第二章　参数估计

一.分类

(一)非参数估计问题：总体的分布类型未知，需要由样本构造统计量去估计总体的分布函数或密度函数。

(二)参数估计问题：总体分布类型已知，其中含有未知参数，需由样本构造统计量来估计未知参数。

例1.学生的成绩$X \sim N(\mu, \sigma^2)$, 但参数μ, σ^2未知，需通过样本来估计。

例2.已知某城市在单位时间内发生交通事故的次数$X \sim P(\lambda)$, 但λ未知，需通过样本来估计。

二.对于参数估计，按问题的性质不同分为两类

1.点估计：适当的选择一个统计量$\hat{\theta}(X_1,\cdots,X_n)$作为未知参数$\theta$的估计.

2.区间估计：求未知参数θ的范围，即求一个区间使以较大的概率包含未知参数.

第一节 矩估计和极大似然估计

第一节 矩估计和极大似然估计

一.矩估计： 用样本k阶矩作为总体k阶矩的估计量，建立含有待估参数的方程，从而解出待估参数.

(一)矩估计的思想：实质是由辛钦大数定律，设$\xi_1, \xi_2, ..., \xi_n, ...$是相互独立同分布的随机变量，且$E\xi_i^k, i = 1, 2, ...$存在，则

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n}\xi_i^k - E\xi_i^k| \geq \epsilon) = 0.$$

所以可用样本矩代替总体矩，求得未知参数θ，即替换原则。

第一节　矩估计和极大似然估计

(二)矩估计的方法：设$X \sim F(x;\theta), \theta = (\theta_1, \cdots, \theta_k)$是未知参数向量,若$F(x;\theta)$的$k$阶矩存在,

$$\alpha_\nu(\theta) = \int_{-\infty}^{\infty} x^\nu \, dF(x;\theta), \qquad 1 \le \nu \le k$$

是$\theta = (\theta_1, \cdots, \theta_k)$的函数.设$(X_1, \cdots, X_n)$是取自总体$X$的一个样本,则由下列方程组

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n X_i = \alpha_1(\theta_1, \cdots, \theta_k); \\ \frac{1}{n}\sum_{i=1}^n X_i^2 = \alpha_2(\theta_1, \cdots, \theta_k); \\ \qquad \vdots \\ \frac{1}{n}\sum_{i=1}^n X_i^k = \alpha_k(\theta_1, \cdots, \theta_k), \end{cases}$$

得到$\theta = (\theta_1, \cdots, \theta_k)$的一组解$\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$, 其中$\hat{\theta}_\nu = \hat{\theta}_\nu(X_1, \cdots, X_n)$, 并以$\hat{\theta}_\nu$ 作为参数θ_ν的估计量, $\nu = 1, \cdots, k$, 则称$\hat{\theta}_\nu$为未知参数θ_ν的**矩估计量**.

第一节 矩估计和极大似然估计

例1:求总体X的均值$EX=\mu$和方差$DX=\sigma^2$的矩估计.
解：$(X_1,\cdots,X_n)$是取自总体X的一个样本,若总体的二阶矩α_2存在,则有$\alpha_2=\sigma^2+\mu^2$,

$$\begin{cases} \overline{X}=\frac{1}{n}\sum_{i=1}^n X_i=\mu; \\ \frac{1}{n}\sum_{i=1}^n X_i^2=\alpha_2=\mu^2+\sigma^2. \end{cases}$$

以此方程组的解作为μ,σ^2的估计

$$\begin{aligned} \hat{\mu} &= \overline{X}; \\ \hat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^n X_i^2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2=\tilde{S}^2. \end{aligned}$$

所以,总体均值μ和方差σ^2的矩估计,分别是样本均值$\overline{X}$和样本二阶中心矩$\tilde{S}^2$. 这个结论对任何总体都成立.

第一节 矩估计和极大似然估计

例2: 设X的密度函数为$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$

解: $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=EX=\frac{1}{\lambda}$, 所以$\hat{\lambda}=\frac{1}{\overline{X}}$为$\lambda$的矩估计.

第一节 矩估计和极大似然估计

例3: 设$X \sim U(\theta_1, \theta_2)$,其概率密度函数为

$$f(x;\theta_1,\theta_2) = \begin{cases} \frac{1}{\theta_2-\theta_1}, \theta_1 < x < \theta_2; \\ 0, \text{其它}, \end{cases}$$

其中$\theta_1 < \theta_2$,求$\theta = (\theta_1, \theta_2)$的矩估计.

解: 由均匀分布$U(\theta_1, \theta_2)$的性质,有$EX = \frac{\theta_1+\theta_2}{2}$, $DX = \frac{(\theta_2-\theta_1)^2}{12}$.
因此可得方程组

$$\begin{cases} \overline{X} = \frac{\theta_1+\theta_2}{2}; \\ \tilde{S}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2 = \frac{(\theta_2-\theta_1)^2}{12}. \end{cases}$$

以此方程组的解作为θ_1, θ_2的估计,得

$$\begin{cases} \hat{\theta}_1 = \overline{X} - \sqrt{3}\tilde{S}; \\ \hat{\theta}_2 = \overline{X} + \sqrt{3}\tilde{S}, \end{cases}$$

此即所求的θ_1, θ_2的矩估计.

第一节 矩估计和极大似然估计

例4:设$X \sim \Gamma(\alpha,\beta), \alpha>0, \beta>0$, 即密度函数为

$$f(x;\alpha,\beta)=\begin{cases}\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\beta x}, & x>0\\ 0, & x\le 0\end{cases}$$

求α,β的矩估计.

解: 由Γ分布的性质, $EX=\frac{\alpha}{\beta}, DX=\frac{\alpha}{\beta^2}$, 因此可得方程组

$$\begin{cases}\overline{X}=\frac{\alpha}{\beta};\\ \tilde{S}^2=\frac{\alpha}{\beta^2}.\end{cases}$$

得

$$\begin{cases}\hat{\alpha}=\frac{\overline{X}^2}{\tilde{S}^2};\\ \hat{\beta}=\frac{\overline{X}}{\tilde{S}^2}.\end{cases}$$

第一节　矩估计和极大似然估计

注：(1)矩估计直观简便

(2)要求总体的原点矩存在，若不存在则不能用，如柯西分布

(3)没有充分利用$F(x;\theta)$对θ所提供的信息

第一节 矩估计和极大似然估计

二.极大似然估计

(一)原理：极大似然估计是建立在极大似然原理基础之上。

引例. 两外形相同的箱子，各装100个球，其中第一箱有99个白球和1个红球，第二箱有1个白球和99个红球，现从两箱中任取一箱，并从箱中任取一球，所取球为白球，问所取球来自哪一箱。

例1: 设总体$X \sim b(1,p)$, $P\{X=1\}=p$, 采用上述思路(极大似然估计的思想)求p的估计值.

