

Homework 7

姓名: 方嘉聪 学号: 2200017849

Problem 1. 给定未知参数 θ 的估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$, 证明

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(\text{Bias}(\hat{\theta})\right)^2 = \text{Var}(\hat{\theta}) + \left(\theta - \mathbb{E}(\hat{\theta})\right)^2.$$

Solution. 我们有

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}[\hat{\theta}^2] - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2 \\ &= \mathbb{E}[\hat{\theta}^2] - \mathbb{E}[\hat{\theta}]^2 + \mathbb{E}[\hat{\theta}]^2 - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2 \\ &= \text{Var}(\hat{\theta}) + \left(\mathbb{E}[\hat{\theta}] - \theta\right)^2 = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2. \end{aligned}$$

证毕.

Problem 2. 令总体 X 服从概率密度函数如下的连续分布, 其中 $\theta > 0$ 为未知参数,

$$f(x) = \begin{cases} \theta/x^2, & x \geq \theta, \\ 0, & x < \theta. \end{cases}$$

给定简单随机样本 X_1, X_2, \dots, X_n , 求 θ 的最大似然估计量.

Solution. 极大似然函数为

$$f(\theta) = \prod_{i=1}^n f(x_i) = \theta^n \left(\prod_{i=1}^n x_i^2 \right)^{-1} \cdot \mathbb{1}_{x_1, x_2, \dots, x_n \geq \theta}.$$

注意到当 $x_1, x_2, \dots, x_n \geq \theta$ 时, $f(\theta)$ 为关于 θ 的单调递增函数, 故最大似然估计量为

$$\hat{\theta}_{\text{MLE}} = \min\{x_1, x_2, \dots, x_n\}.$$

证毕.

Problem 3. 令总体 $X \sim \pi(\lambda)$, 即 X 服从参数为 λ 的泊松分布, λ 为未知参数. 给定简单随机样本 X_1, X_2, \dots, X_n , 本题中, 我们将考虑 $p = e^{-\lambda}$ 的两个不同的估计量:

- (1) 考虑 p 的矩法估计量 $\hat{p}_1 = e^{-\bar{X}}$. 这里, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 为样本均值. 判断 \hat{p}_1 是否为 $p = e^{-\lambda}$ 的最大似然估计 (简要说明原因, 无需严格证明), 判断 \hat{p}_1 是否为无偏估计量, 渐进无偏估计量, 一致估计量, 并计算 \hat{p}_1 的均方误差. 提示: 参考作业二第六题.
- (2) 令 $\hat{p}_2 = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}$. 这里

$$\mathbb{1}_{\{X_i=0\}} = \begin{cases} 1, & X_i = 0, \\ 0, & X_i > 0. \end{cases}$$

判断 \hat{p}_2 是否为无偏估计量, 渐进无偏估计量, 一致估计量, 并计算 \hat{p}_2 的均方误差.

Solution. (1) 极大似然函数为

$$L(p) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = p^n \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} = \frac{1}{\prod_{i=1}^n x_i!} \cdot p^n (-\ln p)^{\sum_{i=1}^n x_i}.$$

对 p 求导, 得到

$$L'(p) = np^{n-1}(-\ln p)^{\sum_{i=1}^n x_i} - p^n \frac{\sum_{i=1}^n x_i}{p} (-\ln p)^{\sum_{i=1}^n x_i - 1} \implies p_{\text{MLE}} = e^{-\bar{X}} = \hat{p}_1.$$

说明 \hat{p}_1 是 $p = e^{-\lambda}$ 的最大似然估计. 由作业二第六题, 对于 $X \sim \pi(\lambda)$ 与任意 $t \in \mathbb{R}$, 我们有

$$\mathbb{E}(e^{tX}) = e^{\lambda(e^t-1)} \implies \mathbb{E}(\hat{p}_1) = \left(\prod_{i=1}^n \mathbb{E}(e^{-X_i}) \right)^{1/n} = e^{\lambda(e^{-1}-1)} \neq e^{-\lambda} = p.$$

故 \hat{p}_1 不是无偏估计量和渐进无偏估计量. 由大数定律知, $\bar{X} \xrightarrow{P} \lambda$, 故由 **Lemma 1** 有,

$$\hat{p}_1 = e^{-\bar{X}} \xrightarrow{P} e^{-\lambda} = p.$$

故 \hat{p}_1 是一致估计量. 下面计算 \hat{p}_1 的均方误差,

$$\mathbb{E}(\hat{p}_1^2) = \mathbb{E}(e^{-2\bar{X}}) = \left(\prod_{i=1}^n \mathbb{E}(e^{-2X_i}) \right)^{1/n} = e^{\lambda(e^{-2}-1)}$$

