

# RANDOM MOTION AND RANDOM TREES

JAY SAGROLIKAR

ABSTRACT. We provide an introduction to probability theory through the lens of Brownian Motion. After constructing Brownian Motion, we outline background behind branching processes and random trees to build a connection between trees and random walks. Through a bijection between certain types of linearly interpolated random walks and uniform plane trees, we show that under certain conditions, the walks converge to a special type of Brownian Motion. This gives us the machinery to state the existence of Aldous' continuum random tree (CRT), which describes the uniform plane tree when the number of vertices grows large.

## CONTENTS

1. Probability Theory	1
2. Brownian Motion	7
3. Graphs and Random Trees	11
4. (Limits of) Uniform Plane Trees	15
Acknowledgments	18
References	18

## 1. PROBABILITY THEORY

In order to introduce some more advanced topics and applications in probability, it will first be useful to rigorously dive into essential definitions in theoretical probability. The definitions that follow all involve outcomes that are not the result of deterministic functions but those that have been chosen at random. We will soon define how these outcomes are chosen.

In some cases (for example, choosing a random number from an interval of the real line), the probability of any individual outcome is zero, yet the sum of the probabilities of all outcomes is 1. To remedy this seeming issue, it is easier to consider events, or sets of outcomes. We use the idea of  $\sigma$ -algebras (sigma algebras) to encode this idea. Informally, a  $\sigma$ -algebra represents a set of events that can be given a probability.

**Definition 1.1** (Sigma Algebra). A  $\sigma$ -algebra  $\mathcal{F}$  is a subset of the sample space  $\Omega$  satisfying

- (1)  $\emptyset \in \mathcal{F}$ .
- (2) If  $A \in \mathcal{F}$ , then  $\Omega \setminus A \in \mathcal{F}$ .
- (3) If  $A_1, A_2, \dots \in \mathcal{F}$ , then  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ .

Elements of  $\sigma$ -algebras are called events, whereas elements of the sample space  $\Omega$  are called outcomes. A crucial example of a  $\sigma$ -algebra is the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$ , which is the  $\sigma$ -algebra generated by all open sets of  $\mathbb{R}$ . A Borel set is an element of the Borel  $\sigma$ -algebra. A  $\sigma$ -algebra can also represent the information available to us after the realization of an outcome  $\omega$  in the sample space.

**Definition 1.2** (Probability Measure). A probability measure  $P$  is a function from  $\mathcal{F} \rightarrow [0, 1]$  such that

- (1)  $P(\emptyset) = 0$  and  $P(\Omega) = 1$ .
- (2) If  $A_1, A_2, \dots$  are pairwise disjoint, then  $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ .
- (3) If  $A, B \in \mathcal{F}$  and  $A \subseteq B$ , then  $P(A) \leq P(B)$ .
- (4) If  $A, B \in \mathcal{F}$ , then  $P(A \cup B) \leq P(A) + P(B)$ .

A probability measure helps us assign probabilities to events in a  $\sigma$ -algebra. Monotonicity and subadditivity of the probability measure will prove to be extremely useful later on. The sample space and the objects defined above form a probability space, which gives us the information necessary to model a random procedure.

**Definition 1.3** (Probability Space). A probability space is a triple  $(\Omega, \mathcal{F}, P)$  consisting of a sample space, a  $\sigma$ -algebra on  $\Omega$ , and a probability measure on  $\mathcal{F}$ .

Events are called  $\mathcal{F}$ -measurable if they are in the  $\sigma$ -algebra  $\mathcal{F}$ . If an event is  $\mathcal{F}$ -measurable, we can measure the probability of the event occurring.

**Definition 1.4** ( $\mathcal{F}$ -Measurable Function). On  $(\Omega, \mathcal{F}, P)$ , a function  $f : \Omega \rightarrow \mathbb{R}$  is called  $\mathcal{F}$ -measurable if for all Borel sets  $B \in \mathcal{B}(\mathbb{R})$ , the preimage  $X^{-1}(B) = \{\omega : f(\omega) \in B\} \in \mathcal{F}$ .

In other words, if the preimage of all Borel sets is  $\mathcal{F}$ -measurable, the function is  $\mathcal{F}$ -measurable.

**Definition 1.5** (Random Variable). A random variable  $X$  is an  $\mathcal{F}$ -measurable function from  $(\Omega, \mathcal{F}, P)$  to the real numbers.

For example, consider a fair coin toss where we gain a dollar if it lands on heads and lose a dollar if it lands on tails. The earning from a single toss of the coin is a random variable. We define it as

$$X = \begin{cases} 1, & \text{if heads} \\ -1, & \text{if tails.} \end{cases}$$

Random variables can take on countable values (as in the example above) or uncountably many (such as a random real number generated from the interval  $[0, 1]$ ). We will define the difference using the following function.

**Definition 1.6** (Cumulative Distribution Function). The cumulative distribution function (CDF or distribution) of a random variable  $X$  on  $(\Omega, \mathcal{F}, P)$  is given by  $F_X : \mathbb{R} \rightarrow [0, 1]$ , where

$$F_X(x) = P(X \leq x).$$

A random variable  $X$  is called continuous if its CDF is continuous and discrete otherwise. Sometimes it is necessary to evaluate the probability of a random variable taking on any value in specific ranges. To do so, we can define a new function: probability density.

**Definition 1.7** (Probability Density Function). Suppose  $X$  is a continuous random variable on  $(\Omega, \mathcal{F}, P)$ . If there exists a function  $f$  such that

$$F(x) = \int_{-\infty}^x f(u) du,$$

$f$  is called the probability density function (PDF or density) of  $X$ . For a discrete random variable  $X$  taking on values  $x_1, \dots, x_n$ , we call the function describing the probability that  $X$  equals any particular  $x_i, i \in \{1, \dots, n\}$  the probability mass function.

Random variables or pairs of random variables have some properties. The first we look at is fairly intuitive to understand – independence, which denotes (informally) whether one random variable affects the outcome of another.

**Definition 1.8** (Independence). Two random variables  $X, Y$  on  $(\Omega, \mathcal{F}, P)$  are independent if

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b).$$

In terms of sets,  $X$  and  $Y$  are independent if for all Borel sets  $A, B \in \mathcal{B}(\mathbb{R})$ ,

$$\begin{aligned} & P\left(\{\omega \in \Omega : X(\omega) \in A\} \cap \{\omega \in \Omega : Y(\omega) \in B\}\right) \\ &= P\left(\{\omega \in \Omega : X(\omega) \in A\}\right)P\left(\{\omega \in \Omega : Y(\omega) \in B\}\right). \end{aligned}$$

Another useful property is the “average” value of a random variable, which is quantified by the expectation, or expected value.

