An introduction to random trees

Christina Goldschmidt
Department of Statistics and Lady Margaret Hall,
University of Oxford

CIMPA-DAAD Research School on Stochastic Processes and their Applications,
National University of Mongolia, Ulaanbaatar, Mongolia, 27th July–7th August 2015

[Version of 28th July 2015]

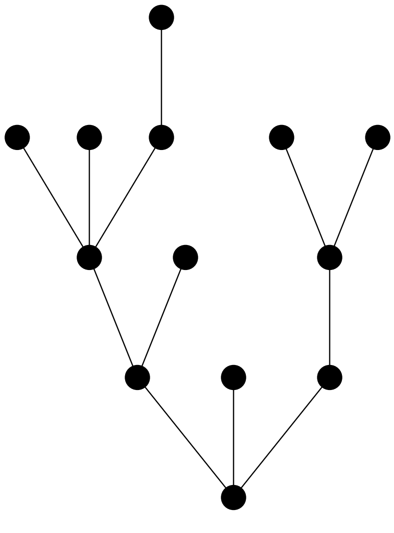
Abstract. Tree structures are very common in nature, and we often want to model trees which are somehow randomly generated. In mathematics, we think of trees as connected graphs (or networks) with no cycles. In these lectures, I will discuss some simple models of random trees and what we can say about large instances of them. The first lecture will focus on Galton-Watson branching processes. We think of an idealised biological population (say of cells) in which each individual lives for a unit of time, gives birth to a random number of children (with a given distribution) and then dies. What can we say about the way the population size evolves? Or about the family tree of that population? In the second lecture, we will take a more combinatorial perspective. What does a tree with n vertices, chosen uniformly at random from all the possibilities, look like for large n? It turns out that we can use Galton-Watson branching processes to help us answer this question. Along the way, we will encounter several beautiful bits of combinatorics and probability, including Cayley's formula, random walks and (finally) Brownian motion.

Chapter 1

Branching processes

1.1 Introduction

Suppose we have a population (say of bacteria). Each individual in the population lives a unit time and, just before dying, gives birth to a random number of children in the next generation. This number of children has probability mass function $p(k)$, $k \geq 0$, called the *offspring distribution* i.e. $p(k)$ is the probability that an individual has k children. Different individuals reproduce independently in the same manner. Here is a possible family tree of such a population:



We start at the bottom of the tree, with a single individual in generation 0. Then there are 3 individuals in generations 1 and 2, 5 individuals in generation 3, a single individual in generation 4 and no individuals in subsequent generations. (Notice that in this case we must have $p(0) > 0$!)

This model is usually attributed to Galton and Watson (1874) and so is often called a *Galton-Watson branching process*, although it was independently studied (much earlier) by Bienaymé (1845). As a simple model of the evolution of a population, it has proved to be enormously powerful. It is also one of the fundamental processes of probability theory. For much more on the topics in this lecture, see Athreya and Ney [4] or Shi [9].

1.2 Generating functions

Let X_n be the size of the population in generation n, so that $X_0 = 1$. Let $C_i^{(n)}$ be the number of children of the ith individual in generation $n \geq 0$, so that we may write

$$X_{n+1} = C_1^{(n)} + C_2^{(n)} + \cdots + C_{X_n}^{(n)}.$$

(We interpret this sum as 0 if $X_n = 0$.) Note that $C_1^{(n)}, C_2^{(n)}, \ldots$ are independent and identically distributed. Consider the probability generating functions $G(s) = \sum_{k=0}^{\infty} p(k) s^k$ (for s such that the right-hand side is absolutely convergent) and $G_n(s) = \mathbb{E}\left[s^{X_n}\right]$ for $n \geq 0$.

Proposition 1.1. *For $n \geq 0$,*

$$G_{n+1}(s) = G_n(G(s)) = \underbrace{G(G(\ldots G}_{n+1 \text{ times}}(s)\ldots)) = G(G_n(s)).$$

Proof. Since $X_0 = 1$, we have $G_0(s) = s$. Also, we get $X_1 = C_1^{(0)}$ which has probability mass function $p(k)$, $k \geq 0$. So $G_1(s) = \mathbb{E}\left[s^{X_1}\right] = G(s)$. Since

$$X_{n+1} = \sum_{i=1}^{X_n} C_i^{(n)},$$

we get

$$\begin{aligned} G_{n+1}(s) = \mathbb{E}\left[s^{X_{n+1}}\right] &= \mathbb{E}\left[s^{\sum_{i=1}^{X_n} C_i^{(n)}}\right] \\ &= \sum_{k=0}^{\infty} \mathbb{E}\left[s^{\sum_{i=1}^{X_n} C_i^{(n)}} \middle| X_n = k\right] \mathbb{P}\left(X_n = k\right) \end{aligned}$$

and, since $C_1^{(n)}, C_2^{(n)}, \ldots$ are independent of X_n,

$$= \sum_{k=0}^{\infty} \mathbb{E}\left[s^{\sum_{i=1}^{k} C_i^{(n)}}\right] \mathbb{P}\left(X_n = k\right).$$

Since $C_1^{(n)}, C_2^{(n)}, \ldots$ are independent and identically distributed, this equals

$$\begin{aligned} &= \sum_{k=0}^{\infty} \mathbb{E}\left[s^{C_1^{(n)}}\right]^k \mathbb{P}\left(X_n = k\right) \\ &= \sum_{k=0}^{\infty} (G(s))^k \mathbb{P}\left(X_n = k\right) \\ &= G_n(G(s)). \end{aligned}$$

Hence, by induction, for $n \geq 1$,

$$G_n(s) = \underbrace{G(G(\ldots G}_{n \text{ times}}(s)\ldots)) = G(G_n(s)). \qquad \square$$

Corollary 1.2. *Suppose that the mean number of children of a single individual is* μ *i.e.* $\sum_{k=1}^{\infty} kp(k) = \mu$. *Then*

$$\mathbb{E}\left[X_n\right] = \mu^n.$$

Proof. We have $\mathbb{E}\left[X_n\right] = G_n'(1)$. By the chain rule,

$$G_n'(s) = \frac{d}{ds} G(G_{n-1}(s)) = G_{n-1}'(s) G'(G_{n-1}(s)).$$

Plugging in $s = 1$, we get

$$\mathbb{E}\left[X_n\right] = \mathbb{E}\left[X_{n-1}\right] G'(1) = \mathbb{E}\left[X_{n-1}\right] \mu = \cdots = \mu^n. \qquad \square$$

Exercise 1.3. *Show that if* $\sigma^2 := \operatorname{var}(X_1) < \infty$ *then*

$$\operatorname{var}(X_n) = \begin{cases} \frac{\sigma^2 \mu^{n-1}(\mu^n - 1)}{\mu - 1} & \text{if } \mu \neq 1 \\ n\sigma^2 & \text{if } \mu = 1. \end{cases}$$

Corollary 1.2 tells us that we get exponential growth on average if $\mu > 1$ and exponential decrease if $\mu < 1$. This raises an interesting question: can the population die out?

