VIII Escuela de Probabilidad y Procesos Estocásticos,
CIMAT, Guanajuato, 12-16 September 2016

Random trees and their scaling limits

Christina Goldschmidt

Department of Statistics and Lady Margaret Hall

UNIVERSITY OF OXFORD

Lecture 1

1. WEAK CONVERGENCE AND SCALING LIMITS

Key reference:

Rick Durrett, **Probability: theory and examples**, 4th edition, Cambridge University Press (2010).

Scaling limits

Suppose we have a sequence of random variables $R_1, R_2, \ldots$ 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 $(R_n, n \geq 1)$.

Scaling limits

Suppose we have a sequence of random variables $R_1, R_2, \ldots$ 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 $(R_n, n \geq 1)$.

Example (The 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 $X \sim N(0, 1)$.

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 $Z_1, Z_2, \ldots$ (as long as the distribution has finite variance).

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

Convergence in distribution

Throughout this minicourse we are going to want to deal with random objects which are not real-valued. Recall that the usual definition of convergence in distribution for a sequence $(X_n)_{n \geq 0}$ of random variables to X is

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) \text{ as } n \to \infty$$

for all x which are points of continuity of the function $x \mapsto \mathbb{P}(X \leq x), x \in \mathbb{R}$.

Problem: this doesn't generalise well to non-real-valued random variables!

Convergence in distribution

Suppose that we want to deal with random variables taking values in some arbitrary metric space (M, d).

Convergence in distribution

Suppose that we want to deal with random variables taking values in some arbitrary metric space (M, d).

Let $\mathcal{C}_b(M, \mathbb{R})$ be the set of bounded continuous functions $f : M \to \mathbb{R}$ (continuous in the sense that if $d(x_n, x) \to 0$ then $f(x_n) \to f(x)$).

Convergence in distribution

Suppose that we want to deal with random variables taking values in some arbitrary metric space (M, d).

Let $\mathcal{C}_b(M, \mathbb{R})$ be the set of bounded continuous functions $f : M \to \mathbb{R}$ (continuous in the sense that if $d(x_n, x) \to 0$ then $f(x_n) \to f(x)$).

Definition

Let $(X_n)_{n\geq 0}$ and X be random variables taking values in M. Then X_n converges in distribution (or converges weakly or converges in law) to X if

$$\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right] \quad \text{as } n \to \infty$$

for every $f \in \mathcal{C}_b(M, \mathbb{R})$.

Convergence in distribution

Exercise

Show that if $M = \mathbb{R}$ *this is equivalent to the usual definition.*

Convergence in distribution

Exercise

Show that if $M = \mathbb{R}$ this is equivalent to the usual definition.

So the convergence in distribution in the CLT also means that

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

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

Convergence in distribution

A very useful theorem:

Theorem (Skorokhod's representation theorem)

Suppose that $(X_n)_{n \geq 0}$ *and* X *are random variables taking values in a Polish space* (M, d)*, each a priori defined on a different probability space. Suppose that*

$$X_n \xrightarrow{d} X$$

as $n \to \infty$*. Then there exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and random variables* $(Y_n)_{n \geq 0}$ *and* Y *defined on it, such that* $X_n \stackrel{d}{=} Y_n$ *for each* $n \geq 0$*,* $Y \stackrel{d}{=} X$ *and*

$$Y_n \to Y \text{ almost surely.}$$

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.

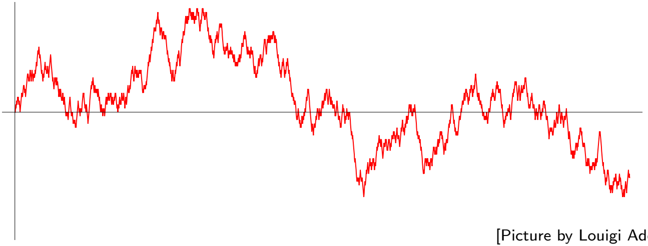
Another (related) scaling limit

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.

Theorem (Donsker's theorem)

Let $(W(t), t \geq 0)$ be a standard Brownian motion. Then as $n \to \infty$,

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



[Picture by Louigi Addario-Berry]

Convergence in distribution

Here, we are thinking of function-valued random variables, where the functions take values in $M = D(\mathbb{R}_+, \mathbb{R})$ and we can specify a metric on M as follows:

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

Convergence in distribution

Here, we are thinking of function-valued random variables, where the functions take values in $M = D(\mathbb{R}_+, \mathbb{R})$ and we can specify a metric on M as follows:

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

This metric encodes uniform convergence on compact time-intervals.

Convergence in distribution

Here, we are thinking of function-valued random variables, where the functions take values in $M = D(\mathbb{R}_+, \mathbb{R})$ and we can specify a metric on M as follows:

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

This metric encodes uniform convergence on compact time-intervals.

Then Donsker's theorem says that for all bounded continuous functions $f : D(\mathbb{R}_+, \mathbb{R}) \to \mathbb{R}$, we have

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

as $n \to \infty$.

2. THE UNIFORM RANDOM TREE

Key references:

David Aldous, **The continuum random tree I**,
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.

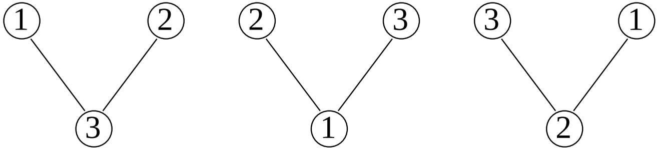
Labelled trees

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

Labelled trees

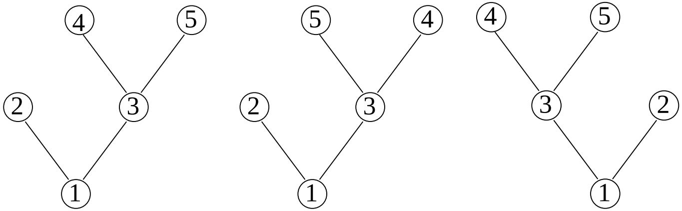
Let $\mathbb{T}_n$ be the set of unordered trees on n vertices labelled by $[n] := \{1, 2, \ldots, n\}$.

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

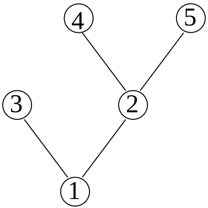


Labelled trees

Unordered means that these trees are all the same:



but this one is different:



Labelled trees

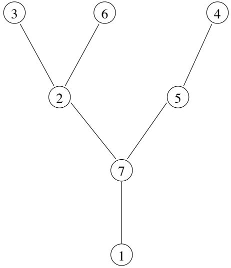
Theorem (Cayley's formula)

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

[Proof due to Jim Pitman, **Coalescent random forests**, *Journal of Combinatorial Theory Series A*, **85**(2) (1999), pp.165–193.]

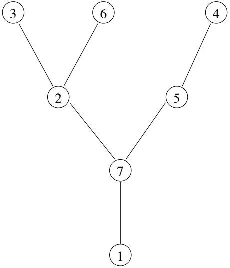
Uniform random trees

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



Uniform random trees

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



Question: what happens as n grows?

Uniform random trees, as $n \to \infty$

There are lots of statistics we might be interested in. For example:

- How many leaves (vertices with a single neighbour) are there?
- More generally, how many vertices of degree k are there (i.e. with exactly k neighbours), for $k \geq 1$?
- What is the diameter of the tree (i.e. the length of the longest path between two vertices in the tree)?
- What is the distance between two uniformly chosen vertices?
- ...

Uniform random trees

It turns out that the first question is not too hard to answer.

Exercise

Prove a limit in probability for the proportion of vertices which are leaves, as $n \to \infty$.

Uniform random trees

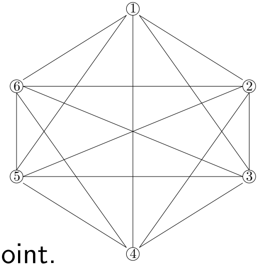
It turns out that the first question is not too hard to answer.

Exercise

Prove a limit in probability for the proportion of vertices which are leaves, as $n \to \infty$.

In order to think about some of the other questions, it useful to have an algorithm for building T_n.

The Aldous-Broder algorithm

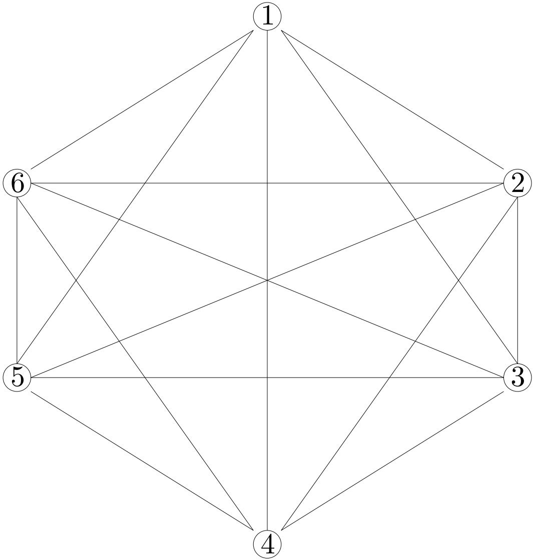


Take the complete graph on n vertices.

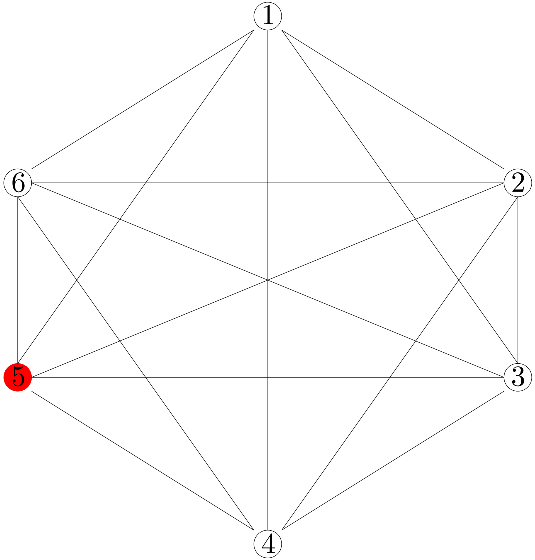
- Pick a uniform vertex to be the starting point.
- Run a simple random walk $(S_k)_{k \geq 0}$ on the graph (i.e. at each step, move to a neighbour chosen uniformly at random).
- Anytime the walk visits a new vertex, keep the edge along which it was reached.
- Stop when all vertices have been visited.

