Problem 4

Imagine a digital card game where there are N unique card types, each labeled from 1 to N. Players can randomly draw cards from an unlimited deck, meaning each draw is independent and has an equal chance of being any of the N card types. This scenario can be considered as sampling with replacement.

Define T as the random variable representing the number of card draws a player needs to complete a full set, with at least one of each card type.

We know that $T = \sum_{i=1}^{N} T_i$, where T_i is the number of additional card draws required to obtain the i-th unique card, given that $i - 1$ different cards have already been drawn.

1. Identify the distribution of T_i and its parameters. Note that the parameters will be in terms of N and i.
2. Develop simulation code to calculate $E(T)$ for values of $N = 15, 25, 35$.

$T_i \sim ?$ Unif, ~~Binom~~, Geom, ~~Poisson~~

↳ Bern

$T_i \sim \text{Geom}(p)$

$T_1 \sim \text{Geom}(1)$

#unseen

$T_2 \sim \text{Geom}\left(\frac{N-1}{N}\right)$

#types

$T_3 \sim \text{Geom}\left(\frac{N-2}{N}\right)$

$T_i \sim \text{Geom}\left(\frac{N-(i-1)}{N}\right)$ ✓

E(T) for N = 15, 25, 35

```
for N in [15, 25, 35]:
  for trial in 1 to num_trials:
    T_is = []
    for i in 1 to N:
      p = (N - (i - 1)) / N # calculate p
      T_i = rgeom(p) # random sample from geom
      T_is.append(T_i)

    T = sum(T_is) # inner simulation step

    T_sum += T # track total across trials to compute average

  T_avg = T_sum / num_trials # compute average
  print(N)
  print(T_avg)
```

Problem 5

In a quality control process for a batch of handcrafted artisanal chocolates, it's known that 3% of the chocolates have imperfections. Each chocolate undergoes an inspection, but the inspection process is not entirely accurate. The probability that an inspector classifies a chocolate as perfect when it actually has an imperfection is 0.04, and the probability that an inspector classifies a chocolate as imperfect when it is actually perfect is 0.03.

1. Define the random variables relevant to this binary classification scenario.
2. Calculate the sensitivity (η), specificity (θ), and prevalence (π) of the inspection process in this context.
3. What is the probability that a chocolate is actually imperfect when the inspector classifies it as such?

1. X: has imperfection $\quad X \begin{cases} 0 \text{ if no imperf.} \\ 1 \text{ if imperf.} \end{cases}$

Y: inspection result $\quad Y \begin{cases} 0 \text{ inspect perf} \\ 1 \text{ inspect imperf} \end{cases}$

$$\eta = \frac{TP}{TP+FN} = \frac{P(Y=1 \cap X=1)}{P(X=1)} = P(Y=1 \mid X=1) = \boxed{.96} \quad \pi = \frac{P}{P+N} = P(X=1) = \boxed{.03}$$

$$\theta = \frac{TN}{TN+FP} = P(Y=0 \mid X=0) = \boxed{.97} = 1 - P(Y=1 \mid X=0)$$

$$P(Y=1 \mid X=0) = .03$$
$$P(Y=0 \mid X=1) = .04$$

$$P(X=1 \mid Y=1) = \frac{P(Y=1 \mid X=1)\,P(X=1)}{P(Y=1)}$$

$$P(Y=1) = \underbrace{P(Y=1 \mid X=1)}_{\eta}\,\underbrace{P(X=1)}_{\pi} + \underbrace{P(Y=1 \mid X=0)}_{.03}\,\underbrace{P(X=0)}_{1-\pi}$$

$$\boxed{\frac{0.96 \times 0.03}{0.96 \cdot 0.03 + 0.03 \times 0.97}}$$