1.3 Extinction

In this section, we investigate the *extinction probability*:

$$q := \mathbb{P}\left(\text{population dies out}\right).$$

If $p(0) = 0$ then every individual has at least one child and so the population clearly grows forever. If $p(0) > 0$, on the other hand, then the population dies out with positive probability because

$$q \geq \mathbb{P}\left(X_1 = 0\right) = p(0) > 0.$$

Notice that this holds even in the cases where $\mathbb{E}\left[X_n\right]$ grows as $n \to \infty$.

Proposition 1.4. *We have*

$$q = \lim_{n \to \infty} \mathbb{P}\left(X_n = 0\right).$$

Proof. The event that the population dies out can be written as

$$\bigcup_{m=1}^{\infty} \{X_m = 0\}.$$

Notice that the events in this union are *not* disjoint: indeed, $X_m = 0$ implies that $X_n = 0$ for all $n > m$. It follows that

$$\bigcup_{m=1}^{n} \{X_m = 0\} = \{X_n = 0\}.$$

We have that $\bigcup_{m=1}^{n} \{X_m = 0\} \subseteq \bigcup_{m=1}^{n+1} \{X_m = 0\} \subseteq \bigcup_{m=1}^{\infty} \{X_m = 0\}$ and so by monotone convergence,

$$\mathbb{P}\left(\bigcup_{m=1}^{\infty} \{X_m = 0\}\right) = \lim_{n \to \infty} \mathbb{P}\left(\bigcup_{m=1}^{n} \{X_m = 0\}\right) = \lim_{n \to \infty} \mathbb{P}\left(X_n = 0\right). \qquad \square$$

Example 1.5. *Suppose that $p(k) = (1/2)^{k+1}$, $k \geq 0$, so that each individual has a geometric number of offspring. We'll find the distribution of X_n.*

First calculate

$$G(s) = \sum_{k=0}^{\infty} s^k \left(\frac{1}{2}\right)^{k+1} = \frac{1}{2-s}.$$

By plugging this into itself a couple of times, we get

$$G_2(s) = \frac{2-s}{3-2s}, \quad G_3(s) = \frac{3-2s}{4-3s}.$$

A natural guess is that $G_n(s) = \frac{n-(n-1)s}{(n+1)-ns}$ which is, in fact, the case, as can be proved by induction. If we want the probability mass function of X_n, we need to expand this quantity out in powers of s. We have

$$\frac{1}{(n+1)-ns} = \frac{1}{(n+1)} \frac{1}{(1-ns/(n+1))} = \sum_{k=0}^{\infty} \frac{n^k s^k}{(n+1)^{k+1}}.$$

Multiplying by $n-(n-1)s$, we get

$$G_n(s) = \sum_{k=0}^{\infty} \frac{n^{k+1}s^k}{(n+1)^{k+1}} - \sum_{k=1}^{\infty} \frac{n^{k-1}(n-1)s^k}{(n+1)^k} = \frac{n}{n+1} + \sum_{k=1}^{\infty} \frac{n^{k-1}s^k}{(n+1)^{k+1}}.$$

We can read off the coefficients now to see that

$$\mathbb{P}(X_n = k) = \begin{cases} \frac{n}{n+1} & \text{if } k = 0 \\ \frac{n^{k-1}}{(n+1)^{k+1}} & \text{if } k \geq 1. \end{cases}$$

Notice that $\mathbb{P}(X_n = 0) \to 1$ as $n \to \infty$, which indicates that the population dies out eventually in this case.

Exercise 1.6. *Suppose now that $p(k) = p(1-p)^k$, $k \geq 0$, for some $p \in (0,1)$.*

- *(a) Find the probability generating function $G(s)$ of this distribution. What is its mean, μ?*
- *(b) Suppose that $p \neq 1/2$. Prove, by induction, that the probability generating function of X_n is*

  $$G_n(s) = p\frac{[(1-p)^n - p^n] - (1-p)s[(1-p)^{n-1} - p^{n-1}]}{[(1-p)^{n+1} - p^{n+1}] - (1-p)s[(1-p)^n - p^n]},$$

  for $n \geq 1$.
- *(c) Find $G_n(0) = \mathbb{P}(X_n = 0)$ and thence show that $q = \min\left\{\frac{p}{1-p}, 1\right\}$. Deduce that the population always dies out if $\mu \leq 1$ whereas it has positive probability of surviving forever if $\mu > 1$.*

Let's now return to the general case. We can find an equation satisfied by q by conditioning on the number of children of the first individual.

$$q = \sum_{k=0}^{\infty} \mathbb{P}(\text{population dies out}|X_1 = k)\,\mathbb{P}(X_1 = k) = \sum_{k=0}^{\infty} \mathbb{P}(\text{population dies out}|X_1 = k)\,p(k).$$

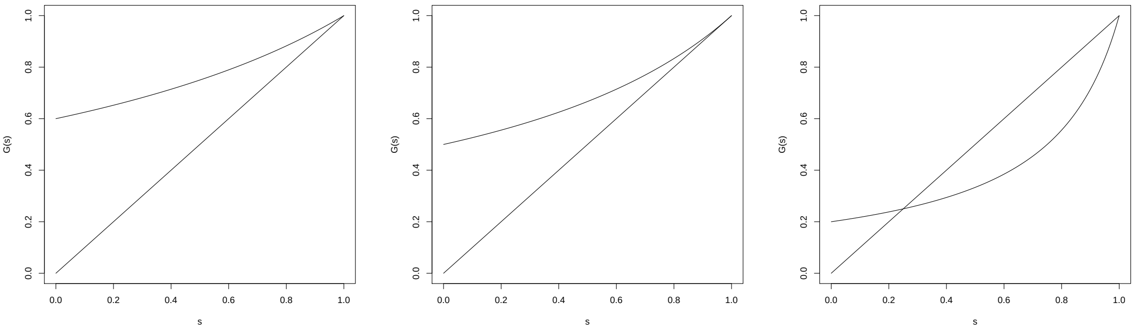


Figure 1.1: Solving $s = G(s)$ for $\mu < 1$, $\mu = 1$ and $\mu > 1$ respectively.

Remember that each of the k individuals in the first generation behaves exactly like the parent. In particular, we can think of each of them starting its own family, which is an independent copy of the original family. Moreover, the whole population dies out if and only if all of these sub-populations die out. If we had k families, this occurs with probability q^k. So

$$q = \sum_{k=0}^{\infty} q^k p(k) = G(q).$$

Exercise 1.7. *Check that for the offspring distribution in Exercise 1.6 the extinction probability q you obtained does, indeed, solve $s = G(s)$.*

The equation $q = G(q)$ doesn't quite enable us to determine q: notice that 1 is always a solution, but it's not necessarily the only solution in $[0, 1]$. It turns out that when there are multiple solutions, the one we want is the *smallest* one in $[0, 1]$.

Theorem 1.8. *Suppose that $p(0) > 0$ and $p(0) + p(1) < 1$.*

(a) The equation $s = G(s)$ has at most two solutions in $[0, 1]$. The extinction probability q is the smallest non-negative root of the equation $G(s) = s$.

