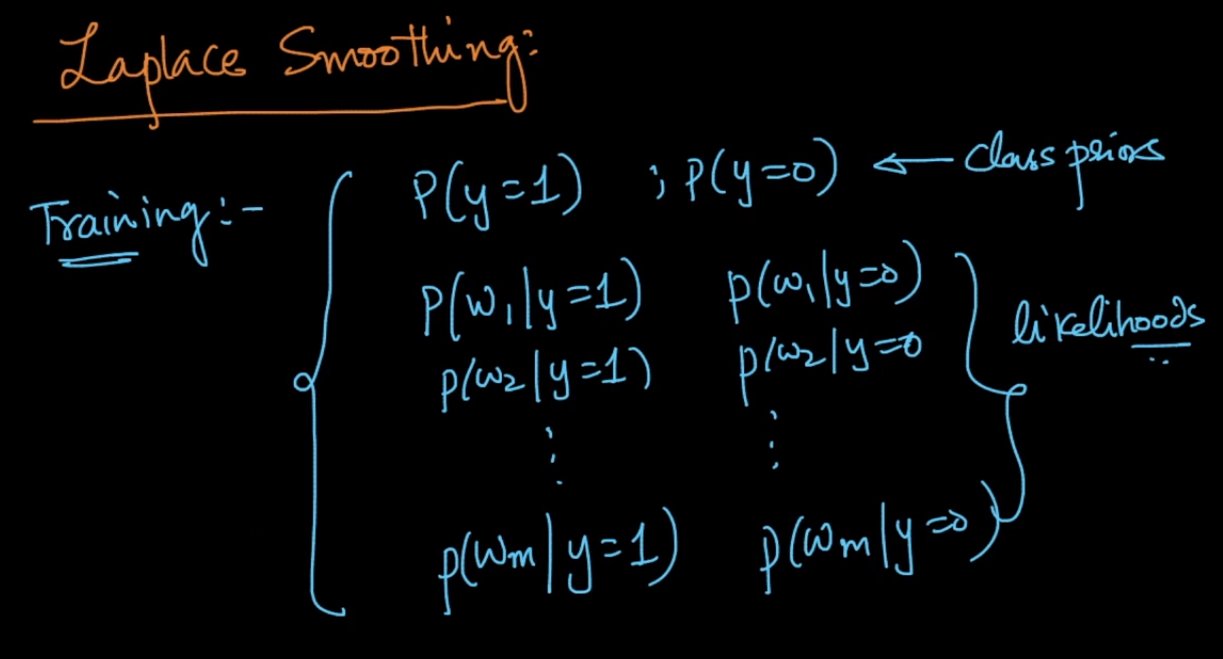
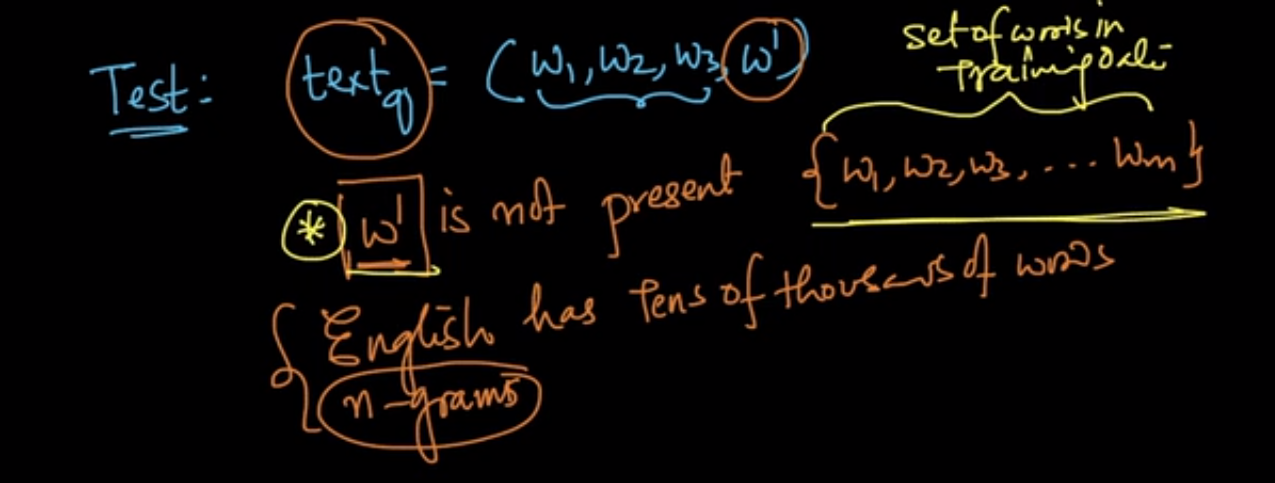
**Laplace Additive Smoothing :**

