Mithilesh Singh

Bayes Theorem

Named after Thomas Bayes.

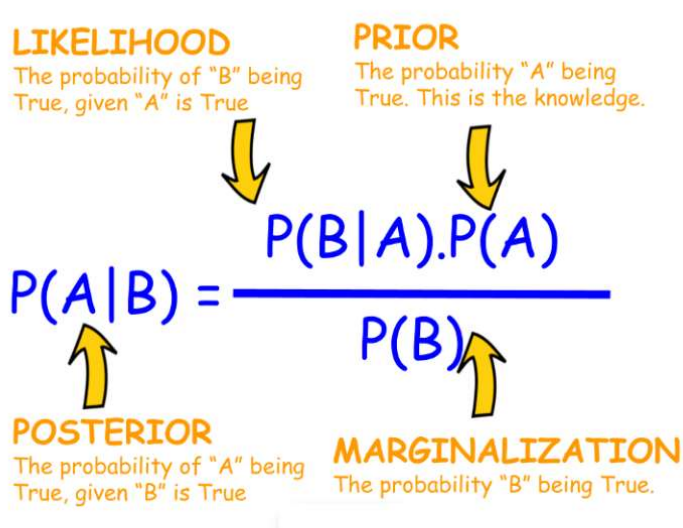
Bayes theorem is actually an extension of conditional probability. It is represented as:

P(A|B) = P(B|A) * P(A)/P(B)

Here, The conditional probability answers the probability of occurrence of A when B has already occurred. In that case, the Bayes theorem answers using prior beliefs and comes to a posterior conclusion.

Describes the probability of an event based on the prior knowledge of conditions that might be related to the event.

The conditional Probability is known as hypothesis. The Hypothesis is calculated through previous evidence or knowledge.



Deriving Bayes Equation from conditional Probability:-

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

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

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

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

Solving above two equations we get Bayes Theorem.

One key to understanding the essence of Bayes' theorem is to recognize that we are dealing with sequential events, whereby new additional information is obtained for a subsequent event, and that new information is used to revise the probability of the initial event.

In this context, the terms prior probability and posterior probability are commonly used.

Definitions

A prior probability is an initial probability value originally obtained before any additional information is obtained.

A posterior probability is a probability value that has been revised by using additional information that is later obtained.

Example 1

What is the probability of a patient having liver disease if they are alcoholic?
Given data(**Prior Information): -**
(1)As per earlier records of the clinic, it states that 10% of the patient's entering the clinic are suffering from liver disease.
(2)Earlier records of the clinic showed that 5% of the patients entering the clinic are alcoholic.
(3)Earlier records of the clinic showed, 7% out of the patient's that are diagnosed with liver disease, are alcoholics.
What is the probability that Patient Being Alcoholic, chances that he is having a liver disease?

Ans:-

As, per Bayes theorem formula,
P(A|B) = (P(B|A) * P(A))/ P(B)

P(A)=Probability that Patient having liver disease =0.10
P(B)= Probability that Patient is alcoholic =0.05
P(A|B)= Probability that Patient having liver disease, it is known that he is alcoholic =?
probability of a patient being alcoholic, given that they have a liver disease is 7%. This defines the B|A:
P(B|A)= 0.07.
As, per Bayes theorem formula,
P(A|B) = (P(B|A) * P(A))/ P(B)
P(A|B) = (0.07 * 0.1)/0.05 = 0.14
Therefore, for a patient being alcoholic, the chances of having a liver disease is 0.14 (14%).
This is a large increase from 10% suggested by past data.

Example 2

In a particular pain clinic , 10% of patients are prescribed narcotic pain killers. Overall 5% of clinic patients are addicted to narcotics (including pain killers and illegal substances)

Out of all the people prescribed pain killers 8% are addicted.

If a patient is addicted what is the probability that he is prescribed pain killer?

ANS:-

As, per Bayes theorem formula,
$P(A|B) = (P(B|A) * P(A))/ P(B)$

P(A)=Event A is being prescribed pain pills= 10%=0.1

P(B)=Event B is being an addict=5%=0.05

$P(B/A)=8\%=0.08$

$P(A/B)=?$

$P(A/B)=P(B/A)*P(A)/P(B)=0.08*0.1/0.05=0.16=16\%.$

Example 2: You are planning a picnic today, but the morning is cloudy. 50% of all rainy days start off cloudy. But cloudy mornings are common (about 40% of days start cloudy). Also this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%). What is the chance of rain during the day?

Ans:-

As, per Bayes theorem formula,
P(A|B) = (P(B|A) * P(A))/ P(B)

P(Rain): 10%
P(Cloudy) : 40%
P(Cloudy|Rain) :50%

P(Rain/Cloudy)=

P(Rain)* P(Cloudy/Rain)/P(Cloudy)

=10%*50%/40%=12.5%.

Example 3

Black box manufactures for aircraft

A →75% production → defect 4%.

B →15% production → defect 6%.

C →10% production → defect 8%.

