LECTURE NOTES

CRM-PIMS PROBABILITY SUMMER SCHOOL 2021

“BEAUTY WAS NOT SIMPLY SOMETHING TO BEHOLD; IT WAS SOMETHING ONE COULD DO.”
TONI MORRISON, *THE BLUEST EYE*

ONLY DEAD MATHEMATICS CAN BE TAUGHT WHERE THE ATTITUDE OF COMPETITION PREVAILS: LIVING MATHEMATICS MUST ALWAYS BE A COMMUNAL POSSESSION.
MARY EVEREST BOOLE

“THE IDEA HOVERED AND SHIMMERED DELICATELY, LIKE A SOAP BUBBLE, AND SHE DARED NOT EVEN LOOK AT IT DIRECTLY IN CASE IT BURST. BUT SHE WAS FAMILIAR WITH THE WAY OF IDEAS, AND SHE LET IT SHIMMER, LOOKING AWAY, THINKING ABOUT SOMETHING ELSE.”
PHILIP PULLMAN, *THE GOLDEN COMPASS*

LOUIGI ADDARIO-BERRY

LECTURE NOTES: 2021 CRM-PIMS PROBABILITY SUMMER SCHOOL

CENTRE DE RECHERCHES MATHÉMATIQUES

SUPPORTED BY CENTRE DE RECHERCHES MATHÉMATIQUES

CRM.MATH.CA

Latest version: August 16, 2023

Contents

List of Figures

List of Tables

Thanks to the CRM, PIMS, and MSRI for
their support of this summer school.

Introduction

Lecture notes to accompany my lectures at the 2021 CRM-PIMS probability summer school.

1

Deterministic and random combinatorial trees

1.1 Preliminaries: graph notation

An *undirected graph* is an ordered pair $G = (V, E)$ where V is the set of vertices and $E \subset \binom{V}{2}$ is the set of edges. Here we write $\binom{V}{2}$ for the set of unordered pairs of elements of V; more generally, for a set S and a positive integer k we write

It's also often convenient to write $\mathrm{v}(G)$ and $\mathrm{e}(G)$ for the vertex and edge sets of graph G, and we do this as well.

$$\binom{S}{k} = \{U \subset S : |U| = k\}$$

for the set of all subsets of S of size k. We often write uv for an edge of a graph G, rather than the more correct but more cumbersome notation $\{u, v\}$.

A *path* in a graph $G = (V, E)$ is a finite sequence $u_0, u_1, \ldots, u_k$ of distinct vertices of G such that for all $0 \le i < k$, $u_i u_{i+1}$ is an edge of G. Two vertices u, v are in the same *connected component* of G if there is a path connecting u and v.

Exercise 1.1.1. *Write $u \overset{G}{\leftrightarrow} v$ if there is a path between u and v in G. Show that $\overset{G}{\leftrightarrow}$ is an equivalence relation.*

We say G is connected if it has a single connected component; otherwise it is *disconnected*.

We write $\mathscr{G}_n$ for the set of graphs $G = (V, E)$ with vertex set $V = [n] := \{1, 2, \ldots, n\}$. For a finite set $\mathscr{S}$ we write $\mathbf{X} \in_u \mathscr{S}$ to mean that X is a uniformly random element of $\mathscr{S}$.

Exercise 1.1.2. *Fix a random graph $\mathbf{G} \in_u \mathscr{G}_n$. Show that*

$$\mathbf{P}\{\mathbf{G} \textit{ is not connected}\} = (1 + o(1))n \cdot 2^{-(n-1)}.$$

A *cycle* in a graph $G = (V, E)$ is a finite sequence $u_0, u_1, \ldots, u_{k+1}$ of vertices such that $u_0, \ldots, u_k$ is a path (so in particular $u_0, u_1, \ldots, u_k$ are all distinct), and $u_{k+1} = u_0$. A graph is a *tree* if it contains no cycles.

Exercise 1.1.3. *Let $G = (V, E)$ be a finite connected graph. Show that G is a tree if and only if $|E| = |V| - 1$.*

Exercise 1.1.4.

Let $T = (V, E)$ be a graph. Prove that the following are equivalent.

1. *T is a tree.*
2. *Any two vertices of T are linked by a* unique *path in T.*
3. *T is* minimally connected, *meaning that T is connected but removing any edge of T disconnects the graph.*
4. *T is* maximally acyclic, *meaning T contains no cycle but for any two non-adjacent vertices $x, y \in T$, the graph $(V, E \cup \{xy\})$ contains a cycle.*

Given a graph $G = (V, E)$, for $u, v \in V$ write $\mathrm{dist}(u, v) = \mathrm{dist}_G(u, v)$ for the graph distance between u and v in G; that is, $\mathrm{dist}(u, v)$ is the smallest number of edges in a path connecting u and v. (If u, v are in different connected components of G then $\mathrm{dist}(u, v) = \infty$. For $u \in V$ and $S \subset V$, we also write

$$\mathrm{dist}(u, S) = \mathrm{dist}_G(u, S) = \inf(\mathrm{dist}(u, v) : v \in S)$$

for the distance from u to S in G.

1.2 *Random combinatorial trees and forests*

In this section, and in much of the rest of the paper, we consider trees that are *rooted*: formally a rooted tree is a triple $\mathrm{t} = (V, E, \rho)$ where (V, E) is a tree and $\rho \in V$.[1] We write

$$\mathscr{T}_n^{\mathrm{unrooted}} = \{\mathrm{t} = (V, E) : \mathrm{t} \text{ is a tree}, V = [n]\}$$

for the set of trees with vertex set $[n]$, and

$$\mathscr{T}_n = \{(V, E, \rho) : (V, E) \in \mathscr{T}_n^{\mathrm{unrooted}} \text{ and } \rho \in V\}.$$

Any tree in $\mathscr{T}_n$ can be rooted in n different ways, which implies that $|\mathscr{T}_n| = n|\mathscr{T}_n^{\mathrm{unrooted}}|$. Cayley's formula (which to the best of knowledge of contemporary academics was first established by Borchardt [2]) provides a very simple formula for $|\mathscr{T}_n^{\mathrm{unrooted}}|$, or equivalently for $|\mathscr{T}_n|$.

Theorem 1.2.1 (Cayley's formula)**.** $|\mathscr{T}_n| = n^{n-1}$.

There are several proofs of Cayley's formula in the literature: one using so called *Prüfer codes*; one discovered by Joyal which considers doubly-rooted trees (called *vertebrates*); one discovered by Pitman which analyzes a coalescent process for building random trees; and the one we present, which is new and is due to Addario-Berry and Donderwinkel.

[1] It's sometimes useful to write $\rho(\mathrm{t})$ rather than ρ, and we will do this when we wish, without further comment.

[2]

Proof of Cayley's formula. The right-hand side naturally counts sequences of integers $\mathrm{v} = (v_1, \ldots, v_{n-1}) \in [n]^{n-1}$. We prove the theorem by showing how to associate a rooted tree $\mathrm{t} \in \mathscr{T}_n$ to each such sequence.

Say that v_i is a *repeated entry* if there is $1 \leq j < i$ with $v_i = v_j$. List the repeated entries in increasing order of index as $v_{i(1)}, \ldots, v_{i(r)}$; so $r = r(\mathrm{v})$ is the number of repeated entries v. (If $v_1, \ldots, v_{n-1}$ are all distinct then $r = 0$.) Set $i(r+1) = n$.

Since v has length $n-1$, there are exactly $r+1$ integers from $[n]$ which do not appear in v; list them in increasing order as $\ell_1, \ldots, \ell_{r+1}$.

Now form a tree $\mathrm{t} = \mathrm{t}(\mathrm{v})$ with vertices $[n]$, edges

$$\left\{v_i v_{i+1} : 1 \leq i \leq n-1, i+1 \notin \{i(1), \ldots, i(r+1)\}\right\} \cup \left\{v_{i(j)-1}\ell_j, 1 \leq j \leq r+1\right\},$$

and root $\rho = v_1$.

Visually, the tree is constructed as follows. Think of v as a path graph, and remove all edges $v_i v_{i+1}$ where v_{i+1} is a repeated entry; if v_{i+1} is the j'th repeated entry then instead attach v_i to ℓ_j. This breaks the path up into pieces. To connect these pieces, glue each repeated entry to the location where it first appears in v.

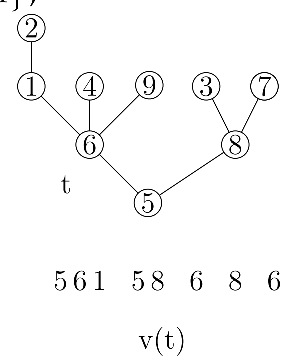


Figure 1.1: A tree t and the sequence v(t).

The graph $\mathrm{t}(\mathrm{v})$ has n vertices and $n-1$ edges, so to prove $\mathrm{t}(\mathrm{v})$ is a tree and therefore an element of $\mathscr{T}_n$, by Exercise 1.1.3 it suffices to show that it is connected. To see this, fix $i \in [n-1]$ with $i > 1$. If v_i is not a repeated entry then $v_{i-1}v_i$ is an edge of $\mathrm{t}(\mathrm{v})$. If v_i is a repeated entry then let j be minimal so that $v_j = v_i$; then $v_{j-1}v_i$ is an edge of $\mathrm{t}(\mathrm{v})$. In either case, there is an edge from v_i to v_k for some $k < i$, which implies by induction that for all $i \in [n-1]$ there is a path from v_i to v_1. Since ℓ_j is attached by an edge to $v_{i(j)}$, it also follows that there is a path from ℓ_j to v_1 for all $1 \leq j \leq r+1$. Thus all vertices are in the same connected component as v_1 and so $\mathrm{t}(\mathrm{v})$ is connected.

To show that the above construction is a bijection, we describe its inverse. Given a rooted tree $\mathrm{t} = (V, E, \rho) \in \mathscr{T}_n$, define a sequence $\mathrm{v}(\mathrm{t}) = (v_1, \ldots, v_{n-1})$ as follows. List the leaves of t in increasing order of label as $\ell_1, \ldots, \ell_{r+1}$. Let t_0 be the one-vertex rooted tree containing only the root ρ of t. For $1 \leq i \leq r+1$, let $\mathrm{p}_i = (p_i(1), \ldots, p_i(h_i))$ be the vertices on the path from t_{i-1} to ℓ_i, so $p_i(1)$ is a vertex of t_{i-1} and $p_i(h_i) = \ell_i$. Let t_i be the forest obtained from t_{i-1} by attaching the path p_i. Then let

$$\mathrm{v} = (v_1, \ldots, v_{n-1}) = (p_1(1), \ldots, p_1(h_1 - 1), p_2(1), \ldots, p_2(h_2 - 1), \ldots, p_{r+1}(1), \ldots, p_{r+1}(h_{r+1} - 1)).$$

Note that $\rho = v_1$. Also, $h_i - 1$ is the number of edges on the path from t_{i-1} to ℓ_i, so v has

$$\sum_{i=1}^{r+1} h_i - 1 = |E| = n - 1$$

entries. Moreover, $v_i v_{i+1}$ is an edge of t with $v_i = \mathrm{par}(v_{i+1})$ if and only if

$$v_i v_{i+1} \notin \{p_j(h_j - 1)p_{j+1}(1), 1 \le j \le r\},$$

and the repeated entries of v are precisely $p_2(1), \ldots, p_{r+1}(1)$, so $v_i v_{i+1}$ is an edge of t if and only if v_{i+1} is not a repeated entry of v. The remaining edges of t are the elements of the set

$$\{p_j(h_j - 1)\ell_j, 1 \le j \le r + 1\};$$

they precisely join the predecessors of repeated entries of v to the leaves $\ell_1, \ldots, \ell_{r+1}$, in increasing order. Thus $\mathrm{v}(\mathrm{t}(\mathrm{v})) = \mathrm{v}$; these two constructions are indeed inverses. Since we have exhibited a bijection between $\mathscr{T}_n$ and $[n]^{n-1}$, it follows that

$$|\mathscr{T}_n| = |[n]^{n-1}| = n^{n-1}. \qquad \square$$

Note that here it's important to require that $v_i = \mathrm{par}(v_{i+1})$, i.e., to think of $v_i v_{i+1}$ as a directed edge directed away from the root. To see this, consider the sequence $\mathrm{v} = (1, 2, 3, 2, 5, 6)$; for this sequence $p_1(h_1 - 1) = p_1(3) = 3$ and $p_2(1) = 2$, so $v_3 v_4 = p_1(3)p_2(1) = 32$ is an edge of the corresponding tree.

The above bijection has nice consequences for random trees. In what follows, for a tree $\mathrm{t} = (V, E, \rho)$ and vertices $u, v \in V$, write $[u, v] = [u, v]_{\mathrm{t}}$ for the unique path from u to v in t.[3] For $v \in V$, write $|v|$ for the graph distance from ρ to v, which equals the number of edges of the path $[\rho, v]$.

[3] $[u, v] = [u, v]_{\mathrm{t}}$: the path between vertices u and v in tree t

Proposition 1.2.2. *Let* $\mathrm{T} \in_u \mathscr{T}_n$ *and let* L *be a uniformly random leaf of* T. *Also, let* $(V_i, i \ge 1)$ *be a sequence of independent uniformly random elements of* $[n]$ *and let* $I = \min(i \ge 1 : V_i \in \{V_1, \ldots, V_{i-1}\})$ *be the index of the first repeated element of the sequence. Then* $|L| + 1 \overset{\mathrm{d}}{=} \min(I, n)$.

Proof. Write $\mathrm{V} = (V_1, \ldots, V_{n-1}) \in_u [n]^{n-1}$. Then $\mathrm{T} = \mathrm{T}(\mathrm{V}) \in_u \mathscr{T}_n$. Moreover, recalling that the repeated entries of $(V_1, \ldots, V_{n-1})$ are $i(1), \ldots, i(r)$ and that $i(r + 1) = n$, we have $\min(I, n) = i(1)$. The first leaf $\ell_1(\mathrm{T})$ is a child of $V_{i(1)-1}$, so

$$|\ell_1(\mathrm{T})| = |V_{i(1)-1}| + 1 = i(1) - 1 = \min(I, n) - 1.$$

But since T is a uniformly random tree, randomly permuting its leaf labels does not change its distribution, so $|\ell_1(\mathrm{T})|$ has the same distribution as $|L|$ for L a uniformly random leaf of T. $\square$

Exercise 1.2.1. *Let* $\mathrm{T} \in_u \mathscr{T}_n$ *and let* L *be a uniformly random leaf of* T. *Show that for all* $1 \le k < n - 1$,

$$\mathbf{P}\{|L| > k\} = \prod_{i=1}^{k} \left(1 - \frac{i}{n}\right)$$

Exercise 1.2.2. *Say that rooted tree t is* binary *if every non-leaf node has exactly two children. Say that a sequence* $\mathrm{v} = (v_1, \ldots, v_{n-1}) \in [n]^{n-1}$ *is binary if every integer* $u \in [n]$ *which appears in* v *appears exactly twice.*

(a) *Fix an odd positive integer $n = 2m + 1$. Show that the set $\mathscr{B}_n$ of rooted binary trees with vertex set $[n]$ is in bijective correspondence with the set of binary sequences*

$$\{\mathrm{v} \in [n]^{n-1} : \mathrm{v} \textit{ is binary}\} .$$

(b) *Let $\mathrm{T} \in_u \mathscr{B}_n$ and let L be a uniformly random leaf of T. Show that for all $1 \le k < m$,*

$$\mathbf{P}\{|L| > k\} = \prod_{i=1}^{k} \left(1 - \frac{i}{2m-i}\right)$$

Exercise 1.2.3. (a) *Let $T_n \in_u \mathscr{T}_n$ and let C_n be the number of children of vertex 1 in T_n. Show that C_n converges in distribution to a* Poisson(1) *random variable.*

(b) *Write π_n for the* empirical child distribution *of T_n: that is,*

$$\pi_n = \frac{1}{n} \sum_{c=0}^{n-1} \delta_c \cdot |\{v \in [n] : v \textit{ has } c \textit{ children in } T_n\}| .$$

Show that π_n converges in probability to the Poisson(1) *distribution with respect to the Prokhorov distance between probability measures. (Suggestion: it suffices to show that for any fixed $k \in \mathbb{N}$, $\pi_n(\{k\}) \to \mathbf{P}\{\mathrm{Poisson}(1) = k\}$ in probability.)*

Exercise 1.2.4. *I would like to write an exercise using the fact that the smallest-labeled and second-smallest-labeled leaves have the same height (in distribution) to formulate a distributional identity involving the first and second repeated elements in a sequence $\mathrm{V} = (V_i, i \ge 1)$ be independent uniformly random elements of $[n]$. But I'm not sure how, since the "second-smallest-labeled leaf" may not exist - the tree may have only one leaf.*

Benjamin Bonnefont suggests: conditionally on the presence of at least 2 leaves at least, one has $I_1 \stackrel{\mathrm{d}}{=} I_2 - I_1 + Uniform([I_1 - 1])$ where the last random variable is independent of the 1st term, conditionally on I_1.

We conclude the section by extending Cayley's formula to a formula for forests with a fixed vertex set. A *forest* is a set of rooted trees with pairwise disjoint vertex sets. Given a forest $F = \{\mathrm{t}_i, i \in I\}$, the *root set* of F is the set $\rho(F) := \{\rho(\mathrm{t}_i), i \in I\}$ of roots of its constituent trees.

Given a set $S \subset [n]$, write $\mathscr{F}_n^S$ for the set of forests F with vertex set $[n]$ and root set S.[4]

[4] Add citations for the next proposition.

Proposition 1.2.3. *For any integers $1 \le k \le n$ and any set $S \subset [n]$ with $|S| = k$,*

$$|\mathscr{F}_n^S| = kn^{n-k-1}$$

Proof. Fix any sequence $\mathrm{v} = (v_1, \ldots, v_{n-k})$ of integers in $[n]$ such that $v_1 \in S$. Say that v_i is a repeated entry if $i > 1$ and either $v_i \in S$ or there is $1 \le j < i$ such that $v_i = v_j$.

List the repeated entries of v in increasing order of index as $v_{i(1)}, \ldots, v_{i(r)}$, set $i(r+1) = n - k + 1$, and list the integers from $[n] \setminus S$ which do not appear in v as $\ell_1, \ldots, \ell_{r+1}$ in increasing order.

Form a graph $F_S = F_S(\mathrm{v})$ with vertices $[n]$, root set S, and edge set

$$\{v_i v_{i+1} : i \in [n-k], i+1 \notin \{i(1), \ldots, i(r+1)\} \cup S\} \cup \{v_{i(j)-1}\ell_j, 1 \le j \le r+1\}.$$

Essentially the same argument as in the proof of Cayley's formula shows that the connected components of F_S are trees and that there are k components of F_S, each containing exactly one vertex of S. Thus, rooting each component of F_S at its unique element of S turns F_S into an element of $\mathscr{F}_n^S$.

The inverse of this construction is also very similar to the one in the proof of Cayley's formula: given $F \in \mathscr{F}_n^S$, list the leaves of F in increasing order of label as $\ell_1, \ldots, \ell_{r+1}$. Let F_0 be the k-vertex forest containing only the root vertices S. For $1 \le i \le r+1$, let $\mathrm{p}_i = (p_i(1), \ldots, p_i(h_i))$ be the path from F_{i-1} to ℓ_i, so $p_i(1)$ is a vertex of F_{i-1} and $p_i(h_i) = \ell_i$. Let F_i be the forest obtained from F_{i-1} by attaching the path p_i. Then let

$$\mathrm{v}(F) = (v_1, \ldots, v_{n-1}) = (p_1(1), \ldots, p_1(h_1 - 1), p_2(1), \ldots, p_1(h_2 - 1), \ldots, p_{r+1}(1), \ldots, p_{r+1}(h_{r+1} - 1)) \, .$$

It is straightforward to see that $\mathrm{v}(F_S(\mathrm{v})) = \mathrm{v}$, so the construction is bijective. It follows that

$$|\mathscr{F}_n^S| = |\{\mathrm{v} = (v_1, \ldots, v_{n-k}) \in [n]^{n-k} : v_1 \in S\}| = kn^{n-k-1} \, . \qquad \square$$

1.3 *Global structure: The line-breaking construction*

In this section, the construction of the sequence of trees $\mathrm{t}_0, \ldots, \mathrm{t}_{r+1}$ arising in the bijection from Theorem 1.2.1 is important, and it's useful to give ourselves a bit more notation.

For a tree $\mathrm{t} = (V, E)$ and a set $S \subset V$, write $\mathrm{t}\langle S \rangle$ for the smallest subtree of t containing the root and all elements of S; equivalently, this tree is the union $\bigcup_{u \in S} [\rho, u]$. [5] Then the sequence of trees from Theorem 1.2.1 can be expressed as $\mathrm{t}_i = \mathrm{t}\langle \{\ell_1, \ldots, \ell_i\} \rangle$. Also, for $v \in \mathrm{t}$ write $\alpha(v, S)$ for the node w of S which minimizes $\mathrm{dist}(v, w)$; this node is unique since t is a tree.

In what follows, we use the "falling factorial" notation $(m)_i = m(m-1) \cdot \ldots \cdot (m - i + 1)$.

[5] $\mathrm{t}\langle S \rangle$: the subtree of t spanned by the root and the vertices in S

Falling factorial notation.

Proposition 1.3.1. *Let $\mathrm{T}_n \in_u \mathscr{T}_n$, and list the leaves of T_n in increasing order of label as $\ell_1, \ldots, \ell_{r+1}$ and set $\ell_m = \rho(\mathrm{T}_n)$ for $m > r + 1$. For $i \ge 1$*

write $D_{n,i}$ for the graph distance from ℓ_i to $\mathrm{T}_n\langle \ell_1,\ldots,\ell_{i-1}\rangle$. Then for any positive integers $k \in [n-1]$ and $g_1,\ldots,g_k$,[6]

$$\mathbf{P}\left\{D_{n,i} = g_i, 1 \le i \le k\right\} = \frac{(n)_{g_1+\ldots+g_k-(k-1)}}{n^{g_1+\ldots+g_k+1}} \cdot \prod_{j=1}^{k}(g_1+\ldots+g_j-(j-1))\,.$$

[6] Implicit in the next probability is the requirement that T has at least k leaves, though our convention about the value of ℓ_m for $m > r+1$ makes the formula hold for all values of k, even $k > r+1$.

Moreover, for each $1 \le i \le r+1$, the node $\alpha(\ell_i, T_n\langle \ell_1,\ldots,\ell_{i-1}\rangle)$ is a uniformly random non-leaf vertex of $T_n\langle \ell_1,\ldots,\ell_{i-1}\rangle$.

Proof. Let $\mathrm{V} = (V_1,\ldots,V_{n-1}) = \mathrm{v}(\mathrm{T})$, so that $\mathrm{V} \in_u [n]^{n-1}$ and $\mathrm{T} = \mathrm{T}(\mathrm{V}) \in_u \mathscr{T}_n$. Then in order to have $D_{n,j} = g_j$ for each $1 \le j \le k$, it must be that $i(j) = 1 + g_1 + \ldots + g_j$ for each $1 \le j \le k$. In turn, for this to occur, it must be that

$$\begin{aligned}
&V_1,\ldots,V_{g_1}\\
&V_{g_1+2},\ldots,V_{g_1+g_2}\\
&V_{g_1+g_2+2},\ldots,V_{g_1+g_2+g_3}\\
&\ldots\\
&V_{g_1+\ldots+g_{k-1}+2},\ldots,V_{g_1+\ldots+g_k}
\end{aligned}$$

are all distinct (call this event A), and that

$$\begin{aligned}
V_{g_1+1} &\in \{V_1,\ldots,V_{g_1}\}\\
V_{g_1+g_2+1} &\in \{V_1,\ldots,V_{g_1+g_2}\}\\
&\ldots\\
V_{g_1+\ldots+g_k+1} &\in \{V_1,\ldots,V_{g_1+\ldots+g_k}\}\,;
\end{aligned}$$

call this event B.

The event A is that $g_1+g_2+\ldots+g_k-(k-1)$ independent random variables, uniformly distributed on $[n]$, are all distinct, so has probability

$$\mathbf{P}\{A\} = \prod_{j=1}^{g_1+\ldots+g_k-(k-1)} \frac{n-(j-1)}{n} = \frac{(n)_{g_1+\ldots+g_k-(k-1)}}{n^{g_1+\ldots+g_k-(k-1)}}\,.$$

To work out the probability of the second event we need to split into sub-events. For $1 \le j \le k$ write B_j for the event that

$$V_{g_1+\ldots+g_j+1} \in \{V_1,\ldots,V_{g_1+\ldots+g_j}\}.$$

Given that A occurs and that $B_1,\ldots,B_{j-1}$ occur, there are $j-1$ repetitions in the set

$$\{V_1,\ldots,V_{g_1+\ldots+g_j}\},$$

and so $g_1+\ldots+g_j-(j-1)$ distinct values in that set. It follows that

$$\mathbf{P}\left\{B_j \;\middle|\; A \cap B_1 \cap \ldots \cap B_{j-1}\right\} = \frac{g_1+\ldots+g_j-(j-1)}{n}\,.$$

Since $B = \bigcap_{j=1}^k B_j$, it follows that

$$\begin{aligned}\mathbf{P}\{B \mid A\} &= \prod_{j=1}^k \mathbf{P}\{B_j \mid A \cap B_1 \cap \ldots \cap B_{j-1}\} \\ &= \prod_{j=1}^k \frac{g_1 + \ldots + g_j - (j-1)}{n}.\end{aligned}$$

Combining the formulas for $\mathbf{P}\{A\}$ and $\mathbf{P}\{B \mid A\}$ gives the identity in the proposition.

Moreover, given that A and B occur, $\alpha(\ell_i, \mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle) = V_{g_1+\ldots+g_i+1}$, and the conditioning precisely tells us that $V_{g_1+\ldots+g_i+1}$ is an element of $\{V_1, \ldots, V_{g_1+\ldots+g_i}\}$, which is the set of non-leaf vertices of the tree $\mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle$. A uniform random variable conditioned to lie in a particular set is uniform on that set, so $\alpha(\ell_i, \mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle)$ is uniformly distributed over $\mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle$. □

Proposition 1.3.1 yields a distributional limit theorem for the vector of distances $(D_{n,i}, i \ge 1)$.

Corollary 1.3.2. *Fix $k \ge 1$ and let $\mathrm{D} = (D_1, \ldots, D_k)$ be a random vector with density*

$$f_{\mathrm{D}}(c_1, \ldots, c_k) = \exp\left(-\frac{(c_1 + \ldots + c_k)^2}{2}\right) \prod_{j=1}^k (c_1 + \ldots + c_j).$$

Then $(n^{-1/2} D_{n,1}, \ldots, n^{-1/2} D_{n,k}) \xrightarrow{\mathrm{d}} \mathrm{D}$, as $n \to \infty$.

Proof. Fix any positive integer k and any positive real values $c_1, \ldots, c_k$, and write $c = c_1 + \ldots + c_k$. Then[7]

$$\frac{(n)_{\lfloor cn^{1/2} \rfloor}}{n^{\lfloor cn^{1/2} \rfloor}} = \prod_{j=1}^{\lfloor cn^{1/2} \rfloor} \left(1 - \frac{j-1}{n}\right) = (1 + o(1)) \exp\left(-\frac{c^2}{2}\right)$$

[7] Exercise: if $k = k(n) = O(n^{1/2})$ then $\prod_{j=1}^k (1 - (j-1)/n) = (1 + o(1)) e^{-k^2/2n}$

The formula from Proposition 1.3.1 then gives that

$$\begin{aligned}&\mathbf{P}\left\{D_{n,i} = \left\lfloor c_i n^{1/2} \right\rfloor, 1 \le i \le k\right\} \\ &= (1 + o(1)) \exp\left(-\frac{c^2}{2}\right) \frac{1}{n^k} \prod_{j=1}^k \left(\left\lfloor c_1 n^{1/2} \right\rfloor + \ldots + \left\lfloor c_j n^{1/2} \right\rfloor - (j-1)\right) \\ &= (1 + o(1)) \exp\left(-\frac{(c_1 + \ldots + c_k)^2}{2}\right) \frac{1}{n^{k/2}} \prod_{j=1}^k (c_1 + \ldots + c_j) \\ &= \frac{1 + o(1)}{n^{k/2}} f_{\mathrm{D}}(c_1, \ldots, c_k).\end{aligned}$$

It follows that for any rectangle $R = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_k, b_k]$,

$$\mathbf{P}\left\{(n^{-1/2} D_{n,1}, \ldots, n^{-1/2} D_{n,k}) \in R\right\} = (1 + o(1)) \int_R f_D(c_1, \ldots, c_k),$$

which implies convergence in distribution. □

Corollary 1.3.3. *For $i \geq 1$ let $J_{n,i}$ be the index of $\alpha(\ell_i, \mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle)$ in* $\mathrm{V}(\mathrm{T})$*, so $\alpha(\ell_i, \mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle) = V_{J_{n,i}}$. Then jointly with the convergence in Corollary 1.3.2, we have*

$$(n^{-1/2} J_{n,i}, 1 \leq i \leq k) \xrightarrow{\mathrm{d}} ((D_1 + \ldots + D_{i-1})U_i, 1 \leq i \leq k)\,,$$

where $(U_i, i \geq 1)$ are independent Uniform$[0,1]$, independent of D.

Proof. The non-leaf vertices of $\mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle$ are

$$\{V_i, 1 \leq i \leq D_{n,1} + \ldots + D_{n,i-1}\} \setminus \{V_{D_{n,1}+1}, V_{D_{n,1}+D_{n,2}+1}, \ldots, V_{D_{n,1}+\ldots+D_{n,i-2}+1}\}.$$

There are thus $D_{n,1} + \ldots + D_{n,i-1} - (i-2)$ vertices in $\mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle$, and by the theorem, $J_{n,i}$ is uniformly distributed among their indices. The claimed convergence in distribution then follows from that in Corollary 1.3.2. □

The preceding corollaries can be beautifully recast using the language of Poisson point processes. Let P be a Poisson process on $[0, \infty)$ with rate $\lambda(x) = x$.[8] List the points of P in increasing order as $(P_i, i \geq 1)$. Set $P_0 = 0$ and for $i \geq 1$ let $D_i = P_i - P_{i-1}$. Then the function

$$f_{(P_1,\ldots,P_k)}(p_1, \ldots, p_k) = \mathbf{1}_{[p_1 < p_2 < \ldots p_k]} \cdot \exp\left(-\frac{p_k^2}{2}\right) \prod_{i=1}^{k} p_i\,,$$

is a joint density for $P_1, \ldots, P_k$. Considering the change of variables $D_i = P_i - P_{i-1}$, writing $c_i = p_i - p_{i-1}$, we have

$$\begin{aligned} f_{(D_1,\ldots,D_k)}(c_1, \ldots, c_k) &= f_{(P_1,\ldots,P_k)}(p_1, \ldots, p_k) \\ &= \mathbf{1}_{[p_1 < p_2 < \ldots p_k]} \cdot \exp\left(-\frac{p_k^2}{2}\right) \prod_{i=1}^{k} p_i \\ &= \mathbf{1}_{[c_1,\ldots,c_k > 0]} \exp - \left(-\frac{(c_1 + \ldots + c_k)^2}{2}\right) \prod_{i=1}^{k} (c_1 + \ldots + c_i). \end{aligned}$$

In view of this calculation, Corollary 1.3.2 states that *the branch lengths in the bijective construction of* T_n *given by the proof of Cayley's formula are asymptotically distributed like the inter-arrival times in a Poisson process on* $[0, \infty)$ *with rate* $\lambda(x) = x$.

Since Corollary 1.3.3 adds that the *attachment location* of the i'th branch is asymptotically uniform over the tree $\mathrm{T}_n\langle \ell_1, \ldots, \ell_{i-1}\rangle$, this means we can construct a tree which is the "distributional limit" of $\mathrm{T}_n\langle \ell_1, \ldots, \ell_k\rangle$ from the Poisson point process as follows. Let $(U_i, i \geq 1)$ be independent Uniform$[0,1]$ random variables, independent of the Poisson process P. Then, starting from the line segments $([P_{i-1}, P_i), 1 \leq i \leq k)$, for each $1 \leq i \leq k$ identify the left endpoint P_{i-1} of the segment $[P_{i-1}, P_i)$ with the point $P_{i-1}U_i$.

[8] A Poisson process on $\mathbb{R}^d$ with rate function $\lambda : \mathbb{R} \to [0, \infty)$ is characterized by two facts. First, for any rectangle $R \subset \mathbb{R}^d$, the number of points $N(R)$ of P falling in R is Poisson($\int_R f(x)\mathrm{d}x$)-distributed. Second, for any $k \in \mathbb{N}$ and any disjoint rectangles $R_1, \ldots, R_k$, the random variables $(N(R_i), 1 \leq i \leq k)$ are mutually independent.

