



清华大学
Tsinghua University



粒子物理与核物理实验中的 数据分析

第九章：置信区间

王喆
清华大学



本章要点

- 统计不确定度中的标准差问题
- 经典置信区间问题
- 利用似然函数或二乘函数确定置信区间
- 多维置信区间
- 物理边界

再论统计分析的目标

假设检验



检验数据是否与某特定理论相符(理论可包含自由参数)。



相符的程度由显著性水平来表示。

参数估计



利用数据确定自由参数的值。



参数的准确程度由对应的不确定度表示。



如何定量估计显著性水平与确定不确定度的大小？

测量结果的表述与含义

实验数据: x_1, \dots, x_n  实验目的: 估计参数 $\theta \rightarrow \hat{\theta}_{\text{obs}}$

还应该给出 $\hat{\theta}$ 的方差, 即 $\hat{\sigma}_{\hat{\theta}}^2$ 。结果应该报告成如下形式

$$\hat{\theta}_{\text{obs}} \pm \hat{\sigma}_{\hat{\theta}} = 5.73 \pm 0.21$$

其真正的含义是什么呢?

如果我们知道 $\hat{\theta}_{\text{obs}}$ 服从某个概率密度 $g(\hat{\theta}; \theta)$, 那么上述结果的正确表述应该是

θ 的估计值为 5.73

$\sigma_{\hat{\theta}}$ 的估计值为 0.21



$\sigma_{\hat{\theta}}$ 描述 $g(\hat{\theta}; \theta)$ 的分布宽度

参数估计值的分布

通常参数估计值服从的概率密度 $g(\hat{\theta}; \theta)$ 是多维高斯分布


$\hat{\theta}$ 和 $\hat{V} = \widehat{\text{cov}}[\hat{\theta}_i, \hat{\theta}_j]$ 综合了我们对 $g(\hat{\theta}; \vec{\theta})$ 的了解或估计



可以用来作为不确定度传递的输入参量，
以及用最小二乘法求平均值等等。

我们可依约定报告不确定度，而不管概率密度 $g(\hat{\theta}; \theta)$ 的形式。
只有当需要对不同实验求平均值时，其的形式就会发挥作用。

如果 $g(\hat{\theta}; \theta)$ 是高斯形式，置信区间可以表述为

$[\hat{\theta}_{\text{obs}} - \hat{\sigma}_{\hat{\theta}}, \hat{\theta}_{\text{obs}} + \hat{\sigma}_{\hat{\theta}}]$  68.3% 置信区间范围。

如果 $g(\hat{\theta}; \theta)$ 不是高斯形式



中心置信区间应给出不对称的不确定度

经典置信区间的定义

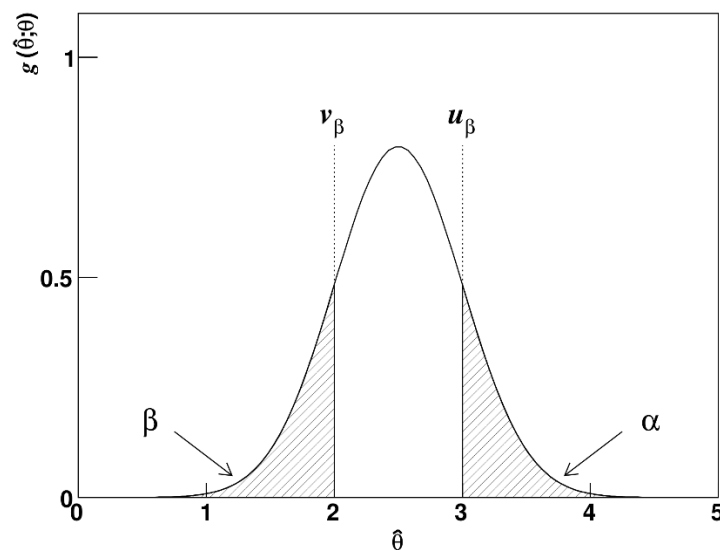
假设参数 θ 的估计量为 $\hat{\theta}$ ，并且估计值为 $\hat{\theta}_{\text{obs}}$ 。

为了正确表述结果，对于所有 θ 仍需要知道 $g(\hat{\theta}; \theta)$ 的形式
首先需要指定“上下分布尾部的概率”，如： $\alpha = \beta = 0.05$ ，
然后找出 $u_{\alpha}(\theta)$ ， $v_{\beta}(\theta)$ ，使得

$$\alpha = P(\hat{\theta} \geq u_{\alpha}(\theta)) = \int_{u_{\alpha}(\theta)}^{\infty} g(\hat{\theta}; \theta) d\hat{\theta} \\ = 1 - G(u_{\alpha}(\theta); \theta)$$

$$\beta = P(\hat{\theta} \leq v_{\beta}(\theta)) = \int_{-\infty}^{v_{\beta}(\theta)} g(\hat{\theta}; \theta) d\hat{\theta} \\ = G(v_{\beta}(\theta); \theta)$$

$G(\hat{\theta}; \theta)$ 为概率密度 $g(\hat{\theta}; \theta)$ 的
累积分布



参数置信带的定义

在 $u_\alpha(\theta)$ 和 $v_\beta(\theta)$ 之间的区域称为**置信带**。

无论 θ 为何值，在置信带找到 $\hat{\theta}$ 的概率为

$$P\left(v_\beta(\theta) \leq \hat{\theta} \leq u_\alpha(\theta)\right) = 1 - \alpha - \beta$$

假设 $u_\alpha(\theta)$ 和 $v_\beta(\theta)$ 单调，则有反函数

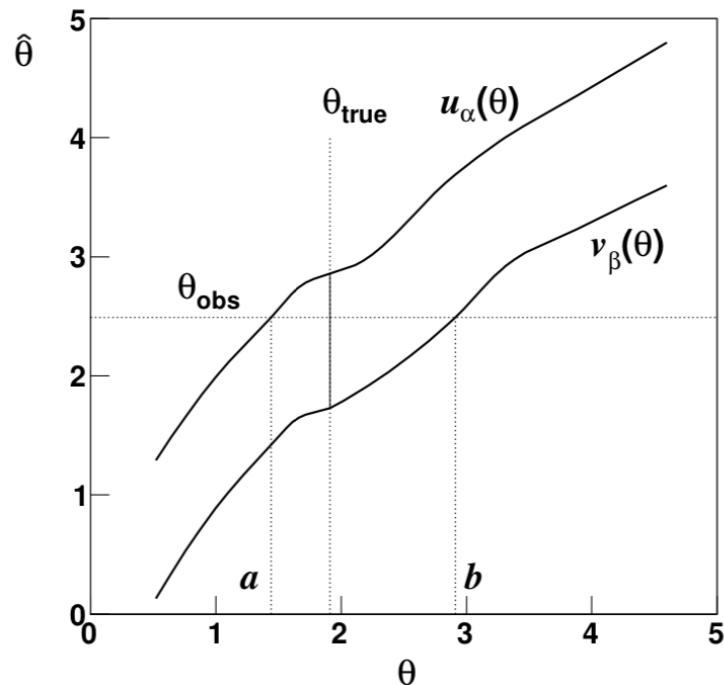
$$\begin{aligned} a(\hat{\theta}) &\equiv u_\alpha^{-1}(\hat{\theta}), \\ b(\hat{\theta}) &\equiv v_\beta^{-1}(\hat{\theta}) \end{aligned}$$

不等式

$$\begin{aligned} \hat{\theta} &\geq u_\alpha(\theta), \\ \hat{\theta} &\leq v_\beta(\theta), \end{aligned}$$

等价于

$$\begin{aligned} a(\hat{\theta}) &\geq \theta, \\ b(\hat{\theta}) &\leq \theta \end{aligned}$$



参数的置信区间确定

根据置信带的定义，有不等式

$$a(\hat{\theta}) \geq \theta, \quad b(\hat{\theta}) \leq \theta$$

$$P(a(\hat{\theta}) \geq \theta) = \alpha, \quad P(b(\hat{\theta}) \leq \theta) \leq \beta$$



或者合并成

$$P(a(\hat{\theta}) \leq \theta \leq b(\hat{\theta})) = 1 - \alpha - \beta$$

在不知道真值 θ 的情况下，通过估计值 $\hat{\theta}$ 与函数 $a(\hat{\theta})$ 和 $b(\hat{\theta})$ 给出 θ 的置信区间。

参数置信区间含义

区间 $[a(\hat{\theta}), b(\hat{\theta})]$ 成为具有置信水平或覆盖概率 $1 - \alpha - \beta$ 的置信区间。

它的深刻含义是 包含真实参数的概率为 $1 - \alpha - \beta$

注意，该区间是随机的，真值 θ 是一个未知常数。

通常情况下，将区间 $[a, b]$ 报道为 $\hat{\theta} \pm_c^d$ ，即

$$c = \hat{\theta} - a, \quad d = b - \hat{\theta}$$

那么 $\hat{\theta} = 80.25 \pm_{0.25}^{0.31}$ 意味着什么呢？

它并不意味着任意一次实验：

$$P(80.00 < \theta < 80.56) = 1 - \alpha - \beta$$

它意味着：重复相同样本容量的实验多次，每次按相同描述构造置信区间，有 $1 - \alpha - \beta$ 比率的实验，置信区间将覆盖 θ 。

单边与中心置信区间

有时，单独指定 α 或 β

→ 单边区间(极限, 上限或下限)

通常，取 $\alpha = \beta = \gamma/2$

→ 覆盖概率为 $1 - \gamma$

→ 中心置信区间

注意：中心置信区间并不意味着区间对于 $\hat{\theta}$ 是对称的，它仅仅是因为 $\alpha = \beta$ 。

粒子与核物理的误差惯例是：68.3%的中心置信区间。

经典置信区间

可以构造置信带:

$$\alpha = \int_{u_\alpha(\theta)}^{\infty} g(\hat{\theta}; \theta) d\hat{\theta} = 1 - G(u_\alpha(\theta); \theta) \Rightarrow u_\alpha(\theta)$$
$$\beta = \int_{-\infty}^{v_\beta(\theta)} g(\hat{\theta}; \theta) d\hat{\theta} = G(v_\beta(\theta); \theta) \Rightarrow v_\beta(\theta)$$

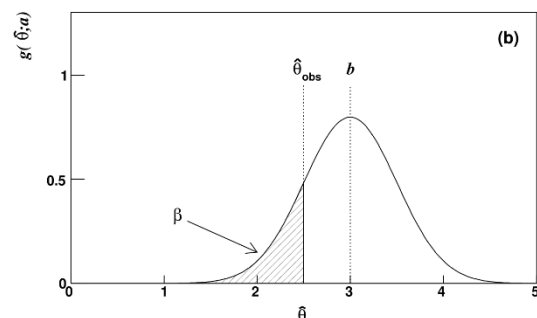
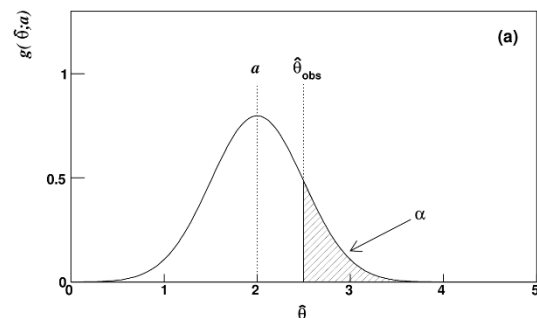
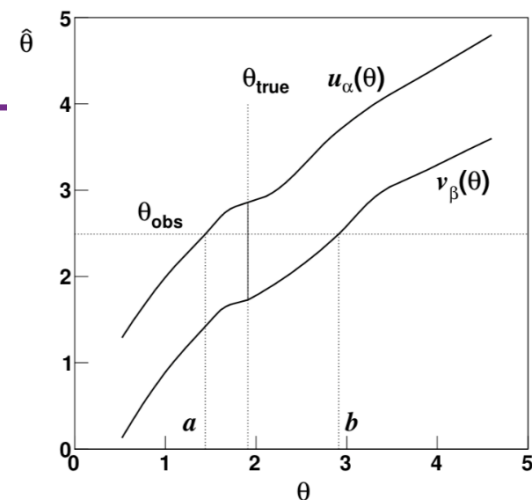
→ $a \equiv u_\alpha^{-1}(\hat{\theta}_{\text{obs}}),$ $b \equiv v_\beta^{-1}(\hat{\theta}_{\text{obs}})$ 置信区间 $[a, b]$

