事例选择、本底估计、信号抽取和效率修正

李刚

LI.Gang@ihep.ac.cn

BESIII Winter School， 2020 Jan 17-19

提纲

1. 简单介绍

   别忘了目标?

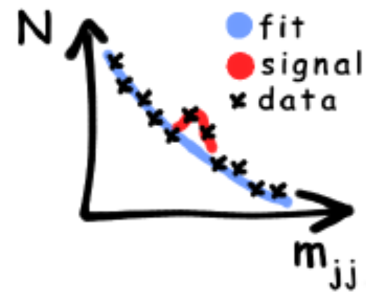
2. 事例选择
3. 本底分析
4. 信号抽取
5. 效率修正
6. 小结

贯穿整个分析过程的是如何减小误差 and/or 提高显著性

数据分析的目的，从物理角度来看
测量作为发现的手段

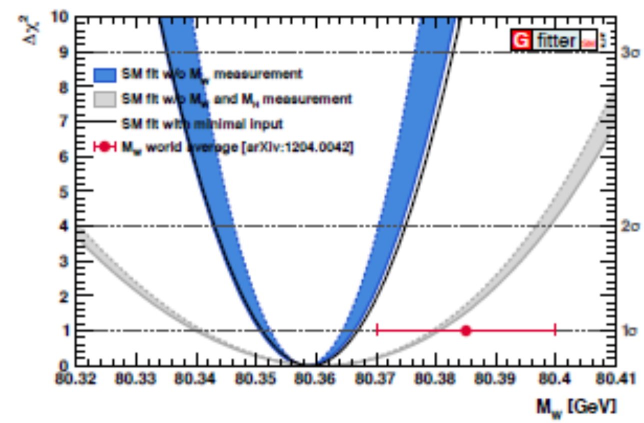
直接

- 看到新粒子或者新现象
- 例子：Higgs, XYZ，Pc, …



间接

- 精确测量关键物理量
- 和理论预期做对比，差别意味着新的东西
- 例子： measure the H、W、Z, etc., precisely



数据分析过程，从统计角度
分类、回归

分类

- **本底、信号**
- 相同末态本底、不同末态本底

回归

- **建立模型，检验模型，确定模型参数**

数据分析的目的，从统计角度来看
统计推断和假设检验

测量关键物理量

- 截面，分支比，常数
- …

检验（毙掉、证实）各种模型

- Zc：分子态，四夸克态等模型
- 标准模型：Higgs 测量
- …

动力学

振幅

M（A）

分支比、截面是振幅积分（信息压缩）的一种形式

- $$B = \frac{\Gamma_i}{\Gamma_{total}} = \frac{N_i}{N_{total}}$$
- $$\sigma \propto \Gamma_i \propto \int |M|^2 d\Omega$$

更实用的公式

$$\boxed{B(X \to f) = \frac{n^{obs} - n^{bkg}}{\varepsilon N_X}}$$

预期精度?

- B：待测分支比，无量纲
- n^{obs}：观测事例数
- n^{bkg}：预期的本底事例数
- N_x：总事例数
- ε：事例选择效率，是主要的输入信息，也是关注最多的核心部分

拿到课题，知道了反应道末态，
写出Br公式，你的事例选择就已经开始了！

事例选择之 ABC

以

$$\boxed{\psi' \to \pi^+\pi^- J/\psi, J/\psi \to l^+l^-}$$

过程为例

事例选择

* 事例选择条件
  * 简：系统误差
  * 松：系统误差
  * 少：系统误差
* 利用统计原理
  * 相对测量，抵消（系统）误差
  * 同时测量，约束（统计）误差
* 利用运动学、动力学特征
  * 动量分布
  * 角分布
  * 中间窄共振

好的选择应该考虑到系统误差

事例选择 —
已有的知识&Monte Carlo

- 一般来说：你的课题的大部分知识都已经含在手头的产生子里了，那就利用起来
- 如果有产生子，哪怕是最简单的相空间产生子，也会给我们提供很多指导
  - 把尽可能多的变量画出来，比较信号、本底
- 如果完全没有产生子，那就得从头来 … …
- 现在假设我们没有产生子

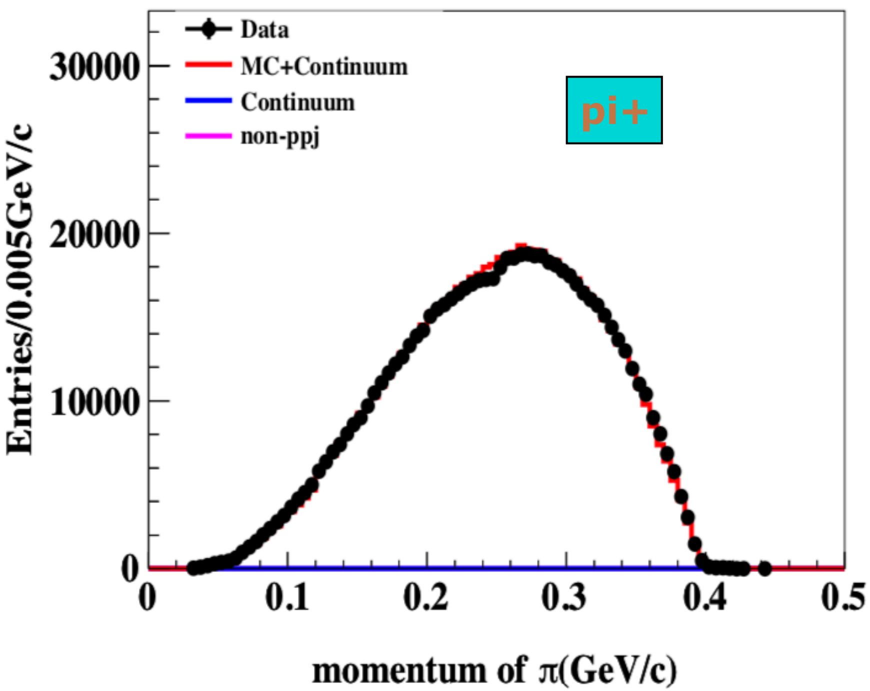
例子第一季 $\psi' \to \pi^+\pi^- J/\psi$

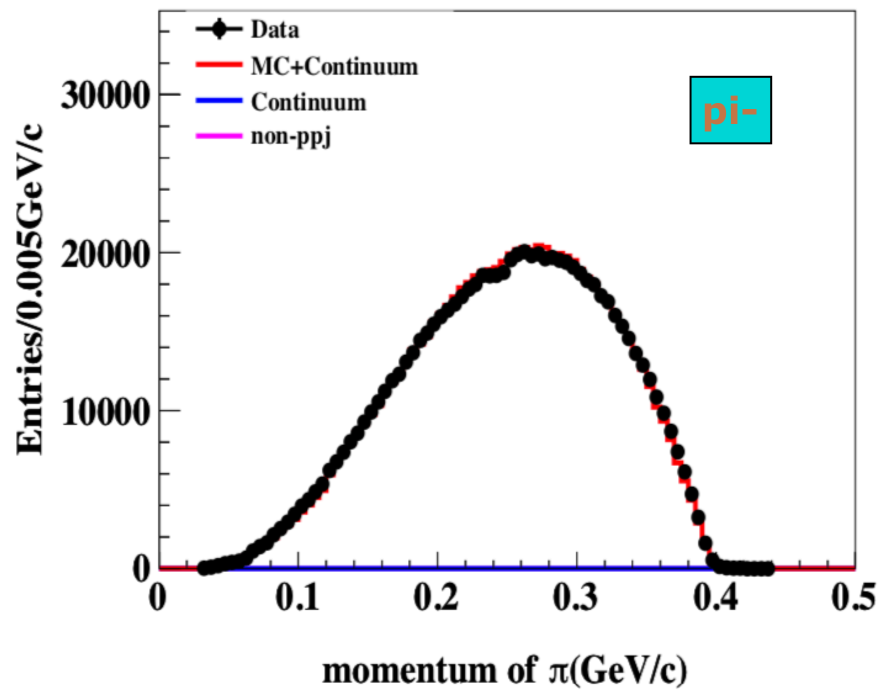
* 这是一个三体衰变，母粒子和子粒子的质量等非常清楚
* 那么末态粒子的动量范围就很容易知道：运动学知识。
* PDG 上也可以查到
  * 末态pi 动量不超过401 MeV，考虑分辨 450 MeV 非常安全：无需考虑这个 cut 的系统误差
  * pi 的动量不高，故其动量的绝对误差也就很小： $\Delta p \propto \sqrt{p}$
  * pipi 的反冲侧是非常窄的 J/psi，可以对 pipi 系统进行很好的约束，故无需 pid

$$\boxed{|\boldsymbol{p}_3|_{\max} = \frac{\left[\left(M^2-(m_1+m_2+m_3)^2\right)\left(M^2-(m_1+m_2-m_3)^2\right)\right]^{1/2}}{2M}}$$

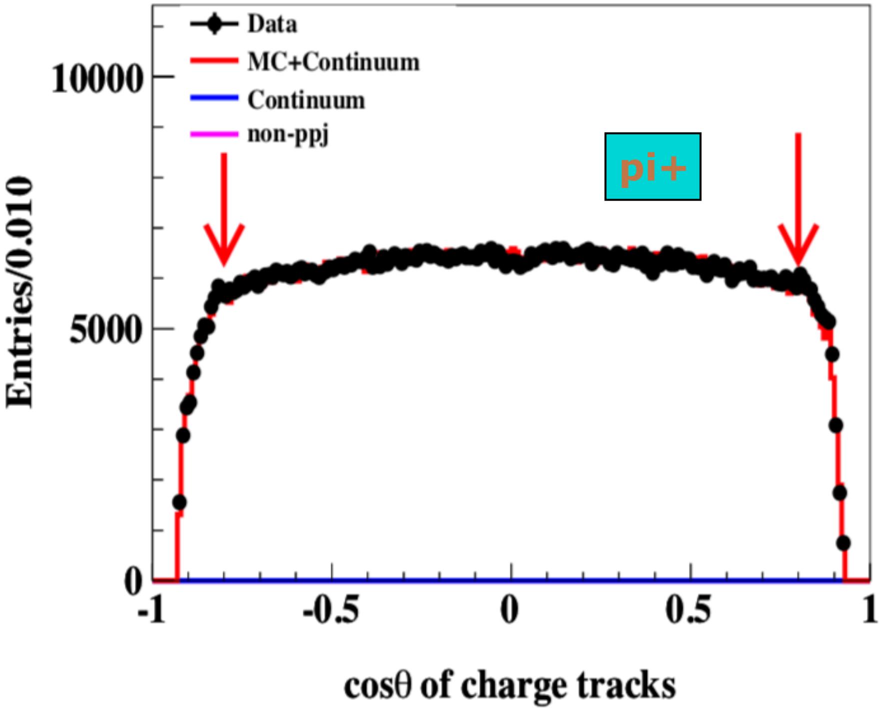
```
[08:32 上午]: ~/config/bin$ threebody.py  3.686 0.14 0.14 3.097
mass0 =     3.68600
mass1 =     0.14000
mass2 =     0.14000
mass3 =     3.09700
The maximum momentum of particl 1 is     0.40055
The maximum momentum of particl 2 is     0.40055
The maximum momentum of particl 3 is     0.47638
[08:32 上午]: ~/config/bin$ _
```

pi 的动量分布

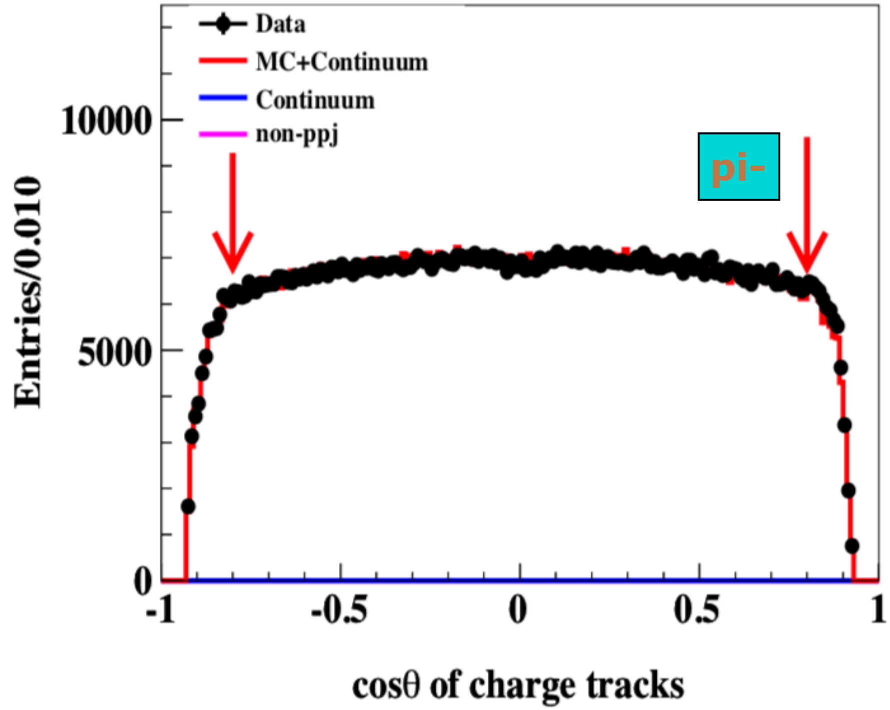

Data
MC+Continuum
Continuum
non-ppj
pi+
Entries/0.005GeV/c
momentum of π(GeV/c)
0
10000
20000
30000
0
0.1
0.2
0.3
0.4
0.5


Data
MC+Continuum
Continuum
non-ppj
pi-
Entries/0.005GeV/c
momentum of π(GeV/c)
0
10000
20000
30000
0
0.1
0.2
0.3
0.4
0.5

pi 的极角

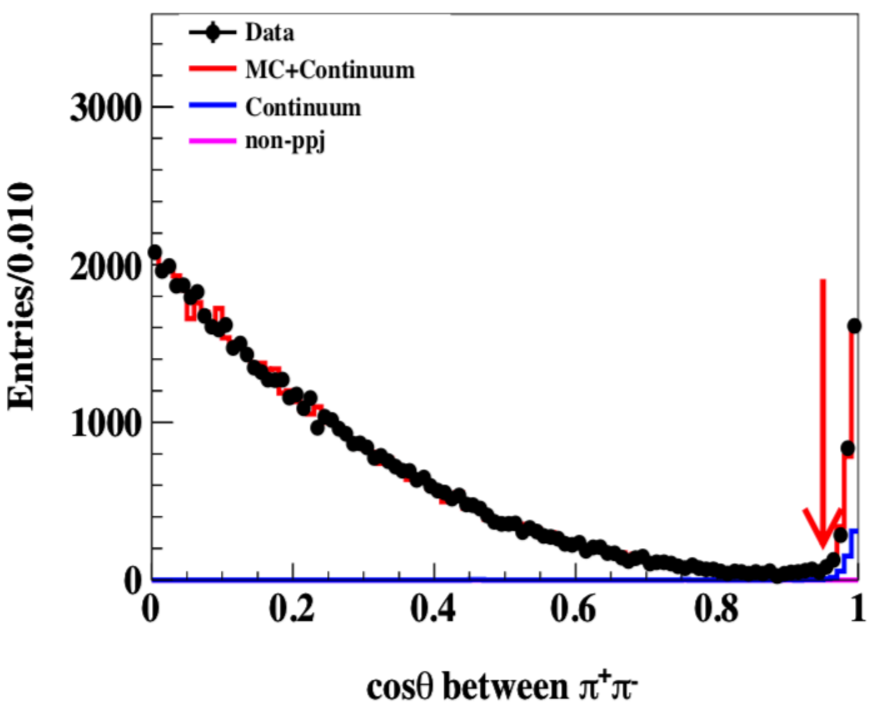


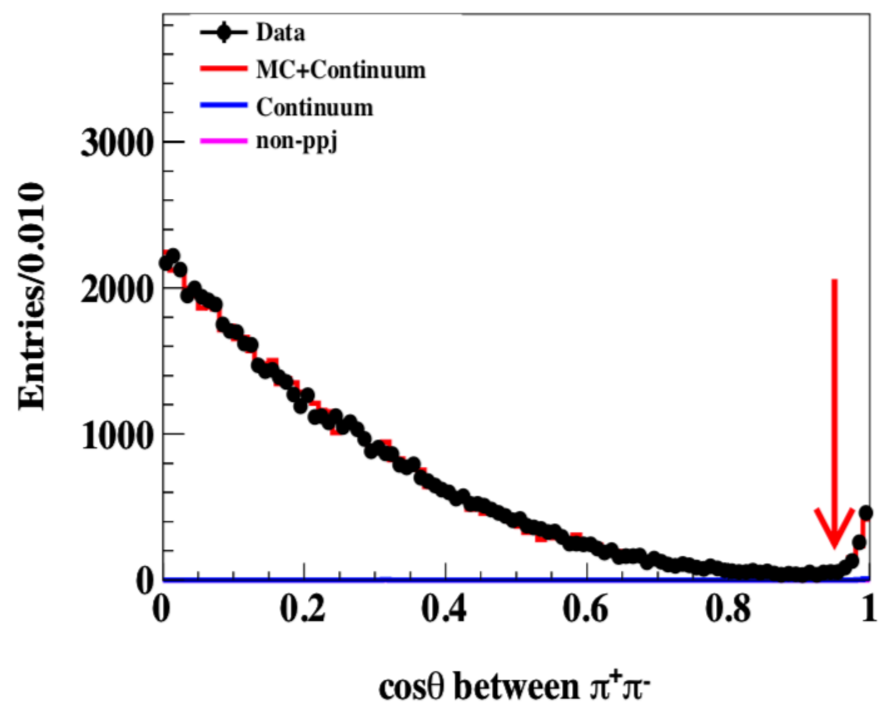
Data
MC+Continuum
Continuum
non-ppj
pi+
10000
5000
0
Entries/0.010
-1
-0.5
0
0.5
1
cosθ of charge tracks



Data
MC+Continuum
Continuum
non-ppj
pi-
10000
5000
0
Entries/0.010
-1
-0.5
0
0.5
1
cosθ of charge tracks

pi pi夹角

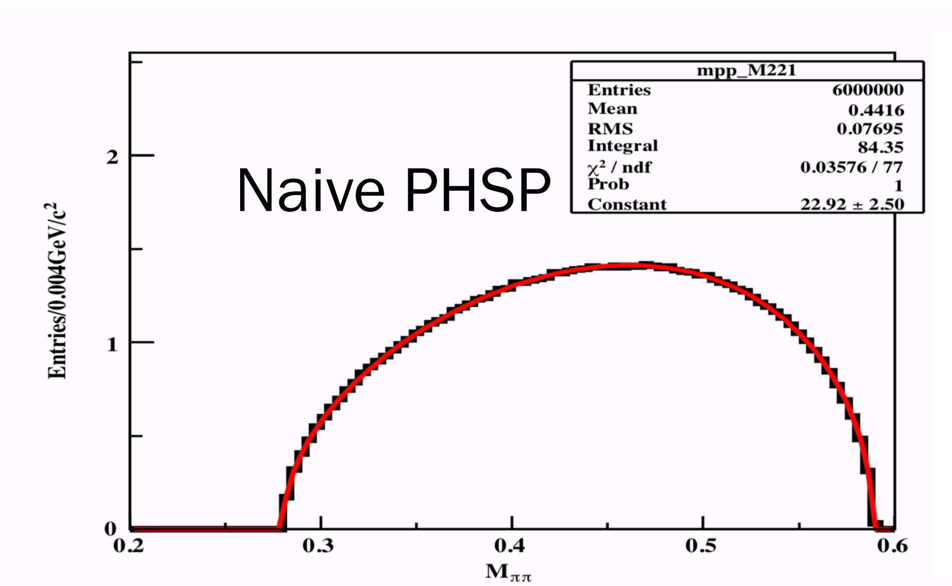

Data
MC+Continuum
Continuum
non-ppj
Entries/0.010
3000
2000
1000
0
0
0.2
0.4
0.6
0.8
1
cosθ between π+π-