It is a bit ambiguous what we mean by the "distributional limit" of $\mathrm{T}_n\langle \ell_1, \ldots, \ell_k\rangle$. One way to give sense to this is to simply consider matrices of pairwise distances between leaves; this perspective is developed in the next exercise.

Exercise 1.3.1. *Fix $k \ge 1$. For $n \ge 1$ let $\mathrm{T}_n \in_u \mathscr{T}_n$, and for $i, j \ge 1$ let $d_n(\ell_i, \ell_j)$ be the distance between ℓ_i and ℓ_j in T_n.*

(a) Show that the matrix $(n^{-1/2} d_n(\ell_i, \ell_j), 1 \le i, j \le k)$ converges in distribution, and describe the limit in terms of the Poisson process P and the uniform random variables $(U_i, 1 \le i \le k)$ given above.

(b) Show that the same convergence in distribution holds if $\ell_1, \ldots, \ell_k$ are replaced by a sequence $L_1, \ldots, L_k$ of independent uniformly random samples from the leaf set of T_n. (Suggestion: since randomly permuting the leaf labels does not change the distribution of T_n, the main step is to show that with high probability $L_1, \ldots, L_k$ are all distinct.)

For the next exercises, we need the notion of the *shape* of a tree. First, by a leaf-labeled tree we mean a rooted tree t whose leaves are labeled by $\{1, \ldots, k\}$, where k is the total number of leaves, and internal nodes are unlabeled.[9]

[9] In a leaf-labeled tree, to distinguish internal vertices we may canonically assign to each internal vertex the set of labels of its descendant leaves.

Given a labeled rooted tree t with leaves $\ell_1, \ldots, \ell_k$, the shape of t is the leaf-labeled rooted tree $\mathrm{shape}(t)$ obtained from t as follows.

1. replace each maximal path containing no internal branch points by an edge.
2. relabel leaves $\ell_1, \ldots, \ell_k$ as $1, \ldots, k$.
3. remove the labels of all non-leaf vertices.

For each edge e of $\mathrm{shape}(t)$, we define the *length* $\mathrm{len}(e) = \mathrm{len}(e; t)$ to be the number of edges of the path in t which gives rise to edge e in $\mathrm{shape}(t)$.

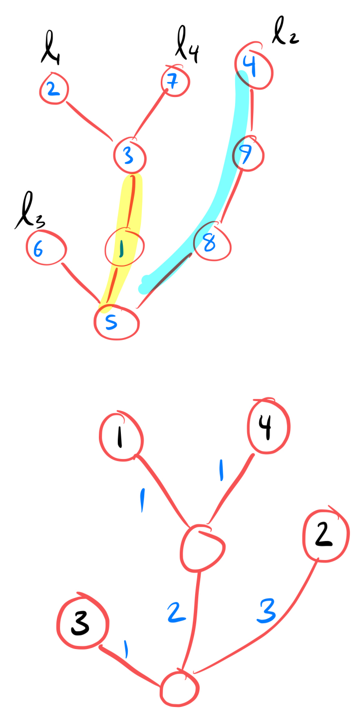


Figure 1.2: Top: a tree t. Bottom: $\mathrm{shape}(t)$ with edges labeled by their lengths.

Exercise 1.3.2. *A leaf-labeled tree is* binary *if the root has exactly one child and all other non-leaf vertices have exactly two children. Show that the number of binary leaf-labeled trees with k leaves is*

$$(2k-3)!! := (2k-3) \cdot (2k-5) \cdot \ldots \cdot 3 \cdot 1 = \frac{(2k-2)!}{(k-1)! 2^{k-1}}.$$

Let

$$\Delta_n = \left\{ \mathbf{x} = (x_1, x_2, \ldots, x_n) : \sum_{j=1}^{n} x_j = 1,\ x_j > 0, 1 \le j \le n \right\}$$

be the $(n-1)$-dimensional simplex. For $(\alpha_1, \ldots, \alpha_n) \in \Delta_n$, the Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_n)$ distribution on Δ_n has density

$$\frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\ldots\Gamma(\alpha_n)} \cdot \prod_{j=1}^{n} x_j^{\alpha_n - 1},$$

with respect to $(n-1)$-dimensional Lebesgue measure Λ_{n-1}, where

$$\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} e^{-s} \mathrm{d}s \,.$$

Exercise 1.3.3. *(a) Show that if $G_1, \ldots, G_m$ are independent and G_i is* $\mathrm{Gamma}(\alpha_i, 1)$*-distributed for $1 \le i \le m$, and $G = E_1 + \ldots + E_m$, then* $(\frac{G_i}{G}, 1 \le i \le m)$ *is* $\mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_m)$*-distributed and is independent of G.*

(b) Show that if $(X_1, \ldots, X_m)$ is $\mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_m)$*-distributed then* $(X_1, \ldots, X_{m-2}, X_{m-1} + X_m)$ *is* $\mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_{m-2}, \alpha_{m-1} + \alpha_m)$*-distributed.*

(c) Let $(Y_1, \ldots, Y_m)$ be a $\mathrm{Dirichlet}(1, 1, \ldots, 1)$ *vector, and let I be the index of a size-biased pick from $Y_1, \ldots, Y_m$, so*

$$\mathbf{P}\left\{ I = i \mid Y_1, \ldots, Y_m \right\} = \frac{Y_i}{Y_1 + \ldots + Y_m} \,.$$

Relabel $(Y_i, i \in [m] \setminus \{I\})$ as $(Y_1^*, \ldots, Y_{m-1}^*)$. Then $(Y_1^*, \ldots, Y_{m-1}^*, Y_I) \overset{\mathrm{d}}{=}$ $\mathrm{Dirichlet}(1, \ldots, 1, 2)$.

(d) Let U be $\mathrm{Uniform}[0,1]$ *be independent of $Y_1, \ldots, Y_m$ and of I. Show that*

$$(Y_1, \ldots, Y_{m-1}, UY_I, (1-U)Y_I)$$

is $\mathrm{Dirichlet}(1, 1, \ldots, 1)$*-distributed.*

Write $\mathscr{T}^{(k)}$ for the set of pairs $(t, (x(e), e \in \mathrm{e}(T)))$, where t is a binary leaf-labeled tree with k leaves and $(x(e), e \in \mathrm{e}(t)) \in [0, \infty)^{\mathrm{e}(t)}$ assigns non-negative lengths to the $2k-1$ edges of t.

Exercise 1.3.4. *(a) Show that for any $k \ge 1$, $(\frac{P_k}{P_{k+1}}, \frac{P_{k+1} - P_k}{P_{k+1}})$ is* $\mathrm{Dirichlet}(2k, 1)$*-distributed and is independent of P_{k+1}.*

(b) Let $T_n \in_u \mathscr{T}_n$. Show that for any $k \ge 1$, writing $T_{n,k} = T_n\langle \ell_1, \ldots, \ell_k \rangle$ for the subtree of T_n spanned by the root and the k smallest-labeled vertices, then

$$(\mathrm{shape}(T_{n,k}), (\mathrm{len}(e), e \in \mathrm{shape}(T_{n,k}))) \overset{\mathrm{d}}{\to} (\mathcal{T}_k, (X(e), e \in \mathrm{e}(\mathcal{T}_k))) \,,$$

where $(\mathcal{T}_k, (X(e), e \in \mathrm{e}(\mathcal{T}_k)))$ is a random element of $\mathscr{T}^{(k)}$ with density

$$f(t, (x(e), e \in \mathrm{e}(T))) = \sum_{e \in \mathrm{e}(T)} x(e) \exp\left(-\Big(\sum_{e \in \mathrm{e}(t)} x(e) \Big)^2 / 2 \right) \,.$$

In part (b) of the exercise, the description of $(\mathcal{T}_k, (X(e), e \in \mathrm{e}(\mathcal{T}_k)))$ can be rephrased as follows. Fix any ordering $(X_1, \ldots, X_{2k-1})$ of $(X(e), e \in \mathrm{e}(\mathcal{T}_k)))$. Then $\mathcal{T}_k$ is independent of $(X_1, \ldots, X_{2k-1})$, and

$$(X_1, \ldots, X_{2k-1}) \overset{\mathrm{d}}{=} G \cdot (Y_1, \ldots, Y_{2k-1}) \,,$$

where G has density

$$\frac{1}{2^{k-1}(k-1)!}x^{2k-1}e^{-x^2/2}\mathbf{1}_{[x\geq 0]}, \tag{1.3.1}$$

and $(Y_1,\ldots,Y_{2k-1})$ is Dirichlet$(1,1,\ldots,1)$-distributed and is independent of G.

Exercise 1.3.5. *Show that for $\ell \geq 0$, if H is* Gamma$((\ell+1)/2,1/2)$*-distributed then $G=\sqrt{H}$ has density*

$$\frac{1}{2^{(\ell-1)/2}\Gamma(\frac{\ell+1}{2})}x^{\ell}e^{-x^2/2}\mathbf{1}_{[x\geq 0]}.$$

Exercise 1.3.6. *[A little informal but shouldn't be too hard to make sense of.] Construct a sequence of binary trees with edge lengths as follows. Let $\mathcal{T}_k$ consist of a single point. For $k \geq 1$, let $\mathcal{T}_k$ be the tree obtained from $\mathcal{T}_{k-1}$ by attaching a line-segment of length $P_k - P_{k-1}$ to a uniform point of $\mathcal{T}_{k-1}$ and giving the new leaf the label k. For $k \geq 1$ and $e \in e(\mathcal{T}_k)$ write* $\mathrm{len}(e)$ *for the edge length of e in $\mathcal{T}_k$. (When new leaves are attached, edges are subdivided and their length is split accordingly.) Note that $\mathcal{T}_k$ has total length P_k.*

(a) Show that $(\mathcal{T}_k,(\mathrm{len}(e),e\in\mathcal{T}_k))$ has density $P_k e^{-P_k^2/2}$.

(b) Show that if edge lengths are ignored, then $\mathcal{T}_{k+1}$ is formed from $\mathcal{T}_k$ by choosing a uniformly random edge, subdividing it, and attaching a leaf with label $k+1$ to the newly formed vertex.

1.4 Local structure

For a node v in a rooted tree $\mathrm{t} = (V,E,\rho)$, write $\mathrm{t}_v^{\uparrow}$ for the subtree of t rooted at v. The nodes of $\mathrm{t}_v^{\uparrow}$ are precisely the nodes w for which $v \in [\rho, w]$.

The next proposition describes the distribution of the subtree rooted at a typical node in a uniformly random tree.

Proposition 1.4.1. *Let $\mathrm{T} \in_u \mathscr{T}_n$. Then for any $v \in [n]$ and any $1 \leq k \leq n$,*

$$\mathbf{P}\left\{|\mathrm{T}_v^{\uparrow}| = k\right\} = \frac{k^{k-1}}{k!}\left(\frac{n-k}{n}\right)^{n-k}\frac{(n-1)_{k-1}}{n^{k-1}}.$$

Moreover, the vertex set of $\mathrm{T}_v^{\uparrow}$, excluding v, is a uniformly random subset of $[n]\setminus\{v\}$ conditional on its size, and $\mathrm{T}_v^{\uparrow}$ is a uniformly random tree rooted at v conditional on its vertex set.

Proof. Let
$$\mathscr{T}_n(v,k) = \{\mathrm{t} \in \mathscr{T}_n : |\mathrm{t}_v^{\uparrow}| = k\}.$$
We may describe a tree t in $\mathscr{T}_n(v,k)$ in four steps.

1. Specify the vertex set of $\mathrm{t}_v^{\uparrow}$.
2. Specify the tree $\mathrm{t}_v^{\uparrow}$.
3. Specify the tree $\mathrm{t} - \mathrm{t}_v^{\uparrow}$ obtained by removing $\mathrm{t}_v^{\uparrow}$ from t
4. Specify the parent of v in $\mathrm{t} - \mathrm{t}_v^{\uparrow}$.

In order to have $\mathrm{t} \in \mathscr{T}_n(v,k)$, the vertex set of $\mathrm{t}_v^{\uparrow}$ must consist of v and $k-1$ other vertices, so there are $\binom{n-1}{k-1}$ choices for the first step.

Having chosen the vertex set of $\mathrm{t}_v^{\uparrow}$, there are then k^{k-2} possibilities for $\mathrm{t}_v^{\uparrow}$ by Cayley's formula (note that the root of $\mathrm{t}_v^{\uparrow}$ is always v).

Finally, having chosen the vertex set of $\mathrm{t}_v^{\uparrow}$, then there are $(n-k)^{n-k-1}$ choices for the tree $\mathrm{t} - \mathrm{t}_v^{\uparrow}$, and $(n-k)$ choices for the parent of v in $\mathrm{t} - \mathrm{t}_v^{\uparrow}$

It follows that

$$|\mathscr{T}_n(v,k)| = \binom{n-1}{k-1} k^{k-2}(n-k)^{n-k}\,.$$

The first equality asserted in the proposition follows by Cayley's formula since $\mathbf{P}\left\{|\mathrm{t}_v^{\uparrow}| = k\right\} = |\mathscr{T}_n(v,k)|/|\mathscr{T}_n|$ and since

$$\frac{1}{n^{n-1}}\binom{n-1}{k-1}k^{k-2}(n-k)^{n-k} = \frac{k^{k-1}}{k!}\left(\frac{n-k}{n}\right)^{n-k}\frac{(n-1)_{k-1}}{n^{k-1}}\,.$$

For the remaining assertions of the proposition, if rather than requiring only that $|\mathrm{t}_v^{\uparrow}| = k$, we require that $\mathrm{t}_v^{\uparrow} = t$ for a specific tree t with k vertices and root v, then to fully describe t we need only specify $\mathrm{t} - \mathrm{t}_v^{\uparrow}$ and the parent of v in $\mathrm{t} - \mathrm{t}_v^{\uparrow}$. Thus, the number of trees $\mathrm{t} \in \mathscr{T}_n$ with $\mathrm{t}_v^{\uparrow} = t$ is

$$|\{\mathrm{t} \in \mathscr{T}_n : \mathrm{t}_v^{\uparrow} = t\}| = (n-|t|)\cdot(n-|t|)^{n-|t|-1} = (n-k)^{n-k}.$$

The other assertions follow since this quantity depends on t only through its size. □

Corollary 1.4.2. *Let* $\mathrm{T} \in_u \mathscr{T}_n$*. Then for fixed* k*, as* $n \to \infty$*, for any* $v \in [n]$*,*

$$\mathbf{P}\left\{|\mathrm{T}_v^{\uparrow}| = k\right\} = (1+o(1))\frac{k^{k-1}e^{-k}}{k!}\,.$$

Proof. This is immediate from the proposition since $(\frac{n-k}{n})^{n-k} = (1+o(1))e^{-k}$ and $(n-1)_{k-1}/n^{k-1} = 1+o(1)$. □

It is not obvious that

$$\sum_{k\geq 1}\frac{k^{k-1}e^{-k}}{k!} = 1\,, \tag{1.4.1}$$

so that the expression on the right-hand side of the display in Corollary 1.4.2 defines a probability distribution, but this is in fact the case.

The distribution is called the *Borel* distribution with parameter 1, or the Borel(1) distribution for short, so Corollary 1.4.2 says that $|\mathrm{T}_v^{\uparrow}|$ converges in distribution to a Borel(1) random variable.

In fact, the Borel(1) distribution is the distribution of the total number of individuals in a Poisson(1) branching process, which is finite almost surely.[10]

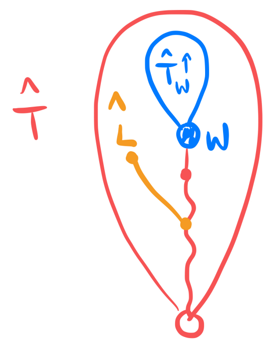
[10] Elaborate on this, here or later...

Exercise 1.4.1. *Let* $\mathrm{T} \in_u \mathscr{T}_n$. *Show that for any fixed* $i \in \mathbb{N}$, *as* $n \to \infty$, *the vector* $(|\mathrm{T}_j^{\uparrow}|, 1 \le j \le i)$ *converges in distribution to a vector of independent Borel*(1) *random variables. In other words, for any fixed positive integers* $n_1, \ldots, n_i$

$$\mathbf{P}\left\{|\mathrm{T}_j^{\uparrow}| = n_j, 1 \le j \le i\right\} \to \prod_{j=1}^{i} \frac{n_j^{n_j-1} e^{-n_j}}{n_j!}$$

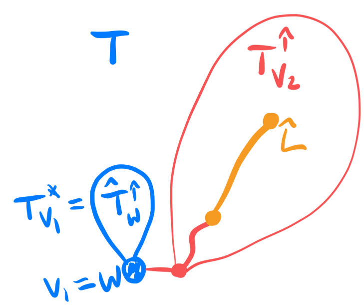
as $n \to \infty$.

Exercise 1.4.2. *Fix* $n \in \mathbb{N}$, *let* $\mathrm{T} \in_u \mathscr{T}_n$ *and let* L *be the smallest-labeled leaf of* T. *Let* $W \in_u [n]$ *be independent of* T_n. *Write* $p(n) = \mathbf{P}\left\{L \in \mathrm{T}_W^{\uparrow}\right\}$. *Show that* $p(n) \to 0$ *as* $n \to \infty$.



We can use the distribution of a random subtree to understand the structure of the random tree $\mathrm{T} \in_u \mathscr{T}_n$ close to the root.

Proposition 1.4.3. *Fix any integer* $k \ge 1$. *Let* $\mathrm{V} \in_u [n]^{n-1}$ *and let* $\mathrm{T} = \mathrm{T}(\mathrm{V})$. *For* $1 \le i < k$ *let* $\mathrm{T}_{V_i}^*$ *be the subtree of* T *consisting of descendants of* V_i *which are not descendants of* V_{i+1}. *Then as* $n \to \infty$, *the trees* $(\mathrm{T}_i^*, 1 \le i < k)$ *are are independent uniformly random trees conditional on their vertex sets and the labels of their roots. Moreover, their sizes* $(|\mathrm{T}_i^*|, 1 \le i < k)$ *are asymptotically independent and Borel*(1)*-distributed.*



Proof. The fact that $\mathrm{T}_{V_1}^*, \ldots, \mathrm{T}_{V_k}^*$ are independent uniformly random trees conditional on their vertex sets and the labels $V_1, \ldots, V_k$ of their roots is immediate from the fact that T is a uniformly random tree. We thus focus on proving that the sizes $(|\mathrm{T}_i^*|, 1 \le i < k)$ are asymptotically independent and Borel(1)-distributed.

Let $\hat{\mathrm{T}} \in_u \mathscr{T}_n$, let $W \in_u [n]$ be independent of $\hat{\mathrm{T}}$, and let T be obtained from $\hat{\mathrm{T}}$ be rerooting T at vertex W; then $\mathrm{T} \in_u \mathscr{T}_n$ as well.

Now write $\mathrm{V} = (V_1, \ldots, V_{n-1}) = \mathrm{v}(\mathrm{T})$, and note that $W = V_1$. Then $\mathrm{T}_{V_1}^* = \mathrm{T}_W^* = \hat{\mathrm{T}}_W^{\uparrow}$ unless the smallest labeled leaf $\hat{L}$ of $\hat{\mathrm{T}}$ is in $\hat{\mathrm{T}}_W^{\uparrow}$. For any $n_1 \ge 1$ we have

$$\begin{aligned}
\mathbf{P}\left\{|\mathrm{T}_{V_1}^*| = n_1\right\} &= \mathbf{P}\left\{|\mathrm{T}_{V_1}^*| = n_1, \hat{L} \notin \hat{\mathrm{T}}_W^{\uparrow}\right\} + \mathbf{P}\left\{|\mathrm{T}_{V_1}^*| = n_1, \hat{L} \in \hat{\mathrm{T}}_W^{\uparrow}\right\} \\
&= \mathbf{P}\left\{|\mathrm{T}_W^{\uparrow}| = n_1, \hat{L} \notin \hat{\mathrm{T}}_W^{\uparrow}\right\} + \mathbf{P}\left\{|\mathrm{T}_{V_1}^*| = n_1, \hat{L} \in \hat{\mathrm{T}}_W^{\uparrow}\right\} \\
&= \mathbf{P}\left\{|\mathrm{T}_W^{\uparrow}| = n_1\right\} - \mathbf{P}\left\{|\mathrm{T}_W^{\uparrow}| = n_1, \hat{L} \in \hat{\mathrm{T}}_W^{\uparrow}\right\} + \mathbf{P}\left\{|\mathrm{T}_{V_1}^*| = n_1, \hat{L} \in \hat{\mathrm{T}}_W^{\uparrow}\right\}
\end{aligned}$$

so

$$|\mathbf{P}\left\{|\mathrm{T}^*_{V_1}| = n_1\right\} - \mathbf{P}\left\{|\hat{\mathrm{T}}^{\uparrow}_{W}| = n_1\right\}| \le \mathbf{P}\left\{\hat{L} \in \hat{\mathrm{T}}^{\uparrow}_{W}\right\}.$$

It follows by Exercise 1.4.2 that

$$\left|\mathbf{P}\left\{|\mathrm{T}^*_{V_1}| = n_1\right\} - \mathbf{P}\left\{|\hat{\mathrm{T}}^{\uparrow}_{W}| = k\right\}\right| \to 0$$

as $n \to \infty$; by Corollary 1.4.2 and since W is independent of $\hat{\mathrm{T}}$, it follows that $|\mathrm{T}^*_{V_1}|$ is asymptotically Borel(1)-distributed.

For $k > 1$, we argue by induction. Let $n' = n - |\mathrm{T}^*_{V_1}|$, let $\mathrm{T}' = \mathrm{T}^{\uparrow}_{V_2}$ and let $\mathrm{V}' = (V'_1, \ldots, V_{n'}) = \mathrm{v}(\mathrm{T}')$. If $V_1, \ldots, V_k$ are all distinct, then $(V'_1, \ldots, V'_{k-1}) = (V_2, \ldots, V_k)$ and

$$(\mathrm{T}^*_{V_2}, \ldots, \mathrm{T}^*_{V_k}) = (\mathrm{T}'^*_{V'_1}, \ldots, \mathrm{T}'^*_{V'_{k-1}})$$

Writing E for the event that $V_1, \ldots, V_k$ are all distinct, then for any integers $n_2, \ldots, n_k \ge 1$,

$$\begin{aligned}
&\Big|\mathbf{P}\left\{ (|\mathrm{T}^*_{V_2}|, \ldots, |\mathrm{T}^*_{V_k}|) = (n_2, \ldots, n_k) \;\Big|\; |\mathrm{T}^*_{V_1}| = n_1\right\} \\
&- \mathbf{P}\left\{ (|\mathrm{T}'^*_{V'_1}|, \ldots, |\mathrm{T}'^*_{V'_{k-1}}|) = (n_2, \ldots, n_k) \;\Big|\; |\mathrm{T}^*_{V_1}| = n_1\right\}\Big| \\
&\le \mathbf{P}\left\{ E^c \mid |\mathrm{T}^*_{V_1}| = n_1\right\} \to 0
\end{aligned}$$

as $n \to \infty$, since $\mathbf{P}\{E\} \to 0$ and $\mathbf{P}\left\{|\mathrm{T}^*_{V_1}| = n_1\right\}$ is bounded away from zero.

Now, for any fixed integer $n_1 \ge 1$, conditionally given that $|\mathrm{T}^*_{V_1}| = n_1$, we have $n' = n - |\mathrm{T}^*_{V_1}| \to \infty$ as $n \to \infty$, so it follows by induction that under this conditioning, $(|\mathrm{T}'^*_{V'_1}|, \ldots, |\mathrm{T}'^*_{V'_{k-1}}|)$ are asymptotically independent and Borel(1)-distributed:

$$\mathbf{P}\left\{ (|\mathrm{T}'^*_{V'_1}|, \ldots, |\mathrm{T}'^*_{V'_{k-1}}|) = (n_2, \ldots, n_k) \;\Big|\; |\mathrm{T}^*_{V_1}| = k\right\} \to \prod_{j=2}^{k} \frac{n_j^{n_j - 1} e^{-n_j}}{n_j!}.$$

This completes the inductive step and the proof. ☐

Exercise 1.4.3. *Let* $\mathrm{T} \in_u \mathscr{T}_n$ *and let* $(V_1, \ldots, V_n) = \mathrm{v}(\mathrm{T})$. *List the children of* V_1 *in increasing order of label as* $U_1, \ldots, U_C$. *Show that for any fixed integers* $1 \le b \le c$,

$$\mathbf{P}\{C = c, V_2 = U_b\} \to \frac{1}{e}\frac{1}{c!}$$

as $n \to \infty$. *(Suggestion: use Proposition 1.2.3 to estimate the degree of* V_1, *and argue that* V_2 *is asymptotically a uniformly random child of* V_1.*)*

The above proposition describes the local structure of a large, uniformly random tree near its root - it consists of a long path (an infinite path in the $n \to \infty$ limit) of independent uniformly random

Borel(1)-sized trees. Exercise 1.4.3 additionally tells us that if we order the children of each node from left to right in increasing order of label, then for each node on the infinite path, the child through which the infinite path continues is uniformly random. This structure was described (in slightly different but essentially equivalent ways by Kennedy[11] and by Grimmett (1980)[12].

[11] https://doi.org/10.2307/3212730,https://sci-hub.do/10.2307/3212730

[12] https://tinyurl.com/grimmett1980

1.5 *Connected graphs beyond trees*

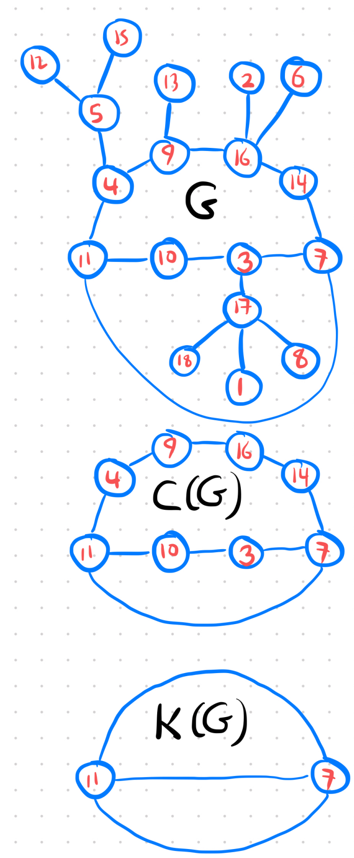
The bijections developed above may also be used to understand the global structure of connected graphs which contain cycles. For this section it's sometimes useful to allow multigraphs, which are graphs which may have multiple edges between vertices, and may also have loop edges. For a multigraph $G = (V, E)$ and an edge $e \in E$, we write $\mathrm{mult}(e) = \mathrm{mult}(e; G)$ for the number of copies (the multiplicity) of e in G. However, "graph" means "simple graph", unless otherwise stated.

Surplus

The *surplus* of a connected multigraph $G = (V, E)$ is the integer $s(G) := 1 + |E| - |V|$, which is the number of edges more than a tree that G has. The goal of this section is to describe the global structure of random connected graphs with large size and with a fixed surplus.

Core

We define a *core* to be a connected graph C with minimum degree 2. Given a connected graph G, the *core of G*, denoted $C(G)$, is the maximum induced subgraph of G with minimum degree 2.



Exercise 1.5.1. *The core of a connected graph is unique.*

Fix a core C with $\mathrm{v}(C) \subset [n]$, and write $k = |\mathrm{v}(C)|$. There is a natural bijection between the set of graphs G with $\mathrm{v}(G) = [n]$ and $C(G) = C$, and the set of forests $\mathscr{F}_n^{\mathrm{v}(C)}$: given such a graph G, form a forest in $\mathscr{F}_n^{\mathrm{v}(C)}$ by removing all edges of C, and rooting each connected component of the resulting graph at its unique element of $\mathrm{v}(C)$. Conversely, given a forest F in $\mathscr{F}_n^{\mathrm{v}(C)}$, one may form a graph G with $\mathrm{v}(G) = [n]$ and $C(G) = C$ by identifying the root of each tree of F with the vertex of C possessing the same label. It follows that

$$|\{\text{Graphs } G \,:\, \mathrm{v}(G) = [n], C(G) = C\}| = kn^{n-k-1},$$

The *kernel* of a connected graph G is the multigraph $K(G)$ obtained from the core $C(G)$ by replacing all maximal length paths all of whose internal vertices have degree 2 by edges. Note that $K(G)$ can have multiple edges and loops (and multiple edges which are loops).

An important point is that for any graph G, the surplus of G, of $C(G)$ and of $K(G)$ are all equal. This can be seen as follows: $C(G)$

can be obtained from G by repeatedly removing leaves (degree-one vertices), and this does not change the surplus. Then, when replacing a path of degree 2 vertices by a single edge, the number of vertices removed is equal to the number of edges removed, so again the surplus does not change.

Suppose C is a core with vertex set $\mathrm{v}(C) = V \subset \mathbb{N}$, and let $K = K(C)$ be its kernel. List the edges of K in lexicographic order as $e(1), \ldots, e(m)$, with $e(m) = (u(m), v(m))$ and $u(m) \leq v(m)$. (Each edge e appears in this list a number of times equal to its $\mathrm{mult}(e; K)$.) Write $P(i)$ for the path between $u(i)$ and $v(i)$ in C which is replaced by the edge $e(i)$ in K, excluding its endpoints.[13] Then C may be recovered from the kernel K together with the paths $(P(i), 1 \leq i \leq m)$.

The reconstruction of C from K and the paths $P(1), \ldots, P(m)$ is not necessarily a one-to-one map. Note that if $e \in \mathrm{e}(K)$ has multiplicity m – say e appears in the list of edges as $e(i+1), \ldots, e(i+m)$ – then permuting the paths $P(i+1), \ldots, P(i+m)$ results in an identical reconstruction. Also, if $e(i)$ is a loop, then there are two choices how to list the path $P(i)$. It follows that

$$\begin{aligned} &|\{\text{Cores } C : \mathrm{v}(C) = V, K(C) = K\}| \\ &= \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathrm{mult}(e;K)\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!} \cdot \sum_{(P_1,\ldots,P_m)} \mathbf{1}_{[(P_1,\ldots,P_m) \text{ valid}]} \, . \end{aligned} \tag{1.5.1}$$

In this expression, the sum is over valid[14] m-tuples of (possibly empty) paths $(P_1, \ldots, P_m)$ with disjoint vertex sets whose union is $V \setminus \mathrm{v}(K)$.

Proposition 1.5.1. *Fix a kernel K with vertex set $[k]$ and with m edges. Then as $\ell \to \infty$, the number of cores C with $\mathrm{v}(C) = [\ell]$ and with $K(C) = K$ is*

$$(1 + o(1)) \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathrm{mult}(e;K)\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!} \cdot \frac{(\ell - k)!\ell^{m-1}}{(m-1)!} \, .$$

Proof. By (1.5.1), it suffices to show that the number of valid m-tuples with disjoint vertex sets whose union is $[\ell] \setminus [k]$ is

$$(1 + o(1)) \frac{(\ell - k)!(\ell - k + 1)^{m-1}}{(m-1)!} = (1 + o(1)) \frac{(\ell - k)!\ell^{m-1}}{(m-1)!}. \tag{1.5.2}$$

The two expressions in (1.5.2) have the same asymptotic behaviour for k fixed as $\ell \to \infty$, but the next construction more naturally relates to the second expression, which is why we provided it.

We may form such an m-tuple $(P_1, \ldots, P_m)$ as follows.

1. Choose an ordering $(s_1, \ldots, s_{\ell-k})$ of $[\ell] \setminus [k]$

2. Choose $(i_1, \ldots, i_{m-1}) \in [\ell - k + 1]^{m-1}$

[13] This means that if $e(i)$ is actually an edge of C then $P(i)$ is empty; otherwise it is non-empty. Note that only one copy of an edge can have an empty path, since C is simple.

[14] By saying $(P_1, \ldots, P_m)$ is valid, we mean that (a) for any $1 \leq i < j \leq m$ such that $e(i), \ldots, e(j)$ are copies of the same edge, at most one of $P_i, \ldots, P_j$ is empty, and (b) for any $1 \leq i \leq m$, if $e(i)$ is a loop then P_i has at least two vertices. This ensures that the reconstruction results in a simple graph.

3. List $(i_1, \ldots, i_{m-1})$ in non-decreasing order as $(i(1), \ldots, i(m-1))$; set $i(0) = 1$ and $i(m) = \ell - k + 1$.

4. For $1 \le j \le m$, let $P_j = s_{i(j-1)}, s_{i(j-1)+1}, \ldots, s_{i(j)-1}$.[15]

[15] If $i(j-1) = i(j)$ then P_i is empty.

