CS5314 Randomized Algorithms

Homework 1 (Suggested solution)

1. **Ans.** The probability that the number of required tosses is odd is:

$$p + (1-p)^2 p + (1-p)^4 p + \ldots = \sum_{i=0}^{\infty} (1-p)^{2i} p = \frac{p}{1-(1-p)^2} = \frac{1}{2-p}.$$

2. **Ans.** Let W and B denote the disjoint events that:

$$W = \text{“the transferred ball from Box 1 to Box 2 is white”}, \quad \text{and}$$

$$B = \text{“the transferred ball from Box 1 to Box 2 is black”}.$$

Then, $\Pr(W) + \Pr(B) = 1$, and

$$\Pr(W) = \frac{a}{a+b} \quad \text{and} \quad \Pr(B) = \frac{b}{a+b}.$$

Let A be the desired event that “the next ball drawn from Box 2 is white”. Hence,

$$\begin{aligned} \Pr(A) &= \Pr\big(A \cap (W \cup B)\big) \\ &= \Pr\big((A \cap W) \cup (A \cap B)\big) \\ &= \Pr(A \cap W) + \Pr(A \cap B) \end{aligned}$$

Since

$$\Pr(A \mid W) = \frac{c+1}{c+d+1} \quad \text{and} \quad \Pr(A \mid B) = \frac{c}{c+d+1} \quad ,$$

we have

$$\Pr(A) = \frac{a(c+1)}{(a+b)(c+d+1)} + \frac{bc}{(a+b)(c+d+1)} = \frac{ac+bc+a}{(a+b)(c+d+1)}.$$

3. (a) **Ans.** From the requirement, we know that

$$\left(\frac{a}{a+b}\right)\left(\frac{a-1}{a+b-1}\right) = \frac{1}{3}.$$

On the other hand, from algebra, we know that

$$\begin{aligned} & \frac{a}{a+b} > \frac{a-1}{a+b-1} \qquad \text{for } a, b \geq 1 \\ \Rightarrow \quad & \left(\frac{a}{a+b}\right)^2 > \left(\frac{a}{a+b}\right)\left(\frac{a-1}{a+b-1}\right) > \left(\frac{a-1}{a+b-1}\right)^2. \end{aligned}$$

So, combining both statements gives:

$$\left(\frac{a}{a+b}\right)^2 > \frac{1}{3} > \left(\frac{a-1}{a+b-1}\right)^2.$$

(b) **Ans.** From the left inequality of part (a), that is

$$\left(\frac{a}{a+b}\right)^2 > \frac{1}{3}$$

we can easily obtain

$$\frac{(\sqrt{3}+1)b}{2} < a.$$

Similarly, from the right inequality of part (a), that is

$$\frac{1}{3} > \left(\frac{a-1}{a+b-1}\right)^2$$

we can easily obtain

$$a < 1 + \frac{(\sqrt{3}+1)b}{2}.$$

Combining these two gives the desired result.

(c) **Ans.** Let W_i denote the event that the ith ball drawn is white. From the requirement, we know that there must be at least one black balls; otherwise,

$$\Pr(W_1 \cap W_2) = 1 \neq \frac{1}{3}.$$

If there are only one black ball (so that $b = 1$), we have $1.36 < a < 2.36$ from part (b). Thus there must be two white balls. By checking

$$\Pr(W_1 \cap W_2) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3},$$

we conclude that one black two whites can yield the desired probability. Thus, the smallest number of balls is 3.

(d) **Ans.** If $b = 2$ (so that b is set to minimum possible value), a would be 3, but

$$\Pr(W_1 \cap W_2) = \frac{3}{10}$$

which is not correct.

If $b = 4$ (so that b is set to the next minimum possible value), a would be 6, and

$$\Pr(W_1 \cap W_2) = \frac{1}{3}.$$

We conclude that 4 blacks and 6 whites can yield the desired probability, and the smallest number of balls (when b is even) is 10.