A defective Black box is randomly chosen, what is the probability that it was manufactured by company A?

Ans:-

Prior Information

P(A)=75%=0.75

P(B)=15%=0.15

P(C)=10%=0.10

P(D/A)=4%=0.04

P(D/B)=6%=0.06

P(D/C)=6%=0.08

P(A/D)=?

P(A/D)=P(A)*P(D/A)/[[P(A)*P(D/A)]+[P(B)*P(D/B)]+[P(C)*P(D/C)]]

=0.75*0.04/(O.75*0.04)+(0.15*0.06)+(0.10+0.08)

=0.03/0.03+0.009+0.008

=0.03/0.047

=30/47

=0.6382

Example 4

A factory produces an item using three machines A,B & C which account for 20%, 30% and 50% of its output respectively.

Defective Item manufactured by machine A, B & C are 5 % , 3 % and 1% respectively. If a randomly selected item is defective, what is the probability is was produced by machine C?

Ans:-

Prior information

P(A)=20%=0.02

P(B)=30%=0.3

P(C)=50%=0.5

P(D/A)=0.05

P(D/B)=0.03

P(D/C)=0.01

Given that item is defective, find the probability of it being Machine C product?

P(C/D)=?

P(C/D) =

P(C)*P(D/C)/[[P(A)*P(D/A)]+[P(B)*P(D/B)]+[P(C)*P(D/C)]]

=0.5*0.01/(0.3*0.03)+(0.2*0.05)+(0.5*0.01)

=0.005/0.009+0.001+0.005

=0.005/0.015=5/15=1/3=0.33.

Example 6

What is the probability of a patient having liver disease if they are alcoholic?

Given data(**Prior Information): -**

(1)As per earlier records of the clinic, it states that 10% of the patient's entering the clinic are suffering from liver disease.

(2)Earlier records of the clinic showed that 5% of the patients entering the clinic are alcoholic.

(3)Earlier records of the clinic showed, 7% out of the patient's that are diagnosed with liver disease, are alcoholics.

This defines the B|A: probability of a patient being alcoholic, given that they have a liver disease is 7%.

What is the probability that Patient Being Alcoholic, chances that he is having a liver disease?

Ans:

P(A)=Probability that Patient having liver disease =0.10

P(B)= Probability that Patient is alcoholic =0.05

P(A|B)= Probability that Patient having liver disease, it is known that he is alcoholic =?

P(B|A)= Probability that Patient is alcoholic having liver disease=0.07.

As, per Bayes theorem formula,

$P(A|B) = (P(B|A) * P(A))/ P(B)$

$P(A|B) = (0.07 * 0.1)/0.05 = 0.14$

Therefore, for a patient being alcoholic, the chances of having a liver disease is 0.14 (14%).

This is a large increase from 10% suggested by past data.

Besides statistics, the Bayes' theorem is also used in various disciplines, with medicine and pharmacology as the most notable examples. In addition, the theorem is commonly employed in different fields of finance. modeling the risk of lending money to borrowers or forecasting the probability of the success of an investment.

These type of computations are actually used for SPAM Filters. we need to know a conditional probability.

Bayesian Spam Filtering

Although Bayes' Theorem is used extensively in the medical sciences, there are other applications. For example, it's used to filter spam. The **event** in this case is that the message is spam. The **test** for spam is that the message contains some flagged words (like "Money, Lottery" or "you have won"). Here's the equation set up (from Wikipedia), read as "The probability a message is spam given that it contains certain flagged words":

$$\Pr(\text{spam}|\text{words}) = \frac{\Pr(\text{words}|\text{spam})\Pr(\text{spam})}{\Pr(\text{words})}$$

The actual equations used for spam filtering are a little more complex; they contain more flags than just content. For example, the timing of the message, or how often the filter has seen the same content before, are two other spam tests.

Example 7.

The probability that the word MONEY word appears in an email, given that the email is spam is 8%.

Probability that an email can be a spam =20%

Probability that Money can appear in an email=2.4%

Find the probability that the email is spam, given that Money word is in the email?

Solution:-

M→Event of Money, S→ Even of SPAM

P(Money/Spam)=0.08, P(Spam)=0.2 , P(Money)=0.024

But We are interested to know the probability that the email is spam, given that Money appears in the email.

P(Spam/Money)=?

P(Spam/Money)=P(Money/Spam)x P(Spam)/P(Money)

=0.08x0.2/0.024=0.67=67%.

Please Note:- There are several forms of Bay's theorem out there, and they are all equivalent (they are written in slightly different ways)

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

$$P(B|A) = \frac{P(A|B)\,.P(B)}{P(A)}$$

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

Example 8.

1% of people have genetic defect.

90% of test correctly detect the Genetic defect (True Positive).

9.6% of the test are False Positive.

If a person gets a positive result, what is the probability that actually the person has the genetic defect?

Ans:-

Assume

A=Positive test report

B=Person having Genetic Disorder.

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

P(A/B)=Probability of finding the Positive report when the person having the Genetic disorder =90%.

