实验物理中的统计方法

第九章：置信区间

杨振伟

回顾

- 矩的定义

  用样本的k阶矩代替总体的k阶矩，解方程组

- 矩方法估计量及其应用
- 矩方法与最大似然法和最小二乘法的比较

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
- 贝叶斯上限
- 物理发现或上限的检验统计量

测量结果的表述与含义

实验数据：$x_1, \ldots, x_n$ ➡ 实验目的：估计参数θ ➔ $\hat{\theta}_{\text{obs}}$

还应该给出$\hat{\theta}$的方差，即$\hat{\sigma}^2_{\hat{\theta}}$。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$g(\hat{\vec{\theta}};\vec{\theta})$是多维高斯分布

$\hat{\vec{\theta}}$和$\hat{V}=\widehat{\mathrm{cov}}[\hat{\theta}_i,\hat{\theta}_j]$综合了我们对$g(\hat{\vec{\theta}};\vec{\theta})$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。

可按照约定报道不确定度，而不管概率密度$g(\hat{\theta};\theta)$的形式。

只有在需要对不同实验求平均时，其形式才会发挥作用。

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

$\left[\hat{\theta}_{\mathrm{obs}}-\hat{\sigma}_{\hat{\theta}},\hat{\theta}_{\mathrm{obs}}+\hat{\sigma}_{\hat{\theta}}\right]$ → 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

区间估计

除了参数的“点估计”，还应当报道反映其统计不确定度的一个区间（interval）。

这个区间（interval）最好具有以下性质：

1）客观地描述实验结果；
2）以给定的概率包含参数真值；
3）为给参数下结论提供必要的信息

经常用“±所得估计量的标准差”表示不确定度，
但有时这种做法不适用：
估计值在物理边界附近，例如观测得到的事例率与零一致。

将简要介绍频率论区间（经典置信区间）和贝叶斯区间。

置信区间的定义

设 θ 是总体的一个待估参数, 其参数空间为 Θ, $X_1,\cdots,X_n$ 为来自总体的样本。对给定的实数 $\alpha\ (0<\alpha<1)$，假设有两个统计量 $T_1(X_1,\cdots,X_n)$, $T_2(X_1,\cdots,X_n)$，若对任意$\theta\in\Theta$，有

$$P(T_1\le\theta\le T_2)\ge 1-\alpha$$

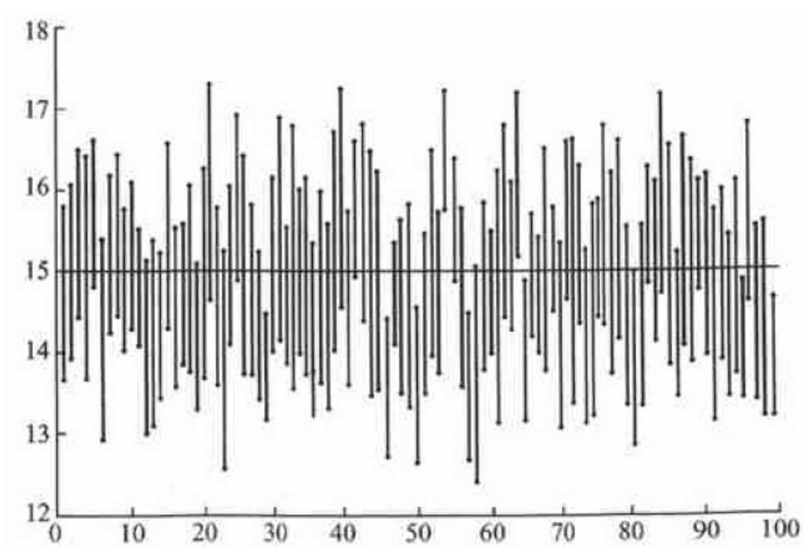
则称随机区间 $[T_1,T_2]$ 为 θ 的置信水平为 $1-\alpha$ 的置信区间或区间估计。

T_1称为(双侧)置信下限，T_2称为(双侧)置信上限。

若满足 $P(T_1\le\theta\le T_2)=1-\alpha$，则称$[T_1,T_2]$ 为 θ 的 $1-\alpha$ 同等置信区间。连续随机变量一般要求取等号。

置信区间的含义

置信水平 $1-\alpha$ 的频率解释：在大量重复估计置信区间时，每次得到的样本观测值不同，从而每次得到的区间也不同。对每个区间而言，要么包含未知参数 θ 的真值, 要么不包含。平均而言，在这些大量的区间中，至少有$100(1-\alpha)\%$个包含θ。



关于置信区间的几点说明

- 置信区间的长度 $T_2 - T_1$ 反映了估计精度
  - $T_2 - T_1$ 越小，估计精度越高

- α 反映了估计的可靠度，α 越小, 越可靠
  - α 越小, $1 - \alpha$ 越大，估计的可靠度越高；
    但此时 $T_2 - T_1$ 往往增大，因而估计精度降低

区间估计三部曲

- 选取枢轴量
- 由分位点的定义建立不等式
- 求解不等式得到下界a和上界b

枢轴量

—— 仅含一个待估参数的
样本的连续函数

枢轴量的分布不依赖于未知参数。

例：正态分布均值的区间估计

- 已知 $X \sim N(\mu, 1)$。设样本容量为$n = 5$：

$$(X_1, X_2, X_3, X_4, X_5)$$

求μ的置信水平为$1-\alpha = 95\%$的置信区间

$$\bar{X} \sim N\left(\mu, \frac{1}{5}\right) \Rightarrow \frac{\bar{X}-\mu}{\sqrt{1/5}} \sim N(0,1)$$

枢轴量

取 $\alpha = 0.05$，　查表或计算得 $Z_{\alpha/2} = 1.96$

$$P\left(\left|\frac{\bar{X}-\mu}{\sqrt{1/5}}\right| \geq 1.96\right) = 0.05$$

称随机区间 $\left(\bar{X} - 1.96\sqrt{1/5},\quad \bar{X} + 1.96\sqrt{1/5}\right)$

为未知参数 μ 的置信度为0.95的置信区间。

若测得 一组样本值, 算得 $\bar{X} = 1.86$,

代入得到区间 $(\mathbf{0.983}, \mathbf{2.737})$

此时, 有:

$$P\left(\bar{X} - 1.96\sqrt{1/5} \le \mu \le \bar{X} + 1.96\sqrt{1/5}\right)$$
$$= P(0.983 \le \mu \le 2.737) = 0.95$$

关于置信区间的构造的两点说明：

- 满足置信度要求的 a 与 b 通常不唯一。若有可能，应选平均长度 $E(T_2 - T_1)$ 达到最短的a与b。枢轴量G满足对称分布时通常容易实现。
- 实际中，选平均长度 $E(T_2 - T_1)$ 尽可能短的a与b往往不易实现。因此，常选择 a与b，使得两个尾部概率各为$\alpha/2$，即$P(G < a) = P(G > b) = \alpha/2$。这样的置信区间称为等尾置信区间。在$G$为偏态分布时常用这种方法。

频率论置信区间

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

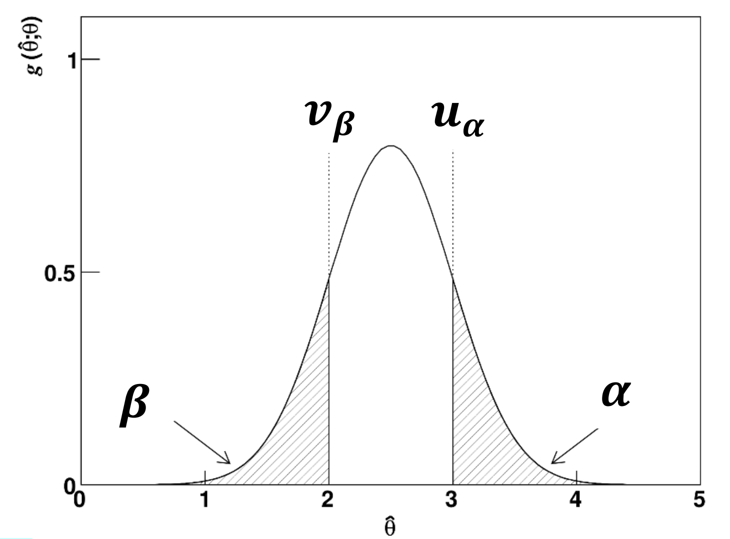
为了正确表述结果，还需要知道 $\hat{\theta}$ 的分布形式 $g(\hat{\theta};\theta)$。

首先指定“上下尾部的概率”，如：$\alpha=\beta=0.05$，

然后找到函数关系 $u_\alpha(\theta)$和$v_\beta(\theta)$，使得

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

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



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

参数置信带的定义

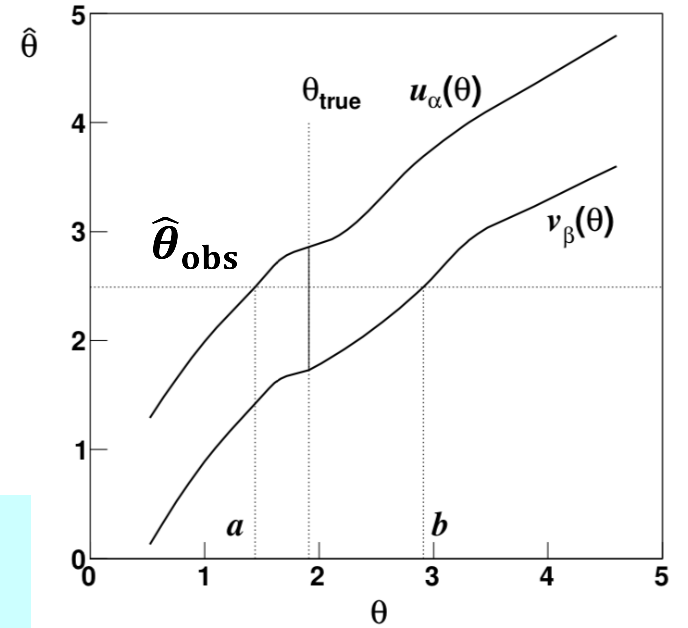
在 $u_\alpha(\theta)$和$v_\beta(\theta)$之间的区域称为置信带。

