

CS203B

Mathematics For Computer
Science -III (Probability)
Indian Institute of Technology, Kanpur

Assignment 3

Name:
Anjali Rana (190147)

Date of Submission:
26th April 2021

Question 1

[2 + 7 points] Suppose there is a town with n persons. Each one of them is lazy but quite fond of spreading fake news. On some day, a person hears some fake news. The next morning, s/he uniformly randomly picks a phone number from the telephone directory, and communicates the fake news to the person with that number. So, now there are 2 persons who know the fake news. As you can imagine, the fake news starts spreading according to the following protocol.

Every morning, each person who knows the fake news, picks a phone number uniformly randomly from the telephone directory, and communicates the fake news to the other person.

You may observe that there is a possibility that a person calls the same person multiple times. Moreover, a single person may get a call from multiple persons on the same day. Hence, every phone call may not necessarily increase the number of persons knowing the fake news.

Let X be a random variable for the number of days it takes for each person of the town to know the fake news. It can be seen that for every m , however large it may be, $P(X > m)$ is non-zero.

Show that $\log_2 n$ is the least number of days to spread the fake news to the entire town. I.e. $P(X < \log_2 n) = 0$.

What is the expected number of days $E[X]$ it will take to spread the fake news? Give a 'nice' expression and a convincing heuristic analysis (as you may find the rigorous proof too difficult).

Solution

We need to show minimum number of days required to spread the news is $\log_2 n$. For this we choose the fastest way in which news can be spread. This can be possible if every person gets the news only once.

Suppose on day 1 one person knows the news. Then on day 1 morning he tells it to a new person so now 2 persons know the news. Now on day 2 morning, these 2 people spread the news to 2 more new people. Now 4 people know the news. On day 3 morning, these 4 people spread the news to more new 4 people and thus, now 8 new people know the news. Hence we see everyday number of people who know the news gets doubled. So at i_{th} day, 2^i people know the news.

So for the entire city to know the news, $\lceil \log_2 n \rceil$ days are the required number of days if the news spreads via the above explained fastest mode. Hence, $P(X < \log_2 n) = 0$.

Given:

X = number of days it takes for each person of the town to know the fake news.

We do the following heuristic analysis to calculate $E[X]$:

We define the following considering starting day no. as Day 0:

D_j = Day no. when no. of people knowing the news becomes j . So,

$D_{j+1} - D_j$: No. of days in which no. of people who know the news increase by one.

Lets define the following terms:

Aware person : Who knows the news

Unaware person : Who does not know the news.

ON DAY D_j :

We define the random variables $A_k \forall k \in [j]$:

$$A_k = \begin{cases} 1, & k_{th} \text{ aware person call another unaware person} \\ 0, & k_{th} \text{ aware person calls another aware person} \end{cases}$$

On Day D_j , $(n - j)$ people are unaware. We define

B_j = number of calls made to unaware people from j aware people. we have

$$B_j = \sum_{k \in [j]} A_k.$$

Since calls are independent, using linearity of expectation we have:

$$E[B_j] = E\left[\sum_{k \in [j]} A_k\right] = \sum_{k \in [j]} E[A_k] \quad (1)$$

Now,

$$E[A_k] = 1 \cdot P(A_k = 1) + 0 \cdot P(A_k = 0) \quad (2)$$

Now, on Day D_j , $(n - j)$ people are unaware and total number of people to whom an aware person can make call on D_j are $(n - 1)$. Hence probability that call is made to an unaware person i.e.

$$P(A_k = 1) = \frac{n-j}{n-1}$$

Using above equation, (1) and (2)

$$E[B_j] = j \cdot \frac{n-j}{n-1}$$

So on Day D_j , $j \cdot \frac{n-j}{n-1}$ expected number of calls are made to unaware people. So $1/E[B_j]$ days after Day D_j , an unaware person becomes aware.

