[Homework] Variance

1. A certain project will be undertaken in 6 stages. There is a 95% chance that each stage will be completed on time independently.

   (a) Compute the probability that all 6 stages are completed on time.
   Sol. Let X_i be the ith stage.

   $$X_i = \begin{cases} 1 & p = 0.95 \\ 0 & q = 0.05 \end{cases}$$

   $$P(X_1) \cdots P(X_6) = (0.95)^6$$

   (b) Compute the expectation and variance of the number of stages that will be completed on time.
   Sol. X be the umber of stages that will be completed on time.

   $$X = X_1 + \cdots + X_6$$

   $$E(X) = E(X_1 + \cdots + X_6) = 6 \cdot 0.95 = 5.7$$

   $$Var(X) = 6 \cdot 0.95 \cdot 0.05 = 0.285$$

2. Find the mean and variance of $1 + 2X + 3Y$ where $E(X) = 1$, $Var(X) = 5$, $E(Y) = 2$, $Var(Y) = 2$, and $\rho = 0.5$.
   Sol.

   $$E[1 + 2X + 3Y] = 1 + 2E[X] + 3E[X] = 1 + 2 \cdot 1 + 3 \cdot 2 = 9$$

   Since

   $$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

   $$Cov(X,Y) = 0.5\sqrt{10}$$

   $$\begin{aligned} Var(1 + 2X + 3Y) &= Var(2X) + Var(3Y) + 2 \cdot Cov(2X, 3Y) \\ &= 2^2 Var(X) + 3^2 Var(Y) + 2 \cdot 2 \cdot 3 Cov(X,Y) \\ &= 38 + 6\sqrt{10} \end{aligned}$$

3. Suppose that a fair coin is rolled twice, resulting 1 (head) or 0 (tail) each time independently.

(a) Calculate the variance of the maximum value X to appear in the two rolls.
Sol.

i	0	1	
$P(X=i)$	$\frac{1}{4}$	$\frac{3}{4}$	1

$$EX = 0 \cdot \frac{1}{4} + 1 \cdot \frac{3}{4} = \frac{3}{4}$$
$$VarX = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{3}{4} - (\frac{3}{4})^2 = \frac{3}{16}$$

(b) Calculate the variance of the minimum value Y to appear in the two rolls.
Sol.

i	0	1	
$P(Y=i)$	$\frac{3}{4}$	$\frac{1}{4}$	1

$$EY = 0 \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} = \frac{1}{4}$$
$$VarY = 0^2 \cdot \frac{3}{4} + 1^2 \cdot \frac{1}{4} - (\frac{1}{4})^2 = \frac{3}{16}$$

(c) Calculate $Cov(X,Y)$ and check the sign of $Cov(X,Y)$ with your intuition.
Sol.

i	0	1	
$P(XY=i)$	$\frac{3}{4}$	$\frac{1}{4}$	1

$$Cov(X,Y) = E(XY) - EX \cdot EY = \frac{1}{16}$$
$$VarX = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{3}{4} - (\frac{3}{4})^2 = \frac{3}{16}$$

(d) Calculate the variance of the sum of the two rolls.
Sol.

i	0	1	2	
$P(X=i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

$$EX = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$
$$VarX = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} - 1^2 = \frac{1}{2}$$

(e) Calculate the variance of the first roll number minus the second roll number.
Sol.

i	-1	0	1	
$P(X = i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

$$EX = (-1) \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 0$$
$$VarX = (-1)^2 \cdot \frac{1}{4} + 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{4} - 0^2 = \frac{1}{2}$$

4. Let X_i be independent with same mean 0, and variance 2. Let $Y_n = X_n + X_{n+1} + X_{n+2}$. For $j \geq 0$, calculate $Cov(Y_n, Y_{n+j})$.
Sol. Since X_i is independent

$$Cov(X_i, X_j) = \begin{cases} 0 & i \neq j \\ 2 & i = j \end{cases}$$

If $j = 0$,

$$Cov(Y_n, Y_n) = Var(Y_n) = Var(X_n + X_{n+1} + X_{n+2}) = Var(X_n) + Var(X_{n+1}) + Var(X_{n+2}) = 6$$

If $j = 1$,
$$Cov(Y_n, Y_{n+1}) = Var(X_{n+1}) + Var(X_{n+2}) = 4$$

If $j = 2$,
$$Cov(Y_n, Y_{n+2}) = Var(X_{n+2}) = 2$$

If $j \geq 3$,
$$Cov(Y_n, Y_{n+3}) = 0$$

5. Consider a graph having n vertices labeled $1, 2, \ldots, n$, and suppose that, between each of the $\binom{n}{2}$ pairs of distinct vertices, an edge is independently present with probability p. The degree of vertex i, designated as D_i, is the number of edges that have vertex i as one of their vertices.

