§2.3 信息不等式

前面主要讨论了如何求UMVUE的问题。但有时候UMVUE并不存在或不易求得，这就需要一个方差下界去评价某个具体的无偏估计的表现。本节讨论无偏估计的方差下界及与之相关的问题。基本结果是瑞典的Cramér 和印度的C. R. Rao 分别在1945-46年对单参数正则分布族证明的一个重要不等式(称为C-R不等式或信息不等式)。以后又有不少学者将条件作了改进，并给出了更为精确的下界，但结果的基本形式并没有重大改变。这些方差下界主要是通过Cauchy-Schwarz 不等式导出的，一般不是下确界，但是它们计算容易，在参数估计理论中有着重要作用。

一、单参数情况下U.E. 的方差下界

考虑参数$\gamma(\theta)$ 的任一无偏估计量δ 的方差下界。取一函数$\psi(X,\theta)$，由Cauchy-Schwarz 不等式知

$$\mathrm{Var}_\theta(\delta) \geq \frac{[\mathrm{Cov}_\theta(\delta,\psi)]^2}{\mathrm{Var}_\theta(\psi)}. \tag{2.3.1}$$

- 若$\mathrm{Var}_\theta(\delta)=\infty$ 或$\mathrm{Var}_\theta(\psi)=\infty$，则不等式自然成立，不起作用。
- 若取合适的ψ，使$\mathrm{Var}_\theta(\psi)<\infty$ 且$\mathrm{Cov}_\theta(\delta,\psi)$ 只与$E_\theta(\delta)=\gamma(\theta)$ 有关，那上式右边就是$\gamma(\theta)$ 的所有无偏估计方差的一个下界。

1. Hammersly-Chapman-Robbins 不等式(1950,1951)

设X 关于控制测度μ 具有p.d.f $p_\theta(x)$, $\theta\in\Theta\subset\mathbb{R}$。给定$\Delta$，若$p_{\theta+\Delta}(x)\neq p_\theta(x)$，且$p_{\theta+\Delta}(\cdot)$ 的支撑包含于$p_\theta(\cdot)$ 的支撑内，那么取

$$\psi(x;\theta)=\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{p_\theta(x)},$$

则由Cauchy-Schwarz 不等式知，

$$\mathrm{Var}_\theta(\delta) \geq \frac{[\gamma(\theta+\Delta)-\gamma(\theta)]^2}{E_\theta\left[\frac{p_{\theta+\Delta}(X)-p_\theta(X)}{p_\theta(X)}\right]^2}. \tag{2.3.2}$$

注：

- 若对(2.3.2) 的右边关于Δ 取最大值，不等式依然成立。

- 在一定条件下，令$\Delta \to 0$，便可由(2.3.2) 得到C-R 不等式。

2. C-R 不等式

定理2.3.1 (C-R不等式, 教材p335, Th7.3.9) 设样本空间$(\mathscr{X}, \mathscr{B}_X)$ 上的分布族$\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ 关于σ 有限测度μ 可控，$dP_\theta(x)/d\mu(x) = p_\theta(x)$, $\theta \in \Theta$。若$\mathscr{P}$ 满足正则条件：

1° 参数空间$\Theta \subset \mathbb{R}$ 是开区间。

2° $\mathscr{P}$ 有共同支撑A；对$\forall x \in A$, $\forall \theta \in \Theta$，有$p_\theta(x) > 0$，$\frac{\partial}{\partial\theta} p_\theta(x)$ 存在、有限。

3° $E_\theta \left[\frac{\partial}{\partial\theta} \log p_\theta(X)\right] = \int_{\mathscr{X}} \frac{\partial}{\partial\theta} p_\theta(x) d\mu(x) = 0, \forall \theta \in \Theta$。

4° $I(\theta) \hat{=} E_\theta \left[\frac{\partial}{\partial\theta} \log p_\theta(X)\right]^2 > 0$，$\forall \theta \in \Theta$。

$\gamma(\theta)$ 是定义在Θ 上的有限实函数，可估且处处可微。若δ 是$\gamma(\theta)$ 的无偏估计且满足条件

5° $\int_{\mathscr{X}} \delta(x) \frac{\partial}{\partial\theta} p_\theta(x) d\mu(x) = \frac{\partial}{\partial\theta} \int_{\mathscr{X}} \delta(x) p_\theta(x) d\mu(x) = \gamma'(\theta)$,

则有下列不等式*(C-R 不等式)* 成立

$$Var_\theta(\delta) \geq [\gamma'(\theta)]^2 / I(\theta). \tag{2.3.3}$$

证明. 不妨设$I(\theta) < \infty$ 且$\mathrm{Var}_\theta(\delta) < \infty$，否则不等式(2.3.3) 显然成立。取

$$\psi(x; \theta) = \frac{\partial}{\partial\theta} \log p_\theta(x).$$

由正则条件$3°, 4°$ 知，$E_\theta[\psi(X;\theta)] = 0$，$0 < \mathrm{Var}_\theta[\psi(X;\theta)] = I(\theta) < \infty$。再由$\delta$