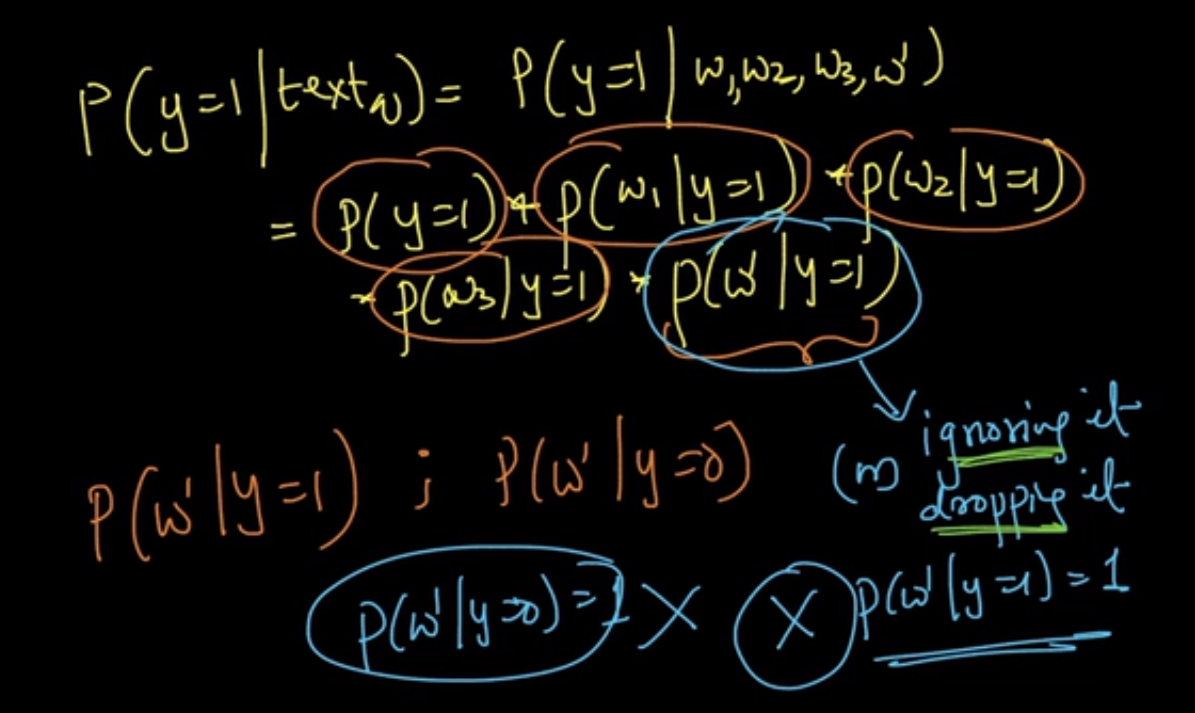
Now at the end of pre-processing of data in last video what we have calculated is shown in below image.



Now lets suppose in our test data we get a word Wx which is not the part of our training corpus which is very common as English has tens of thousands of words and we need to calculate P(Wx |Y = 1) , so how can we calculate this.