Note that if $3 \le i_1, \ldots, i_{m-1} \le \ell + k - 1$, and $i(j) \ge i(j-1) + 2$ for all $1 \le j \le m$, then $P_1, \ldots, P_m$ all contain at least two vertices, so $(P_1, \ldots, P_m)$ is valid. Therefore, if ℓ is large and $i_1, \ldots, i_{k-1}$ are independent, uniform samples from $[\ell - k]$, then with high probability $P_1, \ldots, P_m$ is valid.

If $i_1, \ldots, i_{m-1}$ are all distinct, then there are $(m-1)!$ different orderings of $i_1, \ldots, i_{m-1}$, and each yields the same non-decreasing reordering $(i(1), \ldots, i(m-1))$. This implies that the expression in (1.5.2) is a lower bound on the number of valid m-tuples with disjoint vertex sets whose union is $[\ell] \setminus [k]$.

Finally, the number of choices of $(i_1, \ldots, i_{m-1}) \in [\ell - k + 1]^{m-1}$ with at least two of $i_1, \ldots, i_{m-1}$ taking the same value or neighbouring values is $O((\ell - k + 1)^{m-2}) = O(\ell^{m-2}) = o(\ell^{m-1})$; so m-tuples with at least one empty path contribute a lower-order term to the overall count. It follows that the number of valid m-tuples $(P_1, \ldots, P_m)$ with disjoint vertex sets whose union is $[\ell] \setminus [k]$ is $(1 + o(1))(\ell - k)!(\ell - k + 1)^{m-1}/(m-1)!$, as required. $\square$

For positive integers n and s, write $\mathscr{G}_{n,s}$ for the set of connected graphs with vertex set $[n]$ and with surplus s.

Corollary 1.5.2. *Fix $s \ge 2$, and let $C = C_\ell$ be a uniformly random core with $\mathrm{v}(C) = [\ell]$ and surplus $s(C) = s$. Then as $\ell \to \infty$, with high probability $K(C)$ is a 3-regular multigraph with $2(s-1)$ vertices and $3(s-1)$ edges, and for any fixed such multigraph K,*

$$\lim_{\ell \to \infty} \mathbf{P}\{K(C_\ell) = K\} \propto \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!}.$$

Proof. Any kernel K has minimum degree 3, and so has $|\mathrm{e}(K)| \ge 3|\mathrm{v}(K)|/2$ and surplus

$$1 + |\mathrm{e}(K)| - |\mathrm{v}(K)| \ge 1 + |\mathrm{v}(K)|/2.$$

It follows that if K has surplus s then $|\mathrm{v}(K)| \le 2(s-1)$.

We now divide the set of cores with surplus s according to their kernel. In what follows when we write $\sum_K$ we implicitly mean that K is a kernel. We have

$$\begin{aligned}
&|\{\text{Cores } C : \mathrm{v}(C) = [\ell], s(C) = s\}| \\
&= \sum_K |\{\text{Cores } C : \mathrm{v}(C) = [\ell], s(C) = s, K(C) = K\}| \\
&= \sum_{k=1}^{2(s-1)} \sum_{K : \mathrm{v}(K) = [k]} \binom{\ell}{k} |\{\text{Cores } C : \mathrm{v}(C) = [\ell], s(C) = s, K(C) = [k]\}| .
\end{aligned}$$

A kernel with $\mathrm{v}(K) = [k]$ and with surplus s has $m = k - 1 + s$ edges, so by Proposition 1.5.1, it follows that this number is $(1 + o(1))$ times

$$\sum_{k=1}^{2(s-1)} \sum_{\substack{K:\mathrm{v}(K)=[k] \\ s(K)=s}} \binom{\ell}{k} \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathrm{mult}(e;K)\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!} \cdot \frac{(\ell-k)!\ell^{k+s-2}}{(k+s-2)!}$$
$$= \sum_{k=1}^{2(s-1)} \sum_{\substack{K:\mathrm{v}(K)=[k] \\ s(K)=s}} \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathrm{mult}(e;K)\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!} \cdot \frac{\ell!\ell^{k+s-2}}{k!(k+s-2)!}.$$

For ℓ large, the term $k = 2(s-1)$ dominates, due to the factor ℓ^{k+s-2}, so this sum is in turn $(1 + o(1))$ times

$$\sum_{\substack{K:\mathrm{v}(K)=[2(s-1)] \\ s(K)=s}} \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathrm{mult}(e;K)\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!} \cdot \frac{\ell!\ell^{3s-4}}{(2s-2)!(3s-4)!}$$
$$= \sum_{\substack{K:\mathrm{v}(K)=[2(s-1)] \\ s(K)=s}} \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!} \cdot \frac{\ell!\ell^{3s-4}}{(2s-2)!(3s-4)!}, \tag{1.5.3}$$

the last equality holding since when $\mathrm{v}(K) = [2(s-1)]$ and $s(K) = s$, necessarily K is 3-regular and so any loop edges in K must have multiplicity 1.

In particular, this implies that

$$|\{\text{Cores } C : \mathrm{v}(C) = [\ell], s(C) = s\}|$$
$$= (1 + o(1))|\{\text{Cores } C : \mathrm{v}(C) = [\ell], s(C) = s, |\mathrm{v}(K(C))| = 2(s-1)\}|,$$

which proves the first assertion of the corollary. For the second assertion, it suffices to note that the terms in the sum in (1.5.3) only depend on K via the product term $\prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!}$, so the probability that the kernel equals K must be proportional to this product. $\square$

Corollary 1.5.3. *Fix $s \geq 2$. Then as $\ell \to \infty$,*

$$|\{\text{Cores } C : \mathrm{v}(C) = [\ell], s(C) = s\}|$$
$$= (1 + o(1)) \frac{\ell!\ell^{3s-4}}{(2s-2)!(3s-4)!} \cdot \sum_{\substack{\text{Kernels } K \\ \mathrm{v}(K)=[2(s-1)], s(K)=s}} \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!}.$$

Proof. This is immediate from (1.5.3). $\square$

The above computations now allow us to figure out the typical core size for a large random graph with a fixed surplus. So that the formulas don't get too cumbersome, it's useful to write

$$\kappa(s) = \frac{1}{(2s-2)!(3s-4)!} \sum_{\substack{\text{Kernels } K \\ \mathrm{v}(K)=[2(s-1)], s(K)=s}} \prod_{e \in \mathrm{e}(K)} \frac{2^{-\mathbf{1}_{[e \text{ loop}]}}}{\mathrm{mult}(e;K)!}$$

so that the formula on the right-hand side of the above corollary becomes $(1+o(1))\kappa(s)\ell!\ell^{3s-4}$.

Theorem 1.5.4. *Fix any positive integer $s \geq 1$. Then as $n \to \infty$,*

$$|\mathscr{G}_{n,s}| = (1+o(1))\kappa(s) \cdot n^{n-2+3s/2} \int_0^\infty x^{3s-3} e^{-x^2/2} \mathrm{d}x\,.$$

Moreover, if $G \in_u \mathscr{G}_{n,s}$ then

$$|\mathrm{v}(C(G))|/n^{1/2} \xrightarrow{\mathrm{d}} X$$

where $X \stackrel{\mathrm{d}}{=} \sqrt{\mathrm{Gamma}(\frac{3s-2}{2}, \frac{1}{2})}$ has density

$$\frac{1}{2^{(3s-4)/2}\Gamma((3s-2)/2)} x^{3s-3} e^{-x^2/2} \mathbf{1}_{[x \geq 0]}\,.$$

Proof. We prove the theorem for $s \geq 2$ and leave the case $s = 1$ as an exercise. To specify a graph G in $\mathscr{G}_{n,s}$ we may first specify the vertex set of the core, then specify the core itself, and finally specify the trees which are attached to the core. It follows that

$$\begin{aligned}
&|\{G \in \mathscr{G}_{n,s} : |\mathrm{v}(C(G))| = \ell\}| \\
&= (1+o_\ell(1))\binom{n}{\ell} \cdot \kappa(s)\ell!\ell^{3s-4} \cdot \ell n^{n-\ell-1} \\
&= (1+o_\ell(1))\kappa(s)(n)_\ell n^{n-\ell-1}\ell^{3s-3}.
\end{aligned}$$

We would like to argue that only terms with $\ell = O(n^{1/2})$ contribute asymptotically to $|\mathscr{G}_{n,s}|$. To see this, first note that using the Taylor expansion $\log(1-x) = -x + O(x^2)$ we have

$$(n)_\ell = n^\ell \prod_{i=0}^{\ell-1}(1 - i/n) = n^\ell \exp\left(-\frac{i^2}{2n} + O\Big(\frac{i^3}{n^2}\Big)\right)\,.$$

For $\ell = O(n^{1/2})$ this gives

$$|\{G \in \mathscr{G}_{n,s} : |\mathrm{v}(C(G))| = \ell\}| = (1+o_\ell(1))\kappa(s)n^{n-1}\ell^{3s-3}e^{-\ell^2/2n} \tag{1.5.4}$$

It follows that the total number of graphs $G \in \mathscr{G}_{n,s}$ with $|\mathrm{v}(C(G))| = o(n^{1/2})$ is $o(|\mathscr{G}_{n,s}|)$. Next, summing over graphs G with $|\mathrm{v}(C(G))| = \Theta(n^{1/2})$ also yields the lower bound

$$|\mathscr{G}_{n,s}| = \Omega(n^{n-1/2+(3s-3)/2}).$$

Also, if $\omega(n)$ is any function with $\omega(n) \to \infty$ as $n \to \infty$, then the upper bound $(n)_\ell \leq n^\ell e^{-\ell(\ell-1)/2n}$ gives the bound

$$\begin{aligned}
&|\{G \in \mathcal{G}_{n,s} : |\mathrm{v}(C(G))| \geq \omega(n)\}| \\
&\leq (1+o(1))\kappa(s)n^{n-1} \sum_{\ell \geq \omega(n)} \ell^{3s-3}e^{-\ell(\ell-1)/2n} \\
&= (1+o(1))\kappa(s)n^{n-1+(3s-3)/2} \sum_{\ell \geq \omega(n)} (\ell/n^{1/2})^{3s-3}e^{-\ell(\ell-1)/2n} \\
&= o(n^{n-1/2+(3s-3)/2}) = o(|\mathscr{G}_{n,s}|)\,.
\end{aligned}$$

It follows from these bounds only graphs in $\mathscr{G}_{n,s}$ with core size $\Theta(n^{1/2})$ contribute asymptotically, and therefore (1.5.4) gives

$$|\mathscr{G}_{n,s}| = (1+o(1))\kappa(s)n^{n-2+3s/2}\int_0^\infty y^{3s-3}e^{-y^2/2}\mathrm{d}y\,.$$

When $\ell = \ell(n) = (1+o(1))xn^{1/2}$, equation (1.5.4) also gives

$$|\{G \in \mathscr{G}_{n,s} : |\mathrm{v}(C(G))| = \ell\}| = (1+o(1))\kappa(s)n^{n-2+3s/2}x^{3s-3}e^{-x^2/2}\,.$$

For $\mathrm{G} \in_u \mathscr{G}_{n,s}$, these estimates imply that for $\ell = \theta(n^{1/2})$,

$$\begin{aligned}
&\mathbf{P}\left\{|\mathrm{v}(C(\mathrm{G}))| = \ell\right\}\\
&= (1+o(1))\left(\frac{\ell}{n^{1/2}}\right)^{3s-3} e^{-\ell^2/(2n)}\cdot\left(\int_0^\infty y^{3s-3}e^{-y^2/2}\mathrm{d}y\right)^{-1}.\\
&= \frac{(1+o(1))}{2^{(3s-4)/2}\Gamma(\frac{3s-2}{2})}\left(\frac{\ell}{n^{1/2}}\right)^{3s-3} e^{-\ell^2/(2n)}\,;
\end{aligned}$$

to see the second identity, one may either verify that

$$\int_0^\infty y^k e^{-y^2/2}\mathrm{d}y = 2^{(k-1)/2}\Gamma\Big(\frac{k+1}{2}\Big)\,,$$

or else note that since $x^{3s-3}e^{-x^2/2}/\int_0^\infty y^{3s-3}e^{-y^2/2}\mathrm{d}y$ must be a probability density on $[0,\infty)$, it must be equal to the density from Exercise 1.3.5. □

Exercise 1.5.2. *Fix an integer $s \ge 0$, let $\mathrm{G}_n \in_u \mathscr{G}_{n,s}$, and list the leaves of G_n in increasing order as $(L_i, i \ge 1)$.*[16]

For $k \ge 1$ let C_n^k be the subgraph of G_n containing the core $C(\mathrm{G}_n)$ together with the paths from L_i to $C(\mathrm{G}_n)$ for $1 \le i \le k$. Let K_n^k be the graph obtained from C_n^k by replacing all maximal length paths all of whose internal vertices have degree 2 by edges. For $e \in \mathrm{e}(K_n^k)$ write $\mathrm{len}(e)$ for the number of edges of the path in C_n^k which gives rise to edge e in K_n^k.

[16] Set $L_i = L_1$ when i is greater than the number of leaves of G_n.

(a) Prove that for all $k \ge 1$ there is the joint convergence

$$\begin{aligned}
\frac{1}{n^{1/2}}|\mathrm{v}(C_n^k)| &\xrightarrow{\mathrm{d}} G\\
\frac{1}{n^{1/2}}(\mathrm{len}(e), e \in \mathrm{e}(K_n^k)) &\xrightarrow{\mathrm{d}} G\cdot(X_1, 1 \le i \le 3s-3+2k)\,,
\end{aligned}$$

where G is distributed as $\sqrt{\mathrm{Gamma}((3s-2+2k)/2, 1/2)}$ and $(X_1, 1 \le i \le 3s-3+2k)$ is $\mathrm{Dirichlet}(1,1,\ldots,1)$-distributed and is independent of G.

(b) Prove that

$$\frac{1}{n^{1/2}}\left(\frac{|\mathrm{v}(C_n^{k-1})|}{|\mathrm{v}(C_n^k|}, \frac{|\mathrm{v}(C_n^k)| - |\mathrm{v}(C_n^{k-1})|}{|\mathrm{v}(C_n^k)|}\right) \xrightarrow{\mathrm{d}} \mathrm{Dirichlet}(3s-4+2k, 1),$$

where the limiting vector is independent of the limit G in part (a).

(c) *Let P be a Poisson process on $[0, \infty)$ with rate $\lambda(t) = t$. Independently of P, let H be distributed as $\sqrt{\mathrm{Gamma}(3s-2)/2, 1/2}$, and list the atoms of P which fall in $[H, \infty)$ in increasing order as $H =: P_0 < P_1 < P_2 < \ldots$. Prove that for all $k \geq 1$, P_k is $\sqrt{\mathrm{Gamma}((3s-2+2k)/2, 1/2)}$-distributed.*

(d) *With the notation of part (c), show that for any $k \geq 0$, $(\frac{P_k}{P_{k+1}}, \frac{P_{k+1}-P_k}{P_{k+1}})$ is $\mathrm{Dirichlet}(3s-2+2k, 1)$-distributed and is independent of P_{k+1}.*

2

Bienaymé trees

2.1 *Plane trees and the Ulam-Harris tree*

The *Ulam-Harris tree* $\mathcal{U}$ has nodes labelled by $\bigcup_{n\geq 0}\mathbb{N}^n$, where $\mathbb{N}^0 := \{\emptyset\}$. The node $\emptyset$ is the root. In general, a node at level n is labeled by a string $v = v_1v_2\ldots v_n$; it has parent $\mathrm{par}(v) = v_1v_2\ldots v_{n-1}$ and children $(vi, i \geq 1) = (v_1\ldots v_n i, i \geq 1)$. We think of the children of v as being born one-at-a-time: first $v1$, then $v2$ and so on. If $i < j$ we say vi is an *older sibling* of vj.

We write $\mathcal{U} = \bigcup_{n\geq 0}\mathbb{N}^n$, identifying $\mathcal{U}$ with the set of labels of its nodes. (This is a bit sloppy, since the Ulam-Harris tree is not the only graph with these node labels, but this shouldn't cause any confusion.)

A *subtree* of $\mathcal{U}$ is a set $\mathrm{t} \subset \mathcal{U}$ with the following properties:

(a) $\emptyset \in \mathrm{t}$.

(b) If $v \in \mathrm{t}$ then $\mathrm{par}(v) \in \mathrm{t}$; the ancestors of v are all in t as well.

(c) If $v \in \mathrm{t}$, $v = wi$ then $wj \in \mathrm{t}$ for all $j \leq i$; the older siblings of v are all in t as well.

Given a subtree t of $\mathcal{U}$, for $v \in \mathrm{t}$ we write $c(v;\mathrm{t}) = \max(i : vi \in \mathrm{t})$; this is the *outdegree*, or *number of children* of v in t, and it may be infinite. We also write $\mathrm{t}_n := \mathrm{t} \cap \mathbb{N}^n$, and $\mathrm{t}_{\leq n} = \bigcup_{m=0}^{n}\mathrm{t}_m$ and the like.

A subtree $\mathrm{t} \subset \mathcal{U}$ is *finite* if $|\mathrm{t}| < \infty$. It is *locally finite* if $\mathrm{t}_n := \mathrm{t} \cap \mathbb{N}^n$ is finite for all n. Its *height* is $\mathrm{ht}(\mathrm{t}) := \max(n : \mathrm{t}_n \neq \emptyset)$.

From now on, the phrase "plane tree" means "locally finite subtree of $\mathcal{U}$", and we write $\mathcal{T}$ for the set of plane trees. Equivalently, a plane tree is a rooted tree t in which each node has finitely many children, together with a left-to-right ordering of the children of each vertex of t. The left-to-right orderings endows the nodes of a plane tree with a canonical labeling by strings of positive integers, as follows. The root is labeled by $\emptyset$; recursively, the children of a node with label $i_1,\ldots,i_k$ are labeled in left-to-right order by the elements of the set

$(i_1, \ldots, i_k, i, 1 \leq i \leq c(v; \mathrm{t}))$. Thus, the children of the root $\varnothing$ have labels $1, 2, \ldots, c(\varnothing; \mathrm{t})$, the children of the node with label 1 have labels $11, 12, \ldots, 1c(1; \mathrm{t})$, and so forth. By identifying the nodes of a plane tree with their vertex labels, we realize it as a subtree of the Ulam-Harris tree.

We wish to consider random trees, and for this we need to turn the set of trees into a measurable space.

Definition 2.1.1. *For a plane tree* t *and an integer* $n \geq 0$, *let* $[\mathrm{t}]_{\leq n} = \{\text{plane trees } \mathrm{t}' : \mathrm{t}'_{\leq n} = \mathrm{t}_{\leq n}\}$.

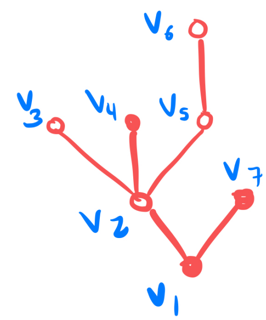
It's useful to also introduce the notation $[\,]_{<n} := [\,]_{\leq n-1}$. The equivalence relation $[\,]_{\leq n}$ partitions the set of trees into countably many equivalence classes; we let $\mathcal{F}_n = \sigma(\{[t]_{\leq n} : \mathrm{t} \in \mathcal{T}\})$, and let $\mathcal{F} = \sigma(\bigcup_{n \geq 0} \mathcal{F}_n)$. Note that $[\,]_{\leq n+1}$ refines $[\,]_{\leq n}$, which implies that $(\mathcal{F}_n, n \geq 0)$ is a filtration. Note that since $[\,]_{\leq n}$ is an equivalence relation, the sets $[\mathrm{t}]_{\leq n}$ are all atoms of $\mathcal{F}_n$.)

Exercise 2.1.1. *Let* $\mathcal{F}^{\mathrm{fin}} = \sigma(\{\{\mathrm{t}\} : \mathrm{t} \in \mathcal{T}, |\mathrm{t}| < \infty\})$ *be the* σ*-algebra generated by finite plane trees. Show that* $\mathcal{F}^{\mathrm{fin}} \subset \mathcal{F}$.

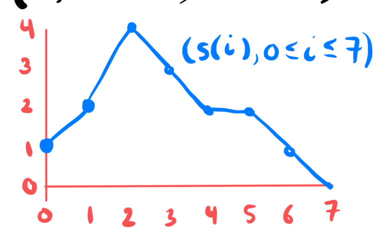
Perhaps: To a rooted tree t with vertex set $[n]$, we associate a plane tree using the convention that the children of each node are listed in increasing order of label from left to right?

2.2 *Plane trees and the cycle lemma*

For a finite plane tree t, by the *lexicographic ordering* of the vertices of T we mean the lexicographic ordering of the vertices according to their Ulam-Harris labels. In this ordering, each vertex appears before all its descendants, and the children of a fixed node appear in left-to-right order.



$$(d_1, d_2, d_3, d_4, d_5, d_6, d_7) = (2, 3, 0, 0, 1, 0, 0)$$



Fix a finite plane tree t, write $n = |\mathrm{t}|$, and list the vertices of t in lexicographic order as $v_1, \ldots, v_n$. Then, for $i \in [n]$ let $d(i) = c(v_i; \mathrm{t})$, and for $0 \leq i \leq n$ let $s(i) = 1 + \sum_{j=1}^{i}(d(j) - 1)$. The sequence $(s(0), \ldots, s(n))$ is called the *depth-first queue process* of t.

Imagine exploring the vertices of tree t in lexicographic order. At time zero, the root of t has been discovered but no vertices have been explored. For $1 \leq i \leq n$, at time i, node v_i is explored, and its set of children is discovered. Then for each $0 \leq i \leq n$, the quantity $s(i)$ is the number of vertices of t which have been discovered but not yet explored. This number is positive until the whole tree has been explored, so $s(i) > 0$ for $0 \leq i < n$ and $s(n) = 0$. Moreover, $s(i+1) \geq s(i) - 1$ for each $0 \leq i < n$, since at each step we explore exactly one vertex and discover $d(i+1) \geq 0$ new vertices.

Conversely, suppose that $(s_0, \ldots, s_n)$ are non-negative integers with $s_0 = 1$ and $s_n = 0$, and with $s_i > 0$ and $s_{i+1} \geq s_i - 1$ for each

$0 \leq i < n$. Then, writing $d_i = s_i - s_{i-1} - 1$ for $i \in [n]$, there is a unique plane tree t with n vertices, such that the degrees of the vertices of t listed in lexicographic order are $(d_1, \ldots, d_n)$. In other words, the tree t can be recovered from its depth-first queue process.

Exercise 2.2.1. *Show carefully that any finite plane tree* t *can be recovered from its depth-first queue process.*

The following combinatorial identity is fundamental for aspects of the study of branching process. It has been rediscovered in various forms by several researchers; the version we present here is more or less that of Dwass[1].

[1] https://doi.org/10.1214/aoms/1177704376

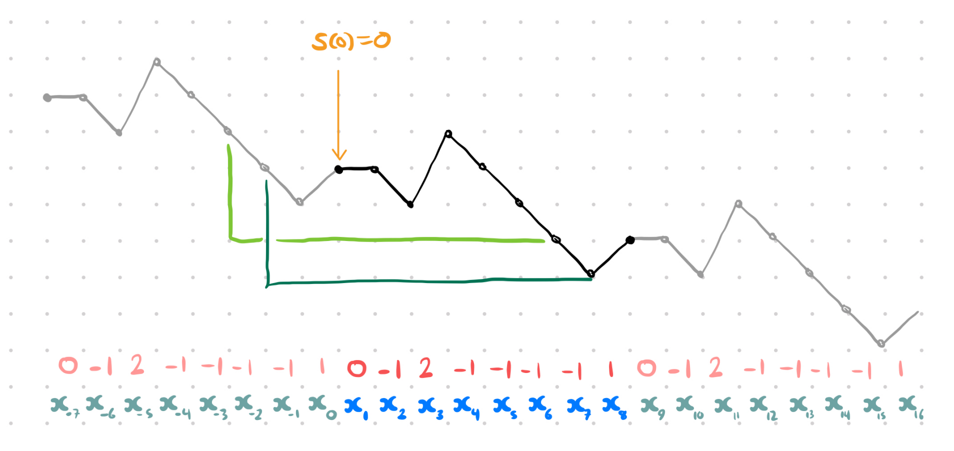
Proposition 2.2.1 (Dwass's cycle lemma). *Fix integers $x_1, x_2, \ldots, x_n \in \{-1, 0, 1, \ldots\}$ with $x_1 + \ldots + x_n = -r \leq 0$. For $j \in \mathbb{Z}$ and $i \in [n]$ let $s_j(i) = x_{j+1 \bmod n} + \ldots + x_{j+i \bmod n}$. Then there are exactly r values of $j \in [n]$ for which $s_k(i) > -r$ for all $0 < i < n$.*

Proof. We closely follow the proof given by Janson[2]. Extend the sequence $(x_k, k \in [n])$ to $\mathbb{Z}$ by evaluating the index modulo n, so $x_k = x_{k+in}$ for all $i \in \mathbb{Z}$. Then extend the definition of $(s(k), 0 \leq k \leq n)$ to $\mathbb{Z}$ accordingly, by setting $s(k) - s(k-1) = x_k$ for $k \in \mathbb{Z}$; since we fix $s(0) = 0$ this uniquely determines $s(k)$ for all $k \in \mathbb{Z}$. More precisely, we have

[2] https://doi.org/10.1214/11-PS188, see Lemma 15.3

$$s(k) = \begin{cases} \sum_{j=1}^{k} x_j & \text{if } k \geq 0 \\ -\sum_{j=k+1}^{0} x_j & \text{if } k < 0\,. \end{cases}$$

Since $x_1 + \ldots + x_n = -r$, this implies that $s(k+n) = s(k) - r$ for all $k \in \mathbb{Z}$. Note also that $s(i+j) - s(i) = s_i(j)$ for all $j \in \mathbb{Z}$ and $i \in [n]$. Next let $m(k) = \min_{-\infty < j \leq k} s(j)$. Since $s(j-n) = s(j) + r$



for all $j \in \mathbb{Z}$, we also have $m(k) = \min_{k-n<j\leq k} s(j)$, and moreover $m(k+n) = m(k) - r$ for all $k \in \mathbb{Z}$. Note also that for all $k \in \mathbb{Z}$, since

$s(k+1) = s(k) + x_{k+1}$,

$$m(k+1) = \begin{cases} m(k) - 1 & \text{if } s(k) = m(k) \text{ and } x_{k+1} = -1 \\ m(k) & \text{otherwise.} \end{cases}$$

It follows that

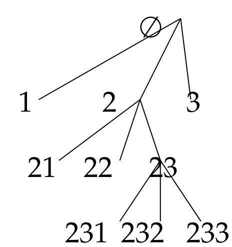
$$\begin{aligned} s_k(i) > -r \text{ for all } 0 < i < n &\Leftrightarrow s(k+i) - s(k) > -r \text{ for all } 0 < i < n \\ &\Leftrightarrow s(k+i) - s(k) + r > 0 \text{ for all } 0 < i < n \\ &\Leftrightarrow s(k+i-n) - s(k) > 0 \text{ for all } 0 < i < n \\ &\Leftrightarrow s(j) > s(k) \text{ for } k-n < j < k \\ &\Leftrightarrow m(k-1) > s(k) \\ &\Leftrightarrow m(k-1) > m(k)\,. \end{aligned}$$

Finally, since $m(n) = m(0) - r$ and $m(i+1) \geq m(i) - 1$ for all i, there are exactly r integers $k \in \{1, \ldots, n\}$ for which $m(k-1) > m(k)$. This completes the proof. □

2.3 *Branching processes, Bienaymé trees and the fundamental theorem*

Fix a probability distribution μ on $\mathbb{R}$ with $\mu(\mathbb{N}) = 1$. A *Bienaymé tree* with offspring distribution μ, or a Bienaymé(μ) tree for short, is a random plane tree T which is the family tree of a branching process with offspring distribution. The law B_μ of T is uniquely determined by the following property: for all $h \geq 1$, for any for any plane tree t of height at most h,

$$\mathbf{P}\left\{T^{\leq h} = t\right\} = \mathrm{B}_\mu(\{\mathrm{t} \in \mathcal{T} : \mathrm{t}^{\leq h} = t\}) = \prod_{v \in t^{\leq h-1}} \mu(\deg_t(v)). \quad (2.3.1)$$



A Bienaymé(μ) tree may be constructed as follows.

- ⋆ Start from the root (call it ∅), let $X_\emptyset$ have law μ.
- ⋆ Give ∅ children $1, \ldots, X_\emptyset$.
- ⋆ Independently for each $i = 1, \ldots, X_\emptyset$, let X_i have law μ.
- ⋆ Give i children $i1, i2, \ldots, iX_i$.
- ⋆ Repeat forever or until done; call the resulting random tree T.

Equivalently: let $(X_v, v \in \mathcal{U})$ be independent with law μ. Then let T be the random subtree of $\mathcal{U}$ in which the root ∅ has $X_\emptyset$ children and more generally, inductively, if $v \in \mathrm{T}$ then $c(v, \mathrm{T}) := X_v$.

Exercise 2.3.1. *Suppose that the random variables $(X_v, v \in \mathcal{U})$ are defined on a common probability space $(\Omega, \mathcal{G}, \mathbf{P})$. Show that the above construction of* T *yields a $\mathcal{G}/\mathcal{F}$-measurable map from Ω to $\mathcal{T}$. In other words,* T *is a $(\mathcal{T}, \mathcal{F})$-valued random variable.*

Let $Z_n = Z_n(\mathrm{T})$ be the number of individuals of T in the n'th generation $\mathbb{N}^n$, and write $|\mathrm{T}| = \sum_{n=0}^{\infty} Z_n$ for the total number of individuals. We say the survival occurs if $Z_n > 0$ for all n, and otherwise that say that extinction occurs. Equivalently, survival occurs if $|\mathrm{T}| = \infty$, and extinction occurs if $|\mathrm{T}| < \infty$.

Theorem 2.3.1 (Fundamental theorem of branching processes)**.** *Let X be a non-negative random variable integer random variable with distribution μ, and let* T *be* B_μ*-distributed. Then $\mathbf{P}\{|\mathrm{T}| = \infty\} > 0$ if and only if one of the following two conditions holds.*

- $\mathbf{P}\{X = 1\} = 1$
- $\mathbf{E}[X] > 1.$

As a warm up, we prove the following lemma.

Lemma 2.3.2. *Let X be μ-distributed. Then for all n, $\mathbf{E}[Z_n] = [\mathbf{E}X]^n$.*

Proof. This is obviously true for $n = 0$. Supposing the equality holds for a given n, we write

$$\mathbf{E}[Z_{n+1}] = \sum_{i=0}^{\infty} \mathbf{E}[Z_{n+1} \mid X_\emptyset = i]\, \mathbf{P}\{X_\emptyset = i\}.$$

Given that $X_\emptyset = i$, the children $1, \ldots, i$ of $\emptyset$ are each the root of an independent copy of the whole process, so

$$\mathbf{E}[Z_{n+1} \mid X_\emptyset = i] = i\mathbf{E}[Z_n].$$

We thus have

$$\mathbf{E}[Z_{n+1}] = \sum_{i=0}^{\infty} i\mathbf{E}[Z_n]\, \mathbf{P}\{X_\emptyset = i\} = \mathbf{E}[Z_n] \cdot \mathbf{E}[X] = [\mathbf{E}[X]]^{n+1},$$

the last step by induction. □

Corollary 2.3.3. *If $\mathbf{E}[X] < 1$ then $\mathbf{E}|\mathrm{T}| < \infty$, so $\mathbf{P}\{|\mathrm{T}| = \infty\} = 0$.*

Proof. If $\mathbf{E}[X] < 1$ then

$$\mathbf{E}|\mathrm{T}| = \sum_{n=0}^{\infty} \mathbf{E}[Z_n] = \sum_{n=0}^{\infty} (\mathbf{E}X)^n = \frac{1}{1 - \mathbf{E}[X]} < \infty.$$

It follows by Markov's inequality that $\mathbf{P}\{|\mathrm{T}| = \infty\} = 0$. □

Here is another approach to the above lemma which gives a little more. For $n \geq 0$ let $\mathcal{F}_n^\circ = \sigma(Z_0, \ldots, Z_n)$. Then $Z_{n+1} = \sum_{v \in \mathrm{T}_n} X_v$, so for any *fixed* subset S of $\mathbb{N}^n$,

$$\mathbf{E}\left\{ Z_{n+1} \mid \mathrm{T}_n = S \right\} = \mathbf{E}\left\{ \sum_{v \in S} X_v \;\middle|\; \mathcal{S} = S \right\} = \sum_{v \in S} \mathbf{E}X = |S| \cdot \mathbf{E}X.$$

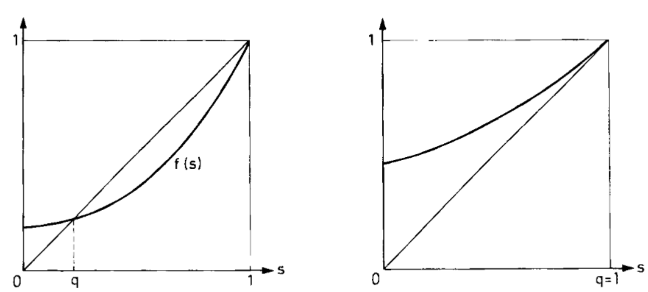
The second inequality holds by linearity of expectation and since $\mathbf{E}\left\{ X_v \mid \mathcal{S} = S \right\} = \mathbf{E}X$. Since $Z_n = |\mathcal{S}|$, it follows that $\mathbf{E}\left\{ Z_{n+1} \mid \mathcal{F}_n^\circ \right\} = Z_n \cdot \mathbf{E}X$. Therefore, if $\mathbf{E}X = 1$ then $(Z_n, n \geq 0)$ is an $\mathcal{F}_n^\circ$-martingale. More generally, setting $M_n = M_n(\mathrm{T}) = Z_n(\mathrm{T})/(\mathbf{E}X)^n$, then $(M_n, n \geq 0)$ is always an $\mathcal{F}_n^\circ$-martingale.

