

Student Information

Name : Onat Özdemir

ID : 2310399

Answer 1

a)

Since $\{Y = 0\}$ and $\{Y = 2\}$ are mutually exclusive and exhaustive events, we can use the Addition Rule to calculate $P\{X = x\}$;

$$P\{X = x\} = \sum_y P(x, y) \quad (1)$$

By using Table 1 and (1);

$$\begin{aligned} P\{X = 0\} &= \sum_y P(0, y) = P(0, 0) + P(0, 2) = 1/12 + 2/12 = 1/4 = 0.25 \\ P\{X = 1\} &= \sum_y P(1, y) = P(1, 0) + P(1, 2) = 4/12 + 2/12 = 1/2 = 0.50 \\ P\{X = 2\} &= \sum_y P(2, y) = P(2, 0) + P(2, 2) = 1/12 + 2/12 = 1/4 = 0.25 \end{aligned} \quad (2)$$

By the definition;

$$E(X) = \sum_x x * P(x) \quad (3)$$

Then, by using (2) and (3);

$$E(X) = \sum_x x * P(x) = 0 * P(0) + 1 * P(1) + 2 * P(2) = 0.50 + 2 * (0.25) = 1 \quad (4)$$

Let $\mu = E(X)$. Then, by the definition;

$$Var(X) = \sum_x (x - \mu)^2 * P(x) \quad (5)$$

We have already found μ in (4);

$$Var(X) = \sum_x (x - 1)^2 * P(x) \quad (6)$$

By using (2) and (6);

$$Var(X) = \sum_x (x - 1)^2 * P(x) = P(0) + 0 + P(2) = 0.50 \quad (7)$$

b)

Let Z be a discrete random variable such that $Z = X+Y$. As can be seen from Table 1, range of Z , $R = \{0, 1, 2, 3, 4\}$. Then,

$$\begin{aligned} P_Z(0) &= P\{X = 0 \cap Y = 0\} = P(0, 0) = 1/12 \\ P_Z(1) &= P\{X = 1 \cap Y = 0\} = P(1, 0) = 4/12 \\ P_Z(2) &= P\{X = 2 \cap Y = 0\} + P\{X = 0 \cap Y = 2\} = P(2, 0) + P(0, 2) = 1/12 + 2/12 = 3/12 \\ P_Z(3) &= P\{X = 1 \cap Y = 2\} = P(1, 2) = 2/12 \\ P_Z(4) &= P\{X = 2 \cap Y = 2\} = P(2, 2) = 2/12 \end{aligned}$$

As an important note, we were able to calculate $P_Z(2)$ as $P\{X = 2 \cap Y = 0\} + P\{X = 0 \cap Y = 2\}$ because $P_Z(2) = P\{(X = 2 \cap Y = 0) \cup (X = 0 \cap Y = 2)\} = P\{X = 2 \cap Y = 0\} + P\{X = 0 \cap Y = 2\}$ since $\{X = 0 \cap Y = 2\}$ and $\{X = 2 \cap Y = 0\}$ are mutually exclusive events.

c)

By the definition,

$$Cov(X, Y) = E(XY) - E(X)E(Y) \quad (8)$$

We have already found $E(X) = 1$ in (4). So let's find $E(Y)$ by using the definition in (3);

$$E(Y) = \sum_y y * P(y) \quad (9)$$

Since $\{X = 0\}$, $\{X = 1\}$ and $\{X = 2\}$ are mutually exclusive and exhaustive events, we can use the Addition Rule to calculate $P\{Y = y\}$;

$$P\{Y = y\} = \sum_x P(x, y) \quad (10)$$

By using Table 1 and (10);

$$\begin{aligned} P\{Y = 0\} &= \sum_x P(x, 0) = P(0, 0) + P(1, 0) + P(2, 0) = 1/12 + 4/12 + 1/12 = 6/12 = 0.50 \\ P\{Y = 2\} &= \sum_x P(x, 2) = P(0, 2) + P(1, 2) + P(2, 2) = 2/12 + 2/12 + 2/12 = 6/12 = 0.50 \end{aligned} \quad (11)$$

