UNIT-V: Probability Distribution

Dr. Pavan Kumar Singh

Probability Distribution

- In Statistics, the **probability distribution** gives the possibility of each outcome of a random experiment or event. It provides the probabilities of different possible occurrences.

- Probability distribution yields the possible outcomes for any random event. It is also defined based on the underlying sample space as a set of possible outcomes of any random experiment. These settings could be a set of real numbers or a set of vectors or a set of any entities. It is a part of probability and statistics.

Types of Probability Distribution

- There are two types of probability distribution which are used for different purposes and various types of the data generation process.

1. Discrete Probability Distribution

(i) Binomial Probability Distribution

(ii) Poisson Probability Distribution

2. Continuous Probability Distribution

Continuous (Cumulative) Probability Distribution

- The cumulative probability distribution is also known as a **continuous probability distribution.** In this distribution, the set of possible outcomes can take on values in a continuous range.
- For example, **a set of real numbers**, is a continuous or normal distribution, as it gives all the possible outcomes of real numbers.

Continuous Probability Distribution

- Normal distribution
- t-Distribution
- Uniform distribution
- Exponential distribution
- Chi-Square Distribution
- F-Distribution
- Beta distribution

Binomial Probability Distribution

- The **binomial distribution** is the discrete probability distribution that gives only two possible results in an experiment, either **Success or Failure**.
- For example, if we toss a coin, there could be only two possible outcomes: heads or tails, and if any test is taken, then there could be only two results: pass or fail.
- This distribution is also called a binomial probability distribution.

Binomial Probability Distribution

The formula

If we consider the probability that in n number of trials, with r successes and $n-r$ failures by $P_r(x=r)$ then,

$$P_r(x=r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

Where p is probability of success and q is the probability of failure.

Recall: $${}^{n}C_{r} = \frac{n!}{(n-r)! \times r!}$$

Note that the events are independent of one another for the number of trials.

Binomial Probability Distribution

The Properties

If we denote the mean and standard deviation of the binomial distribution as μ and σ respectively, then:

(i) mean (μ) $= np$

(ii) standard deviation $(\sigma) = \sqrt{npq}$

(iii) variance $(\sigma^2) = npq$

Binomial Probability Distribution

Example 1

A fair coin is tossed 6 times. Find the probability of obtaining: (a) exactly 4 heads;

(b) at least 5 heads;

(c) at most 2 heads

$$P_r(x = r) = {}^{n}C_r\, p^r q^{n-r}$$

Example 1

A fair coin is tossed 6 times. Find the probability of obtaining: (a) exactly 4 heads;

(b) at least 5 heads;

(c) at most 2 heads

$$P_r(x = r) = {}^nC_r p^r q^{n-r}$$

$p = \frac{1}{2} \quad q = \frac{1}{2} \quad n = 6$

$$\text{(a) } P_r(x = 4) = {}^6C_4 \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^2$$

$$= 15 \times \frac{1}{16} \times \frac{1}{4}$$

Binomial Probability Distribution

Example 1

A fair coin is tossed 6 times. Find the probability of obtaining: (a) exactly 4 heads;

(b) at least 5 heads;

(c) at most 2 heads

$$P_r(x = r) = {}^nC_r\, p^r q^{n-r}$$

$p = \frac{1}{2} \quad q = \frac{1}{2} \quad n = 6$

(b) $P_r(x \geq 5) = P_r(x = 5) \text{ or } P_r(x = 6)$

$$= {}^6C_5 \times \left(\frac{1}{2}\right)^5 \times \left(\frac{1}{2}\right)^1 + {}^6C_6 \times \left(\frac{1}{2}\right)^6 \times \left(\frac{1}{2}\right)^0$$

$$= 6 \times \frac{1}{32} \times \frac{1}{2} + 1 \times \frac{1}{64} \times 1$$

$$= \frac{6}{64} + \frac{1}{64}$$

Example 1

A fair coin is tossed 6 times. Find the probability of obtaining: (a) exactly 4 heads;

(b) at least 5 heads;

