Probability

From the beginning of the semester, we've talked about "chance" as a potential explanation for an observed correlation. For example, suppose I performed a randomized experiment and assigned five subjects to watch an negative campaign ad and 5 subjects to watch apolitical commerical advertising cookies. After the election, I used public records to check whether each participant in my experiment voted. I found that two of the five who viewed the negative ad voted, but three of the five that viewed the cookie commerical voted. Because I randomized the treatment (negative campaign ad or cookie commercial), I can rule out spuriousness and reverse causation. If I can rule out chance, then I can establish causality.

Question for you: Would chance be a reasonable explanation in this case? Is it possible that my treatment doesn't matter at all and I just happened to assign more voters to the watch the cookie commerical?

It is intuitive that with only ten subjects in the experiment, it is hard to rule out chance. But what if we had more subjects? What if 400 of 1,000 subjects who viewed the ad voted and 600 of 1,000 who viewed the cookie commerical voted? Could we more confidently rule out change then? It is intutitive that we can.

We'd like to move beyond intuition though, and think hard about an argument against chance. To do that, we'll need to introduce some basic concepts from probability theory.

Simple Random Variables

Let's start by first thinking about some random variables. We're already familiar with two:

1. coin toss: If you toss a coin, the probability of a head is 0.5 and the probability of a tail is 0.5. If you label a head by the number 1 and a tail be the number 0, then you have 0.5 probability of a 1 and a 0.5 probability of a 0.
2. die roll: If you roll a die, you have six possible outcomes: 1, 2, 3, 4, 5, and 6. Each outcome has a probability of $\frac{1}{6} \approx 0.17$.

We can actually toss a coin or roll a die, or we can have R do the hard work for us.

```r
# toss a coin once
rbinom(n = 1,  # how many repetitions
       size = 1, # how many flips in each repetition
       prob = 0.5)  # the probability of a head
```

```
## [1] 0
```

```r
# roll a die once
rmultinom(n = 1,  # how many repetitions
       size = 1, # how many rolls in each repetition
       prob = rep(1/6, 6))  # the probability of a each side
```

```
##      [,1]
## [1,]    1
## [2,]    0
## [3,]    0
## [4,]    0
## [5,]    0
## [6,]    0
```

```r
# toss a coin 10 times
rbinom(n = 10, size = 10, prob = 0.5)
```

```
##  [1] 5 3 5 5 5 6 4 5 5 3
```

```r
# roll a die 10 times
rmultinom(n = 10, size = 1, prob = rep(1/6, 6))
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    0    0    0    0    0    0    0    1    0     0
## [2,]    0    0    0    1    0    0    0    0    0     0
## [3,]    0    1    0    0    0    0    0    0    0     1
## [4,]    0    0    1    0    0    0    0    0    0     0
## [5,]    0    0    0    0    1    0    0    0    0     0
## [6,]    1    0    0    0    0    1    1    0    1     0
```

Notice that the rmultinom function returns a **matrix**, where each column represents a single repetition and each row represents a possible outcome. If the entry in row i and column j is one, that means that roll j produced a the value i (or the ith outcome). One and only one element in each column can be 1, because each repetition can only produce a single outcome.

Note that a matrix is similar to a vector, but rather than having a single dimension, it has two dimensions, rows and columns, much like a spreadsheet. There is one key difference between a matrix and a vector, and it is how we index. To extract the ith element of a vector `x` in R, we use `x[i]`. To extract the element in the ith row and jth column from a matrix `X`, we use `X[i, j]`. If we want to extract all of row i, we can use `X[i, ]`. Similarly, if we want to extract all of column j, we can use `X[, j]`.

```r
# create a vector
x <- 5:8

# extract the second element
x[2]
```

```
## [1] 6
```

```r
# create a matrix
X <- matrix(x, nrow = 2, ncol = 2)
print(X)
```

```
##      [,1] [,2]
## [1,]    5    7
## [2,]    6    8
```

```r
# extract the element from the second row and first column
X[2, 1]
```

```
## [1] 6
```

```r
# extract the entire second column
X[, 2]
```

```
## [1] 7 8
```

```r
# extract the entire first row
X[1, ]
```

```
## [1] 5 7
```

If we wanted the number each roll produced, rather than the matrix representation, we could get that as well.

```r
# roll a die 10 times
X <- rmultinom(n = 10, size = 1, prob = rep(1/6, 6))

# convert the matrix into a vector of roll numbers
J <- ncol(X)  # the number of columns in x (i.e., the number of rolls)
value <- rep(NA, J)  # a holder for our values
for (j in 1:J) {
  x <- X[, j]  # extract the jth column of X and assign it to x
  value[j] <- which(x == 1)  # find which element of the vector x equals
}

# results
print(value)
```

```
##  [1] 6 3 2 1 2 4 1 1 3 6
```

A Definition of Probability

But what does it mean for an event to have a probability of 0.5? We say that the probability of a head is 0.5 and the probability of rolling a 6 is 0.17, but what does that mean? To define these quantities, we'll rely on the frequentists notion of probability.

Definition: Refer to a a process for producing a random outcome (e.g., flipping a coin, rolling a die, pulling a number from a hat, etc.) as a probabilistic *experiment*.

Note that the probabistic experiment that produces a random outcome is unrelated to the idea of a randomized experiment for drawing causal inferences.

