Artificial Intelligence (AI) for Investments
Prof. Abhinava Tripathi
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture - 11

(Video Starts: 00:15)

With the same setting as before let us try to find the probability of drawing a blue ball in the second row given that you drew a green ball in the first draw without replacement. By now you have understood that since there is no replacement the second row becomes dependent on the first row. In other words, the second row is conditional on the first row, hence the event becomes a conditional probability statement.

Let us break it down further and mathematically define the conditional probability. If we break the conditional probability, we can define event A as the probability of drawing a green ball in draw 1 and event B as the probability of drawing a blue ball in draw 2. The only thing making B conditional on A is the act of not drawing with replacement. Mathematically we can write it as $P(B \mid A)$ or P(B given A).

The formula for $P(B|A) = P(A \cap B) / P(A)$. We already know how to calculate $P(A \cap B)$ that is the intersection of event A and B. Let us try to apply this formula to a problem that is here probability of A = 5/20. Let us now calculate $P(A \cap B)$ .The total number of ways we can draw a green ball in draw 1 and a blue ball in draw 2 is 5x5 that is 25.

The total number of ways of drawing two balls = 20*19 = 380, hence P or $P(A \cap B) = 25/380$. Hence $P(B|A) = P(A \cap B) / P(A) = (25/380)/(5/20) = 5/19$. If you just try to solve the problem as a simple problem you will get 5/19 as the answer.

Since you know that you already have 19 balls left when you draw a green ball in first draw and you have 5 blue balls left in the bin, well that was easy. Was not it? However, let us solve another problem with which we ended in the previous video. What is the probability of drawing two yellow balls without replacement? To solve this, we can define the events as event A as the probability of drawing a yellow ball in draw 1 and event B as a probability of drawing a yellow ball in draw 2.

Now drawing two yellow balls is P(A ∩ B). Since we are doing this without replacement, we can use the conditional probability formula to calculate this. By rearranging the formula that is P(B given A) = P(A ∩ B) / P(A). That is, we get P(A ∩ B) = P(B given A) * P(A); where P(A) = 4/20.

Now what is P(B/A)? This is the probability of drawing a yellow ball in the second row given we have already drawn a yellow ball in the first row and not replaced it. So, we have 19 balls left. Since we have already drawn a yellow ball 3 yellow balls are left now hence P(B/A) = 3/19. So, we get P(A ∩ (intersection) B) = 4/20 * 3/19 = 12/380 = 0.032.

You can check that this is what you have got earlier as well when we try to solve it more intuitively than rather using the formula of conditional probability. Let us end this discussion on this video with two interesting points. First you look at the formula of P(A ∩ B) for dependent and independent events. The only difference is that P(B) gets replaced with P(B given A).Second the condition for the conditional probability formula that is P(B given A) = P(A ∩ B)/ P(A). For this formula to work the condition is that the denominator is non-zero that is P(A) is not equal to 0.

In this video you will learn the addition rule of probability. Let us begin with a simple word problem related to Venn diagram which you might have solved at some point in our life and other discussions.
Let us consider that in your city which has 100 households there are two newspapers that are published the times and the daily news. The circulation department reports that 25 of the cities households have a subscription to the times and 35 subscribed to daily news. A survey reveals that 6 of all households subscribe to both newspapers. How many of the city's households subscribe to either newspaper? Let us start the solution and discussion pertaining to that.

You know that this can be easily represented as the Venn diagram shown here on the screen. Notice the formula probability of P(A or B) equal to P(A) + P(B) + P(A ∩ B). If A and B are mutually exclusive events then this P(A ∩ B) is equal to 0, here A is the number of households reading the times and B is the number of households using the daily news.

The shaded region that overlaps A with B that is A ∩ B indicates the households which have subscription to both. We already know that A = 25, B = 35 and A ∩ B = 6. our objective here is to find A U B. Now let us divide the three regions as a small a those who read only the times, small b those who only read the daily news and c those who read both. Therefore, A = a + c = 25, B = b + c = 35; A U B = a + b + c and A U B = A + B – c. In other words, (A U B) = A + B - intersection of A and B (A ∩ B)=25 +35 - 6 = 54. Now if you divide these numbers by 100 you will get the probabilities of the different events that is P(A union B) = 0.54; P(A) = 0.25, P(B)= 0.35; P(A intersection B)= 0.06.

