Random Variables, Expectations, and Moments - Solved Problems

Random Variables, Expectations, and Moments (Course Textbook's Chp. 4)

Problem 1: Expectation of a Discrete Random Variable Let X be a discrete random variable with the following probability mass function (PMF):

$$P(X = x) = \begin{cases} 0.2 & \text{if } x = 1, \\ 0.5 & \text{if } x = 2, \\ 0.3 & \text{if } x = 3. \end{cases}$$

(a) Find the expected value $E[X]$.

(b) Compute the variance $\text{Var}(X)$.

Solution

(a) The expected value of a discrete random variable is computed by taking the weighted average of all possible values:

$$E[X] = \sum_x xP(X = x).$$

Thus,

$$E[X] = (1)(0.2) + (2)(0.5) + (3)(0.3) = 0.2 + 1.0 + 0.9 = 2.1.$$

(b) The variance of X is given by:

$$\text{Var}(X) = E[X^2] - (E[X])^2.$$

First, calculate $E[X^2]$:

$$E[X^2] = (1^2)(0.2) + (2^2)(0.5) + (3^2)(0.3) = 0.2 + 2.0 + 2.7 = 4.9.$$

Then,

$$\text{Var}(X) = 4.9 - (2.1)^2 = 4.9 - 4.41 = 0.49.$$

Problem 2: Linear Combination of Independent Random Variables Let X_1 and X_2 be two independent random variables with expectations and variances given by:

$$E[X_1] = 3, \quad \text{Var}(X_1) = 2, \quad E[X_2] = 5, \quad \text{Var}(X_2) = 4.$$

Consider a new random variable $Y = 2X_1 + 3X_2$.

(a) Find the expected value $E[Y]$.

(b) Find the variance $\text{Var}(Y)$.

Solution

(a) The expected value of a linear combination of independent random variables is given by:

$$E[Y] = E[2X_1 + 3X_2] = 2E[X_1] + 3E[X_2].$$

Substituting the given values:

$$E[Y] = 2(3) + 3(5) = 6 + 15 = 21.$$

(b) The variance of a linear combination of independent random variables is given by:

$$\text{Var}(Y) = \text{Var}(2X_1 + 3X_2) = 2^2\text{Var}(X_1) + 3^2\text{Var}(X_2).$$

Since X_1 and X_2 are independent, we can add their variances:

$$\text{Var}(Y) = 4 \times 2 + 9 \times 4 = 8 + 36 = 44.$$

Problem 3: Sum of Independent Random Variables Let Z_1 and Z_2 be two independent random variables with expectations and variances given by:

$$E[Z_1] = 4, \quad \text{Var}(Z_1) = 1, \quad E[Z_2] = 6, \quad \text{Var}(Z_2) = 3.$$

Consider the random variable $W = Z_1 + Z_2$.

(a) Find the expected value $E[W]$.

(b) Find the variance $\text{Var}(W)$.

Solution

(a) The expected value of the sum of independent random variables is given by:

$$E[W] = E[Z_1 + Z_2] = E[Z_1] + E[Z_2].$$

Substituting the given values:

$$E[W] = 4 + 6 = 10.$$

(b) The variance of the sum of independent random variables is given by:

$$\text{Var}(W) = \text{Var}(Z_1 + Z_2) = \text{Var}(Z_1) + \text{Var}(Z_2).$$

Since Z_1 and Z_2 are independent, we can add their variances:

$$\text{Var}(W) = 1 + 3 = 4.$$

Problem 4: Chance Game with a Piecewise Random Variable Consider a chance game where a player rolls a fair 6-sided die. Let the random variable X represent the outcome of the roll, where $X \in \{1, 2, 3, 4, 5, 6\}$ with equal probability.

Define a new random variable Y based on the value of X as follows:

$$Y = \begin{cases} 10 & \text{if } X = 1, 2, \text{ or } 3, \\ 20 & \text{if } X = 4 \text{ or } 5, \\ 50 & \text{if } X = 6. \end{cases}$$

(a) Find the probability mass function (PMF) of Y.

(b) Calculate the expected value $E[Y]$.

(c) Determine the variance $\text{Var}(Y)$.

