Random Graphs 30/05/2020

Cayley's Formula

Project work by: *Somak Laha*

Table of contents

1 Introduction

In this report, I shall outline some basics from graph theory and probability theory in an attempt to present a proof of *Cayley's Formula*. *Cayley's Formula* is one of the most simple and elegant results in graph theory that counts number of labelled trees.

Definition 1.1 (Tree)**.** A graph $G = (V, E)$ (V is the set of vertices and E is the set of edges) is said to be a tree if it is connected, acyclic and undirected.

A tree on n vertices can be labelled by assigning each of its vertex a number from $[n] := \{1, 2, \cdots, n\}$. Edge set of a labelled tree has two-element subsets of $[n]$ as its elements.

Theorem 1.2 (Cayley's Formula)**.** There are n^{n-2} different labelled trees on n vertices.

The term *labelled* emphasises that we are not identifying isomorphic trees. We have fixed the set of vertices, namely $[n]$, and two trees are counted as the same if and only if they have same edge set. A *spanning tree* of a connected graph G is a spanning subgraph of G that is a tree. So, alternatively we could have stated the above theorem as: "*The complete graph K_n has n^{n-2} different spanning trees*".

We shall prove this theorem in a probabilistic approach. In fact, we consider simulating a branching process with Poisson(1) offspring distribution. From theory of branching process, one can show that such a process will be extinct with probability 1. As we label the simulated tree, and shall calculate the probability that we get one specific labelled tree l on n vertices, conditioned on the fact that the entire tree has n vertices, we shall see that the conditional probability won't depend on l. That would prove that this conditional probability distribution is uniform on the set of all labelled trees on n vertices. Rest of the proof would be immediate. We need to introduce and prove a few facts from theory of Branching Processes before we get into the proof of *Cayley's Formula*.

2 Introduction to Branching Process

A branching process is a popular and one of the simplest models for a population evolving with time. Consider particles such as bacteria that can generate new particles of the same type. The initial set of objects is referred to as belonging to the 0-th generation. Particles generated from the n-th generation are said to belong to the $(n + 1)$-th generation. As a convention, we shall assume that the 0-th generation consists of one particle only. Shall denote by Z_n the the number of individuals in n-th generation. So, according to convention, $Z_0 = 1$. Also, for $n \geq 1$, Z_n satisfies the recursion relation,

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i},$$

where $(X_{n,i})_{n,i\geq 1}$ is a doubly infinite array of i.i.d. random variables having same distribution as some non-negative integer-valued random variable X. The distribution of X is called the *offspring distribution* of the branching process.
There is a major result involving the extinction probability of a branching process. For a branching process $(Z_n)_{n\geq 0}$, we define its extinction probability as,

$$\eta = \mathbb{P}(\exists n:\ Z_n = 0).$$

Theorem 2.1 (Extinction probability for branching processes)**.** For a branching process with i.i.d. offspring X, $\eta = 1$ when $\mathbb{E}[X] < 1$, while $\eta < 1$ when $\mathbb{E}[X] > 1$. Further, if $\mathbb{E}(X) = 1$ and $\mathbb{P}(X = 1) < 1$, then $\eta = 1$. The extinction probability of η is the smallest solution in $[0, 1]$ of $G_X(t) = t$, where G_X is the probability generating function of X, i.e.,

$$G_X(t) = \mathbb{E}(t^X).$$

The proof of this theorem is skipped here. It can be found in any standard text discussing branching processes. We shall continue by studying laws of *total progeny* T of the branching process, which is defined as,

$$T = \sum_{n=0}^{\infty} Z_n.$$

3 Random Walk perspective of Branching Process and Law of Total Progeny

One can give a general result to give the distribution of T in terms of probabilities involving independent copies of X.

Theorem 3.1 (Laws of Total Progeny)**.** For a branching process with i.i.d. offspring distribution $Z_1 = X$,

$$\mathbb{P}(T = n) = \frac{1}{n}\mathbb{P}(X_1 + \cdots + X_n = n - 1), \tag{1}$$

where $(X_i)_{i=1}^n$ are i.i.d. copies of X.

We shall prove Theorem 3.1 later. In fact, we prove a more general result that states that,

$$\mathbb{P}(T_1 + \cdots + T_k = n) = \frac{k}{n}\mathbb{P}(X_1 + \cdots + X_n = n - k), \tag{2}$$

where $T_1, \cdots, T_k$ are independent copies of T. Its proof will follow from the *Kemperman's formula* (Theorem 3.2). Before stating and proving that, we need to introduce the *Random Walk perspective of Branching Process.*

3.1 Random Walk Perspective

In branching processes, it is quite common to study the number of descendants of each individual in a given generation. For random graph purposes though, it is often convenient to use a different construction of a branching process by sequentially investigating the number of children of each member of the population. Let $X_1^{'},\ X_2^{'},\ \cdots$ be independent and identically distributed random variables having same distribution as the progeny distribution of the branching process. Next, we define $S_0^{'},\ S_1^{'}, \cdots$ recursively as,

$$\begin{aligned} S_0^{'} &= 1 \\ S_i^{'} &= S_{i-1}^{'} + X_i^{'} - 1 = X_1^{'} + \cdots + X_i^{'} - (i-1) \end{aligned}$$

The branching process belonging to the above recursion is as follows. The population starts with only one individual. At time i $(i > 0)$, we select one of the active individuals in the population and give it $X_i^{'}$ children. The children are set to active and the individual itself is set to inactive. One may do this exploration process in breadth-first manner or depth-first manner or even neither.

This exploration process is continued as long as there are active individuals in the population. Then, the process $S_i^{'}$ describes the number of active individuals after the first i individuals has been explored. The process stops when, for the first time, $S_t^{'} = 0$. Let $T^{'}$ be the smallest t for which $S_t^{'} = 0$, i.e.,

$$T^{'} = \inf\{t : S_t^{'} = 0\} = \inf\{t : X_1^{'} + \cdots + X_t^{'} = t - 1\}.$$

In particular, if such a $T^{'}$ does not exist, we define $T^{'} = +\infty$.

Note that, the above defined $T^{'}$ clearly equals the total progeny T of the concerned branching process. This interpretation of total progeny size will be useful in finding its distribution.