注：概率最大的事件最可能发生

第一节 矩估计和极大似然估计

(二)定义: 设$(X_1,\cdots,X_n)$为取自具有概率分布族$\{f(x;\theta):\theta\in\Theta\}$ 的离散型总体X的一个样本,其中$\theta=(\theta_1,\cdots,\theta_k)$是未知的$k$维参数向量, $(X_1,\cdots,X_n)$ 取观测值$(x_1,\cdots,x_n)$ 的概率为

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

$L(x_1,\cdots,x_n;\theta)$称为θ的**似然函数**(Likelihood Function).若$\hat{\theta}$使

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

并以$\hat{\theta}$作为参数θ的估计值,则称$\hat{\theta}$为θ的极大似然估计值,其相应的统计量$\hat{\theta}(X_1,\cdots,X_n)$ 称为θ的**极大似然估计量**.简记为MLE或ML估计.

第一节　矩估计和极大似然估计

(三)求极大似然估计的方法

(1)先求似然函数$L(x_1,\cdots,x_n;\theta)\triangleq\prod_{i=1}^{n}f(x_i;\theta)\triangleq L(\theta)$

(2)因为$L(\theta)$与$\ln L(\theta)$有相同的极大值点，所以$\ln L(\hat{\theta})=\sup_{\theta\in\Theta}\ln L(\theta)$,若L 可微, $\ln L(\theta)$关于$\theta_1,\cdots,\theta_k$分别求导数，并令其等于0，得$\frac{\partial\ln L(\theta)}{\partial\theta_i}=0, i=1,\cdots,k$，称为似然方程组; 求解方程组，得$\hat{\theta}$并证明$\hat{\theta}$使$L(\theta)$达到最大, $\hat{\theta}$即为θ的极大似然估计;
若L不可微, 则用其他方法求出极大似然估计值.

第一节 矩估计和极大似然估计

例2: 设X的密度函数为$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$,求参数$\lambda$的极大似然估计。

解：设$(x_1,\cdots,x_n)$是样本$(X_1,\cdots,X_n)$的一组观测值,似然函数为

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

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

令$\frac{\partial \ln L}{\partial\lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i=0$

经验证,$\hat{\lambda}=\frac{1}{\bar{X}}$是$\lambda$的极大似然估计。

第一节 矩估计和极大似然估计

例2: 设X的密度函数为$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$,求参数$\lambda$的极大似然估计。

解: 设$(x_1,\cdots,x_n)$是样本$(X_1,\cdots,X_n)$的一组观测值,似然函数为

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

则$\ln L(x_1,\cdots,x_n;\lambda)=n\ln\lambda-\lambda\sum_{i=1}^n x_i$
令$\frac{\partial \ln L}{\partial\lambda}=\frac{n}{\lambda}-\sum_{i=1}^n x_i=0$
经验证, $\hat{\lambda}=\frac{1}{\bar{X}}$是$\lambda$的极大似然估计。

第一节 矩估计和极大似然估计

例3:设总体X具有均匀分布,密度函数为

$$f(x,\theta)=\begin{cases}\frac{1}{\theta}, & 0<x<\theta;\\ 0, & \text{其它},\end{cases}$$

其中$\theta>0$是未知参数,求θ的极大似然估计.

解:设$(x_1,\cdots,x_n)$是样本$(X_1,\cdots,X_n)$的一组观测值,似然函数为

$$L(\theta)=\begin{cases}\frac{1}{\theta^n}, & 0<x_i<\theta;\\ 0, & \text{其它},\end{cases}$$

要使L最大,必须使θ最小,当$\theta=\max_{1\le i\le n}x_i=x_{(n)}$时,可使$L$最大,故$\hat{\theta}=X_{(n)}$为参数$\theta$的极大似然估计量.

第一节 矩估计和极大似然估计

例3:设总体X具有均匀分布,密度函数为

$$f(x,\theta)=\begin{cases}\frac{1}{\theta}, & 0<x<\theta;\\ 0, & \text{其它},\end{cases}$$

其中$\theta>0$是未知参数,求θ的极大似然估计.
解:设$(x_1,\cdots,x_n)$是样本$(X_1,\cdots,X_n)$的一组观测值,似然函数为

$$L(\theta)=\begin{cases}\frac{1}{\theta^n}, & 0<x_i<\theta;\\ 0, & \text{其它},\end{cases}$$

要使L最大,必须使θ最小,当$\theta=\max_{1\le i\le n}x_i=x_{(n)}$时,可使$L$最大,故$\hat{\theta}=X_{(n)}$为参数$\theta$的极大似然估计量.

第一节 矩估计和极大似然估计

例4: 设总体$X \sim N(\mu, \sigma^2)$,其中$\theta = (\mu, \sigma^2)$为未知参数向量,参数空间$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$, 求$\mu, \sigma^2$的极大似然估计.

解: 设$(x_1, \cdots, x_n)$是样本$(X_1, \cdots, X_n)$的一组观测值,于是似然函数为

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

两边取对数得

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

分别求上式关于μ和σ^2的偏导数,并令它们为0,得似然方程组

$$\begin{cases} \frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0; \\ \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0, \end{cases}$$

第一节 矩估计和极大似然估计

例4: 设总体$X \sim N(\mu, \sigma^2)$,其中$\theta = (\mu, \sigma^2)$为未知参数向量,参数空间$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$, 求$\mu, \sigma^2$的极大似然估计.

