

分支比多个测量值的合并估计

朱永生

(高能物理研究所、中国科学院大学)

《粒子物理数据分析基础与前沿》研讨会

2016.9.11~14, 北京

1. 引言

- 1.1 问题的提出
- 1.2 测定分支比的一般方法

2. 合并估计的最小二乘法

- 2.1 I 个实验谱为特征变量 $m_i, i = 1, \dots, I$ 直方图
- 2.2 I 个实验谱为特征变量 $m_i, i = 1, \dots, I$ 散点图

3. 合并估计的极大似然法

- 3.1 常规极大似然法
 - 3.1.1 I 个实验谱为特征变量 $m_i, i = 1, \dots, I$ 直方图
 - 3.1.2 I 个实验谱为特征变量 $m_i, i = 1, \dots, I$ 散点图
- 3.2 广义极大似然法
 - 3.2.1 I 个实验谱为特征变量 $m_i, i = 1, \dots, I$ 直方图
 - 3.2.2 I 个实验谱为特征变量 $m_i, i = 1, \dots, I$ 散点图

4. 信度区间和上限的确定，系统误差的考虑

- 4.1 不考虑系统误差时信度区间和上限的确定
- 4.2 考虑系统误差时信度区间和上限的确定

5. 小结

1. 引言

(1.1) 问题的提出

- 同一衰变分支比，多个实验进行测量。
- 多个测量结果进行统计意义上的合并，获得比单个测量结果更高的精度。
- 同一物理量作 I 次独立测量，观测值为 $x_i \pm \sigma_i$, $i=1, \dots, I$ ，假定测量服从正态分布。

合并估计为 $\mu \pm \sigma$ ，其中

$$\mu = \frac{\sum_{i=1}^I \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^I \frac{1}{\sigma_i^2}}, \quad \sigma = \frac{1}{\sqrt{\sum_{i=1}^I \frac{1}{\sigma_i^2}}}. \quad (1.1)$$

- 某些实验由于信号事例数过少，只能给出一定置信水平下的上限。无法利用该公式。

(1.2) 测定分支比的一般方法

- 利用特定的选择判据，选择出信号道的候选信号事例，包含信号事例和本底事例。
- 在一个特征变量的实验观测谱的信号区内，信号和本底谱型具有显著的差异。
- 通过特征变量谱的拟合来求得信号事例数，从而得到待测的分支比。

BES 实验 $\psi(2S) \rightarrow \eta J/\psi$ 分支比测量为例[1]。

末态是 $\gamma\gamma e^+e^-$ 和 $\gamma\gamma\mu^+\mu^-$ 事例。

将轻子对的不变质量约束为 J/ψ 质量，得到候选信号事例的 $M_{\gamma\gamma}$ 的分布。

若信号和本底函数已知，通过 $M_{\gamma\gamma}$ 谱的拟合确定信号事例数，计算出分支比。

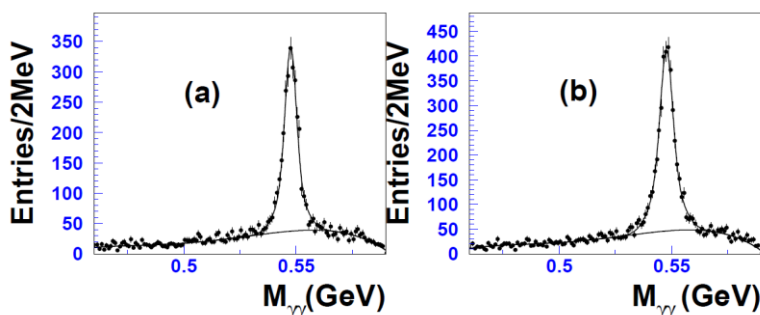


图 1 $\psi' \rightarrow \eta J/\psi$ 候选事例 $M_{\gamma\gamma}$ 的分布. (a) $\gamma\gamma e^+e^-$ 末态 (b) $\gamma\gamma\mu^+\mu^-$ 末态.