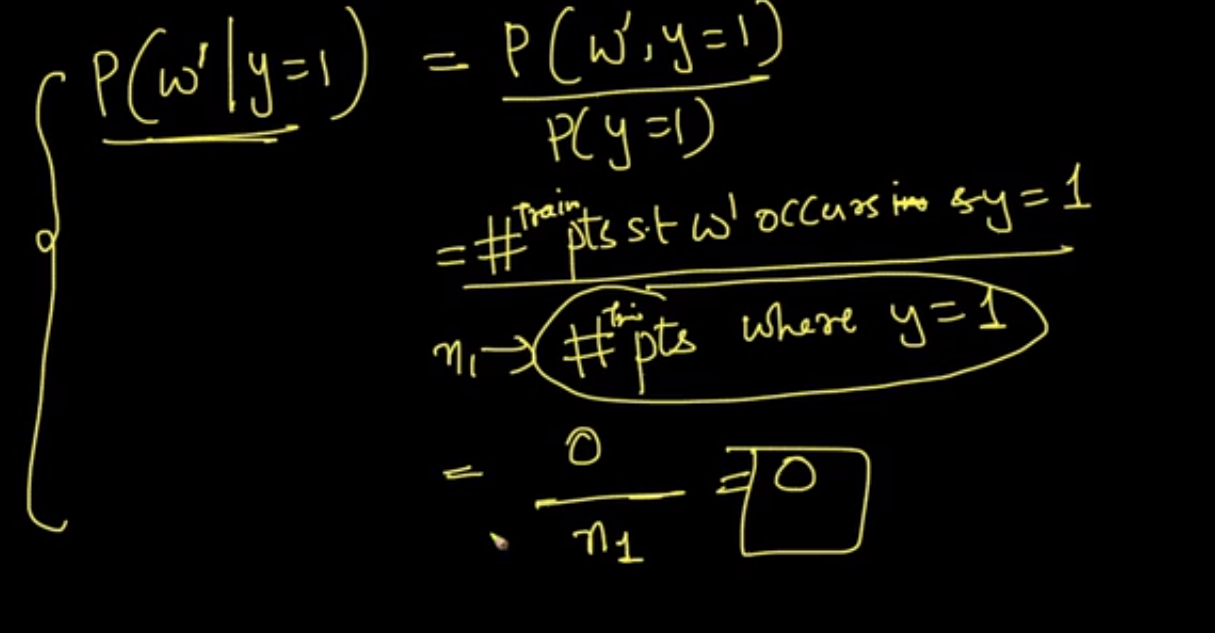


SO for calculating P(Y = 0 |Wx) we need to find P(Wx | y=0)



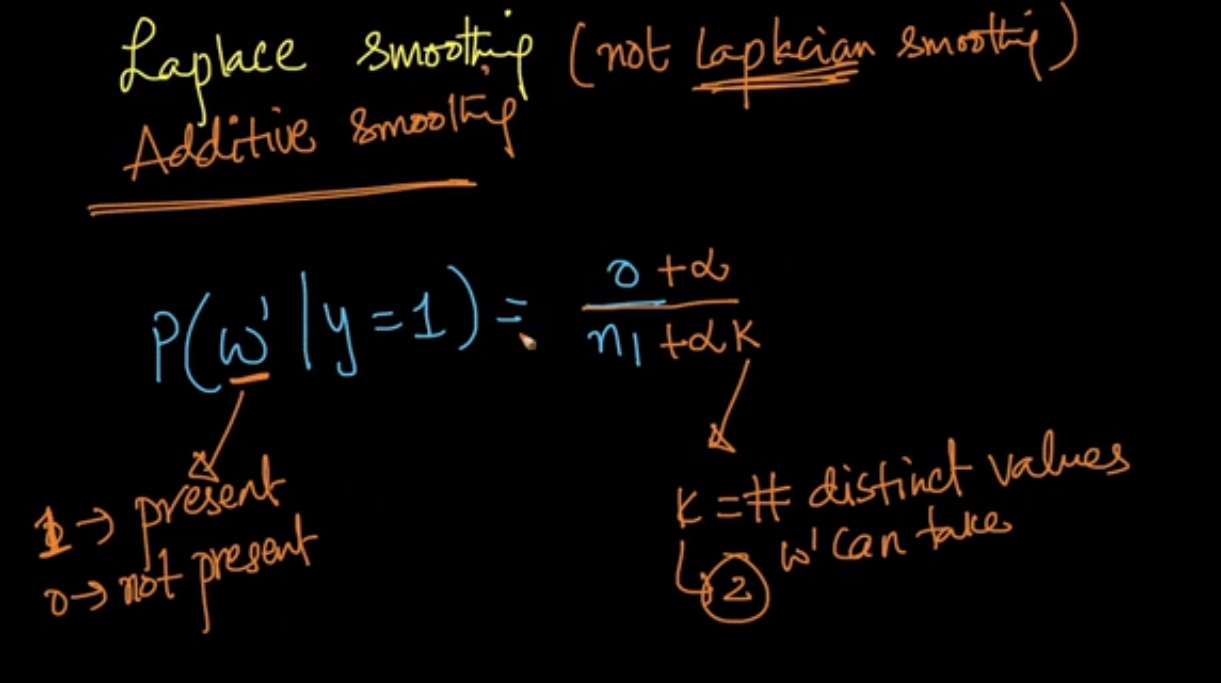
We cant just ignore it or drop it because it will mean that we are taking probability as 1 which is also wrong , there should be some way to handle this kind of situation.

