

## 第8章 盲 分 析

### 8.1 为什么需要盲分析

我们来看一看粒子物理中实际发生的情形.

例子之一: 粒子表<sup>[1]</sup> 中粒子的某些性质参数随着测量年代会发生变化 (参见图 9.2). 这些图显示出同一物理量在一段时间内似乎相对稳定 (误差范围内一致), 然后其中心值会发生突然的跳跃, 然后在新的水平上相对稳定, 然后中心值再次跳跃, …… 这种现象很难说不存在实验者有意或无意的偏差.

例子之二是图 8.1 所示的 LEP 实验不同实验组的  $R_C$  测定值<sup>[123]</sup>. 利用图中上部的 ALEPH、DELPHI、OPAL 实验的 8 个独立测量数据, 通过最小二乘法 (式 (6.2.5)~式 (6.2.8)) 求得最小二乘函数  $Q^2(R_C) \equiv x^2$  为 0.92, 故测量数据的  $\chi^2$ /自由度为  $0.92/7$ , 显著小于各测量相互独立所预期的  $x^2$ /自由度数值. 其可能的原因之一 (尽管实验者恐怕不同意这种看法) 是过低估计了系统误差. 为了检验是否有这种可能性, 略去系统误差再计算  $\chi^2$ /自由度, 其值为  $2.1/7$ , 仍然过小. 造成这种现象的可能原因是, 实验者“有意无意”地使实验结果相互靠近或向标准模型靠近.

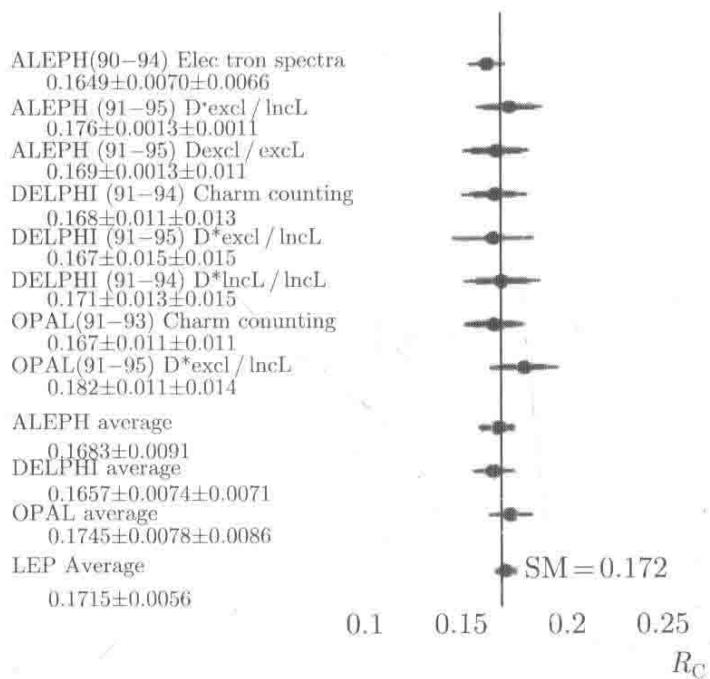


图 8.1 LEP 实验的  $R_C$  测定值

用不着举出更多的实例, 实验者在某种情形下“有意无意”调节实验结果的做法显然是存在的. 这种做法往往出现于如下的情形中:

- (1) 当结果与某种重要的理论 (如标准模型) 的预期值不符时;
- (2) 当结果与有良好声誉的实验组的已有结果不符时;
- (3) 当结果与实验者希望检验的新理论的预期值不符时;
- (4) 当同一实验组利用两个反应道或两种方法对同一课题的研究结果相互不符时.

这些情况下, 实验者往往希望通过分析方法的“合理”调节来“修正”结果或者寻找额外的系统误差来源; 但当实验结果与“预期值”一致时, 一般不会这样做.

实验者将实验结果改变或调节为预想的值可能通过多种“合理”途径. 就粒子物理实验数据分析而言, 可以设想到的“合理”途径至少可以包含以下可能性:

- (1) 信号事例选择条件的不同可以改变结果; 同一组信号事例选择条件但利用不同截断值的也可改变结果.
- (2) 粒子性质的模拟需要用到 MC 模拟, 对同一种物理过程, 存在不同的模拟程序 (如电磁、强子簇射的不同模型及对应的程序包), 使用不同的模拟程序会获得不同的结果; 同一模型使用不同的截断参数也会获得不同的结果.
- (3) 选择不同版本的刻度文件可改变结果.

实验者通过分析方法的“合理”调节来“修正”结果所导致的这类偏差是无法定量化的系统误差, 使得实验结果存在无法控制的偏差, 偏离了测量结果的客观性.

显然, 我们希望能找到实验分析的一种方法或途径, 它能够避免这类人为导致的、无法控制的对于测量结果的偏差, 至少使得这种偏差尽可能地最小化. 盲分析就是可选方法之一.

## 8.2 盲分析需遵循的原则

盲分析是试图避免人为导致的、无法控制的对于测量结果的偏差, 因此就要阻断导致这类偏差的途径. 如前所述, 当测量结果与实验者预想不同时, 往往会企图改变分析方法. 因此盲分析的要素之一是将分析方法与测量结果的数值本身割裂开来. 因为实验测量结果的数值本身并不包含其数值正确性的任何信息, 所以实验测量结果的数值本身对于分析方法的确定不应该起作用. 分析方法的正确性应当用一系列外部的检查来验证. 其次, 数据分析中事例判选的截断值的不同选择, 有可能会导致不同的测量结果 (即有偏差). 盲分析应当确保事例判选的截断值的选择对测量结果具有无偏性.

