COMP 5711: Advanced Algorithms
Fall 2020 Final Exam Solution

Part 1

1. (a) True.

   (b) False.

   (c) False. Recall Q1(a) in the midterm exam. Let op_1 be increment and op_2 be decrement, then $c_1 = O(1)$ and $c_2 = O(1)$. However, the amortized cost is $\Omega(\log n)$ when the sequence consists of both op_1 and op_2.

2. (a) We need to prove $((u, b_1) \oplus (v, b_2)) \oplus (w, b_3) = (u, b_1) \oplus ((v, b_2) \oplus (w, b_3))$. When $b_3 = 1$, it is obvious that both sides are equal to $(w, 1)$. Now consider $b_3 = 0$. If $b_2 = 1$, then the left side is $(v, 1) \oplus (w, 0) = (v + w, 1)$ and the right side is $(u, b_1) \oplus (v + w, 1) = (v + w, 1)$. Now we consider $b_2 = 0$, then both sides are equal to $(u + v + w, b_1)$. Since we have verified all the cases, we conclude that $\oplus$ is associative.

   (b) We can simply use the prefix sum algorithm introduced in class, by simply changing the addition operation to $\oplus$. The algorithm is described as follows:

```
Segment-Prefix-Sum(A, M)
  if |A|=1 then return (A, M);
  for i=1 to |A|/2 in parallel do
    (B[i],N[i]) = (A[i*2-1],M[i*2-1])⊕(A[i*2],M[i*2]);
  (C, P) = Segment-Prefix-Sum(B, N);
  for i=1 to |A| in parallel do
    if i is even (D[i],Q[i])=(C[i/2],P[i/2]);
    else (D[i],Q[i])=(C[(i-1)/2],P[(i-1)/2])⊕(A[i],M[i]);
  return (D, Q);
```

   The work and time of this algorithm is the same with the original prefix sum algorithm, so it has $O(n)$ work and $O(\log n)$ time. Let $\{(B_i, N_i)\}$ be the output of running the above algorithm, then $(B_i, N_i) = \oplus_{j=1}^{i}(A_i, M_i)$. Now we prove $\{B_i\}$ correctly computes segment prefix sum by induction.
   When $i = 1$ this is obviously true. Suppose for $i \le k$ ($k \ge 1$), B_i is the segment prefix sum. Consider $i = k + 1$. If $M_{k+1} = 1$, then a new segment starts from the index $k+1$, so the true value of B_{k+1} should be A_{k+1}, which is correctly computed because $(B_{k+1}, N_{k+1}) = (B_k, N_k) \oplus (A_{k+1}, 1) = (A_{k+1}, 1)$. If $M_{k+1} = 0$, then A_k and A_{k+1} belong to the same segment, so the true value of B_{k+1} should be $B_k + A_{k+1}$, which is also correctly computed because $(B_{k+1}, N_{k+1}) = (B_k, N_k) \oplus (A_{k+1}, 0) = (B_k + A_{k+1}, N_k)$.

3. (a) Let $Y_{ij} = 1$ if $A_i = j$, and $Y_{ij} = 0$ otherwise. Since $\{A_i\}_{i=1}^n$ are mutually independent, we have $\{Y_{ij}\}_{i=1}^n$ are also mutually independent. Since $X_j = \sum_{i=1}^n Y_{ij}$, by Chernoff bound, we have

$$\Pr(X_j \ge \ln(n)) < \left(\frac{e}{\ln(n)}\right)^{\ln(n)} = \frac{n}{n^{\ln \ln n}}.$$

   By union bound, for any constant c,

$$\Pr(X \ge \ln(n)) = \Pr(X_j \ge \ln(n) \text{ for some } j) \le \sum_{j=1}^{n} \Pr(X_j \ge \ln(n)) = \frac{n^2}{n^{\ln \ln n}} = o(\frac{1}{n^c}).$$