Now lets check some of the possible ways to handle this situation



In this way the probability is becoming 0 and when we multiply it with probability of other likelihoods it will become 0 which is also incorrect.

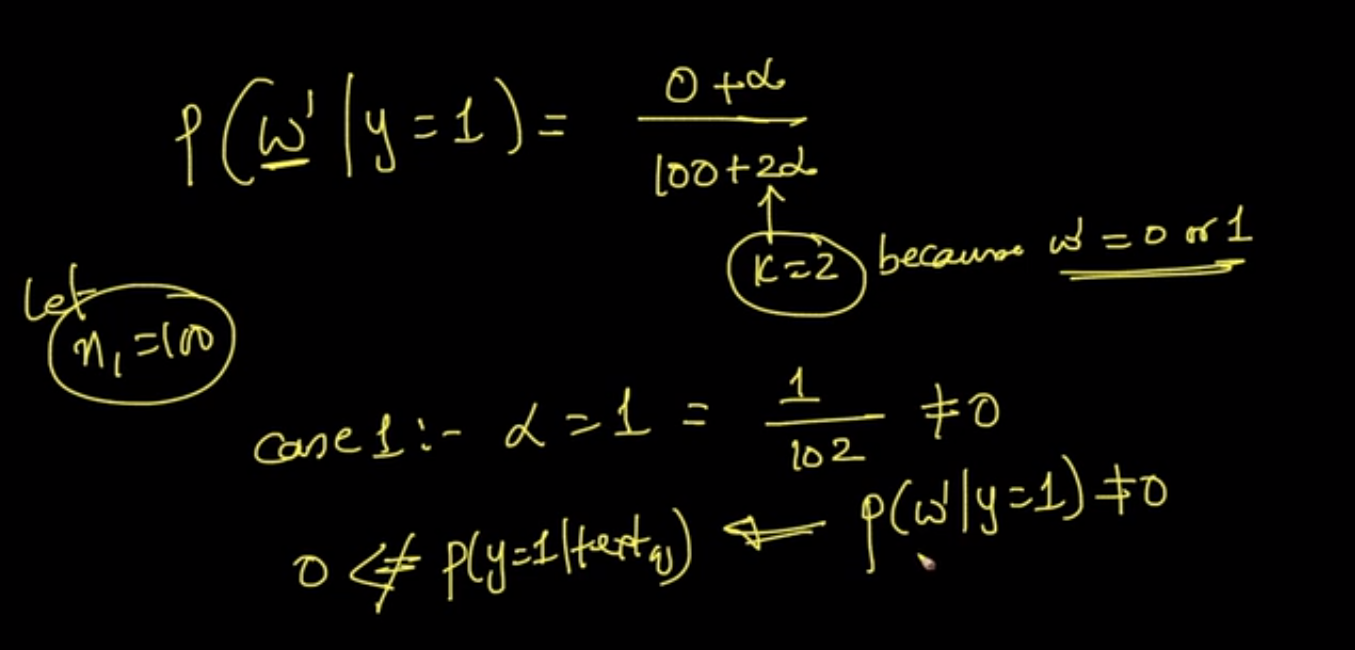
The workaround to this is said to be “**Laplace Smoothing”**  and not laplacian smoothing or also said as **“Additive Smoothing”** ,



It says instead of taking count of word in corpus as 0 lets add some (Alpha) to it and add (alpha\* K) to denominator i.e., n1(number of words in corpus).

