

BESIII数据处理和误差分析

李刚

ligang@ihep.ac.cn

科创计划培训

2026年7月

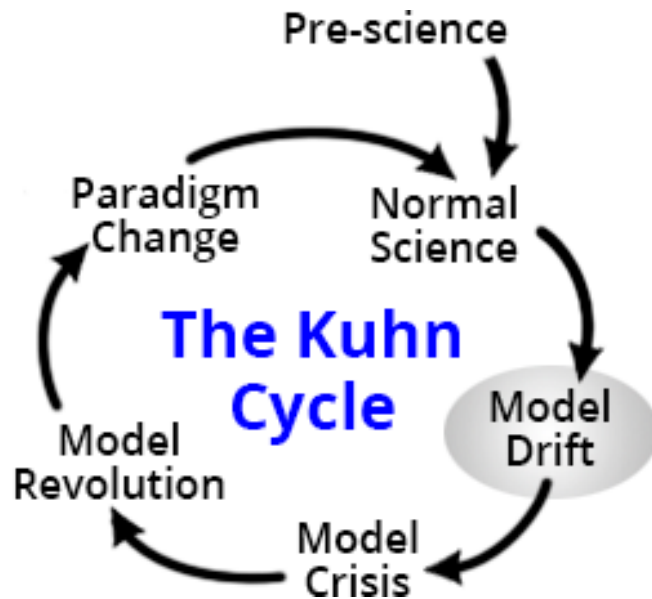
提纲

- ◆ 粒子物理实验
- ◆ 数据分析和测量
 - ◆ 误差分析始于数据分析开始之前
 - ◆ 数据分析优化
- ◆ 误差分析
 - 统计误差和信息
 - 系统误差和修正
- ◆ 举例

为什么研究（粒子）物理？

寻找新动力学 — New physics

* 科学发展：常规科学→迁移→危机→革命→范式改变→新的常规科学



- ✓ 模型迁移
- ✓ 模型危机
- ✓ 模型革命
- ✓ 范式改变

粒子物理有多个反常

多个反常...

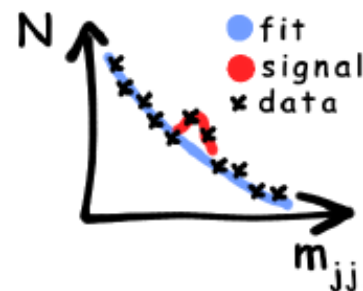
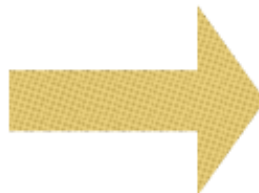
标准模型看起来似乎很完美，尤其是在希格斯玻色子发现后，但是

- ? 基本粒子的质量为什么差别那么大
- ? 希格斯粒子的质量哪里来的
- ? 标准模型的真空为什么不稳定
- ? 基本粒子的味结构的根源是什么
- ? 宇宙中物质-反物质不对称性的根源是什么
- ? 暗物质本质是什么
- ?
- ? 引力，暗能量?

测量作为发现手段：直接和间接

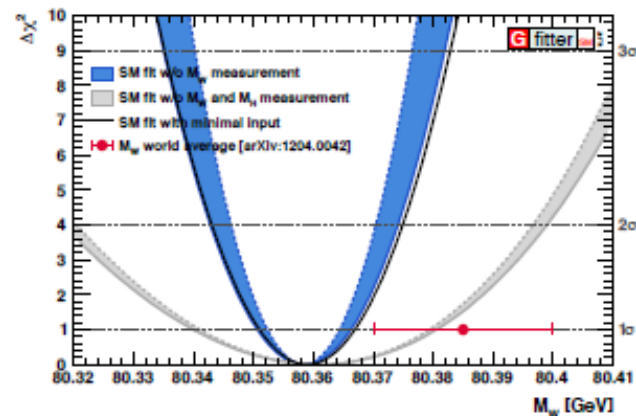
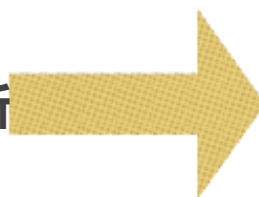
直接

- 看到新粒子或者新现象
- 例子：Higgs, XYZ, Pc, ...



间接

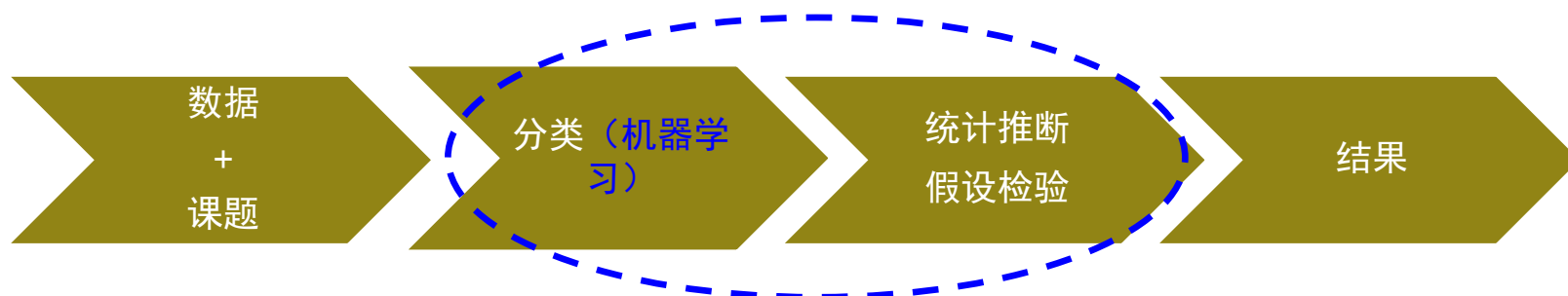
- 精确测量关键物理量
- 和理论预期做对比，差别意味着新的东西
- 例子：measure the H、W、Z, etc., precisely



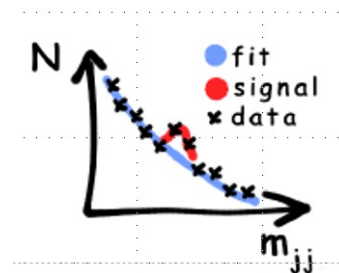
测量和数据分析

粒子物理数据分析

分类+统计推断



TP (S_{sel})	FP (B_{sel})
FN (S_{rej})	TN (B_{rej})



- ◆ 测量不是拿一把尺子简单比划一下就结束的简单任务
- ◆ 测量是个不断理解（系统误差）的过程
- ◆ 测量是个不断理解测量对象的过程
- ◆ 测量的价值在于过程中的修正和最终认识的深化
 - ✓ 定义可能会变
 - ✓ 测量方法可能会改进
 - ✓ 误差一定会减小

任何测量都不简单，关键看精度要求！

- ◆ 测体重：

- ◆ 只是需要天平或者台秤吗？概念+单位
- ◆ 人 vs 苍蝇，精度决定了测量过程

- ◆ 测小行星的质量

- ◆ $F=ma \rightarrow m = F/a$
- ◆ $F=GM_1M_2/r^2$ 或者开普勒定律，测量绕转周期
- ◆ 每种测量方法都利用了某种“新”的物理规律

测量

输出新的知识/信息
必须输入知识/信息

测量的核心

- 精度，精度，还是精度！
- 随着精度的提高，你会发现测量难度和工作量会几何量级的增加
- 最终，甚至开始怀疑测量对象（定义）——恭喜你！

粒子物理的测量是统计推断

例子：测量分支比

$$B(X \rightarrow f) = \frac{n^{obs} - n^{bkg}}{\epsilon N_X}$$

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

知止而后有定，定而后能静，静而后能安，安而后能虑，虑而后能得

➤ 统计误差

- 泊松分布： $1/\sqrt{N}$,
- 二项分布：
 $1/\sqrt{N \cdot p \cdot (1-p)}$

$$B(X \rightarrow f) = \frac{n^{obs} - n^{bkg}}{\epsilon N_X}$$

➤ 一般情况下，统计误差也给系统误差设置了一个下限（知止）

- 接近这个下限非常困难
- 一般会有一个平衡

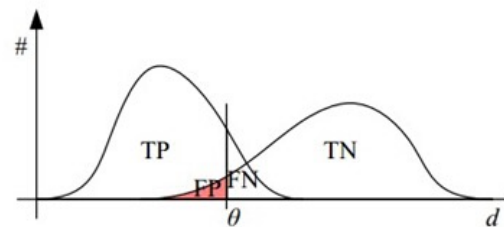
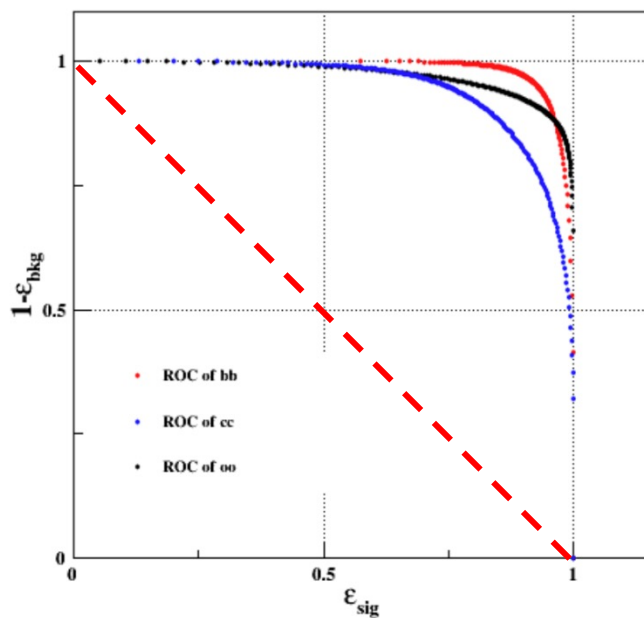
$$\delta B = (\delta n^{obs} + \delta n^{bkg}) + \delta N_X + \delta \epsilon$$

知识、快乐
和痛苦之源

数据分析（事例选择）优化

事例选择：（二）分类

名词：TP, TN, ... 混淆矩阵、ROC、AUC：和分支比无关



<table><tr><td>TP (S_{sel})</td><td>FP (B_{sel})</td></tr><tr><td>FN (S_{rej})</td><td>TN (B_{rej})</td></tr></table>	TP (S_{sel})	FP (B_{sel})	FN (S_{rej})	TN (B_{rej})	<table><tr><td>TP (S_{sel})</td><td>FP (B_{sel})</td></tr><tr><td>FN (S_{rej})</td><td>TN (B_{rej})</td></tr></table>	TP (S_{sel})	FP (B_{sel})	FN (S_{rej})	TN (B_{rej})	<table><tr><td>TP (S_{sel})</td><td>FP (B_{sel})</td></tr><tr><td>FN (S_{rej})</td><td>TN (B_{rej})</td></tr></table>	TP (S_{sel})	FP (B_{sel})	FN (S_{rej})	TN (B_{rej})
TP (S_{sel})	FP (B_{sel})													
FN (S_{rej})	TN (B_{rej})													
TP (S_{sel})	FP (B_{sel})													
FN (S_{rej})	TN (B_{rej})													
TP (S_{sel})	FP (B_{sel})													
FN (S_{rej})	TN (B_{rej})													
$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$	$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$	$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} = 1 - \text{FPR}$												
HEP: “efficiency” $\epsilon_s = \frac{S_{\text{sel}}}{S_{\text{tot}}}$	HEP: “purity” $\rho = \frac{S_{\text{sel}}}{S_{\text{sel}} + B_{\text{sel}}}$	HEP: “background rejection” $1 - \epsilon_b = 1 - \frac{B_{\text{sel}}}{B_{\text{tot}}}$												

事例选择、分类的优化（计数实验）“1 single bin”

- Total cross-section measurement in a counting experiment
- To minimize statistical errors: *maximise efficiency*purity $\epsilon_s*\rho$*
 - *global efficiency $\epsilon_s=S_{\text{sel}}/S_{\text{tot}}$ and global purity $\rho=S_{\text{sel}}/(S_{\text{sel}}+B_{\text{sel}})$*

$$(\sigma_s)_{\text{meas}} = \frac{N_{\text{meas}} - \mathcal{L}\epsilon_b\sigma_b}{\mathcal{L}\epsilon_s}$$

$$\Delta\sigma_s = \frac{\Delta N_{\text{meas}}}{\mathcal{L}\epsilon_s} = \frac{1}{\mathcal{L}\epsilon_s} \sqrt{N_{\text{exp}}}$$

$$N_{\text{exp}} = S_{\text{sel}} + B_{\text{sel}} = \frac{S_{\text{sel}}}{\rho} = \frac{\mathcal{L}\sigma_s\epsilon_s}{\rho}$$

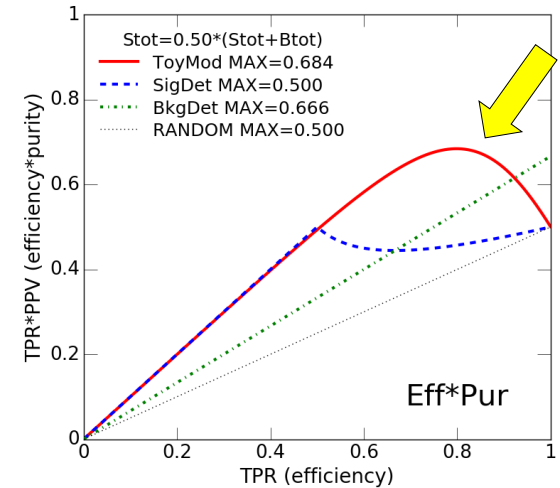
$$\frac{1}{(\Delta\sigma_s)^2} = \frac{1}{\sigma_s} \mathcal{L}\epsilon_s\rho = \frac{1}{\sigma_s^2} S_{\text{tot}}\epsilon_s\rho$$

事例选择、分类的优化 (计数实验)

- 计数实验中的截面测量 σ_s
- 为了最小化统计误差, 我们可以最大化 $\epsilon_s * \rho$
 - *global efficiency* $\epsilon_s = S_{\text{sel}}/S_{\text{tot}}$ and *global purity* $\rho = S_{\text{sel}}/(S_{\text{sel}}+B_{\text{sel}})$ – “1 single bin”

$$\frac{1}{(\Delta\sigma_s)^2} = \frac{1}{\sigma_s} \mathcal{L} \epsilon_s \rho = \frac{1}{\sigma_s^2} S_{\text{tot}} \epsilon_s \rho$$

- For each classifier
 - vary cut $\rightarrow$ vary ϵ_s and ρ
 $\rightarrow$ find maximum of $\epsilon_s * \rho$ (choose “operating point”)
- $\epsilon_s * \rho$: metric between 0 and 1
 - qualitatively relevant: the higher, the better
 - numerically: fraction of Fisher information ($1/\text{error}^2$) available after selecting
 - *correct metric only for σ_s by counting!* $\rightarrow$ 下页表格有更多情形



如果不是计数 → Different metrics (课外)

Binary classifiers for HEP event selection (signal-background discrimination)

Statistical error minimization (or statistical significance maximization)	Cross-section (1-bin counting)	TN, AUC irrelevant Only 2 or 3 global/local variables –	2 variables: global ϵ_s , ρ (given S_{tot})	Maximise $S_{tot} * \epsilon_s * \rho$ (at any S_{tot})
	Searches (1-bin counting)		Simple and CCGV – 2 variables: global S_{sel} , B_{sel} (or equivalently ϵ_s , ρ)	Maximise $\frac{S_{sel}}{\sqrt{S_{sel}+B_{sel}}}$ (i.e. $\sqrt{S_{tot} * \epsilon_s * \rho}$)
				Maximise $\sqrt{2((S_{sel} + B_{sel}) \log \left(1 + \frac{S_{sel}}{B_{sel}}\right) - S_{sel})}$
			HiggsML – 2 variables: global S_{sel} , B_{sel}	Maximise $\sqrt{2((S_{sel} + B_{sel} + K) \log \left(1 + \frac{S_{sel}}{B_{sel}+K}\right) - S_{sel})}$
			Punzi – 2 variables: global ϵ_s , B_{sel}	Maximise $\frac{\epsilon_s}{A/2 + \sqrt{B_{sel}}}$
	Cross-section (binned fits)		2 variables: local $\epsilon_{s,i}$ and ρ_i in each bin (given $s_{tot,i}$ in each bin)	Maximise $\sum_i s_{tot,i} * \epsilon_{s,i} * \rho_i$ Partition in bins of equal ρ_i
				Maximise $\sum_i s_{tot,i} * \epsilon_{s,i} * \rho_i * \left(\frac{1}{s_{tot,i}} \frac{\partial s_{tot,i}}{\partial \theta}\right)^2$ Partition in bins of equal $\rho_i * \left(\frac{1}{s_{tot,i}} \frac{\partial s_{tot,i}}{\partial \theta}\right)$
	Searches (binned fits)		3 variables: local s_{sel} , s_{tot} , S_{sel} in each bin (2 counts or ratios enough?)	Maximise a sum? *
Statistical + Systematic error minimization			3 variables: ϵ_s , ρ , lumi (lumi: tradeoff stat. vs. syst.)	No universal recipe * (may use local S_{sel} , B_{sel} in side band bins)

* Many open questions

误差分析

什么是误差

- * 误差:
 - * Parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand
 - * 与测量结果相关联的参数，表征可合理归因于被测对象的值的离散度
 - * Error/uncertainty
 - * 给出值：测量值，实验值，标称值，计算近似值等
 - * 真值：客观值
- * 一般说来：
 - * 真值是未知的——测量的意义
 - * 因而误差也未知——误差分析的意义
- * 有些情况真值是可以知道的；有些情况真值在相对意义上是知道的——因而误差分析是可能的。
- * 真值已知情况
 - * 理论真值：三角形内角和； $e^+e^- \rightarrow l^+l^-$ 角分布
 - * 定义值：米的定义；绝对零度
 - * 相对真值：研究强相互作用时，弱电理论认为是可以完美忽略的；
BES分析 ($J/\psi, \psi'$ 衰变) 中可以忽略弱作用

误差的分类

- **统计误差**：即随机误差，由于样本大小造成的不确定性。靠增大统计量减小。
- **系统误差**：非随机误差，绝对值及符号恒定，或条件改变而按一定规律变化。靠增大信号本身统计量无法减小。判断误差来源，改善方法或进行测量并修正。
- **粗大误差**：异常测量值。