(b) We define Y_{ij} the same as above. Now Y_{ij} is just pairwise independent, so we use Chebyshev inequality instead. Since Y_{ij} is a Bernoulli variable with parameter $1/n$, we have $\mathbb{E}[Y_{ij}] = 1/n$ and $\mathrm{Var}(Y_{ij}) = 1/n - 1/n^2$, and hence $\mathbb{E}[X_j] = 1$ and $\mathrm{Var}(X_j) = 1 - 1/n$. Therefore, when $n \geq 2$,

$$\Pr(X_j \geq \sqrt{2n} + 1) = \Pr\left(X_j \geq \mathbb{E}[X_j] + \frac{\sqrt{2n}}{\sqrt{1 - 1/n}}\sqrt{\mathrm{Var}(X_j)}\right) \leq \frac{1}{2n}.$$

Similarly, by union bound, $\Pr(X \geq \sqrt{2n} + 1) \leq n \cdot 1/(2n) = 1/2$.

(c) It is direct that $\Pr(A_n = j) = 1/n$. For simplicity, we define the operator $\hat{+}$, where $a \mathbin{\hat{+}} b = a + b$ if $a + b \leq n$, and $a \mathbin{\hat{+}} b = a + b - n$ otherwise. For any $i \in [n-1], j \in [n]$,

$$\begin{aligned} \Pr(A_i = j) &= \Pr(A_i = j \mid K = j) \cdot \Pr(K = j) \\ &\quad + \sum_{k \neq j} \Pr(K = k) \cdot \Pr(\pi(i) \mathbin{\hat{+}} k = j) \cdot \Pr(A_i = k + \pi(i) \mid K = k, \pi(i) \mathbin{\hat{+}} K = j) \\ &= \frac{1}{\sqrt{n}} \cdot \frac{1}{n} + \frac{n-1}{n} \cdot \frac{1}{n-1}(1 - \frac{1}{\sqrt{n}}) = \frac{1}{n}. \end{aligned}$$

(d) Given K, $\mathbb{E}[X_K \mid K] = \sum_{i=1}^n \Pr(A_i = K \mid K) = (n-1)/\sqrt{n} + 1/n \geq 0.9\sqrt{n}$ when n is large enough, and

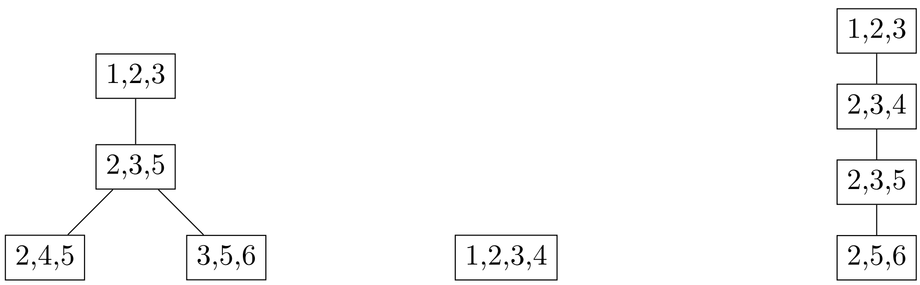
$$\mathrm{Var}(X_K \mid K) = (n-1) \cdot \frac{1}{\sqrt{n}}(1 - \frac{1}{\sqrt{n}}) + \frac{1}{n}(1 - \frac{1}{n}) < \sqrt{n}.$$

By Chebyshev inequality, $\Pr(X_K \leq \sqrt{n}/2 \mid K) \leq \frac{1}{0.4^2\sqrt{n}} = o(1)$. Therefore,

$$\begin{aligned} \Pr(X \leq \sqrt{n}/2) &= \sum_{k=1}^n \Pr(K = k) \Pr(X \leq \sqrt{n}/2 \mid K = k) \\ &\leq \sum_{k=1}^n \Pr(K = k) \Pr(X_K \leq \sqrt{n}/2 \mid K = k) = o(1). \end{aligned}$$

Part 2

1. The tree decompositions with minimal tree width are given below.



(a) (b) (c)

2. For any u, v on the unit circle, let $0 \leq \theta \leq \pi$ be the angle between them. The two vectors fall into the different quadrants if and only if one of the axes falls in between them after rotation. Thus we have

$$\Pr(h(u) = h(v)) = \begin{cases} 1 - 2\theta/\pi, & \text{if } \theta \leq \pi/2; \\ 0, & \text{if } \theta > \pi/2. \end{cases}$$

By $d(u,v) = 2\sin(\theta/2)$, we have $\theta = 2\sin^{-1}(d(u,v)/2)$. Plugging in the given values of r and c, we obtain $p_1 = 1/3$, $p_2 = 0$.

