THE WIENER MEASURE AND DONSKER'S INVARIANCE PRINCIPLE

ETHAN SCHONDORF

ABSTRACT. We give the construction of the Wiener measure (Brownian motion) and prove Donsker's invariance principle, which shows that a simple random walk converges to a Wiener process as time and space intervals shrink. To build up the mechanisms for this construction and proof, we first discuss the weak convergence and tightness of measures, both on a general metric space and on $C[0,1]$, the space of continuous real functions on the closed unit interval. Then we prove the central limit theorem. Finally, we define the Wiener process and discuss a few of its basic properties.

CONTENTS

1. INTRODUCTION

The Wiener process, also called Brownian motion,[1] is one of the most studied stochastic processes. Originally described by botanist Robert Brown in 1827 when he observed the motion of a piece of pollen in a glass of water, it was not until the 1920s that the process was studied mathematically by Norbert Wiener. He gave

Date: November 1, 2019.

[1]While in the physical sciences these two terms have different meanings, here, as in financial disciplines, we use these two terms interchangeably.

the first construction of the Wiener process by defining and proving the existence of a measure that induces the Wiener process on continuous functions and gives the distribution of paths. From this origin the importance of the Wiener process reveals itself. Its ability to describe the random motion of a particle has made it an incredibly useful process for scientific modeling and in turn has motivated much mathematical study. Because this process is used to describe random motion, we intuitively think to look at it as the continuous analogue of the discrete process of a random walk — a process where a walker takes random discrete steps forwards and backwards at discrete time intervals. In 1952, Monroe D. Donsker formally stated what is now known as Donsker's invariance principle, which formalized this idea of the Wiener process as a limit. Donsker's invariance principle shows that the Wiener process is a universal scaling limit of a large class of random walks. This means that the exact distributions of the steps of the random walk do not matter when we take the limit as time and step distance shrink to zero.

In this paper, we aim to construct the Wiener measure and prove Donsker's invariance principle. We start in Section 2, where we define the notions of weak convergence, relative compactness, and tightness. We then go on to prove a number of important convergence theorems, both in a general metric space and in the space $C[0,1]$ — the space of continuous functions on the closed unit interval. These theorems provide the main argument for the construction of the Wiener measure and the proof of Donsker's invariance principle. In Section 3, we prove the classical central limit theorem through Lévy's continuity theorem. Then, in Section 4, we define both a random walk and the Wiener process before examining and proving a few fundamental properties of the latter process. In Section 5, we give a construction of the Wiener measure. Finally, in Section 6, we state and prove Donsker's invariance principle through a tightness argument. This paper assumes the knowledge of basic measure theory and basic probability theory. Our approach here largely follows that of Billingsley [1] and Durrett [2].

2. Convergence of Measures

2.1. **Weak Convergence**

First we will define two notions of convergence: weak convergence of measures and convergence in probability of random variables. These are important notions in measure and probability theory and their definitions are required for the results presented in this paper.

Unless otherwise specified, we assume we are in the measurable space $(S, \mathcal{S})$ where S is a general metric space and $\mathcal{S}$ is the σ-field formed by the Borel sets. We use $\mathbb{P}$ to denote a probability measure and $\mathbb{E}$ to denote the expectation.

Definition 2.1.1. We say a sequence of measures (μ_n) **weakly converges** to the measure μ if for all bounded, continuous functions f,

$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu. \tag{1}$$

We denote this as $\mu_n \Rightarrow \mu$. When we have a sequence of measures with multiple indexes, say n and t, we use the symbol $\Rightarrow_n$ to denote that n is the index we send to infinity.

Consequently, we say that a sequence of random variables (X_n) **converges weakly** or **converges in distribution** to X if the respective cumulative distribution functions, $F_n(x) = \mathbb{P}(X \geq x)$, converges weakly. Equivalently, we can write that (X_n) converges weakly to X if $\mathbb{E}^* f(X_n) \to \mathbb{E} f(X)$ for all closed and bounded functions f, were $\mathbb{E}^*$ denotes the outer expectation (the expectation of the smallest measurable function g that dominates $f(X_n)$). We denote this by $X_n \Rightarrow X$.

Note that we say the sequence probability measures $(\mathbb{P}_n)$ converges weakly to the probability measure $\mathbb{P}$ if (1) holds with $\mathbb{P}_n$ and $\mathbb{P}$ in place of μ_n and μ respectively.

We now define one more term that helps us prove weak convergence.

Definition 2.1.2. We call a set $A \in \mathcal{S}$ a **continuity set of the measure $\boldsymbol{\mu}$** if $\mu(\partial A) = 0$.

Next, we turn to a useful theorem that gives us conditions equivalent to weak convergence that often are easier to show than weak convergence itself.

Theorem 2.1.3. *(The Portmanteau Theorem) Let μ and μ_n be measures on $(S, \mathcal{S})$. The following conditions are equivalent:*

(i) $\mu_n \Rightarrow \mu$.
(ii) $\int f\mu_n(dx) \to \int f\mu(dx)$ *for all bounded, uniformly continuous function* $f : S \to \mathbb{R}$.
(iii) $\limsup_n \mu_n F \leq \mu F$ *for all closed sets* F.
(iv) $\limsup_n \mu_n G \geq \mu G$ *for all open sets* G.
(v) $\mu_n A \to \mu A$ *for all sets* A *that are continuity sets of* μ.

Proof. By the definition of weak convergence (i) implies (ii). To get (iii) from (ii) let us define the function $f(x) = (1 - \epsilon^{-1}\rho(x, F)) \vee 0$ for any positive ϵ. Note that this function is bounded and it is uniformly continuous since $|f(x) - f(y)| \leq \epsilon^{-1}\rho(x, y)$ for any x and y in S. Let us denote by $F^\epsilon := \{x : \rho(x, F) \leq \epsilon\}$ Then, we see that

$$1_F(x) \leq f(x) \leq 1_{F^\epsilon},$$

where 1_F is the indicator function of F. Now (ii) implies,

$$\limsup_n \mu_n F \leq \limsup_n \mu_n f = \mu f \leq \mu F^\epsilon.$$

Letting $\epsilon \downarrow 0$ gives (iii).

(iii) implies (iv) by complementation.

(iii) and (iv) imply

$$\mu\overline{A} \geq \limsup_n \mu_n \overline{A} \geq \limsup_n \mu_n A \geq \liminf_n \mu_n A \geq \liminf_n \mu_n A^\circ \geq \mu A^\circ.$$

If A is a continuity set of μ then $\mu\overline{A} = \mu A^\circ$ and (v) follows.

By linearity, assume that the bounded f satisfies $0 \leq f \leq 1$. Define the set $A_t := \{x : f(x) > t\}$. Then,

$$\int_S f\mu_n(dx) = \int_0^1 \mu_n A_t dt,$$

and similar for μ. Since f is continuous then $\partial A_t = \{x : f(x) = t\}$. These sets are μ-continuity sets except for at countably many t. So, by the bounded convergence theorem,

$$\int_S f\mu_n(dx) = \int_0^1 (\mu_n A_t) dt \to \int_0^1 (\mu A_t) dt = \int_S f\mu_n(dx),$$

and thus (v) implies (i).

□

Now we look at three conditions for weak convergence.

