STOR 831: Weak convergence and local weak convergence	Fall 2015
Lecture 8.3	
Lecture date: Nov 5/Nov 10	*Scribe: Jimmy Jin*

1 Random trees

In this lecture we visit an interesting application of $C[0,1]$ methodology developed previously. The objects of study here are special sorts of graphs called *trees*.

Definition 1 *A graph with no cycles is called a* ***tree****. Notably there is a unique path between any pair of vertices.*

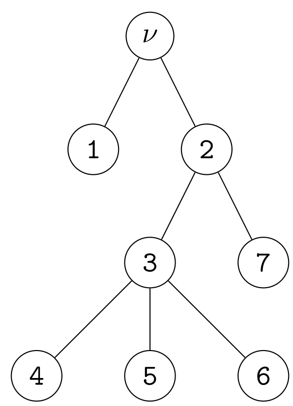


Figure 1: A tree with 8 *vertices and a root* ν*.*

In particular, we will work with *finite, rooted, ordered* trees. In other words, these trees have a finite number of vertices, have a distinguished vertex called the *root*, and have a particular ordering to the vertices. To be precise, first define the set of labels $\mathcal{U}$:

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

where by convention $\mathbb{N}^0 = \{\emptyset\}$. Therefore an element u of $\mathcal{U}$ is a *sequence* of elements of $\mathbb{N}$, i.e.

$$u = (u^1, u^2, \ldots, u^n)$$

We will define the **generation** of $u = (u^1, \ldots, u^n)$ by $|u| = n$. Also, we will need two important operations on elements of $\mathcal{U}$:

1. The **concatenation** of $u, v \in \mathcal{U}$ with $|u| = |v| = n$ by:

$$uv = (u^1, \ldots, u^n, v^1, \ldots, v^n)$$

2. The **parent** of an element u by the mapping $\pi : \mathcal{U} \setminus \emptyset \to \mathcal{U}$ defined by

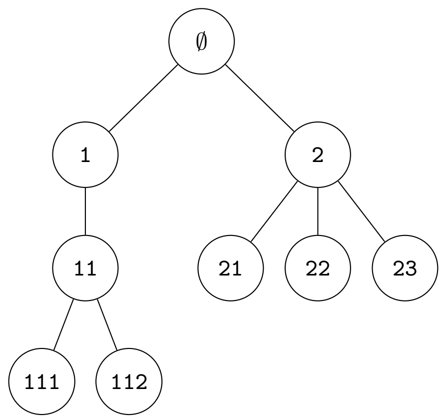
$$\pi(u) = \pi((u^1, \ldots, u^n)) = (u^1, \ldots, u^{n-1})$$

Definition 2 *A* ***finite, rooted, ordered tree*** $\mathcal{T}$ *is a finite subset of* $\mathcal{U}$ *satisfying:*

1. $\emptyset \in \mathcal{T}$
2. *If* $u \in \mathcal{T} \setminus \emptyset$, *then* $\pi(u) \in \mathcal{T}$
3. *For every* $u \in \mathcal{T}$, *there exists integer* $k_u(\mathcal{T}) \geq 0$ *such that for every* $j \in \mathbb{N}$, $uj \in \mathcal{T}$ *if and only if* $1 \leq j \leq k_u(\mathcal{T})$.

We interpret $k_u(\mathcal{T})$ *as the number of children of* u *in* $\mathcal{T}$. *Finally set* $\#(\mathcal{T})$ *to denote the total progeny in the tree* $\mathcal{T}$.

Figure 2: A rooted, ordered tree on 9 vertices.



A tree can be seen as coding a relational structure between the vertices (labels), so this definition essentially specifies a set of rules for the labels which, when followed, will guarantee the right tree structure amongst the labels.

Rule (1) specifies that $\emptyset$ will be considered the root vertex. Rule (2) specifies that if a vertex is in the tree, then it must have a predecessor which is also in the tree. Rule (3) specifies the exact form of the labels with encodes the ordering information. Note that the requirement that $k_u(\mathcal{T})$ be integer forces $\mathcal{T}$ to be finite. For later use, let:

$$\mathcal{A} = \text{ the collection of all finite, rooted, ordered trees}$$

2 Galton-Watson trees

Why do we care about trees? It turns out that trees are one of the most fundamental math objects in all of science. As alluded to above, trees encode a special type of relational structure which occurs in a wide range of settings.