的无偏性及条件5° 知，

$$\begin{aligned}\mathrm{Cov}_\theta[\delta(X),\psi(X,\theta)] &= E_\theta[\delta(X)\psi(X;\theta)] \\ &= \int_{\mathscr{X}} \delta(x)\frac{\partial \log p_\theta(x)}{\partial\theta}p_\theta(x)d\mu(x) \\ &= \int_{\mathscr{X}} \delta(x)\frac{\partial}{\partial\theta}p_\theta(x)d\mu(x) \\ &= \frac{\partial}{\partial\theta}\int_{\mathscr{X}} \delta(x)p_\theta(x)d\mu(x) \\ &= \frac{d}{d\theta}\gamma(\theta)=\gamma'(\theta).\end{aligned}$$

因此，由Cauchy-Schwarz 不等式得

$$\mathrm{Var}_\theta(\delta)\geq\frac{\mathrm{Cov}^2_\theta[\delta(X),\psi(X,\theta)]}{\mathrm{Var}_\theta[\psi(X;\theta)]}=\frac{[\gamma'(\theta)]^2}{I(\theta)}.$$

□

C-R 不等式给出了$\gamma(\theta)$ 的无偏估计方差的一个下界。设δ 是$\gamma(\theta)$ 的一个无偏估计，若在θ_0 处(a) δ 的方差达到C-R 下界，且(b) $\gamma(\theta)$ 的所有无偏估计都满足条件5°，则δ 就是$\gamma(\theta)$ 在θ_0 处的LMVUE；若(a), (b)对所有$\theta\in\Theta$ 都成立，则δ 就是$\gamma(\theta)$ 的UMVUE。

推论2.3.1 (单参数指数族) *(1) 若*$\mathscr{P}=\{P:\ dP(x)/d\mu(x)=A(\eta)\cdot\exp[\eta T(x)]\cdot h(x),\ \eta\in\Xi\}$ *是典则形式的指数族，*Ξ *是自然参数空间。则条件*1° ∼ 5° *成立，C-R 不等式适用。*

(2) 若$\mathscr{P}=\{P:\ dP(x)/d\mu(x)=B(\theta)\cdot\exp[\eta(\theta)T(x)]\cdot h(x),\ \theta\in\Theta\}$ *是一般形式的指数族，*Θ *满足正则条件*1°*，*$B(\theta)$*,* $\eta(\theta)$ *连续可微，则条件*2° ∼ 5° *成立，C-R 不等式适用。*

3. 等号成立的条件

由Cauchy-Schwarz 不等式知，(2.3.3) 中等号成立的充要条件为：存在不全为零、与x 无关的$\alpha(\theta),\beta(\theta)$，使得

$$\alpha(\theta)[\delta(x)-\gamma(\theta)]+\beta(\theta)\psi(x;\theta)=0,\ a.s.\ P_\theta.$$

在正则条件$2°, 4°$ 及$\gamma(\theta)$ 不恒为常数的前提下，该充要条件又等价于：存在非零、与x 无关的$\alpha(\theta)$，使得

$$\frac{\partial}{\partial\theta}\log p_\theta(x) = \alpha(\theta)[\delta(x) - \gamma(\theta)],\ a.e.\ \mu.$$

若两边分别对θ 积分可得

$$p_\theta(x) = B(\theta)\cdot e^{\varphi(\theta)\delta(x)}\cdot h(x),$$

其中$B(\theta) = e^{-\int_{\theta_0}^{\theta}\alpha(\theta)\gamma(\theta)d\theta}$，$\varphi(\theta) = \int_{\theta_0}^{\theta}\alpha(\theta)d\theta$，$h(x)$ 与θ 无关。可见，在正则条件下，只有指数族才可能使C-R 下界处处达到。

定理2.3.2 (教材p341, Cor7.3.15) *在定理2.3.1的记号下，若分布族$\mathscr{P}$ 满足正则条件$1° \sim 4°$ 以及条件*

$6°$ *对任意固定的$x \in A$，$\frac{\partial}{\partial\theta}p_\theta(x)$ 为θ 在Θ 上的连续函数。*

$\gamma(\theta)$ 在Θ 上不恒为常数。δ 是$\gamma(\theta)$ 的一个无偏估计，$Var_\theta(\delta) < \infty$, $\forall\theta \in \Theta$。则C-R 不等式(2.3.3) 中等号对$\forall\theta \in \Theta$ 成立的充要条件是：

$$p_\theta(x) = B(\theta)e^{\varphi(\theta)\delta(x)}h(x),\ \forall\theta \in \Theta.$$

且其中的$B(\theta) > 0$, $\varphi(\theta)$ 在Θ 上连续可微，$\varphi(\theta)$ 在Θ 上严格单调, $h(x) > 0$。

严格证明参见成平等(1985)《参数估计》p.125-126，或陈希孺(1981)《数理统计引论》p.115-117。

即使在指数型分布族中，也不是所有可估函数的无偏估计的方差都能处处达到C-R 下界，只有少量可估函数才有处处达到C-R 下界的无偏估计。

推论2.3.2 *在满足定理2.3.2 条件的指数型分布族中，当且仅当*

$$\gamma(\theta) = E_\theta[a\delta(X) + b],\ \theta \in \Theta$$

时，$\gamma(\theta)$ 才有方差处处达到C-R 下界的无偏估计$\hat{g}(X) = a\delta(X) + b$，其中$a, b$ 为常数。

