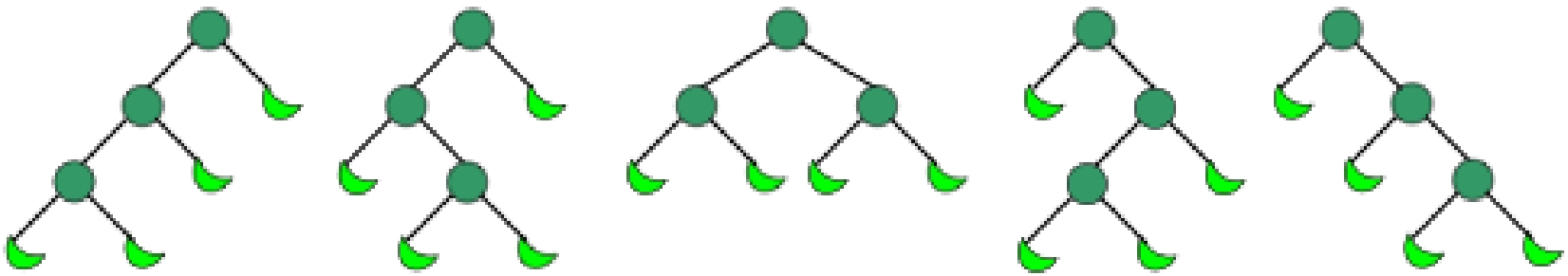
Big trees

Steven N. Evans

February, 2017



Collaborators

Hye Soo Choi (Berkeley)

Rudolf Grűbel (Hannover)

Anton Wakolbinger (Frankfurt)

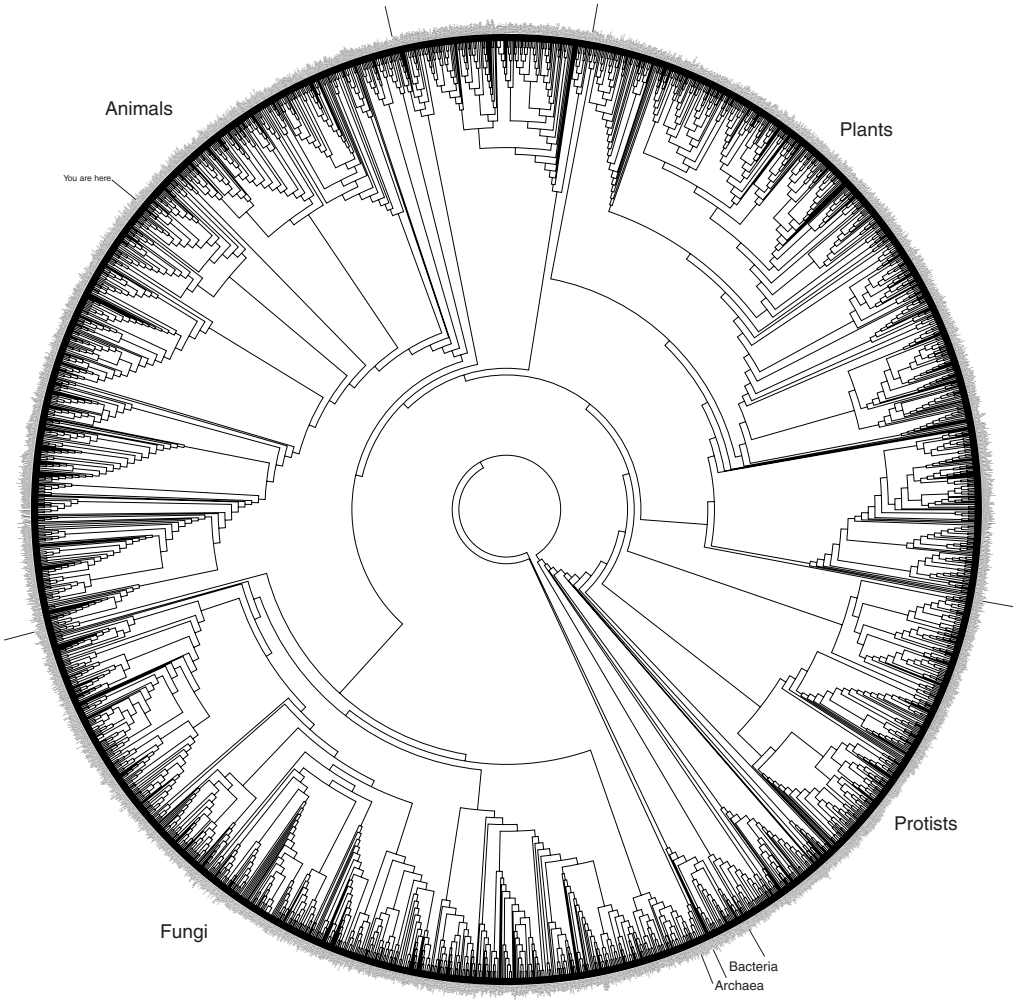


Figure: Big trees abound in biology (and many other disciplines). What does it mean to say that two big trees are similar / different? Is there a PRINICIPLED way to make this distinction?

All who wander are not lost

- A sequence of points on the real line wanders off to infinity if it eventually leaves any bounded set, never to return.
- Can we delineate different ways in which such sequences wander off to infinity?
- Yes! The sequence may converge to $+\infty$, converge to $-\infty$, or neither.

All who wander are not lost

- A sequence of points on the real line wanders off to infinity if it eventually leaves any bounded set, never to return.
- Can we delineate different ways in which such sequences wander off to infinity?
- Yes! The sequence may converge to $+\infty$, converge to $-\infty$, or neither.

All who wander are not lost

- A sequence of points on the real line wanders off to infinity if it eventually leaves any bounded set, never to return.
- Can we delineate different ways in which such sequences wander off to infinity?
- Yes! The sequence may converge to $+\infty$, converge to $-\infty$, or neither.

