Warwick, 20th April 2009

Lecture 1: Random trees and their limits

Christina Goldschmidt

Entirely based on:
Random trees and applications by Jean-François Le Gall, *Probability Surveys* **2** (2005), pp. 245-311. (Very highly recommended. Any errors in this presentation are my responsibility!)

See also:
The continuum random tree I, II, III by David Aldous.

Discrete trees

In this lecture, we will be working with rooted, ordered trees.

Discrete trees

In this lecture, we will be working with rooted, ordered trees.

We will label the vertices by strings which indicate the genealogy. The root is labelled $\emptyset$. Its children are $1, 2, 3, \ldots$. The children of a vertex with label v are given the concatenated labels $v1, v2, v3, \ldots$, e.g. $121, 122, 123, \ldots$ for the children of 12.

Discrete trees

In this lecture, we will be working with rooted, ordered trees.

We will label the vertices by strings which indicate the genealogy. The root is labelled $\emptyset$. Its children are $1, 2, 3, \ldots$. The children of a vertex with label v are given the concatenated labels $v1, v2, v3, \ldots$, e.g. $121, 122, 123, \ldots$ for the children of 12.

The vertices whose labels have v as a prefix are the individuals descended from v.

Coding discrete trees

Consider a rooted ordered tree T, taken to be a set of vertices, since the edges are implied by the vertex-labels.

We will discuss three different encodings of T.

Height function

Suppose that T has n vertices. Let them be $v_0, v_1, \ldots, v_{n-1}$, listed in lexicographical order. Let $|v|$ be the distance of vertex v from the root.

Height function

Suppose that T has n vertices. Let them be $v_0, v_1, \ldots, v_{n-1}$, listed in lexicographical order. Let $|v|$ be the distance of vertex v from the root.

Then the height function is defined by

$$H(i) = |v_i|, \quad 0 \leq i \leq n-1.$$

Height function

Suppose that T has n vertices. Let them be $v_0, v_1, \dots, v_{n-1}$, listed in lexicographical order. Let $|v|$ be the distance of vertex v from the root.

Then the height function is defined by

$$H(i) = |v_i|, \quad 0 \leq i \leq n-1.$$

With a little thought, we see that we can recover the tree from its height function.

Contour function

Trace the "contour" of the tree from left to right at speed 1, so that we pass along each edge twice. Record the distance from the root at each time to get $(C(t), 0 \leq t \leq 2(n-1))$.

Contour function

Trace the "contour" of the tree from left to right at speed 1, so that we pass along each edge twice. Record the distance from the root at each time to get $(C(t), 0 \leq t \leq 2(n-1))$.

In this case, it's easy to see that the tree can be recovered from the contour function (just glue the sides back together).

Depth-first walk

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

Depth-first walk

Let $c(v)$ be the number of children of v, and recall that $v_0, v_1, \ldots, v_{n-1}$ is a list of the vertices of 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

Let $c(v)$ be the number of children of v, and recall that $v_0, v_1, \ldots, v_{n-1}$ is a list of the vertices of 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

It is much less clear that we can recover the tree from the depth-first walk.

Depth-first walk

It is much less clear that we can recover the tree from the depth-first walk.

In fact, for $0 \le i \le n-1$,

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

Galton-Watson trees

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

Galton-Watson trees

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

We will consider the case where the offspring distribution μ is critical or subcritical i.e.

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

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

Galton-Watson trees

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

We will consider the case where the offspring distribution μ is critical or subcritical i.e.

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

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

Since the tree is random, we will refer to the height and contour processes rather than functions.

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

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

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

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

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

Galton-Watson forest

It turns out to be technically easier to deal with a sequence of i.i.d. Galton-Watson trees rather than a single tree. We can concatenate their height processes in order to encode the whole Galton-Watson forest.

Galton-Watson forest

It turns out to be technically easier to deal with a sequence of i.i.d. Galton-Watson trees rather than a single tree. We can concatenate their height processes in order to encode the whole Galton-Watson forest.

For the depth-first walks, we retain the relation $X(i+1) = X(i) + c(v_i) - 1$, so that the first tree ends when the walk first hits -1, the second tree ends when we first hit -2 and so on.

Galton-Watson forest

It turns out to be technically easier to deal with a sequence of i.i.d. Galton-Watson trees rather than a single tree. We can concatenate their height processes in order to encode the whole Galton-Watson forest.

For the depth-first walks, we retain the relation $X(i+1) = X(i) + c(v_i) - 1$, so that the first tree ends when the walk first hits -1, the second tree ends when we first hit -2 and so on.

It can then be checked that we still have

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

Convergence of the depth-first walk

Now specialise to the case where μ is critical and has finite offspring variance $\sigma^2 > 0$.

Convergence of the depth-first walk

Now specialise to the case where μ is critical and has finite offspring variance $\sigma^2 > 0$.

Then $(X(k), k \geq 0)$ is a random walk with no drift and finite-variance step sizes.

Convergence of the depth-first walk

Now specialise to the case where μ is critical and has finite offspring variance $\sigma^2 > 0$.

Then $(X(k), k \geq 0)$ is a random walk with no drift and finite-variance step sizes.

Proposition. As $n \to \infty$,

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

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

Convergence of the height process

Theorem. As $n \to \infty$,

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

where $(\beta(t), t \geq 0)$ is a reflected Brownian motion.

Galton-Watson trees conditioned on their total progeny

Each excursion of the height process of the Galton-Watson forest corresponds to a tree, and the length of the excursion corresponds to the total progeny of that tree. If we condition on the total progeny of the tree to be n, and let $n \longrightarrow \infty$, intuitively we should obtain an excursion of the limit process.

Galton-Watson trees conditioned on their total progeny

Each excursion of the height process of the Galton-Watson forest corresponds to a tree, and the length of the excursion corresponds to the total progeny of that tree. If we condition on the total progeny of the tree to be n, and let $n \to \infty$, intuitively we should obtain an excursion of the limit process.

Let $(H^n(i), 0 \leq i \leq n)$ be the height process of such a conditioned tree.

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

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

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

In fact, more is true.

Theorem. (Marckert and Mokkadem (2003)) As $n \to \infty$,

$$(n^{-1/2}X^n(\lfloor n\cdot \rfloor), n^{-1/2}H^n(\lfloor n\cdot \rfloor), n^{-1/2}C^n(\lfloor 2n\cdot \rfloor)) \xrightarrow{d} \left(\sigma e, \frac{2}{\sigma} e, \frac{2}{\sigma} e\right),$$

All of these results suggest the existence of some sort of limiting tree, coded by the Brownian excursion.