(b) Suppose that the offspring distribution has mean μ. Then

- *if $\mu \leq 1$ then $q = 1$;*
- *if $\mu > 1$ then $q < 1$.*

Proof. We observe that $G(0) = p(0) > 0$, $G(1) = 1$ and that $\mu = G'(1)$ (or $G'(1-)$ if G has radius of convergence precisely 1). We also have

$$G''(s) = \sum_{j=2}^{\infty} j(j-1)p(j)s^{j-2} \quad \text{for } 0 \leq s < 1.$$

Since $p(0) + p(1) < 1$, we have $G''(s) > 0$ for $s \in [0, 1)$. Hence, G is strictly convex.

Consider the case $\mu \leq 1$. Then $G(0) > 0$ and $G'(s) \leq G'(1) = \mu \leq 1$ for all $s \in [0,1]$. It follows that $G(s) > s$ for all $s \in [0,1)$ and so the unique root in $[0,1]$ is $s = 1$, which must necessarily be q. See Figure 1.1.

Consider now the case $\mu > 1$. Then since $G(0) > 0$, $G(1) = 1$, $G'(1) > 1$ and G is strictly convex, there exists precisely one root α of $s = G(s)$ in $[0,1)$. See Figure 1.1. Now

$$G(0) \leq G(\alpha) = \alpha.$$

So $G(G(0)) \leq G(\alpha) = \alpha$ and, iterating, $G_n(0) \leq \alpha$ for all $n \geq 1$. It follows that $q = \lim_{n\to\infty} G_n(0) \leq \alpha$. But since q must be a root of $G(s) = s$, we deduce that $q = \alpha$. □

We refer to the case $\mu < 1$ as *subcritical*, the case $\mu = 1$ as *critical* and the case $\mu > 1$ as *supercritical.*

1.4 Martingales

We have just seen that, in the subcritical and critical cases, we have $X_n \to 0$ as $n \to \infty$. In the supercritical case, on the other hand, there is positive probability that $X_n \not\to 0$. We also have that the population grows exponentially *on average.* What can we say almost surely?

For $n \geq 0$, define $W_n = \mu^{-n} X_n$ and let $\mathcal{F}_n = \sigma(X_m, 0 \leq m \leq n)$.

Proposition 1.9. *$(W_n, n \geq 0)$ is a non-negative martingale (with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$) and so*

$$W := \lim_{n\to\infty} W_n$$

exists almost surely.

Proof. We have already shown that $\mathbb{E}[X_n] = \mu^n$ so that $\mathbb{E}[W_n] = 1$ for all $n \geq 0$. Hence, W_n is integrable. Moreover,

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] = \mu^{-(n+1)}\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mu^{-(n+1)}\mathbb{E}\left[\sum_{i=1}^{X_n} C_i^{(n)}\middle|\mathcal{F}_n\right] = \mu^{-1}W_n\mathbb{E}\left[C_1^{(n)}\right],$$

since X_n is $\mathcal{F}_n$-measurable, $C_i^{(n)}, i \geq 1$ are independent of $\mathcal{F}_n$ and are identically distributed. Since $\mathbb{E}\left[C_1^{(n)}\right] = \mu$, we get

$$\mathbb{E}[W_{n+1}|\mathcal{F}_n] = W_n,$$

as required. The convergence result now follows from the almost sure martingale convergence theorem. □

Proposition 1.9 tells us that, in a rough sense, $W_n \sim \mu^n W$. Using Fatou's lemma, we obtain that $\mathbb{E}[W] \leq \liminf_{n\to\infty} \mathbb{E}[W_n] = 1$. It is, however, still possible that $W \equiv 0$. (Indeed, Proposition 1.9 holds even if $\mu \leq 1$, in which case we *must* have $W \equiv 0$.)

Proposition 1.10. *Suppose that $1 < \mu < \infty$. Then $\mathbb{P}(W = 0)$ is either equal to q or to 1.*

Proof. We have

$$X_{n+1} = \sum_{i=1}^{X_1} X_n^{(i)},$$

where $X_n^{(1)}, X_n^{(2)}, \ldots$ are i.i.d. copies of X_n, independent of X_1. So, similarly,

$$W_{n+1} = \mu^{-1} \sum_{i=1}^{X_1} W_n^{(i)}.$$

Letting $n \to \infty$ and using Proposition 1.9, we see that W has the same distribution as

$$\mu^{-1} \sum_{i=1}^{X_1} W^{(i)}. \tag{1.1}$$

In particular, $\mathbb{P}(W = 0) = \mathbb{E}\left[\mathbb{P}(W = 0)^{Z_1}\right] = G(\mathbb{P}(W = 0))$. So $\mathbb{P}(W = 0)$ is a root of $s = G(s)$. But, by Theorem 1.8 the possible roots are q and 1 and the result follows. $\square$

Exercise 1.11. *Consider again the Geometric offspring distribution of Exercise 1.6. Assume that $p < 1/2$, so that $\mu = \frac{1-p}{p} > 1$. Suppose that W has the distribution specified by*

$$\mathbb{P}(W = 0) = \frac{p}{1-p}, \quad \mathbb{P}(W > w) = \frac{1-2p}{1-p} e^{-\frac{(1-2p)w}{1-p}} \text{ for } w > 0.$$

Show that W solves the recursive distributional equation (1.1).
[Hint: First show that the number of terms in the sum (1.1) which are non-zero has a Geometric distribution, with parameter $p/(1-p)$. Then use moment generating functions.]

A necessary and sufficient condition for $\mathbb{E}[W] = 1$ is given in the following theorem.

Theorem 1.12 (Kesten and Stigum [6]). *Suppose that $1 < \mu < \infty$, and write $\log^+ x = \max\{\log x, 1\}$. Then the following are equivalent:*

- $\mathbb{E}[W] = 1$;
- $\mathbb{P}(W > 0 \mid \text{non-extinction}) = 1$;
- $\sum_{k=0}^{\infty} k p(k) \log^+ k < \infty$.

For much more on this topic, see Shi [9].

1.5 Galton-Watson trees

We have so far just considered the population sizes $X_0, X_1, X_2, \ldots$ in each of the successive generations. But it's also natural to want to think about a richer structure: the *family tree* (also called the *genealogical tree*). In order to do this, it will turn out to be useful to consider a canonical labelling for trees (called the *Ulam-Harris labelling*). Let $\mathbb{N} = \{1, 2, \ldots\}$ and let $\mathcal{U} = \{\emptyset\} \cup \bigcup_{n=1}^{\infty} \mathbb{N}^n$. In general, an element $u \in \mathcal{U}$ is a sequence of natural numbers representing a point in an infinitary tree. We will think of u as a string and write $u_1 u_2 \ldots u_n$. The concatenation of strings $u = u_1 u_2 \ldots u_n$ and $v = v_1 v_2 \ldots v_m$ is written $uv = u_1 u_2 \ldots u_n v_1 v_2 \ldots v_m$.

The label of a vertex $u \in \mathcal{U}$ indicates its genealogy:

- u has *parent* $p(u) = u_1 u_2 \dots u_{n-1}$.
- u has *children* $u1, u2, \dots$.
- Write $|u| = n$ for the *generation* of u.