3.2 Law of Total Progeny

The following theorem is a remarkable result for random walk, say, $(S_n)_{n\geq 0}$. Suppose the walk starts from some $k \geq 0$, i.e., $S_0 = k$. The theorem states that, conditionally on the event $\{S_n = 0\}$, and regardless of the precise distribution of the steps of the walk (though there is some condition the distribution needs to obey), the probability that the walk is at 0 for the first time at n-th step equals k/n. Define,

$$H_0 = \inf\{n \geq 0 : S_n = 0\}.$$

Theorem 3.2 (Kemperman's Formula)**.** For a random walk with i.i.d. steps $(Y_i)_{i\geq 1}$ satisfying that Y_i is integer valued and

$$\mathbb{P}(Y_i \geq -1) = 1,$$

the distribution of H_0 is given by

$$\mathbb{P}_k(H_0 = n) = \frac{k}{n}\ \mathbb{P}_k(S_n = 0). \tag{3}$$

We skip the proof of Theorem 3.2 here and use it as a result in the next proof. It is proved in a later section (Section 5).

Next, we complete the proof of Theorem 3.1.

Proof of Theorem 3.1. We stated earlier that we prove a more general version of the theorem, which is,

$$\mathbb{P}(T_1 + \cdots . + T_k = n) = \frac{k}{n}\,\mathbb{P}(X_1 + \cdots . + X_n = n - k).$$

Define, $H_0^i = \inf\{n \geq 0 : S_n = 0,$ given that $S_0 = i\}$. Say, $H_{0,1}^1, \cdots, H_{0,k}^1$ are k i.i.d. copies of H_0^1. Next note that, by the Markov Property of random walk, H_0^k has the same distribution as $H_{0,1}^1 + \cdots + H_{0,k}^1$. Now, recall the random walk perspective of branching process, from definition of T'(the total progeny), it is clearly a zero-hitting time where the random walk starts from $S_0 = 1$ (which means there is only one particle in the 0-th generation).
Thus, if the random walk in concern is the random walk representation of the branching process, we may say,

$$T_1 + \cdots . + T_k \stackrel{d}{=} H_{0,1}^1 + \cdots + H_{0,k}^1 \stackrel{d}{=} H_0^k.$$

Also, note that, in this random walk the steps are $Y_i = X_i - 1$, where $(X_i)_{i\geq 1}$ are the offspring of visited vertex. So, $\mathbb{P}(Y_i \geq -1) = 1$. Hence, we can apply Kemperman's formula (Theorem 3.2) to write,

$$\begin{aligned}
\mathbb{P}(T_1 + \cdots . + T_k = n) &= \mathbb{P}(H_{0,1}^1 + \cdots + H_{0,k}^1 = n) \\
&= \mathbb{P}(H_0^k = n) \\
&= \frac{k}{n}\mathbb{P}_k(S_n = 0) \\
&= \frac{k}{n}\mathbb{P}(Y_1 + \cdots + Y_n = -k) \\
&= \frac{k}{n}\mathbb{P}(X_1 + \cdots + X_n = n - k).
\end{aligned}$$

This completes the proof. □

4 Proof of Cayley's Formula

As stated earlier, we shall make use of Critical Poisson Branching Process, i.e., Branching Process with Poisson(1) as its progeny distribution. We find out the law of total progeny(T^*) of Poisson Branching Process as a lemma of Theorem 3.1.

Lemma 4.1. For a branching process with i.i.d. offspring $X^* \sim$ Poisson(λ),

$$\mathbb{P}(T^* = n) = \frac{(\lambda n)^{n-1}}{n!}e^{-\lambda n}, \qquad n \geq 1.$$

Proof. Directly from Kemperman's formula, we may write,

$$\begin{aligned}
\mathbb{P}(T^* = n) &= \mathbb{P}(X_1 + \cdots . + X_n = n-1) \qquad [\text{where, } X_1, \cdots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda).] \\
&= \mathbb{P}(Z = n-1) \qquad [\text{where, } Z \sim \text{Poisson}(n\lambda).] \\
&= \frac{(\lambda n)^{n-1}}{n!} e^{-\lambda n}.
\end{aligned}$$

This completes the proof.

□

So, where offspring distribution is Poisson(1), distribution of total progeny T^* is:

$$\mathbb{P}(T^* = n) = \frac{(n)^{n-1}}{n!} e^{-n}, \qquad n \geq 1. \tag{4}$$

Next, we shall prove the Cayley's formula (Theorem 1.2).

Proof of Theorem 1.2. This proof will make use of *family trees*. Each vertex of a family tree will be represented by certain *words*. These words arise inductively. Say, for any vertex j, its number of children is denoted by d_j. The root is represented by ϕ. Then it has d_ϕ may children, thus, its children are $1, 2, \cdots, d_\phi$. Next, inductively the children of vertex 1 are $11, 12, \cdots, 1d_1$. This representation is commonly known as *Ulam-Harris Representation of trees*.

Two family trees are exactly the same if they are represented by same collection of words. Also, for a word w, $|w|$ be its length which denotes the number of steps it is away from the root, i.e., its generation. For the root, $|\phi| = 0$.

Let $\mathcal{T}$ be the random variable denoting the family tree of a branching process with Poisson(1) as its progeny distribution. Note that, number of children of all the vertices are mutually independent. Hence, the probability that $\mathcal{T}$ equals t for some specific family tree t is given by,

$$\begin{aligned}
\mathbb{P}(\mathcal{T} = t) &= \prod_{w \in t} \mathbb{P}(\psi = d_w) \qquad [\text{Where } \psi \text{ is a Poisson(1) random variable.}] \\
&= \prod_{w \in t} \frac{e^{-1}}{d_w!}.
\end{aligned}$$

Once we have an observation of $\mathcal{T}$, we introduce a labelling for it.The root is labelled 1. Conditioned on having total progeny $T^* = n$, rest of the vertices are labelled by $\{2, 3, \cdots, n\}$

uniformly at random and without replacement. Thus we get a labelled tree $\mathcal{L}$. We shall prove that $\mathcal{L}$ is a uniform labelled tree on n vertices.

