ADVANCED PROBABILITY II NOTES

William J. Anderson
McGill University

Contents

Chapter 1

Weak Convergence of Probability Measures.

1.1 Introduction.

Let $(E, \mathcal{E})$ be a measurable space, and let $\mathcal{P}(E)$ be the set of all probabilities on $\mathcal{E}$. In this chapter, we wish to consider the convergence of sequences of probabilities in $\mathcal{P}(E)$.

1.2 Total Variation Measure.

Notation. $\mathcal{E}$ will also denote the set of all measurable functions $f : E \to \bar{\mathbb{R}}$, and $\mathcal{E}_+$ will denotes the set of all non-negative such functions. Recall that a *signed measure* μ on $\mathcal{E}$ is a countably additive real-valued set function on $\mathcal{E}$, which takes at most one of the values $+\infty, -\infty$. For any $x \in E$, we define the *Dirac measure* $\epsilon_x(\cdot)$ by

$$\epsilon_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

ϵ_x is actually a probability measure; if $f \in \mathcal{E}$, then $\int f(y)\epsilon_x(dy) = f(x)$.

Total Variation. Let μ be a signed measure on $\mathcal{E}$. Then by the Jordan-Hahn decomposition theorem, the formulas

$$\mu^+(A) = \sup_{B \subset A, B \in \mathcal{E}} \mu(B), \quad \mu^-(A) = -\inf_{B \subset A, B \in \mathcal{E}} \mu(B), \quad A \in \mathcal{E},$$

define measures μ^+ and μ^- on $\mathcal{E}$ (one of which must be finite) such that $\mu = \mu^+ - \mu^-$ (This is called the *Jordan decomposition* of μ). Moreover, it is possible to find at least one set $E^+ \in \mathcal{E}$ such that

$$\mu^+(A) = \mu(A \cap E^+), \quad \mu^-(A) = -\mu(A \cap E^-), \quad A \in \mathcal{E},$$

where $E^- = E \setminus E^+$. The partition $E = E^+ \cup E^-$ is called the *Hahn* decomposition of E.

The measure $|\mu| = \mu^+ + \mu^-$, defined by

$$|\mu|(A) = \mu^+(A) + \mu^-(A), \quad A \in \mathcal{E},$$

is called the *total variation measure* corresponding to μ. Observe that $|\mu(A)| \leq |\mu|(A)$ for any $A \in \mathcal{E}$. A signed measure μ is called *finite* if $|\mu|(E) < \infty$.

Examples.

(1) Suppose E is countable, and that $\mathcal{E}$ is the family of all subsets of E. Then $\mu^+(\{x\}) = \mu(\{x\}) \vee \mu(\emptyset) = \mu(\{x\}) \vee 0$ and similarly $\mu^-(\{x\}) = -\mu(\{x\}) \wedge 0$. Hence μ^+ and μ^- are given by

$$\mu^+(A) = \sum_{x \in A} \mu(\{x\}) \vee 0, \quad \mu^-(A) = -\sum_{x \in A} \mu(\{x\}) \wedge 0, \quad A \in \mathcal{E},$$

and we can take

$$E^+ = \{x \in E | \mu(\{x\}) \geq 0\}, \quad E^- = \{x \in E | \mu(\{x\}) < 0\}.$$

The total variation measure is given by

$$|\mu|(A) = \sum_{x \in A} |\mu(\{x\})|, \quad A \in \mathcal{E}.$$

(2) Suppose μ is of the form

$$\mu(A) = \int_A f(x)\nu(dx), \quad A \in \mathcal{E},$$

where ν is a measure and f is quasi-ν-integrable. Then

$$\mu^+(A) = \int_A f^+(x)\nu(dx), \quad \mu^-(A) = \int_A f^-(x)\nu(dx), \quad A \in \mathcal{E},$$

where $f^+(x) = f(x) \vee 0$ and $f^-(x) = -[f(x) \wedge 0]$ for all $x \in E$. The total variation measure is given by

$$|\mu|(A) = \int_A |f(x)|\nu(dx), \quad A \in \mathcal{E}.$$

Proposition 1.2.1 *Let μ be a signed measure on $\mathcal{E}$ such that $|\mu|(E) < \infty$. Then*

$$|\mu|(E) = \sup_{f \in \mathcal{E}: |f| \leq 1} |\int f \, d\mu|.$$

Proof. For such an f as described, we have

$$|\int f \, d\mu| = |\int f \, d\mu^+ - \int f \, d\mu^-| \leq \int |f| \, d\mu^+ + \int |f| \, d\mu^- = \int |f| d|\mu| \leq |\mu|(E),$$

and so we have shown that

$$|\mu|(E) \geq \sup_{|f| \leq 1} |\int f \, d\mu|.$$

Equality is actually achieved, since for $f = I_{E^+} - I_{E^-}$, we have $|f| \leq 1$ and

$$\int f \, d\mu = \int I_{E^+} \, d\mu - \int I_{E^-} \, d\mu = \mu^+(E) + \mu^-(E) = |\mu|(E).$$

∎

Definition. The total variation norm of μ is defined to be

$$\|\mu\| = |\mu|(E) = \sup_{|f| \leq 1} |\int f d\,\mu|. \tag{2.1}$$

Remarks.

(1) If μ is a measure, then $\|\mu\| = \mu(E)$.

(2) Let $\mathcal{M}_b(E)$ be the set of all signed measures μ on $\mathcal{E}$ such that $\|\mu\| < \infty$. Then $\mathcal{M}_b(E)$ is a vector space and $\|\cdot\|$ defines a norm on $\mathcal{M}_b(E)$. For we have

$$\|\mu_1 + \mu_2\| = \sup_{|f|\le 1|} |\int f d\,(\mu_1+\mu_2)| = \sup_{|f|\le 1|} |\int f d\,\mu_1 + \int f d\,\mu_2| \le \|\mu_1\| + \|\mu_2\|$$

and

$$\|c\mu\| = \sup_{|f|\le 1|} |\int f d\,(c\mu)| = |c| \sup_{|f|\le 1|} |\int f d\,\mu| = |c|\|\mu\|.$$

It can be shown that $\mathcal{M}_b(E)$ with this norm forms a Banach space. Moreover, the subset $\mathcal{P}(E)$ consisting of probability measures is closed.

(3) If μ is such that $\mu(E) = 0$, then $\mu^+(E) = \mu^-(E)$, and so

$$\|\mu\| = 2 \sup_{A\in\mathcal{E}} \mu(A).$$

In particular, if P_1 and P_2 are probability measures, then

$$\|P_1 - P_2\| = 2 \sup_{A\in\mathcal{E}} |P_1(A) - P_2(A)|. \tag{2.2}$$

(4) The inequality

$$\sup_{|f|\le 1} |\mu(f) - \nu(f)| \le \|\mu - \nu\| \tag{2.3}$$

is useful. In particular, if $\|\mu_n - \mu\| \to 0$ as $n \to \infty$, then $\mu_n(A) \to \mu(A)$ uniformly in $A \in \mathcal{E}$.

Exercise. Let $\mathcal{A}$ be a subset of $\mathcal{E}$ which contains a a subclass $\mathcal{B}$ satisfying (i) $\sigma(\mathcal{B}) = \mathcal{E}$, and (ii) $\mathcal{B}$ is a π-system (that is, $\mathcal{B}$ is closed under finite intersection). Then

$$d(\mu,\nu) = \sup_{A\in\mathcal{A}} |\mu(A) - \nu(A)|$$

defines a metric on $\mathcal{P}(E)$. In particular, if $\mathcal{A} = \mathcal{E}$, then $d(\mu,\nu) = \|\mu - \nu\|/2$ is the total variation metric on $\mathcal{P}(E)$.

Solution. Certainly $d(\cdot,\cdot)$ is non-negative and symmetric and the triangle condition is satisfied. If $d(\mu,\nu) = 0$, then $\mu(A) = \nu(A)$ for all $A \in \mathcal{B}$. Since $\mathcal{B}$ is a π-system which generates $\mathcal{E}$, then $\mu(A) = \nu(A)$ for all $A \in \mathcal{E}$, and so $\mu = \nu$. ∎

Exercise. If $x \ne y$, show that $\|\epsilon_x - \epsilon_y\| = 2$.

Definition. Let $\{\mu_n, n \ge 1\} \subset \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$. We say that $\{\mu_n, n \ge 1\}$ converges to μ in *total variation* if $\|\mu_n - \mu\| \to 0$ as $n \to \infty$.

Suppose that E is a metric space and that $\{x_n, n \ge 1\} \subset E$ and $x \in E$ are such that $x_n \ne x$ for all n, and $x_n \to x$ as $n \to \infty$. Since $\|\epsilon_{x_n} - \epsilon_x\| = 2$, the corresponding sequence of Dirac measures cannot converge in total variation to the Dirac measure at x.

Proposition 1.2.2 (Sheffé's Theorem) *Let λ be a measure on a measurable space $(E,\mathcal{E})$, and let f and $\{f_n, n \ge 1\}$ be non-negative measurable functions on E such that $\int f\,d\lambda = 1$ and $\int f_n\,d\lambda = 1$ for all n, and let*

$$\mu_n(A) = \int_A f_n\,d\lambda, \quad \mu(A) = \int_A f\,d\lambda, \quad A \in \mathcal{E}.$$

If $f_n(x) \to f(x)$ for λ-a.e. $x \in E$, then $\mu_n \to \mu$ in total variation as $n \to \infty$.

Proof. We have $\|\mu - \mu_n\| = 2(\mu - \mu_n)^+(E) = 2\int(f - f_n)^+ \, d\lambda$. Since $(f - f_n)^+ \to 0$ except on a set of λ-measure 0, and $0 \le (f - f_n)^+ \le f$ for all n, then the result follows by the bounded convergence theorem. ∎

Proposition 1.2.3 *Suppose E is a countable set, and that $\mathcal{E}$ is the family of all subsets of E. Let $\{\mu_n, n \ge 1\}$ and μ be probability measures on $\mathcal{E}$. Then $\mu_n \to \mu$ in total variation if and only if $\mu_n(\{x\}) \to \mu(\{x\})$ for every $x \in E$.*

Proof. We have
$$\|\mu_n - \mu\| = \sum_{x\in E} |\mu(\{x\}) - \mu_n(\{x\})| = 2\sum_{x\in E}[\mu(\{x\}) - \mu_n(\{x\})]^+.$$
If $\mu_n \to \mu$ in total variation, then obviously $\mu_n(\{x\}) \to \mu(\{x\})$ for every x. Conversely, if $\mu_n(\{x\}) \to \mu(\{x\})$ for every x, then $[\mu(\{x\}) - \mu_n(\{x\})]^+ \to 0$ and $[\mu(\{x\}) - \mu_n(\{x\})]^+ \le \mu(\{x\})$ for every x, so $\|\mu - \mu_n\| \to 0$ by the bounded convergence theorem. ∎

Proposition 1.2.4 *Let μ_1, ν_1 be distributions on $(E_1, \mathcal{E}_1)$, and let μ_2, ν_2 be distributions on $(E_2, \mathcal{E}_2)$. Then*
$$\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\| \le \|\mu_1 - \nu_1\| + \|\mu_2 - \nu_2\|.$$

Proof. Let $f(x_1, x_2) \in \mathcal{E}_1 \times \mathcal{E}_2$ be such that $|f| \le 1$. Then
$$\begin{aligned}
|\int f\, d(\mu_1 \times \mu_2) - \int f\, d(\nu_1 \times \nu_2)| &\le |\int f\, d(\mu_1 \times \mu_2) - \int f\, d(\nu_1 \times \mu_2)| \\
&\quad + |\int f\, d(\nu_1 \times \mu_2) - \int f\, d(\nu_1 \times \nu_2)| \\
&= |\int \mu_1(dx_1) \int f(x_1, x_2)\mu_2(dx_2) - \int \nu_1(dx_1) \int f(x_1, x_2)\mu_2(dx_2)| \\
&\quad + |\int \mu_2(dx_2) \int f(x_1, x_2)\nu_1(dx_1) - \int \nu_2(dx_2) \int f(x_1, x_2)\nu_1(dx_1)| \\
&\le \|\mu_1 - \nu_1\| + \|\mu_2 - \nu_2\|,
\end{aligned}$$
which implies the desired result. ∎

Proposition 1.2.5 *Let μ_1, ν_1, μ_2, and ν_2 be distributions on $(E, \mathcal{E})$, where E is assumed to be an additive group, so that convolution is defined. Then*
$$\|\mu_1 \star \mu_2 - \nu_1 \star \nu_2\| \le \|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\|.$$

Proof. We use the fact that if μ and ν are measures on $\mathcal{E}$ and if $f \in \mathcal{E}$, then $\mu\star\nu(f) = \int \mu(dx) \int \nu(dy) f(x+y)$. Let $f \in \mathcal{E}$ be such that $|f| \le 1$. Then
$$|\mu_1 \star \mu_2(f) - \nu_1 \star \nu_2(f)| = |\int \mu_1(dx) \int \mu_2(dy) f(x+y) - \int \nu_1(dx) \int \nu_2(dy) f(x+y)| \le \|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\|.$$
∎

1.3 Measures in Metric Spaces.

For the rest of this chapter, we assume that E is a metric space with metric d, and that $\mathcal{E}$ is the Borel σ-algebra, namely the σ-algebra generated by the family of open subsets of E. Let $\mathcal{C}_b(E)$ denote the set of all bounded continuous functions $f : E \to \mathbb{R}$. In this section, we are going to establish some needed properties of probability measures in this setting. We shall need the definition
$$d(x, A) = \inf\{d(x, y) | y \in A\}, \quad x \in E, A \subset E.$$

Proposition 1.3.1 *Let $A \subset E$.*

(1) $|d(x,A) - d(y,A)| \le d(x,y)$, and so $d(x,A)$ is a continuous function of x (in fact a Lipschitz function of x-see below).

(2) $d(x,A) = 0$ if and only if $x \in \bar{A}$.

Proof.

(1) From the triangle inequality, we have $d(x,z) \le d(x,y) + d(y,z)$. Taking the infimum over all $z \in A$ gives $d(x,A) \le d(x,y) + d(y,A)$, so that $d(x,A) - d(y,A) \le d(x,y)$. Interchanging x and y gives $d(y,A) - d(x,A) \le d(y,x) = d(x,y)$. We conclude that $|d(x,A) - d(y,A)| \le d(x,y)$ for all x, y.

(2) By part (1), the set $\{x \in E | d(x,A) = 0\}$ is closed and contains A, so it contains $\bar{A}$. Conversely, if $d(x,A) = 0$, there is a sequence $\{x_n, n \ge 1\} \subset A$ such that $d(x, x_n) \to 0$, and therefore $x \in \bar{A}$.

∎

Definition. Let $f : E \to \mathbb{R}$ be a measurable function. Then f is said to be *Lipschitz continuous* if the number

$$s(f) = \sup_{\substack{x \neq y \\ x,y \in E}} \frac{|f(x) - f(y)|}{d(x,y)}. \tag{3.1}$$

is finite. Note then that we have

$$|f(x) - f(y)| \le s(f) d(x,y), \quad x, y \in E,$$

and so f is necessarily uniformly continuous.

Example. For any $A \subset E$, $d(x,A)$ is Lipschitz continuous in x with $s(d(\cdot, A)) = 1$.

Proposition 1.3.2 *If f and g are Lipschitz continuous, then $f \wedge g$ is Lipschitz continuous, with $s(f \wedge g) \le s(f) \vee s(g)$.*

Proof. Since

$$\max\{|f(x) - f(y)|, |g(x) - g(y)|\} \le [s(f) \vee s(g)] d(x,y),$$

it will be enough to verify that

$$|f \wedge g(x) - f \wedge g(y)| \le \max\{|f(x) - f(y)|, |g(x) - g(y)|\} \tag{3.2}$$

for all x and y. There are four cases to be examined. In the first case, suppose that x and y are such that $f(x) \le g(x)$ and $f(y) \ge g(y)$. Then $f \wedge g(x) - f \wedge g(y) = f(x) - g(y)$ and

$$f(x) - f(y) \le f(x) - g(y) \le g(x) - g(y),$$

which leads to (3.2). The other three cases are just as easily checked, so we conclude that (3.2) holds for all x and y. ∎

Proposition 1.3.3 *Let G be an open set and define*

$$f_n(x) = \min\{1, n d(x, E \setminus G)\}, \quad x \in E, n \ge 1.$$

Then f_n is Lipschitz continuous for all n and $f_n \uparrow I_G$ as $n \to \infty$.

Proof. The fact that f_n is Lipschitz continuous follows from the previous proposition. Certainly $f_n \le f_{n+1} \le I_G$ for all n. If $x \in G$, then $d(x, E \setminus G) > 0$ and we can choose n_0 such that $n_0 d(x, E \setminus G) \ge 1$. Then $n d(x, E \setminus G) \ge 1$ and therefore $f_n(x) = 1$ for all $n \ge n_0$. ∎

Proposition 1.3.4 *Every $\mu \in \mathcal{P}(E)$ is regular; that is, we have*

$$\mu(A) = \sup\{\mu(F) | F \subset A, F \text{ is closed}\}, \quad \mu(A) = \inf\{\mu(G) | A \subset G, G \text{ is open}\}$$

for every $A \in \mathcal{E}$.

Proof. Let

$$\mathcal{A} = \{A \in \mathcal{E} | \text{for every } \epsilon > 0 \text{ there are closed } F \text{ and open } G \text{ with } F \subset A \subset G \text{ and } \mu(G \setminus F) < \epsilon\}.$$

It suffices to show that $\mathcal{A} = \mathcal{E}$. If A is open and $\epsilon > 0$, we take $G = A$ and define $F_n = \{x \in E | f_n(x) = 1\}$, where f_n is as defined in proposition 1.3.3 and $n \geq 1$. F_n is closed for all n and since $F_n \uparrow A$, then $\mu(F_n) \uparrow \mu(A)$ as $n \to \infty$, so there is an n such that $\mu(A \setminus F_n) < \epsilon$. Thus $\mathcal{A}$ contains all open sets. We now show that $\mathcal{A}$ is a σ-algebra. $\mathcal{A}$ is obviously closed under complementation. Suppose $\{A_n, n \geq 1\} \subset \mathcal{A}$, and let $\{F_n, n \geq 1\}$ and $\{G_n, n \geq 1\}$ be closed and open sets respectively such that $F_n \subset A_n \subset G_n$ and $\mu(G_n \setminus F_n) \leq \epsilon/2^{n+1}$ for each n. Define $G = \cup_{n\geq 1} G_n$. Let N be large enough that $\mu(\cup_{n\geq 1} F_n) - \mu(\cup_{n\leq N} F_n) < \epsilon/2$, and set $F = \cup_{n=1}^N F_n$. Then F is closed, G is open, $F \subset \cup_{n\geq 1} A_n \subset G$,

$$\begin{aligned}\mu(G \setminus F) = \mu(G) - \mu(F) &= [\mu(G) - \mu(\cup_{n=1}^\infty F_n)] + [\mu(\cup_{n=1}^\infty F_n) - \mu(\cup_{n\leq N} F_n)] \\ &\leq \mu(G \setminus \cup_{n=1}^\infty F_n) + \epsilon/2 \leq \sum_{n=1}^\infty \mu(G_n \setminus F_n) + \epsilon/2 \leq \epsilon,\end{aligned}$$

and so $\cup_{n\geq 1} A_n \in \mathcal{A}$. (We used the fact that $\cup_{n=1}^\infty G_n \setminus \cup_{n=1}^\infty F_n \subset \cup_{n=1}^\infty (G_n \setminus F_n)$.) ∎

Proposition 1.3.5 *Let $\mu, \nu \in \mathcal{P}(E)$ be such that*

$$\int f\, d\mu = \int f\, d\nu$$

for every $f \in \mathcal{C}_b(E)$ (or only every Lipschitz function in $\mathcal{C}_b(E)$). Then $\mu = \nu$.

Proof. Let G be an open set. Let $f_n(x)$ be as in proposition 1.3.3. Then by the monotone convergence theorem, we have

$$\mu(G) = \lim_{n\to\infty} \int f_n\, d\mu = \lim_{n\to\infty} \int f_n\, d\nu = \nu(G).$$

Hence μ and ν agree on open sets. By regularity, μ and ν agree on $\mathcal{E}$. (Alternatively, since μ and ν agree on the class $\mathcal{C}$ of open sets which is closed under finite intersection and generates $\mathcal{E}$, then $\mu = \nu$ on $\mathcal{E}$.) ∎

Definition. $\mu \in \mathcal{P}(E)$ is said to be *tight* if for every $\epsilon > 0$ there is a compact set K such that $\mu(K) \geq 1 - \epsilon$.

Proposition 1.3.6 *If E is either (i) σ-compact (i.e. a countable union of compact sets), or (ii) separable and complete, then every $\mu \in \mathcal{P}(E)$ is tight.*

Proof. Let $\epsilon > 0$. If E is σ-compact, then $E = \cup_{n=1}^\infty K_n$ where each K_n is compact. Let N be such that $\mu(\cup_{n=1}^N K_n) > 1 - \epsilon$. Then $K = \cup_{n=1}^N K_n$ is compact and $\mu(K) > 1 - \epsilon$. Next, suppose E is separable and complete. Since E is separable, then for each n there is a sequence $A_{n1}, A_{n2}, \ldots$ of open balls of radius $1/n$ that cover E. For each n, let i_n be such that $\mu(\cup_{i=1}^{i_n} A_{ni}) > 1 - \epsilon/2^n$. The set $A = \cap_{n=1}^\infty \cup_{i=1}^{i_n} A_{ni}$ is totally bounded and $\mu(A) > 1 - \epsilon$. Let $K = \bar{A}$. Then K is compact (see Dugundji page 298), and $\mu(K) > 1 - \epsilon$. ∎

1.4 Weak convergence.

In this section, E is a metric space with metric d.

Definition. Let $\{\mu_n, n \geq 1\} \subset \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$. We say that the sequence $\{\mu_n, n \geq 1\}$ *converges weakly* to μ, and write $\mu_n \xrightarrow{w} \mu$, if $\int f\, d\mu_n \to \int f\, d\mu$ for all $f \in \mathcal{C}_b(E)$.

Comparison with convergence in total variation. Suppose that $\{\mu_n, n \geq 1\} \subset \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$ are such that $\mu_n \to \mu$ in total variation. Then for any $f \in \mathcal{C}_b(E)$, we have

$$| \int f d\mu_n - \int f d\mu | = | \int f d(\mu_n - \mu) | = \|f\| | \int (f/\|f\|) d(\mu_n - \mu) | \leq \|f\| \|\mu_n - \mu\|,$$

and so convergence in total variation norm implies weak convergence. The converse is not true, as the following example shows.

Example. Suppose $\{x_n, n \geq 1\} \subset E$ and $x \in E$ are such that $d(x_n, x) \to 0$ as $n \to \infty$. Then

$$\int f \, d\epsilon_{x_n} = f(x_n) \to f(x) = \int f \, d\epsilon_x$$

for all $f \in \mathcal{C}_b(E)$, and so $\epsilon_{x_n} \xrightarrow{w} \epsilon_x$. But recall that $\|\epsilon_{x_n} - \epsilon_x\| = 2$ if $x_n \neq x$.

Example. Suppose that E is countable, and is considered with the metric

$$d(x, y) = \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y. \end{cases}$$

The topology generated is the discrete topology in which every subset of E is open. The Borel σ-algebra $\mathcal{E}$ then consists of all subsets of E. Suppose that $\{\mu_n, n \geq 1\}$ and μ are probabilities on $\mathcal{E}$, and that $\mu_n \xrightarrow{w} \mu$. Since each singleton set is both open and closed, the function $f = I_{\{x\}}$ is continuous and bounded, so $\mu_n(\{x\}) \to \mu(\{x\})$ for every $x \in E$. By the proposition immediately following Sheffé's theorem in the previous section, this implies that $\mu_n \to \mu$ in total variation. Thus weak convergence and convergence in total variation are equivalent for probability measures on countable state spaces.

Background. There is a strong connection between weak convergence, as just defined, and what is called weak*-convergence in normed linear spaces. Specifically, let X be a normed linear space, and let X^* denote the set of all continuous linear functionals on X. Recall that X^* is also a normed linear space with norm

$$\|x^*\| = \sup_{\substack{x \in X \\ \|x\| \leq 1}} |x^*(x)|, \quad x^* \in X^*.$$

We say that a sequence $\{x_n^*, n \geq 1\}$ in X^* converges in norm to $x^* \in X^*$ if $\|x_n^* - x^*\| \to 0$ as $n \to \infty$. We say that a sequence $\{x_n^*, n \geq 1\}$ in X^* is weak*-convergent (or converges vaguely) to $x^* \in X^*$ if $x_n^*(x) \to x^*(x)$ for every $x \in X$. Since

$$|x_n^*(x) - x^*(x)| \leq \|x_n^* - x^*\| \|x\|,$$

it follows that if $x_n^* \to x^*$ in norm, then $x_n^* \to x^*$ vaguely. Moreover, by Alaoglu's theorem, the closed unit ball $\{x^* \in X^* | \|x^*\| \leq 1\}$ is compact in the weak* topology.

Now let us get more specific. Suppose E is a compact Hausdorff space, and let $X = \mathcal{C}_b(E)$. Then X^* can be identified by the isometry

$$x^*(f) = \int f \, d\mu, \quad f \in X,$$

with the space $\mathcal{M}_b(E)$ of all finite regular signed measures μ on $\mathcal{E}$, the Borel subsets of E. It is obvious that weak* convergence in X^* here is just weak convergence as has been defined, and it can be shown that convergence in norm in X^* is simply convergence in total variation.

If E is a metric space and $X = \mathcal{C}_b(E)$, then X^* can be identified with the space of all finitely additive regular set functions μ on $\mathcal{E}$, by the same isometry as above, with the same characterizations of weak* and norm convergence.

Definition. Let $\mu \in \mathcal{P}(E)$. $A \in \mathcal{E}$ is called a continuity set for μ if $\mu(\partial A) = 0$. (Note: $\partial A = \bar{A} \setminus A^o$.)

The following theorem gives three equivalent definitions of weak convergence.

Theorem 1.4.1 (Portmanteau Theorem) *Let $\{\mu_n, n \geq 1\} \subset \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$. The following statements are equivalent.*

(1) $\mu_n \xrightarrow{w} \mu$,

(2) $\int f\, d\mu_n \to \int f\, d\mu$ *for all* $f \in \mathcal{C}_b(E)$ *which are Lipschitz continuous,*

(3) $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *for all open sets* G,

(4) $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for all closed sets* F,

(5) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all* μ*-continuity sets* $A \in \mathcal{E}$.

Proof.

(1) $\Rightarrow$ (2). Obvious.

(2) $\Rightarrow$ (3). Let G be open and let $\epsilon > 0$. Let $f_m(x)$ be as defined in proposition 1.3.3, and let $F_m = \{x \in E : f_m(x) = 1\}$. Then each F_m is closed and $F_m \uparrow G$. Let m be large enough that $\mu(F_m) \geq \mu(G) - \epsilon$. Then

$$\liminf_{n\to\infty} \mu_n(G) \geq \lim_{n\to\infty} \int f_m\, d\mu_n = \int f_m\, d\mu \geq \mu(F_m) \geq \mu(G) - \epsilon.$$

(3) $\Leftrightarrow$ (4). This is obvious by taking complements.

