CLASSICAL TOPICS IN MEASURE-THEORETIC PROBABILITY

MAHNAV PETERSEN

Abstract. This expository paper rigorously develops a number of classical topics in measure-theoretic probability. We first discuss some notions of convergence for random variables and prove two versions of the Law of Large Numbers. We then establish the basic properties of characteristic functions, which we subsequently use to give a short proof of (one version of) the Central Limit Theorem. In the final section, we establish the existence of Brownian motion and prove some of its basic properties, including the striking result that linear Brownian motion on $[0, 1]$ is almost surely nowhere differentiable.

Contents

1. Introduction	1
2. Convergence of Random Variables	2
2.1. Modes of Convergence	2
2.2. Law of Large Numbers	8
3. Characteristic Functions	10
3.1. Basic Properties of Characteristic Functions	10
3.2. Normal Random Variables	12
3.3. Central Limit Theorem	15
4. Brownian Motion	16
4.1. Existence of Brownian Motion	17
4.2. Wiener Measure	20
4.3. Multidimensional Brownian Motion	21
4.4. Differentiability of Brownian Motion	23
Acknowledgments	24
References	24

1. Introduction

In this expository paper, we develop some classical topics in measure-theoretic probability: convergence of random variables, characteristic functions, and the basics of Brownian motion. We assume that the reader knows measure theory, including Dynkin's π-λ theorem and the Fourier transform, and we also assume that the reader has a strong understanding of basic probabilistic concepts like expectation and independence.

We fix some notation that will be used throughout the paper. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $d \in \mathbb{N}$. (We choose the symbol $\mathbb{F}$ since both $\mathbb{R}$ and $\mathbb{C}$ are fields.) We let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and we also let $[n] = \{1, 2, \ldots, n\}$ and $[n]_0 = \{0, 1, 2, \ldots, n\}$ for all $n \in \mathbb{N}_0$. In particular, $[0] = \emptyset$. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and if $A \in \mathcal{F}$ is an event, then we let 1_A denote the indicator random variable associated with A. Also, if X is a random variable, then we let μ_X denote its distribution, and, if X is real-valued, then we let F_X denote its distribution function (so $F_X(x) = \mu_X((-\infty, x])$ for all $x \in \mathbb{R}$).

We conclude this introduction with a few basic results that will be used frequently in the paper.

Proposition 1.1. *Let $(\Omega, \mathcal{F})$ be a measurable space, and let $\mathbb{P}_1$ and $\mathbb{P}_2$ be probability measures on $(\Omega, \mathcal{F})$. Suppose that $\mathcal{P}$ is a π-system in Ω such that $\sigma(\mathcal{P}) = \mathcal{F}$ (which, in particular, implies $\mathcal{P} \subset \mathcal{F}$) and such that $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ for all $A \in \mathcal{P}$. Then $\mathbb{P}_1 = \mathbb{P}_2$.*

Proof. Let $\mathcal{L}$ be the collection of all $A \in \mathcal{F}$ for which $\mathbb{P}_1(A) = \mathbb{P}_2(A)$. It is easy to check that $\mathcal{L}$ is a λ-system in Ω containing $\mathcal{P}$, so, by Dynkin's π-λ theorem, $\mathcal{F} = \sigma(\mathcal{P}) = \mathcal{L}$. QED

Proposition 1.2. *Let $(\Gamma, \mathcal{G})$ and $(\Delta, \mathcal{H})$ be measurable spaces, and let $f : \Gamma \to \Delta$ be a measurable function. Suppose X and Y are Γ-valued random variables that agree in distribution. Then $f(X)$ and $f(Y)$ agree in distribution.*

Proof. Let X be defined on the probability space $(\Omega_X, \mathcal{F}_X, \mathbb{P}_X)$, and let Y be defined on the probability space $(\Omega_Y, \mathcal{F}_Y, \mathbb{P}_Y)$. If $A \in \mathcal{H}$, then $\mathbb{P}_X(f(X) \in A) = \mathbb{P}_X(X \in f^{-1}(A)) = \mathbb{P}_Y(Y \in f^{-1}(A)) = \mathbb{P}_Y(f(Y) \in A)$. QED

The following theorem is a well-known result about independent random variables, so we omit its proof.[1]

Theorem 1.3. *For any collection of probability distributions $\{\mu_\alpha\}_{\alpha \in I}$ (possibly defined on different measurable spaces), there exist a probability space and a collection of independent random variables $\{X_\alpha\}_{\alpha \in I}$, all defined on that probability space, such that each X_α has distribution μ_α (i.e., $\mu_{X_\alpha} = \mu_\alpha$).*

2. Convergence of Random Variables

Throughout this section, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a topological space $\mathcal{X}$.

2.1. Modes of Convergence.

Throughout this subsection, unless stated otherwise, we let (X_n) be a sequence of $\mathbb{F}^d$-valued random variables, all defined on $(\Omega, \mathcal{F}, \mathbb{P})$, and we let X be another $\mathbb{F}^d$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.1. Let E be the set of $\omega \in \Omega$ for which the sequence $(X_n(\omega))$ converges. Then $E = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{(m,n) \in (\mathbb{N} \setminus [N])^2} \{\omega \in \Omega : |X_m(\omega) - X_n(\omega)| < \frac{1}{k}\}$, so E is a measurable set in $\mathcal{F}$. It follows that the set of $\omega \in \Omega$ for which the relation $\lim_{n \to \infty} X_n(\omega) = X(\omega)$ holds is also a measurable set in $\mathcal{F}$; we will denote the probability of this set by $\mathbb{P}(X_n \to X)$. If $\mathbb{P}(X_n \to X) = 1$, then we say that **$(X_n)$ converges to X almost surely (a.s.)**, and we write "$X_n \to X$ a.s. (as $n \to \infty$)" to denote this convergence.

---

[1] This can be proven using the notion of the product probability measure for an arbitrary product of probability spaces; we take the X_α as the coordinate projections from the product space. See [1] for details regarding the construction of the product probability measure.

Definition 2.2. We say that **(X_n) converges to X in probability** if $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$ for all $\epsilon > 0$. We write "$X_n \to X$ in probability (as $n \to \infty$)" to denote this convergence.

Convergence a.s. implies convergence in probability, in the sense if $X_n \to X$ a.s., then $X_n \to X$ in probability as well. The proof is immediate using the reverse Fatou lemma (with the constant function 1 as the dominating function): $\limsup_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = \limsup_{n\to\infty} \mathbb{E}(1_{(|X_n-X|^{-1})(\epsilon,\infty)}) \leq \mathbb{E}(\limsup_{n\to\infty} 1_{(|X_n-X|^{-1})(\epsilon,\infty)}) = 0$ for all $\epsilon > 0$. However, the converse is not true in general, as we show in Example 2.3 below. Hence, convergence a.s. is strictly stronger than convergence in probability.

Example 2.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which there exists an independent sequence of random variables (X_n) such that $\mu_{X_n} = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_1$.[2] (The existence of such a probability space is guaranteed by Theorem 1.3.) Then $X_n \to 0$ in probability. However, for any $N \in \mathbb{N}$, we have by independence that $\mathbb{P}(\{\omega \in \Omega : |X_n(\omega)| < \frac{1}{2} \text{ for all } n \in \mathbb{N} \setminus [N]\}) = \lim_{M\to\infty} \mathbb{P}(X_{N+1} \in (-\frac{1}{2}, \frac{1}{2}), X_{N+2} \in (-\frac{1}{2}, \frac{1}{2}), \ldots, X_M \in (-\frac{1}{2}, \frac{1}{2})) = \lim_{M\to\infty} \prod_{n=N+1}^{M} (1 - \frac{1}{n}) \leq \limsup_{M\to\infty} \prod_{n=N+1}^{M} e^{\frac{-1}{n}} = 0$. (Note that $1 + x \leq e^x$ for all $x \in \mathbb{R}$.) Thus, $\mathbb{P}(\bigcup_{N=1}^{\infty} \{\omega \in \Omega : |X_n(\omega)| < \frac{1}{2} \text{ for all } n \in \mathbb{N} \setminus [N]\}) = 0$, so (X_n) does not converge to X a.s..

Although convergence in probability does not imply convergence a.s., we do have the following useful result:

Proposition 2.4. *$X_n \to X$ in probability if and only if, for every subsequence (X_{n_k}) of (X_n), there is a further subsequence $(X_{n_{k_j}})$ such that $X_{n_{k_j}} \to X$ a.s. as $j \to \infty$.*

Proof. Suppose that $X_n \to X$ in probability. Let (X_{n_k}) be a subsequence of (X_n). Fix $m \in \mathbb{N}$. Since $\mathbb{P}(|X_{n_k} - X| > \frac{1}{m}) \to 0$ as $k \to \infty$, we can find a further subsequence $(X_{n_{k_j}})$ such that $\mathbb{P}(|X_{n_{k_j}} - X| > \frac{1}{m}) < \frac{1}{2^j}$ for all $j \in \mathbb{N}$. By the Borel-Cantelli lemma, there exists $A_m \in \mathcal{F}$ such that $\mathbb{P}(A_m) = 1$ and such that, for all $\omega \in A_m$, there are at most finitely many $j \in \mathbb{N}$ for which $|X_{n_{k_j}}(\omega) - X(\omega)| > \frac{1}{m}$. Then $\mathbb{P}(\bigcap_{m=1}^{\infty} A_m) = 1$, and, if $\omega \in \bigcap_{m=1}^{\infty} A_m$, then $X_{n_{k_j}}(\omega) \to X(\omega)$ as $j \to \infty$. Thus, $X_{n_{k_j}} \to X$ a.s. as $j \to \infty$.

For the converse, fix $\epsilon > 0$, and consider the sequence $(\mathbb{P}(|X_n - X| > \epsilon))$. Since convergence a.s. implies convergence in probability, every subsequence of this sequence has a further subsequence that converges to 0, which implies that $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$. QED

Corollary 2.5. *If $X_n \to X$ in probability, then there is a subsequence (X_{n_k}) such that $X_{n_k} \to X$ a.s. as $k \to \infty$.*

Remark 2.6. For the proof of the converse of Proposition 2.4, we used the fact that, if (x_n) is a sequence in a topological space $\mathcal{X}$, and if $x \in \mathcal{X}$, then $x_n \to x$ as $n \to \infty$ if and only if, for every subsequence (x_{n_k}) of (x_n), there is a further subsequence $(x_{n_{k_j}})$ that converges to x (as $j \to \infty$). Note that, despite this fact, Proposition 2.4 does not show that convergence in probability implies convergence

[2]Note that δ_x denotes the Dirac measure concentrated at $x \in \mathbb{R}$, so $\delta_x(B) = 1$ for every Borel set $B \subset \mathbb{R}$ containing x, while $\delta_x(B) = 0$ for every Borel set $B \subset \mathbb{R}$ that does not contain x.

a.s. (which we know to be false by Example 2.3). The key observation is that a sequence has uncountably many subsequences, so convergence *almost surely* along each subsequence does not imply convergence almost surely for the whole sequence. If "a.s." was replaced by "surely" in Proposition 2.4, then Proposition 2.4 *would* imply that convergence in probability implies pointwise convergence (and hence convergence a.s.).

We now discuss uniqueness of limits:

Proposition 2.7. *Let Y be another $\mathbb{F}^d$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. If $X_n \to X$ in probability and $X_n \to Y$ in probability, then $X = Y$ a.s.. In particular, if $X_n \to X$ a.s. and $X_n \to Y$ a.s., then $X = Y$ a.s. (although this statement about uniqueness of a.s. limits can easily be proven directly).*

Proof. Fix $m \in \mathbb{N}$. Then for $n \in \mathbb{N}$, we have $\mathbb{P}(|X - Y| > \frac{1}{m}) \le \mathbb{P}(|X - X_n| > \frac{1}{2m}) + \mathbb{P}(|X_n - Y| > \frac{1}{2m}) \to 0$ as $n \to \infty$, so $\mathbb{P}(|X - Y| > \frac{1}{m}) = 0$. Hence, $\mathbb{P}(|X - Y| > 0) = 0$, so $X = Y$ a.s.. QED

The significance of the following result will be apparent shortly:

Proposition 2.8. *Let $\mathbb{F}_1 = \mathbb{F}$, and let $\mathbb{F}_2 \in \{\mathbb{R}, \mathbb{C}\}$ and $k \in \mathbb{N}$. Let $f : \mathbb{F}_1^d \to \mathbb{F}_2^k$ be continuous.*

(a) If $X_n \to X$ a.s. as $n \to \infty$, then $f(X_n) \to f(X)$ a.s. as $n \to \infty$. Furthermore, if f is bounded, then $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$.

(b) If $X_n \to X$ in probability as $n \to \infty$, then $f(X_n) \to f(X)$ in probability as $n \to \infty$. Furthermore, if f is bounded, then $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$.

Proof. For (a), the continuity of f implies that $f(X_n) \to f(X)$ a.s. as $n \to \infty$, and, if f is bounded, we can apply the dominated convergence theorem to conclude that $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$.

For (b), let (X_{n_k}) be a subsequence of (X_n). By Proposition 2.4, we can choose a further subsequence $(X_{n_{k_j}})$ that converges a.s. to X_n as $j \to \infty$. By (a), we have $f(X_{n_{k_j}}) \to f(X)$ a.s. as $j \to \infty$, so, by Proposition 2.4 again, we have $f(X_n) \to f(X)$ in probability as $n \to \infty$. If f is bounded, we can apply (a) again to conclude $\mathbb{E}(f(X_{n_{k_j}})) \to \mathbb{E}(f(X))$ as $j \to \infty$. Since every subsequence of $\mathbb{E}(f(X_n))$ has a further subsequence converging to $\mathbb{E}(f(X))$, we have $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$. QED

Definition 2.9. For this definition, let (X_n) be a sequence of $\mathcal{X}$-valued random variables (where $X_1, X_2, \ldots$ *need not* be defined on the same probability space), and let X be another $\mathcal{X}$-valued random variable. We say that **(X_n) converges to X in distribution** if $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$ for every bounded, continuous function $f : \mathcal{X} \to \mathbb{R}$. We write "$X_n \to X$ in distribution (as $n \to \infty$)" to denote this convergence.

We observe that convergence in distribution depends only on the random variables' distributions. By Proposition 2.8, convergence in probability implies convergence in distribution, and hence convergence a.s. implies convergence in distribution. The converse is not true in general, as we show in Example 2.10 below. Therefore, convergence in distribution is strictly weaker than convergence in probability, which is itself strictly weaker than convergence a.s. (although the notion of convergence in distribution applies to a wider class of random variables).

