# 1

## Problem

1. 假设连续型随机变量 $X, Y$ 具有联合密度函数

$$p(x,y)=\begin{cases}(x+y)/3, & 0<x<1, 0<y<2\\ 0, & \text{else}\end{cases}$$

（1）计算 $X, Y$ 各自的数学期望 $E[X], E[Y]$；

（2）计算 $X$ 关于 $Y$ 的条件数学期望 $E[X \mid Y]$。

（3）计算 $X$ 关于 $Y$ 的条件方差 $\mathrm{Var}[X \mid Y]$。

（4）计算 $\mathrm{Var}(3X-2Y+6)$。

## Solution (1)

首先求 $X$ 的边际密度函数：

$$p_X(x)=\int_0^2 \frac{x+y}{3}dy=\frac{1}{3}\left[xy+\frac{y^2}{2}\right]_0^2=\frac{1}{3}(2x+2)=\frac{2(x+1)}{3}, \quad 0<x<1$$

则：

$$E[X]=\int_0^1 x\cdot\frac{2(x+1)}{3}dx=\frac{2}{3}\int_0^1\left(x^2+x\right)dx=\frac{2}{3}\left[\frac{x^3}{3}+\frac{x^2}{2}\right]_0^1=\frac{2}{3}\cdot\frac{5}{6}=\frac{5}{9}$$

再求 $Y$ 的边际密度函数：

$$p_Y(y)=\int_0^1 \frac{x+y}{3}dx=\frac{1}{3}\left[\frac{x^2}{2}+xy\right]_0^1=\frac{1}{3}\left(\frac{1}{2}+y\right)=\frac{1+2y}{6}, \quad 0<y<2$$

则：

$$E[Y]=\int_0^2 y\cdot\frac{1+2y}{6}dy=\frac{1}{6}\int_0^2\left(y+2y^2\right)dy=\frac{1}{6}\left[\frac{y^2}{2}+\frac{2y^3}{3}\right]_0^2=\frac{1}{6}\cdot\frac{22}{3}=\frac{11}{9}$$

因此：

$$E[X]=\frac{5}{9}, \quad E[Y]=\frac{11}{9}$$

# Solution (2)

条件密度函数为：

$$p_{X|Y}(x \mid y)=\frac{p(x,y)}{p_Y(y)}=\frac{\frac{x+y}{3}}{\frac{1+2y}{6}}=\frac{2(x+y)}{1+2y}, \quad 0<x<1$$

则条件期望为：

$$E[X \mid Y=y]=\int_0^1 x \cdot \frac{2(x+y)}{1+2y} dx=\frac{2}{1+2y} \int_0^1\left(x^2+xy\right) dx=\frac{2}{1+2y}\left[\frac{x^3}{3}+\frac{x^2 y}{2}\right]_0^1= \frac{2}{1+2y}\left(\frac{1}{3}+\frac{y}{2}\right)$$

因此:

$$E[X \mid Y]=\frac{2+3Y}{3(1+2Y)}$$

# Solution (3)

首先求 $E\left[X^2 \mid Y\right]$：

$$E\left[X^2 \mid Y=y\right]=\int_0^1 x^2 \cdot \frac{2(x+y)}{1+2y} dx=\frac{2}{1+2y} \int_0^1\left(x^3+x^2 y\right) dx=\frac{2}{1+2y}\left[\frac{x^4}{4}+\frac{x^3 y}{3}\right]_0^1= \frac{2}{1+2y}\left(\frac{1}{4}+\frac{y}{3}\right)$$

根据条件方差公式有：

$$\operatorname{Var}[X \mid Y=y]=E\left[X^2 \mid Y=y\right]-(E[X \mid Y=y])^2=\frac{3+4y}{6(1+2y)}-\frac{(2+3y)^2}{9(1+2y)^2}$$

