CH03 点估计

更新：2026 年 1 月 9 日

目录

1 点估计 **2**

1.1 点估计的含义 2

1.2 无偏性 2

1.3 有效性 3

1.4 相合性 3

2 矩估计 MM **4**

2.1 矩法和矩估计 4

2.2 矩估计的无偏性和渐近无偏性 5

2.3 矩估计的相合性 5

2.4 矩估计的渐近正态性 7

3 极大似然估计 MLE **8**

3.1 定义及求解 8

3.1.1 极大似然估计的定义 8

3.1.2 MLE 求解：根据微积分中求极值 9

3.1.3 MLE 求解：根据定义 11

3.2 极大似然估计的性质 11

3.2.1 极大似然估计的无偏性和相合性 11

3.2.2 极大似然估计与充分统计量 12
3.2.3 极大似然估计的相合渐近正态性 12
4 一致最小均方误差估计 **14**
5 一致最小方差无偏估计 UMVUE **15**
5.1 充分完全统计量法 17
5.2 零无偏估计法 20
5.3 Cramer-Rao 不等式 22
5.3.1 单参数 C-R 不等式 22
5.3.2 C-R 不等式等号成立条件 24
5.3.3 多参数 C-R 不等式 25
5.3.4 有效估计和估计的效率 26

1 点估计

1.1 点估计的含义

定义 1.1 (点估计) 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从某个总体中抽取的样本，$\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\cdots,X_n)$ 是样本的函数，用 $\hat{g}(\boldsymbol{X})$ 作为参数的函数 $g(\theta)$ 的估计，称为**点估计**（point estimation）。

1.2 无偏性

定义 1.2 (无偏性) 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从总体 $\{f(x,\theta):\theta\in\Theta\}$ 中抽取的样本，$g(\theta)$ 是定义在参数空间 Θ 上的已知函数。$\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的一个估计量，若

$$\mathbb{E}_\theta\{\hat{g}(\boldsymbol{X})\}=g(\theta),\quad \theta\in\Theta$$

则称 $\hat{g}(\boldsymbol{X})$ 为 $g(\theta)$ 的一个**无偏估计**(unbiased estimation)，记 $\hat{g}(\boldsymbol{X})=\hat{g}_n(\boldsymbol{X})$。若 $\mathbb{E}_\theta\{\hat{g}_n(\boldsymbol{X})\}\neq g(\theta)$ 但

$$\lim_{n\to\infty}\mathbb{E}_\theta\{\hat{g}_n(\boldsymbol{X})\}=g(\theta),\quad \theta\in\Theta$$

则称 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的**渐近无偏估计**（asymptotically unbiased estimation）。

例 1.1 设 $X_1,\cdots,X_n$ 是取自期望为 μ，方差为 σ^2 的总体的一个样本。显然样本均值 $\bar{X}$ 是 μ 的无偏估计。证明：样本方差 $S^2=\sum_{i=1}^n(X_i-\bar{X})^2/(n-1)$ 是 σ^2 的无偏估计。

证明 显然

$$\begin{aligned}\mathbb{E}(S^2) &= \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathbb{E}(X_i^2) - n\mathbb{E}(\bar{X}^2)\right] = \frac{n}{n-1}[\mathbb{E}(X_1^2) - \mathbb{E}(\bar{X}^2)] \\ &= \frac{n}{n-1}[(\sigma^2+\mu^2) - (\sigma^2/n + \mu^2)] = \sigma^2\end{aligned}$$

故 S^2 是 σ^2 的无偏估计。

1.3 有效性

定义 1.3 (有效性) 设 $\hat{g}_1(\boldsymbol{X})$ 和 $\hat{g}_2(\boldsymbol{X})$ 为 $g(\theta)$ 的两个不同的无偏估计，若

$$\mathbb{D}_\theta\{\hat{g}_1(\boldsymbol{X})\} \le \mathbb{D}_\theta\{\hat{g}_2(\boldsymbol{X})\}, \quad \forall\, \theta \in \Theta$$

且至少存在一个 $\theta \in \Theta$ 使得不等号严格成立，则称 $\hat{g}_1(\boldsymbol{X})$ 比 $\hat{g}_2(\boldsymbol{X})$ 有效。

1.4 相合性

定义 1.4 对 $n \in \mathbb{N}$，$\hat{g}_n(\boldsymbol{X})$ 是 $g(\theta)$ 的一个估计量，若 $\hat{g}_n(\boldsymbol{X})$ 依概率收敛到 $g(\theta)$，即对任意 $\theta \in \Theta$ 及 $\epsilon > 0$ 有

$$\lim_{n\to\infty} P_\theta(|\hat{g}_n(\boldsymbol{X}) - g(\theta)| \ge \epsilon) = 0$$

则称 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的**弱相合估计**（weakly consistent estimation）。若对任意 $\theta \in \Theta$ 有

$$P_\theta\left(\lim_{n\to\infty} \hat{g}_n(\boldsymbol{X}) = g(\theta)\right) = 1$$

即几乎处处收敛，则称 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的**强相合估计**（strongly consistent estimation）。

注 无偏性是小样本性质；相合性和渐近正态性是大样本性质。无偏 $\not\Rightarrow$ 相合；有偏也可以相合 (e.g. $\bar{X} + 1/n$)。强相合 $\Rightarrow$ 弱相合，反之不必对。

例 1.2 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从均匀分布 $U(0,\theta)$ 中抽取的简单随机样本，θ 为未知参数。证明：$T(\boldsymbol{X}) = (\prod_{i=1}^n X_i)^{1/n}$ 是 $g(\theta) = \theta e^{-1}$ 的强相合估计。

证明 令 $Y_i = \log X_i$ 则 Y_i 相互独立，且联合密度为

$$f(y,\theta) = \frac{1}{\theta}e^y \cdot \mathbb{I}_{(-\infty,\log\theta)}(y)$$

且

$$\mathbb{E}(Y_1) = \frac{1}{\theta}\int_{-\infty}^{\log\theta} ye^y\,dy = \frac{1}{\theta}\left[ye^y\Big|_{-\infty}^{\log\theta} - \int_{-\infty}^{\log\theta} e^y\,dy\right] = \log\theta - 1$$

由强大数定律可知

$$\log T(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^n \log X_i = \frac{1}{n}\sum_{i=1}^n Y_i = \bar{Y} \xrightarrow{\text{a.s.}} \mathbb{E}(Y_1) = \log\theta - 1$$

因此有

$$T(\boldsymbol{X}) = \exp(\bar{Y}) \xrightarrow{\text{a.s.}} \exp(\log\theta - 1) = \theta e^{-1}$$

故，$T(\boldsymbol{X}) = (\prod_{i=1}^n X_i)^{1/n}$ 是 $g(\theta) = \theta e^{-1}$ 的强相合估计。

2 矩估计 MM

2.1 矩法和矩估计

定义 2.1 (原点矩和中心矩) 设样本 $X_1, \cdots, X_n$ 是从总体 X 中抽取的简单随机样本。

- 总体 k 阶原点矩为
$$\alpha_k = \mathbb{E}\{X^k\}$$
- 样本 k 阶原点矩为
$$a_{nk} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$
- 总体 k 阶中心矩为
$$\mu_k = \mathbb{E}\{[X - \mathbb{E}(X)]^k\}$$
- 样本 k 阶中心矩为
$$m_{nk} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

定义 2.2 (矩估计) 设有总体分布族 $\{f(x;\theta) : \theta \in \Theta\}$，$\Theta$ 是参数空间，$g(\theta)$ 是定义在 Θ 上参数 θ 的函数，它可以表示为总体分布的某些矩的函数，即
$$g(\theta) = G(\alpha_1, \cdots, \alpha_k; \mu_2, \cdots, \mu_s)$$
设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从上述分布族中抽取的简单样本，用样本矩代替总体矩得到
$$\hat{g}(\boldsymbol{X}) = G(a_{n1}, \cdots, a_{nk}; m_{n2}, \cdots, m_{ns})$$
则 $\hat{g}(\boldsymbol{X})$ 作为 $g(\theta)$ 的估计量，称为 $g(\theta)$ 的矩估计量（moment estimate）。

例 2.1 (Maxwell 分布) 设总体分布有概率密度
$$f(x, \theta) = 2\sqrt{\theta/\pi}\exp(-\theta x^2) \cdot \mathbb{I}(x > 0)$$
其中 θ 为未知参数。设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 为从这个分布中抽取的简单随机样本，求 $g(\theta) = 1/\theta$ 的矩估计量。

解答 求总体一阶矩 α_1
$$\alpha_1 = \mathbb{E}(X) = 2\sqrt{\frac{\theta}{\pi}}\int_0^{\infty} xe^{-\theta x^2}\,dx = \frac{1}{\sqrt{\pi\theta}}$$
解得 $g(\theta) = 1/\theta = \pi\alpha_1^2$，将 α_1 用 $a_{n1} = \bar{X}$ 代替，有矩估计 $\hat{g}(\theta) = \hat{g}_1(\boldsymbol{X}) = \pi\bar{X}^2$。

类似地，求总体二阶矩
$$\alpha_2 = \mathbb{E}(X^2) = 2\sqrt{\frac{\theta}{\pi}}\int_0^{\infty} x^2e^{-\theta x^2}\,dx = \frac{1}{2\theta}$$
所以可以用 $a_{n2} = \sum_{i=1}^{n} X_i^2/n$ 得到另一个矩估计 $\hat{g}(\theta) = \hat{g}_2(\boldsymbol{X}) = 2a_{n2}$。

由此可知，矩估计不唯一。但注意到

$$\mathbb{E}\{\hat{g}_1(\boldsymbol{X})\}=\pi\mathbb{E}\{\bar{X}^2\}=\pi(\mathrm{Var}(\bar{X})+(\mathbb{E}(\bar{X}))^2)=\pi\left(\frac{1}{n}\left(\frac{1}{2\theta}-\frac{1}{\pi\theta}\right)+\frac{1}{\pi\theta}\right)=\left(1-\frac{1}{n}+\frac{\pi}{2n}\right)\frac{1}{\theta}$$

不是无偏的，而

$$\mathbb{E}\{\hat{g}_2(\boldsymbol{X})\}=\frac{2}{n}\mathbb{E}\{\sum_{i=1}^{n}X_i^2\}=\frac{2}{n}n\frac{1}{2\theta}=\frac{1}{\theta}$$

为无偏估计。

2.2 矩估计的无偏性和渐近无偏性

命题 2.1 样本 k 阶原点矩 a_{nk} 是总体 k 阶原点矩 α_k 的无偏估计，即

$$\mathbb{E}\{a_{nk}\}=\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n}X_i^k\right\}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{X_i^k\}=\mathbb{E}\{X_1^k\}=\alpha_k$$

命题 2.2 对 $k\geq 2$，样本 k 阶中心矩 m_{nk} 不是总体 k 阶中心矩 μ_k 的无偏估计，需要调整

$$\begin{aligned}
&\mathbb{E}\{S^2\}=\mathbb{E}\left\{\frac{n}{n-1}\cdot m_{n2}\right\}=\mu_2\\
&\mathbb{E}\{m_{n3}^*\}=\mathbb{E}\left\{\frac{n^2}{(n-1)(n-2)}\cdot m_{n3}\right\}=\mu_3\\
&\mathbb{E}\{m_{n\nu}\}=\mu_\nu+O\left(1/n\right),\quad \forall\,\nu\geq 4
\end{aligned}$$

即矩估计一般具有渐近无偏性。

2.3 矩估计的相合性

定理 2.3 矩估计均有强相合性，即 $a_{nk}\xrightarrow{\text{a.s.}}\alpha_k$ 以及 $m_{nk}\xrightarrow{\text{a.s.}}\mu_k$

证明 由强大数定律可知

$$a_{nk}=\frac{1}{n}\sum\nolimits_{i=1}^{n}X_i^k\xrightarrow{\text{a.s.}}\frac{1}{n}\cdot n\mathbb{E}(X_1^k)=\alpha_k$$

下面证明 m_{nk} 的相合性。由

$$\begin{aligned}
m_{nk}&=\frac{1}{n}\sum\nolimits_{i=1}^{n}(X_i-\bar{X})^k=\frac{1}{n}\sum\nolimits_{i=1}^{n}\sum\nolimits_{r=0}^{k}\binom{k}{r}X_i^r(-\bar{X})^{k-r}\\
&=\frac{1}{n}\sum\nolimits_{r=0}^{k}\binom{k}{r}(-1)^{k-r}\bar{X}^{k-r}\sum\nolimits_{i=1}^{n}X_i^r\\
&=\sum\nolimits_{r=0}^{k}\binom{k}{r}(-1)^{k-r}\bar{X}^{k-r}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}X_i^r\right)\\
&=\sum\nolimits_{r=0}^{k}\binom{k}{r}(-1)^{k-r}a_{n1}^{k-r}a_{nr},\quad a_{n0}=1\\
&:=f(a_{n1},\cdots,a_{nk})
\end{aligned}$$

而 $a_{nk}\xrightarrow{\text{a.s.}}\alpha_k$，由下面的引理 2.4 可知

$$m_{nk}=f(a_{n1},\cdots,a_{nk})\xrightarrow{\text{a.s.}}f(\alpha_1,\cdots,\alpha_k)$$