此处仅限于统计误差与系统误差。

统计误差

统计误差

$\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow \mu^+ \mu^-$

$$X_i \stackrel{i.i.d}{\sim} (m_{J/\psi}, \sigma^2), i = 1, 2, \dots, n,$$

$$m_{J/\psi}, \sigma^2,$$

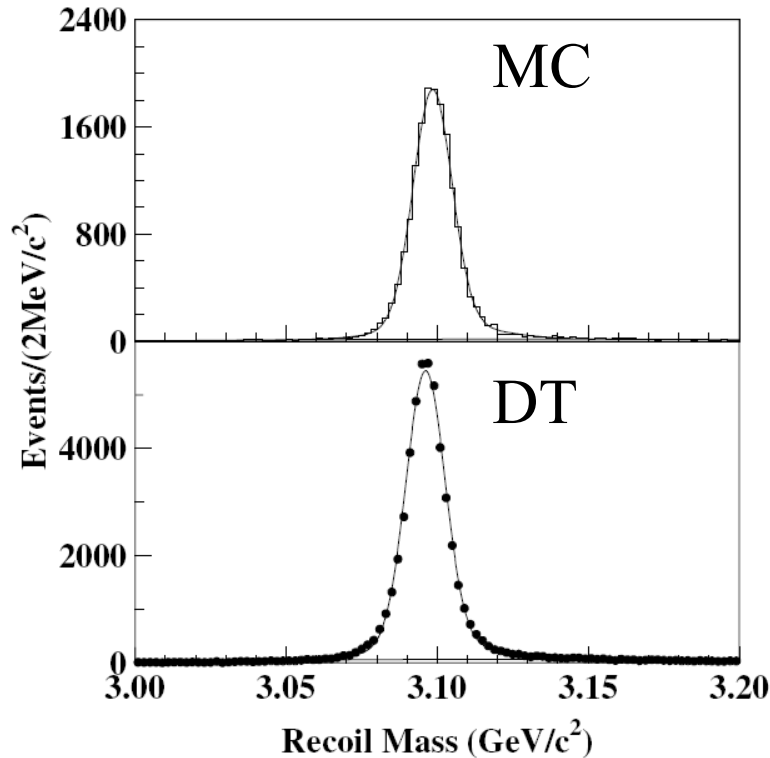
$$\nu_i = x_i - \bar{x}, \quad \bar{x} = \sum_i x_i / n,$$

$$m_{J/\psi} = \bar{x};$$

$$\hat{\sigma}^2 = \sum_i \nu_i^2 / (n-1);$$

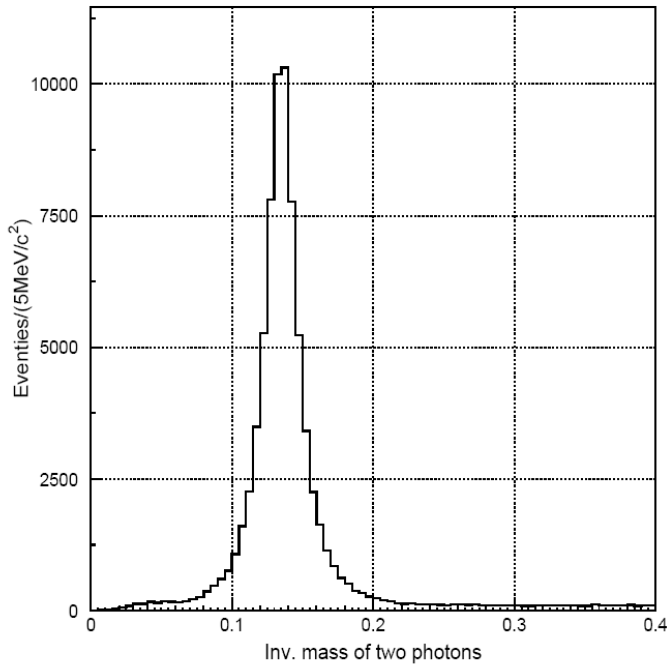
$$\hat{\sigma}^2(\bar{x}) = \hat{\sigma}^2 / n.$$

$$\sigma m_{J/\psi} = \sqrt{\hat{\sigma}^2(\bar{x})}$$



质量的统计误差依赖于统计量 n 和实验分辨 σ

为什么要好的探测器



$$m_{\gamma\gamma}^2 = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos\theta_{\gamma\gamma})$$

$$\Rightarrow \sigma m_{\gamma\gamma} \propto \sigma E_{\gamma}$$

$$\Rightarrow \sigma \bar{m}_{\gamma\gamma} \propto \sigma E_{\gamma} / \sqrt{n}$$

可见能量测量的不确定性要通过增大事例数来补偿!

Performance	BESII	BESIII	CLEOc
$\sigma p/p$	$1.7\%/\sqrt{1+p^2}$	$0.6\% @ p=1\text{GeV}$	$0.6\% @ p=1\text{GeV}$
$\sigma E/E$	$22\% / \sqrt{E}$	$2.5\% @ E=1\text{GeV}$	$2.2\% @ E=1\text{GeV}$
PartID	dE/dx+TOF	dE/dx+TOF	dE/dx+RICH
Coverage	80%	93%	93%

统计误差

- **计数 (1-bin)**
- **参数估计 (Binned fits)**

1-bin 计数

- ◆ 泊松分布(Poisson)

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots$$

$$\bar{x} = \lambda$$

$$V(x) = \lambda$$

- ◆ $\sigma = \text{sqrt}(S)$

- ◆ 存在本底时: $\sigma = \text{sqrt}(S+B)$

$$\frac{1}{(\Delta\sigma_s)^2} = \frac{1}{\sigma_s} \mathcal{L}\epsilon_s \rho = \frac{1}{\sigma_s^2} S_{\text{tot}} \epsilon_s \rho$$

- ◆ $\epsilon_s * \rho$: metric between 0 and 1

- the higher, the better (*only for this specific use case!*)
- Fisher information ($1/\text{error}^2$) available after selecting

- ◆ 少数情况会用到二(多)项分布: $1/\text{sqrt}[N * p * (1-p)]$

参数估计 (binned fits: 课外)

- 数据：一个（多维）分布函数 $f(x)$ （ m 个 bin）每个bin中的事例计数 n_i
 - expected counts $y_i = f(x_i, \theta) dx \rightarrow$ depend on a parameter θ that we want to fit

- 课本告诉我们（Fisher information）

$$(\Delta\hat{\theta})^2 = \text{var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}_\theta} \quad (\text{Cramer-Rao lower bound}), \text{ where } \mathcal{I}_\theta = \sum_{i=1}^m \frac{1}{y_i} \left(\frac{\partial y_i}{\partial \theta} \right)^2 = \int \frac{1}{f} \left(\frac{\partial f}{\partial \theta} \right)^2 dx \quad (\text{Fisher information})$$

- With an ideal classifier (or no background) $\rightarrow y_i = S_i$ and $\mathcal{I}_\theta^{(\text{ideal classifier})} = \sum_{i=1}^m \frac{1}{S_i} \left(\frac{\partial S_i}{\partial \theta} \right)^2$

- With a realistic classifier $\rightarrow y_i(\theta) = \epsilon_i S_i(\theta) + b_i$ and $\mathcal{I}_\theta^{(\text{real classifier})} = \sum_{i=1}^m \epsilon_i \rho_i \times \frac{1}{S_i} \left(\frac{\partial S_i}{\partial \theta} \right)^2$

- ϵ_i and $\rho_i \rightarrow$ local signal efficiency and purity in the i^{th} bin

- 二分类的优化 $\rightarrow$ maximize Information Fraction
 - higher is better

$$(\Delta\hat{\theta}^{(\text{real classifier})})^2 \geq \frac{1}{\text{IF}} (\Delta\hat{\theta}^{(\text{ideal classifier})})^2$$

$$\text{IF} = \frac{\mathcal{I}_\theta^{(\text{real classifier})}}{\mathcal{I}_\theta^{(\text{ideal classifier})}} = \frac{\sum_{i=1}^m \epsilon_i \rho_i \times \frac{1}{S_i} \left(\frac{\partial S_i}{\partial \theta} \right)^2}{\sum_{i=1}^m \frac{1}{S_i} \left(\frac{\partial S_i}{\partial \theta} \right)^2}$$

系统误差

系统误差

- 非统计原因造成的测量结果与真值之间的差别。
- 系统误差的来源：
 1. 测量设备（测量不准）
 2. 实验、分析方法（运动学拟合）
 3. 理论模型（角分布、lund用在低能）
 4. 探测器模拟（核作用的模拟）
 5. 实验者的偏见（期望有新发现）
 6. 理论预期的误导（标准模型总是对的）
 7. 其它（与别的实验不一致心里不塌实）

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

$$\mathcal{E}_{DT} = \mathcal{E}_{DT}^{Geom} \cdot \mathcal{E}_{DT}^{Track} \cdot \mathcal{E}_{DT}^{Ptid} \cdot \mathcal{E}_{DT}^{\gamma ID} \dots$$

$$\mathcal{E}_{MC} = \mathcal{E}_{MC}^{Geom} \cdot \mathcal{E}_{MC}^{Track} \cdot \mathcal{E}_{MC}^{Ptid} \cdot \mathcal{E}_{MC}^{\gamma ID} \dots$$

$$\mathcal{E}_{DT}^i = \mathcal{E}_{MC}^i \cdot f^i$$

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。

Monte Carlo 模拟

- **Monte Carlo 模拟在高能物理实验中广泛应用**
 - **物理事例产生子，产生末态粒子四动量**
 - **粒子与探测器物质发生相互作用**
 - **探测器响应的模拟**
 - **触发、刻度、重建**
 - **与数据进行同样分析，可得到效率、分辨率等**
- **高能物理实验依赖Monte Carlo 模拟**
- **MC 模拟的有效性须数据证明——数据分析和误差处理的重要内容**

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。
 - * $J/\psi \rightarrow e\mu$ 的寻找，bhabha和dimu本底
($e \rightarrow \mu$, $\mu \rightarrow e$ 误判)

系统误差的分析

- ◆ 模型误差、理论误差等较易考虑（至少不必由实验者负责）。
- ◆ 事例选择误差确定是实验者的基本任务。
- ◆ 一言以蔽之，事例选择系统误差表征
“测量数据与Monte Carlo之间的吻合程度”
- ◆ 在“效率修正”部分已经包括：
系统误差即“f因子”的误差。
- ◆ 质量分辨等造成的系统误差同样考虑。

系统误差确定

* 基本考虑

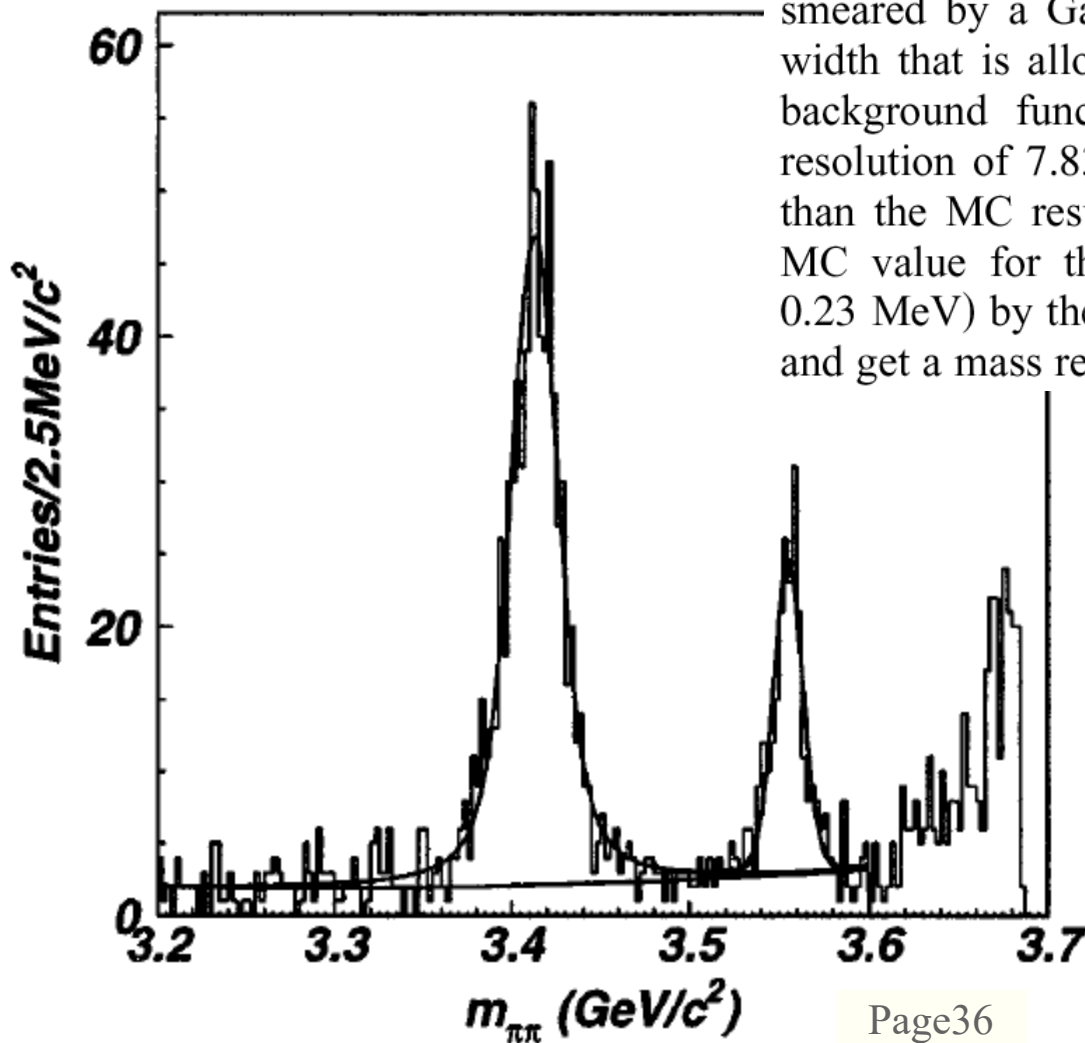
- * 待测物理量与哪些测量量有关，如何将测量量的误差传递给待测物理量

$$m_{\gamma\gamma}^2 = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos\theta_{\gamma\gamma})$$

- * 原则上可以由每个测量量误差求出。
- * 但有时几乎不可能（关系过于复杂），如
 - $J/\psi \rightarrow \pi^+ \pi^- \gamma$: 在运动学拟合后，所有的信息都关联起来了！
- * 寻找变通的办法！对数据分析很重要！

质量分辨对宽度造成的系统误差

For the $\pi^+\pi^-$ channel, we first fit the invariant mass region between 3.5 and 3.6 GeV with a Breit-Wigner resonance with $\Gamma_{\chi_{c2}}$ fixed at the PDG value of 2.00 MeV, smeared by a Gaussian resolution function with an rms width that is allowed to float. We also include a linear background function. The fit results in a χ_{c2} mass resolution of 7.83 ± 1.04 MeV, which is slightly higher than the MC result of 6.31 ± 0.11 MeV. We scale the MC value for the mass resolution at the χ_{c0} (8.12 ± 0.23 MeV) by the ratio of the fitted MC results at the χ_{c2} and get a mass resolution at the χ_{c0} of 10.08 MeV.

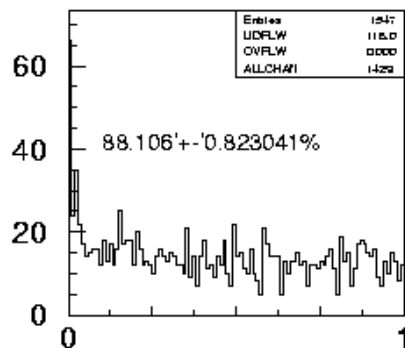


χ_{c0} 宽度测量!

BES, PRL81, 3091(1998)

粒子鉴别效率的系统误差

Table 2: Particle ID efficiencies ($Prob_{ID} > 0.01$) for π^+ , π^- , p and $\bar{p}$ samples selected from $\psi(2S) \rightarrow \pi^+\pi^-J/\psi$, $J/\psi \rightarrow p\bar{p}$ (in %).



J/ψ-prbprp

Particle	data	Monte Carlo
π^-	94.64 ± 0.60	88.55 ± 0.78
π^+	93.69 ± 0.64	87.67 ± 0.81
$\bar{p}$	87.76 ± 0.84	86.50 ± 0.84
p	88.11 ± 0.82	88.44 ± 0.79

数据和MC模拟对 $\pi^+\pi^-$ 和质子、反质子粒子鉴别的效率。
由 $\psi(2S) \rightarrow \pi^+\pi^-J/\psi$, $J/\psi \rightarrow p\bar{p}$ 测得。

MC efficiency corr. factor: 1.154 ± 0.028

相关误差

- * 来源不完全独立的误差。

- * 最简单的例子

寻迹效率、几何接受度

计算事例总数时要计算其误差，在计算分支比时仍要计算。事实上，二者会有一部分抵消。

- * 一个实验组中总数总是100%相关的，在本组内对比时其误差不应考虑在内。相同的软件，如运动学拟合也带来相关的误差。两个独立做同一道分析的人结果仍然不是相互独立的，因为他们的数据样本有很大程度的重叠。

- * 高精度测量时非常重要。

误差合并与传播

* 直接测量误差与间接测量误差

$$Y=f(X_1,X_2,\dots,X_N)$$

$$\sigma^2(Y) = \left[\frac{\partial f}{\partial X_1} \right]^2 \sigma^2(X_1) + \left[\frac{\partial f}{\partial X_2} \right]^2 \sigma^2(X_2) + \dots + \left[\frac{\partial f}{\partial X_N} \right]^2 \sigma^2(X_N) \\ + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial X_i} \frac{\partial f}{\partial X_j} \frac{1}{n} \sum_{k=1}^n \delta X_{ik} \delta X_{jk}$$

关联项各项
独立时为零。

不同的误差合并与传播方式有差别，高能物理实验中常假设误差间无关联，所有误差都较小且近似按正态分布。（add in quadrature）

常用误差传播公式

(1) $U = aX \pm bY$

$$\sigma^2(U) = a^2\sigma_X^2 + b^2\sigma_Y^2 \pm 2ab \operatorname{cov}(X, Y)$$

(2) $U = aXY$

$$\frac{\sigma^2(U)}{U^2} = \frac{\sigma_X^2}{X^2} + \frac{\sigma_Y^2}{Y^2} + \frac{2 \operatorname{cov}(X, Y)}{X \cdot Y}$$

(3) $U = aX^{\pm b}$

$$\frac{\sigma(U)}{U} = \frac{b \cdot \sigma_X}{X}$$

(4) $U = ae^{\pm bX}$

$$\frac{\sigma(U)}{U} = b \cdot \sigma_X$$

(5) $U = a \ln(\pm bX)$

$$\sigma(U) = \frac{a \cdot \sigma_X}{X}$$

**非常有用，
学会推导！**

中场小结

- 高能实验是加速物理理论发展的重要推动力
- 测量的直接目的是最大化实验的输出信息
- 为此，必须有信息的输入，这正是我们工作的重要内容之一
 - 物理信息的输入：动力学、运动学
 - 数学信息：利用各种概率和统计规律
 - 探测器：扬长避短
 - 程序
- 接下来，举例说明 ——

一个 BES 上的例子

物理分析是一门艺术

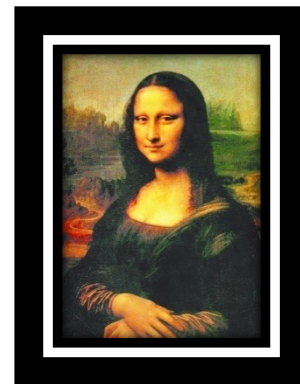
主要内容与目标

- 事例选择与本底分析
 - 目标：掌握事例选择的三大原则
- 信号数
 - 目标：了解三种常用提取信号数的方法
- 效率
 - 目标：与效率损失相关的三个主要问题

以 $\psi(2S) \rightarrow \pi^+\pi^-J/\psi$ 为例



本次课



将来

末态

- BESIII 上能探测到的末态粒子
 - 中性：光子 (γ)、正反中子 (n and $\bar{n}$)
 - 带电粒子：正反电子 ($e^\pm$)、正反缪子 ($\mu^\pm$)、正负 π 介子 ($\pi^\pm$)，正负 K 介子 ($K^\pm$)、正反质子 (p and $\bar{p}$)
- $\psi(2S) \rightarrow \pi^+\pi^-J/\psi, J/\psi \rightarrow anything$ 【半单举 (semi-inclusive)】
 - $\pi^+\pi^- + anything$
- $\psi(2S) \rightarrow \pi^+\pi^-J/\psi, J/\psi \rightarrow l^+l^-$ 【遍举 (exclusive)】
 - $\pi^+\pi^-e^+e^-$ 或者 $\pi^+\pi^-\mu^+\mu^-$

术语园地

inclusive MC, generic MC

exclusive MC, signal MC

分支比测量的实质是反应截面的测量

- 转换为信号事例数的测量，因为 $N_{sig} = \sigma \cdot L \cdot \epsilon$ （截面×亮度×效率）

$$B_{\pi\pi J/\psi} = \frac{\sigma_{\pi\pi J/\psi}}{\sigma_{tot}} = \frac{\frac{N_{\pi\pi J/\psi}}{L \cdot \epsilon_{\pi\pi J/\psi}}}{\frac{N_{tot}}{L}}$$



$$\mathcal{B}_{\pi\pi J/\psi} = \frac{N_{\pi\pi J/\psi}}{\epsilon_{\pi\pi J/\psi} \times N_{tot}}$$

$$\begin{aligned}\mathcal{B}_{ll} &= \frac{\mathcal{B}_{\pi\pi J/\psi} \times \mathcal{B}_{ll}}{\mathcal{B}_{\pi\pi J/\psi}} \\ &= \frac{N_{ll}/(\epsilon_{ll} \times N_{tot})}{N_{\pi\pi J/\psi}/(\epsilon_{\pi\pi J/\psi} \times N_{tot})} \\ &= \frac{N_{ll}/\epsilon_{ll}}{N_{\pi\pi J/\psi}/\epsilon_{\pi\pi J/\psi}}.\end{aligned}$$

分支比测量的通用公式

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X}$$



- B: 待测分支比，无量纲
- N^{sig} : 信号事例数
- N^X : 总事例数
- ϵ : 事例选择效率，通常由蒙卡模拟确定，需特别关注

拿到课题，知道了反应道末态，写出Br公式，你的事例选择就开始了！

事例选择

事例选择 ABC

- 事例选择益简、松、少
 - 有利于减少系统误差
- 利用统计原理减少误差
 - 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

好的选择应该考虑系统误差

- Track level cuts for soft pion candidates

- $|V_z| < 10\text{cm}$

- $|V_r| < 1\text{cm}$

- $|\cos\theta| < 0.80$

- $|\vec{p}| < 0.45\text{GeV}/c$

- all charged tracks satisfy above selection criteria are assumed to be pion.

