贝叶斯统计学基础作业 1

毛沛炫　3220102692

2025 年 3 月 16 日

1. 假设在一个箱子中装有 10 只灯泡，其中 3 只是次品。现在从其中取两次灯泡，每次随机取一个，一种情况采取无放回抽样，一种情况采取放回抽样，定义随机变量如下

- $X=0$，如果如果第一次取出的是正品
- $X=1$，如果第一次取出的是次品
- $Y=0$，如果第二次取出的是正品
- $Y=1$，如果第二次取出的是次品

求无放回抽样和放回抽样条件下的各个联合概率 (4 分)

在无放回抽样条件下，随机变量 X 和 Y 是否独立 (2 分)

在放回抽样条件下，随机变量 X 和 Y 是否独立 (2 分)

解答：

1. 求无放回抽样和放回抽样条件下的各个联合概率

   (a) 有放回

   $P(X=0,Y=0)=0.3\times0.3=0.09,$

   $P(X=0,Y=1)=0.3\times0.7=0.21,$

   $P(X=1,Y=0)=0.7\times0.3=0.21,$

   $P(X=1,Y=1)=0.7\times0.7=0.49$

   (b) 无放回

   $P(X=0,Y=0)=\frac{3}{10}\times\frac{2}{9}=\frac{2}{30},$

   $P(X=0,Y=1)=\frac{3}{10}\times\frac{7}{9}=\frac{7}{30},$

   $P(X=1,Y=0)=\frac{7}{10}\times\frac{3}{9}=\frac{7}{30},$

   $P(X=1,Y=1)=\frac{7}{10}\times\frac{6}{9}=\frac{14}{30}$

2. 在无放回抽样条件下，随机变量 X 和 Y 是否独立

   $P(X=0)=3/10,\ P(Y=0)=3/10,$

   $P(X=0,Y=0)=2/30,$

   $P(X=0,Y=0)\neq P(X=0)\times P(Y=0)$

   所以随机变量 X 和 Y **不独立**

3. 在放回抽样条件下，随机变量 X 和 Y 是否独立

两次伯努利实验，随机变量 X 和 Y **独立**

2. 通过伯努利分布的分布律求其数学期望和方差 (4 分)

解答：

伯努利分布的概率质量函数 (probability mass function)：

$$f_X(x) = p^x(1-p)^{1-x} = \begin{cases} p & \text{if } x = 1, \\ q & \text{if } x = 0. \end{cases}$$

其期望为：

$$\mathrm{E}[X] = \sum_{i=0}^{1} x_i f_X(x) = 0 + p = p$$

方差为：

$$\mathrm{Var}[X] = \sum_{i=0}^{1} (x_i - \mathrm{E}[X])^2 f_X(x) = (0-p)^2(1-p) + (1-p)^2 p = p(1-p) = pq$$

3. 通过在 a 到 b 之间的均匀分布的概率密度函数求其数学期望和方差 (4 分)

解答： 均匀分布的概率密度函数 (probability density function)：

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a,b], \\ 0, & \text{if } x \notin [a,b]. \end{cases}$$

其期望为：

$$\mathrm{E}[X] = \int_a^b x f_X(x) dx = \int_a^b \frac{x dx}{a-b} = \frac{(a-b)^2}{2(a-b)} = \frac{a+b}{2}$$

方差为：

$$\begin{aligned} \mathrm{Var}[X] &= \int_a^b (x - \mathrm{E}[X])^2 f_X(x) dx = \frac{1}{b-a} \int_a^b (x - \frac{a+b}{2})^2 dx \\ &= \frac{1}{b-a} \left(\frac{x^3}{3} - \frac{(a+b)x^2}{2} + \frac{(a+b)^2 x}{4} \right) \Bigg|_a^b = \frac{(b-a)^2}{12} \end{aligned}$$

4. 证明，对于一元线性回归分析，回归系数的最小二乘估计值和极大似然估计值是一致的。(9 分)

解答:

一元线性回归模型为:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

其中假设随机误差 $\varepsilon_1, \varepsilon_2, \varepsilon_3 \dots$ 独立同分布，$\varepsilon \sim N(0, \sigma^2)$

1. 一元线性回归

   一元线性回归的目标是使模型的误差和实际观测数据的误差最小，即求

$$\min \sum_{i=1}^{n} \varepsilon_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

   因此定义损失函数 L 为:

$$L(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

   则求出 L 的驻点即可得到满足条件（模型的误差和实际观测数据的误差最小，在此题下不考虑驻点是误差最大的情况）的 β_0 和 β_1

$$\begin{cases} \frac{\partial L}{\partial \beta_1} = 0 \\ \frac{\partial L}{\partial \beta_0} = 0 \end{cases} \Rightarrow \begin{cases} \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{cases}$$

2. 极大似然估计

   极大似然估计是要让 $\Pi_{i=1}^{n} P(Y = y_i)$ 的值最大，而根据模型，$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$，

$$P(Y = y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{[y_i - (\beta_0 + \beta_1 x_i)]^2}{2\sigma^2} \right\}$$

   因此记极大似然函数 L 为

$$\begin{aligned} L(\beta_0, \beta_1 | y_1, y_2, y_3 \dots) &= \Pi_{i=1}^{n} P(Y = y_i) \\ &= (\sqrt{2\pi}\sigma)^{-n} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \right\} \end{aligned}$$

求 L 的最大值，即求 $[y_i - (\beta_0 + \beta_1 x_i)]^2$ 的最小值，则和一元线性回归相同，

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$