1 第三次作业

习题 10-3

$X_1,\cdots,X_n$ 是抽自总体 $N(\theta,\theta)$ 的随机样本, 其中 $\theta>0$.
(a) 证明 θ 的 MLE, 即 $\hat{\theta}$ 是二次方程 $\theta^2+\theta-W=0$ 的一个根, 其中 $W=\frac{1}{n}\sum_{i=1}^{n}X_i^2$, 并确定哪个根是 MLE.
(b) 用 10.1.3 节中的技术求 $\hat{\theta}$ 的近似方差，并讨论其渐进正态性.

解答（思路）

(a) 设 $l(\theta)$ 为对数似然函数，令 $l'(\theta)=0$，整理可得 $\theta^2+\theta-W=0$. 而 MLE 是 $l(\theta)$ 的最大值点，故还应满足 $l''(\theta)<0$，所以 MLE 为二次方程较大的根 $\frac{-1+\sqrt{1+4W}}{2}$.

(b) 直接用定理 10.1.12，**<u>但是一定要验证条件 A1-A6</u>**(至于此处，可以说“正态分布是指数族”)。总体的 Fisher 信息量 $I(\theta)=\frac{2\theta+1}{2\theta^2}$.
所以 $\sqrt{n}[(\hat{\theta})-(\theta)]\xrightarrow{L}N(0,v(\theta))$，其中 $v(\theta)=\frac{1}{I(\theta)}=\frac{2\theta^2}{2\theta+1}$.
所以 $\hat{\theta}$ 的近似方差为 $\frac{2\theta^2}{n(2\theta+1)}$.

[0] 习题 3-5

设总体 X 服从均匀分布 $U(0,2\theta)$, 令 $X_1,\cdots,X_n$ 为从总体 X 中抽取的简单样本,
(1) 证明 $\hat{\theta}_1=\bar{X}$ 和 $\hat{\theta}_2=(n+1)X_{(n)}/(2n)$ 为 θ 的无偏估计.
(2) 证明 $\hat{\theta}_1$ 为 θ 的强相合估计, $\theta_2^*=X_{(n)}/2$ 为 θ 的弱相合估计.
(3) 求 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的方差, 问哪一个更有效?
(4) 求 $\hat{\theta}_1$ 关于 $\hat{\theta}_2$ 的 ARE.

解答（思路）

(1) 对于 $\hat{\theta}_1$，显然. 对于 $\hat{\theta}_2$，注意次序统计量 $X_{(n)}$ 密度函数为 $\frac{nt^{n-1}}{(2\theta)^n}$ 即可.
(2) 对于 $\hat{\theta}_1$，由 SLLN，显然. 对于 $\theta_2^\star$，注意：依概率收敛于常数 $\iff$ 依分布收敛于常数 $\iff$ $\theta_2^\star$ 的分布函数弱收敛于 θ 的分布函数.