(d) What is the probability that $Y > 10$?

Solution

(a) To find the PMF of Y, we determine the probabilities for each possible value of Y:

- $P(Y = 10) = P(X = 1 \text{ or } 2 \text{ or } 3) = \frac{3}{6} = 0.5$
- $P(Y = 20) = P(X = 4 \text{ or } 5) = \frac{2}{6} = \frac{1}{3} \approx 0.333$
- $P(Y = 50) = P(X = 6) = \frac{1}{6} \approx 0.167$

(b) The expected value of Y is given by:

$$E[Y] = (10)(0.5) + (20)(0.333) + (50)(0.167) = 5 + 6.66 + 8.35 = 20.01.$$

(c) To find the variance of Y, we first need $E[Y^2]$:

$$E[Y^2] = (10^2)(0.5) + (20^2)(0.333) + (50^2)(0.167) = 50 + 133.2 + 417.5 = 600.7.$$

Then,

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2 = 600.7 - (20.01)^2 = 600.7 - 400.4 = 200.3.$$

(d) The probability that $Y > 10$ is given by:

$$P(Y > 10) = P(Y = 20) + P(Y = 50) = 0.333 + 0.167 = 0.5.$$

Problem 5: Evaluating Profitability of a Chance Game Consider a similar chance game where a player rolls a fair 6-sided die. Let the random variable X represent the outcome of the roll, where $X \in \{1, 2, 3, 4, 5, 6\}$ with equal probability.

Define a new random variable Y based on the value of X as follows:

$$Y = \begin{cases} 5 & \text{if } X = 1, 2, 3, \\ 15 & \text{if } X = 4, 5, \\ 30 & \text{if } X = 6. \end{cases}$$

Suppose there is an entry fee of \$10 to play the game. Define the net profit W as:

$$W = Y - 10.$$

(a) Find the probability mass function (PMF) of W.

(b) Calculate the expected value $E[W]$.

(c) Based on your answer in part (b), determine if the game gives positive progress on average.

Solution

(a) To find the PMF of W, we determine the possible values of W and their probabilities:
- $W = 5 - 10 = -5$, which occurs if $X = 1, 2, 3$. Thus,

$$P(W = -5) = P(X = 1 \text{ or } 2 \text{ or } 3) = \frac{3}{6} = 0.5.$$

- $W = 15 - 10 = 5$, which occurs if $X = 4$ or 5. Thus,

$$P(W = 5) = P(X = 4 \text{ or } 5) = \frac{2}{6} = \frac{1}{3} \approx 0.333.$$

- $W = 30 - 10 = 20$, which occurs if $X = 6$. Thus,

$$P(W = 20) = P(X = 6) = \frac{1}{6} \approx 0.167.$$

(b) The expected value of W is given by:

$$E[W] = (-5)(0.5) + (5)(0.333) + (20)(0.167) = -2.5 + 1.665 + 3.34 = 2.505.$$

(c) Since $E[W] = 2.505 > 0$, the game gives a positive expected profit on average. Thus, it is favorable for the player in the long run.

Problem 6: Risk Analysis with Discrete Outcomes An analyst is evaluating the risk associated with a public health intervention. The intervention can result in three outcomes: Improved health ($Z = 0$), No change ($Z = 1$), and Worsened health ($Z = 2$). The probabilities associated with these outcomes are 0.7, 0.2, and 0.1, respectively.

(a) Find the cumulative distribution function (CDF) of Z.

(b) Calculate the probability that the intervention does not worsen health.

Solution

(a) The cumulative distribution function (CDF) for a discrete random variable sums the probabilities of the outcomes up to a certain value. For the random variable Z, the CDF is given by:

$$F(z) = \begin{cases} 0 & \text{if } z < 0, \\ 0.7 & \text{if } 0 \leq z < 1, \\ 0.9 & \text{if } 1 \leq z < 2, \\ 1.0 & \text{if } z \geq 2. \end{cases}$$

This CDF indicates the probability of Z being less than or equal to a certain value.