(4) $\Rightarrow$ (5). For any $A \in \mathcal{E}$, we have

$$\mu(A^o) \leq \liminf_{n\to\infty} \mu_n(A^o) \leq \liminf_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\bar{A}) \leq \mu(\bar{A}).$$

If A is a continuity set, then $\mu(A^o) = \mu(\bar{A})$, so $\mu_n(A) \to \mu(A)$ and (5) holds.

(5) $\Rightarrow$ (1). Let $f \in \mathcal{C}_b(E)$. By scaling and translation, we can assume that $0 < f < 1$. The sets $F_y = \{x \in E | f(x) = y\}$ are pairwise disjoint for different values of $y \in \mathbb{R}$, so $\mu(F_y) > 0$ for only countably many $y \in \mathbb{R}$ (because $\{y \in \mathbb{R} | \mu(F_y) > 0\} = \cup_{n=1}^{\infty} \{y | \mu(F_y) > 1/n\}$ and the sets in the union are finite). It follows that $\{f > y\}$, which has boundary in F_y, is a continuity set for all but countably many y. Hence $\mu_n\{f > y\} \to \mu\{f > y\}$ a.e (dy), and so by the BCT, $\int f d\mu_n = \int_0^\infty \mu_n\{f > y\} dy \to \int_0^\infty \mu\{f > y\} dy = \int f d\mu$.

∎

Example. Let $E = \mathbb{R}$. Let $G = (0, \infty)$ (open) and let $F = (-\infty, 0]$ (closed). Let $\mu_n = \epsilon_{1/n}$, and let $\mu = \epsilon_0$, so that $\mu_n \xrightarrow{w} \mu$. Note that

(1) $\liminf_{n\to\infty} \mu_n(G) = 1$, while $\mu(G) = 0$,

(2) $\limsup_{n\to\infty} \mu_n(F) = 0$, while $\mu(F) = 1$.

Remark. Yet another equivalent condition is the following:

(6) $\liminf_{n\to\infty} \int g\, d\mu_n \geq \int g\, d\mu$ for all bounded lower semicontinuous functions $g : E \to \mathbb{R}$.(Note: g is lower semicontinuous if $\{g > a\}$ is open for any $a \in \mathbb{R}$.)

For if (6) holds then (3) holds since I_G is lower semicontinuous whenever G is open. Conversely, if g is lower semicontinuous and bounded, there is a sequence $\{f_n, n \geq 1\} \subset \mathcal{C}_b(E)$ such that $f_n \uparrow g$. Since $\int g\, d\mu_n \geq \int f_m\, d\mu_n$ for every n, then $\liminf_{n\to\infty} \int g\, d\mu_n \geq \int f_m\, d\mu$. Letting $m \to \infty$ gives (6).

Proposition 1.4.2 *Let $\{\mu_n, n \geq 1\} \subset \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$. Let $\mathcal{U}$ be a subclass of $\mathcal{E}$ such that*

(1) $\mathcal{U}$ is closed under finite intersections,

(2) each open set in $\mathcal{E}$ is a countable union of members of $\mathcal{U}$.

If $\mu_n(A) \to \mu(A)$ for all $A \in \mathcal{U}$, then $\mu_n \xrightarrow{w} \mu$.

Proof. Suppose that $A_1, \dots, A_m \in \mathcal{U}$. Then by the inclusion-exclusion principle, we have

$$\begin{aligned}\mu_n(\cup_{i=1}^m A_i) &= \sum_i \mu_n(A_i) - \sum_{i,j} \mu_n(A_i \cap A_j) + \sum_{i,j,k} \mu_n(A_i \cap A_j \cap A_k) - \dots \\ &\to \sum_i \mu(A_i) - \sum_{i,j} \mu(A_i \cap A_j) + \sum_{i,j,k} \mu(A_i \cap A_j \cap A_k) - \dots \\ &= \mu(\cup_{i=1}^m A_i).\end{aligned}$$

Thus we have $\mu_n(A) \to \mu(A)$ whenever A is a finite union of members of $\mathcal{U}$. Let G be an open set, and let $\epsilon > 0$. Then there is a sequence $\{A_i, i \geq 1\} \subset \mathcal{U}$ such that $\cup_{i=1}^\infty A_i = G$. Let m be large enough that $\mu(\cup_{i=1}^m A_i) > \mu(G) - \epsilon$. Then

$$\mu(G) - \epsilon < \mu(\cup_{i=1}^m A_i) = \lim_{n\to\infty} \mu_n(\cup_{i=1}^m A_i) \leq \liminf_{n\to\infty} \mu_n(G).$$

Since this is true for all $\epsilon > 0$, we have $\mu(G) \leq \liminf_{n\to\infty} \mu_n(G)$. ∎

Definition. Let $\{F_n, n \geq 1\}$ and F be distribution functions on $\mathbb{R}^d$. We say that F_n converges to F in distribution, and write $F_n \xrightarrow{d} F$, if $F_n(x) \to F(x)$ at every point x of continuity of F.

Example. Let F be a distribution function, and define $F_n(x) = F(x - 1/n)$. Then $F_n(x) \to F(x)$ if and only if x is a point of continuity of F.

Theorem 1.4.3 (Helly-Bray Theorem) *Let $\mu_n, n \geq 1$ and μ be laws on $\mathbb{R}^d$ with distribution functions $F_n, n \geq 1$ and F respectively. Then $\mu_n \xrightarrow{w} \mu$ if and only if $F_n \xrightarrow{d} F$.*

Proof. We give the proof only for $d = 1$. Suppose first that $\mu_n \xrightarrow{w} \mu$. If x is a point of continuity of F, then $\mu(\{x\}) = 0$; it follows that $(-\infty, x]$ is a continuity set for μ, so $F_n(x) = \mu_n(-\infty, x] \to \mu(-\infty, x] = F(x)$.

Conversely, suppose that $F_n(x) \to F(x)$ for all x at which F is continuous. Let

$$\mathcal{U} = \{(a, b] | a, b \text{ are points of continuity of } F\}.$$

If $(a, b] \in \mathcal{U}$, then $\mu_n(a, b] = F_n(b) - F_n(a) \to F(b) - F(a) = \mu(a, b]$. By the previous proposition, we need only show that $\mathcal{U}$ satisfies conditions (1) and (2) there. Certainly $\mathcal{U}$ is closed under finite intersection. Since F can have only countably many discontinuities, then points of continuity of F are dense in $\mathbb{R}$, and it follows that the second condition on $\mathcal{U}$ is satisfied. ∎

Definition. A family $\mathcal{A} \subset \mathcal{P}(E)$ is said to be *uniformly tight* (or just tight) if for every $\epsilon > 0$, there is a compact set $K \subset E$ such that $\mu(K) > 1 - \epsilon$ for every $\mu \in \mathcal{A}$.

Theorem 1.4.4 (Helly's Selection Theorem) *Let $\{F_n, n \geq 1\}$ be a sequence of distribution functions on $\mathbb{R}$. Then there exists a subsequence $\{F_{n_i}, i \geq 1\}$ and a possibly substochastic distribution function F such that $F_{n_i} \xrightarrow{d} F$ as $i \to \infty$. If the sequence $\{F_n, n \geq 1\}$ is tight, then F is a proper distribution function.*

Proof. We shall use the diagonalization procedure. Let $\{r_1, r_2, \dots\}$ be an enumeration of the rational numbers. Since $\{F_n(r_1), n \geq 1\} \subset [0,1]$, there is a subsequence $\{n_i^1, i \geq 1\}$ and a number $G(r_1)$ such that $F_{n_i^1}(r_1) \to G(r_1)$. By the same token, there is a subsequence $\{n_i^2, i \geq 1\}$ of $\{n_i^1, i \geq 1\}$ such that $F_{n_i^2}(r_2) \to G(r_2)$, and so on. This defines a non-decreasing function $G(r), r \in \mathbb{Q}$. Define $n_i = n_i^i, i \geq 1$, and $F(x) = \inf\{G(r) | r \text{ is rational and } r > x\}, x \in \mathbb{R}$. Then $\{F_{n_i}, i \geq 1\}$ and F are as required. Note that $F_{n_i}(r) \to G(r)$ for every rational number r.

(1) F is a distribution function. For certainly F is non-decreasing. Suppose that $\{x_n, n \geq 1\}$ are real numbers such that $x_n \downarrow x$. Given $\epsilon > 0$, let $r > x$ be rational such that $F(x) > G(r) - \epsilon$, and let N be such that $x_n < r$ for all $n \geq N$. Then $F(x) > G(r) - \epsilon \geq F(x_n) - \epsilon$ for all $n \geq N$, implying that $F(x) \geq \lim_{n\to\infty} F(x_n)$. Since F is non-decreasing, we certainly have $\lim_{n\to\infty} F(x_n) \geq F(x)$. We have therefore shown that F is right continuous.

(2) $F_{n_i} \xrightarrow{d} F$. For let $x \in \mathbb{R}$ be a point of continuity of F. Given $\epsilon > 0$, let $s \in \mathbb{Q}$ with $x < s$ be such that $G(s) - \epsilon \leq F(x)$ (by def'n of F). Let $r' \in \mathbb{Q}$ be such that $r' < x$ and $F(r') + \epsilon > F(x)$ (since $F(x^-) = F(x)$), and let $r \in \mathbb{Q}$ be such that $r' < r < x$. Then

$$F(x) - \epsilon \leq F(r') \leq G(r) = \lim_{i\to\infty} F_{n_i}(r) \leq \liminf_{i\to\infty} F_{n_i}(x) \leq \limsup_{i\to\infty} F_{n_i}(x) \leq \lim_{i\to\infty} F_{n_i}(s) = G(s) \leq F(x) + \epsilon.$$

Since this is true for all $\epsilon > 0$, we deduce that $\lim_{i\to\infty} F_{n_i}(x) = F(x)$.

Suppose the sequence $\{F_n, n \geq 1\}$ is tight. Let $\epsilon > 0$. Then there is an interval $[a,b]$ such that $F_n[a,b] \geq 1-\epsilon$ for all $n \geq 1$. Let x and y be continuity points of F such that $x < a$ and $y > b$. Then $F_{n_i}(x,y] \geq F_{n_i}[a,b] \geq 1 - \epsilon$ for all n. Letting $i \to \infty$ shows that $F(x,y] \geq 1 - \epsilon$. Since $\epsilon > 0$ was arbitrary, we conclude that $F(\mathbb{R}) = 1$. ∎

Example. Define

$$F_n(x) = aI_{\{x \geq n\}} + bI_{\{x \geq -n\}} + cG(x), \quad x \in \mathbb{R}, n \geq 1,$$

where G is a distribution function and a, b, c are positive numbers such that $a+b+c = 1$. Let $F(x) = b+cG(x)$. Then $F_n(x) \to F(x)$ for all x. However, F may not be a proper distribution function, since $F(+\infty) = b + c$ and $F(-\infty) = b$.

Proposition 1.4.5 *Let $\{\mu_n, n \geq 1\}$ be probability measures on $\mathbb{R}$. If there is a function $\phi : \mathbb{R} \to [0,\infty)$ such that $\phi(x) \to \infty$ as $|x| \to \infty$, and such that $C = \sup_n \int \phi(x)\, d\mu_n < \infty$, then $\{\mu_n, n \geq 1\}$ is tight.*

Proof. We have

$$C \geq \int_{|x|\geq M} \phi(x)\, d\mu_n \geq \inf_{|x|\geq M} \phi(x) \cdot \mu_n\{|x| \geq M\},$$

and so

$$\sup_n \mu_n\{|x| \geq M\} \leq \frac{C}{\inf_{|x|\geq M} \phi(x)}.$$

Now given $\epsilon > 0$, we can choose M large enough that the right-hand side is less than ϵ. ∎

Definition. A family $\mathcal{A} \subset \mathcal{P}(E)$ is said to be *relatively compact* if for every sequence $\{\mu_n, n \geq 1\} \subset \mathcal{A}$, there is a subsequence $\{\mu_{n_i}, i \geq 1\}$ and a $\mu \in \mathcal{P}(E)$ such that $\mu_{n_i} \xrightarrow{w} \mu$ as $i \to \infty$.

Remark. If $\{\mu_n, n \geq 1\}$ is a relatively compact sequence of probability measures, and if every weakly convergent subsequence has the same limit μ, then $\mu_n \xrightarrow{w} \mu$. For suppose that $\mu_n \stackrel{w}{\not\to} \mu$. Then there is $f \in \mathcal{C}_b(E)$ such that $\int f\,d\mu_n \not\to \int f\,d\mu$, and therefore an $\epsilon > 0$ and a subsequence $\{n_i, i \geq 1\}$ such that $|\int f\,d\mu_{n_i} - \int f\,d\mu| > \epsilon$ for all $i \geq 1$. This contradicts the fact that $\{\mu_{n_i}, i \geq 1\}$ must have a subsequence converging weakly to μ.

Proposition 1.4.6 *Let $\mathcal{A} \subset \mathcal{P}(\mathbb{R})$. Then $\mathcal{A}$ is relatively compact if and only if it is uniformly tight.*

Proof. Suppose $\mathcal{A}$ is uniformly tight. Let $\{\mu_n, n \geq 1\} \subset \mathcal{A}$, and let $\{F_n, n \geq 1\}$ be the corresponding distribution functions. By the Helly selection theorem, there is a subsequence $\{F_{n_i}, i \geq 1\} \subset \{F_n, n \geq 1\}$ and a proper (here is where uniform tightness is used) distribution function F such that $F_{n_i}(x)$ converges to $F(x)$ at all points of continuity x of F. Let μ be the probability corresponding to F. Then $\mu_{n_i} \xrightarrow{w} \mu$ as $i \to \infty$. Thus $\mathcal{A}$ is relatively compact.

Suppose $\mathcal{A}$ is not tight. Then there is an $\epsilon > 0$ such that for every compact set K, there is $\mu \in \mathcal{A}$ such that $\mu(K) \leq 1-\epsilon$. For this ϵ, choose for every integer $n \geq 1$ a probability $\mu_n \in \mathcal{A}$ such that $\mu_n(-n, n] \leq 1-\epsilon$. If $\mathcal{A}$ were relatively compact, the sequence $\{\mu_n, n \geq 1\}$ would have a weakly convergent subsequence $\{\mu_{n_i}, i \geq 1\}$, converging say to $\mu \in \mathcal{P}(E)$. Then for every $x \in \mathbb{R}$, $(-x, x) \subset (-n_i, n_i]$ for all large enough i, and so we would have

$$\mu(-x, x) \leq \liminf_{i\to\infty} \mu_{n_i}(-x, x) \leq \liminf_{i\to\infty} \mu_{n_i}(-n_i, n_i] \leq 1 - \epsilon,$$

which would imply $\mu(\mathbb{R}) < 1$. ∎

Problem. Show that if $F_n \xrightarrow{d} F$, where F is continuous, then $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0$ as $n \to \infty$.

In general, the previous proposition is true for any metric space E, as follows.

Theorem 1.4.7 (Prohorov's Theorem) *Let E be a metric space, and let $\mathcal{A} \subset \mathcal{P}(E)$.*

(1) If $\mathcal{A}$ is uniformly tight, then it is relatively compact.

(2) If $\mathcal{A}$ is relatively compact, and if E is separable and complete, then $\mathcal{A}$ is uniformly tight.

Proof. See Billingsley, page 37.

Proposition 1.4.8 *Let $(E, \mathcal{E})$ and $(E', \mathcal{E}')$ be measurable spaces, and let $h : E \to E'$ be a measurable function. Let μ be a probability on $\mathcal{E}$. Define*

$$\mu h^{-1}(A) = \mu[h^{-1}(A)], \quad A \in \mathcal{E}'.$$

Then μh^{-1} is a probability on $\mathcal{E}'$, and for any measurable function $f : E' \to \mathbb{R}$, we have

$$\int f[h(x)]\mu(dx) = \int f(x')\mu h^{-1}(dx'),$$

whenever one, and then both, of these integrals exist.

Proof. If $f = I_A$ where $A \in \mathcal{E}'$, then $\int f[h(x)]\mu(dx) = \mu[h^{-1}(A)] = \mu h^{-1}(A) = \int f(x')\mu h^{-1}(dx')$, so the integral equality is true when f is an indicator function. The proof may be completed via the "usual"method. ∎

Theorem 1.4.9 (The Mapping Theorem) *Let (E, d) and (E', d') be metric spaces, and let $h : E \to E'$ be a measurable function. Let $D_h \subset E$ be the set of points in E at which h is discontinuous. Suppose that $\{\mu_n, n \geq 1\} \subset \mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$ are such that $\mu_n \xrightarrow{w} \mu$ and $\mu(D_h) = 0$. Then $\mu_n h^{-1} \xrightarrow{w} \mu h^{-1}$.*

Proof. Define

$$A_{\epsilon,\delta} = \cup_{\substack{y,z\in E\\ d'[h(y),h(z)]\geq\epsilon}} \{x \in E | d(x,y) < \delta\} \cap \{x \in E | d(x,z) < \delta\},$$

where $\epsilon > 0$ and $\delta > 0$. Then $A_{\epsilon,\delta}$ is the union of open sets, and so is open. Moreover, we have

$$D_h = \cup_\epsilon \cap_\delta A_{\epsilon,\delta},$$

where the union is over all rational $\epsilon > 0$ and the intersection over all rational $\delta > 0$. It follows that $D_h \in \mathcal{E}$.

Let F be a closed subset of E'. Since $\overline{h^{-1}(F)} \subset D_h \cup h^{-1}(F)$, then $\mu[\overline{h^{-1}(F)}] \leq \mu(D_h) + \mu[h^{-1}(F)] = \mu[h^{-1}(F)]$, and so $\mu[\overline{h^{-1}(F)}] = \mu[h^{-1}(F)]$. Hence

$$\limsup_{n\to\infty} \mu_n h^{-1}(F) = \limsup_{n\to\infty} \mu_n [h^{-1}(F)] \leq \limsup_{n\to\infty} \mu_n [\overline{h^{-1}(F)}] \leq \mu[\overline{h^{-1}(F)}] = \mu[h^{-1}(F)] = \mu h^{-1}(F).$$

∎

Corollary 1.4.10 *Suppose that $\mu_n \xrightarrow{w} \mu$ and that $h : E \to \mathbb{R}$ is a bounded measurable function with $\mu(D_h) = 0$. Then $\int h\, d\mu_n \to \int h\, d\mu$.*

Proof. We apply the above proposition with $E' = \mathbb{R}$. Suppose that $|h| \leq M$, and define

$$f(x) = \begin{cases} -M & \text{if } x \leq -M, \\ x & \text{if } -M \leq x \leq M, \\ M & \text{if } x \geq M. \end{cases}$$

Then $f \in \mathcal{C}_b(E)$ and so

$$\int h\, d\mu_n = \int f[h(x)]\, d\mu_n = \int f(x) d\mu_n h^{-1} \to \int f(x) d\mu h^{-1} = \int f[h(x)]\, d\mu = \int h\, d\mu$$

as $n \to \infty$. ∎

1.5 Convergence in Distribution.

Definition. Let $\{\Omega, \mathcal{F}, P\}$ be a probability space, and let $(E, \mathcal{E})$ be a measurable space. Let X be a random variable defined on Ω, and taking values in E. Then the probability measure

$$P_X(A) = PX^{-1}(A) = P\{X \in A\}, \quad A \in \mathcal{E},$$

is called the *distribution* or *law* of X. Observe that for any measurable function $f : E \to \mathbb{R}$, we have

$$Ef(X) = \int f(x)\, dP_X,$$

whenever one, and therefore both, of the integrals exist.

From now on, (E, d) is a metric space with metric d.

Definition. Let $\{X_n, n \geq 1\}$ be a sequence of random variables all taking values in the metric space E (but not necessarily defined on the same probability space), and let X be a random variable with values in E. For each n, let μ_{X_n} denote the distribution of X_n on E, and let μ_X denote the distribution of X. Then we say the sequence $\{X_n, n \geq 1\}$ *converges in distribution* to X, and write $X_n \xrightarrow{d} X$, if $\mu_{X_n} \xrightarrow{w} \mu_X$. A set $A \in \mathcal{E}$ is a continuity set for X if $P\{X \in \partial A\} = 0$.

The following is a direct translation of the Portmanteau theorem.

Theorem 1.5.1 (Portmanteau Theorem) *Let $\{X_n, n \geq 1\}$ and X be random variables with values in E. The following statements are equivalent.*

(1) $X_n \xrightarrow{d} X$,

(2) $Ef(X_n) \to Ef(X)$ *for all* $f \in \mathcal{C}_b(E)$ *which are Lipschitz continuous,*

(3) $\liminf_{n\to\infty} P\{X_n \in G\} \geq P\{X \in G\}$ *for all open sets* G,

(4) $\limsup_{n\to\infty} P\{X_n \in F\} \leq P\{X \in F\}$ *for all closed sets* F,

(5) $\lim_{n\to\infty} P\{X_n \in A\} = P\{X \in A\}$ *for all* X*-continuity sets* $A \in \mathcal{E}$.

Definition. Let $\{X_n, n \geq 1\}$ and X be random variables defined on the same probability space $\{\Omega, \mathcal{F}, P\}$ and taking values in the metric space E. We say that $\{X_n, n \geq 1\}$ *converges in probability* to X and write $X_n \xrightarrow{\mathrm{P}} X$, if for every $\epsilon > 0$, we have

$$P\{d(X_n, X) \geq \epsilon\} \to 0 \text{ as } n \to \infty.$$

Proposition 1.5.2 *(1) If* $X_n \to X$ *in probability, then* $X_n \xrightarrow{d} X$.

(2) If $X_n \xrightarrow{d} a$ *(where* a *is a constant), then* $X_n \to a$ *in probability.*

Proof.

(1) Suppose $X_n \overset{d}{\not\to} X$. Then there is $f \in \mathcal{C}_b(E)$ such that $\int f\, d\mu_{X_n} \not\to \int f\, d\mu_X$; that is, there is an $\epsilon > 0$, and a sequence $\{n_i, i \geq 1\}$ such that $|Ef(X_{n_i}) - Ef(X)| = \left|\int f\, d\mu_{X_{n_i}} - \int f\, d\mu_X\right| \geq \epsilon$ for all $i \geq 1$. This contradicts the fact that since $X_{n_i} \to X$ in probability, there is a further subsequence $\{n_{i_j}, j \geq 1\}$ of $\{n_i, i \geq 1\}$ such that $X_{n_{i_j}} \to X$ a.e., and therefore $Ef(X_{n_{i_j}}) \to Ef(X)$.

(2) Let $f_k(x) = \min\{kd(x, a), 1\}$ where $k \geq 1$ is an integer. Then $f_k \in \mathcal{C}_b(E)$ and $I_{\{d(x,a)\geq 1/k\}} \leq f_k(x)$, so

$$P\{d(X_n, a) \geq 1/k\} = EI_{\{d(X_n,a)\geq 1/k\}} \leq E[f_k(X_n)] = \int f_k(x)\, d\mu_{X_n} \to \int f_k(x)\, d\epsilon_a = f_k(a) = 0$$

as $n \to \infty$.

∎

Proposition 1.5.3 *If* $X_n \xrightarrow{\mathrm{d}} X$ *and* $Y_n \xrightarrow{\mathrm{P}} 0$, *then* $X_n + Y_n \xrightarrow{\mathrm{d}} X$ *and* $X_nY_n \xrightarrow{\mathrm{P}} 0$.

Example. Suppose that X and Y are i.i.d. and assume the values 0 and 1 with probabilities 1/2 each. Define $X_n = Y, n \geq 1$. Then $X_n \xrightarrow{d} X$, but $X_n \overset{P}{\not\to} X$, since $P\{|X_n - X| > \epsilon\} = P\{|Y - X| = 1\} = 1/2$ for any $\epsilon > 0$.

The following are consequences of corollary 1.4.10

Proposition 1.5.4 *Take* $E = \mathbb{R}$. *Suppose that* $X_n \xrightarrow{d} X$.

(1) Then $E|X| \leq \liminf_{n\to\infty} E|X_n|$.

(2) If $\{X_n, n \geq 1\}$ *is uniformly integrable, then* $E(X_n) \to E(X)$ *as* $n \to \infty$.

Proof. Let $D = \{x \in \mathbb{R} | P\{|X| = x\} = 0\} = \{\text{points of continuity of } F_{|X|}\}$. D is dense in $\mathbb{R}$.

(1) For $a \in D$, define

$$h(x) = \begin{cases} |x| & \text{if } |x| \leq a, \\ 0 & \text{if } |x| > a. \end{cases}$$

Then $\mu_X(D_h) = P\{|X| = a\} = 0$, so

$$\int_{\{|X|\leq a\}} |X| dP = Eh(X) = \lim_{n\to\infty} Eh(X_n) = \liminf_{n\to\infty} Eh(X_n) \leq \liminf_{n\to\infty} E|X_n|.$$

The proof is completed by letting $a \to \infty$ through D.

(2) Since $\{X_n, n \geq 1\}$ is uniformly integrable, then by part (1) $E|X| < \infty$. For $a \in D$, define

$$h(x) = \begin{cases} x & \text{if } |x| \leq a, \\ 0 & \text{if } |x| > a, \end{cases}$$

so that $Eh(X_n) \to Eh(X)$. Since

$$\begin{aligned} |E(X_n) - E(X)| &= |Eh(X_n) + \int_{\{|X_n| \geq a\}} X_n \, dP - Eh(X) - \int_{\{|X| \geq a\}} X \, dP| \\ &\leq |Eh(X_n) - Eh(X)| + \int_{\{|X_n| \geq a\}} |X_n| \, dP + \int_{\{|X| \geq a\}} |X| \, dP, \end{aligned}$$

then

$$\limsup_{n\to\infty} |E(X_n) - E(X)| \leq \sup_n \int_{\{|X_n| \geq a\}} |X_n| \, dP + \int_{\{|X| \geq a\}} |X| \, dP.$$

Letting $a \to \infty$ through D now gives $\limsup_{n\to\infty} |E(X_n) - E(X)| = 0$.

∎

Corollary 1.5.5 *Take $E = \mathbb{R}$. Suppose that $g, h : \mathbb{R} \to \mathbb{R}$ are functions such that h is continuous, g is measurable with $g(x) \geq 0$ for all x, and $|h(x)|/g(x) \to 0$ as $|x| \to \infty$. If $X_n \xrightarrow{d} X$, and if $\limsup_{n\to\infty} Eg(X_n) < \infty$, then $Eh(X_n) \to Eh(X)$ as $n \to \infty$.*

Proof. Since $X_n \xrightarrow{d} X$ and h is continuous, then $h(X_n) \xrightarrow{d} h(X)$. Let N be such that

$$C \stackrel{\text{def}}{=} \sup_{n \geq N} \int g(x) \, d\mu_n < \infty.$$

Then $\{h(X_n), n \geq N\}$ is uniformly integrable. For given $\epsilon > 0$, let M be large enough that $|h(x)|/g(x) < \epsilon/C$ for $|x| > M$, and let $\alpha = \sup\{|h(x)| : |x| \leq M\}$. Then if $\beta \geq \alpha$, we have

$$\begin{aligned} \sup_{n \geq N} \int_{\{|h(X_n)| > \beta\}} |h(X_n)| dP &= \sup_{n \geq N} \int_{\{|h(x)| > \beta\}} |h(x)| \, d\mu_n \leq \sup_{n \geq N} \int_{\{|x| > M\}} |h(x)| \, d\mu_n \\ &\leq \frac{\epsilon}{C} \sup_{n \geq N} \int_{\{|x| > M\}} g(x) \, d\mu_n \leq \epsilon. \end{aligned}$$

∎

Chapter 2

Characteristic Functions.

