应用统计

第5讲 参数点估计量的评价

点估计量的评价标准

1)．相合性　估计量随着样本量的增加而逼近参数的真实值。

定义：设$\theta \in \Theta$为未知参数，$\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \cdots, x_n)$是由$n$个样本得到的$\theta$的一个估计量，$n$是样本容量，若对任何$\varepsilon > 0$，有$\lim\limits_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) = 0$，则称$\hat{\theta}_n$为参数$\theta$的相合估计。

依概率收敛，也叫弱相合估计

若是强相合估计则需要几乎处处收敛

定理：　设$\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \cdots, x_n)$是$\theta$的一个估计量，若

$\lim\limits_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta$，　$\lim\limits_{n\to\infty} Var\left(\hat{\theta}_n\right) = 0$，　则$\hat{\theta}_n$是参数$\theta$的相合估计。

证明相合性的时候经常**也会利用大数定律**

相合性就是只要样本足够多，就能保证估计尽可能的精确。

弱相合经常也会用Markov不等式来证明

点估计量的评价标准

2）．无偏性　保证没有系统偏差

设 $\theta \in \Theta$ 为未知参数，$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 是 θ 的一个估计量，若对任意 $\theta \in \Theta$ 有 $E(\hat{\theta})=\theta$，则称 $\hat{\theta}$ 是参数 θ 的无偏估计。

3）．有效性　希望估计围绕参数波动的幅度越小越好。

设 $\hat{\theta}_1$，$\hat{\theta}_2$ 是参数 θ 的无偏估计，如果对任意 $\theta \in \Theta$ 有 $Var(\hat{\theta}_1) \le Var(\hat{\theta}_2)$，且至少有一个 $\theta \in \Theta$ 使得上述不等号严格成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。

简单的应用：样本数量的增加会改善估计。

无偏性与有效性简单的练习

设总体 X 在区间 $[0,\theta]$ 上服从均匀分布，其中 $\theta>0$未知。从总体中抽取样本 X_1,X_2,X_3,X_4 ，则下列统计量均为θ的无偏估计量，比较它们的有效性

$$\hat{\theta}_1=\frac{1}{2}\left(X_1+X_2+X_3+X_4\right),$$

$$\hat{\theta}_2=2X_1+X_2-X_3,$$

$$\hat{\theta}_3=X_1+X_2+X_3-X_4$$

例 1 $X_1, X_2, \cdots, X_n$ 是来自均匀总体 $U(0,\theta)$ 的样本，

参数θ的矩估计量 $\hat{\theta}_1 = 2\bar{X}$，极大似然估计量 $\tilde{\theta} = \max\limits_{1\le k\le n} X_k$，考查无偏性。

$E(\hat{\theta}_1) = E(2\bar{X}) = E(2X) = 2\times\dfrac{\theta}{2} = \theta$，所以 $\hat{\theta}_1 = 2\bar{X}$ 是参数 θ 的无偏估计

$E(\tilde{\theta})$?

计算 $\tilde{\theta} = \max\limits_{1\le k\le n} X_k$ 的分布函数，

当 $0 \le y \le \theta$ 时，由样本的独立同分布性质，可知其分布函数为

$$F_{\tilde{\theta}}(y) = P(\tilde{\theta} \le y) = P\left(\max_{1\le k\le n} X_k \le y\right) = \prod_{k=1}^{n} P(X_k \le y) = \left(\frac{y}{\theta}\right)^n$$

$\tilde{\theta} = \max\limits_{1\le k\le n} X_k$ 的分布函数 $F_{\tilde{\theta}}(y) = \begin{cases} 0, y < 0 \\ \left(\dfrac{y}{\theta}\right)^n, 0 \le y \le \theta \\ 1, y > \theta \end{cases}$，

于是 $\tilde{\theta}$ 的概率密度为 $f_{\tilde{\theta}}(y) = \begin{cases} \dfrac{n}{\theta^n} y^{n-1}, y \in [0,\theta] \\ 0,\text{其它} \end{cases}$， 因此，我们有

$$E(\tilde{\theta}) = \int_{-\infty}^{+\infty} y f_{\tilde{\theta}}(y)dy = \int_0^{\theta} y \frac{n}{\theta^n} y^{n-1} dy$$

$$= \frac{n}{\theta^n}\int_0^{\theta} y^n dy = \frac{n}{\theta^n} \cdot \left.\frac{y^{n+1}}{n+1}\right|_0^{\theta} = \frac{n}{n+1}\theta$$

