实验物理中的统计方法

条件概率 $P(A|B)=\frac{P(A\cap B)}{P(B)}$ 　相互独立：$P(A)=P(A|B)$ 即 $P(A\cap B)=P(A)P(B)$

主观概率（贝叶斯概率）：对假设为真的信心程度.

贝叶斯定理 $P(A|B)=\frac{P(B|A)P(A)}{P(B)}$. 　全概率公式 $P(B)=\sum_i P(B|A_i)P(A_i)$ 其中 A_i 是对 S 的一个划分.

概率密度函数 pdf: $f(x)=\frac{P(x\in[x,x+dx])}{dx}$ 　离散情形 pmf $f(x_i)=p_i$.

cdf（累积分布函数） $F(x)=\int_{-\infty}^{x}f(x')dx'=p(x'<x)$ 　$f(x)=\frac{\partial F(x)}{\partial x}$.

分位数 x_α 定义为：$F(x_\alpha)=\alpha\ (0\le\alpha\le1)$. $x_\alpha=F^{-1}(\alpha)$. 　中位数定义为 $x_{1/2}=F^{-1}(\frac{1}{2})$

众数：pdf 达最大值 $\frac{df}{dx}=0$.

多维分布 　联合概率密度 $f(x,y)=\frac{P(x\sim x+dx,\ y\sim y+dy)}{dxdy}$ 　边缘概率密度 $f_x(x)=\int f(x,y)dy=\frac{P(x\sim x+dx)}{dx}$

条件~~概率密~~ pdf: $h(y|x)=\frac{f(x,y)}{f_x(x)}$ 　$g(x|y)=\frac{f(x,y)}{f_y(y)}$ 　贝叶斯定理：$g(x|y)=\frac{h(y|x)f_x(x)}{f_y(y)}$

若 $f(x,y)=f_x(x)f_y(y)$ 即 $h(y|x)=f_y(y)$ 称 x,y 相互独立.

随机变量函数的 pdf 问题

函数 $a(x)$ 单调，其 pdf 为 $g(a)$ 则 $g(a)da=\left|\int_{x(a)}^{x(a+da)}f(x')dx'\right|=f[x(a)]\left|\frac{dx}{da}\right|da$ 　~~$\Rightarrow g(a)$~~

$\Rightarrow g(a)=f[x(a)]\left|\frac{dx}{da}\right|$.

若 $a(x)$ 不单调，需分段相加. 　$\boxed{\text{另法：从 cdf 求，即 } P(Y\le y)=P(g(x)\le y)=P(x\le g^{-1}(y)) \text{ 再求微分}}$ 　x 的区域

多维情形 $a=a(\vec{x})$. $\vec{x}=(x_1,\cdots,x_n)$ 　$g(a)da=\int_{dV}f(\vec{x})d\vec{x}$. 　dV：满足 $a(\vec{x})\in[a,a+da]$ 的

例 $z=xy$ $(x,y>0)$. $g(z)dz=\iint_{ds}f(x,y)dxdy=\int_0^\infty dx\int_{\frac{z}{x}}^{\frac{z+dz}{x}}f(x,y)dy=\int_0^\infty\left[f(x,\frac{z}{x})\frac{dz}{x}\right]dx=\left[\int_0^\infty f(x,\frac{z}{x})\frac{dx}{x}\right]dz$

$\Rightarrow g(z)=\int_0^\infty f(x,\frac{z}{x})\frac{dx}{x}$.

$x,y>0$ 服从 $f(x,y)$ 对 $z=x+y$, 　$g(z)dz=\iint_{ds}f(x,y)dxdy=\int_0^\infty dx\int_{z-x}^{z+dz-x}f(x,y)dy$.

$g(z)dz=\int_0^\infty dx\left[f(x,z-x)dz\right]=\left[\int_0^\infty f(x,z-x)dx\right]dz$ 　$\Rightarrow g(z)=\int_0^\infty f(x,z-x)dx$

假设 x,y 相互独立，分别服从 $g(x)$ 与 $h(y)$，则 $z=xy$, $f(z)=\int_{-\infty}^{+\infty}g(x)h(\frac{z}{x})\frac{dz}{|x|}$（梅林卷积 $f=g\otimes h$）

$z=x+y$, $f(z)=\int_{-\infty}^{+\infty}g(x)h(z-x)dx$（傅里叶卷积，$f=g\otimes h$）.

雅可比行列式

构造 n 个线性无关函数 $\vec{a}(\vec{x})=(a_1(\vec{x}),\cdots,a_n(\vec{x}))$ 其逆函数 $\vec{x}(\vec{a})$ 存在. 则 $\vec{a}$ 的联合 pdf 为：

$g(\vec{a})=|J|f(\vec{x})$. 　$J=\frac{D(x_1,\cdots,x_n)}{D(a_1,\cdots,a_n)}$

对 $g(\vec{a})$ 积分掉其他不关心的变量，可得任一边缘~~概率密度~~ pdf $g_i(a_i)$.

期待值、方差、标准差

期待值 $E[x]=\int xf(x)dx$ 　常记为 $E[x]=\mu$. 　可以证明 $E[a]=\int g(a)\,a\,da=\int a(x)f(x)dx$

方差 $V[x]=E[(x-E[x])^2]=E[x^2]-\mu^2$. 　记为 $V[x]=\sigma^2$. 　标准差 $\sigma=\sqrt{\sigma^2}$.

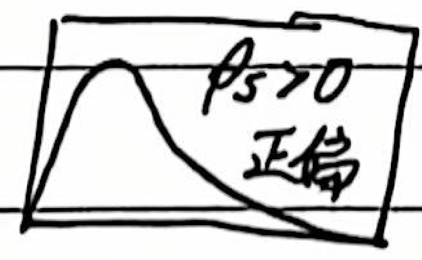
对多维情形，定义协方差 $\mathrm{cov}[x,y]$（或 V_{xy}）$=E[(x-\mu_x)(y-\mu_y)]=E[xy]-\mu_x\mu_y$.