Theorem 2.1.4. *For some measure μ and collection of measures (μ_n), $\mu_n \Rightarrow \mu$ if and only if each subsequence (μ_{n_i}) contains a further subsequence $(\mu_{n_{i_k}})$ that converges weakly to μ as $k \to \infty$.*

Proof. Necessity (the forward direction) is easy by the definition of weak convergence but also not particularly useful. For sufficiency, suppose for contradiction that $\mu_n \not\Rightarrow \mu$. Then, for some bounded and continuous f, $\int_S f(x)\mu_n dx \not\to \int_S f(x)\mu dx$. But then for some positive ϵ and subsequence (μ_{n_i}), $\left\|\int_S f(x)\mu_{n_i}(dx) - \int_S f(x)\mu_n(dx)\right\| > \epsilon$ and thus no further subsequence can converge. This is a contradiction and so $\mu_n \Rightarrow \mu$. □

Theorem 2.1.5. *Suppose the pair (X_n, Y_n) are random elements of $S \times S$. Let $\rho(x, y) = ||x|| = \sup_t |x(t) - y(t)|$ be the supremum norm. Then, if $X_n \Rightarrow X$ and $\rho(X_n, Y_n) \Rightarrow 0$ then $Y_n \Rightarrow X$.*

Proof. Define the set $F^\epsilon := \{x : \rho(x, F) \leq \epsilon\}$ for all ϵ. Then we see that

$$\mathbb{P}(Y_n \in F) \leq \mathbb{P}(\rho(X_n, Y_n) \geq \epsilon) + \mathbb{P}(X \in F^\epsilon).$$

Since F^ϵ is closed, we can apply (iii) of the Portmanteau theorem to give

$$\limsup_n \mathbb{P}(Y_n \in F) \leq \limsup_n \mathbb{P}(X_n \in F^\epsilon) \leq \mathbb{P}(X \in F^\epsilon).$$

If F is closed then $F\epsilon \downarrow F$ as $\epsilon \downarrow 0$. Applying (iii) of the Portmanteau theorem again gives $Y_n \Rightarrow X$ which completes the proof. □

Next we move on to a mapping theorem. We use this theorem primarily to show that weakly convergent probability measures, when restricted to finite dimensions, are still weakly convergent.

Theorem 2.1.6. *(The Mapping Theorem) Let h be a map from $S \to S'$ with metric ρ' and Borel σ-field $\mathcal{S}'$. Furthermore, suppose that h is measurable $\mathcal{S}/\mathcal{S}'$. Finally, let D_h be the set of the discontinuities of h. If $\mathbb{P}_n \Rightarrow \mathbb{P}$ and $\mathbb{P}D_h = 0$, then $\mathbb{P}_n h^{-1} \Rightarrow \mathbb{P}h^{-1}$.*

Proof. Suppose that $x \in \overline{h^{-1}F}$. Then there exists a sequence $x_n \to x$ where $hx_n \in F$ for each n. However, if $x \in D_h^c$ then hx is in $\overline{F}$. Therefore,

$$D_h^c \cap \overline{h^{-1}F} \subset h^{-1}\overline{F}.$$

If F is a closed set in S', then it follows from the hypothesis that

$$\begin{aligned}\limsup_n \mathbb{P}_n(h^{-1}F) &\leq \limsup_n \mathbb{P}_n(\overline{h^{-1}F}) \leq \mathbb{P}(\overline{h^{-1}F}) \\ &= \mathbb{P}\left(D_h^c \cap \overline{h^{-1}F}\right) = \mathbb{P}(h^{-1}\overline{F}) = \mathbb{P}(h^{-1}F).\end{aligned}$$

Therefore condition (iii) of the Portmanteau theorem holds and so $\mathbb{P}h^{-1} \Rightarrow \mathbb{P}h^{-1}$.

□

Theorem 2.1.7. *$X_n \Rightarrow X$ if and only if $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ for every bounded, continuous function g.*

Proof. Suppose $X_n \Rightarrow X$. Let Y_n have the same distribution as X_n but converge to Y almost everywhere. Since g is continuous we have $g(Y_n) \to g(Y)$. The bounded convergence theorem then implies that

$$\mathbb{E}g(X_n) = \mathbb{E}g(Y_n) \to \mathbb{E}g(Y) = \mathbb{E}g(X).$$

This proves the forward direction.

Now for the converse. Define

$$g_{x,\epsilon} = \begin{cases} 1 & y \leq x \\ 0 & y \geq x+\epsilon \\ \frac{y-x}{\epsilon} & x \leq y \leq x+\epsilon \ . \end{cases}$$

By construction this g is continuous, so

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}(X_n) \leq & \limsup_{n\to\infty} \mathbb{E}g_{x,\epsilon}(X_n) \\ & = \mathbb{E}g_{x,\epsilon}(X) \\ & \leq \mathbb{P}(X \leq x+\epsilon). \end{aligned}$$

Letting $\epsilon \to \infty$ gives

$$\limsup_{n\to\infty} \mathbb{P}(X_n) \leq \mathbb{P}(X \leq x).$$

Similarly, using the lower limit and the function $g_{x-\epsilon,\epsilon}$ we can obtain the inequality

$$\liminf_{n\to\infty} \mathbb{P}(X_n) \geq \mathbb{P}(X \leq x).$$

Therefore we see that $\lim_{n\to\infty} \mathbb{P}(X_n) = \mathbb{P}(X \leq x)$ and therefore $X_n \Rightarrow X$. □

2.2. **Tightness**

Here we define a concept of tightness for collections of measures and random variables. Intuitively this ensures that a collection of measures does not have mass that escapes to infinity. Tightness is often used to prove weak convergence.

Definition 2.2.1. Let $(S, \mathcal{S})$ be a measurable space. A collection of measures $\{\mu_i\}$ is **tight** if for all $\epsilon > 0$ there exists a compact set $K \in S$ such that $\sup_i \mu_i(K^c) < \epsilon$ for all i.

If a tight collection contains only one measure we say that the measure is tight.

For a collection of probability measures $\{\mathbb{P}_i\}$, we equivalently say that the collection is tight if for all $\epsilon > 0$ there exists a compact set $K \in S$ such that $\mathbb{P}_i(K) > 1-\epsilon$ for all i.

We say a random variable X is tight if for all $\epsilon > 0$ there is an M_ϵ such that

$$\mathbb{P}(||X|| > M_\epsilon) < \epsilon.$$

Now we introduce the concept of relative compactness, which helps us connect tightness and weak convergence.

Definition 2.2.2. A set is **relatively compact** if its closure is compact.

Let Π be a family of probability measures on $(S, \mathcal{S})$. We call Π relatively compact if every sequence of elements of Π contains a weakly convergent subsequence. Explicitly this means that if Π is relatively compact, then there exists a subsequence $(\mathbb{P}_{n_i}) \in \Pi$ and a probability measure Q, which need not be contained in $(S, \mathcal{S})$, such that $\mathbb{P}_{n_i} \Rightarrow_i Q$.

Theorem 2.2.3. *(Prokhorov's Theorem) If Π is tight, then it is relatively compact.*

The converse of this statement is also true, however we will not need that result in this paper. Before we prove the theorem we quickly state and prove a corollary that more explicitly shows an application of Prokhorov's theorem.

Corollary 2.2.4. *If $(\mathbb{P}_n)$ is tight and each weakly convergent subsequence converges to $\mathbb{P}$, then the entire sequence converges weakly to $\mathbb{P}$.*

Proof. By Prokhorov's theorem and the definition of relative compactness, each subsequence contains further subsequences converging weakly to some limit. By our hypothesis, this limit must be $\mathbb{P}$. Applying Theorem 2.1.4, we see that the sequence $(\mathbb{P}_n)$ converges weakly to $\mathbb{P}$. $\square$

Proof of Prokhorov's Theorem. In order to prove relative compactness we want to find a subsequence $(\mathbb{P}_{n_i})$ and a probability measure $\mathbb{P}$ such that $\mathbb{P}_{n_i} \Rightarrow_i \mathbb{P}$. To show this we use a diagonal argument.

