**1. Define the Bayesian interpretation of probability..**

[In machine learning, the Bayesian interpretation of probability is a learning technique that uses probabilities to define and reason about our beliefs](https://www.wolfram.com/language/introduction-machine-learning/bayesian-inference/)[1](https://www.wolfram.com/language/introduction-machine-learning/bayesian-inference/). [It’s essentially about refreshing our beliefs (priors) with new data (likelihood)](https://www.wolfram.com/language/introduction-machine-learning/bayesian-inference/)[2](https://www.machinelearningplus.com/probability/bayes-theorem/).

[The Bayesian interpretation of probability revolves around Bayes’ theorem, which is used to renew the probability estimate for a hypothesis as more evidence becomes available](https://www.wolfram.com/language/introduction-machine-learning/bayesian-inference/)[2](https://www.machinelearningplus.com/probability/bayes-theorem/). The theorem is defined as:

P (A|B) = \\frac {P (B|A) \\times P (A)} {P (B)}

Where:

* P (A|B): Posterior probability – the revised probability of eventAoccurring after observing eventB.
* P (A): Prior probability – our initial belief inA, prior to the new evidenceB.
* P (B|A): Likelihood – the odds of witnessing the evidenceB, assumingAis valid.
* P (B): Marginal likelihood or evidence – the overall probability of seeing evidenceB[2](https://www.machinelearningplus.com/probability/bayes-theorem/).

The three pillars of Bayesian inference are:

1. Prior (P (A)): Represents our pre-existing knowledge or belief about an event before new data.
2. Likelihood (P (B|A)): Depicts how well our data matches our predictions.
3. Posterior (P (A|B)): Our updated belief after integrating the new evidence. [It evolves from the prior and the likelihood](https://www.wolfram.com/language/introduction-machine-learning/bayesian-inference/)[2](https://www.machinelearningplus.com/probability/bayes-theorem/).

As new data surfaces, the posterior probability from one step can serve as the prior for the subsequent step. [This iterative method sharpens our inferences with accumulating evidence](https://www.wolfram.com/language/introduction-machine-learning/bayesian-inference/)[2](https://www.machinelearningplus.com/probability/bayes-theorem/). [This interpretation of probability is central to many machine learning algorithms, particularly those involved in probabilistic modeling and inference](https://www.wolfram.com/language/introduction-machine-learning/bayesian-inference/)[2](https://www.machinelearningplus.com/probability/bayes-theorem/).

**2. Define probability of a union of two events with equation.**

In machine learning, the probability of the union of two events, denoted byP (A \cup B)[, is the probability that at least one of the two events occurs1](https://www.atnyla.com/tutorial/probability-of-the-union-of-two-events/19/1349).

The formula to calculate the probability of the union of two events A and B is:

P (A \cup B) = P (A) + P (B) - P (A \cap B)

Where:

* P (A \cup B): Probability of the union of events A and B.
* P (A): Probability of event A.
* P (B): Probability of event B.
* P (A \cap B): Probability of the intersection of events A and B (i.e., both events A and B occur).

[This formula accounts for the fact that the sum of the probabilities of A and B individually might include the case where both A and B occur, hence we subtract the intersection to avoid double counting1](https://www.atnyla.com/tutorial/probability-of-the-union-of-two-events/19/1349).

**3. What is joint probability? What is its formula?**

[In machine learning, joint probability is a statistical metric that quantifies the likelihood of multiple events happening simultaneously1](https://www.machinelearningplus.com/probability/joint-probability/). [It’s used to measure the probability of two or more events occurring at the same time1](https://www.machinelearningplus.com/probability/joint-probability/).

[The formula for joint probability depends on whether the events are independent or dependent1](https://www.machinelearningplus.com/probability/joint-probability/):

* [**Independent Events**: If events A and B are independent (the outcome of one doesn’t influence the other), the joint probability is the product of their individual probabilities1](https://www.machinelearningplus.com/probability/joint-probability/):

P(A \cap B) = P(A) \times P(B)

* [**Dependent Events**: If events A and B are dependent (the outcome of one influences the other), the joint probability is given by1](https://www.machinelearningplus.com/probability/joint-probability/):

P(A \cap B) = P(A) \times P(B|A)

Where:

* P(A \cap B)is the joint probability of A and B.
* P(A)is the probability of event A.
* P(B)is the probability of event B.
* P(B|A)[is the conditional probability of event B given that A has already occurred1](https://www.machinelearningplus.com/probability/joint-probability/).

