Advanced Algorithms II (Fall 2020)

Instructor: Chihao Zhang
Scribed by: Erzhi Liu, Siyu Sun

Last modified on Sept 20, 2020

In this lecture we first review some basic properties of random variables. We then examine the linearity of expectations and utilize the property to analyze a few examples of random processes. At last we introduce a useful inequality due to Karp, Upfal and Widgerson for analyzing linear recurrence involving random variables.

1 Random Variables

A probability space is a tuple $(\Omega, \mathscr{F}, \Pr)$ where Ω is the universe, $\mathscr{F}$ is the collection of events and Pr is a probability measure. Today we will assume that Ω is countable, and $\mathscr{F}$ is simply 2^{Ω}. A random variable X is a function $X : \Omega \to \mathbb{R}$. The expectation of X is defined as $\mathbf{E}[X] = \sum_{a:Pr[X=a]>0} a \cdot \mathbf{Pr}[X = a]$.

For any n random variables $X_1, X_2, \dots, X_n$, we have

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i].$$

This property is called the *linearity of expectations* and can be easily verified via the definition:

$$\begin{aligned}
\mathbf{E}[X_1 + X_2] &= \sum_{a,b} (a+b) \cdot \mathbf{Pr}[X_1 = a, X_2 = b] \\
&= \sum_{a,b} a \cdot \mathbf{Pr}[X_1 = a, X_2 = b] + \sum_{a,b} b \cdot \mathbf{Pr}[X_1 = a, X_2 = b] \\
&= \sum_{a} a \cdot \mathbf{Pr}[X_1 = a]] + \sum_{b} b \cdot \mathbf{Pr}[X_2 = b] \\
&= \mathbf{E}[X_1] + \mathbf{E}[X_2]
\end{aligned}$$

Note that we do not need to assume any independence among $\{X_i\}_{i=1,\dots,n}$.

2 The Coupon Collector Problem

The coupon collector problem asks the following question: If each box of a brand of cereals contains a coupon, randomly chosen from n different types of coupons, what is the expected number of boxes one needs to buy to collect all n coupons? An alternative statement is: Given n coupons, how many coupons one expects to draw with replacement before having drawn each coupon at least once?

The expectation can be simply calculated using the linearity property of the expectations.

Let X_i be the number of draws to get the i-th distinct coupon while exactly $i-1$ distinct coupons are already in hand. Then the number of draws X to collect all coupons satisfies

$$X = \sum_{i=0}^{n-1} X_i,$$

so by the linearity of expectations:

$$\mathbf{E}\left[X\right] = \mathbf{E}\left[\sum_{i=0}^{n-1} X_i\right] = \sum_{i=0}^{n-1} \mathbf{E}\left[X_i\right].$$

It is clear that X_i satisfies the geometric distribution with parameter $p_i = \frac{n-i}{n}$, namely X_i is the number of coins one needs to toss before seeing the first HEAD where each independent toss of a coin gives HEAD with probability p_i. We often write $X_i \sim \mathrm{Geom}(p_i)$.

Lemma 1. *Let* $X \sim \mathrm{Geom}(p)$. *Then* $\mathbf{E}\left[X\right] = \frac{1}{p}$.

Proof. We give two proofs.

1. By the definition of the geometric distribution,

$$S \triangleq \mathbf{E}\left[X\right] = \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} \cdot p. \tag{1}$$

   Then

$$(1-p) \cdot S = \sum_{i=1}^{\infty} i \cdot (1-p)^{i} \cdot p \tag{2}$$

   (1)-(2) yields

$$S = \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{1}{p}$$

2. If the first toss is not HEAD, counting from the second toss, the expectation should be the same as S. Therefore the expectation satisfies the following identity:

$$S = p \cdot 1 + (1-p) \cdot (S+1),$$

   which immediately gives $S = \frac{1}{p}$.

□

Now let's come back to the coupon collector:

$$\mathbf{E}\left[X\right] = \mathbf{E}\left[\sum_{i=0}^{n-1} X_i\right] = \sum_{i=0}^{n-1} \mathbf{E}\left[X_i\right] = \sum_{i=0}^{n-1} \frac{n}{n-i} = n \cdot H(n) \overset{n\to\infty}{\longrightarrow} n(\ln n + \gamma),$$

where the constant $\gamma = 0.577\ldots$ is called the Euler constant.

3 Failure of Linearity

In this section, we consider two situations where the linearity of expectation might fail.