- 所谓特征变量谱的拟合，指的是构建最小二乘 (χ^2) 函数或似然函数，通过其极小化或极大化求得信号事例数。
- 分支比计算公式

$$B(R \rightarrow X) = \frac{N_s(R \rightarrow X \rightarrow Y)}{N_R \varepsilon(R \rightarrow X \rightarrow Y) \cdot BR(X \rightarrow Y)}. \quad (1.2)$$

R — 共振态 $\psi(2S)$, X — $\eta J/\psi$, Y — $\gamma\gamma e^+e^-$ 或 $\gamma\gamma\mu^+\mu^-$,

N_R — R 衰变事例数, ε — 信号事例探测效率, N_s — 观测信号事例数,

分支比公式 (符号简化):

$$B_i = \frac{N_{is}}{N_{Ri} \varepsilon_i BR_i} \equiv \frac{N_{is}}{A_i}, \quad i=1(\gamma\gamma e^+e^-), 2(\gamma\gamma\mu^+\mu^-). \quad (1.3)$$

式(1.3)容易推广到多个测量的情形。

- 分支比的合并估计
利用所有单个实验的实验测量谱数据构建联合最小二乘 (χ^2) 函数或联合似然函数, 通过极小化或极大化求得分支比的合并估计。
- 联合 χ^2 函数或联合似然函数的构建取决于各实验的特征变量观测谱形式。

2. 合并估计的最小二乘法

(2.1) I 个实验谱为特征变量 $m_i, i=1, \dots, I$ 直方图

- 每个实验谱的特征变量 m_i 可以各不相同 (或相同)。
- 各直方图的信号区间和子区间划分可以不同 (或相同)。(这两点下同)
- 实验 i 信号区间内总事例数为 N_i ,

直方图子区间数为 J_i ,

子区间 $j_i (=1, \dots, J_i)$ 中的观测事例数为 n_{ij_i} ,

- 联合最小二乘函数 χ^2 定义为

$$\chi^2 = \sum_{i=1}^I \chi_i^2 = \sum_{i=1}^I \sum_{j_i=1}^{J_i} \frac{(n_{ij_i} - N_i p_{ij_i})^2}{N_i p_{ij_i}}, \quad (2.1)$$

p_{ij_i} 是一个事例出现在子区间 j_i 中的概率:

$$p_{ij_i} = p_{ij_i}(\theta_i) = \int_{\Delta m_{j_i}} f_i(m_i | \theta_i) dm_i, \quad (2.2)$$

$f_i(m_i | \theta_i)$ 为实验 i 中信号区间内信号候选事例特征变量 m_i 的概率密度:

$$f_i(m_i | \theta_i) = w_{is} f_{is}(m_i | \theta_{is}) + (1 - w_{is}) f_{ib}(m_i | \theta_{ib}). \quad (2.3)$$

θ_{is} 和 θ_{ib} 是信号和本底的概率密度 f_{is} 和 f_{ib} 的参数 (可以是多个参数),

f_{is} 和 f_{ib} 的形式在实验 i 的数据分析中应当已知。

w_{is} 表示信号区间内信号事例占全部事例的比例, 即 $N_{is} = w_{is} N_i$ 。

由(1.3)式可知实验 i 中信号区间内的信号事例数 N_{is} 为

$$N_{is} = A_i B_i, \quad A_i \equiv N_{Ri} \epsilon_i B_{Ri}, \quad i=1, \dots, I. \quad (2.4)$$

当进行合并估计时, $B = B_i$ 。由 $N_{is} = w_{is} N_i$ 知 $w_{is} = A_i B / N_i$, 代入式 (2.3) 得

$$f_i(m_i|\theta_i) = \frac{A_i}{N_i} B \cdot f_{is}(m_i|\theta_{is}) + \left(1 - \frac{A_i}{N_i} B\right) f_{ib}(m_i|\theta_{ib}). \quad (2.5)$$

式 (2.5) 描述每个实验中特征变量 m_i 的概率密度。

- 式 (2.1) 所示的 χ^2 表式中的未知参数 θ 包括 θ_s , θ_b 和 B

$$\theta = \{B, \theta_s, \theta_b\}, \quad \theta_s = \{\theta_{1s}, \dots, \theta_{Is}\}, \quad \theta_b = \{\theta_{1b}, \dots, \theta_{Ib}\}. \quad (2.6)$$
- 对函数 χ^2 求极小可求得未知参数 θ 的估计值 $\hat{\theta}$ (包括分支比合并估计 \hat{B}) 及其统计 (拟合) 误差 (包括 B 的统计误差 $\sigma_{B,st}$)。

