21 数理统计期中

$\mathscr{NULIOUS}$

一、（20 分）假设 $X_1, X_2, \cdots, X_n$ 为来自 Poisson 总体 $\mathrm{P}(\lambda)$ 的一组简单样本，试

（1）将抽样分布表示为指数族的自然形式，并给出自然参数空间；

（2）证明统计量 $T=\sum_{i=1}^{n} X_i$ 为充分完备统计量。

（3）证明 $g(\lambda)=e^{-\lambda}$ 不存在可以达到 C-R 不等式下界的无偏估计.

解：（1）样本联合密度

$$f\left(x_1, \cdots, x_n ; \lambda\right)=e^{-n \lambda} \frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{x_1 ! \cdots x_n !}$$

写成指数族的自然形式为

$$f\left(x_1, \cdots, x_n ; \lambda\right)=e^{-n \lambda} \cdot \frac{1}{x_1 ! \cdots x_n !} \cdot e^{\sum\limits_{i=1}^{n} x_i \ln \lambda}$$

这里

$$C(\theta)=e^{-n \lambda} \quad h(\boldsymbol{X})=\frac{1}{x_1 ! \cdots x_n !}$$

自然参数为 $\ln \lambda$，又 $\lambda \in(0,+\infty)$

则自然参数空间为

$$\Theta=\{\ln \lambda: \ln \lambda \in \mathbb{R}\}$$

（2）由因子分解定理和自然参数空间有内点，则 $T(\boldsymbol{X})=\sum\limits_{i=1}^{n} X_i$ 为充分完全统计量

（3）无偏估计能达到 C-R 下界 $\Longleftrightarrow$ 分布为单参数指数族且 UMVUE 为充分完全统计量 $T(\boldsymbol{X})$ 的线性函数

先求 UMVUE

解一：注意到

$$e^{-\lambda}=P(X_1=0)$$

再仿照 24 期中 3（1）可以得到 UMVUE

$$\widehat{e^{-\lambda}}=\hat{g}(T)=\left(\frac{n-1}{n}\right)^T$$

解二：首先

$$T(\boldsymbol{X}) \sim \mathrm{P}(n \lambda)$$

设 UMVUE 为 $\hat{g}(T)$，则

$$\mathbb{E}[\hat{g}(T)]=\sum_{k=0}^{\infty} \hat{g}(k) \frac{(n k)^k}{k !} \cdot e^{-n \lambda}=e^{-\lambda}$$

也就是

$$\sum_{k=0}^{\infty}\hat{g}(k)\frac{(n\lambda)^k}{k!}=e^{(n-1)\lambda}=\sum_{k=0}^{\infty}\frac{(n-1)^k\lambda^k}{k!}$$

就能得到

$$\hat{g}(T)=\left(\frac{n-1}{n}\right)^T$$

显然 UMVUE 不是充分完全统计量的线性函数，因此不能达到 C-R 下界

数值验证：

下面求总体（单个样本）的 Fisher 信息量

$$I(\lambda)=-\mathbb{E}\left[\frac{\partial\ln f(x;\lambda)^2}{\partial^2\lambda}\right]=-\mathbb{E}\left[-\frac{X}{\lambda^2}\right]=\frac{1}{\lambda}$$

这里 X 为总体分布，且密度为

$$f(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}$$

因此 C-R 下界为

$$\frac{\left(\frac{\partial\left(e^{-\lambda}\right)}{\partial\lambda}\right)^2}{nI(\lambda)}=\frac{\lambda e^{-2\lambda}}{n}$$

再计算 UMVUE 的方差

$$\begin{aligned}\operatorname{Var}(\hat{g}(T))&=\mathbb{E}\left[\left(\frac{n-1}{n}\right)^{2T}\right]-(\mathbb{E}[\hat{g}(T)])^2\\&=\mathbb{E}\left[\left(\frac{n-1}{n}\right)^{2T}\right]-e^{-2\lambda}\end{aligned}$$

其中

$$\begin{aligned}E\left[\left(\frac{n-1}{n}\right)^{2T}\right]&=\sum_{k=0}^{\infty}\left(\frac{n-1}{n}\right)^{2k}\frac{(n\lambda)^k}{k!}e^{-n\lambda}\\&=e^{-n\lambda}\sum_{k=0}^{\infty}\frac{\left(\frac{(n-1)^2\lambda}{n}\right)^k}{k!}\\&=e^{-n\lambda}\cdot e^{\frac{(n-1)^2\lambda}{n}}\\&=e^{-2\lambda+\frac{\lambda}{n}}\end{aligned}$$