1. $n = \infty$. Consider the St. Petersburg paradox. In each stage of the game, a fair coin is tossed and a gambler guesses the result. He wins the amount he bet if his guess is correct and loses the money if he is wrong. He bets 1 dollar at the first stage. If he loses, he doubles the money and bets again. The game ends when the gambler wins. In stage i, he wins X_i with $\mathbf{E}[X_i] = 0$, so $X = \sum_{i=1}^{\infty} \mathbb{E}[X_i] = 0$. On the other hand, he eventually wins 1 dollar.

$$\mathbf{E}\left[\sum_{i=1}^{\infty} X_i\right] = 1 \neq \sum_{i=1}^{\infty} \mathbf{E}[X_i]$$

2. n is random. Suppose $X_1 = X_2 = \dots = X_N = N$, N is random in $\{1, \dots, 6\}$. Then

$$\mathbf{E}[N] \cdot \mathbf{E}[X_1] = 12.25;$$
$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbf{E}[N \cdot N] = 15.166\ldots$$

We will say more about these two facts later in this course.

4 Quick Select

Now we apply the linearity of expectations to analyze a randomized algorithm. Given an unsorted array A of n *distinct* numbers and an integer $k \in \{1, \dots, n\}$, consider the problem to find the k-th largest number in A. The following `Quick Select` algorithm solves the problem recursively.

Algorithm 1 The `Quick Select` algorithm to find the k-th largest element

Input: *An unsorted array A and a number k.*
Output: *The k-th largest number of A.*

function FIND(A, k)
  Pick $x \in A$ uniformly at random.
  Partition $A - \{x\}$ into A_1, A_2 such that $\forall y \in A_1, y < x, \forall z \in A_2, z > x$
  if $|A_1| = k - 1$ **then**
    return x
  end if
  if $|A_1| > k - 1$ **then**
    return FIND(A_1, k)
  end if
  return FIND($A_2, k - |A_1| - 1$)
end function

We will analyze the expected running time of FIND(A, k).

4.1 The running time of Quick Select

Define X_i as the size of the set A at the i-call to $\textsc{Find}(A, k)$. Then $X_1 = n$. It is clear that the partition step in Line 3 can be implemented in $O(|A|)$ time, so the total running time of the algorithm is proportional to $\sum_{i=1}^{\infty} X_i$. The following lemma shows that on average X_{i+1} cannot be too large comparing to X_i.

Lemma 2. *For every $i \geq 1$, $\mathbf{E}\left[X_{i+1} \mid X_i\right] \leq \frac{3}{4} X_i$.*

Proof. We only prove the lemma for $i = 1$. The same argument applies for larger i. In the algorithm, we use a random x to split the set A into two smaller sets A_1 and A_2. Assuming x is the ℓ-th largest number in A, then $|A_1| = \ell - 1$ and $|A_2| = n - \ell$. In case $\ell \neq k$, the algorithm then recursively calls $\textsc{Find}(A_1, k)$ or $\textsc{Find}(A_2, k - |A_1| - 1)$. So we have

$$\mathbf{E}\left[X_2 \mid X_1\right] \leq \mathbf{E}\left[\max\left\{\ell - 1, n - \ell\right\}\right].$$

To compute the expectation of the maximum of two numbers, we distinguish between whether $\ell - 1 \leq n - x$:

$$\mathbf{E}\left[\max\left\{\ell - 1, n - \ell\right\}\right] = \mathbf{E}\left[x - 1 \mid x - 1 > n - 1\right] \mathbf{Pr}\left[x - 1 > n - 1\right] + \mathbf{E}\left[n - x \mid x - 1 \leq n - x\right] \mathbf{Pr}\left[x - 1 \leq n - x\right].$$

Note that both $\mathbf{E}\left[x - 1 \mid x - 1 > n - 1\right]$ and $\mathbf{E}\left[n - x \mid x - 1 \leq n - x\right]$ are at most $\frac{3}{4}n$, we obtain $\mathbf{E}\left[X_2 \mid X_1\right] \leq \frac{3}{4}n$.

□

It follows from the lemma that

$$\mathbf{E}\left[X_{i+1}\right] = \mathbf{E}\left[\mathbf{E}\left[X_{i+1} \mid X_i\right]\right] \leq \frac{3}{4}\mathbf{E}\left[X_i\right] \leq \left(\frac{3}{4}\right)^i n$$

Let $T = \sum_{i=1}^{\infty} X_i$ be the total running time, then by the linearity of expectations,

$$\mathbf{E}\left[T\right] = \mathbf{E}\left[\sum_{i=1}^{\infty} X_i\right] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}\left[X_i\right] \leq \sum_{i=1}^{n} \left(\frac{3}{4}\right)^{i-1} n = 4n.$$

The second equality above is due to the fact that $X_{i+1} \leq X_i - 1$ for every $i \geq 1$.