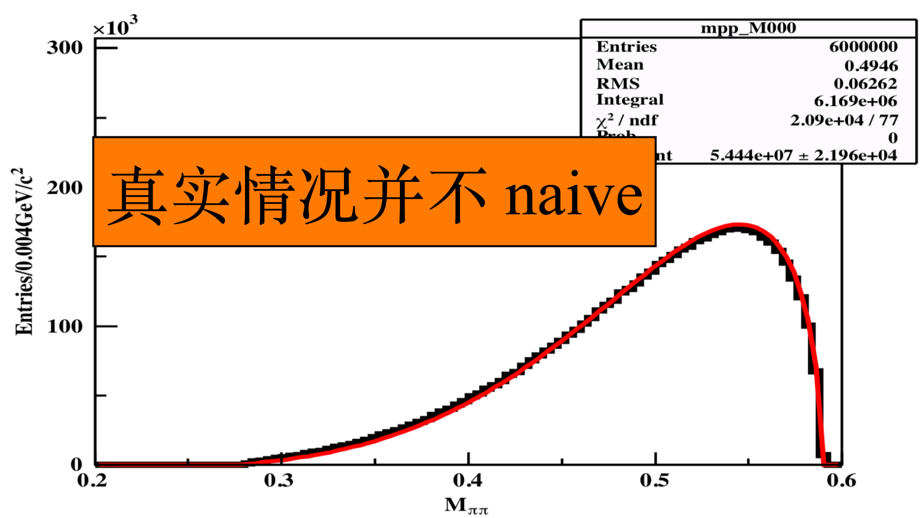

Data
MC+Continuum
Continuum
non-ppj
Entries/0.010
3000
2000
1000
0
0
0.2
0.4
0.6
0.8
1
cosθ between π+π-

$M_{\pi\pi}$ spectrum

BesEvtGen： JPIPI model

Phys. Rev. D 62, 032002

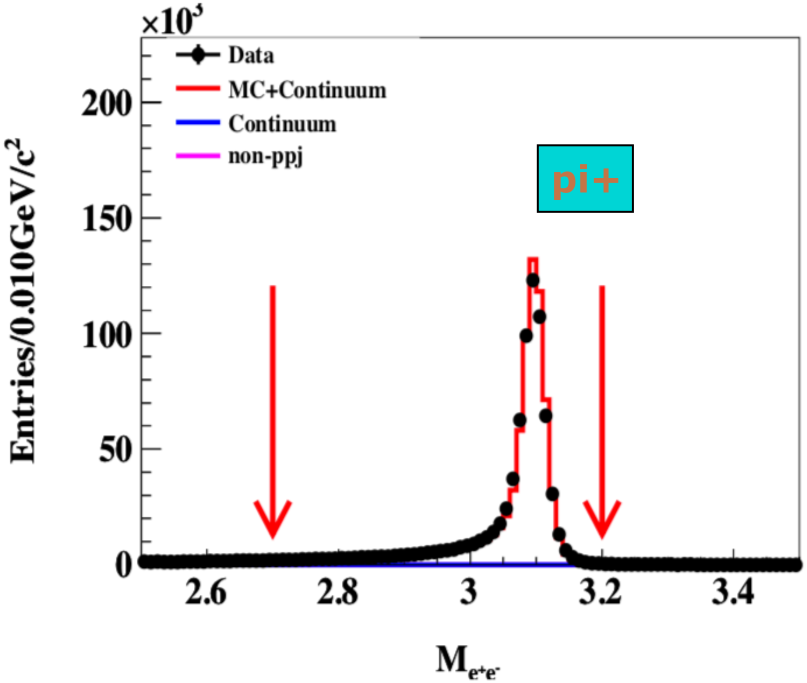


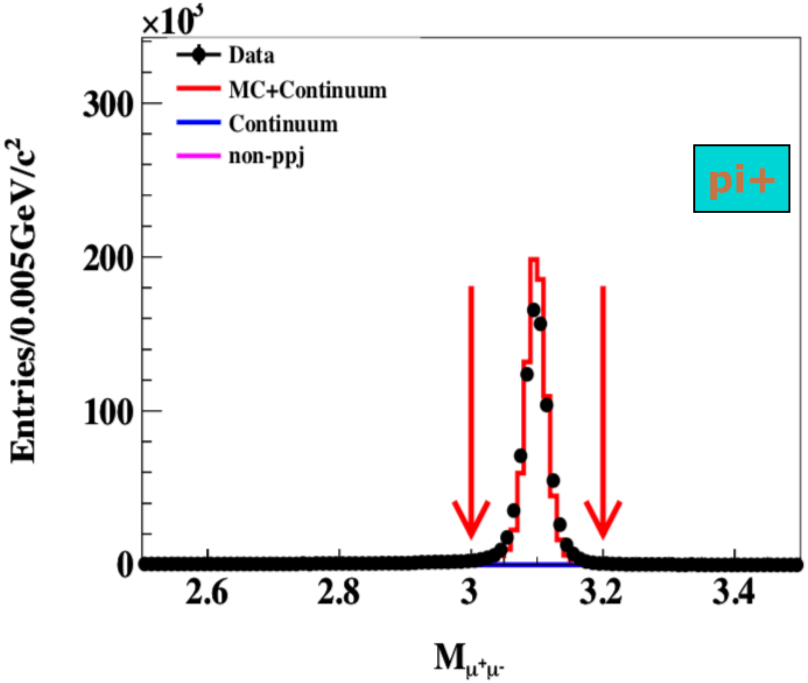


两 pi 倾向于背对背，$M_{\pi\pi}$对效率有显著影响。
这个谱的特点对于去掉光子到电子对的本底非常有用。

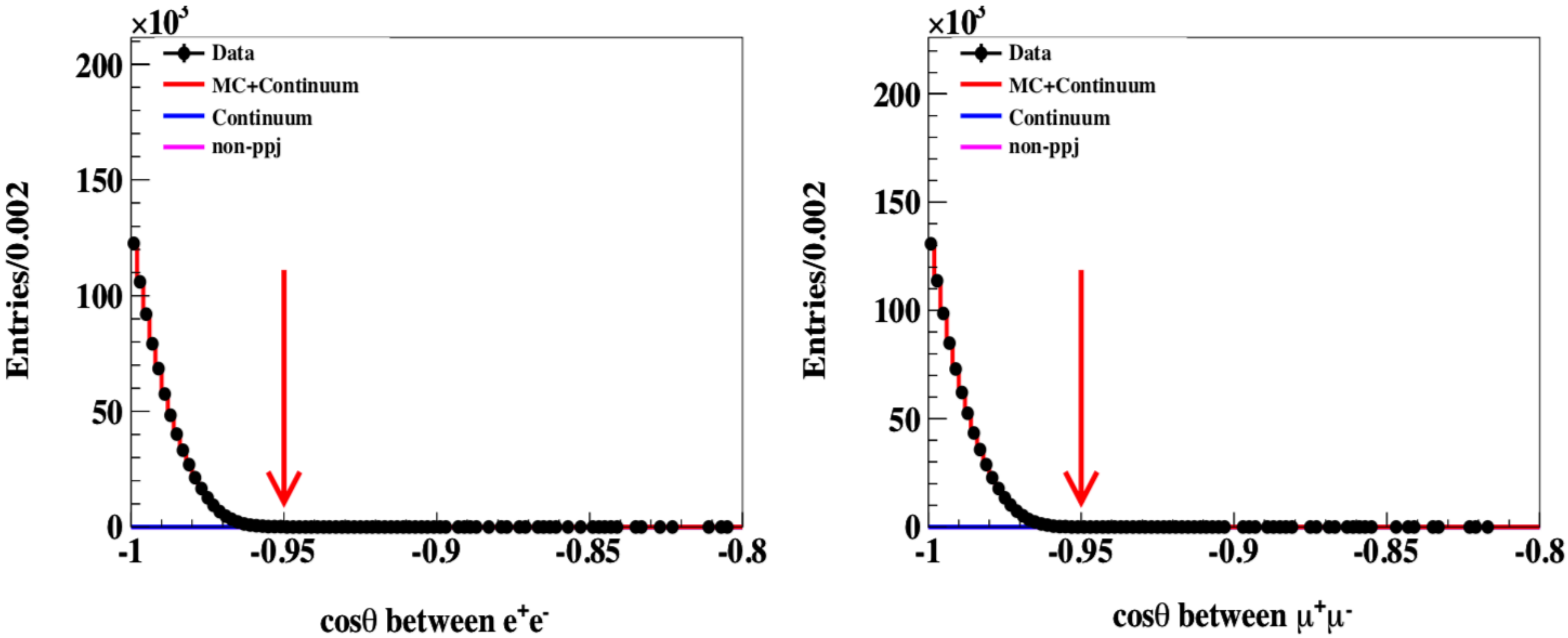
$$\frac{d\sigma}{dm_{\pi\pi}} \propto (s - 2m_{\pi\pi}^2)$$

轻子不变质量


×10³
Data
MC+Continuum
Continuum
non-ppj
pi+
Entries/0.010GeV/c²
200
150
100
50
0
2.6
2.8
3
3.2
3.4
$M_{e^+e^-}$


×10³
Data
MC+Continuum
Continuum
non-ppj
pi+
Entries/0.005GeV/c²
300
200
100
0
2.6
2.8
3
3.2
3.4
$M_{\mu^+\mu^-}$

轻子夹角



×10³
Data
MC+Continuum
Continuum
non-ppj
Entries/0.002
200
150
100
50
0
-1
-0.95
-0.9
-0.85
-0.8
cosθ between e⁺e⁻
cosθ between μ⁺μ⁻

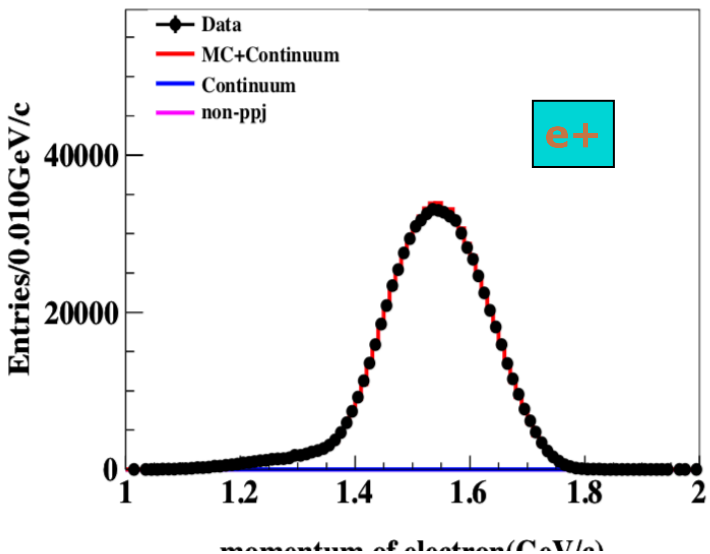
例子第二季：$J/\psi \to l^+l^-$

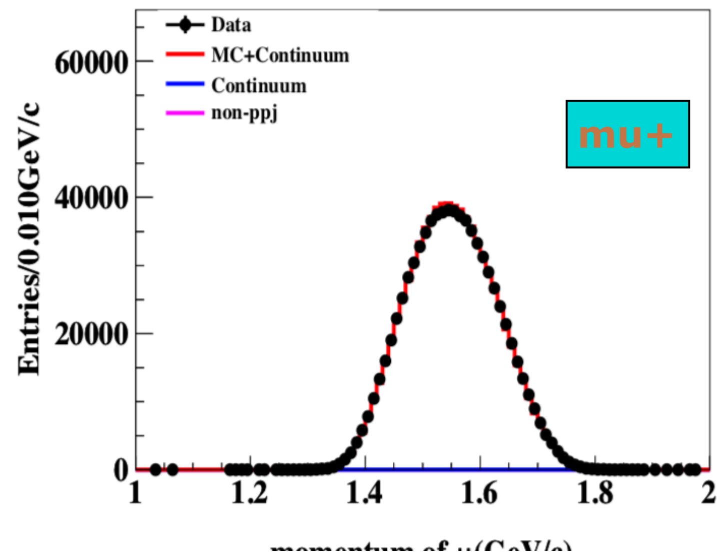
$$|\boldsymbol{p}| = \frac{\left[\left(M^2 - (m_1 + m_2)^2\right)\right]^{1/2}}{2}$$

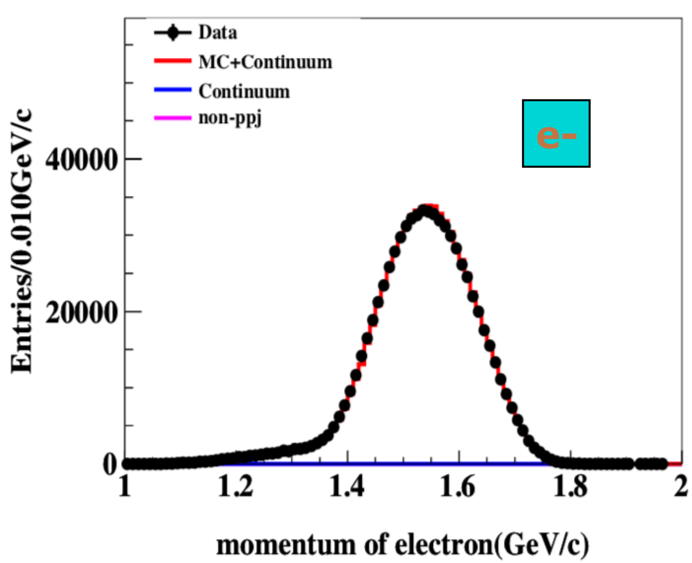
* 两体衰变，在母粒子质心系是单（高）能的：大约1.55 GeV
* 即使 J/psi 运动导致一定的动量展宽，仍然
* 和 pi 的动量区间完全无重叠：无需粒子鉴别，只需区分 e、mu 即可
* 因此这个分析无需用到 dE/dx 和 TOF 的 PID，也就不用分析这项系统误差

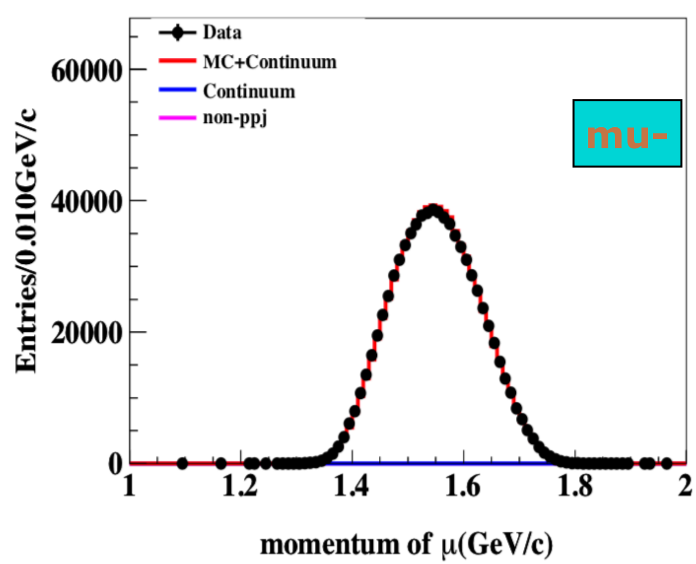
```
[09:08 上午]: ~/config/bin$ twobody.py  3.097 0.106 0.106
mass0 =     3.09700
mass1 =     0.10600
mass2 =     0.10600
The maximum momentum of particl 1/2 is     1.54487
The maximum   energy of particl 1   is     1.54850
The maximum   energy of particl 2   is     1.54850
```

轻子动量分布

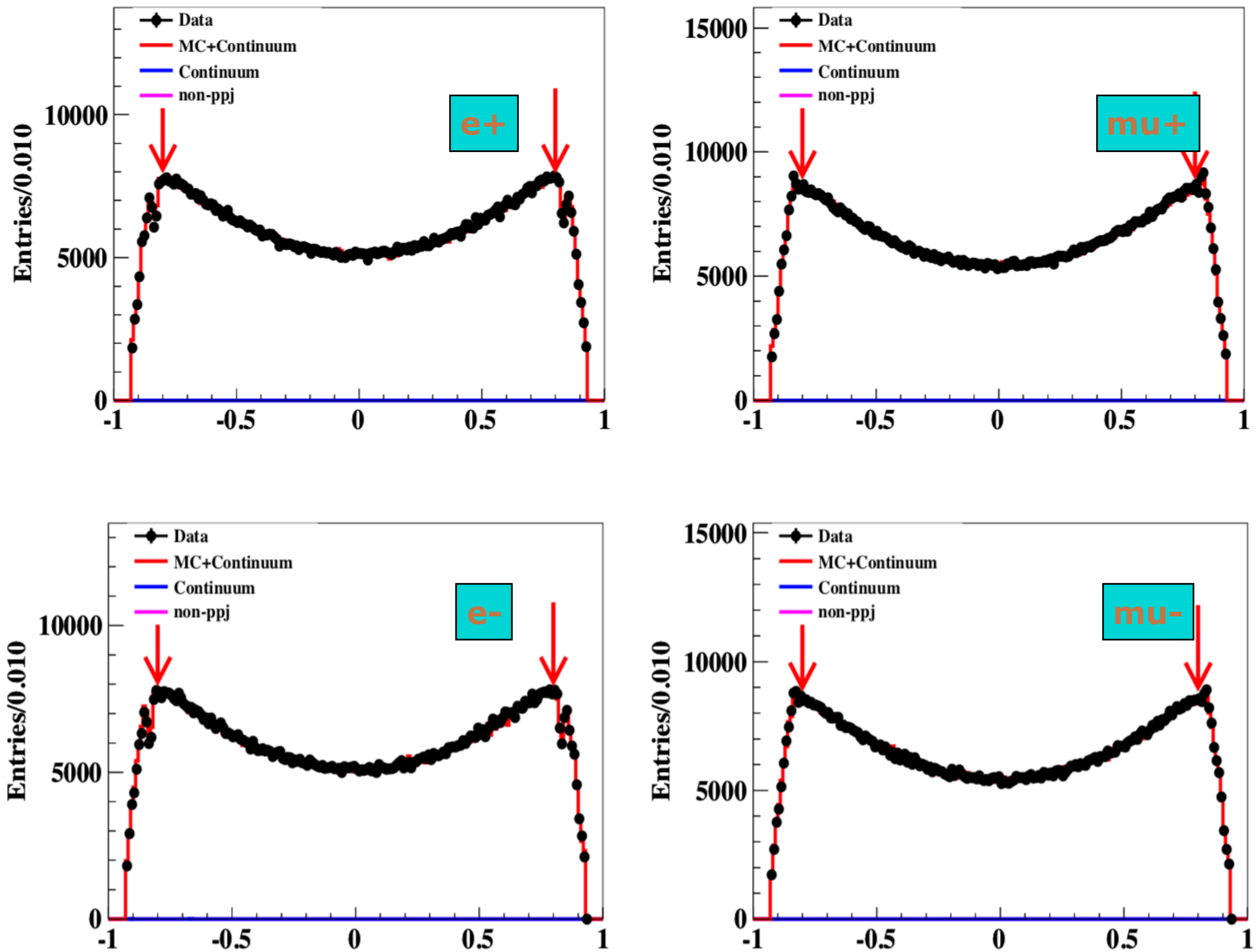

Data
MC+Continuum
Continuum
non-ppj
e+
Entries/0.010GeV/c
40000
20000
0
1
1.2
1.4
1.6
1.8
2


Data
MC+Continuum
Continuum
non-ppj
mu+
Entries/0.010GeV/c
60000
40000
20000
0
1
1.2
1.4
1.6
1.8
2


Data
MC+Continuum
Continuum
non-ppj
e-
Entries/0.010GeV/c
40000
20000
0
1
1.2
1.4
1.6
1.8
2
momentum of electron(GeV/c)

