CS 537: *Randomness in Computing*
Boston University

Prof. Sofya Raskhodnikova
TF: Gavin Brown
February 13, 2020

---

# Solutions to Homework 4

1. **(Geometric distribution)**

   (a) (**in parallel**) *Two faulty machines $M_1$ and $M_2$ are repeatedly run synchronously in parallel (i.e., both machines execute one run, then both machine execute a second run, and so on.) On each run, $M_1$ fails with probability $p_1$ and $M_2$ fails with probability $p_2$, all failure events are independent. Let the random variables $X_1$ and $X_2$ denote the number of runs until the first failure of $M_1$ and $M_2$, respectively. Let $X = \min\{X_1, X_2\}$ denote the number of runs until the first failure of* either *machine. What distributions do $X_1, X_2$ and $X$ have? Give a formal justification for your answer for $X$.*

   Since all failure events of $M_1$ are independent, for all $k = 1, 2, \ldots,$ we have $\Pr[X_1 = k] = (1 - p_1)^{k-1} \cdot p_1$. Hence, $X_1$ has a geometric distribution with parameter $p_1$. Similarly, $X_2$ has a geometric distribution with parameter $p_2$.

   In each run, one of the machines fails with probability $1 - (1 - p_1)(1 - p_2) = p_1 + p_2 - p_1 p_2$. Let $p$ denote this probability. Using the fact that each run is independent, we can see that

   $$\Pr[X = k] = (1 - p)^{k-1} \cdot p.$$

   Hence, $X$ also has a geometric distribution with parameter $p = p_1 + p_2 - p_1 p_2$.

   (b) (**007 style**) *James Bond is imprisoned in cell with three* possible *ways to escape: an air-conditioning duct, a sewer pipe, and the door (which is unlocked). The air-conditioning duct leads to him on a two-hour trip after which he falls through a trap door on his head, much to the amusement of his captors. The sewer pipe is similar, but takes five hours to traverse. Each fall produces temporary amnesia (in fiction, people get this a lot) and he is returned to his cell immediately after his fall. Assume he always chooses one of these three exits with probability* 1/3. *On average, how long does it take before he realizes that the door is unlocked and he just escapes?*

   Let $T$ denote the random variable for the amount of time (in hours) that it takes for Bond to escape. Let $N$ denote the random variable for the number of attempts that he makes until he escapes (including the last attempt; when he opens the door).

   $$\mathbb{E}[T] = \sum_{n=1}^{\infty} \mathbb{E}[T|N = n] \cdot \Pr[N = n]$$

   To calculate $\mathbb{E}[T|N = n]$, observe that if it takes $n$ attempts for Bond to escape, the first $n - 1$ attempts are failed attempts. The expected number of hours that Bond wastes in any particular failed attempt is $\frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 2 = 3.5$, since conditioned on the attempt being a failed one, Bond using the air conditioning duct and sewer pipe are equally likely. Using this, one can see that $\mathbb{E}[T|N = n] = 3.5(n - 1)$. Thus,

   $$\mathbb{E}[T] = 3.5 \cdot \sum_{n=1}^{\infty} (n - 1) \cdot \Pr[N = n] = 3.5 \cdot \mathbb{E}[N - 1] = 7.$$

   This follows from the fact that the number of escape attempts $N$ is a geometric random variable with success probability $\frac{1}{3}$, and thus $\mathbb{E}[N] = 3$.

2. (**Fish**) *In the ocean around Boston there are n kinds of fish. Each catch comes uniformly at random from the n kinds.*

   (a) *What is the expected number of fish you must catch to get all n kinds?*

   This is the same problem as Coupon Collector's. Let $X_i$ denote the number of fish that you need to catch in order to get a new kind of fish after you have caught $i-1$ different kinds of fish. We then have that $X_i \sim \text{Geom}(\frac{n-i+1}{n})$. The total number of fish catches required to get to $n$ different kinds is then given by $\sum_{i=1}^{n} X_i$. By the linearity of expectation,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = nH(n) = n \ln n + O(n).$$

   (Recall that $H(n)$ denotes the $n$-th Harmonic number.)

