1. What is prior probability? Give an example.

Ans: The prior probability is the [probability](https://www.statlect.com/fundamentals-of-probability/probability) assigned to an event before the arrival of some information that makes it necessary to revise the assigned probability.

Example:

Suppose that an individual is extracted at random from a population consisting of two ethnic groups, ABC and XYZ.

We know that:

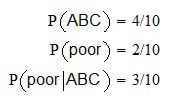
30% of individuals belonging to group ABC have incomes below the poverty line;

the corresponding proportion for the population as a whole is 20%;

40% of the population is made of individuals belonging to group ABC.

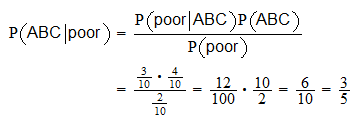
If we extract an individual whose income is below the poverty line, what is the probability that she belongs to group ABC?

This conditional probability can be computed with Bayes' rule.

The quantities involved in the computation are

The prior probability is **P(ABC)** the probability of belonging to group ABC.

The posterior probability **P(ABC|poor)** can be computed thanks to Bayes' rule:



2. What is posterior probability? Give an example.

Ans:

The posterior probability is one of the quantities involved in [Bayes' rule](https://www.statlect.com/fundamentals-of-probability/Bayes-rule).It is the conditional probability of a given event, computed after observing a second event whose conditional and unconditional probabilities were known in advance.

## Definition

The following is a more formal definition.

**Definition** Let A and B be two events whose probabilities P(A) and P(B) are known. If also the [conditional probability](https://www.statlect.com/fundamentals-of-probability/conditional-probability) P(B/A) is known, Bayes' rule gives

[eq4]

The conditional probability P(A/B) thus computed is called posterior probability.

In other words, the posterior probability is the conditional probability P(A/B calculated after receiving the information that the event B has happened.

Example:

Suppose that an individual is extracted at random from a population of men.

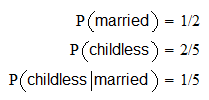
We know the following things::

* the probability of extracting a married individual is 50%;
* the probability of extracting a childless individual is 40%;
* the conditional probability that an individual is childless given that he is married is equal to 20%.

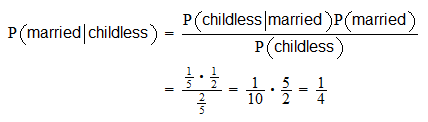
If the individual extracted at random from the population turns out to be childless, what is the conditional probability that he is married?

This conditional probability is called posterior probability and it can be computed by using Bayes' rule above.

The quantities involved in the computation are



The posterior probability is



3. What is likelihood probability? Give an example.

Ans:

Likelihood refers to the process of *determining the best data distribution* given a specific situation in the [data](https://www.simplilearn.com/what-is-data-article).

When calculating the probability of a given outcome, you assume the model's parameters are reliable.

However, when you calculate the likelihood, you’re attempting to determine whether the parameters in a model can be trusted based on the [sample data](https://www.simplilearn.com/tutorials/machine-learning-tutorial/population-vs-sample) you have observed.

Example Scenario

Suppose you have an unbiased coin. If you flip the coin, the probability of getting head and a tail is equal, which is 0.5

Now suppose the same coin is tossed 50 times, and it shows heads only 14 times. You would assume that the likelihood of the unbiased coin is very low. If the coin were fair, it would have shown heads and tails the same number of times.

4. What is Naïve Bayes classifier? Why is it named so?

Ans:

Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.

* It is mainly used in *text classification* that includes a high-dimensional training dataset.
* Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
* **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**.
* Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

**Naïve**: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the basis of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other

5. What is optimal Bayes classifier?

Ans:

Bayes Optimal Classifier is a probabilistic model that finds the most probable prediction using the training data and space of hypotheses to make a prediction for a new data instance.

It is described using the Bayes Theorem that provides a principled way for calculating a conditional probability. It is also closely related to the Maximum a Posteriori: a probabilistic framework referred to as MAP that finds the most probable hypothesis for a training dataset.

6. Write any two features of Bayesian learning methods.

Ans:

**Features of Bayesian learning methods:**

• Each observed training example can incrementally decrease or increase the estimated probability that a hypothesis is correct.

– This provides a more flexible approach to learning than algorithms that completely eliminate a hypothesis if it is found to be inconsistent with any single example.

• Prior knowledge can be combined with observed data to determine the final probability of a hypothesis. In Bayesian learning, prior knowledge is provided by asserting – a prior probability for each candidate hypothesis, and – a probability distribution over observed data for each possible hypothesis.

