Expected size of a tree in the fixed point forest

Samuel Regan[1] and Erik Slivken[2]*

[1] *University of California Davis*
[2] *Dartmouth College*

We study the local limit of the fixed-point forest, a tree structure associated to a simple sorting algorithm on permutations. This local limit can be viewed as an infinite random tree that can be constructed from a Poisson point process configuration on $[0,1]^{\mathbb{N}}$. We generalize this random tree, and compute the expected size and expected number of leaves of a random rooted subtree in the generalized version. We also obtain bounds on the variance of the size.

Keywords: sorting algorithms, random trees, Poisson point processes, random permutations

1 Introduction

We start with a simple sorting algorithm on a deck of cards labeled 1 though n. If the value of the top card is i, place it in the ith position from the top in the deck. Repeat until the top card is a 1. Viewing the deck of cards as a permutation in one-line notation $\pi = \pi(1)\pi(2)\cdots\pi(n)$, we create a new permutation, $\tau(\pi)$, by removing the *value* $\pi(1)$ from beginning of the permutation and putting it into *position* $\pi(1)$. For example, if $\pi = 43512$ then $\tau(\pi) = 35142$. This induces a graph whose vertices are the permutations of $[n] = \{1,\cdots,n\}$ and edges are pairs of permutations $(\pi, \tau(\pi))$. Note that $\tau(\pi)$ has a fixed point at the position $\pi(1)$.

This graph is a rooted forest, which we denote by F_n and call the *fixed point forest*. A rooted forest is a union of rooted trees, and a tree is a graph that does not contain any closed loops involving distinct vertices. A permutation that begins with 1 is called the base of the tree in which they are contained. A thorough introduction to the fixed point forest can be found in Johnson et al. (2017).

The fixed point forest was first studied in McKinley (2015). The largest tree in F_n has size bounded between $(n-1)!$ and $e(n-1)!$ and has as its base the identity permutation. The longest path from a leaf to a base is $2^{n-1} - 1$ and is unique, starting from the permutation $23\cdots n1$ and ending at the identity.

Let $\mathfrak{S}_n$ denote the set of permutations of length n. For $\pi \in \mathfrak{S}_n$, let $\mathcal{F}(\pi)$ denote the collection of fixed points of π other than 1. For each $m \in \mathcal{F}(\pi)$ we create a new permutation $\pi^{(m)}$ such that

$$
\pi^{(m)}(i) = \begin{cases} m, & i = 1 \\ \pi(i-1), & 2 \leq i \leq m \\ \pi(i), & m < i \leq n \end{cases}.
$$

```
┌───────┐
│ 34512 │
└───────┘
┌───────┐
│ 54312 │
└───────┘
┌───────┐                          ┌───────┐
│ 43125 │                          │ 53124 │
└───────┘                          └───────┘
            ┌───────┐
            │ 31245 │
            └───────┘
```
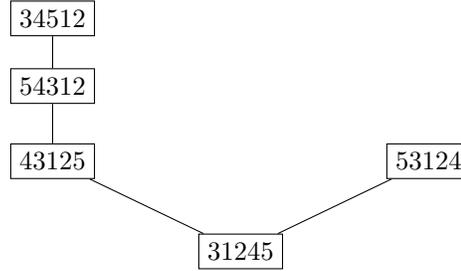
Fig. 1: The descendant tree $desc(\pi)$ for $\pi = 31245$

We say we *bump* the value m in π to create $\pi^{(m)}$ and call $\pi^{(m)}$ a *child* of π. We let $\mathcal{C}(\pi) = \{\pi^{(m)} : m \in \mathcal{F}(\pi)\}$ denote the set of children of π. Every child $\sigma \in \mathcal{C}(\pi)$ satisfies $\tau(\sigma) = \pi$ hence is connected to π in F_n.

Let $N(\pi)$ be the rooted tree in F_n that contains π, with π designated as the root instead of the unique permutation that starts with 1 in $N(\pi)$. Let $desc(\pi)$ be the subtree of $N(\pi)$ rooted at π and consisting of π and its descendants, so that $desc(\pi) \subseteq N(\pi)$. We call this the *descendant tree* of π (See Figure 1). Note that for any permutation $\sigma \in desc(\pi)$, there is some r such that $\tau^r(\sigma) = \pi$.

By Theorem 3.5 in Johnson et al. (2017), there exists a tree, $\mathbf{T}$, such that as $n \to \infty$, for π_n chosen uniformly at random from permutations of size n, the randomly rooted tree $\mathbf{N}_n = N(\pi_n)$, converges in the local weak sense to $\mathbf{T}$. This limiting tree is described in Section 2 of Johnson et al. (2017), and the subtree of $\mathbf{T}$ which corresponds to the local weak limit of $desc(\pi_n)$ has a similar description, denoted by $\mathbf{D}$. In Johnson et al. (2017), they find the distribution for the shortest and longest paths from the root to a leaf in $\mathbf{D}$. The main purpose of the paper is to study the size of $\mathbf{D}$. For $\alpha \in [0, 1]$, we define a generalization of $\mathbf{D}$, denoted $\mathbf{D}_\alpha$ such that $\mathbf{D} = \mathbf{D}_1$. We compute the expected size and expected number of leaves of $\mathbf{D}_\alpha$ and show that they are both unbounded for $\alpha = 1$. Finally we find bounds on the second moment of the size of $\mathbf{D}_\alpha$. We show that the second moment has a phase transition from finite to infinite somewhere between $(3 - \sqrt{5})/2$ and $(\sqrt{5} - 1)/2$.

2 Local limits, point process configurations, and trees

Poisson Point Processes

The following briefly introduces an important probabilistic object: Poisson point processes. A thorough treatment can be found in Kingman (1993).