粒子物理实验中, 一项分析所研究的过程称为信号过程, 信号过程往往集中出现于某几个特征物理量的特定区域内, 该区域称为信号区. 盲分析中, 在分析方法的所有重要步骤确定之前, 根据不同类型的实验测量, 一些实验信息被“隐蔽”起来; 这些实验信息包括: 信号区内某些特征物理量的实验原始数据, 或待测参数的

数值本身, 或者能直接导出待测参数数值的某些特征物理量的分布(图形)等。这样做的动机是为了避免分析者(自觉或不自觉的)采用对最终结果会造成某种偏向的分析方法。如前所述, 实验测量结果的数值本身并不包含其正确性的任何信息。因此, 隐蔽这些实验信息并不妨碍正确地确定实验数据分析方法的整个流程。

粒子物理实验分析中, 通常信号数据利用信号的 MC 模拟数据或实验数据的控制样本来表征, 本底则利用信号边带区数据或本底的 MC 模拟数据来表征。实验中的盲分析意味着, 在对隐蔽了的实验信息“去盲”之前, 分析者已经利用上面提到的这些数据样本确定了该项分析所需的所有方法、步骤和关键参数, 如确定了信号事例判选条件(特征量及其截断值的优化), 确定了信号函数的形状及信号探测效率, 确定了本底函数的形状甚至大小, 检验了拟合方法, 估计了系统误差等。换句话说, 在对隐蔽了的实验信息“去盲”之前, 数据分析的整个流程已经确定; 在“去盲”之后, 所进行的工作仅是按照既定的流程对包含信号区数据在内的全部数据进行一次完整分析, 并由此可直接给出最终结果(如分支比及误差)。

盲分析大致应当遵循以下原则和步骤:

- (1) 若对同一课题有多个分析组同时进行盲分析, 则应对盲分析的策略达成一致。
- (2) 在对隐蔽了的实验信息“去盲”之前, 各分析组应当报告盲分析的详细过程, 合作组其他成员有机会对其分析提出意见和建议。
- (3) 一个专家委员会决定是否需要对现有的分析进行修改。
- (4) 各分析组确定了各自的分析流程后, 专家委员会审查并同意对隐蔽了的实验信息“去盲”, 各组得出各自的测量结果。
- (5) 如果各组的结果不一致, 专家委员会应会同分析组确定改进的盲分析策略; 如果一致, 则确定最终的发表版本。

### 8.3 盲分析方法及实例

对于不同类型的实验测量, 需要采用不同的盲分析方法。

#### 8.3.1 信号盲区方法——稀有衰变和寻找新共振态的盲分析

所谓“稀有衰变”, 指的是衰变分支比尚未进行过测量或只有上限值的情形。这种情形下, 衰变分支比通常很小, 即使信号存在, 信号事例数往往很少而本底很大, 因此分析方法的本底压低因子应当尽可能大, 对信号具有合理的探测效率。这种情形下, 盲分析是非常必要的。

如果信号事例的判选是通过若干特征物理量的截断来实现的, 并且信号事例集中出现在某几个特征物理量的特定区域(称为信号区)、其位置事先可以由先前的

测量或通过信号 MC 模拟数据加以确定的情形下, 则信号盲区方法是十分适宜的, 并且几乎成为一种标准的分析方法。信号盲区是若干个 (通常不多于 2 个) 特征物理量数值处于特定上、下界的一个区域。考虑到探测器分辨及其不确定性的效应, 信号盲区应当比真实的信号区选得大一些, 以便为截断值的进一步优化留出余量。盲分析中, 信号盲区中的实验数据被“隐蔽”起来。信号盲区附近的边带区数据用来确定各特征物理量的本底分布。截断值的优化可以利用各特征物理量的本底分布以及信号 MC 数据样本的分布来进行。这里, 假定各特征物理量是不关联的, 因此某一特征物理量的边带区数据正确地代表了其他特征物理量在信号区中的本底数据。这一假定的正确程度可以用 MC 模拟加以检验。

信号盲区法曾经用于稀有衰变  $K_L^0 \rightarrow \mu^\pm e^\mp$  的寻找<sup>[124]</sup>, 这是最早利用盲分析的实验之一。标准模型预言应当存在这种衰变。该实验的灵敏度为  $10^{11}$  次  $K_L^0$  衰变可观测到一个  $K_L^0 \rightarrow \mu^\pm e^\mp$  衰变事例。两个特征物理量是  $M_{\mu e}$  和  $P_T^2$ 。信号事例预期应当出现于图 8.2 所示的小方框划定的盲区内。盲区内的事例数据在分析的所有重要步骤 (确定信号事例判选的所有截断值, 检验拟合方法, 估计系统误差) 完成之后才对分析者打开。

