A Random Graph Proof for the Irreducible Case of the Markov Chain Tree Theorem

T. I. Fenner and T. H. Westerdale
Department of Computer Science, Birkbeck College,
University of London, Malet Street, London WC1E 7HX,
England
tel. 020-76316704 and 020-76316721
fax 020-76316727
trevor@dcs.bbk.ac.uk and tom@dcs.bbk.ac.uk

1 Introduction

There are several proofs of the Markov chain tree theorem in the literature. Usually the proof proceeds in two steps. The first step proves the theorem for the special case of irreducible Markov chains. (This special case is interesting and useful in itself.) The second step uses the special case theorem as a lemma in the proof of the general case. The proofs in [3] and [1] proceed in that way, and those two references sketch some of the history of the problem.

In this paper we use a random graph approach to obtain a self contained and shorter proof of the special case. It uses no results from random graph theory (or indeed from anywhere else). We feel it is very simple and quite intuitive.[1]

This paper uses only terms, concepts, and notation that are needed to state and prove the special case. For example, even to state the general theorem requires the use of what the literature usually calls an "arborescence". But in the special case, an "arborescence" is obviously simply a spanning in-tree, so we need not even bother to define "arborescence".

2 The Theorem for Irreducible Chains

We think of the state diagram of a finite state Markov chain as a complete digraph, where the nodes correspond to the states and where associated with each edge is a non-negative transition probability. The states are numbered from 1 to n. Edge (i, j) is the edge from state i to state j. The real number p_{ij} is the transition probability associated with edge (i, j).

We call the chain irreducible if for any ordered pair of states there is a directed path of edges from the first state to the second along which all the transition probabilities are positive. Let P be the $n \times n$ matrix of transition probabilities with entries p_{ij}. (The matrix P is called irreducible if the chain is irreducible.)

A spanning in-tree is a rooted spanning tree in which all the edges point toward the root. An i-tree is a spanning in-tree whose root is node i (state i).

If t is a spanning in-tree, we define the number $\hat{t}$ to be the product of all the p_{ij} for which edge (i, j) is in t. We call $\hat{t}$ the value of the spanning in-tree t. For any node i, we define x_i to be the sum of the values of all the i-trees. Let $\bar{x}$ be the sum of the values of all the spanning in-trees, so $\bar{x} = \sum_i x_i$.

Let $\mathbf{x}$ be the row vector whose entries are the x_i's. We shall call $\mathbf{x}$ the *treevalues vector*. If the chain is irreducible then there is at least one spanning in-tree with non-zero value, so $\mathbf{x}$ is not the zero vector.

It is the purpose of this paper to present a proof of the following:

> **Theorem:** If P is the transition probabilities matrix of a irreducible finite state Markov chain, and $\mathbf{x}$ is the treevalues vector, then $\mathbf{x}P = \mathbf{x}$.

Note that the conclusion of the theorem is that $\mathbf{x}$ is an eigenvector of P with eigenvalue 1. It is well known that for an irreducible Markov chain, the subspace of eigenvectors with eigenvalue 1 is one dimensional.[2]

[1] We are grateful to Mark Levene for pointing out some similarities between our proof and the proof in [4].
[2] From the Perron-Frobenius theorem [2, page 65] together with [2, page 100].

The mean limiting absolute probability vector $\tilde{\mathbf{p}}$ is also an eigenvector with eigenvalue 1. [2, page 117] By definition, its i'th entry $\tilde{p}_i$ is the proportion of time the chain spends in state i. Clearly, $\mathbf{x}$ is a scalar multiple of $\tilde{\mathbf{p}}$, and thus, since $\sum_i \tilde{p}_i = 1$, we have $\mathbf{x} = \bar{x}\tilde{\mathbf{p}}$. In other words,

in a irreducible finite state Markov chain we have $\tilde{p}_i = x_i/\bar{x}$, for all i.

This is the usual formulation of the special case of the Markov chain tree theorem. It is more consistent with the usual formulation of the general case, and we see that it is equivalent to our formulation.

3 Proof

Consider the set D of functional digraphs that have the same node set as the complete digraph. (A functional digraph has exactly one edge coming from each node.) For each digraph G in D we define the probability of G to be the product of the p_{ij}'s for each edge (i, j) in G. (Intuitively. we can think of constructing one of these functional digraphs probabilistically, looking at each state i in turn and adding to the graph the edge that comes from i, sending it to node j with probability p_{ij}.) These graph probabilities give us a probability distribution over D. For any subset B of D, we write $\mathbf{P}[B]$ (called the probability of B) for the sum of the probabilities of the digraphs in B. (Obviously, $\mathbf{P}[D] = 1$) .

Note that for any node i, a functional digraph in D can contain at most one i-tree. Suppose t is an i-tree, and consider a functional digraph G that contains t. G contains just the $n-1$ edges in t, plus an additional edge from i. If the edge from i goes to j, then the probability of G is $\hat{t}p_{ij}$.

For any spanning in-tree t, we define T_t to be the set of functional digraphs that contain t. Clearly there are n members in T_t, and $\mathbf{P}[T_t] = \hat{t}$. For any node i, we define S_i to be the set of all functional digraphs that contain an i-tree. Clearly $\mathbf{P}[S_i] = x_i$.

For any ordered pair of nodes i and j, we define E_{ij} to be the set of all the functional digraphs that contain an i-tree and the edge (i, j). If $G \in E_{ij}$ contains the i-tree t, then the probability of G is $\hat{t}p_{ij}$. So it follows that $\mathbf{P}[E_{ij}] = x_i p_{ij}$.

Consider a functional digraph G in D. G must contain at least one cycle, possibly of length 1. If G contains only one cycle it is said to be unicyclic. Clearly a functional digraph contains a spanning in-tree if and only if it is unicyclic. It contains an i-tree if and only if it is unicyclic and i is on the cycle. Thus S_i is the set of all the unicyclic digraphs that have i on the cycle.

Now if a functional digraph G contains both an i-tree and the edge (i, j), then it is unicyclic, i is on the cycle, and the edge (i, j) forms part of the cycle. Thus E_{ij} is the set of all the unicyclic graphs that have the edge (i, j) in the cycle.

Suppose G is a member of S_j , so G is unicyclic with node j on the cycle. We ask which node precedes j on the cycle. If it is i, then G is in E_{ij} . Thus we see that $S_j = \bigcup_i E_{ij}$, and the union is disjoint. So, taking probabilities, we see that $x_j = \sum_i x_i p_{ij}$. Therefore $\mathbf{x}P = \mathbf{x}$. $\square$

References

[1] V. Anantharam and P. Tsoucas. A proof of the Markov chain tree theorem. *Statistics and Probability Letters*, 8:189–192, 1989.

[2] F. R. Gantmacher. *Applications of the Theory of Matrices.* Interscience Publishers, New York, 1959.

[3] F. Thompson Leighton and Ronald Rivest. Estimating a probability using finite memory. *IEEE Trans. on Information Theory*, IT-32(6):733–742, 1986.

[4] Issac Sonin. The state reduction and related algorithms and their applications to the study of Markov chains, graph theory, and the optimal stopping problem. *Advances in Mathematics*, 145:159–188, 1999.