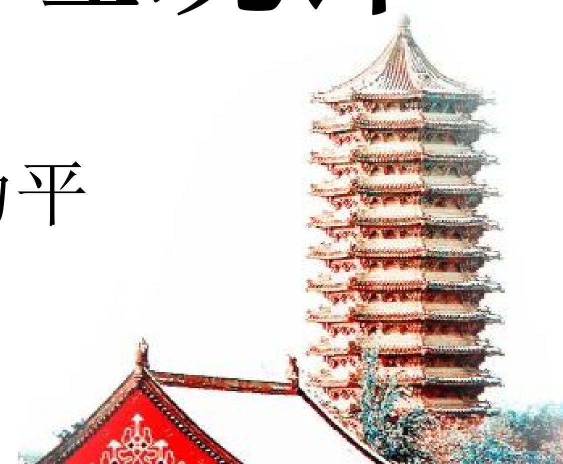
2023-2024春　数理统计

主讲人：数学科学学院 刘力平



第一章 绪论

- 地址：智华楼335
- 电话：13681499285
- Email：liping@math.pku.edu.cn

- 助教： 王啸辰：xcwang1998@126.com，13717689550
  杨云帆：2001110072@pku.edu.cn，18811058616

- 教材：陈家鼎等，《数理统计学讲义》第3版，高等教育出版社。

第一章 绪论

- **参考书:**

1. 陈希孺，《数理统计引论》，科学出版社;
2. D. Freedman等（魏宗舒等译），《统计学》，中国统计出版社;
3. Lehmann，《Theory of Point Estimation》, John Wiley and Sons;
4. Lehmann，《Testing Statistical Hypothesis》, John Wiley and Sons;
5. 陈希孺，《数理统计学简史》，湖南教育出版社。

第一章 绪论

- 课堂纪律：学校要求；不影响他人（但强烈建议上课，ppt太少）。

- 平时答疑时间：周二2-4pm。

- 习题：交法（周四课间，可延至周日晚）；分数；标准；

- 与平行课程关系：教学大纲独立，要求独立，各自单独考试，单独改卷，各自评分。

- 考试：20（习题）+20（期中考试，4月？）+60（期末考试）

第一节 数理统计学的研究对象

- 数学分支/一级学科?

- 与数学/概率论的区别：（统计学家一般不参加国际数学家大会）

1. 更明确的背景；
2. “解”的标准不同，对/错 vs 好/坏；
3. 工作方法不同，更多计算、模拟等；
4. 思维方式不同。
   - 平均身高比较
   - 充分利用统计正则性
   - 读论语得诺贝尔奖?

第一节 数理统计学的研究对象

- 国际/国内数理统计发展

- 北京大学状况、许宝騄先生

  人大等：经济统计；
  北大等：数理统计；
  流行病学统计等。

- 不要为“容易”学统计

第一节 数理统计学的研究对象

● 本课程特点：

基础课，强调常见背景、基本思想和经典方法的掌握和应用；

不强调逻辑证明；

无软件操作。

● 内容：第二、三、四章大部分，第五章前2节，第七章前2节及扩充；第五章第3节至第七章视情况部分简介。

● 统计正则性

大数律、中心极限定理等

- 掷硬币100次，正面45—55次个数？极强的规律性；
- 重对数率（如掷硬币画图）；
- … …。

第一节 数理统计学的研究对象

●几个例子：

1．民意调查：1932年美总统选举，罗斯福vs兰登，文学摘要vs盖洛普。

样本量？抽样方式？Sampling bias。

文学摘要通过超2000万样本预测兰登当选，但预测失败！

因其样本有偏：主要来源于电话号码册与俱乐部名单，多为富人。

第一节 数理统计学的研究对象

2. 孟德尔遗传实验（非常严谨）数据是否存在造假？R. A. Fisher的结论。

数据比例太接近3/4: 1/4 ?

数据分布是客观自然的，自有其分布。

助手之过，凑3/4?

3. 商业交易中正品率的判定?

如何根据部分样品判断本批商品的次品率不超标。

第一节 数理统计学的研究对象

4. 铝厂非法罢工后，资方发现，因停电，某元件的寿命缩短，诉工会赔偿损失。

该赔多少?

法庭上双方均聘请统计学家辩论… …

某统计期刊邀请全球统计学家探讨。

第二节 数理统计学的基本概念

●总体和个体：用分布函数$F(\cdot)$，或随机变量（向量）X表示总体。

●样本：被抽取到的个体，用X_1，$\cdots$, X_n或x_1，$\cdots$, x_n表示，

称n为样本量；

习惯上大小写的区别。

若假设独立同分布（iid），称简单随机样本。

●统计量：样本的（可测）函数$g(x_1,\cdots x_n)$。

●其他的，逐步介绍。

第三节 数理统计学发展简史

请自己阅读。

详细内容请见陈希孺的《数理统计学简史》。

第二章 估计

第一节 参数估计的方法

背景：我们已知总体的分布类型，但不知其中的参数值，希望由数据对参数给出估计。如长期实践中得知，轴承寿命服从Weibull分布，但各种厂家、型号的参数不同。

例如，X是收入，随机抽取样本$x_1, \cdots, x_n$，设近似服从$N(\mu, \sigma^2)$，则$\theta = (\mu, \sigma^2)$，$\Theta \subset R^2$为上半平面。我们的兴趣为：① 平均收入，则$g_1(\theta) = \mu$；② 贫富分化程度，则$g_2(\theta) = \sigma^2$；③ 贫困人口比例，则$g_3(\theta) = P(X < C) = P\left(\frac{X-\mu}{\sigma} < \frac{C-\mu}{\sigma}\right) = \Phi(\frac{C-\mu}{\sigma})$，其中$C$为贫困线。

数学表述：X的分布函数为$F(x, \theta)$，密度函数（或概率函数）为$f(x, \theta)$，其中$\theta = (\theta_1, \cdots, \theta_m) \in \Theta \subset R^m$，$f$的形式已知，问题为通过样本$x_1, \cdots, x_n$估计目标$g(\theta)$。

第一节 参数估计的方法

一、最大似然估计法（Maximum Likelihood Estimate，简记为MLE）

● 似然函数：称

$$L(x_1, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

为参数θ的，关于样本$x_1, \cdots, x_n$的似然函数。

概率论中，θ不变，$x_1, \cdots, x_n$是变元，称联合分布；

数理统计中，$x_1, \cdots, x_n$已知（不变），θ未知，是函数变元。

● 称似然函数的（一个）最大值点（若存在）$\widehat{\theta}_n = \widehat{\theta} = \varphi(x_1, \cdots, x_n)$为$\theta$的最大似然估计。

第一节 参数估计的方法

一、最大似然估计法（Maximum Likelihood Estimate，简记为MLE）

● 似然函数：称

$$L(x_1, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

为参数θ的，关于样本$x_1, \cdots, x_n$的似然函数。

概率论中，θ不变，$x_1, \cdots, x_n$是变元，称联合分布；

数理统计中，$x_1, \cdots, x_n$已知（不变），θ未知，是函数变元。

● 称似然函数的（一个）最大值点（若存在）$\widehat{\theta}_n = \widehat{\theta} = \varphi(x_1, \cdots, x_n)$为$\theta$的最大似然估计。

第一节 参数估计的方法

- 经典统计中最重要的估计方法，首先想到。

- 具有许多好的性质，某些标准下最优。

- 思想：用最可能（most likely）产生数据的参数值作为估计值。

例如：射击10枪，7中，$X_1, \cdots, X_{10}$独立同分布，共同分布为$B(1,\ p)$。若只许你猜p为0.2或0.8，如何猜；若允许你在$(0,\ 1)$中猜，如何猜?