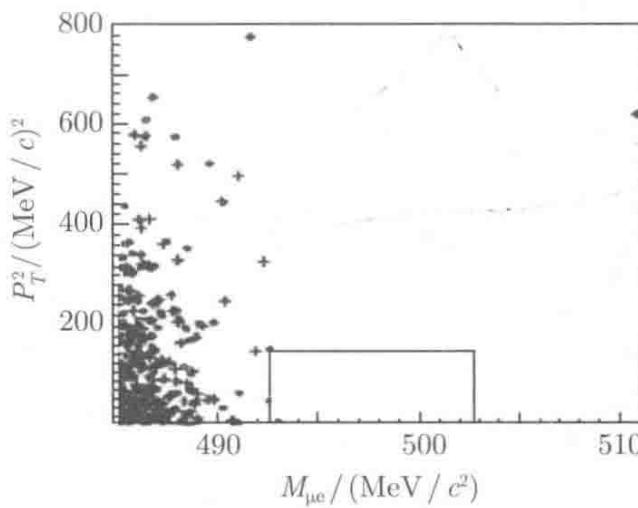


图 8.2 寻找稀有衰变  $K_L^0 \rightarrow \mu^\pm e^\mp$  的信号盲区方法  
正下方的方框是信号盲区

信号盲区法的另一应用实例是 BABAR 实验组对于  $B^0 \rightarrow \rho^\pm \pi^\mp$  衰变事例的寻找<sup>[125]</sup>。图 8.3 显示了两个特征量的散点图:  $m_{ES}$ (横坐标) 和  $\Delta E$ (纵坐标)。对于信号事例,  $m_{ES}$  接近于  $B^0$  介子质量 ( $\sim 5.28 \text{ GeV}$ ),  $\Delta E$  接近于 0。图 8.3 中右部的长方形方框是优化后的信号区, 包围着它的阴影区则是信号盲区。信号盲区上侧、下侧和左侧画斜线的区域为不同的边带区。图 8.4 和图 8.5 是利用这些边带区数据以及本底的 MC 数据确定的本底  $m_{ES}$  和  $\Delta E$  分布。利用这些分布可以确定信号盲区内本底的  $m_{ES}$  和  $\Delta E$  分布的形状。

分析最终采用的各特征量 (包括  $m_{ES}$  和  $\Delta E$ ) 的截断值需要通过某种优化

步骤来确定。优化的标准通常是选择一组截断值使得所谓的信号统计显著性达到极大。信号统计显著性有不同的定义（见 6.3 节），常见的用于截断值优化的量是

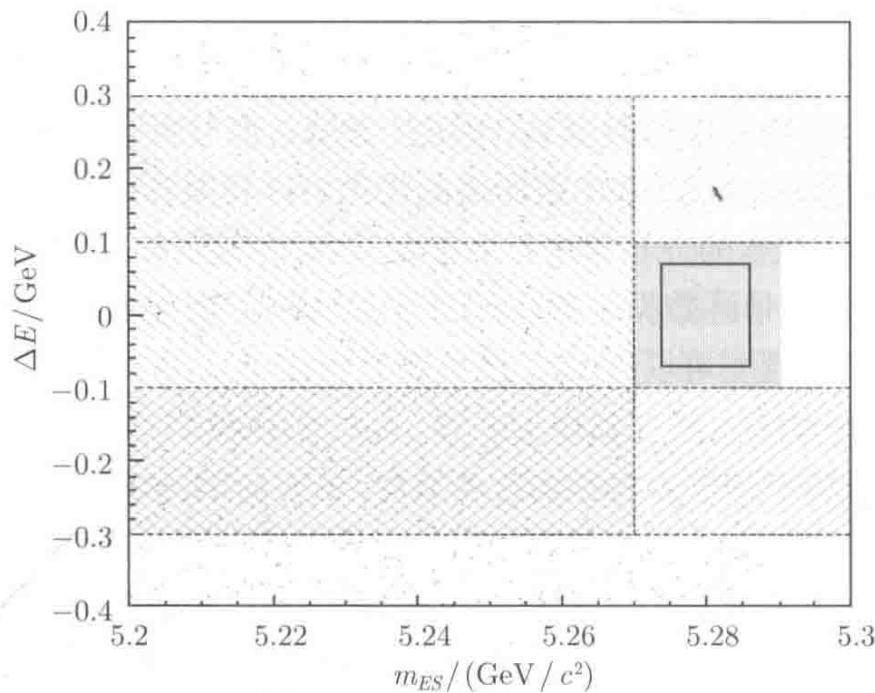


图 8.3 BABAR 实验组寻找  $B^0 \rightarrow \rho^\pm \pi^\mp$  衰变事例的  $m_{ES}$ - $\Delta E$  散点图

