第7章 参数估计

Chpt. 7 Parameter Estimation

统计推断 (Statistical Inference)

[1] 推断什么？

- 推断样本分布中的未知参数
- 根据问题的需要，可以只推断部分参数，推断形式也可以不同，e.g., 估计物件重量a，a是否超过1kg

[2] 怎么推断？

- 从一定的条件和假定（样本和统计模型）出发，按照一定的方法或规则，得出未知事件（未知参数）的某种结论。

[3] 统计推断的特殊性

- 统计推断的结果不能保证不错
- 对比：“等腰三角形底角相等”，“求一元二次方程的根”

统计推断 (Statistical Inference)

[4] 统计推断的意义

- 虽然不能保证在每一种具体情况下所作的统计推断不错，但使用概率论的方法，可以给出种种有意义的指标，去衡量推断的正确程度。
  - E.g., 估计物体重量a，在所设正态模型下，若用样本平均值$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$去估计a，则我们可以算出$\bar{x}$与实际值a的偏离超过给定误差$c > 0$的机会有多大，即$P(|\bar{x} - a| > c)$，可以作为$\bar{x}$这个推断的正确性的一个合理指标。
- 统计推断具备现代数学的严格性。
  - E.g., 估计物体重量a，测量9次，可以证明，当统计模型为正态时，9次测量的算术平均值是对a最优的估计。

统计推断 (Statistical Inference)

[5] 统计推断的研究内容

- 提出各种具体的统计推断方法；
- 计算有关这些方法的性能的各种数量指标；
- 在一定条件和一定意义下寻找最优的推断方法，或者证明某种统计推断方法是最优的。

数理统计学的基础是概率论，统计推断从结论的表述到优良性的刻画，都离不开该概率论。

7.1 参数的点估计

[1] 参数点估计

现象：

很多随机变量/总体的分布是由几个参数完全决定的。

正态分布$N(\mu, \sigma^2)$

Poisson分布$\pi(\lambda)$

假定分布形式已知，知道了参数就可以确定分布

问题：

设总体X的分布函数的形式已知，但他的一个或多个参数未知

借助总体X的样本来估计总体分布中的未知参数θ问题称为参数的点估计问题。

7.1 参数的点估计

[1]参数点估计

概念：

设总体X的分布函数F(X,θ) 的形式为已知，θ是待估参数。

$X_1, X_2, \cdots, X_n$ 是X的一个样本，$x_1, x_2, \cdots, x_n$ 是相应的一个样本观测值。

估计问题就是构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$，用它的观测值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 作为未知参数θ的近似值。

我们称 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 为<u>θ的估计量</u>，$\hat{\theta}(x_1, x_2, \cdots, x_n)$ 为<u>θ的估计值</u>。

[2] 矩估计

设X为连续随机变量，其概率密度为$f(x;\theta_1,\theta_2,\cdots,\theta_k)$，其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数，$X_1,X_2,\cdots,X_n$ 是来自总体X的样本，假设总体 $\boldsymbol{X}$ 的前 k 阶矩存在：

$$\mu_l = E(X^l) = \int_{-\infty}^{\infty} x^l f(x;\theta_1,\theta_2,\cdots,\theta_k)\,dx$$
$$= \mu_l(\theta_1,\theta_2,\cdots,\theta_k)$$

$$l = 1,2,\cdots,k$$

μ_l 是 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数。由此可以求解得到：

[2] 矩估计

$$\begin{cases} \theta_1 = \theta_1(\mu_1, \cdots, \mu_k) \\ \vdots \\ \theta_k = \theta_k(\mu_1, \cdots, \mu_k) \end{cases}$$

但是k 阶矩不能得到，用样本矩 $A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l$ (l=1,2,…,k),

代替 μ_l，形成对θ_l的估计：

$$\begin{cases} \hat{\theta}_1 = \theta_1(A_1, \cdots, A_k) \\ \vdots \\ \hat{\theta}_k = \theta_k(A_1, \cdots, A_k) \end{cases}$$

这样地估计称为矩估计量，得到的值是矩估计（或者说是用矩作估计）。

[Example 7.1] 设总体X的均值μ方差σ^2都存在但未知，又设 $X_1,\cdots,X_n$ 是来自总体的样本，求μ与σ^2的矩估计。

解：
$$\begin{cases} \mu_1 = E(X) = \mu \\ \mu_2 = E(X^2) = D(X) + \left[E(X)\right]^2 = \sigma^2 + \mu^2 \end{cases}$$

得到
$$\begin{cases} \mu = \mu_1 \\ \sigma^2 = \mu_2 - \mu_1^2 \end{cases}$$

分别以一二阶样本矩代替μ_1，μ_2得到，

$$\begin{cases} \hat{\mu} = A_1 = \overline{X} \\ \hat{\sigma}^2 = A_2 - A_1^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{cases}$$

这说明均值、方差的矩估计表达式不因分布的不同而不同。

7.2 极大似然法

[1] 思想方法

极大似然法的想法是，一随机试验，已知有若干个结果

$$A_1,\ A_2,\ \cdots\ ,A_i,\ \cdots$$

如果在一次试验中A_i发生了，则可认为当时的条件最有利于A_i发生，故应如此选择分布的参数，使发生A_i的概率最大。

7.2 极大似然法

[Example 7.2] 已知甲乙两射手命中靶心的概率分别为p_1=0.8和p_2=0.5，今有一张靶纸上表明10枪6中靶心，又知靶子肯定是甲乙之一射的，问究竟是谁所射的可能性最大？

[解] 设事件A＝{10枪6中靶心}

若是甲所射，则A发生的概率为，

$$P_1(A) = C_{10}^{6}(0.8)^{6}(0.2)^{4} = 0.088$$

若是乙所射，则A发生的概率为，

$$P_2(A) = C_{10}^{6}(0.5)^{6}(0.5)^{4} = 0.21$$

显然，$P_1(A) < P_2(A)$，故可认为乙所射的可能性较大.

7.2 极大似然法

[2] 似然函数

含参数θ的总体X的样本 $X_1,\cdots,X_n$，设 $x_1,\cdots,x_n$ 为样本的观测值.

- 当X是离散型时, 设其概率分布为 $P\{X=x\}=p(x,\theta)$, 令

$$L(\theta)=\prod_{i=1}^{n}p(x_i,\theta)$$

$L(\theta)$ 称为似然函数，给定了样本观测值，找出什么样的θ使得样本数显的概率最大

- 当X是连续型时，设其概率密度为$f(x,\theta)$，类似得到似然函数

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

[3] 参数的估计

参数的取值应使所抽到的样本以最大的概率出现. 换言之, 应使似然函数$L(\theta)$达到最大值。

极大似然估计 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 与样本观测值直接相关，它是使得似然函数 $L(\theta)$ 达到最大值的估计值：

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta)$$

相应的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 称为极大似然估计量。

[3] 参数的估计

Remark 1：在很多情况下 $p(x,\theta)$、$f(x,\theta)$关于θ 可微，这时最大似然估计值可在方程 $\dfrac{d}{d\theta}L(x_1,x_2,\cdots,x_n;\theta)=0$ 处求得。

Remark 2：为考虑方便，常用似然函数的对数--**对数似然函数**来代替它。因为lnL是L的单增函数，有相同的最大值

$$\frac{d}{d\theta}\ln L(x_1,x_2,\cdots,x_n;\theta)=0$$

[4] 例题

[Example 7.3] 泊松分布参数的估计。设总体X服从泊松分布

$P\{X=x\}=\dfrac{\lambda^x}{x!}e^{-\lambda}$ 其中参数 λ 未知，求它的极大似然估计值.

解　设 $x_1,\cdots,x_n$ 为其样本观察值，则似然函数为

$$L=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}e^{-n\lambda}$$

$$\ln L=\left(\sum_{i=1}^{n}x_i\right)\ln\lambda-\sum_{i=1}^{n}\ln(x_i!)-n\lambda$$

解方程：
$$\frac{d}{d\lambda}\ln L=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0$$

得到极大似然估计值 $\hat{\lambda}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$

Nankai University

[4] 部分例子

[Example 7.4] 捕鱼问题模拟试验

为了估计湖中鱼的数量 n，先从湖中捕捞出 r 条鱼，并做了记号后放回湖中.

经过适当时间后，可认为有记号的鱼的分布基本均匀.

然后再在湖中捕捞出 s 条鱼，结果发现有 x 条鱼是有记号的.

如何用极大似然估计法从已知数 r, s, x估计出未知数 n?

[Example 7.4] 捕鱼问题模拟试验

[解]：第二次捕捞出有记号的鱼数 Y 服从超几何分布

$$P\{Y=x\}=C_r^x C_{n-r}^{s-x} \Big/ C_n^s$$

未知数 n 应使得 $P\{Y=x\}=C_r^x C_{n-r}^{s-x}/C_n^s$ 最大

未知数 n 应该为 $\hat{n}=\dfrac{rs}{x}$

由于 n 为整数，故此取 $\hat{n}=\left[\dfrac{rs}{x}\right]$。

其实质就是，按第二次捕捞到的有记号的鱼占所捕捞到的全部鱼的比例来估算 n 的值，即认为 $\dfrac{r}{n}=\dfrac{x}{s}$.

[Example 7.5] 均匀分布的参数的极大似然估计

设总体X服从$[a,b]$上的均匀分布， $x_1,\cdots,x_n$ 为样本观测值，求a,b的极大似然估计.

解 记 $\underline{x}=\min(x_1,\cdots,x_n),\ \overline{x}=\max(x_1,\cdots,x_n),$

总体X的密度函数为

$$f(x,a,b)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & other\end{cases}$$

似然函数为

$$L(x,a,b)=\prod_{i=1}^{n}f(x_i,a,b)=\frac{1}{(b-a)^n}$$

[Example 7.5] 均匀分布的参数的极大似然估计

而我们知道

$$a = \underline{x} = \min(x_1, \cdots, x_n)$$

$$b = \overline{x} = \max(x_1, \cdots, x_n)$$

时似然函数达到最大.

故的a,b最大似然估计值为

$$\hat{a} = \min(x_1, \cdots, x_n)$$

$$\hat{b} = \max(x_1, \cdots, x_n)$$

a, b的最大似然估计量

$$\hat{a} = \min(X_1, \cdots, X_n) \qquad \hat{b} = \max(X_1, \cdots, X_n)$$

[Example 7.6] 正态分布的参数的极大似然估计

设总体X服从正态分布 $N(\mu,\sigma^2)$，其中参数 μ,σ^2 未知

求它们的极大似然估计值.

解 设为 $x_1,\cdots,x_n$ 其样本观察值，则似然函数为

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

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

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

[Example 7.6] 正态分布的参数的极大似然估

令

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

$$\frac{\partial}{\partial \sigma}\ln L = \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2-\frac{n}{\sigma}=0$$

解得

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

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

矩估计和极大似然估计对比

	矩估计 Moment Estimation	极大似然估计 Maximum Likelihood Estimation
必要条件	总体的数值特征，这个数值特征通常是包含待估参数的	总体的分布函数
优点	简单易行	利用了分布的信息，估计的质量较好
缺点	未充分利用分布的信息，一般估计量质量较差	计算量较大

例题

例1 设总体X的概率密度为

$$f(x)=\begin{cases}(\alpha+1)x^{\alpha}, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$$

其中$\alpha>-1$为未知参数。求α的矩估计。

解： 先求总体的期望

$$E(X)=\int_0^1 x\cdot(\alpha+1)x^{\alpha}\,\mathrm{d}\,x$$

$$=(\alpha+1)\int_0^1 x^{\alpha+1}\,\mathrm{d}\,x$$

$$=\frac{\alpha+1}{\alpha+2}.$$

上面是数学推导，后面带入样本数据

由矩估计法，令

$$\overline{X} = \frac{\alpha + 1}{\alpha + 2}$$

样本矩

总体矩

解得

$$\hat{\alpha} = \frac{2\overline{X} - 1}{1 - \overline{X}}$$

为α 的矩估计。

注意：要在参数上边加上“^”，表示参数的估计。它是统计量。

例2 设$X_1,X_2,\cdots X_n$是取自总体X的简单随机样本, X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta}, & x\ge\mu,\\ 0, & \text{其他}.\end{cases}$$