而 $P(\theta_2^\star \le t) = 1(t < \theta)(t/\theta)^n + 1(t \ge \theta)$，当 $n \to \infty$ 时，点点收敛于 $1(t \ge \theta)$，即 θ 的分布函数.
(3)$Var(\hat{\theta}_1) = \frac{Var(X)}{n} = \frac{\theta^2}{3n}$. 由 $X_{(n)}$ 密度函数可得 $Var(\hat{\theta}_2) = \frac{\theta^2}{n(n+2)}$. 所以 $\hat{\theta}_2$ 更有效.
(4) 在进入正式讨论之前，先注意以下几个事实：
(a)$\hat{\theta}_2$ 的渐近正态性不能用定理 10.1.12，因为均匀分布的密度函数的支撑集与 θ 有关，不满足条件 A3.
(b)$\sqrt{n}(\hat{\theta}_2 - \theta) \xrightarrow{L^2} 0$. 因为 $\sqrt{n}(\hat{\theta}_2 - \theta)$ 的二阶矩收敛于 0. 由 L^2 收敛可得依分布收敛，所以可以看成 $\sqrt{n}(\hat{\theta}_2 - \theta) \xrightarrow{L} N(0,0)$.
(c) 课件 2 的 13 页的例 2.4 有如下命题：设 $X_1, \cdots, X_n$ i.i.d. $\sim U(0,\theta), \theta > 0$, 则 $\hat{\theta}_{MLE} = X_{(n)}$, 满足 $n\left(\theta - X_{(n)}\right) \xrightarrow{L} \operatorname{Exp}\left(\theta^{-1}\right)$, i.e. $\hat{\theta}_{MLE}$ 的渐近分布为指数分布。（证明过程也不麻烦，算出分布函数，再求极限即可.）
(d) 当 $X_{(n)} \to 2\theta$ 时，$\hat{\theta}_2 \to \frac{n+1}{n}\theta > \theta$. 因此，$n\left(\theta - \hat{\theta}_2\right) \xrightarrow{L} \operatorname{Exp}(...)$ **不成立**！
(e) 但是，由 (c)，我们仍然可知，$n\left(2\theta - X_{(n)}\right) \xrightarrow{L} \operatorname{Exp}\left((2\theta)^{-1}\right)$.
而 $\theta - \hat{\theta}_2 = \frac{1}{2}[2\theta - \frac{n+1}{n}X_{(n)}] = \frac{n+1}{2n}[\frac{2n\theta}{n+1} - X_{(n)}] = \frac{n+1}{2n}[2\theta - X_{(n)} - \frac{2\theta}{n+1}] = \frac{n+1}{2n}[2\theta - X_{(n)}] - \frac{\theta}{n}$.
所以 $k_n(\theta - \hat{\theta}_2) = Z_n - \frac{2n\theta}{n+1}$. 其中 $k_n = \frac{2n^2}{n+1}$，$Z_n = n(2\theta - X_{(n)}) \xrightarrow{L} \operatorname{Exp}((2\theta)^{-1})$.
所以 $n \to \infty$ 时，$k_n(\theta - \hat{\theta}_2)$ 极限分布为“均值为 2θ 的指数分布再向左平移 2θ 的分布”，也即，均值为 0、方差为 $4\theta^2$ 的“指数分布”.

下面回到原题:

由于课件上给的 ARE 的定义是在正态分布下的，所以根据命题 (b)，ARE=0.（此时，$k_n = \sqrt{n}$.）
假如不在正态分布下考虑，还可以在 $k_n = \frac{2n^2}{n+1}$ 下考虑，显然，此时对于 $\hat{\theta}_1$，所以由 CLT 可知 $k_n(\hat{\theta}_1 - \theta)$ **依概率发散到无穷**，即 $\forall M \in \mathbb{R}^+$, $\mathbb{P}(k_n(\hat{\theta}_1 - \theta) \le M) \to 0$.
但是对于取值为无穷的随机变量，方差无法定义.
综上，我们还是认为 ARE=0.

注记 1：为什么“ARE=0”这个结论是合理的呢？这就要涉及到 ARE 背后的统计意义。总的来说，我们可以有如下判断标准：
ARE$\in (0,\infty) \iff$，两个估计量是“同阶”的，差不了太多.

ARE=0 $\Longleftrightarrow$ 前者比后者“弱”.
ARE=∞ $\Longleftrightarrow$ 前者比后者“强”.
这里“强”、“弱”、“同阶”的概念就类似初学微积分的时候遇见的“高阶”、“低阶”、“同阶”的“无穷大（小）量”.
那这里的强弱的含义又是什么呢? 其实就是**收敛速度**. 简单的说，我们可以这样定义一个估计量的收敛速度：若 $k_n(\hat{\theta}-\theta) \to D$，$D$ 为某个分布，则可以认为 $\hat{\theta}$ 的收敛速度为 $O(1/k_n)$.
而具有渐近相合正态性(CAN)这一类统计量, 收敛速度几乎都是 $O(1/\sqrt{n})$.
所以 ARE 的狭义定义就是在同阶的收敛速度下，比较哪个收敛得更快，越快的，方差自然越小（因为常数的方差为 0)，所以我们说它的效率更高.
虽然, 对于不同阶的收敛速度, ARE 的概念自然也适用, 不过此时 ARE 就只能取 0 或 ∞，就不是一个非常精确的量了。
比如，这里的 $\hat{\theta}_1$ 的收敛速度为 $O(1/\sqrt{n})$，但是 $\hat{\theta}_2$ 的收敛速度为 $O(1/n)$，显然后者更强，所以在实际的估计中，后者更加有用。