• Bayesian methods can accommodate hypotheses that make probabilistic predictions

• New instances can be classified by combining the predictions of multiple hypotheses, weighted by their probabilities.

• Even in cases where Bayesian methods prove computationally intractable, they can provide a standard of optimal decision making against which other practical methods can be measured.

7. Define the concept of consistent learners.

Ans:

Consistent Learners. • A learner L using a hypothesis H and training data D is said to be a consistent learner if it always outputs a hypothesis with zero error on D whenever H contains such a hypothesis. • By definition, a consistent learner must produce a hypothesis in the version space for H given D

8. Write any two strengths of Bayes classifier.

Ans:

* It is simple and easy to implement.
* It doesn't require as much training data.
* It handles both continuous and discrete data.
* It is highly scalable with the number of predictors and data points.
* It is fast and can be used to make real-time predictions.

9. Write any two weaknesses of Bayes classifier.

Ans:

* Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. ...
* This algorithm faces the 'zero-frequency problem' where it assigns zero probability to a categorical variable whose category in the test data set wasn't available in the training dataset.

10. Explain how Naïve Bayes classifier is used for

1. Text classification

Ans:

Bayesian statistics gives the leverage of the changing probabilities which can happen prior and post a certain event.

The Naive Bayesian classifier consists of performing the below steps –

* Create a frequency table based on the words
* Calculate the likelihood for each of the classes based on the frequency table
* Calculate the posterior probability for each class
* The highest posterior probability is the outcome of the prediction experiment

All these probabilities are calculated by using the Bayes Theorem. As the Naive Bayes algorithm has the assumption of the “Naive” features it performs much better than other algorithms like Logistic Regression, Tree based algorithms etc. The Naive Bayes classifier is much faster with its probability calculations.

Different kinds of Naive Bayesian implementations exist –

Gaussian Naive Bayes

This is the kind of algorithm used when all features follow a normal distribution. All features are continuous valued. The assumption is that there is no covariance between the independent features.

Multinomial Naive Bayes

It is generally used where there are discrete features(for example – word counts in a text classification problem). It generally works with the integer counts which are generated as frequency for each word. All features follow multinomial distribution. In such cases TF-IDF(Term Frequency, Inverse Document Frequency) also works.

Bernoulli Naive Bayes

This classifier also works with discrete data. The major difference between Multinomial Naive Bayes and Bernoulli is that Multinomial Naive Bayes works with occurrence counts while Bernoulli works with binary/boolean features. For example, the feature values are of the form true/false, yes/no, 1/0 etc. This is best visualized with the help of a histogram.

Different variations of the Naive Bayes classifier all work with the same analogy of independence of features. The way the different types of Naive Bayesian classifiers have been designed they work very well on all kinds of text related problems. Document classification is one such example of a text classification problem which can be solved by using both Multinomial and Bernoulli Naive Bayes. The calculation of probabilities is the major reason for this algorithm to be a text classification friendly algorithm and a top favorite among the masses. This classifier is highly used for predictions in real-time and also used in recommendation systems along with collaborative filtering.

2. Spam filtering

Ans:

A spam filter is a program used to detect unsolicited, unwanted and virus-infected emails and prevent those messages from getting to a user's inbox. Like other types of filtering programs, a spam filter looks for specific criteria on which to base its judgments.

Spam filters all have the same basic objective: to keep unwanted emails out of users’ inboxes. However, there are several different types of spam filters, and they each use different filtering methods to hone in on spam.

Content Filters

Content filters analyze the text inside an email and use that information to decide whether or not to mark it as spam. The content of spam emails is often predictable, particularly because they tend to have the same basic objectives: offer deals, promote explicit material, or otherwise tap into human emotions, feelings, and desires, such as greed or fear.

Content filters may search for words connected to money, such as “discount,” “limited time,” or “offer.” To trigger the filter, there typically would have to be multiple uses of the target word.

Content filters may also examine an email for inappropriate language of a sexual nature that could indicate explicit content. In some campaigns, an attacker may use sexually explicit emails to lure users into opening the email and then clicking on malicious links.

Blacklist Filters

Blacklist email spam filters work by blocking emails from senders that have been put on a list of spammers. Blacklist filters are updated on a regular basis because spammers can change their email addresses relatively easily. If a spammer switches from one email domain to another, the email may still be able to penetrate the filter until it is updated and the sender’s emails once again get labeled as spam.

A company can also use its own blacklist spam filtering to protect its interests. For example, they can use them to target headhunters seeking to attract their employees to other companies. They could also use a blacklist filter to block emails that could waste employees’ time with sales offers and promotions that could distract them from getting their work done.

