Assignment 2: CS 203 (Spring 2020)

Nidhi Hegde (180472)

1. (5+15=20 marks) **Independent Geometric Random Variables**

   A *geometric random variable* counts the number of tosses until you get a head (as defined in notes). Let Y and Z be two independent, geometric random variables with parameter p.

   (a) Interpret the expression $\Pr(Y = i \mid Y + Z = n)$ in terms of tossing only one coin.

   (b) Show that $\Pr(Y = i \mid Y + Z = n) = \frac{1}{n-1}$ for $i = 1, \ldots, n-1$.

   Solution:

   (a) We know that probability of obtaining head in any 2 tosses of a single coin is independent of each other. In terms of tossing only 1 coin, we define Y to be the event that we get the first head on the i^{th} toss if we start considering the outcomes 1^{st} toss onwards, while we can define Z to be the event that we get the first head on the n^{th} toss of the coin, given we start considering the outcomes from $i+1^{th}$ toss. Thus, the above expression is equivalent to the event where we obtain the first head at i^{th} toss, and then the next head at the n^{th} toss, while tossing a single coin n times.

   (b) By the definition of conditional probability, we can say that:

$$\Pr(Y = i \mid Y + Z = n) = \frac{\Pr((Y = i) \cap (Y + Z = n))}{\Pr(Y + Z = n)} \implies \frac{\Pr((Y = i) \cap (Z = n - i))}{\Pr(Y + Z = n)} \tag{1}$$

   We also know that Y and Z are independent. Hence, by definition of independence,

$$\Pr((Y = i) \cap (Z = n - i)) \implies \Pr(Y = i) \cdot \Pr((Z = n - i)$$

   $Y = i \;\; (\forall i \in [1...n-1])$ is the event that we get tails in all the tosses before the i^{th} toss and get the first head exactly at the i^{th} toss. And getting a tail or head in any toss has a probability $1-p$ & p respectively. So, we get:

$$\Pr(Y = i) = (1-p)^{i-1}p \quad and \quad \Pr(Z = n - i) = (1-p)^{n-i-1}p$$
$$\implies \Pr((Y = i) \cap (Z = n - i)) = (1-p)^{n-2} \cdot p^2$$

   Now, the event $\Pr(Y + Z = n)$ can happen in any one of the following ways:
   $(Y = 1,\ Z = n-1)$ *or* $(Y = 2,\ Z = n-2)$ ... *or* $(Y = n-1,\ Z = 1)$.
   So, we get our denominator as:

$$\Pr(Y + Z = n) = \sum_{i=1}^{n-1} \Pr(Y = i) \cdot \Pr((Z = n - i)$$
$$\implies (n-1) \cdot \left((1-p)^{n-2} \cdot p^2\right)$$

   Putting values in equation [1], we get:

$$\Pr(Y = i \mid Y + Z = n) = \frac{(1-p)^{n-2} \cdot p^2}{(n-1) \cdot (1-p)^{n-2} \cdot p^2} \quad = \quad \frac{1}{n-1}$$

□

2. (10+13+5+7=35 marks) **Verifying Matrix Multiplication**

Given three $n \times n$ matrices A, B and C; how fast can we test whether $AB = C$? An obvious answer is to multiply A and B and compare the resulting matrix with C which currently requires $O(n^{2.3728})$ multiplications [1]. We can use a faster method inspired by probabilistic techniques to test $AB = C$ as follows:

1 Pick $x_1, \ldots, x_n \in \{0, 1\}$ randomly, uniformly and independently. Let $\bar{x} = (x_1, \ldots, x_n)$.

2 Test $A(B\bar{x}) = C\bar{x}$? If they match then return **Yes** otherwise **No**.

The above algorithm only requires $O(n^2)$ multiplications. Let us try to prove that the probability of error is 'small'.

(a) Let q be a rational number. Pick a boolean value $u \in \{0, 1\}$ randomly uniformly. Show that $\Pr_u(u = q) \leq \frac{1}{2}$.

