**FSDS MAY BATCH 2022(ML Assignment -10)**

**Submitted by: Shubham Tiwari**

Q1: Define the Bayesian interpretation of probability.

Ans: In the Bayesian interpretation of probability, probability is understood as a measure of belief or degree of certainty about an event. It is subjective, and can be updated as new information becomes available. In this framework, probability is assigned to hypotheses or models, rather than to individual events, and is used to make predictions and decisions based on observed data.

Q2: Define probability of a union of two events with equation.

Ans: The probability of the union of two events A and B, denoted by P(A ∪ B), is defined as the sum of the probabilities of the individual events minus the probability of their intersection. The equation is as follows:

P(A ∪ B) = P(A) + P(B) - P(A ∩ B)

This equation is known as the inclusion-exclusion principle.

Q3: What is joint probability? What is its formula?

Ans: Joint probability is a measure of the likelihood of two or more events occurring together. It is the probability of all the events occurring at the same time. The formula for the joint probability of two events A and B, denoted by P(A and B), is as follows:

P(A and B) = P(A) \* P(B|A)

Where P(A) is the probability of event A, and P(B|A) is the probability of event B, given that event A has occurred.

For more than two events A, B, C, ..., the joint probability is given by:

P(A, B, C, ...) = P(A) \* P(B|A) \* P(C|A, B) \*

It's important to notice that the order of the events is relevant, and the formula change if the order is different.

Q4: What is chain rule of probability?

Ans: The chain rule of probability, also known as the "product rule," states that the probability of a sequence of events happening is the product of the probabilities of each individual event happening. More formally, if A1, A2, ..., An are a sequence of events, then:

P(A1 ∩ A2 ∩ ... ∩ An) = P(A1) \* P(A2 | A1) \* P(A3 | A1 ∩ A2) \* ... \* P(An | A1 ∩ A2 ∩ ... ∩ An-1)

In other words, the probability of all the events happening is equal to the probability of the first event happening, multiplied by the probability of the second event happening given that the first event has already happened, and so on. This rule is used to find the probability of combined events and is useful for solving problems involving conditional probability.

Q5: What is conditional probability means? What is the formula of it?

Ans: Conditional probability is the probability of an event occurring given that another event has already occurred. The formula for conditional probability is:

P(A|B) = P(A ∩ B) / P(B)

where A and B are events, P(A|B) is the probability of event A happening given that event B has already happened, P(A ∩ B) is the probability of events A and B happening together, and P(B) is the probability of event B happening.

It is important to note that P(B) must not be zero, otherwise the conditional probability is not defined.

This definition is useful in understanding the relationship between different events and how they are dependent on each other. For example, using conditional probability we can find the probability of getting a certain disease given some symptoms or getting a certain type of card given a certain suit.

Q6: What are continuous random variables?

Ans: A continuous random variable is a random variable whose set of possible values is an interval of real numbers. The probability that a continuous random variable takes on any particular value is always zero, but the probability that it takes on a value within a certain range of values can be calculated using probability density functions. Examples of continuous random variables include the length of a piece of steel, the weight of a bag of flour, or the temperature of a room.

Q7: What are Bernoulli distributions? What is the formula of it?

Ans: A Bernoulli distribution is a probability distribution that describes the outcome of a single trial of a Bernoulli experiment, which is a statistical experiment with only two possible outcomes, often labeled "success" and "failure". The Bernoulli distribution is a special case of the binomial distribution, where the number of trials is fixed at one.

The formula for the probability mass function of a Bernoulli distribution is:

P(X = x) = p^x \* (1-p)^(1-x)

where:

* P(X = x) is the probability of the outcome being x
* p is the probability of success (x = 1)
* (1-p) is the probability of failure (x = 0)

It is also represented as B(p) where p is the probability of success.

Q8: What is binomial distribution? What is the formula?