For a family tree t and a labelled tree l, we shall write $t \sim l$, if l can be obtained by labelling the family tree t. Now, given a labelled tree l and any family tree t, such that $t \sim l$, let $L(l)$ be the number of ways to label t in different ways (such that the root is labelled as 1). Here, the observation is that, the quantity $L(l)$ does not depend on the choice of t. Thus, given a labelled tree l,

$$\#\{t : t \sim l\} = \frac{1}{L(l)} \prod_{w \in l} d_w!,$$

since, permuting the children of any vertex does not change the tree.

Also, note that, for a labelled tree l and a family tree t, such that $t \sim l$, there are $L(l)$ many ways to label to t in order to get l. Thus, if t is randomly labelled (always labelling the root as 1),

$$\mathbb{P}(t \text{ receives label } l) = \frac{L(l)}{(|l|-1)!}.$$

Using all these observations above, finally we can say,

$$\begin{aligned}
\mathbb{P}(\mathcal{L} = l) &= \sum_{t_l \sim l} \mathbb{P}(\mathcal{T}_l = t_l)\mathbb{P}(t_l \text{ receives label } l) \\
&= \sum_{t_l \sim l} \left[\left(\prod_{w \in t_l} \frac{e^{-1}}{d_w!} \right) \frac{L(l)}{(|l|-1)!} \right] \\
&= \sum_{t_l \sim l} \left[\frac{e^{-|l|}}{\prod_{w \in l} d_w!} \frac{L(l)}{(|l|-1)!} \right] \\
&= \#\{t : t \sim l\} \left[\frac{e^{-|l|}}{\prod_{w \in l} d_w!} \frac{L(l)}{(|l|-1)!} \right] \\
&= \frac{1}{L(l)} \prod_{w \in l} d_w! \left[\frac{e^{-|l|}}{\prod_{w \in l} d_w!} \frac{L(l)}{(|l|-1)!} \right] \\
&= \frac{e^{-|l|}}{(|l|-1)!}.
\end{aligned} \tag{5}$$

Therefore, conditionally on the total progeny size,$T^* = n$,

$$
\begin{aligned}
\mathbb{P}(\mathcal{L} = l \,|\, |\mathcal{L}| = n) &= \frac{\mathbb{P}(\mathcal{L} = l, |\mathcal{L}| = n)}{\mathbb{P}(|\mathcal{L}| = n)} \\
&= \frac{e^{-n}}{(n-1)!} \frac{n!}{n^{n-1}e^{-n}} \\
&\qquad \text{[Follows from equation (4) and equation (5).]} \\
&= \frac{1}{n^{n-2}}.
\end{aligned}
$$

Thus we can see that the obtained probability does not depend on l, except on the fact that $|l| = n$. This implies that the conditional probability distribution of $\mathcal{L}$ is uniform over the set of all labelled trees on $[n]$. Also, *the size of that set must be* $\mathbb{P}(\mathcal{L} = l \,|\, |\mathcal{L}| = n)^{-1} = n^{n-2}$.

This completes the proof of *Cayley's Formula*.

□

5 Cyclic Shifts and Lagrange Inversion

We start with an elementary lemma concerning cyclic shifts. This lemma will be the only useful tool in this entire section.

Definition 5.1 (Cyclic Shift)**.** $\mathbf{x} := (x_1, x_2, \cdots, x_n)$ be a sequence. For $i \in [n]$, $\mathbf{x}^{(i)}$ denotes the i-th cyclic shift of $\mathbf{x}$, that is a sequence of length n whose j-th term is $x_{i+j\,(mod\; n)}$.

Lemma 5.2. Let $\mathbf{x} := (x_1, x_2, \cdots, x_n)$ be a sequence with values in $\{-1, 0, 1, 2, \cdots\}$ such that $\sum_{i=1}^{n} x_i = -k$ for some $1 \le k \le n$. Then there are exactly k many distinct $i \in [n]$ such that $\mathbf{x}^{(i)}$ first hits $-k$ at time n.

Proof. Define, for $1 \le j \le n$, $s_j = \sum_{i=1}^{j} x_i$. Set $i = m$, the least i such that $s_i = \underset{1 \le j \le n}{min}\, s_j$. Then note that the walk with steps from $\mathbf{x}^{(m)}$ hits $-k$ for the first time at time n. Replace $\mathbf{x}$ by $\mathbf{x}^{(m)}$ and from that we construct the other shifts, since any cyclic shift of $\mathbf{x}^{(m)}$ is indeed a cyclic shift of $\mathbf{x}$.
Now we have a $\mathbf{x}$ such that the walk $(s_n)_{n \ge 0}$ with steps from $\mathbf{x}$ hits $-k$ for the first time at time n. For $1 \le b \le k$ define i_b to be the least i such that $s_i = -b$, i.e., the walk hits $-b$ for the first time at time b. Trivially, $i_k = n$. Next, we observe that the walk with steps from $\mathbf{x}^{(i_b)}$ hits $-k$ for the first time at time n, for all $1 \le b \le k$. Also, $\mathbf{x}^{(i_k)} = \mathbf{x}^{(n)} = \mathbf{x}$. Thus, we get k distinct cyclic shifts of $\mathbf{x}$ each of which hits $-k$ for the first time at time n. Also, for any $j \in \{0, \cdots, i_1 - 1\} \cup \{i_1 + 1, \cdots, i_2 - 1\} \cup \cdots \{i_{k-1} + 1, \cdots, i_k - 1\}$, if $\mathbf{x}^{(j)}$ is considered, the walk $(s_n)_{n \ge 0}$ with steps from $\mathbf{x}^{(j)}$ hits $-k$ before time n.

This leaves us with exactly k distinct cyclic shifts of $\mathbf{x}$ that hits $-k$ for the first time at time n.