- 求解方法：一般为数学上求最大值方法，取对数（乘积式变为和式），求导，令其等于0。

- 如解不唯一，任何一个均为MLE。

第一节 参数估计的方法

● 强相合性（strong consistency）：

$$P\left(\lim_{n\to\infty}\widehat{\theta}_n=\theta\right)=1$$ （实变中的几乎处处收敛）

● （弱）相合性（consistency）：$\forall\varepsilon>0$，

$$\lim_{n\to\infty}P\left(\|\widehat{\theta}_n-\theta\|<\varepsilon\right)=1$$ （实变中的依测度收敛）

● 理论上，强相合性可推出弱相合性，反之不成立。应用中，一般无区别（一般说相合性，指的是弱相合性）。

● 直观意义：随样本量增大，估计值可任意接近目标。

● 区别原因：随机变量的收敛比一般收敛复杂。

第一节 参数估计的方法

几种常见分布：

1. 两点（Bernoulli）分布， $X_i \sim B(1, p)$，$x_1, \cdots, x_n$为数据，取值0、1，

$$P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}$$
$$L(x_1, \cdots, x_n, p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$
$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$log\big(L(p)\big) = \sum_{i=1}^{n} x_i log(p) + \left(n - \sum_{i=1}^{n} x_i\right) log(1-p)$$

$$\frac{\partial log(L(p))}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0$$

唯一解为$\hat{p} = \hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} x_i =: \bar{x} = \frac{v}{n}$，概率论中常说的频率。

相合性：由强大数律，$\lim_{n\to\infty} \hat{p}_n = p$ (a.s.)，即频率→概率。

第一节 参数估计的方法

2. 指数分布，$X_i \sim Exp(\lambda)$，

$$p(x) = \lambda e^{-\lambda x} \quad (\text{当} x>0)$$

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

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

$$\frac{\partial log(L(\lambda))}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

唯一解为$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = {}^{1}/_{\overline{x}}$。

相合性：$\overline{x} \to E(X) = 1/\lambda$，故 $\hat{\lambda} \to \lambda$ (a.s.)

$1/\lambda$的直观意义：平均寿命。

第一节 参数估计的方法

3. 正态分布，$X_i \sim N(\mu,\ \sigma^2)$，令$\delta = \sigma^2$，

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

$$log\big(L(\mu, \delta)\big) = nlog(C) - \frac{n}{2}log(\delta) - \frac{1}{2\delta}\sum_{i=1}^{n}(x_i - \mu)^2$$

其中C为某常数，

似然方程组为：

$$\begin{cases} \dfrac{\partial logL}{\partial \mu} = \dfrac{1}{\delta}\displaystyle\sum_{i=1}^{n}(x_i - \mu) = 0 \\ \dfrac{\partial logL}{\partial \delta} = -\dfrac{n}{2\delta} + \dfrac{1}{2\delta^2}\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \end{cases}$$

唯一解为：

$$\begin{cases} \hat{\mu} =: \dfrac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \qquad \text{（样本均值）} \\ \hat{\sigma}^2 =: \hat{\delta} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases}$$

第一节 参数估计的方法

相合性:

由大数率， $P\left(\lim_{n\to\infty}\hat{\mu}=\mu\right)=1$;

又由 $P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_i^2=EX^2\right)=1$,

故而

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_i^2-\lim_{n\to\infty}\hat{\mu}^2=EX^2-(EX)^2\right)=1,$$

即 $P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=var(X)\right)=1$,

即 $P\left(\lim_{n\to\infty}\hat{\delta}=\delta\right)=1$。

注意其中有: $\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}X_i\right)^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2$ 这个常用等式。

第一节 参数估计的方法

4. 威布尔（Weibull）分布

$$F(x, m, \eta) = 1 - exp(-(\frac{x}{\eta})^m) \quad （当 x > 0，参数 m > 0，\eta > 0。）$$

$$\Rightarrow \quad f(x, m, \eta) = \frac{m}{\eta^m} x^{m-1} e^{-(\frac{x}{\eta})^m}$$

故 $$L(m, \eta) = \frac{m^n}{\eta^{mn}} (\prod_{i=1}^{n} x_i)^{m-1} e^{-\frac{\sum_{i=1}^{n} x_i^m}{\eta^m}}。$$

取对数，求偏导，可得似然方程组。计算机求解。

无显式解，如何证明相合性?

第一节 参数估计的方法

5. 均匀分布 $X \sim U[a, b]$（或简化版本 $X \sim U[0, \theta]$）

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

$$L(a, b) = \left(\frac{1}{b-a}\right)^n \prod_{i=1}^{n} I_{[a, b]}(x_i),$$

所以 $L(a, b) = \begin{cases} \left(\dfrac{1}{b-a}\right)^n & 当 \quad a \le \min\limits_{1\le i\le n} x_i \ 且\ b \ge \max\limits_{1\le i\le n} x_i \\ 0 & 否则 \end{cases}$

故 $\hat{a} = \min\limits_{1\le i\le n} x_i$，$\hat{b} = \max\limits_{1\le i\le n} x_i$ （不连续，不可导）。

相合性：教材中有证明，用到实变函数方法。

第一节 参数估计的方法

二、矩估计法（Moment Estimate）

设$\theta=(\theta_1,\cdots,\theta_m)$，$X$的$k$阶矩存在有限，（理论）值为

$$\begin{cases} V_1=E(X^1)=g_1(\theta_1,\cdots,\theta_m) \\ \quad\cdots \quad \cdots \\ V_m=E(X^m)=g_m(\theta_1,\cdots,\theta_m) \end{cases}$$

样本矩为

$$\begin{cases} \widetilde{V}_1=\frac{1}{n}\sum_{i=1}^{n}x_i \\ \quad\cdots \quad \cdots \\ \widetilde{V}_m=\frac{1}{n}\sum_{i=1}^{n}x_i^m \end{cases}$$

由大数律（一定条件下），$\widetilde{V}_k\to V_k$（当样本量→∞），令其相等，得到估计方程组

第一节 参数估计的方法

$$\begin{cases} g_1(\theta_1, \cdots, \theta_m) = \tilde{V}_1 \\ \quad\cdots \quad \cdots \\ g_m(\theta_1, \cdots, \theta_m) = \tilde{V}_m \end{cases}$$

若解为

$$\begin{cases} \tilde{\theta}_1 = f_1(\tilde{V}_1, \cdots, \tilde{V}_m) \\ \quad\cdots \quad \cdots \\ \tilde{\theta}_m = f_m(\tilde{V}_1, \cdots, \tilde{V}_m) \end{cases}$$

则称其为$(\theta_1, \cdots, \theta_m)$的矩估计。（存在性、唯一性）

矩估计历史上曾有重要地位（Pearson），后Fisher力推MLE。

有时MLE收敛速度更快。

第一节 参数估计的方法

几种常见分布：

1. 两点（Bernoulli）分布， $X_i \sim B(1, p)$，

此时$m=1$，而$V_1=EX=p$，$\widetilde{V}_1=\bar{x}$，故$\widetilde{p}=\bar{x}$，与MLE相同。

2. 指数分布，$X_i \sim Exp(\lambda)$，

此时$V_1=EX={}^{1}/_{\lambda}$，$\widetilde{V}_1=\bar{x}$，故$\tilde{\lambda}={}^{1}/_{\bar{x}}$，与MLE一致。

3. 正态分布，$X_i \sim N(\mu, \sigma^2)$

此时$m=2$，$V_1=EX=\mu$，$\widetilde{V}_1=\bar{x}$，

$V_2=EX^2=\sigma^2+\mu^2$，$\widetilde{V}_2=\frac{1}{n}\sum_{i=1}^{n}x_i^2$，

第一节 参数估计的方法

解方程易知，$(\tilde{\mu}, \tilde{\sigma}^2)$与MLE相同。

4. 威布尔（Weibull）分布

$$V_1 = EX = \eta\Gamma\left(\frac{1}{m}+1\right)，V_2 = \eta^2\Gamma\left(\frac{2}{m}+1\right)，$$

故估计方程为：

$$\begin{cases} \eta\Gamma\left(\dfrac{1}{m}+1\right) = \dfrac{1}{n}\sum_{i=1}^{n} x_i \\ \eta^2\Gamma\left(\dfrac{2}{m}+1\right) = \dfrac{1}{n}\sum_{i=1}^{n} x_i^2 \end{cases}$$

与MLE不同。

第一节 参数估计的方法

5. 均匀分布 $X \sim U[0,\ \theta]$

此时$m = 1$，令$EX = \frac{\theta}{2} = \bar{x}$，得$\tilde{\theta} = 2\bar{x}$，

与MLE（$\hat{\theta} = \max\limits_{1 \le i \le n} x_i$）不同。

哪个好，或各自优缺点?

