

# 实验物理中的统计方法

## chap 01 基本概念

1. 随机事件：现象  $A$  有可能发生，也有可能不发生，那么现象  $A$  发生的事例叫做随机事件。

2. 概率

柯尔莫戈夫公理：（概率的定义）概率是事件的函数。

- $P(A) \geq 0$
- $P(S) = 1$
- 若  $A \cap B = \emptyset$ , 则  $P(A) + P(B) = P(A \cup B)$

3. 条件概率与独立性

$$P(A|B) = \frac{P(AB)}{P(B)}$$

$A$ 与 $B$ 独立:  $P(A|B) = P(A)$

$A \cap B = \emptyset$  不能推出  $AB$ 独立无关。

4. 概率的诠释

- 概率是相对频率（**可重复**实验中某个结果出现的次数占所有结果的比例）
- **主观**概率（贝叶斯概率）

$P(A)$  = 对 $A$ 为真的信心程度

5. 贝叶斯定理与全概率公式

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i) = \sum_i P(B|A_i)P(A_i)$$

阐释：

- $P(A)$ 为理论的**先验概率**：“抛掷一枚硬币，正面反面向上的可能性各占一半”这一理论正确的概率
- $P(B|A)$ 为理论预言的实验正确的概率：“硬币正面向上的概率为 $\frac{1}{2}$ ”
- $P(A|B)$ 为理论的**后验概率**

将 $P(A|B)$ 作为下一次的先验概率，进行不断迭代。

例子：三门问题

- 参赛者可看见三扇关闭的门，其中一扇门后面有一辆汽车，选中它即可赢得汽车；另外两扇门后面各藏有一只山羊。参赛者先选定一扇门，但在开启之前，知情的主持人会在其余两扇门中打开一个后面是山羊的门，并讯问参赛者是否更换其选择的门。问题是：参赛者更换选择是否会增加赢得汽车的机会？



### 三门问题的贝叶斯解法：

不失一般性，可以始终将嘉宾选择的门标记为“1”，将主持人打开的门标记为“3”，将剩下的门标记为“2”。

第  $i$  个门后面是汽车的先验概率为

$$P(T_i) = 1/3$$

如果汽车在1号门，主持人打开3号门的条件概率为

$$P(O_3|T_1) = 1/2$$

如果汽车在2号门，主持人打开3号门的条件概率为

$$P(O_3|T_2) = 1$$

如果汽车在3号门，主持人打开3号门的条件概率为

$$P(O_3|T_3) = 0$$

由贝叶斯定理，主持人打开3号门的条件下，在1号门和2号门发现汽车的条件概率分别为：

$$\begin{aligned} P(T_1|O_3) &= \frac{P(O_3|T_1)P(T_1)}{\sum_{i=1}^3 P(O_3|T_i)P(T_i)} = \frac{P(O_3|T_1)}{\sum_{i=1}^3 P(O_3|T_i)} \\ &= \frac{1/2}{1/2 + 1 + 0} = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} P(T_2|O_3) &= \frac{P(O_3|T_2)P(T_2)}{\sum_{i=1}^3 P(O_3|T_i)P(T_i)} = \frac{P(O_3|T_2)}{\sum_{i=1}^3 P(O_3|T_i)} \\ &= \frac{1}{1/2 + 1 + 0} = \frac{2}{3} \end{aligned}$$

所以，嘉宾变更选择后获得汽车的概率更大。

累积分布函数 (c.d.f.)

## 7. $\alpha$ 分位数、中位数、众数

定义:

$$F(x_\alpha) = \alpha$$

按照数值小的一方面积分定义。

即,  $\alpha$ -分位数称为下侧 $\alpha$ -分位数, 当然也可以对应的定义上侧 $\alpha$ -分位数。

## 8. 直方图与概率密度函数

联合概率密度分布函数 $f(x, y)$

边缘概率密度 $f_x(x), f_y(y)$

相互独立:

$$f(x, y) = f_x(x)f_y(y)$$

## 9. 随机变量的函数

- 一维随机变量

函数  $a = a(x)$ ,  $x$  的概率密度函数  $f(x)dx$

$$g(a)da = f(x)dx$$

得到:

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

- 多维随机变量的函数

如:  $z = xy$

Mellin 卷积

$$f(z) = \int_{-\infty}^{\infty} g(x)h\left(\frac{z}{x}\right)\frac{dx}{|x|}$$

Flourier 卷积

$$f(z) = \int_{-\infty}^{\infty} g(x)h(z-x)dx$$

- Jacobi 行列式

做变换, 将我们不关心的函数部分积分掉, 得到我们关心的概率。

## 10. 期待值与方差

期待值、方差、标准差

- 期待值

$$\mu = E[x] = \int x f(x)dx$$

- 协方差

$$\text{cov}[x, y] = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x \mu_y$$

- (Pearson)相关系数

$$\rho_{x,y} = \frac{\text{cov}[x,y]}{\sigma_x \sigma_y}$$

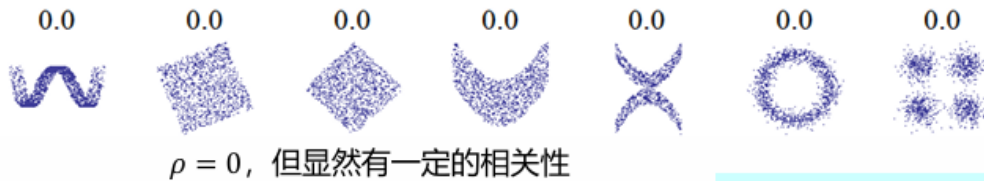
◦ 协方差矩阵  $V$ :

$$V_{ij} = \text{cov}[x_i, x_j]$$

由定义，协方差矩阵为对称矩阵。

相关性定义的局限性：相关系数为0，但是不能说明x与y不相关。

其他的比较方式：秩相关性（只看顺序，不看数值）等等



原点矩与中心矩

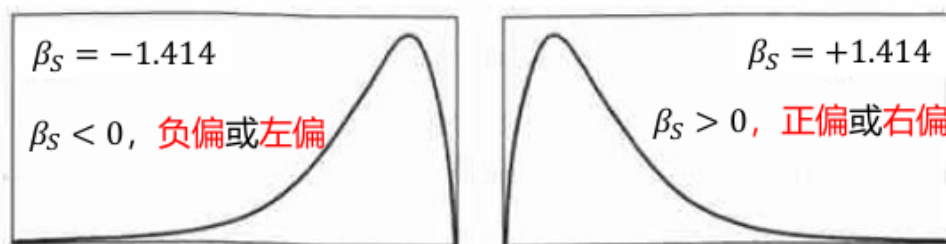
$$\mu_k = E(X^k)$$

$$\nu_k = E[(X - E(X))^k]$$

其他特征数:

偏度系数

$\beta_S$  描述分布偏离对称性的程度



## 11. 特征函数

对于随机变量  $X$ ，称  $e^{itX}$  的数学期望值为随机变量的特征函数。

$$\varphi(t) = E[e^{itX}]$$

对于连续的情况，

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itX} f(x) dx$$

特征函数的性质:

如果  $XY$  相互独立，则特征函数

$$\varphi_{X+Y}(t) = E[e^{it(X+Y)}] = E[e^{itX} e^{itY}] = \varphi_X(t) \varphi_Y(t)$$

一例:

已知随机变量  $X_1, \dots, X_n$  服从相同的概率分布  $f(x)$  且相互独立，则如何计算  $X = \sum_{i=1}^n X_i$  的概率分布函数?

解:

将卷积转化为乘积，都化成特征函数处理。（傅里叶变换的性质）

## 12. 不确定度的传递

问题表述：

假设我们对某个量测量了一组值  $\vec{x} = (x_1, \dots, x_n)$ ，并得到其协方差  $V_{ij} = \text{cov}[x_i, x_j]$ （表征与  $x_i$  有关的测量不确定度）。不确定度的传递<sup>(1) 73</sup> 现考虑一函数  $y = y(\vec{x})$ ，如何求其方差  $V[y]$ ？

例：

考虑随机变量  $X_1, X_2$  与不确定度  $\sigma_1, \sigma_2$ ，计算  $Y = f(X_1, X_2)$  的不确定度？

$$\sigma_Y = \sqrt{\left(\frac{\partial f}{\partial X_1}\right)^2 \sigma_{X_1}^2 + \left(\frac{\partial f}{\partial X_2}\right)^2 \sigma_{X_2}^2 + 2 \frac{\partial f}{\partial X_1} \frac{\partial f}{\partial X_2} \text{cov}[X_1, X_2]}$$

证明：

认为误差不太大，将  $Y$  在平均值附近展开：

$$y(\vec{x}) = y(\vec{\mu}) + \sum_{i=1}^n \frac{\partial y}{\partial x_i} (x_i - \mu_i)$$

计算

$$\sigma_y = E[y^2] - E[y]^2$$

即可得到结果。

随机变量之间的相关性往往与坐标选取有关，可以通过正交变换消除相关性。可以使用类似的方案得到结果：

实验上测量带电粒子动量通常是，测量粒子在探测器中各点的击中坐标  $(x, y)$ ，然后拟合径迹。径迹往往用极坐标  $(r, \theta)$  描述。一般来说， $(x, y)$  的测量不相关。 $(r, \theta)$  是否相关？

两种坐标的变换关系：

$$r^2 = x^2 + y^2, \quad \tan \theta = y/x$$

$$V_{xy} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

由于  $U_{r\theta} = A V_{xy} A^T$

$$A_{ij} = \left[ \frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} \Rightarrow A = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix}$$

坐标变换后的协方差矩阵为

$$U_{r\theta} = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix} \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \begin{bmatrix} \frac{x}{r} & -\frac{y}{r^2} \\ \frac{y}{r} & \frac{x}{r^2} \end{bmatrix} = \frac{1}{r^2} \begin{bmatrix} x^2 \sigma_x^2 + y^2 \sigma_y^2 & xy(\sigma_y^2 - \sigma_x^2) \\ xy(\sigma_y^2 - \sigma_x^2) & x^2 \sigma_x^2 + y^2 \sigma_y^2 \end{bmatrix}$$

除非处处满足  $\sigma_x = \sigma_y$ ，否则  $(r, \theta)$  有相关性。

由于实对称矩阵可以正交相似对角化，所以一定可以找到对角化的协方差矩阵。

## chap 02 常用概率分布

### 二项分布

#### 1. 定义

$N$  次**独立**测量(伯努利试验)，每次**只有**成功(概率始终为  $p$ ) 或失败(概率为  $1 - p$ ) 两种可能。

定义随机变量  $n$  为成功的次数，则  $n$  服从二项分布：

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

记作  $B(N, p)$ 。

## 2. 平均值与方差

随机变量  $n$  的平均值

$$E[n] = Np$$

方差

$$V[n] = Np(1-p)$$

例：计算探测效率与探测效率的不确定度。

探测效率

$$\varepsilon = \frac{N'}{N}$$

假设探测到的粒子数遵循二项分布， $(E(N') = N\varepsilon)$

则

$$\Delta\varepsilon = \frac{\Delta N}{N} = \frac{\sqrt{N\varepsilon(1-\varepsilon)}}{N}$$

