Discrete Distributions

Bernoulli Trials, Binomial Distribution and Geometric Distribution

Noara Razzak

April 30, 2020

Probability Mass Function (PMF)

The **Probability Mass Function (PMF)** is a function that gives the probability that a discrete random variable is exactly equal to some value. It is a fundamental concept in probability theory for discrete random variables.

Definition

For a discrete random variable X taking values in a countable set $\mathcal{X}$, the PMF is defined as:

$$p_X(x) = P(X = x), \quad x \in \mathcal{X}$$

where:

- $p_X(x)$ is the probability mass function of X
- $P(X = x)$ is the probability that the random variable X takes the value x

Properties of PMF

A valid PMF must satisfy the following properties:

1. **Non-negativity**:

$$p_X(x) \geq 0 \quad \text{for all } x \in \mathcal{X}$$

2. **Normalization**:

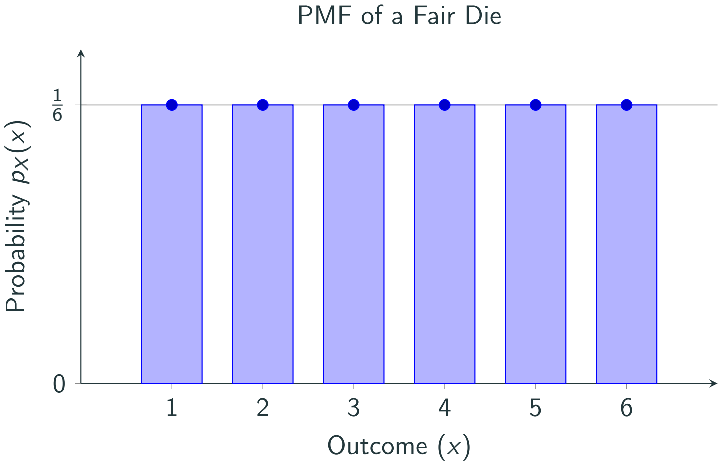
$$\sum_{x \in \mathcal{X}} p_X(x) = 1$$

Example: PMF of a Fair Die

For a fair six-sided die, the random variable X representing the outcome has PMF:

$$p_X(x) = \begin{cases} \frac{1}{6} & \text{for } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

Visualization of PMF



Relation to Other Functions

- For continuous random variables, the analogous concept is the **probability density function (PDF)**
- The **cumulative distribution function (CDF)** can be obtained from the PMF by:

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k)$$

Bernoulli Trials

A **Bernoulli trial** is the simplest probabilistic experiment with only **two outcomes**, **constant probability** and **independent trials**.

Bernoulli Trials Examples

- Distribution: Flipping a coin (Heads = success, Tails = failure).
- Rolling a die and getting a specific number (e.g., a 6).
- Testing a light bulb and determining if it works (success) or not (failure).

Bernoulli Trials

A random variable X follows a **Bernoulli distribution** if:

$$X = \begin{cases} 1 \quad \text{(success)} \quad \text{with probability } p, \\ 0 \quad \text{(failure)} \quad \text{with probability } 1 - p. \end{cases}$$

Bernoulli Trials

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

Probability Mass Function (PMF)

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

Expectation (Mean)

$$E[X] = p$$

Variance

$$Var(X) = p(1-p)$$

Bernoulli Trials

Example 1 Eight balls are drawn from a bag containing 10 white and 10 black balls.

Bernoulli Trials

Example 1

1. Predict whether the trials are Bernoulli trials if the ball drawn is replaced and not replaced.
2. If they are Bernoulli trials,iIn words, define the random variable X.
3. List the values that x may take on.
4. Give the distribution of X

Bernoulli Trials

Example 1 - Solution

1. a. For the first case, when a ball is drawn with replacement, the probability of success (say, white ball) is p=10/20=1/2, which is the same for all eight trials (draws) – Bernoulli trials.
   b. For the second case, when a ball is drawn without replacement, the probability of success (say, white ball) varies with the number of trials. For example, for the first trial, the probability of success, p=10/20. For the second trial, the probability of success is p=9/19, which is not equal to the first trial–not Bernoulli trials.
2. X is the number of balls drawn
3. $x = 0, 1, 2, 3, 4, 5, 6, 7$ and 8.
4. $X \sim B(10/20)$

Example 2: Single Coin Toss

- Let $X = 1$ if **Heads** ($p = 0.5$), else $X = 0$.

