§3.5 Cramér-Rao 不等式—无偏估计方差的下界

Fisher Information: 设随机变量(或向量)X来自分布族$\mathcal{F}=\{p(x;\theta):\theta\in\Theta\}$,其中$p(x;\theta)$为其p.d.f., Θ是开区间, 假设$p(x;\theta)$关于θ可导, 且

$$
\begin{aligned}
0 &= \frac{d}{d\theta}\int_{-\infty}^{\infty} p(x;\theta)dx = \int_{-\infty}^{\infty} \frac{\partial p(x;\theta)}{\partial\theta}dx \\
&= \int_{-\infty}^{\infty} \frac{\partial \log p(x;\theta)}{\partial\theta} p(x;\theta)dx = \mathsf{E}_\theta\left[\frac{\partial \log p(X;\theta)}{\partial\theta}\right].
\end{aligned}
$$

从而

$$
\begin{aligned}
I(\theta) :=& \mathsf{Var}_\theta\left\{\frac{\partial \log p(X;\theta)}{\partial\theta}\right\} = \mathsf{E}_\theta\left[\frac{\partial \log p(X;\theta)}{\partial\theta}\right]^2 \\
=& \int_{-\infty}^{\infty}\left(\frac{\partial \log p(x;\theta)}{\partial\theta}\right)^2 p(x;\theta)dx.
\end{aligned}
$$

$I(\theta)$称为X或分布族$\mathcal{F}$的Fisher Information.

上述隐含了五个条件:

(i) 参数空间Θ是直线上的开区间;

(ii) 导数$\frac{\partial}{\partial\theta}p(x;\theta)$对一切$\theta\in\Theta$都存在;

(iii) 支撑$\{x:p(x;\theta)>0\}$不依赖于θ;

(iv) $p(x;\theta)$的积分和求导可以交换，即

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

(v) 信息函数存在$0<I(\theta)<\infty$.

这些条件称为Cramér-Rao正则条件.将满足这些条件的分布族$\mathcal{F}=\{p(x;\theta):\theta\in\Theta\}$称为Cramér-Rao正则族.

如果$\frac{\partial^2}{\partial\theta^2}p(x;\theta)$对任意的$\theta\in\Theta$都存在,且积分与求导可以交换, 则

$$0=\frac{d^2}{d\theta^2}\int_{-\infty}^{\infty}p(x;\theta)dx=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\left[\frac{\partial\log p(x;\theta)}{\partial\theta}p(x;\theta)\right]dx$$
$$=\int_{-\infty}^{\infty}\frac{\partial^2\log p(x;\theta)}{\partial\theta^2}p(x;\theta)dx+\int_{-\infty}^{\infty}\left[\frac{\partial\log p(x;\theta)}{\partial\theta}\right]^2p(x;\theta)dx.$$

从而

$$I(\theta)=-\mathsf{E}_\theta\left[\frac{\partial^2\log p(X;\theta)}{\partial\theta^2}\right]----\text{计算时较简单}.$$

Example

设$X \sim N(\mu, \sigma^2)$, σ^2已知. 求Fisher信息函数$I(\mu)$.

解: 由于$p(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, 故

$$\log p(x;\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}.$$

从而$\frac{\partial \log p(x;\mu)}{\partial \mu} = \frac{x-\mu}{\sigma^2}$, 那么

$$I(\mu) = \mathsf{E}\left(\frac{\partial \log p(X;\mu)}{\partial \mu}\right)^2 = \mathsf{E}\left(\frac{X-\mu}{\sigma^2}\right)^2 = \frac{1}{\sigma^4}\mathsf{E}(X-\mu)^2 = \frac{1}{\sigma^2}.$$

或者通过$\frac{\partial^2 \log p(x;\mu)}{\partial \mu^2} = -\frac{1}{\sigma^2}$, 得$I(\mu) = -\mathsf{E}\left(\frac{\partial^2 \log p(X;\mu)}{\partial^2 \mu}\right) = \frac{1}{\sigma^2}$.