通常, 我们不构造置信带, 而是解方程得到 a 与 b 的区间极限:

$$\alpha = P(\hat{\theta} \geq \hat{\theta}_{\text{obs}}; a) = \int_{\hat{\theta}_{\text{obs}}}^{\infty} g(\hat{\theta}; a) d\hat{\theta} = 1 - G(\hat{\theta}_{\text{obs}}; a)$$
$$\beta = P(\hat{\theta} \leq \hat{\theta}_{\text{obs}}; b) = \int_{-\infty}^{\hat{\theta}_{\text{obs}}} g(\hat{\theta}; b) d\hat{\theta} = G(\hat{\theta}_{\text{obs}}; b)$$

a 是 θ 的假设值, 使得: $P(\hat{\theta} > \hat{\theta}_{\text{obs}}) = \alpha$

b 是 θ 的假设值, 使得: $P(\hat{\theta} < \hat{\theta}_{\text{obs}}) = \beta$



高斯分布估计量的置信区间

如果 $\hat{\theta}$ 服从

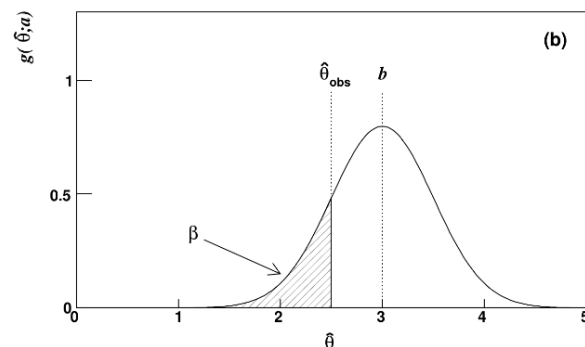
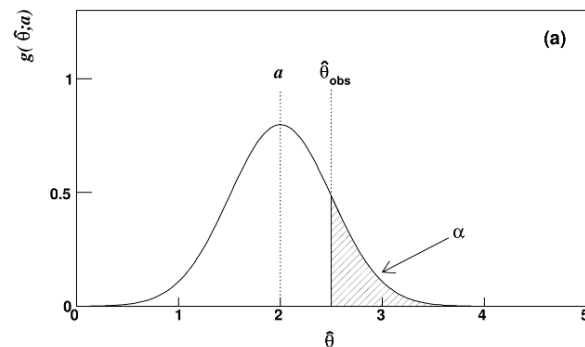
$$g(\hat{\theta}; \theta) = \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp\left(-\frac{(\hat{\theta} - \theta)^2}{2\sigma_{\hat{\theta}}^2}\right)$$

为了找到 θ 的置信区间，解下列方程

$$\alpha = 1 - G(\hat{\theta}_{\text{obs}}; a, \sigma_{\hat{\theta}}) = 1 - \Phi\left(\frac{\hat{\theta}_{\text{obs}} - a}{\sigma_{\hat{\theta}}}\right),$$
$$\beta = G(\hat{\theta}_{\text{obs}}; b, \sigma_{\hat{\theta}}) = \Phi\left(\frac{\hat{\theta}_{\text{obs}} - b}{\sigma_{\hat{\theta}}}\right)$$



a 与 b 的解



高斯分布的累积函数与分位点

前面的函数 G 是 $\hat{\theta}$ 的累积分布，且

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x'^2/2} dx'$$

是标准高斯的累积分布函数，可以证明

$$\begin{aligned} a &= \hat{\theta}_{\text{obs}} - \sigma_{\hat{\theta}} \Phi^{-1}(1 - \alpha), \\ b &= \hat{\theta}_{\text{obs}} + \sigma_{\hat{\theta}} \Phi^{-1}(1 - \beta) \end{aligned}$$

这里 Φ^{-1} 给出标准高斯的分位数（可以调用ROOT中的函数
`Double_t TMath::NormQuantile(Double_t p)`计算）

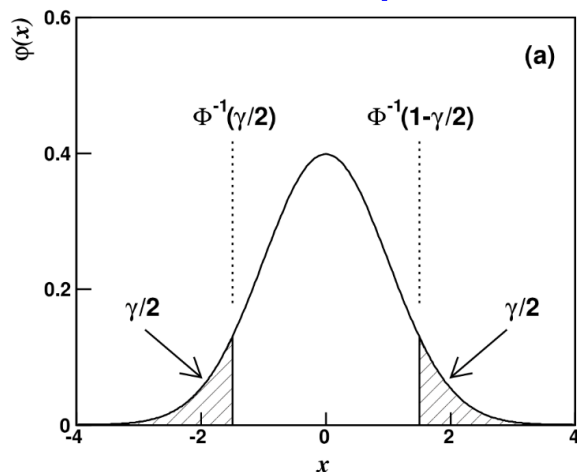


$$\Phi^{-1}(1 - \alpha), \quad \Phi^{-1}(1 - \beta)$$

给出 a 与 b 离 $\hat{\theta}$ 有多少标准差。

标准高斯的分位点

为了找到服从高斯分布的参数估计量的置信区间，需要确定下列分位点（取 $\alpha = \beta = \gamma/2$ ）

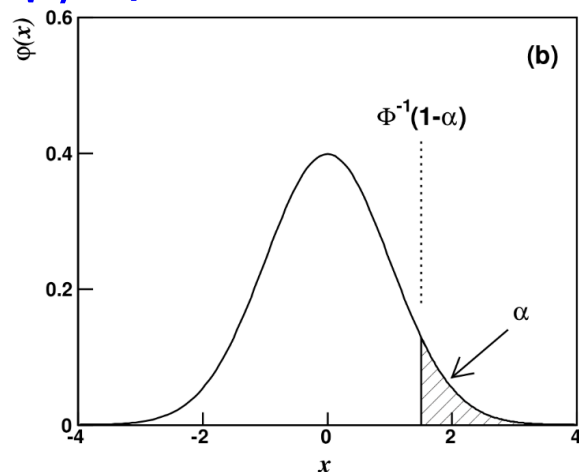


通常对分位点取整

中心

单边

$\Phi^{-1}\left(1 - \frac{\gamma}{2}\right)$	$1 - \gamma$	$\Phi^{-1}(1 - \alpha)$	$1 - \alpha$
1	0.6827	1	0.8413
2	0.9544	2	0.9772
3	0.9973	3	0.9987



有时对概率覆盖率取整

中心

单边

$1 - \gamma$	$\Phi^{-1}\left(1 - \frac{\gamma}{2}\right)$	$1 - \alpha$	$\Phi^{-1}(1 - \alpha)$
0.90	1.645	0.90	1.282
0.95	1.960	0.95	1.645
0.99	2.576	0.99	2.326

泊松分布均值的置信区间

假设 n 是泊松变量, $\hat{v} = n$, 估计值 $\hat{v}_{\text{obs}} = n_{\text{obs}}$,

$$P(n; v) = \frac{v^n}{n!} e^{-v}, \quad n = 0, 1, \dots$$

对于固定的 α, β , 由于 \hat{v} 只能取离散值, 用来确定置信带的函数 $u_\alpha(v)$ 和 $v_\beta(v)$ 不一定对任意 v 都存在, 即不一定能找到整数 \hat{v} , 满足

$P(\hat{v} \geq u_\alpha(v)) = \alpha, P(\hat{v} \leq v_\beta(v)) = \beta$ 。此时, 可考虑把方程变为

$$\begin{aligned} \alpha &= P(\hat{v} \geq \hat{v}_{\text{obs}}; a) = 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{a^n}{n!} e^{-a}, \\ \beta &= P(\hat{v} \leq \hat{v}_{\text{obs}}; b) = \sum_{n=0}^{n_{\text{obs}}} \frac{b^n}{n!} e^{-b}, \end{aligned}$$



得出 a 与 b

a 是 v 的假设值, 使得: $P(\hat{v} \geq \hat{v}_{\text{obs}}) = \alpha$

b 是 v 的假设值, 使得: $P(\hat{v} \leq \hat{v}_{\text{obs}}) = \beta$

泊松分布均值的置信区间确定

利用

$$\sum_{n=0}^m \frac{v^n}{n!} e^{-v} = 1 - F_{\chi^2}(2v; n_d = 2(m + 1))$$

这里 F_{χ^2} 是自由度为 n_d 的卡方变量的累积分布函数。

$$\alpha = 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{a^n}{n!} e^{-a},$$
$$\beta = \sum_{n=0}^{n_{\text{obs}}} \frac{b^n}{n!} e^{-b},$$



$$a = \frac{1}{2} F_{\chi^2}^{-1}(\alpha; n_d = 2n_{\text{obs}}),$$
$$b = \frac{1}{2} F_{\chi^2}^{-1}(1 - \alpha; n_d = 2(n_{\text{obs}} + 1)),$$

这里 $F_{\chi^2}^{-1}$ 是卡方分布的分位数（可以调用ROOT中的函数
Double_t TMath::ChisquareQuantile(Double_t p, Double_t ndf)计算)

泊松分布均值的置信水平上限值

重要特例: $n_{\text{obs}} = 0$

$$\beta = \sum_{n=0}^0 \frac{b^n}{n!} e^{-b} = e^{-b} \quad \Rightarrow \quad b = -\log \beta$$

对于置信水平 $1 - \beta = 95\%$ 的上限,

$$\begin{aligned} b &= -\log(0.05) \\ &= 2.996 \\ &\approx 3 \end{aligned}$$

n_{obs}	lower limit a			upper limit b		
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.01$
0	—	—	—	2.30	3.00	4.61
1	0.105	0.051	0.010	3.89	4.74	6.64
2	0.532	0.355	0.149	5.32	6.30	8.41
3	1.10	0.818	0.436	6.68	7.75	10.04
4	1.74	1.37	0.823	7.99	9.15	11.60
5	2.43	1.97	1.28	9.27	10.51	13.11
6	3.15	2.61	1.79	10.53	11.84	14.57
7	3.89	3.29	2.33	11.77	13.15	16.00
8	4.66	3.98	2.91	12.99	14.43	17.40
9	5.43	4.70	3.51	14.21	15.71	18.78
10	6.22	5.43	4.13	15.41	16.96	20.14

例: 无本底稀有衰变分支比

已知实验对稀有衰变 $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ 的单个事例灵敏度为

$$\text{灵敏度} = \frac{1}{N_{\text{obs}} \times \varepsilon_{\text{eff}}} = \frac{1}{3.93 \times 10^9}$$

如果实验上没有观察到一个事例，要给出90%的置信水平，需计算

$$P(n=0; \mu) = \frac{e^{-\mu} \mu^n}{n!} = 10\%$$



$$\mu = 2.30259$$



$$\begin{aligned} \text{分支比上限} \\ &\leq \frac{2.30259}{3.93 \times 10^9} \\ &= 0.59 \times 10^{-9} \end{aligned}$$

如果实验上观察到一个事例，要给出68%的置信区间的分支比，需要给出重复实验在 $(1-0.68)/2=0.16$ 范围内观察到至少一个事例的下限

$$\sum_{n=1}^{\infty} P(n; \mu) = 1 - P(0; \mu) = 0.16$$



$$\mu = 0.17435$$



$$\begin{aligned} BR &= \frac{1+2.28852}{3.93 \times 10^9} \\ &= (2.54_{-2.10}^{+5.82}) \times 10^{-10} \end{aligned}$$

以及不多于一个事例的均值上限

$$\sum_{n=0}^1 P(n; \mu) = 0.16$$



$$\mu = 3.28852$$

从 $\ln L$ 或 χ^2 近似给出置信区间

若 $\ln L(\theta)$ 呈抛物线状, 通过将 $\ln L(\theta)$ 展开, 则可得到

$$\ln L(\theta) = \log L_{\max} - \frac{(\theta - \hat{\theta})^2}{2\sigma_{\hat{\theta}}^2} \quad \rightarrow \quad \ln L(\hat{\theta} \pm N\sigma_{\hat{\theta}}) = \ln L_{\max} - \frac{N^2}{2}$$

即使 $\ln L(\theta)$ 不呈抛物线状, 也可给出置信区间的近似值

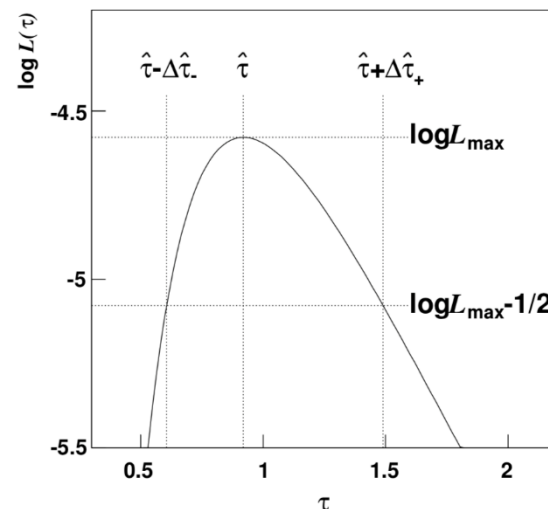
$$\ln(\hat{\theta}_{-c}^{+d}) = \ln L_{\max} - \frac{N^2}{2}, \quad \chi^2(\hat{\theta}_{-c}^{+d}) = \chi_{\min}^2 + N^2$$