   (b) *If you catch 2n fish, what is the expected number of kinds of fish that you did not get?*

   Let $Y_i$ be the indicator random variable of the event: 'after $2n$ catches, I have not caught a fish of kind $i$'. Then

$$\mathbb{E}[Y_i] = \Pr(Y_i = 1) = \left(1 - \frac{1}{n}\right)^{2n}.$$

   Then, the number of kinds of fish that I have not caught is given by $\sum_{i=1}^{n} Y_i$ and hence its expectation (by the linearity of expectation) is:

$$\mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = n\left(1 - \frac{1}{n}\right)^{2n} \approx n \cdot e^{-2n/n} = \frac{n}{e^2}.$$

   (c) *If you catch 3n fish, what is the expected number of kinds of fish that you got exactly once?*

   Let $Z_i$ be the indicator random variable of the event 'You catch exactly one fish of type $i$ in $3n$ catches'. Then, for each $i \in [n]$,

$$\mathbb{E}[Z_i] = \Pr[Z_i = 1] = \binom{3n}{1} \cdot \frac{1}{n} \left(\frac{n-1}{n}\right)^{3n-1} = 3\left(1 - \frac{1}{n}\right)^{3n-1}.$$

   Hence, the expected number of kinds of fish that we catch exactly once is given by

$$\mathbb{E}\left[\sum_{i=1}^{n} Z_i\right] = \sum_{i=1}^{n} \mathbb{E}[Z_i] = 3n\left(1 - \frac{1}{n}\right)^{3n-1} \approx \frac{3n}{e^3}$$

   (d) *What is the expected number of fish you must catch to get n/2 kinds. (You may assume that n is even.)*

   Using the notation from part (a), we are asked to compute the expectation of the sum $\sum_{i=1}^{n/2} X_i$. By the linearity of expectation,

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

*You discovered that one of the kinds of fish can reproduce asexually. One fish of this kind produces one offspring with probability $p_1$, two offsprings with probability $p_2$, and zero offsprings with probability $1 - p_1 - p_2$.*

(e) *You caught one fish of this kind and put it in your fish tank. Assuming no fish die, what is the expected number of fish your tank will eventually have? For which values of $p_1$ and $p_2$ is it bounded?*

Let $F_i$ denote the number of fishes that get added to the population in generation $i$. The original fish is generation 0, that is, $F_0 = 1$. As a fish can reproduce exactly once, all the fish in generation $i$ are offspring of the fishes in generation $i-1$.

By the definition of expectation, $\mathbb{E}[F_1] = 1 \cdot \Pr[F_1 = 1] + 2 \cdot \Pr[F_1 = 2] = p_1 + 2p_2$. By the Law of Total Expectation (compact form), we have, for every $i > 1$,

$$\mathbb{E}[F_i] = \mathbb{E}[\mathbb{E}[F_i \mid F_{i-1}]] = \mathbb{E}[(p_1 + 2p_2)F_{i-1}] = (p_1 + 2p_2) \cdot \mathbb{E}[F_{i-1}] = (p_1 + 2p_2)^i.$$

The expected total is:

$$\mathbb{E}\Big[\sum_{i=0}^{\infty} F_i\Big] = \sum_{i=0}^{\infty} \mathbb{E}[F_i] = \sum_{i=0}^{\infty} (p_1 + 2p_2)^i.$$

This sum converges to $1/(1 - p_1 - 2p_2)$ when $p_1 + 2p_2 < 1$; otherwise, it is unbounded.

3. (Consecutive ones). *You have a $k$-sided fair die. The sides of the die are labeled $1, 2, \ldots, k$.*

(a) *You roll the die until you get a* **pair** *of consecutive ones. What is the expected number of rolls?*

Let $R_i$ denote the random variable for the result of the $i$th roll. Define the following events:
$E_1$: the event that $R_1 = 1$ and $R_2 = 1$.
$E_2$: the event that $R_1 = 1$ and $R_2 \neq 1$.
$E_3$: the event that $R_1 \neq 1$.

The three events described above are mutually exclusive and exhaustive.

Let $X$ denote the random variable for the number of rolls until seeing the first pair of consecutive 1s. Then, $\mathbb{E}[X|E_1] = 2$, as the experiment would have stopped after two rolls. Also, $\mathbb{E}[X|E_2] = \mathbb{E}[2 + X]$ and $\mathbb{E}[X|E_3] = \mathbb{E}[1 + X]$, as the random variable $X$ is memoryless.

