Advance assignment

Q1. What is a random variable in probability theory?
Ans. A random variable is a way of turning the outcomes of a random experiment into numbers. It is really just a function or a map that takes each outcome from a sample space and assigns it a number. We normally use capital letters like X to represent a random variable.

Q2. what are the types of random variables?
Ans. Random Variables can be divided into two broad categories depending upon the type of data available. These are given as follows:

- Discrete random variable
- Continuous random variable

A probability mass function is used to describe a discrete random variable and a probability density function describes a continuous random variable. The upcoming sections will cover these topics in detail.

Q3. What is the difference between discrete and continuous distributions?
Ans. A discrete distribution is one in which the data can only take on certain values, for example integers. A continuous distribution is one in which data can take on any value within a specified range (which may be infinite). For a discrete distribution, probabilities can be assigned to the values in the distribution – for example, “the probability that the web page will have 12 clicks in an hour is 0.15.” In contrast, a continuous distribution has an infinite number of possible values, and the probability associated with any particular value of a continuous distribution is null. Therefore, continuous distributions are normally described in terms of probability density, which can be converted into the probability that a value will fall within a certain range.

Q4. What are probability distribution functions (PDF)?
Ans. To determine the distribution of a discrete random variable we can either provide its PMF or CDF. For continuous random variables, the CDF is well-defined so we can provide the CDF. However, the PMF does not work for continuous random variables, because for a continuous random variable

$P(X=x)=0$

$P(X=x)=0$ for all

$x \in R$

$x \in R$. Instead, we can usually define the probability density function (PDF). The PDF is the density of probability rather than the probability mass. The concept is very similar to mass density in physics: its unit is probability per unit length. To get a feeling for PDF, consider a continuous random variable

X

X and define the function

f

X

(x)

fX(x) as follows (wherever the limit exists):

Q5. How do cumulative distribution functions (CDF) differ from probability distribution functions (PDF)?
Ans.The primary difference between a Cumulative Distribution Function (CDF) and a Probability Distribution Function (PDF) lies in the information they provide. The PDF gives the probability of a specific value, while the CDF gives the probability that a value is less than or equal to a certain point. A PDF, or Probability Density Function, shows the likelihood of a continuous random variable assuming specific values. In contrast, a CDF, or Cumulative Distribution Function, represents the probability that the variable is less than or equal to a particular value

Q6. What is a discrete uniform distribution?
Ans. Uniform (Discrete) Distribution
In fields such as survey sampling, the discrete uniform distribution often arises because of the assumption that each individual is equally likely to be chosen in the sample on a given draw. The PMF of a discrete uniform distribution is given by pX=x=1n+1,x=0,1,…n, which implies that X can take any integer value between 0 and n with equal probability. The mean and variance of the distribution are n2 and nn+212.
To generate a random number from the discrete uniform distribution, one can draw a random number R from the U(0, 1) distribution, calculate S = (n + 1)R, and take the integer part of S as the draw from the discrete uniform distribution.

Q7. What are the key properties of a Bernoulli distribution?
Ans. Bernoulli Distribution is defined as a fundamental tool for calculating probabilities in scenarios where only two choices are present (i.e. binary situations), such as passing or failing, winning or losing, or a straightforward yes or no. Bernoulli Distribution can be resembled through the flipping of a coin. Binary situations involve only two possibilities: success or failure. For example, when flipping a coin, it can land on either heads, representing success; or tails, indicating failure. The likelihood of achieving heads is p, and the likelihood of getting tails is 1-p or q.

Q8. What is the binomial distribution, and how is it used in probability?
Ans. The binomial distribution is a probability distribution that models the probability of a certain number of successes in a fixed number of independent trials, where each trial has only two possible outcomes (success or failure) and a constant probability of success. In probability, it's used to calculate the likelihood of observing a specific number of successes in a series of experiments

Q9. What is the Poisson distribution and where is it applied?
Ans. The Poisson distribution is a discrete probability distribution used to model the probability of a certain number of events occurring within a fixed interval of time or space, given a known average rate. It's particularly useful when events are rare and independent of each other

Q10. What is a continuous uniform distribution?
Ans. Continuous uniform distributions have infinite distribution possibilities. An idealized random number generator would be considered a continuous uniform distribution. With this type of distribution, every point in the continuous range between 0.0 and 1.0 has an equal opportunity of appearing, yet there is an infinite number of points between 0.0 and 1.0.
There are several other important continuous distributions, such as the normal distribution, chi-square, and Student's t-distribution

Q11. What are the characteristics of a normal distribution?
Ans. A normal distribution, also known as a Gaussian distribution or bell curve, is characterized by a symmetric, bell-shaped curve, with the mean, median, and mode all equal and located at the center. It's continuous and unimodal (single peak), with the tails extending indefinitely but never touching the x-axis

