Probability

Vladimir Pozdnyakov

November 21, 2024

Contents

Chapter 1

Basic Set Theory

1.1 Terminology and Notation

- *Sample space* Ω: an arbitrary set representing a list of possible outcomes $\omega \in \Omega$ of a random experiment.
- *Events* $A, B, C, ...$: any subsets ($A \subset \Omega$) of the sample space Ω.
- *Impossible event* $\emptyset$: the empty set.

We say that A is a subset of B ($A \subset B$), iff $\omega \in A$ implies $\omega \in B$, and we say that $A = B$, iff $A \subset B$ and $B \subset A$.

Set operations.

1. *Complementation*:

$$A^c = \{\omega : \omega \notin A\}.$$

2. *Intersection over arbitrary index set* T:

$$\bigcap_{t \in T} A_t = \{\omega : w \in A_t \text{ for all } t \in T\},$$

also, in case of a small number of events, we will use:

$$A \cap B \text{ or } AB.$$

3. *Union over arbitrary index set* T:

$$\bigcup_{t \in T} A_t = \{\omega : w \in A_t \text{ for some } t \in T\},$$

also, in case of a small number of events, we will use:

$$A \cup B.$$

4. *Set difference*:

$$A \setminus B = AB^c.$$

5. *Symmetric difference*:

$$A \bigtriangleup B = (A \setminus B) \cup (B \setminus A).$$

Events A and B are *mutually disjoint or mutually exclusive* if $A \cap B = \emptyset$. In case of mutually disjoint events $A + B$ can be used for $A \cup B$.

Properties of set operations:

1. *Complementation*:

$$(A^c)^c = A, \quad \emptyset^c = \Omega, \quad \Omega^c = \emptyset.$$

2. *Commutativity*:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

3. *Associativity*:

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C).$$

4. *De Morgan's laws*:

$$\left(\bigcup_{t \in T} A_t \right)^c = \bigcap_{t \in T} A_t^c, \quad \left(\bigcap_{t \in T} A_t \right)^c = \bigcup_{t \in T} A_t^c.$$

5. *Distributivity*:

$$B \cap \left(\bigcup_{t \in T} A_t \right) = \bigcup_{t \in T} BA_t, \quad B \cup \left(\bigcap_{t \in T} A_t \right) = \bigcap_{t \in T} (B \cup A_t).$$

We define the *indicator function of A* as

$$1_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \in A^c \end{cases}$$

Note that

$$1_A \leq 1_B \text{ iff } A \subset B$$

and

$$1_{A^c} = 1 - 1_A.$$

1.2 Limits of Sets

We define

$$\inf_{k \geq n} A_k = \bigcap_{k=n}^{\infty} A_k,$$

$$\sup_{k \geq n} A_k = \bigcup_{k=n}^{\infty} A_k,$$

$$\liminf_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega : \omega \in A_n \text{ for all } n \geq n_0(\omega)\},$$

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega : \omega \in A_n \text{ infinitely often}\}.$$

If for some sequence of evens $\{B_n\}$

$$\limsup_{n \to \infty} B_n = \liminf_{n \to \infty} B_n = B,$$

then B is called *limit of* B_n, and we write

$$\lim_{n\to\infty} B_n = B \text{ or } B_n \to B.$$

Properties:

1. $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$
2. $(\liminf_{n\to\infty} A_n)^c = \limsup_{n\to\infty} A_n^c$

We say that a sequence of events $\{A_n\}$ is *monotone non-decreasing* ($A_n \uparrow$) if $A_1 \subset A_2 \subset ...$, and it is *monotone non-increasing* ($A_n \downarrow$) if $A_1 \supset A_2 \supset ...$.

Proposition 1.1 *For a monotone sequence of sets, the limit always exists:*

(1) if $A_n \uparrow$*, then* $\lim_{n\to\infty} A_n = \bigcup_{n=1}^\infty A_n$,

(2) if $A_n \downarrow$*, then* $\lim_{n\to\infty} A_n = \bigcap_{n=1}^\infty A_n$.