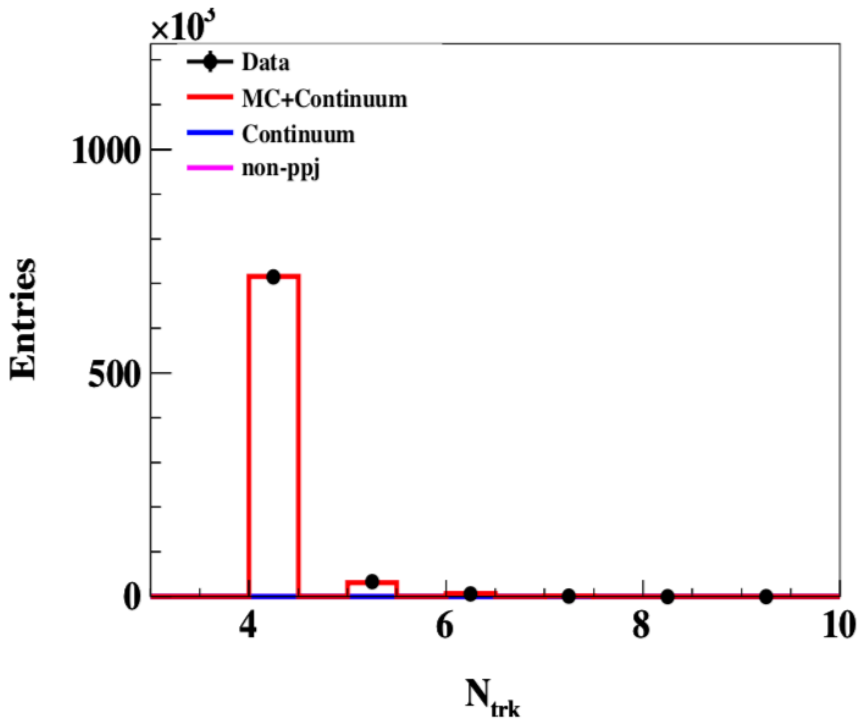

Data
MC+Continuum
Continuum
non-ppj
mu-
Entries/0.010GeV/c
60000
40000
20000
0
1
1.2
1.4
1.6
1.8
2
momentum of μ(GeV/c)

轻子极角

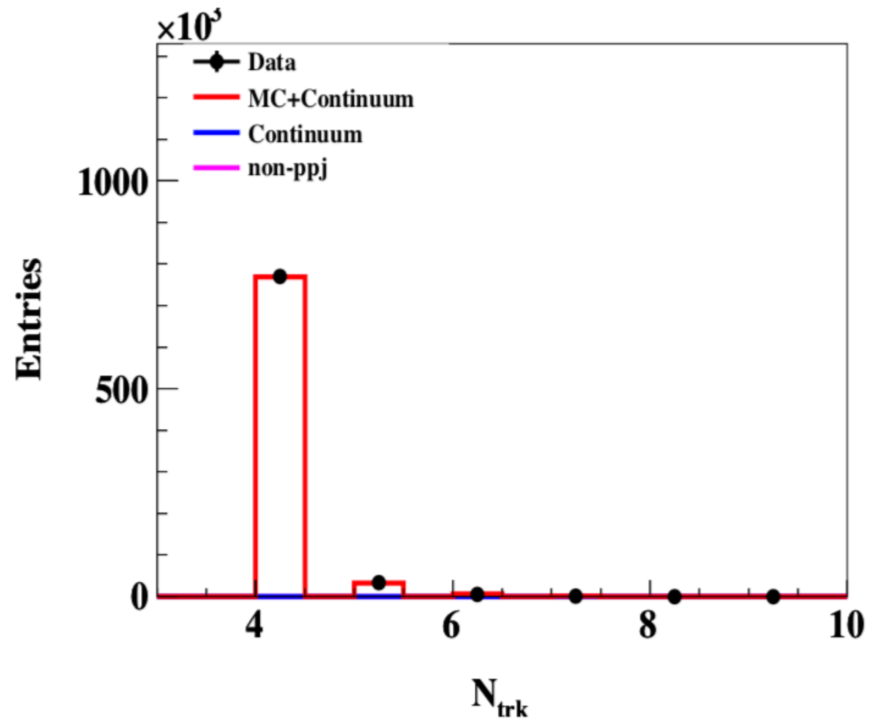


Data
MC+Continuum
Continuum
non-ppj
e+
mu+
e-
mu-
Entries/0.010
0
5000
10000
15000
-1
-0.5
0
0.5
1

带电径迹数

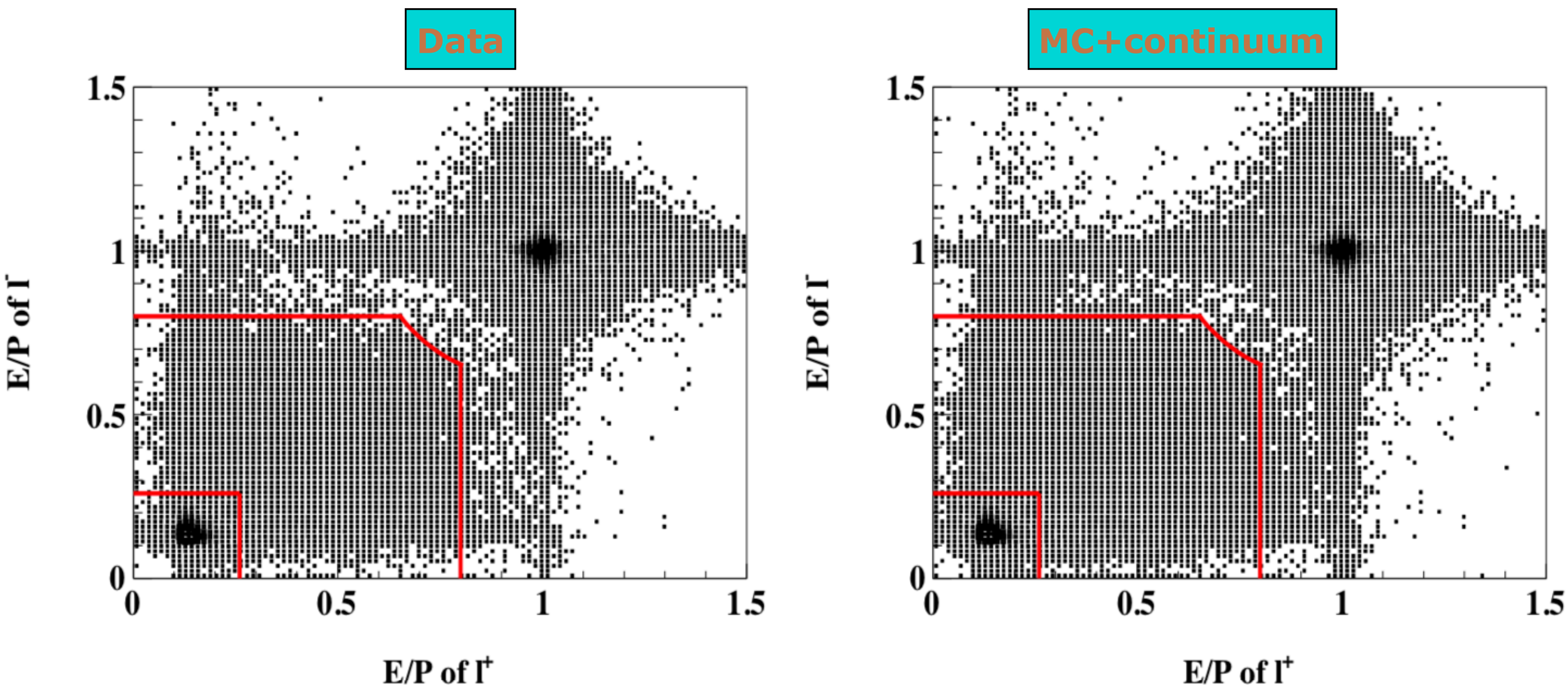


×10³
Data
MC+Continuum
Continuum
non-ppj
Entries
1000
500
0
4
6
8
10
N_trk



×10³
Data
MC+Continuum
Continuum
non-ppj
Entries
1000
500
0
4
6
8
10
N_trk

ep 比



事例选择条件—inclusive

- nGood $>= 4$
- The charged tracks with $|\vec{p}| < 0.45$ GeV/c are assumed to be pion, select the $\pi^+\pi^-$ pair candidates by minimizing $|M^{rec.}_{\pi^+\pi^-} - M_{J/\psi}|$
- $\cos\theta_{\pi^+\pi^-} < 0.95$
- $3.05\mathrm{GeV}/c^2 \le M^{rec.}_{\pi^+\pi^-} \le 3.15\mathrm{GeV}/c^2$
- Take the two fastest positive and negative tracks as lepton candidates, identify the e/μ pair
  - $\mu^+\mu^-$:$[E/p]^+ < 0.26$ and $[E/p]^- < 0.26$
  - e^+e^-:$[E/p]^+ > 0.80$ or $[E/p]^- > 0.80$ or $\sqrt{([E/p]^+ - 1)^2 + ([E/p]^- - 1)^2} < 0.4$
- $\cos\theta_{l^+l^-} < -0.95$
- $2.7\mathrm{GeV}/c^2 < m_{l^+l^-} < 3.2\mathrm{GeV}/c^2$ for $\pi^+\pi^-e^+e^-$ channel and $3.0\mathrm{GeV}/c^2 < m_{l^+l^-} < 3.2\mathrm{GeV}/c^2$ for $\pi^+\pi^-\mu^+\mu^-$ channel

⭐径迹数允许大于 4

⭐高动量的动量无要求，隐含在 mass window里

事例选择条件—inclusive

- Track level cuts for soft pion candidates
  - $|V_z| < 10\text{cm}$
  - $|V_r| < 1\text{cm}$
  - $|\cos\theta| < 0.80$
  - $|\vec{p}| < 0.45\text{GeV}/c$
  - 没有 pid，所有带电粒子都假设为 pi
  - 所有反冲质量在[3.05,3.15]GeV中的entry都存下来使用

$$M^{rec.}_{\pi^+\pi^-} = \sqrt{(p_{ecm} - p_{\pi^+} - p_{\pi^-})^2}$$

☆只用了桶部测量、模拟比较可靠的部分：系统误差

☆完全没看光子的情况：无需考虑光子的模拟，无系统误差

事例选择条件—exclusive

- Track level cuts
  - $|V_z| < 10\mathrm{cm}$
  - $|V_r| < 1\mathrm{cm}$
  - $|\cos\theta| < 0.80$
  - $|\vec{p}| < 2.0\mathrm{GeV}/c$

☆ 也没涉及到光子

事例选择条件—exclusive

- nGood >= 4
- The charged tracks with $|\vec{p}| < 0.45$ GeV/c are assumed to be pion, select the $\pi^+\pi^-$ pair candidates by minimizing $|M^{rec.}_{\pi^+\pi^-} - M_{J/\psi}|$
- $\cos\theta_{\pi^+\pi^-} < 0.95$
- $3.05\mathrm{GeV}/c^2 \leq M^{rec.}_{\pi^+\pi^-} \leq 3.15\mathrm{GeV}/c^2$
- Take the two fastest positive and negative tracks as lepton candidates, identify the e/μ pair
  - $\mu^+\mu^-$:$[E/p]^+ < 0.26$ and $[E/p]^- < 0.26$
  - e^+e^-:$[E/p]^+ > 0.80$ or $[E/p]^- > 0.80$ or $\sqrt{([E/p]^+ - 1)^2 + ([E/p]^- - 1)^2} < 0.4$
- $\cos\theta_{l^+l^-} < -0.95$
- $2.7\mathrm{GeV}/c^2 < m_{l^+l^-} < 3.2\mathrm{GeV}/c^2$ for $\pi^+\pi^-e^+e^-$ channel and $3.0\mathrm{GeV}/c^2 < m_{l^+l^-} < 3.2\mathrm{GeV}/c^2$ for $\pi^+\pi^-\mu^+\mu^-$ channel

- 径迹数允许大于 4
- 高动量的动量无要求，隐含在 mass window里

小结

- 设计了尽可能简单的事例选择
- 本底被尽可能压低
- 尽可能的利用了运动学特征
- 充分考虑了系统误差估计工作量

本底分析

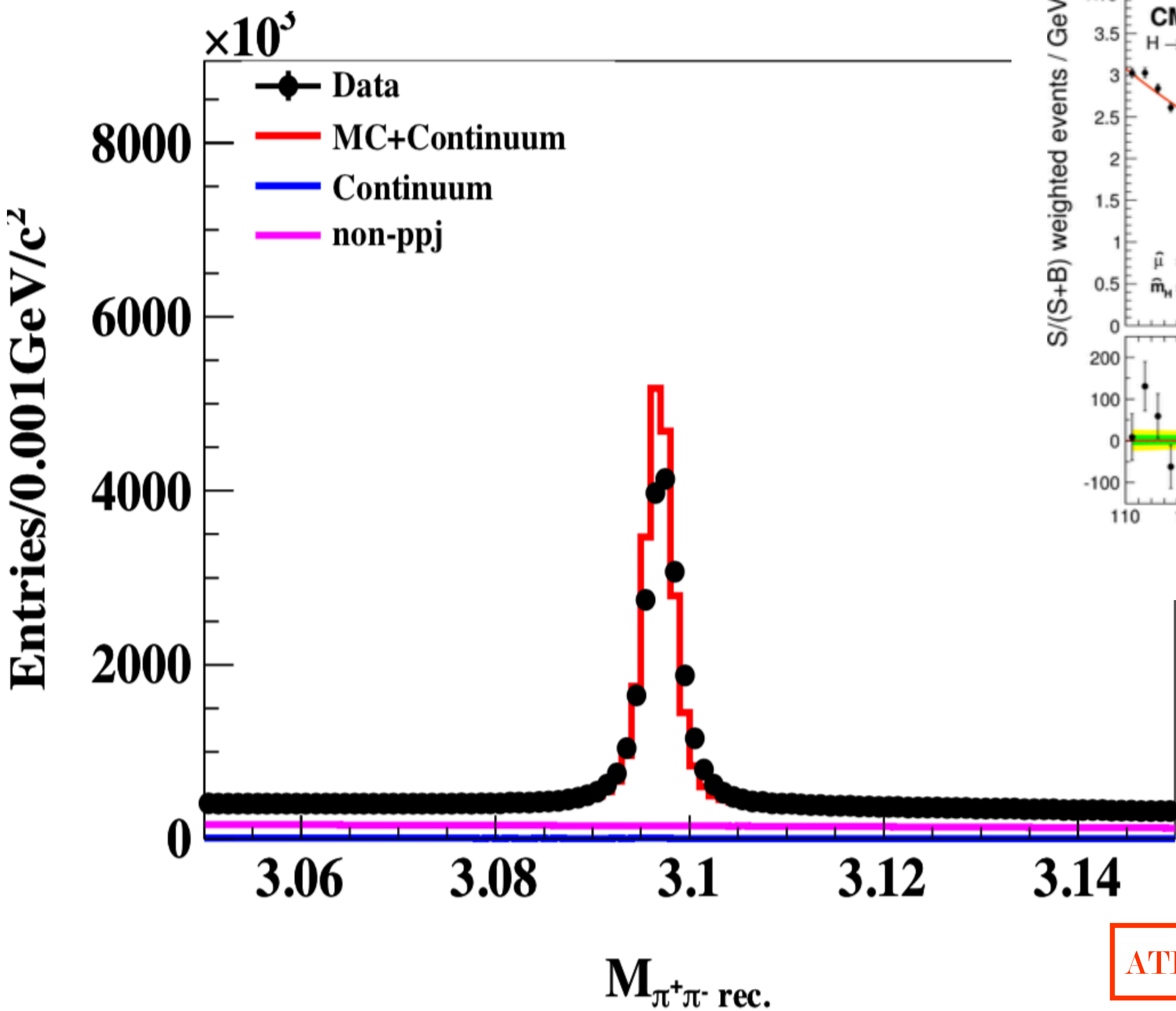
本底分析的意义

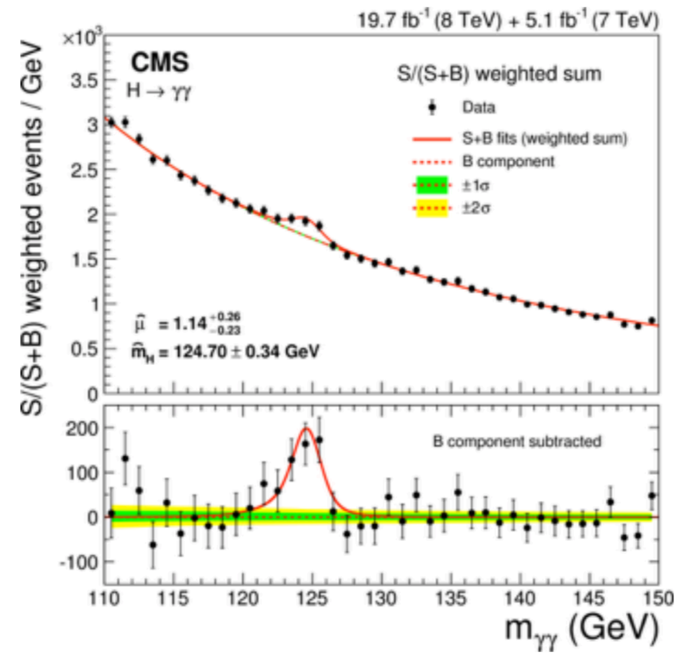
- $$Significance = \frac{s}{\sqrt{s+b}} = \sqrt{s}\sqrt{\frac{s}{s+b}} \propto \sqrt{\epsilon \cdot P}$$
- 如果本底能够被显著压低，significance 就会提高；也意味着统计误差的降低
- 简单的事例选择往往意味着效率高，有效的压低本底则是提高纯度。
- 如果本底无法明显降低，则关于本底的知识（量和形状）则变得更重要。
- 事例选择应该综合考虑，事例选择往往是在本底分析、乃至误差分析的时候仍需优化。

Inclusive 测量的本底

- 信息少（系统误差也少），所以特征少，本底一般都很高
- 理解本底是主要矛盾
  - MC
  - 用数据来 model 数据（data driven，ABCD approach）

Inclusive 过程的信号和本底





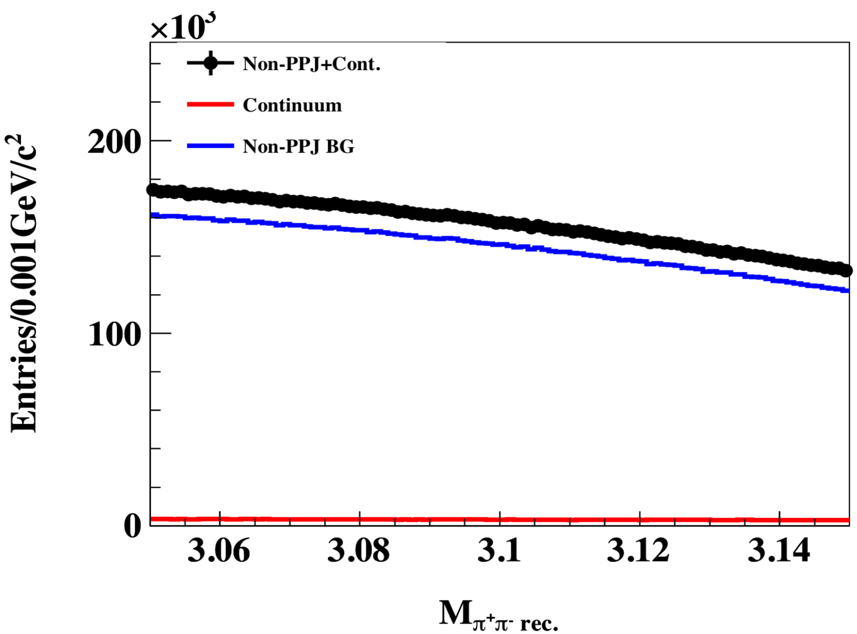
ATLAS&CMS 都是 inclusive 研究

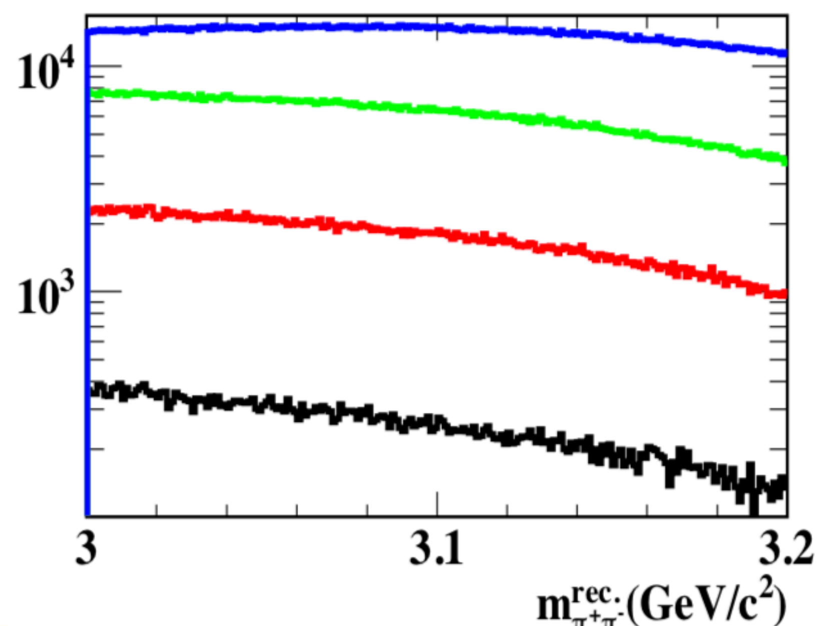
Inclusive 过程的信号和本底

- 既然无法进一步压低，就必须
  - 知道 signal 范围内是平滑的
  - 没有类似信号的起伏，甚至 peak
  - 最好数量是知道的，即使有误差，也可以减少涨落
  - 如果没有 MC 可以很好模拟，那就只能采用 data-driven
    - 统计量受限于数据统计量