Joint probability is foundational in various sectors, from finance to artificial intelligence. [It assists experts in risk assessment, predictive analysis, and decoding complex scenarios with intertwined events1](https://www.machinelearningplus.com/probability/joint-probability/).

**4. What is chain rule of probability?**

[In machine learning, the chain rule (also known as the general product rule) allows the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities1](https://www.hackerearth.com/practice/machine-learning/prerequisites-of-machine-learning/bayes-rules-conditional-probability-chain-rule/tutorial/).

[The chain rule of probability states1](https://www.hackerearth.com/practice/machine-learning/prerequisites-of-machine-learning/bayes-rules-conditional-probability-chain-rule/tutorial/):

P(E\_1 \cap E\_2 \cap ... \cap E\_n) = P(E\_n | E\_1, E\_2, ..., E\_{n-1}) \times P(E\_{n-1} | E\_1, E\_2, ..., E\_{n-2}) \times ... \times P(E\_2 | E\_1) \times P(E\_1)

Where:

* P(E\_1 \cap E\_2 \cap ... \cap E\_n)is the joint probability of eventsE\_1, E\_2, ..., E\_n.
* P(E\_n | E\_1, E\_2, ..., E\_{n-1})is the conditional probability of eventE\_ngiven that eventsE\_1, E\_2, ..., E\_{n-1}have occurred.

[This rule can be used iteratively to calculate the joint probability of any number of events1](https://www.hackerearth.com/practice/machine-learning/prerequisites-of-machine-learning/bayes-rules-conditional-probability-chain-rule/tutorial/). [It’s a fundamental principle in probability theory and provides the basis for the construction of Bayesian networks and other probabilistic models1](https://www.hackerearth.com/practice/machine-learning/prerequisites-of-machine-learning/bayes-rules-conditional-probability-chain-rule/tutorial/).

**5. What is conditional probability means? What is the formula of it?**

[Conditional probability is the likelihood of an event occurring, given that another event has already occurred1](https://byjus.com/maths/conditional-probability/)[2](https://www.geeksforgeeks.org/conditional-probability/). [It quantifies the adjusted probability based on prior knowledge, offering insights into the interdependence of events3](https://testbook.com/maths-formulas/conditional-probability-formula).

Mathematically, conditional probability is represented as

P(A|B)

[, which means the probability of event A occurring when event B has already occurred](https://byjus.com/maths/conditional-probability/)[2](https://www.geeksforgeeks.org/conditional-probability/).

[The formula for conditional probability is given by](https://byjus.com/maths/conditional-probability/)[4](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/conditional-probability-definition-examples/):

P(B|A) = \frac{P(A \cap B)}{P(A)}

Here,

* P(A \cap B)

**6. What are continuous random variables?**

In machine learning (ML) and statistics, random variables are variables whose values are subject to variability or randomness. Random variables can be categorized as either discrete or continuous, depending on the nature of their possible values.

Continuous random variables are variables that can take any value within a certain range or interval. Unlike discrete random variables, which can only take on distinct, separate values, continuous random variables can take on an infinite number of values within a specified range. The possible values of a continuous random variable form a continuous spectrum.

For example, the height of a person, the temperature in a room, or the time it takes for a process to complete are all examples of continuous random variables. These variables can theoretically assume any value within a certain range, and their probability distributions are described using probability density functions (PDFs) rather than probability mass functions.

In the context of machine learning, understanding and modeling continuous random variables is essential for tasks such as regression, where the goal is to predict a continuous outcome rather than a categorical one. Techniques like linear regression or support vector regression are commonly used for predicting continuous variables based on input features.

**7. What are Bernoulli distributions? What is the formula of it?**

[Bernoulli Distribution is a type of discrete probability distribution where every experiment conducted asks a question that can be answered only in yes or no1](https://www.cuemath.com/data/bernoulli-distribution/). [In other words, the random variable can be 1 with a probability p or it can be 0 with a probability (1 - p)1](https://www.cuemath.com/data/bernoulli-distribution/). [Such an experiment is called a Bernoulli trial1](https://www.cuemath.com/data/bernoulli-distribution/).

[The formulas for Bernoulli distribution are given by the probability mass function (pmf) and the cumulative distribution function (CDF)1](https://www.cuemath.com/data/bernoulli-distribution/).