By using (9) and (11);

$$E(Y) = \sum_y y * P(y) = 0 * P(0) + 2 * P(2) = 2 * (0.5) = 1 \quad (12)$$

By the definition,

$$E(XY) = \sum_{(x,y)} x * y * P(x, y) \quad (13)$$

Then by using Table 1,

$$\begin{aligned}
E(XY) &= (0) * (0) * P(0, 0) + (0) * (2) * P(0, 2) + (1) * (0) * P(1, 0) \\
&\quad + (1) * (2) * P(1, 2) + (2) * (0) * P(2, 0) + (2) * (2) * P(2, 2) \\
E(XY) &= (1) * (2) * P(1, 2) + (2) * (2) * P(2, 2) \\
E(XY) &= (1) * (2) * (2/12) + (2) * (2) * (2/12) = 1
\end{aligned} \tag{14}$$

By using (4), (8), (12) and (14)

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 1 - 1 * 1 = 0$$

d)

Let's assume that A and B are independent random variables, then by the definition given in "Probability and Statistics for Computer Scientists, page 45";

$$P_{(A,B)}(a, b) = P_A(a) * P_B(b) \tag{15}$$

for all values of a and b.

Recall the equality given in (13);

$$E(AB) = \sum_{(a,b)} a * b * P(a, b) \tag{16}$$

From (15),

$$E(AB) = \sum_{(a,b)} a * b * P_A(a) * P_B(b) = \left(\sum_a a * P_A(a) \right) \left(\sum_b b * P_B(b) \right) \tag{17}$$

Recall the definition given in (3),

$$E(AB) = \left(\sum_a a * P_A(a) \right) \left(\sum_b b * P_B(b) \right) = E(A)E(B) \tag{18}$$

By using (8) and (18);

$$Cov(A, B) = E(AB) - E(A)E(B) = E(A)E(B) - E(A)E(B) = 0 \tag{19}$$

Since we assumed that A and B are independent random variables and then we have found that $Cov(A, B) = 0$ for these independent random variables in (19), we have proved that "if A and B are independent, then, $Cov(A, B) = 0$ ".

e)

Let's assume that X and Y are independent random variables, then by the definition given in (15),

$$P_{(X,Y)}(x, y) = P_X(x) * P_Y(y) \quad (20)$$

must satisfy for all values of x and y. We have already found $P_X(x)$ and $P_Y(y)$ in (2) and (11), respectively.

By using Table 1,

$$\begin{aligned} P_{(X,Y)}(0, 0) &= 1/12 \\ P_{(X,Y)}(0, 2) &= 2/12 \\ P_{(X,Y)}(1, 0) &= 4/12 \\ P_{(X,Y)}(1, 2) &= 2/12 \\ P_{(X,Y)}(2, 0) &= 1/12 \\ P_{(X,Y)}(2, 2) &= 2/12 \end{aligned} \quad (21)$$

Let's observe that whether for all (x,y) (20) holds or not by using (2), (11) and (21):

$$P_{(X,Y)}(0, 0) = 1/12 \neq P_X(0) * P_Y(0) = (1/4) * (1/2) = 1/8 \quad (22)$$

Since we have already found one (x,y) (which is (0,0)) that doesn't satisfy (20), in (22), this situation contradicts with our assumption. Thus, by using proof by contradiction, we have proved that X and Y are not independent random variables.

Answer 2

a)

Since choosing a broken pen is an independent event and a pen can be either broken or not, testing a pen can be seen as an independent "Bernoulli trial". Since we are testing 12 pens then we have 12 times repeated independent Bernoulli trial. Therefore, the distribution of number of broken pens out of 12 pens will be Binomial Distribution.

Let X be a Binomial random variable which represents the number of broken pens out of 12 pens. Then, in order to find the probability that at least 3 of chosen 12 pens are broken, we should find;

$$P\{X \geq 3\} = \sum_{x=3}^{12} P_X(x) = \sum_{x=0}^{12} P_X(x) - \sum_{x=0}^2 P_X(x) = 1 - \sum_{x=0}^2 P_X(x) \quad (23)$$

By the definition of Binomial cumulative distribution function;

$$\sum_{x=0}^2 P_X(x) = F(2) \quad (24)$$

To find $F(2)$, we should check the Binomial Distribution table which shows the values for cdf, placed in the "Probability and Statistics for Computer Scientists, page 412" by choosing our parameters as;

$$\begin{aligned}n(\text{number of trials}) &= 12 \\x(\text{number of successes}) &= 2 \\p(\text{probability of success}) &= 0.2\end{aligned}$$

