Question 1: What is a random variable in probability theory?

Answer:

A random variable is a function that assigns numerical values to the outcomes of a random experiment. It converts uncertain outcomes in measurable quantities, making them easier to analyze mathematically. In short, it is a set of possible values from a random experiment (values are unknown).

Example:-

If we toss a coin and define x = 1 for head appears, and x = 0 for Tail appears, then in this x is the random variable.

Characteristics:-

(i) it is defined on the sample space of a random experiment.

(ii) it assigns numerical values to outcomes.

(iii) It can be summarized using mean and variance.

(iv) it is linked with a probability distribution.

(v) helpful for converting real life uncertainty into a mathematical form.

Importance:-

(i) it provides a numerical framework for analysing uncertain events.

(ii) it helps for calculate averages, variability and probabilities.

(iii) it is very helpful for learning advanced topics like probability distributions.

(iv) others like make prediction, testing ideas with data etc.

Conclusion:-

A random variable is a way to represent uncertain outcomes in numerical form. It is not random itself but a function that assigns numbers to result of an experiment. This makes it easier to study probabilities and apply mathematical and statistical methods to real life problems.

Question 2: What are the types of random variables?

A random variable is a function that assigns numerical values to the outcomes of a random experiment. In simply we say it is a set of possible values from a random experiment. Random variables are categorized in to two main types, such as: discreate and continuous.

1. Discrete Random Variable: -

A discrete random variable can only take on a countable number of distinct values. These values are often integers and represent outcomes that can be counted, such as the number of heads in a series of coin flips or the number of cars that pass an intersection in an hour. In this series each value has an associated probability, described by the probability mass function or PMF.

Example:-

- Number of customers arriving at a shop in an hour.
- Rolling a dice and getting values 1, 2, 3, 4, 5, 6.

Properties:-

- It can often be shown in tables or bar graphs.
- Probabilities are assigned to exact values.

2. Continuous random Variable: -

A continuous random variable can take uncountably infinite values within a given range or interval. In these continuous random variable probabilities are describes by the Probability Density Function (PDF). The probability of the variable taking any exact value is zero, instead, probabilities are calculated over intervals.

Example:-

- Height or weight of students in a school.
- Time taken to complete a work.

Properties:-

- Probabilities are found using are under the curve of the PDF.
- Represent by smooth curves instead of discreate points.

Conclusion:-

Random variables are an important tool in probability and statistics because they connect uncertain outcomes to numerical values. They are mainly of two types: discrete random variables, which takes countable values like toss and continuous random variables, which take measurable values like height. This classification helps in applying appropriate probability rules and statistical models in real world problems.

Question 3: Explain the difference between discrete and continuous distributions.

A probability distribution shows how probabilities are assigned to the values of a random variable. Based on the type of random variable, the distributions are classified into discrete distributions and continuous distributions.

Discrete Distribution:-

A discrete distribution describes the probabilities of a discrete random variable. Such a variable takes only finite or countable values. Each outcome has a specific probability, and the sum of all probabilities is always equal to 1. The probability function used here is the Probability Mass Function (PMF).

Example:-

- Tossing a coin, roll a dice

Continuous Distribution:-

A continuous distribution describes the probabilities of a continuous random variable. Such a variable can take uncountably infinite values within an interval. The probability of taking any exact value is zero; instead, probabilities are calculated over a range of values. The probability function used here is the Probability Density Function (PDF). C

Example:-

- Height or weight of students, time taken to finish a work

Difference between discrete and continuous distribution:-

- Discrete distribution takes countable values but continuous takes uncountable values in an range or interval.
- Discrete distribution used Probability Mass Function to define things but Continuous distribution used Probability Density function to define things.
- In discreate distribution probability of a specific value can be greater than 0 where as in continuous distribution probability of a specific value is always 0.
- Discrete distribution shown by a table or bar graph but continuous distribution shown by a smooth curve.
- Discreate distribution is associated with a discrete random variable but continuous distribution is associated with a continuous random variable.
- Discrete distribution is specially use for counting problems but continuous distribution is use for measurement problems.

Question 4: What is a binomial distribution, and how is it used in probability?

The binomial distribution is a discrete probability distribution that gives the probability of obtaining a fixed number of successes in a given number of independent trials of a Bernoulli experiment. It has an experiment with only two possible outcomes such as success and failure.

Example:-

- Toss a coin 10 times and counting the number of heads. It says if a coin toss 10 times, what the possibility of K head appear for n times.

Characteristics of Binomial Distribution:-

- There are always fixed number of trials.
- Each trial have only two possible outcomes: success or failure
- All of the trials are completely independent from each other.
- The probability of success or failure remains the same in every trial.