例2.3.1 设随机变量X 的分布族由非负整数集上的离散型分布组成，且$E(X)=\mathrm{Var}(X)=\theta\in(0,\infty)$。若$X$ 作为θ 的U.E. 方差处处达到C-R 下界，则

$$\theta=\mathrm{Var}(X)=\left[\frac{d}{d\theta}E_\theta(X)\right]^2/I(\theta).$$

故$I(\theta)=1/\theta$。又由定理2.3.2知，X 的密度函数形式必为：

$$p_\theta(x)=B(\theta)e^{x\varphi(\theta)}h(x).$$

由$I(\theta)$ 的定义知，

$$I(\theta)=\theta[\varphi'(\theta)]^2.$$

因此，$\varphi'(\theta)=1/\theta$，$\varphi(\theta)=\log\theta$。又由指数族的性质知，

$$\theta=E_\theta(X)=-\frac{d\log B(\theta)}{d\theta}\cdot\frac{1}{\varphi'(\theta)},$$

因此，$B(\theta)=e^{-\theta}$，$h(x)=1/x!$。所以$X\sim\mathrm{Poisson}(\theta)$。

例2.3.2 (1) $X\sim B(n,\theta)$, $\theta\in(0,1)$，哪些估计量的方差可达到C-R 下界？因为$B(n,\theta)$ 的密度函数为

$$p_\theta(x)=\binom{n}{x}\left(\frac{\theta}{1-\theta}\right)^x(1-\theta)^n=(1-\theta)^n e^{x\log\left(\frac{\theta}{1-\theta}\right)}\binom{n}{x}I_{\{0,1,\cdots,n\}}(x),$$

故X 的分布属指数族。由推论2.3.2知，只有$\delta(X)=aX+b$ 的方差可达C-R 下界，即只有$\gamma(\theta)=a\cdot n\theta+b=c\theta+b$ 才有方差处处达到C-R 下界的U.E.。

(2) p172/Example 3.10。对于方差已知的正态分布族，$\overline{X}$ 作为总体均值μ 的UMVUE，其方差处处达到C-R 下界。但是$\overline{X}^2-\sigma^2/n$ 作为μ^2 的UMVUE，方差不能达到C-R下界。

定理2.3.1的一个弱点在于它只能适用于满足条件5° 的δ。若能将加于具体估计量的条件5° 转化为对分布族的限制，那就比较方便了。下面的引理提供了一个充分条件。

引理2.3.1 假定正则条件1°, 2° 成立，$\psi(x;\theta)=\frac{\partial}{\partial\theta}\log p_\theta(x)$，且$\exists b_\theta(x)$，满足

a. 对$\forall\theta\in\Theta$，$b_\theta(x)$ 为x 的$\mathscr{B}_X$ 可测函数；

b. $E_\theta b_\theta^2(X)<\infty$，$\forall\theta\in\Theta$；

c. 对$\forall\theta\in\Theta$，$\exists\epsilon_\theta>0$，使得$\left|\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}\right|\le b_\theta(x),\forall|\Delta|<\epsilon_\theta$。

则

(1) $E_\theta\psi(X;\theta)=0,\ \forall\theta\in\Theta$ （即正则条件3° 满足）。

(2) 若$E_\theta\delta^2<\infty$，则有*(*即条件5° 满足*)*

$$\frac{d}{d\theta}E_\theta[\delta(X)]=E_\theta[\delta(X)\psi(X;\theta)]=Cov_\theta(\delta,\psi).$$

证明. 因为

$$\left|\delta(x)\cdot\frac{p_{\theta+\Delta}(x)-p_\theta(x)}{\Delta p_\theta(x)}\right|\le|\delta(x)||b_\theta(x)|,$$

且

$$E_\theta|\delta(X)\cdot b_\theta(X)|\le[E_\theta\delta^2(X)\cdot E_\theta b_\theta^2(X)]^{1/2}<\infty,$$

故由 Lebesgue 控制收敛定理即可证得该引理。 □

例2.3.3 随机变量$X\sim f(x-\theta)$，$\theta\in(-\infty,\infty)$，$f(\cdot)$ 是自由度为m 的t分布的p.d.f.。则正则条件$1^\circ\sim2^\circ$ 满足，但$3^\circ,5^\circ$ 不能一眼看出。考虑$m=1$ 即Cauchy 分布的情况。因为

$$f(x-\theta)=\frac{1}{\pi}\frac{1}{1+(x-\theta)^2},$$

$$\left|\frac{1}{\Delta}\Big[\frac{1+(x-\theta)^2}{1+(x-\Delta-\theta)^2}-1\Big]\right|\le2+\epsilon,\ \forall|\Delta|<\epsilon,\forall\theta\in(-\infty,\infty).$$

故取$b_\theta(x)=2+\epsilon$，则引理2.3.1的条件符合。

4. 有效率(efficiency)

