信息学中的概率统计：作业七

截止日期：2024 年 12 月 27 日（周五）下课前。**如无特殊情况，请不要提交电子版！**
注意：本次作业第五题为附加题，正确解决该题目可以得到额外 20% 的分数。

第一题

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

$$\mathrm{MSE}(\hat{\theta})=\mathrm{Var}(\hat{\theta})+(\mathrm{Bias}(\hat{\theta}))^2=\mathrm{Var}(\hat{\theta})+(\theta-E(\hat{\theta}))^2$$

答案：

$$\begin{aligned}\mathrm{MSE}(\hat{\theta})=&E((\hat{\theta}-\theta)^2)\\=&E((\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2)\\=&E((\hat{\theta}-E(\hat{\theta}))^2)+(E(\hat{\theta})-\theta)^2+2E((\hat{\theta}-E(\hat{\theta}))\cdot(E(\hat{\theta})-\theta))\\=&\mathrm{Var}(\hat{\theta})+(\mathrm{Bias}(\hat{\theta}))^2\end{aligned}$$

第二题

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，

$$f(x)=\begin{cases}\frac{\theta}{x^2} & x\geq\theta\\ 0 & x<\theta\end{cases}。$$

给定简单随机样本 $X_1,X_2,\ldots,X_n$，给出 θ 的最大似然估计量。
答案：

$$L(\theta)=\prod_{i=1}^{n}f(x_i)=\frac{\theta^n}{\prod_{i=1}^{n}x_i^2}\cdot 1_{\theta\leq\min\{x_1,x_2,\ldots,x_n\}}。$$

注意到当 $\theta\leq\min\{x_1,x_2,\ldots,x_n\}$ 满足时，$L(\theta)$ 关于 θ 为递增函数。因此，θ 的最大似然估计量 $\hat{\theta}=\min\{x_1,x_2,\ldots,x_n\}$。

第三题

令总体 $X\sim\pi(\lambda)$，也即参数为 λ 的泊松分布，λ 为未知参数。给定简单随机样本 $X_1,X_2,\ldots,X_n$，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 $\hat{p}_1=e^{-\overline{X}}$。这里，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 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 $\hat{p}_2 = \frac{1}{n}\sum_{i=1}^n 1_{X_i=0}$。这里
$$1_{X_i=0} = \begin{cases} 1 & X_i = 0 \\ 0 & X_i > 0 \end{cases}。$$
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。

答案:

(1) $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ 是 λ 的最大似然估计，根据最大似然估计的不变性，$e^{-\overline{X}}$ 同样是 $e^{-\lambda}$ 的最大似然估计。$E(\hat{p}_1) = \left(E(e^{-X_i/n})\right)^n$，根据作业二第六题，$E(e^{-X_i/n}) = e^{\lambda(e^{-1/n}-1)}$，因此 $E(\hat{p}_1) = e^{n\lambda(e^{-1/n}-1)}$。注意到 $E(\hat{p}_1) = e^{n\lambda(e^{-1/n}-1)} \neq e^{-\lambda}$，而
$$\lim_{n\to\infty} e^{n\lambda(e^{-1/n}-1)} = e^{-\lambda},$$
因此 $\hat{p}_1$ 不是无偏估计量，但 $\hat{p}_1$ 是渐进无偏估计量。
$E(\hat{p}_1^2) = E(e^{-2\overline{X}}) = \left(E(e^{-2X_i/n})\right)^n$。根据作业二第六题，$E(e^{-2X_i/n}) = e^{\lambda(e^{-2/n}-1)}$，因此 $E(\hat{p}_1^2) = e^{n\lambda(e^{-2/n}-1)}$。

$$\mathrm{MSE}(\hat{p}_1) = E((\hat{p}_1 - e^{-\lambda})^2) = E(\hat{p}_1^2) - 2e^{-\lambda}E(\hat{p}_1) + e^{-2\lambda} = e^{n\lambda(e^{-2/n}-1)} - 2e^{n\lambda(e^{-1/n}-1)-\lambda} + e^{-2\lambda}。$$

注意到
$$\lim_{n\to\infty} \mathrm{MSE}(\hat{p}_1) = e^{-2\lambda} - 2e^{-2\lambda} + e^{-2\lambda} = 0,$$
因此 $\hat{p}_1$ 为一致估计量。