$E\left(\tilde{\theta}\right)=E\left(\max\limits_{1\le k\le n} X_k\right)=\dfrac{n}{n+1}\theta$，统计量 $\max\limits_{1\le k\le n} X_k$ 不是参数 θ 的无偏估计，

$\dfrac{n+1}{n}\max\limits_{1\le k\le n} X_k$ 则为无偏估计，称为无偏校正。

例 2 $X_1, X_2, \cdots, X_n$ 是来自均匀总体 $U(0,\theta)$ 的样本，$\hat{\theta}_1=2\bar{X}$ 和 $\hat{\theta}_2=\dfrac{n+1}{n}\max\limits_{1\le k\le n} X_k$ 都是参数 θ 的无偏估计，比较它们的有效性。

解 均匀总体 $U(0,\theta)$ 的期望、方差分别为 $E(X)=\dfrac{\theta}{2}$, $Var(X)=\dfrac{\theta^2}{12}$

$$Var(\hat{\theta}_1)=Var\left(2\bar{X}\right)=4Var\left(\bar{X}\right)=4\frac{Var(X)}{n}=4\frac{\theta^2}{12n}=\frac{\theta^2}{3n}$$

$$\hat{\theta}_2 = \frac{n+1}{n} \max_{1 \le k \le n} X_k ,$$

计算 $\tilde{\theta} = \max\limits_{1 \le k \le n} X_k$ 的方差，$\tilde{\theta}$ 的概率密度为 $f_{\tilde{\theta}}(y) = \begin{cases} \dfrac{n}{\theta^n} y^{n-1}, y \in [0, \theta] \\ 0,\text{其它} \end{cases}$

$$E(\tilde{\theta}^2) = \int_0^\theta y^2 f_{\tilde{\theta}}(y) dy = \int_0^\theta y^2 \frac{n}{\theta^n} y^{n-1} dy$$

$$= \frac{n}{\theta^n} \int_0^\theta y^{n+1} dy = \frac{n}{\theta^n} \cdot \frac{\theta^{n+2}}{n+2} = \frac{n}{n+2} \theta^2 ,$$

$$Var\left(\tilde{\theta}\right) = E(\tilde{\theta}^2) - \left(E(\tilde{\theta})\right)^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{\left(n+1\right)^2}\theta^2 = \frac{n}{(n+1)^2(n+2)}\theta^2$$

因此 $Var(\hat{\theta}_2) = Var\left(\dfrac{n+1}{n}\tilde{\theta}\right) = \dfrac{(n+1)^2}{n^2} Var\left(\tilde{\theta}\right) = \dfrac{1}{n(n+2)}\theta^2$。

例 2 $X_1,X_2,\cdots,X_n$ 是来自均匀总体 $U(0,\theta)$ 的样本，$\hat{\theta}_1=2\bar{X}$ 和 $\hat{\theta}_2=\dfrac{n+1}{n}\max\limits_{1\le k\le n}X_k$

都是参数 θ 的无偏估计，比较它们的有效性。

$$Var(\hat{\theta}_1)=\frac{\theta^2}{3n},\qquad Var(\hat{\theta}_2)=\frac{1}{n(n+2)}\theta^2$$

当 $n>1$ 时，$\dfrac{1}{n(n+2)}\theta^2<\dfrac{1}{3n}\theta^2$，所以 $\hat{\theta}_2=\dfrac{n+1}{n}\max\limits_{1\le k\le n}X_k$ 比 $\hat{\theta}_1=2\bar{X}$ 更有效。

例3 $X_1,X_2,\cdots,X_n$ 是来自均匀总体 $X\sim U(-a,a)$ 的样本，用矩估计法估计参数a。

$$Var(X)=\frac{a^2}{3},\quad \frac{a^2}{3}=s^2 \quad\Rightarrow\quad \hat{a}=\sqrt{3}s$$ 是不是 a 的无偏估计？

例 4 设 $X_1, X_2, \cdots, X_n$ 是正态总体 $N\left(\mu, \sigma^2\right)$ 的样本，求参数 μ 和 σ^2 的矩估计量和极大似然估计量。

