DSCI 551 Quiz 1 Solutions

Exercise 1

A restaurant serves three fixed-price dinners costing at \$20, \$30 or \$40. Consider a randomly selected parent and child dining at the restaurant. Let X be the cost of the parent's dinner and Y be the cost of the child's dinner. The joint pmf of X and Y is given in the following table:

$p(x,y)$		y = 20	30	40
	20	0.05	0.05	0.10
x	30	0.05	0.10	0.35
	40	0	0.20	0.10

1.1

Find the marginal pmf of X and the marginal pmf of Y

rubric={reasoning:3}

Solution

For $X \in \{20, 30, 40\}$, we write

$$\mathbb{P}\{X = 20\} = \sum_{j=1}^{3} \mathbb{P}(X = 20 \cap Y = y_j) = 0.05 + 0.05 + 0.10 = 0.20;$$
$$\mathbb{P}\{X = 30\} = \sum_{j=1}^{3} \mathbb{P}(X = 30 \cap Y = y_j) = 0.05 + 0.10 + 0.35 = 0.50;$$
$$\mathbb{P}\{X = 40\} = \sum_{j=1}^{3} \mathbb{P}(X = 40 \cap Y = y_j) = 0 + 0.20 + 0.10 = 0.30.$$

Similarly, for $Y \in \{20, 30, 40\}$ we write

$$\mathbb{P}\{Y = 20\} = \sum_{i=1}^{3} \mathbb{P}(X = x_i \cap Y = 20) = 0.05 + 0.05 + 0 = 0.10;$$
$$\mathbb{P}\{Y = 30\} = \sum_{i=1}^{3} \mathbb{P}(X = x_i \cap Y = 30) = 0.05 + 0.10 + 0.20 = 0.35;$$
$$\mathbb{P}\{Y = 40\} = \sum_{i=1}^{3} \mathbb{P}(X = x_i \cap Y = 40) = 0.10 + 0.35 + 0.10 = 0.55.$$

Hence, the pmf of X is given by

$$\mathbb{P}\{X = x\} = \begin{cases} 0.20, & x = 20 \\ 0.50, & x = 30 \\ 0.30, & x = 40 \end{cases},$$

and the pmf of Y is given by

$$\mathbb{P}\{Y=y\} = \begin{cases} 0.10, & y=20 \\ 0.35, & y=30 \\ 0.55, & y=40 \end{cases}.$$

1.2

What is the expected total cost?

rubric={reasoning:3}

Solution

First, we calculate $\mathbb{E}(X)$:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^{3} x_i \cdot \mathbb{P}(X=x_i) \\ &= 20 \cdot 0.20 + 30 \cdot 0.50 + 40 \cdot 0.30 \\ \mathbb{E}(X) &= 31. \end{aligned}$$

Then we calculate $\mathbb{E}(Y)$:

$$\begin{aligned} \mathbb{E}(Y) &= \sum_{j=1}^{3} y_j \cdot \mathbb{P}(Y=y_j) \\ &= 20 \cdot 0.10 + 30 \cdot 0.35 + 40 \cdot 0.55 \\ \mathbb{E}(Y) &= 34.5. \end{aligned}$$

Hence, the expected total cost is

$$E(X+Y) = E(X) + E(Y) = 65.5.$$

1.3.

What is the probability that the parents's and the child's dinner cost at most \$30 each?

rubric={reasoning:3}

Solution

$$\begin{aligned} \mathbb{P}(X \leq 30 \cap Y \leq 30) &= \mathbb{P}(X=20 \cap Y=20) + \mathbb{P}(X=20 \cap Y=30) + \mathbb{P}(X=30 \cap Y=20) + \mathbb{P}(X=30 \cap Y=30) \\ &= 0.05 + 0.05 + 0.05 + 0.10 \\ \mathbb{P}(X \leq 30 \cap Y \leq 30) &= 0.25. \end{aligned}$$

1.4

Given that the child has ordered a \$40 dinner, what is the probability that the parent will order an \$40 dinner as well? rubric={reasoning:3}

Solution

$$\mathbb{P}(X=40|Y=40)=\frac{\mathbb{P}(X=40\cap Y=40)}{\mathbb{P}(Y=40)}=\frac{0.10}{0.55}\simeq 0.18.$$

1.5

Are X and Y independent? Justify your answer.

rubric={reasoning:3}

Solution

X and Y are independent events if $\mathbb{P}(X\cap Y)=\mathbb{P}(X)\cdot\mathbb{P}(Y)$ is satisfied, which only occurs if $\mathbb{P}(X\cap Y)-\mathbb{P}(X)\cdot\mathbb{P}(Y)=0$. Testing for a few outcomes:

$$\begin{aligned}\mathbb{P}(X=20\cap Y=20)-\mathbb{P}(X=20)\cdot\mathbb{P}(Y=20)=0.05-0.20\cdot 0.10\neq 0\\ \mathbb{P}(X=20\cap Y=40)-\mathbb{P}(X=20)\cdot\mathbb{P}(Y=40)=0.10-0.20\cdot 0.55\neq 0.\end{aligned}$$

X and Y are not independent, because the joint probabilities do not equal the product of the marginals.

