Jason Waseq
CSE 107
Lab 4

Suppose you have two coins for which P(Coin 1 = heads) = p and P(Coin 2 = heads) = q. Flip Coin 1 until the first head appears, counting the number of flips. Let N be the number of flips in this first sequence. Next, perform N flips of Coin 2, counting the number of heads. Let Y be the number of heads in this second sequence. Note that N is Geometric with parameter p. If Xi is the random variable Xi = { 1 if the i th flip of Coin 2 is heads 0 if the i th flip of Coin 2 is tails then we see each Xi is Bernoulli with parameter q. We assume of course that all coin flips are independent, so the set {N, X1, X2, … } is independent. Observe that Y = X1 + X2 + ⋯ ⋯ ⋯ + XN is the sum of a random number of independent identically distributed (iid) random variables. We will derive the mean and variance of Y, in terms of p and q, in lecture. Your goal in this assignment will be to compute both $E[Y]$ and Var(Y) experimentally, for various values of p and q. Each approximate mean and variance you compute in this project will be obtained by performing 10,000 trials of the above experiment.

Python Code:

```
import numpy as np

# Define the set of probabilities for p and q
probabilities = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]

# Number of trials
trials = 10000

# Initialize tables for mean and variance
mean_table = np.zeros((len(probabilities), len(probabilities)))
variance_table = np.zeros((len(probabilities), len(probabilities)))

# Function to simulate a geometric distribution (first sequence)
def geometric(p):
    return np.random.geometric(p)

# Function to simulate Bernoulli trials (second sequence)
def bernoulli(q, n):
    return np.random.binomial(n, q)

# Perform the simulation
for i, p in enumerate(probabilities):
    for j, q in enumerate(probabilities):
        Y_values = []
        for _ in range(trials):
```

```
            # Simulate the first sequence (geometric distribution)
            N = geometric(p)

            # Simulate the second sequence (Bernoulli trials)
            Y = bernoulli(q, N)

            Y_values.append(Y)

        # Calculate mean and variance
        mean_table[i, j] = np.mean(Y_values)
        variance_table[i, j] = np.var(Y_values)

# Print the tables
print("Mean Table:")
print(" q: ", end="")
for q in probabilities:
    print(f"{q:8.1f}", end="")
print()
for i, p in enumerate(probabilities):
    print(f"{p:3.1f} | ", end="")
    for j in range(len(probabilities)):
        print(f"{mean_table[i, j]:6.3f}", end="")
    print()

print("\nVariance Table:")
print(" q: ", end="")
for q in probabilities:
    print(f"{q:8.1f}", end="")
print()
for i, p in enumerate(probabilities):
    print(f"{p:3.1f} | ", end="")
    for j in range(len(probabilities)):
        print(f"{variance_table[i, j]:6.3f}", end="")
    print()
```

Output:

Mean Table:

q:	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.996	1.993	3.016	4.094	4.916	6.008	6.873	8.049	9.140
0.2	0.499	0.982	1.507	2.019	2.507	3.007	3.458	4.028	4.439
0.3	0.349	0.657	0.999	1.305	1.697	1.981	2.345	2.673	3.024
0.4	0.252	0.495	0.750	1.006	1.256	1.484	1.737	2.032	2.275
0.5	0.195	0.396	0.592	0.802	0.995	1.206	1.392	1.591	1.799
0.6	0.168	0.327	0.507	0.684	0.826	1.010	1.160	1.335	1.513
0.7	0.144	0.292	0.419	0.566	0.710	0.845	0.996	1.136	1.278
0.8	0.125	0.259	0.372	0.503	0.625	0.746	0.879	0.997	1.119
0.9	0.113	0.217	0.335	0.445	0.555	0.658	0.793	0.879	0.997

Variance Table:

q:	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1.773	5.099	10.760	17.967	23.807	35.104	44.667	58.763	77.242
0.2	0.678	1.545	2.790	4.496	6.344	8.615	10.291	14.058	15.793
0.3	0.399	0.835	1.450	2.023	2.865	3.494	4.660	5.516	6.681
0.4	0.267	0.540	0.876	1.249	1.580	1.903	2.316	2.884	3.383
0.5	0.198	0.392	0.595	0.791	0.987	1.214	1.377	1.570	1.758
0.6	0.165	0.298	0.467	0.594	0.673	0.792	0.877	0.990	1.100
0.7	0.138	0.267	0.352	0.435	0.501	0.549	0.596	0.608	0.600
0.8	0.116	0.219	0.287	0.352	0.392	0.404	0.408	0.396	0.356
0.9	0.104	0.179	0.246	0.291	0.314	0.311	0.297	0.262	0.208

Conjecture:

The mean and variance of Y are expected to increase as p decreases and q increases. A smaller p leads to a longer first sequence, while a larger q increases the likelihood of heads in the second sequence, resulting in more heads overall. This behavior is mathematically reflected in the expected mean and variance formulas: $E[Y] = q/p$ and $Var(Y) = q(1-q)/p^2$. These expressions indicate that both the mean and variance of Y are influenced by the values of p and q with the maximum values occurring when p is small and q is large.