MATH-UA.0235 Probability and Statistics – Worksheet # 6

Problem 1

Carefully verify that the covariance between two random variables X and Y satisfies:

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

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E[(X - E[X])(Y - E[Y])] \\
&= E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\
&= E[XY] - E[YE[X]] - E[XE[Y]] + E[E[X]E[Y]] \\
&= E[XY] - E[X]E[Y] - E[X][Y] + E[X]E[Y] \\
&= E[XY] - E[X]E[Y]
\end{aligned}$$

Problem 2

Consider a discrete random variable X which takes values -1, 0, and 1, with probability $\frac{1}{3}$ for each. We define the random variable $Y = X^2$.

1. Are X and Y independent?

$$P(X = -1, Y = 0) = 0 \neq P(X = -1)P(Y = 0) = \frac{1}{9}$$

X and Y are not independent. Intuitively, that makes perfect sense: the value of X fully determines the value of Y.

2. Are X and Y correlated?
$E[X] = 0$ and $E[Y] = \frac{2}{3}$. Therefore, we have

$$\mathrm{Cov}(X,Y) = E[X(Y - \frac{2}{3})] = \frac{1}{3} \cdot 0 \cdot (-\frac{2}{3}) + \frac{1}{3}(-1) \cdot \frac{1}{3} + \frac{1}{3} 1 \cdot \frac{1}{3} = 0$$

The random variables X and Y have zero covariance. We say that they are uncorrelated.

Problem 3

Let X and Y be two independent geometric random variables with parameters $p \in (0,1)$ and $q \in (0,1)$ respectively. Compute the mean of the random variable $Z = \max(X,Y)$ using Problem 1 of Problem Set 5.

From Problem Set 5, we have the equality

$$E[Z] = \sum_{n=0}^{+\infty} P(Z > n)$$

Now, $P(Z > n) = P((X > n) \cup (Y > n)) = P(X > n) + P(Y > n) - P(X > n, Y > n)$. Since the random variables X and Y are independent, we may write

$$P(Z > n) = P(X > n) + P(Y > n) - P(X > n)P(Y > n)$$

Since X is a geometric random variable with parameter p,

$$P(X > n) = p(1-p)^n \sum_{k=0}^{+\infty} (1-p)^k = p(1-p)^n \cdot \frac{1}{1-(1-p)} = (1-p)^n$$

Likewise,

$$P(Y > n) = (1-q)^n$$

Hence,

$$P(Z > n) = (1-p)^n + (1-q)^n - (1-p)^n(1-q)^n$$

And we can therefore derive

$$E[Z] = \sum_{n=0}^{+\infty} (1-p)^n + \sum_{n=0}^{+\infty} (1-q)^n - \sum_{n=0}^{+\infty} [(1-p)(1-q)]^n = \frac{1}{1-(1-p)} + \frac{1}{1-(1-q)} - \frac{1}{1-(1-p)(1-q)}$$
$$= \frac{1}{p} + \frac{1}{q} - \frac{1}{p+q-pq}$$

Problem 4

Let X and Y be two random variables which take values in $\mathbb{N}$. You are told that X is a Poisson random variable with parameter $\lambda > 0$, and that the probability mass function of Y *given* $X = n$ follows a binomial distribution with parameters n and $p \in (0,1)$.

1. Determine the joint probability mass function of X and Y.

   From the statement of the problem, we have

$$P(Y = k | X = n) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \leq n \\ 0 \quad \text{otherwise} \end{cases}$$

   Furthermore, $P(Y = k, X = n) = P(Y = k|X = n)P(X = n) = P(Y = k|X = n)\frac{\lambda^n}{n!}e^{-\lambda}$. Hence

$$p_{X,Y}(n,k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} & \text{if } k \leq n \\ 0 \quad \text{otherwise} \end{cases}$$

2. What is the marginal distribution of Y?

   The marginal distribution of Y is given by

$$p_Y(k) = \sum_{n=0}^{+\infty} p_{X,Y}(n,k) = \sum_{n=k}^{+\infty} p_{X,Y}(n,k) = \sum_{n=k}^{+\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda}$$

   Hence,

$$p_Y(k) = \frac{p^k \lambda^k}{k!} e^{-\lambda} \sum_{n=k}^{+\infty} (1-p)^{n-k} \frac{\lambda^{n-k}}{(n-k)!} = \frac{p^k \lambda^k}{k!} e^{-\lambda} \sum_{n=0}^{+\infty} (1-p)^n \frac{\lambda^n}{n!} = \frac{p^k \lambda^k}{k!} e^{-\lambda} e^{(1-p)\lambda} = \frac{p^k \lambda^k}{k!} e^{-p\lambda}$$

Problem 5

An urn contains n numbered tokens, with $n > 1$, which are such that they cannot be distinguished if one merely touches them. One picks two tokens in a row, one after the other, and does not put the first token back in the urn before picking the second token. Let X be the random variable corresponding to the value on the first token picked, and Y be the random variable corresponding to the value on the second token picked.