What is the mean and variance?

Example 2 - Solution

- **Mean:** $E[X] = 0.5$
- **Variance:** $Var(X) = 0.5 \times 0.5 = 0.25$

Binomial Distribution

If we perform n independent Bernoulli trials, the total number of successes Y follows a **Binomial distribution**:

$$Y \sim \text{Binomial}(n, p)$$

PMF of Binomial Distribution

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Expectation

$$E[Y] = np,$$

Variance

$$Var(Y) = np(1 - p)$$

Note: $p =$ The probability that the event occurs in a given interval is the same for all intervals.

Binomial Distribution

There are three characteristics of a binomial experiment.

- There are a fixed number of trials. Think of trials as repetitions of an experiment. The letter n denotes the number of trials.
- There are only two possible outcomes, called "success" and "failure", for each trial. The letter p denotes the probability of a success on one trial, and q denotes the probability of a failure on one trial, $p + q = 1$.
- The n trials are independent and are repeated using identical conditions.

Binomial Distribution

Example 3 Approximately 70% of statistics students do their homework in time for it to be collected and graded. Each student does homework independently. In a statistics class of 50 students, what is the probability that 40 will do their homework on time? Students are selected randomly.

Binomial Distribution

Example 3

1. In words, define the random variable X.
2. List the values that x may take on.
3. Find the mean μ number of students who do their homework on time.
4. Give the distribution of X.
5. What is the probability that 40 students complete their homework on time?

Binomial Distribution

Example 3 - Solution

1. X is the number of statistics students who do their homework on time.
2. $x = 0, 1, 2, 3, 4, 5, 6, 7, ..., 50$
3. Mean: $np = 50 \times 0.7$.
4. $X \sim Binomial(50, 0.7)$
5. $P(X = 3) = \binom{50}{40}(0.7)^{40}(0.3)^{10} = 0.0386$

Binomial Distribution

Example 4

It has been stated that about 41% of adult workers have a high school diploma but do not pursue any further education. If 20 adult workers are randomly selected, find the probability that more than 2 of them have a high school diploma but do not pursue any further education. How many adult workers do you expect to have a high school diploma but do not pursue any further education?

Binomial Distribution

Example 3 - *Solution*

First the probability of $x = 0, 1, 2$ and then subtract result from 1

- $P(X = 0) = \binom{20}{0}(0.41)^{20}(0.3)^{0} = 0$
- $P(X = 1) = \binom{20}{1}(0.41)^{1}(0.3)^{19} = 0.0004$
- $P(X = 2) = \binom{20}{2}(0.41)^{2}(0.3)^{18} = 0.0024$
- $P(X > 2) = 1 - (X \leq 2) = 1 - 0.0028 = 0.99972$

$$\mu = np = 20 \times 0.41 = 8.2$$

Geometric Distribution

The **Geometric Distribution** describes the number of trials needed to get the first success in repeated Bernoulli trials with success probability p.

Geometric Distribution

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

Probability Mass Function (PMF)

$$P(X = k) = (1 - p)^{k-1} \cdot p, \quad \text{for } k = 1, 2, 3, \ldots$$

Expected Value

$$E[X] = \frac{1}{p}$$

Variance

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

Example 5

Suppose that you are looking for a student at your college who lives within five miles of you. You know that 55% of the 25,000 students do live within five miles of you. You randomly contact students from the college until one says he or she lives within five miles of you.

Example 5

1. In words, define the random variable X.
2. List the values that x may take on.
3. Find p and q.
4. Give the distribution of X.
5. What is the probability that you need to contact four people?

Example 5 - Solution

1. X is the number of students.
2. $x = 0, 1, 2, 3, 4, 5, 6,..$(total number of students)
3. Mean: $p = 0.55$ and $q = 0.45$.
4. $X \sim Geometric(0.55)$
5. $P(X = 4) = (0.45)^3(0.55)^1 = 0.50$

Next class we will move on to Hyper-geometric and Poisson Distribution.