Inclusive 过程的信号和本底

- 没有不平滑，没有 peak
- 受益于事例选择中：一个事例可以多个 entries
  - 虽然错误组合更多了，但是避免了人为的 bias
- 进一步检查不同的 J/psi 衰变的误组合，仍然平滑





Exclusive 测量的本底

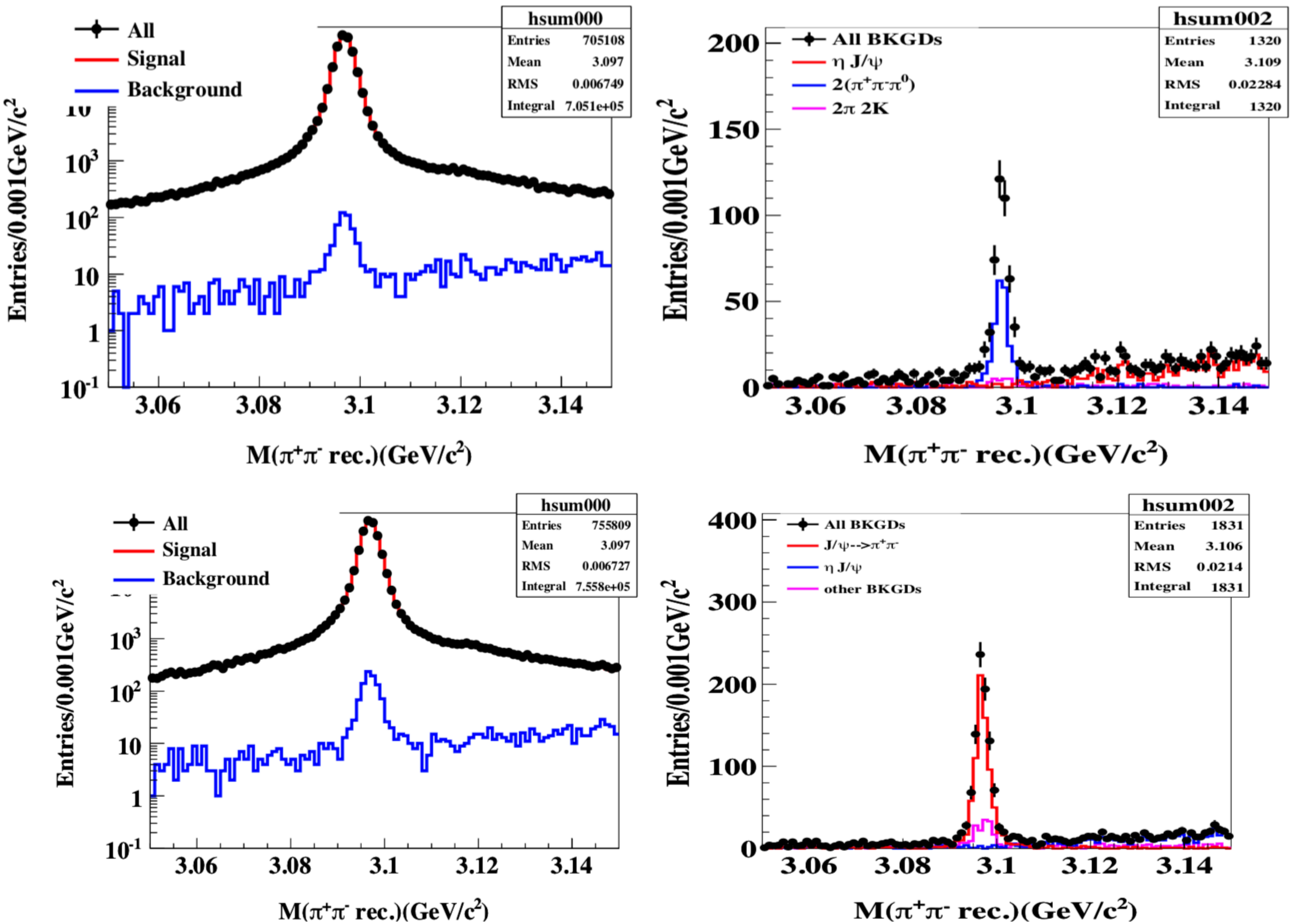
- ☑ 本底水平很低：0.1%水平，因此信号抽取采用了数数的方式
- ☑ 但是仍然和预期精度相当，需要进一步分析和处理减小影响
- ☑ 考察本底，对信号也有进一步的认识（以J/psi -› ee 为例）
- ☑ MC topology 分析工具的改进

No.	decay chain	iTopo	N_{evt}
	$e^-\pi^-e^+\pi^+$(final state)		
0	$\psi' \to \pi^+\pi^- J/\psi, J/\psi \to e^-e^+$	0	566983
1	$\psi' \to \pi^- J/\psi\pi^+, J/\psi \to e^+\gamma_{FSR}e^-$	1	118425
2	$\psi' \to \pi^- J/\psi\pi^+, J/\psi \to \gamma_{FSR}e^-e^+\gamma_{FSR}$	2	16079
3	$\psi' \to \pi^+ J/\psi\pi^-\gamma_{FSR}, J/\psi \to e^-e^+$	3	1802
4	$\psi' \to J/\psi\pi^+\pi^-\gamma_{FSR}, J/\psi \to e^-e^+\gamma_{FSR}$	9	447
11	$\psi' \to J/\psi\gamma_{FSR}\pi^-\pi^+, J/\psi \to e^-\gamma_{FSR}\gamma_{FSR}e^+$	12	51
90	$\psi' \to J/\psi\gamma_{FSR}\gamma_{FSR}\pi^+\pi^-, J/\psi \to e^-e^+\gamma_{FSR}$	90	1
	703788 evts (7 modes)		

以 J/psi -› mumu 作为对比看 FSR

ee 末态含有更多的末态辐射效应

No.	decay chain	iTopo	N_{evt}
	$\mu^-\pi^-\mu^+\pi^+$(final state)		
0	$\psi' \to \pi^- J/\psi \pi^+, J/\psi \to \mu^+\mu^-$	0	717031
1	$\psi' \to J/\psi \pi^+\pi^-, J/\psi \to \gamma_{FSR}\mu^+\mu^-$	1	33801
2	$\psi' \to \pi^-\gamma_{FSR} J/\psi \pi^+, J/\psi \to \mu^-\mu^+$	7	2458
4	$\psi' \to J/\psi \pi^+\pi^-, J/\psi \to \mu^-\gamma_{FSR}\mu^+\gamma_{FSR}$	6	569
7	$\psi' \to \gamma_{FSR} J/\psi \pi^+\pi^-, J/\psi \to \mu^-\gamma_{FSR}\mu^+$	11	113
30	$\psi' \to \pi^+ J/\psi \pi^-\gamma_{FSR}, J/\psi \to \mu^+\mu^-\gamma_{FSR}\gamma_{FSR}$	82	4
48	$\psi' \to \gamma_{FSR}\pi^- J/\psi \gamma_{FSR}\pi^+, J/\psi \to \mu^+\mu^-$	42	2
	753978 evts (7 modes)		


All
Signal
Background
hsum000
Entries 705108
Mean 3.097
RMS 0.006749
Integral 7.051e+05
Entries/0.001GeV/c²
M(π⁺π⁻ rec.)(GeV/c²)
All BKGDs
η J/ψ
2(π⁺π⁻π⁰)
2π 2K
hsum002
Entries 1320
Mean 3.109
RMS 0.02284
Integral 1320
Entries/0.001GeV/c²
M(π⁺π⁻ rec.)(GeV/c²)
All
Signal
Background
hsum000
Entries 755809
Mean 3.097
RMS 0.006727
Integral 7.558e+05
Entries/0.001GeV/c²
M(π⁺π⁻ rec.)(GeV/c²)
All BKGDs
J/ψ-->π⁺π⁻
η J/ψ
other BKGDs
hsum002
Entries 1831
Mean 3.106
RMS 0.0214
Integral 1831
Entries/0.001GeV/c²
M(π⁺π⁻ rec.)(GeV/c²)

本底扣除

- 如果把某个本底直接处理成误差，比如 764 事例，那就直接贡献 764 的误差，如果能用已知的分支比和选择效率进行扣除，就可以大大减小其对误差的贡献

No.	decay chain	iTopo	N_{evt}
	$\pi^-\pi^-\pi^+\pi^+$(final state)		
3	$\psi' \to \pi^-\pi^+J/\psi, J/\psi \to \pi^+\pi^-$	2	764
8	$\psi' \to J/\psi\pi^+\pi^-, J/\psi \to \pi^-\pi^+\gamma_{FSR}$	32	41
10	$\psi' \to \rho^0\pi^-\pi^+, \rho^0 \to \pi^+\pi^-$	31	25
28	$\psi' \to \pi^+a_1^-, a_1^- \to \pi^-\rho^0, \rho^0 \to \pi^-\pi^+$	113	5
32	$\psi' \to a_2^-\pi^+, a_2^- \to \rho^0\pi^-, \rho^0 \to \pi^+\pi^-$	116	4
33	$\psi' \to \rho^0 f_0', \rho^0 \to \pi^-\pi^+, f_0' \to \pi^+\pi^-$	127	4
35	$\psi' \to \pi^-a_2^+, a_2^+ \to \pi^+\rho^0, \rho^0 \to \pi^-\pi^+$	54	3

764 —› 0.1x764， 小一个量级；MC 任意多，效率误差忽略

问题和休息?

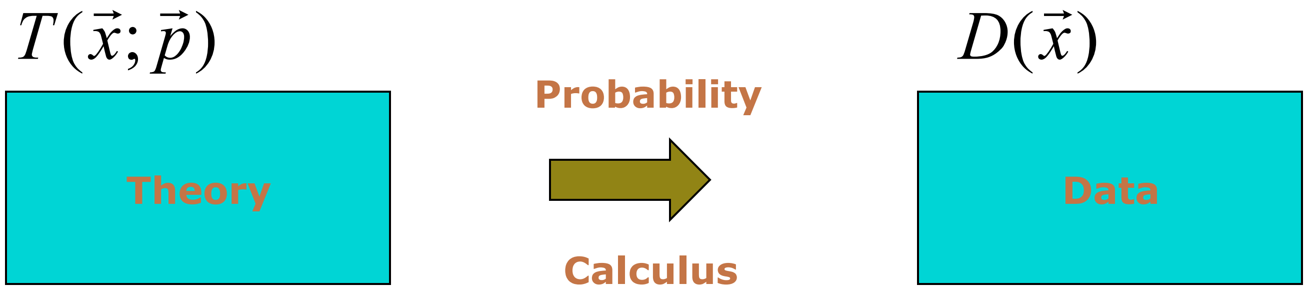
拟合

* 拟合是容易操作的，在数据分析的意义上也最不重要；只要前面的各种研究做得足够细致，拟合顺利成长的结果；在准备不充分之前做拟合是没有意义的；但是这确实最具杀伤力的一招：最终结果就在这里。

拟合的作用在演变

- 拟合的作用在于 combine 已经获得的很多信息：数据、物理模型、探测器等
- 随着技术的发展：分波分析是一个带着复杂的物理模型、参数多达上百的拟合
- ATLAS、CMS 上的会把所有 Higgs 测量结果在某种理论框架下进行 combination，本质上就是一个大的拟合。

Parameter estimation – Introduction



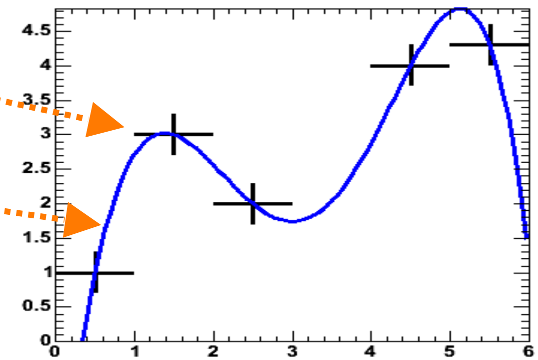
* Given the theoretical distribution parameters p, what can we say about the data

* Need a procedure to estimate *p* from *D*
  * Common technique – fit!

A well known estimator – the χ^2 fit

- Given a set of points and a function *f(x,p)* define the χ^2

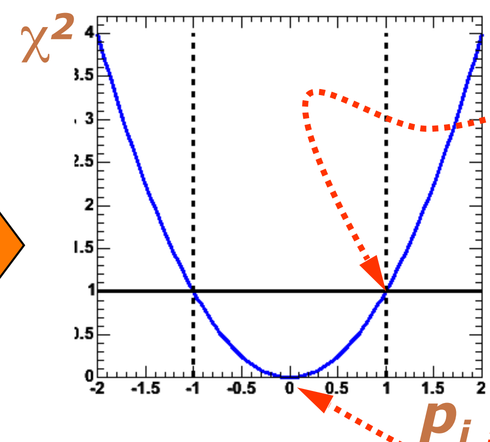
$$\{(\vec{x}_i, y_i, \sigma_i)\}$$



$$\chi^2(\vec{p}) = \sum_i \frac{(y_i - f(\vec{x};\vec{p}))^2}{\sigma_y^2}$$

- Estimate parameters by minimizing the $\chi^2(p)$ with respect to all parameters p_i
  - *In practice, look for*

$$\frac{d\chi^2(p_i)}{dp_i} = 0$$



Error on p_i is given by $\chi 2$ variation of +1

Value of p_i at minimum is estimate for p_i

- Well known: but why does it work? Is it always right? Does it always give the best possible error?

Basics – What is an estimator?

- An *estimator* is a *procedure* giving a value for a parameter or a property of a distribution as a function of the actual data values, i.e.

$$\hat{\mu}(x) = \frac{1}{N}\sum_i x_i$$

← **Estimator of the mean**

$$\hat{V}(x) = \frac{1}{N}\sum_i (x_i - \vec{\mu})^2$$

← **Estimator of the variance**

- A perfect estimator is
  - Consistent: $\lim_{n\to\infty}(\hat{a}) = a$
  - Unbiased – *With finite statistics you get the right answer on average*
  - Efficient $V(\hat{a}) = \left\langle (\hat{a} - \langle \hat{a} \rangle)^2 \right\rangle$ ← This is called the **Minimum Variance Bound**
  - *There is no perfect estimators!*

Likelihood – Another common estimator

Functions used in likelihoods must be Probability Density Functions:

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

- Definition of Likelihood
  - given D(x) and F(x;p)

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

  - 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)

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

Variance on ML parameter estimates

* The estimator for the parameter variance is

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

From Rao-Cramer-Frechet inequality

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

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

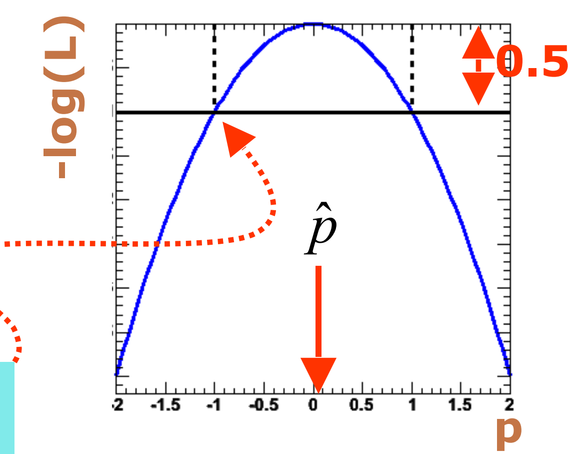
  * I.e. variance is estimated from 2nd derivative of −log(L) at minimum
  * Valid if estimator is efficient and unbiased!

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

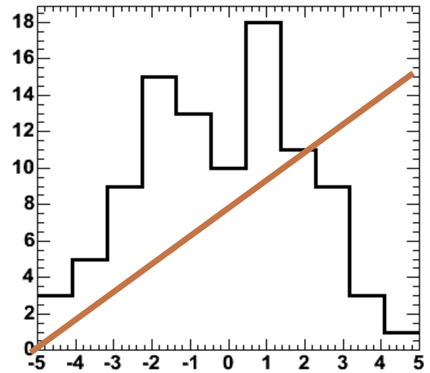
$$\ln L(p) = \ln L(\hat{p}) + \left.\frac{d \ln L}{dp}\right|_{p=\hat{p}} (p-\hat{p}) + \tfrac{1}{2} \left.\frac{d^2 \ln L}{d^2 p}\right|_{p=\hat{p}} (p-\hat{p})^2$$

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

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



Estimating and interpreting Goodness-Of-Fit



'Not good enough'

- Fitting determines best set of parameters of given model to describe data
  - Is 'best' good enough?, i.e.
  - Is it an adequate description, or are there significant and incompatible differences?

- Most common test: the χ^2 test

$$\chi^2 = \sum_i \left(\frac{y_i - f(\vec{x}_i; \vec{p})}{\sigma_i} \right)^2$$

  - If f(x) describes data then $\chi^2 \approx N$, if $\chi^2 >> N$ something is wrong
  - How to quantify meaning of 'large χ^2'?

How to quantify meaning of 'large χ^2'

- Probability distr. for χ^2 is given by

$$\chi^2 = \sum_i \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2 \quad \Longrightarrow \quad p(\chi^2, N) = \frac{2^{-N/2}}{\Gamma(N/2)} \chi^{N-2} e^{-\chi^2/2}$$

- To make judgement on goodness-of-fit, relevant quantity is integral of above:

$$P(\chi^2; N) = \int_{\chi^2}^{\infty} p(\chi^{2\prime}; N) d\chi^{2\prime}$$

- What does χ^2 probability $P(\chi^2,N)$ mean?
  - It is the probability that a function which does genuinely describe the data on N points would give a χ^2 probability as large or larger than the one you already have.
    - Since it is a probability, it is a number in the range [0-1]

Goodness-of-fit – χ^2

- Example for χ^2 probability
  - Suppose you have a function f(x;p) which gives a χ^2 of 20 for 5 points (histogram bins).
  - Not impossible that f(x;p) describes data correctly, just unlikely
  - How unlikely?

$$\int_{20}^{\infty} p(\chi^2, 5)\, d\chi^2 = 0.0012$$

- Note: If function has been fitted to the data
  - Then you need to account for the fact that parameters have been adjusted to describe the data

$$N_{\mathrm{d.o.f.}} = N_{\mathrm{data}} - N_{\mathrm{params}}$$

- Practical tips
  - To calculate the probability in PAW '`call prob(chi2,ndf)`'
  - To calculate the probability in ROOT '`TMath::Prob(chi2,ndf)`'
  - For large N, sqrt($2\chi^2$) has a Gaussian distribution with mean sqrt(2N-1) and σ=1

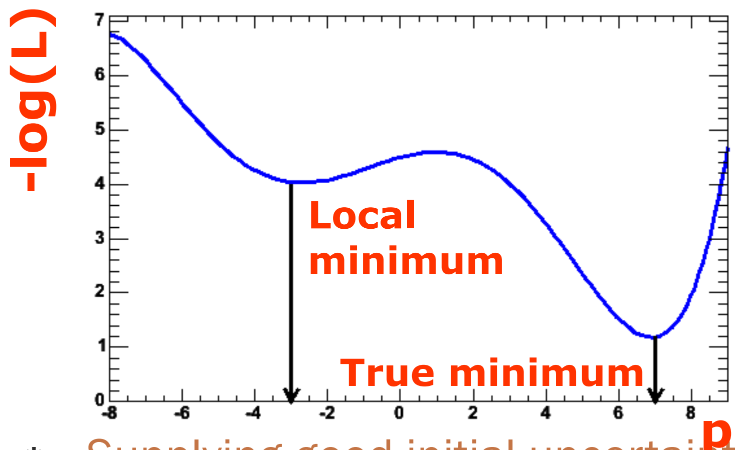
Practical estimation – Numeric χ^2 and -log(L) minimization

- For most data analysis problems minimization of χ^2 or $-\log(L)$ cannot be performed analytically
  - Need to rely on numeric/computational methods
  - In >1 dimension generally a difficult problem!