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

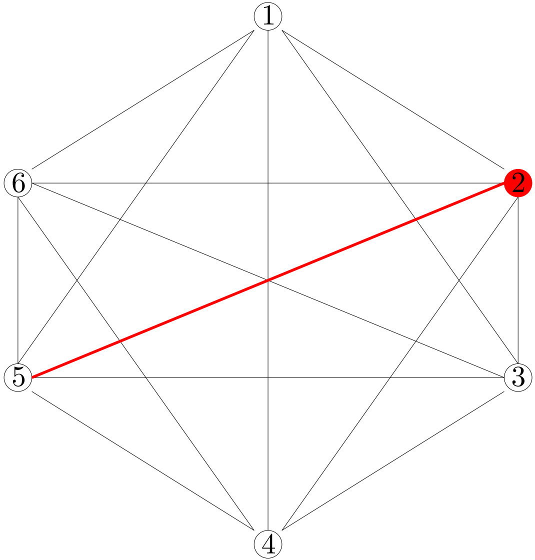
The Aldous-Broder algorithm



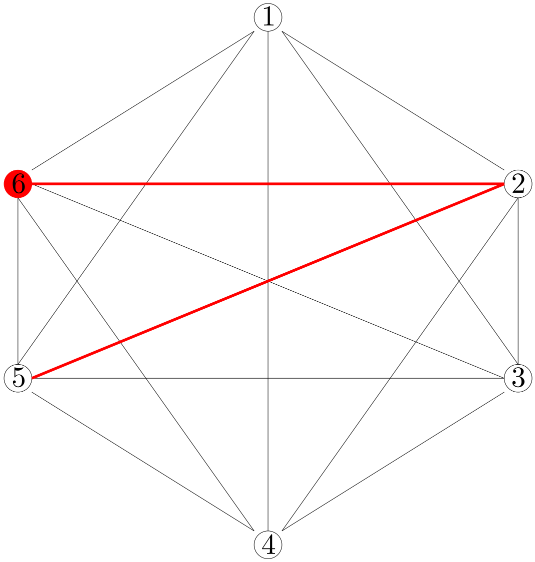
The Aldous-Broder algorithm



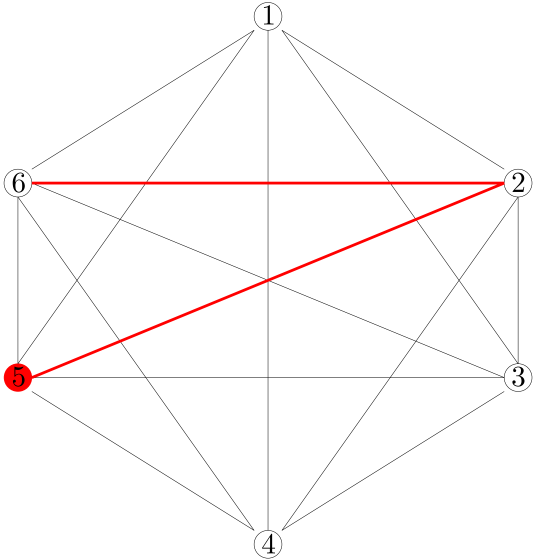
The Aldous-Broder algorithm



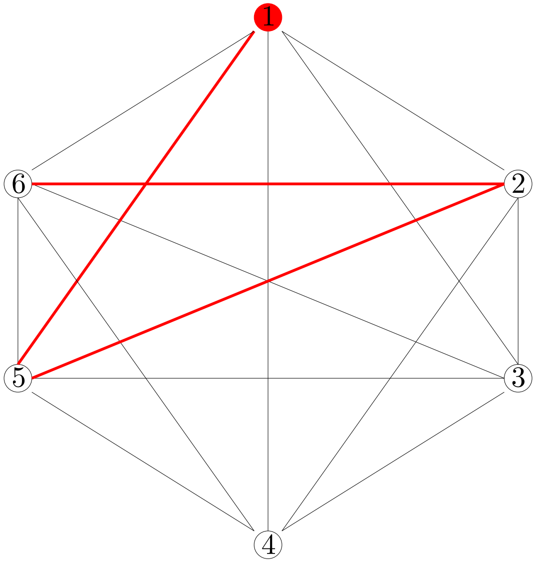
The Aldous-Broder algorithm



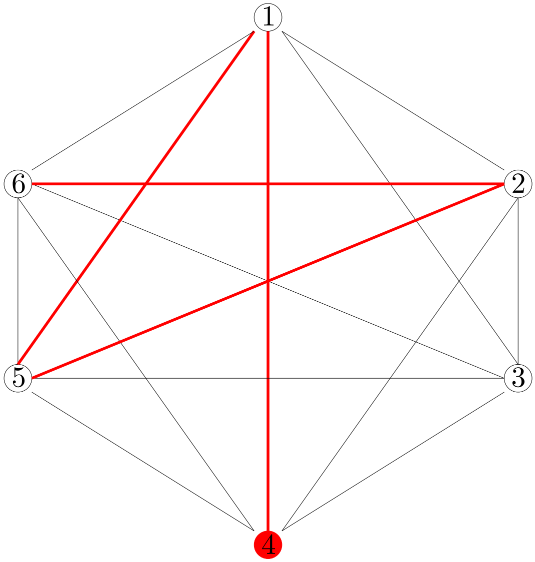
The Aldous-Broder algorithm



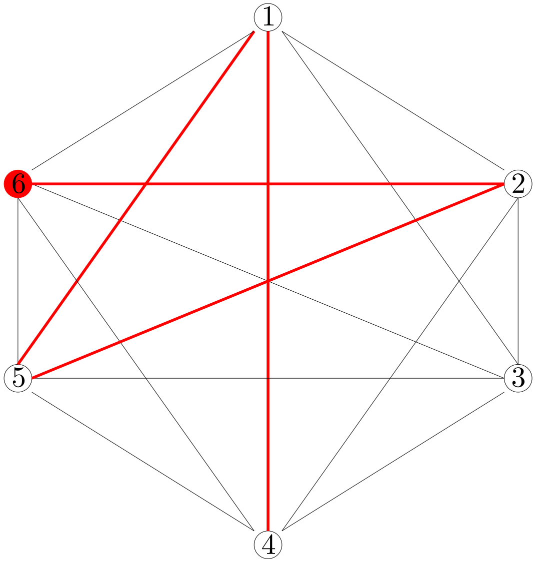
The Aldous-Broder algorithm



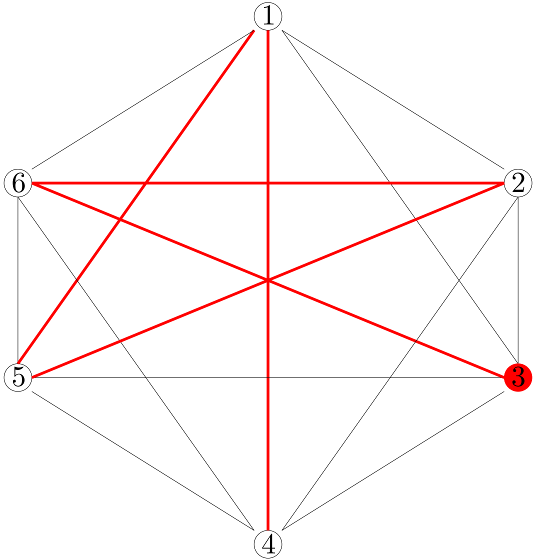
The Aldous-Broder algorithm



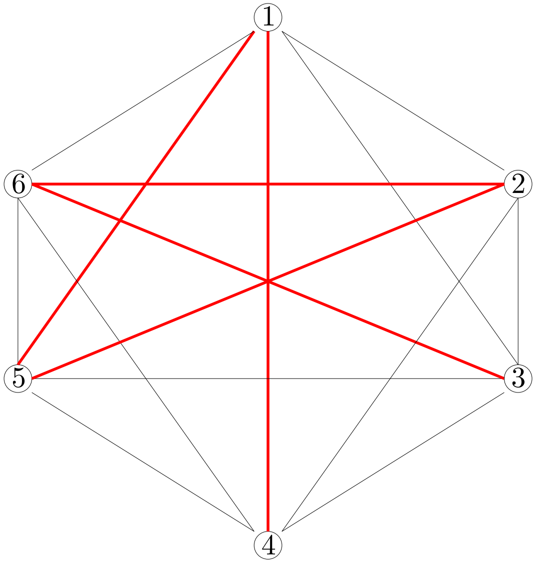
The Aldous-Broder algorithm



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.

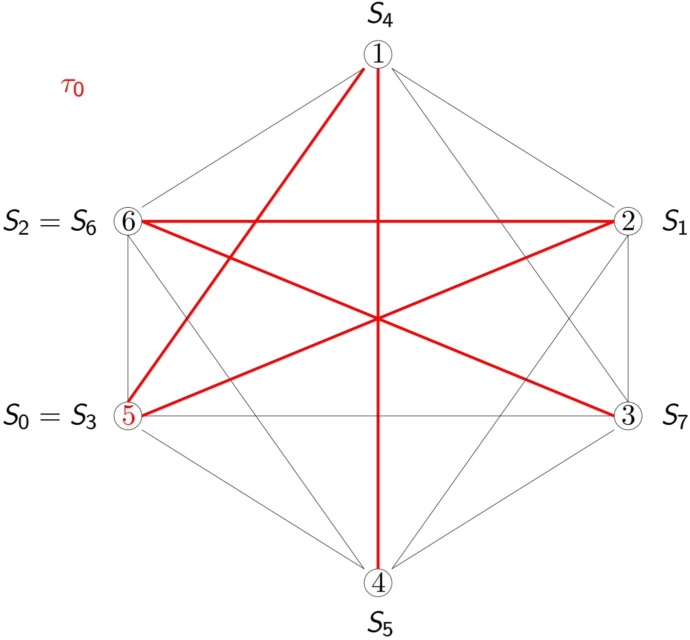
It remains to show that the stationary distribution π for $(\tau_k)_{k\in\mathbb{Z}}$ is uniform on $\mathbb{T}_n^\bullet$. It turns out to be easier to work with the time-reversed chain.

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

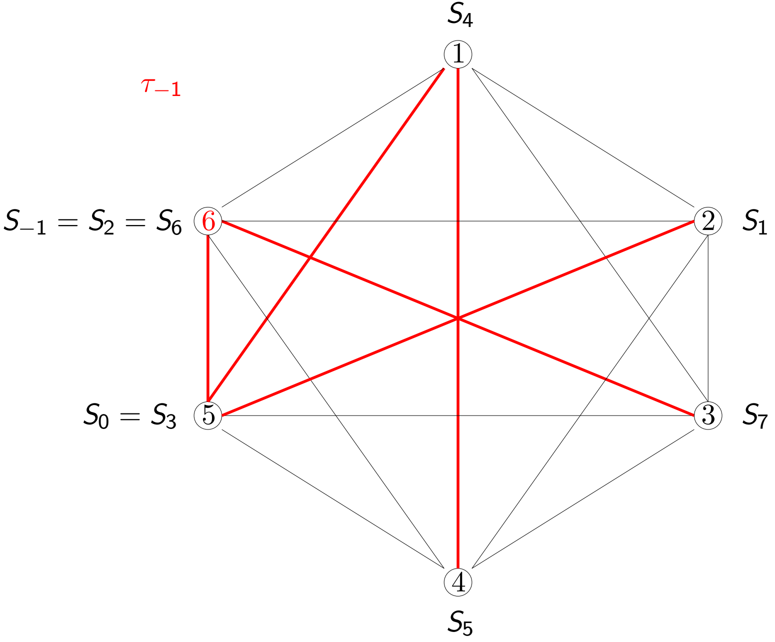
The Aldous-Broder algorithm: proof

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}.

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.

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

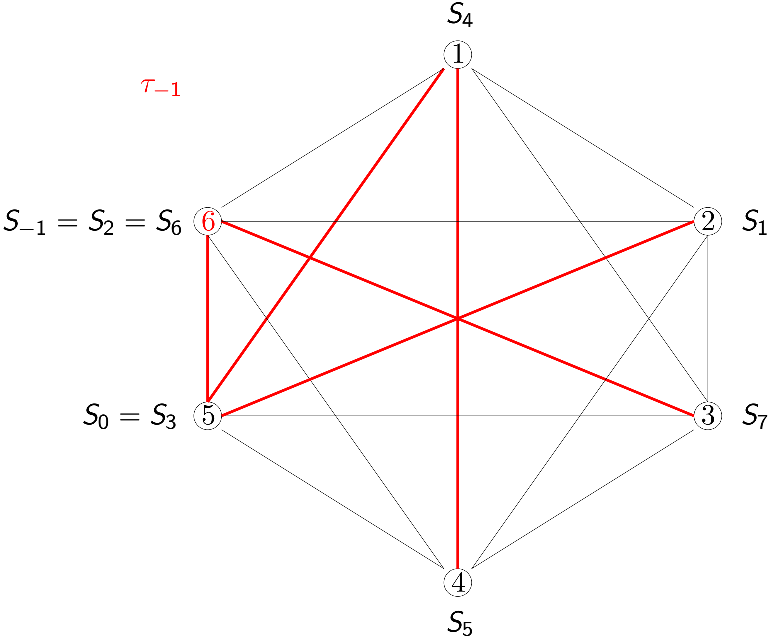
Given τ_{-1}, how many possibilities are there for τ_0?

The Aldous-Broder algorithm: proof

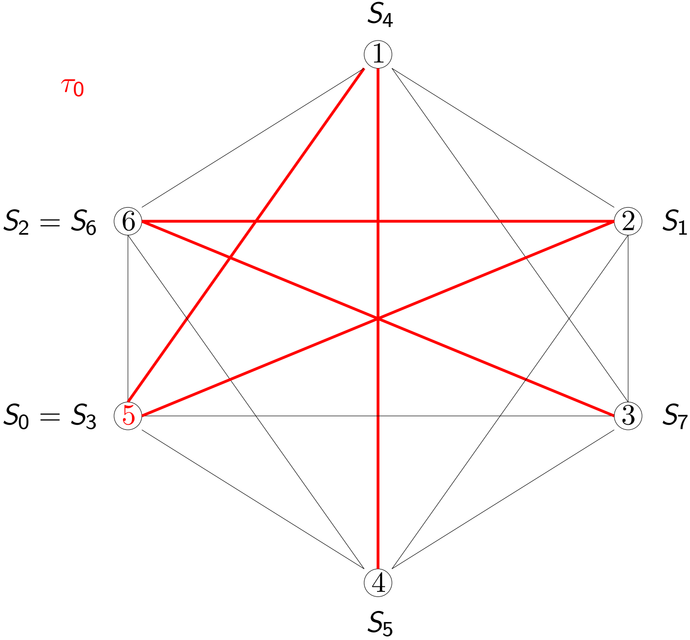
Given τ_{-1}, how many possibilities are there for τ_0?

S_0 must be one of the neighbours of S_{-1}. The possible values for τ_0 are generated by adding one of 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.

The Aldous-Broder algorithm: proof



The Aldous-Broder algorithm: proof



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 of 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 $1/(n-1)$.

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 of 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

Hence, the matrix $Q = (q(\tau, \tau'))_{\tau, \tau' \in \mathbb{T}_n^\bullet}$ is doubly stochastic (its rows and columns all sum to 1).

The Aldous-Broder algorithm: proof

Hence, the matrix $Q = (q(\tau, \tau'))_{\tau, \tau' \in \mathbb{T}_n^\bullet}$ is doubly stochastic (its rows and columns all sum to 1).

So it must have uniform stationary distribution on the set of rooted trees.

The Aldous-Broder algorithm: proof

Hence, the matrix $Q = (q(\tau, \tau'))_{\tau, \tau' \in \mathbb{T}_n^\bullet}$ is doubly stochastic (its rows and columns all sum to 1).

So it 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.

The Aldous-Broder algorithm: proof

Hence, the matrix $Q = (q(\tau, \tau'))_{\tau, \tau' \in \mathbb{T}_n^\bullet}$ is doubly stochastic (its rows and columns all sum to 1).

So it 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 algorithm, for trees with edge-weights.

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

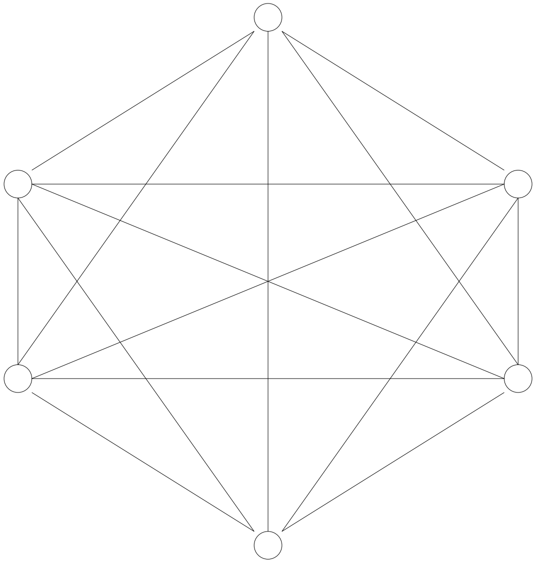
"Do the labelling as we go, then relabel at the end."

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

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

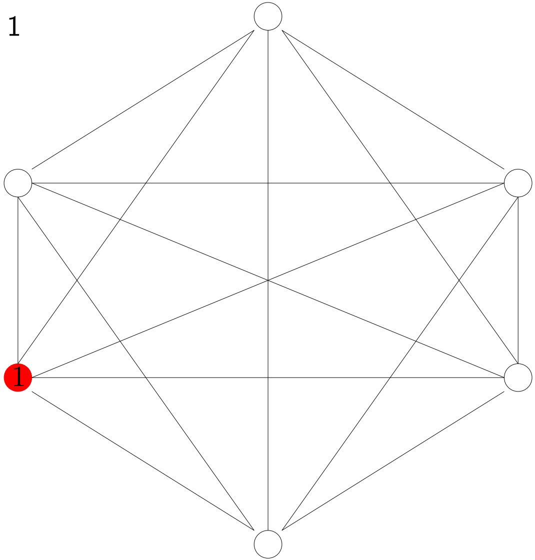
3. Take a uniform random permutation of the labels.