(b) Let $D = (d_{ij})$ be a $n \times n$ matrix with the ith row as D_i. If $D_i \neq \bar{0}$, show that $\Pr_x(D_i\bar{x} = 0) \leq \frac{1}{2}$.

(c) Assume $AB \neq C$. Let $D = AB - C$. Show that the error probability $\Pr_x(D\bar{x} = \bar{0}) \leq \frac{1}{2}$.

(d) How will you change the algorithm to improve its error probability to 2^{-100}? How much overhead does this cause? Give the best possible estimate.

Solution:

(a) Since u is picked randomly uniformly out of 0 or 1, $P(u = 0) = P(u = 1) = \frac{1}{2}$.
Now, given q, the condition $u = q$ can hold only when q has a value in the range of values taken by u. Thus, if q is chosen to be any value other than 0 or 1, $\Pr(u = q) = 0$ by default.
Else, if $q = 1$, then $\Pr_u(u = q) = \Pr_u(u = 1) = \frac{1}{2}$. Similarly, if $q = 0$, then we can say that $\Pr_u(u = q) = \Pr_u(u = 0) = \frac{1}{2}$. Thus, in any case, irrespective of whatever we choose the value of q to be, given a fixed q, $\Pr(u = q)$ is always bounded by $\frac{1}{2}$. Hence, we prove that $\Pr_u(u = q) \leq \frac{1}{2}$.

(b) Since it is given that $D_i \neq \bar{0}$, at least 1 element of D_i must be non-zero.
(Just to be clear, we assume that all the elements of D are already given and fixed, and therefore, we consider randomization only over the elements of $\bar{x}$).
To proceed with the proof, we assume that the k^{th} element of D_i is non-zero. We can say that:

$$S = D_i\bar{x} = \sum_{j=1}^{n} d_{ij}x_j = d_{i1}x_1 + d_{i2}x_2 + ...d_{ik}x_k + ...d_{in}x_n = d_{ik}x_k + S' \tag{2}$$

Now, we separate into 2 cases depending on whether S' is 0 or non-zero. In both the cases, we calculate the probability that $Pr(S = 0)$

- **Case 1: S' is zero** To make $S = 0$ in this case, the contribution of the k^{th} term should be zero. This is only possible if $x_k = 0$ since it is given that $d_{ik} \neq 0$. Now, as x_k is chosen uniformly randomly out of $\{0, 1\}$, $Pr(x_k = 0) = \frac{1}{2}$. Hence, case 1 concludes with getting: $Pr(S = 0 \mid S' = 0) = \frac{1}{2}$.
- **Case 2: S' is non-zero** To make $S = 0$ in this case, the contribution of the k^{th} term should be non-zero and equal to $-S'$ (according to Eq:2) Thus, in order to assure that, x_k must be non-zero and d_{ik} must be $-S'$.
Now, we can say that $Pr(S = 0 \mid S' \neq 0) = Pr(x_k = 1 \cap d_{ik} = -S') \leq Pr(x_k = 1)$. And, we know that $Pr(x_k = 1) = \frac{1}{2}$, by an argument similar to case 1. Hence, case 2 concludes with the following relation: $Pr(S = 0 \mid S' \neq 0) \leq \frac{1}{2}$.

Using Partition theorem of probability and the relations derived from Case 1 and Case 2, we can arrive at the following equation:

$$\begin{aligned} Pr(S=0) &= Pr(S=0 \mid S'=0)\cdot P(S'=0) + Pr(S=0 \mid S'\neq 0)\cdot P(S'\neq 0) \\ &\implies Pr(S=0) \leq \frac{1}{2}\cdot P(S'=0) + \frac{1}{2}\cdot[1-P(S'=0)] \\ &\implies Pr(S=0) \leq \frac{1}{2} + \cancel{\frac{1}{2}\cdot P(S'=0)} - \cancel{\frac{1}{2}\cdot P(S'=0)} \\ &\implies Pr(S=0) \leq \frac{1}{2} \quad (\therefore Hence, proved) \end{aligned}$$