The reason behind those parameters is that we have 12 times repeated Bernoulli Trial so we should choose $n=12$, we are trying to find the probability of having at most 2 broken pens out of 12 pens where choosing a broken pen is successful event for us therefore we should choose $x=2$ and lastly since choosing a broken pen is a successful event for us we should choose $p = \text{probability of choosing a broken pen} = 0.2$. As a result, $F(2) = 0.558$. Hence by using (23) and (24);

$$P\{X \geq 3\} = 1 - F(2) = 0.442$$

Thus, the probability that at least 3 of chosen 12 pens are broken equals to 0.442

b)

By the definition, in a sequence of independent Bernoulli Trials, the number of trials needed to obtain k successes has Negative Binomial Distribution. Since, as described in 2-a, testing a pen can be seen as an independent "Bernoulli Trial", then the number of trials needed to obtain second broken pen has Negative Binomial Distribution.

Let X be a Negative Binomial Random Variable which represents the number of trials needed to obtain second broken pen. Then, we are trying to find;

$$P\{X = 5\}$$

since the question asks "the probability that the fifth pen we test will be the second broken pen we find" which means we need to test 5 pens to obtain second broken pen. By using the formula placed in "Probability and Statistics for Computer Scientists, page 41";

$$P\{X = 5\} = F(5) - F(4) \tag{25}$$

where $F(x)$ is "Negative Binomial cumulative distribution function".

To find $F(5)$ and $F(4)$ we can use the " $\text{nbincdf}(x - R, R, p)$ " Matlab function which computes the negative binomial cdf at x using the corresponding number of successes, R and probability of success p . (the reason behind using " $x-R$ " as the first parameter rather than using x is that nbincdf function uses zero-based distribution) For $F(5)$, our parameters:

$$\begin{aligned}x(\text{number of trials needed to have second broken pen}) &= 5 \\R(\text{number of successes}) &= 2 \\p(\text{probability of success}) &= 0.2\end{aligned}$$

R equals to 2 since we want to obtain second broken pen (which means having 2 broken pens), p equals to 0.2 since choosing a broken pen is the success for us and probability of choosing a broken pen equals to 0.2, x equals to 5 since we are trying to find the probability of obtaining the second broken pen at most 5 trials.

As a result,

$$F(5) = \text{nbincdf}(5 - 2, 2, 0.2) = \text{nbincdf}(3, 2, 0.2) = 0.26272$$

For $F(4)$, our parameters:

$$x(\text{number of trials needed to have second broken pen}) = 4$$

$$R(\text{number of successes}) = 2$$

$$p(\text{probability of success}) = 0.2$$

R equals to 2 since we want to obtain second broken pen (which means having 2 broken pens), p equals to 0.2 since choosing a broken pen is the success for us and the probability of choosing a broken pen equals to 0.2, x equals to 4 since we are trying to find the probability of obtaining the second broken pen at most 4 trials.

As a result,

$$F(4) = \text{nbincdf}(4 - 2, 2, 0.2) = \text{nbincdf}(2, 2, 0.2) = 0.18080$$

By using (25);

$$P\{X = 5\} = F(5) - F(4) = 0.26272 - 0.18080 = 0.08192$$

Thus, the probability that the fifth pen we test will be the second broken pen we find equals to 0.08192

c)

By the definition, in a sequence of independent Bernoulli Trials, the number of trials needed to obtain k successes has Negative Binomial Distribution. Since, as described in 2-a, testing a pen can be seen as an independent "Bernoulli Trial", then the number of trials needed to obtain four broken pens has Negative Binomial Distribution.

Let X be a Negative Binomial Random Variable which represents the number of trials needed to find 4 broken pens. Since the question asks for the average number of tests needed to be done to find 4 broken pens we should find $E(X)$.

Since X is Negative Binomial Random Variable, to find $E(X)$ we can use the formula given in "Probability and Statistics for Computer Scientists, page 63";

$$E(X) = k/p \tag{26}$$

where k is number of successes (in our case it is number of broken pens) and p is the probability of success (probability of choosing a broken pen).

Since we are trying to find 4 broken pens we should choose $k=4$, also since choosing a broken pen is the success for us and the probability of choosing a broken pen equals to 0.2 then we should choose $p=0.2$. To sum up, our parameters will be;

$$\begin{aligned}k(\text{number of successes}) &= 4 \\ p(\text{probability of success}) &= 0.2\end{aligned}$$

Then using (26);

$$E(X) = 4/(0.2) = 20$$

Thus, on average, we are going to test 20 pens to find 4 broken pens.