- But no need to worry – Software exists to solve this problem for you:
  - Function minimization workhorse in HEP many years: MINUIT
  - MINUIT does function minimization and error analysis
  - It is used in the PAW,ROOT fitting interfaces behind the scenes
  - It produces a lot of useful information, that is sometimes overlooked
  - Will look in a bit more detail into MINUIT output and functionality next

Numeric χ^2/-log(L) minimization – Proper starting values

- For all but the most trivial scenarios it is not possible to automatically find reasonable starting values of parameters
  - This may come as a disappointment to some…
  - So you need to supply good starting values for your parameters



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

  - Supplying good initial uncertainties on your parameters helps too
  - Reason: Too large error will result in MINUIT coarsely scanning a wide region of parameter space. It may accidentally find a far away local minimum

Minuit function MIGRAD

* Purpose: find minimum

Progress information, watch for errors here

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

Minuit function MIGRAD

- Purpose: find minimum

```
**********
**   13 **MIGR
**********
(some output o
MIGRAD MINIMIZ
MIGRAD WILL VER                                    RIX.
COVARIANCE MAT      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 minimum

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

```
**********
**   13 **MIGRAD        1000
**********
(some output omitted)
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MA
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 $\frac{d^2L}{dp^2}$

- Purpose: calculate error matrix from

```
**********
**   18 **HESSE        1000
**********
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=257.304 FROM HESSE     STATUS=OK                                   OTAL
                    EDM=2.36534e-06     STRAT                          CURATE
 EXT PARAMETER                                       ERNAL         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
```

Symmetric errors calculated from 2nd derivative of –ln(L) or χ^2

Minuit function HESSE $\frac{d^2L}{dp^2}$

Error matrix (Covariance Matrix) calculated from

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

```
*********
**
****
COV                              UCCESSFULLY
FCN=                             TUS=OK              10 CALLS          42 TOTAL
                                 e-06     STRATEGY= 1       ERROR MATRIX ACCURATE
 EXT                                                INTERNAL        INTERNAL
 NO                                 ERROR           STEP SIZE         VALUE
  1                              3.23861e-01     7.16689e-05    8.84237e-03
  2   sig           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 $\frac{d^2L}{dp^2}$

```
**********
**   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 $\frac{d^2L}{dp^2}$

```
**********
**   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-0
PARAMETER  CORRELAT    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

- Purpose: More rigorous determination of errors
- Warning: Can be very CPU intensive for large number of parameters
- Optional – activated by option "E" in ROOT or PAW

```
**********
**   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)

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

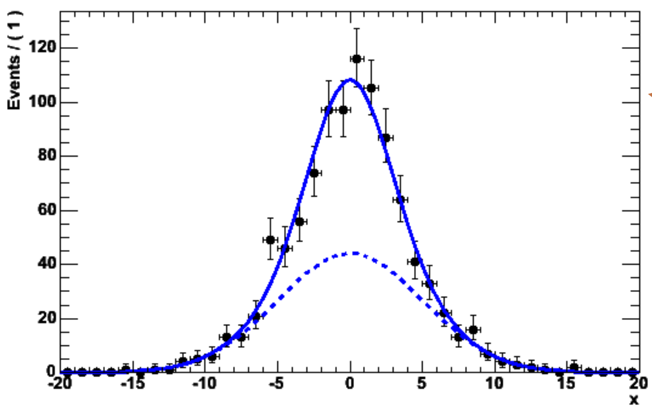
Signs of trouble...