**Definition 1.9** (Expected Value). The expected value of a continuous random variable  $X$  on  $(\Omega, \mathcal{F}, P)$  is defined as

$$\mathbb{E}[X] = \int_{\Omega} X dP.$$

Suppose  $X$  has density  $f$ . Then

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

For a discrete random variable  $X$  taking on values  $x_1, \dots, x_n$ , the expected value of  $X$  is

$$\mathbb{E}[X] = \sum_{i=1}^n x_i P(X = x_i).$$

It is also important to note that the expected value of a random variable is a linear operator on the probability space where it is defined. The next two properties – covariance and variance – give us a metric for how random variables are spread (either with one another or apart from their own expected value).

**Definition 1.10** (Covariance). The covariance of two random variables  $X, Y$  on  $(\Omega, \mathcal{F}, P)$  is

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

**Definition 1.11** (Variance). The variance of a random variable  $X$  on  $(\Omega, \mathcal{F}, P)$  is given by

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The variance of  $X$  can also be expressed as  $Cov(X, X)$ .

We see some distributions more often than others, so it is also useful to name them.

**Definition 1.12** (Normal Distribution). A random variable  $X$  on  $(\Omega, \mathcal{F}, P)$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$  ( $X \sim N(\mu, \sigma^2)$ ) if it has density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

We call the standard normal distribution a normal distribution with  $\mu = 0$  and  $\sigma = 1$ . From here onwards, we define random variables assuming they are well-defined on a probability space  $(\Omega, \mathcal{F}, P)$ .

**Definition 1.13** (Moment Generating Function). The moment generating function of a random variable  $X$  is given by

$$M_X(t) = \mathbb{E}[e^{tX}].$$

Two random variables are identically distributed if their moment generating functions are the same.

Using the definitions above, we can now state some results about the normal distribution that will be central to the construction of Brownian motion in the next section.

**Proposition 1.14.** *The moment generating function of a normally distributed random variable with mean  $\mu$  and variance  $\sigma^2$  is*

$$M_{N(\mu, \sigma^2)}(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2).$$

*Proof.* The proof follows from the definition of expected value and the density of the normal distribution. We can check that if  $X \sim N(\mu, \sigma^2)$ ,

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{xt} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \exp(\mu t + \frac{1}{2}\sigma^2 t^2).$$

□

**Lemma 1.15.** *Suppose  $X \sim N(\mu, \sigma^2)$  and let  $a, b \in \mathbb{R}$ . Then  $Y = aX + b$  is normally distributed with mean  $a\mu + b$  and variance  $a^2\sigma^2$ .*

*Proof.* Consider the moment generating function of  $Y$ . We have

$$\begin{aligned} M_Y(t) &= M_{aX+b}(t) = \mathbb{E}[e^{t(aX+b)}] = e^{tb} M_X(at) \\ &= e^{tb} \exp(a\mu t + \frac{1}{2}\sigma^2 (at)^2) = \exp\left((a\mu + b)t + \frac{1}{2}(a\sigma)^2 t^2\right), \end{aligned}$$

which is the moment generating function of a normally distributed random variable with mean  $a\mu + b$  and variance  $a^2\sigma^2$ . □

The fact that the sum of two normally distributed random variables is also normally distributed is extremely useful. The next few results showcase how the relationship between normally distributed random variables can be utilized.

**Definition 1.16** (Joint Normal Distribution). A finite sequence of random variables  $\{X_1, \dots, X_n\}$  has a joint normal distribution if there exists a finite set of independent random variables  $\{Z_1, \dots, Z_m\}$ , each with a standard normal distribution, such that for any  $i \in \{1, \dots, n\}$ , there exist constants  $a_{i_0}, a_{i_1}, \dots, a_{i_m}$  with

$$X_i = a_{i_0} + a_{i_1}Z_1 + \dots + a_{i_m}Z_m.$$

If the expectation of each random variable in the sequence  $\{X_1, \dots, X_n\}$  is zero (i.e.,  $a_{i_0} = 0$  for all  $i \in \{1, \dots, n\}$ ), then the sequence is said to have a *centered* (or mean-zero) joint normal distribution.

**Lemma 1.17.** *If  $X_1$  and  $X_2$  are random variables with a centered joint normal distribution and  $\mathbb{E}[X_1 X_2] = 0$ , then  $X_1$  and  $X_2$  are independent random variables.*

*Proof.* For the proof, see section 2.2 of [10].  $\square$

**Lemma 1.18.** *Let  $X, Y$  be independent  $N(0, 1)$  random variables, and for some  $\sigma \in \mathbb{R}$ , let*

$$Z = \frac{X + Y}{\sigma}, \quad W = \frac{X - Y}{\sigma}.$$

*Then  $Z, W$  are independent  $N(0, \frac{2}{\sigma^2})$  random variables.*

*Proof.*  $Z, W$  clearly have a joint normal distribution. Since  $X, Y \sim N(0, 1)$ , we know  $\mathbb{E}[X] = 0 = \mathbb{E}[Y]$ , implying  $\mathbb{E}[X^2] = 1 = \mathbb{E}[Y^2]$ . We also have  $\mathbb{E}[Z] = 0 = \mathbb{E}[W]$ , meaning that

$$\text{Var}(Z) = \mathbb{E}[Z^2] = \mathbb{E}\left[\frac{(X + Y)^2}{\sigma^2}\right] = \frac{2}{\sigma^2}.$$

Similarly,  $\text{Var}(W) = \frac{2}{\sigma^2}$ . To show that  $Z, W$  are normally distributed, consider their moment generating functions. We have

$$\begin{aligned} M_Z(t) &= M_{\frac{1}{\sigma}(X+Y)}(t) = M_{\frac{1}{\sigma}X}(t)M_{\frac{1}{\sigma}Y}(t) = \exp\left(-\frac{1}{2\sigma^2}t^2\right)\exp\left(-\frac{1}{2\sigma^2}t^2\right) \\ &= \exp\left(-\frac{1}{\sigma^2}t^2\right), \end{aligned}$$

exactly the moment generating function of a random variable with mean zero and variance  $\frac{2}{\sigma^2}$ . The same result can be derived in a similar manner for  $W$ . It remains to show  $Z$  and  $W$  are independent, which immediately follows from Lemma 1.17 after checking that

$$\mathbb{E}[ZW] = \mathbb{E}\left[\frac{X^2 - Y^2}{\sigma^2}\right] = 0.$$

Hence,  $Z, W \sim N(0, \frac{2}{\sigma^2})$ .  $\square$

We now move from random variables into some results about probabilities. The first gives us a way to bound the occurrence of at least one of a countable set of events.

**Lemma 1.19** (Boole's Inequality). *Suppose  $(\Omega, \mathcal{F}, P)$  is a probability space, and let  $A_1, A_2, \dots \in \mathcal{F}$  be events. Then*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

*Proof.* See 15.1 of [12] for the full proof.  $\square$

**Definition 1.20.** Let  $\{A_n : n \in \mathbb{N}\}$  be a sequence of sets. The limit superior, denoted  $\lim_{n \rightarrow \infty} \sup A_n$  is defined as

$$\lim_{n \rightarrow \infty} \sup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \{x : x \in A_i \text{ for infinitely many } i\}.$$

Boole's Inequality helps us bound individual probabilities, but the next result gives us a way to determine under which conditions events have zero probability (or probability 1).