2个参数，需计算二阶矩

其中 θ,μ 为未知参数，$\theta>0$。求 θ,μ 的矩估计。

解： 先求总体的均值和2阶原点矩。

$$E(X)=\int_{\mu}^{\infty}x\cdot\frac{1}{\theta}e^{-(x-\mu)/\theta}\,\mathrm{d}\,x$$

令$y=(x-\mu)/\theta$

$$=\int_{0}^{\infty}(\theta\, y+\mu)e^{-y}\,\mathrm{d}\,y$$

$$=\theta+\mu.$$

Nankai University 南開大學

$$E(X^2)=\int_{\mu}^{\infty}x^2\cdot\frac{1}{\theta}e^{-(x-\mu)/\theta}\,\mathrm{d}\,x$$

令$y=(x-\mu)/\theta$

$$=\int_0^{\infty}(\theta\, y+\mu)^2 e^{-y}\,\mathrm{d}\,y$$

$$=\int_0^{\infty}(\theta^2\, y^2+2\theta\mu\, y+\mu^2)e^{-y}\,\mathrm{d}\,y$$

$$=\cdots$$

$$=2\theta^2+2\theta\mu+\mu^2$$

$$=\theta^2+(\theta+\mu)^2,$$

令 $\begin{cases} \mu+\theta=\overline{X}, \\ \theta^2+(\mu+\theta)^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2. \end{cases}$

用样本矩
估计总体矩

得

$$\begin{cases} \hat{\theta}=\sqrt{\dfrac{1}{n}\left[\sum\limits_{i=1}^{n}X_i^2-n\overline{X}^2\right]}=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}, \\ \hat{\mu}=\overline{X}-\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}. \end{cases}$$

$\hat{\mu}$, $\hat{\theta}$ 为参数 μ,θ 的矩估计。

例3 设$X_1, X_2, \cdots, X_n$是取自总体 $X\sim B(1,p)$ 的一个样本，求参数 p 的极大似然估计。

解： X的概率分布律为：

$$P\{X=x\}=p^x(1-p)^{1-x}, x=0,1$$

似然函数为

$$L(p)=\prod_{i=1}^{n} f(x_i,p)$$

$$=\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i},$$

对数似然函数为：

$$\ln L(p)=\sum_{i=1}^{n}x_i\ln(p)+(n-\sum_{i=1}^{n}x_i)\ln(1-p),$$

对p求导，并令其等于零，得

$$\frac{d\ln L(p)}{dp}=\frac{1}{p}\sum_{i=1}^{n}x_i-\frac{1}{1-p}(n-\underbrace{\sum_{i=1}^{n}x_i}_{n\bar{x}})=0.$$

上式等价于

$$\frac{\bar{x}}{p}=\frac{1-\bar{x}}{1-p}.$$

解上述方程，得 $p=\bar{x}$. 得 $\hat{p}=\bar{X}$ 为 p 的极大似然估计。

例4 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\theta, & 0<x<1\\ 1-\theta, & 1\le x<2\\ 0, & 其他\end{cases}$$

其中 0< $\theta<1$为未知参数，$(X_1, \cdots, X_n)$是取自总体 X 的样本，试求θ的矩估计量与最大似然估计量

解 $EX=\int_0^1 x\theta dx+\int_1^2 x(1-\theta)dx=\frac{3}{2}-\theta,\ \theta=\frac{3}{2}-EX$

$$\hat{\theta}_{ME}=\frac{3}{2}-\bar{X}$$

记 N 为样本中小于 1 的个数，似然函数为

$$L(\theta)=\theta^N(1-\theta)^{n-N}$$

$$LnL=Nln\theta+(n-N)\ln(1-\theta)$$

似然方程为

$$\frac{dLn\,L}{d\theta}=\frac{N}{\theta}-\frac{n-N}{1-\theta}=0$$

解得

$$\hat{\theta}_{MLE}=\frac{N}{n}$$

例5 设总体X的概率分布为

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中$0<\theta<\frac{1}{2}$是未知参数，利用总体X的如下样本值3,1,3,0,3,1,2,3

求θ的矩估计值和最大似然估计值。

分析：参数的最大似然估计值，就是对给定的观测值$(x_1,\cdots,x_n)$，选取$\hat{\theta}$，使得似然函数$L(\theta)$达到最大，本题中总体X是离散型随机变量，根据其分布确定似然函数$L(\theta)$是求解的关键

解 (1)EX=0× θ^2+1 ×2θ(1-θ)+2× θ^2+3× $(1-2\theta)$ = 3-4θ

$$\theta=\frac{1}{4}(3\text{-EX}),\ \hat{\theta}_{\mathrm{ME}}=\frac{1}{4}(3-\overline{\mathrm{X}})$$

故θ的矩阵估计值为

$$\hat{\theta}_{\mathrm{ME}}=\frac{1}{4}(3-\bar{\mathrm{x}})=\frac{1}{4}(3-2)=\frac{1}{4}$$

(2)对给定的样本值(3,1,3,0,3,1,2,3)，似然函数为

$$L(\theta)=\prod_{i=1}^{8}P\{X_i=x_i\}=P\{X=0\}\cdot\{P(X=1)\}^2\cdot\ P\{X=2\}\cdot\{P(X=3)\}^4$$

$$=\theta^2\cdot[2\theta(1-\theta)]^2\cdot\theta^2\cdot(1-2\theta)^4=4\theta^6(1-\theta)^2(1-2\theta)^4$$

$$\mathrm{Ln}\ L=\ln 4+6\ln\theta+2\ln(1-\theta)+4\ln(1-2\theta)$$

$$\frac{d\ln L}{d\theta}=\frac{6}{\theta}-\frac{2}{1-\theta}-\frac{8}{1-2\theta}=0$$

即有$12\theta^2-14\theta+3=0$

得$\hat{\theta}_{\mathrm{MLE}}=\frac{7-\sqrt{13}}{12}$ ($\hat{\theta}=\frac{7+\sqrt{13}}{12}$ 舍去)

例6 设总体X~U[0,θ]，其中$\theta > 0$为未知参数，$(X_1, \ldots, X_n)$是取自总体X的样本，求θ的矩估计量与最大似然估计量.

解 (1)E(X)=$\frac{\theta}{2}$，θ=2E(X)，$\hat{\theta}_{\mathrm{ME}}=2\overline{\mathrm{X}}$;

（2）X~$f(x;\theta)=\begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{其他} \end{cases}$

似然函数为$L(\theta) = \frac{1}{\theta^n} \quad (0 < x_i < \theta)$

对数似然函数为$\ln L(\theta) = -n ln\, \theta \quad (0 < x_i < \theta)$

似然方程为$\frac{dln\, L}{d\theta} = -\frac{n}{\theta} < 0$

$\boxed{\hat{\theta}_{\mathrm{MLE}}=\max_{1\le i\le n} X_i = X_{(n)}}$

例7 设总体$X\sim U(a,b)$，$(X_1,\ldots,X_n)$是取自总体X的样本，求a,b的矩估计量

例8 设总体X服从如下分布（柯西分布），$(X_1, \ldots, X_n)$是取自总体X的样本，求参数θ的矩估计和极大似然估计

$$f(x;\theta) = \frac{1}{\pi[(1+(x-\theta)^2]}, \qquad x \in (-\infty, +\infty)$$

7.3 衡量估计值优劣的标准

应该存在不同的估计量和估计值

比如 σ^2 的估量。

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$S_*^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

不同的估计量，哪个好，哪个差，这是估计量的评选问题。需要有一些标准

衡量估计值优劣的标准

[1] 无偏性

$X_1, X_2, \cdots, X_n$ 是总体X的样本

$\theta \in \Theta$ 是包含在总体X中的未知参数

若估计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在且等于未知参数 θ，即

$$E(\hat{\theta}) = \theta$$

则称 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 为 θ 的无偏估计量.

此时，用 $\hat{\theta}(x_1, \cdots, x_n)$ 代替 θ 不含系统误差.

样本均值是总体均值的无偏估计；样本方差是总体方差的无偏估计.

设总体X的数学期望为 $E(X)=\mu$，方差为 $D(X)=\sigma^2$

$$E(\overline{X})=\mu, D(\overline{X})=\frac{1}{n}\sigma^2$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2-2X_i\overline{X}+\overline{X}^2\right)$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}X_i^2-2\overline{X}\frac{1}{n-1}\sum_{i=1}^{n}X_i+\frac{n}{n-1}\overline{X}^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}X_i^2-2\frac{n}{n-1}\overline{X}^2+\frac{n}{n-1}\overline{X}^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}X_i^2-\frac{n}{n-1}\overline{X}^2$$

样本均值是总体均值的无偏估计；样本方差是总体方差的无偏估计.

$$E(S^2)=\frac{1}{n-1}\sum_{i=1}^{n}E(X_i^2)-\frac{n}{n-1}E(\overline{X}^2)$$

$$=\frac{n}{n-1}\left[D(X_i)+(EX_i)^2\right]-\frac{n}{n-1}\left[D(\overline{X})+(E(\overline{X}))^2\right]$$

$$=\frac{n}{n-1}(\sigma^2+\mu^2)-\frac{n}{n-1}\left(\frac{1}{n}\sigma^2+\mu^2\right)$$

$$=\sigma^2$$

这也是为什么用 $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 而不用 $Y_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ 的进一步原因。

所以，S^2是σ^2的无偏估计值.

*S*是不是*σ*的无偏估计?

例6（续） 设总体X~U[0,θ]，其中$\theta > 0$为未知参数，$(X_1, \dots, X_n)$是取自总体X的样本，求θ的矩估计量与极大似然估计量.

解 （1） $\hat{\theta}_{\mathrm{ME}}=2\overline{\mathrm{X}}$ （2） $\hat{\theta}_{\mathrm{MLE}}=\max\limits_{1\le i\le n} X_i = X_{(n)}$

1. 验证上述两个估计量是否具有无偏性

$$(1)\ E(\hat{\theta}_{\mathrm{ME}}) = 2E(\bar{X}) = 2E(X) = 2\cdot\frac{\theta}{2} = \theta$$

$$(2)\ F_{\hat{\theta}_{\mathrm{ME}}}(x) = F_{X_{(n)}}(x) = P(X_{(n)} \le x)$$

$$= P(X_1 \le x, \cdots, X_n \le x)$$

$$= P(X_1 \le x)\cdots P(X_n \le x) = F^n(x) \qquad (0 \le x \le \theta)$$

(2) $F_{X_{(n)}}(x) = F^n(x)$

$$f_{X_{(n)}}(x) = nF^{n-1}(x)f(x) = n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta} \qquad (0 \le x \le \theta)$$

$$E\left(X_{(n)}\right) = \int_{-\infty}^{\infty} x f_{X_{(n)}}(x)dx = \int_0^{\theta} x \cdot n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}dx$$

$$= \frac{n}{n+1}\theta$$ 系统偏小

修正 $\hat{\theta}_L = \frac{n+1}{n}\hat{\theta}_{\mathrm{MLE}} = \frac{n+1}{n}X_{(n)}$

[2] 有效性

无偏估计值未必是唯一的，应该选择与 θ 的平均偏差较小者为好，即估计值应有尽量小的方差，这就引出了有效性标准

设 $\hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 与 $\hat{\theta}_2(X_1, X_2, \cdots, X_n)$ 都是 θ 的无偏估计

如果在样本容量相同的情况下，$\hat{\theta}_1$ 较 $\hat{\theta}_2$ 更密集在真值θ附近，即

$$D(\hat{\theta}_1) \le D(\hat{\theta}_2)$$

则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ **有效**；如果对给定的 n，$D(\hat{\theta})$ 的值达到最小，则称 $D(\hat{\theta})$ 为 θ 的**有效估计值**，也称为**最小方差无偏估计（MVUE）**

[2] 有效性

例如 $\overline{X}$ 与 X_i 均为总体均值 μ 的无偏估计，

但是 $$D(\overline{X})=\frac{\sigma^2}{n}<\sigma^2=D(X_i)$$

因此，$\overline{X}$ 较 X_i 有效.

一般时候，虽然都是样本均值，但是随着样本个数的增加，估计的方差会减小，即

$$D(\overline{X}_n)=\frac{1}{n}\sigma^2>\frac{1}{n+k}\sigma^2=D(\overline{X}_{n+k})$$

例6（续） 设总体X~U[0,θ]，其中$\theta > 0$为未知参数，$(X_1, \ldots, X_n)$是取自总体X的样本，求θ的矩估计量与最大似然估计量.

