

RES.TLL-004 STEM Concept Videos, Fall 2013

Transcript – Moments of Distributions Transcript

Why does going to the airport seem to require extra time compared with coming back from the airport even if the traffic is the same in both directions? The answer must somehow depend on more than just the average travel time, which we're assuming is the same and often is. In fact, it depends on the distribution of travel times. Probability distributions are fully described by listing or graphing every probability. For example, how likely is a journey to the airport to be between 10 and 20 minutes? How likely is a 20—30 minute journey? A 30—40 minute journey? And so on. We'll answer the airport question at the end of the video.

This video is part of the Probability and Statistics video series. Many natural and social phenomena are probabilistic in nature. Engineers, scientists, and policymakers often use probability to model and predict system behavior.

Hi, my name is Sanjoy Mahajan, and I'm a professor of Applied Science and Engineering at Olin College. Before watching this video, you should be proficient with integration and have some familiarity with probabilities.

After watching this video, you will be able to:
Explain what moments of distributions are, and
Compute moments and understand what they mean

To illustrate what a probability distribution is, let's consider rolling two fair dice. The probability distribution of their sum is this table. For example, the only way to get a sum of two is to roll a 1 on each die. And, there are 36 possible rolls for a pair of dice. So, getting a sum of two has a probability of 1 over 36. The probability of rolling a sum of 3 is 2 over 36. And so on and so forth. You can fill in a table like this yourself. But the whole distribution, even for something as simple as two dice, is usually too much information.

We often want to characterize the shape of the distribution using only a few numbers. Of course, that throws away information, but throwing away information is the only way to fit the complexity of the world into our brains.

The art comes in keeping the most important information. Finding the moments of a distribution can help us reach our goal. Two moments that you are probably already familiar with are mean and variance. They are the two most important moments of distributions.

Let's define these moments more formally. The mean is the first moment of a distribution. It is also called the expected value and is computed as shown. Expected value of x , that's x with angled brackets around it, is equal to this sum. It's the weighted sum of all of the x 's weighted by their probabilities. Let the x sub i be the possible values of x .

For example, for the rolling of two dice, the possible values for x sub i would be 2,3,4 all the way up through 12. And p sub i would be the corresponding probabilities of rolling those sums - so that was 1 over 36, 2 over 36, and so on.

So, the first moment gives us some idea of what our distribution might look like, but not much. Think about it like this, the center of mass in these two images is in the same place, but the mass is actually distributed very differently in the two cases. We need more information.

The second moment can help us. The second moment is very similar in structure to the first moment. We write it the same way with angled brackets, but now we're talking about the expected value of x squared. So it's still a sum and it's still weighted by the probabilities p sub i , but now we square each possible x value. For the dice example that was the values from two through twelve. This is also called the mean square. First you square the x values, then you take the mean, weighting each x sub i by its probability, p sub i .

In general, the n th moment is defined as follows.

So how does the second moment help us get a better picture of our distribution? Because it can help us calculate something called the variance. The variance measures how spread out the distribution is around the mean. To calculate the variance, you first subtract the mean from each x sub i - this is like finding the distance of each x sub i from the mean - and then you square the result and multiply by p sub i .

What are the dimensions of the variance? The square of the dimensions of x . For example if the dimension is a length, then the variance is a length squared. But we often want a measure of dispersion like the variance, but one that has the same dimensions as x itself. That measure is the standard deviation, σ . σ is defined as the square root of the variance. So if the variable x has dimensions of length, then the variance will have dimensions of length squared, but the standard deviation, σ , will have dimensions of length so it's comparable to x directly.

This expression for the variance looks like a pain to compute, but it has an alternative expression that is much simpler. And you get to show that as one of the exercises after the video. The alternative

expression, the much simpler one, is that the variance is equal to the second moment, our old friend, minus the square of the first moment, or the mean.

Pause the video here to convince yourself that this difference is always non-negative.

This alternative expression for the variance, this much more useful one, is also the parallel axis theorem in mechanics, which says that the moment of inertia of an object about the center of mass is equal to the moment of inertia about an axis shifted by h from the center of mass, a parallel shift, minus mh squared.

So how does this analogy work? This, the dispersion around the mean, which is here at the center of mass, is like the variance. This is like the second moment if we make h equal to the mean. So this is the dispersion around zero or its second moment. So this is like x squared, the expected value. The mass is the sum total of all the weights here for each of x_i which all add up to one. So this is just like one in this problem. And then the h squared, well h is the mean, so this is x squared.

So you can see the exact same structure repeated with h , the shift of axis as the mean, and m the mass, as the sum of all probabilities which is one. So this formula for the variance is also the parallel axis theorem.

Let's use the definitions of the moments, and also of the related quantity, the variance, and practice on a few distributions.