**Lemma 1.21** (Borel-Cantelli Lemma). *Suppose that  $\{A_n\}$  is a sequence of events in a probability space. If*

$$\sum_{n=1}^{\infty} P(A_n) < \infty,$$

then

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

*Proof.* It'll be easiest to consider  $P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right)$ . By monotonicity of probability measures and definition of intersection, we know

$$P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right) \leq P\left(\bigcup_{n=k}^{\infty} A_n\right),$$

so Boole's Inequality implies that

$$P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right) \leq \sum_{n=k}^{\infty} P(A_n).$$

By assumption, the sum  $\sum_{n=1}^{\infty} P(A_n)$  converges, meaning that

$$\lim_{k \rightarrow \infty} \sum_{n=k}^{\infty} P(A_n) = 0.$$

This gives us

$$P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right) \leq 0,$$

and the converse inequality follows from the fact that measures are nonnegative. Hence,

$$P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right) = 0,$$

and the result follows from the definition of limit superior.  $\square$

**Example 1.22.** Let  $\{X_n\}_{n \in \mathbb{N}}$  be a sequence of random variables where for any  $n$ ,  $P(X_n = 0) = \frac{1}{n^2}$ . Since

$$\sum_{n=1}^{\infty} P(X_n = 0) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} < \infty,$$

the Borel-Cantelli Lemma implies that the probability of  $X_n = 0$  occurring for infinitely many  $n$  is zero.

## 2. BROWNIAN MOTION

In this section, we take a closer look at random motion using collections of random variables. These help us define Brownian motion, perhaps the most famous form of random motion. Intuitively, it can be thought of as continuous random motion, or a function of time that is continuous yet rough (differentiable nowhere). We will soon see that Brownian motion arises as the limit of collections of random variables. We call these collections stochastic processes.

**Definition 2.1** (Stochastic Process). A stochastic process is a collection of a random variables  $\{X_t\}$  on a probability space  $(\Omega, \mathcal{F}, P)$  indexed by  $t \in T$  for some set  $T$ .

If  $T$  is countable, we say the stochastic process is discrete; if  $T$  is uncountable, we say the process is continuous.

**Definition 2.2** (Path). For every outcome  $\omega \in \Omega$ , the function

$$f : T \rightarrow \mathbb{R}, f(t) = X_t(\omega)$$

is called a path of the stochastic process  $\{X_t\}$ .

To develop some intuition about how stochastic processes function in practice, consider first a discrete example.

**Example 2.3.** Consider  $n$  fair 6-sided dice rolls. We can let independent and identically distributed random variables  $X_1, \dots, X_n$  represent this discrete stochastic process  $\{X_n\}$ . Here,

$$P(X_i = 1) = P(X_i = 2) = \dots = P(X_i = 6) = \frac{1}{6} \text{ for any } i.$$

The sample space  $\Omega$  consists of all outcomes  $\omega$ , where  $\omega$  encodes the dice roll for each random variable. We can use the paths  $f$  of the process  $\{X_n\}$  to represent this idea; for example,  $\omega = (6, 2, \dots, 1) \in \Omega$  would be assigned the unique function  $f : \{1, \dots, n\} \rightarrow \mathbb{R}$ , where  $f(1) = 6, f(2) = 2, \dots, f(n) = 1$ . This represents rolling a 6, 2, ..., 1 in that particular order. We know that there are  $6^n$  distinct and equally likely outcomes in the sample space. Since the sample space is finite, we can use the power set  $2^\Omega$  as our  $\sigma$ -algebra  $\mathcal{F}$ . Using  $\mathcal{F}$ , define the probability measure  $p : \mathcal{F} \rightarrow [0, 1]$ , where

$$p(E) = \sum_{\omega \in E} \frac{1}{6^n}.$$

For example, let  $E$  be the event of rolling an even number on the first and last roll. There are  $3(6^{n-1})$  outcomes that begin on an even number and  $3(6^{n-1})$  that end on an even number. Of those that begin with an even number, half end with odd numbers. Of those that end with an even number, half begin with odd numbers. Hence,

$$p(E) = \frac{3(6^{n-1}) + 3(6^{n-1})}{(2)(2)6^n} = \frac{1}{4}.$$

Now we move to define the continuous stochastic process Brownian motion. Throughout the construction, we implicitly use the same techniques from the dice roll experiment. We start with a definition.

**Definition 2.4** (Brownian Motion). A stochastic process  $\{B_t\}$  is called Brownian motion starting at  $x \in \mathbb{R}$  if

- (1)  $B_0 = x$ .
- (2) For all times

$$0 \leq t_1 \leq t_2 \leq \dots \leq t_n, \quad B_{t_n} - B_{t_{n-1}}, \dots, B_{t_2} - B_{t_1}$$

are independent random variables.

- (3) If  $s < t$ , then  $B_t - B_s$  is normally distributed with mean zero and variance  $(t - s)$ .
- (4) With probability one, the process  $\{B_t\}$  has continuous paths.

If  $x = 0$ , we call  $\{B_t\}$  standard Brownian motion. Using Lemma 1.15 as a guide, we can check that if  $\{B_t\}$  is standard Brownian motion, then if for any  $t$ ,  $X_t = \sigma B_t + \mu t$ ,  $\{X_t\}$  is also a Brownian motion with incremental expectations  $\mu$  and variances  $\sigma^2$ . This only changes the values in property (3) above. Therefore, we need only construct standard Brownian motion to show existence of this process in the general case.

**Definition 2.5** (Dyadic Rational). A rational  $d \in \mathbb{Q} \cap [0, 1]$  is called a dyadic rational if there exist nonnegative integers  $k, n$  with  $k \leq 2^n$  such that  $d = \frac{k}{2^n}$ .

Let  $\mathcal{D}_n = \{\frac{k}{2^n} : k \in \{0, 1, \dots, 2^n\}\}$  so that  $\mathcal{D} = \bigcup_{n \in \mathbb{N}} \mathcal{D}_n$  represents the set of all dyadic rationals on  $[0, 1]$ . Observe that the set of dyadic rationals is dense in  $\mathbb{R}$ , so to construct Brownian motion, we can first define it on  $\mathcal{D}$  and later extend it to the real line.