解 （1） $\hat{\theta}_{\mathrm{ME}}=2\overline{X}$ （2） $\hat{\theta}_{\mathrm{MLE}}=\max\limits_{1\le i\le n} X_i = X_{(n)}$

1. 两个无偏估计

（1） $\hat{\theta}_{\mathrm{ME}}=2\overline{X}$ （2） $\hat{\theta}_L = \frac{n+1}{n}\hat{\theta}_{\mathrm{MLE}}= \frac{n+1}{n}X_{(n)}$

2. 比较$\hat{\theta}_{\mathrm{ME}}$和$\hat{\theta}_{\mathrm{L}}$的有效性

$$D(\hat{\theta}_{\mathrm{ME}}) = 4D(\bar{X}) = 4\cdot\frac{D(X)}{n} = 4\cdot\frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$

$$D(\hat{\theta}_L) = D\left(\frac{n+1}{n}X_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 D(X_{(n)})$$

$$\Leftarrow D\left(X_{(n)}\right) = E\left({X_{(n)}}^2\right) - [E(X_{(n)})]^2 \qquad = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2 \theta^2$$

$X_{(n)}$概率密度函数

$$f_{X_{(n)}}(x) = nF^{n-1}(x)f(x) = n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}$$

$$E\left({X_{(n)}}^2\right) = \int_0^{\theta} x^2 \cdot \frac{nx^{n-1}}{\theta^n}dx = \frac{n}{n+2}\theta^2$$

$$D(\hat{\theta}_L) = \left(\frac{n+1}{n}\right)^2 D\left(X_{(n)}\right) = \frac{1}{n(n+2)}\theta^2$$

$$D(\hat{\theta}_{\mathrm{ME}}) = \frac{\theta^2}{3n}$$

$$D(\hat{\theta}_L) = \frac{1}{n(n+2)}\theta^2$$

因为 $\frac{1}{3n} \geq \frac{1}{n(n+2)}, n \geq 1$

称$\hat{\theta}_L$相比$\hat{\theta}_{\mathrm{ME}}$具有更好的有效性

Example 7.7 设$(X_1,\cdots,X_n)$是取自总体X的样本，记$\mu=E(X)$，$\sigma^2=D(X)$，试证对任意常数a_i，$\sum_{i=1}^{n} a_i=1, i=1,\cdots,n, \sum_{i=1}^{n} a_i X_i$均是$\mu$的无偏估计，其中$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$最为有效。

证明:

$$E\left(\sum_{i=1}^{n} a_i X_i\right)=\sum_{i=1}^{n} a_i E(X_i)=\sum_{i=1}^{n} a_i E(X)=\mu\sum_{i=1}^{n} a_i=\mu$$

故$\sum_{i=1}^{n} a_i X_i$均是μ的无偏估计。

$$D\left(\sum_{i=1}^{n} a_i X_i\right)=\sum_{i=1}^{n} a_i^2 D(X_i)=\sum_{i=1}^{n} a_i^2 D(X)=\sigma^2\sum_{i=1}^{n} a_i^2$$

欲证$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$最为有效，只需证明当且仅当$a_i=\frac{1}{n},i=1,2,\cdots,n$时，$\sum_{i=1}^{n}a_i^2$达到最小。

已知$\sqrt{\frac{a_1^2+\cdots+a_n^2}{n}}\geq\frac{a_1+\cdots+a_n}{n}$,当且仅当$a_1=\cdots=a_n$等号成立

$$\sum_{i=1}^{n}a_i^2\geq\frac{(\sum_{i=1}^{n}a_i)^2}{n}=\frac{1}{n}\left(\sum_{i=1}^{n}a_i\right)=\frac{1}{\mathrm{n}}$$

且等号成立当且仅当$a_i=\frac{1}{n},i=1,2,\cdots,n$

[3] 相合性（一致性）

由于统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 与n有关，不妨记为 $\hat{\theta}_n$，我们自然希望 n 越大时，对 θ 的估计越精确，于是有相合性（一致性）标准。

$\hat{\theta}(X_1, X_2, \cdots, X_n)$ 是参数 θ 的估计量，如对任意的 $\theta \in \Theta$ ，
当 $n \to \infty$ 时，$\hat{\theta}_n$ 按概率收敛于θ，即对任意$\varepsilon>0$，$\lim\limits_{n\to\infty} P\{|\hat{\theta}_n - \theta| < \varepsilon\} = 1$
则称 $\hat{\theta}_n$ 是θ 的相合估计值（一致估计值）。

样本均值 $\overline{X}$ 是总体均值 μ 的一致估计值

这是因为由大数定理可知 $\lim\limits_{n\to\infty} P\left\{ |\overline{X} - \mu| < \varepsilon \right\} = 1$ 从而，样本均值 $\overline{X}$ 是总体均值 μ 的一致估计值.

方差存在且有限时才成立

样本方差S^2是总体方差σ ^{2}的一致估计值

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}^2$$

$$= \frac{n}{n-1} \bullet \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}^2$$

$$= \frac{n}{n-1}\left(\overline{X^2} - \overline{X}^2\right)$$

当 $n \to \infty$ 时，$\overline{X^2}$ 与 $\overline{X}^2$ 分别按概率收敛于$E(X^2)$与$(EX)^2$

从而S^2按概率收敛于$E(X^2)-(EX)^2 = DX = \sigma^2$

所以，S^2是σ^2的一致估计值.

[4] 均方误差

$$E[(\hat{\theta}-\theta)^2] \qquad = E[((\hat{\theta}-E(\hat{\theta}))-(\theta-E(\hat{\theta})))^2]$$

$$= E\left[\left(\hat{\theta}-E(\hat{\theta})\right)^2\right] - 2E\left[\left(\hat{\theta}-E(\hat{\theta})\right)\left(\theta-E(\hat{\theta})\right)\right] + E[\left(\theta-E(\hat{\theta})\right)^2]$$

$$= D(\hat{\theta}) \ + E[\left(\theta-E(\hat{\theta})\right)^2]$$

有效性　　　　无偏性

7.4 区间估计

[1] 为何要引进参数的区间估计

- 参数点估计方法不能回答估计值的可靠度与精度问题

  假设我们知道 $\hat{\theta}$ 是 θ 的一个无偏估计，即 $E\left(\hat{\theta}\right)=\theta$

  但是“估计值 $\hat{\theta}$ 落在区间 $[\theta-\varepsilon，\theta+\varepsilon]$ 的概率有多大？”

- 许多应用场合不需要对参数给出一个精确估计，而只要大致范围

  例如，要估计一批电子产品的平均寿命，往往不需要一个很精确的数，而只需给出一个不大的范围即可，如8000~9000小时。当然，还要求对这个估计有一定的“可信程度”，比如95%。

引 例

设某厂生产的灯泡使用寿命X~ $N(\mu, 100^2)$，现随机抽取5只，测量其寿命如下：

1455，1502，1370，1610，1430，

则该厂灯泡的平均使用寿命的点估计值为

$$\bar{x} = \frac{1}{5}(1455 + 1502 + 1370 + 1610 + 1430) = 1473.4$$

可以认为该种灯泡的使用寿命在1473. 4个单位时间左右，但范围有多大呢？又有多大的可能性在这“左右”呢？

引 例

如果要求有95%的把握判断 μ 在1473. 4左右，相当于要求对应统计量 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ 以95%概率落在0周围。

$$P\left\{\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le z_{0.025}\right\} = 0.95$$

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

[2] 区间估计的概念

设 $\hat{\theta}_1(X_1,\cdots,X_n)$ 及 $\hat{\theta}_2(X_1,\cdots,X_n)$ 是由样本确定的两个统计量

$$\hat{\theta}_1 < \hat{\theta}_2$$

如果对于给定的α（0<α<1）, 有

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1-\alpha$$

则随机区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 称做参数θ的置信水平为1-α的**置信区间**；其中 $\hat{\theta}_1$ 叫做**置信下限**， $\hat{\theta}_2$ 叫做**置信上限**。

$1-\alpha$ 称为**置信水平**或者**置信度**。

注意：$\hat{\theta}_1 < \hat{\theta}_2$ 的含义是概率意义下成立

置信区间的意义：

从总体X中抽取容量为 n 的样本，共进行N次随机抽样,每次得到的样本值记为 $(x_{1k}, x_{2k}, \cdots, x_{nk})$, $(k=1,2,\ldots, N)$;

由 $\hat{\theta}_1(X_1,\cdots,X_n)$ 及 $\hat{\theta}_2(X_1,\cdots,X_n)$ 得到N个区间

$$(\hat{\theta}_{1k}, \hat{\theta}_{2k}), (k=1,2,\cdots,N)$$

这N 个区间中,有的包含参数θ 的真值，有的不包含。

包含参数θ 的真值的区间大约占100(1-α)%, 不包含参数θ 的真值的区间约占100α%.

区间估计是根据样本算得的，该区间中不一定总是包含目标参数

[2] 区间估计的概念

给定置信水平，根据样本观测值来确定未知参数 θ 的置信区间, 称为参数θ 的区间估计。

满足置信水平1－α的 θ 的置信区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 有无穷多个。

置信区间越小，估计越精确，但置信水平会降低；相反，置信水平越大，估计越可靠，但精确度会降低，置信区间会较长。

对于固定的样本容量，不能同时做到精确度高（置信区间小）可靠程度也高（1- α大）。

如果不降低可靠性，而要缩小估计范围，则必须增大样本容量，增加抽样成本。

在保证可靠性的情况下尽量提升精度。

[Exmaple 7.8] 设总体 $X \sim N(\mu, \sigma^2), (X_1, X_2, \cdots, X_n)$ 是该总体的一个样本，σ^2已知，μ未知，求μ的置信水平为 1 －α的置信区间。

[解]：我们知道 $\bar{X}$ 是μ的无偏估计，且 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，按照标准正态分布$\alpha$分位点的定义

$$P\left\{\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le z_{\alpha/2}\right\} = 1-\alpha$$

$\bar{X}$是随机变量

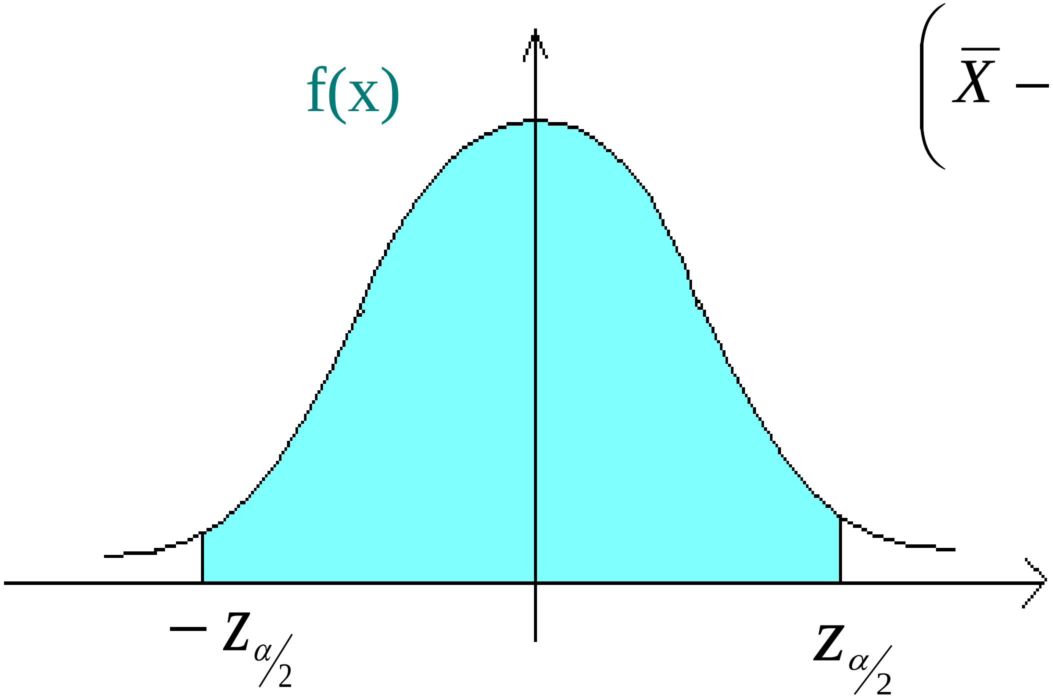
$$P\left\{\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2} \le \mu \le \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right\} = 1-\alpha$$