又注意到

$$\begin{aligned}\mu_k=\mathbb{E}\{[X-\mathbb{E}(X)]^k\}&=\mathbb{E}\left\{\sum_{r=0}^{k}\binom{k}{r}X^r(-\mathbb{E}(X))^{k-r}\right\}\\&=\sum_{r=0}^{k}\binom{k}{r}(-1)^{k-r}\mathbb{E}\{X^r\}\mathbb{E}(X)^{k-r}\\&=\sum_{r=0}^{k}\binom{k}{r}(-1)^{k-r}\alpha_1^{k-r}\alpha_r\\&=f(\alpha_1,\cdots,\alpha_k)\end{aligned}$$

所以 $m_{nk}\xrightarrow{\text{a.s.}}\mu_k$，命题证毕。

引理 2.4 (*) 设函数 $f(y_{n1},\cdots,y_{nk})$ 在 $(c_1,\cdots,c_k)$ 处连续，若 $y_{ni}\xrightarrow{\text{a.s.}}c_i$ 对任意 $i=1,\cdots,k$ 成立，则有 $f(y_{n1},\cdots,y_{nk})\xrightarrow{\text{a.s.}}f(c_1,\cdots,c_k)$。

证明 (*) 由 f 在 $(c_1,\cdots,c_k)$ 处连续，故有 $\forall\,\epsilon>0,\ \exists\,\delta>0$ 使得对 $\forall\,(y_1,\cdots,y_k)$

$$\max_{1\le i\le k}|y_i-c_i|<\delta\quad\Rightarrow\quad|f(y_1,\cdots,y_k)-f(c_1,\cdots,c_k)|<\epsilon \tag{1}$$

对每个 i 和固定的 δ 构造集合

$$B_{i,\delta}=\{\omega:\exists\,N\in\mathbb{N},\ \text{s.t.}\ \forall\,n\ge N,\ |y_{ni}(\omega)-c_i|<\delta\}\subseteq\Omega$$

其中 Ω 为全集。$B_{i,\delta}$ 代表了所有满足下面条件的 ω 构成的集合：

$$\exists\,N\in\mathbb{N},\ \text{s.t.}\ \forall\,n\ge N,\ |y_{ni}(\omega)-c_i|<\delta\quad\Leftrightarrow\quad\lim_{n\to\infty}y_{ni}(\omega)=c_i$$

即 $B_{i,\delta}=\{\omega:\lim_{n\to\infty}y_{ni}(\omega)=c_i\}$。由 $y_{ni}\xrightarrow{\text{a.s.}}c_i$ 对任意 $i=1,\cdots,k$ 成立，有

$$P(\{\omega:\lim_{n\to\infty}y_{ni}=c_i\})=1,\quad i=1,\cdots,k$$

故 $P(B_{i,\delta})=1$ 对任意 $i=1,\cdots,k$。取有限交 $B_\delta=\cap_{i=1}^{k}B_{i,\delta}$，依旧有 $P(B_\delta)=1$。换言之，对任意 $\omega\in B_\delta$ 都有 $\exists\,N_i(\omega)\in\mathbb{N}$ 使得 $\forall\,n\ge N_i(\omega)$ 时 $|y_{ni}(\omega)-c_i|<\delta$，对 $i=1,\cdots,k$。于是，我们取

$$N(\omega)=\max_{1\le i\le k}N_i(\omega)$$

那么对于 $\forall\,n\ge N(\omega)$ 都有 $|y_{ni}(\omega)-c_i|<\delta$，即 $\max_{1\le i\le k}|y_{ni}(\omega)-c_i|<\delta$。由式 1 可知

$$|f(y_{n1}(\omega),\cdots,y_{nk}(\omega))-f(c_1,\cdots,c_k)|<\epsilon \tag{2}$$

综上，对 $\forall\,\epsilon>0$ 都存在 $\delta>0$，使得对 $\forall\,\omega\in B_\delta$ 存在 $N(\omega)$，对 $\forall\,n\ge N(\omega)$ 有式 2 成立。而 $\omega\in B_\delta$ 满足 $P(B_\delta)=1$，即由几乎处处收敛的定义 $\Rightarrow f(y_{n1},\cdots,y_{nk})\xrightarrow{\text{a.s.}}f(c_1,\cdots,c_k)$。

定理 2.5 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从总体 F 中抽取的简单随机样本，待估函数 $g(\theta)=G(\alpha_1,\cdots,\alpha_k,\mu_2,\cdots,\mu_s)$，其矩估计为 $\hat{g}_n(\boldsymbol{X})=G(a_{n1},\cdots,a_{nk},m_{n2},\cdots,m_{ns})$，且 G 为其变元的连续函数，则 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的强相合估计。

2.4 矩估计的渐近正态性

定义 2.3 (相合渐近正态估计) 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从总体 $\{f(x;\theta) : \theta \in \Theta\}$ 中抽取的简单随机样本。若存在与样本空间大小 n 相关的，定义于参数空间 Θ 上的函数 $A_n(\theta)$ 和 $B_n(\theta)$，其中 $B_n(\theta) > 0, \forall\, \theta \in \Theta$，使得

$$\frac{\hat{g}_n(\boldsymbol{X}) - A_n(\theta)}{B_n(\theta)} \xrightarrow{\mathscr{L}} N(0,1), \quad \text{as } n \to \infty$$

且 $\hat{g}_n(\boldsymbol{X})$ 是 $g(\theta)$ 的弱相合估计，则称 $\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的**相合渐近正态估计**（consistent asymptotic normal estimation，CAN 估计）。

定理 2.6 (δ 方法) 设映射 $\phi : D \subseteq \mathbb{R}^k \to S \subseteq \mathbb{R}^m$ 在 θ 处可微。若 $\exists\, r_n \to \infty$ 有

$$r_n(T_n - \theta) \xrightarrow{\mathscr{L}} T$$

则有

$$r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{\mathscr{L}} \phi'(\theta) \cdot T$$

其中 $\phi'(\theta) \in \mathbb{R}^{m \times k}$ 为

$$\phi'(\theta) = \left(\frac{\partial \phi_i}{\partial \theta_j}\right)_{i=1,\cdots,m;\ j=1,\cdots,k} \in \mathbb{R}^{m \times k}$$

证明 下面给出不严格的证明：由于 ϕ 在 θ 处可微，由 Taylor 展开

$$\phi(T_n) \approx \phi(\theta) + \phi'(\theta)(T_n - \theta)$$

所以有 $r_n(\phi(T_n) - \phi(\theta)) = \phi'(\theta) \cdot r_n(T_n - \theta) \xrightarrow{\mathscr{L}} \phi'(\theta) \cdot T$，命题证毕。

定理 2.7 (矩估计的近似正态性) 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从总体 $\{f(x;\theta) : \theta \in \Theta\}$ 中抽取的简单随机样本，$g(\theta)$ 是定义在 Θ 上的实函数，它可以表示为 $g(\theta) = G(\alpha_1, \cdots, \alpha_k)$，这是因为 μ_s 可由 $(\alpha_1, \cdots, \alpha_k)$ 表出。记 $\hat{g}_n(\boldsymbol{X}) = G(a_{n1}, \cdots, a_{nk})$ 为 $g(\theta)$ 的矩估计。再设总体的 $2k$ 阶原点矩存在，且 G 对其各变元的一阶偏导数存在连续，令

$$\begin{aligned} \boldsymbol{B} &= (b_{ij})_{i,j=1,\cdots,k} \in \mathbb{R}^{k \times k} \\ \boldsymbol{d} &= (d_1, \cdots, d_k)^T \in \mathbb{R}^k \end{aligned}$$

其中 $b_{ij} = \alpha_{i+j} - \alpha_i \alpha_j \ (i,j = 1, \cdots, k)$ 及 $d_i = \partial G(\alpha_1, \cdots, \alpha_k)/\partial \alpha_i \ (i = 1, \cdots, k)$。

结论：$\hat{g}_n(\boldsymbol{X})$ 为 $g(\theta)$ 的 CAN 估计，即 $\hat{g}_n(\theta)$ 为 $g(\theta)$ 的弱相合估计，且

$$\sqrt{n}(\hat{g}_n(\boldsymbol{X}) - g(\theta)) \xrightarrow{\mathscr{L}} N(0, b^2) \quad n \to \infty$$

其中 $b^2 = \boldsymbol{d}^T \boldsymbol{B} \boldsymbol{d} \in \mathbb{R}$。

证明 由定理 2.3 可知，矩估计都是强相合估计，那必是弱相合估计。下面证明渐近正态性。

由 $\mathbb{E}\{a_{ni}\} = \alpha_i,\ \forall\, i = 1, \cdots, k$，所以由中心极限定理

$$\sqrt{n}(\boldsymbol{a} - \boldsymbol{\alpha}) \xrightarrow{\mathscr{L}} T \sim N(0, \boldsymbol{A})$$

其中 $\boldsymbol{a} = (a_{n1}, \cdots, a_{nk})^T \in \mathbb{R}^k$ 而 $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k)^T \in \mathbb{R}^k$。而 $\boldsymbol{A} = n[\mathrm{Cov}(a_{ni}, a_{nj})]_{i,j=1,\cdots,k} \in \mathbb{R}^{k\times k}$ 为协方差阵。那么由 δ 方法 2.6

$$\sqrt{n}[G(\boldsymbol{a}) - G(\boldsymbol{\alpha})] \xrightarrow{\mathscr{L}} [G'(\boldsymbol{\alpha})]^T \cdot T \sim N(0, [G'(\boldsymbol{\alpha})]^T \boldsymbol{A}[G'(\boldsymbol{\alpha})])$$

注意到 $\boldsymbol{a} = (a_{n1}, \cdots, a_{nk})^T$ 实则是 $\boldsymbol{X}$ 的函数，故 $G(\boldsymbol{a}) = \hat{g}_n(\boldsymbol{X})$。而 $g(\theta) = G(\alpha_1, \cdots, \alpha_k) = G(\boldsymbol{\alpha})$，故有

$$\sqrt{n}(\hat{g}_n(\boldsymbol{X}) - g(\theta)) \xrightarrow{\mathscr{L}} N(0, [G'(\boldsymbol{\alpha})]^T \boldsymbol{A}[G'(\boldsymbol{\alpha})])$$

于是，我们只需证 $[G'(\boldsymbol{\alpha})]^T \boldsymbol{A}[G'(\boldsymbol{\alpha})] = b^2$。首先注意到

$$G'(\boldsymbol{\alpha}) = [\partial G(\alpha_1, \cdots, \alpha_k)/\partial \alpha_i]^T_{i=1,\cdots,k} = (d_1, \cdots, d_k)^T = \boldsymbol{d}$$

所以 $[G'(\boldsymbol{\alpha})]^T \boldsymbol{A}[G'(\boldsymbol{\alpha})] = \boldsymbol{d}^T \boldsymbol{A} \boldsymbol{d}$，于是我们只需证 $\boldsymbol{A} = \boldsymbol{B}$ 即可。注意到

$$\begin{aligned}
\mathrm{Cov}(a_{ni}, a_{nj}) &= \mathbb{E}(a_{ni}a_{nj}) - \mathbb{E}(a_{ni})\mathbb{E}(a_{nj}) = \mathbb{E}\left\{\left(\frac{1}{n}\sum\nolimits_{l=1}^n X_l^i\right) \cdot \left(\frac{1}{n}\sum\nolimits_{l=1}^n X_l^j\right)\right\} - \alpha_i\alpha_j \\
&= \frac{1}{n^2}\mathbb{E}\left\{\sum\nolimits_{l_1=1}^n \sum\nolimits_{l_2=1}^n X_{l_1}^i X_{l_2}^j\right\} - \alpha_i\alpha_j \\
&= \frac{1}{n^2}\mathbb{E}\left\{\sum\nolimits_{l_1=l_2=l}^n X_l^{i+j} + \sum\nolimits_{l_1\neq l_2} X_{l_1}^i X_{l_2}^j\right\} - \alpha_i\alpha_j
\end{aligned}$$

由于当 $l_1 \neq l_2$ 时 $X_{l_1} \perp\!\!\!\perp X_{l_2}$ 故

$$\begin{aligned}
\mathrm{Cov}(a_{ni}, a_{nj}) &= \frac{1}{n^2}\left[\sum\nolimits_{l_1=l_2=l}^n \mathbb{E}(X_l^{i+j}) + \sum\nolimits_{l_1\neq l_2} \mathbb{E}(X_{l_1}^i)\mathbb{E}(X_{l_2}^j)\right] - \alpha_i\alpha_j \\
&= \frac{1}{n^2}\left[\sum\nolimits_{l_1=l_2=l}^n \alpha_{i+j} + \sum\nolimits_{l_1\neq l_2} \alpha_i\alpha_j\right] - \alpha_i\alpha_j \\
&= [n\alpha_{i+j} + n(n-1)\alpha_i\alpha_j]/n^2 - \alpha_i\alpha_j = [\alpha_{i+j} - \alpha_i\alpha_j]/n
\end{aligned}$$