Consequently,

$$\liminf_{n\to\infty} B_n = \lim_{n\to\infty}\left(\inf_{k\geq n} B_k\right), \quad \limsup_{n\to\infty} B_n = \lim_{n\to\infty}\left(\sup_{k\geq n} B_k\right).$$

Proof. Let us prove (1). We only need to show that $\limsup_n A_n \subset \liminf_n A_n$. First note that monotonicity gives us

$$\liminf_n A_n = \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k = \bigcup_{n=1}^\infty A_n.$$

But also we have

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \subset \bigcup_{n=1}^{\infty} A_n = \liminf_n A_n$$

which finishes proof. $\square$

1.3 Fields

Definition 1.1 *A non-empty class of subsets, $\mathcal{A}$, of Ω is called* field *if*

(1) $\Omega \in \mathcal{A}$*,*

(2) $A \in \mathcal{A}$ *implies* $A^c \in \mathcal{A}$*,*

(3) $A, B \in \mathcal{A}$ *implies* $A \cup B \in \mathcal{A}$*.*

One can show that field is a collection of subsets which is closed under *finite* union, *finite* intersection (note $AB = (A^c \cup B^c)^c$) and complements.

Definition 1.2 *A non-empty class of subsets, $\mathcal{B}$, of Ω is called* σ-field *if*

(1) $\Omega \in \mathcal{B}$*,*

(2) $B \in \mathcal{B}$ *implies* $B^c \in \mathcal{B}$*,*

(3) $B_i \in \mathcal{A}, i \geq 1$ *implies* $\bigcup_{i\geq 1} B_i \in \mathcal{B}$*.*

A σ-field is a collection of subsets which is closed under *countable* union, *countable* intersection and complements.

Examples.

1. The power set. The set of all subsets of Ω, 2^Ω is a σ-field.

2. The trivial σ-field. $\{\emptyset, \Omega\}$ is a σ-field.

3. The countable/co-countable σ-field $\mathcal{C}$. Let $\Omega = \mathbb{R}$, $A \in \mathcal{C}$ iff either A is countable or its complement is.

Exercise 1.1 Let $\Omega = \mathbb{N}$. $A \in \mathcal{C}$ iff either A is finite or its complement is. Is $\mathcal{C}$ a field? σ-field?

Exercise 1.2 *Let $\Omega = \mathbb{N}$. For any A, a subset of $\mathbb{N}$, we define $A_n = A \bigcap [1, n]$. Let a_n is the cardinality of A_n. Consider a collection of subsets of $\mathbb{N}$, $\mathcal{A}$, for which $\lim_{n\to\infty} a_n/n$ exists. Is $\mathcal{A}$ a field? σ-field?*

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be σ-fields, the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$ is $\{B \subset \Omega : B \in \mathcal{B}_1 \text{ and } \mathcal{B}_2\}$.

Proposition 1.2 *An intersection (finite or over an index set) of σ-fields is a σ-field.*

Proposition 1.3 *Let $\mathcal{C}$ be a collection of subsets of Ω. Then there is a smallest σ-field $\sigma(\mathcal{C})$ containing all the sets that are in $\mathcal{C}$.*

Proof. First note that there is at least one σ-field that contains $\mathcal{C}$. Let us define $\sigma(\mathcal{C})$ as a collection of all sets that belong to every σ-field containing $\mathcal{C}$. It is easy to check that $\sigma(\mathcal{C})$ is a σ-field, and it is the smallest. $\square$

1.4 Monotonic Class

Definition 1.3 *A collection $\mathcal{M}$ of subsets of Ω is a* monotonic class *if $A_n \in \mathcal{M}, n = 1, 2...$ and $A_n \uparrow A$ or $A_n \downarrow A$ implies that $A \in \mathcal{M}$.*

Let $\mathcal{C}$ be a collection of subsets of Ω, then by $\mu(\mathcal{C})$ we denote the smallest monotonic class containing all the sets that are in $\mathcal{C}$ (which exists, prove it).

Proposition 1.4 *Let $\mathcal{A}$ be a field of subsets of Ω. The following two conditions are equivalent:*

(1) $\mathcal{A}$ is a σ-field,

(2) $\mathcal{A}$ is a monotonic class.

