Module 2: Thinking About Risk and Uncertainty Through Probability Distributions

Video Transcript

Video 2.1: Module Introduction (4:40)

In the previous module, we discussed the importance of models to enable the use of data within decision processes. In this module, we will introduce the concept of probability distributions as a key concept in modeling, particularly in situations when you have uncertainty in business environments, which is almost always the case.

Now there are many standard probability distributions that are very useful and we will discuss several examples throughout the course and specifically in this module. Distributions helps us to describe uncertainty of important quantities by capturing the relative likelihood or frequency of each possible value. Probability distributions have two elements. The first one is called the support that captures the range of the possible values that the quantity can get. This range or support can be discreet, like the numbers 1, 2, 3 and so forth, or continues essentially all the values, for example, in the interval between zero and one. So, that's the first element. And the second element is the relative likelihood of each value in the range.

We will discuss several examples of both discrete and continuous distributions. Broadly speaking, there are empirical distributions and standard distributions. These are the two main types of probability distributions that you're going to find. Empirical distributions are created in many ways.

For example, by collating expert opinion. For example, let's assume that you are not sure about what the value of a certain quantity is, you can go to multiple experts and ask them to assess what is the value of the quantity. And then you can take the collection of assessments that you received and by essentially put on each one of them an equal or an unequal value and create from that an empirical distribution.

Another way to create empirical distributions is by using existing data that you have collected either by your regular day-to-day systems or by specific experiments that you conducted. So, for example, you can use your sales data. You can take all the values that you've seen in the past and put on each value an equal weight and that will immediately, will provide you with an empirical distribution that you can then use to inform decisions going forward.

And finally, you can also use computers, computer simulations to generate for you values of different scenarios and again take those values and construct empirical distributions. And there are more ways to construct empirical distributions, these are just a few typical examples. In addition, as I mentioned, there are also what we call standard probability distributions. This is another way to create distributions. You can think about standard probability distributions as mini models that were developed sometimes to approximate a quicker empirical distribution or to model physical phenomena. As such, standard distributions, like any model, have parameters or knobs. The input in this case will be a request for certain value from the distribution and the output will be a random draw from the model distribution.

Now, why do we even need standard probability distributions? There are several reasons for that. First, they are often more tractable. They often have close form to express some of the main statistics of the probability distributions. For example, the standard deviation, the median, the mean, and so forth. And, moreover, you can use those close forms to implement in a computer, in a very easy way, a way for you to just tell the computer, "Hey, here's the distribution compute for me the main statistics."

That's something that you often cannot do with empirical distributions, you will have to more numerically compute that. The second benefit of standard distributions is that it allows us to model situations and phenomena, even when we do not have a lot of data or not have a data at all. So, that's a major advantage. There are many standard probability distributions that are very useful, and we will discuss several examples throughout the course. And in fact, in this module, we will already mention a few of the main ones that you also have to read about in the supporting material.

Video 2.2: Decisions Regarding COVID-19 (15:53)

So, next we will come back to the SIR model that we introduced in the previous module and that you had to read about in the preparation for this module. The goal of the discussion is to show one more example on how to use models to inform decisions under uncertainty which, as I mentioned, is a major factor in almost any business decision and particularly in the context of pandemic management such as COVID-19.

In particular, we would like to show how to use the SIR model and probability distributions to model uncertainty in this setting. Recall, what the model capture. The model captures a relationship, and hopefully, you had time to read about the SIR model specifically and develop some intuition regarding how this model can be used and what each of its parameters means.

Now, again what the SIR model is about, remember there are three sets that we call S, I, and R that represents the Susceptible people in the population, the Infected people in the population and the Recovered people in the population. And the relationship between these three sets allows us given today's values of S, I, and R, it allows us to describe or predict what these values are going to be tomorrow or in the next day.

Now, we will start with the assumption that the model's parameters are known deterministically with certainty. And then, we're going to expand the scope to allow the model to capture uncertainty. This will help you understand the concept of distributions of probability distributions and how they can support decisions that are being made under uncertainty.

So, just to make it concrete, we will use the following very simple use cases to illustrate how you may use a model like the SIR model to support some decisions that are relevant to pandemic management. So, imagine that you are managing a health system and let's say that the first question that you have to answer is the following.

So, suppose your healthcare system has a capacity to accommodate maximally 180 infectious persons. And what you would like to know, is this capacity sufficient? And you would like to use the model to try and assess the answer to that question. Now, the second question that we will discuss is a little bit more involved, in which case you currently have capacity to manage only 120 infectious persons.

