6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013
Transcript – Recitation: Hypergeometric Probabilities

In this problem, we're given an urn with n balls in it, out of which m balls are red balls. To visualize it, we can draw a box that represents the set of all n balls. Somewhere in the middle or somewhere else we have a cut, such that to the left we have all the red balls (there are m), and non-red balls. Let's for now call it black balls. That is n minus m.

Now, from this box, we are to draw k balls, and we'd like to know the probability that i out of those k balls are red balls. For the rest of the problem, we'll refer to this probability as p-r, where r stands for the red balls. So from this picture, we know that we're going to draw a subset of the balls, such that i of them are red, and the remaining k minus i are black. And we'll like to know what is the probability that this event would occur.

To start, we define our sample space, omega, as the set of all ways to draw k balls out of n balls. We found a simple counting argument -- we know that size of our sample space has n-choose-k, which is the total number of ways to draw k balls out of n balls.

Next, we'd like to know how many of those samples correspond to the event that we're interested in. In particular, we would like to know c, which is equal to the number of ways to get i red balls after we draw the k balls. To do so, we'll break c into a product of two numbers -- let's call it a times b -- where a is the total number of ways to select i red balls out of m red balls. So the number of ways to get i out of m red balls.

Going back to the picture, this corresponds to the total number of ways to get these balls. And similarly, we define b as the total number of ways to get the remaining k minus i balls out of the set n minus m black balls. This corresponds to the total number of ways to select the subset right here in the right side of the box.

Now as you can see, once we have a and b, we multiply them together, and this yields the total number of ways to get i red balls. To compute what these numbers are, we see that a is equal to m-choose-i number of ways to get i red balls, and b is n minus m, the total number of black balls, choose k minus i, the balls that are not red within those k balls.

Now putting everything back, we have p-r, the probability we set out to compute, is equal to c, the size of the event, divided by the size of the entire sample space. From the previous calculations, we know that c is equal to a times b, which is then equal to m-choose-i times (n minus m)-choose-(k minus i). And on the denominator, we have the entire sample space is a size n-choose-k. And that completes our derivation.

Now let's look at a numerical example of this problem. Here, let's say we have a deck of 52 cards. And we draw a box with n equals 52, out of which we know that there are 4 aces. So we'll call these the left side of the box, which is we have m equals 4 aces. Now if we were to draw seven cards-- call it k equal to 7-- and we'd like to know what is the probability that out of the 7 cards, we have 3 aces.

Using the notation we did earlier, if we were to draw a circle representing the seven cards, we want to know what is the probability that we have 3 aces in the left side of the box and 4 non-aces for the remainder of the deck. In particular, we'll call i equal to 3. So by this point, we've cast the problem of drawing cards from the deck in the same way as we did earlier of drawing balls from an urn.

And from the expression right here, which we computed earlier, we can readily compute the probability of having 3 aces. In particular, we just have to substitute into the expression right here the value of m equal to 4, n equal to 52, k equal to 7, finally, i equal to 3. So we have 4-choose-3 times n minus m, in this case would be 48, choose k minus i, will be 4, and on the denominator, we have 52 total number of cards, choosing 7 cards. That gives us [the] numerical answer [for] the probability of getting 3 aces when we draw 7 cards.