Math Revision Session

Statistics (2): Discrete Random Variables and their famous distributions

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

February 16, 2025

1. Discrete Random Variable
2. Multiple discrete random variables
   - Joint and Marginal
   - Covariance, Correlation, and Independence
   - Conditional ...
3. The rule for calculating the expected value
   - Expectation of a Random Variable
   - Expectation and Conditional Expectation Calculation Rules
4. The rule for calculating the variance
5. Typical discrete random variables
   - Bernoulli Dist.
   - Binomial Dist.
   - Poisson Dist.

Discrete Random Variable

- A discrete random variable takes on a countable number of possible values.
- Examples include the number of heads in a series of coin flips, or the number of customers arriving at a store.
- Unlike continuous random variables, they do not take on every possible value within an interval.

Ex.)

- **Number of Heads in Coin Flips:** If you flip a fair coin three times, the number of heads obtained (0, 1, 2, or 3) is a discrete random variable.
- **Number of Customers Arriving:** The number of customers arriving at a store in an hour follows a discrete distribution like Poisson.
- **Rolling a dice:** The outcome of rolling a six-sided dice (1 to 6) is a discrete random variable.

Probability Function

Ex.) Rolling a dice

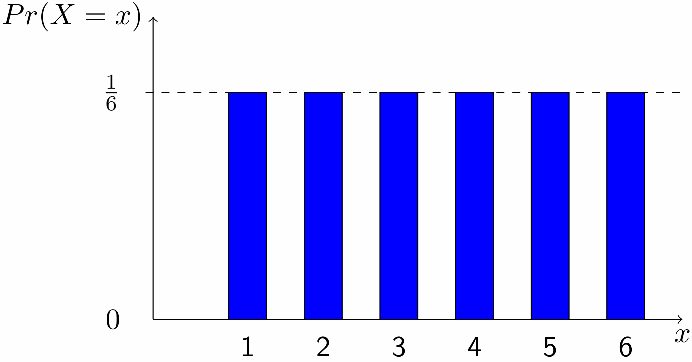
$$X = \begin{cases} 1 & \rightleftarrows \quad Pr(X=1) = 1/6 \\ \vdots & \qquad \vdots \\ 6 & \rightleftarrows \quad Pr(X=6) = 1/6 \end{cases}$$

Of course we have: $Pr(X = 7) = 0$

we can write the probablity as follow:

$$Pr(X = x) = P_X(x)$$

The Probability Mass function (:PMF)



For a discrete random variable X that can take values in ascending order as $x_1, x_2, \ldots$, the following conditions hold:

1. The probability of each possible value is strictly between 0 and 1:

$$0 < P(X = x_i) < 1$$

   for all i.

2. The sum of all probabilities equals 1:

$$\sum P(X = x_i) = 1.$$

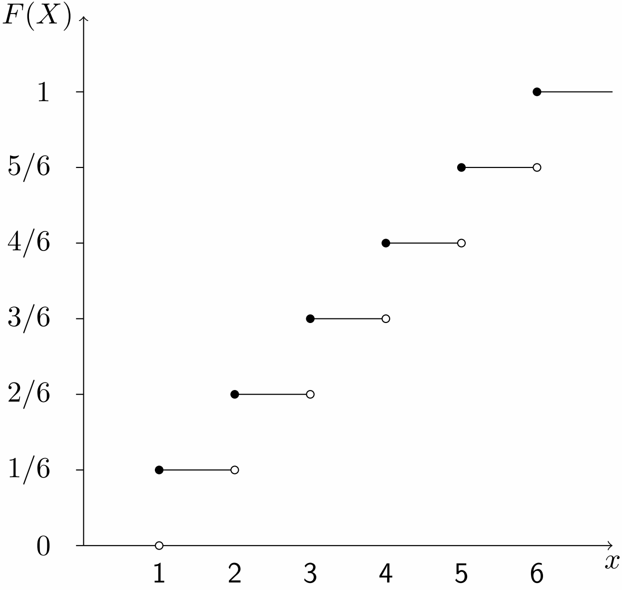
Cumulative Distribution Function (CDF)

- The cumulative distribution function (CDF) of a discrete random variable X is defined as:

$$\begin{aligned} F(X) &= Pr(X \leq x) \\ &= Pr(X = x_1) + \cdots + Pr(X = x_j) \end{aligned}$$

  where $x_j \leq x$.
- It gives the probability that X takes a value less than or equal to x.
- The CDF is a non-decreasing step function.