Fisher信息量的含义

回忆: 极大似然估计的渐近正态性(仅考虑θ为一维的情形)

Theorem

定理设$\mathcal{F}=\{p(x;\theta):\theta\in\Theta\}$是一个概率密度或分布列族，$\Theta$为直线上的非退化区间，$X_1,X_2,\cdots,X_n$是从$\mathcal{F}$中某个总体$X\sim p(x;\theta_0)$产生的样本．当$p(x;\theta)$满足适当正则条件*(regularity condition)*时，则在参数θ的未知真值θ_0为Θ的一个内点的情况下,其似然方程有一个解,记为$\widehat{\theta}_n$ 满足: $\widehat{\theta}_n$依概率收敛于真值θ_0，且

$$\sqrt{n}(\widehat{\theta}_n-\theta_0)\xrightarrow{D}N\left(0,\frac{1}{I(\theta_0)}\right),\quad 当\quad n\to\infty.$$

其中

$$I(\theta)=E_\theta\left[\left(\frac{\partial\log p(X;\theta)}{\partial\theta}\right)^2\right].$$

样本的Fisher信息量 如果总体X来自分布族$\mathcal{F}=\{p(x;\theta):\theta\in\Theta\}$，其Fisher信息量存在，记为$I(\theta)$. $\widetilde{X}=(X_1,X_2,\ldots,X_n)$是来自X的简单随机样本，其联合pdf为$p(\widetilde{x};\theta)$的，那么$\widetilde{X}$的Fisher Information 为

$$I_n(\theta)=\mathsf{Var}_\theta\left\{\frac{\partial\log p(\widetilde{X};\theta)}{\partial\theta}\right\}=\mathsf{Var}_\theta\left\{\sum_{i=1}^n\frac{\partial\log p(X_i;\theta)}{\partial\theta}\right\}$$

$$\overset{\text{独立性}}{=}\sum_{i=1}^n\mathsf{Var}_\theta\left\{\frac{\partial\log p(X_i;\theta)}{\partial\theta}\right\}\overset{\text{代表性}}{=}nI(\theta).$$

Fisher信息量的用途

Theorem

定理3.5.1 (*Cramér-Rao 不等式*) 设总体X来自分布族$\{p(x;\theta),\theta\in\Theta\}$，样本$\widetilde{X}=(X_1,X_2,\ldots,X_n)$的联合密度函数为$p(\widetilde{x};\theta)$，$\widehat{g}=\widehat{g}(\widetilde{X})$是$g(\theta)$无偏估计(其中$g(\cdot)$可微). 假设

$$\frac{d}{d\theta}\int p(\widetilde{x};\theta)d\widetilde{x}=\int\frac{\partial}{\partial\theta}p(\widetilde{x};\theta)d\widetilde{x},$$

$$g'(\theta)=\frac{d}{d\theta}E_\theta\widehat{g}=\int\frac{\partial}{\partial\theta}\widehat{g}(\widetilde{x})p(\widetilde{x};\theta)d\widetilde{x},$$

且$Var_\theta\{\widehat{g}\}<\infty$. 则

$$Var_\theta\{\widehat{g}\}\geq\frac{(g'(\theta))^2}{E_\theta\left[\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right]^2}.$$

此时$\frac{(g'(\theta))^2}{E_\theta\left[\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right]^2}$称为参数$g(\theta)$的无偏估计方差的Cramér-Rao下界(简称C-R下界).

证明:

$$
\begin{aligned}
g'(\theta) =& \frac{d}{d\theta}\mathsf{E}_\theta \widehat{g} = \int \widehat{g}(\widetilde{x})\frac{\partial}{\partial\theta}p(\widetilde{x};\theta)d\widetilde{x} \\
=& \mathsf{E}_\theta\left[\widehat{g}(\widetilde{X})\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right].
\end{aligned}
$$

注意到

$$
\mathsf{E}_\theta\left[\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right] = \int \frac{\partial}{\partial\theta}p(\widetilde{x};\theta)d\widetilde{x} = \frac{d}{d\theta}\int p(\widetilde{x};\theta)d\widetilde{x} = 0,
$$

所以

$$
g'(\theta) = \mathsf{Cov}_\theta\left\{\widehat{g}(\widetilde{X}), \frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right\}.
$$

由Cauchy-Schwarz不等式得

$$\begin{aligned}\{g'(\theta)\}^2 \leq & \mathsf{Var}_\theta\{\widehat{g}(\widetilde{X})\} \cdot \mathsf{Var}_\theta\left\{\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right\} \\ = & \mathsf{Var}_\theta\{\widehat{g}(\widetilde{X})\} \cdot \mathsf{E}_\theta\left\{\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right\}^2.\end{aligned}$$

结论得证.