(数理)相关系数 $\rho_{xy} = \frac{cov[x,y]}{\sigma_x \sigma_y}$ 无量纲. $-1 \le \rho_{xy} \le 1$

若 x,y 相互独立，则 $\rho_{xy}=0$ (反过来不一定).

秩相关性 更好度量相关性 $r_s = \rho_{R(X),R(Y)}$ 即先对 X,Y 进行排序. Spearman's rho

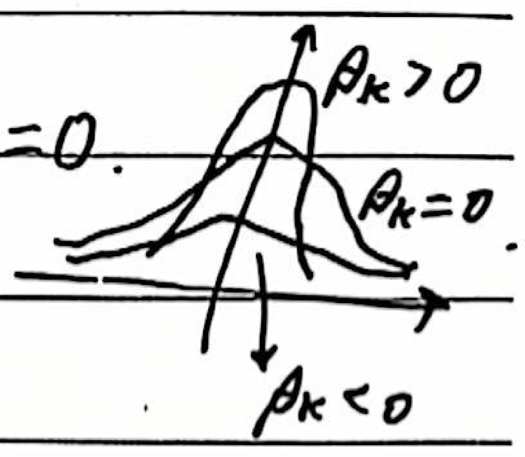
k阶原点矩: $\mu_k = E[X^k]$ k阶中心矩: $\nu_k = E[(X-\mu)^k]$



变异系数 $C_v = \frac{\sigma_x}{\mu_x}$ 无量纲反映波动大小.

偏度系数 $\rho_s = \frac{\nu_3}{\nu_2^{3/2}} = \frac{\nu_3}{\sigma^3}$ (偏度) 描述分布偏离对称性程度 $\rho_s>0$ 正偏/右偏 $\rho_s<0$ 负偏/左偏

峰度系数 $\rho_k = \frac{\nu_4}{\nu_2^2} - 3 = \frac{\nu_4}{\sigma^4} - 3$ (峰度) 描述分布的尖峭程度. 正态分布 $\rho_k=0$.



特征函数 $\varphi(t) = E[e^{itX}]$ $-\infty < t < \infty$.

连续变量 $\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$ 即pdf的傅里叶变换.

性质 $|\varphi(t)| \le \varphi(0) = 1$ $\varphi(-t) = \overline{\varphi(t)}$ $\cancel{\varphi(aX+b)}$ $\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$.

若 X,Y 独立, $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

$\varphi^{(k)}(0) = i^k E[X^k]$.

不确定度的传递. 设 $y = y(\vec{x})$. 已知 $\vec{x}$ 的协方差矩阵 $V_{ij} = cov[x_i, x_j]$.

将 y 在 $E[\vec{x}] = \vec{\mu}$ 线性展开, 得到 $V[y] = \sigma_y^2 \approx \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]\Big|_{\vec{x}=\vec{\mu}} V_{ij}$.

若 x_i 互不相关, $V_{ij} = \sigma_i \sigma_j \delta_{ij}$. $\Rightarrow V[y] \approx \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i}\right]^2_{\vec{x}=\vec{\mu}} \sigma_i^2$.

类似地, 对m组函数 $\vec{y}(\vec{x}) = (y_1(\vec{x}), \dots, y_m(\vec{x}))$. $U_{kl} = cov[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i}\frac{\partial y_l}{\partial x_j}\right]\Big|_{\vec{x}=\vec{\mu}} V_{ij}$.

即 $U = AVA^T$, $A_{ij} = \left[\frac{\partial y_i}{\partial x_j}\right]\Big|_{\vec{x}=\vec{\mu}}$.

局限: 仅在 $\vec{y}(\vec{x})$ 为线性时~~成立~~严格成立, 若函数在 σ_i 范围明显非线性也不行.

正交变换消除相关性. 对 $\vec{x} = (x_1, \dots, x_n)$ 与 $V_{ij} = cov[x_i, x_j]$

作变换 $\vec{y} = A\vec{x}$. $U = AVA^T$ 为对角阵. 待定

二项分布 $f(n;N,p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$ 即 N次独立测量，每次成功概率p，成功n次的概率。

$E[n] = Np \qquad V[n] = Np(1-p)$.

多项分布 有 $m>2$ 种结果，概率为 $p_1, \cdots, p_m$ 。N次试验结果用m维矢量 $\vec{n} = (n_1, \cdots, n_m)$ 表示，其概率为：

$$f(\vec{n};N,p) = \frac{N!}{n_1!\, n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}.$$

由 $E[n_i] = Np_i$，$V[n_i] = Np_i(1-p_i)$ 可得协方差 $V_{ij} = Np_i(\delta_{ij} - p_j)$。即 $i \neq j$ 时 n_i 与 n_j 负相关。

泊松分布 若取 $N \to \infty$，$p \to 0$ 　 $E[n] = Np \to \nu$ 的二项分布"极限

$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu}$ $(n \geq 0)$. 简记为 $\pi(\nu)$. (归一化 $\sum_{n=0}^{\infty} \frac{\nu^n}{n!} e^{-\nu} = 1$).

$E[n] = \nu \qquad V[n] = \nu$. 即标准差是平均值开根号。著名的统计不确定度估计式 $n \pm \sqrt{n}$.

均匀分布 $f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha}, & \alpha < x < \beta \\ 0, & 其他 \end{cases}$ 　简记为 $U(\alpha,\beta)$.

$E[x] = \frac{1}{2}(\alpha+\beta) \qquad V[x] = \frac{1}{12}(\beta-\alpha)^2$

若 $F(x)$ 为连续随机变量 x 的累积分布，则 $y = F(x)$ 服从 $U(0,1)$. (蒙特卡罗方法－反函数变换法)

指数分布 $f(x;\xi) = \begin{cases} \frac{1}{\xi} e^{-x/\xi}, & x \geq 0 \\ 0, & 其他 \end{cases}$ 　简记为 $Exp(\xi)$. 　$E[x] = \xi$; $V[x] = \xi^2$.