(a) What is the distribution of D_i?

Sol. Since each edge occurs independently with probability p, the number of edges connected to a given node is a binomial with parameters $n-1$ and p ($n-1$ is the number of other nodes. hence the number of possible edges).

(b) Find $\rho(D_i, D_j)$, the correlation between D_i and D_j.

Sol. Let $X_{ij} = 1$ if there is an edge between vertices i and j and 0 otherwise. Note that $X_{ij} = X_{ji}$. Then

$$D_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} X_{ij}$$

Thus, for $i \neq j$

$$\begin{aligned} Cov(D_i, D_j) &= Cov(\sum_{k \neq i} X_{ik}, \sum_{k \neq j} X_{jk}) \\ &= \sum_{k \neq i} \sum_{l \neq j} Cov(X_{ik}, X_{jl}) \\ &= Cov(X_{ij}, X_{ji}) = Var(X_{ij}) = p(1-p) \end{aligned}$$

Where the step on the last line follows from noting that $k = j$ and $l = i$ are the only possible values of k and l for which X_{ik} and X_{jl} are not independent. Thus the correlation is then

$$\begin{aligned} \rho(D_i, D_j) &= \frac{Cov(D_i, D_j)}{\sqrt{Var(D_i)}\sqrt{Var(D_j)}} \\ &= \frac{p(1-p)}{(n-1)p(1-p)} \\ &= \frac{1}{n-1} (i \neq j) \end{aligned}$$

for $i = j$, $\rho(D_i, D_j) = 1$

6. A box contains 5 red and 5 blue marbles. Two marbles are withdrawn randomly. If they are the same color, then you win \$1.10; if they are different, then you lose \$1.00. Calculate the mean and variance of the amount you win.

Sol.

$$P(X = 1.1) = \frac{\binom{5}{2} + \binom{5}{2}}{\binom{10}{2}} = \frac{4}{9}$$

$$P(X = -1) = \frac{\binom{5}{1} \cdot \binom{5}{1}}{\binom{10}{2}} = \frac{5}{9}$$

$$E(X) = 1.1 \cdot \frac{4}{9} + (-1) \cdot \frac{5}{9} = -\frac{1}{15}$$
$$Var(X) = 1.1^2 \cdot \frac{4}{9} + (-1)^2 \cdot \frac{5}{9} - (-\frac{1}{15})^2$$

7. A sample of 3 items is selected at random from a box containing 20 items of which 4 are defective. Find the expectation and variance of the number of defective items in the sample.
   Sol. Let X be the number of defective items in the sample.
   Then $X \sim HG(n, m, N) = HG(3, 4, 20)$
   $$EX = n \cdot \frac{m}{N} = 3 \cdot \frac{4}{20} = \frac{3}{5}$$
   $$VarX = n \cdot \frac{m}{N} \cdot (1 - \frac{m}{N}) \cdot (1 - \frac{n-1}{N-1}) = 3 \cdot \frac{4}{20} \cdot (1 - \frac{4}{20}) \cdot (1 - \frac{3-1}{20-1}) = \frac{204}{475}$$
8. Five distinct numbers are randomly distributed to players numbered 1 through 5. Whenever two players compare their numbers, the one with the higher one is declared the winner. Initially, players 1 and 2 compare their numbers; the winner then compares her number with that of player 3, and so on. Let X denote the number of times player 1 is a winner. Find the expectation and variance of X.
   Sol. In Homework4 #4,
   $$P(X = 0) = \frac{1}{2}, P(X = 1) = \frac{1}{6}, P(X = 2) = \frac{1}{12}$$
   $$P(X = 3) = \frac{1}{20}, P(X = 4) = \frac{1}{5}$$
   $$EX = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{12} + 3 \cdot \frac{1}{20} + 4 \cdot \frac{1}{5} = \frac{77}{60}$$
   $$VarX = 0^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{12} + 3^2 \cdot \frac{1}{20} + 4^2 \cdot \frac{1}{5} - (\frac{77}{60})^2$$
9. We mix the ordinary deck of the 52 cards and choose two cards. If we have 2 aces, we stop. Otherwise we mix the deck and choose two cards again. We do this until we get 2 aces. What is the expectation and variance of the number of trials to get the 2 aces?
   Sol.
   $$p = \frac{\binom{4}{2}}{\binom{52}{2}} = \frac{1}{221}$$
   Since $X \sim Geo(p)$
   $$EX = \frac{1}{p} = 221$$

$$VarX = \frac{q}{p^2} = 221 \cdot 220 = 48620$$

10. Someone I know claims to be able to flip a coin in such a way that he can make it land head 90% of the time, on the average. I want to test the hypothesis that he's bluffing against the alternative that he is right. I propose to test this hypothesis by having him flip the coin again and again until it first lands tail. If it takes more than 4 tries, I'll conclude that he's right. Assume that the flips are independent.