解：$E\left(X\right)=\mu$，$Var\left(X\right)=\sigma^2$，所以参数 μ 和 σ^2 的矩估计量为

$$\bar{\mu}=\bar{X},\quad \bar{\sigma}^2=S^2;$$

设 $x_1, x_2, \cdots, x_n$ 为样本观测值， 则似然函数

$$L(x;\mu,\sigma^2)=\prod_{k=1}^{n} f(x_k;\mu)=\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_k-\mu)^2}{2\sigma^2}}=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

对数似然函数为

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

将上式对 μ 求导，并令其等于 0

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

将上式对 σ^2 求导，并令其等于 0

$$\frac{d\ln L(x;\mu,\sigma^2)}{d\sigma^2}=-\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^{n}(x_i-\mu)^2=0$$

解得 $\mu=\dfrac{x_1+\cdots+x_n}{n}=\bar{x}$， $\sigma^2=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

根据题意知必为对数似然函数的极大值点，

所以参数 μ 和 σ^2 的极大似然估计量为 $\hat{\mu}=\bar{X}$， $\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

4). 均方误差 平均偏差

参数θ的估计量无偏估计$\hat{\theta}$的均方误差（mean squared error，简记为 MSE），

是由$E\left((\hat{\theta}-\theta)^2\right)$定义的关于$\theta$的函数

$$E\left((\hat{\theta}-\theta)^2\right)=Var\left(\hat{\theta}\right)-\left(E\left(\hat{\theta}\right)-\theta\right)^2$$

例 5 设$X_1,X_2,\cdots,X_n$是正态总体$N\left(\mu,\sigma^2\right)$的样本，计算例 4 中关于参数$\mu$和$\sigma^2$的矩估计量和极大似然估计量的 MSE。

参数μ和σ^2的矩估计量为 $\bar{\mu}=\bar{X}$， $\bar{\sigma}^2=S^2$

参数μ和σ^2的极大似然估计量为$\hat{\mu}=\bar{X}$， $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

$$E\left(\left(\hat{\sigma}^2-\sigma^2\right)^2\right)<E\left(\left(\bar{\sigma}^2-\sigma^2\right)^2\right)$$

均方误差(Mean square error，MSE)

例 5 设 $X_1,X_2,\cdots,X_n$ 是正态总体 $N\left(\mu,\sigma^2\right)$ 的样本，计算例 4 中关于参数 μ 和 σ^2 的矩估计量和极大似然估计量的 MSE。

参数 μ 和 σ^2 的矩估计量为 $\bar{\mu}=\bar{X}$，$\bar{\sigma}^2=S^2$

$$E\left(\left(\bar{\mu}-\mu\right)^2\right)=E\left(\left(\bar{X}-\mu\right)^2\right)=Var\left(\bar{X}\right)=\frac{\sigma^2}{n}$$

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

$$E\left(\left(\bar{\sigma}^2-\sigma^2\right)^2\right)=E\left(\left(S^2-\sigma^2\right)^2\right)\overset{E\left(S^2\right)=\sigma^2}{=}Var\left(S^2\right)=\frac{2\sigma^4}{n-1}$$

参数μ和σ^2的矩估计量为 $\bar{\mu}=\bar{X}$， $\bar{\sigma}^2=S^2$

参数μ和σ^2的极大似然估计量为$\hat{\mu}=\bar{X}$， $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{n-1}{n}S^2$

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

$$=\left(\frac{n-1}{n}\right)^2 E\left(\left(S^2-\sigma^2\right)^2\right)-2E\left(\frac{n-1}{n}\left(S^2-\sigma^2\right)\cdot\frac{1}{n}\sigma^2\right)+\frac{1}{n^2}\sigma^4$$

$$\overset{E\left(S^2\right)=\sigma^2}{=}\left(\frac{n-1}{n}\right)^2 Var\left(S^2\right)+\frac{1}{n^2}\sigma^4$$

$$=\left(\frac{n-1}{n}\right)^2\frac{2\sigma^4}{n-1}+\frac{1}{n^2}\sigma^4=\frac{2n-1}{n^2}\sigma^4$$

$$E\left(\left(\hat{\sigma}^2-\sigma^2\right)^2\right)=\frac{2n-1}{n^2}\sigma^4<\frac{2}{n-1}\sigma^4=E\left(\left(\bar{\sigma}^2-\sigma^2\right)^2\right)$$