$\square$

5.1 Proof of Kemperman's Formula

Theorem 5.3 (Kemperman's Formula)**.** Suppose a sequence of random variables $\mathbf{X} := (X_1, X_2, \cdots, X_n)$ taking values in $\{-1, 0, 1, 2, \cdots\}$ such that all the X_i's are independent and identically distributed. Define $S_j := \sum_{i=1}^{j} X_i$ and $T_{-k} = \inf\{j > 0 : S_j = -k\}$. Then,

$$\mathbb{P}(T_{-k} = n) = \frac{k}{n}\mathbb{P}(S_n = -k). \qquad [\forall k \in \{1, 2, \cdots, n\}]$$

Remark 5.4. *The above theorem and the version in Theorem 3.2 are equivalent. In Theorem 3.2, the walk started from $S_0 = k$ and was supposed to hit 0 in time n. In this version, the walk starts from 0 and hits $-k$ in time n, i.e., there is a origin change keeping the events otherwise same.*

Proof of Kemperman's Formula. It is equivalent to prove for each $k \in \{1, 2, \cdots, n\}$,

$$\mathbb{P}(T_{-k} = n | S_n = -k) = \frac{k}{n}.$$

Consider the set S to be the set of all sequences of length n taking values in $\{-1, 0, 1, 2, \cdots\}$ such that each sequence sums up to $-k$. Define a relation $\mathcal{R}$ on $S \times S$ as, $(\mathbf{w}, \mathbf{v}) \in \mathcal{R}$ if $\exists i \in [n]$ such that $\mathbf{w}^{(i)} = \mathbf{v}$. Then, clearly this relation $\mathcal{R}$ is an equivalence relation and hence partitions S into equivalence classes. Each equivalence class contains n many different sequences, since each sequence has n many distinct cyclic shifts, one being itself. By Lemma 5.2, we know that each equivalence class contains exactly k members such that the walk with steps from each of them hits $-k$ for the first time at time n. Thus, out of all walks hitting $-k$ at time n, a $\frac{k}{n}$ fraction of them hits $-k$ for the first time at time n. In terms of probability, the same statement can be written as,

$$\begin{aligned} \mathbb{P}(T_{-k} = n | S_n = -k) &= \frac{k}{n} \qquad [\forall k \in \{1, 2, \cdots, n\}] \\ \implies \mathbb{P}(T_{-k} = n) &= \frac{k}{n}\mathbb{P}(S_n = -k) \qquad [\forall k \in \{1, 2, \cdots, n\}]. \end{aligned}$$

This concludes the proof.

$\square$

5.2 Lagrange Inversion Formula

Theorem 5.5 (Lagrange Inversion Formula)**.** Let $F(x) = a_1x + a_2x^2 + \cdots \in xK[[x]]$, where $a_1 \neq 0$ (and *char* $K = 0$), and let $k, n \in \mathbb{Z}$. Then

$$n[x^n]F^{(-1)}(x)^k = k[x^{n-k}]\left(\frac{x}{F(x)}\right)^n. \tag{6}$$

Equivalently, suppose $G(x) \in K[[x]]$ with $G(0) \neq 0$, and let $f(x)$ be defined by

$$f(x) = xG(f(x)).$$

Then,

$$n[x^n]f(x)^k = k[x^{n-k}]G(x)^n. \tag{7}$$

Remark 5.6. *equations* (6) *and* (7) *are equivalent since the statement that* $f(x) = F^{(-1)}(x)$ *is easily seen to mean the same as* $f(x) = xG(f(x))$ *where* $G(x) = \dfrac{x}{F(X)}$.

Shall use Kemperman's Formula to give a proof sketch of this algebraic result.

Suppose the sequence of random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$, as mentioned in statement of Kemperman's Formula takes values in $\{0, 1, 2, \cdots\}$ and $S_j = \sum_{i=1}^{j} X_i$. Then, we concern ourselves with the walk $(S_n)_{n\geq 0}$. Define

$$T_{-k} := \inf\{j > 0 : S_j - j = -k\},$$

Kemperman's Formula can be restated as,

$$\mathbb{P}(T_{-k} = n) = \frac{k}{n}\mathbb{P}(S_n - n = -k)\ \forall k \in \{1, 2, \cdots, n\}$$

Suppose the steps X_i has a common pmf $p_j = \mathbb{P}(X_i = j)$ for $j \geq 0$. Then, the probability generating function of X_i is,

$$g(z) := \mathbb{E}(z^{X_i}) = \sum_{j=0}^{\infty} p_j z^j \quad (|z| < 1)$$

And, for $k = 1, 2, \cdots$, probability generating function for T_{-k} would be,

$$h_k(z) := \mathbb{E}(z^{T_{-k}}) = \sum_{n=1}^{\infty} \mathbb{P}(T_{-k} = n)z^n \quad (|z| < 1)$$

Observe that the walk $(S_n - n)_{n\geq 0}$ can move a maximum of 1 downwards in one step. Thus, for the walk to hit $-k$ for the first time, it needs to hit each of $-1, -2, -3, \cdots$ one by one. Thus, $T_{-k} \stackrel{d}{=}$ sum of k independent copies of T_{-1}. What immediately follows is,

$$h_k(z) = h(z)^k. \qquad [h(z) := h_1(z)]$$

Next, note that,

$$\begin{aligned}
h(z) &= \sum_{k=0}^{\infty} \mathbb{E}(z^{T_{-1}} | X_1 = k)\mathbb{P}(X_1 = k) \\
&= \sum_{k=0}^{\infty} \left[\sum_{n=1}^{\infty} \mathbb{P}(T_{-1} = n | X_1 = k) z^n \right] \mathbb{P}(X_1 = k) \\
&= \sum_{k=0}^{\infty} \left[\sum_{n=1}^{\infty} \mathbb{P}(T_{-k} = n-1) z^n \right] \mathbb{P}(X_1 = k) \\
&\qquad [\because\ X_1 = k, \text{ we need } S_{n-1} - (n-1) = -k.] \\
&= \sum_{k=0}^{\infty} \left[\sum_{n=1}^{\infty} z\mathbb{P}(T_{-k} = n-1) z^{n-1} \right] \mathbb{P}(X_1 = k) \\
&= z\sum_{k=0}^{\infty} h_k(z)\mathbb{P}(X_1 = k) = z\sum_{k=0}^{\infty} \mathbb{P}(X_1 = k) h(z)^k \\
&= z\mathbb{E}(h(z)^{X_1}) = zg(h(z)).
\end{aligned}$$

Next,

$$\begin{aligned}
[z^n]h(z)^k &= [z^n]h_k(z) = \mathbb{P}(T_{-k} = n) = \frac{k}{n}\mathbb{P}(S_n - n = -k) = \frac{k}{n}\mathbb{P}(S_n = n-k) \\
&= \frac{k}{n}[z^{n-k}]g(z)^n \\
&\qquad [\text{Since, } S_n = \text{sum of } n \text{ independent copies of } X_1, \text{ pgf of } S_n = g(z)^n]
\end{aligned}$$

Thus, $h(z) = zg(h(z))$ has a solution for arbitrary non-negative $p_0 (\neq 0), p_1, \cdots p_n$ such that $\sum_{i=0}^{n} p_i \leq 1$. The general conclusion follows by polynomial continuation.