Exercise 2.3.2. *With* T *constructed as in Exercise 2.3.1, show that* $M_n = M_n(\mathrm{T})$ *is a* $\mathbf{P}$*-martingale with respect to* $(\mathcal{F}_n')$*, where* $\mathcal{F}_n' = \sigma(X_v, v \in \mathcal{U}_{<n})$*.*

Now let $F(z) = F_\mu(z) = \mathbf{E}\left[z^X\right] = \sum_{k=0}^\infty \mu(k) z^k$.

Proposition 2.3.4 (Fundamental proposition of branching processes)**.** *If* $\mathbf{P}\left\{X = 1\right\} < 1$ *then*

$$\mathbf{P}\left\{|\mathrm{T}| < \infty\right\} = \min_{x \geq 0}\{F(x) = x\}.$$



Proof. Write $p = \mathbf{P}\left\{|\mathrm{T}| < \infty\right\}$. We prove the proposition in two parts: first we show that $F(p) = p$, and second we show that p is the *smallest* non-negative solution of $F(x) = x$.

The proof of the first part is similar to that of the proof of the lemma. We begin by noting that

$$|\mathrm{T}| < \infty \Leftrightarrow Z_n = 0 \text{ for some } n,$$

so

$$p = \mathbf{P}\left\{ \bigcup_{n=0}^\infty Z_n = 0 \right\}.$$

The events on the right are increasing (if $Z_n = 0$ then $Z_{n+1} = 0$) so it follows that

$$p = \lim_{n\to\infty} \mathbf{P}\{Z_n = 0\}.$$

Now write $F_1(x) = F(x)$ and for $n > 1$ write $F_n(x) = F(F_{n-1}(x))$, so $F_n(x)$ is the result of applying F to x n times.

We claim that for all $n \geq 1$, $\mathbf{P}\{Z_n = 0\} = F_n(0)$. When $n = 1$, we have $F_1(0) = F(0) = \mathbf{P}\{X = 0\} = \mathbf{P}\{Z_1 = 0\}$. For larger n, we apply the same inductive technique as in Lemma 1.

$$\begin{aligned}
\mathbf{P}\{Z_n = 0\} &= \sum_{i=0}^{\infty} \mathbf{P}\{Z_n = 0 \mid Z_1 = i\} \mathbf{P}\{Z_1 = i\} \\
&= \sum_{i=0}^{\infty} \mathbf{P}\{Z_{n-1} = 0\}^i \mathbf{P}\{X = i\} \\
&= \sum_{i=0}^{\infty} F_{n-1}(0)^i \mathbf{P}\{X = i\} \\
&= F(F_{n-1}(0)) \\
&= F_n(0).
\end{aligned}$$

We now have

$$p = \lim_{n\to\infty} F_n(0).$$

Since $F_n(0) \to p$ and F is continuous, we also have $F(F_n(0)) \to F(p)$. Xut $F(F_n(0)) \to p$, so we must have $p = F(p)$.

For the second part, suppose q is any other non-negative solution of $F(x) = x$. By differentiation we see that F is non-decreasing and so since $q \geq 0$ we must have $q = F(q) \geq F(0)$. Repeatedly applying F we see that we must have $q \geq F_n(0)$ for all n, and so $q \geq \lim_{n\to\infty} F_n(0) = p$. □

Proof of Fundamental Theorem. We already saw that if $\mathbf{E}[X] < 1$ then extinction is certain, so we assume that $\mathbf{E}[X] \geq 1$. Case (a) is also obvious so we assume that $\mathbf{P}\{X = 1\} < 1$. Note that $F(0) = \mathbf{P}\{X = 0\} \geq 0$ and that $F''(x) > 0$ for all $x > 0$. Also,

$$F'(z) = \Big(\sum_{n=0}^{\infty} \mathbf{P}\{X = n\} z^n\Big)' = \sum_{n=1}^{\infty} n\mathbf{P}\{X = n\} z^{n-1},$$

so $F'(1) = \sum_{n=1}^{\infty} n\mathbf{P}\{X = n\} = \mathbf{E}[X]$. If $\mathbf{E}[X] > 1$ then by continuity there is $x < 1$ such that $F(x) < x$, so by the intermediate value theorem, there is $0 \leq y < x$ with $F(y) = y$, and we must have $p < 1$. On the other hand, if $\mathbf{E}[X] = 1$ then since $\mathbf{P}\{X = 1\} < 1$ there must be $k > 1$ such that $\mathbf{P}\{X = k\} > 0$. It follows that $F''(x) > 0$ for all $x > 0$, so we must have $F(x) > x$ for all $0 \leq x < 1$, and so $p = 1$. □

Exercise 2.3.3. *Let $(Z_n, n \geq 0)$ be the generation sizes in a Bienaymé(μ) process* T. *Let X be μ-distributed and write $\alpha = \mathbf{E}X$ and $\sigma^2 = \mathbf{Var}\{X\}$.*

We suppose in this question that $\sigma^2 \in (0, \infty)$ *and that* $\alpha > 1$*. Also, write* $M_n = Z_n/(\mathbf{E}X)^n$ *and let* M *be the a.s. martingale limit of* M_n*.*

(a) Prove that for every $n \geq 0$*,*

$$\mathbf{E}\left\{ Z_{n+1}^2 \,\middle|\, \mathcal{F}_n \right\} = (\mathbf{E}X)^2 Z_n^2 + \sigma^2 Z_n.$$

(b) Prove that for every $n \geq 0$*,*

$$\mathbf{E}\left[Z_n^2 \right] = \alpha^{2n} + \frac{\sigma^2(\alpha^n - \alpha^{2n})}{\alpha(1-\alpha)}$$

(c) Prove that $M_n \to M$ *in* L_2 *and that* $\mathbf{Var}\{M\} = \frac{\sigma^2}{\alpha(\alpha-1)}$*.*

Exercise 2.3.4. *Fix* $\lambda \in [0, \infty)$*, let* T *be a Poisson*(λ) *Bienaymé tree, and let* $\theta(\lambda) = \mathbf{P}\{|\mathrm{T}| = \infty\}$*.*

(a) Show that $\theta(\lambda)$ *is the largest real solution* x *of* $e^{-\lambda x} = 1 - x$*.*

(b) Show that θ *is continuous and that* θ *is concave and strictly positive on* $(1, \infty)$*.*

(c) Show that for $0 < \lambda \leq 1$*,* $\theta(\lambda) = 0$*, and for* $\lambda \geq 2$*,* $1 - 2e^{-\lambda} \leq \theta(\lambda) \leq 1 - e^{-\lambda}$*.*

(d) Show that $\theta(\lambda)$ *is increasing and* $\lambda(1 - \theta(\lambda))$ *is decreasing.*

(e) Show that $\frac{\mathrm{d}}{\mathrm{d}\lambda}\theta(\lambda) \uparrow 2$ *as* $\lambda \downarrow 1$*. Conclude that* $2\epsilon(1 - o(1)) \leq \theta(1+\epsilon) \leq 2\epsilon$*, the first inequality holding as* $\epsilon \downarrow 0$*.*

2.4 *Conditioning Bienaymé trees to be finite*

Let μ be a probability distribution with support $\mathbb{N}$, and let T be Bienaymé(μ)-distributed. If $\sum_{i \geq 0} i\mu(i) > 1$ then $\mathbf{P}\{|\mathrm{T}| = \infty\} > 0$. What is the distribution of T conditioned to be finite?

This question only makes sense provided that $p_0 = \mu(0) > 0$; under this condition, we may understand the conditional law by thinking about the distribution of the number of children at the root. Writing $q = \mathbf{P}\{|\mathrm{T}| < \infty\}$, then we have

$$\begin{aligned}\mathbf{P}\{c(\varnothing;\mathrm{T}) = j \mid |\mathrm{T}| < \infty\} &= \frac{\mathbf{P}\{c(\varnothing;\mathrm{T}) = j\}}{\mathbf{P}\{|\mathrm{T}| < \infty\}} \mathbf{P}\{|\mathrm{T}| < \infty \mid c(\varnothing;\mathrm{T}) = j\} \\ &= \frac{\mu(j)}{q} \cdot q^j.\end{aligned}$$

The last equality is because if the root has j children, then each of their subtrees must be finite in order for the whole tree to be finite. Write $\hat{\mu}$ for the probability measure on $\mathbb{N}$ given by $\hat{\mu}(j) = \mu(j)q^{j-1}$.

As a consistency check, note that

$$\sum_{j\geq 0}\hat{\mu}(j) = \sum_{j\geq 0}\mu(j)q^{j-1} = q^{-1}F_\mu(q) = q^{-1}q = 1,$$

since q is a fixed point of F_μ; so $\hat{\mu}$ is indeed a probability measure. Moreover, if $\hat{X}$ is $\hat{\mu}$-distributed then

$$\mathbf{E}\hat{X} = \sum_{j\geq 0} j\mu(j)q^{j-1} = F'_\mu(q) < 1\,,$$

the last inequality holding because F_μ is convex and $F_\mu(t) < t$ for $t \in (q,1)$.

Conditionally given that $|\mathrm{T}| < \infty$ and that $c(\varnothing,\hat{\mathrm{T}}) = j$, the children of $\varnothing$ are themselves the roots of Bienaymé(μ) trees, conditioned to be finite, so their offspring distribution is likewise $\hat{\mu}$. Continuing in this manner, we see that *conditionally given that it is finite,* $|\mathrm{T}|$ *is distributed as a Bienaymé*$(\hat{\mu})$ *tree.*

Exercise 2.4.1 (Poisson Bienaymé process duality). *Fix* $\lambda > 1$ *and let* T *be a Poisson*(λ) *Bienaymé tree conditioned to be finite.*

(a) Show that T *is a Poisson*$(\lambda(1-\theta(\lambda)))$ *Bienaymé tree. (Suggestion: use Exercise 2.3.4 (a).)*

(b) Write $\lambda = 1+\epsilon$*. Show that* $\lambda(1-\theta(\lambda)) = 1-\epsilon(1+o(1))$ *as* $\epsilon \to 0$*. (Suggestion: use Exercise 2.3.4 (e).)*

2.5 *Conditioned Bienaymé trees*

Proposition 2.5.1. *Fix a probability distribution* μ *with support* $\mathbb{N}$*. Let* T *be a Bienaymé tree with offspring distribution* μ*, and let* $(D_i, i \geq 0)$ *be independent* μ*-distributed random variables. Then for any positive integer* n*,*

$$\mathbf{P}\left\{|T| = n\right\} = \frac{1}{n}\mathbf{P}\left\{D_1 + \ldots + D_n = n-1\right\}\,.$$

Proof. Let $\mathcal{D}$ be the set of degree sequences of plane trees with n vertices, and for $0 \leq i < n$ let

$$\mathcal{D}_i = \{(d_{i+1},\ldots,d_n,d_1,\ldots,d_{i-1}) : (d_1,\ldots,d_n) \in \mathcal{D}\}.$$

By the cycle lemma, $\mathcal{D}_0,\ldots,\mathcal{D}_{n-1}$ are disjoint and

$$\bigcup_{i=0}^{n-1}\mathcal{D}_i = \{(d_1,\ldots,d_n) \in \mathbb{N}^n : d_1+\ldots+d_n = n-1\}.$$

Now let t be any plane tree with n vertices, and list the vertices of t in lexicographic order as $v_1,\ldots,v_n$, and list their degrees as

$d(1), \ldots, d(n)$. Then since t is finite, taking $h \geq n-1$ in (2.3.1) gives that

$$\begin{aligned}\mathbf{P}\{T = t\} &= \prod_{v \in t} \mu(\deg_t(v)) \\ &= \prod_{i=1}^{n} \mu(d(i)) \\ &= \mathbf{P}\{(D_1, \ldots, D_n) = (d(1), \ldots, d(n))\}\end{aligned}$$

Summing over degree sequences in $\mathcal{D}$, it follows that

$$\begin{aligned}\mathbf{P}\{|T| = n\} = \mathbf{P}\{(D_1, \ldots, D_n) \in \mathcal{D}\} &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{P}\{(D_1, \ldots, D_n) \in \mathcal{D}_i\} \\ &= \frac{1}{n}\mathbf{P}\{D_1 + \ldots + D_n = n-1\},\end{aligned}$$

as required. □

Let $(X_i, i \geq 1)$ be IID with each X_i distributed as $D_1 - 1$, and for $n \geq 0$ let $S_n = 1 + X_1 + \ldots + X_n$. Write $\tau = \inf(m : S_m = 0)$. Then by the cycle lemma,

$$\mathbf{P}\{\tau = n\} = \mathbf{P}\{S_n = 0, S_m > 0 \text{ for } 0 \leq m < n\} = \frac{1}{n}\mathbf{P}\{S_n = 0\},$$

So Proposition 2.5.1 implies that $\mathbf{P}\{|T| = n\} = \mathbf{P}\{\tau = n\}$. A direct way to see this is by thinking of $(S(i), i \geq 0)$ as (an extension of) the depth-first queue process of a Bienaymé tree. If the tree is infinite, then the process goes on forever (and $\tau = \infty$); and if the tree is finite then its size is precisely τ.

Using the above proposition, we can verify the identity (1.4.1) and so show that the Borel(1) distribution is an honest-to-goodness probability distribution. In fact, we may as well show something slightly more general. Fix $\lambda \in [0, 1]$, let $(D_i, i \geq 1)$ be independent Poisson(λ) random variables, and let T be a Bienaymé tree with Poisson(λ) offspring distribution. By the fundamental theorem of branching processes, $\mathbf{P}\{|T| < \infty\} = 1$. Moreover, for all $k \in \mathbb{N}$, we have

$$\begin{aligned}\mathbf{P}\{|T| = k\} &= \frac{1}{k}\mathbf{P}\{D_1 + \ldots + D_k = k-1\} \\ &= \frac{1}{k}\mathbf{P}\{\text{Poisson}(\lambda k) = k-1\} \\ &= \frac{1}{k}\frac{e^{-\lambda k}(\lambda k)^{k-1}}{(k-1)!}.\end{aligned}$$

Summing over $k \geq 1$ yields that

$$1 = \mathbf{P}\{|T| < \infty\} = \sum_{k \geq 1} \frac{e^{-\lambda k}(\lambda k)^{k-1}}{k!}.$$

The case $\lambda = 1$ of this identity is (1.4.1).

More generally, the Borel(λ) distribution is the probability distribution μ with $\mu(k) = \frac{e^{-\lambda k}(\lambda k)^{k-1}}{k!}$, which describes the law of the total progeny in a Poisson(λ) Bienaymé process.

The next exercise extends Proposition 2.5.1 to forests; you can prove it by considering the depth-first queue process of a forest, which is obtained by concatenating the depth-first queue processes of its constituent trees.

Exercise 2.5.1. *Let μ be a probability distribution with support $\mathbb{N}$, and let $(T_i, i \geq 1)$ be independent Bienaymé(μ) trees. Prove that for any $1 \leq r \leq n$,*

$$\mathbf{P}\left\{|T_1| + \ldots + |T_n| = r\right\} = \frac{r}{n}\mathbf{P}\left\{D_1 + \ldots + D_n = n - r\right\} ,$$

where $(D_1, i \geq 1)$ are independent μ-distributed random variables.

In place of the cycle lemma, one may use the following lemma to prove Proposition 2.5.1. This lemma is weaker than the cycle lemma, but the proof is perhaps more straightforward.

The below proof of Lemma 2.5.2 is from Remco van der Hofstad and Michael Keane, *An Elementary Proof of the Hitting Time Theorem*, American Mathematical Monthly **115**(8): 753-756, 2008. Link.

Exercise: prove this lemma from the cycle lemma!

Exercise 2: Adapt this proof to use only exchangeability of $(X_i, i \geq 1)$ rather than assuming that $(X_i, i \geq 1)$ are IID.

Lemma 2.5.2. *Let $(X_i, i \geq 1)$ be IID random variables taking values in $\{-1, 0, 1, \ldots\}$. Fix $k \geq 0$ and for $n \geq 0$ let $S_n = k + X_1 + \ldots + X_n$. Then for all $n \geq 1$, writing $\tau_0 = \inf(m \in \mathbb{N} : S_m = 0)$,*

$$\mathbf{P}\left\{\tau_0 = n\right\} = \frac{k}{n}\mathbf{P}\left\{S_n = 0\right\}$$

Proof. Write $\mathbf{P}_k\left\{\cdot\right\}$ as shorthand for the measure under which $S_0 = k$. We prove the theorem by induction on n. When $n = k = 1$ the theorem is obvious; in that case

$$\mathbf{P}_1\left\{\tau_0 = 1\right\} = \mathbf{P}\left\{X_1 = -1\right\} = \mathbf{P}_1\left\{S_n = 0\right\}.$$

Also note that for all n, the case $k = 0$ is obvious as in that case both sides of the equation equal zero.

Now fix $n > 1$ and $0 < k \leq n$. Condition on the value of X_1 and use the Markov property: given that $X_1 = i$, the remaining steps look like a random walk started from position $k + i$. After the first step, we have $n - 1$ more steps to get to time n, so for $k \geq 1$,

$$\mathbf{P}_k\left\{\tau_0 = n\right\} = \sum_{i=-1}^{\infty} \mathbf{P}_{k+i}\left\{\tau_0 = n - 1\right\}\mathbf{P}\left\{X_1 = i\right\}.$$

Since $k > 0$, $n > 1$, and $i \geq -1$, we have $k + i \geq 0$ and $n - 1 \geq 1$, so we can apply induction to get

$$\mathbf{P}_k\left\{\tau_0 = n\right\} = \frac{1}{n-1}\sum_{i=-1}^{\infty}(k+i)\mathbf{P}_{k+i}\left\{S_{n-1} = 0\right\}\mathbf{P}\left\{X_1 = i\right\}.$$

Now we use Bayes' formula to obtain

$$\mathbf{P}_{k+i}\left\{S_{n-1} = 0\right\}\mathbf{P}\left\{X_1 = i\right\} = \mathbf{P}_k\left\{X_1 = i, S_n = 0\right\} = \mathbf{P}_k\left\{X_1 = i \mid S_n = 0\right\}\mathbf{P}_k\left\{S_n = 0\right\}.$$

Using this equality in the previous displayed equation, we see that

$$\begin{aligned}\mathbf{P}_k\{\tau_0 = n\} &= \frac{1}{n-1}\mathbf{P}_k\{S_n = 0\}\sum_{i=-1}^{\infty}(k+i)\mathbf{P}_k\{X_1 = i \mid S_n = 0\} \\ &= \frac{1}{n-1}\mathbf{P}_k\{S_n = 0\}(k + \mathbf{E}_k\{X_1 \mid S_n = 0\}).\end{aligned}$$

Given that $S_n = 0$, the average step size from starting from k must be $-k/n$, so since $X_1, \ldots, X_n$ are IID we must have $\mathbf{E}_k\{X_1 \mid S_n = 0\} = -k/n$. Plugging this into the preceding equation we get

$$\begin{aligned}\mathbf{P}_k\{\tau_0 = n\} &= \frac{1}{n-1}\mathbf{P}_k\{S_n = 0\}\left(k - \frac{k}{n}\right) \\ &= \frac{k}{n}\mathbf{P}_k\{S_n = 0\}.\end{aligned}$$

□

We finish the section with an exercise connecting conditioned Poisson Bienaymé processes to uniformly random trees, and another, challenging exercise, concerning the probability of observing a given (large) size for a critical Bienaymé tree.

Exercise 2.5.2. *Fix $\lambda > 0$ and let T be a Bienaymé tree with Poisson(λ) offspring distribution.*

(a) Show that for any rooted plane tree t with n vertices, for any $\lambda > 0$,

$$\mathbf{P}\{\mathrm{T} = \mathrm{t} \mid |\mathrm{T}| = n\} = \frac{1}{n^{n-1}}\frac{n!}{\prod_{v\in \mathrm{t})} c(v;\mathrm{t})!};$$

note that this formula does not depend on λ.

(b) Conditionally given that $|\mathrm{T}| = n$, let T' be the rooted tree obtained from T by labeling the vertices of T uniformly at random with labels from $[n]$ and ignoring the plane structure. Prove that $\mathrm{T} \in_u \mathscr{T}_n$.

(c) Show that if $\lambda = 1$ then $\mathbf{P}\{|\mathrm{T}| = n\} \sim (2\pi n^3)^{-1/2}$. (Suggestion: Stirling's formula.)

(d) Fix $\epsilon \in (0,1)$ and let $\lambda = 1 - \epsilon$. Using the formula

$$\mathbf{P}\{|\mathrm{T}| = n\} = \frac{n^{n-1}}{e^n n!}\frac{(e^{\epsilon}(1-\epsilon))^n}{1-\epsilon}$$

show that[3]

$$\mathbf{P}\{|\mathrm{T}| = n\} \geq \frac{n^{n-1}}{e^n n!}(1-\epsilon^2)^n,$$

that

$$\mathbf{P}\{|\mathrm{T}| = n\} \leq \frac{n^{n-1}}{e^n n!}\left(1 - \frac{\epsilon^2}{3}\right)^n,$$

[3] By Stirling's formula, the lower bound is $(1 + o(1))(2\pi n^3)^{-1/2}(1-\epsilon)^n$ as $n \to \infty$.

and that

$$\mathbf{P}\{|\mathrm{T}| = n\} = \frac{n^{n-1}}{e^n n!}\Big(1 - (1+o(1))\frac{\epsilon^2}{2}\Big)^n,$$

the final asymptotic holding as $\epsilon \downarrow 0$, uniformly in n.

Exercise 2.5.3. *Fix an offspring distribution μ and write $\alpha = \sum_{i\geq 0} i\mu(i)$. Let T be a B_μ-distributed random tree. Assume that* $\gcd(i > 0 : \mu(i) > 0) = 1$.

(a) Show that $\mathbf{P}\{|T| = n\} > 0$ for all n sufficiently large.

(b) **(Harder.)** *Show that if $\alpha = 1$ then*

$$\liminf_{n\to\infty} \frac{\log \mathbf{P}\{|T| = n\}}{n} = 0\,.$$

2.6 Branching processes with immigration

These are very natural extensions of branching processes where at each generation a random number of individuals "immigrate", joining the current population. The generation size process $(U_n)_{n\geq 0}$ of a branching process with offspring distribution μ and immigration distribution ν may be constructed as follows. Let $(X_{n,k}, n, k \geq 1)$ be IID with law μ, and independently let $(Y_n, n \geq 0)$ be IID with law ν. Then set $U_0 = Y_0$ and, for $n \geq 0$ let $U_{n+1} = Y_{n+1} + X_{n,1} + \ldots + X_{n,U_n}$. Note that this construction makes perfect sense with $(Y_n, n \geq 0)$ replaced by a deterministic vector $y = (y_n, n \geq 0)$ of non-negative integers; this will be useful below.

The next theorem characterizes when immigration leads to super-exponential population growth.

Theorem 2.6.1 (Seneta, 1970)**.** *Let Y have law ν. If $\mathbf{E}[\max(\log Y, 0)] < \infty$ then $\lim_{n\to\infty} U_n/\alpha^n$ exists and is almost surely finite. If $\mathbf{E}[\max(\log Y, 0)] = \infty$ then $\lim_{n\to\infty} U_n/c^n$ is almost surely infinite for all $c > 0$.*

Lemma 2.6.2. *Let $(R_n, n \geq 1)$ be IID and non-negative.*

(a) If $\mathbf{E}R_1 < \infty$ then almost surely $\limsup_{n\to\infty} \frac{R_n}{n} = 0$ and $\sum_{n\geq 1} e^{R_n} c^n < \infty$ for all $c \in (0, 1)$.

(b) If $\mathbf{E}R_1 = \infty$ then almost surely $\limsup_{n\to\infty} \frac{R_n}{n} = \infty$ and $\sum_{n\geq 1} e^{R_n} c^n = \infty$ for all $c \in (0, 1)$.

Proof. Suppose $\mathbf{E}R_1 < \infty$ and fix any $\epsilon > 0$. Then

$$\sum_{n>0} \mathbf{P}\{R_n \geq \epsilon n\} = \sum_{n>0} \mathbf{P}\{R_1 \geq \epsilon n\} \leq \frac{1}{\epsilon} \sum_{n\geq 0} \mathbf{P}\{R_1 \geq n\} = \frac{\mathbf{E}R_1}{n} < \infty,$$

so by the first Borel-Cantelli lemma, $\limsup_{n\to\infty} R_n/n \leq \epsilon$ almost surely, and since $\log(1-a) < -a$ for $a \in (0, 1)$, letting $N_0 = \sup(n :$

$R_n \geq \epsilon n$), which is almost surely finite, for all $c \in (0, 1-2\epsilon)$ we have

$$\begin{aligned}\sum_{n\geq 1} e^{R_n} c^n &< \sum_{n\geq 1} e^{R_n + n\log(1-2\epsilon)} \\ &\leq \sum_{n\geq 1} e^{R_n - 2\epsilon n} \\ &\leq \sum_{n\leq N_0} e^{R_n - 2\epsilon n} + \sum_{n > N_0} e^{-\epsilon n} \\ &< \infty\,.\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, the first result follows.

Next suppose $\mathbf{E}R_1 = \infty$ and fix any $C > 1$. Then

$$\sum_{n>0} \mathbf{P}\left\{R_n \geq Cn\right\} = \sum_{n>0} \mathbf{P}\left\{R_1 \geq Cn\right\} \geq \frac{1}{C}\sum_{n\geq C} \mathbf{P}\left\{R_1 \geq n\right\} \geq \frac{\mathbf{E}R_1 - C}{C} = \infty,$$

so by the second Borel-Cantelli lemma, almost surely $R_n/n \geq C$ infinitely often. It follows that almost surely $\limsup_{n\to\infty} R_n/n \geq C$, and for any $c > 1/C$,

$$\sum_{n>0} e^{R_n} c^n \geq \sup_{n>0} e^{R_n} c^n \geq \sup_{n>0} (Cc)^n = \infty.$$

Since $C > 1$ was arbitrary, the second result follows. ☐

Proof of Theorem 2.6.1. First suppose that $\mathbf{E}\left[\max(\log Y_1, 0)\right] = \infty$. Then for all $c > 0$,

$$\limsup_n \frac{U_n}{c^n} \geq \limsup_n Y_n c^n = \infty,$$

the last inequality holding almost surely by Lemma 2.6.2.

Next suppose that $\mathbf{E}\left[\max(\log Y_1, 0)\right] < \infty$. Let $U_{n,k}$ be the number of generation-n descendants of generation-k immigrants. Conditionally given Y_k, $U_{n,k}$ is just distributed as the number of individuals in generation $n-k$ of a branching process started with Y_k individuals. Moreover, $U_{n,k}$ is independent of $(Y_j, j \neq k)$, so if $\mathcal{G} := \sigma(Y_n, n \geq 1)$ then

$$\mathbf{E}\left\{U_{n,k} \;\middle|\; \mathcal{G}\right\} = \mathbf{E}\left\{U_{n,k} \;\middle|\; Y_k\right\} = Y_k \alpha^{n-k}.$$

Since $U_n = \sum_{k=0}^{n} U_{n,k}$ it follows that

$$\begin{aligned}\mathbf{E}\left\{\alpha^{-n} U_n \;\middle|\; \mathcal{G}\right\} &= \sum_{k\leq n} \mathbf{E}\left\{\alpha^{-n} U_{n,k} \;\middle|\; \mathcal{G}\right\} \\ &= \sum_{k\leq n} \frac{Y_k}{\alpha^k}\,. \end{aligned} \tag{2.6.1}$$

Lemma 2.6.2 gives that $\sum_{k\leq n} \frac{Y_k}{\alpha^k} \to \sum_{n\leq 0} \frac{Y_n}{\alpha^n} < \infty$ almost surely, so by the conditional Fatou lemma, almost surely

$$\mathbf{E}\left\{\liminf_{n\to\infty} \alpha^{-n} U_n \;\middle|\; \mathcal{G}\right\} \leq \liminf_{n\to\infty} \mathbf{E}\left\{\alpha^{-n} U_n \;\middle|\; \mathcal{G}\right\} = \sum_{n\leq 0} \frac{Y_n}{\alpha^n} < \infty.$$

Thus, $\mathbf{P}\{\liminf_{n\to\infty} \alpha^{-n}U_n = \infty \mid \mathcal{G}\} = 0$ almost surely. But then

$$\mathbf{P}\left\{\liminf_{n\to\infty} \alpha^{-n}U_n = \infty\right\} = \mathbf{E}\left[\mathbf{P}\left\{\liminf_{n\to\infty} \alpha^{-n}U_n = \infty \mid \mathcal{G}\right\}\right] = 0,$$

so almost surely

$$\liminf_{n\to\infty} \alpha^{-n}U_n < \infty.$$

It remains to show that $\alpha^{-n}U_n$ converges almost surely. For this we use the submartingale convergence theorem, which states that a submartingale which is bounded in expectation converges almost surely. We have

$$\mathbf{E}\{U_{n+1} \mid U_1, \ldots, U_n, \mathcal{G}\} = \alpha U_n + Y_{n+1},$$

so $(\alpha^{-n}U_n, n \geq 0)$ is a submartingale with respect to its natural filtration given $\mathcal{G}$; the fact that it is bounded in expectation (given $\mathcal{G}$) follows from (2.6.1). □

There is a nice construction of branching processes with immigration within the Ulam-Harris tree. A *spinal* tree is a pair (t, p), where $\mathrm{t} \in \mathcal{T}$ and $p = p_0, p_1, \ldots$ is a finite or infinite path in t, starting from the root. We write $p_{\leq n}$ for the truncation of p at level n, so if p has at most n nodes then $p = p_{\leq n}$, and otherwise $p_{\leq n} = p_0, p_1 \ldots p_n$.

Let $X = (X_v, v \in \mathcal{U})$ be IID with distribution μ. Fix a vector $y = (y_n, n > 0)$ of non-negative integers, and another vector $i = (i_n, n > 0)$ of integers with $1 \leq i_n \leq y_n + 1$ for all $n > 0$. Let $P_n = P_n(i) := i_1 i_2 \ldots i_n$, so $P_{n+1} = P_n i_{n+1}$ for $n \geq 0$, and let $P = P_0, P_1, \ldots$. Then define a random tree $T = T(X, y, i)$ containing $P(i)$, as follows.

1. Let $\emptyset \in T$ and let $p_0 = \emptyset$.
2. For $n \geq 0$, given $T_{\leq n}$:
   - Let $c(P_n; T) = y_{n+1} + 1$. (Note that $i_{n+1} \leq c(P_n; T)$ so $P_{n+1} \in T_{n+1}$.)
   - For $v \in T_n$ with $v \neq P_n$, let $c(v; T) = X_v$.

Exercise 2.6.1. *The process $(|T_{n+1}| - 1, n \geq 0)$ is distributed as a branching process with immigration with offspring distribution μ and immigration vector y.*

We next introduce a sigma-field on spinal trees, much the same as we did for the set of trees $\mathcal{T}$. The set of spinal trees is

$$\mathcal{T}^* = \{(\mathrm{t}, p) : \mathrm{t} \in \mathcal{T}, p \text{ a path in t starting at the root}\}.$$

For each $n \geq 0$, for each pair (t, v) where $\mathrm{t} \in \mathcal{T}$ and $v \in \mathrm{t}_n$, we define an equivalence class

$$[(\mathrm{t}, v)]_{\leq n} = \{(\mathrm{t}', p') \in \mathcal{T}^* : \mathrm{t}'_{\leq n} = \mathrm{t}_{\leq n}, p'_n \text{ passes through } v\}.$$

Let $\mathcal{F}_n^* = \sigma(\bigcup_{m=0}^n \{[(\mathrm{t}, p)]_{\le m} : (\mathrm{t}, p) \in \mathcal{T}^*\})$, and let $\mathcal{F}^* = \sigma(\bigcup_{n\ge 0} \mathcal{F}_n^*)$. The reason the definition of $\mathcal{F}_n^*$ has a union over $m \le n$ is that we allow for finite paths, which may end at some level $m \le n$. Again, $(\mathcal{F}_n^*, n \ge 0)$ is a filtration, and it is easy to see that $\mathcal{F}_n^*$ refines $\mathcal{F}_n$ for each n.