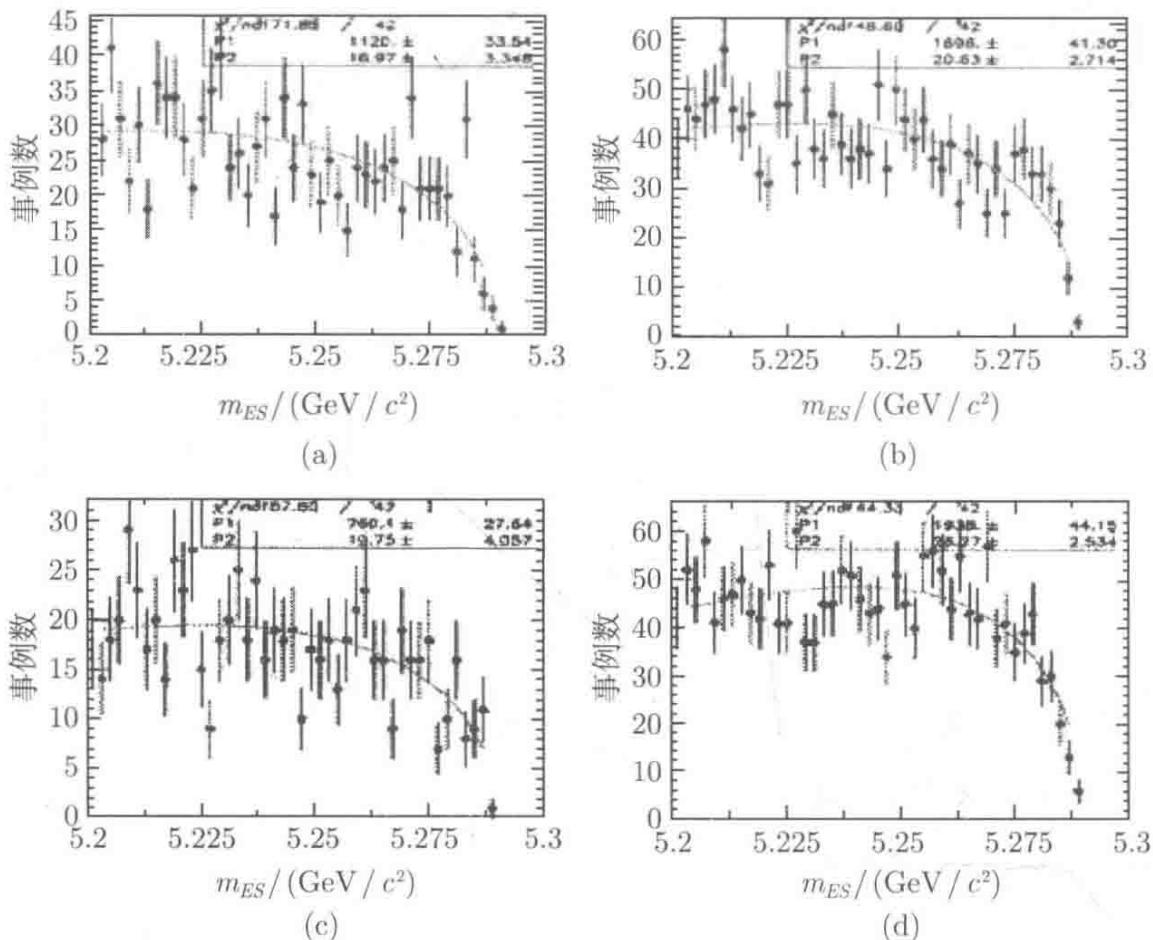
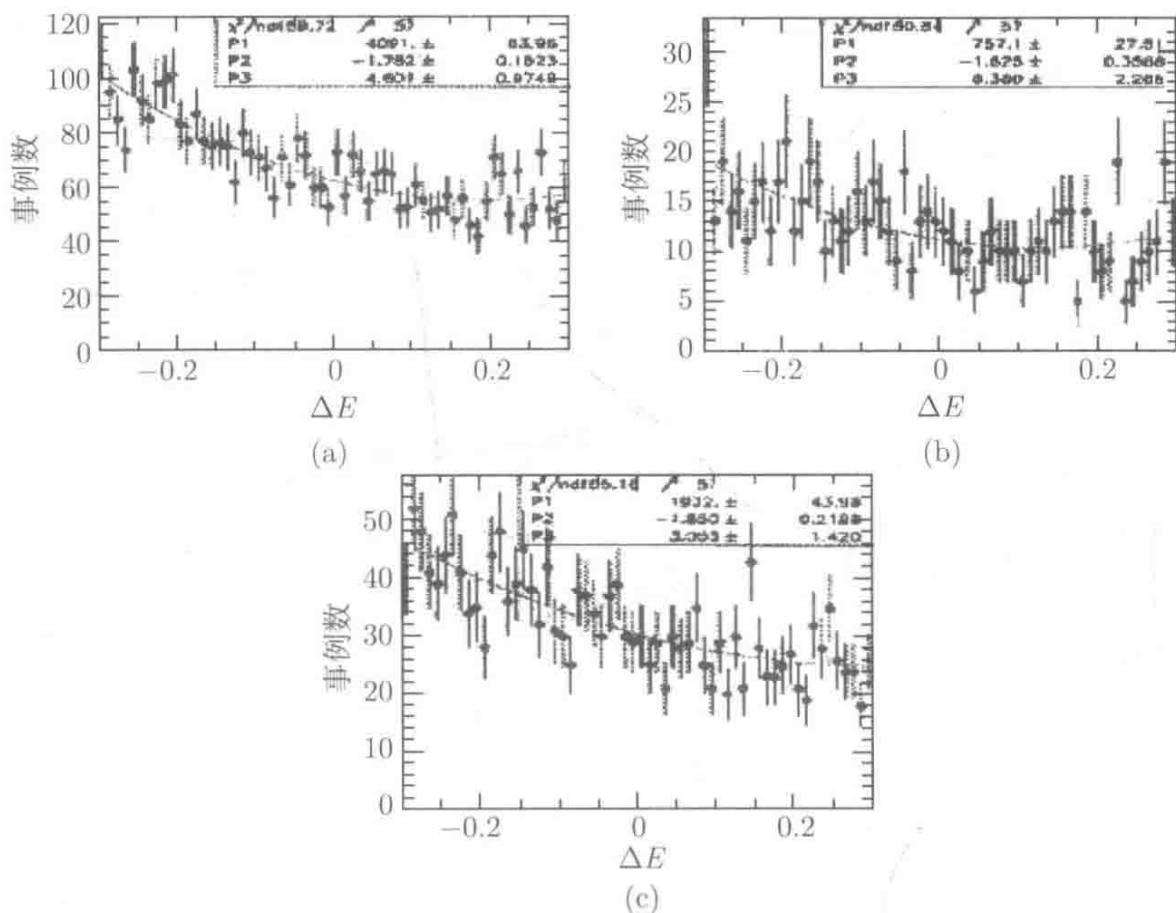