5.3 Galton-Watson Forests

We discussed a random walk perspective of Branching Process in Section 3. The following lemma states the bijection between lattice walks and plane forests formally, and can be easily checked.

Lemma 5.7. Given a plane forest F of k trees with n vertices, let x_i be the number of children of vertex i in order of depth-first search. Then,

$$F \longleftrightarrow (x_1, x_2, \cdots, x_n)$$

sets up a bijection between the set of plane forests with n vertices, k components and sequences of non-negative integers $(x_1, x_2, \cdots, x_n)$ such that the lattice walks with steps

$x_i - 1$ first reaches $-k$ at time n. Moreover, if the trees of the forest are of sizes $n_1, n_2, \cdots, n_k$, then for each $1 \leq i \leq k$, the walk first reaches $-i$ at the time $n_1 + n_2 + \cdots + n_i$, i.e., when the depth-first search of the i-th tree is completed.

As an application of this lemma and cyclic shift lemma, we immediately arrive at these following corollaries.

Corollary 5.8 (Enumeration of plane forests by type). *The type of a forest F is the sequence of of non-negative integers (n_i), where n_i is the number of vertices in F with i children. Let $1 \leq k \leq n$ and let (n_i) be a sequence of non-negative integers with*

$$\sum_i n_i = n \text{ and } \sum_i i n_i = n - k. \tag{8}$$

Then, a forest of type (n_i) has n vertices and $n-k$ non-root vertices, hence k roots and k components. For $1 \leq k \leq n$ and (n_i) subject to equation (8) the number $N^{plane}(n_0, n_1, \cdots)$ of plane forests of type (n_i) with k components and n vertices is,

$$N^{plane}(n_0, n_1, \cdots) = \frac{k}{n} \binom{n}{n_0, n_1, \cdots, n_n}. \tag{9}$$

Proof. We count the number of sequences of $(x_1, x_2, \cdots, x_n)$ of non-negative integers such that,

$$\begin{aligned} n_0 &= \text{number of } 0's \text{ in } (x_1, x_2, \cdots, x_n) \\ n_1 &= \text{number of } 1's \text{ in } (x_1, x_2, \cdots, x_n) \\ &. \\ &. \\ &. \\ n_n &= \text{number of } n's \text{ in } (x_1, x_2, \cdots, x_n) \end{aligned}$$

Number of such sequences= $\binom{n}{n_0, n_1, \cdots, n_n}$
Hence, by cyclic shift lemma, number of such sequences such that the lattice walk with steps $x_i - 1$ hits $-k$ for the first time at time $n = \frac{k}{n} \binom{n}{n_0, n_1, \cdots, n_n}$

Thus, by an application of Lemma 5.7,

$$N^{plane}(n_0, n_1, \cdots) = \frac{k}{n} \binom{n}{n_0, n_1, \cdots, n_n}.$$

□

Corollary 5.9 (Enumeration of labeled forests by type)**.** *For* $1 \leq k \leq n$ *and* (n_i) *subject to equation* (8)*, the number of* $N^{[n]}(n_0, n_1, \cdots)$ *of forests labeled by* $[n]$ *of type* (n_i) *with* k *components and* n *vertices is,*

$$N^{[n]}(n_0, n_1, ...n_n) = \frac{k}{n}\binom{n}{k}\frac{(n-k)!}{\prod_{i\geq 0}(i!)^{n_i}}\binom{n}{n_0, n_1, \cdots, n_n}. \tag{10}$$

Proof. To count labeled forest of type (n_i), we need to count number of ways to label an unlabeled tree of type (n_i) by $[n]$. First, we label the roots of the forest, which can be done in $\binom{n}{k}$ ways. For the non-root vertices, for each permutation of their labels a new labeled tree is listed. Also, permuting the labels of children of same parent does not change the labeled tree. Thus, we conclude,

$$\begin{aligned} N^{[n]}(n_0, n_1, ...n_n) &= \binom{n}{k}\frac{(n-k)!}{\prod_{i\geq 0}(i!)^{n_i}} N^{plane}(n_0, n_1, \cdots) \\ &= \frac{k}{n}\binom{n}{k}\frac{(n-k)!}{\prod_{i\geq 0}(i!)^{n_i}}\binom{n}{n_0, n_1, \cdots, n_n} \end{aligned}$$

□

The next theorem is not a direct consequence of previously stated lemmas. Still it being very similar to the previous results, is enlisted here.

Theorem 5.10 ((Enumeration of labeled forests by numbers of children)**.** *For all sequences of non-negative integers* $(c_1, c_2, \cdots, c_n)$ *with* $\sum_i c_i = n - k$*, the number* $N(c_1, c_2, \cdots, c_n)$ *of forests* F *with vertex set* $[n]$ *in which vertex* i *has* c_i *children for each* $i \in [n]$ *is,*

$$N(c_1, c_2, \cdots, c_n) = \frac{k}{n}\binom{n}{k}\binom{n-k}{c_1, c_2, \cdots, c_n}. \tag{11}$$

Proof. For a labeled forest F with vertex set $[n]$ and $i \in [n]$, let $J_i(F)$ be the set of children of i in F. So F is determined by the sequence of disjoint sets $J_1(F), J_2(F), \cdots, J_n(F)$, and vice-versa. Given a sequence of disjoint subsets $J_1, \cdots, J_n$ of $[n]$, for each $m \in [n]$ let $\boldsymbol{f_m}$ be the relation on $[m] \cup (\cup_{i=1}^{m} J_i)$ defined by,

$$i \xrightarrow{\boldsymbol{f_m}} j \quad \text{iff } i \in [m] \text{ and } j \in J_i.$$