The key for us is to study a certain kind of random tree. The general setup is to fix some $n \geq 1$ and consider a **random tree** $\mathcal{T}_n$ of size n (i.e. $|\mathcal{T}_n| = n$) according to some *distribution* on the space of trees. Then we will study its properties as $n \to \infty$.

The main example we will focus on is that of **Galton-Watson trees**. To do this, fix a distribution μ on $\mathbb{Z}^+ = \{0, 1, \dots, \}$. It turns out that the end behavior does not really depend on the mean of μ so we may assume without loss of generality[1] that

$$\sum_{k \geq 0} k\mu(\{k\}) = 1$$

We will generate a random tree $\mathcal{T}$ using the measure μ in the following way:

1. Start with one individual, the root (labelled $\emptyset$).
2. Every individual $\tilde{u}$ in the population has a random $\xi_{\tilde{u}}$ number of children, where $\xi_{\tilde{u}}$ are independent across individuals with distribution μ.

This puts us squarely in the realm of *finite*, rooted, ordered trees because of the following result from classical branching process theory:

Theorem 3 *If* $\sum_{k \geq 0} k\mu(\{k\}) = 1$*, then* $|\mathcal{T}| < \infty$ *almost surely.*

Up to this point we have not explicitly studied how this randomness in the number of offspring relates to the randomness on the space of trees. Let us illustrate this with an example.

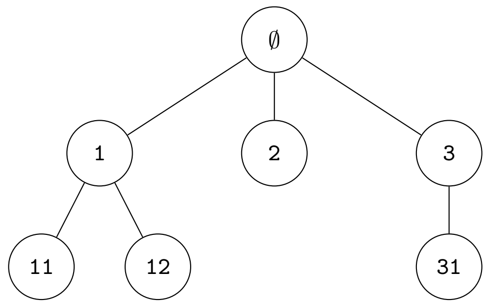
Example 4 *Suppose that we run the process described above and obtain:*

$$\xi_{(\emptyset)} = 3,\ \xi_{(1)} = 2,\ \xi_{(2)} = 0,\ \xi_{(3)} = 1,\ \xi_{(11)} = 0,\ \xi_{(12)} = 0,\ \xi_{(31)} = 0$$

The corresponding tree, call it $\tilde{t}_0$*, looks like:*

[1] At least as long as we stay in the realm of the exponential family—the full picture is slightly more nuanced.

Figure 3: The tree $\tilde{t}_0$ from Example 4.



Now let $\mathcal{T}$ be the random tree that one obtains using the offspring distribution μ, and let Π_μ be the associated distribution on $\mathcal{A}$, the space of trees. Then since the offspring are independent,

$$\Pi_\mu(\tilde{t}_0) = \mu(\{3\})\mu(\{2\})\mu(\{1\})\mu(\{0\})^4$$

In general for a realized tree $\tilde{t}$, this is:

$$\Pi_\mu(\tilde{t}) = \prod_{v\in\tilde{t}} \mu(\{d^+(v)\}) = \prod_{v\in\tilde{t}} \mu(\{k_v(\tilde{t})\})$$

where $d^+(v)$ is the **outdegree** of the vertex $v \in \tilde{t}$. Using this relation we can easily relate the offspring distribution μ to the distribution Π_μ on $\mathcal{A}$.

Example 5 *(GW tree with Geometric offspring)*

Suppose $\mu \sim$ Geometric$(1/2)$, i.e.

$$\mu(\{i\}) = \frac{1}{2^{i+1}}, \quad i \geq 0$$

Let $\mathcal{T} \sim \Pi_\mu(\cdot)$ and fix $\tilde{t} \in \mathcal{A}$. Note that there are exactly $|\tilde{t}| - 1$ children in $\tilde{t}$, so that

$$\Pi_\mu(\tilde{t}) = \prod_{v\in\tilde{t}} \frac{1}{2^{k_v(\tilde{t})+1}} = \left(\frac{1}{2}\right)^{2|\tilde{t}|-1}$$

