CSC265 Homework 4 | Fall 2021
Name: Haojun Qiu | Student Number: 1006900622 |
Who I discussed with: Runshi Yang, Darcy Wang

Probability Space

Let $i \in \mathbb{N}^+$. Let S_i to be the set of valid operation performed at i-th operation. Define events

- A_i = the event that i-th operation is ACCESS = $\{i\text{-th operation be ACCESS}(S, i) : 1 \leq i \leq |S|\}$
- B_i = the event that i-th operation is PREPEND = $\{i\text{-th operation be PREPEND}(S, x) : x \text{ is valid input}\}$.

So, $S_i = A_i \cup B_i$, and A_i and B_i are mutually exclusive. And, notice that $P(A_i) = q$ and $P(B_i) = 1 - q$. Furthermore, elements in A_i are uniformly distributed (since index i is randomly chosen), so does elements in B. So, we have that for all operation $s_i \in A_i$, $P(\{s_i\}) = q/|S|$, where S is the input linked list right before i-th operation.

We can use S_i to define a sample space $S_1 \times S_2 \times \cdots \times S_n$, where each element is a sequence of n operations. For the probability distribution of this sample space, an element (a sequence of operation) $\langle s_1, \ldots, s_n \rangle \in S_1 \times \cdots \times S_n$ has probability being $P(\{s_1\}) \times \cdots \times P(\{s_n\})$ since each operation is chosen independent of others, and each $P(\{s_i\})$ is defined as above with set S_i.

Q1

Let random variable $X_i : S_i \to \mathbb{N}$ to be the indicator random variable of whether i-th operation is PREPEND, for all $i \in \mathbb{N}^+$. That is,

$$X_i(s_i) = \begin{cases} 1, & \text{if operation } s_i \text{ is PREPEND being performed at } i\text{-th operation, i.e., } s \in B_i \\ 0, & \text{if operation } s_i \text{ is not PREPEND being performed at } i\text{-th operation, i.e., } s \in A_i \end{cases}$$

Let random variable $Y_k : S_1 \times S_2 \times \cdots \times S_k \to \mathbb{N}$ to be the length of linked list after k operation, for all $k \in \{0, \ldots, n\}$. Then, since if i-th operation is PREPEND then length of S is incremented by 1 and in which case $X_i = 1$, and S initially has one element, we know that

$$Y_k = 1 + \sum_{i=1}^{k} X_i,$$

hence applying expectation to both side, we have

$$\begin{aligned}
E[Y_k] &= E\left[1 + \sum_{i=1}^{k} X_i\right] \\
&= 1 + \sum_{i=1}^{k} E[X_i] && \text{(by linearity of expectation)} \\
&= 1 + \sum_{i=1}^{k} P(i\text{-th operation performed is PREPEND}) \\
&= 1 + \sum_{i=1}^{k} (1 - q) && \text{(since } P(B_i) = 1 - q \text{ for all } i \in \{1, \ldots, k\}) \\
&= 1 + k(1 - q) = 1 + k - kq.
\end{aligned}$$

Hence, we conclude that for $0 \leq k \leq n$, the expected length of the linked list after k operations have been performed, would be $1 + k - kq$.

Q2

Define the random variable $Z_k : S_1 \times \cdots \times S_k \to \mathbb{N}$ to be the number of steps taken to perform k-th operation. Notice that the length of S after $k-1$ operation is at most k (when all $k-1$ operation are prepend), so Z_k takes value in $\{1, \ldots, k\}$.

So, by definition we have, $E[Z_k] = \sum_{z=1}^{k} zP(Z_k = z)$. Then, notice that A_k and B_k are two mutually exclusive events, and union of them is S_k, so we have $P(Z_k = z) = P(Z_k = z \text{ and } A_k) + P(Z_k = z \text{ and } B_k)$. So,

$$
\begin{aligned}
E[Z_k] = \sum_{z=1}^{k} zP(Z_k = z) &= \sum_{z=1}^{k} z\Big(P(Z_k = z \text{ and } A_k) + P(Z_k = z \text{ and } B_k)\Big) \\
&= \sum_{z=1}^{k} zP(Z_k = z \text{ and } B_k) + \sum_{z=1}^{k} zP(Z_k = z \text{ and } A_k) \\
&= (1 \times P(Z_k = 1 \text{ and } B_k)) + \sum_{z=2}^{k} zP(Z_k = z \text{ and } B_k) + \sum_{z=1}^{k} zP(Z_k = z \text{ and } A_k).
\end{aligned}
$$