那么 \hat{p}_1 的均方误差为

$$\text{MSE}(\hat{p}_1) = \mathbb{E}(\hat{p}_1^2) + p^2 - 2p\mathbb{E}(\hat{p}_1) = e^{\lambda(e^{-2}-1)} + e^{-2\lambda} - 2e^{-\lambda}e^{\lambda(e^{-1}-1)}.$$

(2) 注意到

$$\mathbb{E}(\hat{p}_2) = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(X_i=0) = e^{-\lambda} = p.$$

故 \hat{p}_2 是无偏估计量, 渐进无偏估计量. 那么 $\text{MSE}(\hat{p}_2) = \text{Var}(\hat{p}_2)$, 我们先来计算 $\mathbb{E}(\hat{p}_2^2)$,

$$\begin{aligned} \mathbb{E}(\hat{p}_2^2) &= \mathbb{E}\left[\frac{1}{n^2} \left(\sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}\right)^2\right] \\ &= \frac{1}{n^2} \mathbb{E}\left[\sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}^2 + \sum_{i \neq j} \mathbb{1}_{\{X_i=0 \wedge X_j=0\}}\right] \\ &= \frac{1}{n^2} \left[\sum_{i=1}^n \mathbb{P}(X_i=0) + \sum_{i \neq j} \mathbb{P}(X_i=0) \cdot \mathbb{P}(X_j=0) \right] \\ &= \frac{1}{n^2} \left[ne^{-\lambda} + (n^2 - n)e^{-2\lambda} \right] = e^{-2\lambda} + \frac{e^{-\lambda}}{n}(1 - e^{-\lambda}). \end{aligned}$$

那么有

$$\text{MSE}(\hat{p}_2) = \text{Var}(\hat{p}_2) = \mathbb{E}(\hat{p}_2^2) - \mathbb{E}(\hat{p}_2)^2 = \frac{e^{-\lambda}}{n}(1 - e^{-\lambda}) \xrightarrow{n \rightarrow \infty} 0.$$

故 \hat{p}_2 是一致估计量. 证毕.

Lemma 1. 设 $X_n \xrightarrow{P} X$, g 为 \mathbb{R}^1 上的连续函数, 则 $g(X_n) \xrightarrow{P} g(X)$.

证明. 对于任意的 $\delta > 0$, 存在 $M > 0$ 与 $N_1 \in \mathbb{N}^+$, 当 $n > N_1$ 时, 有

$$\mathbb{P}\left(|X| \geq \frac{M}{2}\right) \leq \frac{\delta}{4}, \quad \mathbb{P}\left(|X_n - X| \geq \frac{M}{2}\right) \leq \frac{\delta}{4}.$$

那么有

$$\begin{aligned} \mathbb{P}(|X_n| \geq M) &= \mathbb{P}\left(|X_n| \geq M, |X| < \frac{M}{2}\right) + \mathbb{P}\left(|X_n| \geq M, |X| \geq \frac{M}{2}\right) \\ &\leq \mathbb{P}\left(|X_n - X| \geq \frac{M}{2}\right) + \mathbb{P}\left(|X| \geq \frac{M}{2}\right) < \frac{\delta}{2}. \end{aligned}$$

由于 $g(x)$ 在 \mathbb{R}^1 上连续, 故 $g(x)$ 在 $[-M, M]$ 上一致连续, 即对于任意 $\varepsilon > 0$, 存在 $\eta > 0$, 对于 $x_1, x_2 \in [-M, M]$, 若 $|x_1 - x_2| < \eta$, 则 $|g(x_1) - g(x_2)| < \varepsilon$. 由于 $X_n \xrightarrow{P} X$, 故存在 $N_2 \in \mathbb{N}^+$, 当 $n > N_2$ 时, 有

$$\mathbb{P}(|X_n - X| \geq \eta) < \frac{\delta}{4}.$$

于是当 $n > \max\{N_1, N_2\}$ 时, 有

$$\begin{aligned} \mathbb{P}(|g(X_n) - g(X)| \geq \varepsilon) &\leq \mathbb{P}(|g(X_n) - g(X)| \geq \varepsilon, |X_n| < M, |X| < M) + \mathbb{P}(\{|X_n| \geq M\} \cup \{|X| \geq M\}) \\ &\leq \mathbb{P}(|X_n - X| \geq \eta) + \mathbb{P}\left(|X| \geq \frac{M}{2}\right) + \mathbb{P}(|X_n| \geq M) \\ &\leq \frac{\delta}{4} + \frac{\delta}{4} + \frac{\delta}{2} = \delta. \end{aligned}$$