- $n\text{Good}+ \geq 1 \ \&\& \ n\text{Good}- \geq 1$

- $\cos\theta_{\pi^+\pi^-} < 0.95$ to remove the photon conversion

- $3.05\text{GeV}/c^2 \leq M_{\pi^+\pi^-}^{\text{rec.}} \leq 3.15\text{GeV}/c^2$

assuming all charge tracks are pions, $M_{\pi^+\pi^-}^{\text{rec.}}$ is defined as:

$$M_{\pi^+\pi^-}^{\text{rec.}} = \sqrt{(p_{\text{ecm}} - p_{\pi^+} - p_{\pi^-})^2},$$

where p is 4 momentum of pion.

例子

$$\psi' \rightarrow \pi^+ \pi^- J/\psi$$

术语园地

soft, slow, low p

hard, fast, high p

注意：无PID，无光子要求，无track数上限要求，允许multi-entries

关于挑选 π 的更多解释

$$\psi' \rightarrow \pi^+ \pi^- J/\psi$$

- 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 π 动量不超过401 MeV（[如何计算得到？](#)），考虑了分辨后450 MeV非常安全：[无需考虑这个 cut 的系统误差](#)；
 - π 的动量不高，故其动量的绝对误差也就很小；
 - $\pi^+ \pi^-$ 的反冲侧是非常窄的 J/ψ ，可以对 $\pi^+ \pi^-$ 系统进行很好的约束，[故无需粒子鉴别，也无需考虑该系统误差](#)。

术语园地
 $\psi(2S)$ 、 ψ' 、 $\psi(3686)$

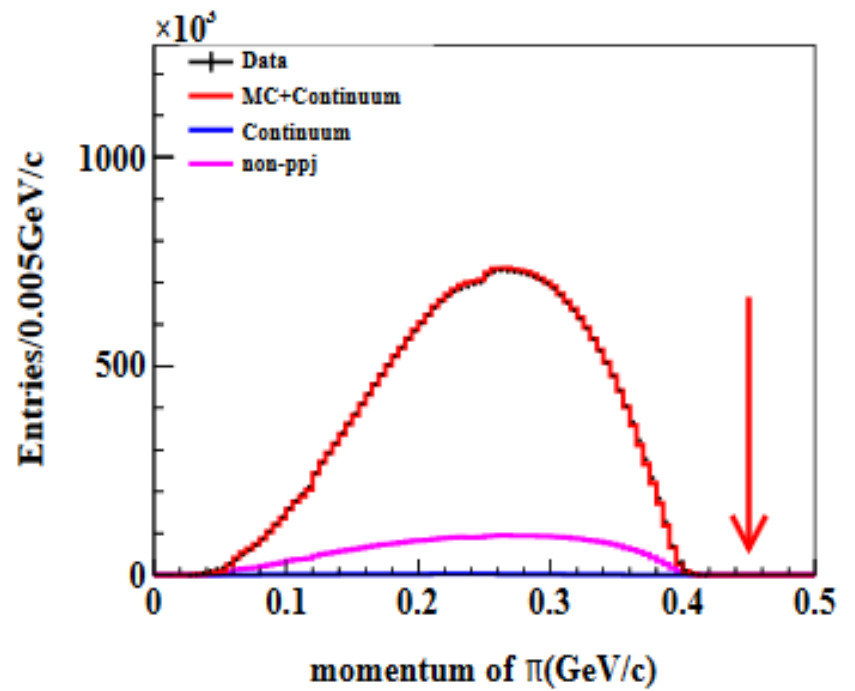
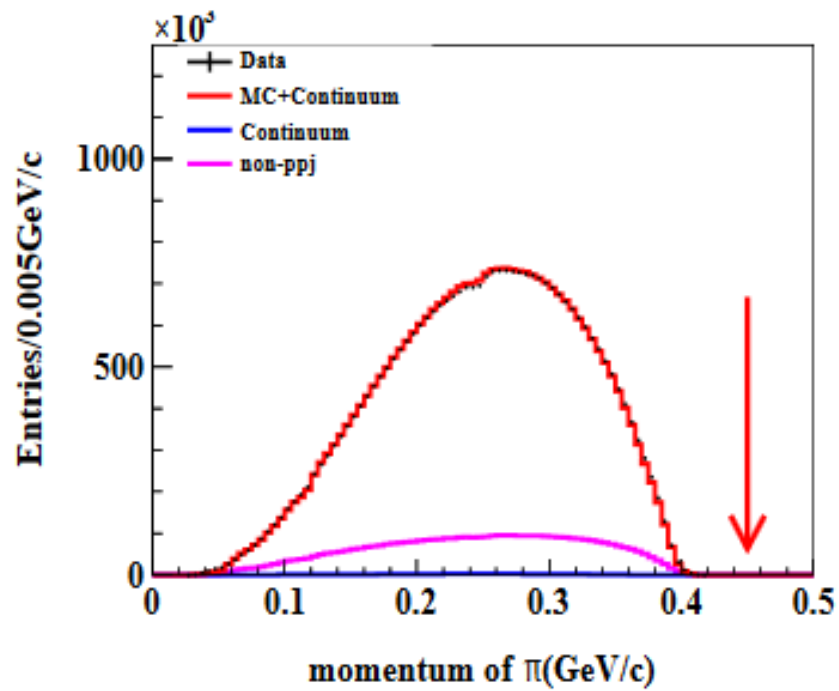


FIG. 2. The momenta of π^+ (left) and π^- (right)

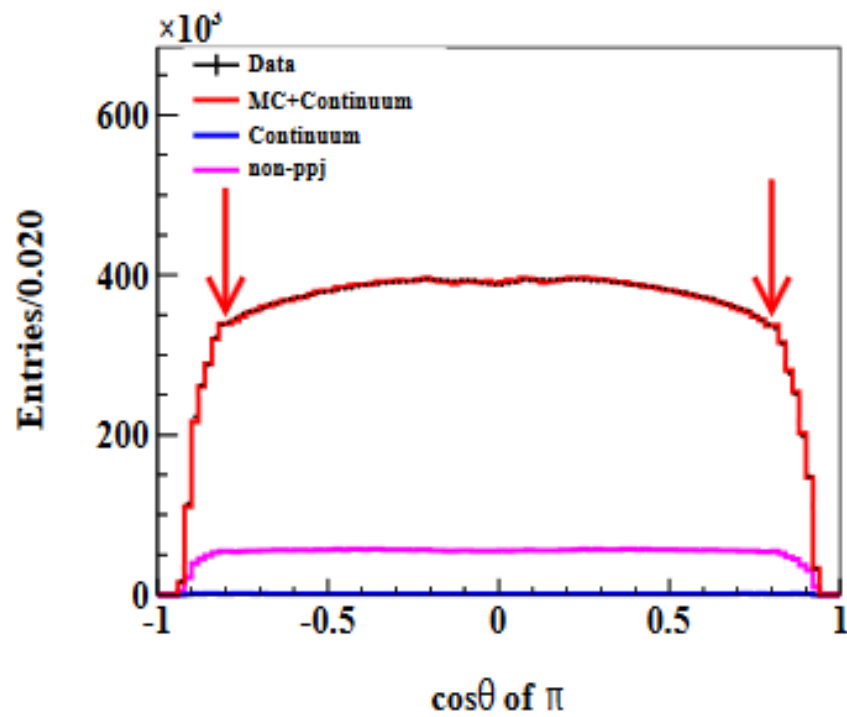
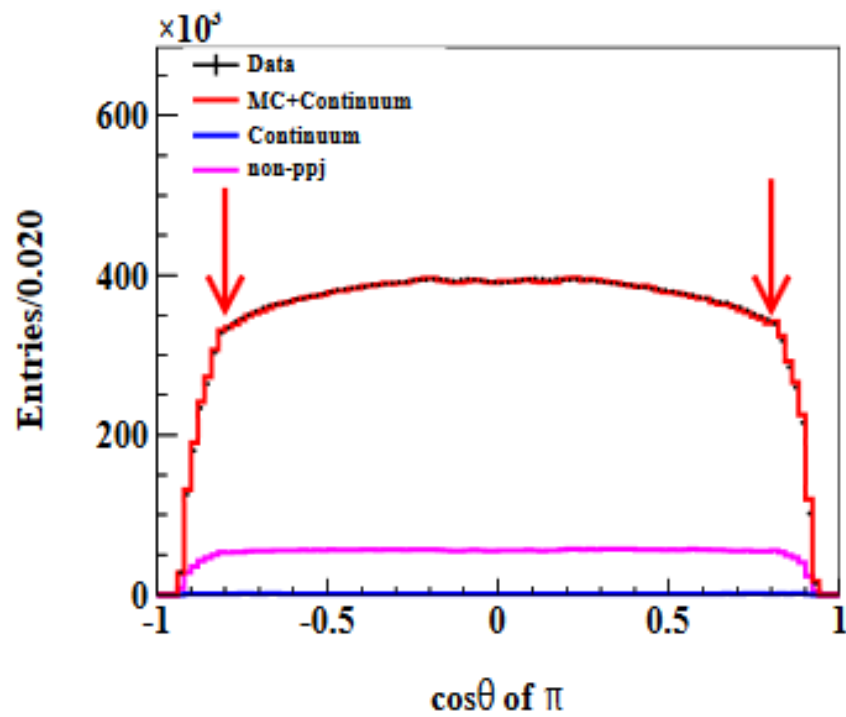


FIG. 3. The cosine of polar angle of π^+ (left) and π^- (right) in lab frame

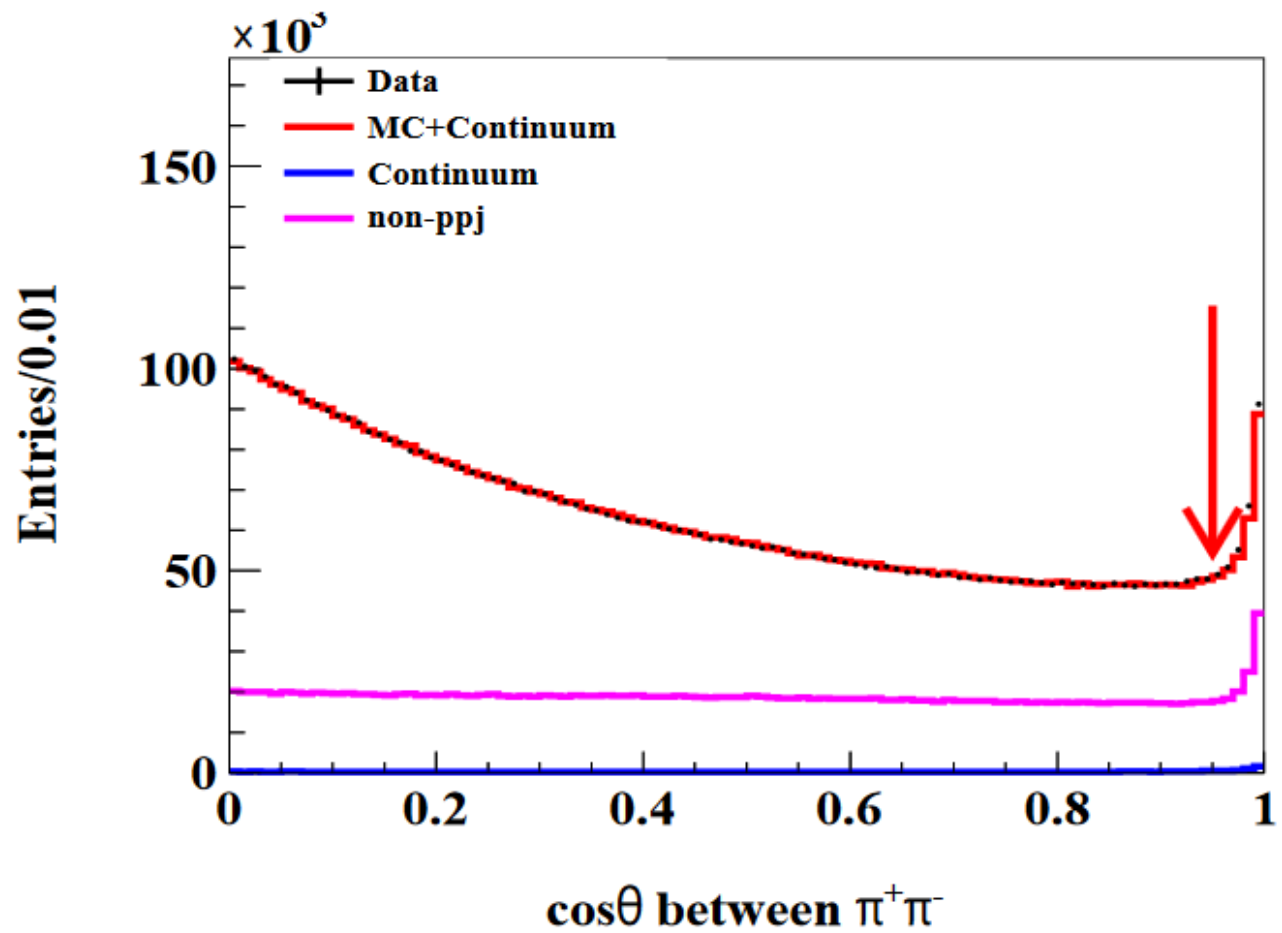
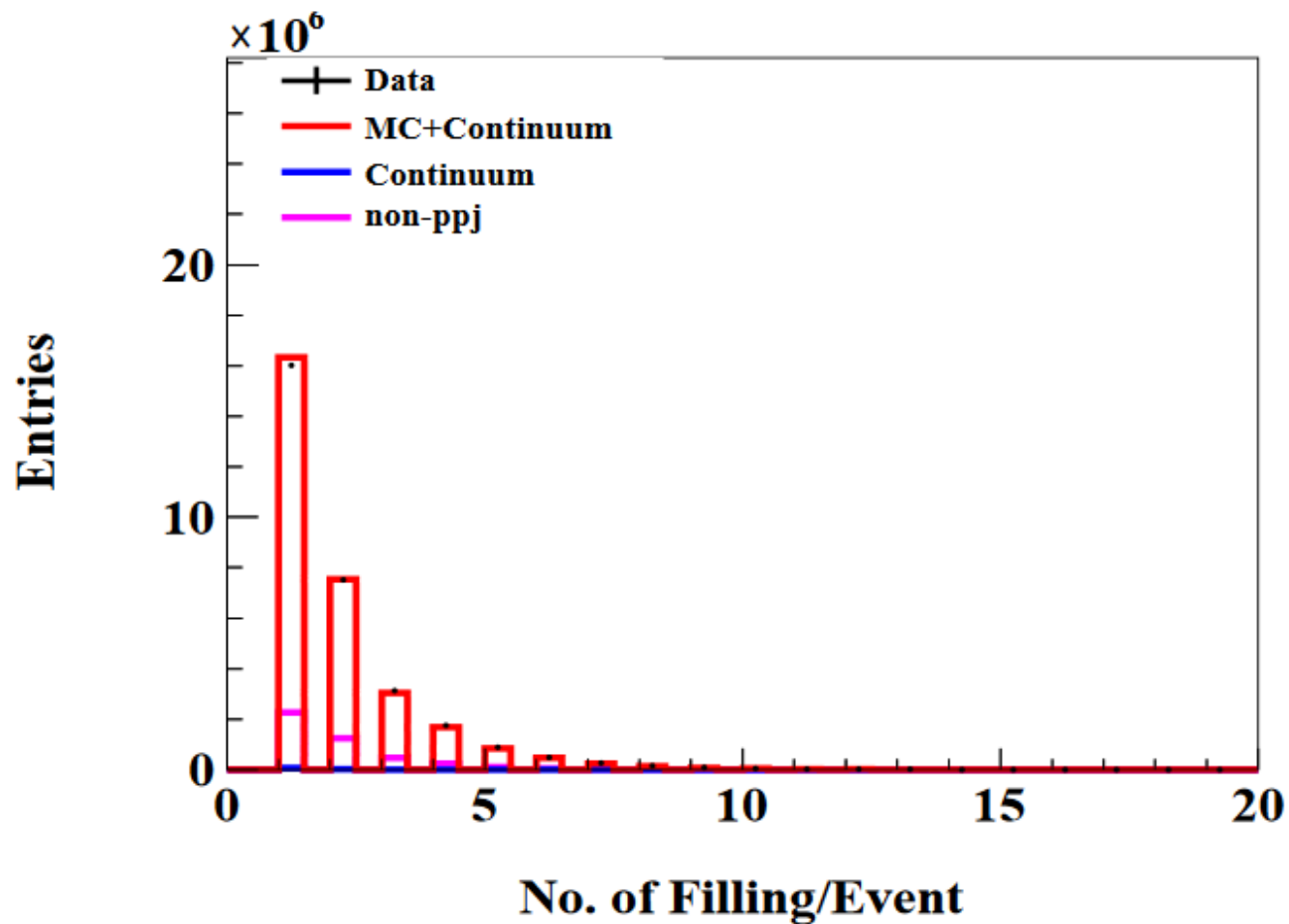


FIG. 4. Cosine distributions of the angle between two pions in $\pi^+\pi^- J/\psi$