这里 $N = \Phi^{-1}\left(1 - \frac{\gamma}{2}\right)$ 是标准正态分布对应于置信水平 $1 - \gamma$ 的分位点, 例如:

$$N = 1 \Rightarrow 1 - \gamma = 0.683$$

在指数函数例子中, 只有 $n = 5$ 个观测值时

$$\hat{\tau} = 0.85_{-0.30}^{+0.52}$$



多维置信区间

研究中常遇到需要给出多参数拟合情况下的多维置信区间

$$\hat{\vec{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$$

参数的联合概率密度函数为

$$g(\hat{\vec{\theta}} | \vec{\theta}) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} \exp \left[-\frac{1}{2} Q(\hat{\vec{\theta}} | \vec{\theta}) \right]$$

其中

$$Q(\hat{\vec{\theta}} | \vec{\theta}) = (\hat{\vec{\theta}} - \vec{\theta})^T V^{-1} (\hat{\vec{\theta}} - \vec{\theta})$$

这里 V^{-1} 为协方差矩阵的逆。当联合概率密度函数值不变时，其等高线对应于常数的 Q 。它们是在参数空间以真值为中央的椭圆（或对于两维以上的超椭圆）。

多维参数置信区间

当参数个数增大时, 由 $Q_\alpha = F_{\chi^2(n)}^{-1}(1 - \alpha)$ 确定的置信区间的置信水平 $CL = 1 - \alpha$ 相应地降低。

Q_α	$1 - \alpha$				
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
1.0	0.683	0.393	0.199	0.090	0.037
2.0	0.843	0.632	0.428	0.264	0.151
4.0	0.954	0.865	0.739	0.594	0.451
9.0	0.997	0.989	0.971	0.939	0.891

多维参数置信区间（续）

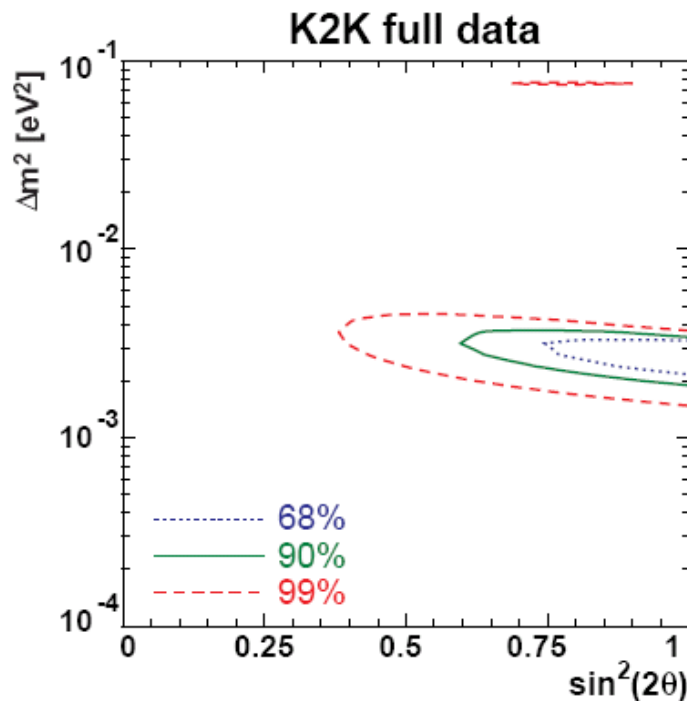
等价地，对于给定 $CL = 1 - \alpha$ ， Q_α 随着 n 增大而增大。

$1 - \alpha$	Q_α				
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
0.683	1.00	2.30	3.53	4.72	5.89
0.90	2.71	4.61	6.25	7.78	9.25
0.95	3.84	5.99	7.82	9.49	11.1
0.99	6.63	9.21	11.3	13.3	15.1

二维参数的置信区间

例如在中微子振荡实验中的双参数拟合问题

$$P(\nu_\mu \rightarrow \nu_\tau) = \sin^2 2\theta \sin^2 \left(\frac{1.27 \Delta m^2 (\text{eV}^2) L (\text{km})}{E_\nu (\text{GeV})} \right)$$



Phys.Rev.D74:072003,2006

物理边界附近的极限

假设 $b = 2.5$ ，观测到 $n = 0$ 。

如果选择 $CL = 0.9$ ，

$$s_{\text{up}} = \frac{1}{2} F_{\chi^2}^{-1}(1 - \alpha; 2(n_{\text{obs}} + 1)) - b$$


$$s_{\text{up}} = -0.197 \quad (CL = 0.90)$$

物理学家：我们一开始就知道 $s \geq 0$ ；不能用负的上限报道昂贵实验的结果！

统计学家：置信区间按照定义只有 90% 的次数涵盖了真值——这无非是没覆盖真值的一次结果。

这种情况并不少见，尤其是在实验灵敏度很低的情况下检验参数值，例如信号 s 非常少的情况。

物理边界附近的极限 (续)

物理学家：我应当选 $CL = 0.95 \rightarrow s_{\text{up}} = 0.496$ 。

“更好的做法”： $CL = 0.917923 \rightarrow s_{\text{up}} = 10^{-4}!$



$$s_{\text{up}} = -0.197 \quad (CL = 0.90)$$

上限怎么会这么小呢？

实际情况： $b = 2.5$ 时， n 的典型泊松涨落至少是 $\sqrt{2.5} = 1.6$ 。

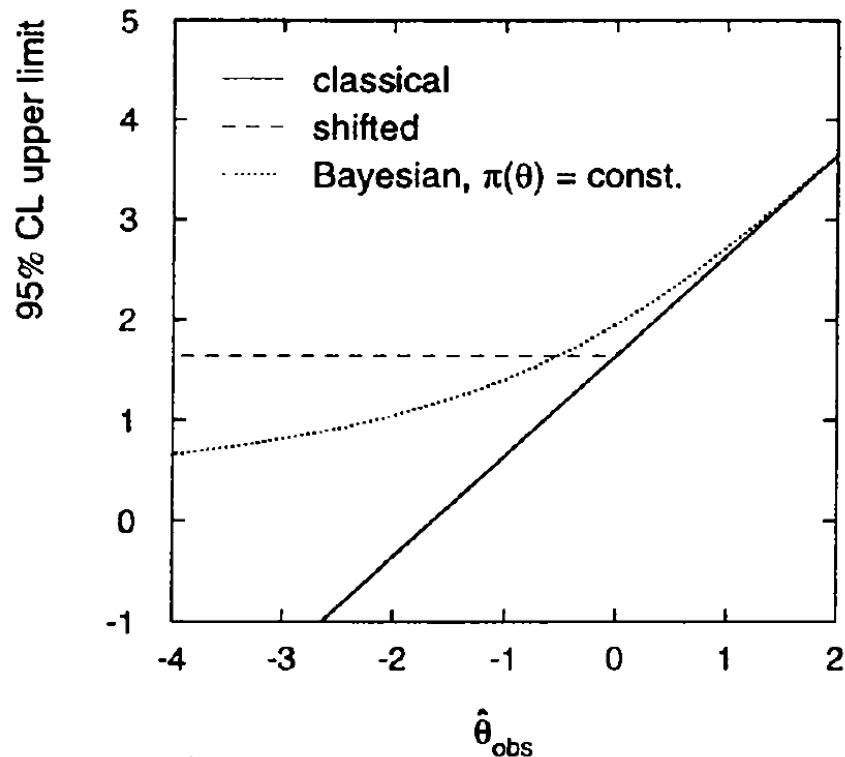
物理边界附近的极限 (续)

1. 经典方法: 不用考虑太多

$$\theta_{\text{up}} = \hat{\theta}_{\text{obs}} + \sigma_{\hat{\theta}} \Phi^{-1}(1 - \beta).$$

2. 偏移方法: 我们稍微把负结果整理一下

$$\theta_{\text{up}} = \max(\hat{\theta}_{\text{obs}}, 0) + \sigma_{\hat{\theta}} \Phi^{-1}(1 - \beta).$$



贝叶斯方法确定上限

贝叶斯统计需要从“验前概率” $\pi(\theta)$ 出发，它反映实验前对于参数的信心程度。

贝叶斯定理给出，观测到数据 x 如何改进我们的信心程度：

$$p(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{-\infty}^{+\infty} L(x|\theta')\pi(\theta')d\theta'} \propto L(x|\theta)\pi(\theta)$$

对验后概率 $p(\theta|x)$ 积分可给出任意期待概率下的区间。

例如，对于 $n \sim Poi(s + b)$ ，95% 置信水平下 s 的上限由下面的积分给出：

$$0.95 = \int_{-\infty}^{s_{\text{up}}} p(s|n) ds$$

贝叶斯验前概率：泊松参数

考虑到 $s \geq 0$ ，可以要求 $s < 0$ 时 $\pi(s) = 0$ 。

一种“验前无倾向”的选择是：

$$\pi(s) = \begin{cases} 1 & s \geq 0 \\ 0 & \text{其他} \end{cases}$$

没有归一化，但只要 s 很大时 $L(s)$ 趋于零即可。

参数变化下不能保持不变。

并不真的反映合理的信心程度，但通常用来作为参考。

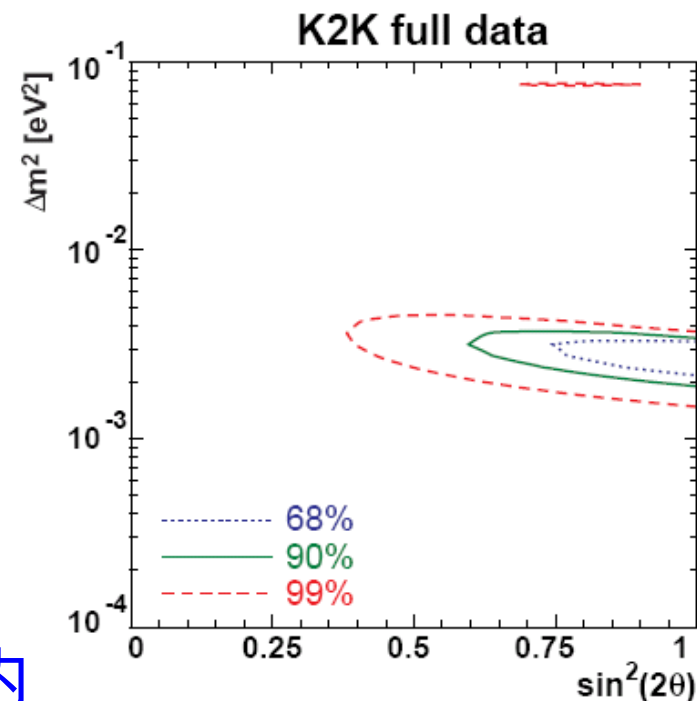
贝叶斯上限的图像！

$$\pi(s) = \begin{cases} 1 & s \geq 0 \\ 0 & \text{其他} \end{cases}$$

$$p(\theta|\mathbf{x}) = \frac{L(\mathbf{x}|\theta) \pi(\theta)}{\int L(\mathbf{x}|\theta') \pi(\theta') d\theta'},$$

极大似然法在物理区间内求极值

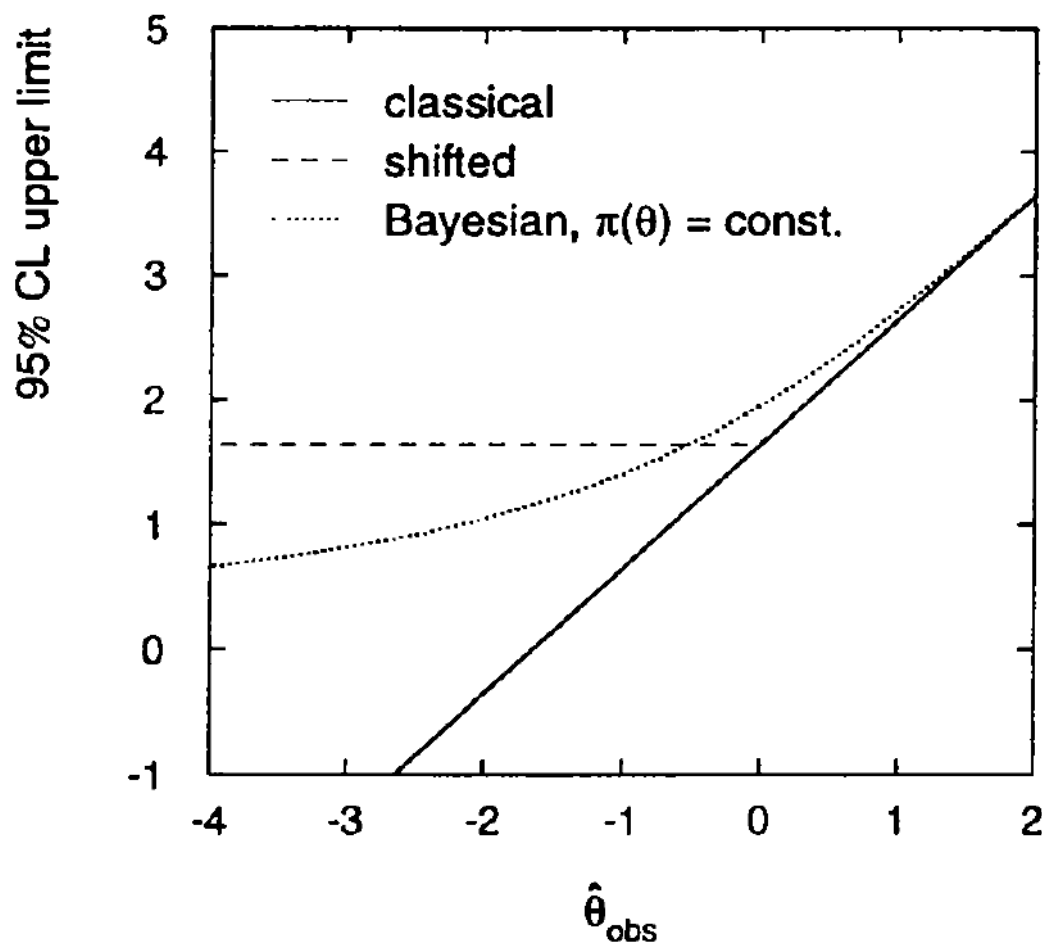
$g(\hat{\theta}; \theta)$ 的 $\hat{\theta}$ 一直是在物理许可的范围内



$$1 - \beta = \int_{-\infty}^{\theta_{\text{up}}} p(\theta|\mathbf{x}) d\theta = \frac{\int_{-\infty}^{\theta_{\text{up}}} L(\mathbf{x}|\theta) \pi(\theta) d\theta}{\int_{-\infty}^{\infty} L(\mathbf{x}|\theta) \pi(\theta) d\theta}.$$

