Pauline Sperry Lecture

UC Berkeley, 7th April 2025

Christina Goldschmidt

UNIVERSITY OF OXFORD
DEPARTMENT OF STATISTICS

SIMONS LAUFER MATHEMATICAL SCIENCES INSTITUTE

"The greatest gift to mankind – the freedom of the mind – is in great peril. If we lost that we lose everything. The universities are its greatest bulwark. They are the first to be attacked. The battle is only just begun."

Pauline Sperry (1953)

Scaling limits in probability, with applications to random trees

Probabilistic limits

Let $(X_n)_{n \geq 1}$ be a sequence of real-valued random variables.

There are several different notions of convergence for such a sequence. You may have encountered, for example, convergence in distribution, convergence in probability and almost sure convergence. The first of these will be our focus today.

Probabilistic limits

Let $(X_n)_{n\geq 1}$ be a sequence of real-valued random variables.

There are several different notions of convergence for such a sequence. You may have encountered, for example, convergence in distribution, convergence in probability and almost sure convergence. The first of these will be our focus today.

Convergence in distribution (weak convergence):
$X_n \xrightarrow{d} X$ if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ at every point of continuity of the limit cumulative distribution function
$F(x) = \mathbb{P}(X \leq x), x \in \mathbb{R}$.

Example

Suppose that G_n models the number of independent trials with probability $1/n$ of success in a single trial until we see the first success. G_n has a Geometric distribution with success probability $p = 1/n$.

Example

Suppose that G_n models the number of independent trials with probability $1/n$ of success in a single trial until we see the first success. G_n has a Geometric distribution with success probability $p = 1/n$.

Then $\frac{G_n}{n} \xrightarrow{d} E$ as $n \to \infty$, where $E \sim \text{Exp}(1)$ has cumulative distribution function

$$F(x) = 1 - \exp(-x), \quad x \geq 0.$$

Example

Proof.
Fix $x \geq 0$. Then

$$\begin{aligned}\mathbb{P}\left(\frac{G_n}{n} \leq x\right) = \mathbb{P}\left(G_n \leq nx\right) &= \mathbb{P}\left(G_n \leq \lfloor nx \rfloor\right) \\ &= 1 - \mathbb{P}\left(G_n > \lfloor nx \rfloor\right) \\ &= 1 - \left(1 - \frac{1}{n}\right)^{\lfloor nx \rfloor} \\ &\to 1 - \exp(-x),\end{aligned}$$

as $n \to \infty$, which is the cumulative distribution function of Exp(1).
(Since the limit cdf is continuous, we need this for every $x \geq 0$.) □

Random mathematical objects

Question: what if we want to deal with random mathematical objects which are not real-valued?

Random mathematical objects

How can we talk about the distribution of a more complicated random mathematical object? Think: a random vector, a random function, a random set, a random surface, . . .

Random mathematical objects

How can we talk about the distribution of a more complicated random mathematical object? Think: a random vector, a random function, a random set, a random surface, ...

The most general set-up is as follows: let (M, d) be a (complete, separable) metric space. We want to characterise the distribution of a random element X of M.

Random mathematical objects

How can we talk about the distribution of a more complicated random mathematical object? Think: a random vector, a random function, a random set, a random surface, ...

The most general set-up is as follows: let (M, d) be a (complete, separable) metric space. We want to characterise the distribution of a random element X of M.

If X can take only countably many possible values, it's straightforward: for each of those values x, we just need to know $\mathbb{P}(X = x)$.

Random mathematical objects

How can we talk about the distribution of a more complicated random mathematical object? Think: a random vector, a random function, a random set, a random surface, ...

The most general set-up is as follows: let (M, d) be a (complete, separable) metric space. We want to characterise the distribution of a random element X of M.

If X can take only countably many possible values, it's straightforward: for each of those values x, we just need to know $\mathbb{P}(X = x)$. But there are lots of interesting situations in which there will be uncountably many possibilities.

Random mathematical objects

How can we talk about the distribution of a more complicated random mathematical object? Think: a random vector, a random function, a random set, a random surface, ...

The most general set-up is as follows: let (M, d) be a (complete, separable) metric space. We want to characterise the distribution of a random element X of M.

If X can take only countably many possible values, it's straightforward: for each of those values x, we just need to know $\mathbb{P}(X = x)$. But there are lots of interesting situations in which there will be uncountably many possibilities.

In general, it turns out that it's sufficient to know the values of $\mathbb{E}[\psi(X)]$ for a sufficiently rich class of test-functions $\psi : M \to \mathbb{R}$.

Random mathematical objects

How can we talk about the distribution of a more complicated random mathematical object? Think: a random vector, a random function, a random set, a random surface, ...

The most general set-up is as follows: let (M, d) be a (complete, separable) metric space. We want to characterise the distribution of a random element X of M.

If X can take only countably many possible values, it's straightforward: for each of those values x, we just need to know $\mathbb{P}(X = x)$. But there are lots of interesting situations in which there will be uncountably many possibilities.

In general, it turns out that it's sufficient to know the values of $\mathbb{E}[\psi(X)]$ for a sufficiently rich class of test-functions $\psi : M \to \mathbb{R}$. This seems a bit obscure at first sight, but you may have already seen a couple of common examples of this idea: generating functions.

Examples

Probability generating functions:

For a random variable X taking values in $\{0, 1, 2, \ldots\}$, let

$$G(s) = \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} p_k s^k, \quad |s| \leq 1,$$

where $p_k = \mathbb{P}(X = k)$, $k \geq 0$. Then G completely determines the distribution of X.

Examples

Probability generating functions:

For a random variable X taking values in $\{0, 1, 2, \dots\}$, let

$$G(s) = \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} p_k s^k, \quad |s| \le 1,$$

where $p_k = \mathbb{P}(X = k)$, $k \ge 0$. Then G completely determines the distribution of X. To see this, let and note that

$$G(s) = p_0 + p_1 s + p_2 s^2 + \cdots,$$

so that we may recover $(p_k)_{k \ge 0}$ by differentiating:
$p_k = G^{(k)}(0)/k!$.

Examples

Probability generating functions:
For a random variable X taking values in $\{0, 1, 2, \ldots\}$, let

$$G(s) = \mathbb{E}\left[s^X\right] = \sum_{k=0}^{\infty} p_k s^k, \quad |s| \leq 1,$$

where $p_k = \mathbb{P}(X = k)$, $k \geq 0$. Then G completely determines the distribution of X. To see this, let and note that

$$G(s) = p_0 + p_1 s + p_2 s^2 + \cdots,$$

so that we may recover $(p_k)_{k \geq 0}$ by differentiating:
$p_k = G^{(k)}(0)/k!$.

Moment generating functions (Laplace transforms):
For a random variable X taking values in $\mathbb{R}_+$, let

$$M(\theta) = \mathbb{E}\left[\exp(-\theta X)\right], \quad \theta \geq 0.$$

Then M completely determines the distribution of X (via Laplace inversion).

Convergence in distribution

For real-valued random variables:

Convergence in distribution (weak convergence):
$X_n \xrightarrow{d} X$ if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ at every point of continuity of the limit cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$.

As it stands, this definition doesn't generalise.

Convergence in distribution

For real-valued random variables:

Convergence in distribution (weak convergence):
$X_n \xrightarrow{d} X$ if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ at every point of continuity of the limit cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$.

As it stands, this definition doesn't generalise. But it turns out to be equivalent to the following:

$\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$ for every bounded continuous function $\phi : \mathbb{R} \to \mathbb{R}$

which is much more amenable to generalisation.

Example continued

Indeed, for non-negative random variables, it's sufficient to have $\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$ for $\phi(x) = \exp(-\theta x)$ and all $\theta \geq 0$. (This is an instance of the continuity theorem for moment generating functions.)

Recall: G_n has a Geometric distribution with success probability $p = 1/n$. Then $\frac{G_n}{n} \xrightarrow{d} E$ where $E \sim \text{Exp}(1)$ as $n \to \infty$.

Alternative proof.

We have

$$\begin{aligned}
\mathbb{E}[\exp(-\theta G_n/n)] = \mathbb{E}\left[(e^{-\theta/n})^{G_n}\right] &= \frac{\frac{1}{n}e^{-\theta/n}}{1 - e^{-\theta/n}(1-1/n)} \\
&= \frac{1}{n(e^{\theta/n} - 1) + 1} \\
&\to \frac{1}{\theta + 1} = \mathbb{E}[\exp(-\theta E)].
\end{aligned}$$

□

Convergence in distribution

Definition. Let (M, d) be a metric space. Suppose that $(X_n)_{n\geq 1}$ and X are random elements of M. Then we say that X_n converges in distribution (or converges weakly) to X if

$$\mathbb{E}\left[\phi(X_n)\right] \to \mathbb{E}\left[\phi(X)\right]$$

for every bounded continuous function $\phi : M \to \mathbb{R}$.