术语园地
event 事例
entry 记录

FIG. 5. The number of filling time of each event

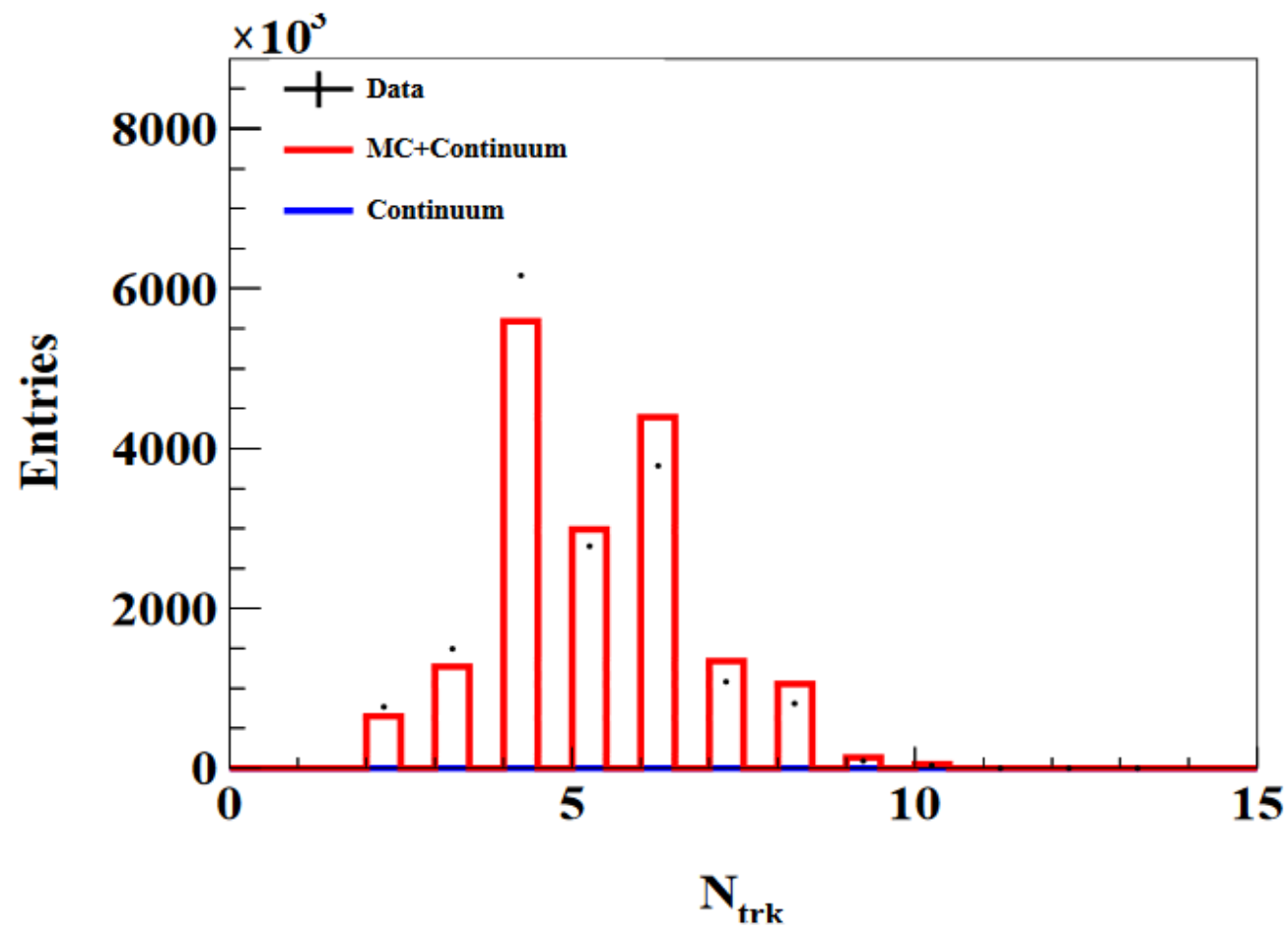
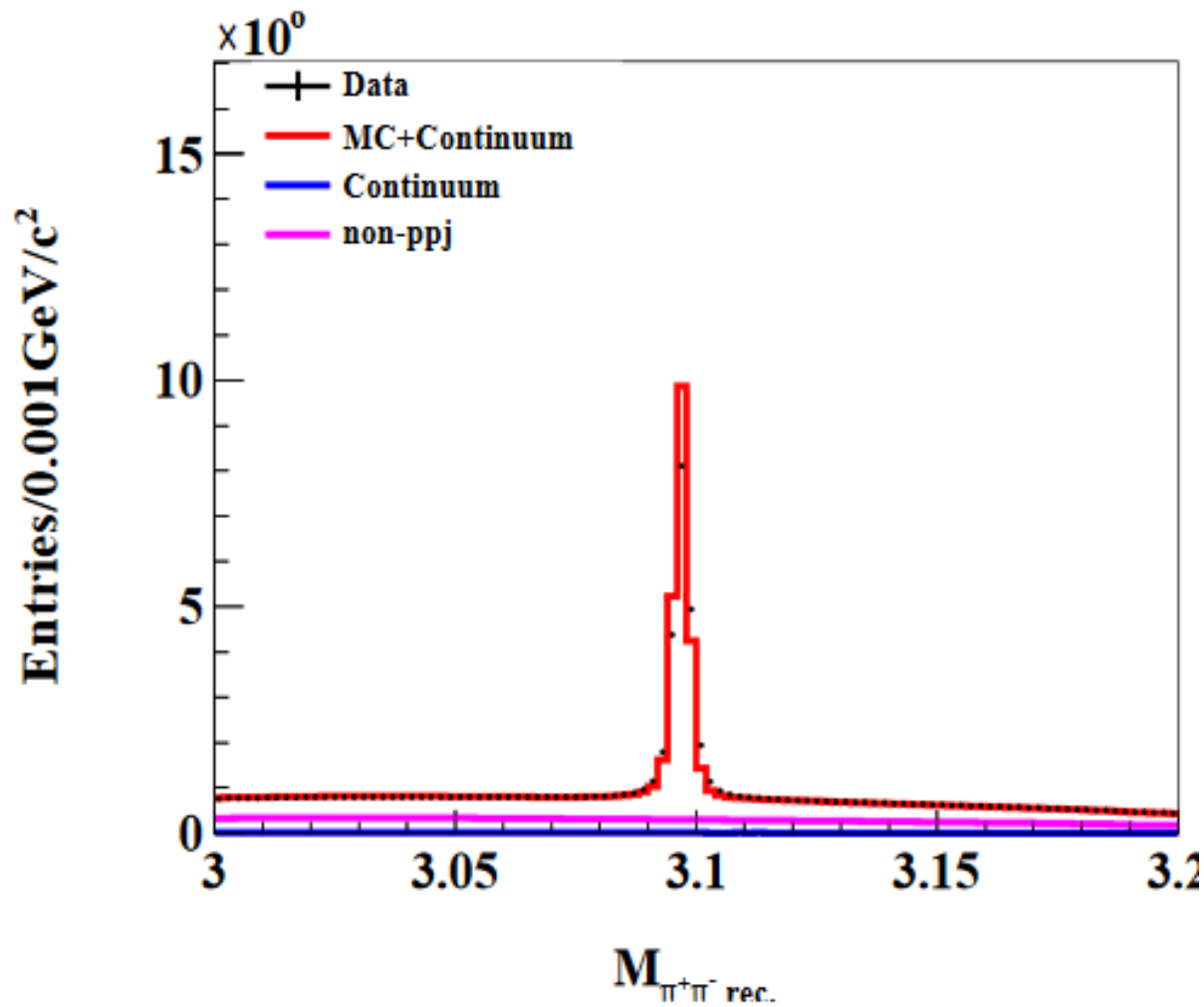


FIG. 6. The number of good tracks of each event

术语园地
multiplicity,
prong



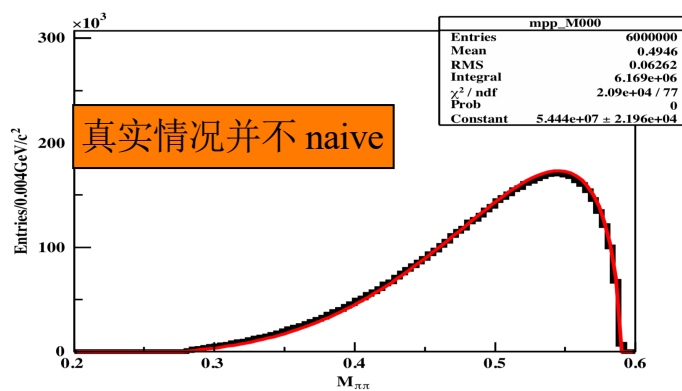
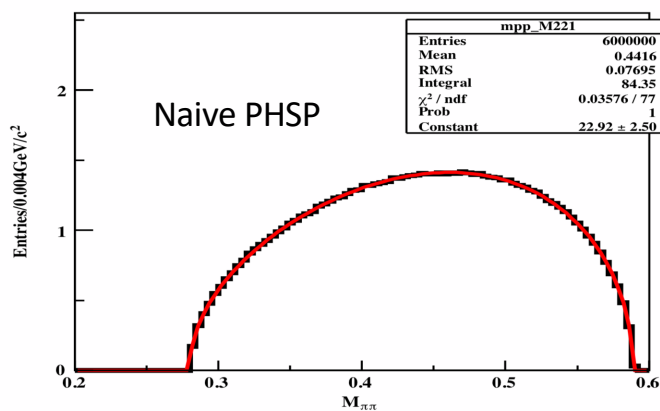
术语园地
信号 signal
本底 backgrounds (BG)

- 在物理分析中，一般用信号MC来代表信号的分布；在数据中，常常是信号与本底同时存在
- 为了方便，常常（不严谨地）将信号中的典型“峰”状分布称为信号

$M_{\pi\pi}$ spectrum

BesEvtGen: JPIPI model

Phys. Rev. D 62, 032002



两 π 倾向于背对背, $M_{\pi\pi}$ 对效率有显著影响。

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

例子

$$\psi(2S) \rightarrow \pi^+ \pi^- J/\psi$$
$$J/\psi \rightarrow l^+ l^-$$

- Track level cuts
 - $|V_z| < 10\text{cm}$
 - $|V_r| < 1\text{cm}$
 - $|\cos\theta| < 0.80$
 - $|\vec{p}| < 2.0\text{GeV}/c$
- $n_{\text{Good}} \geq 4$
- The charged tracks with $|\vec{p}| < 0.45 \text{ GeV}/c$ are assumed to be pion, select the $\pi^+\pi^-$ pair candidates by minimizing $|M_{\pi^+\pi^-}^{\text{rec.}} - M_{J/\psi}|$
- $\cos\theta_{\pi^+\pi^-} < 0.95$
- $3.05\text{GeV}/c^2 \leq M_{\pi^+\pi^-}^{\text{rec.}} \leq 3.15\text{GeV}/c^2$
- 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 $\sqrt{([E/p]^+ - 1)^2 + ([E/p]^- - 1)^2} < 0.4$
- $\cos\theta_{l+l^-} < -0.95$ in lab frame.
- $2.7\text{GeV}/c^2 < m_{l+l^-} < 3.2\text{GeV}/c^2$ for $\pi^+\pi^-e^+e^-$ channel and $3.0\text{GeV}/c^2 < m_{l+l^-} < 3.2\text{GeV}/c^2$ for $\pi^+\pi^-\mu^+\mu^-$ channel

轻子选择的一些说明

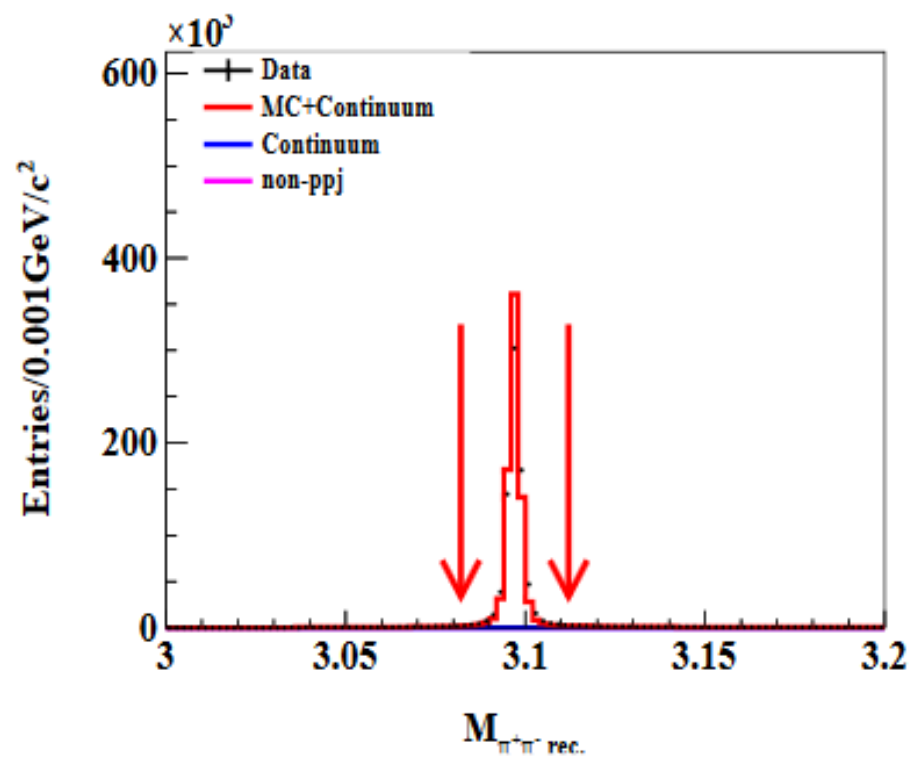
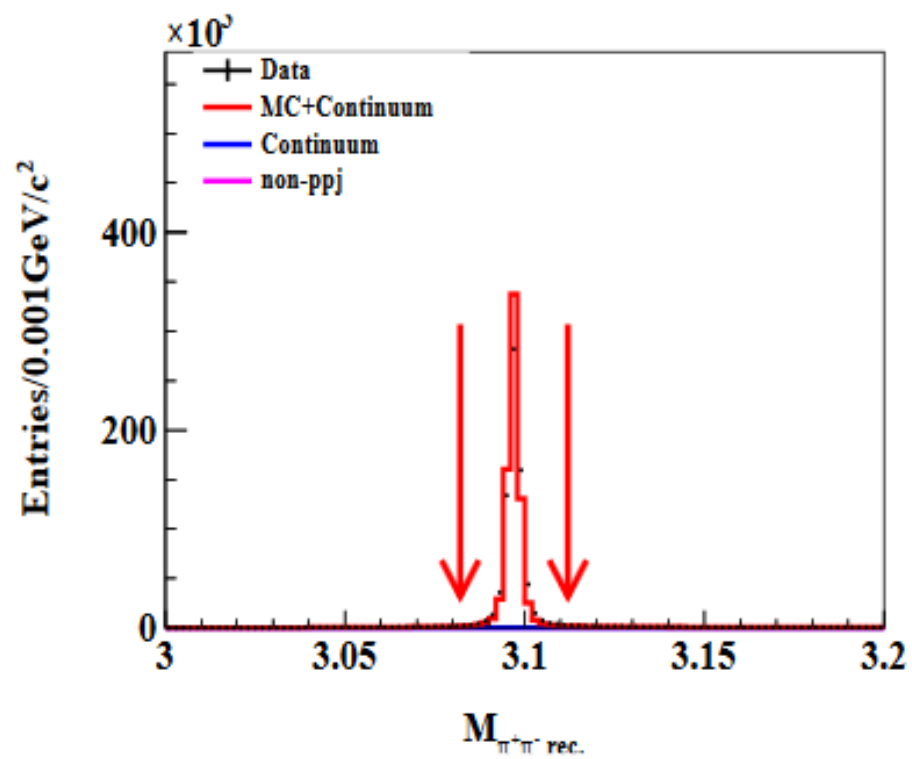
- 两体衰变，在母粒子质心系是单（高）能的：大约**1.55 GeV**
- 即使 J/ψ 运动导致一定的动量展宽，仍然和 π 的动量区间完全无重叠：无需粒子鉴别，只需区分 e 、 μ 即可
- 因此这个分析无需用到 dE/dx 和 TOF 的 PID，也就不用分析这项系统误差

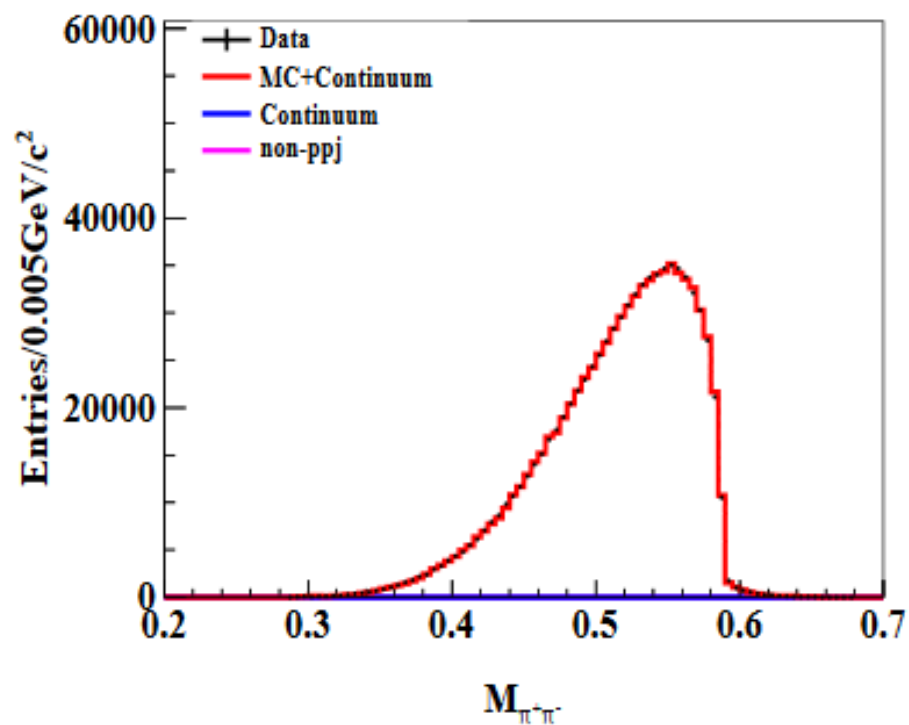
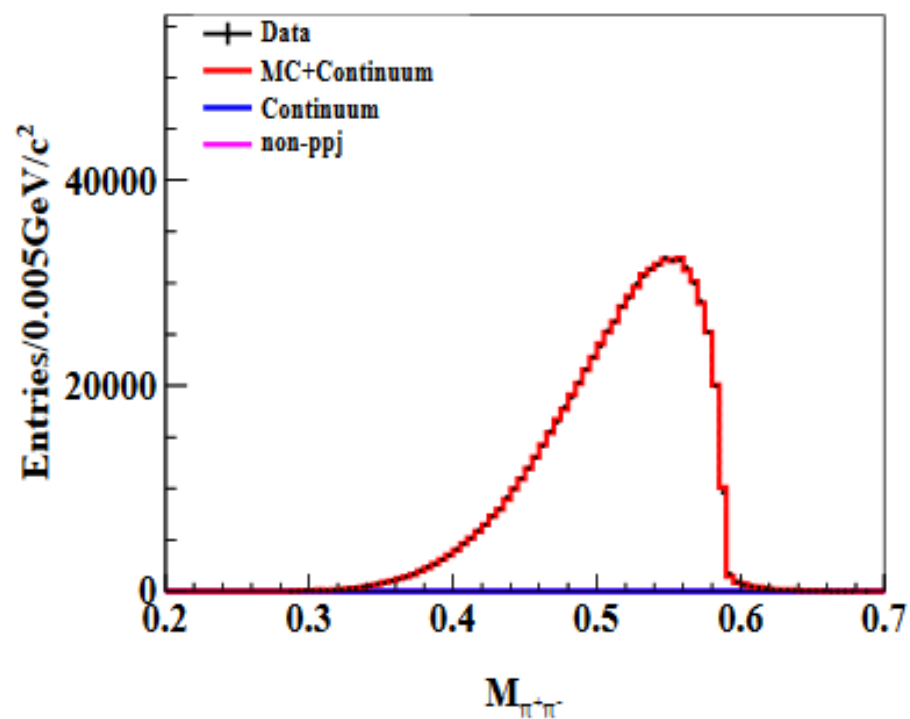
术语园地