  - 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  CORRELATION COEFFICIENTS
      NO.  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
```

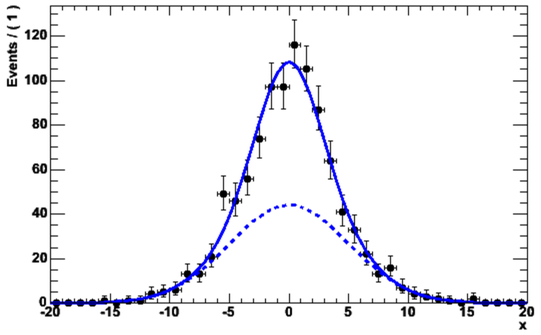
Widths s_1,s_2 strongly correlated fraction f

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



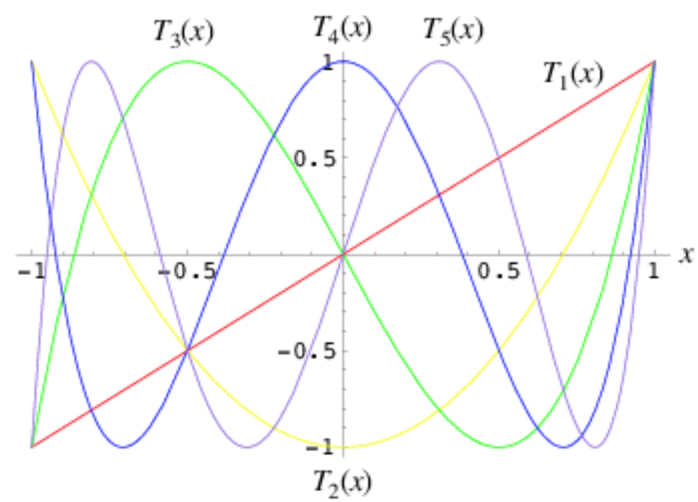
- Correlation of width s2 and fraction f reduced from 0.92 to 0.68
- Choice of parameterization matters!

* 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

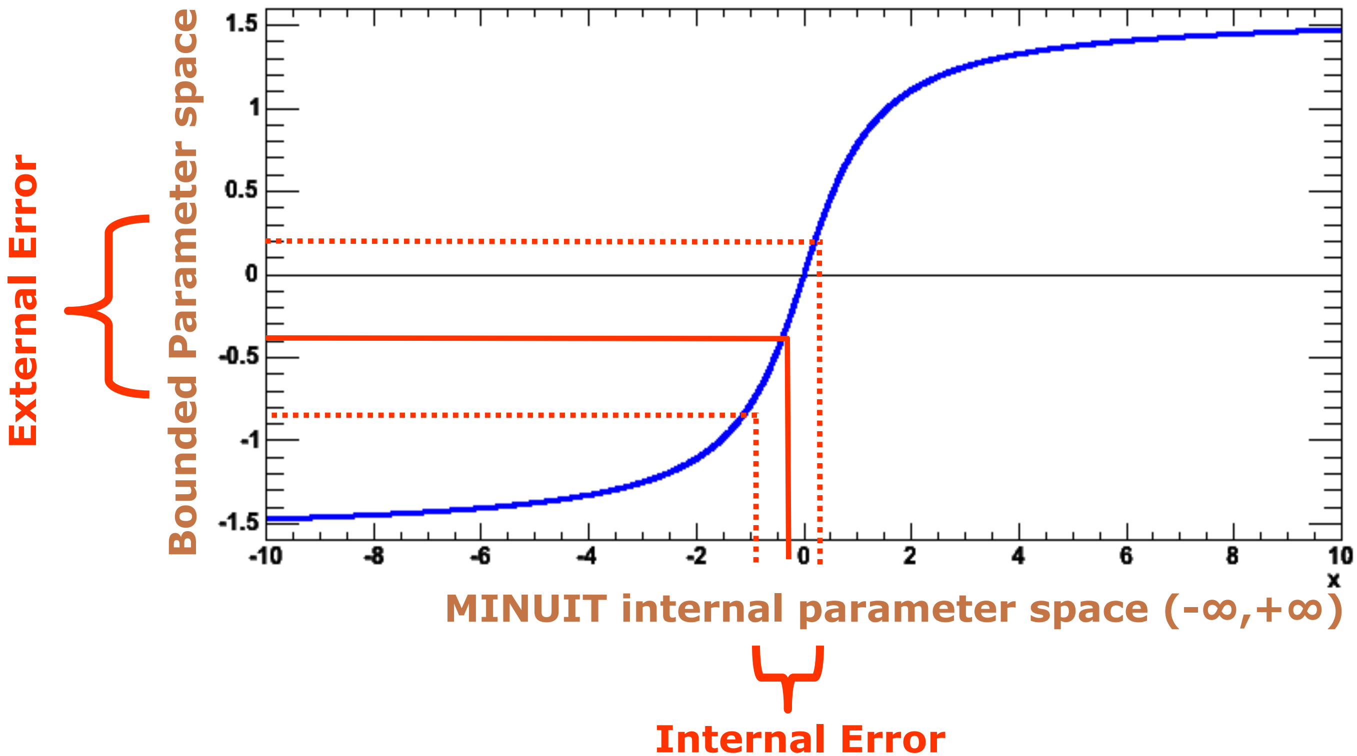
- Warning: Regular parameterization of polynomials $a_0+a_1x+a_2x^2+a_3x^3$ nearly always results in strong correlations between the coefficients a_i.
  - *Fit stability problems, inability to find right solution common at higher orders*
- Solution: Use existing parameterizations of polynomials that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

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



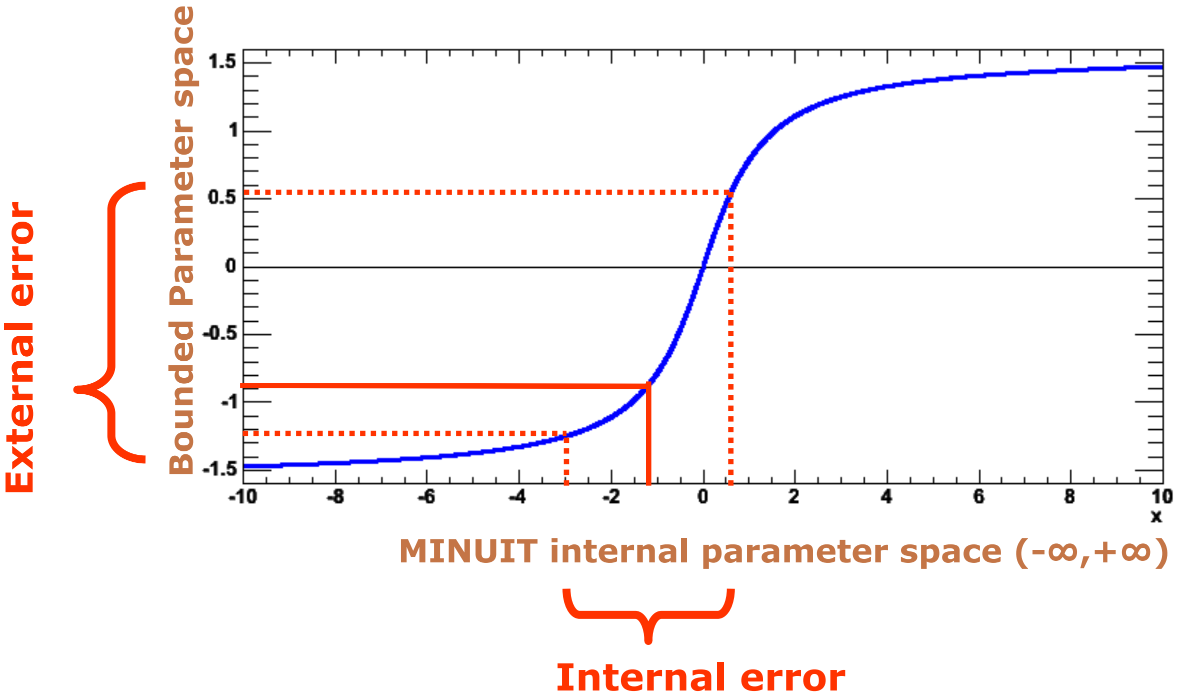
Practical estimation – Bounding fit parameters

* Sometimes is it desirable to bound the allowed range of parameters in a fit
  * Example: a fraction parameter is only defined in the range [0,1]
  * MINUIT option 'B' maps finite range parameter to an internal infinite range using an arcsin(x) transformation:



Practical estimation – Bounding fit parameters

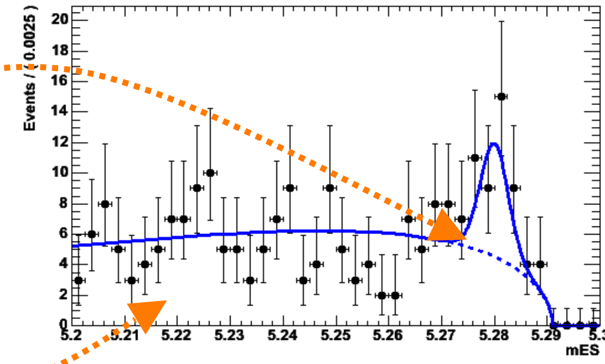
* If fitted parameter values is close to boundary, errors will become asymmetric (and possible incorrect)



* So be careful with bounds!
  * If boundaries are imposed to avoid region of instability, look into other parameterizations that naturally avoid that region
  * If boundaries are imposed to avoid 'unphysical', but statistically valid results, consider not imposing the limit and dealing with the 'unphysical' interpretation in a later stage

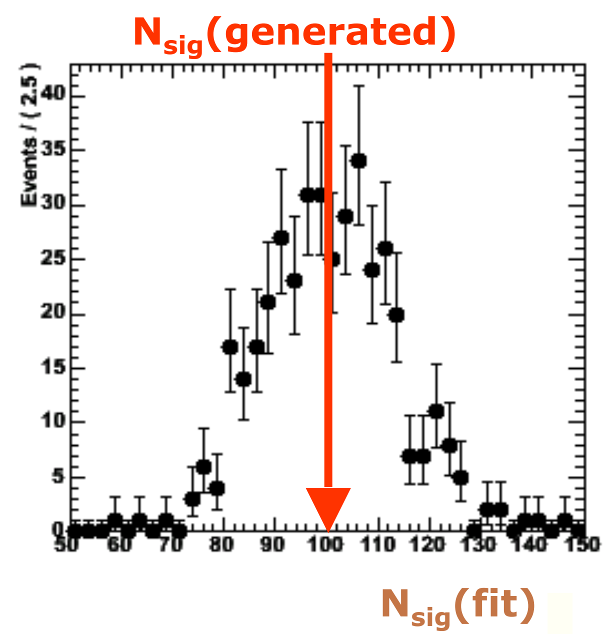
Fit Validation Study – Practical example

- Example fit model in 1-D (B mass)
  - Signal component is Gaussian centered at B mass
  - Background component is Argus function (models phase space near kinematic limit)



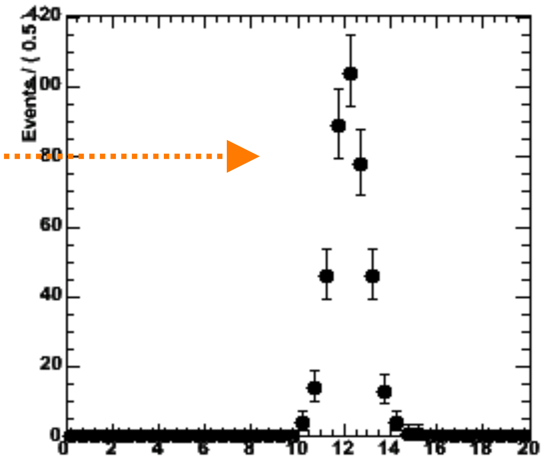
$$F(m; N_{\text{sig}}, N_{\text{bkg}}, \vec{p}_S, \vec{p}_B) = N_{\text{sig}} \cdot G(m; p_S) + N_{\text{bkg}} \cdot A(m; p_B)$$

- Fit parameter under study: N_{sig}
  - Results of simulation study: 1000 experiments with N_{SIG}(gen)=100, N_{BKG}(gen)=200
  - Distribution of N_{sig}(fit)
  - This particular fit looks unbiased…



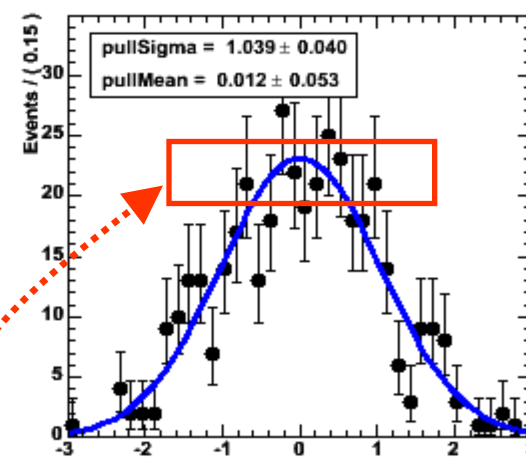
Fit Validation Study – The pull distribution

* What about the validity of the error?
  * Distribution of error from simulated experiments is difficult to interpret…
  * We don't have equivalent of N_{sig}(generated) for the error



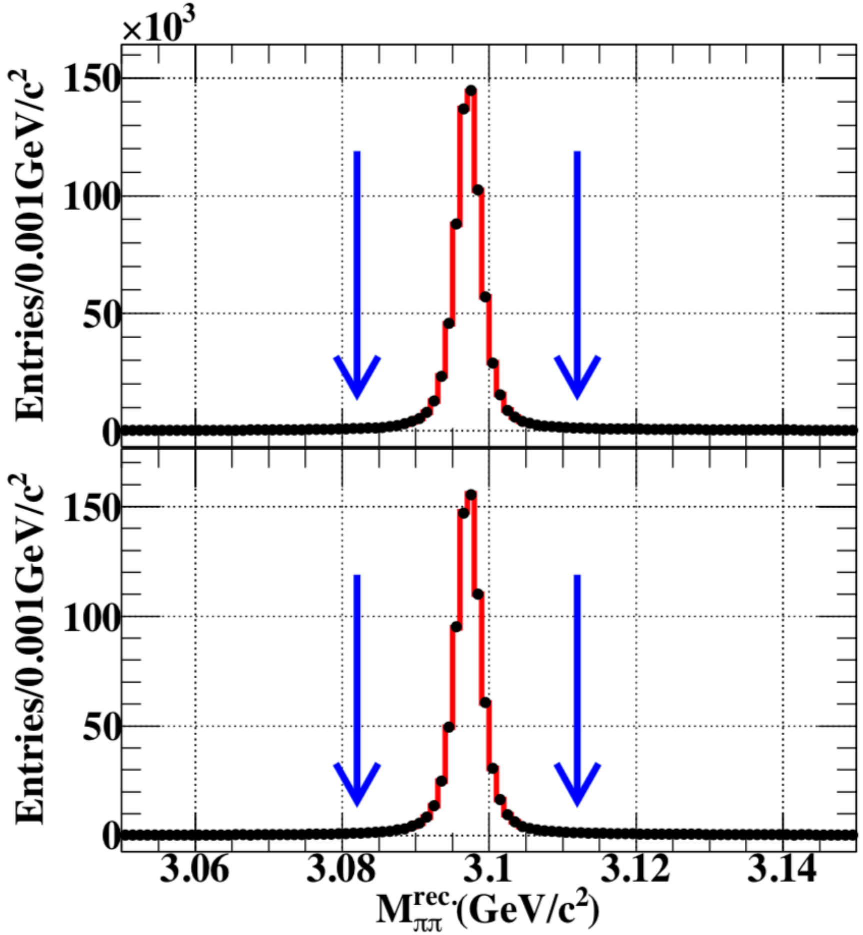
$\sigma(N_{sig})$

* Solution: look at the *pull distribution*
  * Definition: $$\mathrm{pull}(N_{\mathrm{sig}}) = \frac{N_{sig}^{fit} - N_{sig}^{true}}{\sigma_N^{fit}}$$
  * Properties of pull:
    * Mean is 0 if there is no bias
    * Width is 1 if error is correct
  * In this example: no bias, correct error within statistical precision of study



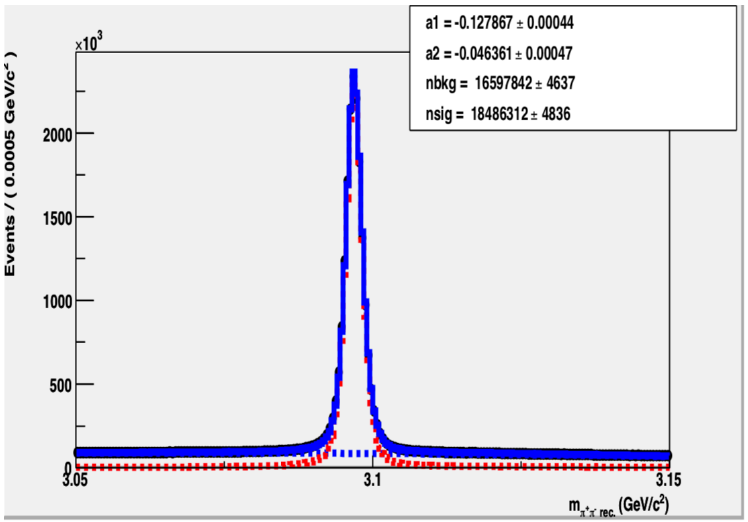
数数是比较幸运的

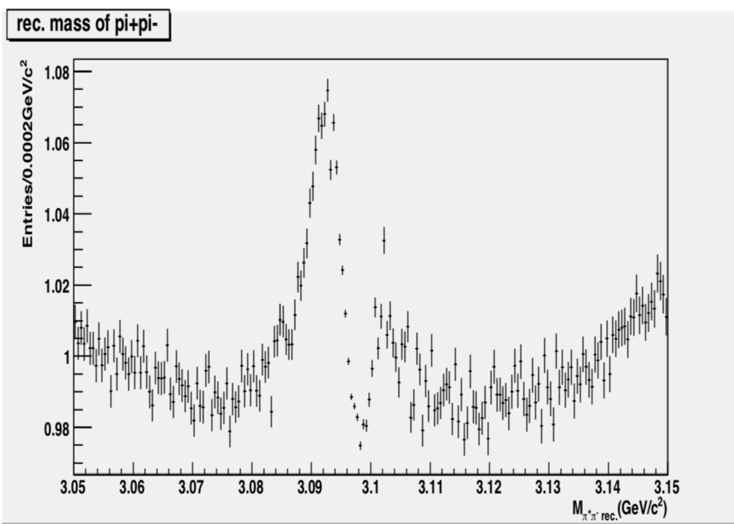
$$N_{sig} = N_{obs} - N_{b1} - N_{b2} - ..$$

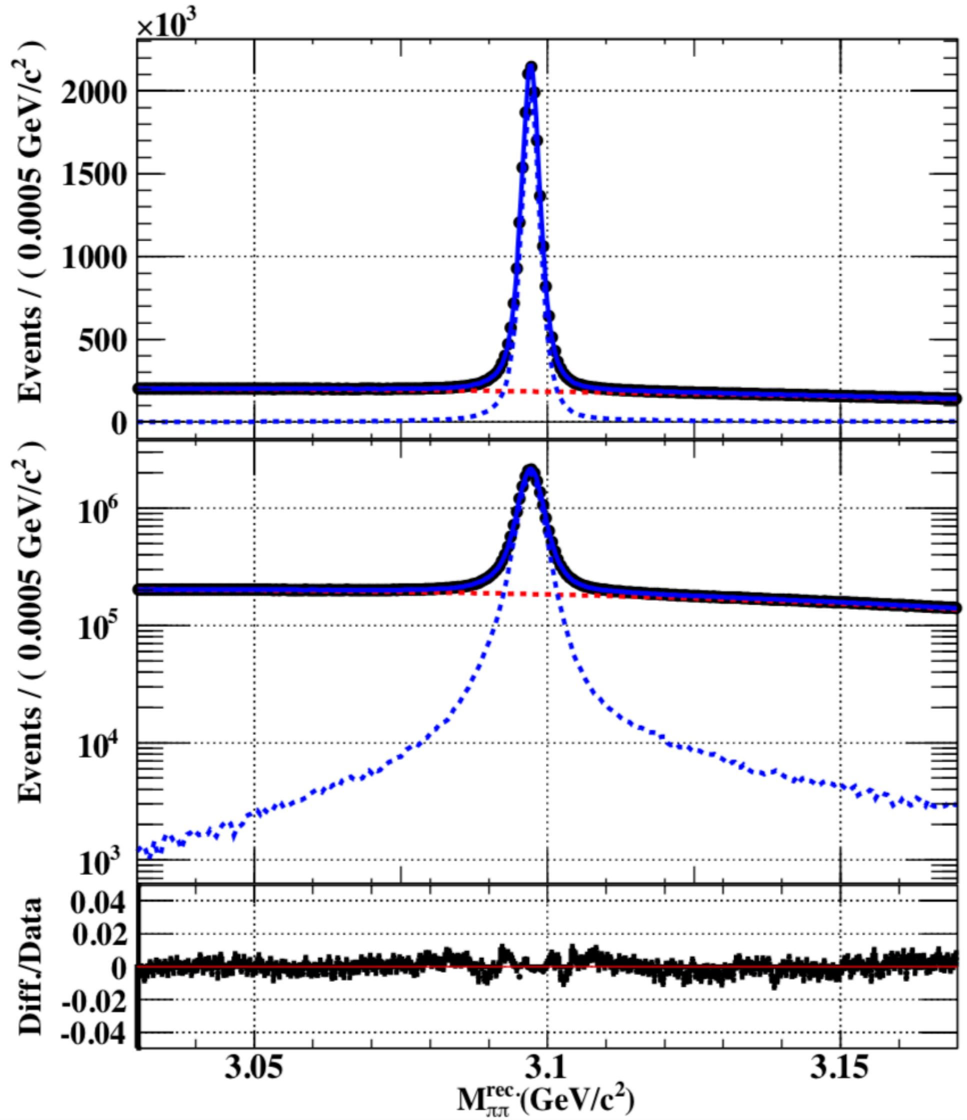


好的 model 并不容易！

用误差来衡量







关于拟合

* 拟合是前期分析结果的一种combination，应该包含分析者对
  * 信号形状的认识：探测器分辨有多大，是否是高斯形式，测量的中心值是否偏移
  * 对本底的了解：来源和大小，是否有未知/不理解的本底
  * 理论模型的知识
  * 多关注如何构建你的模型，这是知识的汇总

* 拟合的好坏：一般用χ^2/NDOF，但这个并不科学
  * χ=(拟合-测量)/误差, χ^2 其实是对所有数据点/bin求和的结果
  * NDOF: number of degree of freedom=数据点/bin数-自由参数个数
* 推荐使用TMath:prob（χ^2,ndof）积分，这个值越大越好。