Then you replace the equation with probabilities you will get P(A union B) = P(A) + P(B) - P(A intersection B). Addition rule of probability can be understood using the image shown here, the intersection region is something we need to subtract. if A and B are mutually exclusive then P(A ∩ B) =0: P(A U B) = P(A) + P(B).

Joint and marginal probability let us discuss different types of probability. The three major types of probability includes joint probability, marginal probability, conditional probability. To understand joint and marginal probability, let us consider an example to know whether an MBA degree could be a possible factor in the success of mutual funds manager. The table here compares mutual fund performance against the rank of MBA programs from where the fund managers earn their MBA degree.

Let us say we have the data of the performance of 400 mutual funds and using that we create this table. Here we have created the table such that the data is divided in four parts. This is known as contingency table. On the left margin we have the fund managers who graduated from a top 30 MBA program and those who did not graduate from the top 30 MBA programs. On the top of the table, we have divided the mutual funds based on whether they have or have not outperformed the market.

This can also be represented in another way as shown here. If you look closely this is very similar to a two-step experiment, where the first experiment is the fund manager graduating from a top 30 MBA program and the second experiment is fund outperforming or not. This is quite similar to what we did in the machine learning project completion example which was also a two-step experiment.

Now before we calculate different types of probabilities, we need to first convert this into a probability table. To do that let us divide all the values by 400 as that is the total number of data points that we have used that gives us the following table as we can see here on the slide. The value 0.1 in the table means that of the complete data 10 percent of the mutual funds were such that the fund manager had graduated from a top 30 MBA program and also that the fund outperformed the market.

Such a probability is known as the joint probability where we are calculating the probability of two or more events occuring simultaneously. This is also the same as calculating the probability of an intersection of two events. Thus, we can say that the joint probability that the mutual fund outperformed the market and the fund manager is from a top 30 MBA school is the probability of the intersection of these two events which is 0.1 pair.
As shown here on the screen we also have given notations to each of the events for simplicity where A1 is for fund manager graduating from a top 30 MBA program, A2 is for the fund manager that did not graduate from a top 30 MBA program, B1 is for the fund that outperforms the market, B2 is for the fund that does not outperform the market. Here we can write the joint probability of event A1 and B1 is equal to P(A1 intersection B1) which is 0.11.The joint probability of events A1 and B2 which is equal to 0.25. The joint probability of event A2 and P1 which is equal to 0.04 and the joint probability of events A2 and B2 is equal to 0.61. Next let us understand the concept of marginal probability. What if I ask you to calculate the probability that a mutual fund outperforms the market.

Can you answer this from the data provided to you? Yes, you can the probability that the fund will outperform the market will nothing but the sum of all the probabilities shown in rows 1 and 3 which is equal to 0.10+0.04 which is 0.14 or 14%. What we have calculated here is the marginal probability. So, basically the marginal probability describes the probability of an event occurring irrespective of the knowledge gained or the effect from previous other external events.

So once again the marginal probability of B1 is nothing but P(B1) which is equal to 0.14. Now let us try to calculate the probability that a mutual fund will outperform the market given that the fund manager graduated from a top MBA program. Is this a conditional probability problem? Well, yes, it is and we will use the following formula P(B1 given A1) = P(A1 ∩ B1) / P(A1).

From the table the probability that a manager graduated from a top MBA program is a marginal probability hence P(A1) = 0.10 +0.25 = 0.35. We also know that the joint probability which is P(A1 ∩ B1) from the table as 0.1. In other words, there is a P(B1 given A1) = 0.10/0.35 = 0.2857 (or 28.57%) chance that a fund would outperform the market given that the manager graduated from a top MBA program.

Let us close this discussion with one interesting lesson. What happens to the conditional probability formula if the events are independent of each other? In that case P(A given B) = P(A) since P(A1 ∩ B1) = P(A) * P(B) for independent events. This formula essentially states that the probability of A given that the B has occurred is independent of event B.