Aldous' algorithm



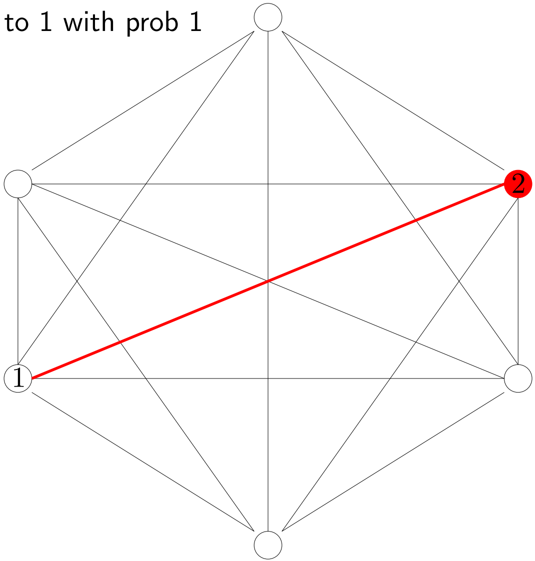
Aldous' algorithm

Start from 1



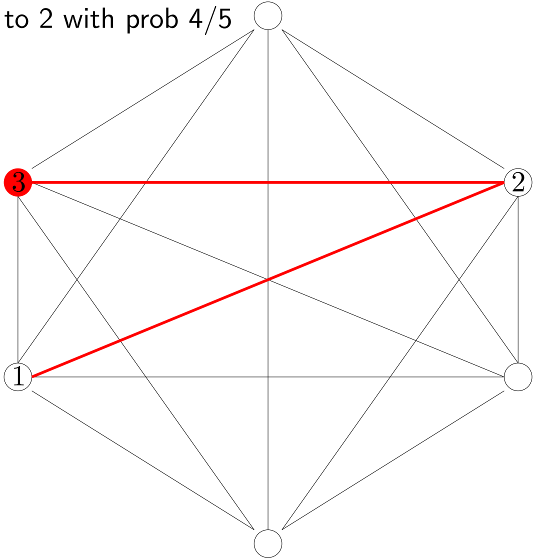
Aldous' algorithm

Connect 2 to 1 with prob 1



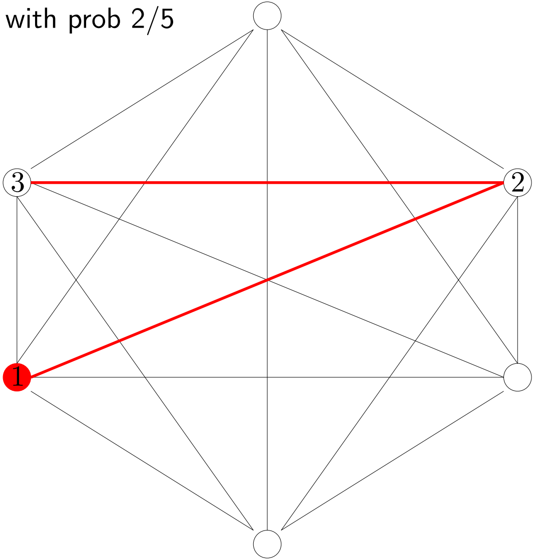
Aldous' algorithm

Connect 3 to 2 with prob 4/5



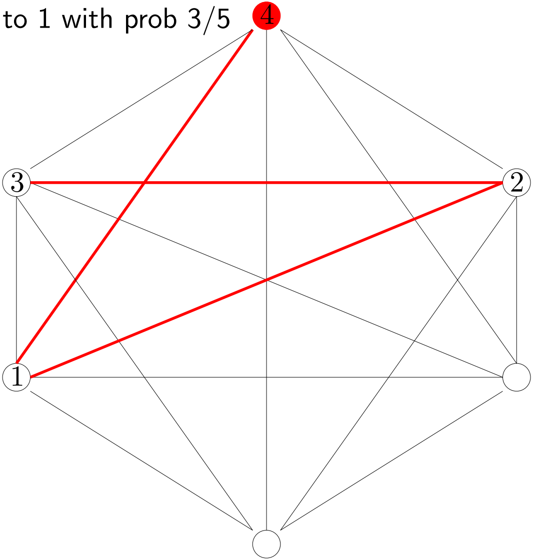
Aldous' algorithm

Jump to 1 with prob 2/5



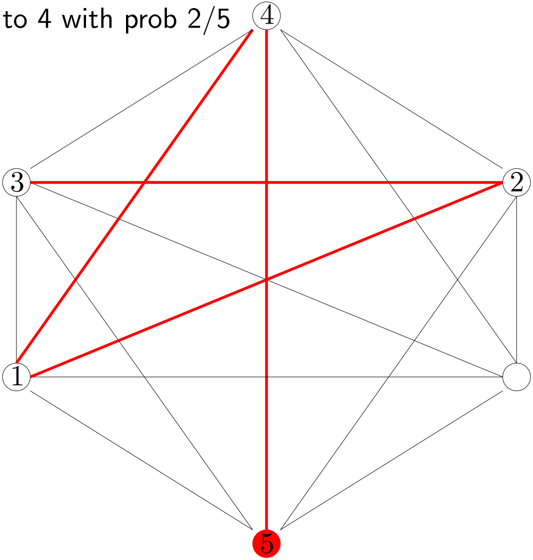
Aldous' algorithm

Connect 4 to 1 with prob 3/5



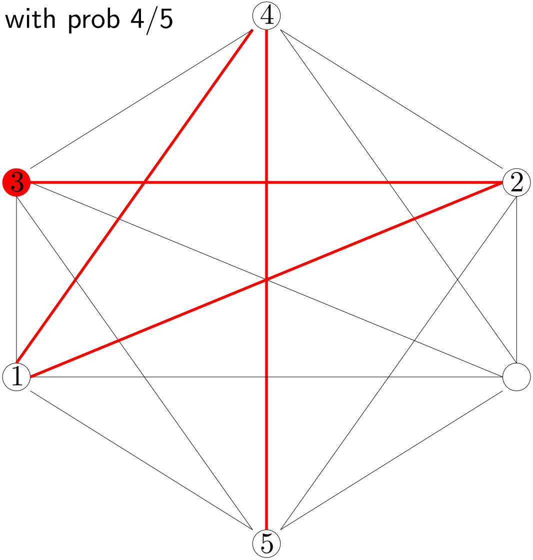
Aldous' algorithm

Connect 5 to 4 with prob 2/5



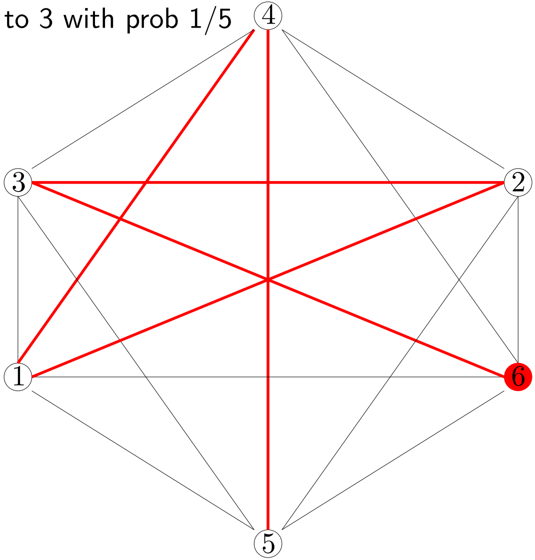
Aldous' algorithm

Jump to 3 with prob 4/5



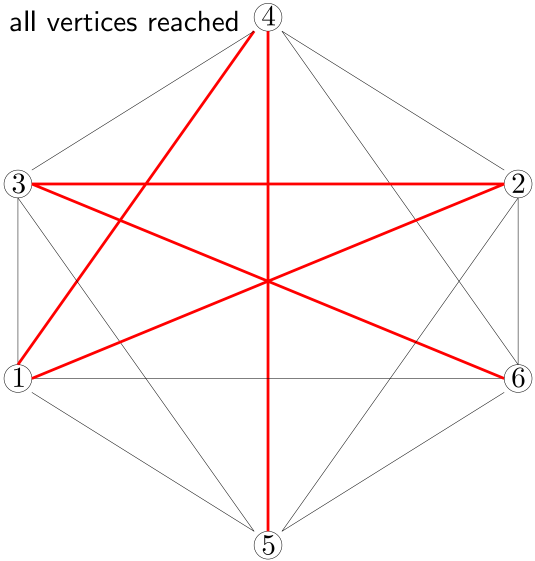
Aldous' algorithm

Connect 6 to 3 with prob 1/5



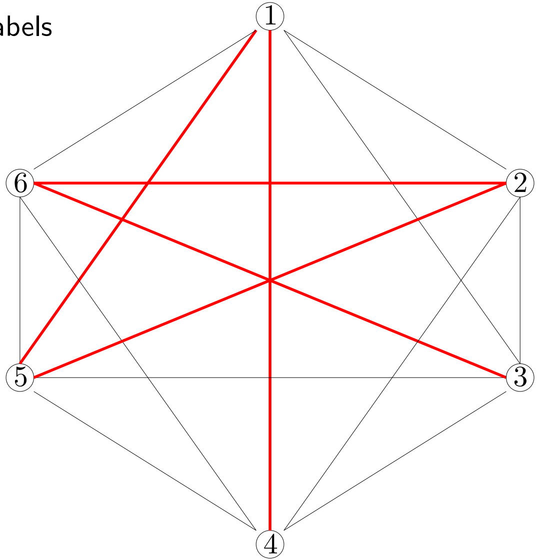
Aldous' algorithm

Stop when all vertices reached



Aldous' algorithm

Permute labels



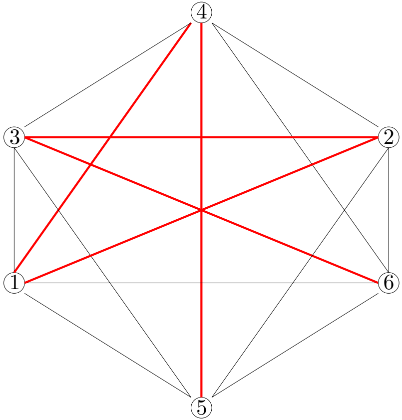
Typical distances

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, $C_1^6 = 4$:



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.

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.

Before we can describe the limiting version of the algorithm, we need a definition.

An inhomogeneous Poisson process

Definition

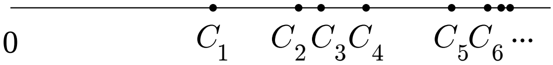
Let $\lambda : [0, \infty) \to [0, \infty)$ *be a continuous function such that* $\int_0^\infty \lambda(s)ds = \infty$ *but* $\int_0^t \lambda(s)ds < \infty$ *for all* $t \geq 0$*.*

We say that an increasing Markov process with càdlàg paths $(P(t), t \geq 0)$ *is an inhomogeneous Poisson process of intensity* λ *if* $P(0) = 0$ *and, given* $P(t) = n \in \mathbb{Z}_+$*, the rate of jumping to* $n + 1$ *is* $\lambda(t)$*.*

Equivalently, the number of points (= jump-times) falling in any interval $[s, t]$ *has a Poisson distribution with mean* $\int_s^t \lambda(r)dr$*, and the numbers of points falling in disjoint intervals are independent.*

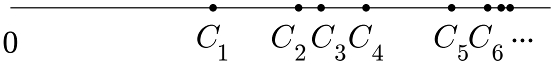
Line-breaking procedure

Let $C_1, C_2, \ldots$ be the points of an inhomogeneous Poisson process on $\mathbb{R}^+$ of intensity $\lambda(t) = t$.



Line-breaking procedure

Let $C_1, C_2, \ldots$ be the points of an inhomogeneous Poisson process on $\mathbb{R}^+$ of intensity $\lambda(t) = t$.



Note that

$$\begin{aligned}\mathbb{P}(C_1 > x) &= \mathbb{P}(\text{no points in [0,x]}) \\ &= \mathbb{P}\left(\text{Poisson}\left(\int_0^x t dt\right) = 0\right) \\ &= \exp\left(-\int_0^x t dt\right) = \exp(-x^2/2).\end{aligned}$$

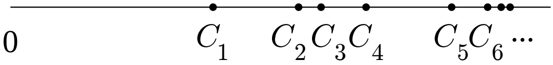
Line-breaking procedure

Exercise

We may equivalently take $E_1, E_2, \ldots$ to be i.i.d. Exponential(1) and set

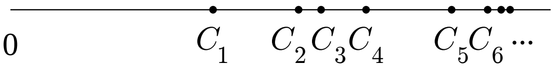
$$C_k = \sqrt{2\sum_{i=1}^{k} E_i}, \; k \geq 1.$$

Line-breaking procedure



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

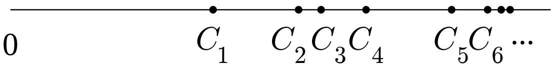
Line-breaking procedure



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

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

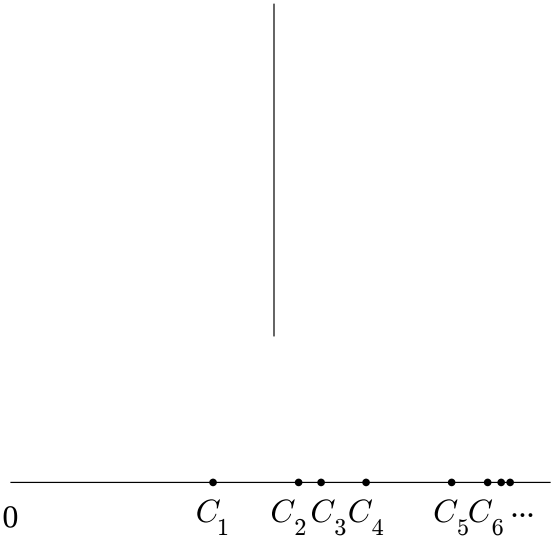
Line-breaking procedure



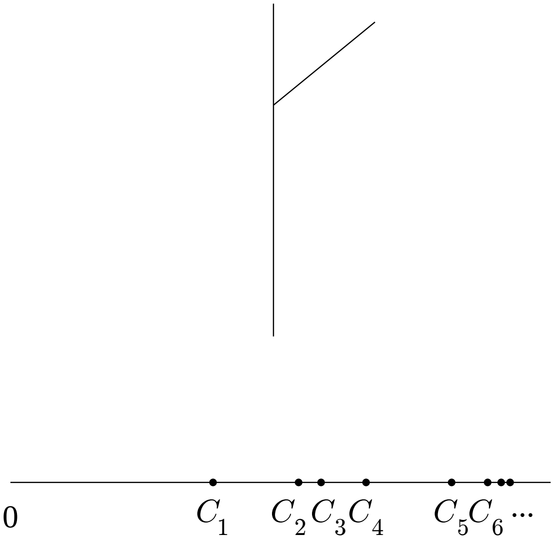
Consider the line-segments $[0, C_1)$, $[C_1, C_2), \ldots$.

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).

