Colloquium of the Institute of Discrete Mathematics, TU Graz

March 11, 2020

Infinite bridges for tree-valued Markov chains

Anton Wakolbinger

Institut für Mathematik, Goethe-Universität Frankfurt

Based on joint work with S. N. Evans (Berkeley) and R. Grübel (Hannover)

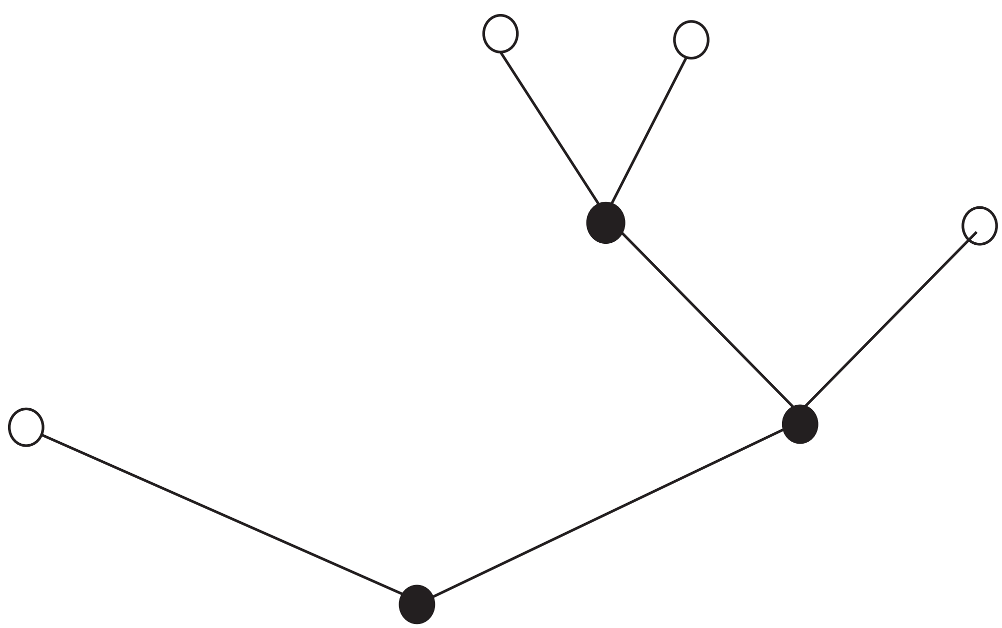
Rémy's tree growth chain

Jean-Luc Rémy (1985),

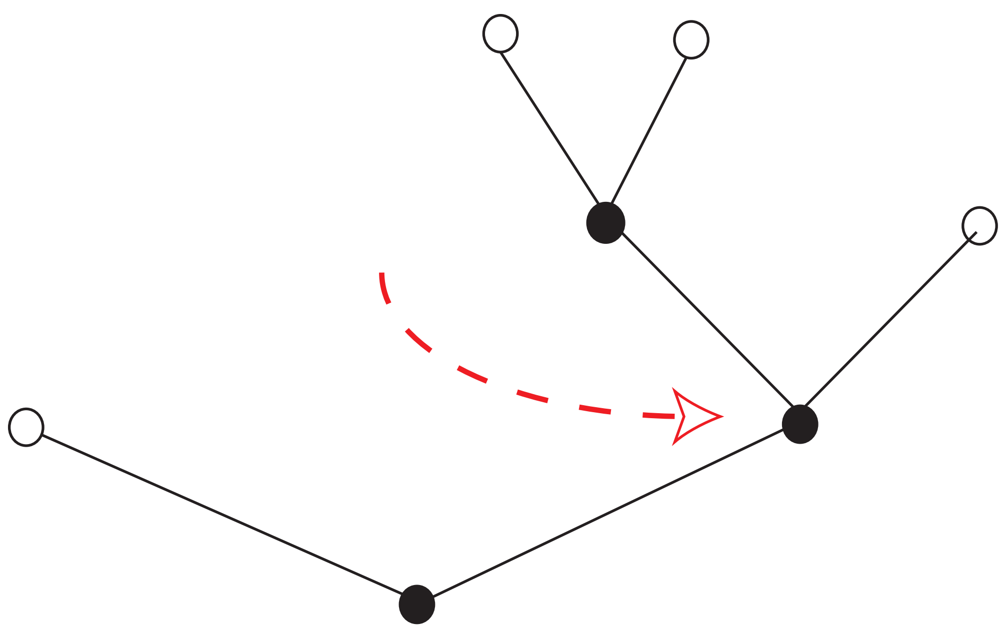
Un procédé iteratif de denobrement d'arbres binaires

et son application à leur generation aléatoire

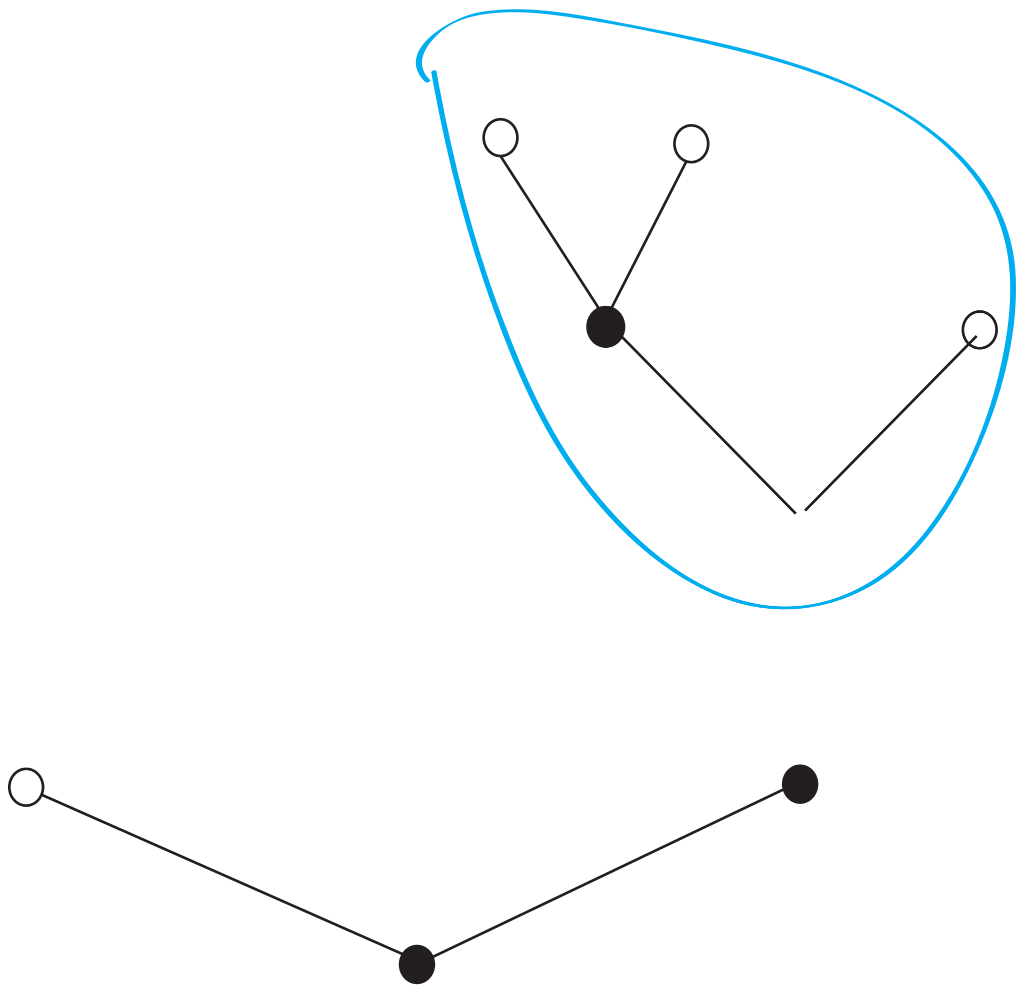
One step of Rémy's tree-valued Markov chain:



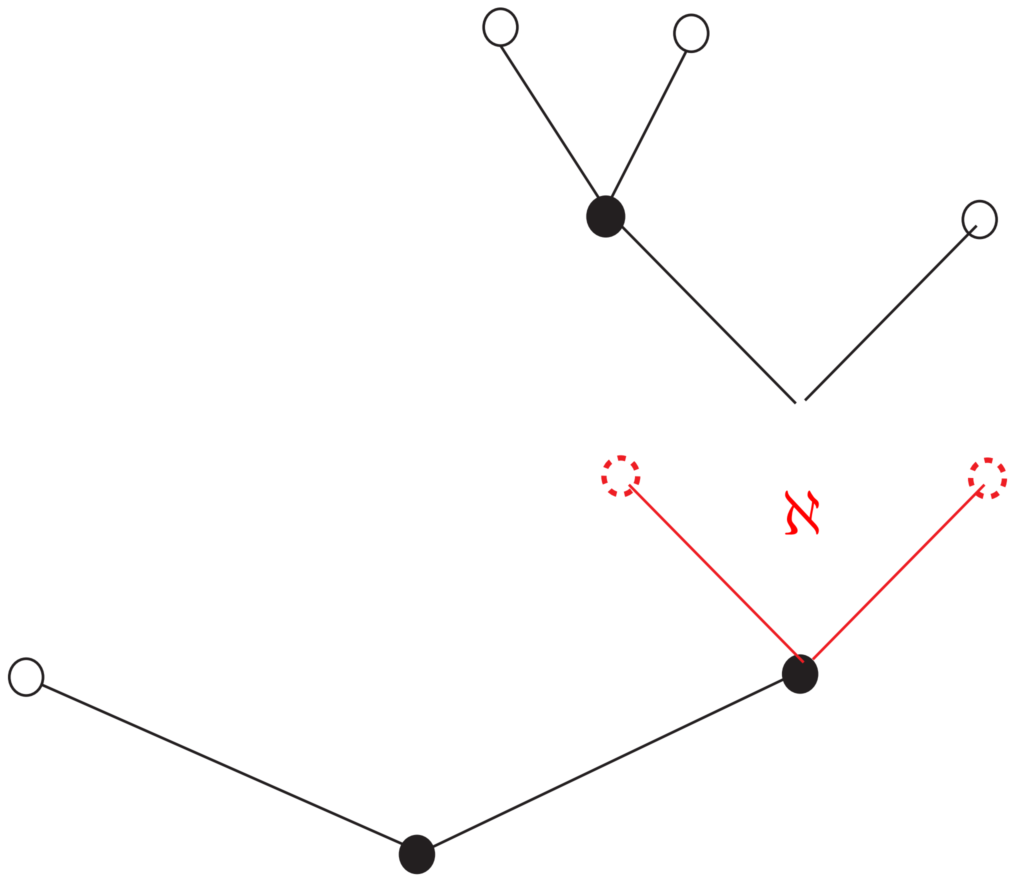
Choose a vertex at random



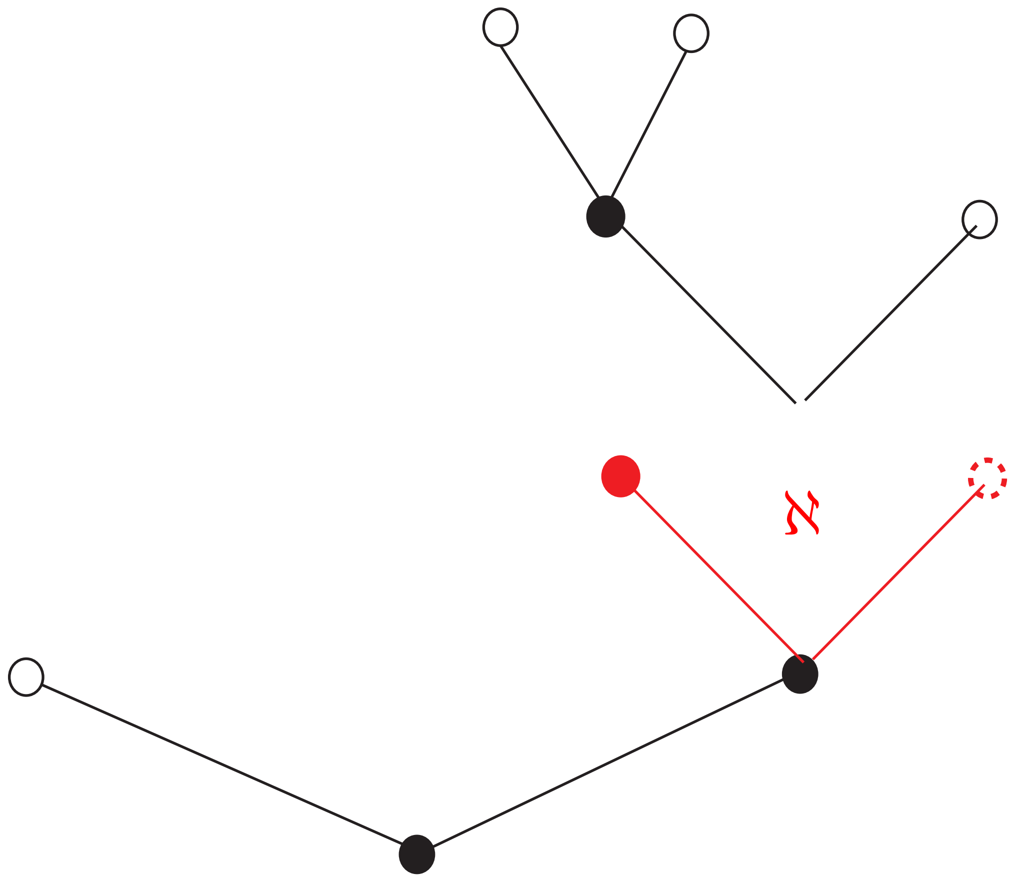
Put the offspring tree of the chosen vertex aside



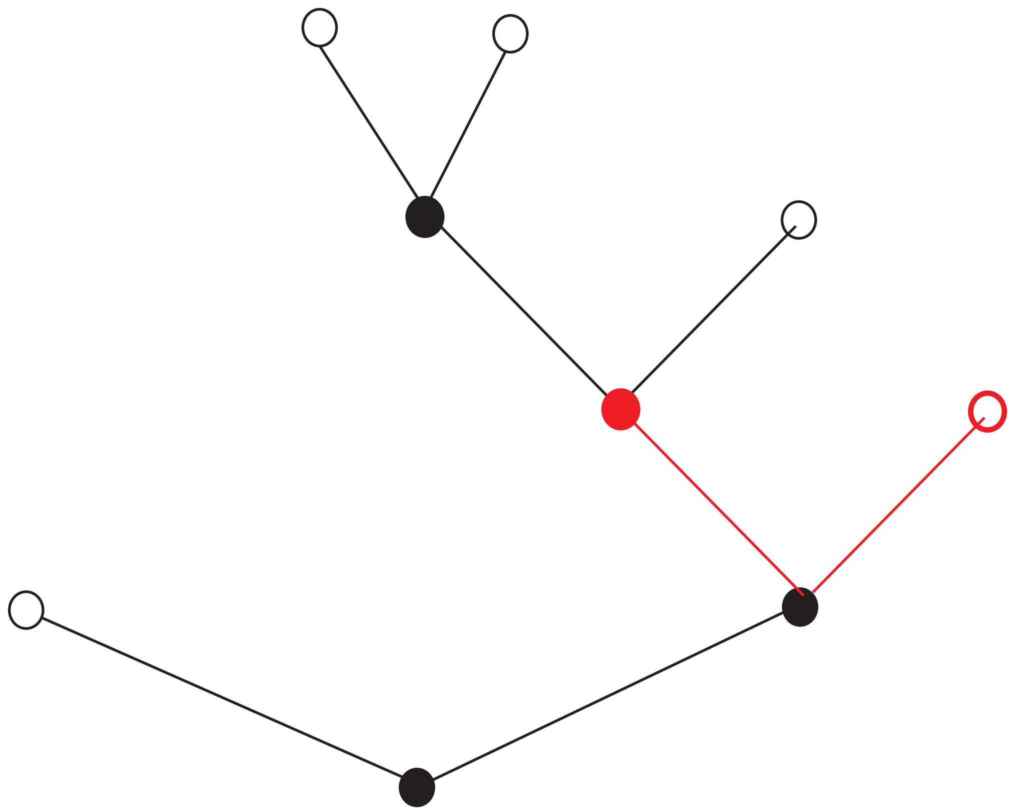
Attach the 2-leaf tree א, rooted in the chosen vertex



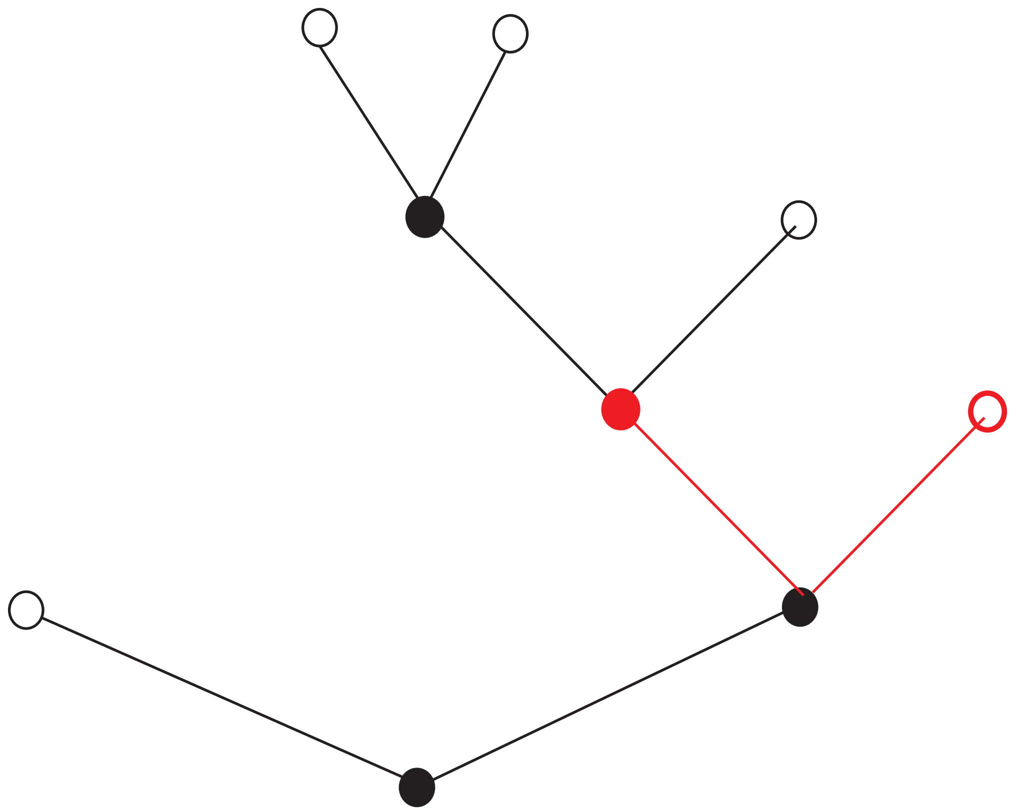
Choose one of the two leaves of ℵ at random