(c) Since $AB \neq C$, at least one element in the matrix D will be non-zero. Let us assume this element resides in row i. Thus $D_i \neq \bar{0}$.
In part(b), we proved that given any row of a matrix containing at least one non-zero element, $\Pr_x(D_i\bar{x}=0) \leq \frac{1}{2}$ holds.
Now, the error probability in this case can be written as $\Pr_x(D\bar{x}=\bar{0})$ which is equivalent to saying that every row when individually multiplied by $\bar{x}$ should give 0. We need to take the intersection of all these events in order to get the required probability.
Next, we know that this combined probability of the intersection of each of the events will be less than equal to the probability of each of the events individually. And, we also know the probability of the i^{th} event, i.e. $\Pr_x(D_i\bar{x}=0) \leq \frac{1}{2}$. Thus, we get,

$$\Pr_x(D\bar{x}=\bar{0}) = \Pr_x((D_1\bar{x}=0)\cap(D_2\bar{x}=0)\cap...(D_i\bar{x}=0)...(D_n\bar{x}=0)) \leq \Pr_x(D_i\bar{x}=0) \leq \frac{1}{2}$$

(d) In order to improve its error probability to $\frac{1}{2^{100}}$, we can repeat the above algorithm 100 times independently. The event of error can be defined as the event of intersection of errors of all these iterations of the algorithm. Since, these executions are independent of each other, $\Pr(error)$ can be expressed as:

$$\begin{aligned} \Pr(error) &= \Pr((D\_\_1^{st}\bar{x}=0)\cap(D\_\_2^{nd}\bar{x}=0)\cap....\cap(D\_\_100^{th}\bar{x}=0)) \\ &= \Pr(D\_\_1^{st}\bar{x}=0)\cdot\Pr(D\_\_2^{nd}\bar{x}=0)...\Pr(D\_\_100^{th}\bar{x}=0) \leq \frac{1}{2^{100}} \end{aligned}$$

Thus, we would get a reduction in original error by a factor of 2^{99}.
Overhead Analysis:
Let the time taken in one run of the original algorithm be t. In the worst case of the new repetitive algorithm, it would take a time $t' = 100t$. Another observation is that for small n, repeating the algorithm 100 times might be a poorer choice (time $\approx 100n^2$) than solving deterministically (time $\approx n^{2.3}$). But as n increases, for large values of n, repetitive randomized algorithm gives a better time complexity.

Another interpretation of this question might be to calculate the expected running time of this repetitive randomised algorithm, and compare with the time required in single run(t). For that:
Let the expected time of this improved algorithm be $E[t']$. For simplicity, we assume the error probability is p in any single run, which can be replaced later by $\frac{1}{2}$ to calculate the best bound
Here, we assume that we stop executing the algorithm, once the algorithm outputs "No". The time taken is the number of iterations upto this iteration. If the algorithm goes on for 100 runs, the 100^{th} iteration can either output "No" or "Yes", both cases taking a time $100t$.

$$E[t'] = p\cdot t + (1-p)^1\cdot p\cdot 2t + (1-p)^2\cdot p\cdot 3t + ... + (1-p)^{99}\cdot p\cdot 100t + (1-p)^{99}\cdot(1-p)\cdot 100t$$

On solving the above $A.G.P$, and putting $p = \frac{1}{2}$ we get

$$E[t'] = \left[\frac{1 + (99 - 100p)(1-p)^{100}}{p}\right] \cdot t$$
$$\implies \left(\frac{1 + (49)(\frac{1}{2})^{100}}{(\frac{1}{2})}\right) \cdot t = 2t \cdot \left(1 + (49)(\frac{1}{2})^{100}\right) \approx 2t$$

□

3. (8+10+7=25 marks) **Improved Chernoff's Bound**

We can improve Chernoff's bound in special cases of random variables as opposed to 0/1 random variables (using simpler proof techniques).

Let X is a sum of n independent random variables $X_1, \ldots, X_n$, each taking values in $\{1, -1\}$, with $\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}$. Then for any $a > 0$, we will prove that

$$\Pr(X \geq a) \leq e^{-\frac{a^2}{2n}}. \tag{3}$$

(a) Prove the inequality $\frac{t^{2i}}{(2i)!} \leq \frac{(t^2/2)^i}{i!}$.