There exists a forest F labeled by $[n]$ such that such that $J_i(F) = J_i$ for all $i \in [n]$ if and only if the J_i's are such that $\boldsymbol{f_m}$ is a forest with vertex set $[m] \cup (\cup_{i=1}^m J_i)$ for every $m \in [n]$. It follows that for each sequence of non-negative integers (c_i) with $\sum_i c_i = n - k$, the number

$$N(c_1, c_2, \cdots, c_n) := \#\{F \in F_{k,n} : \#J_i(F) = c_i \text{ for all } i \in [n]\},$$

(where $F_{k,n}$:=set of all labeled forests with n vertices and k components) is the number of ways to choose a sequence of subsets $(J_1, \cdots, J_n)$ of $[n]$ such that $\#J_m = c_m$ and the relation $\boldsymbol{f_m}$ is a forest on vertex set $[m] \cup (\cup_{i=1}^m J_i)$ for every $m \in [n]$. Clearly, J_1 can be any of the $\binom{n-1}{c_1}$ subsets of $[n] \setminus \{1\}$ of size c_1. For $m \in [n-1]$ make the inductive hypothesis that sets $J_1, \cdots, J_m$ of sizes $c_1, \cdots, c_m$ respectively have been chosen such that $\boldsymbol{f_m}$ is a forest. Which choices of J_{m+1} of size c_{m+1} makes $\boldsymbol{f_{m+1}}$ a forest? There are two cases to consider:

(i) $m + 1 \notin \cup_{i=1}^m J_i$: then J_{m+1} can be any subset of $[n] \setminus \cup_{i=1}^m J_i \setminus \{m + 1\}$;

(ii) $m + 1 \in \cup_{i=1}^m J_i$: then J_{m+1} can be any subset of $[n] \setminus \cup_{i=1}^m J_i \setminus \{r_{m+1}\}$, where $r_{m+1} \notin \cup_{i=1}^m J_i$ is the root of the tree component in $\boldsymbol{f_m}$ that contains $m + 1$.

Either way, regardless of choices of $J_1, \cdots, J_m$ the number of possible choices of J_{m+1} to make $\boldsymbol{f_{m+1}}$ a forest is $\binom{n - \sum_{i=1}^m c_i - 1}{c_{m+1}}$. Consequently, by induction,

$$\begin{aligned}
N(c_1, \cdots, c_n) &= \binom{n-1}{c_1}\binom{n - c_1 - 1}{c_2} \cdots \binom{n - \sum_{i=1}^{n-1} c_i - 1}{c_n} \\
&= \frac{(n-1)!}{c_1!(n - c_1 - 1)!} \frac{(n - c_1 - 1)!}{c_2!\left(n - \sum_{i=1}^{2} c_{i-1}\right)!} \cdots \frac{\left(n - \sum_{i=1}^{n-1} c_i - 1\right)!}{c_n!\left(n - \sum_{i=1}^{n} c_i - 1\right)!} \\
&= \frac{(n-1)!}{(k-1)!} \frac{1}{c_1!c_2! \cdots c_n!} \qquad \left[\text{Since, } \sum_{i=1}^n c_i = n - k\right] \\
&= \frac{(n-1)!}{(k-1)!(n-k)!} \frac{(n-k)!}{c_1!c_2! \cdots c_n!} \\
&= \binom{n-1}{k-1}\binom{n-k}{c_1, c_2, \cdots, c_n} \\
&= \frac{k}{n}\binom{n}{k}\binom{n-k}{c_1, c_2, \cdots, c_n}.
\end{aligned}$$

$\square$

6 A few more Proofs

There are many different methods to prove the Cayley's Formula. Here, we include two more methods that use the theorems and corollaries we have already discussed in the report. Also, a complete proof of the Lagrange Inversion Formula is included in this section.

6.1 A second proof of Cayley's Formula

In view of multinomial theorem, equation (11) amounts to the following identity of polynomials in n commuting indeterminates $x_i, 1 \leq i \leq n$:

$$\sum_{F \in F_{k,n}} \prod_{i=1}^{n} x_i^{C(i,F)} = \frac{k}{n} \binom{n}{k} (x_1 + \cdots + x_n)^{n-k}, \tag{12}$$

where $C(i, F)$ is the number of children of vertex i in forest F. Hence, taking all the x_i's to be identically 1, we obtain,

$$\begin{aligned} \#F_{k,n} =& \#\text{Rooted, labeled forests on } n \text{ vertices and having } k \text{ components} \\ =& k \binom{n}{k} n^{n-k-1}. \end{aligned}$$

In particular, for $k = 1$, number of labeled rooted trees on n vertices$= n^{n-1}$. Equivalently, number of unrooted trees, labeled by $[n] = n^{n-2}$.

6.2 A third proof of Cayley's Formula

Suppose, $t(n)$ denote the number of rooted trees on n vertices who are labelled by $[n]$. Then, the *exponential generating function* $T(x)$ counting the number of rooted labelled trees is defined as,

$$T(x) = \sum_{n \geq 1} t(n) \frac{x^n}{n!}.$$

For this *exponential generating function* $T(x)$, we get this following identity (more details of *exponential generating functions* are given in the Appendix),

$$T(x) = xe^{T(x)}$$

We can use Lagrange Inversion Formula to find out the power series expansion of $T(x)$ explicitly. $T(x)$ is the functional composition inverse of xe^{-x}. In the Lagrange Inversion formula $n[x^n]F^{(-1)}(x)^k = k[x^{n-k}] \left(\frac{x}{F(x)} \right)^n$, take $F(x) = xe^{-x}$ and $k = 1$. Hence,

$F^{(-1)}(x) = T(x)$ and,

$$\begin{aligned}
n[x^n]T(x) &= [x^{n-1}]\left(\frac{x}{xe^{-x}}\right)^n \\
\implies n\ \frac{t(n)}{n!} &= [x^{n-1}]e^{nx} \\
\implies n\ \frac{t(n)}{n!} &= \frac{n^{n-1}}{(n-1)!} \\
\implies t(n) &= n^{n-1}.
\end{aligned}$$

Hence, number of labeled rooted trees on n vertices= n^{n-1}. Equivalently, number of unrooted trees, labeled by $[n] = n^{n-2}$.

6.3 A complete proof of Lagrange Inversion Formula

This proof is not seen as a direct consequence of the Cyclic Shift lemma. Instead we make use of one of its Corollaries.

