

Let  $X$  be a random variable, and let  $g$  be a function.

We know that if  $g$  is linear, then the expected value of the function is the same as that linear function of the expected value.

On the other hand, we know that when  $g$  is nonlinear, in general, these two quantities will not be related.

But there is a special case in which we can establish some relation between these two quantities in the form of an inequality.

This is Jensen's inequality, which we're going to develop.

Jensen's inequality applies to the special case where  $g$  is a convex function.

I'm going to define more precisely what it means to be convex shortly.

But in terms of a picture, it's a function that has a shape of this kind.

So it tends to curve upwards.

So let us look at a simple example.

Suppose that  $X$  is a random variable that can take two values with equal probability.

So these two values have both probability  $1/2$ .

And this is our function,  $g$  of  $x$ .

Since they take the two values with equal probability, the expected value is going to be in the middle.

So this here is the expected value of  $X$ .

And in particular, this value here is going to be  $g$  of the expected value of  $X$ . Now, the random variable  $g$  of  $X$  will take this value with probability  $1/2$ , and it's going to take that value with probability  $1/2$ .

What is the average of  $g$  of  $X$ , the expected value of  $g$  of  $X$ ?

It's going to be  $1/2$  of this value plus  $1/2$  of this value, which you can find as follows.

If you draw a straight line, it's going to be this much.

So this quantity here is the expected value of  $g$  of  $X$ .

And we see that in this example, the expected value of  $g$  of  $X$  is above the value of  $g$  evaluated at the expected value of  $X$ .

So this is what Jensen's inequality says, but for a more general distribution for the random variable  $X$ . Let us now step back and define more precisely what it means for a function to be convex.

The most general definition is the following.

If I take any two points,  $x$  and  $y$ , and I take some number  $p$  between 0 and 1.

So in that case, the number  $px$  plus  $1$  minus  $py$  is a weighted average of  $x$  and  $y$ -- so it's somewhere in the middle.

And if I look at the value of my function at that particular point-- so this value here corresponds to this-- this is less than or equal to the weighted average of the values of  $g$  of  $x$  and  $g$  of  $y$ .

So this is  $g$  of  $x$ .

This is  $g$  of  $y$ .

This value here is the weighted average of the two values.

So this quantity here is  $p$  times  $g$  of  $x$ , which is this value, plus  $1$  minus  $p$ ,  $g$  of  $y$ .

Convexity means that this value is below that value.

Or in other words, whenever I take two points on this curve and join them by a segment, then the function lies underneath that segment.

This is one possible definition.

Now, in terms of a picture, we see that a convex function tends to curve upwards.

This means that the derivative or the slope of the function keeps increasing.

An increasing slope means that the second derivative of that function is non-negative.

And that could be an alternative definition of convexity.

It turns out that if you have a function that's twice differentiable, these two definitions are equivalent.

On the other hand, the first definition is a little more general, because it also applies to functions that are not smooth.

So for example, the function absolute value of  $x$  is a convex one.

But it's not differentiable at zero.

Finally, there's another way of defining convexity, and it is the following property, again for differentiable functions.

What it says is the following.

If I fix a certain point,  $c$ -- to use the same diagram let's say that this is  $c$ -- this value here is going to be  $g$  of  $c$ .

I look at the derivative of that function and take this quantity, which is the derivative times how far I am going.

This quantity here is a first-order Taylor series approximation of my function.

So it corresponds to this black line.

What this condition says is that the function lies on top of that tangent line to my function.

It is not too difficult to show that this condition, non-negative second derivatives, implies this condition.

What you do is that you write down the second order Taylor approximation of your function  $g$ .

And then because the second order term is going to be non-negative because of that condition, you're going to get an inequality in this form.

But in any case, this inequality is pretty intuitive.

And we could take this one just as our definition of convexity-- that is, a function is convex if it has the property that whenever I draw a tangent to my curve, the function lies on top of this linear function.

So now let's move back to probability.

Suppose that  $g$  is convex.

This is true for every  $x$ .

So in particular, if I have my random variable, capital  $X$ , no matter what my random variable is, we're going to get this kind of inequality, no matter what.

And here what I they left blank is  $c$ .

Now,  $c$  is a number.

This is true for any number.

So in particular, it's true if I use as my number the expected value of  $X$ . So we have this inequality that's true now in terms of random variables.

No matter what capital  $X$  happens to be, this will be valid.

And now let us take expectations of both sides.

What we obtain is that the expected value of  $g$  of  $X$  is larger than-- now, this is a number, so the expected value of that number is the number itself plus the expected value of this term.

This quantity is a number.

The expected value of  $X$  minus the expected value of  $X$  is equal to 0.

And we have established this particular fact, which is true for any convex function.

So this is Jensen's inequality.

Let's apply it to some examples.

Let's consider the function  $g$ , which is the quadratic function.

Clearly, this is a convex function.

It has this kind of shape.

And the second derivative of this function is positive.

That's another way of verifying it.

Jensen's inequality is going to tell us something about the expected value of  $X$  squared.

Now, for this expectation, we already know that this is equal to the variance of  $X$  plus the square of the expected value.

And since the variance is always non-negative, we obtain this inequality.

This is consistent with Jensen's inequality.

Jensen's inequality tells us that  $E$  of  $g$  of  $X$ , with  $g$  the quadratic function, is larger than or equal to the square-- that is,  $g$  of the expected value.

So for the case of the square function, Jensen's inequality did not tell us anything that we didn't know.

But it's nice to confirm that it is consistent.

But we could use Jensen's inequality in another setting where the answer might not be as obvious.

For example, take the function  $X$  to the fourth.

This is also a convex function.

And Jensen's inequality is going to tell us that the fourth power of the expected value is less than or equal to the expected value of  $X$  to the fourth.

Another case of a convex function is the negative logarithm.

Remember that the logarithmic function has a shape of this kind, which curves the opposite way.

So it's called a concave function.

But if you take the negative of this function, then you're going to get something that is convex.

So by applying Jensen's inequality to this setting, what we obtain is that  $g$ , which is minus log of the expected value of  $X$ , is less than or equal to the expected value of minus log  $X$ .

And then we can remove the minus signs from both sides.

And that is going to reverse the inequality.

And we will obtain that the logarithm of the expected value of  $X$  is larger than or equal to the expected value of log  $X$ .

So in this case for the logarithmic function, the inequality goes in the opposite direction.

The reason is that the logarithmic function is a concave function, not a convex one.

And by arguing similar to this example, a concave function is the negative of a convex function.

And for concave functions, Jensen's inequality still holds, but the inequality goes the opposite way.

Jensen's inequality turns out to be quite useful.

In many cases, we want to say something about the expected value of  $g$  of  $X$ , and Jensen's inequality allows us to do that.