所以有 $\boldsymbol{A} = n[\mathrm{Cov}(a_{ni}, a_{nj})]_{i,j=1,\cdots,k} = (\alpha_{i+j} - \alpha_i\alpha_j)_{i,j=1,\cdots,k} = \boldsymbol{B}$，综上命题证毕。

3 极大似然估计 MLE

3.1 定义及求解

3.1.1 极大似然估计的定义

定义 3.1 (似然函数) 设 $f(\boldsymbol{x};\boldsymbol{\theta}) = f(x_1, \cdots, x_n;\boldsymbol{\theta})$ 为样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 的概率函数。当 $\boldsymbol{x}$ 固定时，将 $f(\boldsymbol{x};\boldsymbol{\theta})$ 看成 $\boldsymbol{\theta}$ 的函数，称为似然函数（likelihood function）。记为

$$L(\boldsymbol{\theta};\boldsymbol{x}) = f(\boldsymbol{x};\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta,\ \boldsymbol{x} \in \mathscr{X}$$

其中 Θ 为参数空间，$\mathscr{X}$ 为样本空间。称 $\log L(\boldsymbol{\theta};\boldsymbol{x})$ 为对数似然函数，记作 $l(\boldsymbol{\theta};\boldsymbol{x})$。

定义 3.2 (极大似然估计) 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从总体 $\mathscr{F}=\{f(\boldsymbol{x},\boldsymbol{\theta}):\boldsymbol{\theta}\in\Theta\}$ 中抽取的简单随机样本，$L(\boldsymbol{\theta};\boldsymbol{x})$ 是似然函数，若存在统计量 $\hat{\boldsymbol{\theta}}^*=\hat{\boldsymbol{\theta}}^*(\boldsymbol{X})$，满足条件

$$L(\hat{\boldsymbol{\theta}}^*;\boldsymbol{x})=\sup_{\boldsymbol{\theta}\in\Theta}L(\boldsymbol{\theta};\boldsymbol{x}),\quad \boldsymbol{x}\in\mathscr{X}$$

或等价地

$$l(\hat{\boldsymbol{\theta}}^*;\boldsymbol{x})=\sup_{\boldsymbol{\theta}\in\Theta}l(\boldsymbol{\theta};\boldsymbol{x}),\quad \boldsymbol{x}\in\mathscr{X}$$

则称 $\hat{\boldsymbol{\theta}}^*(\boldsymbol{X})$ 为 $\boldsymbol{\theta}$ 的**极大似然估计**（maximum likelihood estimation，MLE）。若待估函数是 $g(\boldsymbol{\theta})$，则定义 $g(\hat{\boldsymbol{\theta}}^*(\boldsymbol{X}))$ 为 $g(\boldsymbol{\theta})$ 的 MLE。

估计 MLE 可以采用 2 种方法：(1) 根据微积分中求极值；(2) 根据极大似然估计的定义。

3.1.2 MLE 求解：根据微积分中求极值

(1) 根据微积分中求极值的方法：设 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)^T\in\mathbb{R}^k$ 为参数向量，若 $l(\boldsymbol{\theta};\boldsymbol{x})$ 的极大值在参数空间 Θ 的内点处（而非边界点）达到，则此点必满足

$$\frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial\theta_i}=0,\quad i=1,\cdots,k$$

除此之外，还需要验证 Hessian 阵的负定性，即

$$\boldsymbol{H}(\boldsymbol{\theta})=\left(\frac{\partial^2 l(\boldsymbol{\theta};\boldsymbol{x})}{\partial\theta_i\partial\theta_j}\right)_{i,j=1,\cdots,k}\in\mathbb{R}^{k\times k}$$

满足 $\boldsymbol{v}^T\boldsymbol{H}\boldsymbol{v}<0$ 对任意 $\boldsymbol{v}\in\mathbb{R}^k$。

定理 3.1 (指数族的 MLE) 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从指数族中抽取的简单随机样本，

$$f(\boldsymbol{x};\boldsymbol{\theta})=C(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^k\theta_iT_i(\boldsymbol{x})\right\}h(\boldsymbol{x}),\quad \boldsymbol{\theta}\in\Theta$$

设 Θ_0 是 Θ 的内点集，记 $L(\boldsymbol{\theta};\boldsymbol{x})$ 是其似然函数，$l(\boldsymbol{\theta};\boldsymbol{x})$ 是对应的对数似然函数。要求 $(T_1,\cdots,T_k)$ 以概率为 1 线性独立。则有若

$$\frac{\partial l(\boldsymbol{\theta};\boldsymbol{x})}{\partial\theta_i}=0,\quad i=1,\cdots,k$$

在 Θ_0 中有解 $\hat{\boldsymbol{\theta}}$，则其必唯一，且为 $\boldsymbol{\theta}$ 的 MLE。

证明 先证明解的唯一性，然后证明其为最优的。

(1) 唯一性 若设 $\boldsymbol{\theta}_0$ 和 $\boldsymbol{\theta}_1$ 为满足 $l'(\boldsymbol{\theta})=0$ 的 2 个不同的解。由之前证明的性质：指数族参数空间的内点集 Θ_0 必是凸集，有

$$\hat{\boldsymbol{\theta}}\triangleq t\cdot\boldsymbol{\theta}_0+(1-t)\cdot\boldsymbol{\theta}_1\in\Theta_0,\quad t\in[0,1]$$

记 $l(t\boldsymbol{\theta}_0+(1-t)\boldsymbol{\theta}_1)=H(t)$，注意到 $l'(\boldsymbol{\theta}_0)=0$，$l'(\boldsymbol{\theta}_1)=0$，以及 $H'(t)=(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T l'(\hat{\boldsymbol{\theta}})$。其中 $l'(\hat{\boldsymbol{\theta}})\in\mathbb{R}^k$。于是有 $H'(0)=H'(1)=0$。由 Rolle 定理 3.2，存在 $t_0\in(0,1)$ 使得 $H''(t_0)=0$。

而 $H''(t)=(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T l''(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)$，其中 $l''(\hat{\boldsymbol{\theta}})\in\mathbb{R}^{k\times k}$。由之前关于指数族性质的证明可知

$$l''(\hat{\boldsymbol{\theta}})=n\left[\frac{\partial^2\log L(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right]_{i,j=1,\cdots,k}=n\left[\frac{\partial\log C(\boldsymbol{\theta})}{\partial\theta_i\theta_j}\right]_{i,j=1,\cdots,k}=n\left[-\mathrm{Cov}(T_i,T_j)\right]_{i,j=1,\cdots,k}\in\mathbb{R}^{k\times k}$$

所以容易得到 $\forall\, t$ 都有 $l''(\hat{\boldsymbol{\theta}})=l''(t\boldsymbol{\theta}_0+(1-t)\boldsymbol{\theta}_1)=n\left[-\mathrm{Cov}(T_i,T_j)\right]_{i,j=1,\cdots,k}\prec\mathbf{0}$。这是因为协方差阵总是半正定，而这里有 $(T_1,\cdots,T_k)$ 以概率为 1 线性独立。

于是若 $\boldsymbol{\theta}_0\neq\boldsymbol{\theta}_1$，则 $H''(t)<0$, $\forall\, t\in(0,1)$。与存在 $t_0\in(0,1)$ 使得 $H''(t_0)=0$ 矛盾。故 $\boldsymbol{\theta}_0=\boldsymbol{\theta}_1$，唯一性证毕。

(2) MLE 设 $\boldsymbol{\theta}_0$ 为唯一解，使得 $l'(\boldsymbol{\theta}_0)=0$。那么对于 $\forall\,\tilde{\boldsymbol{\theta}}\in\Theta$，沿用 (1) 对 $H(t)=l(t\boldsymbol{\theta}_0+(1-t)\tilde{\boldsymbol{\theta}})$ 的定义。

有 $H'(1)=0$（注 $H'(0)$ 的性质不知），以及 $H''(t)<0$ 对 $t\in(0,1)$。于是 H' 在 $(0,1)$ 单调递减，那么 $H'(t)>H'(1)=0$, $\forall\, t\in(0,1)$。于是 H 在 $(0,1)$ 单调递增，那么 $H(1)>H(0)$，即 $l(\boldsymbol{\theta}_0)>l(\tilde{\boldsymbol{\theta}})$ 对任意 $\tilde{\boldsymbol{\theta}}\in\Theta$，即唯一解 $\boldsymbol{\theta}_0$ 是 MLE。

注 对于随机向量 $\boldsymbol{X}=(X_1,\cdots,X_k)\in\mathbb{R}^k$，记 $\boldsymbol{\mu}=\mathbb{E}\{\boldsymbol{X}\}$，则有 $\mathrm{Cov}(\boldsymbol{X})\succeq 0$ 即协方差阵半正定。

证明 对任意向量 $\boldsymbol{v}\in\mathbb{R}^k$，有

$$\begin{aligned}\boldsymbol{v}^T\mathrm{Cov}(\boldsymbol{X})\boldsymbol{v}&=\boldsymbol{v}^T\mathbb{E}\{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\}\boldsymbol{v}=\mathbb{E}\{\boldsymbol{v}^T(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{v}\}\\&=\mathbb{E}\{[(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{v}]^T[(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{v}]\}:=\mathbb{E}\{\boldsymbol{y}^T\boldsymbol{y}\}\geq 0\end{aligned}$$

其中 $\boldsymbol{y}=(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{v}$。所以 $\mathrm{Cov}(\boldsymbol{X})\succeq 0$。

定理 3.2 (Rolle 定理) 令 f 在闭区间 $[a,b]$ 上连续，在开区间 (a,b) 上可导。假如 $f(a)=f(b)$，则在开区间 (a,b) 中存在至少一个点 c，满足 $f'(c)=0$。

证明 由于 f 在 $[a,b]$ 上连续，根据闭区间上连续函数的最值定理，f 在 $[a,b]$ 上必能取得最大值 M 和最小值 m。

如果 $M=m$，则 f 在 $[a,b]$ 上是常数函数，于是在 (a,b) 内任意一点 c 都有 $f'(c)=0$，结论成立。

如果 $M>m$，因为 $f(a)=f(b)$，所以最大值 M 和最小值 m 中至少有一个在 (a,b) 内某点 c 取得。不妨设 $f(c)=M$，其中 $c\in(a,b)$。我们来证 $f'(c)=0$。由于 f 在 c 处可导，考虑导数定义：

$$f'(c)=\lim_{h\to 0}\frac{f(c+h)-f(c)}{h}.$$

当 $h>0$ 时，$\frac{f(c+h)-f(c)}{h}\leq 0$（因为 $f(c)$ 是最大值），所以

$$f'(c)=\lim_{h\to 0^+}\frac{f(c+h)-f(c)}{h}\leq 0.$$

当 $h<0$ 时，$\frac{f(c+h)-f(c)}{h}\geq 0$，所以

$$f'(c)=\lim_{h\to 0^-}\frac{f(c+h)-f(c)}{h}\geq 0.$$

因此 $f'(c)\geq 0$ 且 $f'(c)\leq 0$，只能 $f'(c)=0$。如果最小值 m 在 (a,b) 内取到，同理可证该点处导数为零。综上，总存在 $c\in(a,b)$ 使得 $f'(c)=0$。

使用微积分中求极值的方法，其流程相对比较简单，只需求出似然函数，以最大化似然函数为目标，对参数进行优化即可。

3.1.3 MLE 求解：根据定义

当似然函数不可微，甚至不连续时，只能通过极大似然的定义进行求解。

例 3.1 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从均匀分布族 $\{U(\theta,\theta+1):\theta\in\mathbb{R}\}$ 中抽取的简单随机样本，求 θ 的 MLE。

解答 给定样本 $\boldsymbol{x}$ 时的似然函数为

$$L(\theta;\boldsymbol{x})=\mathbb{I}(\theta<x_{(1)}\leq x_{(n)}<\theta+1)=\mathbb{I}(x_{(n)}-1<\theta<x_{(1)})$$

似然函数最大只能为 1，需要满足 $x_{(n)}-1<\theta<x_{(1)}$，所以有 $\hat{\theta}=t(X_{(n)}-1)+(1-t)X_{(1)}$ 其中 $t\in(0,1)$ 均为 θ 的 MLE。所以此时满足极大似然估计的 $\hat{\theta}$ 有无穷多个。

3.2 极大似然估计的性质

3.2.1 极大似然估计的无偏性和相合性

极大似然估计可能无偏，也可能有偏。极大似然估计可能相合，也可能不相合。下面给出极大似然估计强相合性的结果。

定理 3.3 (*) 设 $\boldsymbol{X}_n=(X_1,\cdots,X_n)$ 是从分布族 $\mathscr{F}=\{f(x;\theta):\theta\in\Theta\}$ 中抽取的简单随机样本，其中 Θ 为任一开区间，若有下面的条件成立：

