Homework1

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

解:

- 求无放回抽样和放回抽样条件下的各个联合概率(4 分)
  对于无放回抽样有:
  $P(X=0,Y=0)=P(\text{第一次取正品})\times P(\text{第二次取正品})=\frac{7}{10}\times\frac{6}{9}=\frac{14}{30}$
  $P(X=0,Y=1)=P(\text{第一次取正品})\times P(\text{第二次取次品})=\frac{7}{10}\times\frac{3}{9}=\frac{7}{30}$
  $P(X=1,Y=0)=P(\text{第一次取次品})\times P(\text{第二次取正品})=\frac{3}{10}\times\frac{7}{9}=\frac{7}{30}$
  $P(X=1,Y=1)=P(\text{第一次取次品})\times P(\text{第二次取次品})=\frac{3}{10}\times\frac{2}{9}=\frac{1}{30}$
  对于放回抽样有:
  $P(X=0,Y=0)=P(\text{第一次取正品})\times P(\text{第二次取正品})=\frac{7}{10}\times\frac{7}{10}=\frac{49}{100}$
  $P(X=0,Y=1)=P(\text{第一次取正品})\times P(\text{第二次取次品})=\frac{7}{10}\times\frac{3}{10}=\frac{21}{100}$
  $P(X=1,Y=0)=P(\text{第一次取次品})\times P(\text{第二次取正品})=\frac{3}{10}\times\frac{7}{10}=\frac{21}{100}$
  $P(X=1,Y=1)=P(\text{第一次取次品})\times P(\text{第二次取次品})=\frac{3}{10}\times\frac{3}{10}=\frac{9}{100}$
- 在无放回抽样条件下，随机变量 X 和 Y 是否独立(2 分)
  如果 $P(X=x,Y=y)=P(X=x)\times P(Y=y)$ 对于所有的 x 和 y 成立，则 X 和 Y 是独立的。在这种情况下，我们可以看到:

  $$P(X=0,Y=0)=\frac{14}{30}\neq\frac{7}{10}\times\frac{7}{10}=P(X=0)\times P(Y=0)$$

  因此，在无放回抽样条件下，随机变量 X 和 Y 不是独立的。

- 在放回抽样条件下，随机变量 X 和 Y 是否独立(2 分)

  $$P(X=0,Y=0)=\frac{49}{100}=\frac{7}{10}\times\frac{7}{10}=P(X=0)\times P(Y=0)$$

  同样地，对于其他的X 和Y 组合，独立性也成立。因此，在放回抽样条件下，随机变量X和Y 是独立的。

2

通过伯努利分布的分布律求其数学期望和方差(4 分)

解:
对于随机变量X，记成功的概率为$p(0{\le}p{\le}1)$，失败的概率为$q=1-p$，其分布律为:

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

数学期望为:

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

方差为:

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

$$=(0-p)^2(1-p)+(1-p)^2p=p(1-p)=pq$$

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:
概率密度函数:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$$

数学期望：

$$E(X) = \int_a^b x \cdot f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{b+a}{2}$$

方差：

$$\mathrm{Var}(X) = E((X - E(X))^2)$$

$$= \int_a^b (x - E(X))^2 \cdot f(x)\,dx = \int_a^b (x - \frac{b+a}{2})^2 \cdot \frac{1}{b-a}\,dx = \frac{(b-a)^2}{12}$$

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：

在一元线性回归中：

$$y = \beta_0 + \beta_1 x + \varepsilon$$

其中β_0、β_1 代表回归系数，后文中 $\hat{\beta}_0$、 $\hat{\beta}_1$ 代表的是回归系数的估计值；ε 代表误差项
关于误差项ε,有$\varepsilon \sim N(0, \sigma^2)$，对于$y|x$,有$y|x \sim N(\beta_0 + \beta_1 x, \sigma^2)$
对最小二乘估计（LSE）：

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

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

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 称为 y_i 的回归拟合值，$e_i = y_i - \hat{y}_i$ 为 y_i 的残差。
要使得上式最小，那么上式对 β_0、β_1 的偏导数均为0，即：

$$\frac{\partial Q}{\partial \beta_0}|_{\beta_0=\hat{\beta}_0} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$
$$\frac{\partial Q}{\partial \beta_1}|_{\beta_1=\hat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0$$

经整理后，得到正规一元二次方程组如下：

$$n\hat{\beta}_0 + (\sum_{i=1}^{n} x_i)\hat{\beta}_1 = \sum_{i=1}^{n} y_i$$
$$(\sum_{i=1}^{n} x_i)\hat{\beta}_0 + (\sum_{i=1}^{n} x_i^2)\hat{\beta}_1 = \sum_{i=1}^{n} x_i y_i$$

解得：

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

对于极大似然估计（MLE）：
此时，因变量的分布为一个联合密度函数，在假设 $\varepsilon \sim N(0, \sigma^2)$ 时，y_i 服从如下正态分布i

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

$y_1, y_2, \ldots, y_n$ 的似然函数为：

$$L(\rho_0, \rho_1, o^{-1}) = \prod_{i=1} f_i(g_i)$$

$$= (2\pi\sigma^2)^{-n/2} exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \beta_0 + \beta_1 x_i]^2\}$$

对数似然函数如下：

$$ln(L) = -\frac{n}{2}ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \beta_0 + \beta_1 x_i]^2$$

再对对数函数求偏导，即回到了最小二乘法（LSE）的求解步骤
综上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