A simple discrete distribution is a single coin flip. Instead of thinking of the coin flip as resulting in heads or tails, let's think about the coin as turning up a zero or one. Let p be the probability of a one.

So the mean is the weighted sum of the x_i 's, weighted by the probabilities. So the mean x is the sum $\sum p_i x_i$ which is equal to one minus p times zero plus p times one which is equal to p .

What about the second moment? X squared, it's equal to the weighted sum of the x_i 's squared so the weights are the same and we can square each value here, the x_i 's, but since they're all zero or one, squaring doesn't change them. So the second moment and the third moment and every higher moment are all p . Pause the video here and compute the variance and sketch it as a function of p .

The variance from our old convenient form of the formula is... variance of x is the mean squared, mean square minus the squared mean and all the moments themselves were just p . So that's p minus p squared which is equal to p times 1 minus p .

What does that look like? We sketch it. P on this axis, variance on that axis and the curve starts at zero

(something I can't understand) and goes back to zero.

This is a p equals 1 and that's p equals zero. Does that make sense?

Yeah, it does... from the meaning of variance as dispersion around the mean. So take the first extreme case of p equals zero. In other words, the coin has no chance of producing a one, always produces a zero every time. There the mean is zero and there is no dispersion because it always produces zero. The same applies when p equals one here at this extreme.

The coin always produces a one with no dispersion. There is no variation, there is no variance and it's plausible that the variance should be a maximum right in between... here at p equals one half which it is on this curve. So everything looks good. Our calculation seems reasonable and checks out in the extreme cases.

Before we go back to the airport problem, let's extend the idea of moments to continuous distributions.

Here, instead of a list of probabilities for each possible x , we have a probability density p as a function of x , where x is now a continuous variable. That's the continuous version for the n th moment was a sum of x_i to the n th weighted by the probabilities. Here, the n th moment, x sub n , is equal to instead of a sum, an integral. Weighted again, as always, by the probability times x sub n , as before and with a dx because p of x times dx is the probability and you add them all up over all possible values of x . So that's the formula for a continuous distribution, for the moments of a continuous distribution.

Let's practice on the simplest continuous distribution, the uniform distribution. X is equally likely to be any real number between zero and one. That's the distribution and we can compute the first and second moments and the variance.

Pause the video here and use the definition of moments for a continuous distribution and compute the mean, the first moment, the second moment, and from those two, the variance.

What you should have found is ... for the mean, it's the integral of one because p of x is one, times x between zero and one dx , which is x squared over two evaluated between zero and one, which equal one half... which makes sense. The mean here, the average value is just one-half right in the middle of the distribution of the possible values of x .

What about x squared? For that, you should have found almost the same calculation, one times x squared dx , which equals x cubed over 3 between zero and one equals one-third. And thus, the variance is equal to one-third, that's the mean square minus the squared mean, which is... one twelfth. And that number is familiar. That's the same $1/12$ that shows up in the moment of inertia of a ruler of length l

and mass m . Its moment of inertia is $1/12 ml^2$ squared which illustrates again the connection between moments of inertia and moments of distributions.

Let's apply our knowledge to understand quantitatively, or in a formal way, what happens to the airport travel – why does it seem so much longer on the way there, than on the way back?

So here, is the ideal travel experience at the airport, the distribution of travel times t . And here is the probability of that particular probability density or that travel time, p of t . In the ideal world, the travel time would be very predictable. Let's say it would almost always be twenty minutes. In that case, you would allow twenty minutes to get to the airport and you would allow twenty minutes on the way back. Going there and coming back would seem the same.

But, here's what travel to the airport actually looks like. Let's say the mean is still the same, but the reality is that there's lots of dispersion. And so the curve actually looks like that. Sometimes the travel time will be 30 minutes, sometimes 40, sometimes 10.

So now, what do you have to do?... this is reality. Well, on the way home, it's no problem. On average, you get home in twenty minutes. You leave whenever you get out of the baggage claim. And while it's true that the trip to the airport follows the same distribution, the risk to you of not making it to the airport on time is much greater. If you just allow twenty minutes, yeah, sometimes you'll get lucky, but every once in a while it will take you twenty-five or thirty minutes.

So what you have to do is allow more time on the way there so that you don't miss your flight - maybe thirty minutes, maybe even forty minutes. It all depends on the dispersion, or standard deviation, of the distribution. On the way to the airport, you are much more aware of the distribution, if you will, than you are on the way back.

In this video, we saw how to calculate the moments of a distribution and how they can help us understand something about a distribution.

When you see a complicated probability distribution, simplify it by finding its moments!

MIT OpenCourseWare
<http://ocw.mit.edu>

RES.TLL.004 STEM Concept Videos
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.