无论θ为何值，在置信带找到$\hat{\theta}_{\mathrm{obs}}$的概率为

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

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

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



不等式

$$\hat{\theta} \ge u_\alpha(\theta), \quad \hat{\theta} \le v_\beta(\theta), \iff a(\hat{\theta}) \ge \theta, \quad b(\hat{\theta}) \le \theta$$

$$P\left(\hat{\theta} \ge u_\alpha(\theta)\right), \quad P\left(\hat{\theta} \le v_\beta(\theta)\right) \iff P\left(a(\hat{\theta}) \ge \theta\right), \quad P\left(b(\hat{\theta}) \le \theta\right)$$

参数的置信区间确定

根据置信带的定义，有不等式

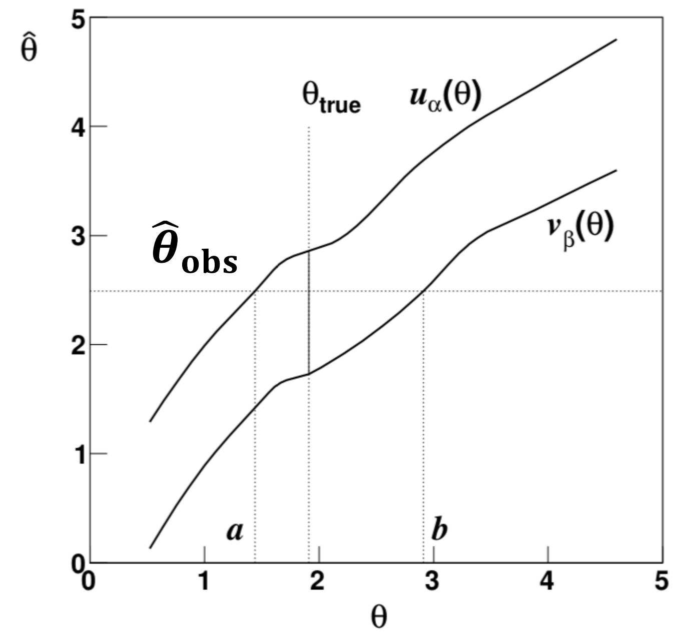
$$a(\hat{\theta}) \geq \theta, \qquad b(\hat{\theta}) \leq \theta$$

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

或者合并成

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

在真值θ未知的情况下，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。



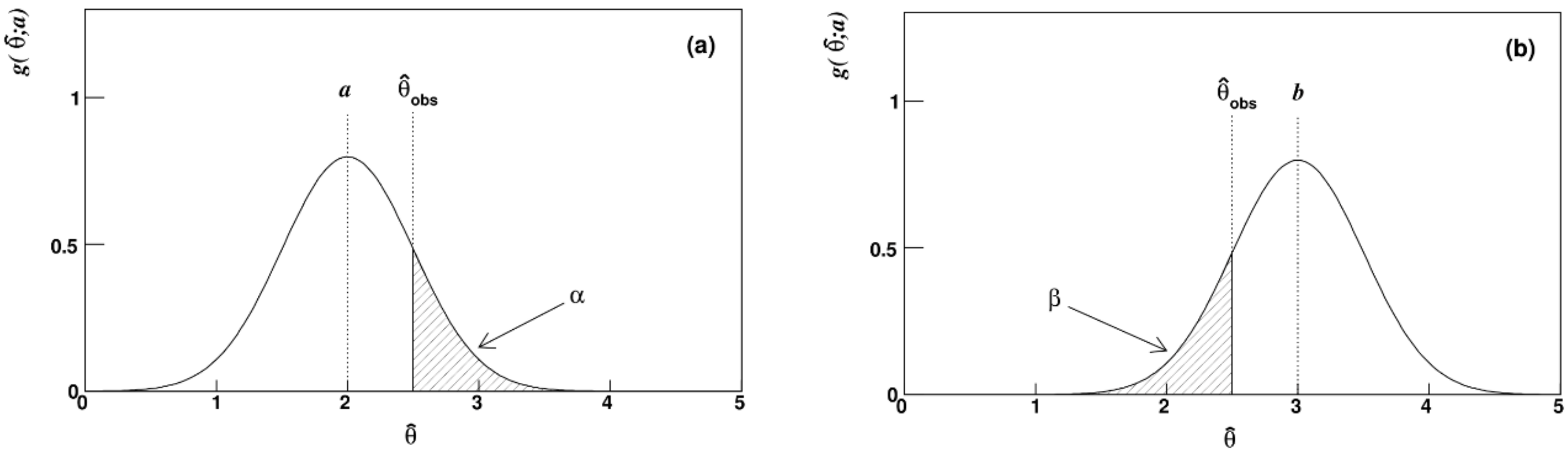
置信区间的求解

求解置信区间 $[a, b]$ 的诀窍是求解下列方程：

$$\alpha = \int_{u_\alpha(\theta)}^{\infty} g(\hat{\theta};\theta)\mathrm{d}\hat{\theta} = \int_{\hat{\theta}_{\mathrm{obs}}}^{\infty} g(\hat{\theta};a)\mathrm{d}\hat{\theta}$$

$$\beta = \int_{-\infty}^{v_\beta(\theta)} g(\hat{\theta};\theta)\mathrm{d}\hat{\theta} = \int_{-\infty}^{\hat{\theta}_{\mathrm{obs}}} g(\hat{\theta};b)\mathrm{d}\hat{\theta}$$

$$a \equiv u_\alpha^{-1}(\hat{\theta}_{\mathrm{obs}}), \quad b \equiv v_\beta^{-1}(\hat{\theta}_{\mathrm{obs}})$$



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

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

参数置信区间含义

注意，得到的区间是随机的，而真值 θ 是一个未知常数。

我们经常将置信区间 $[a, b]$ 报道为$\hat{\theta}_{-c}^{+d}$，即

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

那么 $\hat{\theta} = 80.25_{-0.25}^{+0.31}$ 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：

重复样本容量相同的实验多次，每次按相同方法构造置信区间，那么，覆盖 θ 的置信区间占比为 $1 - \alpha - \beta$。

高斯分布估计量的置信区间

如果$\hat{\theta}$服从

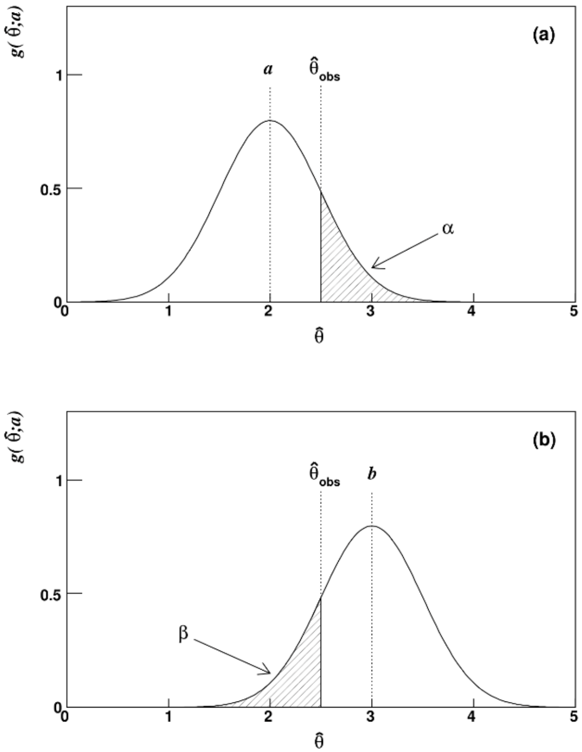
$$g(\hat{\theta};\theta)=\frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}}\exp\left(-\frac{(\hat{\theta}-\theta)^2}{2\sigma_{\hat{\theta}}^2}\right)$$

为了找到 θ 的置信区间，解下列方程

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

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

→ 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

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

是标准高斯的累积分布函数，可以证明

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

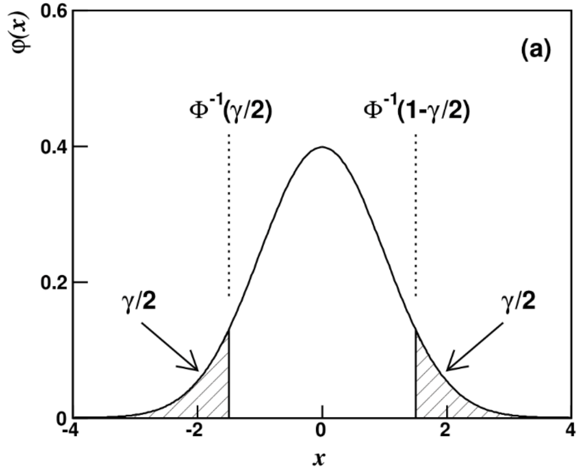
这里Φ^{-1}给出标准高斯的分位数（可以调用ROOT中的函数Double_t TMath::NormQuantile(Double_t p)计算)

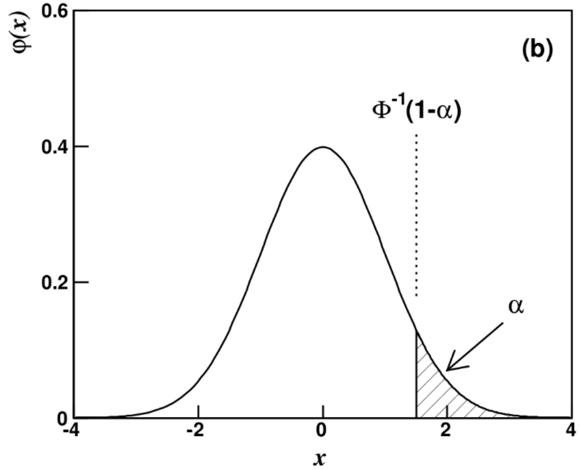
$$\boldsymbol{\Phi^{-1}(1-\alpha)}, \qquad \boldsymbol{\Phi^{-1}(1-\beta)}$$

