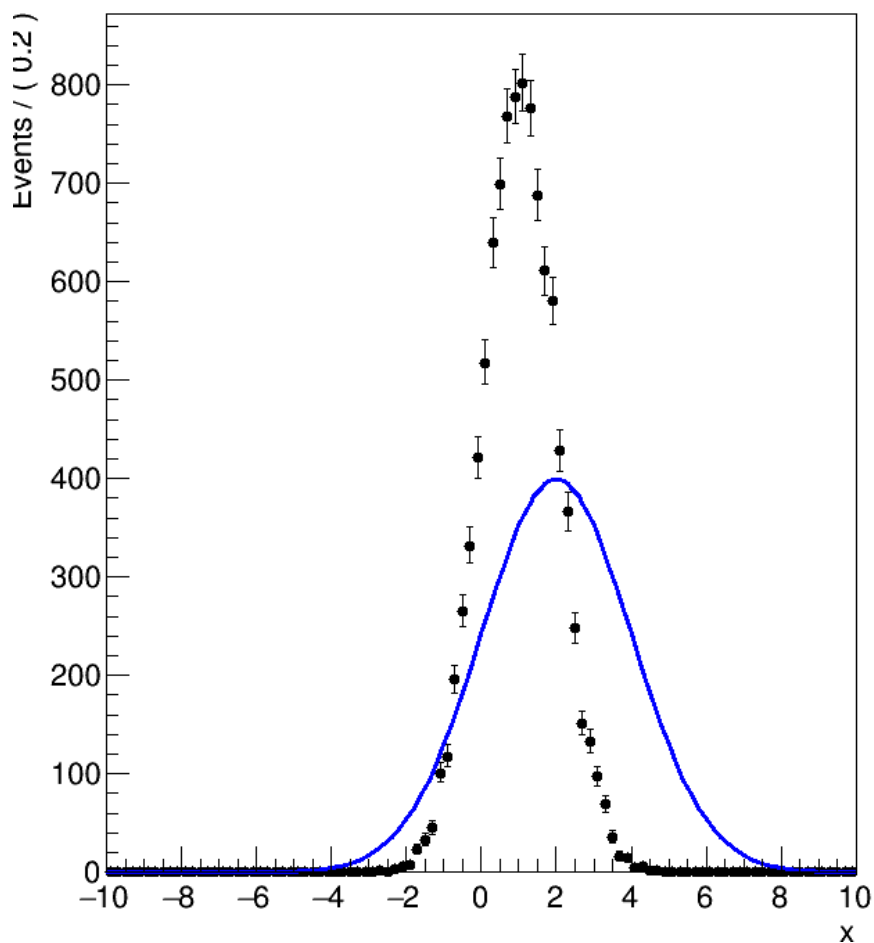
定义参数不一样的高斯pdf (gauss1)