Here k is # of distinct values wx(w dash) can take and (Alpha ) is typically taken as 1.

Lets get back to our example.

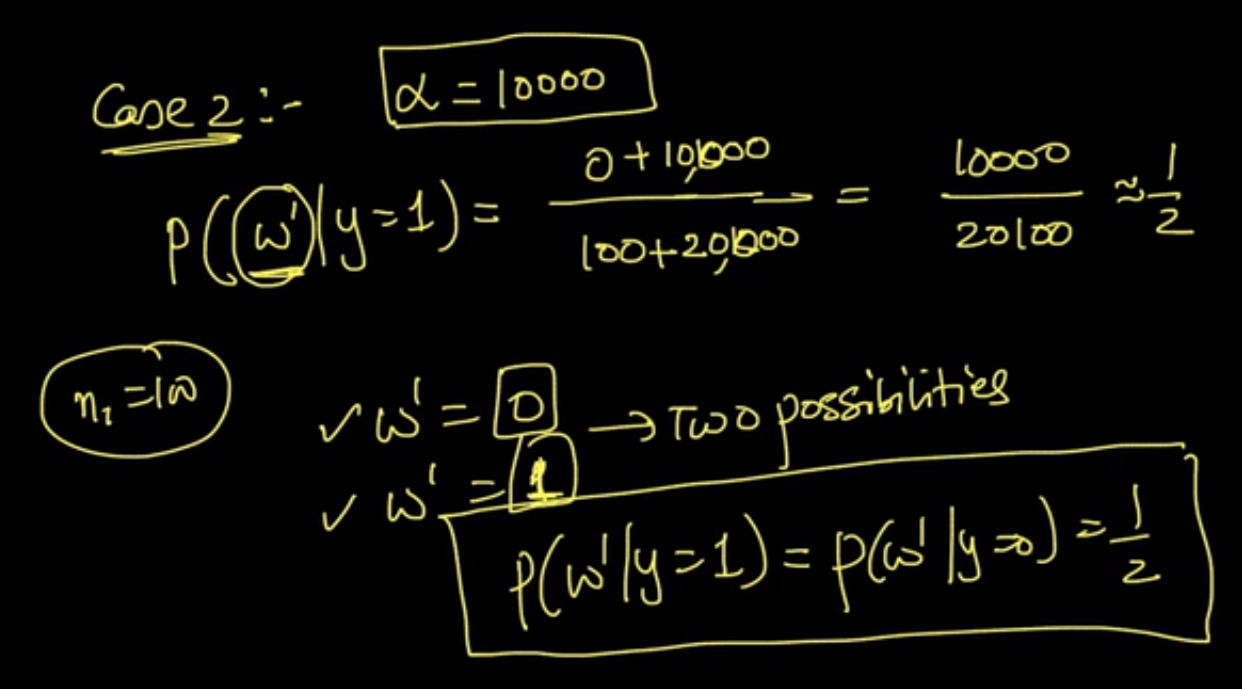


SO here k = 2 because as we can see that word (wx) can either be positive or negative

And now we will get P(wx|y = 1) = 1/102 (Where assumed n1 = 100)

Now the question is why are we making such complex calculation when we could just have replaced “0” with small epsilon value say 0.0001

