[SQUEAKING]

[RUSTLING]

[CLICKING]

YUFEI ZHAO: Let us look at an application of the probabilistic method to extremal set theory. The extremal set theory concerns the study of families of sets that have certain desirable properties, and asking questions such as what is the largest set family that you can have with such families-- with such properties? And in this video, we'll focus on the following property of being an intersecting family. This means a collection of sets, A1 through Al, such that these sets pairwise intersect in a nonempty intersection.

So we want sets A1 through Al, such that whenever you take any two of these sets and take their intersection, it is never the empty set. So here are two basic questions that one can ask. The first one, which will turn out to be an easy warm up, is what is the largest intersecting family of subsets of 1 through n.

So here, there are, in total, 2 to the n subsets of 1 through n. But if I want to only keep a collection of sets that's an intersecting family, what is the largest family I can get? And the second question, which will turn out to be far more interesting, will involve an application of the probabilistic method, is a question, what is the largest intersecting family of K element subsets of 1 through n?

So here, n and K are inputs to the question. So they are given. And we're asked to find what is the largest intersecting family of K element subsets of 1 through n. All right.

So the first question, as I mentioned earlier, turns out to be a fairly easy one, and we'll solve it as a warm. So as an example, just to illustrate what we're talking about, if we look at all sets containing the element 1. Well, this collection of sets is intersecting because they all have the element 1. So every pair intersects in some set that contains the element 1 in particular non-empty.

This collection of sets has size 2 to the n minus 1, because once you fix the element 1, the other elements can be in or not in the sets, and there are 2 to the n minus 1 such possibilities. So a binary choice for each of the elements outside 1.

OK. Now this is just an example. But it turns out this is the best that we can do, because-- so the claim's that we cannot do better, or we cannot get an even larger set. And this is because for every subset A of 1 through n, and most 1 of A and the complement of A, so the elements 1 through n subtract A, so the complement of A. And most one of these can be in the intersecting family.

So what we're doing here is pairing up the sets with their complements. And because it's an intersecting family, you cannot simultaneously have a set and its complement in this intersecting family. And so this eliminates-- so this means that you can have at most half of all the 2 to the n sets, thereby showing that this 2 to the n minus 1 is indeed the best that you can do.

OK. So this finishes part one, question one. And now let's move on to the more interesting and harder question two. Again, let's start with some examples.

First, here's an easy case. If n is less than 2K, then-- so this is the easy setting. If n is less than 2K then by pigeonhole principle, any pair of K element subsets of 1 through n intersect. So we can take all and choose K sets, and they form an intersecting family.

So there's nothing, really, to do here, because we can just take all the K element sets. So the harder part of the question, the more interesting part, is what happens when n is at least 2K? Here, as an example, we can take as earlier all sets containing the element 1.

And this gives us n minus 1 choose K minus 1 sets. And that's before the sets, they all-- this set form-- this collection of sets form an intersecting family because any pair their intersection contains the element 1, and is thus nonempty.

So this is the construction and gives you n minus 1 choose K minus 1 sets. But is this the best you can do? Perhaps through some other methods of construction, through other examples, maybe there are even larger intersecting families of K element subsets.

That turns out not to be the case. And that is the main theorem that we'll prove in this video. And this result is known as the Erdos-Ko-Rado theorem. So this is a seminal and beautiful result in extremal set theory.

And The Erdos-Ko-Rado theorem says that if n is at least 2K, then any intersecting family, and we'll give this family a name, Curly F, of subsets of 1 through n. OK, subsets and any intersecting family of K element subsets of one through n has size at most n minus 1 choose k minus 1.

In other words, the construction that we gave just now is indeed optimal. You cannot get a bigger intersecting family of k element subsets of 1 through n. Let us now prove the Erdos-Ko-Rado theorem. And we'll introduce randomness into the problem and use some beautiful ideas from the probabilistic method.

So here's how we start the proof. Let us order the numbers 1, 2, 3, and so on through n randomly around a circle. So what do I mean by this?

Take a circle and then basically take a uniform random permutation of 1 through n and then place the numbers around the circle in the order of that permutation. So for example, I'm putting down nine numbers, here, n equals to nine, in some circular order chosen uniformly at random.

I'm missing a number here. Six. So here, nine numbers. Uniformly around the circle. And let us call a subset of 1 through n contiguous. Here, I'm defining the term "contiguous."

I'll call the subset contiguous if its elements, the way they're ordered around the circle, form an arc in this ordering. So for example, the set of numbers 4, 3, 9, 5, this four-element set is contiguous. So 3, 4, 5, 9 is contiguous.

Because the order's there-- according to the circular ordering, these four numbers form a contiguous block. On the other hand, the set 1, 3 is not contiguous because they do not form a contiguous block. So that's just a definition.

And now for a given set, sub A, a subset of 1 through n with k elements. So for a given k element subset of 1 through n, what is the probability? So here, A is fixed. A is not random. So we're given this A. What is the probability that A is contiguous?

So the ordering is random. So given a set, what is the probability that under this random ordering is contiguous? Well, let's think about that. So the ordering is random, but here, if A has k elements, you can think about how many different ways these k elements can be placed around the circle.

There are n different positions for where geometrically, this length k arc lies. And once you fix that position, the probability that the elements, the k elements of A actually falls into the desired positions is 1 over k-- 1 over n choose k.

So this is the probability that A is a contiguous set under this random ordering. And now by the linearity of expectations, the expected number of contiguous sets in this collection F is equal to the size of F times the probability that each individual set is contiguous. And that was calculated earlier to be this quantity. So n divided by n choose k.

On the other hand, the property of F being intersecting, I claim, implies that in any given circular ordering there are at most k contiguous sets for-- so in a given circular ordering. So let's pause and think about why this claim is true.

If I give you an ordering and ask, well, how many contiguous sets can there be? And now these contiguous sets, they have to be intersecting. So if we have one set-- so here, let me denote the elements, these circles as-- think about them as positions on the circle. So these dots as positions on the circle.

So if that's one set, well, where can the other sets be? Well, they have to intersect this block. So maybe another set is like this one. But if that set is in, then it excludes-- if this blue set is in,l then it excludes this green set from being a possibility and so on.

OK. So you see that given that this red set is in, well, at most, one of the blue and green on the second line can be in. And at most, one of the blue and green on the third line can be in. And at most, one of blue and green on the fourth line can be in. And these are the all the possibilities of how another contiguous set can intersect the red set.

So that's an argument that can translate into this claim, that F being intersecting implies at most k contiguous sets in any given order. So this argument is very analogous to the argument that we did at the beginning, where we pair up complementary sets. But this time, we're only looking at contiguous sets that intersect a given set.

OK. Well, if you can only get at most k contiguous sets in any given order, then this quantity, which is a random variable, it is always at most k. So in expectation, then, this quantity that we got at the end is also at most k.

Thus the size of F is at most k over n times n choose k, which one can then expand the binomial coefficient to see that the final quantity is equal to n minus 1 choose k minus 1. And that's exactly what we claimed.

All right. So this finishes the proof of the Erdos-Ko-Rado theorem, which is a foundational result in extremal set theory. And it's a beautiful application of the probabilistic method to extremal set theory and to combinatorics in general, where we start with a claim, a theorem, that is not random at all. It is completely deterministic.

It's about finding the maximum possible set size, the size of something that satisfies some constraints. And yet the proof goes by introducing new randomness into the setup. That allows us, then, to derive this beautiful conclusion.