Cumulative distribution function (: CDF, or distribution function)



Properties

1. $0 \leq F_X(x) \leq 1$.
2. $F_X(-\infty) = 0$, $F_X(\infty) = 1$
3. If $x < x^*$, then we have $F_X(x) \leq F_X(x^*)$

Expectation and Variance of a Discrete Random Variable

- The **expected value** (mean) of a discrete random variable X is given by:

$$E[X] = \sum_i x_i P(X = x_i)$$

- It represents the average outcome if the experiment is repeated many times.
- The **variance** measures the spread of a random variable and is given by:

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

- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\text{Var}(X)}$.

Examples of Expectation and Variance

- **Rolling a Fair dice:**
  - X takes values $1, 2, 3, 4, 5, 6$ with equal probability $\frac{1}{6}$.
  - $E[X] = \sum_{i=1}^{6} i \times \frac{1}{6} = 3.5$.
  - $\text{Var}(X) = \sum_{i=1}^{6} (i - 3.5)^2 \times \frac{1}{6} = \frac{35}{12} \approx 2.92$.

- **Bernoulli Trial:**
  - X takes values 0 and 1 with probabilities $1 - p$ and p.
  - $E[X] = p$.
  - $\text{Var}(X) = p(1 - p)$.

Joint Probability Function

- The joint probability function describes the probability that two or more random variables take specific values simultaneously.
- For example, the probability that random variables X and Y take values x and y, respectively, is given by $P(X = x, Y = y)$.
- This is called the joint probability distribution, and it is defined as:

$$P(X = x, Y = y) = p_{X,Y}(x, y)$$

Example: Joint probability for the outcome of two dice rolls

$$P(X = 1, Y = 3) = \frac{1}{36}$$

Here, X represents the outcome of the first dice and Y represents the outcome of the second dice.

Marginal Probability Function

- The marginal probability function describes the probability of a single random variable, regardless of the other variables.
- For example, the marginal probability of X is calculated as:

$$P(X = x) = \sum_{y} P(X = x, Y = y)$$

- The marginal probability function gives us information about one random variable, independent of others.

Example: Marginal probability for one of the dice

$$P(X = 1) = \sum_{y=1}^{6} P(X = 1, Y = y) = \frac{6}{36} = \frac{1}{6}$$

This represents the marginal probability when X equals 1.

Covariance

- Covariance measures the degree to which two random variables change together.
- For two random variables X and Y, the covariance is given by:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

  where μ_X and μ_Y are the means of X and Y, respectively.
- If the covariance is positive, it indicates that as X increases, Y tends to increase as well. If negative, Y tends to decrease as X increases.

Example: Covariance between two variables

$$\text{Cov}(X, Y) = \sum_{i=1}^{n} \frac{(X_i - \mu_X)(Y_i - \mu_Y)}{n}$$

Correlation Coefficient

- The correlation coefficient measures the strength and direction of the linear relationship between two random variables.
- It is defined as the normalized version of covariance:

$$\rho(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

  where σ_X and σ_Y are the standard deviations of X and Y, respectively.
- The value of the correlation coefficient ranges from -1 to 1:
  - $\rho = 1$ indicates a perfect positive linear relationship.
  - $\rho = -1$ indicates a perfect negative linear relationship.
  - $\rho = 0$ indicates no linear relationship.

Example: Correlation coefficient calculation

$$\rho(X, Y) = \frac{\mathsf{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

If $\mathsf{Cov}(X, Y) = 2$, $\sigma_X = 3$, and $\sigma_Y = 4$, then:

$$\rho(X, Y) = \frac{2}{3 \times 4} = \frac{1}{6}$$

Independence of Random Variables

- Two random variables X and Y are said to be **independent** if the occurrence of one does not affect the probability of the other.
- Mathematically, X and Y are independent if and only if:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \quad \text{for all values of } x \text{ and } y.$$

- In terms of expectation, independent random variables satisfy:

$$E[XY] = E[X] \cdot E[Y]$$

- Independence implies that there is no relationship between the variables, unlike covariance or correlation, which may indicate some form of dependency.

