1. Introduction

Welcome to the series of e-learning modules on Bayes' Theorem and its Applications. In this module we will be introduced to Bayes' theorem and the topic will cover the proof of the theorem for two and more number of prior events, difference between prior and posterior probability and the practical applications of Bayes' Theorem.

By the end of this session, you will be able to explain:

- Baye's Theorem
- Conditional & posterior probability
- Extension of the theorem and its proof
- Probability for another future event
- Theorem for two random variables
- Fields of application of the theorem
- Applications of the theorem

One of the most interesting applications of the results of probability theory involves estimating unknown probabilities and making decisions on the basis of new information.
Since World War II , a considerable amount of knowledge has been developed known as the Bayesian Decision Theory whose purpose is to solve problems involving decision making under uncertainty.

The concept of conditional probability takes into account information about the occurrence of one event to predict the probability of another.
This concept can be extended to revise probabilities based on new information and to determine the probability that a particular effect was due to specific cause.
The procedure of revising these probabilities is known as Bayes' theorem.

Bayes' theorem was named after the British Mathematician Reverend Thomas Bayes and published in 1763.
This is one of the most famous memoirs in the history of science.

The so-called 'Bayesian' approach to this problem addresses itself to the question of determining the probability of some event, A, given that another event B, has been observed. That is, determining the value of Probability of 'A', given the value B.
The event A is usually thought of as sample information so that Bayes' rule is concerned with determining the probability of an event given certain sample information.

2. Probability Revision Process

Revising probabilities when new information is obtained is an important phase of probability analysis.
The steps in this probability revision process are shown in the below:
Step 1: Prior Probabilities
Step 2: New Information
Step 3: Application of Bayes' Theorem and
Step 4: Posterior Probabilities

For example, A sample output of 2 defectives in 50 trials, say, event A, might be used to estimate the probability that a machine is not working correctly, say, event B.
OR
You might use the results of your first examination in statistics, again say, event A, as sample evidence in estimating the probability of getting a first class in say, event B.
Bayes' theorem is based on the formula for conditional probability.

Let A_1 and A_2 equal to the sets of events which are manually exclusive and exhaustive and, B equal to a simple event which intersects each of the A events as shows in the diagram.
Observe the diagram.
The part of B which is within A_1 represents the area ' A_1 and B', and the B within A_2 represents the area ' A_2 and B'.

Hence the probability of event A_1 given event B is:
Probability of A_1 intersection B divided by Probability of B.
Similarly the probability of event A_2 given event B is: Probability of A two intersection B divided by Probability of B.

Where, Probability of B is equal to Probability of A_1 intersection B plus probability of A_2 intersection B,
Probability of A_1 intersection B equals to Probability of A_1 into Probability of B given A_1, and Probability of A_2 intersection B equals to Probability of A_2 into Probability of B given A_2.
Hence, Probability of A_1 given B is equal to Probability of A_1 into Probability of B given A_1 divided by Probability of A_1 into Probability of B given A_1 plus Probability of A_2 into Probability of B given A_2.

Let E_1, E_2, E_3 etc… En be a set of 'n' mutually exclusive events with Probability of E_i not equal to zero and i equal to 1 to n, then for any arbitrary event 'A' which is a subset of Union of E_i and i equal to 1 to n such that probability of A is greater than zero, we have:
Probability of E_i given A is equal to Probability of E_i into Probability of A given E_i divided by the sum of Probability of E_i into Probability of A given E_i, i runs from 1 to n.

Since A is the subset of Union of Ei, A is equal to A intersection union Ei equal to 1 to n
Equals to Union i equal to 1 to n A intersection Ei , by distributive law.
Now, since A intersection E_i where i from 1 to n are mutually exclusive events
Applying addition theorem of Probability or Axiom three of Probability we have,
Probability of A is equal to Probability of union (i from 1 to n) A intersection Ei

Is equal to summation (i from 1 to n) Probability of A intersection Ei
Equals to summation (i from 1 to n) Probability of Ei multiplied by Probability of A given Ei.
Call this as star.
Also, we have
Probability of A intersection is equal to Probability of A multiplied by Probability of Ei given A, and
Probability of Ei given A equals to Probability of A intersection Ei divided by Probability of A
This is equal to:
Probability of Ei into Probability of A given Ei divided by the summation of Probability of Ei into Probability of A given Ei where i runs from 1 to n from the equation star.