对于满足正则条件的分布族，因为有C-R 下界，人们就用它建立一个衡量无偏估计优良程度的标准，即有效率。

设$\gamma(\theta)$ 可估，δ 是$\gamma(\theta)$ 的一个无偏估计，称

$$\frac{[\gamma'(\theta)]^2I^{-1}(\theta)}{\mathrm{Var}_\theta(\delta)}$$

为δ 的有效率。如果它等于1，则称δ 为$\gamma(\theta)$ 的有效估计。

这种有效率的定义是值得商榷的，因为有效率能达到1的估计量极少，即便是UMVUE 也往往不是有效估计。

5. Bhattacharyya 不等式

1946年，Bhattacharyya 在更强的正则条件下，利用$p_\theta(x)$ 关于θ 的多阶偏导数，改进了C-R 不等式，得到了一系列越来越大的方差下界。

定理2.3.3 采用定理*2.3.1*的记号，假定分布族满足正则条件：

i),ii) 同正则条件$1^\circ, 2^\circ$；

iii) $\int_{\mathscr{X}} \frac{\partial^i p_\theta(x)}{\partial\theta^i} d\mu(x) = 0, \quad i = 1, \cdots, K,\ \forall\theta \in \Theta$；

iv) $\int_{\mathscr{X}} \frac{1}{p_\theta(x)} \left[\frac{\partial^i p_\theta(x)}{\partial\theta^i}\right]^2 d\mu(x) < \infty, \quad i = 1, \cdots, K,\ \forall\theta \in \Theta$。

设δ 是$\gamma(\theta)$ 的无偏估计，方差有限，且满足条件

v) $$\int_{\mathscr{X}} \delta(x) \frac{\partial^i p_\theta(x)}{\partial\theta^i} d\mu(x) = \frac{\partial^i}{\partial\theta^i} \int_{\mathscr{X}} \delta(x) p_\theta(x) d\mu(x) = \gamma^{(i)}(\theta), i = 1, \cdots, K, \forall\theta \in \Theta.$$

记 $\mathbf{l}(\theta) = (\gamma'(\theta), \gamma^{(2)}(\theta), \cdots, \gamma^{(K)}(\theta))^\tau$，$\mathbf{V}(\theta) = (v_{ij}(\theta))_{K\times K}$，其中

$$v_{ij}(\theta) = E_\theta\left[\frac{1}{p_\theta^2(X)} \frac{\partial^i p_\theta(X)}{\partial\theta^i} \frac{\partial^j p_\theta(X)}{\partial\theta^j}\right], \quad i, j = 1, \cdots, K.$$

则当$\theta \in \Theta$ 处$|\mathbf{V}(\theta)| > 0$ 时，必有

$$Var_\theta(\delta) \geq \mathbf{l}^\tau(\theta)\mathbf{V}^{-1}(\theta)\mathbf{l}(\theta). \tag{2.3.4}$$

证明. 记 $\mathbf{\Psi} = (\psi_1(X;\theta), \cdots, \psi_K(X;\theta))^\tau$，其中

$$\psi_i(x;\theta) = \frac{1}{p_\theta(x)} \frac{\partial^i p_\theta(x)}{\partial\theta^i}, \quad i = 1, \cdots, K.$$

则由条件iii) 知，$E_\theta(\mathbf{\Psi}) = 0$；由条件iv) 知，$\mathrm{Cov}_\theta(\mathbf{\Psi}) = \mathbf{V}(\theta)$ 存在；再由条件v) 知，$\mathrm{Cov}_\theta(\delta, \mathbf{\Psi}) = \mathbf{l}(\theta)$。因此，

$$\mathbf{M} = \mathrm{Cov}_\theta \begin{pmatrix} \delta \\ \mathbf{\Psi} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}_\theta(\delta) & \mathbf{l}^\tau(\theta) \\ \mathbf{l}(\theta) & \mathbf{V}(\theta) \end{pmatrix}.$$

由于

$$|\mathbf{M}| = |\mathbf{V}(\theta)| \cdot [\mathrm{Var}_\theta(\delta) - \mathbf{l}^\tau(\theta)\mathbf{V}^{-1}(\theta)\mathbf{l}(\theta)] \geq 0,$$

再由$|\mathbf{V}(\theta)| > 0$ 即证得定理结论。 □

二、Fisher信息量

定义2.3.1 采用定理2.3.1中的记号，假定分布族满足正则条件$1^\circ \sim 4^\circ$，称

$$I(\theta) = \mathrm{Var}_\theta\Big[\frac{\partial}{\partial\theta}\log p_\theta(X)\Big] = E_\theta\Big[\frac{\partial}{\partial\theta}\log p_\theta(X)\Big]^2$$

为θ 的*Fisher* 信息量。

信息量依赖于参数形式。若$\theta = h(\xi)$，h 可微，则有

$$I^*(\xi) = I(\theta(\xi)) \cdot \Big(\frac{d\theta}{d\xi}\Big)^2.$$

但C-R 不等式与参数形式无关，因为参数$\gamma(\theta)$ 的无偏估计方差的C-R 下界为$[\gamma'(\theta)]^2 I(\theta)^{-1}$；若将该参数视为$\xi$ 的函数，即$\gamma(\theta(\xi))$，则其无偏估计方差的C-R 下界为

$$\frac{(\frac{d\gamma}{d\theta}\cdot\frac{d\theta}{d\xi})^2}{I(\theta(\xi))\cdot(\frac{d\theta}{d\xi})^2} = \frac{[\gamma'(\theta)]^2}{I(\theta)}.$$

