

Statistics in Exp. Data

(小信号测量的区间估计及统计显著性)

朱永生

中国科学院高能物理所, 北京 100049

中国科学院大学, 北京 100049

Madgraph School 2013 Beijing

2013.5.22—26

OUTLINE

- (一) 小信号测量的区间估计
 - 1.1 小信号测量的特点
 - 1.2 区间估计的经典方法
 - 1.3 区间估计的 FC 方法
 - 1.4 系统误差的考虑
- (二) 信号的统计显著性
 - 2.1 实验 P 值
 - 2.2 信号的统计显著性

(一) 小信号测量的区间估计

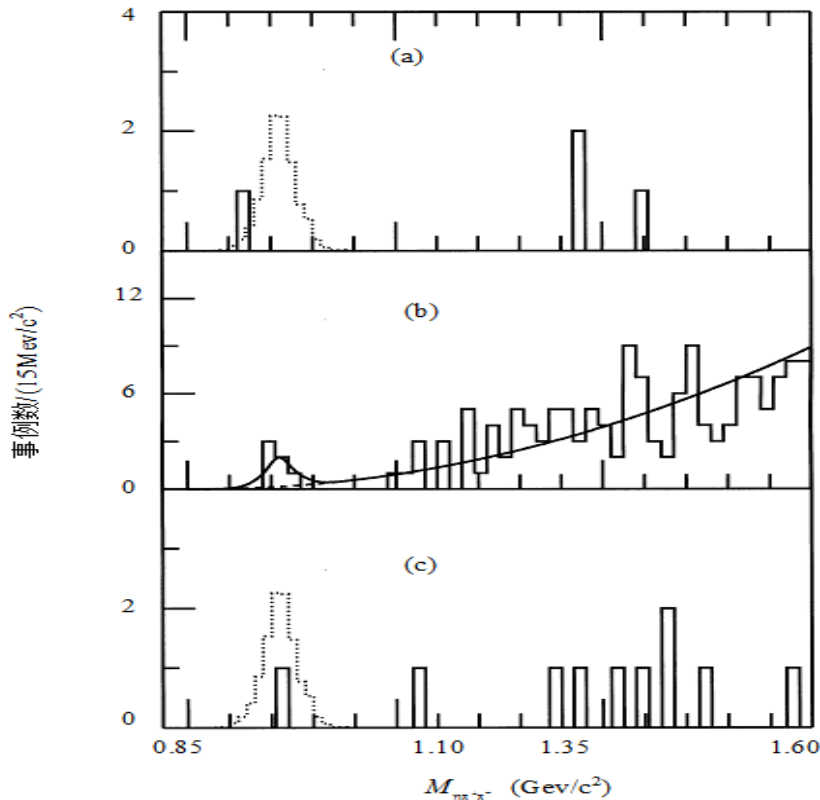
1.1 小信号测量的特点

》小信号——待测物理量数值很小（例如 m_ν ）。

待测现象（信号）出现概率很小（例如小分支比事例）。

测量 $ee \rightarrow \rho\eta'$ 反应截面： $\sigma = n_{sig} / (L\mathcal{E})$ 。

图 1.1 (a) $E_{cm} = 3650$ MeV, (b) 3686 MeV (c) 3773 MeV



- 实测值是随机子样（例 n_{sig} 服从 Poisson 分布, m_τ 服从正态分布）
- 信号事例数为小量， ≈ 0 （分支比上限）
- 信号的期望值存在物理边界（下界 0）
- 测量值 = 本底+信号， 皆有统计涨落和系统误差

》待解问题—根据一次测量值 x （例如 n_{sig} ）及其服从的分布（例 Poisson

分布）对于 x 的期望值 μ 进行统计推断。 $(Br(J/\psi \rightarrow f) = \frac{n_f / \epsilon_f}{N_{J/\psi}})$

1.2 区间估计的经典方法

Neyman 方法 (J.Phil.Trans. Royal Soc. London, Series, A236,333(1937))

从观测值 x 来确定期望值 μ 的一个区间 $\mu \in [\mu_1, \mu_2]$, 满足

$$P(\mu \in [\mu_1, \mu_2]) = \gamma, \quad (1.1)$$

γ 称为置信水平 (Confidence Level -- CL), 或涵盖 (coverage) 概率.

μ_1, μ_2 是 x 的函数. 在 $\mu-x$ 的标绘上, 对于一个确定的置信水平 γ , 满足式(1.1)的置信区间形成一个置信带(confidence belt).

图 1.2 置信水平 γ 的中心置信带

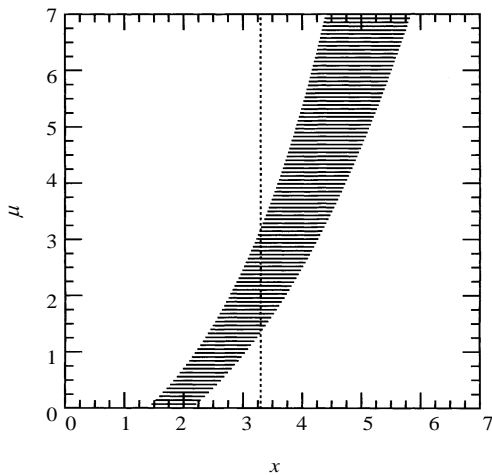
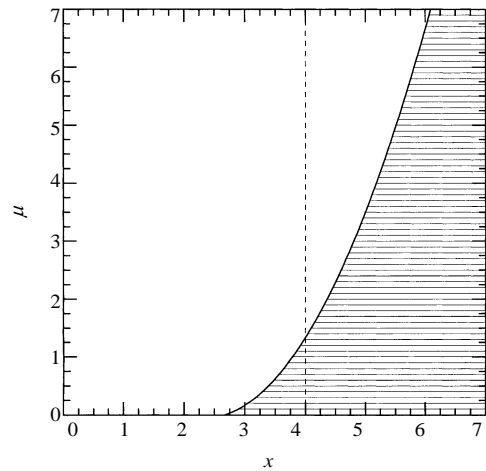