注： *Cauchy-Schwarz*不等式中等号成立当且仅当存在不全为零的$\alpha(\theta)$, $\beta(\theta)$, $\gamma(\theta)$使得

$$P_\theta\left\{\alpha(\theta)\widehat{g}(\widetilde{X})+\gamma(\theta)=\beta(\theta)\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right\}=1.$$

注意到$E_\theta(\widehat{g}(\widetilde{X}))=g(\theta)$, $E_\theta(\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta))=0$, 因此可得$\gamma(\theta)=-\alpha(\theta)g(\theta)$, 这也意味着*Cauchy-Schwarz*不等式中等号成立当且仅当存在不全为零的$\alpha(\theta)$, $\beta(\theta)$使得

$$P_\theta\left\{\alpha(\theta)\left[\widehat{g}(\widetilde{X})-g(\theta)\right]=\beta(\theta)\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right\}=1.$$

Corollary

设$\widetilde{X}=(X_1,X_2,\ldots,X_n)$为取自总体$X\sim p(x;\theta)$的$i.i.d.$样本，$p(x,\theta)$为总体的$pdf$ 满足条件

$$\frac{d}{d\theta}\int p(x;\theta)dx=\int\frac{\partial}{\partial\theta}p(x;\theta)dx.$$

又设$\widehat{g}=\widehat{g}(\widetilde{X})$是一个估计量，$g(\theta)=E_\theta\widehat{g}$，满足条件

$$g'(\theta)=\frac{d}{d\theta}E_\theta\widehat{g}=\int\frac{\partial}{\partial\theta}\left[\widehat{g}(\widetilde{x})\prod_{i=1}^{n}p(x_i;\theta)\right]d\widetilde{x}$$

且$Var_\theta\{\widehat{g}\}<\infty$. 则

$$Var_\theta\{\widehat{g}\}\geq\frac{(g'(\theta))^2}{nE_\theta\left[\frac{\partial}{\partial\theta}\log p(X;\theta)\right]^2}.$$

证明: 这时样本的联合pdf为

$$p(\widetilde{x};\theta)=\prod_{i=1}^{n}p(x_i;\theta).$$

易知

$$\begin{aligned}&\int\frac{\partial}{\partial\theta}p(\widetilde{x};\theta)d\widetilde{x}=\mathsf{E}_\theta\left[\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right]\\ =&\sum_{i=1}^{n}\mathsf{E}_\theta\left[\frac{\partial}{\partial\theta}\log p(X_i;\theta)\right]=n\mathsf{E}_\theta\left[\frac{\partial}{\partial\theta}\log p(X;\theta)\right]\\ =&n\int\frac{\partial}{\partial\theta}p(x;\theta)dx=n\frac{d}{d\theta}\int p(x;\theta)dx=0.\end{aligned}$$

定理中的条件满足.

又

$$
\begin{aligned}
& \mathsf{E}_\theta\left[\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right]^2 = \mathsf{Var}_\theta\left\{\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right\} \\
=& \mathsf{Var}_\theta\left\{\sum_{i=1}^n\frac{\partial}{\partial\theta}\log p(X_i;\theta)\right\} = n\mathsf{Var}_\theta\left\{\frac{\partial}{\partial\theta}\log p(X;\theta)\right\} \\
=& n\mathsf{E}_\theta\left[\frac{\partial}{\partial\theta}\log p(X;\theta)\right]^2.
\end{aligned}
$$

由定理, 推论得证.

Example

设$X_1, X_2, \ldots, X_n$是取自Poisson分布$P(\lambda)$总体的样本，求参数λ的无偏估计方差的C-R下界.

解：对于$x = 0, 1, \ldots$，有

$$\log p(x; \lambda) = x \log \lambda - \log(x!) - \lambda.$$

$$\frac{\partial \log p}{\partial \lambda} = \frac{x}{\lambda} - 1, \quad \frac{\partial^2 \log p}{\partial \lambda^2} = -\frac{x}{\lambda^2}.$$

所以总体的Fisher 信息量为

$$I(\lambda) = -\mathsf{E}_\lambda \left[\frac{\partial^2 \log p}{\partial \lambda^2} \right] = \frac{1}{\lambda^2} \mathsf{E}_\lambda X = \frac{1}{\lambda}.$$

从而C-R下界为$\frac{1^2}{n/\lambda} = \frac{\lambda}{n}$. 无偏估计$\overline{X}$的方差达到了这个下界.

Corollary

如果C-R定理中的条件满足，则达到C-R下界的无偏估计一定是$UMVUE$.

Example

设$X_1, X_2, \ldots, X_n$是取自正态总体$N(\mu, \sigma^2)$(σ已知)的样本，求参数μ的无偏估计方差的C-R下界.

