

1 Geometric distribution

a)

Based on the fact that X_1 and X_2 defines the random variables for machines M1 and M2, who works independently in parallel successfully until they fail with probabilities p_1 and p_2 respectively, it is easy to assume that $X_1 \sim \text{Geom}(p_1)$ and $X_2 \sim \text{Geom}(p_2)$. With regards to X , it denotes the number of successful runs until either machine fails for the first time. So, $X = k$ implies that for $k-1$ runs both machines worked successfully and independently with probability $(1 - p_1) \cdot (1 - p_2)$. We define this probability as $Pr(S) = (1 - p_1) \cdot (1 - p_2)$. The complimentary event of S , \bar{S} denotes the fact that either machine fails. Based on these definition at the k -th run the $Pr(X=k)$ is the following:

$$\begin{aligned} Pr(X = k) &= P(S)^{k-1} \cdot P(\bar{S}) = ((1 - p_1) \cdot (1 - p_2))^{k-1} \cdot (1 - (1 - p_1) \cdot (1 - p_2)) \implies \\ &= ((1 - p_1) \cdot (1 - p_2))^{k-1} \cdot (p_1 + p_2 - p_1 \cdot p_2) \end{aligned}$$

We can easily assume that X follows a geometric distribution, in particular $X \sim \text{Geom}(p_1 + p_2 - p_1 \cdot p_2)$

b)

We define X the random variable that express the time James Bond needs to detect that the door is unlocked to escape. Based on the total law of expectations we will condition the $E[X]$ based on his first choice. Suppose we define the events as:

- C_1 : Choose air conditioning duct as his first choice.
- C_2 : Choose sewer pipe as his first choice.
- C_3 : Choose unlocked door as his first choice.

Then through total law of expectation $E[X]$ is the following:

$$E[X] = E[X|C_1] \cdot Pr(C_1) + E[X|C_2] \cdot Pr(C_2) + E[X|C_3] \cdot Pr(C_3) \quad (1)$$

We know that $Pr(C_1) = Pr(C_2) = Pr(C_3) = \frac{1}{3}$ and it is true that calculating the $E[X|C_1]$ and $E[X|C_2]$ of the first two options C_1 and C_2 , are going to be dependant from the $E[X]$ since he will be caught from his captors and start the process from the start having spent 2 and 5 hours respectively. So:

$$E[X|C_1] = E[X] + 2 \quad E[X|C_2] = E[X] + 5 \quad E[X|C_3] = 1$$

Therefore, since $Pr(C_1) = Pr(C_2) = Pr(C_3) = \frac{1}{3}$ the equation (1) will be:

$$\begin{aligned} E[X] &= E[X|C_1] \cdot Pr(C_1) + E[X|C_2] \cdot Pr(C_2) + E[X|C_3] \cdot Pr(C_3) \implies \\ E[X] &= \frac{(E[X] + 2) + (E[X] + 5) + 1}{3} \implies \\ E[X] &= 8 \end{aligned}$$

It will take 8 hours on average to realize that the door is unlocked.

2 Fish

a)

We define X the random variable that describes how many fish we caught in order to have all n kinds. In particular, $X_k = x_k$ implies that we caught x_k number of fishes to get the k -th new kind of fish. Therefore, it is true that before that number we have already caught $k-1$ kind of fishes. This plot describes a geometric distribution with a probability $p = 1 - \frac{k-1}{n}$ since each catch comes uniformly at random. So the expected value of X is the following:

$$\begin{aligned}
 E[X] &= \sum_{k=1}^n E[X_k] \implies \\
 &= \sum_{k=1}^n \frac{1}{p_k} \implies \\
 &= \sum_{k=1}^n \frac{1}{1 - \frac{k-1}{n}} \implies \\
 &= \sum_{k=1}^n \frac{n}{n - k + 1} \implies \\
 &= n \cdot H_n \implies \\
 &= n \cdot \ln(n) + \Theta(n)
 \end{aligned}$$

where $H_n = \sum_{k=1}^n \frac{1}{k}$ is the harmonic number as we discussed in lecture 6.

b)