## 多项分布

- 测量的结果不止一种，将其中一种视为成功，其他视作失败，则变为二项分布。

## 泊松分布

- 泊松分布可以看作二项分布的极限近似，在  $N \rightarrow \infty, p \rightarrow 0, N \rightarrow \nu$  的情况下，得到：

$$f(n; \nu) = \frac{\nu^n}{n!} e^{-\nu}$$

- 均值

$$E[n] = \nu$$

方差

$$V[n] = \nu$$

泊松分布的特点：  $n \pm \sqrt{n}$

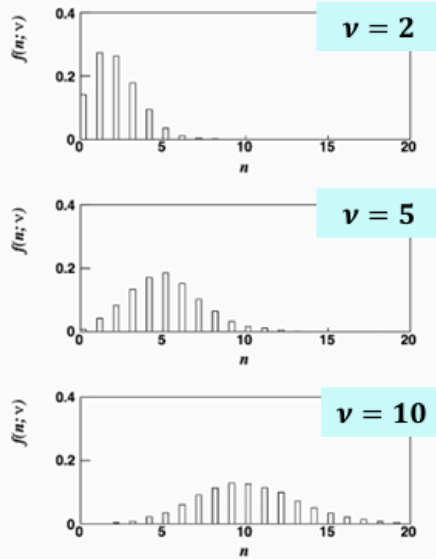
- 泊松过程——泊松分布的产生

条件：

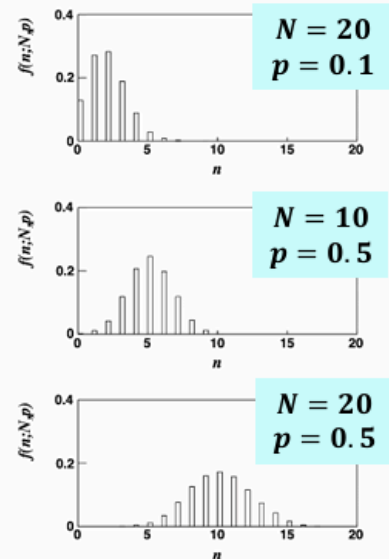
- 单位时间事件发生的概率固定
- 同时发生两次事件几乎不可能（每次事件发生的概率低）
- 每次事件发生是相互独立的

例：

## 泊松分布



## 二项分布



哪两个最相似？

$N$ 越大， $p$ 越小，泊松分布与二项分布越接近！

例：

假设某人站在路边想搭便车。

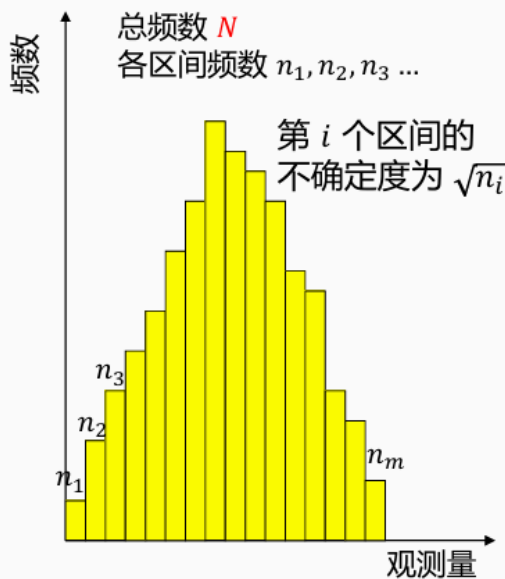
- (1) 每分钟过路的汽车数服从泊松分布，平均每分钟过路一辆，计算一小时通过的汽车数的分布。
- (2) 每分钟过路的汽车数服从泊松分布，平均每分钟过路一辆。假设每辆车让搭便车的概率为 1%，并相互独立。计算过了60辆车后还未能搭上车的概率。

解：

对于问题1，分布显然满足泊松分布的条件（时间可以无限细分，时间无限短的时候通过的汽车数目趋于零）。

对于问题2，分布满足二项分布的条件。

例：



## 直方图可看成

1. 事例总数  $N$  服从泊松分布，每个区间频数  $\vec{n} = (n_1, n_2, \dots, n_m)$  服从多项分布 ( $\sum_{i=1}^m n_i = N$ )
2. 或者，每个区间相互独立的泊松分布

总频数  $N$  的不确定度:  $\Delta N = \sqrt{N}$

或由独立的  $n_i$  的不确定度传递得

$$\begin{aligned} (\Delta N)^2 &= (\Delta n_1)^2 + (\Delta n_2)^2 + \dots + (\Delta n_m)^2 \\ &= n_1 + n_2 + \dots + n_m \\ &= N \end{aligned}$$

注意：当  $N < 5$  时不确定度估计会有很大的偏差。

## 均匀分布

### 1. 函数形式

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha < x < \beta \\ 0, & \text{else.} \end{cases}$$

### 2. 均值与方差

$$\begin{aligned} E[x] &= \frac{\alpha + \beta}{2} \\ V[x] &= \frac{(\beta - \alpha)^2}{12} \end{aligned}$$

均匀分布是蒙特卡洛方法实现的基础！

## 指数分布

### 1. 函数形式

$$f(x; \xi) = \begin{cases} \frac{1}{\xi} e^{-x/\xi}, & x > 0 \\ 0, & \text{else} \end{cases}$$

### 2. 期望与方差

$$E[x] = \xi, V[x] = \xi^2$$

指数分布没有记忆性，可以任意选取开始位置！

## 高斯分布

### 1. 函数形式

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

记作  $N(\mu, \sigma^2)$ 。



## 2. 期望值

$$E[x] = \mu$$

## 方差

$$V[x] = \sigma^2$$

高斯分布是对连续变量概率分布的很好近似。

## 中心极限定理

高斯分布的重要性在于，如果一个随机变量是由大量小贡献随机变量之和构成的，那么它往往服从高斯分布。

对于 $n$ 个独立的随机变量 $x_i$ ，如果每个 $x_i$ 的方差存在，那么这些变量之和构成的随机变量

$$y = \sum_{i=1}^n x_i$$

在 $n \rightarrow \infty$ 的极限下，服从高斯分布 $N(\mu, \sigma^2)$ ，其中

$$\mu = \sum_{i=1}^n \mu_i, \quad \sigma^2 = \sum_{i=1}^n \sigma_i^2$$

测量不确定度通常来自很多贡献之和，所以重复测量的值可以看作服从高斯分布的随机变量。

注意，变量 $y$ 的分布形式与 $x$ 原来的分布形式无关。

多维高斯分布通过协方差矩阵定义：

$$f(\vec{x}; \vec{\mu}, V) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} e^{\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})}$$

对于二维高斯分布可以证明

## 卡方分布

### 1. 定义

对于相互独立的高斯随机变量 $x_1, \dots, x_n$ ，定义

$$z = \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

得到 $z$ 满足的分布：

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

记作 $\chi^2(n)$

### 2. 均值

$$E[z] = n$$

方差

$$V[z] = 2n$$

例：最小二乘法的拟合优度检验

$$y = f(x) = kx + b$$

改变 $k$ 与 $b$ 的取值，使得

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - f(x_i))^2}{\sigma_i^2}$$

极小。此时我们认为估计的结果是最好的。

不断重复实验，得到一系列卡方的值，这个值满足自由度为 $(n - 2)$ 的卡方分布。

拟合优度：

P值（假设检验）判断有多大的概率出错（理论模型与实际情况是否可以认为相符）

## 柯西分布

### 柯西（布莱特-魏格纳）分布

连续随机变量  $x$  的柯西分布为

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$$

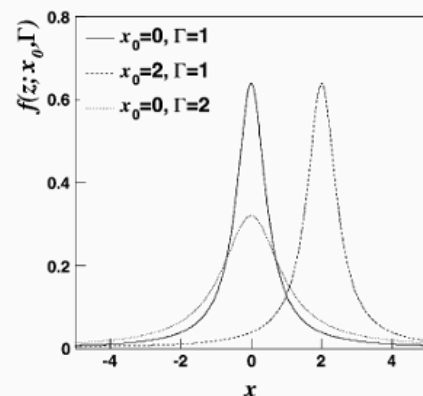
它是布莱特-魏格纳分布的特例

$$f(x; \Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2}$$

$x_0$ ：模（most probable value）

$\Gamma$ ：半高全宽（full with half maximum）

$E[x]$ 没有好的定义， $V[x] \rightarrow \infty$



常用于描述“共振态”粒子的不变质量分布，例如 $\rho, K^*, \phi^0, \dots$   
 $\Gamma$  = 衰变率（平均寿命的倒数）

## 朗道分布

速度为  $\beta = v/c$  的带电粒子穿过厚度为  $d$  的薄物质层，其能量损失  $\Delta$  服从朗道分布：

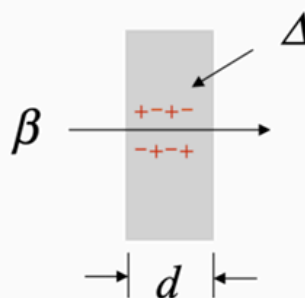
$$f(\Delta; \beta) = \frac{1}{\xi} \phi(\lambda),$$

$$\phi(\lambda) = \frac{1}{\pi} \int_0^\infty e^{-u \ln(u) - \lambda u} \sin \pi u \, du$$

$$\lambda = \frac{1}{\xi} \left[ \Delta - \xi \left( \ln \frac{\xi}{\epsilon'} + 1 - \frac{1}{\sqrt{1 - \beta^2}} \right) \right]$$

$$\xi = \frac{2\pi N_A e^4 Z^2 \rho \Sigma Z}{m_e c^2 \Sigma A} \frac{d}{\beta^2}$$

$$\epsilon' = \frac{I^2 (1 - \beta^2) \exp(\beta^2)}{2 m_e c^2 \beta^2}$$

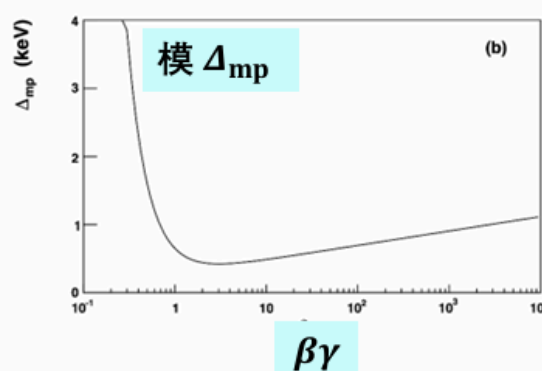
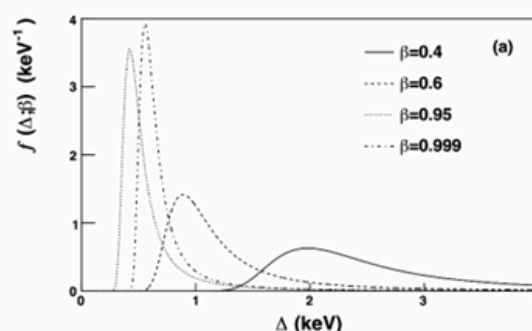


厚度  $d$  增大时，趋于正态分布。

常用于描述粒子的电离能损或能量沉积。

朗道分布有长的“朗道尾部”  
→ 所有的矩都发散

模 (most probable value)  
对  $\beta$  很敏感  
→ 可用于粒子鉴别(PID)