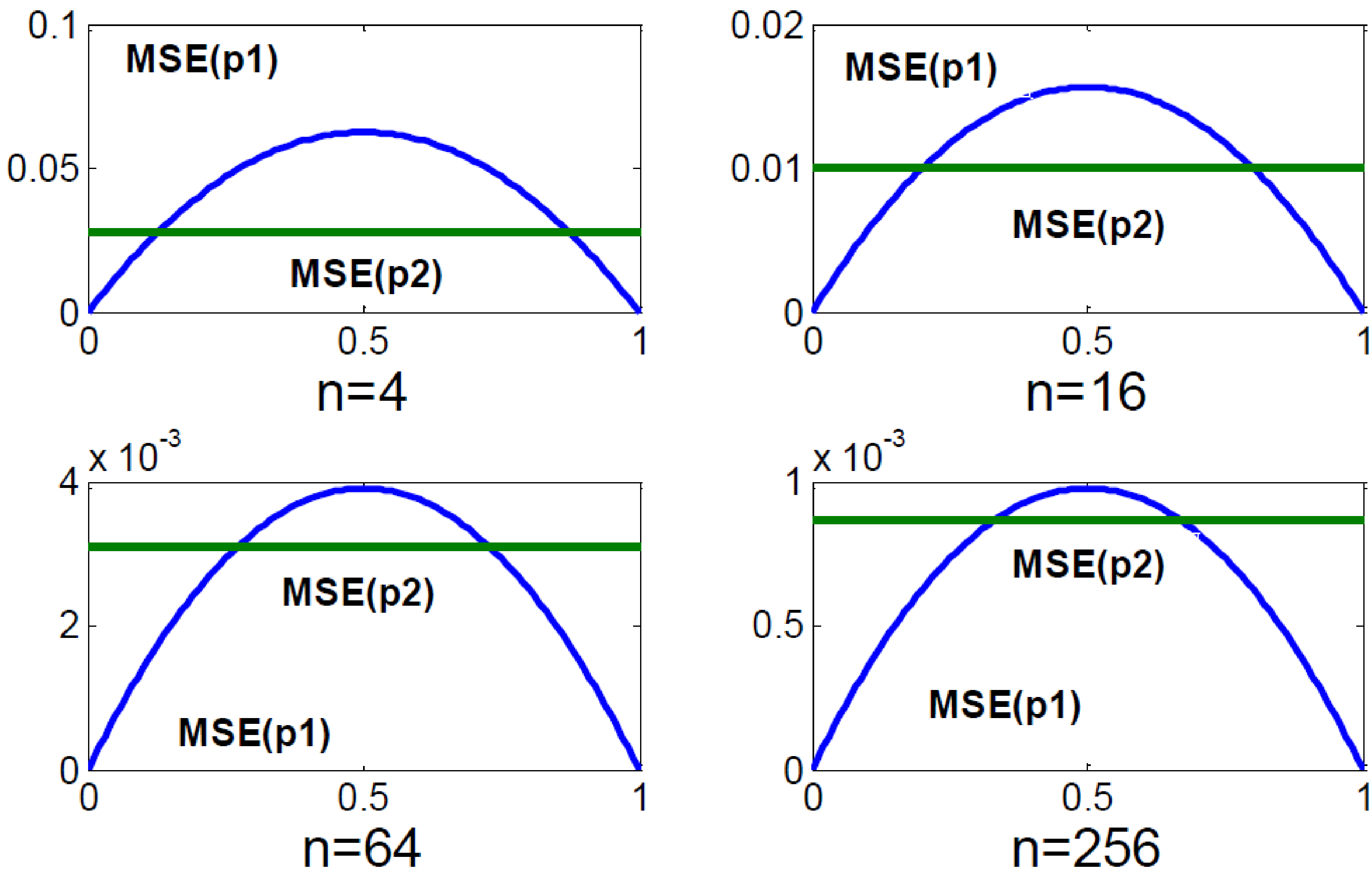
例 6 设 $X_1,X_2,\cdots,X_n$ 是二项总体 $b(1,p)$ 的样本，考虑如下两个关于参数 p 的估计量的 MSE，$\hat{p}_1=\bar{X}$，$\hat{p}_2=\dfrac{n\bar{X}+\alpha}{\alpha+\beta+n}$。

