Part 1 Quiz 1 solutions

February 28, 2025

1 Solutions/answers

Question 1 A continuous random variable can take infinitely many possible values.

solution TRUE
Takes real values.

Question 2 The values of a probability density function (PDF) of a continuous random variable X on a point x, gives the exact probability of the X taking the value x.

solution FALSE
The function value represents probability density. The probability of any single point is 0 regardless of the function value.

Question 3 For a discrete random variable, the probability mass function (PMF) can take negative values.

solution FALSE
The values of the PMF are probabilities, which are always positive.

Question 4 company produces a certain type of electronic component, and the number of defective components in a batch of 5, represented by the discrete random variable X, follows the probability distribution below:

X (Defective Components)	0	1	2	3	4	5
$P(X)$ (Probability)	0.1	0.2	0.3	0.25	0.1	?

- The probability of all 5 components being defective is [A]
- The expected number of defective components in a batch is [B]

solution Since all probabilities must sum to 1 (i.e $\sum_{i=0}^{5} P(X = i) = 1$), we have that:

$$A = P(X = 5) = 1 - 0.1 - 0.2 - 0.3 - 0.25 - 0.1 = 0.05$$

For the expected value we have:

$$B = E[X] = \sum_{i=0}^{5} i \cdot P(X = i) = 0 \cdot 0.1 + 1 \cdot 0.2 + 2 \cdot 0.3 + 3 \cdot 0.25 + 4 \cdot 0.1 + 5 \cdot 0.05 = 2.2$$

Question 5 In a town of [a] residents, [b] of them like coffee, [c] like tea, and 50 of them like both coffee and tea. Suppose a survey is conducted where 50 residents are randomly selected (you may assume this selection is with replacement), and the goal is to determine how many of them like at least one of these two drinks (coffee or tea). Let X be the random variable representing the number of residents in the sample who like at least one of coffee or tea.

What is the expectation of the random variable X?

Hint: First compute the probability that a randomly selected resident likes either coffee or tea, i.e., calculate $P(A \cap B)$, where:

- A: The event that a resident likes coffee.
- B: The event that a resident likes tea.

solution We first compute the probability that someone likes either coffee or tea as follows:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{b}{a} + \frac{c}{a} + \frac{50}{a} = \frac{b + c - 50}{a}$$

Since the sampling is with replacement, X follows the binomial distribution $Bin(50, \frac{b+c-50}{a})$ and has expectation :

$$E[X] = 50 \cdot n \cdot p = 50 \cdot \frac{b + c - 50}{a}$$

Question 6 Consider the random variable X with the following probability mass function (pmf):

X	1	2	3	4	5	6
P(X)	0.1	0.2	0.3	0.2	0.1	0.1

The expected value E(X) of the random variable X is [A]
The variance Var(X) of the random variable X is [B]

solution The expected value of X can be computed as follows:

$$A = E[X] = \sum_{i=1}^{6} i \cdot P(X = i) = 1 \cdot 0.1 + 2 \cdot 0.2 + 3 \cdot 0.3 + 4 \cdot 0.2 + 5 \cdot 0.1 + 6 \cdot 0.1 = 3.3$$

For the variance we can use the formula:

$$B = Var(X) = E[X^2] - (E[X])^2$$

For the first term we get:

$$E[X^2] = \sum_{i=1}^{6} i^2 \cdot P(X = i) = 1^2 \cdot 0.1 + 2^2 \cdot 0.2 + 3^2 \cdot 0.3 + 4^2 \cdot 0.2 + 5^2 \cdot 0.1 + 6^2 \cdot 0.1$$

$$E[X^2] = 0.1 + 0.8 + 2.7 + 3.2 + 2.5 + 3.6 = 12.9$$

Thus,

$$B = Var(X) = E[X^2] - (E[X])^2 = 12.9 - 3.3^2 = 2.01$$

Question 7 A factory produces electronic components with a lifetime modeled by the random variable X. The expected lifetime of a component is 5 hours, and the standard deviation is 2 hours. The factory introduces a new testing process that has been found to multiplies the lifetime by 3 and subtract 2 hours. Let Y=3X-2 represent the adjusted lifetime.