P(B)=Probability the person has genetic defect=1%

P(A/~B)=Probability with positive test result person has not genetic defect=9.6%

P(~B)=1-P(B)=100-01=99% =Probability that Person has not genetic defect=99%

P(B/A)=probability that actually the person has the genetic defect when the test showing the Positive Report.

P(B/A)=?

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

P(B/A)=(0.9x0.01)/[(0.9x0.01)+(0.096x0.99)]

=0.09/[0.09+0.09504]

=0.09/0.18504

=0.4863

=48.63 % chance that in spite of having a positive test report person is patient of Genetic disorder.

Q 9. Epidemiologists claim that the probability of breast cancer among Caucasian women in their mid-50 is 0.005. An established test identified people who had breast cancer and were healthy. A new Mammography test in clinical trials has a probability of 0.85 for detecting for detecting cancer correctly. In women without breast cancer, it has a chance of 0.925 for a negative result. If a 55 year old Caucasian woman tests positive for breast cancer, what is the probability that she,in fact,has breast cancer?

Solution:-

Assume :-

A=Positive test report

B=Person having Disease.

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

P(A/B)=Probability of finding the Positive report when the person having the Disease =0.85

P(B)=Probability the person has Disease=0.005

P(~A/~B)=Probability with negative test result person has not Disease= 0.925

Important Note:-

P(test positive/no disease)+ P(test negative/no disease)=1

P(test positive/ disease)+ P(test negative/ disease)=1

P(~A/~B) + P(A/~B)=1

P(A/~B)=1- P(~A/~B)=1-0.925=0.075.

P(not B)=P(~B)=1-P(B)=1-0.005=0.995

P(B/A)=probability that actually the person has the Disease when the test showing the Positive Report.

P(B/A)=?

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

P(B/A)=0.85*0.005/[(0.85*0.005)+(0.075*0.995)]

=0.00425/[0.00425+0.074625] = 0.00425/0.078875

=0.05388

$$P(B|A) = \frac{P(A|B)\,.P(B)}{P(A)}$$

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

Q 10 . Spam Assassin works by having users train the system. It looks for patterns in the words in emails marked as spam by the user. For example, it may have learned that the word “free” appears in 20% of the mails marked as spam. Assuming 0.1% of non-spam mail includes the word “free” and 50% of all mails received by the user are spam, find the probability that a mail is spam if the word “free” appears in it.

Ans:-

A → FREE

B→SPAM

P(A/B)=0.20

P(A/NOT B)=0.001

P(B)=0.50= P(not B)=0.50

P(B/A)=?

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

P(B/A)=0.50*0.20/(0.50*0.20+0.50*0.001)

=0.995

Try Yourself:-

Question 11:- 1% of the population has certain disease. If an infected person is tested, then there is a 95% chance that the test is positive. If the person is not infected, then there is a 2% chance that the test gives an erroneous positive result ("False Positive").

Given that a person tests positive, what are the chances that she/he has the disease?

Answer:-

The probability that a person has a disease given that the predicted result is Positive.

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

Here A→event Positive.

B→Disease.

P(D/Positive)=?

P(Positive/D)=0.95

P(D)=0.01

P(Positive/Not D)=0.02

P(not D)=1-P(D)=1-0.01=0.99

(as, P(D)+P(Not D)=1)

P(D/positive) =0.95*0.01/ [(0.95*0.01) + (0.02*0.99)]

0.32.

Question 12.

A radar system is designed such that the probability of detecting the presence of an aircraft in its range is 98%. However, if no aircraft is present in its range it still reports (falsely) that an aircraft is present with a probability of 5%. At any time, the probability that an aircraft is present within the range of the radar is 7%.
a) What is the probability that no aircraft is present in the range of the radar given that an aircraft is detected?
b) What is the probability that an aircraft is present in the range of the radar given that an aircraft is detected?
c) What is the probability that an aircraft is present in the range of the radar given that no aircraft is detected?
d) What is the probability that no aircraft is present in the range of the radar given that no aircraft is detected?

Solution:-

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|\text{not } B)\,P(\text{not } B)}$$

D→detecting
A→Aircraft present

Replace in the above formula … the above values and solve.

P(No Aircraft/D)=?
P(D/Aircraft)=?
P(A/ No D)=?
P(No A/ No D)?

Given P(D/A)=0.98
P(D/not A)=0.05
P(A)=0.07
P(not A)=1-0.07=0.93

a) What is the probability that no aircraft is present in the range of the radar given that an aircraft is detected?

P(Not A/D)=?

(b) What is the probability that an aircraft is present in the range of the radar given that an aircraft is detected?

P(A/D)=?

(c) What is the probability that an aircraft is present in the range of the radar given that no aircraft is detected?

P(A/Not D)=?
P(Not D/A)=1- P(D/A)=1-0.98=0.02
P(Not D/Not A)=1-P(D/not A)=1-0.05=0.95

(d) What is the probability that no aircraft is present in the range of the radar given that no aircraft is detected?

P(Not A/Not D)=?