**Theorem 2.6.** *Brownian motion  $\{B_t\}_{t \in \mathcal{D}}$  exists.*

*Proof.* For every  $d \in \mathcal{D}$ , suppose there exists an independent  $N(0, 1)$  random variable labeled  $Z_d$  so that the set  $\{Z_d : d \in \mathcal{D}\}$  indexes  $\mathcal{D}$ . Let  $B_0 = 0$  and  $B_1 = Z_1$  so that  $B_t$  is defined on  $\mathcal{D}_0 = \{0, 1\}$ . The first three properties of Brownian motion are satisfied by these two random variables, so for clarity, we next define  $B_t$  for  $t \in \mathcal{D}_1 \setminus \mathcal{D}_0 = \{\frac{1}{2}\}$ . Define

$$B_{\frac{1}{2}} = \frac{B_0 + B_1}{2} + \frac{Z_{\frac{1}{2}}}{2}.$$

Intuitively,  $B_{\frac{1}{2}}$  can be thought of as the expected value of  $B_{\frac{1}{2}}$  given we know the values of  $B_0$  and  $B_1$ . However, only using this first term of the summand fails to satisfy property (2) of Brownian motion above, since

$$B_1 - \frac{B_0 + B_1}{2} = \frac{B_0 + B_1}{2} - B_0$$

are not independent. The second term provides a way to ensure that increments are both independent and are normally distributed with the correct variance as defined in property (3). By Lemma 1.18,  $B_{\frac{1}{2}}$  and  $B_1 - B_{\frac{1}{2}}$  are independent and normally distributed with mean zero and variance  $\frac{1}{2}$ , as required. Continuing inductively, assume that we have satisfied the first three properties of Brownian motion for all  $t \in \mathcal{D}_{n-1}$ . Let  $d = \frac{2k-1}{2^n} \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$  for  $k \in \{1, \dots, 2^{n-1}\}$ . Then we define

$$B_d = \frac{B_{d-2^{-n}} + B_{d+2^{-n}}}{2} + \frac{Z_d}{2^{\frac{n+2}{2}}}.$$

By this definition, observe that  $B_d - B_{d-2^{-n}}$  (an increment) can be expressed as

$$\frac{B_{d+2^{-n}} - B_{d-2^{-n}}}{2} + \frac{Z_d}{2^{\frac{n+2}{2}}}.$$

The inductive hypothesis confirms that the numerator of the first summand is normally distributed with mean zero and variance  $\frac{1}{2^n}$ , so by using Lemma 1.18, we can rewrite the expression as

$$B_d - B_{d-2^{-n}} = \frac{X + Y}{2^{\frac{n+2}{2}}},$$

where  $X, Y \sim N(0, 1)$ . By the same argument,

$$B_{d+2^{-n}} - B_d = \frac{X - Y}{2^{\frac{n+2}{2}}}.$$

Applying Lemma 1.18 once more gives us the result that

$$B_d - B_{d-2^{-n}} \text{ and } B_{d+2^{-n}} - B_d$$

are independent and normally distributed with mean zero and variance  $\frac{1}{2^{n+1}}$ . By induction, we have defined  $B_t$  on the dyadic rationals to satisfy the first three properties above.  $\square$

**Theorem 2.7.** *There exists a continuous random function  $B_t : [0, 1] \rightarrow \mathbb{R}$  that is a standard Brownian motion.*

*Proof.* We use linear interpolation to connect the points of  $B_t$  defined on  $\mathcal{D}$  above. Define a sequence of functions  $\{F_n\}_{n \in \mathbb{N}}$  with

$$F_0(t) = \begin{cases} Z_1, & t = 1 \\ 0, & t = 0 \\ \text{linear in between} \end{cases} \quad \text{and } F_n(t) = \begin{cases} 2^{-\frac{n+2}{2}} Z_t, & t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1} \\ 0, & t \in \mathcal{D}_{n-1} \\ \text{linear in between.} \end{cases}$$

Since these functions are piecewise linear, they are continuous on  $[0, 1]$ ; further, we claim that for any  $n$  and  $d \in \mathcal{D}_n$ ,

$$B_d = \sum_{i=0}^n F_i(d).$$

By construction, this would imply

$$B_d = \sum_{i=0}^{\infty} F_i(d)$$

on the dyadic rationals. To prove the above, we use induction on  $n$  once again. By the definition of  $F_0$ , the case for  $n = 0$  holds immediately. Assume that for  $d \in \mathcal{D}_{n-1}$ ,

$$B_d = \sum_{i=0}^{n-1} F_i(d) = \sum_{i=0}^{\infty} F_i(d).$$

Since we defined each function to be linear on  $[d - 2^{-n}, d + 2^{-n}]$ , we have that for  $d \in \mathcal{D} \setminus \mathcal{D}_{n-1}$ ,

$$B_d = F_n(d) + \sum_{i=0}^{n-1} F_i(d) = F_n(d) + \sum_{i=0}^{n-1} \frac{F_i(d - 2^{-n}) + F_i(d + 2^{-n})}{2},$$

and finally, by inductive assumption, this equals

$$F_n(d) + \frac{B_{d-2^{-n}} + B_{d+2^{-n}}}{2} = \frac{B_{d-2^{-n}} + B_{d+2^{-n}}}{2} + \frac{Z_d}{2^{\frac{n+2}{2}}},$$

as desired.

Next, we claim that the series  $\sum_{n=0}^{\infty} F_n$  uniformly converges on  $[0,1]$ , which would imply the continuity of  $B_t$  for all values  $t \in [0,1]$  (since  $\mathcal{D}$  is dense in  $\mathbb{R}$ ). First, we look to bound  $Z_d$ . It is known that if a random variable is normally distributed with mean zero and variance 1, then for all  $x > 0$ ,

$$\frac{x}{x^2 + 1} \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right) \leq P\{X > x\} \leq \frac{1}{x} \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right).$$

For the proof of this fact, see 2.9 in the appendix of [11].

Since  $Z_d \sim N(0,1)$ , the inequality above implies that for  $c > 1$  and large  $n$ ,

$$P(|Z_d| \geq c\sqrt{n}) \leq \frac{1}{c\sqrt{n}} \frac{1}{2\pi} \exp\left(-\frac{c^2 n}{2}\right) \leq \exp\left(-\frac{c^2 n}{2}\right).$$

For  $n \in \mathbb{N}$  define the event

$$E_n = \text{there exists } d \in \mathcal{D}_n : |Z_d| \geq c\sqrt{n}.$$

Now we can see that

$$\sum_{n=0}^{\infty} P(E_n) \leq \sum_{n=0}^{\infty} \sum_{d \in \mathcal{D}_n} P(|Z_d| \geq c\sqrt{n}) \leq \sum_{n=0}^{\infty} (2^n + 1) \exp\left(-\frac{c^2 n}{2}\right),$$

which converges by ratio test if  $c > \sqrt{2 \log(2)}$ . Suppose the series above converges. Then the Borel-Cantelli Lemma implies there exists  $N \in \mathbb{N}$  such that for all  $n > N$  and  $d \in \mathcal{D}_n$ ,  $|Z_d| < c\sqrt{n}$ . Hence, we know that for  $n > N$ ,

$$\|F_n\|_{\infty} \leq \frac{c\sqrt{n}}{2^{\frac{n+2}{2}}} < \frac{c\sqrt{n}}{2^{\frac{n}{2}}},$$

so by the Weierstrass M-test,  $\sum_{n=0}^{\infty} F_n$  uniformly converges on  $[0,1]$ . We now have a continuous limit function  $\{B_t\}$  for  $t \in [0,1]$ .