The expected adjusted lifetime Y in [A]

The variance Var(Y) of the adjusted lifetime is [B]

solution From the fact that $Y = 3X - 2$, we get that:

$$A = E[Y] = 3 \cdot E[X] - 2 = 3 \cdot 5 - 2 = 13$$

For the variance, we get:

$$B = Var(X) = 3^2 \cdot Var(X) = 9 \cdot \sigma_X^2 = 9 \cdot 2^2 = 36$$

Question 8 The cumulative distribution function (CDF) is always non-decreasing for both discrete and continuous random variables.

solution TRUE
For $x_1 < x_2$, we have: $P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2) \geq P(X \leq x_1)$ regardless if X is discrete or continuous.

Question 9 Two independent Poisson processes, X_1 and X_2, model the number of customer arrivals at two different service desks. The first service desk receives an average of 1 customers per hour, and the second service desk receives an average of 1.5 customers per hour. Let X_1 and X_2 be the number of customers arriving at the first and second service desks, respectively, during a 2-hour period.

The probability that at least 1 customer arrives at each of the service desks ($X_1 \geq 1$ and $X_2 \geq 1$) in a 2-hour period is [A]

The expected total number of customers arriving at both service desks in a 2-hour period is [B]

The variance of the total number of customers arriving at both service desks in a 2-hour period is [C]

solution Due to the independence of the two processes, we have the following:

$$P(X_1 \geq 1 \text{ and } X_2 \geq 1) = P(X_1 \geq 1) \cdot P(X_2 \geq 1) = (1 - P(X_1 = 0)) \cdot (1 - P(X_2 = 0))$$

$$P(X_1 \geq 1 and X_2 \geq 1) = (1 - e^{-1.5 \cdot 2}) \cdot (1 - 1 - e^{-1 \cdot 2}) = (1 - e^{-3}) \cdot (1 - e^{-2}) = 0.821$$

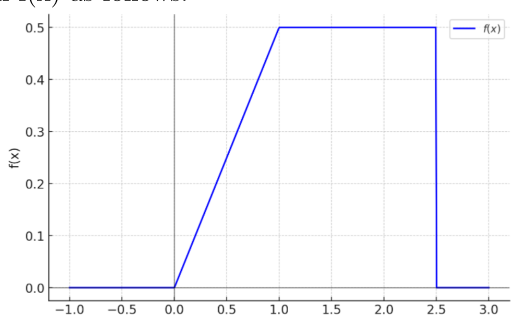
The expected number of customers at the first service desk is $E[X_1] = 1.5 \cdot 2 = 3$ and at the second service desk, it is $E[X_2] = 1 \cdot 2 = 2$. Thus,

$$E[X_1 + X_2] = E[X_1] + E[X_2] = 3 + 2 = 5$$

The variance of a Poisson process has the same numerical value as the expectation. Thus,

$$Var(X_1 + X_2) = 5$$

Question 10 A continuous random variable X, has probability density function f(x) as follows:



$$f(x) = \begin{cases} x/2, & \text{for } 0 < x < 1 \\ 1/2, & \text{for } 1 < x < 2.5 \\ 0, & \text{otherwise} \end{cases}$$

Compute the median (i.e 50th percentile) of the distribution.

solution The cumulative distribution $F(x)$ is equal to the area below the graph of the distribution between 0 and x (since the density is 0 for any $x < 0$). From the graph of the distribution, we see that this area can be computed as follows:

$$F(x) = \begin{cases} \frac{1}{2} \cdot x \cdot x/2, & \text{for } 0 < x < 1 \\ \frac{1}{2} \cdot 1 \cdot \frac{1}{2} + (x-1) \cdot \frac{1}{2}, & \text{for } 1 < x < 2.5 \\ 1, & \text{x¿2.5} \end{cases}$$

After simplifying:

$$F(x) = \begin{cases} \frac{x^2}{4}, & \text{for } 0 < x < 1 \\ \frac{1}{4} + \frac{x-1}{2}, & \text{for } 1 < x < 2.5 \\ 1, & x > 2.5 \end{cases}$$

To find the median, we need to find a value P_{50}, such that: $F(P_50) = \frac{1}{2}$. Thus, we have:

$$\frac{1}{4} + \frac{P_{50} - 1}{2} = \frac{1}{2}$$
$$P_{50} = 2 \cdot \left(\frac{1}{2} - \frac{1}{4}\right) + 1 = 1.5$$