两者相同。

引理2.3.2 若$\log p_\theta(x)$ 关于θ 二阶可导，$\forall x \in A, \forall\theta \in \theta$，且

$$\int \frac{\partial^2}{\partial\theta^2} p_\theta(x) d\mu(x) = \frac{\partial^2}{\partial\theta^2}\int p_\theta(x) d\mu(x) = 0,$$

则

$$I(\theta) = -E_\theta\Big[\frac{\partial^2}{\partial\theta^2}\log p_\theta(X)\Big].$$

证明. 因为

$$\frac{\partial^2}{\partial\theta^2}\log p_\theta(X) = \frac{\frac{\partial^2}{\partial\theta^2}p_\theta(X)}{p_\theta(X)} - \Big[\frac{\frac{\partial}{\partial\theta}p_\theta(X)}{p_\theta(X)}\Big]^2,$$

故结论显然。 □

定理2.3.4 (独立可加性) *1.* $X \sim p_\theta(\cdot)$（关于μ），$Y \sim q_\theta(\cdot)$（关于ν），$\theta \in \Theta$，X, Y 相互独立，且两个分布族均满足正则条件$1^\circ \sim 4^\circ$。若X 中θ 的信息量为$I_1(\theta)$，Y 中θ 的信息量为$I_2(\theta)$，则(X, Y) 中θ 的信息量$I(\theta) = I_1(\theta) + I_2(\theta)$。

2. 若$X_1, \ldots, X_n$ *i.i.d.* $p_\theta(\cdot)$，$\{p_\theta(\cdot) : \theta \in \Theta\}$ 满足正则条件$1^\circ \sim 4^\circ$。其中X_1 的信息量为$I_1(\theta)$，则$(X_1, \ldots, X_n)$ 中θ 的信息量为$nI_1(\theta)$。

例2.3.4 (单参数指数族各种形式参数的Fisher信息量) 设X 的pdf为

$$p_\theta(x) = B(\theta)e^{\eta(\theta)T(x)}h(x),\ \theta \in \Theta \subset \mathbb{R}, \tag{2.3.5}$$

其中$B(\theta), \eta(\theta)$ 连续可微。由指数族的性质可知，

$$\tau(\theta) \hat{=} E_\theta(T(x)) = -\frac{B'(\theta)}{B(\theta)\cdot\eta'(\theta)},\quad \mathrm{Var}_\theta T(x) = \tau'(\theta)/\eta'(\theta).$$

1. θ 的信息量为
$$\begin{aligned} I_1(\theta) &= E_\theta\Big[\frac{\partial \log p_\theta(X)}{\partial\theta}\Big]^2 = E_\theta\Big[\frac{B'(\theta)}{B(\theta)} + \eta'(\theta)T(X)\Big]^2 \\ &= [\eta'(\theta)]^2 E_\theta[T(x)-\tau(\theta)]^2 = [\eta'(\theta)]^2\mathrm{Var}_\theta(T) = \tau'(\theta)\cdot\eta'(\theta). \end{aligned}$$

2. 典则化参数 $\eta = \eta(\theta)$ (与θ 一一对应) 的信息量为
$$\begin{aligned} I_2(\eta) &= I_1(\theta(\eta))\cdot\Big(\frac{d\theta}{d\eta}\Big)^2 = \Big(\frac{d\eta}{d\theta}\Big)^2\mathrm{Var}_\theta(T)\Big(\frac{d\theta}{d\eta}\Big)^2 \\ &= \mathrm{Var}_\theta(T) = \tau'(\theta)/\eta'(\theta). \end{aligned}$$

3. 均值化参数$\tau = \tau(\theta)$ (与θ 一一对应) 的信息量为
$$I_3(\tau) = I_1(\theta(\tau))\cdot\Big(\frac{d\theta}{d\tau}\Big)^2 = \eta'(\theta)\tau'(\theta)\Big(\frac{1}{\tau'(\theta)}\Big)^2 = \frac{1}{\mathrm{Var}_\theta(T)}.$$

例2.3.5 设$X_1, \cdots, X_n$ i.i.d.，服从

(1) Gamma$(\alpha, \theta), \alpha$已知$, \theta > 0$未知;

(2) $N(\theta, \sigma^2), \sigma^2$已知$, \theta \in \mathbb{R}$未知;

(3) Poisson$(\theta), \theta > 0$未知.