2.1 Properties and Examples.

Definition. Let $(E, \mathcal{E})$ be a measurable space. Let μ be a probability measure, and let $f : E \to \mathbb{C}$ be measurable. Then f can be written as $f = \Re(f) + i\Im(f)$ where $\Re(f)$ (the real part of f) and $\Im(f)$ (the imaginary part of f) are measurable functions. We say that f is μ-integrable, and we define

$$\int f \, d\mu = \int \Re(f) \, d\mu + i \int \Im(f) \, d\mu,$$

if both $\Re(f)$ and $\Im(f)$ are integrable. We leave it to the reader to show the following:

(1) if $f, g : E \to \mathbb{C}$ and if $c \in \mathbb{C}$, then $\int f + g \, d\mu = \int f \, d\mu + \int g \, d\mu$ and $\int cf \, d\mu = c \int f \, d\mu$.

(2) if $f_n : E \to \mathbb{C}, n \geq 1$ and $f : E \to \mathbb{C}$, and if $f_n \to f$ a.s.(μ) and $|f_n| \leq g$ where $\int g \, d\mu < \infty$, then $\int f_n \, d\mu \to \int f \, d\mu$.

Proposition 2.1.1 *If $f : E \to \mathbb{C}$ is μ-integrable, then*

$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu.$$

Proof. Suppose $\int f \, d\mu = re^{i\theta}$. Then $\int e^{-i\theta} f \, d\mu = r = |\int f \, d\mu|$. Writing $f(x) = \rho(x) e^{i\phi(x)}$, we have

$$|\int f \, d\mu| = \int e^{-i\theta} f \, d\mu = \int \rho(x) e^{i(\phi(x) - \theta)} \, d\mu = \int \rho(x) \cos[\phi(x) - \theta] \, d\mu \leq \int \rho(x) \, d\mu = \int |f| \, d\mu.$$

∎

Remark. Let $a < b$ be numbers. We shall frequently use the fact that

$$\left| \frac{e^{-ita} - e^{-itb}}{it} \right| = \left| \int_a^b e^{-itx} \, dx \right| \leq \int_a^b 1 \, dx = b - a.$$

Definition. Let μ be a probability measure on $\mathbb{R}$. Then we define the *characteristic function* $\phi(t)$ of μ to be

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \, d\mu(x). \tag{1.1}$$

Let X be a (real) random variable defined on $(\Omega, \mathcal{F}, P)$. Then the *characteristic function* $\phi_X(t)$ of X is

$$\phi(t) = E(e^{itX}) = E[\cos tX] + iE[\sin tX], \quad t \in \mathbb{R}.$$

Note that if μ_X is the distribution of X on $\mathbb{R}$, then $\phi(t) = \int e^{itx} \, d\mu_X$. If μ_X is absolutely continuous with respect to Lebesgue measure, and has density f, then $\phi(t) = \int e^{itx} f(x) \, dx$, the Fourier transform of f.

Proposition 2.1.2 (Basic Properties) *(1)* $\phi(0) = 1$,

(2) $\overline{\phi(t)} = \phi(-t)$,

(3) $|\phi(t)| \leq 1$,

(4) $\phi(t)$ *is uniformly continuous in* t,

(5) $\phi_{aX+b}(t) = e^{ibt}\phi_X(at)$,

(6) X *has the same distribution as* $-X$ *if and only if* $\phi_X(t)$ *is real,*

(7) $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$ *if* X *and* Y *are independent.*

Proof. All except (4) are pretty obvious. For (4), we have

$$|\phi(t+h) - \phi(t)| = |E[e^{i(t+h)X}] - E[e^{itX}]| = |E[e^{itX}[e^{ihX} - 1]]| \leq E|e^{itX}[e^{ihX} - 1]| = E|e^{ihX} - 1|,$$

so that $\lim_{h\to 0}\sup_{t\in\mathbb{R}} |\phi(t+h) - \phi(t)| \leq \lim_{h\to 0} E|e^{ihX} - 1| = 0$. For (6), note that if $\phi_X(t)$ is real, then by part (1), $\phi_X(t) = \phi_X(-t) = \phi_{-X}(t)\forall t$. Then corollary 2.1.4 below gives $X \overset{d}{=} -X$. ∎

Theorem 2.1.3 (The Inversion Formula) *Let* μ *be a probability measure on* $\mathbb{R}$ *and let* $\phi(t)$ *be the characteristic function of* μ*, as defined in (1.1). Let* $a < b$ *be real numbers. Then*

$$\lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)\,dt = \mu(a,b) + \frac{1}{2}\mu\{a,b\}.$$

In particular, the integral on the left is real for all $T > 0$.

Proof. Taking care with the cosine terms,

$$J_T(x) \overset{\text{def}}{=} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}e^{itx}\,dt = \int_{-T}^{T} \frac{\sin t(x-a)}{t}\,dt - \int_{-T}^{T} \frac{\sin t(x-b)}{t}\,dt = R(x-a,T) - R(x-b,T), \tag{1.2}$$

where

$$R(\theta,T) = \int_{-T}^{T} \frac{\sin t\theta}{t}\,dt = \int_{-T\theta}^{T\theta} \frac{\sin w}{w}\,dw = 2\text{sign}(\theta)\int_0^{T|\theta|} \frac{\sin w}{w}\,dw.$$

In particular, $J_T(x)$ is real. Since $\int_0^\infty \frac{\sin w}{w}\,dw = \pi/2$, then $\lim_{T\to\infty} R(\theta,T) = \pi\text{sign}(\theta)$ (note:sign$(0) = 0$). It follows that $J_T(x)$ is bounded and

$$J_T(x) \to \begin{cases} 0 & \text{if } x < a, \\ \pi & \text{if } x = a, \\ 2\pi & \text{if } a < x < b, \\ \pi & \text{if } x = b, \\ 0 & \text{if } x > b, \end{cases} \qquad = 2\pi I_{(a,b)} + \pi I_{\{a,b\}}.$$

Consequently, by Fubini's theorem and then the BCT, we have

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)\,dt = \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\int e^{itx}\,d\mu\,dt = \int J_T(x)d\mu \to 2\pi\mu(a,b) + \pi\mu\{a,b\}.$$

∎

Corollary 2.1.4 *Let* μ_1 *and* μ_2 *be probability measures with characteristic functions* $\phi_1(t)$ *and* $\phi_2(t)$*. If* $\phi_1 = \phi_2$*, then* $\mu_1 = \mu_2$.

Proof. The sets of point masses (i.e. atoms) for μ_1 and μ_2 must each be countable. Hence there is a dense set $D \subset \mathbb{R}$ such that if $x \in D$, then $\mu_1\{x\} = \mu_2\{x\} = 0$. Consider the family

$$\mathcal{U} = \{(a,b) | a, b \in D, a < b\}.$$

By the above theorem, μ_1 and μ_2 coincide on $\mathcal{U}$. Since $\mathcal{U}$ is closed under finite intersection and generates the Borel σ-algebra, then μ_1 and μ_2 coincide on $\mathcal{B}(\mathbb{R})$. ∎

Theorem 2.1.5 *Let $\phi(t)$ be the characteristic function of the probability measure μ. If $\int |\phi(t)|\,dt < \infty$, then μ is absolutely continuous with respect to Lebesgue measure, with bounded continuous density function*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,dt, \quad x \in \mathbb{R}.$$

Proof. Note first that $\overline{f(x)} = \frac{1}{2\pi}\int e^{itx}\phi(-t)\,dt = f(x)$, so f is a real-valued function.

Suppose $b > a$. Since $\left|\frac{e^{-ita}-e^{-itb}}{it}\phi(t)\right| \le |b-a||\phi(t)|$, the BCT implies that the result of the previous theorem can be written as

$$\mu(a,b) + \frac{1}{2}\mu\{a,b\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{-ita}-e^{-itb}}{it}\phi(t)\,dt. \tag{1.3}$$

Fix $a \in \mathbb{R}$, and suppose $b > a$. Then

$$\begin{aligned}\mu(a,b) + \frac{1}{2}\mu\{a\} \le \mu(a,b) + \frac{1}{2}\mu\{a,b\} &= \frac{1}{2\pi}\left|\int \frac{e^{-ita}-e^{-itb}}{it}\phi(t)\,dt\right| \le \frac{1}{2\pi}\int\left|\frac{e^{-ita}-e^{-itb}}{it}\phi(t)\right|\,dt \\ &\le \frac{b-a}{2\pi}\int |\phi(t)|\,dt.\end{aligned}$$

By taking b very close to a, we see that $\mu\{a\} = 0$. Hence μ has no point masses and for $a < b$, (1.3) becomes

$$\mu(a,b) = \frac{1}{2\pi}\int \frac{e^{-ita}-e^{-itb}}{it}\phi(t)\,dt = \frac{1}{2\pi}\left[\int\int_a^b e^{-itx}\,dx\right]\phi(t)\,dt = \int_a^b\left[\frac{1}{2\pi}\int e^{-itx}\phi(t)\,dt\right]dx = \int_a^b f(x)\,dx.$$

Since this is true for all such open intervals (a,b), it follows that $\mu(A) = \int_A f(x)\,dx$ for all $A \in \mathcal{E}$. ∎

Proposition 2.1.6 *Let ϕ be the characteristic function of the random variable X, and suppose that $E(|X|^n) < \infty$. Then $\phi(t)$ has a continuous derivative of order n given by*

$$\phi^{(n)}(t) = \int_{-\infty}^{\infty} (ix)^n e^{itx}\,d\mu(x). \tag{1.4}$$

Proof. The proposition is certainly true for $n = 0$. Suppose it is true for $n = k$, and that X is a random variable with $E|X^{k+1}| < \infty$. Then also $E|X^k| < \infty$, so (1.4) holds with $n = k$, and then

$$\frac{\phi^{(k)}(t+h) - \phi^{(k)}(t)}{h} = \int (ix)^k e^{itx}\frac{e^{ihx}-1}{h}\,d\mu(x).$$

Now $|(e^{ihx}-1)/h|$ is bounded by $|x|$ and $(e^{ihx}-1)/h \to ix$ as $h \to 0$. Letting $h \to 0$, it follows from the bounded convergence theorem that $\phi^{(k+1)}(t)$ exists and is given by (1.4) with $n = k+1$.

Finally, a simple demonstration using the bounded convergence theorem shows that the right-hand side of (1.4) is a continuous function of t. ∎

Example: The Normal Distribution. Suppose that $X \sim N(0,1)$. That is, X has density function

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad x \in \mathbb{R}.$$

Then

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

By the previous proposition, $\phi'(t)$ exists and is continuous and

$$\phi'(t) = \int_{-\infty}^{\infty} ixe^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = -2\int_0^{\infty} x \sin tx \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

Integrating by parts the last integral gives

$$\phi'(t) = -2t\int_0^{\infty} \cos tx \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = -t\phi(t).$$

This means that

$$[\phi(t)e^{t^2/2}]' = \phi'(t)e^{t^2/2} + \phi(t)te^{t^2/2} = [\phi'(t) + t\phi(t)]e^{t^2/2} = 0,$$

so $\phi(t)e^{t^2/2} = c$ (a contant). Since $\phi(0) = 1$, then $c = 1$. Hence

$$\phi(t) = e^{-t^2/2}.$$

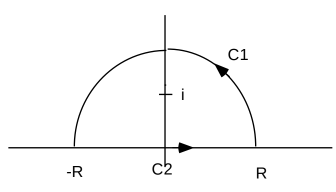
The Cauchy Distribution. A random variable X with density function

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

is said to have the Cauchy distribution on $\mathbb{R}$. The characteristic function is

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \cdot \frac{1}{\pi(1+x^2)}\, dx, \quad t \in \mathbb{R}.$$

We shall evaluate this integral using the residue theorem. First assume $t > 0$, and let C be the closed curve shown below.



Then

$$\oint_C \frac{e^{itz}}{\pi(1+z^2)}\, dz = \frac{1}{\pi}\oint \frac{e^{itz}}{(z+i)(z-i)}\, dz = \frac{1}{\pi} \cdot 2\pi i\{\text{sum of residues}\} = 2i\{\frac{e^{-t}}{2i}\} = e^{-t}.$$

Since for $z \in C_1$ we have

(1) $|1 + z^2| \geq 1 - R^2$, and

(2) $|e^{itz}| = |e^{it(x+iy)}| = |e^{-ty}e^{itx}| \leq e^{-ty} \leq 1$ (here is where $t > 0$ is used),

then

$$\lim_{R\to\infty}\left|\int_{C_1}\frac{e^{itz}}{\pi(1+z^2)}\,dz\right| = \left|\lim_{R\to\infty}\int_0^\pi \frac{e^{itz}}{\pi(1+z^2)}Re^{i\theta}\,d\theta\right| \le \frac{1}{\pi}\lim_{R\to\infty}\int_0^\pi \frac{|e^{itz}|}{|1+z^2|}R\,d\theta$$
$$\le \frac{1}{\pi}\lim_{R\to\infty}\int_0^\pi \frac{R}{|R^2-1|}\,d\theta = 0.$$

Hence

$$e^{-t} = \lim_{R\to\infty}\oint_C \frac{e^{itz}}{\pi(1+z^2)}\,dz = \lim_{R\to\infty}\int_{C_1}\frac{e^{itz}}{\pi(1+z^2)}\,dz + \lim_{R\to\infty}\int_{C_2}\frac{e^{itz}}{\pi(1+z^2)}\,dz = 0+\phi(t),$$

and so $\phi(t) = e^{-t}$ for $t > 0$. Finally, since $X \overset{d}{=} -X$, then ϕ_X is real, so $\phi_X(t) = \phi_X(-t)$. Hence for $t < 0$, we have $\phi_X(t) = e^t$. Consequently, we conclude that

$$\phi(t) = e^{-|t|}, \quad t \in \mathbb{R}.$$

Lemma 2.1.7 *Let ϕ be the characteristic function of the distribution μ. Then*

$$\mu\{x : |x| > 2/u\} \le \frac{1}{u}\int_{-u}^{u}[1-\phi(t)]\,dt, \quad u > 0. \tag{1.5}$$

(Part of the result is that the right-hand side here is real.)

Proof. For any $u > 0$, we have

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

where we note in particular that the left-hand side is real. Dividing both sides by u and integrating with respect to μ gives

$$\frac{1}{u}\int_{-u}^{u}[1-\phi(t)]\,dt = 2\int 1 - \frac{\sin ux}{ux}\,d\mu \ge 2\int_{\{|x|\ge 2/u\}} 1 - \frac{\sin ux}{ux}\,d\mu$$
$$\ge 2\int_{\{|x|\ge 2/u\}} 1 - \frac{1}{|ux|}\,d\mu \ge \mu\{x : |x| > 2/u\},$$

where in the first inequality we used the fact that $|\sin y/y| \le 1$ for all $y \in \mathbb{R}$.

Theorem 2.1.8 (Continuity Theorem) *Let $\{\mu_n, n \ge 1\}$ be probability measures with corresponding characteristic functions $\{\phi_n, n \ge 1\}$.*

(1) Suppose that μ is a probability measure with characteristic function ϕ and that $\mu_n \overset{w}{\to} \mu$. Then $\phi_n(t) \to \phi(t)$ for all $t \in \mathbb{R}$.

(2) Suppose that $\phi_n(t) \to \phi(t)$ for all $t \in \mathbb{R}$, where ϕ is a function such that $\phi(t) \to 1$ as $t \to 0$. Then $\{\mu_n, n \ge 1\}$ is uniformly tight and converges weakly to a probability μ with characteristic function ϕ.

Proof.

(1) Since $\cos tx$ and $\sin tx$ are bounded continuous functions of x, then

$$\phi_n(t) = \int \cos tx\,d\mu_n + i\int \sin tx\,d\mu_n \to \int \cos tx\,d\mu + i\int \sin tx\,d\mu = \phi(t), \quad t \in \mathbb{R}.$$

(2) Let us show that the sequence $\{\mu_n, n \geq 1\}$ is tight. For any $u > 0$, we have from the above lemma that

$$\mu_n\{x : |x| > 2/u\} \leq \frac{1}{u}\int_{-u}^{u}[1-\phi_n(t)]\,dt \to \frac{1}{u}\int_{-u}^{u} 1-\phi(t)\,dt \quad \text{as } n \to \infty,$$

by the BCT, implying that $u^{-1}\int_{-u}^{u} 1-\phi(t)\,dt$ is real and non-negative. Let $\epsilon > 0$. Since $\phi(t) \to 1$ as $t \to 0$, then

$$\frac{1}{u}\int_{-u}^{u} 1-\phi(t)\,dt \to 0 \text{ as } u \to 0.$$

Let u' be such that the left-hand side here is $< \epsilon$. By the BCT, we have

$$\frac{1}{u'}\int_{-u'}^{u'}[1-\phi_n(t)]\,dt \to \frac{1}{u'}\int_{-u'}^{u'}[1-\phi(t)]\,dt < \epsilon,$$

so there is N such that

$$\mu_n\{x : |x| > 2/u'\} \leq \frac{1}{u'}\int_{-u'}^{u'}[1-\phi_n(t)]\,dt < \epsilon$$

for all $n \geq N$. This means that the sequence $\{\mu_n, n \geq 1\}$ is tight, and therefore relatively compact. This means that every subsequence of $\{\mu_n, n \geq 1\}$ has a further subsequence which converges weakly to some probability measure. Because of part (1), the characteristic functions of these limiting probabilities must all coincide with ϕ (implying that ϕ itself is a characteristic function), and therefore all these limiting probabilities in fact are the same, say μ. It is now easy to verify that the entire sequence $\{\mu_n, n \geq 1\}$ converges weakly to μ (see the remark preceding Prohorov's theorem).

∎

Lemma 2.1.9 *For any $n \geq 1$, we have*

$$e^{ix} = \sum_{m=0}^{n}\frac{(ix)^m}{m!} + R(x), \tag{1.6}$$

where

$$|R(x)| \leq \min\left[\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right]. \tag{1.7}$$

Proof. An easy integration by parts gives

$$\int_0^x (x-s)^n e^{is}\,ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1}e^{is}\,ds, \quad n \geq 0. \tag{1.8}$$

Using this with $n = 0$ gives

$$\frac{e^{ix}-1}{i} = \int_0^x e^{is}\,ds = x + i\int_0^x (x-s)e^{is}\,ds,$$

which after a slight rearrangment is

$$e^{ix} = 1 + ix + i^2\int_0^x (x-s)e^{is}\,ds.$$

Now we apply (1.8) (with $n = 1$) to this, and iterate, getting

$$\begin{aligned}
e^{ix} &= 1 + ix + i^2\left[\frac{x^2}{2} + \frac{i}{2}\int_0^x (x-s)^2 e^{is}\,ds\right] = 1 + ix + \frac{(ix)^2}{2} + \frac{i^3}{2}\int_0^x (x-s)^2 e^{is}\,ds \\
&= 1 + ix + \frac{(ix)^2}{2!} + \frac{i^3}{2!}\left[\frac{x^3}{3} + \frac{i}{3}\int_0^x (x-s)^3 e^{is}\,ds\right] = 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{i^4}{3!}\int_0^x (x-s)^3 e^{is}\,ds \\
&= \vdots \\
&= \sum_{m=0}^{n}\frac{(ix)^m}{m!} + R(x),
\end{aligned}$$

where

$$R(x) = \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is}\, ds.$$

Now we show that $R(x)$ satisfies (1.7). An easy calculation gives

$$|R(x)| \le \frac{|x|^{n+1}}{(n+1)!}.$$

On the other hand, since $\int_0^x (x-s)^{n-1}\, ds = x^n/n$, we have (by an integration by parts)

$$\begin{aligned} R(x) &= \frac{i^n}{(n-1)!}\left[\frac{i}{n}\int_0^x (x-s)^n e^{is}\, ds\right] = \frac{i^n}{(n-1)!}\left[-\frac{x^n}{n} + \int_0^x (x-s)^{n-1} e^{is}\, ds\right] \\ &= \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1)\, ds, \end{aligned}$$

and for this last expression, we have

$$\left|\frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1)\, ds\right| \le \frac{2|x|^n}{n!}.$$

∎

Proposition 2.1.10 *Let ϕ be the characteristic function of the random variable X, and suppose that $E|X|^n < \infty$. Then*

$$\phi(t) = \sum_{m=0}^{n} \frac{(it)^m E(X^m)}{m!} + \rho_n(t), \tag{1.9}$$

where

$$|\rho_n(t)| \le E \min\left[\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right], \tag{1.10}$$

and therefore $\rho_n(t) = o(t^n)$.

Proof. By substituting tX for x in (1.6), and then taking expected values, we find that

$$E(e^{itX}) = \sum_{m=0}^{n} \frac{(it)^m E(X^m)}{m!} + ER(tX).$$

Letting $\rho_n(t) = ER(tX)$, we have

$$|\rho_n(t)| \le E|R(tX)| \le E \min\left[\frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!}\right],$$

as required. Next, we have $|\rho_n(t)|/t^n \le E(Y_t)$, where

$$Y_t = \min\left[\frac{t|X|^{n+1}}{(n+1)!}, \frac{2|X|^n}{n!}\right].$$

Since Y_t is less than $2|X|^n/n!$ and tends to 0 as $t \to 0$, then $E(Y_t) \to 0$ as $t \to 0$. ∎

Proposition 2.1.11 (The Hamburger Moment Problem) *Let $\{m_n, n \ge 0\}$ be a sequence of numbers satisfying*

$$r \stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{m_{2n}^{1/2n}}{2n} < \infty. \tag{1.11}$$

Then there is at most one distribution μ (but there may be none) such that $\int x^n\, d\mu = m_n, n \ge 0$.

Proof. Let μ be any distribution with moments $m_n, n \geq 0$, and let $\nu_n = \int |x|^n \, d\mu, n \geq 0$. By the Cauchy-Schwartz inequality,

$$\nu_{2n+1}^2 = \left(\int |x|^n |x|^{n+1} \, d\mu \right)^2 \leq \left(\int |x|^{2n} \, d\mu \right) \left(\int |x|^{2(n+1)} \, d\mu \right) = \nu_{2n}\nu_{2(n+1)} = m_{2n} m_{2(n+1)}.$$

so that

$$\frac{\nu_{2n+1}^{1/(2n+1)}}{2n+1} \leq \left(\frac{m_{2n}^{1/(2n+1)}}{2n+1} \right)^{1/2} \cdot \left(\frac{m_{2(n+1)}^{1/(2n+1)}}{2n+1} \right)^{1/2}.$$

Using this and (1.11), we can show that

$$\limsup_{n\to\infty} \frac{\nu_{2n+1}^{1/(2n+1)}}{2n+1} \leq r,$$

and therefore, together with (1.11) and the fact that $\nu_{2n} = m_{2n}$, that

$$\limsup_{n\to\infty} \frac{\nu_n^{1/n}}{n} = r, \tag{1.12}$$

Replace x by sx in (1.6), and multiply the result by e^{itx}, to get

$$e^{ix(t+s)} = \sum_{m=0}^{n-1} \frac{(isx)^m}{m!} e^{itx} + e^{itx} R(sx).$$

Integrate both sides with respect to μ and use proposition 2.1.6 to find that

$$\phi(t+s) = \sum_{m=0}^{n-1} \frac{s^m}{m!} \phi^{(m)}(t) + \int e^{itx} R(sx) \, d\mu,$$

where

$$|\int e^{itx} R(sx) \, d\mu| \leq \int |R(sx)| \, d\mu \leq \int \frac{|sx|^n}{n!} \, d\mu = \frac{|s|^n}{n!} \nu_n.$$

Since $e^{-n}/n! \leq n^{-n}$, it follows that if $|s| \leq e^{-1}/r'$, where $r' > r$, then

$$|\int e^{itx} R(sx) \, d\mu| \leq \frac{|s|^n}{n!} \nu_n \leq \frac{e^{-n}}{r^n n!} \nu_n \leq \frac{\nu_n}{(r'n)^n} \to 0,$$

because of (1.12). (For we can choose N so that $\nu_n^{1/n}/n < r'' < r'$, and therefore $\nu_n < (nr'')^n$ for all $n \geq N$.) Hence for any t, we have the Taylor series expansion

$$\phi(t+s) = \sum_{m=0}^{\infty} \frac{s^m}{m!} \phi^{(m)}(t), \quad |s| < \frac{1}{er}. \tag{1.13}$$

Now suppose that ν is another probability measure with the same moments as μ, and suppose that ν has characteristic function ψ. Because ϕ and ψ have the same moments, then $\phi^{(m)}(0) = \psi^{(m)}(0)$ for every $m \geq 0$. From (1.13) with $t = 0$, we have

$$\phi(s) = \sum_{m=0}^{\infty} \frac{s^m}{m!} \phi^{(m)}(0) = \sum_{m=0}^{\infty} \frac{s^m}{m!} \psi^{(m)}(0) = \psi(s), \quad |s| \leq \frac{1}{er}.$$

We can now use (1.13) to extend this equality to the whole line, and thereby conclude that $\phi = \psi$, and therefore $\mu = \nu$. ∎

Remark. It can be shown that if the sequence $\{m_n, n \geq 1\}$ satisfies Carleman's condition

$$\sum_{n=1}^{\infty} \frac{1}{m_{2n}^{1/2n}} = \infty,$$

which is weaker that the condition in (1.11), then the same conclusion holds.

Example. We shall give an example of a probability measure which is not determined by its moments. Suppose that $Z \sim N(0,1)$, and define $X = e^Z$. Then X has density function given by

$$f_0(x) = \frac{e^{-(\log x)^2/2}}{x\sqrt{2\pi}}, \quad x > 0.$$

A random variable with this density function is said to have the *lognormal* distribution. Note that since the MGF of Z is $Ee^{tZ} = e^{t^2/2}$, the moments of X are $E(X^n) = E(e^{nZ}) = e^{n^2/2}, n \geq 0$. Now let a be any number with $|a| \leq 1$, and define

$$f(x) = \begin{cases} f_0(x)[1 + a\sin(2\pi \log x)], & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

To show that f is a density function and that f has the same moments as f_0, it suffices to observe that for any integer $r \geq 0$,

$$\int_0^\infty x^r f_0(x) \sin(2\pi \log x)\, dx = \int_{-\infty}^{\infty} e^{r(s+r)} \frac{e^{-(s+r)^2/2}}{\sqrt{2\pi}} \sin(2\pi(s+r))\, ds$$

(where we made the substitution $x = e^{s+r}$ and used the fact that $\sin(\theta + 2\pi r) = \sin\theta$,)

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi}} e^{r^2/2} \int_{-\infty}^{\infty} e^{-s^2/2} \sin(2\pi s)\, ds \\ &= 0, \end{aligned}$$

since the integrand is an odd function of s.

Proposition 2.1.12 (Method of Moments) *Let $\{\mu_n, n \geq 1\}$ be probability measures on $\mathbb{R}$, and suppose that*

$$m_k \stackrel{\text{def}}{=} \lim_{n\to\infty} \int x^k \, d\mu_n$$

exists and is finite for all integers $k \geq 0$. If the numbers $\{m_k, k \geq 0\}$ satisfy (1.11), there is a probability measure μ having moments $\{m_k, k \geq 0\}$ such that $\mu_n \xrightarrow{w} \mu$.