解: 设$(x_1, \cdots, x_n)$是样本$(X_1, \cdots, X_n)$的一组观测值,于是似然函数为

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

两边取对数得

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

分别求上式关于μ和σ^2的偏导数,并令它们为0,得似然方程组

$$\begin{cases} \frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0; \\ \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0, \end{cases}$$

第一节 矩估计和极大似然估计

解之得

$$\begin{cases} \hat{\mu} & = & \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}; \\ \hat{\sigma^2} & = & \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \tilde{s_n^2}. \end{cases}$$

容易验证$\hat{\mu}, \hat{\sigma}^2$满足关系式

$$L(\hat{\mu}, \hat{\sigma}^2) = \sup_{-\infty<\mu<\infty, \sigma^2>0} L(\mu, \sigma^2),$$

所以$\overline{X}$和$\tilde{S_n^2}$分别是μ和σ^2的极大似然估计.

第一节　矩估计和极大似然估计

注：(1)极大似然估计充分利用了总体分布所提供的信息，比矩法估计优，特别是对大样本的情况。

(2)必须知道总体的分布，且有时不易求出极大似然方程组的解。

(3)若函数$g(\theta)$具有单值反函数，$\hat{\theta}$是θ的MLE，则$g(\hat{\theta})$是$g(\theta)$的MLE.

第二节 估计量的优良性准则

第二节 估计量的优良性准则

第二节 估计量的优良性准则

一.无偏估计

1.定义： 设总体X具有分布族$\{F(x;\theta):\theta\in\Theta\}$, $(X_1,\cdots,X_n)$是取自这个总体的一个样本, $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$是未知参数$\theta$的一个估计量,如果对于一切$\theta\in\Theta$,都有

$$E_\theta[\hat{\theta}(X_1,\cdots,X_n)]=\theta$$

则称$\hat{\theta}(X_1,\cdots,X_n)$为$\theta$的**无偏估计**,简记为UE.

第二节 估计量的优良性准则

例1: 设总体X的均值为μ,方差为σ^2,$(X_1,\cdots,X_n)$为取自总体的样本,则
(1)$E(\bar{X})=\mu$;
(2)总体X的k阶原点矩$m_k=E(X^k)$存在,则样本k阶原点矩A_k满足$E(A_k)=m_k$.

解:因为$E(X_i)=E(X)=\mu, D(X_i)=D(X)=\sigma^2, i=1,\cdots,n$,且$X_1,\cdots,X_n$相互独立,则
$E(\bar{X})=\frac{1}{n}\sum_{i=1}^n E(X_i)=\frac{1}{n}\sum_{i=1}^n \mu=\mu$,所以$\bar{X}$是$\mu$的无偏估计。
$E(A_k)=E(\frac{1}{n}\sum_{i=1}^n X_i^k)=\frac{1}{n}\sum_{i=1}^n E(X_i^k)=m_k$,所以$A_k$是$m_k$的无偏估计。

第二节 估计量的优良性准则

例1: 设总体X的均值为μ,方差为σ^2,$(X_1,\cdots,X_n)$为取自总体的样本，则
(1)$E(\bar{X})=\mu$;
(2)总体X的k阶原点矩$m_k=E(X^k)$存在，则样本k阶原点矩A_k满足$E(A_k)=m_k$.

解： 因为$E(X_i)=E(X)=\mu, D(X_i)=D(X)=\sigma^2, i=1,\cdots,n$,且$X_1,\cdots,X_n$相互独立，则
$E(\bar{X})=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu$,所以$\bar{X}$是$\mu$的无偏估计。
$E(A_k)=E(\frac{1}{n}\sum_{i=1}^{n}X_i^k)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=m_k$,所以$A_k$是$m_k$的无偏估计。

第二节 估计量的优良性准则

2.若$E(\hat{\theta}) \neq \theta$,则称它是有偏的，且称函数$b(\theta, \hat{\theta}) = E(\hat{\theta}) - \theta$为$\hat{\theta}$的偏.

3.若有一列θ的估计$\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots, X_n)$,对一切$\theta \in \Theta$满足$\lim_{n\to\infty} E_\theta(\hat{\theta}_n) = \theta$，即$\lim_{n\to\infty} b_n = 0$,则称$\hat{\theta}_n$为$\theta$的渐近无偏估计量。

第二节　估计量的优良性准则

例2：(续)
验证$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$，$\tilde{S}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是否为σ^2的无偏估计。

解：由Th1.2知$E(S^2)=\sigma^2$,所以S^2是σ^2的无偏估计。
$E(\tilde{S}^2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}\sigma^2\neq\sigma^2$，所以$\tilde{S}^2$不是$\sigma^2$的无偏估计，
偏为$E(\tilde{S}^2)-\sigma^2=(\frac{n-1}{n}-1)\sigma^2=-\frac{\sigma^2}{n}$。
但$\lim\limits_{n\to\infty}E(\tilde{S}_n^2)=\sigma^2$，所以$\tilde{S}^2$是$\sigma^2$的渐近无偏估计.

第二节 估计量的优良性准则

例2：(续)

验证$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$，$\tilde{S}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是否为σ^2的无偏估计。

解： 由Th1.2知$E(S^2)=\sigma^2$,所以S^2是σ^2的无偏估计。
$E(\tilde{S}^2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}\sigma^2\neq\sigma^2$，所以$\tilde{S}^2$不是$\sigma^2$的无偏估计，
偏为$E(\tilde{S}^2)-\sigma^2=(\frac{n-1}{n}-1)\sigma^2=-\frac{\sigma^2}{n}$。
但$\lim\limits_{n\to\infty}E(\tilde{S}_n^2)=\sigma^2$，所以$\tilde{S}^2$是$\sigma^2$的渐近无偏估计.

第二节 估计量的优良性准则

4.若对θ的任一实值函数$g(\theta)$,如果存在估计量$T = T(X_1, \cdots, X_n)$, 使得对一切$\theta \in \Theta$,有$E_\theta(T) = g(\theta)$,则$g(\theta)$称为可估计函数.

例： 设$X \sim b(n, p), 0 < p < 1$, X_1是取自这个总体的一个样本,则函数$g(p) = \frac{1}{p}$不可估.