求这三种分布族下参数的Fisher 信息量。

(1) X_1 的pdf为
$$p_\theta(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha}x^{\alpha-1}e^{-\frac{x}{\theta}}I_{(0,\infty)}(x),$$
记 $\eta(\theta) = -\frac{1}{\theta}$, $T(x) = x$, $B(\theta) = \theta^{-\alpha}$，则该pdf具有(2.3.5)的形式，且
$$\tau(\theta) \hat{=} E_\theta(T) = -\frac{\partial \log B(\theta)}{\partial\theta\cdot\eta'(\theta)} = \alpha\theta,\quad \mathrm{Var}_\theta(T) = \alpha\theta^2.$$

因此，由上例及定理2.3.4知，$X_1, \cdots, X_n$中三类参数的Fisher 信息量分别为

$$\begin{aligned} I_1(\theta) &= n(\eta'(\theta))^2 \mathrm{Var}_\theta(T) = n\alpha\theta^{-2}, \\ I_2(\eta) &= n\mathrm{Var}_\theta(T) = n\alpha\eta^{-2}, \\ I_3(\tau) &= n\alpha\tau^{-2} \end{aligned}$$

(2) 由类似的讨论可知，$X_1, \cdots, X_n$中θ 的信息量为$I_1(\theta) = \frac{n}{\sigma^2}$，该信息量与$\theta$ 本身无关。若限制$\theta > 0$，令$\xi = \theta^2$，则$X_1, \cdots, X_n$ 中ξ 的信息量为$I_2(\xi) = \frac{n}{4\xi\cdot\sigma^2}$，与$\xi = \theta^2$ 有关。

(3) $X_1, \cdots, X_n$中θ 的信息量为$I_1(\theta) = \frac{n}{\theta}$.

例2.3.6 (位置族参数的Fisher信息量) 设$X_1, \cdots, X_n$ i.i.d.，X_1 关于Lebesgue 测度μ 具有pdf $f(x-\theta)$，其中$\theta \in \mathbb{R}$。若对于$\forall x \in \mathbb{R}$，$f(x) > 0$，$f'(x)$存在，则条件$1^\circ \sim 2^\circ$ 成立。假定条件$3^\circ \sim 4^\circ$ 也成立。记

$$I_f = \int_{-\infty}^{\infty} \frac{[f'(x)]^2}{f(x)} d\mu(x),$$

则$X_1, \cdots, X_n$ 中θ 的信息量为$I(\theta) = nI_f$，与θ 无关。例如，

(1) $X_1 \sim N(\theta, 1)$，则$I_f = 1,\ I(\theta) = n$；

(2) $X_1 \sim \mathrm{Cauchy}(\theta, 1)$，则$I_f = 1/2,\ I(\theta) = n/2$。

三、多参数情况

定义2.3.2 设X 关于控制测度μ 具有pdf $p_{\boldsymbol{\theta}}(\cdot)$，$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)^\tau \in \boldsymbol{\Theta} \subset \mathbb{R}^s$，$p_{\boldsymbol{\theta}}(\cdot)$满足正则条件

1'. $\boldsymbol{\Theta}$ 为$\mathbb{R}^s$ 中的开凸集。

2'. $p_{\boldsymbol{\theta}}(\cdot)$ 有共同支撑A；$\forall x \in A, \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, p_{\boldsymbol{\theta}}(x) > 0$，且$\frac{\partial p_{\boldsymbol{\theta}}(x)}{\partial \theta_i}, i = 1, \ldots, s$ 存在、有限。

3'. $\int_{\mathscr{X}} \frac{\partial}{\partial \theta_i} p_{\boldsymbol{\theta}}(x) d\mu(x) = 0,\ i = 1, \ldots, s$。

则称矩阵$I(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))_{s\times s}$ 为 $\boldsymbol{\theta}$ 的*Fisher*信息矩阵，其中

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial\theta_i}\log p_{\boldsymbol{\theta}}(X)\cdot\frac{\partial}{\partial\theta_j}\log p_{\boldsymbol{\theta}}(X)\right] = \mathrm{Cov}\left[\frac{\partial}{\partial\theta_i}\log p_{\boldsymbol{\theta}}(X), \frac{\partial}{\partial\theta_j}\log p_{\boldsymbol{\theta}}(X)\right].$$

性质：

- $I(\boldsymbol{\theta})$ 是方差-协方差矩阵，因此$I(\boldsymbol{\theta}) \geq 0$；若各个$\frac{\partial}{\partial\theta_i}\log p_{\boldsymbol{\theta}}(x), i = 1,\ldots,s$ 线性无关，则$I(\boldsymbol{\theta}) > 0$。
- 若$\int_{\mathscr{X}}\frac{\partial^2}{\partial\theta_i\partial\theta_j}p_{\boldsymbol{\theta}}(x)d\mu(x) = 0,\ \forall i,j$，则$I_{ij}(\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log p_{\boldsymbol{\theta}}(X)\right]$。
- 若$\theta_i = h_i(\xi_1,\ldots,\xi_s)$，$i = 1,\ldots,s$，$\boldsymbol{\theta} \overset{1:1}{\longleftrightarrow} \boldsymbol{\xi}$，则 $I^*(\boldsymbol{\xi}) = J^\tau \cdot I(\boldsymbol{\theta}(\boldsymbol{\xi}))\cdot J$，其中
$$J = \left(\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\xi}^\tau}\right) = \begin{pmatrix} \frac{\partial\theta_1}{\partial\xi_1} & \cdots & \frac{\partial\theta_1}{\partial\xi_s} \\ \vdots & \ddots & \vdots \\ \frac{\partial\theta_s}{\partial\xi_1} & \cdots & \frac{\partial\theta_s}{\partial\xi_s}\end{pmatrix}.$$
- (独立可加性) 设$X \sim p_{\boldsymbol{\theta}}(\cdot)$（关于$\mu$），$Y \sim p_{\boldsymbol{\theta}}(\cdot)$（关于$\nu$），且$X, Y$ 独立。若X, Y 中$\boldsymbol{\theta}$ 的信息矩阵分别为$I_1(\boldsymbol{\theta})$，$I_2(\boldsymbol{\theta})$。则(X,Y) 中$\boldsymbol{\theta}$ 的信息矩阵为$I_1(\boldsymbol{\theta}) + I_2(\boldsymbol{\theta})$。