定义两个plot (类似于TH1F) 和一个画布

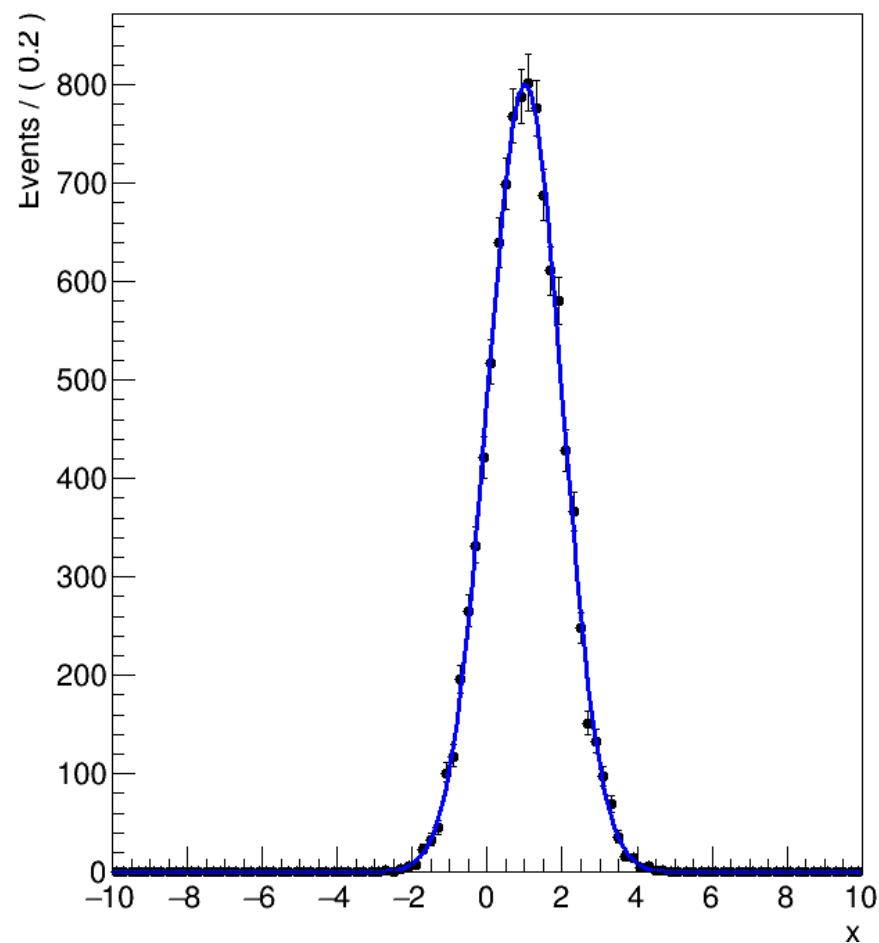
在左边的画布上画出前面产生的dataset, 以及初始条件下的guass1

在右边的画布上画出前面产生的dataset, 以及拟合后的guass1

A RooPlot of "x"



A RooPlot of "x"



Estimate Distance to Minimum, 越小越好
OK表示拟合成功

卡方最小值

拟合结果

相关系数

```
*****
**  18 **HESSE      1000
*****
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=14162.8 FROM HESSE STATUS=OK
EDM=3.21094e-07 STRATEGY= 1 ERROR MATRIX ACCURATE
10 CALLS      66 TOTAL
EXT PARAMETER      INTERNAL      INTERNAL
NO.  NAME      VALUE      ERROR  STEP SIZE      VALUE
1  mean1      1.00665e+00  9.97346e-03  1.64750e-05  1.00836e-01
2  sigma1      9.97346e-01  7.05227e-03  8.15530e-06 -9.59175e-01
ERR DEF= 0.5
EXTERNAL ERROR MATRIX.  NDIM= 25  NPAR= 2  ERR DEF=0.5
9.947e-05  8.427e-09
8.427e-09  4.973e-05
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL      1      2
1  0.00012  1.000  0.000
2  0.00012  0.000  1.000
```



- 在实际工作当中，我们的dataset一般是会从root文件里面引入
- 而且dataset的构成会比较复杂，一般不会是一个高斯分布就能很轻易描述
- 需要各种pdf按不同的权重求和，构成新的pdf以描述更复杂的dataset