## 贝塔分布

连续随机变量  $0 < x < 1$  的贝塔分布为：

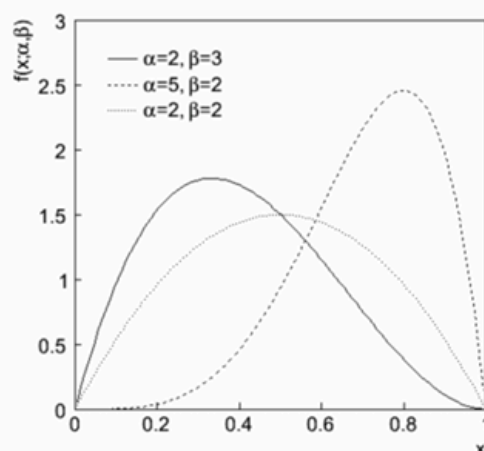
$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

$$E[x] = \frac{\alpha}{\alpha + \beta}$$

简记  $Be(\alpha, \beta)$

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

$$Be(1,1) = U(0,1)$$



常用来表示只在某个有限区间非零的连续随机变量。

$\alpha = \beta = 1$  变为均匀分布。

## 伽马分布

连续随机变量  $x$  的伽马分布为：

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

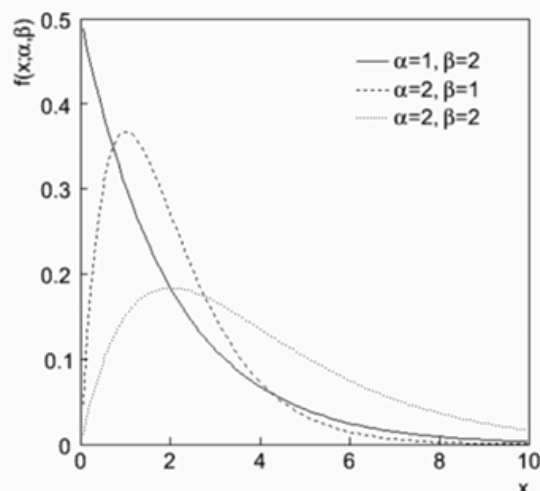
$$E[x] = \alpha\beta$$

简记  $Ga(\alpha, \beta)$

$$V[x] = \alpha\beta^2$$

$$Ga(1, \xi) = Exp(\xi)$$

$$Ga(n/2, 2) = \chi^2(n)$$



常用来表示在  $[0, \infty]$  内不为零的连续随机变量。

例：  $n$  个指数分布随机变量的和，泊松过程第  $n$  个事件发生的时间。

## 学生氏分布

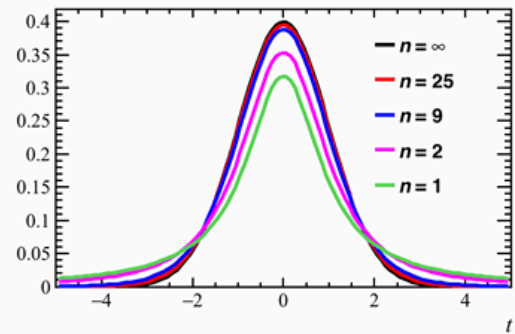
连续随机变量  $x$  的自由度为  $\nu$  的学生氏分布为：

$$f(x; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$E[x] = 0 \quad (\nu > 0)$$

$$V[x] = \frac{\nu}{\nu - 2}$$

简记  $t(\nu)$



$\nu$ ：自由度的数目（可以不是整数）

$t(1)$  = 柯西分布

$t(\infty) = N(0,1)$

## chap 03 蒙特卡罗方法 Monte Carlo Method

### 方法简介

#### 应用场景

1. 对于分布不方便解析求解；
2. 没有必要进行解析求解。
3. 处理高维数值定积分方面具有独特的优势。

#### 使用方法

- 使用MC方法计算定积分：
  1. 生成随机数序列  $r_1, \dots, r_n$ ；
  2. 按照函数  $f(x)$  生成随机数序列  $x_1, \dots, x_n$ ；
  3. 根据函数所占比例计算积分值。
- 使用MC方法进行数据模拟。

### 随机数的产生

- 用物理方法产生真随机数：不可重复，产生速度慢
- 用数学方法产生伪随机数：可重复，产生速度快

随机数产生子

# 函数变换法与舍选法

## 函数变换法

考虑随机变量 $r = F(x) = \int f(x)dx$

$$g(r)dr = f(x)dx$$

$$g(r) = 1$$

故知随机变量 $r$ 为均匀分布。

取 $x = F^{-1}(r)$ ，随机生成0-1之间的均匀分布，就可以得到 $x$ 的分布。

对于指数分布 $f(x) = e^{-x}$ ，进行处理得到：

$$r = F(x) = 1 - e^{-x}$$

$$x = -\ln(1 - r)$$

实现代码如下：

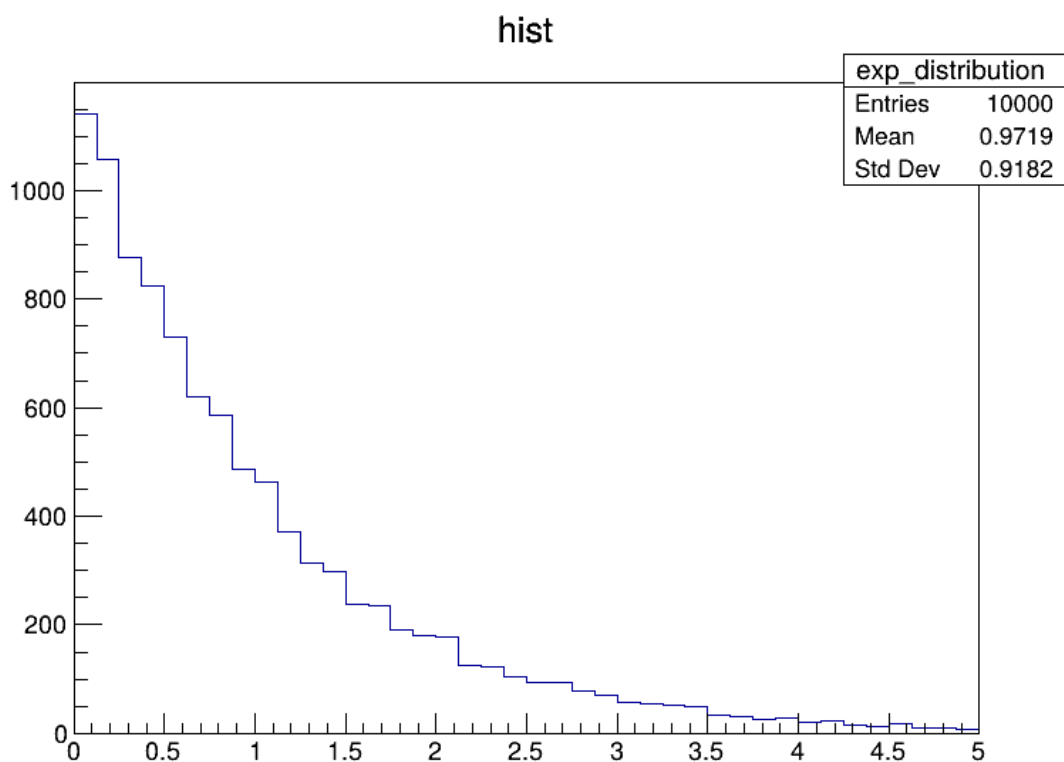
```
import ROOT
import math
rnd = ROOT.TRandom3()

hist = ROOT.TH1F('exp_distribution', 'hist', 40, 0, 5)

for i in range(10000):
    hist.Fill(-math.log(1-rnd.Rndm()))

canvas = ROOT.TCanvas('canvas', 'Canvas', 800, 600)
hist.Draw()

canvas.SaveAs('exp_distribution.png')
```



# 蒙特卡罗方法在物理实验中的应用

## 习题

1. 公平的抽签：袋中有 $r$ 个红球与 $b$ 个黑球，现任意不放回地一一摸出，求事件第 $k$ 次摸出红球的概率。

解：

$$P(K) = \frac{r}{r+b}$$

样本空间

$$n_{\Omega} = (r+b)!$$

成功的次数

$$n_A = r \cdot (r+b-1)!$$

概率

$$P(A) = \frac{n_A}{n_{\Omega}} = \frac{r}{r+b}$$

2. 生日问题：求任意 $n$  ( $n \leq 365$ ) 个人中，至少有两个人生日相同的概率。

解：

将事件转化成对立事件进行计算。 $n$ 个人生日各不相同的概率为

$$P(\bar{A}) = \frac{n_{\bar{A}}}{n_{\Omega}} = \frac{\frac{365!}{(365-n)!}}{365^n} = \left(1 - \frac{1}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right)$$
$$P(A) = 1 - P(\bar{A})$$

3. 约会问题：小明和小红二人约定在 $[0, T]$ 时段内去未名湖会面，规定先到者等候一段时间 $t$  ( $t \leq T$ )再离开。试求小明和小红将会面的概率。

解：

两人出现的时间范围在 $[0, T]$ 之间均匀分布，在两者事件相差的绝对值为 $t$ 之内的情形可以相遇。

$$P(A) = \frac{n_A}{n_{\Omega}} = 1 - \frac{(T-t)^2}{T^2}$$

4. 贝特朗奇论：1889年，法国数学家贝特朗提出下述几何概率问题，并给出三种不同的答案。这就使有的人对当时的概率论中的一些概念与方法产生怀疑，因此被称作“奇论”。请思考：在单位圆上任作一弦，求弦长大于 $\sqrt{3}$ 的概率。

解：

这与任意的方式具有相关性，对于任意的理解不同会导致结果不同。

可以任意先选择圆周上的一个点，然后任意选择一个角度；或者任意选择弦的中点的位置。

5. 有 $n$ 个人参加聚会，每个人都随身带了一份礼物，之后把礼物混合放在一起。会后每个人随机拿取一份礼物。求至少有一个人拿到了自己的礼物的概率。

解：

事件 $A_i$ 代表第 $i$ 个人拿到属于自己的礼物，则

$$P(\cup_{i=1}^n A_i) = \frac{C_n^1}{A_n^1} - \frac{C_n^2}{A_n^2} + \cdots + (-1)^{n+1} \frac{C_n^n}{A_n^n} = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \frac{1}{n!}$$

6. 有n个人参加聚会，每个人都随身带了一份礼物，之后把礼物混合放在一起。会后每个人随机拿取一份礼物。求恰好有 k 个人拿到了自己的礼物的概率。

解：

$$P = \frac{1}{k!} \sum_{i=1}^{n-k} \frac{(-1)^i}{i!}$$

7. 二项分布，泊松分布，高斯分布的特点以及转化关系

8.