第二节 估计量的优良性准则

例3：设$(X_1,\cdots,X_n)$是从正态总体$N(\mu,\sigma^2)$中抽取的一个样本,由定理知$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$, $(n-1)S^2/\sigma^2\sim\chi^2(n-1)$. 虽然$\overline{X}$是$\mu$的一个无偏估计,但$\overline{X}^2$不是$\mu^2$的无偏估计,

$$E_\mu(\overline{X}^2)=D_\mu(\overline{X})+[E_\mu(\overline{X})]^2=\frac{\sigma^2}{n}+\mu^2\neq\mu^2.$$

同样,虽然S^2是σ^2的一个无偏估计,但S也不是σ的无偏估计.事实上

$$\begin{aligned}E_\sigma(\sqrt{n-1}S/\sigma) &= \int_0^\infty\frac{\sqrt{x}}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}e^{-\frac{x}{2}}x^{\frac{n-1}{2}-1}dx\\ &= \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}2^{\frac{n}{2}}\Gamma(\frac{n}{2})\\ &= \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})},\end{aligned}$$

$$E_\sigma(S)=\sigma[\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}]\neq\sigma.$$

第二节 估计量的优良性准则

可见,S不是σ的无偏估计,它的偏为

$$b(\sigma, S) = \sigma[\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} - 1].$$

但容易由这个有偏估计修改得到σ的无偏估计

$$\hat{\sigma} = \sqrt{\frac{n-1}{2}}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}S.$$

第二节 估计量的优良性准则

注：(1)若$\hat{\theta}$是θ的有偏估计，且$E(\hat{\theta}) = a + b\theta$,(a,$b \neq 0$为常数),则可以构造一个$\theta$的无偏估计$\hat{\theta}^* = \frac{\hat{\theta}-a}{b}$。

(2)若$\hat{\theta}$是θ的无偏估计，除了f是线性函数外，并不能推出$f(\hat{\theta})$是$f(\theta)$的无偏估计。

第二节 估计量的优良性准则

例4. 设总体X服从均匀分布,密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}, & 0<x<\theta;\\ 0, & \text{其它},\end{cases}$$

$(X_1,\cdots,X_n)$是取自这个总体的一个样本,
(1)求θ的矩估计, 验证无偏性。
(2)求θ的MLE, 验证无偏性。

解: 由于$EX=\frac{\theta}{2}$,故$\hat{\theta}=2\overline{X}$是$\theta$的矩估计,且它是无偏的.
由前知$\hat{\theta}_L=X_{(n)}$是θ的极大似然估计.
$E_\theta(\hat{\theta}_L)=\int_0^\theta x\cdot\frac{n}{\theta}(\frac{x}{\theta})^{n-1}dx=\frac{n}{n+1}\theta$, 即$\hat{\theta}_L$不是$\theta$的无偏估计,但$\hat{\theta}_L$是$\theta$的渐近无偏估计.
由$\hat{\theta}_L=X_{(n)}$可以构造θ的一个无偏估计量$\hat{\theta}^*=\frac{n+1}{n}X_{(n)}$.

第二节 估计量的优良性准则

例4. 设总体X服从均匀分布,密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}, & 0<x<\theta;\\ 0, & \text{其它},\end{cases}$$

$(X_1,\cdots,X_n)$是取自这个总体的一个样本,
(1)求θ的矩估计, 验证无偏性。
(2)求θ的MLE, 验证无偏性。

解: 由于$EX=\frac{\theta}{2}$,故$\hat{\theta}=2\overline{X}$是$\theta$的矩估计,且它是无偏的.
由前知$\hat{\theta}_L=X_{(n)}$是θ的极大似然估计.
$E_\theta(\hat{\theta}_L)=\int_0^\theta x\cdot\frac{n}{\theta}(\frac{x}{\theta})^{n-1}dx=\frac{n}{n+1}\theta$, 即$\hat{\theta}_L$不是$\theta$的无偏估计,但$\hat{\theta}_L$是$\theta$的渐近无偏估计.
由$\hat{\theta}_L=X_{(n)}$可以构造θ的一个无偏估计量$\hat{\theta}^*=\frac{n+1}{n}X_{(n)}$.

第二节 估计量的优良性准则

考虑θ的这两个无偏估计量$\hat{\theta}$与$\hat{\theta}^*$的方差,

$$\begin{aligned} D_\theta(\hat{\theta}) &= D_\theta(2\overline{X}) = 4D_\theta(\overline{X}) = \tfrac{4}{n}D_\theta(X) = \tfrac{\theta^2}{3n}; \\ D_\theta(\hat{\theta}_L) &= \int_0^\theta x^2 \cdot n\tfrac{x^{n-1}}{\theta^n}dx - (\tfrac{n}{n+1}\theta)^2 \\ &= \tfrac{n\theta^2}{(n+1)^2(n+2)}, \end{aligned}$$

所以

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

显然,$D_\theta(\hat{\theta}^*) \le D_\theta(\hat{\theta})$,而且当$n$很大时,有$\lim_{n\to\infty}\frac{D_\theta(\hat{\theta}^*)}{D_\theta(\hat{\theta})} = \lim_{n\to\infty}\frac{3}{n+2} = 0$. 可见$\hat{\theta}^*$和$\hat{\theta}$的取值都在参数真值$\theta$的周围波动,但$\hat{\theta}^*$比$\hat{\theta}$取值更集中,作为$\theta$的估计量,$\hat{\theta}^*$比$\hat{\theta}$好.

第二节 估计量的优良性准则

二.一致最小方差无偏估计

1.定义：设$T_1(X_1,\cdots,X_n)$为可估函数$g(\theta)$的无偏估计量,若对于任意的$\theta\in\Theta$和任意的$g(\theta)$的无偏估计量$T(X_1,\cdots,X_n)$,都有

$$D_\theta[T_1(X_1,\cdots,X_n)]\le D_\theta[T(X_1,\cdots,X_n)],$$

则称$T_1(X_1,\cdots,X_n)$是$g(\theta)$的 **一致最小方差无偏估计量**,简记为UMVUE.

记$U\triangleq\{T:E_\theta(T)=g(\theta),D_\theta(T)<\infty,\ 对一切\ \theta\in\Theta\}$, 为可估函数$g(\theta)$的方差有限的无偏估计量的集合.
$U_0\triangleq\{T:E_\theta(T)=0,D_\theta(T)<\infty,\ 对一切\ \theta\in\Theta\}$, 为数学期望为零,方差有限的估计量的集合.

第二节 估计量的优良性准则

2.定理: 设$T_1 \in U$,则T_1是$g(\theta)$的一致最小方差无偏估计的充要条件为:对一切$\theta \in \Theta$和$T_0 \in U_0$,有$E_\theta(T_1 T_0) = 0$.