解：由于$\log p(x;\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$，故

$$\frac{\partial \log p(x;\mu)}{\partial \mu} = \frac{1}{\sigma^2}(x-\mu).$$

$$\frac{\partial^2 \log p(x;\mu)}{\partial \mu^2} = -\frac{1}{\sigma^2}.$$

所以总体的Fisher信息量为

$$I(\mu) = -\mathsf{E}\left(\frac{\partial^2 \log p(X;\mu)}{\partial \mu^2}\right) = \frac{1}{\sigma^2}.$$

从而参数μ的无偏估计的C-R下界为σ^2/n. 无偏估计$\overline{X}$的方差达到了这个下界，因此参数μ的UMVUE为$\overline{X}$.

Example

设$X_1, X_2, \ldots, X_n$是取自正态总体$N(\mu, \sigma^2)$(μ已知)的样本，求参数σ^2的无偏估计方差的C-R下界.

解：由于$\log p = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2$，故

$$\frac{\partial \log p}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(x-\mu)^2.$$

$$\frac{\partial^2 \log p}{\partial (\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-\mu)^2.$$

所以总体的Fisher信息量为

$$I(\sigma^2) = -\left[\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\mathsf{E}(X-\mu)^2\right] = \frac{1}{2\sigma^4}.$$

从而参数σ^2的无偏估计的C-R下界为$2\sigma^4/n$.

考察无偏估计

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,\quad \widehat{\sigma_2^2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2.$$

易知

$$\mathsf{Var}\{S^2\}=\frac{2\sigma^4}{n-1},\quad \mathsf{Var}\{\widehat{\sigma_2^2}\}=\frac{2\sigma^4}{n}.$$

$\widehat{\sigma_2^2}$的方差达到了C-R下界，而S^2的没有达到C-R下界.

当μ已知时，$\widehat{\sigma_2^2}$是σ^2的UMVUE.

Example

设$X_1, X_2, \ldots, X_n$是取自均匀总体$U(0,\theta)$的一个样本.

这时

$$p(x;\theta) = \theta^{-1}, \quad 0 \le x \le \theta.$$

$$\frac{\partial \log p(x;\theta)}{\partial \theta} = -\theta^{-1}, \quad 0 \le x \le \theta.$$

所以

$$I(\theta) = \mathsf{E}\left(\frac{\partial \log p(X;\theta)}{\partial \theta}\right)^2 = \int_0^{\theta} \left(-\theta^{-1}\right)^2 \cdot \theta^{-1} dx = \theta^{-2}.$$

C-R下界为

$$\frac{\theta^2}{n}.$$

前面也已经证明$\widehat{\theta} = (1 + \frac{1}{n})X_{(n)}$是$\theta$的UMVUE. 但

$$\mathsf{Var}\{\widehat{\theta}\} = \frac{\theta^2}{n(n+1)} < \frac{\theta^2}{n} = \text{C-R lower bound.}$$

问题出在哪里?

$$\frac{d}{d\theta}\int p(x;\theta)dx = \frac{d}{d\theta}\int_0^{\theta} p(x;\theta)dx = 0$$

但

$$\int \frac{\partial}{\partial\theta} p(x;\theta)dx = \int_0^{\theta} -\theta^{-2}dx = -\theta^{-1} \neq 0$$

求导与求积分不能交换,不满足Cramér-Rao定理中的条件.

一般地,如果支撑$\{x : p(x;\theta) > 0\}$与参数θ有关,则C-R定理的条件不满足,C-R不等式也不成立.

很多参数的无偏估计的C-R下界是达不到的.

Example

设$X_1, \ldots, X_n$ 是取自Poisson分布$P(\lambda)$的简单随机样本，分布列为

$$p(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}, x = 0, 1, 2, \ldots,$$

$\lambda > 0$ 未知. 总体的Fisher信息量为

$$I(\lambda) = \mathsf{E}_\lambda\left[\frac{\partial}{\partial\lambda}\log p(X;\lambda)\right]^2 = \mathsf{E}_\lambda\left[\frac{X}{\lambda} - 1\right]^2 = \frac{1}{\lambda}.$$

若$g(\lambda) = \lambda$, 则$g'(\lambda) = 1$, 于是λ的无偏估计的C-R下界为$\frac{1}{nI(\lambda)} = \frac{\lambda}{n}$. $\widehat{\lambda} = \overline{X}$ 达到了这个下界.

