A1 COMP690 - Fall 2022

Gulliver Häger - 260907422

September 18, 2022

Contents

1 An Exercise on Infinity

1.1 Expected Number of Nodes at Depth l

The number of nodes at depth l in tree T_0 is by definition going to be the number of nodes which have a path of length l to node 0. Every such node with a path of length l is going to contribute 1 to N_l, making it convenient to construct N_l as the sum of indicator functions, where each indicator function corresponds to a possible path of length l.

For a path to have length l from a node i_l to node 0, i_l has to connect to some i_{l-1} which in turn has to connect to some i_{l-2} and so on until i_2 connects with i_1 which ultimately connects to 0. By the given rules for construction of the tree, i_l connecting to i_{l-1} is equivalent to $Z_{i_l} = (i_l - i_{l-1})$. This can then be extended for all i_j where $1 < j \leq l$, where $j = 1$ corresponds to i_1 connecting to node 0, which by definition corresponds to $Z_{i_1} = i_1$

So far, only an arbitrary path has been considered. The expression N_l is going to be the sum of all possible paths, meaning that N_l can be expressed as the following sum, which ranges over all possible paths.

$$N_l = \sum_{1 \leq i_1 < i_2 < \cdots < i_l} \mathbb{1}_{[(Z_{i_1}=i_1)\cap(Z_{i_2}=(i_2-i_1))\cap\cdots\cap(Z_{i_l}=(i_l-i_{l-1}))]} \tag{1}$$

But this sum can clearly also be written as:

$$N_l = \sum_{i_1=1}^{\infty} \sum_{i_2=i_1+1}^{\infty} \cdots \sum_{i_l=i_{l-1}+1}^{\infty} \mathbb{1}_{[(Z_{i_1}=i_1)\cap(Z_{i_2}=(i_2-i_1))\cap\cdots\cap(Z_{i_l}=(i_l-i_{l-1}))]} \tag{2}$$

By linearity of expectation, taking the expectation of N_l as expressed in (2) will be:

$$\mathbb{E}\{N_l\} = \sum_{i_1=1}^{\infty} \sum_{i_2=i_1+1}^{\infty} \cdots \sum_{i_l=i_{l-1}+1}^{\infty} \mathbb{E}\{\mathbb{1}_{[(Z_{i_1}=i_1)\cap(Z_{i_2}=(i_2-i_1))\cap\cdots\cap(Z_{i_l}=(i_l-i_{l-1}))]}\} \tag{3}$$

But the expectation of an indicator is just the probability of the event, and since all Z_j's are i.i.d. they can be written as a product of their probabilities, so (3) can be written as:

$$\mathbb{E}\{N_l\} = \sum_{i_1=1}^{\infty} \sum_{i_2=i_1+1}^{\infty} \cdots \sum_{i_l=i_{l-1}+1}^{\infty} q_{i_1} q_{(i_2-i_1)} \cdots q_{(i_l-i_{l-1})} \tag{4}$$

But the last term in the product of the sum will be the only dependent variable for the innermost sum, and based on the starting value of i_l, $q_{(i_l-i_{l-1})}$ will range over $q_1, q_2, \cdots$ up to infinity. Thus, (4) can be rewritten as:

$$\mathbb{E}\{N_l\} = \sum_{i_1=1}^{\infty} \sum_{i_2=i_1+1}^{\infty} \cdots \sum_{i_{l-1}=i_{l-2}+1}^{\infty} q_{i_1} q_{(i_2-i_1)} \cdots q_{(i_{l-1}-i_{l-2})} \sum_{i_l=1}^{\infty} q_{i_l} \tag{5}$$

But $\sum_{i_l=1}^{\infty} q_{i_l} = 1$ since the sum is a sum of the probabilities of disjoint events whose union comprises the whole probability space - the natural numbers. Therefore, (5) can be rewritten as:

$$\mathbb{E}\{N_l\} = \sum_{i_1=1}^{\infty} \sum_{i_2=i_1+1}^{\infty} \cdots \sum_{i_{l-1}=i_{l-2}+1}^{\infty} q_{i_1} q_{(i_2-i_1)} \cdots q_{(i_{l-1}-i_{l-2})} \tag{6}$$

But this equation is of the exact same form as (4), meaning that the same re-indexing and rearranging can be done for every sum, the outermost included. Thus, the final result of the computation is:

$$E_{\{}N_l\} = 1^l = 1 \tag{7}$$

1.2 Non-Zero Probability of an Infinitely Sized Tree with Proper Root 0

Suppose $q_1 > 0$ and that $\mathbb{E}\{Z\} < \infty$.

Consider the special case of $|T_0| = \infty$ when all nodes are connected to T_0. Since it is a special case - and therefore a subset of all events for which $|T_0| = \infty$ - it is true that:

$$\mathbb{P}\{|T_0| = \infty\} \geq \mathbb{P}\{\text{all nodes are in } T_0\} \tag{1}$$

Now it suffices to show that the RHS in (1) is strictly greater than 0. To do so, an expression for the RHS has to be found. Since all Z_i are i.i.d., the probability of the whole expression is going to be the product of the probabilities that each node is in T_0. But the probability of a node i being in T_0 is the probability that it connects to a node before it - assuming that all prior nodes are in T_0. Thus, the expression will be:

$$\mathbb{P}\{\text{all nodes are in } T_0\} = \prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j \tag{2}$$

Now consider the logarithm of the expression in (2). It is true that:

$$\prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j > 0 \quad \iff \quad \log\left(\prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j\right) > -\infty \tag{3}$$

Applying logarithm rules for products and rewriting the sum:

$$\log\left(\prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j\right) = \sum_{i=1}^{\infty} \log\left(\sum_{j=1}^{i} q_j\right) = \sum_{i=1}^{\infty} \log\left(\left(\sum_{k=1}^{\infty} q_k\right) - \left(\sum_{j=(i+1)}^{\infty} q_j\right)\right) \tag{4}$$

But as explained in 1.1., the first sum of the inner sums equals 1, so (4) can be rewritten as:

$$\log\left(\prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j\right) = \sum_{i=1}^{\infty} \log\left(1 - \left(\sum_{j=(i+1)}^{\infty} q_j\right)\right) \tag{5}$$

From *Bernoulli's Inequality*, the well known inequality $1 + x \leq e^x \quad \forall x \in \mathbb{R}$ can be derived, which then implies that $\log(1 + x) \leq x$, meaning also that $\log(1 - x) \leq -x$. Notice that the only strict equality is when $x = 0$ for all of the above inequalities. Thus, (5) can be bounded as:

$$\log\left(\prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j\right) \geq \sum_{i=1}^{\infty} - \sum_{j=i+1}^{\infty} q_j = -\sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} q_j \tag{6}$$

But (6) almost looks like the expression for the expectation of Z:

$$\log\left(\prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j\right) \geq -\sum_{i=2}^{\infty} i q_i \tag{7}$$

Knowing that $q_1 \geq 0$, (7) can then be rewritten with a strict inequality as:

$$\log\left(\prod_{i=1}^{\infty} \sum_{j=1}^{i} q_j\right) > -\left(q_1 + \sum_{i=2}^{\infty} i q_i\right) = -\sum_{i=1}^{\infty} i q_i = -\mathbb{E}\{Z\} \tag{8}$$

From the original supposition, it is then true that:

$$\log\left(\prod_{i=1}^{\infty}\sum_{j=1}^{i} q_j\right) > -\mathbb{E}\{Z\} > -\infty \tag{9}$$

which was the sufficient condition to prove that $\mathbb{P}\{|T_0| = \infty\} > 0$.

