MATH1550, Winter 2023 **Names:**
Seminar 10

---

1. Suppose X and Y are discrete random variables with joint probability distribution given below.

		x		
		1	2	3
	1	0.10	0.08	0.06
y	2	0.12	0.12	0.06
	3	0.16	0.10	0.20

(a) Find $E(X)$ and $E(Y)$.

(b) Find $\text{var}(X)$ and $\text{var}(Y)$.

(c) Find $\text{cov}(X, Y)$.

(d) The *correlation* between X and Y is defined as $\rho(X, Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$, provided $\text{var}(X), \text{var}(Y) \neq 0$. It provides a measure of the degree of linearity between X and Y (this would appear in course on statistics). Find the correlation between X and Y.

Solution. (a)

$$E(X) = \sum_x \sum_y xf(x, y) = 1\cdot(0.10+0.12+0.16)+2\cdot(0.08+0.12+0.10)+3\cdot(0.06+0.06+0.20) = 1.94$$

$$E(Y) = \sum_x \sum_y yf(x, y) = 1\cdot(0.10+0.08+0.06)+2\cdot(0.12+0.12+0.06)+3\cdot(0.16+0.10+0.20) = 2.22$$

(b)

$$E(X^2) = \sum_x \sum_y xf(x, y) = 1^2\cdot(0.10+0.12+0.16)+2^2\cdot(0.08+0.12+0.10)+3^2\cdot(0.06+0.06+0.20) = 4.46$$

$$\text{var}(X) = E(X^2) - (E(X))^2 = 4.46 - (1.94)^2 = 0.6964.$$

$$E(Y^2) = \sum_x \sum_y yf(x, y) = 1^2\cdot(0.10+0.08+0.06)+2^2\cdot(0.12+0.12+0.06)+3^2\cdot(0.16+0.10+0.20) = 5.58$$

$$\text{var}(Y) = E(Y^2) - (E(Y))^2 = 5.58 - (2.22)^2 = 0.6516.$$

(c)

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyf(x, y) \\ &= 1 \cdot (0.10) + 2 \cdot (0.08) + 3 \cdot (0.06) + 2 \cdot (0.12) + 4 \cdot (0.12) \\ &\quad + 6 \cdot (0.06) + 3 \cdot (0.16) + 6 \cdot (0.10) + 9 \cdot (0.20) \\ &= 4.4 \end{aligned}$$

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 4.4 - (1.94)(2.22) = 0.0932.$$

(d)

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{0.0932}{\sqrt{0.6964}\sqrt{0.6516}} \approx 0.1384.$$

□

2. The joint density function for continuous random variables X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{3}(x+y) & 0 < x < 1, 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}.$$

Find $\text{cov}(X,Y)$. Are X and Y independent?

Solution.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dx\,dy \\ &= \int_0^2\int_0^1 \frac{x}{3}(x+y)\,dx\,dy \\ &= \int_0^2 \frac{x^3}{9} + \frac{x^2y}{6}\Big|_0^1\,dy \\ &= \int_0^2 \frac{1}{9} + \frac{y}{6}\,dy \\ &= \frac{y}{9} + \frac{y^2}{12}\Big|_0^2 \\ &= \frac{5}{9} \end{aligned}$$

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dx\,dy \\ &= \int_0^2\int_0^1 \frac{y}{3}(x+y)\,dx\,dy \\ &= \int_0^2 \frac{x^2y}{6} + \frac{xy^2}{3}\Big|_0^1\,dy \\ &= \int_0^2 \frac{y}{6} + \frac{y^2}{3}\,dy \\ &= \frac{y^2}{12} + \frac{y^3}{9}\Big|_0^2 \\ &= \frac{11}{9} \end{aligned}$$

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)\ dx\ dy \\
&= \int_0^2 \int_0^1 \frac{xy}{3}(x+y)\ dx\ dy \\
&= \int_0^2 \frac{x^3y}{9} \frac{x^2y^2}{6} \Big|_0^1\ dy \\
&= \int_0^2 \frac{y}{9} \frac{y^2}{6}\ dy \\
&= \frac{y^2}{18} \frac{y^3}{18} \Big|_0^2 \\
&= \frac{2}{3}
\end{aligned}
$$

$$
\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{2}{3} - \left(\frac{5}{9}\right)\left(\frac{11}{9}\right) = -\frac{1}{81}.
$$

Since $\mathrm{cov}(X,Y) \neq 0$, it follows that X and Y are not independent. □

3. Let X be a continuous random variable with probability density given by

$$
f(x) = \begin{cases} 1+x & -1 < x \leq 0 \\ 1-x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}.
$$

Let $U = X$ and $V = X^2$. Show that $\mathrm{cov}(U,V) = 0$.

Solution.

$$
\begin{aligned}
E(U) = E(X) &= \int_{-\infty}^{\infty} xf(x)\ dx \\
&= \int_{-1}^{0} x + x^2\ dx + \int_0^1 x - x^2\ dx \\
&= \frac{x^2}{2} + \frac{x^3}{3}\Big|_{-1}^{0} + \frac{x^2}{2} - \frac{x^3}{3}\Big|_0^1 \\
&= -\frac{1}{2} + \frac{1}{3} + \frac{1}{2} - \frac{1}{3} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
E(V) = E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\ dx \\
&= \int_{-1}^{0} x^2 + x^3\ dx + \int_0^1 x^2 - x^3\ dx \\
&= \frac{x^3}{3} + \frac{x^4}{4}\Big|_{-1}^{0} + \frac{x^3}{3} - \frac{x^4}{4}\Big|_0^1 \\
&= \frac{1}{3} - \frac{1}{4} + \frac{1}{3} - \frac{1}{4} \\
&= \frac{1}{6}
\end{aligned}
$$