Packing the real line into an interval

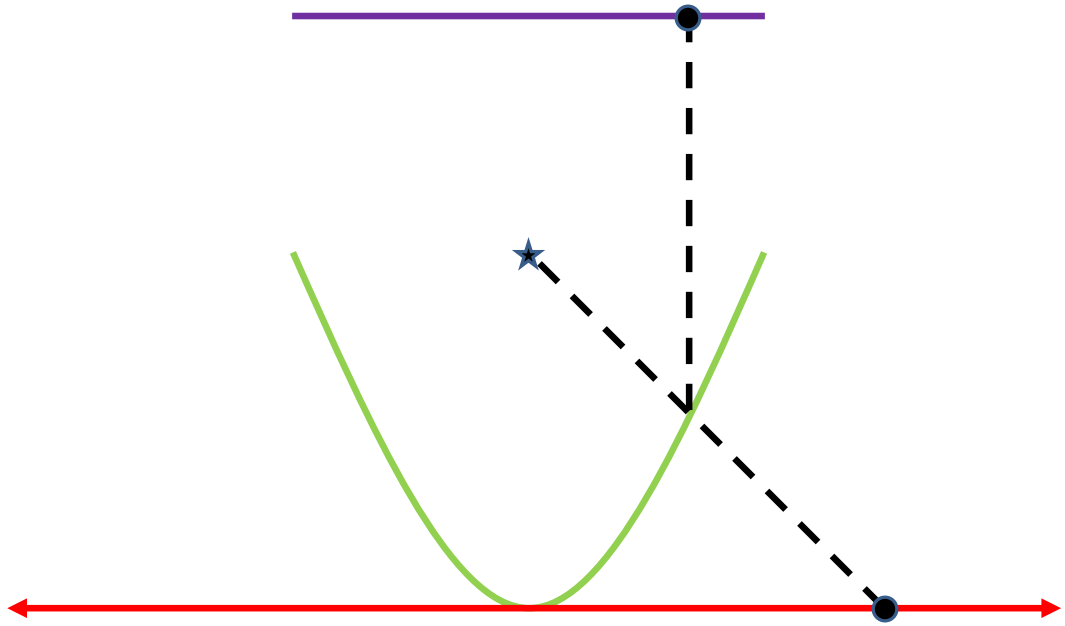
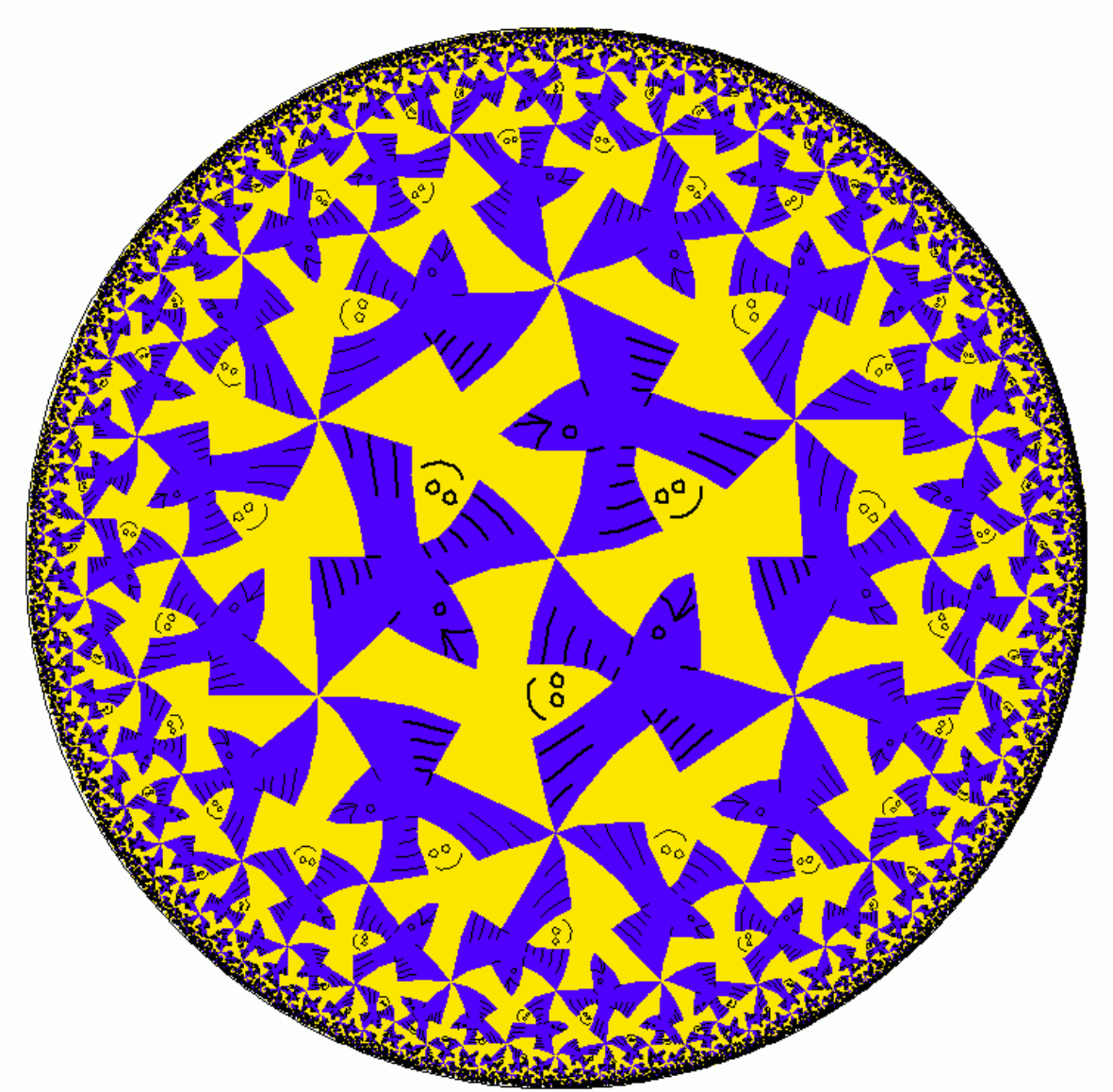


Figure: We can map the whole real line bijectively to an (open) interval. Converging to $\pm\infty$ corresponds to converging (in the usual sense) to one of the end-points of the interval.