We say a random variable X is $\mathrm{Poi}(\alpha)$ if it satisfies $\mathbf{P}(X = k) = \frac{1}{k!}e^{-\alpha}\alpha^k$. If X_0 and X_1 are two independent $\mathrm{Poi}(\alpha_0)$ and $\mathrm{Poi}(\alpha_1)$, respectively, then their sum is $\mathrm{Poi}(\alpha_0 + \alpha_1)$.

A point process on $[0, 1]$ is an integer-valued measure on Borel sets of $[0, 1]$. It may be viewed as a collection of points, which represent the atoms of the measure. A point process configuration on $[0, 1]$ is a collection of point processes, each on $[0, 1]$, and can be viewed as a collection of labelled points on $[0, 1]$.

A Poisson point process on $[0, 1]$ with intensity α is a random integer-valued measure which satisfies two properties: For any Borel subset $E \subset [0, 1]$ with Borel measure λ, the number of atoms of the point process in E is given by $\mathrm{Poi}(\alpha\lambda)$, and for any disjoint Borel subsets of $[0, 1]$ the number of atoms in each are independent. Conditioned on the number of atoms in E the location of each of the atoms is
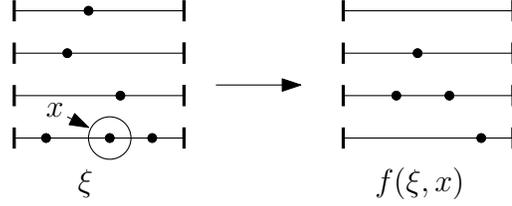
Fig. 2: The bump map $f(\xi, x)$ where ξ_4 is assumed to be empty.

independent and uniform in E.

Collections of Poisson point processes can be merged to create a single poisson point process. Suppose ξ_0 is a $\text{Poi}(\alpha_0)$ point process on $[0, 1]$ and ξ_1 is $\text{Poi}(\alpha_1)$ point process on $[0, 1]$ with ξ_0 and ξ_1 both independent. Then the union of ξ_0 and ξ_1 is distributed like a $\text{Poi}(\alpha_0 + \alpha_1)$ point process. The reverse is also true. Let ξ' be a $\text{Poi}(\alpha_0 + \alpha_1)$ point process on $[0, 1]$ and label each atom 0 with probability $\alpha_0/(\alpha_0 + \alpha_1)$ and 1 otherwise. Let ξ_0 denote the point process consisting of the atoms labeled 0 and ξ_1 the point process of the remaining atoms. Then ξ_0 and ξ_1 are, respectively, independent Poisson(α_0) and Poisson(α_1) point processes on $[0, 1]$. This can be generalized further to $\alpha = \alpha_0 + \cdots + \alpha_{k-1}$. If ξ' is a Poisson(α) point process each atom in ξ' is independently labeled such that the label is i with probability α_i/α for $0 \leq i < k$, then the collection of atoms labeled i is a Poisson(α_i) point process and each ξ_i is independent of the rest.

Let ξ_1 and ξ_2 be two independent Poisson(α) point processes. For $x \in (0, 1)$, define $\xi_1' = \xi_2\big|_{[0,x)} + \xi_1\big|_{(x,1]}$ to be the point process consisting of the atoms from ξ_2 restricted to the interval $[0, x)$ and the atoms from ξ_1 restricted to the interval $(x, 1]$. If x is independent of ξ_1 and ξ_2 then the resulting process ξ_1' is also a Poisson(α) point process.

Weak Convergence

We give a brief definition of the version of local weak convergence that is used to define $\mathbf{T}$ and $\mathbf{D}$. See Aldous and Steele (2004) or Benjamini and Schramm (2001) for a proper discussion of local weak convergence, which is sometimes referred to as Benjamini-Schramm convergence.

Let $G_1, G_2 \cdots$ be a sequence of rooted graphs. For any rooted graph H, the r-neighborhood of the root, denoted $H(r)$, is the subgraph of H induced from all vertices that are distance at most r from the root. The rooted graph G is the *local weak limit* of G_n if for every $r \geq 0$ and every finite graph H,

$$\mathbf{P}[G_n(r) = H] \to \mathbf{P}[G(r) = H].$$

From point process configurations to trees

Let $\xi = (\xi_k)_{k \geq 0}$ be a point process configuration on $[0, 1]^{\mathbb{N}}$ where each ξ_k is a point process on $[0, 1]$. For each atom $x \in \xi_0$ define the bump map $f(\xi, x) = (\xi_k')_{k \geq 0}$ where

$$\xi_k' = \xi_{k+1}\Big|_{[0,x)} + \xi_k\Big|_{(x,1]}.$$

See Figure 2 for an illustration of this map. Given a point process configuration, ξ, the bump map allows us to recursively define a tree with root v_0 whose vertices are point process configurations. Define v_0 to be
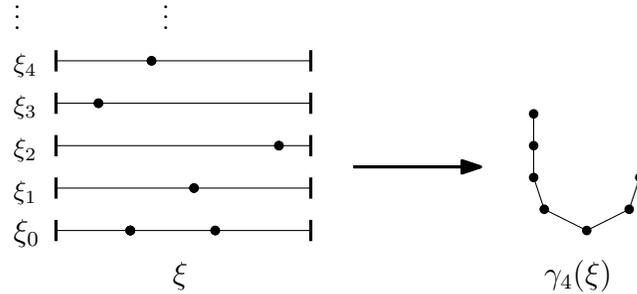
Fig. 3: A point process collection and corresponding 4-neighborhood of the bump tree. Note that any configuration of point processes for ξ_5 and higher will not affect the structure of the bump tree and thus $\gamma_4(\xi) = \gamma(\xi)$.

the root of the tree with corresponding point process configuration $\xi^{v_0} = \xi$. Suppose v is a vertex in the tree with corresponding point process configuration given by ξ^v. For each $x \in \xi_0^v$, create a new vertex $v(x)$ in the tree with point process configuration given by the bump map $\xi^{v(x)} = f(\xi^v, x)$. The newly created vertex $v(x)$ is a considered a child of v. We call this tree the *bump tree* of ξ and denote it by $\gamma(\xi)$. For fixed $r \geq 0$ let $\gamma_r(\xi)$ denote the r-neighborhood of the root in $\gamma(\xi)$. Only the atoms in $(\xi_0, \cdots, \xi_{r-1})$ are necessary to determine the structure of the $\gamma_r(\xi)$, so we may write $\gamma_r(\xi) = \gamma_r(\xi_0, \cdots, \xi_{r-1})$ and assume $\xi_k = \emptyset$ for $k \geq r$. The map γ_r is continuous because a slight perturbation of the atoms will not change the relative order of the points in $(\xi_0, \cdots, \xi_r)$. See Figure 3 for an example of a finite neighborhood of the root of the bump tree for a point process configuration.