图 8.4 本底的  $m_{ES}$  分布

(a) 信号上边带区数据确定; (b) 信号下边带区数据确定; (c) 信号左边带区数据确定;  
(d) 本底 MC 数据确定

图 8.5 本底的  $\Delta E$  分布

(a) 信号左边带区数据确定; (b) 上、下边带区数据确定; (c) 本底 MC 数据确定

$S/\sqrt{S+B}$ , 这里  $S$  和  $B$  是经过各特征量截断值判选后余下的信号和本底事例数的期望值, 其中  $S$  由信号 MC 事例样本或控制数据样本求得,  $B$  由本底 MC 事例样本或边带区数据样本求得. 需要强调指出的是, 截断值优化过程中, 信号盲区中的真实数据仍然是保持隐蔽的. 由于大的事例统计量能够减小统计误差, 所以 MC 事例样本量和边带区数据样本量应当尽可能地大; 当然边带区不能任意地扩大, 对于它的基本限制是其中事例的特征量的分布能够反映信号盲区内本底的特征量的分布, 同时需要适当的归一化处理. 为了避免可能的偏差, 可以只利用这些 MC 事例样本量和边带区数据样本量的一半来进行截断值的优化. 当优化的截断值确定之后, 再利用另一半样本来确定信号探测效率、本底函数的形状和归一化系数, 以及信号函数的形状. 然后根据这些结果, 可以确定拟合方法并估计系统误差.

至此, 数据分析的流程已经基本完成, 余下的工作就是打开信号盲区中的真实数据, 得到最终的物理结果. 对于一项重要的分支比测量, 盲分析打开信号盲区中的真实数据是一件十分郑重的事件. 通常从事该项研究的研究小组需要向一个专门委员会提交一份研究报告, 详细报告其分析方法, 其中至少需要包含以下内容:

- (1) 截断值优化和本底函数形状的确定方法和步骤;
- (2) 信号盲区中的估计本底事例数作为待测分支比的函数;

- (3) 根据信号 MC 样本或控制事例样本估计的信号探测效率;
- (4) 预期的统计灵敏度;
- (5) 系统误差的估计.

只有在专门委员会同意了研究报告 (可视为实验组对于该项研究的权威认可) 的情形下才容许打开信号盲区中的真实数据, 从而得出最终的物理结果. 图 8.6 显示了 BABAR 实验组寻找  $B^0 \rightarrow \rho^\pm \pi^\mp$  衰变事例的分析中, 打开信号盲区后信号事例对于本底事例背景的超出 (信号区在  $m_{ES} \sim 5.28\text{GeV}$ ,  $\Delta E \sim 0$  附近).

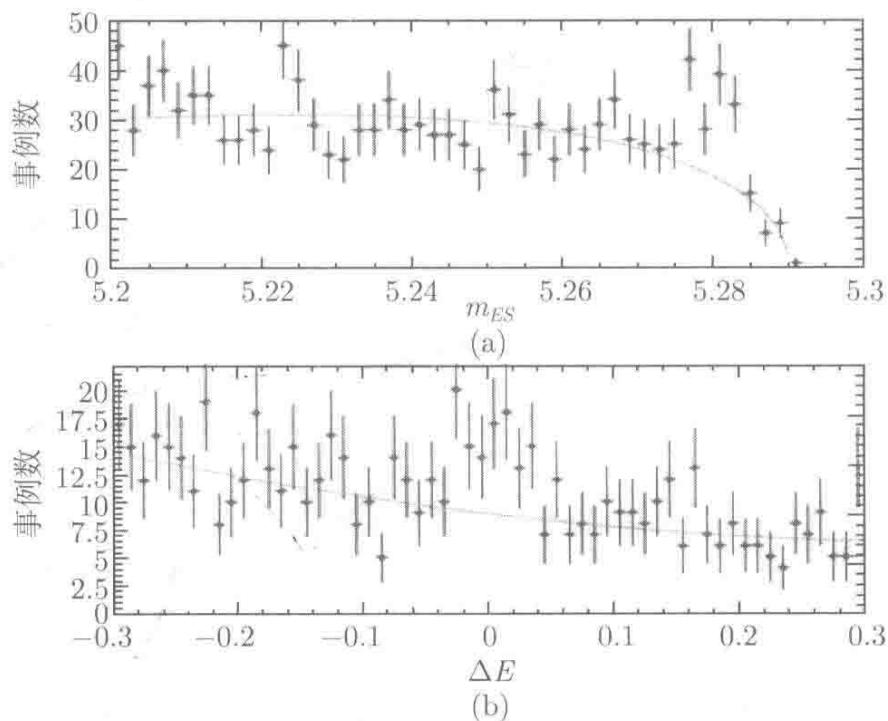


图 8.6 BABAR 实验组寻找  $B^0 \rightarrow \rho^\pm \pi^\mp$  衰变事例的分析中, 打开信号盲区后信号事例出现于  $m_{ES} \sim 5.28\text{GeV}$ ,  $\Delta E \sim 0$  的信号区

另一种情形是寻找新粒子或“隆起”(bump-hunting) 的实验, 这时信号区的位置事先未知, 因而无法设置信号盲区. 而恰恰是这类实验, 特别需要排除实验者的偏差. 历史上曾经观测到某些具有相当高统计显著性的“隆起”, 后来发现是人为制造出来的. 这种情形下的盲分析依赖于对于本底的了解. 如果在“隆起”区域以外的地方对本底进行独立的估计, 并确定分析的方法和步骤, 那么利用后者得到的结果具有盲分析的特征. 也就是说, 如果确定了寻找新粒子或“隆起”的区域, 那么可以将该区域设置为信号盲区, 上面叙述的信号盲区方法就完全适用.