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

Solving the above equation, we get

$$\mathbb{E}[X] = \frac{1 + \frac{1}{k}}{1 - (1 - \frac{1}{k}) \cdot (1 + \frac{1}{k})} = k^2 + k.$$

(b) *You roll the die until you get a* **triple** *of consecutive ones. What is the expected number of rolls?*

Define events:
$E_1$: the event that $R_1 = 1$, $R_2 = 1$, and $R_3 = 1$.
$E_2$: the event that $R_1 = 1$, $R_2 = 1$ and $R_3 \neq 1$.

$E_3$: the event that $R_1 = 1$ and $R_2 \neq 1$.
$E_4$: the event that $R_1 \neq 1$.
Using arguments similar to what was used before, we get

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

Solving for $\mathbb{E}[X]$ gives us $\mathbb{E}[X] = k + k^2 + k^3$.

4. (Selling a plane) *You built a plane and would like to sell it to the buyer who makes the best offer. You received responses from $n$ potential buyers who asked to see the plane. They will come to your factory one at a time and name their price. Each insists on an immediate decision from you whether you accept their price or not. You can compare each offer to all the previous offers (assume there are no ties), but you don't know what subsequent buyers will offer. Assume that buyers come in a random order, chosen uniformly at random from all $n!$ possible orderings. You want to maximize the probability of choosing the best offer.*

   *Consider the following strategy. You do not accept any of the first $m$ offers, only collect information about how much buyers value your plane. After seeing buyer $m$, you change your strategy: you accept the first offer that is better than all the previous ones. Your goal in this problem (broken down into parts) is to find the best value of $m$ as a function of $n$.*

   (a) *Suppose $m < i \leq n$. What is the probability that the best among the first $i-1$ offers is in the first $m$ offers?*
   Since the buyers come in a uniformly random order, the best among the first $(i-1)$ offers is equally likely to be in any of the positions from 1 to $i-1$. Thus, the probability of it being in the first $m$ positions is equal to $\frac{m}{i-1}$.

   (b) *Let $B$ be the event that you choose the best offer. Let $B_i$ be the event that offer $i$ is the best and you choose it. Determine* $\Pr[B_i]$.
   If $i \leq m$, then $\Pr[B_i] = 0$. Assume that $i > m$. Let $E_i$ be the event that the $i^{th}$ offer is the best and let $F_i$ be the event that you choose the best offer. In this notation, $B_i = E_i \cap F_i$. Note that each of the offers is equally likely to be the maximum. Hence, $\Pr[E_i] = \frac{1}{n}$.
   Given that the $i^{th}$ offer is the best one, for you to pick the $i$th offer, the values of all the offers from $m+1$ to $i-1$ must be smaller than the value of the largest among the first $m$ offers. In other words, conditioned on $E_i$, the event $F_i$ that you pick the $i$th offer is equivalent to the event that the best among the first $i-1$ offers being among the first $m$. Hence,

   $$\Pr[B_i] = \Pr[E_i \cap F_i] = \Pr[E_i]\cdot\Pr[F_i \mid E_i] = \frac{1}{n}\cdot\frac{m}{i-1},$$

   where $\Pr[F_i \mid E_i] = \frac{m}{i-1}$ by combining the discussion above with part (a).

   (c) *Show that* $\Pr[B] = \frac{m}{n}\sum_{i=m+1}^{n}\frac{1}{i-1}$.
   Since the events $B_i$ are disjoint, $\Pr[B] = \sum_{i\in[n]}\Pr[B_i] = \frac{m}{n}\cdot\sum_{i=m+1}^{n}\frac{1}{i-1}$.

   (d) *Show that* $\frac{m}{n}\ln\frac{n}{m} \leq \Pr[B] \leq \frac{m}{n}\ln\frac{n-1}{m-1}$.
   Note that $\sum_{i=m+1}^{n}\frac{1}{i-1} = \sum_{j=m}^{n-1}\frac{1}{j}$.

Using arguments similar to those used to prove MU's Lemma 2.10, we can see that $\sum_{j=m}^{n-1} \frac{1}{j} \geq \int_m^n \frac{1}{x}dx = \ln n - \ln m = \ln \frac{n}{m}$. Note that the book uses the bound $\int_1^n \frac{1}{x}dx \leq \sum_{i=1}^{n} \frac{1}{i}$, but they could have stated something tighter: $\int_1^n \frac{1}{x}dx \leq \sum_{i=1}^{n-1} \frac{1}{i}$. Also, $\sum_{j=m}^{n-1} \frac{1}{j} \leq \int_{m-1}^{n-1} \frac{1}{x}dx = \ln(n-1) - \ln(m-1) = \ln \frac{n-1}{m-1}$. The required inequality follows from these two inequalities and part (c).