1. 分布族 $\mathscr{F}$ 是可识别的，即对 Θ 中任意的 $\theta_1\neq\theta_2$，都有 $f(x;\theta_1)\neq f(x;\theta_2)$。
2. 对任意 $x\in\mathscr{X}$，$\theta\in\Theta$ 有 $f(x;\theta)>0$，且一阶偏导数 $\partial f(x;\theta)/\partial\theta$ 存在、连续。那么在任一 P_θ 下，以概率为 1 当 n 充分大时，对数似然函数方程：

$$\sum\nolimits_{i=1}^{n}\frac{\partial f(x_i;\theta)}{\partial\theta}=0$$

存在一个解。

那么解 $\hat{\theta}_n$ 是强相合的，即 $\hat{\theta}_n\xrightarrow{\text{a.s.}}\theta$，$n\to\infty$。

3.2.2 极大似然估计与充分统计量

定理 3.4 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从总体 $\{f(x;\theta):\theta\in\Theta\}$ 中抽取的简单随机样本，$T(\boldsymbol{X})$ 是参数 θ 的充分统计量。若 θ 的极大似然估计 $\hat{\theta}_{\mathrm{MLE}}$ 存在，则它必为充分统计量 T 的函数，即 $\hat{\theta}_{\mathrm{MLE}}=\hat{\theta}_{\mathrm{MLE}}(T(\boldsymbol{X}))$。

证明 由因子分解定理可知样本 $\boldsymbol{X}$ 的概率函数，即似然函数可表示为

$$L(\theta,\boldsymbol{x})=\prod_{i=1}^{n}f(x_i;\theta)=g(T(\boldsymbol{x}),\theta)h(\boldsymbol{x})$$

若 θ 的 MLE 存在，记为 $\hat{\theta}_{\mathrm{MLE}}$，则它满足

$$L(\hat{\theta}_{\mathrm{MLE}},\boldsymbol{x})=\sup_{\theta\in\Theta}L(\theta,\boldsymbol{x})\quad\Rightarrow\quad g(T(\boldsymbol{x}),\hat{\theta}_{\mathrm{MLE}})=\sup_{\theta\in\Theta}g(T(\boldsymbol{x}),\theta)$$

上式中 $g(T(\boldsymbol{x}),\hat{\theta}_{\mathrm{MLE}})=\sup_{\theta\in\Theta}g(T(\boldsymbol{x}),\theta)$ 左右式都只与 $T(\boldsymbol{X})$ 和 $\hat{\theta}_{\mathrm{MLE}}$ 有关，故 $\hat{\theta}_{\mathrm{MLE}}$ 可以写成 $T(\boldsymbol{X})$ 的函数。

3.2.3 极大似然估计的相合渐近正态性

只考虑参数 θ 为一维的情形。设 $\mathscr{F}=\{f(x;\theta):\theta\in\Theta\}$ 为一概率函数族，$\Theta=(a,b)$ 为 $\mathbb{R}$ 上开区间，$\mathscr{F}$ 是可识别的。设 $f(x;\theta)$ 满足下列条件：

(1) 对一切 $\theta\in\Theta$, $x\in\mathscr{X}$ 有 $f(x;\theta)>0$，且 $f(x;\theta)$ 有直到 2 阶的连续偏导数。

(2) 对任一给定的 θ_0，存在它的一个邻域 $U(\theta_0)$ 和可能依赖于该 θ_0 的函数 $g(x)>0$ 和 $G(x)>0$，对该邻域内的任一 θ，有

$$\left|\frac{\partial f(x;\theta)}{\partial\theta}\right|<g(x)\qquad\left|\frac{\partial^2 f(x;\theta)}{\partial\theta^2}\right|<g(x)\qquad\left|\frac{\partial^2 f(x;\theta)}{\partial\theta^2}-\frac{\partial^2 f(x;\theta_0)}{\partial\theta^2}\right|\leq G(x)(\theta-\theta_0)^2$$

且

$$\int_{-\infty}^{\infty}g(x)\,dx<\infty\qquad\mathbb{E}_{\theta_0}[G(x)]=\int_{-\infty}^{\infty}G(x)f(x;\theta_0)\,dx<\infty$$

(3) 对一切 $\theta\in\Theta$，有

$$0<I(\theta)\triangleq\mathbb{E}\left[\left(\frac{\partial\log f(X;\theta)}{\partial\theta}\right)^2\right]=\int_{-\infty}^{\infty}\left(\frac{\partial\log f(X;\theta)}{\partial\theta}\right)^2 f(x;\theta)\,dx<\infty$$

这里 $I(\theta)$ 称为 Fisher 信息量。

定理 3.5 (MLE 的渐近正态性定理) 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自分布族 $\mathscr{F}$ 中抽取的简单随机样本，若上述条件 (1) (2) (3) 满足，则对任何 $\theta_0\in\Theta$，对数似然方程

$$\sum\nolimits_{i=1}^{n}\frac{\partial f(x_i;\theta)}{\partial\theta}=0$$

有一根 $\hat{\theta}_n=\hat{\theta}(\boldsymbol{X})$，则其满足

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{\mathscr{L}}N\left(0,I^{-1}(\theta_0)\right),\quad\theta\in\Theta$$

且 $\hat{\theta}_n$ 为 θ_0 的弱相合估计。

证明 由 Taylor 展开，我们有

$$l'(\hat{\theta}_n) = l'(\theta_0) + l''(\theta_0)(\hat{\theta}_n - \theta_0) + \frac{l'''(\tilde{\theta})}{2}(\hat{\theta}_n - \theta_0)^2$$

其中 $l(\theta)$ 为对数似然函数，$\tilde{\theta}$ 位于 $\hat{\theta}_n$ 和 θ_0 之间。而由 MLE 的定义，有 $l'(\hat{\theta}_n) = 0$，于是将上式变形

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\frac{1}{\sqrt{n}} l'(\theta_0)}{-\frac{1}{n} l''(\theta) - \frac{1}{2n} l'''(\tilde{\theta})(\hat{\theta}_n - \theta_0)} \tag{3}$$

(A) 注意到 $l(\theta) = \sum_{i=1}^n \log f(x_i;\theta)$，而

$$\mathbb{E}\left(\frac{\partial \log f(x;\theta)}{\partial \theta}\right) = \mathbb{E}\left(\frac{f'(x;\theta)}{f(x;\theta)}\right) = \int \frac{f'(x;\theta)}{f(x;\theta)} \cdot f(x;\theta)\, dx = \int f'(x;\theta)\, dx$$

而 $\dfrac{\partial}{\partial \theta} \int f(x;\theta)\, dx = \dfrac{\partial}{\partial \theta} 1 = 0$，故上式 $\int f'(x;\theta)\, dx = 0$。容易推出 $\mathbb{E}[l'(\theta) = 0]$。而 $l(\theta)$ 又是求和形式，故有中心极限定理知

$$\sqrt{n}\left(\frac{l'(\theta_0)}{n} - \mathbb{E}\left[\frac{l'(\theta_0)}{n}\right]\right) \xrightarrow{\mathscr{L}} N\left(0, \mathrm{Var}(l'(\theta_0))\right)$$

注意到 $\mathbb{E}\left[\dfrac{l'(\theta_0)}{n}\right] = 0$，有

$$\mathrm{Var}(l'(\theta_0)) = \mathbb{E}\left[\left(\frac{\partial \log f(x;\theta_0)}{\partial \theta}\right)^2\right] = I(\theta_0)$$

综上就有

$$\frac{1}{\sqrt{n}} l'(\theta_0) \xrightarrow{\mathscr{L}} N(0, I(\theta_0)) \tag{4}$$

(B) 先证明下面的结论

$$-\mathbb{E}\left[\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right] = \mathbb{E}\left[\left(\frac{\partial \log f(x;\theta)}{\partial \theta}\right)^2\right]$$

由密度函数的性质 $\int f(x;\theta)\, dx = 1$ 可以得到 $\int f'(x;\theta)\, dx = \int f''(x;\theta)\, dx = 0$，于是有

$$\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right] &= -\int \left(\frac{f''(x;\theta) f(x;\theta) - [f'(x;\theta)]^2}{[f(x;\theta)]^2}\right) \cdot f(x;\theta)\, dx \\
&= 0 + \int \left(\frac{f'(x;\theta)}{f(x;\theta)}\right)^2 \cdot f(x;\theta)\, dx \\
&= \int \left(\frac{\partial \log f(x;\theta)}{\partial \theta}\right)^2 \cdot f(x;\theta)\, dx \\
&= \mathbb{E}\left[\left(\frac{\partial \log f(x;\theta)}{\partial \theta}\right)^2\right]
\end{aligned}$$

结论得证。又注意到 $l''(\theta_0)$ 是求和的形式，于是根据大数定律有

$$\frac{1}{n} l''(\theta_0) = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} \xrightarrow{P} \mathbb{E}\left[\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right] = -I(\theta_0) \tag{5}$$

(C) 由条件 (2) 和大数定律

$$\frac{1}{n}l'''(\tilde{\theta}) \xrightarrow{P} \mathbb{E}(l'''(\tilde{\theta})) < \infty \tag{6}$$

有限，不妨设 $\frac{1}{n}l'''(\tilde{\theta}) \to M$。

综合 (A) (B) (C)，结合条件 (1) (2) (3) 均满足，容易得到 $\hat{\theta}_n$ 是相合的，故 $\hat{\theta}_n - \theta_0 \to 0,\ n \to \infty$。于是将式 4 式 5 和式 6 代入式 3 中，结合 Slusky 定理

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}} \frac{N(0, I(\theta_0))}{I(\theta_0) - M \cdot (\hat{\theta}_n - \theta_0)} \to N(0, I^{-1}(\theta_0))$$

当 $n \to \infty$ 时，命题证毕。

例 3.2 设 $X_1, \cdots, X_n$ 是取自正态总体 $N(\mu, \sigma^2)$ 的样本，则 μ, σ^2 的 MLE 分别具有渐近正态性。

证明 显然正态分布满足条件 (1) (2) (3)，所以可以使用定理 3.5 证明：

(1) 在 σ^2 已知时，μ 的 MLE 为 $\hat{\mu} = \bar{X}$，由于

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

故有对数概率密度为

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

所以

$$I_1(\mu, \sigma^2) = \mathbb{E}\left[\left(\frac{\partial \log f(x;\mu)}{\partial \mu}\right)^2\right] = \mathbb{E}\left[\frac{(x-\mu)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}$$

故由定理 3.5 可知，

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{\mathscr{L}} N(0, \sigma^2)$$

即 $\hat{\mu}$ 的渐近分布为 $N(\mu, \sigma^2/n)$。

(2) 在 μ 已知时，σ^2 的 MLE 为 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$，而

$$I_2(\mu, \sigma^2) = \mathbb{E}\left[\left(\frac{\partial \log f(x;\sigma^2)}{\partial \sigma^2}\right)^2\right] = \mathbb{E}\left[\left(-\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4}\right)^2\right] = \frac{1}{2\sigma^4}$$

故由定理 3.5 可知，

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{\mathscr{L}} N(0, 2\sigma^4)$$

即 $\hat{\sigma}^2$ 的渐近分布为 $N(\sigma^2, 2\sigma^4/n)$。

4 一致最小均方误差估计

设一参数分布族 $\mathscr{F} = \{f(x;\theta) : \theta \in \Theta\}$，其中 Θ 为参数空间。设 $g(\theta)$ 为定义在 Θ 上的实函数，$\boldsymbol{X} = (X_1, \cdots, X_n)$ 为自 $\mathscr{F}$ 的简单样本，$\hat{g}(\boldsymbol{X}) = \hat{g}(X_1, \cdots, X_n)$ 为 $g(\theta)$ 的一个估计量。

定义 4.1 (均方误差) 设 $\hat{g}(\boldsymbol{X})$ 为 $g(\theta)$ 的估计量，则称 $\mathbb{E}_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2$ 为 $\hat{g}(\boldsymbol{X})$ 的**均方误差**（Mean Square Error, MSE）。

定义 4.2 (一致最小均方误差估计) 设 $\hat{g}_1(\boldsymbol{X})$ 和 $\hat{g}_2(\boldsymbol{X})$ 为 $g(\theta)$ 的两个不同的估计量，若

$$\mathbb{E}_\theta[\hat{g}_1(\boldsymbol{X})-g(\theta)]^2 \le \mathbb{E}_\theta[\hat{g}_2(\boldsymbol{X})-g(\theta)]^2 \quad \forall\,\theta\in\Theta$$

且不等号至少对某个 $\theta\in\Theta$ 成立，则称在 MSE 准则下 $\hat{g}_1(\boldsymbol{X})$ 优于 $\hat{g}_2(\boldsymbol{X})$。

若存在 $\hat{g}^*(\boldsymbol{X})$，使得对 $g(\theta)$ 的任一估计量 $\hat{g}(\boldsymbol{X})$，都有

$$\mathbb{E}_\theta[\hat{g}^*(\boldsymbol{X})-g(\theta)]^2 \le \mathbb{E}_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2 \quad \forall\,\theta\in\Theta$$

则称 $\hat{g}^*(\boldsymbol{X})$ 为 $g(\theta)$ 的**一致最小均方误差估计**。

可惜的是，**一致最小均方误差估计常不存在**。从直观上想，在一个大的估计类中，一致最优估计量不存在，把估计类缩小，就有可能存在一致最优的估计量。我们可以把估计类缩小为**无偏估计类**来考虑。