2.7 *The Kesten-Stigum theorem*

Recall that $M_n = Z_n/\alpha^n$, and that $M = \limsup_{n\to\infty} M_n$ is the a.s. martingale limit of M_n. The goal of this section is to prove the *Kesten-Stigum* theorem, which provides necessary and sufficient conditions for M_n to converge to M in L_1.

Theorem 2.7.1 (Kesten-Stigum Theorem)**.** *Fix an offspring distribution μ with $\alpha = \sum_{i\ge 1} i\mu(i) > 1$, Let T be B_μ-distributed, and let M_n and M be defined as above. Then the following are equivalent.*

(i) $\mathbf{P}\{M = 0\} = \mathbf{P}\{|T| < \infty\}$

(ii) $\mathbf{E}M = 1$

(ii) $\sum_{i\ge 1} \mu(i) \cdot i \log i < \infty$.

Remarks.

- Note that if ω is such that $|T(\omega)| < \infty$ then $M_n(\omega) = 0$ for all n large, so $M(\omega) = 0$. It follows that $\mathbf{P}\{M = 0\} \ge \mathbf{P}\{|T| < \infty\}$.
- By the uniformly integrable martingale convergence theorem, $\mathbf{E}M_n \to \mathbf{E}M$ if and only if (M_n) is uniformly integrable (in which case $M_n \xrightarrow{L_1} M$), so a fourth equivalent condition which can be added to the Kesten-Stigum theorem is that (M_n) is UI.

To prove the Kesten-Stigum theorem we use a beautiful method called a "spinal change of measure". The *size-biasing* $\hat{\mu}$ of μ is the probability distribution with $\hat{\mu}(i) = i\mu(i)/\alpha$. Note that if X has law $\hat{\mu}$ then $\mathbf{P}\{X \ge 1\} = 1$.

Let ν be the probability measure on $\mathbb{Z}^+$ defined by setting $\nu(i) = \hat{\mu}(i-1)$ for all i. Let $X = (X_v, v \in \mathcal{U})$ are independent with law μ, let $Y = (Y_n, n > 0)$ be independent with law ν, and let $U = (U_n, n > 0)$ be independent Uniform$[0,1]$ random variables, with X, Y and U mutually independent. For $n > 0$ let $I_n = \lceil (Y_n + 1)U_n \rceil$, so that I_n is a uniformly random element of $\{1, \ldots, Y_n + 1\}$. Then write BPI_μ^* for the law of the pair $(T, P) = (T(X, Y, I), P(I))$, and let BPI_μ be the law of the tree $T = T(X, Y, I)$ obtained from (T, P) by "ignoring the spine".

Proposition 2.7.2. *For any offspring distribution μ with $\mu(0) < 1$, and any spinal tree (t, p), for all $n \ge 0$,*

$$\mathrm{BPI}_\mu^*(t_{\le n}, p_{\le n}) = \frac{1}{\alpha^n} \mathrm{B}_\mu(t_{\le n}).$$

Proof. Let (T, P) be constructed as above, so that

$$\mathrm{BPI}^*_\mu(\mathrm{t}_{\le n}, p_{\le n}) = \mathbf{P}\left\{(T_{\le n}, P_{\le n}) = (\mathrm{t}_{\le n}, p_{\le n})\right\}.$$

Then write

$$\mathbf{P}\left\{(T_{\le n}, P_{\le n}) = (\mathrm{t}_{\le n}, p_{\le n})\right\} = \prod_{i=0}^{n-1} \mathbf{P}\left\{T_{i+1} = \mathrm{t}_{i+1}, P_{i+1} = p_{i+1} \mid (T_{\le i}, P_{\le i}) = (\mathrm{t}_{\le i}, p_{\le i})\right\}.$$

Now, given that $(\mathrm{t}_{\le i}, p_{\le i})$, in order to have $T_{i+1} = \mathrm{t}_{i+1}$ and $P_{i+1} = p_{i+1}$, the following must occur: p_i must have the right number of children; the correct extension of $p_{\le i}$ must be chosen; and all the other nodes in t_i must also have the right number of children. The probability of all these occurring is

$$\begin{aligned}
&\mathbf{P}\left\{T_{i+1} = \mathrm{t}_{i+1}, P_{i+1} = p_{i+1} \mid (T_{\le i}, P_{\le i}) = (\mathrm{t}_{\le i}, p_{\le i})\right\} \\
&= \hat{\mu}(c(p_i;\mathrm{t})) \cdot \frac{1}{c(p_i;\mathrm{t})} \cdot \prod_{v \in \mathrm{t}_i, v \ne p_i} \mu(c(v;\mathrm{t})) \\
&= \frac{c(p_i;\mathrm{t})\mu(c(p_i;\mathrm{t}))}{\alpha} \cdot \frac{1}{c(p_i;\mathrm{t})} \cdot \prod_{v \in \mathrm{t}_i, v \ne p_i} \mu(c(v;\mathrm{t})) \\
&= \frac{1}{\alpha} \prod_{v \in \mathrm{t}_i} \mu(c(v;\mathrm{t})),
\end{aligned}$$

which combined with the two previous equations gives

$$\mathrm{BPI}^*_\mu(\mathrm{t}_{\le n}, p_{\le n}) = \prod_{i=0}^{n-1} \left(\frac{1}{\alpha} \prod_{v \in \mathrm{t}_i} \mu(c(v;\mathrm{t})) \right) = \frac{1}{\alpha^n} \mathrm{B}_\mu(\mathrm{t}_{\le n}). \qquad \square$$

Corollary 2.7.3. *For all $n \ge 0$,*

$$\frac{d\widehat{\mathrm{B}}_\mu|_{\mathcal{F}_n}}{d\mathrm{B}_\mu|_{\mathcal{F}_n}} = M_n.$$

Proof. For any subtree t of $\mathcal{U}$, by definition,

$$\mathrm{BPI}_\mu(\mathrm{t}_{\le n}) = \sum_p \mathrm{BPI}^*_\mu(\mathrm{t}_{\le n}, p),$$

where the sum is over paths p from the root to generation n in $\mathrm{t}_{\le n}$. But the number of such paths is just $|\mathrm{t}_n|$. Using Proposition 2.7.2 and the fact that $M_n(\mathrm{t}) = |\mathrm{t}_n|/\alpha^n$, we thus have

$$\mathrm{BPI}_\mu(\mathrm{t}_{\le n}) = \frac{|\mathrm{t}_n|}{\alpha^n} \mathrm{B}_\mu(\mathrm{t}_{\le n}) = M_n(\mathrm{t}) \cdot \mathrm{B}_\mu(\mathrm{t}_{\le n}).$$

and the result follows. $\square$

Before proving the Kesten-Stigum theorem, we need one further lemma.

Lemma 2.7.4. *Either* $\mathbf{P}\{M = 0\} = \mathbf{P}\{|T| < \infty\}$ *or* $\mathbf{P}\{M = 0\} = 1$.

Proof. If $i \in T_1$ then the subtree of T rooted at i is itself a B_μ-branching process. Writing

$$M_n^{(i)} = \frac{1}{\alpha^{n-1}} \#\{v \in T_n : 1 \text{ is an ancestor of } v\},$$

then $M_n^{(i)}$ is a martingale; writing $M^{(i)}$ for its almost sure limit, we may decompose M as

$$M = \frac{1}{\alpha}\left(M_n^{(1)} + \ldots + M_n^{(X_\emptyset)}\right).$$

Conditionally given that $X = k$, the limits $M_n^{(1)}, \ldots, M_n^{(k)}$ are independent copies of M, and $M = 0$ if and only if each of $M_n^{(1)}, \ldots, M_n^{(k)}$ equals zero. Thus

$$p := \mathbf{P}\{M = 0\} = \sum_{k \geq 0} \mathbf{P}\{X = k\} \mathbf{P}\{M = 0\}^k = \mathbf{E}\left[p^X\right].$$

The only roots the equation $s = \mathbf{E}\left[s^X\right]$ are $\mathbf{P}\{|T| < \infty\}$ and 1, so the lemma follows. $\square$

Proof of Theorem 2.7.1. Let X have law μ, let Y have law ν where $\nu(i) = \hat{\mu}(i+1)$, and let $L = \log(Y+1)$. It is easy to verify that $\mathbf{E}L < \infty$ if and only if $\mathbf{E}\left[\log^+ Y\right] < \infty$, and

$$\mathbf{E}\left[X \log^+ X\right] = \sum_{i>0} (i \log i)\mu(i) = \sum_{i>0} \log(i)\hat{\mu}(i) = \mathbf{E}L,$$

so by Theorem 2.6.1 $\mathrm{BPI}_\mu(M < \infty) = 1$ if and only if $\mathbf{E}L < \infty$, i.e. if and only if $\mathbf{E}\left[X \log^+ X\right] < \infty$.

We now use that

$$M = \limsup_n M_n = \limsup_n \frac{d\widehat{\mathrm{B}}_\mu|_{\mathcal{F}_n}}{d\mathrm{B}_\mu|_{\mathcal{F}_n}}.$$

by Corollary 2.7.3. Since

$$\mathbf{E}M = \int M(\mathrm{t})\mathrm{B}_\mu(d\mathrm{t}) = \mathrm{B}_\mu(M),$$

It follows by Theorem 4.6.12 that $\mathbf{E}M = 1$ if and only if $\mathrm{BPI}_\mu(M < \infty) = 1$, which occurs if and only if $\mathbf{E}\left[X \log^+ X\right] < \infty$.

Now, if $\mathbf{E}M = 1$ we must have $\mathbf{P}\{M = 0\} < 1$, in which case $\mathbf{P}\{M = 0\} = \mathbf{P}\{|T| < \infty\}$ by Lemma 2.7.4.

Finally, if $\mathbf{E}\left[X \log^+ X\right] = \infty$ then $\mathrm{BPI}_\mu(M = \infty) = 1$, and by Theorem 4.6.12 this implies that $\mathrm{B}_\mu(M = 0) = 1$, or in other words, that $\mathbf{P}\{M = 0\} = 1$; we then have $\mathbf{E}M = 0 < 1$. $\square$

Exercise 2.7.1. *Fix an offspring distribution μ and write $\alpha = \sum_{i \geq 0} i\mu(i)$. Let T be a B_μ-distributed random tree. For $k, n \geq 0$ let $C_n(k) = |\{v \in T_n : v \text{ has } k \text{ children in } T\}|$.*

Prove that for any $k \geq 0$, as $n \to \infty$,

$$\frac{C_n(k)}{|T_n|}\mathbf{1}_{[M\neq 0]} \xrightarrow{\text{a.s.}} \mu(k)\mathbf{1}_{[M\neq 0]}\,, \quad \textit{and} \quad \frac{C_n(k)}{\alpha^n}\mathbf{1}_{[M\neq 0]} \xrightarrow{\text{a.s.}} M\mu(k)\,,$$

where $M := \limsup M_n$ is the almost sure limit of the martingale $M_n = |T_n|/\alpha^n$. Interpret $\frac{C_n(k)}{|T_n|}$ as zero if $|T_n| = 0$.

3
Random graphs

3.1 The Erdős-Rényi process

We begin with the classical Erdős-Rényi processes. Write K_n for the *complete graph*, i.e. the graph with vertices $[n]$ and edges $(\{i,j\}, 1 \le i < j \le n)$.

The Erdős-Rényi process (discrete time). Choose a uniformly random permutation $e_1, \ldots, e_{\binom{n}{2}}$ of $\mathrm{e}(K_n)$. For $0 \le m \le \binom{n}{2}$, let $G_m^{(n)}$ have vertices $[n]$ and edges $\{e_1, \ldots, e_m\}$.

The Erdős-Rényi process (continuous time). Let $(U_e, e \in \mathrm{e}(K_n))$ be independent Uniform$[0,1]$ random variables. For $p \in [0,1]$ let $\mathcal{G}(n,p)$ have vertices $[n]$ and edges $\{e \in \mathrm{e}(K_n) : U_e \le p\}$.

Note that in the continuous time process since the random variables $(U_e, e \in \mathrm{e}(K_n))$ are exchangeable, the ordering $(e_i, 1 \le i \le \binom{n}{2})$ of the edges of K_n in increasing order of U_e-value is a uniformly random permutation. Thus, writing $p_i = U_{e_i}$, the sequence of graphs $(\mathcal{G}(n, p_m), 0 \le m \le \binom{n}{2})$ has the same distribution as the discrete time Erdős-Rényi process. The next exercise contains a closely-related fact.

Exercise 3.1.1. *Prove that for any $p \in (0,1)$ and $0 \le m \le \binom{n}{2}$, given that $|\mathrm{e}(\mathcal{G}(n,p))| = m$, the conditional distribution of $\mathcal{G}(n,p)$ is the same as that of $\mathcal{G}_m^{(n)}$.*

A fair amount of this section is devoted to studying how the component sizes and structures evolve over the course of the Erdős-Rényi process. If we are only interested in component *sizes*, then we might choose to only consider the coalescent at times when the sizes change, or (informally) to simply ignore any edges added by the Erdős-Rényi coalescent that fail to join distinct components. In the discrete time process, we may achieve this as follows. For each $0 \leq m \leq \binom{n}{2}$, let τ_m be the number of edges $e_i = \{U_i, V_i\}$, $0 < i \leq m$ such that U_i and V_i lie in different components of $\mathcal{G}_{i-1}^{(n)}$. (See Figure 3.1 for an example.) Observe that

$$\tau_m + 1 = \begin{cases} \tau_m & \text{if } \mathcal{G}_{m+1}^{(n)} \text{ and } \mathcal{G}_m^{(n)} \text{ have the same number of components} \\ \tau_m + 1 & \text{if } \mathcal{G}_{m+1}^{(n)} \text{ has one fewer component than } \mathcal{G}_m^{(n)}\,. \end{cases}$$

In other words, τ_m increases precisely when the the endpoints of the edge added to $\mathcal{G}_m^{(n)}$ are in different components. Further, the set

$$\{e_m : m \geq 1, \tau_m > \tau_{m-1}\}$$

contains $n-1$ edges, since $\mathcal{G}_0^{(n)}$ has n components and $\mathcal{G}_{\binom{n}{2}}^{(n)}$ almost surely has only one component.

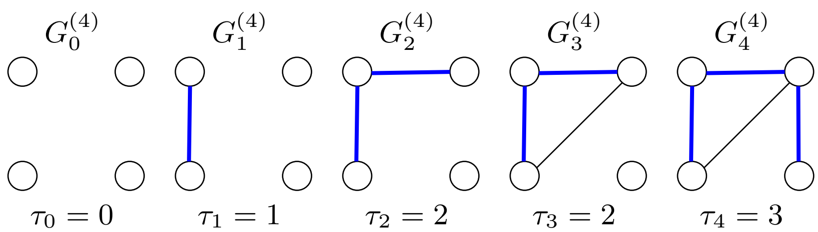


Figure 3.1: An example of the first steps of the Erdős-Rényi coalescent. The edges which join distinct connected components are represented by thicker, blue lines.

Set $I_1 = 0$ and for $1 < k \leq n$ let

$$I_k = \inf\{m \geq 1 : \tau_m = k - 1\}\,.$$

Then for $1 < k \leq n$, the edge e_{I_k} joins distinct components of $\mathcal{G}_{I_k - 1}^{(n)}$, and by symmetry is equally likely to be any such edge.

3.2 *Component sizes when $p < 1/n$.*

For a graph G, and $v \in \mathrm{v}(G)$, let $N(v) = N_G(v)$ be the set of vertices adjacent to v, and let $\mathcal{C}(v) = \mathcal{C}_G(v)$ be the component of G containing v.

Exercise 3.2.1. *(a) Show that in the discrete time Erdős-Rényi process, if for some m, all components of $\mathcal{G}_m^{(n)}$ have size at most s then the probability a uniformly random edge from among the remaining edges has both endpoints in the same component is at most $(s-1)/(n-1)$.*

(b) Show that for all $0 \leq m \leq \binom{n}{2}$, in $\mathcal{G}_m^{(n)}$, $\mathbf{E}|N(v)| = 2m/n$.

(c) Prove by induction that for all $0 \leq m < n/2$, in $\mathcal{G}_m^{(n)}$, $\mathbf{E}|\mathcal{C}(1)| \leq n/(n-2m)$.
*(**Hint.** First condition on $N(1)$, then average.)*

(d) Prove that for all $k \in \mathbb{N}$,

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}_m^{(n)})| \geq k\right\} \leq \frac{n}{k}\mathbf{P}\left\{|\mathcal{C}(1)| \geq k\right\} .$$

(Suggestion. Given that the largest component of $\mathcal{G}_m^{(n)}$ has size s, with probability at least s/n vertex 1 is in such a component.)

Exercise 3.2.2. *This exercise is the continuous-time analogue of the previous one.*

(a) Show that in the continuous-time Erdős-Rényi process, if for some p, all components of $\mathcal{G}(n,p)$ have size at most s then the probability a uniformly random edge from among the remaining edges has both endpoints in the same component is at most $(s-1)/(n-1)$.

(b) Show that for all $p \in (0,1)$, in $\mathcal{G}(n,p)$, $\mathbf{E}|N(v)| = (n-1)p$.

(c) Prove by induction that for all $p \leq 1/(n-1)$, in $\mathcal{G}(n,p)$, $\mathbf{E}|\mathcal{C}(1)| \leq 1/(1-p(n-1))$.
(Suggestion. First condition on $N(1)$, then average.)

(d) Prove that for all $k \in \mathbb{N}$, in $\mathcal{G}(n,p)$,

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}(n,p))| \geq k\right\} \leq \frac{n}{k}\mathbf{P}\left\{|\mathcal{C}(1)| \geq k\right\} .$$

(Suggestion. Given that the largest component of $\mathcal{G}_m^{(n)}$ has size s, with probability at least s/n vertex 1 is in such a component.)

Exercise 3.2.3. *Fix $p \in (0,1)$, $\lambda > 0$ and $n \in \mathbb{N}$. Let B be Binomial(n,p) and P be Poisson(λ).*

(a) For $k \in \mathbb{N}$ let $r(k) = \mathbf{P}\left\{B = k\right\} / \mathbf{P}\left\{P = k\right\}$. Show that $r(k)$ is decreasing in k.

(b) Prove that if $r(0) > 1$ then there is $k^* \in \mathbb{N}$ such that $\mathbf{P}\left\{B = k\right\} \geq \mathbf{P}\left\{P = k\right\}$ for $k \leq k^*$ and $\mathbf{P}\left\{B = k\right\} < \mathbf{P}\left\{P = k\right\}$ for $k > k^*$.

(c) Prove that if $(1-p)^n > e^{-\lambda}$ then $B \preceq_{\mathrm{st}} P$.

(d) Show that $(1-p) > e^{-p/(1-p)}$ for $p \in [0,1)$, and conclude that if $np/(1-p) \leq \lambda$ then $B \preceq_{\mathrm{st}} P$.

For the next exercise, recall that the *total variation distance* between two random variables X and Y is

$$\|X - Y\|_{\mathrm{TV}} := \sup(|\mathbf{P}\left\{X \in S\right\} - \mathbf{P}\left\{Y \in S\right\}| : S \subset \mathbb{R} \text{ Borel}).$$

While not needed for the exercise, we recall that

$$\|X - Y\|_{\mathrm{TV}} = \inf(\mathbf{P}\left\{X' \neq Y'\right\} : (X', Y') \text{ is a coupling of } X \text{ and } Y) .$$

Exercise 3.2.4. *(a) Fix $\epsilon > 0$, let X be Bernoulli(ϵ) and let Y be Poisson(ϵ). Show that $\|X - Y\|_{\mathrm{TV}} \le 2\epsilon^2$.*

(b) Fix $\lambda > 0$ and $n \in \mathbb{N}_+$, let B be Binomial$(n, \lambda/n)$ and let P be Poisson(λ). Show that $\|B - P\|_{\mathrm{TV}} \le 2\lambda^2/n$. (Suggestion: use the fact that total variation distance satisfies the triangle inequality.)

We can use the above exercises to start to understand component sizes in $\mathcal{G}_m^{(n)}$ and in $\mathcal{G}(n,p)$ in more detail. In $\mathcal{G}(n,p)$, by Exercise 3.2.2 parts (c) and (d),

$$\begin{aligned}
\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}(n,p))| \ge k\right\} &\le \frac{n}{k}\mathbf{P}\left\{|\mathcal{C}(1)| \ge k\right\} \\
&\le \frac{n}{k^2}\mathbf{E}|\mathcal{C}(1)| \\
&\le \frac{n}{k^2(1-p(n-1))}\,. \qquad (3.2.1)
\end{aligned}$$

If $p = p(n) = c/n$ with $c \in (0,1)$, this yields that for any function $\omega(n) \to \infty$

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}(n,p))| \ge \omega(n)n^{1/2}\right\} \to 0,$$

so in particular $|\mathcal{C}^{\max}(\mathcal{G}(n,p))|/n \to 0$ in probability. In fact, we can deduce this even for p quite close to $1/n$. Fix any function $\epsilon(n) \to 0$ with $n\epsilon(n) \to \infty$ and let $p = (1-\epsilon(n))/n$. Then $1 - p(n-1) \ge 1 - pn = \epsilon(n)$, so for any $\delta > 0$, the bound (3.2.1) gives

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}(n,p))| \ge \delta n\right\} \le \frac{1}{\delta^2}\frac{1}{n\epsilon(n)} \to 0\,;$$

it again follows that $|\mathcal{C}^{\max}(\mathcal{G}(n,p))|/n \to 0$ in probability.

Exercise 3.2.5. *Show that if $p = p(n)$ is any sequence of values such that $|\mathcal{C}^{\max}(\mathcal{G}(n,p))|/n \to 0$ in probability, then also $\mathbf{E}|\mathcal{C}^{\max}(\mathcal{G}(n,p))|/n \to 0$.*

Though it might seem boring, it's useful to do a third version of this computation, with $p = (1 - \lambda n^{-1/3})/n$ and $\lambda > 0$. For this value of p we have

$$1 - p(n-1) \ge \lambda n^{-1/3}\,,$$

so for any function $\omega(n)$ with $\omega(n) \to \infty$, we have

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}(n,p))| \ge \omega(n)n^{2/3}\right\} \le \frac{n}{(\omega(n)n^{2/3})^2(1-p(n-1))} \le \frac{1}{\omega(n)^2\lambda} \to 0\,;$$

so with high probability the largest component of $\mathcal{G}(n, (1-\lambda n^{-1/3})/n)$ has size $O(n^{2/3})$.

In fact, we can even prove bounds on component sizes when $p = 1/n$ or when p is a little bigger than $1/n$, with a bit of care. Since for any vertex v in $\mathcal{G}(n,p)$, we have $|N(v)| \stackrel{\mathrm{d}}{=} \mathrm{Bin}(n-1,p)$, but neighbourhoods may overlap, it follows that $|\mathcal{C}(v)|$ is stochastically dominated by the size of a Bienaymé process with $\mathrm{Bin}(n-1,p)$ offspring distribution. Using the stochastic relation between Binomial and Poisson random variables given in Exercise 3.2.3, it follows that

$$|\mathcal{C}(v)| \preceq_{\mathrm{st}} |\mathrm{T}|,$$

where $|\mathrm{T}|$ is a Poisson(λ) Bienaymé tree with $\lambda = np/(1-p)$.

The version of the bounds given on this page for $p = (1+\epsilon)/n$ is cleaner and more direct in the hand-written notes than in the typed notes; I haven't had time to update these notes yet, so the handwritten notes are better to look at for this for the time being.

Now suppose $p = (1+\epsilon)/n$ for $\epsilon \in (0,1)$; think of ϵ as close to zero. Then $\lambda = (1+\epsilon)(1+(1+\epsilon)/(n-1-\epsilon)) < 1+\epsilon+2(1+\epsilon)/n$, the second inequality holding for $n \geq 4$.

Therefore,

$$\begin{aligned}\mathbf{P}\{|\mathcal{C}(v)| \geq s\} &\leq \mathbf{P}\{|\mathrm{T}| \geq s\} \\ &\leq \mathbf{P}\{|\mathrm{T}| = \infty\} + \mathbf{P}\{|\mathrm{T}| \geq s \mid |\mathrm{T}| < \infty\}\,.\end{aligned}$$

By Exercise 2.3.4 (e) we know that

$$\mathbf{P}\{|\mathrm{T}| = \infty\} = \theta(\lambda) \leq 2(\lambda - 1) < 2(\epsilon + 2(1+\epsilon)/n)\,.$$

Moreover, conditionally given that $|\mathrm{T}| < \infty$, the tree T is distributed as a Poisson$(\hat{\lambda})$ Bienaymé tree, where $\hat{\lambda} = \lambda(1-\theta(\lambda)) = 1-\epsilon(1+o(1))$ as $\epsilon \to 0$, by Exercise 2.4.1 (b). If ϵ is sufficiently small that $|\hat{\lambda} - (1-\epsilon)| \leq \epsilon/2$, then it follows that

$$\mathbf{P}\{|\mathrm{T}| \geq s \mid |\mathrm{T}| < \infty\} \leq \sum_{m=s}^{\infty} \mathbf{P}\{|\mathrm{T}^*| = m\}\,,$$

where T^* is a Poisson$(1-\epsilon/2)$ Bienaymé tree. Using Exercise 2.5.2 (c), we then have

$$\begin{aligned}\sum_{m=s}^{\infty} \mathbf{P}\{|\mathrm{T}^*| = m\} &\leq \sum_{m=s}^{\infty} \frac{m^{m-1}}{e^m m!}\Big(1 - \frac{\epsilon^2/2}{3}\Big)^m\,, \\ &\leq O(1) \cdot \sum_{m=s}^{\infty} \frac{1}{m^{3/2}} e^{-\epsilon^2 m/6}\,.\end{aligned}$$

When $s = x/\epsilon^2$ for x positive and bounded away from zero, this sum is

$$O(\epsilon) \cdot x^{-1/2} e^{-x/6}.$$

It follows that for any $\delta > 0$ there is x such that for all n sufficiently large, in $\mathcal{G}(n, (1+\epsilon)/n)$,

$$\mathbf{P}\left\{|\mathcal{C}(v)| \geq x/\epsilon^2\right\} \leq (2+\delta)\epsilon\,,$$

and so by Exercise 3.2.2 (d),

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}(n,p))| \geq x/\epsilon^2\right\} \leq (2+\delta)\frac{n\epsilon^3}{x}\,.$$

This bound is not useful when ϵ is too small, but a similar argument does give a bound whatever the value of ϵ. Indeed, conditionally given that $|T| < \infty$ it is distributed as a critical or subcritical

Poisson Bienaymé tree, and for any such Bienaymé tree T^* we have $\mathbf{P}\{|T^*| \geq s\} = O(1)/s^{1/2}$, so

$$\mathbf{P}\{|\mathrm{T}| \geq s \mid |\mathrm{T}| < \infty\} = \frac{O(1)}{s^{1/2}}$$

and we can use this to see that for any sequence $\epsilon(n)$ with $\epsilon(n) \geq 0$ and $(1+\epsilon(n))/n \leq 1$, if $p(n) = (1+\epsilon(n))/n$ then in $\mathcal{G}(n, p(n))$, for all $s \geq 1$,

Fill in more detail here as in handwritten notes

$$\mathbf{P}\{|\mathcal{C}^{\max}| \geq s\} \leq \frac{(2+o(1))\epsilon n}{s} + O(1) \cdot \frac{n}{s^{3/2}}.$$

It follows from this that if $\epsilon(n) = O(n^{1/3})$ then for all $x \geq 1$,

$$\mathbf{P}\left\{|\mathcal{C}^{\max}| \geq xn^{2/3}\right\} = O(1/x),$$

and if $\liminf_{n\to\infty} \epsilon(n)n^{1/3} > 0$ then for all $x \geq 1$,

$$\mathbf{P}\{|\mathcal{C}^{\max}| \geq x\epsilon n\} = O(1/x).$$

Combining these bounds and using the subsubsequence principle yields that for any sequence with $\epsilon(n) \geq 0$ and $(1+\epsilon(n))/n \leq 1$, we have

$$\mathbf{P}\left\{|\mathcal{C}^{\max}| x \cdot \max(\epsilon n, n^{2/3})\right\} = O(1/x).$$

For both $p < 1/n$ and $p \geq 1/n$, our bounds seem to hit a barrier when $\epsilon = O(n^{-1/3})$, and yield an upper bound of the form $|\mathcal{C}^{\max}| = O_P(n^{2/3})$. This is not an artefact of the proof techniques; it corresponds to the actual scaling of the largest component sizes in $\mathcal{G}(n, (1+eps)/n)$ when $\epsilon = O(n^{-1/3})$. [1]

[1] See Luczak, Pittel and Wierman, *The Structure of a Random Graph at the Point of the Phase Transition*, Trans. AMS 341 (2), 1994; Aldous, *Brownian excursions, critical random graphs and the multiplicative coalescent*, Ann. Probab. 25 (2), 1997 for versions of this theorem; Pittel, *On the Largest Component of the Random Graph at a Nearcritical Stage*, J. Comb. Theory B 82, 2001

Theorem 3.2.1. *Fix $c \in \mathbb{R}$ and let $p(n) = (1 + c/n^{1/3})/n$. Then in $\mathcal{G}(n, p(n))$,*

$$\frac{|\mathcal{C}^{\max}|}{n^{2/3}} \xrightarrow{\mathrm{d}} Z(c)$$

for some random variable $Z(c)$ with $\mathbf{E}Z(c) <\in (0,\infty)$.

Exercise 3.2.6. *(a) Show that for $p \leq 1/n$, $(n-1)p/(1-p) \leq np$ and so in $\mathcal{G}(n,p)$ we have $|\mathcal{C}(v)| \leq |T(\mathrm{Poi}(np)|$.*

(b) Show that for all $\delta > 0$ there is $A > 0$ such that for all $\epsilon \in (0,1)$, in $\mathcal{G}(n, (1-\epsilon)/n)$ and all $x > \delta$,

$$\mathbf{P}\left\{|\mathcal{C}(v)| \geq \frac{x}{\epsilon^2}\right\} \leq \frac{A|\epsilon|}{x^{1/2}} e^{-x/2}.$$

I wrote down the proof of (b) hastily so it's possible the statement has minor glitches. If you think you have found a mistake you may be right; but try to prove an upper bound of this form.

(c) Use (b) to show that in $\mathcal{G}(n, (1-\epsilon)/n)$,

$$\mathbf{E}\left[\text{ Number of components of size } \geq \frac{x}{\epsilon^2}\right] = O\left(\frac{n\epsilon^3}{x^{3/2}} e^{-x/2}\right)$$

Suggestion. *Use that the number of vertices in components of size at least k is at least k times the number of such components.*

(d) Use (c) to show that in $\mathcal{G}(n,(1-\epsilon)/n)$,

$$|\mathcal{C}^{\max}| = O_P\left(\frac{2\log(n\epsilon^3)}{\epsilon^2}\right)$$

Exercise 3.2.7. *(a) Suppose p and λ are such that $(n-1)p/(1-p) \le \lambda$. We have seen that in this case, $|\mathcal{C}(v)| \preceq |T|$, where $\mathcal{C}(v)$ is the component of v in $\mathcal{G}(n,p)$ and T is a Poisson(λ) Bienaymé tree. Strengthen this to show that*

$$(|N^k(v)|, k \ge 0) \preceq_{st} (|T_k|, k \ge 0),$$

in that the two sequences can be coupled so that $|N^k(v)| \le |T_k|$ for all $k \ge 0$.

(b) Show that for all n, for any $\epsilon > 0$, with $p = (1-\epsilon)/n$, then in $\mathcal{G}(n,p)$,

$$\mathbf{P}\left\{\max_{w\in\mathcal{C}(v)} \mathrm{dist}(v,w) \ge k\right\} \le (1-\epsilon)^k .$$

(c) (To be developed, but students: it's worth thinking about already!) Show that if $\max_{v\in[n]} \max_{w\in\mathcal{C}(v)} \mathrm{dist}(v,w) \ge k$ then with probability at least k/n vertex 1 satisfies $\max_{w\in\mathcal{C}(1)} \mathrm{dist}(1,w) \ge k/2$. Use this together with part (b) to prove an upper tail bound on $\max_{v\in[n]} \max_{w\in\mathcal{C}(v)} \mathrm{dist}(v,w) \ge k$.