Example 2.10. Let (X_n) be a sequence of real-valued, **independent and identically distributed (i.i.d.)** random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that $\mu_{X_1} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Clearly, $X_n \to X_1$ in distribution. However, $\mathbb{P}(|X_n - X_1| \leq \frac{1}{2}) \leq \mathbb{P}(X_1 = 0, X_n = 0) + \mathbb{P}(X_1 = 1, X_n = 1) = \frac{1}{2}$ by independence for all $n \in \mathbb{N}$, so (X_n) does not converge to X_1 in probability (and hence (X_n) does not converge to X_1 a.s.).

Example 2.11. Let (x_n) be a sequence of real numbers converging to $x \in \mathbb{R}$. For each $n \in \mathbb{N}$, let X_n be a real-valued random variable with distribution δ_{x_n}, and let X be a real-valued random variable with distribution δ_x. Then $X_n \to X$ in distribution, since if $f : \mathbb{R} \to \mathbb{R}$ is a (bounded) continuous function, then $\mathbb{E}(f(X_n)) = f(x_n) \to f(x) = \mathbb{E}(f(X))$ as $n \to \infty$.

Uniqueness of limits is a more subtle topic for convergence in distribution. Since this type of convergence depends only on the random variables' distributions, the strongest form of uniqueness that we can expect is that any two limits have the same distribution. For $\mathbb{R}^d$-valued random variables, this will follow from Corollary 2.16 below. We begin with some preliminary results.

Theorem 2.12. *Let (X_n) be a sequence of $\mathbb{R}^d$-valued random variables (where $X_1, X_2, \ldots$ need not be defined on the same probability space), and let X be another $\mathbb{R}^d$-valued random variable. Suppose that $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ for every smooth (i.e., infinitely differentiable), compactly supported function $f : \mathbb{R}^d \to [0,1]$. (In particular, this hypothesis holds if $X_n \to X$ in distribution.) Then $\mu_{X_n}(R) \to \mu_X(R)$ as $n \to \infty$ for every rectangle[3] R in $\mathbb{R}^d$ such that $\mu_X(bd(R)) = 0$, where $bd(R)$ denotes the topological boundary of R.*

Proof. Let $R = I_1 \times I_2 \times \ldots \times I_d$ be a rectangle in $\mathbb{R}^d$ such that $\mu_X(\mathrm{bd}(R)) = 0$, where $I_1, I_2, \ldots, I_d \subset \mathbb{R}$ are intervals. For $j \in [d]$, let a_j be the left endpoint of I_j and b_j be the right endpoint of I_j.

Fix $k \in \mathbb{N}$. Let $R_k = (a_1 - \frac{1}{k}, b_1 + \frac{1}{k}) \times (a_2 - \frac{1}{k}, b_2 + \frac{1}{k}) \times \cdots \times (a_d - \frac{1}{k}, b_d + \frac{1}{k})$. Let $f_k : \mathbb{R}^d \to [0,1]$ be a smooth, compactly supported function such that $f_k(x) = 1$ for all $x \in R$ and $f_k(x) = 0$ for all $x \in \mathbb{R}^d \setminus R_k$.[4] Then $\mu_X(R_k) = \mathbb{E}(1_{R_k}(X)) \geq \mathbb{E}(f_k(X))$, and $\mathbb{E}(f_k(X_n)) \geq \mathbb{E}(1_R(X_n)) = \mu_{X_n}(R)$. Thus, $\limsup_{n\to\infty} \mu_{X_n}(R) \leq \limsup_{n\to\infty} \mathbb{E}(f_k(X_n)) = \mathbb{E}(f_k(X)) \leq \mu_X(R_k)$. By the continuity of measure and the fact that $\mu_X(\mathrm{bd}(R)) = 0$, we have $\lim_{k\to\infty} \mu_X(R_k) = \mu_X(R \cup \mathrm{bd}(R)) = \mu_X(R)$, so the relation $\limsup_{n\to\infty} \mu_{X_n}(R) \leq \mu_X(R_k)$ implies that $\limsup_{n\to\infty} \mu_{X_n}(R) \leq \mu_X(R)$. An analogous argument (by approximating R using rectangles of the form $(a_1 + \frac{1}{k}, b_1 - \frac{1}{k}) \times (a_2 + \frac{1}{k}, b_2 - \frac{1}{k}) \times \cdots \times (a_d + \frac{1}{k}, b_d - \frac{1}{k})$ for $k \in \mathbb{N}$) yields that $\liminf_{n\to\infty} \mu_{X_n}(R) \geq \mu_X(R)$. QED

Definition 2.13. Fix $j \in [d]$ and $a \in \mathbb{R}$. We define $H_{j,a} \subset \mathbb{R}^d$ as the hyperplane obtained by fixing the j^{th} coordinate at a. More formally, we let $H_{j,a}$ be the set of all $(x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ such that $x_j = a$.

---

[3]By "rectangle," we mean a subset of the form $I_1 \times I_2 \times \cdots \times I_d$ for bounded intervals $I_1, I_2, \ldots, I_d \subset \mathbb{R}$. The representation of a nonempty rectangle as a product of bounded intervals is unique. Note that this is a different use of the word "rectangle" than in Section 4.3.

[4]The existence of such a function is a well-known result from analysis; see Chapter 8.2 of [5] for details.

Lemma 2.14. *Fix $j \in [d]$. Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Then $\mu(H_{j,a}) > 0$ for at most countably many $a \in \mathbb{R}$.*

Proof. This is immediate, since $\sum_{a\in\mathbb{R}} \mu(H_{j,a}) \le \mu(\mathbb{R}^d) < \infty$.[5] QED

Corollary 2.15. *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}^d})$. Let $\mathcal{P}_\mu$ be the collection of all nonempty, open rectangles $R = I_1 \times I_2 \times \cdots \times I_d$ in $\mathbb{R}^d$ such that $\mu(H_{j,a_j}) = \mu(H_{j,b_j}) = 0$ for all $j \in [d]$, where a_j is the left endpoint of I_j and b_j is the right endpoint of I_j (and where $I_1, I_2, \dots, I_d \subset \mathbb{R}$ are bounded intervals). Then $\mathcal{P} \cup \{\varnothing\}$ is a π-system in $\mathbb{R}^d$ such that $\sigma(\mathcal{P}) = \mathcal{B}_{\mathbb{R}^d}$.*

Proof. It is straightforward to check that $\mathcal{P} \cup \{\varnothing\}$ is a π-system. To check that $\sigma(\mathcal{P}) = \mathcal{B}_{\mathbb{R}^d}$, note that, by Lemma 2.14, we can write any open rectangle in $\mathbb{R}^d$ as a countable union of rectangles in $\mathcal{P} \cup \{\varnothing\}$, and that every open set in $\mathbb{R}^d$ is a countable union of open rectangles. QED

Corollary 2.16. *Two $\mathbb{R}^d$-valued random variables X and Y (which need not be defined on the same sample space) agree in distribution if and only if $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$ for every smooth, compactly supported function $f : \mathbb{R}^d \to [0,1]$.*

Proof. Let $\mathcal{P}_{\mu_X}$ be as in Corollary 2.15. Suppose $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$ for every smooth, compactly supported function $f : \mathbb{R}^d \to [0,1]$. Then by Theorem 2.12, we have $\mu_Y(R) = \mu_X(R)$ for every rectangle R in $\mathbb{R}^d$ such that $\mu_X(bd(R)) = 0$. In particular, we have $\mu_Y(R) = \mu_X(R)$ for every rectangle R in $\mathcal{P} \cup \{\varnothing\}$, so $\mu_Y = \mu_X$ by Proposition 1.1 and Corollary 2.15. The converse is clear. QED

The definition for convergence in distribution is rather abstract, so it would be nice to have a more concrete way to characterize this convergence. We will do this for real-valued random variables in Theorem 2.23 below. Specifically, we will give the concrete characterization that, if $X_1, X_2, \dots$ are real-valued random variables, and if X is another real-valued random variable, then $X_n \to X$ in distribution if and only if $F_{X_n}(t) \to F_X(t)$ as $n \to \infty$ for every $t \in \mathbb{R}$ at which F_X is continuous. By taking $x_n = \frac{1}{n}$ for all $n \in \mathbb{N}$ in Example 2.11, we see that the continuity condition is not superfluous. (Note that F_X is increasing and hence has at most countably many points of discontinuity.) However, we need to develop a little more machinery before proving Theorem 2.23.

Definition 2.17. Let $F : \mathbb{R} \to [0,1]$ be an increasing, right-continuous function satisfying $\lim_{t\to\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$. (Thus, F is the distribution function of its Lesbegue-Stieltjes measure.) We define the **generalized inverse of $\boldsymbol{F}$** as the function $G : (0,1) \to \mathbb{R}$, given by $G(x) = \sup\{t \in \mathbb{R} : F(t) < x\}$ for all $x \in \mathbb{R}$. (Note that the relations $\lim_{t\to\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$ imply that G is well-defined.) Observe that G is increasing and hence Borel-measurable.

Lemma 2.18. *Let $F : \mathbb{R} \to [0,1]$ be an increasing, right-continuous function satisfying $\lim_{t\to\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$. Let G be the generalized inverse of F. Fix $x \in (0,1)$ and $t \in \mathbb{R}$. Then $x \le F(t)$ if and only if $G(x) \le t$.*

---

[5]We are using the basic fact that, if an indexed sum of nonnegative numbers is finite, then at most countably many of the terms in the sum are nonzero. See pages 83-84 of [2] and pages xii-xiii of [4] for details.

Proof. If $x \le F(t)$, then the monotonicity of f implies that $G(x) \le t$. If $x > F(t)$, then, by right-continuity, $x > F(t+\epsilon)$ for some $\epsilon > 0$, so $G(x) \ge t + \epsilon > t$. QED

Lemma 2.19. *Let $F : \mathbb{R} \to [0,1]$ be an increasing, right-continuous function satisfying $\lim_{t\to\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$. Let Ω be the collection of all $x \in (0,1)$ such that $\sup\{t \in \mathbb{R} : F(t) < x\} = \inf\{t \in \mathbb{R} : F(t) > x\}$. (Again, the relations $\lim_{t\to\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$ imply that Ω is well-defined.) Then $(0,1) \setminus \Omega$ is countable.*

Proof. For each $x \in (0,1)$, let $a_x = \sup\{t \in \mathbb{R} : F(t) < x\}$ and $b_x = \inf\{t \in \mathbb{R} : F(t) > x\}$; note that $a_x \le b_x$. Suppose $x_1, x_2 \in (0,1) \setminus \Omega$ satisfy $x_1 < x_2$. Then $b_{x_1} \le a_{x_2}$, so (a_{x_1}, b_{x_1}) and (a_{x_2}, b_{x_2}) are disjoint, nonempty intervals. Thus, we can injectively associate each $x \in (0,1) \setminus \Omega$ with a rational number, which implies that $(0,1) \setminus \Omega$ is countable. QED

Proposition 2.20. *Let $X_1, X_2, \ldots$ be real-valued random variables, and let X be another real-valued random variable. For each $n \in \mathbb{N}$, let Y_n be the generalized inverse of F_{X_n}. Also, let Y be the generalized inverse of F_X. Note that we can regard each Y_n as a real-valued random variable on the probability space $((0,1), \mathcal{B}_{(0,1)}, m)$; similarly, we can regard Y as a real-valued random variable on $((0,1), \mathcal{B}_{(0,1)}, m)$.*

(a) $\mu_{X_n} = \mu_{Y_n}$ for all $n \in \mathbb{N}$, and $\mu_X = \mu_Y$.
(b) Suppose $F_{X_n}(t) \to F_X(t)$ as $n \to \infty$ for every $t \in \mathbb{R}$ at which F_X is continuous. Then $Y_n \to Y$ a.s.. In fact, let Ω be the collection of all $x \in (0,1)$ such that $\sup\{t \in \mathbb{R} : F_X(t) < x\} = \inf\{t \in \mathbb{R} : F_X(t) > x\}$. Then $Y_n(x) \to Y(x)$ as $n \to \infty$ for all $x \in \Omega$, which implies that $Y_n \to Y$ a.s. by Lemma 2.19.

Proof. For all $n \in \mathbb{N}$, by Lemma 2.18, we have $F_{Y_n}(t) = m(Y_n \le t) = m(\{x \in (0,1) : x \le F_{X_n}(t)\}) = F_{X_n}(t)$ for all $t \in \mathbb{R}$, so $\mu_{Y_n} = \mu_{X_n}$. Similarly, $\mu_Y = \mu_X$, which gives (a).

We now consider (b). Fix $x \in \Omega$, and fix any $s \in \mathbb{R}$ at which F_X is continuous. Suppose $s > Y(x)$. By Lemma 2.18, we have $x \le F_X(s)$. If $x = F_X(s)$, then $s > Y(x) = \sup\{t \in \mathbb{R} : F_X(t) < x\} = \inf\{t \in \mathbb{R} : F_X(t) > x\} \ge s$, which is absurd. Hence, $x < F_X(s)$, so, whenever $n \in \mathbb{N}$ is sufficiently large, we have $x \le F_{X_n}(s)$ and thus $Y_{X_n}(x) \le s$. Therefore, $\limsup_{n\to\infty} Y_{X_n}(x) \le s$, so, by our choice of s, we conclude $\limsup_{n\to\infty} Y_{X_n}(x) \le Y(x)$.