■

1.3 A Bound for the Probability of a Singleton Tree at root 0

By definition, it is true that:

$$\mathbb{P}\{|T_0| = 1\} = \mathbb{P}\{\text{no nodes has an edge to node } 0\} = \mathbb{P}\{i - Z_i \neq 0 \quad \forall i \geq 1\} \tag{1}$$

By the probability of a complement of an event and since all Z_i are i.i.d., (1) becomes:

$$\mathbb{P}\{|T_0| = 1\} = \prod_{i=1}^{\infty}(1 - \mathbb{P}\{Z_i = i\}) = \prod_{i=1}^{\infty}(1 - q_i) \tag{2}$$

By the AM-GM inequality, (2) can be bounded as:

$$\mathbb{P}\{|T_0| = 1\} = \prod_{i=1}^{\infty}(1 - q_i) \leq \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}(1 - q_i)\right)^n = \lim_{n\to\infty}\left(1 - \frac{\sum_{i=1}^{n} q_i}{n}\right)^n \tag{3}$$

As shown in 1.1, $\sum_{i=1}^{\infty} q_i = 1$, meaning that when applying limit laws, (3) can be written as:

$$\mathbb{P}\{|T_0| = 1\} = \prod_{i=1}^{\infty}(1 - q_i) \leq \left(1 - \frac{1}{\lim_{n\to\infty} n}\right)^{\lim_{n\to\infty}} = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n \tag{4}$$

But the expression in (4) is just the reciprocal of e when defined as a limit as is shown below.

$$\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = \lim_{n\to\infty}\left(\frac{1}{\frac{n}{n-1}}\right)^n \tag{5}$$

$$= \lim_{n\to\infty}\frac{1}{\left(1 + \frac{1}{n-1}\right)^n} \tag{6}$$

$$= \lim_{n\to\infty}\left(\frac{1}{\left(1 + \frac{1}{n-1}\right)^{n-1}}\right)\left(\frac{1}{1 + \frac{1}{n-1}}\right) \tag{7}$$

$$= \lim_{n\to\infty}\left(\frac{1}{\left(1 + \frac{1}{n-1}\right)^{n-1}}\right) \tag{8}$$

$$= \lim_{n\to\infty}\left(\frac{1}{(1 + \frac{1}{n})^n}\right) = \frac{1}{e} \tag{9}$$

Thus, putting (4) and (9) together completes the proof and shows that $\mathbb{P}\{|T_0| = 1\} \leq 1/e$.

■

1.4 The Probability of an Infinitely Large Forest

Firstly, it will be shown that if $\mathbb{E}\{Z\} < \infty$ then $\mathbb{P}\{X < \infty\} = 1$. So assume that $\mathbb{E}\{Z\} < \infty$.

Since X is defined to be the number of trees in the forest, it can be defined as the sum of indicators for every node with the condition for being a proper root. By definition of the forest, this condition is that $Z_i > i$. Thus:

$$X = \sum_{i=1}^{\infty} \mathbb{1}_{[Z_i > i]} \tag{1}$$

Now consider the expectation of Z and how it can be rewritten.

$$\mathbb{E}\{Z\} = \sum_{i=1}^{\infty} i q_i = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} q_j = \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} q_j - \sum_{j=1}^{i-1} q_j \right) = \sum_{i=1}^{\infty} \mathbb{P}\{Z_i > i\} \tag{2}$$

By assumption, it is true that $\mathbb{E}\{Z\} < \infty$ and thus:

$$\sum_{i=1}^{\infty} \mathbb{P}\{Z_i > i\} < \infty \tag{3}$$

This satisfies the condition for *first Borel-Cantelli lemma* which gives that:

$$\mathbb{P}\left\{ \limsup_{i \to \infty} \{Z_i > i\} \right\} = 0 \tag{4}$$

This means that the probability that $Z_i > i$ occurs infinitely often is 0. But the probability of $Z_i > i$ occurring infinitely often is by definition the probability that $X = \infty$. Thus, it can be concluded that:

$$\mathbb{P}\{X = \infty\} = 0 \quad \iff \quad \mathbb{P}\{X \neq \infty\} = 1 \quad \iff \quad \mathbb{P}\{X < \infty\} = 1 \tag{5}$$

Which completes the proof of this implication.

Secondly, it will be shown that if $\mathbb{E}\{Z\} = \infty$ then $\mathbb{P}\{X = \infty\} = 1$. So assume that $\mathbb{E}\{Z\} = \infty$. Reusing (2), it is then true that:

$$\mathbb{E}\{Z\} = \sum_{i=1}^{\infty} \mathbb{P}\{Z_i > i\} = \infty \tag{6}$$

Thus, by the *second Borel-Cantelli lemma*, it is true that:

$$\mathbb{P}\left\{ \limsup_{i \to \infty} \{Z_i > i\} \right\} = 1 \tag{7}$$

This means that the probability that $Z_i > i$ occurs infinitely often is 1. For the exact symmetric reason as the argument above, this then means that $\mathbb{P}\{X = \infty\} = 1$. This completes the proof.

■

2 Expected Bit Complexity

In this problem the following lemma will be used:

Lemma 1: If $X = (0.X_1X_2\cdots)$ and $Y = (0.Y_1Y_2\cdots)$ are two numbers represented in their binary expansions, then $X \geq Y$ if $X_1 = 1$ and $Y_1 = 0$.

Proof of Lemma 1: Assume that $X_1 = 1$ and $Y_1 = 0$. Then clearly $X \geq 1/2$ and $1 \geq Y - (1/2)$ since X_1 or Y_1 contributes $1/2$ to X and Y respectively if they equal 1.

Rearranging the second inequality gives: $1/2 \geq Y$ which then by transativity of inequalities implies that $X \geq Y$.

■

2.1 An Interval Partitioned by One Number

Proposed Algorithm: Starting with $i = 1$, enter the following loop:

Compare U_i and a_i. If $U_i = a_i$, continue to the next iteration of this loop by increasing i by 1. If $U_i = 1$ and $a_i = 0$, output that $U \in [a, 1]$. If $U_i = 0$ and $a_i = 1$, output that $U \in [0, a]$.

Correctness of Algorithm: If the first bits of the numbers differ, then lemma 1 gives an order between the numbers and the algorithm terminates. If the first bits of the numbers do not differ, this means that they both either fall into $[0, (1/2)]$ or $[(1/2), 1]$. But any closed interval on the real line is isomorphic to $[0, 1]$, meaning that removing the first bit from each number makes the current problem the exact same as the original one. The only case in which the algorithm would output a solution in this case, is if the bits differ at some point, in which case lemma 1 has already been used to verify the correctness of the solution.

Expected Bit Complexity of Algorithm: By definition of N, it is clearly equal to the number of iterations of the loop the algorithm uses. Since a is a constant in $\{0, 1\}$ and $U_i \sim \text{Bernoulli}(1/2)$ $\forall i \geq 1$, it is true that $\mathbb{P}\{U_i = a_i\} = 1/2$. The condition for terminating the algorithm is $U_i = a_i$ and the complement of that event is that the algorithm gets rerun. Thus:

$$\mathbb{E}\{N\} = \frac{1}{2} \cdot 1 + \frac{1}{2}\mathbb{E}\{N\} \quad \Rightarrow \quad \mathbb{E}\{N\} = 2 \tag{1}$$

■

2.2 An Interval Partitioned by Two Numbers

Visualize this problem as an infinite binary search tree, where each node at depth l represents a dyadic interval of length 2^{-l}. At a node representing the interval $[\frac{j}{2^l}, \frac{j+1}{2^l}]$ for some $0 \leq j < 2^l$, the left child is going to represent the interval $[\frac{2j}{2^{l+1}}, \frac{2j+1}{2^{l+1}}]$ and the right child the interval $[\frac{2j+1}{2^{l+1}}, \frac{2j+2}{2^{l+1}}]$.