However, you're working hard to build up your capacity and increase it up to 180, but that will take you about two months and again, what you would like to ask yourself in this question is the current capacity is going to be sufficient until I will be able to expand it and then once I expand it, will I have enough capacity to manage the peak of the infections? Okay, so these are two simple questions. Maybe not the most realistic one; reality is always more complex, but hopefully will be realistic enough for us to understand how you can use models to inform typical decisions like this, okay? So, our starting point would be to use the SIR model without uncertainty, right?

So, what we are going to assume here again that the parameters of the model and the input of the model are completely deterministic and are known in advance. So, in particular, we're going to consider a population of 1000 people. Again, this is not a realistic number, but for the sake of the discussion, that will allow us to develop intuition, right? And then, basically, the specific input to the model will consist of the number of people out of these 1000 people that belong at the moment to S, I, and R as initial values in the initial period when we start thinking about the system, right?

And the specific example that we will consider, we'll have 990 people in the set S and then only 10 infected people in the set I and no recovered people. So, the set R has nobody, right? And remember, we are going to consider for the SIR model two important parameters, beta and gamma. Beta reflects or represents the amount or the level of infectivity of the virus. And we are going to consider it as the value of 0.1, as you can see here.

And then gamma is going to reflect how quickly people are recovering from being infected or, in other words, the fraction of people recovered every day. And we're going to consider gamma to be the value of 0.5, okay? So, these are the two parameter values. These are the values of the input and let's see what we can do with that, right.

That basically, it's actually quite easy or not that hard to take the mathematical relationships that the SIR model is capturing and code it in a computer in a way that will allow you given the initial input of S, I, and R and the initial decision about what the parameters beta and gamma values are. And then what you can do with the computer is simply simulate the model output, again as a proxy of what you think is going to happen for the next 200 days, right?

So, you will be able to practice that after the module, but basically, that's what we are going to do. We are going to use the relationship that is being captured by the SIR model to calculate the value of the sets S, I, and R for each day in the next 200 days. And in the following figure, you can see a way, a quite handy way, to visualize the output of the model, right? What you see is that naturally, the total size of the population always sums up to 1000, right?

That's one of the assumptions of the model that maybe you've thought about that nobody dies. So, nobody gets out of the population. You can modify the model to capture that, but we're going to stick with it. This simplified assumption, right? And what is changing is the allocation of those 1000 people between the sets S, I, and R, right?

And this is being captured by the different colors that you see on the figure that represents how many people are currently in S, I, and R, and how does this change over time, right? And what you can see here that we start with a lot of people that are susceptible; they are in the set S, and very few people that are infected, right? And what is happening over time is that the number of infected individuals in the set I is increasing. So, in fact it behaves like a wave that first increases in an accelerated rate, right?

Then the rate of growth slows down until finally, you arrive to the peak of the of the number of people that are infected. That's the peak of the infection wave, sometimes being called like that. And then the number of infected people starts to decline. And then, in the end of the wave, we have a lot of people that were recovered, right? They've moved from the set S, through the set I, through the set R, but we still have some people in the set S that are still susceptible but were not infected during the current infection wave, right? So, that's kind of describing to the output of the model.

Now, we are ready to think about the use case questions that we posed before. So, given the output of the model that you've just looked at, right? And we're looking at this figure, right? I would like us to consider how are we going to use the SIR model output to try and think how are we going to respond to the scenario questions about the hospital capacity, right?

So, the first question was, do I have sufficient capacity? And if you think about the first question, it essentially boils down to asking yourself if the peak number of infectious people is going to be lower or higher than your current capacity, which is 180 people, right? You can take care of 180 patients, right? So, when I look at this picture, right? What you can see here that at the peak, the number of infectious individuals is actually 160, right? Which is essentially lower than your current capacity. So, based on the model, based on this scenario, this is a reassuring answer that you are likely to have sufficient capacity to handle even the peak of the infections.

Now, the second question is a little bit more involved, right? Like you start with reduced capacity of 120, and you are hoping to build up your capacity back up to 180 in about 60 days, right? Now, one thing that you can observe here that we're going to arrive based on the model; the peak is going to happen around day 99 after the infections or after the pandemic wave starts, right? So, that's nice but what you really want to ask yourself is what happens after 60 days, right? Because this is going to be the time that you are assuming is going to take you to build up your capacity, right?