图 1.3 置信水平 γ 的上限置信带



》置信带的构造: 对任一特定的 μ 值, 找到 x 的接受区间 $[x_1, x_2]$ 满足关

系式
$$P(x \in [x_1, x_2] | \mu) = \gamma, \quad (1.2)$$

中心置信区间: $[x_1, x_2]$ 满足

$$P(x < x_1 | \mu) = P(x > x_2 | \mu) = \frac{1-\gamma}{2}; \quad (1.3)$$

上限置信区间 $[x_{\text{up}}, \infty]$ 定义为(图 1.3)

$$P(x > x_{\text{up}} | \mu) = \gamma. \quad (1.4)$$

所有可能的 μ 值对应的 x 接受区间 $[x_1, x_2]$ 的集合即为置信水平 γ 的置信带。

对于任一观测值 x ，这样确定的中心置信区间满足

$$P(\mu < \mu_1 | x) = P(\mu > \mu_2 | x) = \frac{1-\gamma}{2}, \quad (1.5)$$

而上限置信区间满足

$$P(x | \mu < \mu_{\text{up}}) = \gamma. \quad (1.6)$$

》有了置信水平 γ 的置信带，对任一特定的实验观测值 x_0 ，画一条平行于 μ 轴的直线 $x = x_0$ ，由它与置信带的交点求得 μ 的中心置信区间 $[\mu_1, \mu_2]$ 或上限值 μ_{up} 。

因此 μ 的区间估计问题实际上是置信带的构造问题。

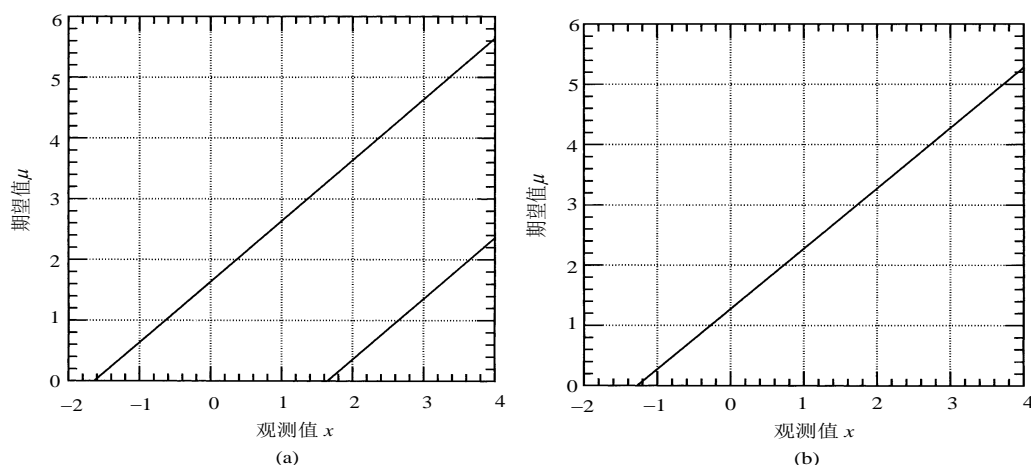
(1) 正态总体

实验观测值 x 服从正态分布，方差 σ^2 已知 (不失一般性，假定 $\sigma=1$)，其期望值是待估计的未知参数 μ ， μ 的物理下界为 0 (μ 为 ≥ 0 的正数)。

概率密度函数为

$$P(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2}\right], \quad (1.7)$$

据式 (1.3)、(1.4) 可画出 $\gamma=90\%$ 的中心置信带和上限置信带。



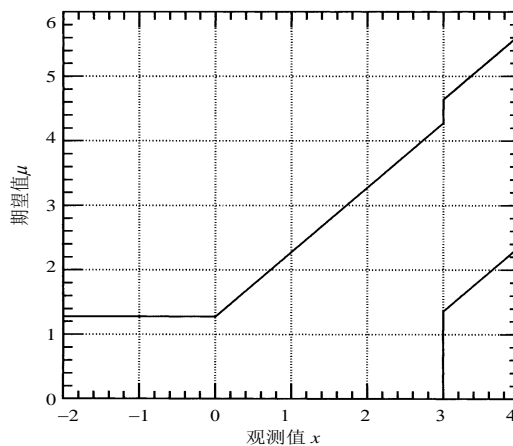
对于一个特定的实验测量值 x ，究竟是报道 μ 的中心区间 $[\mu_1, \mu_2]$ 还是上限区间 $[0, \mu_{\text{up}}]$ ，需要根据某种附加的要求来确定。

flip-flopping 方式: $x < 3\sigma$, 报道 90%上限区间;

$x \geq 3\sigma$ 报道中心区间.

由于 μ 需要 ≥ 0 , 当测定值 x 为负值时, 为了保险起见, 把 x 视为 0 并据此来确定其置信区间. 据此策略确定的置信带见图 1.5.

图 1.5 正态假设下 flip-flopping 策略相应的 $\gamma = 90\%$ 置信带



flip-flopping 置信带的两个缺陷:

- (a) 涵盖概率不足. 对于待估计参量的某些值, 其涵盖概率小于所规定的 γ 值. 例如 $\mu = 2.0$, 图 1.5 确定的接收区间为 $x_1 = 2 - 1.28$ 和 $x_2 = 2 + 1.64$, 该区间内的概率含量 $\int_{x_1}^{x_2} P(x | \mu = 2.0) dx = 0.85$, 没有达到规定的 $\gamma = 90\%$ 置信水平.
- (b) 空集问题. 例如, 当观测值 $x = -1.8$ (实验中是可能出现的), 图 1.4 中找不到相应的 $[\mu_1, \mu_2]$ 或 μ_{up} , 即 μ 的置信水平 90% 的置信区域是空集, 或者说期望值落在了物理不容许的区域.