(2.2) I 个实验谱为特征变量 $m_i, i=1, \dots, I$ 散点图

- 变换成直方图数据, 再通过上述步骤求得合并估计
- 这一变换会损失部分测量信息, 增加估计值的误差。

3. 合并估计的极大似然法

(3.1) 常规极大似然法

(3.1.1) I 个实验谱为特征变量 $m_i, i=1, \dots, I$ 直方图

- 实验 i 信号区间内的总事例数为 N_i , 直方图子区间数为 J_i ,
子区间 $j_i (=1, \dots, J_i)$ 中的观测事例数为 n_{ij_i} .
- 第 j_i 个子区间内观测到 n_{ij_i} 个事例 ($j_i=1, \dots, J_i$) 的联合概率为多项分布:

$$L_i(n_{i1}, \dots, n_{ij_i}) = N_i! \prod_{j_i=1}^{J_i} \frac{p_{ij_i}(m_i|\theta_i)^{n_{ij_i}}}{n_{ij_i}!}, \quad i=1, \dots, I. \quad (3.1)$$

$p_{ij_i}(m_i|\theta_i)$ 是实验 i 中, 子区间 j_i 中观测到一个事例的概率,

由概率密度 $f_i(m_i|\theta_i)$ 在第 j_i 子区间 Δm_{j_i} 中的积分求出:

$$p_{ij_i}(m_i|\theta_i) = \int_{\Delta m_{j_i}} f_i(m_i|\theta_i) dm_i. \quad (3.2)$$

$f_i(m_i|\theta_i)$ 为实验 i 中信号区间内特征变量 m_i 的概率密度, 仍由式(2.5)描述。

- I 个实验的联合似然函数为

$$L(\theta) = \prod_{i=1}^I L_i = \prod_{i=1}^I \left[N_i! \prod_{j_i=1}^{J_i} \frac{p_{ij_i}(m_i|\theta_i)^{n_{ij_i}}}{n_{ij_i}!} \right], \quad (3.3)$$

不考虑与待拟合参数 θ 无关的项, 有

$$\ln L = \sum_{i=1}^I \sum_{j_i=1}^{J_i} n_{ij_i} \ln p_{ij_i}(m_i | \theta_i). \quad (3.4)$$

- 联合似然函数中的未知参数 θ 包括 θ_s, θ_b 和 B :

$$\theta = \{B, \theta_s, \theta_b\}, \quad \theta_s = \{\theta_{1s}, \dots, \theta_{Is}\}, \quad \theta_b = \{\theta_{1b}, \dots, \theta_{Ib}\}. \quad (3.5)$$

- 对 $\ln L$ 求极大可求得未知参数 θ 的估计值 $\hat{\theta}$ (包括分支比合并估计 \hat{B}) 及其统计误差 (包括 B 的统计误差 $\sigma_{B,st}$)。
- θ_s, θ_b 的初值可取为各个实验中确定的值, B 的初值可利用各个实验确定的 B_i 的某种加权平均。
- 当各实验 i 信号区间内的总事例数 N_i 充分大, 且每个子区间内特征变量 m_i 的概率密度 $f_i(m_i | \theta_i)$ 的变化不显著的情形下, 常规极大似然法对于 I 个实验谱为直方图数据的情形可给出相当可靠的分支比合并估计 \hat{B} 及其统计误差 $\sigma_{B,st}$ 。

(3.1.2) I 个实验谱为特征变量 $m_i, i=1, \dots, I$ 散点图

- 实验 i 散点图的事例数为 N_i , 这 N_i 个事例出现在 m_{i1}, \dots, m_{iN_i} , $i=1, \dots, I$ 处。
- 信号区间内的概率密度仍然如式(2.5)所描述。
- I 个实验的观测谱共有 $N \equiv \sum_{i=1}^I N_i$ 个事例。
- N 个事例的联合似然函数为:

$$\begin{aligned} L &= L(m_{11}, \dots, m_{1N_1}; \dots; m_{I1}, \dots, m_{IN_I} | B, \theta_s, \theta_b) = \prod_{i=1}^I L_i(m_{i1}, \dots, m_{iN_i} | B, \theta_{is}, \theta_{ib}) \\ &= \prod_{i=1}^I \prod_{j=1}^{N_i} \left[\frac{A_i}{N_i} B \cdot f_{is}(m_{ij} | \theta_{is}) + \left(1 - \frac{A_i}{N_i} B\right) f_{ib}(m_{ij} | \theta_{ib}) \right]. \end{aligned} \quad (3.6)$$