### 8.3.2 改善分支比测量精度的盲分析

一般而言, 对于已经存在测量值的分支比测量要求有比较高的测量精度, 以改善已有的测量结果. 这种情形下, 需要同时仔细地研究信号的数据样本和 MC 模拟样本. 这样, 前面讨论过的稀有衰变盲分析方法——信号盲区法不再适用, 而需要

寻找其他途径. 一种途径是利用一部分数据来确定分析方法和步骤, 然后将所确定的分析方法和步骤应用于全部数据样本以给出最终结果. 第一个步骤中所使用的数据量的大小是一个需要谨慎考虑的问题. 数据量太小, 本底分布以及其他一些重要分布的细节可能了解得过于粗糙; 数据量太大, 可能 (部分地) 失去了“盲分析”的初衷. 另一种途径是将一定数量的 MC 模拟数据与真实数据混合起来, 利用混合数据确定分析方法和步骤, 然后将后者应用于数据样本以给出最终结果.

### 8.3.3 隐蔽参数法——参数精确测量的盲分析

许多测量中, 待测参数的数值往往是根据某个 (些) 特征变量的分布导出的, 这一分布遍及特征变量的所有物理容许的区域, 因而参数的确定需要利用全部数据来对该分布进行拟合. 有效的分析方法的确定 (如有效的信号事例的判选) 必定会涉及利用真实数据的重新拟合, 因此必定会用到全部实验数据而不可能像信号盲区法中那样将盲区中的数据隐蔽起来.

对于参数 (如  $B$  和  $D$  介子寿命) 的精确测量, 所使用的盲分析方法与 8.3.1 节的信号盲区法不同. 现在, 待测参数已经存在相当精确的测量值, 盲分析的目的是要避免本测量的结果“有意无意”地偏向已有的测量值或粒子数据表给出的平均值的可能性. 对于精确测量, 信号事例统计量很大, 参数的系统误差占据重要地位, 盲分析应当确保系统误差的估计不会受到参数自身数值大小的影响.

为了在进行分析方法的系统检查时免受参数自身数值大小的影响, 可以利用隐蔽参数法. 这一方法中, 待测参数的拟合值  $x$  通过如下变换得到其盲值  $x^*$ :

$$x^* = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} x + R. \quad (8.3.1)$$

其中, 常数偏置值  $R$  由一随机数产生子给定, 该随机数具有适当的均值和标准偏差, 如均值为 0、标准偏差等于  $x$  的实验标准偏差 1~3 倍的正态随机数产生子是比较合理的选择.  $x$  前面的乘因子 1 或 -1 也是随机选定的, 这种做法保证了当参数  $x$  发生变化时, 盲值  $x^*$  的变化方向与  $x$  的变化方向没有固定的关联. 常数偏置值  $R$  与乘因子 1 和 -1 的选取虽然是随机的, 但一旦选定便固定不变了, 因此参数真值与盲值的差值  $\Delta x^* \equiv x^* - x$  是一固定常数, 但其数值对于分析者是隐蔽 (未知) 的. 这些做法都是为了保证分析过程不受参数自身数值大小的影响.

隐蔽参数法的重要特点是: 首先, 它仅改变待测参数的拟合值, 而不改变实验的“原始数据”或 MC 产生的“原始数据”; 其次, 即使参数值  $x$  被改变为盲值  $x^*$ , 但它们有相同的测量误差, 即  $\sigma(x) = \sigma(x^*)$ . 因此整个分析流程 (截断值的优化、效率计算、MC 修正、拟合方法、系统误差估计等) 的确定都可以利用盲值  $x^*$  (即与参数真值有差别且差值的正负未知) 来进行. 例如, 在截断值优化前后求得

的参数盲值分别为  $x_B^* \pm \sigma(x_B^*)$  和  $x_A^* \pm \sigma(x_A^*)$ , 由于  $\Delta x^* = \Delta x_A^* = \Delta x_B^*$ , 故有  $x_A^* - x_B^* = |x_A - x_B|$ . 这表示截断值优化前后参数盲值的变化量与参数真值的变化量是一样的. 同时, 由于  $\sigma(x_B^*) = \sigma(x_B)$ ,  $\sigma(x_A^*) = \sigma(x_A)$ , 所以截断值优化对于参数  $x$  测量精度的改进也是知道的. 根据以上所述, 用盲值  $x^*$  对截断值进行优化的效果与利用参数  $x$  对截断值进行优化是一样的, 但同时使分析者不知道参数  $x$  的数值本身, 这就避免了分析者故意向参数  $x$  的某个预想值调节的企图. 对于系统误差的分析, 用盲值  $x^*$  的效果与利用参数  $x$  也是一样的.

KTeV 实验<sup>[126]</sup> 对参数  $\varepsilon'/\varepsilon$  进行了精确测量, 其中,  $\varepsilon$  是  $K^0 - \bar{K}^0$  混合不对称性参数,  $\varepsilon'$  是  $K \rightarrow \pi\pi$  衰变中的直接 CP 破坏参数. 参数  $\varepsilon'/\varepsilon$  的值依赖于  $K_S^0$  和  $K_L^0$  衰变为  $\pi^+\pi^-$  和  $\pi^0\pi^0$  末态的分支比. 不同的理论模型对于该参数有不同的预期值, 从 0 到  $10^{-3}$  量级, 因而  $\varepsilon'/\varepsilon$  的精确测量可以检验这些理论模型的预期. 实验者希望利用所有的事例对他们获取参数  $\varepsilon'/\varepsilon$  值的分析方法进行系统的检查, 保证其无偏性. KTeV 实验中, 束流打击一个靶同时产生  $K_L^0$  和  $K_S^0$ , 因而可以同时收集  $K_L^0$  和  $K_S^0$  的衰变事例.  $\varepsilon'/\varepsilon$  值由  $K_S^0$  和  $K_L^0$  衰变顶点分布的形状所决定. 显然,  $\varepsilon'/\varepsilon$  值的确定需要用到全部数据, 不存在所谓的信号盲区.