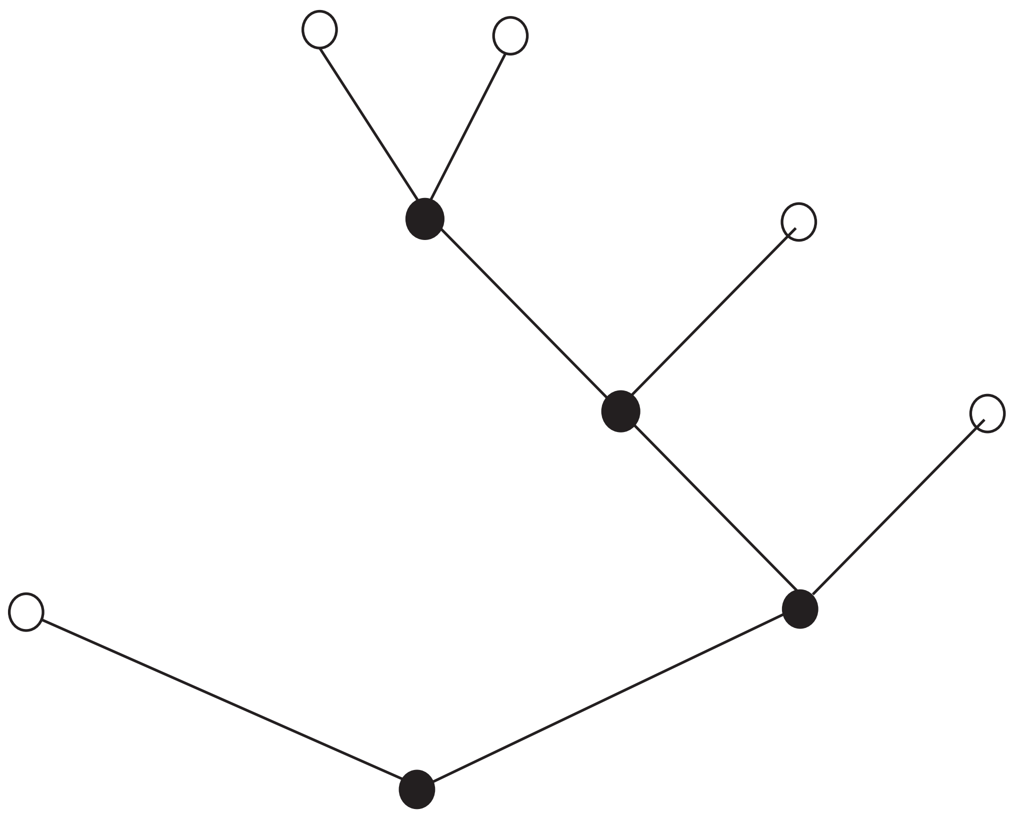


Attach the offspring tree at the the chosen leaf of ℵ



Thus, one additional leaf has been created

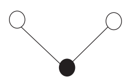




Definition:

The **Rémy chain**

is a binary-tree valued Markov chain $(T_1, T_2, \ldots)$

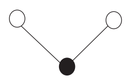
starting in 

and with the following transition mechanism:

Definition:

The **Rémy chain**

is a binary-tree valued Markov chain $(T_1, T_2, \ldots)$

starting in 

and with the following transition mechanism:

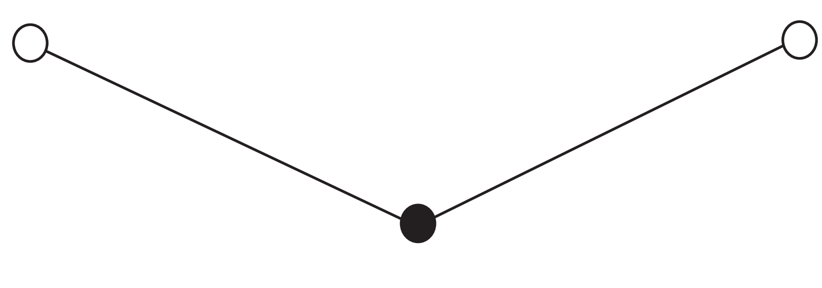
Given $T_n = \mathrm{t}$, choose a vertex v of t at random,

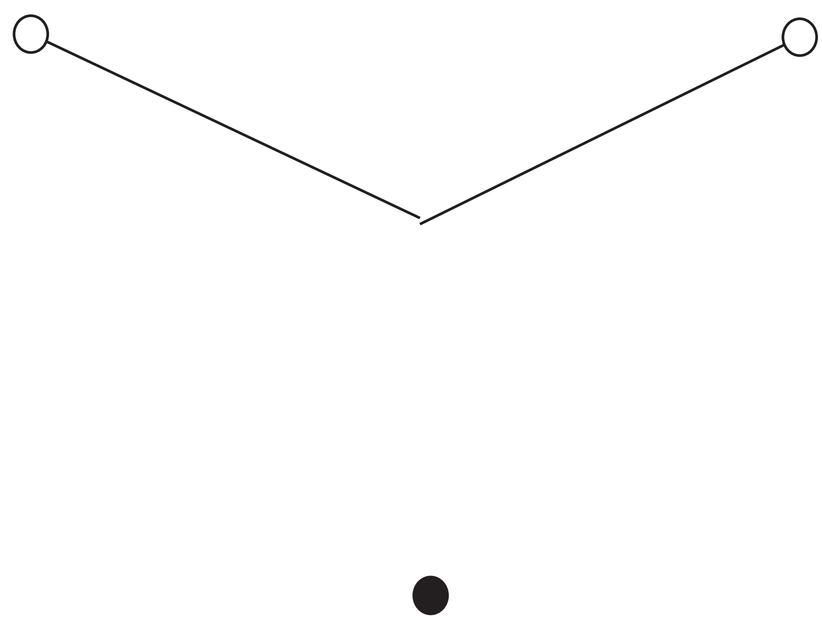
put the offspring tree of v aside,

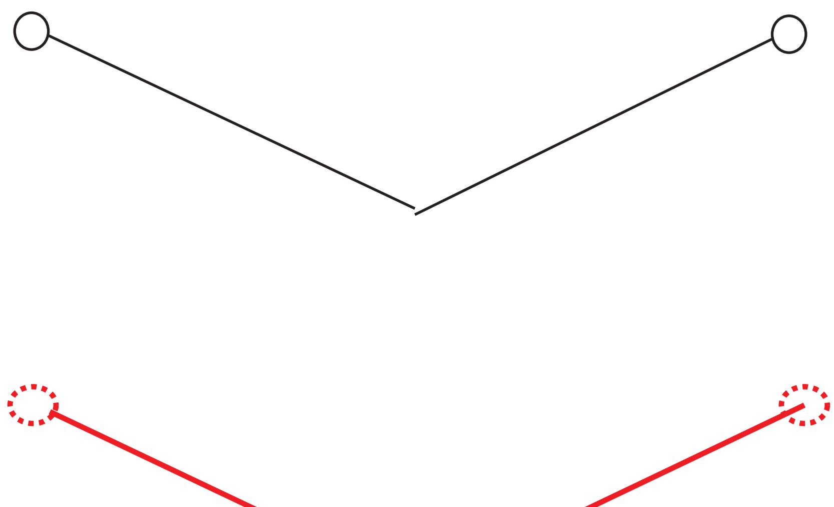
insert 2 children of v,

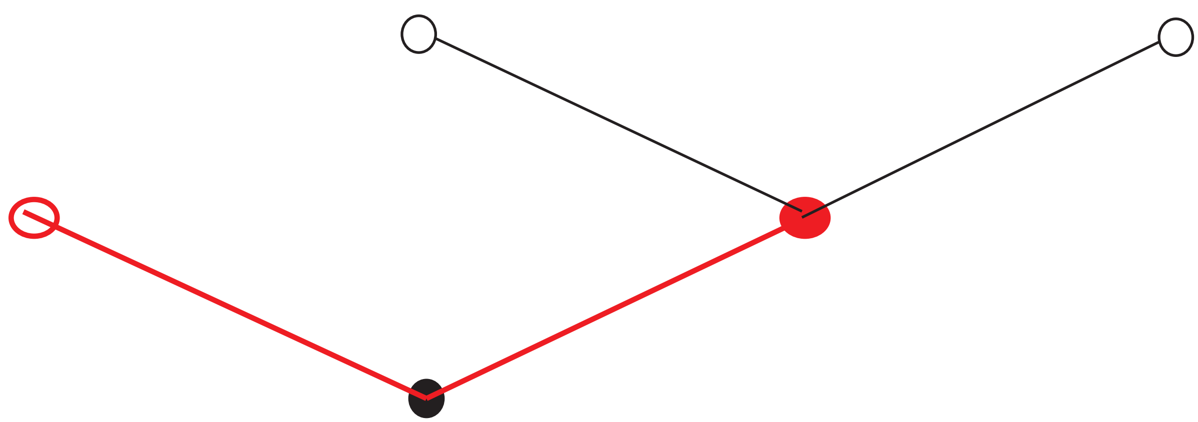
choose one of them at random,

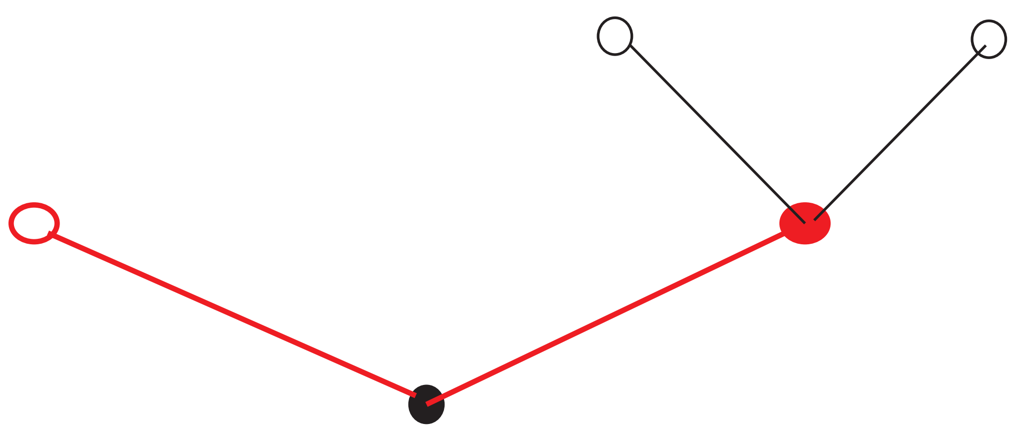
and re-attach to it the previous offspring tree of v.