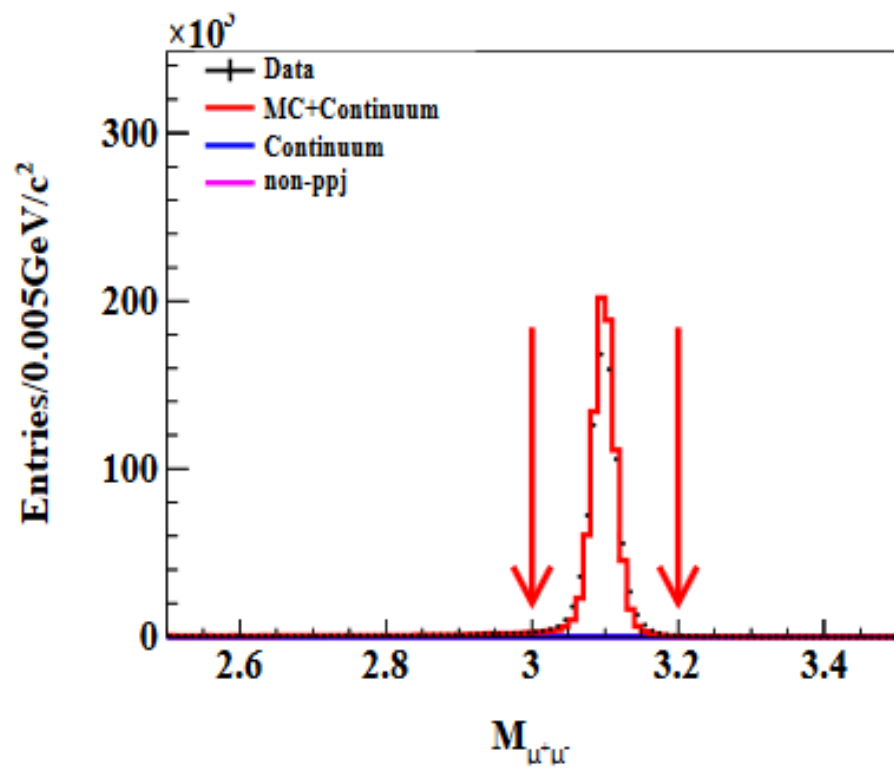
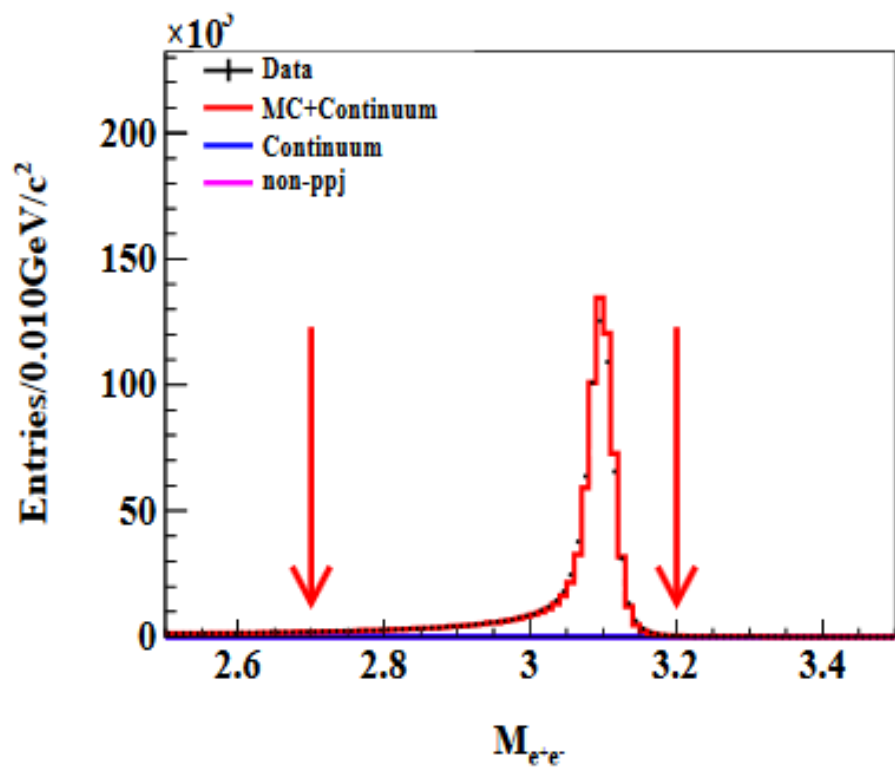
dE/dx 单位距离能损

TOF(time of flight) 粒子飞行时间

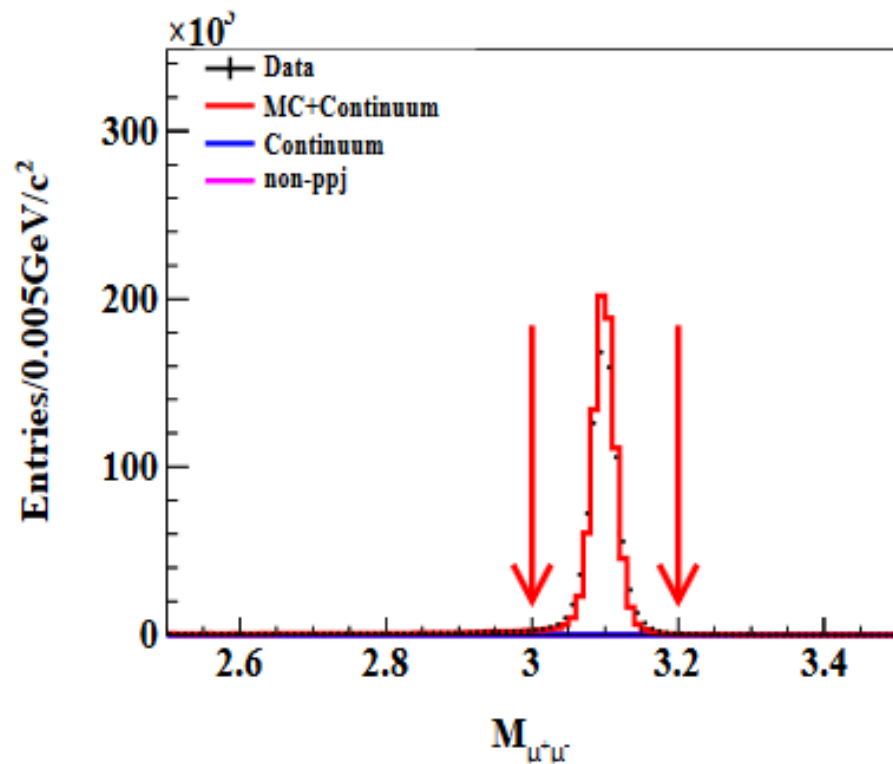
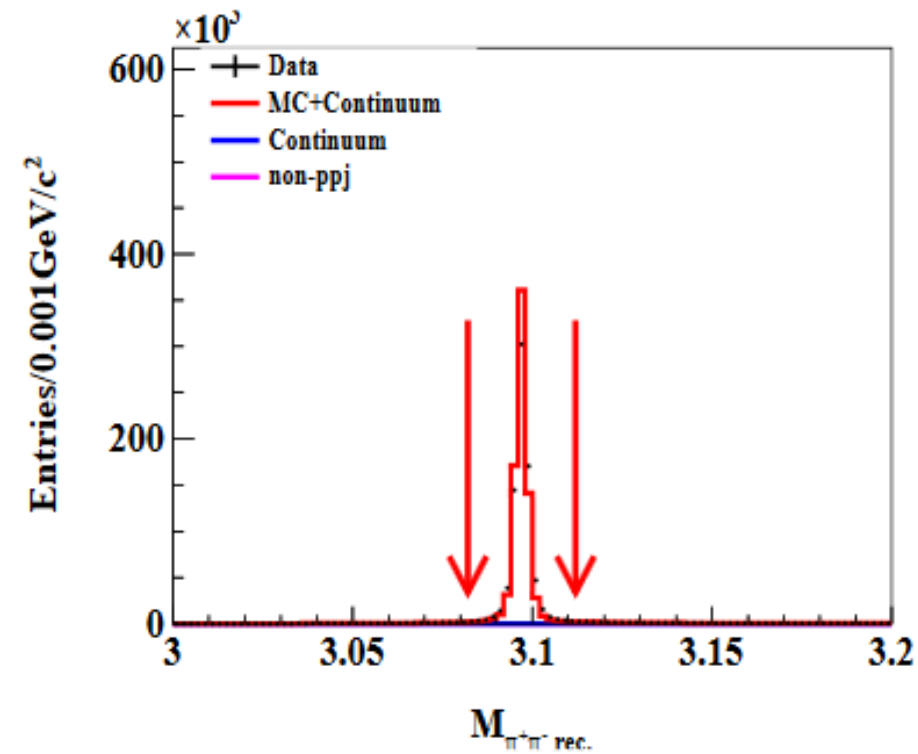
PID (particle identification) 粒子鉴别





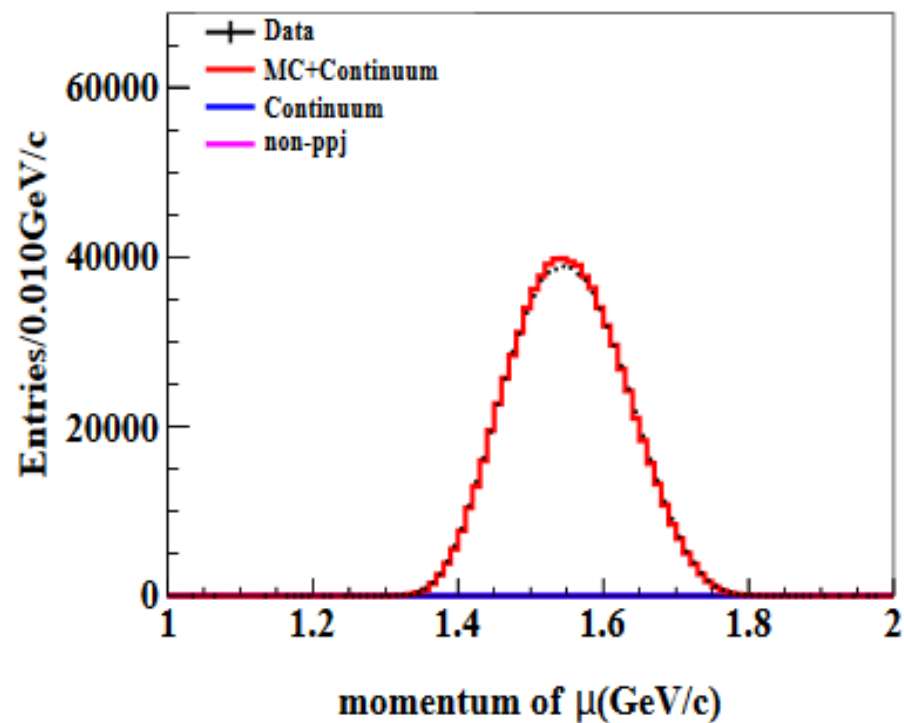
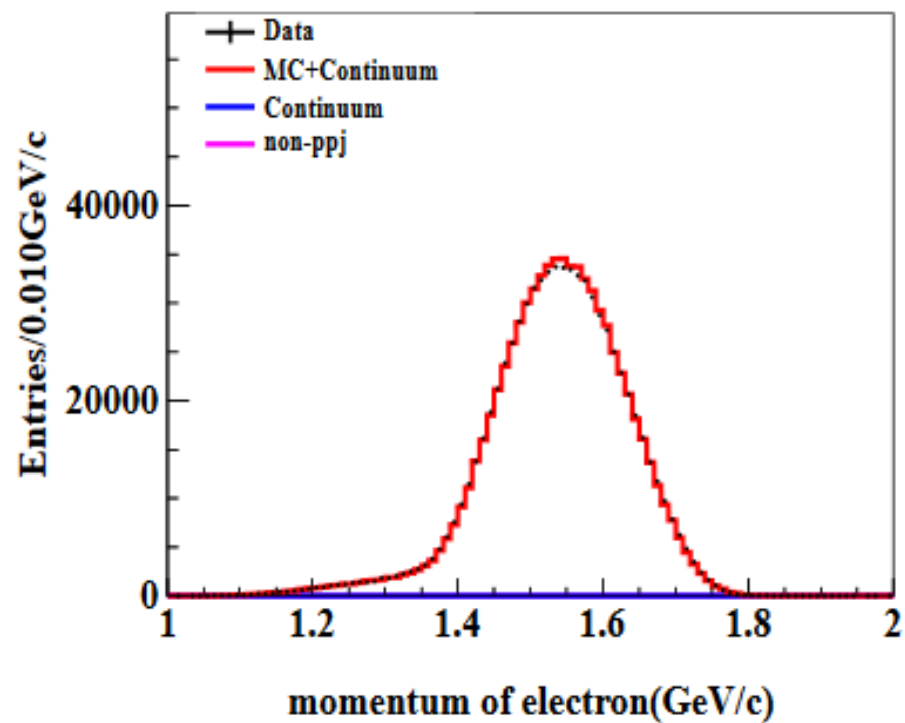


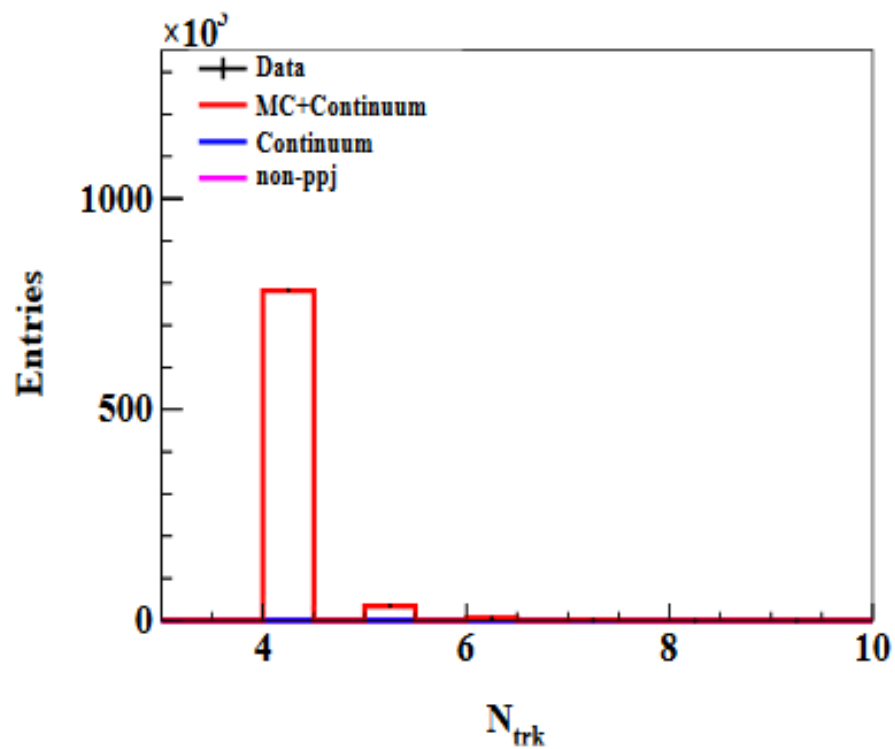
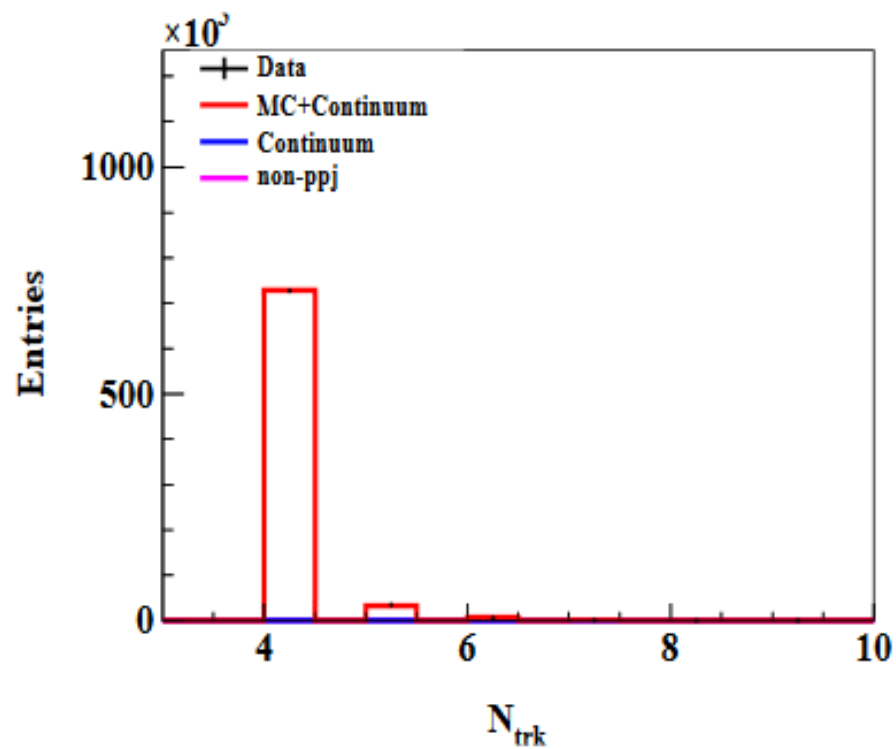
e 与 μ 的不变质量分布有什么差别？为什么？

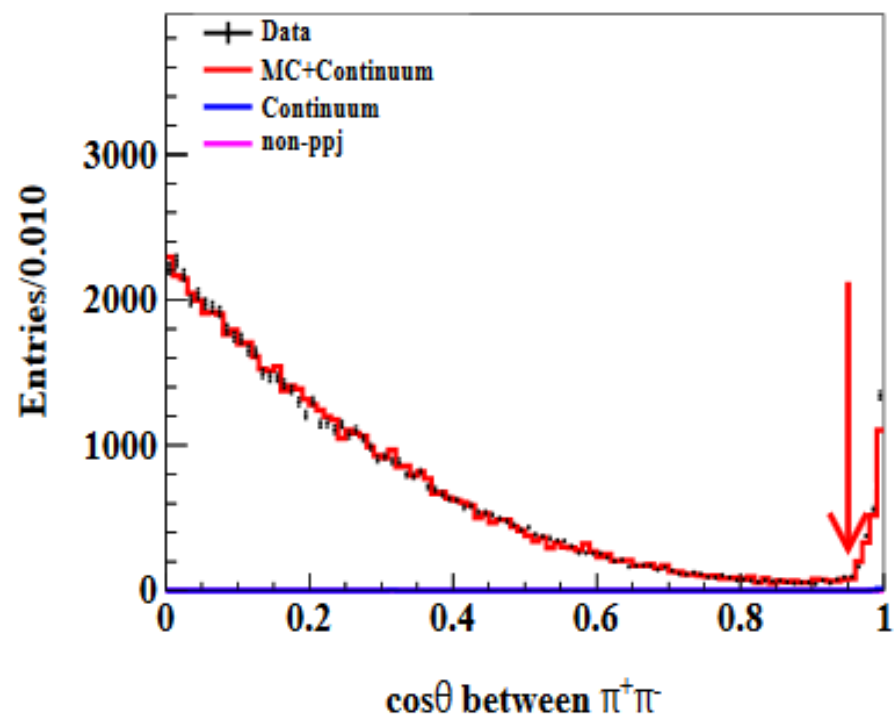
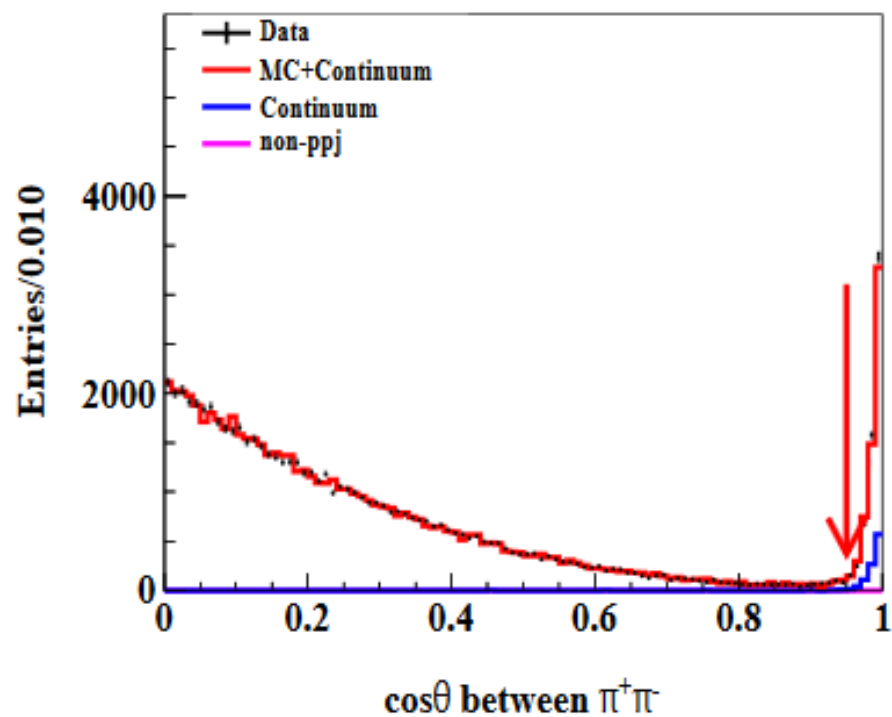