给出a与b离$\hat{\theta}$有多少倍标准差。

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，$\hat{\nu}=n$，估计值 $\hat{\nu}_{\rm obs}=n_{\rm obs}$，

$$P(n;\nu)=\frac{\nu^n}{n!}e^{-\nu},\qquad n=0,1,\ldots$$

对于固定的 α，β，由于 ν 只能取离散值，用来确定置信带的函数$u_\alpha(\nu)$和$v_\beta(\nu)$不一定对任意 ν 都存在，即不一定能找到整数 $\hat{\nu}$，满足$P\big(\hat{\nu}\ge u_\alpha(\nu)\big)=\alpha$，$P\left(\hat{\nu}\le v_\beta(\nu)\right)=\beta$。此时，可类似于前面提到的如下处理，

$$\alpha=P\left(\hat{\theta}\ge u_\alpha(\theta)\right)\qquad \alpha=P\big(a(\hat{\theta})\ge\theta\big)$$
$$\beta=P\left(\hat{\theta}\le v_\beta(\theta)\right)\Longleftrightarrow \beta=P\big(b(\hat{\theta})\le\theta\big)$$

考虑求解

$$\alpha=P(\hat{\nu}\ge\hat{\nu}_{\rm obs};a)=1-\sum_{n=0}^{n_{\rm obs}-1}\frac{a^n}{n!}e^{-a},$$
$$\beta=P(\hat{\nu}\le\hat{\nu}_{\rm obs};b)=\sum_{n=0}^{n_{\rm obs}}\frac{b^n}{n!}e^{-b},$$

$\Rightarrow$ a 与 b

a是ν的假设值，使得：$P(\hat{\nu}\ge\hat{\nu}_{\rm obs})=\alpha$

b是ν的假设值，使得：$P(\hat{\nu}\le\hat{\nu}_{\rm obs})=\beta$

泊松分布均值的置信区间确定

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

这里F_{χ^2}是自由度为n_d的卡方变量的累积分布函数。

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

$\Rightarrow$

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

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）

泊松求和与卡方分布积分关系的推导

$$\sum_{n=0}^{m}\frac{\nu^n}{n!}e^{-\nu}=\int_{2\nu}^{\infty}f_{\chi^2}\big(z;n_d=2(m+1)\big)\mathrm{d}z=1-F_{\chi^2}\big(2\nu;n_d=2(m+1)\big)$$

已知 $f_{\chi^2}(z;n)=\frac{1}{2^{n/2}\Gamma(n/2)}z^{\frac{n}{2}-1}e^{-\frac{z}{2}}$

则 $f_{\chi^2}\big(z;2(m+1)\big)=\frac{1}{2^{m+1}\Gamma(m+1)}z^m e^{-\frac{z}{2}}=\frac{1}{2}\frac{(z/2)^m}{m!}e^{-\frac{z}{2}}$

$$\int_{2\nu}^{\infty}f_{\chi^2}\big(z;n_d=2(m+1)\big)\mathrm{d}z=\int_{2\nu}^{\infty}\frac{1}{2}\frac{(z/2)^m}{m!}e^{-\frac{z}{2}}\mathrm{d}z=\int_{\nu}^{\infty}\frac{x^m}{m!}e^{-x}\mathrm{d}x$$

对 $\int_{\nu}^{\infty}\frac{x^m}{m!}e^{-x}\mathrm{d}x$ 反复用分部积分，可得

$$\int_{\nu}^{\infty}\frac{x^m}{m!}e^{-x}\mathrm{d}x=\frac{\nu^m}{m!}e^{-\nu}+\frac{\nu^{m-1}}{(m-1)!}e^{-\nu}+\cdots+\frac{\nu^1}{1!}e^{-\nu}+\frac{\nu^0}{0!}e^{-\nu}=\sum_{n=0}^{m}\frac{\nu^n}{n!}e^{-\nu}$$

泊松分布均值的置信水平上限值

重要特例: $n_{\text{obs}} = 0$

$$\boldsymbol{\beta} = \sum_{n=0}^{0} \frac{b^n}{n!} e^{-b} = e^{-b}$$

$$b = -\ln\beta$$

对于置信水平 $1-\beta = 95\%$ 的上限,

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

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

反用检验得到置信区间

参数 θ 的置信区间可通过定义假设参数值 θ 的检验得到：

指定 θ “不偏好的”数据值（临界域），使得 P(数据在临界域) $\leq \gamma$，γ为预先指定值，如 0.05或0.1。

反用检验来定义置信区间：

在置信水平 $1-\gamma$ 的检验中不被拒绝的 θ 值的集合。

置信区间包含 θ 真值的概率大于等于 $1-\gamma$。

等价于置信带的构造；置信带是检验的接受域。

反转正态分布均值检验得到置信区间

设$X_1, \dots, X_n$是来自$N(\mu, \sigma^2)$的独立同分布（iid）。考虑检验$H_0: \mu = \mu_0$对$H_1: \mu \neq \mu_0$。对于固定的显著性水平α，一个合理的检验的拒绝域为

$$\left\{\vec{x}: |\bar{x} - \mu_0| > \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right\}$$

当$|\bar{x} - \mu_0| \leq \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$时，$H_0$被接受。

$$P\left(|\bar{x} - \mu_0| \leq \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \middle| \mu = \mu_0\right)$$

$$= P\left(\bar{x} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \middle| \mu = \mu_0\right) = 1 - \alpha$$

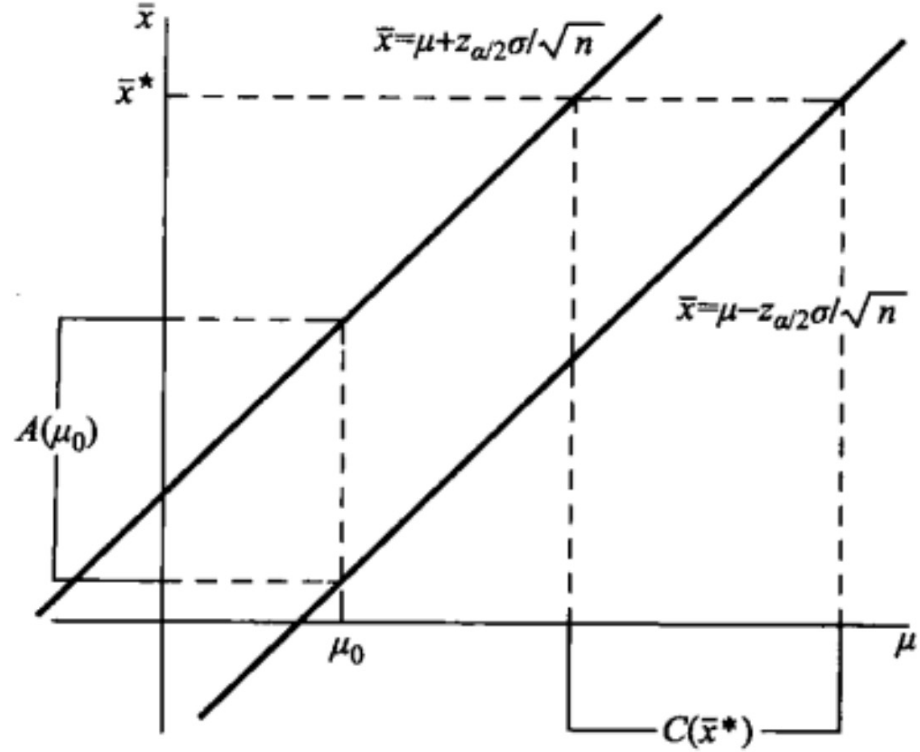
这对所有μ都成立，即

$$P\left(\bar{x} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

置信区间

置信区间与检验的接受域的关系

- 假设检验：固定参数并询问什么样本值（接受域A）与该固定值相符
- 置信区间：固定样本值并询问什么参数值（置信区间C）使该样本值好像最合理



置信区间与p值的关系

等价地，我们可以考虑每个假设的 θ 值的显著性检验，得到相应的 p值，p_θ：

如果 $p_\theta < \gamma$，则拒绝 θ。

置信水平 $\mathrm{CL} = 1 - \gamma$ 的置信区间包含没有被拒绝的 θ 值。

例如，θ 的上限（upper limit）是满足 $p_\theta \geq \gamma$ 的最大 θ 值。

实际应用中，令 $p_\theta = \gamma$ 并求解 θ 即得到最大 θ 值。

中心置信区间与单侧置信区间

有时，单独指定 α 或 β

→ 单边置信区间 (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，它仅仅意味着 $\alpha = \beta$。

粒子与核物理的不确定度惯例是：68.3%的中心置信区间。

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 $\vec{\theta} = (\theta_1, \dots, \theta_n)$：

$$\lambda(\vec{\theta}) = \frac{L(\vec{\theta})}{L(\hat{\vec{\theta}})} \qquad 0 \le \lambda(\vec{\theta}) \le 1$$

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

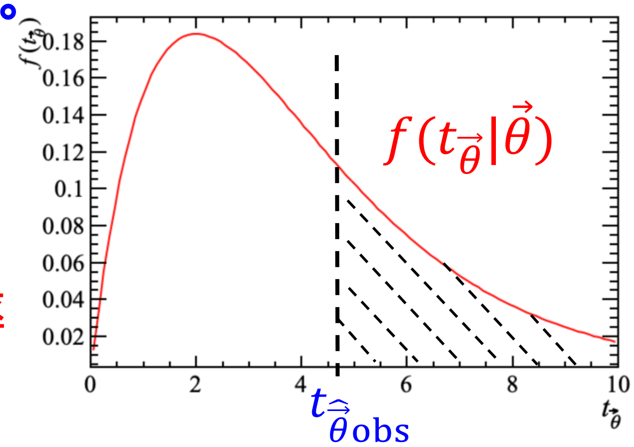
$$t_{\vec{\theta}} = -2\ln\lambda(\vec{\theta})$$

$t_{\vec{\theta}}$ 大 ➔ 数据与假设的 $\vec{\theta}$ 符合得更差。

因此，$\vec{\theta}$ 的 p值为：

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

需要知道概率密度 $f(t_{\vec{\theta}}|\vec{\theta})$



从Wilks定理估计置信区间

Wilks定理（满足大样本极限和其他一些条件）：

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

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Wilks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 $t_{\vec{\theta}}$:

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

注意，$t_{\vec{\theta}}$ 还可以表示为:

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

$\Rightarrow$ $\ln\vec{\theta}$ 可用 $F^{-1}_{\chi^2(n)}$ 表示

从Wilks定理估计置信区间（续）

在 $\vec{\theta}$ 空间，置信区域的边界为

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

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

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

所以，68.3%置信水平的置信区间由下式确定：

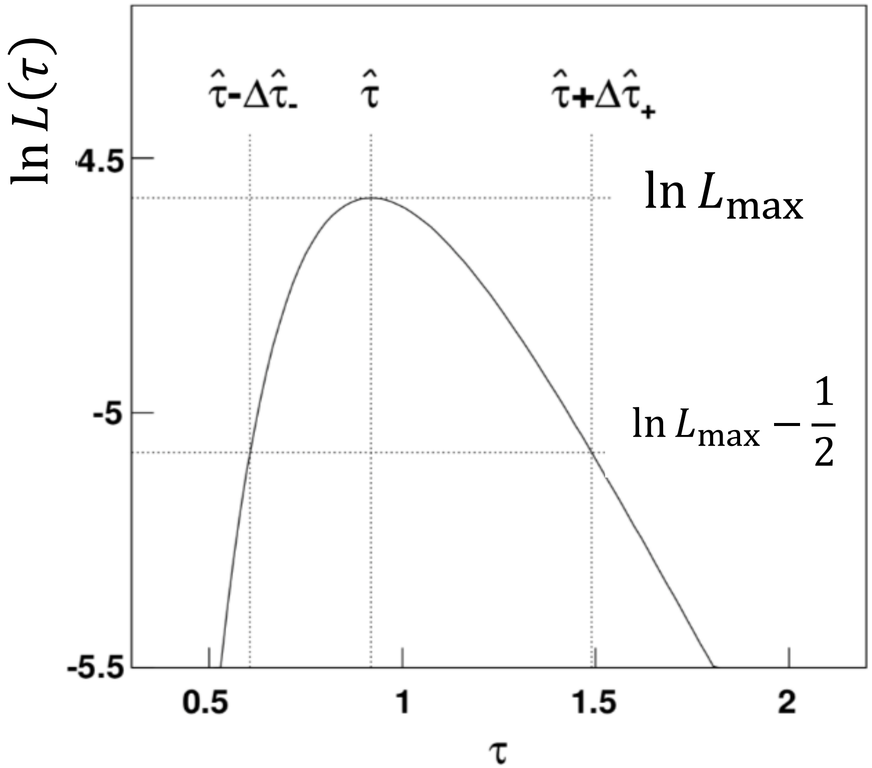
$$\ln L(\theta) = \ln L(\hat{\theta}) - \frac{1}{2}$$

这与求估计量标准差的方法一样，即

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

例：由 $\ln L(\theta)$ 求置信区间

对于 $n=1$个参数， $\mathrm{CL}=68.3\%$ ， $Q_\alpha = F^{-1}_{\chi^2(n)}(1-\alpha)=1$。



在指数函数例子中，只有 $n=5$个观测值时：

可以报道ML估计值，以利用 $\ln L_{\max}-\frac{1}{2}$ 得到的近似置信区间为“不对称误差棒”：

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

多维参数置信区间

当参数个数增大时，由 $Q_\alpha = F^{-1}_{\chi^2(n)}(1-\alpha)$ 确定的置信区间的置信水平 $\mathrm{CL} = 1-\alpha$ 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 $\mathrm{CL} = 1 - \alpha$，Q_α 随着 n 增大而增大。

$\boldsymbol{1 - \alpha}$	$\boldsymbol{Q_\alpha}$				
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
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

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得到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，我们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 $n \sim poi(s+b)$，假设 $b = 4.5$，$n_{\mathrm{obs}} = 5$。求置信水平 $\mathrm{CL} = 95\%$ 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 $P(n \le n_{\mathrm{obs}}; s, b)$，求解下式得 $\mathrm{CL} = 1 - \alpha$ 的上限 s_{up}

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

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

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

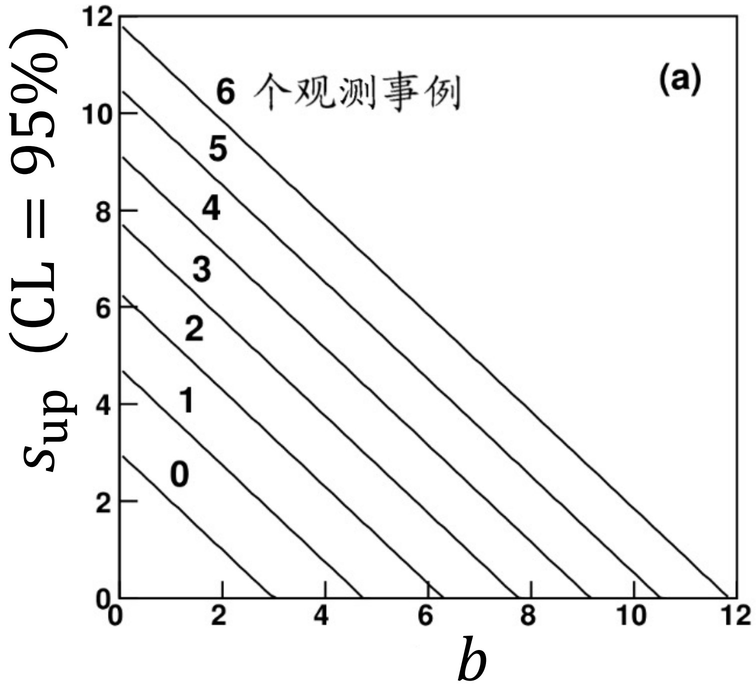
泊松参数s的上限：频率论

对于事例数 n_{obs} 很小的涨落，

$$s_{\mathrm{up}} = \frac{1}{2} F^{-1}_{\chi^2}\big(1-\alpha; 2(n_{\mathrm{obs}}+1)\big) - b$$

可以给出负的结果。

即，置信区间可能是空的。



给定 n_{obs}，95% 置信水平下 s_{up} 随 本底数 b 的变化关系。例如，如果 $n_{\mathrm{obs}} = 1$，$b = 6$，那么 $s_{\mathrm{up}} < 0$。

$\alpha \geq 1 - F_{\chi^2}\big(2b; 2(n_{\mathrm{obs}}+1)\big)$ 时，$s_{\mathrm{up}} \leq 0$。

物理边界附近的极限

假设 $b=2.5$，观测到 $n_{\text{obs}}=0$。

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

$$F_{\chi^2}^{-1}(x;2)=-2\ln(1-x)\quad\Rightarrow\quad s_{\text{up}}=-\ln(\alpha)-b$$

如果选择 $\text{CL}=1-\alpha=0.9$，

$$\Rightarrow\quad s_{\text{up}}=-\ln(0.1)-2.5=-0.197$$

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 ➔ $s_{\mathrm{up}} = 0.496$。

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

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

实际情况：$b = 2.5$时，n 的典型泊松涨落至少是 $\sqrt{2.5} = 1.6$。上限怎么会这么小呢?

零信号假设（$s = 0$）下的平均上限：

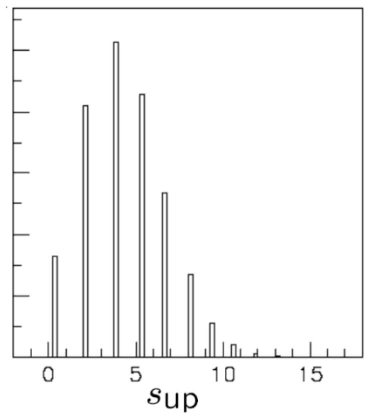
取 $b = 2.5$，CL = 0.95

$n_{\mathrm{obs}} \sim Poi(2.5)$，对于给定的 n_{obs}，上限为

$$s_{\mathrm{up}}(n_{\mathrm{obs}}) = \frac{1}{2} F_{\chi^2}^{-1}(0.95; 2(n_{\mathrm{obs}} + 1)) - 2.5$$

平均上限：$\bar{s}_{\mathrm{up}} = \sum_{n_{\mathrm{obs}}=0}^{\infty} p(n_{\mathrm{obs}}) s_{\mathrm{up}}(n_{\mathrm{obs}}) = 4.44$

$n_{\mathrm{obs}} = 0$ 的概率为 8.2%



$b = 2.5, s = 0$时 CL = 95% 上限 s_{up} 的分布。

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：

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

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：

$$0.95 = \int_{-\infty}^{s_{\mathrm{up}}} p(s|n)\,\mathrm{d}s$$

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：均匀分布的 s 验前概率

将泊松似然函数和均匀分布的验前概率代入贝叶斯定理：

$$p(s|n) \propto \frac{(s+b)^n}{n!} e^{-(s+b)} \quad (s \geq 0)$$

归一化到单位面积：

$$p(s|n) = \frac{(s+b)^n e^{-(s+b)}}{\Gamma(b, n+1)}$$

分母

$= \int_0^\infty (s+b)^n e^{-(s+b)} \mathrm{d}s$

$= \int_b^\infty t^n e^{-t} \, \mathrm{d}t$

$= \Gamma(b, n+1)$

上不完全Γ函数

求解

$$1 - \alpha = \int_0^{s_{\mathrm{up}}} p(s|n) \, \mathrm{d}s$$

确定上限 s_{up}。

$$s_{\mathrm{up}} = \frac{1}{2} F_{\chi^2}^{-1}\big(p; 2(n+1)\big) - b$$

其中 $p = 1 - \alpha\big(1 - F_{\chi^2}[2b, 2(n+1)]\big)$

对特殊情况 $b = 0$，可得 $p = 1 - \alpha$，验前概率均匀的贝叶斯上限从数值上与单侧频率论结果相同（“巧合”）。

推导过程

$$1-\alpha=\int_0^{s_{\text{up}}} p(s|n)\,\mathrm{d}s \Rightarrow s_{\text{up}} \Leftrightarrow \alpha=\int_{s_{\text{up}}}^{\infty} p(s|n)\,\mathrm{d}s=\frac{\int_{s_{\text{up}}}^{\infty}\frac{(s+b)^n}{n!}e^{-(s+b)}\mathrm{d}s}{\int_0^{\infty}\frac{(s+b)^n}{n!}e^{-(s+b)}\mathrm{d}s}$$

分部积分：$\int_0^{\infty}\frac{(s+b)^n}{n!}e^{-(s+b)}\mathrm{d}s=\int_b^{\infty}\frac{t^n}{n!}e^{-t}\mathrm{d}t=\sum_{k=0}^{n}\frac{b^k}{k!}e^{-b}$