We define F_i an indicator function that we do not get the i -th kind of fish after k -th catching. Also, we define as X the number of kind of fish we catch after k tries. We care about the probability $Pr(F_i = 1) = \frac{(n-1)^k}{n^k}$ since after k tries we have successfully get all the other kinds except the i -th kind. So the expected value of X , $E[X]$ is going to be:

$$\begin{aligned}
 E[X] &= \sum_{i=1}^n E[F_i] \implies \\
 &= \sum_{i=1}^n Pr(F_i = 1) \implies \\
 &= \sum_{i=1}^n \frac{(n-1)^k}{n^k} \implies \\
 &= \frac{(n-1)^k}{n^k} \sum_{i=1}^n 1 \implies \\
 &= \frac{(n-1)^k}{n^{k-1}}
 \end{aligned}$$

If we assign $k = 2 \cdot n$ then $E[X] = \frac{(n-1)^{2 \cdot n}}{n^{2 \cdot n - 1}}$ is equal to the expected value of the fishes that we caught in $2 \cdot n$ tries but also the number of fishes that we did not catch after $2 \cdot n$ tries.

c)

Using the same formula for sub question (b), it is easy to compute the expected number of kind of fish we did not get after $k = 3 \cdot n$ tries is:

$$E[X] = \frac{(n-1)^{3 \cdot n}}{n^{3 \cdot n - 1}}$$

So, if all kinds of fish are n then the expected value of kind of fish that we exactly get is $n - \frac{(n-1)^{3 \cdot n}}{n^{3 \cdot n - 1}}$.

d)

Based on sub question a, the expected value $E[X]$ is the following:

$$\begin{aligned} E[X] &= \sum_{k=1}^{\frac{n}{2}} E[X_k] \implies \\ &= \sum_{k=1}^{\frac{n}{2}} \frac{n}{n-k+1} \implies \\ &= n \cdot \left(\sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^{\frac{n}{2}} \frac{1}{k} \right) \implies \\ &= n \cdot (H_n - H_{\frac{n}{2}}) \implies \\ &\approx n \cdot \ln 2 \end{aligned}$$

where $H_n = \sum_{k=1}^n \frac{1}{k}$ is the harmonic number as we discussed in lecture 6.

e)

We define R_i the random variable that describes how many fishes exists in tank after i -th births. In particular, after $i+1$ births the expected value of fishes is the following:

$$\begin{aligned} E[R_{i+1}] &= E[R_i] + E[R_i] \cdot p_1 + E[R_i] \cdot 2 \cdot p_2 + E[R_i] \cdot 0 \cdot (1 - p_1 - p_2) \implies \\ &= E[R_i](1 + p_1 + 2 \cdot p_2) \end{aligned}$$

Assuming that initially for $i = 0$ the expected value $E[R_0] = 1$, then $E[R_i] = (1 + p_1 + 2 \cdot p_2)^i$. The only way to bound the $E[R_i]$ is if $p_1 = p_2 = 0$ so after infinity number of births the expected value $E[R_\infty] = 1$

3 Consecutive ones

a)

We define X the random variable which describes the number of rolls until we have a consecutive number of ones. From the discussion lecture we have been taught that we have to divide the sample space in disjoint events. Here, the disjoint events are the following:

- E_1 : With our first roll we did not get one.
- E_2 : With our first roll we get one but we did not get one in our second roll.
- E_3 : With both rolls we get one.

Through the total law of expectation we can compute the expected value of X as:

$$E[X] = E[X|E_1] \cdot Pr(E_1) + E[X|E_2] \cdot Pr(E_2) + E[X|E_3] \cdot Pr(E_3) \quad (2)$$

Given the fact that our dice has k sides we can compute the probabilities $Pr(E_1)$, $Pr(E_2)$ & $Pr(E_3)$ as:

$$Pr(E_1) = \frac{k-1}{k} \quad Pr(E_2) = \frac{1}{k} \cdot \frac{k-1}{k} \quad Pr(E_3) = \frac{1}{k^2}$$

So equation (2) transforms into:

$$E[X] = \frac{k-1}{k} \cdot E[X|E_1] + \frac{1}{k} \cdot \frac{k-1}{k} \cdot E[X|E_2] + \frac{1}{k^2} \cdot E[X|E_3] \quad (3)$$