Similarly, the equation for P(B given A) = P(B) * P(A) will also hold true for the same reason. As an exercise we can check if the event mutual fund outperforms the market and the event fund manager is from a top MBA school are independent of each other or not. So, coming back to the conditional probability, our objective was to find the probability of an event taking place with the information that the previous known event had already occurred.

Such scenarios are very common in the industry and this is where Bayes theorem comes into picture. In the next set of videos, we will discuss the Bayes theorem in the context of conditional probabilities

Base theorem part 1. Let us go back to our earlier experiment where we were trying to understand whether the MBA degree could be a possible factor in the success of a mutual funds manager.

So, let us see what data we had with us. So, to quickly recap this table remember that 0.1 in the table meant that of the complete data 10 percent of the mutual funds were such that the fund manager had graduated from a top 30 MBA program and also that the fund outperformed the market. Now recall how in the previous discussions we calculated the probability that a mutual fund will outperform the market, given the fund manager graduated from a top MBA program, let us quickly go over the steps here. We use the formula for conditional probability where we calculated probability of B1 given A1 for this video instead of calling them B1 and A1 let us just call B and A events for simplicity. So, the events are defined as event A is for fund manager that graduated from a top 30 MBA program.

And event B is for fund that outperforms the market and we said that P(B given A) = P(A ∩ B) / P(A). From the table that we have seen we have P(A ∩ B) as the joint probability of A and B which is 0.1 and we calculated P(A) = 0.35, P(B given A) = 0.10/0.35 = 0.2857 or 28.6 percent.

So, this 0.286 is the probability that a mutual fund will outperform the market,given the fund manager graduated from a top MBA program. Now suppose I ask you the following question, what is the probability that a fund manager has graduated from a top MBA program given that the mutual fund outperformed the market? If you observe the table, you will notice that this is nothing but P(A given B).

So, why do not we quickly calculate and see what this gives us? Using the formula about conditional probability that you learned earlier we can write P(A given B) = P(A intersection B) / P(B). We already know that P(A ∩ B) = 0.1. What will be probability of B? Remember that B is the event that mutual fund outperform the market.

So, if we go back to our table, we can see that B will be the sum of rows A 1 and 3 which is equal to 0.1 plus 0.04 which comes to 0.14; P(B)= 0.10+0.04 =0.14. So, putting in the values we get P(A given B) = P(A intersection B) / P(B) = 0.1 / 0.14 = 0.714. Now remember that before we calculated P(B given A) = P(A intersection B) / P(A)= 0.2857and therefore, we got the probability as 0.286.

And when we calculated P(A given B) we got the probability of 0.1 / 0.14 = 0.714. So, let us write down both the equations here P(B given A) = P(A intersection B) / P(A)= 0.2857. Similarly, P(A given B) = P(A intersection B) / P(B) = 0.714. Now let us rewrite both these equations that we have here. In the first equation if we multiply both the sides by P(A), we can say that P(B given A) P(A) = P(A intersection B). And for the second equation if we multiply both sides by P(B), we can say that P(A given B) P(B) = P(A intersection B). Now if you observe the screen the right side of both these equations exactly the same and is equal to P(A intersection B).

So, if we equate these equations, we can write P(B given A) = P(A given B) P(B) . We can write this equation at different manner noting that P(B given A) = P(A given B) P(B) / P(A). This equation is the most basic form of the Bayes theorem, in other words you can say that Bayes theorem provides a way of relating different conditional probabilities.

Now you might be wondering, what is the big deal here? All you did was some basic linear algebra and came up with some great looking formula, could not we simply calculate P that is P(A given B) and P(B given A) using the table directly. In many cases the joint probability formula of P(A intersection B)is usually not known to us. That is probability of P(A intersection B)is not known to us.

In many practical scenarios you may only have information about P(A given B) and some knowledge about the events A and B using which you may want to calculate P(B given A). This may not seem very intuitive at first but we will have a look at some examples in the next discussions.

Bayes theorem part 2. So, in the previous video we learned that the Bayes theorem comes into picture when a direct calculation of a conditional probability is not possible due to lack of information.

