ALEA in Europe School, LMU München, February 2016

Scaling limits of random discrete trees

Christina Goldschmidt

UNIVERSITY OF OXFORD

Lecture 1

1. WHAT IS A SCALING LIMIT?

Scaling limits

Suppose we have a sequence of random objects $R_1, R_2, \ldots$ and we can find a sequence $\alpha_1, \alpha_2, \ldots$ such that

$$\alpha_n R_n \stackrel{d}{\to} R$$

as $n \to \infty$ for some limiting random variable R. Then we call R the scaling limit of the sequence $(R_n, n \geq 1)$.

Scaling limits

Suppose we have a sequence of random objects $R_1, R_2, \ldots$ and we can find a sequence $\alpha_1, \alpha_2, \ldots$ such that

$$\alpha_n R_n \xrightarrow{d} R$$

as $n \to \infty$ for some limiting random variable R. Then we call R the scaling limit of the sequence $(R_n, n \geq 1)$.

Central limit theorem. Suppose that $Z_1, Z_2, \ldots$ are independent and identically distributed random variables with mean 0 and variance $0 < \sigma^2 < \infty$. Then as $n \to \infty$,

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} Z_i \xrightarrow{d} X,$$

where $X \sim \mathrm{N}(0, 1)$.

Universality

This scaling limit is universal, in that it doesn't depend on the precise details of the distribution of $Z_1, Z_2, \ldots$ (as long as the distribution has finite variance).

Universality

This scaling limit is universal, in that it doesn't depend on the precise details of the distribution of $Z_1, Z_2, \dots$ (as long as the distribution has finite variance).

(Aside: what happens if $\mathrm{var}(Z_1) = \infty$? Or even if $\mathbb{E}[|Z_1|] = \infty$?)

Another (related) scaling limit

Suppose that $Z_1, Z_2, \ldots$ are independent and identically distributed random variables with mean 0 and variance σ^2. Let $X(0) = 0$ and $X(k) = \sum_{i=1}^{k} Z_i$. Then $(X(k), k \geq 0)$ is a random walk.

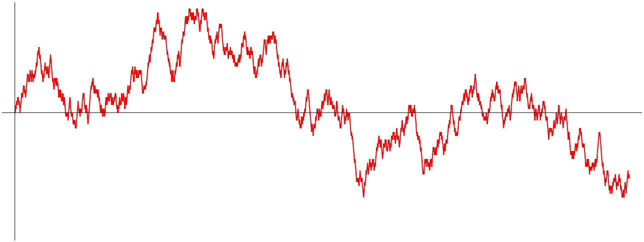
Another (related) scaling limit

Suppose that $Z_1, Z_2, \ldots$ are independent and identically distributed random variables with mean 0 and variance σ^2. Let $X(0) = 0$ and $X(k) = \sum_{i=1}^{k} Z_i$. Then $(X(k), k \geq 0)$ is a random walk.

Donsker's theorem. As $n \to \infty$,

$$\frac{1}{\sigma\sqrt{n}}(X(\lfloor nt \rfloor), t \geq 0) \xrightarrow{d} (W(t), t \geq 0),$$

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



[Picture by Louigi Addario-Berry]

Convergence in distribution

The convergence in distribution in the CLT means that

$$\mathbb{P}\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} Z_i \leq x\right) \to \Phi(x)$$

as $n \to \infty$ for all $x \in \mathbb{R}$.

Convergence in distribution

The convergence in distribution in the CLT means that

$$\mathbb{P}\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} Z_i \leq x\right) \to \Phi(x)$$

as $n \to \infty$ for all $x \in \mathbb{R}$.

However, we are going to want to deal with random objects which are not real-valued; for example, the random walk is a function-valued object. In this case, an equivalent definition generalises better:

$$\mathbb{E}\left[f\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} Z_i\right)\right] \to \mathbb{E}\left[f(X)\right]$$

for all functions $f : \mathbb{R} \to \mathbb{R}$ which are bounded and continuous.

Convergence in distribution

So Donsker's theorem means that for all bounded continuous real-valued functions f,

$$\mathbb{E}\left[f\left(\frac{1}{\sigma\sqrt{n}}(X(\lfloor nt \rfloor), t \geq 0)\right)\right] \to \mathbb{E}\left[f(W(t), t \geq 0)\right].$$

Convergence in distribution

So Donsker's theorem means that for all bounded continuous real-valued functions f,

$$\mathbb{E}\left[f\left(\frac{1}{\sigma\sqrt{n}}(X(\lfloor nt\rfloor), t \geq 0)\right)\right] \to \mathbb{E}\left[f(W(t), t \geq 0)\right].$$

(Here, f is continuous for the metric on càdlàg functions $D(\mathbb{R}_+, \mathbb{R})$ given by

$$d(x,y) = \sum_{k=1}^{\infty} 2^{-k} \left(\sup_{t \in [0,k]} |x(t) - y(t)| \wedge 1 \right),$$

yielding uniform convergence on compact time-intervals).

2. THE UNIFORM RANDOM TREE

Key references:

David Aldous, **The continuum random tree I**,
D. Aldous, *Annals of Probability* **19** (1991) pp.1-28.

David Aldous, **The continuum random tree II. An overview**,
in *Stochastic analysis (Durham 1990)*, vol. 167 of London
Mathematical Society Lecture Note Series (1991) pp.23-70.

Uniform random trees

We start with perhaps the simplest model of a random tree.

Uniform random trees

We start with perhaps the simplest model of a random tree.

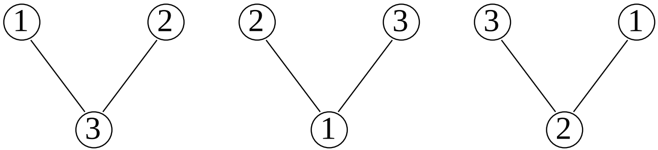
Let $\mathbb{T}_n$ be the set of unordered trees on n vertices labelled by $[n] := \{1, 2, \ldots, n\}$.

Uniform random trees

We start with perhaps the simplest model of a random tree.

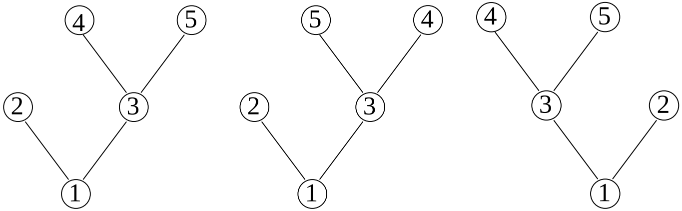
Let $\mathbb{T}_n$ be the set of unordered trees on n vertices labelled by $[n] := \{1, 2, \ldots, n\}$.

For example, $\mathbb{T}_3$ consists of the trees

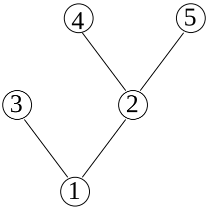


Uniform random trees

Unordered means that these trees are all the same:



but this one is different:



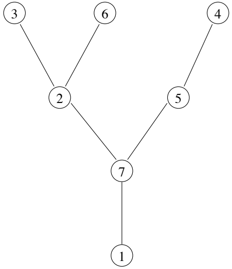
Uniform random trees

Cayley's formula: $|\mathbb{T}_n| = n^{n-2}$.

Uniform random trees

Cayley's formula: $|\mathbb{T}_n| = n^{n-2}$.

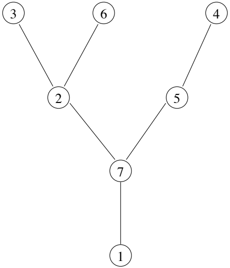
Write T_n for a tree chosen uniformly from $\mathbb{T}_n$.



Uniform random trees

Cayley's formula: $|\mathbb{T}_n| = n^{n-2}$.

Write T_n for a tree chosen uniformly from $\mathbb{T}_n$.

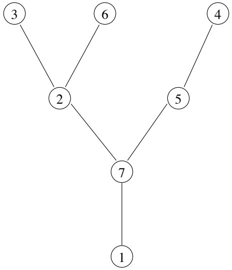


What happens as n grows?

Uniform random trees

Cayley's formula: $|\mathbb{T}_n| = n^{n-2}$.

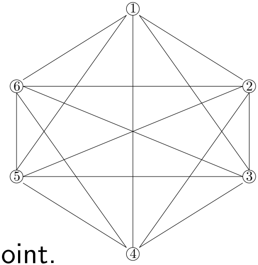
Write T_n for a tree chosen uniformly from $\mathbb{T}_n$.



What happens as n grows?

It's useful to have an algorithm for building T_n.

The Aldous-Broder algorithm

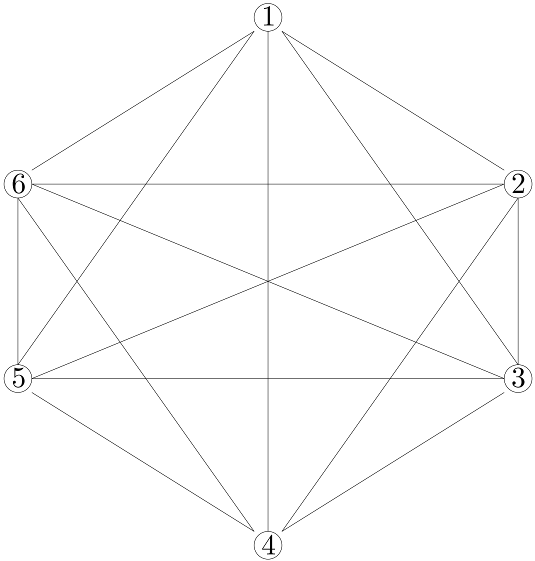


Take the complete graph on n vertices.

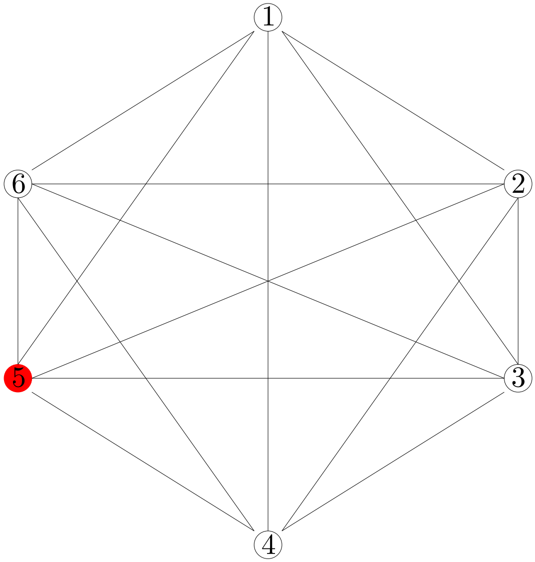
- Pick a uniform vertex to be the starting point.
- Run a simple random walk $(S_k)_{k \geq 0}$ on the graph (i.e. at each step, move to a neighbour chosen uniformly at random).
- Anytime the walk visits a new vertex, keep the edge along which it was reached.
- Stop when all vertices have been visited.

The resulting tree is uniformly distributed on $\mathbb{T}_n$.

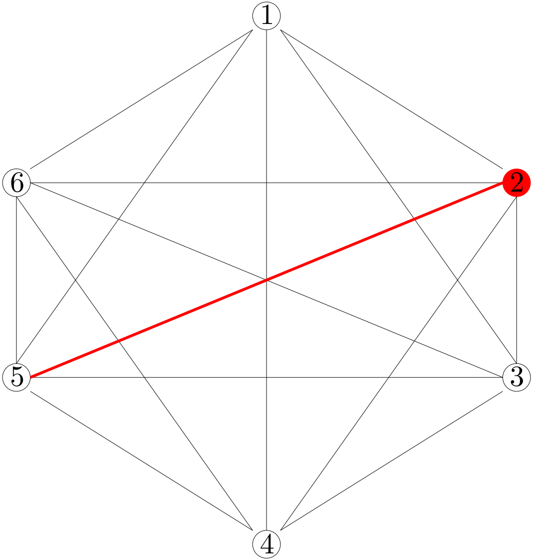
The Aldous-Broder algorithm



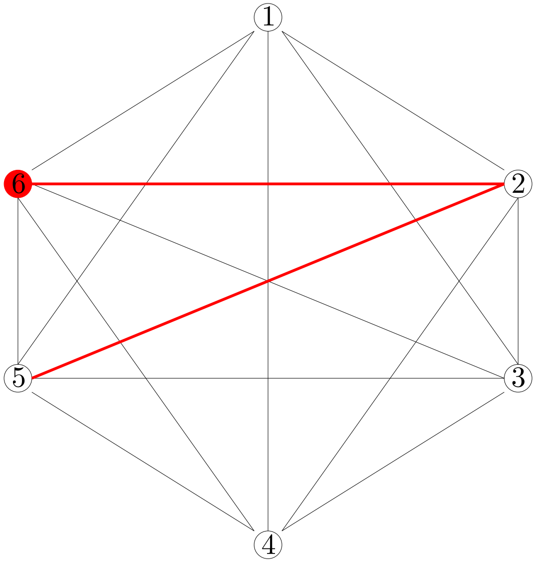
The Aldous-Broder algorithm



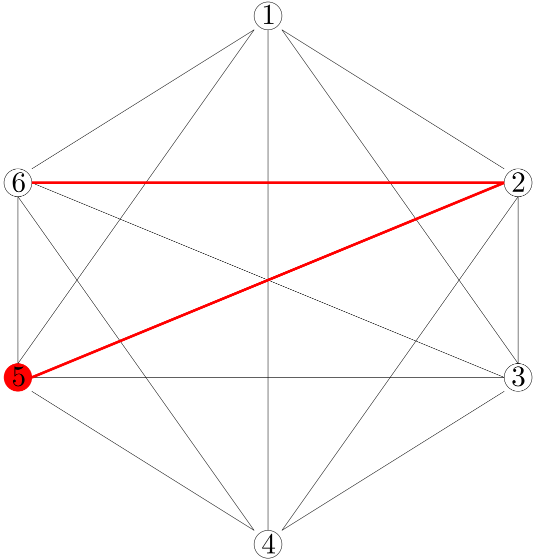
The Aldous-Broder algorithm



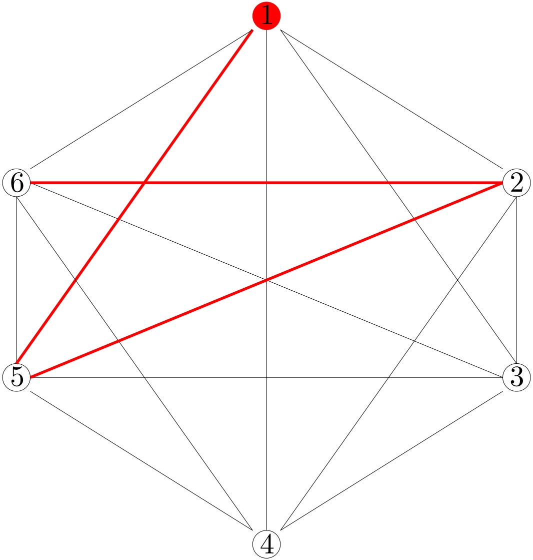
The Aldous-Broder algorithm



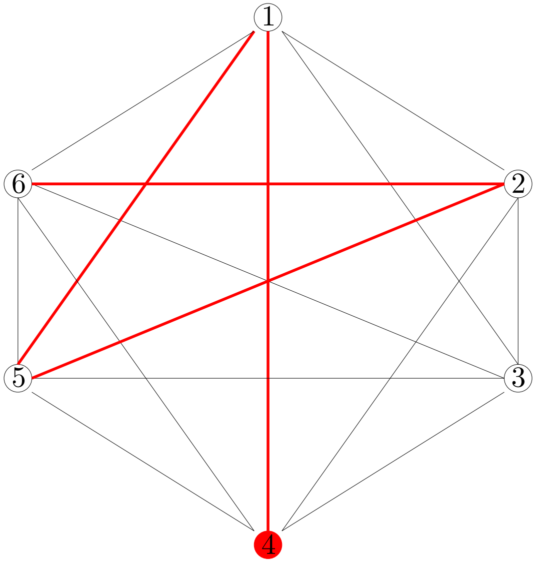
The Aldous-Broder algorithm



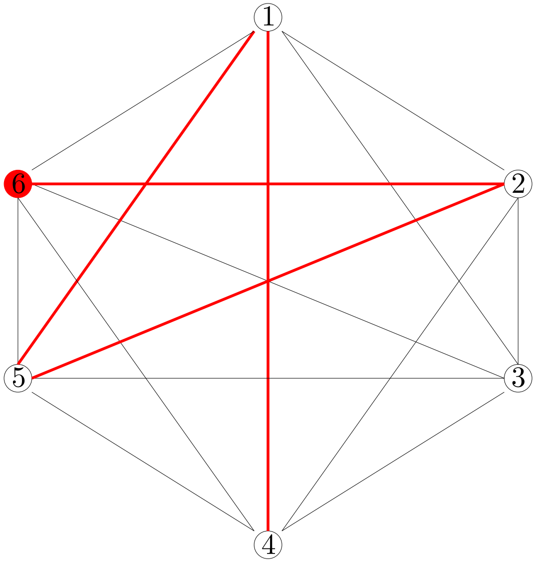
The Aldous-Broder algorithm



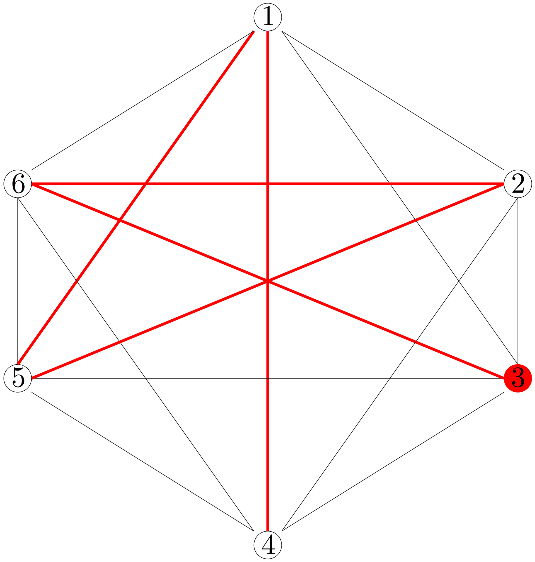
The Aldous-Broder algorithm



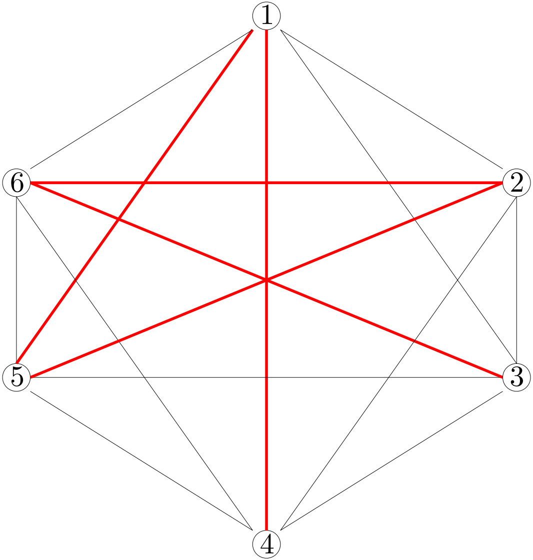
The Aldous-Broder algorithm



The Aldous-Broder algorithm



The Aldous-Broder algorithm



The Aldous-Broder algorithm: proof

The random walk $(S_k)_{k \geq 0}$ has a uniform stationary distribution, and is reversible, so that it makes sense to talk about a stationary random walk $(S_k)_{k \in \mathbb{Z}}$.

The Aldous-Broder algorithm: proof

The random walk $(S_k)_{k \geq 0}$ has a uniform stationary distribution, and is reversible, so that it makes sense to talk about a stationary random walk $(S_k)_{k \in \mathbb{Z}}$.