Since $(\mathbb{P}_n)$ is tight we can choose compact sets such that $K_1 \subset K_2 \subset \dots$ and $\mathbb{P}_n(K_i) > 1-i^{-1}$. Then we can define the set $K := \bigcup_i K_i$ which is separable. Every open cover of each subset of K has a countable subcover (this is a fundamental theorem about separable spaces and is proven in [1, appx. M3]). So, there exists a countable class, $\mathcal{A}$, of open sets with the property that if x is in both K and some open set G, then $x \in A \subset \overline{A} \subset G$ for some $A \in \mathcal{A}$.

Now let us construct $\mathcal{H}$ such that it consists of the empty set and finite unions of sets of the form $\overline{A} \cap K_i$ for $A \in \mathcal{A}$ and $i \geq 1$. Consider the countable class $\mathcal{H} = (H_j)$. For $(\mathbb{P}_n)$ there exists a subsequence $(\mathbb{P}_{n_1})$ such that $\mathbb{P}_{n_1}$ converges as $n \to \infty$. We can continue this process and obtain further convergent subsequences with collection of indexes $(n_{1,i}) \supset (n_{2,i}) \supset \dots$. Then, we see that the subsequence $\mathbb{P}_{n_{i,i}}$ converges and we define

$$\alpha(H) := \lim_{i\to\infty} \mathbb{P}_{n_{i,i}}(H),$$

for $H \in \mathcal{H}$. For open sets $G \subset S$ and an arbitrary $M \subset S$, we define the two functions

$$\beta(G) := \sup_{H \subset G} \alpha(H) \qquad \gamma(M) := \inf_{G \supset M} \beta(G).$$

Our objective is to construct a probability measure on $(S, \mathcal{S})$ such that $\mathbb{P}G = \beta G$. If we can construct such a measure, then for $H \subset G$ we obtain

$$\alpha(H) = \lim_{i\to\infty} \mathbb{P}_{n_{i,i}}(H) \leq \liminf_{i\to\infty} \mathbb{P}_{n_{i,i}}(G).$$

Then, we see that $\mathbb{P}G \leq \liminf_{i\to\infty} \mathbb{P}_{n_{i,i}}(G)$, which implies our desired result, $\mathbb{P}_{n_{i,i}} \Rightarrow_i \mathbb{P}$.

We construct the measure in seven steps:

Step 1: If $F \subset G$ is closed, G is open, and $F \subset H$ for some $H \in \mathcal{H}$, then $F \subset H_0 \subset G$ for some $H_0 \in \mathcal{H}$.

If $F \subset H$, then $F \subset K_{i_0}$ for some i_0. Since it is bounded, the closed set F is compact. For each $x \in F$, choose an $A_x \in \mathcal{A}$ where $x \in A_x \in \overline{A_x} \subset G$. The sets A_x cover the compact F, so there exists a finite subcover, $H_0 = \bigcup_{j=1}^{k}(\overline{A_{x_j} \cap K_{i_0}})$, which has the desired properties.

Step 2: Show that β is finitely subadditive on open sets.

Suppose that $H \subset G_1 \cup G_2$ where $H \in \mathcal{H}$ and G_1, G_2 are open sets. Now define the sets

$$F_1 := \{x \in H : \rho(x, G_1^c) \geq \rho(x, G_2^c)\}$$
$$F_2 := \{x \in H : \rho(x, G_2^c) \geq \rho(x, G_1^c)\}.$$

Note that $H = F_1 \cup F_2$, $F_1 \subset G_1$ and $F_2 \subset G_2$. By the first step and since $F_i \subset H$, we have $F_i \subset H_i \subset G_i$ for some $H_i \in \mathcal{H}$. Now we see that $\alpha(H)$ has the following three properties:

(i) $\alpha(H_1) \leq \alpha(H_2)$ if $H_1 \subset H_2$.
(ii) $\alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2)$ where the equality holds if $H_1 \cap H_2 = \emptyset$.
(iii) $\alpha(H) \leq \alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2) \leq \beta(G_1) + \beta(G_2)$.

From these properties we see that

$$\beta(G_1 \cup G_2) = \sup_{H \subset G_1 \cup G_2} \alpha(H) \leq \beta(G_1) + \beta(G_2),$$

which proves finite subadditivity.

Step 3: Show that β is countably subadditive on open sets.

If $H \subset \bigcup_n G_n$, then since H is compact, $H \subset \bigcup_{n \leq n_0} G_n$ for some n_0. Finite subadditivity implies that

$$\alpha(H) \leq \sum_{n \leq n_0} \beta(G_n) \leq \sum_n \beta(G_n).$$

Taking the supremum over H contained in $\bigcup_n G_n$ gives $\beta(\bigcup_n G_n) \leq \sum_n \beta(G_n)$, which proves countable subadditivity.

Step 4: Show γ is an outer measure.

From the definition of γ the function is monotone and $\gamma(\emptyset) = 0$. It is left to show that γ is countably subadditive. Suppose we have a positive ϵ and an arbitrary $M_n \subset S$. Then choose G_n such that $M_n \subset G_n$ and $\beta(G_n) < \gamma(M_n) + \epsilon/2^n$. Applying step 3,

$$\gamma(\bigcup_n M_n) \leq \beta(\bigcup_n G_n) \leq \sum_n \beta(G_n) \leq \sum_n \gamma(M_n) + \epsilon.$$

Letting ϵ go to zero we get $\gamma(\bigcup_n M_n) \leq \sum_n \gamma(M_n)$ and therefore we have shown countable subadditivity. Thus γ is an outer measure.

Step 5: Show $\beta(G) \geq \gamma(F \cap G) + \gamma(F^c \cap G)$ for a closed set F, and open set G.

Let us choose $H_3, H_4 \in \mathcal{H}$ such that

$$H_3 \subset F^c \cap G \quad \text{and} \quad \alpha(H_3) > \beta(F^c \cap G) - \epsilon$$
$$H_4 \subset H_3^c \cap G \quad \text{and} \quad \alpha(H_4) > \beta(H_3^c \cap G) - \epsilon.$$

Note that by definition H_3 and H_4 are disjoint and contained in G. Therefore,

$$\beta(G) \geq \alpha(H_3 \cup H_4) = \alpha(H_3) + \alpha(H_4)$$
$$> \beta(F^c \cap G) + \beta(H_3^c \cap G) - 2\epsilon \geq \gamma(F^c \cap G) + \gamma(F \cap G) - 2\epsilon.$$

Letting ϵ go to zero gives us our desired result.

Step 6: If $F \subset S$ is closed then F is in the class $\mathcal{M}$ of γ measurable sets.

Suppose F is closed, G is open and we have a set L where $G \supset L$. Taking the infimum over all G by step 5 we find

$$\gamma(L) \geq \gamma(F \cap L) + \gamma(F^c \cap L).$$

This verifies that F is γ-measurable.

Step 7: Show that $S \subset \mathcal{M}$ and the restriction $\mathbb{P}$ of γ to $\mathcal{M}$ is a probability measure satisfying $\mathbb{P}(G) = \gamma(G)a = \beta(G)$ for all open sets $G \subset S$.

Since $\mathscr{M}$ is a σ-algebra every closed set is contained in $\mathscr{M}$ and thus $S \subset \mathcal{M}$. Now to show that $\mathbb{P}$ is a probability measure, note that each K_i has a finite covering by $\mathcal{A}$-sets and so $K_i \in \mathcal{H}$. Therefore,

$$1 \geq (\mathbb{P}(S) = \beta(S) \geq \sup_i \alpha(K_i) \geq \sup_i (1 - i^{-1}) = 1.$$

This shows $\mathbb{P}$ is a probability measure with the desired condition. Since we have constructed this probability measure we have proven Prokhorov's theorem. □

2.3. Weak Convergence in the Space C

The rest of this section is devoted to proving weak convergence theorems in the measurable space $(C, \mathcal{C})$ where $C = C[0,1]$ is the space of continuous functions on the closed unit interval. We use these theorems to prove the existence of the Wiener measure and Donsker's invariance principle.