[The probability mass function (pmf) of a Bernoulli distribution is given by1](https://www.cuemath.com/data/bernoulli-distribution/):

P(X = x) =  
\begin{cases}  
p & \text{if } x = 1 \\  
1 - p & \text{if } x = 0  
\end{cases}

[Here, X is the random variable following a Bernoulli Distribution, p is the probability of success (X = 1), and 1 - p is the probability of failure (X = 0)1](https://www.cuemath.com/data/bernoulli-distribution/).

[The cumulative distribution function (CDF) of a Bernoulli distribution is given by2](https://en.wikipedia.org/wiki/Bernoulli_distribution):

P(X \leq x) =  
\begin{cases}  
0 & \text{if } x < 0 \\  
1 - p & \text{if } 0 \leq x < 1 \\  
1 & \text{if } x \geq 1  
\end{cases}

[The mean of a Bernoulli distribution is E[X] = p and the variance is Var[X] = p(1 - p)](https://www.cuemath.com/data/bernoulli-distribution/)[3](https://codinghero.ai/bernoulli-distribution/).

**8. What is binomial distribution? What is the formula?**

The binomial distribution is a discrete probability distribution that describes the number of successes (usually denoted as "x") in a fixed number of independent and identically distributed Bernoulli trials. A Bernoulli trial is an experiment with only two possible outcomes: success (usually labeled as 1) or failure (usually labeled as 0).

The probability mass function (PMF) of a binomial distribution is given by the binomial probability formula:

\[ P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n - x} \]

Here:

- \( P(X = x) \) is the probability of observing \( x \) successes in \( n \) trials.

- \( \binom{n}{x} \) is the binomial coefficient, representing the number of ways to choose \( x \) successes out of \( n \) trials, and is calculated as \( \frac{n!}{x!(n-x)!} \).

- \( p \) is the probability of success in a single trial.

- \( (1 - p) \) is the probability of failure in a single trial.

- \( n \) is the total number of trials.

In summary, the binomial distribution models the number of successes in a fixed number of independent trials with the same probability of success in each trial. The distribution is denoted as \( B(n, p) \), where \( n \) is the number of trials and \( p \) is the probability of success in each trial.

**9. What is Poisson distribution? What is the formula?**

[The Poisson distribution is a discrete probability distribution that describes the probability of a given number of events occurring in a fixed interval of time or space1](https://byjus.com/maths/poisson-distribution/)[2](https://www.geeksforgeeks.org/poisson-distribution/). [These events must occur with a known constant mean rate and independently of the time since the last event1](https://byjus.com/maths/poisson-distribution/)[2](https://www.geeksforgeeks.org/poisson-distribution/).

[The formula for the Poisson distribution is given by1](https://byjus.com/maths/poisson-distribution/)[2](https://www.geeksforgeeks.org/poisson-distribution/):

P(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}

Here:

* P(x; \lambda) is the probability that an event will occur x times,
* x is the number of times an event occurs,
* λ is the average number of times an event occurs,
* e is Euler’s constant (approximately equal to 2.718).

[In the Poisson distribution, both the mean (average) and variance are equal and are denoted by the parameter λ (lambda)1](https://byjus.com/maths/poisson-distribution/)[2](https://www.geeksforgeeks.org/poisson-distribution/). [This property of equal mean and variance is a distinctive characteristic of the Poisson distribution and simplifies its statistical analysis1](https://byjus.com/maths/poisson-distribution/)[2](https://www.geeksforgeeks.org/poisson-distribution/).

**10. Define covariance.**

[Covariance is a statistical measure that indicates the extent to which two random variables change in tandem1](https://byjus.com/maths/covariance/)[2](https://en.wikipedia.org/wiki/Covariance). [It’s used to gauge the linear relationship between two variables](https://byjus.com/maths/covariance/)[2](https://en.wikipedia.org/wiki/Covariance).

[If greater values of one variable mainly correspond with greater values of the other variable, and the same holds for lesser values (that is, the variables tend to show similar behavior), the covariance is positive](https://byjus.com/maths/covariance/)[2](https://en.wikipedia.org/wiki/Covariance). [In the opposite case, when greater values of one variable mainly correspond to lesser values of the other (that is, the variables tend to show opposite behavior), the covariance is negative](https://byjus.com/maths/covariance/)[2](https://en.wikipedia.org/wiki/Covariance).

[The formula for covariance is given by1](https://byjus.com/maths/covariance/):

Cov(X,Y) = E[(X - \mu\_X)(Y - \mu\_Y)]

Here:

* Cov(X,Y) is the covariance between the variables X and Y,
* E denotes the expected value,
* X and Y are the random variables,
* μ\_X and μ\_Y are the means of X and Y respectively.