The dynamics of the random walk give rise to Markovian dynamics on $\mathbb{T}_n^\bullet$, the set of trees labelled by $[n]$ with a distinguished root.

The Aldous-Broder algorithm: proof

The random walk $(S_k)_{k\geq 0}$ has a uniform stationary distribution, and is reversible, so that it makes sense to talk about a stationary random walk $(S_k)_{k\in\mathbb{Z}}$.

The dynamics of the random walk give rise to Markovian dynamics on $\mathbb{T}_n^\bullet$, the set of trees labelled by $[n]$ with a distinguished root.

Why? Let τ_k be the tree constructed from the random walk started at time k, rooted at S_k.

τ_k depends on $S_k, S_{k+1}, \ldots$ through first hitting times of vertices. These can only occur later if we start from a later time. So, given τ_k, τ_{k+1} is independent of $\tau_{k-1}, \tau_{k-2}, \ldots$.

The Aldous-Broder algorithm: proof

Since the random walk is stationary, the tree must be also.

The Aldous-Broder algorithm: proof

Since the random walk is stationary, the tree must be also.

It remains to show that the stationary distribution π for $(\tau_k)_{k \in \mathbb{Z}}$ is uniform on $\mathbb{T}_n^{\bullet}$. It turns out to be easier to work with the time-reversed chain.

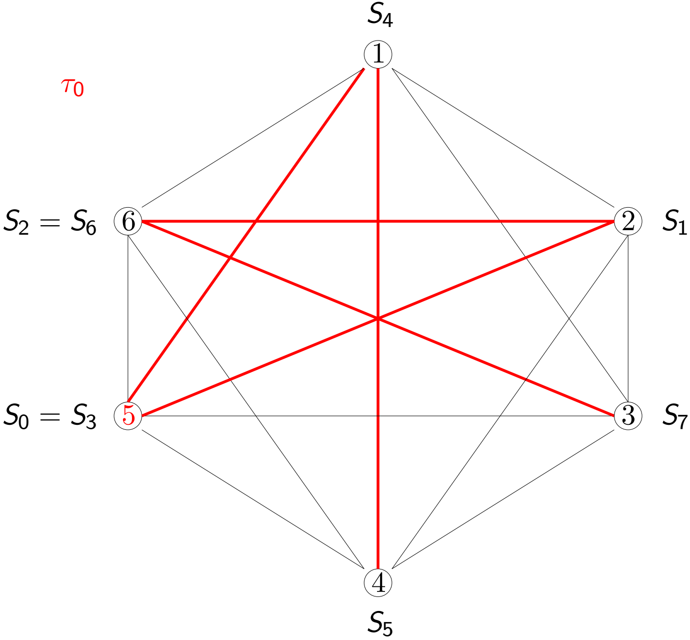
The Aldous-Broder algorithm: proof

Since the random walk is stationary, the tree must be also.

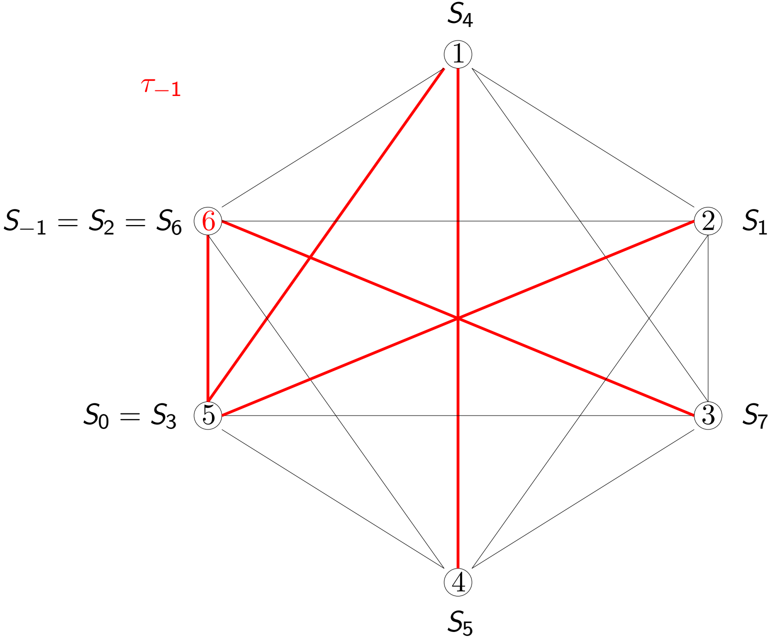
It remains to show that the stationary distribution π for $(\tau_k)_{k\in\mathbb{Z}}$ is uniform on $\mathbb{T}_n^\bullet$. It turns out to be easier to work with the time-reversed chain.

Consider the transition probabilities $q(\tau, \tau')$ for the time-reversed chain (which must have the same stationary distribution).

The Aldous-Broder algorithm: proof



The Aldous-Broder algorithm: proof



The Aldous-Broder algorithm: proof

Taking one step backwards in time (say from time 0 to time -1) inserts an edge from S_0 to S_{-1} in τ_0. This creates a cycle, from which we must delete the unique other edge in that cycle which connects to S_{-1} in order to obtain τ_{-1}.

There are $n-1$ different places that S_0 might move to and so $n-1$ possible rooted trees we can reach going backwards in time, each equally likely.

So for fixed τ, $q(\tau, \tau') = 0$ or $1/(n-1)$.

The Aldous-Broder algorithm: proof

Given τ_{-1}, how many possibilities are there for τ_0? S_0 must be one of the neighbours of S_{-1}. The possible values for τ_0 are generated by adding one the $n-1$ possible edges from S_{-1} to a different vertex. This creates a cycle, from which we remove the edge from S_{-1} to its neighbour in τ_{-1}, which is S_0.

So there are $n-1$ possible trees τ_0 from which we can reach τ_{-1} going backwards in time. These moves all must have probability $1/(n-1)$.

So for fixed τ', $q(\tau, \tau') = 0$ or $1/(n-1)$.

The Aldous-Broder algorithm: proof

So the matrix $Q = (q(\tau, \tau'))$ is doubly stochastic and so must have uniform stationary distribution on the set of rooted trees.

The Aldous-Broder algorithm: proof

So the matrix $Q = (q(\tau, \tau'))$ is doubly stochastic and so must have uniform stationary distribution on the set of rooted trees.

It's straightforward to show that the chain is irreducible and since the root is uniformly distributed, it follows that τ_0 is a uniform random rooted tree. The result follows from forgetting the root.

Remark. There is a more general version of this result, for weighted trees.

A variant due to Aldous

"Do the labelling as we go, then relabel at the end."

Let $U_2, \ldots, U_n$ be uniform on $[n]$.

1. Start from the vertex labelled 1.
2. For $2 \leq i \leq n$, connect vertex i to vertex $V_i = \min\{U_i, i-1\}$.
3. Take a uniform random permutation of the labels.

A variant due to Aldous

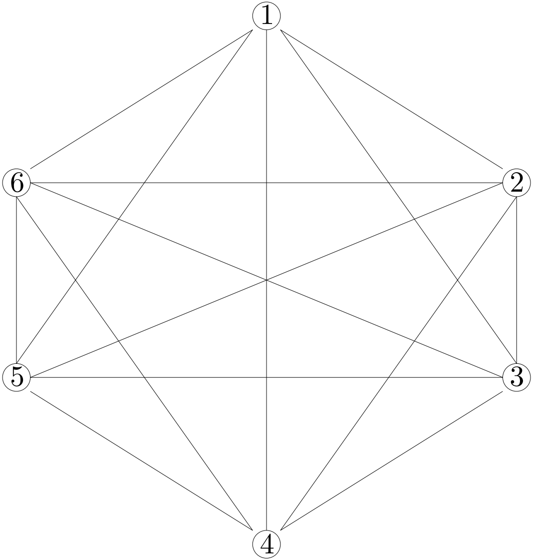
"Do the labelling as we go, then relabel at the end."

1. Start from the vertex labelled 1.
2. For $2 \leq i \leq n$, connect vertex i to vertex V_i such that

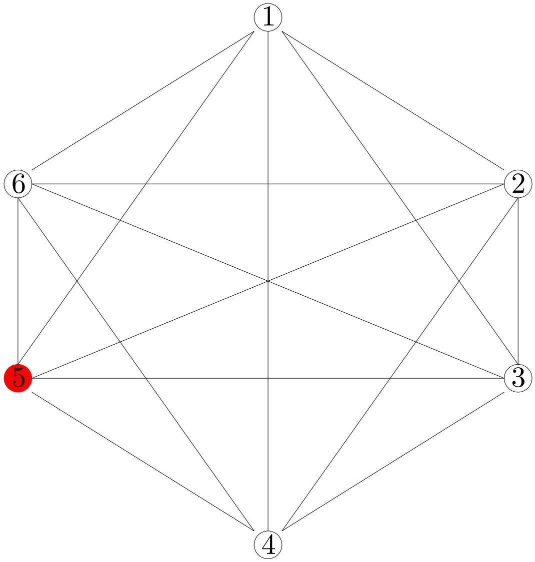
$$V_i = \begin{cases} i-1 \text{ with probability } 1 - \frac{i-2}{n-1} \\ \text{uniform on } \{1, 2, \ldots, i-2\} \text{ otherwise.} \end{cases}$$

3. Take a uniform random permutation of the labels.

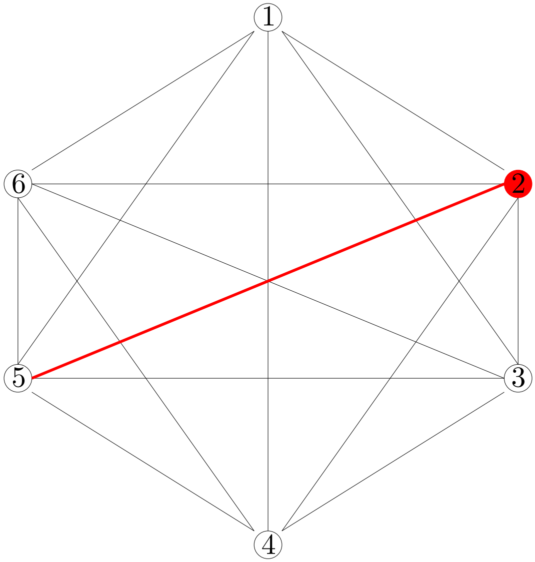
The Aldous-Broder algorithm



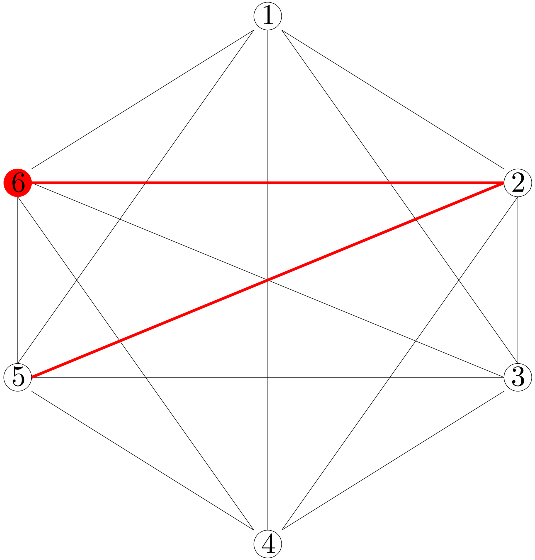
The Aldous-Broder algorithm



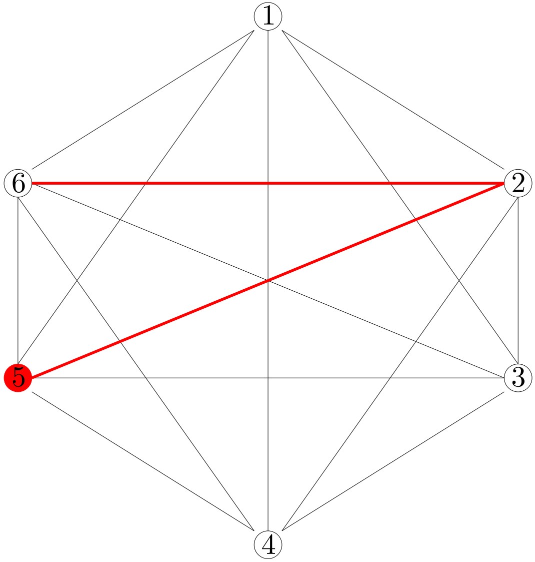
The Aldous-Broder algorithm



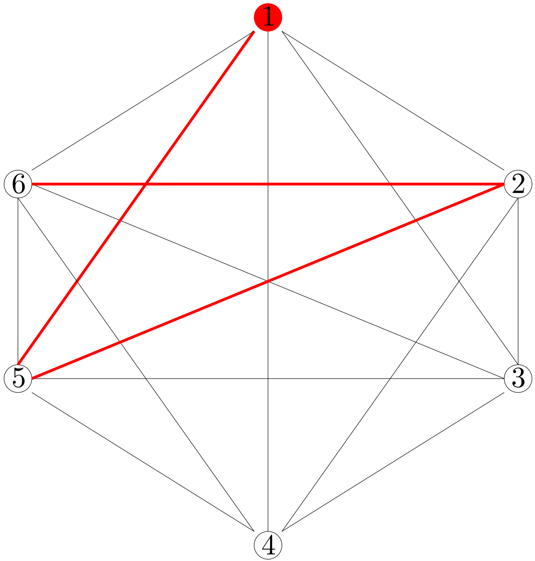
The Aldous-Broder algorithm



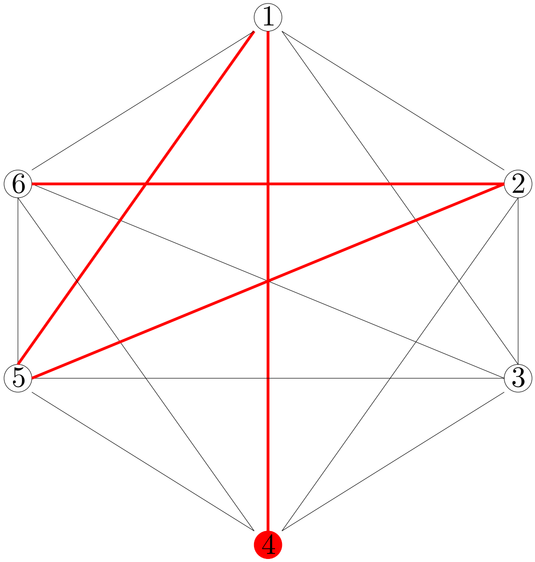
The Aldous-Broder algorithm



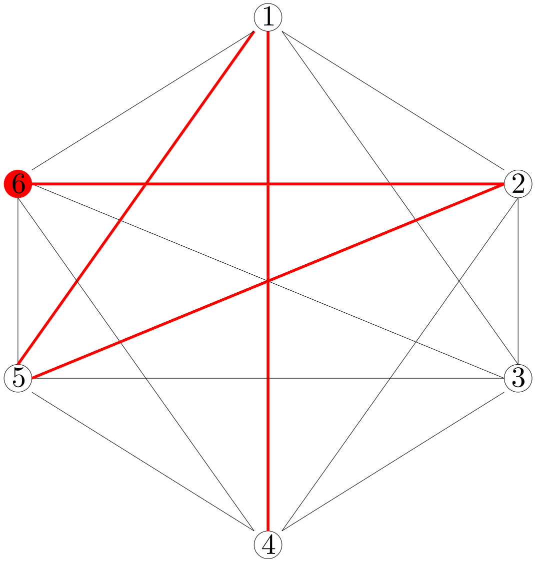
The Aldous-Broder algorithm



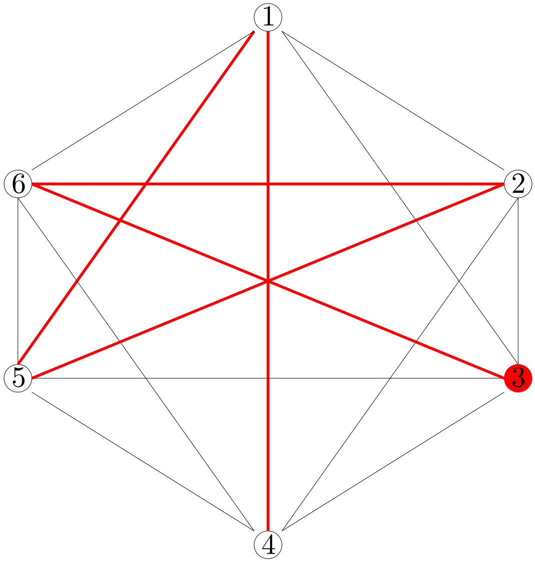
The Aldous-Broder algorithm



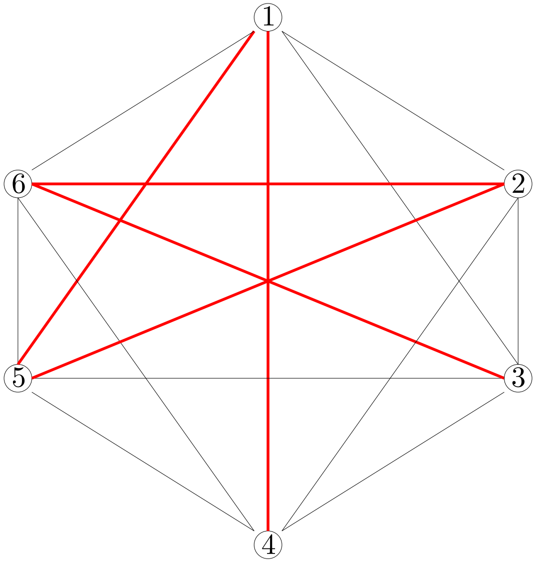
The Aldous-Broder algorithm



The Aldous-Broder algorithm



The Aldous-Broder algorithm

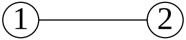


Aldous' algorithm

Consider $n = 10$.

