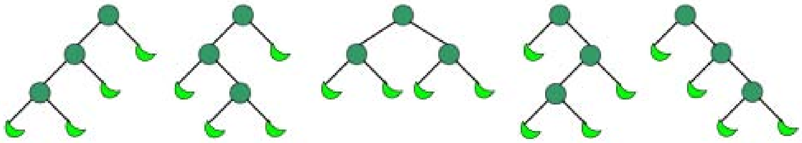
Doob-Martin boundary of Rémy's tree growth chain

Steven N. Evans

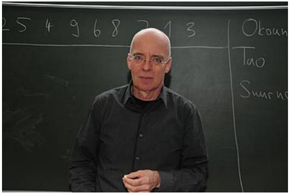
Department of Mathematics & Department of Statistics
Group in Computational and Genomic Biology
Group in Computational Science and Engineering
University of California at Berkeley

August, 2013



Collaborators

Rudolf Grübel
Hannover



Anton Wakolbinger
Frankfurt

Binary trees

- Write $\{0,1\}^\star := \bigsqcup_{k=0}^{\infty} \{0,1\}^k$ for the set of finite words drawn from the alphabet $\{0,1\}$ (with the empty word $\emptyset$ allowed).
- A binary tree is a finite subset $\mathbf{t} \subset \{0,1\}^\star$ with the properties:
  - $v_1 \ldots v_k \in \mathbf{t} \Longrightarrow v_1 \ldots v_{k-1} \in \mathbf{t}$,
  - $v_1 \ldots v_k 0 \in \mathbf{t} \Longleftrightarrow v_1 \ldots v_k 1 \in \mathbf{t}$.

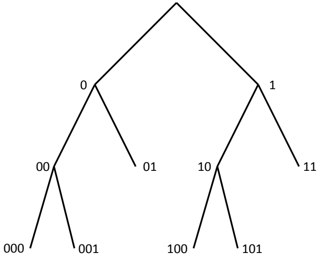


Figure: A binary tree is just a finite rooted tree in which every individual has zero or two children and we can distinguish left from right.

Counting binary trees

- Call the empty word $\emptyset$ the root of the tree.
- A binary tree has $2n + 1$ vertices for some $n \in \mathbb{N}$: $n + 1$ leaves and n interior vertices.
- The number of binary trees with $2n + 1$ vertices is the Catalan number $C_n := \frac{1}{n+1}\binom{2n}{n}$.

Rémy's (1985) algorithm generates a sequence of random binary trees $T_1, T_2, \ldots$ such that T_n is uniformly distributed on the set of binary trees with $2n+1$ vertices.

- Start with T_1 being the unique binary tree $\aleph := \{\emptyset, 0, 1\}$ with 3 vertices.
- Supposing that T_n has been generated, pick a vertex v uniformly at random.
- Cut off the subtree rooted at v and set it aside.
- Attach a copy of the tree $\aleph$ with 3 vertices to the end of the edge that previously led to v.
- Re-attach the subtree rooted at v uniformly at random to one of the two leaves in the copy of $\aleph$.
- Call the two new vertices that have been produced clones of v.

Example of one iteration of Rémy's algorithm

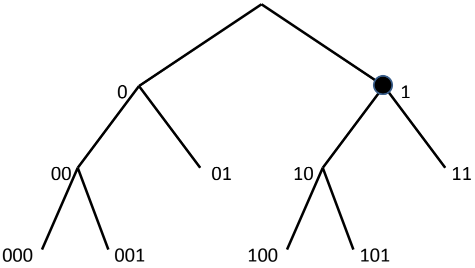


Figure: First step in an iteration of Rémy's algorithm: pick a vertex v uniformly at random.