Packing the plane into a disk

- We can map the plane bijectively to an (open) disk
- A sequence wandering off to infinity converges if it has an asymptotic direction.
- Convergence corresponds to converging (in the usual sense) to one of the boundary points of the disk.

An illustration of the basic idea



What if the plane has a crack in it?

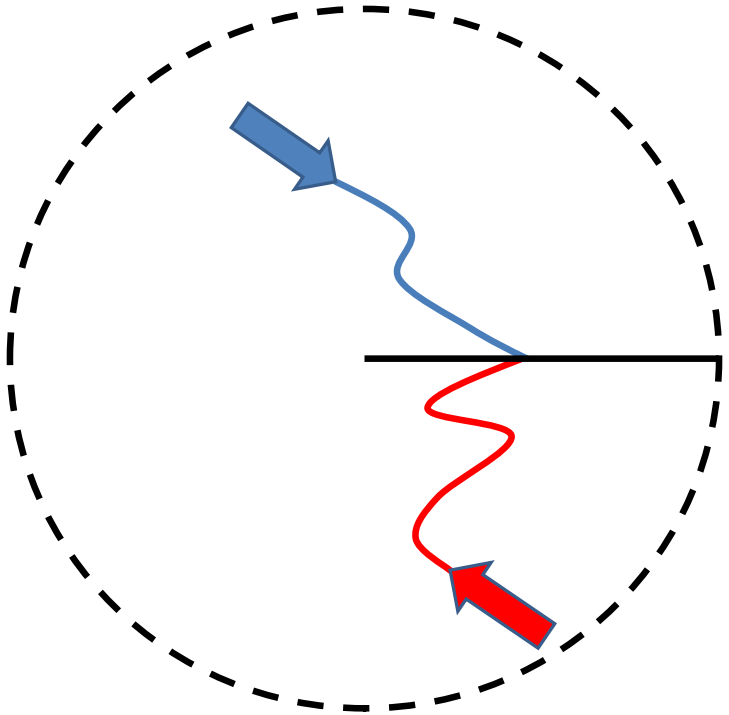


Figure: We remove a half-line from the plane and map the result to the disk. Do these two curves wander off to the same destination or different destinations?

How we want to / should think of the slit disk

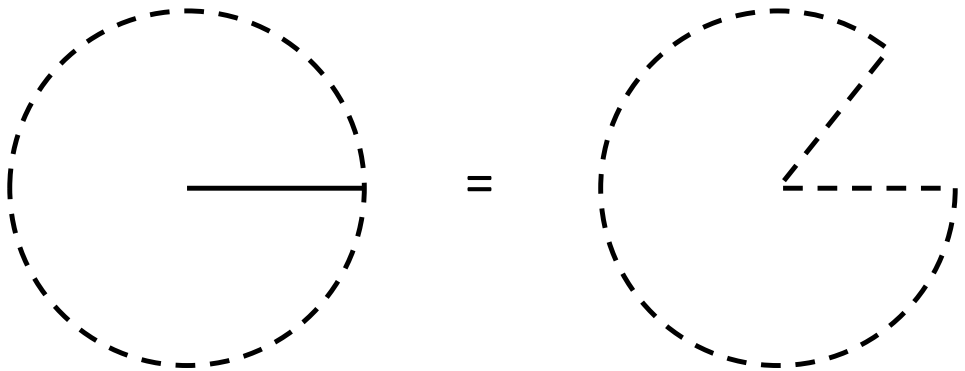


Figure: The disk with a slit removed is "really" the disk with a sector removed.

Brownian motion

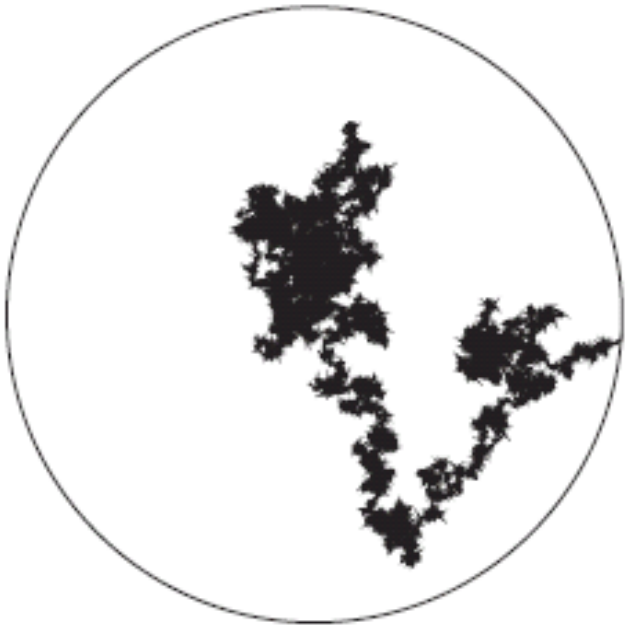


Figure: Brownian motion in d-dimensions is the continuous-time, continuous-space analogue of simple random walk