存在这样的情形，参数 $g(\theta)$ 的无偏估计不存在，如下例：

例 4.1 设样本 X 服从二项分布 $B(n,p)$ 时，其中 n 已知而 p 未知。令 $g(p)=1/p$，则参数 $g(p)$ 的无偏估计不存在。

证明 用反证法：若不然，$g(p)$ 有无偏估计 $\hat{g}(X)$。由于 X 只取 $0,1,\cdots,n$ 这些值，而假设 $\hat{g}(X)$ 无偏，应有

$$\mathbb{E}_p[\hat{g}(X)]=\sum_{x=0}^{n}\hat{g}(x)\binom{n}{x}p^x(1-p)^{n-x}=\frac{1}{p}$$

即多项式方程

$$\sum_{x=0}^{n}\hat{g}(x)\binom{n}{x}p^{x+1}(1-p)^{n-x}-1=0$$

对任意 $p\in(0,1)$ 成立。但这是至多 $n+1$ 阶多项式，根据多项式的性质，至多有 $n+1$ 个根，显然无法满足任意 $p\in(0,1)$ 均成立，矛盾。

5 一致最小方差无偏估计 UMVUE

定义 5.1 (可估) 参数的无偏估计若存在，则称此参数为**可估参数**；若参数函数的无偏估计存在，则称此函数为**可估函数**（estimable function）。因此可估函数的无偏估计类是非空的。

定义 5.2 (一致最小方差无偏估计 UMVUE) 设 $\mathscr{F}=\{f(x;\theta):\theta\in\Theta\}$ 是一个参数分布族，其中 Θ 为参数空间，$g(\theta)$ 为定义在 Θ 上的可估函数。设 $\hat{g}^*(\boldsymbol{X})=\hat{g}^*(X_1,\cdots,X_n)$ 为 $g(\theta)$ 的一个无偏估计，若对 $g(\theta)$ 的任一无偏估计 $\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\cdots,X_n)$，都有

$$\mathrm{Var}_\theta[\hat{g}^*(\boldsymbol{X})]\le \mathrm{Var}_\theta[\hat{g}(\boldsymbol{X})] \quad \theta\in\Theta$$

则称 $\hat{g}^*(\boldsymbol{X})$ 是 $g(\theta)$ 的**一致最小方差无偏估计**（uniformly minimum variance unbiased estimation, UMVUE）。

对给定参数分布族，寻找可估函数的 UMVUE 有如下的几种方法:

1. 零无偏估计法
2. 充分完全统计量法
3. Cramer-Rao 不等式

下列的定理提供了一个改进无偏估计的方法。

定理 5.1 (Rao-Blackwell 定理) 设 $T = T(\boldsymbol{X})$ 是一个充分统计量，而 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的一个无偏估计，则

$$h(T) = \mathbb{E}[\hat{g}(\boldsymbol{X})|T]$$

是 $g(\theta)$ 的一个无偏估计，并且

$$\operatorname{Var}_\theta[h(T)] \leq \operatorname{Var}_\theta[\hat{g}(\boldsymbol{X})] \quad \theta \in \Theta$$

其中等号当且仅当 $P_\theta(\hat{g}(\boldsymbol{X}) = h(T)) = 1$，即 $\hat{g}(\boldsymbol{X}) = h(T)$ a.s. P_θ 成立。

这个定理提供了一个改进无偏估计的方法：即一个无偏估计 $\hat{g}(\boldsymbol{X})$ 对充分统计量 $T(X)$ 的条件期望 $\mathbb{E}[\hat{g}(\boldsymbol{X})|T]$ 将能导出一个新的无偏估计，且它的方差不会超过原估计量 $\hat{g}(\boldsymbol{X})$ 的方差。

这个定理还表明 **UMVUE 一定是充分统计量的函数**，否则可以通过充分统计量，按定理 5.1 的方法构造出一个具有更小方差的无偏估计。

证明 **(A)** 先证 $h(T)$ 的无偏性。由 $T(\boldsymbol{X})$ 为充分统计量，按定义，给定 T 时 $\boldsymbol{X}$ 的条件分布与 θ 无关。因此 $h(T) = \mathbb{E}(\hat{g}(\boldsymbol{X})|T)$ 与 θ 无关。所以 $h(T)$ 是统计量，可作为 $g(\theta)$ 的估计量，且有

$$\mathbb{E}_\theta[h(T)] = \mathbb{E}_\theta[E(\hat{g}(\boldsymbol{X})|T)] = \mathbb{E}_\theta(\hat{g}(\boldsymbol{X})) = g(\theta)$$

（重期望公式）因此 $h(T)$ 是 $g(\theta)$ 的无偏估计。

(B) 再证不等式关系成立，即 $h(T)$ 的方差不增。注意到

$$\begin{aligned}\operatorname{Var}_\theta(\hat{g}(\boldsymbol{X})) &= \mathbb{E}_\theta\{\hat{g}(\boldsymbol{X}) - g(\theta)\}^2 = \mathbb{E}_\theta\{[\hat{g}(\boldsymbol{X}) - h(T)] + [h(T) - g(\theta)]\}^2 \\ &= \mathbb{E}_\theta[\hat{g}(\boldsymbol{X}) - h(T)]^2 + \operatorname{Var}_\theta[h(T)] + 2\mathbb{E}_\theta\{[\hat{g}(\boldsymbol{X}) - h(T)][h(T) - g(\theta)]\}\end{aligned}$$

由 $h(T) = \mathbb{E}(\hat{g}(\boldsymbol{X})|T)$，同样由重期望公式

$$\begin{aligned}&\mathbb{E}_\theta\{[\hat{g}(\boldsymbol{X}) - h(T)][h(T) - g(\theta)]\} = \mathbb{E}_\theta\{\mathbb{E}_\theta[(\hat{g}(\boldsymbol{X}) - h(T))(h(T) - g(\theta))|T]\} \\ &\quad = \mathbb{E}_\theta\{[h(T) - g(\theta)]\mathbb{E}_\theta[(\hat{g}(\boldsymbol{X}) - h(T))|T]\} \\ &\quad = \mathbb{E}_\theta\{[h(T) - g(\theta)][\mathbb{E}(\hat{g}(\boldsymbol{X})|T) - h(T)]\} = 0\end{aligned}$$

代回原式可得

$$\operatorname{Var}_\theta(\hat{g}(\boldsymbol{X})) = \mathbb{E}_\theta[\hat{g}(\boldsymbol{X}) - h(T)]^2 + \operatorname{Var}_\theta[h(T)] \geq \operatorname{Var}_\theta[h(T)] \qquad \square$$

(C) 最后证明等号成立条件。等号成立等价于

$$\mathbb{E}_\theta[\hat{g}(\boldsymbol{X}) - h(T)]^2 = 0$$

即 $\hat{g}(\boldsymbol{X}) = h(T)$ a.s. P_θ 成立。

例 5.1 对于 $X_1,\cdots,X_n \overset{\text{i.i.d.}}{\sim} B(1,p)$ 的简单随机样本，X_1 是 p 的无偏估计。设 $T=\sum_{i=1}^n X_i$，利用统计量 T 构造具有比 X_1 估计更小方差的无偏估计。

解答 使用定理 5.1 定理，容易构造 $h(T)$

$$\begin{aligned} h(t) &= \mathbb{E}(X_1|T=t) = 1\cdot P(X_1=1|T=t)+0\cdot P(X_1=0|T=t) \\ &= \frac{P(X_1=1,T=t)}{P(T=t)} = \frac{P(X_1=1,X_2+\cdots+X_n=t)}{P(T=t)} \\ &= \frac{p\cdot\binom{n-1}{t-1}p^{t-1}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \frac{t}{n} = \bar{x} \end{aligned}$$

显然 $\mathrm{Var}(h(T))=\mathrm{Var}(\bar{X})=p(1-p)/p<p(1-p)=\mathrm{Var}(X_1)$，当 $n>1$ 时 $\bar{X}$ 的方差更小，且为无偏估计。

5.1 充分完全统计量法

定理 5.2 (Lehmann-Scheffe 定理) 设 $X\sim\{f(x;\theta):\theta\in\Theta\}$，$\Theta$ 为参数空间。令 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从总体 X 中抽取的简单样本，$g(\theta)$ 为定义于参数空间 Θ 上的可估函数，$T(\boldsymbol{X})$ 为一个充分完全统计量。若 $h[T(\boldsymbol{X})]$ 为 $g(\theta)$ 的一个无偏估计，则 $h[T(\boldsymbol{X})]$ 是 $g(\theta)$ 的唯一的 UMVUE。唯一性是指：设 $\hat{g}$ 和 $\hat{g}_1$ 是 $g(\theta)$ 的两个估计量，若 $P_\theta(\hat{g}=\hat{g}_1)=1$，对一切 $\theta\in\Theta$，则视 $\hat{g}$ 和 $\hat{g}_1$ 是同一个估计量。

证明 先证唯一性。设 $\hat{g}[T(\boldsymbol{X})],\hat{g}_1[T(\boldsymbol{X})]$ 为 $g(\theta)$ 的无偏估计，令 $\delta[T(\boldsymbol{X})]=\hat{g}[T(\boldsymbol{X})]-\hat{g}_1[T(\boldsymbol{X})]$，则

$$\mathbb{E}_\theta\{\delta[T(\boldsymbol{X})]\}=\mathbb{E}_\theta\{\hat{g}[T(\boldsymbol{X})]\}-\mathbb{E}_\theta\{\hat{g}_1[T(\boldsymbol{X})]\}=0 \quad \theta\in\Theta$$

由于 $T(\boldsymbol{X})$ 为完全统计量，故 $\delta[T(\boldsymbol{X})]=0$ a.s. P_θ 成立，即 $\hat{g}[T(\boldsymbol{X})]=\hat{g}_1[T(\boldsymbol{X})]$ a.s. P_θ 成立，唯一性证毕。

再证一致最小方差性。设 $\varphi(\boldsymbol{X})$ 为 $g(\theta)$ 的任一无偏估计，记 $h[T(\boldsymbol{X})]=\mathbb{E}[\varphi(\boldsymbol{X})|T]$。由 $T(\boldsymbol{X})$ 为充分统计量，故 $h[T(\boldsymbol{X})]$ 与 θ 无关，即为统计量。由定理 5.1 可知，$h[T(\boldsymbol{X})]$ 也是 $g(\theta)$ 的无偏估计，且

$$\mathrm{Var}_\theta\{h[T(\boldsymbol{X})]\}\le\mathrm{Var}_\theta\{\varphi(\boldsymbol{X})\} \quad \theta\in\Theta$$

即 $h[T(\boldsymbol{X})]$ 是最小方差的无偏估计，即 UMVUE，证毕。

例 5.2 设 $X_1,\cdots,X_n \overset{\text{i.i.d.}}{\sim} B(1,p)$，故容易得 $T=T(\boldsymbol{X})=\sum_{i=1}^n X_i$ 服从二项分布 $B(n,p)$，且 T 为充分完全统计量。求 $g(p)=p(1-p)$ 的 UMVUE。

解答 (方法一) 令 $\varphi(\boldsymbol{X})=\mathbb{I}(X_1=1,X_2=0)$，则有 $\mathbb{E}[\varphi(\boldsymbol{X})]=P(X_1=1,X_2=0)=p(1-p)$，即 $\varphi(\boldsymbol{X})$ 为 $g(p)$ 的无偏估计。又注意到 $T=\sum_{i=1}^n X_i\sim B(n,p)$ 和 $\sum_{i=3}^n X_i\sim B(n-2,p)$，于是

可以由定理 5.1 定理改进 $g(p)$ 的无偏估计

$$\begin{aligned} h(t) &= \mathbb{E}[\varphi(\boldsymbol{X})|T=t] = P(X_1, X_2 = 0|T=t) \\ &= \frac{P(X_1=1, X_2=0, \sum_{i=3}^n X_i = t-1)}{P(T=t)} = \binom{n-2}{t-1}\Big/\binom{n}{t} = \frac{t(n-t)}{n(n-1)} \end{aligned}$$

仍然是 $g(p)$ 的无偏估计。而 $h(T)$ 又是充分完全统计量 $T(\boldsymbol{X})$ 的函数，由定理 5.2 可知 $h(T)$ 是 $g(p)$ 的 UMVUE。

解答 (方法二) 设 $\delta(T)$ 为 $g(p)=p(1-p)$ 的一个无偏估计，从而由 $T \sim B(n,p)$ 可以得到

$$\mathbb{E}[\delta(T)] = \sum_{t=0}^{n} \binom{n}{t} \delta(t) p^t (1-p)^{n-t} = p(1-p) \quad p \in (0,1)$$

令 $\rho = p(1-p)$，故有 $p = \rho/(1+\rho)$，代入有

$$\sum_{t=0}^{n} \binom{n}{t} \delta(t) \rho^t = \rho(1+\rho)^{n-2} \quad \rho \in (0,\infty)$$

使用二项式公式展开 $(1+\rho)^{n-2}$ 可以得到

$$\sum_{t=0}^{n} \binom{n}{t} \delta(t) \rho^t = \sum_{t=1}^{n-1} \binom{n-2}{t-1} \rho^t \quad \rho \in (0,\infty)$$