(2) 泊松总体

观测值 x (现改写为观测总事例数 n), 服从期望值 $\mu+b$ 的泊松分布

$$P(n|\mu) = \frac{(\mu+b)^n e^{-(\mu+b)}}{n!}, \quad (1.8)$$

其中待估计的信号事例数服从泊松分布, 期望值为 μ ;

本底事例数服从期望值 b (已知值) 的泊松分布.

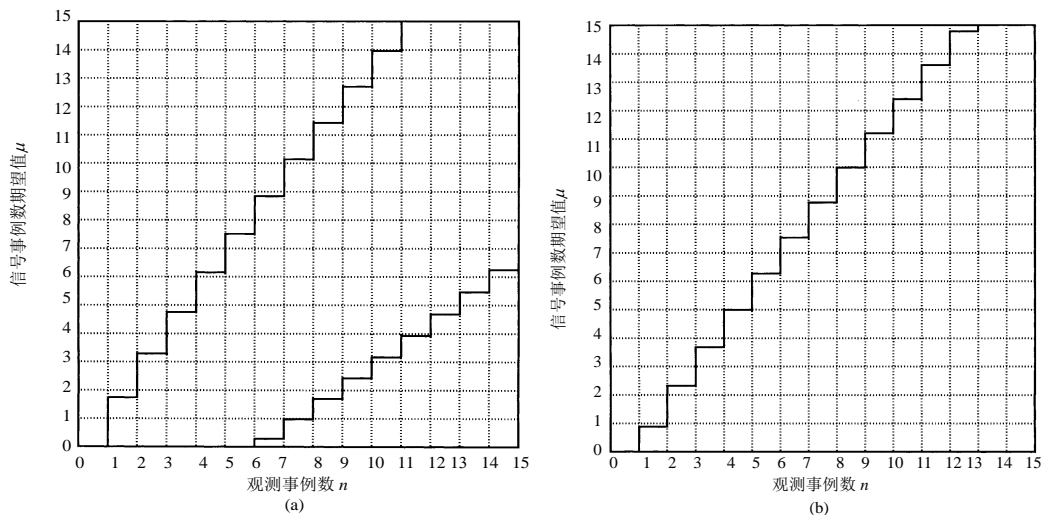
泊松 (离散) 分布, 中心置信带和上限置信带的构成要求是

$$P(\mu \in [\mu_1, \mu_2]) \geq \gamma, \quad P(\mu < \mu_{up}) \geq \gamma, \quad (1.9)$$

即要求实际涵盖概率略大于名义涵盖概率.

利用泊松分布的性质可计算出一定置信水平 γ , 一定 b 值的置信带.

图 1.6 泊松变量的 $\gamma \geq 90\%$ 的中心置信带(a)和上限置信带(b)(本底期望值 $b = 3.0$)



flip-flopping 方式: $x < 3b$, 90%上限区间; $x \geq 3b$ 报道中心区间.

同样出现 undercoverage 和空集问题.

结论: 对正态和泊松观测量, 经典方法存在涵盖概率不足和空集困难, 又不能在中心置信区间和上限置信区间之间作出合理的选择.

1.3 区间估计的 FC 方法

G. Feldman, R. Cousins: PRD 57 (1998) 3873

FC 方法

likelihood ratio ordering 似然比顺序求和方法

unified approach 统一方法

优点: 无空集和涵盖概率不足问题,

自动确定应该报道中心置信区间还是应报道上限。

(1) 泊松总体

按照式(1.8)泊松概率, 对任一给定的观测总事例数 n 和已知的平均本底 b , 使概率 $P(n|\mu)$ 达到极大的那个 μ 值定义为 μ_{best} , 即

$$P(n|\mu_{\text{best}}) > P(n|\mu), \quad \mu \neq \mu_{\text{best}}. \quad (1.10)$$

又根据物理边界要求 $\mu_{\text{best}} \geq 0$ (即 μ 必须 ≥ 0), 得 μ_{best} 的表达式

$$\mu_{\text{best}}(n, b) = \max(0, n - b). \quad (1.11)$$

定义似然比

$$R(\mu, n) \equiv \frac{P(n|\mu)}{P(n|\mu_{\text{best}})} = \left(\frac{\mu + b}{\mu_{\text{best}} + b} \right)^n e^{\mu_{\text{best}} - \mu}, \quad (1.12)$$

对任一特定的 μ 值, 其置信区间 $[n_1, n_2]$ 的求法:

- 用式(1.12)算出一切可能观测值 $n = 0, 1, 2, \dots$ 对应的似然比 $R(\mu, n)$ 值,
- 按 R 值从大到小的顺序决定每个观测值的秩 r (即 R 值最大的 n 值其秩 r 定义为 1, R 值次大的 n 值其 $r = 2$, 等等).