Now, any number in its binary expansion can be thought of as an infinite path from the root, where every new node in the path tightens the bounds of the interval the number can be in, and eventually approaches the number as $l \to \infty$. Thus, in this case, a and b are some infinite paths from the root. On the other hand, U will be a random walk in this tree, with probability $1/2$ of both children at every node since $U_i \sim \text{Bernoulli}(1/2)$.

As soon as the random path U makes, call it P_U, is disjoint from the path a makes, call it P_a, it is possible to determine if $U \leq a$ or $a \leq U$. This is true since for the paths to be disjoint, at some node one path has to take the left edge from the current node and the other path the right edge. But since all intervals in the left subtree are smaller than all elements in the right subtree, it must be true that the number of the path which took the left subtree is smaller than the other number. By smaller, it is meant that any number in an interval in the left subtree is smaller than any number in an interval in the right subtree.

The algorithm is then quite intuitive, and is executed by looking at U_i and determining where in the tree P_U is, and at every point looking if P_U is overlapping with P_a and P_b. As soon as P_U is not overlapping, the algorithm halts and outputs the appropriate ordering between U, a, b, which determines which interval U is in.

The argument for the correctness of the algorithm is given above. By the reasoning above as well, N can be defined as the length of $P_U \cap (P_a \cup P_b)$. But how can the expected value of N be calculated?

Let $c \in \{0, 1, 2, \cdots\}$ be the number of leading common bits for a and b. Thus, $a_i = b_i \quad \forall i.1 \leq i \leq c$ but $a_{c+1} \neq b_{c+1}$. Let $\mathbb{E}_{c-i}\{N\}$ be the expectation of N, after having checked U_i, and given that $U_j = a_j = b_j \quad \forall i.1 \leq j \leq i$. The subscript can be thought of as the remaining number of common bits between a and b at the current node.

Now, clearly $\mathbb{E}_0\{N\}$ is the expected value of N given that a and b take different paths at the beginning of the algorithm, meaning that after having checked U_1, only one of P_a, P_b overlap with P_U, meaning that the problem then reduces to the problem in 2.1, which has an expected value of 2. Thus, $\mathbb{E}_0\{N\} = 1 + 2 = 3$.

Now, consider $\mathbb{E}_i\{N\}$ with some i s.t. $0 \leq i < c$. This corresponds to the expected value after having checked i bits where a and b share one of the children of the current node. Both children of the current node have probability $1/2$, and if the one child shared by P_a, P_b is picked, the expected value of this outcome is $\mathbb{E}_{i+1}\{N\}$. If the other child is picked, both U is either greater or smaller than both a and b, and the algorithm terminates. Thus, $\mathbb{E}_i\{N\} = \frac{1}{2}(1 + \mathbb{E}_{i+1}\{N\})$. But this is then just a simple linear recurrence relation with an initial condition, which can easily be calculated (MATH240) to be: $\mathbb{E}_i = \{N\} = 2^{1-i} + 1$. Note once again that the range of i is $0 \leq i \leq c$. But for every i in that range $\mathbb{E}_i\{N\} \leq 3$ since 2^{1-i} is a decreasing function in i with the initial value 2 when $i = 0$. Thus, by definition of c as a function of a and b, the expected value of N when running the algorithm from the beginning is $\mathbb{E}_c\{N\} = 2^{1-c} + 1 \leq 3$.

■

2.3 An Interval Partitioned by k Numbers

The algorithm in 2.2 extends to k points by adding additional paths and the correctness of it follows from the reasoning in 2.2.

To prove a bound for the expected bit complexity, 2.2 will be used as the base case for an induction argument where $k = 2$. For the induction step, assume that $E\{N\} \leq 3 + \log_2(j) \quad \forall j < k$. Now consider the problem for k. Once again, let c be the last common bit for all k points and let $\mathbb{E}_{c-i}\{N\}$ be the expected value of N after having checked U_i given that U_j is the same as all k points $\forall j.1 \leq j \leq i$. Once again, a recurrence relation will be constructed.

Consider $\mathbb{E}_0\{N\}$. At this point in the binary search tree, c bits have been checked and thus $k - l$ of the numbers are going to be going to one of the children of the current node and l to the other child. Here $1 \leq l \leq (k-1)$ clearly. That P_U is going down to either child has probability $1/2$, meaning that with our the bounds from the IH:

$$\mathbb{E}_0\{N\} \leq \frac{1}{2}\left(3 + \log_2(k-l)\right) + \frac{1}{2}\left(3 + \log_2(l)\right) = 3 + \frac{1}{2}\log_2(kl - l^2) = 3 + \log_2(\sqrt{kl - l^2}) \tag{1}$$

But since $1 \leq l \leq (k-1) < k$, it is true that $kl - l^2 < k^2 - l^2$, which in turn mean that $kl - l^2 < k^2$. Since $\log_2$ is a monotonically increasing function on the positive real line, it is therefore true that $\log_2(\sqrt{kl - l^2}) < \log_2(\sqrt{k^2}) = \log_2(k)$. This inequality together with (1) then shows that:

$$\mathbb{E}_0\{N\} \leq 3 + \log_2(k) \tag{2}$$

The exact same reasoning for building the linear recurrence relation is true as in 2.2, meaning that $\mathbb{E}_i\{N\} = \frac{1}{2}\left(1 + \mathbb{E}_{i+1}\{N\}\right)$. Once again, this recurrence relation can be solved to be $\mathbb{E}_i\{N\} = (\mathbb{E}_0\{N\} - 1)2^{-i} + 1$, So since $\mathbb{E}_0\{N\}$ is proportional to $\mathbb{E}_0\{N\}$ for any i, the upper bound for $\mathbb{E}_0\{N\}$ can be used in the expression for $\mathbb{E}_i\{N\}$. Thus:

$$\mathbb{E}_i\{N\} \leq (3 + \log_2(k) - 1)2^{-i} + 1 = 1 + \left(\frac{1}{2}\right)^i (2 + \log_2(k)) \tag{3}$$

where the range of i is $0 \leq i \leq c$. But $(1/2)^i$ clearly has its max value at $i = 0$ in which case $\mathbb{E}_0\{N\} \leq 3 + \log_2(k)$ as shown before. Thus, $\mathbb{E}_i\{N\} \leq 3 + \log_2(k)$ is true for all i in the range $0 \leq i \leq c$ and thus certainly, $\mathbb{E}_c\{N\} \leq 3 + \log_2(k)$.

This completes the induction argument and therefore shows by the axiom of induction that $\mathbb{E}_c\{N\} \leq 3 + \log_2(k) \quad \forall k \geq 2$. In the case where $k = 1$, it has been shown in 2.1 that $\mathbb{E}\{N\} = 2 \leq 3 + \log_2(k)$. So in the special case where $k = 1$, run the algorithm in 2.1. This completes the proof.

■

3 Convergence in Probability

An equality which will be used in this question is shown below. This equality can easily be shown by induction since the base case is $n = 1$ which is trivial and the induction step is simple arithmetic.

