Cremer-Rao 不等式简明讲稿

★ **Score 函数、Score 统计量、Fisher 信息量**

直观上看，如果一个事件发生的概率小，那么若它发生了就一定会给我们带来更多的信息。

若 $X_1, X_2, \cdots, X_n$ 为 $X \sim f(x;\theta)$ 的一个样本，其观测值为 $x_1, x_2, \cdots, x_n$，若 θ 为参数的真值，则似然函数的值就大，即对数似然函数的导数应该接近于 0（大多数情况下，求 MLE 就是直接令对数似然函数的对参数的导数为 0 得到的）。

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

为方便起见，记 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$，则

$\boldsymbol{X} = (X_1, X_2, \cdots, X_n) \sim f(\boldsymbol{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta) = L(\theta; x_1, \cdots, x_n)$（注意分布与似然的关系）

称对数似然的导数（多个参数时为梯度）称为 Score 函数。

Score 函数：$s(\boldsymbol{x}) = \frac{\partial}{\partial\theta}[\ln f(\boldsymbol{x};\theta)] = \frac{\partial}{\partial\theta}[\ln L(\theta;\boldsymbol{x})]$

Score 统计量：$s(\boldsymbol{X})$（其观测值为 $s(\boldsymbol{x})$）。

Score 函数（或统计量）具有如下性质：

（1）$s(\boldsymbol{x}) = \frac{\partial}{\partial\theta}[\ln f(\boldsymbol{x};\theta)] = \frac{\partial}{\partial\theta}[\ln \prod_{i=1}^{n} f(x_i,\theta)] = \sum_{i=1}^{n} \frac{\partial}{\partial\theta}[\ln f(x_i,\theta)]$，这里 $f(x;\theta)$（x 是小写）是总体的分布（注意与样本的分布 $f(\boldsymbol{x};\theta)$ 的区别与联系）

（2）期望 $E[s(\boldsymbol{X})] = 0$；方差 $D[s(\boldsymbol{X})] = E[s^2(\boldsymbol{X})] \triangleq I_n(\theta)$ 定义为 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 的 Fisher 信息量。

$$E[s(\boldsymbol{X})] = \int_{R^n} s(\boldsymbol{x}) f(\boldsymbol{x},\theta) d\boldsymbol{x} = \int_{R^n} \frac{\partial}{\partial\theta}[\ln f(\boldsymbol{x};\theta)] f(\boldsymbol{x},\theta) d\boldsymbol{x}$$

$$= \int_{R^n} \frac{\frac{\partial}{\partial\theta} f(\boldsymbol{x};\theta)}{f(\boldsymbol{x},\theta)} f(\boldsymbol{x},\theta) d\boldsymbol{x} = \frac{\partial}{\partial\theta} \int_{R^n} f(\boldsymbol{x},\theta) d\boldsymbol{x} = 0$$

Fisher 信息量用来估计 MLE 的方程的方差。数据越多，该量越大，得到的信息越多。

（3）如果二阶导数对一切 $\theta \in \Theta$ 都存在，则

$$I_n(\theta) = D[s(\boldsymbol{X})] = E[s^2(\boldsymbol{X})] = -E[\frac{\partial^2}{\partial\theta^2}[\ln f(\boldsymbol{x};\theta)]]$$

Fisher 信息量为对数似然在参数真值处的负二阶导数的期望，用来估计对数似然的曲率

Pf（上课略）：只给出一维情况，一般情形类似

$$0=\frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{\infty} f\left(x;\theta\right)dx=\int_{-\infty}^{\infty}\frac{\partial^2}{\partial\theta^2}f\left(x;\theta\right)dx$$

$$=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}[\frac{\partial\ln f\left(x;\theta\right)}{\partial\theta}f\left(x;\theta\right)]dx$$

$$=\int_{-\infty}^{\infty}[\frac{\partial^2\ln f\left(x;\theta\right)}{\partial\theta^2}f\left(x;\theta\right)+\frac{\partial\ln f\left(x;\theta\right)}{\partial\theta}\frac{\partial f\left(x;\theta\right)}{\partial\theta}]dx$$