Brownian motion sees the slit

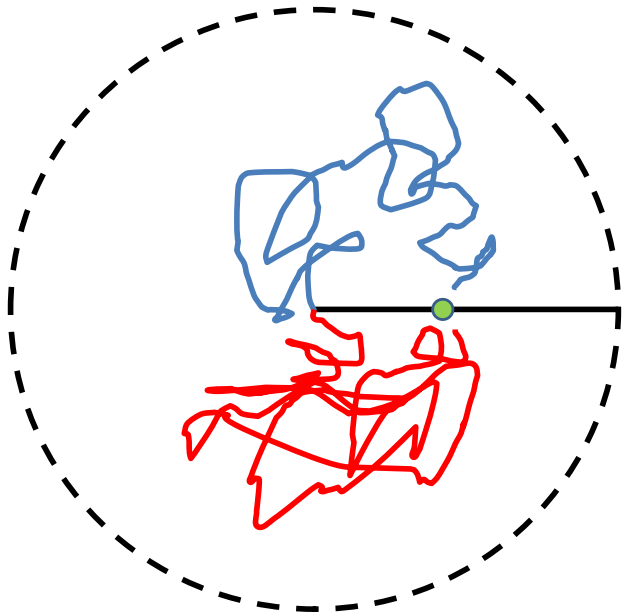


Figure: Brownian motion conditioned to visit a point close to and above the marked point (prior to exiting the slit disk) behaves differently – even at the beginning – to Brownian motion conditioned to visit a point close to and below the marked point (prior to exiting the slit disk).

What is the general philosophy?

- We have some infinite space of objects E.
- We want a way of saying what it means for a sequence of objects (x_n) that wanders off to infinity to "converge".
- We will then have a way of deciding when two large objects are similar / different.

What is the general philosophy?

- We have some infinite space of objects E.
- We want a way of saying what it means for a sequence of objects (x_n) that wanders off to infinity to "converge".
- We will then have a way of deciding when two large objects are similar / different.

What is the general philosophy?

- We have some infinite space of objects E.
- We want a way of saying what it means for a sequence of objects (x_n) that wanders off to infinity to "converge".
- We will then have a way of deciding when two large objects are similar / different.

Implementing the general philosophy

- We take a Markov chain $(X_k)_{k=0}^{\infty}$ with state-space E that wanders off to infinity, define $(X_k^n)_{k=0}^{\infty}$ to be $(X_k)_{k=0}^{\infty}$ conditioned to visit x_n, and say that (x_n) converges if the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ.
- For two sequences, the same / different ensemble of limiting distributions correspond to the same / different limit points.

Implementing the general philosophy

- We take a Markov chain $(X_k)_{k=0}^{\infty}$ with state-space E that wanders off to infinity, define $(X_k^n)_{k=0}^{\infty}$ to be $(X_k)_{k=0}^{\infty}$ conditioned to visit x_n, and say that (x_n) converges if the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ.
- For two sequences, the same / different ensemble of limiting distributions correspond to the same / different limit points.

Infinite bridges

- If the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ, there is an infinite bridge $(X_k^\infty)_{k=0}^\infty$ such that the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ to the joint probability distribution of $(X_0^\infty, \ldots, X_\ell^\infty)$ for every ℓ.
- An infinite bridges has the same backward transition probabilities as the original chain:

$$\mathbb{P}\{X_n^\infty = i \,|\, X_{n+1}^\infty = j\} = \mathbb{P}\{X_n = i \,|\, X_{n+1} = j\}.$$

- Determining the possible limit points in our sense $\Longleftrightarrow$ determining the set of infinite bridge distributions $\Longleftrightarrow$ determining the Markov chains that have the same backward transition probabilities as the original chain $(X_k)_{k=0}^\infty$.

Infinite bridges

- If the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ, there is an infinite bridge $(X_k^\infty)_{k=0}^\infty$ such that the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ to the joint probability distribution of $(X_0^\infty, \ldots, X_\ell^\infty)$ for every ℓ.
- An infinite bridges has the same backward transition probabilities as the original chain:

$$\mathbb{P}\{X_n^\infty = i \,|\, X_{n+1}^\infty = j\} = \mathbb{P}\{X_n = i \,|\, X_{n+1} = j\}.$$

- Determining the possible limit points in our sense $\Longleftrightarrow$ determining the set of infinite bridge distributions $\Longleftrightarrow$ determining the Markov chains that have the same backward transition probabilities as the original chain $(X_k)_{k=0}^\infty$.

Infinite bridges

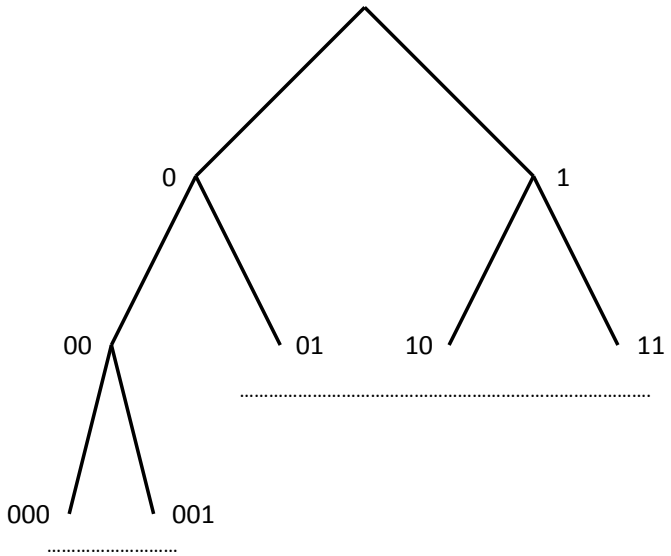
- If the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ, there is an infinite bridge $(X_k^\infty)_{k=0}^\infty$ such that the joint probability distribution of $(X_0^n, \ldots, X_\ell^n)$ converges for every ℓ to the joint probability distribution of $(X_0^\infty, \ldots, X_\ell^\infty)$ for every ℓ.
- An infinite bridges has the same backward transition probabilities as the original chain:

$$\mathbb{P}\{X_n^\infty = i \,|\, X_{n+1}^\infty = j\} = \mathbb{P}\{X_n = i \,|\, X_{n+1} = j\}.$$

- Determining the possible limit points in our sense $\Longleftrightarrow$ determining the set of infinite bridge distributions $\Longleftrightarrow$ determining the Markov chains that have the same backward transition probabilities as the original chain $(X_k)_{k=0}^\infty$.