$$
\begin{aligned}
E(UV) = E(X^3) &= \int_{-\infty}^{\infty} x^3 f(x)\, dx \\
&= \int_{-1}^{0} x^3 + x^4\, dx + \int_{0}^{1} x^3 - x^4\, dx \\
&= \frac{x^4}{4} + \frac{x^5}{5}\Big|_{-1}^{0} + \frac{x^4}{4} - \frac{x^5}{5}\Big|_{0}^{1} \\
&= -\frac{1}{4} + \frac{1}{5} + \frac{1}{4} - \frac{1}{5} \\
&= 0
\end{aligned}
$$

$$
\text{cov}(U, V) = E(UV) - E(U)E(V) = 0 - 0\left(\frac{1}{6}\right) = 0.
$$

□

4. Suppose 2 balls are removed (without replacement) from an urn containing n red balls and m blue balls, with $n, m \geq 2$. For $i = 1, 2$, let $X_i = 1$ if the ith ball removed is red and $X_i = 0$ if it is blue (i.e. not red).

   (a) Do you think $\text{cov}(X_1, X_2)$ is positive, negative or zero?

   (b) Compute $\text{cov}(X_1, X_2)$ to justify your answer to (a).

   (c) Suppose the red balls are numbered 1 through n. Let $Y_i = 1$ if red ball number i is removed, and $Y_i = 0$ otherwise. Do you think $\text{cov}(Y_1, Y_2)$ is positive, negative or zero?

   (d) Compute $\text{cov}(Y_1, Y_2)$ to justify your answer to (c).

Solution. (a) We might guess negative here. If the first ball is red, there is a greater chance the second one will not be red, and vice versa. i.e. there is a greater probability that high values for X_1 occur with low values for X_2 and vice versa.

(b)

$$
\begin{aligned}
E(X_1) &= \sum_{x_1}\sum_{x_2} x_1 f(x_1, x_2) \\
&= 0 \cdot \left(\frac{m(m-1)}{(n+m)(n+m-1)} + \frac{nm}{(n+m)(n+m-1)}\right) \\
&\quad + 1 \cdot \left(\frac{nm}{(n+m)(n+m-1)} + \frac{n(n-1)}{(n+m)(n+m-1)}\right) \\
&= \frac{n}{n+m}
\end{aligned}
$$

$$
\begin{aligned}
E(X_2) &= \sum_{x_1}\sum_{x_2} x_2 f(x_1, x_2) \\
&= 0 \cdot \left(\frac{m(m-1)}{(n+m)(n+m-1)} + \frac{nm}{(n+m)(n+m-1)}\right) \\
&\quad + 1 \cdot \left(\frac{nm}{(n+m)(n+m-1)} + \frac{n(n-1)}{(n+m)(n+m-1)}\right) \\
&= \frac{n}{n+m}
\end{aligned}
$$

$$
\begin{aligned}
E(X_1X_2) &= \sum_{x_1}\sum_{x_2} x_1x_2 f(x_1,x_2) \\
&= 0\cdot\left(\frac{m(m-1)}{(n+m)(n+m-1)} + \frac{nm}{(n+m)(n+m-1)} + \frac{nm}{(n+m)(n+m-1)}\right) \\
&\quad + 1\cdot\left(\frac{n(n-1)}{(n+m)(n+m-1)}\right) \\
&= \frac{n(n-1)}{(n+m)(n+m-1)}
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(X_1,X_2) = E(X_1X_2) - E(X_1)E(X_2) &= \frac{n(n-1)}{(n+m)(n+m-1)} - \left(\frac{n}{n+m}\right)^2 \\
&= \frac{n(n-1)(n+m) - n^2(m+n-1)}{(n+m)^2(n+m-1)} \\
&= \frac{-nm}{(n+m)^2(n+m-1)}
\end{aligned}
$$

Therefore $\operatorname{cov}(X_1,X_2) < 0$.

(c) A tough call? The likelihood of drawing ball 1 and/or 2 is low, and so the expected value for Y_1 and Y_2 should be closer to zero. There is a low probability that both Y_1 and Y_2 are 1 simultaneously, however there is a high probability that they are both 0 simultaneously. We will calculate directly to find out.

(d)

$$
\begin{aligned}
E(Y_1) = \sum_{y_1} y_1 f(y_1) &= 0\cdot\left(\frac{(n+m-1)(n+m-2)}{(n+m)(n+m-1)}\right) + 1\cdot\left(\frac{2(n+m-1)}{(n+m)(n+m-1)}\right) \\
&= \frac{2}{(n+m)}
\end{aligned}
$$

Similarly $E(Y_2) = \frac{2}{(n+m)}$.

$$
\begin{aligned}
E(Y_1Y_2) &= \sum_{y_1}\sum_{y_2} y_1y_2 f(y_1,y_2) \\
&= 0\cdot\left(\frac{(n+m-2)(n+m-3)}{(n+m)(n+m-1)} + \frac{2(n+m-2)}{(n+m)(n+m-1)} + \frac{2(n+m-2)}{(n+m)(n+m-1)}\right) \\
&\quad + 1\cdot\left(\frac{2}{(n+m)(n+m-1)}\right) \\
&= \frac{2}{(n+m)(n+m-1)}
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(Y_1,Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) &= \frac{2}{(n+m)(n+m-1)} - \left(\frac{2}{n+m}\right)^2 \\
&= \frac{2(n+m) - 4(n+m-1)}{(n+m)^2(n+m-1)} \\
&= \frac{4-2n-2m}{(n+m)^2(n+m-1)}
\end{aligned}
$$

Since $m, n \geq 2$, it follows that $\mathrm{cov}(Y_1, Y_2) < 0$.

$\square$