5 Karp-Upfal-Wigderson Inequality

While analyzing randomized algorithms involving recursive calls, if we use $T(n)$ to denote (an upper bound of) the running time on instances of size n, one often meets a recurrence like:

$$T(n) \leq 1 + T(n - X_n)$$

where X_n is a random variable indicating the size reduced in the recursive call. The following useful inequality due to Karp, Upfal and Wigderson [1] provides an upper bound on $\mathbf{E}\left[T(n)\right]$.

Theorem 3. *Let $T : \mathbb{N} \to \mathbb{N}$ be a function. Assuming it satisfies $T(n) \leq 1 + T(n - X_n)$ for certain random variables X_n and moreover*

1. *For some integer a, $T(a) = 0$;*

2. *For any* $n \in \mathbb{N}$, $X_n \in \{0, 1, \ldots, n\}$;

3. *There exists a positive and non-decreasing function* $\mu : \mathbb{N} \to \mathbb{N}$ *such that* $\mathbb{E}[X_n] \geq \mu(n)$ *for all* $n > a$.

Then we have

$$\mathbf{E}[T(n)] \leq \int_a^n \frac{1}{\mu(t)} \, \mathrm{d}t$$

The condition $T(a) = 0$ means that when the input size is equal to or below a, our algorithm can terminate without further recursive calls. So one can imagine that we initially stand at the point $n > a$ on the real line and walk towards the point a. The instantaneous velocity at the point t is X_t, who has a lower bound $\mu(t)$ in expectation. Therefore, if everything goes on as the expectation, the total time one costs to arrive at the point a from the point n should be upper bounded by a term like $\int_a^n \frac{1}{\mu(t)} \, \mathrm{d}t$. However, $\mu(t)$ is only a lower bound for the velocity X_t *in expectation*, and it is possible that $\mathbf{E}\left[\frac{1}{X_t}\right]$ is unbounded. So we cannot obtain the upper bound on time in the most straightforward way. Nevertheless, KUW inequality says that it does hold as we expect.

Proof of Theorem 3. We give a proof by induction on n. If $n = a$, then $\mathbf{E}[T(a)] = 0$ and the theorem trivially holds. So we let $n > a$ and assume the theorem is true for smaller n. Now we bound

$$\mathbf{E}[T(n)] = 1 + \mathbf{E}[T(n - X_n)].$$

The immediate idea is to apply the induction hypothesis to the term $\mathbf{E}[T(n - X_n)]$, but one needs to be careful about the possibility that $X_n = 0$. Let $q = \mathbf{Pr}[X_n \neq 0]$ and $1 - q = \mathbf{Pr}[X_n = 0]$. We distinguish between whether $X_n = 0$:

$$\mathbf{E}[T(n)] = 1 + \mathbf{E}[T(n - X_n)] = 1 + \mathbf{E}[T(n - X_n) \cdot \mathbf{1}[X_n \neq 0]] + \mathbf{E}[T(n - X_n) \cdot \mathbf{1}[X_n = 0]]. \tag{3}$$

Note that

$$\mathbf{E}[T(n - X_n) \cdot \mathbf{1}[X_n = 0]] = \mathbf{E}[T(n - X_n) \mid X_n = 0] \, \mathbf{Pr}[X_n = 0] = \mathbf{E}[T(n)] \cdot (1 - q) \tag{4}$$

Combining (3) and (4), we obtain

$$\mathbf{E}[T(n)] = \frac{1}{q} + \frac{1}{q} \cdot \mathbf{E}[T(n - X_n) \cdot \mathbf{1}[X_n \neq 0]] \tag{5}$$

By the tower rule of expectation, we have

$$\mathbf{E}[T(n - X_n) \cdot \mathbf{1}[X_n \neq 0]] = \mathbf{E}[\mathbf{E}[T(n - X_n) \cdot \mathbf{1}[X_n \neq 0] \mid X_n]].$$

We can apply the induction hypothesis to obtain

$$\mathbf{E}[T(n - X_n) \cdot \mathbf{1}[X_n \neq 0] \mid X_n] \leq \int_a^{n - X_n} \frac{\mathbf{1}[X_n \neq 0]}{\mu(t)} \, \mathrm{d}t. \tag{6}$$

