09 March Assignment

May 10, 2023

```
[ ]: Q1: What are the Probability Mass Function (PMF) and Probability Density␣
       ↪Function (PDF)? Explain with
     an example.
```

```
[ ]: ANS -
```

```
[ ]: Probability Mass Function - Discrete

        The Probability Mass Function (PMF) provides the probability distribution␣
       ↪for discrete variables. For example, rolling dice.
     There are 6 distinct possible outcomes that define the entire sample space {1,␣
       ↪2, 3, 4, 5, 6}. Note that we only have whole numbers,
     i.e. no 1.2 or 3.75. In the PMF, each discrete variable is mapped to its␣
       ↪probability. In an ideal situation rolling a die, each of
     the six variables has a 1/6 probability of being rolled.
          While the ideal distribution is 1/6 for each possible outcome 1 to 6, this␣
       ↪rarely occurs in real-world scenarios.
     If we simulate an increasing number of trials we see that the distribution␣
       ↪doesn't match the ideal even with 100 dice rolls.
     As we increase orders of magnitude the simulated distribution approaches the␣
       ↪theoretical PMF.
```

```
[ ]: Probability Density Function - Continuous

        The probability density function (PDF) shows where observations are more␣
       ↪likely to occur in the probability distribution.
     Perhaps the most important thing to remember to understand PDFs is that the␣
       ↪probability of any specific outcome is 0. We have
     to think in terms of bins or ranges of values to calculate the probability of␣
       ↪seeing those values.
           We can use PDFs to calculate probability by looking at the area under the␣
       ↪curve for our interval. Mathematically, this is
     why a single point has 0 probability, the area of a point is 0. Because we have␣
       ↪a continuous variable, we can use integrals to
     calculate the area under the curve between our interval with the following␣
       ↪equation.
```

[ ]:

[ ]:

[ ]:
```
Q2: What is Cumulative Density Function (CDF)? Explain with an example. Why CDF is used?
```

[ ]:
```
ANS -
```

[ ]:
```
The cumulative distribution function, CDF, or cumulant is a function derived from the probability density function for a
continuous random variable. It gives the probability of finding the random variable at a value less than or equal to a given cutoff.
Many questions and computations about probability distribution functions are convenient to rephrase or perform in terms of CDFs,
e.g. computing the PDF of a function of a random variable.

    The cumulative distribution function (CDF) calculates the cumulative probability for a given x-value. Use the CDF to determine
the probability that a random observation that is taken from the population will be less than or equal to a certain value.
```

[ ]:

[ ]:

[ ]:
```
Q3: What are some examples of situations where the normal distribution might be used as a model?
Explain how the parameters of the normal distribution relate to the shape of the distribution.
```

[ ]:
```
ANS -
```

[ ]:
```
     Normal/Gaussian Distribution is a bell-shaped graph that encompasses two basic terms- mean and standard deviation.
It is a symmetrical arrangement of a data set in which most values cluster in the mean and the rest taper off symmetrically
towards either extreme. Numerous genetic and environmental factors influence the trait.
```

[ ]:
```
1. Height :
           The height of people is an example of normal distribution. Most of the people in a specific population are of average
height. The number of people taller and shorter than the average height people is almost equal, and a very small number of people
```

```
are either extremely tall or extremely short. Several genetic and environmental␣
  ↪factors influence height. Therefore, it follows the
normal distribution.
```

[ ]:
```
2. Rolling A Dice :
                    A fair rolling of dice is also a good example of normal␣
  ↪distribution. In an experiment, it has been found that
when a dice is rolled 100 times, chances to get '1' are 15-18% and if we roll␣
  ↪the dice 1000 times, the chances to get '1' is, again,
the same, which averages to 16.7% (1/6). If we roll two dice simultaneously,␣
  ↪there are 36 possible combinations. The probability of
rolling '1' (with six possible combinations) again averages to around 16.7%, i.
  ↪e., (6/36). More the number of dice more elaborate
will be the normal distribution graph.
```

