

粒子物理与核物理实验中的数据分析

王喆
清华大学

第十讲：最大似然拟合
以及**Minuit**

最大似然拟合与**Minuit**

- ◇ 练习最大似然法
- ◇ 使用极值求解工具**TMinuit**
- ◇ 再次尝试一个测试统计量的概念

例子在 **Lec9** 的课程附件中

例子分解:

```
void getData(vector<double>* xVecPtr)  {  
  
    double TauTruth = 0.5;  
    TRandom r;  
  
    for ( int i=1; i<=30; i++ )  {  
        double x = r.Exp( TauTruth );  
        xVecPtr->push_back(x);  
    }  
}
```

指定输入的数据，该段为利用随机数模拟实验测量结果，实际使用中需要用真实的数据来替代。

待极小化的函数fcn

```
void fcn(int& npar, double* deriv, double& f, double par[], int flag){  
  
    vector<double> xVec = *xVecPtr;           // xVecPtr is global  
    int n = xVec.size();  
  
    double lnL = 0.0;  
    for (int i=0; i<n; i++){  
        double x = xVec[i];  
        double pdf = expPdf(&x, par);  
        if ( pdf > 0.0 ) {  
            lnL += log(pdf);    // need pos  
        }  
        else {  
            cout << "WARNING -- pdf is negative!!!" << endl;  
        }  
    }  
    f = -2.0 * lnL;           // factor of -2 so minuit gets the errors right  
  
    cout<<"par[0] " << par[0] << "\t" << "-2lnL " << f << endl;  
  
}                               // end of fcn
```

npar: 参数个数, par的维数
par: 具体参数值
f: 取当前参数值情况下
函数值, 可以是似然
值, X^2 , 等等
iflag和deriv: 从没有使用过。

配置TMinuit

```
const int npar = 1;           // the number of parameters
TMinuit minuit(npar);
minuit.SetFCN(fcn);
```

生成TMinuit实例，并指定fcn

```
for (int i=0; i<npar; i++){
    minuit.DefineParameter(i, parName[i].c_str(),
        par[i], stepSize[i], minVal[i], maxVal[i]);
}
```

指定参数的起始值，起始步长，上下限

```
// Configure minuit
minuit.SetErrorDef(1);           // delta -2lnL = 1 for Error definition
minuit.SetMaxIterations(500);
```

指定误差的定义方法

开始运行TMinuit, 查看输出

```
minuit.Migrad();           // Minuit's best minimization algorithm
```

TMinuit的配置是很简单的

开始运行之后, **TMinuit**开始在一定区域内调整**fcn**的输入参量, 寻找**fcn**的极小值**f**。

```
double outpar[npar], err[npar];
for (int i=0; i<npar; i++){
    minuit.GetParameter(i,outpar[i],err[i]);
    cout<<"par["<<i<<"] "<<outpar[i]<<" +/- "<<err[i]<<endl;
}
```

极值处的参数值, 及误差

```
// Print results
```

```
Double_t amin,edm,errdef;
Int_t nvpar,nparx,icstat;
minuit.mnstat(amin,edm,errdef,nvpar,nparx,icstat);
```

```
cout<<"Best likelihood "<<amin<<
```

函数极值, 最优似然值

运行过程，结果检查（1）

```
PARAMETER DEFINITIONS:
  NO.   NAME      VALUE      STEP SIZE    LIMITS
    1 xi         2.00000e+00  2.00000e-01  1.00000e-07  1.00000e+08
MINUIT WARNING IN PARAM DEF
===== LIMITS ON PARAM1 TOO FAR APART.
*****
**    1 **SET ERRDEF          1
*****
*****
**    2 **MIGRAD
*****

FIRST CALL TO USER FUNCTION AT NEW START POINT, WITH IFLAG=4.
par[0] 2          -2lnL 58.8236
START MIGRAD MINIMIZATION. STRATEGY 1. CONVERGENCE WHEN EDM .LT. 1.00e-04
par[0] 2.0106     -2lnL 59.05
par[0] 1.98942    -2lnL 58.5971
par[0] 1.98942    -2lnL 58.5971
par[0] 1.9474     -2lnL 57.6899
par[0] 1.84429    -2lnL 55.4156
par[0] 1.55181    -2lnL 48.5778
par[0] 0.825772   -2lnL 30.256
```