Theorem 2.3.1. *Let $\mathbb{P}_n$ and $\mathbb{P}$ be probability measures on $(C, \mathcal{C})$. If the finite dimensional distributions $\mathbb{P}_n$ converge weakly to those of $\mathbb{P}$ and if $(\mathbb{P}_n)$ is tight then $\mathbb{P}_n \Rightarrow \mathbb{P}$.*

Proof. From Prokhorov's theorem, tightness implies relative compactness and thus each subsequence contains a further subsequence weakly converging to some Q. Let π_k be the projection function $\pi_k : \mathbb{R}^\infty \to \mathbb{R}^k$ where for $\mathbf{x} \in R^\infty$, $\pi_k(\mathbf{x}) = (x_1, \ldots, x_k)$. The mapping theorem then implies that $\mathbb{P}_{n_{k_i}} \pi^{-1}_{t_1,\ldots,t_k} \Rightarrow \mathbb{Q}\pi^{-1}_{t_1,\ldots,t_k}$. Since the finite dimensional distributions converge weakly, $\mathbb{P}_n \pi^{-1}_{t_1,\ldots,t_k} \Rightarrow \mathbb{P}\pi^{-1}_{t_1,\ldots,t_k}$. So, by our hypothesis, $\mathbb{P}\pi^{-1}_{t_1,\ldots,t_k} = Q\pi^{-1}_{t_1,\ldots,t_k}$. We define a subclass of $\mathcal{A} \subset \mathcal{S}$ to be *separable* if any two probability measures with $\mathbb{P}(A) = \mathbb{P}'(A)$ for all $A \in \mathcal{A}$ are identical. It can then be shown that the class of finite-dimensional sets $\mathcal{C}_f$ forms a separating class (a proof of this can be found in [1, p. 12]) and therefore $P = Q$. So each subsequence contains a further subsequence converging weakly to the same $\mathbb{P}$. It then follows from Theorem 2.1.4 that $\mathbb{P}_n \Rightarrow \mathbb{P}$. □

Now we define a function that quantifies the uniform continuity of a function.

Definition 2.3.2. A **modulus of continuity** of an arbitrary function x is defined by

$$w(x, \delta) := \sup_{|s-t| \leq \delta} |x(s) - x(t)|,$$

where $\delta \geq 0$.

Next we prove a theorem that gives us a necessary and sufficient condition for relative compactness.

Theorem 2.3.3. *(Arzelà-Ascoli Theorem) The set A is relatively compact if and only if both*

$$\sup_{x \in A} |x(0)| \leq \infty, \tag{2}$$

and

$$\lim_{\delta \to 0} \sup_{x \in A} w(x, \delta) = 0. \tag{3}$$

Proof. First, suppose that A is relatively compact. Then, $\overline{A}$ is compact and thus (2) follows. Now for fixed x, the function $w(x,\delta)$ monotonically converges to zero as $\delta \downarrow 0$. Furthermore, for each δ, the function $w(x,\delta)$ is continuous in x. Therefore, the convergence is uniform over $x \in K$, where K is compact. Setting $K = \overline{A}$, the equation (3) follows.

Now assume that both (2) and (3) hold. Choose k large enough such that $\sup_{x\in A} w(x,k^{-1})$ is finite. Since

$$|x(t)| \leq |x(0)| + \sum_{i=1}^{k} \left| x\left(\frac{it}{k}\right) - x\left(\frac{(i-1)t}{k}\right)\right|,$$

we can then define

$$\alpha := ||x|| < \infty. \tag{4}$$

Our goal is to use (3) and (4) to prove that A is totally bounded. Since the space C is complete, it then follows that $\overline{A}$ is compact and so A is relatively compact.

So, given ϵ, we choose a finite ϵ-net H in the interval $[-\alpha,\alpha]$ on the line and choose a k large enough that $w(x,1/k) < \epsilon$ for all x in A. Now, define B to be the finite set consisting of polygonal functions in C that are linear on the interval $I_{i,k} = \left[\frac{i-1}{k}, \frac{i}{k}\right]$ for $1 \leq i \leq k$ and let B take values of $\alpha_{j_{i-1}}$ and α_{j_i} at the endpoints. If $x \in A$, then $\left|x\left(\frac{i}{k}\right)\right| < \alpha$. Therefore there exists a $y \in B$ such that $\left|x\left(\frac{i}{k}\right) - y\left(\frac{i}{k}\right)\right| < \epsilon$ for $i = 0, 1, \ldots, k$. Now both $y\left(\frac{i}{k}\right)$ and $y\left(\frac{i-1}{k}\right)$ are within 2ϵ of $x(t)$ for $t \in I_{k_i}$. Since $y(t)$ is a convex combination of $y\left(\frac{i-1}{k}\right)$ and $y\left(\frac{i}{k}\right)$, it too is in the 2ϵ-net and thus $\rho(x,y) < 2\epsilon$. Since this holds for all $t \in I_{k_i}$ and all intervals, B is a finite 2ϵ-net for A. Therefore, A is totally bounded and our result follows. □

We can subsequently use the Arzelà-Ascoli Theorem to obtain a condition for tightness.

Theorem 2.3.4. *The sequence $(\mathbb{P}_n)$ is tight if and only if*

(i) For a positive η, there exists an a and N such that

$$\mathbb{P}_n\left[x : |x(0)| \geq a\right] \leq \eta, \tag{5}$$

for $n \geq N$.

(ii) For each positive ϵ and η, there exists a δ where $0 < \delta < 1$ and an N such that

$$\mathbb{P}_n\left[x : w(x,\delta) \geq \epsilon\right] \leq \eta, \tag{6}$$

for $n \geq N$. This can be written in the more compact form

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}_n\left[x : w(x,\delta) \geq \epsilon\right] = 0. \tag{7}$$

Proof. Suppose $(\mathbb{P}_n)$ is tight. Let $\eta > 0$ and choose a compact K such that $\mathbb{P}_n K > 1-\eta$ for all n. By the Arzelà-Ascoli theorem, we have $K \subset \{x : |x(0)| \leq a\}$ for large enough a and $K \subset [x : w(x,\delta) \leq \epsilon]$ for small enough δ. So (i) and (ii) hold with $n_0 := 1$ in either case.

Now note that since C is separable, for each k, there exists a sequence of open $1/k$-balls, $A_{k1}, A_{k2}, \ldots$ that cover C. If we choose n_k large enough, then for $\epsilon > 0$, there exists a finite covering $\mathbb{P}\left[\bigcup_{n\leq n_k} A_{kn}\right] \geq 1-\epsilon/2^k$. Since C is complete, the totally bounded set $\bigcap_{k\geq 1}\bigcup_{n\leq n_k} A_{kn}$ has compact closure K and $\mathbb{P}K > 1-\epsilon$. Thus, every probability measure $\mathbb{P}_n$ is tight on $(C,\mathcal{C})$ for finite n. So, by the necessity statement, for this single measure and for any η there exists an $a_{n,\eta}$ such that

$\mathbb{P}_n[x : |x(0)| \geq a_{n,\eta}] \leq \eta$ and for any given ϵ and η, there exists a $\delta_{n,\epsilon,\eta}$ such that $\mathbb{P}_n[x : w(x, \delta_{n,\epsilon,\eta}) \geq \epsilon] \leq \eta$. Therefore, if $(\mathbb{P}_n)$ satisfies (i) and (ii) we can set $n_0 = 1$ when proving sufficiency. Fix some positive small η and choose a such that if $B := \{x : |x(0)| \leq a\}$, then $\mathbb{P}_n B \geq 1 - \eta$ for all n. Then, choose δ_k such that if $B_k := \{x : w(x, \delta_k) \leq 1/k\}$ then $\mathbb{P}_n B_k \geq 1 - \eta/2^k$ for all n. Let $A = B \cap (\bigcap_k B_k)$ and set $K = \overline{A}$. Then we see that $\mathbb{P}_n K \geq 1 - 2\eta$ for all n. Note that A satisfies the conditions for sufficiency in the Arzelà-Ascoli theorem. Therefore, A is relatively compact and K is compact. Thus $(\mathbb{P}_n)$ is tight. $\square$

Next we prove an extension to the conditions for tightness in Theorem 2.3.4.

