These notes have not received the scrutiny of publication. They could be missing important references, etc.

Sampling and Counting Spanning Trees

Inputs

Let $G = (V, E)$ be a graph, where $V = [n]$ is the vertex set and $E \subset \binom{V}{2}$ is the the edge set.

Goal

We would like to output a random X from the uniform measure on {spanning trees of G}

Spanning trees are acyclic subgraphs (i.e., subsets of edges) which are connected and span all vertices.

A naive runtime for doing this is $2^{|E|}$

In this lecture, we will analyze an algorithm that is not optimal, strictly speaking. However, it will be easier to analyze.

Our strategy will be to calculate the partition function

$$Z = \#\{\text{spanning trees of } G\}$$

We will then provide a sampling algorithm at the end of these notes.

Theorem (Kirchoff matrix tree)

Let A be the adjacency matrix of G.

$$A_{ij} = \begin{cases} 1 & i \sim j \\ 0 & \text{otherwise} \end{cases}$$

Let D be the degree matrix. It is diagonal.

$$D_{ii} = \sum_j A_{ij} =: \deg(i)$$

The Laplacian matrix L will be the difference between these two matrices.

$$L := D - A$$

A naive guess might be $Z = \det(L)$. However, the determinant of L is actually 0. This follows from observing $\mathbf{1}^T L \mathbf{1} = 0$, where $\mathbf{1} \in \mathbb{R}^n$ is a vector of 1's.

We can remove the collinearity simply by deleting the first row and first column, or the j-th row and j-th column for that matter, for arbitrary j. This turns out to be the correct answer.

$$Z = \det(L_{\neg 1, \neg 1}) = \det(L_{\neg j, \neg j}), \;\; \forall j \in [n]$$

An Aside on Determinants

Observe that we can define a determinant as follows, for general matrix $M \in \mathbb{R}^{n \times n}$.

$$\det(M) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^{n} M_{i,\sigma(i)}$$

Note that $\sigma : [n] \to [n]$ is a bijection and that $\text{sgn}(\sigma) \in \{\pm 1\}$ is 1 if and only if "the permutation can be obtained with an even number of transpositions (exchanges of two entries)" (source: Wikipedia).

In practice, a more computationally efficient way of computing the determinant is to perform an LU-decomposition:

$$\det(M) = \det(L) \cdot \det(U) \tag{1}$$

The determinants of L and U are simply the product of their diagonal entries, given their triangular structure.

Intuition on Laplacians

Let $f \in \mathbb{R}^n$ or similarly $f : [n] \to n$. The difference is simply notational.

$$\begin{aligned} f^T L f &= \sum_{i \sim j} (f(i) - f(j))^2 \\ \text{“} &= \text{”} |\nabla f|_2^2 && \text{(The size of the gradient of } f\text{)} \end{aligned}$$

We can think of $\nabla f : E \to \mathbb{R}$, where $(\nabla f)_{ij} = f(i) - f(j)$ for $i < j$.

Note that, in calculus, the Laplacian of f refers to $\Delta f = \sum_i \partial_i^2 f$

We can see further connections via Green's theorem ($f, g : \mathbb{R}^d \to \mathbb{R}$):

$$\int f \Delta g = - \int \langle \nabla f, \nabla g \rangle$$

Therefore, one can view the Laplacian of f as the divergence of the gradient of f.

$$\Delta f = \nabla \cdot (\nabla f)$$

In our context, we will only consider f such that the first entry (or j-th entry) is always 0.

$$f^T L_{\neg 1, \neg 1} f = (0, f)^T L (0, f)$$

Proof of Theorem

Special Case

Let us first consider the special case of where n is equal to the number of edges plus 1.

$$n = |E| + 1$$

This is an interesting special case because, in a tree, the number of edges equals the number of vertices minus 1.

Claim 1. *$Z = 1$ iff G is connected (or, equivalently, iff G is a tree)*

Proof. Observe the following (where e_i is the i-th standard basis vector of $\mathbb{R}^n$):

$$\begin{aligned} L &= \sum_{i \sim j} (e_i - e_j)(e_i - e_j)^T \\ &= BB^T \end{aligned}$$

Note that $B \in \mathbb{R}^{n \times (n-1)}$ and column r of B is $e_i - e_j$, where (i, j) is the r-th edge of G.

We also observe that $L_{\neg n, \neg n} = B_{\neg n} B_{\neg n}^T$, where $B_{\neg n}$ is B with the n-th row removed.

Therefore, $\det(L_{\neg n, \neg n}) = \det(B_{\neg n})^2$

Hence, we want to show $\det(B_{\neg n}) \in \{\pm 1\}$ iff G is a tree.