假设有一系列独立同分布的随机变量  $X_i \sim U(0, 1)$ ,  $i = 1, 2, \dots$ 。定义

$$N_a = \min \left\{ k : \prod_{i=1}^k X_i < a \right\} \quad (0 < a < 1).$$

求  $N_a$  的分布。

解：

$$P(N_a = n_0) = \int_a^1 dx_1 \int_{a/x_1}^1 dx_2 \cdots \int_{a/x_1 \cdots x_{n_0-2}}^1 dx_{n_0-1} \int_0^{a/x_1 \cdots x_{n_0-1}} dx_{n_0} = \frac{a}{(n_0 - 1)!} (-\ln a)^{n_0-1}$$

9.  $Z = XY, X + Y, Y/X, \min X, Y, \max X, Y$  的随机变量分布

实例：计算存在两种衰变方式时的粒子寿命。

分析：

设粒子的寿命为  $x$ ，则粒子寿命小于等于  $a$  的概率（累积分布函数）：

$$H(a) = P(x \leq a) = 1 - P(x > a) = 1 - P(x_1 > a)P(x_2 > a) = 1 - [1 - F(a)][1 - G(a)]$$

同理可以计算出

$$P(x \leq a) = P(x \leq a_1)P(x \leq a_2)$$

下面对实例进行处理：

$$f(x) = \lambda_1 e^{-\lambda_1 x}$$

$$g(x) = \lambda_2 e^{-\lambda_2 x}$$

$$F(x) = 1 - e^{-\lambda_1 x}$$

$$G(x) = 1 - e^{-\lambda_2 x}$$

$$H(x) = 1 - e^{-(\lambda_1 + \lambda_2)x}$$

$$h(x) = (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)x}$$

得到粒子寿命：

$$\tau = \frac{1}{\frac{1}{\tau_1} + \frac{1}{\tau_2}}$$

例：

$$X, Y \sim U(0, 1), Z = \min(X, Y) / \max(X, Y)$$



令  $u = \min, v = \max$

$$f(u, v) = 2, 0 \leq u \leq v \leq 1$$

10. 某粒子在 A 气体中的平均自由程为  $L_A$ , 在 B 气体中的平均自由程为  $L_B$ , 求它在 AB 等量混合气体中的平均自由程

11. 习题 1.13, 1.14, 1.15 讲解

相关性: **线性相关**, 即:  $Y = X^2$  的关系不能得到验证

习题 1.13. 假设  $X$  和  $Y$  是两个连续的随机变量, 且其方差有限。证明相关系数  $\rho \equiv \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \pm 1$  的充要条件是  $X$  与  $Y$  几乎处处有线性关系, 即存在常数  $a \neq 0$  和  $b$  使得  $Y = aX + b$ 。

构造统计量  $U = Y - aX - b$ , 计算  $\sigma[U], E[U]$ , 证明二者均为零即可。

不妨取:

$$b = E$$

12. 重期望法则

$$E[E[X|Y]] = E[X]$$

13. Stein's lemma

14. 一名矿工被困在一个有三扇门的矿井中。如果选择 1 号门, 3 小时后可以到达安全区域; 如果选择 2 号门, 5 小时后会回到原来的位置; 如果选择 3 号门, 7 小时会回到原来的位置。如果假设这名矿工对走过的门没有记忆, 每次都以相等的概率选择每扇门, 那么他平均需要多少时间才能到达安全区域?

$$E[X|Y = 1] = 3$$

$$E[X|Y = 2] = 5 + E[X]$$

$$E[X|Y = 3] = 7 + E[X]$$

$$E[X] = E[E[X|Y]] = \frac{1}{3}(E[X|Y = 1] + E[X|Y = 2] + E[X|Y = 3])$$

可以解出  $E[X]$ 。

15. Skewness (偏度) 与 Kurtosis (峰度)

## chap 04 统计检验

### 假设、检验、显著水平、功效、临界域

- 假设 hypothesis (pl. hypotheses)

假设  $\rightarrow$  物理理论: 预测数据出现的概率, 通过实验进行检验。

简单假设: 给出的假设  $H_0$  可以完全确定概率密度函数  $f(x|H)$ , 称为假设的似然值 (likelihood)  $L(x|H)$ 。

假设的似然值: 假设成立时可观测量  $x$  出现的概率  $f(x|H)$ , 即  $x \sim f(x|H)$ 。

符合假设: 仍然包含未确定的参数。

- 检验

零假设 (null hypothesis)  $H_0$  与备择假设 (alternative hypothesis)  $H_1$ 。

对简单假设  $H_0$  的检验: 在假设  $H_0$  正确的情形下, 在临界域  $W$  内观测到结果的 (小) 概率不超过  $\alpha$ , 即

$$P(x \in W | H_0) \leq \alpha$$

$\alpha$ : 显著性水平/检验大小

一般情形下 $\alpha$ 的取值范围在 $10^{-6} \sim 10^{-7}$ 之间。

临界域: 拒绝域, 补集称为接受域。

- 第一类错误与第二类错误
  - 第一类错误: 假设 $H_0$ 为真, 但是假设被拒绝。这类错误犯错的最大概率 $P(x \in W|H_0)$ 就是显著性水平 $\alpha$ 。
  - 第二类错误: 假设 $H_0$ 为假, 但是假设 $H_1$ 为真。这类错误的概率为 $P(x \in S - W|H_1) = \beta$ ,  $1 - \beta$ 称为功效 (检验对于排除备择假设 $H_1$ 的效力)。
- 一般而言, 调整临界域不可能使得 $\alpha$ 与 $\beta$ 同时减小, 但是增大样本量可以实现这一点。
- 临界域的选择: 单边检验与双边检验

## 信号本底的甄别

对于一次测量的实验结果通常不止是一个数值, 而是一个向量 $\vec{x} = (x_1, x_2, \dots, x_n)$ , 我们需要构造检验统计量来检验假设是否正确。

### 检验统计量 $t$

构造一个标量检验统计量 $t = t(x_1, x_2, \dots, x_n)$ , 对假设进行检验。检验统计量也可以直接取为矢量 $\vec{t} = \vec{t}(x_1, x_2, \dots, x_n)$ 。

### 纯度

事例选择的纯度: 信号事例被正确分类的概率。(换言之, 被判断为信号的所有事例中真正为信号的事例所占的比例。)

$$P(s|x \in W) = \frac{P(x \in W|s)P(s)}{P(x \in W|s)P(s) + P(x \in W|b)P(b)}$$

### 检验效率

信号事例被接受的概率 (被接受的信号事例占有所有信号事例的比例)

即:

$$\varepsilon_s = \int f(t|s) dt = 1 - \alpha$$

## Neyman-Pearson引理与检验统计量的构造

对于多维统计检验量, 应当如何选取最佳的临界域?

### Neyman-Pearson引理

在给定效率的条件下, 想要得到最高纯度的信号样本, 接受域的选择可以为:

$$\frac{g(\vec{t}|H_0)}{g(\vec{t}|H_1)} > c$$

检验统计量可以取为

$$r = \frac{g(\vec{t}|H_0)}{g(\vec{t}|H_1)}$$

引理的证明(from <https://chatgpt.com>)

下面给出内曼-皮尔逊引理 (Neyman-Pearson定理) 的一个经典证明。假设我们要在两种简单假设之间做出检验:

- 零假设  $H_0$ : 随机变量  $X$  的概率密度 (或概率质量) 函数为  $p(x)$ ;
- 备择假设  $H_1$ :  $X$  的密度为  $q(x)$ 。

我们希望构造一个检验函数  $\phi(x)$ , 其中

$$\phi(x) = \begin{cases} 1, & \text{拒绝 } H_0, \\ 0, & \text{接受 } H_0. \end{cases}$$

(为简单起见, 也可以允许在边界点上取介于 0 和 1 的值以调整精确水平。)

定理的结论是: 在所有显著性水平不超过  $\alpha$  的检验中, 以似然比为依据的检验

$$\phi^*(x) = \begin{cases} 1, & \text{当 } \frac{q(x)}{p(x)} > k, \\ 0, & \text{当 } \frac{q(x)}{p(x)} < k, \end{cases}$$

(在  $\frac{q(x)}{p(x)} = k$  的点上可能随机化, 使得  $E_{H_0}[\phi^*(X)] = \alpha$ ) 具有最大的统计功效, 也就是说对于任何其它检验  $\phi(x)$  (其第一类错误率不超过  $\alpha$ ), 都有

$$E_{H_1}[\phi^*(X)] \geq E_{H_1}[\phi(X)].$$

下面是证明步骤:

## 1. 构造最优检验

设我们构造的似然比检验函数为

$$\phi^*(x) = \begin{cases} 1, & \frac{q(x)}{p(x)} > k, \\ \gamma(x), & \frac{q(x)}{p(x)} = k, \\ 0, & \frac{q(x)}{p(x)} < k, \end{cases}$$

其中  $k > 0$  和  $\gamma(x) \in [0, 1]$  取值, 使得该检验的显著性水平满足

$$E_{H_0}[\phi^*(X)] = \int \phi^*(x)p(x) dx = \alpha.$$

## 2. 对任一水平不超过 $\alpha$ 的检验比较功效

设任一其它检验函数为  $\phi(x)$ , 满足

$$E_{H_0}[\phi(X)] = \int \phi(x)p(x) dx \leq \alpha.$$

我们要证明

$$E_{H_1}[\phi^*(X)] - E_{H_1}[\phi(X)] \geq 0.$$

为此, 考虑积分

$$I = \int [\phi^*(x) - \phi(x)] [q(x) - k p(x)] dx.$$

**注意：**根据  $\phi^*(x)$  的构造，

- 当  $\frac{q(x)}{p(x)} > k$  时， $\phi^*(x) = 1$ ；此时  $q(x) - k p(x) > 0$ ，故  $\phi^*(x) - \phi(x) \geq 0$  时乘上正数得非负；
- 当  $\frac{q(x)}{p(x)} < k$  时， $\phi^*(x) = 0$ ；此时  $q(x) - k p(x) < 0$ ，而  $\phi^*(x) - \phi(x) = -\phi(x) \leq 0$ ，乘上负数依然非负；
- 在  $\frac{q(x)}{p(x)} = k$  的点上，由于可以通过随机化调整，积分对这一处的贡献为零（或可以忽略）。

因此可以证明  $I \geq 0$ 。

### 3. 将积分 $I$ 与各检验的期望联系起来

将  $I$  展开得：

$$I = \int \phi^*(x) q(x) dx - k \int \phi^*(x) p(x) dx - \left[ \int \phi(x) q(x) dx - k \int \phi(x) p(x) dx \right].$$

也就是说，

$$E_{H_1}[\phi^*(X)] - k E_{H_0}[\phi^*(X)] - [E_{H_1}[\phi(X)] - k E_{H_0}[\phi(X)]] \geq 0.$$

由于按构造  $E_{H_0}[\phi^*(X)] = \alpha$  而  $E_{H_0}[\phi(X)] \leq \alpha$ ，则

$$E_{H_1}[\phi^*(X)] - k\alpha \geq E_{H_1}[\phi(X)] - k E_{H_0}[\phi(X)] \geq E_{H_1}[\phi(X)] - k\alpha.$$

从而可得

$$E_{H_1}[\phi^*(X)] \geq E_{H_1}[\phi(X)].$$

这说明在所有显著性水平不超过  $\alpha$  的检验中，似然比检验  $\phi^*$  的功效是最大的。

### 4. 总结

以上证明利用了如下关键思想：

- 将两种检验函数之差乘以  $q(x) - k p(x)$  后积分，利用似然比检验的构造保证该乘积非负。
- 从而证明了似然比检验在  $H_1$  下的期望（功效）至少不低于任何其它同水平检验。