3. Bayes’ Theorem for Future Events

Bayes’ theorem for future events:
The probability of the materialization of another event C given by
Probability of C given A intersection E1, Probability of C given A intersection E2, etc
Probability of C given A intersection En is
Probability of C given A is equal to summation, i running from 1 to n
Probability of Ei into Probability of A given Ei into Probability of C given A intersection Ei divided by summation i running from 1 to n Probability of Ei into Probability of A given Ei.

Bayes’ Theorem for two random variables
Consider a sample space omega generated by two random variables X and Y. In principle, Bayes' theorem applies to the events.
A is equal to X equals to x and B is equal to Y equals to y
If X is continuous and Y is discrete,
f of x given Y equals to y is equal to
Probability of Y equals to y given X equals to x into f of x divided by Probability of Y equals to y.

If X is discrete and Y is continuous,
Probability of X equals to x given Y equals to y is equal to f of y given X equals to x into Probability of X equals to x divided by f of y.
If both X and Y are continuous,
f of x given Y equals to y is equal to
f of y given X equals to x into f of x divided by f of y.

Probabilities before revision by Bayes’ rule are called a priori or simply prior probability because they are determined before the simple information is taken into account.
A probability which has undergone revision in the light of sample information using Bayes’ rule is called posterior probability, since it represents the probability calculated after this information is taken into account.

Posterior probabilities are also called revised probabilities because they are obtained by revising the prior probabilities in the light of the additional information gained.
Posterior probabilities are always conditional probabilities, the conditional event being the sample information. Thus a prior probability which is unconditional becomes a posterior probability which is a conditional probability by using Baye’s rule.

The revision of the old probabilities in the light of the additional information supplied by the experiment, or the past records, is of extreme help in arriving at valid decisions in the face of uncertainty.
It may be pointed out that the classical theory is mainly empherical since it employs only sample information as the basis for estimation and testing while, the Bayesian approach employs any and all available information whether it is personal judgment or empirical

evidence

Frequentist and Bayesian approaches both use probability to make Statistical inferences, but they use and interpret probability in very different ways.
A frequency probability is a property which applies to chance events. A Bayesian probability, in contrast, is a mental construct that represents uncertainty. It applies not directly to events, but to our knowledge of them, and can thus be used in determinate situations.

Further inference based on Bayesian rule can be made on prior information alone or on both prior and sample information.
The term ‘prior information’ implies a type of information which a statistician or a decision maker has on an inferential problem before any sampling is conducted.
Prior information often consists of personal judgments of the situation, because of which Bayesian method can actually be considered as an extension of the classical approach.

Some interesting points to be noted are:
Though Bayes’ rule deals with a conditional probability, its interpretation is different from that of the general conditional probability theorem.
The general conditional probability asks:
“What is the probability of the sample or experimental result given the state value”?
Whereas Bayes’ theorem asks:
“What is the probability of the event given the sample or experimental result”

4. Computation Methods

When we use Bayes' theorem, different decision makers may assign different probability to the same set of states of nature. Also we may conduct a new experiment by using posterior probabilities of the preceding experiment as the prior probabilities. As we proceed with the repeated experiments, evidence accumulates and modifies the initial priori probabilities, thereby modifying the intensity of a decision-maker's belief in various states of nature.

The notion of priori and posterior in Bayes' theorem are relative to a given sample outcome. That is, if a posterior distribution has been determined from a particular sample this posterior distribution would be considered the prior distribution relative to a new sample.

The Tabular Approach is helpful in conducting the Bayes' theorem calculations simultaneously for all events Ei.
The computation involves 4 steps.

Step 1: Prepare the three columns:
Column 1 - The mutually exclusive events for which posterior probabilities are desired.
Column 2 - The prior probabilities for the events.
Column 3 - The conditional probabilities of the new information given each event.

Step 2: Compute Joint Probabilities:
In column 4, compute joint probabilities for each event and new information A by using the multiplication law.
To get these joint probabilities, multiply the prior probabilities in column 2 by corresponding conditional probabilities in column 3.
That is, Probability of E_1 intersection A is equal to
Probability of E_1 multiplied by probability of A given E_1.

Step 3: Sum of Joint Probabilities:
In this step, sum up the joint probabilities in column 4 to obtain the probability of the new information, P of A.