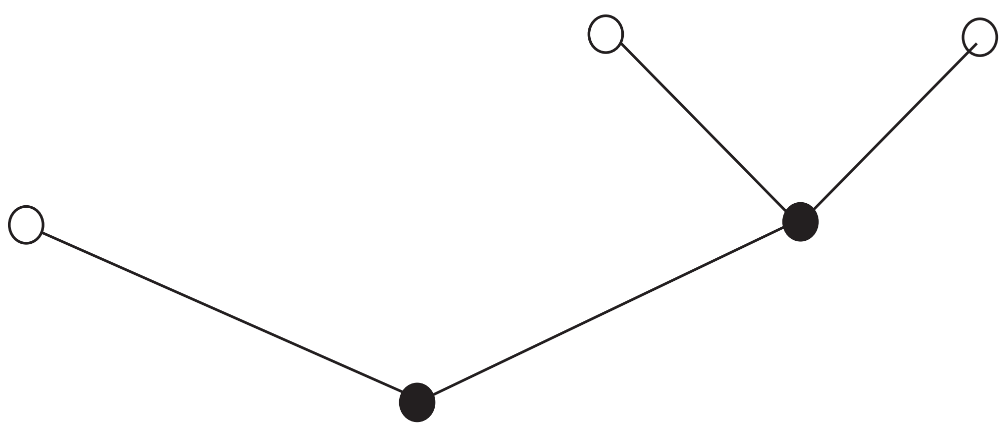


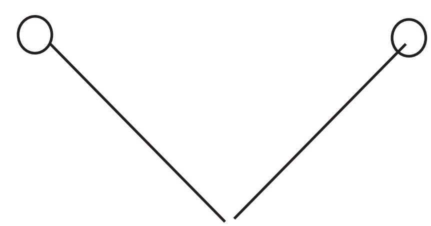


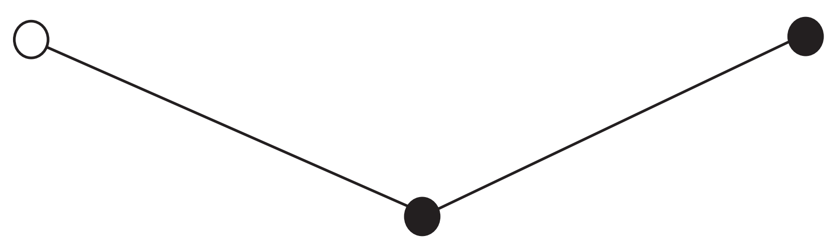


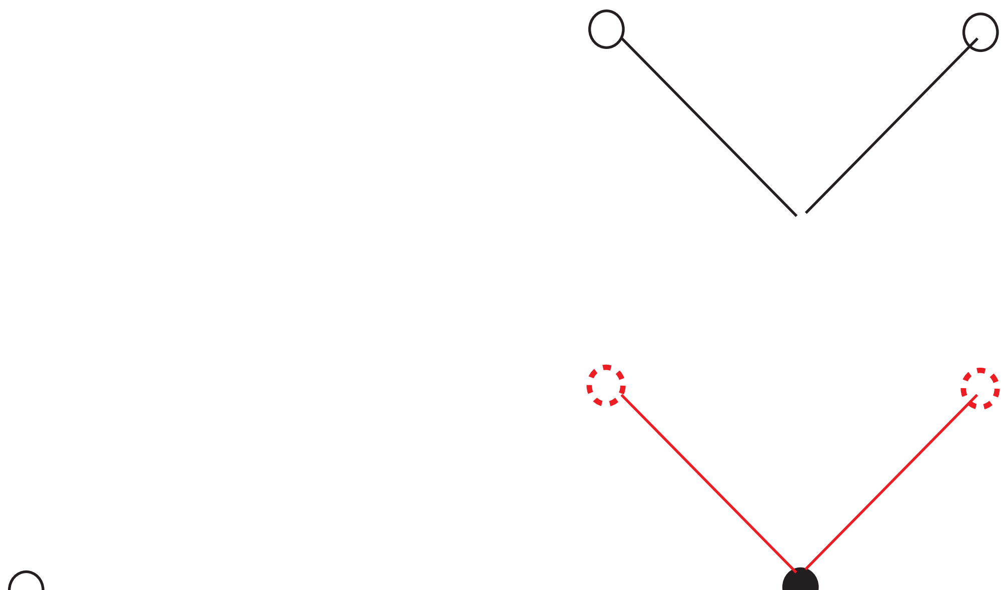


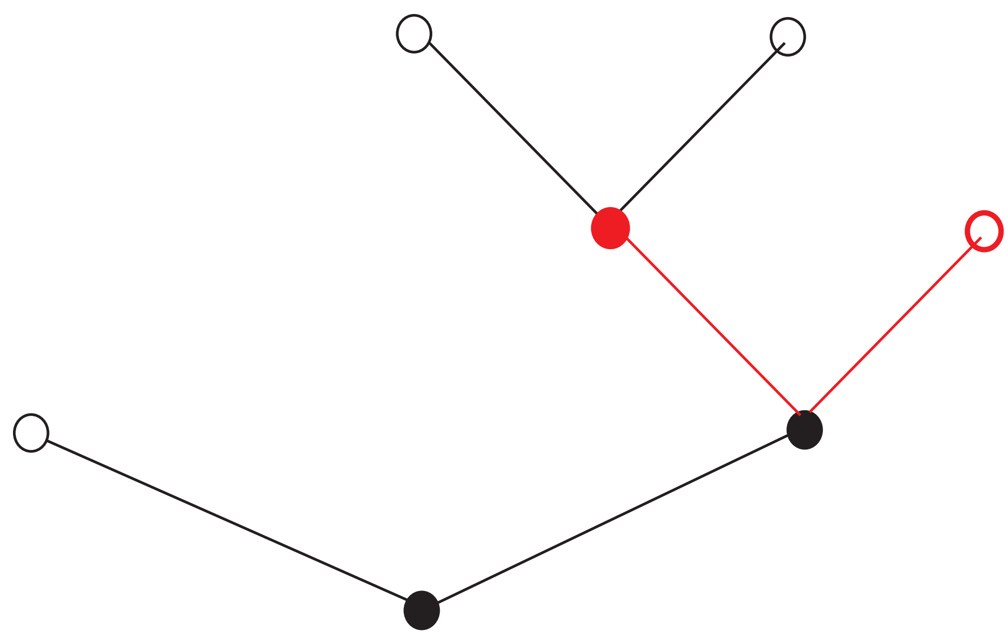


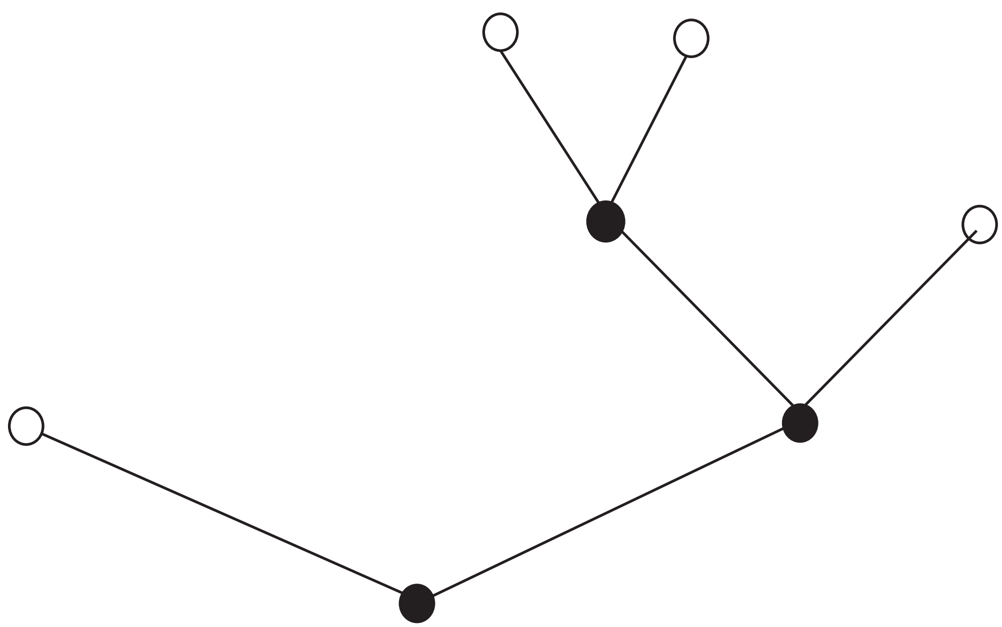


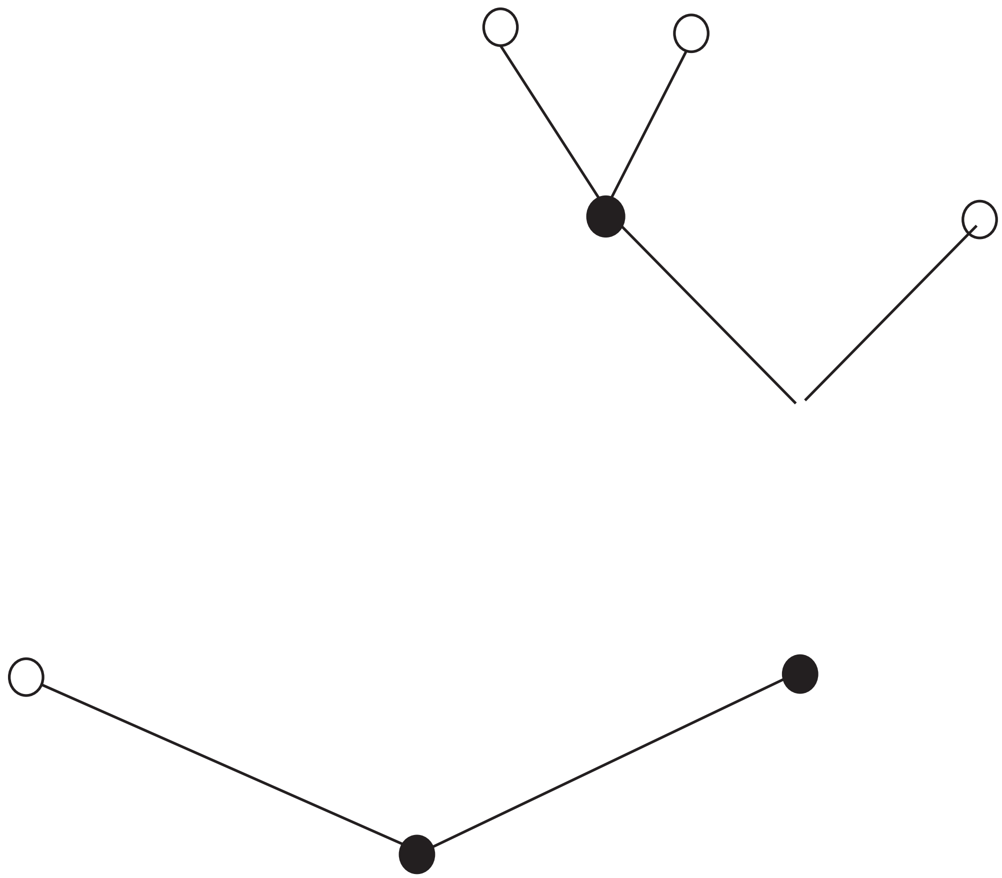


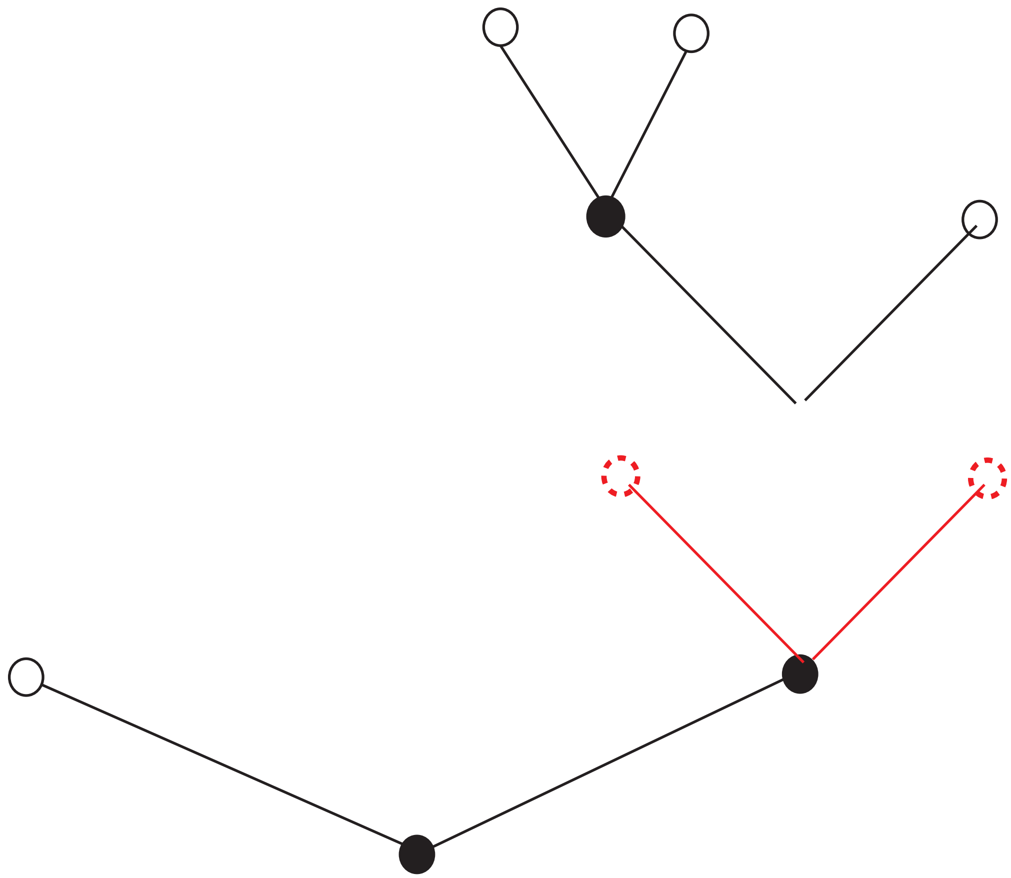


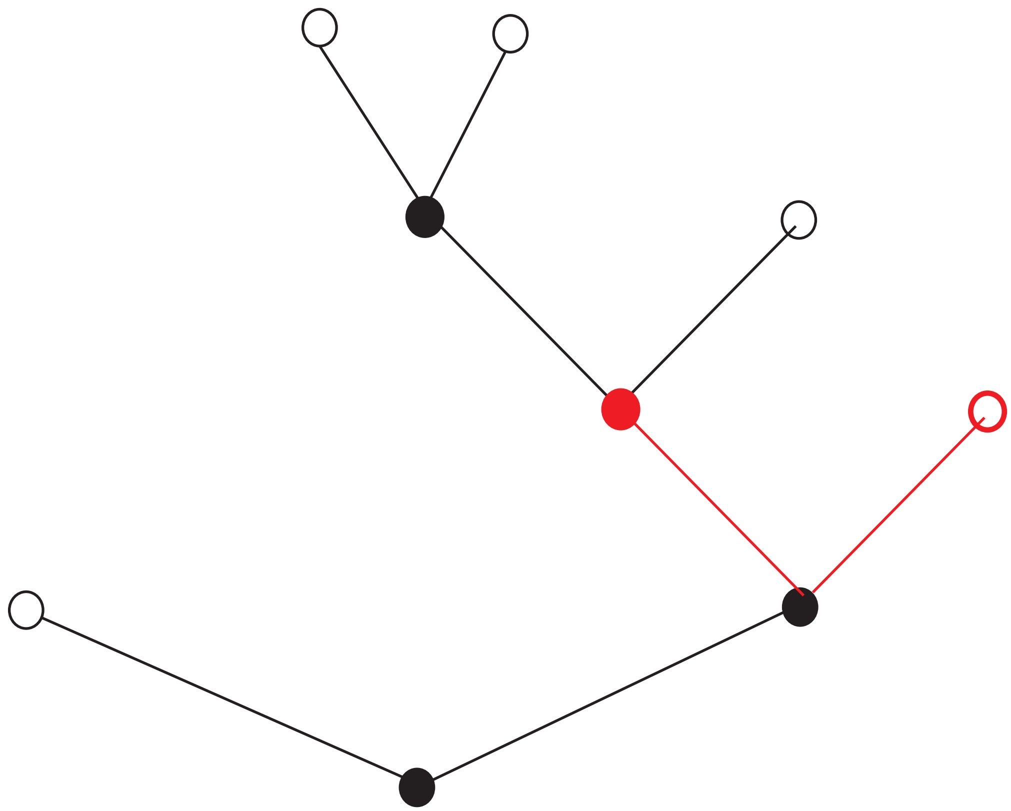


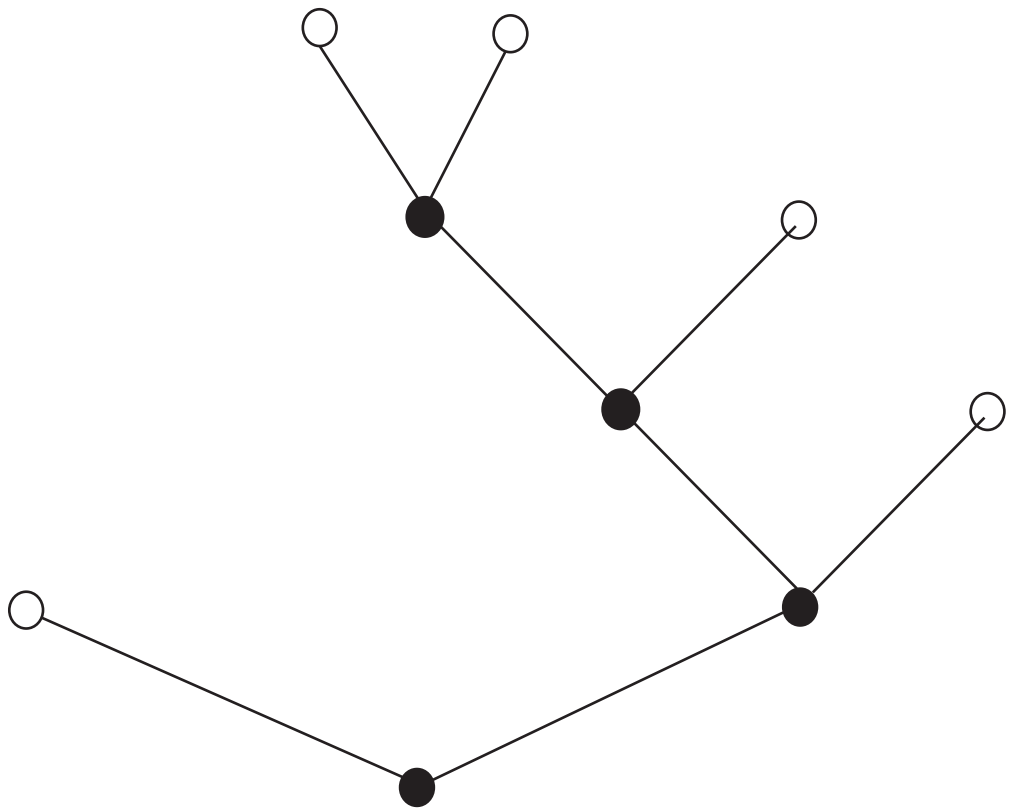


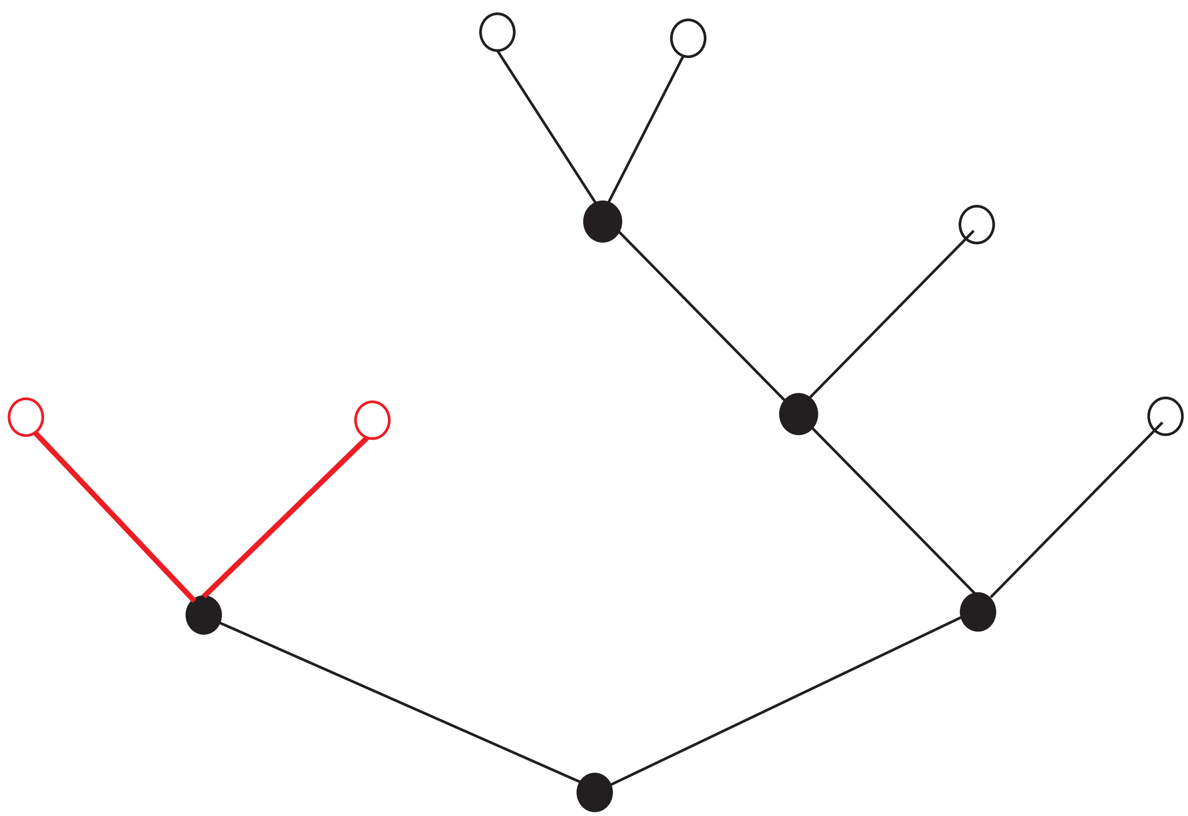


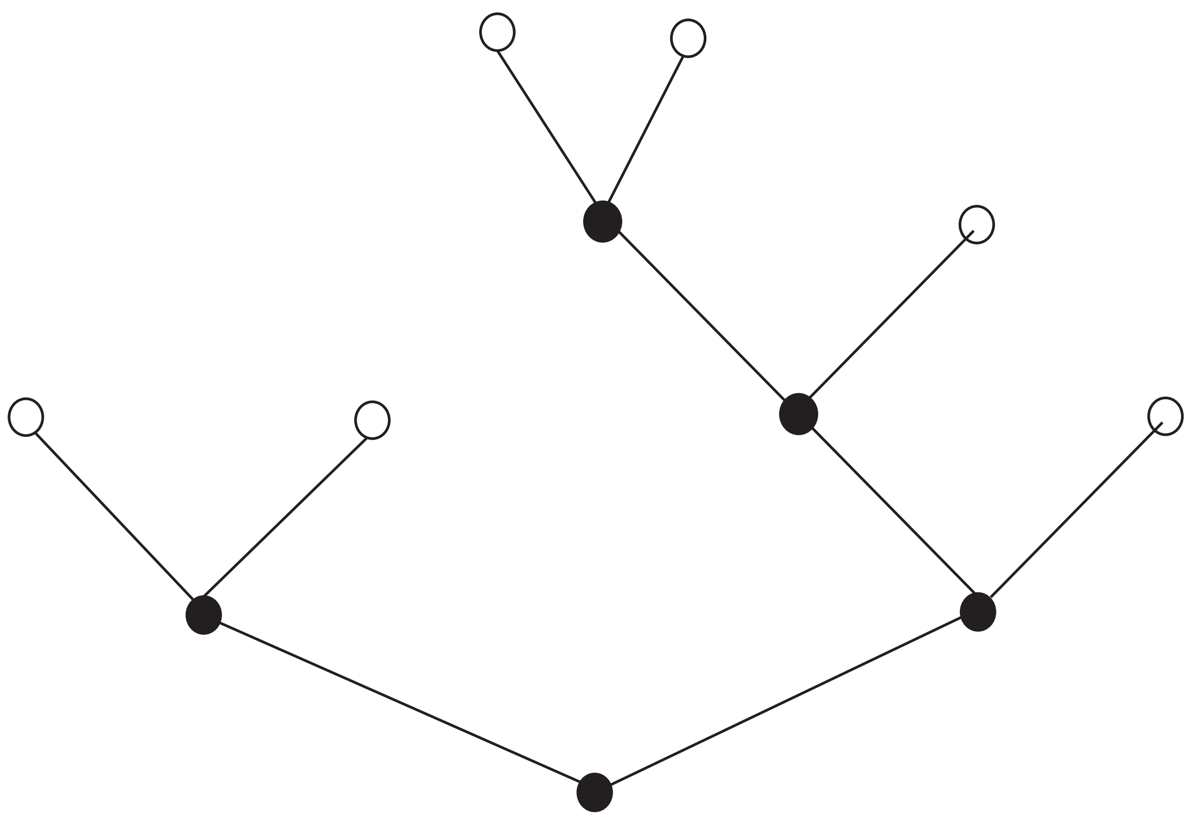










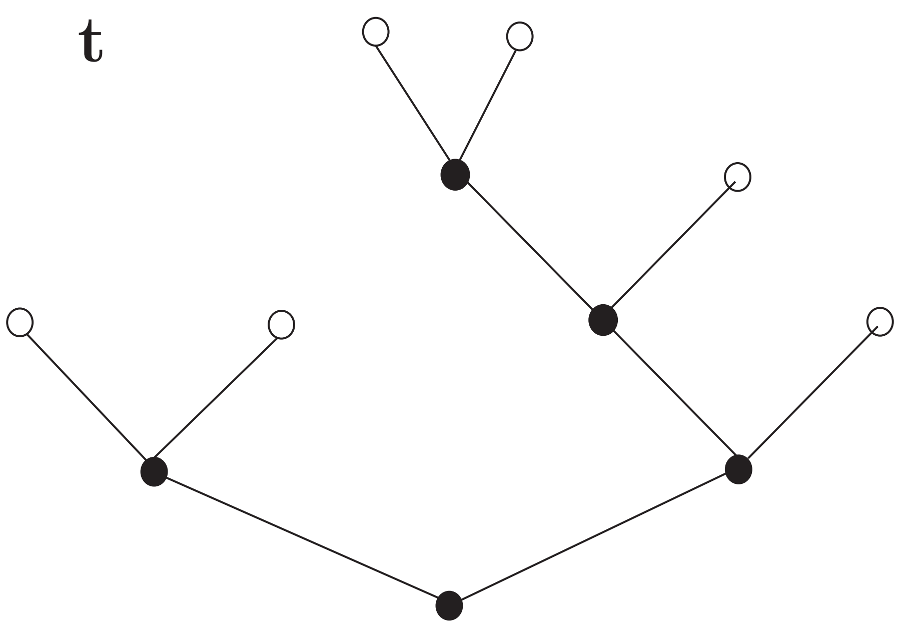


Facts (proved by induction)

T_n is uniformly distributed on the set of binary trees with $n + 1$ leaves.

Facts (proved by induction):

T_n is uniformly distributed on the set of binary trees with $n + 1$ leaves. This set has cardinality $C_n = \frac{1}{n+1}\binom{2n}{n}$.

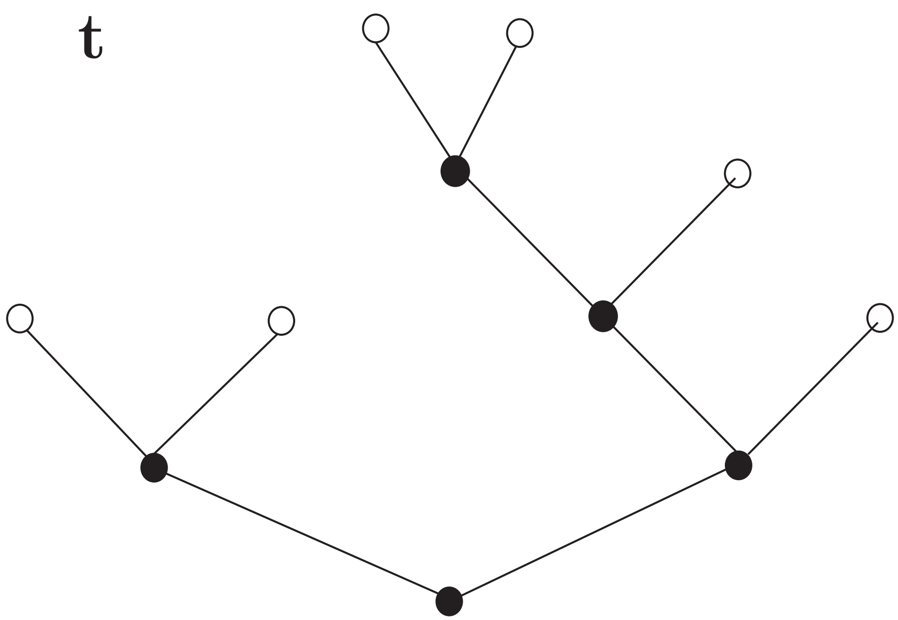


one realisation of T_5

Facts (proved by induction):

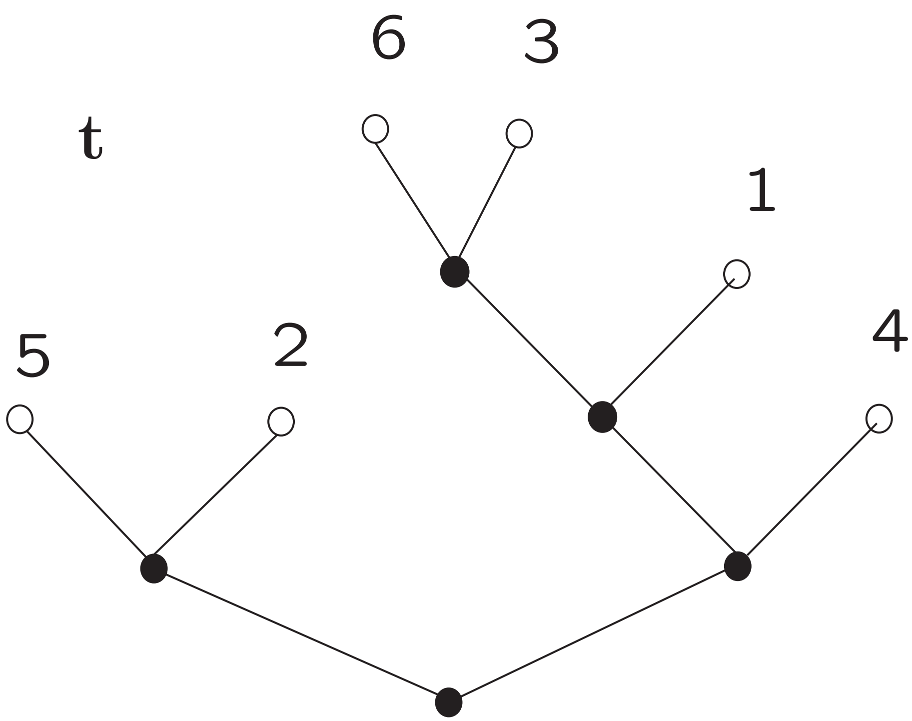
For each n, given $T_n = \mathbf{t}$, the “age order” of the leaves of $\mathbf{t}$ is completely random.