3.推论: 设T_1, T_2分别是参数θ的可估函数$g_1(\theta), g_2(\theta)$的一致最小方差无偏估计量, 则$b_1 T_1 + b_2 T_2$是$b_1 g_1(\theta) + b_2 g_2(\theta)$的一致最小方差无偏估计量, 其中$b_1, b_2$为固定常数.

4.定理: $U \triangleq \{T : E_\theta(T) = g(\theta), Var_\theta(T) < \infty, \forall \theta \in \Theta\}$, 则至多存在一个$g(\theta)$的一致最小方差无偏估计量.

第二节 估计量的优良性准则

三.相合估计量

1.定义： 设$T_n = T_n(X_1, \cdots, X_n)$是$g(\theta)$的估计量，若对任何$\theta \in \Theta$，$T_n$依概率收敛于$g(\theta)$,则称$T_n$是$g(\theta)$的**相合估计**。

注： (1)另一种表述：$\forall \varepsilon > 0, \lim_{n\to\infty} P\{|T_n - g(\theta)| \geq \varepsilon\} = 0.$

(2)相合性是在极限意义下引入的，适用大样本情况。

(3)若T_n以概率1(几乎处处)收敛于$g(\theta)$，即$P_\theta\{\lim_{n\to\infty} T_n(X_1, \cdots, X_n) = g(\theta)\} = 1$，则称$T_n$是$g(\theta)$的**强相合估计**。

(4)若T_n是$g(\theta)$的强相合估计,它也是$g(\theta)$的相合估计.

第二节 估计量的优良性准则

2.证明相合估计的方法

定义，利用切比雪夫不等式$P\{|X-EX|\geq\varepsilon\}\leq\frac{DX}{\varepsilon^2}$.

第二节 估计量的优良性准则

例5： 设总体X服从均匀分布,密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}, & 0<x<\theta;\\ 0, & \text{其它},\end{cases}$$

$(X_1,\cdots,X_n)$是取自这个总体的一个样本，证明：$\hat{\theta}=X_{(n)}$是θ的相合估计。

解：$\hat{\theta}=X_{(n)}$的密度函数为

$$f_n(x;\theta)=\begin{cases}n(\frac{x}{\theta})^{n-1}\frac{1}{\theta}=\frac{n}{\theta^n}x^{n-1}, & 0<x<\theta;\\ 0, & \text{其它},\end{cases}$$

$E(\hat{\theta})=\int_0^\theta x\cdot\frac{n}{\theta^n}x^{n-1}dx=\frac{n}{n+1}\theta,$

$E(\hat{\theta}^2)=\int_0^\theta x^2\cdot\frac{n}{\theta^n}x^{n-1}dx=\frac{n}{n+2}\theta^2,$

$D(\hat{\theta})=E(\hat{\theta}^2)-[E(\hat{\theta})]^2=\frac{n}{n+2}\theta^2-[\frac{n}{n+1}\theta]^2=\frac{n}{(n+1)^2(n+2)}\theta^2$

第二节 估计量的优良性准则

$$
\begin{aligned}
P\{|\hat{\theta}-\theta| \geq \varepsilon\} &= P\{|\hat{\theta}-\tfrac{n}{n+1}\theta-\tfrac{1}{n+1}\theta| \geq \varepsilon\} \\
&\leq P\{|\hat{\theta}-\tfrac{n}{n+1}\theta|+\tfrac{1}{n+1}\theta \geq \varepsilon\} \\
&= P\{|\hat{\theta}-\tfrac{n}{n+1}\theta| \geq \varepsilon-\tfrac{1}{n+1}\theta\} \\
&= P\{|\hat{\theta}-E(\hat{\theta})| \geq \varepsilon-\tfrac{1}{n+1}\theta\} \\
&\leq \frac{D(\hat{\theta})}{(\varepsilon-\frac{1}{n+1}\theta)^2} \\
&= \frac{n\theta^2}{(n+1)^2(n+2)(\varepsilon-\frac{1}{n+1}\theta)^2} \\
&\longrightarrow 0, n \to \infty
\end{aligned}
$$

第二节 估计量的优良性准则

3.定理: 设$(X_1, \cdots, X_n)$是取自具有分布族$\{F(x;\theta):\theta\in\Theta\}$的总体$X$的一个样本, 若$E|X|^p<\infty$, 其中$p$是某一正整数,则样本的$k$阶原点矩$(1\le k\le p)$ $A_k=\frac{1}{n}\sum_{i=1}^n X_i^k$是总体$k$阶矩$\alpha_k=EX^k$的相合估计.

4.定理: 如果T_n是θ的相合估计量,$g(x)$在$x=\theta$连续,则$g(T_n)$也是$g(\theta)$的相合估计量.

第三节 Rao–Cramer 不等式

第三节 Rao–Cramer 不等式

第三节 Rao–Cramer 不等式

一. Rao–Cramer 不等式

1.定义： 若单参数分布密度族(或单参数概率分布族)$\{f(x;\theta):\theta\in\Theta\}$满足如下条件：
(1)参数空间Θ是数轴上的一个开区间;
(2)导数$\frac{\partial f(x;\theta)}{\partial\theta}$对一切$\theta\in\Theta$都存在;
(3)集合$S_\theta\triangleq\{x:f(x;\theta)>0\}$与$\theta$无关($S_\theta$称为支撑);
(4) $\frac{\partial}{\partial\theta}\int f(x;\theta)dx=\int\frac{\partial f(x;\theta)}{\partial\theta}dx=0$,对一切$\theta\in\Theta$;
(5)对任$\theta\in\Theta$,有

$$\begin{aligned}0<I(\theta)&\triangleq E_\theta[\frac{\partial}{\partial\theta}\log f(X;\theta)]^2\\&=\int(\frac{\partial\log f(x;\theta)}{\partial\theta})^2f(x;\theta)dx,\end{aligned}$$

则称这个分布密度族(概率分布族)为Rao–Cramer **正则分布族**,其中条件(1)–(5) 称为**正则条件**, $I(\theta)$称为该分布密度族的**Fisher**信息函数

第三节 Rao–Cramer 不等式

例1： 验证Poisson分布族 $\{f(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}, x=0,1,2,\cdots:\ \lambda>0\}$ 是Rao–Cramer正则分布族.