For a permutation π of length n, we say the index i or the value $\pi(i)$ is k-*separated* if $\pi(i) = i + k$. We define the *separation word* of π point-wise by $\mathbf{W}^\pi(i) := \pi(i) - i$. No two permutations have the same separation word. From this word we can construct a point process configuration $(\xi_k^\pi)_{k \geq 0}$ by placing an atom in ξ_k^π at position i/n if i is a k-separated point in π.

By Proposition 3.4 in Johnson et al. (2017), for fixed $r \geq 0$, as n tends to infinity,

$$(\xi_0^{\pi_n}, \cdots, \xi_{r-1}^{\pi_n}) \longrightarrow_d (\xi_0, \cdots, \xi_{r-1})$$

where ξ_k is a $\mathrm{Poi}(1)$ point process on $[0, 1]$. From the arguments of Theorem 3.5 in Johnson et al. (2017), letting $\xi = (\xi_k)_{k \geq 0}$, we have $\gamma_r(\xi^{\pi_n}) \to \gamma_r(\xi)$ by continuity of γ_r and the Continuous Mapping Theorem [Billingsley (1999)]. Furthermore, it is seen that $\gamma_r(\xi^{\pi_n})$ is the same as the r-neighborhood of the descendant tree $desc(\pi_n)$ with high probability. Therefore $\mathbf{D} := \gamma(\xi)$ is the local weak limit of $desc(\pi_n)$.

We now can state our main results. For $\alpha \in (0, 1]$, let $\xi = (\xi_k)_{k \geq 0}$ be a collection of independent $\mathrm{Poi}(\alpha)$ point processes on $[0, 1]$ and let $\mathbf{D}_\alpha := \gamma(\xi)$ be the corresponding bump tree of ξ. Let D denote the number of vertices and U the number of leaves in $\mathbf{D}_\alpha$. Finally let $\mathbf{E}_\alpha$ and $\mathbf{P}_\alpha$ denote the expectation and probability associated with $\mathrm{Poi}(\alpha)$ point processes. We now may state our main results.

Theorem 1. *For $0 < \alpha < 1$, $\mathbf{E}_\alpha[D] = (1 - \alpha)^{-1}$, and $\mathbf{E}_1[D]$ diverges.*

Theorem 2. *For $0 < \alpha < 1$, $\mathbf{E}_\alpha[U] = e^{-\alpha}(1 - \alpha)^{-1}$, and $\mathbf{E}_1[U]$ diverges.*

Theorem 3. *For $\alpha \geq (\sqrt{5} - 1)/2$, $\mathbf{E}_\alpha(D^2)$ diverges. For $\alpha < (3 - \sqrt{5})/2$, $\mathbf{E}_\alpha(D^2)$ is finite.*
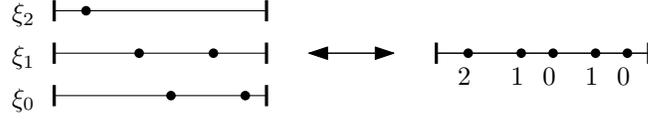
Fig. 4: A collection of point processes corresponding to the word 2 1 0 1 0.

3 Comparison with Galton-Watson trees

In this section we compare our results to the well-studied Galton-Watson tree Watson and Galton (1875); Neveu (1986).

A Galton-Watson tree, $\mathbf{GW}$, can be constructed through a simple random process. Start with a root v_0 and a nonnegative integer-valued random variable X. Create X_{v_0} children of v_0 where X_{v_0} is distributed as and independent copy of X. For each child, v, of v_0 repeat this process, where X_v is an independent copy of X. Depending on the distribution of X, the resulting tree will have drastically different behavior.

Fix a nonnegative integer-valued random variable X with finite expectation $0 < \mathbf{E}[X] < 1$ and finite second moment $\mathbf{E}[X^2] < \infty$. Let $Y = |\mathbf{GW}|$. Let X denote the number of children of the root of $\mathbf{GW}$ and for $1 \le i \le X$, let Y^i denote the number of vertices in the subtree consisting of the ith child and all of its descendants. Each Y^i is distributed identically as an independent copy of $\mathbf{GW}$. We denote the size of $\mathbf{GW}$ conditioned on X by $(Y|X) = 1 + \sum_{i=1}^{X} Y^i$. Taking expectation we have $\mathbf{E}[(Y|X)] = 1 + X\mathbf{E}[Y]$ and thus

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[(Y|X)]] = 1 + \mathbf{E}[X]\mathbf{E}[Y]$$

and so

$$\mathbf{E}[Y] = \frac{1}{1 - \mathbf{E}[X]}.$$

A similar approach for the second moment gives the equation

$$\mathbf{E}[Y^2] = 1 + \mathbf{E}[X]\mathbf{E}[Y] + \mathbf{E}[X]\mathbf{E}[Y^2] + \mathbf{E}[X^2 - X]\mathbf{E}[Y]^2,$$

which can be simplified to

$$\mathbf{E}[Y^2] = \frac{1}{(1 - \mathbf{E}[X])^2} + \frac{\mathbf{E}[X^2] - \mathbf{E}[X]}{(1 - \mathbf{E}[X])^3}. \tag{1}$$

Given that $\mathbf{E}[X] < 1$ and $\mathbf{E}[X^2]$ is finite, (1) shows that $\mathbf{E}[Y^2]$ finite. In particular if X is $\mathrm{Poi}(\alpha)$ then $\mathbf{E}[Y]$ agrees with $\mathbf{E}_\alpha[D]$ from Theorem 1, while Theorem 3 shows the second moment $\mathbf{E}[Y^2]$ cannot agree with the second moment $\mathbf{E}_\alpha[D^2]$ if $\alpha \ge (\sqrt{5} - 1)/2$ since the former is finite while the latter diverges.