(b) Take a variable $t > 0$. Show that $E[e^{tX_i}] \leq e^{t^2/2}$.

(c) Prove inequality 3.

Solution:

(a) The RHS of $\frac{t^{2i}}{(2i)!} \leq \frac{(t^2/2)^i}{i!}$ can be re-written as: $\frac{t^{2i}}{(2i)!} \leq \frac{t^{2i}}{2^i \cdot i!}$. Cancelling t^{2i} and inverting both sides, we see that the above inequality is equivalent to $2^i \cdot i! \leq (2i)!$
Hence we need to just show that $2^i \cdot i! \leq (2i)!$ holds for all $i \geq 1$ for proving the required inequality. Let us prove the following by induction.

$$2^i \cdot i! \leq (2i)! \quad \forall i \geq 1$$

Base Case:
For $i = 0$, both LHS and RHS are equal to 1. Hence, the above relation holds trivially.
Induction Step:
Let us assume that the given Induction Hypothesis (I.H.) holds for i. We need to show that it also holds for $i + 1$.
Given $2^i \cdot i! \leq (2i)!$ is true. Let us consider the L.H.S for $i + 1$

$$L.H.S. = 2^{(i+1)} \cdot (i+1)! = 2 \cdot (i+1) \cdot 2^i \cdot i!$$
$$\implies L.H.S. \leq 2 \cdot (i+1) \cdot (2i)! \quad (since\ i^{th}\ I.H.\ is\ true)$$
$$\implies L.H.S. \leq (2i+2) \cdot (2i)! \quad (Taking\ 2\ inside)$$
$$\implies L.H.S. \leq (2i+2) \cdot (2i+1) \cdot (2i)! \quad (**\ Multiplying\ RHS\ by\ (2i+1))$$
$$\implies L.H.S. \leq (2(i+1))! \quad (Hence,\ proved)$$

** – (Multiplying the larger side by a quantity greater than 1 does not reverse the inequality. Since $i > 0$, $2i + 1$ is a quantity greater than 1.)

(b) By definition of expectation, $E[e^{tX_i}]$ can be written in terms of the random variable X_i as: $\Pr(X_i = 1) \cdot e^t + \Pr(X_i = -1) \cdot e^{-t}$, since the range of R.V. e^{tX_i} is $\{e^t, e^{-t}\}$ and $\Pr(X_i = 1)$ is equivalent to saying that $\Pr(e^{tX_i} = e^t)$.
Also given that $\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}$, we get:

$$E[e^{tX_i}] = \frac{1}{2} \cdot \left[e^t + e^{-t}\right]$$

Applying expansion of e^x, we get

$$\begin{aligned} L.H.S. = E[e^{tX_i}] &= \frac{1}{2} \cdot \left[\left(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!}...\right) + \left(1 - t + \frac{t^2}{2!} - \frac{t^3}{3!}...\right)\right] \\ &= \frac{1}{2} \cdot 2\left[1 + \frac{t^2}{2!} + \frac{t^4}{4!}....\right] = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \end{aligned}$$

Using inequality proved in (a), we can replace each of the terms in this sum and get:

$$\sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{(t^2/2)^k}{k!} = e^{\frac{t^2}{2}} = R.H.S.$$

The term $\sum_{k=0}^{\infty} \frac{(t^2/2)^k}{k!}$ is nothing but the taylor expansion of $e^{\frac{t^2}{2}}$, which is the R.H.S. of the inequation we need to prove. Hence, proved

(c) The condition $X \geq a$ is equivalent to $e^{tX} \geq e^{ta}$ for any $t > 0$ (since exp is an increasing function). Thus, applying Markov's Inequality over the new random variable e^{tX}, we get

$$L.H.S = \Pr(X \geq a) = \Pr(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} \tag{4}$$

Given that X is the sum of independent random variables X_i's, we can re-write the expectation as:

$$\begin{aligned} E[e^{tX}] = E[e^{t(X_1+X_2...X_n)}] &= E[e^{tX_1}e^{tX_2}...e^{tX_n}] \quad (Using **) \\ &= E[e^{tX_1}]E[e^{tX_2}]....E[e^{tX_n}] \leq e^{n\cdot\frac{t^2}{2}} \end{aligned}$$