这样我们就得到μ的置信水平为 1 －α 的置信区间

$$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$$

7.4 区间估计

我们所得到μ的置信水平为 1 －α的置信区间



$$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$$

7.4 区间估计

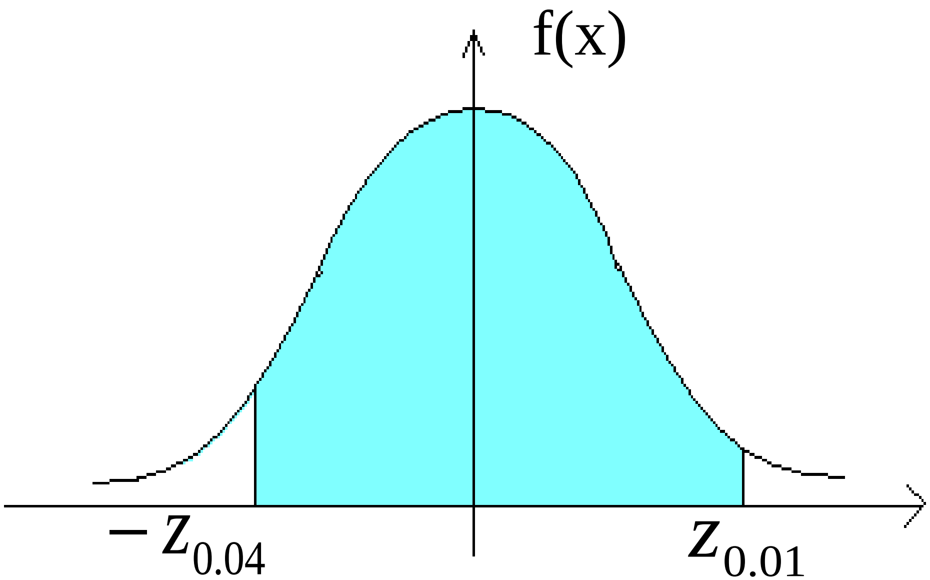
如果 $\alpha=0.05, \sigma=1, n=16$，查表得 $z_{\alpha/2}=z_{0.05}=1.96$，于是得到一个置信水平为0.95的置信区间 $\left(\bar{X}-0.49, \bar{X}+0.49\right)$

如果我们得到样本的一组观测值，计算得到 $\bar{x}=5.20$ ，则得到更为具体的置信区间(4.71, 5.69)。

7.4 区间估计

注意到 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，那么概率为1-α 的区间有无穷多个。

比如α=0.05时，必有 $P\left\{-z_{0.04} \le \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le z_{0.01}\right\} = 0.95$，那么以下区间 $\left(\overline{X}-\dfrac{\sigma}{\sqrt{n}}z_{0.01}, \overline{X}+\dfrac{\sigma}{\sqrt{n}}z_{0.04}\right)$ 是 μ 的置信水平为0.95的置信区间



两个置信水平相等的区间，显然区间长度较短的估计精度更高

如果把估计区间说成(-∞,∞)，那等于什么也没说

上面两个置信水平都为0.95的置信区间的长度分别为

$$2\times\frac{\sigma}{\sqrt{n}}z_{0.025}=3.92\times\frac{\sigma}{\sqrt{n}}$$

$$\frac{\sigma}{\sqrt{n}}(z_{0.04}+z_{0.01})=4.08\times\frac{\sigma}{\sqrt{n}}$$

事实上由于 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，其概率密度是对称的单峰函数，可以断定对称置信区间 $\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$ 长度最短。

[Example 7.9] 某地一旅游者的消费额 X 服从正态分布 $N(\mu,\sigma^2)$，且标准差σ=12元，今要对该地旅游者的平均消费额 $E(X)$ 加以估计，为了能以95%的置信度相信这种估计误差小于2元，问至少要调查多少人？

解 由题意知：消费额 $X \sim N(\mu，12^2)$，设要调查 n 人，使得。

$$P\left\{\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|<1.96\right\}=0.95$$

$$P\left\{\left|\bar{X}-\mu\right|<1.96\times\frac{\sigma}{\sqrt{n}}\right\}=0.95$$

而 $\left|\bar{X}-\mu\right|<2 \Longrightarrow 1.96\times\frac{\sigma}{\sqrt{n}}<2$

解得 $n>\left(\frac{1.96\times 12}{2}\right)^2=138.29$

至少要调查139人

[3] 单侧置信区间

前面讨论的估计量的置信区间都是双侧的，在有些实际问题中，我们只需要考虑单侧区间，例如

某元件的使用寿命，平均寿命没有上限的限制问题，太短就不行，在这种情况下，可将置信上限取为+∞，而只考虑置信下限。

在相反的情况下，则只考虑置信上限。这两种估计方法称为**单侧置信区间**的估计法。

[3] 单侧置信区间

对于给定的(0<α<1),根据样本确定的统计量$\underline{\theta}(X_1,\cdots,X_n)$，对于任意的$\theta\in\Theta$

有$P\{\underline{\theta}<\theta\}=1-\alpha$，则随机区间$(\underline{\theta},+\infty)$ 称做参数θ的置信水平为1-α的单侧置信区间；其中$\underline{\theta}$ 叫做置信水平为1-α的单侧置信下限。

又若，统计量$\overline{\theta}(X_1,\cdots,X_n)$，对于任意的$\theta\in\Theta$ 有 $P\{\theta<\overline{\theta}\}=1-\alpha$,

则随机区间$(-\infty,\overline{\theta})$ 称做参数θ的置信水平为1-α的单侧置信区间；其中$\overline{\theta}$ 叫做置信水平为1-α的单侧置信上限.

Review

总体：对某一数量（或几个）指标进行随机实验、观察，将试验的全部可能的观察值称为总体。

个体：每个可能的观察值称为个体。

抽样：对总体进行一次观察并记录其结果，称为一次抽样；对X独立进行n次观察，并将结果按顺序记为 $X_1,\cdots,X_n$。

样本：随机抽取部分个体，以用于推断总体的特性。

样本与总体是同分布的，样本之间是独立的。

统计量：样本的函数，除了样本、样本的参数外，不含有其他未知量

抽样分布：统计量的分布称为抽样分布。

几类抽样分布（对X~N(μ，σ²)而言的）

样本均值分布 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

T分布 $t=\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

χ^2分布 $\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

F分布 $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1,n_2-1)$

两个总体的样本分布 $N(\mu_1,\sigma_1^2)$ $N(\mu_2,\sigma_2^2)$

[1] $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

[2] $$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} \sim F(n_1-1, n_2-1)$$

[3] 当$\sigma_1=\sigma_2=\sigma$时 $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

其中 $$S_w = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

7.5 单个正态总体均值与方差的区间估计

7.6 两个正态总体均值差与方差比的区间估计

7.7 非正态总体参数的区间估计

7.5 正态总体均值与方差的区间估计

正态总体均值 μ 的区间估计

(1) 正态总体 $X \sim N(\mu, \sigma^2)$ ，已知 $\sigma = \sigma_0$，求μ的区间估计

(2) 正态总体 $X \sim N(\mu, \sigma^2)$，未知 σ，求μ的区间估计

正态总体方差 σ^2 的区间估计

(1) 正态总体 $X \sim N(\mu, \sigma^2)$，已知$\mu = \mu_0$ ，求σ^2 的区间估计

(2) 正态总体 $X \sim N(\mu, \sigma^2)$，未知μ ，求σ^2 的区间估计

正态总体均值μ的区间估计

(1) 设正态总体 $X \sim N(\mu,\sigma^2)$，已知 $\sigma=\sigma_0$ ，求 μ 区间估计

因为 $\overline{X} \sim N(\mu,\dfrac{\sigma^2}{n})$ ，所以 $\overline{Y}=\dfrac{\overline{X}-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$，对于给定的置信水平1−α，

正态分布的密度函数在区间 $\left(-z_{\alpha/2}, z_{\alpha/2}\right)$ 上的概率可以达到要求，即

$$P\left\{\frac{\left|\overline{X}-\mu\right|}{\sigma_0/\sqrt{n}} \le z_{\alpha/2}\right\}=1-\alpha$$

$z_{\alpha/2}$ 是标准正态分布的上 $\alpha/2$ 分位点，即

$$P\left\{X \ge z_{\alpha/2}\right\}=\frac{\alpha}{2}$$

正态总体均值μ的区间估计

（1）设正态总体 $X\sim N(\mu,\sigma^2)$，已知 $\sigma=\sigma_0$ ，求 μ 区间估计

把上述关于事件概率的描述转化为关于均值 μ 的概率描述：

$$P\left\{\overline{X}-\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2}\le\mu\le\overline{X}+\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2}\right\}=1-\alpha$$

由此，求得关于 μ 的置信水平为1−α 的 μ 的置信区间(用观测值)：

$$\left(\bar{x}-\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2},\quad \bar{x}+\frac{\sigma_0}{\sqrt{n}}z_{\alpha/2}\right)$$

正态总体均值μ的区间估计

[Example 7.10]

从一批零件中, 抽取9个零件, 测得其直径(毫米)为

19.7　20.1　19.8　19.9　20.2　20.0　19.9　20.2　20.3

设零件直径服从 $X \sim N(\mu,\sigma^2)$ ，且已知 σ=0.21(毫米), 求这批零件直径的均值 μ 对应于置信水平0.95的置信区间.

解: n=9, 直接计算得 $\bar{x}=20.01$(毫米). 若置信水平1−α , 则 α =0.05,

查表得 $z_{\alpha/2}=z_{0.025}=1.96$

由此得 $\dfrac{\sigma_0}{\sqrt{n}}z_{\alpha/2}=\dfrac{0.21}{\sqrt{9}}\times 1.96=0.14$

得 μ 的置信区间为**(19.87，20.15)**

（2） 正态总体 $X \sim N(\mu,\sigma^2)$，未知σ，求 μ 的区间估计

因σ 未知，在上面的估计中无法使用 σ^2，我们用 S^2 代替 σ^2，得随机变量

$$Y=\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\sim t(n-1)$$

对于给定的置信水平1－α，我们取区间 $\left(-t_{\alpha/2},t_{\alpha/2}\right)$，构造概率事件

$$\left\{\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right|\le t_{\alpha/2}\right\}$$

其满足：

$$P\left\{\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right|\le t_{\alpha/2}\right\}=1-\alpha$$

（2） 正态总体 $X \sim N(\mu,\sigma^2)$，未知σ，求 μ 的区间估计

由此转化为关于 μ 的关系式

$$P\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}\le\mu\le\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}\right)=1-\alpha$$

用观测值带入，求得置信水平为1－α 的 μ 的置信区间(用观测值)：

$$\left(\bar{x}-\frac{s}{\sqrt{n}}t_{\alpha/2},\ \ \bar{x}+\frac{s}{\sqrt{n}}t_{\alpha/2}\right)$$

(2) 正态总体 $X \sim N(\mu,\sigma^2)$，未知σ，求 μ 的区间估计

[Example 7.10续]

未知 σ, 求这批零件直径的均值 μ 对应于置信水平0. 95的置信区间.

解：直接计算得 $\bar{x} = 20.01$(毫米). $s = 2.03$(毫米).

若置信水平1−α =0. 95时， α =0. 05,

选自由度为n−1＝8，查t分布表得 $t_{\alpha/2} = t_{0.025} = 2.31$，由此得，

$$\frac{s}{\sqrt{n}} t_{\alpha/2} = \frac{0.203}{\sqrt{9}} \times 2.316 = 0.16$$

得 μ 的置信区间为(19. 85，20. 17).