贝叶斯上限：均匀分布的 s 验前概率

三种置信区间的对比



含本底泊松分布的经典置信区间

在观察到 n_{obs} 个事例条件下的置信区间的确定

对信号均值 ν_s 的最大似然无偏估计量为: $\hat{\nu}_s = n - \nu_b$

下限:

$$\alpha = P(\hat{\nu}_s \geq \hat{\nu}_s^{\text{obs}}; \nu_s^{\text{low}}) = \sum_{n \geq n_{\text{obs}}} \frac{(\nu_s^{\text{low}} + \nu_b)^n e^{-(\nu_s^{\text{low}} + \nu_b)}}{n!}$$

上限:

$$\beta = P(\hat{\nu}_s \leq \hat{\nu}_s^{\text{obs}}; \nu_s^{\text{up}}) = \sum_{n \leq n_{\text{obs}}} \frac{(\nu_s^{\text{up}} + \nu_b)^n e^{-(\nu_s^{\text{up}} + \nu_b)}}{n!}$$

可能会出现的问题:

如果本底研究给出的预言值 ν_b 与实验观测值 n_{obs} 可比较, 那么可能会出现信号事例上限 ν_s^{up} 只可能取零值的情况。
(参见教材P141图9.9 (a))



如果出现这种情况, 需要采用贝叶斯方法。

贝叶斯方法确定置信区间

在期待值为 ν_s 的条件下观察到 n_{obs} 个事例的似然函数

$$L(n_{\text{obs}}|\nu_s) = \frac{(\nu_s + \nu_b)^{n_{\text{obs}}} e^{-(\nu_s + \nu_b)}}{n_{\text{obs}}!}$$

根据贝叶斯定理，期待值为 ν_s 的后验概率密度函数为

$$p(\nu_s|n_{\text{obs}}) = \frac{L(n_{\text{obs}}|\nu_s)\pi(\nu_s)}{\int_{-\infty}^{+\infty} L(n_{\text{obs}}|\nu'_s)\pi(\nu'_s)d\nu'_s}$$

如果取先验概率密度函数

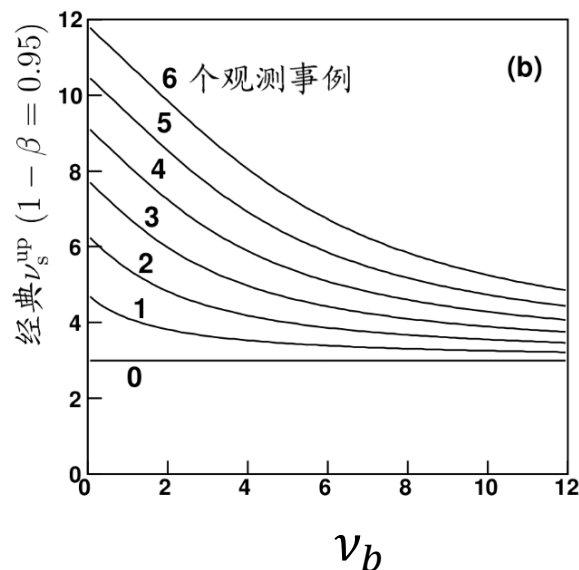
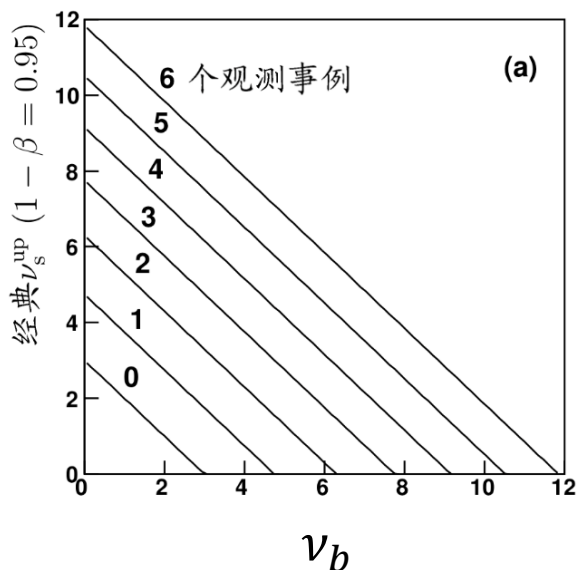
$$\pi(\nu_s) = \begin{cases} \text{常数} & (\nu_s \geq 0) \\ 0 & (\nu_s < 0) \end{cases}$$

贝叶斯上限

因此，上限可表示为

$$\beta = 1 - \frac{\int_0^{\nu_s^{\text{up}}} L(n_{\text{obs}}|\nu_s) d\nu_s}{\int_0^{+\infty} L(n_{\text{obs}}|\nu_s) d\nu_s} = 1 - \frac{\int_0^{\nu_s^{\text{up}}} (\nu_s + \nu_b)^{n_{\text{obs}}} e^{-(\nu_s + \nu_b)} d\nu_s}{\int_0^{+\infty} (\nu_s + \nu_b)^{n_{\text{obs}}} e^{-(\nu_s + \nu_b)} d\nu_s}$$
$$= \frac{e^{-(\nu_s^{\text{up}} + \nu_b)} \sum_{n=0}^{n_{\text{obs}}} \frac{(\nu_s^{\text{up}} + \nu_b)^n}{n!}}{e^{-\nu_b} \sum_{n=0}^{n_{\text{obs}}} \frac{\nu_b^n}{n!}}$$

- 1) 贝叶斯上限 $\nu_s > 0$
- 2) $\nu_s = 0$ 回到经典上限



小结

- 统计不确定度中的标准差问题
- 经典置信区间问题
- 利用似然函数或二乘函数确定置信区间
- 多维置信区间
- 物理边界

从似然函数估计近似的置信区间

假设利用似然比检验参数值 $\vec{\theta} = (\theta_1, \dots, \theta_n)$:

$$\lambda(\vec{\theta}) = \frac{L(\vec{\theta})}{L(\hat{\vec{\theta}})}$$

$$0 \leq \lambda(\vec{\theta}) \leq 1$$

$\lambda(\vec{\theta})$ 小意味着数据与假设的 $\vec{\theta}$ 符合更差。等价地，通常定义

$$t_{\vec{\theta}} = -2 \ln \lambda(\vec{\theta})$$

$t_{\vec{\theta}}$ 大意味着数据与假设的 $\vec{\theta}$ 符合得更差。

因此， $\vec{\theta}$ 的 p 值为:

$$p_{\vec{\theta}} = \int_{t_{\vec{\theta}, \text{obs}}}^{\infty} f(t_{\vec{\theta}} | \vec{\theta}) dt_{\vec{\theta}}$$

修改概率密度 $f(t_{\vec{\theta}} | \vec{\theta})$

从Wicks定理估计置信区间

Wicks定理（满足大样本极限和其他一些条件）：

$$f(t_{\vec{\theta}} | \vec{\theta}) \sim \chi^2(n)$$

自由度数目等于参数 $\vec{\theta} = (\theta_1, \dots, \theta_n)$ 分量个数 n 的卡方分布

假设Wicks定理成立， p 值为

$$p_{\vec{\theta}} = 1 - F_{\chi^2(n)}(t_{\vec{\theta}})$$

要得到置信区间的边界，令 $p_{\vec{\theta}} = \alpha$ 并求解 $t_{\vec{\theta}}$ ：

$$t_{\vec{\theta}} = F_{\chi^2(n)}^{-1}(1 - \alpha)$$

注意， $t_{\vec{\theta}}$ 还可以表示为：

$$t_{\vec{\theta}} = -2 \ln \frac{L(\vec{\theta})}{L(\hat{\vec{\theta}})}$$

从Wicks定理估计置信区间（续）

在 $\vec{\theta}$ 空间，置信区域的边界为

$$L(\vec{\theta}) = L(\hat{\vec{\theta}}) - \frac{1}{2} F_{\chi^2(n)}^{-1}(1 - \alpha)$$

例如，对于 $1 - \alpha = 68.3\%$ 和 $n = 1$ 个参数

$$F_{\chi^2(n)}^{-1}(0.683) = 1$$

所以，68.3%置信水平的置信区间由下式确定：

$$L(\theta) = L(\hat{\theta}) - \frac{1}{2}$$

这与求估计量标准差的方法一样，即

$$[\hat{\theta} - \sigma_{\hat{\theta}}, \hat{\theta} + \sigma_{\hat{\theta}}] \text{ 是 } CL = 68.3\% \text{ 的置信区间。}$$

讨论：检验/置信区间的要素

需要注意的是，我们只能利用观测数据计算的似然函数估计这些置信区间。这种做法可行的原因是，大样本极限下

统计量 $t_{\vec{\theta}} = -2 \ln \frac{L(\vec{\theta})}{L(\hat{\vec{\theta}})}$ 趋于定义明确的分布，与数据无关。

然而，对于有限样本，得到的区间是近似的。

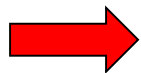
通常来说，要做一个统计检验，需要知道检验统计量 $t(x)$ 的分布函数，这意味着需要知道 $P(x|\vec{\theta})$ 的完整信息。

泊松参数 s 的上限：频率论

考虑观测事例数 $n \sim \pi(s + b)$, 假设 $b = 4.5$, $n_{\text{obs}} = 5$ 。
求置信水平 $\text{CL} = 95\%$ 的上限。

相关的备择假设是 $s = 0$ (小 n 时的临界域), 假设的 s 的
 p 值为 $P(n \leq n_{\text{obs}}; s, b)$, 求解下式得 $\text{CL} = 1 - \alpha$ 的上限 s_{up}

$$\alpha = P(n \leq n_{\text{obs}}; s_{\text{up}}, b) = \sum_{n=0}^{n_{\text{obs}}} \frac{(s_{\text{up}} + b)^n}{n!} e^{-(s_{\text{up}} + b)}$$



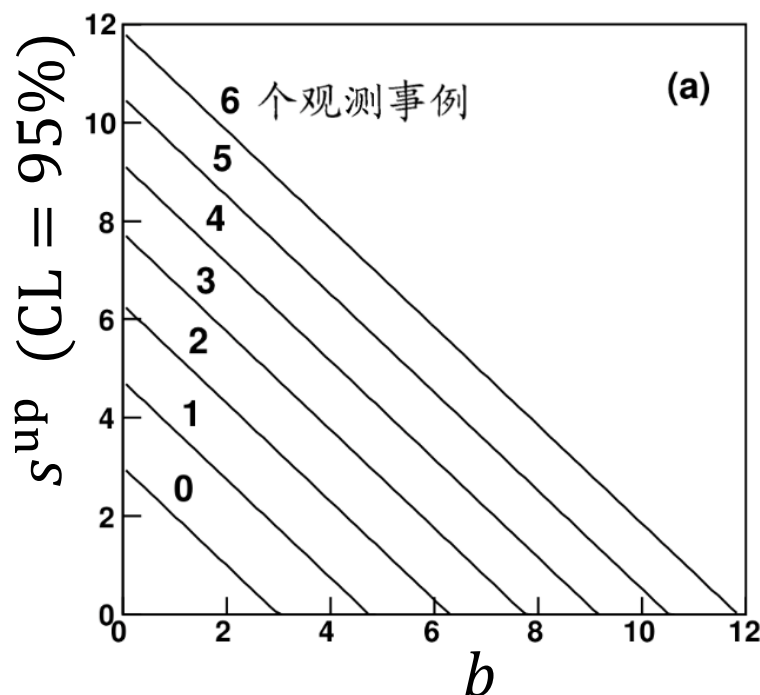
$$\begin{aligned} s_{\text{up}} &= \frac{1}{2} F_{\chi^2}^{-1}(1 - \alpha; 2(n_{\text{obs}} + 1)) - b \\ &= \frac{1}{2} F_{\chi^2}^{-1}(0.95; 2(5 + 1)) - 4.5 \\ &= 6.0 \end{aligned}$$

