1. **Locked Out**

   You just rented a large house and the realtor gave you five keys, one for the front door and the other four for each of the four side and back doors of the house. Unfortunately, all keys look identical, so to open the front door, you are forced to try them at random.

   Find the distribution and the expectation of the number of trials you will need to open the front door. (Assume that you can mark a key after you've tried opening the front door with it and it doesn't work.)

   Solution:

   Let K be a random variable denoting the number of keys you have to try.

$$\Pr[K=1] = \frac{1}{5}$$
$$\Pr[K=2] = \frac{4}{5}\times\frac{1}{4} = \frac{1}{5}$$
$$\Pr[K=3] = \frac{4}{5}\times\frac{3}{4}\times\frac{1}{3} = \frac{1}{5}$$
$$\Pr[K=4] = \frac{4}{5}\times\frac{3}{4}\times\frac{2}{3}\times\frac{1}{2} = \frac{1}{5}$$
$$\Pr[K=5] = \frac{4}{5}\times\frac{3}{4}\times\frac{2}{3}\times\frac{1}{2}\times\frac{1}{1}\times = \frac{1}{5}$$
$$\mathbf{E}[K] = \frac{1}{5}\sum_{i=1}^{5} i = 3$$

   This result may seem surprising at first, but if we consider this experiment as follows: randomly line up keys, then try them in order, we see that this is equivalent to our earlier scheme. Furthermore, the right key is now equally likely to be in any of the five spots.

2. **Linearity**

   Solve each of the following problems using linearity of expectation. Explain your methods clearly.

   (a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability $1/3$ (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability $1/5$, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

(b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears?

(c) A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the expected number of floors the elevator stops at (not counting the ground floor)?

(d) A coin with heads probability p is flipped n times. A "run" is a maximal sequence of consecutive flips that are all the same. (Thus, for example, the sequence $HTHHHTTH$ with $n = 8$ has five runs.) Show that the expected number of runs is $1 + 2(n-1)p(1-p)$. Justify your calculation carefully.

Solution:

(a) Let A_i be the indicator you win the ith time you play game A and B_i be the same for game B. The expected value of A_i and B_i are,

$$\mathbf{E}[A_i] = 1 \cdot 1/3 + 0 \cdot 2/3 = 1/3,$$
$$\mathbf{E}[B_i] = 1 \cdot 1/5 + 0 \cdot 4/5 = 1/5.$$

Let T_A be the random variable for the number of tickets you win in game A, and T_B be the number of tickets you win in game B.

$$\begin{aligned}\mathbf{E}[T_A + T_B] &= 3\mathbf{E}[A_1] + \cdots + 3\mathbf{E}[A_{10}] + 4\mathbf{E}[B_1] + \cdots + 4\mathbf{E}[B_{20}] \\ &= 10\left(3 \cdot \frac{1}{3}\right) + 20\left(4 \cdot \frac{1}{5}\right) = 26\end{aligned}$$

(b) There are $1,000,000 - 4 + 1 = 999,997$ places where "book" can appear, each with a (non-independent) probability of $\frac{1}{26^4}$ of happening. If A is the random variable that tells how many times "book" appears, and A_i is the indicator variable that is 1 if "book" appears starting at the ith letter, then

$$\begin{aligned}\mathbf{E}[A] &= \mathbf{E}[A_1 + \cdots + A_{999,997}] \\ &= \mathbf{E}[A_1] + \cdots + \mathbf{E}[A_{999,997}] \\ &= \frac{999,997}{26^4} \approx 2.19\end{aligned}$$

times.

(c) Let A_i be the indicator that the elevator stopped at floor i.

$$\Pr[A_i = 1] = 1 - \Pr[\text{no one gets off at } i] = 1 - \left(\frac{n-1}{n}\right)^m.$$

If A is the number of floors the elevator stops at, then

$$\begin{aligned}\mathbf{E}[A] &= \mathbf{E}[A_1 + \cdots + A_n] \\ &= \mathbf{E}[A_1] + \cdots + \mathbf{E}[A_n] = n \cdot \left(1 - \left(\frac{n-1}{n}\right)^m\right)\end{aligned}$$

(d) Let A_i be the indicator for the event that a run starts at the i toss. Let $A = A_1 + \cdots + A_n$ be the random variable for the number of runs total. Obviously, $\mathbf{E}[A_1] = 1$. For $i \neq 1$,

$$\begin{aligned}\mathbf{E}[A_i] &= \Pr[A_i = 1] \\ &= \Pr[i = H \mid i-1 = T] \cdot \Pr[i-1 = T] + \Pr[i = T \mid i-1 = H] \cdot \Pr[i-1 = H] \\ &= p \cdot (1-p) + (1-p) \cdot p \\ &= 2p \cdot (1-p).\end{aligned}$$

This gives

$$\begin{aligned}\mathbf{E}[A] &= \mathbf{E}[A_1 + A_2 + \cdots + A_n] \\ &= \mathbf{E}[A_1] + \mathbf{E}[A_2] + \cdots + \mathbf{E}[A_n] = 1 + 2(n-1)p(1-p).\end{aligned}$$

3. **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, $T_i \sim \text{Geom}(p)$, and $\mathbf{E}[T_i] = 1/p = n/(n-i+1)$. 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

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