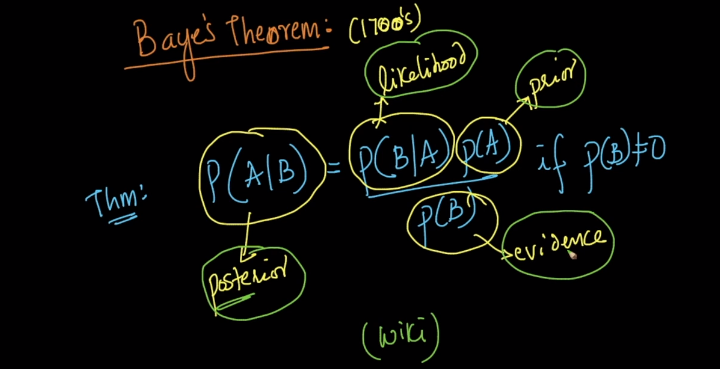
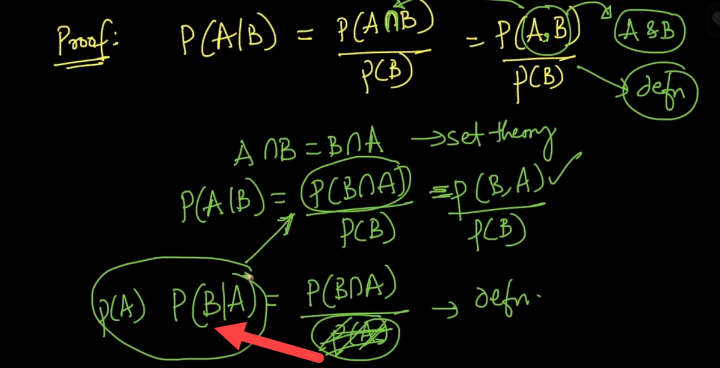
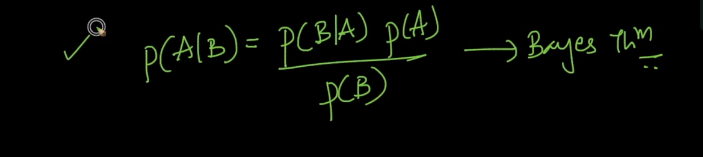
**Bayes Theorem with examples(**<https://en.wikipedia.org/wiki/Bayes%27_theorem>**)**

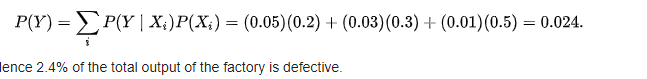
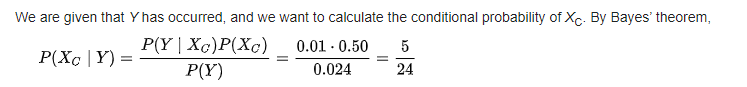
Terminology in statistics as below  


Proof of probability Bayes theorem:



From denominator, we’re moving it to other side and replacing P(B,A) = P(A)P(B/A) in statement P(A/B)  


**Example from Wiki:**  
The entire output of a factory is produced on three machines. The three machines account for 20%, 30%, and 50% of the factory output. The fraction of defective items produced is 5% for the first machine; 3% for the second machine; and 1% for the third machine. If an item is chosen at random from the total output and is found to be defective, what is the probability that it was produced by the third machine?  
**Solution:**  
Let’s say Xa,Xb,Xc are normal probabilities of each machine produced in factory. Simple, just divide it by 100  
we have   
🡪probabilities  
Y Denotes, the **defective** machine we picked randomly. Below statement can be read as Probability of Y for a given Xa = 0.05   
🡪Conditional probabilities

Here, first we need to compute P(Y) which is probability of Y,   
  
  
Total percentage of defective products made are 2.4%  
Next, we going to calculate the probability of getting defective product manufactured by machine 3 when picked randomly.  
we have P(Y|Xc) = 0.0.1, P(Xc) = 0.50 🡪 we get it from question  
P(Y) = 0.024🡪 calculated above  


**Question and reply**

1) Why can’t we add P(B|A1)+P(B|A2)+P(B|A3) to get P(B)?  
 1. My interpretation for P(B|Ai) is the probability of total defective items produced given the condition that it is produced from Ai.  
 2. My interpretation for P(Ai|B) is the probability of machine Ai produced items given the condition that are defective items.  
  
**Reply :**  
Why can’t we add P(B|A1)+P(B|A2)+P(B|A3) to get P(B), We cannot do it this way because you have to weight probabilities. Remember P(B|A1) is the percentage of defective items out of all items produced by machine 1. But here is the catch, we have no idea what amount of items machine 1 has produced. We are just giving, out of all items produced (by all the machines), 20% of the item belongs to machine 1, out of that 20 % only 5% is defective. So we weight it. 0.05 \* 0.2 = This much percentage of the "total item" produced (by all the machines) is defective and belongs to machine 1. So, 0.2 is 20% of the total item which belongs to machine 1 and 5% of those 20% items are defective. And P(B) is, what percentage of total items ( items produced by all the machines) is defective. I hope you get an idea.  
  
2) Bayes theorem is also called reverse probability, where we are usually given an event and we have to calculate the probability where it originated from, right?  
**Reply**: Correct  
  
Multinomial Naive Bayes is used when the input data features consists of numerical discrete values (like counts of words, TF-IDF scores, etc)  
Bernoulli Naive Bayes is used when the input data features consist of only boolean/binary values.