这便是内曼-皮尔逊引理的证明，其结论说明在两个简单假设的情形下，基于似然比的检验是均匀最优的，即在控制第一类错误率（显著性水平）的条件下，其功效最大。

### 参考

本证明的思路与标准教材中给出的证明类似，例如参见

- [维基百科：内曼-皮尔逊引理](#)
- [知乎：Neyman-Pearson基本引理证明](#)

希望这个证明能帮助你理解内曼-皮尔逊定理的基本原理和数学推导过程。

## 线性检验统计量、费舍尔(Fisher)甄别函数

$$J(\mathbf{a}) = \frac{\mathbf{a}^T B \mathbf{a}}{\mathbf{a}^T W \mathbf{a}}$$

注意到这是关于 $\mathbf{a}$ 的齐次函数，所以需要在固定 $\mathbf{a}$ 的模长的情况下进行偏导数运算。

不妨假设  $\mathbf{a}^T W \mathbf{a} = 1$ ，得到

$$J(\mathbf{a}) = \mathbf{a}^T B \mathbf{a} - \lambda(\mathbf{a}^T W \mathbf{a} - 1)$$

对其求偏导数  $\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}}$  得到

$$B \mathbf{a} = \lambda W \mathbf{a}$$

$$W^{-1} B \mathbf{a} = \lambda \mathbf{a}$$

即 $\mathbf{a}$ 是矩阵 $W^{-1} B$ 的特征向量。

再注意到

$$\begin{aligned} B_{ij} a_j &= (\mu_0 - \mu_1)_i \boxed{(\mu_0 - \mu_1)_j a_j} = \lambda_a (\mu_0 - \mu_1)_i \\ &= \lambda_a \end{aligned}$$

最终得到：

$$\mathbf{a} \propto W^{-1}(\mu_0 - \mu_1)$$

在高斯分布的情形下，Fisher检验统计量给出的结果与Neymann-Pearson引理的结果相同。在其他情形下并不相同，但是可以通过做变换实现更好的结果。

(待续：高斯分布情形下证明相同)

神经网络：

- 单层感知器：

输入层->输出层

# 单层感知器

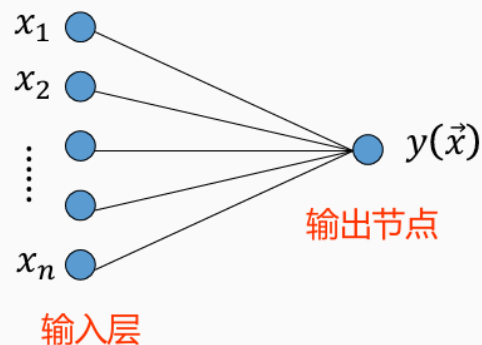
利用非线性函数 $y(\vec{x})$ 定义判别量：
$$y(\vec{x}) = h\left(\omega_0 + \sum_{i=1}^n \omega_i x_i\right)$$

其中 $h$ 是非线性的单调激活函数，例如，逻辑S型函数

$$h(x) = (1 + e^{-x})^{-1}$$

如果激活函数是单调的，得到的 $y(\vec{x})$ 等价于原始的线性判别量。

这是“推广的线性模型”的一个例子，称为单层感知器。



- 多层感知器：

输入层->隐藏层->输出层

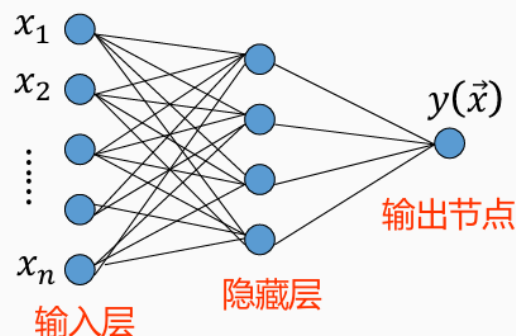
## 多层感知器 (MLP)

利用同样的想法定义，既可以定义输出 $y(\vec{x})$ ，也可以定义构成“隐藏层”的一组变换输入量 $\varphi_1(\vec{x}), \dots, \varphi_m(\vec{x})$ ：

权重中的上标  
表示第几层

$$\varphi_i(\vec{x}) = h\left(\omega_{i0}^{(1)} + \sum_{j=1}^n \omega_{ij}^{(1)} x_j\right)$$

$$y(\vec{x}) = h\left(\omega_{10}^{(2)} + \sum_{j=1}^m \omega_{1j}^{(2)} \varphi_j(\vec{x})\right)$$



这种构造称为多层感知器，是基本的神经网络模型；很容易推广到更多隐藏层。

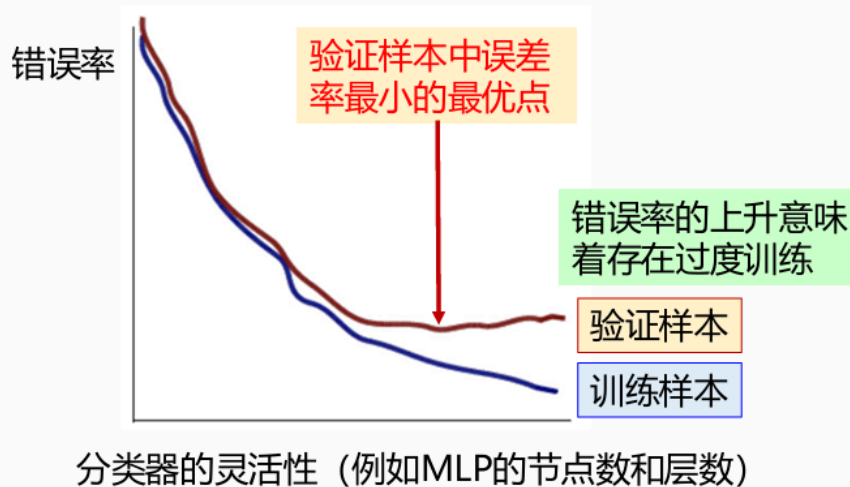
确定中间参数是机器学习核心的内容！

- 过度训练(overfitting)

参数过多，对训练数据的符合非常好，但是反而不能准确预测其他数据的结果。

# 监控过度训练

如果监控验证样本和训练样本中误鉴别事例的比率（或者，误差函数 $E(\omega)$ ），当决策边界变得更灵活时，两个样本的误鉴别比率通常都会下降：



- 输入变量选择

如果输入变量的相关性较强，常常对结果的影响不大。如， $x^2, x^4$ 的结果常常区别不是很大，一般情形下不需要保留高阶矩。

## 检验拟合优度、p值的定义与应用

### p值的定义

对于给定的简单假设 $H_0$ ，我们可以预测出各个结果出现的概率。实验中的结果是其中一种，考虑预期结果不好于实验结果的概率，定义为 $p$ 值。即：

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

例子：在抛硬币实验中，一共抛掷20次硬币，实验结果为17次正面向上。但是，实验预期的最有可能出现的结果为10。所以，与实验结果相比，更糟糕或者一样糟糕的预期结果为0, 1, 2, 3, 17, 18, 19, 20。根据二项式分布的计算可以得到：

$$p = P(N = 0, 1, 2, 3, 17, 18, 19, 20) = 0.0026$$

实验中，想要检验假设是否正确，可以重复实验，检测这些极端事例出现的概率是否与 $p$ 值相符。

（假设 $H_0$ 为真的概率并不是 $p$ 值）

对于这一实验结果，能不能认为假设 $H_0$ 并不成立？换言之，有多少信心（Bayes主观概率的角度）认为原假设 $H_0$ 正确或者错误？这是假设检验中的重要问题。

### $H_0$ 的检验

由 $p$ 值的定义，容易知道：

$$P(p_0 \leq \alpha) = \alpha$$

证明：

$$p(t) = \int_t^\infty f(x)dx$$

$g(p) = 1$  (累积分布的随机变量是均匀分布)

$$\alpha = \int_{t_0}^\infty f(x)dx$$

$$P(p_0 \leq \alpha) = \int_0^\alpha g(p)dp = \alpha$$

## 信号观测的显著程度

考虑泊松计数实验（假设信号与本底的测量结果均遵循泊松分布），设本底数的均值为 $b$ ，信号的均值为 $s$ ， $H_0$ 为没有信号产生，则：

$$P(n|H_0) = \frac{b^n}{n!}e^{-b}$$

假设 $b = 0.5$ ,  $n = 5$ ，问是否可以认为发现了信号？

解：

计算 $p$ 值：

$$p = \sum_{n=5}^{\infty} \frac{b^n}{n!}e^{-b} = 1.7 \times 10^{-4}$$

显著性 $Z$ ：高斯变量在一个方向涨落得到相同 $p$ 值所对应的标准差的倍数。

## Pearson $\chi^2$ 检验

### $\chi^2$ 统计量的定义

实验观测值： $\vec{n} = (n_1, n_2, \dots, n_N)$

理论预期值： $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_N)$

$$\chi^2 = \sum_{i=1}^N \frac{(n_i - \mu_i)^2}{\sigma_i^2}$$

## chap 05 参数估计的一般概念

### 估计量

#### 问题

对于测量结果 $\vec{x}_i$ ，对于概率密度函数的参数有估计 $\hat{\theta}_i$ ，问，应该如何对数据进行处理，得到最佳的估计结果？即，如何通过 $\hat{\theta}$ 估计 $\theta$ ？

如果用更加数学的语言进行表述，在实际操作过程中，我们并不知道概率密度函数 $f(\vec{x}; \theta)$ 的具体形式，只能通过 $n$ 次测量的结果 $\vec{x}_1, \dots, \vec{x}_n$ 对概率密度函数的形式进行估计。

测量结果 $\vec{x}$ 称为**样本**，估计量是样本的函数。经过多次测量，可以得到估计量的分布 $g(\hat{\theta}, \theta)$ 。



## 对估计量的要求

- 相合性：在样本容量无限大的极限下，估计量趋近于真值。
- 无偏性： $b = E[\hat{\theta}] - \theta$  越小越好。
- 有效性：方差  $V[\hat{\theta}]$  越小越好。

定义均方误差MSE

$$MSE = E[(\hat{\theta} - \theta)^2] = V[\hat{\theta}] + b^2$$

对于偏倚为0，方差最小的估计量称为有效估计量。在  $n \rightarrow \infty$  的情形下偏倚趋近于0的估计量称为渐近有效估计量。

然而，在实际操作时，减小估计量的方差与减小估计量的偏移是相互矛盾的，要进行权衡。

## 样本均值

### 问题

对于随机变量  $X$ ，具有未知的概率密度函数  $f(x)$ ，想要通过有限次独立测量得到  $E[X] = \int f(x)dx$  的估计量  $\hat{E}[X]$ ，问应该如何构造函数得到的这一估计值？

显然，这一估计量为无参数估计量。

### 函数构造

构造函数

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

### 函数性质检验

相合性

$$\lim_{n \rightarrow \infty} \mu = \bar{x}$$

(大数弱定律保证)

偏倚

$$E[\mu] = E[x] = \bar{x}$$

估计量的方差

$$\begin{aligned} V[\mu] &= E[\mu^2] - E[\mu]^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

## 样本方差

### 函数构造

假设分布的均值与方差都不能提前确定精确的值，构造函数：

$$s^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n\mu^2 \right)$$

注意，这里的公式与样本的方差不同，分母减去1的作用是，保证估计量是无偏估计量。

## 函数性质检验

### 方差估计量的偏倚

$$E[s^2] = V[X]$$

### 方差估计量的方差

$$s^2 \text{ 的方差} \quad V[s^2] = \frac{1}{n} \left( \mu_4 - \frac{n-3}{n-1} \mu_2^2 \right)$$