$$E\left(\left(\hat{p}_1-p\right)^2\right)=Var\left(\bar{X}\right)=\frac{p(1-p)}{n}$$

$$E\left(\left(\hat{p}_2-p\right)^2\right)=Var\left(\hat{p}_2\right)+\left(E\left(\hat{p}_2\right)-p\right)^2=Var\left(\frac{n\bar{X}+\alpha}{\alpha+\beta+n}\right)+\left(\frac{np+\alpha}{\alpha+\beta+n}-p\right)^2$$

$$=\frac{np(1-p)}{(\alpha+\beta+n)}+\left(\frac{np+\alpha}{\alpha+\beta+n}-p\right)^2$$

当 $\alpha=\beta=\dfrac{\sqrt{n}}{2}$ 时，$\hat{p}_2=\dfrac{n\bar{X}+\sqrt{n}/2}{n+\sqrt{n}}$，$E\left(\left(\hat{p}_2-p\right)^2\right)=\dfrac{n}{4\left(n+\sqrt{n}\right)^2}$


0.1
0.05
0
0
0.5
1
MSE(p1)
MSE(p2)
n=4
0.02
0.01
0
0
0.5
1
MSE(p1)
MSE(p2)
n=16
x 10-3
4
2
0
0
0.5
1
MSE(p2)
MSE(p1)
n=64
x 10-3
1
0.5
0
0
0.5
1
MSE(p2)
MSE(p1)
n=256

最小方差无偏估计
(Minimum Variance Unbiased, MVU)

定义 设 $\hat{\theta}$ 为 $g(\theta)$ 的无偏估计，若对 $g(\theta)$ 的任何无偏估计 $\hat{\theta}_1$，$Var_\theta(\hat{\theta}) \le Var_\theta(\hat{\theta}_1)$ 对 θ 的任何可能的值都成立，则称 $\hat{\theta}$ 为 $g(\theta)$ 的一个最小方差无偏估计（MVU 估计）。

克拉美-劳不等式（Cramer-Rao）满足一定条件，$T = T(x_1, x_2, \cdots, x_n)$ 是 $g(\theta)$ 的任一无偏估计，有 $Var(T) \ge \dfrac{(g'(\theta))^2}{nI(\theta)}$，其中 $I(\theta) = E\left(\left(\dfrac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2\right)$，称为 Fisher 信息量，$n$ 是样本容量。

克拉美-劳不等式成立条件

设总体概率函数 $p(x;\theta)$，$\theta\in\Theta$ 满足下列条件

1. 参数空间 Θ 是直线上的一个开区间；
2. 支撑集 $S=\{x:p(x;\theta)>0\}$ 与 θ 无关；
3. 对一切 $\theta\in\Theta$，$\frac{\partial}{\partial\theta}p(x;\theta)$ 存在；
4. 对 $p(x;\theta)$，积分与微分运算的次序可以交换，即

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}p(x;\theta)\,dx=\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}p(x;\theta)\,dx\text{；}$$

5. 期望 $E\left(\left(\frac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2\right)$ 存在

克拉美-劳不等式（Cramer-Rao）满足一定条件，$T=T(x_1,x_2,\cdots,x_n)$是$g(\theta)$的任一无偏估计，有$Var(T)\ge\frac{(g'(\theta))^2}{nI(\theta)}$，其中$I(\theta)=E\left(\left(\frac{\partial}{\partial\theta}\ln p(x;\theta)\right)^2\right)$，称为 Fisher 信息量，$n$是样本容量。

$$\int_{-\infty}^{+\infty}p(x_k;\theta)dx_k=1 \quad\Rightarrow\quad \frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}p(x_k;\theta)dx_k=\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}p(x_k;\theta)dx_k=0$$

$$\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}p(x_k;\theta)dx_k=\int_{-\infty}^{+\infty}\left[\frac{\partial}{\partial\theta}\ln p(x_k;\theta)\right]p(x_k;\theta)dx_k=E\left(\frac{\partial}{\partial\theta}\ln p(X_k;\theta)\right)=0$$

记 $Z=\frac{\partial}{\partial\theta}\ln\left(\prod_{k=1}^{n}p(X_k;\theta)\right)=\sum_{k=1}^{n}\frac{\partial}{\partial\theta}\ln p(X_k;\theta)$