若$g(\lambda)=e^{-\lambda}$, 则$g'(\lambda)=-e^{-\lambda}$, 于是$e^{-\lambda}$的无偏估计的C-R下界为

$$\frac{[g'(\lambda)]^2}{nI(\lambda)}=\frac{\lambda e^{-2\lambda}}{n},$$

这个下界是不能达到的.

事实上$e^{-\lambda}=\mathsf{P}(X=0)$的UMVUE为$\widehat{g}(\widetilde{X})=\left(1-\frac{1}{n}\right)^{T_n}$, 这里$T_n=X_1+\ldots+X_n$. 下面我们求$\mathsf{Var}\{\widehat{g}(\widetilde{X})\}$.

$$\begin{aligned}\mathsf{Var}\{\widehat{g}(\widetilde{X})\}=&\mathsf{E}\{\widehat{g}(\widetilde{X})\}^2-\{\mathsf{E}\widehat{g}(\widetilde{X})\}^2\\=&\mathsf{E}\left(1-\frac{1}{n}\right)^{2T_n}-\left\{\mathsf{E}\left(1-\frac{1}{n}\right)^{T_n}\right\}^2.\end{aligned}$$

设$Y \sim P(\theta)$, $a > 0$, 则

$$\mathsf{E}a^Y = \sum_{k=0}^{\infty} a^k \frac{\theta^k}{k!} e^{-\theta} = e^{-\theta + a\theta} \sum_{k=0}^{\infty} e^{-a\theta} \frac{(a\theta)^k}{k!} = e^{-\theta(1-a)}.$$

由于$T_n \sim P(n\lambda)$, 所以

$$\begin{aligned} \mathsf{Var}\{\widehat{g}(\widetilde{X})\} =& \mathsf{E}\left(1 - \frac{1}{n}\right)^{2T_n} - \left\{\mathsf{E}\left(1 - \frac{1}{n}\right)^{T_n}\right\}^2 \\ =& e^{-n\lambda\left[1-\left(1-\frac{1}{n}\right)^2\right]} - \left\{e^{-n\lambda\left[1-(1-\frac{1}{n})\right]}\right\}^2 \\ =& e^{-\lambda\left(2-\frac{1}{n}\right)} - e^{-2\lambda} = e^{-2\lambda}\left(e^{\frac{\lambda}{n}} - 1\right) \\ >& \lambda e^{-2\lambda}/n. \end{aligned}$$

C-R不等式中等号成立的条件

Theorem

定理3.5.2. 设总体X来自的分布族$\{p(x;\theta);\ \theta\in\Theta\}$满足$C$-$R$正则条件$(i)$-$(v)$，其中$p(x;\theta)$为其$p.d.f.$，取自该总体的简单随机样本$\widetilde{X}=(X_1,X_2,\ldots,X_n)$的联合$p.d.f.$记为$p(\widetilde{x};\theta)$. 可估参数$g(\theta)$ $(\theta\in\Theta)$ 不恒为常数. 若存在$g(\theta)$的无偏估计$\widehat{g}=\widehat{g}(\widetilde{X})$，使得$\mathsf{Var}_\theta\{\widehat{g}\}=(g'(\theta))^2/I_n(\theta)$，

$$g'(\theta)=\frac{d}{d\theta}\mathsf{E}_\theta\widehat{g}=\int\frac{\partial}{\partial\theta}\widehat{g}(\widetilde{x})p(\widetilde{x};\theta)d\widetilde{x},$$

则$p(\widetilde{x};\theta)$可以表示为下列形式

$$p(\widetilde{x};\theta)=C(\theta)e^{\psi(\theta)\widehat{g}(\widetilde{x})}h(\widetilde{x}),\quad \psi(\theta),C(\theta)\text{可微}.$$

且这时必有，$\psi'(\theta) \neq 0$,

$$g(\theta) = -\frac{1}{\psi'(\theta)} \frac{C'(\theta)}{C(\theta)}.$$

证明: 不妨设分布族中分布的支撑为$(-\infty,+\infty)$. 由于C-R不等式中取等号，必在不全为零的$\alpha(\theta)$, $\beta(\theta)$使得

$$\mathsf{P}_\theta\left\{\alpha(\theta)\left[\widehat{g}(\widetilde{X})-g(\theta)\right]=\beta(\theta)\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)\right\}=1.$$

记

$$A_\theta=\left\{\widetilde{x}:\alpha(\theta)\left[\widehat{g}(\widetilde{x})-g(\theta)\right]\neq\beta(\theta)\frac{\partial}{\partial\theta}\log p(\widetilde{x};\theta)\right\}.$$

则

$$\int_{A_\theta}p(\widetilde{x};\theta)d\widetilde{x}=\mathsf{P}_\theta\left\{\widetilde{X}\in A_\theta\right\}=0.$$

从而

$$\alpha(\theta)\left[\widehat{g}(\widetilde{x})-g(\theta)\right]=\beta(\theta)\frac{\partial}{\partial\theta}\log p(\widetilde{x};\theta),\quad a.e. ----(1)$$

易见$\alpha(\theta)$,$\beta(\theta)$均不为0.