Proof. By applying proposition 1.4.5 with $\phi(x) = x^2$, we see that $\{\mu_n, n \geq 1\}$ is uniformly tight and therefore relatively compact. By applying corollary 1.5.5 with $h(x) = x^k$ and $g(x) = |x|^r$ where $k < r$, we see that for any convergent subsequence $\{\mu_{n_i}, i \geq 1\}$ with limit μ, we have

$$\int x^k \, d\mu = \lim_{n\to\infty} \int x^k \, d\mu_{n_i} = m_k.$$

Since this is true for every k, and since the numbers $\{m_k, k \geq 1\}$ satisfy (1.11), then every convergent subsequence of $\{\mu_n, n \geq 1\}$ has the same limit, namely μ. It follows that $\mu_n \xrightarrow{w} \mu$. ∎

2.2 More Properties of Characteristic Functions.

Definition. A random variable X is said to have a *lattice* distribution if there are constants b and $h \neq 0$ such that

$$P\{X \in b + h\mathbb{Z}\} = 1. \tag{2.1}$$

Note that we can always take $h > 0$. A characteristic function $\phi(t)$ is *periodic* if there are numbers b and $\lambda \neq 0$ with

$$\phi(t + n\lambda) = e^{ibn\lambda}\phi(t) \text{ for all } t \in \mathbb{R} \text{ and } n \in \mathbb{Z}. \tag{2.2}$$

Proposition 2.2.1 *Let X be a random variable with characteristic function $\phi(t)$. Let b and λ be numbers with $\lambda \neq 0$. Then the following statements are equivalent.*

(1) $\phi(\lambda) = e^{ib\lambda}$.

(2) $P\{X \in b + (2\pi/\lambda)\mathbb{Z}\} = 1$ *(that is, X has a lattice distribution).*

(3) $\phi(t + n\lambda) = e^{ibn\lambda}\phi(t)$ *for all $t \in \mathbb{R}$ and $n \in \mathbb{Z}$ (that is, ϕ is periodic).*

Proof. By passing to the random variable $X - b$, it suffices to prove the proposition in the case where $b = 0$. If (1) holds, then $E[1 - \cos\lambda X - i\sin\lambda X] = 0$. This implies that $E[1 - \cos\lambda X] = 0$ and therefore that $P\{\lambda X \in 2\pi\mathbb{Z}\} = 1$, so that (2) holds. Next, if (2) holds, then

$$\phi(t + n\lambda) = E(e^{i(t+n\lambda)X}) = \sum_{m\in\mathbb{Z}} e^{i(t+n\lambda)2\pi m/\lambda} P\{X = 2\pi m/\lambda\} = \sum_{m\in\mathbb{Z}} e^{it2\pi m/\lambda} P\{X = 2\pi m/\lambda\} = \phi(t),$$

so (3) holds. Finally, (3) with $t = 0$ and $n = 1$ obviously implies (1). ∎

Remarks.

(1) The h in (2.1) and the λ in (2.2) are related by $h = 2\pi/\lambda$. The largest $h > 0$ for which (2.1) holds is called the *span* of the distribution. The smallest $\lambda > 0$ for which (2.2) holds is called the *period* of ϕ. We thus have span=2π/period.

(2) Two real numbers h_0 and h_1 are called *incommensurate* if $h_0\mathbb{Z} \cap h_1\mathbb{Z} = \{0\}$ (i.e. if neither is a non-zero rational multiple of the other.) Suppose that h_0 and h_1 are incommensurate and

$$P\{X \in b_0 + h_0\mathbb{Z}\} = 1, \quad P\{X \in b_1 + h_1\mathbb{Z}\} = 1.$$

Let $A_0 = b_0 + h_0\mathbb{Z}$ and $A_1 = b_1 + h_1\mathbb{Z}$. Then $P(X \in A_0 \cap A_1) = 1$. In particular, $A_0 \cap A_1$ is not empty. If there were two members, say x and y of $A_0 \cap A_1$, we would have $b_0 + h_0 m_0 = x = b_1 + h_1 m_1$ and $b_0 + h_0 n_0 = y = b_1 + h_1 n_1$. Then $h_1 m_1 - h_0 m_0 = b_0 - b_1 = h_1 n_1 - h_0 n_0$, and so $h_1(m_1 - n_1) = h_0(m_0 - n_0)$. Since h_0 and h_1 are incommensurate, then $m_1 = n_1$ and $m_0 = n_0$ and so $x = y$. Thus $A_0 \cap A_1$ consists of precisely one member c, and therefore X is *degenerate*; that is, $P\{X = c\} = 1$.

Proposition 2.2.2 *Let $\phi(t)$ be the characteristic function of a random variable X.*

(1) X has a lattice distribution if and only if $|\phi(\lambda)| = 1$ for some $\lambda \neq 0$.

(2) If $|\phi(\lambda_0)| = 1$ and $|\phi(\lambda_1)| = 1$, where λ_0 and λ_1 are non-zero incommensurate numbers, then X is degenerate.

Proof.

(1) This follows from the previous proposition and the fact that if $|\phi(\lambda)| = 1$ for some $\lambda \neq 0$, then $\phi(\lambda) = e^{i\theta}$ for some θ, so there must be a number $b = \theta/\lambda$ such that $\phi(\lambda) = e^{ib\lambda}$. Note that (for the purpose of the proof of the next proposition) θ can be chosen to be in the interval $(-\pi, \pi]$, and so $b \in (-\frac{\pi}{\lambda}, \frac{\pi}{\lambda}]$.

(2) By the above proposition, we would have

$$P\{X \in b_0 + \frac{2\pi}{\lambda_0}\mathbb{Z}\} = 1, \quad P\{X \in b_1 + \frac{2\pi}{\lambda_1}\mathbb{Z}\} = 1.$$

Since $2\pi/\lambda_0$ and $2\pi/\lambda_1$ are incommensurate, then by remark (2), X is degenerate.

∎

Proposition 2.2.3 *Let $\phi(t)$ be the characteristic function of a random variable X. Then there are only three possibilities:*

(1) $|\phi(t)| < 1$ for all $t \neq 0$ (then X is called nonlattice).

(2) there is a $\lambda > 0$ so that $|\phi(\lambda)| = 1$ and $|\phi(t)| < 1$ for $0 < t < \lambda$ (then X has a lattice distribution with span $h = 2\pi/\lambda$).

(3) $|\phi(t)| = 1$ for all t (then X is degenerate).

Proof. Suppose (1) does not hold. Then the set $A = \{s > 0 || \phi(s)| = 1\}$ is non-empty (note that if $u \neq 0$ and $|\phi(u)| = 1$, then $|\phi(-u)| = 1$ as well). Let $\lambda = \inf\{s | s \in A\}$. Then $|\phi(\lambda)| = 1$ by the continuity of ϕ. If $\lambda > 0$, then $|\phi(t)| < 1$ for all $0 < t < \lambda$, and thus (2) holds. Otherwise, we have $\lambda = 0$, and so there exists a sequence $\{\lambda_n, n \geq 1\}$ of strictly positive numbers such that $\lambda_n \to 0$ and $|\phi(\lambda_n)| = 1$ for all n. As noted in part (1) of the proof of proposition 2.2.2, there are b_n's such that $\phi(\lambda_n) = e^{ib_n\lambda_n}$, and therefore

$$P\{X \in b_n + \frac{2\pi}{\lambda_n}\mathbb{Z}\} = 1. \tag{2.3}$$

Furthermore, for each n, we can choose $b_n \in (-\frac{\pi}{\lambda_n}, \frac{\pi}{\lambda_n}]$. Define $A_n = b_n + (2\pi/\lambda_n)[\mathbb{Z} \setminus \{0\}]$, $n \geq 1$. Then $A_n \subset (-\frac{\pi}{\lambda_n}, \frac{\pi}{\lambda_n}]^c$, so $P\{X \in A_n\} \to 0$ as $n \to \infty$. Since $P\{X \in A_n\} + P\{X = b_n\} \geq P[X \in \{b_n\} \cup A_n] = 1$, then $P\{X = b_n\} \geq 1 - P\{X \in A_n\} \to 1$ as $n \to \infty$. Choose N so large that $P\{X = b_n\} > 1/2$ for all $n \geq N$. Then we must have $b_n = b$ for all $n \geq N$, and then $P\{X = b\} = 1$. ∎

Remark. The *Riemann-Lebesgue* lemma is sometimes useful: if $g : \mathbb{R} \to \mathbb{R}$ is integrable, then

$$\lim_{|t| \to \infty} \int e^{-itx} g(x)\, dx = 0.$$

Thus, for the characteristic function $\phi(t)$ of an absolutely continuous distribution, we have $\lim_{|t|\to\infty} \phi(t) = 0$.

2.3 The Multidimensional Case.

If $x = (x_1, \dots, x_d), y = (y_1, \dots, y_d) \in \mathbb{R}^d$, then $\langle x, y\rangle = \sum_{j=1}^d x_j y_j$ denotes their inner product.

Definition. Let $X = (X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$ be a random vector with distribution $\mu(A) = P\{X \in A\}$, $A \in \mathcal{B}(\mathbb{R}^d)$. The characteristic function of X (or of μ) is

$$\phi(t) = Ee^{i\langle t, X\rangle} = \int e^{i\langle t, x\rangle} d\mu(x), \ t \in \mathbb{R}^d.$$

Theorem 2.3.1 (Inversion Formula) *Let $A = [a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$, and suppose $\mu(\partial A) = 0$. Then*

$$\mu(A) = \lim_{T \to \infty} \frac{1}{(2\pi)^d} \int_{[-T,T]^d} \left[\Pi_{j=1}^d \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j}\right] \phi(t) dt.$$

Proof. We have already proved this when $d = 1$. In that proof, we showed that

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\, dt \to \pi[I_{(a,b)}(x) + I_{[a,b]}(x)].$$

By Fubini's theorem, and then the BCT, we then have

$$\begin{aligned}
\int_{[-T,T]^d} \int \left[\Pi_{j=1}^d \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j}\right] e^{it_j x_j} \mu(dx) dt &= \int \left[\Pi_{j=1}^d \int_{-T}^{T} \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} e^{it_j x_j} dt_j\right] \mu(dx) \\
&\to \int \left[\Pi_{j=1}^d \pi[I_{(a_j,b_j)}(x_j) + I_{[a_j,b_j]}(x_j)]\right] \mu(dx) \\
&= (2\pi)^d \int \Pi_{j=1}^d I_{(a_j,b_j)}(x_j) \mu(dx) \\
&= (2\pi)^d \mu[(a_1, b_1) \times \cdots \times (a_d, b_d)] = (2\pi)^d \mu(A),
\end{aligned}$$

where we used $\mu(\partial A) = 0$ in the last two lines. ∎

Corollary 2.3.2 *Let μ_1 and μ_2 be probability measures on $\mathbb{R}^d$ with characteristic functions $\phi_1(t)$ and $\phi_2(t)$. If $\phi_1 = \phi_2$, then $\mu_1 = \mu_2$.*

Proof. Exactly the same proof works. Let F_1 and F_2 be the distribution functions corresponding to μ_1 and μ_2. For each j with $1 \le j \le d$, the functions $F_1(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty)$ and $F_2(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty)$ can each have only countably many discontinuities, so there is a set D_j dense in $\mathbb{R}$ such that D_j contains no discontinuities of these two functions. This means that the hyperplane $x_j = c$ has μ_1-measure 0 and μ_2-measure 0 if $c \in D_j$. Let $D = \cup_{j=1}^d D_j$. Then D is dense in $\mathbb{R}$. A box $\Pi_{j=1}^d (a_j, b_j)$ such that $a_j, b_j \in D$ for all j will have its faces contained in such hyperplanes, and so will have μ_1- and μ_2- measure 0. Hence let $\mathcal{U}$ be the set of all such open boxes. Then $\mathcal{U}$ is closed under finite intersection and generates $\mathcal{B}(\mathbb{R}^d)$, and μ_1 and μ_2 agree on $\mathcal{U}$ by the previous theorem. ∎

Corollary 2.3.3 *Let $X = (X_1, X_2, \ldots, X_d)$ be a random vector and let $\phi_j(t)$ denote the c.f. of X_j, $j = 1, \ldots, d$. Then $X_1, X_2, \ldots, X_d$ are independent if and only if*

$$Ee^{i\langle t, X\rangle} = \Pi_{j=1}^d \phi_j(t_j), \quad \forall t \in \mathbb{R}^d.$$

Proof. The "only if"part follows from the easily checked fact that if X and Y are independent r.v.'s and if $f, g : \mathbb{R} \to \mathbb{C}$, then $Ef(X)g(Y) = Ef(X)Eg(Y)$ (subject to integrability conditions). Conversely, suppose the condition holds. The LHS is $\int_{\mathbb{R}^d} e^{i\langle t,x\rangle} d\mu_X(x)$, where μ_X is the law of X. The RHS is $\Pi_{j=1}^d \int e^{it_j x_j} d\mu_j(x_j) = \int \cdots \int \Pi_{j=1}^d e^{it_j x_j} d\mu_1(x_1) \cdots d\mu_d(x_d) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} d\nu(x)$, where μ_j is the law of X_j and ν is the product probability $\mu_1 \otimes \cdots \otimes \mu_d$. By the previous corollary, we have $\mu_X = \mu_1 \otimes \cdots \otimes \mu_d$ so that $X_1, \ldots, X_d$ are independent. ∎

Theorem 2.3.4 (Continuity Theorem) *Let $\{\mu_n, n \ge 1\} \subset \mathcal{P}(\mathbb{R}^d)$ be probability measures with corresponding characteristic functions $\{\phi_n, n \ge 1\}$, and let $\mu \in \mathcal{P}(\mathbb{R}^d)$ have characteristic function ϕ. Then $\mu_n \xrightarrow{w} \mu$ if and only if $\phi_n(t) \to \phi(t)$ for all $t \in \mathbb{R}^d$.*

Proof. If $\mu_n \xrightarrow{w} \mu$, then trivially $\phi_n \to \phi$. Hence assume that $\phi_n \to \phi$. We will show the sequence $\{\mu_n, n \ge 1\}$ is tight. Let X_n and X be random vectors with distributions μ_n and μ, and let $\theta \in \mathbb{R}^d$ be fixed. $\phi_n(s\theta)$ (where $s \in \mathbb{R}$) is the c.f. of $\langle \theta, X_n\rangle$ and $\phi(s\theta)$ is the c.f. of $\langle \theta, X\rangle$. Since $\phi_n(s\theta) \to \phi(s\theta)$ for all s, then $\langle \theta, X_n\rangle \xrightarrow{d} \langle \theta, X\rangle$. It follows that the distributions of the r.v.'s $\langle \theta, X_n\rangle$ are uniformly tight. Let $\epsilon > 0$. By taking θ to be the jth unit vector in $\mathbb{R}^d$, there is a closed interval $[-c_j, c_j]$ such that $P\{X_n^{(j)} \notin [-c_j, c_j]\} < \epsilon/d$ for all $n \ge 1$ (Here, $X_n^{(j)}$ is the jth component of X_n). Then $P\{X_n \notin [-c_1, c_1] \times \cdots \times [-c_d, c_d]\} \le \sum_{j=1}^d P\{X_n^{(j)} \notin [-c_j, c_j]\} < \epsilon$ for all $n \ge 1$. Hence we have tightness. The rest of the proof is the same as before. ∎

Corollary 2.3.5 (The Cramer-Wold Device) *Let $\{X_n, n \geq 1\}$ and X be d-dimensional random vectors. If $\langle \theta, X_n \rangle \xrightarrow{d} \langle \theta, X \rangle$ for all $\theta \in \mathbb{R}^d$, then $X_n \xrightarrow{d} X$.*

Proof. If $\langle \theta, X_n \rangle \xrightarrow{d} \langle \theta, X \rangle$ for all $\theta \in \mathbb{R}^d$, then $\phi_{X_n}(\theta) = Ee^{i\langle \theta, X_n \rangle} \to Ee^{i\langle \theta, X \rangle} = \phi_X(\theta)$ for all $\theta \in \mathbb{R}^d$, so $X_n \xrightarrow{d} X$ by the theorem. ∎

Proposition 2.3.6 *Suppose that $(X_n, Y_n) \xrightarrow{d} (X, Y)$ as $n \to \infty$, and that $X_n \perp Y_n$ for each n. Then $X \perp Y$.*

Proof. First note that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$. We have $\phi_{X_n}(at)\phi_{Y_n}(bt) = \phi_{aX_n+bY_n}(t) \to \phi_{aX+bY}(t)$ for all a, b, t and $\phi_{X_n}(at)\phi_{Y_n}(bt) \to \phi_X(at)\phi_Y(bt)$ for all a, b, t. Hence $\phi_{aX+bY}(t) = \phi_X(at)\phi_Y(bt)$ for all a, b, t, implying that $X \perp Y$. ∎

Problem. Suppose X is a d dimensional random vector, that C is an $r \times d$ matrix, and that b is an $r \times 1$ vector. Let $Y = CX + b$. Show that $\phi_Y(t) = e^{i\langle t, b \rangle}\phi_X(C^T t),\ t \in \mathbb{R}^r$.

Solution. Follows from fact that $\langle t, CX + b \rangle = \langle t, b \rangle + \langle t, CX \rangle = \langle t, b \rangle + \langle C^T t, X \rangle$.

Definition. A random vector X is said to have the distribution $N(0, \Gamma)$ if X has c.f.

$$\phi(t) = e^{-\frac{1}{2}t^T \Gamma t} = e^{-\frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d \Gamma_{ij} t_i t_j}, \quad t \in \mathbb{R}^d,$$

where Γ is a symmetric non-negative definite d-dimensional matrix.

If $X \sim N(0, \Gamma)$, and if $Y = CX$ where C is an $r \times d$ matrix, then $\phi_Y(t) = \phi_X(C^T t) = e^{-\frac{1}{2}(C^T t)^T \Gamma C^T t} = e^{-\frac{1}{2}t^T (C\Gamma C^T)t}$, so $Y \sim N(0, C\Gamma C^T)$. (Note that $C\Gamma C^T$ is also symmetric and non-negative definite.)

Now let Γ be a given symmetric non-negative definite matrix. Let U be an orthogonal matrix (i.e. $U^T U = I$) such that $\Gamma = U^T D U$, where $D \geq 0$ is a diagonal matrix. Let $X \sim N(0, D)$, and let $Y = U^T X$. Then $Y \sim N(0, U^T D U) = N(0, \Gamma)$. This shows that random vectors "exist"with distribution $N(0, \Gamma)$. Also, $\text{Cov}(Y) = EYY^T = EU^T XX^T U = U^T E(XX^T)U = U^T D U = \Gamma$, which identifies Γ as the covariance matrix of Y.

Chapter 3

The Central Limit Theorem.

3.1 The Central Limit Theorem.

Lemma 3.1.1 *(1) Let $z_1, \dots, z_n$ and $w_1, \dots, w_n$ be complex numbers all of modulus ≤ 1. Then*

$$|\Pi_{m=1}^n z_m - \Pi_{m=1}^n w_m| \leq \sum_{m=1}^n |z_m - w_m|.$$

(2) If b is a complex number with $|b| < 1$, then

$$|e^{-b} - (1-b)| \leq |b|^2.$$

(3) Let $\{a_{n,m}, 1 \leq m \leq r_n < \infty\}$ be non-negative real numbers such that $\sum_{m=1}^{r_n} a_{n,m} \to a$ and $\sum_{m=1}^{r_n} |a_{n,m}|^2 \to 0$ as $n \to \infty$. Then

$$\Pi_{m=1}^{r_n}(1 - a_{n,m}) \to e^{-a}.$$

Proof.

(1) The result is obvious when $n = 1$. Suppose it has been shown to be true when $n = k-1$. Then

$$\begin{aligned}
\left|\Pi_{m=1}^k z_m - \Pi_{m=1}^k w_m\right| &\leq \left|z_k \Pi_{m=1}^{k-1} z_m - z_k \Pi_{m=1}^{k-1} w_m\right| + \left|z_k \Pi_{m=1}^{k-1} w_m - w_k \Pi_{m=1}^{k-1} w_m\right| \\
&\leq \left|\Pi_{m=1}^{k-1} z_m - \Pi_{m=1}^{k-1} w_m\right| + |z_k - w_k| \\
&\leq \sum_{m=1}^k |z_m - w_m|.
\end{aligned}$$

(2) We have

$$e^{-b} - (1-b) = \frac{b^2}{2!} - \frac{b^3}{3!} + \frac{b^4}{4!} - \dots = b^2 \left[\frac{1}{2} - \frac{b}{3!} + \frac{b^2}{4!} - \dots\right],$$

so

$$|e^{-b} - (1-b)| \leq |b|^2 \left[\frac{1}{2} + \frac{1}{3!} + \frac{1}{4!} + \dots\right] = |b|^2(e-2) \leq |b|^2.$$

(3) Let $z_{n,m} = 1 - a_{n,m}$ and $w_{n,m} = e^{-a_{n,m}}$. Then

$$\begin{aligned}
\left|\Pi_{m=1}^{r_n}(1 - a_{n,m}) - e^{-a}\right| &\leq \left|\Pi_{m=1}^{r_n}(1 - a_{n,m}) - \Pi_{m=1}^{r_n} e^{-a_{n,m}}\right| + \left|e^{-\sum_{m=1}^{r_n} a_{n,m}} - e^{-a}\right| \\
&\leq \sum_{m=1}^{r_n} |1 - a_{n,m} - e^{-a_{n,m}}| + \left|e^{-\sum_{m=1}^{r_n} a_{n,m}} - e^{-a}\right| \\
&\leq \sum_{m=1}^{r_n} |a_{n,m}|^2 + \left|e^{-\sum_{m=1}^{r_n} a_{n,m}} - e^{-a}\right| \to 0
\end{aligned}$$

as $n \to \infty$.

Remarks.

(1) When $z_m = z$ and $w_m = w$ for all m, part (1) of this lemma becomes simply

$$|z^n - w^n| \leq n|z - w|. \tag{1.1}$$

(2) If we take $r_n = n$, $a_{n,m} = c_n/n$, and $a = c$ in part(3), we obtain the well-known fact that if $c_n \to c$, then

$$(1 - \frac{c_n}{n})^n \to e^{-c} \text{ as } n \to \infty.$$

Theorem 3.1.2 (The Lindeberg-Lévy or "Classical" Central Limit Theorem) *Let $\{X_n, n \geq 1\}$ be i.i.d. random variables with finite mean μ and finite variance σ^2. Define $S_n = \sum_{i=1}^n X_i, \quad n \geq 1$. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty.$$

Proof. First assume that $\mu = 0$ and $\sigma^2 = 1$. Then $\phi_{S_n/\sqrt{n}}(t) = \phi_X^n(t/\sqrt{n})$ where from proposition 2.1.10, we have $\phi_X(t) = 1 - t^2/2 + o(t^2)$. Using (1.1), we can write

$$\begin{aligned}
|\phi_{S_n/\sqrt{n}}(t) - e^{-t^2/2}| &\leq |\phi_X^n(\frac{t}{\sqrt{n}}) - [1 - \frac{t^2}{2n}]^n| + |[1 - \frac{t^2}{2n}]^n - e^{-t^2/2}| \\
&\leq n|\phi_X(\frac{t}{\sqrt{n}}) - [1 - \frac{t^2}{2n}]| + |[[1 - \frac{t^2}{2n}]^n - e^{-t^2/2}| \\
&= n|o(\frac{t^2}{n})| + |[[1 - \frac{t^2}{2n}]^n - e^{-t^2/2}| \to 0
\end{aligned}$$

as $n \to \infty$. The required result then follows from the continuity theorem. To finish the proof, we observe that

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n \frac{X_i - \mu}{\sigma}}{\sqrt{n}} \xrightarrow{d} N(0,1).$$

∎

Corollary 3.1.3 (Multivariate Version of CLT) *Let $\{X_n, n \geq 1\}$ be a sequence of i.i.d. random vectors with values in $\mathbb{R}^d$, with finite mean vector $\mu = E(X)$ and finite covariance matrix $\Sigma = E(X - \mu)(X - \mu)^T$. Let $S_n = X_1 + \cdots + X_n, n \geq 1$. Then*

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d} N(0, \Sigma)$$

as $n \to \infty$.

Proof. We will use the Cramer-Wold device. Let $\theta \in \mathbb{R}^d$. Let $Y_n = \langle\theta, X_n - \mu\rangle = \theta^T(X_n - \mu)$. Then the Y_n's are i.i.d random variables with mean zero and variance $\text{Var}(Y) = E(YY^T) = E(\theta^T(X - \mu)(X - \mu)^T\theta) = \theta^T\Sigma\theta$, and by the univariate CLT,

$$\langle\theta, \frac{S_n - n\mu}{\sqrt{n}}\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i \xrightarrow{d} N(0, \theta^T\Sigma\theta) \stackrel{d}{=} \langle\theta, W\rangle,$$

where $W \sim N(0, \Sigma)$. ∎

Theorem 3.1.4 (The Lindeberg-Feller Theorem) *For each $n \geq 1$, let $X_{n,m}$ be independent (but not necessarily identically distributed) random variables with $E(X_{n,m}) = 0$ for all $1 \leq m \leq r_n$, and define $S_n = \sum_{m=1}^{r_n} X_{n,m}$. Assume that*

(1) $\sum_{m=1}^{r_n} E(X_{n,m}^2) \to \sigma^2 > 0$ *as* $n \to \infty$,

(2) for all $\epsilon > 0$, $\sum_{m=1}^{r_n} \int_{\{|X_{n,m}| > \epsilon\}} X_{n,m}^2 \, dP \to 0$ *as* $n \to \infty$.

(These two conditions imply that $r_n \to \infty$ as $n \to \infty$.) Then $S_n/\sigma \xrightarrow{d} N(0,1)$ as $n \to \infty$.