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 + 2\left(\sum_{i=1}^{n}\sum_{j=1}^{i-1} x_i x_j\right) \tag{1}$$

Let $X = (x_1, \cdots, x_d)$ be the vector from the origin to the uniformly random point in the d-dimensional hypercube. Clearly $x_i \sim U[-1, 1] \quad \forall i.1 \leq i \leq n$. Let $Y = (y_1, \cdots, y_d)$ be the ray-vector ending in one of the vertices of the hypercube such that the angle between X and Y is the smallest angle. This angle is denoted M_d.

From elementary linear algebra, it is then known that:

$$\cos(M_d) = \frac{X \cdot Y}{||X||\,||Y||} = \frac{\sum_{i=1}^{d} x_i y_i}{\sqrt{\sum_{i=1}^{d} x_i^2}\sqrt{\sum_{i=1}^{d} y_i^2}} \tag{2}$$

But since Y is the vector to one of the vertices, $y_i \in \{-1, 1\} \quad \forall i.1 \leq i \leq d$, meaning that:

$$\sqrt{\sum_{i=1}^{d} y_i^2} = \sqrt{\sum_{i=1}^{d} 1} = \sqrt{d} \tag{3}$$

It is then trivially true that:

$$\sqrt{\sum_{i=1}^{d} y_i^2} \xrightarrow{p} \sqrt{d} \tag{4}$$

Now, to compute the probabilistic convergence of the other term in the denominator, **Chebyshev's inequality** can be used. But to use Chebyshev's inequality, the mean and variance of a related expression first have to be found.

Consider first the mean of x_i^2 which is computed below using linearity of integrals and the symmetry of x^2 about 0.

$$\mathbb{E}\left\{x_i^2\right\} = \int_{-1}^{1} x^2 \cdot \frac{1}{2} dx = \frac{1}{2} \cdot 2 \int_{0}^{1} x^2 dx = \left[\frac{1}{3x^3}\right]_{x=0}^{x=1} = \frac{1}{3} \tag{5}$$

Thus, since x_i are i.i.d. $\forall i.1 \leq i \leq d$, it is true that:

$$\mathbb{E}\left\{\frac{1}{d}\sum_{i=1}^{d} x_i^2\right\} = \frac{1}{d} \cdot d \cdot \frac{1}{3} = \frac{1}{3} \tag{6}$$

The variance of the same expression can be simplified using the well known identity for variance and linearity of expectation.

$$\mathbb{V}\left\{\frac{1}{d}\sum_{i=1}^{d} x_i^2\right\} = \mathbb{E}\left\{\frac{1}{d^2}\left(\sum_{i=1}^{d} x_i^2\right)^2\right\} - \left(\frac{1}{3}\right)^2 = \frac{1}{d^2} \cdot \mathbb{E}\left\{\left(\sum_{i=1}^{d} x_i^2\right)^2\right\} - \frac{1}{9} \tag{7}$$

Now, to simplify this expression, (1), linearity of expectation, and the formula for an arithmetic series can be used:

$$\mathbb{E}\left\{\left(\sum_{i=1}^{d} x_i^2\right)^2\right\} = \mathbb{E}\left\{\sum_{i=1}^{d} x_i^4\right\} + 2\mathbb{E}\left\{\sum_{i=1}^{d}\sum_{j=1}^{i-1} x_i^2 x_j^2\right\} \tag{8}$$

$$= d \cdot \int_{-1}^{1} x^4 \cdot \frac{1}{2} dx + 2\sum_{i=1}^{d}\sum_{j=1}^{i-1} \frac{1}{9} \tag{9}$$

$$= d \cdot \left[\frac{1}{5x^5}\right]_{x=0}^{x=1} + \frac{2}{9}\sum_{i=1}^{d}\sum_{j=1}^{i-1} 1 \tag{10}$$

$$= \frac{d}{5} + \frac{2}{9} \cdot \frac{d(0 + (d-1))}{2} = \frac{d}{5} + \frac{d^2}{9} - \frac{d}{9} \tag{11}$$

So (7) can be simplified using (11) as follows:

$$\mathbb{V}\left\{\frac{1}{d}\sum_{i=1}^{d} x_i^2\right\} = \frac{1}{d^2}\left(\frac{d}{5} + \frac{d^2}{9} - \frac{d}{9}\right) - \frac{1}{9} = \frac{4}{45d} \tag{12}$$

Thus, Chebyshev's inequality gives:

$$\mathbb{P}\left\{\left|\frac{1}{d}\sum_{i=1}^{d} x_i^2 - \frac{1}{3}\right| \geq \epsilon\right\} \leq \frac{\mathbb{V}\left\{\frac{1}{d}\sum_{i=1}^{d} x_i^2\right\}}{\epsilon^2} = \frac{4}{45d\epsilon^2} \quad \forall \epsilon \geq 0 \tag{13}$$

Which means that:

$$\frac{1}{d}\sum_{i=1}^{d} x_i^2 \xrightarrow{p} \frac{1}{3} \tag{14}$$

which can be rearranged using the properties of converged in probability in convergence of fractions given in the question to get:

$$\sqrt{\sum_{i=1}^{d} x_i^2} \xrightarrow{p} \sqrt{\frac{d}{3}} \tag{15}$$

At this point, only the convergence of probability of the numerator of (2) remains to be computed. This computation is simplified by the observation that $\text{sgn}(x_i) = \text{sgn}(y_i) \quad \forall i.1 \leq i \leq d$ since for Y to be the ray which forms the smallest angle with X, both have to be in the same "hyperquadrant" of the hypercube. But since $y_i \in \{-1, 1\}$, this then means that $x_i y_i = |x_i| \quad \forall i 1 \leq i \leq d$. But since $x_i \sim U[-1, 1]$, this then means that $|x_i| \sim U[0, 1]$. Notice that $|x_i|$ are all i.i.d. $U[0, 1]$ R.V.'s.

Let $Z_i \sim U[0, 1] \quad \forall i.1 \leq i \leq d$ and consider the following expectation:

$$\mathbb{E}\left\{\frac{1}{d}\sum_{i=1}^{d} Z_i\right\} = \frac{1}{d}\sum_{i=1}^{d} \mathbb{E}Z_i = \frac{1}{d} \cdot \frac{d}{2} = \frac{1}{2} \tag{16}$$

Consider the variance for the same sum, which is computed analogously to equations (7) to (11):

$$\mathbb{V}\left\{\frac{1}{d}\sum_{i=1}^{d} Z_i\right\} = \frac{1}{d^2}\mathbb{E}\left\{\sum_{i=1}^{d} Z_i\right\} - \frac{1}{4} \tag{17}$$

$$= \frac{1}{d^2}\left(\sum_{i=1}^{d} \mathbb{E}Z_i^2 + 2\sum_{i=1}^{d}\sum_{j=1}^{i-1} Z_i Z_j\right) - \frac{1}{4} \tag{18}$$

$$= \frac{1}{d^2}\left(\frac{d}{3}\frac{d^2-d}{4}\right) - \frac{1}{4} = \frac{1}{12d} \tag{19}$$

So, once again using Chebyshev's inequality:

$$\mathbb{P}\left\{\left|\frac{1}{d}\sum_{i=1}^{d} Z_i - \frac{1}{2}\right| \geq \epsilon\right\} \leq \frac{\mathbb{V}\left\{\frac{1}{d}\sum_{i=1}^{d} Z_i\right\}}{\epsilon^2} = \frac{1}{12d\epsilon^2} \quad \forall \epsilon \geq 0 \quad \Rightarrow \quad \sum_{i=1}^{d} Z_i \xrightarrow{p} \frac{d}{2} \tag{20}$$

But since $Z_i \sim |x_i| = x_i y_i \quad \forall i.1 \leq i \leq d$, (20) gives that

$$\sum_{i=1}^{d} x_i y_u \xrightarrow{p} \frac{d}{2} \tag{21}$$

Combining the three convergences in probability computed so far - equations (4), (14), and (21) - together with the properties of convergence in probability given in the question, the convergence in probability of (2) can be computed:

$$\frac{\sum_{i=1}^{d} x_i y_i}{\sqrt{\sum_{i=1}^{d} x_i^2}\sqrt{\sum_{i=1}^{d} y_i^2}} \xrightarrow{p} \frac{\frac{d}{2}}{\sqrt{d}\sqrt{\frac{d}{3}}} = \frac{\sqrt{3}}{2} \tag{22}$$

But since arccos is a continuous function on $[-1, 1]$, it is then true from equation (2) and (22) that:

$$M_d \xrightarrow{p} \arccos\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{6} \tag{23}$$

Which completes the proof.