泊松参数 s 的上限：频率论

对于事例数 n 很小的涨落,

$$s_{\text{up}} = \frac{1}{2} F_{\chi^2}^{-1}(1 - \alpha; 2(n_{\text{obs}} + 1)) - b \quad \text{可以给出负的结果。}$$

即, 置信区间可能是空的。



由形式规则确定验前概率

由于很难将模糊的信心程度体现在验前概率中，人们经常尝试由形式规则（formal rules）确定验前概率，例如，要求满足某种不变原则，或者对某组测量提供最大的信息增益。

经常称其为“客观验前概率”
这构成了“客观贝叶斯统计”的基础

验前概率不反映信心程度（但可能表示可能的极端情形）。

在“客观贝叶斯”分析中，可以像频率论那样使用置信区间，即，将贝叶斯定理看作产生具有特定覆盖性质的区间的方法。

相关综述参见：R. E. Kass and L. Wasserman, *The Selection of Prior Distributions by Formal rules*, J. Am. Stat. Assoc., Vol. 91, 1343 (1996)

粒子物理中的应用：L. Demortier, S. Jain and H. Prosper, *Reference priors for high energy physics*, Phys. Rev. D82 (2010) 034002 [arXiv:1002.1111]

D. Casadei, Reference analysis of the signal + background model in counting experiments, JINST 7 (2012) 01012 [arXiv:1108.4270]

Jeffrey's 验前概率

根据 “Jeffrey 规则” , 验前概率取为

$$\pi(\vec{\theta}) \propto \sqrt{\det(I(\vec{\theta}))}$$

其中 $I(\vec{\theta})$ 是费舍尔信息矩阵:

$$I_{ij}(\vec{\theta}) = -E \left[\frac{\partial^2 \ln L(\vec{x}|\vec{\theta})}{\partial \theta_i \partial \theta_j} \right] = - \int \frac{\partial^2 \ln L(\vec{x}|\vec{\theta})}{\partial \theta_i \partial \theta_j} L(\vec{x}|\vec{\theta}) d\vec{x}$$

可以证明, 这种验前概率选择给出的推断在参数变换下不变。

对于高斯分布的均值 μ , Jeffrey 验前概率为常数;
对于泊松分布的均值 μ , Jeffrey 验前概率正比于 $1/\sqrt{\mu}$ 。

Jeffrey's 验前概率：泊松分布均值

假设 $n \sim Poi(\mu)$, 要得到 μ 的 Jeffrey 验前概率, 取为

$$L(n|\mu) = \frac{\mu^n}{n!} e^{-\mu}$$

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\mu^2}$$

$$I(\mu) = -E \left[\frac{\partial^2 \ln L}{\partial \mu^2} \right] = \frac{E[n]}{\mu^2} = \frac{1}{\mu}$$

$$\pi(\mu) \propto \sqrt{I(\mu)} = \frac{1}{\sqrt{\mu}}$$

例如, 对于 $\mu = s + b$, 这意味着验前概率 $\pi(s) \sim 1/\sqrt{s + b}$, 依赖于 b 。这不是关于 s 的信心程度。

搜索型分析

在某个相空间寻找信号，结果用变量 x 的直方图表示：

$$\vec{n} = (n_1, \dots, n_N)$$

假设 n_i 服从泊松分布，其均值为 $E[n_i] = \mu s_i + b_i$ ，其中

$$\text{信号 } s_i = s_{\text{tot}} \int_{\text{bin } i} f_s(x; \vec{\theta}_s) dx;$$

μ : 强度参数

$$\text{本底 } b_i = b_{\text{tot}} \int_{\text{bin } i} f_b(x; \vec{\theta}_b) dx$$

通常还有辅属测量，用来约束本底或形状参数：

$$\vec{m} = (m_1, \dots, m_N)$$

假设 m_i 服从泊松分布，其均值为 $E[m_i] = u_i(\vec{\theta})$ 。

似然函数为

冗余参数 $\vec{\theta}_s, \vec{\theta}_b, b_{\text{tot}}$

$$L(\mu, \vec{\theta}) = \prod_{j=1}^N \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^M \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

搜索型实验的统计检验

考虑参数 μ ，正比于某尚未发现的信号过程的发生率。

假设描述数据的模型同时包含 μ 和一组冗余参数 $\vec{\theta}$ 。

为了检验假设的 μ 值，利用轮廓似然比 (profile likelihood ratio)

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\hat{\theta}})}$$

$\hat{\hat{\theta}}$: 对指定的 μ 值使 L 最大

$\hat{\mu}, \hat{\hat{\theta}}$: 使 L 最大

物理发现或上限的检验统计量

物理发现：利用检验统计量 q_0 拒绝纯本底 ($\mu = 0$) 假设

$$q_0 = \begin{cases} -2 \ln \lambda(0), & \hat{\mu} \geq 0 \\ 0, & \hat{\mu} < 0 \end{cases}$$

即，我们把正的 μ 当作相关的备择假设，所以临界域选为对应于 $\hat{\mu}$ 值大的区域。

注意：尽管物理上 $\mu \geq 0$ ，我们在这里允许 $\hat{\mu}$ 为负值。在大样本极限下， $\hat{\mu}$ 服从高斯分布，从而检验统计量的分布可以化成简单的形式。

上限：为了对 μ 设定上限，利用 q_μ 作为检验统计量

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu), & \hat{\mu} \leq 0 \\ 0, & \hat{\mu} > 0 \end{cases}$$

我们取相关的备择假设为 μ 较小，所以临界域定义为对应于 $\hat{\mu}$ 值小的区域。

轮廓似然比的 Wald 近似

为了求 p 值, 需要知道 $f(q_0|0)$ 和 $f(q_\mu|\mu)$ 。

对于备择假设下的中位数显著性, 需要知道 $f(q_\mu|\mu')$ 。

利用 Wald 近似:

$$-2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + o(1/\sqrt{N})$$

$$\hat{\mu} \sim N(\mu', \sigma^2) \Rightarrow E[\hat{\mu}] = \mu'$$

N : 样本容量

σ 可从协方差矩阵得到: $V^{-1} = -E \left[\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right]$

如果可以忽略 $o(1/\sqrt{N})$ 项, 则 $-2 \ln \lambda(\mu)$ 服从非中心的自由度为 1 的卡方分布, 非中心参数为

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}$$

特例: 如果 $\mu' = \mu$, 则 $\Lambda = 0$, $-2 \ln \lambda(\mu) \sim \chi^2(1)$ 【Wilks 定理】。

大样本极限下 q_0 的分布

假设大样本极限下近似成立，可以将 q_0 的完整分布写为

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right) \delta(q_0) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_0}} e^{-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2}$$

$\mu' = 0$ 的特例是 “半卡方分布”：

$$f(q_0|0) = \frac{1}{2} \delta(q_0) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_0}} e^{-q_0/2}$$

在大样本极限下， $f(q_0|0)$ 与冗余参数无关； $f(q_0|\mu')$ 通过 σ 依赖于冗余参数。

大样本极限下 q_0 的累积分布、显著性

根据 q_0 的概率密度，其累积分布为

$$F(q_0|\mu') = \Phi\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)$$

$\mu' = 0$ 的特例:

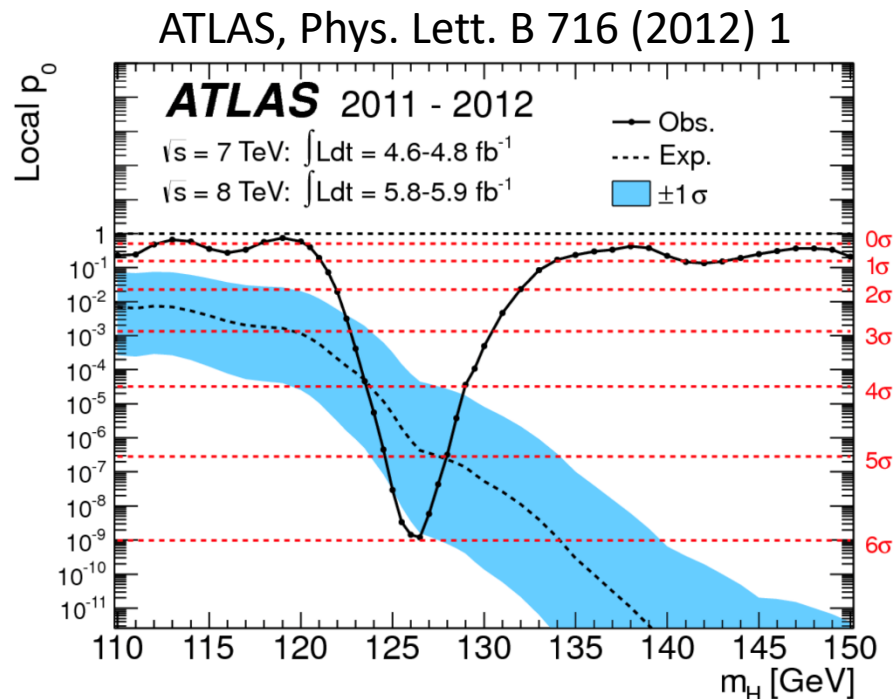
$$F(q_0|0) = \Phi(\sqrt{q_0})$$

$\mu = 0$ 假设的 p 值为

$$p_0 = 1 - F(q_0|0)$$

因此，物理发现的显著性 Z 为

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}$$



G. Cowan, K. Cranmer, E. Gross, O. Vitells, EPJC 71 (2011) 1554 [arXiv:1007.1727]

上限 q_μ 的分布

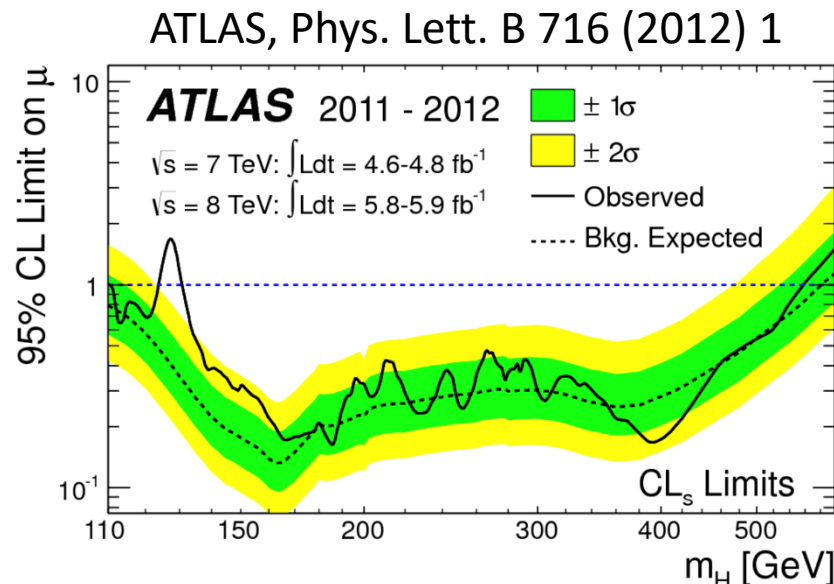
对 q_μ 可以得到类似结果

$$f(q_\mu|\mu') = \Phi\left(\frac{\mu' - \mu}{\sigma}\right) \delta(q_\mu) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_\mu}} e^{-\frac{1}{2}\left(\sqrt{q_\mu} - \frac{(\mu - \mu')}{\sigma}\right)^2}$$

$$f(q_\mu|\mu) = \frac{1}{2} \delta(q_\mu) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_\mu}} e^{-\frac{q_\mu}{2}}$$

$$F(q_\mu|\mu') = \Phi\left(\sqrt{q_\mu} - \frac{(\mu - \mu')}{\sigma}\right)$$

$$p_\mu = 1 - F(q_\mu|\mu) = 1 - \Phi\left(\sqrt{q_\mu}\right)$$



注意：这是 CL_s 上限而不是似然比上限。
参见 G. Cowan, arXiv:1307.2487.