Aldous' algorithm

$V_2 = 1$ with probability 1

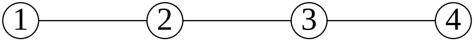


Aldous' algorithm

$$V_3 = \begin{cases} 1 & \text{with probability } 1/9 \\ 2 & \text{with probability } 8/9 \end{cases}$$

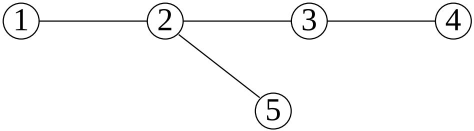
Aldous' algorithm

$$V_4 = \begin{cases} j & \text{with probability } 1/9,\ 1 \leq j \leq 2 \\ 3 & \text{with probability } 7/9 \end{cases}$$



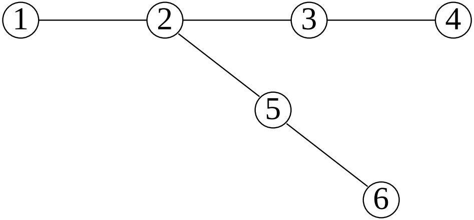
Aldous' algorithm

$$V_5 = \begin{cases} j & \text{with probability } 1/9,\ 1 \le j \le 3 \\ 4 & \text{with probability } 6/9 \end{cases}$$



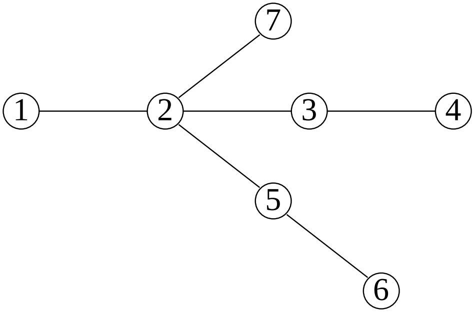
Aldous' algorithm

$$V_6 = \begin{cases} j & \text{with probability } 1/9,\ 1 \le j \le 4 \\ 5 & \text{with probability } 5/9 \end{cases}$$



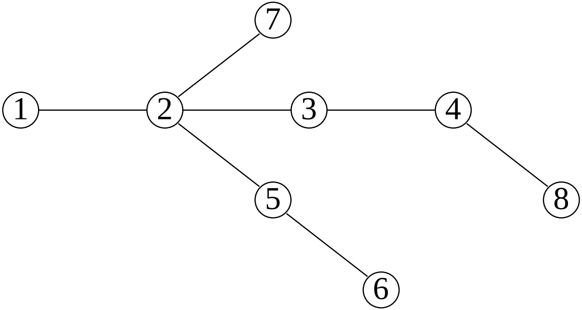
Aldous' algorithm

$$V_7 = \begin{cases} j & \text{with probability } 1/9,\ 1 \le j \le 5 \\ 6 & \text{with probability } 4/9 \end{cases}$$



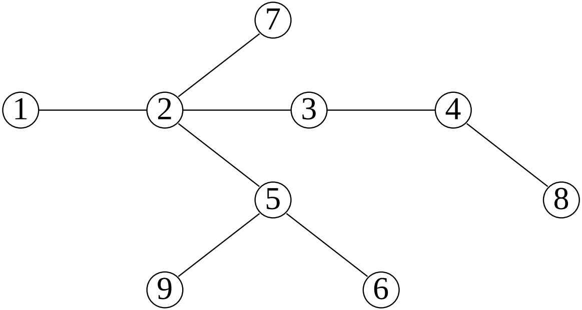
Aldous' algorithm

$$V_8 = \begin{cases} j & \text{with probability } 1/9,\ 1 \le j \le 6 \\ 7 & \text{with probability } 3/9 \end{cases}$$



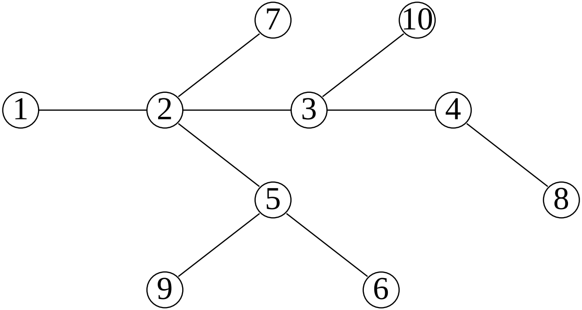
Aldous' algorithm

$$V_9 = \begin{cases} j & \text{with probability } 1/9,\ 1 \le j \le 7 \\ 8 & \text{with probability } 2/9 \end{cases}$$



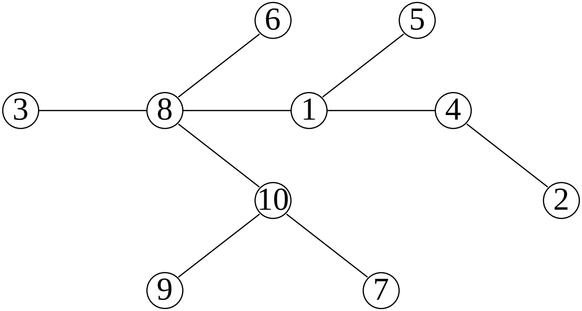
Aldous' algorithm

$$V_{10} = \begin{cases} j & \text{with probability } 1/9,\ 1 \le j \le 8 \\ 9 & \text{with probability } 1/9 \end{cases}$$



Aldous' algorithm

Permute.



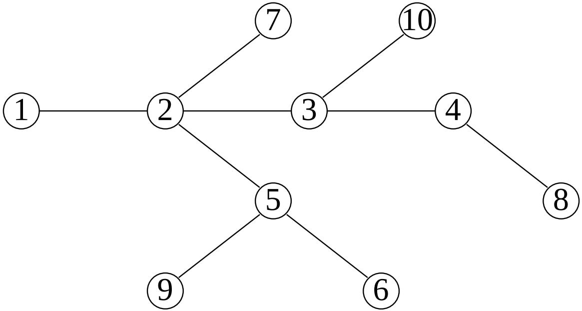
Typical distances

Consider the tree before we permute. Let

$$C_1^n = \inf\{i \geq 2 : V_i \neq i-1\}.$$

We can use C_1^n to give us an idea of typical distances in the tree.

In our example, $C_1^{10} = 5$:



Typical distances

For $2 \leq i \leq n$, connect vertex i to vertex V_i such that

$$V_i = \begin{cases} i-1 \text{ with probability } 1 - \frac{i-2}{n-1} \\ \text{uniform on } \{1, 2, \ldots, i-2\} \text{ otherwise.} \end{cases}$$

$C_1^n = \inf\{i \geq 2 : V_i \neq i-1\}$

Proposition. $n^{-1/2} C_1^n$ converges in distribution as $n \to \infty$.

Typical distances

Once we have built this first stick of consecutive labels, we pick a uniform starting point along that stick and attach a new stick with a random length, and so on.

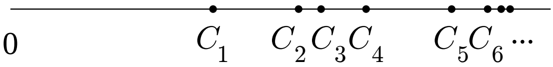
Typical distances

Once we have built this first stick of consecutive labels, we pick a uniform starting point along that stick and attach a new stick with a random length, and so on.

Imagine now that edges in the tree have length 1. The proposition suggests that rescaling edge-lengths by $n^{-1/2}$ will give some sort of limit for the whole tree. The limiting version of the algorithm is as follows.

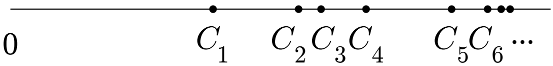
Line-breaking procedure

Let $C_1, C_2, \ldots$ be the points of an inhomogeneous Poisson process on $\mathbb{R}^+$ of intensity t at t. Equivalently, take $E_1, E_2, \ldots$ to be i.i.d. Exponential(1) and set $C_k = \sqrt{2\sum_{i=1}^{k} E_k}$.



Line-breaking procedure

Let $C_1, C_2, \ldots$ be the points of an inhomogeneous Poisson process on $\mathbb{R}^+$ of intensity t at t. Equivalently, take $E_1, E_2, \ldots$ to be i.i.d. Exponential(1) and set $C_k = \sqrt{2 \sum_{i=1}^{k} E_k}$.



(Note that
$\mathbb{P}\left(C_1 > x\right) = \exp\left(-\int_0^x t dt\right) = \exp(-x^2/2) = \mathbb{P}\left(E_1 > x^2/2\right)$.)

Line-breaking procedure

Consider the line-segments $[0, C_1)$, $[C_1, C_2), \ldots$.

Line-breaking procedure

Consider the line-segments $[0, C_1)$, $[C_1, C_2), \ldots$.

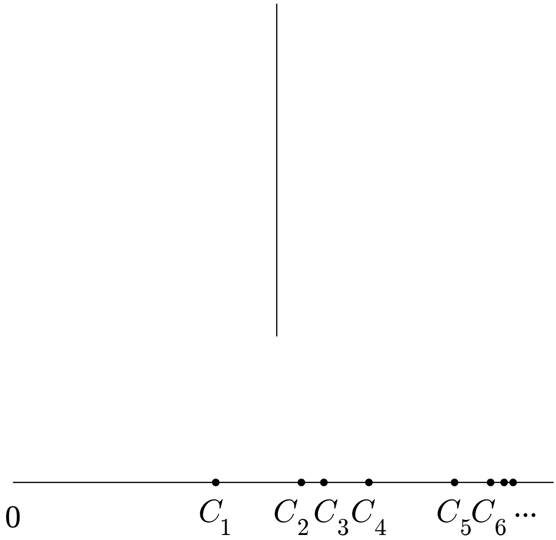
Start from $[0, C_1)$ and proceed inductively.

Line-breaking procedure

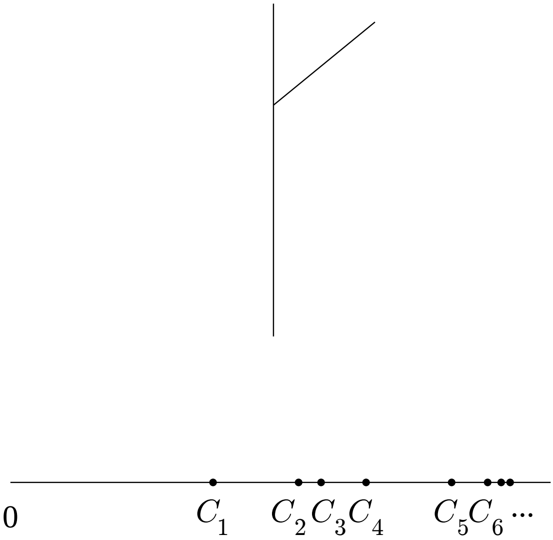
Consider the line-segments $[0, C_1)$, $[C_1, C_2), \ldots$.

Start from $[0, C_1)$ and proceed inductively. For $i \geq 1$, sample B_i uniformly from $[0, C_i)$ and attach $[C_i, C_{i+1})$ at the corresponding point of the tree constructed so far (this is a point chosen uniformly at random over the existing tree).

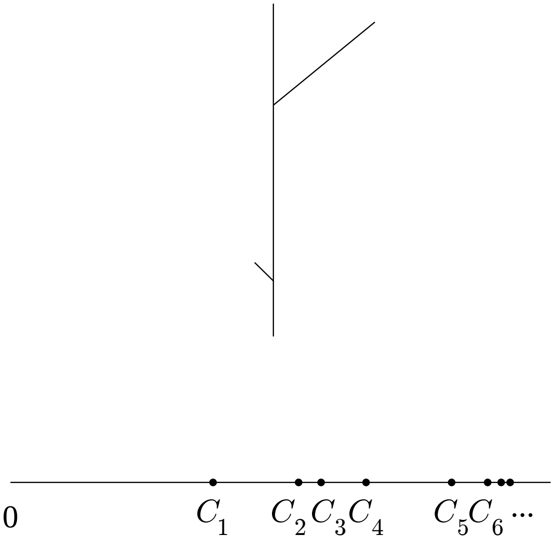
Line-breaking procedure



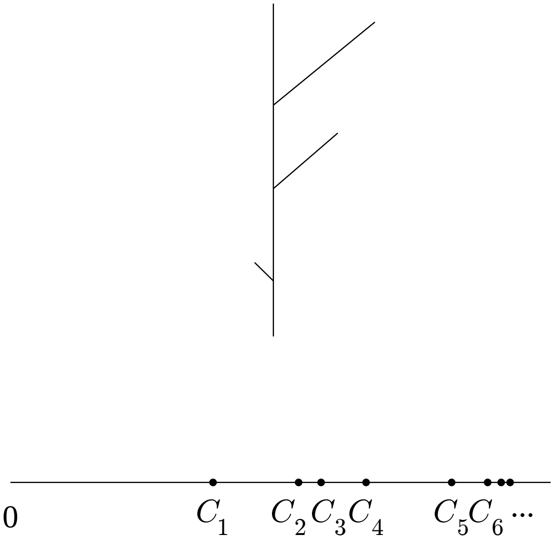
Line-breaking procedure



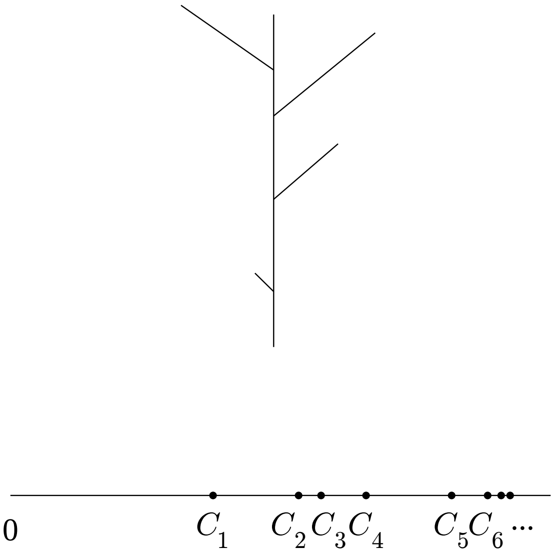
Line-breaking procedure



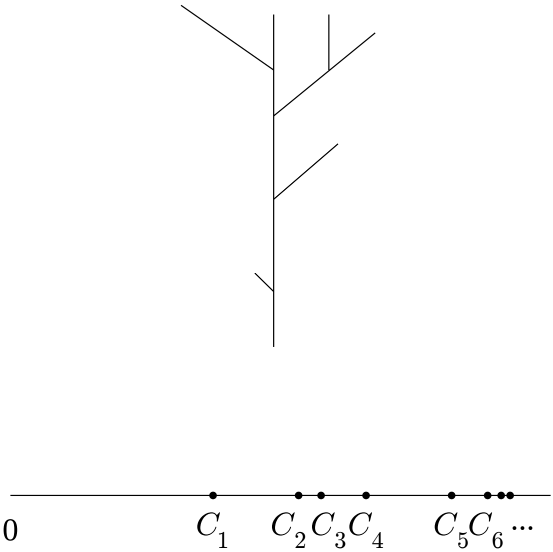
Line-breaking procedure



Line-breaking procedure



Line-breaking procedure



Line-breaking procedure

Start from $[0, C_1)$ and proceed inductively.

For $i \geq 1$, sample B_i uniformly from $[0, C_i)$ and attach $[C_i, C_{i+1})$ at the corresponding point of the tree constructed so far (this is a point chosen uniformly at random over the existing tree).

Take the union of all the branches, thought of as a metric space, and then take its completion.

Line-breaking procedure

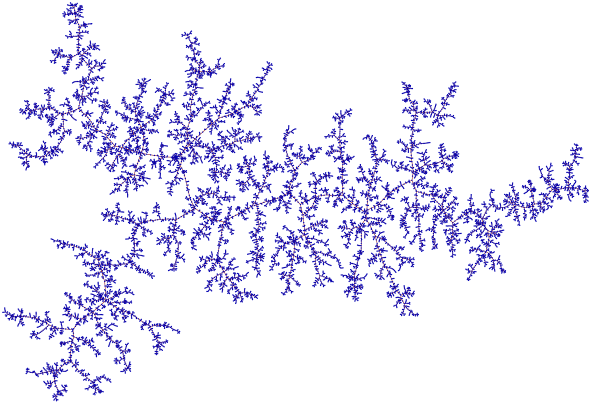
Start from $[0, C_1)$ and proceed inductively.

For $i \geq 1$, sample B_i uniformly from $[0, C_i)$ and attach $[C_i, C_{i+1})$ at the corresponding point of the tree constructed so far (this is a point chosen uniformly at random over the existing tree).

Take the union of all the branches, thought of as a metric space, and then take its completion.

This procedure gives (slightly informally expressed) definition of Aldous' Brownian continuum random tree (CRT) which will be the key object in this minicourse.

A first look at the Brownian CRT



[Picture by Igor Kortchemski]

The scaling limit of the uniform random tree

Theorem. (Aldous (1991)) As $n \to \infty$,

$$\frac{1}{\sqrt{n}} T_n \xrightarrow{d} \mathcal{T},$$

where $\mathcal{T}$ is the Brownian CRT.

A very brief idea of a proof

Recall that we had

$$C_1^n = \inf\{i \geq 2 : V_i \neq i - 1\}.$$

More generally, for $k \geq 1$, define C_k^n to be the kth element of the set $\{i \geq 2 : V_i \neq i - 1\}$ i.e. the kth cut-time.

Let $B_k^n = V_{C_k^n}$, the kth branch-point.

Then the heart of the proof is the fact that

$$\left(\frac{1}{\sqrt{n}}(C_1^n, B_1^n), \frac{1}{\sqrt{n}}(C_2^n, B_2^n), \ldots \right) \xrightarrow{d} ((C_1, B_1), (C_2, B_2), \ldots)$$

as $n \to \infty$.

The scaling limit of the uniform random tree

Theorem. (Aldous (1991)) As $n \to \infty$,

$$\frac{1}{\sqrt{n}} T_n \xrightarrow{d} \mathcal{T},$$

where $\mathcal{T}$ is the Brownian CRT.

Of course, before we can really make sense of this theorem, we need to know what sort of objects we're really dealing with, and what is the topology in which the convergence occurs.

The scaling limit of the uniform random tree

Theorem. (Aldous (1991)) As $n \to \infty$,

$$\frac{1}{\sqrt{n}} T_n \xrightarrow{d} \mathcal{T},$$

where $\mathcal{T}$ is the Brownian CRT.

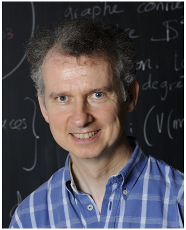
Of course, before we can really make sense of this theorem, we need to know what sort of objects we're really dealing with, and what is the topology in which the convergence occurs.

We will, in fact, sketch a proof of a more general result.

2. DISCRETE TREES

Key reference:

Jean-François Le Gall, **Random trees and applications**, *Probability Surveys* **2** (2005) pp.245-311.



Ordered trees

It turns out to be helpful to work with rooted, ordered trees (also called plane trees).

Ordered trees

It turns out to be helpful to work with rooted, ordered trees (also called plane trees).

This is not too much of a restriction if what we're really interested in is labelled unordered trees, since it's always possible to obtain a rooted ordered tree from a labelled one: for example, root at the vertex labelled 1 and order the children of a vertex from left to right in increasing order of label.

Ordered trees

We will use the Ulam-Harris labelling. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ and

$$\mathcal{U} = \bigcup_{n=0}^{\infty} \mathbb{N}^n,$$

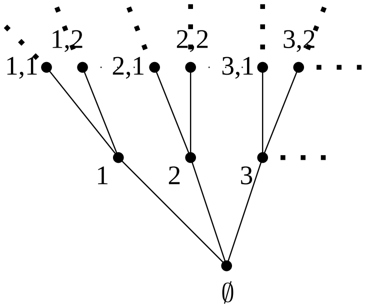
where $\mathbb{N}^0 = \{\emptyset\}$.

Ordered trees

We will use the Ulam-Harris labelling. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ and

$$\mathcal{U} = \bigcup_{n=0}^{\infty} \mathbb{N}^n,$$

where $\mathbb{N}^0 = \{\emptyset\}$. An element $u \in \mathcal{U}$ is a sequence $u = (u^1, u^2, \ldots, u^n)$ of natural numbers representing a point in an infinitary tree:



Thus the label of a vertex indicates its genealogy.

Ordered trees

Write $|u| = n$ for the generation of u.

u has parent $p(u) = (u^1, u^2, \dots, u^{n-1})$.

u has children $u1, u2, \dots$ where, in general,
$uv = (u^1, u^2, \dots, u^n, v^1, v^2, \dots, v^m)$ is the concatenation of sequences $u = (u^1, u^2, \dots, u^n)$ and $v = (v^1, v^2, \dots, v^m)$.

We root the tree at $\emptyset$.

Ordered trees

A (finite) rooted, ordered tree $\mathbf{t}$ is a finite 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 $c(u) \in \mathbb{Z}_+$ such that for $j \in \mathbb{N}$, $uj \in \mathbf{t}$ iff $1 \leq j \leq c(u)$.

Ordered trees

A (finite) rooted, ordered tree $\mathbf{t}$ is a finite 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 $c(u) \in \mathbb{Z}_+$ such that for $j \in \mathbb{N}$, $uj \in \mathbf{t}$ iff $1 \leq j \leq c(u)$.

$c(u)$ is the number of children of u in $\mathbf{t}$.

Ordered trees

A (finite) rooted, ordered tree $\mathbf{t}$ is a finite 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 $c(u) \in \mathbb{Z}_+$ such that for $j \in \mathbb{N}$, $uj \in \mathbf{t}$ iff $1 \leq j \leq c(u)$.