So, let's just look at what happens at day 60, right? We can do that, and what we can see there, that the number of infected individuals is expected to be about 112, which again is good news because that's still lower than your current capacity, which is 120, right? So, taking all of that into account, the answers that the model is providing are quite reassuring. And the next question is, are we done? Are we now happy? Can we just finish the discussion and settle it, or maybe there are some reasons to still be concerned?

Now, you always have to remember that the answers that the model provides hold true in what we call the reality of the model, right? And there is no immediate guarantee that they also hold in the real world, right? You always have to ask yourself to what extent and how likely it is that the model that you're using or the model reality that you're using, right? Translate to the true reality in the real world. And this is a hugely important question.

So, one way to answer this question or answer these questions, right? Is to look on different scenarios, stress your model, right? So, let's just look at the following table that shows you to what extent the model's output is sensitive to changes in both the model inputs and the model's parameters, right?

So, every row in this table corresponds to a different scenario, right? The first row is the baseline scenario we just discussed, right? With the current input values and the current parameter values of beta and gamma, right? Now, let's just take a look at the second scenario, right? In the second scenario, we essentially assume that the virus is slightly more infectious. And that means that beta is now having a higher value, right? It's 0.125 instead of instead of 0.1, like we like we had in the baseline scenario, right? And what are the implications; right? And if you look on the implication, you can see that there is a very large change in both the peak number of infected individuals, right? It's now 241 versus 160, but also, with respect to how quickly you reach that peak, and now you reach the peak in 67 days versus the prior value, that was 89 days, right?

So, most importantly, under this scenario, you are not faring that well at all, right? In this scenario, you are likely to run out of capacity, maybe by a lot. So, when you look at this table, your confidence in this model's singular outcome under the baseline scenario should be more than a little shaky, right? You actually see that this model is very sensitive to both the inputs' values and the parameters' values. So, what is the general lesson; right? So, unless you are absolutely sure in your current estimates of the parameter’s beta and gamma, you cannot rely merely on the model's output with respect to the baseline scenario. In all likelihood, you cannot be sure to have accurate estimates in this case, right?

We still, even today, struggle to understand what is the infectivity level of the Coronavirus, and in fact, it could be very, very different, right? Different variants can have different level of infectivity, how people behave can have impact on the level of infectivity. So, this specific parameter is actually in all likelihood, in reality, it's not going to be easy to estimate very accurately, right?

So, you don't only have a theoretical reason to be concerned here. You actually have a very practical reason to be concerned here because if you think about the parameters in this specific model, and that's very typical in many other models, it's actually not that easy to estimate them very, very precisely, right?

Now, what is nice about the scenario analysis that we just explored, is that it can help guiding your understanding about which of the parameters is most sensitive to estimation errors and will have the biggest impact on the model's output and that should make you alert to where you put your attention in trying to understand how the model behaves under different scenarios. What we're going to do next is going to think about how you deal with this situation in a more rigorous way, where you're not just looking at what the possible values are, but you also want to understand which one of them is more likely. That's coming next.

Video 2.3: Applying the SIR Model With Uncertainty: Part One (7:22)

So next, we will discuss a situation when we have uncertainty about the models' parameters and in that case we're going to model the parameters as probability distributions instead of fixed values. And why is that important as opposed to just simply take different scenarios like we did before. It is important because in such situations it's really important for the decision maker, not only to understand what the possible values are, but also to understand what the respective frequencies or likelihood of each value occurring are. So, that's going to be something that we're going to do with the help of probability distributions.

But one thing that is very important to understand that they get going is that when the input variables or parameters into the model can take a range of values and have specific respective likelihood or probabilities. Then also the output of the model is likely to take a range of values. So, when you think about uncertainty is an input, we model and uncertainty in the input parameters in this case specifically, then that will induce an uncertainty on the output.

And that's what we need to understand better in the following discussion. How do you think about the connection between the uncertainty in the input and the uncertainty in the output? its simplest manner, a distribution summarizes how often different values or scenario of, scenarios of care. And let's just look on some examples next. So here's in it, one way to model the uncertainty of the parameter beta. We are going to use one of the most commonly used standard distributions that is called the uniform distribution.

This mini model has two parameters: The maximum value and the minimum value. And what the assumption is that over that interval that is being defined by the minimum and maximum values the likelihood of each value in that range, which is a continuous range, it's the entire interval, is equally likely. And this is what is called again, the uniform distribution. And just to develop intuition, we mentioned that uniform or standard distributions approximate empirical distributions. And why does it make sense or when does it make sense to use a uniform distribution?

