[SQUEAKING]

[RUSTLING]

[CLICKING]

PROFESSOR: In this video, we'll look at a powerful tool in the probabilistic method known as a bounded differences inequality, which also goes by the name Azuma-Hoeffding inequality.

This inequality tells us that if we have a function of independent random variables and such that this function doesn't change very much if we just change one of the inputs of the function, then the output of this function, given the random input, is fairly concentrated around its mean.

Let's look at the actual statement of this theorem. We have x1, a random variable taking values in the set or in the probability space omega 1, and so on. So x1, x2, to xn. So these are independent random variables. And it's very important that they're independent.

We have a function, f, which takes as input n coordinates and outputs a real number. And the hypothesis of this theorem is that this function f satisfies the following property-- that if we look at two different outputs of f on inputs x and y such that x and y differ on exactly one coordinate. So by only changing a single coordinate of f, the assumption is that the value of f changes by no more than one.

So f in some sense is fairly smooth to the fluctuations in the input. If you change only one input coordinate, then the output of f does not change too much, does not change by more than one.

OK, what is the conclusion of the theorem? It says that the random variable obtained by evaluating f on these independent random input coordinates-- that's what we call this output, z-- and this random variable z satisfies the following concentration inequality. For every non-negative real lambda, the probability that z exceeds its expectation by at least lambda is at most this quantity here, which goes down rapidly as lambda gets large.

And also, we also have a lower tail concentration bound, which says that the probability that z is significantly below its expectation by more than lambda, at least lambda, is this probability is upper bounded by this same quantity which, again, decays extremely quickly when lambda is large. So this is the statement of the bounded differences inequality.

Again, the intuition here is that given independent inputs to a function which satisfies this property, that it does not change by more than one, upon changing any single coordinate, the output random variable is very concentrated around its mean.

In the rest of this video, I want to present three applications of this inequality. The first application is meant to illustrate a very simple example of a function where we can apply the theorem. And this function is simply the function taking Boolean input, so n Boolean inputs, and outputs a real number obtained by simply adding up the input numbers.

In other words, this is a sum of n different coin tosses, each coin toss resulting in a zero or a one. And you can check that this function has the property that we require in this theorem, namely that if you flip just one of the input coordinates, the output changes by-- well, in this case, exactly one, but certainly no more than one.

Now we can apply the Azuma-Hoeffding inequality or the bounded differences inequality to this function, and that gets us a tail bound on the binomial distribution. This case, you know that the expectation of z is exactly n over 2, and it tells you that the deviation cannot exceed very much beyond something on the order of the square root of n.

This bound is also known as Chernoff bound. And in fact, the proof of the bounded differences inequality is very similar to the proof of the Chernoff bound, which you can view in a different video. Let me now go on to the next example.

The next example concerns a problem called the coupon collector problem. The setup is that we have independent random numbers s1 through sn, and they're each chosen uniformly from the numbers 1 through n. So these numbers are uniformly and independently chosen.

So you can imagine a setup as having n different coupons. And each time you draw a random coupon, see what it is, return it to the box, and draw it again. And these s1 through sn's is a list of coupons that you draw from this box.

And the random variable that we're interested in, z, is the number of missing coupons, the number of coupons that you have not seen through this process.

In other words, this is the number of elements of 1 through n that are not among the elements s1 through sn. So this is a number of missing coupons. It is a random variable, because s1 through sn are random.

We wish to understand how z is concentrated around its mean. And for that, we can apply the bounded differences inequality. Note that this function, given viewing this quantity as a function from s1 through sn as inputs to the output number, this function satisfies the required hypothesis. Namely, if you change one of the si's, you do not change the number of missing coupons by more than one.

Number of coupons, missing coupons might not change at all. But it cannot change by more than one.

So this means that we can apply the bounded differences inequality to deduce the conclusion that the probability that z deviates from its expectation by more than lambda is at most 2, because here we're using upper and lower bounds simultaneously. Times 2 to the minus e 2 lambda squared over n.

And the expectation is something that we can calculate pretty easily using linearity of expectations of these n different numbers, these n different coupons. Each single coupon is missing with probability 1 minus 1 over n to the n, because this is the probability that a specific coupon, coupon i, is missing, which is the event that coupon i is not drawn in each of the n different random draws. And this quantity is very close to n over e.

