6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013
Transcript – Recitation: Sampling People on Buses

Hi. In this problem, we're dealing with buses of students going to a job convention. And in the problem, we'll be exercising our knowledge of PMFs-- probability mass functions. So we'll get a couple of opportunities to write out some PMFs, and also calculating expectations or expected values. And also, importantly, we'll actually be exercising our intuition to help us not just rely on numbers, but also to just have a sense of what the answers to some probability questions should be.

So the problem specifically deals with four buses of students. So we have buses, and in each one carries a different number of students. So the first one carries 40 students, the second one 33, the third one has 25, and the last one has 50 students for a total of 148 students. And because these students are smart, and they like probability, they are interested in a couple questions.

So suppose that one of these 148 students is chosen randomly, and so we'll assume that what that means is that each one has the same probability of being chosen. So they're chosen uniformly at random. And let's assign a couple of random variables. So we'll say x corresponds to the number of students in the bus of the selected student.

OK, so one of these 148 students is selected uniformly at random, and we'll let x correspond to the number of students in that student's bus. So if a student from this bus was chosen, then x would be 25, for example.

OK, and then let's come up with another random variable, y, which is almost the same thing. Except instead of now selecting a random student, we'll select a random bus. Or equivalently, we'll select a random bus driver.

So each bus has one driver, and instead of selecting one of the 148 students at random, we'll select one of the four bus drivers also uniformly at random. And we'll say the number of students in that driver's bus will be y. So for example, if this bus driver was selected, then y would be 33.

OK, so the main problem that we're trying to answer is what do you expect the expectation-- which one of these random variables do you expect to have the higher expectation or the higher expected value? So, would you expect x to be higher on average, or y to be higher? And what would be the intuition for this?

So obviously, we can actually write out the PMFs for x and y. These are just discrete random variables. And we can actually calculate out what the expectation is. But it's also useful to exercise your intuition, and your sense of what the answer should be.

So it might not be immediately clear which one would be higher, or you might even say that maybe it doesn't make a difference. They're actually the same. But a useful way to approach some of these questions is to try to take things to the extreme and see how that plays out.

So let's take the simpler example and take it to the extreme and say, suppose a set of four buses carrying these number of students. We have only two buses-- one bus that has only 1 student, and we have another bus that has 1,000 students. OK. And suppose we ask the same question.

Well, now if you look at it, there's a total of 1,001 students now. If you select one of the students at random, it's overwhelmingly more likely that that student will be one of the 1,000 students on this huge bus. It's very unlikely that you'll get lucky and select the one student who is by himself.

And so because of that, you have a very high chance of selecting the bus with the high number of students. And so you would expect x, the number of students, to be high-- to be almost 1,000 in the expectation. But on the other hand, if you selected the driver at random, then you have a 50/50 chance of selecting this one or that one. And so you would expect the expectation there to be roughly 500 or so. And so you can see that if you take this to the extreme, then it becomes more clear what the answer would be.

And the argument is that the expectation of x should be higher than the expectation of y, and the reason here is that because you select the student at random, you're more likely to select a student who is in a large bus, because that bus just has more students to select from. And because of that, you're more biased in favor of selecting large buses, and therefore, that makes x higher in expectation. OK, so that's the intuition behind this problem. And now, as I actually go through some of the more mechanics and write out what the PMFs and the calculation for the expectation would be to verify that our intuition is actually correct.

OK, so we have two random variables that are defined. Now let's just write out what their PMFs are. So the PMF-- we write it as little P of capital X and little x. So the random variable-- what we do is we say the probability that it will take on a certain value, right? So what is the probability that x will be 40?

Well, x will be 40 if a student from this bus was selected. And what's the probability that a student from this bus is selected? That probability is 40/148, because there's 148 students, 40 of whom are sitting in this bus. And similarly, x will be 33 with probability 33/148, and x will be 25 with probability 25/148. And x will be 50 with probability 50/148. And it will be 0 otherwise.

OK, so there is our PMF for x, and we can do the same thing for y. The PMF of y-- again, we say what is the probability that y will take on certain values? Well, y can take on the same values as x can, because we're still dealing with the number of students in each bus. So y can be 40.

But the probability that y is 40, because we're selecting the driver at random now, is 1/4, right? Because there's a 1/4 chance that we'll pick this driver. And the probability that y will be 33 will also be 1/4, and the same thing for 25 and 50. And it's 0 otherwise.

OK, so those are the PMFs for our two random variables, x and y. And we can also draw out what the PMFs look like. So if this is 25, 30, 35, 40, 45, and 50, then the probability that it's 25 is 25/148. So we can draw a mass right there.

For 33, it's a little higher, because it's 33/148 instead of 25. For 40, it's even higher still. It's 40/148. And for 50, it is still higher, because it is 50/148. And so you can see that the PMF is more heavily favored towards the larger values.

We can do the same thing for y, and we'll notice that there's a difference in how these distributions look. So if we do the same thing, the difference now is that all four of these masses will have the same height. Each one will have height 1/4, whereas this one for x, it's more heavily biased in favor of the larger ones. And so because of that, we can actually now calculate what the expectations are and figure out whether or not our intuition was correct.

OK, so now let's actually calculate out what these expectations are. So as you recall, the expectation is calculated out as a weighted sum. So for each possible value of x, you take that value and you weight it by the probability of the random variable taking on that value. So in this case, it would be 40 times 40/148, 33 times 33/148, and so on.

48 plus 25 times 25/148 plus 50 times 50/148. And if you do out this calculation, what you'll get is that it is around 39. Roughly 39.

And now we can do the same thing for y. But for y, it's different, because now instead of weighting it by these probabilities, we'll weight it by these probabilities. So each one has the same weight of 1/4.

So now we get 40 times 1/4 plus 33 times 1/4. That's 25 times 1/4 plus 50 times 1/4. And if you do out this arithmetic, what you get is that this expectation is 37. And so what we get is that, in fact, after we do out the calculations, the expected value of x is indeed greater than the expected value of y, which confirms our intuition.

OK, so this problem, to summarize-- we've reviewed how to write out a PMF and also how to calculate expectations. But also, we've got a chance to figure out some intuition behind some of these problems. And so sometimes it's helpful to take simpler things and take things to the extreme and figure out intuitively whether or not the answer makes sense.

It's useful just to verify whether the numerical answer that you get in the end is correct. Does this actually make sense? It's a useful guide for when you're solving these problems. OK, so we'll see you next time.

6.041SC Probabilistic Systems Analysis and Applied Probability
Fall 2013