无记忆性: $f(t - t_0 | t \geq t_0) = f(t)$

高斯分布 (正态分布) $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$ 　简记为 $N(\mu,\sigma^2)$ 　$E[x] = \mu$ 　$V[x] = \sigma^2$.

标准正态分布 $N(0,1)$ 　$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 　$\Phi(x) = \int_{-\infty}^{x} \varphi(x')dx'$

若 $x \sim N(\mu,\sigma^2)$ 则 $y = \frac{x-\mu}{\sigma} \sim N(0,1)$

中心极限定理 n个独立随机变量 $\{x_i\}$，方差均存在，则 $y = \sum_{i=1}^{n} x_i$ 在 $n \to \infty$ 下服从 $N(\mu,\sigma^2)$

$\mu = \sum_{i=1}^{n} \mu_i \qquad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

泊松分布 ν 较大时，与高斯分布逐渐靠近 (实际中 $\nu \geq 5$)

多维高斯分布 $f(\vec{x};\vec{\mu},V) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1} (\vec{x}-\vec{\mu})\right]$. 　$E[x_i] = \mu_i$ 　$cov[x_i,x_j] = V_{ij}$

$n=2$ 时，$f(x_1,x_2;\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right)$

$\rho = \frac{cov[x_1,x_2]}{\sigma_1 \sigma_2}$ 是相关系数. 对二维正态分布，$\rho = 0 \Leftrightarrow x_1, x_2$ 相互独立.

卡方 (χ^2) 分布 $x_1, \cdots, x_n$ 是相互独立的服从 ~~$N(\mu$~~ $N(0,1)$. 　$z = \sum_{i=1}^{n} x_i^2$.

$z(\geq 0)$ 服从 χ^2 分布 (自由度为n) 　$f(z;n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{\frac{n}{2}-1} e^{-z/2}$ 　简记为 $\chi^2(n)$

$E[z] = n \qquad V[z] = 2n$.

柯西分布 $f(x)=\frac{1}{\pi}\frac{1}{1+x^2}$.

布莱特-魏格纳分布 $f(x;\Gamma,x_0)=\frac{1}{\pi}\frac{\Gamma/2}{\Gamma^2/4+(x-x_0)^2}$. x_0：模 Γ：半高全宽.

$E[x]$没有好的定义，$V[x]\to\infty$.

~~朗道分布 $f(\Delta;\beta)=\frac{1}{\xi}\phi(\lambda)$~~

$(0<x<1)$

β分布 $f(x;\alpha,\beta)=B(\alpha,\beta)\,x^{\alpha-1}(1-x)^{\beta-1}$. 记为 $Be(\alpha,\beta)$.

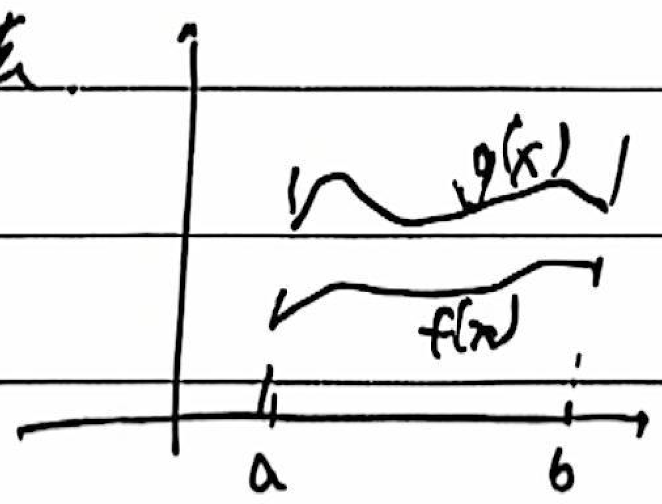
$E[x]=\frac{\alpha}{\alpha+\beta}$ $V[x]=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $B(1,1)=U(0,1)$.

Γ分布 $f(x;\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$ $Ga(\alpha,\beta)$ $(x>0)$ $E[x]=\alpha\beta$ $V[x]=\alpha\beta^2$.

学生氏分布 $f(x;\nu)=\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ 记为 $t(\nu)$ $E[x]=0$, $V[x]=\frac{\nu}{\nu-2}$.

蒙特卡罗方法 变换法 $r\sim U(0,1)$ $r=\int_{-\infty}^{x}f(x)dx \Rightarrow x=F^{-1}(r)$

舍选法



生成$g(x)$区域的随机分布点，再舍选出$f(x)$区域的点.

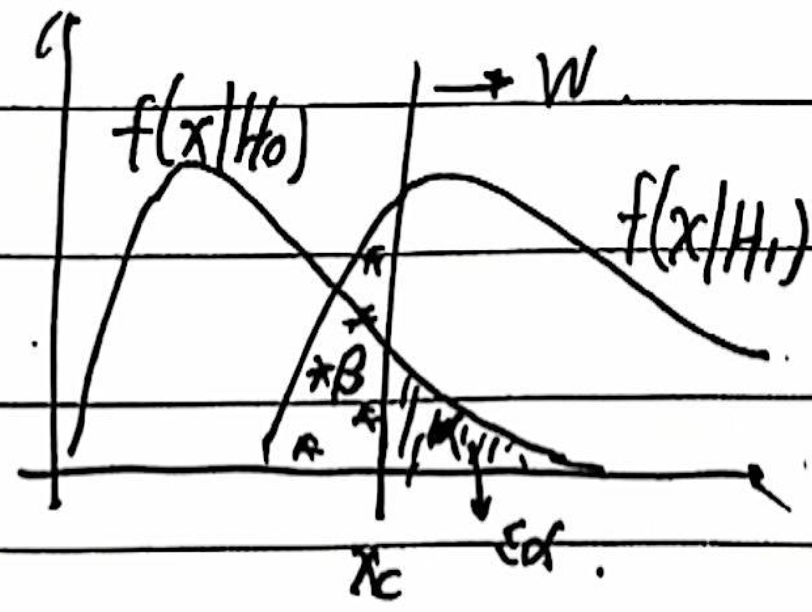
$x\sim g(x)$ $y\sim U(0,1)\cdot g(x)$

假设 假设H可以预测数据的概率，即观测的结果(用x表示) $x\sim f(x|H)$. 简单假设/复合假设.