Notably this measure depends only on the size of the tree and not the structure. This implies that if we fix some progeny size $n \geq 1$ and let $\mathcal{A}_n \subset \mathcal{A}$ be the trees of size n, i.e. $\mathcal{A}_n = \{\tilde{t} \in \mathcal{A} : |\tilde{t}| = n\}$, then the distribution of the conditioned tree $\mathcal{T}_n = \mathcal{T}|\{|\mathcal{T}| = n\}$ is uniform over $\mathcal{A}_n$, i.e.

$$\mathbb{P}_\mu(\mathcal{T}_n = \tilde{t}) = \frac{1}{|\mathcal{A}_n|}$$

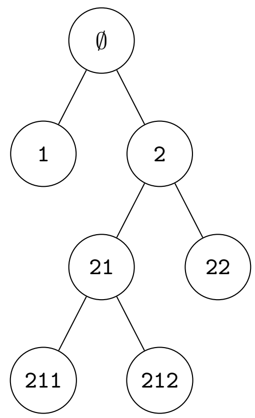
Example 6 *(GW tree with 0 or 2 offspring)*

If we take the measure μ *to be defined by*

$$\mu(\{0\}) = \mu(\{2\}) = \frac{1}{2}$$

then it is not difficult to see that Π_μ *is the uniform measure on the space of ordered, rooted, strictly binary trees.*

Figure 4: A strictly binary tree on 7 vertices.

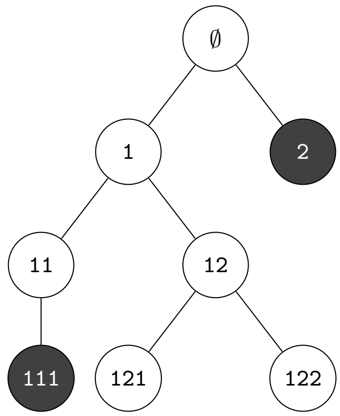


These types of trees are heavily used in phylogenetics, as speciation can be viewed as one species specializing into exactly two subspecies (iterating to get speciation into more than 2 subspecies).

To summarize, we have shown that specifying an offspring distribution in a Galton-Watson process induces a measure on a space of trees. We say that families of random trees which can be obtained via conditioning Galton-Watson processes on some fixed size are **simply generated trees.** Many (but not all) families of random trees belong to this class.

Eventually our goal will be to broaden our view of trees to that of metric spaces. In short any tree can be viewed as a metric space using the graph distance on the tree, where each edge has length 1:

Figure 5: The distance between nodes (111) *and* (2) *is 4.*



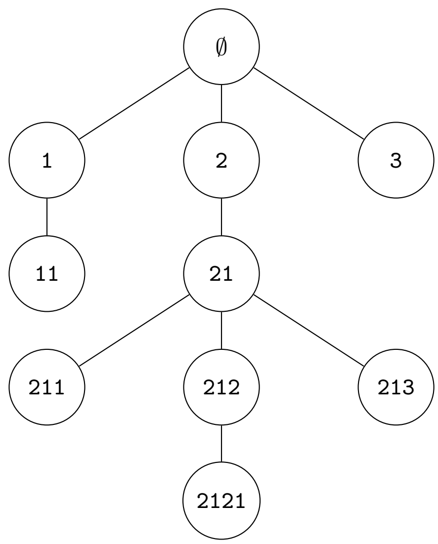
In general, the picture will be

- **Finite tree** $\longrightarrow$ compact metric space
- **Finite random tree** $\longrightarrow$ compact *random* metric space
- **Sequence of finite random trees** $\{\mathcal{T}_n : n \geq 1\}$ $\longrightarrow$ *sequence* of compact random metric spaces

3 Three different encodings of trees

To bring this discussion back to the realm of classical weak convergence, we would like a way to represent trees as functions in $C[0,1]$. Fortunately, there are three difference 1-1 maps from $\mathcal{A}$ to $C[0,1]$. In all that follows we will work with the tree:

Figure 6: The tree $\tilde{t}$.