The approach used to compute $\mathbf{E}[Y]$ and $\mathbf{E}[Y^2]$ cannot be used to compute $\mathbf{E}_\alpha[D]$ and $\mathbf{E}_\alpha[D^2]$ because the subtrees from the root in $\mathbf{D}_\alpha$ are not independent of each other.

4 Words from point process configurations

For a collection of point processes on $[0,1]$, $\xi = \{\xi_k\}_{k \ge 0}$, let $w_r(\xi)$ be the word constructed from the relative order of the atoms in $(\xi_0, \cdots, \xi_{r-1})$. For example see Figure 4. Assuming that no two atoms of ξ

are in the same location, the structure of the r-neighborhood of the root in the tree $\gamma_r(\xi)$ can be constructed directly from this word. Let Ω_r denote the space of finite words with letters from $\{0, \cdots, r-1\}$.

If ξ is a $\mathrm{Poi}(\alpha)$ point process configuration, this induces a probability measure $\mathbf{P}_{\alpha,r}$ on Ω_r for every $r \geq 0$. The following lemma describes this distribution.

Lemma 4. *Let ξ be a $\mathrm{Poi}(\alpha)$ point process configuration and $W = w_r(\xi)$ the word given by the relative order of the first r point processes of ξ. Let w denote a fixed word of length n in Ω_r. Then*

$$\mathbf{P}_{\alpha,r}(|W| = n) = \frac{1}{n!} e^{-\alpha r} \alpha^n r^n \tag{2}$$

and

$$\mathbf{P}_{\alpha,r}(W = w) = \frac{1}{n!} e^{-\alpha r} \alpha^n. \tag{3}$$

Proof:

Construct the r independent $\mathrm{Poi}(\alpha)$ point processes from a single $\mathrm{Poi}(r\alpha)$ point process by labeling each atom independently from $\{0, \cdots, r-1\}$, choosing the label uniformly at random. The probability that $|W| = n$ is precisely the probability that a $\mathrm{Poi}(r\alpha)$ point process has n atoms in $[0, 1]$, the right hand side of (2). As the labeling is independent for each atom, each of the r^n possible labelings is equally likely, so the probability that $W = w$ for a fixed w of length n is computed by dividing the right hand side of (2) by r^n, giving (3). □

For $W \in \Omega_r$ of length n we write $W = W_1 \cdots W_n$ in one line notation. For a fixed subset of indices $A = (i_1, \cdots, i_j)$ let $W_A = W_{i_1} \cdots W_{i_j}$. We may refine Lemma 4 even further.

Lemma 5. *Let $u = u_1 \cdots u_j$ be a word in Ω_r. Let $W \in \Omega_r$, and $A = (i_1, \cdots, i_j)$ be a set of indices such that $1 \leq i_1 < \cdots < i_j \leq n$. Then,*

$$\mathbf{P}_{\alpha,r}\left(\{W_A = u\} \cap \{|W| = n\}\right) = \frac{1}{n!} e^{-\alpha r} \alpha^n r^{n-j}.$$

Proof:

Conditioned on $|W| = n$, the labels of the atoms indexed by A are chosen independently so

$$\mathbf{P}_{\alpha,r}(W_A = u \,|\, |W| = n) = r^{-j}$$

and the statement follows. □

The tree $\gamma_r(\xi)$ with word $w_r(\xi)$ will agree up to a relabeling of the vertices of the tree $\gamma_r(\xi')$ if $w_r(\xi) = w_r(\xi')$. A vertex in the tree corresponds to bumping a particular set of atoms in a particular order. Therefore the measure $\mathbf{P}_{\alpha,r}$ on words in Ω_r is exactly the measure we need to understand the $\gamma_r(\xi)$.

We can translate our language of bumping atoms in ξ to bumping letters in words. Let $W \in \Omega_r$. For each $0 \in W$, we construct a new word by removing the chosen 0 and reducing every letter to the left of it by 1. We say the index of this letter 0 is *bumped* and indices less than the bumped index are *shifted*. The set of indices of the 0s in a word are called the *bumpable* indices. The set of words that can be constructed by bumping a single 0 in W are called the children of W and denoted $\mathcal{C}(W)$. For example the word 2 1 0 1 0 has has two children, 1 0 □ 1 0 and 1 0 □ 0 □, where □ is used to indicate bumped indices
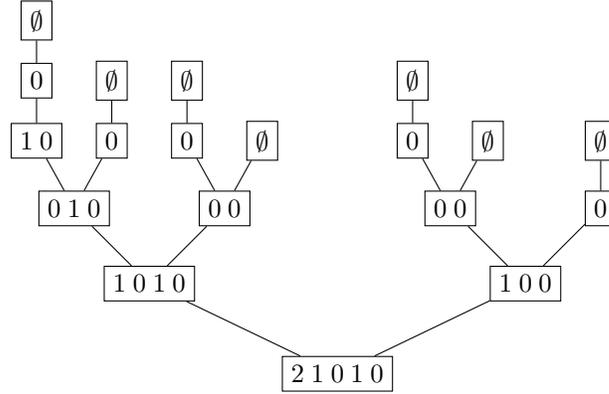
Fig. 5: The tree, $\gamma(w)$, for the root word $w = 2\,1\,0\,1\,0$

or indices shifted below zero. Once the letter at an index becomes $\square$ in a word it can never become 0 in one of its descendants. We construct a rooted tree, denoted $\gamma(W)$, following a process that mirrors our construction of $\gamma(\xi)$ for point process configurations. We let $\gamma_j(W)$ denote the j-neighborhood of the root in $\gamma(W)$.

