TRINITY COLLGE DUBLIN

School of Computer Science and Statistics

Extra Questions ST3009: Statistical Methods for Computer Science

NOTE: There are many more example questions in Chapter 4 of the course textbook "A First Course in Probability" by Sheldon Ross, and also some in Chapters 6 and 7 (but ignore the questions involving continuous random variables).

Question 1. Let X denote a random variable that takes on any of the values -1, 0, and 1 with respective probabilities $P(X = -1) = 0.2$, $P(X = 0) = 0.5$, $P(X = 1) = 0.3$. Compute $E[X]$ and $E[X^2]$.

Solution

- $E[X] = -1 \times 0.2 + 0 \times 0.5 + 1 \times 0.3 = 0.1$
- $E[X^2] = (-1)^2 \times 0.2 + 0^2 \times 0.5 + 1^2 \times 0.3 = 0.5$

Question 2. There are 6 cards in a hat: $1\heartsuit$, $1\spadesuit$, $1\diamondsuit$, $2\heartsuit$, $2\spadesuit$, $2\clubsuit$. You draw one card uniformly at random. If its suit is $\diamondsuit$ then you draw one more card, otherwise you stop. Let X be the sum of the ranks on the one or two cards drawn. What is the probability that one card is drawn ? What is the probability that two cards are drawn ? Find the PMF of X and $E[X]$.

Solution

- Probability of one card is probability that do not draw $1\diamondsuit$. That is, five possible cards out of six so the probability is 5/6.
- Probability of two cards is probability that draw $1\diamondsuit$. The probability is 1/6.
- Let E be the event that one card is drawn.
  - $P(X = 1) = P(X = 1|E)P(E) + P(X = 1|E^c)P(E^c)$. $P(E) = 5/6$, $P(E^c) = 1/6$. When one card is drawn then $X = 1$ when $1\heartsuit$, $1\spadesuit$ are drawn i.e. $P(X = 1|E) = 2/5$ (given one card is drawn, there are five possible cards i.e excluding $1\diamondsuit$). When two cards are drawn then the first card must be $1\diamondsuit$ and so $X > 1$ i.e. $P(X = 1|E^c) = 0$. So $P(X = 1) = 2/5 \times 5/6 = 0.3333$.
  - $P(X = 2) = P(X = 2|E)P(E) + P(X = 2|E^c)P(E^c)$. When one card is drawn $X = 2$ when $2\heartsuit$, $2\spadesuit$, $2\clubsuit$ are drawn i.e. with probability 3/5. When two cards are drawn then $X = 2$ with probability $P(X = 1|E^c) = 2/5$ (since $1\diamondsuit$ since that's already been drawn). So $P(X = 2) = 3/5 \times 5/6 + 2/5 \times 1/6 = 0.5667$
  - $P(X = 3) = P(X = 3|E)P(E) + P(X = 3|E^c)P(E^c)$. When one card is drawn $X = 3$ cannot happen so $P(X = 3|E) = 0$. When two cards are drawn then $X = 3$ when the second card is $2\heartsuit$, $2\spadesuit$, $2\clubsuit$ i.e. with probability $P(X = 1|E^c) = 3/5$. So $P(X = 3) = 3/5 \times 1/6 = 0.1$
- Therefore $E[X] = 1 \times 0.3333 + 2 \times 0.5667 + 3 \times 0.1$.

Question 3. Consider the following game. A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears i times, $i = 1, 2, 3$, then the player wins i euros; if the number bet by the player does not appear on any of the dice, then the player loses €1. What is the expected win/loss for the player ?

Solution If we assume that the dice are fair and act independently of each other, then the number of times that the number bet appears is a binomial random variable with parameters (3,1/6). Hence, letting X denote the players winnings in the game, we have:

$$P(X=-1) = \binom{3}{0}(1/6)^0(5/6)^3 = 125/216$$
$$P(X=1) = \binom{3}{1}(1/6)^1(5/6)^2 = 75/216$$
$$P(X=2) = \binom{3}{2}(1/6)^2(5/6)^1 = 15/216$$
$$P(X=3) = \binom{3}{3}(1/6)^3(5/6)^0 = 1/216$$

and so $E[X] = -1 \times 125/216 + 2 \times 75/216 + 2 \times 15/216 + 3 \times 1/216 = -17/216$

Question 4. Find the expected total number of successes that result from n trials when trial i is a success with probability p_i, $i = 1, \dots, n$.