Theorem 6.1 (Lagrange Inversion Formula)**.** Suppose $G(x) \in K[[x]]$ with $G(0) \neq 0$, and let $f(x)$ be defined by

$$f(x) = xG((f(x))).$$

Then,

$$n[x^n]f(x)^k = k[x^{n-k}]G(x)^n.$$

Proof. Let $G(x)$ be the exponential generating function of plane forests on n vertices, labeled by $[n]$.

$$G(x) = \sum_{n \geq 0} t_n \frac{x^n}{n!}.$$

If σ is a labelled forest on $[n]$, then let $r_i(\sigma)$ be the number of vertices with i children, then define, $t^\sigma := \prod t_i^{r_i(\sigma)}$. Now set

$$s_n = \sum_{\tau} t^\tau,$$

where the sum runs over all rooted trees on $[n]$. Suppose,

$$\begin{aligned}
f(x) &= \sum_{n \geq 1} s_n \frac{x^n}{n!} \\
&= t_0 x + 2t_0 t_1 \frac{x^2}{2!} + (6t_0 t_1^2 + 3t_0^2 t_2)\frac{x^3}{3!} + \cdots .
\end{aligned}$$

If τ is a rooted tree on $[n]$, whose root has k children, then τ is obtained by choosing a root $r \in [n]$ and then placing k rooted trees on remaining vertices $[n] \setminus \{r\}$. Hence, by Multiplication principle of Exponential Generating Function,

$$f(x)^k = \sum_{n \geq 1} \left(\sum_{\zeta} t^{\zeta} \right) \frac{x^n}{n!},$$

where ζ runs over all *ordered* k-tuples of rooted trees with total vertex set $[n]$. Thus (since rooted trees are non-empty, so there are $k!$ ways to order k of them on $[n]$),

$$\frac{1}{k!} f(x)^k = \sum_{n \geq 1} \left(\sum_{\sigma} t^{\sigma} \right) \frac{x^n}{n!}, \tag{13}$$

where σ runs over all planted forests on $[n]$ with k components. Next, by Multiplication principle,

$$\frac{t_k}{k!} x f(x)^k = \sum_{n \geq 1} \left(\sum_{\zeta} t^{\zeta} \right) \frac{x^n}{n!},$$

where now ζ runs over all rooted trees on $[n]$ whose root has k children. Summing over all $k \geq 1$ yields, $f(x) = xG((f(x)))$.

Now let k be any positive integer. Then, from equation (13) and recalling equation (10), we obtain,

$$\begin{aligned}
\left[\frac{x^n}{n!}\right] \frac{1}{k!} f(x)^k &= \left[\frac{x^n}{n!}\right] \sum_{n \geq 1} \left(\sum_{\sigma} t^{\sigma} \right) \frac{x^n}{n!} \\
&= \frac{k}{n} \binom{n}{k} \sum_{r_0, r_1, \cdots} \frac{(n-k)!}{\prod_{i \geq 0} (i!)^{r_i}} \binom{n}{r_0, r_1, \cdots} t_0^{r_0} t_1^{r_1} \cdots,
\end{aligned}$$

summed over all $\mathbb{N}$-sequences $r_0, r_1, \cdots$ satisfying $\sum r_i = n$ and $\sum i r_i = n - k$. Equivalently,

$$[x^n] f(x)^k = \frac{k}{n} \sum_{(r_0, r_1, \cdots): \sum r_i = n \,\&\, \sum i r_i = n-k} \binom{n}{r_0, r_1, \cdots} \frac{t_0^{r_0} t_1^{r_1} \cdots}{0!^{r_0} 1!^{r_1} \cdots}.$$

But,

$$\begin{aligned}
G(x)^n &= \left(t_0 + t_1 \frac{x}{1!} + t_2 \frac{x^2}{2!} + \cdots \right)^n \\
&= \sum_{r_0 + r_1 + \cdots = n} \binom{n}{r_0, r_1, \cdots} \frac{t_0^{r_0} t_1^{r_1} \cdots}{0!^{r_0} 1!^{r_1} \cdots} x^{\sum i r_i}.
\end{aligned}$$

Thus, $[x^n] f(x)^k = \dfrac{k}{n} [x^{n-k}] G(x)^n$, as desired.

Note that the above is a general and complete proof of Lagrange Inversion Formula. Though, the used $G(x)$ and $f(x)$ has some special structure counting trees, they are written in terms of indeterminates $t_0, t_1, \cdots$. Any two power series $G_1(x)$ and $f_1(x)$ obeying $f_1(x) = xG_1((f_1(x)))$, can be written in form of the used $G(x)$ and $f(x)$ by assigning certain values to the indeterminates. This completes the proof. $\square$

7 List of Referred Books and Notes

I referred to some parts of these following books and notes to prepare this report.

(i) Remco van der Hofstad. RANDOM GRAPHS AND COMPLEX NETWORKS Volume I (Chapter 3).

(ii) Richard P. Stanley. Enumerative Combinatorics Volume II (Chapter 5).

(iii) J. Pitman. Combinatorial Stochastic Processes Ecole d'Eté de Probabilités de Saint-Flour XXXII – 2002 (Chapter 6).

(iv) Lecture notes of a course in Combinatorics by Mark Haiman. (Available here.)

8 Conclusion

Here, I list the topics in order that I read and learnt while doing this project work.

(i) I revised the basics of Branching Processes before learning the Random Walk perspective of Branching Processes and the *Kemperman's Formula*.

(ii) I read a proof of *Kemperman's Formula* using mathematical induction. Next, I learnt the law of total progeny of a branching process.

(iii) The main topic, proof of *Cayley's formula* came next.

(iv) For further reading, I consulted Stanley's book (Enumerative Combinatorics Volume II) to read tree counting in more detail and find how Lagrange Inversion Formula is related. I also learnt exponential generating function here.

(v) Next, from Pitman's book (Combinatorial Stochastic Processes Ecole d'Eté de Probabilités de Saint-Flour XXXII – 2002), I read about cyclic shifts and Lagrange Inversion. It gave some nice insight. I found a more intuitive proof of *Kemperman's Formula* here.

(vi) Finally, I wrapped with more detailed tree counting as consequences of cyclic shift lemma.