Now suppose $s < Y(x)$. By Lemma 2.18, $F_X(s) < x$, so, whenever $n \in \mathbb{N}$ is sufficiently large, $F_{X_n}(s) < x$ and hence $s < Y_n(x)$. Thus, $\liminf_{n\to\infty} Y_n(s) \ge s$, so, by our choice of s, we conclude $\liminf_{n\to\infty} Y_n(s) \ge Y(x)$. QED

Lemma 2.21. *Let $(\mathcal{X}, d)$ be a metric space. For each $m, n \in \mathbb{N}$, let $a_{mn} \in \mathcal{X}$. Fix $L \in \mathcal{X}$. Suppose that, for every $\epsilon > 0$. there exists $N \in \mathbb{N}$ such that $d(a_{mn}, L) < \epsilon$ whenever $m, n \ge N$. Suppose also that $\lim_{n\to\infty} a_{mn}$ exists for all $m \in \mathbb{N}$. Then $\lim_{m\to\infty} \lim_{n\to\infty} a_{mn}$ exists and equals L.*

Proof. Fix $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that $d(a_{mn}, L) < \epsilon$ whenever $m, n \ge N$. Fix $m \ge N$. Choose $n_m \in \mathbb{N}$ such that $n_m \ge N$ and $d(\lim_{n\to\infty} a_{mn}, a_{mn_m}) < \epsilon$. Then $d(\lim_{n\to\infty} a_{mn}, L) \le d(\lim_{n\to\infty} a_{mn}, a_{mn_m}) + d(a_{mn_m}, L) < 2\epsilon$. QED

Corollary 2.22. *Let $(\mathcal{X}, d)$ be a metric space. For each $m, n \in \mathbb{N}$, let $a_{mn} \in \mathcal{X}$. Fix $L \in \mathcal{X}$. Suppose that, for every $\epsilon > 0$. there exists $N \in \mathbb{N}$ such that $d(a_{mn}, L) < \epsilon$ whenever $m, n \geq N$. Suppose also that $\lim_{n\to\infty} a_{mn}$ and $\lim_{m\to\infty} a_{mn}$ exist for all $m, n \in \mathbb{N}$. Then $\lim_{m\to\infty}\lim_{n\to\infty} a_{mn} = \lim_{n\to\infty}\lim_{m\to\infty} a_{mn} = L$ (so, in particular, we can switch the order of the limits).*

Theorem 2.23. *Let $X_1, X_2, \ldots$ be real-valued random variables, and let X be another real-valued random variable. The following are equivalent:*

(a) $X_n \to X$ in distribution (i.e., $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$ for every bounded, continuous function $f : \mathbb{R} \to \mathbb{R}$).
(b) $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$ for every smooth, compactly supported function $f : \mathbb{R} \to [0, 1]$.
(c) $\mu_{X_n}([a, b]) \to \mu_X([a, b])$ as $n \to \infty$ for every $a, b \in \mathbb{R}$ such that $a \leq b$ and $\mu_X(\{a, b\}) = 0$.
(d) $F_{X_n}(t) \to F_X(t)$ as $n \to \infty$ for every $t \in \mathbb{R}$ at which F_X is continuous.

Proof. Clearly, (a) implies (b). By Theorem 2.12, (b) implies (c).

Suppose (c) holds. Fix any $t \in \mathbb{R}$ at which F_X is continuous. Since F_X has at most countably many discontinuities, we can choose a sequence of strictly decreasing real numbers (a_k) with $a_1 < t$ such that $\lim_{k\to\infty} a_k = -\infty$ and such that F_X is continuous at each a_k. By continuity, $\mu_X(\{a_k, t\}) = 0$ for all $k \in \mathbb{N}$.

We first check that the hypotheses of Corollary 2.22 hold for the "double sequence" $(\mu_{X_n}([a_k, t]))_{n,k\in\mathbb{N}}$, with $L = \mu_X((-\infty, t])$. Clearly, $\lim_{k\to\infty} \mu_{X_n}([a_k, t])$ and $\lim_{n\to\infty} \mu_{X_n}([a_k, t])$ exist for all $n, k \in \mathbb{N}$. Now fix $\epsilon > 0$. Choose $c, d \in \mathbb{R}$ with $c \leq t$ and $c \leq d$ such that $\mu_X([c, d]) \geq 1 - \epsilon$ and such that F_X is continuous at c and d. Choose $N \in \mathbb{N}$ such that, if $n \geq N$, then $|\mu_{X_n}([c, t]) - \mu_X([c, t])| \leq \epsilon$ and $|\mu_{X_n}([c, d]) - \mu_X([c, d])| \leq \epsilon$ (and hence $\mu_{X_n}([c, d]) \geq 1 - 2\epsilon$), and such that $a_k \leq c$ for $k \geq N$. Then if $n, k \geq N$, we have $|\mu_{X_n}([a_k, t]) - \mu_X((-\infty, t])| = |(\mu_{X_n}([a_k, c)) + \mu_{X_n}([c, t])) - (\mu_X((-\infty, c)) + \mu_X([c, t]))| \leq |\mu_{X_n}([c, t]) - \mu_X([c, t])| + |\mu_{X_n}([a_k, c))| + |\mu_X((-\infty, c))| \leq 4\epsilon$, which completes the verification of the hypotheses of Corollary 2.22.

Then we have $F_X(t) = \mu_X((-\infty, t]) = \lim_{k\to\infty} \mu_X([a_k, t]) = \lim_{k\to\infty}\lim_{n\to\infty} \mu_{X_n}([a_k, t]) = \lim_{n\to\infty}\lim_{k\to\infty} \mu_{X_n}([a_k, t]) = \lim_{n\to\infty} \mu_{X_n}((-\infty, t]) = \lim_{n\to\infty} F_{X_n}(t)$, which gives (d).

Suppose (d) holds. Define Y and $Y_1, Y_2, \ldots$ as in Proposition 2.20. Then for any bounded, continuous function $f : \mathbb{R} \to \mathbb{R}$, we have $\mathbb{E}(f(X_n)) = \mathbb{E}(f(Y_n)) \to \mathbb{E}(f(Y)) = \mathbb{E}(f(X))$ as $n \to \infty$ by Proposition 2.8, so (a) holds. QED

2.2. **Law of Large Numbers.** We now arrive at one of the most important theorems in probability: the Law of Large Numbers. Technically, the Law of Large Numbers is not a single theorem but rather a collection of theorems, all of which have slightly different hypotheses and conclusions but express the same general principle. We will prove two of these theorems (Theorems 2.24 and 2.25 below).

To get some intuition for the Law of Large Numbers, consider a physics student who is measuring the time it takes for a ball to drop from a particular height. Suppose that the student performs the same ball-drop experiment n times (where $n \in \mathbb{N}$), obtaining n measurements $X_1, X_2, \ldots, X_n$, each of which is a real-valued random variable. We assume that the ball drops are performed independently of each other; then it is reasonable to assume that the random variables $X_1, X_2, \ldots, X_n$

are independent. It is also reasonable to assume that $X_1, X_2, \dots, X_n$ have the same distribution (and hence the same mean), since the ball drops are meant to be performed in the same way. From experience, we expect that, when n is large, the average $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ is a good estimate for the "true value" of the ball-drop time, since a couple outlier measurements would not skew the average $\frac{1}{n}(X_1+X_2+\cdots+X_n)$ too much for large n. (This intuition is reflected in the proof of Theorem 2.24.) The Law of Large Numbers essentially formalizes this empirical principle that averaging together more data points gives a better estimate for the true value of a quantity (where we will model the "true value" as the expected value), provided that the data points are collected independently of each other.

Theorem 2.24. ***(Weak Law of Large Numbers)*** *Let $X_1, X_2, \dots$ be a sequence of real-valued, pairwise uncorrelated random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with the same mean μ. Suppose that $\mathbb{E}(|X_n|^2) < \infty$ for all $n \in \mathbb{N}$ and that the sequence of variances $(\mathrm{Var}(X_n))$ is bounded. Then $\frac{1}{n}(X_1 + X_2 + \cdots + X_n) \to \mu$ in probability as $n \to \infty$.*

Proof. Choose $\sigma > 0$ such that $\mathrm{Var}(X_n) \leq \sigma^2$ for all $n \in \mathbb{N}$. Fix $\epsilon > 0$. Then by Chebyshev's inequality, we have $\mathbb{P}(|\frac{1}{n}(X_1+X_2+\cdots+X_n)-\mu| > \epsilon) \leq \frac{1}{\epsilon^2}\mathrm{Var}(\frac{1}{n}(X_1+X_2+\cdots+X_n)) = \frac{1}{n^2\epsilon^2}\sum_{j=1}^n \mathrm{Var}(X_j) \leq \frac{n\sigma^2}{n^2\epsilon^2} \to 0$ as $n \to \infty$. QED

Theorem 2.25. ***(Strong Law of Large Numbers)*** *Let $X_1, X_2, \dots$ be a sequence of real-valued, 4-wise independent random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with the same mean μ. Suppose that $\mathbb{E}(|X_n|^4) = \mathbb{E}(X_n^4) < \infty$ for all $n \in \mathbb{N}$ and that the sequence of ordinary 4^{th} moments $(\mathbb{E}(X^4))$ is bounded. Then $\frac{1}{n}(X_1 + X_2 + \cdots + X_n) \to \mu$ a.s. as $n \to \infty$. (Thus, the Strong Law of Large Numbers has a more robust conclusion than the Weak Law of Large Numbers, at the cost of more stringent hypotheses.)*

Proof. Choose $C > 0$ such that $\mathbb{E}(X_n^4) < C$ for all $n \in \mathbb{N}$. We may assume without loss of generality that $\mu = 0$, since, once the $\mu = 0$ case is proven, the general case follows easily by considering the sequence $X_1 - \mu, X_2 - \mu, \dots$.

Fix $n \in \mathbb{N}$. We assert that $\mathbb{E}((X_1 + X_2 + \cdots + X_n)^4) \leq 3Cn^2$. For proof, first observe that $\mathbb{E}((X_1 + X_2 + \cdots + X_n)^4) = \sum_{i,j,k,l \in [n]} \mathbb{E}(X_i X_j X_k X_l)$. Since $\mu = 0$, we observe that, by 4-wise independence, the only nonzero terms in the sum $\sum_{i,j,k,l \in [n]} \mathbb{E}(X_i X_j X_k X_l)$ occur when $i = j = k = l$ or when we can pair up the four indices into two pairs of two indices, such that two indices in the same pair are the same number in $[n]$ whereas two indices in different pairs are different numbers in $[n]$. There are n terms of the first kind, and each of these terms is bounded by C. There are $3n(n-1)$ terms of the second kind, and each of these terms is bounded by C (by Hölder's inequality). Thus, $\sum_{i,j,k,l \in [n]} \mathbb{E}(X_i X_j X_k X_l) \leq nC + 3n(n-1)C \leq 3n^2C$, which proves the asserion.

Let $A_n = \{\frac{1}{n}|X_1 + X_2 + \cdots + X_n| \geq n^{-1/8}\} = \{|X_1 + X_2 + \cdots + X_n| \geq n^{7/8}\}$. Let $\{A_n \ i.o.\}$ be the collection of $\omega \in \Omega$ that are in infinitely many of the A_n. (Indeed, "i.o." stands for "infinitely often.") Choose any $\omega \in \Omega$ such that $\frac{1}{n}(X_1(\omega)+X_2(\omega)+\cdots+X_n(\omega))$ does not converge to 0 as $n \to \infty$. Then there exists $\epsilon > 0$ such that $\frac{1}{n}|X_1(\omega)+X_2(\omega)+\cdots+X_n(\omega)| \geq \epsilon$ for infinitely many $n \in \mathbb{N}$, which implies that $\omega \in \{A_n \ i.o.\}$. Thus, it suffices to show that $\mathbb{P}(A_n \ i.o.) = 0$. By the Borel-Cantelli lemma, it suffices to show that $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$, which is indeed the case by Markov's inequality and the preceding paragraph since $\mathbb{P}(A_n) = \mathbb{P}((X_1 + X_2 + \cdots + X_n)^4 \geq n^{7/2}) \leq \frac{1}{n^{7/2}}\mathbb{E}((X_1 + X_2 + \cdots + X_n)^4) \leq \frac{3Cn^2}{n^{7/2}} = \frac{3C}{n^{3/2}}$. QED

3. Characteristic Functions

3.1. **Basic Properties of Characteristic Functions.** We assume that the reader is familiar with the basic theory of the Fourier transform. However, because there are many conventions for the Fourier transform, we state our convention here:

Definition 3.1. Let μ be a complex measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. We define the **Fourier-Stieltjes transform of μ** as the function $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ given by $\hat{\mu}(\zeta) = \int_{\mathbb{R}^d} e^{ix\cdot\zeta}\, d\mu(x)$ for all $\zeta \in \mathbb{R}^d$. Also, if $f \in L^1((\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, m), \mathbb{C})$, then we define the **Fourier transform of f** as the function $\hat{f} : \mathbb{R}^d \to \mathbb{C}$ given by $\hat{f}(\zeta) = \int_{\mathbb{R}^d} f(x)e^{ix\cdot\zeta}\, dx$ for all $\zeta \in \mathbb{R}^d$. By identifying $f \in L^1((\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, m), \mathbb{C})$ with the complex measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ defined by mapping $A \in \mathcal{B}_{\mathbb{R}^d}$ to $\int_A f(x)\, dx$, we can view the Fourier transform as a special case of the Fourier-Stieltjes transform.

We also recall the well-known inversion theorem from analysis:

Theorem 3.2. ***(Fourier inversion theorem.)*** *Let $f \in L^1((\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, m), \mathbb{C})$. Suppose that $\hat{f} \in L^1((\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, m), \mathbb{C})$ as well. Then $f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{f}(\zeta)e^{-i\zeta\cdot x}\, d\zeta$ for almost all $x \in \mathbb{R}^d$ (with respect to the Lesbegue measure). Furthermore, if f is continuous, then the relation $f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{f}(\zeta)e^{-i\zeta\cdot x}\, d\zeta$ holds for every $x \in \mathbb{R}^d$.*

Corollary 3.3. *Let $f : \mathbb{R}^d \to \mathbb{C}$ be a Schwartz function. Then we have $f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{f}(\zeta)e^{-i\zeta\cdot x}\, d\zeta$ for every $x \in \mathbb{R}^d$.*

Let X be an $\mathbb{R}^d$-valued random variable. We define the **characteristic function of X** by $\phi_X = \widehat{\mu_X}$ (i.e., ϕ_X is the Fourier-Stieltjes transform of the distribution of X). Explicitly, $\phi_X(\zeta) = \mathbb{E}(e^{iX\cdot\zeta}) \in \mathbb{C}$ for all $\zeta \in \mathbb{R}^d$. Note that ϕ_X depends only on the distribution of X.

We now list some basic properties about characteristic functions.

(a) If X has a density f, then $\phi_X = \hat{f}$ is the Fourier transform of the density. Note, however, that ϕ_X always exists, even if X does not have a density.

(b) $\phi_X(0) = 1$, and $|\phi_X(\zeta)| \leq \mathbb{E}(|e^{iX\cdot\zeta}|) = 1$ for all $\zeta \in \mathbb{R}^d$. In particular, ϕ_X is bounded.

(c) If $X_1, X_2, \ldots$ is a sequence of $\mathbb{R}^d$-valued random variables converging to X in distribution, then the characteristic functions ϕ_{X_n} converge pointwise to ϕ_X (since $e^{ix\cdot\zeta}$ is a bounded, continuous function of $x \in \mathbb{R}^d$ for any fixed $\zeta \in \mathbb{R}^d$).