Uses of Binomial Distribution Probability:-

- Helps in problems where outcomes are binary.
- Used to check the probability of finding defective items in a sample taken from production.
- Helps for to predict customer behaviour or satisfaction.
- Used in drug testing to estimate how many patients will show improvement after treatment.
- Determines the probability of a certain number of voters supporting a candidate in a sample survey.
- Calculates probabilities such as the chance of a player scoring a certain number of goals.
- Used to predict inheritance of traits.
- Estimates the probability of success/failure in repeated tasks of a project.

Question 5: What is the standard normal distribution, and why is it important?

The standard normal distribution is a special type of normal distribution where the mean is 0 and the standard deviation is 1. It is a bell-shaped, symmetric probability distribution centred at zero.

Key Features of Standard Normal Distribution:-

- Symmetrical about the mean (0).
- Mean value is 0, Standard Deviation value is 1.
- The total area under the curve is 1.
- Most values lie close to the mean, with probabilities decreasing as we move away.

Importance of Standard Normal Distribution:-

- Any normal distribution can be converted into the standard normal distribution using the z-score which is very helpful for calculating.
- Used in hypothesis testing, confidence intervals, and z-tests.
- Different normal distributions can be compared by standardizing data into z-scores.
- Standard normal distribution plays a key role in the Central Limit Theorem, which states that the distribution of sample means tends toward normality.

Examples of use cases:-

- Estimating the likelihood of stock returns within a given range.
- Finding the probability of a student scoring above a certain mark in exams.
- Determining quality control limits in manufacturing.

Conclusion:-

The standard normal distribution is the most important distribution in statistics. By standardizing data into **z-scores**, it allows us to calculate probabilities, compare data sets, and perform statistical tests. Its widespread applications in science, business, and research make it a central concept in probability theory.

Question 6: What is the Central Limit Theorem (CLT), and why is it critical in statistics?

The Central Limit Theorem (CLT) is one of the most important concepts in probability and statistics. It states that when we take independent random samples from any population that has a finite mean and variance, the distribution of the sample mean will become approximately normal as the sample size increases, regardless of the shape of the original population. This means that even if the population is skewed or irregular, the average of a large enough sample will follow a normal distribution.

Key Conditions for CLT:-

- The samples must be independent of each other.
- The sample size should be large, typically 30 or more, for the approximation to be accurate.
- The population must have a finite mean and finite variance.

Importance of CLT:-

- Forms the foundation of inferential statistics.
- Allows normal approximation even for non-normal populations.
- Basis for hypothesis testing (z-tests, t-tests, chi-square tests).
- Enables construction of confidence intervals.
- Applied in fields like quality control, opinion polls, economics, and medical research.

Example:-

- Using a class sample to estimate average student marks.
- Predicting election results from survey samples.
- Estimating the average income of households from a survey sample in economics.

Conclusion:-

The Central Limit Theorem is critical because it connects sample statistics to population parameters in a reliable way. It allows statisticians and researchers to make meaningful inferences, predictions, and decisions, even when the original data is not normally distributed. Its wide applicability in science, business, and research makes it one of the most fundamental concepts in statistics.

Question 7: What is the significance of confidence intervals in statistical analysis?

A confidence interval (CI) is a range of values calculated from sample data that is likely to contain the true value of a population parameter, such as the mean or proportion. It is usually expressed with a percentage like 90%, 95%, or 99%, which shows how confident we are that the interval contains the true value. For example, a 95% confidence interval for the average height of students means we are 95% sure that the true average height lies within that range.

Significance of Confidence Intervals:

- Estimate Population Values: They give a range instead of a single number, which shows possible values for the population.
- Measure Uncertainty: They show how precise or reliable our sample estimate is.
- Help in Decision Making: Used in business, medicine, and social studies to make informed decisions.
- Support Hypothesis Testing: Help check if a certain value is likely or not based on data.
- Compare Groups: Allow comparison between different groups to see if differences are significant.
- Easy to Interpret: Give both the range and likely direction of results, making findings easier to understand.

Examples:

- Estimating the average income of households in a city.
- Predicting weight loss in a medical study.
- Finding the proportion of voters supporting a candidate.

Conclusion:

Confidence intervals are important because they show a likely range for population values, help understand uncertainty, and support better decisions in real-life situations.

Question 8: What is the concept of expected value in a probability distribution?

The expected value is the average value you would expect from a random experiment if it is repeated many times. It represents the center or typical outcome of a probability distribution and helps us understand what is likely to happen on average.

For a discrete random variable, the expected value is found by multiplying each possible outcome by its probability and then adding them all together. For a continuous random variable, it is calculated using a formula that involves integrating over all possible values.

Significance of Expected Value:

- Shows the long-term average result of an experiment.
- Helps in decision making in business, finance, and everyday life.
- Summarizes a probability distribution with one meaningful number.
- Helps calculate variance and standard deviation, which measure how spread out the outcomes are.
- Assists in planning and risk assessment, such as knowing the average gain or loss.
- Useful in games of chance and gambling to decide fair bets.