Notice that $P(Z_k = 1 \text{ and } B_k) = P(Z_k = 1 | B_k)P(B_k) = 1(1-q) = 1-q$. And, since prepend takes 1 step only, for all $z > 1$, $P(Z_k = z \text{ and } B_k) = P(Z_k = z | B_k)P(B_k) = 0(1-q) = 0$, thus the second term above is 0. We have

$$
E[Z_k] = (1-q) + \sum_{z=1}^{k} zP(Z_k = z \text{ and } A_k).
$$

Notice that for Y_{k-1}, $k \geq 1$, length of S after $(k-1)$ operations takes values in $\{1, \ldots, k\}$ (equals k when all $k-1$ operations are PREPEND), so by the joint probability of two random variables,

$$
\begin{aligned}
E[Z_k] &= (1-q) + \sum_{z=1}^{k} z\left(\sum_{y=1}^{k} P(Z_k = z \text{ and } A_k \text{ and } Y_{k-1} = y)\right) \\
&= (1-q) + \sum_{y=1}^{k}\sum_{z=1}^{k} zP(Z_k = z \text{ and } A_k \text{ and } Y_{k-1} = y), \qquad \text{(by reorder terms of the summation)}
\end{aligned}
$$

and since $P(Z_k = z \text{ and } A_k \text{ and } Y_{k-1} = y) = 0$ when $z > y$ (steps taken for any access operation can not exceed current length), we can simplify the above as

$$
\begin{aligned}
E[Z_k] &= (1-q) + \sum_{y=1}^{k}\sum_{z=1}^{y} zP(Z_k = z \text{ and } A_k \text{ and } Y_{k-1} = y) \\
&= (1-q) + \sum_{y=1}^{k}\sum_{z=1}^{y} zP(Z_k = z \text{ and } A_k | Y_{k-1} = y)P(Y_{k-1} = y)
\end{aligned}
$$

Given that $|S| = y$ after $(k-1)$ operations, the probability that step taken equals to z with access operation, is the probability of operation ACCESS(S, z) being called in the k-the operation (in set S_k), as discussed before this is $\frac{q}{|S|} = \frac{q}{y}$). So, $P(Z_k = z \text{ and } A_k | Y_{k-1} = y) = \frac{q}{y}$, and thus

$$
\begin{aligned}
E[Z_k] &= (1-q) + \sum_{y=1}^{k}\sum_{z=1}^{y} z\frac{q}{y}P(Y_{k-1} = y) \\
&= (1-q) + \sum_{y=1}^{k}\left(\frac{q}{y}P(Y_{k-1} = y)\sum_{z=1}^{y} z\right) \\
&= (1-q) + \sum_{y=1}^{k}\left(\frac{q}{y}P(Y_{k-1} = y)\frac{y(y+1)}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= (1-q) + \frac{q}{2}\sum_{y=1}^{k}\Big(P(Y_{k-1}=y)(y+1)\Big) \\
&= (1-q) + \frac{q}{2}\Big(\sum_{y=1}^{k} yP(Y_{k-1}=y) + P(Y_{k-1}=y)\Big) \\
&= (1-q) + \frac{q}{2}\Big(\sum_{y=1}^{k} yP(Y_{k-1}=y) + \sum_{y=1}^{k} P(Y_{k-1}=y)\Big).
\end{aligned}
$$

Then, notice that since Y_{k-1} only takes values in $\{1,\dots,k\}$, by definition $\sum_{y=1}^{k} yP(Y_{k-1}=y) = E[Y_{k-1}]$, and $\sum_{y=1}^{k} P(Y_{k-1}=y) = 1$. Therefore,

$$
\begin{aligned}
E[Z_k] &= (1-q) + \frac{q}{2}\big(E[Y_{k-1}]+1\big) \\
&= (1-q) + \frac{q}{2}\big(1+(k-1)(1-q)+1\big) && \text{(by Q1)} \\
&= (1-q) + \frac{q}{2}\big((k-1)(1-q)+2\big) = 1 + \frac{q(k-1)(1-q)}{2},
\end{aligned}
$$

which is the number of expected number of steps taken to perform the k'th operation.

Q3

Define the random variable $W_n : S_1 \times \cdots \times S_n \to \mathbb{N}$ to be the number of steps taken to perform all n operations. Then, we know that $W_n = \sum_{k=1}^{n} Z_k$, the sum of number of steps taken to perform 1 to k operations, so

$$
\begin{aligned}
E[W_n] &= E\left[\sum_{k=1}^{n} Z_k\right] \\
&= \sum_{k=1}^{n} E[Z_k] && \text{(by linearity of expectation)} \\
&= \sum_{k=1}^{n}\left(1 + \frac{q(k-1)(1-q)}{2}\right) && \text{(by Q2)} \\
&= n + \sum_{k=1}^{n}\left(\frac{q(k-1)(1-q)}{2}\right) \\
&= n + \frac{q(1-q)}{2}\sum_{k=1}^{n}(k-1) \\
&= n + \frac{q(1-q)}{2}\left(\frac{n(n+1)}{2} - \frac{2n}{2}\right) \\
&= n + \frac{q(1-q)}{2}\left(\frac{n^2+n-2n}{2}\right) \\
&= n + \frac{q(1-q)(n^2-n)}{2},
\end{aligned}
$$

which is the expected number of steps taken to perform all n operations.