事实上，若$\alpha(\theta)=0$, $\beta(\theta)\neq 0$, 那么$\frac{\partial}{\partial\theta}\log p(\widetilde{x};\theta)=0,\ a.e.$,

从而$I(\theta)=0$, 与正则条件(v)矛盾.

若$\beta(\theta)=0$, $\alpha(\theta)\neq 0$, 那么$\widehat{g}(\widetilde{x})=g(\theta)\ \ a.e.$ 故

$$g(\theta')=\mathsf{E}_{\theta'}(\widehat{g}(\widetilde{X}))=\mathsf{E}_{\theta'}(g(\theta))=g(\theta),\quad 对\forall\theta'\in\Theta成立,$$

与$g(\theta)$在Θ上不恒为常数矛盾.

所以$\alpha(\theta)$, $\beta(\theta)$均不为0. 因此(1)可改写为

$$\frac{\partial}{\partial\theta}\log p(\widetilde{x};\theta)=\frac{\alpha(\theta)}{\beta(\theta)}\left(\widehat{g}(\widetilde{x})-g(\theta)\right),\quad a.e. ----(2)$$

因此

$$\begin{aligned}\log p(\widetilde{x};\theta) &= \int \frac{\alpha(\theta)}{\beta(\theta)}\big(\widehat{g}(\widetilde{x}) - g(\theta)\big)d\theta + q(\widetilde{x})\\ &= -\int \frac{\alpha(\theta)}{\beta(\theta)}g(\theta)d\theta + \int \frac{\alpha(\theta)}{\beta(\theta)}d\theta \cdot \widehat{g}(\widetilde{x}) + q(\widetilde{x}),\end{aligned}$$

记$C(\theta) = \exp\left\{-\int \frac{\alpha(\theta)}{\beta(\theta)}g(\theta)d\theta\right\}$, $\psi(\theta) = \int \frac{\alpha(\theta)}{\beta(\theta)}d\theta$, $h(\widetilde{x}) = \exp\{q(\widetilde{x})\}$, 则可得

$$p(\widetilde{x};\theta) = C(\theta)e^{\psi(\theta)\widehat{g}(\widetilde{x})}h(\widetilde{x}).$$

显然$\psi'(\theta)=\alpha(\theta)/\beta(\theta)\neq 0$. 由$p(\widetilde{x};\theta)=C(\theta)e^{\psi(\theta)\widehat{g}(\widetilde{x})}h(\widetilde{x})$得

$$\frac{\partial}{\partial\theta}\log p(\widetilde{X};\theta)=\frac{C'(\theta)}{C(\theta)}+\psi'(\theta)\widehat{g}(\widetilde{X}).$$

上式两边取数学期望得

$$0=\frac{C'(\theta)}{C(\theta)}+\psi'(\theta)g(\theta).$$

即

$$g(\theta)=-\frac{1}{\psi'(\theta)}\frac{C'(\theta)}{C(\theta)}.$$

反过来

Theorem

定理3.5.3. 设样本$\tilde{X}=(X_1,X_2,\ldots,X_n)$的联合密度函数可以表示为$p(\tilde{x};\theta)=C(\theta)e^{\psi(\theta)U(\tilde{x})}h(\tilde{x})$，其中$C(\theta)>0$，$\psi(\theta)$是连续可微函数，且$\psi'(\theta)\neq 0$. 则当且仅当

$$g(\theta)=-\alpha\frac{1}{\psi'(\theta)}\frac{C'(\theta)}{C(\theta)}+\beta=\alpha \mathsf{E}_\theta(U(\tilde{X}))+\beta$$

时，才有达到C-R下界的无偏估计$\widehat{g}(\tilde{X})$，且

$$\mathsf{P}_\theta\left\{\widehat{g}(\tilde{X})=\alpha U(\tilde{X})+\beta\right\}=1,$$

其中，α，β 是与θ无关的两个常数.

