Week 1 Problem Set: Probability & R Intro

DUE: Before class (i.e., before 8:00 am) on Tuesday

[NOTE: I have not necessarily left enough space for the answers. Simply cut and paste in the Word document as needed. I would prefer that equations be typed in with Word’s Equation Editor but you can print the problem set, write out your answers, and scan back into to a pdf if you prefer.]

Before we launch into the problem set, we will go over a few toy examples to get you comfortable with combinatorics.

Let’s say you have a bag of M&Ms in which only two colors were possible: Red and Blue. Each bag contains 5 M&Ms. We can list all possible bags as follows:

5 Red (1 way to get this: RRRRR)

4 Red + 1 Blue (5 ways to get this: RRRRB, RRRBR, RRBRR, RBRRR, BRRRR)

3 Red + 2 Blue (10 ways to get this: RRRBB, RRBRB, RBRRB, BRRRB, RRBBR, etc.)

2 Red + 3 Blue (10 ways to get this: BBBRR, BBRBR, BRBBR, RBBBR, BBRRB, etc.)

1 Red + 4 Blue (5 ways to get this: BBBBR, BBBRB, BBRBB, BRBBB, RBBBB)

5 Blue (1 way to get this: BBBBB)

The number of ways to get Red and Blue from a bag containing Red-Blue-only M&Ms is given by the “binomial coefficient”

(Remember that 0! = 1, and 1! = 1.)

The *probability* of getting Red and Blue from a bag containing Red-Blue-only M&Ms is just this coefficient times the joint probability of getting a Red times and a Blue times, and is given by

Make sure you understand why this is equivalent to

If we have equal probability of Red and Blue M&Ms, than

and so

If you add up all the possible ways in which you could pull out Red and Blue M&Ms from a bag with 5 M&Ms, you get exactly 32 different ways. If Red and Blue are equally likely, all 32 options are equally likely (drawing RRRRR in that order is equally likely to drawing RRBRB in that order), but since there are 10 different ways to get 3 Red and 2 Blue, this is a more likely bag than all Reds.

The probability of getting Red and Blue if the probability of getting either Red or Blue is goes as:

Red and Blue: Probability = 1/32 = 0.03125

Red and Blue: Probability = 5/32 = 0.15625

Red and Blue: Probability = 10/32 = 0.3125

Red and Blue: Probability = 10/32 = 0.3125

Red and Blue: Probability = 5/32 = 0.15625

Red and Blue: Probability = 1/32 = 0.03125

Now what if the probability of getting Red is actually 0.10 and the probability of Blue only 0.90. Make sure you understand why the probabilities are now:

Red and Blue: Probability = 1/100000 = 0.00001

Red and Blue: Probability = 45/100000 = 0.00045

Red and Blue: Probability = 810/100000 = 0.00810

Red and Blue: Probability = 7290/100000 = 0.07290

Red and Blue: Probability = 32805/100000 = 0.32805

Red and Blue: Probability = 59049/100000 = 0.59049

The multinomial distribution is just an extension of the binomial distribution in which more than two outcomes are possible.

In class you got a bag of M&M’s :

**How many Brown M&Ms: \_\_\_\_\_\_\_**

**How many Yellow M&Ms: \_\_\_\_\_\_\_**

**How many Green M&Ms: \_\_\_\_\_\_\_**

**How many Red M&Ms: \_\_\_\_\_\_\_**

**How many Orange M&Ms: \_\_\_\_\_\_\_**

**How many Blue M&Ms: \_\_\_\_\_\_\_**

We will treat the colors as being distributed randomly (i.i.d. = independent and identically distributed) with each M&M as a “trial”, and the bag of M&Ms as a “sample” drawn from a larger population of M&Ms (say, all the M&Ms produced in a given year).

**1 (4 pts). What is the FORMULA (i.e., an equation) for the probability of obtaining any given combination of colors? Please define all variables used.**

**2a (2 pts). What is the actual PROBABILITY (i.e., a number, not an equation) of obtaining your combination of colors assuming that all colors are equally likely?**

**2b (2 pts). What is the PROBABILITY of obtaining your combination of colors given the actual distribution of colors (according to the company): 24% blue, 14% brown, 16% green, 19% orange, 13% red, 14% yellow?**

**2c (2 pts). What is the PROBABILITY of obtaining your combination of colors given the actual distribution of colors in your bag of M&Ms?**

**2d (2 pts) Comparing 2a-2c, what distribution of colors (equal probability, company-stated probability, sample-based probability) makes your bag of M&Ms more likely to have occurred.**

For the problems that follow, you are going to be writing R code that involves a lot of loops and can be slow to run. R is famously slow at loops, and we’ll learn over the course of the semester ways to avoid using them. But part of the point of this exercise is to get comfortable with loops, so using loops in this context is a feature, not a bug, of the problem set. (If you are interested in knowing more about how R manages memory, you can find a nice explanation here: <http://adv-r.had.co.nz/memory.html>) Note that the default memory allocated to R is not that big, and if you get an error that you have run out of memory, you can google around for ways to increase the memory available to R (the procedure depends on your operating system).

**3 (12 pts). Write a short R script to simulate the combinations of colors that would have been possible from your bag of M&Ms assuming the company-stated color distribution (from 2b) and use this script to calculate the probability of obtaining the combination in your bag. There are many ways to solve this problem, but I am looking for you to solve it by sampling hypothetical bags of M&Ms and using that collection of bags to calculate the probability of obtaining your specific bag of M&Ms. [Hint 1: Try the R function ‘sample’. Hint 2: Make sure to write your script in a way that makes it possible to also answer Question 4] (Copy and paste your script below.)**

**4 (2 pts). According to your simulation, what was the most likely combination of colors? What was the probability of getting that most-likely combination?**

**5 (2 pts). If each bag of M&Ms had contained 50 M&Ms, would your combination of colors (whatever they were) be more or less likely to have occurred and why?**

**6 (2 pts). Lets say you have a bag of 20 M&Ms and find the first 19 are all blue. What is the probability that the 20th is also blue?**