This problem set covers material from Week 2, dates 2/16- 2/20. Textbook problems (if assigned) can be found at the end of the corresponding chapter.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Monday 2/16

1. We are continuously flipping a fair coin. We stop when the absolute value of the difference between the number of Heads flips and the number of Tails flipped is 3 (e.g. TTT or THTHTTT). Let's find the probability p that we stop in **at most 6 tosses**.

   Define the event A_i as the event that we stop on the i-th toss, for $i \geq 1$.

   (a) For each i of interest, find $P(A_i)$ (provide some brief justification/reasoning for each).

   (b) Using part (a), find p. Provide justification where necessary.

2. 1.55 (No need to simplify)

Wednesday 2/18

Part (a) of the following two problems should be done on paper. The remaining parts need to be completed in the associated `.Rmd` file.

3. Suppose each of n balls is independently placed into one of n boxes at random, with all boxes equally likely for each ball.

   (a) What is the theoretical probability that *exactly one* box will be empty for $n > 1$?

   (b) Use `R` to evaluate the probability you obtained in (a) for $n = 9$ balls.

   (c) Write a function called `sim_one_empty` that estimates the probability of interest in (a) using simulation. Your function should take in as input i) the number of balls/boxes n, and ii) the number of simulations/iterations to run B. Your function should report back the probability.

   The following may be useful:

   - The function `unique()` takes in as input a vector, and returns the subset of unique entries of a vector. For example, if `x = (1, 3, 3, 2)`, then `unique(x) = (1, 3, 2)`.

- The function `length()` takes in as input a vector, and returns the number of elements in that vector. Using the same `x` as above, `length(x) = 4`.

(d) Verify your theoretical probability from (a) by using your function `sim_one_empty` for $n = 9$ and 5000 simulations.

4. A group of 8 fair four-sided dice are thrown.

   (a) What is the probability that exactly 2 of each of the values 1, 2, 3, and 4 appear?

   (b) Use `R` to evaluate the probability you obtained in (a)

   (c) Write code that estimates the probability of interest in (a) using 10000 simulations. The following may be useful:

   - The function `sort()` takes in as input a vector, and returns a vector that is sorted in ascending order . For example, if `x = (1, 3, 3, 2)`, then `sort(x) = (1, 2, 3, 3)`.
   - The function `rep(x, each = #)` takes in as input a vector `x` and repeats each element of `x` the number of times specified by the argument `each`. For example, using the same `x` as above, `rep(x, each = 2) = c(1, 1, 3, 3, 3, 3, 2, 2)`

Friday 2/20

Please define events!
Note: you can do some of these problem parts through "reasoning"/thinking conditionally or by using formulas/theorems. Both are fine, and I encourage you to think through multiple ways of solving the problem.

5. 2.20. Depending on how you approach this problem, it might be helpful to remember that intersections distribute over unions: for any events A, B, C: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

6. 2.21

7. Suppose there are 3 blood types in the population, named type 1 through type 3. A person has blood type i with probability p_i for $i = 1, 2, 3$. Consider the scenario:

   - A crime was committed, and it is known that two people together committed the crime.
   - A suspect was identified, and they have blood type 1. This suspect has probability p of being guilty (though they may not actually be guilty).
   - Evidence at the crime scene is collected, and it's later found that one of the criminals has type 1 blood, and the other criminal has type 2 blood.

   (a) Given the evidence, what is the probability that the suspect is guilty? *Simplify your solution by cancelling out common factors.*

(b) Based on your solution in (a), does the evidence make it more or less likely that the suspect is guilty when compared to the probability without evidence? Does it depend on the values of p_1, p_2, p_3? If so, give a simple criterion for when the evidence makes it more likely that the suspect is guilty.

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.