给定H时x的概率称作假设H的似然值 $L(x|H)$ (朴素称为接受域).

考虑简单假设H_0和备择假设H_1，对H_0的检验定义为：对样本指定一个临界域W(拒绝域)，使得在H_0正确的情况下，观测到这数据概率不超过某个(小)概率α. $P(x\in W|H_0)\leq\alpha$ 一显著性水平/检验的大小.

第一类错误(弃真错误) 假设H_0为真而被拒绝 $P(x\in W|H_0)\leq\alpha$



第二类错误(取伪错误) H_0为假而H_1为真，但接受了H_0.

$P(x\in S-W|H_1)=\beta$.

$1-\beta$ 称为相对于备择假设H_1的检验的功效.

构造检验统计量 n维数据空间 $\vec{x}=(x_1,\cdots,x_n)$ 的临界域边界 $t(x_1,\cdots,x_n)=t_{cut}$ 其中t是一个标量检验统计量. $\Rightarrow$ n维问题化为一维问题

若H_0为本底事例，H_1为信号事例 本底选择效率：α 信号选择效率：$1-\beta$

信号纯度 $=P(s|\vec{x}\in W)=\frac{P(\vec{x}\in W|s)P(s)}{P(\vec{x}\in W|s)P(s)+P(\vec{x}\in W|b)P(b)}=\frac{(1-\beta)P(s)}{(1-\beta)P(s)+\alpha P(b)}=1-$信号误鉴别率

信号选择效率

信噪比 $\frac{P(s|\vec{x}\in W)}{P(b|\vec{x}\in W)} = \frac{1-\beta}{\alpha}\frac{P(s)}{P(b)}$

宽松选择：效率高，纯度低.

严格选择：纯度高，效率低.

奈曼-皮尔逊引理：在给定效率条件下，要得到最高纯度的信号样本，或在给定显著性水平下得到最高功效，可以选择下列接受域实现：$\frac{g(\vec{t}|H_0)}{g(\vec{t}|H_1)} > c$. (c为常数，与效率有关)

对不含未定参量的最佳的一致检验统计量为 $r = \frac{g(\vec{t}|H_0)}{g(\vec{t}|H_1)}$ → 简单假设 H_0 与 H_1 的似然比.

实际中 r 为单调函数为宜. 奈-皮方法在多变量情形下不实用.

线性检验统计量　线性变换 $t(\vec{x}) = \sum_{i=1}^{n} a_i x_i = \vec{a}^T\vec{x}$.　$\tau_k = \vec{a}^T\vec{\mu}_k$　$\Sigma_k^2 = \vec{a}^T V_k \vec{a}$.　$k=0,1$ (假设)

费舍尔定义鉴别函数 $J(\vec{a}) = \frac{(\tau_0-\tau_1)^2}{\Sigma_0^2+\Sigma_1^2}$　实际中要让 $|\tau_0-\tau_1|$ 尽可能大，Σ_0^2 与 Σ_1^2 尽可能小，即 $J(\vec{a})$ 尽可能大

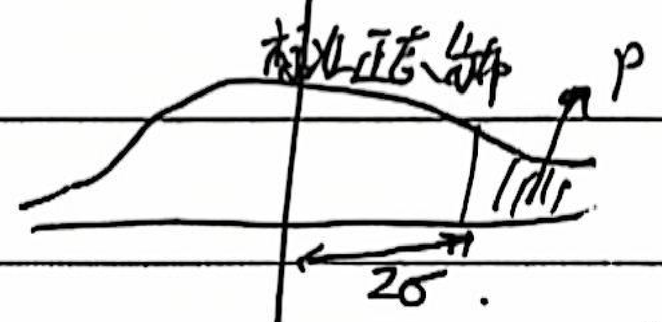
令 $B = (\vec{\mu}_0-\vec{\mu}_1)(\vec{\mu}_0-\vec{\mu}_1)^T$　$W = V_0+V_1$　则 $J(\vec{a}) = \frac{\vec{a}^T B\vec{a}}{\vec{a}^T W\vec{a}}$　令 $\frac{\partial J}{\partial a_i}=0 \Rightarrow \vec{a} \propto W^{-1}(\vec{\mu}_0-\vec{\mu}_1)$

将 $t(\vec{x})$ 写成 $t(\vec{x}) = a_0 + \sum_{i=1}^{n} a_i x_i$ 可以用 a_0 来固定 τ_0 与 τ_1，于是求 $J(\vec{a})$ 最大值即最小化其分母

$$\Sigma_0^2+\Sigma_1^2 = E_0(t-\tau_0)^2 + E_1(t-\tau_1)^2.$$

p值　定义：观测到数据 $\vec{x}$ 与假设 H 的符合程度不好于实际数据 $\vec{x}_{obs}$ 与 H 的符合程度的概率.

