# The Subtlety of Probability Spaces

Consider the following two problems, which at first glance look very similar:

1. Elizabeth is decorating cookies for a Christmas party. She has baked 500 indistinguishable cookies and has 6 frosting colors: red, green, blue, white, pink and mocha mousse. Out of all possible configurations of frosting the cookies, Elizabeth picks one uniformly at random, and frosts the cookies following that configuration. What is the probability she ends up frosting exactly 100 cookies in red?

2. Elizabeth is decorating cookies for a Christmas party. She has baked 500 indistinguishable cookies and has 6 frosting colors: red, green, blue, white, pink and mocha mousse. She chooses a frosting color uniformly at random for each cookie, and the color she chooses for each cookie is independent of the colors she chose for the other cookies. What is the probability she ends up frosting exactly 100 cookies in red?

Despite their similarities, these questions have two different answers. So what is causing this difference? It's that the wordings imply different underlying probability spaces!

**Version 1**: In the first wording of the problem, Elizabeth is choosing uniformly among all possible configurations of frosting the cookies. Since the cookies are indistuingishable, then each configuration is just the number of cookies with each frosting flavor: i.e. one configuration might be having 300 green cookies, 50 red, 50 blue, 50 white, 30 pink, and 20 mocha. Formally, we could say

$$\Omega = \{(r, g, b, w, p, m) \in \mathbb{N}^6 : r + g + b + w + p + m = 500\}$$

where $r$ represents the number of red cookies, $g$ represents the number of green cookies, etc.

Because Elizabeth is choosing among these configurations uniformly at random, then for any possible configuration $\omega \in \Omega$, we have $\mathbb{P}(\omega) = \frac{1}{|\Omega|}$. If $A$ is the event of having exactly 100 red cookies, then $|A|$ is the number of natural number solutions so $g + b + w + p + m = 400$, which is $\binom{404}{4}$ by Stars and Bars. Likewise, $|\Omega| = \binom{505}{5}$. We therefore get

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\binom{404}{4}}{\binom{505}{5}} \approx 0.00408$$

**Version 2**: In the second wording of the problem, Elizabeth is making a random choice for each possible cookie. Just like with the Hi-Chew problem on HW10, where it helped to label the candies, in this case defining the sample space is easier if we imagine lining up the cookies in a row (otherwise we are dealing with the multinomial distribution, which you might learn about in future classes). Then we can let

$$\Omega = \{r, g, b, w, p, m\}^{500}$$

where the entries of the 500-tuples tell us which flavors Elizabeth frosted each cookie with (i.e. the outcome $\omega = (r, r, g, b, ...)$ would represent Elizabeth frosting the first cookie in the line red, the second cookie red, the third cookie green, the fourth cookie blue, etc).

Note that now for every $\omega \in \Omega$, we have $\omega = \left(\frac{1}{6}\right)^{500}$ (since there is a $\frac{1}{6}$ chance of any one cookie having a specific color, and the cookie colors are independent). By Binomial probability, we can view frosting each cookie as its own "trial", and the chance of any one cookie being red is $\frac{1}{6}$, so the chance of exactly 100 red cookies is

$$\mathbb{P}(A) = \binom{500}{100} \cdot \left(\frac{1}{6}\right)^{100} \cdot \left(\frac{5}{6}\right)^{400} \approx 0.0066.$$

**Comparing the Two**: We get different probabilities in these two cases because the way we calculate the probability of certain outcomes is not the same. In the first probability space, note that having all 500 cookies be red has exactly the same chance of having 300 red and 200 green cookies (because these are just two different configurations of frosting cookies, and each configuration is equally likely). In the second probability space, having 500 cookies be red has a much lower probability than having 300 red and 200 blue cookies (in fact, the probability is $\binom{500}{300}$ times lower!).

Neither of these probability spaces is more "right" than the other in terms of cookie decorating (if Elizabeth ever is randomly decorating 500 cookies, she can choose in which manner she wants to do so), instead the underlying probability space should be inferred from the wording of the problem. We did not phrase the original review session problem as clearly as we could have, which is why there was confusion.