Convergence in distribution

Definition. Let (M, d) be a metric space. Suppose that $(X_n)_{n \geq 1}$ and X are random elements of M. Then we say that X_n converges in distribution (or converges weakly) to X if

$$\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$$

for every bounded continuous function $\phi : M \to \mathbb{R}$.

Examples of metric spaces:

Convergence in distribution

Definition. Let (M, d) be a metric space. Suppose that $(X_n)_{n \geq 1}$ and X are random elements of M. Then we say that X_n converges in distribution (or converges weakly) to X if

$$\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$$

for every bounded continuous function $\phi : M \to \mathbb{R}$.

Examples of metric spaces:

1. $\mathbb{R}^n$ with the Euclidean distance: random vectors.

Convergence in distribution

Definition. Let (M, d) be a metric space. Suppose that $(X_n)_{n \geq 1}$ and X are random elements of M. Then we say that X_n converges in distribution (or converges weakly) to X if

$$\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$$

for every bounded continuous function $\phi : M \to \mathbb{R}$.

Examples of metric spaces:

1. $\mathbb{R}^n$ with the Euclidean distance: random vectors.
2. $C([0,1], \mathbb{R})$ with the supremum norm: random continuous functions.

Convergence in distribution

Definition. Let (M, d) be a metric space. Suppose that $(X_n)_{n \geq 1}$ and X are random elements of M. Then we say that X_n converges in distribution (or converges weakly) to X if

$$\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$$

for every bounded continuous function $\phi : M \to \mathbb{R}$.

Examples of metric spaces:

1. $\mathbb{R}^n$ with the Euclidean distance: random vectors.
2. $C([0, 1], \mathbb{R})$ with the supremum norm: random continuous functions.
3. Compact sets in $\mathbb{R}^n$ with the Hausdorff distance: random compact sets.

What is a scaling limit?

We are given random elements $(X_n)_{n \geq 1}$ of a metric space (M, d) which are "growing" in some sense, and we have a notion of a scaling operation (i.e. we can "stretch" and "shrink" elements of M).

What is a scaling limit?

We are given random elements $(X_n)_{n \geq 1}$ of a metric space (M, d) which are "growing" in some sense, and we have a notion of a scaling operation (i.e. we can "stretch" and "shrink" elements of M).

A random element X of M is called the scaling limit of $(X_n)_{n \geq 1}$ if there exists some deterministic sequence of real numbers $(a_n)_{n \geq 1}$ such that $a_n \to 0$ and

$$a_n \, X_n \xrightarrow{d} X$$

as $n \to \infty$.

What is a scaling limit?

We are given random elements $(X_n)_{n\geq 1}$ of a metric space (M, d) which are "growing" in some sense, and we have a notion of a scaling operation (i.e. we can "stretch" and "shrink" elements of M).

A random element X of M is called the scaling limit of $(X_n)_{n\geq 1}$ if there exists some deterministic sequence of real numbers $(a_n)_{n\geq 1}$ such that $a_n \to 0$ and

$$a_n \, X_n \xrightarrow{d} X$$

as $n \to \infty$.

We've actually already seen an example: we had that if $G_n \sim \text{Geom}(1/n)$ then $\frac{1}{n} G_n \xrightarrow{d} E$ where $E \sim \text{Exp}(1)$ as $n \to \infty$.

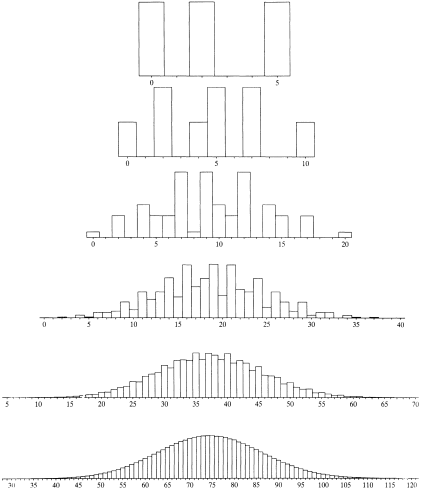
A prototypical example

The central limit theorem. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with $\mathbb{E}[X_1] = 0$ and $\mathrm{var}(X_1) = \sigma^2 \in (0, \infty)$. Let $S_n = \sum_{i=1}^{n} X_i$. Then

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, 1).$$

The CLT

Figure 5. Distribution of S_n for $n = 1, 2, 4, 8, 16, 32$.



[Taken from Jim Pitman's *Probability*]

The CLT

Some comments:

The CLT

Some comments:

- We say that the scaling limit of $(S_n)_{n \geq 1}$ is normally distributed.

The CLT

Some comments:

- We say that the scaling limit of $(S_n)_{n\geq 1}$ is normally distributed.
- S_n can take both positive and negative values for any n, so it's not growing in the simplest sense of the word!

The CLT

Some comments:

- We say that the scaling limit of $(S_n)_{n\geq 1}$ is normally distributed.
- S_n can take both positive and negative values for any n, so it's not growing in the simplest sense of the word!
- We have $\text{var}(S_n) = n\sigma^2$, so the variance *is* growing.

The CLT

Some comments:

- We say that the scaling limit of $(S_n)_{n\geq 1}$ is normally distributed.
- S_n can take both positive and negative values for any n, so it's not growing in the simplest sense of the word!
- We have $\mathrm{var}(S_n) = n\sigma^2$, so the variance *is* growing. Our normalisation has precisely the effect of making the variance equal to 1 for every n.

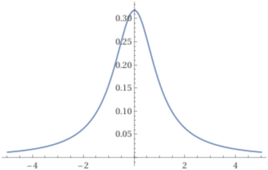
The CLT

Some comments:

- We say that the scaling limit of $(S_n)_{n\geq 1}$ is normally distributed.
- S_n can take both positive and negative values for any n, so it's not growing in the simplest sense of the word!
- We have $\text{var}(S_n) = n\sigma^2$, so the variance *is* growing. Our normalisation has precisely the effect of making the variance equal to 1 for every n.
- The distribution of the X_i's appears only through the variance. This phenomenon is referred to as universality.

What if the conditions of the CLT aren't satisfied?

The classic example is the so-called Cauchy distribution, which has probability density function $f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$.



Although the density is symmetric around 0, it doesn't have a well-defined expectation (since $\int_{-\infty}^{\infty} |x| f(x) dx = \infty$) and therefore doesn't have a finite variance either.

Moreover, if $X_1, X_2, \ldots$ are i.i.d. Cauchy random variables then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \quad \text{has the same distribution as } X_1.$$

(So there's no chance that $(X_1 + \cdots + X_n)/\sqrt{n}$ will converge.)

What if the variance doesn't exist?

There are several other possible scaling limits for sums of i.i.d. random variables!

What if the variance doesn't exist?

There are several other possible scaling limits for sums of i.i.d. random variables!

For example, suppose that X takes values in $\{-1, 0, 1, 2, \ldots\}$ and is such that

- $\mathbb{E}[X_1] = 0$
- $\mathbb{P}(X_1 = k) \sim ck^{-\alpha-1}$ as $k \to \infty$ for $c > 0$ and $\alpha \in (1, 2)$.

Then $\text{var}(X_1) = \sum_{k=-1}^{\infty} k^2 \mathbb{P}(X_1 = k)$. We have $\sum_{k=k_0}^{\infty} k^{1-\alpha} = \infty$ for any $k_0 \geq 1$, and $1 - \alpha \in (-1, 0)$, so the variance is infinite.

What if the variance doesn't exist?

There are several other possible scaling limits for sums of i.i.d. random variables!

For example, suppose that X takes values in $\{-1, 0, 1, 2, \ldots\}$ and is such that

- $\mathbb{E}[X_1] = 0$
- $\mathbb{P}(X_1 = k) \sim ck^{-\alpha-1}$ as $k \to \infty$ for $c > 0$ and $\alpha \in (1, 2)$.

Then $\text{var}(X_1) = \sum_{k=-1}^{\infty} k^2 \mathbb{P}(X_1 = k)$. We have $\sum_{k=k_0}^{\infty} k^{1-\alpha} = \infty$ for any $k_0 \geq 1$, and $1 - \alpha \in (-1, 0)$, so the variance is infinite. (Note: the asymptotic behaviour of the probability mass function is consistent with the existence of the mean, since $-\alpha \in (-2, -1)$ and so $\sum_{k=k_0}^{\infty} k^{-\alpha} < \infty$.)

What if the variance doesn't exist?

There are several other possible scaling limits for sums of i.i.d. random variables!

For example, suppose that X takes values in $\{-1, 0, 1, 2, \dots\}$ and is such that

- $\mathbb{E}[X_1] = 0$
- $\mathbb{P}(X_1 = k) \sim ck^{-\alpha-1}$ as $k \to \infty$ for $c > 0$ and $\alpha \in (1, 2)$.