● MLE（$\max\limits_{1 \le i \le n} x_i$）概率为1地偏小。

● 矩估计在明知对某个$i \le n$，$\theta \ge x_i$（x_i较大）时，会仍用偏小的$2\bar{x}$作为估计。

第一节 参数估计的方法

三、一个实例

例1. “序列号”估计方法（背景：二战）

模型为N未知，从$\{1, 2, \cdots, N\}$中随机抽取n个，记为$\{0 < k_1 < \cdots < k_n \le N\}$，希望利用此数据估计$N$。

首先可以用$\frac{1}{n-1}\sum_{i=2}^{n}(k_i - k_{i-1} - 1)$对两个数据间的平均间隔进行估计，从而$N$的一个自然的估计为：

$$\widehat{W}_1 = k_n + \frac{1}{n-1}\sum_{i=2}^{n}(k_i - k_{i-1} - 1)$$
$$= k_n + \frac{1}{n-1}(k_n - k_1) - 1 = \frac{n}{n-1}k_n - \frac{1}{n-1}k_1 - 1$$

可以证明，$E(\widehat{W}_1) = N$。

第一节 参数估计的方法

郑忠国提出了改进的估计：

$$\widehat{W}_2 = k_n + \frac{1}{n}\sum_{i=1}^{n}(k_i - k_{i-1} - 1)$$
$$= \frac{n+1}{n}k_n - 1$$

其中$k_0 =: 0$。亦可证明，$E\left(\widehat{W}_2\right) = N$，但其方差更小。

期望、方差表示什么？概率空间如何定义？

离散型均匀分布。

第二节 估计的优良性标准

当用$\varphi(x_1,\cdots,x_n)$估计目标$g(\theta)$时，希望$\varphi(x_1,\cdots,x_n)$与目标$g(\theta)$距离越近越好。但因为$\varphi(x_1,\cdots,x_n)$是随机变量，$g(\theta)$是θ的函数，也可以取不同的值，所以如何定义“距离”很重要。

在数理统计中，对同一随机事件，当参数θ取不同的值时，对应的概率测度也不同。为明确起见，记可测空间$(\Omega,\mathcal{F})$上的概率分布族为$\{P_\theta,\theta\in\Theta\}$，记相应的期望、方差等为$E_\theta()$、$Var_\theta()$等。

定义1. 称$\varphi(X_1,\cdots,X_n)$为$g(\theta)$的无偏估计，若

$$E_\theta\varphi(X_1,\cdots,X_n)=g(\theta),\qquad \forall\theta\in\Theta$$

这是较正常的要求。

第一节的常用分布中，两点分布p的MLE、正态分布中μ的MLE等都是无偏的，均匀分布的矩估计也是，但正态分布中σ^2的MLE不是无偏的!

第二节 估计的优良性标准

定义2. 设$\varphi(X_1,\cdots,X_n)$为$g(\theta)$的一个估计，称

$$M_\theta(\varphi)=E_\theta[\varphi(X_1,\cdots,X_n)-g(\theta)]^2$$

为φ的均方误差（Mean square error，MSE）。

若φ是无偏的，则$M_\theta(\varphi)=Var_\theta(\varphi)$。

定义3. 对于$g(\theta)$的两个估计φ_1，φ_2，若$\forall\theta\in\Theta$，有$M_\theta(\varphi_1)\le M_\theta(\varphi_2)$，则称$\varphi_1$不次于$\varphi_2$；若还存在 $\theta_0\in\Theta$，使得 $M_{\theta_0}(\varphi_1)<M_{\theta_0}(\varphi_2)$，则称$\varphi_1$比 φ_2有效。

例如，若 X的方差存在有限，θ是X的均值，令$\varphi_1(X_1,\cdots,X_n)=\frac{1}{n}\sum_{i=1}^{n}X_i$，$\varphi_2(X_1,\cdots,X_n)=\sum_{i=1}^{n}\lambda_iX_i$，其中$\lambda_i$满足$\sum_{i=1}^{n}\lambda_i=1$，则$\varphi_1$，$\varphi_2$均无偏；$\varphi_1$不次于$\varphi_2$；若$\lambda_i$不全相等，则$\varphi_1$比$\varphi_2$有效。

第二节 估计的优良性标准

●两个估计并不一定总能比较！一般地，不能找到不次于所有其他估计的估计量。（？）例如用常数$\varphi(X_1,\cdots,X_n)\equiv g(\theta_0)$估计$g(\theta)\cdots$。

●缩小范围，仅考虑无偏估计。

定义4. 称$\varphi(X_1,\cdots,X_n)$为$g(\theta)$的（一致）最小方差无偏估计（MVUE），如果它无偏，且对任意的$g(\theta)$的无偏估计$\psi(X_1,\cdots,X_n)$，有$M_\theta(\varphi)\le M_\theta(\psi)$（对$\forall\theta\in\Theta$）。（$\varphi$：\phi；$\psi$：\psi）

●最小方差无偏估计在许多常见情形下存在唯一，并可以求出。

●是一定标准下的“最优良”的估计。还有其他标准。

●如何求？可以通过利用充分统计量。

第二节 估计的优良性标准

定义5. 设$X_1, \cdots, X_n$为简单随机样本，称统计量$U=\varphi(X_1, \cdots, X_n)$是$\theta$的充分（sufficient）统计量，若似然函数$L(x_1, \cdots, x_n, \theta)$可表示为

$$q[\varphi(x_1, \cdots, x_n), \theta]\cdot h(x_1, \cdots, x_n)$$

其中$h(x_1, \cdots, x_n)$是不依赖于θ的非负函数。

●直观意义：充分统计量包含了样本$X_1, \cdots, X_n$中关于参数θ的全部信息。

●$\varphi_0(X_1, \cdots, X_n)=(X_1, \cdots, X_n)$一定是$\theta$的充分统计量。

●引入定义5的目的：降低样本的维数及复杂度时，不丢失关于θ的信息。

●等价表示：给定u_0及可测集A，$P_\theta((X_1, \cdots, X_n)\in A|U=u_0)$与$\theta$无关。

----------------end 2024.02.22

有了充分统计量U，构造$g(\theta)$的估计时，可以仅考虑利用U即可。易知U的维数越低越好。

第二节 估计的优良性标准

定义5. 设$X_1, \cdots, X_n$为简单随机样本，称统计量$U=\varphi(X_1, \cdots, X_n)$是$\theta$的充分（sufficient）统计量，若似然函数$L(x_1, \cdots, x_n, \theta)$可表示为

$$q[\varphi(x_1, \cdots, x_n), \theta] \cdot h(x_1, \cdots, x_n)$$

其中$h(x_1, \cdots, x_n)$是不依赖于θ的非负函数。

●直观意义：充分统计量包含了样本$X_1, \cdots, X_n$中关于参数θ的全部信息。

●$\varphi_0(X_1, \cdots, X_n)=(X_1, \cdots, X_n)$一定是$\theta$的充分统计量。

●引入定义5的目的：降低样本（函数）的维数及复杂度时，不丢失关于θ的信息。

●等价表示：给定u_0及可测集A，$P_\theta((X_1, \cdots, X_n) \in A | U=u_0)$与$\theta$无关。

有了充分统计量U，构造$g(\theta)$的估计时，可以仅考虑利用U即可。易知U的维数越低越好。

第二节 估计的优良性标准

几个例子：

1. 两点分布

$$L(x_1, \cdots, x_n, p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$
$$= (1-p)^n\left(\frac{p}{1-p}\right)^{\sum_{i=1}^{n} x_i} \cdot 1$$

即$\sum_{i=1}^{n} X_i$与p的函数乘1，所以$\varphi(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i$是$p$的充分统计量。（$n$次射击后，欲估计射击命中率，仅需知道总击中次数，不需知道具体结果。）

对任意$1 \le k < n$，（$\sum_{i=1}^{k} x_i, \sum_{i=k+1}^{n} x_i$）也是充分统计量。

2. 指数分布

$$L(x_1, \cdots, x_n, \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

所以$\sum_{i=1}^{n} X_i$是λ的充分统计量。同理，（$\sum_{i=1}^{k} x_i, \sum_{i=k+1}^{n} x_i$）也是。

第二节 估计的优良性标准

3. 正态分布

$$L(x_1, \cdots, x_n, \mu, \delta) = \left(\frac{1}{\sqrt{2\pi\delta}}\right)^n e^{-\frac{1}{2\delta}\sum_{i=1}^{n}(x_i-\mu)^2}$$
$$= (2\pi\delta)^{-\frac{n}{2}} e^{\left\{-\frac{1}{2\delta}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\mu)^2\right]\right\}}$$

故$\varphi(X_1, \cdots, X_n) = (\overline{X}, \sum_{i=1}^{n}(X_i-\overline{X})^2)$是$(\mu, \delta)$的充分统计量。

定义6. 若存在（非负可测）函数$h(x)$、$S(\theta)$，正整数k，及$C_j(\theta)$，$T_j(x)$ $(1 \le j \le k)$，使随机变量X的密度函数（或概率函数）可表示为

$$f(x, \theta) = S(\theta)h(x)exp\left\{\sum_{j=1}^{k} C_j(\theta)T_j(x)\right\}$$