效率修正

关于效率

* 分支比公式中的效率为事例的选择效率，包括探测器几何接收度、选择条件的效率、探测末态的分支比等。

* 效率的估计一般用Monte Carlo模拟。基本的假设是Monte Carlo可以很好地模拟数据。一般情况下这个假设不成立。

* 应当用实际数据检验Monte Carlo模拟的可靠性，必要时对Monte Carlo模拟效率进行修正。

关于效率

* **一般地**

$$\varepsilon_{DT} = \varepsilon_{DT}^{Geom} \cdot \varepsilon_{DT}^{Track} \cdot \varepsilon_{DT}^{Ptid} \cdot \varepsilon_{DT}^{\gamma\ ID} \cdots$$

$$\varepsilon_{MC} = \varepsilon_{MC}^{Geom} \cdot \varepsilon_{MC}^{Track} \cdot \varepsilon_{MC}^{Ptid} \cdot \varepsilon_{MC}^{\gamma\ ID} \cdots$$

$$\varepsilon_{DT}^{i} = \varepsilon_{MC}^{i} \cdot f^{i}$$

测量“f^i”，即可修正Monte Carlo效率得到真实数据对应的效率。这是物理分析的重要内容之一。

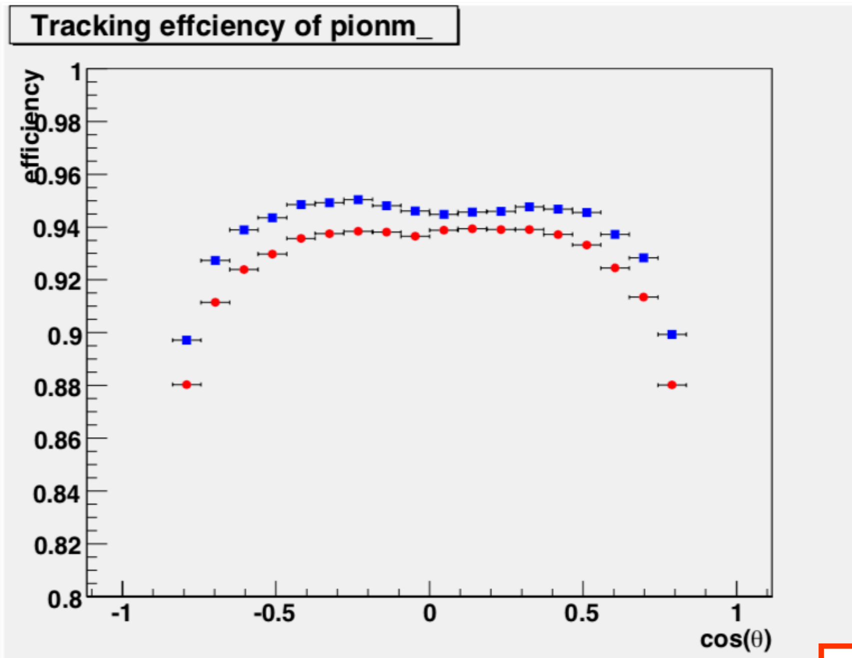
关于效率估计的说明

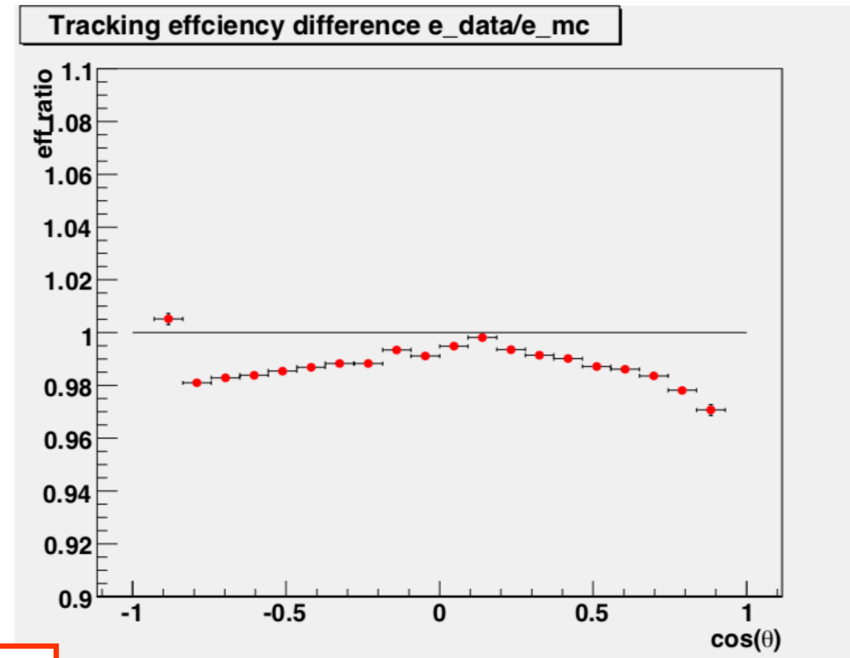
* 数据与Monte Carlo的不一致永远存在，因此利用Monte Carlo得到的效率永远需要修正，但对于不同的过程要求也不同。
* 对于小统计量情况，统计误差为主，数据与模拟的差别对结果影响很小，但对本底估计时需特别关注。
* 对于大统计量情形，期望高精度结果，数据与模拟的差别不容忽视。
* 对于稀有事例的寻找，本底情况的模拟可能会与真实情况有很大差别。
    * 各种 rare decay 的寻找
    * $J/\psi \to e\mu$的寻找，Bhabha和dimu本底（$e \to \mu$，$\mu \to e$误判）

两种方式：修正和抵消

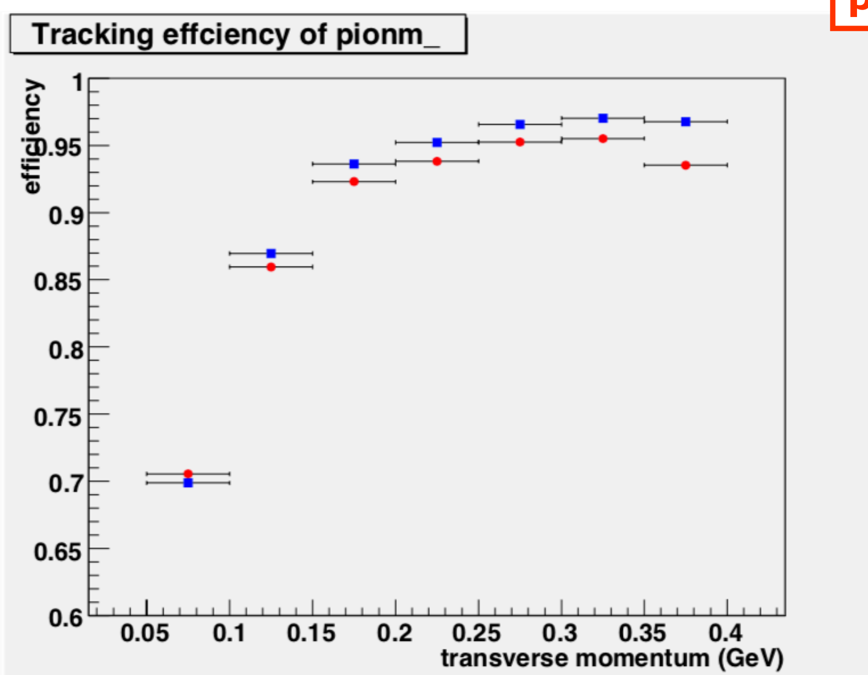
$$B_{\pi\pi J/\psi} = \frac{N_{\pi\pi J/\psi}}{\epsilon_{MC} N^{tot}} \qquad \epsilon_{MC} = \epsilon_{\pi^+}\epsilon_{\pi^-}\epsilon_{m_{\pi\pi}}\ldots$$

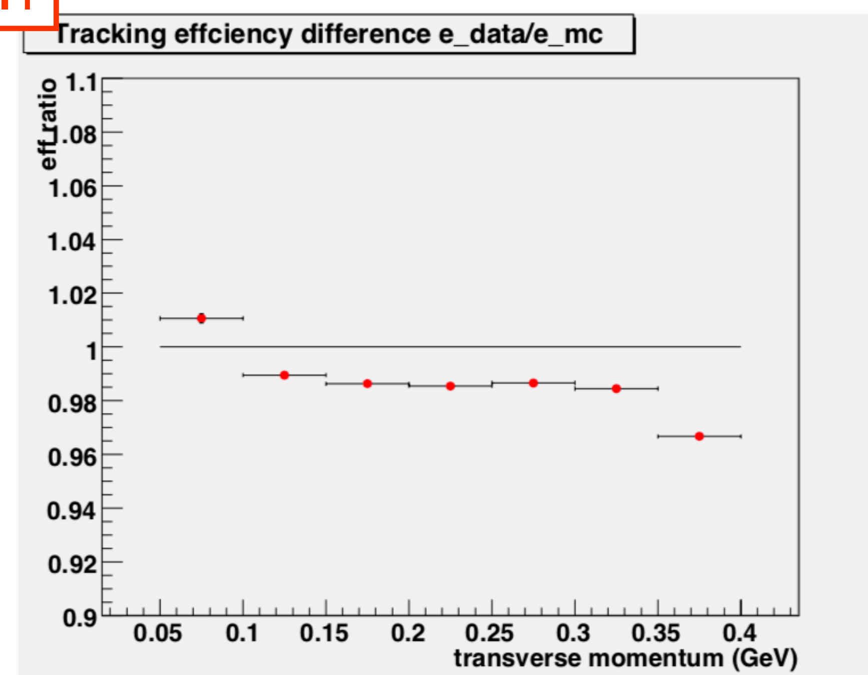
- 这里主要是修正
  - 朱凯老师做 tracking 效率研究
  - 给出 data vs mc of cos theta and pt 的二维分布
  - 在分析中用 data/MC 的 ratio 对 MC 做了 reweight （修正）
  - 精度控制在 0.1%


Tracking effciency of pionm_
efficiency
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0.82
0.8
-1
-0.5
0
0.5
1
cos(θ)

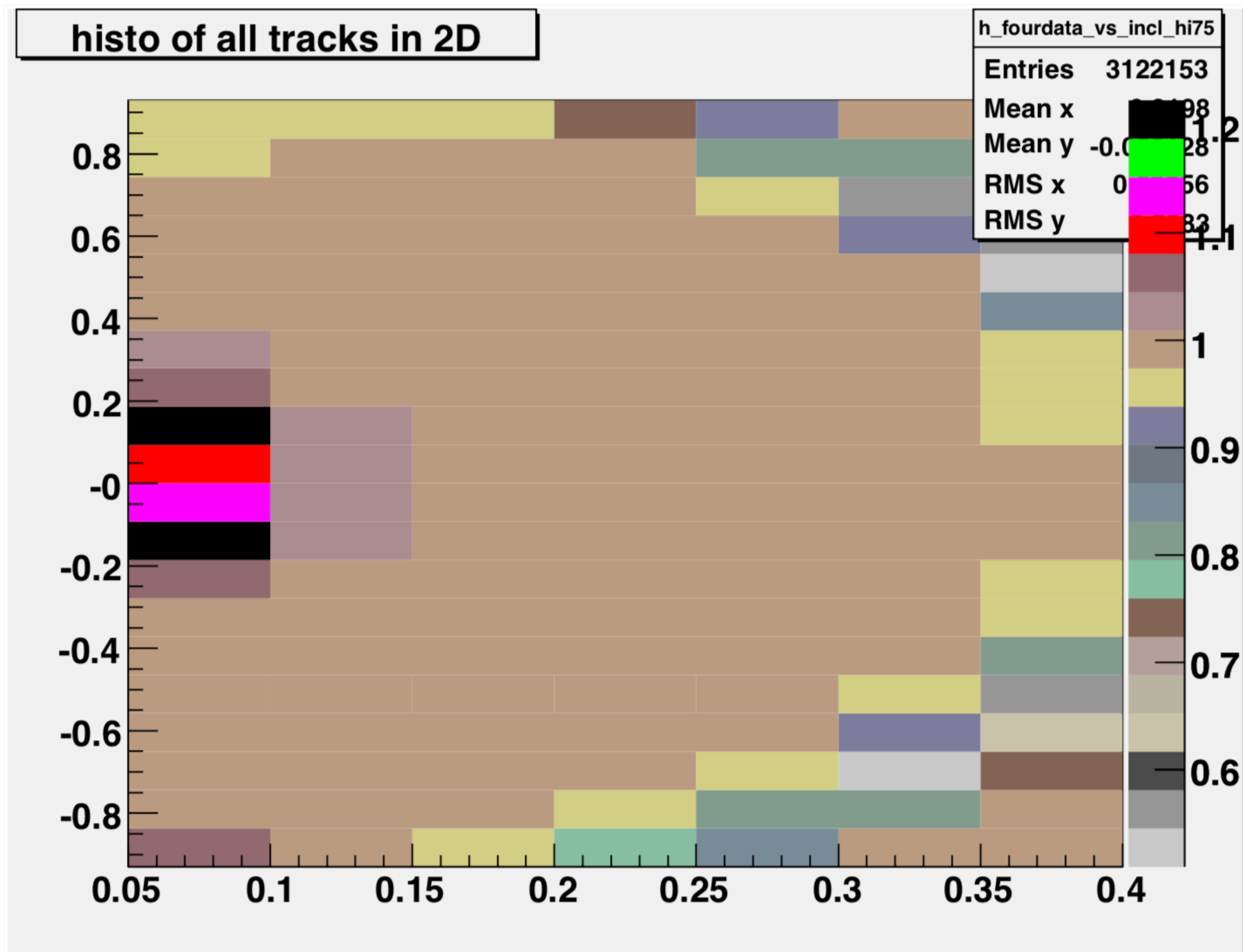

Tracking effciency difference e_data/e_mc
eff_ratio
1.1
1.08
1.06
1.04
1.02
1
0.98
0.96
0.94
0.92
0.9
-1
-0.5
0
0.5
1
cos(θ)

pi+


Tracking effciency of pionm_
efficiency
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
transverse momentum (GeV)


Tracking effciency difference e_data/e_mc
eff_ratio
1.1
1.08
1.06
1.04
1.02
1
0.98
0.96
0.94
0.92
0.9
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
transverse momentum (GeV)

Data vs mc of cos theta and pt



$M_{\pi\pi}$ specctrum

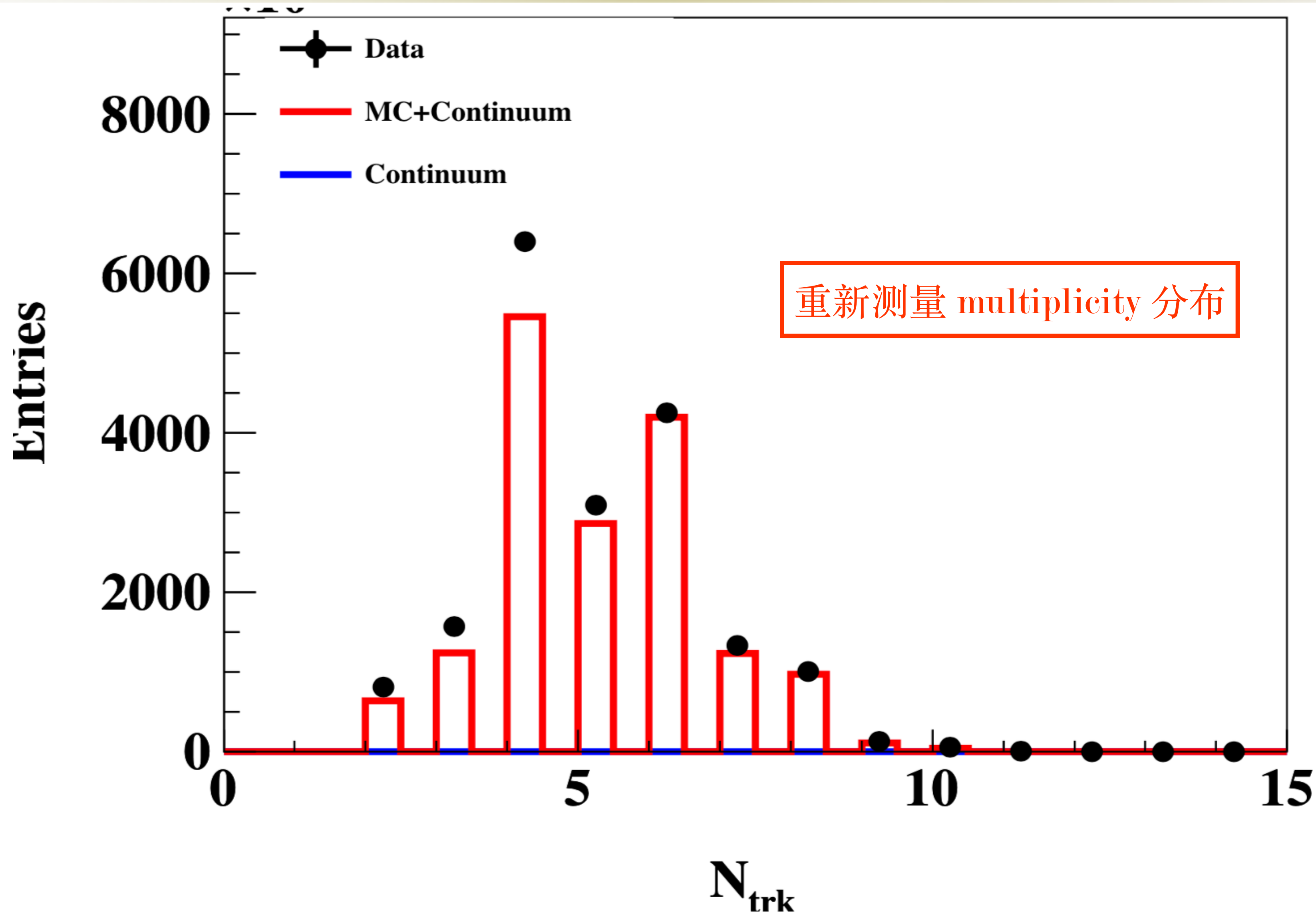
- MC 采用了 JPIPI，g 和 g1 来自 BESI 的测量：

$$g(s-2m_{\pi\pi}^2)+2g_1E_{\pi^+}E_{\pi^-}\epsilon_{\psi'}^*(M)\cdot\epsilon_\psi(\Lambda)$$

- 最终采用数据（轻子道的干净样本）来 reweight MC $M_{\pi\pi}$ 分布
- 高统计量的时候会出现新的问题
- 探测器里出现 2 tracks、4tracks，6tracks, ...时，每种情形的每根 track 的重建效率并不同相同：track 越多，效率越低，...
- 然鹅 ...

prongs of J/ψ	0	2	4	6	8	10	Total
number of events(n_i or n)	607255	11885458	14889294	6462789	688523	11853	34549656 [a]
w_k from EvtGen	0.0175	0.3440	0.4310	0.1871	0.0199	0.0003	53.505%
w_k from Mehtod-I	0.0225	0.3899	0.4024	0.1667	0.0169	0.0015	53.549%
w_k from Mehtod-II	0.0230	0.3960	0.4010	0.1624	0.0164	0.0013	53.574%
ϵ_k (MC)	55.98%	55.02%	53.20%	51.20%	48.61%	44.70%	

数据、MC 的 multiplicity 分布符合很差

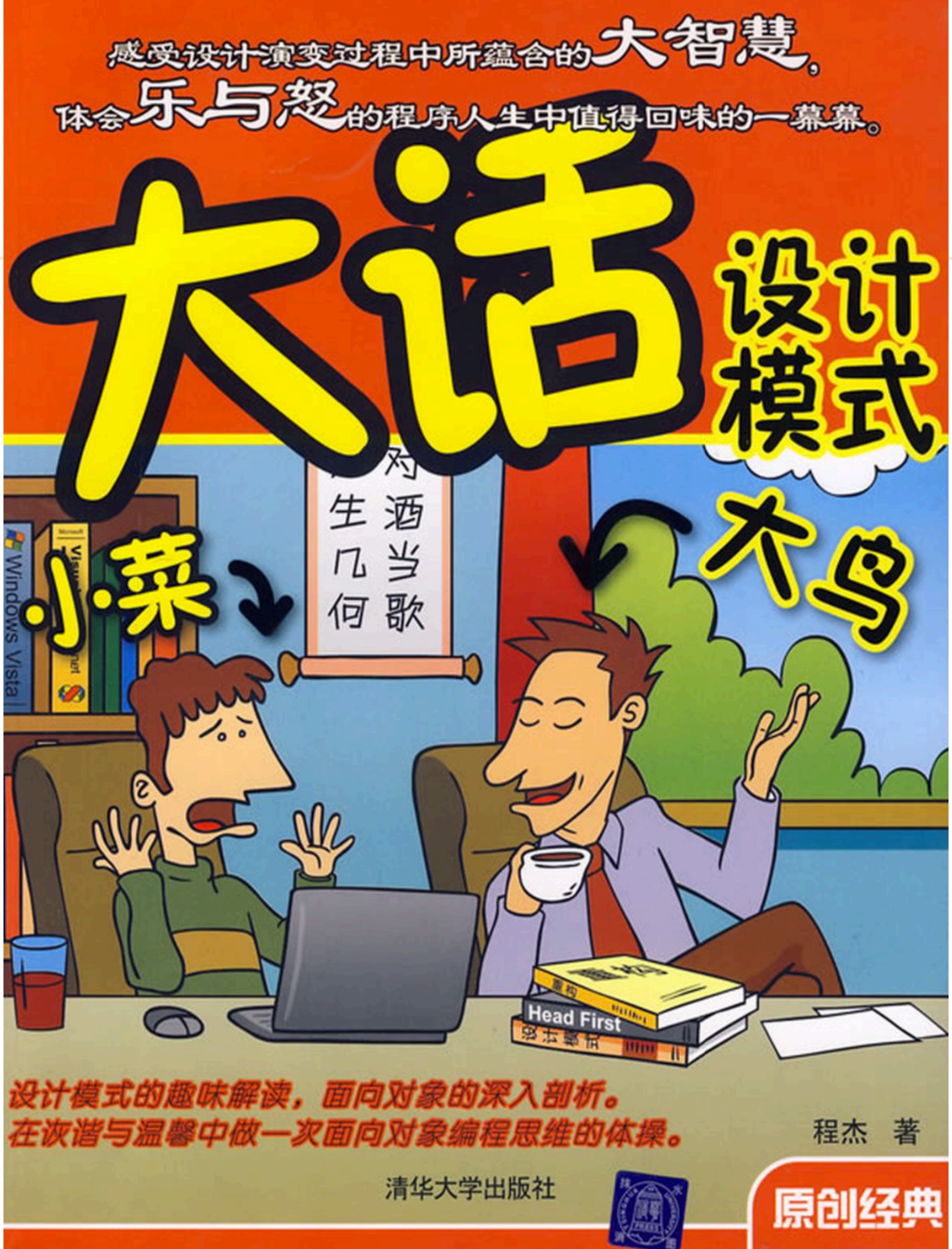


解方程

$$\begin{pmatrix} 884138.2 \\ 1691653.1 \\ 6947052.0 \\ 3295541.7 \\ 4557824.2 \\ 1395702.8 \\ 1059873.7 \\ 134480.3 \\ 55084.1 \\ 4179.1 \\ 835.5 \end{pmatrix} = \begin{pmatrix} 52.8927 & 2.8744 & 0.1369 & 0.0179 & 0.0022 & 0.0000 \\ 1.7753 & 9.5437 & 1.2390 & 0.1800 & 0.0247 & 0.0291 \\ 0.9905 & 41.1825 & 6.3557 & 1.0455 & 0.2037 & 0.0571 \\ 0.0434 & 0.9993 & 18.4289 & 4.1838 & 0.9895 & 0.5728 \\ 0.0104 & 0.3829 & 25.8516 & 11.2145 & 3.4419 & 1.8181 \\ 0.0006 & 0.0148 & 0.9357 & 18.6851 & 8.4610 & 4.7885 \\ 0.0003 & 0.0026 & 0.2430 & 15.2442 & 13.9761 & 9.5418 \\ 0.0001 & 0.0001 & 0.0090 & 0.5556 & 14.2377 & 12.1279 \\ 0.0000 & 0.0000 & 0.0010 & 0.0810 & 6.9984 & 10.1813 \\ 0.0000 & 0.0000 & 0.0001 & 0.0023 & 0.2583 & 4.4639 \\ 0.0000 & 0.0000 & 0.0000 & 0.0003 & 0.0230 & 1.0272 \end{pmatrix} \begin{pmatrix} w_0 \\ w_2 \\ w_4 \\ w_6 \\ w_8 \\ w_{10} \end{pmatrix}$$

