1. Random Walks and Wiener Measure

I will try to make precise that 'Brownian motion is the limit of random walks'.

I'll start with what actually is a random walk...

Definition 1.1. Let $(\epsilon_k)_{\mathbf{N}}$ be iid random variables[1] with zero mean and variance $\sigma^2 > 0$. With $\sum^n \epsilon_k := S_n$ and making $S_0 = 0$ almost surely, call $(S_n)_{\geq 0}$ a random walk.

The case where the ϵ_n are coin tosses is the 'usual' random walk we're probably used to.

Given an $n \in \mathbf{N}$, $\omega \in \Omega$, we can use the random variable S_n to get an element $X^n\omega : t \mapsto X^n_t\omega$ of C as follows:

Divide $[0, 1]$ equally into n subintervals. At the endpoints of each interval, which are the $\frac{i}{n}$, $i \in \{0, \dots n\}$, make $X^n_{\frac{i}{n}}\omega = \frac{S_i\omega}{\sigma\sqrt{n}}$. Within each interval, make $X^n_t\omega$ to be the line joining $X^n_{\frac{i}{n}}\omega$ and $X^n_{\frac{i+1}{n}}\omega$. $X^n\omega$ is continuous because it is piecewise linear.

But more than this is true. If we give C the uniform topology, it becomes a measurable space with the Borel σ-algebra, $\mathcal{C}$. Now we can ask whether $X^n : \omega \mapsto X^n\omega$ is a random element of C.

If Y is a random element of C, then[2] $\pi_{t_1 \dots t_k}Y =: (Y_{t_1}, \dots Y_{t_k})$ is a random vector, and so each $\pi_t Y =: Y_t$, $t \in [0, 1]$ has to be a random variable. Conversely, if Y maps Ω into C and each Y_t is a random variable, $t \in [0, 1]$, then $\pi_{t_1 \dots t_k}Y$ is a random vector, ie if $A \in \mathcal{B}\mathbf{R^k}$ then $(\pi_{t_1 \dots t_k}Y)^{-1} A = Y^{-1}\pi^{-1}_{t_1 \dots t_k}A \in \mathcal{F}$. But[3] the *finite dimensional sets* $\pi^{-1}_{t_1 \dots t_k}A$ actually generate $\mathcal{C}$. So Y is a random element iff each Y_t is a random variable, $t \in [0, 1]$.

Now by writing out X^n explicitly as $X^n_t\omega = \frac{S_{\lfloor nt \rfloor}\omega}{\sigma\sqrt{n}} + (nt - \lfloor nt \rfloor)\frac{\epsilon_{\lfloor nt \rfloor + 1}\omega}{\sigma\sqrt{n}}$, we can see that each X^n_t is a random variable, and so X^n is a random element of C.

Now that we know this, we can start putting measures on C. Call P_n the push-forward by X^n of $\mathbf{P}$ onto C. We can now ask whether our sequence of measures (P_n) have a limit, ie whether they converge weakly to some measure on C (giving meaning to the statement 'limit of random walks').

If indeed the $\mathsf{P_n}$ have a limit P, then because[4] the projections $\pi_{t_1 \dots t_k}$ are continuous, $\mathsf{P}_n\pi^{-1}_{t_1 \dots t_k} \Rightarrow \mathsf{P}\pi^{-1}_{t_1 \dots t_k}$. This does work in the other direction, but only if the P_n are 'tight'. More on tightness later, but the first step would be to find limits of the measures $\mathsf{P}_n\pi^{-1}_{t_1 \dots t_k}$ on $\mathbf{R^k}$. The $\mathsf{P}_n\pi^{-1}_{t_1 \dots t_k}$ are the distributions of the random vectors $\left(X^n_{t_1}, \dots, X^n_{t_k}\right)$, the *finite dimensional distributions* of X^n. So we are looking for the limiting finite dimensional distributions of X^n. Let's begin by just seeing whether $X^n_t = \mathsf{P}\pi^{-1}_t$ has a limiting distribution, $t \in [0, 1]$.