$$=\int_{-\infty}^{\infty}[\frac{\partial^2\ln f\left(x;\theta\right)}{\partial\theta^2}]f\left(x;\theta\right)dx+\int_{-\infty}^{\infty}[\frac{\partial\ln f\left(x;\theta\right)}{\partial\theta}]^2 f\left(x;\theta\right)dx$$

$$=E[\frac{\partial^2\ln f\left(X;\theta\right)}{\partial\theta^2}]+E[\frac{\partial\ln f\left(X;\theta\right)}{\partial\theta}]^2=E[\frac{\partial^2\ln f\left(X;\theta\right)}{\partial\theta^2}]+I(\theta)$$

★ **CRLB**、有效估计

定理（**Cramer-Rao** 不等式）：设样本 $(X_1,X_2,\cdots,X_n)\triangleq\boldsymbol{X}$ 的联合分布密度（或分布律）

$f\left(\boldsymbol{x};\theta\right)$，$\theta\in\Theta\subset R$，满足（对参数求导）可交换条件：

$$\int_{R^n}\frac{\partial f(\boldsymbol{x},\theta)}{\partial\theta}d\boldsymbol{x}=\frac{\partial}{\partial\theta}\int_{R^n}f(\boldsymbol{x},\theta)d\boldsymbol{x},\theta\in\Theta$$

（注：等式右边必为 **0**，理由密度函数积分为 **1**）

记 $T(\boldsymbol{X})$ 为任方差有限的估计量，满足如下条件

$$\frac{\partial}{\partial\theta}[E(T(\boldsymbol{X}))]=\underbrace{\frac{\partial}{\partial\theta}\int_{R^n}T(\boldsymbol{x})f(\boldsymbol{x},\theta)d\boldsymbol{x}=\int_{R^n}T(\boldsymbol{x})\frac{\partial}{\partial\theta}f(\boldsymbol{x},\theta)d\boldsymbol{x}}_{\text{可交换求导}},\theta\in\Theta$$

则 $D(T(\boldsymbol{X}))\geq\dfrac{(\frac{\partial}{\partial\theta}[E(T(\boldsymbol{X}))])^2}{I_n(\theta)}$。

证明（上课略）：**Cauchy-Schwartz** 不等式：$DZ\geq\dfrac{(Cov(Z,Y))^2}{DY}$

取 $Z=T(\boldsymbol{X}),Y=s(\boldsymbol{X})=\dfrac{\partial}{\partial\theta}[\ln f\left(\boldsymbol{X};\theta\right)]$

$EY=0,DY=I_n(\theta)$

由于

$$Cov(Z,Y)=E(ZY)=E(T(\boldsymbol{X})s(\boldsymbol{X}))=E(T(\boldsymbol{X})\frac{\partial}{\partial\theta}[\ln f\left(\boldsymbol{X};\theta\right)])$$

$$=\int_{R^n}T(\boldsymbol{x})\frac{\partial}{\partial\theta}[\ln f\left(\boldsymbol{x};\theta\right)]f(\boldsymbol{x},\theta)d\boldsymbol{x}$$

$$=\int_{R^n}T(\boldsymbol{x})\frac{\frac{\partial}{\partial\theta}f\left(\boldsymbol{x};\theta\right)}{f(\boldsymbol{x},\theta)}f(\boldsymbol{x},\theta)d\boldsymbol{x}=\frac{\partial}{\partial\theta}\int_{R^n}T(\boldsymbol{x})f(\boldsymbol{x},\theta)d\boldsymbol{x}$$

$$=\frac{\partial}{\partial\theta}E(T(\boldsymbol{X})$$

故$D(T(\boldsymbol{X}))\geq\dfrac{(\frac{\partial}{\partial\theta}[E(T(\boldsymbol{X}))])^2}{I_n(\theta)}$。

注:

（1）$I_n(\theta)$是度量样本$(X_1,X_2,\cdots,X_n)\triangleq\boldsymbol{X}$ 中含θ的信息量，X_1的 Fisher 信息量是