则称其服从指数型分布（或称指数分布族）。

● 它包括许多常用分布，如两点分布、二项分布、指数分布、正态分布、Poisson分布等。可同时研究。k一般不大。

第二节 估计的优良性标准

●因为相乘即指数相加，所以

$$L(\theta) = S(\theta)^n \cdot \prod_{i=1}^{n} h(x_i) \cdot exp\left\{\sum_{j=1}^{k} C_j(\theta) \sum_{i=1}^{n} T_j(x_i)\right\}$$

故而$\left(\sum_{i=1}^{n} T_1(x_i), \cdots, \sum_{i=1}^{n} T_k(x_i)\right)$是$\theta$的一个充分统计量。

4. 均匀分布（不属于指数分布族）

$$L(a,\ b) = \left(\frac{1}{b-a}\right)^n I_{[a,\ \infty)}\left(\min_{1\le i\le n} x_i\right) I_{(-\infty,\ b]}\left(\max_{1\le i\le n} x_i\right)$$

故$\left(\min_{1\le i\le n} x_i, \max_{1\le i\le n} x_i\right)$是$(a,\ b)$的充分统计量。

定义7. 称统计量$U = \varphi(X_1, \cdots, X_n)$是$\theta$的完全（complete）统计量，若任给（可测）函数$u(\cdot)$，如果

$$E_\theta\{u(\varphi(X_1, \cdots, X_n))\} = 0 \qquad (\forall \theta \in \Theta)$$

则有$P_\theta\big(u\big(\varphi(X_1, \cdots, X_n)\big) = 0\big) = 1$，对$\forall \theta \in \Theta$。

第二节 估计的优良性标准

●两点分布中，$\varphi(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i$是完全的：

对任意$u(\cdot)$，因为$\varphi(X_1, \cdots, X_n) \backsim B(n, p)$，如果

$$E_p u(\varphi) = 0 \qquad \forall p \in (0, 1)$$

则$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} u(k) = 0$，对$\forall p \in (0, 1)$。

记$y = {}^{p}/_{(1-p)}$，则$\sum_{k=0}^{n} u(k) C_n^k y^k = 0$，对$\forall y \in (0, \infty)$，

故有$u(k) = 0$，（$k = 0, \cdots, n$），即φ是完全的。

●两点分布中，$\tilde{\varphi}(X_1, \cdots, X_n) = (\varphi_1, \varphi_2) = (\sum_{i=1}^{k} X_i, \sum_{i=k+1}^{n} X_i)$不是完全的（$\forall 1 \le k < n$）：

令$u(\tilde{\varphi}) = u(\varphi_1, \varphi_2) = \frac{\varphi_1}{k} - \frac{\varphi_2}{n-k}$，则$E_p u(\tilde{\varphi}) = 0$，但$u(\tilde{\varphi}) \neq 0$。

●两点分布中，$\varphi(X_1, \cdots, X_n) = (X_1, \cdots, X_n)$是充分的，不是完全的。

第二节 估计的优良性标准

●数据压缩（降维）或数据预处理的优良性标准：

① 该保留的信息都保留—充分性；
② 该丢掉的信息都丢掉—完全性。

两点分布中，$(X_1, \cdots, X_n)$、$(\sum_{i=1}^{k} X_i, \sum_{i=k+1}^{n} X_i)$、$\sum_{i=1}^{n} X_i$都是充分的，但仅$\sum_{i=1}^{n} X_i$是完全的。

●可以证明，若参数空间Θ有内点，则指数分布族中的充分统计量$(\sum_{i=1}^{n} T_1(x_i), \cdots, \sum_{i=1}^{n} T_k(x_i))$是完全的。

定理（Blackwell-Lehmann-Scheffe）若$\varphi(X_1, \cdots, X_n)$是$\theta$的完全的充分统计量，$\psi(\varphi(X_1, \cdots, X_n))$是$g(\theta)$的无偏估计，则它就是$g(\theta)$的（一致）最小方差无偏估计，且在概率为1相等的意义下，是唯一的。

●定理非常重要。

●定理说明，最小方差无偏估计仅需通过完全的充分统计量构造。

第二节 估计的优良性标准

对于正态分布，（$\bar{X}$，$\sum_{i=1}^{n}(X_i-\bar{X})^2$）是完全的充分统计量，故而，

因为$E\bar{X}=\mu$，所以$\bar{X}$就是μ的最小方差无偏估计；

令$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$（称为样本方差），则

$$E\left(S^2\right)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2-n\bar{X}^2\right)\right]$$
$$=\frac{1}{n-1}\sum_{i=1}^{n}E\left(X_i^2\right)-\frac{n}{n-1}E\left(\bar{X}^2\right)$$
$$=\frac{n}{n-1}\left(\sigma^2+\mu^2\right)-\frac{n}{n-1}\left(Var(\bar{X})+(E\bar{X})^2\right)$$
$$=\frac{n}{n-1}\sigma^2+\frac{n}{n-1}\mu^2-\frac{n}{n-1}\left(\frac{\sigma^2}{n}+\mu^2\right)$$
$$=\sigma^2$$

故S^2是σ^2的最小方差无偏估计。注意其MLE和矩估计均为$\frac{n-1}{n}S^2$。

第二节 估计的优良性标准

● 无偏估计有时不存在：例如，两点分布中，因为对任意的估计ψ，

$$E_p\left(\psi(X_1, \cdots, X_n)\right) = \sum_{x_i=0,1} \psi(x_1, \cdots, x_n)\, p^{x_i}(1-p)^{1-x_i}$$

是p的（最高n次）多项式，故而当$g(p)$为有理函数时，不存在其无偏估计，故也不存在其最小方差无偏估计。见P27例2.5（其中有具体的$g(p) = \frac{1}{1+p^2}$）。

●无偏估计有时虽然存在，但不合理（例如P27例2.6）：设X服从Poisson分布，参数为θ，样本量为1，而$g(\theta) = e^{-2\theta}$。可以证明，$g(\theta)$的唯一的无偏估计为

$$\hat{g}(X) = (-1)^X$$

即当X为偶数时，用1估计$g(\theta)$；而当X为奇数时，用-1估计$g(\theta)$。但我们已知$g(\theta) = e^{-2\theta} > 0$，所以此例中，坚持无偏性不合理。

第二节 估计的优良性标准

● 从一般的统计应用中看，上页的有时并不特别常见（上面两个例子显然都是人造的）。许多情况下，无偏性是较自然、合理的要求。（数学思维 vs 统计思维）

● 结合大样本性质，还有“渐近无偏”的概念。

●最小方差无偏估计的方差有多小？下面是一个非常重要的定理。

第二节 估计的优良性标准

定理（Cramer-Rao不等式）：设X的密度函数为$f(x,\theta)$，$\theta\in(a,b)$，$(-\infty\le a<b\le\infty)$，$(X_1,\cdots,X_n)$是$X$的简单随机样本，$\varphi(X_1,\cdots,X_n)$是$g(\theta)$的无偏估计，则在下列条件下，

(1) X的支撑$E=:\{x:f(x,\theta)>0\}$与θ无关；

(2) $g'(\theta)$与$\frac{\partial f(x,\theta)}{\partial\theta}$都存在，且

$$\int_E\frac{\partial f(x,\theta)}{\partial\theta}dx=\frac{\partial}{\partial\theta}\left(\int_E f(x,\theta)dx\right)=0$$

$$\frac{\partial}{\partial\theta}\int_E\cdots\int_E\varphi(x_1,\cdots,x_n)\prod_{i=1}^n f(x_i,\theta)dx_1\cdots dx_n$$

$$=\int_E\cdots\int_E\varphi(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^n f(x_i,\theta)\,dx_1\cdots dx_n$$

有不等式

$$var_\theta(\varphi(X_1,\cdots,X_n))\ge\frac{\left(g'(\theta)\right)^2}{nI(\theta)}\qquad\forall\theta\in\Theta$$

其中$I(\theta)=:\int_E\left(\frac{\partial logf(x,\theta)}{\partial\theta}\right)^2f(x,\theta)dx=E_\theta\left(\frac{\partial logf(X,\theta)}{\partial\theta}\right)^2>0$称为$\theta$的Fisher信息量。

第二节 估计的优良性标准

证明：不妨设$var_\theta\left(\varphi(X_1, \cdots, X_n)\right) < \infty$，且$I(\theta) < \infty$。因为

$$\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i, \theta) = \sum_{j=1}^{n}\left[\prod_{i\neq j} f(x_i, \theta)\frac{\partial f(x_j, \theta)}{\partial\theta}\right]$$

$$= \sum_{j=1}^{n}\left[\prod_{i=1}^{n} f(x_i, \theta)\frac{\frac{\partial f(x_j, \theta)}{\partial\theta}}{f(x_j, \theta)}\right]$$

$$= \prod_{i=1}^{n} f(x_i, \theta)\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[logf(x_i, \theta)]$$