G. Cowan, K. Cranmer, E. Gross, O. Vitells, EPJC 71 (2011) 1554 [arXiv:1007.1727]

本底 b 不确定的轮廓似然函数

测量两个泊松变量:

$$\begin{array}{ll} n \sim \text{Poi}(s + b) & \text{(主测量或“搜索”测量)} \\ m \sim \text{Poi}(\tau b) & \text{(控制测量, } \tau \text{ 已知)} \end{array}$$

似然函数为

$$L(s, b) = \frac{(s + b)^n}{n!} e^{-(s+b)} \frac{(\tau b)^m}{m!} e^{-\tau b}$$

利用 L 构造轮廓似然比 (b 为冗余参数):

$$\lambda(0) = \frac{L(0, \hat{\hat{b}}(0))}{L(\hat{s}, \hat{b})}$$

需要的要素包括:

$$\hat{s} = n - m/\tau$$

$$\hat{b} = m/\tau$$

$$\hat{\hat{b}}(s) = \frac{n + m - (1 + \tau)s + \sqrt{(n + m - (1 + \tau)s)^2 + 4(1 + \tau)sm}}{2(1 + \tau)}$$

对于物理发现 ($s = 0$) 的检验:

$$\hat{\hat{b}}(0) = \frac{n + m}{(1 + \tau)}$$

渐进显著性

利用:

$$\hat{s} = n - m/\tau$$

$$\hat{b} = m/\tau$$

$$\hat{b}(0) = \frac{n + m}{(1 + \tau)}$$



$$\lambda(0) = \frac{L(0, \hat{b}(0))}{L(\hat{s}, \hat{b})} = \left(\frac{n + m}{(1 + \tau)n} \right)^n \left(\frac{\tau(n + m)}{(1 + \tau)m} \right)^m$$



$$Z = \sqrt{q_0} = \sqrt{-2 \ln \lambda(0)}$$

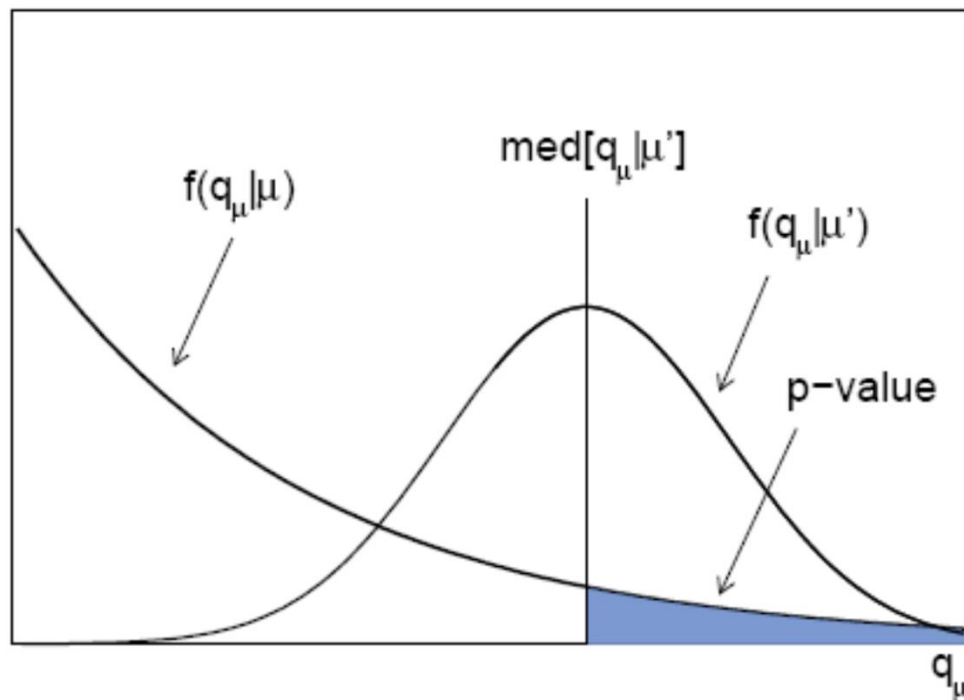
$$= \left[-2 \left(n \ln \left[\frac{n + m}{(1 + \tau)n} \right] + m \ln \left[\frac{\tau(n + m)}{(1 + \tau)m} \right] \right) \right]^{1/2}$$

$$n > \hat{b}$$

$$Z = 0 \quad n \leq \hat{b}$$

期待的（或中位数）显著性/灵敏度

筹划实验时，我们希望量化实验对可能的发现的灵敏度，例如，在假定某个非零的强度参数 μ' 的情况下，通过给定的中位数显著性来量化。



求 p 值需要 $f(q_0 | 0)$; 求灵敏度需要 $f(q_0 | \mu')$ 。

含本底的计数实验中期待的发现显著性

1. 本底 b 已知的计数实验的发现灵敏度:

(a) $\frac{s}{\sqrt{b}}$

(b) 轮廓似然比及Asimov近似: $\sqrt{2((s+b)\ln(1+s/b) - s)}$

2. 本底 b 不确定度为 σ_b 的发现灵敏度:

(a) $\frac{s}{\sqrt{b + \sigma_b^2}}$

(b) 轮廓似然比及Asimov近似:

$$\sqrt{2 \left((s+b) \ln \left(\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right) - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right)}$$

本底已知的计数实验

实验记录事例数 $n \sim Poi(s + b)$, 其中

s = 期待的信号数

b = 期待的本底数

要检验信号的发现, 需要计算 $s = 0$ 假设的 p 值:

$$p = P(n \geq n_{\text{obs}} | b) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{b^n}{n!} e^{-b} = 1 - F_{\chi^2}(2b; 2n_{\text{obs}})$$

通常将 p 值转换为等价的显著性: $Z = \Phi^{-1}(1 - p)$, 其中 Φ 是标准正态的累积分布, 例如, $Z > 5$ (5σ 效应) 意味着 $p < 2.9 \times 10^{-7}$ 。

要表征对发现的灵敏度, 可以给出给定 s 的假定下期望的 (均值或中位数) 显著性 Z 。

期待的发现显著性: s/\sqrt{b}

$s + b$ 很大时, $n \rightarrow x \sim N(\mu, \sigma^2)$, $\mu = s + b$, $\sigma^2 = s + b$ 。

对于观测值 x_{obs} , $s = 0$ 假设的 p 值为 $P(x > x_{\text{obs}} | s = 0)$:

$$p_0 = 1 - \Phi\left(\frac{x_{\text{obs}} - b}{\sqrt{b}}\right)$$

因此, 拒绝 $s = 0$ 的显著性为:

$$Z_0 = \Phi^{-1}(1 - p_0) = \frac{x_{\text{obs}} - b}{\sqrt{b}}$$

假定信号率为 s 时, 期待的 (中位数) 显著性为:

$$\text{median}[Z_0 | s + b] = \frac{s}{\sqrt{b}}$$

显著性的更好的近似

参数 s 的泊松似然函数为 $L(s) = \frac{(s+b)^n}{n!} e^{-(s+b)}$

为了检验某个发现，利用轮廓似然比：

$$q_0 = \begin{cases} -2 \ln \lambda(0), & \hat{s} \geq 0 \\ 0, & \hat{s} < 0 \end{cases}$$

$$\lambda(s) = \frac{L(s, \hat{\theta}(s))}{L(\hat{s}, \hat{\theta})}$$

此时没有
冗余参数

所以，检验 $s = 0$ 的似然比统计量为

$$q_0 = -2 \ln \frac{L(0)}{L(\hat{s})} = 2 \left(n \ln \frac{n}{b} + b - n \right)$$

($n > b$, 其他情况下 $q_0 = 0$)

$s + b$ 足够大时，利用 Wilks 定理

$$q_0 = \sqrt{2 \left(n \ln \frac{n}{b} + b - n \right)}$$

($n > b$, 其他情况下 $Z = 0$)

为了求 $\text{median}[Z|s]$ ，令 $n \rightarrow s + b$ (即Asimov数据集)：

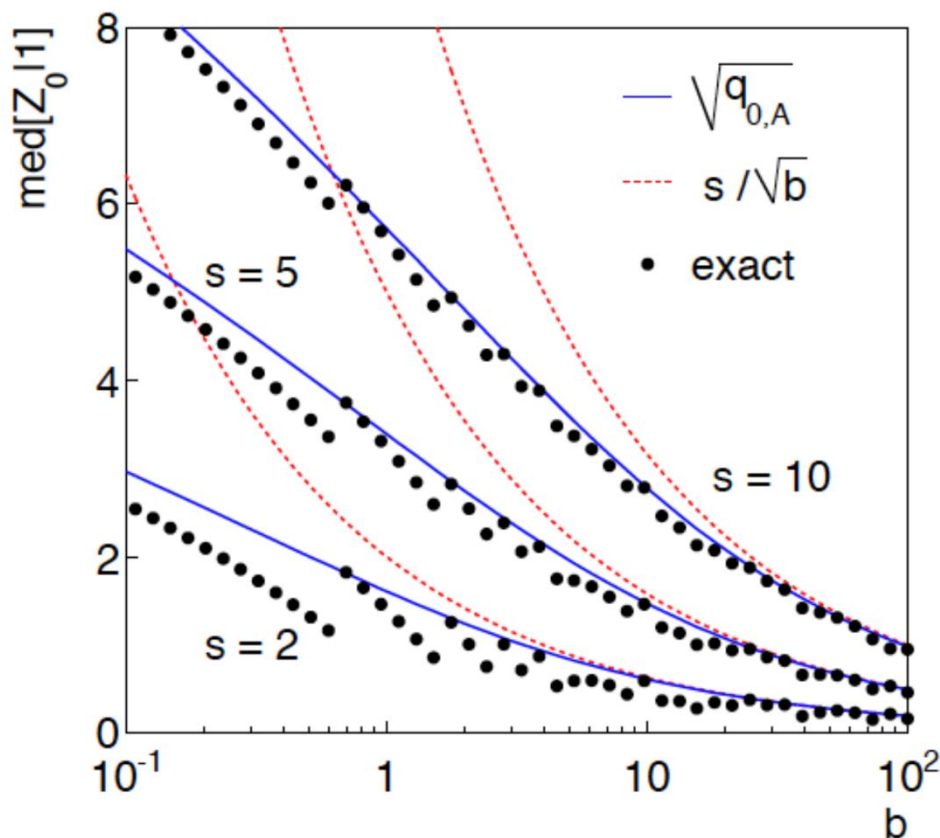
$$Z_A = \sqrt{2((s+b) \ln(1+s/b) - s)}$$

$s \ll b$ 时，回到 s/\sqrt{b} 。

$s = 0$ 假设的中位数显著性的对比

$n \sim \text{Poi}(s + b)$, 假定 s 条件下 $s = 0$ 假设的中位数显著性:

CCGV, EPJC 71 (2011) 1554, arXiv:1007.1727



“Exact” :
来自MC, 数据离散导致跳跃。

$\sqrt{q_{0,A}}$:
在很大范围的 s 和 b 都是很好的近似。

s/\sqrt{b} :
仅在 $s \ll b$ 时符合得才较好近。

s/\sqrt{b} 推广到 b 不确定的情形

s/\sqrt{b} 的直观解释是，它比较的是信号 s 与零信号假设下 n 的标准差 \sqrt{b} 。

现在假设 b 的值不确定，其标准差为 σ_b 。

一个合理的猜测是，将 \sqrt{b} 替换为 \sqrt{b} 和 σ_b 的平方和，即

$$\text{median}[Z|s] = \frac{s}{\sqrt{b + \sigma_b^2}}$$

人们有时用这种方法优化某些分析，例如， σ_b 不可忽略的一些分析。

中位数显著性的Asimov近似

为了得到物理发现的中位数显著性，假定本底+信号模型，并用它们的期待值替换 n 和 m ： $n \rightarrow s + b$, $m \rightarrow \tau b$

$$\rightarrow Z_A = \left[-2 \left((s+b) \ln \left[\frac{s + (1+\tau)b}{(1+\tau)(s+b)} \right] + \tau b \ln \left[1 + \frac{s}{(1+\tau)b} \right] \right) \right]^{1/2}$$