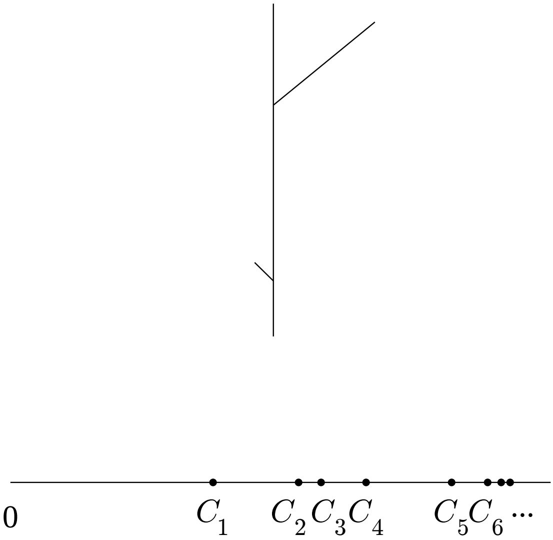
Line-breaking procedure



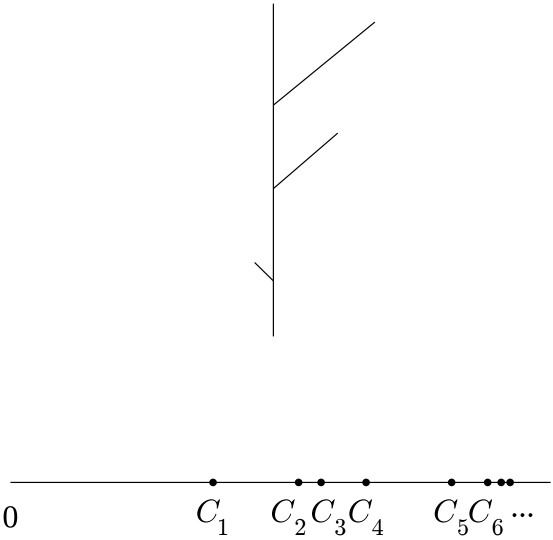
Line-breaking procedure



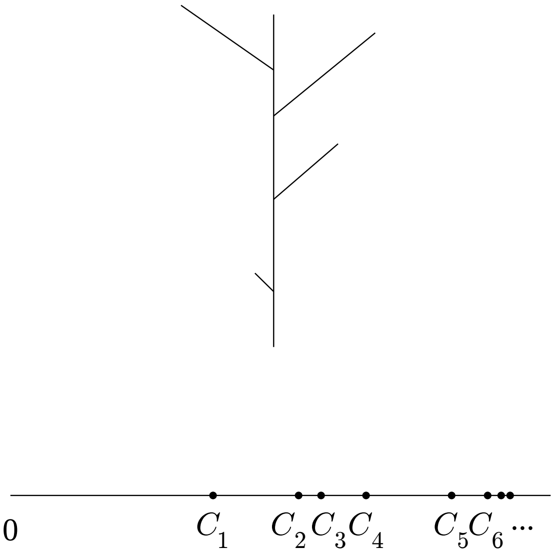
Line-breaking procedure



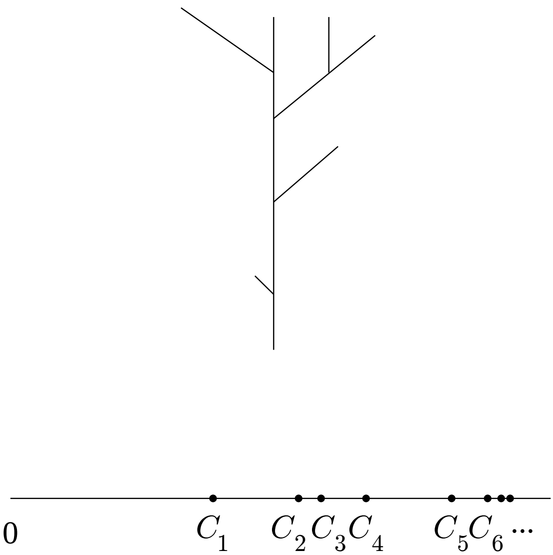
Line-breaking procedure



Line-breaking procedure



Line-breaking procedure



Line-breaking procedure

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

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).

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).

For $n \geq 1$, consider the union of all the line-segments making up the first n branches. We can think of this as a metric space (M_n, d_n) in a natural way. These metric spaces are nested as n increases, so it makes sense to think about the space $M = \cup_{n \geq 1} M_n$.

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).

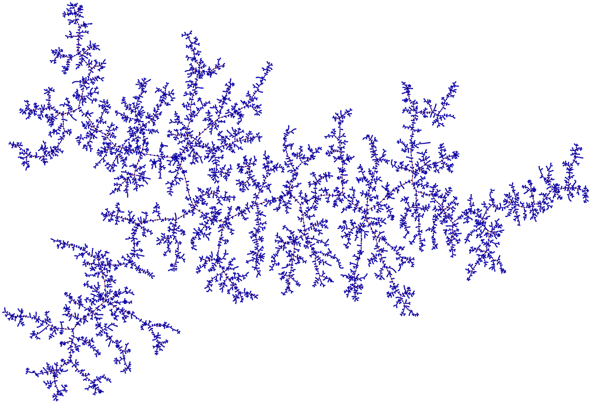
For $n \geq 1$, consider the union of all the line-segments making up the first n branches. We can think of this as a metric space (M_n, d_n) in a natural way. These metric spaces are nested as n increases, so it makes sense to think about the space $M = \cup_{n \geq 1} M_n$.

(Finally, take the metric completion $(\bar{M}, \bar{d})$, formed by adding in all limit points of Cauchy sequences $(x_k)_{k \geq 1}$ in M.)

Line-breaking procedure

The line-breaking procedure gives a (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)) Let T_n be a uniform random labelled tree. 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))

Let T_n be a uniform random labelled tree. 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))

Let T_n be a uniform random labelled tree. 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!

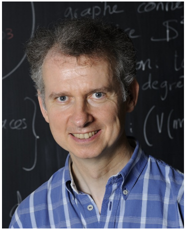
We will, in fact, sketch a proof of a much more general result.

Lecture 2

3. GALTON-WATSON 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: some notation

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

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

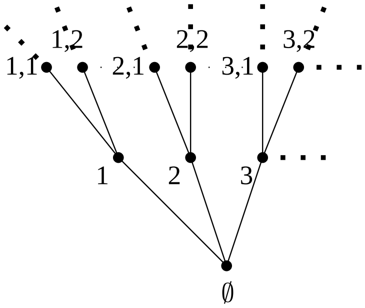
where $\mathbb{N}^0 = \{\emptyset\}$.

Ordered trees: some notation

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

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

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



Thus the label of a vertex indicates its genealogy.

Ordered trees: some notation

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

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

u has children $u1, u2, \ldots$.

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.

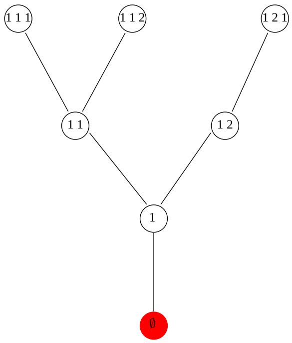
Height function

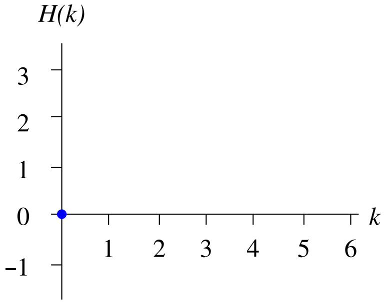
Suppose that $\mathbf{t}$ has n vertices. Let them be $v_0, v_1, \ldots, v_{n-1}$, listed in lexicographical order.

Then the height function is defined by

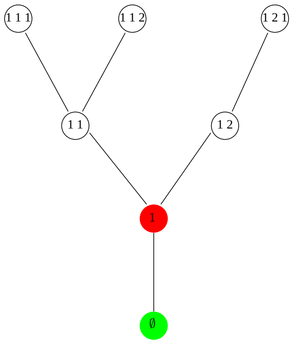
$$H(k) = |v_k|, \quad 0 \leq k \leq n-1.$$

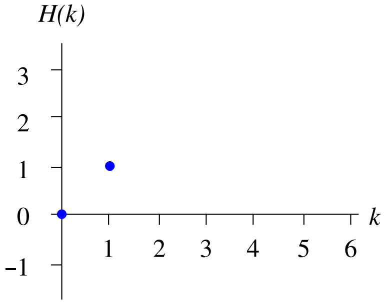
Height function



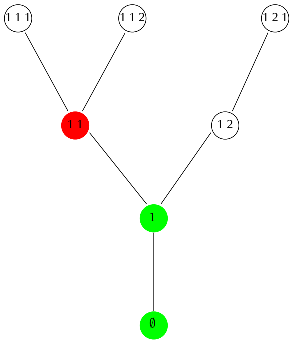


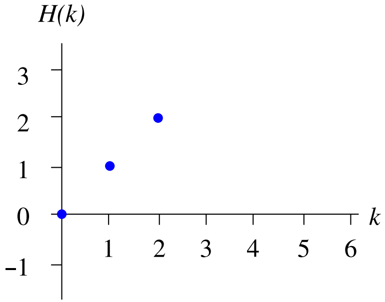
Height function



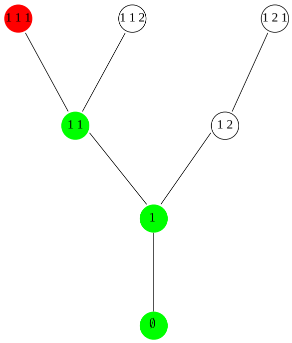


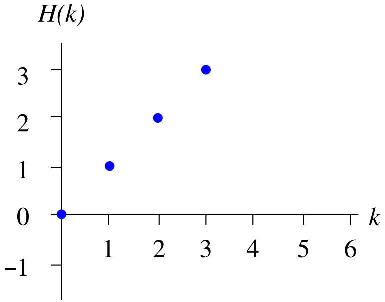
Height function



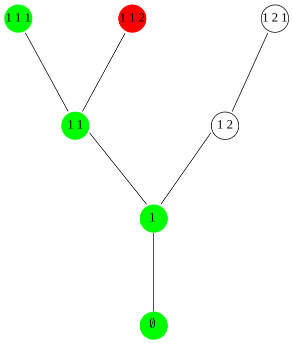


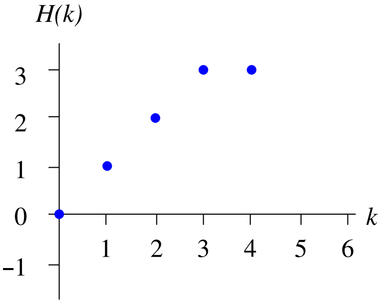
Height function



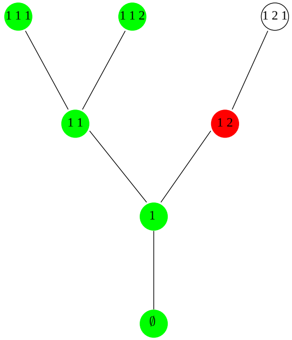


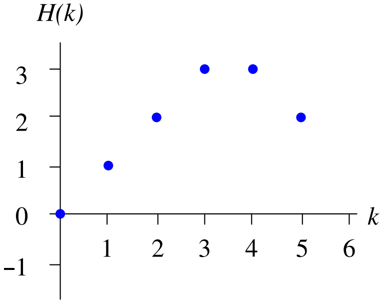
Height function



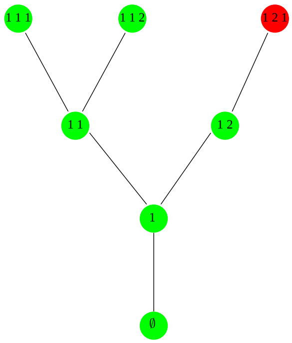


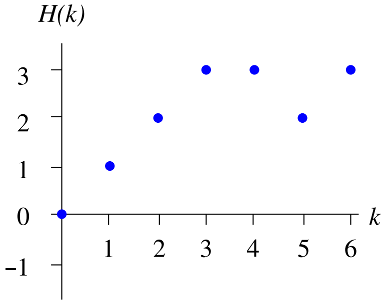
Height function



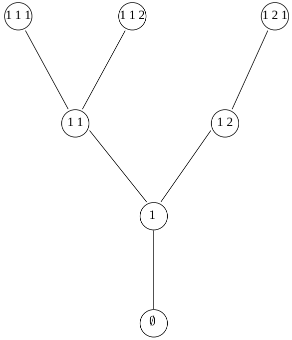


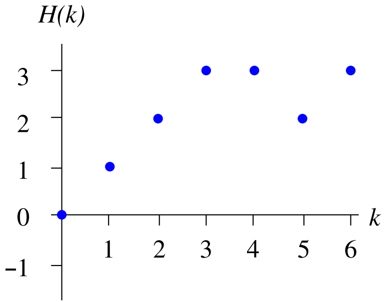
Height function





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

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

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.

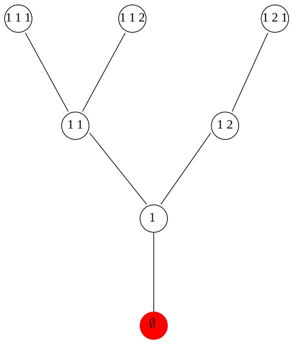
Define

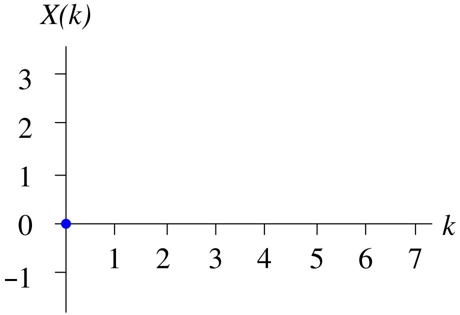
$$X(0) = 0,$$
$$X(i) = \sum_{j=0}^{i-1}(c(v_j) - 1), \text{ for } 1 \leq i \leq n.$$

In other words,

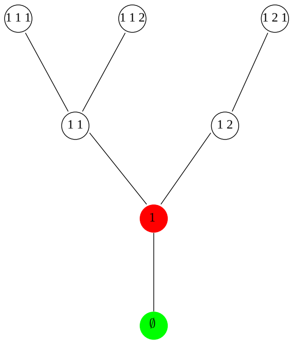
$$X(i+1) = X(i) + c(v_i) - 1, \quad 0 \leq i \leq n-1.$$