Theorem 2.3.5. *Suppose* $0 = t_0 < t_1 < \ldots < t_k = 1$ *and*

$$\min_{1<i<k} (t_i - t_{i-1}) \geq \delta. \tag{8}$$

If we define $I_i := [t_{i-1}, t_i]$, *then for arbitrary* x,

$$w(x, \delta) \leq 3 \max_{1\leq i\leq k} \sup_{s\in I_i} |x(s) - x(t_{i-1})|, \tag{9}$$

and, for arbitrary $\mathbb{P}$,

$$\mathbb{P}[x : w(x, \delta) \geq 3\epsilon] \leq \sum_{i=1}^{k} \mathbb{P}[x : \sup_{s\in I_i} |x(s) - x(t_{i-1})| \geq \epsilon]. \tag{10}$$

Proof. Let m be defined to be the maximum in (9). We only need to look at the differences of points s and t that are no more than δ apart since $w(x, \delta) := \sup_{|s-t|<\delta} |x(s) - x(t)|$. Since each time interval I_i is at least δ wide, we only need to look at s and t in the same I_i and in adjacent I_i. In the former case

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t_{i-1}) - x(t)| \leq 2m.$$

In the latter case

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t_{i-1}) - x(t_i) + |x(t_i) - x(t)| \leq 3m.$$

Therefore,

$$w(x, \delta) \leq 3m,$$

and (9) follows. Lastly, (10) follows by the subadditivity of $\mathbb{P}$. $\square$

Finally, we come to a condition for weak convergence for finite weakly converging random variables that we later use to prove Donsker's invariance principle.

Theorem 2.3.6. *If both*

$$(X^n_{t_1}, \ldots, X^n_{t_k}) \Rightarrow_n (X_{t_1}, \ldots, X_{t_k}), \tag{11}$$

for all $t_1, \ldots, t_k$, *and*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left[w\left(X^n, \delta\right) \geq \epsilon\right] = 0, \tag{12}$$

for all positive ϵ, *then* $X^n \Rightarrow_n X$.

Proof. Let $\mathbb{P}$ and $\mathbb{P}_n$ be the distributions of X and X_n respectively. We can reformulate (12) as $\mathbb{P}_n \pi^{-1}_{t_1,\ldots,t_k} \Rightarrow_n \mathbb{P}\pi^{-1}_{t_1,\ldots,t_k}$. To prove $X^n \Rightarrow_n X$ we instead equivalently prove that $\mathbb{P}_n \Rightarrow \mathbb{P}$. To show this we only need to prove that $\mathbb{P}_n$ is tight and then the result follows by Theorem 2.3.1. Since we have the weak convergence of the finite distributions, $X^n_0 \Rightarrow_n X_0$ implies $(\mathbb{P}_n \pi_0^{-1})$ is tight. This implies the first condition of Theorem 2.3.4. We also see that (12) can be written as (7). Since both conditions of Theorem 2.3.4 are satisfied, $(\mathbb{P}_n)$ is tight, which proves the desired result. $\square$

3. Central Limit Theorem

The objective of this section is to prove the Lindeberg-Lévy central limit theorem. In order to do this we must define a characteristic function, which is a function that completely defines a random variable's distribution. We then prove a continuity theorem about characteristic functions that allows us to prove the central limit theorem.

Definition 3.1.1. A **characteristic function** φ of a random variable X is defined to be

$$\varphi(t) = \mathbb{E}[e^{itX}].$$

The basis for our proof of the central limit theorem is the following version of a continuity theorem of characteristic functions due to Lévy.

Theorem 3.1.2. *(Lévy Continuity Theorem) Let $\mu_n, 1 \le n \le \infty$ be probability measures with characteristic functions φ_n respectively.*

(i) If $\mu_n \Rightarrow \mu$, then, for all t, $\varphi_n(t)$ converges to $\varphi(t)$.
(ii) If $\varphi_n(t)$ converges pointwise to $\varphi(t)$, which is continuous at 0, then the associated sequence of measures (μ_n) is tight and converges weakly to μ, which has the characteristic function φ.

Proof. For (i), since e^{itx} is bounded and $\mu_n \Rightarrow \mu$, it follows from Lemma 2.1.7 that $\varphi_n(t) \to \varphi(t)$ for all t.

For (ii) we first want to show tightness. To do this we start with a seemingly unrelated integral in an attempt to bound μ_n by a sequence converging to 0. So, consider

$$\int_{-u}^{u} (1 - e^{itx})dt = 2u - \int -u^u (\cos tx + i \sin tx)dt = 2u - \frac{2 \sin ux}{x}.$$

Dividing by u, integrating over $\mu_n(dx)$ on both sides, and using Fubini's theorem on the left hand side we see that

$$\frac{1}{u}\int_{-u}^{u} (1 - \varphi_n(t))dt = 2 \int \left(1 - \frac{\sin ux}{ux}\right) dt.$$

Now to bound the right hand side, first note that $|x| \ge |\sin x|$ for all x and so $1 - (\sin ux)/ux \ge 0$. Note that $|\sin ux| \le 1$ outside of $(-2/u, 2/u)$. Discarding this interval we see that the right hand side becomes

$$2 \int \left(1 - \frac{\sin ux}{ux}\right) dt \ge 2 \int_{|x| \ge \frac{2}{u}} \left(1 - \frac{1}{|ux|}\right) \mu_n(dx) \ge \mu_n \left(\left\{x : |x| \ge \frac{2}{u}\right)\right\}.$$

Note that as $t \to 0$, $\varphi t \to 1$, so

$$\frac{1}{u}\int_{-u}^{u} (1 - \varphi_n(t))dt \to 0$$

as $u \to 0$. Therefore, we can choose u such that the integral is less than ϵ. Since $\varphi_n(t) \to \varphi(t)$ for all t, it follows from the bounded convergence theorem that

$$2\epsilon \ge \frac{1}{u}\int_{-u}^{u} (1 - \varphi_n(t))dt \ge \mu_n \left(\left\{x : x \ge \frac{2}{u}\right)\right\},$$

for $n \geq N$. So μ_n is tight.

So, since the sequence of measures is tight, every subsequence has a further subsequence that converges to the same μ. This then implies that $\mu_n \Rightarrow \mu$. If not, there would then exist some subsequence (μ_{n_k}) such that for some $\delta > 0$ and some bounded, continuous f, $\int f d\mu_{n_k} - \int f d\mu > \delta$. There would then be no further subsequence, which is a contradiction. By (1), μ has characteristic function $\varphi(t)$. This completes the proof.

□

Before we move on to the central limit theorem we state without proof a proposition concerning the characteristic functions of random variables with finite second moments. A proof can be found in [2, p. 99-100].

Proposition 3.1.3. *If $\mathbb{E}|X|^2 < \infty$, then*

$$\varphi(t) = 1 + dt\mathbb{E}X - \frac{t^2\mathbb{E}(X^2)}{2} + \mathbf{o}(t^2)$$

Theorem 3.1.4. *(Central Limit Theorem) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with $\mathbb{E}X_i = \mu$ and Var$(X_i) = \sigma^2 \in (0, \infty)$. If $S_n = X_1 + X_2 + \ldots$, then*

$$\frac{S_n - n\mu}{\sigma n^{1/2}} \Rightarrow \chi,$$

where χ has the standard normal distribution.

