1. 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.)

   Answer: Let K be a random variable denoting the number of trials until we find the right key.

$$\begin{aligned}
Pr[K=1] &= \frac{1}{5} \\
Pr[K=2] &= \frac{4}{5}\times\frac{1}{4}=\frac{1}{5} \\
Pr[K=3] &= \frac{4}{5}\times\frac{3}{4}\times\frac{1}{3}=\frac{1}{5} \\
Pr[K=4] &= \frac{4}{5}\times\frac{3}{4}\times\frac{2}{3}\times\frac{1}{2}=\frac{1}{5} \\
Pr[K=5] &= \frac{4}{5}\times\frac{3}{4}\times\frac{2}{3}\times\frac{1}{2}\times\frac{1}{1}\times=\frac{1}{5} \\
\mathbb{E}(K) &= \frac{1}{5}\sum_{i=1}^{5} i = 3
\end{aligned}$$

   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. 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?

   Answer: Let A_i be the indicator you win the i^{th} 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.

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

□

3. 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?

Answer: 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 i^{th} letter, then

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

times. □

4. 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)?

Answer: 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

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

□

5. 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.

Answer: 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, $E[A_1]=1$. For $i\neq 1$,

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

This gives

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

□

6. Consider a random graph (undirected, no multi-edges, no self-loops) on n nodes, where each possible edge exists independently with probability p. Let X be the number of isolated nodes (nodes with degree 0). What is $E(X)$? Why isn't X a binomial distribution?

Answer: Let's first pause and ask ourselves why X is not binomial. If we consider a trial as adding an edge, which happens with probability p, we will have $\frac{n(n-1)}{2}$ trials. If we were interested in the number of edges that the resulting graph has, then it would be binomial. But unfortunately, that is not the random variable we're looking for (Star Wars reference here).

Since we are interested in the number of isolated nodes, we must instead consider a trial creating an isolated node, which happens with probability $(1-p)^{n-1}$. However, now our trials are not independent. For example, given that a node is not isolated, the conditional probability of all nodes connected to that node being isolated becomes 0.

So how can we solve this problem? Let's introduce some indicator variables $X_1, X_2, \ldots, X_n$, where $X_i = 1$ if node i is isolated.
Note that $\Pr[X_i = 1] = (1-p)^{n-1}$, and thus $\mathbf{E}(X_i) = (1-p)^{n-1}$.
Now, we can rewrite X as

$$X = X_1 + X_2 + \cdots + X_n$$

Using the Linearity of Expectation, we know

$$\begin{aligned}
\mathbf{E}(X) &= \mathbf{E}(X_1 + X_2 + \cdots + X_n) \\
&= \mathbf{E}(X_1) + \mathbf{E}(X_2) + \cdots + \mathbf{E}(X_n) \\
&= (1-p)^{n-1} + (1-p)^{n-1} + \cdots + (1-p)^{n-1} \\
&= n(1-p)^{n-1}
\end{aligned}$$

What happened here? We ended up with the same expectation as a binomial distribution, even though it wasn't binomial. In general, many different distributions can have the same expectation, but can vary greatly. This is one such example. Another is the two following distributions: one that is always $\frac{1}{2}$, and another is a fair coin toss. Both have the same expectation, but are very different.