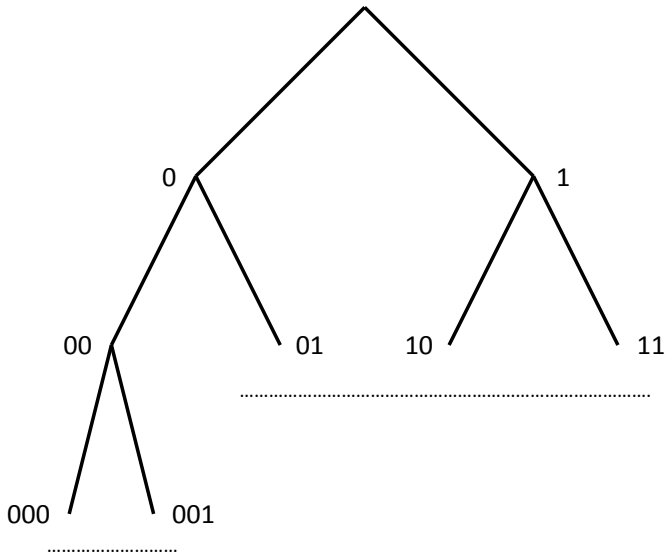
Trees at last, trees at last, thank God almighty ...

- For concreteness sake, let's consider trees that are rooted, binary (each vertex has 0 or 2 children), and ordered (we distinguish between a left child and a right child).
- We can identify such a tree as a subtree of the complete rooted binary tree.



Trees at last, trees at last, thank God almighty ...

- For concreteness sake, let's consider trees that are rooted, binary (each vertex has 0 or 2 children), and ordered (we distinguish between a left child and a right child).
- We can identify such a tree as a subtree of the complete rooted binary tree.



How many binary trees are there?

The number of binary trees with $2n+1$ vertices (and hence $n+1$ leaves) is the n^{th} Catalan number $\frac{1}{n+1}\binom{2n}{n}$.

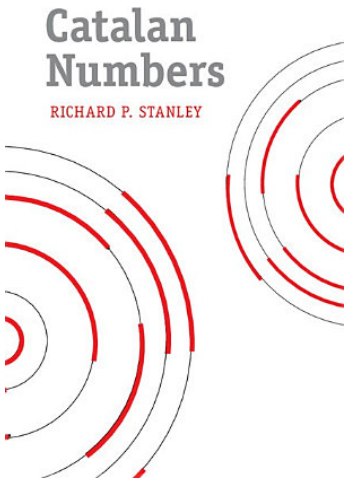


Figure: *Catalan numbers are probably the most ubiquitous sequence of numbers in mathematics. This book gives for the first time a comprehensive collection of their properties and applications to combinatorics, algebra, analysis, number theory, probability theory, geometry, topology, and other areas. [...] the book presents 214 different kinds of objects counted by them [...]*

Rémy's algorithm

Rémy's (1985) algorithm iteratively generates a sequence of random binary trees $(U_n)_{n=1}^{\infty}$ such that U_n is uniformly distributed on the set of binary trees with $2n+1$ vertices.

Example of one iteration of Rémy's algorithm

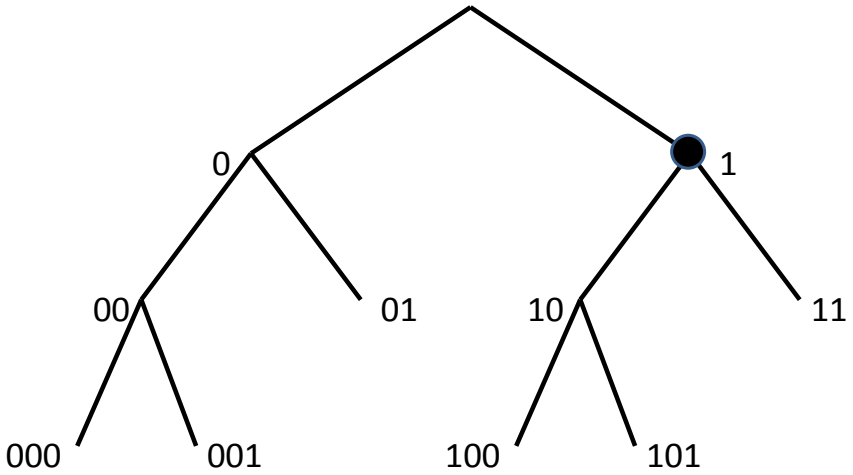


Figure: First step in an iteration of Rémy's algorithm: pick a vertex v uniformly at random.