证明： 这时

$$\frac{\partial \log p(\widetilde{X};\theta)}{\partial \theta} = \frac{C'(\theta)}{C(\theta)} + \psi'(\theta)U(\widetilde{X}). \tag{1}$$

上式两边取数学期望得

$$0 = \frac{C'(\theta)}{C(\theta)} + \psi'(\theta)E_\theta[U(\widetilde{X})]. \tag{2}$$

即

$$E_\theta[U(\widetilde{X})] = -\frac{1}{\psi'(\theta)}\frac{C'(\theta)}{C(\theta)}. \qquad ------------(*)$$

将(1), (2)相减得

$$\frac{\partial \log p(\widetilde{X};\theta)}{\partial \theta} = \psi'(\theta)\left[U(\widetilde{X}) - E_\theta[U(\widetilde{X})]\right]. \tag{3}$$

从而

$$I_n(\theta) = \mathsf{Var}\left(\frac{\partial \log p(\widetilde{X};\theta)}{\partial \theta}\right) = nI(\theta) = (\psi'(\theta))^2 \mathsf{Var}\{U(\widetilde{X})\}, \tag{4}$$

$$\frac{d}{d\theta}\mathsf{E}_\theta[U(\widetilde{X})] = \mathsf{E}_\theta\left[(U(\widetilde{X}) - E_\theta[U(\widetilde{X})])\frac{\partial \log p(\widetilde{X};\theta)}{\partial \theta}\right] = \psi'(\theta)\mathsf{Var}\{U(\widetilde{X})\}. \tag{5}$$

所以

$$\mathsf{Var}\{U(\widetilde{X})\} = \frac{\left(d\mathsf{E}_\theta[U(\widetilde{X})]/d\theta\right)^2}{nI(\theta)}.$$

即$U(\widetilde{X})$ 为$-\frac{1}{\psi'(\theta)}\frac{C'(\theta)}{C(\theta)}$的无偏估计(根据(*))，并且达到了C-R下界.

因此对一般情形，若$\mathsf{P}_\theta\left\{\widehat{g}(\widetilde{X})=\alpha U(\widetilde{X})+\beta\right\}=1$，则$\widehat{g}(\widetilde{X})$是

$$g(\theta)=\alpha\left[-\frac{1}{\psi'(\theta)}\frac{C'(\theta)}{C(\theta)}\right]+\beta$$

的无偏估计,并且达到了C-R下界.

反过来，假设$\widehat{g}(\widetilde{X})$是$g(\theta)$的达到C-R下界的无偏估计，则由定理3.5.2的证明知，存在$k(\theta)\neq 0$ 使得

$$\widehat{g}(\widetilde{x})-g(\theta)=k(\theta)\frac{\partial\log p(\widetilde{x};\theta)}{\partial\theta}.$$

从而

$$\widehat{g}(\widetilde{x})=k(\theta)\psi'(\theta)U(\widetilde{x})+g(\theta)+k(\theta)\frac{C'(\theta)}{C(\theta)}.$$

由于$\widehat{g}(\widetilde{X})$为统计量，应与$\theta$无关，

所以可以令

$$k(\theta)\psi'(\theta) = \alpha,$$

$$g(\theta) + k(\theta)\frac{C'(\theta)}{C(\theta)} = \beta,$$

其中α, β为与θ无关的常数. 故

$$g(\theta) = -\alpha\frac{1}{\psi'(\theta)}\frac{C'(\theta)}{C(\theta)} + \beta,$$

$$\mathsf{P}_\theta\left\{\widehat{g}(\widetilde{X}) = \alpha U(\widetilde{X}) + \beta\right\} = 1.$$

定理的必要性得到证明.

Example

$X_1, \ldots, X_n$ 是取自Poisson分布$P(\lambda)(\lambda > 0)$的样本, 联合分布列为

$$p(\tilde{x}; \lambda) = \frac{\lambda^{x_1+\ldots+x_n}}{x_1! \ldots x_n!} e^{-n\lambda} I\{x_i = 0, 1, \ldots, , i = 1, 2, \ldots, n\}$$
$$= e^{-n\lambda} e^{(\log \lambda) U(\tilde{x})} \frac{I\{x_i = 0, 1, \ldots, i = 1, 2, \ldots, n\}}{x_1! \ldots x_n!}.$$

其中$U(\tilde{x}) = x_1 + \ldots + x_n$.