复杂一些的例子1.0



- vi omegabw.cxx

```
1 using namespace RooFit;
2 void omegabwfit()
3 {
4
5     //here import my data
6     TFile *f_data = new TFile("sample.root");
7     TTree *t_data = (TTree*)f_data->Get("GammaOmegaPhi");
8     ///////////////////////////////////////////////////
9     RooRealVar momega("momega","M_{#pi^{+}#pi^{-}}#pi^{0}} (GeV/c^{2})",0.582,0.982);
10    RooDataSet data("data","data",t_data,momega);
11    ///////////////////////////////////////////////////bw conv gauss/////////////////////////////////
12    RooRealVar mean0("mean0","mean0",0,-0.1,0.1);
13    RooRealVar sigma0("sigma0","sigma0",0.1,0,0.1);
14    RooGaussian gauss0("gauss0","gauss0",momega,mean0,sigma0);
15    RooRealVar m0("m0","m0",0.782,0.582,0.982);
16    m0.setConstant();
17    RooRealVar g0("g0","g0",0.00868,0.0,0.020);
18    g0.setConstant();
19    RooBreitWigner bw0("bw0","bw0",momega,m0,g0);
20    RooFFTConvPdf conv1("conv1","conv1",momega,bw0,gauss0);
21    ///////////////////////////////////////////////////poly/////////////////////////////////
22    RooRealVar p0("p0","",1.3,-2,2);
23    RooRealVar p1("p1","",0.6,-2,2);
24    RooRealVar p2("p2","",0.2,-2,2);
25    RooChebychev bkgo0("bkgo0","bkgo0",momega,RooArgList(p0,p1,p2));
26    ///////////////////////////////////////////////////
27    RooRealVar nsig("nsig","nsig",19000,0,50000);
28    RooRealVar nbkg("nbkg","nbkg",13000,0,50000);
29
30    RooAddPdf summomega("sumomega","sumomega",RooArgList(conv1,bkgo0),RooArgList(nsig,nbkg));
31    summomega.fitTo(data);
32
33
34
35    RooPlot* momegaframe = momega.frame();
36    TCanvas *c1 = new TCanvas("c1","a canvas",0,0,800,600);
37    ///////////////////////////////////////////////////
38    c1->cd();
39    data.plotOn(momegaframe);
40    summomega.plotOn(momegaframe,LineColor(4));
41    summomega.plotOn(momegaframe,Components(conv1),LineStyle(4),LineColor(2),LineWidth(3));
42    summomega.plotOn(momegaframe,Components(bkgo0),LineStyle(2),LineColor(7),LineWidth(3));
43    momegaframe->Draw();
44 }
```

