# Homework 1: Random Events and Probability

University of Arizona CSC 380: Principles of Data Science

Homework due at 11:59pm on September 7, 2021

This assignment will build your knowledge of random events and random variables. It will further exercise your familiarity with measures of probability such as probability mass functions (PMFs) and probability density functions (PDFs). The questions in this assignment will build on lecture material as well as the assigned reading (Wasserman, L. 2004. "All of Statistics" : Chapters 1 & 2).

**Deliverables** Submit your homework as a PDF along with code to D2L by the stated deadline. **Show all work along with answers.** This is for your benefit as incorrect answers may receive partial credit if the work demonstrates understanding.

### Problem 1: Coinflips (1pt)

Suppose we flip a fair coin 10 times. What is the probability that the following events occur:

a) *The number of heads and the number of tails are equal*

b) *There are more heads than tails*

c) *There are less heads than tails*

d) *The $i$th flip and the $(11 - i)$the flip are the same for $i = 1 \dots 5$*

e) *We observe at least four consecutive heads*

### Problem 2: Random Monkey (1pt)

A monkey types on a 26-letter keyboard, using only lowercase letters. It is well-known that monkeys type uniformly at random from the alphabet. Answer the following:

a) *Assume that the monkey types 1,000,000 letters, what is the probability that the word "proof" appears at least once?*

b) ~~*Again, assuming that the monkey types 1,000,000 letters, what is the expected number of times the word "proof" appears?*~~

c) *On average, how many letters would the monkey need to type for the word "proof" to appear?*

**Problem 3: Random Dice (2pts)**

This problem will compare the theoretical properties of a fair die to empirical results from simulation. It will further familiarize you with the `numpy.random` library.

a) *Assume that we roll two fair six-sided dice. What is the probability of observing double sixes?*

b) *Let the random variable $X$ represent the result of rolling two fair six-sided dice. For example $X = (1, 5)$ is the event that we roll $(1, 5)$. What distribution does $X$ follow and what are the parameters of the distribution?*

c) *Initialize the random seed to 0 using `numpy.random.seed`. Using `numpy.random.randint` simulate 1,000 throws of two fair six-sided dice. From these simulations, what is the empirical probability of double sixes (the percentage of times this event occurred in simulation)?*

d) *Reset the random seed to 0 and repeat the above simulation a total of 10 times and report the empirical probability of double sixes for each run.*

e) *The empirical probability of double sixes from each simulation will differ. Yet, the probability of double sixes is fixed and was calculated in part (a) above. Why do these numbers disagree?*

f) *In the above we flipped $1,000$ coins $10$ times each. How would our results change if we flip $5,000$ coins $10$ times each? How would they change if we flip $1,000$ coins $100$ times each?*

## Problem 4: Conditional Probability (2pts)

a) *Assume that we roll two fair six-sided dice. What is $P(\text{sum is } 5 \mid \text{first die is } 3)$? What is $P(\text{sum is } 5 \mid \text{first die is } 1)$?*

b) *Assume that we roll two fair four-sided dice. What is $P(\text{sum is at least } 3)$? What is $P(\text{First die is } 1)$? What is $P(\text{sum is at least } 3 \mid \text{first die is } 1)$?*

c) *Suppose two players each roll a die, and the one with the highest roll wins. Each roll is considered a "round" and further suppose that ties magically don't happen (or those rounds are simply ignored) so there is always a winner. The best out of 7 rounds wins the match (in other words the first to win 4 rounds wins the match). Let $W$ be the event that you win the whole match. Let $S = (i, j)$ be the current score where you have $i$ wins and the opponent has $j$ wins. Determine the probability that you win the matchm, given the current score $P(W \mid S = (i, j))$ for each of the 16 possible values of $S = (i, j)$.*

To help you out a bit with part (c) above I am providing 4 big hints:

1) *If the score is equal–that is $S = (k, k)$–then there is equal chance of either player winning the match, $P(W \mid S = (k, k)) = \frac{1}{2}$ for $k = 0, 1, 2, 3$.*

*2) Let* $R_i$ *be a random variable where* $R_i = 1$ *if you win round* $i$ *and* $R_i = 0$ *if you lose that round. Note that* $P(R_i = 0) = P(R_i = 1) = \frac{1}{2}$.

*3) Recall that by the law of total probability* $P(W \mid S = (i,j)) = P(W, R_{i+j+1} = 1 \mid S = (i,j)) + P(W, R_{i+j+1} = 0 \mid S = (i,j))$.

*4) By the probability chain rule* $P(W, R_{i+j+1} \mid S = (i,j)) = P(W \mid R_{i+j+1}, S = (i,j)) P(R_{i+j+1} \mid S = (i,j))$.