可以看出与上述置信区间与已知 σ时得到的估计相差不多。

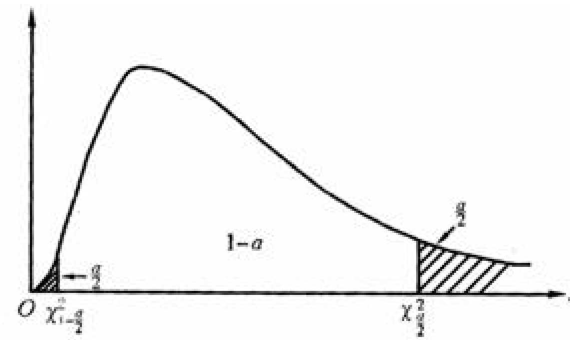
正态总体方差 σ^2的区间估计

(1) 正态总体 $X \sim N(\mu,\sigma^2)$，已知 $\mu=\mu_0$，求 σ^2的区间估计

利用随机变量 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$ 进行估计

由于此分布曲线不对称，故对于给定的置信水平1−α，很难找到最短的置信区间. 通常模仿前面的做法，取区间 $\left(\chi^2_{1-\alpha/2}, \ \chi^2_{\alpha/2}\right)$ 使得：

$$P\left\{\chi^2_{1-\alpha/2} \le \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \le \chi^2_{\alpha/2}\right\}=1-\alpha$$



正态总体方差 σ^2的区间估计

(1) 设正态总体 $X \sim N(\mu,\sigma^2)$，已知 $\mu=\mu_0$，求 σ^2的区间估计

转化为关于 σ^2的概率描述，

$$P\left\{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2}}\right\}=1-\alpha$$

得置信水平为1−α的σ^2的置信区间：

$$\left(\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{\alpha/2}}, \frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\chi^2_{1-\alpha/2}}\right)$$

（2）正态总体 $X \sim N(\mu,\sigma^2)$，未知 μ，求 σ^2 的区间估计

由于 $\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 中 μ 未知，用 $\overline{X}$ 代替，得到

$$\chi^2 = \frac{n-1}{\sigma^2}\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2 \xrightarrow{\mu\Rightarrow\overline{X}} \frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$$

此只是与上面的差一个自由度，对给定的置信水平1-α，我们取区间 $\left(\chi^2_{1-\alpha/2},\ \chi^2_{\alpha/2}\right)$，使

$$P\left\{\chi^2_{1-\alpha/2} \le \frac{n-1}{\sigma^2}S^2 \le \chi^2_{\alpha/2}\right\} = 1-\alpha$$

得置信水平为1-α 的 σ^2 的置信区间：

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right)$$

(2) 正态总体 $X\sim N(\mu,\sigma^2)$，未知 μ，求 σ^2 的区间估计

[Example 7.11]

从一批零件中, 抽取9个零件, 测得其直径(毫米)为

19.7, 20.1, 19.8, 19.9, 20.2, 20.0, 19.9, 20.2, 20.3

设零件直径 $X\sim N(\mu,\sigma^2)$， 且未知 μ

求这批零件直径的方差 σ^2对应于置信水平**0.95**的置信区间.

解: 已知**α= 0.05, n = 9, s^2 = 0.411**, 按自由度 **k = 8** 查表得,

$$\chi^2_{0.975}=2.18,\quad \chi^2_{0.025}=17.5$$

所求置信区间为: $\left(\dfrac{8\times 0.411}{17.5},\dfrac{8\times 0.411}{2.18}\right)$

即（**0.188**，**1.508**）.

Summary on Estimation for Normal Parameters

STEP 1：确定一个合适的样本统计量

[1] 其分布是已知的；

[2] 统计量中含有待估计的参数

STEP2：对给定的置信水平1-α构造满足其的一个随机事件

（一般用区间表示）

STEP3：把关于事件的概率描述转化为关于参数的概率描述

STEP4：用满足置信水平1-α的参数区间作为置信区间

（用样本观测值）

Summary on Estimation for Normal Parameters

均值	方差已知	$\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
	方差未知	$\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$
方差	均值已知	$\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$
	均值未知	$\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$

7.6 两个正态总体均值差与方差比的区间估计

不同工艺生产的两批同类产品，可以认为是来自两个相互独立的不同总体. 有时我们要对其某个质量指标作比较，分析它们是否有显著的差异. 这时可观察 $\mu_1-\mu_2$ 和 σ_1^2/σ_2^2 的置信区间.

以下讨论两个正态总体均值差与方差比的区间估计问题.

设总体 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$ 且 $(x_1,x_2,\cdots,x_{n_1})$ 和 $(y_1,y_2,\cdots,y_{n_2})$ 分别是总体 X 与 Y 的样本观察值.

两个正态总体均值差的区间估计

(1) 设两个正态总体 $X \sim N(\mu_1,\sigma_1^2)$，$Y \sim N(\mu_2,\sigma_2^2)$ 已知 σ_1,σ_2，求 $\mu_1-\mu_2$ 的区间估计

选择包含 $\mu_1-\mu_2$ 随机变量 $V=\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$

对于给定的置信水平1-α，取区间 $(-z_{\alpha/2},z_{\alpha/2})$，使

$$P\left\{-z_{\alpha/2} \le V \le z_{\alpha/2}\right\}=1-\alpha$$

$$P\left\{-z_{\alpha/2} \le \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \le z_{\alpha/2}\right\}=1-\alpha$$

南開大學 Nankai University

两个正态总体均值差的区间估计

(1) 设两个正态总体 $X \sim N(\mu_1,\sigma_1^2)$，$Y \sim N(\mu_2,\sigma_2^2)$ 已知 σ_1,σ_2，求 $\mu_1-\mu_2$ 的区间估计

把关于随机事件的概率描述转化为关于 $\mu_1-\mu_2$ 的概率描述

$$P\left\{\overline{X}-\overline{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1-\mu_2 \le \overline{X}-\overline{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right\}=1-\alpha$$

得置信水平为1−α 的 $\mu_1-\mu_2$ 的置信区间：

$$\left(\bar{x}-\bar{y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \ \bar{x}-\bar{y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

注：若 σ_1,σ_2 未知且 n_1 与 n_2 很大时，可用 s_1^2,s_2^2 分别代替 σ_1^2,σ_2^2，仍使用上式作 $\mu_1-\mu_2$ 的区间估计.

两个正态总体均值差的区间估计

(2) 设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$ 未知 σ_1, σ_2 ，但是 $\sigma_1 = \sigma_2$ ，求 $\mu_1 - \mu_2$ 的区间估计。

使用随机变量 $$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

其中 $$S_w = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

对于给定的置信水平1- α，我们取区间 $(-t_{\alpha/2},\ t_{\alpha/2})$，使

$$P\left\{ \frac{\left|(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)\right|}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \le t_{\alpha/2} \right\} = 1 - \alpha$$

两个正态总体均值差的区间估计

(2) 设两个正态总体 $X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$ 未知 σ_1,σ_2 ，但是 $\sigma_1=\sigma_2$ ，求 $\mu_1-\mu_2$ 的区间估计。

把关于随机事件的概率描述转化为关于 $\mu_1-\mu_2$

$$P\left\{\bar{X}-\bar{Y}-t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le \mu_1-\mu_2\le \bar{X}-\bar{Y}+t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right\}=1-\alpha$$

得置信水平为1-α 的 $\mu_1-\mu_2$ 的置信区间（用样本观测值）

$$\left(\bar{x}-\bar{y}-t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\bar{x}-\bar{y}+t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

两个正态总体均值差的区间估计

[Example 7.12] 两台机床生产同一个型号的滚珠，从甲机床生产的滚珠中取8个，从乙机床生产的滚珠中取9个，测得这些滚珠的直径(毫米)如下：

甲机床： **15.0, 14.8, 15.2, 15.4, 14.9, 15.1, 15.2, 14.8**；

乙机床： **15.2, 15.0, 14.8, 15.1, 15.0, 14.6, 14.8, 15.1, 14.5**.

设两台机床生产的滚珠直径服从正态分布

求： 这两台机床生产的滚珠直径均值差 $\mu_1-\mu_2$ 的对应于置信水平**0.90**的置信区间， 如果：

(1) 已知两台机床生产的滚珠直径的标准差分别是

σ_1=0. 18(毫米)及 σ_2=0. 24(毫米)；

(2) 未知 σ_1和 σ_2，但假设 $\sigma_1=\sigma_2$.

解（1） σ_1及σ_2已知，估计 $\mu_1-\mu_2$，采用统计量

$$V=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

置信水平为1−α 的 $\mu_1-\mu_2$ 的置信区间

$$\left(\bar{x}-\bar{y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \ \bar{x}-\bar{y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

查正态分布表得 $z_{0.05}$=1.645，代入上式得所求的置信区间

（**-0.018，0.318**）

解（2）σ_1，σ_2未知，但$\sigma_1=\sigma_2$, 估计 $\mu_1-\mu_2$，采用统计量

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

置信水平为1−α 的 $\mu_1-\mu_2$ 的置信区间

$$\left(\overline{x}-\overline{y}-t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\ \overline{x}-\overline{y}+t_{\alpha/2}S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

取自由度$k=8+9-2=15$, 查t 分布表得$t_{0.05}=1.753$，再计算$S_w=0.228$，

代入上式得所求的置信区间为（**-0.044**，**0.344**）.

两个正态总体方差比的区间估计

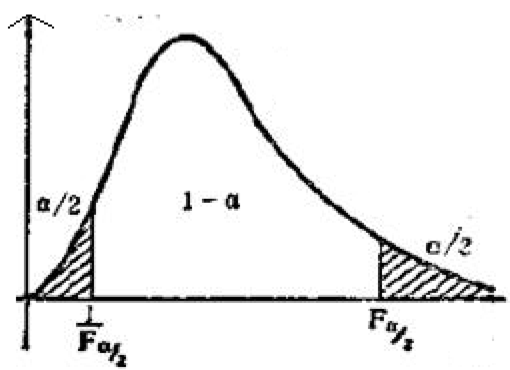
(1) 设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ 已知 μ_1, μ_2 ， 求 σ_1^2 / σ_2^2 的区间估计

利用随机变量

$$F = \frac{\dfrac{\sum_{i=1}^{n_1}(X_i - \mu_1)^2}{n_1\sigma_1^2}}{\dfrac{\sum_{j=1}^{n_2}(Y_j - \mu_2)^2}{n_2\sigma_2^2}} \sim F(n_1, n_2)$$

对于给定的置信水平1−α， 构造置信区间 $(F_{1-\alpha/2}, F_{\alpha/2})$

$$P\left\{ F_{1-\alpha/2} \le \frac{\sum_{i=1}^{n_1}(X_i - \mu_1)^2 \Big/ n_1\sigma_1^2}{\sum_{j=1}^{n_2}(Y_j - \mu_2)^2 \Big/ n_2\sigma_2^2} \le F_{\alpha/2} \right\} = 1-\alpha$$



两个正态总体方差比的区间估计

(1) 设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ 已知 μ_1, μ_2 ， 求 σ_1^2 / σ_2^2 的区间估计

把上述对于事件的描述转化为关于对 方差比 的描述

$$P\left(\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2 \Big/ n_1}{F_{\alpha/2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2 \Big/ n_2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2 \Big/ n_1}{F_{1-\alpha/2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2 \Big/ n_2} \right) = 1-\alpha$$

得置信水平为1-α 的 σ_1^2 / σ_2^2 的置信区间：

$$\left(\frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2 \Big/ n_1}{F_{\alpha/2}\sum_{j=1}^{n_2}(y_j-\mu_2)^2 \Big/ n_2}, \ \frac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2 \Big/ n_1}{F_{1-\alpha/2}\sum_{j=1}^{n_2}(y_j-\mu_2)^2 \Big/ n_2} \right)$$

两个正态总体方差比的区间估计

(2) 设两个正态总体 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$ 未知 μ_1, μ_2 ，求 σ_1^2 / σ_2^2 的区间估计

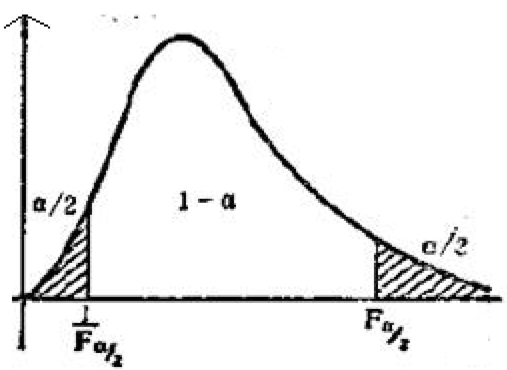
由于二者的总体均值未知，替代为样本均值，采用随机变量：

$$F = \frac{S_1^2}{\sigma_1^2} \Big/ \frac{S_2^2}{\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

对于给定的置信水平1-α，构造置信区间 $(F_{1-\alpha/2}, F_{\alpha/2})$

得置信水平为1-α 的 σ_1^2 / σ_2^2 的置信区间

$$\left(\frac{s_1^2}{F_{\alpha/2} s_2^2}, \frac{s_1^2}{F_{1-\alpha/2} s_2^2} \right)$$



两个正态总体方差比的区间估计

[Example 7.13] 两台机床生产同一个型号的滚珠，从甲机床生产的滚珠中取8个，从乙机床生产的滚珠中取9个，测得这些滚珠的直径（毫米）如下：

甲机床：**15.0, 14.8, 15.2, 15.4, 14.9, 15.1, 15.2, 14.8**；

乙机床：**15.2, 15.0, 14.8, 15.1, 15.0, 14.6, 14.8, 15.1, 14.5**.