Example of one iteration of Rémy's algorithm – continued

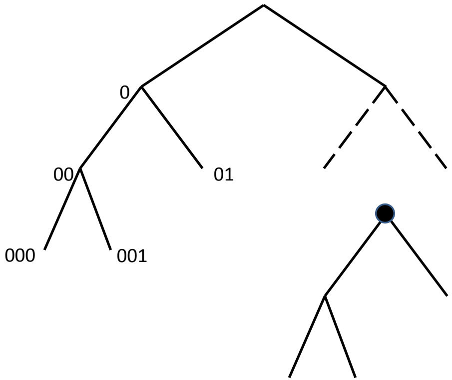


Figure: Second step in an iteration of Rémy's algorithm: cut off the subtree rooted at v and attach a copy of $\aleph$ to the end of the edge that previously led to v.

Example of one iteration of Rémy's algorithm – continued

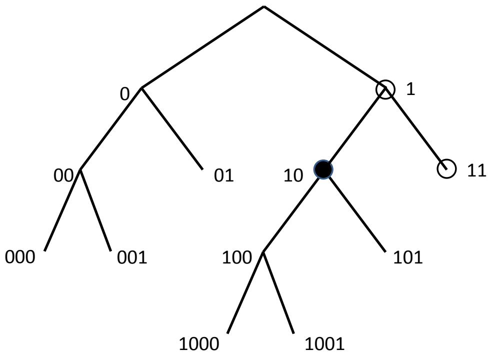


Figure: Third step in an iteration of Rémy's algorithm: re-attach the subtree rooted at v to one of the two leaves of the copy of $\aleph$, and re-label the vertices appropriately. The solid circle is the new location of v and the open circles are the clones of v.

Connections with Aldous' Brownian CRT

Marchal (2003) showed that the Rémy trees thought of as real trees with unit edge lengths converge almost surely in some sense to Aldous' Brownian continuum random tree after suitable rescaling.

Conversely, Le Gall (1999) showed that if one successively samples points in a conditionally independent manner from the CRT using the associated mass measure on the leaves and thinks of the trees induced by the sampled leaves and the root as (combinatorial) binary trees, then the resulting process is Rémy's chain.

It follows from Hewitt-Savage that the CRT generates the tail σ-field of the Rémy chain up to null sets. In other words, the CRT is the Poisson boundary of the Rémy chain.

What are the multi-step transition probabilities?

- Condition on T_m.
- Say that a vertex of T_{m+n} is a clonal descendant of a vertex $v \in T_m$ if it is v itself, a clone of v, a clone-of-a-clone of v, etc.
- We can decompose T_{m+n} into connected pieces according to clonal descent from the vertices of T_m.

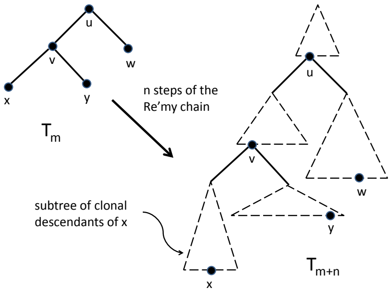


Figure: Decomposition of T_{m+n} via clonal descent from the vertices of T_m.

What are the multi-step transition probabilities? – continued

- The numbers of clonal descendants of the $2m+1$ vertices is the result of n steps in a Polya urn that starts with $2m+1$ balls of different colors and at each stage a ball is chosen uniformly at random and replaced along with two balls of the same color.
- Conditional on the numbers of clonal descendants, the binary trees of clonal descendants are independent and uniformly distributed.
- Conditional on the trees of clonal descendants, the ancestors from T_m are located at independently and uniformly chosen leaves of their respective trees of clonal descendants.

