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：本次作业第五题第二问为附加题，正确解决该问可以得到额外 15% 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

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

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。

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。

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 $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ 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 $1-\alpha$ 置信区间。最终结果不应依赖于标准正态分布的分布函数。

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为 $0 \le p_i \le 1$，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 和 p_2 均为未知参数。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 $P(p_o = \max\{p_1, p_2\}) \ge 2/3$, 这里 $o = 1$ 或 $o = 2$ 为策略返回的选择。

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

本问为附加问，正确解决该问可以得到额外 15% 的分数。