例2.3.7 (指数族的信息矩阵) 设X 的分布为指数族，关于控制测度μ 具有pdf

$$p_{\boldsymbol{\theta}}(x) = B(\boldsymbol{\theta})\exp\{\sum_{i=1}^s \eta_i(\boldsymbol{\theta})T_i(x)\}h(x),\ \boldsymbol{\theta}\in\boldsymbol{\Theta},$$

其中$\boldsymbol{\Theta}$ 是$\mathbb{R}^s$ 中的开凸集，$B(\boldsymbol{\theta}), \eta_i(\boldsymbol{\theta}), i = 1,\cdots,s$ 连续可微。记$\tau_i = E_{\boldsymbol{\theta}}[T_i(x)],\ i = 1,\ldots,s$，$\boldsymbol{\tau} = (\tau_1,\ldots,\tau_s)'$。若$\boldsymbol{\tau} \overset{1:1}{\longleftrightarrow} \boldsymbol{\theta}$，则 $I(\boldsymbol{\tau}) = C^{-1}$，其中$C$ 为$(T_1,\ldots,T_s)$ 的协方差阵。

例2.3.8 (位置-尺度族参数的信息矩阵) 设X 关于Lebesgue 测度μ 具有pdf

$$p_{\boldsymbol{\theta}}(x) = \frac{1}{\theta_2}f\Big(\frac{x-\theta_1}{\theta_2}\Big),$$

其中$\boldsymbol{\theta} = (\theta_1,\theta_2)^\tau, \theta_2 > 0, \theta_1\in\mathbb{R}$，$f(x)$ 为一已知的pdf，对于$\forall x\in\mathbb{R}$，$f(x) > 0$ 且可导，则

$$\begin{aligned} I_{11}(\boldsymbol{\theta}) &= \frac{1}{\theta_2^2}\int_{-\infty}^{\infty}\Big[\frac{f'(y)}{f(y)}\Big]^2 f(y)d\mu(y), \\ I_{22}(\boldsymbol{\theta}) &= \frac{1}{\theta_2^2}\int_{-\infty}^{\infty}\Big[\frac{yf'(y)}{f(y)}+1\Big]^2 f(y)d\mu(y), \\ I_{12}(\boldsymbol{\theta}) &= \frac{1}{\theta_2^2}\int_{-\infty}^{\infty}y\Big[\frac{f'(y)}{f(y)}\Big]^2 f(y)d\mu(y). \end{aligned}$$

定理2.3.5 (C-R不等式) *在定义2.3.2 的条件下，若还满足 4'. $I(\boldsymbol{\theta})$ 正定。设δ 为一个估计量，满足$E_{\boldsymbol{\theta}}\delta^2<\infty$，且下列二者之一成立*

5'. $\frac{\partial}{\partial\theta_i}E_{\boldsymbol{\theta}}\delta,\ i=1,\ldots,s$ 存在，且求偏导与期望可交换次序。

5". $\exists b_{\boldsymbol{\theta}}^{(i)}, i=1,\ldots,s$，满足$E_{\boldsymbol{\theta}}[b_{\boldsymbol{\theta}}^{(i)}(X)]^2<\infty$；且$\exists\boldsymbol{\theta}$ 的一个邻域$U(\boldsymbol{\theta})$，使得

$$\left|\frac{p_{\boldsymbol{\theta}+\Delta e_i}(x)-p_{\boldsymbol{\theta}}(x)}{\Delta p_{\boldsymbol{\theta}}(x)}\right|\le b_{\boldsymbol{\theta}}^{(i)}(x),\ \forall\boldsymbol{\theta}+\Delta e_i\in U(\boldsymbol{\theta}),$$

其中e_i 为$\mathbb{R}^s$ 中的单位向量。

则

$$Var_{\boldsymbol{\theta}}(\delta)\ge\boldsymbol{\gamma}^\tau I^{-1}(\boldsymbol{\theta})\boldsymbol{\gamma},$$

其中$\boldsymbol{\gamma}=\left[\frac{\partial}{\partial\theta_1}E_{\boldsymbol{\theta}}\delta,\ldots,\frac{\partial}{\partial\theta_s}E_{\boldsymbol{\theta}}\delta\right]^\tau$。