(2) 注意到 $P(X_i = 0) = e^{-\lambda}$，因此 $E(\hat{p}_2) = e^{-\lambda}$，也即 $\hat{p}_2$ 是无偏估计量。因此，
$$\mathrm{MSE}(\hat{p}_1) = \mathrm{Var}(\hat{p}_1) = \frac{1}{n}\cdot e^{-\lambda}\cdot(1 - e^{-\lambda})。$$
由于 $\lim_{n\to\infty}\mathrm{MSE}(\hat{p}_2) = 0$，因此 $\hat{p}_2$ 为一致估计量。

第四题

给定样本 $X_1, X_2, \ldots, X_n \sim N(\mu_1, \sigma_1^2)$，$Y_1, Y_2, \ldots, Y_m \sim N(\mu_2, \sigma_2^2)$，满足 $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$ 相互独立。

(1) 令 $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$，$\overline{Y} = \frac{1}{m}\sum_{i=1}^m Y_i$。给出 $X - Y$ 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 $1-\alpha$ 置信区间。最终结果应依赖于 $\Phi^{-1}(1-\alpha/2)$，其中 $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx$ 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 $1-\alpha$ 置信区间。最终结果不应依赖于标准正态分布的分布函数。

答案:

(1) $\overline{X} \sim N(\mu_1, \sigma_1^2/n)$，$\overline{Y} \sim N(\mu_2, \sigma_2^2/m)$。由独立性，
$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \sigma_1^2/n + \sigma_2^2/m)。$$

(2) 令枢轴量

$$G = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}},$$

则有 $G \sim N(0,1)$。因此，

$$P(-\Phi^{-1}(1-\alpha/2) \le G \le \Phi^{-1}(1-\alpha/2)) = 1-\alpha,$$

也即

$$P(\overline{X}-\overline{Y}-\Phi^{-1}(1-\alpha/2)\cdot\sqrt{\sigma_1^2/n+\sigma_2^2/m} \le \mu_1-\mu_2 \le \overline{X}-\overline{Y}+\Phi^{-1}(1-\alpha/2)\cdot\sqrt{\sigma_1^2/n+\sigma_2^2/m}) = 1-\alpha。$$

(3) 注意到

$$\overline{X} - \overline{Y} - (\mu_1 - \mu_2) \sim N(0, \sigma_1^2/n + \sigma_2^2/m),$$

由 Chernoff bound，有

$$P(\left|\overline{X} - \overline{Y} - (\mu_1 - \mu_2)\right| \ge k\sqrt{\sigma_1^2/n + \sigma_2^2/m}) \le 2e^{-k^2/2}。$$

令 $k = \sqrt{2\ln(2/\alpha)}$，有

$$P(\left|\overline{X} - \overline{Y} - (\mu_1 - \mu_2)\right| \le \sqrt{2\ln(2/\alpha)}\sqrt{\sigma_1^2/n + \sigma_2^2/m}) \ge 1-\alpha,$$

也即

$$P(\overline{X} - \overline{Y} - \sqrt{2\ln(2/\alpha)}\sqrt{\sigma_1^2/n + \sigma_2^2/m} \le \mu_1 - \mu_2 \le \overline{X} - \overline{Y} + \sqrt{2\ln(2/\alpha)}\sqrt{\sigma_1^2/n + \sigma_2^2/m}) \ge 1-\alpha。$$

第五题

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 p_i，且 p_i 均为未知参数。在第 t 轮中，选择一台游戏机 $1 \le i \le n$，并观测到结果 $X_t \sim B(1, p_i)$。这里 $X_1, X_2, \ldots$ 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 $N = O(\ln n/\epsilon^2))$，则有 $P(p_o \ge \max p_i - \epsilon) \ge 2/3$，这里 $1 \le o \le n$ 为算法返回的选择。

本题中，我们考虑 $n = 2$ 的情况，也即考虑给定两台游戏机，中奖概率分别为 p_1 和 p_2。令 $\Delta = |p_1 - p_2|$。

(1) 若 Δ 为已知参数且 $\Delta > 0$，证明采取均匀采样策略并令 $N = O(1/\Delta^2)$，则有 $P(p_o = \max\{p_1, p_2\}) \ge 2/3$，这里 $o = 1$ 或 $o = 2$ 为算法返回的选择。

(2) 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 $o = 2$ 为算法返回的选择；
- 策略的总观测次数与 $1/\Delta$ 为多项式关系。

