151-128 Binomial Dist. on Cookie Question

Akira Van De Groenendaal

December 2024

1 Cookie Decorating

In the probability review session, someone brought up a great question about why we can't do the following question with a binomial distribution, which I didn't fully address in the live session. So here's a summary of the problem and more detailed answer about why a binomial model doesn't quite work.

Problem: 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 at random for each cookie. What is the probability she ends up frosting exactly 100 cookies in red?

Approach: We do this with stars and bars. We calculate the size of the whole sample space, which is all the ways to frost the cookies with 6 colors. Then, we calculate the number of ways to have exactly 100 red cookies. We then divide this by the size of the sample space.

1.1 Why not binomial distribution?

This is a great idea, since we can define "success" to mean a cookie is red, and "failure" means a cookie is not red. It follows that success probability $p = \frac{1}{6}$.

The issue is that a binomial distribution assumes order (a series of coin flips or some other experiment) and distinguishability (each coin is different). Then we would model $\mathbb{P}(100 \text{ reds}) = \binom{500}{100}(\frac{1}{6})^{100}(\frac{5}{6})^{400}$. However, neither order nor distinguishability holds in our cookie experiment. Intuitively, this means we can't line up our cookies and decide for each one, "you're red" or "you're not red".

After this, it was brought up that, why can't we ditch the $\binom{500}{100}$? This is another great point. However, in a binomial distribution, this essentially counts all the possible sequences of success and failure. In our case in particular, we're left with $(\frac{1}{6})^{100}(\frac{5}{6})^{400}$. However, this expression gives us the probability of **any one** sequence of 100 reds and 400 "not reds". As such, it still assumes order, but we're simply undercounting by only considering one possible sequence.