最后

$$\operatorname{Var}(\hat{g}(T))=e^{-2\lambda+\frac{\lambda}{n}}-e^{-2\lambda}$$

不能达到 C-R 下界

二．（30 分）设 $X_1,\ldots,X_n$ 为来自 0-1 分布 $\mathrm{B}(1,p),0<p<1$ 的一组简单样本，试

（1）求 $g(p)=(1-p)^2$ 的矩估计量和极大似然估计量，并说明是否为无偏估计。

（2）求 $g(p)$ 的 UMVUE，其方差是否达到 C-R 不等式的下界？

（3）证明 $g(p)$ 的极大似然估计量具有渐近正态性.

解：（1）由于 $\mathbb{E}[X]=p$，矩估计量为

$$\hat{g(p)}_M=(1-\overline{X})^2$$

又样本联合密度

$$f\left(x_1,\cdots,x_n;p\right)=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

求导有

$$\frac{\partial \ln f}{\partial p}=\frac{\sum_{i=1}^n x_i}{p}-\frac{n-\sum_{i=1}^n x_i}{1-p}=0.$$

解得

$$\hat{p}_{MLE}=\overline{X}$$

再由 MLE 的不变性，有

$$g\hat{(p)}_{MLE}=\left(1-\overline{X}\right)^2$$

无偏性：

$$\begin{aligned}\mathbb{E}\left[(1-\overline{X})^2\right]&=\mathbb{E}\left[1-2\overline{X}+\overline{X}^2\right]\\&=1-2\mathbb{E}[\overline{X}]+\frac{1}{n^2}\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right]\\&=1-2p+p^2+\frac{p(1-p)}{n}\\&\neq(1-p)^2\end{aligned}$$

这就说明两个估计均不无偏

（2）注意二项分布为指数族，且自然参数空间有内点，易知 $T(\boldsymbol{X})=\sum\limits_{i=1}^n X_i$ 为充分完全统计量

先找一个无偏估计

$$\mathbb{I}_{\{X_1=0,X_2=0\}}$$

则 UMVUE 为

$$\begin{aligned}\hat{g}(p)_{UMVUE}&=\mathbb{E}\left[\mathbb{I}_{\{X_1=0,X_2=0\}}\mid\sum_{i=1}^n x_i=t\right]\\&=P\left(X_1=0,X_2=0\mid\sum_{i=1}^n X_i=t\right)\\&=\frac{P\left(X_1=0,X_2=0,\sum_{i=3}^n X_i=t\right)}{P\left(\sum_{i=1}^n X_i=t\right)}\\&=\frac{(n-t)(n-t-1)}{n(n-1)}\end{aligned}$$

也就是

$$\hat{g}(p)_{UMVUE}=\frac{(n-T)(n-T-1)}{n(n-1)}$$

它不是 $T(\boldsymbol{X})=\sum\limits_{i=1}^n X_i$ 的线性函数，因此不能达到 C-R 下界

数值验证：先求总体（单个样本）的 Fisher 信息量

$$\begin{aligned}I\left(p\right)&=-\mathbb{E}\left[\frac{\partial^2}{\partial p^2}\ln f\left(x;p\right)\right]\\&=\frac{1}{p(1-p)}\end{aligned}$$

这里总体的密度为

$$f(x;p)=p^x(1-p)^{1-x}$$

因此 C-R 下界为

$$\frac{\frac{\partial g(p)}{\partial p}}{nI(p)}=\frac{4p(1-p)^3}{n}$$

又

$$\operatorname{Var}(\hat{g}(p)_{UMVUE})=\operatorname{Var}\left(\frac{(n-T)(n-T-1)}{n(n-1)}\right)=\frac{\operatorname{Var}\left(n^2-n+(1-2n)T+T^2\right)}{n^2(n-1)^2}$$