$\blacksquare$

Assignment 7 COMP690

Gulliver Häger and Tommy He

November 2022

Contents

1 A Special DeBruijn Graph

1.1 Ergodicity and Uniform Stationary Distribution

Clearly, the given Markov chain is aperiodic since there exists states such as $(0, 0, \cdots, 0, 0)$ with self-loops with non-zero probability. It is also true that the given Markov chain is irreducible, since given any state for X_i, all states are equiprobable in X_{i+n}, i.e. $\mathbb{P}(X_{i+n} = s) = 2^{-n} \quad \forall s \in S$, meaning that with non-zero probability all states are reachable given enough time. Since $|S| = 2^n < \infty$, we can also conclude that the given Markov Chain is ergodic.

Left to show is that the stationary distribution is uniform. To do this consider any state $s \in S$ where s will be of the form $s = (a_1, a_2, \cdots, a_n) \quad a_i \in \{0, 1\}$. By definition of the Markov chain, there is exactly two other states $s_0, s_1 \in S$ which has non-zero probability of transitioning to a. These two states are $S_0 = (0, a_1, a_2, \cdots, a_{n-1})$ and $s_1 = (1, a_1, a_2, \cdots, a_{n-1})$. By definition, we have that $p_{s_0 s} = p_{s_1 s} = 1/2$. By definition of the stationary distribution, we must have that:

$$\pi_s = \sum_{s' \in S} \pi_{s'} p_{s's} = \pi_{s_0} \frac{1}{2} + \pi_{s_1} \frac{1}{2} \quad \forall s \in S \tag{1}$$

We can see that letting $\pi_s = \pi_{s_0} = \pi_{s_1} = 2^{-n}$ satisfies the above equation for every $s \in S$. But the stationary distribution is unique for any ergodic Markov chain, we know that $\pi_s = 2^{-n} \quad \forall s \in S$.

■

1.2 Upper Bound for Maximum Hitting Times

We will show that given any $x, y \in S$, the expected hitting time is upper bounded by 2^n, meaning that $E[T_{xy}] = O(2^n) \quad \forall x, y \in S$ which certainly proves the desired claim.

So fix $x, y \in S$ and let $X_0 = x$. Notice that for any given X_i, there is a probability of 2^{-n} to reach y in X_{i+n}. So consider the new random variable $Y_j = \mathbb{1}_{\{X_{j \cdot n} = y\}} \quad j \in \mathbb{N}$. By definition of Y_j, if $Y_j = 1$, then we have reached y in the original Markov chain. But Y_j is clearly distributed as Bernoulli(2^{-n}), meaning that the we can upper bound the $\mathbb{E}[T_{xy}]$ by the expected time it takes to see the first success in $\{Y_1, Y_2, Y_3, \cdots\}$. Thus we have deduced that $\mathbb{E}[T_{xy}]$ is upper bounded by a geometric random variable with success probability 2^{-n}, giving that:

$$\mathbb{E}[T_{x,y}] \leq 2^n \quad \forall x, y \in S \tag{1}$$

And thus we can conclude that:

$$\max_{x,y \in S} \mathbb{E}[T_{xy}] = O(2^n) \tag{2}$$

■

1.3 Upper Bound For Covering Time

By Mathew's inequality and using part 2 of this question and letting $A = S$, we have that:

$$\mathbb{E}[T_i] \leq O(2^n) \cdot \log(|S|) = O(2^n) \cdot \log(2^n) = O(n2^n) \quad \forall i \in S \tag{1}$$

where T_i is the covering time for S starting from $i \in S$.

■

1.4 Lower Bound For Covering Time

Let:

$$A = \{(a_1, \cdots, a_n) \in S \mid a_1, \cdots, a_{n-\log(n)-1} \in \{0,1\},\ a_{n-\log(n)}, \cdots, a_{n-1} = 0, a_n = 1\} \tag{1}$$

To prove the lower bound for the covering time, we apply Matthew's on this set and notice that $|A| = 2^{n-\log(n)}$, meaning that:

$$\log(|A|) = \log(2)(n - \log(n)) = \Omega(n) \tag{2}$$

Fix $x, y \in A$. The minimum hitting time between any two sequences in this set can be upper bounded by the probability that we hit y going from x in less than n steps in the Markov chain, plus the probability that we hit y going from x in more than n steps. Notice by the nature of the strings in A, there are only $n/\log(n)$ possible points of getting to y in less than n steps. If we do not hit y within n steps, we have reached a uniform distribution on S by part 1. Thus we have that:

$$\mathbb{E}[T_{xy}] \leq n \cdot \mathbb{P}(T_{xy} \leq n) + 2^n \cdot \mathbb{P}(T_{xy} > n) \tag{3}$$

To proceed, we bound $P(T_{xy} \leq n)$ and get the desired minimum hitting time since we have uniformly bounded the hitting times within A from below to get that $\mu_- = \Omega(2^n)$

Since we have to cover at least A in order to cover all of A, we get by Mathews inequality:

$$\mu_-(A) \cdot \log(|A|) = \Omega(n2^n) \leq \mathbb{E}[T_i] \quad \forall i \in S \tag{4}$$

■

2 Time-Reversible Markov Chains

For this question, sub-questions 2 and 4 were taken off the assignment by Luc after finding an issue in the formulation of them. Below are the remaining two sub-questions.

2.1 Equality in Distribution

Let $(i_0, \cdots, i_t) \in S^t$. Then we have by the Markov property that and the fact that X_0 has the distribution π:

$$\mathbb{P}((X_0, \cdots X_t) = (i_0, \cdots, i_t)) = \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{t-1} i_t} \tag{1}$$

But by time reversibility we have that $\pi_k p_{kj} = \pi_j p_{jk}$ for any $j, k \in S$, meaning that $\pi_{i_0} p_{i_0 i_1} = \pi_{i_1} p_{i_1 i_0}$ which can be used to rewrite equation (1) as:

$$\begin{aligned} \mathbb{P}((X_0, \cdots X_t) = (i_0, \cdots, i_t)) = (\pi_{i_0} p_{i_0 i_1}) p_{i_1 i_2} \cdots p_{i_{t-1} i_t} &= (\pi_{i_1} p_{i_1 i_0}) p_{i_1 i_2} \cdots p_{i_{t-1} i_t} &(2)\\ &= p_{i_1 i_0} (\pi_{i_1} p_{i_1 i_2}) \cdots p_{i_{t-1} i_t} &(3) \end{aligned}$$

Applying this property recursively and reordering the terms then gives that:

$$\begin{aligned} \mathbb{P}((X_0, \cdots X_t) = (i_0, \cdots, i_t)) &= p_{i_1 i_0} p_{i_2 i_1} \cdots p_{i_t i_{t-1}} \pi_t &(4)\\ &= \pi_t p_{i_t i_{t-1}} p_{i_{t-1} i_{t-2}} \cdots p_{i_1 i_0} &(5) \end{aligned}$$