So, let us take an example of such a case, this is one of the most common applications in the medical domain .there is a study going on to understand the likelihood of middle-aged women who develop breast cancer. So, let us say that the past research suggests that the probability of a middle-aged female developing breast cancer is 0.01 or 1 percent. Let us name this event and call this event B, so probability of P (B) = 0.01.

Before we move on let us ask a question do we think the medical test that we do are 100% accurate? Do they always predict correctly? Well, the answer is the accuracy varies from test to test but the breast cancer detection test at times are incorrect. A person can be tested positive for breast cancer but in actuality one may not have it. The vice versa is also possible and these are the major types of misdiagnoses.

We already know that P(B) = 0.01 where B is the event that the middle age female develops breast cancer. Now we want to know how this probability changes if the female tests positive. So, basically, we want to find the probability that a female actually develops breast cancer if she test positive, we cannot treat this as 100 percent because there can be cases of misdiagnosis as well. So, in other words we are looking or asking how reliable is this testing methodology.

Can you tell what type of probability we are looking at here? This is a type of conditional probability because we are measuring the probability of breast cancer given the condition that the test came positive. So, let us name some of these events to avoid confusion later on earlier we called the event of a middle-aged female developing breast cancer as event B. So, let us call the event that a female test positive as event A.

Then the probability that we are interested in measuring is basically P (B given A). To calculate this, we have decided to perform an experiment and here are the results of that experiment. We have a sample of women who already had breast cancer and found that only 90 percent of these women tested positive. This means that the probability of women testing positive given she has best cancer is 0.9.

So, what is this 0.9? Is not this P(A given B)? Remember that event A is the event that a woman test positive and event B is that event that a woman has breast cancer. So, we have P(A given B) is 0.9 and we wanted to calculate P(B given A). So, let us see if we have all the information we need from Bayes rule, as per bayes' rule P(B given A) = P(A given B) P(B) / P(A),so we have P of P(A given B) and P(B) with us, but now we need to calculate P(A).

P(A) is basically the probability that a woman tests positive now this information may not be available directly but we have another piece of information that might help us. In our experiment we also considered women who did not actually have breast cancer and found that 8 percent of these women were misdiagnosed and has tested positive. So, I can say that 8 percent or 0.08 is basically the probability that P(A given B'), think about this.

B is the event that a woman has breast cancer this means that B' which is the complement of B will be an event that a woman does not have breast cancer. And we are measuring the probability that a woman will test positive given the condition that she does not have breast cancer. Hence this probability is nothing but P(A given B'), so let us play a little bit with P(A). It may get a little mathematical here, so let us try to understand each term that we talk about.

We can say that $P(A) = P(A \cap B) + P(A \cap B')$. How did I get here? Remember the joint probability table but this time let us put everything in terms of A and B, for example as we can see here in the table $P(A) = P(A \cap B)$ and $P(A \cap B')$. Similarly, $P(A') = P(A' \cap B)$ and $P(A' \cap B')$.

Let us not be overwhelmed looking at this table, we have just replaced all the joint probability values in their mathematical notation. For example, the joint probability of A and B is written as $P(A \cap B)$ and the joint probability of A and B dash is $P(A \cap B')$. So, how do we calculate the marginal probability of A? Marginal probability of A will be the sum of the values in column B and B'.

This is how we will get $P(A) = P(A \cap B) + P(A \cap B')$. Now let us try to use our conditional probability knowledge to modify this equation a little bit. We can write $P(A \cap B)$ = P(A given B)* P(B) and similarly, we can also write $P(A \cap B')$ as P(A given B')* P(B').

So, let us plug in these values and see what we get. We had P(A) = P(A ∩ B) + P(A ∩ B’) and now we can write this as P(A) = P(A given B)* P(B) + P(A given B’) *P(B’). So, if we plug this in the original equation, we will get P(B given A) = P(A given B)* P(B) / P(A).

Now let us replace P(A) with the equation that we have derived here. So, we get P(B | A)= (P(A given B) P(B)) / (P(A given B)* P(B) + P(A given B’)* P(B’)). So, let us see if we have the data to calculate P(B given A). What are these A and B here? Event A is the event that women test positive and event B is given that a woman has breast cancer.

We already saw that we have P(A given B) = 0.9 and P(B) as 0.01, now we are left with P(A given B’)*P(B’). Remember that in the place of having information about P(A) we instead had the information about P(A given B’) which is nothing but the probability that a woman will test positive given the condition that she does not have breast cancer and we saw that this value is 0.08.

So, finally, we are left with P(B’) this is simple it will be nothing but the complement of B. Recall the complement rule here therefore P(B’)=1-P(B) which is 1 - 0.01 which is equal to 0.99. So, let us plug in these values and we have P(B | A)= (P(A given B) P(B)) / (P(A given B) P(B) + P(A given B’) P(B’))= P(B | A)= (0.9 * 0.01) /((0.9 * 0.01) + (0.08 * 0.99)) = 0.102 or 10.2%.

Hence there is only a 10.2 percent chance that a woman will develop breast cancer if she test positive. What did we make of the test? Yes, we are right it is not reliable at all. So, what did we learn in this video? We saw a simple example where we were able to make use of a few non-conditional probabilities and calculate some new conditional probabilities. And we were able to do this with the help of Bayes theorem which is given by P(B | A)= (P(A given B) P(B)) / (P(A given B) P(B) + P(A given B’) P(B’)). Now let us try to understand the significance of this equation. Generally, we start off with the set of initial prior probabilities you can say that prior probability is the probability of an event before we did our experiment and collected our new data.

In our case the probability of event B which was the probability of a female developing breast cancer is the prior probability this value of 0.01 or 1 percent and we knew this because of our past. Then from certain sources we obtain new information about all the events for which we had the prior probabilities. In our case these sources were the experiments that we performed and got two important pieces of information about the relation between actually having breast cancer and being tested positive for breast cancer.

This information was given to us in the form of conditional probabilities that is P(A given B) and P(A given B’). Our objective was basically to update this our prior knowledge about females developing breast cancer with this new information and calculate P (B given A). This new probability that we calculated is referred to as posterior probabilities and Bayes theorem is used for making these probability calculations.

In our case the probability that a woman has or will develop breast cancer if she test positive was the posterior probability. Now this formula for Bayes theorem is actually most basic form of Bayes theorem because we are assuming that B is dichotomous that we can only break it into B and B’. And therefore, P(B | A)= (P(A given B) P(B)) / (P(A given B) P(B) + P(A given B’) P(B’)).

However, in reality that may not be the case, with this we conclude this video where we saw the application of Bayes theorem.

Recap Bayes theorem, so we have learned that base theorem can be used to calculate posterior probabilities and the formula is given as P(B | A)= (P(A given B) P(B)) / (P(A given B) P(B) + P(A given B’) P(B’)).

In the previous example where we considered event B as the event where a female develops breast cancer there were only two possible outcomes to this experiment either the female had breast cancer or that she does not have breast cancer and the probabilities of these two events happening were denoted by P (B) and P (B’) but what if the experiment could have more than two outcomes. Let us say I break down B into B1, B2 and B3. Can I still apply Bayes theorem in that case? Of course, we can, here we are presenting an interesting case study based on three hospital centres which are being subjected to a malpractice lawsuit. You have the information about a patient being treated in one of the three hospital centres let us name them hospital 1, hospital 2 and hospital 3 and suddenly if we inform you that a malpractice route has been filed.

But we do not know which hospital has the lawsuit that has been filed. Lawsuit indicates malpractice and you do not want your patient to be treated at such a hospital. Hence you decide to find the probability that this lawsuit was filed against which hospital whether hospital 1, 2 or 3. To understand the problem better let us look at the information that we have at hand, let us say we are given the following prior probabilities.

So, let us consider event Bi as the event that a patient greeted at hospital i, thus the probability that a patient was treated in hospital one is P(B1) = 0.6. Similarly, the probability that a patient was treated in hospital 2 is P(B2) which is 0.3. And the probability that a patient was treated in hospital 3 is P(B3) which is 0.1. So, these are our prior probabilities this varies due to various factors like better availability, number of recognized doctors, reputation and various other factors.