Definition 1.13. *A* rooted ordered tree $\mathbf{t}$ *is a subset of* $\mathcal{U}$ *such that*

- $\emptyset \in \mathbf{t}$*;*
- *for all* $u \in \mathbf{t}$ *such that* $u \neq \emptyset$*,* $p(u) \in \mathbf{t}$*;*
- *for all* $u \in \mathbf{t}$*, there exists* $c(u) \in \{0, 1, 2, \dots\}$ *such that for* $j \in \mathbb{N}$*,* $uj \in \mathbf{t}$ *if and only if* $1 \leq j \leq c(u)$*.*

The tree is rooted at $\emptyset$*. The number of children of* u *in* $\mathbf{t}$ *is given by* $c(u)$*. We write* $\#(\mathbf{t})$ *for the size (number of vertices) of* $\mathbf{t}$ *and note that*

$$\#(\mathbf{t}) = 1 + \sum_{u \in \mathbf{t}} c(u).$$

We can think of the family tree of a branching process as a random rooted ordered tree which is obtained by sampling independent and identically distributed random variables $C(u), u \in \mathcal{U}$ each having the offspring distribution. The corresponding *Galton-Watson tree* $\mathbf{T}$ is then obtained by finding the largest subset of $\mathcal{U}$ such that the conditions of Definition 1.13 are fulfilled. Note that the tree is finite if and only if the branching process becomes extinct.

1.6 Relationship to random walks

Trees are slightly awkward objects to manipulate and so we often find it useful to encode them in terms of *functions*. This will be useful in the next lecture. There are several different ways to do this; we will concentrate on one which works well with our labelling. See Sections 1.1 and 1.2 of Le Gall [7] for more details.

For a rooted ordered tree $\mathbf{t}$, let $v_1 = \emptyset$ and, for $i \geq 2$, let v_i be the ith vertex of $\mathbf{t}$ when the vertices are listed in lexicographic order of label.

Definition 1.14. *The* Lucasiewicz path *associated with a tree* $\mathbf{t}$ *of size* $\#(\mathbf{t}) = n$ *is the function* $\ell : \{0, 1, \dots, n\} \to \{-1, 0, 1, \dots\}$ *defined by* $\ell(0) = 0$ *and, for* $1 \leq i \leq n$*,*

$$\ell(i) = \ell(i-1) + c(v_i) - 1;$$

i.e. $\ell(i) = \sum_{j=1}^{i} (c(v_j) - 1)$*. Note that by (1.13), we must have* $\ell(n) = -1$*. Moreover,* $\ell(i) \geq 0$ *for* $1 \leq i \leq n-1$*.*

Exercise 1.15. *Show that a given Lucasiewicz path* $(\ell(i), 0 \leq i \leq n)$ *encodes a unique tree.*

In other words, there is a bijection between rooted ordered trees and Lucasiewicz paths.

In the case of Galton-Watson trees, this correspondence gives us something particularly nice. Write $(L(i), 0 \leq i \leq \#(\mathbf{T}))$ for the Lucasiewicz path associated with a Galton-Watson tree $\mathbf{T}$. Let $(S(i), i \geq 0)$ be a random walk with step-sizes having distribution $p(k+1), k \geq -1$ and let $\tau = \inf\{i : S(i) = -1\}$.

Proposition 1.16. *The processes* $(L(i), 0 \leq i \leq \#(\mathbf{T}))$ *and* $(S(i), 0 \leq i \leq \tau)$ *have the same distribution.*

Sketch proof. We have $L(i) = \sum_{j=1}^{i}(C(v_j) - 1)$, so that $L(i) \geq 0$ for $0 \leq i \leq \#(\mathbf{T})$ and $L(\#(\mathbf{T})) = -1$. Since the random variables $C(u) - 1$, $u \in \mathcal{U}$ are independent, take values in $\{-1, 0, 1, \ldots\}$ and have probability mass function $p(k+1), k \geq -1$, the result seems intuitively clear. The problem is that the labels $v_1, v_2, \ldots$ are random and *depend* on $(C(u), u \in \mathcal{U})$, so it is not, in fact, obvious that $C(v_1) - 1, C(v_2) - 1, \ldots$ are independent and identically distributed. The point is that, at step $n+1$ of the lexicographic exploration of the tree, the labels $v_1, \ldots, v_n$ and the numbers $C(v_1), \ldots, C(v_n)$ enable us to determine (a) if we have reached the end of the tree and (b) if not, which is the label v_{n+1}. But, in case (b), the information we already possess can't influence the value of $C(v_{n+1})$. See Proposition 1.5 of Le Gall [7] for the details. □

Note that we can (and will) think of the random walk as continuing on beyond the first time it hits -1.

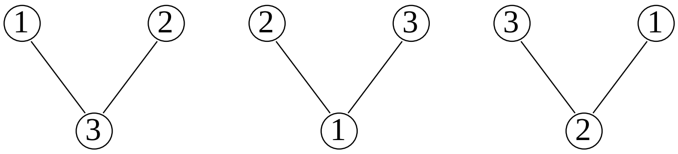
Remark 1.17. *Proposition 1.16 entails that we can think of the extinction probability of a branching process as the probability that the random walk ever hits* -1*. We observe that the step sizes of the random walk have mean* $\mu - 1$*. Assuming additionally that* $p(0) > 0$ *and* $p(0) + p(1) < 1$*, we see that it is, indeed, the case that the walk hits* -1 *with probability 1 if* $\mu \leq 1$ *(corresponding to extinction) and has positive probability of never hitting* -1 *if* $\mu > 1$*.*

Chapter 2

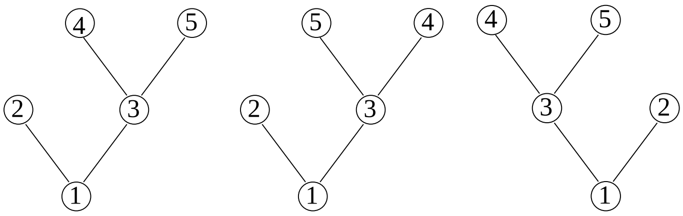
Uniform random trees

2.1 Introduction

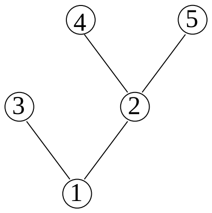
In this lecture, we're going to take a more combinatorial perspective on random trees, but it will eventually connect back to what we discussed in the first lecture. Here we will concentrate on *labelled trees*. We let $\mathbb{T}_n$ be the set of (unordered) labelled trees on n vertices (labelled by $[n] := \{1, 2, \ldots, n\}$. For example, $\mathbb{T}_3$ consists of the trees



Unordered means that these trees are all the same:



but this one is different:



Suppose we take all of the trees on n labelled vertices and pick one uniformly at random; call it T_n. What does T_n look like? Phrased another way: what does a *typical* tree on n vertices look like?

2.2 Cayley's formula

A good starting place is to ask how many elements $\mathbb{T}_n$ has.

Theorem 2.1 (Cayley's formula)**.** *For $n \geq 1$,*

$$|\mathbb{T}_n| = n^{n-2}.$$

This theorem was first proved by Borchardt (1860) and later extended by Cayley (1889).