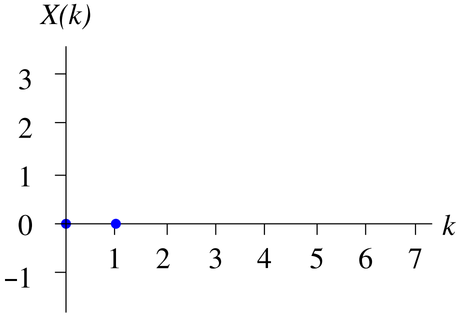
Depth-first walk



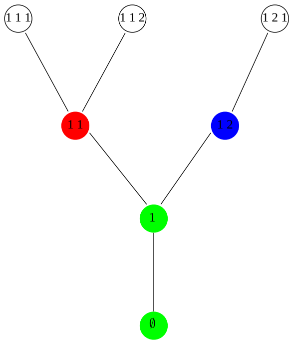


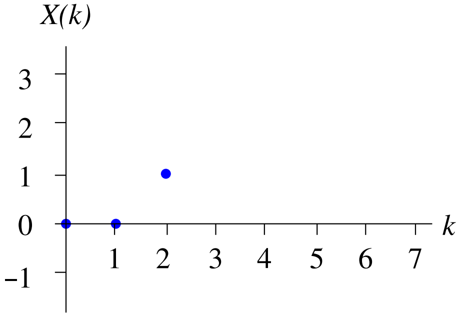
Depth-first walk



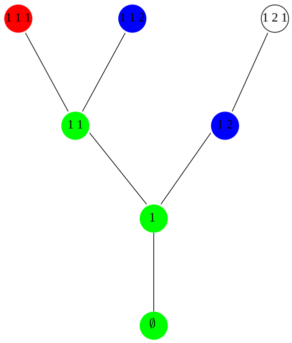


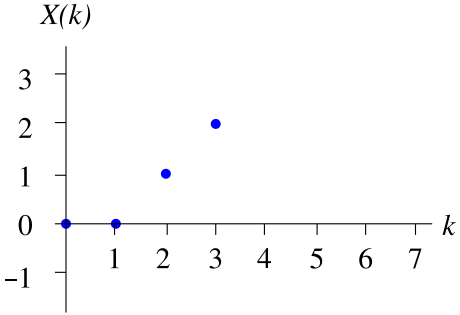
Depth-first walk



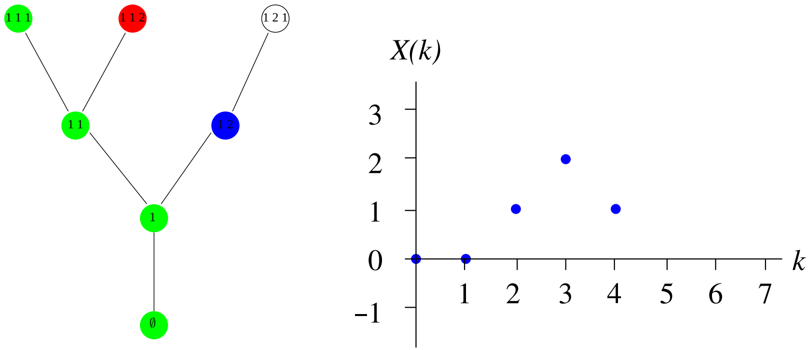


Depth-first walk

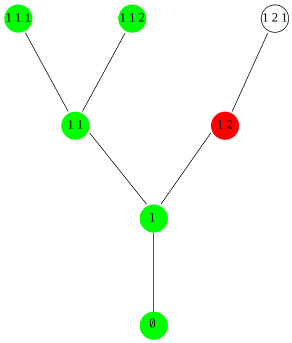


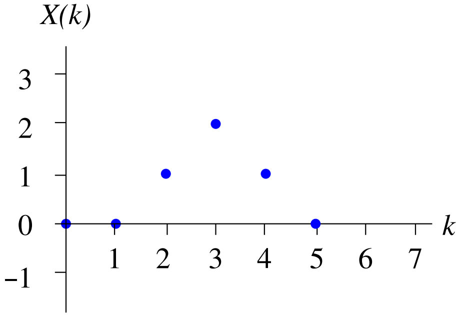


Depth-first walk

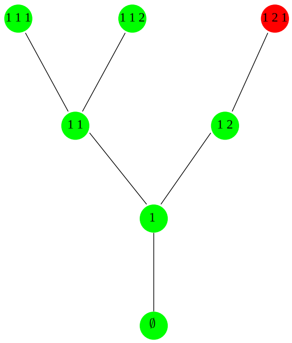


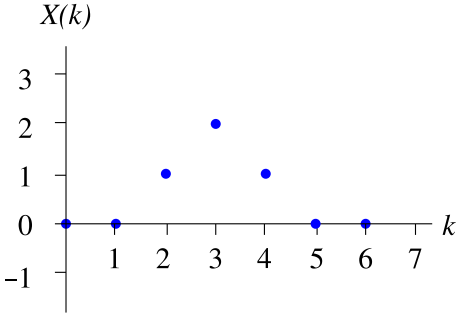
Depth-first walk



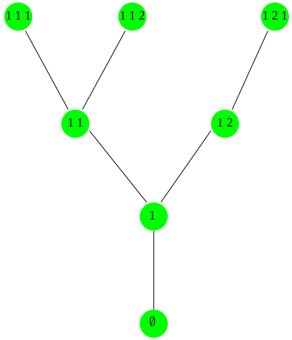


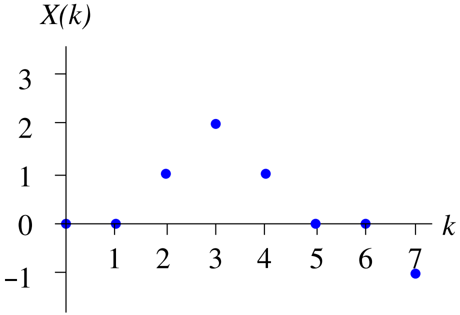
Depth-first walk



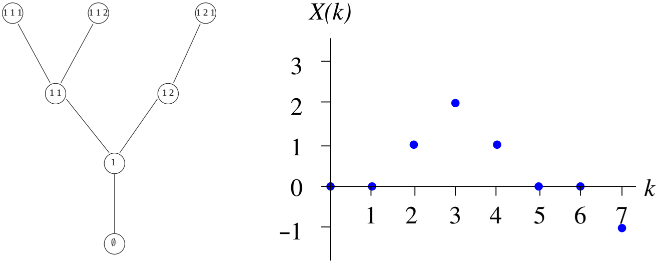


Depth-first walk





Depth-first walk



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

Proposition

For $0 \le i \le n-1$,

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

Random discrete trees

From a probabilistic perspective, a natural probability measure on trees is that generated by a so-called 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 processes

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

Galton-Watson processes

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 processes

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 processes

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 processes

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 processes

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, $Z_0 = 1$). The process $(Z_n)_{n\geq 0}$ is a Markov chain with an absorbing state at 0.

Galton-Watson processes

In order to avoid special cases, we will assume that $p(0) > 0$ and $p(0) + p(1) < 1$. This means that it's always possible for the branching process to die out and we won't have every individual that gives birth just deterministically having a single offspring.

Generating functions

Probability generating functions play a key role in the analysis of branching processes. Let

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

be the probability generating function of the offspring distribution.

Generating functions

Probability generating functions play a key role in the analysis of branching processes. Let

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

be the probability generating function of the offspring distribution.

Let $C_i^{(n)}$ denote the number of children of individual $i \geq 1$ in generation $n \geq 0$, so that

$$Z_n = C_1^{(n-1)} + C_2^{(n-1)} + \cdots + C_{Z_{n-1}}^{(n-1)}.$$

Generating functions

Probability generating functions play a key role in the analysis of branching processes. Let

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

be the probability generating function of the offspring distribution.

Let $C_i^{(n)}$ denote the number of children of individual $i \geq 1$ in generation $n \geq 0$, so that

$$Z_n = C_1^{(n-1)} + C_2^{(n-1)} + \cdots + C_{Z_{n-1}}^{(n-1)}.$$

Then if $G_n(s) = \mathbb{E}\left[s^{Z_n}\right]$, we get $G_1(s) = G(s)$ and, for $n \geq 2$,

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

Extinction probability

Let $q = \mathbb{P}\,(\text{population dies out}) = \mathbb{P}\,(\cup_{n=1}^{\infty}\{Z_n = 0\})$. Since these events are nested ($Z_n = 0$ implies that $Z_{n+1} = 0$), we have

$$q = \lim_{n\to\infty} \mathbb{P}\,(Z_n = 0)\,.$$

Extinction probability

Let $q = \mathbb{P}(\text{population dies out}) = \mathbb{P}(\cup_{n=1}^{\infty}\{Z_n = 0\})$. Since these events are nested ($Z_n = 0$ implies that $Z_{n+1} = 0$), we have

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

Recall that each of the individuals in the first generation behaves exactly like the parent. We can think of each of them starting its own family, which is an independent copy of the original family. Moreover, the whole population dies out if and only if all of the subpopulations die out. If there are k individuals in the first generation, this occurs with probability q^k. So

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

Extinction probability

So q solves the equation $s = G(s)$. Notice that $s = 1$ is always a solution, but it may not be the only one.

Extinction probability

So q solves the equation $s = G(s)$. Notice that $s = 1$ is always a solution, but it may not be the only one.

Theorem

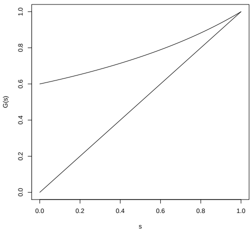
Suppose that $p(0) > 0$ and $p(0) + p(1) < 1$.

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

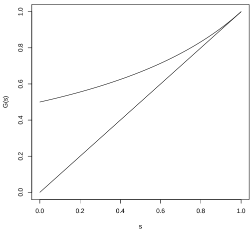
(b) *Suppose that the offspring distribution has mean μ. Then*
- *if $\mu \leq 1$ then $q = 1$;*
- *if $\mu > 1$ then $q < 1$.*

Proof by picture

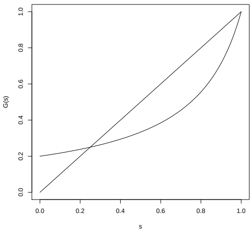
Solving $s = G(s)$:



$\mu < 1$ (subcritical)



$\mu = 1$ (critical)



$\mu > 1$ (supercritical)

Note that $\mathbb{P}(Z_n = 0) = G_n(0)$ and so $q = \lim_{n\to\infty} G_n(0)$.

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.

As before, call its depth-first walk X. Because the numbers of children of different individuals are i.i.d. X has a particularly nice form.

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

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 tree is a stopped random walk

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 tree is a stopped random walk

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 (2005).]

Critical Galton-Watson trees

We will restrict our attention to the case where the offspring distribution p is critical i.e.

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

Critical Galton-Watson trees

We will restrict our attention to the case where the offspring distribution p is critical i.e.

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

Then $q = 1$ and the resulting tree, T, is finite.

Critical Galton-Watson trees

We will restrict our attention to the case where the offspring distribution p is critical i.e.

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

Then $q = 1$ and the resulting tree, T, is finite.

Another way to see this: the depth-first walk has the law of a random walk with step sizes of mean 0 stopped when it first hits -1. Since such a random walk is recurrent, we get $N < \infty$ with probability 1.

Critical Galton-Watson trees

We will restrict our attention to the case where the offspring distribution p is critical i.e.

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

Then $q = 1$ and the resulting tree, T, is finite.

Another way to see this: the depth-first walk has the law of a random walk with step sizes of mean 0 stopped when it first hits -1. Since such a random walk is recurrent, we get $N < \infty$ with probability 1.

The critical case turns out to be the right one to consider in order to capture various natural combinatorial models.

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.

Exercise

1. *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.*
2. *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.*

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

Suppose now that we have offspring variance $\sigma^2 := \sum_{k=1}^{\infty}(k-1)^2 p(k) \in (0, \infty)$ (which is the case for all the examples we have seen so far).

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

Suppose now that we have offspring variance $\sigma^2 := \sum_{k=1}^{\infty}(k-1)^2 p(k) \in (0, \infty)$ (which is the case for all the examples we have seen so far).

Then 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.

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

Suppose now that we have offspring variance $\sigma^2 := \sum_{k=1}^{\infty}(k-1)^2 p(k) \in (0, \infty)$ (which is the case for all the examples we have seen so far).

Then 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 $\sigma^2 := \sum_{k=1}^{\infty}(k-1)^2 p(k) \in (0,\infty)$ (which is the case for all the examples we have seen so far).

Then 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. (This is true if, for example, $p(1) > 0$.)

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 (Kaigh (1976))

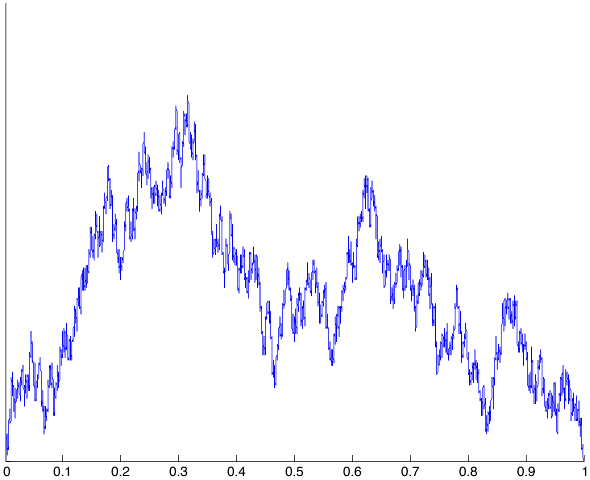
As $n \to \infty$*,*

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

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

[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]

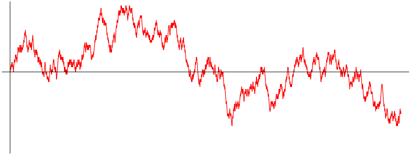
Brownian excursion

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

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

Theorem

As $n \to \infty$,

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

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

Actually, I'm going to cheat...

Consider the unconditioned random walk $(X(k), k \geq 0)$ (without stopping). This is the depth-first walk of a sequence of i.i.d. unconditioned Galton-Watson trees: the random walk X begins encoding a new tree every time it attains a new minimum. It is technically easier to work without the conditioning.

Actually, I'm going to cheat...

Consider the unconditioned random walk $(X(k), k \geq 0)$ (without stopping). This is the depth-first walk of a sequence of i.i.d. unconditioned Galton-Watson trees: the random walk X begins encoding a new tree every time it attains a new minimum. It is technically easier to work without the conditioning.

Define the height process H for all $i \geq 0$ via $H(0) = 0$ 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). This is the depth-first walk of a sequence of i.i.d. unconditioned Galton-Watson trees: the random walk X begins encoding a new tree every time it attains a new minimum. It is technically easier to work without the conditioning.

Define the height process H for all $i \geq 0$ via $H(0) = 0$ and, for $i \geq 1$,

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

This is the height process of the 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). This is the depth-first walk of a sequence of i.i.d. unconditioned Galton-Watson trees: the random walk X begins encoding a new tree every time it attains a new minimum. It is technically easier to work without the conditioning.

Define the height process H for all $i \geq 0$ via $H(0) = 0$ and, for $i \geq 1$,

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

This is the height process of the 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 3

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]

An unconditioned result

From this to get that

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

we need to know that the sequence of processes on the left-hand side is tight. See Le Gall (2005) for the details.

An unconditioned result

From this to get that

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

we need to know that the sequence of processes on the left-hand side is tight. See Le Gall (2005) for the details.

(By a theorem of Lévy, the process on the right-hand side is a reflected Brownian motion, i.e.

$$\left(W(t) - \min_{0 \leq s \leq t} W(s), t \geq 0 \right) \stackrel{d}{=} (|W(s)|, 0 \leq s \leq t),$$

but we won't need this.)

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 $n \to \infty$,

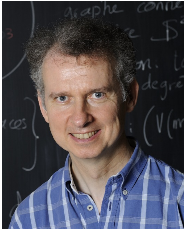
$$\frac{\sigma}{\sqrt{n}} \left(H^n(\lfloor nt \rfloor), 0 \leq t \leq 1\right) \xrightarrow{d} 2 \left(e(t), 0 \leq t \leq 1\right),$$

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

4. $\mathbb{R}$-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 $f_{x,y} : [0, d(x, y)] \to \mathcal{T}$ such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x, y)) = y$. 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 $g : [0, 1] \to \mathcal{T}$ is a continuous injective map 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 $x, y \in \mathcal{T}$,

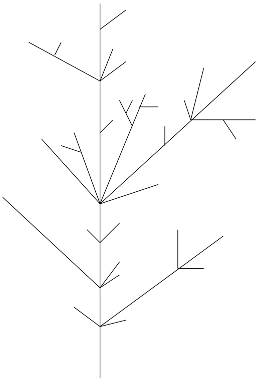
- *There exists a unique shortest path $[[x, y]]$ from x to y (of length $d(x, y)$). (There is a unique isometric map $f_{x,y} : [0, d(x, y)] \to \mathcal{T}$ such that $f_{x,y}(0) = x$ and $f_{x,y}(d(x, y)) = y$. We write $f_{x,y}([0, d(x, y)]) = [[x, y]]$.)*
- *The only non-self-intersecting path from x to y is $[[x, y]]$. (If $g : [0, 1] \to \mathcal{T}$ is a continuous injective map 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 $\mathcal{T} \setminus \{v\}$ is connected.
More generally, the degree of v is the number of connected components of $\mathcal{T} \setminus \{v\}$.