因此：

$$\operatorname{Var}[X \mid Y]=\frac{1+6Y+6Y^2}{18(1+2Y)^2}$$

# Solution (4)

先求 $\operatorname{Var}(X)$ 和 $\operatorname{Var}(Y)$ 。
已知 $E[X]=\frac{5}{9}$ ，求 $E\left[X^2\right]$ ：

$$E\left[X^2\right]=\int_0^1 x^2 \cdot \frac{2(x+1)}{3} dx=\frac{2}{3} \int_0^1\left(x^3+x^2\right) dx=\frac{2}{3}\left[\frac{x^4}{4}+\frac{x^3}{3}\right]_0^1=\frac{2}{3} \cdot \frac{7}{12}=\frac{7}{18}$$

所以：

$$\operatorname{Var}(X)=E\left[X^2\right]-(E[X])^2=\frac{7}{18}-\left(\frac{5}{9}\right)^2=\frac{7}{18}-\frac{25}{81}=\frac{63}{162}-\frac{50}{162}=\frac{13}{162}$$

已知 $E[Y]=\frac{11}{9}$，求 $E\left[Y^2\right]$：

$$E\left[Y^2\right]=\int_0^2 y^2\cdot\frac{1+2y}{6}dy=\frac{1}{6}\int_0^2\left(y^2+2y^3\right)dy=\frac{1}{6}\left[\frac{y^3}{3}+\frac{y^4}{2}\right]_0^2=\frac{1}{6}\cdot\frac{32}{3}=\frac{16}{9}$$

所以：

$$\operatorname{Var}(Y)=E\left[Y^2\right]-(E[Y])^2=\frac{16}{9}-\left(\frac{11}{9}\right)^2=\frac{16}{9}-\frac{121}{81}=\frac{144}{81}-\frac{121}{81}=\frac{23}{81}$$

再求 $\operatorname{Cov}(X,Y)=E[XY]-E[X]E[Y]$：

$$E[XY]=\iint xyp(x,y)dxdy=\int_0^2\int_0^1 xy\cdot\frac{x+y}{3}dxdy=\frac{1}{3}\int_0^2\int_0^1\left(x^2y+xy^2\right)dxdy$$

先对 $x$ 积分：

$$\int_0^1\left(x^2y+xy^2\right)dx=y\left[\frac{x^3}{3}+\frac{x^2y}{2}\right]_0^1=y\left(\frac{1}{3}+\frac{y}{2}\right)=\frac{y}{3}+\frac{y^2}{2}$$

再对 $y$ 积分：

$$E[XY]=\frac{1}{3}\int_0^2\left(\frac{y}{3}+\frac{y^2}{2}\right)dy=\frac{1}{3}\left[\frac{y^2}{6}+\frac{y^3}{6}\right]_0^2=\frac{1}{3}\cdot\frac{12}{6}=\frac{2}{3}$$

则：

$$\operatorname{Cov}(X,Y)=\frac{2}{3}-\frac{5}{9}\cdot\frac{11}{9}=\frac{2}{3}-\frac{55}{81}=\frac{54}{81}-\frac{55}{81}=-\frac{1}{81}$$

代入方差公式：

$$\operatorname{Var}(3X-2Y+6)=9\cdot\frac{13}{162}+4\cdot\frac{23}{81}-12\cdot\left(-\frac{1}{81}\right)=\frac{117}{162}+\frac{92}{81}+\frac{12}{81}$$

最终有：

$$\operatorname{Var}(3X-2Y+6)=\frac{117}{162}+\frac{184}{162}+\frac{24}{162}=\frac{325}{162}$$

# 2

# Problem

考虑 $\bar{X}$ 为来自 $\mu=8,\sigma=18$ 的总体的样本的均值。运用概率论知识回答下述问题。
（1）当样本容量为 $n=81$ 时，计算 $P(\bar{X}>6)$；

（2）当样本容量为 $n=64$ 时，计算 $P(3<\bar{X}<9)$ 。

# Solution (1)

标准误差为：

$$\sigma_{\bar{X}}=\frac{\sigma}{\sqrt{n}}=\frac{18}{\sqrt{81}}=\frac{18}{9}=2$$

因此，$\bar{X}\sim N\left(8,2^2\right)$ 。标准化：

$$Z=\frac{\bar{X}-\mu}{\sigma_{\bar{X}}}=\frac{\bar{X}-8}{2}$$

则：

$$P(\bar{X}>6)=P\left(Z>\frac{6-8}{2}\right)=P(Z>-1)$$

由于标准正态分布的对称性，$P(Z>-1)=P(Z<1)=0.8413$ 。因此：

$$P(\bar{X}>6)=0.8413$$

# Solution (2)

标准误差为：

$$\sigma_{\bar{X}}=\frac{\sigma}{\sqrt{n}}=\frac{18}{\sqrt{64}}=\frac{18}{8}=2.25$$

