

Random Target Lemma

Sucharita Jayanti

February 21, 2013

Abstract

The well known *Random Target Lemma* states that if one starts at a node i in a connected graph, selects a target node j at random (according to the probability distribution given by the stationary distribution) and performs a random walk until j is first reached, the expected length of the random walk is independent of the node i at which the walk is started. This paper, written as part of a class project for Dartmouth's Math 100 course, presents a proof of this lemma. This proof is the same as the one in [1], but is rephrased and restructured to make the notation more uniform and the proof more readable and understandable.

1 Introduction

Imagine a graph, where the ten most common Paris tourist attractions (Eiffel Tower, Louvre, Notre Dame *etc*) are the nodes and the metro connections between them are the edges. Assume, as is the case, that this graph is connected. Now, suppose a tourist, Bob, has this list of attractions in Paris and wishes to stay at a hotel near one of them. Bob, as a newcomer to Paris, is naturally concerned about the number of subway connections it would take for him to visit all the attractions on his list, and wishes to minimize this number. Each morning, he randomly chooses a destination (based on the steady state probabilities of the attractions) and, since he lacks a metro map, embarks on a random walk to it. The Random Target Lemma saves Bob a lot of trouble in choosing his hotel by guaranteeing that the expected number of subway hops that he will end up taking each day to reach his destination is the same, regardless of which attraction he chooses to stay near!

In the following sections, we will introduce helpful notation and definitions, establish lemmas useful for the proof of the Random Target Lemma, and ultimately, prove the Random

Target Lemma. This proof mimics the one in [1], but is rephrased and restructured to make the notation more uniform and the proof more readable and understandable.

2 Notation and Definitions

Definition 1. *If j is a node in a connected graph. T_j denotes any random walk that terminates the first time node j is reached; in particular, if the start node is j , then T_j is empty (since j is reached without taking any steps).*

Definition 2. *Let G be a connected graph and ρ be a probability distribution on V . Then:*

- $E_i S$ denotes the expected number of steps in a random walk S that starts at node i .
- $E_i N_j S$ denotes the expected number of visits to node j in a random walk S that starts at node i , not counting the last visit to j if j is the last node on S .
- $E_\rho S = \sum_{i \in V} \rho(i) E_i S$
- $E_\rho N_j S = \sum_{i \in V} \rho(i) E_i N_j S$

For an illustration of the above definition, $E_i T_j$ is the expected number of steps in a random walk that starts at node i , and terminates the first time node j is reached. The following observation is immediate from the definitions.

Observation 1. *For all nodes i and j , $E_i N_j T_j = 0$*

Definition 3. $Z_{ij} = \sum_{t=0}^{\infty} (p_{ij}^{(t)} - \pi_j)$, where $p_{ij}^{(t)}$ is the probability that a random walk goes from node i to node j in exactly t steps and π_j is the steady state probability of node j (and π is the stationary distribution).

3 Auxiliary Lemmas

Lemma 1. *If S is a random walk from i to i , then $E_i N_j S = \pi_j E_i S$.*

We do not prove this lemma, but note that it follows from the Reward-Renewal Theorem.

Lemma 2. $E_i N_i T_j = \pi_i (E_i T_j + E_j T_i)$

Proof. Let S be a random walk that terminates as soon as i is reached *after* the first visit to j . Then, S can be expressed as a concatenation of a random walk T_j that starts from i and a random walk T_i that starts from j . Therefore, we can write $E_i S = E_i T_j + E_j T_i$ and $E_i N_i S = E_i N_i T_j + E_j N_i T_i$. Combining these with $E_i N_i S = \pi_i E_i S$ (from Lemma 1), we have $E_i N_i T_j + E_j N_i T_i = \pi_i (E_i T_j + E_j T_i)$. Since $E_j N_i T_i = 0$ (by Observation 1), we have $E_i N_i T_j = \pi_i (E_i T_j + E_j T_i)$. \square

Lemma 3. Let S be a random walk of length t_0 starting from node i . Then, $E_i N_j S = \sum_{t=0}^{t_0-1} p_{ij}^{(t)}$.

Proof. Let $x(t)$ be an indicator random variable whose value is 1 if S visits j in the t^{th} step, and 0 otherwise. The expected number of visits to j on S is:

$$E_i N_j S = E(\sum_{t=0}^{t_0-1} x(t)) = \sum_{t=0}^{t_0-1} E(x(t)) = \sum_{t=0}^{t_0-1} p_{ij}^{(t)}.$$

\square

Corollary 1. $E_j N_j T_i + \sum_{t=0}^{t_0-1} p_{ij}^{(t)} = \pi_j (E_j T_i + t_0 + E_\rho T_j)$

if ρ is the distribution that results after t_0 steps from node i .