Proof. We give a beautiful proof due to Pitman [8]. Call the quantity we're after $a(n)$. It's useful for the purposes of this proof to think of rooted trees with directed edges, where the edges are all directed away from the root. (We can simply forget the root and the directions of the arrow at the end.) Start from the empty graph on our n vertices and count the number of different sequences of directed edges that we can add to form a rooted tree.

First way to count.
Start with one of the $a(n)$ unrooted trees on n vertices. There are n ways to pick a root. Direct all edges away from the root. There are $(n-1)!$ different orders in which we might add the $n-1$ directed edges. So, we have $a(n) \times n \times (n-1)! = a(n)n!$ different sequences of directed edges.

Second way to count.
Start by thinking of each of the n vertices as a root. At each step, we have a forest of rooted trees, with edges directed away from the roots. Add an edge from an arbitrary vertex v to the root of a different tree, and root the resulting tree at the root of v's old tree. If we have added k edges already, then we have a forest containing $n-k$ trees (we reduce the number of trees by 1 at each step). So there are $n(n-k-1)$ choices for the edge we add: the starting vertex can be any one of the n vertices and there are $n-k-1$ roots of other trees to which it can connect.

So the total number of choices is

$$\prod_{k=0}^{n-2} n(n-k-1) = n^{n-1}(n-1)! = n^{n-2}n!.$$

So $a(n)n! = n^{n-2}n!$ and cancelling gives $a(n) = n^{n-2}$. □

There are many questions we might ask about our uniform random tree T_n. For example,

- How many *leaves* (i.e. vertices with only one neighbour) are there?
- How many vertices have two neighbours? Or three? Or more?

- What is the *diameter* of the tree (i.e. the length of the longest path between two points in the tree)?
- Pick a vertex uniformly at random to be the root. How many vertices are there at distance d from the root?

Because the tree is random, all of these quantities are random variables.

2.3 Leaves

Theorem 2.2. *Let N_n denote the number of leaves in T_n. Then*

$$\frac{N_n}{n} \xrightarrow{p} e^{-1}$$

as $n \to \infty$.

Proof. Note that

$$N_n = \sum_{i=1}^{n} I_i$$

where

$$I_i = \begin{cases} 1 & \text{if the vertex labelled } i \text{ is a leaf} \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\mathbb{E}\left[I_i\right] = \mathbb{P}\left(i \text{ is a leaf}\right).$$

In order to calculate this probability, we need to determine how many trees on $[n]$ have i as a leaf. Each such tree can be thought of as a tree on $[n] \setminus \{i\}$ plus an edge from one of the other vertices to i. Since there are $(n-1)^{n-3}$ trees on $n-1$ labelled vertices, it follows that there are $(n-1)^{n-2}$ trees which have i as a leaf. So

$$\mathbb{P}\left(i \text{ is a leaf}\right) = \frac{(n-1)^{n-2}}{n^{n-2}} = \left(1 - \frac{1}{n}\right)^{n-2} \to e^{-1},$$

as $n \to \infty$.

If the indicator random variables $I_1, I_2, \ldots, I_n$ were independent, we would now be able to apply the weak law of large numbers. But this is not quite the case. Let $i \neq j$. Then

$$\mathbb{P}\left(i \text{ is a leaf and } j \text{ is a leaf}\right) = \frac{(n-2)^2(n-2)^{n-4}}{n^{n-2}} = \left(1 - \frac{2}{n}\right)^{n-2} \to e^{-2},$$

so that I_i and I_j are *asymptotically* independent. In particular,

$$\operatorname{cov}\left(I_i, I_j\right) = \left(1 - \frac{2}{n}\right)^{n-2} - \left(1 - \frac{1}{n}\right)^{2(n-2)} \to 0,$$

as $n \to \infty$. We also have

$$\operatorname{var}\left(I_i\right) = \left(1 - \left(1 - \frac{1}{n}\right)^{n-2}\right)\left(1 - \frac{1}{n}\right)^{n-2} \to (1 - e^{-1})e^{-1}.$$

Hence, using symmetry,

$$\mathrm{var}\left(\frac{N_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}\,(I_i) + \frac{2}{n^2}\sum_{i<j}\mathrm{cov}\,(I_i, I_j) = \frac{1}{n}\mathrm{var}\,(I_1) + \frac{n-1}{n}\mathrm{cov}\,(I_1, I_2) \to 0,$$

as $n \to \infty$. Finally, for any $\epsilon > 0$, we can apply Chebyshev's inequality to obtain

$$\mathbb{P}\left(\left|\frac{N_n}{n} - e^{-1}\right| > \epsilon\right) \leq \frac{\mathrm{var}\,(N_n/n)}{\epsilon^2} \to 0,$$

as $n \to \infty$, as desired. □

2.4 The uniform random tree as a Galton-Watson tree

We now make the link between our study of the uniform random tree and the last lecture on branching processes.

Proposition 2.3. *The uniform random tree T_n has the same distribution as a tree generated as follows:*

- *Take a Galton-Watson tree with Poisson(1) offspring distribution;*
- *Condition it to have total progeny precisely n;*
- *Assign the vertices random labels chosen from $[n]$ and forget the original ordering.*

Proof. Recall the standard labelling of the vertices of a Galton-Watson tree $\mathbf{T}$, and let $\mathbf{t}$ be a particular tree with $\#(\mathbf{t}) = n$ and numbers of offspring $c(v_1) = c_1, \ldots, c(v_n) = c_n$. Then

$$\mathbb{P}\,(\mathbf{T} = \mathbf{t}) = \prod_{i=1}^{n}\frac{e^{-1}}{c_i!} = e^{-n}\prod_{i=1}^{n}\frac{1}{c_i!}.$$

Now observe that $\mathbb{P}\,(\#(\mathbf{T}) = n)$ is a function only of n. Hence,

$$\mathbb{P}\,(\mathbf{T} = \mathbf{t}|\#(\mathbf{T}) = n) = f(n)\prod_{i=1}^{n}\frac{1}{c_i!}$$

for some function f. Now consider labelling the vertices of $\mathbf{T}$ with $[n]$. There are $n!$ different ways to do this, of which $\prod_{i=1}^{n} c_i!$ give rise to the same unordered labelled tree, once we forget the ordering. Hence, the probability of obtaining a particular labelled unordered tree t is $f(n)/n!$. Since this depends only on n, and not on any other feature of the tree, it must be the case that the tree is uniformly distributed on $\mathbb{T}_n$. □

Remark 2.4. *Other classes of "combinatorial" trees may also be obtained as Galton-Watson trees conditioned on their total progeny. For example, taking $p(k) = \left(\frac{1}{2}\right)^{k+1}, k \geq 0$, gives a uniform plane tree; taking $p(0) = 1/2$ and $p(2) = 1/2$ gives a uniform complete binary tree (as long as the total progeny is odd).*

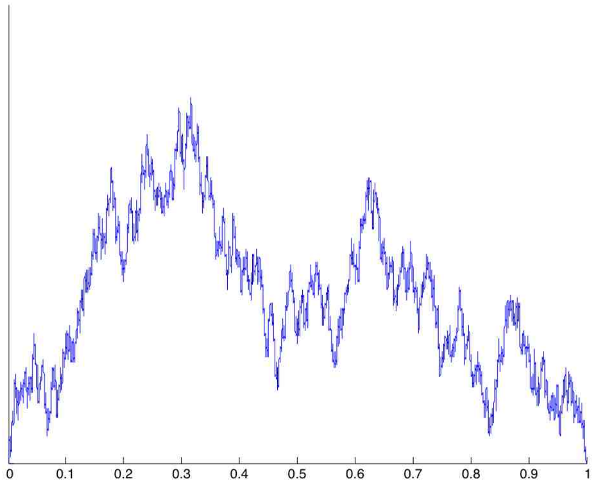


Figure 2.1: A standard Brownian excursion.