## 样本协方差

### 协方差与相关系数的估计量

协方差  $V_{xy} = \text{cov}[x, y]$  的估计量为

$$\hat{V}_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{n}{n-1} (\overline{xy} - \bar{x} \cdot \bar{y}) \quad (\text{无偏})$$

相关系数  $\rho = V_{xy}/(\sigma_x \sigma_y)$  的估计量为

$$\hat{\rho} = r = \frac{\hat{V}_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left( \sum_{i=1}^n (x_i - \bar{x})^2 \sum_{k=1}^n (y_k - \bar{y})^2 \right)^{1/2}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}$$

$r$  有偏倚。但是当  $n \rightarrow \infty$  时，该偏倚趋于零。

一般而言，概率密度  $g(r; \rho, n)$  形式复杂；对于高斯变量  $x, y$

$$E[r] = \rho - \frac{\rho(1-\rho^2)}{2n} + \mathcal{O}(n^{-2}); \quad V[r] = \frac{1}{n}(1-\rho^2)^2 + \mathcal{O}(n^{-2})$$

## chap 06 极大似然法

### 似然函数、极大似然估计量

#### 最大似然法的基本思想

假设函数分布的形式  $f(x; \theta)$ ，对于多次测量  $\vec{x} = (x_1, \dots, x_n)$ ，如何估计  $\vec{\theta} = (\theta_1, \dots, \theta_m)$ ?

$$P(x_i \in [x_i, x_i + dx_i]) = \prod_{i=1}^n f(x_i; \theta) dx_i$$

我们直观上有理由相信，观测到的结果具有相对较高的概率值。

定义似然函数：

$$L(\vec{x}, \theta) = \prod_{i=1}^n f(x_i; \theta)$$

对数似然函数：

$$\ln L(\vec{x}, \theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

给定一组实验数据，似然函数就是估计量的函数。

最大似然估计量 (Maximum Likelihood Estimator) 定义为：使得似然函数取最大值的估计量，满足：

$$\frac{\partial L(\vec{\theta})}{\partial \theta_i} = 0$$

准确来说，上式求出的是平稳值的位置，一般情况下从中挑选使得L最大的值即可。

为了方便计算，一般情形下，可以选取参数的函数进行处理，即：

$$\frac{\partial \ln L(h)}{\partial h} = 0, h = h(\theta)$$

最大似然法的好处：利用了所有数据的信息，与区间的划分无关，对测量值不做任何的修正。

关于“与分区间是否有关”这一点，可以通过与最小二乘法的对比来理解。对于最小二乘法在分区间的过程中，距离较近的点的值被合并，每个点的作用被削弱。但是最大似然法着重于“计数”的性质，合并区间并不会造成实质性的影响。

最大似然法的缺点：

一般情况下最大似然法给出的结果不会非常糟糕（渐进  $(n \rightarrow \infty)$  有效），但是在个别极端情况下会带来问题（如  $U[X]$ ，估计随机变量的范围）。同时，在参数较多的情形，这一方案的计算难度较大。

## 例子：指数分布与高斯分布的估计量确定

### 指数分布

分布形式：

$$f(t; \tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

给定实验样本之后：

$$\ln L(\tau) = - \sum_{i=1}^n \frac{t_i}{\tau} - n \ln \tau$$

对于实验样本，要求为简单数据样本（意味着各次测量独立同分布）。

由最大似然法：

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n t_i$$

### 估计量的偏倚

由于形式是平均值的分布，所以应该是无偏估计量。

### 另一种指数分布形式的表达

$$f(t; \lambda) = \lambda e^{-\lambda t}$$

显然估计量为：

$$\hat{\lambda} = \frac{1}{\frac{1}{n} \sum_{i=1}^n t_i}$$

但是这一估计量仅仅是渐近无偏的。

使用特征函数的方法可以硬算得到概率密度分布函数，可以验证。

## 高斯分布

$$\ln L(\mu, \sigma^2) = \sum_{i=1}^n \left( \ln \frac{1}{\sqrt{2\pi}} + \frac{1}{2} \ln \frac{1}{\sigma^2} - \frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

得到：

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

显然，均值的估计量是无偏估计量，方差的估计量是渐近无偏估计量。

## 估计量的方差

### 使用MC方法确定估计量的方差

具体而言，将估计值作为真值，进行多次MC模拟，计算数据样本的方差 $s^2$ 近似作为估计量的方差的值。

### 使用解析方法确定估计量的方差

对于常见的分布形式，直接积分计算估计量的方差是可行的，但是一般情形下的计算是困难的。

### 使用RCF边界确定估计量的方差

信息不等式：

$$V[\theta] = \frac{(1 + \frac{\partial b}{\partial \theta})^2}{E[-\frac{\partial^2 \ln L}{\partial \theta^2}]}$$

如果最大似然估计量(ML估计量)为有效估计量，则 $V[\hat{\theta}] \approx \frac{1}{-E[-\frac{\partial^2 \ln L}{\partial \theta^2}]}$ 。

对于高维情形：

# 多维参数的信息不等式：费舍尔信息矩阵

对于  $m$  个参数  $\vec{\theta} = (\theta_1, \dots, \theta_m)$ ，最小方差界由费舍尔信息矩阵给出：

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

信息不等式：  $V - I^{-1}$  是半正定矩阵，其中  $V_{ij} = \text{cov}[\hat{\theta}_i, \hat{\theta}_j]$ 。

因此，  $V[\hat{\theta}_i] \geq (I^{-1})_{ii}$

经常用  $I^{-1}$  近似协方差矩阵，利用  $L$  在最大值处的二阶微分矩阵来估计。

$$(V^{-1})_{ij} = -\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \bigg|_{\vec{\theta} = \hat{\vec{\theta}}}$$

费舍尔信息矩阵

一维：

$$\hat{V}(\hat{\theta}) = \left( -\frac{\partial^2 \ln L}{\partial \theta^2} \right)^{-1} \bigg|_{\theta = \hat{\theta}}$$

## 使用图解法确定估计量的方差

用ML估计量处的值估计不确定度，可以看出：

$$\begin{aligned} \ln L(\theta) &= \ln L(\hat{\theta}) + \frac{\partial \ln L(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2} \frac{\partial^2 \ln L}{\partial \theta^2} \bigg|_{\theta = \hat{\theta}} (\theta - \hat{\theta})^2 \\ &= \ln L(\hat{\theta}) - 0.5 \frac{(\theta - \hat{\theta})^2}{V[\theta]} \end{aligned}$$

可以看出，在图中  $\ln L$  下降为0.5时的半宽度为方差的估计值。

## 扩展的最大似然估计

解决的是样本容量不固定的情形。在实验过程中，样本空间的大小  $n$  也是随机变量（一定时间内衰变事例发生的次数）。容易看出，事例发生的次数（样本空间的大小）满足均值为  $\nu$  的泊松分布。

$$L(\nu; \theta) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^n f(x_i; \theta)$$

## 均值与未知参数相关的情形

$$\begin{aligned} \ln L(\nu; \theta) &= n \ln \nu(\theta) - \nu(\theta) - \ln(n!) + \sum_{i=1}^n \ln f(x_i; \theta) \\ &= c - \nu(\theta) + \sum_{i=1}^n \ln(\nu(\theta) \cdot f(x_i, \theta)) \end{aligned}$$

扩展的最大似然函数利用了样本空间大小的信息，理论上可以减小不确定度。

$\vec{x} = (x_1, x_2, \dots, x_n)$ ,  $X_i \sim N(\mu, \sigma^2)$ , 则均值  $\sim N(\mu, \sigma^2/n)$ 。

## 均值与未知参数无关的情形

此时：

$$\ln L(\nu, \theta) = c + \sum_{i=1}^n \ln \nu f(x_i, \theta)$$

形式与样本空间的大小固定时相同，但是这样的形式有利于处理 $f(x)$ 已知有不同成分叠加的情形。

例如， $f(x) = \frac{\mu_s}{\mu_s + \mu_b} f_s(x) + \frac{\mu_b}{\mu_s + \mu_b} f_b(x)$ ，使用扩展的极大似然估计的结果具有更加直观的“事例数”的估计的意义。

对于非物理结果，（本底信号较大时信号的估计值是负数），也应该按照统计涨落来正常记录。

## 分区间数据的极大似然估计

在数据分区间的情形下，可以对分区间数据进行拟合。根据假设，得到预期的区间内的事例数：

$\vec{\nu} = (\nu_1, \dots, \nu_n)$ ，似然函数为多项分布，对数似然函数的结果为：

$$\ln L = \sum_{i=1}^n n_i \nu_i(\theta)$$

相比于点估计(unbinned ML)，误差稍大，但是影响不大。

## ML拟合优度检验

对于拟合优度检验，一般可以直接利用ML方法对 $\ln L$ 进行模拟，p值的确定就是 $P(\ln L \leq \ln L_{actual})$ 。

## 不等精度似然结果的合并

假设对 $\mu$ 进行独立测量 $n$ 次，测量结果为 $x_i \pm \sigma_i$ ，且 $x_i \sim N(\mu, \sigma_i^2)$ 。使用最大似然法对 $\mu$ 进行估计，得到：

$$\ln L(x_i, \mu) = c - \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_i^2}$$

$$\frac{\partial \ln L}{\partial \mu} = 0 \Rightarrow \sum_{i=1}^n \frac{x_i - \hat{\mu}}{\sigma_i^2} = 0$$

$$\hat{\mu} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, w_i = \frac{1}{\sigma_i^2}$$

由各个测量量相互独立，可以方便的计算出

$$\sigma_{\hat{\mu}} = \frac{1}{\sum_{i=1}^n w_i}$$

对所有测量结果的合并，不确定度一定比任何一次测量的结果更低。

对于权重因子也可以按照这种方式进行定义，本质是完全相同的。

## 不等精度观测结果的合并：另一种权重

对某一固定量  $\mu$  作  $n$  次不等精度测量，测量值为  $x_1, \dots, x_n$ ，对应的标准差分别为  $\sigma_1, \dots, \sigma_n$ 。每个测量值  $x_i \sim N(\mu, \sigma_i^2)$ ，且各测量值相互独立，方差已知。则该样本的似然函数为

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma_i}\right)^2\right]$$

取对数并解似然方程

$$\frac{\partial \ln L(\mu)}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma_i} = 0 \rightarrow \hat{\mu} = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} = \sum_{i=1}^n \omega_i x_i$$

权重因子

$$\omega_i = \frac{1/\sigma_i^2}{\sum_{i=1}^n 1/\sigma_i^2}$$

$$\widehat{\sigma^2_{\hat{\mu}}} = \left( -1 / \frac{\partial^2 \ln L}{\partial \mu^2} \right) \bigg|_{\mu=\hat{\mu}} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}$$

$$\sum_{i=1}^n \omega_i = 1$$

## chap 07 最小二乘法

### 最小二乘法与最大似然法之间的关系

对于  $n$  个相互独立的高斯变量  $Y_i$ ，对数似然函数可以直接等价于：

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \lambda(x_i; \theta))^2}{\sigma_i^2} \text{取最小值。}$$