I. Height function: List the vertices in "lexicographic" order as $\emptyset = u_0, u_1, \ldots$. So the vertices above would be

$$\emptyset, 1, 11, 2, 21, 211, 212, 2121, 213, 3$$

Then plot the height (i.e. generation #) = $|u_i|$ for each i in the above order. Call this function $h_{\tilde{t}}(\cdot)$:

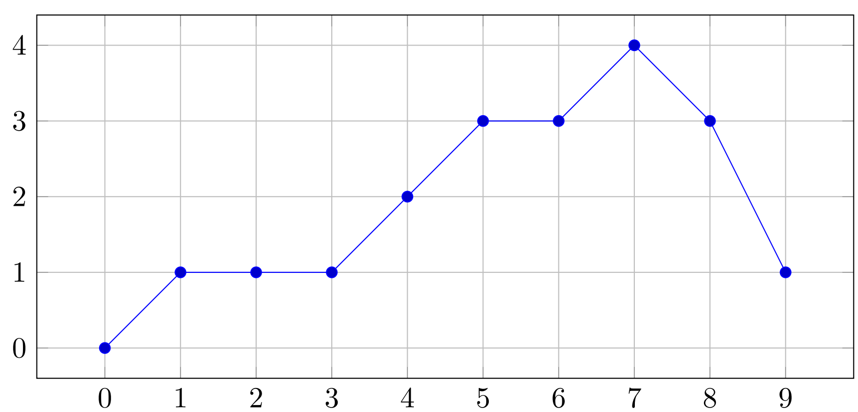
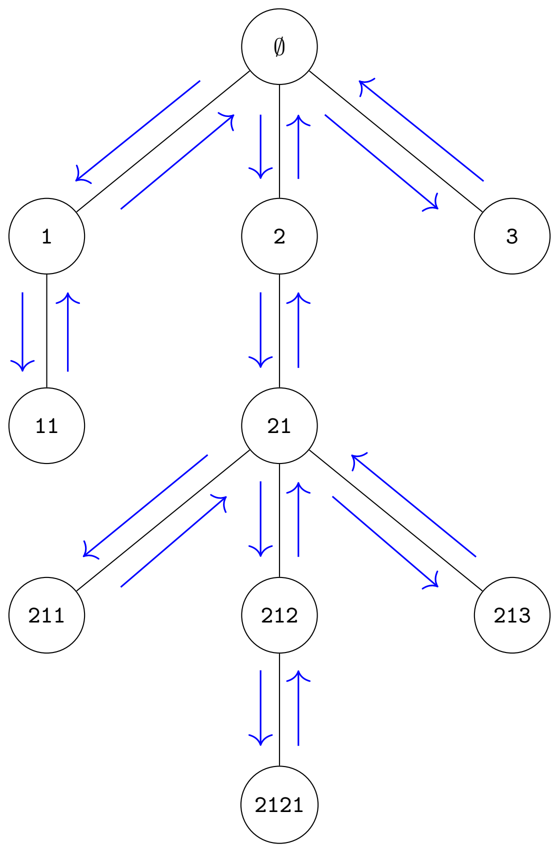


Figure 7: $h_{\tilde{t}}(\cdot)$, the height function of $\tilde{t}$

II. Contour function (Dyck path): Think of an ant exploring the tree starting from the root and crawling along the edges at rate 1 until the ant finally comes back to the root. The ant goes from left to right and explores in a depth-first fashion:

Figure 8: The exploration of the Contour function ant on $\tilde{t}$.



Plot the distance of the ant from the root at any given time and denote this path $c_{\tilde{t}}(\cdot)$. Note that every edge is traversed twice so the ant finishes exploration of the entire tree at time $(2\times \#\text{ edges in } \tilde{t}) = (2(|\tilde{t}|-1))$:

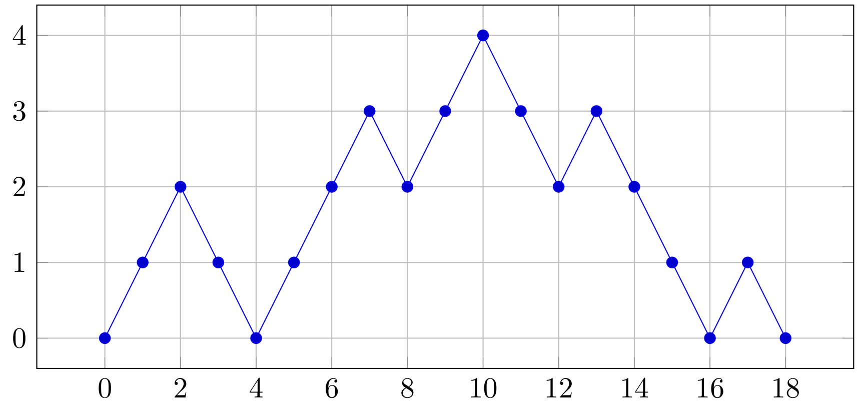


Figure 9: $c_{\tilde{t}}(\cdot)$, *the contour function of* $\tilde{t}$

III. Lukasiewicz path: This encoding is less pictorial than the previous 2 encodings but much more tractable. Recall the order used when constructing the height function $\emptyset = u_0, u_1, \ldots, u_{|\tilde{t}|-1}$.

Consider the map Φ which keeps track of the number of *children* as we explore the tree sequentially:

$$\Phi : \tilde{t} \to (k_{u_0}(\tilde{t}), k_{u_1}(\tilde{t}), \ldots, k_{u_{|\tilde{t}|-1}}(\tilde{t})$$

For our tree this map would give us the list

$$(3, 1, 0, 1, 3, 0, 1, 0, 0, 0)$$

We do not plot the picture of this list because it will turn out that a certain other function of $\Phi(\tilde{t})$ will turn out to be the right formulation for our purposes.

4 A random walk in the Lukasiewicz path

The gist of this entire setup is that the these encodings converge to various functionals of Brownian motion as $n \to \infty$. At this point we understand fairly well how well-behaved random walks converge to Brownian motion. Therefore to understand the convergence of these encodings to functionals of Brownian motion, we will show how one can find a random walk inside these encodings.

We will work mostly the Lukasiewicz path and the height function. We start with a fundamental fact about the Lukasiewicz path.

Let $\mathcal{S}$ = the collection of all finite sequences of non-negative integers $(m_1, \ldots, m_p), p \geq 1$ such that:

$$m_1 + \ldots + m_i \geq i \quad \text{and} \quad m_1 + \ldots + m_p = p - 1$$

Proposition 7 *Φ is a 1-1 map from $\mathcal{A}$ to $\mathcal{S}^2$.*

Proof: omitted, see Le Gall's survey for details. ∎

In light of this fact now let $\tilde{t} \in \mathcal{A}$ with $p = |\tilde{t}|$ and suppose $\Phi(\tilde{t}) = (m_1, \dots, m_p)$. Define the "walk" $\{x_k\}_{0 \le k \le p}$ by $x_0 = 0$ and:

$$x_l = \sum_{i=1}^{l} (m_i - 1), \quad o \le l \le p$$

Proposition 8 *The walk $\{x_k\}$ satisfies the following:*

1. $x_0 = 0$, $x_p = -1$
2. $x_n \ge 0$ *for all* $0 \le n \le p - 1$
3. $x_i - x_{i-1} \ge -1$ *for all* $1 \le i \le p$

In English, this means that the walk starts at 0 and ends at -1, is non-negative for all time before $p - 1$, and never decreases by more than 1 in any given time step.

Since the height function and Lukasiewicz path are both 1-1 maps from the space of trees, it is perhaps unsurprising that the height function can be recovered from this walk:

Theorem 9 *The height function of a tree is given by:*

$$h_{\tilde{t}}(k) = \#\Big\{ j \in \{0, 1, \dots, l-1\} : x_j = \inf_{j \le l \le k} x_l \Big\}, \quad 0 \le k \le p \qquad (\star)$$

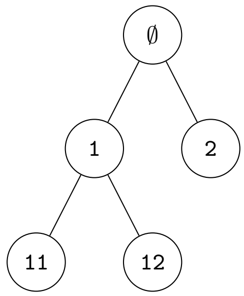
In other words, the height function at time k is the number of time points j before k where x_j was the minimum value of the walk in time from j until k. To see why this is true we give a "proof" by picture:

Proof via picture:

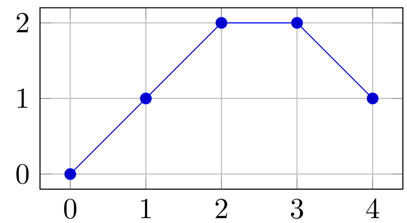
Consider the following tree and height function:

[2] For a proof of this fact, see Le Gall, *Random Trees and Applications*, Proposition 1.1.

