Randomized Algorithms (600.464/664) - Assignment 2: Solutions

P.C. Shyamshankar

Due: September 20

1. (a) Here, $X, Y \in \{0, 1, 2\}$. The joint distribution of $X$ and $Y$ is:

<table>
<thead>
<tr>
<th>$X \setminus Y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{8}$</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{16}$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(b) $X$ and $Y$ are not independent, since:

$$P[X = 2, Y = 2] = 0 \neq \frac{1}{16} \cdot \frac{1}{16} = P[X = 2] \cdot P[Y = 2]$$ (2)

(c)

$$E[X + Y] = E[X] + E[Y]$$ (3)

$$= \sum_{x=0}^{2} xP[X = x] + \sum_{y=0}^{2} yP[Y = y]$$ (4)

$$= \left( \frac{3}{8} + \frac{1}{8} \right) + \left( \frac{3}{8} + \frac{1}{8} \right)$$ (5)

$$= 1$$ (6)

$$E[XY] = \sum_{x=0}^{2} \sum_{y=0}^{2} x \cdot y \cdot P[X = x, Y = y]$$ (7)

$$= \frac{1}{8}$$ (8)

$$E[e^{X + Y}] = E[e^{X}] + E[Y]$$ (9)

$$= \sum_{x=0}^{2} e^{x} P[X = x] + E[Y]$$ (10)

$$\approx 2.044 + 0.5$$ (11)

$$= 2.544$$ (12)
\[ P [X \neq Y] = 1 - P [X = Y] \quad (13) \]
\[ = 1 - \frac{1}{4} - \frac{1}{8} \quad (14) \]
\[ = \frac{5}{8} \quad (15) \]

\[ E [X|Y] \] is a function which maps possible values of \( Y \) to the conditional expectation \( E [X|Y = y] \).

\[ E [X|Y = 0] = \sum_{x=0}^{2} x \cdot P [X = x|Y = 0] \quad (16) \]
\[ = \left( 0 + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{16} \right) \cdot \frac{16}{9} \quad (17) \]
\[ = \frac{2}{3} \quad (18) \]

\[ E [X|Y = 1] = \sum_{x=0}^{2} x \cdot P [X = x|Y = 1] \quad (19) \]
\[ = \left( 0 + 1 \cdot \frac{1}{8} + 0 \right) \cdot \frac{8}{3} \quad (20) \]
\[ = \frac{1}{3} \quad (21) \]

\[ E [X|Y = 2] = \sum_{x=0}^{2} x \cdot P [X = x|Y = 2] \quad (22) \]
\[ = (0 + 0 + 0) \cdot 16 \quad (23) \]
\[ = 0 \quad (24) \]

2. (a) (Problem Set 3, Question 8)

\( X \) is a binomial random variable, with parameters \( 2n \) and \( 1/2 \). Therefore, \( \mu_X = n, \sigma_X = \sqrt{n/2} \).

Now,

\[ P [|X - n| < \sqrt{n}] = 1 - P [|X - n| \geq \sqrt{n}] \quad (25) \]
\[ P [|X - n| \geq \sqrt{n}] = P [|X - \mu_X| \geq \sqrt{n}] \quad (26) \]

Using Chebyshev’s Inequality,

\[ P [|X - n| \geq \sqrt{n}] \leq \frac{\sigma_X^2}{\sqrt{n^2}} \quad (27) \]
\[ \leq \frac{n}{\sqrt{2n}} \quad (28) \]
\[ = \frac{n}{2n} \quad (29) \]
\[ = \frac{1}{2} \quad (30) \]
And therefore,

\[ P [ |X - n| < \sqrt{n}] \geq 1 - \frac{1}{2} = \frac{1}{2} \]  \quad (31)

By the definition of the binomial distribution,

\[ P [X = k] = \binom{2n}{k} \frac{1}{2^k} \left(1 - \frac{1}{2}\right)^{2n-k} \]  \quad (32)

\[ \frac{1}{2^{2n}} \]  \quad (33)

\[ P [|X - n| < \sqrt{n}] = P [n - \sqrt{n} < X < n + \sqrt{n}] \]  \quad (34)

\[ = \sum_{i=n-\sqrt{n}}^{n+\sqrt{n}} \binom{2n}{i} \frac{1}{2^{2n}} \]  \quad (35)

\[ = \frac{1}{2^{2n}} \sum_{i=n-\sqrt{n}}^{n+\sqrt{n}} \binom{2n}{i} \]  \quad (36)