(d) Fix $n \in \mathbb{N}$, and let $X_1, X_2, \ldots, X_n$ be independent, $\mathbb{R}^d$-valued random variables (all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$). We assert that $\phi_{X_1+X_2+\cdots+X_n} = \phi_{X_1}\phi_{X_2}\cdots\phi_{X_n}$ (i.e., the characteristic function of a finite sum of independent $\mathbb{R}^d$-valued random variables is the product of their characteristic functions). Indeed, we have

$$\begin{aligned}\phi_{X_1+X_2+\cdots+X_n}(\zeta) &= \mathbb{E}(e^{i(X_1+X_2+\cdots+X_n)\cdot\zeta}) \\ &= \mathbb{E}(e^{iX_1\cdot\zeta}e^{iX_2\cdot\zeta}\cdots e^{iX_n\cdot\zeta}) \\ &= \mathbb{E}(e^{iX_1\cdot\zeta})\mathbb{E}(e^{iX_2\cdot\zeta})\cdots\mathbb{E}(e^{iX_n\cdot\zeta}) \\ &= \phi_{X_1}(\zeta)\phi_{X_2}(\zeta)\cdots\phi_{X_n}(\zeta)\end{aligned}$$

for all $\zeta \in \mathbb{R}^d$.

(e) If $a \in \mathbb{R}$ and $b \in \mathbb{R}^d$, then $\phi_{aX+b}(\zeta) = \mathbb{E}(e^{i(aX+b)\cdot\zeta}) = \mathbb{E}(e^{iX\cdot(a\zeta)}e^{ib\cdot\zeta}) = e^{ib\cdot\zeta}\phi_X(a\zeta)$ for all $\zeta \in \mathbb{R}^d$.

Proposition 3.4. *Let $X = (X_1, X_2, \ldots, X_d)$ be an $\mathbb{R}^d$-valued random variable. Fix $k \in \mathbb{N}_0$ and $(j_1, j_2, \ldots, j_k) \in [d]^k$, where we interpret $(j_1, j_2, \ldots, j_k)$ as the empty list when $k = 0$. Suppose $\mathbb{E}(|X_{j_k} \cdots X_{j_2} X_{j_1}|) < \infty$, where we interpret the empty product as 1 (so that this hypothesis is trivially satisfied when $k = 0$). Then $\frac{\partial}{\partial\zeta_{j_k}\cdots\partial\zeta_{j_2}\partial\zeta_{j_1}}\phi_X(\zeta) = i^k\mathbb{E}(X_{j_k} \cdots X_{j_2}X_{j_1}e^{iX\cdot\zeta})$ for all $\zeta \in \mathbb{R}^d$, where we interpret a 0^{th}-order partial derivative as the original function, and, furthermore, $\frac{\partial}{\partial\zeta_{j_k}\cdots\partial\zeta_{j_2}\partial\zeta_{j_1}}\phi_X$ is uniformly continuous. In particular, if $\mathbb{E}(|X|^k) < \infty$, then $\phi_X \in C^k(\mathbb{R}^d, \mathbb{C})$ (i.e., all the k^{th}-order partial derivatives of ϕ_X exist on $\mathbb{R}^d$ and are continuous).*

Proof. We prove differentiability by induction on k. The $k = 0$ case is trivial. Now suppose that $k \in \mathbb{N}$ and that the result holds for $k-1$. Since $\mathbb{E}(|X_{j_k} \cdots X_{j_2}X_{j_1}|) < \infty$, we have $\mathbb{E}(|X_{j_{k-1}} \cdots X_{j_2}X_{j_1}|) < \infty$. Hence, we have $\frac{\partial}{\partial\zeta_{j_{k-1}}\cdots\partial\zeta_{j_2}\partial\zeta_{j_1}}\phi_X(\zeta) = i^{k-1}\mathbb{E}(X_{j_{k-1}} \cdots X_{j_2}X_{j_1}e^{iX\cdot\zeta})$ for all $\zeta \in \mathbb{R}^d$. Let $\psi = \frac{\partial}{\partial\zeta_{j_{k-1}}\cdots\partial\zeta_{j_2}\partial\zeta_{j_1}}\phi_X$. Let u_{j_k} be the standard basis vector in $\mathbb{R}^d$ with a 1 in the j_k^{th} coordinate and zeroes in the other coordinates. Then for $h \in \mathbb{R} \setminus \{0\}$, we have $\frac{1}{h}(\psi(\zeta + hu_{j_k}) - \psi(\zeta)) = i^{k-1}\mathbb{E}(X_{j_{k-1}} \cdots X_{j_2}X_{j_1}e^{iX\cdot\zeta}(\frac{e^{iX_{j_k}h}-1}{h}))$. Since $|e^{it} - 1| \leq |t|$ for any $t \in \mathbb{R}$, we get $|X_{j_{k-1}} \cdots X_{j_2}X_{j_1}e^{iX\cdot\zeta}(\frac{e^{iX_{j_k}h}-1}{h})| \leq |X_{j_k}X_{j_{k-1}} \cdots X_{j_2}X_{j_1}|$, so taking $h \to 0$ and using the dominated convergence theorem gives $\frac{1}{h}(\psi(\zeta + hu_{j_k}) - \psi(\zeta)) \to i^k\mathbb{E}(X_{j_k}X_{j_{k-1}} \cdots X_{j_2}X_{j_1}e^{iX\cdot\zeta})$, as desired.

For uniform continuity, fix $s \in \mathbb{R}^d$. Let $\gamma = \frac{\partial}{\partial\zeta_{j_k}\cdots\partial\zeta_{j_2}\partial\zeta_{j_1}}\phi_X$. We define $\gamma_s(\zeta) = \gamma(\zeta + s)$ for all $\zeta \in \mathbb{R}^d$. We must show that $||\gamma_s - \gamma||_{\sup} \to 0$ as $s \to 0$ (where $||\cdot||_{\sup}$ indicates the supremum norm). For each $\zeta \in \mathbb{R}^d$, we have $|\gamma(\zeta + s) - \gamma| \leq \mathbb{E}(|X_{j_k} \cdots X_{j_2}X_{j_1}|(|e^{iX\cdot s}-1|))$, so $||\gamma_s-\gamma||_{\sup} \leq \mathbb{E}(|X_{j_k} \cdots X_{j_2}X_{j_1}|(|e^{iX\cdot s}-1|)) \to 0$ as $s \to 0$ by the dominated convergence theorem. QED

Corollary 3.5. *ϕ_X is uniformly continuous for any $\mathbb{R}^d$-valued random variable X.*

We have already seen that the distribution of an $\mathbb{R}^d$-valued random variable X determines the characteristic function ϕ_X. We now show that the (highly nontrivial) converse holds in the special case that X is real-valued; i.e., the characteristic function of a real-valued random variable completely determines its distribution. This will follow from Theorem 3.6 below, which allows us to reduce questions about convergence in distribution into questions about convergence of characteristic functions (and also adds to the list of equivalences in Theorem 2.23):

Theorem 3.6. *Let $X_1, X_2, \ldots$ be real-valued random variables, and let X be another real-valued random variable. The following are equivalent:*

(a) $X_n \to X$ in distribution as $n \to \infty$.
(b) The sequence of characteristic functions (ϕ_{X_n}) converges pointwise to ϕ_X.
(c) The sequence of characteristic functions (ϕ_{X_n}) converges to ϕ_X a.e. (with respect to Lesbegue measure).

Proof. Clearly, (a) implies (b), and (b) implies (c). Now suppose that (c) holds. To show that (a) holds, it suffices by Theorem 2.23 to prove that $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ for every smooth, compactly supported function $f : \mathbb{R} \to [0, 1]$. Note that any such

f is Schwartz, so, by Corollary 3.3 and Fubini's theorem, we have $\mathbb{E}(f(X_n)) = \int_{\mathbb{R}} f(x)\, d\mu_{X_n}(x) = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \hat{f}(\zeta)e^{-i\zeta x}\, d\zeta d\mu_{X_n}(x) = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \hat{f}(\zeta)e^{-i\zeta x}\, d\mu_{X_n}(x)d\zeta = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\zeta)\phi_{X_n}(-\zeta)d\zeta$. (Note the importance of the fact that the relation $f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\zeta)e^{-i\zeta x}\, dx$ holds for every $x \in \mathbb{R}^d$, not just almost every x.) Similarly, $\mathbb{E}(f(X)) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\zeta)\phi_X(-\zeta)d\zeta$. Thus, $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ as $n \to \infty$ by the dominated convergence theorem. QED

Corollary 3.7. *Two real-valued random variables agree in distribution if and only if their characteristic functions are equal (which, by Corollary 3.5, occurs if and only if their characteristic functions are equal a.e. with respect to Lesbegue measure).*

3.2. **Normal Random Variables.** We now consider **normal random variables**, which are of fundamental importance to probability theory.

Definition 3.8. Let $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R} \setminus \{0\}$. We define $N(\mu, \sigma^2)$ as the probability distribution on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ with density $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$ for $x \in \mathbb{R}$. (Using the famous identity $\int_{-\infty}^{\infty} e^{-\pi x^2} dx = 1$, it is easy to show that $\int_{-\infty}^{\infty} f(x)\, dx = 1$, so $N(\mu, \sigma^2)$ is well-defined.)

Proposition 3.9. *Let $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R} \setminus \{0\}$. Also, let $a \in \mathbb{R} \setminus \{0\}$ and $b \in \mathbb{R}$, and let X be a real-valued random variable with distribution $N(\mu, \sigma^2)$. Then $aX + b$ has distribution $N(a\mu + b, (a\sigma)^2)$.*

Proof. It suffices to show that aX has distribution $N(a\mu, (a\sigma)^2)$ and that $X + b$ has distribution $N(\mu + b, \sigma^2)$. Fix $c, d \in \mathbb{R}$ with $c \le d$. Then $\mu_{X+b}((c, d]) = \mu_X((c-b, d-b]) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{c-b}^{d-b} e^{-(x-\mu)^2/(2\sigma)^2}\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_c^d e^{-(x-(b+\mu))^2/(2\sigma)^2}\, dx = (N(\mu+b, \sigma^2))((c, d])$, so $\mu_{X+b} = N(\mu+b, \sigma^2)$ by Proposition 1.1 (since the collection of half-open intervals of the form $(c, d]$ with $c \le d$ form a π-system).

We observe that $\mu_{-X}((c, d]) = \mu_X([-d, -c)) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-d}^{-c} e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_c^d e^{-(x-(-\mu))^2/(2\sigma^2)}\, dx = (N(-\mu, \sigma^2))((c, d])$, so $\mu_{-X} = N(-\mu, \sigma^2)$, again by Proposition 1.1. Thus, to prove that aX has distribution $N(a\mu, (a\sigma)^2)$, we may assume without loss of generality that $a > 0$, since the general case follows from this special case and the preceding sentence. Then we observe that $\mu_{aX}((c, d]) = \mu_X((c/a, d/a]) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{c/a}^{d/a} e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi(a\sigma)^2}} \int_c^d e^{-(x-a\mu)^2/(2(a\sigma)^2)}\, dx = (N(a\mu, (a\sigma)^2))((c, d])$, so $\mu_{aX} = N(a\mu, (a\sigma)^2)$, yet again by Proposition 1.1 QED

We call $N(0, 1)$ the **standard normal distribution**, so $N(0, 1)$ has density $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for $x \in \mathbb{R}$. Let X be a **standard normal random variable** (i.e., a real-valued random variable with $\mu_X = N(0, 1)$). Since the characteristic function of X is the Fourier transform of f, we have $\phi_X(t) = e^{-t^2/2}$ for all $t \in \mathbb{R}$.[6] Thus, if X is a real-valued random variable with distribution $N(\mu, \sigma^2)$ for $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R} \setminus \{0\}$, then $\phi_X(t) = e^{i\mu t - (\sigma t)^2/2}$ for all $t \in \mathbb{R}$, because, by Proposition 3.9, $\frac{1}{\sigma}X - \frac{\mu}{\sigma}$ is a standard normal random variable, so $\phi_X(t) = \phi_{\sigma(\frac{1}{\sigma}X - \frac{\mu}{\sigma}) + \mu}(t) = e^{i\mu t}\phi_{\frac{1}{\sigma}X - \frac{\mu}{\sigma}}(\sigma t) = e^{i\mu t} e^{-(\sigma t)^2/2}$ for all $t \in \mathbb{R}$.

[6]The Fourier transform of f (where $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for $x \in \mathbb{R}$) is a well-known result from analysis; see Proposition 16.5 of [3] for details.

We now ask a seemingly simple question: if X is a real-valued random variable with distribution $N(\mu, \sigma^2)$ for $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R} \setminus \{0\}$, then what is the mean and variance of X? (It is not hard to show that $\mathbb{E}(|X|^n) < \infty$ for all $n \in \mathbb{N}_0$, so the mean and variance of X exist.) It is not too difficult to directly compute that $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$; for this reason, $N(\mu, \sigma^2)$ is often called the **normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma^2}$**. However, the direct proof of these facts is somewhat messy, so we give a more slick proof using Proposition 3.4. For $t \in \mathbb{R}$, since $\phi_X(t) = e^{i\mu t - (\sigma t)^2/2}$, we compute $\phi'(t) = (i\mu - \sigma^2 t)e^{i\mu t-(\sigma t)^2/2}$ and $\phi''_X(t) = (i\mu - \sigma^2 t)^2 e^{i\mu t-(\sigma t)^2/2} - \sigma^2 e^{i\mu t-(\sigma t)^2/2}$. Then by evaluating at $t = 0$ and using Proposition 3.4, we immediately obtain $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

The preceding paragraph demonstrates some of the power of characteristic functions. The next result further demonstrates their efficacy.

Theorem 3.10. *Fix $n \in \mathbb{N}$, and let $X_1, X_2, \ldots, X_n$ be independent, real-valued random variables (all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$). For each $j \in [n]$, suppose that X_j has distribution $N(\mu_j, \sigma_j^2)$ for some $\mu_j \in \mathbb{R}$ and $\sigma \in \mathbb{R}\setminus\{0\}$. Then $X_1 + X_2 + \cdots + X_n$ has distribution $N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$.*

Proof. For $t \in \mathbb{R}$, we observe that $\phi_{X_1+X_2+\cdots+X_n}(t) = \phi_{X_1}(t)\phi_{X_2}(t)\cdots\phi_{X_n}(t) = e^{i\mu_1 t-(\sigma_1 t)^2/2} e^{i\mu_2 t-(\sigma_2 t)^2/2} \cdots e^{i\mu_n t-(\sigma_n t)^2/2} = e^{i(\mu_1+\mu_2+\cdots+\mu_n)t-(\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2)t^2/2}$, so the result follows from Corollary 3.7. QED

The following result is a simple but helpful estimate.