Answer 3

a)

Since "The time until the first phone call and the times between two consecutive calls are independent exponential random variables" then the distribution of the time until the first phone call will be Exponential Distribution. Let X be an Exponential random variable which represents "the time until the first phone call (measured in hours)". Then we can represent the probability that Bob gets a phone call before 2 hours as;

$$P\{X < 2\}$$

Since for continuous distributions $P\{X = x\} = 0$ holds (Probability and Statistics for Computer Scientists, page 75) and exponential distribution is a continuous distribution;

$$P\{X < 2\} = P\{X \leq 2\} = F(2)$$

Since "Bob gets a phone call before 2 hours" and "Bob doesn't get a phone call for at least 2 hours" mutually exclusive and exhaustive events, then by the complement rule, "the probability that Bob doesn't get a phone call for at least 2 hours" can be represented as,

$$P\{X \geq 2\} = P\{X > 2\} + P\{X = 2\} = P\{X > 2\} = 1 - F(2) \quad (27)$$

In order to find $F(2)$ where $F(x)$ is "Exponential Cumulative Distribution Function", we can use a Matlab function which is $\text{expcdf}(x, \sigma)$ that represents the cumulative distribution function (CDF) at x of the exponential distribution with scale parameter σ such that $\sigma = 1/\lambda$ where λ is rate parameter. Since question says that "Bob gets a phone call every 4 hours on average", then the average number of calls that Bob gets per hour which equals to rate parameter is 0.25 calls per hour, therefore $\sigma = 1/\lambda = 1/(0.25) = 4$ By using $\text{expcdf}(x, \sigma)$ with the parameters;

$$\begin{aligned}x &= 2 \text{ (to calculate cdf at } x = 2\text{)} \\ \sigma &= 4\end{aligned}$$

$$F(2) = \text{expcdf}(2, 4) = 0.39347 \quad (28)$$

By using (27) and (28);

$$P\{X \geq 2\} = 1 - 0.39347 = 0.60653$$

Thus, the probability that Bob doesn't get a phone for at least 2 hours equals to 0.60653.

b)

Since it is extremely unlikely to get two phone calls simultaneously, getting phone calls are rare events (rare event definition is taken from "Probability and Statistics for Computer Scientists, page 64, part 3.4.5"). So, since getting phone calls are rare events and 10 hours is a fixed time period, distribution of the number of phone calls that Bob gets for the first 10 hours is a Poisson Distribution.

Let X be a Poisson Random Variable which represents the number of phone calls that Bob gets for the first 10 hours.

Then, the probability that for the first 10 hours, Bob gets at most 3 phone calls can be represented as;

$$P\{X \leq 3\} = F(3)$$

where $F(x)$ is the "cumulative distribution function" of Poisson Distribution.

Let λ_0 be the average number of calls per hour. Since the question says that Bob gets a phone call every 4 hours on average, then the average number of calls per hour equals to $\lambda_0 = 1/4 = 0.25$ calls per hour.

Let λ_1 be the average number of calls per 10 hours. To find λ_1 we should multiply the average number of calls per hour with 10 since our new time period is 10 hours. Then, $\lambda_1 = 10 * \lambda_0 = 2.5$ calls every 10 hours.

Then, we can find $F(3)$ from the Poisson Distribution Table given in "Probability and Statistics for Computer Scientists, Page 415" by choosing our parameters as;

$$\begin{aligned} \lambda \text{ (frequency)} &= \lambda_1 = 2.5 \\ x \text{ (number of calls in first 10 hours)} &= 3 \end{aligned}$$

Hence, $F(3) = 0.758$. Thus, the probability that for the first 10 hours, Bob gets at most 3 phone calls equals to 0.758.

c)

Since getting phone calls qualify as rare events(as explained in 3-b), probability distributions of the number of phone calls for the first 10 hours, and the number of phone calls for the first 16 hours are Poisson Distributions. Let X_1 be a Poisson Random Variable which represents the

number of phone calls for the first 10 hours and X_2 be a Poisson Random Variable which represents the number of phone calls for the first 16 hours. Then, given that Bob did not get more than 3 phone calls for the first 10 hours, the probability that he does not get more than 3 phone calls for the first 16 hours can be represented as;

$$P\{X_2 < 4 \mid X_1 < 4\} \quad (29)$$

Using the conditional probability formula given in "Probability and Statistics for Computer Scientists, Page 27";

$$P\{X_2 < 4 \mid X_1 < 4\} = \frac{P\{(X_2 < 4) \cap (X_1 < 4)\}}{P\{X_1 < 4\}} \quad (30)$$