起始值
误差定义
最小化方法

查看每一步迭代的
参数值和对应的函
数值（似然值）

运行过程，结果检查（2）

```
FCN=26.7553 FROM MIGRAD      STATUS=INITIATE      13 CALLS      14 TOTAL
                        EDM= unknown      STRATEGY= 1      NO ERROR MATRIX
EXT PARAMETER      CURRENT GUESS      STEP      FIRST
NO.   NAME        VALUE      ERROR      SIZE      DERIVATIVE
  1   xi          5.85846e-01  2.00000e-01  -1.29762e-04** at limit **
par[0] 0.574235 -2lnL 26.7439
par[0] 0.579924 -2lnL 26.7466
par[0] 0.568574 -2lnL 26.7472
par[0] 0.574523 -2lnL 26.7439
par[0] 0.580214 -2lnL 26.7469
par[0] 0.56886 -2lnL 26.7469
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
par[0] 0.574523 -2lnL 26.7439
par[0] 0.580214 -2lnL 26.7469
par[0] 0.56886 -2lnL 26.7469
par[0] 0.577365 -2lnL 26.7447
par[0] 0.571688 -2lnL 26.7447
par[0] 0.575091 -2lnL 26.744
par[0] 0.573956 -2lnL 26.744
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=26.7439 FROM MIGRAD      STATUS=CONVERGED      26 CALLS      27 TOTAL
                        EDM=1.64469e-07      STRATEGY= 1      ERROR MATRIX ACCURATE
EXT PARAMETER      ERROR      STEP      FIRST
NO.   NAME        VALUE      SIZE      DERIVATIVE
  1   xi          5.74523e-01  1.04894e-01  7.48942e-07** at limit **
EXTERNAL ERROR MATRIX.      NDIM= 25      NPAR= 1      ERR DEF=1
1.100e-02
par[0] 0.574523+/-0.104894
Best likelihood 26.7439
```

初始化

拟合收敛！
下面求误差

拟合成功！

不能放过任何一个warning信息！



作业：继续研究之前的一个测试统计量

- ▶ 产生1000个有100个事例的高斯分布， $\mu=0$ ， $\sigma=1$
 1. 利用Minuit，用高斯函数来做似然值拟合，全部参数自由，记录每次的拟合结果， μ ， σ ，likelihood
 - 画上述三个变量的分布
 2. 仍用高斯函数拟合，固定 $\mu=0.05$ ， σ 自由，记录得到的 σ ，likelihood
 - 画上述两个变量的分布
 3. 固定 $\mu=0.2$ ， σ 自由，记录得到的 σ ，likelihood
 4. 固定 $\mu=0.5$ ， σ 自由，记录得到的 σ ，likelihood
- ▶ 对比上述四种拟合的 σ ，likelihood的分布，画在一起



计算并统计test statistic的分布

$L(\hat{\mu}, \hat{\sigma})$ μ 和 σ 两个参数都是自由的情况下，拟合得到的似然值

$L(\mu, \hat{\sigma})$ 只有 σ 一个参数是自由的情况下，拟合得到的似然值，这时 μ 可以取真值0，也可以取0.05, 0.1, 0.2分别计算

$$q_{\mu} = -2 \ln \lambda(\mu) = -2 \ln \frac{L(\mu, \hat{\sigma})}{L(\hat{\mu}, \hat{\sigma})}$$

画出 μ 可以取真值0, 0.05, 0.1, 0.2时的 q_{μ} 分布