Proof. We first simplify our proof by shifting our random variables to have an expected value of 0 by defining $Y_i := X_i - \mu$. From the above proposition,

$$\varphi(t) = \mathbb{E}e^{etY_1} + \mathbf{o}(t^2).$$

Now consider,

$$\mathbb{E}\left[\exp\left(\frac{itS_n}{\sigma n^{1/2}}\right)\right] = \left(1 - \frac{t^2}{2n} + \mathbf{o}(n^{-1})\right)^n.$$

As $n \to \infty$ the last term goes to 0 and the rest of the expression goes to $\exp\left(\frac{-t^2}{2}\right)$. This is the characteristic function of the standard normal distribution. Therefore, it follows by (ii) of Lévy's continuity theorem that $(S_n - n\mu)/\left(\sigma n^{1/2}\right)$ converges weakly to the standard normal distribution. □

4. Wiener Processes

4.1. **Random Walk**

Now that we have finished our discussion of convergence theorems and the central limit theorem, we turn our focus to the Wiener process. Wiener processes are used to describe random motion and so it is natural to think of this as a continuous version of the discrete process of random walk on the integer lattice $\mathbb{Z}^d$, which we define here.

Definition 4.1.1. (Simple Random Walk) Let $\mathbf{e_1} \ldots \mathbf{e_d}$ be the standard orthonormal basis for the d-dimensional lattice $\mathbb{Z}^d$ and $(X_i)_{i \in \mathbb{N}}$ be a sequence of independent and identically distributed random variables where $\mathbb{P}(X_i = \mathbf{e_k}) = \mathbb{P}(X_i = \mathbf{e_k}) = 1/(2d)$. The **simple random walk on $\mathbb{Z}^\mathbf{d}$ starting at $\mathbf{x}$** is

$$S_n = x + \sum_{i=0}^{n} X_i.$$

Our goal for the rest of this paper is to use the measure and probability theory we have proven to define the Wiener process and show a few of its basic properties, then to construct the Wiener measure, and finally, to prove Donsker's invariance principle, which shows that indeed the Wiener process is the limit of a large class of random walks.

4.2. **Definition**

Definition 4.2.1. A **one-dimensional Wiener process** is a real valued process $W_t,\ t \geq 0$ with the following properties:

(i) W_t has independent increments. So, if $t_0 < t_1 < \ldots < t_n$, then $W(t_0)$, $W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$ are independent.
(ii) If $s, t \geq 0$, then $W(s+t) - W(s)$ has normal distribution with mean 0 and variance t.
(iii) W has continuous paths. That is, with probability one, $t \to W_t$ is continuous.

We say a one-dimensional Wiener process is **standard** if it additionally has the property that:

(iv) $W_0 = 0$.

4.3. **Fundamental Properties**

Proposition 4.3.1. *(Translation Invariance) For $t \geq 0$, if W is a Wiener process, then so is $W_{t+h} - W_h$ for some fixed $h \geq 0$.*

Proof. To prove this we must check all the parts of the definition of a Wiener process (Def. 4.2.1) for $W'_t := W_{t+h} - W_h$.

(i) Let $t_0 < t_1 < \ldots < t_n$. Note that $h < t_0 + h < t_1 + h < \ldots < t_n + h$. Therefore,

$$\begin{aligned} W'(t_0) &= W(t_0 + h) - W(h) \\ W'(t_1) - W'(t_0) &= W(t_1 + h) - W(t_0 + h) \\ &\vdots \\ W'(t_n) - W'(t_{n-1}) &= W(t_n + h) - W(t_{n-1} + h). \end{aligned}$$

These are independent because of the independent increments of W.
(ii) Now we want to look at the increment $W'_{t+s} - W'_s$. Expanding this out we see that $W'_{t+s} - W'_s = W_{t+s+h} - W_{s+h}$. Since part (ii) of the definition holds for any $s \geq 0$, we define $s' := s + h$ and see that by the definition of the Wiener process $W_{t+s+h} - W_{s+h} = W_{t+s'} - W_{s'}$, which has normal distribution with mean zero and variance t.
(iii) Note that W' is a composition consisting of a shift and a difference of W. Therefore, since W by definition is continuous, W' is also continuous.

Since W' satisfies all three parts of the definition it is a Wiener process. Thus, the Wiener process is translation invariant. □

Remark 4.3.2. One can note that the translation invariance of the Wiener process is a property it shares with the simple random walk. Next we turn to a scaling relation that random walks do not possess.

Proposition 4.3.3. *(Scaling Relation) For a standard Wiener process and any $t > 0$,*

$$\{W_{st} : s \geq 0\} \stackrel{d}{=} \{t^{1/2} W_s : s \geq 0\}.$$

Proof. The two families of random variables must have the same finite dimensional distributions. So for $k_1, \ldots, k_n$, we want $(W_{k_1,t}, \ldots, W_{k_n,t}) \stackrel{d}{=} (t^{1/2} B_{k_1}, \ldots, t^{1/2} B_{k_n})$. First, we show this for when $n = 1$. We see that multiplying $t^{1/2}$ by a normal distribution with expectation zero and variance s gives us

$$\mathbb{E}[t^{1/2} \cdot \mathcal{N}(0, s)] = t^{1/2} \cdot 0 = 0$$

$$\mathrm{Var}[t^{1/2} \cdot \mathcal{N}(0, s)] = \mathbb{E}[(t^{1/2} \cdot \mathcal{N}(0, s))^2] = ts$$

So, $t^{1/2} \cdot \mathcal{N}(0, s)$ has a normal distribution with mean zero and variance ts. For $n > 1$, the result follows from part (i) of the definition of the Wiener process (independent increments). □

5. Wiener Measure

In the study of stochastic processes (objects defined as a family of random variables), the process's *law* is defined to be the measure that induces the process on a family of functions. We wish to find a law that defines the Wiener process on the family of functions $C = C[0, 1]$ in order to help us study and understand the process. By defining and constructing this law, we find a distribution of the paths of the Wiener process. For the rest of the paper we restrict ourselves to the space $(C, \mathcal{C})$.

5.1. **Definition**

We define the **Wiener measure**, $\mathbb{W}$, on $(C, \mathcal{C})$ as the measure with the following two properties:

(i) For each t, the random variable X_t is normally distributed under $\mathbb{W}$ with mean 0 and variance t:

$$\mathbb{W}[X_t \leq \alpha] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} \exp\left(\frac{-u^2}{2}\right) du, \tag{13}$$

where we interpret $t = 0$ to mean $\mathbb{W}[X_0 = 0] = 1$.

(ii) The stochastic process $[X_t : 0 \leq t \leq 1]$ has independent increments under $\mathbb{W}$, meaning that if

$$0 \leq t_0 \leq t_1 \ldots \leq t_k = 1, \tag{14}$$

then the random variables,

$$X_{t_1} - X_{t_0}, \ldots, X_{t_k} - X_{t_{k-1}}, \tag{15}$$

are independent with respect to the measure $\mathbb{W}$. The standard Wiener process W is the random element on $(C, \mathcal{C}, \mathbb{W})$ defined by $W_t(x) = x(t)$.

5.2. **Construction**

Theorem 5.2.1. *There exists on $(C, \mathcal{C})$ a probability measure, $\mathbb{W}$, with the finite dimensional distribution specified by (13).*

Proof. We start our construction with a sequence of independent and identically distributed random variables, $Y_1, Y_2, \ldots$, on some probability space with mean 0 and positive finite variance. We define $S_n := \sum_{i=1}^n Y_i$ and define $X^n(w)$ to be the element of C where

$$X_t^n(w) := \frac{1}{\sigma\sqrt{n}} S_{\lfloor nt \rfloor}(w) + (nt - \lfloor nt \rfloor) \frac{1}{\sigma\sqrt{n}} Y_{\lfloor nt \rfloor + 1}(w) \tag{16}$$

at t. The function $X^n(w)$ is a linear interpolation (a linear mapping between points) between values at $X_{i/n}^n(w) = S_i(w)/(\sigma\sqrt{n})$ at points i/n.

