

第7章 假设检验和统计显著性

在实验测量中,除了要得到未知参数的估计值之外,往往希望了解被测物理量所服从的分布.被观测的物理量(随机变量)的分布函数的确切形式通常是未知的,只能以假设的方式提出它所服从的分布,并从统计的观点根据观测值来判断这一假设的合理性.这类问题称为**统计假设的检验**.假设检验的内容十分丰富,希望比较全面了解这方面知识的读者推荐阅读统计书籍中的有关章节(如文献[2]~[9]).本章只讨论实验分析中经常用到的似然比检验和拟合优度检验,7.3节讨论粒子物理实验中与假设检验有关的一个重要概念:信号的统计显著性(statistical significance).

7.1 似然比检验

设随机变量 x 的概率密度为 $f(x|\theta)$, 未知参数 $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$, $\theta \in \Omega$, Ω 为参数空间. 假定零假设 H_0 是对 $\theta_1, \theta_2, \dots, \theta_k$ 中 r 个参数加上 r 个约束条件(如等于某些常数), 使得 θ 被限制在参数空间 Ω 的一个子空间 ω 中. 我们的问题是根据随机变量 x 的容量 n 的子样 $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ 来检验假设

$$H_0: \theta \in \omega, \quad H_1: \theta \in \Omega - \omega. \quad (7.1.1)$$

也就是说, 似然比检验是来检验对于待测的未知参数 θ , 究竟是零假设 H_0 为真, 还是备择假设 H_1 为真.

对于子样值 $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, 似然函数为

$$L = \prod_{i=1}^n f(x_i|\theta).$$

记似然函数在参数空间 Ω 中的极大值为 $L_m(\Omega)$, 而在零假设 H_0 为真的条件下得到的子空间 ω 中似然函数的极大值记为 $L_m(\omega)$. 似然比 λ 定义为

$$\lambda \equiv \frac{L_m(H_0)}{L_m(H_1)} \equiv \frac{L_m(\omega)}{L_m(\Omega)}. \quad (7.1.2)$$

由概率密度的非负性可知, λ 为非负值; 同时, 在子空间 ω 中的极大值 $L_m(\omega)$ 不可能大于整个参数空间 Ω 中的极大值 $L_m(\Omega)$, 故有 $0 \leq \lambda \leq 1$.

不失一般性, 假设 (7.1.1) 通常可写成如下形式:

$$H_0: \theta_i = \theta_{i0}, \quad i = 1, 2, \dots, r \quad (\text{表示为 } \theta_r = \theta_{r0})$$

θ_j 不限定, $j = 1, 2, \dots, s$ (表示为 θ_s 不限定);

$H_1: \theta_i \neq \theta_{i0}, i = 1, 2, \dots, r$ (表示为 $\theta_r \neq \theta_{r0}$)

θ_j 不限定, $j = 1, 2, \dots, s$ (表示为 θ_s 不限定). (7.1.3)

这里原假设限定了参数集的一个子集的数值 ($\theta_r = \theta_{r0}$), 而备择假设 H_1 中这一子集可取 H_0 限定的参数值之外的任意值. 也就是说, r 是备择假设 H_1 与零假设 H_0 的待估计的独立参数个数之差, 或自由度之差. 于是, 现在式 (7.1.2) 定义的似然比是固定 $\theta_r = \theta_{r0}$ 而对 θ_s 求极大的 $L(x|\theta)$ 值与对于全部参数求极大的 $L(x|\theta)$ 值之比. 利用上述记号, 似然比可写为

$$\lambda \equiv \frac{L_m(H_0)}{L_m(H_1)} = \frac{L(x|\theta_{r0}, \theta_s'')}{L(x|\theta_r', \theta_s')}, \quad (7.1.4)$$

式中, θ_s'' 是原假设 H_0 限定的参数空间中的极大似然函数对应的 θ_s 值, 而 θ_r', θ_s' 是全部参数空间 θ 中的极大似然函数对应的 θ_r, θ_s 值.

似然比 λ 是观测值 x_1, x_2, \dots, x_n 的函数. 若 λ 的观测值接近于 1, 那么 H_0 为真时的极大值 $L_m(\omega)$ 与整个参数空间的极大值 $L_m(\Omega)$ 相接近, 这表示 H_0 为真的可能性很大; 反之, 若 λ 很小, 则 H_0 为真的可能性很小. 直观地就可知道, 似然比 λ 是零假设 H_0 的合理的检验统计量.

似然比检验的方法可陈述如下: 令 H_0 为真时, 似然比 λ 的概率密度为 $g(\lambda|H_0)$, 对于给定的显著性水平 α , λ 的临界域由

$$0 < \lambda < \lambda_\alpha \quad (7.1.5)$$

确定, 其中 λ_α 满足

$$\alpha = \int_0^{\lambda_\alpha} g(\lambda|H_0) d\lambda. \quad (7.1.6)$$

如果 λ 的观测值 λ_{obs} 大于 λ_α , 则在水平 α 上接受零假设 H_0 ; 反之, 则拒绝 H_0 .

如果函数 $g(\lambda|H_0)$ 为未知, 只要知道 λ 的某个单调函数的概率密度, 仍然可以进行似然比检验. 设 $y = y(\lambda)$ 是 λ 的单调函数, y 的概率密度 $h(y|H_0)$ 为已知, 则有

$$\alpha = \int_0^{\lambda_\alpha} g(\lambda|H_0) d\lambda = \int_{y(0)}^{y(\lambda_\alpha)} h(y|H_0) dy. \quad (7.1.7)$$

求出变量 y 的临界域

$$y(0) < y < y(\lambda_\alpha), \quad (7.1.8)$$

通过 λ 与 $y(\lambda)$ 的逆变换容易求出临界值 λ_α .

通常似然比 λ (或它的函数) 的严格分布是很难找到的, 这时, H_0 的检验问题相当复杂, 一般需要借助于近似方法. 可以证明^[3], 如果 H_0 对于全部参数 θ 中的 $s+r$ 个参数施加了 r 个约束, 当零假设 H_0 为真时, 如子样容量 n 很大, 则统计量

$$-2\ln\lambda \equiv 2[\ln L_m(H_1) - \ln L_m(H_0)] \quad (7.1.9)$$

渐近地服从 $\chi^2(r)$ 分布. 因此, 可利用统计量 $-2\ln\lambda$ 来检验零假设 H_0 , 通过 $\chi^2(r)$ 分布概率密度函数的积分来确定 λ 的临界域.