But we know that by the Markov property:

$$\mathbb{P}((X_0, \cdots, X_t) = (i_t, i_{t-1}, \cdots, i_0)) = \pi_t p_{i_t i_{t-1}} p_{i_{t-1} i_{t-2}} \cdots p_{i_1 i_0} \tag{6}$$

meaning that:

$$\mathbb{P}((X_0, \cdots X_t) = (i_0, \cdots, i_t)) = \mathbb{P}((X_0, \cdots, X_t) = (i_t, i_{t-1}, \cdots, i_0)) \tag{7}$$

which by relabeling gives that:

$$\mathbb{P}((X_0, \cdots X_t) = (i_0, \cdots, i_t)) = \mathbb{P}((X_t, \cdots X_0) = (i_0, \cdots, i_t)) \tag{8}$$

But since $(i_0, \cdots, i_t) \in S^t$ was picked arbitrarily, this is equivalent to the claim we want to show:

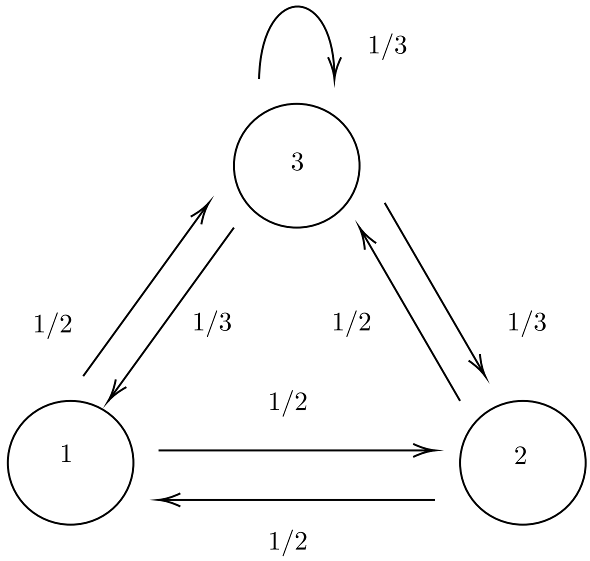
$$(X_0, \cdots, X_t) \stackrel{\mathcal{L}}{=} (X_t, \cdots, X_0) \tag{9}$$

$\blacksquare$

2.2 A Counterexample

Below is a diagram of the Markov chain defined by the transition matrix P.

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \tag{1}$$



First, we find the stationary distribution by solving the equation $\pi = P\pi$. This equation gives us the system of equations:

$$\pi_1 = \frac{1}{2}\pi_2 + \frac{1}{3}\pi_3 \tag{2}$$
$$\pi_2 = \frac{1}{2}\pi_3 + \frac{1}{3}\pi_3 \tag{3}$$
$$\pi_3 = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{3}\pi_3 \tag{4}$$
$$1 = \pi_1 + \pi_2 + \pi_3 \tag{5}$$

which we easily can solve to get that:

$$\pi = (\pi_1, \pi_2, \pi_3) = \left(\frac{2}{7}, \frac{2}{7}, \frac{3}{7}\right) \tag{6}$$

Now, we will find the expected hitting times between states 1 and 3. Notice that states 1 and 2 are going to behave completely symmetrically and thus we already know that $\mathbb{E}[T_{12}] = \mathbb{E}[T_{21}]$, $\mathbb{E}[T_{13}] = \mathbb{E}[T_{23}]$, and $\mathbb{E}[T_{31}] = \mathbb{E}[T_{32}]$. So we can compute $\mathbb{E}[T_{13}]$ using this:

$$\mathbb{E}[T_{13}] = \frac{1}{2} + \frac{1}{2}\left(1 + \frac{1}{2}\mathbb{E}[T_{23}]\right) = 1 + \frac{1}{2}\mathbb{E}[T_{13}] \tag{7}$$

and thus we get that $\mathbb{E}[T_{13}] = 2$.

In order to computer $\mathbb{E}[T_{31}]$ we first computer $\mathbb{E}[T_{21}]$ since the former depends on the latter.

$$\mathbb{E}[T_{21}] = \frac{1}{2} + \frac{1}{2}\left(1 + \mathbb{E}[T_{31}]\right) = 1 + \frac{1}{2}\mathbb{E}[T_{31}] \tag{8}$$

Proceeding with $\mathbb{E}[T_{31}]$:

$$\mathbb{E}[T_{31}] = \frac{1}{3} + \frac{1}{3}\left(1 + \mathbb{E}[T_{21}]\right) + \frac{1}{3}\left(1 + \mathbb{E}[T_{31}]\right) \tag{9}$$
$$= \frac{4}{3} + \frac{1}{2}\mathbb{E}[T_{31}] \tag{10}$$

From which we get that $\mathbb{E}[T_{31}] = \frac{8}{3}$.

Thus we can conclude that:

$$\pi_1 \cdot \mathbb{E}[T_{13}] = \frac{4}{7} \tag{11}$$

$$\pi_3 \cdot \mathbb{E}[T_{31}] = \frac{8}{7} \tag{12}$$

which are clearly not equal, meaning that the given Markov chain is a counterexample to the statement that $\pi_i \cdot \mathbb{E}[T_{ij}] = \pi_j \cdot \mathbb{E}[T_{ji}]$.

$\blacksquare$

3 Eigenvalues

3.1 Magnitude of Eigenvalues is at most 1

As we have seen in class, π is an eigenvector with eigenvalue 1. Suppose for a contradiction that there exists some non-zero row vector $\mathbf{x}$ which is an eigenvector with eigenvalue $|\lambda| > 1$. By definition we know that:

$$\mathbf{x}P^n = |\lambda|^n \mathbf{x} \tag{1}$$

Which means that for big enough n, all non-zero coordinates of $|\lambda|^n\mathbf{x}$ are strictly greater than 1, which by the above equality then means that all non-zero coordinates of $\mathbf{x}P^n$ are strictly greater than 1. Since it was assumed that $\mathbf{x}$ is a non-zero vector, this means that there exists at least one such coordinate. But every coordinate in $\mathbf{x}P^n$ is a convex combination of the coordinates in $\mathbf{x}$ with coefficients from the column in P^n, meaning that there must exist an entry in every column in P^n which is strictly greater than 1. But this clearly contradicts the fact that P^n is a transition matrix itself (as seen in class) since there will exist rows in P^n which sum to a number strictly greater than 1.

Thus we can concluded that any eigenvalue has to have magnitude at most 1.

■

3.2 Eigenvalues in Lazy Matrix are Real and Nonnegative

Let λ be the eigenvalues for the lazy version. Then characteristic equation for the eigenvalues is given by

$$\left|\frac{P+I}{2} - \lambda I\right| = 0$$
$$\Rightarrow |P - (2\lambda - 1)I| = 0$$

For 3.1, let λ' be the eigenvalues of the normal chain. They corresponded to the characteristic equation $|P - \lambda' I| = 0$. Notice for a solution λ', we must have $\lambda' = 2\lambda - 1$. If λ' lies in $[-1, 1]$, then the lazy λ will lie in $\frac{[-1,1]+1}{2} = [0, 1]$, which is indeed real and nonnegative.