他们的分析方法与其他实验的常规分析相同, 但参数  $\varepsilon'/\varepsilon$  的值被隐蔽, 代之以式 (8.3.1) 所示的参数  $\varepsilon'/\varepsilon$  (现在是  $x$ ) 的盲值  $x^*$ , 但所有的“原始数据”, 特别是  $K_S^0$  和  $K_L^0$  衰变顶点分布的“原始数据”并没有改变, 因此这种情形下, 仍然可以确定完整的分析流程. 利用 MC 模拟数据可以验证这样确定的分析流程得到的参数  $\varepsilon'/\varepsilon$  值的正确性. 利用常数偏置值  $R$  与乘因子 1 和  $-1$  的两组随机选取值平行地进行两组盲分析, 以避免可能的偏向性, 两组盲分析的拟合结果的比较是一种全盲的比较. 两组盲分析完成之后, 常数偏置值  $R$  与乘因子 1 和  $-1$  的随机选取值被固定为相同, 以进行比较, 由此可检查分析方法的一致性. 当确认了分析方法的正确性之后, 对分析实验数据拟合得到的参数  $\varepsilon'/\varepsilon$  值“去盲”作为测量结果发表.

### 8.3.4 隐蔽不对称性分布和不对称参数法——不对称性测量盲分析

某些测量中, 仅隐蔽待测参数的值本身是不够的, 某些特征量的分布图形也需要隐蔽, 因为根据这些图形可能对待测参数的值作出粗略的估计, 而这一点对于保证分析过程的无偏性可能带来不利影响.

这类测量的一个实例是不对称性的测量, 如 BABAR 实验中性  $B$  介子衰变的时间依赖 CP 不对称性的测量<sup>[127]</sup>.  $\Upsilon(4S)$  衰变为一对中性  $B$  介子  $B^0\bar{B}^0$ , 实验中衰变到包含粲偶素的 CP 本征态的那个中性  $B$  介子被完全重建, 记为  $B_{\text{rec}}$ ; 另一个反冲的中性  $B$  介子 ( $B^0$  或  $\bar{B}^0$ ) 记为  $B_{\text{tag}}$ , 称为味标记 (flavor tagging)  $B$  介子, 该介子的味 (flavor) 根据它的衰变产物确定. CP 不对称性  $A_{\text{CP}}(\Delta t)$  的符号和大致数值通过查看图 8.7(a) 所示的  $B^0$  和  $\bar{B}^0$  衰变为 CP 本征态的  $\Delta t$  分布可以大致确

定. 这里  $\Delta t = t_{\text{rec}} - t_{\text{tag}}$ ,  $t_{\text{rec}}$  是  $B_{\text{rec}}$  介子的衰变时间,  $t_{\text{tag}}$  是  $B_{\text{tag}}$  介子的衰变时间. 时间依赖的 CP 不对称性定义为

$$A_{\text{CP}}(\Delta t) = \frac{N_+(\Delta t) - N_-(\Delta t)}{N_+(\Delta t) + N_-(\Delta t)}, \quad (8.3.2)$$

其中,  $N_+(\Delta t)$  和  $N_-(\Delta t)$  分别为味标记粒子是  $B^0$  和  $\bar{B}^0$  的事例数. 由于存在关系式

$$A_{\text{CP}}(\Delta t) = -\eta_f \sin 2\beta \sin(\Delta m_{B^0} \Delta t), \quad (8.3.3)$$

故利用  $A_{\text{CP}}(\Delta t)$  立即可导出 CP 破坏不对称性参数  $\sin 2\beta$  的数值.

显然, 由图 8.7(a) 所示的味标记粒子  $B^0$  和  $\bar{B}^0$  的事例数  $N_+(\Delta t)$  和  $N_-(\Delta t)$  的分布, 我们立即可以按照式 (8.3.2) 估计出  $A_{\text{CP}}(\Delta t)$  的数值, 并据此导出 CP 破坏不对称性参数  $\sin 2\beta$  的数值. 所以要想避免可能的人为偏向性, 仅隐蔽所测参数  $A_{\text{CP}}(\Delta t)$  的值是不够的, 还需要隐蔽可推断出不对称性的  $\Delta t$  分布 (图 8.7(a)).

BABAR 实验中, 通过实验数据的  $\Delta t$  分布拟合得到的  $A_{\text{CP}}(\Delta t)$  值利用式 (8.3.1) 的方式加以隐蔽, 代之以盲值  $A'_{\text{CP}}(\Delta t)$ . 与 8.3.3 节中所讨论的一样, 常数偏置值  $R$  与乘因子 1 和  $-1$  的选取虽然是随机的, 但一旦选定便固定不变了, 因此参数真值与盲值的差值  $\Delta A'_{\text{CP}}(\Delta t) \equiv A'_{\text{CP}}(\Delta t) - A_{\text{CP}}(\Delta t)$  是一固定常数, 但其数值对于分析者是隐蔽 (未知) 的. 这种变换不改变实验的“原始数据”或 MC 产生的“原始数据”; 其次,  $A_{\text{CP}}(\Delta t)$  与  $A'_{\text{CP}}(\Delta t)$  有相同的测量误差. 因此整个分析流程 (截断值的优化、效率计算、MC 修正、拟合方法、系统误差估计等) 的确定都可以利用盲值  $A'_{\text{CP}}(\Delta t)$  (即与参数真值有差别且差值的正负未知) 来进行.