Then $\text{var}(X_1) = \sum_{k=-1}^{\infty} k^2 \mathbb{P}(X_1 = k)$. We have $\sum_{k=k_0}^{\infty} k^{1-\alpha} = \infty$ for any $k_0 \geq 1$, and $1 - \alpha \in (-1, 0)$, so the variance is infinite. (Note: the asymptotic behaviour of the probability mass function is consistent with the existence of the mean, since $-\alpha \in (-2, -1)$ and so $\sum_{k=k_0}^{\infty} k^{-\alpha} < \infty$.) In this case, it turns out that

$$\frac{S_n}{n^{1/\alpha}} \xrightarrow{d} S^{(\alpha)}$$

where $S^{(\alpha)}$ has a so-called α-stable distribution. (Note that we're dividing by something much bigger than $\sqrt{n}$!)

Random continuous functions

Suppose we want to make a random continuous function $F : [0, 1] \to \mathbb{R}$ with $F(0) = 0$.

Random continuous functions

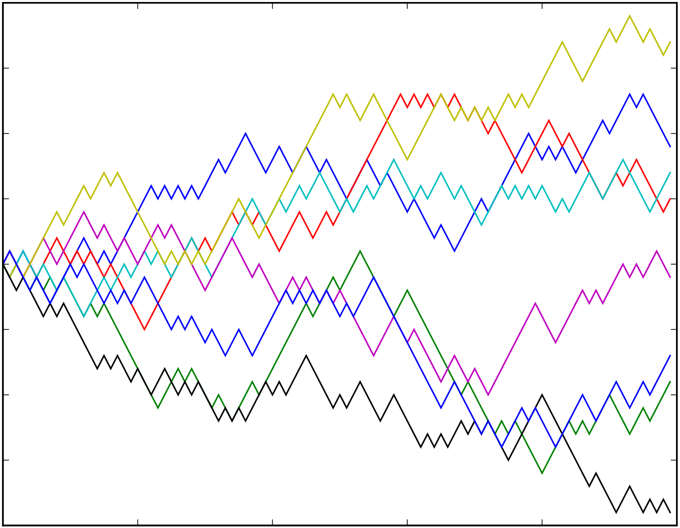
Suppose we want to make a random continuous function $F : [0, 1] \to \mathbb{R}$ with $F(0) = 0$.

One nice way to do this is to use a simple symmetric random walk and interpolate linearly between its steps.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and let $S_k = \sum_{i=1}^{k} X_i$. Then let

$$F_n(t) = S_{\lfloor nt \rfloor} + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}, \text{ for } t \in [0, 1).$$

Random continuous functions



[Picture from Wikipedia, by Morn. Created with Matplotlib.
GFDL, `https://commons.wikimedia.org/w/index.php?curid=9398546`]

Random continuous functions

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and let $S_k = \sum_{i=1}^{k} X_i$. Let

$$F_n(t) = S_{\lfloor nt \rfloor} + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}, \text{ for } t \in [0, 1).$$

Random continuous functions

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and let $S_k = \sum_{i=1}^k X_i$. Let

$$F_n(t) = S_{\lfloor nt \rfloor} + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}, \text{ for } t \in [0, 1).$$

What happens as $n \to \infty$?

Random continuous functions

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and let $S_k = \sum_{i=1}^{k} X_i$. Let

$$F_n(t) = S_{\lfloor nt \rfloor} + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}, \text{ for } t \in [0,1).$$

What happens as $n \to \infty$?

By the CLT, we have $\frac{S_{\lfloor nt \rfloor}}{\sqrt{\lfloor nt \rfloor}} \xrightarrow{d} \mathrm{N}(0,1)$ for each $t \in [0,1]$, so it seems reasonable to rescale F_n by $1/\sqrt{n}$.

Random continuous functions

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and let $S_k = \sum_{i=1}^{k} X_i$. Let

$$F_n(t) = S_{\lfloor nt \rfloor} + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}, \text{ for } t \in [0,1).$$

What happens as $n \to \infty$?

By the CLT, we have $\frac{S_{\lfloor nt \rfloor}}{\sqrt{\lfloor nt \rfloor}} \xrightarrow{d} \mathrm{N}(0,1)$ for each $t \in [0,1]$, so it seems reasonable to rescale F_n by $1/\sqrt{n}$. Then $\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0,t)$.

Random continuous functions

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and let $S_k = \sum_{i=1}^k X_i$. Let

$$F_n(t) = S_{\lfloor nt \rfloor} + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}, \text{ for } t \in [0, 1).$$

What happens as $n \to \infty$?

By the CLT, we have $\frac{S_{\lfloor nt \rfloor}}{\sqrt{\lfloor nt \rfloor}} \xrightarrow{d} \mathrm{N}(0, 1)$ for each $t \in [0, 1]$, so it seems reasonable to rescale F_n by $1/\sqrt{n}$. Then $\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, t)$.

Notice that $\left|\left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}\right| \leq 1$ so if we divide by $1/\sqrt{n}$ this term becomes negligible as $n \to \infty$.

Random continuous functions

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, and let $S_k = \sum_{i=1}^{k} X_i$. Let

$$F_n(t) = S_{\lfloor nt \rfloor} + \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1}, \text{ for } t \in [0, 1).$$

What happens as $n \to \infty$?

By the CLT, we have $\frac{S_{\lfloor nt \rfloor}}{\sqrt{\lfloor nt \rfloor}} \xrightarrow{d} \mathrm{N}(0, 1)$ for each $t \in [0, 1]$, so it seems reasonable to rescale F_n by $1/\sqrt{n}$. Then $\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0, t)$.

Notice that $\left| \left(t - \frac{\lfloor nt \rfloor}{n}\right) X_{\lfloor nt \rfloor + 1} \right| \leq 1$ so if we divide by $1/\sqrt{n}$ this term becomes negligible as $n \to \infty$.

We also have that $S_{\lfloor nt_1 \rfloor}, S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor}, \ldots, S_{\lfloor nt_r \rfloor} - S_{\lfloor nt_{r-1} \rfloor}$ are independent for any $0 \leq t_1 < t_2 < \cdots < t_r \leq 1$ and any $r \geq 2$.

So we're looking for a random continuous function F which is such that

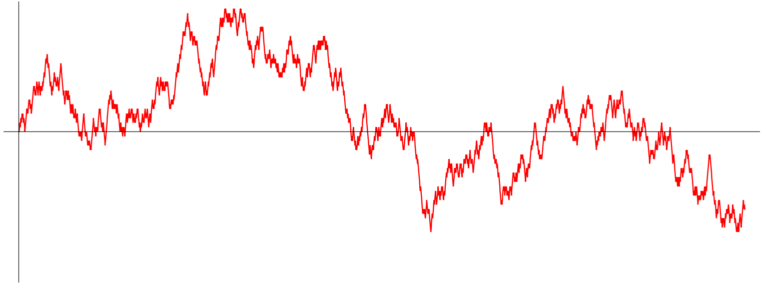
- $F(0) = 0$
- $F(t) - F(s) \sim \mathrm{N}(0, t - s)$ for any $0 \leq s < t \leq 1$
- $F(t_1), F(t_2) - F(t_1), \ldots, F(t_r) - F(t_{r-1})$ are independent for any $0 \leq t_1 < t_2 < \cdots < t_r \leq 1$ and any $r \geq 2$.

So we're looking for a random continuous function F which is such that

- $F(0) = 0$
- $F(t) - F(s) \sim \mathrm{N}(0, t - s)$ for any $0 \leq s < t \leq 1$
- $F(t_1), F(t_2) - F(t_1), \ldots, F(t_r) - F(t_{r-1})$ are independent for any $0 \leq t_1 < t_2 < \cdots < t_r \leq 1$ and any $r \geq 2$.

It turns out that there is a unique random function satisfying these conditions: Brownian motion.

Brownian motion



Scaling limit

Theorem. As $n \to \infty$,

$$\left(\frac{F_n(t)}{\sqrt{n}}, 0 \leq t \leq 1 \right) \xrightarrow{d} (F(t), 0 \leq t \leq 1),$$

where F is a Brownian motion.

Scaling limit

Theorem. As $n \to \infty$,

$$\left(\frac{F_n(t)}{\sqrt{n}}, 0 \leq t \leq 1 \right) \xrightarrow{d} (F(t), 0 \leq t \leq 1),$$

where F is a Brownian motion.

Recall: this means that for any bounded continuous functional $\phi : C([0,1], \mathbb{R}) \to \mathbb{R}$ we have

$$\mathbb{E}\left[\phi(F_n/\sqrt{n})\right] \to \mathbb{E}\left[\phi(F)\right]$$

as $n \to \infty$.

Scaling limit

Theorem. As $n \to \infty$,

$$\left(\frac{F_n(t)}{\sqrt{n}}, 0 \leq t \leq 1 \right) \xrightarrow{d} (F(t), 0 \leq t \leq 1),$$

where F is a Brownian motion.