4. **Ans.** Let X denote the number of inversions in an array. Let X_{ij} be the indicator for the pair (i, j) being an inversion, where $i < j$. Then we have:

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i<j} X_{ij}\right] = \sum_{i<j} \mathrm{E}[X_{ij}] = \sum_{i<j} \frac{1}{2} = \frac{n(n-1)}{4}.$$

5. **Ans.** Let X be the random variable which counts the number of pairs which are coupled. Let X_i be an indicator such that

$$\begin{aligned} \Pr(X_i) &= 1 \quad \text{if the } i\text{th pair are coupled} \\ \Pr(X_i) &= 0 \quad \text{otherwise} \end{aligned}$$

Then, $X = X_1 + X_2 + \cdots + X_{20}$, and $\mathrm{E}[X_i] = 1/20$. By linearity of expectation,

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{20} X_i\right] = \sum_{i=1}^{20} \mathrm{E}[X_i] = \sum_{i=1}^{20} \frac{1}{20} = 1.$$

6. (a) **Ans.** $\min(X, Y)$ is equivalent to the number of times we need to perform two experiments together, each with success probability p and q, in order to obtain a success from at least one of these experiments. In other words, $\min(X, Y)$ is a geometric random variable with parameter $1-(1-p)(1-q)$, i.e.,

$$\Pr(\min(X, Y) = k) = ((1-p)(1-q))^{k-1}\,(1-(1-p)(1-q))$$

(b) **Ans.** We will give two approaches to solve this problem.

i. By definition,

$$\begin{aligned} \mathrm{E}[X \mid X \le Y] &= \sum x \Pr(X = x \mid X \le Y) \\ &= \sum x \frac{\Pr(X = x \cap X \le Y)}{\Pr(X \le Y)} \\ &= \frac{1}{\Pr(X \le Y)} \sum x \Pr(X = x)\Pr(Y \ge x). \end{aligned}$$

Now, we calculate $\Pr(X \le Y)$ at first.

$$\begin{aligned} \Pr(X \le Y) &= \sum_i \Pr(X = i \cap Y \ge i) \\ &= \sum_i \Pr(X = i)\Pr(Y \ge i) \\ &= \sum_i (1-p)^{i-1} p (1-q)^{i-1} \\ &= \frac{p}{p+q-pq}. \end{aligned}$$

Next, we have:

$$\begin{aligned} \sum x \Pr(X = x)\Pr(Y \ge x) &= \sum x(1-p)^{x-1} p \cdot \Pr(Y \ge x) \\ &= \sum x(1-p)^{x-1} p (1-q)^{x-1} \\ &= p \sum x((1-p)(1-q))^{x-1} \\ &= \frac{p}{(1-(1-p)(1-q))^2} = \frac{p}{(p+q-pq)^2}. \end{aligned}$$

Combining the above results, we have

$$\mathrm{E}[X \mid X \le Y] = \frac{1}{p+q-pq}.$$

ii. By memoryless property.

At first, we define an indicator Z, such that

$$Z = \begin{cases} 1, & \text{if first trial of first coin succeeds;} \\ 0, & \text{otherwise} \end{cases}$$

Whenever $Z = 0$ (i.e., the first trial of coin fails), we use X^* to denote the number of further trials to have the first coin succeeds. Thus, whenever $Z = 0$, $X = 1 + X^*$, and X^* is a geometric random variable with parameter p by the memoryless property.

$$\begin{aligned} \mathrm{E}[X \mid X \leq Y] &= 1 \cdot \Pr(Z = 1 \mid X \leq Y) + \mathrm{E}[1 + X^* \mid X \leq Y] \Pr(Z = 0 \mid X \leq Y) \\ &= \Pr(Z = 1 \mid X \leq Y) + (1 + \mathrm{E}[X \mid X \leq Y]) \Pr(Z = 0 \mid X \leq Y) \\ &= 1 + \mathrm{E}[X \mid X \leq Y] \Pr(Z = 0 \mid X \leq Y). \end{aligned}$$

Note that $\Pr(Z = 0 \mid X \leq Y) = \Pr(Z = 0 \cap X \leq Y)/\Pr(X \leq Y)$, and the cases that correspond to the event $(X \leq Y)$ are the followings:

$$\begin{array}{cccc} (X = 1, Y = 1), & (X = 1, Y = 2), & (X = 1, Y = 3), & \cdots, \\ (X = 2, Y = 2), & (X = 2, Y = 3), & (X = 2, Y = 4), & \cdots, \\ & \vdots & & \\ (X = k, Y = k), & (X = k, Y = k+1), & (X = k, Y = k+2), & \cdots, \\ & \vdots & & \end{array}$$

whereas the cases for the event $(Z = 0 \cap X \leq Y)$ are obtained from the above by removing those cases with $X = 1$. Hence,

$$\Pr(Z = 0 \cap X \leq Y) = \Pr(X \leq Y) - \sum_i \Pr(X = 1 \cap Y = i) = \Pr(X \leq Y) - p.$$