或者利用 $\hat{b} = m/\tau$ 的方差 $V[\hat{b}] \equiv \sigma_b^2 = b/\tau$ 消除 τ ：

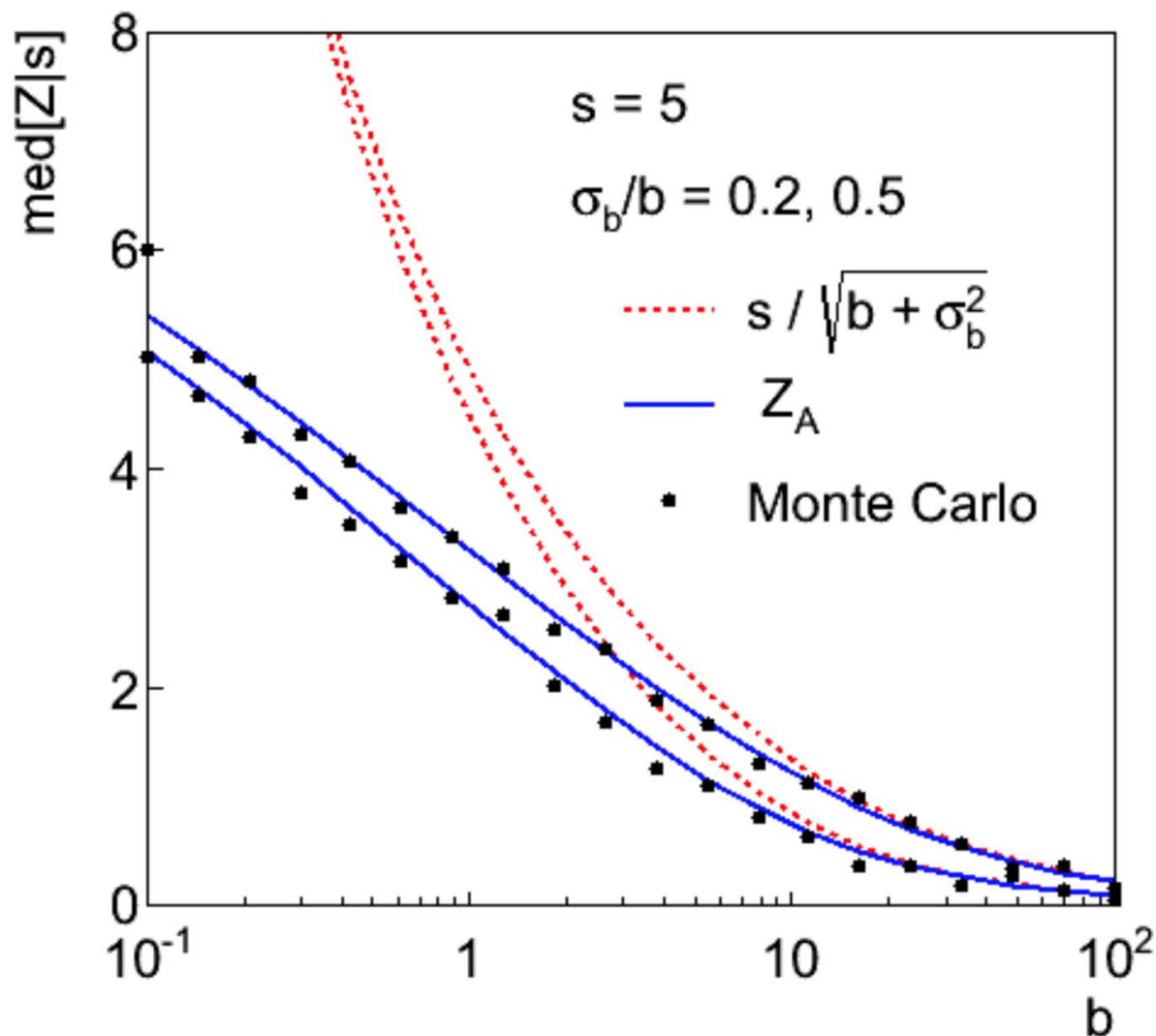
$$Z_A = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

按 s/b 和 $\sigma_b^2/b (= 1/\tau)$ 展开得到：

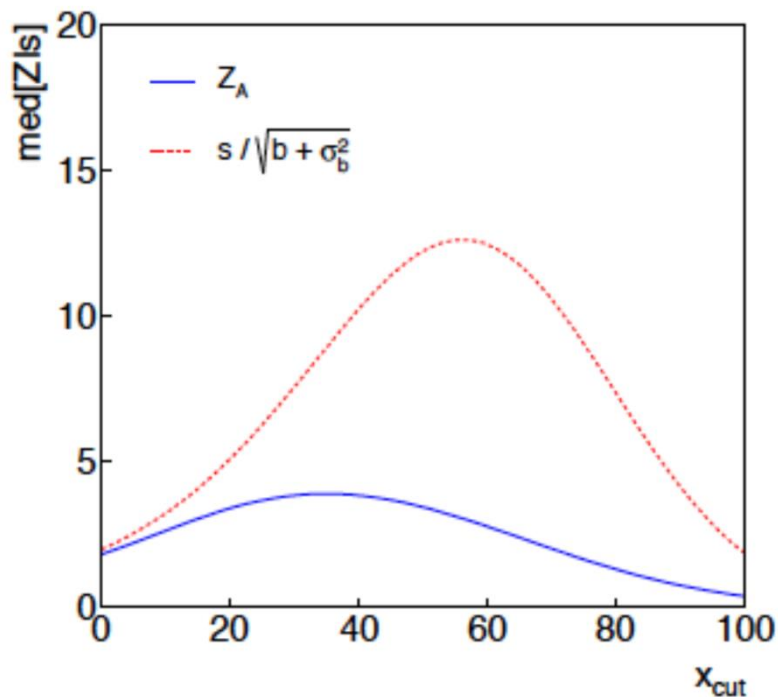
$$Z_A = \frac{s}{\sqrt{b + \sigma_b^2}} \left(1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b) \right)$$

所以，“直观的”公式可以解释为一种极限情形，即将轮廓似然比检验用于Asimov数据集。

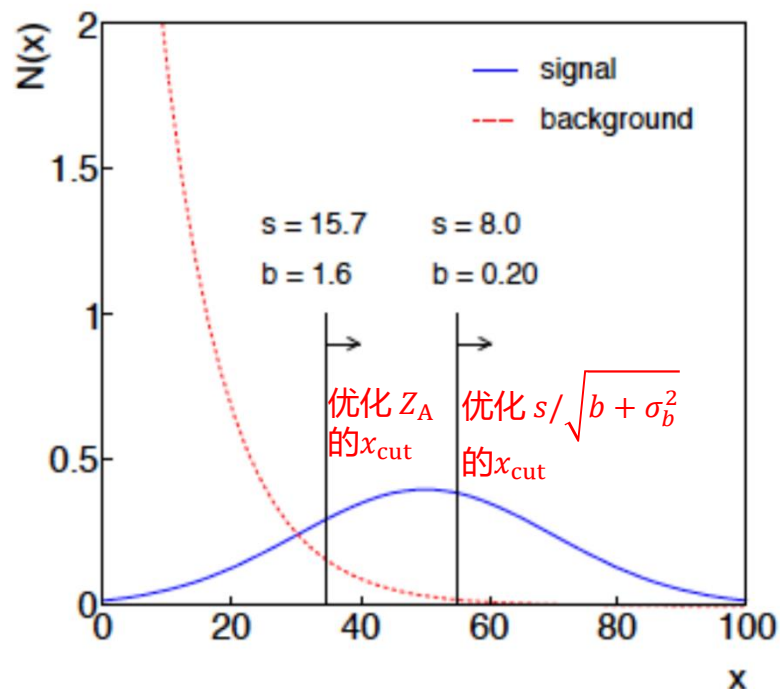
几个公式的对比: $s = 5$



用显著性/灵敏度优化事例筛选



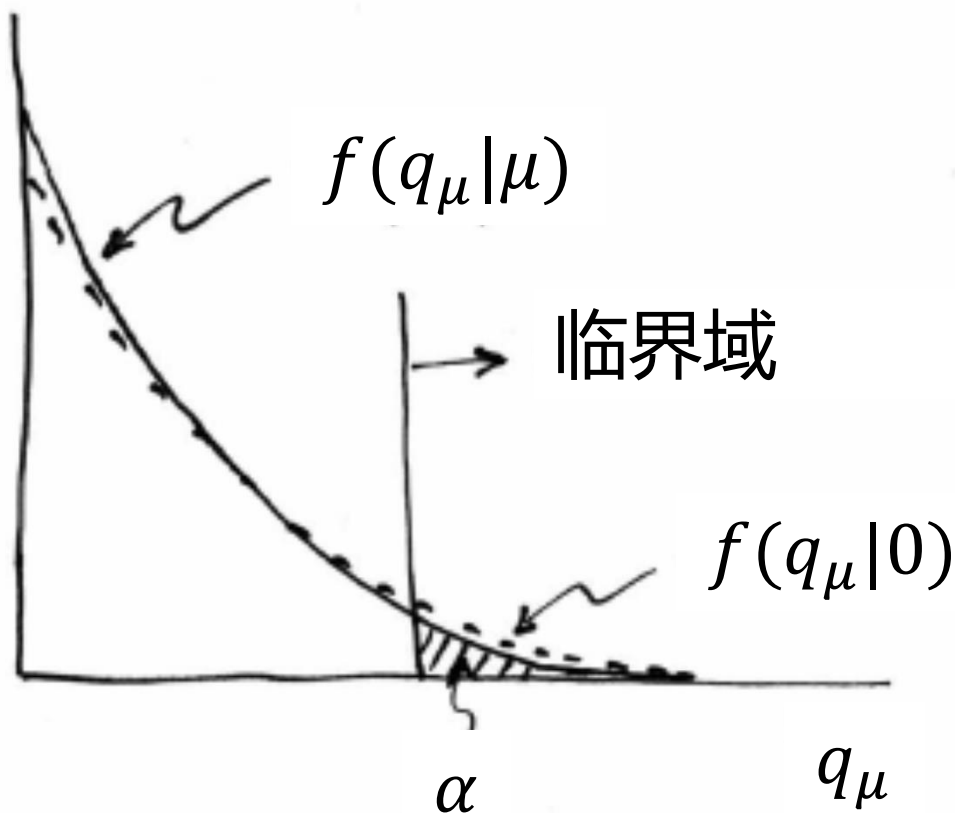
期待的显著性随筛选条件
 x_{cut} 的变化。



信号与本底的分布及不同的
事例筛选结果。

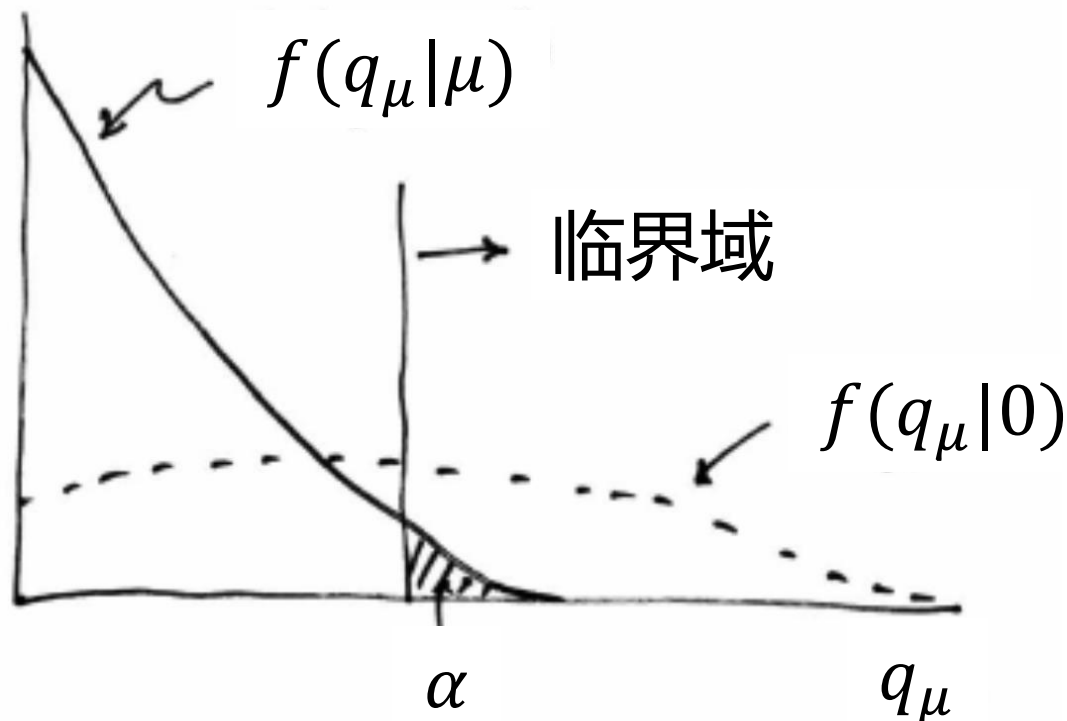
对 μ 的检验灵敏度很低的情形

有时，给定假设的 μ 的效应相对于纯本底 ($\mu = 0$) 预言的效应很小。这意味着 $f(q_\mu|\mu)$ 和 $f(q_\mu|0)$ 几乎相同：