Recall: this means that for any bounded continuous functional $\phi : C([0,1], \mathbb{R}) \to \mathbb{R}$ we have

$$\mathbb{E}\left[\phi(F_n/\sqrt{n})\right] \to \mathbb{E}\left[\phi(F)\right]$$

as $n \to \infty$.

Bounded continuous functionals capture all sorts of different things. For example, for $f \in C([0,1], \mathbb{R})$, we could take

- $\phi(f) = \exp(-\max_{0 \leq t \leq 1} f(t))$
- $\phi(f) = \sin(f(1/4)f(1/2)f(3/4))$.

Universality

It turns out that this isn't only true for simple random walk. It works also for any random walk with independent identically distributed step-sizes as long as they have mean 0 and variance 1. (And if they have variance σ^2, we just get a constant scaling factor σ.)

Brownian motion is weird and wonderful!

Let B be a Brownian motion. The following statements are true with probability 1:

Brownian motion is weird and wonderful!

Let B be a Brownian motion. The following statements are true with probability 1:

- B is Hölder continuous of exponent α (i.e. $|B(t) - B(s)| \leq C|t - s|^\alpha$ for constants C and α) for every $\alpha < 1/2$ but there is no interval on which it is Hölder continuous of exponent $\alpha \geq 1/2$.

Brownian motion is weird and wonderful!

Let B be a Brownian motion. The following statements are true with probability 1:

- B is Hölder continuous of exponent α (i.e. $|B(t) - B(s)| \leq C|t - s|^{\alpha}$ for constants C and α) for every $\alpha < 1/2$ but there is no interval on which it is Hölder continuous of exponent $\alpha \geq 1/2$.
- In fact, B is nowhere differentiable.

Brownian motion is weird and wonderful!

Let B be a Brownian motion. The following statements are true with probability 1:

- B is Hölder continuous of exponent α (i.e. $|B(t) - B(s)| \leq C|t - s|^\alpha$ for constants C and α) for every $\alpha < 1/2$ but there is no interval on which it is Hölder continuous of exponent $\alpha \geq 1/2$.
- In fact, B is nowhere differentiable.
- The zero set $\mathscr{Z} = \{t : B(t) = 0\}$ has "length" (Lebesgue measure) 0.

Brownian motion is weird and wonderful!

Let B be a Brownian motion. The following statements are true with probability 1:

- B is Hölder continuous of exponent α (i.e. $|B(t) - B(s)| \leq C|t - s|^{\alpha}$ for constants C and α) for every $\alpha < 1/2$ but there is no interval on which it is Hölder continuous of exponent $\alpha \geq 1/2$.
- In fact, B is nowhere differentiable.
- The zero set $\mathcal{Z} = \{t : B(t) = 0\}$ has "length" (Lebesgue measure) 0.
- $\mathcal{Z}$ is a perfect set: it is closed and has no isolated points. ($\mathcal{Z}$ is therefore uncountable.)

Brownian motion is weird and wonderful!

Let B be a Brownian motion. The following statements are true with probability 1:

- B is Hölder continuous of exponent α (i.e. $|B(t) - B(s)| \leq C|t - s|^{\alpha}$ for constants C and α) for every $\alpha < 1/2$ but there is no interval on which it is Hölder continuous of exponent $\alpha \geq 1/2$.
- In fact, B is nowhere differentiable.
- The zero set $\mathcal{Z} = \{t : B(t) = 0\}$ has "length" (Lebesgue measure) 0.
- $\mathcal{Z}$ is a perfect set: it is closed and has no isolated points. ($\mathcal{Z}$ is therefore uncountable.)
- $\mathcal{Z}$ is a random fractal set, with fractal dimension equal to $1/2$.

Brownian motion is easy to calculate with

Brownian motion sits at the intersection of many different classes of stochastic processes: it's a Markov process, it's a Gaussian process and it's a martingale. So there are many different tools and techniques available for its study.

Brownian motion is easy to calculate with

Brownian motion sits at the intersection of many different classes of stochastic processes: it's a Markov process, it's a Gaussian process and it's a martingale. So there are many different tools and techniques available for its study.

This means that even if what we're interested in is actually a fact about random walks (for example, how much time the random walk spends above the x-axis), it's often much easier to do those calculations in the continuum and use the answer as an approximation.

Brownian motion is easy to calculate with

Brownian motion sits at the intersection of many different classes of stochastic processes: it's a Markov process, it's a Gaussian process and it's a martingale. So there are many different tools and techniques available for its study.

This means that even if what we're interested in is actually a fact about random walks (for example, how much time the random walk spends above the x-axis), it's often much easier to do those calculations in the continuum and use the answer as an approximation. To continue the example, the amount of time in $[0,1]$ that a Brownian motion spends above the x-axis has density

$$\frac{1}{\pi\sqrt{x(1-x)}}, \quad 0 < x < 1.$$

Brownian motion is useful!

There are many real-world applications in which random walks or Brownian motion are used as a model. For example,

- stock prices
- animal movements
- genetic evolution in a population
- particle motion in physics,

and as a component part of many many more!

Brownian motion is useful!

There are many real-world applications in which random walks or Brownian motion are used as a model. For example,

- stock prices
- animal movements
- genetic evolution in a population
- particle motion in physics,

and as a component part of many many more!

There are also deep links to the theory of PDE's.

Trees

I now want to move on to talk about something which looks completely unrelated: trees.

Trees

I now want to move on to talk about something which looks completely unrelated: trees.

Mathematically, a tree is a connected acyclic graph.

Trees

I now want to move on to talk about something which looks completely unrelated: trees.

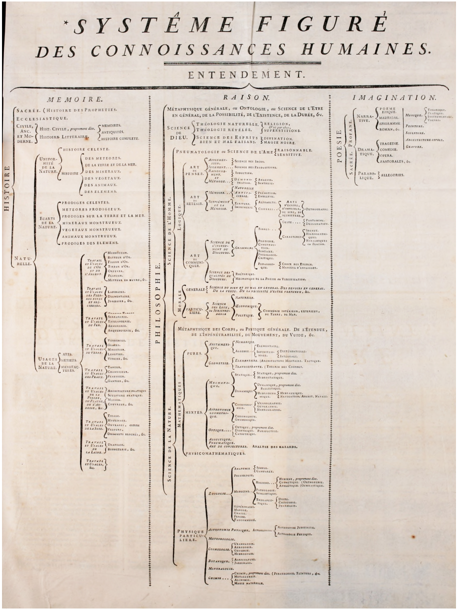
Mathematically, a tree is a connected acyclic graph.

But tree-structures are also ubiquitous in nature.

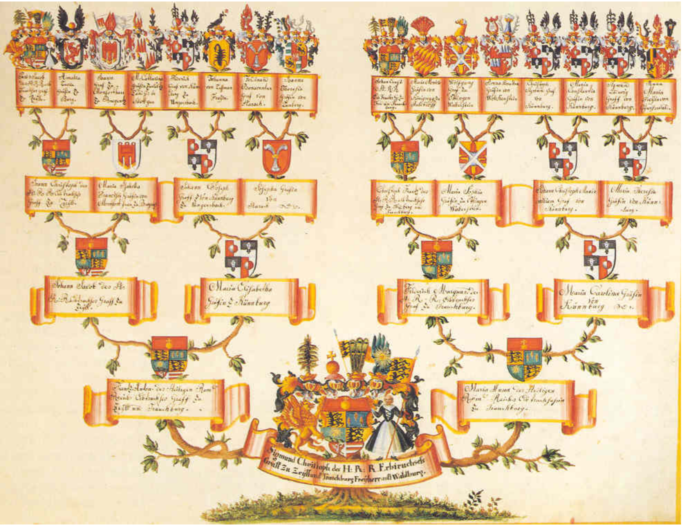
[“Lichtenberg figure in block of plexiglas” by Bert Hickman.
(Licensed under Attribution via Wikimedia Commons.)]

["Unique snow flake" by Pen Waggener - Flickr: Unique.
(Licensed under CC BY 2.0 via Wikimedia Commons.)]

[“Yarlung Tsangpo river, Tibet” by NASA, `http://photojournal.jpl.nasa.gov/catalog/PIA03708`. (Licensed under Public Domain via Wikimedia Commons.)]



[http://ets.lib.uchicago.edu/ARTFL/OLDENCYC/images.
(Licensed under Public Domain via Wikimedia Commons.)]

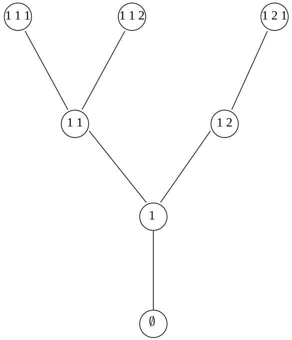


["Waldburg Ahnentafel", http://www.ahneninfo.com/de/ahnentafel.htm. (Licensed under Public Domain via Wikimedia Commons.)]