读取ROOT文件

定义gauss0, bw0

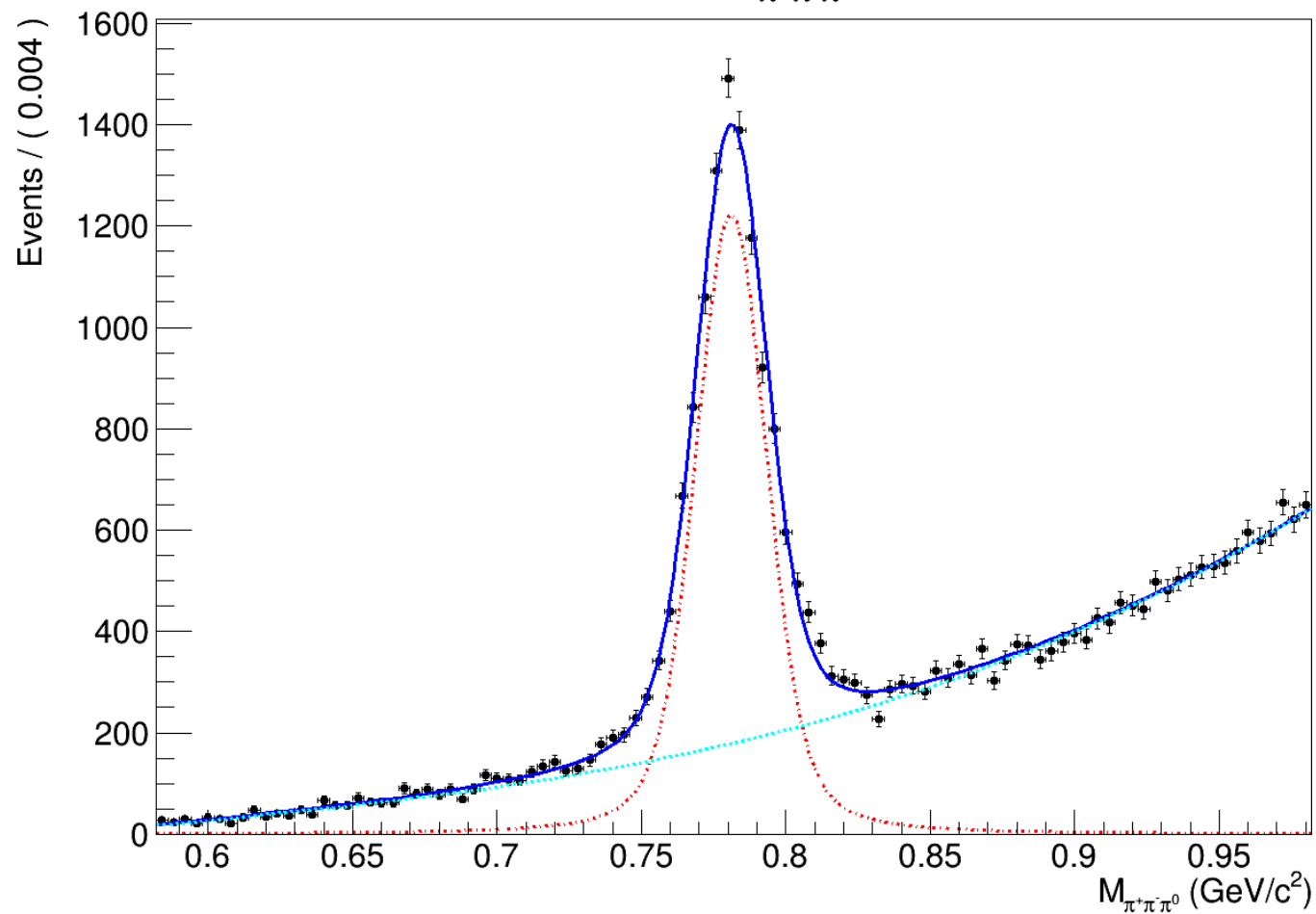
卷积gauss0、bw0为conv1
(描述信号)

定义切比雪夫多项式
(描述本底)

将 conv1 和 bkgo0 相加
RooArgList为信号和本底
的事例数，可以通过拟合
得到具体的事例数

画出拟合结果

A RooPlot of " $M_{\pi^+\pi^-\pi^0}$ (GeV/c²)"





```

*****
**      9 **HESSE          3500
*****
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=-360148 FROM HESSE      STATUS=OK          62 CALLS          314 TOTAL
              EDM=2.59207e-05  STRATEGY= 1      ERROR MATRIX ACCURATE

EXT PARAMETER          INTERNAL          INTERNAL
NO.   NAME      VALUE      ERROR      STEP SIZE      VALUE
1  mean0      -6.92184e-04  1.78943e-04  7.32605e-04  -6.92190e-03
2  nbkg       2.30344e+04  1.78863e+02  2.69154e-03  -7.87064e-02
3  nsig       1.05995e+04  1.39871e+02  2.43186e-03  -6.13852e-01
4  p0         1.17803e+00  7.98795e-03  1.47813e-03  6.29839e-01
5  p1         2.94004e-01  1.16601e-02  1.41953e-03  1.47537e-01
6  p2         4.54275e-02  8.42510e-03  1.23438e-03  2.27157e-02
7  sigma0     1.03433e-02  2.07064e-04  2.55629e-03  -2.22566e+00
              ERR DEF= 0.5
EXTERNAL ERROR MATRIX.   NDIM= 25  NPAR= 7  ERR DEF=0.5
3.202e-08 -1.886e-03  1.888e-03  3.484e-08  1.246e-07  1.607e-07 -4.077e-09
-1.886e-03  3.199e+04 -8.953e+03 -3.401e-01 -6.805e-01 -3.206e-01  1.003e-02
1.888e-03 -8.953e+03  1.956e+04  3.401e-01  6.805e-01  3.206e-01 -1.004e-02
3.484e-08 -3.401e-01  3.401e-01  6.381e-05  6.390e-05  3.638e-05 -3.736e-07
1.246e-07 -6.805e-01  6.805e-01  6.390e-05  1.360e-04  6.861e-05 -7.696e-07
1.607e-07 -3.206e-01  3.206e-01  3.638e-05  6.861e-05  7.098e-05 -3.710e-07
-4.077e-09  1.003e-02 -1.004e-02 -3.736e-07 -7.696e-07 -3.710e-07  4.288e-08
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL      1      2      3      4      5      6      7
1  0.15548  1.000 -0.059  0.075  0.024  0.060  0.107 -0.110
2  0.42533 -0.059  1.000 -0.358 -0.238 -0.326 -0.213  0.271
3  0.51472  0.075 -0.358  1.000  0.304  0.417  0.272 -0.347
4  0.69237  0.024 -0.238  0.304  1.000  0.686  0.541 -0.226
5  0.81386  0.060 -0.326  0.417  0.686  1.000  0.698 -0.319
6  0.70743  0.107 -0.213  0.272  0.541  0.698  1.000 -0.213
7  0.42079 -0.110  0.271 -0.347 -0.226 -0.319 -0.213  1.000
    
```