Definition: Refer to each instance of the experiment as a *repetition*.

Definition: Refer the proportion of times an event happens when the probabilistic experiment is repeated many (i.e., infinite) times, independently and under the same conditions, as *probability* of that event.

We can write this out mathematically, where $\Pr(\text{event}) = \frac{\#\text{ events}}{\#\text{repetitions}}$ for a very large number of repetitions.

Let's make some observations. First, note that the smallest possible number of events is zero and the largest possible number of events is the number of repetitions. This fact leads to the first observation.

Observation: A probability is bounded below by zero and above by one, so that $0 \leq \Pr(\text{event}) \leq 1$.

Second, note that the number of non-events (i.e., an event not happening) is equal to the number of repetitions minus the number of events. A little alebra shows something interesting: $\Pr(\text{non-event}) = \frac{\#\text{repetitions} - \#\text{ events}}{\#\text{repetitions}} = \frac{\#\text{repetitions}}{\#\text{repetitions}} - \frac{\#\text{ events}}{\#\text{repetitions}} = 1 - \Pr(\text{event})$. This fact leads to the second observation.

Observation: The probability of a non-event is one minus the probability of an event.

The Urn Model

Question for you: (taken from Freedman et al. 2007, pp. 223) A box contains red marbles and blue marbles. One marble is drawn at random from the box (each marble has an equal chance of being drawn). If it is red, you win $1. If it is blue, you win nothing. You can choose between two boxes: box A contains 3 red marbles and 2 blue ones, and box B contains 30 red marbles and 20 blue ones. Which box offers the better chance of winning, or are they the same?

You might prefer box A because it contains fewer blue marbles or box B because it contains more red marbles, but both lines of thinking are incorrect. Imagine drawing many times from box A. Each marble has an equal probability of being selected, so the probability of picking any particular marble is $\frac{1}{5} = 0.2$. Then the probability of picking a red marble is $\frac{3}{5} = 0.6$. Similarly, the probability of picking a red marble from box B is $\frac{30}{50} = 0.6$.

The important ratio in this problem is the $\frac{\text{\# red marbles}}{\text{\# marbles}}$. Notice that this ratio is the same for each box.

Many problems in probability can be represented as drawing colored marbles or numbered tickets from a box or "urn", so this is an important concept to understand.The key assumption is that all marbles or tickets in the urn have an equal probability of being selected with each draw.

Suppose we have a urn with three total tickets numbered 1, 2, and 3. Conceptually, we can imagine drawing in two different ways.

Sampling WITH and WITHOUT Replacement

First, you might draw two tickets WITH replacement from the urn. Using this approach, you would:

1. Stir the urn well.
2. Draw one ticket and note the number.
3. Replace the ticket and stir the urn well (again).
4. Draw a second ticket and note the number.

Second, you might draw two tickets WITHOUT replacement from the urn. Using this approach, you would:

1. Stir the urn well.
2. Draw one ticket and note the number.
3. Draw a second ticket and note the number without replacing the first draw.

The key distinction is that when drawing WITHOUT replacement, a ticket cannot be drawn multiple times.

We could actually use an urn with tickets or marbles, or we can create a vector that has the desired elements and use R to sample from that.

```
# collect 3 red marbles into a vector
red_marbles <- rep("red", 3)

# collect 2 blue marbles into a vector
blue_marbles <- rep("blue", 2)

# collect the red and blue marbles into a single vector
marbles <- c(red_marbles, blue_marbles)

# sample twice WITHOUT replacement
sample(marbles, # the vector to sample from
       size = 2, # the number of times to sample
       replace = FALSE)  # sample with replacement? (default is FALSE)
```

```
## [1] "red" "red"
```

```r
# sample twice WITH replacement
sample(marbles, size = 2, replace = TRUE)
```

```
## [1] "blue" "blue"
```

Practice Problems

1. Define "probability." Be sure to define probabilistic "experiment" and "repetition" and include each in your definition.
2. Explain why a probability cannot be less than zero or greater than one.
3. If the probability of an event is π, what is the probability of a non-event?
4. You flip a coin 10,000 times. About how many heads would you expect, on average?
5. You roll a die 10,000 times. About how many 6's would you expect, on average?
6. Suppose I have an unusually shaped die with 10 sides and each side numbered 1 to 10. Because the sides are not the same size, the probability of rolling a 1 is not necessarily 1/10. Suppose I roll this die 100,000 times and I get 5,600 1's. Use the definition of probability to estimate the probability of rolling a 1 with my unusual die.
7. Explain the difference between drawing with and without replacement.
8. (from Freedman et al. 2007, p. 226) One hundred tickets will be draw at random WITH replacement from a box of tickets. On each draw, you will be paid the amount printed on the ticket. You can choose between two boxes: box A has tickets 1 and 2 and box B has tickets 1 and 3. Which box do you choose?
9. Suppose you sample two marbles WITHOUT replacement from an urn with 7 red marbles, 3 blue marbles, and 5 green marbles. What is the probability of drawing 2 red marbles? Use a loop to repeatedly sample two marbles WITHOUT replacement from an urn with 7 red marbles, 3 blue marbles, and 5 green marbles. Use the definition of probability to estimate the answer. (Hint: You need to store the necessary information on each interaction of the loop in some way.)