对数似然函数为

$$\ln L = \sum_{i=1}^I \sum_{j=1}^{N_i} \ln \left[\frac{A_i}{N_i} B \cdot f_{is}(m_{ij} | \theta_{is}) + \left(1 - \frac{A_i}{N_i} B\right) f_{ib}(m_{ij} | \theta_{ib}) \right]. \quad (3.7)$$

- 联合似然函数中的未知参数 $\theta = \{B, \theta_s, \theta_b\}$ 如式(3.5)所示。

(3.2) 广义极大似然法

(3.2.1) I 个实验谱为特征变量 $m_i, i=1, \dots, I$ 直方图

- 实验 i 信号区间内的总事例数为 N_i , 直方图子区间数为 J_i , 子区间 $j_i (=1, \dots, J_i)$ 中的观测事例数为 n_{ij_i} 。
- 广义极大似然法中, n_{ij_i} 是期望值 λ_{ij_i} 的泊松变量. 第 j_i 个子区间内观测到 n_{ij_i} 个事例 ($j_i = 1, \dots, J_i$) 的联合概率为

$$L_i(n_{i1}, \dots, n_{iJ_i}) = \prod_{j_i=1}^{J_i} \frac{1}{n_{ij_i}!} \lambda_{ij_i}^{n_{ij_i}} e^{-\lambda_{ij_i}}, \quad (3.8)$$

λ_{ij_i} 由信号区间内的概率密度 $f_i(m_i | \theta_i)$ 在第 j_i 子区间中的积分求出:

$$\lambda_{ij_i} = \lambda_i \int_{\Delta m_{j_i}} f_i(m_i | \theta_i) dm_i, \quad (3.9)$$

λ_i 是实验 i 直方图的总事例数 N_i (泊松变量) 的期望值:

$$\lambda_i = \sum_{j_i=1}^{J_i} \lambda_{ij_i}. \quad (3.10)$$

- 广义极大似然法中, 式(2.4)实验 i 中的信号事例数 N_{is} 应考虑为一个泊松变量, 故

$$\text{式(2.4)} \quad N_{is} = A_i B_i, \quad A_i \equiv N_{Ri} \varepsilon_i B R_i, \quad i=1, \dots, I.$$

需用下式代替

$$\underline{\lambda_{is} = A_i B_i, \quad A_i \equiv N_{Ri} \varepsilon_i B R_i, \quad i=1, \dots, I.} \quad (3.11)$$

式 (2.3) 特征变量 m_i 的概率密度 $f_i(m_i | \theta_i)$ 中的 w_{is} 现在为:

$$w_{is} = \lambda_{is} / \lambda_i \quad (3.12)$$

当进行合并估计时, $B = B_i$ 。

由 $\lambda_{is} = w_{is} \lambda_i$, 知 $w_{is} = A_i B / \lambda_i$, 代入式 (2.3) 得

$$\underline{f_i(m_i | \theta_i) = \frac{A_i}{\lambda_i} B \cdot f_{is}(m_i | \theta_{is}) + \left(1 - \frac{A_i}{\lambda_i} B\right) f_{ib}(m_i | \theta_{ib})} \quad (3.13)$$

I 个实验的联合似然函数定义为

$$L(\theta) = \prod_{i=1}^I L_i = \prod_{i=1}^I \prod_{j_i=1}^{J_i} \frac{\lambda_{ij_i} (m_i | \theta_i)^{n_{ij_i}}}{n_{ij_i}!} e^{-\lambda_{ij_i}}, \quad (3.14)$$

不考虑与待拟合参数 θ 无关的项, 有

$$\underline{\ln L(\theta) = \sum_{i=1}^I \left[\sum_{j_i=1}^{J_i} (n_{ij_i} \ln \lambda_{ij_i}) - \lambda_i \right].} \quad (3.15)$$

- 联合似然函数中 $\lambda \equiv \sum_{i=1}^I \lambda_i$ 是否保留取决于 λ 是否依赖于 θ (除 λ 之外的未知参数)。
- 联合似然函数中的未知参数 θ 包括 θ_s, θ_b 和 λ, B :
 $\theta = \{B, \lambda, \theta_s, \theta_b\}, \quad \lambda = \{\lambda_1, \dots, \lambda_I\}, \quad \theta_s = \{\theta_{1s}, \dots, \theta_{Is}\}, \quad \theta_b = \{\theta_{1b}, \dots, \theta_{Ib}\}.$ (3.16)
- 最优化计算中, $\lambda = \{\lambda_1, \dots, \lambda_I\}$ 的初值可取为 $\{N_1, \dots, N_I\}$, θ_s, θ_b 的初值可取为各个实验中确定的值, 而 B 的初值可利用各个实验确定的 B_i 的某种加权平均。