If α is chosen uniformly at random from $[0, 2\pi/3]$, the collision probability doesn't only depend on $d(u,v)$, but also the positions of u, v, and we should consider the worst case. For $d(u,v) = 1$, a worst case is $u = (0,1), v = (\cos(\pi/6), \sin(\pi/6))$. In this case, they collide when $\alpha \in [\pi/3, \pi/2]$. So the collision probability is $p_1 = (\pi/6)/(2\pi/3) = 1/4$. When $d(u,v) = cr = \sqrt{2}$, they still never collide, so $p_2 = 0$.

3. (a) Let $s_1, \cdots, s_n$ be all the different elements in the stream. We compute

$$\begin{aligned}\Pr(X > t) &= \Pr(h(s_i) > t \text{ for } i \in [n]) + \sum_{i=1}^{n} \Pr(h(s_i) \le t) \cdot \Pr(h(s_j) > t \text{ for all } j \ne i) \\ &= (1-t)^n + nt(1-t)^{n-1}.\end{aligned}$$

Therefore $F_X(t) = 1 - \Pr(X > t) = 1 - (1-t)^n - nt(1-t)^{n-1}$ and $f_X = n(n-1)t(1-t)^{n-2}$.

(b) We compute

$$\mathbb{E}[Y] = \mathbb{E}[1/X] = \int_0^1 \frac{f_X(t)}{t} \mathrm{d}t = n \int_0^1 (n-1)(1-t)^{n-2} \mathrm{d}t = n,$$

which means Y is an unbiased estimator.

(c)

$$\begin{aligned}\Pr[|Y - n| > c_1 n] &= \Pr[|1/X - n| > c_1 n] \\ &= \Pr[1/X < (1-c_1)n] + \Pr[1/X > (1+c_1)n] \\ &= \Pr[X > \frac{1}{(1-c_1)n}] + \Pr[X < \frac{1}{(1+c_1)n}].(*)\end{aligned}$$

For the second term in $(*)$, define $p = \frac{1}{(1+c_1)n}$. Let $Z_i = 1$ if the hash value of the i-th distinct element is smaller than p. Then $\Pr[Z_i = 1] = p$ and $\{Z_i\}_{i=1}^n$ are pairwise independent. Let $Z = \sum_{i=1}^n Z_i$, then $\Pr(X < p) = \Pr(Z \ge 2)$. Since $\mathbb{E}[Z] = pn$ and $\mathrm{Var}(Z) = p(1-p)n$, by Chebyshev inequality,

$$\begin{aligned}\Pr(Z \ge 2) &= \Pr\left(Z - \mathbb{E}[Z] \ge \frac{2 - pn}{\sqrt{p(1-p)n}} \cdot \sqrt{\mathrm{Var}(Z)}\right) \\ &\le \frac{p(1-p)n}{(2-pn)^2} \le pn = 1/(1+c_1),\end{aligned}$$

where the last inequality holds because $pn < 1$. Similarly, for the first term in $(*)$, we set $p = \frac{1}{(1-c_1)n}$. Then,

$$\begin{aligned}\Pr(X > p) = \Pr(Z \le 1) &= \Pr\left(Z - \mathbb{E}[Z] \le -\frac{pn - 1}{\sqrt{p(1-p)n}} \cdot \sqrt{\mathrm{Var}(Z)}\right) \\ &\le \frac{p(1-p)n}{(pn-1)^2} \le \frac{pn}{(pn-1)^2} = (1-c_1)/c_1^2.\end{aligned}$$

So we can set $c_2 = \frac{1}{1+c_1} + \frac{1-c_1}{c_1^2}$. Note that as c_1 goes to 1, the first term goes to 1/2 and the second term goes to 0. So such constants $0 < c_1, c_2 < 1$ must exist. For example, we may take $c_1 = 0.9$ and $c_2 = 0.65$.