解：(1)参数空间$\{\lambda>0\}$是一开区间；

(2) $\frac{\partial f(x;\theta)}{\partial\lambda}=\frac{\lambda^{x-1}}{x!}(x-\lambda)e^{-\lambda}$,对一切$\lambda>0$存在；

(3) $f(x;\lambda)>0$,对一切非负整数x及λ成立；

(4)

$$\sum_{x=0}^{\infty}\frac{\lambda^{x-1}}{x!}(x-\lambda)e^{-\lambda}=\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}e^{-\lambda}-\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}e^{-\lambda}=1-1=0;$$

(5) $\log f(x;\lambda)=x\log\lambda-\log x!-\lambda$, $\frac{\partial\log f(x;\lambda)}{\partial\lambda}=\frac{x}{\lambda}-1$, 因此

$$E_\lambda[\frac{\partial\log f(X;\lambda)}{\partial\lambda}]=0, I(\lambda)=E_\lambda[\frac{\partial\log f(X;\lambda)}{\partial\lambda}]^2=D_\lambda(\frac{X}{\lambda}-1)=\frac{1}{\lambda},$$

故Poisson分布族是Rao–Cramer正则分布族.

第三节 Rao–Cramer 不等式

例2： 正态分布族$\{N(\mu,1):-\infty<\mu<\infty\}$是Rao–Cramer正则分布族.

解:对所考虑的正态分布密度

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

条件(1)–(4)是显然满足的,只需验证条件(5)也满足.事实上,由于

$$\frac{\partial\log f(x;\mu)}{\partial\mu}=x-\mu,$$

故

$$E_\mu[\frac{\partial\log f(X;\mu)}{\partial\mu}]^2=E_\mu(X-\mu)^2=D_\mu X=1.$$

所以正态分布族$\{N(\mu,1):-\infty<\mu<\infty\}$是Rao–Cramer正则分布族.

第三节 Rao–Cramer 不等式

注：

(1)常用的单参数分布族都是Rao–Cramer正则分布族,但均匀分布族$\{U(0,\theta):\theta>0\}$ 不是Rao–Cramer正则分布族，因为它的支撑$S_\theta=\{x:f(x;\theta)>0\}$是一个依赖于$\theta$的开区间.

(2)计算$I(\theta)$的一个简便方法：$I(\theta)=-E_\theta\left(\frac{\partial^2\log f(X;\theta)}{\partial\theta^2}\right)$

第三节 Rao–Cramer 不等式

2.定理：(Rao–Cramer不等式) 设总体X的概率分布密度族$\{f(x;\theta):\theta\in\Theta\}$是Rao–Cramer正则分布族,可估函数$g(\theta)$是$\Theta$上的可微函数,$(X_1,\cdots,X_n)$是取自总体$X$的一个样本.若$T(X_1,\cdots,X_n)$是$g(\theta)$的一个无偏估计量,且对一切$\theta\in\Theta$,满足

$$\begin{aligned}&\frac{\partial}{\partial\theta}\int T(x_1,\cdots,x_n)L(x_1,\cdots,x_n;\theta)dx_1\cdots dx_n\\&=\int T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}L(x_1,\cdots,x_n;\theta)dx_1\cdots dx_n,\end{aligned}\qquad(*)$$

其中$L(x_1,\cdots,x_n;\theta)=\prod_{i=1}^n f(x_i;\theta)$是样本$(X_1,\cdots,X_n)$的联合分布密度函数. 则对一切$\theta\in\Theta$,有

$$D_\theta(T)\geq\frac{[g'(\theta)]^2}{nI(\theta)}\qquad(1)$$

其中$I(\theta)$是该分布族的Fisher信息函数.

第三节 Rao–Cramer 不等式

特别当$g(\theta)=\theta$时，$D_\theta(T)\geq\frac{1}{nI(\theta)}$.
如果(1)中等号成立，则$\exists C(\theta)\neq 0$, s.t.

$$\frac{\partial\ln L(x_1,\cdots,x_n;\theta)}{\partial\theta}=C(\theta)[T(x_1,\ldots,x_n)-g(\theta)]a.s.$$

注： (1)称满足上述定理中正则条件的估计量T为正规估计量。

(2)称Rao—Cramer不等式所规定的下界为**C–R**下界.

(3)若一个无偏估计T的方差达到这个下界,且$g(\theta)$的一切无偏估计都满足条件(*),则T就是$g(\theta)$的UMVUE.

(4)离散性总体：密度函数→概率分布，积分→求和

第三节 Rao–Cramer 不等式

(5)$nI(\theta)$越大，$Var_\theta(T)$可达下界越低，即$g(\theta)$可能被估计得越准确．所以$I(\theta)$在一定意义上表示对X进行观测所提供的关于未知参数θ的平均信息．

(6)C-R不等式只对C-R正则族成立．如定理中条件之一不满足，C-R不等式可能不成立．

第三节 Rao–Cramer 不等式

例: 设$(X_1,...,X_n)$是总体X的一个样本, X的密度函数为

$$f(x;\theta)=\mathrm{e}^{-(x-\theta)}, x>\theta.$$

因为$\{x: f(x;\theta)>0\}=\{x>\theta\}$, 所以不是C-R正则族.
形式上计算C-R下界:

$$I(\theta)=E\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right]^2=E(1)=1,\quad \text{下界}\frac{1}{n}$$

$$\hat{\theta}=X_{(1)}-\frac{1}{n},\quad f(t;\theta)=n\mathrm{e}^{-(nt-n\theta+1)},\quad t>\theta-\frac{1}{n}.$$

$$E(\hat{\theta})=\theta,\quad E(\hat{\theta}^2)=\theta^2+\frac{1}{n^2},\quad D(\hat{\theta})=\frac{1}{n^2}<\frac{1}{n}$$

第三节 Rao–Cramer 不等式

例：设$(X_1, ..., X_n)$是总体X的一个样本，X的密度函数为

$$f(x;\theta) = \mathrm{e}^{-(x-\theta)}, x > \theta.$$

因为$\{x : f(x;\theta) > 0\} = \{x > \theta\}$，所以不是C-R正则族.
形式上计算C-R下界:

$$I(\theta) = E\left[\frac{\partial \ln f(X;\theta)}{\partial\theta}\right]^2 = E(1) = 1, \quad \text{下界}\frac{1}{n}$$

$$\hat{\theta} = X_{(1)} - \frac{1}{n}, \quad f(t;\theta) = n\mathrm{e}^{-(nt-n\theta+1)}, \quad t > \theta - \frac{1}{n}.$$

$$E(\hat{\theta}) = \theta, \quad E(\hat{\theta}^2) = \theta^2 + \frac{1}{n^2}, \quad D(\hat{\theta}) = \frac{1}{n^2} < \frac{1}{n}$$

第三节 Rao–Cramer 不等式

例3:设总体X具有Bernoulli分布族$\{b(1,p):\ p\in(0,1)\}$, $(X_1,\cdots,X_n)$是取自这一总体的样本,求p的正规无偏估计的方差下界.