Proof. By dividing all the $X_{n,m}$'s by σ, we can assume that $\sigma = 1$. First, note that for any $\epsilon > 0$, we have

$$E(X_{n,m}^2) = \int_{\{|X_{n,m}| \le \epsilon\}} X_{n,m}^2 \, dP + \int_{\{|X_{n,m}|>\epsilon\}} X_{n,m}^2 \, dP \le \epsilon^2 + \sum_{m=1}^{r_n} \int_{\{|X_{n,m}|>\epsilon\}} X_{n,m}^2 \, dP,$$

implying that

$$\sup_{1\le m\le r_n} E(X_{n,m}^2) \to 0 \text{ as } n \to \infty. \tag{1.2}$$

Let $a_{n,m} = t^2 E(X_{n,m}^2)/2$ and $a = t^2/2$. Then by (1), we have $\sum_{m=1}^{r_n} a_{n,m} \to a$, and by (1.2) and (2) we have

$$\sum_{m=1}^{r_n} a_{n,m}^2 = \frac{t^4}{4} \sum_{m=1}^{r_n} [E(X_{n,m}^2)]^2 \le \frac{t^4}{4} \sup_{1\le m\le r_n} E(X_{n,m}^2) \sum_{m=1}^{r_n} E(X_{n,m}^2) \to 0 \quad \text{as } n \to \infty.$$

Let $\phi_{n,m}(t)$ be the characteristic function of $X_{n,m}$, so that $\phi_{S_n}(t) = \Pi_{m=1}^{r_n} \phi_{n,m}(t)$. Since

$$\left|\phi_{S_n}(t) - e^{-t^2/2}\right| \le |\Pi_{m=1}^{r_n}\phi_{n,m}(t) - \Pi_{m=1}^{r_n}(1-a_{n,m})| + \left|\Pi_{m=1}^{r_n}(1-a_{n,m}) - e^{-a}\right|,$$

and since by part (3) of lemma 3.1.1, the second term on the right tends to 0, we only have to show that

$$|\Pi_{m=1}^{r_n}\phi_{n,m}(t) - \Pi_{m=1}^{r_n}(1-a_{n,m})| \to 0 \text{ as } n \to \infty. \tag{1.3}$$

To do this, suppose that $\epsilon > 0$. Then by part (1) of lemma 3.1.1,

$$\begin{aligned}
|\Pi_{m=1}^{r_n}\phi_{n,m}(t) - \Pi_{m=1}^{r_n}(1-a_{n,m})| &\le \sum_{m=1}^{r_n} |\phi_{n,m}(t) - (1-a_{n,m})| \le \sum_{m=1}^{r_n} E[\min\{\frac{|tX_{n,m}|^3}{3!}, \frac{2|tX_{n,m}|^2}{2!}\}] \\
&\le \sum_{m=1}^{r_n} \left[\int_{\{|X_{n,m}|\le\epsilon\}} \frac{|tX_{n,m}|^3}{6} \, dP + \int_{\{|X_{n,m}|>\epsilon\}} |tX_{n,m}|^2 \, dP\right] \\
&\le \frac{\epsilon|t|^3}{6} \sum_{m=1}^{r_n} \int_{\{|X_{n,m}|\le\epsilon\}} X_{n,m}^2 \, dP + t^2 \sum_{m=1}^{r_n} \int_{\{|X_{n,m}|>\epsilon\}} X_{n,m}^2 \, dP \\
&\le \frac{\epsilon|t|^3}{6} \sum_{m=1}^{r_n} EX_{n,m}^2 + t^2 \sum_{m=1}^{r_n} \int_{\{|X_{n,m}|>\epsilon\}} X_{n,m}^2 \, dP
\end{aligned}$$

so that

$$\limsup_{n\to\infty} |\Pi_{m=1}^{r_n}\phi_{n,m}(t) - \Pi_{m=1}^{r_n}(1-a_{n,m})| \le \frac{\epsilon|t|^3}{6}.$$

Since ϵ is arbitrary, this proves (1.3). ∎

Remark. We show why conditions (1) and (2) imply that $r_n \to \infty$ as $n \to \infty$. By assumption (1), $r_n \cdot \sup_{1\le m\le r_n} EX_{n,m}^2 \ge \sum_{m=1}^{r_n} EX_{n,m}^2 \ge \frac{\sigma^2}{2} \forall n \ge N$, and then (1.2) implies $r_n \to \infty$.

The following is a non-triangular version of the Lindeberg-Feller theorem.

Theorem 3.1.5 *Let $\{X_n, n \ge 1\}$ be a sequence of independent (but not necessarily identically distributed) random variables with finite means $\mu_n = E(X_n)$ and finite variances $\sigma_n^2 = Var(X_n)$ for all n. Define $S_n = \sum_{m=1}^n X_m$ and $s_n^2 = Var(S_n) = \sum_{m=1}^n \sigma_m^2$. Assume that for all $\epsilon > 0$,*

$$\frac{1}{s_n^2} \sum_{m=1}^{n} \int_{\{|X_m-\mu_m|>\epsilon s_n\}} (X_m - \mu_m)^2 \, dP \to 0 \quad \text{as } n \to \infty. \tag{1.4}$$

Then

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty. \tag{1.5}$$

Proof. In the Lindeberg-Feller theorem, take $r_n = n$, $X_{n,m} = (X_m - \mu_m)/s_n$. Then for condition (1) in the Lindeberg-Feller theorem, we have

$$\sum_{m=1}^{r_n} E(X_{n,m}^2) = \sum_{m=1}^{n} \frac{E(X_m - \mu_m)^2}{s_n^2} = 1,$$

and so is obviously satisfied, and condition (2) there obviously becomes the condition in (1.4) of this theorem. ∎

Applications. We shall consider several cases in which Lindeberg's condition in (1.4) is satisfied, and therefore for which the convergence in (1.5) takes place.

(1) *Identically Distributed Case.* Assume the X_n's are i.i.d. with finite mean μ and variance σ^2. Then $s_n^2 = n\sigma^2$ and for the condition in (1.4), we have

$$\begin{aligned}\frac{1}{s_n^2}\sum_{m=1}^{n}\int_{\{|X_m-\mu_m|>\epsilon s_n\}}(X_m-\mu_m)^2\,dP &= \frac{1}{n\sigma^2}\sum_{m=1}^{n}\int_{\{|X_1-\mu|>\epsilon\sigma\sqrt{n}\}}(X_1-\mu)^2\,dP \\ &= \frac{1}{\sigma^2}\int_{\{|X_1-\mu|>\epsilon\sigma\sqrt{n}\}}(X_1-\mu)^2\,dP \to 0 \quad \text{as } n\to\infty.\end{aligned}$$

Consequently, the Lindeberg-Feller theorem contains the "ordinary"central limit theorem as a special case.

(2) *Uniformly Bounded Case.* Suppose the X_n's are independent, that $|X_n| \le M < \infty$ for all n, and that $s_n^2 \to \infty$ as $n \to \infty$. Then by Chebychev,

$$\int_{\{|X_m-\mu_m|>\epsilon s_n\}}(X_m-\mu_m)^2\,dP \le (2M)^2 P\{|X_m-\mu_m|>\epsilon s_n\} \le \frac{(2M)^2\sigma_m^2}{\epsilon^2 s_n^2},$$

so for the condition in (1.4), we have

$$\frac{1}{s_n^2}\sum_{m=1}^{n}\int_{\{|X_m-\mu_m|>\epsilon s_n\}}(X_m-\mu_m)^2\,dP \le \frac{(2M)^2}{\epsilon^2 s_n^2} \to 0 \quad \text{as } n\to\infty.$$

(3) *Lyapunov's Condition.* Suppose the X_n's are independent and that there is a $\delta > 0$ such that

$$\frac{1}{s_n^{2+\delta}}\sum_{m=1}^{n} E|X_m-\mu_m|^{2+\delta} \to 0 \text{ as } n\to\infty.$$

Then

$$E|X_m-\mu_m|^{2+\delta} \ge \int_{\{|X_m-\mu_m|\ge\epsilon s_n\}}|X_m-\mu_m|^{\delta}|X_m-\mu_m|^2\,dP \ge \epsilon^\delta s_n^\delta\int_{\{|X_m-\mu_m|\ge\epsilon s_n\}}(X_m-\mu_m)^2\,dP,$$

so for the condition in (1.4), we have

$$\frac{1}{s_n^2}\sum_{m=1}^{n}\int_{\{|X_m-\mu_m|>\epsilon s_n\}}(X_m-\mu_m)^2\,dP \le \frac{1}{s_n^2}\sum_{m=1}^{n}\frac{E|X_m-\mu_m|^{2+\delta}}{\epsilon^\delta s_n^\delta} \to 0 \quad \text{as } n\to\infty.$$

(4) *The DeMoivre-Laplace Limit Theorem.* Suppose that X is a binomial random variable with parameters n and p. Since X can be written as $X = X_1 + \cdots + X_n$, where the X_i's are independent Bernoulli random variables taking values 0 and 1 with probabilities q and p respectively, then

$$\frac{X-np}{\sqrt{npq}} \xrightarrow{d} N(0,1) \text{ as } n\to\infty.$$

Example. Suppose $X \sim$ binomial with $n = 100$ and $p = 1/2$. Then

$$P\{48 < X \le 51\} \approx P\{48.5 < X \le 51.5\} = P\{\frac{48.5 - 50}{5} < \frac{X - np}{\sqrt{npq}} \le \frac{51.5 - 50}{5}\} = P\{-.3 < Z \le .3\}$$
$$= 0.2358(\text{ from tables of the normal distribution}).$$

The exact value, as calculated in Mathematica, is 0.235647.

3.2 The Berry-Esseen Theorem.

Remark. For the next lemma, we recall the *Riemann-Lebesgue* lemma: if $g : \mathbb{R} \to \mathbb{R}$ is integrable, then

$$\lim_{|t|\to\infty} \int e^{-itx} g(x)\, dx = 0.$$

Lemma 3.2.1 *Let F and G be distribution functions with integrable characteristic functions ϕ and ψ respectively. Suppose there exists an $\epsilon > 0$ such that $\int_{\{|t|\le\epsilon\}} \left|\frac{\phi(t)-\psi(t)}{t}\right| dt < \infty$ (this would be true, for example, if the means of F and G both exist). Then*

$$F(x) - G(x) = \frac{1}{2\pi} \int \frac{-e^{-itx}}{it} [\phi(t) - \psi(t)]\, dt.$$

Proof. Recall that F has a bounded density given by $f(x) = \frac{1}{2\pi} \int e^{-itx} \phi(t)\, dt$. It follows from Fubini's theorem that

$$F(x) - F(a) = \int_a^x f(w)\, dw = \int_a^x \frac{1}{2\pi} \int e^{-itw} \phi(t)\, dt\, dw = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itx}}{it} \phi(t)\, dt.$$

A similar identity holds for G. Hence we can write

$$[F(x) - G(x)] - [F(a) - G(a)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itx}}{it} [\phi(t) - \psi(t)]\, dt.$$

Letting $a \to -\infty$ and using the Riemann-Lebesgue lemma with $g(t) = [\phi(t) - \psi(t)]/t$, we get the required result. Note that $\int |g(t)|\, dt < \infty$ since we can write

$$\begin{aligned}
\int |g(t)|\, dt &\le \int_{\{|t|\le\epsilon\}} \left|\frac{\phi(t) - \psi(t)}{t}\right| dt + \int_{\{|t|>\epsilon\}} \left|\frac{\phi(t) - \psi(t)}{t}\right| dt \\
&\le \int_{\{|t|\le\epsilon\}} \left|\frac{\phi(t) - \psi(t)}{t}\right| dt + \frac{1}{\epsilon} \left[\int |\phi(t)| dt + \int |\psi(t)| dt\right].
\end{aligned}$$

Also note that $\int_{\{|t|\le\epsilon\}} \left|\frac{\phi(t)-\psi(t)}{t}\right| dt \le \int_{\{|t|\le\epsilon\}} \left|\frac{\phi(t)-1}{t}\right| dt + \int_{\{|t|\le\epsilon\}} \left|\frac{1-\psi(t)}{t}\right| dt$, so that if the means of F and G exist, then such an ϵ does exist. ∎

Definition. Let $L > 0$.

$$h_L(x) = \frac{1}{\pi} \frac{1 - \cos Lx}{Lx^2}, \quad x \in \mathbb{R},$$

is called *Polya's density function.* It has characteristic function

$$w_L(t) = \begin{cases} 1 - \frac{|t|}{L} & \text{if } |t| \le L, \\ 0 & \text{if } |t| > L. \end{cases}$$

Let H_L be the distribution function corresponding to h_L. Note that if $\phi(t)$ is any characteristic function, then $\phi(t)w_L(t)$ is an integrable characteristic function. This is the main purpose of Polya's density function.

Lemma 3.2.2 *Let F and G be distribution functions, and suppose that G is differentiable with $G'(x) \le \lambda < \infty$ for all x.*

(1) Let

$$\Delta(x) = F(x) - G(x), \quad \eta = \sup_x |\Delta(x)|, \quad \Delta_L = \Delta \star H_L, \quad \eta_L = \sup_x |\Delta_L(x)|.$$

Then

$$\eta \le 2\eta_L + \frac{24\lambda}{\pi L} \quad \forall L > 0. \tag{2.1}$$

(2) Suppose that F and G have finite means. Let ϕ and ψ be the characteristic functions of F and G. Then

$$\sup_x |F(x) - G(x)| \le \frac{1}{\pi} \int_{-L}^{L} \left| \frac{\phi(t) - \psi(t)}{t} \right| dt + \frac{24\lambda}{\pi L} \quad \forall L > 0. \tag{2.2}$$

Proof. Let $\{x_n, n \ge 1\}$ be such that $|\Delta(x_n)| \uparrow \eta$. Since $\Delta(x) \to 0$ as $x \to \pm\infty$, we can assume that the x_n's belong to some compact interval. Hence there exists a convergent subsequence $\{x_{n_i}, i \ge 1\}$, converging say to x_0. Then we have $|\Delta(x_{n_i})| \uparrow \eta$ and $x_{n_i} \to x_0$ as $i \to \infty$. Let us show that either

$$\Delta(x_0) = \eta \text{ or } \Delta(x_0^-) = -\eta. \tag{2.3}$$

First, if F is continuous at x_0, then $|\Delta(x_0)| = \eta$, so that (2.3) holds. Hence suppose that F is not continuous at x_0, and that $\Delta(x_0) \ne \eta$. If infinitely many of the x_{n_i}'s lie at or above x_0, then there is a subsequence of $\{x_{n_i}, i \ge 1\}$ tending to x_0 from above, so $|\Delta(x_0)| = \eta$ by right continuity of F. This would mean $\Delta(x_0) = -\eta$, so that $F(x_0^-) - G(x_0) < \Delta(x_0) = -\eta$, contradicting the definition of η. Hence, except for at most finitely many terms, the sequence $\{x_{n_i}, i \ge 1\}$ lies below x_0. This means that $|F(x_0^-) - G(x_0)| = \eta$. If $F(x_0^-) - G(x_0) = \eta$, then we would have $F(x_0) - G(x_0) > \eta$; hence we can only have $F(x_0^-) - G(x_0) = -\eta$. Thus the statement in (2.3) is proved.

Assume that $\Delta(x_0) = \eta$. Since $G'(x) \le \lambda$ and F is non-decreasing, then for $x_1 > x_0$,

$$\Delta(x_1) = F(x_1) - \left[G(x_0) + \int_{x_0}^{x_1} G'(x)\, dx \right] \ge F(x_0) - G(x_0) - \lambda(x_1 - x_0) = \eta - \lambda(x_1 - x_0).$$

Letting $\delta = \eta/2\lambda$ and $t = x_0 + \delta$, we have (taking $x_1 = t - x$ in the above equation)

$$\Delta(t - x) \ge \begin{cases} \eta - \lambda(t - x - x_0) = \eta - \lambda(\delta - x) = \frac{\eta}{2} + \lambda x, & \text{if } |x| \le \delta. \\ -\eta & \text{otherwise.} \end{cases}$$

Since also

$$\int_{|x|>\delta} h_L(x)\, dx = 2 \int_{\delta}^{\infty} h_L(x)\, dx \le 2 \int_{\delta}^{\infty} \frac{2}{\pi L x^2}\, dx = \frac{4}{\pi L \delta},$$

then

$$\begin{aligned} \eta_L \ge \Delta_L(t) &= \int \Delta(t - x) h_L(x)\, dx = \int_{|x|\le\delta} \Delta(t - x) h_L(x)\, dx + \int_{|x|>\delta} \Delta(t - x) h_L(x)\, dx \\ &\ge \int_{|x|\le\delta} \left[\frac{\eta}{2} + \lambda x \right] h_L(x)\, dx - \eta \int_{|x|>\delta} h_L(x)\, dx = \frac{\eta}{2} \int_{|x|\le\delta} h_L(x)\, dx - \eta \int_{|x|>\delta} h_L(x)\, dx \\ &= \frac{\eta}{2} \left[1 - \int_{|x|>\delta} h_L(x)\, dx \right] - \eta \int_{|x|>\delta} h_L(x)\, dx \ge \frac{\eta}{2} \left[1 - \frac{4}{\pi L \delta} \right] - \eta \frac{4}{\pi L \delta} \\ &= \frac{\eta}{2} - \frac{12\lambda}{\pi L}, \end{aligned}$$

which is equivalent to (2.1).

Next, suppose that $\Delta(x_0^-) = -\eta$. Let X and Y be random variables with distribution functions F and G. Define $X^* = -X$ and $Y^* = -Y$, and let F^* and G^* be the corresponding distribution functions. Then

$$F^*(x) = P\{-X \le x\} = P\{X \ge -x\} = 1 - P\{X < -x\} = 1 - F((-x)^-), \quad G^*(x) = 1 - G(-x),$$

and so $\Delta^*(x) = F^*(x) - G^*(x) = G(-x) - F((-x)^-)$. Observe that $|\Delta^*(x)| \le \eta$ for all x and $\Delta^*(-x_0) = G(x_0) - F((x_0)^-) = \eta$. Hence we have the scenerio of the previous paragraph with $\eta^* = \eta$, and so we have

$$\eta_L^* \ge \frac{\eta}{2} - \frac{12\lambda}{\pi L}.$$

Finally, $\Delta^*(x) = -\Delta(-x)$ except at discontinuities of F, so

$$\Delta_L^*(x) = \int_{-\infty}^{\infty} \Delta^*(x-y)h_L(y)\,dy = -\int_{-\infty}^{\infty} \Delta(y-x)h_L(y)\,dy = -\int_{-\infty}^{\infty} \Delta(-x-w)h_L(w)\,dw = -\Delta_L(-x),$$

and so $\eta_L^* = \eta_L$. This finishes the proof of part (1).

For part (2), we apply part (1) and use the previous lemma to find that

$$\begin{aligned}
\sup_x |F(x) - G(x)| &\le 2\sup_x |F * H_L(x) - G * H_L(x)| + \frac{24\lambda}{\pi L} \\
&= \frac{1}{\pi}\sup_x \left| \int_{-\infty}^{\infty} \frac{-e^{-itx}}{it}[\phi(t) - \psi(t)]w_L(t)\,dt \right| + \frac{24\lambda}{\pi L} \\
&\le \frac{1}{\pi}\sup_x \int_{-\infty}^{\infty} \left| \frac{\phi(t) - \psi(t)}{t} \right| w_L(t)\,dt + \frac{24\lambda}{\pi L},
\end{aligned}$$

and therefore the required result. Notice that the requirements of the lemma are satisfied because $\phi(t)w_L(t)$ and $\psi(t)w_L(t)$ are integrable, and $\int_{\{|t|\le\epsilon\}} \left| \frac{\phi(t)w_L(t) - \psi(t)w_L(t)}{t} \right| dt = \int_{\{|t|\le\epsilon\}} \left| \frac{\phi(t)-\psi(t)}{t} \right| |w_L(t)|\,dt < \infty$. ∎

Theorem 3.2.3 (The Berry-Esseen Theorem) *Let $\{X_n, n \ge 1\}$ be i.i.d. random variables with mean 0 and variance σ^2, and assume that $E|X|^3 = \rho < \infty$. Let F_n be the distribution function of $S_n/\sigma\sqrt{n}$ and let Φ be the standard normal distribution function. Then*

$$\sup_x |F_n(x) - \Phi(x)| \le \frac{3\rho}{\sigma^3\sqrt{n}}, \quad n \ge 1. \tag{2.4}$$

(Note: Since $\rho^{1/3} \ge \sigma$, the smallest the RHS can be is $3/\sqrt{n}$.)

Proof. By considering the random variables $X' = X/\sigma$, we can assume $\sigma = 1$. Then $\rho \ge 1$ and the right-hand side of (2.4) is ≥ 1 for $n \le 9$. Hence in the remainder of this proof, we can assume $n \ge 10$. By applying part (2) of the preceding lemma, we obtain

$$\sup_x |F_n(x) - \Phi(x)| \le \frac{1}{\pi}\int_{-L}^{L} \left| \frac{\phi^n(t/\sqrt{n}) - e^{-t^2/2}}{t} \right| dt + \frac{24\lambda}{\pi L}, \tag{2.5}$$

where ϕ is the characteristic function of X. The rest of the proof is to show that the right-hand side in (2.5) is majorized by the right-hand side of (2.4).

We will require the following two facts:

Fact 1: $|\phi(t) - 1 + t^2/2| \le \rho|t|^3/6$ (from (1.10) of chapter 2), and

Fact 2: if α and β are complex numbers and γ is a positive number such that $|\alpha|, |\beta| \le \gamma$, then

$$|\alpha^n - \beta^n| = |\alpha - \beta||\alpha^{n-1} + \alpha^{n-2}\beta + \cdots + \alpha\beta^{n-2} + \beta^{n-1}| \le n\gamma^{n-1}|\alpha - \beta|.$$

We will take $L = 4\sqrt{n}/3\rho$ in the rest of the proof. We are going to apply fact 2 to the integrand of the integral on the right-hand side of (2.5) with

$$\alpha = \phi(\frac{\theta}{\sqrt{n}}), \quad \beta = e^{-\theta^2/2n}, \quad \gamma = e^{-5\theta^2/18n},$$

where $|\theta| \le L$, and therefore $\theta^2/2n \le L/2n = 8/(9\rho^2) \le 1$. This is justified because obviously $\beta \le \gamma$, and by fact 1 we have

$$|\alpha| \le \left|\phi(\frac{\theta}{\sqrt{n}}) - 1 + \frac{\theta^2}{2n}\right| + \left|1 - \frac{\theta^2}{2n}\right| \le \frac{\rho|\theta|^3}{6n^{3/2}} + 1 - \frac{\theta^2}{2n} = \frac{\rho|\theta|}{\sqrt{n}} \cdot \frac{|\theta|^2}{6n} + 1 - \frac{\theta^2}{2n} \le \frac{4}{3} \cdot \frac{|\theta|^2}{6n} + 1 - \frac{\theta^2}{2n} = 1 - \frac{5\theta^2}{18n} \le \gamma,$$

where we used the facts that $\rho|\theta|/\sqrt{n} \le \rho L/\sqrt{n} \le 4/3$ and $1 - x \le e^{-x}$ for $x \ge 0$. We will also need to know that since $n \ge 10$, then $\frac{5(n-1)}{18n} \ge \frac{1}{4}$, so $\gamma^{n-1} \le e^{-\theta^2/4}$. Thus, applying fact 2, we have

$$\begin{aligned}\left|\phi^n(\frac{\theta}{\sqrt{n}}) - e^{-\theta^2/2}\right| &= |\alpha^n - \beta^n| \le ne^{-\theta^2/4}|\alpha - \beta| \le ne^{-\theta^2/4}\left[\left|\phi(\frac{\theta}{\sqrt{n}}) - 1 + \frac{\theta^2}{2n}\right| + \left|1 - \frac{\theta^2}{2n} - e^{-\theta^2/2n}\right|\right] \\ &\le ne^{-\theta^2/4}\left[\frac{\rho|\theta|^3}{6n^{3/2}} + \frac{\theta^4}{8n^2}\right]\end{aligned}$$

where we used the fact that $|1 - x - e^{-x}| \le x^2/2$ for $|x| \le 1$. Then for the integrand in (2.5), we have

$$\left|\frac{\phi^n(\frac{\theta}{\sqrt{n}}) - e^{-\theta^2/2}}{\theta}\right| \le \frac{ne^{-\theta^2/4}}{|\theta|}\left[\frac{\rho|\theta|^3}{6n^{3/2}} + \frac{\theta^4}{8n^2}\right] = e^{-\theta^2/4}\left[\frac{\rho|\theta|^2}{6n^{1/2}} + \frac{|\theta|^3}{8n}\right] \le \frac{1}{L}e^{-\theta^2/4}\left[\frac{2|\theta|^2}{9} + \frac{|\theta|^3}{18}\right],$$

where we used the facts that $\rho/\sqrt{n} = 4/3L$ and $1/n = 1/\sqrt{n} \cdot 1/\sqrt{n} \le 4/3L \cdot 1/3$. Since

$$\int_{-\infty}^{\infty} x^2 e^{-x^2/2a^2}\, dx = a^3\sqrt{2\pi}, \qquad \int_{-\infty}^{\infty} |x|^3 e^{-x^2/2a^2}\, dx = 4a^4,$$

and $24\lambda = 24 \sup_x G'(x) = 24/\sqrt{2\pi} < 9.6$, then the right-hand side of (2.5) is

$$\begin{aligned}\frac{1}{\pi}\int_{-L}^{L}\left|\frac{\phi^n(\frac{\theta}{\sqrt{n}}) - e^{-\theta^2/2}}{\theta}\right| d\theta + \frac{24\lambda}{\pi L} &\le \frac{1}{\pi}\int_{-L}^{L}\frac{1}{L}e^{-\theta^2/4}\left[\frac{2|\theta|^2}{9} + \frac{|\theta|^3}{18}\right] d\theta + \frac{9.6}{\pi L} \\ &\le \frac{1}{\pi L}\int_{-\infty}^{\infty} e^{-\theta^2/4}\left[\frac{2|\theta|^2}{9} + \frac{|\theta|^3}{18}\right] d\theta + \frac{9.6}{\pi L} \\ &= \frac{1}{\pi L}\left[\frac{2}{9}2\sqrt{4\pi} + \frac{16}{18} + 9.6\right] \\ &= \frac{1}{\pi}\frac{3}{4}\left[\frac{2}{9}2\sqrt{4\pi} + \frac{16}{18} + 9.6\right]\frac{\rho}{\sqrt{n}} \\ &\le \frac{3\rho}{\sqrt{n}},\end{aligned}$$

as required. ∎

Remark. Consider the situation of the Demoivre-Laplace limit theorem, where $X_i \sim$ bernoulli for every i. Then $\rho = E|X_i - p|^3 = q|-p|^3 + p(1-p)^3 = pq(p^2 + q^2)$. Since $\sigma = \sqrt{pq}$, then the Berry-Esseen bound is

$$\frac{3\rho}{\sigma^3\sqrt{n}} = \frac{3(p^2 + q^2)}{\sqrt{npq}}.$$

For $p = q = 1/2$, this is $3/\sqrt{n}$.

3.3 Martingale Central Limit Theorem.

Reference: Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.

Notation. Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \le i \le k_n\}$ be a zero-mean, square-integrable martingale for each fixed $n \ge 1$, and let $X_{ni} = S_{n,i} - S_{n,i-1}$ (where $S_{n0=0}$) denote the martingale differences. We assume that $k_n \uparrow +\infty$ as $n \to \infty$. The double sequence $\{S_{ni}, \mathcal{F}_{ni}, 1 \le i \le k_n, n \ge 1\}$ is called a *martingale array.*

Martingale arrays are frequently derived from ordinary zero-mean martingales $\{S_i, \mathcal{F}_i, i \ge 1\}$ as follows: define $k_n = n$, $\mathcal{F}_{ni} = \mathcal{F}_i$, and $S_{ni} = S_i/s_n,\ 1 \le i \le n$, where $s_n^2 = \mathrm{Var}(S_n)$.

Theorem 3.3.1 *Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \le i \le k_n, n \ge 1\}$ be a zero-mean, square-integrable martingale array with differences X_{ni}. Suppose that*

$$\max_i |X_{ni}| \xrightarrow{P} 0, \tag{3.1}$$

$$\sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{P} \eta^2, \tag{3.2}$$

where η^2 is an a.s. finite r.v.,

$$E\left(\max_i |X_{ni}|^2\right) \quad \textit{is bounded in } n, \tag{3.3}$$

and the σ-fields are nested, that is

$$\mathcal{F}_{n,i} \subset \mathcal{F}_{n+1,i}, 1 \le i \le k_n, n \ge 1. \tag{3.4}$$

Then $S_{nk_n} = \sum_{i=1}^{k_n} X_{ni} \xrightarrow{d} Z$, where Z has c.f. $\phi(t) = Ee^{-\eta^2 t^2/2}$.