7.2 拟合优度检验

拟合优度检验的是被测物理量分布函数是否可用零假设规定的函数形式来描述. 其中皮尔逊 χ^2 检验适用于大样问题的拟合优度检验, 即被测物理量的样本容量很大的情形; 而柯尔莫哥洛夫 (Kolmogolov) 检验和斯米尔诺夫-克拉默-冯·迈希斯 (Smirnov-Cramer-von Mises) 检验则适用于任意样本容量.

7.2.1 皮尔逊 χ^2 检验

假定我们有随机变量 x 的 n 个观测值 x_i ($i = 1, 2, \dots, n$), 它们落入互不相容的 N 个子区间之内 (第 i 子区间的变量值在 $\Delta x_i = (x_{i-1} \rightarrow x_i)$ 之间), 落入 i 子区间的观测值个数记为 n_i (观测频数), 且有 $\sum_{i=1}^N n_i = n$. 一个观测值落入第 i 个子区间的概率为 (当 x 为连续变量)

$$p_i = \int_{\Delta x_i} f(x) dx, \quad i = 1, 2, \dots, N \quad (7.2.1)$$

或 (当 x 为离散变量)

$$p_i = \sum_{j, x_j \in \Delta x_i} q_j, \quad i = 1, 2, \dots, N, \quad (7.2.2)$$

其中, $f(x)$ 是变量 x 的概率密度; $q_j = P(x = x_j)$.

我们欲检验的假设是

$$H_0: p_i = p_{0i}, \quad i = 1, 2, \dots, N, \quad (7.2.3)$$

即 p_i 是否可用某个已知的分布 p_{0i} 来描述. 其中 p_{0i} 应当是归一的:

$$\sum_{i=1}^N p_{0i} = 1, \quad (7.2.4)$$

并有

$$p_{0i} = \int_{\Delta x_i} f_0(x) dx \quad \text{或} \quad p_{0i} = \sum_{j, x_j \in \Delta x_i} q_{0j}, \quad i = 1, 2, \dots, N. \quad (7.2.5)$$

如果假设 H_0 成立, 第 i 子区间的理论频数为 np_{0i} , 它与实际观测到的频数 n_i 的差别应当合理地小, 因此统计量

$$X^2 \equiv \sum_{i=1}^N \frac{(n_i - np_{0i})^2}{np_{0i}} \equiv \frac{1}{n} \sum_{i=1}^N \frac{n_i^2}{p_{0i}} - n \quad (7.2.6)$$

可以作为子样观测值与假设 H_0 的分布 p_{0i} 一致性的检验统计量.

为了确定临界域, 必须知道统计量 X^2 的分布, 我们可定性地作如下考虑: 在每一子区间中的观测频数 n_i 可考虑为泊松变量, 当 H_0 为真时, 其期望值和方差均为 np_{0i} . 当 np_{0i} 足够大时, 泊松变量近似于期望值和方差等于 np_{0i} 的正态变量 $N(np_{0i}, np_{0i})$, 因此

$$X_i = \frac{n_i - np_{0i}}{\sqrt{np_{0i}}} \sim N(0, 1), \quad i = 1, 2, \dots, N. \quad (7.2.7)$$

X^2 正是这 N 个标准正态变量的平方和, 由于条件 (7.2.4) 的存在, 这 N 个变量中只有 $N-1$ 个是独立的. 因此, X^2 近似地服从 $\chi^2(N-1)$ 分布. 统计量的这一性质称为皮尔逊定理: 不论分布 $f_0(x)$ 是何种函数, 当 H_0 为真时, 统计量 X^2 的渐近分布 (当 $n \rightarrow \infty$ 时) 是自由度 $N-1$ 的 χ^2 分布. 这一性质与随机变量的分布函数 $f(x)$ 的形式无关, 因此, 这种检验是分布自由的, 适用于任何总体分布的随机变量的拟合优度检验问题.

当 H_0 为真时, X^2 近似服从 $\chi^2(N-1)$ 分布. 当 H_0 不为真时, 则观测到的 X^2 值 X_{obs}^2 平均地大于 H_0 为真时的 X_{obs}^2 值, 因此过大的 X_{obs}^2 值意味着应当拒绝假设 H_0 . 当满足以下条件时, 在显著性水平 α 上拒绝假设 H_0 :

$$X_{\text{obs}}^2 > \chi_{\alpha}^2(N-1), \quad (7.2.8)$$

其中, 临界值 $\chi_{\alpha}^2(N-1)$ 由 $\chi^2(N-1)$ 分布的累积分布函数确定:

$$1 - \alpha = \int_0^{\chi_{\alpha}^2(N-1)} f(y; N-1) dy. \quad (7.2.9)$$

通常, 假设 H_0 所设定的物理量 x 的分布 p_{0i} (式 (7.2.5)) 中会包含 L 个未定参数. 当利用最小二乘法估计这些参数时, 是通过使式 (7.1.6) 所示的统计量 X^2 达到极小 X_{\min}^2 得到的. 这时 X_{\min}^2 服从 χ^2 分布, 其自由度等于独立观测数 $(N-1)$ 减去待估计的独立参数个数 (L) . 这一关系对于观测数个数无穷多、线性最小二乘估

计的情形严格正确, 在其他情形下则只是一种近似. 因此, 当假设 H_0 中包含 L 个未定参数, 且用最小二乘法估计这些参数值时, 皮尔逊 χ^2 检验在显著性水平 α 上拒绝假设 H_0 的判别条件为

$$X_{\text{obs}}^2 > \chi_{\alpha}^2(N - 1 - L). \quad (7.2.10)$$

由前面的讨论可见, 皮尔逊 χ^2 检验依赖于变量

$$X_i = \frac{n_i - np_{0i}}{\sqrt{np_{0i}}}$$

渐近地服从标准正态分布这一假定 (见式 (7.2.7)), 这就要求落在每一子区间中的事例观测频数 n_i 充分大; 但另一方面, 将 n_i 个观测值归并到同一子区间中, 相当于用随机变量 x 在该子区间中的平均值代替 n_i 个不同的观测值, 这必定导致数据信息的某种损失. 这两者是相互矛盾的. 在实际选择子区间的大小时, 一般遵循的原则是, 使每个子区间内的理论频数 ≥ 5 的条件下, 子区间数目以较多为宜. 当子区间数目不太小 (如 $N - 1 \geq 6$, N 为子区间数) 时, 在一个或两个子区间中理论频数甚至可以小于 5.