Proposition 3.11. *Let X be a standard normal random variable. Then for all $a > 0$, we have $\mu_X([a, \infty)) \leq \frac{1}{a}\frac{1}{\sqrt{2\pi}} e^{-a^2/2}$.*

Proof. Since $\frac{x}{a} \geq 1$ whenever $x \geq a$, we have $\mu_X([a,\infty)) = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-x^2/2}\, dx \leq \frac{1}{\sqrt{2\pi}} \int_a^\infty \frac{x}{a} e^{-x^2/2}\, dx = \frac{1}{a}\frac{1}{\sqrt{2\pi}} \int_{a^2/2}^\infty e^{-x} dx = \frac{1}{a}\frac{1}{\sqrt{2\pi}} e^{-a^2/2}$. QED

Next, we discuss normal distributions in higher dimensions.

Definition 3.12. We define the ***d*-dimensional standard normal distribution** as the probability distribution on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ with density $f : \mathbb{R}^d \to \mathbb{R}$ given by $f(x) = \frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2} = \prod_{j=1}^d \frac{1}{\sqrt{2\pi}} e^{-x_j^2/2}$ for $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$. (Using Tonelli's theorem, it is easy to show that $\int_{\mathbb{R}^d} f(x)\, dx = 1$, so this definition makes sense.) Note that the 1-dimensional standard normal distribution is $N(0, 1)$, so this definition is consistent with our prior terminology. Also note that, if we say "standard normal distribution" (without specifying the dimension), we always mean "1-dimensional standard normal distribution."

Definition 3.13. A ***d*-dimensional standard normal random variable** is an $\mathbb{R}^d$-valued random variable X whose distribution μ_X is the d-dimensional standard normal distribution. Note that, if we say "standard normal random variable," then we mean "1-dimensional standard normal random variable."

The next result (3.14) is quite useful; its proof is easy and hence omitted.

Proposition 3.14. *An $\mathbb{R}^d$-valued random variable $X = (X_1, X_2, \ldots, X_d)$ is a d-dimensional standard normal random variable if and only if $X_1, X_2, \ldots, X_d$ are independent, standard normal random variables.*

Proposition 3.15. *Let X be a d-dimensional standard normal random variable, and let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a linear isometry (i.e., $T : \mathbb{R}^d \to \mathbb{R}^d$ is a linear map such that $|Tx| = |x|$ for all $x \in \mathbb{R}^d$). Then TX is a d-dimensional standard normal random variable.*

Proof. We know from linear algebra that T is invertible, that T^{-1} is a linear isometry, and that $|\det(T^{-1})| = 1$. Thus, if V is an open set in $\mathbb{R}^d$, we see (by the change of variables theorem) that $\mu_{TX}(V) = \mu_X(T^{-1}(V)) = \frac{1}{(2\pi)^{d/2}} \int_{T^{-1}(V)} e^{-|x|^2/2}\,dx = \frac{1}{(2\pi)^{d/2}} \int_V e^{|T^{-1}(x)|^2/2}\,dx = \frac{1}{(2\pi)^{d/2}} \int_V e^{|x|^2/2}\,dx$. Since the map sending $B \in \mathcal{B}_{\mathbb{R}^d}$ to $\frac{1}{(2\pi)^{d/2}} \int_B e^{|x|^2/2}\,dx$ is a probability measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, we conclude using Proposition 1.1 that the function $f : \mathbb{R}^d \to \mathbb{R}$ given by $f(x) = \frac{1}{(2\pi)^{d/2}} e^{-|x|^2/2}$ for $x \in \mathbb{R}^d$ is a density of TX. QED

Corollary 3.16. *Let X and Y be independent real-valued random variables (defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$) such that $\mu_X = \mu_Y = N(0, \sigma^2)$ for some $\sigma \in \mathbb{R} \setminus \{0\}$. Then $X + Y$ and $X - Y$ are independent real-valued random variables such that $\mu_{X+Y} = \mu_{X-Y} = N(0, 2\sigma^2)$.*

Proof. By Propositions 3.9 and 3.14, it suffices to show that $(\frac{X+Y}{\sqrt{2}\sigma}, \frac{X-Y}{\sqrt{2}\sigma})$ is a 2-dimensional standard normal random variable. Since $(\frac{X}{\sigma}, \frac{Y}{\sigma})$ is a 2-dimensional standard normal random variable, and since the map sending $(x, y) \in \mathbb{R}^2$ to $(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}})$ is a linear isometry, Proposition 3.15 gives the result. QED

Definition 3.17. Let $X = (X_1, X_2, \ldots, X_d)$ be an $\mathbb{R}^d$-valued random variable. Suppose that $\mathbb{E}(|X|^2) < \infty$ (or, equivalently, $\mathbb{E}(|X_j|^2) < \infty$ for all $j \in [d]$). We define the **covariance matrix of X** as the d-by-d matrix whose entry in row i, column j is $\mathrm{Cov}(X_i, X_j)$, where $i, j \in [d]$. We let $\mathrm{Cov}(X)$ denote the covariance matrix of X.

Let Y be an $\mathbb{R}^d$-valued random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that Y is a ***d*-dimensional normal random variable** if, for some $k \in \mathbb{N}$, we have $Y = AX + b$ for some k-dimensional standard normal random variable X (defined on $(\Omega, \mathcal{F}, \mathbb{P})$), some linear map $A : \mathbb{R}^k \to \mathbb{R}^d$ (which we can view as a d-by-k matrix with respect to the standard bases of $\mathbb{R}^k$ and $\mathbb{R}^d$), and some constant vector $b \in \mathbb{R}^d$. In this case, we have $\mathbb{E}(|Y|^2) < \infty$, and it is not hard to show that $\mathbb{E}(Y) = b$ and $\mathrm{Cov}(Y) = AA^T$. Also, any d-dimensional standard normal random variable is a d-dimensional normal random variable, and, if Y is a real-valued random variable with $\mu_Y = N(\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R} \setminus \{0\}$, then Y is a 1-dimensional normal random variable.

We next show (in Theorem 3.19) that the distribution of a d-dimensional normal random variable is uniquely determined by its expectation and covariance matrix. We first state a lemma.

Lemma 3.18. *Let Y be a d-dimensional normal random variable with $\mathbb{E}(Y) = 0$, which means we can choose $k \in \mathbb{N}$ such that $Y = AX$ for some k-dimensional standard normal random variable X (defined on the same probability space as Y) and some linear map $A : \mathbb{R}^k \to \mathbb{R}^d$. Fix $l \in \mathbb{N}$ with $l \geq k$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space such that there exists an l-dimensional standard normal random variable Z on $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mu_Y = \mu_{BZ}$ for some linear map $B : \mathbb{R}^l \to \mathbb{R}^d$.*

Proof. Define B as the matrix obtained from A by adjoining $l-k$ columns of zeroes to the end of it. The result then follows easily by checking equality on measurable rectangles and using Proposition 1.1. QED

Theorem 3.19. *Let X and Y be independent d-dimensional normal random variables (not necessarily defined on the same sample space) such that $\mathbb{E}(X) = \mathbb{E}(Y)$ and $\mathrm{Cov}(X) = \mathrm{Cov}(Y)$. Then $\mu_X = \mu_Y$.*

Proof. It suffices to prove this when $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, since the general case follows easily from this special case. Also, by Lemma 3.18, we can assume without loss of generality that X and Y are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, that $X = B_1Z$ for some k-dimensional standard normal random variable Z (where $k \in \mathbb{N}$) and linear map $B_1 : \mathbb{R}^k \to \mathbb{R}^d$, and that $Y = B_2Z$ for some linear map $B_2 : \mathbb{R}^k \to \mathbb{R}^d$. Since $\mathrm{Cov}(X) = \mathrm{Cov}(Y)$, we have $B_1B_1^T = B_2B_2^T$. Also, by Propositions 1.2 and 3.15, we just need to show that $B_1 = B_2U$ for some linear isometry $U : \mathbb{R}^k \to \mathbb{R}^k$.

Let V_1 be the row space of B_1, and let V_2 be the row space of B_2. Also, let $v_1, v_2, \ldots, v_d \in \mathbb{R}^k$ be the rows of B_1, and let $w_1, w_2, \ldots, w_d \in \mathbb{R}^k$ be the rows of B_2. Finally, let $v_{j_1}, v_{j_2}, \ldots, v_{j_t}$, where $t \in [d]$ and $j_1 < j_2 < \cdots < j_t$, be a basis of V_1. Define a linear map $U : V_1 \to V_2$ by $U(v_{j_i}) = w_{j_i}$ for all $i \in [l]$. The relation $B_1B_1^T = B_2B_2^T$ implies that U preserves inner products between the basis vectors $v_{j_1}, v_{j_2}, \ldots, v_{j_t}$, which implies that U preserves inner products between any vectors in V_1 and hence is a linear isometry. In particular, U is injective, so $\dim(V_1) \le \dim(V_2)$. A symmetric argument yields $\dim(V_2) \ge \dim(V_1)$, so $\dim(V_1) = \dim(V_2) = t$.

Let $e_1, e_2, \ldots, e_{k-t}$ be an orthonormal basis of $V_1^\perp$ (the orthogonal complement of V_1), and let $f_1, f_2, \ldots, f_{k-t}$ be an orthonormal basis of $V_2^\perp$. Linearly extend U to have domain $\mathbb{R}^k$ by setting $U(e_i) = f_i$ for each $i \in [k-t]$. Then $U : \mathbb{R}^k \to \mathbb{R}^k$ is still a linear isometry. Note that $V_1^\perp = \ker(B_1)$ and $V_2^\perp = \ker(B_2)$. Using these facts along with the relation $B_1B_1^T = B_2B_2^T$ yields the relation $B_1 = B_2U$ (which can be proven by checking equality on the basis $v_{j_1}, v_{j_2}, \ldots, v_{j_t}, e_1, e_2, \ldots, e_{k-t}$), thus completing the proof. QED

Corollary 3.20. *Let $Y = (Y_1, Y_2, \ldots, Y_d)$ be a d-dimensional normal random variable. Then $Y_1, Y_2, \ldots, Y_d$ are independent if and only if they are pairwise uncorrelated (i.e., $\mathrm{Cov}(Y)$ is a diagonal matrix).*

Proof. One implication is immediate. For the converse, let $X = (X_1, X_2, \ldots, X_d)$ be a d-dimensional standard normal random variable. Let $W = (W_1, W_2, \ldots, W_d) = (\sqrt{\mathrm{Var}(Y_1)}X_1, \sqrt{\mathrm{Var}(Y_2)}X_2, \ldots, \sqrt{\mathrm{Var}(Y_d)}X_d) + \mathbb{E}(Y)$. We have $\mathbb{E}(W) = \mathbb{E}(Y)$ and $\mathrm{Cov}(W) = \mathrm{Cov}(Y)$, so, by Theorem 3.19, we have $\mu_Y = \mu_W = \mu_{W_1} \times \mu_{W_2} \times \cdots \times \mu_{W_d} = \mu_{Y_1} \times \mu_{Y_2} \times \cdots \times \mu_{Y_d}$ (where the last equality uses Proposition 1.2). QED

3.3. **Central Limit Theorem.** We now arrive at another one of the most important theorems in probability: the Central Limit Theorem. Like the Law of Large Numbers, the Central Limit Theorem is technically not a single theorem but rather a collection of similar theorems that vary slightly in their hypotheses and conclusions but express the same general principle. We will prove one of these theorems (Theorem 3.21).

To give some intuition for the Central Limit Theorem, recall the physics student who is measuring the time it takes for a ball to drop from a particular height. We again suppose that the student performs the same ball-drop experiment n times (where $n \in \mathbb{N}$), obtaining n independent measurements $X_1, X_2, \ldots, X_n$, each of which is a real-valued random variable, and all of which have the same distribution. By normalizing the random variables appropriately, we can assume without losing much generality that $X_1, X_2, \ldots, X_n$ all have mean zero and variance one. By the Law of Large Numbers, we know that $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ converges in probability, and hence in distribution, to the mean 0. However, one may wonder about the "shape" of the distribution of $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ for large n. If we do not scale $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ at all, then the distribution will just look like that of δ_0, so we have to scale $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ by some (nonconstant) function of n to obtain any new insight into the shape. It turns out that, if we scale $\frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ by $\sqrt{n}$ (so that we are considering the random variables $\frac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)$), then, amazingly, the distribution of $\frac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)$ will look like a standard normal distribution when n is large! (In particular, there is no dependence on the distributions of the μ_{X_j}, besides the requirements that $X_1, X_2, \ldots, X_n$ are independent and that $\mu_{X_1} = \mu_{X_2} = \cdots = \mu_{X_n}$.) This is the essence of the Central Limit Theorem, which we now prove using the machinery of characteristic functions (and hence using the machinery of Fourier analysis, since Theorem 3.6 was proven using Corollary 3.3):

Theorem 3.21. ***(Central Limit Theorem.)*** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean 0 and variance 1. Then $\frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n)$ converges in distribution to a standard normal random variable as $n \to \infty$.*

Proof. Fix $t \in \mathbb{R}$. Let $\phi = \phi_{X_1} = \phi_{X_2} = \cdots$. By Theorem 3.6, it suffices to show that $\phi_{\frac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)}(t) \to e^{-t^2/2}$ as $n \to \infty$. Note that $\phi_{\frac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)}(t) = \phi_{X_1+X_2+\cdots+X_n}(\frac{t}{\sqrt{n}}) = \phi(\frac{t}{\sqrt{n}})^n$. By Taylor's theorem (which is justified by Proposition 3.4), $\phi(t) = 1 - \frac{t^2}{2} + R(t)$, where $R(t)$ is the remainder satisfying $\frac{R(t)}{t^2} \to 0$ as $t \to 0$. In particular, $\frac{1}{t^2}(nR(\frac{t}{\sqrt{n}})) = \frac{R(t/\sqrt{n})}{(t/\sqrt{n})^2} \to 0$ as $n \to \infty$. Thus, $\phi_{\frac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)}(t) = \phi(\frac{t}{\sqrt{n}})^n = (1 - \frac{(t/\sqrt{n})^2}{2} + R(\frac{t}{\sqrt{n}}))^n = (1 + \frac{-t^2/2 + nR(t/\sqrt{n})}{n})^n \to e^{-t^2/2}$ as $n \to \infty$.[7] QED

4. Brownian Motion

Throughout this section, fix $\mathcal{T} \subset [0, \infty)$ such that $0 \in \mathcal{T}$, and fix $x_0 \in \mathbb{R}^d$, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R} \setminus \{0\}$. We let $C(\mathcal{T}, \mathbb{R}^d)$ denote the collection of continuous functions with domain $\mathcal{T}$ and codomain $\mathbb{R}^d$. For $t \in \mathbb{R}$ and $B \in \mathcal{B}_{\mathbb{R}^d}$, we call $C_{t,B} = \{f \in C(\mathcal{T}, \mathbb{R}^d) : f(t) \in B\}$ the **cylinder set** associated with t and B, and we let $C(\mathcal{T}, \mathbb{R}^d)_{x_0} = C_{0,\{x_0\}}$. We also let $\mathcal{C}_{\mathcal{T},\mathbb{R}^d}$ be the σ-algebra in $C(\mathcal{T}, \mathbb{R}^d)$ generated by all cylinder sets (i.e., $\mathcal{C}_{\mathcal{T},\mathbb{R}}$ is generated by $\{C_{t,B} : t \in \mathcal{T}, B \in \mathcal{B}_{\mathbb{R}^d}\}$). (This is analogous to the definition of the product σ-algebra.) Henceforth, we think of $C(\mathcal{T}, \mathbb{R}^d)$ as a measurable space equipped with the σ-algebra $\mathcal{C}_{\mathcal{T},\mathbb{R}^d}$.