(3.2.2) I 个实验谱为特征变量 $m_i, i=1, \dots, I$ 散点图

- 对于 I 个实验谱为散点图数据的情形，实验 i 散点图的事例数为 N_i ，这 N_i 个事例出现在 m_{i1}, \dots, m_{iN_i} ， $i=1, \dots, I$ ， $j=1, \dots, N_i$ 处。

实验 i 的广义似然函数为

$$L_i(m_i | \theta_i, \lambda_i) = e^{-\lambda_i} \frac{\lambda_i^{N_i}}{N_i!} \prod_{j=1}^{N_i} f_i(m_{ij} | \theta_i),$$

将概率密度 $f_i(m_i | \theta_i)$ 式(3.13)代入上式并略去不依赖于 θ_i 的项，得

$$L_i(m_i | \theta_i, \lambda_i) = e^{-\lambda_i} \prod_{j=1}^{N_i} \left[A_i B \cdot f_{is}(m_{ij} | \theta_{is}) + (\lambda_i - A_i B) f_{ib}(m_{ij} | \theta_{ib}) \right]. \quad (3.17)$$

I 个实验谱的联合广义似然函数为

$$L(\theta) = \prod_{i=1}^I L_i(m_i | \theta_i, \lambda_i) = \prod_{i=1}^I e^{-\lambda_i} \prod_{j=1}^{N_i} \left[A_i B \cdot f_{is}(m_{ij} | \theta_{is}) + (\lambda_i - A_i B) f_{ib}(m_{ij} | \theta_{ib}) \right]. \quad (3.18)$$

从而有

$$\begin{aligned} \ln L(\theta) &= \sum_{i=1}^I \ln L_i \\ &= -\lambda + \sum_{i=1}^I \sum_{j=1}^{N_i} \ln \left[A_i B \cdot f_{is}(m_{ij} | \theta_{is}) + (\lambda_i - A_i B) f_{ib}(m_{ij} | \theta_{ib}) \right]. \end{aligned} \quad (3.19)$$

- 联合似然函数中 $\lambda \equiv \sum_{i=1}^I \lambda_i$ 是否保留取决于 λ 是否依赖于 θ （除 λ 之外的未知参数）。

4. 信度区间和上限的确定，系统误差的考虑

(4.1) 不考虑系统误差时信度区间和上限的确定

- 现在，有了未知参数 θ 的最小二乘估计或极大似然估计 $\hat{\theta}$
(包括分支比合并估计值 \hat{B} 及其统计误差 $\sigma_{B,st}$)。
- 面临的问题是如何给出分支比的测量结果，
即实验报道合并估计值 B 的 CL = 68.27% 区间还是 CL = 90% 的上限？
需要额外根据某种 flip-flopping 策略[2]来决定。
- 用贝叶斯最大后验密度区间确定分支比 B 的 CL = 68.27% 区间或 CL = 90% 的上限。
需要利用贝叶斯后验密度作为 flip-flopping 策略的依据，这就要用到似然函数。
(当利用最小二乘法进行合并估计时，需要额外构建似然函数，这是一个缺陷。)

- $h(B|\mathbf{n})$ 是 B 的后验概率密度:

$$h(B|\mathbf{n}) = \frac{L(\mathbf{n}|B)\pi(B)}{\int L(\mathbf{n}|B)\pi(B)dB} \quad (4.1)$$

\mathbf{n} 表示 I 个实验的观测谱数据。对于直方图和散点图这两种情形, 有

$$\mathbf{n} = \{n_{11}, \dots, n_{1J_1}; \dots; n_{I1}, \dots, n_{IJ_I}\}, \quad (4.2)$$

$$\mathbf{n} = \{m_{11}, \dots, m_{1N_1}; \dots; m_{I1}, \dots, m_{IN_I}\}.$$

$\pi(B)$ 为先验概率密度, 常用的 $\pi(B)$ 是 B 的物理允许变量域 $[0,1]$ 内的均匀分布。

代入上式即得:

$$h(B|\mathbf{n}) = \frac{L(\mathbf{n}|B)}{\int_0^1 L(\mathbf{n}|B)dB} \quad (4.3)$$

