Lecture 10: Discrete Distributions

For each question below, write the R code you would use and record the result when appropriate. Answer in complete sentences when asked to explain.

Part 1: Probability, Simulations, and Random Variables

1. In your own words, explain what probability is. Your explanation should mention why probabilities are always between 0 and 1.

2. In your own words, explain the difference between theoretical probability and empirical probability.

3. Suppose we flip a fair coin one time.

   (a) What are the possible outcomes?

   (b) What is the probability of getting Heads?

   (c) What is the probability of getting Tails?

   (d) Explain why this is an example of a Bernoulli trial.

4. Suppose we flip a fair coin 10 times and count the number of Heads.
   (a) What is the random variable in this situation?

   (b) Why is this no longer just a Bernoulli distribution?

   (c) What distribution would be more appropriate? Explain briefly.

Part 2: Using `matrix()` and `apply()`

5. Run the following code.

```
matrix(1:10, nrow=3)
```

   (a) Record the output.

   (b) How many rows does the matrix have?

   (c) What do you notice about how R fills the matrix?

6. Run the following code.

```
x_mat <- matrix(1:10, nrow=3)

apply(x_mat, MARGIN=1, sum)
apply(x_mat, MARGIN=2, sum)
```

   (a) What does `MARGIN=1` do?

   (b) What does `MARGIN=2` do?

   (c) Why might `apply()` be useful when running simulations?

Part 3: Coin Flips and the Binomial Distribution

7. You flip a fair coin 100 times. We want to determine the probability of flipping exactly 48 Heads.

   (a) Identify n, k, and p for this situation.

   (b) Use R to calculate the theoretical probability.

   ```
   choose(100, 48)*0.5^48*(1-0.5)^52
   ```

   (c) Interpret this probability in one sentence.

8. Now estimate the same probability empirically using 25 simulated trials.

```
trials_25 <- rbinom(25, size=100, prob=0.5)

table(trials_25)
barplot(table(factor(trials_25, levels=0:100)))

sum(trials_25 == 48)/25
```

   (a) Record the empirical probability.

   (b) Did your empirical probability match the theoretical probability exactly?

   (c) Why might 25 trials not be enough to get a very accurate estimate?

9. Repeat the simulation using 100 trials and 10,000 trials.

   (a) Record the empirical probability for 100 trials.

   (b) Record the empirical probability for 10,000 trials.

   (c) What happens to the empirical probability as the number of trials increases?

Part 4: Guessing on a Multiple-Choice Quiz

10. A multiple-choice quiz has 10 questions. Each question has 3 options, only one of which is correct. A student guesses randomly on each question.

    (a) What is the probability of getting one question correct?

    (b) What is the random variable in this situation?

    (c) Why is this an example of a Binomial distribution?

11. We want to know the probability that the student gets 5 or more questions correct.

    (a) Use R to calculate the theoretical probability.

    ```
    sum(dbinom(5:10, size=10, prob=1/3))
    ```

    (b) Use R to estimate the probability empirically with 10000 simulated students.

    ```
    quiz_scores <- rbinom(10000, size=10, prob=1/3)
    barplot(table(factor(quiz_scores, levels=0:10)))
    mean(quiz_scores >= 5)
    ```

    (c) Compare the theoretical and empirical probabilities.

12. In context, explain what this probability tells us about guessing on the quiz.

Part 5: M&M's and the Uniform Discrete Distribution

13. The M&M Factory produces 6 different color candies: Red, Orange, Yellow, Green, Blue, and Brown. Suppose the factory makes a large and equal amount of each color.
    (a) What are the possible outcomes for one randomly selected M&M?
    (b) What is the probability of getting each color?
    (c) Explain why this is an example of a Uniform Discrete distribution.

14. Simulate a bag containing 30 M&M's.

```
colors <- c("Red", "Orange", "Yellow", "Green", "Blue", "Brown")
bag_30 <- sample(colors, 30, replace=TRUE)
table(bag_30)
barplot(table(factor(bag_30, levels=colors)))
```

    (a) Record the count table.
    (b) Should you expect each color to appear exactly the same number of times in a bag of 30? Explain briefly.

15. Now simulate a Family Pack containing 250 M&M's.

```
bag_250 <- sample(colors, 250, replace=TRUE)
table(bag_250)
prop.table(table(bag_250))
barplot(table(factor(bag_250, levels=colors)))
```

(a) Record the proportions from `prop.table()`.

(b) Are the proportions closer to $\frac{1}{6}$ than they were in the bag of 30?

(c) Explain how this connects to the Law of Large Numbers.

Part 6: IT Help Desk Calls and the Poisson Distribution

16. You were recently hired to run the IT Customer Assistance desk at the Mount. You found that the rate of phone calls asking for help is 10 calls per hour.

(a) What is the value of λ in this situation?

(b) What does one random value from this distribution represent?

(c) Why might a Poisson distribution be reasonable for this situation?

17. You currently have enough people to answer 18 calls per hour. We want to know the probability that you receive 19 or more calls in an hour.

    (a) Use R to calculate the theoretical probability.

    ```
    1 - ppois(18, lambda=10)
    ```

    (b) Use R to estimate the probability empirically with 10000 simulated hours.

    ```
    calls <- rpois(10000, lambda=10)
    table(calls)
    barplot(table(calls))
    sum(calls >= 19)/10000
    ```

    (c) Compare the theoretical and empirical probabilities.

    (d) Based on this probability, would you be very worried about receiving 19 or more calls in one hour? Explain briefly.

18. Create a sentence explaining what the visualization tells us about the number of calls the help desk usually receives in an hour.