按 r 从小到大的顺序对观测值 n 的概率 $P(n|\mu)$ 求和, 直到满足

$$\sum_r P(n(r)|\mu) \geq \gamma, \quad (1.13)$$

$n(r)$ 的最小值 n_1 和最大值 n_2 构成该 μ 值的置信水平 γ 的置信区间。

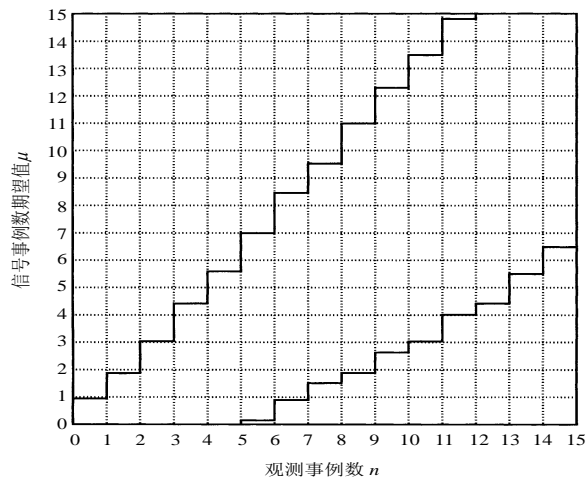
对一切 μ 值算出的 $[n_1, n_2]$, 构成置信水平 γ 的置信带。

》例：信号期望值 $\mu=0.5$ ，本底 $b=3$ 的泊松变量置信水平 $\gamma=90\%$

置信区间构造。 置信区间 $(n_1, n_2) = (0, 6)$.

n	$P(n \mu)$	μ_{best}	$P(n \mu_{best})$	R	Rank	置信区间
0	0.030	0.	0.050	0.607	6	y
1	0.106	0.	0.149	0.708	5	y
2	0.185	0.	0.224	0.826	3	y
3	0.216	0.	0.224	0.963	2	y
4	0.189	1.	0.195	0.966	1	y
5	0.132	2.	0.175	0.753	4	y
6	0.077	3.	0.161	0.480	7	y
7	0.039	4.	0.149	0.259		
8	0.017	5.	0.140	0.121		
9	0.007	6.	0.132	0.050		
10	0.002	7.	0.125	0.018		
11	0.001	8.	0.119	0.006		

图 1.7 平均本底 $b = 3.0$ ，置信水平 $\gamma = 90\%$ 泊松变量的置信带



》利用 FC 方法，编制程序计算了

$$\gamma = 0.6827, 0.90, 0.95, 0.99; \quad b = 0 \sim 15; \quad n = 0 \sim 20$$

情况下的信号事例期望值 μ 的置信区间，可查表。

置信区间上、下限的精度好于 0.01.

》泊松变量 FC 方法与经典方法的置信带比较，

对于任何 n ，FC 方法确定的置信区间的上、下限是唯一的，

不会出现空集和 undercoverage 的困难。

对于大的观测值 n ，FC 方法与经典方法结果相近。

对 n 比较小，与本底期望值 b 接近时，

FC 方法自动给出 μ 的上限，即 μ 的下限为 0。

TABLE IV. 90% C.L. intervals for the Poisson signal mean μ , for total events observed n_0 , for known mean background b ranging from 0 to 5.

$n_0 \setminus b$	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	5.0
0	0.00, 2.44	0.00, 1.94	0.00, 1.61	0.00, 1.33	0.00, 1.26	0.00, 1.18	0.00, 1.08	0.00, 1.06	0.00, 1.01	0.00, 0.98
1	0.11, 4.36	0.00, 3.86	0.00, 3.36	0.00, 2.91	0.00, 2.53	0.00, 2.19	0.00, 1.88	0.00, 1.59	0.00, 1.39	0.00, 1.22
2	0.53, 5.91	0.03, 5.41	0.00, 4.91	0.00, 4.41	0.00, 3.91	0.00, 3.45	0.00, 3.04	0.00, 2.67	0.00, 2.33	0.00, 1.73
3	1.10, 7.42	0.60, 6.92	0.10, 6.42	0.00, 5.92	0.00, 5.42	0.00, 4.92	0.00, 4.42	0.00, 3.95	0.00, 3.53	0.00, 2.78
4	1.47, 8.60	1.17, 8.10	0.74, 7.60	0.24, 7.10	0.00, 6.60	0.00, 6.10	0.00, 5.60	0.00, 5.10	0.00, 4.60	0.00, 3.60
5	1.84, 9.99	1.53, 9.49	1.25, 8.99	0.93, 8.49	0.43, 7.99	0.00, 7.49	0.00, 6.99	0.00, 6.49	0.00, 5.99	0.00, 4.99
6	2.21, 11.47	1.90, 10.97	1.61, 10.47	1.33, 9.97	1.08, 9.47	0.65, 8.97	0.15, 8.47	0.00, 7.97	0.00, 7.47	0.00, 6.47
7	3.56, 12.53	3.06, 12.03	2.56, 11.53	2.09, 11.03	1.59, 10.53	1.18, 10.03	0.89, 9.53	0.39, 9.03	0.00, 8.53	0.00, 7.53
8	3.96, 13.99	3.46, 13.49	2.96, 12.99	2.51, 12.49	2.14, 11.99	1.81, 11.49	1.51, 10.99	1.06, 10.49	0.66, 9.99	0.00, 8.99
9	4.36, 15.30	3.86, 14.80	3.36, 14.30	2.91, 13.80	2.53, 13.30	2.19, 12.80	1.88, 12.30	1.59, 11.80	1.39, 11.30	0.43, 10.30
10	5.50, 16.50	5.00, 16.00	4.50, 15.50	4.00, 15.00	3.50, 14.50	3.04, 14.00	2.63, 13.50	2.27, 13.00	1.94, 12.50	1.19, 11.50
11	5.91, 17.81	5.41, 17.31	4.91, 16.81	4.41, 16.31	3.91, 15.81	3.45, 15.31	3.04, 14.81	2.67, 14.31	2.33, 13.81	1.73, 12.81
12	7.01, 19.00	6.51, 18.50	6.01, 18.00	5.51, 17.50	5.01, 17.00	4.51, 16.50	4.01, 16.00	3.54, 15.50	3.12, 15.00	2.38, 14.00
13	7.42, 20.05	6.92, 19.55	6.42, 19.05	5.92, 18.55	5.42, 18.05	4.92, 17.55	4.42, 17.05	3.95, 16.55	3.53, 16.05	2.78, 15.05
14	8.50, 21.50	8.00, 21.00	7.50, 20.50	7.00, 20.00	6.50, 19.50	6.00, 19.00	5.50, 18.50	5.00, 18.00	4.50, 17.50	3.59, 16.50
15	9.48, 22.52	8.98, 22.02	8.48, 21.52	7.98, 21.02	7.48, 20.52	6.98, 20.02	6.48, 19.52	5.98, 19.02	5.48, 18.52	4.48, 17.52
16	9.99, 23.99	9.49, 23.49	8.99, 22.99	8.49, 22.49	7.99, 21.99	7.49, 21.49	6.99, 20.99	6.49, 20.49	5.99, 19.99	4.99, 18.99
17	11.04, 25.02	10.54, 24.52	10.04, 24.02	9.54, 23.52	9.04, 23.02	8.54, 22.52	8.04, 22.02	7.54, 21.52	7.04, 21.02	6.04, 20.02
18	11.47, 26.16	10.97, 25.66	10.47, 25.16	9.97, 24.66	9.47, 24.16	8.97, 23.66	8.47, 23.16	7.97, 22.66	7.47, 22.16	6.47, 21.16
19	12.51, 27.51	12.01, 27.01	11.51, 26.51	11.01, 26.01	10.51, 25.51	10.01, 25.01	9.51, 24.51	9.01, 24.01	8.51, 23.51	7.51, 22.51
20	13.55, 28.52	13.05, 28.02	12.55, 27.52	12.05, 27.02	11.55, 26.52	11.05, 26.02	10.55, 25.52	10.05, 25.02	9.55, 24.52	8.55, 23.52