子区间的划分有两种方法. 一种是等宽度划分, 即各子区间宽度相等. 当随机变量 X 的取值域为 $(-\infty, \infty)$ 时, 一般在两端概率密度极低, 故两端的两个子区间取为 $(-\infty, x_1)$ 和 (x_{N-1}, ∞) . 这种划分方法的好处是十分简单, 采用得较为广泛. 另一种是等概率划分方法, 即在零假设 H_0 成立的条件下, 使每个子区间内的理论概率相等 $p_{01} = p_{02} = \cdots = p_{0N}$. 当子样容量 n 充分大时, 在等概率划分的条件下, 可以确定出子区间数目 N 的一个最佳值, 使得皮尔逊 χ^2 检验的近似势函数达到极大 (参考文献 [2]). 因此一般说来, 等概率方法在理论上较为优越, 但它的计算较为繁杂, 实际上使用不多.

7.2.2 柯尔莫哥洛夫检验

皮尔逊 χ^2 检验无疑是物理学家使用得最普遍的非参数检验方法. 但它只适用于子样容量大的场合, 而且由于将数据归并到 N 个子区间中, 导致某种程度的信息损失, 降低了检验的有效性. 本节所要阐述的柯尔莫哥洛夫检验方法, 避免了对数据的分组划分, 因而更充分地利用了数据的信息; 同时该方法对任何子样容量都适用, 故对小样问题明显地比皮尔逊 χ^2 检验优越. 但是应当注意, 柯尔莫哥洛夫检验仅适用于连续随机变量, 并且零假设给定的分布不能包含未知参数.

设将被测物理量 x 的 n 个测量值按数值递增的次序排列:

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n,$$

子样分布函数 $S_n(x)$ 定义为

$$S_n(x) = \begin{cases} 0, & x < x_1 \\ \frac{k}{n}, & x_k \leq x < x_{k+1}, \quad k = 1, 2, \dots, n-1 \\ 1, & x_n \leq x \end{cases} \quad (7.2.11)$$

$S_n(x)$ 是一上升的阶梯函数, 在每一观测值 x_1, x_2, \dots, x_n 处阶梯增高 $1/n$.

柯尔莫哥洛夫检验是利用子样分布函数 $S_n(x)$ 来检验随机变量 x 的分布函数 $F(x)$ 是否具有特定的形式 $F_0(x)$

$$H_0: F(x) = F_0(x). \quad (7.2.12)$$

格里汶科定理证明了, 当 $n \rightarrow \infty$ 时, $S_n(x)$ 依概率 1 收敛于累积分布函数 $F(x)$. 可以预期, 当 H_0 为真, 即 x 的累积分布 $F(x) = F_0(x)$ 时, 在任一 x_k ($k = 1, 2, \dots, n$) 处, $S_n(x)$ 与 $F_0(x)$ 的值应当十分接近; 相反, 若在某些 x_k 处 $S_n(x)$ 与 $F_0(x)$ 的值差异很大, 则 x 的分布 $F(x)$ 与给定分布 $F_0(x)$ 处处相符的可能性很小, 即原假设很可能不为真. 因此, $S_n(x)$ 与 $F_0(x)$ 的差异可以用来衡量测量数据与假设 H_0 之间的一致性.

我们定义如下三个随机变量来表征 $S_n(x)$ 与 $F_0(x)$ 的差值:

$$\begin{aligned} D_n^+ &= \max_{-\infty < x < +\infty} (S_n(x) - F_0(x)), \\ D_n^- &= \max_{-\infty < x < +\infty} (F_0(x) - S_n(x)), \\ D_n &= \max_{-\infty < x < +\infty} |S_n(x) - F_0(x)| = \max(D_n^+, D_n^-). \end{aligned} \quad (7.2.13)$$

当总体 X 的分布 $F(x)$ 为连续函数时, 在子样容量 n 很大的情形下, D_n 和 D_n^+ 的极限分布为 (D_n^+ 与 D_n^- 有相同的分布)

$$\lim_{n \rightarrow \infty} P\left(D_n \leq \frac{z}{\sqrt{n}}\right) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 z^2}, \quad z > 0, \quad (7.2.14)$$

$$\lim_{n \rightarrow \infty} P\left(D_n^+ \leq \frac{z}{\sqrt{n}}\right) = 1 - e^{-2z^2}, \quad z > 0. \quad (7.2.15)$$

由式 (7.2.15) 立即有

$$\lim_{n \rightarrow \infty} \{4n(D_n^+)^2 \leq z\} = 1 - e^{-\frac{z}{2}}.$$

与 $\chi^2(n)$ 的累积分布函数对比即知, 当 $n \rightarrow \infty$ 时

$$D_n^+ \sim \sqrt{\frac{\chi^2(2)}{4n}}. \quad (7.2.16)$$

可见, 统计量 D_n, D_n^+, D_n^- 的分布与总体分布 $F(x)$ 无关, 且这一结论对任意子样容量 n 皆为正确. 因此, 利用 D_n, D_n^+, D_n^- 作为检验统计量的柯尔莫哥洛夫检验是分布自由的, 适用于任何连续总体.

对于不同的备择假设, 需选用不同的检验统计量, 见表 7.1. 当 H_0 为真时, $D_n(D_n^+, D_n^-)$ 接近于 0; 当 H_0 不为真而备择假设 H_1 为真时, $D_n(D_n^+, D_n^-)$ 有增大的趋势. 因此, $D_n(D_n^+, D_n^-)$ 的临界域在其分布的上侧. 给定显著性水平 α , 其临界值 $D_{n,\alpha}^{(+,-)}$ 由

$$P\{D_n^{(+,-)} > D_{n,\alpha}^{(+,-)}\} \leq \alpha \quad (7.2.17)$$

给出. 对于五种不同的显著性水平 α 值, $D_{n,\alpha}^{(+,-)}$ 的临界值列于附表 13, 表的最后一行给出了 n 趋于 ∞ 时的临界值 $D_{n,\alpha,n \rightarrow \infty}$, 并且总有

$$D_{n,\alpha} < D_{n,\alpha,n \rightarrow \infty}.$$

对于单侧备择假设的情形, 当 n 很大时, 还可利用式 (7.2.16) 来确定临界值

$$D_{n,\alpha}^{+,-} = \sqrt{\frac{\chi_{\alpha}^2(2)}{4n}}, \quad (7.2.18)$$

其中, $\chi_{\alpha}^2(2)$ 是 $\chi^2(2)$ 分布的上侧 α 分位数.

表 7.1 柯尔莫哥洛夫检验中的备择假设和检验统计量

原假设	备择假设	检验统计量	注
$H_0 : F(x) = F_0(x)$	$H_1 : F(x) \neq F_0(x)$	D_n	双侧备择假设
$H_0 : F(x) = F_0(x)$	$H_1 : F(x) > F_0(x)$	D_n^+	单侧备择假设
$H_0 : F(x) = F_0(x)$	$H_1 : F(x) < F_0(x)$	D_n^-	单侧备择假设

这样, 根据一组观测值 x_1, x_2, \dots, x_n 来检验零假设 H_0 就变得十分简单. 对于所需要的备择假设选择适当的统计量 D_n 或 D_n^+, D_n^- , 将观测值 x_1, x_2, \dots, x_n 代入式 (7.2.13) 得到 $D_n^{(+,-)}$ 的实际观测值 $D_{n,\text{obs}}^{(+,-)}$, 如果 $D_{n,\text{obs}}^{(+,-)} > D_{n,\alpha}^{(+,-)}$, 则在显著性水平 α 上拒绝 H_0 而接受备择假设; 反之, 则接受原假设 H_0 .