    (a) Under the null hypothesis $p = 0.5$ that he cannot influence the outcome, identify the distribution of the number of spins until the coin lands tail.
    Sol. Let X be the number of spins until the coin lands tail under the null hypothesis.
    Then $X \sim Geo(0.5)$

    (b) What is the expectation and variance of the number of spins to the first tail under the alternative hypothesis $p = 0.9$(head)?
    Sol. Let Y be the number of spinsto the first tail under the alternative hypothesis.
    Then $Y \sim Geo(0.1)$
    $$EY = \frac{1}{p} = 10$$
    $$VarY = \frac{q}{p^2} = 90$$

Extra

1. We flips a p-coin many times, where $p = 0.40$.

$$\begin{array}{rl} X_i & i^{th} \text{ flip record, where } H \text{ and } T \text{ are recorded as 1 and 0} \\ Y_i := 2X_i - 1 & i^{th} \text{ flip record, where } H \text{ and } T \text{ are recorded as 1 and } -1 \end{array}$$

Calculate the mean and variance of the following related random variables, i.e., fill up blanks of the below table.

Random variable	Mean	Variance
Y_i		
$\sum_{i=1}^{n} Y_i$		
$\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}}$		

2. On a multiple-choice exam with 4 possible answers for each of the 20 questions, what is the mean and variance of the number of correct answers that a student will get just by guessing?

3. 20 people consisting of 10 couples are in an island. Each person lives after 1 year with probability 0.5, independently. Let X be the number of surviving couples after 1 year.

   (a) Identify the distribution of X.
   Sol.
   $$X_i = \begin{cases} 1 & i\text{th couple survive } p = \frac{1}{2}\cdot\frac{1}{2} \\ 0 & \text{otherwise } q = 1 - \frac{1}{4} \end{cases}$$
   Then $X_i \sim B(\frac{1}{4})$
   $$X = X_1 + X_2 + \cdots + X_{10}$$
   $$X \sim B(10, \frac{1}{4})$$

   (b) Calculate the mean and variance of X.
   Sol.
   $$E(X) = np = 10 \cdot \frac{1}{4} = \frac{5}{2}$$
   $$Var(X) = npq = 10 \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{15}{8}$$

4. An urn contains 4 white and 4 black balls. We randomly choose 4 balls. If 2 of them are white and 2 are black, we stop. If not, we replace the balls in the urn and again randomly select 4 balls. This continues until exactly 2 of the 4 chosen are white. What is the mean and variance of the number of trials?

5. Let X_i, $1 \le i \le 4$, have mean 3 and variance 2 and the correlation ρ_{ij} between two are all 0.5. Compute

   (a) $Cov(X_1 + X_2, X_2 + X_3)$.
   Sol.
   $$\begin{aligned} Cov(X_1+X_2, X_2+X_3) &= Cov(X_1,X_2) + Cov(X_1,X_3) + Cov(X_2,X_2) + Cov(X_2,X_3) \\ &= \rho\{\sqrt{Var(X_1)Var(X_2)} + \sqrt{Var(X_1)Var(X_3)} \\ &\quad + Var(X_2) + \sqrt{Var(X_2)Var(X_3)}\} \\ &= 0.5(2+2+2+2) = 4 \end{aligned}$$

(b) $Cov(X_1 + X_2, X_3 + X_4)$.

Sol.

$$\begin{aligned}
Cov(X_1 + X_2, X_3 + X_4) &= Cov(X_1, X_3) + Cov(X_1, X_4) + Cov(X_2, X_3) + Cov(X_2, X_4) \\
&= \rho\{\sqrt{Var(X_1)Var(X_3)} + \sqrt{Var(X_1)Var(X_4)} \\
&\quad + \sqrt{Var(X_2)Var(X_3)} + \sqrt{Var(X_2)Var(X_4)}\} \\
&= 0.5(2 + 2 + 2 + 2) = 4
\end{aligned}$$