A mathematical abstraction: ordered trees

Consider a rooted ordered tree on n vertices ("ordered" means that the left-to-right ordering matters).

Example: $n = 7$



Random trees

The set $\mathcal{T}_n$ of ordered trees on n vertices is one of the (many!) combinatorial families enumerated by the Catalan numbers:

$$|\mathcal{T}_n| = \frac{1}{n+1}\binom{2n}{n}.$$

Random trees

The set $\mathcal{T}_n$ of ordered trees on n vertices is one of the (many!) combinatorial families enumerated by the Catalan numbers:

$$|\mathcal{T}_n| = \frac{1}{n+1}\binom{2n}{n}.$$

Using Stirling's approximation, we get that

$$|\mathcal{T}_n| \sim \frac{4^n}{n^{3/2}\sqrt{\pi}} \quad \text{as } n \to \infty.$$

Random trees

The set $\mathcal{T}_n$ of ordered trees on n vertices is one of the (many!) combinatorial families enumerated by the Catalan numbers:

$$|\mathcal{T}_n| = \frac{1}{n+1}\binom{2n}{n}.$$

Using Stirling's approximation, we get that

$$|\mathcal{T}_n| \sim \frac{4^n}{n^{3/2}\sqrt{\pi}} \quad \text{as } n \to \infty.$$

Let T_n be a tree picked uniformly at random from $\mathcal{T}_n$.

Random trees

The set $\mathcal{T}_n$ of ordered trees on n vertices is one of the (many!) combinatorial families enumerated by the Catalan numbers:

$$|\mathcal{T}_n| = \frac{1}{n+1}\binom{2n}{n}.$$

Using Stirling's approximation, we get that

$$|\mathcal{T}_n| \sim \frac{4^n}{n^{3/2}\sqrt{\pi}} \quad \text{as } n \to \infty.$$

Let T_n be a tree picked uniformly at random from $\mathcal{T}_n$.

Question: What can we say about the properties of T_n as n gets large?

Things we might want to know

- What is the largest distance between the root and another vertex?
- What is the diameter of the tree? (i.e. what is the length of the longest path between two points in the tree?)
- How many vertices are there at distance d from the root?
- How many leaves (i.e. vertices with only one neighbour) are there?
- How many vertices have two neighbours? Or three? Or more?

Things we might want to know

- What is the largest distance between the root and another vertex?
- What is the diameter of the tree? (i.e. what is the length of the longest path between two points in the tree?)
- How many vertices are there at distance d from the root?
- How many leaves (i.e. vertices with only one neighbour) are there?
- How many vertices have two neighbours? Or three? Or more?

Because the tree is random, all of these quantities are random variables with (complicated) distributions which depend on n.

Things we might want to know

- What is the largest distance between the root and another vertex?
- What is the diameter of the tree? (i.e. what is the length of the longest path between two points in the tree?)
- How many vertices are there at distance d from the root?
- How many leaves (i.e. vertices with only one neighbour) are there?
- How many vertices have two neighbours? Or three? Or more?

Because the tree is random, all of these quantities are random variables with (complicated) distributions which depend on n.

Can we take a limit as $n \to \infty$, in a sensible way?

Things we might want to know

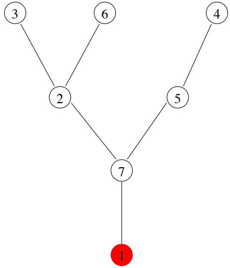
- What is the largest distance between the root and another vertex?
- What is the diameter of the tree? (i.e. what is the length of the longest path between two points in the tree?)
- How many vertices are there at distance d from the root?
- How many leaves (i.e. vertices with only one neighbour) are there?
- How many vertices have two neighbours? Or three? Or more?

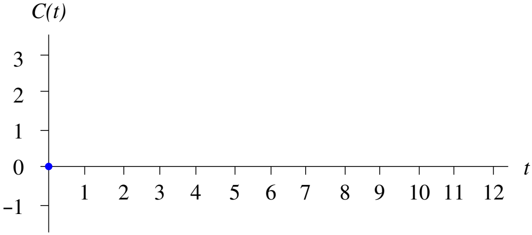
Because the tree is random, all of these quantities are random variables with (complicated) distributions which depend on n.

Can we take a limit as $n \to \infty$, in a sensible way?

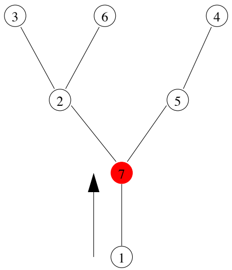
It's useful to have a way of "getting our hands" on T_n. We do this via a functional encoding.

