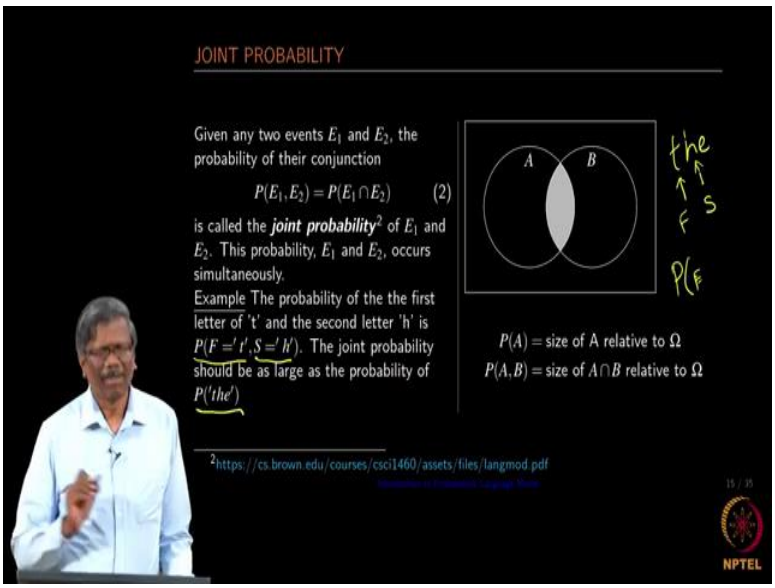
Applied Natural Language Processing
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture - 22
Joint and conditional probabilities, Independence with examples

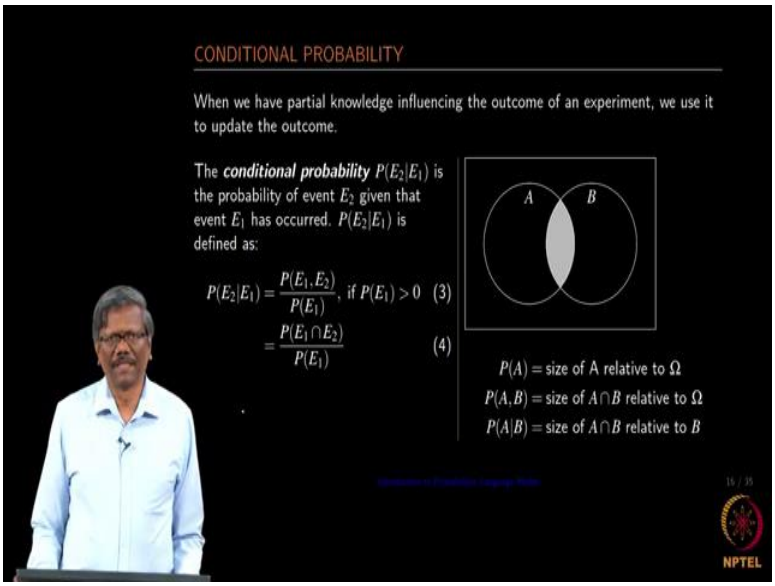
(Refer Slide Time: 00:15)



So, let us get on to the next term which is the joint probability. So, now, we have defined what is an event. So, now, we want to find out whether the two events can be combined to get one more probability number. For example, I want to combine the probability of the event and the probability of the event pack. So, it is defined as in set notation as given it 2 OKs. So, this is being called as the joint probability of E_1 and E_2

Probability E_1 and E_2 occur simultaneously. So, if we will now define or try to explain this with respect to certain examples ok. So, here the probability of the first letter t and the probability of the second letter h is the probability of F equal to t comma S equal to h. So, this is the joint probability and this probability should be as large as the probability of the event probability the ok.

(Refer Slide Time: 02:00)

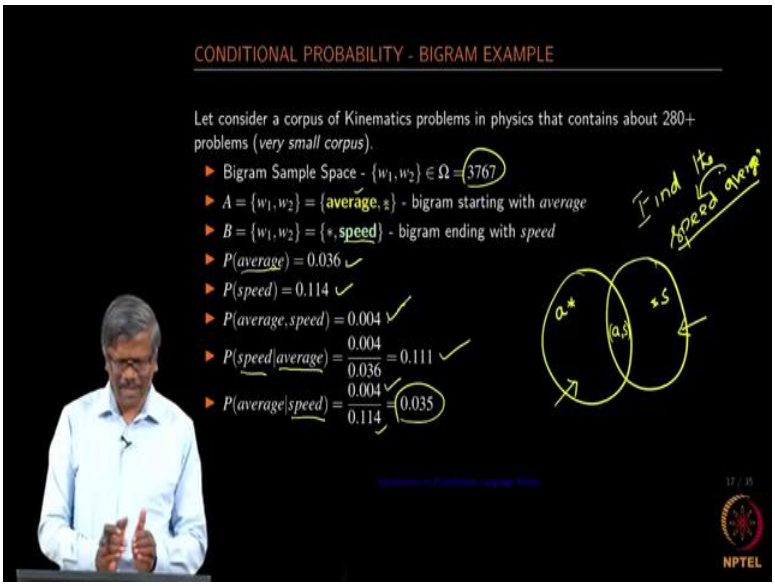


So, next we will define the conditional probability. So, in many cases rightly so, you know certain input with respect to what we are doing to be doing as an experiment right. So, we have some a priori knowledge or partial knowledge about the outcome of the experiments. So, and that we want to use in our probability experiments $P(E_1,E_2)$ is defined below