Proof. (1) $\Rightarrow$ (2) Any σ-field is obviously a monotonic class (recall $A_n \uparrow A = \cup_n A_n$).

(2) $\Rightarrow$ (1) Let $A_n \in \mathcal{A}$. Consider $B_n = \bigcup_{i=1}^{n} A_i$. Since $\mathcal{A}$ is a field, $B_n \in \mathcal{A}$. But also $B_n \subset B_{n+1}$, and $B_n \uparrow \bigcup_{i=1}^{\infty} A_i$, therefore, by definition of a monotonic class $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ $\square$

The following theorem shows that by taking monotonic limits we can turn a field into a σ-field.

Theorem 1.1 *Let $\mathcal{A}$ be a field of subsets of Ω. Then*

$$\mu(\mathcal{A}) = \sigma(\mathcal{A})$$

Proof. By Proposition 1.4 $\sigma(\mathcal{A})$ is a monotonic class, therefore, $\mu(\mathcal{A}) \subset \sigma(\mathcal{A})$. Thus, it would be enough to establish that $\mu(\mathcal{A})$ is a field (and, therefore, a σ-field, by Proposition 1.4 again).

First, we show that $A \in \mu(\mathcal{A})$ implies $A^c \in \mu(\mathcal{A})$. Consider

$$\mathcal{M} = \{B : B \in \mu(\mathcal{A}), B^c \in \mu(\mathcal{A})\}.$$

Note that $\mathcal{A} \subset \mathcal{M} \subset \mu(\mathcal{A})$. It is easy to prove that $\mathcal{M}$ is a monotonic class.[1] Since $\mu(\mathcal{A})$ is the smallest monotonic class, $\mathcal{M} = \mu(\mathcal{A})$, and, as consequence, we have $A \in \mu(\mathcal{A}) \Rightarrow A \in \mathcal{M} \Rightarrow A^c \in \mu(\mathcal{A})$

Next, we prove that $\mu(\mathcal{A})$ is closed under taking finite unions. Consider

$$\mathcal{M}_1 = \{A : A \cup B \in \mu(\mathcal{A}) \text{ for all } B \in \mathcal{A}.\}$$

Then $\mathcal{M}_1$ is a monotonic class and $\mathcal{A} \subset \mathcal{M}_1$, therefore, $\mu(\mathcal{A}) \subset \mathcal{M}_1$. Let

$$\mathcal{M}_2 = \{B : A \cup B \in \mu(\mathcal{A}) \text{ for all } A \in \mu(\mathcal{A})\}.$$

Then $\mathcal{M}_2$ is also a monotonic class. Now, if $B \in \mathcal{A}$ and $A \in \mu(\mathcal{A}) \subset \mathcal{M}_1$, then, by definition of $\mathcal{M}_1$ we get $A \cup B \in \mu(\mathcal{A})$, that is, $B \in \mathcal{M}_2$. Thus, $\mathcal{A} \subset \mathcal{M}_2$, and by minimality of $\mu(\mathcal{A})$ we get $\mu(\mathcal{A}) \subset \mathcal{M}_2$. Finally, if $B \in \mu(\mathcal{A}) \subset \mathcal{M}_2$ and $A \in \mu(\mathcal{A})$, by definition of $\mathcal{M}_2$ we obtain $A \cup B \in \mu(\mathcal{A})$. $\square$

Exercise 1.3 Prove that $\mathcal{M}$, $\mathcal{M}_1$, and $\mathcal{M}_2$ are monotonic classes.

1.5 Dynkin's theorem

Definition 1.4 *A collection $\mathcal{P}$ of subset of Ω is a π*-system *if it is closed under finite intersection.*

Definition 1.5 *A non-empty class of subsets, $\mathcal{L}$, of Ω is called λ*-system *if*

[1] For example, we need to show that if $B_n \in \mathcal{M}$ and $B_n \uparrow B$ then $B \in \mathcal{M}$. By definition of $\mathcal{M}$, $B_n \in \mathcal{M}$ implies that $B_n \in \mu(\mathcal{A})$ and $B_n^c \in \mu(\mathcal{A})$. Since $\mu(\mathcal{A})$ is a monotonic class, we obtain that $B = \lim \uparrow B_n \in \mu(\mathcal{A})$ and $B^c = \lim \downarrow B_n^c \in \mu(\mathcal{A})$, i.e. B, indeed, belongs to $\mathcal{M}$.