Example of one iteration of Rémy's algorithm – continued

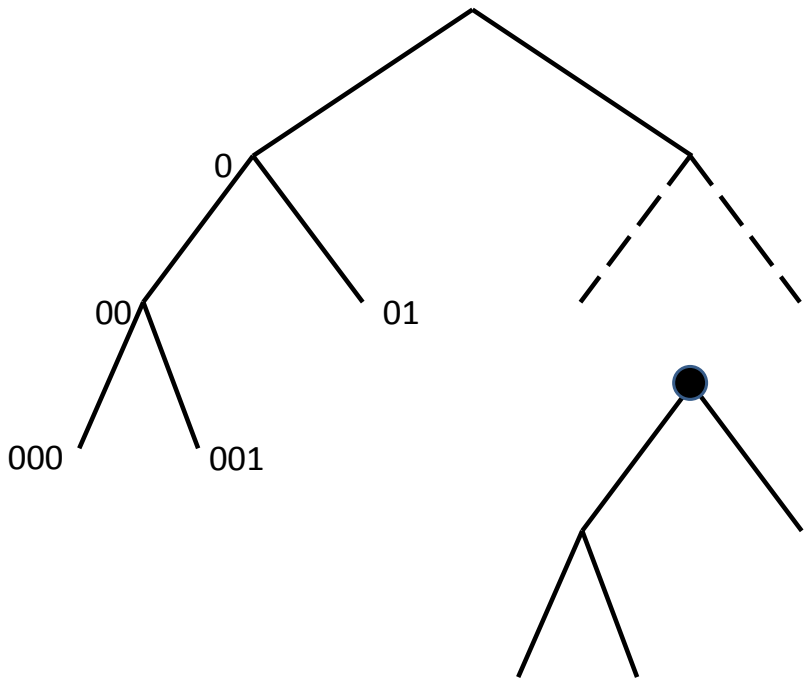


Figure: Second step in an iteration of Rémy's algorithm: cut off the subtree rooted at v and attach a copy of the binary tree with 3 vertices to the end of the edge that previously led to v.

Example of one iteration of Rémy's algorithm – continued

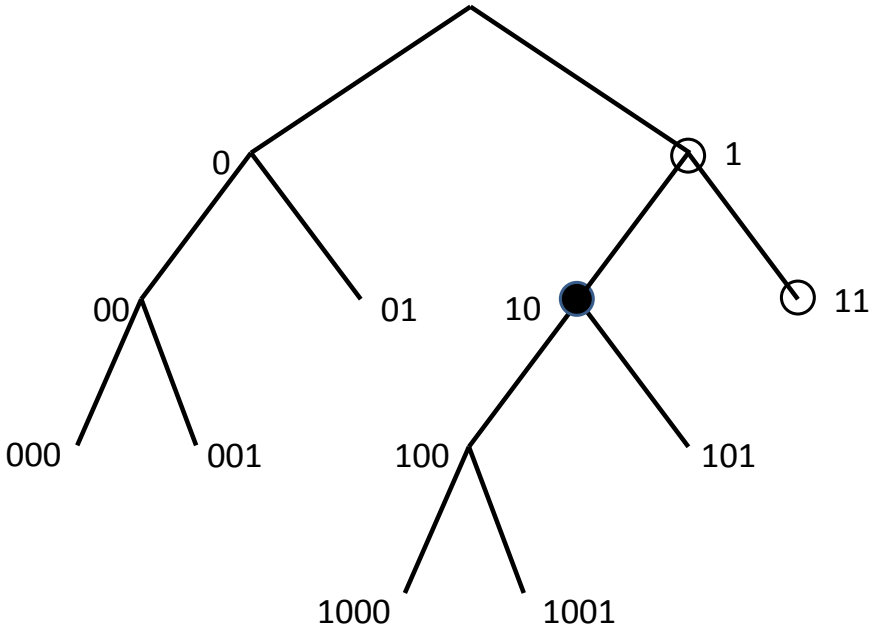


Figure: Third step in an iteration of Rémy's algorithm: re-attach the subtree rooted at v to one of the two leaves of the copy of the 3-vertex tree, and re-label the vertices appropriately. The solid circle is the new location of v and the open circles are the clones of v.

Example of a backward transition for Rémy's algorithm

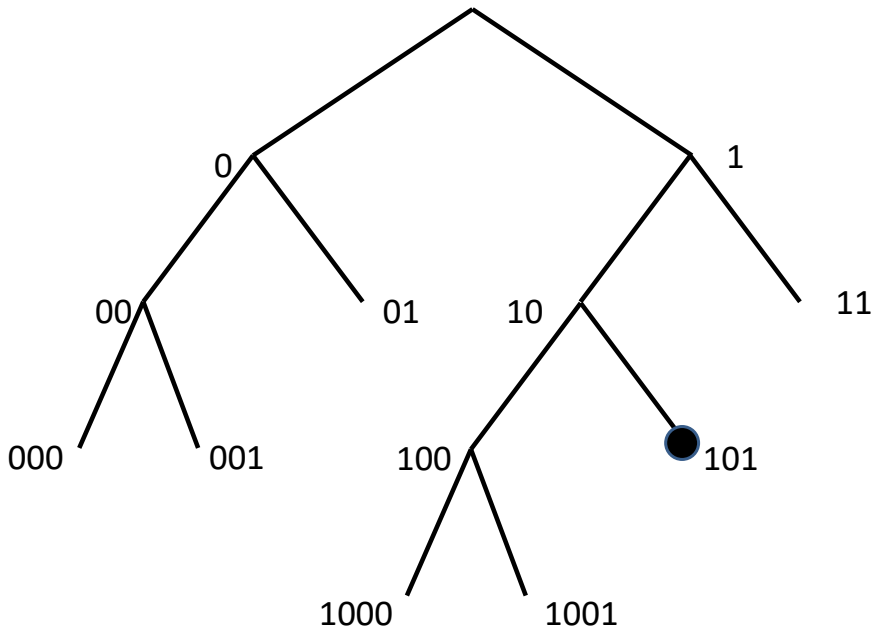


Figure: First step in a backward transition of Rémy's algorithm: pick a leaf w uniformly at random.