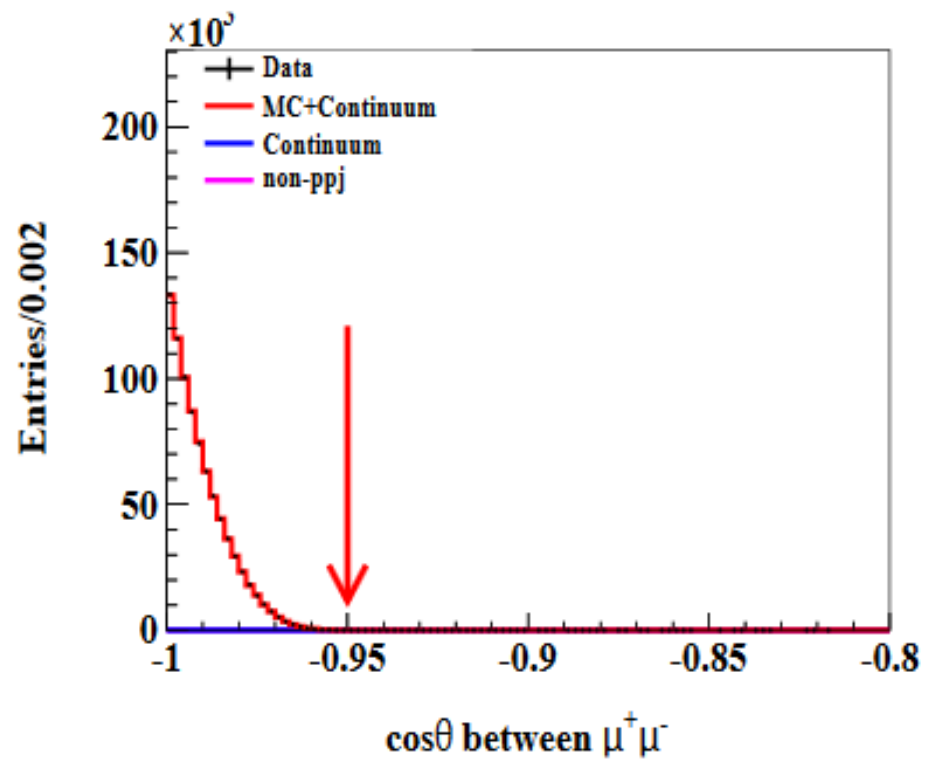
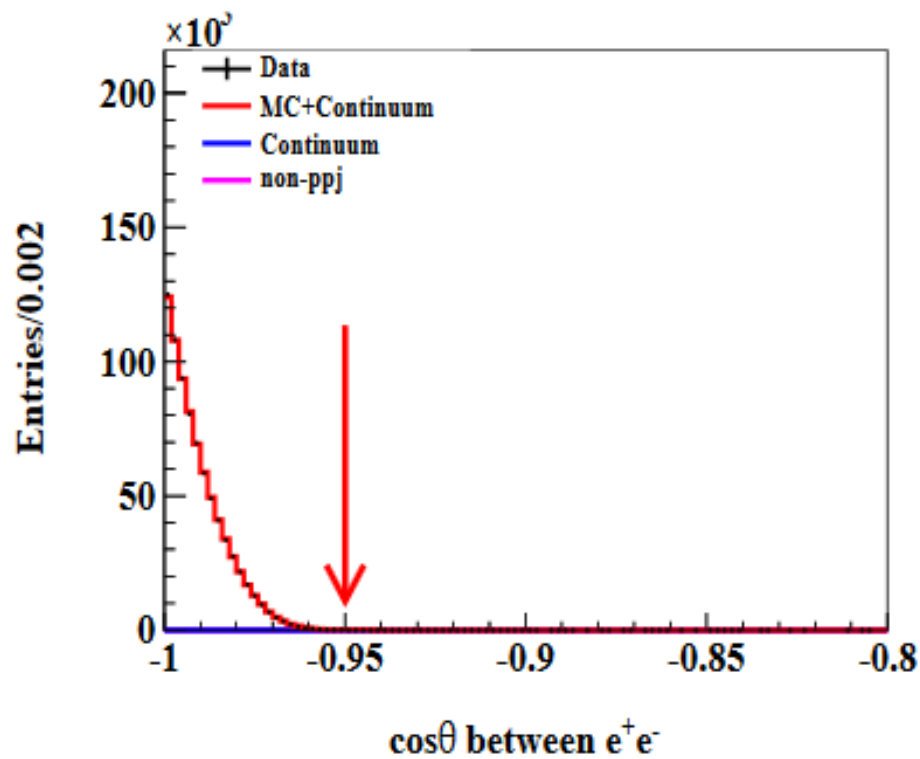
$$\frac{\delta m_{12}}{m_{12}} = \sqrt{\left[\frac{\delta E_1}{2E_1}\right]^2 + \left[\frac{\delta E_2}{2E_2}\right]^2 + \left[\cot \frac{\theta}{2} \frac{\delta \theta}{2}\right]^2}$$

与 $\pi\pi$ 反冲比较，轻子对的不变质量分布有什么不同？
低动量的 track 分辨好
分辨差一个量级！

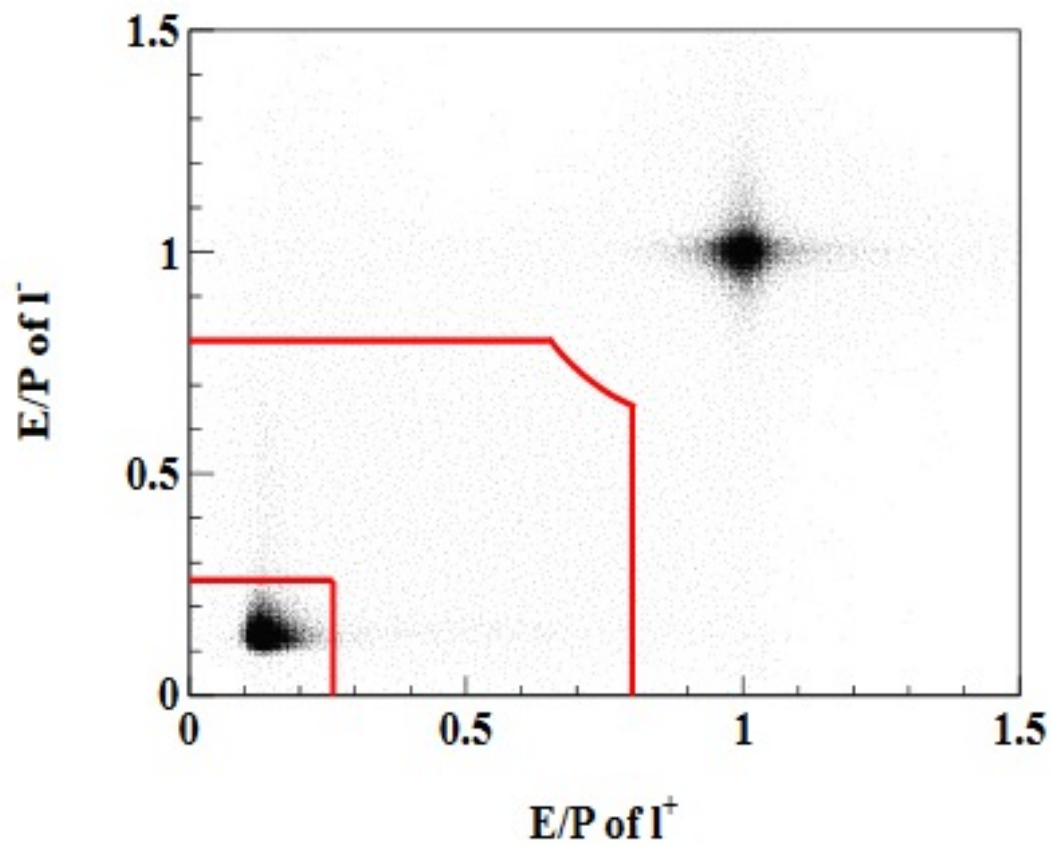








轻子对之间的夹角与 $\pi^+\pi^-$ 之间的夹角有什么明显不同？为什么？



术语园地
EMC
(ElectroMagnetic Calorimeter)
电磁量能器

小结

- ☑ 事例选择的目的是：提高信噪比
- ☑ 三条原则：
 - ☑ 尽可能简单的事例选择
 - ☑ 尽可能在保证效率的同时压低本底
 - ☑ 尽可能的利用运动学特征
- ☑ 作为小白该如何入手？
 - ☑ 思考
 - ☑ 阅读备忘录（memo）与文章
 - ☑ 积极参加组会、研讨会、合作组会议
 - ☑ 有疑问时向老师与同学请教

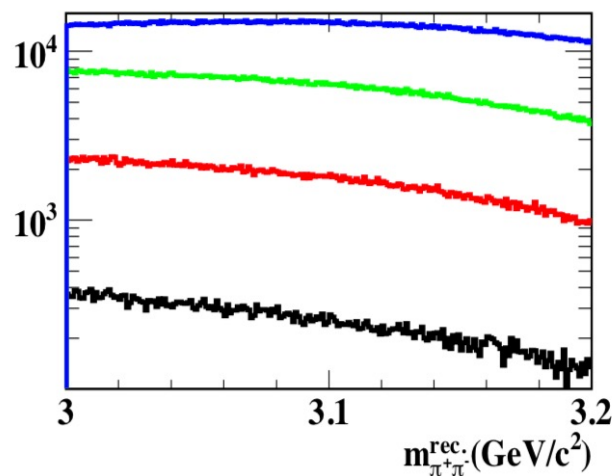
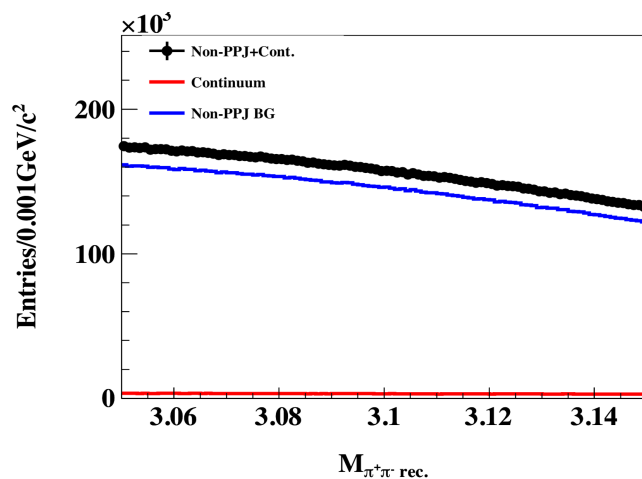
本底分析

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X} = \frac{N^{obs} - N^{bg}}{\epsilon \cdot N^X}$$

- 本底分析与事例选择常常是一个迭代的过程
 - 针对特定的本底，有时可以用简单有效的方法去除（veto）
- 本底须知：
 - 形状是否平滑？有没有类似信号的起伏（峰状本底）？
 - 本底数量大小？
- 确定本底的主要方法：
 - Inclusive MC: **topology**（a fantastic tool at BESIII）
 - 数据驱动（data driven）：其他数据、边带法（sideband）、检查分布
 - 猜测（经验依赖）

$\pi^+\pi^-J/\psi$ 过程的本底

- ✓ 平滑，没有 peak
- ✓ 受益于事例选择中：一个事例（event）可以有多条记录（entries）
- ✓ 虽然错误组合更多了，但是避免了人为的偏差（bias）
- ✓ 进一步检查不同的 J/ψ 衰变的误组合，仍然平滑



$J/\psi \rightarrow e^+ e^-$ 测量的本底

☑ MC topology 分析工具（首创自杜书先，由李刚等改进）

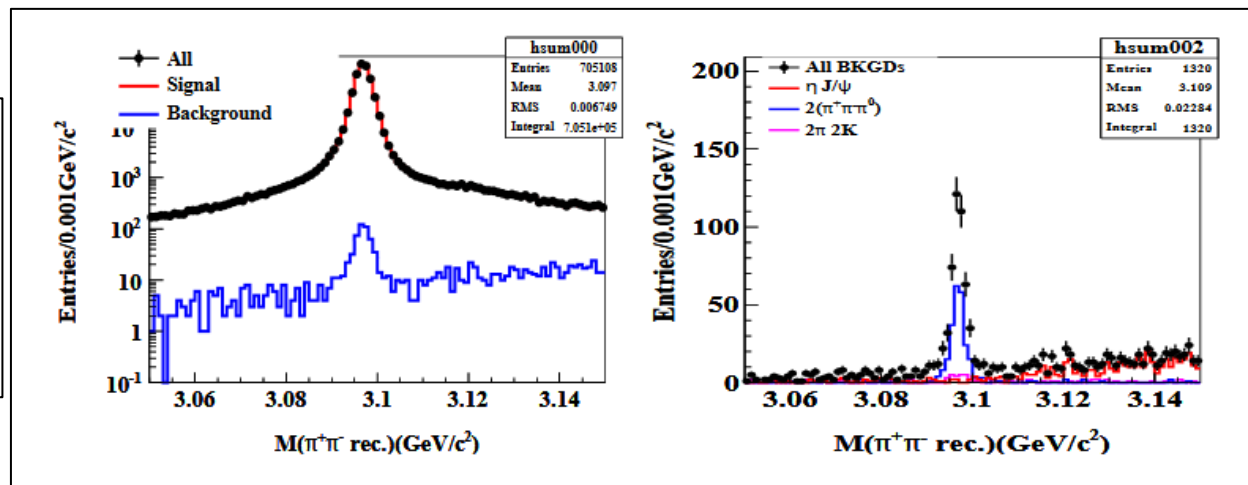
No.	decay chain	iTopo	N_{evt}
$e^- \pi^- e^+ \pi^+$ (final state)			
0	$\psi' \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow e^- e^+$	0	566983
1	$\psi' \rightarrow \pi^- J/\psi \pi^+, J/\psi \rightarrow e^+ \gamma_{FSR} e^-$	1	118425
2	$\psi' \rightarrow \pi^- J/\psi \pi^+, J/\psi \rightarrow \gamma_{FSR}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

No.	decay chain	iTopo	N_{evt}
$e^- \pi^- e^+ \pi^+ \gamma$ (final state)			
5	$\psi' \rightarrow \eta J/\psi, \eta \rightarrow \gamma \pi^+ \pi^-, J/\psi \rightarrow e^- e^+$	7	269
10	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow e^- \gamma_{FSR} e^+, \eta \rightarrow \gamma \pi^+ \pi^-$	25	54
45	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow \gamma_{FSR} e^+ \gamma_{FSR} e^-, \eta \rightarrow \pi^- \pi^+ \gamma$	192	3
326 evts (3 modes)			

Peaking backgrounds!

在本例中无需恐慌

本底水平很低：0.1%水平



以 $J/\psi \rightarrow \mu^+ \mu^-$ 作为对比看 FSR

比 ee 末态含有更少的末态辐射效应

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

思考题

- 分析换成测量 $\psi(2S) \rightarrow \pi^0\pi^0 J/\psi$ 的分支比

1. 这个过程的末态是什么？
2. 分支比的计算公式？
3. 事例选择？

提示：一个 π^0 衰变成两个光子 ($\pi^0 \rightarrow 2\gamma$)

提取信号数

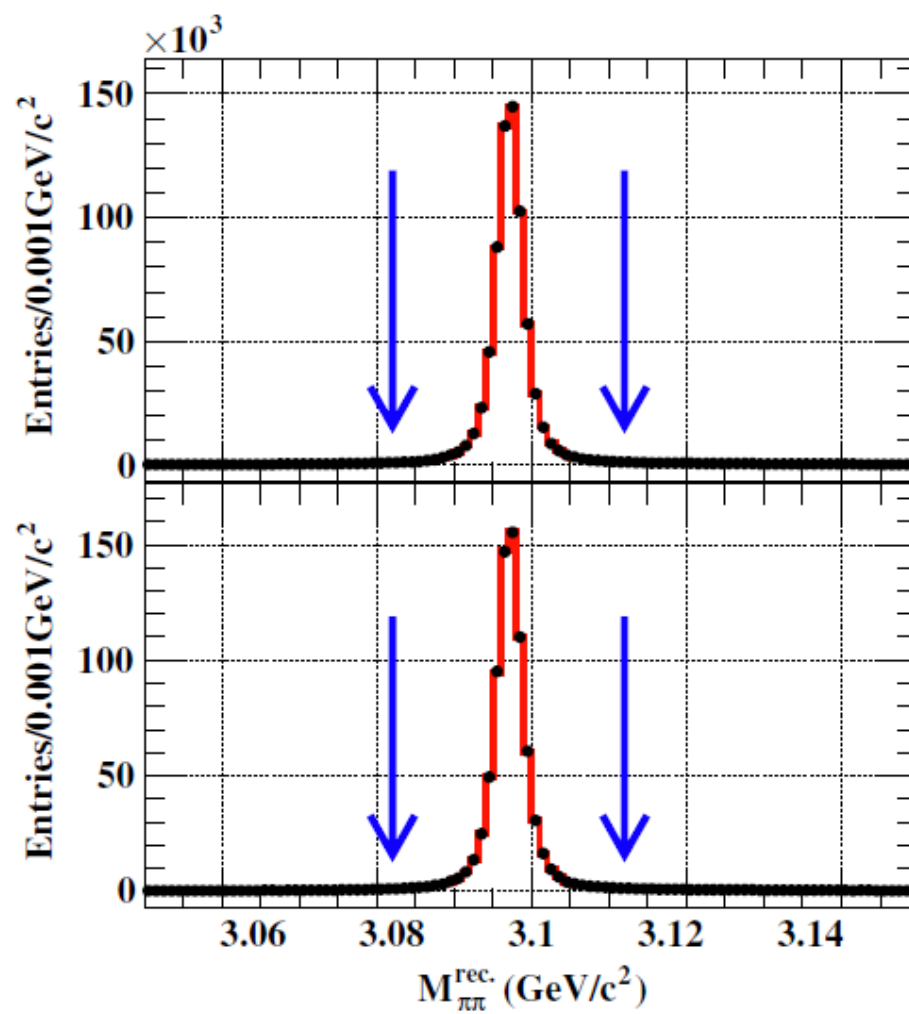
- 观测数减去本底数 ($N_{sig} = N_{obs} - N_{bg}$)
 - Inclusive MC
 - 边带法 (sideband)
- 拟合