(1) $\Omega \in \mathcal{L}$,

(2) $A, B \in \mathcal{L}$, $A \subset B$ *implies* $B \setminus A \in \mathcal{L}$,

(3) $B_i \in \mathcal{L}, i \geq 1$, $B_i \subset B_{i+1}$ *implies* $\bigcup_{i\geq 1} B_i \in \mathcal{L}$.

Definition 1.6 *A non-empty class of subsets,* $\mathcal{L}$*, of* Ω *is called* λ'-system *if*

(1) $\Omega \in \mathcal{L}$,

(2) $B \in \mathcal{L}$ *implies* $B^c \in \mathcal{L}$,

(3) if $B_i \in \mathcal{L}, i \geq 1$ *and they are disjoint, then* $\sum_{i\geq 1} B_i \in \mathcal{L}$.

Exercise 1.4 Prove that postulates of λ-system and λ'-system are equivalent.

Let $\mathcal{C}$ be a collection of subsets of Ω, then by $\lambda(\mathcal{C})$ we denote the smallest λ-system containing all the sets that are in $\mathcal{C}$.

Theorem 1.2 (Dynkin's Theorem) *Let* $\mathcal{P}$ *be a* π*-system of subsets of* Ω*. Then*

$$\lambda(\mathcal{P}) = \sigma(\mathcal{P})$$

Proof. Every σ-field is a λ-system, therefore, $\lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$. Now, if we prove that $\lambda(\mathcal{P})$ is closed under finite intersection, then $\lambda(\mathcal{P})$ is a σ-field, and, as a result, $\sigma(\mathcal{P}) \subset \lambda(\mathcal{P})$.

Let us define

$$\mathcal{L}_1 = \{B \in \lambda(\mathcal{P}) : B \cap A \in \lambda(\mathcal{P}) \text{ for all } A \in \mathcal{P}\}.$$

It is easy to see that $\mathcal{P} \subset \mathcal{L}_1$, $\mathcal{L}_1$ is a λ-system. Therefore, by minimality of $\lambda(\mathcal{P})$ we have $\lambda(\mathcal{P}) \subset \mathcal{L}_1$. By definition of $\mathcal{L}_1$ we also have $\mathcal{L}_1 \subset \lambda(\mathcal{P})$. That is,

$\mathcal{L}_1 = \lambda(\mathcal{P})$.

Now let

$$\mathcal{L}_2 = \{A \in \lambda(\mathcal{P}) : B \cap A \in \lambda(\mathcal{P}) \text{ for all } B \in \lambda(\mathcal{P})\}.$$

Again $\mathcal{L}_2$ is a λ-system.

Now, if $B \in \mathcal{P}$ and $A \in \lambda(\mathcal{P}) = \mathcal{L}_1$, then, by definition of $\mathcal{L}_1$ we get $A \cap B \in \lambda(\mathcal{P})$, that is, $B \in \mathcal{L}_2$. Thus, $\mathcal{P} \subset \mathcal{L}_2$, and by minimality of $\lambda(\mathcal{P})$ and definition of $\mathcal{L}_2$ we get $\lambda(\mathcal{P}) = \mathcal{L}_2$. That is, whenever $A \in \lambda(\mathcal{P}) = \mathcal{L}_2$ and $B \in \lambda(\mathcal{P})$, by definition of $\mathcal{L}_2$ we obtain $A \cap B \in \lambda(\mathcal{P})$. $\square$

Exercise 1.5 Show that $\mathcal{L}_1$ and $\mathcal{L}_2$ are λ-systems.

Exercise 1.6 Prove that if a λ-system is closed under intersection, then it is a σ-field.

Exercise 1.7 Let $\mathcal{P}$ be a π-system, and $\mathcal{L}$ be a λ-system. Prove that if $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$

Exercise 1.8 Give an example of λ-system which is not a σ-field.