Solution Let $X_i = 1$ is trial i is a success and 0 otherwise. The total number of successes is $X = \sum_{i=1}^n X_i$ and the expected total number of successes is $E[X] = E[\sum_{i=1}^n X_i] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p_i$.

Question 5. 100 snails are competing in a relay race. Each snail takes on average 10 minutes to complete the race course. What is the expected time for the race to complete (i.e. for all 100 snails to complete the course, one after another) ? Explain your reasoning.

Solution $100 \times 10 = 1000$ minutes. Using the linearity of expectation.

Question 6. A bag contains 100 chips, 75 chips with 1 marked on them and 25 with -1 marked on them. Pick a chip from the bag and let random variable X be the value of the chip. What is $E[X]$? What is $Var(X)$?

Suppose now that you draw 50 chips from the bag, with replacement. Let Y be the sum of values of the chips drawn. What is $E[Y]$? Hint: use linearity of the expectation.

What is $Var(Y)$? Hint: use the fact that the chips are drawn independently.

What is $E[\frac{1}{50}Y]$ and $Var(\frac{1}{50}Y]$? Hint: linearity of the expectation again, and that $Var(aX) = a^2 Var(X)$.

Solution

- $P(X=1) = 75/100$, $P(X=-1) = 25/100$ so $E[X] = 1 \times 75/100 - 1 \times 25/100 = 0.5$. Also $E[X^2] = 1^2 \times 75/100 + (-1)^2 \times 25/100 = 1$ and so $Var(X) = E[X^2] - E[X]^2 = 1 - 0.5^2 = 0.75$.

- $Y = \sum_{i=1}^{50} X_i$ where X_i is the value of the ith chip drawn. $E[Y] = E[\sum_{i=1}^{50} X_i] = \sum_{i=1}^{50} E[X_i] = \sum_{i=1}^{50} 0.5 = 25$.

- $Var(Y) = Var(\sum_{i=1}^{50} X_i) = \sum_{i=1}^{50} Var(X_i) = \sum_{i=1}^{50} 0.75 = 37.5$.

- Using the linearity of the expectation once again, $E[\frac{1}{50}Y] = \frac{1}{50}E[Y] = \frac{25}{50} = 0.5$. For the variance, $Var(\frac{1}{50}Y) = \frac{1}{50^2}Var(Y) = \frac{37.5}{50^2} = 0.015$.

Question 7. You select one student uniformly at random and let random variable $X = 1$ if they have brown hair and $X = 0$ otherwise. Suppose there are 100 students of whom 40 have brown hair. What is $Prob(X = 1)$? What is $E[X]$? Are these the same or different, and why ?

You now carry out a poll of students in the class to try to estimate the number with brown hair. You do this by selecting n students uniformly at random and checking their hair colour, letting $X_i = 1$ if it is brown for student i. Let $Y = \sum_{i=1}^{n} X_i$. What is $E[Y]$? Is it the same as $E[X]$ or different ? What is $E[\frac{1}{n}Y]$? What is the variance of $\frac{1}{n}Y$ (express in terms of $Var(X)$) ? Hints: use linearity of the expectation and the fact that students are sampled independently.

Solution

- $P(X = 1) = 40/100$. $E[X] = 1 \times P(X = 1) + 0 \times P(X = 0) = P(X = 1) = 40/100 = 0.4$. Since this is an indicator random variable, we always have $E[X] = P(X = 1)$.

- $E[Y] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} 0.4 = 0.4n$ using linearity of the expectation (so $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$) and the fact that students are sampled independently (so $E[X_i] = E[X]$ for all i). $E[\frac{1}{n}Y] = \frac{1}{n}E[Y] = 0.4$. This is the same as $E[X]$.

- $Var(\frac{1}{n}Y) = \frac{1}{n^2}Var(Y) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i)$. Since the students are sampled independently, $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = nVar(X)$. Hence, $Var(\frac{1}{n}Y) = \frac{1}{n^2}Var(X)$.

Question 8. Compute the variance of a binomial random variable X with parameters n and p.