$\mathbb{R}$-trees

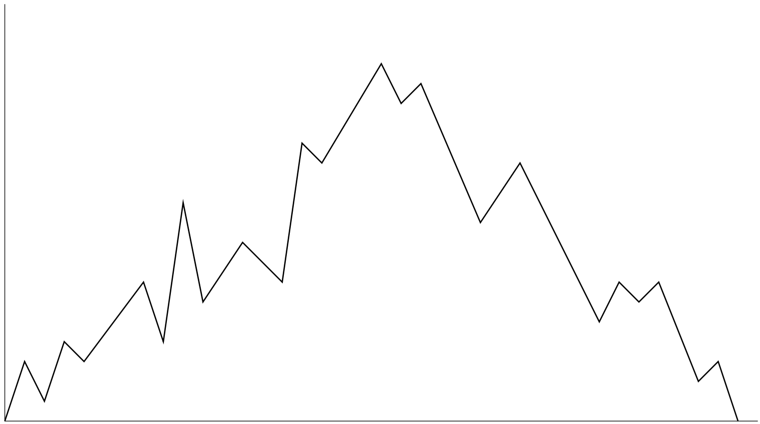
Example

A metric space obtained by glueing together finitely many finite line-segments is an $\mathbb{R}$*-tree.*



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

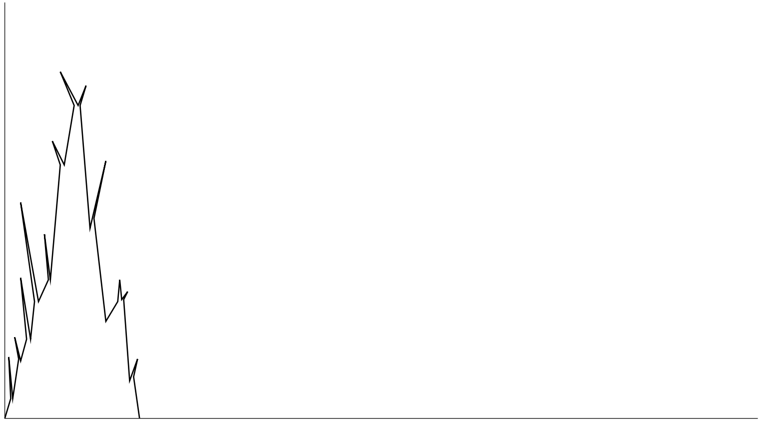
Now put glue on the underside of the excursion and push the two sides together...

Coding $\mathbb{R}$-trees

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

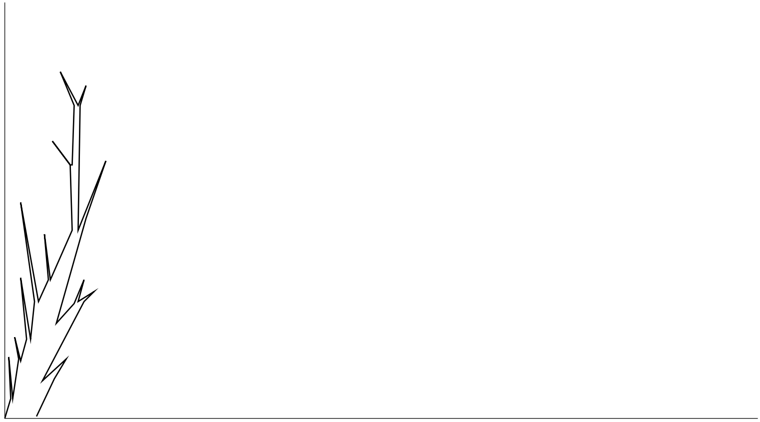
Coding ℝ-trees

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



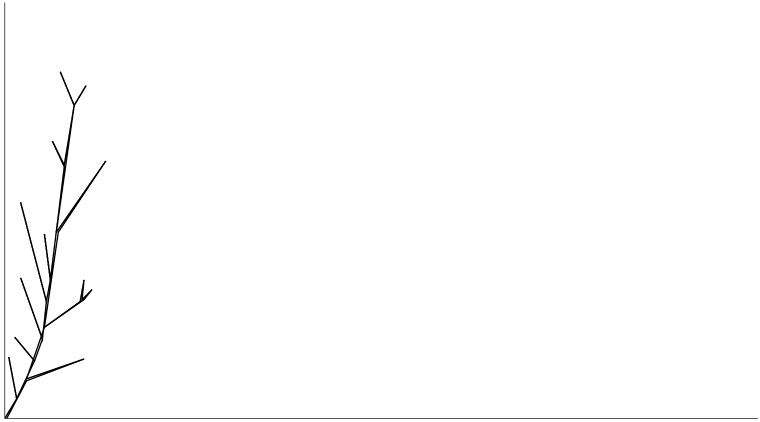
Coding $\mathbb{R}$-trees

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



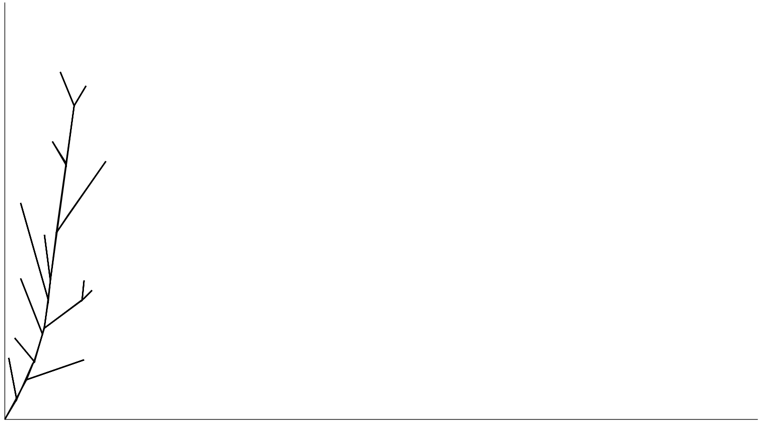
Coding $\mathbb{R}$-trees

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



Coding $\mathbb{R}$-trees

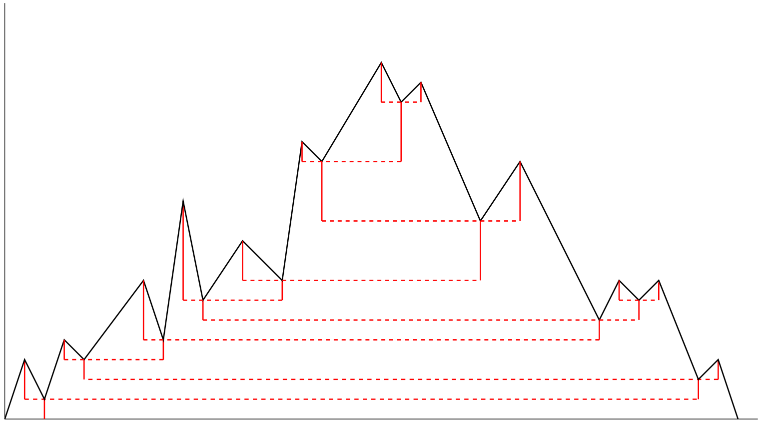
Now put glue on the underside of the excursion and push the two sides together to get a tree.



Coding $\mathbb{R}$-trees

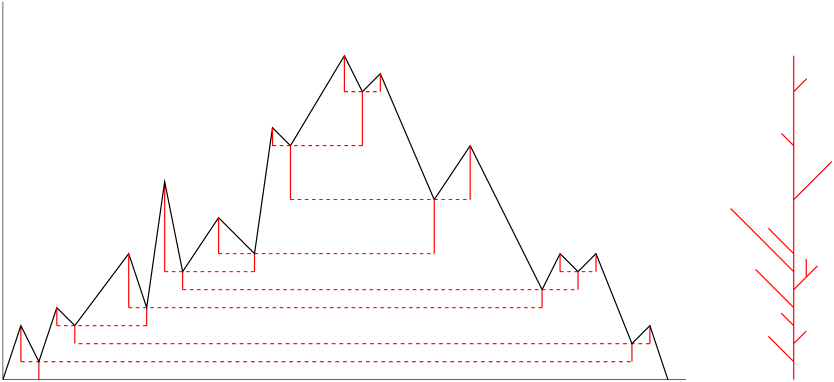
Formally, use h to define a distance:

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



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. Conversely, any (rooted) $\mathbb{R}$-tree can be represented in the form $(\mathcal{T}_g, d_g)$ for some excursion g.

[Proof: see Le Gall (2005).]

Coding $\mathbb{R}$-trees

Theorem

For any excursion h, $(\mathcal{T}_h, d_h)$ is an $\mathbb{R}$-tree. Conversely, any (rooted) $\mathbb{R}$-tree can be represented in the form $(\mathcal{T}_g, d_g)$ for some excursion g.

[Proof: see Le Gall (2005).]

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. Conversely, any (rooted) $\mathbb{R}$-tree can be represented in the form $(\mathcal{T}_g, d_g)$ for some excursion g.

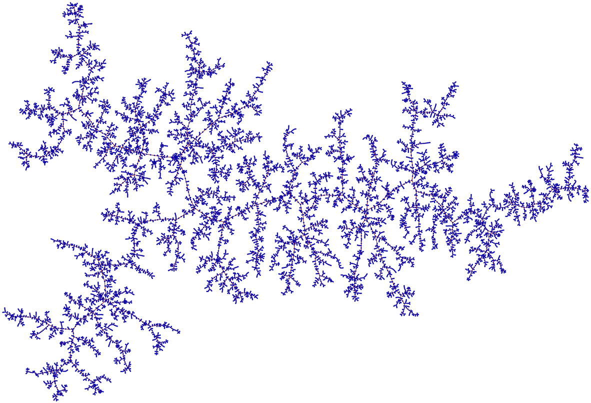
[Proof: see Le Gall (2005).]

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

Definition

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

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

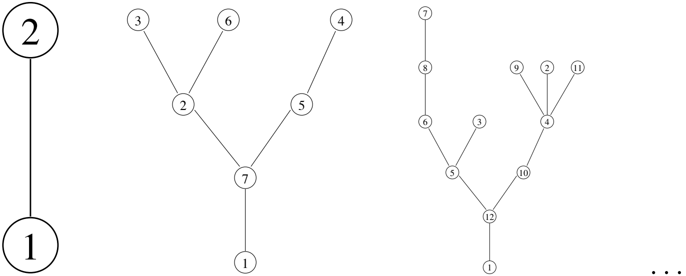


[Picture by Igor Kortchemski]

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 $\mathrm{d_{GH}}$ on $\mathbb{M}$ in a moment. Recall that then

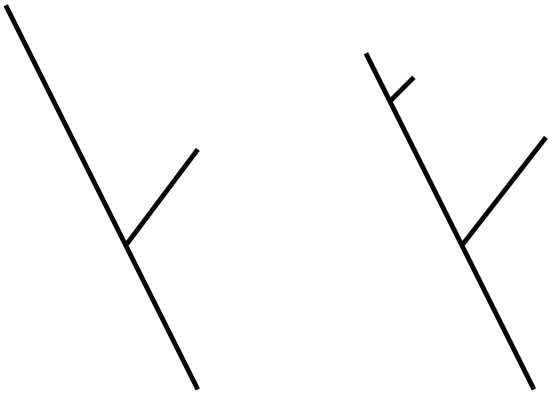
$$(T_n, \sigma d_{\mathrm{gr}}/\sqrt{n}) \xrightarrow{d} (\mathcal{T}_{2e}, d_{2e}) \quad \text{as } n \to \infty$$

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

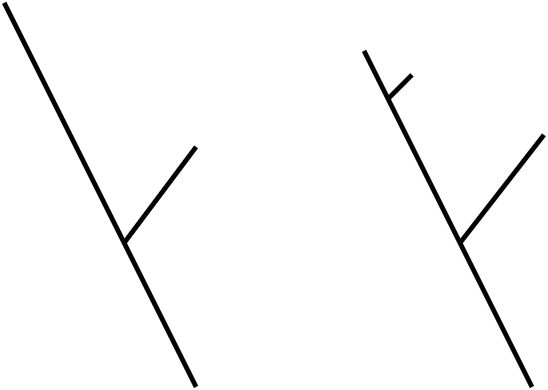
Suppose that (X, d) and (X', d') are compact metric spaces.



Measuring the distance between compact metric spaces

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

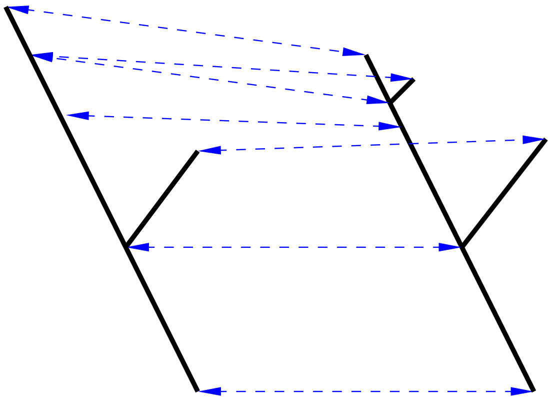
A correspondence R is a subset of $X \times X'$ such that for every $x \in X$, there exists $x' \in X'$ with $(x, x') \in R$ and vice versa.



Measuring the distance between compact metric spaces

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

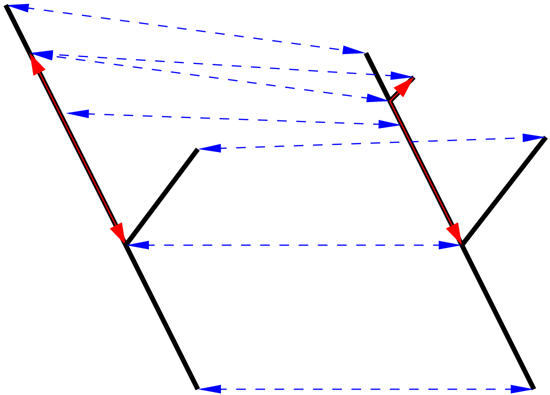
A correspondence R is a subset of $X \times X'$ such that for every $x \in X$, there exists $x' \in X'$ with $(x, x') \in R$ and vice versa.



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

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

(There exists an equivalent definition which more closely resembles that of the usual Hausdorff distance, but this one is easier to use. See Burago, Burago and Ivanov, **A course in metric geometry** for more details.)

Measuring the distance between compact metric spaces

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

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

(There exists an equivalent definition which more closely resembles that of the usual Hausdorff distance, but this one is easier to use. See Burago, Burago and Ivanov, **A course in metric geometry** for more details.)

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

Measuring the distance between compact metric spaces

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

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

(There exists an equivalent definition which more closely resembles that of the usual Hausdorff distance, but this one is easier to use. See Burago, Burago and Ivanov, **A course in metric geometry** for more details.)

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

Theorem

$(\mathbb{M}, d_{GH})$ *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

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

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

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.

Theorem (Aldous (1993), Le Gall (2005))

As $n \to \infty$,

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

where convergence is in the Gromov-Hausdorff sense.

[Approach due to Grégory Miermont.]

Uniform ordered trees

Exercise

There is a simpler argument, using a different functional encoding of the tree, the so-called contour function, which proves the convergence to the Brownian CRT for uniform ordered trees.

Some simple consequences

Let T_n be any of the conditioned Galton-Watson trees to which the theorem applies. Let D_n be the diameter of T_n and let R_n be the distance between two uniformly chosen points. Let D and R be the corresponding quantities for the Brownian CRT.