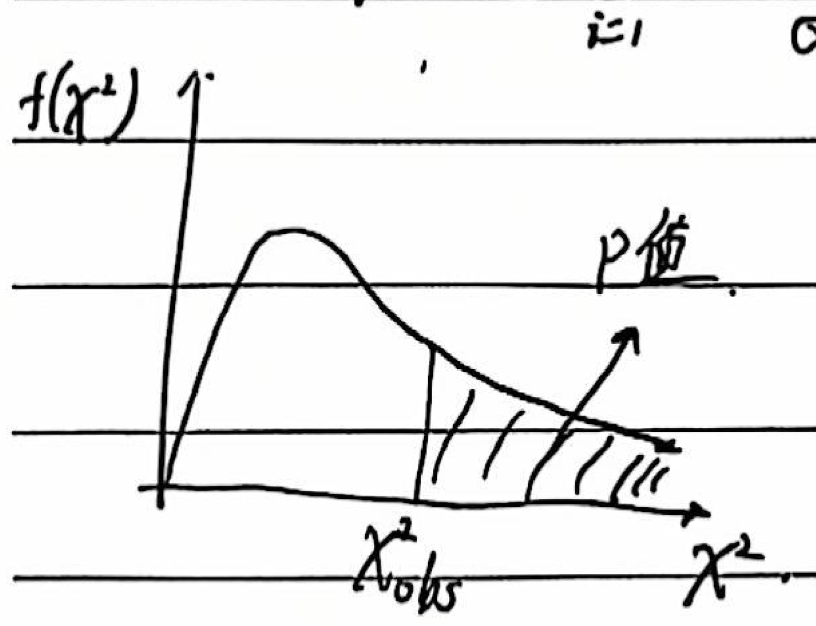
显著性 Z:　即 $p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx$.



查看别处效应：~~失去~~ 对信号位置的先验信息不足，考虑该效应后p值会变大.

皮尔逊卡方统计量　比较观测数据(实验): $\vec{n}$ 与预期均值(模型) $\vec{\nu}$ 是否相符.　(n_i 相互独立)

$\chi^2 = \sum_{i=1}^{N} \frac{(n_i-\nu_i)^2}{\sigma_i^2}$　$\sigma_i^2 = V[n_i]$.　若 $n_i \sim \pi(\nu_i)$ 则 $V[n_i]=\nu_i$.　此时 $\chi^2 = \sum_{i=1}^{N} \frac{(n_i-\nu_i)^2}{\nu_i}$



有 $p = \int_{\chi^2}^{\infty} f(z;N)\,dz$　若 $n_i \sim N(\nu_i, \sigma_i^2)$ 或 $n_i \sim \pi(\nu_i)$ 且 $\nu_i \gg 1$，则 χ^2 服从卡方分布.

评判估计量好坏　相合性(一致性)　$\lim_{n\to\infty}\hat{\theta} = \theta$　$\lim_{n\to\infty} P(|\hat{\theta}-\theta|>\varepsilon) = 0, \forall \varepsilon>0$ 成立

无偏性　$b = E[\hat{\theta}] - \theta = 0$.

有效性　对任何估计量 $\hat{\theta}'$，有 $\lim_{n\to\infty} \frac{V[\hat{\theta}_n]}{V[\hat{\theta}_n']} \le 1$.　称 $\hat{\theta}$ 为渐进有效估计量

~~有效估计量~~ 均方误差 $MSE = E[(\hat{\theta}-\theta)^2] = V[\hat{\theta}] + (E[\hat{\theta}-\theta])^2$.

无偏估计量中方差最小的估计量为有效估计量.

~~均值的估计量~~ 统计量：数据样本的函数（不含未知参数）　估计量：用来估计pdf的某些属性的统计量，记为$\hat\theta$.

估计值：估计量的观测值.

均值~~统计~~的估计量　$\hat\mu = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$（样本均值）.

若$V[x]$有限，$\bar x$是与μ相合的估计量，即　$\forall\varepsilon>0,\ \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|\geq\varepsilon\right)=0$（弱大数定理）

无偏性：$E[\bar x] = E[\frac{1}{n}\sum_{i=1}^{n}x_i] = \frac{1}{n}\sum_{i=1}^{n}E[x_i] = \frac{1}{n}\sum_{i=1}^{n}\mu=\mu$.　$V[\bar x]=\sigma^2/n$.

~~还~~可以证明，$\bar x$是μ的有效估计量.

方差的估计量　假设均值μ和方差$V[x]=\sigma^2$都是未知量，样本方差定义为　$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar x)^2$.

$s^2=\frac{n}{n-1}\left[\overline{x^2}-\bar{x}^2\right]$　因子$\frac{1}{n-1}$保证s^2无偏.

假设$\mu=E[x]$先验已知，则$S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=\overline{x^2}-\bar{x}^2$.　s^2与S^2都是σ^2的无偏估计量.

s^2的方差　$V[s^2]=\frac{1}{n}\left[\mu_4-\frac{n-3}{n-1}\mu_2^2\right]$　k阶中心矩μ_k估计：$\hat\mu_k=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar x)^k$.

协方差的估计量　$\hat V_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar x)(y_i-\bar y)=\frac{n}{n-1}(\overline{xy}-\bar x\cdot\bar y)$.（无偏）.

相关系数$\rho=V_{xy}/(\sigma_x\sigma_y)$ ~~估~~　$\hat\rho=r=\frac{\hat V_{xy}}{s_x s_y}=\frac{\overline{xy}-\bar x\bar y}{\sqrt{(\overline{x^2}-\bar x^2)(\overline{y^2}-\bar y^2)}}$　r有偏倚。但$n\to\infty$时该偏倚趋于0.

考虑样本$\vec x=(x_1,\cdots,x_n)$，x_i独立同分布，$x_i\sim f(x;\theta)$　目标：估计θ. 更一般地，估计$\vec\theta=(\theta_1,\cdots,\theta_m)$

若$f(x;\theta)$为真，则P(观测到第i个数据处于$[x_i,x_i+dx_i]$，$\forall i=1,\cdots,n$)$=\prod_{i=1}^{n}f(x_i,\theta)\,dx_i$.