设两台机床生产的滚珠直径服从正态分布

求：这两台机床生产的滚珠直径方差比 σ_1^2/σ_2^2 的对应于置信水平 **1-α= 0.90**的置信区间，如果：

（1）已知两台机床生产的滚珠直径的均值分别是 μ_1=15.0（毫米）及 μ_2=14.9（毫米）；

（2）未知 μ_1及 μ_2.

解 已知 $n_1=8, n_2=9, \alpha=0.10$,

（1）取自由度 $n_1=8, n_2=9$，查F分布表得 $F_{\alpha/2}=F_{0.05}(8,9)=3.23$

利用F分布的性质计算 $F_{1-\alpha/2}=F_{0.95}(8,9)=\dfrac{1}{F_{0.05}(9,8)}=\dfrac{1}{3.39}=0.295$

再计算

$$\sum_{i=1}^{8}(x_i-\mu_1)^2=0.34,\ \sum_{j=1}^{9}(y_j-\mu_2)^2=0.46$$

代入求得置信区间（**0.257，2.819**）.

解 已知 $n_1=8, n_2=9$, $\alpha=0.10$,

（2）取自由度 $n_1=8-1, n_2=9-1$，查 F 分布表得 $F_{\alpha/2}=F_{0.05}(7,8)=3.50$

利用 F 分布的性质计算

$$F_{1-\alpha/2}=F_{0.95}(7,8)=\frac{1}{F_{0.05}(8,7)}=\frac{1}{3.73}=0.268$$

再计算

$$s_1^2=0.0457,\quad s_2^2=0.0575$$

代入求得置信区间（**0.227**，**2.966**）.

均值 $\mu_1-\mu_2$	方差已知	$V=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$
	方差未知	$T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$
方差 σ_1^2/σ_2^2	均值已知	$\dfrac{\dfrac{\sum_{i=1}^{n_1}(x_i-\mu_1)^2}{n_1\sigma_1^2}}{\dfrac{\sum_{j=1}^{n_2}(y_j-\mu_2)^2}{n_2\sigma_2^2}}\sim F(n_1,n_2)$
	均值未知	$\dfrac{S_1^2}{\sigma_1^2}/\dfrac{S_2^2}{\sigma_2^2}\sim F(n_1-1,n_2-1)$

正态分布的单侧区间估计

在有些实际问题中，我们只需要考虑单侧区间，例如

- 某元件的使用寿命，平均寿命没有上限的限制问题，但太短不行，在这种情况下，可将置信上限取为+∞，而只考虑置信下限。

$$P\left(\theta \geq \underline{\theta}(X_1, \cdots, X_n)\right) = 1 - \alpha$$

  称$\underline{\theta}$是θ的一个置信水平为$1-\alpha$的置信下界。

- 考虑空气污染问题时，通常更关心单位空间污染物含量的上限，污染物越少越好，这时将置信下限取为−∞，只考虑置信上限。

$$P\left(\theta \leq \overline{\theta}(X_1, \cdots, X_n)\right) = 1 - \alpha$$

  称$\overline{\theta}$是θ的一个置信水平为$1-\alpha$的置信上界。

正态分布的单侧区间估计

与双侧置信区间的定义相比较，可以看出置信上、下界是一种特殊的置信区间，其一端为$+\infty$或$-\infty$。因此，前面用于求区间估计的方法可以直接应用于此。

例如，求$N(\mu,\sigma^2)$中μ的置信水平为$1-\alpha$的置信下界，假设σ已知

对$\overline{X}$进行标准化$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

根据上分位点的定义 $$P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\leq z_\alpha\right)=1-\alpha$$

$$P\left(\mu\geq\overline{X}-\sigma/\sqrt{n}\cdot z_\alpha\right)=1-\alpha$$

注意：这是置信下界

枢轴变量法

区间估计的一般方法——枢轴变量法

（1）找一个与要估计的参数 θ 有关的统计量 T，一般是其一个良好的点估计；

（2）设法找出T和θ的某一函数$S(T,\theta)$, 其分布F要与θ无关，S称为**枢轴变量**；

（3）对任何常数$a<b$, 不等式$a \le S(T,\theta) \le b$要能改写为等价的形式$A \le \theta \le B$, A、B只与T，a，b有关，而与θ无关；

（4）取分布F的上$\frac{\alpha}{2}$分位点$x_{\frac{\alpha}{2}}$和上$1-\frac{\alpha}{2}$分位点$x_{1-\frac{\alpha}{2}}$，则有$F(x_{\frac{\alpha}{2}}) - F\left(x_{1-\frac{\alpha}{2}}\right) = 1-\alpha$, 因此$P\left(x_{1-\frac{\alpha}{2}} \le S(T,\theta) \le x_{\frac{\alpha}{2}}\right) = 1-\alpha$;

根据第(3)条，不等式 $x_{1-\frac{\alpha}{2}} \le S(T,\theta) \le x_{\frac{\alpha}{2}}$可解出$A \le \theta \le B$，因此$[A,B]$就是$\theta$的一个置信水平为$1-\alpha$的区间估计。

正态总体区间估计总结

（1）**χ^2分布**：i.i.d. $X_1,\cdots,X_n \sim N(0,1)$, $Y=\sum_{i=1}^{n} X_i{}^2$, 则$Y\sim\chi^2(n)$

（2）**t分布**：X,Y独立，$X\sim N(0,1)$，$Y\sim\chi^2(n)$, 则$\frac{X}{\sqrt{Y/n}}\sim t(n)$

（3）**F分布**: X,Y独立，$X\sim\chi^2(n_1)$, $Y\sim\chi^2(n_2)$, 则$\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$

表1. 一个正态总体

均值	方差已知	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
	方差未知	$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$
方差	均值已知	$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$
	均值未知	$\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$

表2. 两个正态总体

均值差	方差已知	$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)$
	方差未知	$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$
方差比	均值已知	$\frac{[\sum_{i=1}^{n_1}(x_i-\mu_1)^2]/n_1\sigma_1^2}{[\sum_{j=1}^{n_2}(y_j-\mu_2)^2]/n_2\sigma_2^2} \sim F(n_1,n_2)$
	均值未知	$\frac{S_1^2}{\sigma_1^2}\Big/\frac{S_2^2}{\sigma_2^2} \sim F(n_1-1,n_2-1)$

7.7 非正态总体参数的区间估计——大样本法

若总体不服从正态分布时，一般是很难确定总体中的未知参数的；

但当样本容量 n 很大时，中心极限定理告诉我们 $\dfrac{\overline{X}-\mu(\theta)}{\sigma(\theta)/\sqrt{n}}$ 近似

服从标准正态分布 **N(0,1).** 可利用此作出近似的区间估计.

设总体X服从某一分布，其概率函数或密度中含有未知参数θ，则总

体均值与方差都依赖于参数θ. 接下来，我们考虑样本 $X_1, X_2, \cdots, X_n$，它们

相互独立且与总体同分布，而且

$$E(X_i)=\mu(\theta),\quad D(X_i)=\sigma^2(\theta),\quad (i=1,2,\cdots,n)$$

7.7 非正态总体参数的区间估计——大样本法

当样本容量**n**充分大(≥50)时，由中心极限定理知，$\dfrac{\overline{X}-\mu(\theta)}{\sigma(\theta)/\sqrt{n}}$ 近似服从标准正态分布**N(0,1)**.

因此，对给定的置信水平1−α， 有

$$P\left\{\frac{\left|\overline{X}-\mu(\theta)\right|}{\sigma(\theta)/\sqrt{n}} \le z_{\alpha/2}\right\} \approx 1-\alpha$$

若能从不等式 $\dfrac{\left|\overline{x}-\mu(\theta)\right|}{\sigma(\theta)/\sqrt{n}} \le z_{\alpha/2}$ 解出参数$\boldsymbol{\theta}$，把关于随机事件的概率描述转换为关于参数$\boldsymbol{\theta}$ 的描述，则得参数 θ 的近似置信水平为1−α 的置信区间。

服从“0-1”分布的总体参数p的区间估计

设总体X服从“ 0−1 ”分布：x=0 或者 x=1，其中参数 p 未知. 则 $E(X)=p$

$D(X)=p(1-p)$ 对给定的置信水平1−α，得

$$P\left\{\frac{\left|\overline{X}-p\right|}{\sqrt{p(1-p)}\Big/\sqrt{n}}\le z_{\alpha/2}\right\}\approx 1-\alpha$$

把不等式 $\dfrac{\left|\overline{x}-p\right|}{\sqrt{p(1-p)}\Big/\sqrt{n}}\le z_{\alpha/2}$ 两边平方并整理得

$$P\left(n(\overline{x}-p)^2\le p(1-p)z_{\alpha/2}^2\right)\approx 1-\alpha$$

再化作关于p的二次不等式

$$(n+z_{\alpha/2}^2)p^2-(2n\overline{x}+z_{\alpha/2}^2)p+n\overline{x}^2\le 0$$

服从“0-1”分布的总体参数p的区间估计

令： $a = n + z_{\alpha/2}^2$， $b = -(2n\overline{x} + z_{\alpha/2}^2)$， $c = n\overline{x}^2$

因为各x_i取值0或1，故 $0 \le \overline{x} \le 1$ ，从而判别式

$$b^2 - 4ac = 4n\overline{x}(1-\overline{x})z_{\alpha/2}^2 + z_{\alpha/2}^4 > 0$$

得：

$$\hat{p}_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \qquad \hat{p}_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

即 $\hat{p}_1, \hat{p}_2$ 是两个实根. 则参数 p 的近似置信区间为 $(\hat{p}_1, \hat{p}_2)$.

注：如果对区间估计的精度要求不高，当n比较大时，可以用

$$\left(\overline{x} - z_{\frac{\alpha}{2}}\sqrt{\overline{x}(1-\overline{x})/n}, \overline{x} + z_{\alpha/2}\sqrt{\overline{x}(1-\overline{x})/n}\right)$$

作为参数p的区间估计结果。

服从“0-1”分布的总体参数p的区间估计

[Example 7.14] 从一批产品中, 抽取**100**个样品, 发现其中有**75**个优质品

求 这批产品的优质品率 p 对应于置信水平 **0.95** 的置信区间.

解 已知 **n=100, α=0.05**，设总体 $X=\begin{cases}0 & \text{取到非优质品}\\ 1 & \text{取到优质品}\end{cases}$

则 X 服从 “0−1”分布: $P\{X=x\}=p^x(1-p)^{1-x}$

$x=0$ 或者 $x=1$; 其中 p 为这批产品的优质品率.

按题意，样本容量$n=100$, 在样本观测中恰有25个0与75个1, 所以 $\bar{x}=0.75$

查表得 $z_{\alpha/2}=z_{0.025}=1.96$，于是代入公式计算得

$$a=100+1.96^2=103.8406$$

$$b=-(2\times100\times0.75+1.96^2)=-153.8416$$

$$c=100\times0.75^2=56.25$$

由此得 $\hat{p}_1=0.657$, $\hat{p}_2=0.825$.

服从指数分布的总体参数θ的区间估计

总体X服从指数分布 $Exp(\theta)$，其中参数θ未知，则$E(X)=\theta$, $D(X)=\theta^2$.

对给定的置信水平1−α，有 $P\left\{\dfrac{\left|\overline{X}-\theta\right|}{\theta/\sqrt{n}}\le z_{\alpha/2}\right\}\approx 1-\alpha$

$$P\left\{\frac{\overline{X}}{1+\dfrac{z_{\alpha/2}}{\sqrt{n}}}\le\theta\le\frac{\overline{X}}{1-\dfrac{z_{\alpha/2}}{\sqrt{n}}}\right\}\approx 1-\alpha$$

如何做精确的估计?

故参数θ的近似置信区间为

$$\left(\frac{\overline{x}}{1+\dfrac{z_{\alpha/2}}{\sqrt{n}}},\ \frac{\overline{x}}{1-\dfrac{z_{\alpha/2}}{\sqrt{n}}}\right)$$

[Example 7.15]从一批电子元件中, 随机抽取 **50** 个样品, 测得它们的平均寿命为**1200** 小时, 设电子元件的使用寿命服从指数分布 $\boldsymbol{Exp(\theta)}$, 求参数θ相应于置信水平 **0.99** 的置信区间.