$\mathbf{t}$

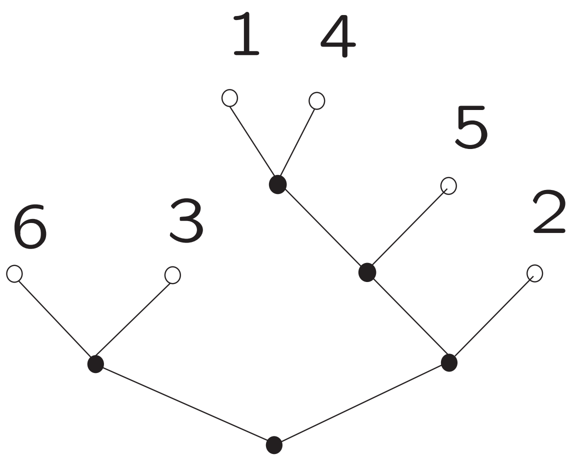


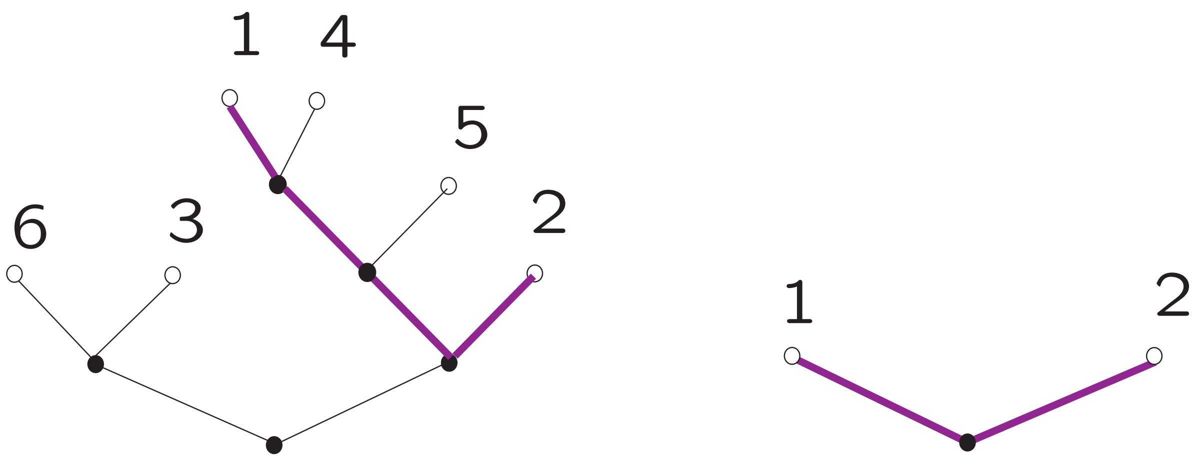
Facts (proved by induction):

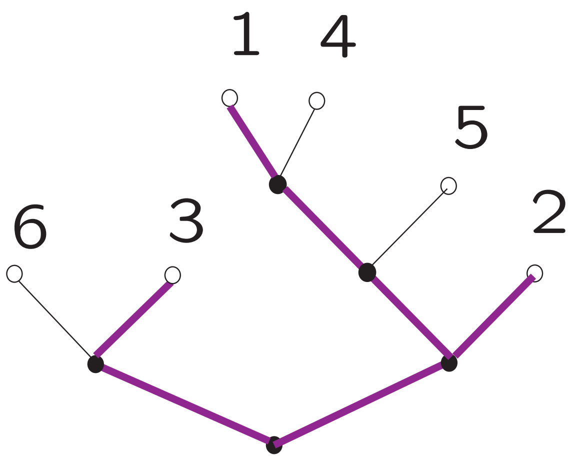
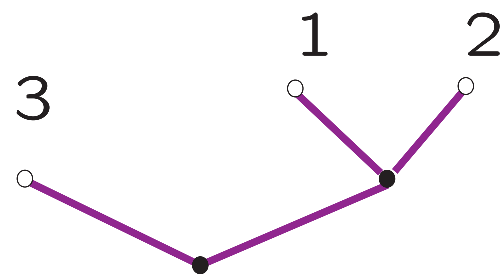
For each n, given $T_n = \mathbf{t}$, the “age order” of the leaves of $\mathbf{t}$ is completely random.

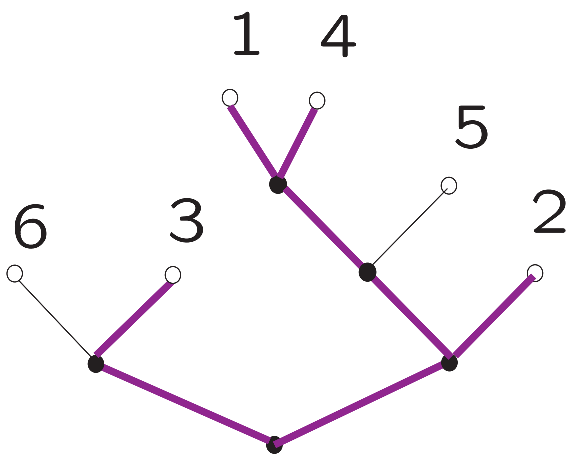
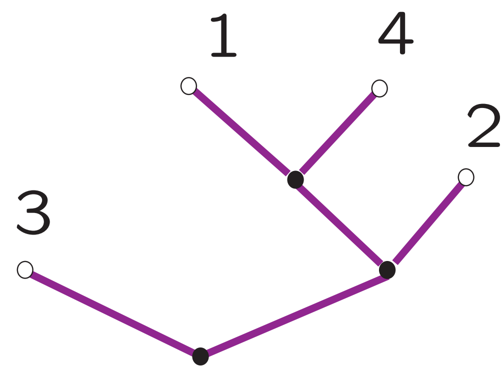


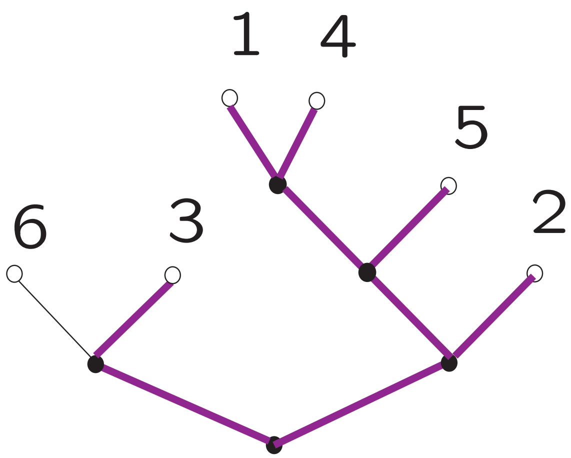
The age order of the leaves determines the path from $\aleph$ to $\mathbf{t}$


1
4
5
6
3
2

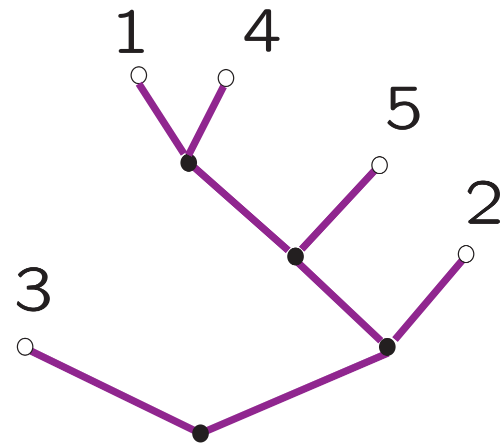

1
4
5
6
3
2
1
2


1
4
5
6
3
2

3
1
2

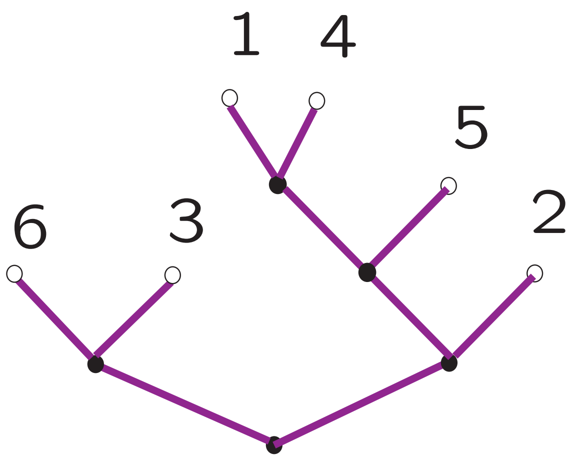

1
4
5
6
3
2

1
4
2
3

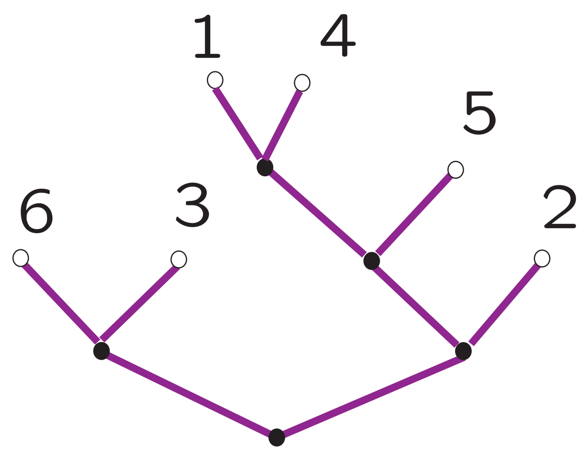


1
4
5
6
3
2



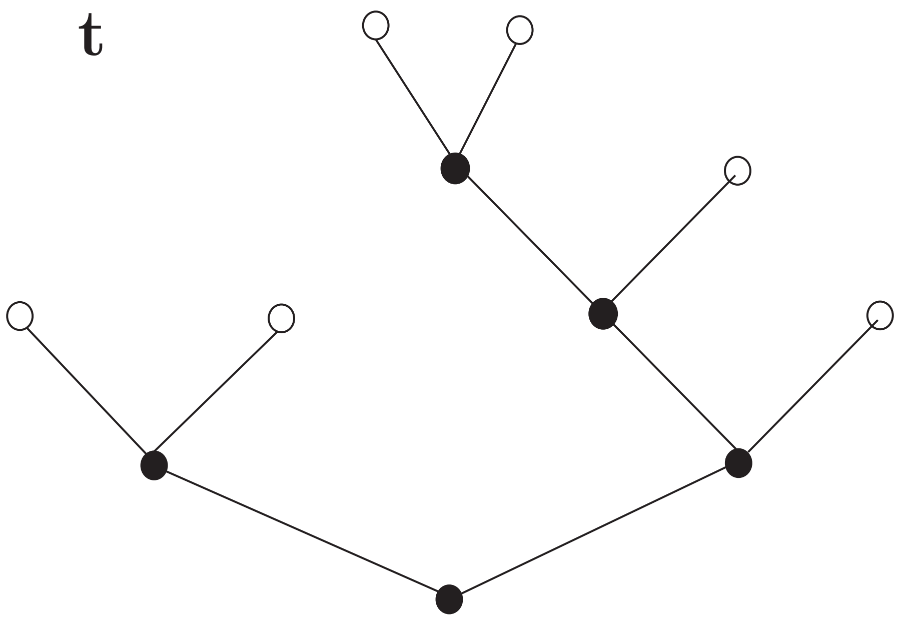
1
4
5
2
3


1
4
5
6
3
2


1
4
5
6
3
2

Thus, for given $\mathbf{t}$,

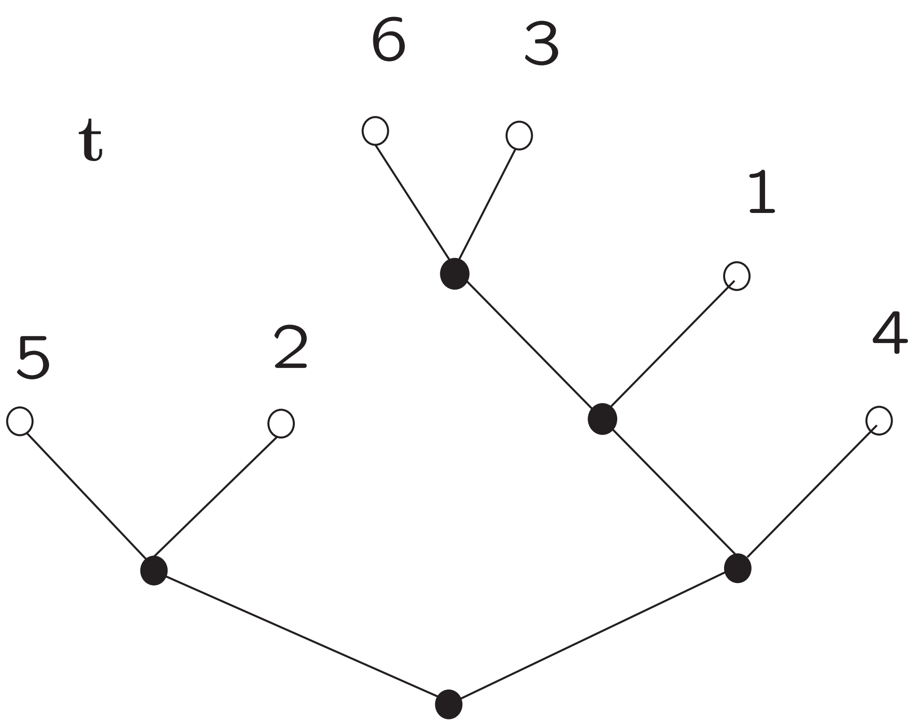
we obtain the *random bridge* from $\aleph$ to $\mathbf{t}$



Thus, for given $\mathfrak{t}$,

we obtain the *random bridge* from $\aleph$ to $\mathfrak{t}$

through a random labeling of the leaves of $\mathfrak{t}$



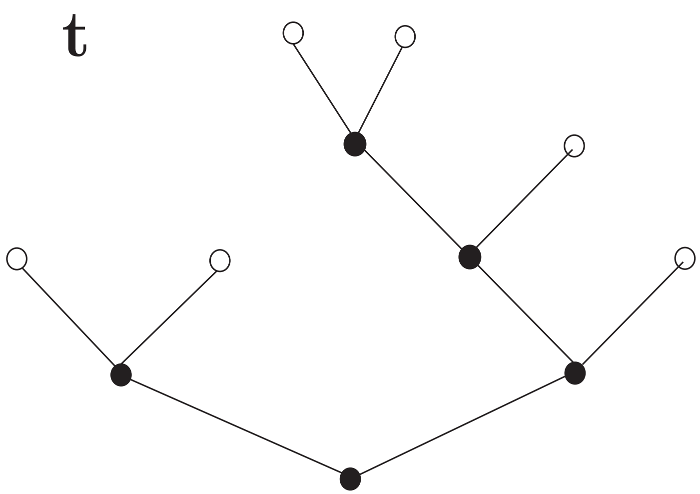
which then determines

the path from $\aleph$ to $\mathfrak{t}$

as we just saw.

For a binary tree $\mathbf{t}$ with $n+1$ leaves,

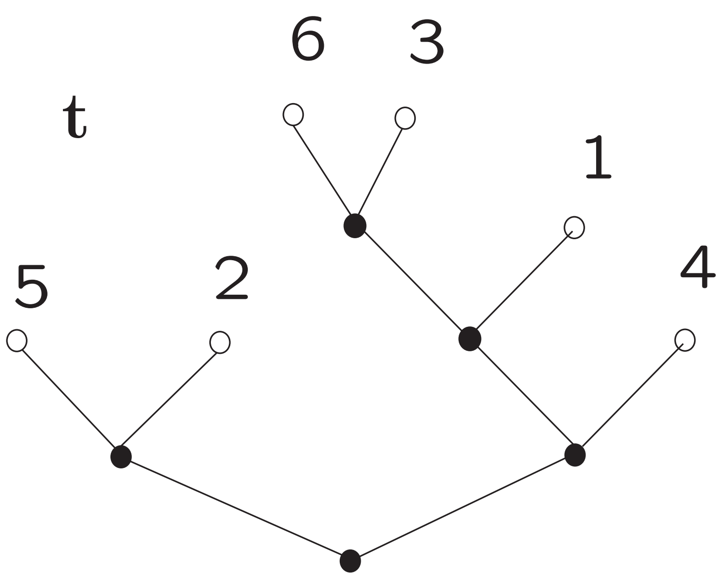
the *Rémy bridge* $(T_1^{\mathbf{t}}, \ldots T_n^{\mathbf{t}})$ from $\aleph$ to $\mathbf{t}$ arises as follows



For a binary tree $\mathbf{t}$ with $n+1$ leaves,

the *Rémy bridge* $(T_1^{\mathbf{t}}, \ldots T_n^{\mathbf{t}})$ from $\aleph$ to $\mathbf{t}$ arises as follows

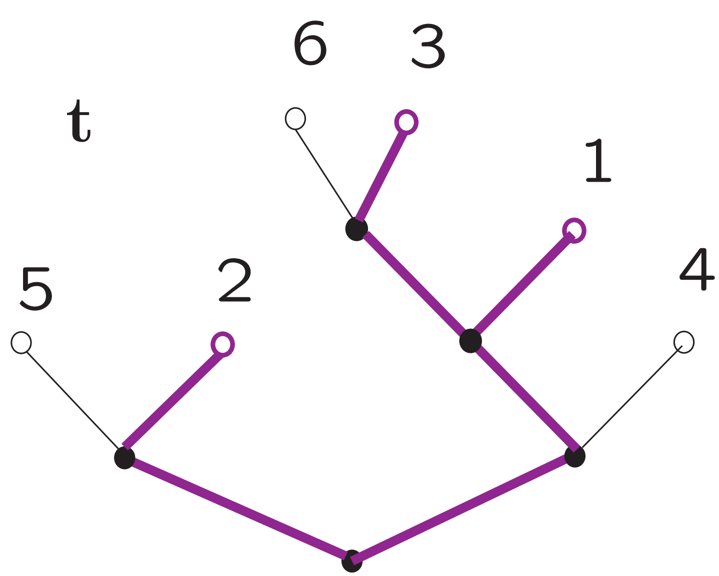
- label the leaves of $\mathbf{t}$ randomly by $1, 2, \ldots, n+1$

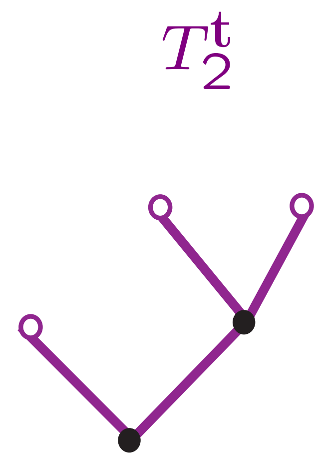


For a binary tree $\mathbf{t}$ with $n+1$ leaves,

the *Rémy bridge* $(T_1^{\mathbf{t}}, \ldots T_n^{\mathbf{t}})$ from $\aleph$ to $\mathbf{t}$ arises as follows

- label the leaves of $\mathbf{t}$ randomly by $1, 2, \ldots, n+1$
- for $k = 1, \ldots, n$, let $T_k^{\mathbf{t}}$ be the (reduced binary) subtree of $\mathbf{t}$ spanned by the leaves labeled by $1, \ldots, k+1$