Let X be a $C(\mathcal{T}, \mathbb{R}^d)$-valued random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each $t \in \mathcal{T}$, let $X_t(\omega) = (X(\omega))(t)$ for all $\omega \in \Omega$. Observe that X_t is an $\mathbb{R}^d$-valued measurable function on $(\Omega, \mathcal{F}, \mathbb{P})$ (and hence an $\mathbb{R}^d$-valued random

[7] We have used the well-known fact that, if (c_n) is a sequence of complex numbers converging to $c \in \mathbb{C}$, then $(1 + \frac{c_n}{n})^n \to e^c$ as $n \to \infty$. See Theorem 3.4.2 of [6] for a proof of this fact.

variable). In this way, we can represent the random variable X as a collection of $\mathbb{R}^d$-valued random variables $(X_t)_{t\in\mathcal{T}}$, such that the map sending $t \in \mathcal{T}$ to $X_t(\omega)$ is continuous for all $\omega \in \Omega$. Conversely, if $(X_t)_{t\in\mathcal{T}}$ is a collection of $\mathbb{R}^d$-valued random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and if the map sending $t \in \mathcal{T}$ to $X_t(\omega)$ is continuous for all $\omega \in \Omega$, then we can consider $(X_t)_{t\in\mathcal{T}}$ as a ($\mathcal{C}_{\mathcal{T},\mathbb{R}^d}$-measurable) $C(\mathcal{T}, \mathbb{R}^d)$-valued random variable, where $(X_t)_{t\in\mathcal{T}}(\omega)$ is the continuous function sending $t \in \mathcal{T}$ to $X_t(\omega)$ for all $\omega \in \Omega$. Thus, it is common to represent $C(\mathcal{T}, \mathbb{R}^d)$-valued random variables as collections of $\mathbb{R}^d$-valued random variables $(X_t)_{t\in\mathcal{T}}$, all defined on the same probability space, satisfying the above continuity condition (i.e., that the map sending $t \in \mathcal{T}$ to $X_t(\omega)$ is continuous for all $\omega \in \Omega$).

Note that $d = 1$ for the first two subsections.

4.1. **Existence of Brownian Motion.** We say that a $C(\mathcal{T}, \mathbb{R})_{x_0}$-valued random variable $(X_t)_{t\in\mathcal{T}}$ is a **(one-dimensional) Wiener process on the set of times $\boldsymbol{\mathcal{T}}$ with starting point $\boldsymbol{x_0}$, drift $\boldsymbol{\mu}$, and variance $\boldsymbol{\sigma^2}$** if the following hold:

(a) **(Normally distributed increments)** If $s, t \in \mathcal{T}$ and $s < t$, then $X_t - X_s$ has distribution $N(\mu(t-s), \sigma^2(t-s))$. (In particular, X_t has distribution $N(t\mu + x_0, \sigma^2 t)$ for all $t \in \mathcal{T} \setminus \{0\}$.)

(b) **(Independent increments)** If $n \in \mathbb{N}$, and if $t_0, t_1, \ldots, t_n \in \mathcal{T}$ satisfy $t_0 < t_1 < \cdots < t_n$, then $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent. (Hence, if $n \in \mathbb{N}$, and if $t_0, t_1, \ldots, t_n \in \mathcal{T}$ satisfy $t_0 \le t_1 \le \cdots \le t_n$, then $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.)

In the case that $x_0 = 0$, that $\mu = 0$, and that $\sigma^2 = 1$, we say that this Wiener process $(X_t)_{t\in\mathcal{T}}$ is **standard**. In the case that $\mathcal{T}$ is an interval I, we say that this Wiener process $(X_t)_{t\in I}$ is a **(one-dimensional) Brownian motion on $\boldsymbol{I}$** (with starting point x_0, drift μ, and variance σ^2). Brownian motion models random continuous motion, and it is an important concept both in mathematics and in many fields that use mathematics, like physics (for, e.g., modeling the motion of a particle in a cloud of dust) and finance (for, e.g., modeling the fluctuations of stock prices).

Let $(X_t)_{t\in\mathcal{T}}$ be a Wiener process with starting point x_0, drift μ, and variance σ^2. The following facts are easy to prove:

(a) If $0 \in \mathcal{S} \subset \mathcal{T}$, then $(X_t)_{t\in\mathcal{S}}$ is a Wiener process with starting point x_0, drift μ, and variance σ^2.

(b) If $a \in \mathbb{R} \setminus \{0\}$, then $(aX_t)_{t\in\mathcal{T}}$ is a Wiener process with starting point ax_0, drift $a\mu$, and variance $(a\sigma)^2$.

(c) If $b \in \mathbb{R}$, then $(X_t + b)_{t\in\mathcal{T}}$ is a Wiener process with starting point $x_0 + b$, drift μ, and variance σ^2.

(d) If $d \in \mathbb{R}$, then $(X_t + dt)_{t\in\mathcal{T}}$ is a Wiener process with starting point x_0, drift $\mu + d$, and variance σ^2.

Hence, if $(B_t)_{t\in[0,\infty)}$ is a standard Brownian motion on $[0, \infty)$, then $(\sigma B_t + \mu t + x_0)_{t\in T}$ is a Wiener process with starting point x_0, drift μ, and variance σ^2.

We now show that there exists a Wiener process $(X_t)_{t\in\mathcal{T}}$ with starting point x_0, drift μ, and variance σ^2. By the preceding paragraph, it suffices to establish the existence of a standard Brownian motion $(B_t)_{t\in[0,\infty)}$ on $[0, \infty)$.

For each $n \in \mathbb{N}_0$, let $\mathcal{D}_n = \{\frac{k}{2^n} : k \in \mathbb{N}_0\}$. Note that $\mathbb{N}_0 = \mathcal{D}_0 \subset \mathcal{D}_1 \subset \mathcal{D}_2 \subset \cdots$. The union $\mathcal{D} = \bigcup_{n=0}^{\infty} \mathcal{D}_n$ is known as the set of **(nonnegative) dyadic rationals**;

observe that $\mathcal{D}$ is a countable dense subset of $[0, \infty)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be any probability space such that there exists a countable collection $\{Z_t\}_{t \in \mathcal{D}}$ of independent standard normal random variables, all defined on $(\Omega, \mathcal{F}, \mathbb{P})$. (Such a probability space exists by Theorem 1.3.) We now inductively define a collection of real-valued random variables $(B_t)_{t \in \mathcal{D}}$ by defining B_t first for $t \in \mathcal{D}_0$, then for $t \in \mathcal{D}_1 \setminus \mathcal{D}_0$, then for $t \in \mathcal{D}_2 \setminus \mathcal{D}_1$, and so forth, such that, for each $n \in \mathbb{N}_0$, the $C(\mathcal{D}_n, \mathbb{R})$-valued random variable $(B_t)_{t \in \mathcal{D}_n}$ is a standard Wiener process.

For $k \in \mathbb{N}_0$, define $B_k = \sum_{j=1}^k Z_j$. (Hence, $B_0 = 0$, because the empty sum is zero.) Then it is easy to check using Theorem 3.10 that $(B_t)_{t \in \mathcal{D}_0}$ is a standard Wiener process.

Now fix $n \in \mathbb{N}$, and suppose that we have defined $(B_t)_{t \in \mathcal{D}_{n-1}}$ such that, when viewed as a $C(\mathcal{D}_{n-1}, \mathbb{R})$-valued random variable, $(B_t)_{t \in \mathcal{D}_{n-1}}$ is a standard Wiener process. For $t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$, we observe that $t - 2^{-n}, t + 2^{-n} \in \mathcal{D}_{n-1}$, so we can define $B_t = \frac{B_{t-2^{-n}} + B_{t+2^{-n}}}{2} + 2^{-(n+1)/2} Z_t$. (Thus, we are linearly interpolating the values of $B_{t-2^{-n}}$ and $B_{t+2^{-n}}$, and then adding in $2^{-(n+1)/2} Z_t$, so that, informally speaking, the interpolation itself has some randomness associated with it that is independent of the randomness from $B_{t-2^{-n}}$ and $B_{t+2^{-n}}$.) We now show that $(B_t)_{t \in \mathcal{D}_n}$ is a standard Wiener process. (Note that the continuity condition here is trivial since $\mathcal{D}_n$ has the discrete topology.)

Fix $m \in \mathbb{N}$. Consider the random vector $(B_t - B_{t-2^{-n}} : t \in [m2^n])$. Using induction on n, we observe that each entry in this random vector is a linear combination of the entries of $(Z_t : t \in \mathcal{D}_n)$, so $(B_t - B_{t-2^{-n}} : t \in [m2^n])$ is a $(m2^n)$-dimensional normal random variable. Thus, by Corollary 3.20, to show that the collection $\{B_t - B_{t-2^{-n}} : t \in [m2^n]\}$ is independent, it suffices to show that it is pairwise independent.[8]

Fix $t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$. It is easy to show that $B_t - B_{t-2^{-n}} = \frac{B_{t+2^{-n}} - B_{t-2^{-n}}}{2} + 2^{-(n+1)/2} Z_t$ and $B_{t+2^{-n}} - B_t = \frac{B_{t+2^{-n}} - B_{t-2^{-n}}}{2} - 2^{-(n+1)/2} Z_t$. Using Proposition 3.9 and Corollary 3.16, we see that $B_t - B_{t-2^{-n}}$ and $B_{t+2^{-n}} - B_t$ are independent random variables such that $\mu_{B_t - B_{t-2^{-n}}} = \mu_{B_{t+2^{-n}} - B_t} = N(0, 2^{-n})$.

Fix $j, j' \in [m2^n]$, with $j < j'$. Suppose that j is even or that $j' \neq j + 1$ (or both). Then there exists $s, s' \in \mathcal{D}_{n-1} \cap (0, m)$ with $s \leq s'$ such that $[\frac{j-1}{2^n}, \frac{j}{2^k}] \subset [s - 2^{-(n-1)}, s]$ and $[\frac{j'-1}{2^n}, \frac{j'}{2^n}] \subset [s', s' + 2^{-(n-1)}]$. Using the increment relations from the preceding paragraph, we observe that $B_{\frac{j}{2^k}} - B_{\frac{j-1}{2^n}}$ can be written as a linear combination of $B_s - B_{s-2^{-(n-1)}}$ and $Z_{s-2^{-n}}$. Similarly, we observe that $B_{\frac{j'}{2^n}} - B_{\frac{j'-1}{2^n}}$ can be written as a linear combination of $B_{s'+2^{-(n-1)}} - B_{s'}$ and $Z_{s'+2^{-n}}$. Because $B_s - B_{s-2^{-(n-1)}}$ and $B_{s'+2^{-(n-1)}} - B_{s'}$ are themselves linear combinations of random variables in $(Z_t : t \in \mathcal{D}_{n-1} \cup [0, m])$, and because $(B_t)_{t \in \mathcal{D}_{n-1}}$ is a discrete Wiener process, we see that $B_s - B_{s-2^{-(n-1)}}, B_{s'+2^{-(n-1)}} - B_{s'}, Z_{s+2^{-n}}, Z_{s-2^{-n}}$ are pairwise independent. By Corollary 3.20, we see that these four random variables are independent, so we conclude that $B_{\frac{j}{2^k}} - B_{\frac{j-1}{2^n}}$ and $B_{\frac{j'}{2^n}} - B_{\frac{j'-1}{2^n}}$ are independent.

Therefore, we have established that the entries of $(B_t - B_{t-2^{-n}} : t \in [m2^n])$ are independent and have the desired distribution of $N(0, 2^{-n})$. By writing arbitrary increments as sums of increments of the form $B_t - B_{t-2^{-n}}$ for $t \in \mathcal{D}_n \setminus \{0\}$ (and

[8]By "pairwise independent," we mean that, for any distinct $t_1, t_2 \in [m2^n]$, the random variables $B_{t_1} - B_{t_1-2^{-n}}$ and $B_{t_2} - B_{t_2-2^{-n}}$ are independent.

recalling that $m \in \mathbb{N}$ was arbitrary), we easily conclude using Theorem 3.10 that $(B_t)_{t\in\mathcal{D}_n}$ is a standard Wiener process, as desired.

Now that we have defined $(B_t)_{t\in\mathcal{D}}$ in the desired fashion, we examine the regularity of $(B_t)_{t\in\mathcal{D}}$. For the remainder of the subsection, fix any $\alpha \in (0, 1/2)$. Also, let $j, m \in \mathbb{N}$ and $k \in \mathbb{N}_0$ satisfy $\frac{j}{2^k} \leq m$. Since $B_{\frac{j}{2^k}} - B_{\frac{j-1}{2^k}}$ has distribution $N(0, 2^{-k})$ (and hence $2^{k/2}(B_{\frac{j}{2^k}} - B_{\frac{j-1}{2^k}})$ is a standard normal random variable), a simple computation using Proposition 3.11 (and the fact that $2\mu_X([a,\infty)) = \mu_X((-\infty, -a] \cup [a, \infty))$ for any $a > 0$ and any standard normal random variable X) yields that $\mathbb{P}(|B_{\frac{j}{2^k}} - B_{\frac{j-1}{2^k}}| \geq 2^{-k\alpha}) \leq C\exp(-c2^{k(1-2\alpha)})$ for constants $C = \frac{2}{\sqrt{2\pi}}$ and $c = \frac{1}{2}$. Note that $C\exp(-c2^{k(1-2\alpha)})$ is summable as (j,k) runs over all pairs satisfying $\frac{j}{2^k} \leq m$, so, by the Borel-Cantelli lemma, there is an event $\Omega_0^m \in \mathcal{F}$ with $\mathbb{P}(\Omega_0^m) = 1$ such that, for all $\omega \in \Omega_0^m$, we have $|B_{\frac{j}{2^k}}(\omega) - B_{\frac{j-1}{2^k}}(\omega)| < 2^{-k\alpha}$ for all but finitely many pairs (j,k) satisfying $\frac{j}{2^k} \leq m$. Hence, for all $\omega \in \Omega_0^m$, there exists $K_m(\omega) > 0$ such that $|B_{\frac{j}{2^k}}(\omega) - B_{\frac{j-1}{2^k}}(\omega)| \leq K_m(\omega)2^{-k\alpha}$ for all pairs (j,k) satisfying $\frac{j}{2^k} \leq m$. Let $\Omega_0 = \cap_{m=1}^{\infty}\Omega_0^m$; note that $\mathbb{P}(\Omega_0) = 1$. (Note also that Ω_0 depends on α, but this dependence does not matter.)