Question 11 A continuous random variable X has cummulative distribution function (CDF) F(x) as follows:

$$F(x) = \begin{cases} 0, & \text{for } x < 1 \\ 1 - \frac{1}{x^4}, & \text{for } x \geq 1 \end{cases}$$

The expectation of X is: [A]
The variance of X is: [B]

solution We first compute the probability density function as follows:

$$f(x) = \frac{d}{dx}F(x) = \begin{cases} 0, & \text{for } x < 1 \\ -(-4) \cdot x^{-5}, & \text{for } x \geq 1 \end{cases}$$

$$f(x) = \begin{cases} 0, & \text{for } x < 1 \\ \frac{4}{x^5}, & \text{for } x \geq 1 \end{cases}$$

The expectation of X can be computed as follows:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_{0}^{\infty} 4 \cdot x^{-4}dx = 4 \cdot \left[-\frac{1}{3} \cdot x^{-3}\right]_0^{\infty} = \frac{4}{3}$$

The variance of X can be computed as follows:

$$Var(X) = E[X^2] - (E[X])^2$$

We first compute:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_{0}^{\infty} 4 \cdot x^{-3}dx = 4 \cdot \left[-\frac{1}{2} \cdot x^{-2}\right]_0^{\infty} = 2$$

Thus, we get:

$$Var(X) = E[X^2] - (E[X])^2 = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9} = 0.222$$

Question 12 A quality control inspector is testing two independent machines. Each machine produces defective items with probabilities p_1=[a] and p_2=[b], respectively. The inspector counts the number of items produced by each machine until the first defective item is found. Let X_1 and X_2 be the number of items produced by Machine 1 and Machine 2 until their first defect, where X_1 and X_2 are independent and follow the corresponding geometric distributions: $X_1 \sim G(p_1)$ and $X_2 \sim G(p_2)$.

Let $T = X_1 + X_2$ be the total number of items inspected until both machines have produced their first defect.

What is the variance of T?

solution Since the distributions are independent, we know that:

$$Var(T) = Var(X_1 + X_2) = Var(X_1) + Var(X_2) = \frac{1 - p_1}{p_1^2} + \frac{1 - p_2}{p_2^2}$$

Question 13 A bakery sells a special kind of pastry, and each customer who enters has a probability p=0.01 of buying it. On a busy day, 200 customers visit the bakery. Let X be the number of pastries sold that day.

Use the Poisson approximation to estimate the probability that at least 2 of the pastries are sold.

solution The random variable X follows the Binomial distribution: $X \sim Bin(200, 0.01)$, which has mean: $E[X] = 200 \cdot 0.01 = 2$.

Due to the Poisson approximation, X approximately follows the distribution $Poi(2)$. That is,

$$P(X = k) \simeq \frac{e^{-2} \cdot 2^k}{k!}$$

Thus, we get:

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-2} - 2 \cdot e^{-2} = 1 - 3 \cdot e^{-2} = 0.594$$

Question 14 A university cafeteria serves lunch to 600 students daily. Historically, 40% of students choose the vegetarian option. Let X be the number of students who choose the vegetarian option on a given day.

Use the normal approximation to the binomial distribution to estimate the probability that between 230 and 250 students (inclusive) choose the vegetarian option.

Apply a continuity correction in your calculation.

You are required to use closest value in the following standard normal distribution table below for $P(Z < z)$, $0 \le z \le 2.08$ to determine your answer.

z	0.00	0.02	0.04	0.06	0.08
0.00	0.5000	0.5080	0.5160	0.5239	0.5319
0.20	0.5793	0.5871	0.5948	0.6026	0.6103
0.40	0.6554	0.6628	0.6700	0.6772	0.6844
0.60	0.7257	0.7324	0.7389	0.7454	0.7517
0.80	0.7881	0.7939	0.7995	0.8051	0.8106
1.00	0.8413	0.8461	0.8508	0.8554	0.8599
1.20	0.8849	0.8888	0.8925	0.8962	0.8997
1.40	0.9192	0.9222	0.9251	0.9279	0.9306
1.60	0.9452	0.9474	0.9495	0.9515	0.9535
1.80	0.9641	0.9656	0.9671	0.9686	0.9699
2.00	0.9772	0.9783	0.9793	0.9803	0.9812

solution The random variable X follows the Binomial distribution: $X \sim Bin(600, 0.4)$, which has mean $E[X] = 600 \cdot 0.4 = 240$, and variance $Var(X) = \sigma_x^2 = 600 \cdot 0.4 \cdot (1 - 0.4) = 600 \cdot 0.24 = 144$.