同时, 为了隐蔽不对称性, 图 8.7(a) 所示的数据  $\Delta t$  分布也被隐蔽起来而代之以盲值  $\Delta t'$  的分布 [128]:

$$\Delta t' = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} \cdot S_{\text{Tag}} \cdot \Delta t + R. \quad (8.3.4)$$

对于  $B^0$  和  $\bar{B}^0$  的味标记,  $S_{\text{Tag}}$  值分别是 1 和  $-1$ . 因为  $B^0$  和  $\bar{B}^0$  的不对称性  $A_{\text{CP}}(\Delta t)$  数值几乎相等而符号相反, 所以式 (8.3.4) 右边第一项中乘因子 1 和  $-1$  的随机选择可以隐蔽  $\Delta t$  分布的不对称性信息. 此外, 从  $B^0$  和  $\bar{B}^0$  的  $\Delta t$  分布相对于  $\Delta t = 0$  点的不对称性就可以目视出不对称性  $A_{\text{CP}}$  数值和符号的信息, 而式 (8.3.4) 右边第二项的随机选取的偏置值  $R$  具有隐蔽  $\Delta t = 0$  点的效果. 这样, 在图 8.7(b) 所示的数据  $\Delta t'$  分布中,  $B^0$  和  $\bar{B}^0$  的不对称性  $A_{\text{CP}}(\Delta t)$  的数值和符号的信息已经不能从  $\Delta t'$  分布图形加以判断 ( $B^0$  和  $\bar{B}^0$  曲线的差异来源于两者寿命的差异).

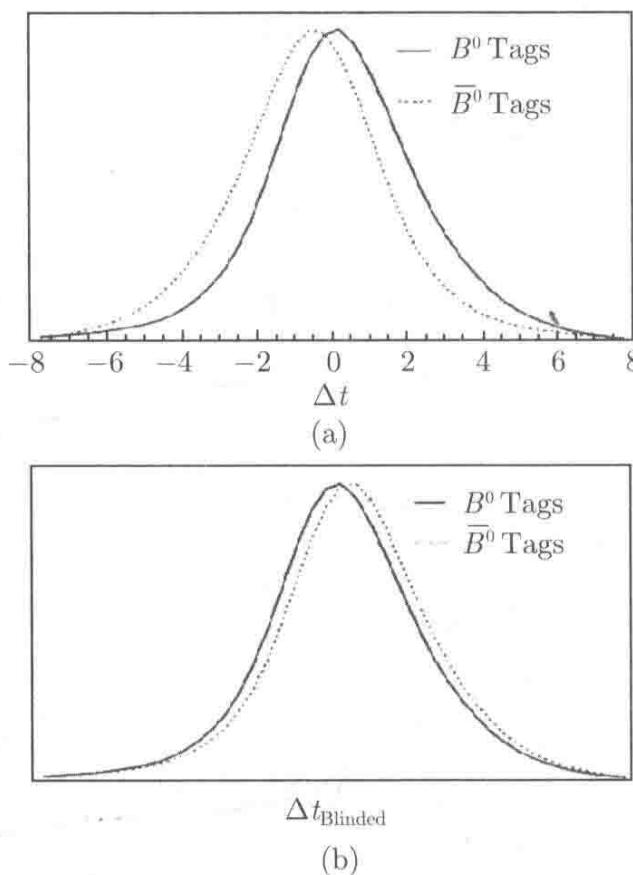


图 8.7 味标记粒子  $B^0$  和  $\bar{B}^0$  衰变事例数  $N_+$  和  $N_-$  的  $\Delta t$  分布

$\sin 2\beta = 0.75$ , 时间分辨率为 BABAR 实验中的典型值. (a)  $B^0$ (实线) 和  $\bar{B}^0$ (虚线) 衰变事例数  $N_+$  和  $N_-$  与  $\Delta t$  的关系; (b)  $B^0$ (实线) 和  $\bar{B}^0$ (虚线) 衰变事例数  $N_+$  和  $N_-$  与  $\Delta t_{\text{Blind}}$  的关系

因此, 实验者就可以利用盲值  $A'_{\text{CP}}(\Delta t)$  来确定分析方法并验证所采用的分析方法和步骤的正确性, 而这种验证与不对称性是否存在、数值大小是不相关联的.

## 8.4 结语和讨论

上面的讨论告诉我们, 偏离测量结果的客观性的偏差利用盲分析可以至少部分地加以消除或减小, 但是实验者通过分析方法的“合理”调节来“修正”结果所导致的偏差不一定通过盲分析能够完全消除 [129]. 例如, 我们在 8.1 节中指出的数据刻度文件和物理过程 MC 模型的不同选择会导致不同的结果就不能用盲分析来消除, 因为不论是否进行盲分析, 刻度文件和物理过程 MC 模型总是要用到的. 为了尽可能排除对于测量结果的人为偏差, 分析方法应当遵循公认为有效的一些规则, 可以大致列举如下:

- (1) 在利用真实数据之前, 首先利用 MC 模拟数据确定完整的数据分析流程.
- (2) 对于粒子识别等所有实验分析都会遇到的问题, 利用合作组已经详细研究过的标准判选判据, 除非你的测量对象对于该问题的判选判据特别敏感. 在后者的情形下, 应当对它进行更详尽的研究得出对于标准判选判据的修改, 该研究中应当

避免利用修改后的判选判据来观察你的测量对象的结果, 其结论应当获得合作组的认可.

- (3) 利用正规的优化步骤来确定特征量的截断值.
- (4) 通过控制样本确定事例判选特征量的种类和截断值是一种相对无偏的方法. 应当选择适合于你的研究对象的、统计量尽可能大的控制样本.
- (5) 选择合作组公用的数据刻度文件版本.
- (6) 对于物理过程的描述存在不同模型的情形, 应当充分考虑由此导致的系统误差.