Hence,

$$E[D_{j+1} - D_j] = \frac{n-1}{(n-j)j}$$

Now,

Since $D_n = \sum_{j=1}^{n-1} (D_{j+1} - D_j)$, we have:

$$E[D_n] = E\left[\sum_{j=1}^{n-1} (D_{j+1} - D_j)\right]$$

Using linearity of expectation,

$$\begin{aligned} &= \sum_{j=1}^{n-1} \frac{n-1}{(n-j)j} \\ &= \sum_{j=1}^{n-1} \left(\frac{n-1}{n} \left[\frac{1}{j} + \frac{1}{n-j} \right] \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{n-1}{n} \sum_{j=1}^{n-1} \left[\frac{1}{j} + \frac{1}{n-j} \right] \\
&= 2 \frac{n-1}{n} \sum_{j=1}^{n-1} \left[\frac{1}{j} \right] \quad (\because \sum_{j=1}^{n-1} \frac{1}{j} = \sum_{j=1}^{n-1} \frac{1}{n-j})
\end{aligned} \tag{3}$$

Now we calculate the sum for a large population (i.e $n \rightarrow \infty$)

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{n-1} \frac{1}{j} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{n-1} \frac{1}{j/n}$$

Replace j/n with y to make it Reimann Sum

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{y=1/n}^{(n-1)/n} \frac{1}{y} \\
&= \int_{1/n}^1 \frac{1}{y} dy \\
&= \log(1) - \log(1/n) \\
&= \log n
\end{aligned}$$

So,

$$\begin{aligned}
\sum_{j=1}^{n-1} \frac{1}{j} &\rightarrow \log n \text{ for large } n \\
\frac{n-1}{n} &\rightarrow 1 \text{ for large } n
\end{aligned}$$

Hence, from eq. 3,

$$\begin{aligned}
E[D_n] &= 2 \frac{n-1}{n} \sum_{j=1}^{n-1} \left[\frac{1}{j} \right] \\
&\sim 2 \log n
\end{aligned}$$

Now, D_n is the day no. on which no. of people who know the fake news becomes n , which is equal to the random variable X . So,

$$E[X] \sim 2\log n$$

Question 2

[9 points] On a long straight road lies a house at 0-th milestone and a bar at n -th milestone. There is a drunkard standing at i -th milestone. He decides to walk the next mile towards the bar or the home with equal probability. He repeats this on reaching each milestone. The decision (to walk the next mile towards home or bar) at each milestone is independent of all the past moves (due to the large amounts of Vodka he had). The walk stops as soon as the drunkard reaches home or bar.

Let $e(i)$ be the expected number of steps in which the walk stops starting from the i -th milestone. Let $p(i)$ be the probability that the drunkard starting from the i -th milestone reaches home.

Fix $n = 4$ for an easier calculation. Find $e(i)$ and $p(i)$, $i \in [3]$.

(Optional question) What are these values for general n ?

Solution

We define the following events:

S : Drunkard takes step in positive direction

\bar{S} : Drunkard takes step in negative direction

R : Drunkard has eventually reached home (0-th milestone)

D_x : Position coordinate of drunkard after taking x steps where $x \in [n] \cup 0$

Suppose Y be the random variable such that:

$Y :=$ Number of steps the drunkard walks before he stops (reaches home or the bar).

As per the question,

$e(i) =$ Expected value of Y given that drunkard starts from the i -th milestone.

So,

$$\begin{aligned} e(i) &= E[Y | D_0 = i] \\ &= P(S | D_0 = i) \cdot E[Y | D_0 = i \wedge S] + P(\bar{S} | D_0 = i) \cdot E[Y | D_0 = i \wedge \bar{S}] \\ &= P(S) \cdot E[Y | D_1 = i + 1] + P(\bar{S}) \cdot E[Y | D_1 = i - 1] \end{aligned}$$

We have,

$P(S|D_0 = i) = P(\bar{S}|D_0 = i) = \frac{1}{2}$ (each step is independent of history).

$$\begin{aligned} e(i) &= \frac{1}{2} \cdot (1 + E[Y|D_0 = i + 1]) + \frac{1}{2} \cdot (1 + E[Y|D_0 = i - 1]) \\ &= 1 + \frac{1}{2} \cdot e(i + 1) + \frac{1}{2} \cdot e(i - 1) \\ \implies e(i + 1) - 2e(i) + e(i - 1) + 2 &= 0 \quad \dots(1) \end{aligned}$$