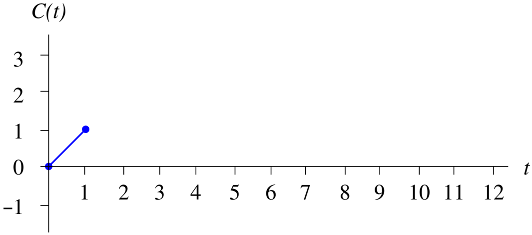
Contour function



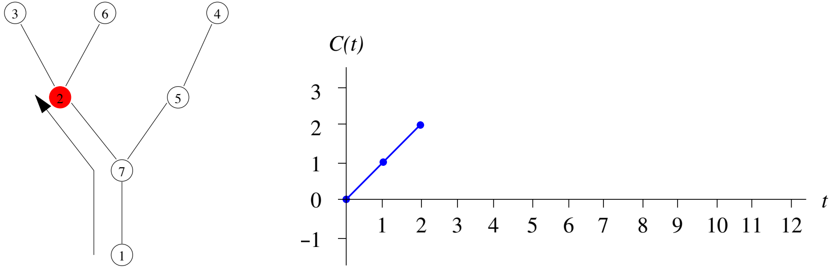


Contour function

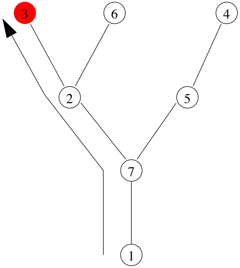


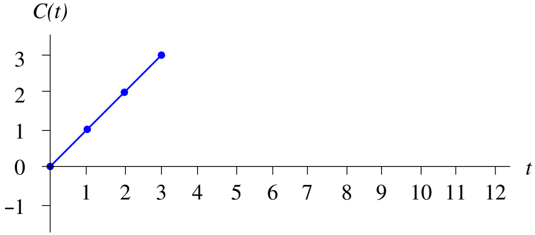


Contour function

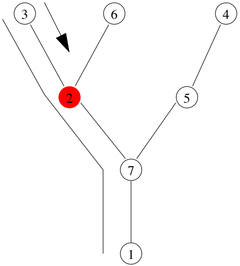


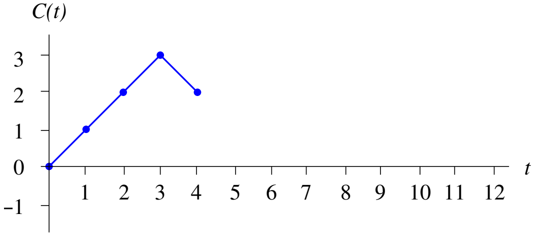
Contour function



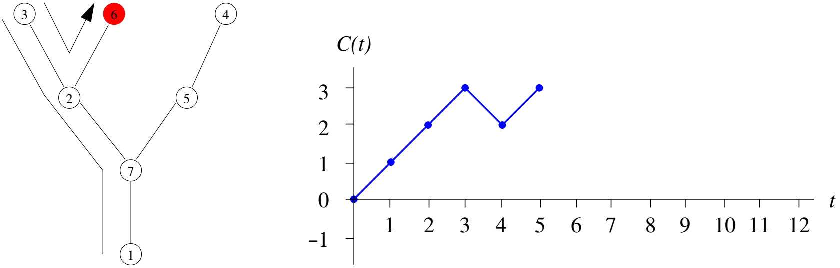


Contour function

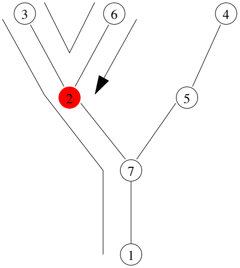


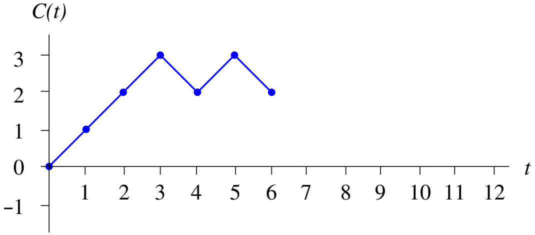


Contour function

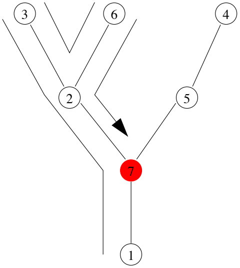


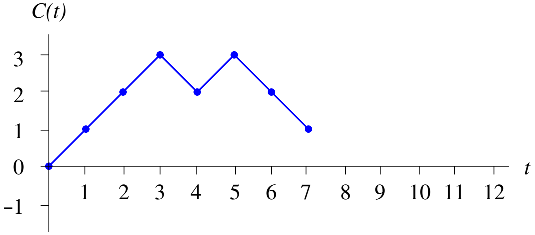
Contour function



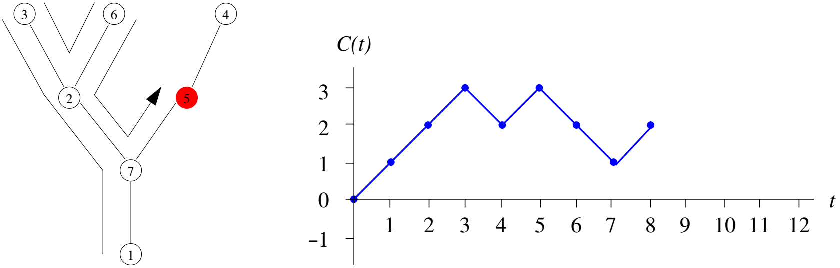


Contour function

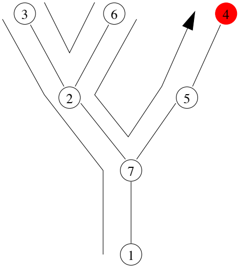


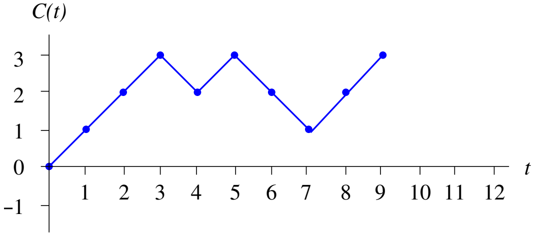


Contour function

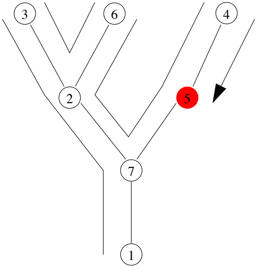


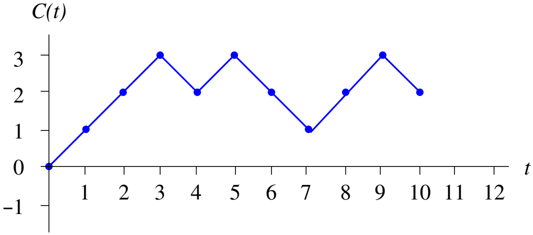
Contour function



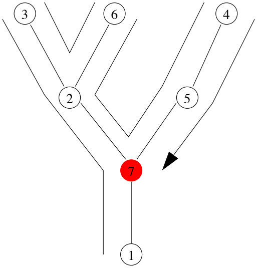


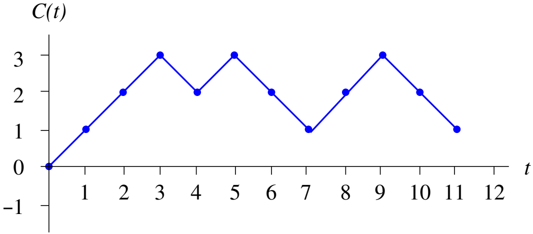
Contour function



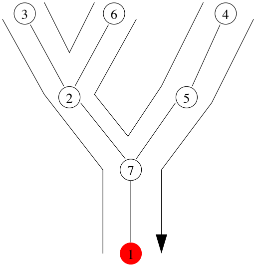


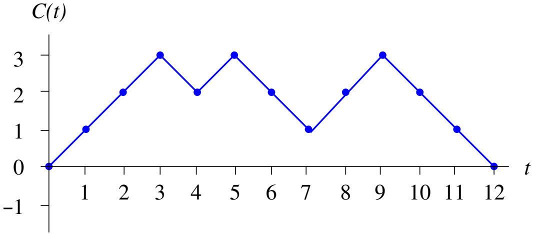
Contour function



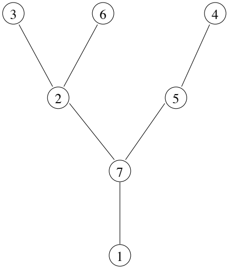


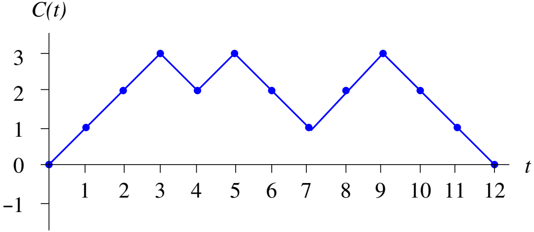
Contour function





Contour function





The contour function is a sort of “expanded” version of the tree.

A bijection

Indeed, there is a bijection between the set $\mathcal{T}_n$ of ordered trees with n vertices and the set $\mathcal{W}_n$ of discrete walks with $2(n-1)$ steps in $\{-1, +1\}$ which start and end at 0 and remain non-negative in between.

A bijection

Indeed, there is a bijection between the set $\mathcal{T}_n$ of ordered trees with n vertices and the set $\mathcal{W}_n$ of discrete walks with $2(n-1)$ steps in $\{-1, +1\}$ which start and end at 0 and remain non-negative in between. We'll call these excursions.

A bijection

Indeed, there is a bijection between the set $\mathcal{T}_n$ of ordered trees with n vertices and the set $\mathcal{W}_n$ of discrete walks with $2(n-1)$ steps in $\{-1, +1\}$ which start and end at 0 and remain non-negative in between. We'll call these excursions.

In fact, a uniformly random element of $\mathcal{W}_n$ is the same as a simple random walk conditioned to start and end at 0 and remain non-negative in between. To see this, we just need to observe that every particular sequence of $+1$ and -1 steps which satisfies the conditions has the same probability for the random walk.

A bijection

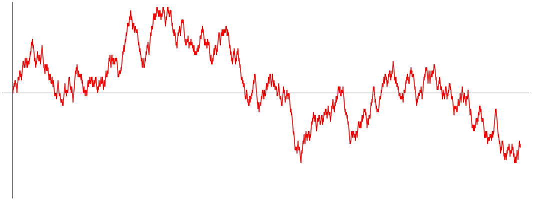
Indeed, there is a bijection between the set $\mathcal{T}_n$ of ordered trees with n vertices and the set $\mathcal{W}_n$ of discrete walks with $2(n-1)$ steps in $\{-1, +1\}$ which start and end at 0 and remain non-negative in between. We'll call these excursions.

In fact, a uniformly random element of $\mathcal{W}_n$ is the same as a simple random walk conditioned to start and end at 0 and remain non-negative in between. To see this, we just need to observe that every particular sequence of $+1$ and -1 steps which satisfies the conditions has the same probability for the random walk.

If we ignore the conditioning, then we know that we get Brownian motion as the scaling limit of this path.

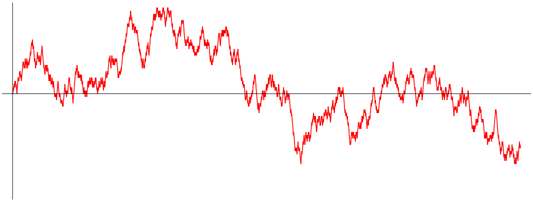
Conditioning Brownian motion

The Brownian motion path is made up of excursions away from 0:

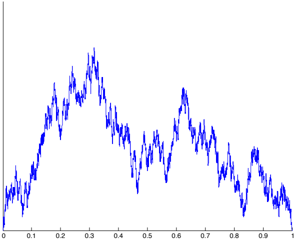


Conditioning Brownian motion

The Brownian motion path is made up of excursions away from 0:



If we take one of these excursions conditioned to have length 1, we get a standard Brownian excursion, $(\mathbb{e}(t), 0 \leq t \leq 1)$.

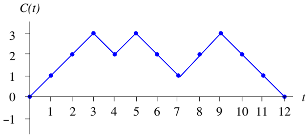


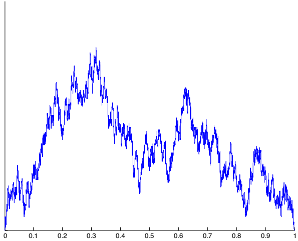
Scaling limit of a random walk excursion

Let $(C(t), 0 \leq t \leq 2(n-1))$ be a simple random walk excursion of $2(n-1)$ steps, linearly interpolated.

Theorem.
As $n \to \infty$,

$$\frac{1}{\sqrt{2n}}\Big(C(2(n-1)s), 0 \leq s \leq 1\Big) \xrightarrow{d} (\mathbb{e}(s), 0 \leq s \leq 1).$$





Scaling limit for the random tree

This suggests that there should be some sort of scaling limit for the tree, which is somehow encoded by the Brownian excursion.