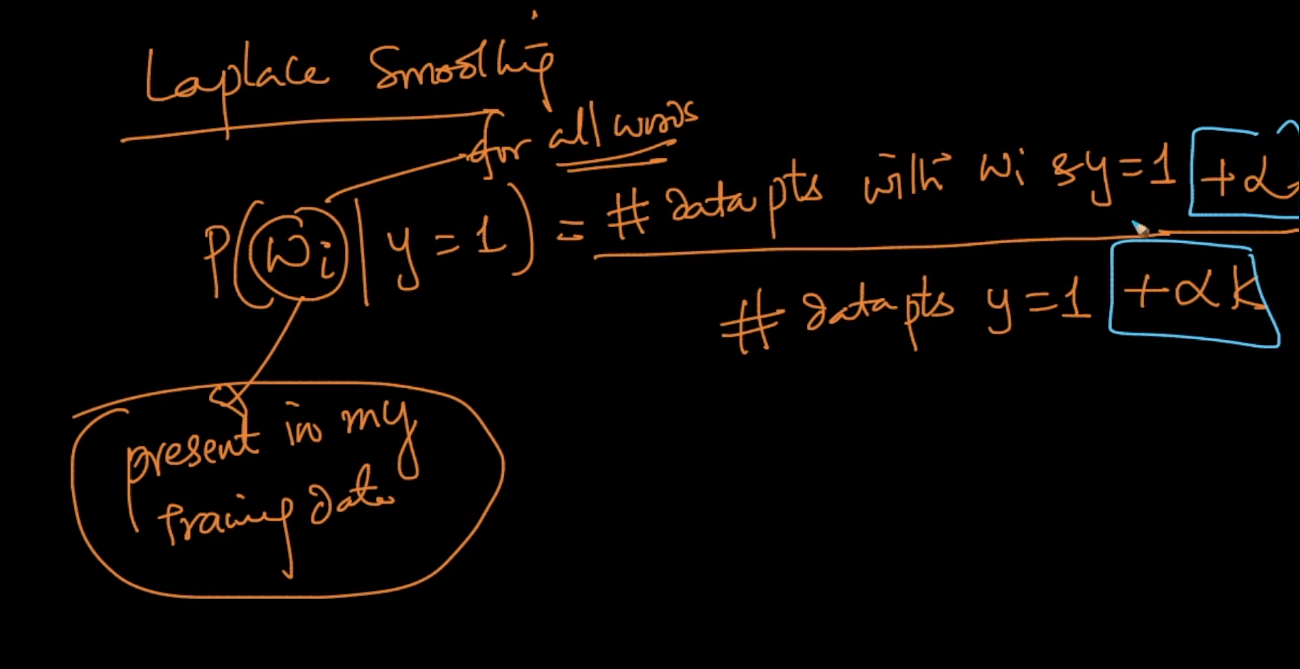
But there is some strong mathematical calculation behind all this.



Now lets say our n1 = 10000 so value for P(wx|y = 1 or y=0) = 1/2

Because we don’t know anything about out wx.

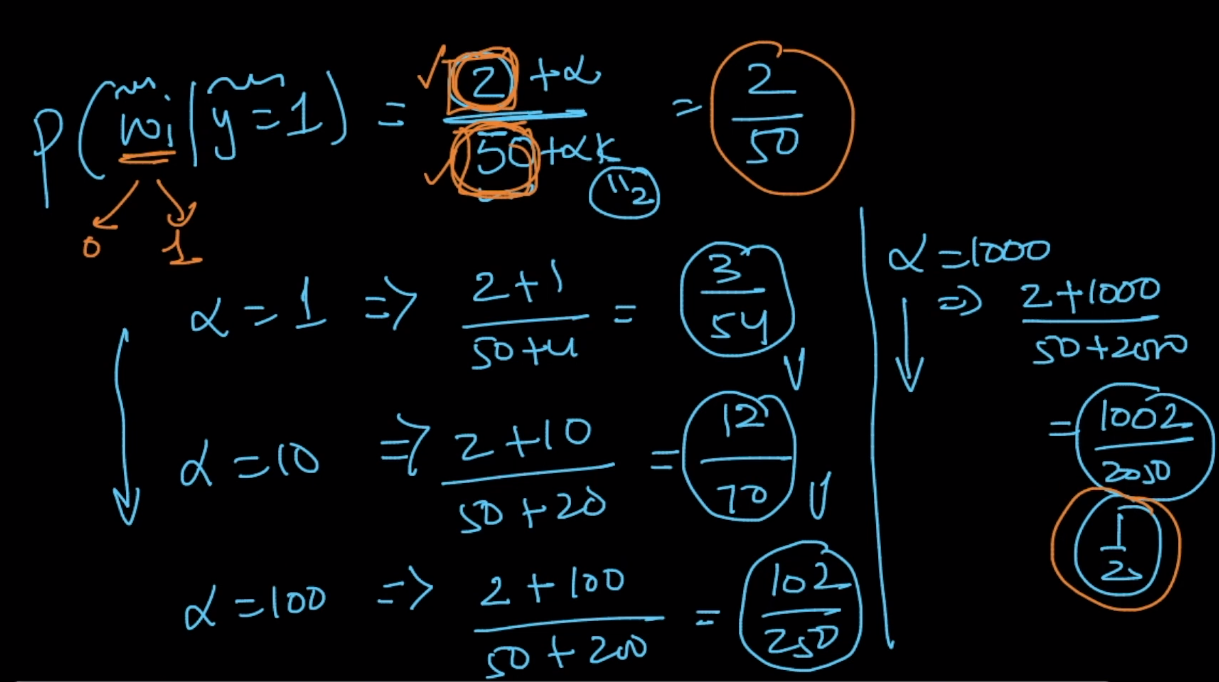
Now what actually we do in laplace smoothing is we do same procedure of adding (Alpha) to numerator and (K\*Alpha) to denominator for all the words weather they are present in corpus or not.