对 μ 的检验灵敏度很高的情形

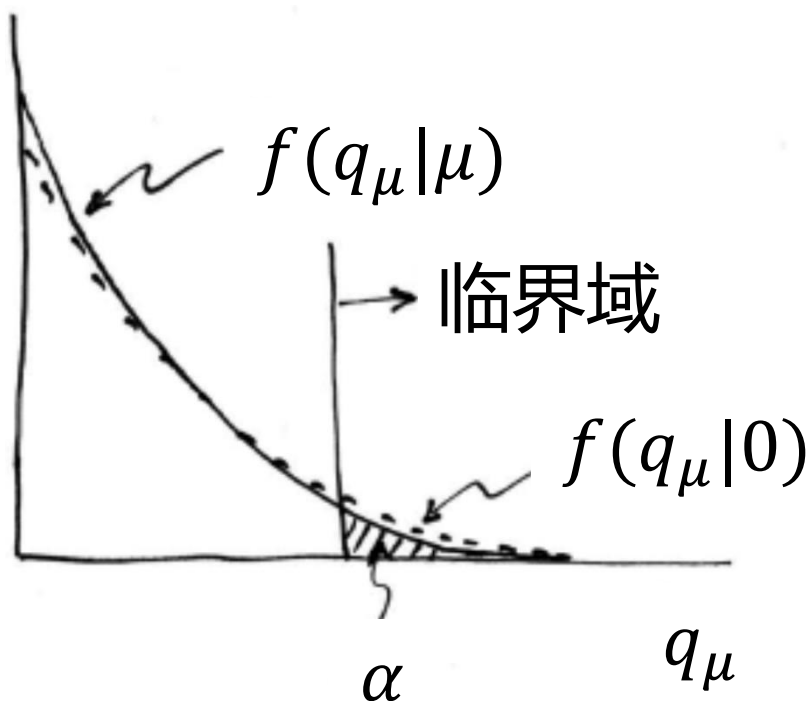
如果对 μ 的检验灵敏度很高, 意味着 $f(q_\mu|\mu)$ 和 $f(q_\mu|0)$ 明显不同:



即, 检验效力 ($\mu = 0$ 时拒绝 μ 的概率) 显著高于 α 。
这个效力可以表征灵敏度。

不可靠的假设排除情形

考虑低灵敏度。 μ 为真时拒绝 μ 的概率为 α (例如5%) , 而 $\mu = 0$ 时拒绝 μ 的概率 (效力) 稍大于 α 。



这意味着, 我们将以稍大于 $\alpha = 5\%$ 的概率拒绝 μ 为真的假设, 但几乎没有灵敏度。

这称为 “不可靠的假设排除”。

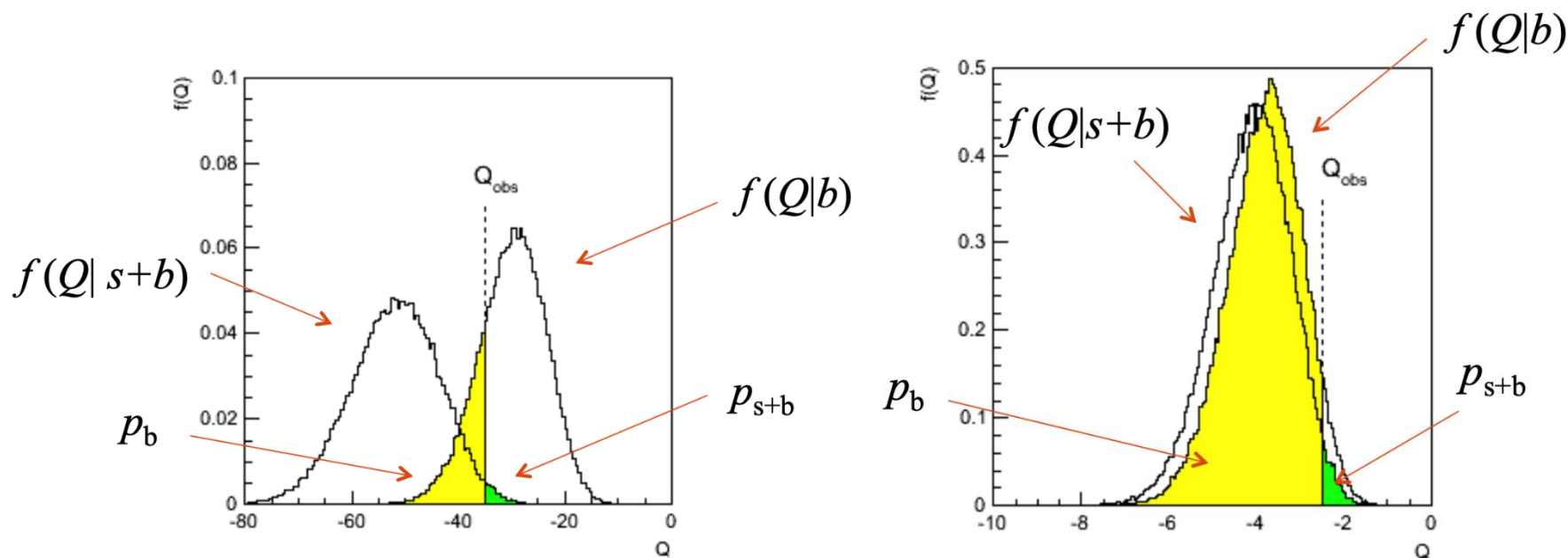
为了解决这个问题, 提出了 “ CL_s ” 方法设置上限。

T. Junk, Nucl. Instrum. Methods Phys. Res., Sec. A 434, 435 (1999);
A. L. Read, J. Phys. G 28, 2693 (2002).

CL_s 方法

在通常的CL_s表示中，我们用同一个统计量 $Q = -2 \ln L_{s+b} / L_b$ 同时检验假设 $\mu = 0$ (b) 和 $\mu > 0$ ($\mu s + b$)。

“灵敏度低”意味着在 b 和 $s + b$ 假设下 Q 的分布非常接近。

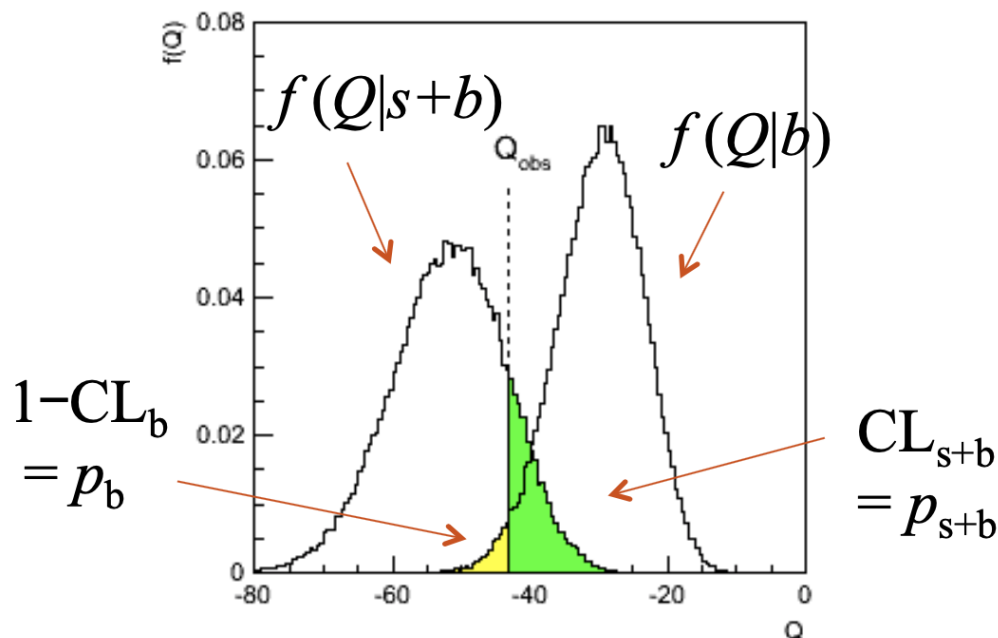


CL_s 方法 (续)

CL_s的做法是，检验不基于通常的 p 值 (CL_{s+b}) 而是基于 CL_{s+b} 与 CL_b 的比值，即定义

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1 - p_b}$$

如果 $CL_s \leq \alpha$ ，
则拒绝 $s + b$ 假设。

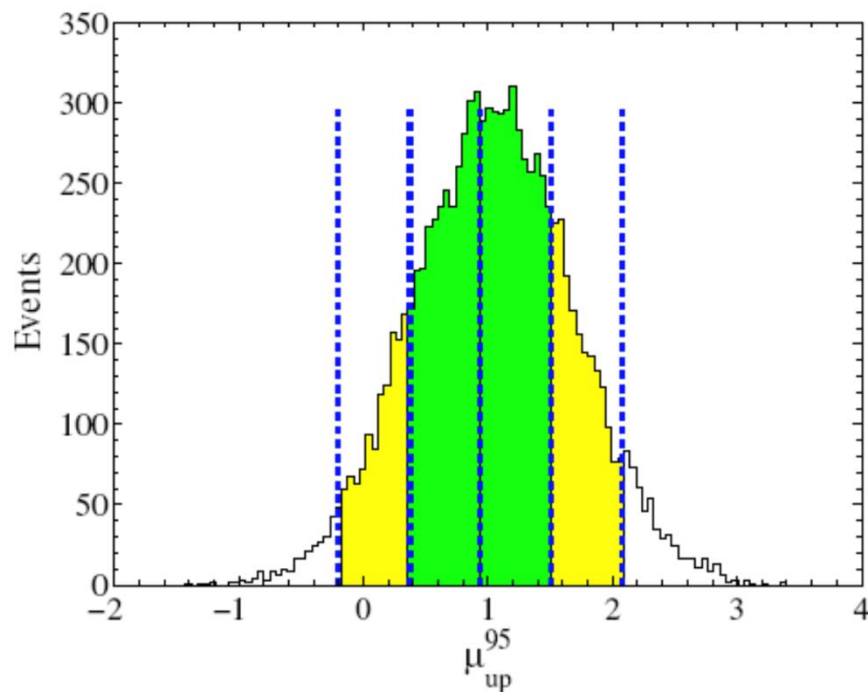


当两个分布离得越来越近时，则提高“有效的” p 值（防止灵敏度太低时得出排除结论）。

对 $\mu = \sigma/\sigma_{\text{SM}}$ 设置上限

对参数 $\mu = \sigma/\sigma_{\text{SM}}$ 实施 CL_s 步骤，得到上限 μ_{up} 。

例如，寻找希格斯粒子，定义 $\mu = m_H/m_H^{\text{SM}}$ 。对于每个给定的 m_H 值，有一个 μ_{up} 的观测值，也可以得到分布 $f(\mu_{\text{up}}|0)$ 。



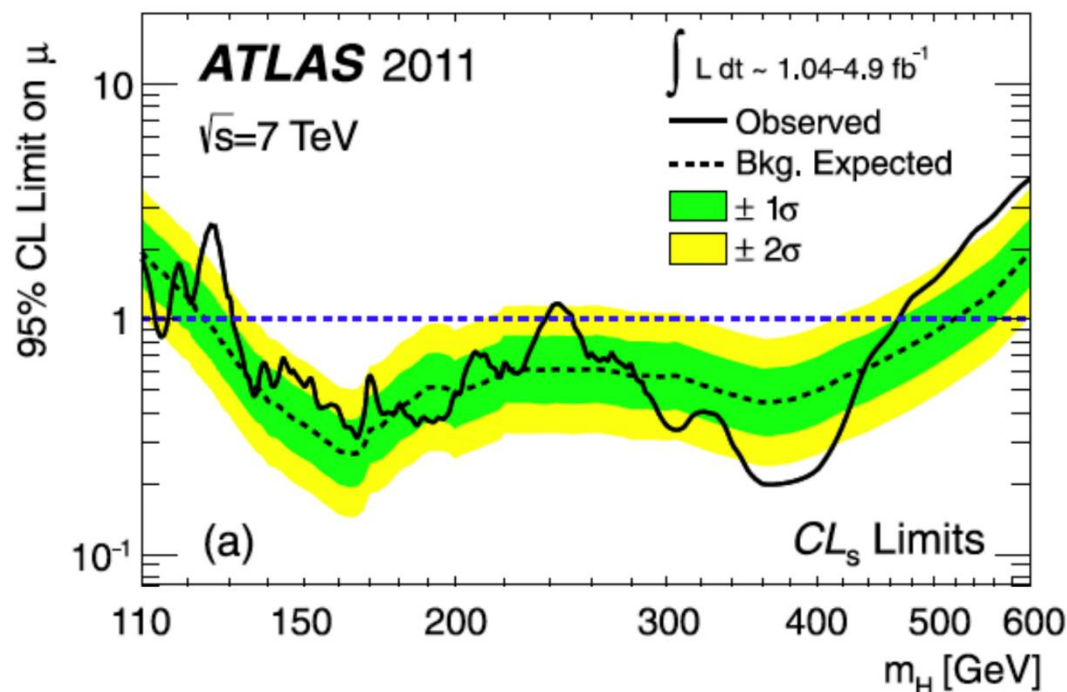
$\pm 1\sigma$ (绿色)和 $\pm 2\sigma$ (黄色)条带
来自简易模拟 (toy MC) ;

竖线来自渐进公式。

如何理解上限图

对每个 m_H 值, 求 μ 的 CL_s 上限, 并确定 $\mu = 0$ 假设下得到的上限 μ_{up} 的分布。

虚线是中位数 μ_{up} , 绿色 (黄色) 条带给出这个分布的 $\pm 1\sigma$ (2σ) 范围。



ATLAS, Phys. Lett. B 710 (2012) 49