其中

$$\begin{aligned}\operatorname{Var}\left(n^2-n+(1-2n)T+T^2\right)&=\operatorname{Var}\left((1-2n)T+T^2\right)\\&=\mathbb{E}\left[\left[(1-2n)T+T^2\right]^2\right]-\left(\mathbb{E}\left[(1-2n)T+T^2\right]\right)^2\end{aligned}$$

这里由于 $T\sim \mathrm{B}(n,p)$

$$\begin{aligned}\mathbb{E}\left[\left[(1-2n)T+T^2\right]^2\right]&=\mathbb{E}\left[(1-2n)^2T^2+(2-4n)T^3+T^4\right]\\&=\mathbb{E}\left[(1-2n)^2T^2\right]+\mathbb{E}\left[(2-4n)T^3\right]+\mathbb{E}\left[T^4\right]\\&=(1-2n)^2np(1-p+np)+(2-4n)(1-3p+3np+(2-3n+n^2)p^2)\\&+np(1-p)(1-6p+6p^2)+7np(1-p)(n-1)p+6n(n-1)(n-2)p^3+n^3p^4\end{aligned}$$

另一方面

$$\mathbb{E}\left[(1-2n)T+T^2\right]=(1-2n)p+np(1-p+np)$$

这就说明了 UMVUE 不能达到 C-R 下界

（3）解一：直接利用 MLE 的渐进正态性

$$\sqrt{n}\left(\hat{p}_{MLE}-p\right)\xrightarrow{D}N\left(0,\frac{1}{I(p)}\right)$$

也就是

$$\sqrt{n}\left(\hat{p}_{MLE}-p\right)\xrightarrow{D}N\left(0,p(1-p)\right)$$

再使用 Δ 方法

$$\sqrt{n}\left(g(\hat{p})_{MLE}-g(p)\right)\xrightarrow{D}N\left(0,4p(1-p)^3\right)$$

解二：$\overline{X}$ 有 CLT

$$\sqrt{n}\left(\overline{X}-p\right)\xrightarrow{D}N\left(0,p(1-p)\right)$$

再使用 Δ 方法

$$\sqrt{n}\left(g(\hat{p})_{MLE}-g(p)\right)\xrightarrow{D}N\left(0,4p(1-p)^3\right)$$

三．（15 分）设 $X_1,\ldots,X_n$ 来自正态总体 $N(\mu,1)$ 的一组简单随机样本，试 $(0<\alpha<1)$

（1）证明样本平均值 $\overline{X}$ 与统计量 $X_1-\overline{X}$ 相互独立.

（2）求概率 $P\left(X_1\leq 0\right)$ 的 UMVUE 以及其置信水平为 $1-\alpha$ 的置信区间.

解：（1）注意方差已知，由 $N(\mu,1)$ 为指数族, 容易得到 $T(\boldsymbol{X})=\overline{X}$ 为充分完全统计量

设

$$Y_i=X_i-\mu\sim N(0,1)$$

则

$$X_1-\overline{X}=(X_1-\mu)-\frac{\sum\limits_{i=1}^{n}(X_i-\mu)}{n}=Y_1-\overline{Y}$$

分布与 μ 无关，是辅助量，由 Basu 定理知二者独立

（2）显然 $\mathbb{I}_{\{X_1\leq 0\}}$ 就是无偏估计，设要求的 UMVUE 为 $h(T)$

$$\begin{aligned}h(T)&=\mathbb{E}\left[\mathbb{I}_{\{X_1\leq 0\}}\mid T(X)\right]\\&=P\left(X_1\leqslant 0\mid T(X)\right)\\&=P\left(X_1-\overline{X}\leqslant-\overline{X}\mid\overline{X}\right)\\&\overset{\text{独立性}}{=}P\left(X_1-\overline{X}\leqslant-\overline{X}\right)\end{aligned}$$

注意 Y_1 与 $\overline{Y}$ 不独立，有

$$X_1-\overline{X}\sim Y_1-\overline{Y}\sim N\left(1,1-\frac{1}{n}\right)$$

则

$$h(T)=\Phi\left(-\frac{\overline{X}}{\sqrt{1-\frac{1}{n}}}\right)$$

这里 $\Phi(x)$ 为 $N(0,1)$ 的分布函数，又

$$\overline{X}\sim N\left(\mu,\frac{1}{n}\right)\quad\sqrt{n}(\overline{X}-\mu)\sim N(0,1)$$