[This formula calculates the average of the product of the differences of each variable from their mean1](https://byjus.com/maths/covariance/). [The units of measurement of the covariance are those of X times those of Y2](https://en.wikipedia.org/wiki/Covariance).

**11. Define correlation**

[Correlation is a statistical measure that describes the degree of association between two or more variables1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation). [It quantifies the strength and direction of the linear relationship between these variables1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation).

[The correlation coefficient, often denoted by r, is a summary measure that describes the extent of this relationship1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation). [The correlation coefficient ranges from -1 to +11](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation).

* [A correlation coefficient close to +1 indicates a strong positive relationship, meaning that as one variable increases, the other variable also tends to increase1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation).
* [A correlation coefficient close to -1 indicates a strong negative relationship, meaning that as one variable increases, the other variable tends to decrease](https://www.bing.com/aclk?ld=e8dywAupo034navdQFDFCgSzVUCUzjj70w7N_fJVhkPEn9U7L_-gwUXlo_GRGv44agMmalXiq7-xaxfEp1hlc0mJkW_VSl6Z04htA1hssVzfzp733h4qyGQIHhZjmnYnWLMwDxbCdHEbPmdBgvoNSr52FuYmcfVJsPs8SDnXdLE1GYw1ed&u=aHR0cHMlM2ElMmYlMmZsZWFybi5maW5hbmNlc3RyYXRlZ2lzdHMuY29tJTJmZmluYW5jZS10ZXJtcyUyZmNvcnJlbGF0aW9uJTJm&rlid=718b4cf9b6e110fecae7293451d8ffb0)[1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation).
* [A correlation coefficient close to 0 indicates little or no linear relationship between the variables](https://www.bing.com/aclk?ld=e8dywAupo034navdQFDFCgSzVUCUzjj70w7N_fJVhkPEn9U7L_-gwUXlo_GRGv44agMmalXiq7-xaxfEp1hlc0mJkW_VSl6Z04htA1hssVzfzp733h4qyGQIHhZjmnYnWLMwDxbCdHEbPmdBgvoNSr52FuYmcfVJsPs8SDnXdLE1GYw1ed&u=aHR0cHMlM2ElMmYlMmZsZWFybi5maW5hbmNlc3RyYXRlZ2lzdHMuY29tJTJmZmluYW5jZS10ZXJtcyUyZmNvcnJlbGF0aW9uJTJm&rlid=718b4cf9b6e110fecae7293451d8ffb0)[1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation).

[It’s important to note that correlation measures association, not causation](https://www.bing.com/aclk?ld=e8dywAupo034navdQFDFCgSzVUCUzjj70w7N_fJVhkPEn9U7L_-gwUXlo_GRGv44agMmalXiq7-xaxfEp1hlc0mJkW_VSl6Z04htA1hssVzfzp733h4qyGQIHhZjmnYnWLMwDxbCdHEbPmdBgvoNSr52FuYmcfVJsPs8SDnXdLE1GYw1ed&u=aHR0cHMlM2ElMmYlMmZsZWFybi5maW5hbmNlc3RyYXRlZ2lzdHMuY29tJTJmZmluYW5jZS10ZXJtcyUyZmNvcnJlbGF0aW9uJTJm&rlid=718b4cf9b6e110fecae7293451d8ffb0)[1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation). [Therefore, a high correlation between two variables does not imply that one variable causes the other to change1](https://byjus.com/maths/correlation/)[2](https://en.wikipedia.org/wiki/Correlation).

**12. Define sampling with replacement. Give example.**

[Sampling with replacement is a method used in statistical sampling where an individual or item, once selected, can be chosen again1](https://www.statology.org/sampling-with-vs-without-replacement/)[2](https://www.statisticshowto.com/sampling-with-replacement-without/). [This means that each selection is independent of the others1](https://www.statology.org/sampling-with-vs-without-replacement/)[2](https://www.statisticshowto.com/sampling-with-replacement-without/). [The probability of choosing any individual or item remains constant and does not change over the course of the sampling process1](https://www.statology.org/sampling-with-vs-without-replacement/)[2](https://www.statisticshowto.com/sampling-with-replacement-without/).

For example, suppose we have a bag with 5 balls, each of a different color: Red, Blue, Green, Yellow, and Black. If we were to sample with replacement, we would select a ball, note its color, and then put it back into the bag before drawing again. [So, it’s possible that we could select the same ball more than once1](https://www.statology.org/sampling-with-vs-without-replacement/)[2](https://www.statisticshowto.com/sampling-with-replacement-without/).