Let $\mathcal{F}_0 = \mathcal{F} \cap \mathcal{P}(\Omega_0)$ and $\mathbb{P}_0 = \mathbb{P}|_{\mathcal{F}_0}$. Henceforth, we restrict our attention to the probability space $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$, and we consider the domain of each random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ to now be restricted to Ω_0. (For example, if $t \in \mathcal{D}$, we consider the domain of B_t to be Ω_0 instead of Ω, even though we are still using the symbol B_t instead of $B_t|_{\Omega_0}$. Note that this does not affect the measurability of any of the random variables, and $(B_t)_{t\in\mathcal{D}_n}$ is still a Wiener process for each $n \in \mathbb{N}_0$.)

Fix $\omega \in \Omega_0$ and $a \in \mathcal{D}$. We now show that the map sending $t \in \mathcal{D} \cap [a, a+1)$ to $B_t(\omega)$ is α-Hölder continuous.[9] Choose $m \in \mathbb{N}$ such that $a + 2 \leq m$. Fix $s, t \in [a, a+1)$ with $s < t$. For each $k \in \mathbb{N}_0$, let s_k be the smallest element of $\mathcal{D}_k \cap [s, \infty)$, and let t_k be the largest element of $\mathcal{D}_k \cap [0, t]$. Choose the unique $n \in \mathbb{N}$ such that $t - s \in [2^{-n}, 2^{-n+1})$. Note that $[s,t]$ contains either one or two elements of $\mathcal{D}_n$. In the former case, $s_n = t_n$ is the unique element of $\mathcal{D}_n \cap [s,t]$; in the latter case, $s_{n-1} = t_{n-1}$ is the unique element of $\mathcal{D}_{n-1} \cap [s,t]$. Thus, we can always choose $l \in \{n-1, n\}$ such that $s_l = t_l$. Note that $t - s \geq 2^{-n} \geq 2^{-(l+1)}$. For each $k \in \mathbb{N}_0$, we observe that s_k and s_{k+1} are either equal or consecutive elements

[9] If (X, d_X) and (Y, d_Y) are metric spaces, then recall that a function $f : X \to Y$ is said to be α-Hölder continuous if there exists a constant $C > 0$ such that $d_Y(f(x), f(y)) \leq C(d_X(x,y)^\alpha)$ for all $x, y \in X$. It is easy to show that, if f is α-Hölder continuous, then f is uniformly continuous.

in $\mathcal{D}_{k+1}$, and similarly for t_k and t_{k+1}, so

$$\begin{aligned}|B_t(\omega)-B_s(\omega)| &= \left|\left(\sum_{k=l+1}^{\infty}(B_{t_k}(\omega)-B_{t_{k-1}}(\omega))\right)-\left(\sum_{k=l+1}^{\infty}(B_{s_k}(\omega)-B_{s_{k-1}}(\omega))\right)\right| \\ &\le \left(\sum_{k=l+1}^{\infty}|B_{t_k}(\omega)-B_{t_{k-1}}(\omega)|\right)+\left(\sum_{k=l+1}^{\infty}|B_{s_k}(\omega)-B_{s_{k-1}}(\omega)|\right) \\ &\le 2K_m(\omega)\sum_{k=l+1}^{\infty}2^{-k\alpha} = 2K_m(\omega)2^{-\alpha(l+1)}\left(\sum_{k=0}^{\infty}(2^{-\alpha})^k\right) \\ &\le \left(2K_m(\omega)\left(\sum_{k=0}^{\infty}(2^{-\alpha})^k\right)\right)(t-s)^{\alpha},\end{aligned}$$

which yields the desired α-Hölder continuity.

This α-Hölder continuity allows us to extend this map (i.e., the map sending $t\in\mathcal{D}$ to $B_t(\omega)$) to have domain $[0,\infty)$ instead of $\mathcal{D}$, such that this extension is continuous.[10] We now show that the $C([0,\infty),\mathbb{R})_0$-valued random variable $(B_t)_{t\in[0,\infty)}$ is a standard Brownian motion on $[0,\infty)$. Let $n\in\mathbb{N}$, and let $t_0,t_1,\dots,t_n\in[0,\infty)$ satisfy $t_0<t_1<\cdots<t_n$. By Proposition 3.14, it suffices to show that $(\frac{B_{t_1}-B_{t_0}}{\sqrt{t_1-t_0}},\frac{B_{t_2}-B_{t_1}}{\sqrt{t_2-t_1}},\dots,\frac{B_{t_n}-B_{t_{n-1}}}{\sqrt{t_n-t_{n-1}}})$ is an n-dimensional standard normal random variable. For $j\in[n]$, let $(t_{j,k})_{k\in\mathbb{N}}$ be a sequence of numbers in $\mathcal{D}\cap(t_{j-1},t_j]$ converging to t_j as $k\to\infty$. Also, let $(t_{0,k})_{k\in\mathbb{N}}$ be a sequence of numbers in $\mathcal{D}\cap[0,t_0]$ converging to t_0 as $k\to\infty$. Then $(\frac{B_{t_{1,k}}-B_{t_{0,k}}}{\sqrt{t_{1,k}-t_{0,k}}},\frac{B_{t_{2,k}}-B_{t_{1,k}}}{\sqrt{t_{2,k}-t_{1,k}}},\dots,\frac{B_{t_{n,k}}-B_{t_{n-1,k}}}{\sqrt{t_{n,k}-t_{n-1,k}}})$ is an n-dimensional standard normal random variable for each $k\in\mathbb{N}$ (because $(B_t)_{t\in\mathcal{D}_l}$ is a Wiener process for each $l\in\mathbb{N}_0$), so $(\frac{B_{t_{1,k}}-B_{t_{0,k}}}{\sqrt{t_{1,k}-t_{0,k}}},\frac{B_{t_{2,k}}-B_{t_{1,k}}}{\sqrt{t_{2,k}-t_{1,k}}},\dots,\frac{B_{t_{n,k}}-B_{t_{n-1,k}}}{\sqrt{t_{n,k}-t_{n-1,k}}})$ converges in distribution to an n-dimensional standard normal random variable as $k\to\infty$. Furthermore, by continuity, $(\frac{B_{t_{1,k}}-B_{t_{0,k}}}{\sqrt{t_{1,k}-t_{0,k}}},\frac{B_{t_{2,k}}-B_{t_{1,k}}}{\sqrt{t_{2,k}-t_{1,k}}},\dots,\frac{B_{t_{n,k}}-B_{t_{n-1,k}}}{\sqrt{t_{n,k}-t_{n-1,k}}})$ converges pointwise, and hence in distribution, to $(\frac{B_{t_1}-B_{t_0}}{\sqrt{t_1-t_0}},\frac{B_{t_2}-B_{t_1}}{\sqrt{t_2-t_1}},\dots,\frac{B_{t_n}-B_{t_{n-1}}}{\sqrt{t_n-t_{n-1}}})$ as $k\to\infty$. By the uniqueness of limits for convergence in distribution for $\mathbb{R}^n$-valued random variables, we conclude that $(\frac{B_{t_1}-B_{t_0}}{\sqrt{t_1-t_0}},\frac{B_{t_2}-B_{t_1}}{\sqrt{t_2-t_1}},\dots,\frac{B_{t_n}-B_{t_{n-1}}}{\sqrt{t_n-t_{n-1}}})$ is indeed an n-dimensional standard normal random variable, which completes the proof of the existence of a standard Brownian motion $(B_t)_{t\in[0,\infty)}$.

4.2. **Wiener Measure.** The following two results establish that Wiener processes are unique with respect to distribution.

Theorem 4.1. *Let $(X_t)_{t\in\mathcal{T}}$ and $(Y_t)_{t\in\mathcal{T}}$ be Wiener processes (not necessarily defined on the same probability space) on the set of times $\mathcal{T}$ with starting point x_0, drift μ, and variance σ^2. Then $(X_t)_{t\in\mathcal{T}}$ and $(Y_t)_{t\in\mathcal{T}}$ agree in distribution. (This distribution on $C(\mathcal{T},\mathbb{R})_{x_0}$ is called the* ***(one-dimensional) Wiener measure associated with $\boldsymbol{\mathcal{T}, x_0, \mu}$, and $\boldsymbol{\sigma^2}$****, and we will denote this probability distribution by $w_{\mathcal{T}}(x_0,\mu,\sigma^2)$.)*

[10]Here, we are using the following basic result from analysis: if (X,d_X) and (Y,d_Y) are metric spaces, if A is a dense subset of X, if Y is complete (i.e., every Cauchy sequence in Y converges), and if $f:A\to Y$ is uniformly continuous, then f can be extended to a continuous map $\tilde{f}:X\to Y$. (In fact, $\tilde{f}$ is uniformly continuous.) Also, note that a continuous function $f:X\to Y$ is uniquely determined by its values on a dense subset A, so the extension $\tilde{f}$ is unique.

Proof. We can consider $(X_t)_{t\in\mathcal{T}}$ and $(Y_t)_{t\in\mathcal{T}}$ to be $C(\mathcal{T},\mathbb{R})$-valued random variables (without affecting the measurability of the random variables), and it suffices to prove that they induce the same probability distribution on $C(\mathcal{T},\mathbb{R})$. Let $n\in\mathbb{N}$, and let $t_0,t_1,\ldots,t_n\in\mathcal{T}$ satisfy $0=t_0<t_1<\cdots<t_n$. Also, let $B_0,B_1,B_2,\ldots,B_n\in\mathcal{B}_\mathbb{R}$. The collection of all sets of the form $C_{t_0,B_0}\cap C_{t_1,B_1}\cap\cdots\cap C_{t_n,B_n}$ is a π-system in $C(\mathcal{T},\mathbb{R})_{x_0}$ that generates $\mathcal{C}_{\mathcal{T},\mathbb{R}}$, so, by Proposition 1.1, it suffices to show that $\mu_{(X_t)_{t\in\mathcal{T}}}$ and $\mu_{(Y_t)_{t\in\mathcal{T}}}$ agree on $C_{t_0,B_0}\cap C_{t_1,B_1}\cap\cdots\cap C_{t_n,B_n}$.

For each $j\in[n]$, let $dX_j=X_{t_j}-X_{t_{j-1}}$, and let $dY_j=Y_{t_j}-Y_{t_{j-1}}$. Thus, $\mu_{dX_j}=N(\mu(t_j-t_{j-1}),\sigma^2(t_j-t_{j-1}))=\mu_{dY_j}$, so, by the independence of the increments, we have $\mu_{(dX_1,dX_2,\ldots,dX_n)}=\mu_{dX_1}\times\mu_{dX_2}\times\cdots\times\mu_{dX_n}=\mu_{dY_1}\times\mu_{dY_2}\times\cdots\times\mu_{dY_n}=\mu_{(dY_1,dY_2,\ldots,dY_n)}$. Using Proposition 1.2, we deduce that $\mu_{(X_{t_0},X_{t_1},\ldots,X_{t_n})}=\mu_{(Y_{t_0},Y_{t_1},\ldots,Y_{t_n})}$, which implies that $\mu_{(X_t)_{t\in\mathcal{T}}}$ and $\mu_{(Y_t)_{t\in\mathcal{T}}}$ indeed agree on $C_{t_0,B_0}\cap C_{t_1,B_1}\cap\cdots\cap C_{t_n,B_n}$. QED

Theorem 4.2. *Let $(X_t)_{t\in\mathcal{T}}$ be a $C(\mathcal{T},\mathbb{R})_{x_0}$-valued random variable with $\mu_{(X_t)_{t\in\mathcal{T}}}=w_\mathcal{T}(x_0,\mu,\sigma^2)$. Then $(X_t)_{t\in\mathcal{T}}$ is a Wiener process on the set of times $\mathcal{T}$ with starting point x_0, drift μ, and variance σ^2.*

Proof. Let $(Y_t)_{t\in\mathcal{T}}$ be a Wiener process on the set of times $\mathcal{T}$ with starting point x_0, drift μ, and variance σ^2. (We established the existence of $(Y_t)_{t\in\mathcal{T}}$ in the preceding subsection.) Then $\mu_{(X_t)_{t\in\mathcal{T}}}=\mu_{(Y_t)_{t\in\mathcal{T}}}$. Fix $n\in\mathbb{N}$, and let $t_0,t_1,\ldots,t_n\in\mathcal{T}$ satisfy $t_0<t_1<\cdots<t_n$. For each $j\in[n]$, let $dX_j=X_{t_j}-X_{t_{j-1}}$ and $dY_j=Y_{t_j}-Y_{t_{j-1}}$. It suffices to show that $\mu_{(dX_1,dX_2,\ldots,dX_n)}=\mu_{(dY_1,dY_2,\ldots,dY_n)}$ (by taking $n=1$ for normally distributed increments and using product measures for independent increments), and this follows immediately from Proposition 1.2. QED

4.3. **Multidimensional Brownian Motion.** We now briefly discuss Wiener processes in higher dimensions. For simplicity (and with very little loss of generality), we only consider standard Wiener processes.

Definition 4.3. A **(standard) d-dimensional Wiener process on the set of times $\mathcal{T}$** is a $C(\mathcal{T},\mathbb{R}^d)_0$-valued random variable of the form $(X_1,X_2,\ldots,X_d)$, where $X_1,X_2,\ldots,X_d$ are independent standard one-dimensional Wiener processes on the set of times $\mathcal{T}$. (Note that, if $X_1,X_2,\ldots,X_d$ are measurable $C(\mathcal{T},\mathbb{R})$-valued functions, then $(X_1,X_2,\ldots,X_d)$ is a measurable $C(\mathcal{T},\mathbb{R}^d)$-valued function.)