Scaling limit for the random tree

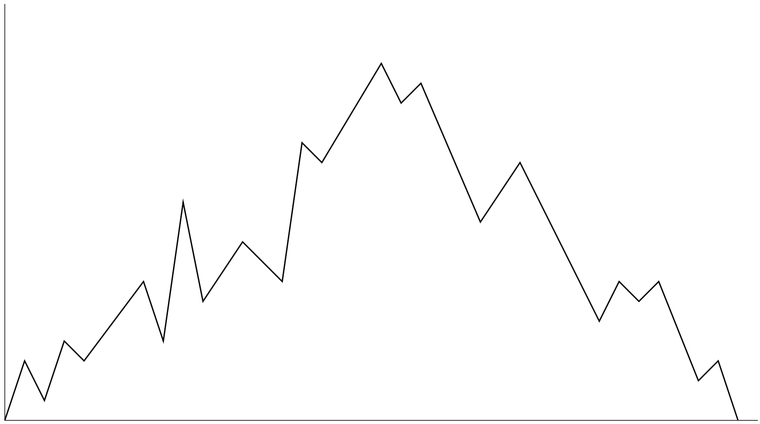
This suggests that there should be some sort of scaling limit for the tree, which is somehow encoded by the Brownian excursion.

This is indeed true, and it's called the Brownian continuum random tree. It was discovered by David Aldous, who was a member of faculty at Berkeley from 1979 until his retirement in 2018.

Scaling limit for the random tree

In order to give a (slightly informal) definition of the Brownian continuum random tree, we need to think about how to get back from excursions to trees.

From excursions back to trees



From excursions back to trees

Now put glue on the underside of the excursion and push the two sides together...

From excursions back to trees

Now put glue on the underside of the excursion and push the two sides together...

From excursions back to trees

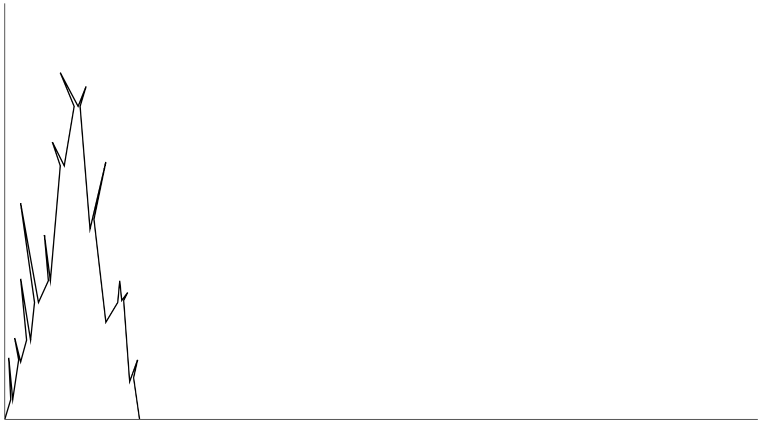
Now put glue on the underside of the excursion and push the two sides together...

From excursions back to trees

Now put glue on the underside of the excursion and push the two sides together...

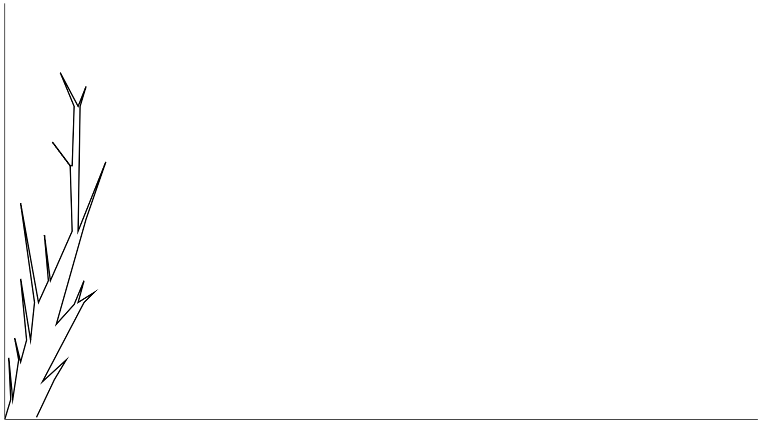
From excursions back to trees

Now put glue on the underside of the excursion and push the two sides together...



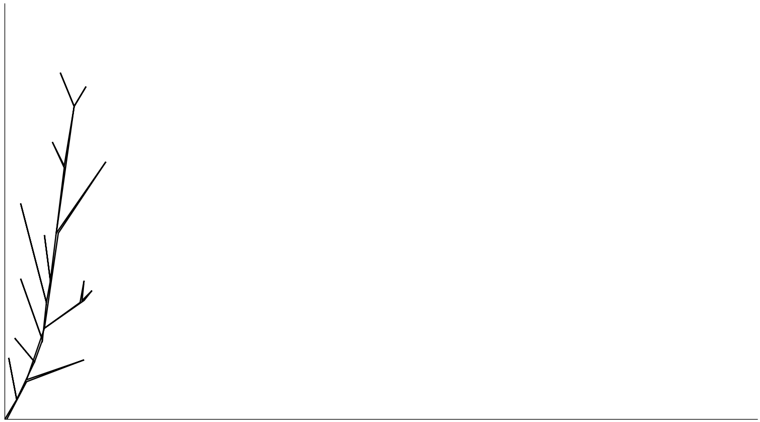
From excursions back to trees

Now put glue on the underside of the excursion and push the two sides together...



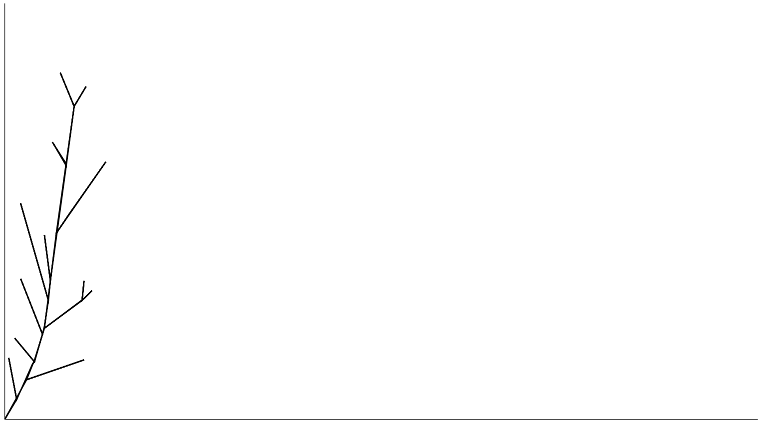
From excursions back to trees

Now put glue on the underside of the excursion and push the two sides together...



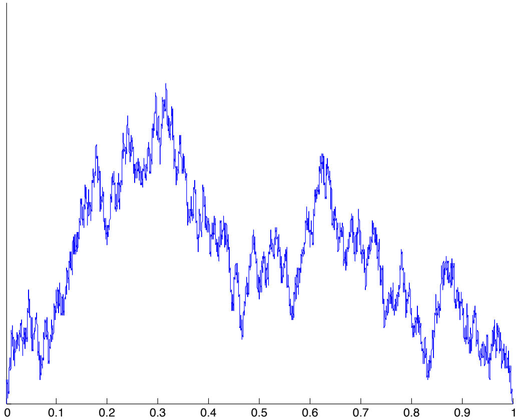
From excursions back to trees

Now put glue on the underside of the excursion and push the two sides together to get a tree.



Brownian continuum random tree

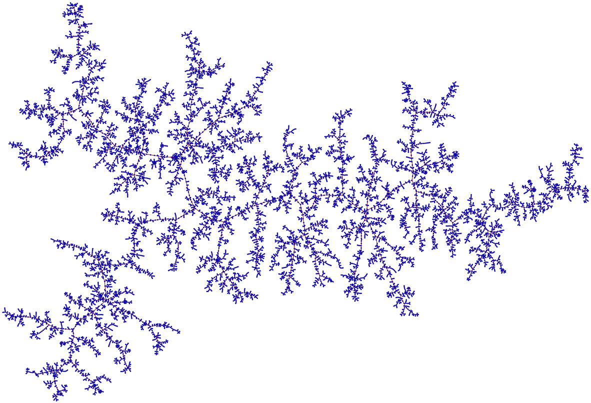
If you do this gluing operation to a Brownian excursion, you get the Brownian continuum random tree.



[Picture by Igor Korchemski]

Brownian continuum random tree

If you do this gluing operation to a Brownian excursion, you get the Brownian continuum random tree.



[Picture by Igor Korchemski]

Scaling limit theorem

Theorem (Aldous).

As $n \to \infty$,

$$\frac{1}{\sqrt{2n}} T_n \xrightarrow{d} \mathscr{T},$$

where $\mathscr{T}$ is the Brownian continuum random tree.