对 $H_1 : F(x) \neq F_0(x)$ 的情形, 式 (7.2.17) 结合式 (7.2.13) 可重新表述为

$$P\{S_n(x) - D_{n,\alpha} \leq F_0(x) \leq S_n(x) + D_{n,\alpha}\} \geq 1 - \alpha, \quad \text{对所有 } x. \quad (7.2.19)$$

式 (7.2.19) 表示累积分布函数 $F_0(x)$ 大于 $S_n(x) - D_{n,\alpha}$ 而小于 $S_n(x) + D_{n,\alpha}$ 的概率大于 $1 - \alpha$, 即真的累积分布函数 $F_0(x)$ 落在 $[S_n(x) - D_{n,\alpha}, S_n(x) + D_{n,\alpha}]$ 区域内的置信概率大于 $1 - \alpha$. 这一关系可以用来估计为了使测量值能反映总体分布达到一定的精度, 需要多少次测量 (即子样容量 n). 例如, 要求在置信水平 90% 上, 实验测量得到的分布反映真实分布 $F_0(x)$ 的精度好于 0.20. 因为 D_n 是子样分布与真分布 $F_0(x)$ 的差值, 它是测量值反映 $F_0(x)$ 的精度度的直接表示, 所以上述要求相当于 $\alpha = 10\%$, $D_n \leq 0.20$. 从附表 13 可以查到, 为了使 $D_n \leq 0.20$, 子样容量必须满足 $n > 35$. 也就是说, 36 次测量构成的子样分布与总体真分布的差别在置信水

平 90%时小于 20%. 类似地, 当要求精度好于 0.05 时, 由附表 13 可查得 n 需大于等于 600.

需要强调指出, 柯尔莫哥洛夫检验中零假设给定的分布不能包含未知参数. 当总体理论分布包含未知参数时, 即为 $F_0(x, \theta)$ 的情形, 需要用子样 x_1, x_2, \dots, x_n 来估计参数 θ 得到其估计值 $\hat{\theta}_n$, 但利用 $F_0(x, \hat{\theta}_n)$ 作为理论分布得到的 D_n, D_n^+, D_n^- 统计量不再是分布自由的, 而是与 $F_0(x, \theta)$ 有关. 这样, 式 (7.2.14) 和式 (7.2.15) 的极限分布不再成立, 无法据此确定临界域.

作为柯尔莫哥洛夫检验的一个例子, 我们来考察一个低统计实验. 假定观测了中性 K 介子衰变为 $\pi^+e^-\nu$ 的 30 个事例, 其衰变时间谱如图 7.1(a) 中折线 $S_{30}(t)$ 所示. 希望检验该中性 K 介子是不是 \bar{K}^0 . 当认为这些介子是 \bar{K}^0 时 (零假设 H_0), 衰变时间谱如图 7.1(a) 中的连续曲线 $F_0(t)$ 所示.

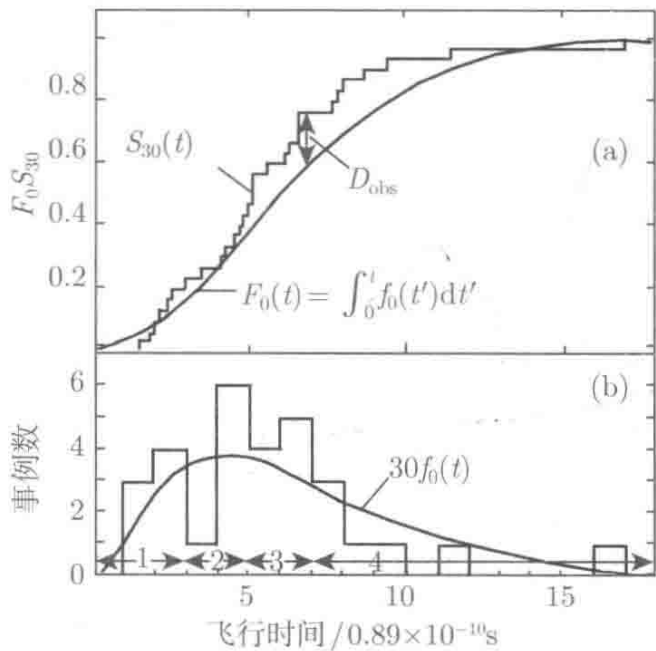


图 7.1 衰变时间谱的理论预期和实验测量值

(a) 累积分布 (柯尔莫哥洛夫检验); (b) 微分分布 (皮尔逊 χ^2 检验)

从测量和理论曲线确定了

$$D_{30,\text{obs}} = \max_{0 < t < \infty} |S_{30}(t) - F_0(t)| = 0.17.$$

选定显著性水平 $\alpha = 0.010$, 从附表 13 查得

$$D_{30,\alpha=0.10} = 0.2176,$$

因此不能拒绝零假设 H_0 .

作为对比, 利用同样的观测数据对 H_0 作皮尔逊 χ^2 检验. 将衰变时间划分为 4 个子区间: $0\sim3, 3\sim5, 5\sim7, 7\sim18$ (单位 $0.89\times10^{-10}\text{s}$), 如图 7.1(b) 所示, 这样的划

分满足每一子区间内观测频数 ≥ 5 的要求, 而且每一子区间内的理论概率积分值大致相等. 这时, 式 (7.2.6) 表示的统计量 X^2 渐近地服从 $\chi^2(3)$ 分布, 由 χ^2 分布表查得

$$\chi_{\alpha=0.10}^2(3) = 6.251,$$

而观测值算出

$$X_{\text{obs}}^2 = 3.0 < \chi_{0.10}^2(3).$$

因此, 皮尔逊检验同样得到不能拒绝原假设 H_0 的结论.