Since there are at most \(2\sqrt{n} + 1\) terms in the sum, each of which is bounded above by the largest binomial coefficient, namely \(\binom{2n}{n}\), we can approximate it to:

\[ P [|X - n| < \sqrt{n}] \leq \frac{1}{2^{2n}} (2\sqrt{n} + 1) \binom{2n}{n} \]  \quad (37)

Combining (31) and (37), we get:

\[ \frac{1}{2^{2n}} (2\sqrt{n} + 1) \binom{2n}{n} \geq \frac{1}{2} \]  \quad (38)

\[ \binom{2n}{n} \geq \frac{2^{2n}}{4\sqrt{n} + 2} \]  \quad (39)

(b) (Problem Set 3, Question 19)

Consider a single path of length \(\ln n\). Let \(X_i\) be a random variable defined as 1 if ball \(i\) falls in a bin on the path, and 0 otherwise. Here, \(P [X_i = 1] = \frac{\ln n/\alpha^2}{n} = E [X_i]\). Then \(X = X_1 + X_2 + \cdots + X_n^2\) is the random variable denoting the number of balls that fall in bins on the given path. We can then compute \(E [X] = n^2 \cdot \ln n/\alpha^2 = \ln n\).

Applying the Chernoff Bound,

\[ P [X \geq c \ln n] = P [X \geq \ln n (1 + \delta)] \]  \quad (40)

\[ \leq \exp - \frac{\ln n \cdot \delta \cdot \ln \delta}{2} \]  \quad (41)

\[ = n^{-2\ln a/2} \]  \quad (42)
If we require this probability to be at most $n^{-k}$ for some $k$, then

$$n^{-\frac{4 \ln \delta}{2k}} \leq n^{-k} \quad (43)$$

$$\delta \ln \delta \geq 2k \quad (44)$$

$$\delta \geq \frac{2k}{\ln 2k} \quad (45)$$

We can upper-bound the total number of monotone connected paths to $n^3$, since there are $n^2$ possible starting points, and at most $2^{\ln n} \approx n$ possible paths starting at each point. In this case, in order to achieve the result of no path containing more than $c \ln n$ balls with high probability, we require:

$$P[X \geq c \ln n] \cdot n^3 \leq \frac{1}{n} \quad (46)$$

$$P[X \geq c \ln n] \leq \frac{1}{n^4} \quad (47)$$

Setting $k = 4$, we get $\delta = \frac{8}{\ln 8}$, with $c = 1 + \delta$.

The only difference in the case of non-monotone connected paths is that there are now at most $n^2 \cdot 4^{\ln n} \approx n^4$ such paths, in which case $k = 5$, $\delta = 10/\ln 10$ and $c = 1 + \delta$.

In both cases, a constant $c$ exists such that with high probability, no path of length $\ln n$ contains more than $c \ln n$ balls.

(c) (Problem Set 3, Question 23a)

Let $X_1, X_2, \cdots, X_n$ be uniformly distributed $\{0, 1\}$ random variables, and let $X = X_1 + X_2 + \cdots + X_n$. Here, $E[X_i] = 1/2$, and $E[X] = n/2$.

By applying the Chernoff Bound, we get:

$$P\left[X \geq \frac{3n}{4}\right] = P\left[X \geq \frac{n}{2} \left(1 + \frac{1}{2}\right)\right] \quad (48)$$

$$\leq e^{-n/2(1/2)^2} \quad (49)$$

$$= e^{-n/8} \quad (50)$$

Alternatively,

$$P\left[X \geq \frac{3n}{4}\right] = \sum_{j=3n/4}^{n} P[X = j] \quad (51)$$

$$= \sum_{j=3n/4}^{n} \binom{n}{j} \frac{1}{2^n} \quad (52)$$

$$= \frac{1}{2^n} \sum_{j=3n/4}^{n} \binom{n}{j} \quad (53)$$
Combining these results, we get

\[
\frac{1}{2^n} \sum_{j=3n/4}^{n} \binom{n}{j} \leq e^{-n/8} \quad (54)
\]

\[
\sum_{j=3n/4}^{n} \binom{n}{j} \leq 2^n \cdot e^{-n/8} \quad (55)
\]