[ ]:
```
3. Blood Pressure :
                Blood pressure generally follows a Gaussian distribution␣
  ↪(normal) in the general population, and it makes Gaussian
mixture models a suitable candidate for modelling blood pressure behaviour.
```

[ ]:
```
4. IQ :
       In this scenario of increasing competition, most parents, as well as␣
  ↪children, want to analyze the Intelligent Quotient level.
Well, the IQ of a particular population is a normal distribution curve; where␣
  ↪the IQ of a majority of the people in the population
lies in the normal range whereas the IQ of the rest of the population lives in␣
  ↪the deviated range.
```

[ ]:

[ ]:

[ ]:
```
Q4: Explain the importance of Normal Distribution. Give a few real-life␣
  ↪examples of Normal
Distribution.
```

[ ]:
```
ANS -
```

[ ]:
```
      The normal distribution is the most important probability distribution in␣
  ↪statistics because it fits many natural phenomena.
For example, heights, blood pressure, measurement error, and IQ scores follow␣
  ↪the normal distribution. It is also known as the
Gaussian distribution and the bell curve.
```

[ ]:
```
1. Height :
```

```
         The height of people is an example of normal distribution. Most of␣
 ↪the people in a specific population are of average
height. The number of people taller and shorter than the average height people␣
 ↪is almost equal, and a very small number of people
are either extremely tall or extremely short. Several genetic and environmental␣
 ↪factors influence height. Therefore, it follows the
normal distribution.
```

[ ]:
```
2. Rolling A Dice :
                   A fair rolling of dice is also a good example of normal␣
 ↪distribution. In an experiment, it has been found that
when a dice is rolled 100 times, chances to get '1' are 15-18% and if we roll␣
 ↪the dice 1000 times, the chances to get '1' is, again,
the same, which averages to 16.7% (1/6). If we roll two dice simultaneously,␣
 ↪there are 36 possible combinations. The probability of
rolling '1' (with six possible combinations) again averages to around 16.7%, i.
 ↪e., (6/36). More the number of dice more elaborate
will be the normal distribution graph.
```

[ ]:
```
3. Blood Pressure :
                Blood pressure generally follows a Gaussian distribution␣
 ↪(normal) in the general population, and it makes Gaussian
mixture models a suitable candidate for modelling blood pressure behaviour.
```

[ ]:
```
4. IQ :
       In this scenario of increasing competition, most parents, as well as␣
 ↪children, want to analyze the Intelligent Quotient level.
Well, the IQ of a particular population is a normal distribution curve; where␣
 ↪the IQ of a majority of the people in the population
lies in the normal range whereas the IQ of the rest of the population lives in␣
 ↪the deviated range.
```

[ ]:

[ ]:

[ ]:
```
Q5: What is Bernaulli DistributionGive an Example. What is the difference␣
 ↪between Bernoulli
Distribution and Binomial Distribution?
```

[ ]:
```
ANS -
```

[ ]:
```
      The Bernoulli distribution is a discrete probability distribution that␣
 ↪describes the probability of a random variable with only
two outcomes.
```

```
In the random process called a Bernoulli trial, the random variable can take
↪one outcome, called a success, with a probability p,
or take another outcome, called failure, with a probability q = 1-p.
 The success outcome is denoted as 1 and the failure outcome is denoted as 0.
The Bernoulli distribution is a special case of the binomial distribution where
↪a single trial is conducted and the binomial
distribution is the sum of repeated Bernoulli trials.
```

[ ]:
```
- Example :
           Tossing a coin can result in only two possible outcomes (head or
↪tail). We call one of these outcomes (head) a success and
the other (tail), a failure.
 The probability of success (p) or head is 0.5 for a fair coin. The probability
↪of failure (q) or tail = 1-p = 1-0.5 = 0.5.
    We have two outcomes:
Tail or 0 with a probability of 0.5.
Head or 1 with a probability of 0.5 also.
This is an example of a probability mass function where we have the probability
↪for each outcome.
```