Let us define the right hand term of (16) to be the function ψ. By Chebyshev's inequality, we see that the ψ converge weakly to 0 as $n \to \infty$. Furthermore, by the central limit theorem, Theorem 2.1.5, and the fact that $\lfloor nt \rfloor / n \to t$ as $n \to \infty$, we get that $X_t^n \Rightarrow_n \sqrt{t} \cdot \mathcal{N}(0,1)$.

If $s \leq t$, we can rewrite (16) as

$$(X_s^n, X_t^n - X_s^n) = \frac{1}{\sigma\sqrt{n}} (S_{\lfloor ns \rfloor}, S_{\lfloor nt \rfloor} - S_{\lfloor ns \rfloor}) + (\psi_{ns}, \psi_{nt} - \psi_{ns}) \underset{n}{\Rightarrow} (N_1, N_2) \tag{17}$$

where

$$\operatorname{Var}(N_1) = s \quad \text{and} \quad \operatorname{Var}(N_2) = t - s.$$

By the mapping theorem, $(X_s^n, X_t^n) \Rightarrow_n (N_1, N_1 + N_2)$. We see that if $\mathbb{P}_n$ is the distribution of X^n on C, then, for each (t_i) for $i = 1, 2, \ldots, k$, the finite distribution $\mathbb{P}_n \pi_{t_1, \ldots t_k}^{-1}$ converges weakly to what we want $\mathbb{W}\pi_{t_1, \ldots t_k}^{-1}$ to be: the normal distribution with independent increments.

If we can show that the sequence $(\mathbb{P}_n)$ is tight, it then follows from Prokhorov's theorem that some subsequence $(\mathbb{P}_{n_k})$ converges weakly to a limit that we will call $\mathbb{W}$. Then, $\mathbb{P}_{n_k} \pi_{t_1, \ldots t_k}^{-1} \Rightarrow_k \mathbb{W}\pi_{t_1, \ldots t_k}^{-1}$, which, by what we have already shown, is the probability measure on $\mathbb{R}^k$ that we want. First, we show the tightness of the probability measure of X^n on C when the sequence (Y_n) is stationary (meaning the distribution of $(Y_k, \ldots, Y_{k+i})$ is the same for all k).

Lemma 5.2.2. *Suppose X^n is defined by (16), that (Y_n) is stationary, and that*

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} \lambda^2 \mathbb{P} \left[\max_{k \leq n} |S_k| \geq \lambda \sigma \sqrt{n} \right] = 0. \tag{18}$$

Then, the sequence of probability measures $(\mathbb{P}_n)$ on C corresponding to (X^n) is tight.

Proof. To prove the sequence of probability measures is tight, we use Theorem 2.3.4. First, we want to show condition (i). Note that $X_0^n = 0$ and letting $a > 0$ we see that $\mathbb{P}[X^n : X_0^n > a] = 0$, so the condition follows. We can translate condition (ii) into the requirement that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}[w(X^n, \delta) \geq \epsilon] = 0 \tag{19}$$

for each ϵ. Applying Theorem 2.3.5,

$$\mathbb{P}[w(X^n, \delta) \geq 3\epsilon] \leq \sum_{i=1}^{v} \mathbb{P} \left(\sup_{t_{i-1} \leq s \leq t_i} \left| X_s^n - X_{t_{i-1}}^n \right| \geq \epsilon \right) \tag{20}$$

if $\min_{1<i<v}(t_i - t_{i-1}) \geq \delta$.

To make this easier to analyze, let $t_i = m_i/n$ for some $m_i \in \mathbb{N}$ satisfying $0 = m_0 < m_1 < \ldots < m_v = n$. Note the polygonal character of X^n as it is a linear interpolation of points. Therefore, if the t_i have the above form, then the supremum in (20) becomes $\left|S_k - S_{m_{i-1}}\right|/(\sigma\sqrt{n})$. So, (20) can be rewritten as

$$\begin{aligned}\mathbb{P}[w(X^n,\delta) \geq 3\epsilon] &\leq \sum_{i=1}^{v} \mathbb{P}\left(\max_{m_{i-1}\leq k\leq m_i} \frac{\left|S_k - S_{m_{i-1}}\right|}{\sigma\sqrt{n}} \geq \epsilon\right) \\ &= \sum_{i=1}^{v} \mathbb{P}\left(\max_{k\leq m_i - m_{i-1}} |S_k| \geq \epsilon\sigma\sqrt{n}\right), \end{aligned} \tag{21}$$

where the equality holds because of the assumed stationarity of the (Y_n). The inequality holds if the condition of (20) does. The inequality (20) can be reformulated as

$$\frac{m_i}{n} - \frac{m_{i-1}}{n} \geq \delta$$

for $1 < i < v$. Simplifying further, take $m_i = im$ where $0 \leq i < v$ and m is an integer chosen by the following criteria: (1) we need $m_i - m_{i-1} = m \geq n\delta$ for $i < v$. So take $m = \lceil n\delta \rceil$. (2) We also require $(v-1)m < n < vm$, so let $v = \lceil n/m \rceil$. Now, we see that

$$\begin{gathered} m_v - m_{v-1} \leq m \\ v = \left\lceil \frac{n}{m} \right\rceil \underset{n}{\rightarrow} \frac{1}{\delta} < \frac{2}{\delta} \\ \frac{n}{m} \underset{n}{\rightarrow} \frac{1}{\delta} > \frac{1}{2\delta}. \end{gathered}$$

For large n, (21) can be reformulated as

$$\begin{aligned}\mathbb{P}[w(X^n,\delta) \geq 3\epsilon] &\leq v\mathbb{P}\left(\max_{k\leq m} |S_k| \geq \epsilon\sigma\sqrt{n}\right) \\ &\leq \frac{2}{\delta}\mathbb{P}\left(\max_{k\leq m} |S_k| \geq \epsilon\sigma\frac{\sqrt{m}}{\sqrt{2\delta}}\right). \end{aligned} \tag{22}$$

Let us define λ and δ to be $\lambda = \epsilon/\sqrt{2\delta}$. Then, we see that (22) becomes

$$\mathbb{P}[w(X^n,\delta) \geq 3\epsilon] \leq \frac{4\lambda^2}{\epsilon^2}\mathbb{P}\left(\max_{k\leq m} |S_k| \geq \lambda\sigma\sqrt{m}\right).$$

Now we have an expression that looks like (18), so we can choose ϵ and η such that

$$\frac{4\lambda^2}{\epsilon^2}\mathbb{P}\left(\max_{k\leq m} |S_k| \geq \lambda\sigma\sqrt{m}\right).$$

Fixing δ, λ, and sending m and n to infinity, we obtain (19). By Theorem 2.3.4 we have shown that the probability measures of X^n on C are tight. $\square$

We complete the construction of the Wiener measure by using the independence of the random variables Y_n in (16). Etemadi's inequality (for statement see Appendix A), implies that

$$\mathbb{P}\left[\max_{k\leq m} |S_k| \geq \alpha\right] \leq 3\max_{k\leq m}\mathbb{P}\left[|S_k| \geq \alpha/3\right].$$

In order to meet the conditions of Lemma 5.2.2, we must satisfy (18), which we can reformulate with Etemadi's inequality to be

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \lambda^2 \max_{k\leq n} \mathbb{P}\left[|S_k| \geq \lambda\sigma\sqrt{n}\right] = 0. \tag{23}$$