$c(u)$ is the number of children of u in $\mathbf{t}$.

Write $\#(\mathbf{t})$ for the size (number of vertices) of $\mathbf{t}$ and note that

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

Ordered trees

A (finite) rooted, ordered tree $\mathbf{t}$ is a finite 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 $c(u) \in \mathbb{Z}_+$ such that for $j \in \mathbb{N}$, $uj \in \mathbf{t}$ iff $1 \leq j \leq c(u)$.

$c(u)$ is the number of children of u in $\mathbf{t}$.

Write $\#(\mathbf{t})$ for the size (number of vertices) of $\mathbf{t}$ and note that

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

Write $\mathbf{T}$ for the set of all rooted ordered trees.

Two ways of encoding a tree

Consider a rooted ordered tree $\mathbf{t} \in \mathbf{T}$.

It will be convenient to encode this tree in terms of discrete functions which are easier to manipulate.

Two ways of encoding a tree

Consider a rooted ordered tree $\mathbf{t} \in \mathbf{T}$.

It will be convenient to encode this tree in terms of discrete functions which are easier to manipulate.

We will do this is two different ways:

- the height function
- the depth-first walk.

Height function

Suppose that $\mathbf{t}$ has n vertices. Let them be $v_0, v_1, \ldots, v_{n-1}$, listed in lexicographical order.

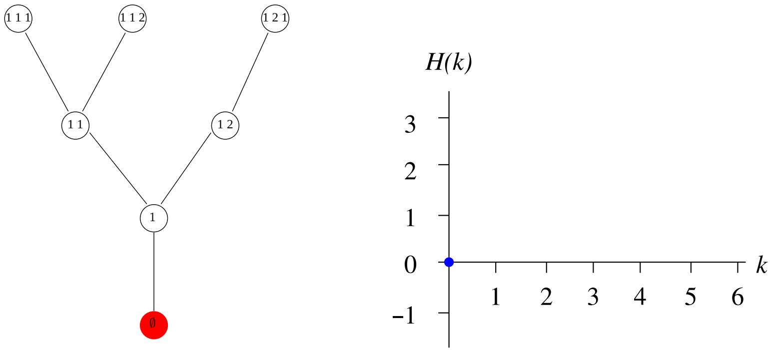
Height function

Suppose that $\mathbf{t}$ has n vertices. Let them be $v_0, v_1, \ldots, v_{n-1}$, listed in lexicographical order.

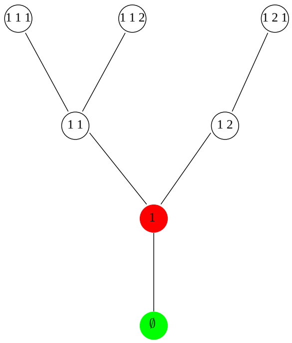
Then the height function is defined by

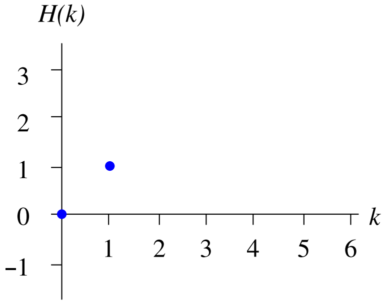
$$H(k) = |v_k|, \quad 0 \leq k \leq n-1.$$

Height function

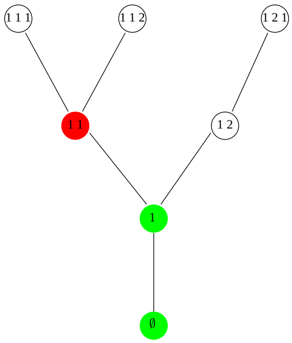


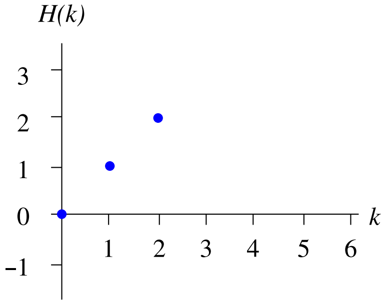
Height function



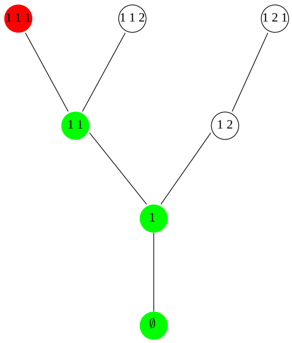


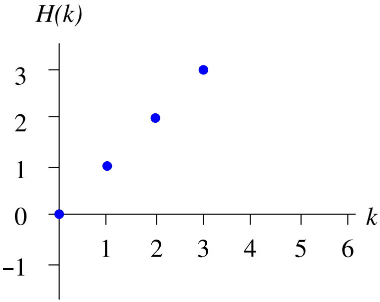
Height function



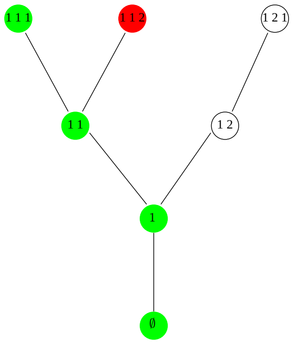


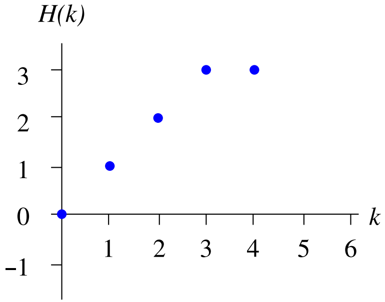
Height function



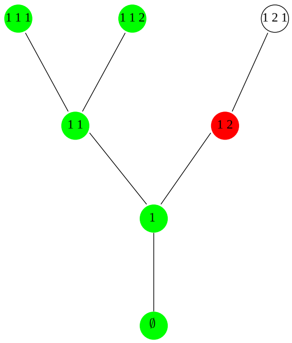


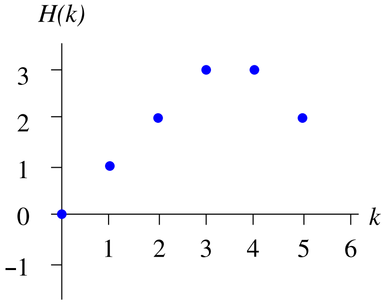
Height function



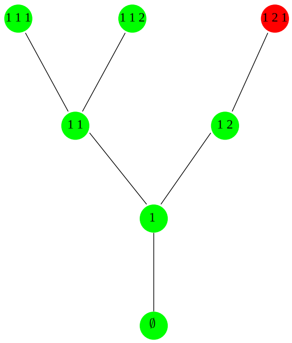


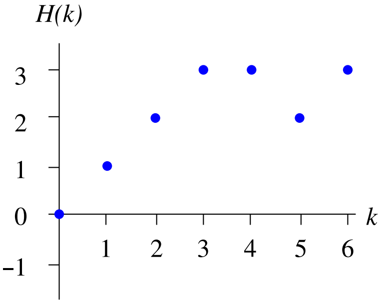
Height function



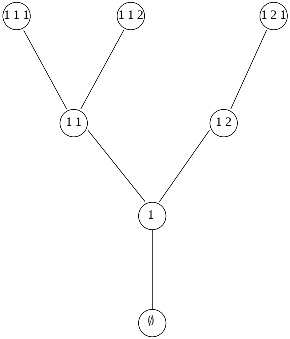


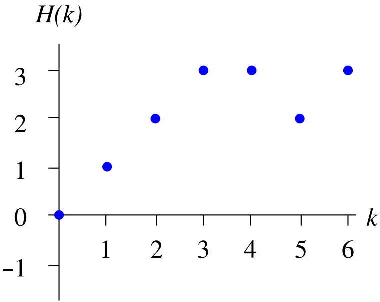
Height function





Height function





We can recover the tree from its height function.

Depth-first walk

Recall that $c(v)$ is the number of children of v, and that $v_0, v_1, \ldots, v_{n-1}$ is a list of the vertices of $\mathbf{t}$ in lexicographical order.

Depth-first walk

Recall that $c(v)$ is the number of children of v, and that $v_0, v_1, \ldots, v_{n-1}$ is a list of the vertices of $\mathbf{t}$ in lexicographical order.

Define

$$\begin{aligned} X(0) &= 0, \\ X(i) &= \sum_{j=0}^{i-1} (c(v_j) - 1), \text{ for } 1 \leq i \leq n. \end{aligned}$$

Depth-first walk

Recall that $c(v)$ is the number of children of v, and that $v_0, v_1, \ldots, v_{n-1}$ is a list of the vertices of $\mathbf{t}$ in lexicographical order.

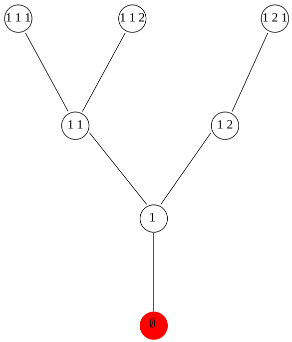
Define

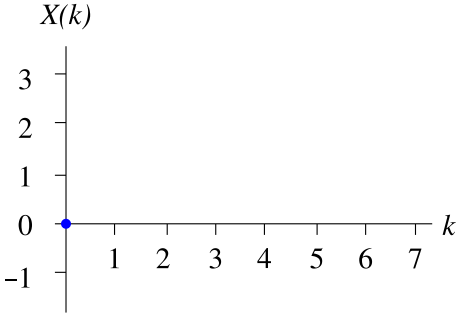
$$X(0) = 0,$$
$$X(i) = \sum_{j=0}^{i-1} (c(v_j) - 1), \text{ for } 1 \leq i \leq n.$$

In other words,

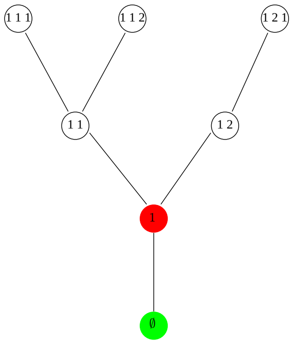
$$X(i+1) = X(i) + c(v_i) - 1, \quad 0 \leq i \leq n-1.$$

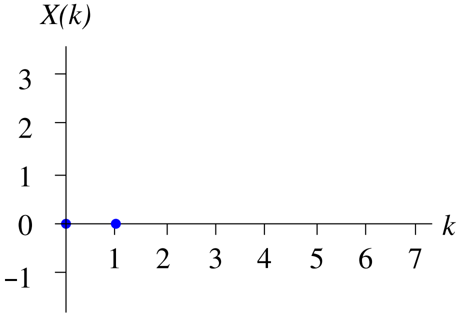
Depth-first walk



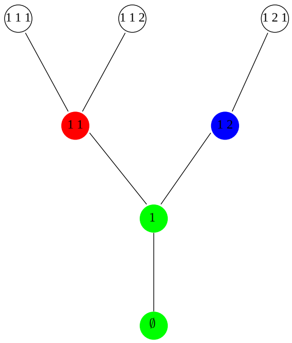


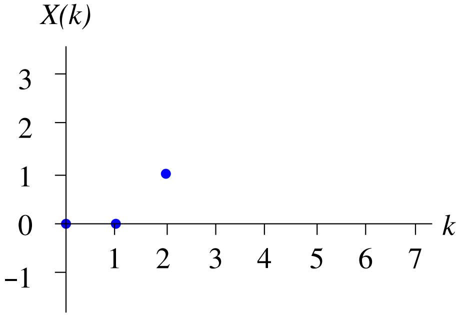
Depth-first walk



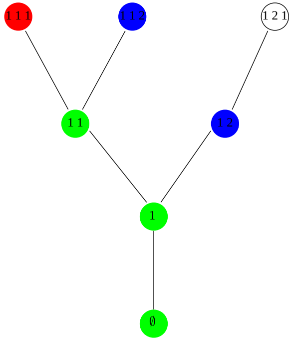


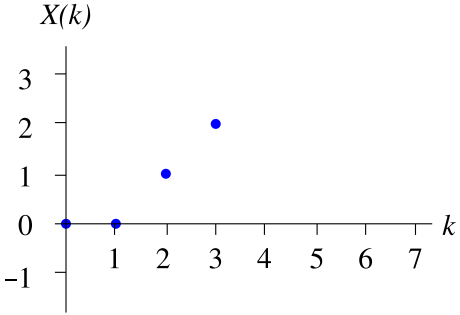
Depth-first walk



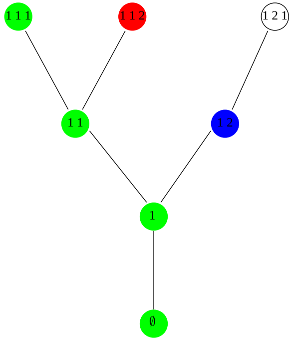


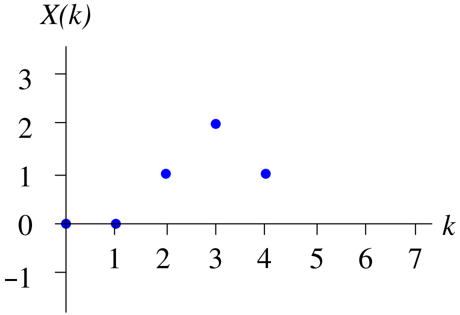
Depth-first walk



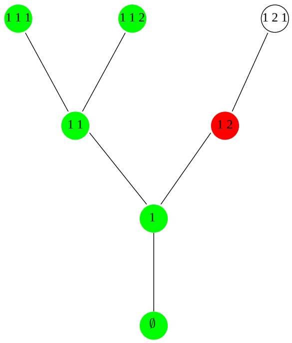


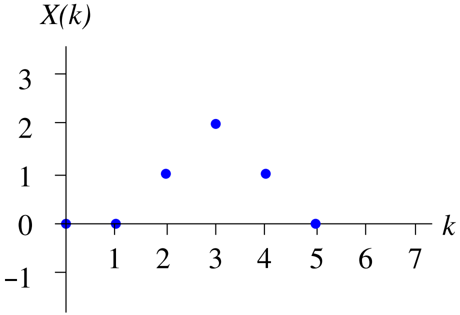
Depth-first walk



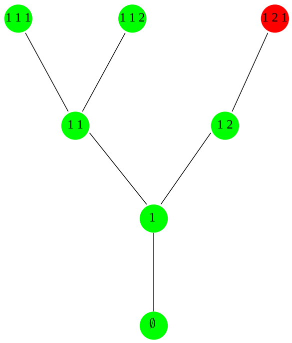


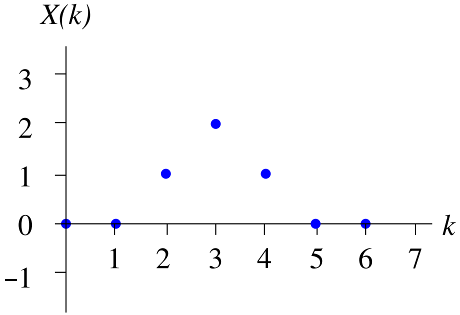
Depth-first walk



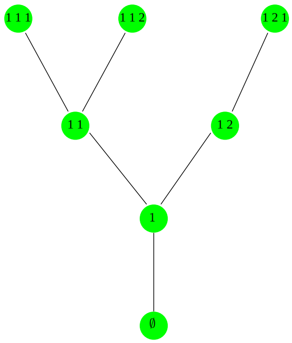


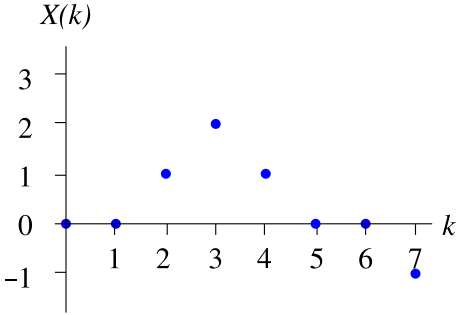
Depth-first walk



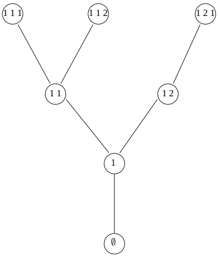


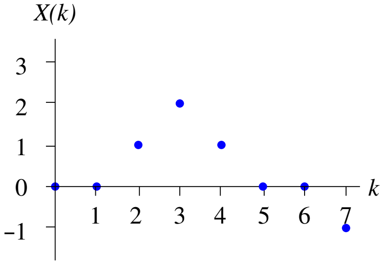
Depth-first walk





Depth-first walk





It is less easy to see that the depth-first walk also encodes the tree.

Proposition. For $0 \leq i \leq n-1$,

$$H(i) = \#\left\{0 \leq j \leq i-1 : X(j) = \min_{j \leq k \leq i} X(k)\right\}.$$

Random discrete trees

From a probabilistic perspective, a natural probability measure on trees is that generated by the usual Galton-Watson branching process. We will see in a moment that this is a good thing to do from a combinatorial perspective too!

Galton-Watson process

A Galton-Watson branching process $(Z_n)_{n \geq 0}$ describes the size of a population which evolves as follows:

Galton-Watson process

A Galton-Watson branching process $(Z_n)_{n \geq 0}$ describes the size of a population which evolves as follows:

- Start with a single individual.

Galton-Watson process

A Galton-Watson branching process $(Z_n)_{n \geq 0}$ describes the size of a population which evolves as follows:

- Start with a single individual.
- This individual has a number of children distributed according to the offspring distribution p, where $p(k)$ gives the probability of k children, $k \geq 0$.

Galton-Watson process

A Galton-Watson branching process $(Z_n)_{n \geq 0}$ describes the size of a population which evolves as follows:

- Start with a single individual.
- This individual has a number of children distributed according to the offspring distribution p, where $p(k)$ gives the probability of k children, $k \geq 0$.
- Each child reproduces as an independent copy of the original individual.

Galton-Watson process

A Galton-Watson branching process $(Z_n)_{n \geq 0}$ describes the size of a population which evolves as follows:

- Start with a single individual.
- This individual has a number of children distributed according to the offspring distribution p, where $p(k)$ gives the probability of k children, $k \geq 0$.
- Each child reproduces as an independent copy of the original individual.

Z_n gives the number of individuals in generation n (in particular, $Z_0 = 1$).

Galton-Watson trees

A Galton-Watson tree is the family tree arising from a Galton-Watson branching process.

Galton-Watson trees

A Galton-Watson tree is the family tree arising from a Galton-Watson branching process. We will think of this as a rooted ordered tree.

Galton-Watson trees

A Galton-Watson tree is the family tree arising from a Galton-Watson branching process. We will think of this as a rooted ordered tree.

We will consider the case where the offspring distribution p is critical i.e.

$$\sum_{k=1}^{\infty} kp(k) = 1.$$

This ensures, in particular, that the resulting tree, T, is finite.

Galton-Watson trees

A Galton-Watson tree is the family tree arising from a Galton-Watson branching process. We will think of this as a rooted ordered tree.

We will consider the case where the offspring distribution p is critical i.e.

$$\sum_{k=1}^{\infty} kp(k) = 1.$$

This ensures, in particular, that the resulting tree, T, is finite.

Since the tree is random, we will refer to the height process rather than function.

Uniform random trees revisited

Proposition. Let T be a (rooted, ordered) Galton-Watson tree, with Poisson(1) offspring distribution and total progeny N.

Uniform random trees revisited

Proposition. Let T be a (rooted, ordered) Galton-Watson tree, with Poisson(1) offspring distribution and total progeny N. Assign the vertices labels uniformly at random from $\{1, 2, \ldots, N\}$ and then forget the ordering and the root. Let $\tilde{T}$ be the labelled tree obtained.