What are the multi-step transition probabilities? – continued

- Label the vertices of $T_m = \mathbf{s}$ with $1, \ldots, 2m+1$.
- The probability of evolving to $T_{m+n} = \mathbf{t}$ enhanced with a particular clonal descent decomposition is

$$\frac{n!}{n_1! \cdots n_{2m+1}!} \frac{\prod_{j=1}^{2m+1}[1 \times 3 \times \cdots \times (2n_j - 1)]}{(2m+1) \times (2m+3) \times \cdots \times (2(m+n)-1)}$$
$$\times \prod_{j=1}^{2m+1} \frac{1}{C_{n_j}} \prod_{j=1}^{2m+1} \frac{1}{n_j + 1}$$
$$= \frac{n!}{(2m+1) \times (2m+3) \times \cdots \times (2(m+n)-1)} \frac{1}{2^n}.$$

What are the multi-step transition probabilities? – continued

The probability $p(\mathbf{s}, \mathbf{t})$ of transitioning from $\mathbf{s}$ to $\mathbf{t}$ is thus

$$\frac{n!}{(2m+1) \times (2m+3) \times \cdots \times (2(m+n)-1)} \frac{1}{2^n} N(\mathbf{s}, \mathbf{t}),$$

where $N(\mathbf{s}, \mathbf{t})$ is the number of ways of embedding $\mathbf{s}$ into $\mathbf{t}$ such that:

- Leaves are mapped to leaves.
- If u, v are vertices of $\mathbf{s}$ such that v is below and to the left (resp. right) of u, then the image of v in $\mathbf{t}$ is below and to the left (resp. right) of the image of u in $\mathbf{t}$.

A remark about $N(\mathbf{s}, \mathbf{t})$

Note that an embedding of $\mathbf{s}$ into $\mathbf{t}$ is determined by the images of the leaves of $\mathbf{s}$, and so $N(\mathbf{s}, \mathbf{t})$ is just the number of subsets of cardinality $m+1$ drawn from the $n+1$ leaves of $\mathbf{t}$ such that the tree induced by the chosen leaves is isomorphic to $\mathbf{s}$.

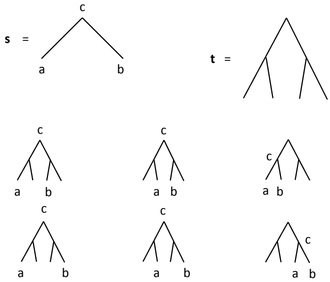


Figure: All the embeddings of the unique binary tree $\mathbf{s}$ with 3 vertices into a particular tree $\mathbf{t}$ with 7 vertices.

Doob-Martin kernel

- Recall that $\aleph$ is the binary tree with 3 vertices.
- If $\mathbf{s}$ and $\mathbf{t}$ are binary trees with $2m+1$ and $2(m+n)+1$ leaves, then the corresponding Doob-Martin kernel is

$$\begin{aligned} K(\mathbf{s},\mathbf{t}) &:= \frac{p(\mathbf{s},\mathbf{t})}{p(\aleph,\mathbf{t})} \\ &= \frac{1}{\mathbb{P}\{T_m = \mathbf{s}\}} \mathbb{P}\{T_m = \mathbf{s} \,|\, T_{m+n} = \mathbf{t}\} \end{aligned}$$

- If m is fixed, and $n \to \infty$, then

$$K(\mathbf{s},\mathbf{t}) \sim 2^m (1 \times 3 \times \cdots \times (2m-1)) \frac{1}{n^{m+1}} N(\mathbf{s},\mathbf{t}).$$

- A sequence $(\mathbf{t}_k)_{k\in\mathbb{N}}$ of binary trees converges in the Doob-Martin topology if $\lim_{k\to\infty} K(\mathbf{s},\mathbf{t}_k)$ exists for all binary trees $\mathbf{s}$.

Why do we care?

- We obtain an interesting compactification of the space of binary trees that contains information about different ways in which a sequence of trees can "go to infinity".
- All positive harmonic functions of the Rémy chain are positive linear combinations of functions of the form $\mathbf{s} \mapsto \lim_{k\to\infty} K(\mathbf{s}, \mathbf{t}_k)$.
- We understand all the ways it is possible to condition the Rémy chain to "do something at infinity" (conditioning $\iff$ Doob h-transforms $\iff$ positive harmonic functions).
- There is an interesting connection with the recent theory of graph limits developed by Lovász, Szegedy, Borgs, Chayes, Sós, Vesztergombi, Diaconis, Janson, Tao, Austin, ...