In constructing the Wiener measure, we may use any sequence of random variables (Y_n) that is convenient since we are only trying to prove the existence of the measure. Therefore, we choose Y_n that are independent and have the standard normal distribution. By the central limit theorem, the partial sum $S_k/\sqrt{k}$ also has the standard normal distribution denoted here by $\mathcal{N}$. Note that, by Markov's inequality,

$$\mathbb{P}\left[|\mathcal{N}| \geq \lambda\right] = \mathbb{P}\left[\mathcal{N}^4 \geq \lambda^4\right] \leq \frac{\mathbb{E}\mathcal{N}^4}{\lambda^4} = \frac{3}{\lambda^4}.$$

Therefore,

$$\mathbb{P}\left[|S_k| \geq \lambda\sigma\sqrt{n}\right] = \mathbb{P}\left[\sqrt{k}\,|\mathcal{N}|\right] \geq \lambda\sigma\sqrt{n}] \leq \frac{3k^2}{n^2\sigma^4\lambda^4}.$$

For $k \leq n$, this implies that

$$\mathbb{P}\left[|S_k| \geq \lambda\sigma\sqrt{n}\right] \leq \frac{3}{\sigma^4\lambda^4}.$$

Letting λ go to infinity we get (23). Therefore, we obtain the condition for Lemma 5.2.2 and so we see that this family of probability measures is tight. By Prokhorov's theorem, we know that some subsequence converges and by the work we did before the lemma we know that this limit converges to the measure with the properties that we want. Therefore, we have constructed the Wiener measure. □

6. Donsker's Invariance Principle

We devote the final section of this paper solely to proving Donsker's invariance principle, which show us that Brownian motion is the limit of random walk. Having constructed the Wiener measure and proven Theorem 2.3.6 we have already done most of the work.

Theorem 6.1.1. *(Donsker's Invariance Principle) If $Y_1, Y_2, \ldots$ are independent and identically distributed with mean 0 and variance σ^2 and if X^n is a random variable defined by (16), then $X^n \Rightarrow W$ as $n \to \infty$.*

Proof. Having proved the existence of the Wiener measure, $\mathbb{W}$, and thus the corresponding random function, W, we can write (17) as

$$(X_s^n, X_t^n - X_s^n) \underset{n}{\Rightarrow} (W_s, W_t - W_s),$$

which implies

$$(X_s^n, X_t^n) \underset{n}{\Rightarrow} (W_s, W_t).$$

Therefore, we can obtain the condition for Theorem 2.3.6

$$\left(X_{t_1}^n, \ldots, X_{t_k}^n\right) \underset{n}{\Rightarrow} \left(W_{t_1}, \ldots, W_{t_k}\right).$$

We now just need to show the second condition for Theorem 2.3.6,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left[w(X^n, \delta) \geq \epsilon\right] = 0.$$

We have shown this for normally distributed Y_i, however we must now extend it. In the proof of Lemma 5.2.2. we showed that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left[w(X^n,\delta)\geq\epsilon\right]\leq v\mathbb{P}\left[\max_{k\leq m}|S_k|\geq\epsilon\sigma\sqrt{n}\right]. \tag{24}$$

By Etemadi's inequality, we then see that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left[w(X^n,\delta)\geq\epsilon\right]\leq 3v\max_{k\leq m}\mathbb{P}\left[|S_k|\geq\frac{\epsilon\sigma\sqrt{n}}{3}\right]. \tag{25}$$

It thus suffices for us to check (23) when $\sigma=1$. And so we split our analysis into two cases. First, for large k, we use the central limit theorem to show that the partial sum converges to the standard normal distribution. So, by the central limit theorem, if k_λ in the maximum of (23) is large enough and $k_\lambda\leq k\leq n$, then

$$\mathbb{P}\left[|S_k|\geq\lambda\sigma\sqrt{n}\right]\leq\mathbb{P}\left[|S_k|\geq\lambda\sigma\sqrt{k}\right]<\frac{3}{\lambda^4}.$$

In the second case, for small $k\leq k_\lambda$, we use Chebyshev's inequality to show that

$$\mathbb{P}\left[|S_k|\geq\lambda\sigma\sqrt{n}\geq\frac{k_\lambda}{\lambda^2 n}.\right.$$

Therefore, the maximum in (23) is dominated by $(3/\lambda^4)\vee(k_\lambda/(\lambda^2 n))$. Thus, we have satisfied the conditions of Theorem 2.3.6 and therefore $X^n\Rightarrow_n W$. □

Acknowledgments

I would first and foremost like to thank my mentor Yucheng Deng for his constant help and guidance throughout this project. I would also like to thank Gregory Lawler for his lecture series "Random Walk and the Heat Equation," which provided a starting point for this project. Furthermore, I'd like to thank Vivian Healey and Abdalla Dali Nimer for reading drafts of this paper and providing feedback. Last, but certainly not least, I would like to thank Peter May for organizing the REU.

References

[1] Patrick Billingsley. Convergence of Probability Measures 2e. John Wiley Sons, Inc. 1999.
[2] Rick Durrett. Probability: Theory and Examples 4.1e. Cambridge University Press. 2013.
[3] Gregory F. Lawler and Vlada Limic. Random Walk: A Modern Introduction. Cambridge University Press. 2010.
[4] Peter Mörters and Yuval Peres. Brownian Motion https://www.stat.berkeley.edu/ aldous/205B/bmbook.pdf.
[5] Serik Sagitov. Weak Convergence of Probability Measures. http://www.math.chalmers.se/ serik/C-space.pdf

Appendix A. Inequalities

Stowed away here are a few non-trivial inequalities used in the paper.

Theorem A.0.1. *(Markov's Inequality) Given a probability space* (X,Σ,μ), *a measurable extended real-valued function* f, *a set* $B\in X$, *and an* $\epsilon>0$,

$$\mu\left(\{x\in B: f(x)\geq\epsilon\}\right)\leq\frac{1}{\epsilon}\int_B f\,d\mu.$$

Proof. Assume without loss of generality that $f(x) \geq 0$. Let K be the indicator function on the set $A = (\{x \in B : f(x) \geq \epsilon\})$ for $B \in X$. Note that for $\epsilon > 0$, if $x \in A$, then

$$\epsilon K(x) = \epsilon \leq f(x),$$

and if $x \notin A$, then

$$\epsilon K(x) = 0 \leq f(x).$$

So $\epsilon K(x) < f(x)$. By the properties of the integral,

$$\int_B \epsilon K(x) d\mu \leq \int_B f(x) d\mu.$$

Finally, observe that

$$\int_B \epsilon K(x) d\mu = \epsilon \int_B K(x) d\mu = \epsilon \mu(A).$$

Therefore,

$$\mu\left(\{x \in X : f(x) \geq \epsilon\}\right) \leq \frac{1}{\epsilon} \int_B f d\mu.$$

For functions f that are not strictly positive, let $f = f^+ - f^-$ and the apply the same proof to both parts. □

Corollary A.0.2. *(Chebyshev's Inequality) Let X be a random variable with finite expectation μ and finite, non-zero σ^2. Then,*

$$\mathbb{P}\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}.$$

Proof. Note that we can reformulate Markov's inequality as

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}x}{a}.$$

Taking $Y = (X - \mu)^2$ and $a = (\sigma k)^2$ and noting that $\mathrm{Var}(Y) = \sigma^2 = \mathbb{E}(X - \mu)^2$, the inequality follows immediately. □

Proposition A.0.3. *(Etemadi's Inequality) Suppose $(S_k)_{k=1}^n$ is a sequence of partial sums of independent random variables $X_1, \ldots, X_n$ where $S_k = \sum_{i=0}^k X_i$. Then,*

$$\mathbb{P}\left[\max_{k \leq n} |S_k| \geq 3\alpha\right] \leq 3 \max_{k \leq n} \mathbb{P}\left[|S_k| \geq \alpha\right].$$

A proof of Etemadi's inequality can be found in [1, appx. M19].