---

[1] any mention of 'random element' should be understood as being a measurable function defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and the term 'random variable/vector' shall be used when the random element is $\mathbf{R}/\mathbf{R}^k$-valued

[2] for $f \in C[0, 1]$, $\pi_{t_1 \dots t_k}f$ is the random vector $(f(t_1), \dots, f(t_k))$, and $t_1, \dots t_k$ are always so that $0 \leq t_1 \leq \dots \leq t_k \leq 1$ and $k \in \mathbf{N}$

[3] I won't prove it, but I don't think it is too far from being routine

[4] this comes from a change of variable in the definition of weak convergence

First write $X^n_t = \sqrt{\frac{\lfloor nt\rfloor}{n}}\frac{S_{\lfloor nt\rfloor}}{\sigma\sqrt{\lfloor nt\rfloor}} + \xi^n_t$, where $\xi^n_t := (nt - \lfloor nt\rfloor)\frac{\epsilon_{\lfloor nt\rfloor+1}}{\sigma\sqrt{n}}$, $t \in [0,1]$. Chebyshev's inequality shows that the ξ^n_t go to zero in probability (in n), and so weakly too. $\frac{S_n}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}$, by the Central Limit theorem, where $\mathcal{N}$ is the standard Gaussian distribution[5]. Since $\left(\mathbf{E}f\left(\frac{S_{\lfloor nt\rfloor}}{\sigma\sqrt{\lfloor nt\rfloor}}\right)\right)_n$ is a subsequence of $\left(\mathbf{E}f\left(\frac{S_n}{\sigma\sqrt{n}}\right)\right)_n$, f bounded continuous, $\frac{S_{\lfloor nt\rfloor}}{\sigma\sqrt{\lfloor nt\rfloor}} \Rightarrow \mathcal{N}$ too. Lastly, note that $\sqrt{\frac{\lfloor nt\rfloor}{n}} \to \sqrt{t}$.

Result 1.1. *Let (a_n) be a real sequence, (Y_n), Y be random variables/vectors. Suppose $a_n \to a$ and $Y_n \Rightarrow Y$. Then $a_nY_n \Rightarrow aY$.*

Result 1.2. *Let (Y_n), Y, and (Z_n) be random variables/vectors. Suppose $Y_n \Rightarrow Y$, $|Z_n| \Rightarrow 0$. Then $Y_n + Z_n \Rightarrow Y$.*

So using Result 1.1, $\sqrt{\frac{\lfloor nt\rfloor}{n}}\frac{S_{\lfloor nt\rfloor}}{\sigma\sqrt{\lfloor nt\rfloor}} \Rightarrow \sqrt{t}N \sim \mathcal{N}_t$, where N is a standard Gaussian random variable (so $\mathcal{N}_t$ is the distribution of a 0-mean t-variance Gaussian random variable) and finally using Result 1.2, $X^n_t \Rightarrow \mathcal{N}_t$.

Now let $s, t \in [0,1]$, $s \le t$. Instead of looking at the random vector (X^n_s, X^n_t), which might seem like the obvious extension towards finite dimensional distributions, let's look at $(X^n_s, X^n_t - X^n_s) = \left(\frac{S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}}, \frac{S_{\lfloor nt\rfloor}-S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}}\right) + (\xi^n_s, \xi^n_t - \xi^n_s)$, the key reason being that $S_{\lfloor ns\rfloor}$ and $S_{\lfloor nt\rfloor} - S_{\lfloor ns\rfloor}$ are *independent*. First, $(\xi^n_s, \xi^n_t - \xi^n_s) \Rightarrow 0$ by Chebyshev again. Now $\frac{S_{\lfloor nt\rfloor}-S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}}$ has the same distribution as $\frac{S_{\lfloor nt\rfloor-\lfloor ns\rfloor}}{\sigma\sqrt{n}}$, and similarly to before, $\frac{S_{\lfloor nt\rfloor-\lfloor ns\rfloor}}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}_{t-s}$, so that $\frac{S_{\lfloor nt\rfloor}-S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}_{t-s}$ too. Now since $\frac{S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}_s$ and $\frac{S_{\lfloor nt\rfloor}-S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}_{t-s}$, and they are ***independent*** for each n, $\left(\frac{S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}}, \frac{S_{\lfloor nt\rfloor}-S_{\lfloor ns\rfloor}}{\sigma\sqrt{n}}\right) \Rightarrow (N_s, N_{t-s})$, where N_s and N_{t-s} are also independent, and have distributions $\mathcal{N}_s$ and $\mathcal{N}_{t-s}$ respectively. Applying Result 1.2 concludes that $(X^n_s, X^n_t - X^n_s) \Rightarrow (N_t, N_{t-s})$.

The independence of $S_{\lfloor ns\rfloor}$ and $S_{\lfloor nt\rfloor} - S_{\lfloor ns\rfloor}$ was key in finding the limiting distribution of $(X^n_s, X^n_t - X^n_s)$. Similarly,

$$\left(X^n_{t_1}, X^n_{t_2} - X^n_{t_1}, \ldots, X^n_{t_k} - X^n_{t_{k-1}}\right) \Rightarrow \left(N_{t_1}, N_{t_2-t_1}, \ldots, N_{t_k-t_{k-1}}\right),$$

where the components of this limiting distribution are independent. Since the map $(x_1, \ldots, x_k) \mapsto (x_1, x_2 + x_1, \ldots, x_k + x_{k-1})$ is continuous, $(X_{t_1}, \ldots, X_{t_k}) \Rightarrow \left(N_{t_1}, N_{t_2-t_1} + N_{t_1}, \ldots, N_{t_k-t_{k-1}} + N_{t_{k-1}}\right)$.