We now need to verify property (2) and (3) for  $t \in [0,1]$ . Suppose  $t_1 < t_2 < \dots < t_n \in [0,1]$ . Since  $\mathcal{D}$  is dense in  $\mathbb{R}$ , choose sequences in  $\mathcal{D}$  labeled  $t_{1,k} \leq t_{2,k} \leq \dots \leq t_{n,k}$  such that for any  $i \in \{1, \dots, n\}$ ,

$$\lim_{k \rightarrow \infty} t_{i,k} = t_i.$$

Since  $B_t$  is continuous we have for any  $i$  with  $1 \leq i \leq n-1$ ,

$$\mathbb{E}[B_{t_{i+1}} - B_{t_i}] = \mathbb{E}\left[\lim_{k \rightarrow \infty} (B_{t_{i+1,k}} - B_{t_{i,k}})\right] = \lim_{k \rightarrow \infty} \mathbb{E}[B_{t_{i+1,k}} - B_{t_{i,k}}] = 0.$$

Checking variances of increments is similar and the full details can be seen on page 11 of [11].  $\square$

**Theorem 2.8.** *There exists a continuous random function  $B_t : [0, \infty) \rightarrow \mathbb{R}$  that is a standard Brownian motion.*

*Proof.* We can stitch together the pieces of many intervals of Brownian motion with length one. Consider a sequence  $B^0, B^1, \dots$  of independent random variables with the same distribution of  $B_t$  above. To satisfy the four properties that define Brownian motion, let

$$B_t = B_{t - \lfloor t \rfloor}^{[\lfloor t \rfloor]} + \sum_{i=0}^{[\lfloor t \rfloor] - 1} B_i.$$

Then properties (2) and (3) of Brownian Motion above follow from the distributions of the sequence  $\{B^n\}_{n \in \mathbb{N}}$ , and continuity follows from the continuity of the pieces.  $\square$

### 3. GRAPHS AND RANDOM TREES

Before introducing trees, we must first develop some graph theory. This background will give us the tools to understand random trees and eventually the connection between random trees and random motion.

**Definition 3.1** (Graph). A graph is a pair  $G = (V, E)$  where  $V$  is a set of vertices and  $E \subseteq \{(u, v) : u, v \in V\}$  is a set of edges that join two vertices.

For the remainder of this section, we let  $G$  be a graph. In reference to  $G$ , we denote  $V(G)$  and  $E(G)$  as the set of vertices and edges, respectively. For this paper, we consider the edge  $(u, v)$  and  $(v, u)$  to be indistinguishable (in other words, we are considering undirected graphs). The following definitions provide some notation.

**Definition 3.2** (Adjacency). Two vertices  $u, v \in V$  are adjacent in  $G$  if  $(u, v) \in E(G)$ .

**Definition 3.3** (Finite Walk, Trail, Path). A finite walk in  $G$  is an alternating sequence of vertices and edges  $(v_1, e_1, v_2, e_2, \dots, e_n, v_{n+1})$  where for any  $i \in \{1, \dots, n\}$ ,  $e_i$  connects  $v_i$  and  $v_{i+1}$ . An infinite walk has no first or last vertex. A trail in  $G$  is a walk with all edges being distinct. A path in  $G$  is a trail with all vertices being distinct.

**Definition 3.4** (Connected).  $G$  is called connected if for every  $u, v \in V(G)$  there exists a path between  $u$  and  $v$  in  $G$ .

**Definition 3.5** (Cycle). A cycle in  $G$  is a path that starts and ends with the same vertex. In other words, let  $P = (v_1, e_1, \dots, e_{n-1}, v_n)$  be a path. Then  $P$  is a cycle if  $v_1 = v_n$ .

**Definition 3.6** (Discrete Tree). A discrete tree is a connected graph with no cycles.

**Lemma 3.7.** *If  $T$  is a discrete tree, then for any  $u, v \in V(T)$ , there exists a unique path between  $u$  and  $v$  in  $T$ .*

*Proof.*  $T$  must be connected with no cycles. Hence, there exists at least one path between  $u$  and  $v$ . Suppose two distinct paths  $P_1$  and  $P_2$  connect  $u$  and  $v$ . Let  $P_1$  traverse the vertices  $(u = a_1, \dots, a_n = v)$  and  $P_2$  traverse  $(u = b_1, \dots, b_m = v)$ . Let  $i$  be the smallest integer with  $a_i \neq b_i$ . Since  $b_m = a_n$ , there exists a set

$$J = \{j \in \{i, \dots, m\} : b_j = a_k \text{ for some } k \in \{i, \dots, n\}\}.$$

Let  $j$  be the minimum element of  $J$ , with  $b_j = a_k$  for some  $k$ . Then the path traversing  $a_{i-1}, a_i, \dots, a_k = b_j, b_{j-1}, \dots, b_{i-1} = a_{i-1}$  is a cycle in  $G$ , implying  $T$  is not a tree, a contradiction. Hence, the path between  $u$  and  $v$  must be unique.  $\square$

**Definition 3.8** (Degree). The degree  $d(v)$  of a vertex  $v$  in a graph is the number of edges incident to  $v$ .

**Definition 3.9** (Leaf). A leaf in a tree is a vertex adjacent to only one other vertex. A vertex  $v$  is a leaf if and only if it has degree  $d(v) = 1$ .

**Definition 3.10 (Root).** A root is a distinguished vertex  $v$  in a tree  $T$ . We denote  $(T, v)$  for a tree with root  $v$ .

**Definition 3.11 (Random Tree).** A random tree is a tree formed by a stochastic process.

Though we can define what it means to be a random tree, it still is not clear how random trees grow. Intuitively, we can think of random trees as a branching process that relies on relationships between vertices. We imagine starting from a root vertex  $v$  that branches out into  $k$  offspring via some probability distribution. In the same fashion, each of those  $k$  children of  $v$  branch into some number of offspring. We codify this branching process using offspring distributions.

**Definition 3.12 (Offspring Distribution).** The offspring distribution  $p(k), k \geq 0$  of a branching process is a probability mass function giving the probability of an individual vertex having  $k$  children (where different individuals reproduce independently).

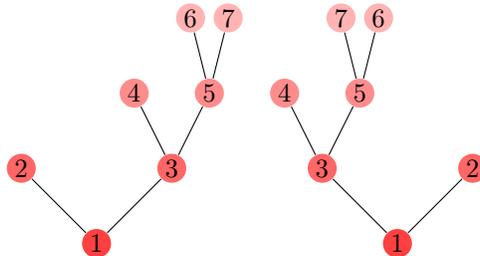
These types of processes are sometimes called Galton-Watson branching processes. When we consider trees built from Galton-Watson branching processes, they are called Galton-Watson trees. With or without a branching process, our trees still have no labeling system (or way to distinguish vertices), so let's consider the set of unordered labeled trees with  $n$  vertices  $\mathbb{T}_n$ . A tree  $t \in \mathbb{T}_n$  is labeled by elements of the set  $\{1, 2, \dots, n\}$ , where each vertex receives a unique label. We can easily embed any element of  $\mathbb{T}_n$  in the plane, and we define trees to stay unique up to orientation (or ordering of offspring). From this definition, we can see that  $\mathbb{T}_2$  has just one element:



In a similar manner,  $\mathbb{T}_3$  has three elements, and the elements of  $\mathbb{T}_4$  consist of



with the labels shuffled around. There are  $\frac{4!}{2} = 12$  ways to uniquely draw the first configuration and only 4 ways to uniquely draw the second configuration, meaning there are 16 total unordered labeled trees with 4 vertices. For clarity, the following two trees are considered the same since ancestor relationships are preserved (and the order of vertices does not matter).