Now, the notion of distributions codifies very common-sense idea. At its simplest manner, a distribution summarizes how often different values or scenario of, scenarios of care. And let's just look on some examples next.

So, here's in it, one way to model the uncertainty of the parameter beta. We are going to use one of the most commonly used standard distributions that is called the uniform distribution. This mini model has two parameters: The maximum value and the minimum value. And what the assumption is that over that interval that is being defined by the minimum and maximum values the likelihood of each value in that range, which is a continuous range, it's the entire interval, is equally likely. And this is what is called again, the uniform distribution. And just to develop intuition, we mentioned that uniform or standard distributions approximate empirical distributions. And why does it make sense or when does it make sense to use a uniform distribution?

It makes sense to use uniform distribution when we don't have very strong sense of what the specific value of the parameter is other than understanding some lower and upper bounds. So, when you are in that situation where you have an interval that you quite convinced that the value should be in that interval. But you're not very sure within that interval which value is more likely, uniform distribution is very intuitive to use.

Now let's just think about the way to model the other parameter that we have in the SIR model, which is the gamma parameter. And here we are going to use another commonly used standard distribution, that is called the triangular distribution. And like the uniform distribution, it also gets two parameters: The minimum value and the maximum values value. And it also can take values over a continuous interval that is defined by the minimum and the maximum values, parameters.

But unlike the uniform distribution that essentially within the defined range gave every possible value the same likelihood. The triangular distribution essentially gives the highest likelihood to the middle point. And then as you go to the right or to the left towards the upper and lower bounds. This the likelihood is going down right. And what is the intuition that this distribution is capturing.

Here, the intuition is, well, we are quite sure that the value has to be around 0.05. But we are still considering the fact that it could be either higher or lower. But as we go either far to the right or far to the left, we are now considering that is less likely to happen. And again, like the uniform distribution case, this can be an approximation of an empirical distribution that will be more discreet, and we'll have these rectangles that we've seen before capturing the respective histogram of that discrete distribution. So, what we're going to do next, we are going to think about the SIR model, but unlike this scenario that we had before where all everything was deterministic.

Now, the two parameters will be captured through probability distributions, uniform distribution for beta and triangle distribution for gamma. The mean of the two distributions will be equal to the deterministic values that we had before the 0.1 and the 0.05, respectively. And then what we're going to try and see how does this impact about the uncertainty of the output of the model. So, we're going to see the connection between the uncertainty that is going in and the uncertainty that is coming out of the model.

So specifically, we're going to discuss next, how do we analyze the results or the output of the SIR model when we have uncertainty in the parameters. And again, if I go back to the input, right, we are replacing the two deterministic values that we had before, for beta and gamma, with distributions. And the way we're going to use that, if before we simulated the outcome by taking these two specific values including the relationship in the computer and simulate the outcome for 200 days.

Now we're going to do something that is a little bit more involved. What we're going to do every time is going to draw value of beta and gamma from the respective distributions and then run and simulate 200 days for these specific values. And then we're going to repeat that many many times, and every time we're going to record the statistics of the run that we had. And we can run, since we have a computer that can do that, we can run quite many scenarios every time sampling.

That's called sampling or drawing values of beta and gamma from the distributions that we assigned to them and then run and and look at the output rate.

Video 2.4: Applying the SIR Model With Uncertainty: Part Two (12:44)

So, the next interesting question would be how do we analyze the output of such a model, right? And what you can see here, is one particular interesting statistic that we looked at before, which is the peak number of infections, right? And as you can see, this is not one number like we had before. This is not anymore 160, right? Like, we had a specific number for the specific values of beta and gamma, right?

And now you have a distribution on the value of the peak number of infections, right? So, it's now being captured through another distribution. This is not a very standard distribution, but it's an empirical distribution. Going back to this concept, that basically what you see here is the histogram is the likelihood for each value of peak infections. This is the X-axis mapped against the height of the rectangle on that value which represents the likelihood of that scenario to materialize, right? And that's kind of the Y-axis that you see that gives you the likelihood, right?

Now, looking at this, I hope it's now becoming clear to you that the answer whether we have sufficient capacity or not, right? With respect to the peak, that's kind of use case question number one, cannot be answered anymore by yes or no answer, right? Like it's we need a different answer or different way or different language to answer this question, right? And what we're going to try and do here is we're going to focus on, let's say, our capacity, which is 180, right? And what we would like to ask ourselves is a slightly different question.