1.6

Find the correlation coefficient between X and Y.

rubric={reasoning:3}

Solution

First, we calculate $\mathbb{E}(XY)$:

$$\begin{array}{rcc}\mathbb{E}(XY) & = & \sum_{i=1}^{3}\sum_{j=1}^{3}x_i\cdot y_j\cdot\mathbb{P}(X=x_i\cap Y=y_j)\\ \\ & = & 20\,(20\cdot 0.05+30\cdot 0.05+40\cdot 0.10)\\ & + & 30\,(20\cdot 0.05+30\cdot 0.10+40\cdot 0.35)\\ & + & 40\,(20\cdot 0.00+30\cdot 0.20+40\cdot 0.10)\\ \\ \mathbb{E}(XY) & = & 1070.\end{array}$$

Using $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ from 1.2, we calculate $\mathrm{Cov}(X,Y)$:

$$\begin{array}{rcc}\mathrm{Cov}(X,Y) & = & \mathbb{E}(XY)-\mathbb{E}(X)\cdot\mathbb{E}(Y)\\ & = & 1070-31\cdot 34.5\\ \mathrm{Cov}(X,Y) & = & 0.5.\end{array}$$

Now, calculating $\mathbb{E}(X^2)$

$$\begin{array}{rcc}\mathbb{E}(X^2) & = & \sum_{i=1}^{3}x_i^2\cdot\mathbb{P}(X=x_i)\\ & = & 20^2\cdot 0.20+30^2\cdot 0.50+40^2\cdot 0.30\\ \mathbb{E}(X^2) & = & 1010\end{array}$$

and $\mathbb{E}(Y^2)$

$$\begin{aligned}\mathbb{E}(Y^2) &= \sum_{j=1}^{3} y_j^2 \cdot \mathbb{P}(Y = y_j) \\ &= 20^2 \cdot 0.10 + 30^2 \cdot 0.35 + 40^2 \cdot 0.55 \\ \mathbb{E}(Y^2) &= 1235,\end{aligned}$$

we have

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 1010 - 31^2 = 49; \\ \mathrm{Var}(Y) &= \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = 1235 - 34.5^2 = 44.75.\end{aligned}$$

Finally, the correlation is given by

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{0.5}{\sqrt{49 \cdot 44.75}} \simeq 0.01.$$

Exercise 2

2.1

A recent study suggested that 65% of all eligible voters will vote in the next federal election. Suppose 20 eligible voters were randomly selected from the population of all eligible voters and let X be the number of the 20 eligible who will vote. Is X a binomial random variable? Justify your answer.

rubric={reasoning:3}

Solution

X is a binomial random variable, because we're summing an unordered sample of binary random variables ("will vote" or "will not vote") from a sufficiently large population without replacement. Let x_i be defined by

$$x_i = \begin{cases} 1 & \text{individual } i \text{ votes} \\ 0 & \text{otherwise} \end{cases}$$

For each i, the probability that we select an individual who votes (*i.e.*, $x_i = 1$), uniformly at random from the population, is estimated by our survey to be $p = 0.65$. We select each of our 20 individuals independently of the others. Hence, $X := \sum_i x_i$ is a sum of $n = 20$ iid Bernoulli random variables, so it is distributed as $\mathrm{Bin}(n, p)$.

Notes

1. it is **untrue** that each individual has a 0.65 probability of voting (compare to real life). In fact, the question seems to imply that each individual either **will vote** (votes with probability 1) or **will not vote** (votes with probability 0). It is the fact that we select them uniformly at random from a large population that allows us to conclude that on average for each selection we will choose a voting individual with probability 0.65.
2. it is **untrue** that individuals' voting preferences are independent (*e.g.*, eligible voters who are attending university are more likely to vote than a member of the general population). The selections are independent because they are made independently [uniformly at random] from the total population.

2.2

Explain the difference between the Binomial random variable and the Negative Binomial random variable.

rubric={reasoning:3}

Solution

A binomial distribution $\text{Bin}(n, p)$ gives the probabilities of "winning" k out of n (iid) games when the probability of winning each game is p. Therefore, a binomial random variable gives the number of times one has "won" out of n trials.

A negative binomial distribution $\text{NB}(r, p)$ gives the distribution of the number of trials in a sequence of iid Bernoulli trials before r events occur. Therefore, a negative binomial random variable is a number of times that one has "lost" before having "won" r times, or another way, the number of "failures" until the rth success.

Accordingly, the difference between the distributions is whether n trials have elapsed or r events have occurred. In some sense, then, negative binomial tells us "how long" we should expect to wait in order to attain r events, while a binomial tells us how many events occur for a *fixed* number of trials.