$\lambda_i$  为待确定的真值， $\lambda_i = \lambda(x_i; \theta)$ ， $x_i$  认为是精确测量的结果。

对于并不相互独立的高斯变量，但是可以用高维高斯密度函数分布描述的结果，可以写为：

$$\chi^2 = \sum_{i,j=1}^n (y_i - \lambda(x_i; \theta)) V_{ij}^{-1} (y_j - \lambda(x_j; \theta)) \text{取最小值。}$$

对于不是高斯分布的情形（一般情形下偏离的不会太多），也可以使用  $\chi^2$  最小值来确定参数的取值。

### 线性情形下的最小二乘估计

对于  $\lambda$  对各个参数  $\theta_j$ ， $j = 1, \dots, m$  线性的情形：

$$\lambda_i = \sum_{j=1}^m a_j(x_i) \theta_j, a_j(x) \text{ 是任意线性独立函数。}$$

此时得到的估计量是有效估计量。（无偏且方差最小）

设  $\mathbf{A} : A_{ij} = a_j(x_i)$ ，则  $\vec{\lambda} = \mathbf{A} \vec{\theta}$ 。

$$\begin{aligned} \chi^2 &= (\vec{y} - \mathbf{A} \vec{\theta})^T \mathbf{V}^{-1} (\vec{y} - \mathbf{A} \vec{\theta}) \\ \frac{\partial \chi^2}{\partial \vec{\theta}} &= -\mathbf{A}^T \mathbf{V}^{-1} (\vec{y} - \mathbf{A} \vec{\theta}) - (\vec{y} - \mathbf{A} \vec{\theta})^T \mathbf{V}^{-1} \mathbf{A} = 0 \\ \Rightarrow \hat{\vec{\theta}} &= (\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{V}^{-1} \vec{y} =: \mathbf{B} \vec{y} \end{aligned}$$

计算参数估计的协方差矩阵：

$$\mathbf{U} = \mathbf{BVB}^T = (\mathbf{AV}^{-1}\mathbf{A}^T)^{-1} = \left(\frac{1}{2}\nabla\nabla\chi^2\right)^{-1}$$

换言之，

$$\chi^2(\vec{\theta} \pm \sigma_{\hat{\theta}}) = \chi_{min}^2 + 1$$

在高斯分布的形式下，由于 $\chi^2 = -2\ln L + c$ ，可以对应于最大似然法的图解法条件。

这一范围确定下来的 $\vec{\theta}$ 区间称为置信区间。

## 多项式的最小二乘拟合

作为线性估计量的一个实例，考虑多项式的假设为：

$$\lambda(x_i; \vec{\theta}) = \sum_{i=1}^m \theta_i x_i^{i-1}$$

当多项式自由变量 $m + 1 = n$ 时， $\chi^2 = 0$ 恒成立，估计无意义。（过拟合）

非线性最小二乘法拟合估计：一般情形 $\lambda(x; \vec{\theta})$ 下只能进行数值求解。

有约束情形下的最小二乘拟合

为了简单起见，仅考虑无参数的最小二乘拟合。使用Lagrange乘子法计算

$$\chi^2(\vec{x}, \vec{\alpha}) = (\vec{x}' - \vec{x})^T \mathbf{V}^{-1} (\vec{x}' - \vec{x}) + 2\vec{\psi}^T \alpha$$

的最小值即可。

## 检验最小二乘法的拟合优度

$\chi_{min}^2$ 的大小决定了数据与假设之间的符合程度，进而可以直接用来检验拟合优度。

如果 $n$ 个变量相互独立的高斯分布变量，由定义 $\chi^2$ 显然满足自由度为 $n - m$ 的卡方分布。

如果 $\chi_{min}^2$ 较小，只能说明实验数据与当前假设相对符合。

**“较小”的标准： $E[\chi^2(n_d)] = n_d$** ，认为是可以接受的结果。

拟合优度与统计不确定度之间的关系：统计不确定度对应于 $\chi^2$ 在最小值附近的变化快慢，拟合优度对应 $\chi_{min}^2$ 的大小。换言之，拟合优度与测量的不确定度直接并**没有直接的关系**。

例如，如果对某次测量的结果进行上下“平移”的同时保持误差不变，对于参数的估计值的误差保持不变，但是 $\chi_{min}^2$ 的值可以变大或变小。

## 最小二乘法处理分区数据

对直方图进行拟合。假设 $X$ 的概率密度分布函数 $f(x; \theta)$ ， $Y_i$ 为第 $i$ 个区间的频数，则

$\lambda_i = n \int_{x_{min}}^{x_{max}} f(x; \theta) dx = np_i(\vec{\theta})$ 。假设区间的宽度足够小，则 $Y_i$ 可以看作是泊松变量。

$$\chi^2(\theta) = \sum_{i=1}^n \frac{(y_i - \lambda_i)^2}{\sigma_i^2}$$

换言之，假定 $y_i$ 服从均值为 $\mu_i(\theta)$ ，方差为 $\sigma_i^2(\theta)$ 的卡方分布，这种方法称为最小二乘估计(LS)。在实际操作中，常常使用 $\hat{\sigma}_i^2 = y_i$ 代替 $\sigma_i^2$ ，这种方法称为修正最小二乘估计(MLS)。MLS估计方法的优点是，计算比较方便，但缺点在于，如果区间的数据量太小， $\sigma_i^2$ 的估计会相对不准。一般情形下， $y_i > 5$ 就可以认为是相对不错的估计量。



## 归一化问题

对比扩展的最大似然法（事例总数作为泊松变量进行估计），在最小二乘估计中使用类似的方案，得到：

$$\begin{aligned}\lambda_i(\theta) &= \nu p_i(\theta) \\ \chi^2(\theta) &= \sum_{i=1}^n \frac{(y_i - \lambda_i)^2}{\sigma_i^2} \\ \frac{\partial \chi_{MLS}^2}{\partial \nu} = 0 &\Rightarrow \sum_{i=1}^n \frac{(y_i - \lambda_i)(-p_i)}{y_i} = 0, \hat{\nu}_{MLS} = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n p_i^2 / y_i} = \frac{1}{\sum_{i=1}^n p_i^2 / y_i} \\ \frac{\partial \chi_{LS}^2}{\partial \nu} = 0 &\Rightarrow \sum_{i=1}^n \frac{y_i^2}{\nu p_i} + p_i = 0, \hat{\nu}_{LS}^2 = \sum_{i=1}^n \frac{y_i^2}{p_i}\end{aligned}$$

代入实际的估计量的值的结果，得到：

$$\begin{aligned}\hat{\nu}_{MLS} &= n - \chi^2 \\ \hat{\nu}_{LS} &= n + \chi^2 / 2\end{aligned}$$

不是好的估计量，有偏倚。

## 测量结果的合并

对于 $n$ 次测量，测量结果为 $\vec{y} = (y_1, \dots, y_n)$ ，协方差矩阵为 $\mathbf{V}^{n \times n}$ 。如果各个测量结果相互独立，用最小二乘法估计真值：

$$\chi^2 = \frac{(y_i - \lambda)^2}{V_{ii}}, \frac{\partial \chi^2}{\partial \lambda} = 0 \Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^n \frac{y_i}{V_{ii}}}{\sum_{i=1}^n \frac{1}{V_{ii}}}$$

这与使用ML方法得到的结果相同。

而对于相关的测量结果：

$$\begin{aligned}\chi^2 &= \vec{y}'^T \mathbf{V}^{-1} \vec{y} = \sum_{i,j=1}^n (y_i - \lambda) V_{ij}^{-1} (y_j - \lambda), \frac{\partial \chi^2}{\partial \lambda} = 0 \\ \Rightarrow \hat{\lambda} &= \frac{\sum_{i=1}^n \omega_i y_i}{\sum_{i=1}^n \omega_i}, \omega_i = \sum_{j=1}^n V_{ij}^{-1}, V[\lambda] = \frac{\sum_{i=1}^n \omega_i V_{ij} \omega_j}{\sum_{i=1}^n \omega_i}\end{aligned}$$

例子：

# 两个相关实验的平均值

假设有两个相关的测量量  $y_1$  和  $y_2$ ，且

$$V = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \rightarrow \begin{aligned} \hat{\lambda} &= w_1 y_1 + (1 - w_1) y_2, \quad w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \\ V[\hat{\lambda}] &= \frac{(1 - \rho^2)\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \equiv \sigma \end{aligned}$$

因第二个测量导致方差倒数的增加为

$$\frac{1}{\sigma^2} - \frac{1}{\sigma_1^2} = \frac{1}{1 - \rho} \left( \frac{\rho}{\sigma_1} - \frac{1}{\sigma_2} \right)^2 > 0 \rightarrow \text{第二个测量结果对平均值总是有帮助。}$$

如果  $\rho > \sigma_1/\sigma_2 \rightarrow w_1 < 0$ ,

加权平均的结果将不在  $y_1$  和  $y_2$  之间!

如果相关性由使用相同数据引起，不可能发生这种情况；  
如果相关性来自共同的随机效应，有可能发生这种情况。  
如果  $\rho, \sigma_1, \sigma_2$  不正确，结果将很不可信，需要自己检查。

考虑  $X_1 = x_1 \pm \sigma_1^{stat} \pm \sigma_1^{syst}$ ,  $X_2 = x_2 \pm \sigma_2^{stat} \pm \sigma_2^{syst}$ ，则协方差矩阵：

$$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

## chap 08 矩方法

### 矩的定义

原点矩（代数矩）

$$\mu'_k = \int f(x) x^k dx$$

中心矩

$$\mu_k = \int f(x) (x - E[X])^k dx$$

中心矩与原点矩之间的关系

$$\mu_k = \sum_{i=0}^k (-1)^{n-i} \mu'_i E[X]^{k-i}$$

一般情形下，可以使用给出的特征函数方便的计算出各阶原点矩。

$$\varphi(t) = E[e^{itX}] = \int f(x) e^{itx} dx$$

$$\frac{d^k \varphi(t)}{dt^k} = i^k \int f(x) x^k e^{itx} dx$$

故

$$\mu'_k = \frac{\varphi^{(k)}(0)}{i^k}$$

## 矩方法的估计量

$$E[a_i(x)] = \int f(x) a_i(x; \vec{\theta}) dx = e_i(\vec{\theta})$$

使用 $a_i(x)$ 的样本平均对 $e_i(\theta)$ 做出估计，用数学的语言表述：

$$\hat{e}_i(\vec{\theta}) = \frac{1}{n} \sum_{j=1}^n a_i(x_j)$$

得到 $m$ 个方程，原则上可以确定 $m$ 个参数的估计值。一般情形下，可以直接 $a_i = x^i$ ，这也是矩方法的来源。

## 协方差的估计

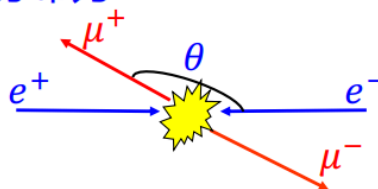
一般情形下，可以使用协方差传递的方法计算出结果。对于矩的估计值（注意平均值确定还是未定），可以直接使用方差来进行估计。