(2) 正态总体

泊松变量的 FC 方法能以十分相似的方法应用于正态变量.

对于正态分布, 对任一给定的观测值 x , 使 $P(x|\mu)$ 达到极大的那个 μ 值定义为 μ_{best} , 即

$$P(x|\mu_{\text{best}}) = \max P(x|\mu); \quad (1.14)$$

根据物理边界要求 $\mu_{\text{best}} \geq 0$, 得 μ_{best} 的表达式

$$\mu_{\text{best}} = \max(0, x), \quad (1.15)$$

于是有条件概率

$$P(x|\mu_{\text{best}}) = \begin{cases} \frac{1}{\sqrt{2\pi}}, & x \geq 0, \\ \exp\left(-\frac{x^2}{2}\right) / \sqrt{2\pi}, & x < 0. \end{cases} \quad (1.16)$$

似然比 $R(x)$ 定义为

$$R(x) = \begin{cases} \frac{P(x|\mu)}{P(x|\mu_{\text{best}})} = \exp\frac{(x-\mu)^2}{2}, & \text{当 } x \geq 0, \\ \exp\left(x\mu - \frac{\mu^2}{2}\right), & \text{当 } x < 0. \end{cases} \quad (1.17)$$

对任一给定的 μ 值, 置信水平 γ 的置信区间 $[x_1, x_2]$ 由下式求出:

$$\int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2}\right] dx = \gamma, \quad R(x_1) = R(x_2). \quad (1.18)$$

一切可能的 μ 值对应的置信区间 $[x_1, x_2]$, 构成置信水平 γ 的置信带.

利用数值方法求解式(1.18)。

》对于观测值 $x \in (-3, 3)$ 的情形, $\gamma = 68.27\%$, 90% , 95% , 99%

的置信区间 $[\mu_1, \mu_2]$ 的数值给出了列表.

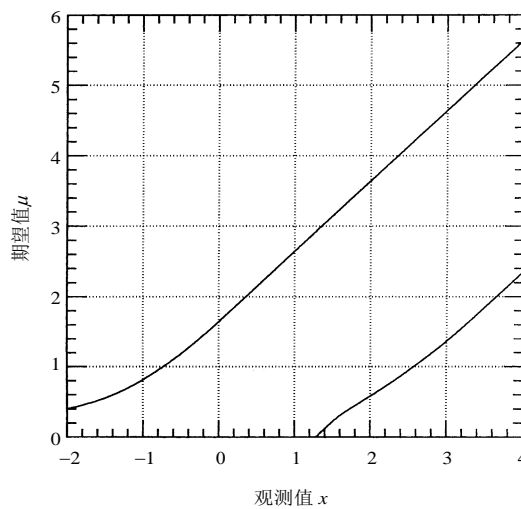
$\mu_1 = 0$ 相当于上限置信区间.

表 FC 方法求得的正态总体期望值 μ 置信水平 68.27%, 90%, 95%, 99% 的置信区间 $[\mu_1, \mu_2]$.

x_0 是总体的实验观测值. 表中所有数字均以正态总体标准偏差 σ 为单位

x_0	68.27% C.L.	90% C.L.	95% C.L.	99% C.L.
-3.0	0.00 0.04	0.00 0.26	0.00 0.42	0.00 0.80
-2.0	0.00 0.07	0.00 0.40	0.00 0.62	0.00 1.10
-1.0	0.00 0.27	0.00 0.81	0.00 1.10	0.00 1.68
-0.0	0.00 1.00	0.00 1.64	0.00 1.96	0.00 2.58
1.0	0.24 2.00	0.00 2.64	0.00 2.96	0.00 3.58
2.0	1.00 3.00	0.58 3.64	0.35 3.96	0.00 4.58
3.0	2.00 4.00	1.37 4.64	1.14 4.96	0.67 5.58

图 1.8: 正态变量期望值 μ 的置信水平 $\gamma = 0.90$ 的置信带.