3.3 Explorations of graphs

Before analyzing component sizes further, we need to introduce some more tools: *exploration process* for graphs. Let G be a graph with vertex set $[n]$. We will explore the components of G via a procedure called depth-first search; this is closely linked to the depth-first queue process seen above. In depth-first search, at each step one vertex is "explored": its neighbours are revealed, and the previously undiscovered neighbours are added to the queue for later exploration.

Formally, in the depth-first queue process of graph G with $v(G) = [n]$, at step i the vertex set is partitioned into sets E_i, A_i and U_i, respectively containing *explored*, *active*, and *undiscovered* vertices. We initialize the process by taking $E_0 = \emptyset$, $A_0 = \emptyset$ and $U_0 = [n]$. The process will conclude with $E_n = [n]$ and $A_0 = U_0 = \emptyset$. We define the *priority* of a vertex to be the step at which it leaves the collection of undiscovered vertices, which is $\max(i : v \in U_i)$, so vertices that are activated later have higher priority.[2]

[2] This is what makes the process a depth-first search process; if vertices discovered earlier had higher priority we would instead obtain breadth-first search.

Depth-first queue process for G.

STEP t $(1 \le t \le n)$, going from time $t-1$ to time i:

$\star$ If A_{t-1} is non-empty then let v_t be the vertex of A_t with

> highest priority (in case of a tie, choose the vertex with the smallest label from among top-priority vertices).
>
> $\star$ If A_{t-1} is empty then let v_t be the smallest-labeled vertex of U_{t-1}. $\star$ Let $A_t = (A_{t-1} \setminus v_t) \cup (N(v_t) \cap U_{t-1})$. $\star$ Let $E_t = E_{t-1} \cup \{v_t\} = \{v_1, \ldots, v_t\}$. $\star$ Let $U_t = [n] \setminus (A_t \cup E_t) = (U_{t-1} \setminus v_t) \setminus (N(v_t) \cap U_{t-1})$.

Exercise 3.3.1. *Show that the whole process can be recovered from either* $(A_i, i \in [n])$ *or* $(U_i, i \in [n])$.

At all times $0 \leq t \leq n$, the explored, active and unexplored vertices partition $v(G) = [n]$, so $|A_t| + |E_t| + |U_t| = n$. Since also $|E_t| = t$, it follows that

$$|A_t| + |U_t| = n - t.$$

An example of the depth-first queue process of a graph with four connected components appears in Figure 3.2.

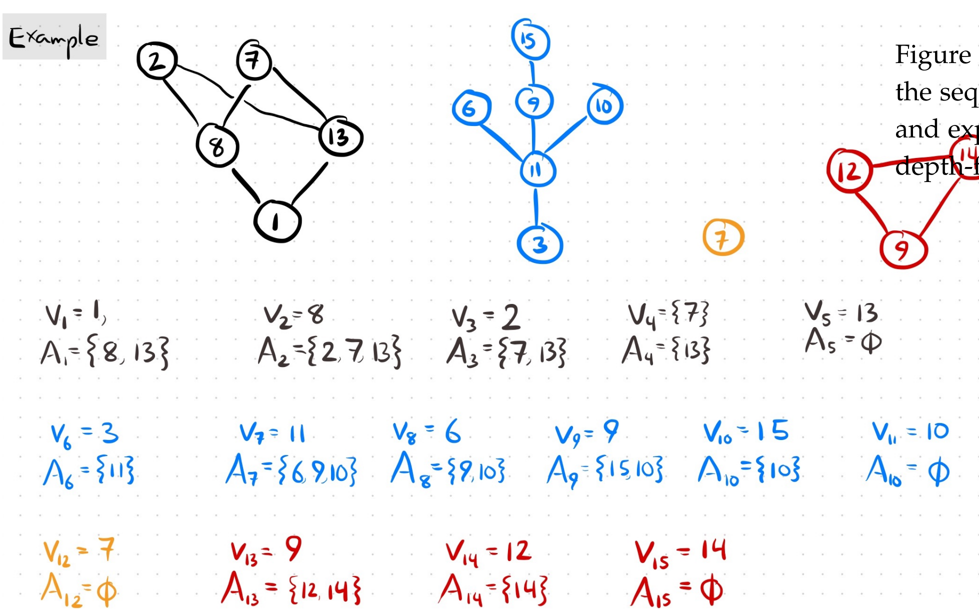


Figure 3.2: A graph G and the sequence of active sets A_t and explored vertices v_t for its depth-first queue process.

A component exploration concludes at step t precisely if A_t is empty, which is to say that after exploring v_t, every vertex which has been discovered has also been explored.

Write $\kappa(t)$ for the number of components discovered by step t, so $\kappa(t) = |\{0 \leq i < t : A_i = \emptyset\}|$. Then $\kappa(0) = 0$, and $\kappa := \kappa(n)$ is the number of components of G.

Let $(T(j), 1 \leq j \leq \kappa)$ be the ordered sequence of times at which a component exploration concludes. Writing $X(i) = X_G(i) = |N(v_i) \cap$

$U_i| - 1$ and $S(i) = \sum_{j=1}^{i} X(j)$, then we may re-express $T(j)$ as

$$T(j) = \min\left(t \geq 0 : \sum_{i=0}^{t} X(i) \leq -j\right) = \min(t \geq 0 : S(t) \leq -j)\,.$$

Setting $T(0) = 0$, the component sizes of G, in the order they are discovered by the depth-first search, are then

$$(T(j+1) - T(j), 0 \leq j < \kappa).$$

The sequence $(S(t), 0 \leq t \leq n)$ corresponding to the DFQP of the graph from Figure 3.2 is shown in Figure 3.4.

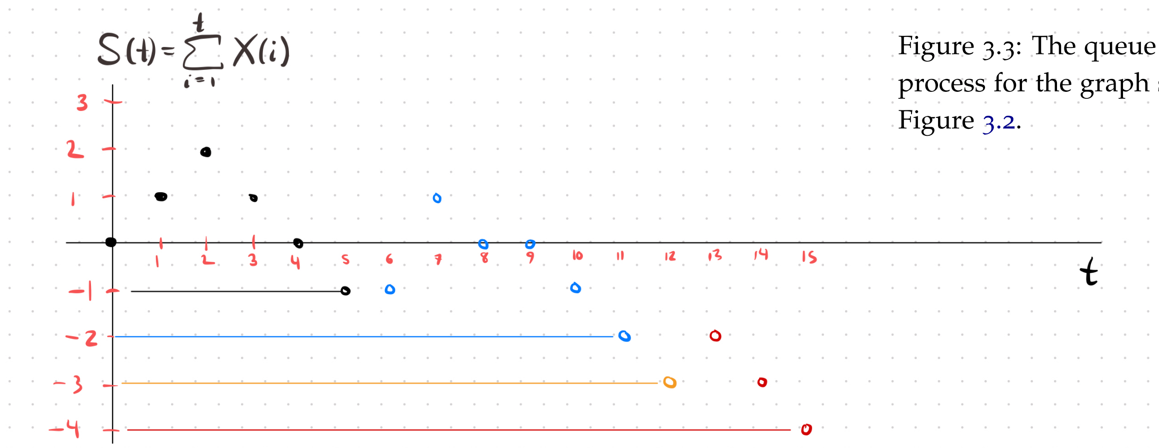


Figure 3.3: The queue length process for the graph shown in Figure 3.2.

Proposition 3.3.1. *For all $t \geq 0$, $|A_t| = S(t) + \kappa(t)$.*

Proof. This is true at time $t = 0$, when all three quantities are 0. For $t > 0$, there are two cases.

1. If $|A_{t-1}| \neq 0$ then $A_t = (A_{t-1} \setminus v_t) \cup (N(v_t) \cap U_{t-1})$ and $\kappa(t) = \kappa(t-1)$, so

$$|A_t| = |A_{t-1}| - 1 + |N(v_t) \cap U_{t-1}| = |A_{t-1}| + X(t) = |A_{t-1}| + X(t) + \kappa(t) - \kappa(t-1),$$

so inductively $|A_t| = S(t) + \kappa(t)$.

2. If $|A_{t-1}| = 0$ then $\kappa(t) = \kappa(t-1) + 1$, and

$$|A_t| = |N(v_t) \cap U_{t-1}| = X(t) + 1 = |A_{t-1}| + X(t) + 1 = |A_{t-1}| + X(t) + \kappa(t) - \kappa(t-1)$$

and again $|A_t| = S(t) + \kappa(t)$ by induction.

$\square$

So far the discussion of depth-first queue process for graphs relates to an arbitrary graph G; we now focus on random graphs. Suppose that $\mathcal{G} = \mathcal{G}(n,p)$; then the sets A_i, E_i and U_i are random. We write $\mathcal{F}_t = \sigma(A_i, E_i, U_i, 0 \le i \le t)$ for the σ-algebra generated by the process up to time t.

It's useful to write $U'_{t-1} = U_{t-1} \setminus \{v_t\}$; then $U'_{t-1} = U_{t-1}$ if $A_{t-1} \neq \emptyset$, since in this case, $v_t \in A_{t-1}$. However, if $A_{t-1} = \emptyset$ then $v_t \in U_{t-1}$ and so $|U'_{t-1}| = |U_{t-1}| - 1$.

Note that, conditionally given $\mathcal{F}_{t-1}$, the set $N(v_t) \cap U_{t-1}$ is a binomial random subset of U'_{t-1}. This means in particular that

$$\mathbf{P}\{|N(v_t) \cap U_{t-1}| = k \mid \mathcal{F}_{t-1}\} = \binom{|U'_{t-1}|}{k} p^k (1-p)^{|U'_{i-1}|-k}. \quad (3.3.1)$$

On the event that $|A_t| > 0$, we obtain that

$$\mathbf{E}\{U_{t+1} \mid \mathcal{F}_t\} = |U_t|(1-p).$$

Heuristically, this suggests that perhaps $|U_t|$ tracks the function $n(1-p)^t$. On the other hand, $|A_t| = n - t - |U_t|$, so this suggests that perhaps

$$|A_t| \approx n - t - n(1-p)^t.$$

This heuristic may not be convincing for general p, in particular when $np < 1$ so that the process S has a downward drift, ignoring the times when A_t is empty seems suspect. But for $np > 1$, so the drift is positive (at least for small times t), the heuristic may seem more reasonable.

When $p = p(n) = o(1)$, we may summarize these heuristics as

$$|U_t| \approx n(1-p)^t \approx ne^{-pt} \quad \text{and } |A_t| \approx n - t - n(1-p)^t \approx n - t - ne^{-pt} =: f_n(t).$$

Exercise 3.3.2. *Show that for any $A > 0$,*

$$\sup_{0 \le p \le A/n} \sup_{0 \le t \le n} |n(1-p)^t - ne^{-pt}| = O(1)$$

We will prove a law of large numbers for $|\mathcal{C}^{\max}|$ in $\mathcal{G}(n,p)$, when $(1 + \omega(n)/n^{1/3})/n \le p = O(1/n)$, by showing that $|U_t|$ indeed closely tracks $f(t) = f_n(t)$ until time $n\theta(np)$.

Theorem 3.3.2. *Fix $A > 0$. Let $\epsilon = \epsilon(n)$ be a non-negative sequence with $\epsilon(n)n^{1/3} \to \infty$ and $\epsilon(n) \le A < \infty$. Take $p = p(n) = (1 + \epsilon(n))/n$ and write $\lambda = \lambda(n) = np = 1 + \epsilon$. Then in $\mathcal{G}(n,p)$,*

$$\frac{|\mathcal{C}^{\max}|}{n\theta(\lambda)} \xrightarrow{\mathrm{P}} 1.$$

We begin the proof by recording some of the important properties of f which we will use in the argument. First, with

$$d = d(n,p) = n\theta(np) = n\theta(\lambda),$$

then d is the unique positive solution to $f(t) = 0$. Second,

$$f'(t) = npe^{-pt} - 1 = p(n - t - f(t)) - 1,$$

so

$$f'(d) = p(n-d-f(d)) - 1 = p(n-n\theta(np) - 1 = \lambda(1-\theta(\lambda)) - 1 = \hat{\lambda} - 1\,,$$

where $\hat{\lambda}$ is the Poisson dual parameter to λ. Third,

$$f''(t) = -np^2 e^{-pt} \geq -\frac{(A+1)^2}{n}.$$

Using these bounds, we can show that $f(t)$ grows reasonably quickly when t is not too large. In what follows we fix a function $\omega(n) \to \infty$ which grows slowly enough that $\omega(n)^2 = o(\epsilon(n)n^{1/3})$, and set

$$g = g(n) = \omega(n)\sqrt{n/\epsilon}.$$

Proposition 3.3.3. *For all $t \leq n\epsilon/(2(A+1)^2)$ we have $f(t) \geq \epsilon t/2$. Moreover, $f(t) \geq \epsilon t/2$ for $t \leq g$, for n large.*

Proof. First, $f'(0) = np - 1 = \epsilon$, so by the above bound on $f''(t)$ we have

$$f'(t) \geq f'(0) - t(A+1)^2/n \geq \epsilon/2$$

the second bound holding for $t \leq n\epsilon/(2(A+1)^2)$. It follows that for $t \leq n\epsilon/(2(A+1)^2)$,

$$f(t) = f(0) + \int_0^t f'(s)\mathrm{d}s \geq t\epsilon/2.$$

The second assertion of the proposition follows from the first, since

$$g = \omega(n)\sqrt{n/\epsilon} = o(n^{2/3}) = o(\epsilon n),$$

where for the first bound we have used that $\omega(n) = o(\epsilon^{1/2}n^{1/2})$ and for the second we have used that $\epsilon n^{1/3} \to \infty$. In particular, this implies that $g \leq n\epsilon/(2(A+1)^2)$ for n large, so the first bound applies at time $t = g$. □

This proposition establishes adequate control on f near time zero. We also need to understand the behaviour of f at times near d, and at intermediate times. The behaviour near time d is the subject of the next proposition. Here and below we write

$$\sigma = \sigma(n) = \sqrt{n \cdot \epsilon(n)}\,;$$

note that $g \cdot \epsilon(n) = \omega(n) \cdot \sigma$.

Proposition 3.3.4. *Let $\sigma = \sqrt{\epsilon n}$. Then with $d^- = d - g$ and $d^+ = d + g$, it holds that $f(d^-) \geq 10\sigma\sqrt{\omega}$ and $f(d^+) \leq -10\sigma\sqrt{\omega}$ for n large.*

Proof. We saw that $f'(d) = \hat{\lambda} - 1$ and that $0 \geq f''(t) \geq -(A+1)^2/n$, so

$$f'(t) \leq \hat{\lambda} - 1 + \frac{g(A+1)^2}{n}$$

whenever $t \in [d^-, d]$. Since $\epsilon \le A < \infty$ and $\lambda = np = 1+\epsilon$, there is some $\delta = \delta(A) > 0$ such that $\hat{\lambda} - 1 \le -2\delta\epsilon$. On the other hand,

$$\frac{g(A+1)^2}{n} = \frac{\omega(n)}{\sqrt{n\epsilon}} = o(\epsilon)\,,$$

the second bound holding by the assumption that $\omega^2(n) = o(n^{1/3}\epsilon)$, so for n sufficiently large, the preceding upper bound on $f'(t)$ is at most $-\delta\epsilon$. It follows that

$$f(d^-) = \int_{d-g}^{d} f'(t)\mathrm{d}t \le -\delta\epsilon g = -\delta\omega\sigma \le -10\sigma\sqrt{\omega}.$$

The last bound holds since $\omega = \omega(n) \to \infty$. The bound on $f(d^+)$ is similar but easier. $\square$

Corollary 3.3.5. *For n large, $f(t) \ge 10\sigma\sqrt{\omega(n)}$ for all $t \in [g, d^-]$.*

Proof. We just saw that $f(d^-) \ge 10\sigma\sqrt{\omega(n)}$. We also have $f(g) \ge \epsilon g/2 = \sigma\omega(n)/2 \ge 10\sigma\sqrt{\omega(n)}$, the last inequality for n large. The result then follows since f is concave. $\square$

The information about f derived above is summarized in the following picture.

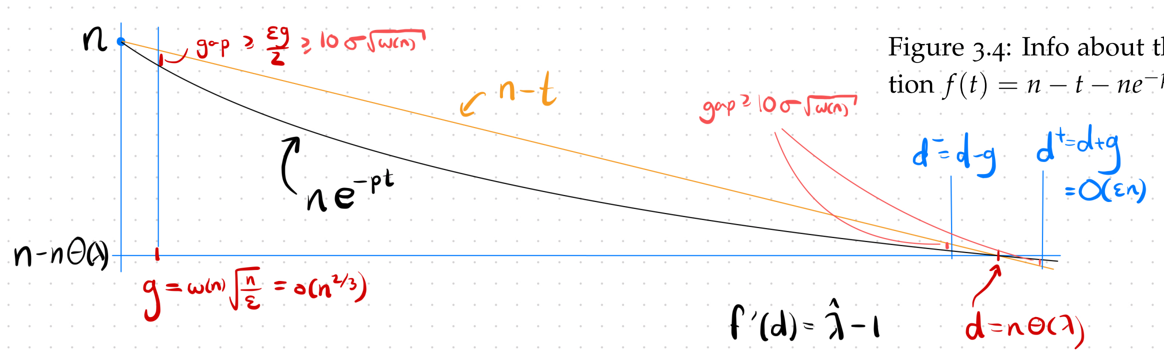


Figure 3.4: Info about the function $f(t) = n - t - ne^{-pt}$.

Now let $Z = \kappa(g) - 1$ be the number of components of $\mathcal{G}(n,p)$ completely explored by time g, and set $L = T_Z$ and $R = T_{Z+1}$.

Theorem 3.3.6. *Under the assumptions of Theorem 3.3.2, $\frac{L-R}{n\theta(\lambda)} \xrightarrow{\mathbb{P}} 1$.*

Before proving the theorem, we use it to prove Theorem 3.3.2. The following straightforward fact, which is used in the proof, is left as an exercise.

Exercise 3.3.3. *For any $A > 0$ there is $\alpha > 0$ such that for all $1 \le \lambda \le (A+1)$, it holds that*

$$\hat{\lambda} = \lambda(1-\theta(\lambda)) \le 1 - \alpha(\lambda - 1).$$

With $\lambda = 1+\epsilon$, the exercise shows that that $\hat{\lambda} < 1-\alpha\epsilon$. When $\lambda = 1+\epsilon$ for $\epsilon = o(1)$, then in fact $\hat{\lambda} = 1-\epsilon(1+o(1))$; the bound on $\hat{\lambda}$ from the exercise enlarges the range of ϵ allowed at the cost of weakening the control on $\hat{\lambda}$.

Proof of Theorem 3.3.2. Between times L and R, a connected component of $\mathcal{G}(n,p)$ is explored, so $|\mathcal{C}^{\max}| \geq R-L$. This establishes one of the two bounds needed to prove the theorem. For the other, we need to bound the largest size among the components explored before time R and after time L.

Since $L \leq g = o(n^{2/3})$ and $n\theta(\lambda)/n^{2/3} \to \infty$, it is immediate that any component explored before time L has size $o(n\theta(\lambda))$.

To control the component sizes after step R, we need a small auxiliary computation. Fix $\delta > 0$ to be chosen shortly[3], and set $n' = n-(1-\delta)n\theta(\lambda)$. Using the bound from Exercise 3.3.3 we have

$$\begin{aligned} p &= \frac{\lambda}{n} \\ &\leq \frac{\lambda(1-(1-\delta)n\theta(\lambda))}{n'} \\ &= \frac{\lambda(1-\theta(\lambda))+\delta\lambda\theta(\lambda)}{n'} \\ &= \frac{\hat{\lambda}+\delta\lambda\theta(\lambda)}{n'} \\ &\leq \frac{1-\alpha\epsilon+\delta(1+A)2\epsilon}{n'} \\ &\leq \frac{1-(\alpha/2)\epsilon}{n} \end{aligned}$$

the last bound holding with the choice $\delta = \alpha/(4(1+A))$. Let $\mathcal{G}'$ be the subgraph of $\mathcal{G}(n,p)$ with vertices $[n]\setminus\{v_1,\ldots,v_L\}$. Then $\mathcal{G}'$ is distributed as $\mathcal{G}(n-R,p)$, up to relabeling of the vertices.

Writing E for the the event that $L \geq (1-\delta)n\theta(\lambda)$ so that $n-L \leq n'$, and letting T be a Poisson$(1-(\alpha/2)\epsilon)$ Bienaymé tree, we then have

$$\begin{aligned} \mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}')| \geq s \mid E\right\} &\leq \mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}(n',p))| \geq s\right\} \\ &\leq \frac{n}{s}\mathbf{P}\left\{|\mathcal{C}_{\mathcal{G}(n',p)}(1)| \geq s\right\} \\ &\leq \frac{n}{s^2}\mathbf{E}\left[|T|\right], \\ &= \frac{n}{s^2}\frac{1}{(\alpha/2)\epsilon} \end{aligned}$$

Taking $s = g = \omega(n)\sqrt{n/\epsilon}$ we obtain that

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}')| \geq g \mid E\right\} = o(1).$$

But since also $\mathbf{P}\{E\} \to 1$, it follows that

$$\mathbf{P}\left\{|\mathcal{C}^{\max}(\mathcal{G}')| \geq g\right\} \to 0.$$

[3] We will take $\delta = \alpha/(4(1+A))$, in fact.

Since $g = o(\epsilon n)$ it follows that with high probability no component of $\mathcal{G}'$ is the largest of $\mathcal{G}(n, p)$. □

Exercise 3.3.4. *Let $(T_i, i \geq 1)$ be a sequence of independent Poisson(1) Bienaymé trees. Using the identity*

$$\mathbf{P}\{|T_1| + \ldots + |T_k| = n\} = \frac{k}{n}\mathbf{P}\{\text{Poisson}(n) = n - k\},$$

prove that

$$\frac{|T_1| + \ldots + |T_k|}{k^2}$$

converges in distribution. Give a formula for the cumulative distribution function of the limiting distribution.

To prove Theorem 3.3.6, we analyze the process $S(t) = \sum_{i=1}^{t} X(i)$. The next proposition is the key to the analysis; it expresses the difference between $S(t)$ and x_t as effectively a weighted sum of martingale differences, which we can then control using fairly standard arguments. Recall that $\mathcal{F}_t$ is the σ-algebra generated by the first t steps of the process.

Proposition 3.3.7. *Let $D_t = \mathbf{E}\{X(t) \mid \mathcal{F}_{t-1}\}$ and write*

$$\Delta_t = X(t) - D_t = X(t) - \mathbf{E}\{X(t) \mid \mathcal{F}_{t-1}\}.$$

Then

$$S(t) - x_t = \sum_{i=1}^{t}(1-p)^{t-i}(\Delta_i - p\kappa(i)).$$

Moreover, with $\tilde{S}(t) = x_t + \sum_{i=1}^{t}(1-p)^{t-i}\Delta_i$, then

$$|S(t) - \tilde{S}(t)| \leq pt\kappa(t).$$

We prepare for the proof of the proposition by deriving an identity for D_{t+1} in terms of $S(t)$ and $\kappa(t+1)$.

Lemma 3.3.8. *For all $t \geq 0$, $D_{t+1} = p(n - t - S(t) - \kappa(t+1)) - 1$.*

Proof. We saw in (3.3.1) that conditionally given $\mathcal{F}_t$,

$$X(t) + 1 = |N(v_{t+1}) \cap U_t|$$

is $\text{Bin}(|U_t'|, p)$-distributed, where $U_t' = U_t \setminus \{v_{t+1}\}$, so

$$|U_t'| = |U_t \setminus \{v_{t+1}\}| = \begin{cases} |U_t| & \text{if } |A_t| \neq \emptyset \\ |U_t| - 1 & \text{if } |A_t| = \emptyset. \end{cases}$$

since $\kappa(t+1) - \kappa(t) = \mathbf{1}_{[A_t = \emptyset]}$, this implies that

$$\begin{aligned}
D_{t+1} &= \mathbf{E}\{X(t+1) \mid \mathcal{F}_t\} \\
&= \mathbf{E}\{|N(v_{t+1}) \cap U_t| - 1 \mid \mathcal{F}_t\} \\
&= |U_t'|p - 1 \\
&= (|U_t| - (\kappa(t+1) - \kappa(t)))p - 1 \\
&= (n - t - S(t) - \kappa(t+1))p - 1,
\end{aligned}$$

the last equality holding since $U_t = n - t - |A_t| = n - t - S(t) - \kappa(t)$. $\square$

Proof of Proposition 3.3.7. First, using the definition of Δ_{t+1} and D_{t+1} and the lemma, we have

$$\begin{aligned}
S(t+1) &= S(t) + X(t+1) \\
&= S(t) + \Delta_{t+1} + D_{t+1} \\
&= S(t) + \Delta_{t+1} + (n - t - S(t) - \kappa(t+1))p - 1 \\
&= (1-p)S(t) + \Delta_{t+1} - p(n-t) - 1 - p\kappa(t+1).
\end{aligned}$$

Next,

$$x_{t+1} - (1-p)x_t = n - t - 1 - (1-p)(n-t) = p(n-t) - 1,$$

so $x_{t+1} = (1-p)x_t + p(n-t) - 1$. Subtracting this identity from the identity for $S(t+1)$ gives

$$S(t+1) - x_{t+1} = (1-p)(S(t) - x_t) + \Delta_{t+1} - p\kappa(t+1),$$

and the claimed identity follows by induction.

For the bound, simply note that since $\kappa(i)$ is increasing and $(1-p) < 1$,

$$|S(t) - \tilde{S}(t)| = |\sum_{i=1}^{t} (1-p)^{t-i} p\kappa(i)| \le \sum_{i=1}^{t} p\kappa(t) = pt\kappa(t).$$

$\square$

The bound on the difference between $S(t)$ and $\tilde{S}(t)$ means we can get a handle on the behaviour of $S(t)$ if we can (a) understand that of $\tilde{S}(t)$ and (b) bound the upper tail probabilities of $\kappa(t)$. We tackle these points in turn.

Proposition 3.3.9. *Write*

$$M(t) = \frac{\tilde{S}(t) - x_t}{(1-p)^t} = \sum_{i=1}^{t} (1-p)^{-i} \Delta_i.$$

Then for any $1 \leq t \leq n$*,*

$$\sup_{i \leq t} M(i) = O_P(\sqrt{t})\,.$$

As a consequence, for any $1 \leq t \leq n$*,*

$$\max_{i \leq t} |\tilde{S}(t) - f(t)| = O_P(\sqrt{t}).$$

The first bound of the proposition may be restated as follows. For any $\delta > 0$ there is $k > 0$ such that

$$\sup_{n \geq 1} \sup_{1 \leq t \leq n} \mathbf{P}\left\{\sup_{i \leq t} M(i) \geq K\sqrt{n}\right\} \leq \delta\,.$$

In proving the proposition, we will use the following straightforward martingale identity and bound.

Lemma 3.3.10. *Let* $(M_t)_{t \geq 0}$ *be a martingale relative to a filtration* $(\mathcal{F}_t)_{t \geq 0}$*, with* $M_0 = 0$*, and let* $I_t = M_t - M_{t-1}$ *for* $t \geq 1$*. Then*

$$\mathbf{Var}\{M_t\} = \sum_{i=1}^{t} \mathbf{Var}\{I_i\} = \sum_{i=1}^{t} \mathbf{E}\left[\mathbf{Var}\{I_i \mid \mathcal{F}_{i-1}\}\right]\,.$$

Moreover, for any $K > 0$*,*

$$\mathbf{P}\left\{\max_{i \leq t} |M_i| \geq K\right\} \leq \mathbf{Var}\{Z_t\}/K^2\,.$$

Proof. For the identity, since $\mathbf{E}M_t = \mathbf{E}M_0 = 0$, and $M_t = \sum_{i=1}^{t} I_i$,

$$\mathbf{Var}\{M_t\} = \mathbf{E}\left[M_t^2\right] = \sum_{i,j=1}^{t} \mathbf{E}\left[I_i I_j\right]\,.$$

But for $i < j$,

$$\mathbf{E}\left[I_i I_j\right] = \mathbf{E}\left[\mathbf{E}\left\{I_i I_j \mid \mathcal{F}_{j-1}\right\}\right] = \mathbf{E}\left[I_i \mathbf{E}\left\{I_j \mid \mathcal{F}_{j-1}\right\}\right] = 0\,,$$

so

$$\mathbf{Var}\{M_t\} = \sum_{i=1}^{t} \mathbf{E}\left[I_i^2\right]\,.$$

Also, since $\mathbf{E}\{I_i \mid \mathcal{F}_{i-1}\} = 0$, $\mathbf{Var}\{I_i \mid \mathcal{F}_{i-1}\} = \mathbf{E}\left\{I_i^2 \mid \mathcal{F}_{i-1}\right\}$ so

$$\mathbf{E}\left[\mathbf{Var}\{I_i \mid \mathcal{F}_{i-1}\}\right] = \mathbf{E}\left[\mathbf{E}\left\{I_i^2 \mid \mathcal{F}_{i-1}\right\}\right] = \mathbf{E}\left[I_i^2\right]\,,$$

which combined with the previous identity for $\mathbf{Var}\{M_t\}$ establishes the equality in the lemma. The bound is immediate from Doob's maximal inequality. $\square$

Proof of Proposition 3.3.9. First, note that $M(t)$ is indeed a martingale: since $\Delta_t = X(t) - \mathbf{E}\{X_t \mid \mathcal{F}_{t-1}\}$, $\mathbf{E}\{\Delta_t \mid \mathcal{F}_{t-1}\} = 0$ and $\mathbf{E}\{\Delta_i \mid \mathcal{F}_{t-1}\} = \Delta_i$ for $i < t$, so

$$\begin{aligned}\mathbf{E}\{M(t) \mid \mathcal{F}_{t-1}\} &= \sum_{i=1}^{t}(1-p)^{-i}\mathbf{E}\{\Delta_i \mid \mathcal{F}_{t-1}\} \\ &= \sum_{i=1}^{t-1}(1-p)^{-i}\Delta_i = M(t-1)\,.\end{aligned}$$

Now recall that conditionally given $\mathcal{F}_{i-1}$, $X(i)+1$ is $\mathrm{Bin}(|U_i'|, p)$-distributed, so Δ_i is (conditionally) simply a binomial shifted by its mean. Thus

$$\mathbf{Var}\{\Delta_i \mid \mathcal{F}_{i-1}\} = |U_{i-1}'|p(1-p) \le np(1-p)$$

and so

$$\begin{aligned}\mathbf{Var}\left\{(1-p)^{-i}\Delta_i\right\} &= (1-p)^{-2i}\mathbf{Var}\{\Delta_i\} \\ &\le (1-p)^{-2n} \cdot np \\ &\le \left(1 - \frac{A+1}{n}\right)^{-2n} \cdot (A+1) < C < \infty\,.\end{aligned}$$

Thus

$$\mathbf{Var}\{M(t)\} = \sum_{i=1}^{t}(1-p)^{-2i}\mathbf{Var}\{\Delta_i\} \le Ct,$$

so by Doob's maximal inequality (the second bound in Lemma 3.3.10, for any $\delta > 0$, with $K = (C/\delta)^{1/2}$,

$$\mathbf{P}\left\{\max_{i\le t}|M(i)| > Kt^{1/2}\right\} \le \frac{Ct}{K^2 t} = \delta\,.$$

This proves the first assertion of the proposition. For the second, we have

$$\max_{i\le t}|\tilde{S}(t) - f(t)| \le \max_{i\le t}|\tilde{S}(t) - x_t| + \max_{i\le t}|x_t - f(t)|$$

The second term is $O(1) = O(\sqrt{t})$, and the first is bounded from above by $\max_{i\le t}|M(i)|$ since $|M(i)| = |\tilde{S}(i) - x_i|/(1-p)^i \ge |\tilde{S}(i) - x_i|$. □

Combining the bound

$$|S(t) - \tilde{S}(t)| \le pt\kappa(t)$$

from Proposition 3.3.7 with the bound

$$\max_{i\le t}|\tilde{S}(i) - f(i)| = O_P(\sqrt{t}).$$

from Proposition 3.3.9, using that $\kappa(i)$ is increasing in i we obtain that

$$\max_{i \le t} |S(i) - f(i)| \le pt\kappa(t) + O_P(\sqrt{t}). \tag{3.3.2}$$

The main step remaining in the proof is to bound $\kappa(t)$, and we now turn to this. Recall that $Z = \kappa(g) - 1$ is the number of components fully explored by time g, and that $L = T_Z \le g$.