定义似然函数$L=\prod_{i=1}^{n}f(x_i,\theta)$　$\Rightarrow$ $\hat\theta$使L取最大.　$\ln L(\hat\theta)$ 对数似然函数

最大似然估计量$\hat\theta$：$\frac{\partial L(\hat\theta)}{\partial\theta_i}=0$.　实际应用通常给出最好的实用解.

例外：$X\sim U(0,a)$　$\vec x=(x_1,\cdots,x_n)$　则由最大似然法给出$\hat a=\max\{x_1,\cdots,x_n\}$.　有偏倚

最大似然估计量的唯一性　选取另一等价参数$h(\theta)$，只要$\frac{\partial h}{\partial\theta}\neq0$，则$\hat h=h(\hat\theta)$.

信息不等式为任何估计量方差设定了下界　$V[\hat\theta]\geq\left(1+\frac{\partial b}{\partial\theta}\right)^2\Big/E\left[-\frac{\partial^2\ln L}{\partial\theta^2}\right]$　→ 最小方差界(MVB)

(RCF不等式)　b为偏倚量$E[\hat\theta]-\theta$.

如果$b=0$且等式成立，称$\hat\theta$是有效估计量　通常b很小，等式近似成立，则$V[\hat\theta]\approx-1/E\left[\frac{\partial^2\ln L}{\partial\theta^2}\right]$.

对于m个参数$\vec\theta=(\theta_1,\cdots,\theta_m)$　最小方差界由Fisher信息矩阵给出：

$$I_{ij}=E\left[-\frac{\partial^2\ln L}{\partial\theta_i\partial\theta_j}\right]=-n\int f(x;\vec\theta)\frac{\partial^2\ln f(x;\vec\theta)}{\partial\theta_i\partial\theta_j}dx.$$

信息不等式：$V-I^{-1}$是半正定矩阵，其中$V_{ij}=\mathrm{cov}[\hat\theta_i,\hat\theta_j]$.　因此$V[\hat\theta_i]\geq(I^{-1})_{ii}$.

经常用I^{-1}近似协方差矩阵，并进一步估计：$(\hat V^{-1})_{ij}=-\frac{\partial^2\ln L}{\partial\theta_i\partial\theta_j}\Big|_{\vec\theta=\hat{\vec\theta}}$.

考虑单参数θ的情况，将$\ln L(\theta)$在$\hat\theta$展开　$\ln L(\theta)=\ln L(\hat\theta)+\frac{1}{2}\left[\frac{\partial^2\ln L}{\partial\theta^2}\right]_{\theta=\hat\theta}(\theta-\hat\theta)^2+\cdots$

$\Rightarrow$ $\ln L(\theta)\approx\ln L_{max}-\frac{(\theta-\hat\theta)^2}{2\hat\sigma_{\hat\theta}^2}$　即$\ln L(\hat\theta\pm\hat\sigma_{\hat\theta})\approx\ln L_{max}-\frac{1}{2}$.

$\Rightarrow$ 为了得到$\hat\sigma_{\hat\theta}$，让θ偏离$\hat\theta$，使得$\ln L$减小0.5　（图解法）.

扩展的最大似然法　有时样本容量 $n \sim \pi(\nu)$．实验结果定义为 $n, x_1, \cdots, x_n$．

则 $L(\nu,\vec\theta) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^{n} f(x_i;\vec\theta)$．假设理论给出 $\nu = \nu(\vec\theta)$．去掉与 $\vec\theta$ 无关的项后．

$\ln L(\vec\theta) = n \ln \nu(\vec\theta) - \nu(\vec\theta) + \sum_{i=1}^{n} \ln(f(x_i;\vec\theta))$．$= -\nu(\vec\theta) + \sum_{i=1}^{n} \ln[\nu(\vec\theta)\cdot f(x_i;\vec\theta)]$

假设 ν 与 $\vec\theta$ 相互独立　$\frac{\partial L}{\partial \nu} = 0$　$\Rightarrow \hat\nu = n$．

不等精度观测结果合并　对 μ 作 n 次独立测量，结果为 $x_i \pm \sigma_i$．即 $x_i \sim N(\mu, \sigma_i^2)$．

$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma_i^2}\right]$．　$\frac{\partial \ln L}{\partial \mu} = 0$．$\Rightarrow \hat\mu = \frac{1}{\omega}\sum_{i=1}^{n} \omega_i x_i$．　$\omega_i = 1/\sigma_i^2$．　$\omega = \sum_{i=1}^{n} \omega_i$．

$\hat{\sigma^2_{\hat\mu}} = \left(-1\Big/\frac{\partial^2 \ln L}{\partial \mu^2}\right)\Big|_{\mu=\hat\mu} = \frac{1}{\omega}$．小于任何一个 σ_i^2．

最小二乘法．设 Gauss 变量 y_i $(i=1,\cdots,N)$　$E[y_i] = \lambda_i = \lambda(x_i;\vec\theta)$　$x_1,\cdots,x_N$ 与 $V[y_i] = \sigma_i^2$ 已知　估计 $\vec\theta$

$g(\vec y;\vec\lambda,\vec\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(y_i-\lambda_i)^2}{2\sigma_i^2}\right)$．　$\ln L(\vec\theta) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i - \lambda_i(x_i;\vec\theta))^2}{\sigma_i^2}$

即求 $\chi^2(\vec\theta) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(x_i;\vec\theta))^2}{\sigma_i^2}$ 的最小值．

最小二乘估计量　设 y_i 为多维 Gauss 变量，协方差矩阵为 V 已知．$g(\vec y;\vec\lambda,V) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} \exp\left[-\frac{1}{2}(\vec y-\vec\lambda)^T V^{-1}(\vec y-\vec\lambda)\right]$．

$\ln L(\vec\theta) = -\frac{1}{2}\sum_{i,j=1}^{N} [y_i - \lambda(x_i;\theta)] (V^{-1})_{ij} [y_j - \lambda(x_j;\theta)]$．即求下式的最小值：