解 已知 $n=50,\ \bar{x}=1200,\ \alpha=0.01$， 查正态分布表得 $Z_{0.005}=2.576$.

$$\theta_1=\frac{\bar{x}}{1+\frac{z_{\alpha/2}}{\sqrt{n}}}=\frac{1200}{1+\frac{2.576}{\sqrt{50}}}=879.571$$

$$\theta_2=\frac{\bar{x}}{1-\frac{z_{\alpha/2}}{\sqrt{n}}}=\frac{1200}{1-\frac{2.576}{\sqrt{50}}}=1887.687$$

故所求参数θ的置信区间为 **(879.571, 1887.687).**

服从泊松分布的总体参数λ的区间估计

总体$\boldsymbol{X}$ 服从泊松分布$\pi(\lambda)$，其中参数λ未知，则$\boldsymbol{E(X)} = \lambda$, $\boldsymbol{D(X)} = \lambda$.

对给定的置信水平$1-\alpha$，求λ的区间估计

$$\left(\bar{X}+\frac{z_{\frac{\alpha}{2}}^{2}}{2n}-z_{\frac{\alpha}{2}}\sqrt{\frac{z_{\frac{\alpha}{2}}^{2}}{4n^{2}}+\frac{\bar{X}}{n}},\bar{X}+\frac{z_{\frac{\alpha}{2}}^{2}}{2n}+z_{\frac{\alpha}{2}}\sqrt{\frac{z_{\frac{\alpha}{2}}^{2}}{4n^{2}}+\frac{\bar{X}}{n}}\right)$$

贝叶斯估计

Bayes Estimation

为什么引入贝叶斯估计?

频率统计派的不足

对B(1, p)的进行极大似然参数估计，得到 $\hat{p} = \bar{X}$

只进行一次抽样， 观测到 Tail（$x_1 = 0$），有估计值$\hat{p} = 0$
即Tail出现概率为100%

增加实验次数n=3, 连出三次Tail（在真实 p=0.5情况下，仍然有1/8的概率），有$\hat{p} = 0$
且我们有更高的信心，因为$D(\overline{X_3}) < D(\overline{X_1})$。

计算B(1, p)的基于$\bar{X}$置信区间，有95%置信区间为

$$\left(\bar{x} - 1.96\sqrt{\bar{x}(1-\bar{x})/n}, \bar{x} + 1.96\sqrt{\bar{x}(1-\bar{x})/n}\right) = (0,0)$$

即有95%的信心认为，p落于(0,0) 区间，即$\hat{p} = 0$

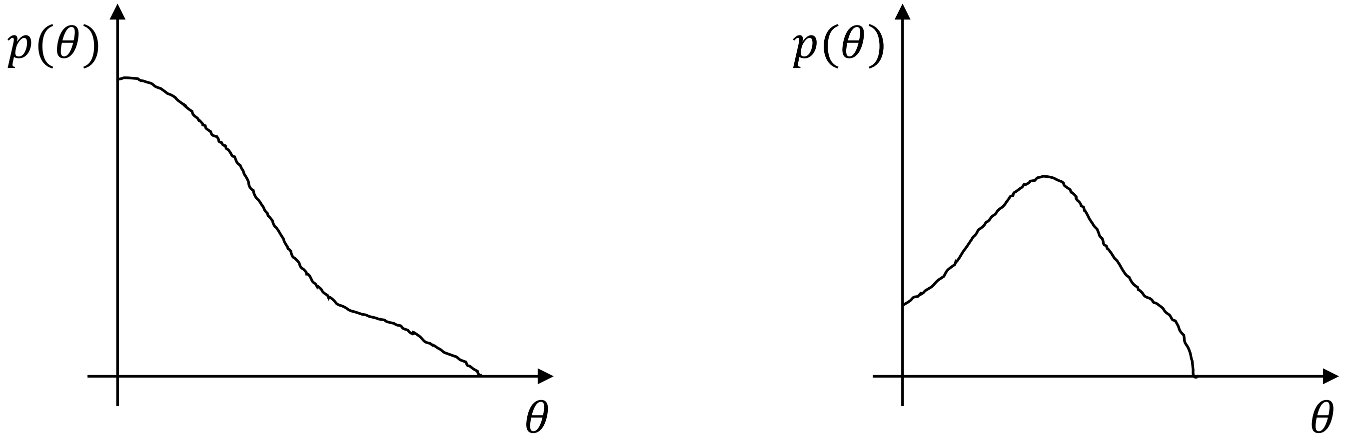
违背常识(counter intuitive)!

为什么引入贝叶斯估计?

- 经典学派（频率学派）
  - 无论是矩估计还是极大似然估计，或者其他方法，**认为参数 θ是一个简单的未知数**
  - 在抽样前，我们对θ没有任何了解，**所有的信息都来自于样本**
- 贝叶斯学派
  - 在抽样前，我们对θ已经有了一些了解，叫做“**先验知识**”
  - 贝叶斯学派进一步要求，**这种先验知识必须用θ的某种概率分布表达出来**，称之为先验分布

贝叶斯估计

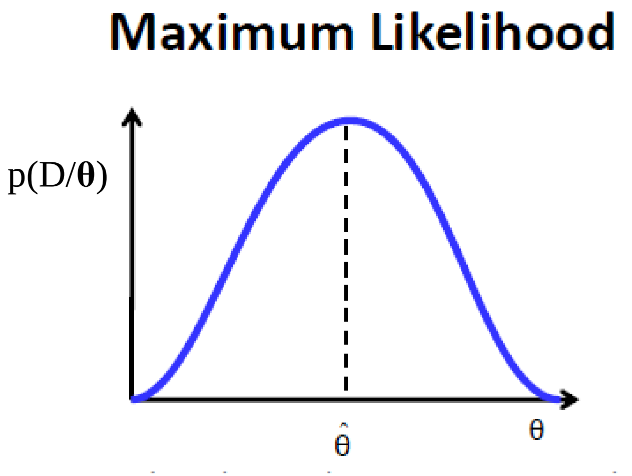
[Example 7.16] 设工厂每日生产一大批某种产品，我们想要估计当日的废品率θ，该厂在以前已经生产过很多批产品，假设过去的检验有记录保存，则确实能提供关于废品率的有效信息，据此可以画出θ的概率密度曲线

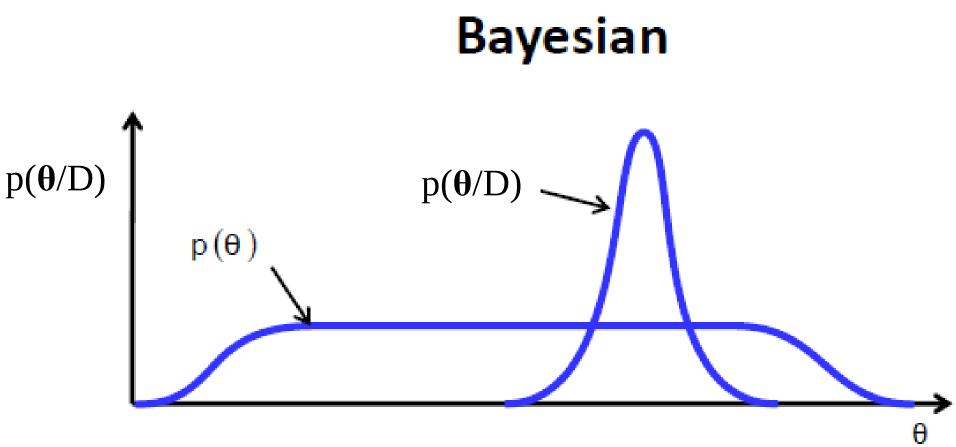


(a) (b)

- 这种先验知识，应在当前估计废品率时适当的使用，而不应弃之不顾
- 这种思想符合我们日常处理事务的习惯：当我们面临一个问题时，除考虑当前的情况外，往往还要结合以前的先例和经验

贝叶斯估计





- 在观察数据之前，参数θ用一个先验分布$p(\theta)$来描述
- 在观察数据以后，我们利用贝叶斯公式来得到后验分布$p(\theta|D)$
- 理想情况下，我们希望通过观察数据能降低我们对参数的不确定性，即使得图中的分布函数看起来更加集中

贝叶斯估计

如何确定先验概率分布$p(\theta)$?

- 客观法（如前面废品率的例子）
- 主观概率法
- 同等无知原则（如废品率$\theta \sim U(0,1)$）
- 共轭先验分布
  - 先验分布和后验分布属于同一个函数族
  - Conjugate prior – Wikipedia
- ……

陈希孺、倪国熙 《数理统计学教程》p203

贝叶斯估计

给定参数θ先验密度$p(\theta)$，总体的概率密度为$f(X|\theta)$；

因此$(\theta, X_1, \cdots, X_n)$的联合密度为

$$f(\theta, X_1, \cdots, X_n) = f(\theta) f(X_1, \cdots, X_n|\theta)$$

$$= f(\theta) \prod_{i=1}^{n} f(X_i|\theta)$$

根据贝叶斯公式，可得

后验分布　　　　先验分布　　似然函数

$$\underbrace{\boxed{f(\theta|X_1, \cdots, X_n)}}_{\text{后验分布}} = \frac{f(\theta, X_1, \cdots, X_n)}{f(X_1, \cdots, X_n)} = \frac{\underbrace{\boxed{f(\theta)}}_{\text{先验分布}} \underbrace{\boxed{\prod_{i=1}^{n} f(X_i|\theta)}}_{\text{似然函数}}}{\int f(\theta) \prod_{i=1}^{n} f(X_i|\theta)\, d\theta}$$

后验分布代表了我们在抽样后对参数θ的知识。

与贝叶斯公式的关系

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)} \qquad i = 1,2,\cdots,n$$

	贝叶斯公式	贝叶斯估计
问题	事件$B_1,\cdots,B_n$中哪一个发生了？	$\theta = ?$
先验知识	$P(B_1),\cdots,P(B_n)$	$p(\theta)$
当前知识	事件A发生了	样本$X_1,\cdots,X_n$
后验知识	$P(B_1\|A),\cdots,P(B_n\|A)$	$p(\theta\|X_1,\cdots,X_n)$

贝叶斯估计

$$f(\theta|X_1,\cdots,X_n)=\frac{f(\theta,X_1,\cdots,X_n)}{f(X_1,\cdots,X_n)}=\frac{f(\theta)\prod_{i=1}^{n}f(X_i|\theta)}{\int f(\theta)\prod_{i=1}^{n}f(X_i|\theta)\,d\theta}$$

贝叶斯估计的优化目标是估计**一个分布**，在得到该后验分布后，对参数θ的任何统计推断都只能基于这个后验分布。

如：θ的点估计为该后验分布的期望

注：按照前面的描述， $f(\theta)$必须是一个概率（密度）函数，即$\int f(\theta)\,d\theta=1$。但贝叶斯估计中只需要保证$p(\theta|X_1,\cdots,X_n)$是一个密度函数即可，即使$\int f(\theta)\,d\theta\neq 1$也无妨，这样的$f(\theta)$被称为广义先验密度。

贝叶斯估计

[Example 7.17] 做n次独立重复试验，每次观察事件A是否发生，A在每次试验中发生的概率为p，求p的贝叶斯点估计

解：引入$X_i = 1$或0，表示第i次试验事件A是否发生，$i = 1,2,\cdots,n$

假设p的先验分布为$f(p)$，则p的后验分布为

$$f(p|X_1,\cdots,X_n) = \frac{f(p)p^X(1-p)^{n-X}}{\int_0^1 f(q)q^X(1-q)^{n-X}dq}, 0 \le p \le 1$$

其中$X = \sum_{i=1}^n X_i$。

此分布的期望就是p的贝叶斯点估计

$$\hat{p} = \int_0^1 pf(p|X_1,\cdots,X_n)dp = \int_0^1 \frac{f(p)p^{X+1}(1-p)^{n-X}}{\int_0^1 f(q)q^X(1-q)^{n-X}dq}dp$$

贝叶斯估计

$$\hat{p}=\int_0^1 pf(p|X_1,\cdots,X_n)dp=\frac{\int_0^1 f(p)p^{X+1}(1-p)^{n-X}dp}{\int_0^1 f(q)q^X(1-q)^{n-X}dq}$$