Examples:

- Rolling a die: The expected value is 3.5, which is the average of all possible outcomes (1, 2, 3, 4, 5, 6).
- Buying a lottery ticket: Expected winnings tell if the ticket is likely worth buying.
- Stock market: Expected return shows the average profit or loss an investor can expect.
- Production: Expected number of defective items helps plan quality control.

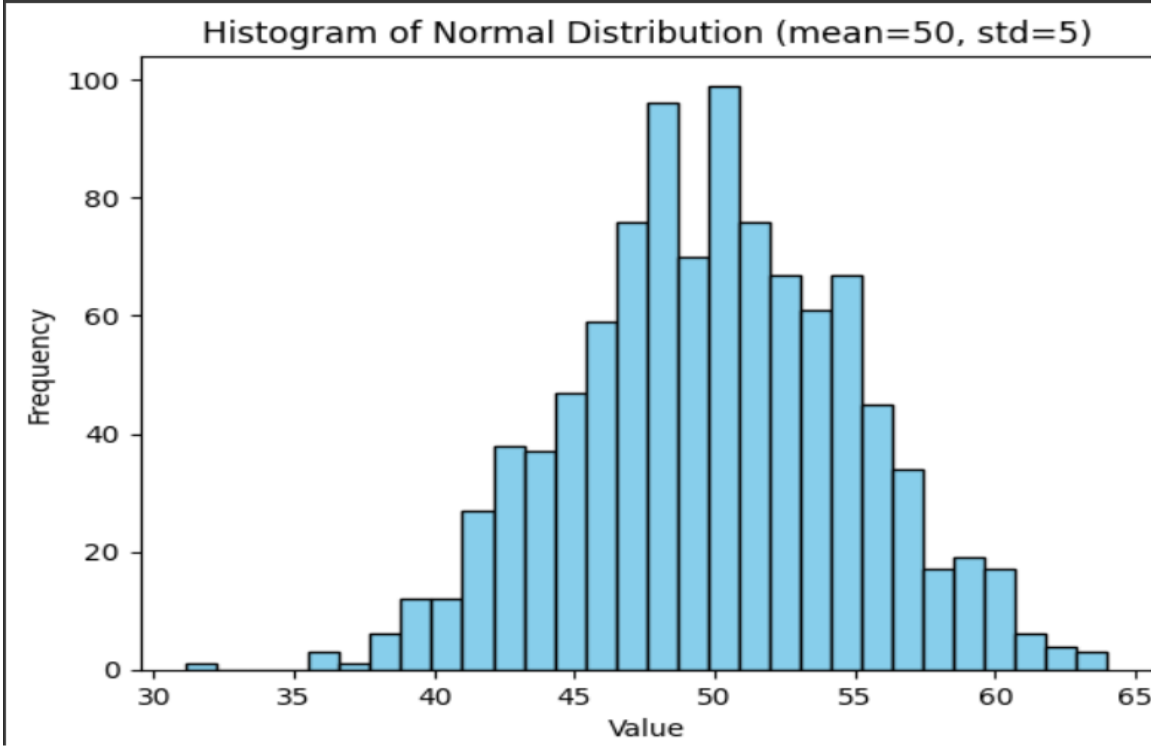
Conclusion:

The expected value is important because it gives a clear idea of the average outcome of a random process. It helps in decision making, planning, and understanding probability distributions, making it one of the most useful concepts in statistics.

Question 9: Write a Python program to generate 1000 random numbers from a normal distribution with mean = 50 and standard deviation = 5. Compute its mean and standard deviation using NumPy, and draw a histogram to visualize the distribution.

```python
import numpy as np
import matplotlib.pyplot as plt

mean = 50
std_dev = 5
n_samples = 1000

data = np.random.normal(loc=mean, scale=std_dev, size=n_samples)

calculated_mean = np.mean(data)
calculated_std = np.std(data)

print("Calculated Mean:", calculated_mean)
print("Calculated Standard Deviation:", calculated_std)

plt.hist(data, bins=30, color='skyblue', edgecolor='black')
plt.title("Histogram of Normal Distribution (mean=50, std=5)")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```

```
Calculated Mean: 49.867788599823804
Calculated Standard Deviation: 5.037726779796727
```



Question 10: You are working as a data analyst for a retail company. The company has collected daily sales data for 2 years and wants you to identify the overall sales trend. daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255, 235, 260, 245, 250, 225, 270, 265, 255, 250, 260] ● Explain how you would apply the Central Limit Theorem to estimate the average sales with a 95% confidence interval. ● Write the Python code to compute the mean sales and its confidence interval.

```python
import numpy as np
import scipy.stats as stats

daily_sales = [220, 245, 210, 265, 230, 250, 260, 275, 240, 255,
               235, 260, 245, 250, 225, 270, 265, 255, 250, 260]

data = np.array(daily_sales)
n = len(data)

mean_sales = np.mean(data)
```