计算似然函数 $L(\mathbf{n}|B)$ 时, 给定 B 值, 参数 θ 中除 B 之外的参数都取 $\ln L$ 达到极大时的估计值 $\hat{\theta}$ 。

- 给定信度水平 $CL = \gamma$, 贝叶斯统计下的最优区间称为最大后验密度 (HPD — Highest posterior Density) 区间[3]。参数 B 的信度水平 γ 的 HPD 区间 I 满足以下条件:

$$P(B \in I | \mathbf{n}) = \int_I h(B|\mathbf{n})dB = \gamma, \quad (4.4)$$

且对任意 $B_1 \in I, B_2 \notin I$, 总有

$$h(B_1|\mathbf{n}) \geq h(B_2|\mathbf{n}). \quad (4.5)$$

HPD 区间是给定信度水平 γ 的长度最短的区间。

参数 B 的信度水平 γ 的上限 B_{up} 则为:

$$P(B \leq B_{up} | \mathbf{n}) = \int_{B \leq B_{up}} h(B|\mathbf{n})dB = \gamma. \quad (4.6)$$

- 贝叶斯后验密度 flip-flopping 策略如下:

» 如果满足式(4.7)的 $\gamma = 90\%$ 的 HPD 区间 $R_{0.9}$ 存在:

$$R_{0.9} \in [\tilde{B}_l, \tilde{B}_u], \quad \tilde{B}_l < \tilde{B}_u, \quad h(\tilde{B}_l | \mathbf{n}) = h(\tilde{B}_u | \mathbf{n}), \quad (4.7)$$

则实验结果报道估计值 B 的 68.27% 信度区间 $R_{0.6827}$:

$$R_{0.6827} \in [B_l, B_u], \quad B_l < B_u, \quad h(B_l | \mathbf{n}) = h(B_u | \mathbf{n}), \quad (4.8)$$

且有

$$B = \hat{B}_{-\sigma_-}^{+\sigma_+}, \quad \sigma_+ = B_u - \hat{B}, \quad \sigma_- = \hat{B} - B_l \quad (4.9)$$

其中 \hat{B} 是参数 B 的极大似然估计。

» 如果满足式(4.7)的 $\gamma = 90\%$ 的 HPD 区间 $R_{0.9}$ 不存在:

则根据式(4.6)给出 $CL = 90\%$ 的上限。

- $R_{0.9}$ 存在与否的判据

» 对于后验密度 $h(B|\mathbf{n})$ 为分支比物理区间 $B \in [0,1]$ 内单峰凸函数的情形 (大多数分支比测量实验满足这一条件), 满足式(4.7)的 HPD 区间 $R_{0.9}$ 是否存在可由下列判据来判定:

$$\begin{aligned} h(B=0|\mathbf{n}) &\geq h(B_{up}(CL=0.9)|\mathbf{n}), & R_{0.9} & \text{不存在;} \\ h(B=0|\mathbf{n}) &< h(B_{up}(CL=0.9)|\mathbf{n}), & R_{0.9} & \text{存在。} \end{aligned}$$

- 》 这种贝叶斯后验密度 flip-flopping 策略可用图 2 来表示。
 图(a)为 $R_{0.9}$ 不存在的情形，实验报道分支比 B 的 90% 上限 $B_{up}(CL=0.9)$ ；
 图(b)为 $R_{0.9}$ 存在的情形，实验报道分支比 B 的 68.27% 信度区间 $R_{0.6827} \in [B_l, B_u]$ 。