则 $E(Z)=\sum_{k=1}^{n}E\left(\frac{\partial}{\partial\theta}\ln p(X_k;\theta)\right)=0$

$$E(Z^2)=Var(Z)=\sum_{k=1}^{n}Var\left(\frac{\partial}{\partial\theta}\ln p(X_k;\theta)\right)=\sum_{k=1}^{n}E\left(\left(\frac{\partial}{\partial\theta}\ln p(X_k;\theta)\right)^2\right)=nI(\theta)$$

$$g(\theta)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x_1,x_2,\cdots,x_n)\prod_{k=1}^{n}p(x_k;\theta)\,dx_1dx_2\cdots dx_n$$

$$g'(\theta)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x_1,x_2,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{k=1}^{n}p(x_k;\theta)\,dx_1dx_2\cdots dx_n$$

$$=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x_1,x_2,\cdots,x_n)\left[\frac{\partial}{\partial\theta}\ln\left(\prod_{k=1}^{n}p(x_k;\theta)\right)\right]\prod_{k=1}^{n}p(x_k;\theta)\,dx_1dx_2\cdots dx_n$$

$$=E\left(T(X_1,X_2,\cdots,X_n)\frac{\partial}{\partial\theta}\ln\left(\prod_{k=1}^{n}p(X_k;\theta)\right)\right)=E(T\cdot Z)$$

$$g'(\theta)=E(T\cdot Z)=E\big((T-g(\theta))\cdot Z\big)=Cov\big(T-g(\theta),Z\big)$$

$$[g'(\theta)]^2\le E\left((T-g(\theta))^2\right)\cdot E(Z^2)=Var(T)\cdot Var(Z)\quad\Rightarrow\quad Var(T)\ge\frac{(g'(\theta))^2}{nI(\theta)}$$

Fisher信息量

Fisher 信息量 ：

$$I(\theta)=E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2=E\left[\frac{\left(\frac{\partial}{\partial\theta}p(x;\theta)\right)}{p(x;\theta)}\right]^2=\int\left[\frac{\left(\frac{\partial}{\partial\theta}p(x;\theta)\right)^2}{p(x;\theta)}\right]dx\text{。}$$

Fisher 信息量$I(\theta)$越大，克拉美-劳不等式中的下界越低，表示$g(\theta)$的无偏估计更有可能达到较小的方差，即更有可能被估计的准确一些。$g(\theta)$的估计完全来自于样本，$g(\theta)$能估得更准，表示样本所含的信息量更大。n是样本数目，如把总信息看做分母$nI(\theta)$，则一个样本恰好占有信息量$I(\theta)$。$I(\theta)$越大意味着参数越容易估计，它反映模型“易于认识”的程度。这个量在统计学中有多方面重要的应用。

最小方差无偏估计（MVU）的判断

例 1. 设样本 $X_1, X_2, \cdots, X_n$ 来自指数总体 $Exp(\lambda)$，估计其期望 $g(\lambda)=\frac{1}{\lambda}$。

$$E(\bar{X})=g(\lambda)=\frac{1}{\lambda},\quad Var(\bar{X})=\frac{1}{n\lambda^2},\quad g(\lambda)=\frac{1}{\lambda}$$

$$(g'(\lambda))^2=\frac{1}{\lambda^4},$$

$$I(\lambda)=E\left(\left(\frac{\partial}{\partial\lambda}\ln p(X;\lambda)\right)^2\right)=E\left(\left(\frac{\partial}{\partial\lambda}\ln\left(\lambda e^{-\lambda X}\right)\right)^2\right)=E\left(\left(\frac{1}{\lambda}-X\right)^2\right)=\frac{1}{\lambda^2}$$

$$\frac{(g'(\lambda))^2}{nI(\lambda)}=\frac{1}{n\lambda^2}=Var(\bar{X})$$

最小方差无偏估计（MVU）的判断

例 2. 设样本 $X_1, X_2, \cdots, X_n$ 来自泊松总体 $P(\lambda)$，估计其期望 $g(\lambda)=\lambda$

$E(\bar{X})=\lambda,\quad Var(\bar{X})=\frac{\lambda}{n},\qquad E(S^2)=\lambda,\quad Var(S^2)=?$