1.6 Borel σ-fields

Consider the real line $\mathbb{R}$ and let $\mathcal{I}$ be the collection of intervals of form:

$$(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$$

for all a and b, $-\infty \le a \le b \le \infty$, with a convention that $(a, \infty] = (a, \infty)$. Let $\mathcal{A}$ be the system of *finite sums of disjoint intervals* of the form $(a, b]$, i.e.,

$$A \in \mathcal{A} \text{ iff} A = \sum_{i=1}^{n} (a_i, b_i].$$

Exercise 1.9 Prove that $\mathcal{A}$ is a field, but not a σ-field.

Definition 1.7 *The* Borel σ-field on the real line*,* $\mathcal{B}(\mathbb{R})$*, is the smallest σ-field that contains $\mathcal{A}$, and its sets are called* Borel sets.

Exercise 1.10 Observe that $\mathcal{I}$ is a π-system. Show that $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$.

Note that

$$(a, b) = \bigcup_{n=1}^{\infty} (a, b - 1/n],$$

$$[a, b] = \bigcap_{n=1}^{\infty} (a - 1/n, b],$$

$$\{a\} = \bigcap_{n=1}^{\infty} (a - 1/n, a].$$

Thus the Borel σ-field contains singletons $\{a\}$ and all sets of these six forms

$$(a, b), [a, b], [a, b), (-\infty, b), (-\infty, b], (a, \infty).$$

One can show that that construction of the Borel σ-filed can be based on any of these six types of intervals.

Exercise 1.11 Let $\mathcal{I}_1$ is the collection of intervals of form $[a, b]$. Show that $\sigma(\mathcal{I}_1) = \mathcal{B}(\mathbb{R})$.

Proposition 1.5 *Let $\mathcal{C}$ be a collection of subsets of Ω, let $B \subset \Omega$, and define collection of subsets of B:*

$$\mathcal{C} \cap B = \{A \cap B : A \in \mathcal{C}\}.$$

Then

$$\sigma(\mathcal{C} \cap B) = \sigma(\mathcal{C}) \cap B,$$

as σ-fields on B.

Proof. Obviously $\mathcal{C} \cap B \subset \sigma(\mathcal{C}) \cap B$. Also it is clear that $\sigma(\mathcal{C}) \cap B$ is a σ-field (on B), therefore, by minimality of $\sigma(\mathcal{C} \cap B)$ we get that

$$\sigma(\mathcal{C} \cap B) \subset \sigma(\mathcal{C}) \cap B.$$

Now let us define

$$\mathcal{C}_B = \{A \in \sigma(\mathcal{C}) : A \cap B \in \sigma(\mathcal{C} \cap B)\}.$$

Note that $\mathcal{C}_B$ is a σ-field (on Ω), and

$$\mathcal{C} \subset \mathcal{C}_B \subset \sigma(\mathcal{C}).$$

Therefore,

$$\sigma(\mathcal{C}) \subset \sigma(\mathcal{C}_B) = \mathcal{C}_B \subset \sigma(\mathcal{C}),$$

and $\sigma(\mathcal{C}) = \mathcal{C}_B$. Thus, if $A \in \sigma(\mathcal{C}) = \mathcal{C}_B$, then by definition of $\mathcal{C}_B$ we get $A \cap B \in \sigma(\mathcal{C} \cap B)$ which means that $\sigma(\mathcal{C}) \cap B \subset \sigma(\mathcal{C} \cap B)$. $\square$

Exercise 1.12 If collections $\mathcal{C}_1$ and $\mathcal{C}_2$ are such that $\mathcal{C}_1 \subset \mathcal{C}_2$, then $\sigma(\mathcal{C}_1) \subset \sigma(\mathcal{C}_2)$.