Doob-Martin convergence and bridges

- Given a binary tree $\mathbf{t}$ with $2M(\mathbf{t})+1$ vertices, write $T_1^{\mathbf{t}}, \ldots, T_{M(\mathbf{t})}^{\mathbf{t}}$ for the bridge obtained by conditioning the Rémy chain (started at $\aleph$) to hit $\mathbf{t}$ at time $M(\mathbf{t})$.
- It follows from a general remark of Föllmer (1975) that $(\mathbf{t}_k)_{k\in\mathbb{N}}$ with $M(\mathbf{t}_k) \to \infty$ converges in the Doob-Martin topology if for each $\ell \in \mathbb{N}$ the random ℓ-tuple $(T_1^{\mathbf{t}_k}, \ldots, T_\ell^{\mathbf{t}_k})$ converges in distribution (i.e. initial segments of the bridge to $\mathbf{t}_k$ converge in distribution).
- The limits define an infinite bridge $(T_n^\infty)_{n\in\mathbb{N}}$ with $T_1^\infty = \aleph$.

Doob-Martin convergence and bridges – continued

- Note that if $\mathbf{s}, \mathbf{t}$ are binary trees with $2m+1$ and $2m+3$ vertices, respectively, then

$$\begin{aligned}
\mathbb{P}\{T_m^{\mathbf{t}_k} = \mathbf{s} \,|\, T_{m+1}^{\mathbf{t}_k} = \mathbf{t}\} &= \frac{p(\aleph, \mathbf{s}) p(\mathbf{s}, \mathbf{t}) p(\mathbf{t}, \mathbf{t}_k)}{p(\aleph, \mathbf{t}) p(\mathbf{t}, \mathbf{t}_k)} \\
&= C_m^{-1} \frac{1}{2m+1} \frac{1}{2} N(\mathbf{s}, \mathbf{t}) / C_{m+1}^{-1} \\
&= \frac{(m+1)! m!}{(2m)!} \frac{1}{2m+1} \frac{1}{2} N(\mathbf{s}, \mathbf{t}) \frac{(2(m+1))!}{(m+2)!(m+1)!} \\
&= \frac{1}{m+2} N(\mathbf{s}, \mathbf{t}).
\end{aligned}$$

- Therefore, any limit bridge evolves backwards in time as follows:
  - Pick a leaf uniformly at random.
  - Delete the chosen leaf and its sibling.
  - Close up the gap if there is one.
- To understand the Doob-Martin compactification we need to understand all processes $(T_n^\infty)_{n \in \mathbb{N}}$ with $T_1^\infty = \aleph$ that have this description.

A simple example

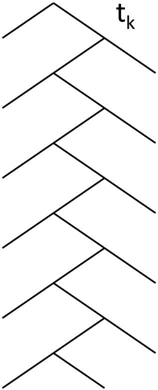


Figure: The binary tree $\mathbf{t}_k$ has $2k+1$ vertices and consists of a single spine with leaves hanging off to the left and right alternately.

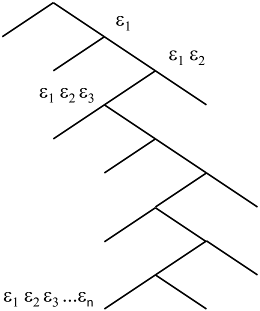


Figure: The value at time n of the infinite bridge arising from the sequence of trees depicted in Figure 7. The tree consists of leaves hanging of a single spine that moves to the left or right according to successive tosses of a fair coin.