(c) at most 2 heads

$$P_r(x = r) = {}^nC_r p^r q^{n-r}$$

$p = \frac{1}{2}$ $q = \frac{1}{2}$ $n = 6$

(c) $P_r(x \leq 2) = P_r(x = 0)$ or $P_r(x = 1)$ or $P_r(x = 2)$

$$= {}^6C_0 \times \left(\frac{1}{2}\right)^0 \times \left(\frac{1}{2}\right)^6 + {}^6C_1 \times \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^5 + {}^6C_2 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{1}{2}\right)^4$$

$$= 1 \times 1 \times \frac{1}{64} + 6 \times \frac{1}{2} \times \frac{1}{32} + 15 \times \frac{1}{4} \times \frac{1}{16}$$

$$= \frac{1}{64} + \frac{6}{64} + \frac{15}{64}$$

Binomial Probability Distribution

Example 3

A test contains 10 multiple choice questions comprising of 4 options in which only one option is correct. Find the probability that a candidate can guess 7 out of the 10 questions correctly.

$$P_r(x = r) = {}^nC_r p^r q^{n-r}$$

Binomial Probability Distribution

Example 3

A test contains 10 multiple choice questions comprising of 4 options in which only one option is correct. Find the probability that a candidate can guess 7 out of the 10 questions correctly.

$$\mathrm{P_r}(x = r) = {}^{n}\mathrm{C}_r p^r q^{n-r}$$

$$p = \frac{1}{4} \quad q = \frac{3}{4} \quad n = 10 \quad r = 7$$

$$\mathrm{P_r}(x = 7) = {}^{10}\mathrm{C}_7 \times \left(\frac{1}{4}\right)^7 \times \left(\frac{3}{4}\right)^3$$

$$= 120 \times \frac{1}{16384} \times \frac{27}{64}$$

$$= 0.0030899$$

Poisson Probability Distribution

Introduction

When the number of trials is relatively large and the probability of success is comparatively small, the most appropriate approach to such random experiment is Poisson probability distribution.

The Poisson probability distribution formula is given as:

$$\Pr(x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$$

Where

$x = 0, 1, 2, 3, \ldots$

$\lambda = np$

$e = 2.718$ **(Euler's constant)**

Poisson Probability Distribution

The Properties

If we denote the mean and standard deviation of the poisson distribution as μ and σ respectively, then:

(i) mean (μ) = λ

(ii) standard deviation (σ) = $\sqrt{\lambda}$

(iii) Variance (σ^2) = λ

Practical considerations

In practice, we can use the Poisson distribution to very closely approximate the binomial distribution provided that the product np is constant with

$$n \geq 100 \qquad \text{and} \qquad p \leq 0.05$$

Note that this is not a hard-and-fast rule and we simply say that

> 'the larger n is the better and the smaller p is the better provided that np is a sensible size.'

The approximation remains good provided that $np < 5$ for values of n as low as 20.

Example 1

3 in every 1000 H-mobile phones are discovered to have fault. Find the probability that out of 5 000 H-mobile phones, exactly 8 will have fault.

$$p = \frac{3}{1000} = 0.003,\ n = 5\,000,\ \lambda = 5\,000 \times 0.003 = 15 \qquad \Pr(x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$$

$$P_r(x = 8) = \frac{15^{8} \times 2.718^{-15}}{8!}$$

$$= 0.01947$$

- **Prob.** In the manufacture of glassware, bubbles can occur in the glass which reduces the status of the glassware to that of a ‘second’. If, on average, one in every 1000 items produced has a bubble, calculate the probability that exactly six items in a batch of three thousand are seconds.

Solution

Suppose that X = number of items with bubbles, then $\quad X \sim B(3000, 0.001)$

Since $n = 3000 > 100$ and $p = 0.001 < 0.005$ we can use the Poisson distribution with $\lambda = np = 3000 \times 0.001 = 3$. The calculation is:

$$\mathrm{P}(X = 6) = \mathrm{e}^{-3}\frac{3^6}{6!} \approx 0.0498 \times 1.0125 \approx 0.05$$

The result means that we have about a 5% chance of finding exactly six seconds in a batch of three thousand items of glassware.