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{p(E_2)} \quad \text{if } P(E_2) > 0$$

$$= \frac{P(E_1 \cap E_2)}{p(E_1)}$$

(Refer Slide Time: 03:03)



Let us take a small example of conditional probability using a bigram ok; I am sure you know how bigrams are formed a given a document you take two words at a time you just move the window which contains two slots and then you move continuously over each word and then you get a set of bigram. I have taken a corpus where kinematics problems are available and there are about 280 problems there it is a very small corpus and the bigram samples are around 3767.

So, I have taken two words at a time and then I counted that and this is what we have. So, that is your sample space ok. So, now, we want to find out the words starting with average and the words ending with a speed ok. So, let us get a set where the average is used as the first word ok. So, we get the bigrams and then all the bigrams that start with average and the second could be anything ok.

In the same way we get another set where speed is used as the second word and any word could be part of the bigram; that means, it is mandatory to have speed as the second word, and then the first could be anything. So, then you have two such combinations there is a possibility that there is the average speed may occur in this set let us assume that this is the average star this is the star speed ok.

So, there is a chance that we could get a and s here right. So, in this case since it is talking about the kinematic problem. So, there are situations where these two words are available as part of that.

You get the probability for the word average you have the word average and then you have the entire tokens available and then find out what is the probability of that word you know from by distributing that probability mass of one across the entire token set in the same fashion you get this.

So, if you look at these two words the average and speed; speed occur more times than average so; that means, you will find this particular set a bigger than this set average. So, let us find out how many times average and speed occur together ok.

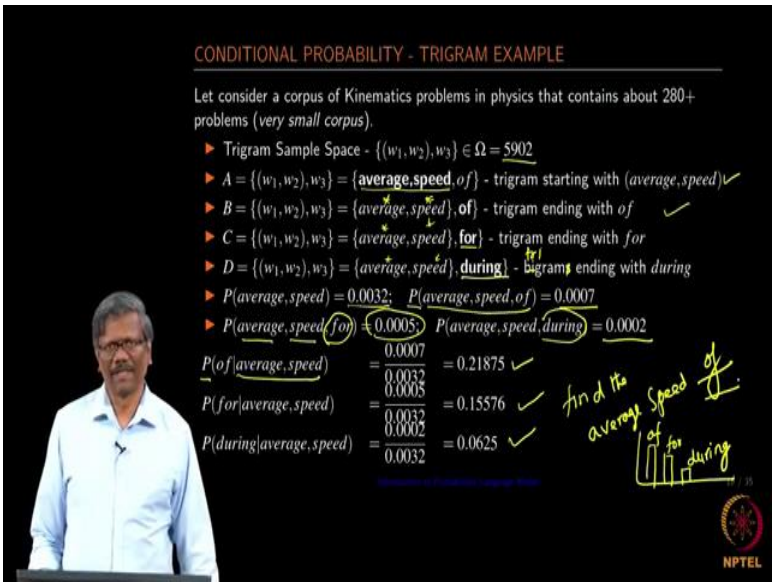
So, you get this number again from the bigram values; you go and then find out and count how many times the average and speed occur together, and then that is equal to 0.004 ok. So, now, finding going to be finding the conditional probability of the average is the first and speed being the second word.

So, now, I have the joint probability and now I have the count of the word or the probability of the word average and then you remember it is the ratio of the joint probability to the probability of that word that occurs as the first one.

So, we have average speed as the probability of 0.111 ok. So, let us find out whether it is at the same when speed occurs as the first word ok. So, in this case again we know this and we also know the probability of the speeds and then you look at the value it is smaller; that means, you will find more of average speed than speed average in the given documents. So, this is how we familiarize ourselves with the structure in the document and the sentence.

For example, if you are asked to construct a sentence and then you are given a sentence like find the speed average you are likely to come back and then say that this is not the right formation and this would be the right formation of the sentence based on what we had seen here ok. So, we have some information about these bigrams and then using the bigram logic and the information that is available to us with respect to the probability we find that the particular sentence finds the speed average need not occur in this but the right sentence could find the average speed. So, in this way we will be able to find out whether the given sentences the right sentence or wrong sentences ok.