Another example

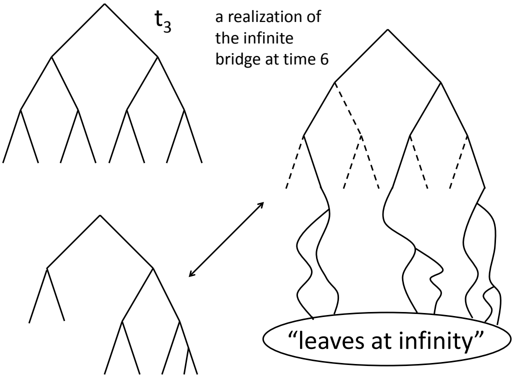


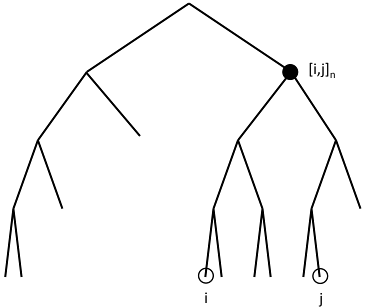
Figure: If $\mathbf{t}_k$ is the complete binary tree with 2^k leaves, then $\lim_k \mathbf{t}_k$ exists in the Doob-Martin topology and the resulting infinite bridge at time n can be built by choosing $n+1$ points uniformly from the leaves at infinity of the infinite complete binary tree.

Given an infinite bridge $(T_n^\infty)_{n\in\mathbb{N}}$, it is possible (by Kolmogorov consistency) to label the $n+1$ leaves of T_n^∞ with $\{1,\dots,n+1\}$ so that the following hold.

- All labelings are equally likely.
- In passing from T_{n+1}^∞ to T_n^∞:
  - The leaf labeled $n+2$ is deleted, along with its sibling.
  - If the sibling of the leaf labeled $n+2$ is also a leaf, then the common parent (which is now a leaf) is assigned the sibling's label.

Most recent common ancestors

- We want to use the labeling to build an infinite binary-tree-like structure for which $\mathbb{N}$ plays the role of the leaves.
- If $i, j \in \mathbb{N}$ are the labels of two leaves T_n^∞ that are represented as the words $u_1 \dots u_k$ and $v_1 \dots v_\ell$ in $\{0,1\}^\star$, then set $[i,j]_n := u_1 \dots u_m = v_1 \dots v_m$, where $m := \max\{h : u_h = v_h\}$.
- That is, $[i,j]_n$ is the most recent common ancestor in T_n^∞ of the leaves labeled i and j.



Interior vertices

- Define an equivalence relation $\equiv$ on the Cartesian product $\mathbb{N} \times \mathbb{N}$ by declaring that $(i', j') \equiv (i'', j'')$ if and only if $[i', j']_n = [i'', j'']_n$ for some (and hence all) n such that $i', j', i'', j'' \in [n+1]$.
- Write $\langle i, j \rangle$ for the equivalence class of the pair (i, j).
- Think of $\langle i, j \rangle$ as the being the most recent common ancestor of the leaves i and j and of such points being interior vertices of a tree-like object.

Ordering subtrees

- Define a partial order $<_L$ on the set of equivalence classes by declaring for $(i', j'), (i'', j'') \in \mathbb{N} \times \mathbb{N}$ that $\langle i', j' \rangle <_L \langle i'', j'' \rangle$ if and only if for some (and hence all) n such that $i', j', i'', j'' \in [n+1]$ we have $[i', j']_n = u_1 \dots u_k$ and $[i'', j'']_n = u_1 \dots u_k 0 v_1 \dots v_\ell$ for some $u_1, \dots, u_k, v_1, \dots, v_\ell \in \{0, 1\}$.
- Interpret the ordering $\langle i', j' \rangle <_L \langle i'', j'' \rangle$ as the "vertex" $\langle i'', j'' \rangle$ being below and to the left of the "vertex" $\langle i', j' \rangle$.