- 在实际工作当中，有时候某些pdf是没法用公式去描述的，找不到或者不存在公式去描述(或者说我就是不想用公式)，而我们需要这些PDF，我们就会想办法在物理上产生相近的PDF。比如我们可以从信号MC，收集一些事例构成dataset，然后由dataset去产生这个PDF。这种方法可以由任意的dataset来生成对应的PDF。

复杂例子1.1



- vi omegasigmcf1.cxx

```
1 using namespace RooFit;
2 void omegasigmcf1()
3 {
4
5 //here import my data
6 TFile *f_data = new TFile("sample.root");
7 TTree *t_data = (TTree*)f_data->Get("GammaOmegaPhi");
8 ///////////////////////////////////////////////////
9 RooRealVar momega("momega","M_{#pi^{+}#pi^{-}}#pi^{0}} (GeV/c^{2})",0.582,0.982);
10 RooDataSet data("data","data",t_data,momega);
11 ///////////////////////////////////////////////////sigpdf////////////////////////////////////
12 TFile *f_mc = new TFile("myshape.root");
13 TTree *t_mc = (TTree*)f_mc->Get("GammaOmegaPhi");
14 RooDataSet signalmc("signalmc","signalmc",t_mc,momega);
15 RooKeysPdf somegapdf("sigPdf","sigpdf",momega,signalmc,RooKeysPdf::MirrorBoth,1);
16
17
18
19 RooRealVar mean0("mean0","mean0",0,-0.1,0.1);
20 RooRealVar sigma0("sigma0","sigma0",0.1,0,0.1);
21 RooGaussian gauss0("gauss0","gauss0",momega,mean0,sigma0);
22 RooFFTConvPdf conv1("conv1","conv1",momega,somegapdf,gauss0);
23 //////////////////////////////////////////////////poly////////////////////////////////////
24 RooRealVar p0("p0","",1.3,-2,2);
25 RooRealVar p1("p1","",0.6,-2,2);
26 RooRealVar p2("p2","",0.2,-2,2);
27 RooChebychev bkgo0("bkgo0","bkgo0",momega,RooArgList(p0,p1,p2));
28 ///////////////////////////////////////////////////
29 RooRealVar nsig("nsig","nsig",19000,0,50000);
30 RooRealVar nbkg("nbkg","nbkg",13000,0,50000);
31
32 RooAddPdf summomega("sumomega","sumomega",RooArgList(conv1,bkgo0),RooArgList(nsig,nbkg));
33 summomega.fitTo(data);
34
35
36
37
38
39 RooPlot* momegaframe = momega.frame();
40
41
42 TCanvas *c1 = new TCanvas("c1","a canvas",0,0,1200,800);
43
```