Now we have B1, B2 and B3 such that the probabilities of all these three events sum up to 1. Hence in this example B said to be a polytomous in nature while in our previous example B was dichotomous in nature. Next let us learn about the new information available to us. Let us say that the hospitals have shared the probability that a malpractice suit was filed against them in the past. This means that we have the conditional probability of a malpractice would being filed given that the hospital is known to us.

For example, if you see the table provided here you can see that the probability that the malpractice suit is filed given it is hospital 1 is 0.002. The probability that the malpractice suit is file given it is hospital 2 it is 0.005 and the probability that malpractice suit is file given is hospital 3 is 0.007. So, what is the objective of this problem? Let us say malpractice suit is filed today then what is the probability that the suit was filed against hospital 1.

In other words, we want to calculate P(B1 given A) that is probability of B1 given A. So, how do we calculate this? Recall our question of Bayes theorem which was P(B1 | A)= (P(A given B1) P(B1)) / P(A). Now remember that whatever problem you are given you can always start from this formula and then derive the final equation from there.

We already know that P(A given B1) and P(B1) what we do not have with us is P(A). P(A) is nothing but a marginal probability .we had created a joint probability table as we are seeing here and we said that we can calculate the joint probability of A as P(A∩B) plus P(A∩B’) but now we have B1, B2 and B3.

Again, we need to do is simply rewrite this table we have the joint probability of A and B1 as P(A∩B1). The joint probability of A and B 2 is P(A∩B2). In this scenario we have the joint probability of A and B1 as P(A∩B1), the joint probability of A and B2 as P(A∩B2) and so on. Hence the marginal probability of A comes out as P(A) = P(A∩B1) + P(A∩B2) + P(A∩B3).

And we have already seen how to write any of the joint probabilities in terms of their conditional probabilities. So, we can write P(A∩B1) = P(A given B1)P(B1), P(A∩B2) = P(A given B2)P(B2), P(A∩B3) = P(A given B3)P(B3), hence substituting these values back into P(A).

We get P(A) = P(A given B1)P(B1) + P(A given B2)P(B2) + P(A given B3)P(B3). Now let us say we want to calculate P(B1 | A) that is probability of B1 given A = (P(A given B1) P(B1)) / P(A) and we know that P(A) = P(A given B1)P(B1) + P(A given B2)P(B2) + P(A given B3)P(B3).

Hence P(B1 | A)= (P(A given B1) P(B1)) / P(A given B1)P(B1) + P(A given B2)P(B2) + P(A given B3)P(B3). Similarly, we can also write P(B2 | A)= (P(A given B2) P(B2)) / P(A given B1)P(B1) + P(A given B2)P(B2) + P(A given B3)P(B3).

Now let us see if we have all the values with this? We have P(A given B1) as 0.002, P(A given B2) is 0.005 and P(A given B3) is 0.007. we also have probability of B1, B2, and B3. So, what we get P(B1 | A) = 0.002*0.6/(0.002*0.6 + 0.005*0.3 + 0.007*0.1) =0.0012/(0.0012+0.0015+0.0007) = 0.3529 or 35.29 percent. This means if malpractice would take place so that there is a 35.29 percent chance that this suit took place against hospital 1.

Now you can similarly calculate the probability of suit taking place at hospital 2, this is nothing but P(B2 | A) and can also calculate P(B3 | A). Now why did we consider a problem where we could have actually been broken into B1, B2 and B3? The reason is that these scenarios are more common in the industry and real life and that even though you may have seen a lot of complex calculation being made in the segment.

Bayes theorem is not that complex and you can actually extend this to a more complex scenario. Suppose we have multiple events such as B1, B2, B3, B4 and so on up to Bn which are mutually exclusive. Simply modify the denominator of the right side of the bayes' theorem if you are calculating P(Bi given A) then probability of P(Bi given A) = P(A given Bi) P(Bi) / P(A).

Now depending upon the number of events that you have it could be B1, B2, B3 which means three events that could be B1, B2, B3, B4 and all the way up to Bn. Then you can generalize the P A as follows, the generalized form of Bayes theorem comes out to be P(Bi given A) = P(A given Bi) P(Bi) / (P(A given B1) P(B1) + P(A given B2) P(B2) + … + P(A given Bn) P(Bn)).

(Video Ends: 33:55)