PROFESSOR: In this segment, we will discuss a little bit the union bound and then discuss a counterpart, which is known as the Bonferroni inequality. Let us start with a story.

Suppose that we have a number of students in some class. And we have a set of students that are smart, let's call that set a1. So this is the set of smart students. And we have a set of students that are beautiful. And let's call that set a2. So a2 is the set of beautiful students. If I tell you that the set of smart students is small, and the set of beautiful students are small, then you can probably conclude that there are very few students that are either smart or beautiful.

What does this have to do with probability? Well, when we say very few are smart, we might mean that if I pick a student at random, there is only a small probability that I pick a smart student, and similarly for beautiful students. Can we make this statement more precise? Indeed we can. We have the union bound that tells us that the probability that I pick a student that is either smart or beautiful is less than or equal to the probability of being a smart student plus the probability of picking a beautiful student. So, if this probability is small and that probability is 1, then this probability will also be small, which means that there is only a small number of students that are either smart or beautiful. Now let us try to turn the statement around its head.

Suppose that most of the students are smart and most of the students are beautiful. So in this case, I'm telling you that these sets a1 and a2 are big. Now, if the set a1 is big, then it means that this set here, the complement of a1, is a small set. And if I tell you that the set a2 is big, then it means that this set here, which is a complement of a2, is also small. So everything outside here is a small set, which means that whatever is left, which is the intersection of a1 and a2, should be a big set. So we should be able to conclude that, in this case, most of the students belong to the intersection. So they're both smart and beautiful.

How can we turn this into a mathematical statement? It's the following inequality that we will prove shortly. But what it says is that the probability of the intersection is larger than or equal to something. And if this probability is close to 1, which says that most of the students are smart, and this probability is close to 1, which says that more students are beautiful, then this difference here is going to be close to 1 plus 1 minus 1, which is 1. Therefore, the probability of the intersection is going to be larger than or equal to some number that's close to 1. So this one will also be close to 1, which is the conclusion that indeed most students fall in this intersection and they're both smart and beautiful.

So what we will do next will be to derive this inequality and actually generalize it. So here is the relation that we wish to establish. We want to show that the probability of a certain event is bigger than something. How do we show that? One way is to show that the probability of the complement of this event, namely this event here, we want to show that this event has small probability. Now what is this event? Here we can use DeMorgan's laws, which tell us that this event is the same as this one. That is, the complement of an intersection is the union of the complements. Since these two sets or events are identical, it means that their probabilities will also be equal.

And next we will use the union bound to write this probability as being less than or equal to the sum of the probabilities of the two events whose union we are taking. Now we're getting close, except that here we have complements all over, whereas up here we do not have any complements. What can we do? Well, the probability of a complement of an event is the same as 1 minus the probability of that event. And we do the same thing for the terms that we have here. This probability here is equal to 1 minus the probability of a1. And this probability here is equal to 1 minus the probability of a2. And now if we take this inequality, cancel this term with that term, and then moved terms around, what we have is exactly this relation that we wanted to prove.

It turns out that this inequality has a generalization to the case where we take the intersection of n events. And this has, again, the same intuitive content. Suppose that each one of these events a1 up to a(n) is almost certain to occur. That is, it has a probability close to 1. In that case, this term will be close to n. We subtract n minus 1, so this term on the right hand side will be close to 1. Therefore, the probability of the intersection will be larger than or equal to something that's close to 1. So this is big. Essentially what it's saying is that we have big sets and we take their intersection, then that intersection will also be big in terms of having large probability.

How do we prove this relation? Exactly the same way as it was proved for the case of two sets. Namely, instead of looking at this event, we look at the complement of this event. And we use DeMorgan's laws to write this complement as the union of the complements. These two are the same sets or events, so they have the same probability. And then we use the union bound to write this as being less than or equal to the probabilities of all of those sets.

Now this is equal to 1 minus the probability of the intersection. This side here is equal to 1 minus the probability of a1. This is one term. We get n such terms, the last one being 1 minus the probability of a(n). And we still have an inequality going this way. We collect those 1s that we have here. There's n of them, and one here, so we're left with n minus 1 terms that are equal to 1. And this gives rise to this term. We have all the probabilities of the various events that appear with the same sign. This gives rise to this term. And finally, this term here will correspond to that term.

Namely, if we started with this inequality and just rearrange a few terms, we obtain this inequality up here. So these Bonferroni inequalities are a nice illustration of how one can combine DeMorgan's laws, set theoretical operations, and the union bound in order to obtain some interesting relations between probabilities.