则

$$P(X_1\leq 0)=\Phi(-\mu)$$

即 $P(X_1\leq 0)$ 关于 μ 单调递减，只需要求 μ 的置信区间即可另外 μ 有置信系数为 $1-\alpha$ 的置信区间

$$\left[\overline{X}-\frac{u_{\frac{\alpha}{2}}}{\sqrt{n}},\overline{X}+\frac{u_{\frac{\alpha}{2}}}{\sqrt{n}}\right]$$

则 $P(X_1\leq 0)$ 有置信区间

$$\left[\Phi\left(-\left(\overline{X}+\frac{u_{\frac{\alpha}{2}}}{\sqrt{n}}\right)\right),\Phi\left(-\left(\overline{X}-\frac{u_{\frac{\alpha}{2}}}{\sqrt{n}}\right)\right)\right]$$

四．（15 分）设 $X_1,X_2,\cdots,X_n$ 为来自如下分布的一组简单样本

X	0	1	2
P	$\frac{1}{2}\left[(1-\theta)^2+\theta^2\right]$	$2\theta(1-\theta)$	$\frac{1}{2}\left[(1-\theta)^2+\theta^2\right]$

其中 $0<\theta<1/2$ 为参数。试利用重参数化方法和极大似然估计的不变性

（1）求 θ 的极大似然估计量，并求其渐近分布.

（2）由此给出 θ 的一个（渐近）置信水平为 $1-\alpha$ 的置信区间 $(0<\alpha<1)$.

解：（1）重参数化

$$P(X=1)=P(X=3)\triangleq\mu=\frac{1}{2}\left[(1-\theta)^2+\theta^2\right]$$

则

$$P(X=2)=1-2\mu$$

同时令 n_0,n_1,n_2 为 n 个样本中取值为 0，1，2 的样本个数，有 $n_0+n_1+n_2=n$，则样本联合密度为

$$f\left(x_1,\cdots,x_n;\mu\right)=\mu^{n_0}(1-2\mu)^{n_1}\mu^{n_2}$$

进而

$$\ln f\left(x_1,\cdots,x_n;\mu\right)=n_0\ln\mu+n_1\ln(1-2\mu)+n_2\ln\mu$$

$$\frac{\partial\ln f}{\partial\mu}=\frac{n_0}{\mu}+\frac{n_2}{\mu}-\frac{2n_1}{1-2\mu}=0$$

得到

$$\hat{\mu}_{MLE}=\frac{n_0+n_2}{2n}$$

注意 $\theta<\frac{1}{2}$，反解 θ 有

$$\hat{\mu}_{MLE}=\frac{1}{2}\left[(1-\hat{\theta}_{MLE})^2+\hat{\theta}_{MLE}\right]$$

得到

$$\hat{\theta}_{MLE}=\frac{1-\sqrt{1-\frac{2n_1}{n}}}{2}$$

由 MLE 的渐进正态性

$$\sqrt{n}\left(\hat{\theta}_{MLE}-\theta\right)\xrightarrow{D}N\left(0,\frac{1}{I(\theta)}\right)$$

这里总体（单个样本）的 Fisher 信息量为

$$\begin{aligned}I(\theta)=&-\mathbb{E}\left[\ln'' f(x;\theta)\right].\\=&-\big[P(x=0)\cdot\ln'' f(x=0;\theta)+P(x=1)\cdot\ln'' f(x=1;\theta)\\&+P(x=2)\cdot\ln'' f\left(x=2;\theta\right)\big]\end{aligned}$$

对 $X=0$ 或 2

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

进而

$$\frac{\partial^2}{\partial\theta^2}\ln f=\frac{4\left(2\theta^2-2\theta+1\right)-2(2\theta-1)(4\theta-2)}{\left(2\theta^2-2\theta+1\right)^2}$$

而对 $X=1$

$$\frac{\partial^2}{\partial\theta^2}\ln f(x=1;\theta)=-\frac{1}{\theta^2}-\frac{1}{(1-\theta)^2}$$

则

$$I(\theta)=\frac{2}{2\theta^2-2\theta+1}+\frac{1}{\theta(1-\theta)}$$

就有

$$\sqrt{n}\left(\hat{\theta}_{MLE}-\theta\right) \xrightarrow{D} N\left(0, \frac{1}{\frac{2}{2\theta^2-2\theta+1}+\frac{1}{\theta(1-\theta)}}\right)$$