7.2.3 斯米尔诺夫-克拉默-冯·迈希斯检验

与柯尔莫哥洛夫检验利用子样分布函数与原假设总体累积分布函数之差的极大值作为拟合优度的检验统计量不同, 斯米尔诺夫-克拉默-冯·迈希斯检验利用子样分布函数与原假设总体累积分布函数的偏差 W^2 作为检验统计量, 其定义为

$$W^2 = \int_{-\infty}^{\infty} [S_n(x) - F_0(x)]^2 f_0(x) dx. \quad (7.2.20)$$

其中, $f_0(x)$ 是原假设的概率密度; $F_0(x)$ 是原假设的分布函数. 将式 (7.2.11) 的子样分布函数代入式 (7.2.20), 注意到 $F(-\infty) \equiv 0, F(+\infty) \equiv 1$, 可得^[3]

$$\begin{aligned} W^2 &= \int_{-\infty}^{x_1} F_0^2(x) dF_0(x) + \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} \left[\frac{i}{n} - F_0(x) \right]^2 dF_0(x) \\ &\quad + \int_{x_n}^{\infty} [1 - F_0(x)]^2 dF_0(x) \\ &= \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^n \left[F_0(x_i) - \frac{2i-1}{2n} \right]^2. \end{aligned} \quad (7.2.21)$$

该式可用于检验统计量 W^2 的具体计算.

对于一特定的 x 值, $nS_n(x)$ 为期望值 $nF_0(x)$ 的二项分布变量, 故有

$$\begin{aligned} E[S_n(x)] &= F_0(x), \\ E\{[S_n(x) - F_0(x)]^2\} &= \frac{1}{n} F_0(x) [1 - F_0(x)]. \end{aligned}$$

式 (7.2.21) 所示的统计量具有均值和方差:

$$\begin{aligned} E(W^2) &= \frac{1}{n} \int_0^1 F_0(x) [1 - F_0(x)] dF_0(x) = \frac{1}{6n}, \\ V(W^2) &= E(W^4) - [E(W^2)]^2 = \frac{4n-3}{180n^3}. \end{aligned} \quad (7.2.22)$$

即使对于有限的 n 值, 统计量 W^2 的分布也与总体 X 的分布完全无关, 这一点容易证明. 作变量代换 $Y = F_0(x)$, 统计量 W^2 变成

$$W^2 = \int_0^1 [S_n(Y) - Y]^2 dY,$$

它完全独立于原假设的概率密度 $f_0(x)$, 也即统计量 W^2 的分布与原假设的分布相独立.

斯米尔诺夫证明了 [3], 在 $n \rightarrow \infty$ 的极限情形下, 当原假设为真时, nW^2 的特征函数为

$$\varphi(t) = \left\{ (2it)^{1/2} / \sin[(2it)^{1/2}] \right\}^{1/2}. \tag{7.2.23}$$

Anderson 和 Darling 通过 $\varphi(t)$ 的反演计算了 nW^2 的临界值, 对于某些常用的显著性水平 α , nW^2 临界值列于表 7.2. α 及其对应的 nW^2 临界值更详细的表见书末的附表 14. Marshall 证明了, nW^2 趋于其极限分布的速度很快, 以至于在 $n \geq 3$ 的情形下, 表 7.2 所列的 nW^2 的临界值都是精确的. 因此, 可利用此表基于统计量 W^2 进行关于原假设 $F(x) = F_0(x)$ 的检验. 由 W^2 的定义 (7.2.20) 可知, nW^2 的临界域在其临界值的上侧.

表 7.2 斯米尔诺夫-克拉默-冯·迈希斯检验统计量 nW^2 的临界值

α	nW^2 的临界值
0.10	0.347
0.05	0.461
0.01	0.743
0.001	1.168

与柯尔莫哥洛夫检验相同, 斯米尔诺夫-克拉默-冯·迈希斯检验中零假设给定的分布不能包含未知参数, 否则统计量 nW^2 不再是分布自由的, 式 (7.2.21) 的分布不再成立, 无法据此确定临界域.

7.3 信号的统计显著性

实验测量值通常同时包含信号和本底的贡献, 而且信号和本底的测量都存在统计涨落, 即它们都是随机变量. 当在信号区观测到的事例数 n 明显地高于预期的本底事例数 b 时, 我们会判断观测到了信号事例. 例如, 在图 5.7(b) 中不变质量 $m_{\eta\pi^+\pi^-} \approx 958\text{MeV}$ 附近观测到了 η' 粒子的信号. 显然, n 比 b 大出越多, 对于“观测到了信号事例”这一判断的可信程度越高. 信号的统计显著性是物理学家对“观测到了信号事例”这一判断的定量化表征.

在讨论信号的统计显著性问题时, 零假设 H_0 通常表示为观察到的实验现象可以只用已知的现象或本底函数圆满地描述, 备择假设 H_1 则表示观察到的实验现象需要用已知的现象或本底函数, 加上未知、待寻找的新信号过程的贡献才能完整地描述. 信号的统计显著性就是观察到的实验现象偏离已知的现象或本底函数, 发现新信号、新过程的定量表征. 信号的统计显著性越高, 发现新过程的可信度越大. 在粒子物理实验中, 基本上达成了一种共识, 如果实验中观察到一种新信号其显著性 $S \geq 5$, 则可以认为“发现”了一种新信号; 当新信号的显著性 $S \geq 3$ ($S \geq 2$) 时, 则只能说新信号的存在有强(弱)的证据. 所以在寻找新现象的实验测量中, 信号的统计显著性尤为重要.

7.3.1 实验 P 值

粒子物理实验中, 对是否观测到信号事例这一判断的定量表征方法之一是给出实验 P 值, 它定义为

$$P(u_{\text{obs}}) = P(u \geq u_{\text{obs}} | H_0) = \begin{cases} \sum_{u \geq u_{\text{obs}}} p_u(H_0), & u \text{ 离散型,} \\ \int_{u \geq u_{\text{obs}}} f(u|H_0) du, & u \text{ 连续型.} \end{cases} \quad (7.3.1)$$

其中, u 是实验观测量(随机变量)或用实验观测量构造的统计量; u_{obs} 是某个实验测量到的 u 值; $p_u(H_0)$ 是零假设 H_0 为真时离散型变量 u 的概率分布; $f(u|H_0)$ 是零假设 H_0 为真时连续型变量 u 的概率密度函数; 实验 P 值是 H_0 为真时 $u \geq u_{\text{obs}}$ 的概率. 不失一般性, 我们设定 u 值越大, H_0 为真的可能性越小. 那么, $P(u_{\text{obs}})$ 值就是该实验测量值 u_{obs} 与零假设 H_0 不一致性的某种定量表征, $P(u_{\text{obs}})$ 值越小, H_0 为真的可能性越小.