测量值 $x \leq 1.28$ 时, μ 置信区间下界为 0, 应报道 90% 置信水平上限;

$x > 1.28$ 时, 应报道 $\mu_{-\sigma_e}^{+\sigma_u}$ 的实验结果, σ_u, σ_e 是相应的正、负误差.

》正态变量 FC 方法与经典方法的置信带比较:

x 大的区域, 两者的置信区间 $[\mu_1, \mu_2]$ 相近;

$x \leq 0$ 和 $x \approx 0$ 的区域两者有明显的差别.

FC 方法确定的置信区间的上、下限是唯一的,

上限和中心置信区是自然形成的, 无空集和 undercoverage 困难.

1.4 系统误差的考虑

J.Conrad: PRD67, (2003) 012002

》前面讨论中，对于泊松总体， $P(n|\mu) \equiv P(n)_{\mu+b} = \frac{(\mu+b)^n e^{-(\mu+b)}}{n!}$

假定了信号区内本底事例数期望值 b 为已知常数。

》实际上， b 具有不确定性（ b 存在系统误差，应考虑为随机量）。

对信号事例数期望值 μ 作区间估计时，其泊松概率应修改为

$$q(n)_{\mu+b} = \frac{1}{\sqrt{2\pi}\sigma_b} \int_0^\infty p(n)_{\mu+b'} e^{-(b-b')^2/2\sigma_b^2} db', \quad (1.19)$$

其中 σ_b 是 b 的标准差。

》反应截面由 $\sigma = n_{\text{sig}}/L\varepsilon$ 确定，探测效率 ε 的确定也会有系统误差，

泊松概率应进一步修改为

$$q(n)_{\mu+b} = \frac{1}{\sqrt{2\pi}\sigma_b\sigma_\varepsilon} \int_0^\infty \int_0^\infty p(n)_{\mu\varepsilon'+b'} e^{-(b-b')^2/2\sigma_b^2} e^{-(1-\varepsilon')^2/2\sigma_\varepsilon^2} db' d\varepsilon', \quad (1.20)$$

其中 σ_ε 是信号事例探测效率 ε 的(相对)系统误差。

》利用式(1.19)、(1.20)的概率密度函数，再按照 FC 方法，即可求得信号事例数期望值的置信区间。

》在式(1.19)、(1.20)的概率密度函数表式中，系统误差的分布被假定为正态分布。原则上，系统误差也可考虑为其他分布。

》按照以上原则，J.Conrad 等编制了计算机程序包 POLE

(参见 <http://www3.tsl.uu.se/~conrad/pole.html>)，可计算 $n \leq 100$ ，

$\mu \leq 50$ 情形下，经典方法、FC 方法或改进的 FC 方法构造的置信带，

系统误差的分布可以是正态分布、对数-正态分布或均匀分布。

(二) 信号的统计显著性

》信号的统计显著性 (statistical significance) 是物理学家对“观测到了信号事例”这一判断的定量表征。

》信号统计显著性问题时，零假设 H_0 ：观测量只用本底函数描述，备择假设 H_1 ：观测量用本底函数+信号函数描述。

》信号的统计显著性：

观察到的现象偏离本底函数、发现新信号的定量表征。

信号的统计显著性越高，发现新信号的可信度越大。

共识：信号显著性 $s \geq 5$ ，认为“发现”了一种新信号；

$s \geq 3$ 信号的存在有强的证据。

$s \geq 2$ ， 信号的存在有弱的证据。

2.1 实验 P 值

设 u 是观测量(随机变量)或用它构造的统计量，

u_{obs} 是某个实验测量到的 u 值，

实验 P 值定义： H_0 为真时 $u \geq u_{\text{obs}}$ 的概率。

(不失一般性，设定 u 值越大， H_0 为真的可能性越小)。

$$P(u_{\text{obs}}) = P(u \geq u_{\text{obs}} | H_0) = \begin{cases} \sum_{u \geq u_{\text{obs}}} p_u(H_0), & u \text{ 离散型} \\ \int_{u \geq u_{\text{obs}}} f(u | H_0) du, & u \text{ 连续型} \end{cases} \quad (2.1)$$

$P(u_{\text{obs}})$ 值是该实验测量值 u_{obs} 与零假设 H_0 不一致性的定量表征，

$P(u_{\text{obs}})$ 值越小， H_0 为真的可能性越小 (有信号的可能性大)。

》 观测量为泊松变量时的实验 P 值

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。

信号区内信号和本底事例数可视为期望值 s 和 b 的泊松变量，信号区内观测到总事例数 n 的概率为

$$f(n; s, b) = \frac{(s+b)^n}{n!} e^{-(s+b)}. \quad (2.2)$$

n 即为式 (2.1) 中的统计量 u 。

若 n 的观测值为 n_{obs} ，则实验 P 值为

$$P(n_{\text{obs}}) = P(n \geq n_{\text{obs}} | H_0) = \sum_{n=n_{\text{obs}}}^{\infty} f(n; s=0, b) = 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b}. \quad (2.3)$$

P 值越小， H_0 为真(观测到的事例仅由于本底的贡献)的可能性越小。

举例：假定 $b=0.6, n_{\text{obs}}=5$ ，实验 P 值 $\cong 4 \times 10^{-4}$ ，

这表示实验观测到 $n \geq n_{\text{obs}}=5$ 个事例仅仅是由于本底的贡献
(即仅由本底涨落导致) 的概率小到只有 4×10^{-4} 。

(即很可能存在新的信号事例。)

