

Tyler Thatcher
INF503
10/30/2018

Homework #4

Problem 1A:

- Below is the screenshot from the output of the code (using a geometric distribution), meaning that the average number of failures before success was 177.

Failures: 177

Code Used:

```
avg_num_of_failuress = np.random.geometric(p=0.005, size=100)
print("Failures: {}".format(int(avg_num_of_failuress.sum() / 100)))
```

Problem 1B:

- Below is the screenshot from the code (using a binomial distribution), meaning that the average number of successes in 10000 trials was 43.

Successes: 43

Code Used:

```
avg_num_of_successes = np.random.binomial(10000, p=0.005)
print("Successes: {}".format(int(avg_num_of_successes)))
```

Problem 2A:

- Below is the Average number of unique fragments from 100 trials which is 9860. This makes sense because you have zero error so every 16-mer you generate is going to be unique because the sequence should repeat producing close to the length of the genome.

Avg. Number of Unique Fragments (100 trials): 9860

Problem 2B:

- Below is the Average number of unique fragments with 1% error rate which is 54396. This is 5 times the amount of the unique sequences generated from zero error. Meaning that with only 1% error rate, it exploded in complexity making the amount of unique sequences way more.

Avg. Number of Unique Fragments (100 trials): 54396

Problem 2C:

- The program didn't finish. I let it run for 12 hours, I will ask you about this in class. I have no idea why it took so much longer.