如何选择$f(p)$？这里我们利用同等无知原则，事先认为p取[0, 1]内的一切值都有同等可能，即$p\in U(0,1)$作为p的先验分布。

这时$f(p)=1,0\le p\le 1$，带入可得：

$$\hat{p}=\frac{\int_0^1 p^{X+1}(1-p)^{n-X}dp}{\int_0^1 q^X(1-q)^{n-X}dq}=\frac{B(X+2,n-X+1)}{B(X+1,n-X+1)}$$

$$=\frac{\Gamma(X+2)\Gamma(n-X+1)/\Gamma(n+3)}{\Gamma(X+1)\Gamma(n-X+1)/\Gamma(n+2)}=\frac{X+1}{n+2}$$

贝叶斯估计

贝叶斯估计 $\hat{p} = \frac{X+1}{n+2}$

这个估计与频率$\frac{X}{n}$有差别，当n很大时不显著，而在n很小时较显著

换一个角度看，当n很小时，用贝叶斯估计比用频率更合理，因为当n很小时，试验结果出现 $X = 0$ 或者 $X = n$ 的可能性较大（不可忽视），这时，$\frac{X}{n}$ 把p估计为0或者1，就太极端了；若按照贝叶斯估计，在这两种情况下，分布可以给出估计值$\frac{1}{n+2}$和$\frac{n+1}{n+2}$，相对而言更合理一些。

共轭先验（conjugate prior)

https://en.wikipedia.org/wiki/Conjugate_prior

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters[note 1]	Interpretation of hyperparameters	Posterior predictive[note 2]
Bernoulli	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + n - \sum_{i=1}^{n} x_i$	α successes, β failures[note 3]	$p(\tilde{x}=1) = \frac{\alpha'}{\alpha'+\beta'}$

$$\hat{p} = \frac{\alpha'}{\alpha' + \beta'} = \frac{X+1}{n+2}$$

- 数学形式上的简洁性
- 超参数可解释性

贝叶斯估计

[Example 7.18] 设$X_1, \cdots, X_n$ 是来自正态总体$N(\theta, 1)$的简单随机样本，已知θ的先验分布为正态分布$N(\mu, \sigma^2)$（μ, σ^2 已知），求θ的贝叶斯估计

解 θ的先验分布和样本的分布为

$$p(\theta) = (\sqrt{2\pi}\sigma)^{-1} \exp\left[-\frac{1}{2\sigma^2}(\theta - \mu)^2\right]$$

$$f(x|\theta) = (\sqrt{2\pi})^{-1} \exp\left[-\frac{1}{2}(x - \theta)^2\right]$$

θ的后验密度为

$$p(\theta|X_1, \cdots, X_n) = \exp\left[-\frac{1}{2\sigma^2}(\theta - \mu)^2 - \frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2\right]/I$$

其中I是与θ无关的变量

将exp中的部分关于θ配方

$$-\frac{1}{2\sigma^2}(\theta-\mu)^2-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2=-\frac{1}{2\eta^2}(\theta-t)^2+J$$

其中

$$t=\left(n\bar{X}+\mu/\sigma^2\right)/\left(n+1/\sigma^2\right)$$
$$\eta^2=1/\left(n+1/\sigma^2\right)$$

而J与θ无关，带入后验分布中可得

$$p(\theta|X_1,\cdots,X_n)=I_1\exp\left[-\frac{1}{2\eta^2}(\theta-t)^2\right] \qquad I_1=Ie^J$$

因为 $\int_{-\infty}^{\infty}p(\theta|X_1,\cdots,X_n)\mathrm{d}\theta=1$ 所以 $I_1=(\sqrt{2\pi}\eta)^{-1}$

因此，θ的后验分布就是正态分布$N(t,\eta^2)$

θ的点估计为

$$\hat{\theta} = t = \underbrace{\frac{n}{n+1/\sigma^2}\bar{X}}_{\text{来自于样本（观察）}} + \underbrace{\frac{1/\sigma^2}{n+1/\sigma^2}\mu}_{\text{来自于先验}}$$

上述θ的点估计可以看作是$\overline{X}$和μ的一种加权平均（折衷）

- n为样本容量，n越大，样本信息越多，$\overline{X}$的权重越大
- 对μ而言，σ^2越大，表示先验信息越不确定θ在μ附近，反之，σ^2越小，表示先验信息比较确定θ在μ附近，因此μ的权重和σ^2成反比

最大后验估计

Maximum a posteriori estimation （MAP）

完整的贝叶斯估计是估计参数的分布，涉及两个分布的运算，对于稍复杂的学习过程，计算难度太大

实际上，当取估计到的参数分布概率最大的点作为最佳参数，那么分布估计就变为了参数的点估计，即MAP

$$\hat{\theta}_{MAP}(x) =$$

$$\hat{\theta} = \mathrm{argmax}_{\theta} f(\theta) = \arg\max_{\theta} \frac{f(\theta) f(x|\theta)}{\underbrace{\int_{\beta} f(\beta) f(x|\beta)\, d\beta}_{\text{与}\theta\text{无关}}} = \mathrm{argmax}_{\theta} f(\theta) f(x|\theta)$$

根据上式，最大后验估计转变为求函数极值的问题

最大似然估计是最大后验估计的特殊情况， $\theta \sim U(a, b)$

求Example 7.18 中θ的最大后验估计

贝叶斯区间估计

在有了先验分布$p(\theta)$和样本$X_1,\cdots,X_n$后，算出后验分布$p(\theta|X_1,\cdots,X_n)$

与前面讲的区间估计类似，我们可以找两个统计量$\hat{\theta}_1$和$\hat{\theta}_2$，使

$$\int_{\hat{\theta}_1}^{\hat{\theta}_2} p(\theta|X_1,\cdots,X_n)\,d\theta = 1-\alpha$$

区间$[\hat{\theta}_1,\hat{\theta}_2]$的意思是：在所得的后验分布情形下，$\theta$落在这个区间的概率是$1-\alpha$。

$1-\alpha$被称为后验信度，可解释为：在观察到样本后，我对区间$[\hat{\theta}_1,\hat{\theta}_2]$能包含未知参数$\theta$的相信程度为$1-\alpha$。

后验信度与前面讲的置信度含义相似，但理论架构不同

贝叶斯区间估计

[Example 7.18 续] 设$X_1,\cdots,X_n$ 是来自正态总体$N(\theta,1)$的简单随机样本，已知θ的先验分布为正态分布$N(\mu,\sigma^2)$（μ,σ^2 已知），求θ的后验信度为$1-\alpha$的贝叶斯区间估计

解：θ的后验分布就是正态分布$N\left(t,\eta^2\right)$

$$t=\left(n\bar{X}+\mu/\sigma^2\right)/\left(n+1/\sigma^2\right)$$
$$\eta^2=1/\left(n+1/\sigma^2\right)$$

θ的后验信度为$1-\alpha$的贝叶斯区间估计为

$$\left[t-\eta z_{\alpha/2},t+\eta z_{\alpha/2}\right]$$

注：满足后验信度为$1-\alpha$的区间有很多，仍然选择$\hat{\theta}_2-\hat{\theta}_1$最小的那个

贝叶斯区间估计

[Example 7.17 续]做n次独立重复试验，每次观察事件A是否发生，A在每次试验中发生的概率为p，求p的后验信度为$1-\alpha$的贝叶斯区间估计

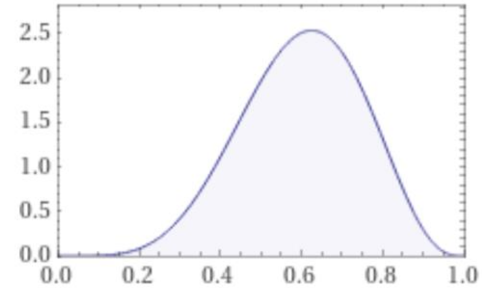
解：当取$U(0,1)$为先验分布时，p的后验分布为

$$f(p|X_1,\cdots,X_n)=\frac{p^X(1-p)^{n-X}}{\int_0^1 q^X(1-q)^{n-X}dq}=\frac{p^X(1-p)^{n-X}}{B(X+1,n-X+1)}$$

Beta分布

要找到$\hat{p}_1,\hat{p}_2$使得

$$\int_{\hat{p}_1}^{\hat{p}_2}\frac{p^X(1-p)^{n-X}}{B(X+1,n-X+1)}dp=1-\alpha$$



类似χ^2和F分布，我们采用以下方式计算$\hat{p}_1,\hat{p}_2$

$$\int_{-\infty}^{\hat{p}_1}\frac{p^X(1-p)^{n-X}}{B(X+1,n-X+1)}dp=\frac{\alpha}{2},\int_{\hat{p}_2}^{\infty}\frac{p^X(1-p)^{n-X}}{B(X+1,n-X+1)}dp=\frac{\alpha}{2}$$

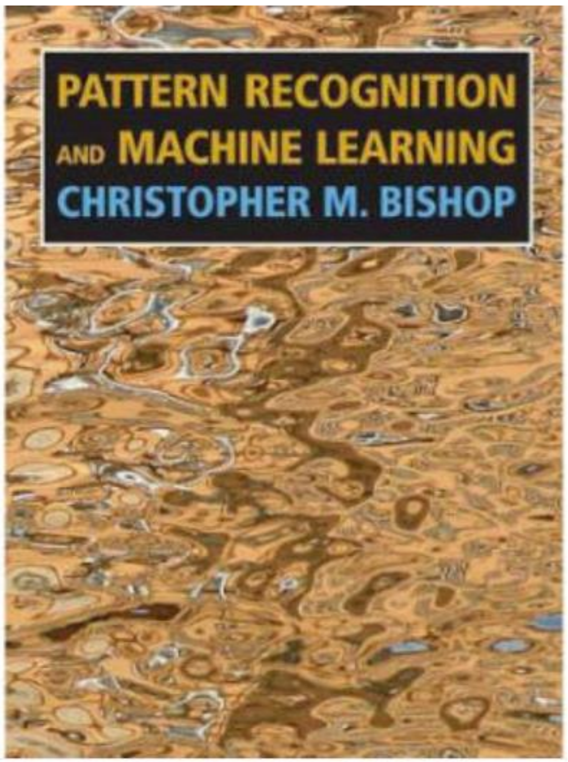
贝叶斯区间估计

与前面的区间估计相比，贝叶斯区间估计无疑是容易的，虽然在求$\hat{\theta}_1, \hat{\theta}_2$时比较繁琐，但是可以用计算机实现。

前面讲的区间估计，基于几个统计量刚好服从t分布，χ^2分布和F分布，这只是少见的几个特例（刚好实际中用的比较多）。

往往由于分布问题无法解决，只好用大样本理论；如果样本容量不大，容易产生比较大的误差，而且我们无法控制由此而产生的误差。贝叶斯方法则不存在这样的问题

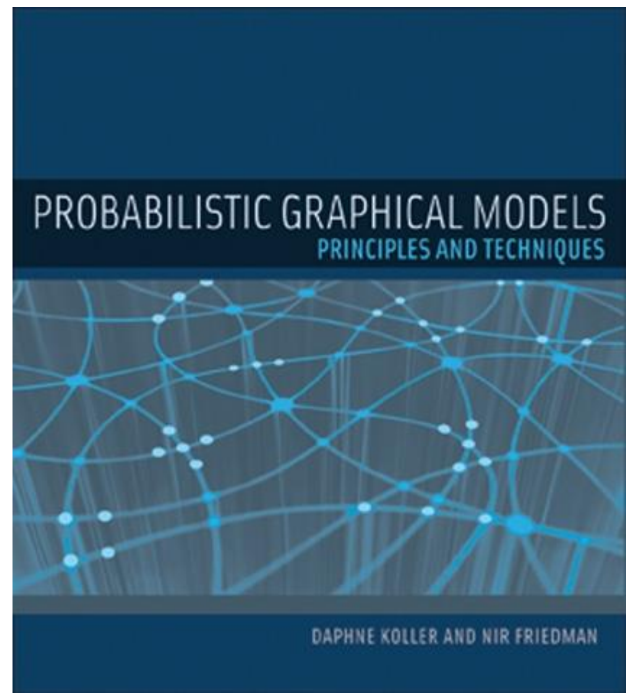
进一步学习



Daphne Koller

概率图模型 专项课程

https://www.coursera.org/specializations/probabilistic-graphical-models



谢谢！