Hence we have

$$\begin{aligned} \mathrm{E}[X \,|\, X \leq Y] &= 1 + \mathrm{E}[X \,|\, X \leq Y] \left(1 - \frac{p}{\Pr(X \leq Y)}\right) \\ \mathrm{E}[X \,|\, X \leq Y] &= \frac{\Pr(X \leq Y)}{p} = \frac{1}{p + q - pq} \end{aligned}$$

(c) **Ans.** The probability $\Pr(X = Y)$ can be calculated as follows:

$$\Pr(X = Y) = \sum_{k=1}^{\infty} \Pr(X = Y = k) = \sum_{k=1}^{\infty} (1 - p)^{k-1} p (1 - q)^{k-1} q = \frac{pq}{p + q - pq}.$$

(d) **Ans.** Since $\max(X, Y) + \min(X, Y) = X + Y$, we have:

$$\mathrm{E}[\max(X, Y)] = \mathrm{E}[X] + \mathrm{E}[Y] - \mathrm{E}[\min(X, Y)] = \frac{1}{p} + \frac{1}{q} - \frac{1}{p + q - pq}.$$

7. (a) **Ans.** If the ith candidate is the best of all candidates, we will choose him if and only if `(1)` $i > m$ and `(2)` the best of the first $i-1$ candidates (say y) is among the first m candidates (otherwise, if y is not among the first m candidates, we will choose y by our interview strategy).

Let B_i denote the event that ith candidate is the best, and Y_i denote the event that the best of first $i-1$ candidates is among the first m candidates. Then, we have

$$\Pr(E_i) = \begin{cases} 0 & \text{for } i \le m \\ \Pr(B_i \cap Y_i) & \text{for } i > m \end{cases}$$

The term $\Pr(B_i \cap Y_i)$ is equal to

$$\Pr(B_i)\Pr(Y_i \mid B_i) = \frac{1}{n} \cdot \frac{m}{i-1} = \frac{m}{n} \cdot \frac{1}{i-1}.$$

From our definition, we can see that $\Pr(E) = \sum_{i=1}^{n} \Pr(E_i)$. Thus, we have

$$\Pr(E) = \sum_{i=1}^{n} \Pr(E_i) = \sum_{i=m+1}^{n} \Pr(E_i) = \frac{m}{n} \sum_{i=m+1}^{n} \frac{1}{i-1}.$$

(b) **Ans.** Consider the curve $f(x) = 1/x$. The area under the curve from $x = j-1$ to $x = j$ is less than $1/(j-1)$. Thus,

$$\sum_{j=m+1}^{n} \frac{1}{j-1} \ge \int_{m}^{n} f(x)dx = \log_e n - \log_e m.$$

Similarly, the area under the curve from $x = j-2$ to $x = j-1$ is greater than $1/(j-1)$. Thus,

$$\sum_{j=m+1}^{n} \frac{1}{j-1} \le \int_{m-1}^{n-1} f(x)dx = \log_e (n-1) - \log_e (m-1).$$

Combining these two inequalities with part (a) gives the desired result.

(c) **Ans.** Let $g(m) = m(\log_e n - \log_e m)/n$. By differentiating $g(m)$, we get

$$g'(m) = \frac{\log_e n - \log_e m}{n} - \frac{1}{n},$$

which is 0 when $m = n/e$. Also, if we differentiate $g'(m)$, we get

$$g''(m) = \frac{-1}{mn} < 0,$$

which indicates that $g(m)$ attains maximum when $m = n/e$.

By substituting $m = n/e$ in the inequality of part(b), we get

$$\Pr(E) \ge \frac{m(\log_e n - \log_e m)}{n} = \frac{n\,(\log_e n - \log_e (n/e))}{ne} = \frac{n\,(\log_e e)}{ne} = \frac{1}{e}.$$