Example: Tossing two fair coins. If X is the outcome of the first coin and Y is the outcome of the second, the variables are independent because:

$$P(X = H, Y = T) = P(X = H) \cdot P(Y = T) = \frac{1}{4}.$$

Correlation and Independence

- **Correlation coefficient (ρ)** measures the strength and direction of the linear relationship between two variables.
- A correlation coefficient of 0 suggests no linear relationship, but this does not necessarily imply independence.
- **Independence** between two variables implies a correlation of 0, but the reverse is not true.
- Two variables having a correlation of 0 does not mean they are independent; they might have a nonlinear relationship.

Example: If X and Y are independent, then:

$$\text{Cov}(X, Y) = 0 \quad \Rightarrow \quad \rho(X, Y) = 0$$

However, if $\rho(X, Y) = 0$, this only means there is no linear relationship, but the variables may still be dependent in a nonlinear way.

Conditional Probability

- **Conditional probability** is the probability of an event occurring given that another event has already occurred.
- The conditional probability of A given B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

  where $P(A \cap B)$ is the probability that both A and B occur.
- Conditional probability helps to update the probability of an event based on the occurrence of another event.

Example: If we have two dice rolls, and we know that the first dice shows a 4, what is the probability that the second dice shows a 6?

$$P(\text{second dice} = 6 \mid \text{first dice} = 4) = \frac{P(\text{first dice} = 4, \text{second dice} = 6)}{P(\text{first dice} = 4)}$$

$$= \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

Conditional Expectation

- **Conditional expectation** is the expected value of a random variable given that certain conditions are known.
- The conditional expectation of X given $Y = y$ is defined as:

$$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$$

  where $P(X = x|Y = y)$ is the conditional probability of $X = x$ given $Y = y$.
- Conditional expectation gives the expected value of X based on the information available about Y.

Example: Suppose X represents the outcome of a dice roll, and we know that Y is the outcome of a second dice roll. The conditional expectation of X given $Y = 6$ is:

$$E[X|Y = 6] = \sum_{x=1}^{6} x \cdot P(X = x|Y = 6)$$

Since X and Y are independent, the conditional probability $P(X = x|Y = 6) = P(X = x)$, and thus:

$$E[X|Y = 6] = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$

Expectation of a Random Variable

- The **expectation** (or **expected value**) of a random variable X, denoted as $E[X]$, is the weighted average of all possible values of X, weighted by their probabilities.
- For a discrete random variable, the expectation is calculated as:

$$E[X] = \sum_{x} x \cdot P(X = x)$$

  where x represents possible values and $P(X = x)$ is the probability of $X = x$.
- For a continuous random variable, the expectation is given by:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \, dx$$

  where $f_X(x)$ is the probability density function of X.

Linearity of Expectation

- The **linearity of expectation** states that the expectation of the sum of random variables is the sum of their expectations, regardless of whether the random variables are independent or not.
- Mathematically, for any random variables X and Y, and constants a, b, and c:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

- This property holds even if X and Y are not independent.

Example: If X and Y are random variables with $E[X] = 2$ and $E[Y] = 3$, then:

$$E[2X + 3Y] = 2E[X] + 3E[Y] = 2 \times 2 + 3 \times 3 = 4 + 9 = 13$$

Expectation of a Product of Independent Random Variables

- For two independent random variables X and Y, the expectation of their product is the product of their expectations:

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

- This rule holds only when X and Y are independent. If they are not independent, the expectation of their product is not necessarily the product of their expectations.

Example: If X and Y are independent with $E[X] = 2$ and $E[Y] = 3$, then:

$$E[X \cdot Y] = E[X] \cdot E[Y] = 2 \times 3 = 6$$

Expectation and Conditional Expectation Calculation Rules

- **Linearity of Expectation:** If X and Y are random variables, and a and b are constants, then:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

- **Conditional Expectation of a Sum:** If X and Y are random variables, then:

$$\mathbb{E}[X + Y \mid Z] = \mathbb{E}[X \mid Z] + \mathbb{E}[Y \mid Z]$$

- **Tower Property (Law of Total Expectation):** For a random variable X and a sigma-algebra $\mathcal{G}$, the law of total expectation states:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]]$$

  This means that the expectation of X can be obtained by first conditioning on $\mathcal{G}$ and then taking the expectation of the conditional expectation.

- **Iterated Expectation:** If X and Y are random variables, then:

$$\mathbb{E}[X \mid Y] = \mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Z]$$

  This rule allows breaking down expectations into simpler conditional expectations.
- **Conditional Expectation of a Product:** If X and Y are random variables, then:

$$\mathbb{E}[XY \mid Z] = \mathbb{E}[X \mid Z]\mathbb{E}[Y \mid Z]$$

  when X and Y are conditionally independent given Z.