Not if we do or do not have sufficient capacity, but rather than what is the probability that we are going to have sufficient capacity at the peak number of infections, right? Now, to do that, we need to take the 180 on the X-axis here, and we want to basically understand what is the cumulative weight of all the rectangles that stand left to it, right?

These are all the scenarios with their respective likelihoods in which the peak number of infections is going to be 180 or less, right? Now, what is going to be to the right of the 180 value, right? This is going to be all the scenarios in which we are not going to have sufficient capacity because we are not going to have enough capacity because the peak number of infections is going to be higher than our capacity value, which is 180, right? So, when we do that, and we sum all the rectangles to the left of 180, we see that the likelihood that this is going to happen is slightly over or over 53%. So, that means that 53% or 53.7% of the time, we are going to have sufficient capacity, but at the same time, we also know that 46% or 46.3% of the time, we are not going to have sufficient capacity, right?

And this is a far more involved answer because it's not a yes or no answer. And now the decision maker has to decide whether she is comfortable with this, right? Is that a risk? Right, is that uncertainty, right? There is an uncertainty in the output here that she's willing to take, or maybe she wants to do something about this, right? So, again, this is with respect to the peak number of infections, right?

Now, let's just take a look about something that is even more comprehensive than that, right? Like if I want to change now my capacity maybe, maybe I want to see what I can do to change the current capacity to 180; I can actually answer similarly to what we just answered for 180 for every possible value, right? This is going to be the graph that you see here. This is again the cumulative distribution of the peak number of infections, right? That basically plots for every level of on the X value, right?

What is the likelihood that you were going to have sufficient capacity?

Namely, what is the cumulative heights of all the rectangles that stand left to that value? This is being going to be captured by the Y-axis on the right-hand side here, right? So, it basically translates the histogram that we see here into basically answering for every level of capacity you can consider what is the likelihood that you go that is going to be sufficient for you. And again, this is something that the computer can create for you. And again, we're going to come back to how to use this type of analysis to answer different questions, but later in this module, but also in the assignments that you will have to think about after this module.

Let's just look on something a little bit more involved. What we see here is the, what we call the joint distribution of the two things that we were interested in, in the prior example, right? On the X-axis, we are going to look at on the days until the peak of infections arrive, right? On the Y-axis, we're going to look on what is the peak itself. What is the peak number of infections that we're going to have, right? And each point here, each point that you see on the graph will be a scenario, right? That corresponds to the X value, will correspond to the, under this scenario, how quickly the peak will arrive, and the Y value of that point will correspond to what will be the peak itself under this scenario, right?

So, we now look on the output statistics, these two statistics together. And this is again called the joint distribution of these two random quantities. And we're going to get back to this type of drawing, we're going to talk about correlation in one of the next modules, right? But for the sake of the discussion today, let's just see how can we use this kind of visualization to further understand what is the uncertainty in the output and what does it tell us about our decisions, right? Now the two dark lines that we see here, this is the average number of infections, and the average time until you will arrive to the peak across all the scenarios. And now that divides the space of all of these points into four quadrants, right? And what I would like you to ask yourself is, what are the worst scenarios for you? Now, if you think about that, the very worst scenarios for you are the scenarios in which the peak is high and arrives relatively quickly, right?

Because that will stress your system the most and also will make your response time much more problematic if it's going to take you more time to build more capacity, right? And that is being captured by the all the red points that are on the left upper quadrant, right? These are the scenarios where basically the all the cases where you're going to have potentially over 600 infections, right? In many of these cases, you're going to have more than 600 infections that can happen maybe in less than 60 days, right? So, there are many scenarios in which you're going to be really in trouble, right? And when I look on the total weight of all of these points, right? It's more than 25% of the time, you're going to be in that bad regime, right?

Now, what is the most favorable set of scenarios for you, right? The most favorable set of scenarios for you is actually the scenarios in which the peak is low, and it takes a lot of time for the peak to arrive. And this is exactly the diagonal to the upper left. This is the right lower quadrants in which you see all of these points where basically, you can have the peak less than 100, right? And it might occur in more than 80 days. That's kind of a lot of these scenarios fall in that category, right? And that's going to happen with 33%, about 30% chance when you actually sum all of these scenarios in that regime, right? And then of course there are the other regimes that are kind of in between. But again, what I hoped you see that now the answers are far more complex, right? And also, it requires a judgment call from the decision maker. What is the risk they are willing to take?