Uniform random trees revisited

Proposition. Let T be a (rooted, ordered) Galton-Watson tree, with Poisson(1) offspring distribution and total progeny N. Assign the vertices labels uniformly at random from $\{1, 2, \ldots, N\}$ and then forget the ordering and the root. Let $\tilde{T}$ be the labelled tree obtained. Then, conditional on $N = n$, $\tilde{T}$ has the same distribution as T_n, a uniform random tree on n vertices.

Other combinatorial trees (in disguise)

Let T be a Galton-Watson tree with offspring distribution p and total progeny N.

Other combinatorial trees (in disguise)

Let T be a Galton-Watson tree with offspring distribution p and total progeny N.

- If $p(k) = 2^{-k-1}$, $k \geq 0$ (i.e. Geometric(1/2) offspring distribution) then conditional on $N = n$, the tree is uniform on the set of ordered trees with n vertices.

Other combinatorial trees (in disguise)

Let T be a Galton-Watson tree with offspring distribution p and total progeny N.

- If $p(k) = 2^{-k-1}$, $k \geq 0$ (i.e. Geometric(1/2) offspring distribution) then conditional on $N = n$, the tree is uniform on the set of ordered trees with n vertices.
- If $p(0) = 1/2$ and $p(2) = 1/2$ then, conditional on $N = n$ (for n odd), the tree is uniform on the set of (complete) binary trees.

The depth-first walk of a Galton-Watson process is a stopped random walk

Recall that p is a distribution on $\mathbb{Z}_+$ such that $\sum_{k=1}^{\infty} kp(k) = 1$.

The depth-first walk of a Galton-Watson process is a stopped random walk

Recall that p is a distribution on $\mathbb{Z}_+$ such that $\sum_{k=1}^{\infty} kp(k) = 1$.

Proposition. Let $(R(k), k \geq 0)$ be a random walk with initial value 0 and step distribution $\nu(k) = p(k+1), k \geq -1$. Set

$$M = \inf\{k \geq 0 : R(k) = -1\}.$$

The depth-first walk of a Galton-Watson process is a stopped random walk

Recall that p is a distribution on $\mathbb{Z}_+$ such that $\sum_{k=1}^{\infty} kp(k) = 1$.

Proposition. Let $(R(k), k \geq 0)$ be a random walk with initial value 0 and step distribution $\nu(k) = p(k+1), k \geq -1$. Set

$$M = \inf\{k \geq 0 : R(k) = -1\}.$$

Now suppose that T is a Galton-Watson tree with offspring distribution p and total progeny N. Then

$$(X(k), 0 \leq k \leq N) \stackrel{d}{=} (R(k), 0 \leq k \leq M).$$

The depth-first walk of a Galton-Watson process is a stopped random walk

Recall that p is a distribution on $\mathbb{Z}_+$ such that $\sum_{k=1}^{\infty} kp(k) = 1$.

Proposition. Let $(R(k), k \geq 0)$ be a random walk with initial value 0 and step distribution $\nu(k) = p(k+1), k \geq -1$. Set

$$M = \inf\{k \geq 0 : R(k) = -1\}.$$

Now suppose that T is a Galton-Watson tree with offspring distribution p and total progeny N. Then

$$(X(k), 0 \leq k \leq N) \stackrel{d}{=} (R(k), 0 \leq k \leq M).$$

[Careful proof: see Le Gall.]

Galton-Watson trees conditioned on their total progeny: finite variance case

Suppose now that we have offspring variance
$\sigma^2 := \sum_{k=1}^{\infty}(k-1)^2 p(k) \in (0, \infty)$.

The depth-first walk X is a random walk with step mean 0 and variance σ^2, stopped at the first time it hits -1. The underlying random walk has a Brownian motion as its scaling limit, by Donsker's theorem.

The total progeny N is equal to $\inf\{k \geq 0 : X(k) = -1\}$. We want to condition on the event $\{N = n\}$.

Galton-Watson trees conditioned on their total progeny: finite variance case

Suppose now that we have offspring variance $\sigma^2 := \sum_{k=1}^{\infty}(k-1)^2 p(k) \in (0, \infty)$.

The depth-first walk X is a random walk with step mean 0 and variance σ^2, stopped at the first time it hits -1. The underlying random walk has a Brownian motion as its scaling limit, by Donsker's theorem.

The total progeny N is equal to $\inf\{k \geq 0 : X(k) = -1\}$. We want to condition on the event $\{N = n\}$.

Standing assumption: $\mathbb{P}(N = n) > 0$ for all n sufficiently large.

Galton-Watson trees conditioned on their total progeny: finite variance case

Write $(X^n(k), 0 \leq k \leq n)$ for the depth-first walk conditioned on $\{N = n\}$. Then there is a conditional version of Donsker's theorem:

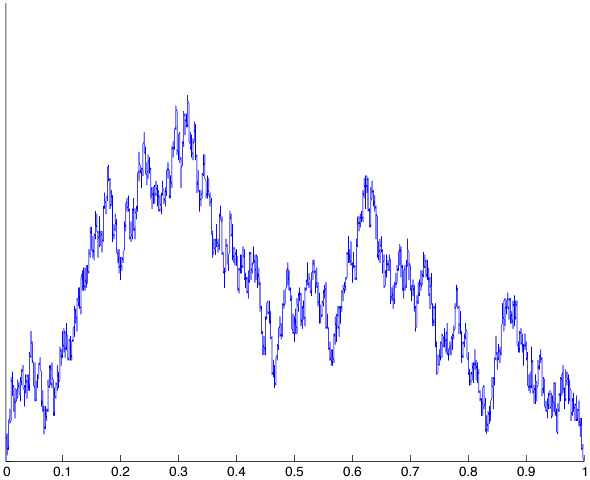
Theorem. As $n \to \infty$,

$$\frac{1}{\sigma\sqrt{n}}(X^n(\lfloor nt \rfloor), 0 \leq t \leq 1) \xrightarrow{d} (e(t), 0 \leq t \leq 1),$$

where $(e(t), 0 \leq t \leq 1)$ is a standard Brownian excursion.

[See W.D. Kaigh, **An invariance principle for random walk conditioned by a late return to zero**, *Annals of Probability* **4** (1976) pp.115-121.]

Brownian excursion



[Picture by Igor Kortchemski]

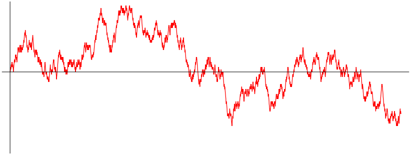
Brownian excursion

There are several (equivalent) definitions of this process.

Brownian excursion

There are several (equivalent) definitions of this process.

For example, let W be a standard Brownian motion.



Fix $s > 0$. Let

$$g_s = \sup\{t \leq s : W(t) = 0\} \text{ and } d_s = \inf\{t \geq s : W(t) = 0\}.$$

Note that $W(s) \neq 0$ with probability 1, so that $\mathbb{P}(g_s < s < d_s) = 1$. Then for $t \in [0, 1]$ define

$$e(t) = \frac{|W(g_s + t(d_s - g_s))|}{\sqrt{d_s - g_s}}.$$

It turns out that the distribution of $(e(t), 0 \leq t \leq 1)$ is independent of s.

Convergence of the coding processes

Let $(H^n(i), 0 \leq i \leq n)$ be the height process of a critical Galton-Watson tree with offspring variance $\sigma^2 \in (0, \infty)$, conditioned to have total progeny n.

Convergence of the coding processes

Let $(H^n(i), 0 \leq i \leq n)$ be the height process of a critical Galton-Watson tree with offspring variance $\sigma^2 \in (0, \infty)$, conditioned to have total progeny n.

Theorem. As $n \to \infty$,

$$\frac{\sigma}{\sqrt{n}} \left(H^n(\lfloor nt \rfloor), 0 \leq t \leq 1 \right) \xrightarrow{d} 2 \left(e(t), 0 \leq t \leq 1 \right),$$

where $(e(t), 0 \leq t \leq 1)$ is a standard Brownian excursion.

Actually, I'm going to cheat...

Consider the unconditioned random walk $(X(k), k \geq 0)$ (without stopping) and, as usual, let the height process be H, where $H(0) = 0$ and, for $i \geq 1$,

$$H(i) = \# \left\{ 0 \leq j \leq i - 1 : X(j) = \min_{j \leq k \leq i} X(k) \right\}.$$

Actually, I'm going to cheat...

Consider the unconditioned random walk $(X(k), k \geq 0)$ (without stopping) and, as usual, let the height process be H, where $H(0) = 0$ and, for $i \geq 1$,

$$H(i) = \# \left\{ 0 \leq j \leq i - 1 : X(j) = \min_{j \leq k \leq i} X(k) \right\}.$$

(A little thought shows that this is the height process of a sequence of i.i.d. unconditioned Galton-Watson trees.)

Actually, I'm going to cheat...

Consider the unconditioned random walk $(X(k), k \geq 0)$ (without stopping) and, as usual, let the height process be H, where $H(0) = 0$ and, for $i \geq 1$,

$$H(i) = \# \left\{ 0 \leq j \leq i - 1 : X(j) = \min_{j \leq k \leq i} X(k) \right\}.$$

(A little thought shows that this is the height process of a sequence of i.i.d. unconditioned Galton-Watson trees.)

We have

$$\frac{1}{\sigma \sqrt{n}} (X(\lfloor nt \rfloor), t \geq 0) \xrightarrow{d} (W(t), t \geq 0)$$

as $n \to \infty$.

An unconditioned result

Proposition. As $n \to \infty$,

$$\frac{\sigma}{\sqrt{n}}(H(\lfloor nt \rfloor), t \geq 0) \to 2\left(W(t) - \min_{0 \leq s \leq t} W(s), t \geq 0\right)$$

in the sense of finite-dimensional distributions, i.e. if $0 \leq t_1 \leq t_2 \leq \cdots \leq t_m$ then

$$\frac{\sigma}{\sqrt{n}}(H(\lfloor nt_1 \rfloor), \ldots, H(\lfloor nt_m \rfloor))$$
$$\xrightarrow{d} 2\left(W(t_1) - \min_{0 \leq s \leq t_1} W(s), \ldots, W(t_m) - \min_{0 \leq s \leq t_m} W(s)\right).$$

[Approach due to Marckert & Mokkadem, **The depth first processes of Galton-Watson trees converge to the same Brownian excursion**, *Annals of Probability* **31** (2003), pp.1655-1678]

Lecture 2

Aim: the scaling limit of a critical Galton-Watson tree with finite offspring variance

Theorem. (Aldous (1993), Le Gall (2005)) Let T_n be a Galton-Watson tree with critical offspring distribution and finite offspring variance $\sigma^2 \in (0, \infty)$, conditioned to have total progeny n. Then as $n \to \infty$,

$$\frac{\sigma}{\sqrt{n}} T_n \xrightarrow{d} \mathcal{T},$$

where $\mathcal{T}$ is the Brownian CRT.

Recap

$(X(k), k \geq 0)$ is the depth-first walk of a sequence of i.i.d. Galton-Watson trees.

$(H(k), k \geq 0)$ is the height process, defined by $H(0) = 0$ and, for $i \geq 1$,

$$H(i) = \# \left\{ 0 \leq j \leq i-1 : X(j) = \min_{j \leq k \leq i} X(k) \right\}.$$

We have

$$\frac{1}{\sigma\sqrt{n}}(X(\lfloor nt \rfloor), t \geq 0) \xrightarrow{d} (W(t), t \geq 0)$$

as $n \to \infty$.

Recap

Proposition. As $n \to \infty$,

$$\frac{\sigma}{\sqrt{n}}(H(\lfloor nt \rfloor), t \geq 0) \to 2\left(W(t) - \min_{0 \leq s \leq t} W(s), t \geq 0\right)$$

in the sense of finite-dimensional distributions, i.e. if $0 \leq t_1 \leq t_2 \leq \cdots \leq t_m$ then

$$\frac{\sigma}{\sqrt{n}}(H(\lfloor nt_1 \rfloor), \ldots, H(\lfloor nt_m \rfloor))$$
$$\xrightarrow{d} 2\left(W(t_1) - \min_{0 \leq s \leq t_1} W(s), \ldots, W(t_m) - \min_{0 \leq s \leq t_m} W(s)\right).$$

[Approach due to Marckert & Mokkadem, **The depth first processes of Galton-Watson trees converge to the same Brownian excursion**, *Annals of Probability* **31** (2003), pp.1655-1678]

Conditioned version

Let $(H^n(i), 0 \leq i \leq n)$ be the height process of a critical Galton-Watson tree with offspring variance $\sigma^2 \in (0, \infty)$, conditioned to have total progeny n.

Theorem. As $n \to \infty$,

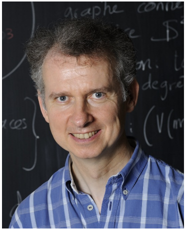
$$\frac{\sigma}{\sqrt{n}} \left(H^n(\lfloor nt \rfloor), 0 \leq t \leq 1 \right) \xrightarrow{d} 2 \left(e(t), 0 \leq t \leq 1 \right),$$

where $(e(t), 0 \leq t \leq 1)$ is a standard Brownian excursion.

3. $\mathbb{R}$-TREES

Key reference:

Jean-François Le Gall, **Random trees and applications**,
Probability Surveys **2** (2005) pp.245-311.



Continuous trees

We want a continuous notion of a tree. We don't really care about vertices: the important aspects are the shape of the tree and the distances. So it makes sense to think in terms of metric spaces.

$\mathbb{R}$-trees

Definition. A compact metric space $(\mathcal{T}, d)$ is an $\mathbb{R}$-tree if for all $x, y \in \mathcal{T}$,

- There exists a unique shortest path $[[x, y]]$ from x to y (of length $d(x, y)$).

- The only non-self-intersecting path from x to y is $[[x, y]]$.

$\mathbb{R}$-trees

Definition. A compact metric space $(\mathcal{T}, d)$ is an $\mathbb{R}$-tree if for all $x, y \in \mathcal{T}$,

- There exists a unique shortest path $[[x, y]]$ from x to y (of length $d(x, y)$). (There is a unique isometric map $f_{x,y}$ from $[0, d(x, y)]$ into $\mathcal{T}$ such that $f(0) = x$ and $f(d(x, y)) = y$. We write $f_{x,y}([0, d(x, y)]) = [[x, y]]$.)
- The only non-self-intersecting path from x to y is $[[x, y]]$. (If g is a continuous injective map from $[0, 1]$ into $\mathcal{T}$, such that $g(0) = x$ and $g(1) = y$, then $g([0, 1]) = [[x, y]]$.)

$\mathbb{R}$-trees

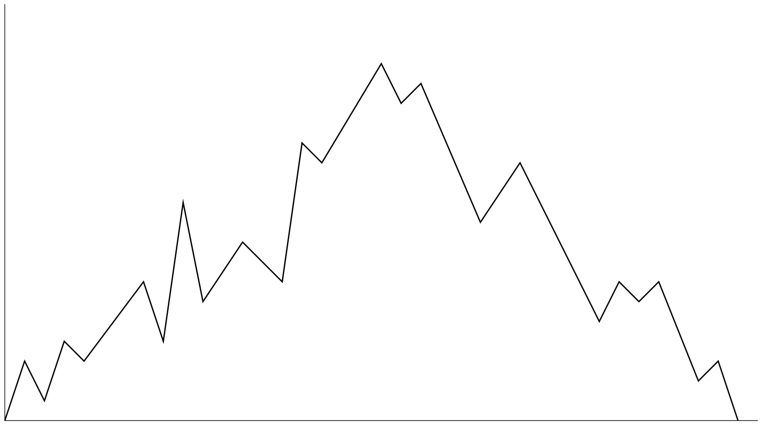
Definition. A compact metric space $(\mathcal{T}, d)$ is an $\mathbb{R}$-tree if for all $x, y \in \mathcal{T}$,

- There exists a unique shortest path $[[x, y]]$ from x to y (of length $d(x, y)$). (There is a unique isometric map $f_{x,y}$ from $[0, d(x, y)]$ into $\mathcal{T}$ such that $f(0) = x$ and $f(d(x, y)) = y$. We write $f_{x,y}([0, d(x, y)]) = [[x, y]]$.)
- The only non-self-intersecting path from x to y is $[[x, y]]$. (If g is a continuous injective map from $[0, 1]$ into $\mathcal{T}$, such that $g(0) = x$ and $g(1) = y$, then $g([0, 1]) = [[x, y]]$.)

An element $v \in \mathcal{T}$ is called a vertex.
A rooted $\mathbb{R}$-tree has a distinguished vertex ρ called the root.
The height of a vertex v is its distance $d(\rho, v)$ from the root.
A leaf is a vertex v such that $v \notin [[\rho, w]]$ for any $w \neq v$.

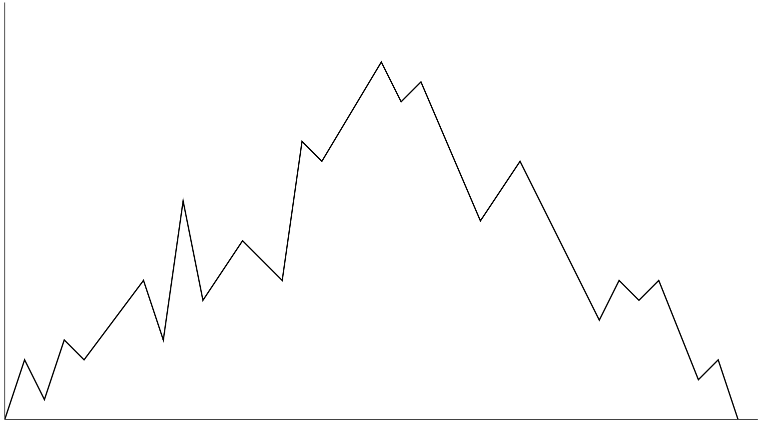
Coding $\mathbb{R}$-trees

Let $h : [0, 1] \to \mathbb{R}^+$ be an excursion, that is a continuous function such that $h(0) = h(1) = 0$ and $h(x) > 0$ for $x \in (0, 1)$. h will play the role of the height process for an $\mathbb{R}$-tree.



Coding $\mathbb{R}$-trees

Now put glue on the underside of the excursion and push the two sides together...



Coding $\mathbb{R}$-trees

Now put glue on the underside of the excursion and push the two sides together...

Coding $\mathbb{R}$-trees

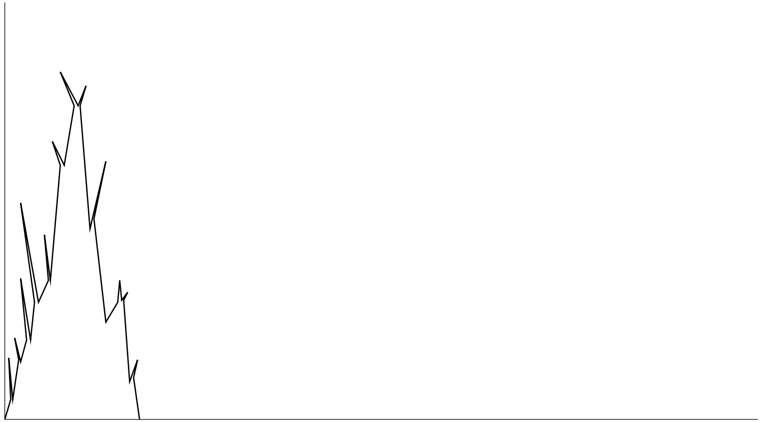
Now put glue on the underside of the excursion and push the two sides together...

Coding $\mathbb{R}$-trees

Now put glue on the underside of the excursion and push the two sides together...

