(*): Assigned to weekly problem set.

# Multinomial Distribution

1. Regular M&M candies are produced by Mars, Inc. company. Each candy takes one of six color: red, orange, yellow, green, blue, or brown. (Fun fact: in 1995, the tan colored M&M was replaced by blue!) In 2008, Mars published that the color of the candies are produced in the proportions:

| **Color** | Red | Orange | Yellow | Green | Blue | Brown |
|---|---|---|---|---|---|---|
| **Proportion** | 13% | 20% | 14% | 16% | 24% | 13% |

It is assumed that each individual candy's color is determined by the above proportions, independent of the colors of other candy pieces.

Everyone has a bag of $n =$ ___ candies. Let the vector $(R, O, Y, G, Bl, Br)$ denote the number of candy pieces of the respective colors in your particular bag.

(a) Explain why the vector $(R, O, Y, G, Bl, Br)$ has a multinomial distribution, and explicitly write down the parameters of this distribution.

(b) What is the probability that a single bag of ___ candies contains at least one candy of each color? Write this out as a probability statement, and explain in words how you would calculate it.

(c) Suppose that I am red-green color blind, and so I treat the red and green candies as the same brown-ish color, which unfortunately also look like the brown M&Ms. Let $S$ denote the number of red, green, and brown candies. What is the marginal distribution of $S$? What is the joint distribution of $(B, O, Y, S)$ (note the ordering!)? Be sure to specify the parameters for each distribution.

(d) Red dye #2 is a known carcinogen. While the red M&M candies are not produced with this dye, Mars decides to temporarily stop producing red candies due to customer concerns. (This actually happened in 1976!) What is the conditional distribution of $(O, Y, G, Bl, Br)$ given this temporary change?

2. Suppose a bowl contains 700 pieces of candy, with 100 pieces of each color (Red, Orange, Yellow, Green, Blue, Indigo, Violet). Moreover, suppose you scoop out 20 pieces of candy at random, so that every subset of size 20 is equally likely. Let $R, O, Y, G, B, I, V$ denote the number of candies in a the scoop of the respective colors.

   (a) Explain why the distribution of $(R, O, Y, G, B, I, V)$ is **not** multinomial.

   (b) Find a formula for joint PMF of $(R, O, Y, G, B, I, V)$. *Hint: It may be helpful to first think of a slightly different scenario where there are only 2 colors of candy.*

   (c) Explain how to modify the sampling procedure so the resulting joint distribution **IS** Multinomial.