We may omit the $\square$ symbol in the labeling of the tree. The $\square$ symbol is used to emphasize that the set of indices is the same for each word in the same tree. See Figure 5 for the rooted tree in Ω_3 associated with the word $2\,1\,0\,1\,0$. The sequence of indices that are bumped to reach the vertex v in $\gamma(W)$ is called the bumping sequence of v.

For $j \geq 1$ and every vertex $v \in \gamma_j(W) \backslash \gamma_{j-1}(W)$ there is a corresponding set of j atoms that must be bumped in a particular order to reach v. This sequence of atoms induces an ordered set of indices $A = \{a_1 < \cdots < a_j\}$ and permutation, σ, of length j such that v is obtained by bumping the atoms at the indices in order $\{a_{\sigma_1}, \cdots, a_{\sigma_j}\}$ where each of the indices must be 0 when they are bumped. We say the set of indices A reaches v by the order σ. Since $\gamma(W)$ is a tree, any such v is reachable by a unique pair (A, σ).

For a set of indices $A = \{a_1 < \cdots < a_j\}$, we say A is *complete* in W if there exists an order $\sigma \in \mathfrak{S}_j$ and a sequence of words $W = W^0, \cdots, W^j$ such that for $1 \leq i \leq j$, $W^i \in \mathcal{C}(W^{i-1})$ is obtained by bumping the index a_{σ_i} in W^i. Whether or not A is complete in W is independent of the letters not in A. The following lemma gives conditions on when A is complete in W.

Lemma 6. *If A is complete in $W \in \Omega_r$ with $|A| = j$, there is a unique $\sigma \in \mathfrak{S}_j$ such that a vertex in $\gamma(W)$ is reachable by (A, σ). If $r \geq j$, then for each $\sigma \in \mathfrak{S}_j$ there is a unique sequence of values $u = u_1 \cdots u_j$ such if $W_A = u$ then there exists a vertex in $\gamma(W)$ that is reachable by (A, σ).*

Finally, A is complete with respect to W if and only if $W_{a_i} \leq \min(j - i, r - 1)$ for $1 \leq i \leq j$.

Proof:

Since A is complete in W there is at least one $\sigma \in S_j$ and v in $\gamma(W)$ such that v is reachable by (A, σ). First a_{σ_1} is bumpable if and only if $W_{a_{\sigma_1}} = 0$. In order for $a_{\sigma_{i+1}}$ to be bumpable after bumping a_{σ_1} up to a_{σ_i}, the label of $a_{\sigma_{i+1}}$ must be 0, and therefore index must be shifted exactly $W_{a_{\sigma_{i+1}}}$ times by bumping indices larger then $a_{\sigma_{i+1}}$. For this to occur there must be exactly $W_{a_{\sigma_{i+1}}}$ integers m such that $m < i + 1$

and $\sigma_m > \sigma_{i+1}$. In terms of σ^{-1} we have for $1 \le i \le j$,

$$W_{a_i} = \#\{i < m \le j | \sigma_i^{-1} > \sigma_m^{-1}\}.$$

The sequence of values $W_{a_1} \cdots W_{a_j}$ is the unique inversion table (Knuth (1998)) for the permutation σ^{-1}. No two permutations have the same inversion table and thus σ must be unique. Given a $\sigma \in \mathfrak{S}_j$, if W_A is the inversion table for σ^{-1} then A will be complete with respect to W.

Finally we have that $W_{a_1} \cdots W_{a_j}$ is an inversion table if and only if $W_{a_i} \le j - i$ for $1 \le i \le j$. We also have that $W_{a_i} \le r - 1$ by definition. □

Define the following truncated factorial function:

$$f_y(x) = \left\{ \begin{array}{ll} x!, & x \le y, \\ y! y^{x-y}, & y < x. \end{array} \right.$$

Note that $\lim_{y \to \infty} f_y(x) = x!$.

Let $\beta_r(j)$ denote the set of subwords of length j such such that A is complete in W if and only if $W_A \in \beta_r(j)$. For any $r \ge 0$ and $j \ge 0$, by Lemma 6,

$$|\beta_r(j)| = f_r(j)$$

and for $r \ge j$, this simplifies to

$$|\beta_r(j)| = j!.$$

5 Expectation of D and U

Let $D^{(r)}$ denote the number of vertices in $\gamma_r(\xi)$. Let $U^{(r)}$ denote the number of leaves in $\gamma_r(\xi)$ that are distance less than r from the root. Note that a leaf in $\gamma_r(\xi)$ that is distance r from the root may not be a leaf in $\gamma_{r+1}(\xi)$. By Theorem 5.1 in Johnson et al. (2017), the longest path to a leaf in $\gamma(\xi)$ is almost surely finite and therefore $\gamma_r(\xi)$ is identical to $\gamma(\xi)$ for large enough r. To compute the expectation of D and U it suffices to compute the expectation of $D^{(r)}$ and $U^{(r)}$ and let r tend to infinity.

Let W be chosen from Ω_r. For $j \le r$ let $D_j^{(r)} = |\gamma_j(W) \backslash \gamma_{j-1}(W)|$. Similarly let $\mathcal{L}_j$ denote the set of leaves in $\gamma_j(W)$, so that for $j \le r - 1$, $U_j^{(r)} = |\mathcal{L}_j(W) \backslash \mathcal{L}_{j-1}(W)|$, the number of leaves in $\gamma_j(W)$ exactly distance j from the root. By linearity of expectation

$$\mathbf{E}_{\alpha,r}[D^{(r)}] = \sum_{j=0}^{r} \mathbf{E}_{\alpha,r}[D_j^{(r)}]$$

and

$$\mathbf{E}_{\alpha,r}[U^{(r)}] = \sum_{j=0}^{r-1} \mathbf{E}_{\alpha,r}[U_j^{(r)}].$$

For a fixed $j \le n$, let $\mathcal{A}$ be the set of all subsets of j indices $A \subseteq [n]$. Consider a fixed $A \in \mathcal{A}$ and a word u of length j with letters less than r. If a word $W \in \Omega_r$ has length n, there are r^{n-j} possible fillings of the indices in $[n] \setminus A$ and there are $f_r(j)$ ways to fill the indices of A so that A is complete in W.