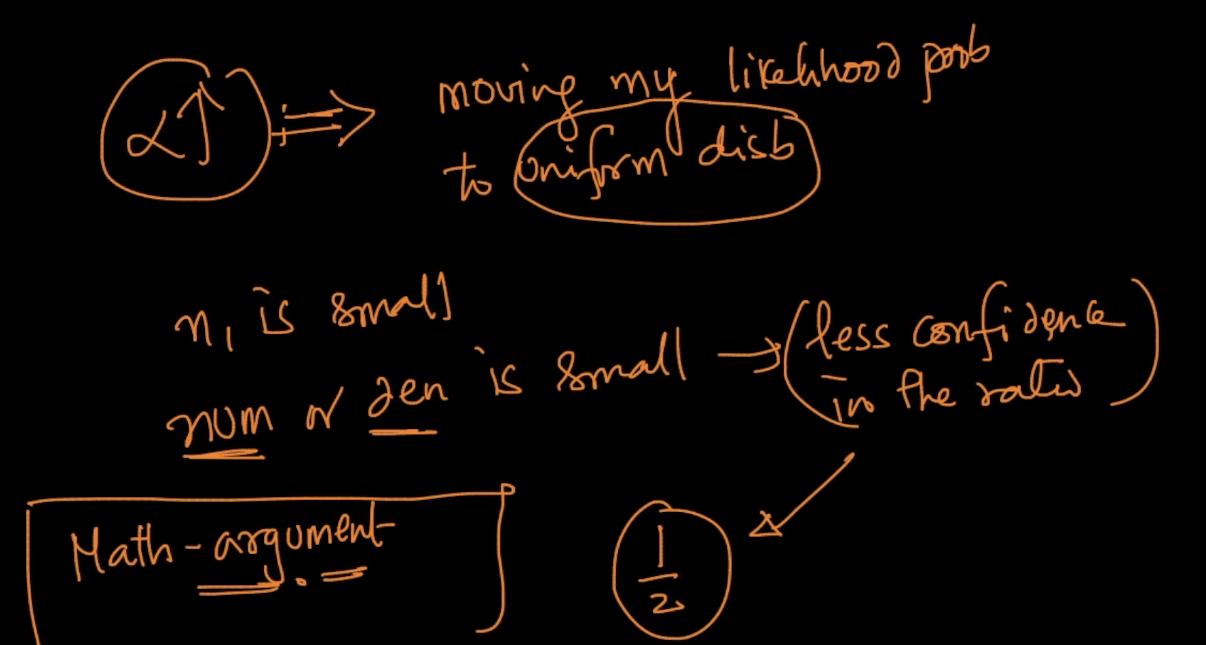
And since we are adding this values it is also called “**Additive Smoothing"**

Why it is called Smoothing?

This is because when the value of (Alpha) increases we are moving our likelihood probabilities towards uniform distribution .



Here (Alpha 1 <Alpha 10 < Alpha 100 < Alpha 1000)

As we can see that at alpha = 1000 we are getting value of 1/2 which is very well uniformly distributed as we know that the word Wx can only take two outcomes i.e., 1 or 0 so it is very well uniformly distributed. 

So if the value of numerator or denominator is less then we have less confidence on the ratio.

Comments:

