Artificial Intelligence (AI) for Investments

Prof. Abhinava Tripathi
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture - 10

(Video Starts: 00:18)

In this lesson we will discuss probability distributions and different types of probability distribution namely binomial distribution and normal distribution. Probability distributions such as normal distribution are often employed to draw inferences from the sample data. For example, how does one make the following inference. What is the probability that a data scientist in the U.S earns between 97000 dollars 215000 dollars.

We will try to explore such questions in this lesson. We will start the discussion with basic rule of probability such as single and multiple events rule of addition and multiplication and joint and marginal probabilities. We will also discuss conditional probability and base theorem. In the context of probability models, we will discuss random variables expected value of random variables variance and cumulative probabilities.

These building blocks on probability models will also help us understand more advanced topics related to sampling statistical inference and hypothesis testing. We will be introduced to important probability concepts such as intersection and union of events, mutually exclusive events and complex events. Probability models empower statisticians to draw inferences from sample data and make statements like the average salary of a data scientist in the US is 97000 plus minus 39000 dollars with 95 percent of confidence.

The exit polls for this election show us with 95 percent confidence that party X will be 53 plus minus 3 seats in the senate. How do we get to these conclusions? In this video we will discuss the theoretical underpinnings regarding basics of probability, probability distributions and statistical inference. We will also discuss the formula and an understanding for using the probability driven inferential statistics in everyday life.

We will also discuss in detail the theoretical foundation to be able to understand these concepts. In the section of basics of probability, we will go through the building concepts probably already heard or have a good sense of the general formula for probability but you will need to learn some more advanced concepts to be able to estimate probability for more complex cases. For example, estimating the probability of multiple events can be more complex than just a single event.

Let us say the probability that a data scientist has a master's degree is relatively easy to calculate. But consider the following questions. What is the probability that a data scientist has a master's degree and earns more than 120000 dollars? What is the probability that a data scientist either has a master's degree or earns more than 120,000 dollars? Are these same or not? If not, then are they at least related and if yes in what way.

What kind of information do we need to calculate these numbers? A discussion of single and multiple events associated rules of probability that is the addition and multiplication rules and joint and marginal probabilities will enable us to take these calls. Moreover, events can interact in other ways too. For example, given that a data scientist earns more than 120000 dollars what is the probability that they have a master's degree.

Is it the same as the reverse condition that is given a data scientist has a master's degree? What is the probability that may earn more than 120000? Are the two at least related? If yes then in what pay? Such conditional event relations will be covered in the section on conditional probability and the topic of base theorem will give us a very powerful mathematical formula for calculating them.

All of these building blocks will help us move on to the next building block which is probability distributions. Probability distributions as the name implies help us visualize and quantify how probabilities are distributed across the range of possible values. For example, what is the probability that a data scientist earns more than 97000 dollars and less than 115000 dollar. What is the probability that done between under 115000 dollars and 133000 dollars?

As you can see from this distribution even though both the intervals are about 18000 dollars long the probability of earning 115000 dollars to 133000 is much lower. That is because of how the probabilities are distributed and in the section of probability distributions we will deep dive into these concepts of discrete variables and continuous variables. Having learned all these building blocks you will be able to proceed to the final section on statistical inference which is the one that will give you all the tools to draw inferences from samples for entire populations.

We will start out with the primer on samples and sampling going through the nuances of how to sample and ensure that our sample for example 2000 data scientists whether it is a good representation of the entire population which here would be all data scientists in the US or not. Once you have the samples, we will go through the process of making sampling distributions from this data.

And then using the central limit theorem to extrapolate from it and make inferences for the entire population. In other words, we will use the sample data to estimate the average data scientist salary which will be in the form of an interval estimate. For example, 97000 dollars plus minus 39000 dollars which is margin and 97000 dollars is average and with an associated degree of confidence we will learn to estimate the confidence intervals as well.

In the industry there are a lot of tools that can be used to directly apply different statistical techniques. Some of these tools are like python, R, SAS, SPSS, Excel and Minitab. However, the most common tools used in the industry are python and R and this is true for a variety of reasons. Both python and R provide a wide variety of statistical techniques such as linear and non-linear modelling etcetera.

Also, you can easily create quality plots and include mathematical symbols and formally with these languages. Unlike many other tools these python and R are free to use. In particular R is very popular among statisticians. So, in this module and in the course, we will be using R as a primary tool for demonstrating different concepts. At some point in life, you would have heard that the probability of something happening.

For example, a stock giving returns above certain level is 80 percent or a particular party winning election is 50 percent. But what is this probability and what do these numbers mean to us. Let us try to understand this with a simple example. In most of the sports that you watch whether it is cricket, football, soccer or basketball, tossing a coin is a very common feature. In cricket the coin might be tossed to decide which team gets to bat first or ball first.

While in soccer it might be to decide who kicks off in the first half. Now in these coin tosses are you always sure whether it is going to be a heads or tails? Well, no assuming that it is a fair coin you may not. So, let us take an example of a fair coin. A fair coin is a coin that will have the structure or the make identical for both the heads and tail sides. So, that when you toss it the chance of getting head is same as the chance of getting tails.

Now let us go back to the example where a coin toss is happening between the Liverpool team and Manchester United team in the English Premier League. The captain of Liverpool calls for heads what do you think will be the chance that Liverpool captain wins. Let us try to calculate. It you can have only two outcomes heads or tails. Since both are equally likely to happen as in the case of a fair coin, the chance that the Liverpool captain wins is calculated like this.

Number of favorable outcome / Total outcomes = 1(heads)/2(heads and tails) =0.5
The chances of something happening like this is known as probability. In other words, in this particular case, you can say that the probability of winning the coin toss is 0.5 or 50 percent. Thus, probability is a language of uncertainty. Now let us formalize some of these terms related to probability.

The coin toss is known as an experiment here. The set of possible outcomes that is head and tails is known as the sample space. Sample space can be mathematically represented as Ω (omega) symbol which is equal to two possible outcomes that is heads and tails. In our example the probability of the Liverpool captain winning is known as the event here. Let us consider rolling a dice that has six faces. What is the random experiment here, it is the rolling of a dice.

What are the possible outcomes they can be 1, 2, 3, 4, 5 or 6. Hence we can write our sample space as follows Ω (Omega) = {1, 2, 3, 4, 5, 6}, all six events are possible. Now what if I ask you what

is the probability of getting 1 in the rolling of a dice. You can easily calculate this with the formula: Number of favorable outcome / Total outcomes = 1(getting a 1)/ 6(1, 2, 3, 4,5, 6) = ⅙.
What about the probability of getting an even number?

Here the number of favourable outcomes are 3 that is 2, 4 or 6 and the probability becomes 3/6 = 0.5. In this example rolling a dice is a random event or random experiment that is rolling one or rolling an even number here are the events for which we find the probability. A random experiment is a process that leads to one of several possible outcomes. An event is just a subset of all the possible outcomes of a random experiment.

I hope this helps us differentiate between a random experiment and an event. Now what if I say that the formula of probability that we have discussed till now that is the number of favourable outcomes divided by total outcomes may not apply at all such instances. Let us consider the marks scored out of 100 by a student in a math test. Here the random experiment is that the student scoring marks in a math test what can be the outcomes of the test.

The possible outcomes can be 10, 20, 99 or even 0 or 100 essentially any number between 0 and 100. Let us consider that the evaluation happens only in whole numbers and not in fractions. so that the sample space becomes the set of whole numbers between 0 and 100 which is {0, 1, 2, ….,100} that is 101 possible outcomes. Now if I ask you what is the probability of the event that a student score 80 in the math test.

Will you apply the probability formula that we have just seen that is number of favourable outcomes divided by total outcomes. Let us try it out. According to the formula that will be : Number of favourable outcome / Total outcomes = 1(getting 80)/ 101(0 to 100) = 1/101, will be the probability of a student scoring 80 in that test. Do we think this is right? May not be so because not all the experiments are designed to make all outcomes equally likely. Just like you can have an unfair point you can have a test that is designed such that the people at least get a passing rate or at least a certain set of minimum grade.

So, how would we calculate the probability in that scenario? We can always use past data to find out empirically. For example, if I tell you that there were 50 students in the class who took the math test out of which 10 students scored 80, 10 students scored 90, and the rest score 60. Then what is the probability of a student scoring 80 marks? If you use the formula that is number of favorable outcomes/ Total outcomes = 1(getting 80)/ 3(60, 80,90) = ⅓ because the sample space becomes {60, 80 , 90}, this would be incorrect.

Hence in this case the formula has to change a little bit and you can consider it as frequency of favourable outcome divided by total number of outcomes. Frequency is essentially the number of times an experiment is repeated with the outcome. In our case the experiment is repeated for 50 students and the frequency of favourable outcome is 10. Hence the probability becomes 10/50= 0.20.

Now there are certain important things that we should note here. The formula that is number of favourable outcomes divided by total outcome worked in the case of coin tossing and rolling a die experiment because first the different outcomes of experiments do not depend on what happened in the past in the case of dice and coin. Second different outcomes for example heads and tails or the case of coin toss and 1, 2, 3, 4, 5, 6 in the case of die roll are equally likely.

The chance of one coming up is same as any other number in a die roll. Similarly, the chance of heads is equally likely as the chance of getting tails in a coin toss. However, this is usually not the case with most real-life experiments. As noted earlier this is also not the case with the math exam which is designed in a way that one outcome that is scoring zero probably and most likely is not same as scoring 100 or 60 or 80.

By now you have understood that we calculate probability when we want to understand the chances of an event happening in a random experiment. Probability is associated with events that have unsure outcomes. In fact, most things in life have unsure outcomes. For example, it might or might not rain on a given day hence there are two possible events rain and no rain. And each of these two events have a certain probability associated with them.

Probabilities are always between 0 and 1. Let us look at a formula to understand this better. Probability = Frequency of favorable outcomes / Total frequency of all outcomes. Now the frequency of favourable outcomes can be zero for example getting 7 in a die roll or maximum equal to the total frequency of all possible outcomes. For example, if you define the event as getting 1,2, 3, 4, 5 or 6 in a die roll the event will have a total probability of 1.

Hence the probability of an event in a random experiment can only be between 0 and 1. A probability of zero means that us with a certainty that event will not occur. A probability of 1 means that an event is certain and it will occur definitely. For example, the probability that a day will follow the night is always one unless some cosmic event occurs where the sun disappears. let us hope that does not happen.

At times probability is also represented as percentages. So, 0 correspond to 0 percent and one corresponds to 100 percent. You may have watched the weather channel and they will tell you that there is a 20 percent or 30 percent chance of rainfall on any given day. Someone may give you an investment advice and they may tell you that investment has a high probability of succeeding or not succeeding. In medicine you have certain probabilities.

You may have a certain probability for a surgery being a success. In business you may have a certain probability of customer churn out. A probability that a customer may pay a loan back or default on the loan. If you are in sales or marketing, you may assign a certain probability to a customer purchasing a product and there are various ways of calculating probabilities. There are so many examples as if life itself is a game of probabilities, is not it?

Combining two or more experiments. Let us say that instead you flip two coins, what would be the sample space here. Let us figure it out. The experiment of tossing two coins can be thought of as two steps in which step one is the tossing of the first coin and step two is the tossing of second coin. Now if you use H to denote a head and T to denote a tail then H, H indicates that the experiment outcome with the head on the first coin and another head on the second coin.

So, if we write down all the possible outcomes, we will get Ω (Omega) = HH, HT, TH, and TT,and therefore the total possible outcomes would be 4 that is 2 x 2. What do you notice? 2 is the sample space of first coin HT and also for the next two it is the sample space of second coin toss as well in step two that is HT. Now if we are to generalize this for a multi-step experiment, we can say that if an experiment can be described as a sequence of K steps where the possible outcomes are n1, n2,…nk then the total number of possible experimental outcomes would be Ω (Omega) = n1 x n2 x …x nk. In such multi-step experiments we can represent this in the form of a tree diagram. Step 1 corresponds to the first coin toss and step 2 corresponds to the second coin toss.

For each step there are two possible outcomes head or tail. On the right end of the tree, you can see all the possible outcomes of this experiment. You would notice some key points here. First one is that experiments need not be only a single step they can be multi-step. Consequently, the sample outcome can be a complex representation of the multi-step experiment and it need not be just numbers. Now let us extend our learning to a more real life use case.

Consider that you are working in a data scientist role and you are starting a project that involves two stages. One stage is the data preparation stage and the other stage is the model development stage. Your task is to estimate the possible completion time for the project. Here is some more information about the project. An analysis of similar projects in the past revealed possible completion time for the data preparation stages five six or seven weeks which is equally likely.

And possible completion time for the model development stages second two three or four weeks again it will be equally likely. What is the random experiment here? It is the two-step process as part of the machine learning project which involved the data preparation and model development stage. What do you think can be different events here. For example, can completing the data preparation stage in five weeks be an event here? Yes, why not.

Can completing the machine learning project in exactly eight weeks be an event? Yes of course. Can completing the model development staging three weeks is an event? Yes, it is. Now let us try to find out the possibilities and probability of the occurrence of these events. For that let us create a decision tree like we created in the coin tossing example discussed earlier. Here in the last column, we added the numbers in the sample outcome to get the completion time of the whole project.

As seen earlier The number of sample outcomes= (number of outcomes in step 1) * (number of outcomes in step2) = 3 * 3 = 9. Now let us find the probability of completing the data preparation stage exactly in five weeks. You can see that the frequency of favourable outcomes are three the first three hence the probability is C by 9 which is 1 by 3. Similarly, what do you think is the probability of completing the model development stage in exactly three weeks.

Again, since there are three possibilities that is two three and four weeks out of which 1 is 3 weeks the answer would be 1/3. Then what is the probability that the project will be completed in exactly eight weeks? We can see that in two of the nine experimental outcomes the project gets completed in eight weeks thus the probability is 2/9 or 22.22 percent. Similarly, we can also compute the probability that the project will completed in nine weeks or less.

Types of events. In this video let us compute and estimate the probability of finishing the project. For example, in this particular case if we are estimating the probability that project will be finished in nine weeks or less, we can see that there are six possible outcomes out of total nine outcomes that satisfy this condition that the project is being completed in nine weeks or less. Hence the probability of project being completed in nine weeks or less is 6 /9 that is 66.67 percent.

Now what is the probability that work will be completed in more than nine weeks? Also is this question somehow related to the previous question. There are three possible outcomes out of the total nine possible outcomes where this condition is satisfied and therefore the probability works out to 1 by 3. Now if I can subtract this one by three from total probability one, we get 2 by 3 which is 66.67 percent.

This is precisely the probability of getting the project completed in nine weeks or less. How did we do this? We could do this because between these two events let us say the project being completed in nine weeks or less. We call it event A and project being completed in nine weeks more is event B; we can call them complementary events. In other words, the sample outcomes or sample points that are not in event A that is 10, 11, 12 are the sample point of event B and vice versa.

Hence, we can write event B as event A^C or the complement to event A. This can very well be represented as a Venn diagram. As you can see in this Venn diagram the area enclosed within the rectangle is the complete sample space for the experiment and the area shaded in the yellow represents event A. Thus, the complement of event a that is event A^C which in our case is also event B is represented by the area that is not in A which is the white portion within the rectangle.

Since the probability of complete sample space is 1 we can say that $P(A) + P(A^C) = 1$ which is the probability of event A plus probability of event A^C or even B is equal to 1 where P(A) represents the probability of event A and $P(A^C)$ is the complementary area to A or event B. This rule also extends to the basic rule of probability that is the sum of probabilities of all the events in a given experiment always add up to 1.

We learned that an event is a subset of all possible outcomes of a random experiment or a subset of the sample space. Within events we can further classify events into simple events and compound events. A simple event has only one outcome whereas the compound event has two or more simple events as outcomes. Hence in our machine learning project completion problem which is a multi-step experiment, compound events like completing the data preparation stage in exactly five weeks can be broken down into three simple events. Completing the data preparation stage in five weeks and the model development stage into completing the data preparation in five weeks and the model development stage in three weeks. And lastly the third one is completing the data preparation stage in five weeks and the model development stage in four weeks.

Note that all the three events have just one outcome in the samples. In the example of rolling a dice the event rolling a one is an example of a simple event. But they went rolling an odd number will be an example of a compound event as it can be broken into three simple events rolling a one rolling a three or rolling a five. Now let us consider the event completing the data preparation stage in five weeks and model development stage in two weeks.

We can see that it has a composite or intersection nature of two events. The first event is completing the data preparation stage in five weeks which can be said as event A and the second event may be completing the model development stage in two weeks for event B. For Event A, the sample outcomes are (5, 2), (5, 3) and (5, 4). For Event B, the sample outcomes are (5, 2), (6, 2) and (7, 2). The intersection or the common of these two events is 5, 2 The intersection or the common of these 2 events is (5, 2): A ∩ B.

Mathematically this is represented as A intersection B, intersection is represented by inverted U symbol and in this case that outcome is 5, 2. In the Venn diagram shown here on the screen events A and B are two events that have some overlapping occurrence hence the area shaded in green is the event representing A intersection B. The probability of completing the data preparation stage in five weeks and the model development in stage 2 is 1.

So, there is one possible outcome 5, 2 and therefore the probability one by nine that is the possible outcome and the divide by total outcomes. Let us try to understand this using a simple example where two dice are rolled simultaneously. Can you tell how many outcomes will be there in the sample space of course it is 6x6=36. Let us define two particular events, event A and even B as follows. Event A where die rolls and the first die has a number as one.

There are six possible outcomes for event A: {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1,6)}. Here you must note that each outcome has two values as they represent the individual outcomes of each of the two dices. So, if I consider 1, 4 as event we mean that the first die is rolled as 1 and the second die rolled four. Similarly, you can see all the possible outcome of event a as listed here. Now, we can define event B also where the dice rolls as the number five.

Again, the possible outcomes are: {(1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (6, 5)}. The intersection of events A and B (A ∩ B) of these events is (1,5). Going back to our machine learning project completion example what if I ask you to determine the probability of the event. Completing the data preparation stage in five weeks or completing the model development stage in two weeks.

As seen earlier; let us name these two events first completing the data preparation stage in five weeks as event A and completing the model development stage in two weeks as event B. For Event A, the sample outcomes are (5, 2),(5, 3) and (5, 4). For Event B, the sample outcomes are (5, 2), (6, 2) and (7, 2). The or criteria represents the union of these events and we determine this by writing all the possible outcomes that belong to these two events which are which are {(5, 2), (5, 3), (5, 4), (6, 2) and (7, 2)}.

Mathematically the union of events A and B can be represented as A ∪ (union) B or ‘A or B’ which is 5 2, 5 3, 5 4, 6 2 and 7 2. If we use the Venn diagram between event A and B then as shown here the shaded area in green represents the union of A and B. Now let us again consider the dice example where we are tossing the die two dice simultaneously. Event A is that the roles where that first die has number one.

So, the sample outcomes or sample space is 1 1, 1 2, 1 3, 1 4, 1 5, 1 6 that is six possible outcomes. Event B is the rows where second die has number five and hence the sample outcomes or sample space is six possible outcomes that is 1 5, 2 5, 3 5, 4 5, 5 5 and 6 5. So, for the outcomes of event A union B or A or B we have listed all the possible outcomes present in event A, all the outcomes event B and the common event which is 1, 5 which is common to both event A and B.

And therefore, the overall number of sample space events are {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2,5), (3,5), (4,5),(5,5),(6,5)}. These are all events that are part of A ∪ B or A or B so, now that we have understood the event complements and finding the intersection and the union as well for two events let us move on to another type of event which is mutually exclusive events in the next video. Mutually exclusive events are also known as disjoint events.

So, we can say that two events are mutually exclusive when they do not occur at the same time. For example, if a student has been awarded grade C in a particular subject in an examination, then they cannot be awarded grade B in the same subject in the same examination. So, the events number one a student being awarded grade B in a subject in an exam and number two a student being awarded grade C in the same subject in the same examination are mutually exclusive or disjoint events.

We can see that in the Venn diagram shown on the screen here you can see that events A and B do not have an overlapping region. This is a clear indication that the two events are mutually exclusive or disjoint. One interesting fact for you to work out is that complement events are always mutually exclusive but not the other way around. Let us take the rolling two dice example and define the following events.