》系统误差的考虑

信号区宽度的不确定性

不同的信号区宽度给出不同的 n_{obs} 和 b 值，以及不同的 P 值。

信号区的宽度应足以包含绝大部分信号事例。

例，信号为正态分布，信号区取为 $x_c \pm 3\sigma$ ，包含 99.7% 信号事例。

本底事例数期望值 b 可能存在系统误差。

b 由信号区附近的边带区 (sideband) 的事例数分布确定

(隐含假设：边带区内不含信号事例，且很好地反映信号区内的本底贡献。)

b 存在系统误差的原因

边带区内的事例数分布不平滑

边带区的宽度没有严格的规则，存在不确定性。

利用不同的本底函数形式、不同的边带区宽度，确定 b 的值不同。

报导保守的结果

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 0.5~0.7，则相应的实验 P 值范围为

$2 \times 10^{-4} \sim 8 \times 10^{-4}$ 。

最后结果：报道 P 值的这一范围，或报道 P 值为 8×10^{-4} 。

2.2 信号的统计显著性

》实验 P 值—实验观测到的现象仅仅是由于本底贡献的概率，

（但实验观测量与待寻找的信号间的关系不直观）

》观测量与信号间的关系通常用信号的统计显著性 s 表示。

》数据分析中，使用了不同的统计显著性 s 的定义。

(1) “计数实验”中统计显著性 s 的定义

文献 S.I.Bityukov et al, NIMA452, 2000, 518

I.Narsky, NIMA450,2000,444

信号统计显著性 s 有如下各种定义：（角标用来区分）

（以下定义都基于信号区内的计数，称为“计数实验 counting experiment”）。

$$S_1 = \frac{n-b}{\sqrt{b}} = \frac{s}{\sqrt{b}}, \quad (2.4)$$

$$S_2 = \frac{n-b}{\sqrt{n}} = \frac{s}{\sqrt{s+b}}, \quad (2.5)$$

$$S_{12} = \sqrt{n} - \sqrt{b}, \quad (2.6)$$

$$S_{B1} = S_1 - k(\alpha)\sqrt{n/b}, \quad (2.7)$$

$$S_{B12} = 2S_{12} - k(\alpha), \quad (2.8)$$

$$\int_{-\infty}^{S_N} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \sum_{i=0}^{n-1} \frac{b^i}{i!} e^{-b}, \quad (2.9)$$

其中 n 是信号区内观测到的总事例数，

b 是信号区内本底事例数的（已知）期望值。

$k(\alpha)$ 是一个与观察到信号事例的确定程度相关的系数。

S.I.Bityukov et al, Proceedings of PHYSTAT05, 12-15, Sep.2005, p.106:

S_1 只考虑本底的涨落, 本底 $\sim N(b,b)$

S_2 考虑正态变量 $n=s+b$ 的涨落, b 是本底期望值

S_{12} 考虑信号 $\sim N(s,s)$ 和本底 $\sim N(b,b)$ 的涨落

S_N 考虑观测值为泊松变量 $P(n)$, 本底为泊松变量 $P(b)$ 的涨落

(2) 用似然函数定义信号的统计显著性

似然比方法: 在许多情形下, 用**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 比用计数实验计算 S 要精确。

* 用似然比方法进行检验时, 拟合的范围不只是“信号区”, 而应该更宽。(与计数法不同)

* $L(b)$ 、 $L(s+b)$ 为用原假设 H_0 (本底) 和备择假设 H_1 (本底+信号) 的 pdf 构造的似然函数,

$$L(b) = \prod_{i=1}^N f_b(u_i), \quad (2.10)$$

$$L(s+b) = \prod_{i=1}^N [\omega_s f_s(u_i) + (1-\omega_s) f_b(u_i)], \quad (2.11)$$

拟合范围内的 (总、信号、本底) 事例数为 N , s , b ($N=s+b$)
 ω_s 为信号事例数的权因子 .

* $L_m(b)$ 和 $L_m(s+b)$ 为似然函数极大值,

* I.Narsky 在 $-2\ln[L_m(b)/L_m(s+b)]$ 服从 $\chi^2(1)$ 分布的假定下给出了信号的统计显著性

$$S = [2(\ln L_m(s+b) - \ln L_m(b))]^{1/2}. \quad (2.12)$$

(3) 我对信号显著性的处理

朱永生：高能物理与核物理 30（2006）331

》信号显著性的定义

实验 P 值越小，备择假设为真（存在信号）的可能性越大。

因此，信号显著性 S_Z 定义为：

$$\int_{-S_Z}^{S_Z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - P(u \geq u_{\text{obs}} | H_0) \equiv 1 - P(u_{\text{obs}}). \quad (2.13)$$

该式的左边是正态分布在 $\pm S_Z$ 个标准差 ($\pm S_Z \sigma$) 内的积分概率。

此定义下，相对应的 S_Z 值和实验 P 值如下表。

S_Z 值	1	2	3	4	5	6
实验 P 值	0.3173	0.0455	0.0027	6.3×10^{-5}	5.7×10^{-7}	2.0×10^{-9}

》计数实验的实验 P 值和信号显著性

$$P(n_{\text{obs}}) = P(n \geq n_{\text{obs}} | H_0) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{b^n}{n!} e^{-b} = 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b}. \quad (2.14)$$

代入式(2.13)，立即有

$$\int_{-S_Z}^{S_Z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b}. \quad (2.15)$$

与 Narsky 的定义式(2.9)比较，积分下限不同：

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

》似然比方法的实验 P 值和信号显著性

H_0 的本底似然函数 $L(b)$ 有 m 个待定参数 ($f_b(u)$ 的参数),

H_1 备择假设似然函数 $L(s+b)$ 有 $k(>m)$ 个待定参数

($f_b(u)$ 和 $f_s(u)$ 的参数及两者的相对权重).

r 是 H_1 与 H_0 的待估计独立参数个数之差, 或自由度之差

$$r = k - m$$

可以证明, 当 H_0 为真, 在子样容量很大的情形下, 统计量

$$u = -2 \ln \lambda = 2(\ln L_m(s+b) - \ln L_m(b)) \quad (2.16)$$

$$\lambda = L_m(b)/L_m(s+b) \quad (2.17)$$

渐近地服从 $\chi^2(r)$ 分布.