Variance and Expectation

- The **variance** of a random variable X, denoted $\text{Var}(X)$, measures the spread of values around the expected value.
- The variance is related to the expectation as follows:

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

where $E[X^2]$ is the expectation of the square of X.

Example: If X is a random variable with $E[X] = 5$ and $E[X^2] = 30$, then:

$$\text{Var}(X) = E[X^2] - (E[X])^2 = 30 - 5^2 = 30 - 25 = 5$$

Variance Calculation Rules

- The variance of a random variable X is defined as:

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

- For a linear transformation of the form $Y = aX + b$, the variance is calculated as:

$$\mathsf{Var}(Y) = a^2\mathsf{Var}(X).$$

- The variance of the sum of two independent random variables X and Y is:

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) \quad \text{(if } X \text{ and } Y \text{ are independent).}$$

- For the variance of a sum of dependent random variables, we use:

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2 \cdot \mathsf{Cov}(X, Y),$$

where $\mathsf{Cov}(X, Y)$ is the covariance between X and Y.

Example: Variance of a Linear Transformation

- Suppose X is a random variable with $\mathsf{Var}(X) = 4$.
- If we perform a linear transformation $Y = 3X + 2$, then:

$$\mathsf{Var}(Y) = 3^2 \times \mathsf{Var}(X) = 9 \times 4 = 36.$$

- Thus, $\mathsf{Var}(Y) = 36$.

Example: Variance of the Sum of Random Variables

- If X and Y are independent random variables with:

$$\mathsf{Var}(X) = 5, \quad \mathsf{Var}(Y) = 3,$$

then the variance of their sum is:

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) = 5 + 3 = 8.$$

Bernoulli Distribution

- The **Bernoulli distribution** is a discrete probability distribution for a random variable that takes only two possible values: success (: 1) and failure (: 0).
- A random variable X follows a Bernoulli distribution with parameter p if:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p, \quad \text{where } 0 \leq p \leq 1.$$

  This parameter p, for example, represents the probability of success.
- The probability mass function (PMF) is given by:

$$P(X = x) = p^x (1 - p)^{1-x}, \quad x \in \{0, 1\}.$$

Example: Tossing a biased coin with probability $p = 0.7$ of landing heads ($X = 1$) and $1 - p = 0.3$ of landing tails ($X = 0$).

The probability function is:

$$P_X(1) = p, \quad P_X(0) = 1 - p.$$

The CDF is:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

If a random variable X follows the Bernoulli distribution, then we write as follows:

$$X \sim \text{Bernoulli}(p)$$

Expectation and Variance of a Bernoulli Distribution

- The **expected value** (mean) of a Bernoulli-distributed random variable X is:

$$E[X] = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = p.$$

- The **variance** of X is:

$$\text{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p).$$

Example: If X follows a Bernoulli distribution with $p = 0.7$, then:

$$E[X] = 0.7, \quad \text{Var}(X) = 0.7 \times (1 - 0.7) = 0.21.$$

Binomial Distribution

- The **binomial distribution** describes the number of successes in n independent Bernoulli trials, each with success probability p.
- A random variable X follows a binomial distribution if:

$$X \sim \mathsf{Bin}(n, p),$$

  where:
  - n is the number of trials,
  - p is the probability of success in each trial.

The binomial distribution is useful for describing uncertain quantities that take discrete values, such as counts or numbers of people.

- The probability of obtaining exactly k successes in n trials is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

- Here, $\binom{n}{k}$ is the binomial coefficient:

$$\binom{n}{k} = {}_nC_k = \frac{n!}{k!(n-k)!}.$$

Example: Suppose we flip a biased coin ($p = 0.6$) 5 times. The probability of getting exactly 3 heads is:

$$P(X = 3) = \binom{5}{3} (0.6)^3 (0.4)^2 = 10 \times 0.216 \times 0.16 = 0.3456.$$

In the binomial distribution, considering **combinations** is essential because we are interested in the probability of obtaining a certain number of successes regardless of their order.

For example, when flipping a coin 10 times and obtaining 3 heads, the specific positions where heads appear (e.g., first, second, and third flips, or second, fifth, and ninth flips) form different combinations, each with the same probability of occurring. Therefore, to determine the total probability of obtaining exactly 3 heads, we must count all possible arrangements and sum their probabilities.