Scaling limit theorem

Theorem (Aldous).

As $n \to \infty$,

$$\frac{1}{\sqrt{2n}} T_n \xrightarrow{d} \mathscr{T},$$

where $\mathscr{T}$ is the Brownian continuum random tree.

What does this convergence capture? It's particularly good for thinking about distances in the tree.

Scaling limit theorem

Theorem (Aldous).

As $n \to \infty$,

$$\frac{1}{\sqrt{2n}} T_n \xrightarrow{d} \mathscr{T},$$

where $\mathscr{T}$ is the Brownian continuum random tree.

What does this convergence capture? It's particularly good for thinking about distances in the tree.

For example, the largest distance from the root to another vertex in T_n is given by the maximum of the corresponding random walk excursion.

Scaling limit theorem

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

$$\frac{1}{\sqrt{2n}} T_n \xrightarrow{d} \mathscr{T},$$

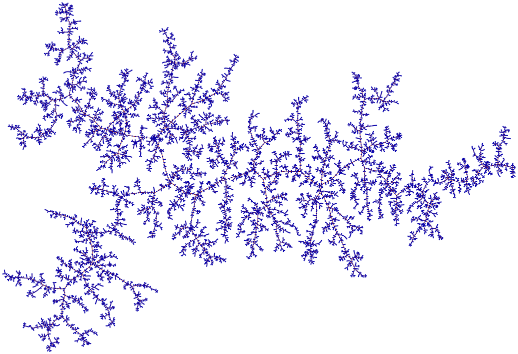
where $\mathscr{T}$ is the Brownian continuum random tree.

What does this convergence capture? It's particularly good for thinking about distances in the tree.

For example, the largest distance from the root to another vertex in T_n is given by the maximum of the corresponding random walk excursion. This converges in distribution, on rescaling, to the equivalent quantity for $\mathscr{T}$, which is the maximum of the Brownian excursion, and has cumulative distribution function

$$\mathbb{P}\left(\max_{0 \le t \le 1} \mathbb{e}(t) \le x\right) = 1 - 2\sum_{k=1}^{\infty}(4x^2k^2 - 1)\exp(-2x^2k^2), \quad x \ge 0.$$

Brownian continuum random tree



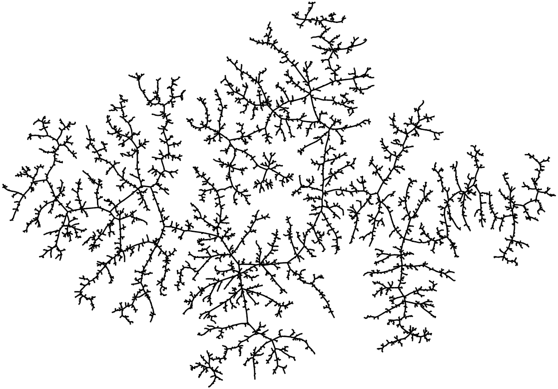
Like Brownian motion, the Brownian continuum random tree is a fascinating mathematical object! In particular, it is a random fractal with fractal dimension 2, and has lots of nice distributional properties.

Universality

It turns out that many different families of "uniform-like" trees have the Brownian continuum random tree as their scaling limit.

Universality

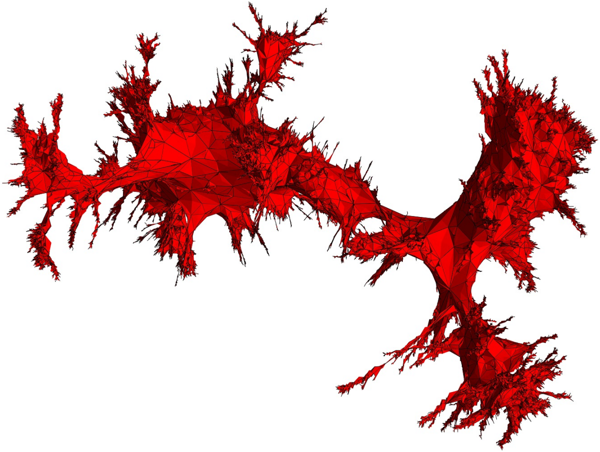
It turns out that many different families of "uniform-like" trees have the Brownian continuum random tree as their scaling limit. It also shows up as a building block in the scaling limits of other more complicated random discrete structures, for example random graphs.



[Picture by Nicolas Broutin]

Universality

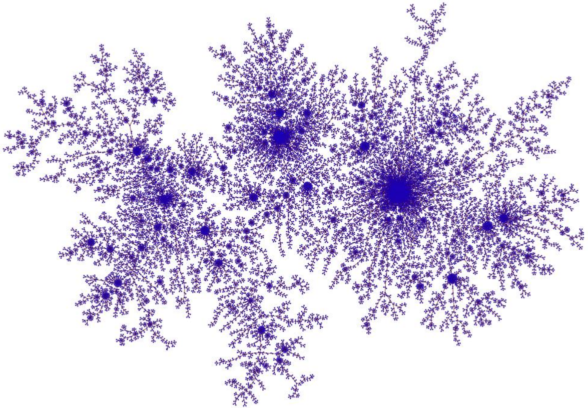
It turns out that many different families of "uniform-like" trees have the Brownian continuum random tree as their scaling limit. It also shows up as a building block in the scaling limits of other more complicated random discrete structures, for example random graphs or random planar maps.



[Picture by Jérémie Bettinelli]

Random trees and random graphs

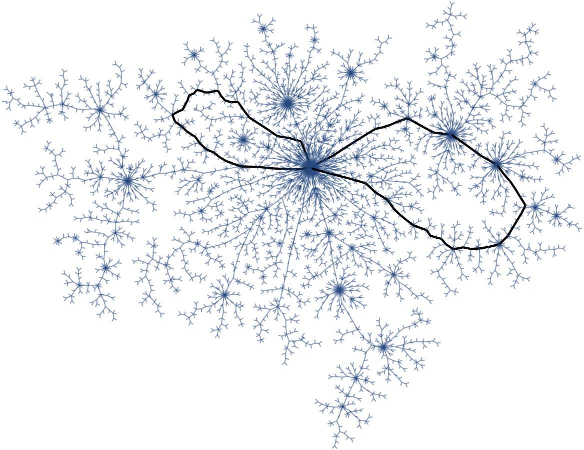
Lots of my recent work has focussed on the scaling limits of various models of random trees and random graphs, particularly those with hubs (which are related to the stable distributions I mentioned earlier).



[Picture by Igor Kortchemski]

Random trees and random graphs

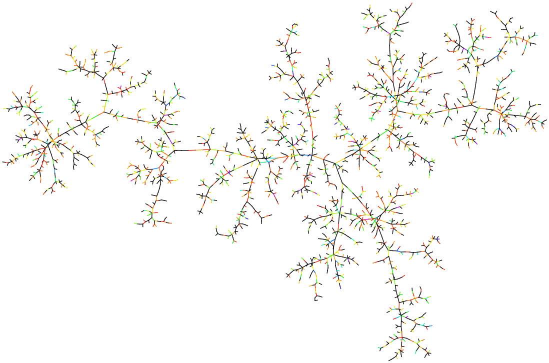
Lots of my recent work has focussed on the scaling limits of various models of random trees and random graphs, particularly those with hubs (which are related to the stable distributions I mentioned earlier).



[Picture by Delphin Sénizergues]

Random trees and random graphs

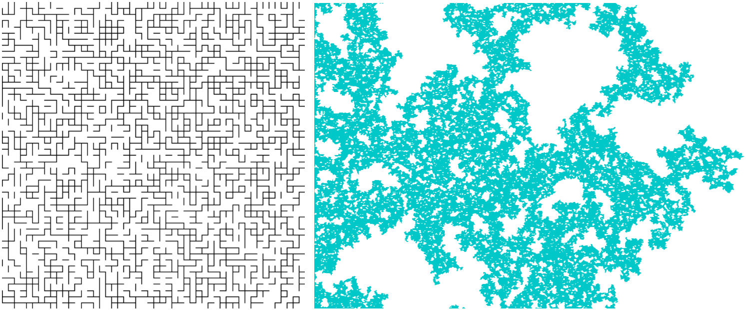
Another nice example is the scaling limit of the minimum spanning tree of the complete graph. Here, the scaling turns out to be $n^{1/3}$ rather than $\sqrt{n}$, and the fractal dimension is 3 almost surely.



[Picture by Louigi Addario-Berry]

Scaling limits are everywhere!

Scaling limits turn up all over probability theory. There are some particularly famous examples in the context of percolation and other models coming from statistical mechanics.



[L: Picture by James Martin; R: "The fractal dimension of the percolation by invasion cluster at the percolation threshold is 91/48=1.89" by Alexis Monnerot-Dumaine. (Licensed under Attribution via Creative Commons.)]

Thank you for your attention!