而

$$g(\theta) = \int_E \cdots \int_E \varphi(x_1, \cdots, x_n)\prod_{i=1}^{n} f(x_i, \theta)\, dx_1 \cdots dx_n$$

（因是无偏估计）

第二节 估计的优良性标准

所以

$$g'(\theta)=\int_E\cdots\int_E\varphi(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i,\theta)\,dx_1\cdots dx_n$$

$$=\int_E\cdots\int_E\varphi(x_1,\cdots,x_n)\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[logf(x_i,\theta)]\prod_{i=1}^{n}f(x_i,\theta)\,dx_1\cdots dx_n$$

$$=E_\theta\left[\varphi(X_1,\cdots,X_n)\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[logf(X_i,\theta)]\right]$$

又因为

$$E_\theta\left\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[logf(X_i,\theta)]\right\}=nE_\theta\left[\frac{\partial}{\partial\theta}[logf(X_1,\theta)]\right]$$

$$=n\int_E\frac{\partial f(x,\theta)}{\partial\theta}dx=0$$

所以$g'(\theta)=E_\theta\left[(\varphi(X_1,\cdots,X_n)-g(\theta))\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[logf(X_i,\theta)]\right]$

第二节 估计的优良性标准

故

$$[g'(\theta)]^2 \le E_\theta\left[\left(\varphi(X_1,\ \cdots,\ X_n)-g(\theta)\right)\right]^2 E_\theta\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[logf(X_i,\ \theta)]\right]^2$$

$$= var_\theta(\varphi)\ \ var_\theta\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[logf(X_i,\ \theta)]\right)$$

$$= var_\theta(\varphi)\, n\, var_\theta\left(\frac{\partial}{\partial\theta}[logf(X_1,\ \theta)]\right) = var_\theta(\varphi)\, n\, I(\theta)$$

证毕。

● 由C-R不等式，参数无偏估计的方差有下界，一般情况下最好的阶为 $1/n$（中心极限定理的阶为 $1/\sqrt{n}$）。

● 如果$g(\theta)$的某无偏估计φ的方差达到了C-R不等式给出的下界，则φ就是$g(\theta)$的最小方差无偏估计（第二种判定方法）。

第二节 估计的优良性标准

例1. 两点分布，$X \sim B(1, p)$，

$$f(x, p) = p^x (1-p)^{1-x}$$

$$\frac{\partial(logf(x, p))}{\partial p} = \frac{x}{p} - \frac{1-x}{1-p}$$

$$I(p) = E_p\left(\frac{X}{p} - \frac{1-X}{1-p}\right)^2$$

$$= \left(\frac{0}{p} - \frac{1}{1-p}\right)^2 (1-p) + \left(\frac{1}{p} - \frac{0}{1-p}\right)^2 p = \frac{1}{p(1-p)}$$

- 当$p = {}^1/_2$时，Fisher信息量最小。
- $\overline{X}$的方差为$var_p(\overline{X}) = \frac{1}{n} p(1-p) = \frac{1}{nI(\theta)}$，达到了C-R不等式的下界。

第二节 估计的优良性标准

例2. 正态分布，$X \sim N(\mu,\ \sigma_0^2)$，其中$\sigma_0^2$已知。

则

$$f(x,\ \mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2}$$

$$\frac{\partial(\log f(x,\ \mu))}{\partial\mu} = \frac{x-\mu}{\sigma_0^2}$$

$$I(\mu) = \frac{E_\mu(X-\mu)^2}{\sigma_0^4} = \sigma_0^{-2}$$

● σ_0^2越大，Fisher信息量越小。

● $\overline{X}$的方差为${\sigma_0^2}/{n} = {1}/{nI(\mu)}$，达到了C-R不等式的下界，是$\mu$的最小方差无偏估计。

第二节 估计的优良性标准

例3. （P30例2.9）（其中计算提示：$\sum_{i=1}^{n} X_i$服从参数为$n\lambda$的Poisson分布）

说明有时最小方差无偏估计也不可能达到C-R不等式的下界！（明显是人造的例子）

例4. 若$X \sim U(0,\theta)$，则θ的MLE为$\max\limits_{1\le i\le n} X_i$，矩估计为$2\bar{X}$，均有不足。

令$\hat{\theta}=\frac{n+1}{n}\max\limits_{1\le i\le n} X_i$，则$\hat{\theta}$为$\theta$的无偏估计（实际上是最小方差无偏估计）。容易验证，$Var_\theta(\hat{\theta})=\frac{\theta^2}{n(n+2)}$，以$O(\frac{1}{n^2})$的速度趋向于0，快于C-R不等式中的$O(\frac{1}{n})$的阶。

原因：X的支撑与参数θ有关。

第二节 估计的优良性标准

● 请查找$I(\theta)$的等价定义。

● 若参数θ是m维的，也有相应的C-R不等式，可用于一般的正态分布。此时$I(\theta)$变为$m\times m$正定矩阵，称为Fisher信息矩阵，$g'(\theta)$变为列向量（梯度），$\frac{(g'(\theta))^2}{nI(\theta)}$变为二次型$\frac{1}{n}g'(\theta)^T I^{-1}(\theta)g'(\theta)$。请自行推导、补充。

● 大样本性质：当样本量n趋向∞时的性质，如（强）相合性，还有收敛速度等。常有

$$\sqrt{n}\left(\widehat{\theta}_n-\theta\right)\to N\left(0,\ \Sigma(\theta)\right)。$$

渐近协差阵$\Sigma(\theta)$与$I(\theta)$的关系?

● 小样本性质：无偏性、最小方差、C-R不等式等，不要求n趋向∞。

● 是否可以互推？相合估计⇒无偏？无偏估计⇒相合？证明或举反例。

第二节 估计的优良性标准

★ 统计方法的优良性标准？如高考录取：

状元；

前10（20）名录取数；

最低录取分数；

平均录取分数…。

是否应一致？

第三节 置信区间

● 点估计：用$\varphi(X_1, \cdots, X_n)$估计$g(\theta)$。一般一个完整的估计方法还要求给出或估计估计量的标准差（或方差），之后利用中心极限定理……

● 区间估计：用随机区间$[\varphi_1(X_1, \cdots, X_n), \varphi_2(X_1, \cdots, X_n)]$估计$g(\theta)$。

定义1. 设$\gamma \in (0, 1)$，$\varphi_1(X_1, \cdots, X_n) \le \varphi_2(X_1, \cdots, X_n)$是两个统计量，称$[\varphi_1(X_1, \cdots, X_n), \varphi_2(X_1, \cdots, X_n)]$是$g(\theta)$的置信水平为$\gamma$的置信区间，若对任意的$\theta \in \Theta$，有

$$P_\theta\{\varphi_1(X_1, \cdots, X_n) \le g(\theta) \le \varphi_2(X_1, \cdots, X_n)\} \ge \gamma$$

● 一般地，φ_1和φ_2皆为有限。

● 若刻意取$\varphi_1 = -\infty$或0，则称φ_2为置信上限。如污染浓度等。

● 若刻意取$\varphi_2 = +\infty$，则称φ_1为置信下限。如寿命、航天等。

第三节 置信区间

● 按照定义1，$(-\infty, +\infty)$永远是$g(\theta)$的置信水平为γ的置信区间！

● 优良性标准（非正式定义），一般地，用同一组数据，构造的置信区间越窄越好；置信上限越小越好；置信下限越大越好。

● 一般地，置信水平越大，置信区间越宽。

正态分布参数置信区间的几种基本类型：

一、已知方差$\sigma^2 = \sigma_0^2$，求期望μ的置信区间。

二、未知方差σ^2，求期望μ的置信区间。

三、未知期望μ，求方差σ^2的置信区间。

第三节 置信区间

一、已知方差$\sigma^2=\sigma_0^2$，求期望μ的置信区间

相对简单：因为$X_i \sim N(\mu,\ \sigma_0^2)$，故而$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \sim N(\mu,\ \frac{\sigma_0^2}{n})$，所以

$$\eta =: \frac{\overline{X}-\mu}{\sqrt{\sigma_0^2/n}} \sim N(0,\ 1)$$

若取$\gamma=0.95$，则$P(|\eta|\le 1.96)=\gamma$，因此，解不等式$|\eta|\le 1.96$，得μ的置信水平为95%的置信区间为

$$\left[\overline{X}-1.96\sqrt{\sigma_0^2/n},\overline{X}+1.96\sqrt{\sigma_0^2/n}\right]$$

● 习惯上，γ常取0.90，0.95或0.99，此时对应的常数为1.65，1.96或2.58。

● 给定$\gamma=0.95$，是否可以有别的水平恰为γ的置信区间?