由任意性可知 $g(X_n) \xrightarrow{P} g(X)$. 证毕. □

◁

Problem 4. 给定样本 $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y_1, Y_2, \dots, Y_m \sim \mathcal{N}(\mu_2, \sigma_2^2)$, 满足 $\{X_i\}_{i=1}^n, \{Y_j\}_{j=1}^m$ 相互独立.

- (1) 令 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i$. 给出 $\bar{X} - \bar{Y}$ 服从的分布.
- (2) 假定 σ_1^2 和 σ_2^2 均已知, 利用上一问中的结果构造枢轴量并给出 $\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的置信区间. 最终结果应该依赖于 $\Phi^{-1}(1 - \alpha/2)$, 其中 $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ 为标准正态分布的分布函数.
- (3) 同样假定 σ_1^2 和 σ_2^2 均已知, 利用 Chernoff Bound, 给出 $\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的置信区间. 最终结果不应依赖于标准正态分布的分布函数. ◀

Solution. (1) $\bar{X} \sim \mathcal{N}(\mu_1, \sigma_1^2/n), \bar{Y} \sim \mathcal{N}(\mu_2, \sigma_2^2/m)$, 且 \bar{X} 与 \bar{Y} 相互独立. 故

$$\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2/n + \sigma_2^2/m)$$

(2) 构造枢轴量

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \sim \mathcal{N}(0, 1).$$

记 $\sigma' = \sqrt{\sigma_1^2/n + \sigma_2^2/m}$, 那么有

$$\begin{aligned} \mathbb{P}\left[-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq Z \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right] &= 1 - \alpha. \\ \mathbb{P}\left[\bar{X} - \bar{Y} - \sigma' \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + \sigma' \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right] &= 1 - \alpha. \end{aligned}$$

故 $\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的置信区间为

$$\left[\bar{X} - \bar{Y} - \sigma' \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{X} - \bar{Y} + \sigma' \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right].$$

(3) 对于 $Z \sim \mathcal{N}(0, 1)$, 有 $M_Z(t) = \mathbb{E}(e^{tZ}) = e^{t^2/2}$. 由 Chernoff Bound, 对于任意 $t > 0$, 有

$$\mathbb{P}(|Z| \geq \varepsilon) \leq \inf_{t>0} \{e^{-t\varepsilon} M_Z(t)\} = 2e^{-\varepsilon^2/2}.$$

$2e^{-\varepsilon^2/2} = \alpha \implies \varepsilon = \sqrt{2\ln(2/\alpha)}$. 于是有

$$\begin{aligned} \mathbb{P}\left[-\varepsilon \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma'} \leq \varepsilon\right] &\geq 1 - \alpha. \\ \mathbb{P}\left[\bar{X} - \bar{Y} - \varepsilon \cdot \sigma' \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + \varepsilon \cdot \sigma'\right] &\geq 1 - \alpha. \end{aligned}$$

故 $\mu_1 - \mu_2$ 的置信水平为 $1 - \alpha$ 的置信区间为

$$\left[\bar{X} - \bar{Y} - \sqrt{2\ln(2/\alpha)} \cdot \sigma', \bar{X} - \bar{Y} + \sqrt{2\ln(2/\alpha)} \cdot \sigma'\right].$$

证毕. ◁

Problem 5. 在课上, 我们考虑了下述模型: 给定 n 台游戏机, 第 i 台游戏机的中奖概率为 $0 \leq p_i \leq 1$, 且 p_i 均为未知参数. 在第 t 轮中, 选择一台游戏机 $1 \leq i \leq n$, 并观测到结果 $X_t \sim B(1, p_i)$. 这里 X_1, X_2, \dots 相互独立.

在课上, 我们考虑了下述均匀采样策略: 对每台游戏机进行 N 次观测, 并返回样本均值最大的游戏机. 若取 $N = O(\ln n / \varepsilon^2)$, 则有 $\mathbb{P}(p_o \geq \max p_i - \varepsilon) \geq 2/3$, 这里 $1 \leq o \leq n$ 为策略返回的选择.