Since the time until the first phone call and the times between two consecutive calls are independent exponential random variables, then the distribution of the time until the 4th call is Gamma Distribution. Let T be a Gamma Random variable which represents "the time until the 4th call (measured in hours)". Then, by using Gamma-Poisson formula given in "Probability and Statistics for Computer Scientists, Page 88";

$$\begin{aligned} P\{X_1 < 4\} &= P\{T > 10\} \\ P\{X_2 < 4\} &= P\{T > 16\} \end{aligned} \quad (31)$$

Then, if we replace the terms in (30) with the terms in (31);

$$P\{X_2 < 4 \mid X_1 < 4\} = \frac{P\{(T > 16) \cap (T > 10)\}}{P\{T > 10\}} \quad (32)$$

Since $\{T > 16\} \subset \{T > 10\}$ then by the Set Theory;

$$\{T > 16\} \cap \{T > 10\} = \{T > 16\}$$

Therefore,

$$P\{X_2 < 4 \mid X_1 < 4\} = \frac{P\{T > 16\}}{P\{T > 10\}} \quad (33)$$

If replace the values in (33) by using the equations given in (31);

$$P\{X_2 < 4 \mid X_1 < 4\} = \frac{P\{X_2 < 4\}}{P\{X_1 < 4\}} \quad (34)$$

For the poisson random variables X_1 and X_2 ;

$$\begin{aligned} P\{X_2 < 4\} &= P\{X_2 \leq 3\} = F_2(3) \\ P\{X_1 < 4\} &= P\{X_1 \leq 3\} = F_1(3) \end{aligned}$$

holds where $F_2(x)$ and $F_1(x)$ are "Poisson cumulative distribution functions" for X_2 and X_1 , respectively. Then, we can transform (34) to;

$$P\{X_2 < 4 \mid X_1 < 4\} = \frac{F_2(3)}{F_1(3)} \quad (35)$$

We can find the values of $F_2(3)$ and $F_1(3)$ from the Poisson Distribution Table given in "Probability and Statistics for Computer Scientists, Page 415". To find $F_2(3)$ we should choose our parameters as;

$$\begin{aligned}\lambda \text{ (frequency)} &= 4 \\ x \text{ (number of calls in first 16 hours)} &= 3 \\ F_2(3) &= 0.433\end{aligned}$$

Let me explain the logic behind these values. Since the question says that Bob gets a phone call every 4 hours on average, then the average number of calls per hour equals to $\lambda_0 = 1/4 = 0.25 \text{ calls per hour}$. To find λ we should multiply the average number of calls per hour with 16 since our new time period is 16 hours (because we defined X_2 as the number of phone calls for the first 16 hours). Then, $\lambda = 16 * \lambda_0 = 4$ calls every 16 hours. And since we are trying to find the probability of getting at most 3 calls for the first 16 hours, then we should choose $x = 3$

To find $F_1(3)$ we should choose our parameters as;

$$\begin{aligned}\lambda \text{ (frequency)} &= 2.5 \\ x \text{ (number of calls in first 10 hours)} &= 3 \\ F_1(3) &= 0.758\end{aligned}$$

Since the question says that Bob gets a phone call every 4 hours on average, then the average number of calls per hour equals to $\lambda_0 = 1/4 = 0.25 \text{ calls per hour}$. To find λ we should multiply the average number of calls per hour with 10 since our new time period is 10 hours (because we defined X_1 as the number of phone calls for the first 10 hours). Then, $\lambda = 10 * \lambda_0 = 2.5$ calls every 10 hours. And since we are trying to find the probability of getting at most 3 calls for the first 10 hours, then we should choose $x = 3$

Then, by using (35);

$$P\{X_2 < 4 \mid X_1 < 4\} = \frac{F_2(3)}{F_1(3)} = \frac{0.433}{0.758} = 0.57124010554$$

Thus, the probability that Bob does not get more than 3 phone calls for the first 16 hours, given that he did not get more than 3 phone calls for the first 10 hours equals to 0.57124010554.