假定存在某种信号事例, 它的 u 值集中地出现在一个特定的区域, 称为信号区. 实验中信号区内观测到的信号事例数可视为期望值 s 的泊松变量, 信号区内观测到的本底事例数可视为期望值 b 的泊松变量, 则信号区内观测到总事例数 n 的概率为

$$f(n; s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)}. \quad (7.3.2)$$

假定实验观测到的信号区内总事例数为 n_{obs} , 令零假设 H_0 为观测到的事例仅仅是由于本底的贡献, 则相应的实验 P 值为

$$P(n_{\text{obs}}) = P(n \geq n_{\text{obs}} | H_0) = \sum_{n=n_{\text{obs}}}^{\infty} f(n; s=0, b) = 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b}. \quad (7.3.3)$$

实验 P 值越小, H_0 为真(观测到的事例仅仅是由于本底的贡献)的可能性越小. 举一个数值例子: 假定 $b = 0.6$, $n_{\text{obs}} = 5$, 则相应的实验 P 值约等于 4×10^{-4} , 这表示实验观测到 $n \geq n_{\text{obs}} = 5$ 个事例仅仅是由于本底涨落导致的概率小到只有 4×10^{-4} .

在实际应用中需要注意信号区内平均本底事例数 (即期望值) b 可能存在系统误差. 通常信号区内平均本底事例数是由信号区外附近的本底区 (称为边带区) 的事例数分布确定的, 这实际上隐含着两个假设: 边带区内不包含信号事例, 边带区内事例分布的函数形式能相当好地反映信号区内的本底贡献. 然而实际的测量数据中边带区内观测到的事例数分布往往不够平滑, 边带区的宽度也没有严格的规则加以判定, 而是依赖于实验者对“边带区内不包含信号事例, 且其事例分布的函数形式能相当好地反映信号区内的本底贡献”这一假设的主观判断予以确定, 因此存在不确定性, 通常利用不同的本底函数形式、不同的边带区宽度来确定 b 的系统误差. 例如, 在上面的例子中, $b = 0.6$, 假定考虑了系统误差后信号区内平均本底事例数 b 的范围为 $0.5 \sim 0.7$, 则相应的实验 P 值范围为 $2 \times 10^{-4} \sim 8 \times 10^{-4}$. 作为最后结果, 可以报道实验 P 值的这一范围, 或者保守地仅报道实验 P 值为 8×10^{-4} .

另一个问题是信号区的宽度. 不同的信号区宽度给出不同的 n_{obs} 和 b 值, 从而得到不同的实验 P 值. 一般信号区的宽度取得足以包含绝大部分的信号事例 (如果存在). 例如, 对于信号的实验分布为正态函数的情形, 信号区间取为 $x_c \pm 2.5\sigma$ 或 $x_c \pm 3\sigma$ 都是合理的选择, 这里 x_c 和 σ 分别是正态函数的中心值和标准偏差, 它们分别包含了全部信号事例的 98.8% 和 99.7%. 报道最终结果时推荐同时给出信号区间的宽度和实验 P 值.

7.3.2 信号的统计显著性

实验 P 值表示的是实验观测到的现象仅仅是由于本底涨落导致的概率, 但是实验观测到的现象与待寻找的新现象、新信号之间的关系表达得不直观, 后者通常用信号的统计显著性 S 来表示. 然而在粒子物理实验数据分析的发展过程中, 使用了不同的统计显著性 S 的定义.

1. “计数实验”中信号统计显著性的定义

所谓的**计数实验** (counting experiment), 是指利用信号区内的信号事例数和本底事例数来确定信号统计显著性的一种做法. 文献 [119]、[120] 对于信号统计显著性 S 有如下各种定义:

$$S_1 = \frac{s}{\sqrt{b}} \quad (7.3.4)$$

$$S_2 = \frac{s}{\sqrt{s+b}} \quad (7.3.5)$$

$$S_{12} = \sqrt{s+b} - \sqrt{b} \quad (7.3.6)$$

$$S_{B1} = S_1 - k(\alpha) \sqrt{1 + \frac{s}{b}} \quad (7.3.7)$$

$$S_{B12} = 2S_{12} - k(\alpha) \quad (7.3.8)$$

$$\int_{-\infty}^{S_N} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \sum_{i=0}^{n-1} \frac{b^i}{i!} e^{-b}. \quad (7.3.9)$$

其中, s 是信号区内信号事例数的期望值, b 是信号区内本底事例数的期望值, 后者被认为是一个已知值; S 的不同角标表示来源于不同的定义; $k(\alpha)$ 是一个与观察到信号事例的确定程度相关的系数. 对于它们的详细说明请阅读文献 [120].

统计显著性的以上不同定义对应于 s 和 b 的不同假定^[121]. 其中, S_1 只考虑本底的涨落, 并假定本底服从期望值和方差均等于 b 的正态分布, 即 $b \sim N(b, b)$. S_2 认为信号区内观测到的事例总数 n 服从期望值 $s+b$ 、方差 b 等于本底期望值的正态分布, 即 $n \sim N(s+b, b)$. S_{12} 则假定信号服从 $N(s, s)$ 分布而本底服从 $N(b, b)$ 分布. S_N 假定 n 、 s 、 b 是服从期望值 n 、 s 、 b 的泊松变量. 因此, 应该按照实验中 n 、 s 、 b 所服从的实际分布来选择统计显著性的计算公式.

2. 用似然函数定义的信号统计显著性

在许多情形下, 通过实验测量值或它的统计量的分布来检验零假设 H_0 和备择假设 H_1 比简单地通过信号区内的计数来检验要更为准确和精确, 这时, 需要用似然比方法进行检验. 假定 $L(b)$ 和 $L(s+b)$ 分别为用零假设 H_0 和备择假设 H_1 的概率密度构造的似然函数, $L_m(b)$ 和 $L_m(s+b)$ 分别为用 $L(b)$ 和 $L(s+b)$ 拟合实验数据得到的似然函数极大值, I. Narsky^[119] 在 $-2 \ln[L_m(b)/L_m(s+b)]$ 服从自由度为 1 的 χ^2 分布的假定下给出了信号的统计显著性

$$S = [2(\ln L_m(s+b) - \ln L_m(b))]^{1/2}. \quad (7.3.10)$$

似然函数可以取为标准的形式

$$L(s+b) = \prod_{i=1}^N [\omega_s f_s(u_i) + (1 - \omega_s) f_b(u_i)], \quad (7.3.11)$$

$$L(b) = \prod_{i=1}^N f_b(u_i), \quad (7.3.12)$$

其中, N 为实验中信号拟合区间内观察到的总事例数 (拟合区间的概念见 2.2.5 节的讨论). 需要特别指出, N 与前面提到的信号区内观测到的总事例数 n 是不同的, 因为实验的信号拟合区间总是比信号区要宽, 所以 N 总大于 n ; 式 (7.3.10)~式 (7.3.12) 中 s 和 b 分别表示信号拟合区内的信号和本底事例数, 实验测量值或它的统计量表示为 u , u_i 为第 i 个事例的测量值, $f_s(u)$ 和 $f_b(u)$ 分别表示信号拟合区内信号和本底的概率密度函数, ω_s 为信号事例数的权因子 (占全部 N 个事例的比例).