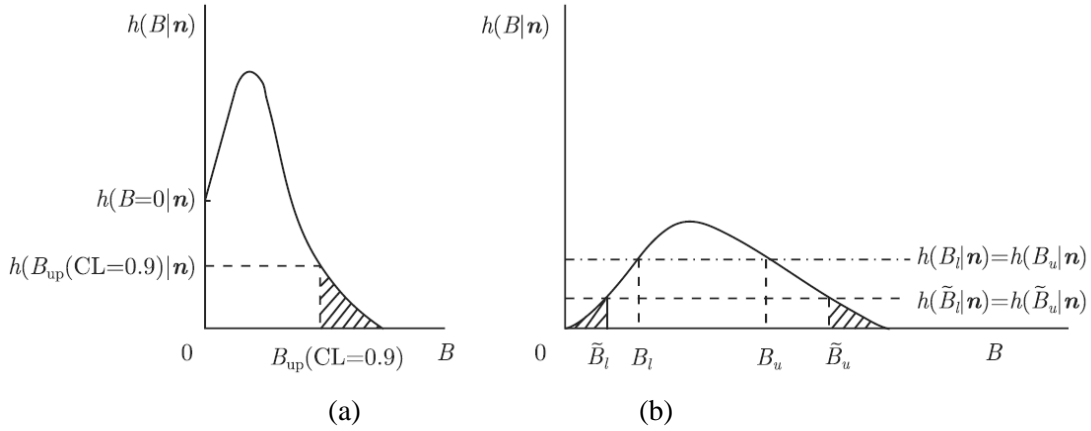


图 2 贝叶斯后验密度 flip-flopping 策略的图示
 (a) $R_{0.9}$ 不存在，报道分支比 B 的 90% 上限 $B_{up}(CL=0.9)$ 。

图中划斜线的区域面积等于 $1 - CL = 0.1$ 。

- (b) $R_{0.9}$ 存在，报道分支比 B 的 68.27% 信度区间 $R_{0.6827} \in [B_l, B_u]$ 。

图中 $h(B|n)$ 在区间 $[B_l, B_u]$ 内的积分等于信度 $CL = 0.6827$ ，

划斜线的两个区域面积之和等于 $1 - CL = 0.1$ 。

- 这样确定的信度区间或上限没有包括 B 的系统误差的贡献。

(4.2) 考虑系统误差时信度区间和上限的确定

(4.2.1) B 的合并估计的系统误差

为了求得 B 的合并估计的系统误差，需要考虑 I 个实验测量的关联性。

- I 个实验是独立测量

B 的系统误差可用多个独立测量的误差合并公式来估计：

$$\sigma_{B,\text{sys}}^2 = \left(\sum_{i=1}^I \sigma_{B_i,\text{sys}}^{-2} \right)^{-1}. \quad (4.10)$$

$$\frac{\sigma_{B_i,\text{sys}}^2}{B_i^2} = \frac{\sigma_{N_{Ri}}^2}{N_{Ri}^2} + \frac{\sigma_{BR_i}^2}{BR_i^2} + \frac{\sigma_{\varepsilon_i}^2}{\varepsilon_i^2} + \frac{\sigma_{N_{ib}}^2}{N_{ib}^2}, \quad (4.11)$$

$\sigma_{B_i,\text{sys}}, \sigma_{\varepsilon_i}, \sigma_{N_{ib}}$ 是第 i 个实验对于 $B_i, \varepsilon_i, N_{ib}$ 系统误差的估计，

N_{ib} 是第 i 个实验中的本底事例数 $N_{ib} = N_i - A_i B_i$ 。

式(4.11)右边的各个量在第 i 个实验的数据分析中应当已经知道。

就其性质，第 i 个实验的 $\sigma_{N_{Ri}}, \sigma_{BR_i}, \sigma_{\varepsilon_i}, \sigma_{N_{ib}}$ 是相互独立的。

- I 个实验测量不独立，

$\sigma_{B_i,\text{sys}}^2$ 中有相互独立的成分 $(\sigma_{B_i,\text{sys}}^2)_{\text{uncom}}$ 和相同的成分 $(\sigma_{B,\text{sys}}^2)_{\text{com}} \equiv (\sigma_{B_i,\text{sys}}^2)_{\text{com}}$ ，

则 B 的系统误差可由下式估计：

$$\sigma_{B,\text{sys}}^2 = \left(\sigma_{B,\text{sys}}^2\right)_{\text{uncom}} + \left(\sigma_{B,\text{sys}}^2\right)_{\text{com}} , \quad (4.12)$$

$$\left(\sigma_{B,\text{sys}}^2\right)_{\text{uncom}} = \left(\sum_{i=1}^I \left(\sigma_{B_i,\text{sys}}^{-2}\right)_{\text{uncom}}\right)^{-1} . \quad (4.13)$$