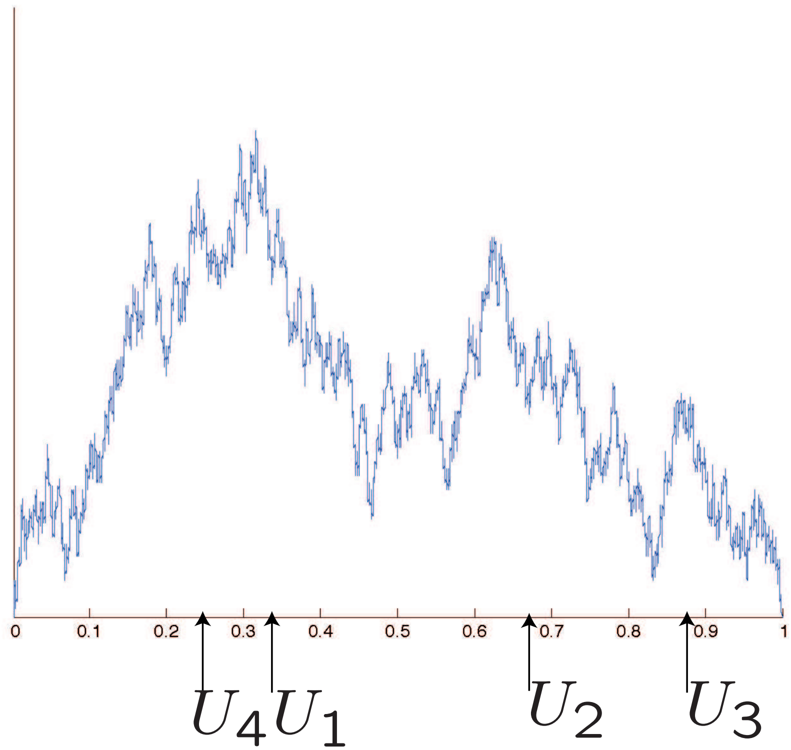




Le Gall (1999) noticed the following fact:

Let $B : [0, 1] \mapsto \mathbb{R}_+$ be a standard Brownian excursion,

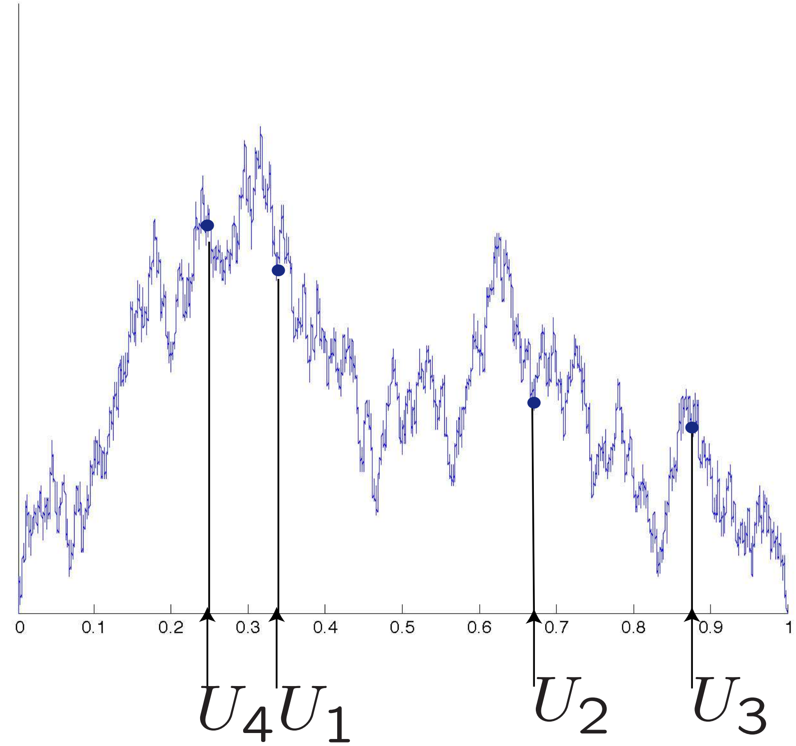
let $U_1, U_2, \ldots$ be i.i.d. uniform in $[0, 1]$



Le Gall (1999) noticed the following fact:

For each n let T_n be the binary tree drawn into B

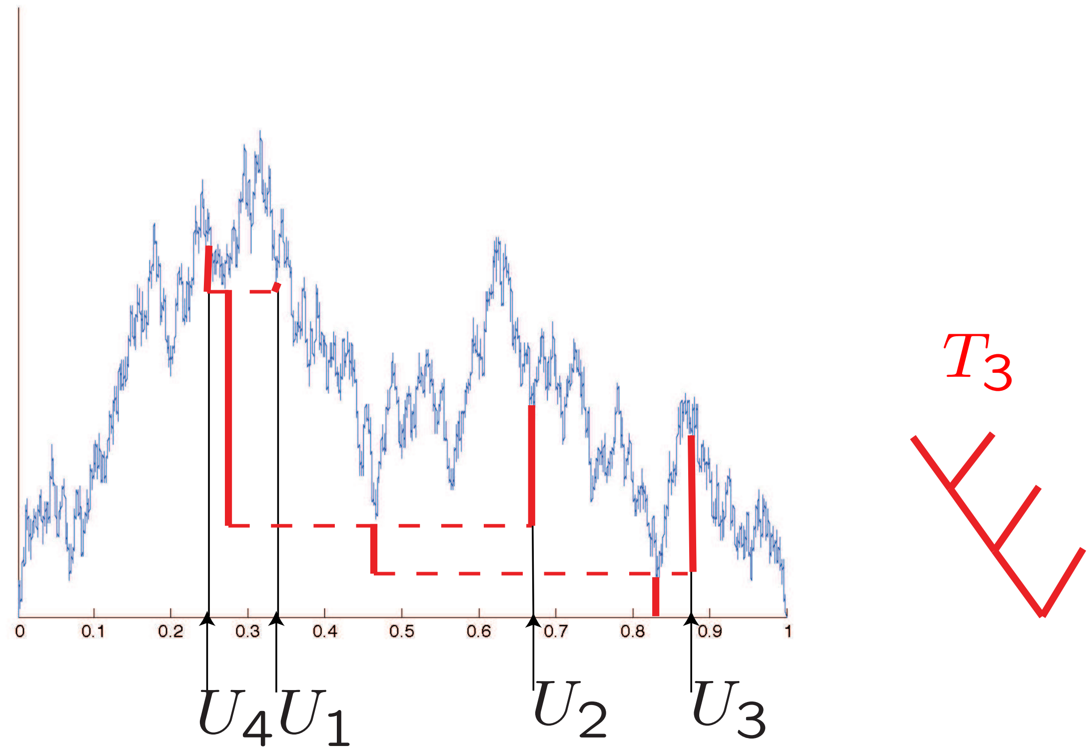
below the points $(U_1, B(U_1)), \ldots, (U_{n+1}, B(U_{n+1}))$.



Le Gall (1999) noticed the following fact:

For each n let T_n be the binary tree drawn into B

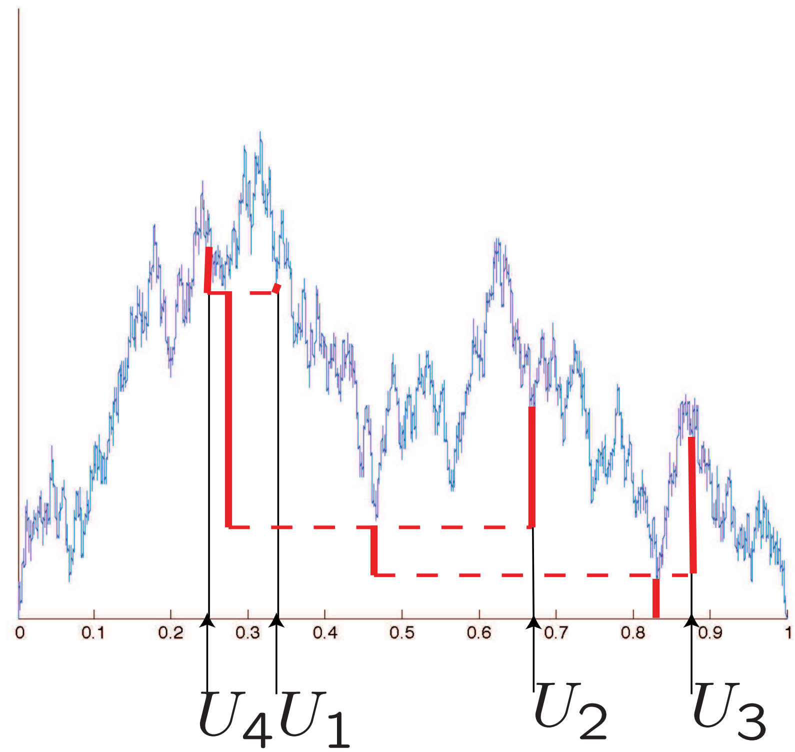
below the points $(U_1, B(U_1)), \ldots, (U_{n+1}, B(U_{n+1}))$.

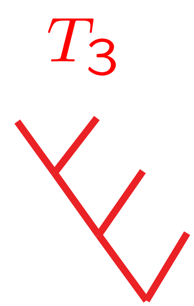


Le Gall (1999) noticed the following fact:

Then, randomized over B and $U_1, U_2, \ldots$,

the distribution of $(T_1, T_2, \ldots)$ is that of the Rémy chain.

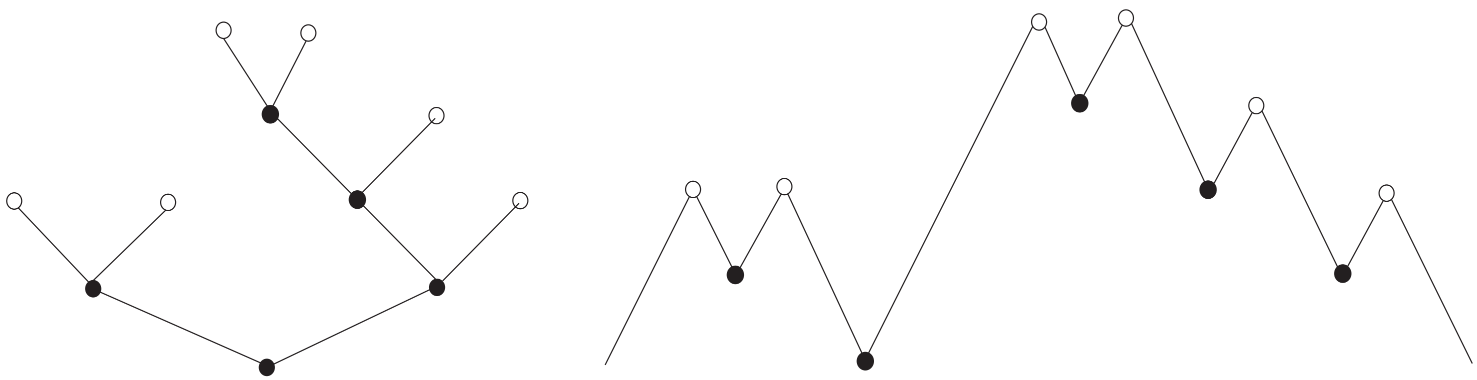




Let us discuss this first in the discrete world:

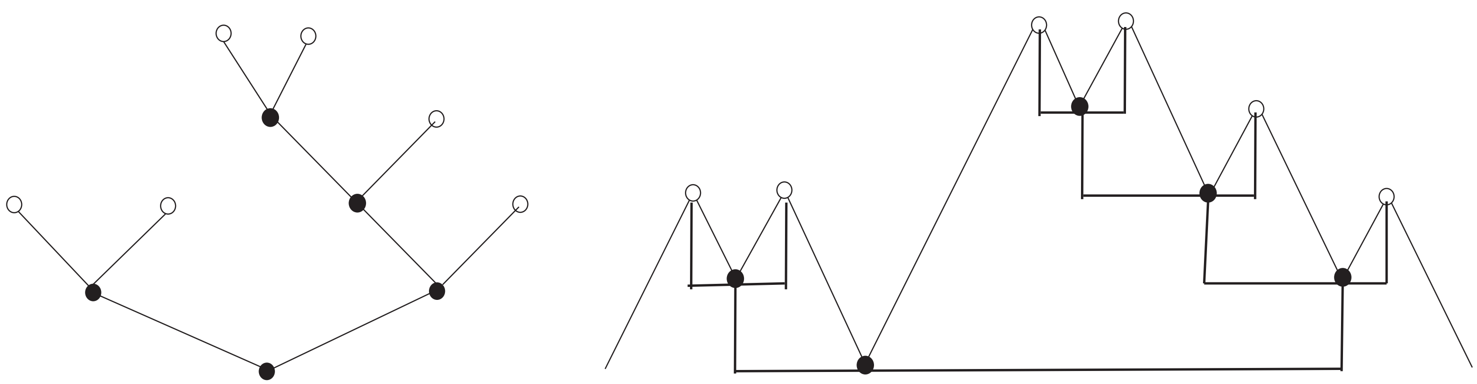
Harris coding of a binary tree with $n + 1$ leaves

by a function $f : \{0, 1, \ldots, 4n\} \to \mathbb{N}_0$



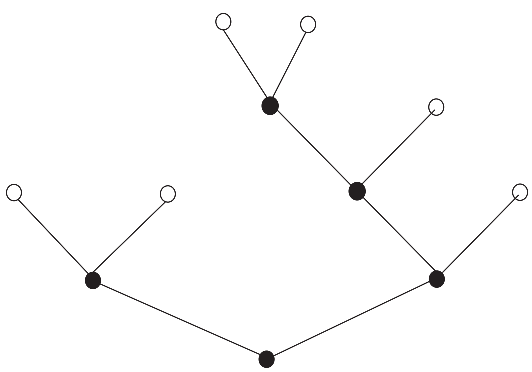
Harris coding of a binary tree with $n + 1$ leaves

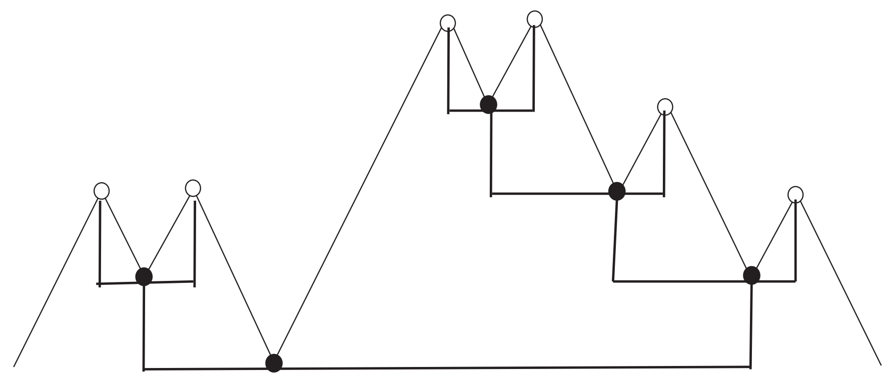
by a function $f : \{0, 1, \ldots, 4n\} \to \mathbb{N}_0$



A random labeling of the leaves of $\mathfrak{t}$

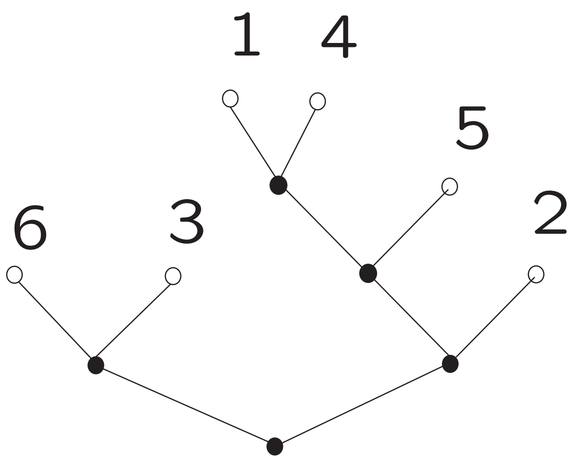
corresponds to a random labeling of the maxima of f

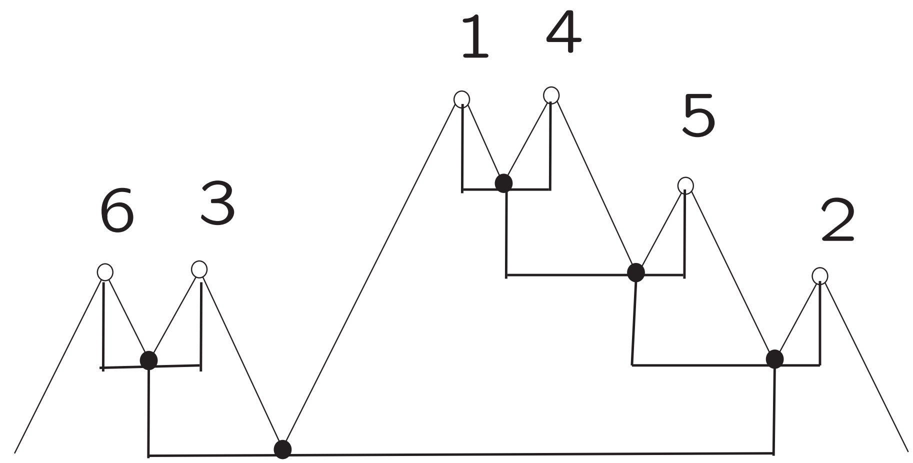




A random labeling of the leaves of $\mathbf{t}$

coresponds to a random labeling of the maxima of f





Philippe Marchal (2003) *Constructing a sequence of random walks strongly converging to Brownian motion*

The rescaling $u \mapsto \frac{1}{2\sqrt{2n}} F_n(u \cdot 4n)$, $u \in [0, 1]$,
of the random functon F_n that codes the random tree T_n
converges almost surely to a standard Brownian excursion B.

The Brownian excursion B encodes a continuum tree $\mathscr{T}^B$
that is rooted, is endowed with a left-right order
and carries a measure that is supported by the leaves of $\mathscr{T}^B$.

(The unordered version of $\mathscr{T}^B$ represents the so-called
Aldous' *Continuum Random Tree*.)

Thus, Le Gall's representation of the Rémy chain corresponds
to the sequence $(T_1, T_2, ...)$ of binary trees
that arise through a successive i.i.d. sampling of leaves $\xi_1, \xi_2, \ldots$
from $\mathscr{T}^B$.

Let Π^{b} be the distribution of the Rémy chain $(T_1, T_2, \ldots)$ drawn from a standard Brownian excursion B, *given* $B = \mathbf{b}$.

Π^{b} is different from $\Pi :=$ the distribution of the Rémy chain.

However,

Π^{b} has the same Markovian backward dynamics as Π:

Under Π^{b}, given $T_n = \mathbf{t}$,

$(T_1, \ldots, T_n)$ is a Rémy bridge from $\aleph$ to $\mathbf{t}$.

In this sense, Π^{b} is an *infinite Rémy bridge distribution*.

Definition. An *infinite Rémy bridge* is a Markov chain $(T_n^\infty)_{n\in\mathbb{N}}$ such that $T_1^\infty = \aleph$ and

$$\mathbb{P}(T_n^\infty = \mathbf{s} \mid T_{n+1}^\infty = \mathbf{t}) = \mathbb{P}(T_n = \mathbf{s} \mid T_{n+1} = \mathbf{t})$$

for all $n \geq 1$ and finite binary trees $\mathbf{s}, \mathbf{t}$,

i.e. $(T_n^\infty)_{n\in\mathbb{N}}$ has the same backward transition probabilities as the Rémy chain (T_n).

Fact: The infinite Rémy bridge distributions

are in 1-1 correspondence with the nonnegative functions h

defined on the space of finite binary trees

that are harmonic for the Rémy transition matrix

and satisfy $h(\aleph) = 1$.

Here the keyword is the h-transform

$$\mathbb{P}(T_n^\infty = \mathbf{t}) = \frac{1}{h(\aleph)} \mathbb{P}(T_n = \mathbf{t}) h(\mathbf{t}).$$

Facts:

- The set of infinite Rémy bridge distributions is convex.