似然函数也可以取为广义似然函数的形式, 即实验中观察到的总事例数 N 不认为是一常数而认为是一泊松变量, 这就考虑了观察事例数泊松分布的不确定性, 这时的似然函数形式为

$$L(s + b) = \frac{e^{-(s+b)}}{N!} \prod_{i=1}^N [sf_s(u_i) + bf_b(u_i)], \tag{7.3.13}$$

$$L(b) = \frac{e^{-b}b^N}{N!} \prod_{i=1}^N f_b(u_i). \tag{7.3.14}$$

3. 用实验 P 值定义的信号统计显著性

由于实验 P 值表示的是实验数据与零假设之间的不一致性, 实验 P 值越小, 零假设为真的可能性越小, 备择假设为真的可能性越大. 因此文献 [122] 将信号显著性与实验 P 值联系起来, 对信号显著性作了如下的定义:

$$\int_{-S_P}^{S_P} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - P(u \geq u_{\text{obs}} | H_{\text{bg}}) \equiv 1 - P(u_{\text{obs}}). \tag{7.3.15}$$

该式的左边是正态分布在 $\pm S_P$ 个标准差 ($\pm S_P \sigma$) 内的积分概率. 在这样的定义下, 相对应的 S 值和实验 P 值列于表 7.3.

表 7.3 实验 P 值与统计显著性 S 的对应关系

S 值	实验 P 值
1	0.3173
2	0.0455
3	0.0027
4	6.3×10^{-5}
5	5.7×10^{-7}

对于计数实验的情形, 实验 P 值为

$$P(n_{\text{obs}}) = P(n \geq n_{\text{obs}} | H_0) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{b^n}{n!} e^{-b} = 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b}. \tag{7.3.16}$$

代入式 (7.3.15), 立即有

$$\int_{-S_P}^{S_P} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b}. \tag{7.3.17}$$

与式 (7.3.9) 比较, 仅积分下限有所不同.

对于似然比方法的情形, 如前所述, 假定零假设 H_0 是实验数据仅由本底似然函数 $L(b)$ 描述, 备择假设 H_1 是实验数据由 $L(s + b)$ 描述, 假定 $L(b)$ 有 m 个待定

参数 (描述本底概率密度 $f_b(u)$ 的参数), $L(s+b)$ 有 $k(>m)$ 个待定参数 (描述本底概率密度 $f_b(u)$ 和信号概率密度 $f_s(u)$ 的参数以及它们之间的相对权因子). H_0 (似然函数 $L(b)$) 是对 H_1 (似然函数 $L(s+b)$) 中 k 个待定参数中的 r 个参数加以固定, 或加上了 r 个约束条件; 或者说, r 是备择假设 H_1 与零假设 H_0 的待估计的独立参数个数之差, 或自由度之差, 于是似然比统计量

$$\lambda = L_m(b)/L_m(s+b) \quad (7.3.18)$$

是零假设 H_0 的合理的检验统计量, 这里 $L_m(b)$ 是利用 $L(b)$ 拟合实验数据得到的极大似然函数值. 文献 [3] 证明了, 当 H_0 为真时, 在子样容量很大的情形下, 统计量

$$u = -2 \ln \lambda = 2(\ln L_m(s+b) - \ln L_m(b)) \quad (7.3.19)$$

渐近地服从 $\chi^2(r)$ 分布 (Wilks 定理). 当 λ 的观测值接近于 1 时, H_0 为真的可能性很大; 当 λ 的观测值接近于 0 时, H_0 为真的可能性很小. 所以, λ 的临界域在 λ 值接近 0 的区域, 相应地, u 的临界域在 u 值大的区域. 若 u 的实验观测值为 u_{obs} , 则由式 (7.3.1) 知实验 P 值为

$$P(u_{\text{obs}}) = \int_{u_{\text{obs}}}^{\infty} \chi^2(u; r) du. \quad (7.3.20)$$

代入式 (7.3.15) 立即得到利用似然比统计量计算信号统计显著性 S 的表达式

$$\int_{-S_P}^{S_P} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_0^{u_{\text{obs}}} \chi^2(u; r) du. \quad (7.3.21)$$

对于 $r=1$ 的特殊情形, 有

$$\int_{-S_P}^{S_P} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_0^{u_{\text{obs}}} \chi^2(u; 1) du = 2 \int_0^{\sqrt{u_{\text{obs}}}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

立即可得

$$S_P = \sqrt{u_{\text{obs}}} = [2(\ln L_m(s+b) - \ln L_m(b))]^{1/2}. \quad (7.3.22)$$

它与 I. Narsky 的式 (7.3.10) 完全一致.

由以上讨论可知, 式 (7.3.15) 定义的信号统计显著性无论对于计数实验或似然比方法都是适用的, 因此避免了多重定义. 利用似然比统计量计算信号统计显著性的式 (7.3.21) 的优点在于避免了确定信号区和本底边带区时带来的不确定因素, 但由以上的推导过程可知必须满足大子样容量 (即 N 充分大) 的要求. 对于小子样容量的情形, 仍然应当用式 (7.3.17) 计算信号统计显著性 S .

由于 S 值有多种定义, 所以实验结果的报道中, 在陈述某个信号的统计显著性 S 值为多大时, 应当说明其明确的定义.

4. 示例

假定不变质量 $m_{\eta'}$ 分布的实验数据如表 7.4 所示, 试计算信号 η' (958) 的统计显著性 S .

表 7.4 不变质量 $m_{\eta'}$ 分布实验数据

区间	事例	区间	事例	区间	事例	区间	事例	区间	事例
855		1005		1155	2	1305	3	1455	5
870	1	1020		1170	1	1320	2	1470	4
885		1035	1	1185	3	1335	3	1485	7
900		1050		1200	1	1350	2		
915	1	1065	1	1215	3	1365	4		
930		1080		1230	2	1380	3		
945	3	1095	1	1245	2	1395	5		
960	2	1110	1	1260	3	1410	4		
975	1	1125	2	1275	2	1425	4		
990		1140	1	1290	4	1440	5		

注: 表中“区间”中的数字表示区间下限 (MeV), 上限是该数字加 15MeV. “事例”中无数字表示事例数为 0.

