Random walks on trees and matchings

MATH285K - Spring 2010 *Presenter: Ngoc Chi Le*

Reference: [1].

1 Introduction

The paper is based on the bijection between the set of phylogenetic tree with l leaves and the set of perfect matchings on $2n$ points, $\mathcal{M}_n$, (where $n = l - 1$). Using such bijection and analization a natural walk on Mn the paper gave sharp rates of convergence for a natural Markov chain on the space of phylogenetic trees. Roughly, the results show that $\frac{1}{2}n \log n$ steps are necessary and suffice to achieve randomess.

2 Background and needed tools

2.1 Phylogenetic trees and random matchings

A phylogenetic tree with l leaves is a rooted binary tree with l labeled leaves. Let G be a graph with vertex set V and edge set E. A perfect matching is a set of disjoint edges containing all vertices. In the paper, we consider the perfect matchings on $2n$ points, as the way to divide $1, 2, ..., 2n$ into n couples. Here, we briefly describe the correspondence between matchings and trees. Begin with a tree with l labeled leaves. Label the internal vertices sequentially with $l{+}1, l{+}2, ..., 2(l{-}1)$ chosing at each stage the ancestor which has both children labeled and who has the descendant lowest possible available label. When all nodes are labeled, create a matching on $2n = 2(l - 1)$ vertices by grouping siblings. To go backwward, given a perfect matching of $2n$ points, note that at least one matched pair has both entries from $1, 2, 3, ..., n + 1$. All such labels are leaves; if there are several leaf-labeled pairs, choose the pair with the smallest label. Give the next available label $n + 2 = l + 1$ to there parent node. There are then a new set of available labeled pairs. Choose again the pair with the smallest label to take the next available label for its parent and so on.

2.2 Marrkov Chain

For matchings in $\mathcal{M}_n$, a step in the walk is obtained by picking two matched pairs at random, a random entry of each pair and transposing these entries.. For general $n \geq 2$ and $x, y \in \mathcal{M}_n$, define:

$$K(x,y) = \begin{cases} \frac{1}{n(n-1)} & \text{if } x \text{ and } y \text{ differ by a transposition} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The Markov chain (1) has the uniform distribution $\pi(x) = 2^n n!/(2n)!$ as unique stationary distribution. Since K is symmetric, and so, reversible. Because of reversibility, K has an orthonormal basis of eigenvectors $f_i(x)$ with

$$Kf_i(x) = \sum K(x,y)f_i(y) = \beta_i f_i(x)$$

Here β_i is the associated eigenvalue and both f_i and β_i are real. We may express the chi-square distance as:

$$\|K_x^m - \pi\|_2^2 = \sum_y \frac{|K^m(x,y) - \pi(y)|^2}{\pi(y)} = \sum_{i,\beta_i \neq 1} f_i^2(x)\beta_i^{2m} \tag{2}$$

In (2) $K_x^m(y) = K^m(x,y) = \sum_z K^{m-1}(x,z)K(z,y)$ and a is a universal constant. The result is sharp; if $m = \frac{1}{2}n(\log n + c)$ for c positive, there is $x*$ and positive $\epsilon = \epsilon(c)$ such that

$$\|K_{x*}^m - \pi\| \geq \epsilon \qquad \text{for all } n \tag{3}$$

2.3 Group theory

Let $\mathcal{P}_n$ be the partition of n. Partition are written as $\lambda \vdash n$ with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r), \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$ Clearly, S_{2n}, the symmetric group on $2n$ letters acts transitively on matchings coordinate wise.

$$\sigma((i_1, i_2), (i_3, i_4), \ldots, (i_{2n-1}, i_{2n})) = (\sigma(i_1), \sigma(i_2)), \ldots, (\sigma(i_{2n-1}), \sigma(i_{2n}))$$

So () is in force. More over, since S_2n acts transitively on the space of matchings, we have a permutation representation of S_{2n} on $\mathcal{L}(\mathcal{M}_n) = f : \mathcal{M}_n \to \mathbb{R}$. Matchings may be thought of as a product of n disjoint transposition and so as fixed-point free idempotent mappings of $1, 2, \ldots, 2n$ to itself, or as the elements of the conjugacy class of S_2n with all cycles of length two. If B_n is the subgroup of S_{2n} fixing the matching $(1, 2)$(3,4)…(2n-1,2n), then B_n is isomorphic to the hyperocatahedral group of order $2^n n!$. Matchings may be identified with elements of quotient S_{2n}/B_n. The irreducible representation of S_{2n} are indexed by partitions μ of $2n$. They will be denoted S^μ. A crucial fact is that the decomposition of $\mathcal{L}(\mathcal{M}_n)$ is known:

Theorem 1. *Let $\mathcal{M}_n = S_{2n}/B_n$. Let $\mathcal{L}(\mathcal{M}_n)$ be all real functions on $\mathcal{M}_n$, considered as a representation of S_2n. Then*

$$\mathcal{L}(\mathcal{M}_n) = \bigoplus_{\lambda \vdash n} S^{2\lambda}$$

where the direct sum is over all partitions λ of n, $2\lambda = (2\lambda_1, 2\lambda_2, \ldots, 2\lambda_k)$ and $S^{2\lambda}$ is the associated irreducible representation of the symmetric group S_{2n}

Proposition 2. *The transition matrix K of (1) and T_n satisfy*

$$K = \frac{2n-1}{2n-2}\left(T_n - \frac{1}{2n-1}I\right)$$

Corollary 3. *The transition matrix K of (1) has an eigenvalue β_λ for each $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ of n, given by*

$$\beta_\lambda = \frac{1}{n(n-1)} \sum_{j=1}^{k} \lambda_j^2 - j\lambda_j$$

The multiplicity of β_λ is determined by $\mu = 2\lambda$:

$$mult(\lambda) = \frac{(2n)!}{\prod_{(i,j)\in\mu} h(i,j)} \tag{4}$$

with the product being over the cells of the shape μ, and $h(i,j)$ hook length $\mu_i + \mu'_j - i - j + 1$ where μ' is the transposed diagram

3 Main result

From the preaparations, we have the main result:

Theorem 4. *For the Markove chain $K(x,y)$ of (1) on $\mathcal{M}_n$ the space of perfect matchings on $2n$ points, for any starting state x, if $m = \frac{1}{2}n(\log n + c)$, with $c > 0$, then*

$$\|K_x^m - \pi\| \leq ae^{-c} \tag{5}$$

References

[1] Persi Diaconis and Susan P. Holmes, Random walks on trees and matchings, *Electronic Journal of Probability*, Vol. 7 (2002) Paper no. 6, pages 1-17