因此，$\bar{X}\sim N\left(8,2.25^2\right)$ 。标准化：

$$Z_1=\frac{3-8}{2.25}=\frac{-5}{2.25}\approx-2.2222,\quad Z_2=\frac{9-8}{2.25}=\frac{1}{2.25}\approx 0.4444$$

则：

$$P(3<\bar{X}<9)=P(-2.2222<Z<0.4444)$$

查标准正态分布表：

- $P(Z<0.4444)\approx 0.6715$
- $P(Z<-2.2222)\approx 0.0131$

所以：

$$P(-2.2222<Z<0.4444)=P(Z<0.4444)-P(Z<-2.2222)\approx 0.6715-0.0131=0.65$$

# 3

## Problem

设 $Y, X, Z$ 为随机变量，求解如下优化问题的显式解：
$\min_{\alpha,\beta,\gamma} E\left[(Y-\alpha-\beta X-\gamma Z)^2\right].$

## Solution

对 $\alpha$ 求偏导：

$\frac{\partial L}{\partial \alpha}=-2E[Y-\alpha-\beta X-\gamma Z]=0$

整理得：

$E[Y]-\alpha-\beta E[X]-\gamma E[Z]=0 \quad \Rightarrow \quad \alpha=E[Y]-\beta E[X]-\gamma E[Z]$ (1)

对 $\beta$ 求偏导：

$\frac{\partial L}{\partial \beta}=-2E[X(Y-\alpha-\beta X-\gamma Z)]=0$

即：

$E[XY]-\alpha E[X]-\beta E\left[X^2\right]-\gamma E[XZ]=0$ (2)

对 $\gamma$ 求偏导：

$\frac{\partial L}{\partial \gamma}=-2E[Z(Y-\alpha-\beta X-\gamma Z)]=0$

即：

$E[ZY]-\alpha E[Z]-\beta E[ZX]-\gamma E\left[Z^2\right]=0$ (3)

解线性方程组 (1), (2), (3) 可得：

$$\alpha=E[Y]-\beta E[X]-\gamma E[Z]$$
$$\beta=\frac{\operatorname{Cov}(X,Y)\operatorname{Var}(Z)-\operatorname{Cov}(X,Z)\operatorname{Cov}(Z,Y)}{\operatorname{Var}(X)\operatorname{Var}(Z)-\operatorname{Cov}(X,Z)^2}$$
$$\gamma=\frac{\operatorname{Cov}(Z,Y)\operatorname{Var}(X)-\operatorname{Cov}(X,Z)\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)\operatorname{Var}(Z)-\operatorname{Cov}(X,Z)^2}$$

# 4

## Problem

证明在考虑一元线性回归模型中
$R^2 = r_{XY}^2$

## Solution

考虑一元线性回归模型：

$$Y = \alpha + \beta X + \varepsilon$$

我们首先求解 $R^2$。

回归系数的最小二乘估计为：

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$
$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

由于 $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i = \bar{Y} + \hat{\beta}\left(X_i - \bar{X}\right)$，我们有：

$$\hat{Y}_i - \bar{Y} = \hat{\beta}\left(X_i - \bar{X}\right)$$

因此：

$$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 = \hat{\beta}^2 \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

代入 $\hat{\beta}$ 的表达式：

$$SSR = \left[\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right]^2 \cdot \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

因此有：

$$R^2 = \frac{SSR}{SST} = \frac{\frac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} = \frac{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \cdot \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}$$

与此同时，有：

$$r_{XY}^2 = \left[\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}}\right]^2 = \frac{\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)\right]^2}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}$$

比较 $R^2$ 和 $r_{XY}^2$ 的表达式，我们得到：

$$R^2 = r_{XY}^2$$

□