利用Inclusive MC 提供的信息

No.	decay chain	iTopo	N_{evt}
	$e^- \pi^- e^+ \pi^+ \gamma$ (final state)		
5	$\psi' \rightarrow \eta J/\psi, \eta \rightarrow \gamma \pi^+ \pi^-, J/\psi \rightarrow e^- e^+$	7	269
10	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow e^- \gamma_{FSR} e^+, \eta \rightarrow \gamma \pi^+ \pi^-$	25	54
45	$\psi' \rightarrow J/\psi \eta, J/\psi \rightarrow \gamma_{FSR} e^+ \gamma_{FSR} e^-, \eta \rightarrow \pi^- \pi^+ \gamma$	192	3
326 evts (3 modes)			

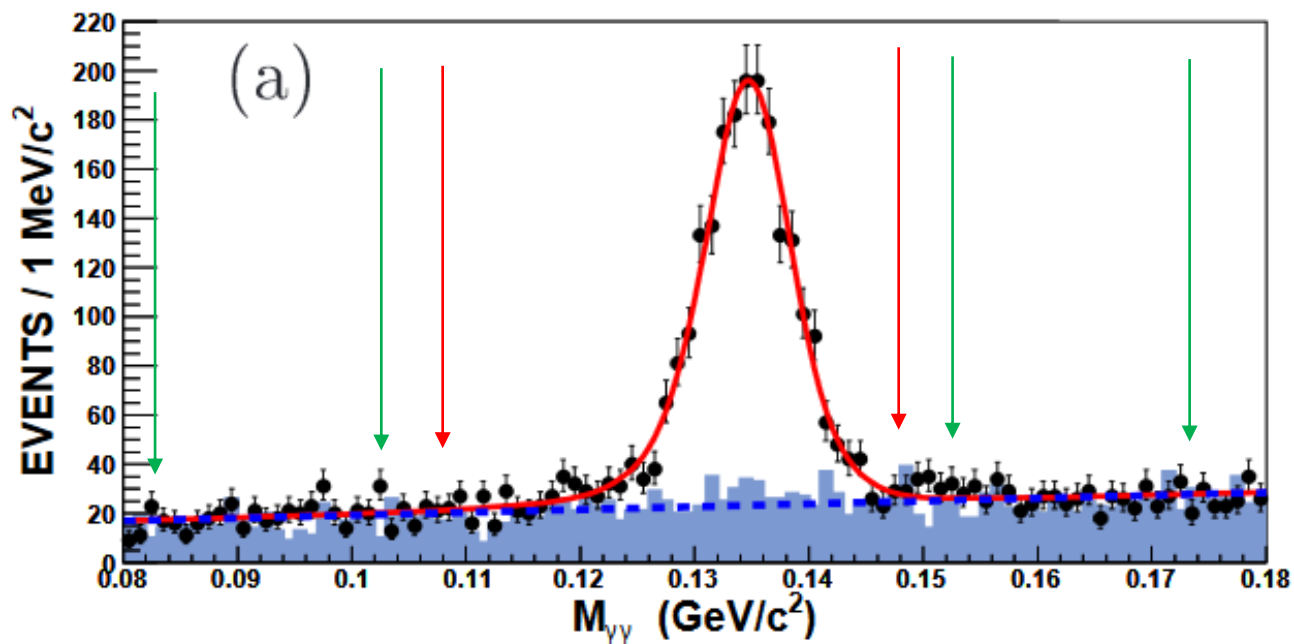
$$J/\psi \rightarrow l^+ l^-:$$

$$N_{obs} - N_{bg}$$

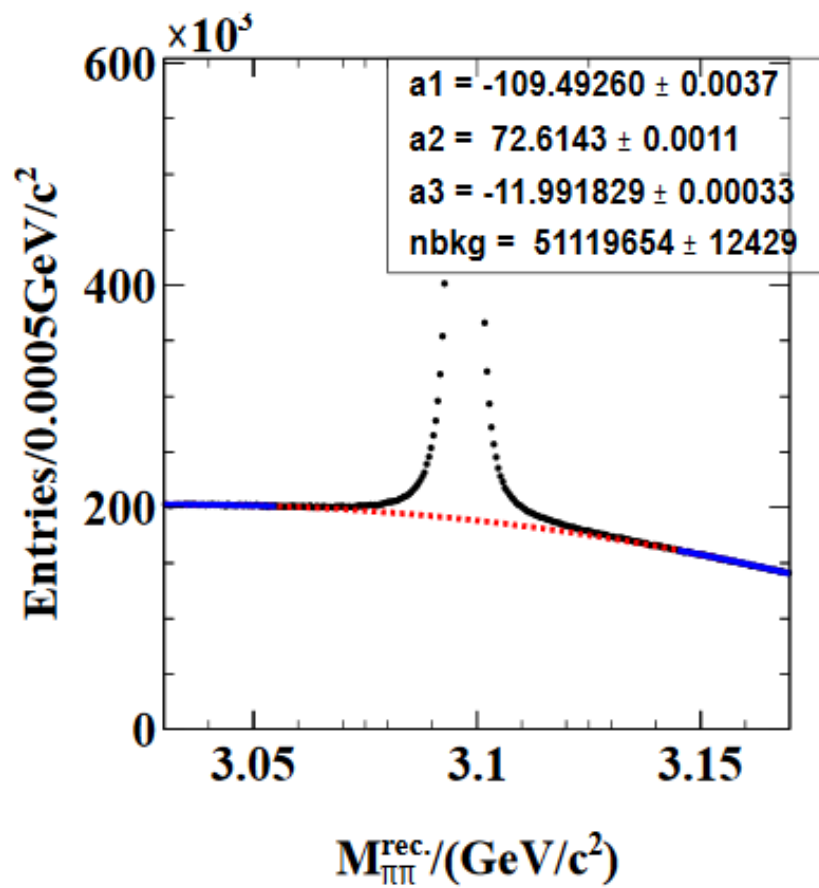


边带法 (sideband)

- 注意：
仅适用于平滑本底



拟合边带（非典型） $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi$



二维边带

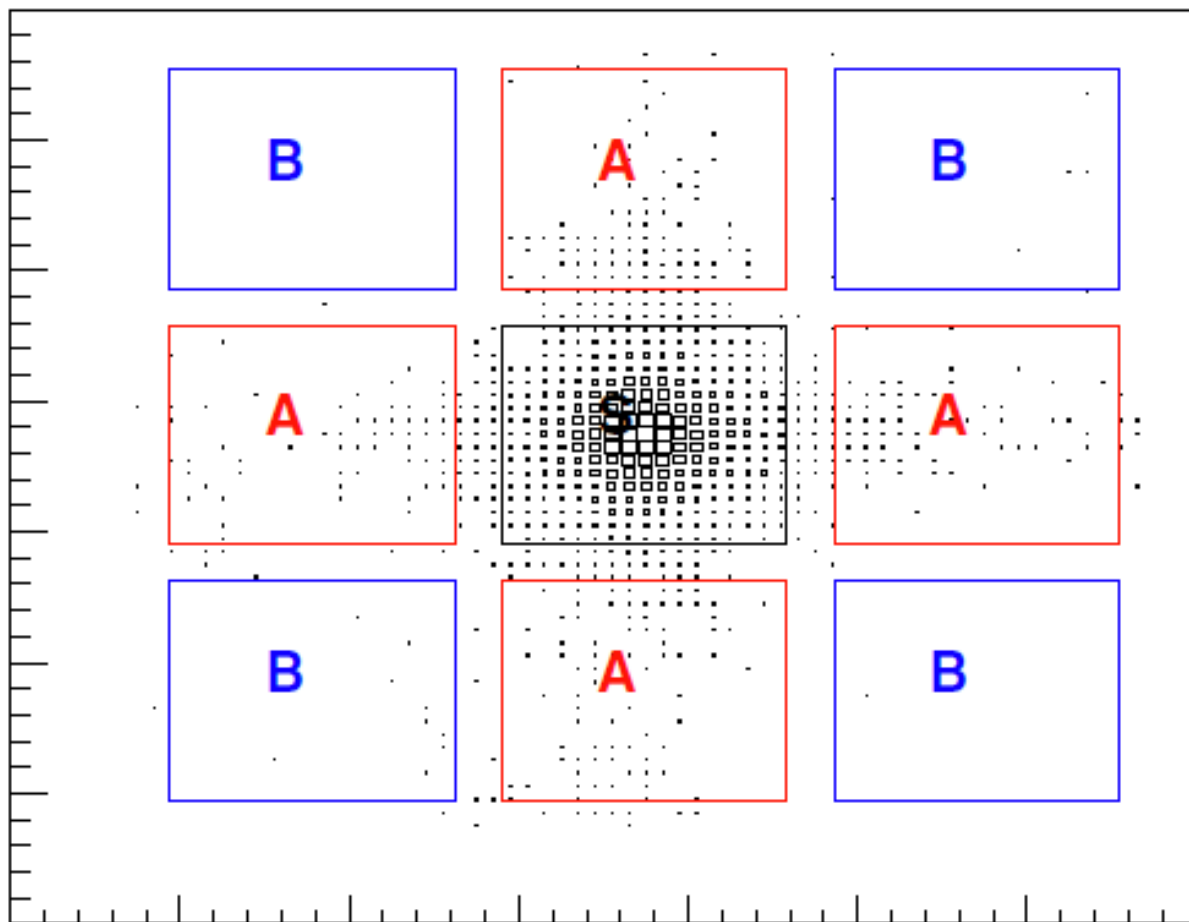
每一维对应一个共振态
的不变质量谱分布

常用公式：

$$S = \frac{A}{2} + \frac{B}{4}$$

(S、A、B面积相等)

为什么是这个公式？



拟合法

$$M_E(x) = N_S \cdot S(x) + N_B \cdot B(x)$$

- RooFit 几乎能满足一切需求
- 主要工作：

确定描述信号与本底的函数

需要对物理与探测器有一定程度的了解

术语园地

PDF (probability density function)
对应于量子力学里的 $|\mathbf{A}|^2$

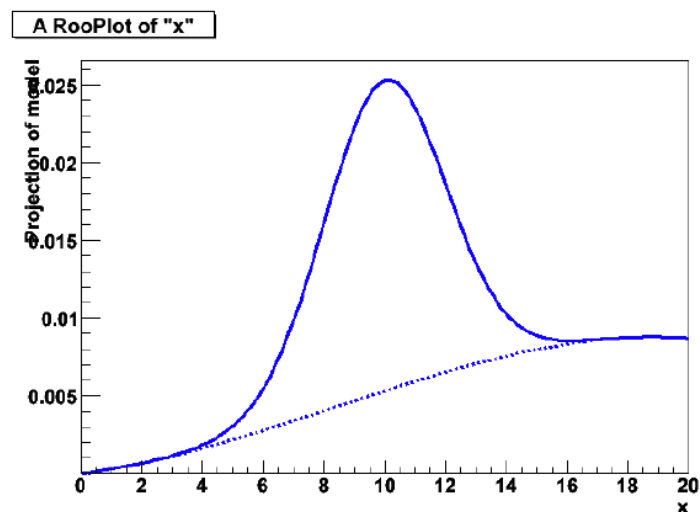


Figure 6 – Drawing of composite model and its components

Name	Functional form	Class name
Gaussian	$\exp\left(-0.5\left(\frac{x-m}{s}\right)^2\right)$	RooGaussian(name,title,x,m,s)
Exponential	$\exp(a \cdot x)$	RooExponential(name,title,x,a)
Polynomial	$1 + \sum_{i=1,n} a_i x^i$	RooPolynomial(name,title,x,alist)
Chebyshev polynomial	$1 + \sum_{i=1,n} a_i T_i(x)$	RooChebychev(name,title,x,alist)

Table 1 – Basic functions implemented in RooFit

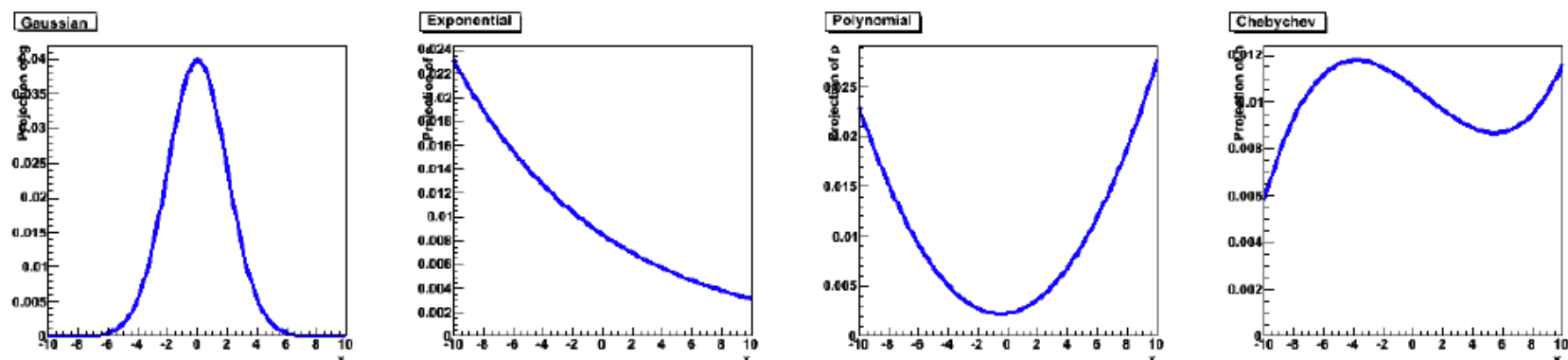
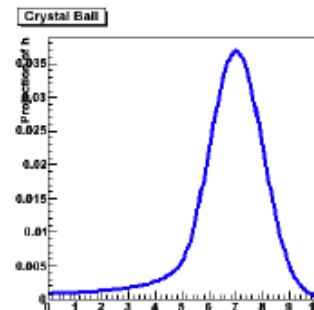
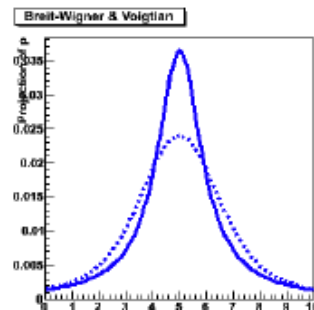
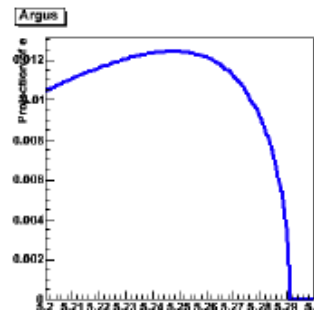
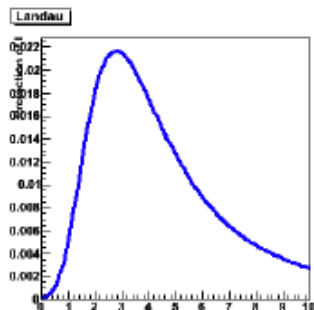


Figure 7 – Basic p.d.f shapes: Gaussian, Exponential, Polynomial and Chebyshev polynomial

Landau	$\text{TMath::Landau}(x, \text{mean}, \text{sigma})$	$\text{RooLandau}(\text{name}, \text{title}, x, \text{mean}, \text{sigma})$
Argus	$x \left(1 - \left(\frac{x}{m}\right)^2\right)^p \cdot \exp\left(c \left(1 - \left(\frac{x}{m}\right)^2\right)\right)$	$\text{RooArgusBG}(\text{name}, \text{title}, x, m, c, p)$
Breit-Wigner	$\frac{1}{(x-m)^2 + \frac{1}{4}g^2}$	$\text{RooBreigWigner}(\text{name}, \text{title}, x, m, g)$
Voigtian	$\frac{1}{(x-m)^2 + \frac{1}{4}g^2} \otimes \exp\left(-\frac{1}{2}\left(\frac{x}{s}\right)^2\right)$	$\text{RooVoigtian}(\text{name}, \text{title}, x, m, g, s)$
Crystal Ball	$\frac{\left(\frac{n}{ a }\right)^n e^{-\frac{1}{2}a^2}}{\left(\frac{n}{ a } - a - x\right)^n} \Big _{x < - a }, \exp\left(-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right) \Big _{x > - a }$	$\text{RooCBSShape}(\text{name}, \text{title}, x, m, s, a, n)$
Decay	$\exp(- x /\tau) \otimes R(x)$	$\text{RooDecay}(\text{name}, \text{title}, x, \text{tau}, R)$



$\psi(2S) \rightarrow \pi^+ \pi^- J/\psi$ 的拟合

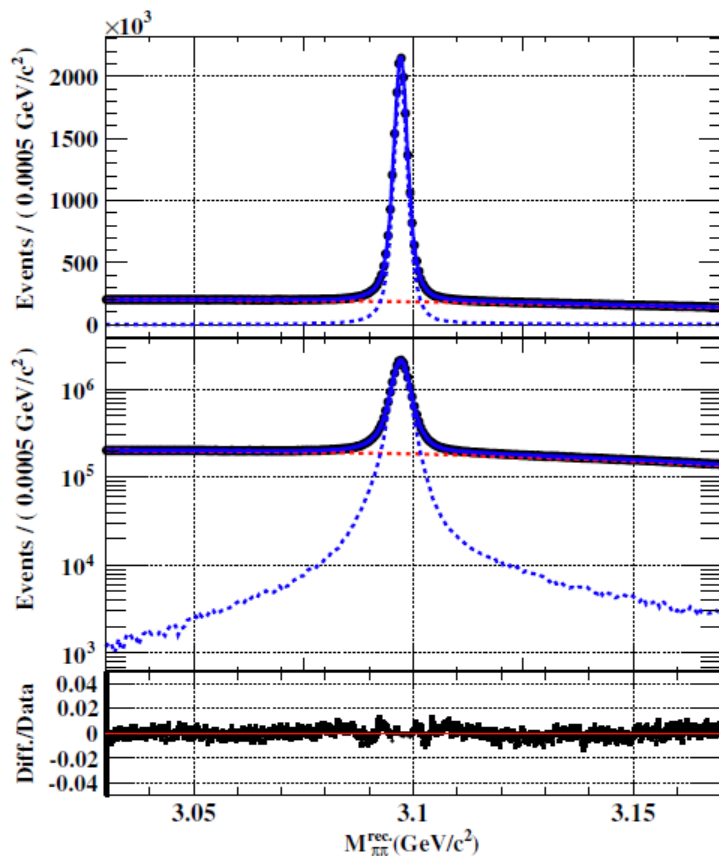


FIG. 3 (color online). The dipion recoil mass spectrum for $\psi(3686) \rightarrow \pi^+ \pi^- J/\psi$, $J/\psi \rightarrow \text{anything}$. Top: Data points (black) overlaid with the fit result (solid blue curve) obtained using the signal shape from $\psi(3686) \rightarrow \pi^+ \pi^- J/\psi$, $J/\psi \rightarrow e^+ e^-$ (blue dashed curve) and a second-order polynomial background shape (red dashed curve). Middle: The same plot as the top but with a log scale. Bottom: The fractional difference between the fit and the data.