$I_1(\theta)$。$I_1(\theta)=E[s^2(X_1)]=-E[\frac{\partial^2}{\partial\theta^2}[\ln f\left(x;\theta\right)]]$

（2）总体的两个独立样本$\boldsymbol{X},\boldsymbol{Y}$，则$(\boldsymbol{X},\boldsymbol{Y})$的 Fisher 信息量为其各自 Fisher 信息量之和。

（3）（与教材上的定义的联系）若总体$X\sim f\left(x;\theta\right)$的一个样本，其 Fisher 信息量是

$I_1(\theta)$（同教材），故$X_1,X_2,\cdots,X_n$的 Fisher 信息量$I_n(\theta)=nI_1(\theta)$。

此时$I_1(\theta)=E[s^2(X_1)]=E[(\frac{\partial}{\partial\theta}[\ln f\left(x;\theta\right)])^2]=-E[\frac{\partial^2}{\partial\theta^2}[\ln f\left(x;\theta\right)]]$

该结论的另一证明（略）:

$$I_n(\theta)=E[s^2(\boldsymbol{X})]=E[(\frac{\partial}{\partial\theta}[\ln f\left(\boldsymbol{X};\theta\right)])^2]=E[(\frac{\partial}{\partial\theta}[\ln\prod_{i=1}^n f\left(X_i;\theta\right)])^2]$$

$$=E[(\sum_{i=1}^n\frac{\partial}{\partial\theta}[\ln f\left(X_i;\theta\right)])^2]$$

$$=\sum_{i=1}^n E(\frac{\partial}{\partial\theta}[\ln f\left(X_i;\theta\right)])^2+2\sum_{i<j}E[\frac{\partial}{\partial\theta}\ln f\left(X_i;\theta\right)\frac{\partial}{\partial\theta}\ln f\left(X_j;\theta\right)]$$

$$=\sum_{i=1}^n E(\frac{\partial}{\partial\theta}[\ln f\left(X_i;\theta\right)])^2=nI_1(\theta)$$

（4）与 MLE 联系（不要求）$\sqrt{n}(\hat{\theta}_{MLE}-\theta)\xrightarrow{D}N(0,\frac{1}{I_n(\theta)})$

（5）若$T(\boldsymbol{X})$为$g(\theta)$的无偏估计，即$E(T(\boldsymbol{X}))=g(\theta)$，则

$$D(T(\boldsymbol{X}))\geq\frac{(g'(\theta))^2}{nI_1(\theta)}$$ （即教材结果）

若无偏估计达到 CRLB，则该估计称为有效估计，故有效估计必为 UMVUE。

（6）（不要求）多参量情况：$\boldsymbol{\theta}\in R^k$

$$D(T(\boldsymbol{X}))\geq(\frac{\partial}{\partial\theta}[E(T(\boldsymbol{X}))])^T[I_n(\theta)]^{-1}\frac{\partial}{\partial\theta}[E(T(\boldsymbol{X}))]$$

$I_n(\boldsymbol{\theta})=E((\frac{\partial}{\partial\boldsymbol{\theta}}[\ln f(\boldsymbol{x};\boldsymbol{\theta})])(\frac{\partial}{\partial\boldsymbol{\theta}}[\ln f(\boldsymbol{x};\boldsymbol{\theta}))^T]))$是$k\times k$ Fisher 信息矩阵。

例：Poisson 总体$X\sim P(\lambda)$，$\bar{X}$是λ的有效估计，也是 UMVUE。

Pf：（这里用$I_1(\lambda)$计算，注意$I_n(\lambda)=nI_1(\lambda)$ ）$X\sim P(\lambda)$的 p.m.f.为

$$f(x;\lambda)=\frac{\lambda^x e^{-\lambda}}{x!},x=0,1,\cdots\Rightarrow\ln f(x;\lambda)=\ln\frac{\lambda^x e^{-\lambda}}{x!}=x\ln\lambda-\lambda-\ln(x!)$$

$$\frac{\partial}{\partial\lambda}\ln f(x;\lambda)=\frac{x}{\lambda}-1,\qquad(=s(x))$$

$$\frac{\partial^2}{\partial\lambda^2}\ln f(x;\lambda)=-\frac{x}{\lambda^2},\quad(=\frac{\partial}{\partial\lambda}s(x))$$

$$I_1(\lambda)=-E(\frac{\partial^2}{\partial\lambda^2}\ln f(X;\lambda))=-E(-\frac{X}{\lambda^2})=\frac{1}{\lambda}$$