### 例：角分布理论的简单验证

在  $e^+e^- \rightarrow \mu^+\mu^-$  实验中，理论预言角分布为

$$\frac{dn}{d\cos\theta} \propto n(1 + \cos^2\theta)$$

$n =$  事例数



$$E[\cos^2\theta] = \frac{\int_{-1}^{+1} \cos^2\theta \cdot n(1 + \cos^2\theta) d\cos\theta}{\int_{-1}^{+1} n(1 + \cos^2\theta) d\cos\theta} = 0.4$$

二阶原点矩

为了验证理论，计算样本的二阶原点矩：

$$\overline{\cos^2\theta} = \frac{1}{n} \sum_{i=1}^n \cos^2\theta_i$$

假设检验可通过比较二阶原点矩与二阶样本原点矩来完成。

# 角分布理论的简单验证：不确定度估计

不含参数，已知 $\cos^2 \theta$ 的真值：

$$E[\cos^2 \theta] = 0.4$$

$\cos^2 \theta$ 的样本方差：

$$S^2 = \frac{1}{n} \sum_{i=1}^n (\cos^2 \theta_i - E[\cos^2 \theta])^2$$

样本的二阶矩 $\overline{\cos^2 \theta}$ 的方差：

$$V[\overline{\cos^2 \theta}] = \frac{S^2}{n}$$

假设实验观测1000次， $\overline{\cos^2 \theta} = 0.39$ ，并计算得到 $S^2 = 0.15$ 。则实验结果报告为： $0.39 \pm 0.01$

观测值在一个标准误差范围内与理论预期相符。

## 例：含参数的矩方法估计

假设理论预言 $e^+e^- \rightarrow \mu^+\mu^-$ 的角分布含未知参数 $\alpha$ ：

$$\frac{dn}{d \cos \theta} \propto n(1 + \alpha \cos^2 \theta)$$

计算 $\cos \theta$  二阶原点矩的数学期望值

$$E[\cos^2 \theta] = \frac{\int_{-1}^{+1} \cos^2 \theta \cdot n(1 + \alpha \cos^2 \theta) d \cos \theta}{\int_{-1}^{+1} n(1 + \alpha \cos^2 \theta) d \cos \theta} = \frac{5 + 3\alpha}{5(3 + \alpha)}$$

则参数 $\alpha$  与二阶原点矩的关系为

$$\alpha = \frac{5(3E[\cos^2 \theta] - 1)}{3 - 5E[\cos^2 \theta]}$$



$$\hat{\alpha} = \frac{5(3\overline{\cos^2 \theta} - 1)}{3 - 5\overline{\cos^2 \theta}}$$

只要函数是可积的，采用矩方法原则上就可以确定参数。

# 最大似然法、最小二乘法和矩方法

	矩方法	最大似然法	最小二乘法
数据输入	单个事例	单个事例	直方图
多维问题	最容易	归一化较复杂	较难
充分性	会有信息丢失	最具充分性	有时与区间大小有关
相合性	收敛于真值	收敛于真值	收敛于真值
有效性	不是最有效	通常最有效	基本上与似然法一样
无偏性	渐进无偏	渐进无偏	渐进无偏
拟合优度	较难评估	较难评估	很容易

**充分性：**估计量应包含观测值对于未知参数的全部信息；  
**相合性：**样本容量增大时，估计值收敛于真值；  
**有效性：**估计量的分布对其期望值具有最小方差；  
**无偏性：**无论样本容量多大，估计值与真值无系统偏差。

## chap 09 统计误差、置信区间和极限

### 统计不确定度中的标准差问题

例子： $\hat{\theta} = \hat{\theta}(\vec{x})$ 是随机变量，那么

$$\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}$$

意味着 $\theta$ 的估计值是 $\hat{\theta}$ ， $\sigma_{\theta}$ 的估计值是 $\hat{\sigma}_{\theta}$ 。从方差的估计情况可以判断出 $\theta$ 的分布情况。

### 点估计与区间估计

- 点估计：在某一方法下最好的估计值。
- 区间估计：真值落在某一区间的范围大。

### 经典置信区间问题

#### 置信区间

对于给定的小量 $\alpha$ ， $\theta$ 为待定的估计量， $\Theta$ 为所有的估计量取值构成的集合。对于 $\theta \in \Theta$ 满足

$$P(T_1 < \theta < T_2) \geq 1 - \alpha$$

那么随机区间 $[T_1, T_2]$ 称为置信区间。等号成立时称为**同等置信区间**。

$T_1$ (置信下界),  $T_2$ (置信上界)是数据样本的随机变量。

## 置信区间的含义

频率论的观点：对于足够多次实验，每次用相同的方法构造置信区间，真值落在置信区间内的概率为  $1 - \alpha$ 。

贝叶斯的观点：有  $1 - \alpha$  的信心确定真值落在置信区间中。

置信区间的精度与可靠度：

区间长度越小，说明区间的精度越高；置信区间的  $\alpha$  越小，说明置信区间的可靠度越高。

## 区间估计的基本步骤

- 确定枢轴量

枢轴量：仅仅是样本的函数，不包含待估计的参数。同时，枢轴量的分布不能依赖于待估计的参数。

- 构建不等式
- 解出置信区间

## 例子：正态分布均值的区间估计

对于独立同分布的正态分布随机变量  $X \sim N(\mu, 1^2)$ 。样本空间为  $X_i, i = 1, 2, 3, 4, 5$ 。容易注意到：

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

根据给定的  $\alpha$ ，可以给出置信区间为  $[\bar{X} - \delta(\alpha), \bar{X} + \delta(\alpha)]$ 。

另一种方法：对于高斯随机变量

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_i^2}$$

服从自由度为  $n$  的卡方分布。可以验证，采用不同的枢轴量和方法，给出的置信区间是可以不同的。

## 置信区间的构造

- 一般情况下，置信区间的端点取值并不唯一。如果有可能，应当选取最短的置信区间。
- 对于偏态分布的情形，选取最短的置信区间往往是不容易做到的，常常选取两侧对称的概率（均为  $\frac{\alpha}{2}$ ），称为等尾置信区间。

## 对于一般情况的讨论

对于构造的估计量，可以给出估计量随真值的变化函数  $g(\hat{\theta}; \theta)$ 。对于双侧边界给出的  $\alpha, \beta$ ：

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

由两个积分就可以方便的求出上下界  $u_{\alpha}(\theta), v_{\beta}(\theta)$ 。在  $u_{\alpha}(\theta)$  到  $v_{\beta}(\theta)$  之间的区域称为置信带。

对于某次实验的估计量  $\hat{\theta}$ ，利用方程：

$$u_{\alpha}(a) = \hat{\theta}$$
$$v_{\beta}(b) = \hat{\theta}$$

确定置信区间为 $[b, a]$ 。

## 高斯分布估计量的置信区间

如果 $g(\hat{\theta}; \theta)$ 服从高斯分布，则：

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

## 泊松分布估计量的置信区间

## 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

## chap 10 解谱法

## 附录：概率论沉思录

### chap 01 合情推理

1. 强三段论（演绎推理）：A真  $\rightarrow$  B真，A真，则B真  
弱三段论（合情推理）：A真  $\rightarrow$  B真，B真，则A更合情
2. 布尔代数：  
积：AB（A与B同时为真，则为真）  
和：A+B（A或B为真，则为真）  
否： $\overline{A}$ （A为假）

- 写成加法与乘法的形式，为了直接运用加法与乘法的运算律。
- 运算的实例：

- $AA = A, A\overline{A} = False(0), 0 + A = A, \overline{\overline{A}} = A, A + A = A$
- $A + \overline{A} = True(1), True(1)A = A$
- $\overline{A} + A\overline{B} = \overline{AB}$

### 3. 蕴含关系

$A \Rightarrow B$ （A为真，则B为真；B为假，则A为假）

用布尔代数表达： $A = AB$ ，可以验证满足上面的蕴含关系的定义。

$A \Rightarrow B$ 的布尔值与 $\overline{A} + B$ 的布尔值相同。

### 4. 逻辑函数

对 $n$ 个命题，最多有 $2^{n+1}$ 个不同的概率函数。（ $n$ 个命题有 $2^n$ 种结果，函数可能有两种取值）

$2^{n+1}$ 种函数，可以简化为规范析取范式（形如 $A_1 \dots A_n, \overline{A_1} \dots A_n$ 的合取（交/乘积）式的析取（或/和）式）

### 5. 完备集：合取(AND)、析取(OR)与否定(NOT)

实际上这三种逻辑关系中，两者可以表出第三者，如  $A + B = \overline{\overline{A} \overline{B}}$

,  $AB = \overline{\overline{A} + \overline{B}}$ 。进而可以定义与非 ( $\uparrow$ ) 运算:  $A \uparrow B = \overline{A + B}$  (合取式的否定), 或非 ( $\downarrow$ ) 运算:  $A \downarrow B = \overline{A \overline{B}}$  (析取式的否定)

#### 6. 合情条件

- 合情程度用实数表示( $A|B$ )
- 定性地与常识相符
- 一致性推理 (结果与路径无关)

## chap 02 定量规则

### 1. 乘法规则

$$(AB|C) = F[(B|C), (A|BC)]$$

$$(AB|C) = F[(A|C), (B|AC)]$$

进一步确定  $F(x, y)$  的形式:

- 由合情条件:  $F_x \geq 0, F_y \geq 0$
- 由迭代关系:  $F[x, F(y, z)] = F[F(x, y), z]$

记  $u = F(x, y), v = F(y, z)$ , 对上面的表达式分别对  $x, y$  求偏导:

$$F_1(u, z)F_1(x, y) = F_1(x, v)$$

$$F_1(u, z)F_2(x, y) = F_2(x, v)F_1(y, z)$$

$$\text{记 } G(x, y) = \frac{F_2(x, y)}{F_1(x, y)}$$

$$G(x, y) = G(x, v)F_1(y, z) \dots (\#)$$

或

$$G(x, y)G(y, z) = G(x, v)F_2(y, z) \dots (\#\#)$$

对 (#) 式, 对  $z$  偏导:

$$0 = G_2(x, v)F_2(y, z)F_1(y, z) + G(x, v)F_{12}(y, z)$$

对 (##) 式, 对  $y$  求偏导:

$$[G(x, y)G(y, z)]_y = G_2(x, v)F_1(y, z)F_2(y, z) + G(x, v)F_{21}(y, z)$$

连续性给出:

$$F_{12}(y, z) = F_{21}(y, z)$$

代入得到:

$$[G(x, y)G(y, z)]_y = 0$$

进而可以确定函数  $G(x, y)$  的形式:

$$G(x, y) = r \frac{H(x)}{H(y)}$$

将分离变量的形式代入 (#) (##) 式, 略去推导, 积分得到:

$$\phi(v) = c\phi(y)\phi^r(z)$$



其中：

$$\phi(x) = \int \frac{dx}{H(x)}$$

将这一结果代入迭代关系的表达式，化简得到：

$$\phi(F(x, y)) = c \cdot \phi(x)\phi(y)$$

令  $w(x) = c \cdot \phi(x)$ ，得到：

$$w(F(x, y)) = w(x)w(y)$$

回到原来的问题，得到结论：

$$w(AB|C) = w(A|BC)w(B|C)$$

## 2. 加法规则

略去推导，用类似的方式可以推出：

$$P(A|B) + P(\bar{A}|B) = 1$$

## 3. 推理规则与演绎推理、合情推理之间的关系（合情推理的合情程度用概率函数表达）