Thus, $\sigma_X = 12$.

Using the normal approximation, we get that X can be approximated by: $X \simeq Y \sim N(240, 12^2)$.

Therefore, after taking the continuity correction into account, we get:

$$P(230 \le X \le 250) \simeq P(229.5 < Y < 250.5) = P\left(\frac{229.5 - \mu_Y}{\sigma_Y} < Z < \frac{250.5 - \mu_Y}{\sigma_Y}\right)$$

$$P(230 \le X \le 250) \simeq P\left(\frac{229.5 - \mu_Y}{\sigma_Y} < Z < \frac{250.5 - \mu_Y}{\sigma_Y}\right) = P(-0.875 < Z < 0.875)$$

for some standard normal random variable $Z \sim N(0, 1)$. Thus,

$$P(230 \le X \le 250) \simeq P(Z < 0.833) - P(Z < -0.875) = P(Z < 0.875) - (1 - P(Z < -0.875))$$

$$P(230 \le X \le 250) \simeq 2 \cdot P(Z < -0.875) - 1 = 2 \cdot 0.8106 - 1 = 0.6212$$

We used the yellow highlighted entry of the table as shown below:

z	0.00	0.02	0.04	0.06	0.08
0.00	0.5000	0.5080	0.5160	0.5239	0.5319
0.20	0.5793	0.5871	0.5948	0.6026	0.6103
0.40	0.6554	0.6628	0.6700	0.6772	0.6844
0.60	0.7257	0.7324	0.7389	0.7454	0.7517
0.80	0.7881	0.7939	0.7995	0.8051	0.8106
1.00	0.8413	0.8461	0.8508	0.8554	0.8599
1.20	0.8849	0.8888	0.8925	0.8962	0.8997
1.40	0.9192	0.9222	0.9251	0.9279	0.9306
1.60	0.9452	0.9474	0.9495	0.9515	0.9535
1.80	0.9641	0.9656	0.9671	0.9686	0.9699
2.00	0.9772	0.9783	0.9793	0.9803	0.9812

Question 15 For each of the following probability distributions, fill in the blanks:

(1) Binomial distribution: $X \sim B(50, 0.1)$, the variance of X is [A].
(2) Uniform distribution: $X \sim U(15, 27)$, the variance of X is [B].
(3) Geometric distribution: $X \sim G(0.2)$, the expected value of X is [C].

solution

1. The variance of the binomial distribution is:

$$Var(X) = np(1-p) = 50 \cdot 0.1 \cdot (1 - 0.1) = 5 \cdot 0.9 = 4.5$$

2. The variance of the uniform distribution is:

$$Var(X) = \frac{(B-A)^2}{12} = \frac{(27-15)^2}{12} = 12$$

3. The expected value of the Geometric distribution is:

$$E[X] = \frac{1}{\lambda} = \frac{1}{0.2} = 5$$

Question 16 A factory produces metal rods whose lengths are uniformly distributed between 90 cm and 110 cm. A construction project requires rods that are at least 95 cm but no longer than 105 cm.

Let X represent the length of a randomly selected rod.

Given that a rod is longer than [a] cm, what is the probability that it still meets the project's length requirements?

(Hint: Use conditional probability)

solution Let $L \sim U([90, 110])$ be a random variable following the Uniform distribution and representing the length of the rod. We want to compute the following conditional probability:

$$P(95 \le L \le 105 | L > a) = \frac{P(95 \le L \le 105 \text{ and } L > a)}{P(L > a)} = \frac{P(\max\{95, a\} \le L \le 105)}{P(L > a)}$$

We distinguish 3 cases as follows (assuming $a < 110$, since it can't be $L > a$ otherwise):