证明. 记 $\psi_i(x;\boldsymbol{\theta})=\frac{\partial}{\partial\theta_i}\log p_{\boldsymbol{\theta}}(x),\ i=1,\ldots,s$。由定义2.3.2 知

$$I_{ij}(\boldsymbol{\theta})=\mathrm{Cov}_{\boldsymbol{\theta}}[\psi_i(X;\boldsymbol{\theta}),\psi_j(X;\boldsymbol{\theta})].$$

由正则条件3', 4', 5'或5"，以及$E_{\boldsymbol{\theta}}\delta^2<\infty$ 知，

$$\mathrm{Cov}_{\boldsymbol{\theta}}[\delta,\psi_i(X;\boldsymbol{\theta})]=\frac{\partial}{\partial\theta_i}E_{\boldsymbol{\theta}}\delta,\quad i=1,\cdots,s.$$

任取常向量$\boldsymbol{a}=(a_1,\ldots,a_s)^\tau\ne0$，由Cauchy-Schwarz不等式知，

$$\mathrm{Var}_\theta(\delta)\ge\frac{[\mathrm{Cov}(\delta,\sum_{i=1}^s a_i\psi_i(X;\boldsymbol{\theta}))]^2}{\mathrm{Var}(\sum_{i=1}^s a_i\psi_i(X;\boldsymbol{\theta}))}=\frac{(\boldsymbol{a}^\tau\boldsymbol{\gamma})^2}{\boldsymbol{a}^\tau I(\boldsymbol{\theta})\boldsymbol{a}}.$$

因此，

$$\mathrm{Var}_\theta(\delta)\ge\max_{\boldsymbol{a}\ne0}\frac{\boldsymbol{a}^\tau\boldsymbol{\gamma}\boldsymbol{\gamma}^\tau\boldsymbol{a}}{\boldsymbol{a}^\tau I(\boldsymbol{\theta})\boldsymbol{a}}=\boldsymbol{\gamma}^\tau I(\boldsymbol{\theta})^{-1}\boldsymbol{\gamma}.$$

□

注：

(1) 若除θ_i 外$\boldsymbol{\theta}$ 的其余元素均已知，则由定理2.3.1知，δ 方差的C-R 下界为$\left[\frac{\partial}{\partial\theta_i}E_{\boldsymbol{\theta}}\delta\right]^2 I_{ii}^{-1}(\boldsymbol{\theta})$。这也正是定理2.3.5证明过程中取$\boldsymbol{a}=\boldsymbol{e}_i$ 的情况。因此，由定理2.3.5所得的方差下界$\boldsymbol{\gamma}^\tau I(\boldsymbol{\theta})^{-1}\boldsymbol{\gamma}$ 不小于由定理2.3.1所得的方差下界。同时可知，

$$I_{ii}^{-1}(\boldsymbol{\theta})\le[I^{-1}(\boldsymbol{\theta})]_{ii},$$

等号当且仅当$I(\boldsymbol{\theta})$ 为对角阵时成立，此时称未知参数之间是正交的。

(2) 当且仅当δ（不一定是统计量）具有形式

$$\delta(X) = g(\boldsymbol{\theta}) + [\nabla g(\boldsymbol{\theta})]^\tau I^{-1}(\boldsymbol{\theta})[\nabla \log p_{\boldsymbol{\theta}}(X)]$$

时，C-R 下界可达，且此时相应的分布族必定是指数族，其中$E_{\boldsymbol{\theta}}\delta = g(\boldsymbol{\theta})$，

$$\nabla g(\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial\theta_1}g(\boldsymbol{\theta}), \ldots, \frac{\partial}{\partial\theta_s}g(\boldsymbol{\theta})\Big)^\tau,$$
$$\nabla \log p_{\boldsymbol{\theta}}(x) = \Big(\frac{\partial}{\partial\theta_1}\log p_{\boldsymbol{\theta}}(x), \ldots, \frac{\partial}{\partial\theta_s}\log p_{\boldsymbol{\theta}}(x)\Big)^\tau.$$

(3) 根据定理2.3.3的思路，当$C-R$下界达不到时，还可考虑采用对数密度函数的高阶导数来改进$C-R$ 下界。记

$$\psi_{i_1,\ldots,i_s} = \frac{1}{p_\theta(x)}\frac{\partial^{i_1+\ldots+i_s}p_\theta(x)}{\partial\theta_1^{i_1}\cdots\partial\theta_s^{i_s}},$$

设$\mathbb{S}$ 为若干$\psi_{i_1,\ldots,i_s}$ 的集合，其中的每个元素均满足正则条件 3'，且它们的协方差阵$K(\boldsymbol{\theta})$ 正定，则由类似于定理2.3.3 或定理2.3.5 的证明方法即可得下列Bhattacharyya不等式

$$\mathrm{Var}_\theta\delta \geq \boldsymbol{\alpha}' K^{-1}(\theta)\boldsymbol{\alpha},$$

其中$\boldsymbol{\alpha}$ 的元素为

$$\frac{\partial^{i_1+\ldots+i_s}}{\partial\theta_1^{i_1}\cdots\partial\theta_s^{i_s}}E_\theta\delta = \mathrm{Cov}_{\boldsymbol{\theta}}(\delta, \psi_{i_1,\ldots,i_s}).$$