Okay, so we're going to continue to talk about that, and we're going to continue to talk about how to interpret it and what kind of additional question you can ask about this picture or pictures like this. But one thing that I would like to emphasize before we move on is the blue point that essentially shows you again the output that we've seen before of the baseline scenario, right? And what that output represents? It represents the output of the average values of beta and gamma, right? Now the 0.1 and 0.05 are the average values of beta and gamma. And what you can see here is the output of that particular scenarios in this blue point.

And I want you to understand or kind of pay attention to the fact that the output of the average value here does not capture the average output. And this is a very important insight for you to remember because often, what we tend to think about that if we plug into our model the average values of the input parameters, we're going to get the average performance or the average output. And that's almost always not true, okay? So, remember that as a takeaway. So, in the examples that we discussed, we used a simple model that has a lot of assumptions that could limit its practical, descriptive, and predictive power, right?

So, as much as you might be impressed by what the model can give you, I want you to remember that this is a relatively simple model, right? And there are more complex versions of these models that you can read about in the supplementary reading material. So, I want you to understand that the real life, when you actually want to model very real and practical decisions, you need to have much more complex models than this, right? But I do you want to highlight, before we conclude this discussion, several very important takeaways, right?

So, the first takeaway is that regardless of anything when you use a model, you need to acknowledge the fact that it provides only an approximation of the real world and thus you need to use it with great level of care and judgment calls, right? The second takeaway is that, speaking of that, you have to understand what uncertainty exists with respect to the model's parameters and input, right? This is key to understand, especially with respect to the risk and the impact on the related outcomes of the model.

And here distributions, probability distributions can be used very effectively to model the input and the parameters' uncertainty, as well as the induced uncertainty that they imply on the outcomes of the model. This is really a key concept. And then another takeaway is that taking point estimates, right? These are taking the parameters' means, right? The input parameters' means does not yield the average model's outcomes. This is very important to understand because that's a common mistake that people make in the field.

And finally, the last takeaway is thinking about uncertainty requires very different questions. It's not a yes, no question. It's more like what is the likelihood of something to happen, and respectively, right? The answers are also very different, and they're also probabilistic answers. So, we are going to move now to ask more probabilistic questions and have more probabilistic answers than more deterministic questions versus deterministic answers.

Video 2.5: Retail Inventory Management (14:38)

Next, we're going to discuss another example that will help us illustrate how do you use probability distributions to inform decisions under uncertainty. And the next example is going to be about inventory management, which is a very important area in which people make decisions under uncertainty, essentially making decisions how much stock to build up of different products. And again, the use case is going to be simple, but I want to assure you that this is an area in the real world in which people on a regular basis have to deal with uncertainty and they use often models including probability distributions in a way that is very instrumental for them to make decisions.

So, the simple example that we are going to consider is talks about a retailer that plans to use its warehouse in the vicinity of the Boston area to supply its store in Newbury Street, anticipating a new iPhone model that is going to be launched during the holiday season. And the retailer guaranteed the supply of 1000 phones but he's still worried that that might not be enough to satisfy the expected or anticipated demand. And frankly speaking, this is a new product.

So, it's not unlikely that the demand in the store is not only uncertain, but there's not too much historical information about what that demand might look like. So, while we think about it is a random variable, we might not have a lot of data to model that. So, maybe the retailer is only able to estimate what the mean demand is going to be, and let's say that that's 800 units. And maybe in addition to that, the retailer is able to estimate the standard deviation of the demand, and let's say that that's 250.

So, this is not an typical situation where you have relatively little data about the past demand or very little understanding of how the past demand reflects on the demand of a new product. And you maybe only are able to estimate some basic statistics of the demand distribution, the uncertain and demand distribution. And again, I want to remind you that we're going to use the concept or the term random variable to capture random quantities. So, we're going to think about the demand for the iPhone as a random variable.

So, again, the retailer is wondering whether 1,000 units of supply is going to be sufficient inventory in the face of this uncertain demand, this random variable. And let's just think about how can you model this decision. And I would like to highlight the fact that the first question you have to ask yourself is, what is the precise metrics that the retailer should consider in determining whether 1,000 units are sufficient or not? So, the notion of what sufficient is is not obvious here. And in fact, it takes a choice of the retailer about the business metrics she or he would like to consider in order to inform what model we would like to build.