第三节 置信区间

例1. 设滚珠直径X服从正态分布$N(\mu, 0.05)$，随机抽取6个，测得其直径为：14.70，15.21，14.90，14.91，15.32，15.32，试求μ的95%的置信区间。

解：$\overline{x}=\frac{1}{6}\sum_{i=1}^{6}x_i=15.06$，$1.96\sqrt{0.05/6}\approx 0.18$，故所求为

$[15.06-0.18, 15.06+0.18]=[14.88, 15.24]$。

● 能否说，μ以95%的概率落入$[14.88, 15.24]$?

● μ是非随机的，$[14.88, 15.24]$是已知的，无随机性，无概率问题。

● 抽样前，$\left[\overline{X}-1.96\sqrt{\sigma_0^2/n}, \overline{X}+1.96\sqrt{\sigma_0^2/n}\right]$是随机区间，以概率$\gamma=0.95$覆盖$\mu$!

第三节 置信区间

例1. 设滚珠直径X服从正态分布$N(\mu, 0.05)$，随机抽取6个，测得其直径为：14.70，15.21，14.90，14.91，15.32，15.32，试求μ的95%的置信区间。

解：$\overline{x}=\frac{1}{6}\sum_{i=1}^{6}x_i=15.06$，$1.96\sqrt{0.05/6}\approx 0.18$，故所求为

$[15.06-0.18, 15.06+0.18]=[14.88, 15.24]$。

- 能否说，μ以95%的概率落入$[14.88, 15.24]$?

- μ是非随机的，$[14.88, 15.24]$是已知的，无随机性，无概率问题。

- 抽样前，$\left[\overline{X}-1.96\sqrt{\sigma_0^2/n}, \overline{X}+1.96\sqrt{\sigma_0^2/n}\right]$是随机区间，以概率$\gamma=0.95$覆盖$\mu$!

第三节 置信区间

二、未知方差σ^2，求期望μ的置信区间。

因为σ^2未知，“一”的方法无法使用，需要用其无偏估计S^2代替它。但如此一来，得到的$T=\frac{\bar{X}-\mu}{\sqrt{S^2/n}}$不再服从正态分布，1.96也不能用了。

下面主要研究T的分布。

定理3.1. 设$X_1,\cdots,X_n$独立同分布，共同分布为$N(0,1)$，则$\xi=:\sum_{i=1}^n X_i^2$服从Γ分布$\Gamma(\frac{n}{2},\frac{1}{2})$。

证明：（直接证明，如数学归纳法，略）

定义2： 若$\xi\sim\Gamma(\frac{n}{2},\frac{1}{2})$，则称其服从$n$个自由度的$\chi^2$分布，记为$\xi\sim\chi^2(n)$。

其密度函数（不重要），草图（重要），为……。

第三节 置信区间

● 自由度是统计中非常重要的概念，来源于物理/力学，如火车有1个自由度，汽车有2个，R^n中的任一点有n个自由度。（$R^1, R^2, R^3, \cdots$）

● 称ξ的自由度为n，是因为按照其统计背景，来源于R^n中随机向量$(X_1, \cdots, X_n)$。但ξ自身是一维的。

● χ^2分布是统计中3个重要分布之一，主要描述方差整体分布。

推论1. 若$\xi \sim \chi^2(n)$，则$E\xi = n$。

证明：略。

推论2. 若$\xi \sim \chi^2(m)$，$\eta \sim \chi^2(n)$，ξ与η相互独立，则$\xi + \eta \sim \chi^2(m+n)$。

证明：取$X_1, \cdots, X_{m+n}$独立同分布，共同分布为$N(0, 1)$，令⋯ ⋯

第三节 置信区间

定理3.2. 设$X_1, \cdots, X_n$相互独立，$X_i \sim N(\mu_i, \sigma^2)$，$A=(a_{ij})$为$n$阶正交矩阵，令$(Y_1, \cdots, Y_n)^T = A(X_1, \cdots, X_n)^T$，则$Y_1, \cdots, Y_n$相互独立，且$Y_i \sim N(\sum_{k=1}^n a_{ik}\mu_k，\sigma^2)$。

证明：（1）首先考虑$\mu_i = 0\ (i=1, \cdots, n)$的情形。此时$(X_1, \cdots, X_n)^T$的联合密度为

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\}$$

对给定的实数$t_1, \cdots, t_n$，令

$$D=\left\{(x_1, \cdots, x_n):\ \sum_{j=1}^n a_{ij}x_j \le t_i,\ i=1, \cdots, n\right\}$$

则$P(Y_1 \le t_1, \cdots, Y_n \le t_n) = \int_D \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\} dx_1 \cdots dx_n$，做变量替换$(y_1, \cdots, y_n)^T = A(x_1, \cdots, x_n)^T$，则$(x_1, \cdots, x_n)^T = A^T(y_1, \cdots, y_n)^T$，且$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$。

因为A是正交矩阵，故其雅可比行列式绝对值为1。故

$$
\begin{aligned}
&P(Y_1 \le t_1, \cdots, Y_n \le t_n) \\
&= \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2\right\} dy_1 \cdots dy_n \\
&= \prod_{i=1}^{n} \left[\int_{-\infty}^{t_i} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{y_i^2}{2\sigma^2}\right\} dy_i\right]
\end{aligned}
$$

由此可知，$Y_1, \cdots, Y_n$相互独立，皆服从$N(0, \sigma^2)$。

（2）一般情形（μ_i不全为0）。令$Z_i = X_i - \mu_i$，则$Z_1, \cdots, Z_n$满足（1）的条件，故$\sum_{j=1}^{n} a_{1j} Z_j, \cdots, \sum_{j=1}^{n} a_{nj} Z_j$ 相互独立，共同分布为$N(0, \sigma^2)$。

而

$$
Y_i = \sum_{j=1}^{n} a_{ij} Z_j + \sum_{j=1}^{n} a_{ij}\mu_j \qquad (i = 1, \cdots, n)
$$

所以$Y_1, \cdots, Y_n$相互独立，且$Y_i \sim N(\sum_{k=1}^{n} a_{ik}\mu_k, \sigma^2)$。证毕。

第三节 置信区间

定理3.3. 设$X_1, \cdots, X_n$相互独立，$X_i \sim N(\mu, \sigma^2)$，则

（1）$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$；

（2）$\frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \chi^2(n-1)$；

（3）$\bar{X}$与$\sum_{i=1}^{n} (X_i - \bar{X})^2 = (X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2$相互独立。

证明：此时$\mu_i \equiv \mu$。取n阶正交阵A，其第一行元素皆为$1/\sqrt{n}$，令$Y_i = \sum_{j=1}^{n} a_{ij} X_j$，由定理3.2，$Y_1, \cdots, Y_n$相互独立，$Y_i \sim N(\sum_{k=1}^{n} a_{ik}\mu, \sigma^2)$。故$Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$。

当$i \geq 2$时，因为A正交，故$\sum_{k=1}^{n} a_{ik} = 0$，从而$Y_i \sim N(0, \sigma^2)$。因此，

$$\bar{X} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}} = \frac{1}{\sqrt{n}} Y_1 \sim N(\mu, \frac{\sigma^2}{n})$$

又因为$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n(\bar{X})^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2$，

第三节 置信区间

所以

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=2}^{n}\left(\frac{Y_i}{\sigma}\right)^2\sim\chi^2(n-1)$$

最后，由于$Y_1, \cdots, Y_n$相互独立，故$\bar{X}=\frac{1}{\sqrt{n}}Y_1$与$\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=2}^{n}Y_i^2$相互独立。

● 此定理非常重要：如何利用正交变换将S^2中的n项相关的随机变量平方和变为相互独立的$n-1$项平方和。

● 关于自由度：虽然$(X_1-\bar{X}, \cdots, X_n-\bar{X})^T\in R^n$，但因为$\sum_{i=1}^{n}(X_i-\bar{X})\equiv 0$，实际上分布它在$R^n$中的某$n-1$维子空间中，故仅有$n-1$个自由度。（或，$\bar{X}$占了一个自由度）

● 此处自由度的统计学解释为：独立信息个数。

第三节 置信区间

现在研究$T=\frac{\bar{X}-\mu}{\sqrt{S^2/n}}$，显然

$$T=\frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{\frac{1}{(n-1)\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2}}$$