(b) The probability that the health intervention does not worsen health, $P(Z < 2)$, can be found directly from the CDF:

$$P(Z < 2) = F(1.99) = 0.9.$$

This means there is a 90% chance that the intervention will not worsen health, corresponding to the probabilities of improved health or no change in health.

Problem 7: Evaluating Policy Impact Using Binomial Distribution A policy intervention was implemented to improve employment rates in a region. After the intervention, the probability that an individual is employed is $p = 0.7$.

A random survey of 4 individuals is conducted to assess the employment rate after the policy change. Let Y be the random variable representing the number of employed individuals among the 4 surveyed. Assume Y follows a binomial distribution with parameters $n = 4$ and $p = 0.7$.

(a) Calculate the expected number of employed individuals out of the 4 surveyed.

(b) Find the probability that exactly 3 individuals are employed.

(c) What is the probability that at least 3 individuals are employed?

Solution

(a) The expected number of employed individuals for a binomial distribution is given by:

$$E[Y] = np = 4 \times 0.7 = 2.8.$$

On average, we expect 2.8 out of the 4 individuals to be employed after the policy intervention.

(b) The probability that exactly 3 individuals are employed is given by the binomial probability formula:

$$P(Y = 3) = \binom{4}{3} p^3 (1-p)^{4-3} = \binom{4}{3} (0.7)^3 (0.3)^1.$$

First, calculate the binomial coefficient:

$$\binom{4}{3} = \frac{4!}{3!(4-3)!} = 4.$$

Then, substitute the values of $p = 0.7$ and $1 - p = 0.3$:

$$P(Y = 3) = 4 \times (0.7)^3 \times (0.3) = 4 \times 0.343 \times 0.3 = 0.4116.$$

Thus, the probability that exactly 3 individuals are employed is approximately 0.4116.
(c) The probability that at least 3 individuals are employed is given by:

$$P(Y \geq 3) = P(Y = 3) + P(Y = 4).$$

We already know $P(Y = 3) = 0.4116$. Now, we calculate $P(Y = 4)$ using the binomial formula:

$$P(Y = 4) = \binom{4}{4} p^4 (1-p)^{4-4} = 1 \times (0.7)^4 \times (0.3)^0 = (0.7)^4 = 0.2401.$$

Therefore,

$$P(Y \geq 3) = P(Y = 3) + P(Y = 4) = 0.4116 + 0.2401 = 0.6517.$$

Thus, the probability that at least 3 individuals are employed is approximately 0.6517.

Problem 9: Modeling Public Opinion Using a Continuous Uniform Distribution A government agency conducts a public opinion poll to measure support for a new policy. The agency assumes that public opinion is uniformly distributed over the interval [0, 100], where 0 represents complete opposition to the policy, and 100 represents complete support.

Let the random variable X represent an individual's level of support for the policy, and assume $X \sim \text{Uniform}(0, 100)$. The cumulative distribution function (CDF) of a uniform distribution $X \sim \text{Uniform}(a, b)$ is given by:

$$F(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b, \\ 1 & \text{if } x > b. \end{cases}$$

(a) Find the probability density function (PDF) of X.

(b) Calculate the expected level of support for the policy.

(c) Determine the probability that an individual's support is between 40 and 60.

(d) What is the probability that an individual is strongly supportive of the policy, defined as having a support level greater than 80?

(e) Do you think the assumption that public opinion is uniformly distributed is reasonable in a divided society, where opinions might be clustered around extremes? Justify your answer.

Solution

(a) The probability density function (PDF) for a continuous uniform distribution $X \sim \text{Uniform}(a, b)$ is given by:

$$f(x) = \frac{1}{b-a}, \quad \text{for } a \leq x \leq b.$$

In this case, $a = 0$ and $b = 100$, so the PDF of X is:

$$f(x) = \frac{1}{100}, \quad \text{for } 0 \leq x \leq 100.$$

(b) The expected value $E[X]$ of a uniform distribution Uniform(a, b) is given by:

$$E[X] = \frac{a+b}{2}.$$

Substituting $a = 0$ and $b = 100$:

$$E[X] = \frac{0+100}{2} = 50.$$

Thus, the expected level of support is 50.
(c) The probability that an individual's support is between 40 and 60 is given by the area under the PDF between these values:

$$P(40 \leq X \leq 60) = \frac{60-40}{100} = \frac{20}{100} = 0.2.$$

Therefore, the probability that an individual's support is between 40 and 60 is 0.2 (or 20%).
(d) The probability that an individual is strongly supportive of the policy (with a support level greater than 80) is:

$$P(X > 80) = \frac{100-80}{100} = \frac{20}{100} = 0.2.$$

Thus, the probability that an individual has a support level greater than 80 is 0.2 (or 20%).
(e) The assumption of a uniform distribution implies that every level of support, from 0 to 100, is equally likely. However, in a divided society, opinions might be clustered around extremes, such as strong support (near 100) or strong opposition (near 0), with fewer individuals holding moderate views (in the middle).
In such cases, a uniform distribution might not be appropriate because it assumes that moderate opinions are just as likely as extreme opinions. A more suitable model might be a bimodal distribution, where there are peaks near the extremes, reflecting the polarization in the population. Thus, the uniform assumption may oversimplify the true distribution of opinions in a divided society.

Problem 10: Modeling Public Opinion Response Time Using an Exponential Distribution A government agency is conducting a public opinion survey on a new policy and is interested in modeling the time it takes for individuals to respond to the survey. Based on prior data, the agency assumes that the time T (in hours) it takes for an individual to respond follows an exponential distribution with parameter $\lambda = 0.5$. This means that on average, individuals respond within 2 hours (since $\frac{1}{\lambda} = 2$).

The probability density function (PDF) for an exponential distribution is given by:

$$f(t) = \lambda e^{-\lambda t}, \quad \text{for } t \geq 0.$$

The cumulative distribution function (CDF) is given by:

$$F(t) = 1 - e^{-\lambda t}, \quad \text{for } t \geq 0.$$

(a) Write down the PDF of T given that $\lambda = 0.5$.

(b) Find the expected time for an individual to respond to the survey.

(c) Calculate the probability that an individual responds within 1 hour.

(d) Determine the probability that it takes more than 3 hours for an individual to respond.

(e) In real-world situations, do you think it's reasonable to model the response time with an exponential distribution? What assumptions does the exponential model make about the response time, and are these assumptions likely to hold in practice?

Solution

(a) The probability density function (PDF) of an exponential distribution with $\lambda = 0.5$ is:

$$f(t) = 0.5e^{-0.5t}, \quad \text{for } t \geq 0.$$

(b) The expected value $E[T]$ of an exponential distribution is given by:

$$E[T] = \frac{1}{\lambda}.$$

Substituting $\lambda = 0.5$:

$$E[T] = \frac{1}{0.5} = 2.$$

So, the expected response time is 2 hours.

(c) The probability that an individual responds within 1 hour is given by the cumulative distribution function (CDF):

$$P(T \leq 1) = F(1) = 1 - e^{-0.5\times 1} = 1 - e^{-0.5}.$$

Using the value of $e^{-0.5} \approx 0.6065$:

$$P(T \leq 1) = 1 - 0.6065 = 0.3935.$$

Thus, the probability that an individual responds within 1 hour is approximately 0.3935 (or 39.35%).

(d) The probability that it takes more than 3 hours for an individual to respond is:

$$P(T > 3) = 1 - P(T \leq 3) = 1 - F(3) = 1 - (1 - e^{-0.5\times 3}).$$

Simplifying:

$$P(T > 3) = e^{-0.5\times 3} = e^{-1.5}.$$

Using the value $e^{-1.5} \approx 0.2231$:

$$P(T > 3) \approx 0.2231.$$

Therefore, the probability that it takes more than 3 hours to respond is approximately 0.2231 (or 22.31%).

(e) The exponential distribution assumes that the probability of responding is memoryless, meaning that the probability of responding in the next moment does not depend on how long an individual has already waited to respond. While this might be a reasonable assumption for processes like equipment failure or radioactive decay, it may not always hold for human behavior, where individuals might be more likely to respond at specific times (e.g., during work hours or after a reminder). In real-world survey situations, response times could cluster around particular time periods, and thus, other distributions (such as a Weibull or normal distribution) might be more appropriate in some cases.