解: 容易验证Bernoulli分布族是Rao–Cramer正则分布族,有

$$f(x;p)=p^x(1-p)^{1-x},\quad x=0,1,\ 0<p<1,$$

且

$$I(p)=\sum_{x=0}^{1}[\frac{\partial \log f(x;p)}{\partial p}]^2 f(x;p)=\frac{1}{p(1-p)},$$

故p的正规无偏估计的方差下界为

$$\frac{1}{nI(p)}=\frac{p(1-p)}{n}.$$

$\hat{p}=\overline{X}$是p的无偏估计,而$D(\overline{X})=\frac{p(1-p)}{n}$达到Rao–Cramer不等式的下界.

第三节　Rao–Cramer 不等式

例3:设总体X具有Bernoulli分布族$\{b(1,p):\ p\in(0,1)\}$, $(X_1,\cdots,X_n)$是取自这一总体的样本,求p的正规无偏估计的方差下界.

解: 容易验证Bernoulli分布族是Rao–Cramer正则分布族,有

$$f(x;p)=p^x(1-p)^{1-x},\quad x=0,1,\ 0<p<1,$$

且

$$I(p)=\sum_{x=0}^{1}[\frac{\partial\log f(x;p)}{\partial p}]^2 f(x;p)=\frac{1}{p(1-p)},$$

故p的正规无偏估计的方差下界为

$$\frac{1}{nI(p)}=\frac{p(1-p)}{n}.$$

$\hat{p}=\overline{X}$是p的无偏估计,而$D(\overline{X})=\frac{p(1-p)}{n}$达到Rao–Cramer不等式的下界.

第三节 Rao–Cramer 不等式

例4： 设X的密度函数为$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & x\le 0,\end{cases}$，$(X_1,\cdots,X_n)$ 是取自总体X的样本,未知参数$\lambda>0$, 求待估函数$u(\lambda)=\frac{1}{\lambda}$的正规无偏估计量的C–R下界.

解：指数分布族是Rao–Cramer正则分布族，且当$x>0$时，

$$\log f(x;\lambda)=-\lambda x+\log\lambda$$

$$\frac{\partial^2\log f(x;\lambda)}{\partial\lambda^2}=\frac{\partial^2(-\lambda x+\log\lambda)}{\partial\lambda^2}=\frac{\partial}{\partial\lambda}(-x+\frac{1}{\lambda})=-\frac{1}{\lambda^2}$$

所以$I(\lambda)=-E(\frac{\partial^2\log f(X;\lambda)}{\partial\lambda^2})=-E(-\frac{1}{\lambda^2})=\frac{1}{\lambda^2}$

C-R下界: $\frac{[u'(\lambda)]^2}{nI(\lambda)}=\frac{1}{n\lambda^2}$.

$E(\bar{X})=E(X)=\frac{1}{\lambda}=u(\lambda)$, $D(\bar{X})=\frac{1}{n}D(X)=\frac{1}{n}\frac{1}{\lambda^2}=\frac{1}{n\lambda^2}$,
$\bar{X}$的方差达到C-R下界.

第三节 Rao–Cramer 不等式

例4： 设X的密度函数为$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & x\le 0,\end{cases}$，$(X_1,\cdots,X_n)$ 是取自总体X的样本,未知参数$\lambda>0$, 求待估函数$u(\lambda)=\frac{1}{\lambda}$的正规无偏估计量的C–R下界.

解：指数分布族是Rao–Cramer正则分布族，且当$x>0$时，

$$\log f(x;\lambda)=-\lambda x+\log\lambda$$

$$\frac{\partial^2\log f(x;\lambda)}{\partial\lambda^2}=\frac{\partial^2(-\lambda x+\log\lambda)}{\partial\lambda^2}=\frac{\partial}{\partial\lambda}(-x+\frac{1}{\lambda})=-\frac{1}{\lambda^2}$$

所以$I(\lambda)=-E(\frac{\partial^2\log f(X;\lambda)}{\partial\lambda^2})=-E(-\frac{1}{\lambda^2})=\frac{1}{\lambda^2}$

C-R下界: $\frac{[u'(\lambda)]^2}{nI(\lambda)}=\frac{1}{n\lambda^2}$.

$E(\bar{X})=E(X)=\frac{1}{\lambda}=u(\lambda)$, $D(\bar{X})=\frac{1}{n}D(X)=\frac{1}{n}\frac{1}{\lambda^2}=\frac{1}{n\lambda^2}$,

$\bar{X}$的方差达到C-R下界.

第三节 Rao–Cramer 不等式

二.有效估计量

1.定义： 设总体X的概率分布密度族$\{f(x;\theta):\theta\in\Theta\}$是Rao–Cramer正则分布族,$g(\theta)$是可估函数,$T=T(X_1,\cdots,X_n)$是$g(\theta)$的估计量.
(1)若$E_\theta(T)=g(\theta)$,则称

$$e_\theta(T)=\frac{[g'(\theta)]^2}{nI(\theta)}/D_\theta(T)$$

为$g(\theta)$的无偏估计量T的**效率**.
(2)若$E_\theta(T)=g(\theta)$,且$e_\theta(T)=1$,则称T为$g(\theta)$的**有效估计量**.
(3)设$T_n=T_n(X_1,\cdots,X_n)$是$g(\theta)$的一列无偏估计,若

$$\lim_{n\to\infty}e_\theta(T_n)=e_0,$$

则称e_0为T_n的**渐近效率**;当$e_0=1$时,称T_n是$g(\theta)$的**渐近有效估计**.