Q12. What is the standard normal distribution, and why is it important?
Ans. The standard normal distribution is a specific normal distribution with a mean of 0 and a standard deviation of 1. It's crucial in statistics because it allows for easy comparisons and calculations across different normal distributions. The standard normal distribution serves as a reference point for understanding and interpreting the properties of other normal distributions

Q13. What is the Central Limit Theorem (CLT), and why is it critical in statistics?
Ans. The Central Limit Theorem (CLT) states that the distribution of sample means will approximate a normal distribution as the sample size increases, regardless of the original population distribution. This theorem is crucial in statistics because it allows us to apply many statistical methods that rely on the normal distribution, even when the underlying data is not normally distributed

Q14. How does the Central Limit Theorem relate to the normal distribution?
Ans.
The Central Limit Theorem (CLT) establishes a fundamental relationship between sample means and the normal distribution. It states that as the sample size increases, the distribution of sample means will approximate a normal distribution, regardless of the original population's distribution

Q15. What is the application of Z statistics in hypothesis testing?
Ans. Z-statistics are a cornerstone of hypothesis testing, providing a way to determine if there's a significant difference between a sample statistic (like the sample mean) and a population parameter (like the population mean) when the population variance is known, or if the sample size is large. They help researchers decide whether to reject the null hypothesis or accept the alternative hypothesis

Q16. How do you calculate a Z-score, and what does it represent?

Ans. To calculate a Z-score, you use the formula: $z = (x - \mu) / \sigma$, where 'x' is the raw score, 'μ' is the population mean, and 'σ' is the population standard deviatio
Explanation:

- x: The individual score you want to convert to a Z-score.
- μ: The mean (average) of the population.
- σ: The standard deviation of the population.
-

In simpler terms: A Z-score tells you how many standard deviations a particular score is away from the mean of its distribution.

Q17. What are point estimates and interval estimates in statistics?
Ans. In statistics, point estimation is using a single value from a sample to estimate a population parameter, while interval estimation involves creating a range of values within which the population parameter is likely to fall, with a certain level of confidence

Q18. What is the significance of confidence intervals in statistical analysis?
Ans. Confidence intervals are crucial in statistical analysis as they provide a range of plausible values for an unknown population parameter, based on sample data, along with a degree of confidence in that estimate. This range, often expressed with a 95% or 99% confidence level, offers more information about the precision of an estimate than a simple point estimate or p-value alone

Q19. What is the relationship between a Z-score and a confidence interval?
Ans. A Z-score and a confidence interval are closely related concepts in statistics. A Z-score is a standardized score that indicates how many standard deviations a data point is away from the mean, while a confidence interval is a range of values that is likely to contain the true population parameter. The Z-score is used in calculating the margin of error, which is a component of the confidence interval

Q20. How are Z-scores used to compare different distributions?
Ans. Z-scores are used to compare different distributions by standardizing data, meaning they transform data points to a common scale based on the mean and standard deviation of each distribution. This allows for a meaningful comparison of data points from different distributions, regardless of their original means and standard deviations.
Here's how z-scores facilitate comparison:

1. Standardization:
   A z-score indicates how many standard deviations a data point is from the mean of its respective distribution. For example, a z-score of 2 means the data point is two standard deviations above the mean.

2. Common Scale:
By converting data points to z-scores, they are all expressed on a common scale (z-scores), allowing for direct comparison across different distributions.
3. Identifying Extremes:
Z-scores help identify data points that are unusually high or low within their respective distributions, which can be useful in outlier detection or statistical analysis.
4. Probability Calculations:
Z-scores can be used to estimate the probability of a data point falling within a certain range of the mean, which can be helpful in making inferences about the data.

Q21. What are the assumptions for applying the Central Limit Theorem?
Ans. The Central Limit Theorem (CLT) has a few key assumptions to ensure its validity in practical applications. Primarily, the samples should be drawn independently and randomly from the population, and the sample size needs to be sufficiently large (often considered 30 or more). Additionally, the population from which the samples are drawn should have a finite variance

Q22. What is the concept of expected value in a probability distribution?
Ans. In a probability distribution, the expected value, also known as the mean or average, is a weighted average of all possible values of a random variable. It's the theoretical average you'd expect to get if you repeated the experiment or process generating the random variable many times. The weights in this average are the probabilities of each value occurring.

Q23. How does a probability distribution relate to the expected outcome of a random variable?
Ans. A probability distribution describes how probabilities are allocated to the possible values of a random variable. The expected value of a random variable is calculated by weighting each possible outcome by its probability, as defined by the probability distribution, and then summing these weighted outcomes. In essence, the probability distribution provides the framework for calculating the expected value, which represents the long-term average outcome of the random variable