

Bags and balls: problem of the month

Jonathan Dushoff

2021 February

Question

You have 1001 bags. Each bag has 1000 balls in it. Each ball is either black or white and each bag has a different number of black balls.

You pick a bag at random, and from it select 19 different balls at random. What is the probability that exactly 7 of the balls are black?

Answer

Let $T=1000$ be the number of balls in each bag, $t=19$ be the number of balls you select, and $b=7$ be the target number of black balls.

If B is the number of black balls in the bag you select, then the number of ways to select b black balls and $t-b$ white balls is:

$$\binom{B}{b} \binom{T-B}{t-b}$$

and the associated probability is

$$\frac{\binom{B}{b} \binom{T-B}{t-b}}{\binom{T}{t}}$$

The answer if we pick a bag at random is just the average over the $T+1$ possible values of B :

$$\frac{1}{T+1} \sum_{B=0}^T \frac{\binom{B}{b} \binom{T-B}{t-b}}{\binom{T}{t}}$$

We can separate out the parts that don't contain B :

$$\frac{1}{(T+1)\binom{T}{t}} \sum_{B=0}^T \binom{B}{b} \binom{T-B}{t-b}$$

It's simple enough to do the sum by pattern-finding and prove it by induction.

$$\sum_{B=0}^T \binom{B}{b} \binom{T-B}{t-b} = \binom{T+1}{t+1}$$

But that's no fun! If there's a simple pattern (as here), then there should be a cool combinatoric explanation, and we should pretend we found it directly.

The first factor in the summand describes ways of choosing b balls from a set of B , and the second ways of choosing $t-b$ from a set of $T-B$. How would that map to a set of $T+1$ balls? Well, we imagine that we have an extra ball to divide the two sets. Now each term in the sum counts the ways to choose $t+1$ from $T+1$ with b to the left of our divider, 1 at the divider, and the remainder to the right. The divider thus determines the position of ball $b+1$. The sum adds up over all of the places where ball $b+1$ can be (and some places where it can't, but that doesn't hurt us, we get zeroes there). So the sum counts all of the ways to choose $t+1$ from $T+1$ (as we already knew, but this is the fun way to think about it).

Now, substitute for the sum and expand factorials:

$$\frac{\binom{T+1}{t+1}}{(T+1)\binom{T}{t}} = \frac{1}{t+1}.$$

In other words, all of the $t+1$ outcomes of drawing t balls are equally likely, and thus *independent of both T and b* . This means there must be an even cooler explanation, and we should ignore everything above and pretend we found *that* directly.

Here it is. I did not find it directly, but instead by looking for it for a long time after finding the formula above and knowing it must be there.

The slick answer: Choosing a random bag from our set is equivalent to having a single bag and uniformly choosing how many of them will be black. This is in turn equivalent to putting $T + 1$ balls in order, and then picking one of them to be the divider between black (on the left, say) and white. We then pick t more at random from the bag and see how many of those are black (i.e., how many are to the left of the divider). We could generate an identical random process by first picking $t + 1$ balls from the $T + 1$ in the augmented bag and *then* randomly deciding which was the divider. In other words, the divider is equally likely to be any of the $t + 1$ chosen balls, and therefore each of the $t + 1$ possible outcomes is equally likely.