故其 CRLB 为$\frac{1}{nI_1(\lambda)}=\frac{\lambda}{n}$

由于$E(\bar{X})=\lambda,D(\bar{X})=\frac{\lambda}{n}$，达到 CRLB，故是有效估计，也是 UMVUE。

例：均匀总体$U(0,\theta)$的有关问题，$f(x;\theta)=\frac{1}{\theta}1_{\{0<\theta<1\}}$

直接计算如下：$\frac{\partial}{\partial\theta}\ln f(x;\theta)=-\frac{1}{\theta}\Rightarrow I_1(\theta)=E[(\frac{\partial}{\partial\theta}\ln f(X;\theta))^2]=\frac{1}{\theta^2}$

CRLB 应为 $\frac{1}{nI_1(\theta)}=\frac{\theta^2}{n}$

前面的课上，我们已经证明了 $\frac{n+1}{n}X_{(n)}$ 为 θ 的无偏估计，且

$$D(\frac{n+1}{n}X_{(n)})=\frac{\theta^2}{n(n+2)}<\frac{\theta^2}{n}=CRLB$$

问题出在哪里：（对参数求导的可交换性不满足，CR 定理不能用），事实上

$$\frac{d}{d\theta}\int_0^\theta t(x)f(x;\theta)dx=\frac{d}{d\theta}\int_0^\theta t(x)\frac{1}{\theta}dx=\frac{t(\theta)}{\theta}+\int_0^\theta t(x)\frac{d}{d\theta}(\frac{1}{\theta})dx$$
$$\neq\int_0^\theta t(x)\frac{d}{d\theta}f(x;\theta)dx$$

注：

（1）一般若 p.d.f 的支撑集依赖于待估参数，可交换条件不满足，不能用 Cramer-Rao 定理。好在于若是指数族分布，均可以满足，即可以直接用 Cramer-Rao 定理。

（2）上例中 $\frac{n+1}{n}X_{(n)}$ 确为 θ 的 UMVUE，事实上，$X_{(n)}$ 为 θ 的充分统计量，且其分布密度为

$$f_{X_{(n)}}(x)=\begin{cases}\frac{nx^{n-1}}{\theta^n}, & 0<x<\theta,\\ 0, & elsewhere.\end{cases}$$

该分布族是完备的，因为根据完备的定义，只需证明：$X\sim f_{X_{(n)}}(x)$，若当 $E(g(X))=0$ 时必有 $P(g(X)=0)=1$，由于

$$E(g(X))=\int_0^\theta g(x)\frac{nx^{n-1}}{\theta^n}dx=0,\forall\theta>0$$
$$\Leftrightarrow\int_0^\theta g(x)x^{n-1}dx=0,\forall\theta>0$$

两边对 θ 求导得，$g(\theta)\theta^{n-1}=0$，故 $g(\theta)=0,\forall\theta>0$，从而完备，又由于 $E(\frac{n+1}{n}X_{(n)})=\theta$（无偏），因此由 Lemann-Scheffe 定理知，θ 的 UMVUE 为

$$E(\frac{n+1}{n}X_{(n)}\mid X_{(n)})=\frac{n+1}{n}X_{(n)}$$

（3）关于完备统计量，我们有如下定理：

定理：若总体的分布是指数族分布，即

$$f(x;\boldsymbol{\theta})=h(x)c(\boldsymbol{\theta})\exp\{\sum_{j=1}^{k}w(\theta_j)T_j(x)\},\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)\in\Theta$$

则只要参数空间Θ包含R^k中的一个开集，则统计量

$(\sum_{i=1}^{n}T_1(X_i),\sum_{i=1}^{n}T_2(X_i),\cdots,\sum_{i=1}^{n}T_k(X_i))$ 是完备统计量。