(Refer Slide Time: 09:06)



So, let us take another example of the trigram. I am taking the same corpus it is very small but in this case now I have taken three words at a time. So, now, I am forming the trigrams. So, I have to use the window size with three slots and then move these slots along the document and then I am capturing all the trigrams. So, these spaces now containing 5902 trigrams. So, if you remember the previous case for the bigram it is only about 3767. So, now, in this case it is 5902.

So, as you increase the gram from 2 to 3 to 4; the sample space would increase you also would see that the complexities increasing slowly as you increase the size of the n-grams. So, in this case, what we are doing is we are taking two words as occurring together like what we had seen average speed and then we want to find out what is the third word there is going up occur after average speed ok; in this case now we capture all the trigrams that starting with the average speed I am restricting my size here.

And then I am going to find all the trigrams ending with. So, it is not going to be just average and speed it is going to contain any other word here right; just to show some idea I have just use those towards again. So, again we going to find the word for the same operation again I am doing the same operation. So, now, we have about four sets this is a trigram ending with during ok.

So, we have also found that the average speed has occurred and that the probability and then the probability of these three occurring together is pointed naught naught naught 7

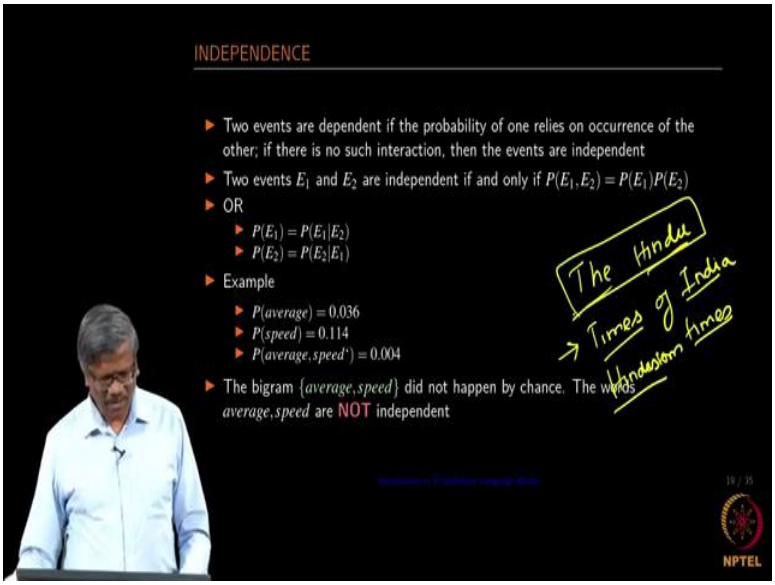
and then all the three words along with for like the last one we have the probability and then for the word during along with the average and speed this is the probability. So, now, we want to estimate what would be the best example of what is the trigram that most likely to happen you know with these three words ok

So, in this case again as we have done in the earlier computation we are going to be computing the conditional probability where average and speed are not right. So, if you use the formula for the conditional probability we are getting the probability of the word of occurring after average and speed is 0.21875 and for the word for it is 0.15576 and for the during this is the.

So, now what would be the most likely sentence if you use average and speed ok. So, find the average. So, we can say that it is going to be the most likely to occur ok. So, if you have drawn the graph like we had seen in the earlier case here I am going to be done using a bar chart. So, we will have for here and then we will have I am sorry this is off, this is for and this is a during ok.

So, given the knowledge that this is how the probability is available. So, most likely a word that we will pick is of ok. So, I wish you have understood the concept of conditional probability with respect to what we are doing in the natural language processing ok.

(Refer Slide Time: 14:29)



Let us move on to be next one see for example, in many cases right we have a seen certain words occur together for example, New Delhi, New York, Hong Kong and so on if you look at the cities; in they do not happen by chance because they are the names of the cities. In the same way, there are certain words that always happen together; we want to find out how many, how often they occur together and then do they really happen by chance or is it because of the number of time they occur; they always occur together and we have to use them always as a set of words.

If you look at the examples in the case of newspapers; The Hindu; right; Times of India, Hindustan Times. So, even though these two all these words are different; they going to be occurring together if you are talking about newspapers right. So, we need to find out whether these words in the given corpus occur as a chance. So, we have to find out whether they occur by chance or they are dependent on each other right; so, in some way or other.

Since the machine is going to do the job automatically we have to give some kind of a tool. So, it identifies certain words that are always together. So, we can probably use the idea of independence in some way and there are also other mechanisms available; it is a very simple rudimentary approach. I just want to explain the concept of independence through that. Two events are dependent then the probability relies on the occurrences of either one of those rights.

So, we have seen that in the conditional probability and if they are independent then the probability of those events would be the multiplication of the probability of individual events or in the other case we can say that the probability of the event E 1 is not really affected by the condition that we have here even in this case it is not really affected by the condition here.

So, let us take the example of average and speed like we are seen; the probability of the average is 0.036 and the probability of speed is 0.114. Do they occur by chance ok? They do not occur by chance if you look at it. They occur as a pair; hence they are not independent events they depend on each other.