Before proceeding, I think it's worth noting that whilst we have now found limiting finite dimensional distributions, they are not actually that convenient to work with: whilst we can say the $N_{t_i-t_{i-1}} + N_{t_{i-1}}$ have distributions $\mathcal{N}_{t_i}$ (the $N_{t_i-t_{i-1}}$ and $N_{t_{i-1}}$ *are* independent), we can't say the $\mathsf{P}_\mathsf{n}\pi^{-1}_{t_1\ldots t_k} \Rightarrow \mathcal{N}_{t_1} \times \cdots \times \mathcal{N}_{t_k}$, because the N_{t_i} are not independent! Perhaps unsurprisingly now, the independence of the increments makes it more convenient to work with their limiting distributions instead: $\left(X^n_{t_1}, X^n_{t_2} - X^n_{t_1}, \ldots, X^n_{t_k} - X^n_{t_{k-1}}\right) \Rightarrow \mathcal{N}_{t_1} \times \mathcal{N}_{t_2-t_1} \times \cdots \times \mathcal{N}_{t_k-t_{k-1}}$.

[5]this 'hybrid' notation ($\mathcal{N}$ is a measure) is sometimes convenient because you don't have to specify the random variable whose distribution it is, which could introduce questions of independence

Anyway, whatever they look like, we now have measures $\mu_{t_1\dots t_k}$ on C which are the limiting finite dimensional distributions of X^n. Said another way, $\mathsf{P}_n\pi^{-1}_{t_1\dots t_k} \Rightarrow \mu_{t_1\dots t_k}$. Now we can return to the problem of finding a limiting measure for P_n itself..

Question 1.3. *Is there a measure, say* P*, on* C *so that* $\mathsf{P}\pi^{-1}_{t_1,\dots t_k} = \mu_{t_1\dots t_k}$ *and if that's the case, then does* $\mathsf{P}_n \Rightarrow \mathsf{P}$*?*

We'll answer the first part of the question first. Let (μ_n) be sequence of probability measures on C.

Definition 1.2. If $\forall\varepsilon > 0$ there is a compact set $K \subset C$ so that $\forall n \in \mathbf{N}$ $\mu_n K > 1 - \varepsilon$, call (μ_n) *tight*.

Result 1.4. (P_n) *is tight.*

Definition 1.3. If every subsequence of (μ_n) has a further subsequence which is convergent, call (μ_n) *precompact*.

Result 1.5. *Suppose* (μ_n) *is tight. Then* (μ_n) *is precompact.*

So, by taking the subsequence which is just $(1, 2, \dots)$, we know there is some sub(sub)sequence (P_{n_i}) so that $\mathsf{P}_{n_i} \Rightarrow \mathsf{W}$, for some W. Now the projections are continuous, so the $\mathsf{P}_{n_i}\pi^{-1}_{t_1\dots t_k} \Rightarrow \mathsf{W}\pi^{-1}_{t_1\dots t_k}$, and since the $\mathsf{P}_n\pi^{-1}_{t_1\dots t_k} \Rightarrow \mu_{t_1\dots t_k}$ and that subsequences of a convergent sequence must converge to the same limit as the sequence itself, it must be that the $\mathsf{W}\pi^{-1}_{t_1\dots t_k} = \mu_{t_1\dots t_k}$. That is, there *is* a measure, here we called it W, whose finite dimensional distributions are exactly those of the limiting finite dimensional distributions of X^n.

Now let's answer the second part of the question: does $\mathsf{P}_n \Rightarrow \mathsf{W}$? We know that each $\mathsf{P}_{n_{i_j}} \Rightarrow \mathsf{P}^{i_j}$, for some P^{i_j}, the superscript because in principle the particular P^{i_j} depends on the subsubsequence. But then similarly to the previous paragraph, the $\mathsf{P}_{n_{i_j}}\pi^{-1}_{t_1\dots t_k} \Rightarrow \mathsf{P}^{i_j}\pi^{-1}_{t_1\dots t_k}$, so it must be that the $\mathsf{P}^{i_j}\pi^{-1}_{t_1\dots t_k} = \mathsf{W}\pi^{-1}_{t_1\dots t_k}$. But now the P^{i_j} and W agree on finite dimensional sets, so in fact the $\mathsf{P}^{i_j} = \mathsf{W}$, and so every subsubsequence of (P_n) converges weakly to W. This property of (P_n) means that $\mathsf{P}_n \Rightarrow \mathsf{W}$.