Ordering subtrees – continued

- Similarly, define another partial order $<_R$ by declaring that $\langle i', j' \rangle <_R \langle i'', j'' \rangle$ if and only if for some (and hence all) n such that $i', j', i'', j'' \in [n+1]$ we have $[i', j']_n = u_1 \ldots u_k$ and $[i'', j'']_n = u_1 \ldots u_k 1 v_1 \ldots v_\ell$ for some $u_1, \ldots, u_k, v_1, \ldots, v_\ell \in \{0, 1\}$.
- Interpret the ordering $\langle i', j' \rangle <_R \langle i'', j'' \rangle$ as the "vertex" $\langle i'', j'' \rangle$ being below and to the right of the "vertex" $\langle i', j' \rangle$.

Yet another partial order

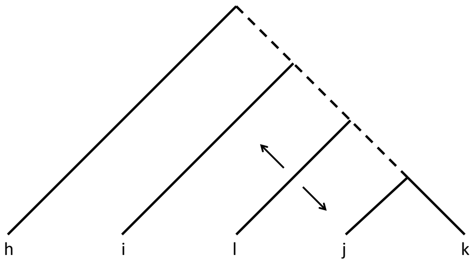
- Define a third partial order $<$ on the set of equivalence classes of $\mathbb{N} \times \mathbb{N}$ by declaring that $\langle i', j' \rangle < \langle i'', j'' \rangle$ if either $\langle i', j' \rangle <_L \langle i'', j'' \rangle$ or $\langle i', j' \rangle <_R \langle i'', j'' \rangle$.
- Interpret the ordering $\langle i', j' \rangle < \langle i'', j'' \rangle$ as the "vertex" $\langle i'', j'' \rangle$ being below the "vertex" $\langle i', j' \rangle$.

Most recent common ancestors

Any two equivalence classes $\langle h, i\rangle$ and $\langle j, k\rangle$ have a unique most recent common ancestor $\langle h, i\rangle \wedge \langle j, k\rangle$: the element x of $\{\langle h, i\rangle, \langle h, j\rangle, \langle h, k\rangle, \langle i, j\rangle, \langle i, k\rangle, \langle j, k\rangle\}$ such that $x \leq \langle h, i\rangle$ and $x \leq \langle j, k\rangle$.

Assigning distances

- For $\ell \in \mathbb{N}$, write write the equivalence class $\langle \ell, \ell \rangle$ as just ℓ.
- For $h, i, j, k \in \mathbb{N}$, such that $\langle h, i \rangle < \langle j, k \rangle$, set $I_\ell := \mathbb{1}\{\langle h, i \rangle \leq \langle j, k \rangle \wedge \ell\}$ for $\ell \in \mathbb{N} \setminus \{h, i, j, k\}$.
- The sequence of random variables $(I_\ell)_{\ell \notin \{h,i,j,k\}}$ is exchangeable.



Assigning distances – continued

- Recall we are assuming $\langle h, i\rangle < \langle j, k\rangle$.
- By de Finetti's theorem and the strong law of large numbers,

$$d(\langle h, i\rangle, \langle j, k\rangle) := \lim_{n\to\infty} \frac{1}{n} \sum_{1\le \ell\le n,\, \ell\notin\{h,i,j,k\}} I_\ell$$

  exists almost surely.
- Extend the definition to general pairs $\langle h, i\rangle$, $\langle j, k\rangle$ by

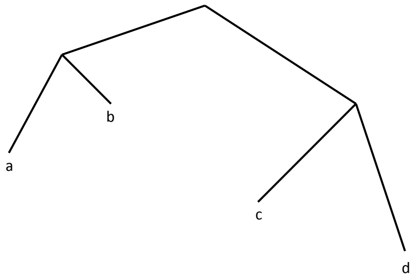
$$d(\langle h, i\rangle, \langle j, k\rangle) := d(\langle h, i\rangle, \langle h, i\rangle \wedge \langle j, k\rangle) + d(\langle h, i\rangle \wedge \langle j, k\rangle, \langle j, k\rangle).$$

Constructing a real tree

Unfortunately, d may be be just a pseudo-metric on the equivalence classes, but it satisfies the four point condition

$$d(a,b) + d(c,d) \le [d(a,c) + d(b,d)] \vee [d(a,d) + d(b,c)],$$

so the equivalence classes embed into a unique, minimal, complete real tree $(\mathbf{T}, d)$ that is compact.