当 u 的实验观测值为 u_{obs} , 则由式(2.1)知实验 P 值为

$$P(u_{\text{obs}}) = \int_{u_{\text{obs}}}^{\infty} \chi^2(u; r) du, \quad (2.18)$$

代入式(2.13)得到信号统计显著性 S_Z 的表达式

$$\int_{-S_Z}^{S_Z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - P(u_{\text{obs}}) = \int_0^{u_{\text{obs}}} \chi^2(u; r) du. \quad (2.19)$$

对于 $r=1$ 的特殊情形, 有

$$S_Z = \sqrt{u_{\text{obs}}} = [2(\ln L_m(s+b) - \ln L_m(b))]^{1/2}, \quad (2.20)$$

与 I. Narsky 的式(2.12)完全一致.

由以上讨论可知, 式(2.13)定义的信号统计显著性

$$\int_{-S_Z}^{S_Z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - P(u \geq u_{\text{obs}} | H_0) \equiv 1 - P(u_{\text{obs}}). \quad (2.13)$$

对计数实验或似然比方法都适用, 避免了多重定义.

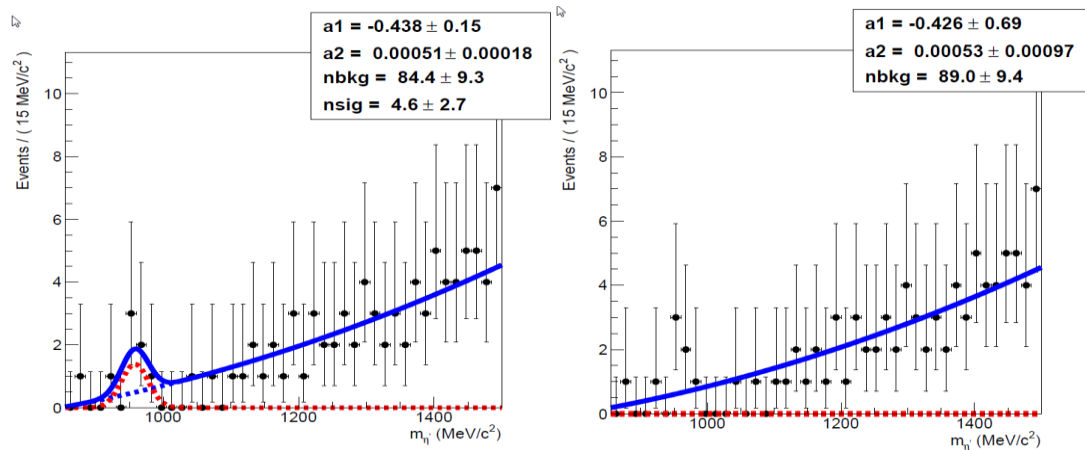
示例

数据

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

表中“区间”中的数字表示下限 (MeV)，上限是该数字加 15MeV。

“事例”中无数字表示事例数为 0。



本底的 pdf: 多项式函数 $f_b(m) = c_1 + c_2 m + c_3 m^2$

信号 pdf: 正态函数

$$f_s(m) = N(m_0, \sigma^2), \quad m_0 = 958 \text{ MeV}, \quad \sigma = 20 \text{ MeV}.$$

解:

》拟合区间 $[840, 1500] \text{ MeV} = [a, b]$ 。

》极大似然法应用于直方图数据,

第 i 区间出现事例数为 $n_i (i = 1, 2, \dots, N)$ 的联合概率(似然函数)由多项分布给出:

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

p_i 是第 i 个子区间中出现一个事例的概率. 由概率密度 $f(x|\mathcal{G})$ 在

第 i 子区间中的积分求出:
$$p_i = p_i(\mathcal{G}) = \int_{\Delta m_i} f(m|\mathcal{G}) dm.$$

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

》对于本底函数
$$p_i = p_i(\mathcal{G}) = \int_{\Delta m_i} f_b(m|\mathcal{G}) dm.$$

$$\mathcal{G} = c_1, c_2, c_3, \quad \text{求极大值 } \ln L_m(b).$$

》对于信号+本底函数
$$p_i = p_i(\mathcal{G}) = \int_{\Delta m_i} [w_s f_s + (1-w_s) f_b] dm.$$

$$\mathcal{G} = c_1, c_2, c_3, w_s, \quad \text{求极大值 } \ln L_m(s+b).$$

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

$$u = -2 \ln \lambda = 2(\ln L_m(s+b) - \ln L_m(b)) \quad (2.16)$$

渐近地服从 $\chi^2(r=4-3=1)$ 分布. 按式 (2.20) 计算统计显著性 S_Z :

$$S_Z = \sqrt{u_{\text{obs}}} = [2(\ln L_m(s+b) - \ln L_m(b))]^{1/2}, \quad (2.20)$$

》 计算结果

对于本底极大值 $\ln L_m(b)$,

$$c_1 = 6.14 \times 10^{-6}, \quad c_2 = 2.62 \times 10^{-6}, \quad c_3 = 3.25 \times 10^{-9}.$$

对于信号+本底函数极大值 $\ln L_m(s+b)$

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

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

作为比较，利用计数法，假定信号区间为 [900, 1020] MeV,

(中心值 $\pm 3\sigma$)

观测总事例数为 $n = 7$, 本底期望值为 $b = 3.8$,

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

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

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

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

(4) 报道信号统计显著性的注意要点

由于 S 值有多种定义，因此实验结果的报道中，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。

(结束)