HW01 — STAT/CS 287  
NAME: Cecily Page  
DATE: September 6th 2018

## P1.1.a

Sets only have keys while dictionaries have keys and values

## P1.1.b

Sets and dictionaries are both stored with hash tables and both use {Curly Braces}

## P1.1.c

You could try with a dictionary and have the values be arbitrary, or you could quasi simulate a set by using a list, but before adding any element have it go through a loop and make sure that element isn’t already in the list. This wouldn’t be perfect because lists have order while sets do not.

## P1.2

See problem1\_crpage.py

## P1.3

## {2, 3, 4, 5, 7, 8} {34, 5, 7, 8, 84, 52}

## similarity: 0.3333333333333333

## {34, 5, 7, 8, 84, 52} {2, 43, 'Octopus', 54, 84}

## similarity: 0.1

## {2, 43, 'Octopus', 54, 84} {'chicken', 'Octopus', 'Kangaroo'}

## similarity: 0.14285714285714285

## P2

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of Heads in a row -> Probability | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 0.2 | 117 | 36 | 6 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0.4 | 145 | 72 | 20 | 12 | 5 | 0 | 0 | 0 | 0 | 0 |
| 0.6 | 109 | 50 | 38 | 24 | 12 | 4 | 6 | 1 | 4 | 0 |
| 0.8 | 27 | 18 | 13 | 21 | 18 | 12 | 4 | 8 | 9 | 7 |

## P2 Bonus

## P3.1

It’s quicker to read from a file than to download the information every time. It also means you’re always using the same data – if the downloaded file is edited you could get different results (sometimes not what you want)

## P3.2

[('the', 8186), ('and', 4993), ('of', 4125), ('to', 3543), ('a', 2976), ('in', 2638), ('it', 2013), ('his', 2005), ('i', 1917), ('that', 1904), ('he', 1833), ('was', 1764), ('you', 1455), ('with', 1351), ('had', 1297), ('as', 1148), ('at', 1045), ('her', 1038), ('for', 972), ('him', 965), ('on', 932), ('not', 860), ('is', 842), ('be', 780), ('have', 742), ('said', 660), ('were', 658), ('but', 654), ('my', 653), ('mr', 620), ('this', 588), ('so', 582), ('by', 578), ('all', 571), ('there', 567), ('they', 564), ('no', 548), ('from', 529), ('me', 522), ('if', 471), ('she', 459), ('out', 446), ('one', 438), ('been', 435), ('when', 434), ('or', 434), ('which', 409), ('them', 393), ('who', 375), ('what', 371), ('an', 349), ('your', 345), ('would', 341), ('lorry', 336), ('are', 333), ('into', 319), ('their', 318), ('do', 315), ('up', 309), ('will', 295), ('upon', 291), ('could', 282), ('defarge', 280), ('man', 279), ('little', 265), ('any', 261), ('more', 261), ('its', 261), ('time', 260), ('now', 256), ('then', 253), ('hand', 247), ('miss', 232), ('down', 232), ('before', 232), ('know', 230), ('some', 229), ('again', 227), ('am', 225), ('himself', 219), ('very', 217), ('than', 216), ('two', 212), ('good', 209), ('like', 198), ('see', 198), ('looked', 193), ('other', 193), ('never', 192), ('long', 192), ('doctor', 192), ('madame', 191), ('face', 187), ('these', 187), ('old', 186), ('made', 185), ('here', 184), ('night', 182), ('much', 181), ('way', 180)]

## P3 Bonus

The library ‘collections’ has a class ‘Counter’ that takes in a list and creates an ordered dictionary that counts the number of times each item in the list appears. The dictionary is {item1: times\_appeared\_1, item2: times\_appeared+2… item\_n: times\_appeared\_n}

## P4.1

It’s important to keep your code DRY because if you need made a mistake in a function you only have to change it in one place as opposed to many and possibly missing one.

## P4.2

Yes, if q1 = q2 then the second flip has nothing to do with the first flip making it independent.

## P4.3

Count of Unique Events p=.5, q1=.4, q2=.6

('H', 'H') 189

('T', 'H') 306

('T', 'T') 183

('H', 'T') 322

Count of Unique Events p=.2, q1=.1, q2=.5

('H', 'H') 27

('T', 'H') 376

('T', 'T') 412

('H', 'T') 185

For my first dependent coin flip I chose the first coin to be 50/50 like a real coin. Then, I made the second coin 40% for a head if the first one was heads because that’s how it seems it should be (if you get heads on the first flip you might be less likely to get one on the second)

For the second, I made it even more unlikely to get double heads.

## P4.4

The simplest way to estimate the probability of each event is by using the expected value of each event. This can be calculated using the condition probability equation.

## P4.5

Probabilities for p=.5, q1=.4, q2=.6

('T', 'T') 0.2

('H', 'T') 0.3

('H', 'H') 0.2

('T', 'H') 0.3

Probabilities for p=.2, q1=.1, q2=.5

('T', 'T') 0.4

('T', 'H') 0.4

('H', 'H') 0.020000000000000004

('H', 'T') 0.18000000000000002

## P4 Bonus 1

## P4 Bonus 2