By Lemma 5 we have

$$\mathbf{P}_{\alpha,r}\left(\{A \text{ is complete in } W\} \cap \{|W| = n\}\right) = e^{-\alpha r}\alpha^n r^{n-j}f_r(j)/n!. \tag{4}$$

By the one-to-one correspondence with complete indices A in W of size j with vertices in $\gamma(W)$ exactly distance j from the root, the expectation of $D_j^{(r)}$ is

$$\mathbf{E}_{\alpha,r}[D_j^{(r)}\mathbf{1}_{|W|=n}] = \sum_{A \in \mathcal{A}} e^{-\alpha r}\alpha^n r^{n-j}f_r(j)/n! = e^{-\alpha r}\alpha^n r^{n-j}f_r(j)/(j!(n-j)!). \tag{5}$$

For $r \geq j$,

$$\mathbf{E}_{\alpha,r}[D_j^{(r)}\mathbf{1}_{|W|=n}] = e^{-\alpha r}\alpha^n r^{n-j}/(n-j)!, \tag{6}$$

and $\mathbf{E}_{\alpha,r}[D_j^{(r)}] = \sum_{n \geq j}\mathbf{E}[D_j^{(r)}\mathbf{1}_{|W|=n}]$, so

$$\mathbf{E}_{\alpha,r}[D_j^{(r)}] = \alpha^j e^{-\alpha r}\sum_{n \geq j}\frac{(\alpha r)^{n-j}}{(n-j)!} = \alpha^j. \tag{7}$$

Proof of Theorem 1: From (7), $\mathbf{E}_{\alpha,r}[D_j^{(r)}] = \alpha^j$ for $j \leq r$ and $\mathbf{E}_{\alpha,r}[D^{(r)}] = \sum_{j=0}^r \alpha^j$. Then $\lim_{r \to \infty} D^{(r)} = D$ and by Monotone Convergence Theorem

$$\mathbf{E}_{\alpha}[D] = \lim_{r \to \infty}\mathbf{E}_{\alpha,r}[D^{(r)}] = \lim_{r \to \infty}\frac{1 - \alpha^{r+1}}{1 - \alpha} = \frac{1}{1 - \alpha}.$$

$\square$

Expected number of leaves

For a set of indices A of size j that are complete in W, let X denote the word obtained after bumping every index in A. The vertex labelled with X is a leaf if it contains no bump-able indices, that is X has no 0s. Let $a_0 = 0$ and $a_{j+1} = |W| + 1$. For $0 \leq i \leq j$, an index $b_i \in (a_i, a_{i+1})$ is bump-able in X if and only if $W_{b_i} = j - i$. If $r \leq j$ and $i \leq j - r$, $W_{b_i} < r \leq j - i$ and hence b_i cannot be bump-able. Otherwise if $i > j - r$, there are $r - 1$ choices for W_{b_i} so that b_i is not bump-able.

Let $\ell(r, n, A)$ denote the number words, w of length n in Ω_r such that A corresponds to a leaf in $\gamma(w)$. There are $f_r(j)$ possible ways to fill in the indices of A. For $r \leq j$,

$$\ell(r, n, A) = f_r(j)r^{\sum_{i=0}^{j-r}(a_{i+1}-a_i-1)}(r-1)^{\sum_{i=j-r+1}^{j}(a_{i+1}-a_i-1)}. \tag{8}$$

For $j < r$ this simplifies to
$$\ell(r, n, A) = j!(r-1)^{n-j}. \tag{9}$$

Thus for $j < r$ we have

$$\mathbf{P}_{\alpha,r}\left(\Big\{|W| = n\Big\} \bigcap \Big\{X \text{ is a leaf}\Big\}\right) = e^{-\alpha r}\alpha^n(r-1)^{n-j}j!/n!. \tag{10}$$

For $j < r$ the expectation of $U_j^{(r)} \mathbf{1}_{\{|W|=n\}}$ is

$$\mathbf{E}_{\alpha,r}[U_j^{(r)}\mathbf{1}_{\{|W|=n\}}] = \sum_{A \in \mathcal{A}} e^{-\alpha r}\alpha^n(r-1)^{n-j}j!/n! = e^{-\alpha r}\alpha^n(r-1)^{n-j}/(n-j)!. \quad (11)$$

Summing over $n \geq j$ gives

$$\mathbf{E}_{\alpha,r}[U_j^{(r)}] = e^{-\alpha r}\alpha^j \sum_{n \geq j}(\alpha(r-1))^{n-j}/(n-j)! = e^{-\alpha}\alpha^j. \quad (12)$$

Proof of Theorem 2:

From (12), $\mathbf{E}_{\alpha,r}[U_j^{(r)}] = e^{-\alpha}\alpha^j$ for $j < r$ and $\mathbf{E}_{\alpha,r}[U^{(r)}] = \sum_{j=0}^{r-1} e^{-\alpha}\alpha^j$. Then $\lim_{r \to \infty} U^{(r)} = U$ and by Monotone Convergence Theorem

$$\mathbf{E}_\alpha[U] = \lim_{r \to \infty}\mathbf{E}_{\alpha,r}[U^{(r)}] = \lim_{r \to \infty}e^{-\alpha}\frac{1-\alpha^r}{1-\alpha} = \frac{e^{-\alpha}}{1-\alpha}. \quad (13)$$