We get the last inequality by what we proved in $part(b)$ for any random variable X_i. Thus, putting the above inequality in $eq4$:

$$L.H.S \leq \frac{E[e^{tX}]}{e^{ta}} \leq \frac{e^{n\cdot\frac{t^2}{2}}}{e^{ta}}$$

Now put $t = \frac{a}{n}$ to get:

$$\frac{e^{n\cdot\frac{t^2}{2}}}{e^{ta}} = \frac{e^{\frac{a^2}{2n}}}{e^{\frac{a^2}{n}}} = e^{\frac{-a^2}{2n}} = R.H.S.$$

Hence, we proved $L.H.S. \leq R.H.S.$.

(Proof for $**$:)

$$\begin{aligned} E[e^{tX_1}e^{tX_2}...e^{tX_n}] &= \sum \Pr(X_1 = x1, X_2 = x2...X_n = xn).e^{tx1}.e^{tx2}....e^{txn} \\ = \sum \Pr(X_1 = x1)P(X_2 = x2)...p(X_n = xn).e^{tx1}.e^{tx2}....e^{txn} &\quad (by\ definition\ of\ independence) \\ = \sum (\Pr(X_1 = x1).e^{tx1}).(\Pr(X_2 = x2).e^{tx2})...(\Pr(X_n = xn).e^{txn}) & \\ = E[e^{tX_1}]E[e^{tX_2}]....E[e^{tX_n}] & \end{aligned}$$

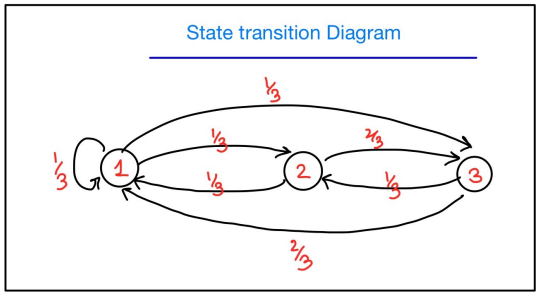

Figure 1: State Transition Diagram

□

4. (5+15=20 marks) **Markov Chain**

A homogeneous Markov chain has state space $S = \{1, 2, 3\}$ with the transition matrix M as follows:

$$M = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}$$

(a) Draw the state transition diagram corresponding to M.

(b) Let $\Pr(X_0 = 1) = \frac{1}{2}$ and $\Pr(X_0 = 2) = \frac{1}{4}$. Find $\Pr(X_0 = 3, X_1 = 2, X_2 = 1)$.

Solution:

(a) See Fig:1

(b) We know that for any Markov Chain, the following property holds:

$$\Pr(X_0 = 3, X_1 = 2, X_2 = 1) = \Pr(X_0 = 3) \cdot \Pr(X_1 = 2|X_0 = 3) \cdot \Pr(X_2 = 1|X_1 = 2)$$

Since it is a homogeneous Markov Chain, we know that $\Pr(X_2 = 1|X_1 = 2) = \Pr(X_1 = 1|X_0 = 2)$. We also know that the initial probability distribution along the entire state space, must sum to 1.

$$\Pr(X_0 = 1) + \Pr(X_0 = 2) + \Pr(X_0 = 3) = 1$$

Hence, we get $\Pr(X_0 = 3) = \frac{1}{4}$.
We can also get the required transition probabilities using the transition matrix M, as:

$$\Pr(X_0 = 3, X_1 = 2, X_2 = 1) = \Pr(X_0 = 3) \cdot \Pr(X_1 = 2|X_0 = 3) \cdot \Pr(X_1 = 1|X_0 = 2)$$
$$\implies \frac{1}{4} \cdot M_{32} \cdot M_{21} = \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{36}$$

□

References

[1] François Le Gall. *Powers of tensors and fast matrix multiplication. International Symposium on Symbolic and Algebraic Computation, ISSAC'14, Kobe, Japan, July 23-25, 2014.*