- There is a unique root $\rho \in \mathbf{T}$ that is the limit of $\bigwedge_{1 \le i < j \le n+1} \langle i, j \rangle$.
- There is a partial order $\prec$ on $\mathbf{T}$ given by $x \prec y$ if and only if $x \neq y$ and x is on the segment between ρ and y.
- If $\langle h, i \rangle \prec \langle j, k \rangle$, then $\langle h, i \rangle < \langle j, k \rangle$.
- Any two points $x, y \in \mathbf{T}$ have a unique most recent common ancestor $x \curlywedge y$, the furthest point z from ρ such that $\rho \preceq z \preceq x$ and $\rho \preceq z \preceq y$.
- For $i, j \in \mathbb{N}$, $i \curlywedge j = \langle i, j \rangle = i \wedge j$.

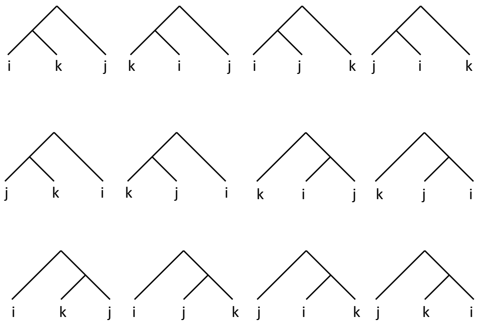
Building a probability measure on the real tree

Again by exchangeability, de Finetti, and the strong law of large numbers, there is a unique Borel probability measure μ on $\mathbf{T}$ such that

$$\mu\{y \in \mathbf{T} : x \prec y\} = \lim_{n \to \infty} \frac{1}{n} \#\{1 \le k \le n : x \prec k\}.$$

NOTE: The construction of $\mathbf{T}$, d and μ is superficially similar to a construction in a paper by Haulk & Pitman 2011 on de Finetti-like representations of exchangeable hierarchies. They proceed more concretely by building the real tree with its metric as a subset of ℓ^1, but applying their procedure in our setting doesn't always yield compact trees and the interpretation of their measure isn't as straightforward.

We have yet to incorporate the partial orders $<_L$ and $<_R$.
Observe that the order structures $<_L$ and $<_R$ on $\mathbb{N}$ are completely determined by a knowledge for all distinct i, j, k of the isomorphism class of the subtree spanned by i, j, k.



Triplet puzzling – continued

Consequently, the order structures $<_L$ and $<_R$ on $\mathbb{N}$ are completely determined by a knowledge for all distinct $i, j \in \mathbb{N}$ of the distances

$$d(i, i \wedge j) \text{ and } d(j, i \wedge j),$$

and whether

$$\langle i, j \rangle <_L i \text{ and } \langle i, j \rangle <_R j$$

or

$$\langle i, j \rangle <_R i \text{ and } \langle i, j \rangle <_R i.$$

An exchangeable array

- Write
$$I_{ij} := \mathbb{1}\{\langle i,j\rangle <_L i \,\&\, \langle i,j\rangle <_R j\}.$$
- The array of triplets
$$((d(i, i\wedge j), d(j, i\wedge j), I_{ij}))_{i,j\in\mathbb{N}}$$
is exchangeable.
- By the Aldous–Hoover–Kallenberg theory, there exists i.i.d. random variables U, $(U_i)_{i\in\mathbb{N}}$, and $(U_{ij})_{i,j\in\mathbb{N},\, i<j}$ that are uniform on $[0,1]$ and a function F such that
$$(d(i, i\wedge j), d(j, i\wedge j), I_{ij}) = F(U, U_i, U_j, U_{ij}),$$
where $U_{ij} = U_{ji}$ for $i > j$ (here $<$ is the usual order on $\mathbb{N}$).