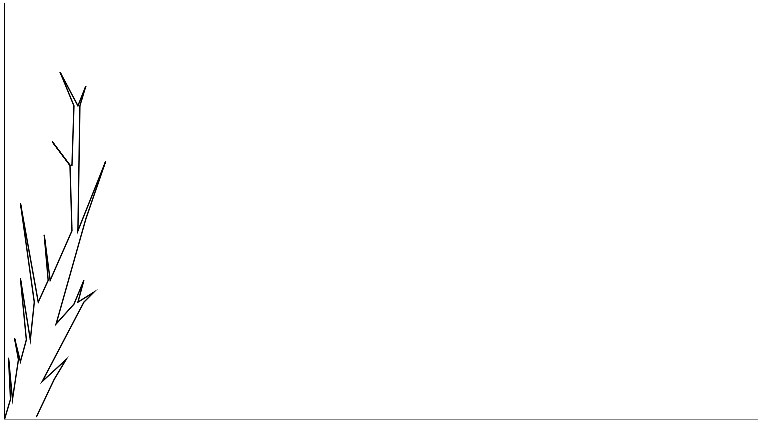
Coding $\mathbb{R}$-trees

Now put glue on the underside of the excursion and push the two sides together...



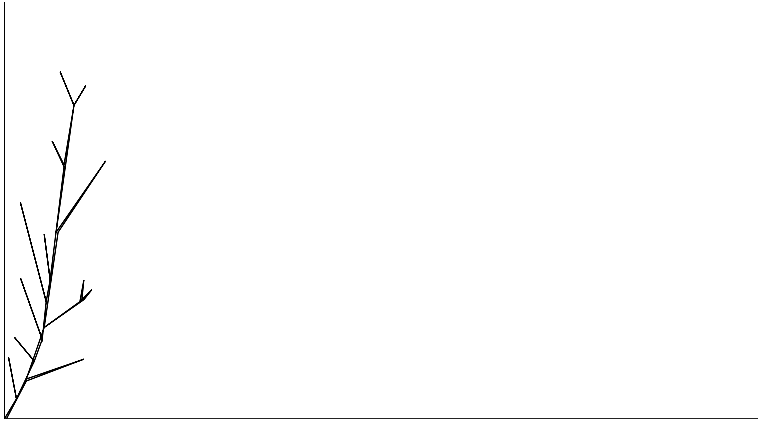
Coding $\mathbb{R}$-trees

Now put glue on the underside of the excursion and push the two sides together...



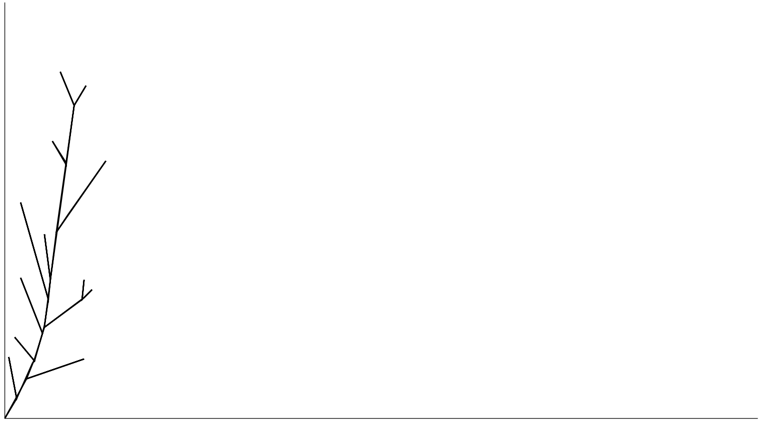
Coding $\mathbb{R}$-trees

Now put glue on the underside of the excursion and push the two sides together...



Coding $\mathbb{R}$-trees

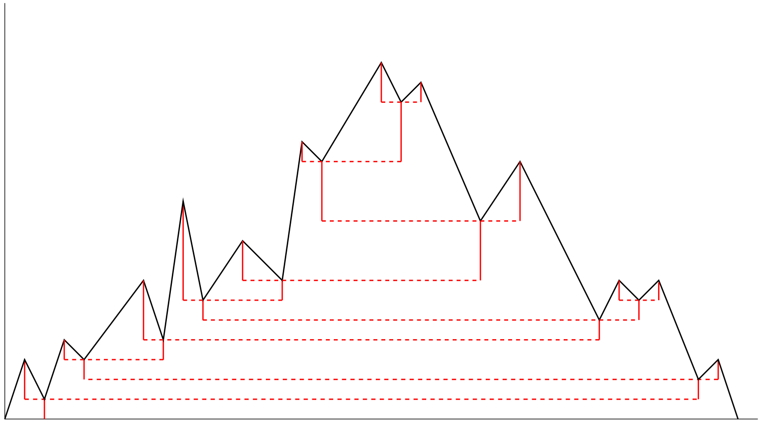
Now put glue on the underside of the excursion and push the two sides together to get a tree.



Coding $\mathbb{R}$-trees

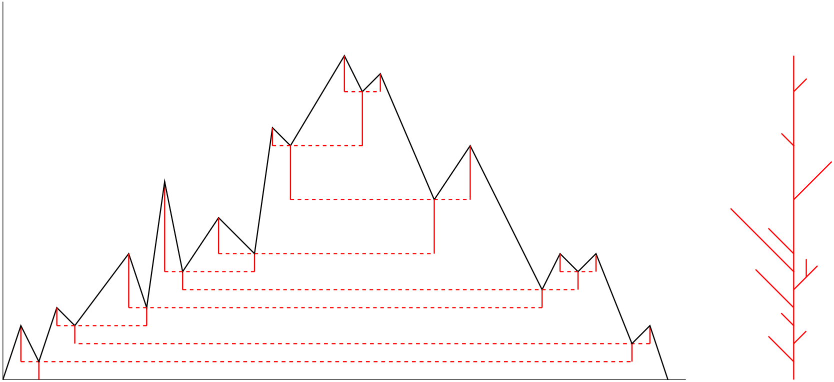
Formally, use h to define a distance:

$$d_h(x, y) = h(x) + h(y) - 2 \inf_{x \wedge y \leq z \leq x \vee y} h(z).$$



Coding $\mathbb{R}$-trees

Let $y \sim y'$ if $d_h(y, y') = 0$ and take the quotient $\mathcal{T}_h = [0, 1]/\sim$.



Coding $\mathbb{R}$-trees

Theorem. For any excursion h, $(\mathcal{T}_h, d_h)$ is an $\mathbb{R}$-tree.

[Proof: see Le Gall.]

Coding $\mathbb{R}$-trees

Theorem. For any excursion h, $(\mathcal{T}_h, d_h)$ is an $\mathbb{R}$-tree.

[Proof: see Le Gall.]

We will always take the equivalence class of 0 to be the root, ρ.

Coding $\mathbb{R}$-trees

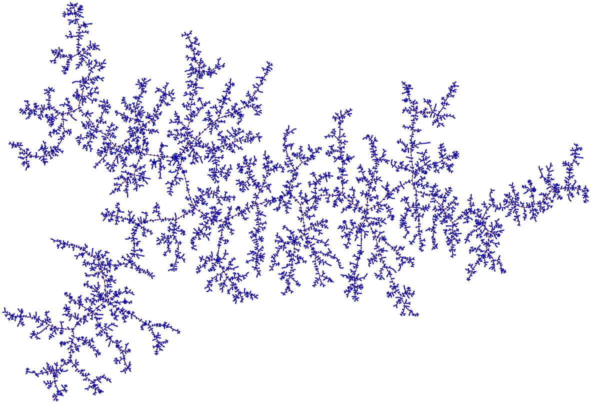
Theorem. For any excursion h, $(\mathcal{T}_h, d_h)$ is an $\mathbb{R}$-tree.

[Proof: see Le Gall.]

We will always take the equivalence class of 0 to be the root, ρ.

Definition. The Brownian continuum random tree is $(\mathcal{T}_{2e}, d_{2e})$, where e is a standard Brownian excursion.

The Brownian continuum random tree $\mathcal{T}_{2e}$

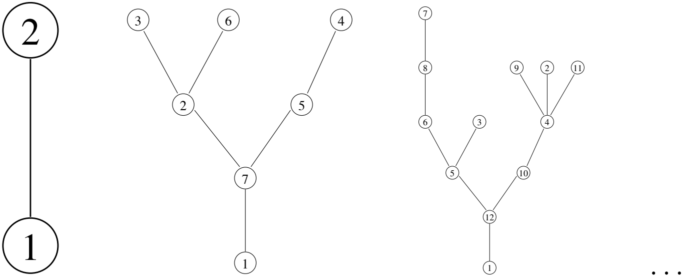


[Picture by Igor Kortchemski]

Discrete trees as metric spaces

We want to think of $(T_n, n \geq 1)$ as metric spaces.

The vertices of T_n come equipped with a natural metric: the graph distance d_{gr}.



We sometimes write aT_n for the metric space (T_n, ad_{gr}) given by the vertices of T_n with the graph distance scaled by a.

Convergence in distribution

What is the the sense of the convergence in distribution

$$(T_n, \sigma d_{\mathrm{gr}}/\sqrt{n}) \xrightarrow{d} (\mathcal{T}_{2e}, d_{2e}) \quad \text{as } n \to \infty?$$

Convergence in distribution

Consider the space, $\mathbb{M}$, of compact metric spaces up to isometry. We'll define a metric d_{GH} on $\mathbb{M}$ in a moment. Then

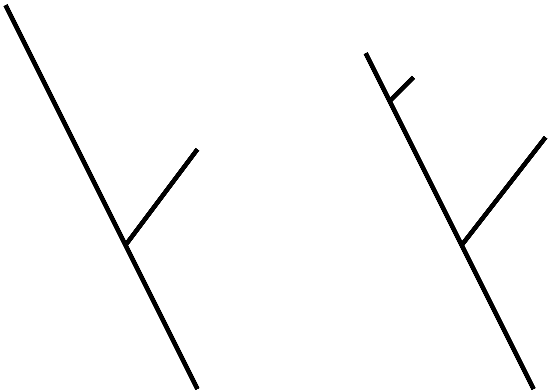
$$(T_n, \sigma d_{\mathrm{gr}}/\sqrt{n}) \xrightarrow{d} (\mathcal{T}_{2e}, d_{2e}) \quad \text{as } n \to \infty$$

will mean that for any bounded function $f : \mathbb{M} \to \mathbb{R}$ which is continuous with respect to d_{GH}, we have

$$\mathbb{E}\left[f\left((T_n, \sigma d_{\mathrm{gr}}/\sqrt{n})\right)\right] \to \mathbb{E}\left[f\left((\mathcal{T}_{2e}, d_{2e})\right)\right] \quad \text{as } n \to \infty.$$

Measuring the distance between compact metric spaces

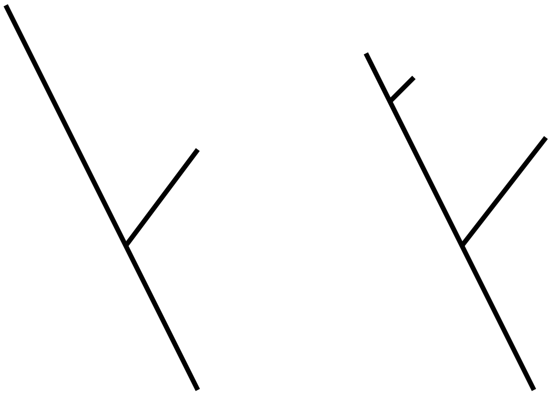
Suppose that (X, d) and (X', d') are compact metric spaces.



Measuring the distance between compact metric spaces

Suppose that (X, d) and (X', d') are compact metric spaces.

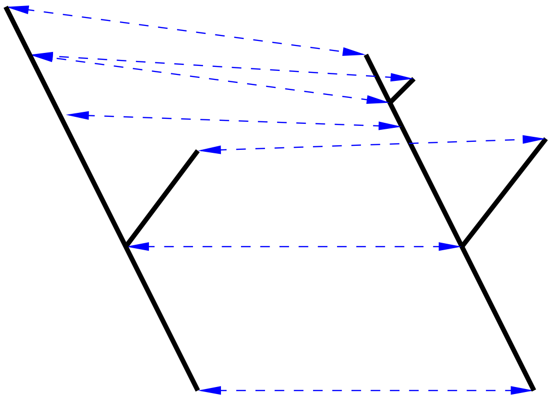
A correspondence R is a subset of $X \times X'$ such that for every $x \in X$, there exists $x' \in X'$ with $(x, x') \in R$ and vice versa.



Measuring the distance between compact metric spaces

Suppose that (X, d) and (X', d') are compact metric spaces.

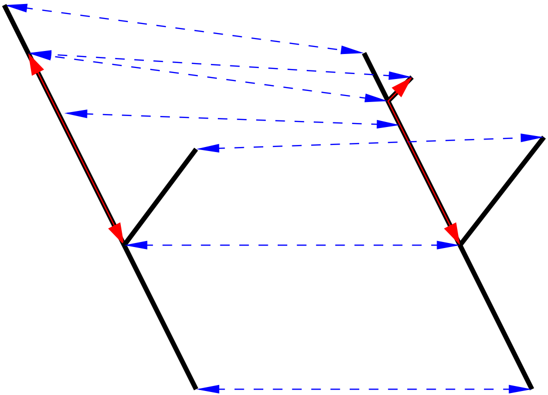
A correspondence R is a subset of $X \times X'$ such that for every $x \in X$, there exists $x' \in X'$ with $(x, x') \in R$ and vice versa.



Measuring the distance between compact metric spaces

The distortion of R is

$$\operatorname{dis}(R) = \sup\{|d(x,y) - d'(x',y')| : (x,x'),(y,y') \in R\}.$$



Measuring the distance between compact metric spaces

The Gromov-Hausdorff distance between (X, d) and (X', d') is

$$\mathrm{d}_{\mathrm{GH}}((X, d), (X', d')) = \frac{1}{2} \inf_{R} \mathrm{dis}(R).$$

Measuring the distance between compact metric spaces

The Gromov-Hausdorff distance between (X, d) and (X', d') is

$$\mathrm{d}_{\mathrm{GH}}((X, d), (X', d')) = \frac{1}{2} \inf_{R} \mathrm{dis}(R).$$

(There exists an equivalent definition which more closely resembles that of the usual Hausdorff distance, but this one is easier to use.)

Measuring the distance between compact metric spaces

The Gromov-Hausdorff distance between (X, d) and (X', d') is

$$\mathrm{d_{GH}}((X, d), (X', d')) = \frac{1}{2} \inf_R \mathrm{dis}(R).$$

(There exists an equivalent definition which more closely resembles that of the usual Hausdorff distance, but this one is easier to use.)

Recall that $\mathbb{M}$ is the space of compact metric spaces, up to isometry.

Theorem. $(\mathbb{M}, \mathrm{d_{GH}})$ is a complete separable metric space.

[Proof: see Evans, Pitman and Winter, **Rayleigh processes, real trees, and root growth with re-grafting**, *Probability Theory and Related Fields* **134** (2006) pp.81-126.]

Convergence to the Brownian CRT

Let T_n be our Galton-Watson tree conditioned to have size n.

Write H^n for its height process and recall that

$$\frac{\sigma}{\sqrt{n}}(H^n(\lfloor nt \rfloor), 0 \le t \le 1) \xrightarrow{d} 2(e(t), 0 \le t \le 1),$$

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

Convergence to the Brownian CRT

Let T_n be our Galton-Watson tree conditioned to have size n.

Write H^n for its height process and recall that

$$\frac{\sigma}{\sqrt{n}}(H^n(\lfloor nt \rfloor), 0 \leq t \leq 1) \xrightarrow{d} 2(e(t), 0 \leq t \leq 1),$$

where $(e(t), 0 \leq t \leq 1)$ is a standard Brownian excursion.

Theorem. (Aldous (1993), Le Gall (2005)) As $n \to \infty$,

$$\left(T_n, \frac{\sigma}{\sqrt{n}} d_{\mathrm{gr}} \right) \xrightarrow{d} (\mathcal{T}_{2e}, d_{2e}),$$

where convergence is in the Gromov-Hausdorff sense.

[Approach due to Grégory Miermont.]

Universality

We started with the uniform random labelled tree, and then generalised to conditioned critical Galton-Watson trees with finite offspring variance. So the Brownian CRT is the universal scaling limit of a whole class of trees. In fact, this class is much larger!

Universality

Some other examples of trees (and graphs!) with the Brownian CRT as their scaling limit are:

- uniform unordered unlabelled rooted trees
- uniform unordered unlabelled unrooted trees
- critical multi-type Galton-Watson trees
- random trees with a prescribed degree sequence satisfying certain conditions
- random dissections
- random graphs from subcritical classes.

See Benedikt Stufler's talk on Thursday for more details.

Applications

Universal scaling limits often show up in other places, and the Brownian CRT is no exception. It appears, for example, in

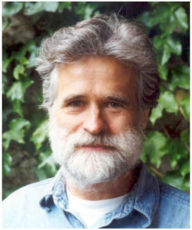
- the scaling limit of random planar maps [Le Gall, Miermont];
- the scaling limit of the critical Erdős-Rényi random graph [Addario-Berry, Broutin, G. (2010)].

4. THE BROWNIAN CONTINUUM RANDOM TREE

Key references:

David Aldous, **The continuum random tree III**, *Annals of Probability* **21** (1993) pp.248-289.

Jim Pitman, **Combinatorial stochastic processes**, *Lecture notes in mathematics* **1875**, Springer-Verlag, Berlin (2006).



What is a continuum random tree?!

A continuum tree is a triple $(\mathcal{T}, d, \mu)$ where $(\mathcal{T}, d)$ is an $\mathbb{R}$-tree with leaves $\mathcal{L}(\mathcal{T})$ and μ is a Borel probability measure on $\mathcal{T}$ which is non-atomic and satisfies

- $\mu(\mathcal{L}(\mathcal{T})) = 1$;
- for every $v \in \mathcal{T}$ of degree $k \geq 2$, let $\mathcal{T}_1, \ldots, \mathcal{T}_k$ be the connected components of $\mathcal{T} \setminus \{v\}$. Then $\mu(\mathcal{T}_i) > 0$ for all $1 \leq i \leq k$.

What is a continuum random tree?!

We can endow the set of continuum trees with a generalisation of the Gromov-Hausdorff topology, the Gromov-Hausdorff-Prokhorov topology, which takes account of the measure also.

What is a continuum random tree?!

We can endow the set of continuum trees with a generalisation of the Gromov-Hausdorff topology, the Gromov-Hausdorff-Prokhorov topology, which takes account of the measure also.

Idea: take two compact measured metric spaces, and find a correspondence between them. In addition to minimising the distortion of the correspondence, find a coupling of the two measures which puts as small mass as possible outside the correspondence.

What is a continuum random tree?!

We can endow the set of continuum trees with a generalisation of the Gromov-Hausdorff topology, the Gromov-Hausdorff-Prokhorov topology, which takes account of the measure also.

Idea: take two compact measured metric spaces, and find a correspondence between them. In addition to minimising the distortion of the correspondence, find a coupling of the two measures which puts as small mass as possible outside the correspondence.

A continuum random tree (CRT) is a random variable taking values in the set of continuum trees.

The mass measure of the Brownian CRT

Let μ_{2e} be the push-forward of Lebesgue measure on $[0, 1]$ onto $\mathcal{T}_{2e}$.

Consider a uniform random tree T_n. Put mass $1/n$ at each vertex. Call the resulting probability measure μ_n. It is not hard to show that

$$\left((T_n, d_{\mathrm{gr}}/\sqrt{n}, \mu_n\right) \xrightarrow{d} (\mathcal{T}_{2e}, d_{2e}, \mu_{2e})$$

as $n \to \infty$.

The mass measure of the Brownian CRT

Let μ_{2e} be the push-forward of Lebesgue measure on $[0,1]$ onto $\mathcal{T}_{2e}$.