Corollary 3.3.2 *If (3.1) and (3.3) are replaced by the conditional Lindeberg condition*

$$\sum_{i=1}^{k_n} E[X_{ni}^2 I_{\{|X_{ni}|>\epsilon\}} | \mathcal{F}_{n,i-1}] \xrightarrow{P} 0 \quad \textit{as } n \to \infty, \textit{ for all } \epsilon > 0, \tag{3.5}$$

if (3.2) is replaced by an analogous condition on the conditional variance:

$$V_{nk_n}^2 \stackrel{\text{def}}{=} \sum_{i=1}^{k_n} E(X_{ni}^2 | \mathcal{F}_{n,i-1}) \xrightarrow{P} \eta^2 \ \textit{as } n \to \infty, \tag{3.6}$$

and if (3.4) holds, then the conclusion of the theorem remains true.

If the martingale differences are independent, this corollary obviously reduces to the Lindeberg-Feller CLT.

3.4 Poisson Convergence.

Definition. A random variable Z has the Poisson distribution with mean $\lambda > 0$ if

$$P\{Z = z\} = \frac{\lambda^z e^{-\lambda}}{z!}, \quad z = 0, 1, 2, \ldots .$$

Remarks.

(1) The characteristic function of Z is $\psi(t) = e^{\lambda(e^{it}-1)}$.

(2) Suppose $\{Z_n, n \geq 1\}$ are Poisson random variables with means λ_n, and that $\lambda_n \to \lambda$, where $\lambda > 0$. Then

$$e^{\lambda_n(e^{it}-1)} \to e^{\lambda(e^{it}-1)},$$

so that $Z_n \xrightarrow{d} Z$, where Z has a Poisson distribution with mean λ.

Theorem 3.4.1 *For each $n \geq 1$, let $X_{n,m}, 1 \leq m \leq r_n$ be independent random variables such that*

$$P\{X_{n,m} = 1\} = p_{n,m}, \quad P\{X_{n,m} = 0\} = 1 - p_{n,m},$$

where

(1) $\sum_{m=1}^{r_n} p_{n,m} \to \lambda > 0$,

(2) $\max_{1 \leq m \leq r_n} p_{n,m} \to 0$ *as* $n \to \infty$.

Let $S_n = \sum_{m=1}^{r_n} X_{n,m}$ for $n \geq 1$. Then $S_n \xrightarrow{d} Z$, where Z has a Poisson distribution with parameter λ.

Remarks.

(1) This is the same triangular scheme as in the Lindeberg-Feller theorem. Assumption (1) here is the same as assumption (1) there, since $EX_{nm}^2 = p_{nm}$. But assumption (2) is different.

(2) Once again, we observe that assumptions (1) and (2) imply that $r_n \to \infty$ as $n \to \infty$. For suppose there is a subsequence $\{n_i, i \geq 1\}$ such that $r_{n_i} \leq r < \infty$ for all $i \geq 1$. Then assuming (2) holds, we would have

$$\sum_{m=1}^{r_{n_i}} p_{n_i,m} \leq \sum_{m=1}^{r_{n_i}} \max_{1 \leq m \leq r_{n_i}} p_{n_i,m} \leq r \max_{1 \leq m \leq r_{n_i}} p_{n_i,m} \to 0,$$

contradicting (1).

(3) Because of assumptions (1) and (2), we have

$$\sum_{m=1}^{r_n} p_{n,m}^2 \leq \left(\max_{1 \leq m \leq r_n} p_{n,m} \right) \cdot \sum_{m=1}^{r_n} p_{n,m} \to 0 \text{ as } n \to \infty.$$

Proof 1. Let $\phi_{n,m}(t)$ and $\phi_n(t)$ be the characteristic functions of $X_{n,m}$ and S_n, so that

$$\phi_{n,m}(t) = (1 - p_{n,m}) + p_{n,m}e^{it}, \quad \phi_n(t) = \Pi_{m=1}^{r_n}[(1 - p_{n,m}) + p_{n,m}e^{it}].$$

Let $\lambda_n = \sum_{m=1}^{r_n} p_{n,m}$. Let $\psi_n(t)$ be the characteristic function of a Poisson distribution with mean λ_n, and let ψ be the characteristic function of a Poisson with mean λ. Then by lemma 3.1.1, we have

$$\begin{aligned} |\phi_n(t) - \psi_n(t)| &= |\Pi_{m=1}^{r_n}[1 + p_{n,m}(e^{it} - 1)] - \Pi_{m=1}^{r_n} e^{p_{n,m}(e^{it}-1)}| \\ &\leq \sum_{m=1}^{r_n} |1 + p_{n,m}(e^{it} - 1) - e^{p_{n,m}(e^{it}-1)}| \end{aligned}$$

(by part (2) of lemma 3.1.1 with $b = p_{n,m}(1 - e^{it})$. Note that by assumption (2), n may be chosen large enough that $|b| < 1$.)

$$\leq \sum_{m=1}^{r_n} p_{n,m}^2 |e^{it} - 1|^2 \leq 4 \sum_{m=1}^{r_n} p_{n,m}^2 \to 0$$

as $n \to \infty$. Since by a remark above, we also have $\psi_n \to \psi$, then by the triangle inequality $\phi_n \to \psi$. ∎

Remarks. For the second proof, we require the following facts:

(1) if $\mu_1, \ldots, \mu_n$ and $\nu_1, \ldots, \nu_n$ are distributions, then by propositions 1.2.4 and 1.2.5,

$$\|\mu_1 \star \cdots \star \mu_n - \nu_1 \star \cdots \star \nu_n\| \leq \sum_{m=1}^{n} \|\mu_m - \nu_m\|.$$

(2) Let μ be the measure with $\mu(0) = 1 - p$ and $\mu(1) = p$, and let ν be the Poisson measure with mean p. Then using the fact that $1 - x \leq e^{-x} \leq 1$ for $x \geq 0$, we have

$$\begin{aligned}
\|\mu - \nu\| &= \sum_{n=0}^{\infty} |\mu(n) - \nu(n)| = |\mu(0) - \nu(0)| + |\mu(1) - \nu(1)| + \sum_{n=2}^{\infty} \nu(n) \\
&= |1 - p - e^{-p}| + |p - pe^{-p}| + 1 - e^{-p} - pe^{-p} = [e^{-p} + p - 1] + [p - pe^{-p}] + 1 - e^{-p} - pe^{-p} \\
&= 2p(1 - e^{-p}) \leq 2p^2.
\end{aligned}$$

Proof 2. Let $\mu_{n,m}$ be the distribution of $X_{n,m}$, and let μ_n be the distribution of S_n. Let $\nu_{n,m}, \nu_n$, and ν be Poisson distributions with means $p_{n,m}, \lambda_n = \sum_{m=1}^{r_n} p_{n,m}$, and λ, respectively. Since $\mu_n = \mu_{n,1} \star \cdots \star \mu_{n,r_n}$ and $\nu_n = \nu_{n,1} \star \cdots \star \nu_{n,r_n}$, then by remarks (1) and (2),

$$\|\mu_n - \nu_n\| \leq \sum_{m=1}^{r_n} \|\mu_{n,m} - \nu_{n,m}\| \leq 2 \sum_{m=1}^{r_n} p_{n,m}^2 \to 0 \text{ as } n \to \infty.$$

As noted at the beginning of this section, we also have $\|\nu_n - \nu\| \to 0$ as $n \to \infty$. Hence $\|\mu_n - \nu\| \leq \|\mu_n - \nu_n\| + \|\nu_n - \nu\| \to 0$. ∎

Example. Let $\{p_n, n \geq 1\}$ be numbers in $[0,1]$ such that $p_n \to 0$ in such a way that $\lambda_n =^{\text{def}} np_n \to \lambda > 0$. In the above theorem, take $r_n = n$ and $p_{n,m} = p_n$. Then S_n has the binomial distribution with parameters n and p_n, and $S_n \xrightarrow{d} \text{Poisson}(\lambda)$.

3.5 Infinite Divisibility.

Question. For each $n \geq 1$, let $X_1^n, \ldots, X_n^n$ be i.i.d., and put

$$S_n = X_1^n + \cdots + X_n^n. \tag{5.1}$$

What distributions (other than the Normal or the Poisson) can appear as limits in distribution of the S_n's?

Definition. An r.v. X is *infinitely divisible* if for every $n \geq 1$, there are i.i.d. r.v.'s $X_1^n, \ldots, X_n^n$ such that $X \stackrel{d}{=} X_1^n + \cdots + X_n^n$.

Proposition 3.5.1 *Let X be a r.v. with c.f. $\phi(t)$. Then X is infinitely divisible if and only if for each $n \geq 1$, there is a c.f. $\phi_n(t)$ such that $\phi(t) = (\phi_n(t))^n$.*

Proof. $\Leftarrow$ Let $X_1^n, \ldots, X_n^n$ be i.i.d. each with c.f. $\phi_n(t)$. Then $X_1^n + \cdots + X_n^n$ has c.f. $\phi(t)$, so equals X in distribution.

Proposition 3.5.2 *Answer to above question: X is the limit in distribution of sums as in (5.1) if and only if X is infinitely divisible.*

Proof. If X is infinitely divisible, this is obvious. Hence suppose X is such that $S_n \xrightarrow{d} X$. Write $S_{2n} = Y_n + Y_n'$, where $Y_n = X_1^{2n} + \cdots + X_n^{2n}$ and $Y_n' = X_{n+1}^{2n} + \cdots + X_{2n}^{2n}$. Then $Y_n \stackrel{d}{=} Y_n'$ and $P\{Y_n > y\}^2 = P\{Y_n > y, Y_n' > y\} \leq P\{S_{2n} > 2y\} \to P\{X > 2y\}$ and similarly $P\{Y_n < -y\}^2 = P\{Y_n < -y, Y_n' < -y\} \leq$

$P\{S_{2n} < -2y\} \to P\{X < -2y\}$, from which it follows that the sequence $\{Y_n, n \geq 1\}$ is tight. Hence $\exists$ a subsequence $\{Y_{n_i}, i \geq 1\}$ which is convergent in distribution, say to Y. Then also $Y'_{n_i} \xrightarrow{d} Y'$, where $Y' \stackrel{d}{=} Y$, and is independent of Y. Then $Y_{n_i} + Y'_{n_i} \xrightarrow{d} Y + Y'$, so $X \stackrel{d}{=} Y + Y'$ where Y and Y' are i.i.d. Thus the definition of infinite divisibility holds with $n = 2$. The same proof can be extended to the case $n > 2$. ∎

Lemma 3.5.3 *Let $\phi(t)$ be a c.f. Then*

$$1 - \Re(\phi(2t)) \leq 4[1 - \Re(\phi(t))], \quad t \in \mathbb{R}.$$

Proof. Since $\cos 2\theta = 2\cos^2\theta - 1$, then

$$1-\Re(\phi(2t)) = E[1-\cos 2Xt] = 2E[1-\cos^2 tX] = 2E(1-\cos tX)(1+\cos tX) \leq 4E(1-\cos tX) = 4[1-\Re(\phi(t))].$$

∎

Proposition 3.5.4 *An infinitely divisible c.f. $\phi(t)$ never vanishes.*

Proof. Since $|\phi(t)|^2$ is also an infinitely divisible c.f. ($|\phi(t)|^2 = \phi(t)\phi(-t)$ is the c.f. of $X + X'$ where $X \stackrel{d}{=} X'$ and X and X' are independent) and vanishes if and only if $\phi(t)$ vanishes, we can assume $\phi(t)$ is real and non-negative. Since $\phi(0) = 1$, then ϕ is strictly positive on some interval $(-b, b)$ where $b > 0$. Hence we need only prove that if $a > 0$ is such that $r =^{\text{def}} \inf_{t\in(-a,a)} \phi(t) > 0$, then $\inf_{t\in(-2a,2a)} \phi(t) > 0$.

Let $0 < \epsilon < 1/4$, and let n be so large that $1 - r^{1/n} < \epsilon$. Then $1 - \phi_n(t) = 1 - \phi(t)^{1/n} < \epsilon$ for all $t \in (-a, a)$, and so $1 - \phi_n(2t) \leq 4[1 - \phi_n(t)] < 4\epsilon$ for all $t \in (-a, a)$. Then $\phi(2t) = \phi_n(2t)^n \geq (1 - 4\epsilon)^n$ for all $t \in (-a, a)$. ∎

Theorem 3.5.5 (Lévy-Khintchine, 1934) *Let $\phi(t)$ be a c.f. Then $\phi(t)$ is infinitely divisible if and only if*

$$\log \phi(t) = i\gamma t - \frac{\sigma^2 t^2}{2} + \int \left[e^{itx} - 1 - \frac{itx}{1+x^2}\right] \frac{1+x^2}{x^2} \mu(dx),$$

where $\gamma \in \mathbb{R}$, $\sigma^2 > 0$, and μ is a finite measure on $\mathbb{R}$ such that $\mu\{0\} = 0$. This representation is unique in terms of γ, σ^2, and μ.

Examples.

(1) $N(\gamma, \sigma^2)$. Here $\mu(dx) = 0$.

(2) Poisson(λ). Take $\gamma = \lambda/2$, $\sigma^2 = 0$, and let $\mu(\{1\}) = \lambda/2$ and $\mu(\mathbb{R} \setminus \{1\}) = 0$. Then

$$i\gamma t - \frac{\sigma^2 t^2}{2} + \int \left[e^{itx} - 1 - \frac{itx}{1+x^2}\right] \frac{1+x^2}{x^2} \mu(dx) = \frac{i\lambda t}{2} + \left[e^{it} - 1 - \frac{it}{2}\right] \lambda = \lambda(e^{it} - 1).$$

(3) Gamma, Cauchy. (Note: we say $X \sim$ Cauchy(a, b) where $a > 0$ if X has density function $f(x) = \frac{a}{\pi} \frac{1}{a^2+(x-b)^2}$ for $x \in \mathbb{R}$ or if X has c.f. $\phi(t) = e^{itb-a|t|}$).

3.6 The Convergence of Types Theorem.

Problem. Suppose that $U_n \xrightarrow{d} U$ and that $\{\alpha_n, n \geq 1\}$ and $\{\beta_n, n \geq 1\}$ are real numbers with $\alpha_n \to \alpha$ and $\beta_n \to \beta$, where α and β are finite. Then $\alpha_n U_n + \beta_n \xrightarrow{d} \alpha U + \beta$.

Solution This follows from proposition 1.5.3. Since $\alpha_n - \alpha \to 0$, then $(\alpha_n - \alpha)U_n \xrightarrow{P} 0$. Since $\alpha U_n \xrightarrow{d} \alpha U$, then $\alpha_n U_n = (\alpha_n - \alpha)U_n + \alpha U_n \xrightarrow{d} \alpha U$. Then $\alpha_n U_n + (\beta_n - \beta) \xrightarrow{d} \alpha U$, so $\alpha_n U_n + \beta_n \xrightarrow{d} \alpha U + \beta$.

Remark. In particular, if ϕ_n and ϕ are the characteristic functions of U_n and U, then $\phi_n(\alpha_n t) \to \phi(\alpha t)$ for every t. We will use this frequently in the proof of the next theorem.

Theorem 3.6.1 (Convergence of Types Theorem) *Suppose that*

(1) $U_n \xrightarrow{d} U$, *where* U *is non-degenerate,*

(2) $\exists$ *numbers* α_n, β_n *with* $\alpha_n > 0$ *such that* $\alpha_n U_n + \beta_n \xrightarrow{d} V$, *where* V *is non-degenerate.*

Then the limits $\alpha = \lim_{n\to\infty} \alpha_n$ *and* $\beta = \lim_{n\to\infty} \beta_n$ *exist (and are finite) with* $\alpha > 0$, *and* $V \stackrel{d}{=} \alpha U + \beta$.

Proof. Let $V_n \stackrel{d}{=} \alpha_n U_n + \beta_n$. Let $\phi_n(t), \psi_n(t), \phi(t)$, and $\psi(t)$ be the characteristic functions of U_n, V_n, U, and V respectively. Then

$$\phi_n(t) \to \phi(t), \quad \psi_n(t) = e^{i\beta_n t}\phi_n(\alpha_n t) \to \psi(t). \tag{6.1}$$

Let $\{\alpha_{n_i}, i \geq 1\}$ be a subsequence which converges to say $\alpha \in [0, \infty]$. (Note that such a subsequence always exists. Either we can find a subsequence which tends to ∞, or the α_n's all belong to a compact set.) If α were 0, we would have $\phi_{n_i}(\alpha_{n_i} t) \to 1$ (since $\alpha_{n_i} U_{n_i} \xrightarrow{d} 0$) and therefore $|\psi_{n_i}(t)| \to 1$ as $i \to \infty$, so that $|\psi(t)| = 1$ for all t, implying that V is degenerate. Hence we have shown $\alpha > 0$. If α were ∞, then since

$$\phi_n(u) = e^{-iu\beta_n/\alpha_n}\psi_n(u/\alpha_n), \quad u \in \mathbb{R},$$

we would have $|\phi_{n_i}(u)| \to 1$ as $i \to \infty$, so that $|\phi(t)| = 1$ for all t. Thus we have shown that $\alpha < \infty$. Finally, observe that $|\psi_{n_i}(t)| \to |\phi(\alpha t)|$ as $i \to \infty$, so that $|\psi(t)| = |\phi(\alpha t)|$ for all t. If there were two distinct subsequential limits, say $\alpha' < \alpha$, then $|\phi(\alpha t)| = |\psi(t)| = |\phi(\alpha' t)|$, so that $|\phi(u)| = |\phi(u\alpha'/\alpha)|$ for all u. Iterating this, we find that for every u, $|\phi(u)| = |\phi(u(\alpha'/\alpha)^m)|$ for every $m \geq 1$. Letting $m \to \infty$ shows that $|\phi(u)| = 1$ for all u, so U is degenerate. Hence we conclude that every convergent subsequence of the α_n's converges to the same limit. This means that the sequence $\{\alpha_n, n \geq 1\}$ converges to $\alpha > 0$ as stated.

Now we turn to the β_n's. Let $\delta > 0$ be such that $|\phi(\alpha t)| > 0$ for $|t| \leq \delta$. Then

$$e^{i\beta_n t} = \frac{\psi_n(t)}{\phi_n(\alpha_n t)} \to \frac{\psi(t)}{\phi(\alpha t)}, \quad |t| \leq \delta.$$

Next, since $\psi(t)/\phi(\alpha t) \to 1$ as $t \to 0$, then

$$\frac{1}{u}\int_{-u}^{u} 1 - \frac{\psi(t)}{\phi(\alpha t)}\, dt \to 0 \text{ as } u \to 0.$$

Let $u' < \delta$ be such that the left-hand side here is $< 1/2$. Since $|1 - e^{-i\beta_n t}| \leq 2$, then by (1.5) of chapter 2 and the BCT, we have

$$\epsilon_{\beta_n}\{x : |x| > 2/u'\} \leq \frac{1}{u'}\int_{-u'}^{u'} [1 - e^{i\beta_n t}]\, dt \to \frac{1}{u'}\int_{-u'}^{u'} [1 - \frac{\psi(t)}{\phi(\alpha t)}]\, dt < \frac{1}{2},$$

so there is N such that $\epsilon_{\beta_n}\{x : |x| > 2/u'\} < 1$ (and is therefore zero) for all $n \geq N$. This means that $|\beta_n| \leq \frac{2}{u'} \forall n \geq N$ and therefore the sequence $\{\beta_n, n \geq 1\}$ is bounded. If β_{n_i} is a convergent subsequence which converges say to β, then $e^{i\beta t} = \psi(t)/\phi(\alpha t)$, $|t| \leq \delta$. So if β and β' were two subsequential limits, we would have $e^{i\beta t} = e^{i\beta' t}$ for all $|t| \leq \delta$, implying that $\beta = \beta'$. Thus every convergent subsequence has the same limit, so the sequence $\{\beta_n, n \geq 1\}$ is itself convergent as claimed. By letting $n \to \infty$ in the equality $\psi_n(t) = e^{i\beta_n t}\phi_n(\alpha_n t)$, we get $\psi(t) = e^{i\beta t}\phi(\alpha t)$, implying that $V \stackrel{d}{=} \alpha U + \beta$. ∎

Corollary 3.6.2 *Suppose that* $\{S_n, n \geq 1\}$ *are random variables and that there exist sequences of numbers*

(1) $\{a_n, n \geq 1\}$ *and* $\{b_n, n \geq 1\}$ *with* $a_n > 0$ *for all* n *such that*

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} U$$

where U *is nondegenerate,*

(2) $\{a_n', n \geq 1\}$ *and* $\{b_n', n \geq 1\}$ *with* $a_n' > 0$ *for all* n *such that*

$$\frac{S_n - b_n'}{a_n'} \xrightarrow{d} V$$

where V *is nondegenerate.*

Then U *and* V *are of the same type. If* U *is normal, then so is* V*.*

Proof. Let $U_n = (S_n - b_n)/a_n$, $V_n = (S_n - b_n')/a_n'$, $\alpha_n = a_n/a_n'$, and $\beta_n = (b_n - b_n')/a_n'$. Then

$$V_n = \frac{S_n - b_n'}{a_n'} = \frac{a_n}{a_n'} \cdot \frac{S_n - b_n}{a_n} + \frac{b_n - b_n'}{a_n'} = \alpha_n U_n + \beta_n, \quad n \geq 1.$$

By the convergence of types theorem, we have $V \stackrel{d}{=} \alpha U + \beta$. ∎

3.7 Stable Laws.

Definition. R.v.'s X and Y are said to be *of the same type* if there are $a > 0$ and b such that $Y \stackrel{d}{=} \frac{X-b}{a}$. That is, if X and Y have the same distribution except for scale and location parameters.

Definition. A non-degenerate r.v. Y is said to have a *stable law* if for every $k \geq 1$, Y is of the same type as $Y_1 + \cdots + Y_k$ where $Y_1, Y_2, \ldots, Y_k$ are i.i.d with the same distribution as Y. That is, if $\exists$ constants $a_k > 0, b_k$ so that

$$Y \stackrel{d}{=} \frac{Y_1 + \cdots + Y_k - b_k}{a_k}.$$

Y is said to have a *strictly* stable law if the b_k's can be chosen to be zero.

Problem. Show that stable laws are infinitely divisible.

Solution. Suppose X has a stable law, so that $X \stackrel{d}{=} \frac{X_1 + \ldots + X_n - b_n}{a_n}$ for every n. Let

$$X_m^n = \frac{X_m - \frac{b_n}{n}}{a_n}.$$

Then $X_1^n, \ldots, X_n^n$ are i.i.d. and $X_1^n + \cdots + X_n^n \stackrel{d}{=} X$ for every n.

However, the family of stable laws is *strictly* contained in the family of infinitely divisible laws, since stable laws are absolutely continuous (a problem) and the Poisson law is discrete.

Proposition 3.7.1 *Y is the limit in distribution of $\frac{X_1 + \cdots + X_k - B_k}{A_k}$ for some i.i.d. sequence $\{X_i, i \geq 1\}$, some sequence $A_k > 0$, and some sequence B_k if and only if Y has a stable law. (Note: the A_k's and B_k's are not necessarily the same as the a_k's and b_k's in the above definition of a stable law.)*

Proof. If Y has a stable law, we can take $X_1, X_2 \ldots$ to be i.i.d. with the same distribution as Y. Conversely, fix k and let $Z_k = \frac{X_1 + \cdots + X_k - B_k}{A_k}$ and $S_n^j = X_{(j-1)n+1} + \cdots + X_{jn}$ for $j, n \geq 1$. Then

$$\begin{aligned}
Z_{nk} &= \frac{S_n^1 + \cdots + S_n^k - B_{nk}}{A_{nk}}, \\
A_{nk} Z_{nk} &= (S_n^1 - B_n) + \cdots + (S_n^k - B_n) + (kB_n - B_{nk}), \\
\frac{A_{nk} Z_{nk}}{A_n} &= \frac{S_n^1 - B_n}{A_n} + \cdots + \frac{S_n^k - B_n}{A_n} + \frac{kB_n - B_{nk}}{A_n}.
\end{aligned}$$

The first k terms on the RHS of the last identity tend in distribution to $Y_1 + \cdots + Y_k$ as $n \to \infty$, where $Y_1, \ldots, Y_k$ are i.i.d with the same distribution as Y. On the left hand side, Z_{nk} converges in distribution to Y. It now follows from the convergence of types theorem that $\exists\, \alpha > 0$ and β such that $\alpha Y \stackrel{d}{=} Y_1 + \cdots + Y_k - \beta$. ∎

Lemma 3.7.2 *Suppose $\{a_n, n \geq 1\}$ is an increasing sequence of strictly positive numbers such that $a_2 > 1$ and $a_{mn} = a_m a_n$ for all $m, n \geq 1$. Then $\exists \beta > 0$ such that $a_n = n^\beta \forall n \geq 1$.*

Proof. Define β by $a_2 = 2^\beta$. Since $a_{mn} = a_m a_n$, it follows that $a_{m^r} = a_m^r \forall m, r \geq 1$. In particular, if n is of the form 2^r, then $a_n = a_{2^r} = a_2^r = 2^{r\beta} = n^\beta$. Define $m_{n,q} = \left[\frac{q \log n}{\log 2}\right]$ (where $[x]$ denotes the integer part of x) for all $q \geq 1$, so that $m_{n,q} \leq \frac{q \log n}{\log 2} < m_{n,q} + 1$, and therefore $m_{n,q} \log 2 \leq \log n^q < (m_{n,q} + 1) \log 2$, so $2^{m_{n,q}} \leq n^q < 2^{(m_{n,q}+1)}$. Since a_n is increasing, then $(2^{m_{n,q}})^\beta = a_{2^{m_{n,q}}} \leq a_{n^q} \leq a_{2^{(m_{n,q}+1)}} = (2^{(m_{n,q}+1)})^\beta$. Now $a_{n^q} = a_n^q$, so $(2^{m_{n,q}})^\beta \leq a_n^q \leq (2^{(m_{n,q}+1)})^\beta$, and taking logs and then dividing by $q \log 2$ gives us $\frac{\beta m_{n,q}}{q} \leq \frac{\log a_n}{\log 2} \leq \frac{\beta(m_{n,q}+1)}{q}$. Now letting $q \to \infty$, and using the fact that $\lim_{q\to\infty} \frac{[qx]}{q} = x$, both endpoints of this double inequality tend to $\frac{\beta \log n}{\log 2}$, so $\frac{\log a_n}{\log 2} = \frac{\beta \log n}{\log 2}$, and therefore $\log a_n = \log n^\beta \forall n \geq 1$. ∎

Remark. If $Y = U + V$ where U and V are independent, then

$$F_Y(y) = \int_{-\infty}^{+\infty} F_U(y - v) dF_V(v) \leq \int_{(-\infty,0]} dF_V(v) + \int_{(0,+\infty)} F_U(y) dF_V(v) = F_V(0) + F_U(y)[1 - F_V(0)].$$

Proposition 3.7.3 *Suppose Y has a stable law. Then the normalizing constants are given by $a_n = n^{1/\alpha}$ for all $n \geq 1$, where $0 < \alpha \leq 2$. (α is called the* index.*)*

Proof. Let $\phi(t)$ be the c.f. of Y. It is a little easier to work with $\beta = 1/\alpha$. Define $W \stackrel{d}{=} Y - Y'$, where Y' is independent of Y and has the same distribution. If $W_1, \ldots, W_k$ are independent copies of W, then

$$\frac{W_1 + \cdots + W_k}{a_k} \stackrel{d}{=} \frac{Y_1 + \cdots + Y_k - b_k}{a_k} - \frac{Y_1' + \cdots + Y_k' - b_k}{a_k} \stackrel{d}{=} Y - Y' \stackrel{d}{=} W.$$

Thus W, which is symmetric, is stable with location constants zero and the same normalizing constants a_k as Y. Note also that $\phi_W(t)$ is real and strictly positive, the latter since W is infinitely divisible.