Definition 1.8 *Let $\mathcal{I}$ is a collection of all closed intervals of $[a, b]$. The* Borel σ-field on $[a, b]$, *$\mathcal{B}([a, b])$, is is the smallest σ-field that contains $\mathcal{I}$, and, by Proposition 1.5*

$$\mathcal{B}([a, b]) = \mathcal{B}(\mathbb{R}) \cap [a, b].$$

Chapter 2

Probability Space

2.1 Definition and Basic Properties

Definition 2.1 *A* probability space *is a triple* $(\Omega, \mathcal{F}, \mathbf{P})$*, where*

(a) Ω *is a set of points* ω*;*

(b) $\mathcal{F}$ *is a* σ*-field of subsets of* Ω*;*

(c) $\mathbf{P}$ *is a* σ*-additive* probability *measure, i.e.*

1. $\mathbf{P} : \mathcal{F} \mapsto [0, 1]$,
2. $\mathbf{P}(\Omega) = 1$,
3. *If* $\{A_n\}_{n\geq 1}$ *are disjoint events from* $\mathcal{F}$*, then*

$$\mathbf{P}(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} \mathbf{P}(A_n).$$

Properties of probability measures.

1. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.

   Proof. $1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c)$. $\square$

2. $\mathbf{P}(\emptyset) = 0$.

   Proof. $\mathbf{P}(\emptyset) = \mathbf{P}(\Omega^c) = 1 - \mathbf{P}(\Omega) = 1 - 1$. $\square$

3. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$.

   Proof. Note first that

$$\mathbf{P}(A) = \mathbf{P}(AB^c) + \mathbf{P}(AB),$$

$$\mathbf{P}(B) = \mathbf{P}(BA^c) + \mathbf{P}(AB).$$

   Thus

$$\begin{aligned}
\mathbf{P}(A \cup B) &= \mathbf{P}(AB^c \cup BA^c \cup AB) \\
&= \mathbf{P}(AB^c) + \mathbf{P}(BA^c) + \mathbf{P}(AB) \\
&= \mathbf{P}(A) - \mathbf{P}(AB) + \mathbf{P}(B) - \mathbf{P}(AB) + \mathbf{P}(AB) \\
&= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB).
\end{aligned}$$

   $\square$

4. *Inclusion-Exclusion Formula*:

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i) = \sum_i \mathbf{P}(A_i) - \sum_{i<j} \mathbf{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbf{P}(A_i \cap A_j \cap A_k)$$

$$+\cdots+(-1)^{n+1}\mathbf{P}(A_1\cap\cdots\cap A_n).$$

Proof. By induction. The induction base, case of $n=2$, has already been proved. Transition from $n-1$ to n is based on the following observation:

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i)=\mathbf{P}(\bigcup_{i=1}^{n-1} A_i\cup A_n)=\mathbf{P}(\bigcup_{i=1}^{n-1} A_i)+\mathbf{P}(A_n)-\mathbf{P}(\bigcup_{i=1}^{n-1} A_iA_n).$$

After applying the inclusion-exclusion formula (twice) for $n-1$ we obtain the needed result. $\square$

5. *Monotonicity Property*:

$$\mathbf{P}(A_1\cup A_2\cdots)\le\mathbf{P}(A_1)+\mathbf{P}(A_2)+\cdots.$$

Proof. Let $B_1=A_1$, $B_n=A_1^c\cdots A_{n-1}^cA_n$, $n\ge 2$. It is easy to see that $B_i\cap B_j=\emptyset$ and $\bigcup_{n\ge1}A_n=\sum_{n\ge1}B_n$. Therefore,

$$\begin{aligned}\mathbf{P}(A_1\cup A_2\cdots) &= \mathbf{P}(B_1+B_2+\cdots)\\ &= \mathbf{P}(B_1)+\mathbf{P}(B_2)+\cdots\\ &\le \mathbf{P}(A_1)+\mathbf{P}(A_2)+\cdots.\end{aligned}$$

$\square$

Theorem 2.1 *Let* $\mathbf{P}$ *be a* finitely *additive measure of sets from* σ*-field* $\mathcal{F}$[1] *with* $\mathbf{P}(\Omega)=1$*. The following four conditions are equivalent:*

[1]i.e., for every disjoint sets A and B in $\mathcal{F}$ we have $\mathbf{P}(A+B)=\mathbf{P}(A)+\mathbf{P}(B)$.