Consider a uniform random tree T_n. Put mass $1/n$ at each vertex. Call the resulting probability measure μ_n. It is not hard to show that

$$\left((T_n, d_{\mathrm{gr}}/\sqrt{n}, \mu_n\right) \xrightarrow{d} (\mathcal{T}_{2e}, d_{2e}, \mu_{2e})$$

as $n \to \infty$.

Lemma. $\mu_{2e}(\mathcal{L}(\mathcal{T}_{2e})) = 1$.

[Intuition: non-leaf vertices of T_n are typically at distance $o(\sqrt{n})$ from a leaf. Proof: see Aldous (1991).]

The root of the Brownian CRT

Moreover, since the law of T_n is invariant under uniform random re-rooting (i.e. choosing a new root according to μ_n), the same must be true for $\mathcal{T}_{2e}$ if we re-root according to a sample from μ_{2e}.

The branch-points of the Brownian CRT

The branch-points of $\mathcal{T}_{2e}$ correspond to the local minima of the Brownian excursion e. With probability 1, there are no repeated local minima, which tells us that the branch-points all have degree 3 i.e. the tree is binary.

(Note that T_n is not binary. The fact that $\mathcal{T}_{2e}$ is tells us that there cannot be more than two children of a vertex in T_n whose family trees grow to $\sqrt{n}$ height.)

Characterising a CRT via sampling

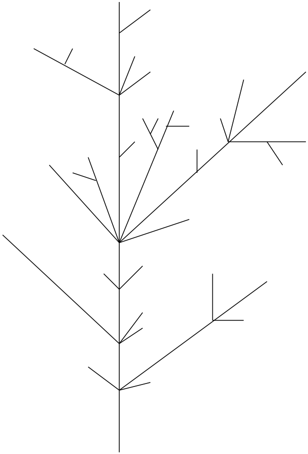
Take a CRT $(\mathcal{T}, d, \mu)$ and suppose that $U_1, U_2, \ldots$ are i.i.d. samples from the measure μ.

Characterising a CRT via sampling

Take a CRT $(\mathcal{T}, d, \mu)$ and suppose that $U_1, U_2, \ldots$ are i.i.d. samples from the measure μ. (Note: these are a.s. leaves.)

Characterising a CRT via sampling

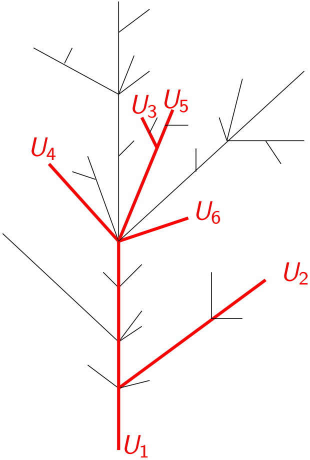
Take a CRT $(\mathcal{T}, d, \mu)$ and suppose that $U_1, U_2, \ldots$ are i.i.d. samples from the measure μ. (Note: these are a.s. leaves.) For $m \geq 2$, let $\mathcal{R}(m)$ be the subtree of $\mathcal{T}$ spanned by $U_1, U_2, \ldots, U_m$.



Characterising a CRT via sampling

Take a CRT $(\mathcal{T}, d, \mu)$ and suppose that $U_1, U_2, \ldots$ are i.i.d. samples from the measure μ. (Note: these are a.s. leaves.) For $m \geq 2$, let $\mathcal{R}(m)$ be the subtree of $\mathcal{T}$ spanned by $U_1, U_2, \ldots, U_m$.

$\mathcal{R}(6)$



Characterising a CRT via sampling

For every $m \geq 2$, $\mathcal{R}(m)$ can be regarded as a discrete tree with edge-lengths and labelled leaves, and so its distribution is specified by its tree-shape, $\mathbf{t}$, an unrooted unordered tree with m labelled leaves, and its edge-lengths. The reduced trees are clearly consistent, in that $\mathcal{R}(m)$ is a subtree of $\mathcal{R}(m+1)$.

Characterising a CRT via sampling

For every $m \geq 2$, $\mathcal{R}(m)$ can be regarded as a discrete tree with edge-lengths and labelled leaves, and so its distribution is specified by its tree-shape, $\mathbf{t}$, an unrooted unordered tree with m labelled leaves, and its edge-lengths. The reduced trees are clearly consistent, in that $\mathcal{R}(m)$ is a subtree of $\mathcal{R}(m+1)$.

Theorem. (Aldous (1993)) The law of $(\mathcal{T}, d, \mu)$ is specified by its random finite-dimensional distributions, that is the laws of $(\mathcal{R}(m), m \geq 2)$.

The random fdds of the Brownian CRT

Observe that $\mathcal{R}(m)$ must be binary since $\mathcal{T}_{2e}$ is. So the tree-shape of $\mathcal{R}(m)$ has $2m-2$ vertices and $2m-3$ edges.

Let $\mathbf{t}$ be this tree-shape and let $x_1, x_2, \ldots, x_{2m-3}$ be the edge-lengths listed in any (arbitrary but fixed) order.

The random fdds of the Brownian CRT

Observe that $\mathcal{R}(m)$ must be binary since $\mathcal{T}_{2e}$ is. So the tree-shape of $\mathcal{R}(m)$ has $2m-2$ vertices and $2m-3$ edges.

Let $\mathbf{t}$ be this tree-shape and let $x_1, x_2, \ldots, x_{2m-3}$ be the edge-lengths listed in any (arbitrary but fixed) order.

Theorem. (Aldous (1993)) $\mathcal{R}(m)$ has density

$$f(\mathbf{t}; x_1, x_2, \ldots, x_{2m-3}) = \left(\sum_{i=1}^{2m-3} x_i\right) \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{2m-3} x_i\right)^2\right).$$

[See Le Gall (2005) for a direct proof from the Brownian excursion.]

The random fdds of the Brownian CRT

Observe that $\mathcal{R}(m)$ must be binary since $\mathcal{T}_{2e}$ is. So the tree-shape of $\mathcal{R}(m)$ has $2m-2$ vertices and $2m-3$ edges.

Let $\mathbf{t}$ be this tree-shape and let $x_1, x_2, \ldots, x_{2m-3}$ be the edge-lengths listed in any (arbitrary but fixed) order.

Theorem. (Aldous (1993)) $\mathcal{R}(m)$ has density

$$f(\mathbf{t}; x_1, x_2, \ldots, x_{2m-3}) = \left(\sum_{i=1}^{2m-3} x_i\right) \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{2m-3} x_i\right)^2\right).$$

[See Le Gall (2005) for a direct proof from the Brownian excursion.]

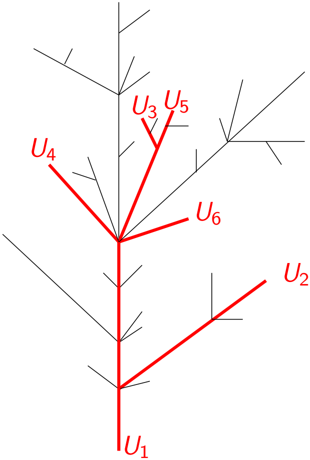
This implies that $\mathbf{t}$ is uniform on the set of binary unordered trees with m labelled leaves and that the edge-lengths are exchangeable.

Lecture 3

Recap: characterising the Brownian CRT by sampling

Take a Brownian CRT $(\mathcal{T}, d, \mu)$ and suppose that $U_1, U_2, \ldots$ are i.i.d. samples from the measure μ. (Note: these are a.s. leaves.) For $m \geq 2$, let $\mathcal{R}(m)$ be the subtree of $\mathcal{T}$ spanned by $U_1, U_2, \ldots, U_m$.

$\mathcal{R}(6)$



Recap: the random fdds of the Brownian CRT

Observe that $\mathcal{R}(m)$ must be binary since $\mathcal{T}_{2e}$ is. So the tree-shape of $\mathcal{R}(m)$ has $2m-2$ vertices and $2m-3$ edges.

Let $\mathbf{t}$ be this tree-shape and let $x_1, x_2, \ldots, x_{2m-3}$ be the edge-lengths listed in any (arbitrary but fixed) order.

Theorem. (Aldous (1993)) $\mathcal{R}(m)$ has density

$$f(\mathbf{t}; x_1, x_2, \ldots, x_{2m-3}) = \left(\sum_{i=1}^{2m-3} x_i\right) \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{2m-3} x_i\right)^2\right).$$

[See Le Gall (2005) for a direct proof from the Brownian excursion.]

This implies that $\mathbf{t}$ is uniform on the set of binary unordered trees with m labelled leaves and that the edge-lengths are exchangeable.

The Dirichlet distribution

Write

$$\mathcal{S}_n = \left\{ (s_1, s_2, \dots, s_n) \in \mathbb{R}_+^n : \sum_{i=1}^{n} s_i = 1 \right\}.$$

Definition. The Dirichlet distribution with parameters $a_1, a_2, \dots, a_n > 0$ (written $\mathrm{Dir}(a_1, a_2, \dots, a_n)$) has density

$$\frac{\Gamma(a_1 + a_2 + \cdots + a_n)}{\Gamma(a_1) \cdots \Gamma(a_n)} x_1^{a_1 - 1} \dots x_n^{a_n - 1}$$

with respect to ($(n-1)$-dimensional) Lebesgue measure on $\mathcal{S}_n$.

The Dirichlet distribution

Write

$$\mathcal{S}_n = \left\{ (s_1, s_2, \ldots, s_n) \in \mathbb{R}_+^n : \sum_{i=1}^{n} s_i = 1 \right\}.$$

Definition. The Dirichlet distribution with parameters $a_1, a_2, \ldots, a_n > 0$ (written $\mathrm{Dir}(a_1, a_2, \ldots, a_n)$) has density

$$\frac{\Gamma(a_1 + a_2 + \cdots + a_n)}{\Gamma(a_1) \cdots \Gamma(a_n)} x_1^{a_1 - 1} \ldots x_n^{a_n - 1}$$

with respect to ($(n-1)$-dimensional) Lebesgue measure on $\mathcal{S}_n$.

Note: If $B \sim \mathrm{Beta}(a_1, a_2)$ then $(B, 1-B) \sim \mathrm{Dir}(a_1, a_2)$.

The Dirichlet distribution

Write

$$\mathcal{S}_n = \left\{ (s_1, s_2, \ldots, s_n) \in \mathbb{R}_+^n : \sum_{i=1}^n s_i = 1 \right\}.$$

Definition. The Dirichlet distribution with parameters $a_1, a_2, \ldots, a_n > 0$ (written $\mathrm{Dir}(a_1, a_2, \ldots, a_n)$) has density

$$\frac{\Gamma(a_1 + a_2 + \cdots + a_n)}{\Gamma(a_1) \cdots \Gamma(a_n)} x_1^{a_1 - 1} \ldots x_n^{a_n - 1}$$

with respect to ($(n-1)$-dimensional) Lebesgue measure on $\mathcal{S}_n$.

Note: If $B \sim \mathrm{Beta}(a_1, a_2)$ then $(B, 1 - B) \sim \mathrm{Dir}(a_1, a_2)$.

$\mathrm{Dir}(1, 1, \ldots, 1)$ is the uniform distribution on the simplex $\mathcal{S}_n$, and is the law of the lengths of the sub-intervals into which $[0, 1]$ is split by $n - 1$ independent $\mathrm{U}(0, 1)$ random variables.

Dirichlet distribution facts (size-biased sampling)

Proposition. Let $\mathbf{D} = (D_1, D_2, \ldots, D_n) \sim \mathrm{Dir}(a_1, a_2, \ldots, a_n)$ and

$$\mathbb{P}\left(I = i | \mathbf{D}\right) = D_i$$

(i.e. sample a size-biased co-ordinate). Then, conditionally on the event $\{I = i\}$, we have

$$(D_1, \ldots, D_i, \ldots, D_n) \sim \mathrm{Dir}(a_1, \ldots, a_i + 1, \ldots, a_n).$$

Dirichlet distribution facts (beta-gamma algebra)

Proposition. If $\mathbf{D} \sim \mathrm{Dir}(a_1, a_2, \ldots, a_n)$ and $G \sim \mathrm{Gamma}(\sum_{i=1}^n a_i, 1)$ are independent then

$$G \times (D_1, D_2, \ldots, D_n) \stackrel{d}{=} (G_1, G_2, \ldots, G_n),$$

where
$G_1 \sim \mathrm{Gamma}(a_1, 1), G_2 \sim \mathrm{Gamma}(a_2, 1), \ldots, G_n \sim \mathrm{Gamma}(a_n, 1)$
are independent.

Moreover,

$$\left(\frac{G_1}{\sum_{i=1}^n G_i}, \frac{G_2}{\sum_{i=1}^n G_i}, \ldots, \frac{G_n}{\sum_{i=1}^n G_i}\right) \stackrel{d}{=} (D_1, D_2, \ldots, D_n)$$

and is independent of $\sum_{i=1}^n G_i \sim \mathrm{Gamma}(\sum_{i=1}^n a_i, 1)$.

Dirichlet distribution facts (beta-gamma algebra)

A consequence that will be useful for us in a moment:

Proposition. If $B \sim \mathrm{Beta}(k, 1)$ and $(D_1, \ldots, D_k) \sim \mathrm{Dir}(\underbrace{1, 1, \ldots, 1}_{k})$ are independent then

$$(BD_1, \ldots, BD_k, 1 - B) \sim \mathrm{Dir}(\underbrace{1, 1, \ldots, 1}_{k+1}).$$

The random fdds of the Brownian CRT

Recall that the edge-lengths of $\mathcal{R}(m)$ have joint density

$$f(x_1, x_2, \ldots, x_{2m-3}) \propto \left(\sum_{i=1}^{2m-3} x_i\right) \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{2m-3} x_i\right)^2\right). \quad (\star)$$

The random fdds of the Brownian CRT

Recall that the edge-lengths of $\mathcal{R}(m)$ have joint density

$$f(x_1, x_2, \ldots, x_{2m-3}) \propto \left(\sum_{i=1}^{2m-3} x_i\right) \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{2m-3} x_i\right)^2\right). \quad (\star)$$

If $m = 2$, we get density $x \exp(-x^2/2)$ for the length of the single branch. This is the density of $\sqrt{2 \times \mathrm{Exp}(1)}$ i.e. same as the density of the first length in the line-breaking construction...

The random fdds of the Brownian CRT

Recall that the edge-lengths of $\mathcal{R}(m)$ have joint density

$$f(x_1, x_2, \ldots, x_{2m-3}) \propto \left(\sum_{i=1}^{2m-3} x_i\right) \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{2m-3} x_i\right)^2\right). \quad (\star)$$

If $m = 2$, we get density $x \exp(-x^2/2)$ for the length of the single branch. This is the density of $\sqrt{2 \times \mathrm{Exp}(1)}$ i.e. same as the density of the first length in the line-breaking construction...

Proposition. The line-breaking construction realises the random fdds of the Brownian CRT.

The random fdds of the Brownian CRT

Proof. For $m \geq 2$, a simple change-of-variables argument shows that $(\star)$ is the same as the density of

$$\sqrt{2 \sum_{i=1}^{m-1} E_i} \times (D_1, D_2, \ldots, D_{2m-3}),$$

where the factors are independent and

$$E_1, E_2, \ldots, E_{m-1} \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$$

and

$$(D_1, D_2, \ldots, D_{2m-3}) \sim \mathrm{Dir}(1, 1, \ldots, 1).$$

Line-breaking revisited

Recall the line-breaking construction:

Take $E_1, E_2, \ldots$ to be i.i.d. Exp(1) and set $C_k = \sqrt{2 \sum_{i=1}^{k} E_k}$.

Consider the line-segments $[0, C_1)$, $[C_1, C_2), \ldots$.

Line-breaking revisited

Recall the line-breaking construction:

Take $E_1, E_2, \ldots$ to be i.i.d. Exp(1) and set $C_k = \sqrt{2 \sum_{i=1}^{k} E_k}$.

Consider the line-segments $[0, C_1)$, $[C_1, C_2), \ldots$.

Start from $[0, C_1)$ and proceed inductively.

Line-breaking revisited

Recall the line-breaking construction:

Take $E_1, E_2, \ldots$ to be i.i.d. Exp(1) and set $C_k = \sqrt{2 \sum_{i=1}^{k} E_k}$.

Consider the line-segments $[0, C_1)$, $[C_1, C_2), \ldots$.

Start from $[0, C_1)$ and proceed inductively. For $i \geq 1$, sample B_i uniformly from $[0, C_i)$ and attach $[C_i, C_{i+1})$ at the corresponding point of the tree constructed so far (this is a point chosen uniformly at random over the existing tree).

Line-breaking revisited

The points $B_1, C_1, B_2, C_2, \ldots, B_{m-2}, C_{m-2}$ split the interval $[0, C_{m-1})$ into $2m - 3$ sub-intervals.

Line-breaking revisited

The points $B_1, C_1, B_2, C_2, \ldots, B_{m-2}, C_{m-2}$ split the interval $[0, C_{m-1})$ into $2m-3$ sub-intervals.

$C_{m-1} \overset{d}{=} \sqrt{2 \sum_{i=1}^{m-1} E_i}$.

Line-breaking revisited

The points $B_1, C_1, B_2, C_2, \dots, B_{m-2}, C_{m-2}$ split the interval $[0, C_{m-1})$ into $2m-3$ sub-intervals.

$C_{m-1} \stackrel{d}{=} \sqrt{2\sum_{i=1}^{m-1} E_i}$.

Claim. The sub-intervals into which the values

$$\frac{B_1}{C_{m-1}}, \frac{C_1}{C_{m-1}}, \dots, \frac{B_{m-2}}{C_{m-1}}, \frac{C_{m-2}}{C_{m-1}}$$

(put in increasing order) split $[0,1)$ have $\mathrm{Dir}(1,1,\dots,1)$ distribution, independently of C_{m-1}.

Line-breaking revisited

The points $B_1, C_1, B_2, C_2, \ldots, B_{m-2}, C_{m-2}$ split the interval $[0, C_{m-1})$ into $2m-3$ sub-intervals.

$C_{m-1} \stackrel{d}{=} \sqrt{2\sum_{i=1}^{m-1} E_i}$.

Claim. The sub-intervals into which the values

$$\frac{B_1}{C_{m-1}}, \frac{C_1}{C_{m-1}}, \ldots, \frac{B_{m-2}}{C_{m-1}}, \frac{C_{m-2}}{C_{m-1}}$$

(put in increasing order) split $[0, 1)$ have $\mathrm{Dir}(1, 1, \ldots, 1)$ distribution, independently of C_{m-1}.

This proves the proposition.

Line-breaking revisited

The points $B_1, C_1, B_2, C_2, \ldots, B_{m-2}, C_{m-2}$ split the interval $[0, C_{m-1})$ into $2m-3$ sub-intervals.

$C_{m-1} \stackrel{d}{=} \sqrt{2 \sum_{i=1}^{m-1} E_i}$.

Claim. The sub-intervals into which the values

$$\frac{B_1}{C_{m-1}}, \frac{C_1}{C_{m-1}}, \ldots, \frac{B_{m-2}}{C_{m-1}}, \frac{C_{m-2}}{C_{m-1}}$$

(put in increasing order) split $[0,1)$ have $\mathrm{Dir}(1,1,\ldots,1)$ distribution, independently of C_{m-1}. ▶

This proves the proposition.

Indeed, we can recover a Brownian CRT by taking the metric space completion of the object constructed by line-breaking.

Line-breaking revisited

The points $B_1, C_1, B_2, C_2, \ldots, B_{m-2}, C_{m-2}$ split the interval $[0, C_{m-1})$ into $2m-3$ sub-intervals.

$C_{m-1} \stackrel{d}{=} \sqrt{2 \sum_{i=1}^{m-1} E_i}$.