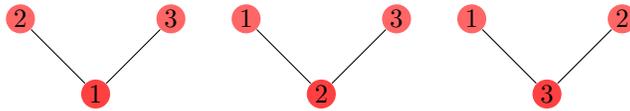
We can generalize this with the following result.

**Proposition 3.13** (Cayley's Formula). *The set of unordered labeled trees with  $n$  vertices has  $n^{n-2}$  elements.*

*Proof.* We claim there exists a bijection between  $\mathbb{T}_n$  and sequences of natural numbers  $\{1, \dots, n\}$  with length  $n - 2$ . Suppose  $t \in \mathbb{T}_n$  with vertex labels  $\{v_1, v_2, \dots, v_n\}$  where  $v_1 < v_2 < \dots < v_n$ . Construct a sequence of natural numbers with length  $n - 2$  as follows:

- Remove the leaf with the lowest label from the tree, recording the leaf's unique neighbor as an element of the sequence.
- Repeat until there are two vertices left.

Consider the case  $n = 3$  and the set  $\mathbb{T}_3$ , which consists of the three trees



The process above yields the three sequences  $\{1\}$ ,  $\{2\}$ , and  $\{3\}$ , which are clearly the only sequences feasible using  $\{1, 2, 3\}$  and having length 1. Continuing inductively, suppose that the mapping is a bijection for all  $k$  with  $3 \leq k < n$ . Further, let  $t_1, t_2 \in \mathbb{T}_n$  be different. Observe that for any vertex  $v_i$  in a tree, the number of appearances of  $v_i$  in the sequence must be  $d(v_i) - 1$  by construction. To show that the sequences constructed by  $t_1, t_2$  must be different, consider the following:

- If the lowest labeled leaves of  $t_1$  and  $t_2$  ( $v_1$  and  $v_2$ , respectively) are different, suppose without loss of generality that  $v_1 < v_2$ . The lowest labeled leaf of  $t_1$  won't appear in the sequence for  $t_1$ , but that  $v_1 < v_2$  implies the label  $v_1$  is not a leaf in  $t_2$ , meaning it must appear at least once in the sequence constructed by  $t_2$ .
- If the lowest labeled leaves are the same, but the neighbors are different, by construction the first number in the sequence will be different.
- If both the lowest labeled leaf and associated neighbor are the same, our iterative process removes the leaf, and the resulting smaller trees are still different. By the inductive hypothesis, therefore, the sequences are different.

Hence, the mapping is injective. Now suppose for some  $n \geq 4$  we have an arbitrary sequence of length  $n - 2$

$$A = (a_1, a_2, \dots, a_{n-2})$$

constructed with numbers in  $\{v_1, \dots, v_n\}$ . We need to show there exists a tree that maps to the sequence via the previous construction. First, find the lowest number in  $\{v_1, \dots, v_n\}$  that does not appear in the sequence – call it  $v_i$ . Define

$$A' = (a_2, \dots, a_{n-2}),$$

which is a sequence with length  $n - 3$  constructed with numbers in  $\{v_1, \dots, v_n\} \setminus \{v_i\}$ . By the inductive hypothesis, there exists a tree  $T'$  that maps to  $A'$ . We know that  $a_1$  can be found in  $T'$  as a label, so attach the vertex  $v_i$  as a leaf with the unique neighbor labeled  $a_1$  in  $T'$ . Call this new tree  $T$ . The sequence associated with  $T$  must start with  $a_1$  (since  $v_i$  is the lowest labeled leaf in  $T$  by construction and its unique neighbor is the vertex labeled with  $a_1$ ) and continue with  $A'$ , meaning it is indeed  $A$ . Hence, the mapping is surjective. The number of possible sequences

of natural numbers  $\{1, \dots, n\}$  of length  $n - 2$  is  $n^{n-2}$ , which implies  $\mathbb{T}_n$  has  $n^{n-2}$  elements.  $\square$

**Notation 3.14.** The uniform random tree  $t \in \mathbb{T}_n$  is a tree drawn at random from the set  $\mathbb{T}_n$ .

We have seen that picking a tree at random from  $\mathbb{T}_n$  is the same as a uniform random sample from sequences built with the numbers  $\{1, \dots, n - 2\}$ . We now aim to show a connection between branching processes and the uniform random tree.

**Definition 3.15** (Poisson Distribution). The Poisson distribution is a discrete probability distribution with probability mass function

$$\frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \mathbb{N} \cup \{0\}.$$

The probability mass function can intuitively be thought of the probability of finding  $k$  events in a given length of time if the events happen independently at an average rate of  $\lambda$ . We say a random variable has a *Poisson*( $\lambda$ ) distribution. The connection between Galton-Watson branching processes and the uniform random tree relies on the Poisson distribution in the following proposition.

**Proposition 3.16.** *The uniform random tree sampled from  $\mathbb{T}_n$  has the same distribution as a Galton-Watson tree with*

- *Poisson(1) offspring distribution*
- *$n$  total progeny*
- *Vertices labeled randomly from the set  $\{1, \dots, n\}$*

*Proof.* Let  $G$  be an ordered tree generated by the Galton-Watson process with *Poisson*(1) offspring distribution, and suppose  $t$  is a fixed, ordered, and labeled tree with  $n$  vertices. Let the vertices of  $t$   $\{v_1, \dots, v_n\}$  have offspring  $c(v_1) = c_1, \dots, c(v_n) = c_n$ . We know that

$$P(G = t) = \prod_{i=1}^n \frac{e^{-1}}{c_i!} = e^{-n} \prod_{i=1}^n \frac{1}{c_i!}.$$

Since the probability that the Galton-Watson tree has  $n$  vertices only depends on  $n$ , we know that

$$P(G = t \mid G \text{ has } n \text{ total vertices}) = f(n) \prod_{i=1}^n \frac{1}{c_i!}$$

for some function  $f$ . Let  $T$  be a random variable representing a uniform random sample from  $\mathbb{T}_n$ , and define

$t' = t$  above, but with randomly assigned vertex labels.