Example of a backward transition – continued

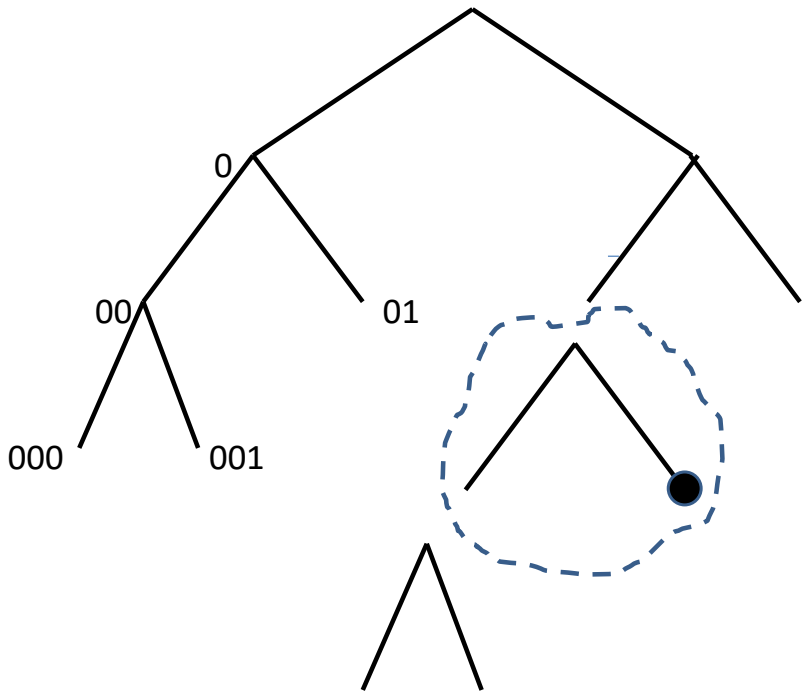


Figure: Second step in a backward transition of Rémy's algorithm: delete the chosen leaf w and its sibling.

Example of a backward transition – continued

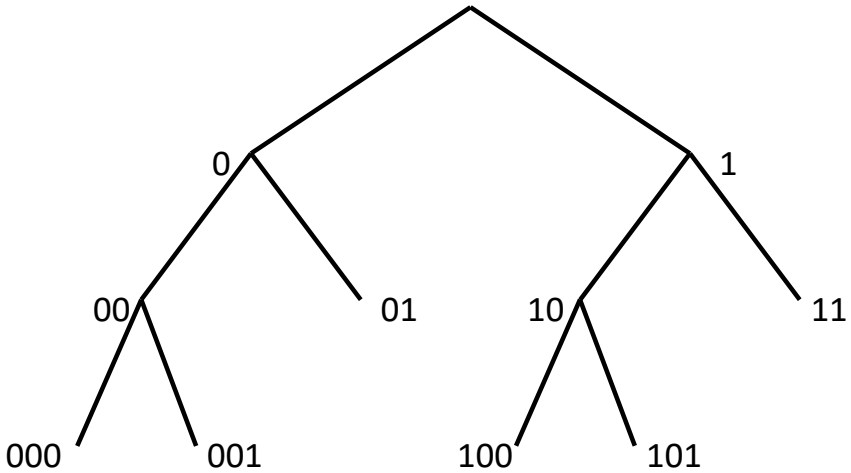


Figure: Third step in a backward transition of Rémy's algorithm: close up the gap.

Embedding one binary tree into another

An embedding of a binary tree $\mathbf{s}$ into a binary tree $\mathbf{t}$ is a map from the vertex set of $\mathbf{s}$ into the vertex set of $\mathbf{t}$ such that the following hold.

- The image of a leaf of $\mathbf{s}$ is a leaf of $\mathbf{t}$.
- 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 $\mathbf{t}$.

Examples of embeddings

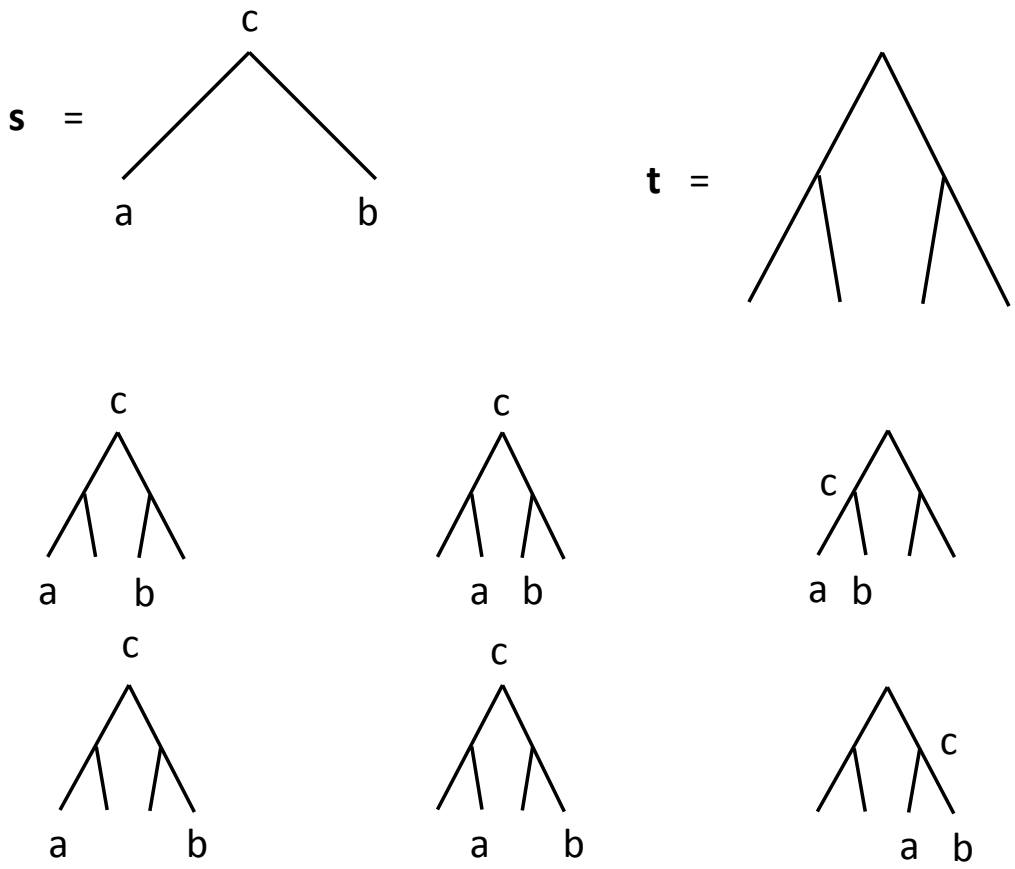


Figure: All the embeddings of the unique binary tree $\mathbf{s}$ with 3 vertices into a particular tree $\mathbf{t}$ with 7 vertices.