So, that should remind you that analytics has a lot of technical aspects; but at the end of the day, it has to interact with managerial choices. So, one very common metric that you can choose in this specific setting is the notion of the fill rate. The fill rate is essentially the likelihood, the probability that you're going to be able to satisfy all demand that you're going to have. So, it's essentially saying, what is the likelihood that I'm going to be never stocked out or will not stock out?

That's the fill rate. So, that maybe is a reasonable metric in this particular setting. The next question is how exactly should the retailer think about the uncertain demand, how that should be modeled? And here you have to ask yourself different questions. So, what data is available, what kind of other expert inputs you have, how tractable you want the model that you create to be, and how does it capture the business metrics you're interested in.

And I'm going to assume in this particular setting that we're going to model, it's an assumption, we're going to model the demand distribution. We're going to model this random variable as another standard distribution that is very, very important; that you read about, hopefully, in the supporting material. And that is the normal distribution that is specified by two parameters: The mean that in this case we're going to estimate to be 800; and the standard deviation; which is 250. And again, this is a very, very important distribution that the high level captures many physical phenomena that we see in real life.

Now, I'm going to ask specifically two questions, or the retailer is going to ask specifically two questions in this context.

The first question we already mentioned, which is what is the fill rate of an inventory stocking level of 1,000 units? But additionally, the retailer might be also interested in the following question, which is what is the inventory level that I need to build in order to guarantee that 95% of the time I'm going to be able to satisfy all my customers? So, that's essentially asking, if I want to use the terminology of the field rate, what is the stocking level, what is the inventory level that will guarantee, will secure a 95% fill rate? Now, before we move on, I want to go back, step back and highlight some of the steps and the considerations that we took in the discussion that we had so far.

So, first observe the fact that, again, we use data to create models. And we had to match the business metrics and the choices that the decision makers are making to the data through models. And we use the data to estimate the model's parameter in this case, this is the demand model. The normal distribution, we estimated the mean and the standard deviation which were the two input parameters into this specific model.

Now we are going to use the normal distribution as a descriptive model in this case to describe the uncertainty of the demand. And now we will show how to use that sub model, this mini model, to create a prescriptive model that can inform how much inventory the retailer should hold. But that said, I would like to again highlight to you that no model is perfect. And while there are many reasons and advantages to use the normal distribution, for example, you only need to estimate two parameters in order to have a complete model: the mean and the standard deviation.

And in many cases, it does capture the dynamics of the demand that is consist of many independent decisions that different customers are making whether to purchase or not, that model is not perfect. For example, if you read about the normal distribution, theoretically speaking, it can take negative values. And that doesn't make a lot of sense in the context of a demand for iPhones. So, this model is not perfect. But again, it is a very typical model that we are using in practice, and it does give you, in most cases, very good answers, but it's not perfect. I would like you to remember that.

And again, before we move on, I would like to highlight the fact that probability distributions allow the managers to take a measured analytics approach to risk and it helps them to answer questions in both directions.

The first question is, given a policy, given the decision to hold 1,000 units of inventory in your warehouse, what is the performance or what is the likely performance of that policy is going to be? The model is helping you to assess that. At the same time, the model allows you to ask a different question, which is given a desired business performance, specifically fill rate of 95%, what is the necessary stocking level or inventory level that you need to hold? So, it allows you to go in two directions here.

And that's what makes these kinds of models very, very useful for managers because it allows them to ask questions of different types and get reasonable answers that inform their decision making. So, let's see how we can answer the two questions that we post. And what you see here is the probability distribution of the normal distribution.

So, unlike the uniform distribution, that's another standard distribution that we've seen before, that has this flat curve. The normal distribution has a very typical bell-shaped curve that is symmetric around the mean of the distribution, and it's either skinny and tall or fat and short. And that is being determined by the second parameter, which is the standard deviation. The higher the standard deviation, the more flat, short, and fat the curve is. And the smaller the standard deviation, the distribution would be far more concentrate around the mean. And in all likelihood, by the way, you're not likely to see too many values when you consider normal distribution more than three standard deviations, left and right to the mean. So, that's some properties that, hopefully, you read more about in the supporting material preparing for this module.

Now, how are we going to answer the question given what we see in front of us? So, what was the first question? The first question was, what is the fill rate that is induced by having 1,000 units of inventory? And to some extent, that's similar to asking in the SIR model, is capacity of 180 going to be sufficient in face of the number of infected individuals?