■

4 Random Walk on an Odd Cycle

We want to find eigenvalues and eigenvectors λ_j, v_j such that $Pv_j = \lambda_j v_j$. When we expand out P, we find it to be

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \cdots & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \cdots & 0 & 0 \\ \vdots & & & \ddots & & & \\ \frac{1}{2} & 0 & 0 & 0 & \cdots & \frac{1}{2} & 0 \end{pmatrix}$$

where $P_{ij} = \frac{1}{2}$ if and only if $|i-j| = 1$. When we attempt to verify the eigenvalues and eigenvectors given, we get the equation

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \cdots & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \cdots & 0 & 0 \\ \vdots & & & \ddots & & & \\ \frac{1}{2} & 0 & 0 & 0 & \cdots & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} \cos\left(\frac{2\pi 0 j}{n}\right) \\ \cos\left(\frac{2\pi 1 j}{n}\right) \\ \cos\left(\frac{2\pi 2 j}{n}\right) \\ \vdots \\ \cos\left(\frac{2\pi(n-1)j}{n}\right) \end{pmatrix} = \cos\left(\frac{2\pi j}{n}\right) \begin{pmatrix} \cos\left(\frac{2\pi 0 j}{n}\right) \\ \cos\left(\frac{2\pi 1 j}{n}\right) \\ \cos\left(\frac{2\pi 2 j}{n}\right) \\ \vdots \\ \cos\left(\frac{2\pi(n-1)j}{n}\right) \end{pmatrix}$$

for the $j$the eigenvector. For the kth column, this exands to verifying

$$\cos\left(\frac{2\pi kj}{n}\right)\cos\left(\frac{2\pi j}{n}\right) = \frac{1}{2}\cos\left(\frac{2\pi(k-1)j}{n}\right) + \frac{1}{2}\cos\left(\frac{2\pi(k+1)j}{n}\right)$$

by using the cyclic nature of cos every 2π. However, notice this is true by the sum to product formula for cos, and so these are indeed our eigenvectors and eigenvalues. Since there are n of them, they form the entire eigenspace, and we are done. The second highest eigenvalue, in terms of absolute value, will be the one closest to 1 or -1. This will be given by

$$\cos\left(\frac{2\pi(n-1)/2}{n}\right) = \cos\left(\frac{(n-1)\pi}{n}\right)$$

and so the spectral gap will be

$$1 - \left|\cos\left(\frac{(n-1)\pi}{n}\right)\right|$$

and the relaxation time

$$\frac{1}{1-\cos\left(\frac{(n-1)\pi)}{n}\right)}$$

We can conclude from the relaxation time an upper bound of

$$\frac{\log\left(\frac{e}{\pi_{\min}}\right)}{1-\cos\left(\frac{(n-1)\pi)}{n}\right)} = \frac{\log(ne)}{1-\cos\left(\frac{(n-1)\pi)}{n}\right)}$$

for the mixing time since the stationary vector is uniform for a cycle.

USE COUPLING FOR IMPROVEMENT

5 Random Walk on a Grid

5.1 Lower Bound

Let $A' = \{1, \cdots, \lfloor \frac{n}{2} \rfloor\} \times n \times n$ be a subset of the torus. Then, $\pi_{A'}$, the probability that the random walk is in A', equals 1 by symmetry. We know that

$$\begin{aligned} \tau &\geq \max_{\forall A, \forall u \notin A, A: \pi_A > 1/(2e)} \left\lfloor (\pi_A - \frac{1}{2e}) \mathbb{E} T_{u,A} \right\rfloor \\ &\geq \max_{u \notin A'} \left\lfloor (\pi'_A - \frac{1}{2e}) \mathbb{E} T_{u,A'} \right\rfloor \end{aligned}$$

Let u be a midpoint on A'^c i.e. $(\frac{3}{4}n, \_, \_)$ and create a markov chain where you change states if and only if you change the x-value. Then, we get a lazy chain where each state has self loop probability of $\frac{10}{12}$ and transition probability of $\frac{1}{12}$ for going towards positive or negative. We want the expected time before it hits the yz-plane of 0 or $\frac{n}{2}$. Since it's a lazy version of the typical chain graph we know, it will only be greater by a constant factor, and so is still $\Omega(n^2)$ because the hitting time of a random walk on the chain graph of length $n/2$ started at the midpoint of the chain graph to the endpoints is $\Omega(n^2)$

■

5.2 Upper Bound Using Coupling

To find an upper found for the mixing time we will use coupling. Let X_i, Y_i be the Markov chain realizations we are coupling. Let them be coupled such that they start as a random walk independently, but as soon as they coincide in one coordinate they transition in the same direction for that coordinate if that coordinate is the coordinate changing for the current transition. Clearly then, the coupling time for the n^3 torus can then be upper bounded by the product of the coupling times of X_i and Y_i in each of their dimensions. But since we are on the torus, each coordinates coupling is going to correspond to a coupling on a C_n graph. Using the exact same coupling method to get the upper bound for mixing time of the chain graph, with slight modifications, we get that the mixing time on the C_n graph is upper bounded by n^2. Thus, by the previous reduction, we get that:

$$\tau \leq O(3 \cdot n^3) = O(n^3) \tag{1}$$

■

COMP 690 Assignment 8

Massimo Vicenzo, Caelan Atamanchuk, Gulliver Häger and Tommy He

5^{th} December 2022

Note: Conditional Entropy

For the assignment to make calculations much simpler, we use the definition of conditional entropy, which we derive in the following way:

$$\begin{aligned}
\varepsilon(X,Y) &= \sum_{(i,j)} \mathbb{P}(X=i,Y=j)\log_2(\frac{1}{\mathbb{P}(X=i,Y=j)}) \\
&= \sum_{(i,j)} \mathbb{P}(X=i,Y=j)\log_2(\frac{1}{\mathbb{P}(Y=j|X=i)\mathbb{P}(X=i)}) \\
&= \sum_{(i,j)} \mathbb{P}(X=i,Y=j)\log_2(\frac{1}{\mathbb{P}(Y=j|X=i)}) + \sum_{(i,j)} \mathbb{P}(X=i,Y=j)\log_2(\frac{1}{\mathbb{P}(X=i)}) \\
&= \varepsilon(Y|X) + \sum_{(i,j)} \mathbb{P}(Y=j|X=i)\mathbb{P}(X=i)\log_2(\frac{1}{\mathbb{P}(X=i)}) \\
&= \varepsilon(Y|X) + \sum_{i} \mathbb{P}(X=i)\log_2(\frac{1}{\mathbb{P}(X=i)}) \sum_{j} \mathbb{P}(Y=j|X=i) \\
&= \varepsilon(Y|X) + \sum_{i} \mathbb{P}(X=i)\log_2(\frac{1}{\mathbb{P}(X=i)})(1) \\
&= \varepsilon(Y|X) + \varepsilon(X)
\end{aligned}$$

With this, we will succinctly prove the next two questions.

1 Entropy Calculation

First we relate the entropy of $(X_0, X_1, \ldots, X_i)$ to that of $(X_0, X_1, \ldots, X_{i-1})$:

$$\begin{aligned}\varepsilon(X_0, X_1, \ldots, X_i) &= \varepsilon(X_i|X_0, X_1, \ldots, X_{i-1}) + \varepsilon(X_0, X_1, \ldots, X_{i-1}) \\ &= \varepsilon(X_i|X_{i-1}) + \varepsilon(X_0, X_1, \ldots, X_{i-1}) \\ &= \varepsilon(X_1|X_0) + \varepsilon(X_0, X_1, \ldots, X_{i-1}) \\ &= (\varepsilon(X_0, X_1) - \varepsilon(X_0)) + \varepsilon(X_0, X_1, \ldots, X_{i-1})\end{aligned}$$

The second equality comes from the fact that $\mathbb{P}(X_i = s_i|X_{i-1} = s_{i-1}, \ldots, X_0 = s_0) = \mathbb{P}(X_i = s_i|X_{i-1} = s_{i-1})$. This is due to the fact that conditioning on more than just the previous state in the Markov chain is useless for determining where it is.