How does our philosophy play out for the Rémy chain?

- Suppose that $(\mathbf{t}_k)_{k=1}^{\infty}$ is a sequence of binary trees, where $\mathbf{t}_k$ has $n_k + 1$ leaves and $n_k \to \infty$ as $k \to \infty$. Then $(\mathbf{t}_k)_{k=1}^{\infty}$ converges in the sense we have been discussing if and only if for each binary tree $\mathbf{s}$ the following limit exists

$$\pi(\mathbf{s}; (\mathbf{t}_k)_{k=1}^{\infty}) := \lim_{k\to\infty} \frac{\#\text{ of embeddings of } \mathbf{s} \text{ into } \mathbf{t}_k}{\binom{n_k+1}{m+1}},$$

  where $\mathbf{s}$ has $m + 1$ leaves.

- Two sequences $(\mathbf{t}'_k)_{k=1}^{\infty}$ and $(\mathbf{t}''_k)_{k=1}^{\infty}$ converge to the same limit if and only if for all binary trees $\mathbf{s}$

$$\pi(\mathbf{s}; (\mathbf{t}'_k)_{k=1}^{\infty}) = \pi(\mathbf{s}; (\mathbf{t}''_k)_{k=1}^{\infty}).$$

How does our philosophy play out for the Rémy chain?

- Suppose that $(\mathbf{t}_k)_{k=1}^{\infty}$ is a sequence of binary trees, where $\mathbf{t}_k$ has $n_k + 1$ leaves and $n_k \to \infty$ as $k \to \infty$. Then $(\mathbf{t}_k)_{k=1}^{\infty}$ converges in the sense we have been discussing if and only if for each binary tree $\mathbf{s}$ the following limit exists

$$\pi(\mathbf{s}; (\mathbf{t}_k)_{k=1}^{\infty}) := \lim_{k\to\infty} \frac{\#\text{ of embeddings of } \mathbf{s} \text{ into } \mathbf{t}_k}{\binom{n_k+1}{m+1}},$$

  where $\mathbf{s}$ has $m + 1$ leaves.

- Two sequences $(\mathbf{t}'_k)_{k=1}^{\infty}$ and $(\mathbf{t}''_k)_{k=1}^{\infty}$ converge to the same limit if and only if for all binary trees $\mathbf{s}$

$$\pi(\mathbf{s}; (\mathbf{t}'_k)_{k=1}^{\infty}) = \pi(\mathbf{s}; (\mathbf{t}''_k)_{k=1}^{\infty}).$$

The sampling perspective

- Equivalently, $(\mathbf{t}_k)_{k=1}^{\infty}$ converges if and only if for every m the random subtree spanned by $m+1$ leaves of $\mathbf{t}_k$ chosen uniformly at random without replacement converges in distribution as $k \to \infty$.

Graph limits and metric measure spaces connection / analogy

- The notion of convergence witnessed by convergence in distribution of randomly sampled sub-objects is also a key idea in the area of graph limits: Borgs, Chayes, Diaconis, Janson, Lovász, Sós, Szegedy, Tao, Vesztergombi, ...
- There is a concrete characterization of the possible limit objects in terms of $\mathbb{R}$-trees with extra structure. This is analogous to the appearance of graphons in the theory of graph limits.
- Convergence in Gromov and Vershik's theory of metric measure spaces is also witnessed by convergence in distribution of randomly sampled sub-objects.

Graph limits and metric measure spaces connection / analogy

- The notion of convergence witnessed by convergence in distribution of randomly sampled sub-objects is also a key idea in the area of graph limits: Borgs, Chayes, Diaconis, Janson, Lovász, Sós, Szegedy, Tao, Vesztergombi, ...
- There is a concrete characterization of the possible limit objects in terms of $\mathbb{R}$-trees with extra structure. This is analogous to the appearance of graphons in the theory of graph limits.
- Convergence in Gromov and Vershik's theory of metric measure spaces is also witnessed by convergence in distribution of randomly sampled sub-objects.

Graph limits and metric measure spaces connection / analogy

- The notion of convergence witnessed by convergence in distribution of randomly sampled sub-objects is also a key idea in the area of graph limits: Borgs, Chayes, Diaconis, Janson, Lovász, Sós, Szegedy, Tao, Vesztergombi, ...
- There is a concrete characterization of the possible limit objects in terms of $\mathbb{R}$-trees with extra structure. This is analogous to the appearance of graphons in the theory of graph limits.
- Convergence in Gromov and Vershik's theory of metric measure spaces is also witnessed by convergence in distribution of randomly sampled sub-objects.

Convergence notions induced by other Markov chains

Similar results, but with different notions of convergence arise if the Rémy chain is replaced by other tree–valued Markov chains such as:

- random binary search trees,
- random digital search trees,
- random radix sort trees,
- preferential attachment trees (a.k.a. nested Chinese restaurant processes),
- trees associated with shuffle products and Hopf algebras,...

Future

Turning these perspectives into statistical procedures that can be used with data is uncharted territory. For example, for each binary tree $\mathbf{s}$ we have the statistic

$$\mathbf{t} \mapsto \# \text{ of embeddings of } \mathbf{s} \text{ into } \mathbf{t}.$$

- How do we choose between / combine these for a particular purpose?
- Can we use them to build tractable exponential families?
- Is there a version of this using Markov chains more adapted to phylogenies?