Ans: The binomial distribution is a probability distribution that describes the number of successes in a fixed number of Bernoulli trials, where each trial has only two possible outcomes: success or failure. The Bernoulli trials are assumed to be independent, and the probability of success is the same for each trial.

The formula for the probability mass function of a binomial distribution is:

P(X = k) = (n choose k) \* p^k \* (1-p)^(n-k)

where:

* P(X = k) is the probability of getting k successes in n trials
* n is the total number of trials
* k is the number of successes
* p is the probability of success on a single trial
* (n choose k) is the binomial coefficient, which is equal to n! / (k! \* (n-k)!)

It is also commonly represented as B(n,p) where n is the total number of trials and p is probability of success.

Q9: What is Poisson distribution? What is the formula?

Ans: The Poisson distribution is a probability distribution that describes the number of times an event occurs within a fixed interval of time or space, assuming the rate of events is constant and the events are independent of each other. It is typically used for modeling the number of occurrences of rare events, such as the number of telephone calls received by a call center in a given hour, or the number of customers arriving at a store in a given hour.

The formula for the probability mass function of a Poisson distribution is:

P(X = k) = (e^(-λ) \* λ^k) / k!

where:

* P(X = k) is the probability of k events occurring in the given interval.
* λ (lambda) is the average number of events that occur in the interval.
* e is the mathematical constant approximately equal to 2.718
* k! (k factorial) is the product of all the integers from 1 to k
* It is also represented as P(λ) where λ is the average rate of events in the interval.

Q10: Define covariance.

Ans: Covariance is a measure of the degree to which two random variables change together. It is used to indicate the strength and direction of the linear relationship between two variables.

Covariance is calculated by multiplying the difference of each variable's value from its mean, and then averaging those products. The resulting value can be positive, negative or zero.

A positive covariance indicates that the two variables tend to increase or decrease together, while a negative covariance indicates that the variables tend to move in opposite directions. A zero covariance indicates that there is no linear relationship between the two variables.

For example, the covariance between the variables x and y can be represented mathematically as Cov(x,y) = E((x - E(x))(y - E(y)))

Q11: Define correlation.

Ans: Correlation is a statistical measure that describes the strength and direction of a linear relationship between two variables. It is a value between -1 and 1 that indicates the extent to which the variables are related.

A correlation of 1 indicates a perfect positive correlation, meaning that as one variable increases, the other variable also increases, and as one variable decreases, the other variable also decreases. A correlation of -1 indicates a perfect negative correlation, meaning that as one variable increases, the other variable decreases, and vice versa. A correlation of 0 indicates no correlation, meaning that there is no linear relationship between the two variables.

Correlation is usually represented by the Greek letter rho (r), and it can be calculated using the formula for the sample correlation coefficient.It is useful to note that correlation doesn't imply causation, it only indicate that two variable are related in some way.

Q12: Define sampling with replacement. Give example.

Ans: Sampling with replacement refers to the process of randomly selecting items from a population and then allowing those items to be selected again in future draws.

**For example**, if we have a bag containing 5 red marbles and 3 blue marbles, and we wanted to randomly select 2 marbles with replacement, we would reach into the bag, randomly select a marble (e.g. red), place it aside, reach into the bag again and randomly select another marble (e.g. blue). The marbles that were selected (red and blue) are then returned to the bag before the next draw is made.

Q13: What is sampling without replacement? Give example.

Ans: Sampling without replacement refers to the process of selecting a certain number of items from a set, where each item can only be chosen once. An example would be randomly selecting 5 names from a list of 100 names without allowing any of the names to be chosen more than once.

Q14: What is hypothesis? Give example.

Ans: A hypothesis is a statement or assumption about a relationship or phenomenon that can be tested through scientific experimentation or observation. It is a proposed explanation for an observable event or pattern in nature. An example of a hypothesis could be "Plants grown with more sunlight will grow taller than those grown with less sunlight." This hypothesis can be tested by setting up an experiment where plants are grown under different levels of sunlight and measuring their heights to see if there is a correlation.