(e) *Show that $\frac{m}{n} \ln \frac{n}{m}$ is maximized when $\frac{m}{n} = \frac{1}{e}$. Give a lower bound on $\Pr[B]$ when $m = \frac{n}{e}$.*
Let $x$ be a shorthand for $\frac{m}{n}$. The first and second order derivatives of $x \log \frac{1}{x}$ with respect to $x$ are equal to $\log \frac{1}{x} - 1$ and $\frac{-1}{x}$ respectively. We get the value $x = \frac{1}{e}$ by solving $\log \frac{1}{x} - 1 = 0$. Substituting this in $\frac{-1}{x}$, we get $-e \leq 0$. Hence, $x = \frac{1}{e}$ is a maximizer for the function $x \log \frac{1}{x}$. From part (d), we have $\Pr[B] \geq \frac{m}{n} \ln \frac{n}{m} \geq \frac{1}{e} \ln \ln e = \frac{1}{e}$.

5. (Optional, no collaboration). *Konstantinos asks you to sort the homework assignments you graded in alphabetical order. You are trying to convince him that the assignments are already nearly sorted. You pull out a set of assignments (of your choice) in order they appear, and Konstantinos checks that they are indeed in alphabetical order. Suppose there are n assignments, and they are actually in a uniformly random order. Let L be the largest number of assignments for which the check that Konstantinos performs passes. (For example, if the assignments are in order Feng, Ali, Palak, Will, Dina, Miles, Iden, then L is 3: Ali, Dina, and Iden appear in the correct order, but there is no larger set of assignments that are in alphabetical order.) Prove that the expectation of L is at least $\sqrt{n}$. (For simplicity, you may assume that $\sqrt{n}$ is an integer, and don't actually play this trick on Konstantinos!)*

Instead of names, we'll think about the equivalent setting of uniformly random permutations on the integers $1, \ldots, n$. Then $L$ is the length of the longest increasing subsequence. Let random variable $D$ be the length of the longest *decreasing* subsequence. By symmetry, $\mathbb{E}[L] = \mathbb{E}[D]$. For a fixed sequence $s$, let $L(s)$ and $D(s)$ be the length of the longest increasing and decreasing subsequences, respectively. We'll use the following lemma, which roughly says that there's always either a long increasing or a long decreasing sequence:

**Lemma 1.** *For any sequence $s = (s_1, \ldots, s_n)$ of distinct[1] elements, $L(s) \cdot D(s) \geq n$.*

*Proof.* Let $l_i$ be the length of the longest increasing subsequence of $s$ that ends with $s_i$. (Note that this, in general, is different than the length of the longest increasing subsequence of the first $i$ elements.) Similarly, let $d_i$ be the length of the longest decreasing subsequence that ends with $s_i$. For each $i \in [n]$, we have a pair $(l_i, d_i)$. These pairs are distinct: for every $i < j$, either $s_j > s_i$ or $s_j < s_i$; in the former case, $l_j > l_i$ and, in the latter case, $d_j > d_i$. We have $L(s) = \max_{i \in [n]} l_i$, and similarly for $D(s)$, so $(l_i, d_i) \in \{1, \ldots, L(S)\} \times \{1, \ldots, D(s)\}$. Since there are $n$ distinct values for $(l_i, d_i)$,

$$n \leq \left|\{1, \ldots, L(S)\} \times \{1, \ldots, D(s)\}\right| = L(s) \cdot D(s).$$

□

We use Lemma 1, the symmetry between $L$ and $D$, the linearity of expectation, and the inequality of arithmetic and geometric means to to give a lower bound on the expectation:

$$\mathbb{E}[L] = \frac{\mathbb{E}[L] + \mathbb{E}[D]}{2} = \mathbb{E}\left[\frac{L+D}{2}\right] \geq \mathbb{E}\left[\sqrt{LD}\right] \geq \sqrt{n}.$$

[1]This assumption can be removed, but the current statement suffices for solving the problem.