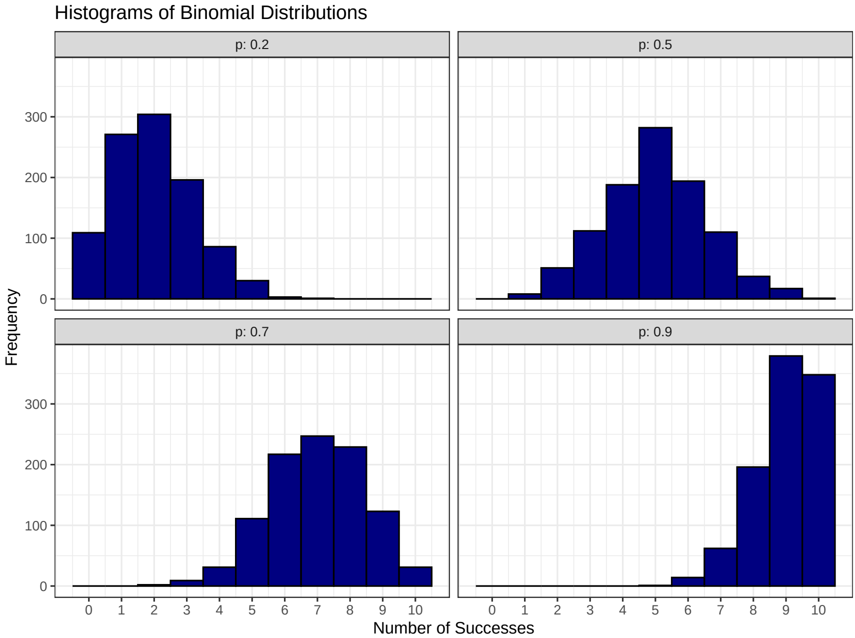
Expectation and Variance

- The **expected value** of a binomially distributed random variable $X \sim \text{Bin}(n, p)$ is:
$$E[X] = np.$$
- The **variance** of X is:
$$\text{Var}(X) = np(1-p).$$

Example: If $X \sim \text{Bin}(10, 0.3)$, then:

$$E[X] = 10 \times 0.3 = 3, \quad \text{Var}(X) = 10 \times 0.3 \times 0.7 = 2.1.$$

$X \sim Bin(10, p)$, Samplilng 1000 data points from this distribution.



Problem

JH's father goes fishing every day for 31 days during the summer vacation in August. When the sea is rough, the cruiser shakes and it becomes dangerous, so he reluctantly gives up on fishing. Based on past climate data, the probability of the sea being rough is 0.2. Answer the following questions:

1. What is the expected number of days JH's father will be able to go fishing?
2. What is the probability that he will be able to go fishing all 31 days?
3. What is the probability that the number of fishing days will be 20 or fewer?

Note: The only reason JH's father will stop fishing is if the sea is rough.

Answer:

1. Let X represent the number of days JH's father can actually go fishing out of the 31 days. The probability that the sea will not be rough is p, which is calculated as:

$$p = 1 - 0.2 = 0.8$$

Since X is a random variable, it follows a binomial distribution:

$$X \sim \text{Bin}(31, 0.8)$$

The expected value of X can be calculated as:

$$E[X] = n \times p = 31 \times 0.8 = 24.8$$

2. The probability that JH's father can go fishing all 31 days is calculated as follows:

$$Pr(X = 31) = {}_{31}C_{31} \times 0.8^{31} \times 0.2^{31-31} \simeq 0.001$$

3.

$$\begin{aligned} Pr(X \leq 20) &= Pr(X = 0) + \cdots + Pr(X = 20) \\ &= 1 - \{Pr(X = 21) + \cdots + Pr(X = 31)\} \end{aligned}$$

Which is better for your calculation?

Relation to the Binomial Distribution

- The Bernoulli distribution is a special case of the **binomial distribution**.
- If we perform n independent Bernoulli trials with probability p, the total number of successes follows a binomial distribution:

$$X \sim \mathsf{Bin}(n, p).$$

- The Bernoulli distribution is simply the binomial distribution with $n = 1$:

$$X \sim \mathsf{Bernoulli}(p) \quad \Leftrightarrow \quad X \sim \mathsf{Bin}(1, p).$$

Example: A single coin flip follows a Bernoulli distribution, while the number of heads in 10 flips follows a binomial distribution.