[In this scenario, the probability of drawing any particular color remains constant at 1/5 for each draw, regardless of what colors have been drawn previously1](https://www.statology.org/sampling-with-vs-without-replacement/)[2](https://www.statisticshowto.com/sampling-with-replacement-without/). [This is because the ball, once drawn, is returned to the bag, keeping the total number of balls constant1](https://www.statology.org/sampling-with-vs-without-replacement/)[2](https://www.statisticshowto.com/sampling-with-replacement-without/).

**13. What is sampling without replacement? Give example.**

[Sampling without replacement is a statistical method where an individual or item, once selected, cannot be chosen again1](https://stats.libretexts.org/Courses/Diablo_Valley_College/Math_142%3A_Elementary_Statistics_%28Kwai-Ching%29/Math_142%3A_Course_Material/05%3A_Chapter_5/Ch_3.4_Sampling_with_without_replacement)[2](https://www.statology.org/sampling-with-vs-without-replacement/). [This means that each selection is dependent on the others](https://stats.libretexts.org/Courses/Diablo_Valley_College/Math_142%3A_Elementary_Statistics_%28Kwai-Ching%29/Math_142%3A_Course_Material/05%3A_Chapter_5/Ch_3.4_Sampling_with_without_replacement)[2](https://www.statology.org/sampling-with-vs-without-replacement/). [The probability of choosing any individual or item changes over the course of the sampling process1](https://stats.libretexts.org/Courses/Diablo_Valley_College/Math_142%3A_Elementary_Statistics_%28Kwai-Ching%29/Math_142%3A_Course_Material/05%3A_Chapter_5/Ch_3.4_Sampling_with_without_replacement)[2](https://www.statology.org/sampling-with-vs-without-replacement/).

For example, suppose we have a hat with 5 names: Andy, Karl, Tyler, Becca, and Jessica. If we were to sample without replacement, we would select a name, note it, and then leave it out of the hat before drawing again. So, on the first draw, we might select the name Tyler. We would then leave his name out of the hat. On the second draw, we might select the name Andy. [Thus our sample would be: {Tyler, Andy}](https://stats.libretexts.org/Courses/Diablo_Valley_College/Math_142%3A_Elementary_Statistics_%28Kwai-Ching%29/Math_142%3A_Course_Material/05%3A_Chapter_5/Ch_3.4_Sampling_with_without_replacement)[2](https://www.statology.org/sampling-with-vs-without-replacement/).

In this scenario, the probability of drawing any particular name changes with each draw. [For example, the probability of choosing the name Tyler is 1/5 on the first draw and the probability of choosing the name Andy is 1/4 on the second draw2](https://www.statology.org/sampling-with-vs-without-replacement/). [This is because the name, once drawn, is not returned to the hat, reducing the total number of names2](https://www.statology.org/sampling-with-vs-without-replacement/).

**14. What is hypothesis? Give example.**

[A hypothesis is an educated guess or proposed explanation for a phenomenon, based on some initial observations or data1](https://researchmethod.net/what-is-a-hypothesis/). [It is a tentative statement that can be tested and potentially proven or disproven through further investigation and experimentation1](https://researchmethod.net/what-is-a-hypothesis/). [Hypotheses are often used in scientific research to guide the design of experiments and the collection and analysis of data1](https://researchmethod.net/what-is-a-hypothesis/).

[For example, a researcher studying the effects of sugar on dental health might propose the following hypothesis: “If a person consumes a high amount of sugar, then they will have a higher risk of developing cavities.” In this case, the consumption of sugar is the independent variable, and the risk of developing cavities is the dependent variable](https://researchmethod.net/what-is-a-hypothesis/)[2](https://www.indeed.com/career-advice/career-development/what-is-a-hypothesis). [The researcher can then design an experiment to test this hypothesis by comparing the dental health of individuals with high and low sugar diets](https://researchmethod.net/what-is-a-hypothesis/)[2](https://www.indeed.com/career-advice/career-development/what-is-a-hypothesis).

[There are several types of hypotheses, including the research hypothesis, null hypothesis, alternative hypothesis, directional hypothesis, non-directional hypothesis, statistical hypothesis, composite hypothesis, and empirical hypothesis1](https://researchmethod.net/what-is-a-hypothesis/). [Each type of hypothesis serves a different purpose and is used in different contexts within scientific research1](https://researchmethod.net/what-is-a-hypothesis/).