In this case, what we would like to see is how much area there is left to the value 1,000 here on the X value. When we try to answer the question with respect to the SIR model, instead of calculating the area, we have to calculate the sum of rectangles because that was a discrete distribution. The normal distribution is a continuous distribution. So instead, we need to calculate the area under the curve, lying left to the value 1,000.

Now, don't be too worried about that, you don't need to do that yourself. You can actually have this being calculated for you in one command in Python as you can see here. And when you plug it in, you plug the value 1,000, the parameters' values 800 and 200, you get the answer which is 0.79. Which means that if you have in stock 1,000 units in 79% of the time, you're going to have enough capacity; and in 21% of the time, you're not going to have capacity. So, that gives you an answer to the first question. Now, what is the second question about?

The second question is about what is the value of inventory that you need to hold in order to secure 95% chance of fill rate? To be able to, in 95% of the time, to satisfy all the demand that you're going to face. And in order to do that, you're going to now have to take the 1,000 value and try to start pushing it to the right until the moment in time when the area under the curve will be 95% or 0.95. Again, you don't need to do that yourself, you can actually ask the computer to do that for you. And you can specify the desired level of field rate 8.95, as you can see here in the command. And again, the parameters of the distribution 800 and 250, in this case, and what you're going to get back from the computer is the value. In this case, the value is 1,212, which is the value for which, if I will calculate the area left to that value under the curve, it's going to be 0.95.

Now, what we see here is that having the intuition about what is going on and using the computer's ability to code these standard distributions like the normal distribution, you can easily answer many, many types of questions. Now, before we conclude this example and go to a more complicated version of this example and discuss that, I would like you to note that when I look on the 1,212 as the answer to the second question, and I compare it to the mean demand; which is 800, the difference between them is equal to 1.64 times the standard deviation; which is, in this case, 250.

So, one easy way or one handy way to think about fill rate is by how much you are ordering over the mean demand in order to secure the fill rate, and that's often being measured by the number of standard deviations that you need to add on top of the mean in order to secure the fill rate. More to come on this in a more complex example that we're going to discuss next.

Video 2.6: Module Wrap-Up (3:54)

So, let me try and summarize some of the important takeaways of this module. We talked about probability distributions as a key modeling tool to capture uncertainty in business environments. Specifically, we showed how distributions can be used to quantify risk of the current state of the system, namely, given a current policy, what is likely to be the performance of the policy, and at the same time, we also showed how it can be used to determine what policy is needed to achieve a certain risk tolerance or a certain performance target. We saw so in two examples, right?

Both in the inventory example, the retailer example, and the appointment scheduling example, that sometimes you need to tradeoff between different risk levels and make a choice. So, for example, in the appointment scheduling example, you have to tradeoff between having high utilization of your clinic and not having too much overtime. So, these are tradeoffs that the model allows, models allow decision makers to make but they don't wave them from having to make choices about what the performance metrics are and what is their appetite with respect to the performance and the tradeoff that can be done, that can be achieved.

So, another important takeaway from this module is to understand and be able to use the phenomena of risk pooling. Risk pooling is a very, very important physical phenomena that you can use to your advantage in many situations when you have to address uncertainty in business environments.

What is it saying? It's saying that when you sum multiple random variables, when you are able to pool multiple random quantities, what you get is that the relative risk is decreasing. What do we mean by that? The relative risk is allowing you to essentially manage uncertainty better, and that's a very powerful strategy to mitigate risk more effectively. The quantity or the notion that captures the effect of risk pooling is the coefficient of variation that is defined as the standard deviation over the mean.

And what we've seen is that when we do risk pooling, when we sum random variables together, the mean is scaled as linearly in the number of random variables that we sum, but the standard deviation is scaling only as a square root. And that implies that the relative risk captured by the coefficient of variation is approaching zero.

And finally, maybe the most important takeaway from this module is the fact that it's very, very risky and can lead to very bad decisions to think about uncertainty in deterministic terms. That includes thinking about the uncertainty about the parameters of the model as the average of these parameters, and just make decisions based on that. But also it's very risky to think about the average performance of the system and make decisions based on that.

So, what you need to do is to transform to be able to ask yourself probabilistic questions and give probabilistic answers that will allow you to understand the implications of uncertainty in business environments in a much more effective and the right way. But that, at the same time, requires managers to make choices and to make decisions about what is their risk appetite and how they want to accept or not accept given levels of risk.