Definition 4.4. Let $\mathcal{T}$ be an interval I. Then a d-dimensional Wiener process on the set of times I is called a **d-dimensional Brownian motion on I**. If $d=1$, we sometimes refer to this as a **linear Brownian motion on I**; if $d=2$, we sometimes refer to this as a **planar Brownian motion on I**.

We have already established the existence of a (standard) one-dimensional Wiener process on the set of times $\mathcal{T}$. Hence, we have established the existence of the associated Wiener measure $w_\mathcal{T}(0,0,1)$. Thus, Theorems 1.3 and 4.2 imply the existence of a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}$.

Lemma 4.5. *Let $(X_t)_{t\in\mathcal{T}}$ be a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}$. Then the following hold:*

(a) ***(Normally distributed increments)*** *If $s,t\in\mathcal{T}$ and $s<t$, then $\frac{1}{\sqrt{t-s}}(X_t-X_s)$ is a d-dimensional standard normal random variable.*

(b) ***(Independent increments)*** *If $n\in\mathbb{N}$, and if $t_0,t_1,\ldots,t_n\in\mathcal{T}$ satisfy $t_0<t_1<\cdots<t_n$, then $X_{t_1}-X_{t_0},X_{t_2}-X_{t_1},\ldots,X_{t_n}-X_{t_{n-1}}$ are independent.*

(Hence, if $n \in \mathbb{N}$, and if $t_0, t_1, \ldots, t_n \in \mathcal{T}$ satisfy $t_0 \le t_1 \le \cdots \le t_n$, then $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.)

Proof. For each $t \in \mathcal{T}$, write $X_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,d})$ for real-valued random variables $X_{t,1}, X_{t,2}, \ldots, X_{t,d}$.

The proof of the first condition (normally distributed increments) is easy. For the second condition (independent increments), independence implies that the joint distribution of $(X_{t_1,1} - X_{t_0,1}, X_{t_2,1} - X_{t_1,1}, \ldots, X_{t_n,1} - X_{t_{n-1},1}), (X_{t_1,2} - X_{t_0,2}, X_{t_2,2} - X_{t_1,2}, \ldots, X_{t_n,2} - X_{t_{n-1},2}), \ldots, (X_{t_1,d} - X_{t_0,d}, X_{t_2,d} - X_{t_1,d}, \ldots, X_{t_n,d} - X_{t_{n-1},d})$ is the product of the d marginal distributions, and that each of these marginal distributions $\mu_{(X_{t_1,j} - X_{t_0,j}, X_{t_2,j} - X_{t_1,j}, \ldots, X_{t_n,j} - X_{t_{n-1},j})}$, where $j \in [d]$, is itself equal to the product $\mu_{X_{t_1,j} - X_{t_0,j}} \times \mu_{X_{t_2,j} - X_{t_1,j}} \times \cdots \times \mu_{X_{t_n,j} - X_{t_{n-1},j}}$. Using the associativity of the product measure, we see that $\{X_{t_i,j} - X_{t_{i-1},j}\}_{(i,j) \in [n] \times [d]}$ is independent, which implies the result. QED

Proposition 4.6. *Let $(X_t)_{t \in \mathcal{T}}$ and $(Y_t)_{t \in \mathcal{T}}$ be $C(\mathcal{T}, \mathbb{R}^d)_0$-valued random variables satisfying the two conditions stated in Lemma 4.5 (i.e., normally distributed increments and independent increments). (In particular, $(X_t)_{t \in \mathcal{T}}$ and $(Y_t)_{t \in \mathcal{T}}$ could be standard d-dimensional Wiener processes on the set of times $\mathcal{T}$.) Then $(X_t)_{t \in \mathcal{T}}$ and $(Y_t)_{t \in \mathcal{T}}$ agree in distribution. (This distribution on $C(\mathcal{T}, \mathbb{R}^d)_0$ is called the* ***d-dimensional Wiener measure associated with $\mathcal{T}$****, and we will denote this probability distribution by $w_{\mathcal{T}}(\mathbb{R}^d)$. Note that $w_{\mathcal{T}}(\mathbb{R}) = w_{\mathcal{T}}(0, 0, 1)$.)*

Proof. This follows by essentially the exact same argument as in the proof of Theorem 4.1. QED

Theorem 4.7. *Let $(X_t)_{t \in \mathcal{T}}$ be a $C(\mathcal{T}, \mathbb{R}^d)_0$-valued random variable with $\mu_{(X_t)_{t \in \mathcal{T}}} = w_{\mathcal{T}}(\mathbb{R}^d)$. Then $(X_t)_{t \in \mathcal{T}}$ is a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}$.*

Proof. Fix $j \in [d]$. Consider the map sending $f = (f_1, f_2, \ldots, f_d) \in C(\mathcal{T}, \mathbb{R}^d)$ to $f_j \in C(\mathcal{T}, \mathbb{R})$. It is not hard to check that the preimage of a cylinder set $C_{t,B}$ (for $t \in \mathcal{T}$ and $B \in \mathcal{B}$) is measurable in $C(\mathcal{T}, \mathbb{R}^d)$, so this map is measurable. It follows (by comparing $(X_t)_{t \in \mathcal{T}}$ to a standard d-dimensional Wiener process and using Proposition 1.2) that the one-dimensional (random) component functions of $(X_t)_{t \in \mathcal{T}}$ all have $w(\mathbb{R})$ as their distribution, so, by Theorem 4.2, the component functions of $(X_t)_{t \in \mathcal{T}}$ are Wiener processes on the set of times $\mathcal{T}$.

Now consider the map sending $f = (f_1, f_2, \ldots, f_d) \in C(\mathcal{T}, \mathbb{R}^d)$ to the tuple of functions $(f_1, f_2, \ldots, f_d) \in C(\mathcal{T}, \mathbb{R}) \times C(\mathcal{T}, \mathbb{R}) \times \cdots \times C(\mathcal{T}, \mathbb{R})$. (Note that the notation $(f_1, f_2, \ldots, f_d)$ is being used in two different ways here.) We first note that the collection of all sets of the form $C_{t_1,B_1} \times C_{t_2,B_2} \times \cdots \times C_{t_d,B_d}$, where $t_1, t_2, \ldots, t_d \in \mathcal{T}$ and $B_1, B_2, \ldots, B_d \in \mathcal{B}_{\mathbb{R}}$, generates the product σ-algebra in $C(\mathcal{T}, \mathbb{R}) \times C(\mathcal{T}, \mathbb{R}) \times \cdots \times C(\mathcal{T}, \mathbb{R})$. (To prove this, use induction on j to show that, for all $j \in [d]_0$, the generated σ-algebra contains all sets of the form $D_1 \times D_2 \times \cdots \times D_j \times C_{t_{j+1},B_{j+1}} \times C_{t_{j+2},B_{j+2}} \times \cdots \times C_{t_d,B_d}$, where $D_1, D_2, \ldots, D_j \in \mathcal{C}_{\mathcal{T},\mathbb{R}}$.) It is not hard to show that the preimage of any such set $C_{t_1,B_1} \times C_{t_2,B_2} \times \cdots \times C_{t_d,B_d}$ is measurable in $C(\mathcal{T}, \mathbb{R}^d)$, so this map is measurable. Thus, by comparing $(X_t)_{t \in \mathcal{T}}$ to a standard d-dimensional Wiener process and using Proposition 1.2, it follows that the joint distribution of the one-dimensional (random) component functions of $(X_t)_{t \in \mathcal{T}}$ on $C(\mathcal{T}, \mathbb{R}) \times C(\mathcal{T}, \mathbb{R}) \times \cdots \times C(\mathcal{T}, \mathbb{R})$ is the product of the marginal distributions, which gives the desired independence. QED

Corollary 4.8. *Let $(X_t)_{t\in\mathcal{T}}$ be a $C(\mathcal{T},\mathbb{R}^d)_0$-valued random variable. Then $(X_t)_{t\in\mathcal{T}}$ is a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}$ if and only if $(X_t)_{t\in\mathcal{T}}$ satisfies the two conditions stated in Lemma 4.5 (i.e., normally distributed increments and independent increments).*

Proof. One implication is directly given by Lemma 4.5. The other implication follows from Proposition 4.6, and Theorem 4.7. QED

Proposition 4.9. ***(Scaling invariance)*** *Fix $a > 0$. Suppose $(B_t)_{t\in\mathcal{T}}$ is a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}$. Then $(\frac{1}{\sqrt{a}}B_{at})_{t\in\mathcal{T}/a}$ is a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}/a = \{t/a : t \in \mathcal{T}\}$.*

Proof. This is straightforward using Corollary 4.8. QED

Proposition 4.10. *Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a linear isometry, and let $(X_t)_{t\in\mathcal{T}}$ be a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}$. Then $(TX_t)_{t\in\mathcal{T}}$ is a (standard) d-dimensional Wiener process on the set of times $\mathcal{T}$.*

Proof. This is straightforward using Proposition 3.15 and Corollary 4.8. QED

4.4. **Differentiability of Brownian Motion.** We first recall the Weierstrass Approximation Theorem from analysis, which states that any continuous function $f \in C([0,1],\mathbb{R})$ can be uniformly approximated arbitrarily closely by a polynomial, in the sense that, for any $\epsilon > 0$, there exists a polynomial $p : [0,1] \to \mathbb{R}$ such that $|f(x) - p(x)| < \epsilon$ for all $x \in [0,1]$.[11] Thus, we see that all continuous functions in $C([0,1],\mathbb{R})$ are quite "well-behaved" in the sense that they are very close (in the supremum norm) to being a polynomial (and polynomials are well-behaved in the sense that they are smooth).

We now establish a striking result (Theorem 4.11) that contrasts with the comments in the preceding paragraph. Informally speaking, Theorem 4.11 says that, if we randomly choose a continuous function in $C([0,1],\mathbb{R})$ in some reasonably uniform manner, then, almost surely, our chosen function will be nowhere differentiable! (Thus, almost all continuous functions from $[0,1]$ to $\mathbb{R}$ are like the famous Weierstrass function, in the sense that they are continuous everywhere but differentiable nowhere.) Since shifting a function does not affect differentiability, we can assume without loss of generality that our randomly chosen function $f \in C([0,1],\mathbb{R})$ satisfies $f(0) = 0$. Thus, it is reasonable to model our randomly chosen function in $C([0,1],\mathbb{R})$ as a (one-dimensional) standard Brownian motion on $[0,1]$.

We now formally state and prove the theorem:

Theorem 4.11. *Let $(B_t)_{t\in[0,1]}$ be a (one-dimensional) standard Brownian motion on $[0,1]$, defined on the probability space $(\Omega,\mathcal{F},\mathbb{P})$. For $\omega \in \Omega$, consider the function sending $t \in [0,1]$ to $B_t(\omega) \in \mathbb{R}$. Then this function is a.s. nowhere Lipschitz continuous*[12] *and hence a.s. nowhere differentiable.*

Proof. Fix $m, n \in \mathbb{N}$. Let $A_{m,n} = \{\omega \in \Omega : \text{there exists } s \in [0,1] \text{ such that } |B_t(\omega) - B_s(\omega)| \leq m|t-s| \text{ whenever } t \in [0,1] \text{ satisfies } |t-s| \leq \frac{3}{n}\}$. It suffices to show that $A_{m,n}$ is contained within an event of probability zero, since then the union of the $A_{m,n}$ over all m, n is contained within an event of probability zero.

[11]For a proof of this, see Theorem 7.26 in [7].

[12]If (X, d_X) and (Y, d_Y) are metric spaces, and if $x \in X$, then recall that a function $f : X \to Y$ is said to be Lipschitz continuous at x if there exist $C > 0$ and $\epsilon > 0$ such that $d_Y(f(x), f(y)) \leq Cd_X(x,y)$ for all $y \in X$ satisfying $d_X(x,y) < \epsilon$.

For $k \in [n-2]$, let $X_{k,n} = \max\{|B_{\frac{k+j}{n}} - B_{\frac{k+j-1}{n}}| : j \in [2]_0\}$. Let $E_{m,n} = \{\omega \in \Omega : X_{k,n}(\omega) \leq \frac{5m}{n}$ for some $k \in [n-2]\} \in \mathcal{F}$. It is not hard to show that $A_{m,n} \subset E_{m,n}$ and that $A_{m,1} \subset A_{m,2} \subset A_{m,3} \subset \cdots$. Hence, we have $A_{m,n} \subset \bigcap_{k=n}^{\infty} E_{m,k}$, so it suffices to show that $\mathbb{P}(E_{m,n}) \to 0$ as $n \to \infty$.

Using the union bound and the properties of the increments of Brownian motion, we see that $\mathbb{P}(E_{m,n}) \leq (n-2)(|\mathbb{P}(B_{\frac{1}{n}}| \leq \frac{5m}{n}))^3 \leq n(\mathbb{P}(|B_{\frac{1}{n}}| \leq \frac{5m}{n}))^3$. By scaling invariance (4.9) with $a = \sqrt{n}$, we have $\mathbb{P}(E_{m,n}) \leq n(\mathbb{P}(|B_{\frac{1}{n}}| \leq \frac{5m}{n}))^3 = n(\mathbb{P}(|B_1| \leq \frac{5m}{\sqrt{n}}))^3 \leq n(\frac{1}{\sqrt{2\pi}} \frac{10m}{\sqrt{n}})^3 \to 0$ as $n \to \infty$. QED

Acknowledgments

I would like to thank Prof. Rick Durrett, Prof. Ewain Gwynne, Prof. Greg Lawler, Prof. Peter Mörters, Prof. Ron Peled, Prof. Yuval Peres, Prof. Sadahiro Saeki, Prof. Terence Tao, and Prof. Tomasz Tkocz, all of whose texts or notes on probability were invaluable as I was learning this material.

I am very grateful to my mentor Minjae Park for his support and guidance throughout the REU. Also, I am very grateful to Prof. Peter May for organizing this REU and for his feedback on this paper, and I would like to thank everyone who lectured in the program. Finally, I would like to thank my friends and family for their constant support.

References

[1] Sadahiro Saeki. *A Proof of the Existence of Infinite Product Probability Measures.* Taylor & Francis. https://www.jstor.org/stable/2974880.
[2] Walter Rudin. *Real and Complex Analysis.*
[3] Richard F. Bass. *Real Analysis for Graduate Students.* 2020.
[4] Terrance Tao. *An Introduction to Measure Theory.*
[5] Gerald Folland. *Real Analysis: Modern Techniques and Their Applications.*
[6] Rick Durrett. *Probability: Theory and Examples.*
[7] Walter Rudin. *Principles of Mathematical Analysis.*