Proposition 3.3.11. *With probability tending to one, $Z \le \sigma/\omega$ and $L \le \sigma/(\epsilon\omega)$.*

Proof. First,

$$|S(L) - \tilde{S}(L)| \le pL\kappa(L) = pLZ \le pgZ.$$

But

$$S(L) = S(T_Z) = A_{T_Z} - \kappa(T_Z) = -\kappa(T_Z) = -Z,$$

so

$$\tilde{S}(L) \le (pg - 1)Z. \tag{3.3.3}$$

Since $g = o(n^{2/3})$ we have $pg < 1/2$ for n large. Also, $f(L) > 0$ since $L \le g < n\theta(\lambda)$, so the preceding bound implies that

$$|\tilde{S}(L) - f(L)| > |-Z/2 - 0| = Z/2.$$

It follows that

$$\begin{aligned} \mathbf{P}\{Z \ge \sigma/\omega\} &\le \mathbf{P}\left\{|\tilde{S}(L) - f(L)| \ge \sigma/(2\omega)\right\} \\ &\le \mathbf{P}\left\{\sup_{0 \le t \le g} |\tilde{S}(t) - f(t)| \ge \sigma/(2\omega)\right\} \end{aligned} \tag{3.3.4}$$

By Proposition 3.3.9,

$$\sup_{0 \le t \le g} |\tilde{S}(t) - f(t)| = O_P(\sqrt{g}),$$

so as long as $\sqrt{g} = o(\sigma/\omega)$ we deduce that the probability in (3.3.4) tends to zero and thus $Z \le \sigma/\omega$ with high probability. But

$$\frac{\sigma^2}{\omega^2 g} = \frac{n\epsilon}{\omega^2(\omega(n/\epsilon)^{1/2})} = \frac{n\epsilon^3}{\omega} \to \infty,$$

so indeed $g = o(\sigma^2/\omega^2)$.

Next, by definition $L \le g$ so to prove the second assertion it suffices to show that, with $I = [\sigma/(\epsilon\omega), g]$, then $\mathbf{P}\{L \in I\} \to 0$. But for $t \in I$, by Proposition 3.3.3,

$$f(t) \ge \frac{\epsilon t}{2} \ge \frac{\sigma}{2\omega},$$

so if

$$\sup_{0\leq t\leq g} |\tilde{S}(t) - f(t)| < \sigma/(2\omega)$$

then

$$\inf_{t\in I} \tilde{S}(t) > 0$$

for all $t \in I$. On the other hand, if $L \in I$ then by (3.3.3),

$$\tilde{S}(L) < 0$$

This implies that

$$\mathbf{P}\{L \in I\} \leq \mathbf{P}\left\{\sup_{0\leq t\leq g} |\tilde{S}(t) - f(t)| \geq \sigma/(2\omega)\right\},$$

which we already saw tends to zero (recall (3.3.4)). □

Proof of Theorem 3.3.6. Recall that

$$d^+ = d + g = n\theta(\lambda) + \omega(n/\epsilon)^{1/2} = O(n\theta(\lambda)) = o(\sigma^2).$$

Let E be the event that

$$\sup_{t\leq\min(R,d^+)} |S(t) - f(t)| \leq \sigma\sqrt{\omega}.$$

Since $\kappa(R) = Z$, the bound (3.3.2) and Proposition 3.3.11 together give

$$\max_{t\leq\min(R,d^+)} |S(t) - f(t)| \leq npZ + O_P(\sqrt{d^+}) = O_P(\sigma),$$

so $\mathbf{P}\{E\} \to 1$.

Recall from Corollary 3.3.5 that $f(t) \geq 10\sigma\sqrt{\omega(n)}$ for $t \in [g, d^-]$ when n is large. In view of this, if E occurs then $S(t) > 0$ for $t \in [g, d^-]$. On the other hand, $S(R) = -(Z+1) = -\kappa(g) < 0$ so this implies that when E occurs, $R \geq d^-$.

Next, if E occurs and also $Z \leq \sigma$, then by the bound on $f(d^+)$ from (3.3.4) we have

$$S(d^+) \leq \sigma\sqrt{\omega} - 10\sigma\sqrt{\omega} = -9\sigma\sqrt{\omega} < -(Z+1) < -\kappa(g),$$

so $R = T_{Z+1} \leq d^+$. Since $\mathbf{P}\{E \cap \{Z \leq \sigma\}\} \to 1$, it follows that $\mathbf{P}\{R \leq d^+\} \to 1$.

We have established that $\mathbf{P}\{R \in [d^-, d^+]\} \to 1$; since $d^- = d - g$, $d^+ = d + g$ and $g = o(n\epsilon) = o(n\theta(\lambda))$, it follows that $R/(n\theta(\lambda)) \xrightarrow{\mathrm{P}} 1$. Since $L \leq g = o(n\theta(\lambda))$, this also implies that $(R-L)/(n\theta(\lambda)) \xrightarrow{\mathrm{P}} 1$. □

Exercise 3.3.5. *For a tree t with vertex set* $[n]$*, let* $S_{\mathrm{t}} = (S_{\mathrm{t}}(i), 0 \le i \le n)$ *be the depth-first queue process of T (started from the vertex with label* 1*), and write* $a(\mathrm{t}) = \sum_{i \le i < n} S_{\mathrm{t}}(i)$*. For distinct* $u, v \in [n]$ *say that edge* $e = uv$ *is allowable for tree* t *if* $e \notin e(\mathrm{t})$ *and the depth first tree of* $\mathrm{t} + e$ *is* t*, i.e., adding the edge e to* t *doesn't change the depth-first tree*

Fix any $n \in \mathbb{N}$ *and* $s \in \mathbb{N}$ *such that* $s \le \binom{n}{2} - (n-1)$*. Build a connected graph G with vertex set* $[n]$ *and surplus s as follows. First let* T *be a random tree with vertex set* $[n]$ *such that*

$$\mathbf{P}\left\{\mathrm{T} = \mathrm{t}\right\} \propto \binom{a(\mathrm{t})}{k}.$$

Conditionally given T*, let S be a set of s allowable edges, chosen uniformly at random over all such sets. Let G be obtained by adding the edges in S to the tree T.*

Show that G is a uniformly random element of $\mathcal{G}_{n,s}$*.*

4
Minimum spanning trees

4.1 Basics

The first thing to do is define minimum spanning trees. Let $G = (V, E)$ be a graph. A *spanning subgraph* of G is a graph $H = (V, E')$ with $E' \subseteq E$. A *spanning tree* of G is a spanning subgraph of G which is a tree.

Exercise 4.1.1. *Every finite connected graph has at least one spanning tree.*

Now suppose the edges of G are given weights $\mathrm{w} = (\mathrm{w}(e), e \in E)$. The *weight* of a subgraph $H = (V', E')$ of G is

$$\mathrm{w}(H) = \sum_{e \in E'} \mathrm{w}(e) .$$

A spanning subtree T of G is a *minimal* spanning tree of G with respect to w if $\mathrm{w}(T) \leq \mathrm{w}(T')$ for every spanning tree T' of G. When the weights w are clear, we simply say the tree T is a minimal spanning tree of G.

Proposition 4.1.1. *If G is a finite connected graph, and the weights $\mathrm{w} = (\mathrm{w}(e), e \in e(G))$ are all distinct, then G has a unique minimal spanning tree.*

Proof. Suppose for contradiction that T and T' are distinct minimal spanning trees. Let $e = uv$ be the smallest weight edge which is in exactly one of the trees T and T'. By symmetry we may assume that e is an edge of T and not of T'

Let γ be the unique $u - v$ path in T'. Since T has no cycles, there must be an edge e' of γ which is not in T. By our choice of e and since the edge weights are all distinct, this implies that $\mathrm{w}(e') > \mathrm{w}(e)$. Then the spanning subgraph H obtained from T' by removing e' and adding e is then a spanning tree with smaller weight than T'. $\square$

4.2 *Kruskal's algorithm*

The idea at the heart of Kruskal's algorithm is contained in the following observation.[1]

[1] Joseph B. Kruskal, Jr. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.*, 7:48-50, 1956. URL http://dx.doi.org/10.2307/2033241

Lemma 4.2.1. *Fix a finite connected graph $G = (V, E)$ and weights $\mathrm{w} = (\mathrm{w}(e), e \in E)$ which are all distinct. List the edges in increasing order of weight as $e_1, \ldots, e_{|E|}$. For $1 \le i \le |E|$, let $E_i = \{e_1, \ldots, e_i\}$, and let $G_i = (V, E_i)$. Also, let $G_0 = (V, \emptyset)$. For all $1 \le i \le |E|$, if the endpoints of e_i are in distinct connected components of G_{i-1} then e_i is an edge of the minimum spanning tree.*

Proof. Write u and v for the endpoints of e_i. Let γ be the unique $u - v$ path in the MST. There is no path from u to v in G_{i-1}. It follows that if e_i is not in the MST then γ must contain an edge e_j with $j > i$. In this case, the spanning tree obtained from the MST by removing e_j and adding e_i has smaller weight than the MST, a contradiction. $\square$

Corollary 4.2.2. *The edges of the minimum spanning tree of G are precisely the elements of the set*

$$S = \{e_i : e_i \text{ joins distinct connected components of } G_{i-1}\}.$$

Proof. The lemma states that all elements of S are edges of the MST. A tree with $|V|$ vertices has $|V| - 1$ edges, so if $|S| = |V| - 1$ then the corollary follows.

For $1 \le i \le |V|$, if $e_i \in S$ then G_i has one fewer connected component of G_{i-1}, and if $e_i \notin S$ then G_{i-1} and G_i have the same number of connected components. Since G_0 has $|V|$ connected components and $G = G_{|V|}$ has 1 connected, component, it follows that $|S| = |V| - 1$. $\square$

Kruskal's algorithm is a procedural description of the above lemma.

Kruskal's algorithm.
Input: A graph $G = (V, E)$ and weights $\mathrm{w} = (\mathrm{w}(e), e \in E)$.
1. List the edges in increasing order of weight as $e_1, \ldots, e_{|E|}$.
2. Let $S_0 = \emptyset$, let $F_0 = (V, S_0)$.
3. For $1 \le i \le |E|$:
   – If e_i joins distinct trees of F_{i-1} then let $S_i = S_{i-1} \cup \{e_i\}$. Otherwise, let $S_i = S_{i-1}$.
   – Let $F_i = (V, S_i)$.

Ouput: $F_{|E|}$.

Kruskal's algorithm adds edge e_i if it joins distinct trees of F_{i-1}. The next exercise implies this is equivalent to adding e_i if it joins distinct connected components of G_{i-1}.

Exercise 4.2.1. *Show that for each* $0 \le j \le |E|$ *and each connected component* H *of* G_j*, there is a connected component* T_H *of* F_j *with vertex set* $\mathrm{v}(H)$*.*

Thus, if G is connected then the set of edges of $F_{|E|}$ is precisely the set S from Corollary 4.2.2, so the output of Kruskal's algorithm is the MST of G with weights w. The next exercise strengthens this conclusion slightly.

Exercise 4.2.2. *Show that for each* $0 \le j \le |E|$*, and each connected component* H *of* G_j*, the connected component* T_H *of* F_j *with vertex set* U *is in fact the MST of* H *with weights* $(\mathrm{w}(e), e \in e(H))$*.*

It is reasonable to call Kruskal's algorithm the greedy algorithm for constructing a spanning tree of G, since at each step it adds the "first" edge it can which does not create a cycle. When encountering such a clean, simple, provably correct procedure, it is natural to ask whether it applies in other settings. In this case, the answer is "yes".

Definition 4.2.3. *An* independence system *is a pair* $(X, \mathcal{I})$*, where* X *is a finite set and* $\mathcal{I}$ *is a nonempty set of subsets of* X *satisfying the* hereditary property*:*

1. If $A \in \mathcal{I}$ *and* $B \subset A$ *then* $B \in \mathcal{I}$*.*

An independence system $(X, \mathcal{I})$ *is a* matroid *if* $\mathcal{I}$ *also satisfies the* augmentation property*:*

2. If $A, B \in \mathcal{I}$ *and* $|A| < |B|$ *then there exists* $x \in B \setminus A$ *such that* $A \cup \{x\} \in \mathcal{I}$*.*

If $(X, \mathcal{I})$ is an independence system (so in particular if $(X, \mathcal{I})$ is a matroid) then X is called the *ground set* of (X, I), and the elements of $\mathcal{I}$ are called the *independent sets* of (X, I).

An independent set B is *maximal* if for all $x \in X \setminus B$, the set $B \cup \{x\}$ is not an independent set. The set of maximal independent sets in an independence system $M = (X, \mathcal{I})$ is denoted

$$\mathcal{B} = \mathcal{B}(M) = \{B \in \mathcal{I} : B \text{ is maximal within } \mathcal{I}\}.$$

If M is a matroid then its maximal independent sets are also called *bases*.

Exercise 4.2.3. *Let* $\mathcal{B}$ *be the set of bases of a matroid* $M = (X, \mathcal{E})$*. Prove the following facts.*

1. Any two bases $A, B \in \mathcal{B}$ *have the same size.*

2. The set $\mathcal{B}$ *satisfies the* basis exchange property*: for all* $A, B \in \mathcal{B}$*, there exists* $a \in A \setminus B$ *and* $b \in B \setminus A$ *such that* $(A \setminus \{a\}) \cup \{b\} \in \mathcal{B}$*.*

3. Any non-empty set $\mathcal{A}$ *of subsets of* X *which satisfies the basis exchange property is the set of bases of some matroid.*

For a finite connected graph $G = (V, E)$, the set $\mathcal{E}$ of edge sets of spanning trees of G satisfies the basis exchange property.[2] By the preceding exercise, $\mathcal{E}$ is thus the set of bases of some matroid. The matroid is called the *graphic matroid* of G; it's ground set is E, and its independent sets are precisely the acyclic spanning subgraphs (or *spanning forests*) of G.

[2] To see this, fix any two spanning trees T, T', with edge sets $E, E' \in \mathcal{E}$. Let $e = uw$ be any edge of T which is not an edge of T', and let γ be the $u - v$ path in T'. Since T contains no cycle, there must be an edge e' of γ which is not in T. Then $E'' = (E' \setminus \{e'\}) \cup \{e\}$ is the edge set of a spanning tree T'' so $E'' \in \mathcal{E}$.

Now suppose the elements of X are given weights $\mathrm{w} = (\mathrm{w}(x), x \in X)$. The weight of a set $A \subset X$ is $\mathrm{w}(A) = \sum_{x \in A} \mathrm{w}(x)$. The greedy algorithm builds a maximal independent set when applied to any independence system.

Greedy algorithm for independence systems.
Input: An independence system $(X, \mathcal{I})$ and weights $\mathrm{w} = (\mathrm{w}(x), x \in X)$.
1. List the elements of X in increasing order of weight as $x_1, \ldots, x_{|X|}$.
2. Let $I_0 = \emptyset$.
3. For $1 \le i \le |X|$:
   – If $I_{i-1} \cup \{x_i\} \in \mathcal{I}$ then let $I_i = I_{i-1} \cup \{x_i\}$. Otherwise, let $I_i = I_{i-1}$.

Ouput: $I_{|X|}$.

Let $I_{|X|}$ be the output of the greedy algorithm run on input $(X, \mathcal{I})$ and w. Observe that for any $0 \le i \le j \le |X|$, if $x_i \notin I_j$ then $I_j \cup \{x_i\} \notin \mathcal{I}$ and so $I_{|X|} \cup \{x_i\} \notin \mathcal{I}$. It follows that $I_{|X|}$ is indeed a maximal independent set (so a basis, if $(X, \mathcal{I})$ is a matroid). The next proposition states that if $(X, \mathcal{I})$ is a matroid, then the greedy algorithm in fact finds a minimum-weight basis.

Proposition 4.2.4. *For any matroid $M = (X, \mathcal{I})$ and non-negative weights $\mathrm{w} = (\mathrm{w}(x), x \in X)$, on input M and w, the greedy algorithm outputs a minimum weight basis of M with respect to w. Moreover, if the weights are all distinct then the minimum weight basis is unique.*

Proof. Let A be any other basis different from $I_{|X|}$. First, observe that $|A \cap \{x_1, \ldots, x_i\}| \le |I_i|$ for all $0 \le i \le |X|$. Indeed, if not then by the augmentation property there is $x_j \in |A \cap \{x_1, \ldots, x_i\}|$ such that $I_i \cup \{x_j\} \in I$, contradicting the above observation.

List the elements of $I_{|X|}$ as $i_1, \ldots, i_m$ and the elements of A as $a_1, \ldots, a_m$, in both cases in increasing order of weight. The claim established in the preceding paragraph implies that $\mathrm{w}(a_j) \ge \mathrm{w}(i_j)$ for each $1 \le j \le m$, so $\mathrm{w}(A) \ge \mathrm{w}(I_{|X|})$. Thus, $I_{|X|}$ is a minimum weight basis.

Finally, if the weights are all distinct then there must be $1 \le j \le m$ such that $\mathrm{w}(a_j) > \mathrm{w}(i_j)$, so $\mathrm{w}(A) > \mathrm{w}(I_{|X|})$, proving uniqueness. □

Perhaps more remarkable than the preceding proposition, is that the success of the greedy algorithm *characterizes* matroids among independence systems.

Proposition 4.2.5. *Let (X, I) be an independence system. Suppose that for any non-negative weights $\mathrm{w} = (\mathrm{w}(x), x \in X)$, the greedy algorithm outputs a maximal independent set of minimum weight. Then (X, I) is a matroid.*

Proof. First observe that the supposition implies all maximal independent sets have the same size. To see this, suppose for a contradiction that A and B are two maximal independent sets with $|A| < |B|$, and fix $c \in (|A \setminus B|/|B \setminus A|, 1)$. Then define weights $\mathrm{w} = (\mathrm{w}(x), x \in X)$ as follows:

$$\mathrm{w}(x) = \begin{cases} 0 & \text{if } x \in A \cap B \\ c & \text{if } x \in B \setminus A \\ 1 & \text{if } x \in A \setminus B \\ 2 & \text{otherwise.} \end{cases}$$

Then $\mathrm{w}(A) = |A \setminus B| \leq |A|$ and $\mathrm{w}(B) = c|B \setminus A| > |A \setminus B|$. However, the greedy algorithm first adds all of $A \cap B$, then adds all of $B \setminus A$, so the output is the set B, a contradiction.

Now fix any two sets $A, B \in \mathcal{I}$ with $|B| = |A| + 1$. To prove $(X, \mathcal{I})$ is a matroid, it suffices to show that there exists $x \in B \setminus A$ with $A \cup \{x\} \in \mathcal{I}$, since this implies $(X, \mathcal{I})$ has the augmentation property.

Write m for the size of the maximal independent sets in $\mathcal{I}$. Then define a weight function as follows:

$$\mathrm{w}(x) = \begin{cases} 0 & \text{if } x \in A \\ 1 & \text{if } x \in B \setminus A \\ m+1 & \text{otherwise.} \end{cases}$$

For these weights, the greedy algorithm first adds all of A. Since maximal independent sets have size $m \geq |B| > |A|$, the greedy algorithm must then add at least one further element before returning a maximal independent set C.

Observe that any maximal independent set D containing B has

$$\mathrm{w}(D) \leq |B \setminus A| + (m+1)(m - |B|).$$

Next, any set S of size m containing A and disjoint from $B \setminus A$ has

$$\mathrm{w}(S) = (m+1)(m?|A|) = (m+1) + (m+1)(m?|B|) > \mathrm{w}(D).$$

Finally, C contains A, and by the assumption about the behaviour of the greedy algorithm, $\mathrm{w}(C) \leq \mathrm{w}(D)$. Thus, necessarily C contains at least one element of $B \setminus A$, so there is some $x \in B \setminus A$ such that $A \cup \{x\} \in \mathcal{I}$. $\square$

Exercise 4.2.4. *Show that the greedy algorithm need not output a maximal independent set of minimum weight if $(X, \mathcal{I})$ is not an independence system.*

Exercise 4.2.5. *Let $M = (X, \mathcal{I})$ be a matroid.*

1. Fix $Y \subset X$ and let $\mathcal{I}|_Y := \{I \in \mathcal{I} : I \subset Y\}$. Show that $(Y, \mathcal{I}|_Y)$ is a matroid.

2. Fix non-negative weights $\mathrm{w} = (\mathrm{w}(x), x \in X)$, and let $(I_j, 0 \leq j \leq |X|)$ be the sequence of sets constructed by the greedy algorithm run on input M and w. Show that for all $1 \leq i \leq |X|$, I_j is a minimum weight basis of $(\{x_1, \ldots, x_j\}, \mathcal{I}|_{\{x_1,\ldots,x_j\}})$.

4.3 *Cycle breaking*

Kruskal's algorithm adds edges in increasing order of weight, provided their addition does not create cycles. Cycle breaking is very similar: it *removes* edges, in decreasing order of weight, provided their deletion does not disconnect the graph.

In the language of matroids, cycle breaking is a special instance of the greedy algorithm: cycle breaking on G is the greedy algorithm applied to the dual matroid of the graphic matroid of G.

The dual of a matroid, or more generally of an independence system $(X, \mathcal{I})$, is defined as follows. First note that I can be recovered from the set B of its maximal elements, due to the hereditary property. Precisely,

$$\mathcal{I} = \bigcup_{B \in \mathcal{B}} : \{S \subset X : S \text{ is a subset of } B\}.$$

The *dual* system of $(X, \mathcal{I})$ is the pair $(X, \mathcal{I}^*)$, where

$$\mathcal{I}^* = \bigcup_{B \in \mathcal{B}} \{S \subset S : S \text{ is a subset of } B^c\}.$$

For a matroid $M = (X, \mathcal{I})$, write $M^* = (X, \mathcal{I}^*)$ and call M^* the *dual matroid* of M.

Exercise 4.3.1. *Prove that if $(X, \mathcal{I})$ is an independence system then $(X, \mathcal{I}^*)$ is an independence system, and if $(X, \mathcal{I})$ is a matroid then $(X, \mathcal{I}^*)$ is a matroid.*

Next, let M be the graphic matroid of a connected graph $G = (V, E)$. Then M^* is the *co-graphic* matroid, whose independent sets are the sets $S \subset E$ for which $(V, E \setminus S)$ is connected. In particular, for a given $S \in \mathcal{I}^*$, the edges $e \in E \setminus S$ such that $S \cup \{e\} \in \mathcal{I}^*$ are precisely those for which $(V, E \setminus (S \cup \{e\}))$ is connected.

Given non-negative weights $\mathrm{w} = (\mathrm{w}(e), e \in E)$, let $W = \max\{\mathrm{w}(e) : e \in E\}$, and define new weights $\mathrm{w}^* = (\mathrm{w}^*(e), e \in E)$ by

$$\mathrm{w}^*(e) = W - \mathrm{w}(e).$$

The cycle breaking algorithm on G with weights w is equivalent to the greedy algorithm on M^* with weights w^*, and yields a minimum weight basis S of M^*.

Note that $B = E \setminus S$ is a basis of M, so is the edge set of a spanning tree $T = (V, B)$ of G. For any other spanning tree $T' = (V, C)$ of G, the set $E \setminus C$ is also a basis of M^*, so

$$\mathrm{w}(C) = \mathrm{w}(E) - \mathrm{w}(E \setminus C) \leq \mathrm{w}(E) - \mathrm{w}(E \setminus S) = \mathrm{w}(B)\,.$$

In other words, T is a minimal spanning tree of B, so cycle breaking indeed builds an MST.

4.4 *Prim's algorithm*

The greedy algorithms seen above are global, in the sense that they optimize the choice of which edge to add over all possible edges. Prim's algorithm is also a greedy procedure, but it only optimizes *locally*, growing a spanning tree one vertex at a time.

Prim's algorithm.
Input: A finite connected graph $G = (V, E)$, a starting node $\rho \in V$, and weights $\mathrm{w} = (\mathrm{w}(e), e \in E)$.
1. Let $E_0 = \emptyset$ and let $v_0 = \rho$.
2.For $1 \leq i < |V|$:
– let $e_i \in E$ minimize $\{\mathrm{w}(e) : e = uv, u \in \{v_0, \ldots, v_{i-1}\}, v \notin \{v_0, \ldots, v_{i-1}\}$;
– let $v_i = v$ and let $E_i = E_{i-1} \cup \{e_i\}$.
Ouput: $(V, E_{|V|})$.

Proposition 4.4.1. *If the weights* w *are non-negative then the output of Prim's algorithm is a minimal spanning tree of G with weights* w.

Proof. It suffices to prove that for all $0 < i \leq |V|$, the set of edges $\{e_j, 1 \leq j < i\}$ is a subset of the edge set of some MST.

The proof proceeds by induction. For $i = 0$ there is nothing to prove. Now fix $i \geq 1$ and suppose $e_1, \ldots, e_{i-1}$ are all edges of some MST T. Write $e_i = u_i v_i$, where u_i is the endpoint of e_i which is an element of $\{v_0, \ldots, v_{i-1}\}$; its other endpoint is v_i. Let γ be the path from u_i to v_i in T. If γ contains e_i then e_i is an edge of T so $e_1, \ldots, e_i$ are all edges of T as required. If γ does not contain the edge e_i then γ contains another edge e connecting $\{v_0, \ldots, v_{i-1}\}$ with the rest of the vertices. By the definition of e_i, it must be that $\gamma(e) \geq \gamma(e_i)$. Thus the tree T' obtained from T by removing e and adding e_i is a spanning tree with weight $\mathrm{w}(T') \leq \mathrm{w}(T)$. Thus T' is a minimum tree containing $e_1, \ldots, e_i$. This completes the induction and the proof. $\square$

Prim's algorithm (which one might call the local greedy algorithm) is also a clean, simple, provably correct procedure, and as with Kruskal's algorithm, it is natural to ask in what generality it yields optimal results. The answer to this question is also known; it is a class of problems called *matroid embeddings*.[3] This class contains a diverse enough range of problems that a general theory of random instances of matroid embedding problems seems unlikely.

[3] Paul Helman, Bernard M. E. Moret, and Henry D. Shapiro. An exact characterization of greedy structures. *SIAM J. Discrete Math.*, 6(2):274? 283, 1993. http://dx. doi.org/10.1137/0406021

4.5 Euclidean MSTs

4.6 The mean-field MST

In this section, we study the structure of $T_n = \mathrm{MST}(K_n, \mathrm{U})$, where $\mathrm{U} = (U_e, e \in e(K_n))$ are independent Uniform$[0,1]$ edge weights. Writing $\mathcal{F}(n,p)$ for the subgraph of T_n consisting of only edges of weight at most p, then Kruskal's algorithm couples $\mathcal{F}(n,p)$ with $\mathcal{G}(n,p)$ so that the vertex sets of the connected components of the two random graphs are identical; more strongly, each connected component of $\mathcal{F}(n,p)$ is the minimum weight spanning tree of the corresponding component of $\mathcal{G}(n,p)$, for each $0 \le p \le 1$.

For $p = (1+\epsilon)/n$ with $\epsilon > 0$ fixed, with high probability the random graph process $\mathcal{G}(n,p)$ contains a unique component of linear size, and all other components are of logarithmic size. Thus, in some sense the global structure of T_n is already essentially determined by that of the largest component of $\mathcal{F}(n,(1+\epsilon)/n)$. The next subsection develops this remark in some detail; the following subsection focusses on the local structure of T_n.

Global properties

Subsection to be written.

The local limit

Recall that for $\mathrm{T} \in_u \mathscr{T}_n$, we described the asymptotic local structure of T as a sequence of independent Poisson(1) Bienaymé trees, glued along an infinite path. This section similarly addresses the local structure of the minimum spanning tree T_n. However, as opposed to the situation for random combinatorial trees, here we also wish to record local information about the edge weights as well (they are, after all, an essential aspect of the minimum weight spanning tree!). We saw earlier in the section that the structure of T_n is essentially determined by edges of weight $O(1/n)$, so in order to obtain non-trivial asymptotic information it is useful to scale the weights, by a factor of n.

Given $v \in [n]$, write $(U_i^v, 1 \leq i \leq n-1)$ for the edge weight incident to v in K_n, listed in increasing order, and let $(P_1^v, \ldots, P_{n-1}^v) = (nU_1^v, \ldots, nU_{n-1}^v)$.

Proposition 4.6.1. *Let $\mathcal{P}$ be a homogeneous rate-one Poisson point process on $[0, \infty)$; list its atoms in increasing order as $P_1 < P_2 < \ldots$. Then for any k, $(P_1^v, \ldots, P_k^v) \xrightarrow{\mathrm{d}} (P_1, \ldots, P_k)$.*

We leave the detailed proof to the reader; the key point is that for any $x > 0$,

$$|\{i : P_i^v \leq x\}| = |\{w \in [n] : w \neq v, U_{vw} \leq x/n\}| \stackrel{\mathrm{d}}{=} \mathrm{Bin}(n-1, x/n) \xrightarrow{\mathrm{d}} \mathrm{Poisson}(x)\,.$$

Definition 4.6.2. *The* Poisson-weighted infinite tree *(PWIT) is the random weighting of the Ulam-Harris tree given by weighting the child edges of each vertex with the ranked sequence of atoms $(P_{v,vi}, i \geq 1)$ of a rate-one Poisson process on $[0, \infty)$, the processes for different vertices being mutually independent.*

We'd like to describe the local structure of T_n in terms of the PWIT. For this, we require a brief digression, to introduce the notion of local weak convergence in more detail.

We begin with unweighted graphs. Let $G = (V, E, \rho)$ and $\hat{G} = (\hat{V}, \hat{E}, \hat{\rho})$ be connected rooted graphs with all degrees finite; we call such graphs *locally finite graphs*. A *rooted isomorphism* between G and $\hat{G}$ is a bijection $\varphi : V \to \hat{V}$ with $\varphi(\rho) = \hat{\rho}$ such that for $u, v \in V$, $uv \in E$ if and only if $\varphi(u)\varphi(v) \in \hat{E}$. If there is a rooted isomorphism between G and $\hat{G}$ then we write $G \simeq \hat{G}$.

For $r > 0$, write G^r for the induced subgraph of G with vertex set $V^r = \{v \in V : \mathrm{dist}_G(v, \rho) \leq r\}$, still rooted at ρ; then let

$$D(G, \hat{G}) = \frac{1}{\sup(r : G^r \simeq \hat{G}^r)}\,.$$

Exercise 4.6.1. *D satisfies the triangle inequality.*

Write $\mathcal{G}$ for the set of isomorphism-equivalence classes of locally finite rooted graphs. Abusing notation by continuing to write D for the push-forward of the above function to $\mathcal{G}$, then $(\mathcal{G}, D)$ is a complete separable metric space.

As is usual in the setting of probability on complete separable metric spaces, for a sequence $(G_n, 1 \leq n \leq \infty)$ of $\mathcal{G}$-valued random variables, we say that G_n converges in distribution to G_∞, and write $G_n \xrightarrow{\mathrm{d}} G_\infty$, if

$$\mathbf{E} f(G_n) \to \mathbf{E} f(G_\infty)$$

for all bounded continuous functions $f : \mathcal{G} \to \mathbb{R}$.

With this notation, then $\mathrm{T}(n) \in_u \mathscr{T}_n$ indeed converges in distribution to $\mathrm{T}(\infty)$, where the limiting tree is an infinite path to which are attached a sequence of independent Poisson(1) Bienaymé trees.

Exercise 4.6.2. *Let G_n be a uniformly random 3 regular graph with vertex set $[n]$, rooted at vertex 1. Then G_n converges in distribution to an infinite 3-regular tree.*

Exercise 4.6.3. *Let B_n be a complete binary tree with n levels, rooted at a uniformly random vertex. Then B_n converges in distribution to the* canopy tree*; see Figure 4.1.*

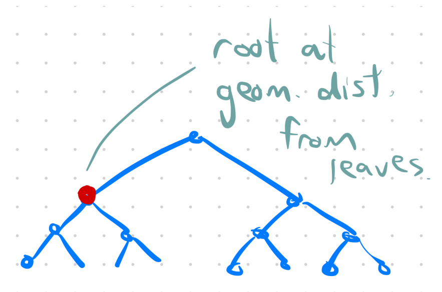


Figure 4.1: A canopy tree

We next turn to weighted graphs. For a weighted rooted graph $G = (V, E, \rho, \mathrm{w})$ and vertices $u, v \in V$, write

$$\mathrm{dist}_g(u,v) := \inf\Big(\sum_{e\in\gamma} \mathrm{w}(e) : \gamma \text{ a path from } u \text{ to } v\Big)\,.$$

Say G is *locally finite* if $V^r := \{v \in V \mathrm{dist}_G(\rho, v)\}$ is finite for all r. Note that we can think of unweighted graphs as weighted graphs in which all edges have weight 1.