Eliminating the constant in eq.1 by substituting $i = i + 1$ and then subtracting both the equations gives:

$$e(i + 2) - 3e(i + 1) + 3e(i) - e(i - 1) = 0$$

Now, to solve the above equation, we use the characteristic polynomial of above equation:

$$\begin{aligned} x^3 - 3x^2 + 3x - 1 &= 0 \\ \implies (x - 1)^3 &= 0 \end{aligned}$$

The solution of the equation will be,

$$e(i) = (P + Qi + Ri^2) \cdot (1)^i = P + Qi + Ri^2$$

If $i=0$, drunkard is home:

$$e(0) = P = 0$$

If $i=n$, drunkard reached bar and has already stopped.

$$e(n) = Qn + Rn^2 = 0 \implies R = -\frac{Q}{n}$$

Substituting values of P and R in eq. 1 we get,

$$Q(i + 1) - \frac{Q}{n}(i + 1)^2 - 2Qi + 2\frac{Q}{n}i^2 + Q(i - 1) - \frac{Q}{n}(i - 1)^2 + 2 = 0$$

Comparing constant terms both sides

$$Q - \frac{Q}{n} - Q - \frac{Q}{n} + 2 = 0$$

$$\implies Q = n$$

Therefore, $P = 0, Q = n, R = -1$

So,

$$e(i) = ni - i^2$$

For $n = 4$, we get

$$e(1) = 3, e(2) = 4, e(3) = 3$$

As per the question,

$p(i)$ =probability that the drunkard starting from the i -th milestone reaches home

$$\begin{aligned} p(i) &= P(R|D_0 = i) = P(R \wedge S|D_0 = i) + P(R \wedge \bar{S}|D_0 = i) \\ &= P(R|S \wedge D_0 = i) \cdot P(S|D_0 = i) + P(R|\bar{S} \wedge D_0 = i) \cdot P(\bar{S}|D_0 = i) \quad \dots(2) \end{aligned}$$

Since no. of steps don't affect probability of reaching home:

$$P(R|S \wedge D_0 = i) = P(R|D_0 = i + 1) = p(i + 1)$$

$$P(R|\bar{S} \wedge D_0 = i) = P(R|D_0 = i - 1) = p(i - 1)$$

Substituting values in (2) we get:

$$p(i) = \frac{1}{2} \cdot p(i + 1) + \frac{1}{2} \cdot p(i - 1)$$

$$\implies p(i + 1) - 2p(i) + p(i - 1) = 0 \quad \dots(3)$$

Now, to solve the above equation, we use the characteristic polynomial of above equation:

$$x^2 - 2x + 1 = 0$$

We get $x = 1, 1$

Since there is a repeated root so the solution of equation (3) is:

$$\begin{aligned}
 p(i) &= (P + Qi) \cdot (1)^i \\
 &= P + Qi \quad \dots(4)
 \end{aligned}$$

If $i=0$, means drunkard is already home so,

$$p(0) = P = 1$$

If $i=n$ means drunkard has reached bar and cannot reach home.

$$p(n) = 1 + Qn = 0 \implies Q = -\frac{1}{n}$$

Put these values in eq. 4,

$$p(i) = 1 - \frac{i}{n}$$

Therefore, for $n = 4$, we get

$$p(1) = \frac{3}{4}, p(2) = \frac{1}{2}, p(3) = \frac{1}{4}$$

Question 3

[3 + 6 points] When an event E happens, you often feel an emotion of 'surprise'. Let us try to measure this emotion mathematically. Assume that the surprise depends only on the probability $p := P(E)$, that the mind has somehow estimated. Denote the surprise by a non-negative function $S(p)$. It is defined only for p in the range $[0, 1]$. Unsurprisingly, the surprise function should satisfy the following natural axioms.

Axiom 1. $S(1) = 0$

Since there is no surprise when an event E happens, that has probability $p = 1$.