2.5 Taking limits

We've already seen that if we look at the proportion of the vertices in T_n which are leaves then we obtain a meaningful limit as $n \to \infty$. What can we say more generally about T_n as n gets large? We will make crucial use of the link to Galton-Watson trees outlined in the last section. We won't prove the results in this section; see Aldous [2] and Le Gall [7] for more details.

Recall that if $\mathbf{T}$ is a Galton-Watson tree with offspring distribution $p(k), k \geq 0$, then its Lucasiewicz path $(L(i), 0 \leq i \leq \#(\mathbf{T}))$ has the same distribution as a random walk $(S(k), k \geq 0)$ with step-size distribution $p(k+1), k \geq -1$, stopped at τ, the first time S hits -1.

We will concentrate on the critical case, where $\mu = 1$. The fundamental idea that we will make use of in this section comes from the following theorem.

Theorem 2.5 (Donsker's theorem). *Suppose that $(S(k), k \geq 0)$ is a random walk with step-sizes of mean 0 and finite variance σ^2. Then, as $n \to \infty$,*

$$\left(\frac{1}{\sigma\sqrt{n}} S(\lfloor nt \rfloor), t \geq 0\right) \xrightarrow{d} (B(t), t \geq 0),$$

where $(B(t), t \geq 0)$ is a standard Brownian motion.

In other words, in the critical case, there is a good way to rescale the random walk in such a way that we obtain a non-trivial limiting process. If we have a Poisson(1) offspring distribution, we get that the step-sizes precisely have mean 0 and variance 1.

Rather than a critical Galton-Watson of random size, we would like to consider a critical Galton-Watson tree conditioned to have size n. The corresponding Lucasiewicz path is a random walk conditioned to hit -1 for the first time at time n. Let $(S^{(n)}(k), 0 \leq k \leq n)$ have the same distribution as $(S(k), 0 \leq k \leq n)$ conditioned on the event $\{\tau = n\}$.

Theorem 2.6 (Kaigh [5]). *As* $n \to \infty$,

$$\left(\frac{1}{\sigma\sqrt{n}}S^{(n)}(\lfloor nt \rfloor), 0 \le t \le 1\right) \xrightarrow{d} (\mathbb{e}(t), 0 \le t \le 1),$$

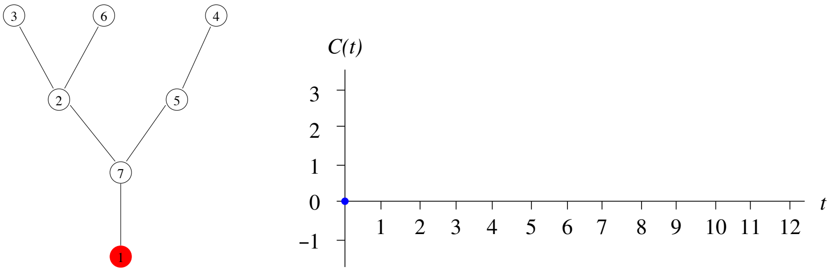
where $(\mathbb{e}(t), 0 \le t \le 1)$ *is a* standard Brownian excursion.

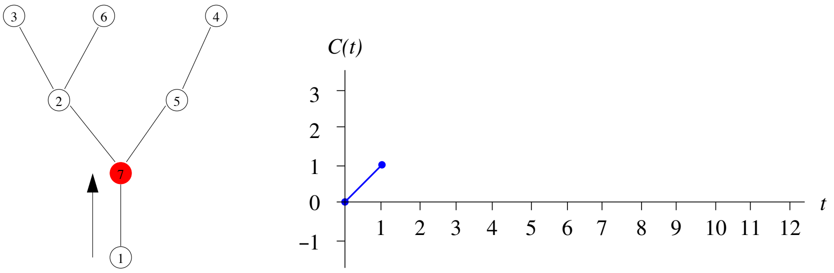
We don't have time to go into a formal definition of $\mathbb{e}$ but, informally, it is a Brownian motion conditioned to stay positive in the interval $(0, 1)$ and to come back to 0 at time 1 (see Figure 2.1).

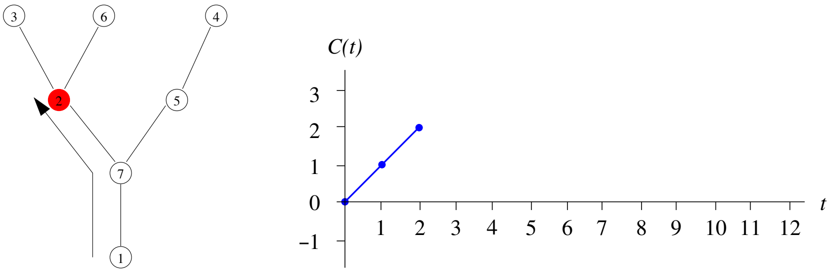
This theorem suggests that there should be a limit for the tree itself, and that limit should somehow be encoded by $\mathbb{e}$. This is, indeed, the case, as was proved by Aldous in a series of papers in the early 1990's [1, 2, 3]. The limit is known as the *Brownian continuum random tree.*

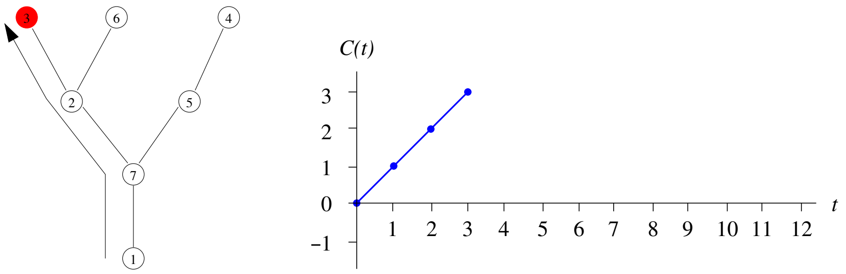
2.6 The Brownian continuum random tree

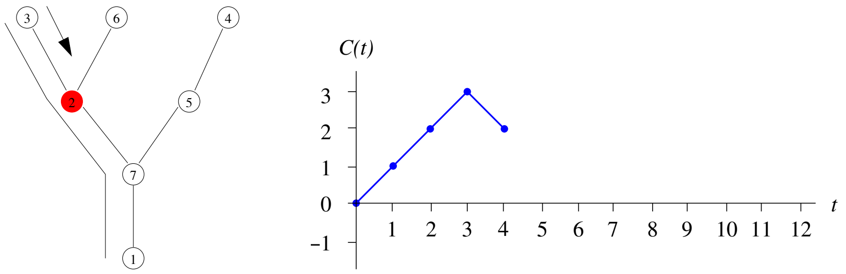
In order to see how the limit tree arises, we will go through the construction in the special case where the offspring distribution is given by $p(k) = \left(\frac{1}{2}\right)^{k+1}, k \ge 0$ i.e. when we have a uniform plane tree. We will use an encoding for our tree which is somewhat easier to visualise than the Lucasiewicz path. The *contour function* is obtained by simply tracing the "contour" of the tree from left to right at speed 1, so that we pass along each edge twice. Record the distance from the root at each time to get $(C(k), 0 \le k \le 2n - 2)$.