Hence we can assume that Y is symmetric. Since $Y_1 + \cdots + Y_{m+n} = (Y_1 + \cdots + Y_m) + (Y_{m+1} + \cdots + Y_{m+n})$, we see that

$$a_{m+n} Y \stackrel{d}{=} a_m Y' + a_n Y", \tag{7.1}$$

where Y' and $Y"$ are independent and have the same distribution as Y. If one takes $m = n$, then one obtains $a_{2n} Y \stackrel{d}{=} a_n(Y' + Y") \stackrel{d}{=} a_n a_2 Y'''$, so $a_{2n} = a_2 a_n$. A simple extension of the argument leading to (7.1) then tells us that $a_{mn} = a_m a_n$, and then that $a_{r^k} = a_r^k$ for all $r, k \geq 1$.

From (7.1), $Y \stackrel{d}{=} \frac{a_m}{a_{m+n}} Y' + \frac{a_n}{a_{m+n}} Y"$, so by the remark preceding this proposition, with $U = \frac{a_m}{a_{m+n}} Y'$ and $V = \frac{a_n}{a_{m+n}} Y"$, and using the fact that $F_U(y) = F_Y(\frac{a_{m+n}}{a_m} y)$ and $F_V(0) = F_Y(0)$, we have

$$F_Y(y) \leq F_Y(0) + F_Y(\frac{a_{m+n}}{a_m} y)[1 - F_Y(0)].$$

If there are sequences $\{m_i, i \geq 1\}$ and $\{n_j, j \geq 1\}$ such that $\frac{a_{m_i+n_j}}{a_{m_i}} \downarrow 0$, then $F_Y(y) \leq F_Y(0) + F_Y(0)[1 - F_Y(0)]$ for all $y > 0$, which would in turn imply that $1 \leq F_Y(0) + F_Y(0)[1 - F_Y(0)]$, so that $F_Y(0) = 1$, and then $P\{Y = 0\} = 1$ (since Y is symmetric). It follows that the numbers $\{\frac{a_m}{a_{m+n}} : m, n \geq 1\}$ are bounded, say by M. If we take $m = k^r$ and $m + n = (k+1)^r$, we have $\left(\frac{a_k}{a_{k+1}}\right)^r = \frac{a_m}{a_{m+n}} \leq M$ for all r. If there were k so that $\frac{a_k}{a_{k+1}} > 1$, then taking r large enough would contradict the validity of the bound M. Hence the sequence $\{a_n, n \geq 1\}$ is non-decreasing. The fact that $a_2 = a_2 a_1$ implies that $a_1 = 1$. If a_2 were 1, then $Y_1 + Y_2 \stackrel{d}{=} Y$, implying that $\phi(t)^2 = \phi(t)$, and therefore that $\phi(t) = 1$, so Y is degenerate at 0. This means $a_2 > 1$, and then by the lemma, that $a_n = n^\beta \forall n$.

Lastly, we prove that $\alpha \leq 2$. Suppose to the contrary that $b = 2\beta < 1$. Since $Y_1 + \cdots + Y_n \stackrel{d}{=} n^\beta Y$, then $\phi(t)^n = \phi(n^\beta t)$, so $\phi(n^{-\beta}) = \phi(1)^{1/n}$. Recall that $\phi(t)$ is real and strictly positive. By putting $t = n^{-\beta}$ and $s = 1/n$, we have

$$\liminf_{t\to0} \frac{1 - \phi(t)}{t^2} \leq \lim_{n\to\infty} \frac{1 - \phi(n^{-\beta})}{n^{-2\beta}} = \lim_{n\to\infty} \frac{1 - \phi(1)^{1/n}}{n^{-2\beta}} = \lim_{s\to0} \frac{1 - \phi(1)^s}{s^b} = 0,$$

where the last equality is a result of l'Hopital's rule and the fact that $b < 1$. Again by l'Hopital's rule we have $\lim_{t\to 0} \frac{1-\cos tx}{t^2} = \frac{x^2}{2}$, and so by the Fatou-Lebesgue lemma, $EY^2 = E\left[2\lim_{t\to 0} \frac{1-\cos tY}{t^2}\right] \le 2\liminf_{t\to 0} \frac{1-\phi(t)}{t^2} = 0$. This implies that Y is degenerate, a contradiction. ∎

Proposition 3.7.4 *Y has a stable law if and only if its c.f. $\phi(t)$ has the form $\phi(t) = e^{g(t)}$, where*

$$g(t) = itc - b|t|^\alpha \left[1 + i\kappa sign(t) w_\alpha(t)\right], \tag{7.2}$$

where $b > 0$, $c \in \mathbb{R}$ (c is just a centering constant), $-1 \le \kappa \le 1$ and

$$w_\alpha(t) = \begin{cases} \tan(\frac{\pi\alpha}{2}) & \text{if } 0 < \alpha \le 2, \alpha \ne 1, \\ \frac{2}{\pi}\log|t| & \text{if } \alpha = 1. \end{cases}$$

Proof. Suppose $\phi(t)$ has the given form. Let $\beta = 1/\alpha$. Then

$$w_\alpha(n^\beta t) = \begin{cases} w_\alpha(t) & \text{if } 0 < \alpha \le 2 \text{ and } \alpha \ne 1, \\ w_\alpha(t) + \frac{2}{\pi}\log n & \text{if } \alpha = 1, \end{cases}$$

so

$$\begin{aligned} g(n^\beta t) &= in^\beta tc - bn|t|^\alpha\left[1 + i\kappa \text{sign}(t) w_\alpha(n^\beta t)\right] = \begin{cases} ng(t) - itcn + itcn^\beta & \text{if } 0 < \alpha \le 2 \text{ and } \alpha \ne 1, \\ ng(t) - ibt\kappa\frac{2}{\pi} n\log n & \text{if } \alpha = 1, \end{cases} \\ &= ng(t) - itb_n, \end{aligned}$$

where $b_n = \begin{cases} c(n - n^\beta) & \text{if } 0 < \alpha \le 2 \text{ and } \alpha \ne 1 \\ \frac{2b\kappa}{\pi} n\log n & \text{if } \alpha = 1 \end{cases}$. Hence $\phi(t)^n = e^{ng(t)} = e^{g(a_n t) + itb_n} = \phi(a_n t)e^{itb_n}$ where $a_n = n^\beta$, which implies that ϕ is the c.f. of a stable law. ∎

Remark. We will denote the distribution whose c.f. is given in the above proposition by $\text{Stab}_\alpha(b, c, \kappa)$.

Examples.

(1) $\alpha = 2$. Then $g(t) = itc - bt^2$ and so $Y \sim N(c, b)$.

(2) $\alpha = 1, \kappa = 0$. Then $g(t) = itc - b|t|$, so $\frac{Y-c}{b}$ has the Cauchy distribution with density $\frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$.

(3) $\alpha = 1/2, \kappa = 1$. Then $g(t) = itc - b|t|^{1/2}[1 + i\text{sign}(t)]$. Then $\frac{Y-c}{b^2}$ has the inverse Gaussian distribution with density $(2\pi x^3)^{-1/2}e^{-1/2x}$, $x > 0$.

These are the only cases where the density is known in closed form. Hoffmann-Jorgensen (Probability with a View Toward Statistics, Volume I, p. 409) gives the densities in terms of hypergeometric functions in all other cases except when $\alpha = 1, \kappa \ne 0$.

Theorem 3.7.5 (The Stable Central Limit Theorem) *Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s such that*

$$\lim_{x\to\infty} x^\alpha P\{X_1 > x\} = u, \quad \lim_{x\to\infty} x^\alpha P\{X_1 < -x\} = v$$

where $0 < \alpha < 2$ and $u, v \in [0, \infty)$ are such that $u + v > 0$. Set

$$m_n = \begin{cases} 0 & \text{if } 0 < \alpha < 1, \\ nE\sin(\frac{X_1}{n}) & \text{if } \alpha = 1, \\ EX_1 & \text{if } 1 < \alpha < 2, \end{cases}$$

and

$$b = \alpha(u + v)C(\alpha), \qquad \kappa = \frac{v - u}{v + u},$$

where

$$C(\alpha) = \begin{cases} \frac{\Gamma(2-\alpha)\cos\frac{\alpha\pi}{2}}{\alpha(1-\alpha)} & \text{if } 0 < \alpha < 2,\ \alpha \neq 1, \\ \frac{\pi}{2} & \text{if } \alpha = 1. \end{cases}$$

Let $U_n = \frac{1}{n^{1/\alpha}} \sum_{j=1}^{n} (X_j - m_j)$ *for* $n \geq 1$*. Then* $U_n \xrightarrow{d} Stab_\alpha(b, 0, \kappa)$ *as* $n \to \infty$*.*

Problem. Show that Y has a symmetric stable law iff $\phi_Y(t) = e^{-b|t|^\alpha}$ where $b > 0$ and $0 < \alpha \leq 2$.

Problem. Prove that the stable distributions are all absolutely continuous with bounded continuous density functions.

Remarks.

(1) If X is stable with index $0 < \alpha < 2$, then $E|X|^p < \infty$ for $0 \leq p < \alpha$ and $E|X|^p = +\infty$ for $p \geq \alpha$.

(2) Let X and Y be independent strictly stable r.v.'s with indices α and β. If $Y \geq 0$ (so that $\beta < 1$), then $XY^{1/\alpha}$ is stable with index $\alpha\beta$ (Feller Vol. II, page 176).

(3) If $X \sim \mathrm{Stab}_\alpha(b, 0, \kappa)$ with $0 < \alpha < 1$, then $-X \sim \mathrm{Stab}_\alpha(b, 0, -\kappa)$. $X \sim \mathrm{Stab}(0, b, \alpha, -1)$ with $0 < \alpha < 1$, the support of X is $(-\infty, 0)$.

(4) Suppose $0 < \alpha < 1$. If $X \sim \mathrm{Stab}_\alpha(b, 0, \kappa)$, the support of X is

$$\mathrm{Supp}(X) = \begin{cases} (-\infty, 0) & \text{if } \kappa = -1, \\ (-\infty, +\infty) & \text{if } -1 < \kappa < 1, \\ (0, +\infty) & \text{if } \kappa = 1. \end{cases}$$

Reference: Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes.* Chapman and Hall, Boca Raton.

Chapter 4

Weak Convergence in $C[0,1]$ and Weiner Measure.

4.1 Relative Compactness in $C[0,1]$.

Notation. $C = C[0,1]$ will be the set of all continuous functions $x : [0,1] \to \mathbb{R}$. If $x, y \in C$, we define $d(x,y) = \sup_{t\in[0,1]} |x(t) - y(t)|$. With this metric, C is a separable complete metric space. Let $\mathcal{C}$ be the σ-algebra of Borel subsets of C. Given $x \in C$, we define the *modulus of continuity* of x to be $w_x(\delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|$, $0 < \delta < 1$.

Notation. Given $0 \le t_1 < t_2 \cdots < t_k \le 1$, we denote by $\pi_{t_1,\cdots,t_k} : C \to \mathbb{R}^k$ the projection operator defined by $\pi_{t_1,\cdots,t_k}(x) = (x(t_1), \ldots, x(t_k))$. Sets of the form $\pi^{-1}_{t_1,\cdots,t_k}(A)$, where $A \in \mathcal{B}(\mathbb{R}^k)$, are called (finite dimensional) cylinder sets. Let $\mathcal{C}_f$ denote the family of all cylinder sets. Then $\mathcal{C}_f \subset \mathcal{C}$ since the projections $\pi_{t_1,\cdots,t_k}$ are continuous and therefore measurable. If $P \in \mathcal{P}(C)$, the distribution $P\pi^{-1}_{t_1,\cdots,t_k}$ on $\mathcal{B}(\mathbb{R}^k)$ defined by $P\pi^{-1}_{t_1,\cdots,t_k}(A) = P[\pi^{-1}_{t_1,\cdots,t_k}(A)]$ is called a finite dimensional distribution of P.

Proposition 4.1.1 *$\mathcal{C}_f$ is closed under finite intersection, and $\sigma(\mathcal{C}_f) = \mathcal{C}$.*

Proof. Suppose that $U \in \mathcal{C}_f$ is of the form $U = \pi^{-1}_{s,u}(A)$, where $A \in \mathcal{B}(\mathbb{R}^2)$. Let t be such that $s < t < u$, and let $p : \mathbb{R}^3 \to \mathbb{R}^2$ denote the projection defined by $p(a,b,c) = (a,c)$. Then $\pi_{s,u} = p \circ \pi_{s,t,u}$, so $U = \pi^{-1}_{s,t,u} p^{-1}(A) = \pi^{-1}_{s,t,u}(D)$ where $D = p^{-1}(A) \in \mathbb{R}^3$. By embellishing this argument, one can show that if $U, V \in \mathcal{C}_f$, there is a common set $0 \le t_1 < t_2 \cdots < t_k \le 1$ of indices and $A, B \in \mathcal{B}(\mathbb{R}^k)$ such that $U = \pi^{-1}_{t_1,\cdots,t_k}(A)$ and $V = \pi^{-1}_{t_1,\cdots,t_k}(B)$. Then $U \cap V = \pi^{-1}_{t_1,\cdots,t_k}(A \cap B)$, so $\mathcal{C}_f$ is closed under finite intersection.

Next, for any $x \in C$ and $\epsilon > 0$, $\{y \in C : d(x,y) \le \epsilon\} = \cap_{r\in\mathbb{Q}\cap[0,1]} |\{y : |x(r) - y(r)| \le \epsilon\} = \cap_{r\in\mathbb{Q}\cap[0,1]} \pi_r^{-1}[x(r) - \epsilon, x(r) + \epsilon] \in \sigma(\mathcal{C}_f)$. That is, every closed ball, and therefore every open ball, belongs to $\sigma(\mathcal{C}_f)$. By separability, all the open sets in C belong to $\sigma(\mathcal{C}_f)$, so $\mathcal{C} \subset \sigma(\mathcal{C}_f)$. We noted above that $\mathcal{C}_f \subset \mathcal{C}$ and hence $\sigma(\mathcal{C}_f) \subset \mathcal{C}$. ∎

Theorem 4.1.2 *Let $\{P_n, n \ge 1\} \subset \mathcal{P}(C)$, and let $P \in \mathcal{P}(C)$. If*

(1) $\{P_n, n \ge 1\}$ is tight, and

(2) the finite dimensional distributions of P_n converge weakly to those of P,

then $P_n \xrightarrow{w} P$.

Proof. Let $\{P_{n_i}, i \ge 1\}$ be a weakly convergent subsequence of $\{P_n, n \ge 1\}$, and suppose $Q \in \mathcal{P}(C)$ is such that $P_{n_i} \xrightarrow{w} Q$. By the mapping theorem, the finite dimensional distributions of P_{n_i} converge weakly to those of Q. Since they must also converge to those of P, then P and Q have the same finite dimensional distributions, so $P = Q$ on $\mathcal{C}_f$, so $P = Q$. ∎

Theorem 4.1.3 (Arzela-Ascoli) *A subset A of $C[0,1]$ has compact closure if and only if*

(1) $\sup_{x\in A}|x(0)|<\infty$*, and*

(2) $\lim_{\delta\to 0}\sup_{x\in A} w_x(\delta)=0$ *(i.e. A is equicontinuous).*

Proposition 4.1.4 *Let $\{P_n, n\geq 1\}$ be probability measures on $C[0,1]$. Then $\{P_n, n\geq 1\}$ is tight if and only if*

(1) for each $\eta>0$, $\exists a>0$ and $n_0\geq 1$ such that $P_n\{x:|x(0)|>a\}\leq\eta\ \forall n\geq n_0$, and

(2) for each $\epsilon>0,\eta>0$, $\exists 0<\delta<1$ and $n_0\geq 1$ such that $P_n\{x:w_x(\delta)\geq\epsilon\}\leq\eta\ \forall n\geq n_0$.

Proof. Suppose $\{P_n\}$ is tight. Given η, choose K compact $\ni P_n(K)>1-\eta\forall n$. By the A-A theorem, $a=^{\text{def}}\sup_{x\in K}|x(0)|<\infty$ is such that $K\subset\{x:|x(0)|\leq a\}$ (so (1) holds), and given $\epsilon>0$, $\exists\delta>0\ni \sup_{x\in K}w_x(\delta)<\epsilon$ or equivalently $K\subset\{x:w_x(\delta)<\epsilon\}$ (so (2) holds), with $n_0=1$ in each case (which is important for the converse).

Conversely, suppose conditions (1) and (2) hold, and let $\eta>0$. Since C is separable and complete, the finite family $\{P_n, 1\leq n\leq n_0\}$ is tight, and so by the previous paragraph $\exists a\ni P_n\{x:|x(0)|\geq a\}\leq\eta\forall 1\leq n\leq n_0$, and for each $\epsilon>0, \exists 0<\delta<1\ni P_n\{x:w_x(\delta)\geq\epsilon\}\leq\eta\forall 1\leq n\leq n_0$. The point is that we can assume (1) and (2) hold with $n_0=1$. Thus, given $\eta>0$, let a be such that $P_n\{x:|x(0)|\geq a\}\leq\eta/2\forall n\geq 1$, and choose a sequence $\delta_k\downarrow 0$ such that $P_n\{x:w_x(\delta_k)\geq 1/k\}\leq\eta/2^{k+1}\forall n\geq 1$. Let $D=\{x:|x(0)|\geq a\}\cup_{k=1}^{\infty}\{x:w_x(\delta_k)\geq 1/k\}$. Then by subadditivity, $P_n(D)\leq\eta\forall n\geq 1$. Since D^c satisfies conditions (1) and (2) in the A-A theorem, then $K=\overline{D^c}$ is compact. Since $P_n(K)\geq P_n(D^c)\geq 1-\eta\forall n\geq 1$, we are finished. ∎

Lemma 4.1.5 *Suppose that $0=t_0<t_1<\cdots<t_m=1$ and $\min_{1<i<m}(t_i-t_{i-1})\geq\delta$. Then for any $x\in C$,*

$$w_x(\delta)\leq 3\max_{1\leq i\leq m}\sup_{t_{i-1}\leq s\leq t_i}|x(s)-x(t_{i-1})|. \tag{1.1}$$

For any $P\in\mathcal{P}(\mathcal{C})$,

$$P\{x|w_x(\delta)\geq 3\epsilon\}\leq\sum_{i=1}^{m}P\{x:\sup_{t_{i-1}\leq s\leq t_i}|x(s)-x(t_{i-1})|\geq\epsilon\}. \tag{1.2}$$

Proof. Let M be the maximum on the RHS of (1.1), and let $I_i=[t_{i-1},t_i],\ 1\leq i\leq m$. If $|s-t|\leq\delta$, then either s and t must lie in the same interval, say I_i, in which case $|x(s)-x(t)|\leq|x(s)-x(t_{i-1})|+|x(t)-x(t_{i-1})|\leq 2M$, or s and t must lie in adjacent intervals, say I_i and I_{i+1} respectively, in which case $|x(s)-x(t)|\leq|x(s)-x(t_{i-1})|+|x(t_i)-x(t_{i-1})|+|x(t)-x(t_i)|\leq 3M$. Hence (1.1). Finally,

$$\{x|w_x(\delta)\geq 3\epsilon\}\subset\{x|\max_{1\leq i\leq m}\sup_{t_{i-1}\leq s\leq t_i}|x(s)-x(t_{i-1})|\geq\epsilon\}=\cup_{i=1}^{m}\{x:\sup_{t_{i-1}\leq s\leq t_i}|x(s)-x(t_{i-1})|\geq\epsilon\},$$

from which the last statement follows by subadditivity of P. ∎

Proposition 4.1.6 *Condition (2) of proposition 4.1.4 holds if for each $0<\epsilon,\eta<1$, $\exists\ 0<\delta<1$ and $n_0\geq 1$ such that*

$$\frac{1}{\delta}P_n\{\sup_{t\leq s\leq t+\delta}|x(s)-x(t)|\geq\epsilon\}\leq\eta\ \forall n\geq n_0,$$

for every $t\in[0,1]$.

Proof. In the lemma, take $t_i=i\delta$ for $1\leq i<m=[1/\delta]$. Then $P_n\{x:w_x(\delta)\geq 3\epsilon\}\leq\sum_{i=1}^{m}P_n\{x:\sup_{t_{i-1}\leq s\leq t_i}|x(s)-x(t_{i-1})|\geq\epsilon\}\leq m\delta\eta\ \forall n\geq n_0$. ∎

4.2 Random Elements with Values in $C[0,1]$.

Notation. Let $\xi_1, \xi_2, \dots$ be a sequence of r.v.'s on some probability space, let $S_n = \xi_1 + \cdots + \xi_n$ for all n, and define

$$X_n(t) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}, \quad 0 \le t \le 1, n \ge 1, \tag{2.1}$$

where $\sigma > 0$. Here, $[nt]$ denotes the integer part of the number nt. In particular, $X_n(i/n) = \frac{1}{\sigma\sqrt{n}} S_i$ for points $i/n \in [0,1]$ where i is an integer; $X_n(t)$ is just the function of t obtained by linear interpolation of the values $X_n(i/n)$, and so $X_n(t)$ is a continuous function of t.

Proposition 4.2.1 *Suppose that for each $\epsilon > 0$, $\exists \lambda > 1$ and $n_0 \ge 1$ such that if $n \ge n_0$, then*

$$P\{\max_{i \le n} |S_{k+i} - S_k| \ge \lambda\sigma\sqrt{n}\} \le \frac{\epsilon}{\lambda^2} \quad \forall k \ge 1.$$

Then $\forall 0 < \epsilon, \eta < 1$, $\exists 0 < \delta < 1$ and $n_0 \ge 1$ such that

$$\frac{1}{\delta} P\{\sup_{t \le s \le t+\delta} |X_n(s) - X_n(t)| \ge \epsilon\} \le \eta, \quad n \ge n_0,$$

for every $t \in [0,1]$.

Proof. By hypothesis, $\exists \lambda > 1$ and m_0 such that $P\{\max_{i \le m} |S_{k+i} - S_k| \ge \lambda\sigma\sqrt{m}\} \le \frac{\eta\epsilon^2}{\lambda^2} \quad \forall m \ge m_0,\ k \ge 1$. Let $\delta = \epsilon^2/\lambda^2$ and let n_0 be an integer larger than m_0/δ. For all n, we have $[n\delta] \le n\delta = n\epsilon^2/\lambda^2$, and so $\lambda\sqrt{[n\delta]} \le \epsilon\sqrt{n}$. Suppose $n \ge n_0$. Then $[n\delta] \ge [n_0\delta] \ge m_0$, and so

$$\begin{aligned} P\{\max_{i \le [n\delta]} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k| \ge \epsilon\} &= P\{\max_{i \le [n\delta]} |S_{k+i} - S_k| \ge \sigma\epsilon\sqrt{n}\} \le P\{\max_{i \le [n\delta]} |S_{k+i} - S_k| \ge \lambda\sigma\sqrt{[n\delta]}\} \\ &\le \frac{\eta\epsilon^2}{\lambda^2} = \delta\eta. \end{aligned} \tag{2.2}$$

Next, for a given pair s, t with $t \le s \le t + \delta$, let $0 \le k < j$ be integers such that

$$\frac{k}{n} \le t < \frac{k+1}{n}, \quad \frac{j-1}{n} \le t + \frac{\delta}{2} < \frac{j}{n}. \tag{2.3}$$

Because of the polygonal character of $X_n(t)$, we have (draw a picture) $|X_n(u) - X_n(\frac{k}{n})| \le \max_{0 \le i \le j-k} |X_n(\frac{k+i}{n}) - X_n(\frac{k}{n})|$ $\forall u$ with $\frac{k}{n} \le u \le \frac{j}{n}$. Writing $|X_n(s) - X_n(t)| \le |X_n(s) - X_n(\frac{k}{n})| + |X_n(t) - X_n(\frac{k}{n})|$, we obtain

$$\sup_{t \le s \le t+\frac{\delta}{2}} |X_n(s) - X_n(t)| \le 2 \max_{0 \le i \le j-k} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k|.$$

An easy calculation from (2.3) gives $j - k < \frac{\delta n}{2} + 2$. If $n \ge \frac{4}{\delta}$, then $2 \le \frac{n\delta}{2}$, so $j - k < n\delta$, and then

$$\sup_{t \le s \le t+\frac{\delta}{2}} |X_n(s) - X_n(t)| \le 2 \max_{0 \le i \le [n\delta]} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k|.$$

Combining this with (2.2) gives

$$P\{\sup_{t \le s \le t+\frac{\delta}{2}} |X_n(s) - X_n(t)| > \epsilon\} \le P\{\max_{0 \le i \le [n\delta]} \frac{1}{\sigma\sqrt{n}} |S_{k+i} - S_k| > \frac{\epsilon}{2}\} \le \frac{\eta\epsilon^2}{4\lambda^2} = \frac{\delta\eta}{4}, \quad n \ge n_0 \stackrel{\text{def}}{=} \frac{\max\{m_0, 4\}}{\delta},$$

and therefore $\frac{2}{\delta} P\{\sup_{t \le s \le t+\frac{\delta}{2}} |X_n(s) - X_n(t)| > \epsilon\} \le \frac{\eta}{2}$, which is equivalent to the required result. ∎

Proposition 4.2.2 *Let $\xi_1, \xi_2, \dots, \xi_n$ be independent random variables with mean 0 and finite variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$. Let $S_n = \xi_1 + \cdots + \xi_n$ and $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$. Then*

$$P\{\max_{i\le n} |S_i| \ge \lambda s_n\} \le 2P\{|S_n| \ge (\lambda - \sqrt{2})s_n\}.$$

Proof. We can assume $\lambda > \sqrt{2}$ since otherwise the result is trivial. Define $E_i = \{\max_{j<i} |S_j| < \lambda s_n \le |S_i|\}$, $i = 1, \dots, n-1$. Then

$$\begin{aligned} P\{\max_{i\le n} |S_i| \ge \lambda s_n\} &= P\{\max_{i\le n} |S_i| \ge \lambda s_n, |S_n| \ge (\lambda - \sqrt{2})s_n\} + P\{\max_{i\le n} |S_i| \ge \lambda s_n, |S_n| < (\lambda - \sqrt{2})s_n\} \\ &\le P\{|S_n| \ge (\lambda - \sqrt{2})s_n\} + \sum_{i=1}^{n-1} P[E_i \cap \{|S_n| < (\lambda - \sqrt{2})s_n\}] \end{aligned}$$

On the set $E_i \cap \{|S_n| < (\lambda - \sqrt{2})s_n\}$, we have $|S_i| - |S_n| \ge \lambda s_n - (\lambda - \sqrt{2})s_n = \sqrt{2}s_n$, and hence $|S_n - S_i| \ge ||S_n| - |S_i|| \ge \sqrt{2}s_n$. This gives $P[E_i \cap \{|S_n| < (\lambda - \sqrt{2})s_n\}] \le P[E_i \cap \{|S_n - S_i| \ge \sqrt{2}s_n\}] = P(E_i)P\{|S_n - S_i| \ge \sqrt{2}s_n\}] \le P(E_i)\frac{\sum_{k=i+1}^n \sigma_k^2}{2s_n^2} \le \frac{P(E_i)}{2}$, where we used independence and then Chebychev's inequality. Since $\sum_{i=1}^{n-1} P(E_i) = P\{\max_{i\le n-1} |S_i| \ge \lambda s_n\}$, we then have

$$P\{\max_{i\le n} |S_i| \ge \lambda s_n\} \le P\{|S_n| \ge (\lambda - \sqrt{2})s_n\} + \frac{1}{2}\sum_{i=1}^{n-1} P(E_i) \le P\{|S_n| \ge (\lambda - \sqrt{2})s_n\} + \frac{1}{2}P\{\max_{i\le n} |S_i| \ge \lambda s_n\}.$$

∎

Corollary 4.2.3 *Suppose that $\xi_1, \xi_2, \dots$ are i.i.d. random variables with mean 0 and finite variance σ^2. Then for any $\epsilon > 0$, $\exists \lambda > 1$ and $n_0 \ge 1$ such that $P\{\max_{i\le n} |S_i| \ge \lambda\sigma\sqrt{n}\} \le \frac{\epsilon}{\lambda^2}$ $\forall n \ge n_0$.*

Proof. In this case, $s_n = \sigma\sqrt{n}$. Let $\lambda > 2\sqrt{2}$ (so that $\lambda - \frac{\lambda}{2} > \sqrt{2}$, so that $\lambda - \sqrt{2} > \frac{\lambda}{2}$) be such that $\frac{32\sqrt{2}}{\lambda\sqrt{\pi}} \le \frac{\epsilon}{2}$. Then

$$\begin{aligned} 2P\{|S_n| \ge (\lambda - \sqrt{2})\sigma\sqrt{n}\} &\le 2P\{|S_n| \ge \frac{\lambda\sigma\sqrt{n}}{2}\} = 2P\{\frac{|S_n|}{\sigma\sqrt{n}} \ge \frac{\lambda}{2}\} \\ &\to 2P\{|N| \ge \frac{\lambda}{2}\} \le \frac{16E|N^3|}{\lambda^3} = \frac{32\sqrt{2}}{\lambda^3\sqrt{\pi}} \le \frac{\epsilon}{2\lambda^2}. \end{aligned}$$

by the CLT, where $N \sim N(0,1)$ and we used the fact that $E|N|^3 = 2\sqrt{\frac{2}{\pi}}$.