Axiom 2. $p > q \implies S(p) < S(q)$.

Since the surprise on event-1 happening is less, if its probability is known to be larger than event-2.

Axiom 3. $S(p)$ is a continuous function of p .

Axiom 4. $S(pq) = S(p) + S(q)$. Why is this justified?

Show that the surprise function $S(p)$ is given by $S(1/2) \cdot \log_2(1/p) \geq 0$, for $p \in [0, 1]$. It is customary to define $S(1/2) := 1$; making $S(p) = \log_2(1/p)$.

Solution

Axiom 4 justification:

Suppose there are two independent events M and N . $P(M) = p$ and $P(N) = q$. The surprise on seeing $M \cap N$ (i.e. $S(pq)$) might be taken to be surprise on seeing M ($S(p)$) plus remaining additional surprise on seeing N given that M has already been seen. Since M and N are independent events $S(q|p) = S(q)$. So that additional surprise is $S(q)$

To prove : $S(p) = S(1/2) \cdot \log_2(1/p) \geq 0$.

Proof. Put $p = q = \frac{1}{2}$ in axiom 4

$$S\left(\frac{1}{4}\right) = S\left(\frac{1}{2}\right) + S\left(\frac{1}{2}\right)$$

$$S\left(\frac{1}{4}\right) = 2S\left(\frac{1}{2}\right) \quad \dots(1)$$

Now put $q = \frac{1}{4}$ in axiom 4

$$S\left(\frac{p}{4}\right) = S(p) + S\left(\frac{1}{4}\right) \quad \dots(2)$$

Then put $q = \frac{1}{2}$ in axiom 4

$$S\left(\frac{p}{2}\right) = S(p) + S\left(\frac{1}{2}\right) \quad \dots(3)$$

From eq. 1 and eq. 2,

$$S\left(\frac{p}{4}\right) = S(p) + 2S\left(\frac{1}{2}\right) \quad \dots(4)$$

Now subtract 2 times eq.3 from eq.4 to get,

$$S\left(\frac{p}{4}\right) + S(p) = 2S\left(\frac{p}{2}\right) \quad \dots(5)$$

Now we define following function (where x is a real number)

$$G(x) = S\left(\left(\frac{1}{2}\right)^x\right) \quad \dots(6)$$

Put $x = x + 1$

$$G(x + 1) = S\left(\frac{1}{2} \left(\frac{1}{2}\right)^x\right) \quad \dots(7)$$

Put $x = x + 2$

$$G(x + 2) = S\left(\frac{1}{4} \left(\frac{1}{2}\right)^x\right) \quad \dots(8)$$

Substitute $\left(\frac{1}{2}\right)^x = p$ in equations 6,7,8. Put it in 5:

$$G(x + 2) - 2G(x + 1) + G(x) = 0$$

Characteristic polynomial of above equation will be,

$$x^2 - 2x + 1 = 0$$

$$\implies x = 1, 1$$

So the solution will be,

$$G(x) = (P + Qx)(1)^x$$

$$\implies G(x) = P + Qx \quad \dots(9)$$

Substitute $x = 0$ and $x = 1$ in eq.9

$$G(0) = P = S\left(\left(\frac{1}{2}\right)^0\right) = S(1) = 0$$

$$G(1) = Q = S\left(\left(\frac{1}{2}\right)^1\right) = S\left(\frac{1}{2}\right)$$

$$\implies G(x) = S\left(\frac{1}{2}\right)x \quad \dots(9)$$

Now,

$$\text{Since, } \left(\frac{1}{2}\right)^x = p$$

$$\implies x = \log_2\left(\frac{1}{p}\right)$$

Therefore,

$$G(x) = S(p) = S\left(\frac{1}{2}\right) \log_2\left(\frac{1}{p}\right) \quad (\text{from (9)})$$

