Simulation and de Montmort's Matching Problem

1. (*) We will revisit the following famous problem due to de Montmort several times throughout the term:

   > Consider a well-shuffled deck of n cards, labeled 1 through n. You flip over the cards one by one, saying the numbers 1 through n as you do so. You win the game if, at some point, the number you say aloud is the same as the number on the card being flipped over (for example, if the 7th card in the deck has the label 7). For large n, the probability of winning is approximately $1 - e^{-1}$.

   (a) For each of $n = 5, 10, 50, 100$, write a program in R that uses the `sample` function to play one iteration of the game and output whether the game results in a win or a loss. Then use the `replicate` function to simulate a large number of trials of the game in order to approximate the probability of winning. Verify that the probability of winning indeed approaches $1 - e^{-1}$ as n gets larger.

   (b) Modify the program you designed in the previous part to instead output the number of matches in one iteration of the game (again, for each of $n = 5, 10, 50, 100$).

   (c) Recall that the **mean** of a data set is the arithmetic average of the values in the set, while the **standard deviation** measures the variability of values in the set. Use `replicate` to simulate a large number of games using the program from the previous part. Then use the `mean` and `sd` functions to compute the mean and standard deviation of the number of matches in a large number of trials. What do you observe about the mean and standard deviation as n gets larger?

Non-simple Sampling

2. The `sample` function can generate outcomes from a sample space with unequal probability. If v is a vector of outcomes, and p is a vector of probabilities of those outcomes (i.e. the first entry of p is the probability of drawing the first entry of v), then `sample(v,n, p, replace = T)` generates n random outcomes from v with probabilities corresponding to p.

   (a) Suppose two teams (A and B) compete in a best of 5 tournament. In each game, team A has a 55% chance of winning (assume that the outcomes of games are independent). Approximate the probability that team A wins the tournament.

   (b) Experiment using the `sample` function with the argument `replace = F` instead of `replace = T`. If a particular entry in v is not selected in the first draw, is its probability of being selected in the second draw still given by the corresponding entry of p? If not, how does the probability appear to be determined?

(*) Indicates problems that will also be collected for homework.