W is usually called 'Wiener measure', and if I write W for a random element whose distribution is W (eg W is the identity between C), then (W_t) is called a Wiener process, or a Brownian motion. From our construction of W, it is specified as the measure on C under which the random variables π_t each have respective distributions $\mathcal{N}_t$, and that the $\pi_{t_i} - \pi_{t_{i-1}}$ are independent. This specification is unique because it forces the the finite dimensional distributions $\mathsf{W}\pi^{-1}_{t_1\dots t_k}$ to be what they are. Said another way, a process (B_t) is a Brownian motion (has distribution W when viewed as a random element of C), if each B_t has distribution $\mathcal{N}_t$ and (B_t) has independent increments (the $B_{t_i} - B_{t_{i-1}}$ are independent).

This perspective affords the interpretation that if $B = (B_t)$ is a Brownian motion, then it can be thought of as a random function whose likelihood of being in an interval $[a, b]$ at time t can be read off from $\mathcal{N}_t$, and whose 'displacements' between any times $t_1, \dots, t_k$ have nothing to do with each other. I think it's from this that the object we've constructed takes the name of Robert Brown, a Botanist who in 1828 studied the movement of plant pollen particles in water.

2. Circle Doubling Walk and Brownian Motion

A simple random walk (S_n), where the ϵ_k are iid fair coin tosses, is the process tracking the position after n, say, seconds, of "I'll toss a coin each second, heads I take a step forwards, tails I take a step backwards". From the previous section, (S_n) gives us a random function that converges in distribution to a Brownian motion. From the construction of this random function, moving further along in the limit is sort of making the steps smaller and smaller, whilst tossing the coin more and more frequently.

Now the ϵ_k didn't have to be coin tosses, they just needed to be iid and L^2 (so that they have a variance). So in some sense, convergence to Brownian motion of these types of processes is a bit like the central limit theorem but for functions: because of the random function it affords, the convergence says something about the entire process (S_n). Since there seems to be this 'invariance' — it doesn't matter what the ϵ_k actually are, the random function will always go to a Brownian motion — what if we start with a random walk which doesn't actually seem random at first?

The circle doubling map is the map $T : x \mapsto \{2x\}$, where $x \in [0, 1)$ and $\{x\}$ is the fractional part of x, which is also a member of the circle. So, let us pick a point X on the circle, say uniformly at random, apply T, then depending on whether we're in $\left[0, \frac{1}{2}\right)$ or in $\left[\frac{1}{2}, 1\right)$, take a step forwards or backwards respectively. Take $f_k = \mathbf{1}\left\{T^k X \in \left[0, \frac{1}{2}\right)\right\} - \mathbf{1}\left\{T^k X \in \left[\frac{1}{2}, 1\right)\right\}$, then we'll get the $S_n := \sum^n f_k$, which are tracking the position of this walk, and can give us a random function in exactly the same way as before.

Whilst $f_k \in \{-1, 1\}$, it's not clear whether they might fit into the framework of the last section, ie are they iid? (They are L^2.)

Conjecture. *The f_k are iid coin tosses*[6]*, ie* $\mathbf{P}\{f_k = -1\} = \mathbf{P}\{f_k = 1\} = \frac{1}{2}$.

This would mean that the walk we get is actually a proper random walk (it would be the simple random walk), and that the random function goes to a Brownian motion in distribution! All this coming from the circle doubling system, which itself is deterministic.

If $x \in [0, 1)$, then x almost surely has a unique binary expansion $x = \sum b_k 2^{-k}$, where the $b_k \in \{0, 1\}$. So, almost surely, $X = \sum B_k 2^{-k}$, where the random variables $B_k \in \{0, 1\}$. Then since $T : X \mapsto 2X$, $TX = \sum B_{k+1} 2^{-k-1}$, so T is like a 'left-shift'. Also, since $x \in \left[0, \frac{1}{2}\right)$ iff b_1 in x's binary expansion is 1, f_k can be written as $f_k = \mathbf{1}\{B_{k+1} = 1\} - \mathbf{1}\{B_{k+1} = 0\}$. Since X is uniform, it seems reasonable to expect...

Question 2.1. *Is it that the B_k are iid with* $\mathbf{P}\{B_k = 0\} = \mathbf{P}\{B_k = 1\} = \frac{1}{2}$*?*

If this is indeed true, then it would prove the conjecture, since $f_k = 1$ iff $B_{k+1} = 1$.

[6] here, the $\mathbf{P}$ is the measure of the probability space on which X is defined