（2）置信区间为

$$\left[\hat{\theta}_{MLE}-\frac{u_{\frac{\alpha}{2}}}{\sqrt{nI(\theta)}}, \hat{\theta}_{MLE}+\frac{u_{\frac{\alpha}{2}}}{\sqrt{nI(\theta)}}\right]$$

五.（20 分）设 $X_1,\ldots,X_n, i.i.d \sim U(0,\theta)$ ，其中 $\theta>1$ 为未知参数. 记 $X_{(n)}=\max\limits_{1\le i\le n} X_i$ ，试

（1）证明 $X_{(n)}$ 是充分但不完备的统计量.

（2）求 θ 的 UMVUE.

解:（1）样本联合密度为

$$f(x_1,\cdots,x_n;\theta)=\frac{1}{\theta^n}\mathbb{I}_{\left\{X_{(n)}<\theta\right\}}$$

取 $g(T(\boldsymbol{X});\theta)=\theta^{-n}\cdot I_{\left\{x_{(n)}<\theta\right\}}, h(\boldsymbol{X})=1$, 由因子分解定理知道 $T(\boldsymbol{X})=X_{(n)}$ 为充分统计量

注意 $\theta>1$，为了证明 $T(\boldsymbol{X})=X_{(n)}$ 不是 θ 的完全统计量. 只需寻找 t 的某一实函数 $\varphi(t)$ ，满足 $\mathbb{E}_\theta[\varphi(T)]=0$ ，即

$$\int_0^\theta \varphi(t)t^{n-1}\,\mathrm{d}t=0,\quad \theta>1$$

但 $\mathbb{P}_\theta\left(\varphi\left(X_{(n)}\right)=0\right)<1$

注意到上式对 θ 求导后只能得到 $\varphi(t)=0, t>1$. 因此我们只要构造合适的 $\varphi(t), t\le 1$ ，使得

$$\int_0^1 \varphi(t)t^{n-1}\,\mathrm{d}t=0$$

即可，例如

$$\varphi(t)=\begin{cases} t^{1-n} & t\le\frac{1}{2}\\ -t^{1-n} & \frac{1}{2}<t\le 1\\ 0 & 1<t\le\theta\end{cases}$$

（2）统计量只充分不完全，考虑零无偏法

$X_{(n)}$ 有密度

$$g(t)=n\theta^{-n}t^{n-1}\mathbb{I}_{(0,\theta)}(t)$$

设 $\mathbb{E}[\delta(t)]=0$，即

$$\int_0^\theta \delta(t)n\theta^{-n}t^{n-1}dt=0$$

也就是

$$\int_0^\theta \delta(t)t^{n-1}dt=0$$

进而

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

求导有

$$\delta(t)\equiv 0\quad \forall t>1$$

设要求的 UMVUE 为 $h(T)$，它要满足以下条件

$$\mathbb{E}[h(T)\delta(T)] = 0$$

$$\int_0^\theta \delta(t)h(t)n\theta^{-n}t^{n-1}dt = 0$$

注意 $\delta(t) \equiv 0 \quad \forall t > 1$，上式也就说明

$$\int_0^1 \delta(t)h(t)n\theta^{-n}t^{n-1}dt = 0$$

又 $h(T)$ 无偏，即 $\mathbb{E}[h(T)] = \theta$

$$\theta = \int_0^\theta h(t)n\theta^{-n}t^{n-1}dt$$

待定 $h(T) = c, 0 \le T < 1$，有

$$c\int_0^1 \frac{nt^{n-1}}{\theta^n}dt = c\theta^{-n}$$

想要

$$\int_1^\theta h(t)\frac{nt^{n-1}}{\theta^n}dt$$

含有 θ^{-n} 及 θ 项，待定剩下部分为

$$h(T) = \begin{cases} c & 0 \le T < 1 \\ bT + d & T \ge 1 \end{cases}$$

代入积分的值则有

$$c = 1 \quad b = \frac{n+1}{n} \quad d = 0$$

最后

$$h(T) = \begin{cases} 1 & 0 \le T < 1 \\ \frac{n+1}{n}T & T \ge 1 \end{cases}$$