

In this segment, we derive and discuss the Markov inequality, a rather simple but quite useful and powerful fact about probability distributions.

The basic idea behind the Markov inequality as well as many other inequalities and bounds in probability theory is the following.

We may be interested in saying something about the probability of an extreme event.

By extreme event, we mean that some random variable takes a very large value.

If we can calculate that probability exactly, then, of course, everything is fine.

But suppose that we only have a little bit of information about the probability distribution at hand.

For example, suppose that we only know the expected value associated with that distribution.

Can we say something?

Well, here's a statement, which is quite intuitive.

If you have a non-negative random variable, and I tell you that the average or the expected value is rather small, then there should be only a very small probability that the random variable takes a very large value.

This is an intuitively plausible statement, and the Markov inequality makes that statement precise.

Here is what it says.

If we have a random variable that's non-negative and you take any positive number, the probability that the random variable exceeds that particular number is bounded by this ratio.

If the expected value of  $X$  is very small, then the probability of exceeding that value of  $a$  will also be small.

Furthermore, if  $a$  is very large, the probability of exceeding that very large value drops down because this ratio becomes smaller.

So that's what the Markov inequality says.

Let us now proceed with a derivation.

Let's start with the formula for the expected value of  $X$ , and just to keep the argument concrete, let us assume that the random variable is continuous so that the expected value is given by an integral.

The argument would be exactly the same as in the discrete case, but in the discrete case, we would be using a sum.

Now since the random variable is non-negative, this integral only ranges from 0 to infinity.

Now, we're interested, however, in values of  $X$  larger than or equal to  $a$ , and that tempts us to consider just the integral from  $a$  to infinity of the same quantity.

How do these two quantities compare to each other?

Since we're integrating a non-negative quantity, if we're integrating over a smaller range, the resulting integral will be less than or equal to this integral here, so we get an inequality that goes in this direction.

Now let us look at this integral here.

Over the range of integration that we're considering,  $X$  is at least as large as  $a$ .

Therefore, the quantity that we're integrating from  $a$  to infinity is at least as large as  $a$  times the density of  $X$ .

And now we can take this  $a$ , which is a constant, pull it outside the integral.

And what we're left with is the integral of the density from  $a$  to infinity, which is nothing but the probability that the random variable takes a value larger than or equal to  $a$ .

And now if you compare the two sides of this inequality, that's exactly what the Markov inequality is telling us.

Now it is instructive to go through a second derivation of the Markov inequality.

This derivation is essentially the same conceptually as the one that we just went through except that it is more abstract and does not require us to write down any explicit sums or integrals.

Here's how it goes.

Let us define a new random variable  $Y$ , which is equal to 0 if the random variable  $X$  happens to be less than  $a$  and it is equal to  $a$  if  $X$  happens to be larger than or equal to  $a$ .

How is  $Y$  related to  $X$ ?

If  $X$  takes a value less than  $a$ , it will still be a non-negative value, so  $X$  is going to be at least as large as the value of 0.

that  $Y$  takes.

If  $X$  is larger than or equal to  $a$ ,  $Y$  will be  $a$ , so  $X$  will again be at least as large.

So no matter what, we have the inequality that  $Y$  is always less than or equal to  $X$ . And since this is always the case, this means that the expected value of  $Y$  will be less than or equal to the expected value of  $X$ .

But now what is the expected value of  $Y$ ?

Since  $Y$  is either 0 or  $a$ , the expected value is equal to  $a$  times the probability of that event, which is  $a$  times the probability that  $X$  is larger than or equal to  $a$ .

And by comparing the two sides of this inequality, what we have is exactly the Markov inequality.

Let us now go through some simple examples.

Suppose that  $X$  is exponentially distributed with parameter or equal to 1 so that the expected value of  $X$  is also going to be equal to 1, and in that case, we obtain a bound of  $1/a$ .

To put this result in perspective, note that we're trying to bound a probability.

We know that the probability lies between 0 and 1.

There's a true value for this probability, and in this particular example because we have an exponential distribution, this probability is equal to  $e^{-a}$ .

The Markov inequality gives us a bound.

In this instance, the bound takes the form of  $1/a$ , and the inequality tells us that the true value is somewhere to the left of here.

A bound will be considered good or strong or useful if that bound turns out to be quite close to the correct value so that it also serves as a fairly accurate estimate.

Unfortunately, in this example, this is not the case because the true value falls off exponentially with  $a$ , whereas the bound that we obtained falls off at a much slower rate of  $1/a$ .

For this reason, one would like to have even better bounds than the Markov inequality, and this is one motivation for the Chebyshev inequality that we will be considering next.

But before we move there, let us consider one more example.

Suppose that  $X$  is a uniform random variable on the interval from minus 4 to 4, and we're interested in saying something about the probability that  $X$  is larger than or equal to 3.

So we're interested in this event here.

So the value of the density, because we have a range of length 8, the value of the density is  $1/8$ .

So we know that this probability has a true value of  $1/8$ , which we can indicate on a diagram here.

Probabilities are between 0 and 1.

We have a true value of  $1/8$ .

Lets us see what the Markov inequality is going to give us.

There's one difficulty that  $X$  is not a non-negative random variable, so we cannot apply the Markov inequality right away.

However, the event that  $X$  is larger than or equal to 3 is smaller than the event that the absolute value of  $X$  is larger than or equal to 3.

That is, we take this blue event and we also add this green event, and we say that the probability of the blue event is less than or equal to the probability of the blue together with the green event, which is the event that the absolute value of  $X$  is larger than or equal to 3.

So now we have a random variable, which is non-negative, and we can apply the Markov inequality and write that this is less than or equal to the expected value of the absolute value of  $X$  divided by 3.

What is this expectation of the absolute value of  $X$ ?

$X$  is uniform on this range.

The absolute value of  $X$  will be taking values only between 0 and 4.

And because the original distribution was uniform, the absolute value of  $X$  will also be uniform on the range from 0 to 4.

And for this reason, the expected value is going to be equal to 2, and we get a bound of  $2/3$ .

This is a pretty bad bound.

It is true, of course, but it is quite far from the true answer.

Could we improve this bound?

In this particular example, we can.

Because of symmetry, we know that the probability of being larger than or equal to 3 is equal to the probability of being less than or equal to minus 3.

Or the probability of this event, which is the blue and the green, is twice the probability of just the blue event.

Or to put it differently, this probably here is equal to  $1/2$  of the probability that the absolute value of  $x$  is larger than or equal to 3, and therefore, by using the same bound as here, we will obtain an answer of  $1/3$ .

So by being a little more clever and exploiting the symmetry of this distribution around 0, we get a somewhat better bound of  $1/3$ , which is, again, a true bound.

It is more informative than the original bound, but still it is quite far away from the true answer.