BAYES THEOREM

- Named after Thomas Bayes.
- Bayes theorem is actually an extension of conditional probability. It is represented as:

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

- 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.

Ex You've been planning a picnic for your family. You're trying to decide whether to postpone due to rain. The chance of rain on any day is 15%. The morning of the picnic, it's cloudy. The prob. of it being cloudy is 25% and on days where it rains, it's cloudy in the morning 80% of the time. Should you postpone the picnic?

$P(\text{rain}) = 0.15$

$P(\text{cloudy}) = 0.25$

$P(\text{cloudy}|\text{rain}) = 0.80$

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

$$P(\text{rain}|\text{cloudy}) = \frac{P(\text{cloudy}|\text{rain}) \cdot P(\text{rain})}{P(\text{cloudy})} = \frac{0.8 \cdot 0.15}{0.25}$$

$$\boxed{P(\text{rain}|\text{cloudy}) = 0.48}$$

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 Patients 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 of a Patient Being Alcoholic, when 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(B|A)= Probability that Patient is alcoholic having liver disease=0.07.

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

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

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

Therefore, the chances of having a liver disease given that he is an alcoholic is 0.14 (14%).

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 painkillers 8% are addicted.
If a patient is addicted what is the probability that he is prescribed pain killer?

P(A) = Probability of patient being prescribed pain killers = 0.10
P(B) = Probability of patient being addicted to pain killers = 0.05
P(B|A) = Probability of a prescribed patient being addicted = 0.08

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

P(A|B) = 0.08*0.10/0.05 = 0.16

If a patient is addicted to narcotics, the probability that they are prescribed narcotic painkillers is 0.16 or 16%.

Example 3: You are planning a picnic today, but its cloudy in the morning. 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?
Sol:- The chance of rain during the day given that it's cloudy in the morning is 0.125 or 12.5%.

Bayesian SPAM(simultaneously posted advertising message) Filtering

Bayes theorem for filter the email as SPAM or NOT.
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").

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 probability that the word MONEY word appears in an email, given that the email is spam is 8%. Probability that an email can be spam is 20%. Probability that Money can appear in an email is 2.4%. Find the probability that the email is spam, given that Money word is in the email?

P(Money|Spam)=0.08, P(Spam)=0.2 , P(Money)=0.024
P(Spam|Money) = 0.08*0.2/0.024 = 0.67 = 67%

Conclusion: There is a 67% chance that the email is spam given that there is the word Money in it.

Generalized form of Bayes Theorem

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|\bar{A}) * P(\bar{A})}$$

Example 1:

Epidemiologists claim that the probability of breast cancer among Caucasian women in their mid -50s is 0.005. An established test identified people who had breast cancer and those that were healthy. A new mammography test in clinical trials has a probability of 0.85 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?

A=Positive test report

B=Person having cancer.

P(Cancer) = 0.005

P(+ve|Cancer) = 0.85

P(~Cancer) = 0.995

P(+ve|~Cancer) = 0.0075

P(Cancer|+ve) = ?

$$= \frac{P(+ve|Cancer) * P(Cancer)}{P(Cancer)*P(+ve|Cancer)* + P(\sim Cancer) *P(+ve|\sim Cancer)}$$

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

Example 2:

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.

Data Given:

- P(Free | Spam) = 0.20
- P(Free | Non Spam) = 0.001
- P(Spam) = 0.50 => P(Non Spam) = 0.50
- P(Spam | Free) = ?

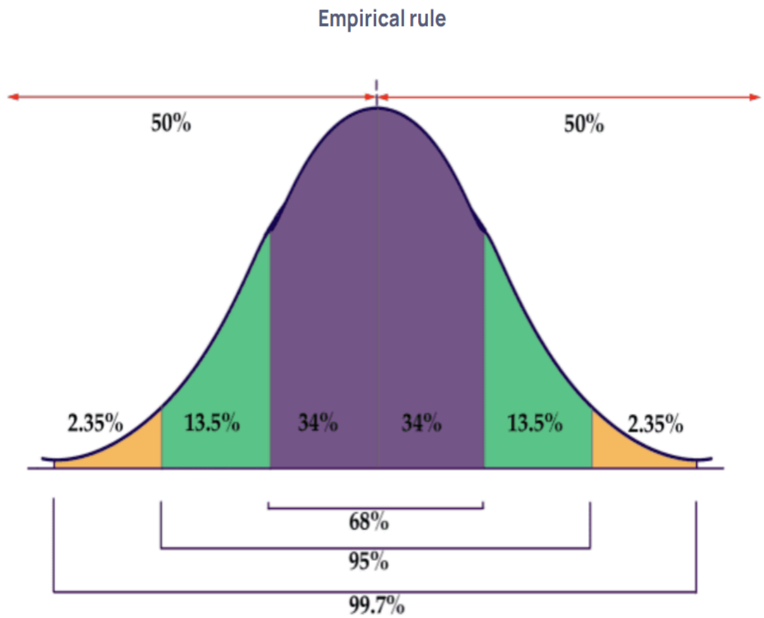
Using Bayes' Theorem:

- P(Spam | Free) = P(Spam) * P(Free | Spam) / P(Free)
- P(Spam | Free) = 0.50 * 0.20 / (0.50 * 0.20 + 0.50 * 0.001)
- P(Spam | Free) = 0.995

Example 3 :- 1% of the population has a 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?

Ans:
Given that a person tests positive, the chances that they have the disease is approximately 0.324 or 32.4%.

Empirical Rule.



About 68.26% data lies within 1 SD of the mean (μ),

About 95.44% data lies within 2 SD of the mean (μ),

About 99.72% data lies within 3 SD of the mean (μ)

& the rest 0.28% of the data lies outside 3 SD (>3σ) of the mean (μ), And this part of the data is considered as outliers.

Example:

(1) The Normal distribution has a standard deviation of 10 and mean 70. Approximately what area is contained between 70 and 90?

(2) The mean life of a tire is 30,000 km. The standard deviation is 2000 km.
Then, 68% of all tires will have a life between ____________ km and ___________ km.