Observe the following:

$$\begin{aligned} \det(B) &= \det(B^T) \\ &= \sum_{\sigma \in S_{n-1}} \operatorname{sgn}(\sigma) \prod_{a=1}^{n-1} B_{\sigma(a), a} \end{aligned}$$

We know the following about $B_{\sigma(a),a}$:

$$B_{\sigma(a),a} = \begin{cases} \pm 1 & \text{if } \sigma(a) \text{ is a neighbor of edge } a \\ 0 & \text{o/w} \end{cases}$$

The following must hold:

- Each edge a picks a neighbor $\sigma(a)$
- No two edges can pick the same vertex
- No edges can pick vertex n

Only 1 σ can satisfy all of these criteria. As a result, $\det(B^T) = \pm 1$ □

General Case

Let $L = BB^T$, where $B \in \mathbb{R}^{n \times |E|}$

We introduce the following notation, where $B_{\neg 1} \in \mathbb{R}^{(n-1) \times |E|}$:

$$M := L_{\neg 1, \neg 1} = B_{\neg 1} B_{\neg 1}^T$$

Using the Cauchy-Binet formula (which we will prove later):

$$\begin{aligned} \det(M) &= \sum_{\substack{S \subset E \\ |S| = n-1}} \det[(B_{\neg 1})_{\cdot, S} (B_{\neg 1}^T)_{S, \cdot}] \qquad \text{(Laplacian for } G = (V, S)\text{, i.e. removing one vertex)} \\ &= \sum_{\substack{S \subset E \\ |S| = n-1}} \mathbb{1}_{S \text{ is a spanning tree}} \\ &= \# \text{ of spanning trees} \end{aligned}$$

This concludes the proof of the theorem.

Aside: Proof Sketch of Cauchy-Binet

Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times n}$, where $n \leq m$

Let $\lambda_i(\cdot)$ refer to the i-th eigenvalue

Observe that $\det(AB) = \sum_{S:|S|=n} \det(A_{\cdot,S}B_{\cdot,S})$. One way to see this is to observe that the left-hand side equals $\prod_{i=1}^{n} \lambda_i(AB)$ (since a determinant is simply the product of the eigenvalues), and this product in turn equals $\prod_{i=m-n+1}^{m} \lambda_i(BA)$. This follows from the general fact that $\lambda_i(AB) = \lambda_{m-n+i}(BA)$[1].

To help clarify how we index our eigenvalues, note that $\text{rank}(AB) = \text{rank}(BA) \leq n$ and $\lambda_1, \ldots, \lambda_{m-n} = 0$

We can then conclude that $\prod_{i=m-n+1}^{m} \lambda_i(BA) = \sum_{S:|S|=n} \det(A_{\cdot,S}B_{\cdot,S})$ by making the following observations about the characteristic polynomial (swapping addition for where we normally see subtraction):

$$\begin{aligned}\det(zI + BA) &= \prod_{i=1}^{m}(z + \lambda_i)\\ &= z^{m-n} \prod_{i=m-n+1}^{m} (z + \lambda_i)\end{aligned}$$

To provide intuition on our earlier statement that $\lambda_i(AB) = \lambda_{m-n+i}(BA)$ in general, one can observe the following, where v is an eigenvector of AB with eigenvalue λ:

$$BA(Bv) = B(ABv) = \lambda Bv \tag{2}$$

Therefore, Bv is an eigenvector for BA with eigenvalue λ.

How to Sample

We will work with an autoregressive sampler, i.e., of the form

$$P(X) = \prod_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1})$$

Let's sample an intermediate quantity, $Y_1 \sim \text{Bernoulli}(P(\text{edge 1 is in } T))$. If $Y_1 = 0$, we will delete edge 1. Otherwise, we will keep it.

Note that

$$P(\text{edge 1 is in } T) = \frac{\#\text{ of spanning trees with edge 1}}{\text{total }\#\text{ of spanning trees}}$$

Then, we will sample $Y_2 \sim \text{Bernoulli}(P(\text{edge 2 is in } T \mid Y_1))$, where

$$P(\text{edge 2 is in } T \mid Y_1 = 1) = \frac{\#\text{ of spanning trees with edge 2 and edge 1}}{\text{total }\#\text{ of spanning trees with edge 1}}$$

Similarly,

$$P(\text{edge 2 is in } T \mid Y_1 = 0) = \frac{\#\text{ of spanning trees with edge 2 but without edge 1}}{\text{total }\#\text{ of spanning trees without edge 1}}$$

[1] This fact will be very useful in subsequent lectures

We continue this sampling procedure for $Y_1, \ldots, Y_{|E|}$ and have thus sampled a spanning tree.

Note: counting the number of spanning trees containing edge 1 can be done by merging the adjacent vertices and summing the corresponding rows of the adjacency matrix. This uses the fact that we can count spanning trees on weighted graphs, which we will discuss more next lecture.