Event A is Rolls where the first die has the number 1. Sample space: (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6). Event B is Rolls where the first die has the number 3 and the second die has number 5. Sample space: {(3, 5)}. Looking at these two events event A and even B of die rolls you can see that two events have nothing in common that is there is no intersection between these two events. Hence these two events are called as mutually exclusive.

Probability for complex events. Until now you will learn what complement events and mutually exclusive events are and how to find the intersection or union of events. We have worked a lot on the machine learning project completion example. Let us now work on a different classic problem. Let us say that you have a bin containing four balls of different colours yellow, green, red and blue.

Now let us begin with a simple problem in how many ways you can draw two balls from the bin with replacement. This means that once you draw a ball you put it back in the bin for the second draw. In the first draw you have four balls so four base. And similarly, you can draw four ways in the second row. Hence you have 4* 4 = 16 ways of drawing the two balls with replacement. Now what if I ask you how many ways you can draw two balls with replacement such that both are yellow.

In each draw you have only one yellow ball so there is only one way in which both the draws are yellow balls. This can also be understood by drawing a tree. So, now having known the two values, those two values are number one number of ways to draw two balls with replacement is equal to 16 and number of ways to draw two balls with replacement such that both draws are yellow is equal to 1.

Thus, we can say that the probability of drawing two balls with replacement such that both draws are yellow =1/16. Now let us look at another way to find the probability of complex events. What is the probability of drawing a yellow ball? It is P= ¼. Now if you remember drawing two yellow balls with replacement is a compound event and is the intersection of two simple events.

First event being drawing a yellow ball in first draw that is event A and drawing yellow ball in the second draw that is event B. For both the events the probability is 1/4. And if you observe closely the final answer is nothing but ¼* ¼=1/16. In other words, P(A∩ (intersection) B) = P(A) * P(B).

This is known as the multiplication rule of probabilities. This multiplication rule works only when there are two events and those are independent. Events are independent if and only if the occurrence of one event does not affect the other. So, if there are four independent events A, B, C and D then the probability of all these four events happening together = P(A)*P(B)*P(C)*P(D).

This multiplication rule can also be applied for dependent events and the formula changes slightly. Now you must not confuse independent events with mutually exclusive events. Mutually exclusive events are events that cannot happen together and we can represent this with the help of a Venn diagram. On the other hand, independent events are such that the occurrence of one event is not dependent on the other event and we cannot represent the same using a Venn diagram. So, here is a simple question for us.

What would happen to the probability of A and B happening together if the events are mutually exclusive? As you know mutually exclusive events cannot happen together thus the probability of A and B i.e. P(A and B) or P(A∩B) becomes zero. In practice if you want to figure out whether two events are independent or not you can find the probability by creating a tree and cross verify by applying the multiplication rule.

Now let us extend this example to a more complex one. Let us consider that the bin now has 20 balls and that out of all the 20 balls, there are four yellow balls, five green balls, six red balls and five blue balls. What is the probability of drawing a yellow ball? You can draw four ways out of 20 so it is 1 by 5 which is equal to 0.2. Now what is the probability of drawing two yellow balls with replacement? Since we know that it is this event is an intersection of two events.

The event first or event A being drawing yellow ball in the first draw and event B for the second event being drawing yellow ball in the second row. Both of these are independent events. Let us apply the multiplication rule that is $P(A \cap B) = P(A) * P(B) = 4/20 * 4/20 = 1/25 = 0.04$, when the multiplication rule made our life very easy. But now let us take the same sample to see what happens if the events are now dependent.

What is the probability of drawing two yellow balls if we do not replace that ball after we have drawn it, well of course that changes the probability. So, we still have one fifth chance of drawing the first yellow ball but for the second yellow ball the probability is no longer one-fifth. This is because now we have a bin that has 19 balls in it and three of them are yellow. So, the probability of drawing two yellow balls without replacing it now is $=1/5 * 3/19 = 0.032$

So, this was a simple example of applying the multiplication rule for the dependent events. Because when we drew the ball and did not replace it the probability of the second event got affected.

(Video Ends: 34:26)