分部积分：$\int_{s_{\text{up}}}^{\infty}\frac{(s+b)^n}{n!}e^{-(s+b)}\mathrm{d}s=\int_{s_{\text{up}}+b}^{\infty}\frac{t^n}{n!}e^{-t}\mathrm{d}t=\sum_{k=0}^{n}\frac{\left(s_{\text{up}}+b\right)^k}{k!}e^{-(s_{\text{up}}+b)}$

利用 $\boldsymbol{\sum_{k=0}^{m}\frac{\nu^k}{k!}e^{-\nu}=1-F_{\chi^2}(2\nu;n_d=2(m+1))}$

$$\Rightarrow \alpha=\frac{\sum_{k=0}^{n}\frac{\left(s_{\text{up}}+b\right)^k}{k!}e^{-(s_{\text{up}}+b)}}{\sum_{k=0}^{n}\frac{b^k}{k!}e^{-b}}=\frac{1-F_{\chi^2}\left(2\left(s_{\text{up}}+b\right);2(n+1)\right)}{1-F_{\chi^2}\left(2b;2(n+1)\right)}$$

$$\Rightarrow F_{\chi^2}\left(2\left(s_{\text{up}}+b\right);2(n+1)\right)=1-\alpha\left(1-F_{\chi^2}\left(2b;2(n+1)\right)\right)\equiv p$$

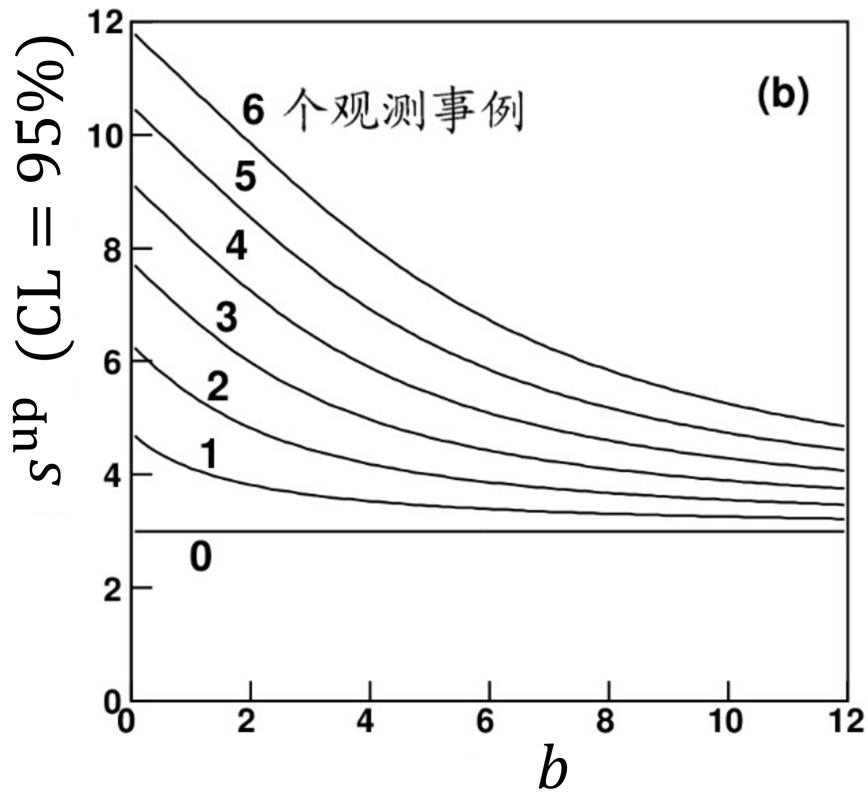
$$\Rightarrow 2\left(s_{\text{up}}+b\right)=F_{\chi^2}^{-1}\left(p;2(n+1)\right) \Rightarrow s_{\text{up}}=\frac{1}{2}F_{\chi^2}^{-1}\left(p;2(n+1)\right)-b$$

贝叶斯上限：均匀分布的 s 验前概率

对于 $b > 0$，贝叶斯上限处处都大于频率论上限。

1）贝叶斯上限 $s_{\text{up}} > 0$

2）$n = 0$ 时，贝叶斯上限与 b 无关



$n = 0$时：

$$\alpha = \frac{\sum_{k=0}^{n} \frac{(s_{\text{up}}+b)^k}{k!} e^{-(s_{\text{up}}+b)}}{\sum_{k=0}^{n} \frac{b^k}{k!} e^{-b}}$$

$$= \frac{e^{-(s_{\text{up}}+b)}}{e^{-b}} = e^{-s_{\text{up}}}$$

$$\Rightarrow s_{\text{up}} = -\ln(\alpha) > 0$$

对于经典上限，$n = 0$ 时：

$$s_{\text{up}} = -\ln(\alpha) - b$$

$\alpha \geq e^{-b}$ 时，$s_{\text{up}} \leq 0$

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

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

其中 $I(\vec{\theta})$ 是费舍尔信息矩阵：

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

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

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

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

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

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

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

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

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

μ：强度参数

通常还有辅助测量，用来约束本底或形状参数：

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

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

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

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）

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

$\hat{\hat{\vec{\theta}}}$：对指定的 μ 值使 L 最大

$\hat{\mu}, \hat{\vec{\theta}}$：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：

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

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

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 $\mathcal{O}(1/\sqrt{N})$ 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。

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

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：

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

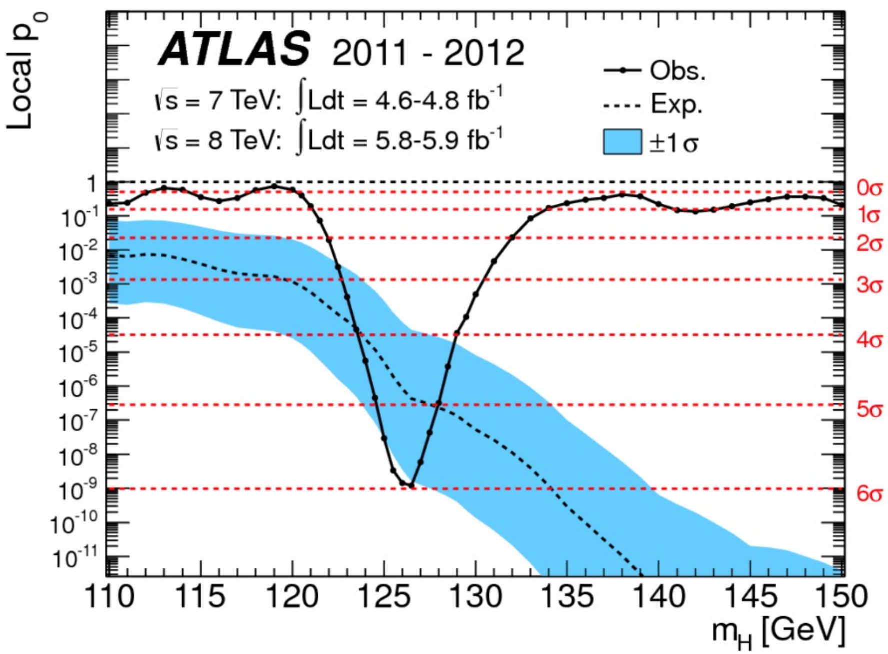
$\mu = 0$ 假设的 p值为

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

因此，物理发现的显著性 Z 为

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

ATLAS, Phys. Lett. B 716 (2012) 1



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

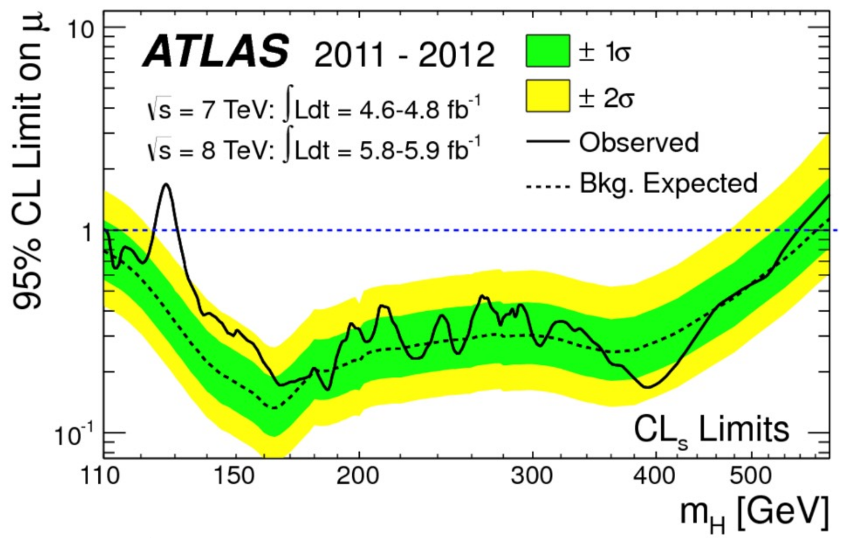
$$f(q_\mu|\mu') = \Phi\left(\frac{\mu'-\mu}{\sigma}\right)\delta(q_\mu) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_\mu}}e^{-\frac{1}{2}\left(\sqrt{q_\mu}-\frac{(\mu-\mu')}{\sigma}\right)^2}$$

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

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

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

ATLAS, Phys. Lett. B 716 (2012) 1



注意：这是 $\mathrm{CL_s}$ 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：

$n \sim Poi(s+b)$ （主测量或“搜索”测量）
$m \sim Poi(\tau b)$ （控制测量，τ 已知）

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)：

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

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： $\hat{s} = n - m/\tau$ $\hat{b} = m/\tau$ $\hat{\hat{b}}(0) = \dfrac{n+m}{(1+\tau)}$