$\chi^2(\vec\theta) = \sum_{i,j=1}^{N} [y_i - \lambda(x_i;\vec\theta)] (V^{-1})_{ij} [y_j - \lambda(x_j;\vec\theta)]$．其最小值定义了最小二乘估计量 $\hat\theta$．

即使 y_i 不是 Gauss 分布，该定义仍适用

线性情况的最小二乘估计　$\lambda(x;\vec\theta) = \sum_{j=1}^{m} a_j(x)\theta_j$　其中 $a_j(x)$ 是 x 的任意线性独立函数．

高斯-马尔科夫定理：$\hat\theta$ 没有偏倚，且得到的方差最小．

令 $A_{ij} = a_j(x_i)$．$\chi^2(\vec\theta) = (\vec y - \vec\lambda)^T V^{-1}(\vec y - \vec\lambda) = (\vec y - A\vec\theta)^T V^{-1}(\vec y - A\vec\theta)$．

$\nabla\chi^2 = -2(A^T V^{-1}\vec y - A^T V^{-1}A\vec\theta) = 0$．$\rightarrow \hat\theta = \cancel{A^T V^{-1} A}\ (A^T V^{-1} A)^{-1} A^T V^{-1}\vec y = B\vec y$．

$U_{ij} \triangleq \mathrm{cov}[\hat\theta_i, \hat\theta_j]$．$\Rightarrow U = BVB^T = (A^T V^{-1} A)^{-1}$

由RCF下界：$\frac{\partial \chi^2}{\partial \vec\theta} = -2\frac{\partial(A^T V^{-1}\vec y - A^T V^{-1} A\vec\theta)}{\partial\vec\theta} = 2A^T V^{-1} A$．　$(U^{-1})_{ij} = \frac{1}{2}\left[\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right]\Big|_{\vec\theta=\hat\theta} = \sum_{k,l=1}^{N} a_i(x_k)(V^{-1})_{kl}\, a_j(x_l)$

$\rightarrow$ 若 y_i 是 Gauss 变量，其方差与 RCF 边界一致．

$\chi^2(\vec\theta) = \chi^2(\hat\theta) + \frac{1}{2}\sum_{i,j=1}^{m}\left[\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right]_{\vec\theta=\hat\theta} (\theta_i - \hat\theta_i)(\theta_j - \hat\theta_j)$　令 $\chi^2(\hat\theta) = \chi^2_{min}$　上式给出 $\chi^2(\hat\theta \pm \hat\sigma_{\hat\theta}) = \chi^2_{min} + 1$．

可把 $\chi^2(\vec\theta) \leq \chi^2_{min} + 1$ 看作"置信区间"给出含真值 $\vec\theta$ 的可能性．

多项式的最小二乘拟合　$\lambda(x;\theta_0,\cdots,\theta_m) = \sum_{j=0}^{m} \theta_j x^j$

注　如果数据服从 Gauss 分布，那么 $\chi^2(\theta) = -2\ln L(\theta)$．　因此 $\hat\sigma^2_{\hat\theta} \approx 2\left[\frac{\partial^2\chi^2}{\partial\theta^2}\right]^{-1}_{\theta=\hat\theta}$．

或由图解法　$\chi^2(\hat\theta_0 \pm \hat\sigma_{\hat\theta_0}) = \chi^2_{min} + 1$．

检验最小二乘法的拟合优度　假设 $y_i\ (i=1,\cdots,N)$ 为独立Gauss变量（σ_i 已知），且 $\lambda(x_i;\vec{\theta})$ 是 θ_i 的正确函数，那么 $\vec{\theta}=\hat{\vec{\theta}} \Rightarrow \chi^2_{min}=\sum_{i=1}^{N}\frac{(y-\lambda(x_i;\hat{\vec{\theta}}))^2}{\sigma_i^2}$　$\chi^2_{min}\sim\chi^2(n_d)$. $n_d=N-m$. m 为参数个数.

χ^2_{min} 可用作拟合优度统计量，检验函数形式 $\lambda(x_i;\hat{\theta})$ 的好坏.（单边检验）.

拟合优度与统计不确定度的关系　统计不确定度小不意味着是一个好的拟合（反之亦然）.

用最小二乘法合并实验结果　~~对~~已知 λ 的 N 个测量结果 $y_1,\cdots,y_N$　$\sigma_i^2=V[y_i]$ 已知.

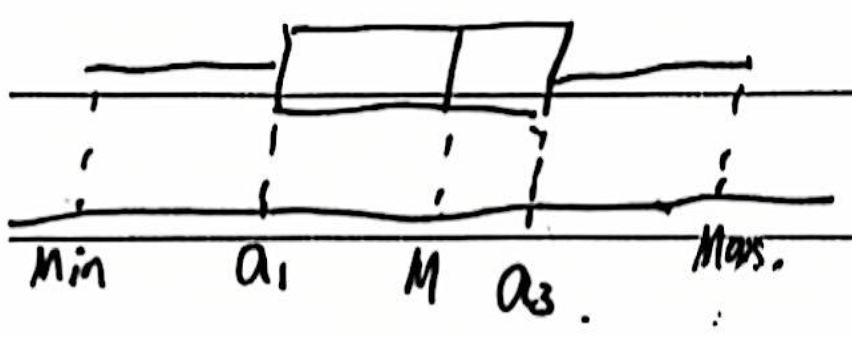
若各测量量之间不相关　$\chi^2(\lambda)=\sum_{i=1}^{N}\frac{(y_i-\lambda)^2}{\sigma_i^2} \Rightarrow \hat{\lambda}=\sum_{i=1}^{N}w_iy_i$.　$w_i=\frac{1/\sigma_i^2}{\sum_{j=1}^{N}1/\sigma_j^2}$　$V[\hat{\lambda}]=\sum_{i=1}^{N}w_i\sigma_i$

若各测量量相关　$cov[y_i,y_j]=V_{ij}$.　$\chi^2(\lambda)=\sum_{i,j=1}^{N}(y_i-\lambda)(V^{-1})_{ij}(y_j-\lambda)$.