- The distribution of an infinite Rémy bridge $(T_n^\infty)_{n\in\mathbb{N}}$ is extremal in the set of of infinite Rémy bridge distributions if and only if $\mathscr{F}^\infty := \bigcap_m \sigma(T_m^\infty, T_{m+1}^\infty, \ldots)$ is trivial.

- Every infinite Rémy bridge distribution has a unique integral representation as a mixture of extremal ones.

This fits into the general theory of Gibbs specifications
(H. Föllmer (1975), *Phase transition and Martin boundary*)

Example. Let Π be the distribution of the Rémy chain $(T_1, T_2, ...)$.
Think of a two-stage experiment:
First take a standard Brownian excursion B, then
given $\{B = \mathbf{b}\}$, sample sucessively from the leaves of $\mathscr{T}^{\mathbf{b}}$.

With $\Pi^{\mathbf{b}}$ being the distribution of the arising sequence $(T_1^{\mathbf{b}}, T_2^{\mathbf{b}}, ...)$
we have the desintegration

$$\Pi(\cdot) = \int \Pi^{\mathbf{b}}(\cdot)\ \mathbf{P}(B \in \mathrm{d}\mathbf{b}).$$

By the Hewitt-Savage 0-1 law, for $\mathbf{P}(B \in (.))$- almost all $\mathbf{b}$,
$\Pi^{\mathbf{b}}$ is $\mathscr{F}^{\infty}$-trivial (and hence extremal).

A sequence of finite binary trees $\mathbf{t}_k$ is said *to converge in the Doob-Martin topology* of the Rémy chain if the sequence of Rémy bridge distributions $\Pi^{\mathbf{t}_k}$ converges in the sense of finite-dimensional distributions.

The *Doob-Martin boundary* is then the set of all Doob-Martin limit points outside the space of *finite* binary trees.

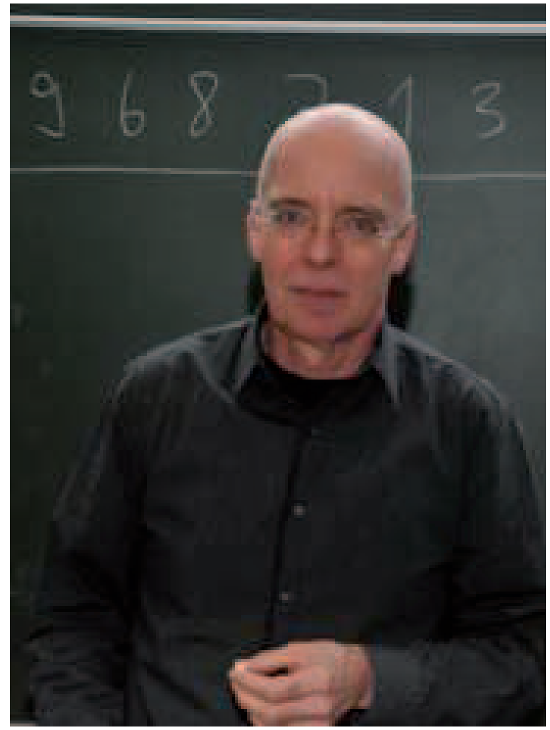
Fact: Every point in the D-M boundary of the Rémy chain corresponds to some infinite Rémy bridge distribution.

Questions:

1. Does every point in the D-M boundary of the Rémy chain correspond to an *extremal* infinite Rémy bridge distribution?
(In more analytic terms: Is the full D-M boundary equal to the so-called *minimal* boundary ?)

2. What does the D-M boundary of the Rémy chain look like?
(The latter question was asked to us by J.F. Le Gall)

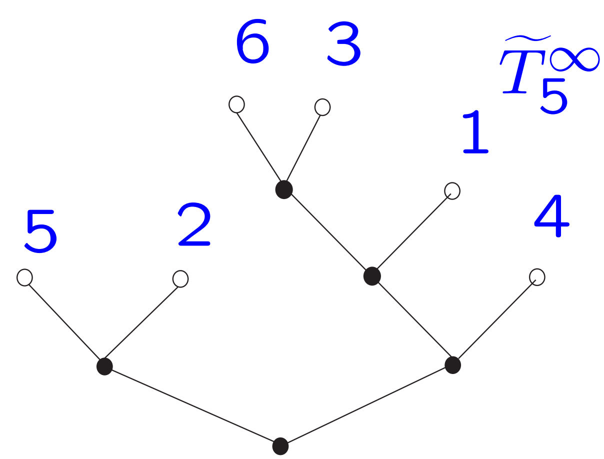
Steve Evans



Rudolf Grübel

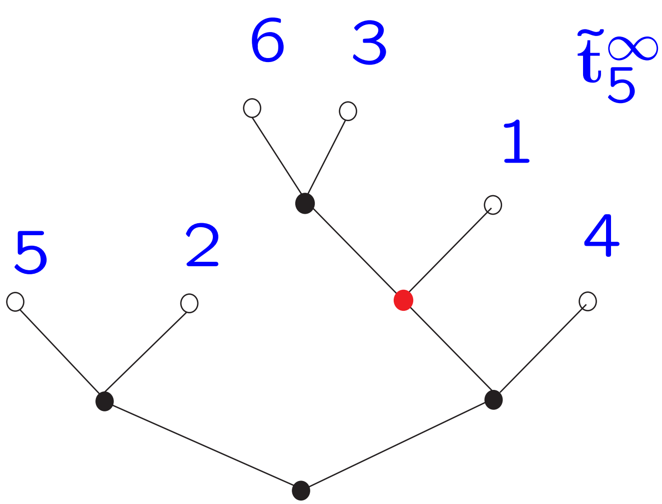
Leaf-labeled infinite Rémy bridge distributions:

Proposition [EGW17]. There is a 1-1 correspondence between the infinite Rémy bridge distributions and the leaf-labeled infinite Rémy bridge distributions, i.e. the distributions of sequences $(\widetilde{T}_n^\infty)$ of leaf-labelled binary trees $\widetilde{T}_n^\infty$ whose leaf set is randomly and bijectively labeled by $[n+1]$, and where $\widetilde{T}_{n-1}^\infty$ is obtained from $\widetilde{T}_n^\infty$ by removing the leaf $n+1$ and its sibling and closing the (potential) gap - as we have seen before.



The combinatorial tree encoded by a path of a leaf-labeled infinite Rémy bridge:

Let $(\tilde{\mathbf{t}}_n^\infty)$ be a path of a leaf-labeled infinite Rémy bridge. From $(\tilde{\mathbf{t}}_n^\infty)$ one can read off of the most recent common ancestor $\langle i, j \rangle$ of the leaves labeled by i and j.



$$\langle 1, 3 \rangle = \langle 1, 6 \rangle$$

The combinatorial tree encoded by a path of a leaf-labeled infinite Rémy bridge:

Let $(\tilde{\mathbf{t}}_n^\infty)$ be a path of a leaf-labeled infinite Rémy bridge. From $(\tilde{\mathbf{t}}_n^\infty)$ one can read off of the most recent common ancestor $\langle i, j \rangle$ of the leaves labeled by i and j.
Formally, the $\langle i, j \rangle$ are equivalence classes w.r. to an equivalence relation $\equiv$ on $\mathbb{N} \times \mathbb{N}$:
two pairs of leaves (i, j), (k, ℓ) are equivalent if they have the same MRCA in $\tilde{\mathbf{t}}_n^\infty$ for $n > \max(i, j, k, \ell)$.
The equivalence classes $\langle i, j \rangle$ represent all the vertices of the tree, the $\langle i, i \rangle =: i$ are singletons and represent the leaves.

Ordering the tree encoded by $(\tilde{\mathfrak{t}}_n^\infty)$:

$(\tilde{\mathfrak{t}}_n^\infty)$ induces

a partial order "by descent" on the vertices $\langle i,j\rangle$, $\langle i,k\rangle, \ldots$

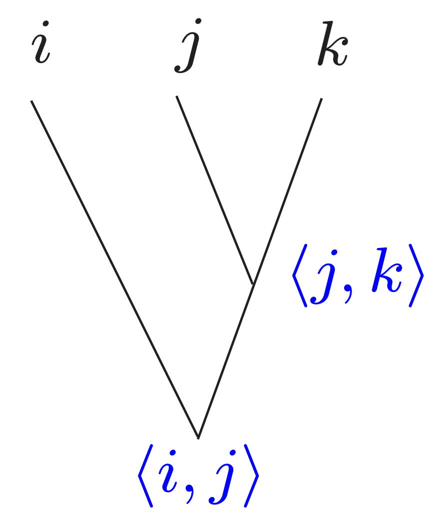
and a left-right order on the leaves $i, j, \ldots$

E.g. in the figure:

$$\langle i,j\rangle = \langle i,k\rangle < \langle j,k\rangle$$

$$w(i,j) = w(i,k) = w(j,k) = \curvearrowright$$

$$w(j,i) = w(i,k) = w(k,j) = \curvearrowleft$$



Ordering the tree encoded by $(\tilde{\mathfrak{t}}_n^\infty)$:

$(\tilde{\mathfrak{t}}_n^\infty)$ induces

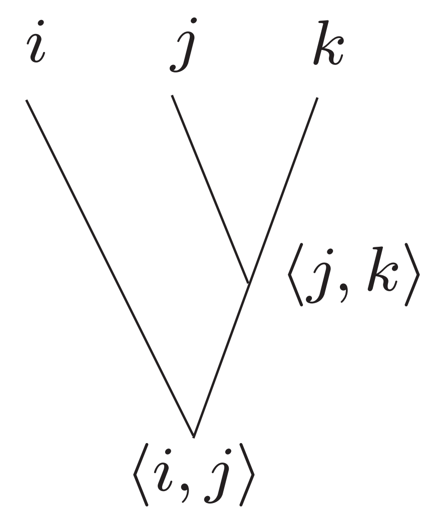
a partial order "by descent" on the vertices $\langle i, j\rangle$, $\langle i, k\rangle, \ldots$

and a left-right order on the leaves $i, j, \ldots$

In any case:

$$\langle i, j\rangle = \langle i, k\rangle < \langle j, k\rangle$$

$$\Longrightarrow \; w(i,j) = w(i,k)$$



Ordering the tree encoded by $(\tilde{\mathfrak{t}}_n^\infty)$:

$(\tilde{\mathfrak{t}}_n^\infty)$ induces

a partial order "by descent" on the vertices $\langle i,j\rangle$, $\langle i,k\rangle, \ldots$

and a left-right order on the leaves $i, j, \ldots$

The equivalence classes $\langle i,j\rangle$ and the two orders $<$ and $w(\cdot,\cdot)$ in the system $\mathscr{D}^{(\tilde{\mathfrak{t}}_n^\infty)} := (\equiv, <, w)$ obey natural axioms which define what we call a didendritic system with label set $\mathbb{N}$ ([EW20, Def. 6.1 and Prop. 6.7])

Leaf-labeled infinite Rémy bridge distributions correspond to exchangeable didendritic systems:

For a leaf-labeled infinite Rémy bridge $(\widetilde{T}_n^\infty)$, the random didrentic system $\mathscr{D}^{(\widetilde{T}_n^\infty)}$ is exchangeable in the sense that its distribution is invariant under finite permutations of the label set $\mathbb{N}$.

Conversely, to every exchangeable didendritic system there corresponds a (leaf-labeled) infinite Rémy bridge, and this corresponcence is 1-1 [EGW17, Lemma 5.12].

Extremal infinite Rémy bridge distributions correspond to ergodic exchangeable didendritic systems [EGW17, Prop. 5.19]

Definition: An exchangeable didentritic system $\mathscr{D}$ is ergodic :

$\iff$ its distribution is trivial

on the σ-field of finite-permutation-invariant events.

By a criterion of Aldous on the ergodicity of exchangeable arrays, $\mathscr{D}$ is ergodic iff for all finite disjoint $H_1, \ldots, H_s \subset \mathbb{N}$ the restrictions $\mathscr{D}_{H_1}, \ldots, \mathscr{D}_{H_s}$ are independent.

The elements of the D-M-boundary constitute extremal infinite bridges [EGW17, Cor. 5.21] :

D-M convergence of $(\mathfrak{t}_k)$ corresponds to f.d.d. convergence of $\left(\mathscr{D}^{\tilde{\mathfrak{t}}_k}\right)$ to an exchangeable didentritic system $\mathscr{D}$.

We apply Aldous' criterion to $\mathscr{D}$.

Let ℓ be so large that $H_1 \cup \cdots \cup H_s \subset [l+1]$. The restrictions $\mathscr{D}^{\tilde{\mathfrak{t}}_k}_{H_1}, \ldots, \mathscr{D}^{\tilde{\mathfrak{t}}_k}_{H_s}$ can be generated from the first $\ell + 1$ draws in a drawing without replacement from the leaves of $\mathfrak{t}_k$.

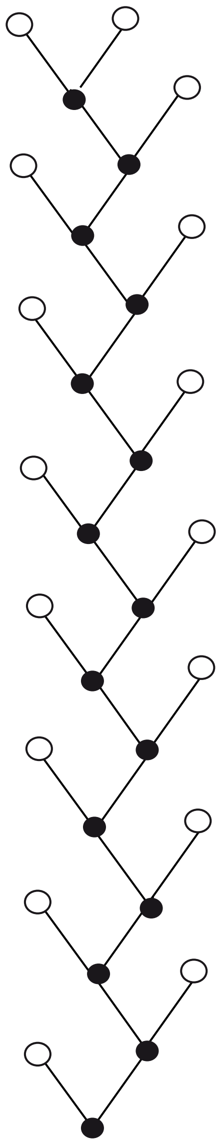
For $k \to \infty$, this can be coupled with high probability to a drawing *with* replacement. For the latter, the resulting restrictions to $H_1, \ldots, H_s$ are clearly independent. $\square$

We just saw:

1. Every point in the D-M boundary corresponds to an infinite Rémy bridge distribution and thus to an exchangeable didentritic system.

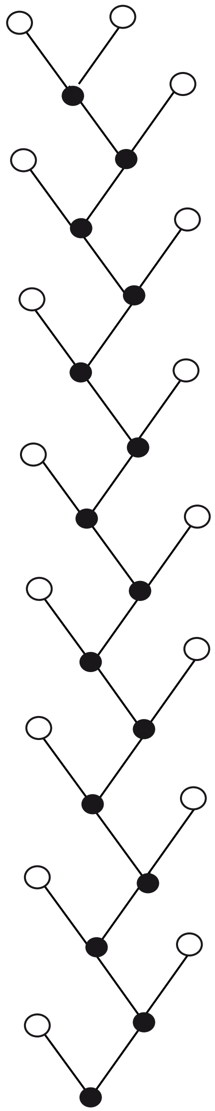
2. **Every** point in the D-M boundary is extremal (resp. ergodic). (In more analytic terms: The **full** D-M boundary is equal to the **minimal** boundary)

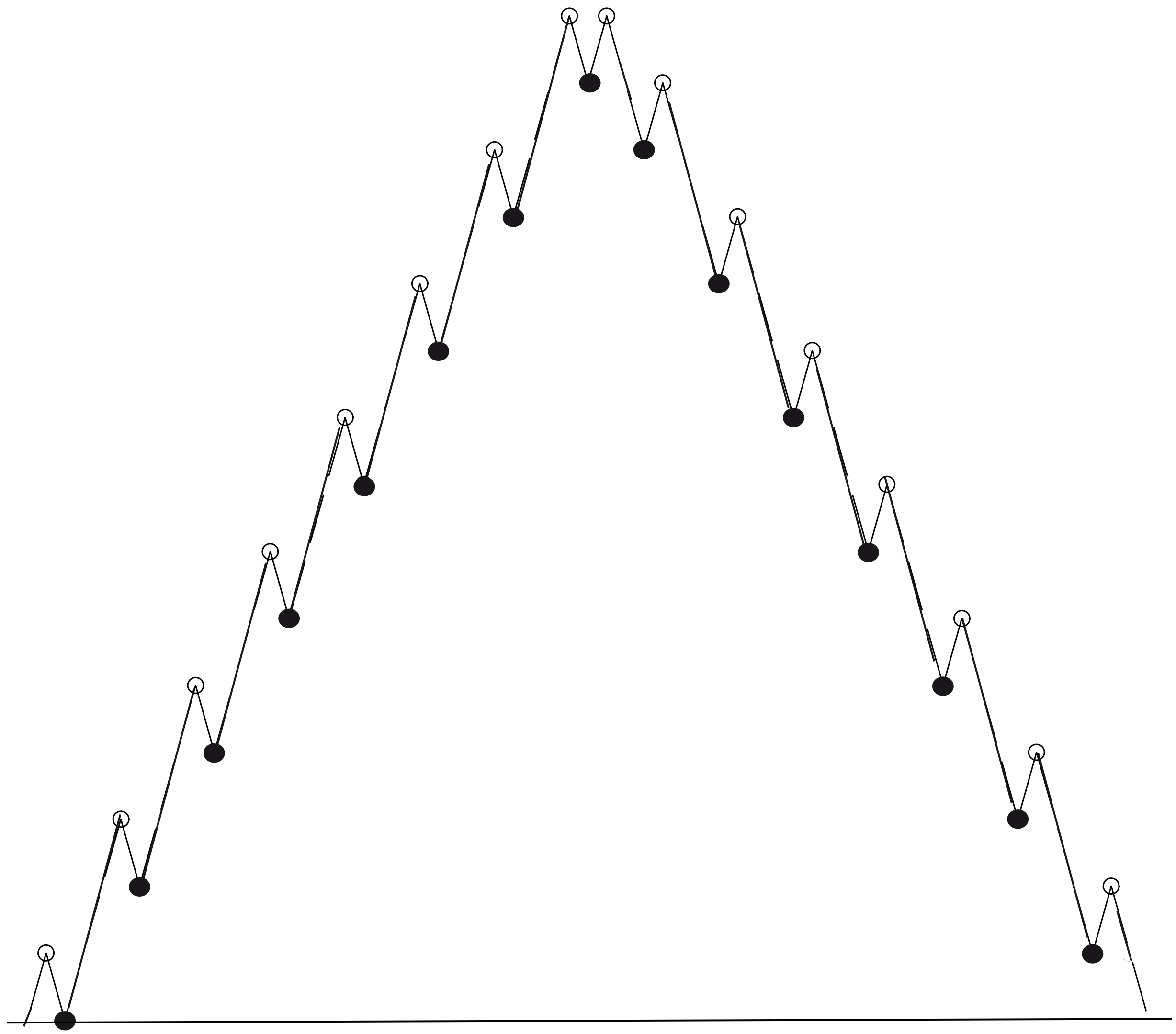
Le Gall's question: What does the D-M-boundary of the Rémy chain look like? thus asks for a representation of the ergodic exchangeable didentritic systems.