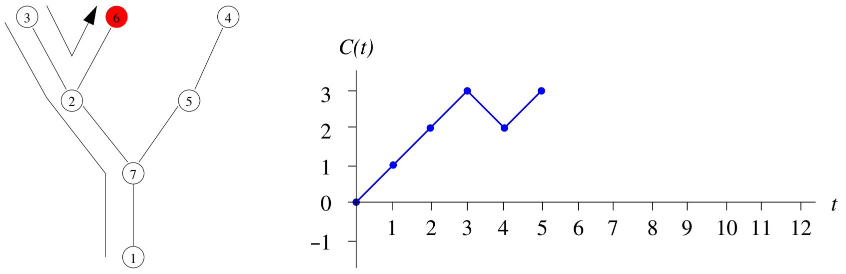


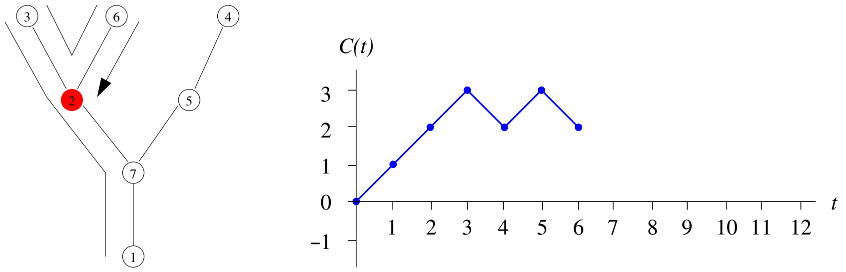


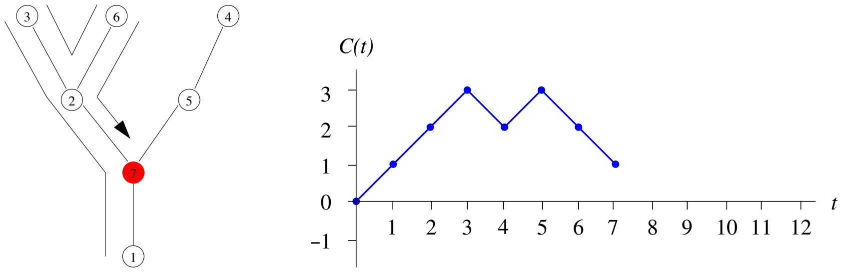


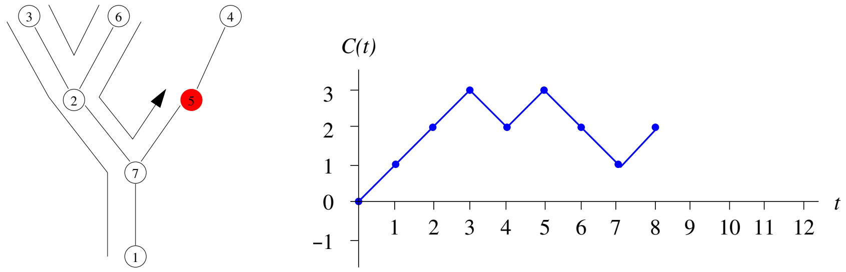

3
6
4
2
5
7
1
C(t)
3
2
1
0
−1
t
1 2 3 4 5 6 7 8 9 10 11 12

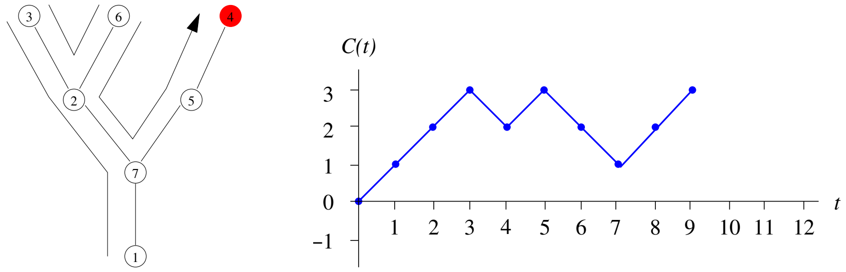

3
6
4
2
5
7
1
C(t)
3
2
1
0
−1
t
1 2 3 4 5 6 7 8 9 10 11 12

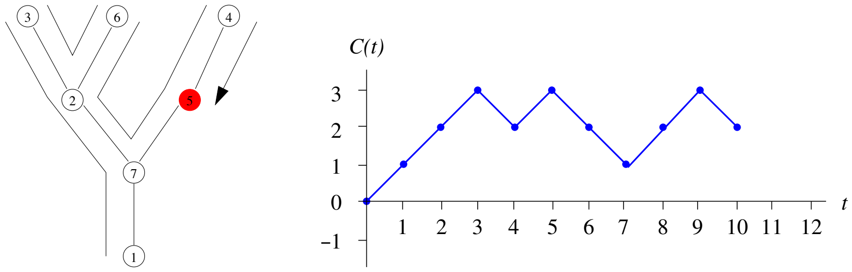

3
6
4
2
5
7
1
C(t)
3
2
1
0
−1
t
1 2 3 4 5 6 7 8 9 10 11 12

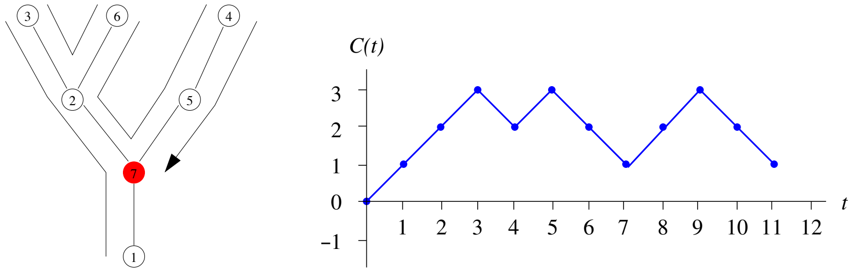

3
6
4
2
5
7
1
C(t)
3
2
1
0
−1
t
1 2 3 4 5 6 7 8 9 10 11 12

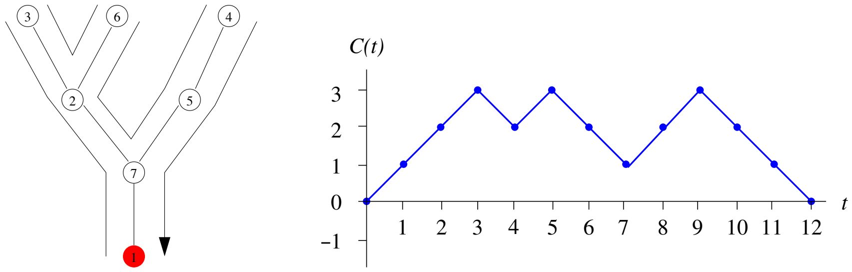

3
6
4
2
5
7
1
C(t)
3
2
1
0
−1
t
1 2 3 4 5 6 7 8 9 10 11 12











The contour function is a sort of “expanded” version of the tree. For convenience, we will define a slightly shifted version: let $\tilde{C}(0) = 0$, $\tilde{C}(2n) = 0$ and, for $1 \leq k \leq 2n-1$, $\tilde{C}(k) = 1 + C(k-1)$.