Since  $P(T = t') = \frac{1}{n^{n-2}}$  only depends on  $n$  by Cayley's Formula, it suffices to show that the probability of  $G = t'$  conditioning on

- $G$  having  $n$  total vertices
- Assigning  $G$  random vertex labels from  $\{1, \dots, n\}$

only depends on  $n$ . Call this outcome  $E$ . We've already conditioned on the first item, so consider labeling  $G$  with the set  $\{1, \dots, n\}$ . There are  $n!$  total ways to do so, but  $\prod_{i=1}^n c_i!$  of them give rise to  $t'$ , so we have that

$$P(E) = \left( f(n) \prod_{i=1}^n \frac{1}{c_i!} \right) \left( \frac{\prod_{i=1}^n c_i!}{n!} \right) = \frac{f(n)}{n!}.$$

This probability does not depend on a choice of  $t$ , meaning the distribution is uniform. Hence, the conditioned Galton-Watson tree has the same distribution as the uniform random tree sampled from  $\mathbb{T}_n$ .  $\square$

#### 4. (LIMITS OF) UNIFORM PLANE TREES

Until now, we have only dealt with discrete trees. However, just as discrete random motion has a continuous analogue (in Brownian motion), so too do branching processes. In this section, we aim to show how random motion can encode random trees, and near the end of the section, we elucidate the connection between Brownian motion and a special type of random tree. As is often the case when trying to turn discrete processes into continuous ones, convergence is key.

**Definition 4.1** (Convergence in Distribution). A sequence of random variables  $X_1, X_2, \dots, X_n$  converges in distribution to a random variable  $X$  if the sequence of CDFs  $F_{X_1}, F_{X_2}, \dots, F_{X_n}$  converges to  $F_X$ . In other words,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

for all  $x \in \mathbb{R}$  where  $F_X$  is continuous.

To set up the convergence later in the section, we use a Galton-Watson tree with a specific offspring distribution called a uniform (rooted) plane tree.

**Definition 4.2** (Uniform Plane Tree). A uniform (rooted) plane tree is a Galton-Watson tree formed with offspring distribution  $p(k) = (\frac{1}{2})^{k+1}$ .

**Definition 4.3** (Walk Excursion). A walk excursion of size  $2n$  is a function

$$f : \{0, \dots, 2n\} \rightarrow \mathbb{N} \cup \{0\}$$

such that  $f(0) = 0 = f(2n)$  and for any  $j \in \{1, \dots, 2n\}$ ,

$$f(j) - f(j-1) \in \{-1, 1\}.$$

**Theorem 4.4.** *There exists a bijection between the set of uniform plane trees with  $n$  edges and the set of walk excursions with  $2n$  steps.*

*Proof.* Consider  $(T, v)$ , a uniform plane tree  $T$  with root  $v$  and  $n$  edges. We can embed  $T$  into the plane so that the edges have length one. Imagine a particle starting at the root of  $T$  at time  $t = 0$  and exploring  $T$  at unit speed via a left-to-right depth-first traversal. A depth-first traversal is a tree-searching algorithm that, beginning at the root, travels as far down the tree as possible on a given branch before backtracking to find new paths. A left-to-right depth-first traversal is one that starts at the leftmost branch in the plane embedding of the tree and moves to the right.

Let the distance from the particle to the root of the tree be encoded by a contour function  $C_T : \mathbb{R} \rightarrow \mathbb{R}$  that maps a time  $t$  to the distance on the tree the particle is from the root vertex. Since each edge is traversed exactly twice in the depth-first traversal, we must have  $t \in [0, 2n]$ . We also know that for any  $j \in \{1, \dots, 2n\}$ ,

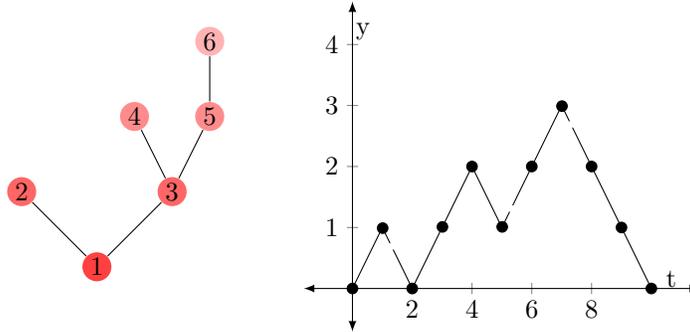
$$C_T(j) - C_T(j-1) = 1 \text{ if visiting an edge for the first time at step } j \text{ and}$$

$$C_T(j) - C_T(j-1) = -1 \text{ if visiting an edge for the second time at step } j.$$

Furthermore, since the depth-first search starts and ends at the root,

$$C_T(0) = 0 = C_T(2n).$$

By definition, then,  $C_T$  is a linearly interpolated walk excursion with  $2n$  steps. Below is an example tree and its contour function (with labels only to emphasize the order vertices are seen in the depth-first traversal).



Now consider a walk excursion with  $2n$  steps. First, linearly interpolate the points on its graph. We can then encode the graph by labeling steps upward with a  $u$  and steps downward with a  $d$ . For example, the contour function above would correspond to the string

$$uduuduudd.$$

To construct a uniform plane tree, first draw a root. Then pass through the string with these rules:

- If the next item in the encoding is a  $u$ , create an edge from current vertex to a new vertex.
- If the next item is a  $d$ , travel (back) across the only edge possible.

Walk excursions are never negative, so being able to execute the second step is guaranteed.  $\square$

By this encoding, leaves are local maximums of the contour function. We next showcase another interesting result, which tells us the size of the set of uniform plane trees.

**Theorem 4.5.** *Let  $A$  be the set of uniform plane trees with  $n$  edges. Then*

$$|A| = \frac{1}{n+1} \binom{2n}{n}.$$

*Proof.* By Theorem 4.4, we can count the number of walk excursions of size  $2n$ . The total number of paths with  $y$ -increments in  $\{-1, 1\}$  that begin at  $(0, 0)$  and end at  $(2n, 0)$  is  $\binom{2n}{n}$  since  $n$  upward steps need to be present. To find  $|A|$ , we need to remove the paths that ever have a  $y$ -coordinate of  $-1$ . Let  $B$  contain all paths  $f$  that satisfy this condition. For any such path  $f$ , let  $T$  be the first time that the path has a  $y$ -coordinate of  $-1$ . Define

$$g(t) = \begin{cases} -2 - f(t) & t < T \\ f(t) & T \geq t. \end{cases}$$

We see that  $g$  also either has  $y$ -steps of  $1$  or  $-1$ , except every path  $g$  defined in this way starts at  $(0, -2)$  and ends at  $(2n, 0)$ . Since every path in  $B$  can be manipulated this way, there exists a bijection between paths in  $B$  and paths with  $y$ -increments in  $\{-1, 1\}$  and starting at  $(0, -2)$  and ending at  $(2n, 0)$  (to recover  $f$  from  $g$ , we

define  $T$  in the same way as above and apply the inverse transformation). The number of paths in  $B$ , therefore, must be  $\binom{2n}{n+1}$  by the same logic as above. Hence, we have

$$|A| = \binom{2n}{n} - \binom{2n}{n+1} = \frac{1}{n+1} \binom{2n}{n}.$$

□

We've seen that if a tree is a random sample from the set of uniform plane trees, it has a corresponding contour function, which is a uniform random sample from paths starting and ending at the origin conditioned to always stay nonnegative. Now consider a more general random walk. Suppose  $X_1, X_2, \dots$  are independent and identically distributed random variables with

$$P(X_i = 1) = \frac{1}{2} = P(X_i = -1), \quad i \geq 1.$$

Let  $X_0 = 0$ . Define

$$S_n = \sum_{i=1}^n X_i$$

as a random walk.