由多项式的性质，比较系数有

$$\delta(t) = \begin{cases} 0 & t = 0, n \\ \binom{n-2}{t-1}\Big/\binom{n}{t} & t = 1, \cdots, n-1 \end{cases}$$

既有

$$\delta(T) = \frac{T(n-T)}{n(n-1)}$$

为 $g(p)=p(1-p)$ 的无偏估计，而其为充分统计量的函数，由定理 5.2 可知 $\delta(T)$ 为 UMVUE。

例 5.3 设 $X = (X_1, \cdots, X_n)$ 为从 Poisson 分布 $\mathrm{Poi}(\lambda)$ 中抽取的简单样本，求

1. $g_1(\lambda) = \lambda^r$，$r \in \mathbb{N}^+$ 的 UMVUE。
2. $g_2(\lambda) = P_\lambda(X_1 = x)$ 的 UMVUE。

解答 (1) 容易得到 $T = T(\boldsymbol{X}) = \sum_{i=1}^n X_i \sim \mathrm{Poi}(n\lambda)$ 为充分完全统计量。设 $h_1(T)$ 为 $g_1(\lambda) = \lambda^r$ 的无偏估计，故有

$$\mathbb{E}[h_1(T)] = \sum_{t=0}^{\infty} h_1(t) \frac{e^{-n\lambda}(n\lambda)^t}{t!} = \lambda^r$$

将 $e^{n\lambda}$ 作 Taylor 展开有

$$\lambda^r e^{n\lambda} = \sum_{l=r}^{\infty} \frac{n^{t-r}\lambda^t}{(t-r)!} = \sum_{t=0}^{\infty} h_1(t) \frac{n^t \lambda^t}{t!}$$

比较系数可得

$$h_1(T)=\frac{T(T-1)\cdots(T-r+1)}{n^r}$$

且其为 $g_1(\lambda)=\lambda^r$ 的无偏估计，又为充分完全统计量的函数，由定理 5.2 可知 $h_1(T)$ 为 $g_1(\lambda)$ 的 UMVUE。

解答 (2) 设 $\varphi(X_1)=\mathbb{I}(X_1=x)$，则 $\mathbb{E}\lambda[\varphi(X_1)]=P_\lambda(X_1=x)$，因此 $\varphi(X_1)$ 为 $g_2(\lambda)$ 的无偏估计。又注意到 $T=\sum_{i=1}^n X_i\sim \mathrm{Poi}(n\lambda)$ 和 $\sum_{i=2}^n X_i\sim \mathrm{Poi}((n-1)\lambda)$，故有

$$\begin{aligned}h_2(t)=\mathbb{E}[\varphi(X_1)|T=t]=P(X_1=x|T=t)&=\frac{P(X_1=x,T=t)}{P(T=t)}\\&=\frac{P(X_1=x)P(\sum_{i=2}^n X_i=t-x)}{P(\sum_{i=1}^n X_i=t)}=\frac{(n-1)^{t-x}t!}{n^t(t-x)!x!}\end{aligned}$$

由定理 5.1 可知，$h_2(T)$ 是 $g_2(\lambda)$ 的无偏估计。又因为其为充分完全统计量 T 的函数，由定理 5.2 可知 $h_2(T)$ 是 $g_2(\lambda)$ 的 UMVUE。

例 5.4 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从正态分布 $N(a,\sigma^2)$ 中抽取的简单随机样本，记 $\boldsymbol{\theta}=(a,\sigma^2)$。求

1. $g_1(\boldsymbol{\theta})=\sigma^r$，$r>0$ 的 UMVUE。
2. $g_2(\boldsymbol{\theta})=a/\sigma^2$ 的 UMVUE。

解答 (1) 容易得 $\boldsymbol{T}=(T_1,T_2)=(\bar{X},\sum_{i=1}^n(X_i-\bar{X})^2)$ 是 $\boldsymbol{\theta}$ 的充分完全统计量。又注意到 $T_2/\sigma^2\sim\chi^2_{n-1}$，于是

$$\begin{aligned}\mathbb{E}\left(\frac{T_2}{\sigma^2}\right)^{r/2}&=\int_0^\infty y^{\frac{r}{2}}\cdot\frac{1}{2^{\frac{n-1}{2}}\Gamma((n-1)/2)}y^{\frac{n-1}{2}-1}e^{-\frac{y}{2}}\,dy\\&=2^{\frac{r}{2}}\,\Gamma\left(\frac{n+r-1}{2}\right)\Big/\,\Gamma\left(\frac{n-1}{2}\right)\triangleq\frac{1}{K_{n-1,r}}\end{aligned}$$

因此构造

$$h_1(\boldsymbol{T})=K_{n-1,r}\cdot T_2^{r/2}=T_2^{r/2}\cdot\Gamma\left(\frac{n-1}{2}\right)\Big/\left[2^{\frac{r}{2}}\Gamma\left(\frac{n+r-1}{2}\right)\right]$$

为 σ^r 的无偏估计，又为充分完全统计量 $\boldsymbol{T}$ 的函数，故由定理 5.2 可知 $h_1(\boldsymbol{T})$ 是 $g_1(\boldsymbol{\theta})$ 的 UMVUE。

解答 (2) 断言：若 $Y\sim\chi^2_n$ 则 $\mathbb{E}(1/Y)=1/(n-2)$。注意到 $T_1\sim N(a,\sigma^2/n)$，$T_2/\sigma^2\sim\chi^2_{n-1}$ 且 $T_1\perp T_2$，故

$$\mathbb{E}\left(\frac{T_1}{T_2}\right)=\mathbb{E}(T_1)\cdot\mathbb{E}\left(\frac{1}{T_2}\right)=\frac{a}{(n-3)\sigma^2}$$

于是构造 $h_2(\boldsymbol{T})=(n-3)T_1/T_2$ 即为 $g_2(\boldsymbol{\theta})$ 的无偏估计，而又是充分完全统计量 $\boldsymbol{T}$ 的函数。故由定理 5.2 可知 $h_2(\boldsymbol{T})$ 是 $g_2(\boldsymbol{\theta})$ 的 UMVUE。

5.2 零无偏估计法

即使使用充分完全统计量构造无偏估计，配合 Lehmann-Scheffe 定理 5.2 寻找 UMVUE 十分常见。但仍然存在难以构造的情况，例如无法找到充分完全统计量，构造无偏估计困难等。下面的例子就展示了这种情况。

例 5.5 设 $X_1,\cdots,X_n \overset{\text{i.i.d.}}{\sim} U(0,\theta)$，但 $\theta>1$，求 θ 的 UMVUE。

解答 这里我们证明：$T=X_{(n)}$ 不是 θ 的完全统计量。假设 $\varphi(T)$ 满足 $\mathbb{E}[\varphi(T)]=0$，即

$$\begin{aligned}\mathbb{E}[\varphi(T)] &= \int_0^\theta \varphi(t)\cdot\frac{nt^{n-1}}{\theta^n}\,dt=0\\ \Rightarrow \int_0^1 \varphi(t)t^{n-1}\,dt &+ \int_1^\theta \varphi(t)t^{n-1}\,dt=0\end{aligned}$$

可以构造 $\varphi(T)=[(n+1)T-n]\cdot\mathbb{I}_{(0,1)}(T)$ 就有

$$\int_0^1 \varphi(t)t^{n-1}\,dt+\int_1^\theta \varphi(t)t^{n-1}\,dt = t^{n+1}-t^n|_0^1+\int_1^\theta 0\cdot t^{n-1}\,dt\equiv 0$$

于是 $\varphi(T)$ 满足 $\mathbb{E}[\varphi(T)]=0$，但 $\varphi(T)\neq 0$ a.s. P，即 $T=X_{(n)}$ 是不完全的。此时无法继续使用定理 5.2 构造 UMVUE。

于是，这里引入求解 UMVUE 的另一个方法，即**零无偏估计法**。

定理 5.3 (零无偏估计法) 设 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的一个无偏估计，$\mathrm{Var}_\theta(\hat{g}(\boldsymbol{X}))<\infty$，$\forall\,\theta\in\Theta$。令

$$\mathscr{U}=\{U(\boldsymbol{X}):\mathbb{E}_\theta[U(\boldsymbol{X})]=0,\forall\,\theta\in\Theta\}$$

为零无偏估计的集合。则 $\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的 UMVUE 的充分必要条件为

$$\mathrm{Cov}_\theta[\hat{g}(\boldsymbol{X}),U(\boldsymbol{X})]=\mathbb{E}_\theta[\hat{g}(\boldsymbol{X})\cdot U(\boldsymbol{X})]=0 \qquad \forall\,\theta\in\Theta,\forall\,U(\boldsymbol{X})\in\mathscr{U}$$

证明 **(A)** 先证明 $\Leftarrow$，即已知 $\mathrm{Cov}[\hat{g}(\boldsymbol{X}),U(\boldsymbol{X})]=0$, $\forall\,U(\boldsymbol{X})\in\mathscr{U}$。设 $\hat{g}_1(\boldsymbol{X})$ 为 $g(\theta)$ 的任一无偏估计，且 $\hat{g}_1(\boldsymbol{X})\neq\hat{g}(\boldsymbol{X})$，记 $U(\boldsymbol{X})=\hat{g}_1(\boldsymbol{X})-\hat{g}(\boldsymbol{X})$，则有 $\mathbb{E}[U(\boldsymbol{X})]=0$，即 $U(\boldsymbol{X})\in\mathscr{U}$。那么根据条件有

$$\begin{aligned}\mathrm{Var}[\hat{g}_1(\boldsymbol{X})] &= \mathrm{Var}[\hat{g}(\boldsymbol{X})+U(\boldsymbol{X})]=\mathrm{Var}[\hat{g}(\boldsymbol{X})]+\mathrm{Var}[U(\boldsymbol{X})]+2\mathrm{Cov}[\hat{g}(\boldsymbol{X}),U(\boldsymbol{X})]\\ &= \mathrm{Var}[\hat{g}(\boldsymbol{X})]+\mathrm{Var}[U(\boldsymbol{X})]\geq\mathrm{Var}[\hat{g}(\boldsymbol{X})]\end{aligned}$$

故 $\hat{g}(\boldsymbol{X})$ 方差最小，且无偏，为 UMVUE。

(B) 再证明 $\Rightarrow$，即已知 $\hat{g}(\boldsymbol{X})$ 为 UMVUE。若对任意 $U(\boldsymbol{X})\in\mathscr{U}$，即 $\mathbb{E}[U(\boldsymbol{X})]=0$，那么有 $\hat{g}(\boldsymbol{X})+\lambda\cdot U(\boldsymbol{X})$ 仍然无偏，对任意 λ 成立。又因为 $\hat{g}(\boldsymbol{X})$ 为 UMVUE，有

$$\begin{aligned}&\mathrm{Var}[\hat{g}(\boldsymbol{X})]\leq\mathrm{Var}[\hat{g}(\boldsymbol{X})+\lambda\cdot U(\boldsymbol{X})]\\ \Rightarrow&\mathrm{Var}[\hat{g}(\boldsymbol{X})]\leq\mathrm{Var}[\hat{g}(\boldsymbol{X})]+\lambda^2\mathrm{Var}[U(\boldsymbol{X})]+2\lambda\mathrm{Cov}[\hat{g}(\boldsymbol{X}),U(\boldsymbol{X})]\\ \Rightarrow&\mathrm{Var}[U(\boldsymbol{X})]\cdot\lambda^2+2\mathrm{Cov}[\hat{g}(\boldsymbol{X}),U(\boldsymbol{X})]\cdot\lambda\geq 0 \quad \forall\,\lambda\end{aligned}$$

注意到这是关于 λ 的二次函数，且有根 $\lambda_1 = 0$ 使得函数取 0。则另一个根必须为 0，使得函数恒正，即

$$\lambda_2 = -\frac{2\mathrm{Cov}[\hat{g}(\boldsymbol{X}), U(\boldsymbol{X})]}{\mathrm{Var}[U(\boldsymbol{X})]} = 0$$

故有 $\mathrm{Cov}[\hat{g}(\boldsymbol{X}), U(\boldsymbol{X})] = 0$，命题证毕。

推论 1 设 $T = T(\boldsymbol{X})$ 为 θ 的充分统计量，设 $h(T) = h[T(\boldsymbol{X})]$ 是 $g(\theta)$ 的一个方差有限的无偏估计。令 $\mathscr{U}_T$ 是基于充分统计量 T 的零无偏估计的集合

$$\mathscr{U}_T = \{U(T) : \mathbb{E}_\theta[U(T)] = 0, \forall\, \theta \in \Theta\}$$

则 $h(T)$ 是 $g(\theta)$ 的 UMVUE 的充分必要条件为

$$\mathrm{Cov}_\theta[h(T), \delta(T)] = \mathbb{E}_\theta[h(T) \cdot \delta(T)] = 0 \qquad \forall\, \theta \in \Theta, \forall\, \delta(T) \in \mathscr{U}_T$$

例 5.6 设 $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} U(0, \theta)$，但 $\theta > 1$，求 θ 的 UMVUE。