信号形状采用的是无参数PDF

术语园地

χ^2 与 likelihood 拟合

$$\chi^2: \sum_i \frac{(m_i - f_i)^2}{\sigma_i^2}, \quad \mathcal{L}: \prod_i P_i$$

检查拟合是否成功、结果是否合理

```
*****
** 13 **MIGRAD      1000      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=25139.4 FROM MIGRAD  STATUS=INITIATE  10 CALLS  11 TOTAL
EDM= unknown STRATEGY= 1 NO ERROR MATRIX
EXT PARAMETER      VALUE      ERROR      STEP      FIRST
NO.  NAME          VALUE      ERROR      SIZE      DERIVATIVE
1  mean            -1.00000e+00  1.00000e+00  1.00000e+00  -6.53357e+01
2  sigma           3.00000e+00  1.00000e+00  1.00000e+00  -3.60009e+01
ERR DEF= 0.5
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=25137.2 FROM MIGRAD  STATUS=CONVERCED  33 CALLS  34 TOTAL
EDM=8.3048e-07 STRATEGY= 1 ERROR MATRIX ACCURATE
EXT PARAMETER      VALUE      ERROR      STEP      FIRST
NO.  NAME          VALUE      ERROR      SIZE      DERIVATIVE
1  mean            -9.40910e-01  3.03997e-02  3.32893e-03  -2.95416e-02
2  sigma           3.01575e+00  2.22446e-02  2.43807e-03  5.98751e-03
ERR DEF= 0.5
EXTERNAL ERROR MATRIX.  NDIM= 25  NPAR= 2  ERR DEF=0.5
9.241e-04 -1.762e-05
-1.762e-05 4.948e-04
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL  1  2
1  0.02606  1.000 -0.026
2  0.02606 -0.026 1.000
*****
** 18 **HESSE      1000
*****
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=25137.2 FROM HESSE  STATUS=OK  10 CALLS  44 TOTAL
EDM=8.30707e-07 STRATEGY= 1 ERROR MATRIX ACCURATE
EXT PARAMETER      VALUE      ERROR      STEP      FIRST
NO.  NAME          VALUE      ERROR      SIZE      DERIVATIVE
1  mean            -9.40910e-01  3.04002e-02  6.65786e-04  -9.40910e-01
2  sigma           3.01575e+00  2.22449e-02  9.75228e-05  3.01575e+00
ERR DEF= 0.5
EXTERNAL ERROR MATRIX.  NDIM= 25  NPAR= 2  ERR DEF=0.5
9.242e-04 -1.807e-05
-1.807e-05 4.948e-04
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL  1  2
1  0.02672  1.000 -0.027
2  0.02672 -0.027 1.000
```

1. 检查输出信息
2. 估计统计误差
3. 检查拟合优度

术语园地

拟合优度 goodness of fit

对一个好的拟合有 $\chi^2/NDF \approx 1$

自由度: NDF (number of freedom)

NDF = No. of Bin – No. of paras

检查拟合是否成功、结果是否合理 II

MIGRAD MINIMIZATION HAS CONVERGED.

MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.

COVARIANCE MATRIX CALCULATED SUCCESSFULLY

FCN=25137.2 FROM MIGRAD STATUS=CONVERGED 33 CALLS 34 TOTAL
EDM=8.3048e-07 STRATEGY= 1 ERROR MATRIX ACCURATE

EXT	PARAMETER			STEP	FIRST
NO.	NAME	VALUE	ERROR	SIZE	DERIVATIVE
1	mean	-9.40910e-01	3.03997e-02	3.32893e-03	-2.95416e-02
2	sigma	3.01575e+00	2.22446e-02	2.43807e-03	5.98751e-03

ERR DEF= 0.5

EXTERNAL ERROR MATRIX. NDIM= 25 NPAR= 2 ERR DEF=0.5

9.241e-04 -1.762e-05

-1.762e-05 4.948e-04

PARAMETER CORRELATION COEFFICIENTS

NO.	GLOBAL	1	2
1	0.02606	1.000	-0.026
2	0.02606	-0.026	1.000

拟合小结

- 拟合是对前期分析工作已获得信息的集大成，包含了分析者对
 - 信号形状的认识：物理模型、数据样本的大小、探测器分辨、效率曲线、测量的中心值是否有偏等
 - 本底的了解：来源与大小、是否有未知/不理解的本底
- 多关注如何构建模型，这是知识的汇总

术语园地

分波分析（PWA: partial wave analysis）
带着复杂的物理模型、参数可多达上百的拟合

效率

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X}$$

什么是效率

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot B_{int} \cdot N^X}$$

把中间态分支比单列出来的定义

- 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、中间态的分支比等。
- 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。
- 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。

术语园地

接收度 acceptance

效率 efficiency

探测效率 detection efficiency

什么是效率 (II)

- 一般地

$$\epsilon_{DT} = \epsilon_{DT}^{Geom} \cdot \epsilon_{DT}^{Track} \cdot \epsilon_{DT}^{Ptid} \cdot \epsilon_{DT}^{\gamma ID} \dots$$

$$\epsilon_{MC} = \epsilon_{MC}^{Geom} \cdot \epsilon_{MC}^{Track} \cdot \epsilon_{MC}^{Ptid} \cdot \epsilon_{MC}^{\gamma ID} \dots$$

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

通过研究得到 f^i 即可对 MC 所得到的效率进行修正，使之与数据的真实效率更符合

一个 cut flow 的例子

估一估

- $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi$
 - $J/\psi \rightarrow l^+ l^-$
- 的效率各是多少？

Criteria	$\psi(2s) \rightarrow \Sigma^0 \bar{\Lambda}$	$\psi(2s) \rightarrow \bar{\Sigma}^0 \Lambda$
Total Events	80000	80000
$N_{good} = 4, N_{charge} = 0$	44198	43983
$N_p = 1, N_{\bar{p}} = 1, N_{\pi^+} = 1, N_{\pi^-} = 1$	43856	43643
$1 \leq N_\gamma \leq 10$	38075	38146
$\theta(\gamma, \bar{p}) > 20^\circ$	36491	34538
$\Lambda, \bar{\Lambda}$ reconstruction	34354	32564
Pass 6C	25192	22390
$\chi^2_{6C}(\gamma) < \chi^2_{6C}(\gamma\gamma/0\gamma)$	24681	21850
$\chi^2_{6C} < 25$	18368	15447
$L_\Lambda/\sigma > 2 (\bar{\Lambda})$	16084	13502
$M_{\Lambda\bar{\Lambda}} > 3.48 \text{ GeV}/c^2$	15891	13498
$M_{\gamma\bar{\Lambda}}(M_{\gamma\Lambda}) \geq 1.15 \text{ GeV}$	15870	13488
Efficiency ($\epsilon\%$)	19.84%	16.86%

Table 2: Cut flow of $\psi(2S) \rightarrow \bar{\Sigma}^0 \Lambda + c.c.$

关于效率估计的说明

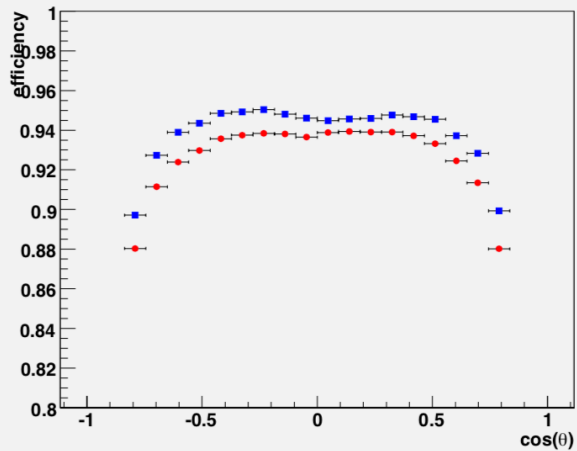
- 数据与 MC 的不一致永远存在，因此利用 MC 得到的效率永远需要修正，但对于不同的过程要求也不同
 - 对于小统计量情况，统计误差为主，数据与信号模拟的差别对结果影响很小，但对本底估计时需特别关注（稀有过程、新物理）
 - 对于大统计量情形，期望高精度结果，数据与信号模拟的差别不容忽视

径迹效率修正

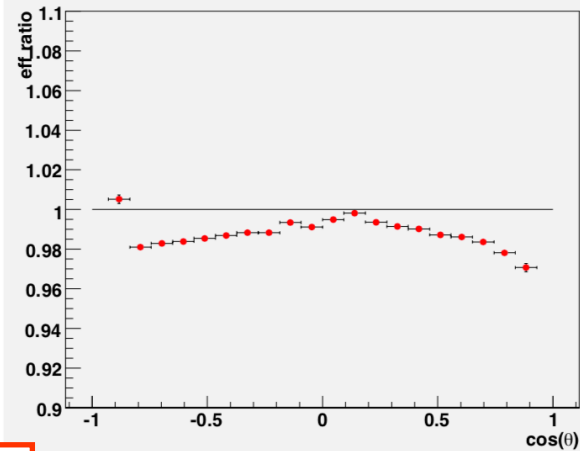
$$B_{\pi\pi J/\psi} = \frac{N_{\pi\pi J/\psi}}{\epsilon_{MC} N^{tot}} \quad \epsilon_{MC} = \epsilon_{\pi^+} + \epsilon_{\pi^-} + \epsilon_{m_{\pi\pi}} \dots ?$$

- ☑ 径迹效率研究
 - ☑ 给出数据与MC径迹效率关于动量与角度的二维分布
 - ☑ 在分析中用 data/MC 之比对 MC 做了 reweight（修正）
 - ☑ 精度控制在 0.1%

Tracking efficiency of pionm_

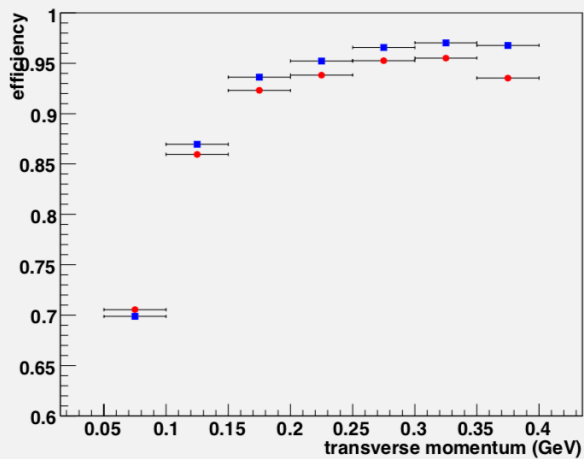


Tracking efficiency difference e_data/e_mc

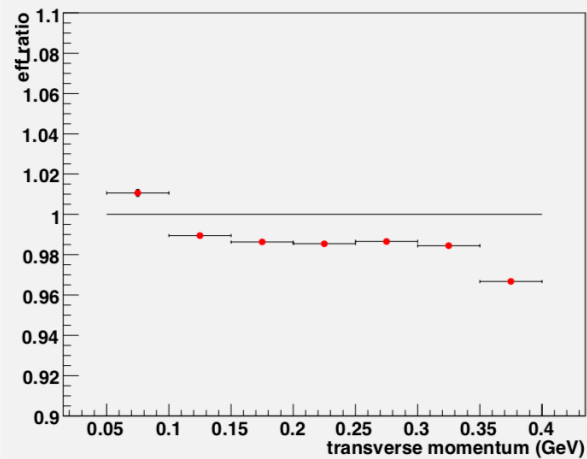


pi+

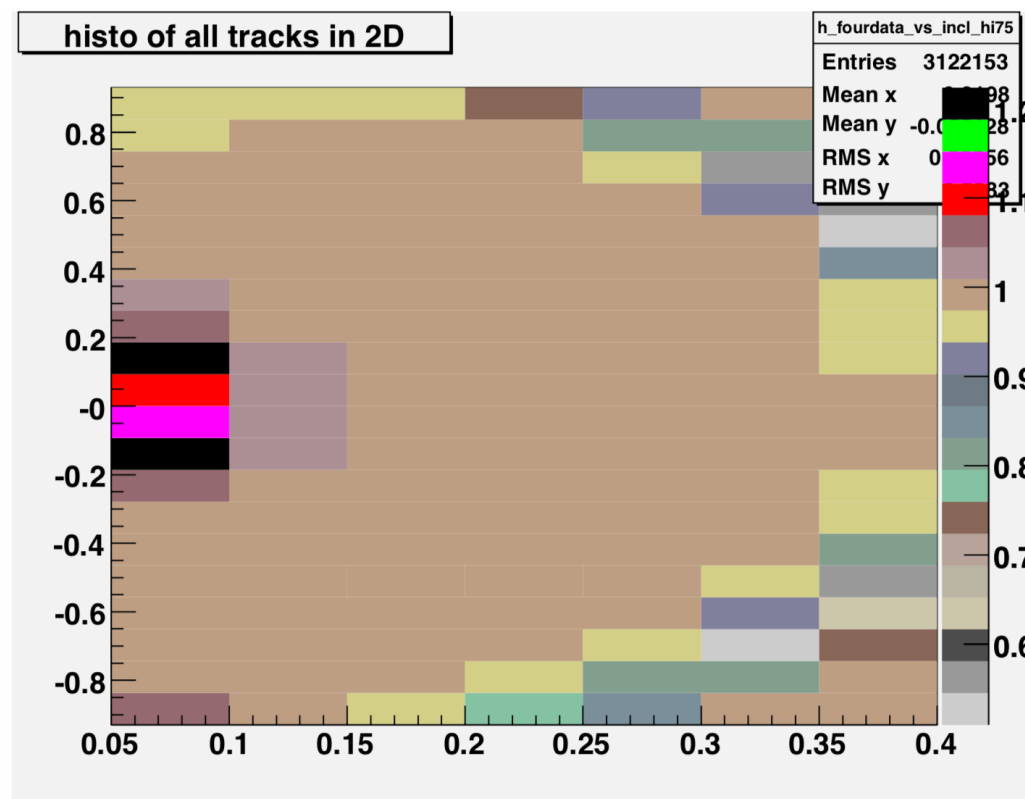
Tracking efficiency of pionm_



Tracking efficiency difference e_data/e_mc

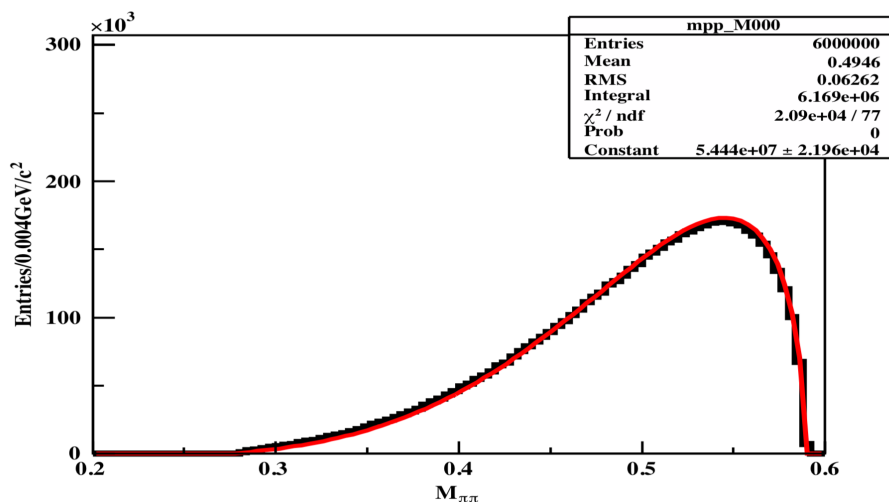


数据与MC二维径迹效率比（reweight 的依据）



谱形分布的效率修正

谱形影响动量分布进而影响效率



- MC 采用了 JPIPI, g 和 g_1 来自 BES1 的测量:

$$g(s - 2m_{\pi\pi}^2) + 2g_1 E_{\pi^+} E_{\pi^-} \epsilon_{\psi'}^*(M) \cdot \epsilon_{\psi}(\Lambda)$$

- 最终采用数据（轻子道的干净样本）来 reweight MC $M_{\pi\pi}$ 分布

多重数分布的效率修正

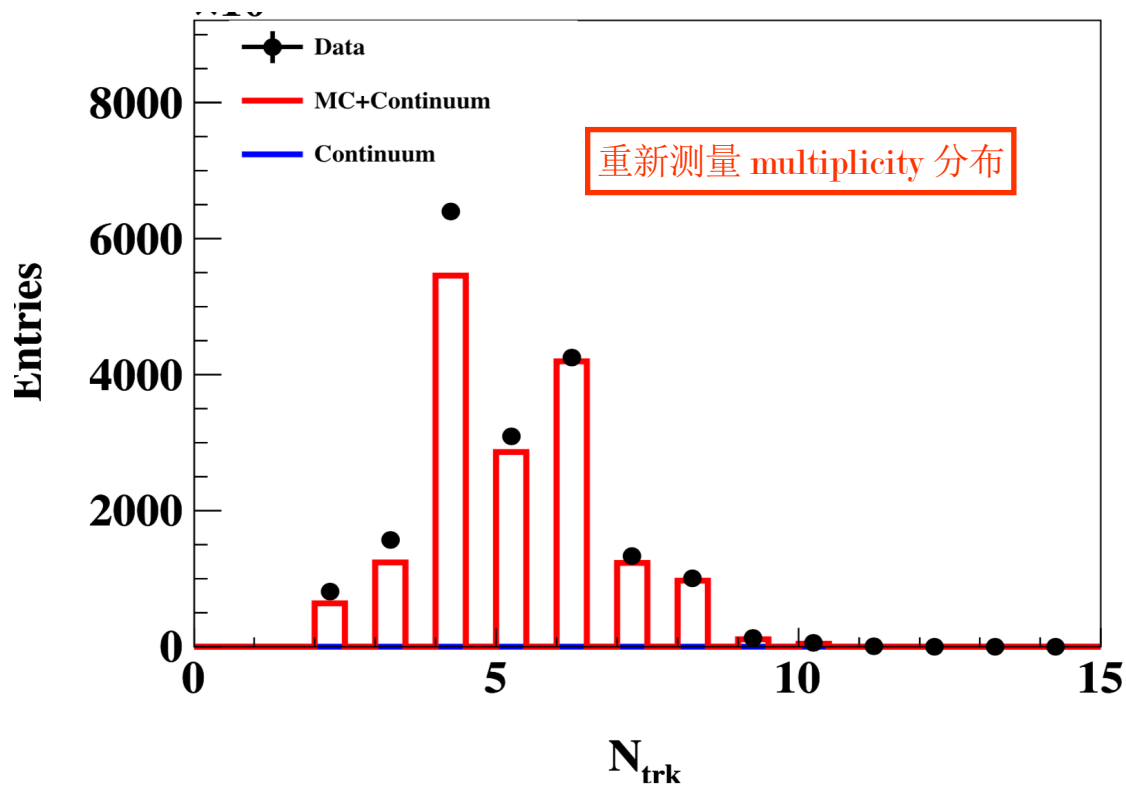
○高统计量的新问题:

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,

...

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 的多重数分布符合不佳



解方程

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

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

测量 $B(J/\psi \rightarrow l^+ l^-)$ 时的相消法

- 常规方法: $B = \frac{N^{obs}}{\epsilon N^{tot}}$
- 这里 $B_{ll} = \frac{N_{\pi\pi ll}}{\epsilon_{\pi\pi ll} N_{tot}} / \frac{N_{\pi\pi}}{\epsilon_{\pi\pi} N_{tot}} = \frac{N_{\pi\pi ll}}{N_{\pi\pi}} \frac{\epsilon_{\pi\pi}}{\epsilon_{\pi\pi ll}}$
- 因此
 - 总数 N_{tot} 可以消掉
 - 原则上 π 的径迹效率也可以相消, $\epsilon_{\pi\pi ll} = \epsilon_{\pi\pi} \epsilon_{ll}$
 - 只剩下高动量的轻子需要仔细考虑, 而高动量的粒子效率高 (因此误差小)

回顾

$$B(X \rightarrow f) = \frac{N^{sig}}{\epsilon \cdot N^X}$$

- 推荐阅读：我关于 $\pi^+\pi^-J/\psi$ 分析的 memo 与 draft
- 在哪？上 HyperNews 或 Documentation 去找
- 它们又在哪？bes3.ihep.ac.cn 主页右下角有链接

当分析工作遇到困难时

重读《西游记》