(1) $\mathbf{P}$ *is* σ*-additive (i.e.,* $\mathbf{P}$ *is a probability measure),*

(2) $\mathbf{P}$ *is continuous from below, i.e., if* $A_n \uparrow A$*,* $A_n \in \mathcal{F}$*, then* $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$*,*

(3) $\mathbf{P}$ *is continuous from above, i.e., if* $A_n \downarrow A$*,* $A_n \in \mathcal{F}$*, then* $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$*,*

(4) $\mathbf{P}$ *is continuous at* $\emptyset$*, i.e., if* $A_n \downarrow \emptyset$*,* $A_n \in \mathcal{F}$*, then* $\mathbf{P}(A_n) \downarrow 0$*.*

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

Let $A_n \uparrow A$, $A_n \in \mathcal{F}$.

$$\begin{aligned}
\mathbf{P}(\bigcup_{n\geq 1} A_n) &= \mathbf{P}(A_1 + A_2 \setminus A_1 + A_3 \setminus A_2 + \cdots) \\
&= \mathbf{P}(A_1) + \mathbf{P}(A_2 \setminus A_1) + \mathbf{P}(A_3 \setminus A_2) + \cdots \\
&= \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1) + \mathbf{P}(A_3) - \mathbf{P}(A_2) + \cdots \\
&= \lim_n \mathbf{P}(A_n)
\end{aligned}$$

(2) $\Rightarrow$ (3)

Let $A_n \downarrow A$, $A_n \in \mathcal{F}$. Consider sequence $\{A_n^c\}_{n\geq 1}$. It is nondecreasing, therefore, by (2)

$$\lim_n \mathbf{P}(A_n^c) = \mathbf{P}(\bigcup_{n\geq 1} A_n^c).$$

Now,

$$\begin{aligned}
\lim_n \mathbf{P}(A_n) &= \lim_n (1 - \mathbf{P}(A_n^c)) = 1 - \lim_n \mathbf{P}(A_n^c) \\
&= 1 - \mathbf{P}(\bigcup_{n\geq 1} A_n^c) = 1 - \mathbf{P}([\bigcap_{n\geq 1} A_n]^c) \\
&= 1 - 1 + \mathbf{P}(\bigcap_{n\geq 1} A_n) = \mathbf{P}(\bigcap_{n\geq 1} A_n)
\end{aligned}$$

(3) $\Rightarrow$ (4)

Obvious.

(4) $\Rightarrow$ (1)

Let $\{A_i\}_{i\geq 1}$ be disjoint events from $\mathcal{F}$.

$$\begin{aligned}
\sum_{i=1}^{\infty}\mathbf{P}(A_i) &= \lim_n \sum_{i=1}^{n}\mathbf{P}(A_i) = \lim_n \mathbf{P}(\sum_{i=1}^{n} A_i) \\
&= \lim_n [\mathbf{P}(\sum_{i=1}^{\infty} A_i) - \mathbf{P}(\sum_{i=n+1}^{\infty} A_i)] \\
&= \mathbf{P}(\sum_{i=1}^{\infty} A_i) - \lim_n \mathbf{P}(\sum_{i=n+1}^{\infty} A_i) = \mathbf{P}(\sum_{i=1}^{\infty} A_i),
\end{aligned}$$

because $\sum_{i=n+1}^{\infty} A_i \downarrow \emptyset$ (why?), and by (4) $\lim_n \mathbf{P}\left(\sum_{i=n+1}^{\infty} A_i\right) = 0$. $\square$

Exercise 2.1 Consider a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Suppose that $A = \lim A_n$ exists, where $A_n \in \mathcal{F}$. Show that

$$\lim_n \mathbf{P}(A_n) = \mathbf{P}(A).$$

Exercise 2.2 Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Consider set function on $\mathcal{F} \times \mathcal{F}$:

$$\rho(A, B) = \mathbf{P}(A \triangle B).$$

Show that $\rho(\cdot, \cdot)$ satisfy the triangle inequality, i.e., for any $A, B, C \in \mathcal{F}$

$$\rho(A, C) \leq \rho(A, B) + \rho(B, C).$$

Exercise 2.3 Let μ be a *finitely* additive finite measure on a field $\mathcal{A}$, let $A_i, i \geq$

1 be disjoint sets from $\mathcal{A}$ such that $A = \sum_{i\geq 1} A_i$ also belongs $\mathcal{A}$. Which one,

$$\text{(a) } \mu(A) \geq \sum_{i\geq 1} \mu(A_i) \quad \text{or} \quad \text{(b) } \mu(A) \leq \sum_{i\geq 1} \mu(A_i),$$

is true? Prove that one, and give a counterexample for another.

