DIS 10A

1 Coupon Collection

Suppose you take a deck of n cards and repeatedly perform the following step: take the current top card and put it back in the deck at a uniformly random position (the probability that the card is placed in any of the n possible positions in the deck — including back on top — is $1/n$).

Consider the card that starts off on the bottom of the deck. What is the expected number of steps until this card rises to the top of the deck?

[*Hint*: Let T be the number of steps until the card rises to the top. We have $T = T_n + T_{n-1} + \cdots + T_2$, where the random variable T_i is the number of steps until the bottom card rises from position i to position $i-1$. Thus, for example, T_n is the number of steps until the bottom card rises off the bottom of the deck, and T_2 is the number of steps until the bottom card rises from second position to top position. What is the distribution of T_i?]

Solution:

Since a card at location i moves to location $i-1$ when the current top card is placed in any of the locations $i, i+1, \ldots, n$, it will rise with probability $p = (n-i+1)/n$. Thus, $T_i \sim \text{Geometric}(p)$, and $\mathbb{E}[T_i] = 1/p = n/(n-i+1)$. We now can see how this is exactly the coupon collector's problem, but with one fewer term (namely, without T_1). Finally, we can apply linearity of expectation to compute

$$\mathbb{E}[T] = \sum_{i=2}^{n} \mathbb{E}[T_i] = \sum_{i=2}^{n} \frac{n}{n-i+1} = n\sum_{i=2}^{n} \frac{1}{n-i+1} = n\sum_{i=1}^{n-1} \frac{1}{i} \approx n\ln(n-1).$$

2 Rolling Dice

(a) If we roll a fair 6-sided die, what is the expected number of times we have to roll before we roll a 6? What is the variance?

(b) Suppose we have two fair n-sided dice labeled Die 1 and Die 2. If we roll the two dice until the value on Die 1 is smaller than the value on Die 2, what is the expected number of times that we roll? What is the variance?

Solution:

(a) Since rolling a die until obtaining a 6 involves independent rolls with a constant probability of success per roll, the expected number of times we roll follows a geometric distribution.

This question seeks to review basic formulas for the geometric distribution. The probability of rolling a 6 is $1/6$. Recall that the expectation is the inverse of the probability, and that the variance is $(1-p)/p^2$. Thus the expectation is $1/(1/6) = 6$ rolls. Thus the variance is $(1-p)/p^2 = (1-1/6)/(1/6)^2 = 30$ rolls.

(b) If we roll the two dice, three outcomes are possible: the two dice show the same number, Die 1 is greater than Die 2, or Die 2 is greater than Die 1. The last two events occur with the same likelihood and the first event occurs with chance $n/n^2 = 1/n$, since there are n^2 possible rolls and n different numbers for which there could to be duplicates. Thus the number of ways that Die 1 is smaller than Die 2 on a given roll is $(n^2-n)/2$, so the probability that this occurs on a given roll is $(n^2-n)/(2n^2) = 1/2 - 1/(2n)$.

The expected number of times we roll is therefore geometrically distributed with

$$p = \frac{1}{2} - \frac{1}{2n}.$$

Plugging this into the formulas for expectation and variance yields the answer.

3 Unreliable Servers

In a single cluster of a Google competitor, there are a huge number of servers, each with a uniform and independent probability of going down in a given day. On average, 4 servers go down in the cluster per day.

(a) What is an appropriate distribution by which the number of servers that crash can be modeled?

(b) Compute the expected value and variance of the number of crashed servers for a certain cluster.

(c) Compute the probability that less than 3 servers crashed.

(d) Compute the probability at least 3 servers crashed.

Solution:

(a) Because each server goes down independently of the other servers, and with the same probability, the number of servers that crash on a given day follows a binomial distribution. We are given that the number of servers in the cluster is large, and only 4 servers crash on average per day, so $n \gg p$ and we can model the number of servers that crash as a Poisson distribution with $\lambda = 4$.

(b) Recall that the expectation and variance are both equal to $\lambda = 4$.

(c) To compute the probability that less than 3 servers went down, we must add the probabilities that 0 servers go down, 1 server goes down, and the probability that 2 servers go down. The PMF of the Poisson distribution is

$$\mathbb{P}[X = i] = \frac{\lambda^i}{i!} e^{-\lambda}.$$

Thus

$$\mathbb{P}[X = 0 \text{ or } X = 1 \text{ or } X = 2] = e^{-4} + 4e^{-4} + \frac{4^2}{2} e^{-4} = e^{-4} + 4e^{-4} + 8e^{-4} = 13e^{-4}.$$

(d) $1 - \mathbb{P}[\text{less than 3 servers crashed}] = 1 - 13e^{-4}$.

4 Will I Get My Package?

A delivery guy in some company is out delivering n packages to n customers, where $n \in \mathbb{N}$, $n > 1$. Not only does he hand a random package to each customer, he opens the package before delivering it with probability $1/2$. Let X be the number of customers who receive their own packages unopened.

(a) Compute the expectation $\mathbb{E}(X)$.

(b) Compute the variance $\text{var}(X)$.

Solution:

(a) Define

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th customer gets his/her package unopened,} \\ 0, & \text{otherwise.} \end{cases}$$

By linearity of expectation, $\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{E}(X_i)$. We have

$$\mathbb{E}(X_i) = \mathbb{P}[X_i = 1] = \frac{1}{2n},$$

since the ith customer will get his/her own package with probability $1/n$ and it will be unopened with probability $1/2$ and the delivery guy opens the packages independently.
Hence,

$$\mathbb{E}(X) = n \cdot \frac{1}{2n} = \boxed{\frac{1}{2}}.$$

(b) To calculate $\text{var}(X)$, we need to know $\mathbb{E}(X^2)$.
By linearity of expectation:

$$\mathbb{E}(X^2) = \mathbb{E}\big((X_1 + X_2 + \cdots + X_n)^2\big) = \mathbb{E}\Big(\sum_{i,j} X_i X_j\Big) = \sum_{i,j} \mathbb{E}(X_i X_j).$$

Then we consider two cases, either $i = j$ or $i \neq j$.
Hence $\sum_{i,j} \mathbb{E}(X_i X_j) = \sum_i \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_i X_j)$.

$$\mathbb{E}(X_i^2) = \mathbb{E}(X_i) = \frac{1}{2n}$$

for all i. To find $\mathbb{E}(X_i X_j)$, we need to calculate $\mathbb{P}[X_i X_j = 1]$.

$$\mathbb{P}[X_i X_j = 1] = \mathbb{P}[X_i = 1]\mathbb{P}[X_j = 1 \mid X_i = 1] = \frac{1}{2n} \cdot \frac{1}{2(n-1)}$$

since if customer i has received his/her own package, customer j has $n-1$ choices left.
Hence,

$$\mathbb{E}(X^2) = n \cdot \frac{1}{2n} + n \cdot (n-1) \cdot \frac{1}{2n} \cdot \frac{1}{2(n-1)} = \frac{3}{4},$$

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{3}{4} - \frac{1}{4} = \boxed{\frac{1}{2}}.$$