Proof. Let S be the following walk:

1. Start at node j and go on a random walk until node i is first reached.
2. Go on a random walk for t_0 steps.
3. Continue the random walk, if necessary, until the walk first reaches j .

From the definition of S and Lemma 3, the left hand side of the identity in the lemma is $E_j N_j S$.

From Lemma 1, it follows that $E_j N_j S = \pi_j E_j S = \pi_j \times \text{expected \# of steps of } S$
 $= \pi_j (E_j T_i + t_0 + E_\rho T_j)$ \square

Lemma 4. $\pi_i E_\pi T_i = Z_{ii}$

Proof. Let $t_0 \geq 0$ and S be the following walk:

1. Start at i and go on a random walk for t_0 steps.

2. Continue the random walk until the walk first reaches i .

Let ρ be the probability distribution after the first t_0 steps. Then, $E_i S = t_0 + E_\rho T_i$.

By Lemma 1, we have $E_i N_i S = \pi_i E_i S = \pi_i(t_0 + E_\rho T_i)$.

Observe that $E_\rho N_i T_i = 0$, since i is not visited on T_i (being at i at the end does not count as a visit of i). Therefore, $E_i N_i S =$ number of times i is visited during the first t_0 steps, which (by Lemma 3) equals $\sum_{t=0}^{t_0-1} p_{ii}^{(t)}$.

From the relationships concluded above, we have: $\pi_i(t_0 + E_\rho T_i) = \sum_{t=0}^{t_0-1} p_{ii}^{(t)}$. Therefore, $\sum_{t=0}^{t_0-1} (p_{ii}^{(t)} - \pi_i) = \pi_i E_\rho T_i$. Letting $t_0 \rightarrow \infty$ and noting that $\rho \rightarrow \pi$ as $t_0 \rightarrow \infty$, we get $\sum_{t=0}^{\infty} (p_{ii}^{(t)} - \pi_i) = \pi_i E_\pi T_i$. In other words, $Z_{ii} = \pi_i E_\pi T_i$. \square

Lemma 5. $\sum_j Z_{ij} = 0$

Proof.

$$\begin{aligned} \sum_j Z_{ij} &= \sum_j \sum_{t=0}^{\infty} (p_{ij}^{(t)} - \pi_j) && \text{by definition of } Z_{ij} \\ &= \sum_{t=0}^{\infty} \sum_j (p_{ij}^{(t)} - \pi_j) \\ &= \sum_{t=0}^{\infty} \left(\sum_j p_{ij}^{(t)} - \sum_j \pi_j \right) \\ &= \sum_{t=0}^{\infty} (1 - 1) \\ &= 0 \end{aligned}$$

\square

4 Proof of Random Target Lemma

We are finally ready to prove the Random Target Lemma.

Lemma 6 (Random Target Lemma).

$$\sum_j \pi_j E_i T_j = \sum_j Z_{jj}$$

Thus, $\sum_j \pi_j E_i T_j$ is the same for all i .

Proof.

$$\begin{aligned}
E_j N_j T_i + \sum_{t=0}^{t_0-1} p_{ij}^{(t)} &= \pi_j(E_j T_i + t_0 + E_\rho T_j) && \text{By Corollary 1} \\
\Rightarrow \pi_j(E_j T_i + E_i T_j) + \sum_{t=0}^{t_0-1} p_{ij}^{(t)} &= \pi_j(E_j T_i + t_0 + E_\rho T_j) && \text{By Lemma 2} \\
\Rightarrow \sum_{t=0}^{t_0-1} (p_{ij}^{(t)} - \pi_j) &= \pi_j(E_\rho T_j - E_i T_j) \\
\Rightarrow Z_{ij} = \pi_j(E_\rho T_j) - \pi_j(E_i T_j) &&& \text{By Def of } Z_{ij} \text{ and letting } t_0 \rightarrow \infty \\
\Rightarrow Z_{ij} = Z_{jj} - \pi_j E_i T_j &&& \text{By Lemma 4} \\
\Rightarrow Z_{ij} - Z_{jj} = -\pi_j E_i T_j \\
\Rightarrow Z_{jj} - Z_{ij} = \pi_j E_i T_j
\end{aligned}$$

By summing both sides over j , we get:

$$\begin{aligned}
\sum_j \pi_j E_i T_j &= \sum_j (Z_{jj} - Z_{ij}) \\
&= \sum_j Z_{jj} - \sum_j Z_{ij} \\
&= \sum_j Z_{jj} - 0 && \text{By Lemma 5} \\
&= \sum_j Z_{jj}
\end{aligned}$$

Hence we have the Random Target Lemma. □

References

- [1] Aldous, David, and James Fill. "General Markov Chains." *Reversible Markov Chains and Random Walks on Graphs*. 1999. 1-36. Digital file.