Now,

$$0 \leq p \leq 1$$

$$\implies 1 \leq \frac{1}{p} < \infty$$

$$\implies 0 \leq \log_2 \frac{1}{p} < \infty$$

Now, using axiom 2

$$\begin{aligned} S(1) = 0 &\implies S(1/2) > 0 && \text{(since } 1 > 1/2 \text{ so } S(1) < S(1/2)) \\ \implies 0 \leq S\left(\frac{1}{2}\right) \log_2\left(\frac{1}{p}\right) &< \infty \\ \implies 0 \leq S(p) &< \infty \end{aligned}$$

Hence proved,

$$S(p) = S(1/2) \cdot \log_2(1/p) \geq 0$$

□

Question 4

[8points] Suppose that every vertex of an n -vertex bipartite graph is given a personalized list of $> \log_2 n$ possible colors. Prove that it is possible to give each vertex a color, from its list, such that no two adjacent vertices receive the same color.

[Hint: Why is this a probability question?]

Solution

We will do a probabilistic proof. We will prove that probability of failing is less than 1, hence there is a positive probability of success.

Suppose A : Set containing n vertices of the bipartite graph.

Since the graph is bipartite, it can be divided into two independent and disjoint subsets, say A_1 and A_2 .

Let C_a be the set containing personalized list of colors of vertex $a \in A$.

Given $|C_a| > \log_2 n \forall a \in A$

Let U be the set of all colors used in the graph.

$$U = \bigcup_{a \in A} C_a$$

Now, let us define a function

$$h : U \mapsto \{1, 2\}$$

$\forall u \in U$, value of $h(u)$ is randomly selected to be 1 or 2 with equal probability. Hence, this function partitions the set A into two partitions. So we have, $\forall u \in U$

$$\begin{aligned} P(h(u) = 1) &= \frac{1}{2} \\ P(h(u) = 2) &= \frac{1}{2} \end{aligned}$$

Now, let us define another function ($\mathcal{P}(U)$ denotes power set of U)

$$q : \{1, 2\} \mapsto \mathcal{P}(U)$$

$q(1)$: set of all those colors $u \in U$ for which $h(u) = 1$
 $q(2)$: set of all those colors $u \in U$ for which $h(u) = 2$.

In order to prove no two adjacent vertices have same colour, we need to prove that there exists a function h such that $\forall i \in \{1, 2\}$

$$q(i) \cap C_a \neq \emptyset \quad \forall a \in A$$

i.e. we need to prove

$$P(q(i) \cap C_a \neq \emptyset \forall a \in A) > 0$$

So, we need to prove,

$$P(\exists a \in A : q(i) \cap C_a = \emptyset) < 1$$

Now,

$$\begin{aligned} P(\exists a \in A : q(i) \cap C_a = \emptyset) &= P\left(\bigcup_{a \in A} (q(i) \cap C_a = \emptyset)\right) \\ &\leq \sum_{i=1}^2 \sum_{a \in A_i} P(q(i) \cap C_a = \emptyset) \end{aligned}$$

Now, colours in C_a are mapped with probability 1/2 to k^{th} partition such that $k \neq i$,
so

$$\begin{aligned} P(\exists a \in A : q(i) \cap C_a = \emptyset) &\leq \sum_{i=1}^2 \sum_{a \in A_i} \left(\frac{1}{2}\right)^{|C_a|} \\ &= \sum_{a \in A} \left(\frac{1}{2}\right)^{|C_a|} \\ &< \sum_{a \in A} \frac{1}{n} \quad (\because |C_a| > \log_2 n) \\ &< 1 \end{aligned}$$

Therefore,

$$\begin{aligned} P(\exists a \in A : q(i) \cap C_a = \emptyset) &< 1 \\ \implies P(h \text{ fails for some } a \in A) &< 1 \\ \implies P(h \text{ does not fail for any } a \in A) &> 0 \end{aligned}$$

Hence proved that there is a positive probability of existence of a function h which satisfies the conditions. So there exists such a colouring.

REFERENCES:

<https://hal.archives-ouvertes.fr/hal-01652777/document>

<https://people.maths.bris.ac.uk/maaig/teaching/complexnets/rumours.pdf>

<https://math.stackexchange.com/questions/748776/colouring-bipartite-graph-with-sets-of-possible-colors-to-each-vertex>