Poisson Distribution

- The **Poisson distribution** models the number of events occurring in a fixed interval of time or space, given that these events happen with a known constant mean rate and independently of the time since the last event.
- A random variable X follows a Poisson distribution with parameter λ (the average rate of occurrence), denoted as:

$$X \sim \text{Poisson}(\lambda),$$

  where $\lambda > 0$ is the rate of events per interval.
- This distribution is used to investigate the frequency or count of events that occur with an extremely low probability, meaning events that happen only rarely.

The Poisson distribution is not related to fish. The name "Poisson" comes from the French mathematician Siméon-Denis Poisson. Poisson was a 19th-century mathematician who made significant contributions to probability theory and statistics. The Poisson distribution is one of his key results, particularly used to model the occurrence of random events within a fixed period of time or space.

So, "Poisson" refers to the mathematician, not a fish!

Probability Mass Function (PMF)

- The probability of observing exactly k events in an interval is given by the Poisson PMF:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

where e is Euler's number (approximately 2.71828), and k is the number of events observed.

Example: If $\lambda = 3$, the probability of observing exactly 2 events is:

$$P(X = 2) = \frac{3^2 e^{-3}}{2!} = \frac{9e^{-3}}{2} \approx 0.2241.$$

Expectation and Variance

- The **expected value** of a Poisson-distributed random variable $X \sim \text{Poisson}(\lambda)$ is:
$$E[X] = \lambda.$$
- The **variance** of X is also:
$$\text{Var}(X) = \lambda.$$

Example: If $X \sim \text{Poisson}(5)$, then:

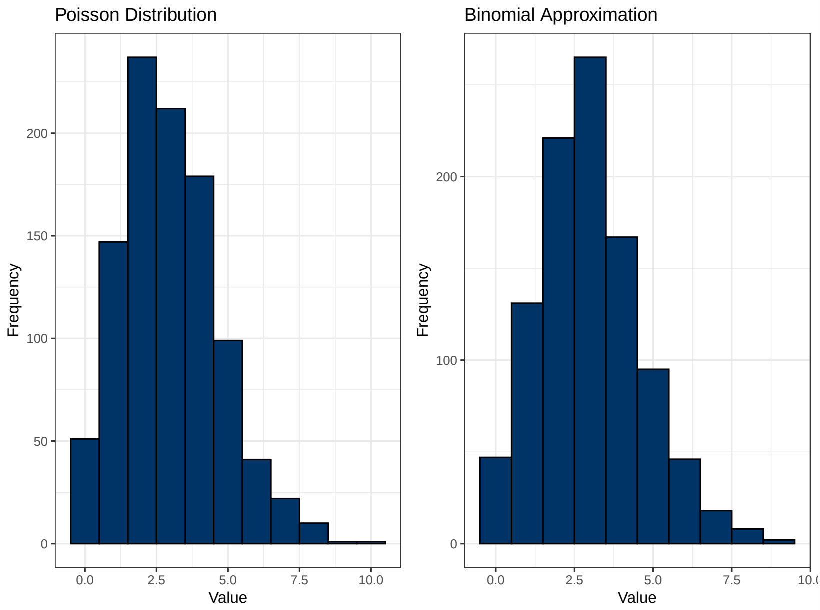
$$E[X] = 5, \quad \text{Var}(X) = 5.$$

Relation to the Binomial Distribution

- The Poisson distribution can be seen as a limiting case of the binomial distribution.
- When the number of trials n becomes large and the probability of success in each trial p becomes small, such that $np = \lambda$ remains constant, the binomial distribution approximates the Poisson distribution.

Example: If $n = 1000$ and $p = 0.005$, the binomial distribution Bin$(1000, 0.005)$ can be approximated by Poisson(5).

$\lambda = 3$, sample size $= 1000$, and λ is approximated by $n = 30, p = \lambda/n$.



Problem

There is a machine in the production department of a company that has been in use for quite a long time. This machine is very old, and it has been found that it produces defective products 5 times out of every 1000. Calculate the probability that this machine will not produce any defective products when manufacturing 50 items.

Answer

- By binomial distribution assumption:

$$\begin{aligned} Pr(X=0) &= {}_{50}C_0 0.005^0 (1-0.005)^{50-0} \\ &= 0.995^{50} \simeq 0.7783 \end{aligned}$$

- By Poisson distribution assumption:

$$\begin{aligned} Pr(X=0) &= \frac{0325^0 e^{-0.25}}{0!} \\ &= e^{-0.25} \simeq 0.7788 \end{aligned}$$