答案：

(1) 根据课上分析，令 $\epsilon = \Delta/2$，$n = 2$，则有 $N = O(1/\Delta^2)$，且 $P(p_o \ge \max p_i - \Delta/2) \ge 2/3$。因为 $|p_1 - p_2| = \Delta$，$p_o \ge \max p_i - \Delta/2$ 等价于 $p_o = \max\{p_1, p_2\}$。

(2) 令 $\ell = 1, 2, \ldots$。在第 ℓ 个阶段中，令 $\epsilon_\ell = 2^{-\ell}$，且 $\delta_\ell = 1/6 \cdot 2^{-\ell}$。在第 ℓ 个阶段中，采取均匀采样策略并令 $N = O(1/\epsilon_\ell^2 \cdot \log(1/\delta_\ell))$。令 $\hat{p}_1^\ell$ 和 $\hat{p}_2^\ell$ 为两台游戏机在第 ℓ 个阶段中观测的样本均值。

- 若 $\hat{p}_1^\ell - \epsilon_\ell \ge \hat{p}_2^\ell + \epsilon_\ell$ 则返回第一台游戏机；
- 若 $\hat{p}_2^\ell - \epsilon_\ell \ge \hat{p}_1^\ell + \epsilon_\ell$ 则返回第二台游戏机；

- 否则进入下一个阶段 $\ell+1$。

令事件

$$E=\bigcup_{\ell}\bigcup_{i\in\{1,2\}} p_i \in [\hat{p}_i^{\ell}-\epsilon_\ell, \hat{p}_i^{\ell}+\epsilon_\ell])。$$

根据课上分析，

$$P(p_1 \in [\hat{p}_1^{\ell}-\epsilon_\ell, \hat{p}_1^{\ell}+\epsilon_\ell]) \geq 1-\delta_\ell = 1-1/6\cdot 2^{-\ell},$$
$$P(p_2 \in [\hat{p}_2^{\ell}-\epsilon_\ell, \hat{p}_2^{\ell}+\epsilon_\ell]) \geq 1-\delta_\ell = 1-1/6\cdot 2^{-\ell}。$$

由 Union bound，有

$$P(E) \geq 1-\sum_{\ell=1}^{\infty} 1/6\cdot 2^{-\ell}\cdot 2 = 2/3。$$

下面我们证明，若 E 发生，则有 $p_o=\max\{p_1,p_2\}$，且总观测次数与 $1/\Delta$ 为多项式关系。

首先证明 $p_o=\max\{p_1,p_2\}$。注意到，若事件 E 发生，则当算法结束时，若 $o=1$，则

$$p_1 \geq \hat{p}_1^{\ell}-\epsilon_\ell \geq \hat{p}_2^{\ell}+\epsilon_\ell \geq p_2。$$

若 $o=2$，则

$$p_2 \geq \hat{p}_2^{\ell}-\epsilon_\ell \geq \hat{p}_1^{\ell}+\epsilon_\ell \geq p_1。$$

因此 $P(p_o=\max\{p_1,p_2\}) \geq 2/3$。

下面分析总观测次数。当事件 E 发生时，如果 $\epsilon_\ell \leq \Delta/4$ 时，若 $p_1 \geq p_2$，则

$$\hat{p}_1^{\ell}-\epsilon_\ell \geq p_1-2\epsilon_\ell \geq p_2+2\epsilon_\ell \geq \hat{p}_2^{\ell}+\epsilon_\ell;$$

若 $p_2 \geq p_1$，则

$$\hat{p}_2^{\ell}-\epsilon_\ell \geq p_2-2\epsilon_\ell \geq p_1+2\epsilon_\ell \geq \hat{p}_1^{\ell}+\epsilon_\ell。$$

因此当事件 E 发生时，若 $\ell \geq \log_2(1/\Delta)+10$，则策略一定停止。因此，总观测次数为

$$\sum_{\ell=1}^{\log_2(1/\Delta)+10} 1/\epsilon_\ell^2\cdot\log(1/\delta_\ell) = \sum_{\ell=1}^{\log_2(1/\Delta)+10} O(4^{\ell}\cdot\ell) = O(1/\Delta^2\cdot\log(1/\Delta))。$$

思考：如何得到更好的结果？比如取 $\delta_l=\Theta(1/\ell^2)$?