Some simple consequences

Let T_n be any of the conditioned Galton-Watson trees to which the theorem applies. Let D_n be the diameter of T_n and let R_n be the distance between two uniformly chosen points. Let D and R be the corresponding quantities for the Brownian CRT.

Corollary

We have

$$\frac{\sigma}{\sqrt{n}} D_n \xrightarrow{d} D \quad \textit{and} \quad \frac{\sigma}{\sqrt{n}} R_n \xrightarrow{d} R$$

as $n \to \infty$.

Some simple consequences

Let T_n be any of the conditioned Galton-Watson trees to which the theorem applies. Let D_n be the diameter of T_n and let R_n be the distance between two uniformly chosen points. Let D and R be the corresponding quantities for the Brownian CRT.

Corollary

We have

$$\frac{\sigma}{\sqrt{n}} D_n \xrightarrow{d} D \quad \text{and} \quad \frac{\sigma}{\sqrt{n}} R_n \xrightarrow{d} R$$

as $n \to \infty$.

It turns out that

$$\mathbb{P}(D > x) = \sum_{k=1}^{\infty} e^{-2k^2x^2}(8k^2x^2 - 2), \ x \geq 0$$

and

$$\mathbb{P}(R > x) = \exp(-x^2/2), \ x \geq 0.$$

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 rooted trees [Haas & Miermont (2012)]
- uniform unordered unrooted trees [Stufler (2014+)]
- critical multi-type Galton-Watson trees [Miermont (2008)]
- random trees with a prescribed degree sequence satisfying certain conditions [Broutin & Marckert (2014)]
- random dissections [Curien, Haas & Kortchemski (2015)]
- random graphs from subcritical classes [Panagiotou, Stufler & Weller (2014+)]
- the range of a Brownian bridge in a hyperbolic space [Chen & Miermont (2016+)]
- the trace of a random walk bridge on an infinite d-regular tree for $d \geq 3$ [Stewart (2016++)]

Applications

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

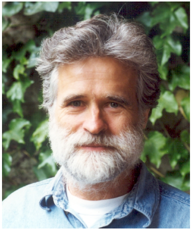
- the scaling limit of random planar maps [Le Gall (2013), Miermont (2013)];
- the scaling limit of the critical Erdős-Rényi random graph [Addario-Berry, Broutin, G. (2010, 2012)].

5. 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 $1 \leq i \leq k$.

Gromov-Hausdorff-Prokhorov distance

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

Gromov-Hausdorff-Prokhorov distance

We can endow the set of continuum trees with a generalisation of the Gromov-Hausdorff distance, the Gromov-Hausdorff-Prokhorov distance, 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 probability measures which puts as small mass as possible outside the correspondence.

Gromov-Hausdorff-Prokhorov distance

More formally, suppose we have compact measured metric spaces (X, d, μ) and (X', d', μ').

Gromov-Hausdorff-Prokhorov distance

More formally, suppose we have compact measured metric spaces (X, d, μ) and (X', d', μ').

Define a coupling of μ and μ' to be a probability measure m on $X \times X'$ such that for $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(X')$,

$$m(A, X') = \mu(A) \quad \text{and} \quad m(X, B) = \mu'(B).$$

Gromov-Hausdorff-Prokhorov distance

More formally, suppose we have compact measured metric spaces (X, d, μ) and (X', d', μ').

Define a coupling of μ and μ' to be a probability measure m on $X \times X'$ such that for $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(X')$,

$$m(A, X') = \mu(A) \quad \text{and} \quad m(X, B) = \mu'(B).$$

Then define the Gromov-Hausdorff-Prokhorov distance to be

$$\mathrm{d}_{\mathrm{GHP}}((X, d, \mu), (X', d', \mu')) = \inf_{R,m} \max \left\{ \frac{1}{2}\mathrm{dis}(R), m(R^c) \right\},$$

where the infimum is over all possible correspondences $R \subseteq X \times X'$ and all possible couplings m of μ and μ'.

Gromov-Hausdorff-Prokhorov distance

More formally, suppose we have compact measured metric spaces (X, d, μ) and (X', d', μ').

Define a coupling of μ and μ' to be a probability measure m on $X \times X'$ such that for $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(X')$,

$$m(A, X') = \mu(A) \quad \text{and} \quad m(X, B) = \mu'(B).$$

Then define the Gromov-Hausdorff-Prokhorov distance to be

$$\mathrm{d_{GHP}}((X, d, \mu), (X', d', \mu')) = \inf_{R,m} \max \left\{ \frac{1}{2}\mathrm{dis}(R), m(R^c) \right\},$$

where the infimum is over all possible correspondences $R \subseteq X \times X'$ and all possible couplings m of μ and μ'.

If $\mathbb{M}^*$ is the set of compact measured metric spaces, up to measured isometry, then

$$(\mathbb{M}^*, \mathrm{d_{GHP}})$$

is again a complete separable metric space.

What is a continuum random tree?!

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

What is a continuum random tree?!

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

As its name would suggest, the Brownian CRT is a continuum random tree! In order to make this precise, we need a measure.

What is a continuum random tree?!

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

As its name would suggest, the Brownian CRT is a continuum random tree! In order to make this precise, we need a measure.

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

The mass measure of the Brownian CRT

Extra exercise (for the keen!)

Consider a uniform random tree T_n. Put mass $1/n$ at each vertex. Call the resulting probability measure μ_n. Show that

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

as $n \to \infty$, in the sense of the Gromov-Hausdorff-Prokhorov distance.

The mass measure of the Brownian CRT

Extra exercise (for the keen!)

Consider a uniform random tree T_n. Put mass $1/n$ at each vertex. Call the resulting probability measure μ_n. Show that

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

as $n \to \infty$, in the sense of the Gromov-Hausdorff-Prokhorov distance.

Lemma

1. $\mu_{2e}(\mathcal{L}(\mathcal{T}_{2e})) = 1$.
2. *For every $v \in \mathcal{T}_{2e}$ of degree $k \geq 2$, if $\mathcal{T}_1, \ldots, \mathcal{T}_k$ are the connected components of $\mathcal{T}_{2e} \setminus \{v\}$ then $\mu(\mathcal{T}_i) > 0$ for all $1 \leq i \leq k$.*

[Intuition: non-leaf vertices of T_n are typically at distance $o(\sqrt{n})$ from a leaf, and the leaves are spread "uniformly" over the tree. Proof: see Aldous (1991).]

Lecture 4

The root of the Brownian CRT

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

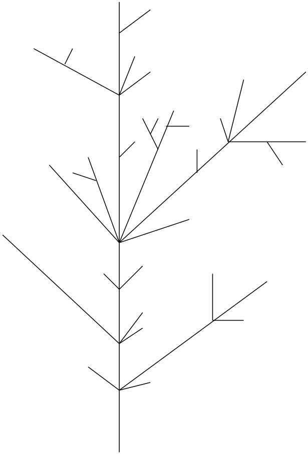
Take a CRT $(\mathcal{T}, d, \mu)$ and suppose that $U_1, U_2, \ldots$ are i.i.d. samples from the measure μ.

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.)

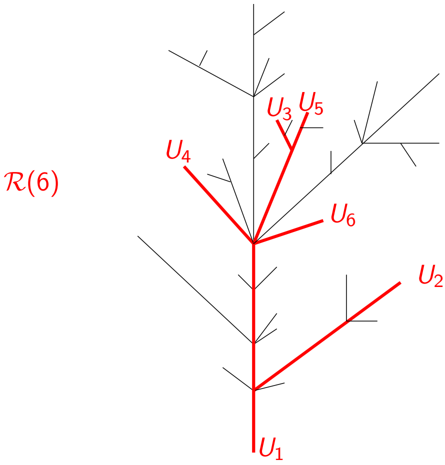
Characterising a CRT via sampling

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 $U_1, U_2, \ldots, U_m$.



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.

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* $Dir(a_1, a_2, \ldots, a_n)$*, 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* $\mathcal{S}_n$*.*

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 $Dir(a_1, a_2, \ldots, a_n)$, 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

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

Definition

The Dirichlet distribution with parameters $a_1, a_2, \dots, a_n > 0$, written $Dir(a_1, a_2, \dots, a_n)$, 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 $\mathcal{S}_n$.

Note: If $B \sim \text{Beta}(a_1, a_2)$ then $(B, 1-B) \sim \text{Dir}(a_1, a_2)$. $\text{Dir}(1, 1, \dots, 1)$ 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 $\text{U}(0, 1)$ random variables.

Dirichlet distribution facts (size-biased sampling)

Proposition

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

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

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

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

Dirichlet distribution facts (beta-gamma algebra)

Exercise

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

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

where
$G_1 \sim Gamma(a_1, 1), G_2 \sim Gamma(a_2, 1), \ldots, G_n \sim Gamma(a_n, 1)$ *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 $\sum_{i=1}^{n} G_i \sim Gamma(\sum_{i=1}^{n} a_i, 1)$.

Dirichlet distribution facts (beta-gamma algebra)

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

Proposition

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

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

Note: $\text{Beta}(1, 1) = \text{U}[0, 1]$.

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 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 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 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,

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

and

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

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.

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

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.

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

So it remains to prove the following 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}.

Line-breaking revisited

Sketch proof of claim:

Line-breaking revisited

Sketch proof of claim:

1. For any $m \geq 2$, $\left(\frac{C_{m-1}}{C_m}, \frac{C_m - C_{m-1}}{C_m}\right) \sim \text{Dir}(2m-2, 1)$ independently of C_m.

Line-breaking revisited

Sketch proof of claim:

1. For any $m \geq 2$, $\left(\frac{C_{m-1}}{C_m}, \frac{C_m - C_{m-1}}{C_m}\right) \sim \text{Dir}(2m-2, 1)$ independently of C_m.
2. For $m = 2$, $\left(\frac{B_1}{C_2}, \frac{C_1 - B_1}{C_2}, \frac{C_2 - C_1}{C_2}\right) \stackrel{d}{=} \left(\frac{UC_1}{C_2}, \frac{(1-U)C_1}{C_2}, \frac{C_2 - C_1}{C_2}\right)$, where $U \sim \text{U}[0,1]$ is independent of everything else. Combining with 1. and the previous proposition, we get that this has $\text{Dir}(1,1,1)$ law.

Line-breaking revisited

Sketch proof of claim:

1. For any $m \geq 2$, $\left(\frac{C_{m-1}}{C_m}, \frac{C_m - C_{m-1}}{C_m}\right) \sim \mathrm{Dir}(2m-2, 1)$ independently of C_m.
2. For $m = 2$, $\left(\frac{B_1}{C_2}, \frac{C_1 - B_1}{C_2}, \frac{C_2 - C_1}{C_2}\right) \stackrel{d}{=} \left(\frac{UC_1}{C_2}, \frac{(1-U)C_1}{C_2}, \frac{C_2 - C_1}{C_2}\right)$, where $U \sim \mathrm{U}[0,1]$ is independent of everything else. Combining with 1. and the previous proposition, we get that this has $\mathrm{Dir}(1,1,1)$ law.
3. Now proceed by induction: suppose that the given subintervals have lengths $(L_1, \ldots, L_{2m-3}) \sim \mathrm{Dir}(1, 1, \ldots, 1)$. Sampling B_{m-1} takes a size-biased pick from among these intervals, and splits it at a uniform position. This gives back lengths $(\tilde{L}_1, \ldots, \tilde{L}_{2m-2}) \sim \mathrm{Dir}(1, 1, \ldots, 1)$.

Line-breaking revisited

Sketch proof of claim:

1. For any $m \geq 2$, $\left(\frac{C_{m-1}}{C_m}, \frac{C_m - C_{m-1}}{C_m}\right) \sim \mathrm{Dir}(2m-2, 1)$ independently of C_m.
2. For $m = 2$, $\left(\frac{B_1}{C_2}, \frac{C_1 - B_1}{C_2}, \frac{C_2 - C_1}{C_2}\right) \overset{d}{=} \left(\frac{UC_1}{C_2}, \frac{(1-U)C_1}{C_2}, \frac{C_2 - C_1}{C_2}\right)$, where $U \sim \mathrm{U}[0,1]$ is independent of everything else. Combining with 1. and the previous proposition, we get that this has $\mathrm{Dir}(1,1,1)$ law.
3. Now proceed by induction: suppose that the given subintervals have lengths $(L_1, \ldots, L_{2m-3}) \sim \mathrm{Dir}(1,1,\ldots,1)$. Sampling B_{m-1} takes a size-biased pick from among these intervals, and splits it at a uniform position. This gives back lengths $(\tilde{L}_1, \ldots, \tilde{L}_{2m-2}) \sim \mathrm{Dir}(1,1,\ldots,1)$.
4. Then the lengths we want are $\left(\frac{C_{m-1}}{C_m}(\tilde{L}_1, \ldots, \tilde{L}_{2m-2}), \frac{C_m - C_{m-1}}{C_m}\right)$ which has distribution $\mathrm{Dir}(1,1,\ldots,1)$ by 1. and the previous proposition.

Line-breaking revisited

Sketch proof of claim:

1. For any $m \geq 2$, $\left(\frac{C_{m-1}}{C_m}, \frac{C_m - C_{m-1}}{C_m}\right) \sim \text{Dir}(2m-2, 1)$ independently of C_m.
2. For $m = 2$, $\left(\frac{B_1}{C_2}, \frac{C_1 - B_1}{C_2}, \frac{C_2 - C_1}{C_2}\right) \stackrel{d}{=} \left(\frac{UC_1}{C_2}, \frac{(1-U)C_1}{C_2}, \frac{C_2 - C_1}{C_2}\right)$, where $U \sim \text{U}[0,1]$ is independent of everything else. Combining with 1. and the previous proposition, we get that this has $\text{Dir}(1,1,1)$ law.
3. Now proceed by induction: suppose that the given subintervals have lengths $(L_1, \ldots, L_{2m-3}) \sim \text{Dir}(1, 1, \ldots, 1)$. Sampling B_{m-1} takes a size-biased pick from among these intervals, and splits it at a uniform position. This gives back lengths $(\tilde{L}_1, \ldots, \tilde{L}_{2m-2}) \sim \text{Dir}(1, 1, \ldots, 1)$.
4. Then the lengths we want are $\left(\frac{C_{m-1}}{C_m}(\tilde{L}_1, \ldots, \tilde{L}_{2m-2}), \frac{C_m - C_{m-1}}{C_m}\right)$ which has distribution $\text{Dir}(1, 1, \ldots, 1)$ by 1. and the previous proposition.

This completes the proof of the proposition.

Line-breaking revisited

Proposition

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

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

Line-breaking revisited

Proposition

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

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 can be shown to be 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* $(\Delta_1, \Delta_2, \Delta_3) \sim Dir(1/2, 1/2, 1/2)$.
- *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* $i = 1, 2, 3$.

A random fractal

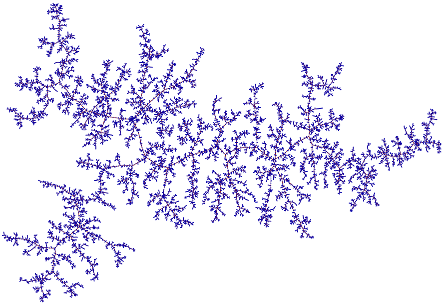
The self-similarity of the Brownian CRT tells us, in particular, that it is a random fractal.