根据定理, 知只有型如

$$\alpha \mathsf{E}_\lambda U(\tilde{X}) + \beta = \alpha \mathsf{E}_\lambda (X_1 + \ldots + X_n) + \beta = n\lambda\alpha + \beta$$

的参数才存在达到C-R下界的无偏估计

$$\alpha(X_1 + \ldots + X_n) + \beta = n\alpha\overline{X} + \beta.$$

Example

$X_1, \ldots, X_n$ 是取自正态分布$\{N(\mu, 1)(-\infty < \mu < \infty\}$的样本, 其联合密度函数为

$$p(\widetilde{x}; \mu) = (2\pi)^{-n/2} \exp\left\{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2}\right\}$$
$$= (2\pi)^{-n/2} \exp\left\{-\frac{n}{2}\mu^2\right\} \exp\left\{\mu \sum_{i=1}^{n} x_i\right\} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\}.$$

$U(\widetilde{x}) = x_1 + \ldots + x_n$. 根据定理, 知只有型如

$$\alpha \mathsf{E}_\mu U(\widetilde{X}) + \beta = \alpha \mathsf{E}_\mu (X_1 + \ldots + X_n) + \beta = n\mu\alpha + \beta$$

的参数才存在达到C-R下界的无偏估计

$$\alpha(X_1 + \ldots + X_n) + \beta = n\alpha\overline{X} + \beta.$$

Example

$X_1, \ldots, X_n$ 是取自指数分布$E(\lambda)(\lambda > 0)$的样本，其联合密度函数为

$$p(\widetilde{x}; \lambda) = \lambda^n \exp\left\{-\lambda \sum_{i=1}^n x_i\right\} I\{\min x_i > 0\}.$$

$U(\widetilde{x}) = x_1 + \ldots + x_n$. 根据定理，知只有型如

$$\alpha \mathsf{E}_\lambda U(\widetilde{X}) + \beta = \alpha \mathsf{E}_\lambda (X_1 + \ldots + X_n) + \beta = n\lambda\alpha + \beta$$

的参数才存在达到C-R下界的无偏估计

$$\alpha(X_1 + \ldots + X_n) + \beta = n\alpha\overline{X} + \beta.$$

有效估计和估计的效率

达到C-R下界的无偏估计,称为**有效估计**.

Definition

定义3.5.2 设$\widehat{g}(\widetilde{X})$ 是$g(\theta)$的无偏估计, 称

$$e_n = \frac{[g'(\theta)]^2/[nI(\theta)]}{\mathsf{Var}_\theta\{\widehat{g}(\widetilde{X})\}}$$

为无偏估计$\widehat{g}(\widetilde{X})$的效率(efficiency);如果$e_n = 1$, 则称$\widehat{g}(\widetilde{X})$ 是$g(\theta)$的有效估计; 如果$e_n \to 1$,当$n \to \infty$, 则称$\widehat{g}(\widetilde{X})$ 是$g(\theta)$的渐近有效估计.

如果$\widehat{g}(\widetilde{X})$ 是$g(\theta)$的估计，满足

$$\sqrt{n}\left\{\widehat{g}(\widetilde{X})-g(\theta)\right\} \xrightarrow{D} N(0,V(\theta)),\quad \text{当} \quad n\to\infty.$$

也把

$$e=\frac{[g'(\theta)]^2/[nI(\theta)]}{V(\theta)/n}=\frac{[g'(\theta)]^2/I(\theta)}{V(\theta)}$$

称为渐近效率.

Example

设$X_1, X_2, \ldots, X_n$是取自正态总体$N(\mu, \sigma^2)$的样本(σ^2已知)，前面已证样本方差$\overline{X}$是μ的有效估计，即效率$e_n = 1$. 下面考察样本中位数m_e的渐近效率.

总体$N(\mu, \sigma^2)$的中位数也是μ，且$p(x)|_{x=\mu} = \frac{1}{\sqrt{2\pi}\sigma}$. 已知

$$\sqrt{n}(m_e - \mu) \xrightarrow{D} N(0, \frac{\pi\sigma^2}{2}).$$

m_e的渐近效率为

$$e = \frac{1/I(\mu)}{\pi\sigma^2/2} = \frac{2}{\pi}.$$

在这个例子中

- 样本均值$\overline{X}$是MLE，也是LSE，它有高的效率，但不Robust.
- 样本中位数m_e是LADE，它是Robust估计，但相对而言它效率低.

评价一个统计量优劣常常有不同的准则.

You can not have everything at the same time.

(鱼和熊掌不可兼得)