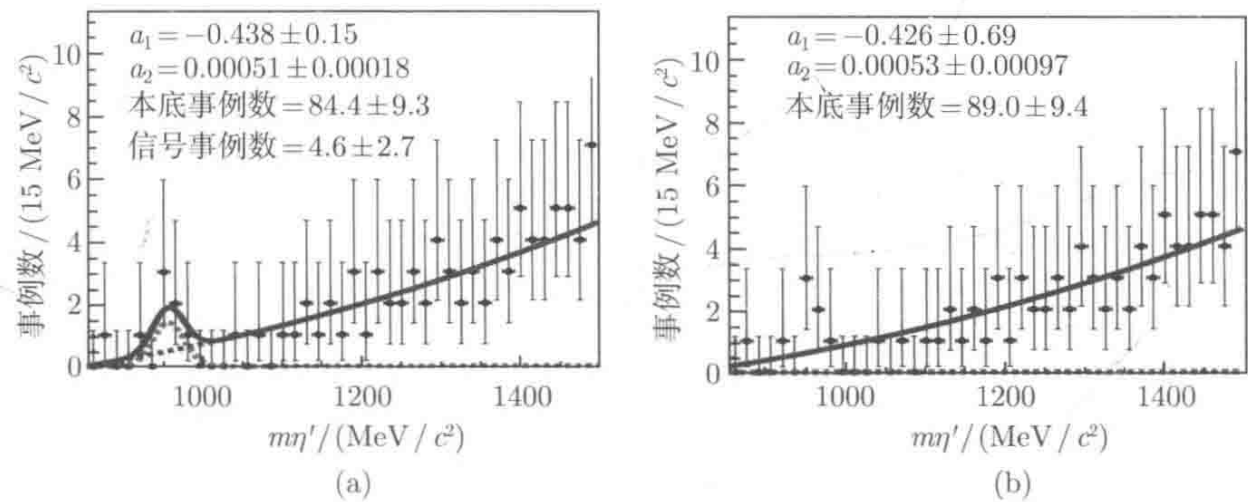


图 7.2 信号 η' (958) 统计显著性计算的图示

(a) 有信号 η' (958) 的极大似然拟合; (b) 无信号 η' (958) 的拟合

首先用似然比方法计算信号的统计显著性. 拟合区间 $[a,b] = [855,1500](\text{MeV})$, 子区间宽度为 15MeV. 本底函数和信号函数取为

$$f_b(m) = c_1 + c_2m + c_3m^2,$$

$$f_s(m) = N(m_0, \sigma^2) = N(958, 20^2).$$

极大似然法应用于直方图数据, 第 i 个子区间内事件数为 n_i ($i = 1, 2, \dots, N$) 的联合概率 (似然函数) 由多项分布给出:

$$L(n_1, n_2, \dots, n_N | \vartheta) = n! \prod_{i=1}^N \frac{1}{n_i!} p_i^{n_i},$$

式中, p_i 是第 i 个子区间中出现一个事例的概率. p_i 可由概率密度 $f(x|\vartheta)$ 在第 i 个子区间内的积分求出:

$$p_i = p_i(\vartheta) = \int_{\Delta m_i} f(m|\vartheta) dm,$$

于是

$$\ln L(n_1, n_2, \dots, n_N | \vartheta) = \sum_{i=1}^N n_i \ln p_i(\vartheta) - \sum_{i=1}^N \ln n_i! + \ln n!.$$

求 $\ln L(n_1, n_2, \dots, n_N | \vartheta)$ 的极大值 $\ln L_m$ 时可去除与参数 ϑ 无关的项 $\sum_{i=1}^N \ln n_i!$ 和 $\ln n!$:

$$\left. \frac{\partial \ln L}{\partial \theta} \right|_{\theta=\hat{\theta}} = \frac{\partial}{\partial \theta} \left[\sum_{i=1}^N n_i \ln p_i(\theta) \right]_{\theta=\hat{\theta}} = 0.$$

对于零假设 H_0 , 实验数据仅由本底似然函数 $L(b)$ 描述, 由

$$p_i = p_i(\vartheta) = \int_{\Delta m_i} f_b(m|\vartheta) dm, \quad \vartheta = c_1, c_2, c_3$$

求出极大值 $\ln L_m(b) = 241.975$, $c_1 = 6.14 \times 10^{-6}$, $c_2 = -2.62 \times 10^{-6}$, $c_3 = 3.25 \times 10^{-9}$.

对于备择假设 H_1 , 实验数据由 $L(s+b)$ (信号 + 本底) 函数描述, 由

$$p_i = p_i(\vartheta) = \int_{\Delta m_i} [w_s f_s + (1 - w_s) f_b] dm, \quad \vartheta = c_1, c_2, c_3, w_s$$

求出极大值 $\ln L_m(s+b) = 244.577$ 以及

$$w = 0.0545, \quad c_1 = 7.38 \times 10^{-6}, \quad c_2 = -3.23 \times 10^{-6}, \quad c_3 = 3.76 \times 10^{-9}.$$

$L(s+b)$ 有 4 个待定参数, $L(b)$ 有 3 个待定参数, 统计量

$$u = -2 \ln \lambda = 2(\ln L_m(s+b) - \ln L_m(b))$$

渐近地服从 $\chi^2(1)$ 分布. 按式 (7.3.22) 计算统计显著性 S :

$$\begin{aligned} S &= \sqrt{u_{\text{obs}}} = [2(\ln L_m(s+b) - \ln L_m(b))]^{1/2} \\ &= \sqrt{2(244.577 - 241.975)} = 2.28. \end{aligned}$$

作为比较, 利用计数法计算信号的统计显著性. 假定信号区间为 $[900, 1020]$ (MeV)(中心值 $\pm 3\sigma$), 信号区间内观测总事例数为 $n = 7$, 本底期望值可以利用实验数据由 $L(s+b)$ 描述时求出的 c_1, c_2, c_3 计算得到 $b = 3.8$, 由此求得

$$S_1 = (n - b)/\sqrt{b} = \frac{7 - 3.8}{\sqrt{3.8}} = 1.64,$$

$$S_2 = \frac{n-b}{\sqrt{n}} = \frac{7-3.8}{\sqrt{7}} = 1.21,$$

$$S_{12} = \sqrt{n} - \sqrt{b} = 0.70,$$

$$\int_{-\infty}^{S_N} N(0,1)dx = \sum_{i=0}^{n-1} \frac{b^i}{i!} e^{-b} = 0.9091, \quad S_N = 1.34,$$

$$\int_{-S_P}^{S_P} N(0,1)dx = \sum_{i=0}^{n-1} \frac{b^i}{i!} e^{-b} = 0.9091, \quad S_P = 1.69.$$

可见, 利用不同的信号显著性定义会得到相当不同的数值, 因此在陈述某个信号的统计显著性 S 值为多大时, 说明其明确的定义是十分必要的.