Step 4: Create Column 5:
In column 5, compute the posterior probabilities by using the basic relationship of conditional probability.

Bayesian methods are increasingly being applied in a diverse assortment of fields, including science, business, law, medicine, engineering, sociology, psychology, artificial intelligence, and philosophy.
Bayes' Theorem can be used in Internet marketing to track profile visitors to a Website.
Bayes' methods are applied to both virtual screening and the chemical biology arena, where applications range from bridging phenotypic and mechanistic space of drug action to the prediction of ligand-target interactions.

5. Areas of Application

Bayesian statistical analysis is useful in business and political applications involving actual and opportunity costs where decisions often must be made under uncertain conditions.
Bayesian theorem is particularly helpful in financial modelling, solving pricing problems in foreign exchange trading and to solve social policy-making problems.
Bayesian rule holds similar promise for researchers of business and management problems.

Here is an example.
Assume that a factory has two machines.
Past records shows that Machine-I produces 30% of the items of the output and Machine-II produces 70% of the items.
Further 5% of the items produced by Machine-I were defective and only 1% produced by Machine -II were defective.
If a defective item is drawn at random, what is the probability that the defective item was produced by Machine-I or Machine-II?

Solution:
Let, E1 be the event of drawing an item produced by Machine-I,
E2 be the event of drawing an item produced by Machine-II, and
A be the event of drawing a defective item produced by the machines
Then from the given information
Prior probabilities are:
Probability of E1 is equal to 30 percent, that is 0.30, and
Probability of E2 is equal to 70 percent that is 0.70.

From the additional information
Conditional probabilities
Probability of A given E1 is equal to 5 percent equal to 0.05
Probability of A given E2 is equal to 1 percent equal to 0.01
Probability of E1 given A is equal to Probability of E1 into Probability of A given E1 divided by the sum of Probability of Ei into Probability of A given Ei where i runs from 1 to 2.
Probability of E1 given A equals to 0.30 into 0.05 divided by the product of 0.30 and 0.05 plus the product of 0.70 and 0.01
Which equals to 0.682, which is 68.2 percent.

Probability of E2 given A is equal to Probability of E2 into Probability of A given E2 divided by the sum of Probability of Ei into Probability of A given Ei where i runs from one to two
Probability of E2 given A equals to 0.70 into 0.01 divided by the product of 0.30 and 0.05 plus the product of 0.70 and 0.01 .
Which equals to 0.318, which is 31.8 percent.
Here, note that without the additional information we may be inclined to say that the defective item is drawn from Machine-II output since Probability of E2 is 70 percent, which is larger than Probability of E1 which is 30 percent.

With the additional information we may give a better answer
The probability that the defective item was produced by Machine-I is 68.2 percent,

and that by Machine-II is only 31.8 percent.
Hence, we may now say that the defective item is more likely drawn from the output produced by Machine-I.

Here is another example.
Suppose a drug test is 99% sensitive and 99% specific. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability he or she is a user?

Now, let E1 be the event that people are drug-users
E2 be the event that people are non- drug users, and
A be the event that individual tests positive.
Then from the given information:
Prior probabilities are
Probability of E1 is equal to 0.5 percent, equals to .005
Probability of E2 is equal to .095 percent equals to 0.995.
From the additional information,
Conditional probabilities:
Probability of A given E1 is equal to .99
Probability of A given E2 is equal to .01

Probability that the selected individual is a drug user whose test turns out to be positive can be computed as follows:
Probability of E1 given A is equal to Probability of E1 into Probability of A given E1 divided by the sum of Probability of Ei into Probability of A given Ei where i runs from one to two
Probability of E1 given A equals to .005 into .99 divided by the product of .005 and .99 plus the product of .995 and 0.01.
Which is equal to 0.332
That is, approximately 33.2 percent.

Despite the apparent accuracy of the test, if an individual tests positive, it is more likely that they do not use the drug than that they do.
This surprising result arises because the number of non-users is very large compared to the number of users, such that the number of false positives which is 0.995 percent outweighs the number of true positives which is 0.495 percent.

Here's a summary of our learning in this session:

- Statement and proof of Bayes' Theorem
- Proof of theorem for the extended form
- Theorem for two random variables
- Difference between conditional and posterior probability
- Tabular approach to compute posterior probabilities using Bayes' rule
- Fields of applications of the theorem