Plugging (6) into (5), we obtain

$$\begin{aligned}
\mathbf{E}\left[T(n)\right] &\le \frac{1}{q} + \frac{1}{q} \cdot \mathbf{E}\left[\int_a^{n-X_n} \frac{\mathbf{1}[X_n \neq 0]}{\mu(t)}\, \mathrm{d}t\right] \\
&= \frac{1}{q} + \mathbf{E}\left[\int_a^{n-X_n} \frac{1}{\mu(t)}\, \mathrm{d}t \,\middle|\, X_n \neq 0\right] \\
&= \frac{1}{q} + \mathbf{E}\left[\int_a^{n} \frac{1}{\mu(t)}\, \mathrm{d}t \,\middle|\, X_n \neq 0\right] - \mathbf{E}\left[\int_{n-X_n}^{n} \frac{1}{\mu(t)}\, \mathrm{d}t \,\middle|\, X_n \neq 0\right] \\
&= \frac{1}{q} + \int_a^{n} \frac{1}{\mu(t)}\, \mathrm{d}t - \mathbf{E}\left[\int_{n-X_n}^{n} \frac{1}{\mu(t)}\, \mathrm{d}t \,\middle|\, X_n \neq 0\right]. \qquad (7)
\end{aligned}$$

Note that $\mu(n)$ is positive and non-decreasing, we have

$$\mathbf{E}\left[\int_{n-X_n}^{n} \frac{1}{\mu(t)}\, \mathrm{d}t \,\middle|\, X_n \neq 0\right] \ge \mathbf{E}\left[\int_{n-X_n}^{n} \frac{1}{\mu(n)}\, \mathrm{d}t \,\middle|\, X_n \neq 0\right] = \mathbf{E}\left[X_n \mid X_n \neq 0\right] \cdot \frac{1}{\mu(n)}. \qquad (8)$$

Since $\mathbf{E}[X_n] = \mathbf{E}[X_n \mid X_n \neq 0] \cdot q + \mathbf{E}[X_n \mid X_n = 0] \cdot (1-q)$, we have $\mathbf{E}[X_n \mid X_n \neq 0] = \frac{\mathbf{E}[X_n]}{q}$. Combining with (8), we obtain

$$\mathbf{E}\left[\int_{n-X_n}^{n} \frac{1}{\mu(t)}\, \mathrm{d}t \,\middle|\, X_n \neq 0\right] \ge \frac{\mathbf{E}[X_n]}{q \cdot \mu(n)} \ge \frac{1}{q}.$$

Plugging this into (7) proves the theorem.

□

In the following we apply the KUW inequality for the examples met today.

5.1 Mean of Geometric Variables

Assuming $X \sim \mathrm{Geom}(p)$, we use KUW to upper bound $\mathbf{E}[X]$. We use $T(1)$ to denote the number of tosses before seeing the first HEAD. Then clearly

$$T(1) = 1 + T(1 - X_1) \quad \text{for} \quad X_1 \sim \mathrm{Ber}(p).$$

Since $\mathbf{E}[X_1] = p$, we can pick $\mu(t) = p$. Then by KUW,

$$\mathbf{E}[T(1)] \le \int_0^1 \frac{1}{p}\, \mathrm{d}t = \frac{1}{p}.$$

5.2 Coupon Collector

We use $T(m)$ to denote the number of draws when exactly m types of coupons are not collected yet. Then

$$T(m) = 1 + T(m - X_m) \text{ for } X_m \sim \mathrm{Ber}(m/n)$$

Then for an integer m,

$$\mathbf{E}[X_m] = \frac{m}{n} = \frac{\lceil m \rceil}{n} \triangleq \mu(m).$$

Note that we use $\frac{\lceil m \rceil}{n}$ instead of $\frac{m}{n}$ here to avoid the divergence of the integral in KUW. It then follows that

$$\mathbf{E}\left[T(n)\right] \leq \int_{0^+}^{n} \frac{n}{\lceil t \rceil} \, \mathrm{d}t = n \cdot \sum_{i=1}^{n} \frac{1}{i} = nH(n).$$

We remark that we integrated from 0^+ here to avoid unnecessary discussions. This is fine by slightly tweaking the proof of KUW.

5.3 Quick Select

Let $T(n)$ denote the upper bound on the number of rounds (or equivalently the depth of the recursion) of our Quick Select algorithm when $|A| = n$. Then

$$T(n) = 1 + \max\left\{T(n - X_n), T(X_n - 1)\right\} \quad \text{where} \quad X_n \in_R \{1, 2, \dots, n\}.$$

On the other hand, we already showed that

$$\max\left\{T(n - X_n), T(X_n - 1)\right\} = T(n - Y_n)$$

for certain Y_n satisfying

$$\mathbf{E}\left[Y_n\right] \geq \frac{n}{4} \triangleq \mu(n).$$

So by KUW,

$$\mathbf{E}\left[T(n)\right] \leq \int_{0}^{n} \frac{4}{t} \, \mathrm{d}t = 4 \log n.$$

References

[1] R. M. Karp, E. Upfal, and A. Wigderson, *The complexity of parallel search*, Journal of Computer and System Sciences, 36 (1988), pp. 225–253. 4