1. What is the joint probability mass function of X and Y?

   The random variables X and Y can each take values $1, 2, \ldots, n$. Since the first token is not put back into the urn, we have

$$p_{X,Y}(i,j) = P(Y=j|X=i)P(X=i) \begin{cases} \frac{1}{n(n-1)} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

2. What is the covariance between X and Y?

   For this question, we may use the expression $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$. To compute $E[X]$ and $E[Y]$, we compute the marginal distributions of X and Y. By symmetry of the joint probality mass function, these marginal distributions are the same. We have

$$p_X(i) = \frac{n-1}{n(n-1)} = \frac{1}{n} \quad , \quad i = 1, 2, \ldots, n$$

   and the same marginal distribution for Y. Hence

$$E[X] = E[Y] = \sum_{i=1}^{n} i\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2}$$

   Considering the table of possible values for XY, and its symmetry, we can write

$$E[XY] = 2\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} \frac{ij}{n(n-1)} = 2\frac{1}{n(n-1)}\sum_{j=1}^{n-1} j\frac{(n-j)(n+j+1)}{2} = \frac{1}{n(n-1)}\sum_{j=1}^{n-1}[(n^2+n)j - j^3 - j^2]$$

   We have the equalities

$$\sum_{j=1}^{n-1} j = \frac{n(n-1)}{2} \qquad \sum_{j=1}^{n-1} j^2 = \frac{(n-1)n(2n-1)}{6} \qquad \sum_{j=1}^{n-1} j^3 = \frac{(n-1)^2n^2}{4}$$

   Thus

$$E[XY] = \frac{n^2+n}{2} - \frac{2n-1}{6} - \frac{(n-1)n}{4} = \frac{6n^2+6n-4n+2-3n^2+3n}{12} = \frac{3n^2+5n+2}{12}$$

   We conclude that

$$\mathrm{Cov}(X,Y) = \frac{3n^2+5n+2}{12} - \frac{n^2+2n+1}{4} = \frac{-n-1}{12}$$

   X and Y are negatively correlated.

3. Compute the correlation coefficient $\rho(X,Y)$

   We need to compute $\mathrm{Var}(X) = \mathrm{Var}(Y)$.

$$E[X^2] = \sum_{i=1}^{n} i^2\frac{1}{n} = \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6} = E[Y^2]$$

   Therefore,

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{(n+1)(2n+1)}{6} - \frac{n^2+2n+1}{4} = \frac{n^2-1}{12}$$

   We can then conclude

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-n-1}{12}\frac{12}{n^2-1} = -\frac{1}{n-1}$$

   We see that the correlation coefficient goes to 0 as $n \to +\infty$, as makes intuitive sense.

   For $n = 3$, the coefficient is equal to -50%, which is not a strong correlation. Its magnitude decreases with higher n, falling already to -10% as $n = 11$.

Problem 6

Students take a two-part standardized exam, consisting of a verbal part, and a quantitative part. Let X be the random variable for a student's normalized score on the verbal test, and let Y be the random variable for the same student's normalized score on the quantitative test.

So many students take the test that X and Y are well approximated by continuous random variables, and the organization in charge of the test finds the following joint probability density function for X and Y:

$$f_{X,Y}(x,y) = \begin{cases} 2xy + 0.5 & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the covariance $\mathrm{Cov}(X,Y)$.

For this problem, we will practice with the original definition for the covariance:

$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$. We first need to compute $E[X]$ and $E[Y]$. Since x and y play the same role in the expression for f_{XY}, we have $E[X] = E[Y]$. We may now write

$$E[X] = E[Y] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f_{X,Y}(x,y)dxdy = \int_0^1\int_0^1 y(2xy+0.5)dxdy = \int_0^1 \left[x^2y^2 + 0.5y\right]_0^1 dy = \int_0^1 (y^2 + 0.5y)dy$$
$$= \left[\frac{y^3}{3} + \frac{y^2}{4}\right]_0^1 = \frac{7}{12}$$

We are now ready to calculate

$$\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - \frac{7}{12})(y - \frac{7}{12}) f_{X,Y}(x,y)dxdy = \int_0^1\int_0^1 (x-\frac{7}{12})(y-\frac{7}{12})(2xy+0.5)dx$$
$$= \int_0^1 (y - \frac{7}{12})\left[\int_0^1 \left(2x^2y + 0.5x - \frac{7}{6}xy - \frac{7}{24}\right)dx\right]dy = \int_0^1 (y-\frac{7}{12})\left[\frac{2}{3}x^3y + \frac{x^2}{4} - \frac{7}{12}xy - \frac{7}{24}x\right]_0^1 dy$$
$$= \int_0^1 \left(y - \frac{7}{12}\right)\left(\frac{y}{12} - \frac{1}{24}\right)dy = \left[\frac{y^3}{36} - \frac{13y^2}{288} + \frac{7y}{288}\right]_0^1 = \frac{1}{144}$$

Problem 7

Consider random, uniformly distributed points on the unit circle, and X the random variable corresponding to their x coordinate, and Y the random variable corresponding to their y coordinate. Compute the correlation coefficient $\rho(X,Y)$.

By symmetry of the problem, $E[X] = E[Y]$.

$$E[Y] = E[X] = \frac{1}{\pi}\int_{-1}^1 \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} xdy\right)dx = \frac{1}{\pi}\int_{-1}^1 2x\sqrt{1-x^2}dx = -\frac{2}{3\pi}\left[(1-x^2)^{3/2}\right]_{-1}^1 = 0$$

We are now ready to compute the covariance between X and Y:

$$\mathrm{Cov}(X,Y) = E[XY] = \frac{1}{\pi}\int_{-1}^1 \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} xydy\right)dx = \frac{1}{\pi}\int_{-1}^1 x\left[\frac{y^2}{2}\right]_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dx = 0$$

We conclude that the random variables X and Y are uncorrelated. We can then infer that the correlation coefficient is also zero.