Exercise 2.7. *Show that* $(\tilde{C}(k), 0 \leq k \leq 2n)$ *has the same distribution as a simple symmetric random walk conditioned to return to the origin for the first time at time* $2n$.

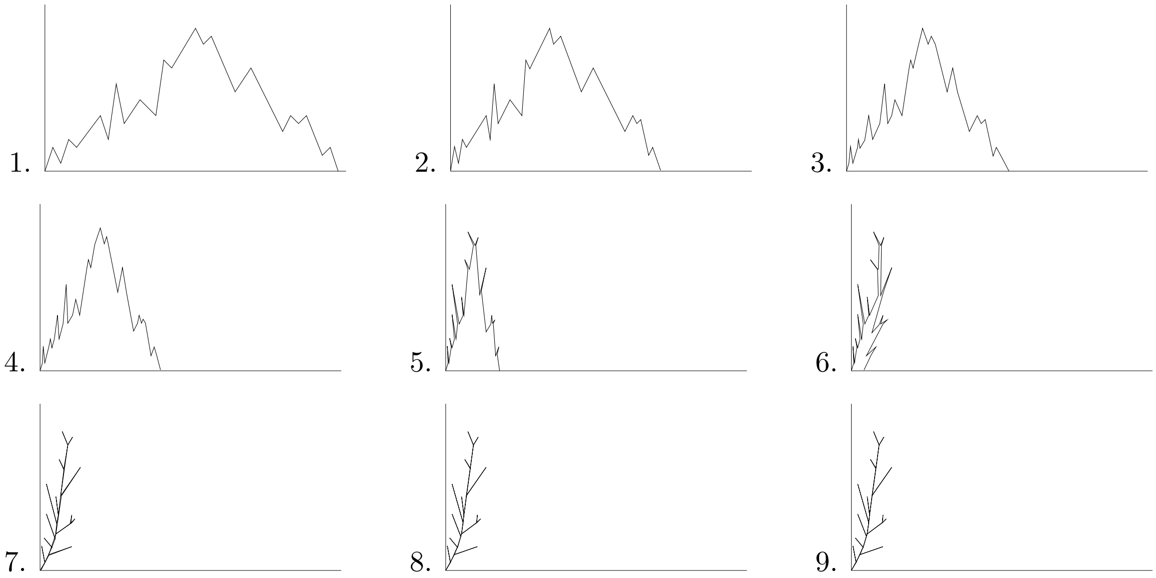
It's easy to linearly interpolate to give a continuous function $\tilde{C} : [0, 2n] \to [0, \infty)$, and similarly for C. Since the variance of the offspring distribution is 2, it follows from Theorem 2.6 that

$$\frac{1}{\sqrt{2}\sqrt{2n}}(\tilde{C}(2nt), 0 \le t \le 1) \xrightarrow{d} (\mathbb{e}(t), 0 \le t \le 1)$$

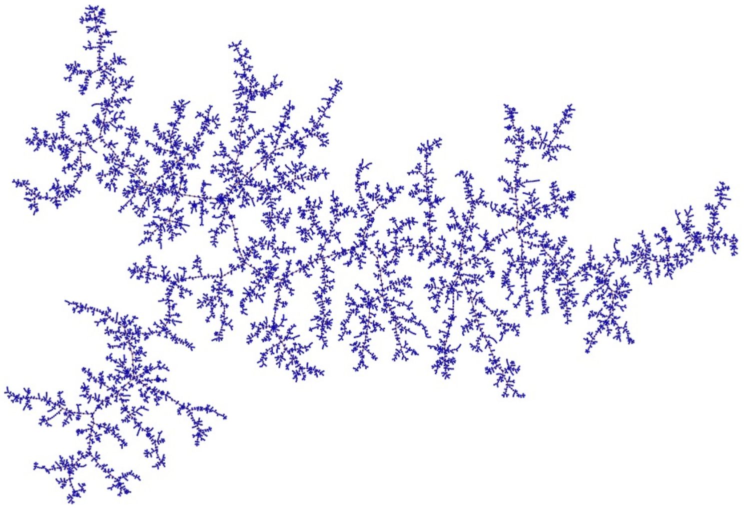
as $n \to \infty$ and, hence,

$$\frac{1}{\sqrt{n}}(C(2nt), 0 \le t \le 1) \xrightarrow{d} (2\mathbb{e}(t), 0 \le t \le 1).$$

It's also relatively straightforward to see how to get back a tree from a contour function. Intuitively speaking, we put glue on the underside of the function and then push the two sides together until they meet:



The Brownian continuum random tree we obtain by doing this gluing procedure to the function $(2\mathbb{e}(t), 0 \le t \le 1)$:



For a general critical Galton-Watson tree with finite offspring variance σ^2 conditioned to have total progeny n, it turns out that the convergence of the (suitably rescaled) Lucasiewicz path to a standard Brownian excursion actually implies the convergence of the contour function of the tree (although this is not trivial to see). Hence, *any* such conditioned tree converges in distribution as $n \to \infty$ to the Brownian continuum random tree. See Aldous [2] and Le Gall [7] for more details, in particular the specification of the topology in which convergence occurs.

Acknowledgements

The pictures of a standard Brownian excursion (Figure 2.1) and the Brownian continuum random tree were made by Igor Kortchemski. I am very grateful to him for letting me use them.

Bibliography

[1] D. Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[2] D. Aldous. The continuum random tree. II. An overview. In *Stochastic analysis (Durham, 1990)*, volume 167 of *London Math. Soc. Lecture Note Ser.*, pages 23–70. Cambridge University Press, Cambridge, 1991.

[3] D. Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[4] K. Athreya and P. Ney. *Branching processes.* Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[5] W. D. Kaigh. An invariance principle for random walk conditioned by a late return to zero. *Ann. Probability*, 4(1):115–121, 1976.

[6] H. Kesten and B. Stigum. A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Statist.*, 37:1211–1223, 1966.

[7] J.-F. Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.

[8] J. Pitman. Coalescent random forests. *J. Combin. Theory Ser. A*, 85(2):165–193, 1999.

[9] Zhan Shi. Random walks and trees. In *X Symposium on Probability and Stochastic Processes and the First Joint Meeting France-Mexico of Probability*, volume 31 of *ESAIM Proc.*, pages 1–39. EDP Sci., Les Ulis, 2011.