本题中, 我们考虑 $n = 2$ 的情况, 也即给定两台游戏机, 中奖概率分别为 p_1 和 p_2 , 且 p_1, p_2 为未知参数. 令 $\Delta = |p_1 - p_2|$.

(1) 若 Δ 为已知参数且 $\Delta > 0$, 证明采用均匀采样策略并令 $N = O(1/\Delta^2)$, 则有

$$\mathbb{P}(p_o = \max\{p_1, p_2\}) \geq \frac{2}{3}$$

这里 $o = 1$ 或 $o = 2$ 为策略返回的选择.

(2) (**Bonus 15%**) 若 Δ 为未知参数且 $\Delta > 0$, 设计策略, 使得以至少 $2/3$ 的概率, 下述事件同时成立:

- $p_o = \max\{p_1, p_2\}$, 这里 $o = 1$ 或 $o = 2$ 为策略返回的选择.

- 策略的总观测次数于 $1/\Delta$ 为多项式关系.



Solution. (1) 令 $\varepsilon = \Delta/2$, 由已证的结论, 对于均匀采样策略, 取 $N = O(\ln 2/\varepsilon^2) = O(1/\Delta^2)$, 有

$$\mathbb{P}(p_o \geq \max_{i=1,2} p_i - \varepsilon) = \mathbb{P}(p_o \geq \max_{i=1,2} p_i - \Delta/2) = \mathbb{P}(p_o = \max\{p_1, p_2\}) \geq \frac{2}{3}$$

证毕.

- (2) 考虑如下算法, 主要思想是每一轮将两个游戏机均运行若干次, 如果采样均值低于某个临界值, 则将该游戏机从备选游戏机中删去, 每一轮将临界值缩小 2 倍 (这个算法对于 n 台游戏机的情况也成立).

Algorithm 1 Successive Reject Algorithm

Require: 输入 $n = 2$, 待定参数 $\delta > 0$ 和每一轮运行的次数 T

Ensure: 输出 o

```

1:  $S_0 \leftarrow [n], t \leftarrow 0$ 
2: while  $|S_t| > 1$  do
3:    $t \leftarrow t + 1, \varepsilon_t \leftarrow 2^{-t}$ .
4:   对于每个  $i \in S_{t-1}$ , 运行  $T$  次.
5:    $S_t \leftarrow \{i \in S_{t-1} : \hat{p}_{i,t} \geq \max_{j \in S_{t-1}} \hat{p}_{j,t} - \varepsilon_t\}$ 
6: return  $S_t$ 

```

其中 $\hat{p}_{i,t}$ 为第 t 轮中第 i 台游戏机的样本均值. 注意到 $p_1, p_2 \in [0, 1]$, 在第 t 轮, 由 Hoeffding 不等式, 有

$$\mathbb{P}\left(|p_i - \hat{p}_{i,t}| > \frac{\varepsilon_t}{2}\right) \leq 2e^{-2T\varepsilon_t^2}$$

取 $T = \log(nt^2/\delta)/\varepsilon_t^2$, 那么有

$$\mathbb{P}\left(|p_i - \hat{p}_{i,t}| > \frac{\varepsilon_t}{2}\right) \leq \frac{2\delta}{nt^2}, \quad \forall i \in [n], \forall t \geq 1.$$

由 Union Bound, 有

$$\mathbb{P}\left(\forall i \in [n], |p_i - \hat{p}_{i,t}| \leq \frac{\varepsilon_t}{2}\right) \geq 1 - \frac{2\delta}{t^2}. \quad (1)$$

记事件 (不妨设 $p_1 > p_2$)

$$E_t = \{1 \in S_t \text{ and } \forall j, p_j < p_1 - 2\varepsilon_t \Rightarrow j \notin S_t\}.$$

对于 E_0 , $\varepsilon_0 = 1$, $S_0 = [2]$, $p_2 > p_1 - 2\varepsilon_0 = p_1 - 2$, 故 $\mathbb{P}(E_0) = 1$.

记事件 $E = \bigwedge_{t=1}^{\infty} E_t$. 注意到 E 表示 S_t 中剩下的游戏机是 1. 故我们来分析 $\mathbb{P}(E)$. 注意到¹

$$\mathbb{P}(E_t | E_{t-1}) \geq \mathbb{P}\left(\forall i \in S_{t-1}, |p_i - \hat{p}_{i,t}| \leq \frac{\varepsilon}{2} \middle| E_{t-1}\right) \geq 1 - \frac{2\delta}{t^2}. \quad (2)$$

¹第一个不等号需要分析事件 $E_t | E_{t-1}$, 证明在作业最后一页.