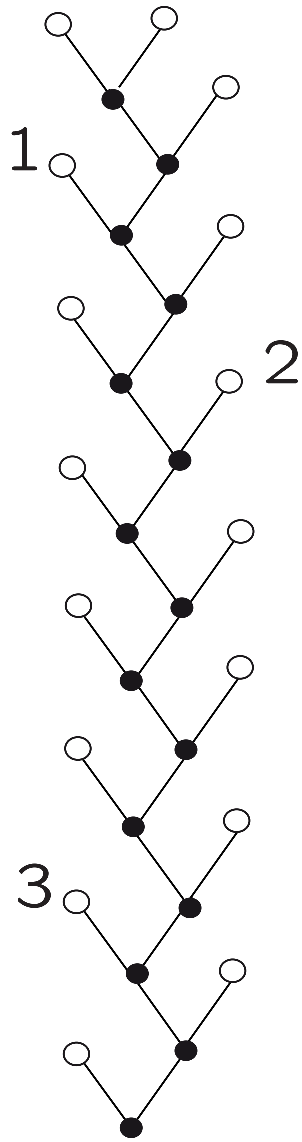


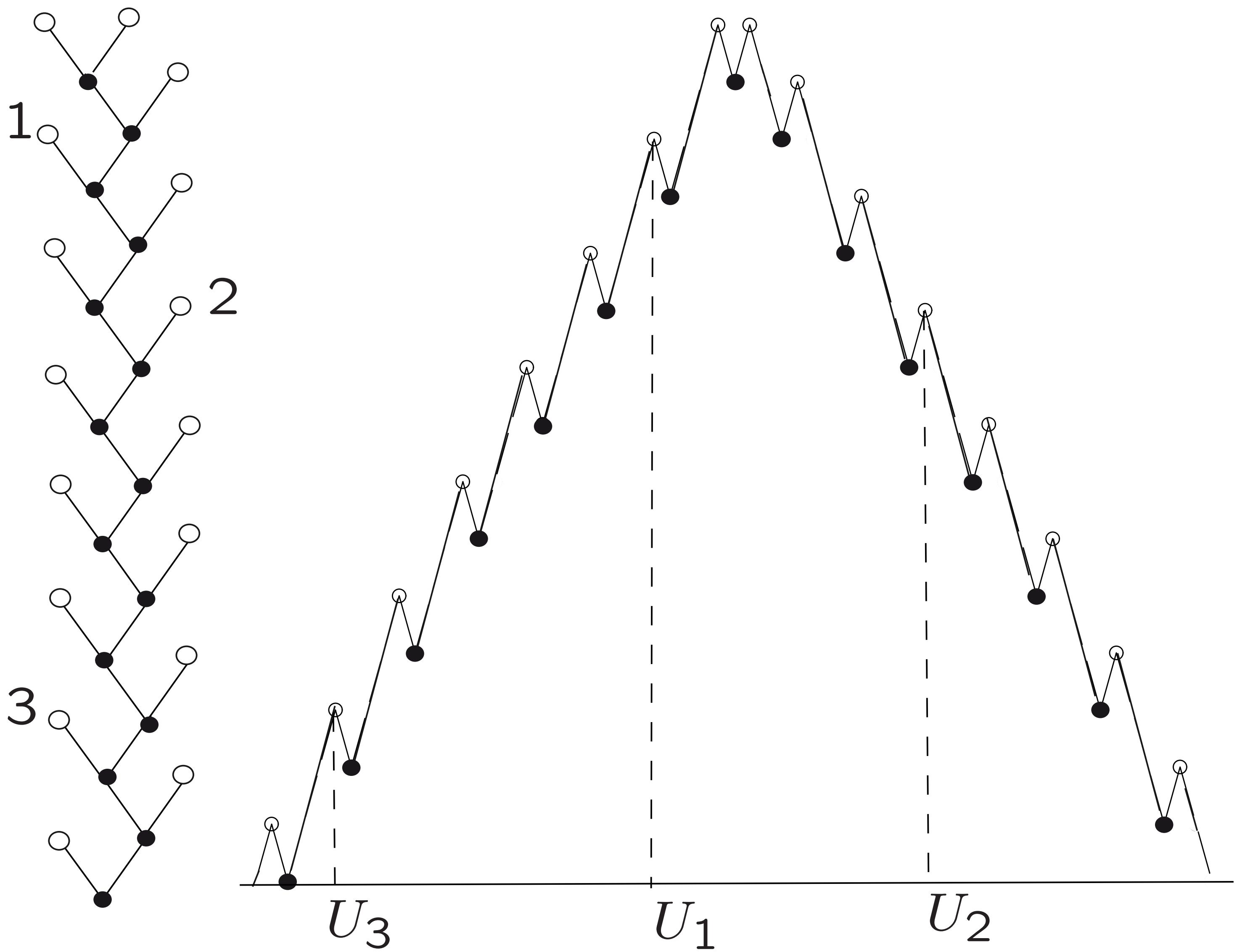
$\mathbf{t}_{16}^{\mathsf{zigzag}}$

A guiding example to our representation theorem is a sequence of "spinal trees" $\mathbf{t}_k^{\mathsf{zigzag}}$, $k = 1, 2, \ldots$

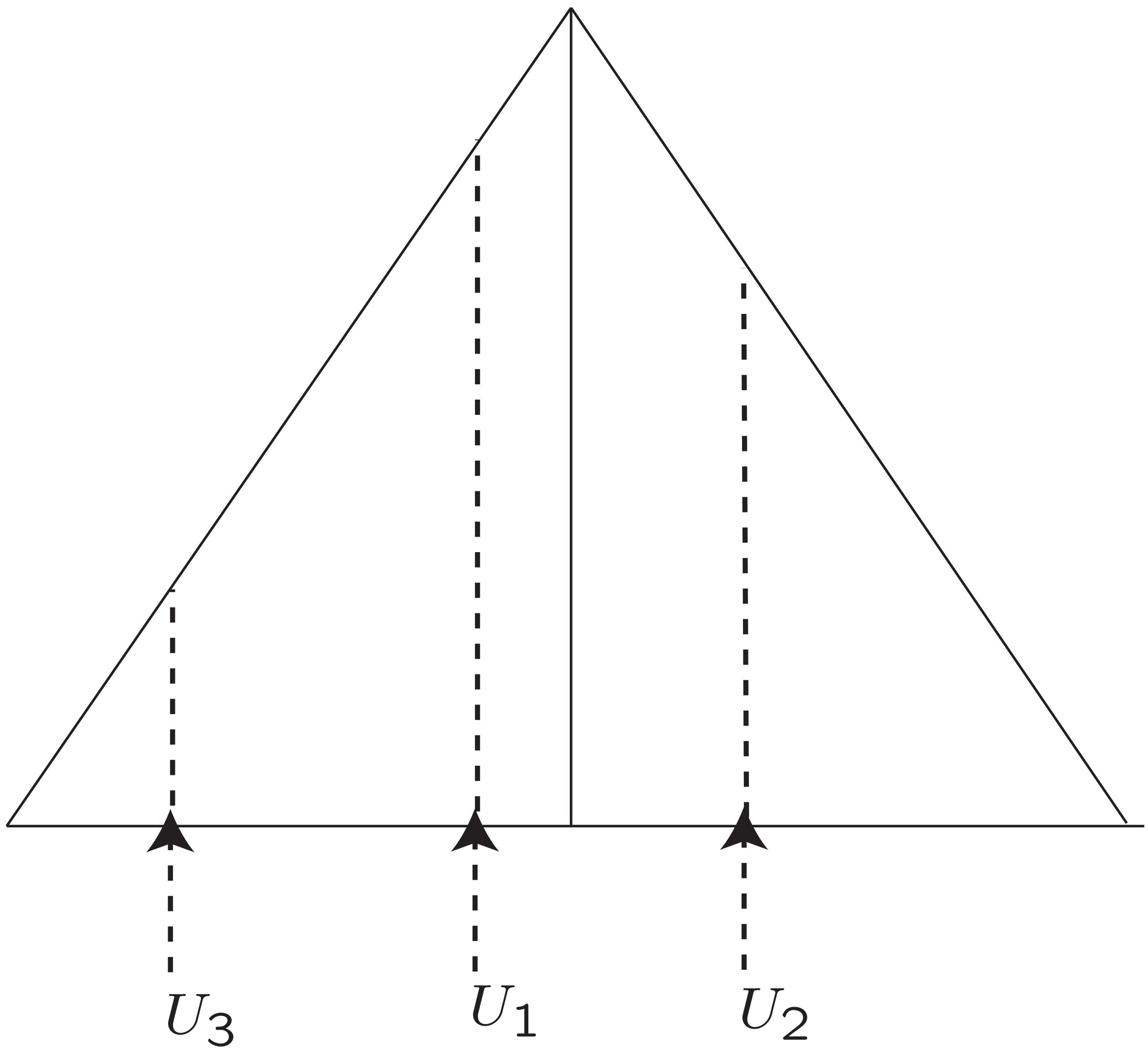


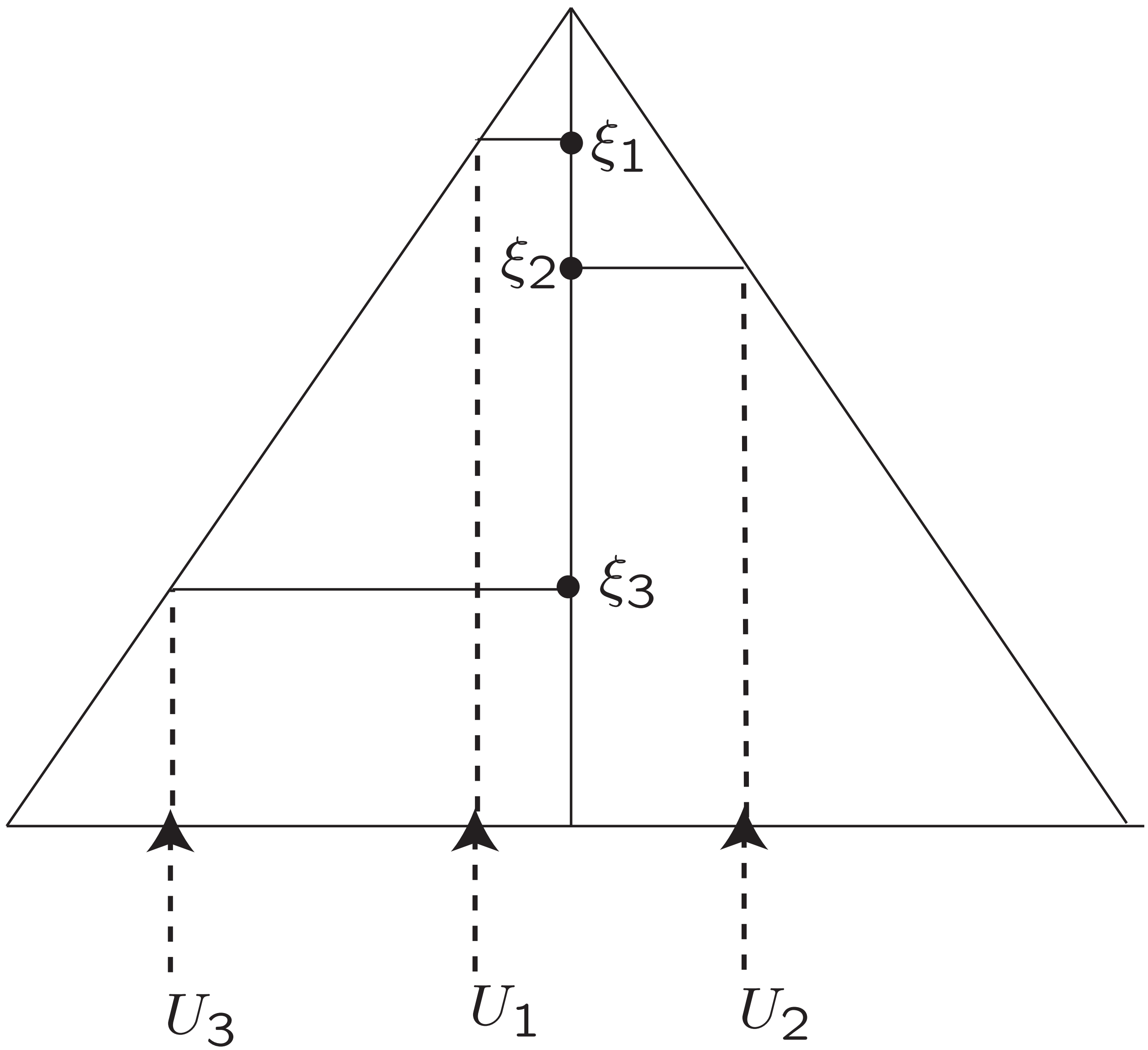


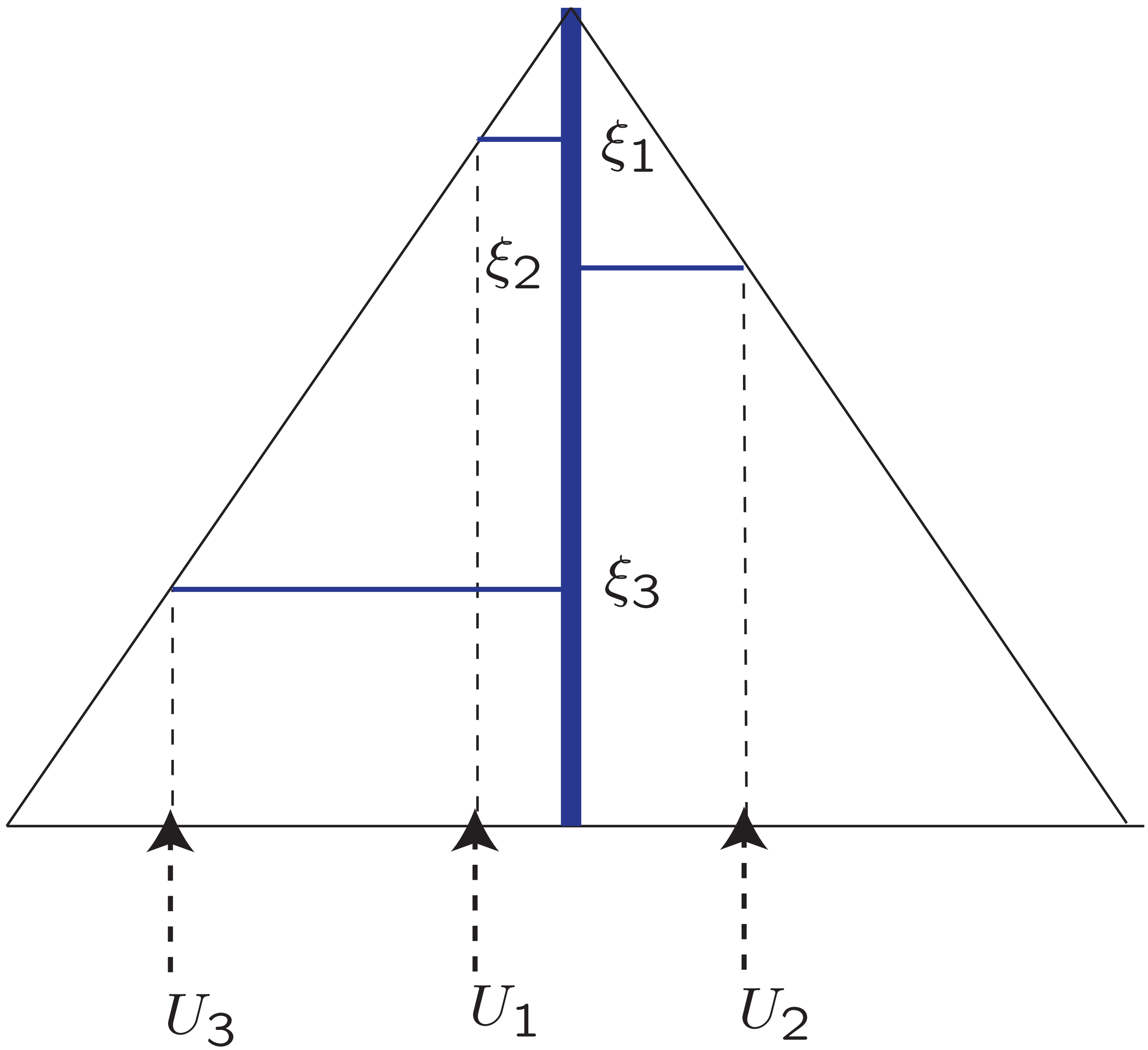

1
2
3



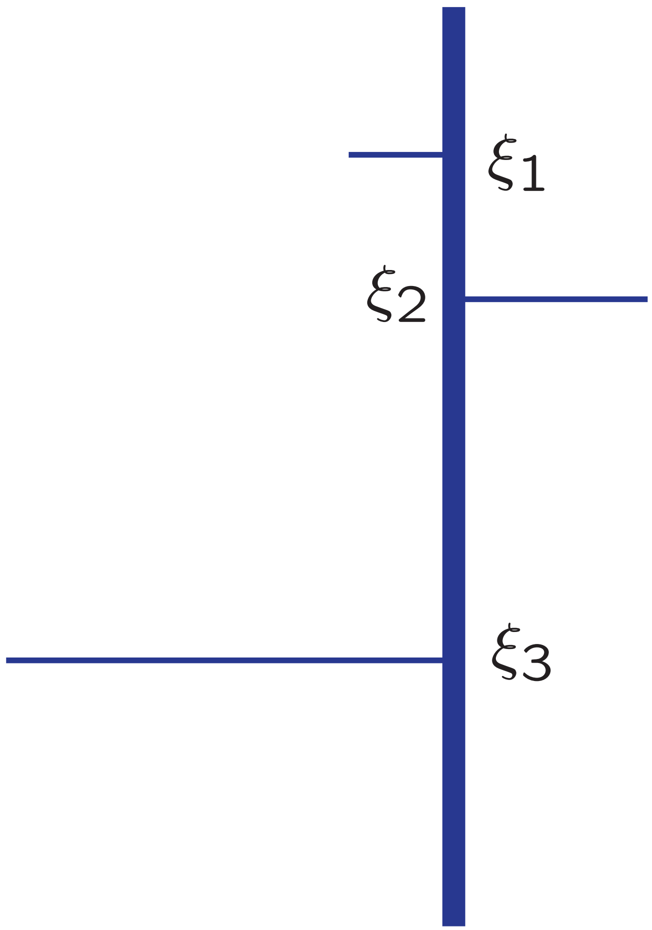
1
2
3
U_3
U_1
U_2


U_3
U_1
U_2


ξ_1
ξ_2
ξ_3
U_3
U_1
U_2



ξ_1
ξ_2
ξ_3
U_3
U_1
U_2



a vertical *core*,

with Lebesgue measure as

sampling measure,

and a continuum of

isolated leaves dangling off

randomly to left or right

Measuring the set of leaves of an ergodic exchangeable didendritic system $\mathscr{D}$

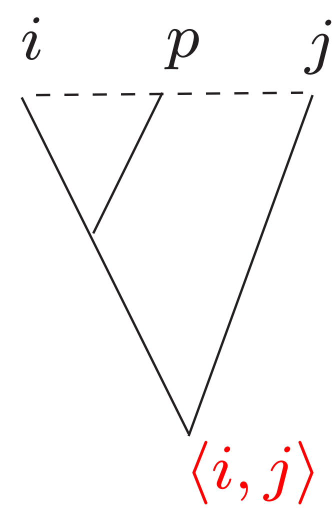
Fix $i, j \in \mathbb{N}$ (and think of i and j as two leaves). How big is the fraction of leaves "between" i and j, i.e. above $\langle i, j \rangle$?

$$I_p := \mathbf{1}_{\{\langle i,j \rangle \leq \langle p,p \rangle\}}, \quad p \in \mathbb{N}$$

is an exchangeable sequence of r.v.'s.

$$d(i,j) := \lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} I_p \quad \text{exists a.s.}$$

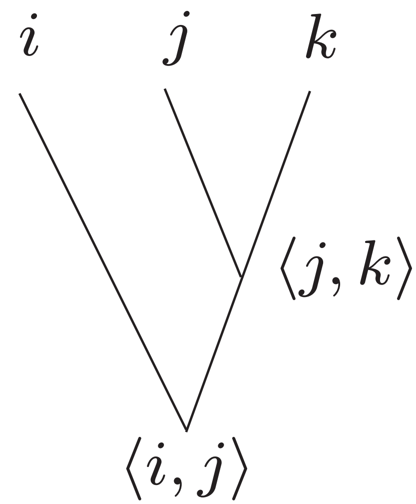
(by de Finetti), is a.s. constant (by ergodicity)



Embedding an ergodic exchangeable DDS $\mathscr{D}$ into an ultrametric tree $\mathbf{T}$:

$$d(i,j) := \lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} \mathbf{1}_{\{\langle i,j\rangle \le \langle p,p\rangle\}},$$

$$i, j \in \mathbb{N}$$



is an exchangeable ultrametric on $\mathbb{N}$:

$$d(i,j) = \max\{d(i,k),\, d(j,k)\}$$

Interpret $\frac{1}{2}d(i,j)$ as the depth of the MRCA of the images of i and j, with an embedding $\mathscr{J} : \mathscr{D} \to \mathbf{T}$ into an ultrametric tree $\mathbf{T}$.