This is a set of points plotted in the plane, but we can view  $S$  as a continuous function  $S : [0, \infty) \rightarrow \mathbb{R}$  by linearly interpolating (piecewise) the points of  $S$ . This defines a function  $S(t), t \geq 0$ .

**Theorem 4.6** (Donsker's Invariance Principle). *Let  $B_t$  be a standard Brownian motion. Then as  $n \rightarrow \infty$ , the process*

$$\frac{1}{\sqrt{n}}(S(\lfloor nt \rfloor), t \geq 0)$$

*converges in distribution to  $(B_t, t \geq 0)$ .*

*Proof.* The proof of this theorem is quite involved; for a rigorous construction, see section 6 of [13]. □

Since this connection appears between simple random walks and Brownian motion, it intuitively seems as if there should exist a continuous analogue to walk excursions. We define a standard Brownian excursion  $\mathcal{E}_t$  informally as Brownian motion conditioned to stay nonnegative on the interval  $[0, 1]$  where  $\mathcal{E}_0 = 0 = \mathcal{E}_1$ . This is by no means precise, but for the purposes of this paper, the intuitive explanation will do. For a more detailed treatment of this topic, see section 1.3 of [3]. Using this definition, we can extract a similar connection.

**Theorem 4.7.** *Let  $\mathcal{E}_t$  be a standard Brownian excursion. Then if  $C_n$  is a random walk conditioned to have a  $y$ -coordinate of  $-1$  for the first time at step  $n$  (i.e., a walk excursion), as  $n \rightarrow \infty$  we have*

$$\left( \frac{1}{\sqrt{n}} C_n(\lfloor nt \rfloor), 0 \leq t \leq 1 \right) \rightarrow (\mathcal{E}_t, 0 \leq t \leq 1),$$

*where the convergence is in distribution.*

*Proof.* See [8] for the proof, which can be seen as an extension to Donsker's Invariance Principle. □

In the discrete case, we constructed an explicit bijection between walk excursions of size  $2n$  and uniform plane trees with  $n$  edges. If, as  $n \rightarrow \infty$ , walk excursions converge in distribution to a type of Brownian motion, we should also be able to find a way to extract a continuous analogue of a uniform plane tree using  $\mathcal{E}_t$  as defined above. Aldous [1] showed that this indeed can be done, and it has been named the Brownian continuum random tree (CRT).

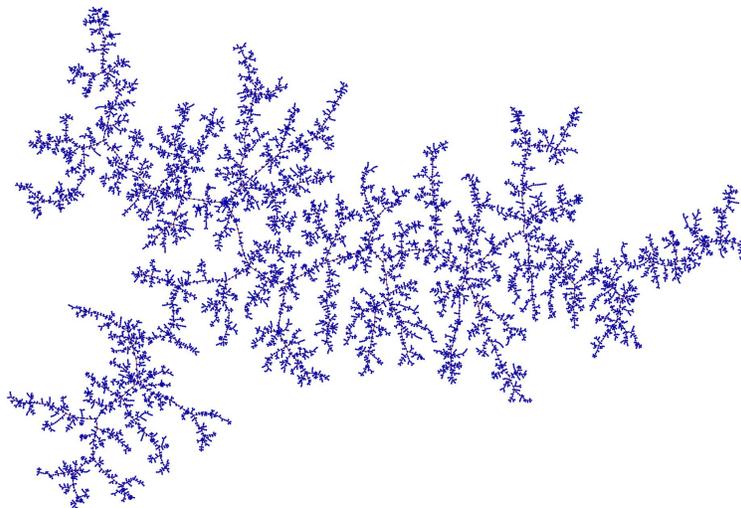


FIGURE 1. A simulation from [9] of Aldous' continuum random tree.

Though we can show the convergence of the contour function to the Brownian excursion and extract the CRT, it cannot immediately be concluded that the limit of the uniform plane tree as the number of edges (or vertices) grows large converges to the Brownian continuum random tree. In fact, we must define the conditions under which the tree converges; see section 2 of [3] for more details. Just as discrete random walks cleanly connect to trees and branching processes, under these conditions, there also exists a neat link between continuous random motion and the continuum random tree.

#### ACKNOWLEDGMENTS

I would like to thank Katie Gravel, my mentor, for her invaluable support in answering my questions and working through results with me. I also express my gratitude to Professor Ewain Gwynne for his lecture series and notes on random trees, which inspired the latter sections of this paper. Finally, I am also pleased to thank Professor Daniil Rudenko for his lectures and coordinating of the apprentice program and Professor Peter May for his organization of the UChicago Math REU.

#### REFERENCES

- [1] David Aldous. The continuum random tree. 1991. <https://doi.org/10.1214/aop/1176990534>.

- [2] Allen Dickson. An introduction to graph theory. 2006. [https://www.math.utah.edu/mathcircle/notes/MC\\_Graph\\_Theory.pdf](https://www.math.utah.edu/mathcircle/notes/MC_Graph_Theory.pdf).
- [3] Jean-Francois Le Gall. Random trees and applications. 2005. <https://www.imo.universite-paris-saclay.fr/~jfllegall/Cornell.pdf>.
- [4] David Galvin. Prufer codes and cayley's formula. 2012. [https://www3.nd.edu/~dgalvin1/40210/40210\\_F12/Prufer.pdf](https://www3.nd.edu/~dgalvin1/40210/40210_F12/Prufer.pdf).
- [5] Christina Goldschmidt. An introduction to random trees. 2015. <https://people.bath.ac.uk/ak257/talks/Goldschmidt.pdf>.
- [6] David Guichard. Combinatorics and graph theory. 2022. [https://whitman.edu/mathematics/cgt\\_online/book.pdf](https://whitman.edu/mathematics/cgt_online/book.pdf).
- [7] Ewain Gwynne. Notes from CRT talk. 2022.
- [8] W.D. Kaigh. An invariance principle for random walk conditioned by a late return to zero. 1976. <https://www.jstor.org/stable/2959376>.
- [9] Igor Kortchemski. Random trees. 2021. <https://igor-kortchemski.perso.math.cnrs.fr/images.html>.
- [10] Gregory F. Lawler. Stochastic calculus: An introduction with applications. 2014. <http://www.math.uchicago.edu/~lawler/finbook.pdf>.
- [11] Peter Morters and Yuval Peres. Brownian motion. 2022. <https://people.bath.ac.uk/maspm/book.pdf>.
- [12] University of Connecticut OER. Probability inequalities. 2020. <https://probability.oer.math.uconn.edu/wp-content/uploads/sites/2187/2020/08/ch15M.pdf>.
- [13] Ethan Schondorf. The wiener measure and dontsker's invariance principle. 2019. <https://math.uchicago.edu/~may/REU2019/REUPapers/Schondorf.pdf>.