Header Filters

Header filters examine the header of an email to see if it may be coming from an illegitimate source. This could include Internet Protocol (IP) addresses that spammers tend to use. It may also include information that indicates the email is just one copy of many emails sent at the same time to pre-organized groups of recipients.

Language Filters

Sometimes spammers target people from other countries, and the email is therefore in a different language than that of the recipient. In most cases, a user will only want to receive emails in languages in which they are fluent.

However, if a business connection or customer from another country reaches out, there exists the chance that the language filter could categorize that legitimate email as spam, so users may have to be instructed to check their spam folders when expecting these kinds of messages.

Rule-based Filters

You can use a filter to set up specific rules that can be applied to all emails coming into your system. If the email’s content or origin matches one of the rules, it can be automatically sent to a spam folder. For example, you can set the filter to look for specific words or phrases in the body of an email. If these words are present, the message gets sent to the spam folder.

You can also set the filter so it looks for particular words or phrases in the header. This can be useful for emails associated with memberships that, while still useful, result in unwanted messages from time to time.

Rule-based spam filtering is also useful for targeting specific senders. You can set them up to look for information in the domain the email is coming from or the name of the person sending it.

Bayesian Filter

A Bayesian filter can learn your preferences by examining the emails that you send to spam. It observes the content of the emails you mark as spam and then sets up rules accordingly. These rules are then applied to future emails trying to get into your inbox.

* For example, if you constantly mark all emails from a specific sender as spam, a Bayesian filter can recognize this pattern. It will then look for emails from that sender and move them to your spam folder automatically.

3. Market sentiment analysis

And:

Opinion mining is used as scrutiny of public opinions. The growth of social network has put onward the views of the general public on a larger scale and in an open manner. The comments, views and opinions act as deciding factors whether these are positive opinion or negative opinion. Guessing about the opinions' polarity is not a good idea, so, an intelligent system need to be introduced to categorize the views. Sentiment analysis thus emerged as a highlighted area in data mining. The opinions are judged on the basis of unsupervised and supervised learning. Supervised learning has unwavering to be superior to unsupervised mode of view verdict

Market sentiment is found through sentiment analysis, also known as opinion mining, which is the use of natural language processing methods to extract the attitude of a writer from source materials.

Supervised classification methods, such as Support Vector Machines , Naïve Bayes or ensembles have been deployed to perform sentiment analysis in multiple research projects. Machine learning techniques mainly use the bag-of-words model. In the bag-of-words model, a text is represented as the collection of its words, disregarding the order of those words in their sentences. However, the order of the words in a sentence can change the sentiment of a word. For example, consider the word “underestimate”. This word potentially has a negative connotation, but if we consider it beside other words like “underestimated stock” it can become positive.

Supervised classification methods, such as Support Vector Machines, Naïve Bayes or ensembles have been deployed to perform sentiment analysis in multiple research projects. Machine learning techniques mainly use the bag-of-words model. In the bag-of-words model, a text is represented as the collection of its words, disregarding the order of those words in their sentences. However, the order of the words in a sentence can change the sentiment of a word. For example, consider the word “underestimate”. This word potentially has a negative connotation, but if we consider it beside other words like “underestimated stock” it can become positive.

Recently, Deep Learning approaches have emerged as a powerful tool in sentiment analysis in Big Data due to the advantages they provide over other methods. One of these advantages is that features are learned hierarchically during the process of Deep Learning instead of the feature engineering that is required in data mining. Additionally, in Deep Learning methods, each word is considered as part of a sentence. In this way, relevant information contained in word order, proximity, and relationships is not lost. Furthermore, Deep Learning benefits from a similarity model. Word embedding creates a vector representation of words with a much lower dimensional space compared to the bag of the words model . The vectors representing similar words in vector space are therefore closer together. One of the other main concepts in Deep Learning algorithms is the automatic extraction of representation (abstractions) . To achieve this goal Deep Learning uses a massive amount of unsupervised data (Big Data) and extracts complex representations automatically. One of the advantages of abstract representation extracted with Deep Learning algorithms is their generalization. Features extracted from a given dataset can be used successfully for a discriminative task on another dataset. Deep Learning is an important aspect of artificial intelligence because it provides a complex representation of Big Data and also makes the machine independent of human knowledge.

Deep Learning constructs complicated representations for image and video data with a high level of abstraction. High-level data representations provided by Deep Learning can be used for simpler linear models for Big Data. This representation can be useful for image indexing and retrieval. In other words