$\Rightarrow \hat{\lambda}=\sum_{i=1}^{N}w_iy_i$.　$w_i=\sum_{j=1}^{N}(V^{-1})_{ij}\Big/\sum_{k,l=1}^{N}(V^{-1})_{kl}$.　$V[\hat{\lambda}]=\sum_{i,j=1}^{N}w_iV_{ij}w_j$.

LS方法得到 $\hat{\lambda}$ 无偏，且方差最小（高斯－马科夫定理）

箱线图　~~样分~~ 样本 p 分位数 $(0<p<1)$ 记为 x_p，至少有 np 个样本 $\le x_p$　至少有 $n(1-p)$ 个值 $\ge x_p$



简单随机样本　$X_1,\cdots,X_n$ 独立同分布.　以后简称样本

~~对~~ 用不放回抽样代替放回抽样所得样本的代替条件：$N/n\ge10$. N 为总体中个体总数，n 为样本容量.

抽样分布：统计量 $T_n=g(X_1,\cdots,X_n)$ 的分布　(1) χ^2 分布　(2) t 分布　(3) F 分布　(4) 正态总体样本均值和样本方差分布.

F 分布　设 $X\sim\chi^2(n)$　$Y\sim\chi^2(m)$　且 X,Y 相互独立。令 $F=\frac{X/n}{Y/m}$　称 F 服从第一自由度为 n，第二自由度为 m 的 F 分布.

性质：1. 若 $F\sim F(n,m)$ 则 $1/F\sim F(m,n)$.

2. $F_{1-\alpha}(n,m)=\frac{1}{F_\alpha(m,n)}$　其中 $F_\alpha(m,n)$ 为 $F(m,n)$ 的上 α 分位数.

t 分布　设 $X\sim N(0,1)$　$Y\sim\chi^2(n)$　X,Y 相互独立　称 $T=\frac{X}{\sqrt{Y/n}}$ 为服从自由度为 n 的 t 分布.

$n\to\infty$ 时　~~F~~ t 分布 $\longrightarrow N(0,1)$.

性质：1° $f_n(t)$ 为偶函数　$n\to\infty$　$f_n(t)\to\varphi(t)$.

2° 满足 $P(|T|>t_{\alpha/2})=\alpha$ 的点 $t_{\alpha/2}$ 称为 t 分布的双侧 α 分位数.

正态总体样本均值和样本方差的分布　设总体 $X\sim N(\mu,\sigma^2)$.　样本均值 $\bar{X}\sim N(\mu,\sigma^2/n)$.

或 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$.

样本方差分布 —— σ^2 已知　$\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2\sim\chi^2(n-1)$.

σ^2 未知但 μ 已知　则 $\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$.

围绕某固定值 x_0 的 k阶矩/简单矩 $\mu_k = \int (x-x_0)^k f(x)\,dx$　　$x_0=0$ 时称为k阶原点矩 μ_k' 或 e_k.

$x_0 = E[x] = \mu$ 时称为 k阶中心矩 μ_k.

原点矩与中心矩关系、　$\mu_k = \sum_{l=0}^{k} \binom{k}{l} \mu_{k-l}' (-\mu_1')^l$,　　$\mu_k' = \sum_{l=0}^{k} \binom{k}{l} \mu_{k-l} (\mu_1')^l$.

矩的一般表达　设 $x \sim f(x;\hat{\theta})$. 构造 m个线性独立函数 $a_i(x)$.

$E[a_i(x)] = \int a_i(x) f(x,\hat{\theta})\,dx = e_i(\hat{\theta})$　(矩的一般表达式).

其中 $e_i(\hat{\theta})$ 可用无偏样本均值估计 $\hat{e}_i = \overline{a_i} = \frac{1}{n}\sum_{j=1}^{n} a_i(x_j)$.　令 $e_i(\hat{\theta}) = \overline{a_i}$，则 $\hat{\theta}$ 可通过求解 m个 $e_i(\hat{\theta})$ 方程来确定.

常见选择：$a_i = x^i$　$E[a_i(x)] = \mu_i'$　i阶原点矩.

置信区间　设θ为待估参数，对~~某~~给定实数 α $(0<\alpha<1)$，存在两个统计量 $T_1(X_1,\cdots,X_n)$ 与 $T_2(X_1,\cdots,X_n)$

若对 $\forall\theta\in\Theta$，有 $P(T_1 \le \theta \le T_2) \ge 1-\alpha$　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.

若上式可以取等，称 $[T_1, T_2]$ 为θ的 $1-\alpha$ 同等置信区间.

枢轴量：仅含1个参数的样本的连续函数，且分布不依赖于未知参数. 如 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ 其中σ已知.

等尾置信区间　使两个尾部概率各为 $\alpha/2$

~~高斯分布估计量的置信区间　若$\hat{\theta}$服从 $g(\hat{\theta};\theta) = \frac{1}{\sqrt{2\pi\sigma_{\hat{\theta}}^2}} \exp\left(-\frac{(\hat{\theta}-\theta)^2}{2\sigma_{\hat{\theta}}^2}\right)$ 求θ的置信区间~~

后验 Jeffrey's 验前概率　$\pi(\hat{\theta}) \propto \sqrt{\det(I(\hat{\theta}))}$.　其中 $I(\hat{\theta})$ 是Fisher信息矩阵

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

这样的验前概率选择给出的推断在参数变换下不变

$n \sim N(\mu,\sigma^2)$.　$\pi(\mu) \propto \frac{1}{\mu}$.　　$n \sim \underset{\text{泊松分布}}{\pi(\lambda)}$　$\pi(\lambda) \sim \frac{1}{\sqrt{\lambda}}$.