A random fractal

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

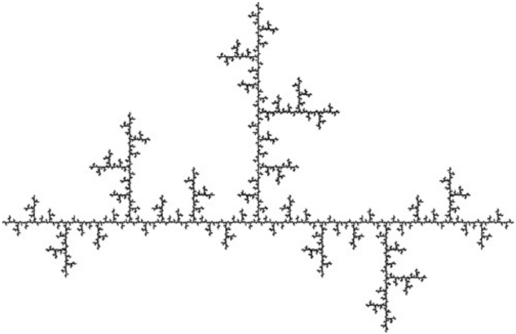
Theorem (Haas & Miermont (2004), Duquesne & Le Gall (2005))

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.

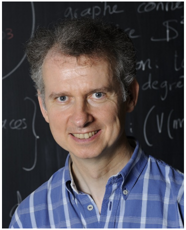


[Image from Croydon & Hambly (2008)]

6. THE STABLE TREES

Key reference:

Thomas Duquesne & Jean-François Le Gall, **Random trees, Lévy processes and spatial branching processes**,
Astérisque **281** (2002)



Infinite variance

Write T_n for a Galton-Watson tree with critical offspring distribution $(p(k), k \geq 0)$, 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 $(p(k), k \geq 0)$, 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?

It's going to be important to understand what happens to sums of i.i.d. random variables with mean 0 and infinite variance. We will treat a particular special case.

A generalised central limit theorem

Theorem

Let $Z_1, Z_2, \ldots$ be i.i.d. random variables such that $\mathbb{P}(Z_1 \geq -1) = 1$, $\mathbb{E}[Z_1] = 0$ and, for some $\alpha \in (1, 2)$,

$$\mathbb{P}(Z_1 = k) \sim ck^{-\alpha-1} \text{ as } k \to \infty,$$

for some constant $c > 0$. Then as $n \to \infty$,

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

where S_α is a random variable with Laplace transform

$$\mathbb{E}[\exp(-\lambda S_\alpha)] = \exp(C_\alpha \lambda^\alpha), \ \lambda \geq 0,$$

for $C_\alpha = \frac{c\Gamma(2-\alpha)}{\alpha(\alpha-1)}$.

A generalised central limit theorem

Notice that we can include the case $\alpha = 2$: if $\mathbb{E}\left[Z_1^2\right] = \sigma^2 < \infty$ then we get that

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} Z_i \xrightarrow{d} S_2,$$

where S_2 has a $N(0,1)$ distribution, with Laplace transform

$$\mathbb{E}\left[\exp(-\lambda S_2)\right] = \exp(C_2\lambda^2),\ \lambda \in \mathbb{R}.$$

Stable laws

We say that the random variables S_α, $\alpha \in (1, 2]$ have stable laws. There is, in fact, a two-parameter family of such distributions, which have the property that for every $n \geq 1$, there exist constants a_n and b_n such that if Y has such a distribution then Y satisfies the recursive distributional equation

$$Y \stackrel{d}{=} \frac{Y_1 + Y_2 + \cdots + Y_n - a_n}{b_n}$$

where $Y_1, Y_2, \ldots$ are i.i.d. copies of Y.

Stable laws

We say that the random variables S_α, $\alpha \in (1, 2]$ have stable laws. There is, in fact, a two-parameter family of such distributions, which have the property that for every $n \geq 1$, there exist constants a_n and b_n such that if Y has such a distribution then Y satisfies the recursive distributional equation

$$Y \stackrel{d}{=} \frac{Y_1 + Y_2 + \cdots + Y_n - a_n}{b_n}$$

where $Y_1, Y_2, \ldots$ are i.i.d. copies of Y.

In our case, $a_n = 0$ and $b_n = n^{1/\alpha}$ for $n \geq 1$.

Stable laws

We say that the random variables S_α, $\alpha \in (1, 2]$ have stable laws. There is, in fact, a two-parameter family of such distributions, which have the property that for every $n \geq 1$, there exist constants a_n and b_n such that if Y has such a distribution then Y satisfies the recursive distributional equation

$$Y \stackrel{d}{=} \frac{Y_1 + Y_2 + \cdots + Y_n - a_n}{b_n}$$

where $Y_1, Y_2, \ldots$ are i.i.d. copies of Y.

In our case, $a_n = 0$ and $b_n = n^{1/\alpha}$ for $n \geq 1$.

[Reference: see Durrett, **Probability theory and examples** for a very beautiful presentation of the stable laws and how they arise.]

Functional convergence

In order to understand the behaviour of a single conditioned Galton-Watson tree, we again start by understanding the depth-first walk X corresponding to a sequence of i.i.d. unconditioned Galton-Watson trees.

Functional convergence

In order to understand the behaviour of a single conditioned Galton-Watson tree, we again start by understanding the depth-first walk X corresponding to a sequence of i.i.d. unconditioned Galton-Watson trees.

The functional convergence is as follows.

Theorem

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

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

where L is an α-stable Lévy process with no negative jumps, having Laplace transform

$$\mathbb{E}\left[\exp(-\lambda L(t))\right] = \exp(C_\alpha \lambda^\alpha t), \ \lambda \geq 0.$$

The Lévy process L

$L = (L(t), t \geq 0)$ is a process with stationary independent increments. We have $L(0) = 0$, and for fixed $t \geq 0$, $L(t)$ has Laplace transform

$$\mathbb{E}\left[\exp(-\lambda L(t))\right] = \exp(C_\alpha \lambda^\alpha t), \ \ \lambda \geq 0.$$

The Lévy process L

$L = (L(t), t \geq 0)$ is a process with stationary independent increments. We have $L(0) = 0$, and for fixed $t \geq 0$, $L(t)$ has Laplace transform

$$\mathbb{E}\left[\exp(-\lambda L(t))\right] = \exp(C_\alpha \lambda^\alpha t), \ \ \lambda \geq 0.$$

Because of the stationary independent increments, this determines all the finite-dimensional distributions of the process:

$$\begin{aligned}
&\mathbb{E}\left[\exp(-(\lambda_1 - \lambda_2)L(t_1) - (\lambda_2 - \lambda_3)L(t_2) - \cdots - \lambda_n L(t_n))\right] \\
&= \mathbb{E}\left[\exp(-\lambda_1 L(t_1) - \lambda_2[L(t_2) - L(t_1)] - \cdots - \lambda_n[L(t_n) - L(t_{n-1})]\right] \\
&= \exp(C_\alpha[\lambda_1^\alpha t_1 + \lambda_2^\alpha(t_2 - t_1) + \cdots + \lambda_n^\alpha(t_n - t_{n-1})])
\end{aligned}$$

and so determines its law.

The Lévy process L

$L = (L(t), t \geq 0)$ is a process with stationary independent increments. We have $L(0) = 0$, and for fixed $t \geq 0$, $L(t)$ has Laplace transform

$$\mathbb{E}\left[\exp(-\lambda L(t))\right] = \exp(C_\alpha \lambda^\alpha t),\ \ \lambda \geq 0.$$

Because of the stationary independent increments, this determines all the finite-dimensional distributions of the process:

$$\begin{aligned}
&\mathbb{E}\left[\exp(-(\lambda_1 - \lambda_2)L(t_1) - (\lambda_2 - \lambda_3)L(t_2) - \cdots - \lambda_n L(t_n))\right] \\
&= \mathbb{E}\left[\exp(-\lambda_1 L(t_1) - \lambda_2[L(t_2) - L(t_1)] - \cdots - \lambda_n[L(t_n) - L(t_{n-1})]\right] \\
&= \exp(C_\alpha[\lambda_1^\alpha t_1 + \lambda_2^\alpha(t_2 - t_1) + \cdots + \lambda_n^\alpha(t_n - t_{n-1})])
\end{aligned}$$

and so determines its law.

Recall that we had $\mathbb{E}\left[\exp(-\lambda S_\alpha)\right] = \exp(C_\alpha \lambda^\alpha)$, which entails that

$$t^{-1/\alpha} L(t) \stackrel{d}{=} L(1) \stackrel{d}{=} S_\alpha,\ \ t \geq 0.$$

So $L(t)$ has a stable law for each t, and the process is self-similar with index α.

Height process convergence

Recall that

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

Height process convergence

Recall that

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

In the case of finite variance, the limiting relationship between X and H was multiplication by a constant. That is no longer the case here.

Height process convergence

Recall that

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

In the case of finite variance, the limiting relationship between X and H was multiplication by a constant. That is no longer the case here.

The limiting analogue $(H_t^{(\alpha)}, t \geq 0)$ is defined as a (suitably normalised) local time at level 0 of the process

$$\left(L_s - \inf_{s \leq r \leq t} L_r, 0 \leq s \leq t \right).$$

(The local time is a measure of how much time this process spends at 0.)

Height process convergence

Recall that

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

In the case of finite variance, the limiting relationship between X and H was multiplication by a constant. That is no longer the case here.

The limiting analogue $(H_t^{(\alpha)}, t \geq 0)$ is defined as a (suitably normalised) local time at level 0 of the process

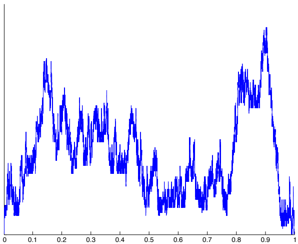
$$\left(L_s - \inf_{s \leq r \leq t} L_r, 0 \leq s \leq t \right).$$

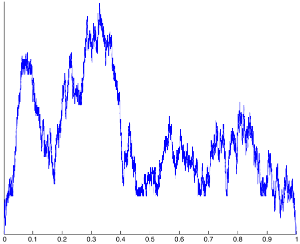
(The local time is a measure of how much time this process spends at 0.)

It turns out that $(H_t^{(\alpha)}, t \geq 0)$ is a continuous process (but it has some pretty weird properties!).

An excursion $e^{(\alpha)}$ of the limiting height process

There is an excursion theory for the α-stable Lévy process L, which enables us to think about a single tree, and we can again make sense of an excursion $e^{(\alpha)}$ of $H^{(\alpha)}$ of length 1.





[Pictures by Igor Kortchemski]

Height process convergence

Theorem (Duquesne & Le Gall (2002); Duquesne (2003))

As $n \to \infty$,

$$n^{-\frac{(\alpha-1)}{\alpha}}(H^n(\lfloor nt \rfloor), 0 \leq t \leq 1) \xrightarrow{d} C(e^{(\alpha)}(t), 0 \leq t \leq 1).$$

Height process convergence

Theorem (Duquesne & Le Gall (2002); Duquesne (2003))

As $n \to \infty$,

$$n^{-\frac{(\alpha-1)}{\alpha}}(H^n(\lfloor nt \rfloor), 0 \le t \le 1) \xrightarrow{d} C(e^{(\alpha)}(t), 0 \le t \le 1).$$

As before, this is the key result that enables us to deduce the convergence of the trees.

The stable trees

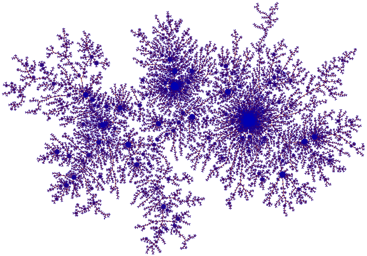
Theorem (Duquesne & Le Gall (2002); Duquesne (2003))

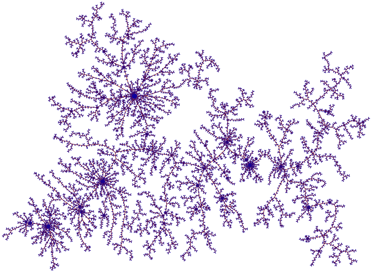
Suppose that the offspring distribution satisfies $p(k) \sim ck^{-1-\alpha}$ *as* $k \to \infty$ *for* $\alpha \in (1, 2)$*. Then as* $n \to \infty$*,*

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

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]

The stable trees

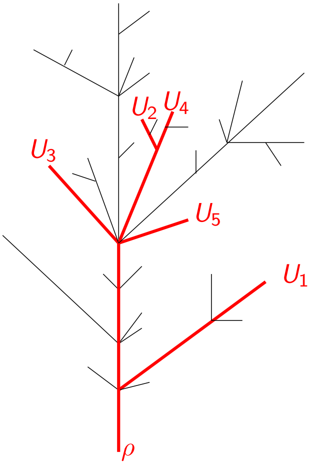
An important difference between the stable trees for $\alpha \in (1, 2)$ and the Brownian CRT is that the Brownian CRT is binary. The stable trees, on the other hand, have only branch-points of infinite degree.

A uniform measure

For $\alpha \in (1,2)$, the stable tree $\mathcal{T}_\alpha$ is again naturally endowed with a "uniform" probability measure μ_α, which is the push-forward of the Lebesgue measure on $[0,1]$ onto the tree. It is also the limit of the discrete uniform measure on T_n. As in the Brownian case, μ_α is supported by the set of leaves of $\mathcal{T}_\alpha$, and the law of the tree is invariant under random re-rooting according to μ_α.

Reduced trees

Let $U_1, U_2, \ldots$ be leaves sampled independently from $\mathcal{T}_\alpha$ according to μ_α, and let $\mathcal{T}_{\alpha,n}$ be the subtree spanned by the root ρ and $U_1, \ldots, U_n$:



Characterising the law of a stable tree

As usual, $\mathcal{T}_{\alpha,n}$ can be thought of in two parts: its tree-shape $T_{\alpha,n}$ (a rooted unordered tree with n labelled leaves) and its edge-lengths.

Characterising the law of a stable tree

As usual, $\mathcal{T}_{\alpha,n}$ can be thought of in two parts: its tree-shape $T_{\alpha,n}$ (a rooted unordered tree with n labelled leaves) and its edge-lengths.

Moreover, $\mathcal{T}_\alpha$ is the completion of $\bigcup_{n\geq 1} \mathcal{T}_{\alpha,n}$.

Line-breaking construction

We had that Aldous' line-breaking construction precisely gives the random finite-dimensional distributions for the Brownian CRT, i.e. if $\tilde{T}_n$ is the nth tree in the line-breaking construction, we have

$$(\tilde{\mathcal{T}}_n, n \geq 1) \stackrel{d}{=} (\mathcal{T}_{2,n}, n \geq 1)\,.$$

Line-breaking construction

We had that Aldous' line-breaking construction precisely gives the random finite-dimensional distributions for the Brownian CRT, i.e. if $\tilde{T}_n$ is the nth tree in the line-breaking construction, we have

$$(\tilde{\mathcal{T}}_n, n \geq 1) \stackrel{d}{=} (\mathcal{T}_{2,n}, n \geq 1)\,.$$

Question: does there exist a similar line-breaking construction for the stable trees with $\alpha \in (1, 2)$?

Line-breaking construction

We had that Aldous' line-breaking construction precisely gives the random finite-dimensional distributions for the Brownian CRT, i.e. if $\tilde{T}_n$ is the nth tree in the line-breaking construction, we have

$$(\tilde{\mathcal{T}}_n, n \geq 1) \stackrel{d}{=} (\mathcal{T}_{2,n}, n \geq 1)\,.$$

Question: does there exist a similar line-breaking construction for the stable trees with $\alpha \in (1, 2)$?

Answer: yes!

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

¡ Muchas gracias !