Write $G^r = (V^r, E^r, \rho, \mathrm{w}^r)$, where $E^r = \{uv \in E : u, v \in V^r\}$ and $\mathrm{w}^r = (\mathrm{w}(e), e \in E^r)$. Say that locally finite graphs $G = (V, E, \rho, \mathrm{w})$ and $\hat{G} = (\hat{V}, \hat{E}, \hat{\rho}, \hat{\mathrm{w}})$ are *r-close* if there exist $x, y \ge r$ with $|x - y| < 1/r$ and a rooted isomorphism φ of (V^x, E^x, ρ) and $(\hat{V}^y, \hat{E}^y, \hat{\rho})$ such that $|\mathrm{w}(e) - \hat{\mathrm{w}}(\varphi(e))| \le 1/r$ for all $e \in E^x$. Also, say that G and $\hat{G}$ are isomorphic if there exists and isomorphism φ from (V, E, ρ) to $(\hat{V}, \hat{E}, \hat{\rho})$ such thatw$(uv) = \hat{\mathrm{w}}(\varphi(u)\varphi(v))$ for all $uv \in E$.

Let $\mathcal{G}_*$ be the set of isomorphism-equivalence classes of locally finite weighted rooted graphs, and for $G, \hat{G} \in \mathcal{G}_*$, write

$$D_*(G, \hat{G}) = \frac{1}{\sup(r : G \text{ and } \hat{G} \text{ are } r\text{-close})}\,.$$

In this definition, there is again a slight abuse of notation; we should really take G and $\hat{G}$ to be representatives from equivalence classes of $\mathcal{G}_*$, but the definition does not depend on the choice of representatives.

A couple minor issues with the definition at the moment. a) Perhaps there is an issue for values r which are not continuity points for the graph structure of G^r. b) In order to have convergence of of the complete graph to the PWIT, we in fact need to "ignore" high-weight edges (we should e.g. remove them from G^r) before checking whether the graphs are locally isomorphic). This issue is in fact already present in Aldous-Steele (Objective Method); Aldous-Lyons (Processes on Unimodular Random Networks). It doesn't arise in Benjamini-Schramm, because they restrict to locally finite *unweighted* graphs.

Proposition 4.6.3. *$(\mathcal{G}_*, D_*)$ is a complete separable metric space.*

With this definition, we can now state the key local convergence result for the randomly weighted complete graph.

Proposition 4.6.4. *Let G_n be the complete graph K_n, rooted at vertex 1, with edge weights $(nU_e, e \in e(K_n))$, where $(U_i, e \in e(K_n))$ are independent Uniform$[0,1]$ random variables. Then G_n converges in distribution to the Poisson-weighted infinite tree.*

Proof idea. Let $G_n^{(k)}$ be the weighted subgraph of G_n consisting of the k new smallest-weight neighbours of each vertex, out to (unweighted) distance k from vertex 1. More formally, $G_n^{(0)}$ consists only of the vertex 1. Then, for $0 \le i < k$, conditionally given $G_n^{(0)}, \ldots, G_n^{(i)}$, for each vertex $v \in \mathrm{v}(G_n^{(i)} \setminus \mathrm{v}(G_n^{(i-1)})$, let $S(v)$ be the k weighted nearest neighbours of v in $[n] \setminus \mathrm{v}(G_n^{(i-1)})$, and let $G_n^{(i+1)}$ be the induced subgraph of G_n with vertex set

$$\mathrm{v}(G_n^{(i)} \cup \bigcup_{v \in \mathrm{v}(G_n^{(i)} \setminus \mathrm{v}(G_n^{(i-1)})} S(v)\,.$$

Next, define a subtree $\mathrm{PWIT}^{(k)}$ of the PWIT in the same way. In the PWIT this is easier, because the children of a node are already labeled in increasing order of weight; so

$$\mathrm{v}(\mathrm{PWIT}^{(k)}) = \{\varnothing\} \cup \bigcup_{i=1}^{k} [k]^i.$$

For $i < k$, conditionally given $G_n^{(i)}$, for a vertex v of $G_n^{(i)}$ at unweighted distance i from 1, the ranked weights $W_{v,1}, \ldots, W_{v,k}$ on edges from v to $S(v)$ are distributed as the k smallest values among $n - |v(G_n^{(i)})|$ independent random variables distributed as $n \cdot \mathrm{Uniform}[0,1]$. It follows that

$$(W_{v,1}, \ldots, W_{v,k})$$

is asymptotically distributed as the first k points of a rate-one Poisson process, i.e., as the first k weights on the child edges of a vertex of the PWIT. Moreover, conditionally given $G_n^{(i)}$, the sets $\{S(v), v \in v \in \mathrm{v}(G_n^{(i)} \setminus \mathrm{v}(G_n^{(i-1)})\}$ are independent size-k subsets of $[n] \setminus \mathrm{v}(G_n^{(i)})$. There are at most k^i sets in this collection, so with high probability they are all disjoint and therefore $G_n^{(i)}$ is a tree. □

That this fact is useful for the study of minimum spanning trees is witnessed by the following beautiful result of Aldous and Steele[4].

[4] DJ Aldous and J.M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. *Probability on Discrete Structures*, 110:1–72, 2004

Theorem 4.6.5. *Let $(G_n, 1 \le n \le \infty)$ be a sequence of $\mathcal{G}_*$-valued random variables, and write $G_n = (V_n, E_n, \rho_n, \mathrm{w}_n)$. Suppose that for n finite, G_n almost surely has n vertices, and the root ρ_n of G_n is uniformly random. Suppose also that almost surely no two edges of G_∞ have the same weight.*

Under these conditions, if $G_n \xrightarrow{\mathrm{d}} G_\infty$ in $\mathcal{G}_*$ then we have the joint convergence in distribution

$$(G_n, \mathrm{MST}(G_n)) \xrightarrow{\mathrm{d}} (G_\infty, \mathrm{WMSF}(G_\infty))$$

in $\mathcal{G}_* \times \mathcal{G}_*$,

$$\deg_{G_n}(\rho_n) \xrightarrow{\mathrm{d}} \deg_{G_\infty}(\rho_\infty),$$

This theorem doesn't quite make sense as written, because $\mathrm{WMSF}(G_\infty)$ need not be a connected graph. We should really augment the notion of convergence: the space $\mathcal{G}_\infty$ to include "binary" information about edge presence and absence. In other words, the convergence of the second coordinate should be in an augmented space $\mathcal{G}_*^b$, corresponding to local convergence of weighted graphs with distinguished subgraphs; the subgraph can be recorded by marking each edge *where the limiting random degree satisfies* $\mathbf{E}\deg_{G_\infty}(\rho_\infty) = 2$, *and*

$$\sum_{e \in \mathrm{MST}(G_n):\rho_n \in e} \mathrm{w}_n(e) \xrightarrow{\mathrm{d}} \sum_{e \in \mathrm{WMSF}(G_\infty):\rho_\infty \in e} \mathrm{w}_\infty(e)\,.$$

In view of this theorem, in order to understand the asymptotic local structure of the MST of K_n, we are allowed to focus on the structure of the WMSF of the PWIT, and this is what we now do. More specifically, we will focus on the structure of the component of the WMSF containing the root, which we recall has label $\emptyset$. We hereafter denote this component by M. We begin our analysis by focussing on an important subtree of M; the tree T^∞ built by running Prim's algorithm on the PWIT, starting from the root $\emptyset$.

Recall that a graph G is *one-ended* if for any two paths γ, γ' the symmetric difference

So far the definition of one-endedness hasn't actually appeared earlier, so "recall" is inaccurate; but it should appear in the section on Euclidean MSTs earlier, once that is written...

$$\gamma \triangle \gamma' = \{e \in \gamma : e \notin \gamma'\} \cup \{e' \in \gamma' : e' \notin \gamma\}$$

has finite size.

Theorem 4.6.6. *The tree* T^∞ *is one-ended.*

We begin the proof with a lemma.

Lemma 4.6.7. *Let* $\Lambda^0 = \sup(P_e : e \in e(\mathrm{T}^\infty))$*. then* Λ^0 *has density* $\theta'(x)\mathbf{1}_{[x\geq 1]}$*, where* $\theta(x) = \mathbf{P}\{|\mathrm{T}(\mathrm{Poi}(x))| = \infty\}$*. Moreover, almost surely, there is a unique edge of* T^∞ *with weight* Λ^0*.*

Proof. For $u \in \mathcal{U}$ and $x \geq 0$, write

$$\mathrm{T}_u(x) = \{v \in \mathcal{U} : u \preceq v, \text{ all edges on the path from } u \text{ to } v \text{ have weight } \leq x\}.$$

Then $\Lambda^0 \leq x$ if and only if $|\mathrm{T}_\emptyset(x)| = \infty$, and $\mathrm{T}_\emptyset(x)$ is distributed as a Poisson(x) Bienaymé tree; so $\mathbf{P}\left\{\Lambda^0 \leq x\right\} = \theta(x)$. This shows that Λ^0 has the density claimed in the lemma.

To see that a.s. there is an edge with weight Λ^0, we argue as follows. Let $f(1)$ be the smallest weight edge leaving $\mathrm{T}_\emptyset(1)$ and let $X(1) = P_{f(1)}$. Then inductively, for $k \geq 1$, if $\mathrm{T}_\emptyset(X(k))$ is finite then let $f(k+1)$ be the smallest-weight edge leaving $\mathrm{T}_\emptyset(X(k))$ and set $X(k+1) = P_{f(k+1)}$.

Conditionally given $\mathrm{T}_\emptyset(X(k))$, independently for each vertex $v \in \mathrm{T}_\emptyset(X(k))$, the weight of the smallest weight edge from v to $\mathcal{U} \setminus \mathrm{T}_\emptyset(X(k))$ has distribution $X(k) + \mathrm{Exp}(1)$. It follows that conditionally given $\mathrm{T}_\emptyset(X(k))$,

$$X(k+1) \stackrel{\mathrm{d}}{=} X(k) + \mathrm{Exp}(|\mathrm{T}_\emptyset(X(k))|).$$

Moreover, if uv is the unique edge from $\mathrm{T}_\emptyset(X(k))$ to $\mathcal{U} \setminus \mathrm{T}_\emptyset(X(k))$ with weight $X(k+1)$, then $\mathrm{T}_v(X(k+1))$ is distributed as a Poisson$(X(k+$

1)) Bienaymé tree. Since $X(k+1) > X(k) \geq X(0)$, it follows that

$$|\mathrm{T}_\emptyset(X(k+1))| \stackrel{\mathrm{d}}{=} |\mathrm{T}_\emptyset(X(k))| + |\mathrm{T}(\mathrm{Poi}(X(k+1))| \succeq_{\mathrm{st}} |\mathrm{T}_\emptyset(X(k))| + |\mathrm{T}(\mathrm{Poi}(X(0))|.$$

It follows that conditionally given $X(0)$,

$$\sigma = \inf(k : |\mathrm{T}_\emptyset(X(k))| = \infty) \preceq_{\mathrm{st}} \mathrm{Geom}(\theta(X(0))).$$

Since $X(0)$ is almost surely greater than 1, it follows that σ is almost surely finite, and that $f(\sigma)$ is the unique edge of T^∞ of weight $\Lambda^0 = X(\sigma)$. □

Proof of Theorem 4.6.6. Write $e^0 = u^0 v^1$ for the unique weight-Λ^0 edge of T^∞, with u^0 the parent of v^1. Then let

- $\mathrm{T}^{(0)}$ be the component of $\mathrm{T}^\infty - e^0$ containing u^0
- $\mathrm{T}^{\infty,1}$ be the component of $\mathrm{T}^\infty - e^0$ containing v^1.

Note that after exploring e^0, all edges explored by Prim's algorithm lie within $\mathrm{T}^{\infty,1}$. Also, $\mathrm{T}_{v^1}(\Lambda^0)$ is distributed as a Poisson(Λ^0) Bienaymé tree conditioned to survive. Writing

$$\Lambda^1 = \sup(P_e, e \in e(\mathrm{T}^{\infty,1}))\,,$$

it follows that

$$\begin{aligned}
\mathbf{P}\left\{\Lambda^1 \leq x \;\middle|\; \Lambda^0\right\} &= \mathbf{P}\left\{|\mathrm{T}_{v^1}(x)| = \infty \;\middle|\; |\mathrm{T}_{v^1}(\Lambda^0)| = \infty\right\} \\
&= \frac{\mathbf{P}\left\{|\mathrm{T}_{v^1}(x)| = \infty\right\}}{\mathbf{P}\left\{|\mathrm{T}_{v^1}(\Lambda^0)| = \infty\right\}} \\
&= \frac{\theta(x)}{\theta(\Lambda^0)}\,,
\end{aligned}$$

so conditionally given Λ^0, the random variable Λ^1 has density

$$\frac{\theta'(x)}{\theta(\Lambda^0)}\mathbf{1}_{[x \leq \Lambda^0]}.$$

It also follows as in the lemma that $\mathrm{T}_{v^1}(\Lambda^1)$ contains a unique edge $e^1 = u^1 v^2$ of weight Λ^1.

Prove a more general version of the lemma, so that it actually follows *from* the lemma rather than *as in* the lemma.

Continuing in this way, we may inductively define a random sequence of edges $(e^k, k \geq 0)$, with $e^k = u^k v^{k+1}$ and u^k the parent of v^{k+1}, such that with

- $\mathrm{T}^{(k)}$ the component of $\mathrm{T}^\infty - \{e^0, \ldots, e^k\}$ containing u_k
- $\mathrm{T}^{\infty,k+1}$ the component of $\mathrm{T}^\infty - \{e^0, \ldots, e^k\}$ containing v^{k+1},

then

$$\Lambda^k := P_{e^k} = \max(P_e, e \in e(\mathrm{T}^{\infty,k})),$$

all edges of $T^{\infty,k}$ have weight strictly less than Λ^{k-1}, and u^k is a descendent of v^k for all $k \geq 1$.

This implies that for each $k \geq 0$, Prim's algorithm first explores $T^{(k)}$, then explores edge e^k, and after exploring e^k, all vertices added by Prim's algorithm are in $T^{\infty,k+1}$, i.e., they are descendants of v^{k+1}. This in particular implies that T^∞ is one-ended. $\square$

Corollary 4.6.8. *For any $x > 1$, T^∞ almost surely contains at most finitely many edges of weight greater than x.*

Proof. The proof of the theorem shows that the sequence $(\Lambda_i, i \geq 0)$ is a Markov process with transition density

$$p(\ell, x) = \frac{\theta'(x)}{\theta(\ell)} \mathbf{1}_{[x \leq \ell]} \, .$$

It follows that $\Lambda_i \to 1$ almost surely. Since if $\Lambda^i < x$ then all edges of weight $\geq x$ belong to $T^{(0)}, \ldots, T^{(i)}$, the result follows. $\square$

Exercise 4.6.4. *Prove that for all $k \geq 0$, the following all hold.*

(a) Conditionally given that $\Lambda^k = h$, the random variable Λ^{k+1} has density $(\theta'(y)/\theta(h))\mathbf{1}_{[1 \leq y \leq h]}$.

(b) Prove that

$$\mathbf{P}\left\{|T^{(k+1)}| = m \;\middle|\; \Lambda^{k+1} = y\right\} = \frac{\theta(y)}{\theta'(y)} \frac{e^{-my}(my)^{m-1}}{(m-1)!} = \frac{\theta(y)}{\theta'(y)} \cdot m\mathbf{P}\left\{|T(\mathrm{Poi}(\lambda))| = m\right\} \, .$$

- *Write $F_y(x) = \mathbf{P}\left\{|T^{(k+1}| \leq x/(y-1)^2 \;\middle|\; \Lambda^{k+1} = y\right\}$. Show that $F_y(x) \to F(x)$ as $y \downarrow 1$, where $F(x)$ is a cumulative distribution function. Describe $F(x)$.*

It follows with not too much work from the third part of the exercise exercise that T^∞ has quadratic volume growth.

Elaborate ... in probability? In expectation?

Having described T^∞, we now turn to explaining how to build M from T^∞.

To understand the relation between M and T^∞, we define an iterative way to construct M from T^∞ via an aggregation process, in which finite trees attach themselves to T^∞ (or to other finite trees which are already part of the aggregate).

Every vertex $u \in T^\infty$ lies in one of the finite trees $T^{(i)}$; if $u \in T^{(i)}$ then write $\Lambda(u) = \Lambda^i$. This is the weight of the largest-weight edge added by Prim's algorithm when building T^∞ after u is added.

Next, for $v \in \mathcal{U}$ let

$$\tau(v) = \sup(\lambda : |T_v(\lambda)| < \infty).$$

If uv is an edge of $\mathcal{U}$ with $u \in T^\infty$ and $v \notin T^\infty$, then $\mathrm{w}(uv) > \Lambda(u)$. Now remember that uv is an edge of M if and only if, in the subgraph of $\mathcal{U}$ obtained by keeping only edges of weight less than $\mathrm{w}(uv)$, the components containing u and v are both finite. Since $\mathrm{w}(uv) > \Lambda(u)$, the component containing u is infinite (it contains all but finitely many elements of T^{∞}), so

- if $|\tau(v)| < \mathrm{w}(uv)$ then $|\mathrm{T}_v(\mathrm{w}(uv))| = \infty$ and $uv \notin e(\mathrm{M})$, and
- if $|\tau(v)| > \mathrm{w}(uv)$ then $\mathrm{T}_v(\mathrm{w}(uv))| < \infty$ and $uv \in e(\mathrm{M})$.

Now let $\mathrm{M}^0 = \mathrm{T}^{\infty}$, write

$$\partial \mathrm{M}^0 = \{uv \in e(\mathcal{U}) : u \in \mathrm{M}^0, v \notin \mathrm{M}^0\},$$

and let

$$\mathrm{M}^1 = \mathrm{T}^{\infty} \cup \{\mathrm{T}_v(\mathrm{w}(uv)) : uv \in \partial \mathrm{M}^0, \tau(v) > \mathrm{w}(uv)\}.$$

Inductively, given M^k, set

$$\mathrm{M}^{k+1} = \mathrm{M}^k \cup \{\mathrm{T}_v(\mathrm{w}(uv)) : uv \in \partial M^k : \tau(v) > \mathrm{w}(uv)\}.$$

Then

$$\mathrm{M} = \mathrm{M}^{\infty} := \lim_{k \to \infty} \mathrm{M}^k.$$

Writing M_v for the subtree of M rooted at v, when v is a "boundary vertex" - the root of one of the trees added when passing from M^{ℓ} to $\mathrm{M}^{\ell+1}$ for some ℓ - then we can describe the expected size of M_v conditional on the weight of the boundary edge.

Proposition 4.6.9. *For $\ell \geq 0$, for $uv \in \partial \mathrm{M}^{\ell}$, for any $\lambda > 1$,*

$$\mathbf{E}\{|\mathrm{M}_v| \mid \mathrm{w}(uv) = \lambda\} = \sum_{k \geq 1} \int_{\lambda < \lambda_1 < \ldots < \lambda_k} \prod_{i=1}^{k} \frac{\exp(-(\lambda_i - \hat{\lambda}_i))}{1 - \hat{\lambda}_i} \mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_k.$$

We omit the proof of this result as it is somewhat involved. Informally, the sum indexes the "number of steps" of the aggregation process. The numerator in the product accounts for the probability that a tree at the far end of a boundary edge of weight λ_i is finite, and the denominator gives the expected size of such a tree given that it is finite. (Recall: expected size of $\mathrm{T}(\mathrm{Poi}(\lambda)) = 1/(1-\lambda)$ for $\lambda < 1$.)

I should add it in later though!

Corollary 4.6.10. M *is almost surely one-ended.*

Proof. Since $\partial \mathrm{M}^0$ almost surely has only countably many edges, it suffices to show that for each $uv \in \partial \mathrm{M}^0$, almost surely $|\mathrm{M}_v| < \infty$. This in turn follows if

$$\sum_{k \geq 1} \int_{\lambda < \lambda_1 < \ldots < \lambda_k} \prod_{i=1}^{k} \frac{\exp(-(\lambda_i - \hat{\lambda}_i))}{1 - \hat{\lambda}_i} \mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_k$$

is finite for all $\lambda > 1$, since the weight of any boundary edge is strictly greater than 1 almost surely.

Fix $\lambda > 1$, and write S_k for the k'th summand in the above sum. This is in fact fairly straightforward to show: since $\hat{\lambda}_i$ decreases as λ_i increases, we have

$$\begin{aligned} S_k &\leq \frac{1}{k!}\frac{1}{(1-\hat{\lambda})^k}\int_{(\lambda_1,\ldots,\lambda_k)\in[\lambda,\infty)} \exp(-\sum_{i=1}^{k}(\lambda_i-\hat{\lambda}_i))\mathrm{d}\lambda_1\ldots\mathrm{d}\lambda_k \\ &\leq \frac{1}{k!}\frac{1}{(1-\hat{\lambda})^k}\exp(-k(\lambda-\hat{\lambda}))\,, \end{aligned}$$

and so

$$\sum_{k\geq 1} S_k \leq \sum_{k\geq 1}\frac{1}{k!}\frac{1}{(1-\hat{\lambda})^k}\exp(-k(\lambda-\hat{\lambda})\,,$$

which is indeed finite, as required. □

It is possible to derive more detailed information about M.[5]

[5] See "The local weak limit of the minimum spanning tree of the complete graph", https://arxiv.org/abs/1301.1667.

Theorem 4.6.11. M *has cubic volume growth: almost surely*

$$\frac{\log|\{u\in \mathrm{M}:|u|\leq r\}|}{\log r}\to 3.$$

For the upper bound in the proof, one may analyze the above aggregation process, pretending that the trees of $\mathrm{M}\setminus\mathrm{M}^0$ each have diameter 1, using the formula for the sizes of trees of $\mathrm{M}\setminus\mathrm{M}^0$. The lower bound is more involved: it uses an analysis of Prim's algorithm on K_n, plus the fact that the minimum spanning tree of K_n converges in distribution to M.

In fact, a more quantitative version of Theorem 4.6.11 is known; it's possible to show that there exists $C > 0$ so that almost surely

$$\frac{r^3}{\log^C r}\leq|\{u\in\mathrm{M}:|u|\leq r\}|\leq r^3\exp(C\sqrt{\log r})\,,$$

for all r sufficiently large.

Conjecture 1. $\frac{|\{u\in\mathrm{M}:|u|\leq r\}|}{r^3}$ *converges in distribution.*

Martingale limits and changes of measure

The next theorem describes how the dichotomy between absolute continuity and mutual singularity of measures manifests when observed along a filtration.

Theorem 4.6.12. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a measurable space and let $\mathbf{Q}$ be a finite measure on $(\Omega, \mathcal{F})$. Fix an increasing sequence of sub-σ-field $(\mathcal{F}_n)_{n\geq 1}$ with $\sigma(\bigcup_n \mathcal{F}_n) = \mathcal{F}$. Write $\mathbf{P}_n := \mathbf{P}|_{\mathcal{F}_n}$ and $\mathbf{Q}_n := \mathbf{Q}|_{\mathcal{F}_n}$. Suppose that $\mathbf{Q}_n \ll \mathbf{P}_n$ for all n, and write $X_n = d\mathbf{Q}_n/d\mathbf{P}_n : \Omega \to [0,\infty)$ for the corresponding Radon-Nikodym derivatives. Then setting $X = \limsup_{n\to\infty} X_n$, it holds that*

$$\mathbf{Q} = X\mathbf{P} + \mathbf{Q}\mathbf{1}_{[X=\infty]}. \tag{4.6.1}$$

Exercise 4.6.5. *In the notation of Theorem 4.6.12, show that $(X_n, n \geq 1)$ is an $\mathcal{F}_n$-martingale for $\mathbf{P}$.*

Remark. Since the X_n are non-negative, Exercise 4.6.5 implies that X_n converges $\mathbf{P}$-almost surely, so we must have $\mathbf{P}\{\lim_{n\to\infty} X_n = X\} = 1$. But it is standard that if $Z_n \xrightarrow{\mathrm{p}} Z_\infty$ and $\mathbf{E}|Z_n| < \infty$ for all n, then $\mathbf{E}|Z_n| \to \mathbf{E}|Z|$ if and only if $(Z_n, n \geq 1)$ is uniformly integrable. This means that there is another equivalent property which may be added to (1) in Theorem 4.6.12: that $(X_n, n \geq 1)$ is $\mathbf{P}$-uniformly integrable.

Exercise 4.6.6. *Let $(X_n, 1 \leq n \leq \infty)$ be random variables in $L_1(\Omega, \mathcal{F}, \mathbf{P})$ such that $X_n \xrightarrow{\mathrm{p}} X_\infty$. Prove that the following are equivalent: (a) $X_n \xrightarrow{L_1} X_\infty$; (b) $(X_n, 1 \leq n \leq \infty)$ is uniformly integrable; (c) $\mathbf{E}|X_n| \to \mathbf{E}|X_\infty|$.*

Lemma 4.6.13. *In the setting of Theorem 4.6.12, if $\mathbf{Q} \ll \mathbf{P}$ then $\mathbf{Q} = X\mathbf{P}$.*

Recall that $\mathbf{Q} = X\mathbf{P}$ is shorthand for the statement that for all $E \in \mathcal{F}$,

$$\mathbf{Q}(E) = \int_E d\mathbf{Q} = \int_E X d\mathbf{P} = \mathbf{E}_{\mathbf{P}}\left\{X\mathbf{1}_{[E]}\right\}.$$

Proof. First suppose $\mathbf{Q}$ is absolutely continuous with respect to $\mathbf{P}$. Then the Radon-Nikodým derivative $\tilde{X} = d\mathbf{Q}/d\mathbf{P}$ exists and satisfies $\mathbf{Q}(\tilde{X} = \infty) = 0$, so we just want to show that for all $E \in \mathcal{F}$,

$$\mathbf{Q}(E) = \mathbf{E}_{\mathbf{P}}\left\{X\mathbf{1}_{[E]}\right\}.$$

For all $E \in \mathcal{F}$, by the definition of the Radon-Nikodým derivative,

$$\mathbf{E}_{\mathbf{Q}}\left\{\mathbf{1}_{[E]}\right\} = \int_E 1 d\mathbf{Q} = \int_E \tilde{X} d\mathbf{P} = \mathbf{E}_{\mathbf{P}}\left\{\tilde{X}\mathbf{1}_{[E]}\right\}. \tag{4.6.2}$$

For all $E \in \mathcal{F}_n$, we also have

$$\begin{aligned}
\mathbf{E}_{\mathbf{P}}\left\{X_n \mathbf{1}_{[E]}\right\} &= \int_E X_n d\mathbf{P} \\
&= \int_E X_n d\mathbf{P}_n && \text{(Homework)} \\
&= \int_E 1 d\mathbf{Q}_n && \text{(Since } X_n = d\mathbf{Q}_n/d\mathbf{P}_n\text{)} \\
&= \int_E 1 d\mathbf{Q} \\
&= \mathbf{E}_{\mathbf{Q}}\left\{\mathbf{1}_{[E]}\right\},
\end{aligned}$$

so X_n is a version of $\mathbf{E}\left[\tilde{X} \mid \mathcal{F}_n\right]$. Since $\mathcal{F}_\infty := \sigma(\bigcup_{n\to\infty} \mathcal{F}_n) = \mathcal{F}$, the non-negative martingale convergence theorem then gives that $\mathbf{P}$-almost surely

$$X_n \to \mathbf{E}\left[\tilde{X} \mid \mathcal{F}_\infty\right] = \mathbf{E}\left[\tilde{X} \mid \mathcal{F}\right] = \tilde{X}.$$

But $X = \limsup_{n\to\infty} X_n$ by definition, so $\mathbf{P}$-almost surely $X = \tilde{X}$. Thus $\mathbf{E}_{\mathbf{P}}\left\{\tilde{X}\mathbf{1}_{[E]}\right\} = \mathbf{E}_{\mathbf{P}}\left\{X\mathbf{1}_{[E]}\right\}$, and the result follows from (4.6.2). □

Proof of Theorem 4.6.12. Let π be the average of $\mathbf{P}$ and $\mathbf{Q}$, so $\pi(E) = (\mathbf{P}(E) + \mathbf{Q}(E))/2$ for $E \in \mathcal{F}$. For $n \geq 1$ let $\pi_n = \pi|_{\mathcal{F}_n} = (\mathbf{P}_n + \mathbf{Q}_n)/2$. Then both $\mathbf{P}$ and $\mathbf{Q}$ are absolutely continuous with respect to π, and likewise $\mathbf{P}_n$ and $\mathbf{Q}_n$ are absolutely continuous with respect to π_n for all $n \geq 1$.

Write $U_n = d\mathbf{Q}_n/d\pi_n$ and $V_n = d\mathbf{P}_n/d\pi_n$ and let $U = \limsup_{n\to\infty} U_n$ and $V = \limsup_{n\to\infty} V_n$. Since $\mathbf{Q} \ll \pi$ it follows by Lemma 4.6.13 (applied with π in place of $\mathbf{P}$) that π-almost surely $U_n \to U$ and that $\mathbf{Q} = U\pi$. Likewise, applying Lemma 4.6.13 with π in place of $\mathbf{P}$ and $\mathbf{P}$ in place of $\mathbf{Q}$, it follows that π-almost surely $V_n \to V$ and that $\mathbf{P} = V\pi$.

Next, π-almost surely we have

$$U_n + V_n = \frac{d\mathbf{Q}_n}{d\pi_n} + \frac{d\mathbf{P}_n}{d\pi_n} = 2\frac{d\pi_n}{d\pi_n} = 2.$$

It follows that

$$\pi(U + V = 0) = \pi(\limsup_n (U_n + V_n) = 0) = 0,$$

so π-almost surely, U/V is well-defined (and equal to ∞ if $U = \infty$

and $V = 0$), and

$$\begin{aligned}\frac{U}{V} &= \frac{\lim_{n\to\infty} U_n}{\lim_{n\to\infty} V_n}\\ &= \lim_{n\to\infty} \frac{U_n}{V_n}\\ &= \lim_{n\to\infty} X_n && \text{(chain rule)}\\ &= X.\end{aligned}$$

Finally, we already know $\mathbf{Q} = U\pi$ and $\mathbf{P} = V\pi$. We may also write $U = XV + U\mathbf{1}_{[V=0]} = XV + U\mathbf{1}_{[X=\infty]}$, so

$$\mathbf{Q} = U\pi = XV\pi + \mathbf{1}_{[X=\infty]}U\pi = X\mathbf{P} + \mathbf{1}_{[X=\infty]}\mathbf{Q}\,,$$

as claimed. □

Corollary 4.6.14. *In the setting of Theorem 4.6.12, we have the following.*

1. $\mathbf{Q} \ll \mathbf{P} \Leftrightarrow \mathbf{Q}(X = \infty) = 0 \Leftrightarrow \mathbf{E_P}X = 1.$

2. $\mathbf{Q} \perp \mathbf{P} \Leftrightarrow \mathbf{Q}(X = \infty) = 1 \Leftrightarrow \mathbf{E_P}X = 0.$

Proof. If $\mathbf{Q} \ll \mathbf{P}$ then by Lemma 4.6.13 we have $\mathbf{Q} = X\mathbf{P}$ so clearly $\mathbf{Q}(X = \infty) = 0$. We now repeatedly use (4.6.1). If $\mathbf{Q}(X = \infty) = 0$ then by (4.6.1) we have

$$\mathbf{E_P}\{X\} = \mathbf{E_Q}\{1\} - \mathbf{E_Q}\left\{\mathbf{1}_{[X=\infty]}\right\} = 1.$$

If $\mathbf{E_P}\{X\} = 1$ then again by (4.6.1), $\mathbf{Q}(X = \infty) = 0$ so $\mathbf{Q} = X\mathbf{P}$ and thus $\mathbf{Q} \ll \mathbf{P}$. This proves the first line of equivalences of the theorem.

Note that by Exercise 4.6.5, X_n is an $\mathcal{F}_n$-martingale for $\mathbf{P}$ so $\mathbf{E_P}\{X\} \le \liminf_{n\to\infty} \mathbf{E_P}\{X_n\} < \infty$. It follows that $\mathbf{P}(X = \infty) = 0$.

If $\mathbf{Q} \perp \mathbf{P}$ then $\mathbf{Q}$ has no absolutely continuous part with respect to $\mathbf{P}$. On the other hand, $X\mathbf{P} \ll \mathbf{P}$, so by (4.6.1) we must have $\mathbf{Q} = \mathbf{1}_{[X=\infty]}\mathbf{Q}$; this in turn implies that $\mathbf{Q}(X = \infty) = 1$.

If $\mathbf{Q}(X = \infty) = 1$ then by (4.6.1), $\mathbf{E_P}\{X\} = \int X d\mathbf{P} = X\mathbf{P} = 0$. Finally, if $X\mathbf{P} = 0$ then by (4.6.1) we have $\mathbf{Q}(X = \infty) = 1$. But $\mathbf{P}X = \infty = 0$, which implies $\mathbf{Q} \perp \mathbf{P}$. □

Bibliography

DJ Aldous and J.M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. *Probability on Discrete Structures*, 110:1–72, 2004.