可以看出，分子为标准正态$N(0,1)$，而分母中的$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2$服从$\chi^2(n-1)$，故分子、分母的分布皆为完全已知（与未知参数$\mu$、$\sigma^2$无关）！

且相互独立，故T的分布也是完全已知的，不依赖于未知参数μ、σ^2。

定义3. 称随机变量T服从n个自由度的t分布，记为$T\sim t(n)$，若T的分布密度为

$$p_n(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

第三节 置信区间

● 密度函数：不用背，但应知道其草图。

● 与标准正态密度区别：重尾（heavy-tailed）。

● 对$\forall t \in R$，有$\lim_{n\to\infty} p_n(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$。

定理3.4. 设ξ与η相互独立，$\xi \sim N(0,\ 1)$，$\eta \sim \chi^2(n)$，则

$$T = \frac{\xi}{\sqrt{\eta/n}} \sim t(n)$$

证明：略（随机变量函数的分布）。

● t分布也称学生分布（Student distribution，Gosset笔名Student。）是统计中3个重要分布之一。

第三节 置信区间

回到我们的问题，则有重要结论：

$$T=\frac{\bar{X}-\mu}{\sqrt{S^2/n}}\sim t(n-1)$$

● 利用上述结论，很快得到μ的置信区间构造方法：给定置信水平γ（如0.95），查t分布表（P354），自由度$df=n-1$，得临界值λ，满足

$$P(|T|\le\lambda)=\gamma$$

解不等式得，μ的置信水平为γ的置信区间为

$$\left[\bar{X}-\lambda\frac{S}{\sqrt{n}},\bar{X}+\lambda\frac{S}{\sqrt{n}}\right]$$

● 由于t分布是已知分布，对给定的n，及常用置信水平0.90、0.95和0.99，前人早已完成制表工作。

第三节 置信区间

例2. 用仪器间接测量高温，重复5次，得到数据1250，1265，1245，1260，1275。设测量值服从正态分布$N(\mu, \sigma^2)$，求μ的95%的置信区间。

解：$n=5$，$\bar{X}=1259$，$S^2=\frac{1}{n-1}\sum_{i=1}^{5}(X_i-\bar{X})^2=\frac{570}{4}$。查$t$分布表（$\gamma=0.95$，df=4），得$\lambda=2.776$，故置信区间为：$\left[\bar{X}-\lambda\frac{S}{\sqrt{n}},\bar{X}+\lambda\frac{S}{\sqrt{n}}\right]=[1244.2, 1273.8]$。

三、未知期望μ，求σ^2的置信区间。

由定理3.3的（2），$Y=:\frac{(n-1)S^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\sim\chi^2(n-1)$，对于给定的$\gamma$，查$\chi^2$分布表得$\lambda_1<\lambda_2$（画图 ），使得$P(Y<\lambda_1)=P(Y>\lambda_2)=\frac{1-\gamma}{2}$，则$P(\lambda_1\le Y\le\lambda_2)=\gamma$，即

第三节 置信区间

$$P\left(\frac{(n-1)S^2}{\lambda_2} \le \sigma^2 \le \frac{(n-1)S^2}{\lambda_1}\right) = \gamma$$

即

$$\left[\frac{1}{\lambda_2}\sum_{i=1}^{n}(X_i-\bar{X})^2, \frac{1}{\lambda_1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$$

是σ^2的置信水平为γ的置信区间。

例2（续）查χ^2分布表（$\gamma = 0.95$，df=4）得λ_1 =0.484，$\lambda_2 = 11.1$，故σ^2的95%的置信区间$[51.35, 1177.69]$。

如果取$\widetilde{\lambda_1} = 0$（类似于$-\infty$），则$\widetilde{\lambda_2} = 9.49$，故$\sigma^2$的95%的置信下限为60.06；同理，$\sigma^2$的95%的置信上限为$\frac{1}{\widetilde{\lambda_1}}\sum_{i=1}^{n}(X_i-\bar{X})^2$，本书$\widetilde{\lambda_1}$查不到。

● 该置信区间的优良性：未提及。

第三节 置信区间

- 枢轴量方法：找枢轴量（非统计量）满足
  - ① 其表达式中包含且仅包含我们感兴趣的参数；
  - ② 其分布已知，与参数无关。

对情况一，枢轴量为$\eta=\frac{\bar{X}-\mu}{\sqrt{\sigma_0^2/n}}$；

对情况二，枢轴量为$T=\frac{\bar{X}-\mu}{\sqrt{S^2/n}}$；

对情况三，枢轴量为$Y=\frac{(n-1)S^2}{\sigma^2}$。

通过解不等式，可以得到感兴趣参数的置信区间。

思考：若$\mu=\mu_0$已知，如何求σ^2的置信区间?

- 枢轴量方法不足：① 如何找？缺乏统一方法；② 优良性?

第三节 置信区间

● 统计量方法*（多用于可靠性中寻找置信下限等）

设$X_i \sim F(x,\theta)$，$g(\theta)$为Θ上的实值函数（目标），又设$\varphi(X_1,\cdots,X_n)$为任一实值统计量，令

$$G(u,\theta)=P_\theta(\varphi(X_1,\cdots,X_n)\ge u)=1-P_\theta(\varphi(X_1,\cdots,X_n)<u)$$

则易证，$G(u,\theta)$是u的左连续减函数（$1-G(u+0,\theta)$是分布函数）。给定置信水平$0<\gamma<1$，对固定样本$x_1,\cdots,x_n$，记$u=\varphi(x_1,\cdots,x_n)$，令

$$g_L(u)=inf\{g(\theta):\ \theta\in\Theta,\ G(u,\theta)\ge 1-\gamma\}$$

$$g_U(u)=sup\{g(\theta):\ \theta\in\Theta,\ G(u+0,\theta)\le \gamma\}$$

如同通常情况，当$\{\theta\in\Theta,\ G(u,\theta)\ge 1-\gamma\}$为空集时，定义$g_L(u)=sup\{g(\theta):\ \theta\in\Theta\}$；当$\{\theta\in\Theta,\ G(u+0,\theta)\le\gamma\}$为空集时，定义$g_U(u)=inf\{g(\theta):\ \theta\in\Theta\}$。

第三节 置信区间

定理3.5. $g_L(\varphi(X_1,\cdots,X_n))$是$g(\theta)$的水平为$\gamma$的置信下限，即

$$P_\theta\left(g(\theta)\geq g_L(\varphi(X_1,\cdots,X_n))\right)\geq\gamma\qquad\forall\theta\in\Theta$$

证明：固定$\theta\in\Theta$，令$c=sup\{u:G(u,\theta)\geq 1-\gamma\}$，则（画图）

$$G(c,\theta)\geq 1-\gamma\geq G(c+0,\theta)\qquad(1)$$

由g_L的定义可知，

$$g(\theta)<g_L(u)\ \Longrightarrow G(u,\theta)<1-\gamma\quad（g_L(u)\text{是下确界}）\qquad(2)$$

再由（1）式及G对u单调减知，

$$G(u,\theta)<1-\gamma\ \Longrightarrow\ u>c\qquad(3)$$

因此，

$$\begin{aligned}&P_\theta\left(g(\theta)<g_L(\varphi(X_1,\cdots,X_n))\right)\\ &\leq(\text{由}2)P_\theta(G(\varphi(X_1,\cdots,X_n),\theta)<1-\gamma)\\ &\leq(\text{由}3)P_\theta(\varphi(X_1,\cdots,X_n)>c)\\ &\leq G(c+0,\theta)\leq(\text{由}1)1-\gamma\end{aligned}$$

证毕。

证明不长，但看起来较难。一是不熟悉，二是“下确界”等。

第三节 置信区间

● 同理可知，$g_U(\varphi(X_1,\cdots,X_n))$是$g(\theta)$的水平为$\gamma$的置信上限，即

$$P_\theta\left(g(\theta)\le g_U\left(\varphi(X_1,\cdots,X_n)\right)\right)\ge\gamma\qquad\forall\theta\in\Theta$$

● $[g_L(\varphi(X_1,\cdots,X_n)),\ g_U(\varphi(X_1,\cdots,X_n))]$是$g(\theta)$的水平为$2\gamma-1$的置信区间。

例3. 设$X\sim N(\mu,\sigma^2)$，数据为$X_1,\cdots,X_n$，求σ^2的置信上限（例2（续）后已提及）。

解：取$\varphi(X_1,\cdots,X_n)=\sum_{i=1}^n(X_i-\bar{X})^2$，则

$$G(u,\sigma^2)=P\left(\sum_{i=1}^n(X_i-\bar{X})^2\ge u\right)$$

$$=P\left(\frac{1}{\sigma^2}\sum_{i=1}^n(X_i-\bar{X})^2\ge\frac{u}{\sigma^2}\right)=1-H(\frac{u}{\sigma^2})$$