- **Prob.** A manufacturer produces light-bulbs that are packed into boxes of 100. If quality control studies indicate that 0.5% of the light-bulbs produced are defective, what percentage of the boxes will contain:
- (a) no defective?
- (b) 2 or more defectives?

Solution

As n is large and p, the P(defective bulb), is small, use the Poisson approximation to the binomial probability distribution. If X = number of defective bulbs in a box, then

$$X \sim \mathsf{P}(\mu) \text{ where } \mu = n \times p = 100 \times 0.005 = 0.5$$

(a) $\mathsf{P}(X = 0) = \dfrac{e^{-0.5}(0.5)^0}{0!} = \dfrac{e^{-0.5}(1)}{1} = 0.6065 \approx 61\%$

(b) $\mathsf{P}(X = 2 \text{ or more}) = \mathsf{P}(X = 2) + \mathsf{P}(X = 3) + \mathsf{P}(X = 4) + \ldots$ but it is easier to consider:

$$\mathsf{P}(X \geq 2) = 1 - [\mathsf{P}(X = 0) + \mathsf{P}(X = 1)]$$

$$\mathsf{P}(X = 1) = \frac{e^{-0.5}(0.5)^1}{1!} = \frac{e^{-0.5}(0.5)}{1} = 0.3033$$

i.e. $\quad \mathsf{P}(X \geq 2) = 1 - [0.6065 + 0.3033] = 0.0902 \approx 9\%$

- **Prob:** If a Poisson variate X is such that P[X = 1] = 2P[X = 2], find the mean and variance of the distribution.

Question:

Find the mean, variance, and standard deviation of the binomial distribution with the given values of n and p.

$n = 121, p = 0.44$

Solution: Let λ be the mean of the distribution, hence by Poisson distribution,

$$P[X = x] = \frac{e^{-\lambda}\lambda^{x}}{\lfloor\underline{x}}; x = 0, 1, 2, \ldots$$

Now, $P[X = 1] = 2P[X = 2]$

$$\Rightarrow \quad \frac{e^{-\lambda}.\lambda^{1}}{\lfloor\underline{1}} = 2\frac{e^{-\lambda}.\lambda^{2}}{\lfloor\underline{2}}$$

$$\Rightarrow \lambda = \lambda^{2} \Rightarrow \lambda^{2} - \lambda = 0 \Rightarrow \lambda(\lambda - 1) = 0 \Rightarrow \lambda = 0, 1$$

But $\lambda = 0$ is rejected

[$\because$ if $\lambda = 0$ then either $n = 0$ or $p = 0$ which implies that Poisson distribution does not exist in this case.]

$\therefore \lambda = 1$

Hence mean $= \lambda = 1$, and

Variance $= \lambda = 1$.

Problem 1:

In a multiple-choice test, each question has 4 choices, and only one is correct. If you guess the answers randomly, what is the probability of getting at least 7 correct answers out of 10 questions?

Solution 1:

Let X be the number of correct answers out of 10 questions. Since each question has 4 choices and only one is correct, the probability of guessing a correct answer is $p = \frac{1}{4}$. Thus, $p = 0.25$ and $q = 1 - p = 1 - 0.25 = 0.75$.

We're interested in finding $P(X \geq 7)$.

$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$

Using the binomial probability formula:

$P(X = k) = \binom{n}{k} \times p^k \times q^{(n-k)}$

Where:

- $n = 10$ (number of trials/questions)
- $p = 0.25$ (probability of success)
- $q = 0.75$ (probability of failure)

$P(X \geq 7) = \sum_{k=7}^{10} \binom{10}{k} \times 0.25^k \times 0.75^{(10-k)}$

Calculate each term and sum them up to get the probability.

- **Prob.** The number of accidents at a particular intersection follows a Poisson distribution with a mean of 0.5 accidents per day. What is the probability that there are no accidents at this intersection tomorrow?
- Hint:

Here, we have:

- Average number of accidents (λ) = 0.5
- Number of accidents we're interested in (k) = 0

We apply the Poisson probability formula to find the probability of $k = 0$.