例如 BES 实验 $\psi(2S) \rightarrow \eta J/\psi$ 分支比测量中, $\sigma_{N_{Ri}}^2$ 是相同的成分,

而 $\sigma_{BR_i}, \sigma_{\varepsilon_i}, \sigma_{N_{ib}}$ 则是相互独立的成分。

(4.2.2) 考虑 B 的系统误差时, 对于似然函数的修正

- 为了精确地描述测量数据, 似然函数不但依赖于感兴趣的参数 B , 同时还依赖于称为冗余参数的附加参数 ν , 即似然函数同时为 B 和 ν 的函数, 用 $L(\mathbf{n} | B, \nu)$ 表示[3]。
- B 的系统误差的效应利用冗余参数 ν 的概率密度 $\pi(\nu)$ 来描述, $\pi(\nu)$ 的合理模型通常为均值为 B 的正态分布 $N(B, \sigma_\nu^2)$, σ_ν 为某个标准离差值, 这里取分支比为 ν 时的系统误差值 $\sigma_{\nu,\text{sys}}$ 。于是 $L(\mathbf{n} | B, \nu)$ 可表示为

$$L(\mathbf{n} | B, \nu) = L(\mathbf{n} | \nu) \cdot \frac{1}{\sqrt{2\pi}\sigma_{\nu,\text{sys}}} \exp\left(-\frac{(\nu - B)^2}{2\sigma_{\nu,\text{sys}}^2}\right), \quad (4.14)$$

这时, 后验密度 $h(B | \mathbf{n})$ 式(4.3)中的似然函数 $L(\mathbf{n} | B)$ 需要代之以

$$\tilde{L}(\mathbf{n} | B) = \int_0^1 L(\mathbf{n} | \nu) \cdot \frac{1}{\sqrt{2\pi}\sigma_{\nu,\text{sys}}} \exp\left(-\frac{(\nu - B)^2}{2\sigma_{\nu,\text{sys}}^2}\right) d\nu. \quad (4.15)$$

- 因此, 当考虑系统误差时, 后验密度 $h(B | \mathbf{n})$ 的表式现在为

$$h(B | \mathbf{n}) = \frac{\tilde{L}(\mathbf{n} | B)}{\int_0^1 \tilde{L}(\mathbf{n} | B) dB}. \quad (4.16)$$

- $\pi(\nu)$ 的选择, $\pi(\nu)$ 的截断, 下端和上端截断

» 当 $\sigma_{\nu,\text{sys}}$ 取为某个常数, 例如 ν 等于 B 的极大似然估计值 \hat{B} 时的系统误差 $\sigma_{\hat{B},\text{sys}}$, 正态分布的 $\pi(\nu)$ 当 $\nu \rightarrow 0$ 和 $\nu \rightarrow 1$ 时出现下端和上端截断。

» 取 $\sigma_{\nu,\text{sys}} = \tilde{\sigma}_{\nu,\text{sys}} \cdot \nu$, $\tilde{\sigma}_{\nu,\text{sys}}$ 是一个常数, 表示分支比为 ν 时的相对误差, 则当 $\nu \rightarrow 0$ 正态分布 $N(\nu, \sigma_\nu^2) \rightarrow 0$, 从而避免了下端截断。

» 选择对数正态或伽玛分布作为 $\pi(\nu)$ 能够避免下端截断, 因为此时 $\nu \in [0, \infty)$;

» 选择贝塔分布作为 $\pi(\nu)$ 能够避免下端和上端截断, 因为此时 $\nu \in [0, 1]$ 。

» 当 $\hat{B} \ll 1$, 上端截断可以忽略。

5. 小结

- 多个实验结果的合并可以提高分支比测量的精度，但是对仅给出上限的实验结果的合并是一个困难的问题。
- 本节提出的合并估计方法对于单个测量结果是否用上限来表示没有限制。
- 利用所有单个实验的实验测量谱数据构建联合 χ^2 函数或联合似然函数，通过极小化/极大化求得分支比的合并估计。
- 利用极大似然法，特别是广义极大似然法进行合并估计能给出更为可靠的结果。
- 对于考虑和不考虑系统误差这两种情形，利用贝叶斯方法给出了确定分支比信度水平 $CL=\gamma$ 的 HPD 区间和上限的相应公式。

文献

- [1] BES collab. Phys. Rev. 2004, (D70) : 012006.
[2] Feldman G, Cousins R. Phys. Rev. 1998(D57): 3873
[3] Olive K A et al. (Particle Data Group), Chinese Physics, 2014,(C38): 090001;
<http://pdgLive.lbl.gov>