The third equality comes from noticing that since X_0 is distributed like the stationary vector π, then so are all the following states, as per the property of the stationary distribution: $\pi_j = \sum_i p_{ij}\pi_i$. Because of this, $\mathbb{P}(X_i = s|X_{i-1} = t) = \mathbb{P}(X_1 = s|X_0 = t)$.

Now, putting it all together,

$$\varepsilon(X_0, X_1, \ldots, X_t) = \varepsilon(X_0, X_1) + \sum_{i=2}^{t} \varepsilon(X_0, X_1) - \varepsilon(X_0)) = t\varepsilon(X_0, X_1) - (t-1)\varepsilon(X_0)$$

To show that $\varepsilon(X) \leq \varepsilon(Y)$, first notice that, as Y is made up of t independent variables all distributed like X_0: $\varepsilon(Y) = (t+1)\varepsilon(X_0)$. So we need to just show that:

$$\begin{aligned}\varepsilon(X) \leq \varepsilon(Y) &\iff t\varepsilon(X_0, X_1) - (t-1)\varepsilon(X_0) \leq (t+1)\varepsilon(X_0) \\ &\iff t\varepsilon(X_0, X_1) \leq 2t\varepsilon(X_0) \\ &\iff \varepsilon(X_0, X_1) \leq 2\varepsilon(X_0)\end{aligned}$$

Using conditional entropy: $\varepsilon(X_0, X_1) = \varepsilon(X_0) + \varepsilon(X_1|X_0)$. So we only need to show that $\varepsilon(X_1|X_0) \leq \varepsilon(X_0)$. We do so by expanding them and showing that the difference is greater than 0.

$$\begin{aligned}\varepsilon(X_0) - \varepsilon(X_1|X_0) &= \sum_i \mathbb{P}(X_0 = i)\log_2(\frac{1}{X_0 = i}) - \sum_{i,j} \mathbb{P}(X_0, X_1 = i, j)\log_2(\frac{\mathbb{P}(X_0 = i)}{\mathbb{P}(X_0, X_1 = i, j)}) \\ &= \sum_{i,j} \mathbb{P}(X_1 = j|X_0 = i)\mathbb{P}(X_0 = i)\log_2(\frac{1}{\mathbb{P}(X_0 = i)}) - \sum_{i,j} \mathbb{P}(X_0, X_1 = i, j)\log_2(\frac{\mathbb{P}(X_0 = i)}{\mathbb{P}(X_0, X_1 = i, j)}) \\ &= \sum_{i,j} \mathbb{P}(X_0, X_1 = i, j)\log_2(\frac{\mathbb{P}(X_0, X_1 = i, j)}{\mathbb{P}(X_0 = i)^2}) \geq 0\end{aligned}$$

The last step comes from noticing that we have the Kullback-Leibler divergence, which is always greater than 0.

2 Guessing A Random Pair of Integers

2.1 (i)

$$
\begin{aligned}
&\mathcal{E}(X)+\mathcal{E}(Y)-\mathcal{E}(X,Y)=\mathcal{E}(Y)-\mathcal{E}(Y\mid X)\\
&\qquad=\sum_j \mathbb{P}(Y=j)\log\left(\frac{1}{\mathbb{P}(Y=j)}\right)-\sum_{(i,j)}\mathbb{P}((X,Y)=(i,j))\log\left(\frac{\mathbb{P}(X=i)}{\mathbb{P}((X,Y)=(i,j))}\right)\\
&\qquad=\sum_i\left(\sum_j \mathbb{P}((X,Y)=(i,j))\log\left(\frac{1}{\mathbb{P}(Y=j)}\right)\right)-\sum_{(i,j)}\mathbb{P}((X,Y)=(i,j))\log\left(\frac{\mathbb{P}(X=i)}{\mathbb{P}((X,Y)=(i,j))}\right)\\
&\qquad=\sum_{(i,j)}\mathbb{P}((X,Y)=(i,j))\left(\log\left(\frac{1}{\mathbb{P}(Y=j)}\right)-\log\left(\frac{\mathbb{P}(X=i)}{\mathbb{P}((X,Y)=(i,j))}\right)\right)\\
&\qquad=\sum_{(i,j)}\mathbb{P}((X,Y)=(i,j))\left(\log\left(\frac{\mathbb{P}((X,Y)=(i,j))}{\mathbb{P}(X=i)\mathbb{P}(Y=j)}\right)\right)\\
&\qquad=\mathrm{KL}((X,Y),XY)\\
&\qquad\geq 0
\end{aligned}
$$

2.2 (ii)

Our algorithm will create a Huffman tree for X, and then we proceed to ask questions using the tree, which has an upper bound of $\varepsilon(X)+1$ from class. After this, we have now determined X and so we will create the Huffman tree for Y given the value we found for X.

Let Q_Z for a random variable Z be the number of questions of the form "is $Z\in A$?" for guessing Z. Then, notice that

$$\mathbb{E}Q_{(X,Y)}=\mathbb{E}Q_X+\mathbb{E}Q_{(Y|X)}$$

since if we know X, then we have a new conditional random variable given by $Y\mid X$ for Y. From the definition of conditional entropy from above, we can find

$$\mathcal{E}(Y\mid X)=\mathbb{E}\log\frac{1}{p(y\mid x)}$$

which corresponds precisely to the entropy of the conditional random variable given by $Y\mid X$. So, we can analyze the upper bound of this since viewing the entropy of the 2 random variables by

$$\mathbb{E}Q_X\leq\mathcal{E}(X)+1$$

$$\mathbb{E}Q_{(Y|X)}\leq\mathcal{E}(Y\mid X)+1$$

with the bounds given by Shannon. Adding gives

$$\mathbb{E}Q_X+\mathbb{E}Q_{(Y|X)}\leq\mathcal{E}(X)+\mathcal{E}(Y\mid X)+2=\mathcal{E}(X,Y)+2$$

as desired. Algorithmically, we can implement this with Huffman trees for the questionnaires for determining the values of each random variable with bounded questions. First, we would construct a Huffman tree for X, then ask at most $\mathcal{E}(X)+1$ questions to find its value x_0. With its value, we construct a Huffman tree for $Y\mid X=x_0$, then ask at most $\mathcal{E}(Y\mid X=x_0)+1$ questions to our desired values.

■

2.3 (iii)

Proceeding exactly as in the previous question, we determine the $(X_1,\cdots,X_d)$ one by one given the already known random variables, i.e. at every step of our algorithm, we are going to use Huffman trees to determine $(X_i|X_{i-1},\cdots,X_1)$ with less than $\mathcal{E}(X_i|X_{i-1},\cdots,X_1)+1$ number of questions in expectation. Applying the previous part of this question inductively and using the fact that $\mathcal{E}(X, Y) = \mathcal{E}(X) + \mathcal{E}(Y|X)$ just as in the previous part. Thus we get by induction:

$$\begin{aligned} \mathbb{E}[Q_{(X_1,\cdots,X_d)}] &\le \mathbb{E}[Q_{X_1}] + E[Q_{((X_2,\cdots,X_d)|X_1)}] && (1)\\ &\le \mathcal{E}(X_1) + 1 + \mathcal{E}(((X_2,\cdots,X_d)|X_1)) + (d-1) && (2)\\ &= \mathcal{E}((X_1,\cdots,X_d)) + d && (3) \end{aligned}$$

■