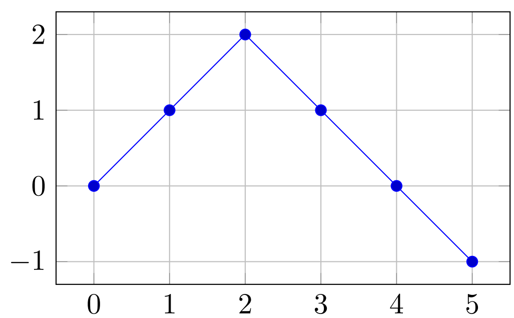
(a) The tree $\tilde{t}$



(b) The height function $h_{\tilde{t}}(\cdot)$



Then Φ is given by $\Phi(\tilde{t}) = (2, 2, 0, 0, 0)$ and the "$m_i - 1$" sequence $= (1, 1, -1, -1, -1)$. The associated Lukasiewicz path is:



Then following the formula $(\star)$ one can see that it matches the height function above. ■

So far everything has been deterministic, but now we show precisely where a random walk arises in this context. Let μ be a critical offspring distribution. The following algorithm generates a random Galton-Watson tree $\mathcal{T}$ with offspring distribution μ:

1. Generate $M_1, M_2, \ldots$ iid with distribution μ
2. Let $T = \inf\{p \geq 1 : M_1 + \ldots + M_p < p\}$
3. Let $\mathcal{T} =$ the unique tree with

$$\Phi(\mathcal{T}) = (M_1, M_2, \ldots, M_T)$$

The definition of T is enforced because otherwise the sequence $(M_1, \ldots, M_T)$ may not be in $\mathcal{S}$, the target space of the mapping Φ. In other words T guarantees that there exists a tree which will correspond to the sequence $(M_1, \ldots, M_T)$.

Example 10 *Suppose the sequence $\{M_k\}_{k\geq 1}$ is generated and given by*

$$(3, 0, 2, 0, 0, 0, 0, 5, 1, 2)$$

The partial sums of this sequence are:

p	*1*	*2*	*3*	*4*	*5*	*6*
sum	*3*	*3*	*5*	*5*	*5*	*5*

So in this case $T = 6$. Therefore the Lukasiewicz path and corresponding tree are

$$\Phi(\mathcal{T}) = (3, 0, 2, 0, 0, 0)$$

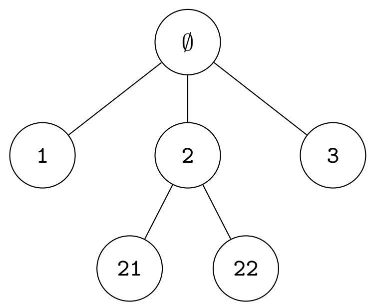


Figure 11: The tree corresponding to $\Phi(\mathcal{T}) = (3, 0, 2, 0, 0, 0)$.

Now use the slightly shifted sequence $(M_1 - 1, M_2 - 1, \ldots)$. Write $Y_j = M_j - 1$ and view these as increments of a random walk process. More precisely, let $\{S_k\}_{k\geq 1}$ be the process defined by $S_0 = 0$ and

$$S_j = \sum_{i=1}^{j} Y_i, \quad j \geq 1$$

Then clearly $\{S_k\}_{k\geq 1}$ is a random walk with increment distribution ν defined in terms of the original offspring distribution μ, by:

$$\nu(k) = \mu(k+1), \quad k \geq -1$$

Note in particular that, if μ has mean 1, then ν has mean zero so we are in a very classical regime. This finally shows that:

Theorem 11 *If $T = \inf\{n \geq 1 : S_n = -1\}$, then the Lukasiewicz path of a Galton-Watson tree with offspring distribution μ has the same distribution as the above random walk process up to T:*

$$(S_0, S_1, \ldots, S_T)$$