解答 在例 5.5 中，无法根据充分完全统计量法求 θ 的 UMVUE，于是，这里使用零无偏估计法求 θ 的 UMVUE。

注意到 $T = X_{(n)}$ 是一个充分统计量。设 $U(T)$ 满足 $\mathbb{E}_\theta[U(T)] = 0$ 对任意的 $\theta > 1$ 成立。根据例 5.5 的结论，有

$$\int_0^1 U(t) t^{n-1}\, dt + \int_1^\theta U(t) t^{n-1}\, dt = 0$$

对 θ 求导有 $U(\theta) = 0$ 任意的 $\theta > 1$ 成立。

下面考虑：构造 $\hat{g}(T)$ 使得 $\mathbb{E}_\theta[\hat{g}(T), U(T)] = 0$，即

$$\int_0^1 \hat{g}(t) U(t) t^{n-1}\, dt + \int_0^1 \hat{g}(t) \cdot 0 \cdot t^{n-1}\, dt = 0 \quad \Rightarrow \quad \int_0^1 \hat{g}(t) U(t) t^{n-1}\, dt = 0$$

不妨构造

$$\hat{g}(T) = C \cdot \mathbb{I}_{(0,1)}(T) + Bt \cdot \mathbb{I}_{(1,\theta)}(T)$$

其中 C, B 待定。于是易得 $\int_0^1 \hat{g}(t) U(t) t^{n-1}\, dt = 0$，即 $\mathbb{E}_\theta[\hat{g}(T), U(T)] = 0$。继续，寻找 C, B 使得 $\mathbb{E}_\theta[\hat{g}(T)] = \theta$，即

$$\begin{aligned} \int_0^1 Cn\frac{t^{n-1}}{\theta^n}\, dt + \int_i^\theta Btn\frac{t^{n-1}}{\theta^n}\, dt = \theta \\ \Rightarrow \frac{C}{\theta^n} + \frac{n}{n+1} B \left(\theta - \frac{1}{\theta^n}\right) = \theta \end{aligned}$$

取 $B = \frac{n+1}{n}$，$C = 1$ 即成立。故根据定理 5.3 的推论 1 可知，$\hat{g}(T) = \mathbb{I}_{(0,1)}(X_{(n)}) + \frac{n+1}{n} X_{(n)} \cdot \mathbb{I}_{(1,\theta)}(X_{(n)})$ 是 $g(\theta) = \theta$ 的 UMVUE。

5.3 Cramer-Rao 不等式

Cramer-Rao 不等式是判别一个无偏估计量是否为 UMVUE 的主要方法之一。这一方法的思想如下：设 $\mathscr{U}_g$ 是 $g(\theta)$ 的一切无偏估计构成的类。$\mathscr{U}_g$ 中估计量的方差有一个下界，如果 $g(\theta)$ 的一个无偏估计量 $\hat{g}$ 的方差达到这个下界，则 $\hat{g}$ 就是 $g(\theta)$ 的一个 UMVUE。

定义 5.3 (C-R 正则分布族，C-R 正则条件) 若单参数概率函数族 $\mathscr{F} = \{f(x;\theta) : \theta \in \Theta\}$ 满足下列条件：

(1) 参数空间 Θ 是直线上的某个开区间

(2) 对任何 $x \in \mathscr{X}$ 及 $\theta \in \Theta$，$f(x;\theta) > 0$，即分布族具有共同支撑

(3) 对任何 $x \in \mathscr{X}$ 及 $\theta \in \Theta$，$\partial f(x;\theta)/\partial\theta$ 存在

(4) 概率函数 $f(x;\theta)$ 的积分与微分运算可交换，即

$$\frac{\partial}{\partial\theta}\int f(x;\theta)\,dx = \int \frac{\partial}{\partial\theta} f(x;\theta)\,dx$$

若 $f(x;\theta)$ 为离散随机变量的概率分布，上述条件改为无穷级数和微分运算可交换

(5) 下列数学期望存在，且

$$0 < I(\theta) = \mathbb{E}_\theta\left[\frac{\partial \log f(X;\theta)}{\partial\theta}\right]^2 < \infty$$

则称该分布族为 **C-R 正则分布族**，其中 (1)-(5) 称为 **C-R 正则条件**。$I(\theta)$ 称为该分布的 **Fisher 信息量**（或称为 Fisher 信息函数）。

5.3.1 单参数 C-R 不等式

定理 5.4 (Cramer-Rao 不等式) 设 $\mathscr{F} = \{f(x;\theta) : \theta \in \Theta\}$ 是 C-R 正则分布族，$g(\theta)$ 是定义于参数空间 Θ 上的可微函数。设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是由总体 $f(x;\theta) \in \mathscr{F}$ 中抽取的简单随机样本，$\hat{g}(\boldsymbol{X})$ 是 $g(\theta)$ 的任一无偏估计，且满足下列条件

(6) 积分

$$\int\cdots\int \hat{g}(\boldsymbol{x}) f(\boldsymbol{x};\theta)\,d\boldsymbol{x}$$

可在积分号下对 θ 求导数（此处 $d\boldsymbol{x} = dx_1\cdots dx_n$）

则有不等式成立

$$\mathrm{Var}_\theta[\hat{g}(\boldsymbol{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)} \quad \forall\,\theta \in \Theta$$

称为 **Cramer-Rao 不等式**，简称 **C-R 不等式**。

证明 记 $S(\boldsymbol{x};\theta)=\partial\log f_n(\boldsymbol{x};\theta)/\partial\theta$，其中 $f_n(\boldsymbol{x};\theta)$ 为联合密度函数。容易得

$$\begin{aligned}
\mathbb{E}[S(\boldsymbol{x};\theta)] &= \mathbb{E}\left[\frac{\partial\log f_n(\boldsymbol{x};\theta)}{\partial\theta}\right] = \mathbb{E}\left[\sum_{i=1}^n\frac{\partial\log f(x_i;\theta)}{\partial\theta}\right] = \mathbb{E}\left[\sum_{i=1}^n\frac{\partial f(x_i;\theta)}{\partial\theta}\frac{1}{f(x_i;\theta)}\right]\\
&= \sum_{i=1}^n\int_{\mathscr{X}}\frac{\partial f(x_i;\theta)}{\partial\theta}\cdot\frac{1}{f(x_i;\theta)}\cdot f(x_i;\theta)\,dx_i = \sum_{i=1}^n\int_{\mathscr{X}}\frac{\partial f(x_i;\theta)}{\partial\theta}\,dx_i\\
&= \sum_{i=1}^n\frac{\partial}{\partial\theta}\int_{\mathscr{X}} f(x_i;\theta)\,dx_i = \sum_{i=1}^n\frac{\partial}{\partial\theta}1 = 0
\end{aligned}$$

又注意到

$$\begin{aligned}
\mathrm{Var}_\theta[S(\boldsymbol{x};\theta)] &= \mathrm{Var}_\theta\left[\sum_{i=1}^n\frac{\partial\log f(x_i;\theta)}{\partial\theta}\right] = \sum_{i=1}^n\mathrm{Var}_\theta\left[\frac{\partial\log f(x_i;\theta)}{\partial\theta}\right]\\
&= \sum_{i=1}^n\mathbb{E}_\theta\left[\frac{\partial\log f(x_i;\theta)}{\partial\theta}\right]^2 = nI(\theta)
\end{aligned}$$

而

$$\begin{aligned}
\mathrm{Cov}_\theta[\hat g(\boldsymbol{X}),S(\boldsymbol{X};\theta)] &= \mathbb{E}_\theta[\hat g(\boldsymbol{X})\cdot S(\boldsymbol{X};\theta)] = \int_{\mathscr{X}}\hat g(\boldsymbol{x})\left[\frac{1}{f_n(\boldsymbol{x};\theta)}\frac{\partial f_n(\boldsymbol{x},\theta)}{\partial\theta}\right]f_n(\boldsymbol{x};\theta)\,d\boldsymbol{x}\\
&= \int_{\mathscr{X}}\hat g(\boldsymbol{x})\frac{\partial f_n(\boldsymbol{x},\theta)}{\partial\theta}\,d\boldsymbol{x} = \frac{\partial}{\partial\theta}\int_{\mathscr{X}}\hat g(\boldsymbol{x})f_n(\boldsymbol{x},\theta)\,d\boldsymbol{x}\\
&= \frac{\partial}{\partial\theta}\mathbb{E}_\theta[\hat g(\boldsymbol{X})] = \frac{\partial}{\partial\theta}g(\theta) = g'(\theta)
\end{aligned}$$

然后，由 Cauchy-Schwarz 不等式

$$\mathrm{Var}_\theta[\hat g(\boldsymbol{X})]\cdot\mathrm{Var}_\theta[S(\boldsymbol{X};\theta)]\geq\{\mathrm{Cov}_\theta[\hat g(\boldsymbol{X}),S(\boldsymbol{X};\theta)]\}^2$$

代入既有

$$\begin{aligned}
&\mathrm{Var}_\theta[\hat g(\boldsymbol{X})]\cdot[nI(\theta)]\geq[g'(\theta)]^2\\
\Rightarrow\ &\mathrm{Var}_\theta[\hat g(\boldsymbol{X})]\geq\frac{[g'(\theta)]^2}{nI(\theta)}\quad\forall\,\theta\in\Theta
\end{aligned}$$

□

等号成立条件，由 Cauchy-Schwarz 不等式决定。

注 若 $g(\theta)$ 的无偏估计 $\hat g(\boldsymbol{X})$ 的方差 $\mathrm{Var}_\theta[\hat g(\boldsymbol{X})]$ 达到了 C-R 不等式的下界，则一定是 UMVUE；但未达到下界，不能说明不是 UMVUE。

例 5.7 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从 $N(a,\sigma^2)$ 中抽取的简单随机样本，其中 σ^2 已知。证明 $\bar X$ 为 a 的 UMVUE。

证明 由正态分布为指数族，C-R 正则条件满足。$N(a,\sigma^2)$ 的密度函数为

$$f(x;a)=(2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}$$

从而，Fisher 信息量为

$$I(a)=\mathbb{E}_a\left[\frac{\partial\log f(X;a)}{\partial a}\right]^2=\mathbb{E}_a\left[\frac{(X-a)^2}{\sigma^4}\right]=\frac{\mathrm{Var}_a(X)}{\sigma^4}=\frac{1}{\sigma^2}$$

故 C-R 下界为 $1/[nI(a)]=\sigma^2/n$。而 $\mathrm{Var}(\bar X)=\sigma^2/n$ 达到 C-R 下界，且为无偏估计。由定理 5.4 可知，$\bar X$ 是 a 的 UMVUE。

5.3.2 C-R 不等式等号成立条件

定理 5.5 (C-R 不等式等号成立条件) C-R 不等式等号成立条件，需考虑特定的分布族：

1. 若样本分布族**非指数族**，则其关于 $g(\theta)$ 的无偏估计之方差 **一定不能达到** C-R 下界($\forall\theta\in\Theta$)。
2. 若样本分布族为**指数族**，则其关于 $g(\theta)$ 的无偏估计之方差 **不一定能达到** C-R 下界($\forall\theta\in\Theta$)。

只有当样本分布族为指数族 $f(\boldsymbol{x};\theta)=C(\theta)\exp\{Q(\theta)T(\boldsymbol{x})\}h(\boldsymbol{x})$，且

$$\hat{g}(\boldsymbol{X})=aT(\boldsymbol{X})+b \qquad \mathbb{E}_\theta[\hat{g}(\boldsymbol{X})]=\mathbb{E}_\theta[aT(\boldsymbol{X})+b]=g(\theta)$$

此时方可取到 C-R 下界，其中 $a\neq 0$ 和 b 是与 θ 无关的常数。

评论 非指数族 $\Rightarrow$ 一定不能达到 C-R 下界；指数族 $\Rightarrow$ 不一定能达到 C-R 下界。

证明 由 Cauchy-Schwarz 不等式的取等条件，Cramer-Rao 不等式取等，即有

$$S(\boldsymbol{X};\theta)=a(\theta)\hat{g}(\boldsymbol{X})+b(\theta) \qquad a(\theta)\neq 0$$

下面，我们断言：

$$S(\boldsymbol{X};\theta)=a(\theta)\hat{g}(\boldsymbol{X})+b(\theta) \quad\Leftrightarrow\quad f_n(\boldsymbol{X};\theta)=C(\theta)\exp\{Q(\theta)\hat{g}(\boldsymbol{X})\}h(\boldsymbol{X})$$

其中 $f_n(\boldsymbol{X};\theta)$ 为概率密度函数。

(A) 先证 $\Rightarrow$。注意到 $S(\boldsymbol{X};\theta)=\partial\log f_n(\boldsymbol{X};\theta)/\partial\theta$，于是对任意 $\theta\in\Theta$，作积分

$$\begin{aligned}
\int_{\theta_0}^{\theta} S(\boldsymbol{X};\theta)\,d\theta &= \int_{\theta_0}^{\theta}\frac{\partial\log f_n(\boldsymbol{X};t)}{\partial\theta}\,dt=\hat{g}(\boldsymbol{X})\int_{\theta_0}^{\theta}a(t)\,dt+\int_{\theta_0}^{\theta}b(t)\,dt\\
\Rightarrow \log f_n(\boldsymbol{X};\theta)-\log f_n(\boldsymbol{X};\theta_0) &:= Q(\theta)\hat{g}(\boldsymbol{X})+R(\theta)\\
\Rightarrow f_n(\boldsymbol{X};\theta) &= e^{R(\theta)}\exp\{Q(\theta)\hat{g}(\boldsymbol{X})\}f_n(\boldsymbol{X};\theta_0)
\end{aligned}$$