Claim. The sub-intervals into which the values

$$\frac{B_1}{C_{m-1}}, \frac{C_1}{C_{m-1}}, \ldots, \frac{B_{m-2}}{C_{m-1}}, \frac{C_{m-2}}{C_{m-1}}$$

(put in increasing order) split $[0, 1)$ have $\mathrm{Dir}(1, 1, \ldots, 1)$ distribution, independently of C_{m-1}. ▶

This proves the proposition.

Indeed, we can recover a Brownian CRT by taking the metric space completion of the object constructed by line-breaking. Note: completion can only add leaves.

Rémy's algorithm

Consider the tree shapes in the line-breaking construction: at step $m-1$ we have an unordered tree with m labelled leaves. We have seen that it is uniform on the set of binary trees with m labelled leaves, for $m \geq 2$.

Rémy's algorithm

Consider the tree shapes in the line-breaking construction: at step $m-1$ we have an unordered tree with m labelled leaves. We have seen that it is uniform on the set of binary trees with m labelled leaves, for $m \geq 2$.

Implicit in the line-breaking construction, then, is an algorithm (originally due to Rémy (1985)) for generating these trees:

- Start from an edge with end-points labelled 1 and 2.
- For $m \geq 3$, pick an edge from the existing tree uniformly at random, subdivide it into two edges and attach another edge to the new vertex, with label m at its other end.

Rémy's algorithm

If T_n is the nth tree in Rémy's algorithm, and μ_n is the uniform distribution on the leaves, then it's not hard to show that

$$\left(T_n, \frac{1}{\sqrt{2n}} d_{\mathrm{gr}}, \mu_n \right) \xrightarrow{d} (\mathcal{T}_{2e}, d_{2e}, \mu_{2e}).$$

(In fact, this time the convergence is almost sure.)

Self-similarity

Consider picking three independent points U_1, U_2, U_3 from $\mathcal{T}_{2e}$ according to μ_{2e}. There is a unique branch-point between these three points, and it splits the tree into three subtrees, $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$.

Self-similarity

Consider picking three independent points U_1, U_2, U_3 from $\mathcal{T}_{2e}$ according to μ_{2e}. There is a unique branch-point between these three points, and it splits the tree into three subtrees, $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$.

Write d_1, d_2, d_3 and μ_1, μ_2, μ_3 for the restrictions of d_{2e} and μ_{2e} to each of these subtrees respectively. Let
$\Delta_1 = \mu_{2e}(\mathcal{T}_1), \Delta_2 = \mu_{2e}(\mathcal{T}_2), \Delta_3 = \mu_{2e}(\mathcal{T}_3)$.

Self-similarity

Consider picking three independent points U_1, U_2, U_3 from $\mathcal{T}_{2e}$ according to μ_{2e}. There is a unique branch-point between these three points, and it splits the tree into three subtrees, $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$.

Write d_1, d_2, d_3 and μ_1, μ_2, μ_3 for the restrictions of d_{2e} and μ_{2e} to each of these subtrees respectively. Let
$\Delta_1 = \mu_{2e}(\mathcal{T}_1), \Delta_2 = \mu_{2e}(\mathcal{T}_2), \Delta_3 = \mu_{2e}(\mathcal{T}_3)$.

Theorem. (Aldous (1993))

- We have $(\Delta_1, \Delta_2, \Delta_3) \sim \mathrm{Dir}(1/2, 1/2, 1/2)$.
- The rescaled subtrees $(\mathcal{T}_1, d_1/\sqrt{\Delta_1}, \mu_1/\Delta_1)$, $(\mathcal{T}_2, d_2/\sqrt{\Delta_2}, \mu_2/\Delta_2)$, $(\mathcal{T}_3, d_3/\sqrt{\Delta_3}, \mu_3/\Delta_3)$ are i.i.d. Brownian CRTs, independent of $(\Delta_1, \Delta_2, \Delta_3)$.
- U_i and the original branch-point are independent samples from μ_i/Δ_i in subtree $i = 1, 2, 3$.

An operator on (laws of) CRTs

Let $\mathcal{M}$ be the set of probability measures on continuum trees. Define an operator $\mathfrak{F} : \mathcal{M} \to \mathcal{M}$ as follows: for $M \in \mathcal{M}$,

- Sample independent trees (T_1, d_1, μ_1), (T_2, d_2, μ_3), (T_3, d_3, μ_3) having distribution M;
- For $1 \leq i \leq 3$, sample U_i according to μ_i;
- Independently sample $(\Delta_1, \Delta_2, \Delta_3) \sim \mathrm{Dir}(1/2, 1/2, 1/2)$;
- Rescale to obtain $(T_1, \Delta_1^{1/2} d_1, \Delta_1 \mu_1)$, $(T_2, \Delta_2^{1/2} d_2, \Delta_2 \mu_2)$, $(T_3, \Delta_3^{1/2} d_3, \Delta_3 \mu_3)$.
- Identify the vertices U_1, U_2, U_3 in order to obtain a single larger tree T with a marked branch-point B; the metrics and measures naturally induce a metric d and a measure μ on T.
- Forget the branch-point in order to obtain (T, d, μ).

$\mathfrak{F}(M)$ is the distribution of (T, d, μ).

The Brownian CRT as a unique fixed point

The previous theorem told us that the law of the Brownian CRT is a fixed point of $\mathfrak{F}$.

The Brownian CRT as a unique fixed point

The previous theorem told us that the law of the Brownian CRT is a fixed point of $\mathfrak{F}$.

Theorem. (Albenque & G. (2015))

- [Unique fixed point] Suppose that M is a law on continuum trees which is a fixed point of $\mathfrak{F}$. Then there exists $\alpha > 0$ such that if $(T, d, \mu) \sim M$ then $(T, \alpha d, \mu)$ is a Brownian CRT.

The Brownian CRT as a unique fixed point

The previous theorem told us that the law of the Brownian CRT is a fixed point of $\mathfrak{F}$.

Theorem. (Albenque & G. (2015))

- [Unique fixed point] Suppose that M is a law on continuum trees which is a fixed point of $\mathfrak{F}$. Then there exists $\alpha > 0$ such that if $(T, d, \mu) \sim M$ then $(T, \alpha d, \mu)$ is a Brownian CRT.
- [Attractive] Suppose that M is a law on continuum trees such that if $(T, d, \mu) \sim M$, given (T, d, μ), V_1, V_2 are sampled independently from μ, then $\mathbb{E}\left[d(V_1, V_2)\right]$ exists and is equal to $\pi/2$. Let $M_n = \mathfrak{F}^n M$. Then M_n converges weakly to the law of the Brownian CRT in the sense of the Gromov-Prokhorov topology.

The Brownian CRT as a unique fixed point: sketch proof

Lemma. Suppose that $M \in \mathcal{M}$ is a fixed point of $\mathfrak{F}$. Let $(T, d, \mu) \sim M$ and, conditionally on (T, d, μ), let $V_1, V_2 \overset{\text{i.i.d.}}{\sim} \mu$. Then there exists a constant $\alpha > 0$ such that $\alpha d(V_1, V_2) \sim$ Rayleigh.

The Brownian CRT as a unique fixed point: sketch proof

Lemma. Suppose that $M \in \mathcal{M}$ is a fixed point of $\mathfrak{F}$. Let $(T, d, \mu) \sim M$ and, conditionally on (T, d, μ), let $V_1, V_2 \overset{\text{i.i.d.}}{\sim} \mu$. Then there exists a constant $\alpha > 0$ such that $\alpha d(V_1, V_2) \sim$ Rayleigh.

Lemma. Suppose that $M \in \mathcal{M}$ is a fixed point of $\mathfrak{F}$. Then the random finite dimensional distributions of M are the same as those of the Brownian CRT, up to a strictly positive scaling factor α.

The Brownian CRT as a unique fixed point: sketch proof

Lemma. Suppose that $M \in \mathcal{M}$ is a fixed point of $\mathfrak{F}$. Let $(T, d, \mu) \sim M$ and, conditionally on (T, d, μ), let $V_1, V_2 \overset{\text{i.i.d.}}{\sim} \mu$. Then there exists a constant $\alpha > 0$ such that $\alpha d(V_1, V_2) \sim$ Rayleigh. ▶

Lemma. Suppose that $M \in \mathcal{M}$ is a fixed point of $\mathfrak{F}$. Then the random finite dimensional distributions of M are the same as those of the Brownian CRT, up to a strictly positive scaling factor α. ▶

[See Marie Albenque & Christina Goldschmidt, **The Brownian continuum random tree as the unique solution to a fixed point equation**, *Electronic Communications in Probability* **20** (2015), paper no. 61, pp.1-14.]

A random fractal

The self-similarity of the Brownian CRT tells us, in particular, that it is a random fractal.

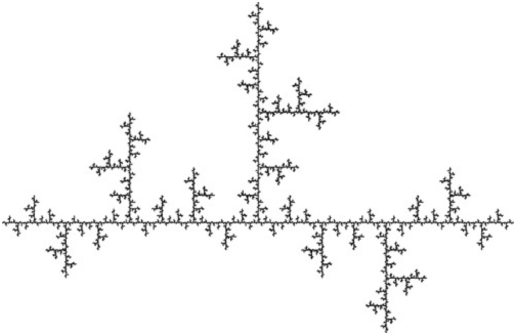
A random fractal

The self-similarity of the Brownian CRT tells us, in particular, that it is a random fractal.

Theorem. (Haas & Miermont (2004), Duquesne & Le Gall (2005)) The Brownian CRT has Hausdorff dimension 2, almost surely.

A random fractal

Croydon & Hambly (2008) showed that it is a familiar deterministic fractal endowed with a random metric.

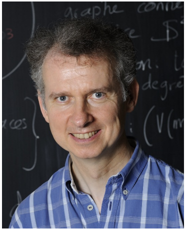


[Image from Croydon & Hambly (2008)]

5. THE STABLE TREES

Key reference:

Jean-François Le Gall, **Random trees and applications**, *Probability Surveys* **2** (2005) pp.245-311.



Infinite variance

Write T_n for a Galton-Watson tree with critical offspring distribution, conditioned to have total progeny n. We have so far focussed on the case where the offspring distribution also has finite variance. What if this is not true?

Infinite variance

Write T_n for a Galton-Watson tree with critical offspring distribution, conditioned to have total progeny n. We have so far focussed on the case where the offspring distribution also has finite variance. What if this is not true?

[Reference: Durrett, **Probability theory and examples** for a very beautiful presentation of the material in the next few slides.]

Generalised central limit theorems

Theorem. Let $Z_1, Z_2, \ldots$ be i.i.d. random variables such that, for some $\alpha \in (0, 2)$,

- $\lim_{t\to\infty} \mathbb{P}(Z_1 > t) / \mathbb{P}(|Z_1| > t) = \theta \in [0, 1]$;
- $\mathbb{P}(|Z_1| > t) = t^{-\alpha} L(t)$ where L is a slowly varying function.

Let $a_n = \inf\{t : \mathbb{P}(|Z_1| > t) \leq n^{-1}\}$ and $b_n = n\mathbb{E}\left[Z_1 \mathbb{1}_{\{|Z_1| \leq a_n\}}\right]$.

Generalised central limit theorems

Theorem. Let $Z_1, Z_2, \ldots$ be i.i.d. random variables such that, for some $\alpha \in (0, 2)$,

- $\lim_{t\to\infty} \mathbb{P}(Z_1 > t) / \mathbb{P}(|Z_1| > t) = \theta \in [0, 1]$;
- $\mathbb{P}(|Z_1| > t) = t^{-\alpha} L(t)$ where L is a slowly varying function.

Let $a_n = \inf\{t : \mathbb{P}(|Z_1| > t) \leq n^{-1}\}$ and $b_n = n\mathbb{E}\left[Z_1 \mathbb{1}_{\{|Z_1| \leq a_n\}}\right]$.

Then as $n \to \infty$,

$$\frac{1}{a_n}\left(\sum_{i=1}^{n} Z_i - b_n\right) \xrightarrow{d} S_\alpha,$$

where S_α is an α-stable random variable with

$$\mathbb{E}\left[e^{itS_\alpha}\right] = \exp(itc - b|t|^\alpha(1 + i(2\theta - 1)\mathrm{sgn}(t)w_\alpha(t))),$$

where b, c are constants and $w_\alpha(t) = \begin{cases} \tan(\pi\alpha/2) & \text{if } \alpha \neq 1 \\ (2/\pi)\log|t| & \text{if } \alpha = 1. \end{cases}$

Generalised central limit theorems

Recall: we are interested in Galton-Watson trees with critical offspring distribution $p(k), k \geq 0$. The corresponding depth-first walk has step distribution $\nu(k) = p(k+1), k \geq -1$. We want $Z_1 \sim \nu$.

Generalised central limit theorems

Recall: we are interested in Galton-Watson trees with critical offspring distribution $p(k), k \geq 0$. The corresponding depth-first walk has step distribution $\nu(k) = p(k+1), k \geq -1$. We want $Z_1 \sim \nu$.

It's natural then to consider $p(k) \sim k^{-1-\alpha}$ since then $\mathbb{P}(Z_1 > k) \sim k^{-\alpha}$. Since we need $\mathbb{E}[Z_1] = 0$, the mean must exist, which restricts us to the case $\alpha \in (1, 2]$ ($\alpha = 2$ gives finite variance.)

Generalised central limit theorems

Recall: we are interested in Galton-Watson trees with critical offspring distribution $p(k), k \geq 0$. The corresponding depth-first walk has step distribution $\nu(k) = p(k+1), k \geq -1$. We want $Z_1 \sim \nu$.

It's natural then to consider $p(k) \sim k^{-1-\alpha}$ since then $\mathbb{P}(Z_1 > k) \sim k^{-\alpha}$. Since we need $\mathbb{E}[Z_1] = 0$, the mean must exist, which restricts us to the case $\alpha \in (1, 2]$ ($\alpha = 2$ gives finite variance.)

- As $\mathbb{P}(|Z_1| > k) \sim k^{-\alpha}$, we have $a_n = \inf\{t : \mathbb{P}(|Z_1| > t) \leq n^{-1}\} \sim n^{1/\alpha}$.

Generalised central limit theorems

Recall: we are interested in Galton-Watson trees with critical offspring distribution $p(k), k \geq 0$. The corresponding depth-first walk has step distribution $\nu(k) = p(k+1), k \geq -1$. We want $Z_1 \sim \nu$.

It's natural then to consider $p(k) \sim k^{-1-\alpha}$ since then $\mathbb{P}(Z_1 > k) \sim k^{-\alpha}$. Since we need $\mathbb{E}[Z_1] = 0$, the mean must exist, which restricts us to the case $\alpha \in (1, 2]$ ($\alpha = 2$ gives finite variance.)

- As $\mathbb{P}(|Z_1| > k) \sim k^{-\alpha}$, we have $a_n = \inf\{t : \mathbb{P}(|Z_1| > t) \leq n^{-1}\} \sim n^{1/\alpha}$.
- As $\alpha \in (1, 2)$, we may replace b_n by $n\mathbb{E}[Z_1] = 0$.

Generalised central limit theorems

Recall: we are interested in Galton-Watson trees with critical offspring distribution $p(k), k \geq 0$. The corresponding depth-first walk has step distribution $\nu(k) = p(k+1), k \geq -1$. We want $Z_1 \sim \nu$.

It's natural then to consider $p(k) \sim k^{-1-\alpha}$ since then $\mathbb{P}(Z_1 > k) \sim k^{-\alpha}$. Since we need $\mathbb{E}[Z_1] = 0$, the mean must exist, which restricts us to the case $\alpha \in (1, 2]$ ($\alpha = 2$ gives finite variance.)

- As $\mathbb{P}(|Z_1| > k) \sim k^{-\alpha}$, we have $a_n = \inf\{t : \mathbb{P}(|Z_1| > t) \leq n^{-1}\} \sim n^{1/\alpha}$.
- As $\alpha \in (1, 2)$, we may replace b_n by $n\mathbb{E}[Z_1] = 0$.
- As $\mathbb{P}(Z_1 < t) = 0$ for $t \leq -1$, we have $\theta = \lim_{t\to\infty} \mathbb{P}(Z_1 > t) / \mathbb{P}(|Z_1| > t) = 1$.

Generalised central limit theorems

So, in our setting, we have

$$\frac{1}{n^{1/\alpha}} \sum_{i=1}^{n} Z_i \xrightarrow{d} S_\alpha,$$

where

$$\mathbb{E}\left[e^{itS_\alpha}\right] = \exp(-ib|t|^\alpha \mathrm{sgn}(t) \tan(\pi\alpha/2)).$$

Functional convergence

The corresponding functional convergence is as follows.

Theorem. Let $X(k) = \sum_{i=1}^{k} Z_k$. Then

$$\frac{1}{n^{1/\alpha}}(X(\lfloor nt \rfloor), t \geq 0) \xrightarrow{d} (S_\alpha(t), t \geq 0),$$

where S_α is a α-stable Lévy process with no negative jumps.

Functional convergence

The corresponding functional convergence is as follows.

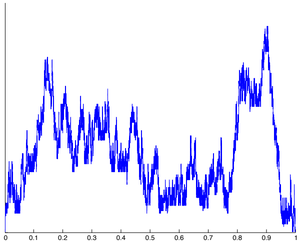
Theorem. Let $X(k) = \sum_{i=1}^{k} Z_k$. Then

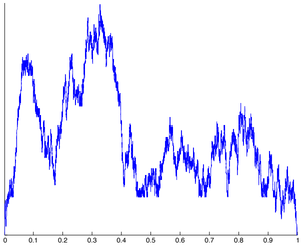
$$\frac{1}{n^{1/\alpha}}(X(\lfloor nt \rfloor), t \geq 0) \xrightarrow{d} (S_\alpha(t), t \geq 0),$$

where S_α is a α-stable Lévy process with no negative jumps.

There is also an excursion theory for these processes. However, the limiting relationship between X and the corresponding height process is no longer multiplication by a constant! In particular, an excursion of a stable Lévy process is far from a continuous function; the corresponding limiting height process, however, is continuous.

The limiting height process





[Pictures by Igor Kortchemski]

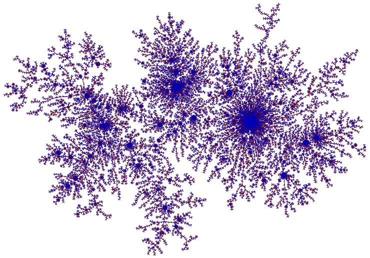
The stable trees

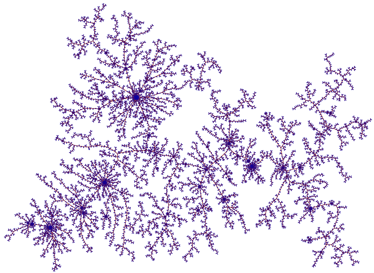
Theorem. (Duquesne & Le Gall (2002); Duquesne (2003))
Suppose that $p(k) \sim ck^{-1-\alpha}$ as $k \to \infty$ for $\alpha \in (1,2)$. Then as $n \to \infty$,

$$\frac{1}{n^{1-1/\alpha}} T_n \xrightarrow{d} c_\alpha \mathcal{T}_\alpha,$$

where $\mathcal{T}_\alpha$ is the stable tree of parameter α and c_α is a strictly positive constant. (The convergence is in the sense of the Gromov–Hausdorff distance.)

The stable trees





[Pictures by Igor Kortchemski]

Some futher reading

A natural question, in view of what we have seen for the Brownian CRT, is does there exist a line-breaking construction of the stable trees? This question is resolved in:

Christina Goldschmidt and Bénédicte Haas, **A line-breaking construction of the stable trees**,
Electronic Journal of Probability **20** (2015), paper no. 16, pp.1-24.

Thank you for listening!