Appendices

1 Introduction to Exponential Generating Functions and Rooted labelled trees

Definition 1.1 (Exponential Generating Functions)**.** *$f(n)$ be the number of ways to impose a certain structure on a set of n elements. Then, $F(n)$, the exponential generating function of that specific structure is defined as,*

$$F(n) = \sum_n f(n)\,\frac{x^n}{n!}.$$

Example 1.2. Consider the trivial structure, which is the structure of being a "set". Trivially, for any given set there can be only one such structure which is the one we already have. Thus, for all n, $f(n) = 1$. Thus, the exponential generating function is,

$$F(x) = 1 + 1.\frac{x}{1!} + 1.\frac{x^2}{2!} + 1.\frac{x^3}{3!} + \cdots = e^x.$$

1.1 Addition and Multiplication Principle

In order to be able to apply exponential generating functions, we need to know what it means to add or multiply them. Say we have some kinds of structures, f-structures, g-structures, h-structures. We let $f(n)$, $g(n)$ and $h(n)$ be respectively the number of f, g and h-structures on a set of size n. Their generating functions are respectively $F(x)$, $G(x)$ and $H(x)$.

Result 1.3 (Addition principle for exponential generating functions)**.** Suppose that the set of f-structures on each set is the disjoint union of the sets of g-structures and h-structures on that set. Then,

$$F(x) = G(x) + H(x).$$

Example 1.4. The addition principle is rather obvious. To illustrate this, we may again use the structure of being a "set". Trivially, for any given set, it can be either empty or non-empty and these two structures are mutually exclusive. So, if one defines $f =$trivial structure, $g =$empty set structure and $h =$non-empty set structure, it follows that, $F(x) = G(x) + H(x)$. Now, $g(n) = 1$ if $n = 0$, and $g(n) = 0$ for any other $n \geq 1$. On the other hand, $h(n) = 0$ if $n = 0$, and $h(n) = 1$ for any other $n \geq 1$. Thus, $G(x) = 1$ and $H(x) = e^x - 1$. That makes, $F(x) = e^x$ which matches our previous example.

Result 1.5 (Multiplication principle for exponential generating functions)**.** Say, X is any given set. Then, number of f-structures on X follows the following rule,

$$f(|X|) = \sum_{(S,T)} g(|S|)h(|T|),$$

where (S, T) runs over all weak ordered partitions of X into two blocks, i.e., $S \cap T = \phi$ and $S \cup T = X$. Then,

$$F(x) = G(x)H(x).$$

Example 1.6. We aim to find the exponential generating function counting the number of non-empty subsets of a n-element set, A. We might partition A in A_1 and A_2 such that, A_1 will be considered the subset. Then A_1 will need to have the structure of "non-empty set", while A_2 is required to have the trivial structure of being a "set". Generating functions associated with the structure of A_1 and A_2 are respectively $G(x) = e^x - 1$ and $H(x) = e^x$. Thus, the required generating function is,

$$F(x) = G(x)H(x) = (e^x - 1)(e^x) = e^{2x} - e^x = \sum_{n\geq 0}(2^n - 1)\frac{x^n}{n!}.$$

This is the expected answer, since we knew that there are $(2^n - 1)$ many subsets of an n-element set.

1.2 Composition Principle

This principle exhibits the real strength of exponential generating function. We shall assume f, g and h-structures and their generating functions to be $F(x), G(x)$ and $H(x)$ as before. We have one additional restriction that, $h(0) = 1$.

Result 1.7 (Composition principle for exponential generating functions)**.** Say, X is any given set. Then, number of f-structures on X follows the rule,

$$\begin{aligned} f(|X|) &= \sum_{\pi\ \in \Pi(X)} g(|B_1|)g(|B_2|)\cdots g(|B_k|)h(k), \qquad |X| > 0, \\ f(0) &= 1, \end{aligned}$$

where the sum ranges over all partitions $\pi = \{B_1, B_2, \cdots, B_k\}$ of the finite set X. Then,

$$F(x) = H(G(x)).$$

(Here, $G(x) = \sum\limits_{n\geq 1} g(n)\dfrac{x^n}{n!}$, since g is defined on positive integers only.)

Example 1.8. Let $f(n)$ be the number of ways for n persons to form into non-empty lines and then arrange the lines in a circular order. There are $g(j) = j!$ ways to arrange j people in a line and $h(k) = (k-1)!$ ways to circularly arrange $k(> 0)$ lines. Thus,

$$\begin{aligned} G(x) &= \sum_{n\geq 1} n!\frac{x^n}{n!} = \frac{x}{1-x} \\ H(x) &= 1 + \sum_{n\geq 1}(n-1)!\frac{x^n}{n!} = 1 + \frac{1}{log(1-x)}. \end{aligned}$$

Hence,

$$\begin{aligned}
F(x) &= H(G(x)) \\
&= 1 + log\left(1 - \frac{x}{1-x}\right)^{-1} \\
&= 1 + log(1-2x)^{-1} - log(1-x)^{-1} \\
&= 1 + \sum_{n\geq 1}(2^n - 1)(n-1)!\frac{x^n}{n!}.
\end{aligned}$$

Thus, $f(n) = (2^n - 1)(n-1)!$. This result can be verified by direct combinatorial argument.

1.3 Exponential Generating Function for Rooted Labelled Trees

Let $t(n)$ denote the number of rooted labelled trees on an n-element set. We also assume $t(0) = 0$, i.e., the empty tree is not counted as a labelled tree. Then the corresponding exponential generating function is,

$$T(x) = \sum_{n\geq 1} t(n)\frac{x^n}{n!}.$$

Now, we observe that a rooted forest structure consists of smaller rooted forests each of whose roots are joined to the root of the entire forest. So, let us name some smaller structures. Say, g-structure = singleton set, h-structure = set. Then, $G(x) = x$ and $H(x) = e^x$. Then, using the multiplication and composition principles,

$$T(x) = G(x)H(T(x)) = xe^{T(x)}. \tag{14}$$

Using the above identity, $T(x)e^{-T(x)} = x$. Here, substituting $T^{-1}(x)$ in place of x yields $T^{-1}(x) = xe^{-x}$. In order to obtain an explicit formula for $t(n)$, we shall have to find a power series expansion for $T(x)$. *Lagrange Inversion Formula* helps us obtain the power series expansion. The inversion formula is discussed in detail in a Subsection 5.2. The proof is given in short in Subsection 6.2.