那么有

$$\mathbb{P}(E) = \mathbb{P}\left(\bigwedge_{t=1}^{\infty} E_t\right) = \prod_{t=1}^{\infty} \mathbb{P}(E_t|E_{t-1}) \geq \prod_{t=1}^{\infty} \left(1 - \frac{2\delta}{t^2}\right) \geq 1 - \sum_{i=1}^{\infty} \frac{2\delta}{i^2} \geq 1 - 2\delta \cdot \frac{\pi^2}{6}$$

令 $\delta = 1/\pi^2$, 那么有 $\mathbb{P}(E) \geq 2/3$. 下面来证明算法的总观测次数与 $1/\Delta$ 为多项式关系.

在 $n = 2$ 的情况下, 游戏机 1 的运行次数不会超过游戏机 2 的运行次数. 而对于游戏机 2, 由(1)知, 有至少 $2/3$ 的概率有

$$|p_2 - \hat{p}_{2,t}| \leq \frac{\varepsilon_t}{2}, \quad |p_1 - \hat{p}_{1,t}| \leq \frac{\varepsilon_t}{2} \implies \hat{p}_{1,t} - \hat{p}_{2,t} - \varepsilon_t \leq \Delta \leq \hat{p}_{1,t} - \hat{p}_{2,t} + \varepsilon_t.$$

故当 $t = \log_2(1/\Delta) + 2$ 时, 有

$$\hat{p}_{1,t} - \hat{p}_{2,t} \geq \Delta - \varepsilon_t > \varepsilon_t \implies \hat{p}_{2,t} < \hat{p}_{1,t} - \varepsilon_t \implies 2 \notin S_t.$$

故游戏机 2 的总运行次数不超过

$$\sum_{t=1}^{\lceil \log_2(1/\Delta) + 2 \rceil} \frac{\log(2t^2/\delta)}{2^{-2t}} < \sum_{t=1}^{\lceil \log_2(4/\Delta) \rceil} \frac{\log(2 \log^2(4/\Delta))}{2^{-2t}} = O\left(\frac{1}{\Delta^2} (\log \log^2(1/\Delta) + C)\right)$$

其中 C 为与 Δ 无关的常数. 故总的观测次数为 $\text{poly}(1/\Delta)$. 证毕.

注: 对于(2)中的第一个不等号, 我们来证明 $\{\forall i \in S_{t-1}, |p_i - \hat{p}_{i,t}| \leq \varepsilon/2 |E_{t-1}\} \subseteq E_t | E_{t-1}$.

- 由于 $\forall i \in S_{t-1}, \hat{p}_{i,t} \in p_i \pm \varepsilon/2$, 有

$$\begin{aligned} \max_{j \in S_{t-1}} \hat{p}_{j,t} - \varepsilon_t &\leq \max_{j \in S_{t-1}} p_j - \frac{\varepsilon}{2} < p_1 - \varepsilon \\ \implies \hat{p}_{1,t} &\geq p_1 - \frac{\varepsilon}{2} > \max_{j \in S_{t-1}} \hat{p}_{j,t} - \varepsilon_t \implies 1 \in S_t. \end{aligned}$$

- 假设存在 $j_0, p_{j_0} < p_1 - 2\varepsilon$, 且 $j_0 \in S_t$. 那么有

$$\hat{p}_{j_0,t} \geq \max_{j \in S_{t-1}} \hat{p}_{j,t} - \varepsilon_t, \quad \hat{p}_{j_0,t} \leq p_{j_0} + \frac{\varepsilon}{2} < p_1 - \frac{3\varepsilon}{2} \implies p_1 > \max_{j \in S_{t-1}} \hat{p}_{j,t} + \frac{\varepsilon}{2}$$

与 $\forall i \in S_{t-1}, |p_i - \hat{p}_{i,t}| \leq \varepsilon/2$ 矛盾. 故 $\forall j, p_j < p_1 - 2\varepsilon_t \Rightarrow j \notin S_t$.

综上, $\{\forall i \in S_{t-1}, |p_i - \hat{p}_{i,t}| \leq \varepsilon/2 |E_{t-1}\} \subseteq E_t | E_{t-1}$. 证毕. ◁