$\square$

6   Expectation of D^2

For $a,b,c,m \geq 0$ let $n = a+b+c+m$. Let $\mathcal{B}(a,b,c,m)$ be the set of all ordered pairs of subsets of $[n]$, (A,B), such that $|A \setminus B| = a$, $|B \setminus A| = b$, and $|A \cap B| = c$ and let $\mathcal{B}(a,b,c) = \bigcup_m \mathcal{B}(a,b,c,m)$. We denote the set of distinct subwords u on the indices $A \cup B$ such that and both u_A and u_B are complete by $\chi_r(A,B)$. The size of $\chi_r(A,B)$ is denoted by $x_r(A,B)$ and only depends on the relative order of A and B. Suppose $(A,B) \in \mathcal{B}(a,b,c)$. For both subwords to be complete, each index $a_i \in A \setminus B$ must have letters strictly less than $\min(a+c-i,r)$, each index $b_j \in B \setminus A$ must have letters strictly less than $\min(b+c-j,r)$, and each index $a_i = b_j \in A \cap B$ must have letters strictly less than $\min(a+c-i,b+c-j,r)$. Thus

$$x_r(A,B) = \frac{f_r(a+c)f_r(b+c)}{\prod_{a_i=b_j}\min(r,\max(a+c-i,b+c-j))}. \quad (14)$$

The following lemma provides uniform bounds of $x_r(A,B)$ for all $(A,B) \in \mathcal{B}(a,b,c)$.

Lemma 7. *Fix a,b,c and $r \geq 0$. For $(A,B) \in \mathcal{B}(a,b,c)$, if $a \leq b$, then*

$$f_r(a+c)f_r(b) \leq x_r(A,B) \leq (a+c)!(b+c)!/c!$$

Otherwise if $a > b$, then

$$f_r(b+c)f_r(a) \leq x_r(A,B) \leq (a+c)!(b+c)!/c!.$$

Proof:

For a fixed a,b,c and r, $x_r(A,B)$ will reach its minimum value over $\mathcal{B}(a,b,c)$ when the product in the denominator is maximized in the right hand side of (14). The denominator of $x_r(A,B)$ is maximized

when every index in $A \cap B$ is less than every index in $A \cup B \setminus A \cap B$ so $A \cap B = \{a_1 = b_1, \cdots, a_c = b_c\}$. In this case for $a \leq b$ the denominator of the right hand side of (14) is given by

$$\prod_{i=1}^{c} \min(r, b + i) = f_r(b + c)/f_r(b)$$

and

$$x_r(A, B) = f_r(a + c)f_r(b).$$

Otherwise for $a > b$

$$x_r(A, B) = f_r(b + c)f_r(a).$$

For the other direction $x_r(A, B)$ is maximized when the denominator in the right-hand side of (14) is minimized. This occurs when every index in $A \cap B$ is greater than every index in $A \cup B \setminus A \cap B$. In this case,

$$x_r(A, B) = \frac{f_r(a + c)f_r(b + c)}{f_r(c)} \leq \frac{(a + c)!(b + c)!}{c!}. \tag{15}$$

$\square$

These bounds on $x_r(A, B)$ will give us bounds on $\mathbf{E}_\alpha[D^2]$. Let $V_r = 1 + \sum_{j=1}^{\infty} D_j^{(r)}$. For a fixed set of indices $A \in \mathbb{Z}_+$ let $\mathbf{1}_A(W)$ denote the indicator function that is 1 if W_A is complete and 0 if W_A is not complete or A is not a subset of indices of W. Then

$$V_r = \sum_{A \subset \mathbb{Z}_+} \mathbf{1}_A(W)$$

with $\lim_{r \to \infty} V_r = D$. We also have

$$V_r^2 = \sum_{(A,B) \subset \mathbb{Z}_+^2} \mathbf{1}_A(W)\mathbf{1}_B(W)$$

$$= \sum_{a,b,c} \sum_{\mathcal{B}(a,b,c)} \mathbf{1}_A(W)\mathbf{1}_B(W)$$

$$= \sum_{a,b,c,m} \sum_{\mathcal{B}(a,b,c,m)} \mathbf{1}_A(W)\mathbf{1}_B(W)\mathbf{1}_{|W|=a+b+c+m}.$$

For a fixed pair $(A, B) \in \mathcal{B}(a, b, c, m)$, using Lemma 5 we have

$$\mathbf{E}_{\alpha,r}[\mathbf{1}_A(W)\mathbf{1}_B(W)\mathbf{1}_{\{|W|=a+b+c+m\}}]$$

$$= \sum_{u \in \chi_r(A,B)} \mathbf{P}_{\alpha,r}\left(\{W_{A \cup B} = u\} \cap \{|W| = a + b + c + m\}\right)$$

$$= \frac{1}{(a + b + c + m)!}e^{-\alpha r}\alpha^{a+b+c+m}r^m x_r(A, B). \tag{16}$$

The value of $x_r(A, B)$ depends on (A, B) but the upper and lower bounds from Lemma 7 only depend on $a, b,$ and c. Thus we have bounds of (16) that are uniform for all $(A, B) \in \mathcal{B}(a, b, c, m)$. For each m the size of $\mathcal{B}(a, b, c, m)$ is $\binom{a+b+c+m}{a,b,c,m} = \frac{(a+b+c+m)!}{a!b!c!m!}$. Thus

$$\sum_{\mathcal{B}(a,b,c,m)} \mathbf{E}_{\alpha,r}\left[\mathbf{1}_A(W)\mathbf{1}_B(W)\mathbf{1}_{|W|=a+b+c+m}\right]$$

$$\geq \frac{\alpha^{a+b+c}}{a!b!c!} f_r(\max(a,b) + c) f_r(\min(a,b)) \frac{1}{m!}(\alpha r)^m e^{-\alpha r} \quad (17)$$