第三节 Rao–Cramer 不等式

例5：设X服从$\Gamma(r,\frac{1}{\lambda})$,即

$$f(x;\theta)=\begin{cases}\frac{1}{\lambda^r\Gamma(r)}x^{r-1}e^{-\frac{x}{\lambda}}, x>0;\\ 0, x\le 0,\end{cases}$$

其中$r>0$已知,$\lambda>0$未知，$(X_1,\cdots,X_n)$为X的样本，证明：$\frac{1}{nr}\sum_{i=1}^n X_i$ 是λ的有效估计。

证： 因为$E(\frac{1}{nr}\sum_{i=1}^n X_i)=\frac{1}{nr}nr\lambda=\lambda$,所以$\frac{1}{nr}\sum_{i=1}^n X_i$是$\lambda$的无偏估计。
验证条件，它是Rao-Cramer正则分布族，$\ln f(x,\lambda)=-r\ln\lambda-\ln\Gamma(r)+(r-1)\ln x-\frac{x}{\lambda}$, 则

$$\frac{\partial\ln f(x;\lambda)}{\partial\lambda}=-\frac{r}{\lambda}+\frac{x}{\lambda^2},\frac{\partial^2\ln f(x;\lambda)}{\partial\lambda^2}=\frac{r}{\lambda^2}-\frac{2x}{\lambda^3}$$

第三节 Rao–Cramer 不等式

例5： 设X服从$\Gamma(r,\frac{1}{\lambda})$,即

$$f(x;\theta)=\begin{cases}\frac{1}{\lambda^r\Gamma(r)}x^{r-1}e^{-\frac{x}{\lambda}}, x>0;\\ 0, x\leq 0,\end{cases}$$

其中$r>0$已知,$\lambda>0$未知，$(X_1,\cdots,X_n)$为X的样本，证明：$\frac{1}{nr}\sum_{i=1}^n X_i$ 是λ的有效估计。

证： 因为$E(\frac{1}{nr}\sum_{i=1}^n X_i)=\frac{1}{nr}nr\lambda=\lambda$,所以$\frac{1}{nr}\sum_{i=1}^n X_i$是$\lambda$的无偏估计。
验证条件，它是Rao-Cramer正则分布族，$\ln f(x,\lambda)=-r\ln\lambda-\ln\Gamma(r)+(r-1)\ln x-\frac{x}{\lambda}$，则

$$\frac{\partial\ln f(x;\lambda)}{\partial\lambda}=-\frac{r}{\lambda}+\frac{x}{\lambda^2},\frac{\partial^2\ln f(x;\lambda)}{\partial\lambda^2}=\frac{r}{\lambda^2}-\frac{2x}{\lambda^3}$$

第三节 Rao–Cramer 不等式

所以

$$I(\lambda) = -E(\frac{r}{\lambda^2} - \frac{2X}{\lambda^3}) = -\frac{r}{\lambda^2} + \frac{2\lambda r}{\lambda^3} = \frac{r}{\lambda^2}$$

所以C-R下界为$\frac{1}{nI(\lambda)} = \frac{1}{n\frac{r}{\lambda^2}} = \frac{\lambda^2}{nr}$

而$D(\frac{1}{nr}\sum_{i=1}^n X_i) = \frac{1}{n^2r^2}\sum_{i=1}^n D(X_i) = \frac{nr\lambda^2}{n^2r^2} = \frac{\lambda^2}{nr}$达到C-R下界，所以$\frac{1}{nr}\sum_{i=1}^n X_i$ 是λ的有效估计。

第三节 Rao–Cramer 不等式

2.定义： 对可估函数$g(\theta)$的任意两个无偏估计量T_1和T_2, 称

$$e(T_1|T_2)=\frac{D_\theta(T_1)}{D_\theta(T_2)}$$

为估计量T_1关于T_2的**相对效率**;如果$e(T_1|T_2)<1$,则称T_1比T_2有效.

注： 也就是若$D(T_1)<D(T_2)$,则T_1比T_2有效。

第四节 Rao–Blackwell定理

1.定理(Rao-Blackwell定理) 设$T(X_1,...,X_n)$是分布族$\{F(x;\theta):\theta\in\Theta\}$ 的一个充分统计量，V是$g(\theta)$的一个无偏估计量，则

$$V_0=V_0(T)=E(V|T)$$

是$g(\theta)$的一个无偏估计量且$Var_\theta(V_0)\le Var_\theta(V),\forall\theta\in\Theta$. 其中等号成立的充要条件是$V$在a.s.意义下是$T$的函数.

第四节 Rao–Blackwell定理

注: (1) 定理提供了一种改善估计的方法，若已知一无偏估计量，则可构造一新的无偏估计量且其方差比原估计量的方差小；
(2) 一致最小方差无偏估计量(UMVUE)一定是充分统计量的函数，否则按定理的方法可构造比原估计量有更小方差的估计；
(3) 若$g(\theta)$的由充分统计量构造的无偏估计量是a.s.唯一的，则此无偏估计量一定是一致最小方差无偏估计量.
(4) 由充分统计量构造的无偏估计量为充分无偏估计量.

第四节　Rao–Blackwell定理

2.定理 在Rao-Cramer正则条件下, $g(\theta)$的无偏估计量$T(X_1, ..., X_n)$是有效估计量，则$T(X_1, ..., X_n)$是充分统计量.

3.定理(Lehmann-Scheffe定理) 设$(X_1, ..., X_n)$是取自分布族$\{F(x;\theta) : \theta \in \Theta\}$的一个样本，$T(X_1, ..., X_n)$是$\theta$的充分完备统计量. 如果一个只依赖于$T(X_1, ..., X_n)$的估计量$h[T(X_1, ..., X_n)]$为$g(\theta)$的无偏估计，则它是$g(\theta)$的唯一的一致最小方差无偏估计量.

第四节 Rao–Blackwell定理

例1：设$(X_1, ..., X_n)$是取自两点分布族$\{b(1, p) : 0 < p < 1\}$的一个样本，求p的一致最小方差无偏估计量.

例2：设$(X_1, ..., X_n)$是取自均匀分布族$\{U(0, \theta) : 0 < \theta < 1\}$的一个样本，求$\theta$的一致最小方差无偏估计量.

例3：设$(X_1, ..., X_n)$是取自参数为λ的Poisson分布的一个样本，其中$\lambda > 0$为未知参数，求$P_\lambda(k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, k = 0, 1, 2...$的一致最小方差无偏估计量.