- As usual, instances in which there is dependence on the r.v. U correspond to mixtures over extreme points in the set of infinite bridge distributions, so we suppose there is no dependence on U.
- In this case, the isomorphism type of the rooted, compact real tree $(\mathbf{T}, d)$ equipped with the probability measure μ is almost surely constant, and so we can treat $(\mathbf{T}, d, \mu)$ as a fixed real tree equipped with a fixed probability measure.

Representation by sampling

- If $X_1, X_2, \ldots$ are i.i.d. $\mathbf{T}$-valued random variables distributed as μ, then we can suppose that for some Borel bijection $\varphi : \mathbf{T} \to [0,1]$ we have $U_i = \varphi(X_i)$.
- If $I_{ij} = \mathbb{1}\{\langle i,j\rangle <_L i \;\&\; \langle i,j\rangle <_R j\}$ is representable as $\Psi(U_i, U_j, U_{ij})$, then $W : \mathbf{T} \times \mathbf{T} \to [0,1]$ defined by $W(x,y) = \mathbb{E}[\Psi(\varphi(x), \varphi(y), U_{ij})]$ has the properties:
  - $W(x,y) = 1 - W(y,x)$,
  - If z is a point whose deletion disconnects $\mathbf{T}$ into 3 components A, B, C, with A containing the root, then either:
    - $W(x,y) = 1$ for all $x \in B$ and $y \in C$,

    or
    - $W(x,y) = 0$ for all $x \in B$ and $y \in C$.
  - If x is a point whose deletion disconnects $\mathbf{T}$ into 2 components A, B, with A containing the root, then:
    - $W(x, \cdot)$ is constant on B.

A little more concretely ...

We can construct a realization of $\mathbb{N}$ with its partial orders $<_L$ and $<_R$, and hence the infinite bridge as follows.

- Pick i.i.d. $\mathbf{T}$-valued random variables $X_1, X_2, \ldots$ with common distribution μ.
- If the deletion of X_i disconnects $\mathbf{T}$ into 2 components, then toss a coin that comes up heads with probability the common value of $W(X_i, y)$ on the component not containing the root.
- If the coin comes up heads (resp. tails), declare $\langle i,j\rangle <_L i$ and $\langle i,j\rangle <_R j$ (resp. $\langle i,j\rangle <_R i$ and $\langle i,j\rangle <_L j$) for every j such that X_j falls into that component.
- If the deletion of neither X_i nor X_j disconnects $\mathbf{T}$, then $W(X_i, X_j) \in \{0,1\}$. If $W(X_i, X_j) = 1$ (resp. 0), then $\langle i,j\rangle <_L i$ & $\langle i,j\rangle <_R j$ (resp. $\langle i,j\rangle <_R i$ & $\langle i,j\rangle <_L j$).

Simplest example again

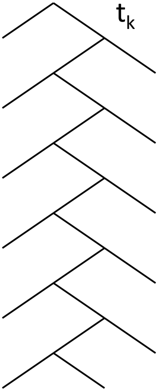


Figure: The binary tree $\mathbf{t}_k$ has $2k+1$ vertices and consists of a single spine with leaves hanging off to the left and right alternately.

Simplest example again – continued

Here we can take:

- $\mathbf{T}$ to be $[0,1]$ with d the usual metric,
- ρ to be 0,
- the partial order $\prec$ to be the usual order on $[0,1]$,
- $x \curlywedge y$ to be the usual minimum of x and y,
- μ to be Lebesgue measure.

In this case, $W(x,y) = \frac{1}{2}$ for $x \curlywedge y$.