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

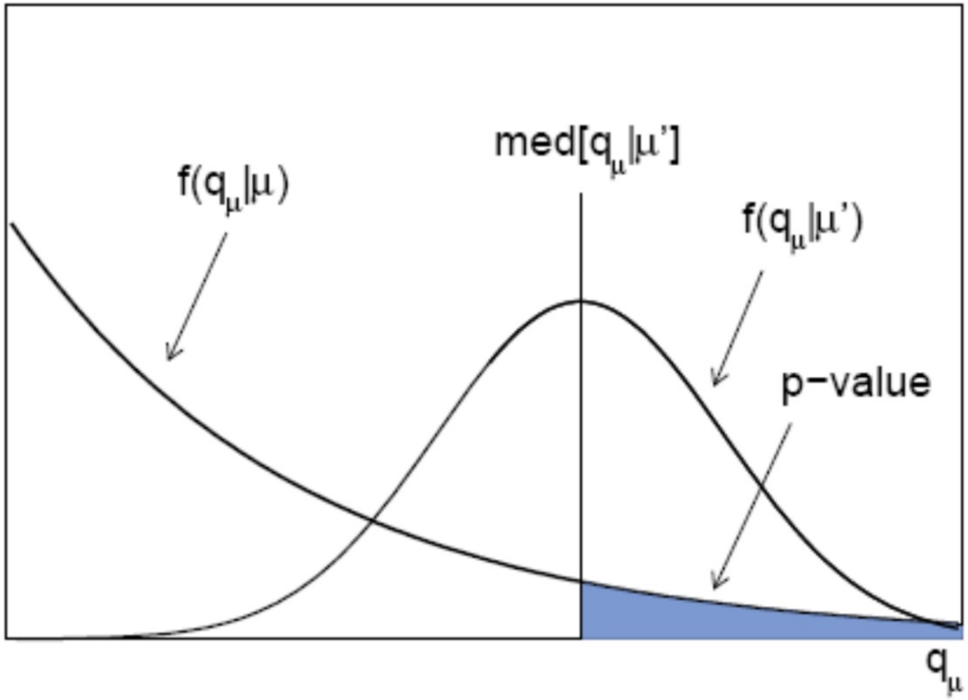
$$Z = \sqrt{q_0} = \sqrt{-2\ln\lambda(0)}$$
$$= \left[-2\left(n\ln\left[\frac{n+m}{(1+\tau)n}\right] + m\ln\left[\frac{\tau(n+m)}{(1+\tau)m}\right]\right)\right]^{1/2}$$

$n > \hat{b}$

$Z = 0$ $n \le \hat{b}$

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) $\dfrac{s}{\sqrt{b}}$

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

2. 本底 b 不确定度为 σ_b 的发现灵敏度：

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

(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：

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

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 $p < 2.9\times10^{-7}$。

要表征对发现的灵敏度，可以给出给定 s 的假定下期待的（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，$n \to x \sim N(\mu,\sigma^2)$，$\mu = s+b$，$\sigma^2 = s+b$。

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

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

因此，拒绝 $s=0$ 的显著性为：

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

假定信号率为 s时，期待的（中位数）显著性为：

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

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：

此时没有冗余参数

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

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

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 $\mathrm{median}[Z|s]$，令 $n \to s + b$ （即Asimov数据集）：

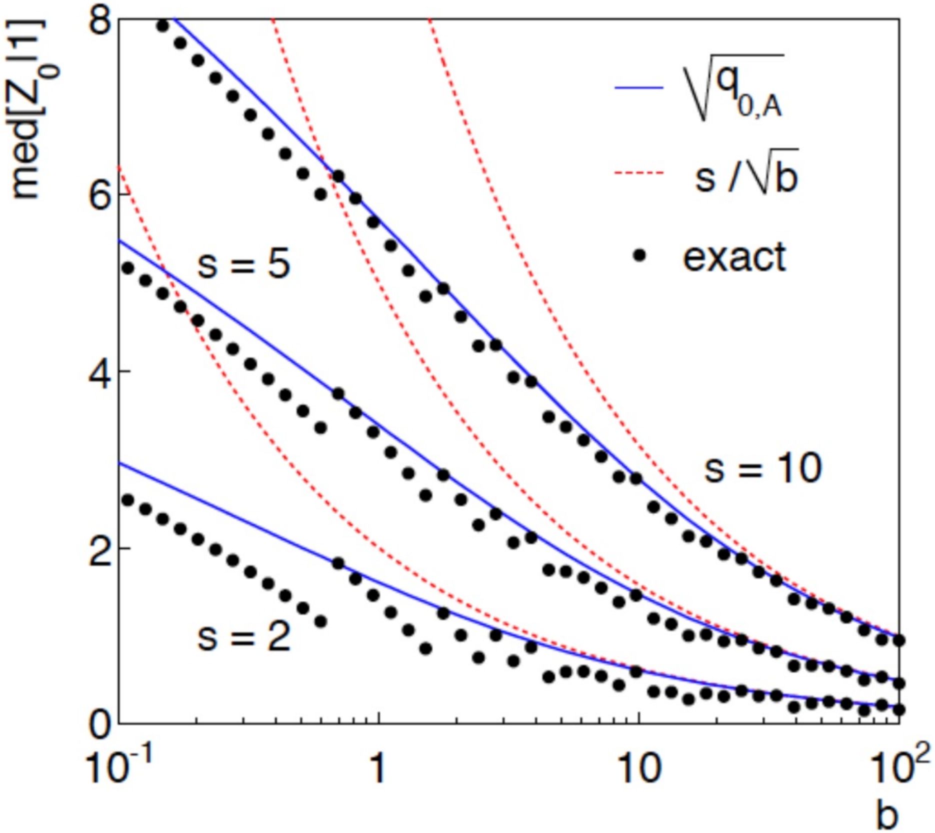
$$Z_A = \sqrt{2((s+b)\ln(1 + s/b) - s)}$$

$s \ll b$ 时，回到 $s/\sqrt{b}$。

$s=0$假设的中位数显著性的对比

$n \sim Poi(s+b)$，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。

$\sqrt{q_{0,\mathrm{A}}}$：
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

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

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：$n \to s + b$，$m \to \tau b$

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

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

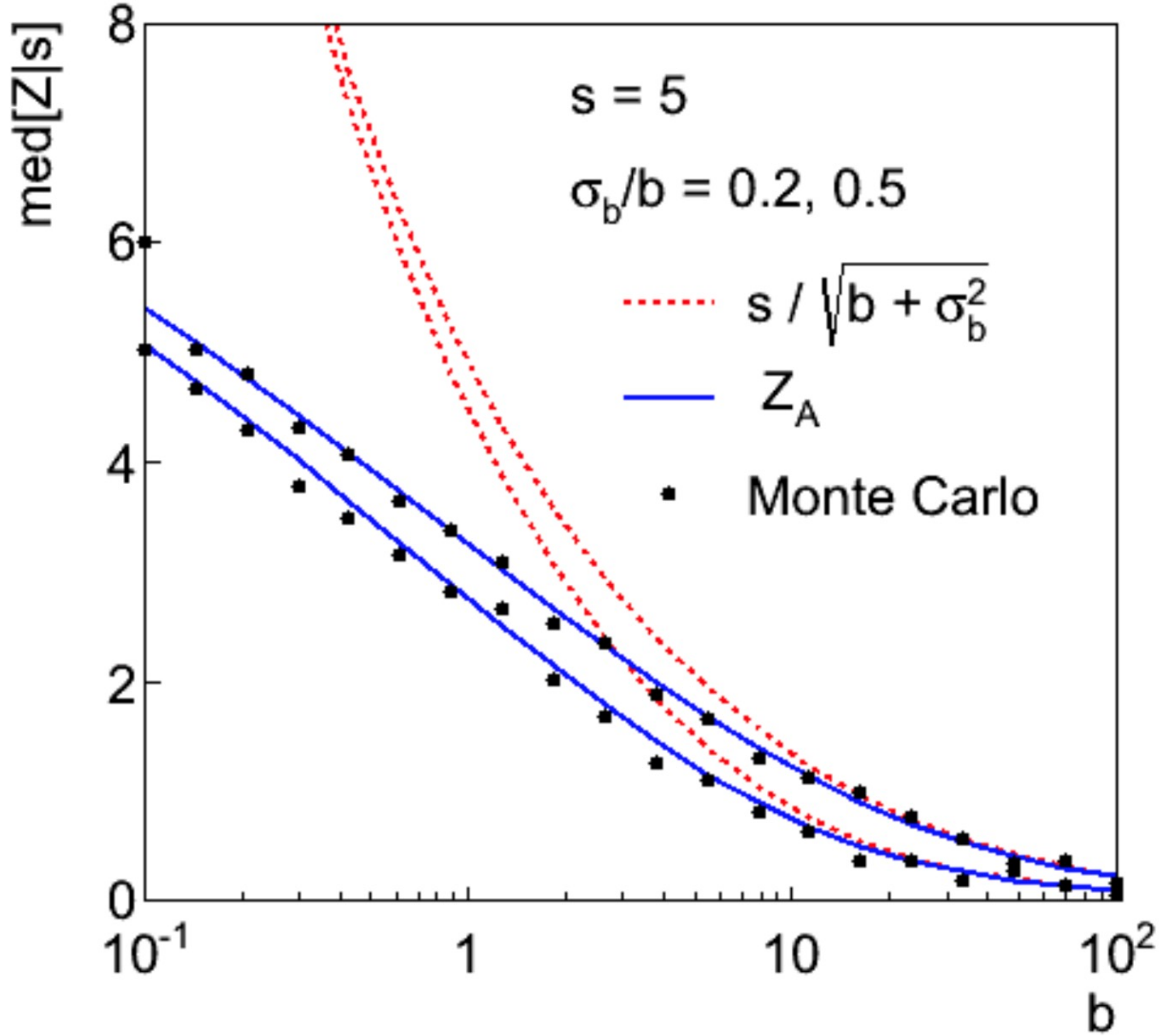
$$Z_A = \left[2\left((s+b)\ln\left[\frac{(s+b)\left(b+\sigma_b^2\right)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right] \right) \right]^{1/2}$$