Summing over $m \geq 0$ in (17) gives the lower bound

$$\sum_{\mathcal{B}(a,b,c)} \mathbf{E}_{\alpha,r}\left[\mathbf{1}_A(W)\mathbf{1}_B(W)\right] \geq \frac{\alpha^{a+b+c}}{a!b!c!} f_r(\min(a,b) + c) f_r(\max(a,b)). \quad (18)$$

Similarly for the upper bound we have

$$\sum_{\mathcal{B}(a,b,c)} \mathbf{E}_{\alpha,r}\left[\mathbf{1}_A(W)\mathbf{1}_B(W)\right] \leq \alpha^{a+b+c}\binom{a+c}{c}\binom{b+c}{c}. \quad (19)$$

Proof of Theorem 3:

In this section we make repeated use of the identity

$$\sum_{n \geq 0} \binom{n+k}{n} x^n = \frac{1}{(1-x)^{k+1}}.$$

See Wilf (2006) for a variety of similar identities.

By Fatou's Lemma $\lim_{r \to \infty} \mathbf{E}_{\alpha,r}[V_r^2] \leq \mathbf{E}_\alpha[\lim_{r \to \infty} V_r^2] = \mathbf{E}_\alpha[D^2]$ so

$$\lim_{r \to \infty} \sum_{a < b, c} \frac{\alpha^{a+b+c}}{a!b!c!} f_r(\min(a,b) + c) f_r(\max(a,b)) \leq \sum_{0 \leq a < b, 0 \leq c} \binom{a+c}{a}\alpha^{a+b+c} \quad (20)$$

$$\leq \mathbf{E}_\alpha[\lim_{r \to \infty} V_r^2]$$

$$= \mathbf{E}_\alpha[D^2].$$

The right hand side of (20) can be simplified further. Suppose $1/2 < \alpha < 1$. Then

$$\sum_{0 \leq a < b, 0 \leq c} \binom{a+c}{c}\alpha^{a+b+c} = \sum_{0 \leq a < b} \frac{\alpha^b}{1-\alpha}\left(\frac{\alpha}{(1-\alpha)}\right)^a \quad (21)$$

$$= \frac{1}{2\alpha - 1}\sum_{b > 0} \alpha^b\left(\left(\frac{\alpha}{1-\alpha}\right)^b - 1\right) \quad (22)$$

$$= \frac{1}{2\alpha - 1}\sum_{b > 0}\left(\frac{\alpha^2}{1-\alpha}\right)^b - \alpha^b. \quad (23)$$

There is an issue when $\alpha = 1/2$ in (22) and (23). But in this case $\frac{\alpha}{1-\alpha} = 1$ in (21), so (22) becomes $\sum_{b \geq 0} \frac{b\alpha^b}{1-\alpha}$, which is finite. Otherwise (23) diverges precisely when $\alpha^2/(1-\alpha) \geq 1$ which occurs if $(\sqrt{5} - 1)/2 \leq \alpha < 1$. For the other direction we have

$$
\begin{aligned}
\mathbf{E}_\alpha[D^2] &= \mathbf{E}_\alpha[\lim_{r \to \infty} V_r^2] \\
&\leq \sum_{a,b,c \geq 0} \binom{a+c}{c}\binom{b+c}{c}\alpha^{a+b+c} \\
&= \sum_{b,c \geq 0} \binom{b+c}{c}\frac{\alpha^{b+c}}{(1-\alpha)^{c+1}} \\
&= \frac{1}{(1-\alpha)^2} \sum_{c \geq 0} \left(\frac{\alpha}{(1-\alpha)^2}\right)^c
\end{aligned}
\tag{24}
$$

The last line (24) converges when $\alpha/(1-\alpha)^2 < 1$, which occurs when $0 < \alpha < (3 - \sqrt{5})/2$. $\qquad\square$

As α increases from $(3 - \sqrt{5})/2$ to $(\sqrt{5} - 1)/2$ a phase transition occurs where $\mathbf{E}_\alpha[D^2]$ becomes infinite. With a more precise analysis of the size of $x_r(A, B)$ that depends more closely on the relative order of A and B, one might be able to obtain the exact location where this phase transition occurs.

Acknowledgements

References

D. Aldous and J. M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004. doi: 10.1007/978-3-662-09444-0_1. URL http://dx.doi.org/10.1007/978-3-662-09444-0_1.

I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp. (electronic), 2001. ISSN 1083-6489. doi: 10.1214/EJP.v6-96. URL http://dx.doi.org/10.1214/EJP.v6-96.

P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. ISBN 0-471-19745-9. doi: 10.1002/9780470316962. URL http://dx.doi.org/10.1002/9780470316962. A Wiley-Interscience Publication.

T. Johnson, A. Schilling, and E. Slivken. Local limit of the fixed point forest. *Electron. J. Probab.*, 22: Paper No. 18, 26, 2017. ISSN 1083-6489. doi: 10.1214/17-EJP36. URL https://doi.org/10.1214/17-EJP36.

J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1993. ISBN 0-19-853693-3. Oxford Science Publications.

D. E. Knuth. *The Art of Computer Programming, Volume 3: (2Nd Ed.) Sorting and Searching*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 1998. ISBN 0-201-89685-0.

G. McKinley. A problem in card shuffling, UC Davis Undergraduate Thesis, 2015. `https://www.math.ucdavis.edu/files/1114/3950/6599/McKinley_UG_Thesis_SP15.pdf`.

J. Neveu. Arbres et processus de Galton-Watson. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(2):199–207, 1986. ISSN 0246-0203. URL `http://www.numdam.org/item?id=AIHPB_1986__22_2_199_0`.

H. W. Watson and F. Galton. On the probability of the extinction of families. *J. Anthr. Inst. Great Britain*, 4:138–144, 1875.

H. S. Wilf. *generatingfunctionology*. A K Peters, Ltd., Wellesley, MA, third edition, 2006. ISBN 978-1-56881-279-3; 1-56881-279-5.