1. $a < 90$
   In this case, we have $P(L > a) = 1$ and
   $$P(95 \le L \le 105 | L > a) = P(95 \le L \le 105) = 10/20 = 0.5$$
2. $90 \le a < 95$
   In this case, we have $P(L > a) = \frac{110-a}{20}$ and
   $$P(95 \le L \le 105 | L > a) = \frac{P(95 \le L \le 105)}{P(L > a)} = \frac{\frac{10}{20}}{\frac{110-a}{20}} = \frac{10}{110-a}$$
3. $a \ge 95$
   In this case, we have $P(L > a) = \frac{110-a}{20}$ and
   $$P(95 \le L \le 105 | L > a) = \frac{P(a \le L \le 105)}{P(L > a)} = \frac{\frac{105-a}{20}}{\frac{110-a}{20}} = \frac{105-a}{110-a}$$

Question 17 A box contains an unlimited supply of red, yellow, and green balls. Each time you randomly draw a ball, the probability of getting a red ball is p_1=[a], the probability of getting a yellow ball is p_2=[b], and the draws are independent (with replacement).

Let X be the total number of draws until you get the first red ball that comes later than the first yellow ball.

What is the variance of X?

solution (As announced during the quiz, due to the fact that the question (in blue) was accidentally missing, this problem will not count for the scoring of the quiz.)

Let X_1 be the the number of draws in the first phase of the experiment (i.e until we see the first yellow ball) and X_2 be the number of subsequent draws (from the $(X_1+1)st$ draw onwards) until we the first red ball (excluding the first X_1 draws). We observe that both random variables follow Geometric distributions as follows: $X_1 \sim G(p_1)$ and $X_2 \sim G(p_2)$. Furthermore, X_1 and X_2 are independent since they depend on two disjoint sets of independent draws.

Given the above and since $X = X_1 + X_2$, we have:

$$Var(X) = Var(X_1 + X_2) = Var(X_1) + Var(X_2) = \frac{1-p_1}{p_1^2} + \frac{1-p_2}{p_2^2}$$

Question 18 A student has two bags of marbles:

- Bag A contains 4 red marbles and 6 blue marbles.
- Bag B contains 7 red marbles and 3 blue marbles.

The student randomly picks one bag (with equal probability) and then draws one marble from it. The marble drawn is red.

What is the probability that the marble was drawn from Bag A?

solution Define the following events:

- A: Bag A was picked
- B: Bag B was picked
- Red: a red marble was drawn
- Blue: a blue marble was drawn

We will use Bayes theorem to compute the following:

$$P(A|Red) = \frac{P(Red|A)P(A)}{P(Red)}$$

For the denominator, we have:

$$P(Red) = P(Red \cap A) + P(Red \cap B) = P(Red|A)P(A) + P(Red|B)P(B)$$

$$P(Red) = 0.4 \cdot 0.5 + 0.7 \cdot 0.5 = 0.55$$

Thus,

$$P(A|Red) = \frac{P(Red|A)P(A)}{P(Red)} = \frac{0.2}{0.55} \simeq 0.364$$

Question 19 In a company, there are:

- 2 departments with 1 employee
- 4 departments with 2 employees
- 1 department with 3 employees
- 3 departments with 4 employees

A department is randomly selected. Let X be the number of employees in that department.

The expected value of X, the number of employees in the selected department, is [A].

An employee is randomly selected from the entire company.

Let Y be the number of employees in the department where the selected employee works.

The expected value of Y, the number of employees in the selected employee's department, is [B]

solution The expected value of X can be computed as follows:

$$E[X] = \sum x_i \cdot p_i = 1 \cdot 2/10 + 2 \cdot 4/10 + 3 \cdot 1/10 + 4 \cdot 3/10 = 25/10 = 2.5$$

The expected value of Y can be computed as follows:

$$E[Y] = \sum x_i \cdot p_i = 1 \cdot (2 \cdot 1)/25 + 2 \cdot (4 \cdot 2)/25 + 3 \cdot (1 \cdot 3)/25 + 4 \cdot (3 \cdot 4)/25 = 75/25 = 3$$

Question 20 Which of the following statements about continuous random variables is TRUE?

- The probability that a continuous random variable takes any specific single value is always greater than zero.
- The cumulative distribution function (CDF) of a continuous random variable is a decreasing function.
- The area under the probability density function (PDF) over its entire range is always equal to 1.
- Continuous random variables can only take integer values within a given interval.

solution The area under the probability density function (PDF) over its entire range is always equal to 1