$\left(g'(\lambda)\right)^2=1,$

$$I(\lambda)=E\left(\left(\frac{\partial}{\partial\lambda}\ln\left(e^{-\lambda}\frac{\lambda^X}{X!}\right)\right)^2\right)=E\left(\left(\frac{X}{\lambda}-1\right)^2\right)=\frac{1}{\lambda}$$

$$\frac{\left(g'(\lambda)\right)^2}{nI(\lambda)}=\frac{\lambda}{n}=Var(\bar{X})$$

最小方差无偏估计（MVU）的判断

例 3. 设样本 $X_1, X_2, \cdots, X_n$ 来自正态总体 $N(\mu, \sigma^2)$，σ^2 已知，估计 μ

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

$$I(\mu) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma^4} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}$$

按照克拉美-劳不等式，参数 μ 的无偏估计量方差的下界为 $\frac{(g'(\mu))^2}{nI(\mu)} = \frac{\sigma^2}{n}$，

$\bar{X}$ 是 μ 的无偏估计，方差恰好为 $Var(\bar{X}) = \frac{\sigma^2}{n}$，因此 $\bar{X}$ 是 μ 的 MVU 估计。

无偏估计的改善

定理　设X、Y是两个随机变量，$E(X)=\mu$，$Var(X)>0$，令

$\varphi(y)=E(X|\,Y=y)$，则有$E(\varphi(Y))=\mu$，$Var(\varphi(Y))\le Var(X)$，其中等号成立

的充分必要条件是X与$\varphi(Y)$几乎处处相等。

$E(X)=E\left(E(X|Y)\right)$，　$Var(X)=Var\left(E(X|Y)\right)+E\left(Var(X|Y)\right)$

其中等号成立的充分必要条件是X与$E(X|Y)$几乎处处相等。

Rao-Blackwell 定理 设总体的概率密度函数为$p(x;\theta)$，$x_1,x_2,\cdots,x_n$是其

样本。$T=T(x_1,\cdots,x_n)$是θ的充分统计量，则对θ的任意无偏估计

$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$，$\tilde{\theta}=E(\hat{\theta}|T)$也是$\theta$的无偏估计，且$Var(\tilde{\theta})\le Var(\hat{\theta})$。

充分统计量

- 统计量是把样本中的信息进行加工处理的结果，它可简化数据规律，便于统计推断。人们自然希望这种加工处理不损失原样本的信息。包含关于未知参数全部信息的统计量就叫充分统计量。

1920 年，R. A. Fisher 与天文学家 Eddington 之间有一场争论。

Eddington 用 $D=\sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{i=1}^{n}\left|X_i-\bar{X}\right|$ 估计正态总体 $N\left(\mu,\frac{\sigma^2}{n}\right)$ 的标准差 σ，Fisher

主张用 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 中的 S。Fisher 的理由是：S 包含了有关 σ 的全

部信息，而 D 则不是。因为只要算出 S 这个统计量的值，就是把原信息丢掉，

对参数 σ 的估计也无任何损失。

定义 设 $X_1,X_2,\cdots,X_n$ 是来自某个总体的样本，总体分布函数为 $F(x;\theta)$，统计量 $T=T(X_1,X_2,\cdots,X_n)$ 称为 θ 的充分统计量，如果在给定 T 的取值后，$X_1,X_2,\cdots,X_n$ 的条件分布与 θ 无关。

例 1. 设 $X_1,X_2,\cdots,X_n$ 来自总体 $b(1,p)$，则 $T=n\bar{X}$ 是参数 p 的充分统计量。

$$P\left(X_1=x_1,X_2=x_2,\cdots,X_n=x_n \middle| T=t\right)$$

$$=\frac{P\left(X_1=x_1,X_2=x_2,\cdots,X_n=t-\sum_{k=1}^{n-1}x_k\right)}{P(T=t)}=\frac{\prod_{k=1}^{n-1}P(X_k=x_k)\cdot P\left(X_n=t-\sum_{k=1}^{n-1}x_k\right)}{\binom{n}{t}p^t(1-p)^{n-t}}$$

$$=\frac{\prod_{k=1}^{n-1}p^{x_k}(1-p)^{x_k}\cdot p^{t-\sum_{k=1}^{n-1}x_k}(1-p)^{1-t+\sum_{k=1}^{n-1}x_k}}{\binom{n}{t}p^t(1-p)^{n-t}}=\frac{p^t(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}}=\frac{1}{\binom{n}{t}}$$