[ ]:
```
     The difference between Bernoulli distribution and binomial distribution is
↪given below: Bernoulli distribution is used when we
want to model the outcome of a single trial of an event. If we want to model
↪the outcome of multiple trials of an event, Binomial
distribution is used. It is represented as X   Bernoulli (p).
```

[ ]:
```

```

[ ]:
```

```

[ ]:
```
Q6. Consider a dataset with a mean of 50 and a standard deviation of 10. If we
↪assume that the dataset
is normally distributed, what is the probability that a randomly selected
↪observation will be greater
than 60? Use the appropriate formula and show your calculations.
```

[ ]:
```
ANS -
```

[ ]:
```
z = (x -  ) /
```

[4]:
```
z = (60 - 50) / 10
```

[5]:
```
z
```

[5]: 1.0

[ ]:
```
15.87
```

[ ]:
```
```

[ ]:
```
```

[ ]:
```
Q7: Explain uniform Distribution with an example.
```

[ ]:
```
ANS -
```

[ ]:
```
     A continuous probability distribution is a Uniform distribution and is related to the events which are equally likely to occur.
It is defined by two parameters, x and y, where x = minimum value and y = maximum value. It is generally denoted by u(x, y).
            OR
If the probability density function or probability distribution of a uniform distribution with a continuous random variable X is
f(b)=1/y-x, then It is denoted by U(x,y), where x and y are constants such that x<a<y. It is written as
X
U(a,b)
```

[ ]:
```
```

[ ]:
```
```

[ ]:
```
Q8: What is the z score? State the importance of the z score.
```

[ ]:
```
ANS -
```

[ ]:
```
   A Z-score is an indicator of how closely a value relates to the mean of a set of values. Z-score is quantified by the standard
deviations from the mean. The mean score and the data points score are equal when the Z-score is zero. A Z-score of 1.0 indicates a
result that is one standard deviation from the mean. Z-scores are binary, with a positive value denoting a score above the mean and a
negative value denoting a score below the mean.
 Z-scores within the financial sector can be described as measures of the variability of an observation that traders can use to gauge
market volatility.
 The Z-score can also be known as the Altman Z-score model.
```

[ ]:
```
importance of the z score :

        The Z-score is important because it lets statisticians and traders know whether a score falls within the norm for a given data
```

```
set or deviates from it. Analysts can also modify scores from multiple data
 ↪sets using Z-scores to create scores that are more
accurately comparable to one another.
  Because they allow for a comparison between two scores that do not belong to
 ↪the same normal distribution, z-scores are significant.
They are also employed to determine the likelihood that a z-score will occur
 ↪inside a normal distribution. A negative z-score
indicates that the raw data is smaller than the mean. The raw score is higher
 ↪than the average if the z-score is positive.
  Among Z-scores benefits are:
    A z-score transformation takes into account the mean value as well as the
 ↪variability of a group or raw scores.Even though the raw
data came from various tests, it is still possible to compare them.
```

[ ]:

[ ]:

[ ]:
```
Q9: What is Central Limit Theorem? State the significance of the Central Limit
 ↪Theorem.
```

[ ]:
```
ANS -
```

[ ]:
```
        Central Limit Theorem Definition The Central Limit Theorem (CLT) states
 ↪that the distribution of a sample mean that
approximates the normal distribution, as the sample size becomes larger,
 ↪assuming that all the samples are similar, and no matter
what the shape of the population distribution.
     The central limit theorem is useful when analyzing large data sets because
 ↪it allows one to assume that the sampling distribution
of the mean will be normally-distributed in most cases. This allows for easier
 ↪statistical analysis and inference.
```

[ ]:

[ ]:

[ ]:
```
Q10: State the assumptions of the Central Limit Theorem.
```

[ ]:
```
ANS -
```

[ ]:
```
Assumptions of Central Limit Theorem :

    Assumptions of central limit theorem are stated below:

-The sample should be drawn randomly following the condition of randomization.
```

```
-The samples drawn should be independent of each other. They should not influence the other samples.
-When the sampling is done without replacement, the sample size shouldn't exceed 10% of the total population.
-The sample size should be sufficiently large.
```

[ ]:

```

```

[ ]:

```

```