OK. So this is an application of the bounded differences inequality to a function which is not linear, like before. So the first example is a much simpler example because the function f is simply a sum of its inputs.

And here, there's a more complicated function. Our last example involves an even subtler function. And here, the result is a classic theorem in probabilistic combinatorics due to Shamir and Spencer from the '80s.

The theorem concerns the chromatic number of a random graph. So let z be the chromatic number of the random graph, G n, p OK, so G n, p is the Erdos-Rényi random graph obtained by taking n vertices and putting an edge between every pair of vertices with probability p. So throw a probability p coin for each possible edge independently and construct a random graph this way.

And then z is the chromatic number of this random graph. So it is the minimum number of colors required to color all the vertices so that no two adjacent vertices receive the same color.

This is some random variable. And this random variable is pretty complicated. It's pretty hard to analyze. But nevertheless, using the bounded differences inequality, let us deduce the following concentration bound, showing that z typically is not too far away from its expectation. And specifically, we have the bound saying that z deviates from expectation by more than this quantity here. And this event has probability at most 2 times e to the minus 2 lambda squared.

OK, so let's prove this theorem.

The interesting part of this proof is how to set up the function, f, so that we can apply the bounded differences inequality. So right now, z is some quantity which seems kind of complicated, and it's based on something that is random. So how can we phrase it in terms of independent random variables in a way so that we can apply the bounded differences inequality?

Well, one way to do it, and it's a natural first attempt, is to view z as a function with n, choose two inputs, one input for each possible edge of the random graph.

That is a valid choice in the sense that it satisfies the bounded differences condition, but it will not give the correct bound because it will turn out to have way too many variables in this function. We'll take a look at a different method that will be able to provide us with the desired bound. And there is some neat idea here on how to cluster the random variables together.

So we will represent our graph on n vertices labeled 1 through n. And this graph, which has some edges and we'll need to represent these edges, we'll represent them as an element of the following product set-- omega 1 times omega 2-- so these are Cartesian products-- times omega 3. So 1, 2, omega n minus one.

And how we're going to encode the graph using the element of this product set is as follows-- omega 1 will be the set 0 comma 1, and this set will record-- either 0 or 1, which one gets chosen-- will record whether there is an edge between the first vertex and the second vertex.

So if we choose 0, then there's no edge between 1 and 2. If we choose 1, then there is an edge between 1 and 2.

Omega 2 is the product set of 0, 1 with itself. And the two bits encodes whether there is an edge between 1 and 3, and also whether there is an edge between 2 and 3. That information is encoded in the element of omega 2 that we choose.

And omega 3 is now 0 comma 1, raise to the third power. And here, these three bits record whether the three edges from 1, 2, and 3, to the fourth vertex are included or not included in this graph and so on.

OK, so every graph on n-labeled vertices can be represented as an element of this product set and vice versa. Basically, we are clustering the edges together according to the right end point of its edge according to this vertex order of the vertices.

So why is this useful? If we have two graphs, g and g prime, that differ on edges around only one vertex-- so if g and g prime differ on edges around only one vertex, then the chromatic number, or chi of g minus chi of g prime, this, in absolute value, is at most one.

So if you have two graphs and one can be obtained from the other by modifying edges around a single vertex, then their chromatic numbers cannot differ by more than one. And the reason is that we could have just chosen a new color for the vertex that is involved, so we would not need to use more than one color change to go from one graph to another.

OK, so this is a important fact that turns out to be incredibly useful. And it's useful because if we have a graph represented as an element of this product set, then having another graph that's obtained by just changing one coordinate results in a graph where the only changes to the graph itself are edges around a single vertex.

And therefore, the function, f, that sends an element of this to an element of the reals-- basically, the chromatic number of the graph-- satisfies the bounded differences hypothesis. And by applying the theorem, we can then deduce the concentration bound here.

So that finishes the proof of this theorem on the concentration of the chromatic number of a random graph.

All right. So in this video, we saw the statement of the bounded differences inequality, which is a important and versatile tool that is used all over probabilistic combinatorics. It says that the concentration of a function of independent inputs-- or this, as long as the function has the property that it doesn't change very much if you only change a single coordinate, then the output as a random variable is highly concentrated around its mean. And we saw three different examples applying this bounded differences inequality.