例 1. 设 $X_1, X_2, \cdots, X_n$ 来自总体 $b(1, p)$,

若 $T = X_1 + X_2$,

$$P\left(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n \middle| T = t\right)$$

$$= \frac{P\left(X_1 = x_1, X_2 = t - x_2, \cdots, X_n = x_n\right)}{P\left(X_1 + X_2 = t\right)}$$

$$= \frac{p^{t + x_3 + \cdots + x_n} (1 - p)^{n - t - (x_3 + \cdots + x_n)}}{\binom{2}{t} p^t (1 - p)^{2 - t}} = \frac{p^{x_3 + \cdots + x_n} (1 - p)^{n - 2 - (x_3 + \cdots + x_n)}}{\binom{2}{t}}$$

充分统计量的判定

因子分解定理 设总体的概率函数为 $F(x;\theta)$，$X_1,X_2,\cdots,X_n$ 为样本，则

$T=T(X_1,X_2,\cdots,X_n)$ 为 θ 的充分统计量的充分必要条件是：存在两个函数

$g(t,\theta)$和 $h(x_1,x_2,\cdots,x_n)$使得对任意的θ和任一组观察值$x_1,x_2,\cdots,x_n$，有

$f(x_1,x_2,\cdots,x_n;\theta)=g(T(x_1,x_2,\cdots,x_n),\theta)\cdot h(x_1,x_2,\cdots,x_n)$，其中 $g(t,\theta)$是通过统计量T的取值而依赖于样本的。（概率函数是指密度函数或概率分布列）

例 2. 设 $X_1, X_2, \cdots, X_n$ 来自正态总体 $N(\mu, \sigma^2)$，则 $(\bar{X}, S^2)$ 是参数 $\theta = (\mu, \sigma^2)$ 的充分统计量。

$$p(x_1, x_2, \cdots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right)$$

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

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right) \exp\left(-\frac{(n-1) s^2}{2\sigma^2}\right)$$

无偏估计的改善

例 1. 设 X_1, X_2 来自正态总体 $N(\mu, 1)$， $\bar{X} = \frac{X_1 + X_2}{2}$， $E(\bar{X}) = \mu$, $Var(\bar{X}) = \frac{1}{2}$

定义 $\varphi(X_1) = E(\bar{X} | X_1)$，则

$$E(E(\bar{X}| X_1)) = E(\bar{X}) = \mu \text{，} \quad Var(E(\bar{X}| X_1)) \le Var(\bar{X})$$

$$\varphi(X_1) = E(\bar{X}| X_1) = E\left(\frac{X_1 + X_2}{2} \middle| X_1\right)$$

$$= \frac{1}{2}E(X_1| X_1) + \frac{1}{2}E(X_2| X_1) = \frac{1}{2}X_1 + \frac{1}{2}\mu$$

无偏估计的改善

例 2. 设有某商场顾客到来服从泊松分布，即每分钟到达的人数服从参数为 λ 泊松分布，$X_1, X_2, \cdots, X_n$ 为一个样本，分别记录连续 n 分钟观察到的每分钟顾客到达的人数。希望估计每分钟没有顾客到来的概率 $e^{-\lambda}$。

$\theta_0 = \begin{cases} 1 & \text{if } x_1 = 0 \\ 0 & \text{otherwise} \end{cases}$ 为参数 $e^{-\lambda}$ 的一个无偏估计量，

$S_n = \sum_{k=1}^{n} X_k$ 是 λ 的充分统计量，

取充分统计量下的条件期望保证仍然是统计量

$\theta_1 = E(\theta_0 | S_n) = \left(1 - \frac{1}{n}\right)^{S_n}$ 为 $e^{-\lambda}$ 改进的无偏估计量。

作业

习题二

15，18，19

其中19题第一问利用讲义方法构造并给出改进

补充题，已知 $X_1, X_2, \cdots, X_n$为总体$X \sim Exp(\lambda)$的样本，$E(X)=\frac{1}{\lambda}$。

$\overline{X}$为样本均值，用$\frac{1}{\overline{X}}$估计参数λ是否为无偏估计，给出解释。