Solutionm variable The binomial random variable $X = X_1 + X_2 + \cdots + X_n$ is the sum of n independent bernoulli random variables X_i, with $P(X_i = 1) = p$. Now $Var(X_i) = E[X_i^2] - E[X_i]^2 = p - p^2 = p(1 - p)$. Therefore $Var(X) = Var(X_1 + X_2 + \cdots + X_n) = np(1 - p)$.

Question 9. Suppose that the son of a man of height x (in inches) attains a height that randomly distributed with mean $x + 10$ and variance 4. What is the expected height at full growth of the son of a man who is 170cm tall?

Solution Let X be the height of the man and Y be the height of the son. Then $E[Y|X = x] = x + 10$ and so $E[Y|X = 170] = 180$.

Question 10. Suppose that N people throw their hats into the center of a room. The hats are mixed up, and each person randomly selects one. Find the expected number of people that select their own hat.

Solution Letting Xdenote the number of matches, $X = X_1 + X_2 + \cdots + X_N$ where $X_i = 1$ if person i selects his own hat and 0 otherwise. Since for each i the ith person is equally likely to select any of the N hats, $E[X_i] = P(X_i = 1) = 1/N$ and so $E[X] = E[X_1] + E[X_2] + \cdots + E[X_N] = N \times 1/N = 1$. So on average exactly one person selects his own hat.

Question 11. I toss a coin. If it comes up heads I then throw a six-sided die and win the amount that comes up in euros. If tails then I pay €1. If I play the game once what is my expected win/loss ?

I repeat the game 10 times. What is my expected win/loss ?

Solution

- Let X be the amount I win and C the event that the coin comes up heads. Then $E[X|C] = 1\times1/6+2\times1/6+3\times1/6+4\times1/6+5\times1/6+6\times1/6 = 3.5$ and $E[X|C^c] = -1$. And so $E[X] = E[X|C]P(C) + E[X|C^c]P(C^c) = 3.5 \times 1/2 - 1 \times 1/2 = 1.25$ (using marginalisation of the conditional probability).

- Repeating 10 times, the expected win/loss is $10 \times 1.25 = 12.5$ (using linearity of the expectation).

Question 12. When I travel to work by bike it takes on average 20 minutes. When I take the bus it takes on average 45 minutes. I take the bus when it rains, and on any given day suppose it rains with probability 0.25. What is my overall expected travel time to work ? Hint: use marginalisation of conditional expectation.

Solution Let T be the time taken to travel to work, C be the event that I cycled and B that I took the bus. Then $E[T|C] = 20$, $E[T|B] = 45$ and so $E[T] = E[T|C]P(C) + E[T|B]P(B) = 20 \times (1 - 0.25) + 45 \times 0.25$.

Question 13. A product that is sold seasonally yields a net profit of 5 dollars for each unit sold and a net loss of 1 dollar for each unit left unsold when the season ends. The number of units of the product that could be sold at a specific department store during any season is a random variable having probability mass function $p(i)$, $i = 0, 1, 2, \ldots$. If the store must stock this product in advance, what is the expected profit (express in terms of $p(i)$ and the number n of units stocked) ? Suppose $p(i) = 1/10$ for $i = 1, \ldots, 10$ and $p(i) = 0$ for $i > 10$. Using matlab plot the expected profit vs n and determine the number of units n the store should stock so as to maximize its expected profit.

Solution

- Let n be the number stocked, X be the number of units that could be sold and M the profit. The profit/loss is $M = 5X - (n - X) = 6X - n$ when $n \geq X$, and $5n$ when $X > n$ (the sales cannot exceed n). The expected profit is therefore $E[M] = \sum_{i=1}^{n}(6i - n)p(i) + 5n\sum_{i=n+1}^{\infty} p(i)$.
- For $n \leq 10$, $E[M] = \sum_{i=1}^{n}(6i - n)/10 + 5n\sum_{i=n+1}^{10} 1/10$. The maximum is when $n = 10$.

Question 14. I am listening to new songs on spotify. Suppose the probability that I don't like each suggested song is 0.75, independently for each song. On average how many songs do I have to listen to before I find one that I like.