其中$H(\cdot)$为$\chi^2(n-1)$的分布函数。

第三节 置信区间

易知G是σ^2的连续增函数，设$g_U(u)$满足方程$G(u, g_U(u)) = \gamma$，即

$$1 - H\left(\frac{u}{g_U(u)}\right) = \gamma$$

$H\left(\frac{u}{g_U(u)}\right) = 1 - \gamma$，查表（0.95查不到！）得临界值$\lambda$，则$g_U(u) = \frac{u}{\lambda}$，即

$g_U(X_1, \cdots, X_n) = \frac{1}{\lambda}\sum_{i=1}^{n}(X_i - \overline{X})^2$为$\sigma^2$的置信上限（与枢轴量方法相同）。

例4. 某型号火箭发射3次，皆成功，求成功率p的95%的置信下限。

--------------end 20240304

解：取统计量$\varphi(X_1, \cdots, X_3) = \sum_{i=1}^{3} X_i$，则$G(u, p) = P_p(\sum_{i=1}^{3} X_i \geq u)$，故

$$G(3, p) = P_p\left(\sum_{i=1}^{3} X_i \geq 3\right) = p^3,$$

$$G(2, p) = P_p\left(\sum_{i=1}^{3} X_i \geq 2\right) = 3p^2 - 2p^3, \cdots\cdots$$

易知G是σ^2的连续增函数，设$g_U(u)$满足方程$G(u, g_U(u)) = \gamma$，即

$$1 - H\left(\frac{u}{g_U(u)}\right) = \gamma$$

$H\left(\frac{u}{g_U(u)}\right) = 1 - \gamma$，查表（0.95查不到！）得临界值$\lambda$，则$g_U(u) = \frac{u}{\lambda}$，即

$g_U(X_1, \cdots, X_n) = \frac{1}{\lambda}\sum_{i=1}^{n}(X_i - \bar{X})^2$为$\sigma^2$的置信上限（与枢轴量方法相同）。

例4. 某型号火箭发射3次，皆成功，求成功率p的95%的置信下限。

解：取统计量$\varphi(X_1, \cdots, X_3) = \sum_{i=1}^{3} X_i$，则$G(u, p) = P_p(\sum_{i=1}^{3} X_i \geq u)$，故

$$G(3, p) = P_p\left(\sum_{i=1}^{3} X_i \geq 3\right) = p^3,$$

$$G(2, p) = P_p\left(\sum_{i=1}^{3} X_i \geq 2\right) = 3p^2 - 2p^3, \cdots\cdots$$

第三节 置信区间

现在数据为$\sum_{i=1}^{3} X_i = 3$，故由方程$G(3, p_L) = 1 - \gamma = p_L^3$，得到

$$p_L = \sqrt[3]{1-\gamma} = \sqrt[3]{0.05} \approx 0.368$$

可以证明，这是最好的置信下限!

- 统计量方法如何选统计量？可选充分统计量、或MLE。
- 参数空间多维时，可以有置信域（Confidence region）。
- 近现代统计中，难以构造精确置信区间，常用

$$[\hat{\theta} - 1.96\hat{s}, \hat{\theta} + 1.96\hat{s}]$$

其覆盖率约为95%。

第四节 分布函数与密度函数的估计

- 参数的（Parametric）：参数空间有限维；
- 非参数的（Non-parametric)：参数空间无穷维，如估计函数；
- 半参数的（Semi-parametric）：有限维部分+（或乘）无穷维部分。

一、分布函数$F(x)$的估计（不假定X的分布类型，非参数问题）

设样本为简单随机样本$X_1, \cdots, X_n$，对于任意固定的x，因为$F(x) = P(X \le x)$，用频率估计概率，得$F(x)$的无偏估计量$\frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x]}(X_i) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i, \infty)}(x)$。由强大数律，$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x]}(X_i) = F(x)\right) = 1$。

定义1. 称$F_n(x) =: \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x]}(X_i) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i, \infty)}(x)$为$X$的经验分布函数。

- $F_n(x)$为右连左极阶梯函数，是以$\frac{1}{n}$（若X_i不相等）概率取X_i的随机变量的分布函数。

第四节 分布函数与密度函数的估计

定理4.1.（Glivenko-Cantelli）

$$P\left(\lim_{n\to\infty} sup_{x\in(-\infty,\,\infty)} |F_n(x)-F(x)|=0\right)=1$$

证明：略。

定理说明，$F_n(x)$一致收敛于$F(x)$。

二、密度函数$f(x)$的估计

（一）、直方图法

对任意$x\in(c,\,d]$，若$f(x)$较大，则样本$X_1,\cdots,X_n$落入$(c,\ d]$的会较多。估计的思想将其反过来。由于

$$P(c<X\le d)=\int_c^d f(t)\,dt\approx f(x)(d-c)$$

第四节 分布函数与密度函数的估计

再用频率估计概率$P(c < X \le d)$，得到$f(x)$的估计量

$$\widehat{f(x)} = \frac{1}{n(d-c)}\sum_{i=1}^{n} I_{(c,d]}(X_i)$$

直方图法就是利用上述想法，但c、d的取法等有规范。

● 将样本值重新排序，得顺序统计量$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$。

● 取常数a略小于$X_{(1)}$，b略大于$X_{(n)}$，将$[a, b]$等分为m个区间$(t_i, t_{i+1}]$，其中$i = 0, \cdots, m-1$，$t_i = a + i \cdot \frac{b-a}{m}$，$t_0 = a$，$t_m = b$（$m$的选取依赖$n$，区间“左开右闭”是为了规范）。

● 记v_i为$X_1, \cdots, X_n$中落入$(t_i, t_{i+1}]$的频数，$h = \frac{b-a}{m}$。

● 在数轴上做以$[t_i, t_{i+1}]$为底，以$\frac{v_i}{nh}$为高的m个矩形，得直方图。

第四节 分布函数与密度函数的估计

- m 的选取依赖n，n越大，m相应也越大。
- 得到的估计是阶梯函数，一定条件下，是相合的。
- 优点：简单、方便、易于理解。

例1. 教材P52，例4.2……（有了直方图后，多年前的统计软件就可以在你输入不同分布类型后，画出最好的拟合曲线，直观上帮助你建模）

（二）、核估计法

希望对直方图方法进行改进，更精细一些，且估计量可以有较好的数学性质，如连续。

定义2. 设$K(x) \geq 0$，$\int_{-\infty}^{\infty} K(x)dx = 1$，称其为核函数，称

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

为$f(x)$的核估计，称$h > 0$为其窗宽。（加权平均，权重可随x随时变化）

第四节 分布函数与密度函数的估计

● 核估计为一类估计，其中的核函数$K(x)$、窗宽h由操作者选取，选的好，估计效果也好。

● 一般地， $K(x)$常常选为关于原点对称的（单峰）函数，如:

$$K_0(x) = \begin{cases} \frac{1}{2} & 当 -1 \le x \le 1 \\ 0 & 否则 \end{cases}$$

$$K_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$K_2(x) = \frac{1}{\pi(1+x^2)}$$

等。

第四节 分布函数与密度函数的估计

● 一般当数据离x越近时，$K\left(\frac{x-X_i}{h}\right)$越大，说明数据$X_i$对估计的贡献大。

● 窗宽h越小，说明越重视靠近x的数据，而轻视远离x的数据。一般地，若数据量大，窗宽可取的较小；数据量小，窗宽不得不取的较大。

● 样本固定时，窗宽h大，核估计平滑；窗宽h小，核估计波动大。

● 当取$K(x)=K_0(x)$时，

$$f_n(x)=\frac{1}{2hn}\sum_{i=1}^{n} I_{(x-h,\,x+h]}(X_i)=\frac{1}{2h}[F_n(x+h)-F_n(x-h)]$$

其中$F_n(\cdot)$为经验分布函数。由微积分基本定理，一定条件下，约等于$f(x)$。此时的$f_n(x)$称为Rosenblatt估计（最早的核估计）。

第四节 分布函数与密度函数的估计

- 一定条件下有强相合性。

- 当$K(x)$与h选的好时，核估计比直方图估计好。

- 有寻找最优窗宽的相关成熟方法。

- 容易验证，$\int_{-\infty}^{\infty} f_n(x)\,dx \equiv 1$。

- $f_n(x)$是非参数的，其收敛速度达不到一般参数估计的$O(n^{-\frac{1}{2}})$收敛速度。