prongs of J/ψ	0	2	4	6	8	10	Total
number of events(n_i or n)	607255	11885458	14889294	6462789	688523	11853	34549656 [a]
w_k from EvtGen	0.0175	0.3440	0.4310	0.1871	0.0199	0.0003	53.505%
w_k from Mehtod-I	0.0225	0.3899	0.4024	0.1667	0.0169	0.0015	53.549%
w_k from Mehtod-II	0.0230	0.3960	0.4010	0.1624	0.0164	0.0013	53.574%
ϵ_k (MC)	55.98%	55.02%	53.20%	51.20%	48.61%	44.70%	

修正是笨办法，应该从方法设计出发



测量 B（J/psi->l+l-）

- 常规方法： $B(J/\psi \to l^+l^-) = \dfrac{N_{obs}^{\pi\pi ll}}{\epsilon_{\pi\pi ll} N_{tot}}$
- $\epsilon_{\pi\pi ll} = \epsilon_{\pi\pi}\epsilon_{ll}$
- 那么 $B_{ll} = \dfrac{N_{\pi\pi ll}}{\epsilon_{\pi\pi ll} N_{tot}} / \dfrac{N_{\pi\pi}}{\epsilon_{\pi\pi} N_{tot}} = \dfrac{N_{\pi\pi ll}}{N_{\pi\pi}} \dfrac{\epsilon_{\pi\pi}}{\epsilon_{\pi\pi ll}}$
- So
  - 无需总数
  - 原则上 pi 的 tracking 可以 cancel
  - 只剩下高动量的轻子需要仔细考虑，而高动量的比较容易处理

Sources	$\pi^+\pi^- J/\psi$	e^+e^-	$\mu^+\mu^-$
Tracking	0.80	0.20	0.20
Multiplicity of J/ψ	0.20	0.20	0.20
$M_{\pi^+\pi^-}$ distribution	0.35	0.01	0.01
Background Shape	0.03	0.03	0.04
Fit/Count Range	0.06	0.14	0.14
Bin Size	0.06	0.06	0.06
E/p	—	0.18	0.09
$\cos\theta_{\pi^+\pi^-}$	0.13	0.07	0.07
$\cos\theta_{l^+l^-}$	—	0.04	0.05
FSR effect of l^+l^-	—	0.10	0.23
Fit Method	0.37	0.37	0.37
Trigger	0.10	0.30	0.30
Number of $\psi(3686)$	0.81	—	—
Sum in quadrature	1.28	0.62	0.63

多说一点

- CEPC、FCC 上 处采用$e^+e^- \to ZH \to \mu^+\mu^-H$
- 模型无关：不用看 Higgs 的衰变，也就不依赖于 Higgs 的衰变机制
- 绝对测量：产生的 Higgs 数是可以 taged
- BESIII 上的 charm 介子对产生，Lambda_c 都是

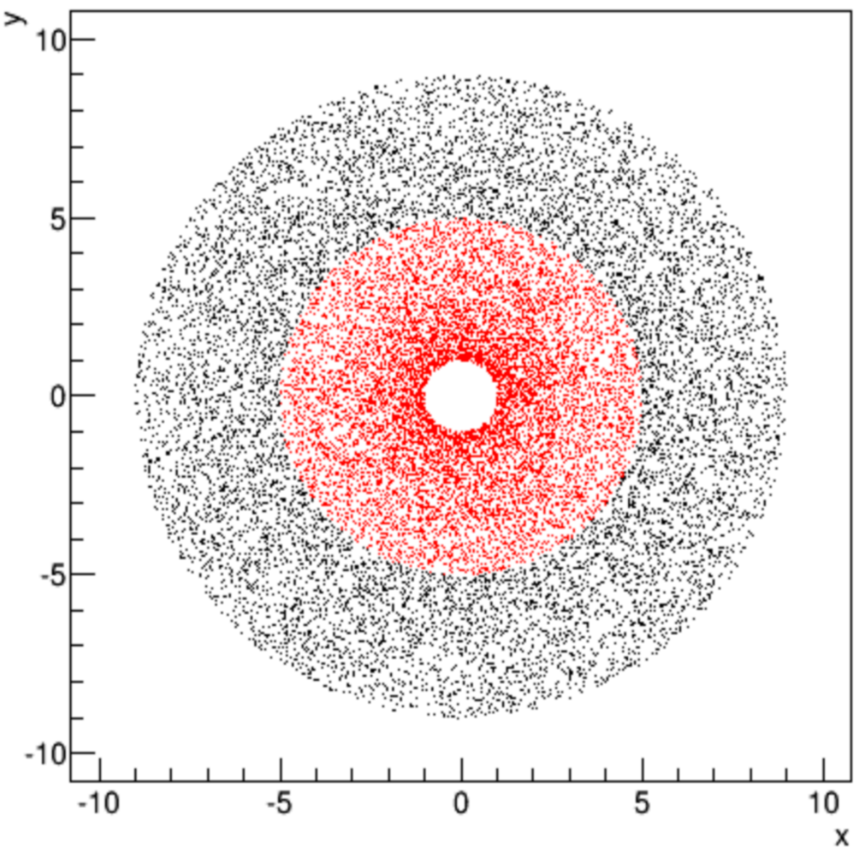
机器学习在数据分析中的应用

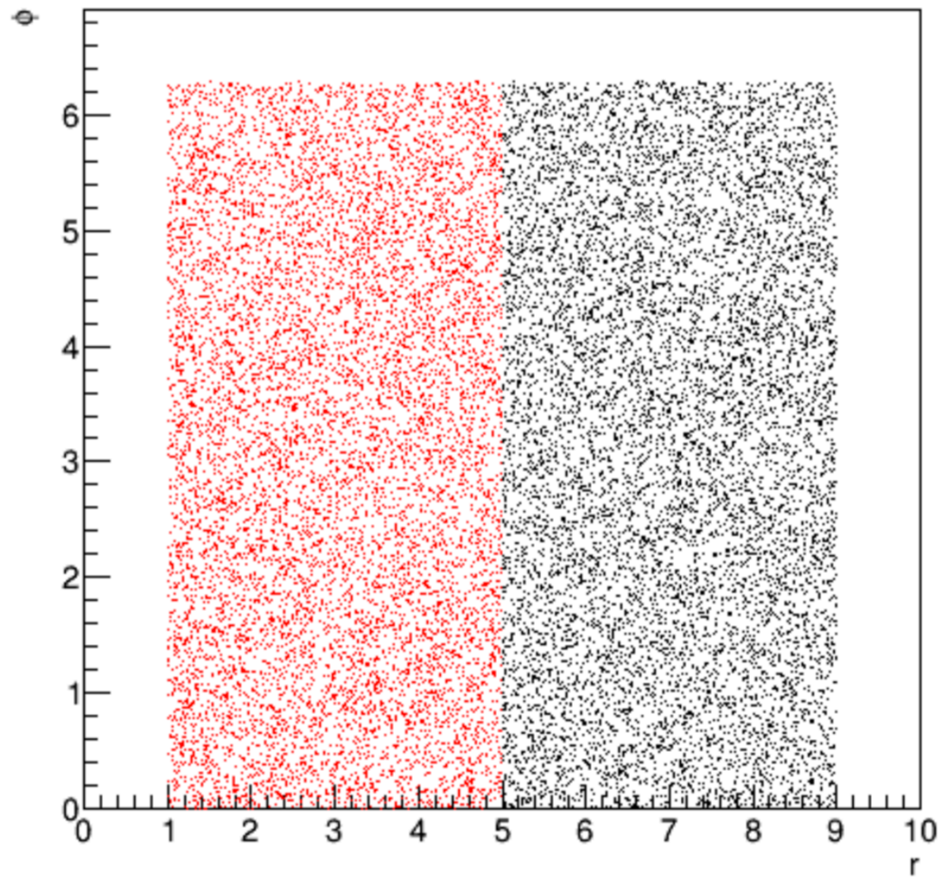
什么是机器学习?

* 传统的算法都需要编程实现
* 机器学习的不同是：不需要编程，或者只需要极少代码，可以通过数据学习（训练）“经验”并应用到新的数据上（测试、应用）。

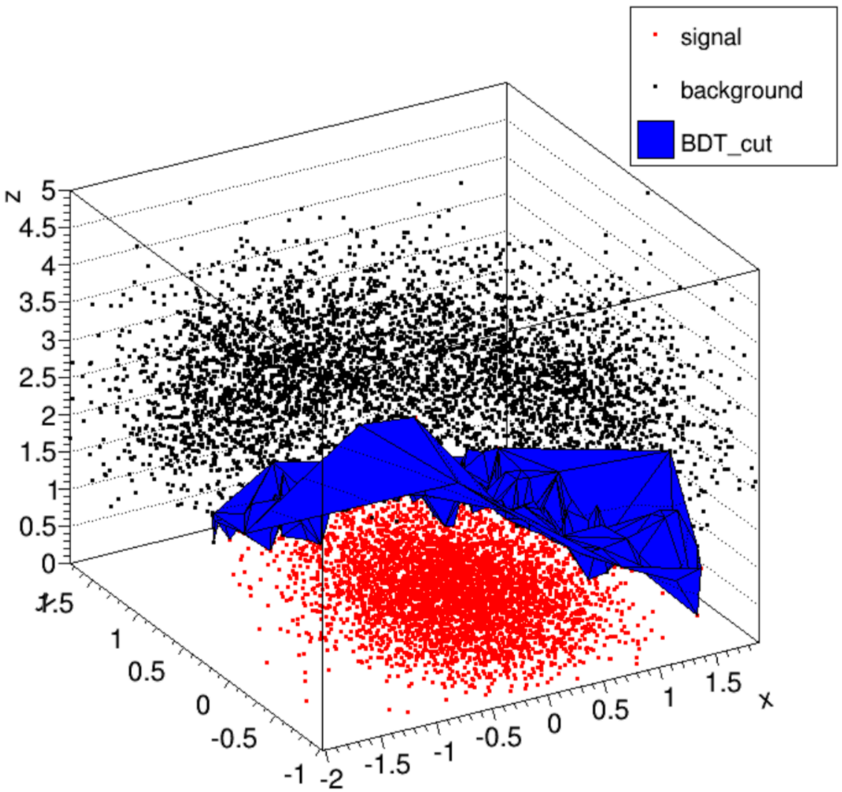
* 高能物理的机器学习叫“MVA”，采用的软件工具叫做“TMVA”，包含了大部分传统机器学习算法：BDT，神经网络 ...
* 深度学习是机器学习的一部分：特征是网络模型的层数和复杂度大幅增加 ...
* 写代码是一种高级脑力劳动，如果能避免或者减少，将会大大提高效率和效果 ...

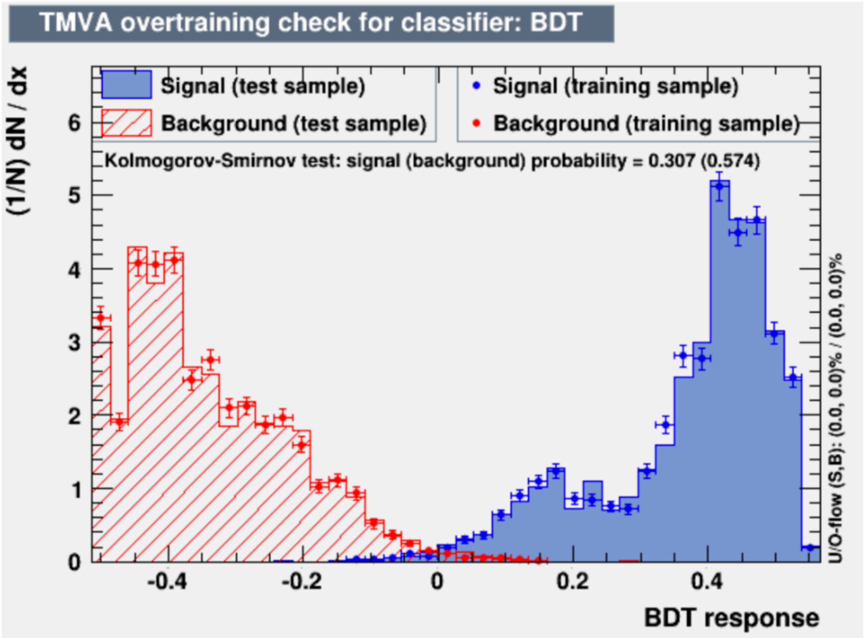
示例：简单分类问题
线性 到 非线性





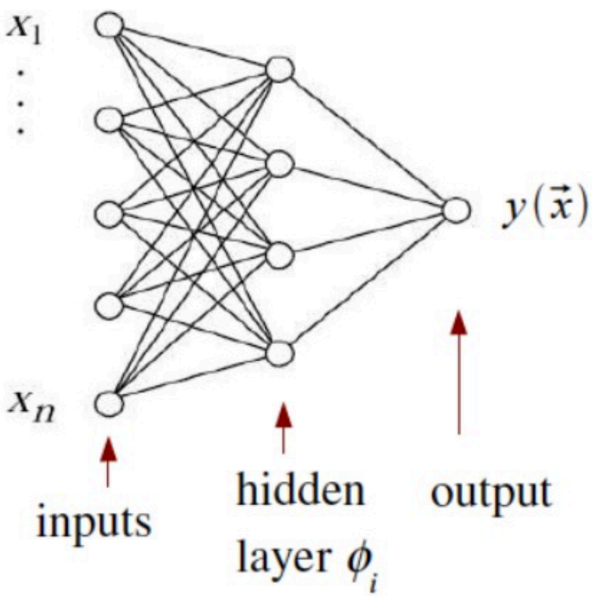
示例：分类问题


signal
background
BDT_cut
z
x


TMVA overtraining check for classifier: BDT
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.307 (0.574)
(1/N) dN / dx
BDT response
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

人工神经网络
Multilayer perceptron , MLP

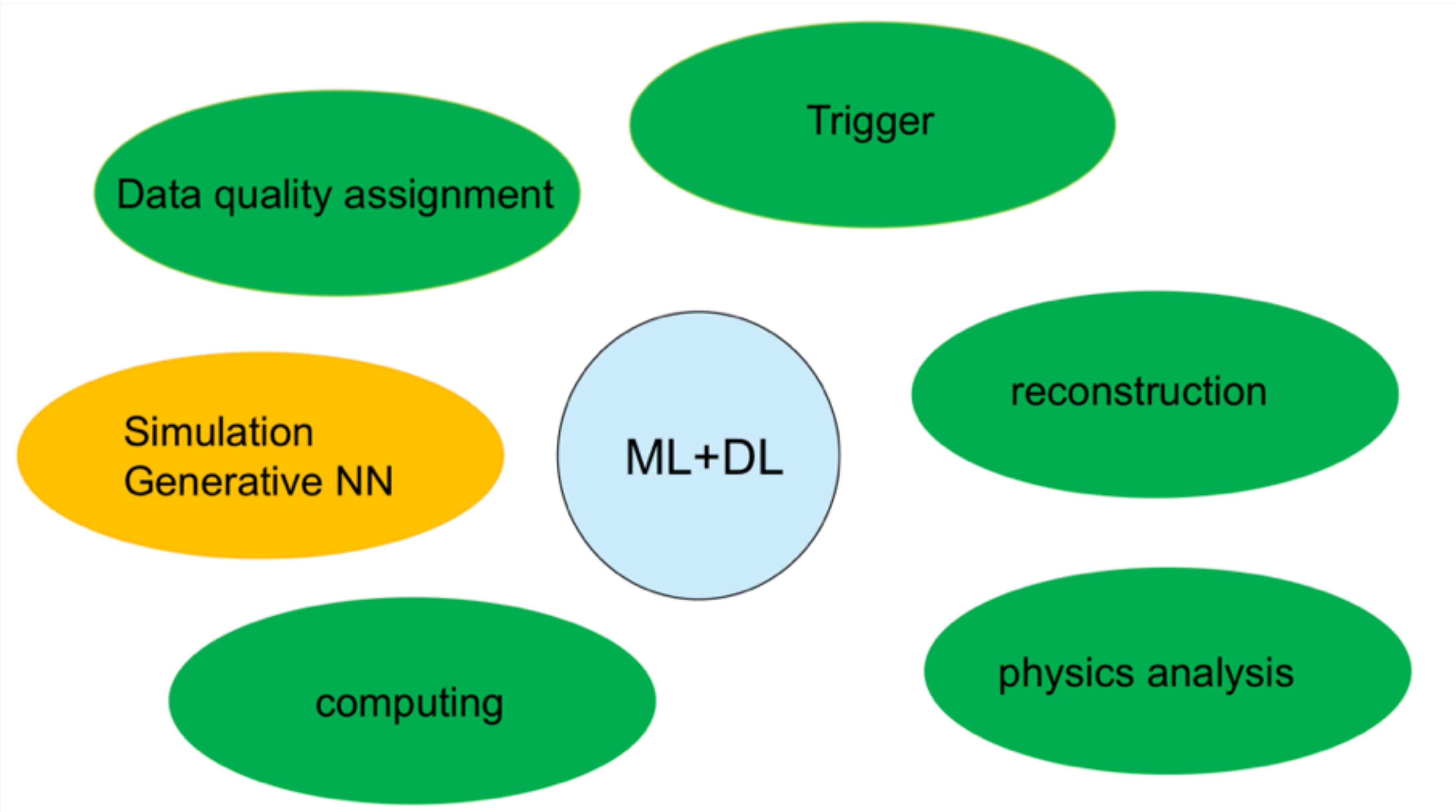
- 由大量神经元相互连接而成的复杂系统
- 全互联型神经网络:每一个神经元都可以与其它所有神经元发生相互作用
- 阶层型神经网络:每一个神经元只能与相邻层 的神经元发生相互作用，而与本层的神经元不 发生信息传递



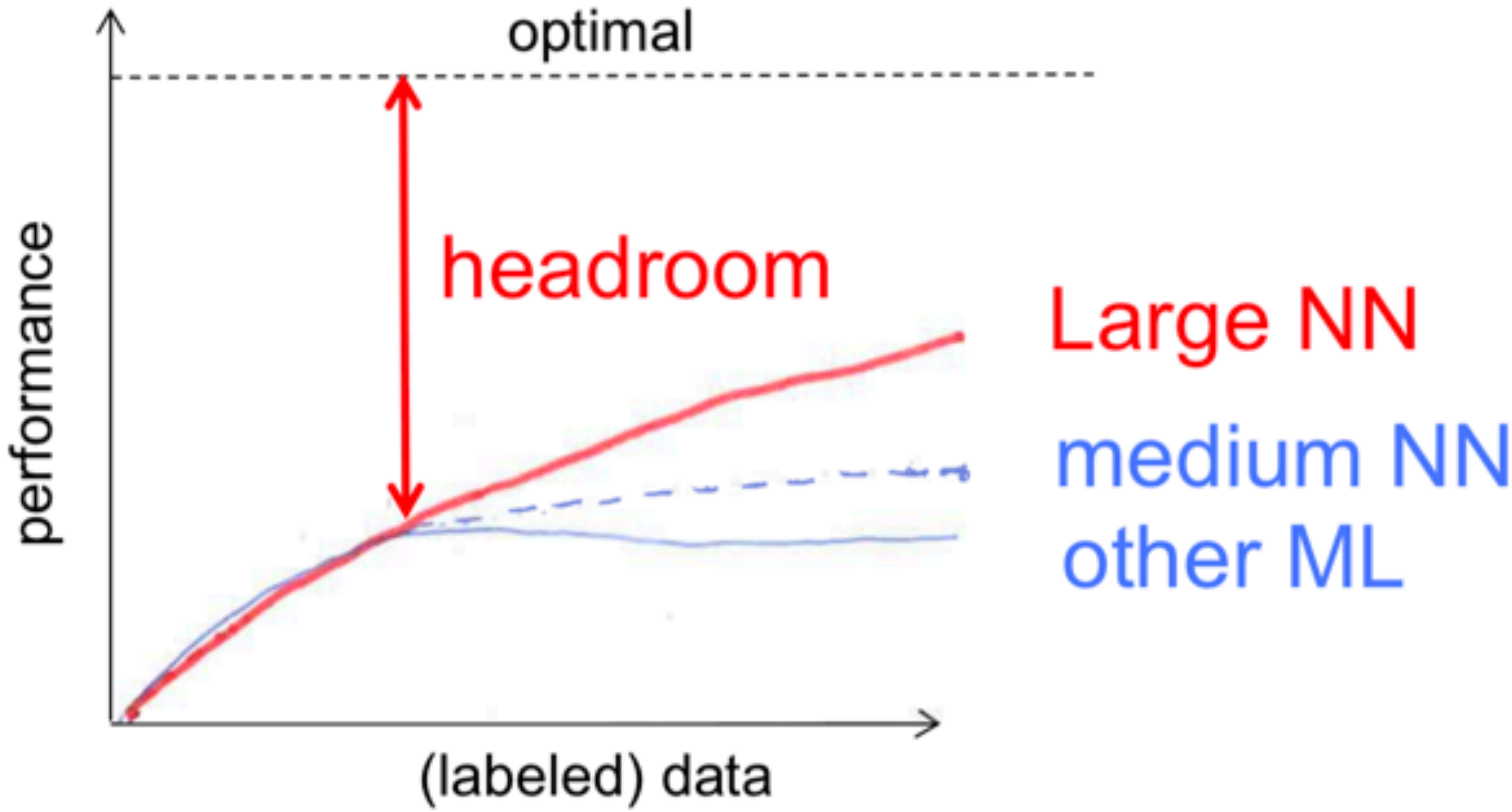
$$\varphi_i(\vec{x})=h\left(w_{i0}^{(1)}+\sum_{j=1}^{n} w_{ij}^{(1)} x_j\right)$$

$$y(\vec{x})=h\left(w_{10}^{(2)}+\sum_{j=1}^{n} w_{1j}^{(2)} \varphi_j(\vec{x})\right)$$

高能物理实验中有大量应用场景



深度学习的潜力？



由于模拟技术的广泛应用，带标签的数据在高能物理中非常易于获得！

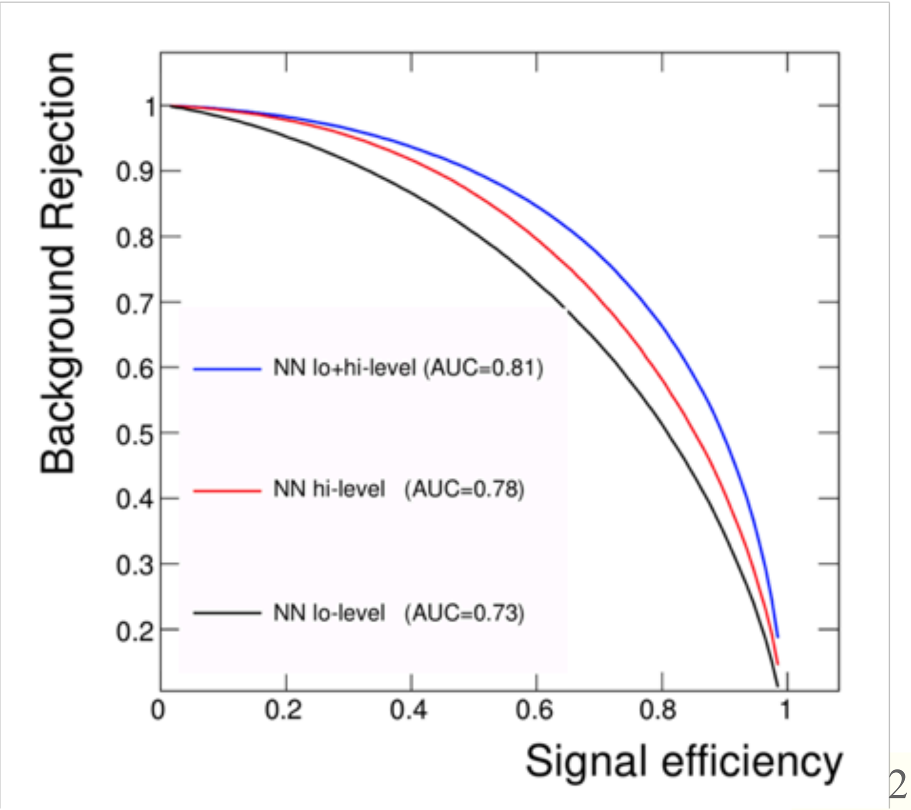
DL 在高能物理中的首次应用

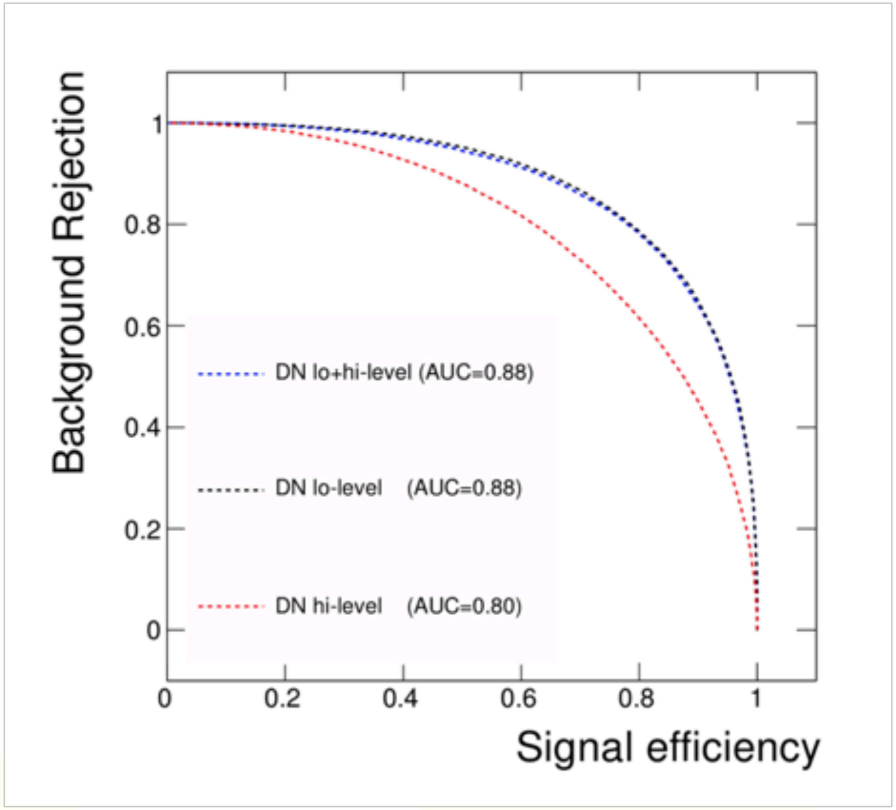
Searching for Exotic Particles in High-Energy Physics with Deep Learning

Pierre Baldi, Peter Sadowski (UC, Irvine) , Daniel Whiteson (UC, Irvine & Pennsylvania U. & UC, Irvine)

Feb 19, 2014 - 9 pages

Nature Commun. 5 (2014) 4308
DOI: 10.1038/ncomms5308
e-Print: arXiv:1402.4735 [hep-ph] | PDF





机器学习小结

- 线性判别方法:利用样本的线性函数作为样本类别的判别函数，方法简单，容易实现，计算量和数据存储量小;对于线性不可分样本，计算复杂，错分率较大。
- 人工神经网络:对非线性复杂关联数据具有很强判别能力，基本思想简单;设计、 训练复杂费时，计算量和存储量很大，需要有足够统计量的训练样本集
- 决策树方法:具有物理直观性，程序设计和调试简单，计算速度快;当信号和本底样本相互重叠或数据存在非线性关联时，判选效率下降。与人工神经网络相比，决策树方法计算量相对小，对训练样本量要求不是特别大;理论最优性能略逊
- Toolkit for multivariate data analysis (TMVA) 是一个多元统计分析的工具性程序包
- 大型高能物理实验是典型的复杂大系统的科学研究工作，原始数据集合样本数量巨大，利用MVA方法对多维复杂数据集合进行分析。
- 深度学习具有更大的参数空间，可以描述更复杂的问题，具有更大的潜力和前景，但是计算量比较大、学习曲线比较缓。

总结

- ☑数据分析是个充满设计感的活动
- ☑好的物理知识背景和丰富的经验、想象力会让你对目标、过程和结果有比较好的把握
- ☑测量的本质是大量信息的输入和一点点关键信息获得
- ☑测量的整个过程别忘了不定期的重估一下对精度的预期
- ☑多学习概率统计知识和工具
- ☑技术让生活更轻松