这里和上面的例子的区别是，把BW分布，换成了由我们自己的dataset生成的一个PDF

复杂例子1.2



- vi omegasigmcfits.cxx

```
1 using namespace RooFit;
2 void omegasigmcfits()
3 {
4     //here import my data
5     TFile *f_data = new TFile("sample.root");
6     TTree *t_data = (TTree*)f_data->Get("GammaOmegaPhi");
7     RooRealVar momega("momega", "M_{#omega} (GeV/c^{2})", 0.582, 0.982);
8     RooDataSet data("data", "data", t_data, momega);
9     //sighistpdf
10    TH1F *moss = new TH1F("moss", "", 200, 0.582, 0.982); //momega signalmc shape
11    TFile *f_mc = new TFile("myshape.root");
12    TTree *t_mc = (TTree*)f_mc->Get("GammaOmegaPhi");
13    double smomega;
14    t_mc->SetBranchAddress("momega", &smomega);
15    int EntNum = t_mc->GetEntries();
16    for(int i = 0; i < EntNum; i++)
17    {
18        t_mc->GetEntry(i);
19        moss->Fill(smomega);
20    }
21    RooDataHist somegahist("somegahist", "somegahist", momega, moss);
22    RooHistPdf somegapdf("somegapdf", "somegapdf", momega, somegahist, 4);
23
24    RooRealVar mean0("mean0", "mean0", 0, -0.1, 0.1);
25    RooRealVar sigma0("sigma0", "sigma0", 0.1, 0, 0.1);
26    RooGaussian gauss0("gauss0", "gauss0", momega, mean0, sigma0);
27    RooFFTConvPdf conv1("conv1", "conv1", momega, somegapdf, gauss0);
28    //poly
29    RooRealVar p0("p0", "", 1.3, -2, 2);
30    RooRealVar p1("p1", "", 0.6, -2, 2);
31    RooRealVar p2("p2", "", 0.2, -2, 2);
32    RooChebychev bkgo0("bkgo0", "bkgo0", momega, RooArgList(p0, p1, p2));
33    //
34    RooRealVar nsig("nsig", "nsig", 19000, 0, 50000);
35    RooRealVar nbkg("nbkg", "nbkg", 13000, 0, 50000);
36
37    RooAddPdf sumomega("sumomega", "sumomega", RooArgList(conv1, bkgo0), RooArgList(nsig, nbkg));
38    sumomega.fitTo(data);
39
```

这是另一种生成PDF的方法，即先把数据填到一个hist里面，再由这个hist来生成一个PDF。该方法较上一页的方法相比稍微粗糙一些，但是快，在生成PDF的数据量很大的时候适用。



结果都一样的就不放了，大家可以自己跑着玩



- 拟合还有很多可以的操作，比如
 - ✓ 需要拟合的数据量巨大的时候，可以把数据分成bin拟合，速度会大大加快
 - ✓ 如果需要同时进行多个量拟合的时候，会有二维单位的拟合，即那种联合概率密度函数。
 - ✓ dataset很难拟合上的话，可以进行多次迭代拟合(拟合本身就是一种迭代！)，以得到更有的结果。
 - ✓ 还可以自己编写PDF函数，写成头文件或者写成包之类的，后面可以自己用。
 - ✓ blablabla



- 需要进行资料查阅的话，可以到ROOT的官网里面的reference里面去搜索，也可以直接在bing搜索某个函数。有问题解决不了的话，不要自己一直纠结，狠狠抓师兄们！（root作为工具我们的要求是熟练地用，而不需要自己去浪费太多时间）
- ROOT自带的tutorial就很好用，里面有很多东西,大家可以把整个tutorial目录cp到自己的目录下，随便运行进行尝试(里面有roofit的部分，也有各种别的部分):
 - ✓ `cp $ROOTSYS/tutorials ./`



本次入门教学到此圆满结束！！
感谢各位的参加！！