4.3 Weiner Measure and the Invariance Principle.

Definition. A probability measure W on $(C = C(0,1), \mathcal{C})$ such that

(1) $W\{x : x(0) = 0\} = 1$,

(2) for each $t > 0$, $x(t) \sim N(0,t)$ under W,

(3) for any $0 \le t_0 < t_1 < \cdots < t_k \le 1$, $x(t_1) - x(t_0), x(t_2) - x(t_1), \dots, x(t_k) - x(t_{k-1})$ are independent under W,

is called Weiner measure. Note that, if $s < t$, then $x(t) = x(s) + [x(t) - x(s)]$ independent, so $\phi_{[x(t)-x(s)]}(u) = \phi_{x(t)}(u)/\phi_{x(s)}(u) = e^{tu^2/2}/e^{su^2/2} = e^{(t-s)u^2/2}$, so $x(t) - x(s) \overset{d}{=} N(0, t-s)$ under W. Hence $(x(t_1), x(t_2) - x(t_1), \dots, x(t_k) - x(t_{k-1})) \overset{d}{=} (\sqrt{t_1}N_1, \sqrt{t_2 - t_1}N_2, \dots, \sqrt{t_k - t_{k-1}}N_k)$, where $N_1, \dots, N_k$ are independent $N(0,1)$ r.v.'s.

Suppose that the r.v.'s $\xi_1, \xi_2, \ldots$ in the definition of X_n in (2.1) are i.i.d. with mean 0 and variance σ^2. If $0 \le s < t \le 1$, then

$$X_n(t) - X_n(s) = \frac{S_{[nt]} - S_{[ns]}}{\sigma\sqrt{n}} + \psi_n(t) - \psi_n(s),$$

where $\psi_n(t) = (nt - [nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}$. Firstly, by the Lindeberg-Lévy theorem,

$$\frac{S_{[nt]} - S_{[ns]}}{\sigma\sqrt{n}} = \frac{S_{[nt]} - S_{[ns]}}{\sigma\sqrt{[nt]-[ns]}} \cdot \sqrt{\frac{[nt]}{n} - \frac{[ns]}{n}} \xrightarrow{\mathrm{d}} \sqrt{t-s}N,$$

where $N \sim N(0,1)$ and we used the fact that $\frac{[nt]}{n} \to t$ as $n \to \infty$. Next, by Chebychev,

$$P\{|\psi_n(t)| > \epsilon\} \le \frac{(nt-[nt])^2 \mathrm{Var}(\xi_{[nt]+1})}{n\sigma^2\epsilon^2} = \frac{(nt-[nt])^2}{n\epsilon^2} \le \frac{1}{n\epsilon^2} \to 0,$$

so $\psi_n(t) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. Consequently, by proposition 1.5.3, $X_n(t) - X_n(s) \xrightarrow{\mathrm{d}} \sqrt{t-s}N$. Next, if $U_n \xrightarrow{\mathrm{d}} U$ and $V_n \xrightarrow{\mathrm{d}} V$, and if U_n and V_n are independent r.v.'s for each n, then it is easy to see that $(U_n, V_n) \xrightarrow{\mathrm{d}} (U, V)$ where U and V are independent. It follows from the k-dimensional version of this fact that if $0 < t_1 < \cdots < t_k \le 1$, then

$$\begin{aligned}(X_n(t_1), X_n(t_2) - X_n(t_1), \ldots, X_n(t_k) - X_n(t_{k-1})) &\xrightarrow{\mathrm{d}} (\sqrt{t_1}N_1, \sqrt{t_2 - t_1}N_2, \ldots, \sqrt{t_k - t_{k-1}}N_k) \\ &= \text{ the } W\text{-distribution of } (x(t_1), x(t_2) - x(t_1), \ldots, x(t_k) - x(t_{k-1})).\end{aligned}$$

By the mapping theorem, with $h(x) = Ax$, where

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

we then have $(X_n(t_1), X_n(t_2), \ldots, X_n(t_k)) \xrightarrow{\mathrm{d}}$the W-distribution of $(x(t_1), x(t_2), \ldots, x(t_k))$. Consequently, the finite dimensional distributions of X_n converge weakly to those of W.

For each n, let P_n be the distribution of X_n on C. By combining corollary 4.2.3 with proposition 4.2.1 and proposition 4.1.6, we see that condition (2) of proposition 4.1.4 holds. Since $X_n(0) = 0$, then $P_n\{|x(0)| = 0\} = 1 \forall n$, so condition (1) of proposition 4.1.4 also holds. Hence the sequence $\{P_n, n \ge 1\}$ is tight.

Theorem 4.3.1 *Weiner measure exists.*

Proof. Since $\{P_n, n \ge 1\}$ is tight, $\exists Q \in \mathcal{P}(C)$ and a subsequence $\{P_{n_i}, i \ge 1\} \ni P_{n_i} \xrightarrow{\mathrm{w}} Q$ as $i \to \infty$. Since the finite dimensional distributions of P_{n_i} converge weakly to those of Q, and also (by the preceding paragraphs) to those of the postulated measure W, then Q has the same finite dimensional distributions as for the intended W. Since Q is uniquely determined by its finite dimensional distributions, then Q must be Weiner measure. ∎

Theorem 4.3.2 (Donsker's Invariance Principle) *Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d. r.v.'s on some probability space, each having finite variance σ^2. For each n, let P_n be the distribution of X_n, as defined in (2.1), on $C[0,1]$. Then $P_n \xrightarrow{\mathrm{w}} W$.*

Proof. This follows directly from theorem 4.1.2 and the two paragraphs before the preceding theorem. ∎

Remark. Taking $t = 1$ gives $X_n(1) = \frac{S_n}{\sigma\sqrt{n}}$, so Donsker's IP contains the classical CLT.

Chapter 5

Brownian Motion

5.1 Definition and Properties.

Definition. A stochastic process $\{W_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in $\mathbb{R}$ is a *Brownian Motion on* $\mathbb{R}$ if

(1) $W_0 \equiv 0$,

(2) given "times" $0 \leq t_0 < t_1 < t_2 < \cdots < t_n < \infty$, the "increments" $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent,

(3) $W_t - W_s \sim N(0, t-s)$ whenever $0 \leq s < t$,

(4) for P-a.s. $\omega \in \Omega$, the path $W_t(\omega)$ is continuous in t.

Remark. Properties (2) and (3) are equivalent to

(5) $\{W_t, t \geq 0\}$ is a Gaussian process (i.e. for any $0 < t_1 < t_2 < \cdots < t_n < \infty$, the joint distribution of $W_{t_1}, \ldots, W_{t_n}$ is Gaussian) with mean 0 (i.e. $EW_t = 0 \forall t \geq 0$) and covariance function $\mathrm{Cov}(W_s, W_t) = \min\{s, t\} \forall s, t > 0$.

For if (2) and (3) hold, then $\{W_t, t \geq 0\}$ is obviously Gaussian with mean zero, and if $0 < s < t$, we have $EW_sW_t = EW_s(W_t - W_s) + EW_s^2 = EW_s^2 = s$. Conversely, if (5) holds, and if $0 \leq s < t \leq u < v$, then $\mathrm{Cov}(W_v - W_u, W_t - W_s) = \mathrm{Cov}(W_v, W_t) - \mathrm{Cov}(W_v, W_s) - \mathrm{Cov}(W_u, W_t) + \mathrm{Cov}(W_u, W_s) = t - s - t + s = 0$. It follows from this that the increments in (2), which must be jointly Gaussian, are independent. Finally, we have $\mathrm{Var}(W_t - W_s) = \mathrm{Var}(W_t) - 2\mathrm{Cov}(W_t, W_s) + \mathrm{Var}(W_s) = t - 2s + s = t - s$, finishing (3).

Existence. Consider the probability space $(C[0,1], \mathcal{C}, W)$ of the previous section, and the coordinate functions $x(t), t \in [0,1]$. Define $U_t = tx(1/t),\ t \geq 1$ and $W_t = U_{t+1} - U_1, t \geq 0$. Then $\{W_t, t \geq 0\}$ is a Brownian motion.

(1) Clearly $W_0 = 0$.

(5) Since the finite dimensional distributions of the x-process are multivariate Gaussian, then so are those of the U-process, and therefore those of $\{W_t, t \geq 0\}$. So $\{W_t, t \geq 0\}$ is a Gaussian process. Obviously $EW_t = 0 \forall t \geq 0$. If $1 \leq u \leq v$, then $\mathrm{Cov}(U_u, U_v) = uv\mathrm{Cov}[x(1/u), x(1/v)] = uv \min\{1/u, 1/v\} = u$. Suppose $0 \leq s \leq t$. Then $EW_sW_t = E(U_{s+1} - U_1)(U_{t+1} - U_1) = EU_{s+1}U_{t+1} - EU_{s+1}U_1 - EU_1U_{t+1} + EU_1^2 = (s+1) - 1 - 1 + 1 = s$. Hence $\mathrm{Cov}(W_s, W_t) = \min\{s, t\}$.

(4) Obvious.

Proposition 5.1.1 (Simple Properties of Brownian Motion) *Suppose* $\{W_t, t \geq 0\}$ *is Brownian motion on* $\mathbb{R}$.

(1) Then so also are

(a) $\{W_{t+s} - W_s, t \geq 0\}$ *(change of origin),*

(b) $\{\frac{1}{c} W_{c^2 t}, t \geq 0\}$*, where* $c \neq 0$ *(scale change),*

(c) $\{B_t, t \geq 0\}$*, where* $B_t = \begin{cases} tW(1/t) & \text{if } t > 0, \\ 0 & \text{if } t = 0, \end{cases}$ *(time inversion),*

(d) $\{W_\tau - W_{\tau - t}, 0 \leq t \leq \tau\}$ *where* $\tau > 0$ *is fixed (time reversal).*

(2) $\{W_t, t \geq 0\}$ *is a martingale,*

(3) $\{W_t, t \geq 0\}$ *is a Markov process.*

Proof.

(1) All four processes are Gaussian and have mean zero. One needs only to check the covariance functions.

(2) Let $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t) = \sigma(W_v - W_u, 0 \leq u \leq v \leq t)$. Then $E[W_t|\mathcal{F}_s] = E[W_t - W_s|\mathcal{F}_s] + E[W_s|\mathcal{F}_s] = 0 + W_s = W_s$ a.s.

(3) Also because of independent increments.

Definition. A function $f : [a, b] \to \mathbb{R}$ is *of bounded variation* if there is $M > 0$ such that $\sum_{i=1}^{n} |f(t_i) - f(t_{i-1})| \leq M < \infty$ for all partitions $a = t_0 < t_1 < \cdots < t_n = b$ of $[a, b]$.

Proposition 5.1.2 *Let* $\{W_t, t \geq 0\}$ *be a Brownian motion defined on a probability space* $(\Omega, \mathcal{F}, P)$*, and let* $[a, b]$ *be a finite subinterval of* $\mathbb{R}$*. Then almost every path* $W_t(\omega)$ *is not of bounded variation on* $[a, b]$*.*

Proof. It suffices to prove this for $[0, 1]$. Define $v_n(\omega) = \sum_{i=1}^{2^n} |W(\frac{i}{2^n}) - W(\frac{i-1}{2^n})|$, and $v(\omega) = \lim_{n\to\infty} \uparrow v_n(\omega)$. We will show that $v = +\infty$ for a.s. ω. Since $E|N(0, 1)| = b = \sqrt{\frac{2}{\pi}}$ and $\mathrm{Var}(|N|) = c = 1 - \frac{2}{\pi}$, and since $W(\frac{i}{2^n}) - W(\frac{i-1}{2^n}) \sim N(0, \frac{1}{2^n}) \sim \frac{1}{2^{n/2}} N(0, 1)$, then $Ev_n = 2^{n/2} b$ and $\mathrm{Var}(v_n) = c$. Then

$$P\{v_n > \alpha\} \geq P\{|v_n - Ev_n| \leq Ev_n - \alpha\} \geq 1 - \frac{\mathrm{Var}(v_n)}{(Ev_n - \alpha)^2} = 1 - \frac{c}{(2^{n/2} b - \alpha)^2} \to 1 \tag{1.1}$$

as $n \to \infty$. Then $P\{v > \alpha\} = \lim_{n\to\infty} P\{v_n > \alpha\} = 1$ for all $\alpha > 0$. ∎

Note: In the first inequality in (1.1), we used the fact that if $|x - y| \leq y - \alpha$, then $x \geq \alpha$.

Proposition 5.1.3 *Almost every path* $W_t(\omega)$ *is nowhere differentiable on* $[0, +\infty)$*.*

Proof.

Chapter 6

The Kolmogorov Extension Theorem.

Notation. Let T be a set (called the index set), and for each $t \in T$, let $(\Omega_t, \mathcal{F}_t)$ be a measurable space. Let $\Pi_{t\in T}\Omega_t$ denote the set of all functions $\omega : T \to \cup_{t\in T}\Omega_t$ such that $\omega(t) \in \Omega_t$ for all $t \in T$. $\Pi_{t\in T}\Omega_t$ is called the *product space*.

If T is a finite set, we already know how to define the product σ-algebra $\otimes_{t\in T}\mathcal{F}_t$ of subsets of $\Pi_{t\in T}\Omega_t$. We now wish to do the same when T is infinite.

Let S be a subset of T. A subset of $\Pi_{t\in T}\Omega_t$ of the form $C = A \times \Pi_{t\in T\setminus S}\Omega_t$, where $A \in \otimes_{t\in S}\mathcal{F}_t$, is called a cylinder set with base A. If S is finite, C is called a finite-dimensional cylinder set. Let $\mathcal{F}_S$ denote the family of all cylinder sets with bases in $\otimes_{t\in S}\mathcal{F}_t$. Since a cylinder set with base A can be identified with A, we can identify $\mathcal{F}_S$ with $\otimes_{t\in S}\mathcal{F}_t$.

Proposition 6.0.4 *Let $\mathcal{C}_f = \cup\{\mathcal{F}_S : S$ is a finite subset of $T\}$ be the family of all finite dimensional cylinder sets. Then $\mathcal{C}_f$ is an algebra of subsets of $\Pi_{t\in T}\Omega_t$.*

Proof. First note that if S_1 and S_2 are finite subsets of T with $S_1 \subset S_2$, and if $C = A_1 \times \Pi_{t\in T\setminus S_1}\Omega_t$ is a cylinder set with base $A_1 \in \otimes_{t\in S_1}\mathcal{F}_t$, then C can also be represented as $C = A_2 \times \Pi_{t\in T\setminus S_2}\Omega_t$ where $A_2 = A_1 \times \Pi_{t\in S_2\setminus S_1}\Omega_t \in \otimes_{t\in S_2}\mathcal{F}_t$. This means that any two cylinder sets can be represented with respect to the same finite subset of T. Thus, if C_1 and C_2 are two cylinder sets, we can write $C_1 = A_1 \times \Pi_{t\in T\setminus S}\Omega_t$ and $C_2 = A_2 \times \Pi_{t\in T\setminus S}\Omega_t$ where $A_1, A_2 \in \otimes_{t\in S}\mathcal{F}_t$, and then $C_1 \cup C_2 = (A_1 \cup A_2) \times \Pi_{t\in T\setminus S}\Omega_t$ and $C_1 \cap C_2 = (A_1 \cap A_2) \times \Pi_{t\in T\setminus S}\Omega_t$ are cylinder sets as well.

If $C = A \times \Pi_{t\in T\setminus S}\Omega_t$ is a cylinder set, then $C^c = A^c \times \Pi_{t\in T\setminus S}\Omega_t$ is also a cylinder set. Finally, $\Pi_{t\in T}\Omega_t$ is itself a cylinder set since $\Pi_{t\in T}\Omega_t = A \times \Pi_{t\in T\setminus S}\Omega_t$ with $A = \Pi_{t\in S}\Omega_t$ for any finite subset S of T. ∎

Definition. We take the product σ-algebra $\otimes_{t\in T}\mathcal{F}_t$ to be the σ-algebra of subsets of $\Pi_{t\in T}\Omega_t$ generated by the algebra $\mathcal{C}_f$ of finite dimensional cylinder sets.

Notation. For brevity, we let $\Omega = \Pi_{t\in T}\Omega_t$ and $\mathcal{F} = \otimes_{t\in T}\mathcal{F}_t$. For each $s \in T$, let $X_s : \Omega \to \Omega_s$ denote the sth coordinate mapping defined by $X_s(\omega) = \omega(s)$. Note that X_s is measurable with respect to $\mathcal{F}$ and $\mathcal{F}_s$. In fact, $\mathcal{F}$ is the smallest σ-algebra which makes all the coordinate mappings measurable.

Proposition 6.0.5 *For each finite subset S of T, let P_S be a probability defined on $\otimes_{t\in S}\mathcal{F}_t$. Suppose the P_S's satisfy the consistency condition: if $S_1 \subset S_2$, then P_{S_2} restricted to $\otimes_{t\in S_1}\mathcal{F}_t$ coincides with P_{S_1} (that is, if $A \in \otimes_{t\in S_1}\mathcal{F}_t$, then $P_{S_2}[A \times \Pi_{t\in S_2\setminus S_1}\Omega_t] = P_{S_1}(A)$. Then there exists an additive set function P on $\mathcal{C}_f$ such that $P(\Omega) = 1$ and which coincides with P_S on every $\mathcal{F}_S$.*

Proof. If $B \in \mathcal{C}_f$, then $B \in \mathcal{F}_S$ for some finite S; define $P(B) = P_S(B)$. If $B \in \mathcal{F}_{S_1}$ and $B \in \mathcal{F}_{S_2}$ where S_1 and S_2 are finite, then $P_{S_1}(B) = P_{S_2}(B)$ because of the consistency condition. Thus P is unambiguously defined. If $A, B \in \mathcal{C}_f$, and $A \cap B = \emptyset$, there is a finite S so that $A, B \in \mathcal{F}_S$. Then $A \cup B \in \mathcal{F}_S$ and $P(A \cup B) = P_S(A \cup B) = P_S(A) + P_S(B) = P(A) + P(B)$. ∎

In order that P be σ-additive on $\mathcal{C}_f$, we need additional conditions on the sets Ω_t.

Definition. A class $\mathcal{C}$ of subsets of a set E is said to be compact if for every sequence $\{C_n, n \geq 1\}$ in $\mathcal{C}$ such that $\cap_{n=1}^{\infty} C_n = \emptyset$, there is an integer N such that $\cap_{n=1}^{N} C_n = \emptyset$. A probability space $(E, \mathcal{E}, \mu)$ is said to satisfy a compactness condition if there is a compact subclass $\mathcal{C}$ of $\mathcal{E}$ such that $\mu(A) = \sup\{\mu(C) : C \in \mathcal{C},\ C \subset A\} \forall A \in \mathcal{E}$.

Remark. If E is a Polish space (i.e. a complete separable metric space), $\mathcal{E}$ is the Borel σ-algebra, and μ is any probability on $\mathcal{E}$, then $(E, \mathcal{E}, \mu)$ satisfies a compactness condition. Note that $\mathbb{R}^n$ and any countable set are examples of Polish spaces.

Theorem 6.0.6 (Kolmogorov's Extension Theorem) *For each finite subset S of T, let P_S be a probability defined on $\otimes_{t \in S} \mathcal{F}_t$. Suppose the P_S's satisfy the consistency condition of the preceding proposition and that, for each $t \in T$, $(\Omega_t, \mathcal{F}_t, P_{\{t\}})$ satisfies a compactness condition. Then there exists a unique probability P on the product σ-algebra $\mathcal{F}$ which extends all the P_S's. Stated another way, there exists a probability space $(\Omega, \mathcal{F}, P)$ and on it a stochastic process $\{X_t, t \in T\}$ (the coordinate mappings) having the P_S's as its finite-dimensional distributions (i.e. $P\{(X_{t_1}, \dots, X_{t_n}) \in A\} = P_{\{t_1, \dots, t_n\}}(A)$ for all $A \in \otimes_{i=1}^{n} \mathcal{F}_{t_i}$ and finite subsets $\{t_1, \dots, t_n\}$ of T).*

Example. Take $T = \mathbb{N} = \{1, 2, 3, \dots\}$, $\Omega_t = \mathbb{R}$, $\mathcal{F}_t = \mathcal{B}(\mathbb{R})$ for all $t \in T$. Then $\Omega = \mathbb{R}^{\mathbb{N}}$. For each $n \geq 1$, let P_n be a probability on the Borel subsets of $\mathbb{R}^n$ such that P_{n+1} restricted to $\mathbb{R}^n$ coincides with P_n (i.e. $P_{n+1}(A \times \mathbb{R}) = P_n(A)$ for all Borel subsets A of $\mathbb{R}^n$). Then there exists a unique probability P on the product σ-algebra $\mathcal{F}$ which extends each of the P_n's. Stated another way, there exists a probability space $(\Omega, \mathcal{F}, P)$ and on it a stochastic process $\{X_n, n \geq 1\}$ (the coordinate mappings) having the P_n's as its finite-dimensional distributions (i.e. $P\{(X_1, \dots, X_n) \in A\} = P_n(A)$ for all Borel subsets A of $\mathbb{R}^n$ and all $n \geq 1$.

In particular, if $F_n, n \geq 1$ are distribution functions on $\mathbb{R}$, and if for every $n \geq 1$, P_n is product probability $dF_1 \times \cdots \times dF_n$, there exists a probability space $(\Omega, \mathcal{F}, P)$ and on it a sequence $\{X_n, n \geq 1\}$ (the coordinate mappings) of independent random variables such that $X_n \sim F_n$ for every $n \geq 1$.

Example. Let T be any set, $\Omega_t = \mathbb{R}$, $\mathcal{F}_t = \mathcal{B}(\mathbb{R})$ for all $t \in T$. Then $\Omega = \mathbb{R}^T$. Let $\Gamma(s,t) : T \times T \to [0, +\infty)$ be symmetric ($\Gamma(s,t) = \Gamma(t,s)$ and non-negative definite ($\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \Gamma(T_i, t_j) \geq 0$ for all finite subsets $\{t_1, \dots, t_n\}$ of T). For each finite subset $\{t_1, \dots, t_n\}$ of T, let $P_{\{t_1, \dots, t_n\}}$ be the distribution of a Gaussian random vector $(Y_{t_1}, \dots, Y_{t_n})$ with mean vector zero and covariance matrix $[\Gamma(t_i, t_j)]_{i,j=1,\dots,n}$. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and on it a stochastic process $\{X_t, tT\}$ (the coordinate mappings) having the $P_{\{t_1, \dots, t_n\}}$'s as its finite-dimensional distributions. $\{X_t, t \in T\}$ is called a Gaussian family.

If $T = [0, +\infty)$, $\{X_t, t \geq 0\}$ is called a Gaussian process. In particular, if $\Gamma(s,t) = \min\{s,t\}$, then $\{X_t, t \geq 0\}$ is called a Brownian motion.

Remark. Let $\mathcal{C}_c$ be the family of all cylinder sets $C = A \times \Pi_{t \in T \setminus S} \Omega_t$, where $A \in \otimes_{t \in S} \mathcal{F}_t$ and S is a countable subset of T . Then $\mathcal{C}_c$ is a σ-algebra of subsets of $\Pi_{t \in T} \Omega_t$ and it is easy to show that $\mathcal{C}_c$ coincides with $\otimes_{t \in T} \mathcal{F}_t$. This means that every member of $\mathcal{F}$ depends on only countably many values from T.

Theorem 6.0.7 *Suppose that $T = [0, +\infty)$, $\Omega_t = \mathbb{R}$, $\mathcal{F}_t = \mathcal{B}(\mathbb{R})$ for all $t \in T$. Let $\{X_t, t \geq 0\}$ be the process constructed on $(\Omega, \mathcal{F}, P)$ in Kolmogorov's Extension Theorem. Suppose there exist constants $a, b, c > 0$ such that*

$$E|X_t - X_s|^a \leq c|t-s|^{b+1} \quad \forall s, t \geq 0. \tag{0.1}$$

Let $\Omega_c = C[0, +\infty)$ =all continuous functions $[0, +\infty) \to \mathbb{R}$, let $\mathcal{C}_f$ denote the family of finite dimensional cylinder subsets of Ω_c, and let $\mathcal{F}_c$ be the σ-algebra of subsets of Ω_c generated by $\mathcal{C}_f$. Then there exists a probability P_c on $\mathcal{F}_c$ such that

$$P_c(B) = P(A) \quad \forall B \in \mathcal{F}_c \text{ such that } B = A \cap \Omega_c \text{ for some } A \in \mathcal{F}.$$

(Ref: Hida, T. and M. Hitsuda.(1991). Gaussian Processes*. AMS.*

Note that if $A = \Omega$, then $B = \Omega_c$, and we see that $P_c(\Omega_c) = 1$. Moreover, the coordinate process $\{X_t, t \geq 0\}$ restricted to Ω_c has the same finite dimensional distributions under P_c as it had under P. Hence we have constructed a process $\{X_t, t \geq 0\}$ on a probability space $(\Omega_c, \mathcal{F}_c, P_c)$ which has the required finite dimensional distributions, and which has continuous sample paths.

For a Brownian motion process, the condition in (0.1) is satisfied with $a = 4, b = 1, c = 3$.