Solution Let X be the number of songs that I need to listen to before finding one like. $P(X = 1) = 0.25$, $P(X = 2) = 0.25(1 - 0.25)$, $P(X = 3) = 0.25(1 - 0.25)^2$ and so on. Therefore, $E[X] = 1 \times 0.25 + 2 \times 0.25(1 - 0.25) + 3 \times 0.25(1 - 0.25)^2 + \cdots = \sum_{i=1}^{\infty} i \times 0.25 \times (1 - 0.25)^{i-1}$.

Question 15. Suppose that the number of people entering a department store on a given day is a random variable with mean 50. Suppose further that the amounts of money spent by these customers are independent random variables having a common mean of €9. Finally, suppose also that the amount of money spent by a customer is also independent of the total number of customers who enter the store. What is the expected amount of money spent in the store on a given day?

Solution If we let N denote the number of customers that enter the store and X_i the amount spent by the ith such customer, then the total amount of money spent can be expressed as $\sum_{i=1}^{N} X_i$. Now $E[\sum_{i=1}^{N} X_i | N = n] = \sum_{i=1}^{n} E[X_i] = 9n$. And

$$
\begin{aligned}
E[\sum_{i=1}^{N} X_i] &= E[\sum_{i=1}^{N} X_i | N = 1]P(N = 1) + E[\sum_{i=1}^{N} X_i | N = 2]P(N = 2) + E[\sum_{i=1}^{N} X_i | N = 3]P(N = 3) + \ldots \\
&= 9(1 \times P(N = 1) + 2 \times P(N = 2) + 3 \times P(N = 3) + \ldots) \\
&= 9P(N) = 9 \times 50 = 450.
\end{aligned}
$$

Question 16. Suppose X and Y are random variables with the following joint pmf. Are X and Y independent?

X/Y	1	2	3
1	1/18	1/9	1/6
2	1/9	1/6	1/18
3	1/6	1/18	1/9

Solution From the table we compute the marginal probabilities $P(X = 1) = 1/3$, $P(Y = 1) = 1/3$. Since $P(X = 1 \text{ and } Y = 1) = 1/18$ and $P(X = 1)P(Y = 1) = 1/9$ X and Y are not independent.

Question 17. Let I_A and I_B be indicator variables for the events A and B. Express $Cov(I_A, I_B)$ in terms of $P(A)$, $P(B)$ and $P(A \cap B)$.

Solution $Cov(I_A, I_B) = E[I_A I_B] - E[I_A]E[I_B] = P(A \cap B) - P(A)P(B) = P(B)(P(A|B) - P(A))$.

Question 18. Suppose X and Y are random variables with $P(X = -1) = 1/2$, $P(X = 1) = 1/2$, $P(Y = -1) = 1/2$, $P(Y = 1) = 1/2$. Let $c = P(X = 1 \text{ and } Y = 1)$. Determine the joint PMF of X and Y. Also $Cov(X, Y)$ and $Corr(X, Y)$.

For what values of c are X and Y independent ? For what values of c are X and Y 100% correlated ?

Solution

- The joint PMF is:

Y/X	1	-1
1	c	0.5-c
-1	0.5-c	c

- $E[X] = 0$, $E[Y] = 0$, $Var(X) = 1$, $Var(Y = 1)$. So $E[XY] = 1 \times 1 \times c + (-1) \times 1 \times (0.5 - c) + 1 \times (-1) \times (0.5 - c) + (-1) \times (-1) \times c = 4c - 1$. $Cov(XY) = E[XY] - E[X]E[Y] = E[XY] = 4c - 1$. $Corr(X, Y) = Cov(X, Y)/\sqrt{Var(X)Var(Y)} = 4c - 1$.

- We must have $Cov(X, Y) = 0$ for X and Y to be independent, and so $c = 1/4$. It is easy to check in this case that all four probabilities in the table are 0.5 and they are independent.

- When $c = 0.5$ the correlation is 1 which means that X and Y are 100% correlated.

Question 19. A computer game draws a rectangle with random length X and breadth Y, with $E[X] = 1$ and $E[Y] = 2$. The length and breadth are selected independently. The area of the rectangle is XY. What is the expected area of the rectangle ?

Solution We use the fact that $E[XY] = E[X]E[Y]$ when X and Y are independent. The area of a rectangle is XY and so the expected area is $E[XY] = E[X]E[Y] = 1 \times 2$.