Embedding of an ergodic exchangeable DDS $\mathscr{D}$ into a metric measure space $(\mathbf{S}, d, \mu)$

Let $\mathbf{S}$ be the closure of the subtree of $\mathbf{T}$ that is spanned by $\{\mathscr{J}(\langle i, j\rangle) : i \neq j \in \mathbb{N}\}$.

$(\mathbf{S}, d)$ is an $\mathbb{R}$-tree and

$$\mu(\{x \in \mathbf{S} : \mathscr{J}(\langle i, j\rangle) \leq x\}) := d(i, j)$$

defines a diffuse probability measure μ on $\mathbf{S}$.

Theorem [EGW17] *All* extremal infinite Rémy bridges arise through a successive sampling of leaves $\xi_1, \xi_2, \ldots$ from some complete separable rooted $\mathbb{R}$-tree $\mathbf{S}$, equipped with a diffuse measure μ (the “sampling measure”) and a (randomized) left-right ordering given by a measurable function $W : (\mathbf{S} \times [0,1])^2 \to \{\curvearrowright, \curvearrowleft\}$, and random variables $U_1, U_2, \ldots$ that are independent and uniform on $[0, 1]$. Here,

$$W((\xi_i, U_i), (\xi_j, U_j)) = \curvearrowright$$

prescribes that the i-th sampled leaf is to the left of the common ancestor of leaves i and j.

Here, $\mathbb{S}$, μ and W satisfy the consistency conditions:

(**A**) The tree $\mathbb{S}$ is "almost binary":

Almost surely for distinct $i, j, k \in \mathbb{N}$, precisely one of

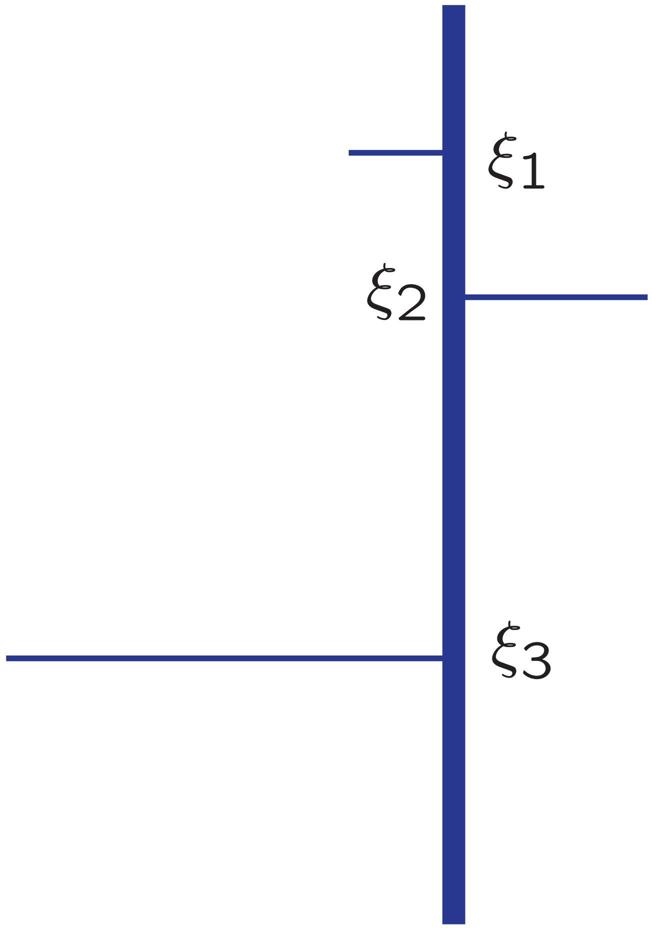
(i) $\quad \xi_i \curlywedge \xi_j = \xi_i \curlywedge \xi_k \prec \xi_j \curlywedge \xi_k,$

(ii) $\quad \xi_j \curlywedge \xi_k = \xi_j \curlywedge \xi_i \prec \xi_k \curlywedge \xi_i,$

(iii) $\quad \xi_k \curlywedge \xi_i = \xi_k \curlywedge \xi_j \prec \xi_i \curlywedge \xi_j$ holds.x

(**B**) The left-right ordering is consistent along the tree:

$$(*) \begin{cases} W(\xi_i, U_i, \xi_j, U_j) = \curvearrowright \iff W(\xi_j, U_j, \xi_i, U_i) = \curvearrowleft \\ \\ \xi_i \curlywedge \xi_j = \xi_i \curlywedge \xi_k \prec \xi_j \curlywedge \xi_k \\ \qquad \Longrightarrow W(\xi_i, U_i, \xi_j, U_j) = W(\xi_i, U_i, \xi_k, U_k) \end{cases}$$



a vertical *core*,

with Lebesgue measure as

sampling measure,

and a continuum of

isolated leaves dangling off

randomly to left or right

$$\mathbf{S} = [0, \tfrac{1}{2}]$$

μ:= Lebesgue measure on S

$$W(\xi_i, U_i, \xi_j, U_j) = \curvearrowright$$

if ξ_i falls below ξ_j and $U_i < \frac{1}{2}$

A different Markov chain – and the same boundary

Let ν be a diffuse measure on $\{0, 1\}^{\mathbb{N}}$.

The PATRICIA(ν)-chain is a

full binary tree - valued Markov chain built from an

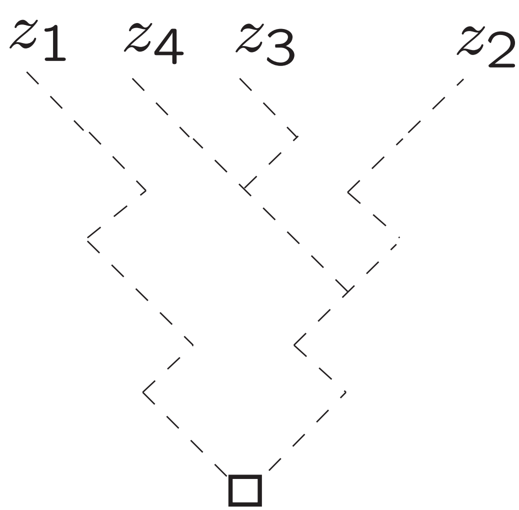
i.i.d. sequence $Z_1, Z_2, \ldots$ of infinite binary words

with common distribution ν

as follows:

For distinct infinite binary words $z_1, z_2, \ldots$, the **radix sort tree** $\mathrm{s} = R(z_1, \ldots, z_n)$ is the minimal binary tree whose n leaves correspond to initial segments of $z_1, \ldots, z_n$.

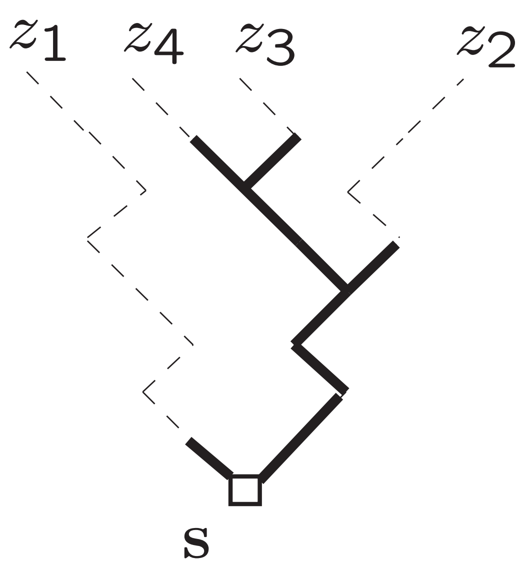
Example (with $n = 4$):



$$z_1 = 00100100\ldots$$
$$z_2 = 11011011\ldots$$
$$z_3 = 11010010\ldots$$
$$z_4 = 11010000\ldots$$

For distinct infinite binary words $z_1, \ldots, z_n$, the **radix sort tree** $\mathbf{s} = \mathbf{R}(z_1, \ldots, z_n)$ is the minimal binary tree whose n leaves correspond to initial segments of $z_1, \ldots, z_n$.

Example (with $n = 4$):



$$z_1 = 00100100\ldots$$

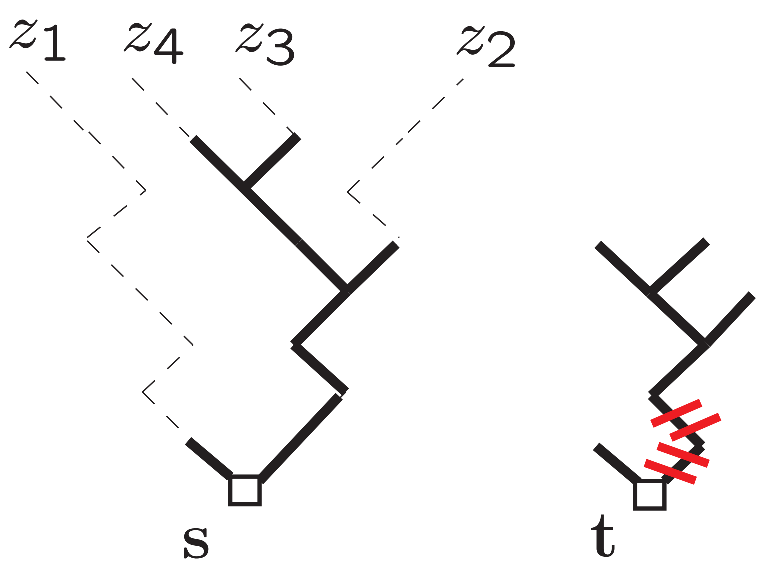
$$z_2 = 11011011\ldots$$

$$z_3 = 11010010\ldots$$

$$z_4 = 11010000\ldots$$

The PATRICIA contraction $\mathbf{\Phi}$ maps the radix sort tree $\mathbf{s}$ into the full binary tree $\mathbf{t} = \mathbf{\Phi}(\mathbf{s})$ by deleting the out-degree 1 vertices and closing up the gaps.

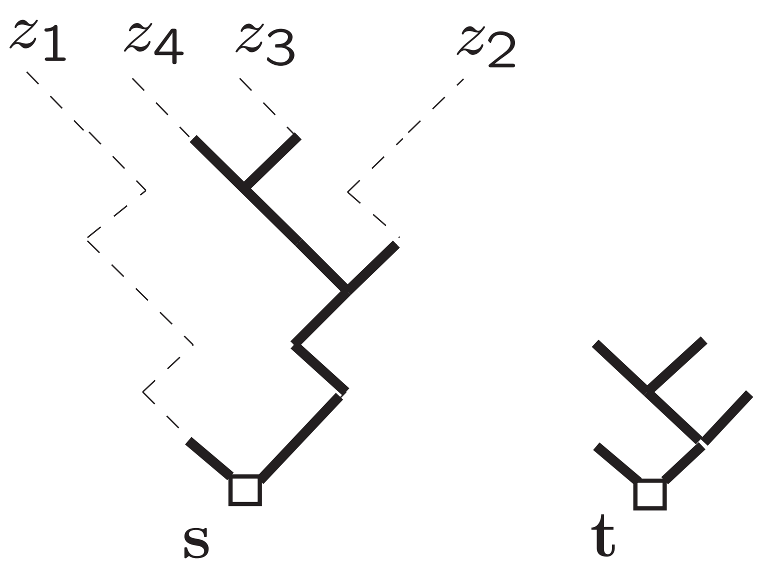
Example (with $n = 4$):



$\mathbf{t} := \mathbf{\Phi}(\mathbf{R}(z_1, \ldots, z_n))$ is the **PATRICIA tree** obtained from the distinct infinite binary words $z_1, \ldots, z_n$.

The PATRICIA contraction Φ maps the radix sort tree s into the full binary tree $\mathbf{t} = \Phi(\mathrm{s})$ by deleting the out-degree 1 vertices and closing up the gaps.

Example (with $n = 4$):



$\mathbf{t} := \Phi(\mathbf{R}(z_1, \ldots, z_n))$ is the **PATRICIA tree** obtained from the distinct infinite binary words $z_1, \ldots, z_n$.

Radix Sort Chains and PATRICIA chains

Let ν be a diffuse measure on $\{0, 1\}^{\mathbb{N}}$
and $Z_1, Z_2, \ldots$ be i.i.d. with distribution ν.

Fact ([EW20, Prop. 3.7])
The radix sort chain ${}^{\nu}R_n := R_n(Z_1, \ldots, Z_n), n = 1, 2, \ldots$,
and the PATRICIA chain ${}^{\nu}P_n := P_n(Z_1, \ldots, Z_n), n = 1, 2, \ldots$
are both Markov chains.

This can be seen from their time reversal. E.g., a backward step of the PATRICIA chain is just an inverse Rémy move.

Fact (Corollary to [EW20, Lemma 3.13]):
$\mathbb{P}({}^{\nu}P_n = \mathbf{t}) > 0$ for any full binary tree $\mathbf{t}$ with n leaves.

Thus the previous observation (on the backward transitions) has a nice consequence:

The D-M boundary of the PATRICIA(ν)-chain equals the D-M boundary of the Rémy chain !

Infinite PATRICIA(ν)-bridges are the same
as infinite Rémy bridges.

Fact: $\mathbf{R}(z_1, \ldots, z_n) = \mathbf{R}(z_{\sigma(1)}, \ldots, z_{\sigma(n)})$ for any permutation of $[n]$, hence by Hewitt-Savage the distribution of $({}^\nu P_n)_{n\in\mathbb{N}}$ is trivial on the tail field $\mathscr{F}^\infty$.

Consequently, any PATRICIA(ν)-chain $({}^\nu P_n)_{n\in\mathbb{N}}$ is an extremal PATRICIA (or Rémy) bridge.

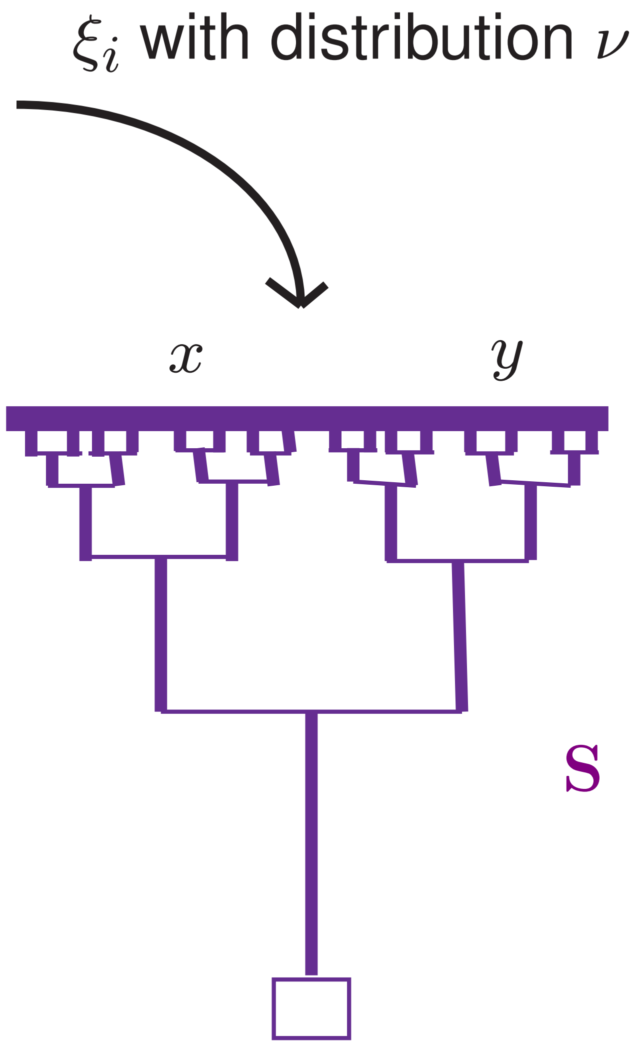
Because of the just stated fact that

$\mathbb{P}({}^\nu P_n = \mathbf{t}) > 0$ for any full binary tree $\mathbf{t}$ with n leaves,

the above described "zigzag bridge" (though being an extremal Rémy (and thus also PATRICIA)-bridge) cannot be a PATRICIA(ν)-chain.

Being an extremal Rémy bridge, any PATRICIA(ν)-chain must have a sampling representation in terms of some $(\mathbf{S}, d, \mu)$. This is as follows:

- $\mathbf{S}$:= completion of the tree with vertices $\in \bigcup_{n \in \mathbb{N}_0} \{0,1\}^n$,
- ultrametric distance d given by
$$d(v_1 \dots v_n, v_1 \dots v_n v_{n+1}) := 2^{-n-2}$$
- sampling measure $\mu := \nu$ on the tree boundary $\{0,1\}^\infty$.



$\mathbf{S}$ . . . the complete ultrametric binary tree with $d(\square, \text{leaves}) = \frac{1}{2}$

ν is a diffuse probability measure on the leaves of $\mathbf{S}$.

$W(\xi_i, U_i, \xi_j, U_j) = \curvearrowleft$

if ξ_i falls to the left of ξ_j

The Doob-Martin boundary of the radix-sort chain

Reference measure γ := fair coin tossing measure on $\{0, 1\}^\infty$.

For any diffuse measure ν on $\{0, 1\}^\infty$, the radix-sort(ν)-chain

... is a radix-sort(γ)-bridge

(because it has the radix-sort(γ) backward dynamics)

... and is even an *extremal* radix-sort(γ)-bridge

(because its tail field is trivial due to the Hewitt-Savage 0-1 law).

Theorem (EW17)

a) Every extremal radix-sort(γ)-bridge

is a radix-sort(ν)-chain for some diffuse ν.

(Consequently, the radix-sort(ν)-chains

are *precisely* the extremal radix-sort(γ)-bridges.)

b) Every limit of a sequence of radix-sort(γ)-bridges

from $\square$ to $\mathbf{t}_m$, $m \to \infty$, is an extremal radix-sort(γ)-bridge.

(In other words: The minimal boundary equals

the full D-M boundary.)

Theorem (EW17)

a) Every extremal radix-sort(γ)-bridge

is a radix-sort(ν)-chain for some diffuse ν.

Idea of proof: Let (R_n^∞) be an extremal RST bridge,

and $(\tilde{R}_n^\infty)$ its labeled version. The latter induces a random

sequence $(\langle i\rangle)_{i\in\mathbb{N}}$ of leaves of the complete binary tree.

By exchangeabilty of this sequence and by extremality of (R_n),

the $\langle i\rangle$, $i = 1, 2, \ldots$, are i.i.d.

For their distribution ν we have $(R_n^\infty) \overset{d}{=} ({}^\nu R_n)$. $\square$

Theorem (EW17)

b) Every limit of a sequence of radix-sort(γ)-bridges from $\square$ to $\mathbf{t}_m$, $m \to \infty$, is an extremal radix-sort(γ)-bridge.

Idea of proof: Let (R_n^∞) be the limit of a sequence of radix-sort(γ) bridges from $\square$ to $\mathbf{t}_m$, $m \to \infty$, and $(\tilde{R}_n^\infty)$ its labeled version. It suffices to show that the associated exchangeable sequence $(\langle i \rangle)_{i \in \mathbb{N}}$ of leaves of the complete binary tree is ergodic. For this, we again use the ergodicity criterion of Aldous, this time with the array

$A_{ij} := \langle i \rangle \wedge \langle j \rangle$ ... the "most recent common ancestor" of the leaves $\langle i \rangle$ and $\langle j \rangle$. $\square$

To summarize:

The D-M boundary of the radix-sort tree chain
with fair coin tossing strings as input
is in 1-1 correspondence to the radix-sort tree chains
with i.i.d. (ν)-input, with ν a diffuse measure on $\{0, 1\}^\infty$.

[EGW17] S. N. Evans, R. Grübel and A.W.,
Doob-Martin boundary of Rémy’s tree growth chain.
Ann. Probab. 45 (2017), 225-277, arXiv:1411.2526 [math.PR]

[EW17] S.N. Evans and A.W., Radix sort trees in the large.
Electron. Commun. Probab. 22 (2017)

[EW20], S.N. Evans and A.W., PATRICIA bridges.
In: Genealogies of Interacting Particle Systems,
eds. M. Birkner, R. Sun and J. Swart, pp. 233-267,
World Scientific 2020, arXiv:1806.06256 [math.PR]