按 s/b 和 $\sigma_b^2/b(=1/\tau)$ 展开得到：

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

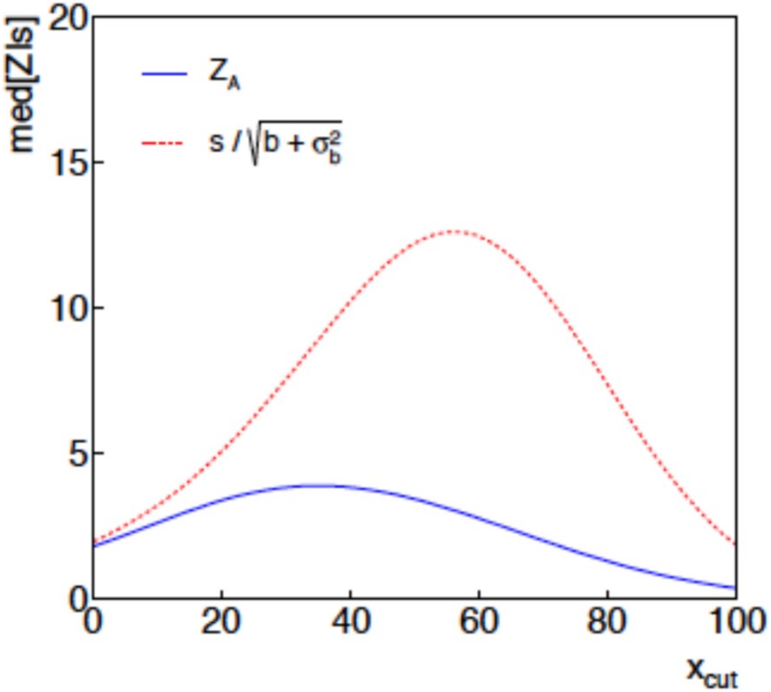
所以，“直观的”公式可以解释为一种极限情形，即将轮廓似然比检验用于Asimov数据集。

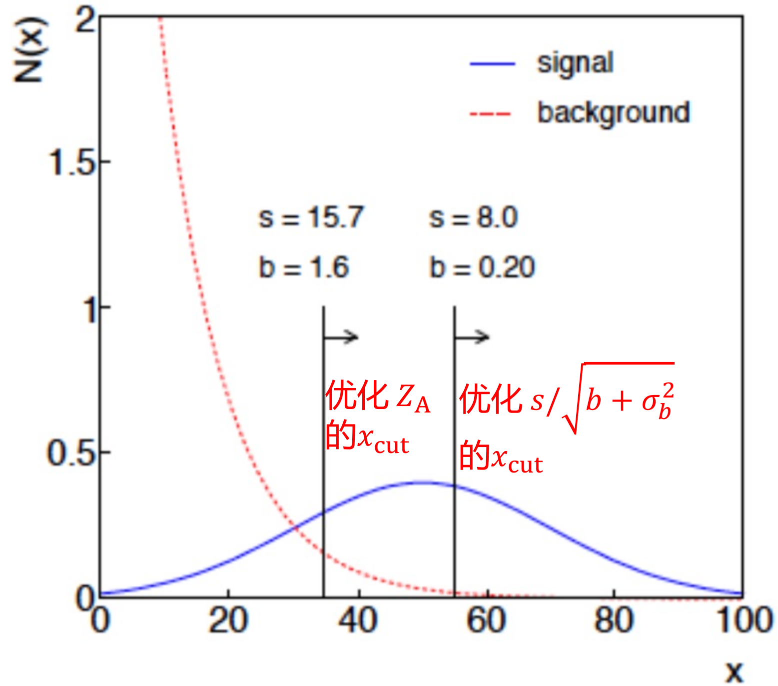
几个公式的对比：$s = 5$



med[Z|s]
8
6
4
2
0
10^{-1}
1
10
10^{2}
b
s = 5
σ_b/b = 0.2, 0.5
$s / \sqrt{b + \sigma_b^2}$
Z_A
Monte Carlo

用显著性/灵敏度优化事例筛选



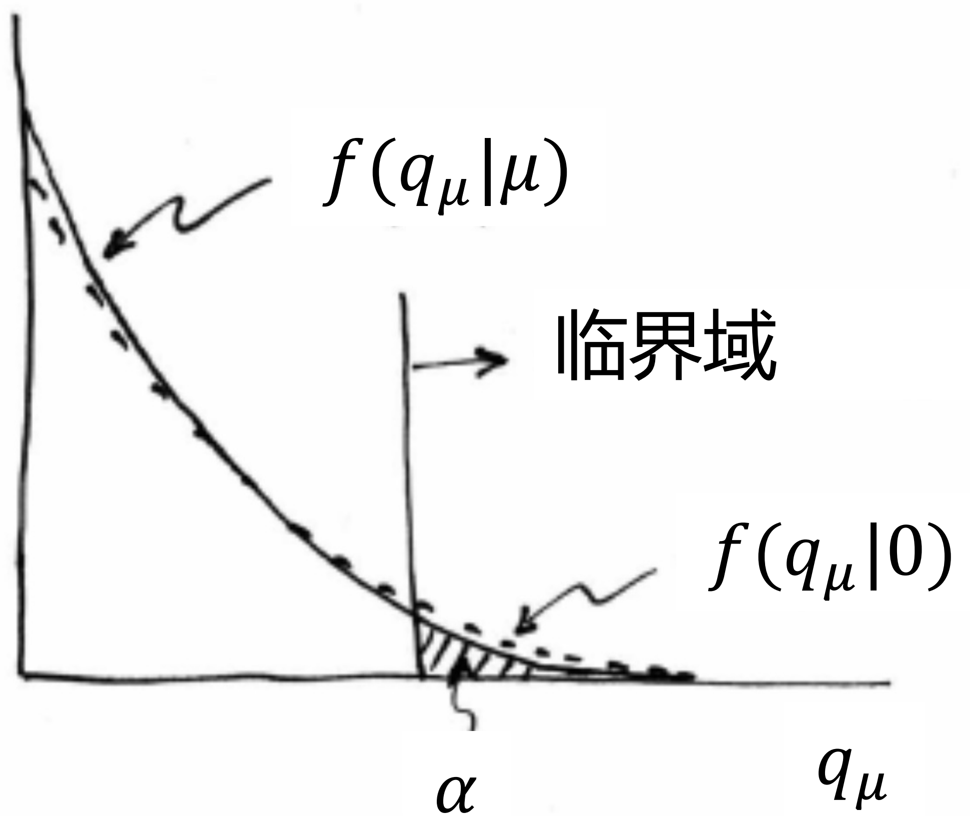


期待的显著性随筛选条件x_{cut}的变化。

信号与本底的分布及不同的事例筛选结果。

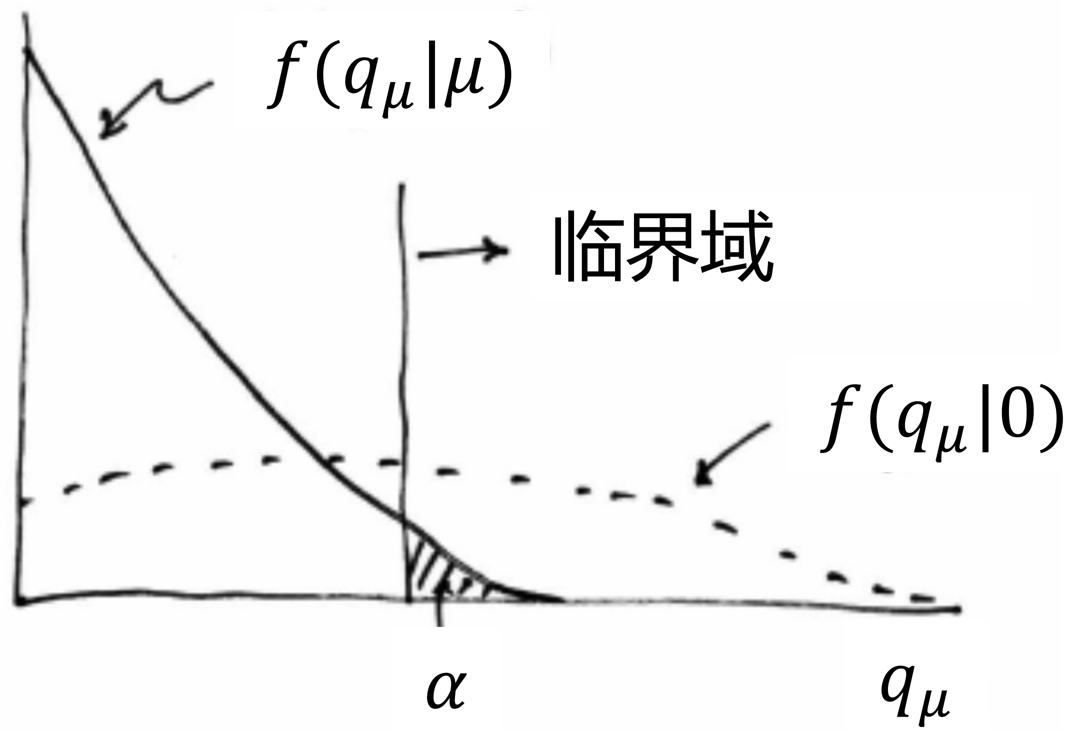
对 μ 的检验灵敏度很低的情形

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

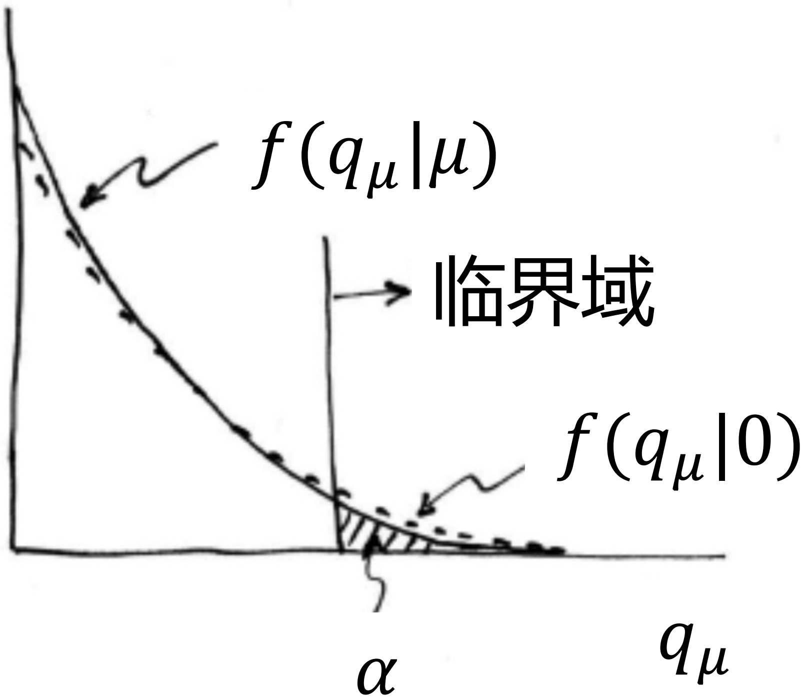
如果对 μ 的检验灵敏度很高，意味着 $f(q_\mu|\mu)$ 和 $f(q_\mu|0)$ 明显不同：