其中 $Q(\theta)=\int_{\theta_0}^{\theta}a(t)\,dt$，$R(\theta)=\int_{\theta_0}^{\theta}b(t)\,dt$。于是 $f_n(\boldsymbol{X};\theta)$ 能写成指数族的形式，证毕。

(B) 再证 $\Leftarrow$。已知 $f_n(\boldsymbol{X};\theta)$，于是可求 $S(\boldsymbol{X};\theta)$

$$\begin{aligned}
S(\boldsymbol{X};\theta)&=\frac{\partial\log f_n(\boldsymbol{X};\theta)}{\partial\theta}=\frac{\partial}{\partial\theta}\left[\log C(\theta)+Q(\theta)\hat{g}(\boldsymbol{X})+\log h(\boldsymbol{X})\right]\\
&=\frac{\partial\log C(\theta)}{\partial\theta}+Q'(\theta)\hat{g}(\boldsymbol{X})\triangleq a(\theta)\hat{g}(\boldsymbol{X})+b(\theta)
\end{aligned}$$

其中 $a(\theta)=Q'(\theta)\neq 0$，$b(\theta)=\partial\log C(\theta)/\partial\theta$，证毕。

例 5.8 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为自 Poisson 分布 $\mathrm{Poi}(\lambda)$ 中抽取的简单样本。证明只有 $g(\lambda)$ 是 λ 的线性函数时，才存在 $g(\lambda)$ 的无偏估计，其方差能处处达到 C-R 下界（即 $\forall\theta\in\Theta$）。

证明 样本 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 的联合概率函数为

$$\begin{aligned}
f(\boldsymbol{x};\lambda)=\prod_{i=1}^{n}f(x_i;\lambda)&=\frac{\lambda^{\sum_{i=1}^{n}x_i}e^{-n\lambda}}{x_1!\cdots x_n!}\\
&=\frac{e^{-n\lambda}\exp\{n\bar{x}\log\lambda\}}{x_1!\cdots x_n!}\\
&=C(\lambda)\exp\{Q(\lambda)T(\boldsymbol{x})\}h(\boldsymbol{x})
\end{aligned}$$

为指数族，其中 $C(\lambda)=e^{-n\lambda}$，$Q(\lambda)=\log\lambda$，$T(\boldsymbol{x})=n\bar{x}=\sum_{i=1}^n x_i$，$h(\boldsymbol{x})=1/(x_1!\cdots x_n!)$。由 C-R 不等式取等条件，即定理 5.5 可知，只有当

$$g(\lambda)=\mathbb{E}_\lambda[aT(\boldsymbol{X})+b]=a\mathbb{E}_\lambda[\sum_{i=1}^n x_i]+b=an\cdot\lambda+b$$

为 λ 的线性函数时，无偏估计 $\hat{g}(\boldsymbol{X})=aT(\boldsymbol{X})+b$ 的方差才能达到 C-R 下界，且为 $g(\lambda)$ 的 UMVUE。

例如：取 $g(\lambda)=\lambda$，此时取 $a=1/n$，$b=0$，就有 $\hat{g}(\boldsymbol{X})=\bar{X}$ 为 UMVUE。

5.3.3 多参数 C-R 不等式

定理 5.6 (多参数 C-R 不等式) 设 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)\in\mathbb{R}^k$，总体概率函数记作 $f(x;\boldsymbol{\theta})$，设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从总体 $f(x;\boldsymbol{\theta})$ 中抽取的简单随机样本。设 $\hat{\boldsymbol{\theta}}=\hat{\theta}(\boldsymbol{X})=(\hat{\theta}_1,\cdots,\hat{\theta}_k)\in\mathbb{R}^k$ 是 $\boldsymbol{\theta}$ 的一个无偏估计。记 $\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ 为 $\hat{\boldsymbol{\theta}}$ 的协方差阵

$$\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})=\mathbb{E}_{\boldsymbol{\theta}}\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right]\in\mathbb{R}^{k\times k}$$

为非负定阵。则此时的 Cramer-Rao 不等式为

$$\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\geq[n\boldsymbol{I}(\boldsymbol{\theta})]^{-1}$$

其中 $\boldsymbol{I}(\boldsymbol{\theta})$ 是总体的 Fisher 信息矩阵

$$\boldsymbol{I}(\boldsymbol{\theta})=\mathbb{E}_{\boldsymbol{\theta}}\left[\left(\frac{\partial\log f(X;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\log f(X;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^T\right]\in\mathbb{R}^{k\times k}$$

特别地，若记 $\boldsymbol{I}^*(\boldsymbol{\theta})=[\boldsymbol{I}(\boldsymbol{\theta})]^{-1}$，则有

$$\mathrm{Var}_{\boldsymbol{\theta}}(\hat{\theta}_i)\geq\frac{\boldsymbol{I}^*_{ii}(\boldsymbol{\theta})}{n}\qquad i=1,2,\cdots,k \tag{7}$$

其中 $\boldsymbol{I}^*_{ii}(\boldsymbol{\theta})$ 表示 $\boldsymbol{I}^*(\boldsymbol{\theta})$ 的第 i 个对角元素。

注 Fisher 信息矩阵还可以表示为

$$\boldsymbol{I}(\boldsymbol{\theta})=\mathbb{E}_{\boldsymbol{\theta}}\left[-\left(\frac{\partial^2\log f(X;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)\right]\in\mathbb{R}^{k\times k}$$

注 式 7 成立，需要证明如下结论

$$\boldsymbol{A}\geq 0\quad\Rightarrow\quad\boldsymbol{A}_{ii}\geq 0$$

这个结论是显然的：由非负定的性质，任意向量 $\boldsymbol{v}$ 都有 $\boldsymbol{v}\boldsymbol{A}\boldsymbol{v}^T\geq 0$。只要取 $\boldsymbol{v}=\boldsymbol{e}_i=(0,\cdots,1,\cdots,0)$，其中 $\boldsymbol{e}_i$ 表示第 i 个分量为 1，其他分量为 0。于是有 $\boldsymbol{e}_i\boldsymbol{A}\boldsymbol{e}_i^T=A_{ii}\geq 0$，证毕。

例 5.9 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从正态总体 $N(a,\sigma^2)$ 中抽取的简单样本，记 $\boldsymbol{\theta}=(a,\sigma^2)$，其中 $\theta_1=a$，$\theta_2=\sigma^2$。求 $\boldsymbol{\theta}$ 的两个分量无偏估计方差的 C-R 下界，并将其与 θ_1 和 θ_2 的无偏估计 $\bar{X}$ 和 S^2 的方差进行比较。

解答 正态随机变量的密度函数为

$$f(x;\boldsymbol{\theta})=(2\pi\theta_2)^{-1/2}\exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2}\right\}$$

可知

$$\frac{\partial\log f(x;\boldsymbol{\theta})}{\partial\theta_1}=\frac{x-\theta_1}{\theta_2}\qquad\frac{\partial\log f(x;\boldsymbol{\theta})}{\partial\theta_2}=\frac{-\theta_2+(x-\theta_1)^2}{2\theta_2^2}$$

由此计算信息矩阵

$$\boldsymbol{I}_{11}(\boldsymbol{\theta})=\frac{1}{\sigma^2}\quad\boldsymbol{I}_{22}(\boldsymbol{\theta})=\frac{1}{2\sigma^4}\quad\boldsymbol{I}_{12}(\boldsymbol{\theta})=\boldsymbol{I}_{21}(\boldsymbol{\theta})=0$$

故有

$$n\boldsymbol{I}(\boldsymbol{\theta})=\begin{pmatrix}\dfrac{n}{\sigma^2}&0\\0&\dfrac{n}{2\sigma^4}\end{pmatrix}\qquad[n\boldsymbol{I}(\boldsymbol{\theta})]^{-1}=\begin{pmatrix}\dfrac{\sigma^2}{n}&0\\0&\dfrac{2\sigma^4}{n}\end{pmatrix}$$

记 $\hat{\theta}_1=\bar{X}$，$\hat{\theta}_2=S^2$，则由 Cramer-Rao 不等式可知

$$\mathrm{Cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})=\mathrm{Cov}_{\boldsymbol{\theta}}\begin{pmatrix}\hat{\theta}_1\\\hat{\theta}_2\end{pmatrix}\geq\begin{pmatrix}\dfrac{\sigma^2}{n}&0\\0&\dfrac{2\sigma^4}{n}\end{pmatrix}$$

注意到 $\mathrm{Var}_{\boldsymbol{\theta}}(\hat{\theta}_1)=\sigma^2/n$ 达到 C-R 下界，故 $\hat{\theta}_1=\bar{X}$ 是 $\theta_1=a$ 的 UMVUE。而根据 $(n-1)S^2/\sigma^2\sim\chi^2_{n-1}$ 知

$$\mathrm{Var}_{\boldsymbol{\theta}}(\hat{\theta}_2)=2(n-1)\cdot\frac{\sigma^4}{(n-1)^2}=\frac{2\sigma^4}{n-1}>\frac{2\sigma^4}{n}$$

故 $\hat{\theta}_2=S^2$ 的方差大于 C-R 下界，不是 $\theta_2=\sigma^2$ 的 UMVUE。

5.3.4 有效估计和估计的效率

定义 5.4 (有效估计) 设 $\hat{g}_n(\boldsymbol{X})$ 是 $g(\theta)$ 的无偏估计，则比值

$$e_{\hat{g}_n}(\theta)=\frac{[g'(\theta)]^2/[nI(\theta)]}{\mathrm{Var}_\theta[\hat{g}_n(\boldsymbol{X})]}$$

称为无偏估计 $\hat{g}_n(\boldsymbol{X})$ 的效率（efficiency）。显然 $0<e_{\hat{g}_n}(\theta)\leq1$，当 $e_{\hat{g}_n}(\theta)=1$ 时，称 $\hat{g}_n(\boldsymbol{X})$ 是 $g(\theta)$ 的**有效估计**（effective estimation）。若 $\hat{g}_n(\boldsymbol{X})$ 不是 $g(\theta)$ 的有效估计，但 $\lim_{n\to\infty}e_{\hat{g}_n}(\theta)=1$，则称 $\hat{g}_n(\boldsymbol{X})$ 是 $g(\theta)$ 的**渐近有效估计**（asymptotically effective estimation）。

例 5.10 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从 $N(a,\sigma^2)$ 中抽取的简单随机样本。

1. 当 a 未知时，证明样本方差 S^2 不是 σ^2 的有效估计，但是渐近有效估计。
2. 当 a 已知时，求 σ^2 的有效估计。

证明 (1) 当 a 未知时，由例 5.9 可知 S^2 的方差为 $2\sigma^4/(n-1)$ 达不到 C-R 下界 $2\sigma^4/n$，故不是有效估计。估计的效率为 $e_{S^2}(\sigma^2)=(n-1)/n<1$，但是

$$\lim_{n\to\infty}e_{S^2}(\sigma^2)=\lim_{n\to\infty}\frac{n-1}{n}=1$$

故 S^2 是 σ^2 的渐近有效估计。

解答 (2) 当 a 已知，设 $S_n^2 = \sum_{i=1}^n (X_i - a)^2/n$，容易得 $nS_n^2/\sigma^2 \sim \chi_n^2$，于是有

$$\mathrm{Var}(S_n^2) = \frac{\sigma^4}{n^2} \cdot \mathrm{Var}\left(\frac{nS_n^2}{\sigma^2}\right) = \frac{\sigma^4}{n^2} \cdot 2n = \frac{2\sigma^4}{n}$$

达到了 C-R 下界，故此时 σ^2 的 UMVUE 为 S_n^2。

例 5.11 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 是从下列含有位置参数的指数分布族中抽取的简单样本，

$$f(x;\theta) = e^{-(x-a)} \cdot \mathbb{I}(x > a)$$

求 a 的 UMVUE。

解答 容易证 $X_{(1)}$ 是 a 的充分完全统计量，且密度函数为 $f_{X_{(1)}}(x) = ne^{-n(x-a)} \cdot \mathbb{I}(x > a)$，于是

$$\mathbb{E}_a(X_{(1)}) = n \int_a^{\infty} xe^{-n(x-a)}\, dx = a + \frac{1}{n}$$

故 $X_{(1)} - 1/n$ 是 a 的无偏估计，由 Lehmann-Scheffe 定理 5.2 可知 $X_{(1)} - 1/n$ 是 a 的 UMVUE。

但是，概率函数 $f_n(\boldsymbol{x}; a) = \exp\{-\sum_{i=1}^n (x_i - a)\} \cdot \mathbb{I}(x_{(1)} > a)$ 的支撑集 $\{\boldsymbol{x} : f_n(\boldsymbol{x}; a) > 0\} = \{\boldsymbol{x} : \forall\, x_i > a\}$ 与参数 a 有关。所以，不满足 C-R 正则条件，不能讨论有效性。