Moreover, for the first event in which we did not get one at our first attempt, we start again the process and we know that $E[X]$ will be the same plus one, our first try, which we have already missed to have one. So $E[X|E_1] = E[X] + 1$. The same idea works for E_2 with $E[X|E_2] = E[X] + 2$, while for the third disjoint event E_3 we succeed to roll two consecutive ones so $E[X|E_3] = 2$. Assuming all that, the equation 3 is the following:

$$\begin{aligned} E[X] &= \frac{k-1}{k} \cdot E[X|E_1] + \frac{1}{k} \cdot \frac{k-1}{k} \cdot E[X|E_2] + \frac{1}{k^2} \cdot E[X|E_3] \implies \\ E[X] &= \frac{k-1}{k} \cdot (E[X] + 1) + \frac{1}{k} \cdot \frac{k-1}{k} \cdot (E[X] + 2) + \frac{1}{k^2} \cdot 2 \implies \\ E[X] &= k^2 + k \end{aligned}$$

b)

Following the same idea with subquestion (a), the disjoint events of our sample space are:

- E_1 : With our first roll we did not get one.
- E_2 : With our first roll we get one but we did not get one in our second roll.
- E_3 : With our first and second roll we get one but we did not get one in our third roll.

- E_4 : With first, second and third roll we get one.

Through the total law of expectation we can compute the expected value of X as:

$$E[X] = E[X|E_1] \cdot Pr(E_1) + E[X|E_2] \cdot Pr(E_2) + E[X|E_3] \cdot Pr(E_3) + E[X|E_4] \cdot Pr(E_4) \quad (4)$$

Given the fact that our dice has k sides we can compute the probabilities $Pr(E_1), Pr(E_2), Pr(E_3)$ & $Pr(E_4)$ as:

$$Pr(E_1) = \frac{k-1}{k} \quad Pr(E_2) = \frac{1}{k} \cdot \frac{k-1}{k} \quad Pr(E_3) = \frac{1}{k} \cdot \frac{1}{k} \cdot \frac{k-1}{k} \quad Pr(E_4) = \frac{1}{k^3}$$

So equation (4) transforms into:

$$E[X] = \frac{k-1}{k} \cdot E[X|E_1] + \frac{k-1}{k^2} \cdot E[X|E_2] + \frac{k-1}{k^3} \cdot E[X|E_3] + \frac{1}{k^3} \cdot E[X|E_4] \quad (5)$$

Moreover, for the first event in which we did not get one at our first attempt, we start again the process and we know that $E[X]$ will be the same plus one, our first try, which we have already missed to have one. So $E[X|E_1] = E[X] + 1$. The same idea works for E_2 with $E[X|E_2] = E[X] + 2$ and for E_3 with $E[X|E_3] = E[X] + 3$, while for the fourth disjoint event E_4 we succeed to roll three consecutive ones so $E[X|E_4] = 3$. Assuming all that, the equation 5 is the following:

$$\begin{aligned} E[X] &= \frac{k-1}{k} \cdot E[X|E_1] + \frac{k-1}{k^2} \cdot E[X|E_2] + \frac{k-1}{k^3} \cdot E[X|E_3] + \frac{1}{k^3} \cdot E[X|E_4] \implies \\ E[X] &= \frac{k-1}{k} \cdot (E[X] + 1) + \frac{k-1}{k^2} \cdot (E[X] + 2) + \frac{k-1}{k^2} \cdot (E[X] + 3) + \frac{1}{k^3} \cdot 3 \implies \\ E[X] &= k^3 + k^2 + k \end{aligned}$$

4 Randomized Quicksort

a)

Following the same process as in randomized quicksort which was learned at lecture, we define as x_1, x_2, \dots, x_n the input values of graded exams and as y_1, y_2, \dots, y_n the same input values but ordered. Also, for $i < j$ we note as an indicator random variable $X_{i,j}$ that takes value 1 if y_i and y_j are compared from us and not from the other TFs, and 0 otherwise. Then the total number of comparisons X satisfies:

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_{i,j}$$

So given the fact that $X_{i,j}$ is a indicator variable the the expected value of all the calculations, $E[X]$ would be:

$$\begin{aligned} E[X] &= E\left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n X_{i,j}\right] \implies \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n E[X_{i,j}] \implies \end{aligned}$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^n Pr(X_{i,j})$$

Hence all we have to do is to compute the probability that two elements y_i and y_j being computed from us. The probability of having value 1 for $X_{i,j}$, $Pr(X_{i,j} = 1)$ for the official randomized quicksort is calculated as $\frac{2}{j-i+1}$ since the only way to compared the values y_i and y_j is either choose as first pivot y_i or y_j over the range of set $\{y_i, \dots, y_j\}$. In all the other cases they will be divided and they will never be compared. We do not care about the set $\{1, \dots, y_i\}$ and $\{y_j, \dots, n\}$ since if the first pivot is chosen in either of those sets there is still a chance to compare them in future iterations.

In our case, the analysis is the same with one exception. Our sample space is not the set $\{y_i, \dots, y_j\}$ but the set $\{1, \dots, y_j\}$. The reason is that if the first pivot is chosen in the set $\{1, \dots, y_i\}$ then y_i and y_j will not be computed from us but from other TFs. The only way we compared the values y_i and y_j is either choose as first pivot y_i or y_j over the range of set $\{1, \dots, y_j\}$, so the probability of $Pr(X_{i,j} = 1) = \frac{2}{j}$. So the expected value $E[X]$ would be:

$$\begin{aligned} E[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n Pr(X_{i,j}) \implies \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{2}{j} \implies \\ &= 2 \cdot \sum_{i=1}^{n-1} \left(\sum_{j=1}^n \frac{1}{j} - \sum_{j=i}^n \frac{1}{j} \right) \implies \\ &= 2 \cdot (n - H_n) \end{aligned}$$

where $H_n = \sum_{i=1}^n \frac{1}{i}$ is the harmonic number as we discussed in lecture 6.

b)

Hurry's intuition is wrong since he claims that on average you give half of the remaining exams to other TFs. If each and every time you choose the median element as the pivot then you will have on average $\log_2(x)$ pivots. However, if you pick as a pivot an element bigger than the median then you just add one more pivot to the total pivots you are going to use, while if you pick as a pivot an element smaller than the median then you will terminate quicker since every time you take only the preceding elements of the pivot. Intuitively, choosing a pivot smaller than the median element has more impact than choosing an element bigger than median. So, my opinion is that since the choice of a pivot smaller of the median compared with the choice of a pivot bigger or equal than the median reduces drastically the total number of pivots, combined with the assumption that we take each time the preceding elements of the pivot, the expected number of pivots will be fewer on average compared to the assumption of Hurry, which depicts the case where we choose every time the median element in order to have on average $\log_2(x)$ pivots.

c)

We follow the same idea as sub question (a). We define as x_1, x_2, \dots, x_n the input values of graded exams and as y_1, y_2, \dots, y_n the same input values but ordered. Also, we note as an indicator random variable Z_i

that takes value 1 if z_i is the pivot, and 0 otherwise. Then the total number of pivots Z satisfies:

$$Z = \sum_{z=1}^n Z_i$$

Hence all we have to do is to compute the probability that element Z_i is the pivot. The sample space is the set $\{1, \dots, y_i\}$ since if we choose a pivot Z_k such that $\forall k, i < k \leq n$, we still have the chance to choose Z_i as a pivot but in a future iteration since we take each time to compute the grades preceding the current pivot. Focusing on the set $\{1, \dots, y_i\}$, the probability to choose as pivot z_i is $Pr(Z_i = 1) = \frac{1}{i}$. So the expected value $E[Z]$ would be:

$$\begin{aligned} E[Z] &= E\left[\sum_{i=1}^n Z_i\right] \implies \\ &= \sum_{i=1}^n E[Z_i] \implies \\ &= \sum_{i=1}^n Pr(Z_i = 1) \implies \\ &= \sum_{i=1}^n \frac{1}{i} \implies \\ &= H_n \approx \ln(n) \end{aligned}$$

where $H_n = \sum_{i=1}^n \frac{1}{i}$ is the harmonic number as we discussed in lecture 6.