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”。

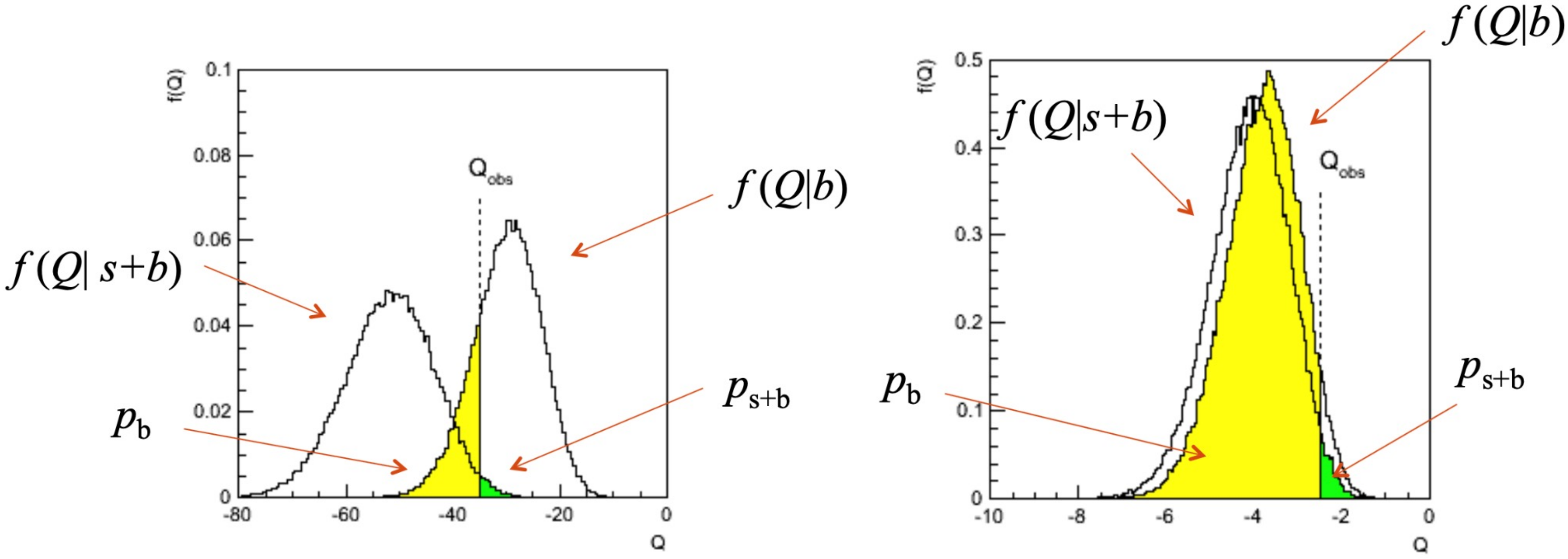
为了解决这个问题，提出了“$\mathrm{CL_s}$”方法设置上限。

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

$\mathrm{CL_s}$ 方法

在通常的$\mathrm{CL_s}$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。

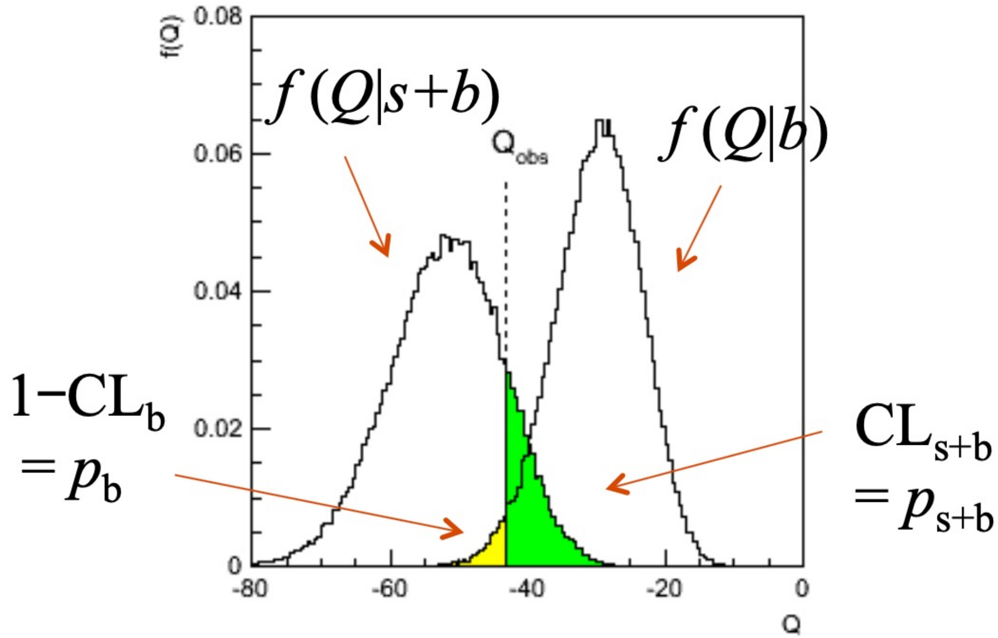


$\mathrm{CL_s}$ 方法（续）

$\mathrm{CL_s}$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

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \frac{p_{s+b}}{1 - p_b}$$

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

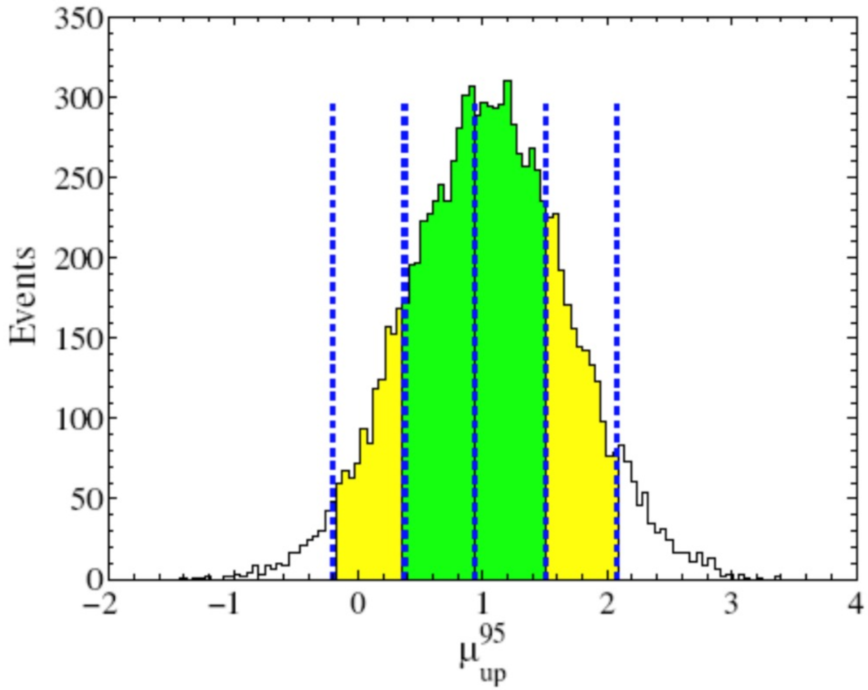


当两个分布离得越来越近时，则提高“有效的”p值（防止灵敏度太低时得出排除结论）。

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

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

例如，寻找希格斯粒子，定义 $\mu = m_H/m_H^{\mathrm{SM}}$ 。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 $f(\mu_{\mathrm{up}}|0)$。



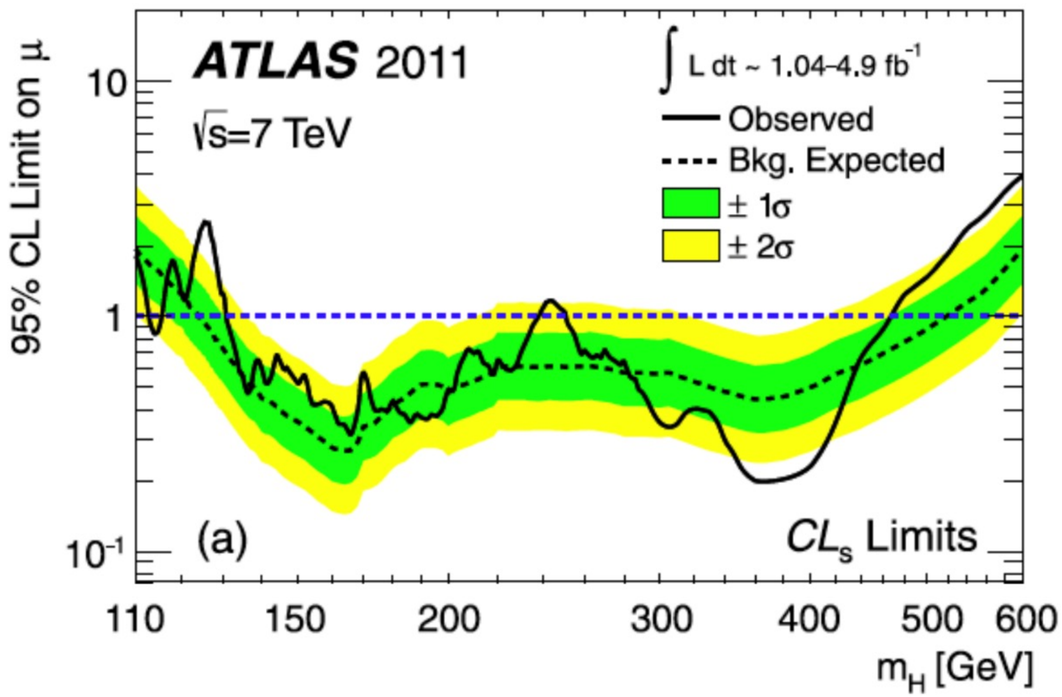
$\pm1\sigma$ (绿色)和 $\pm2\sigma$ (黄色)条带来自简易模拟（toy MC）；

竖线来自渐进公式。

如何理解上限图

对每个 m_H 值，求 μ 的 $\mathrm{CL_s}$ 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

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
- 贝叶斯上限
- 物理发现或上限的检验统计量