2.2 Cumulative Distribution Function

Consider the probability space $\{\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P}\}$. Let

$$F(x) = \mathbf{P}((\infty, x]).$$

Proposition 2.1 *The function $F(x)$ has the following properties:*

(1) F is right continuous[2] *and has a limit on the left,*

(2) F is monotone non-decreasing,

(3) $F(-\infty) = \lim_{x\downarrow -\infty} F(x) = 0$, $F(\infty) = \lim_{x\uparrow\infty} F(x) = 1$.

Proof. The continuity property of $\mathbf{P}$ implies (1) and (3) (check it!). Monotonicity of $\mathbf{P}$ gives (2). □

Definition 2.2 *A function $F : \mathbb{R} \mapsto [0, 1]$ that satisfies (1)-(3) is called a* cumulative distribution function.

Three types of distribution functions:

- Discrete distribution function. If F is piecewise constant, and it changes its values at points $x_1, x_2, ...$ by jumps of size $p_1, p_2, ...$ ($p_k > 0$, $\sum_k p_k =$

[2] If $x_n \downarrow x$, then $F(x_n) \downarrow F(x)$

1), then F is called *discrete*. Examples: Discrete Uniform, Bernoulli, Binomial, Poisson, Geometric etc.

- Absolutely continuous distribution function. If F has the representation

$$F(x) = \int_{-\infty}^{x} f(y)dy,$$

for some nonnegative function, then F is called *absolutely continuous*, and function f is called the *density* of the distribution function F. Examples: Uniform on $[a, b]$, Normal, Exponential, χ^2, Cauchy etc.

- Singular distribution function. If all the points of increases of the continuous distribution function F[3] belongs to a set with zero Lebesgue measure (see later), then F is called *singular*. It is a very strange type of distribution one example of which can be found in Shiryaev (1995, p. 156).

One can show that any distribution function F can be represented as a *mixture*

$$F(x) = \alpha_d F_d(x) + \alpha_c F_c(x) + \alpha_s F_s(x),$$

where $\alpha_d, \alpha_c, \alpha_s \geq 0$, $\alpha_d + \alpha_c + \alpha_s = 1$, and cdfs F_d, F_c, and F_s are discrete, absolutely continuous and singular, respectively.

2.3 Set Induction: Dynkin's Theorem Again

Set induction is the most important application of Dynkin's Theorem.

[3] x is a point of increase of F if for any $\epsilon > 0$ we have $F(x + \epsilon) - F(x - \epsilon) > 0$

Proposition 2.2 *Let $\mathbf{P}_1$, $\mathbf{P}_2$ be two probability measures on $\{\Omega, \mathcal{F}\}$. The collection*

$$\mathcal{L} = \{A \in \mathcal{F} : \mathbf{P}_1(A) = \mathbf{P}_2(A)\}$$

is a λ system.

Proof. Let us use here λ'-system postulates.

First, note that $\Omega \in \mathcal{L}$, because $\mathbf{P}_1(\Omega) = \mathbf{P}_2(\Omega) = 1$.

Second, $A \in \mathcal{L}$, i.e., $\mathbf{P}_1(A) = \mathbf{P}_2(A)$ implies $A^c \in \mathcal{L}$, i.e., $\mathbf{P}_1(A^c) = \mathbf{P}_2(A^c)$, because

$$\mathbf{P}_1(A^c) = 1 - \mathbf{P}_1(A) = 1 - \mathbf{P}_2(A) = \mathbf{P}_2(A^c).$$

Finally, if $\{A_i\}$ is a sequence of mutually disjoint events from $\mathcal{L}$, then

$$\mathbf{P}_1(\bigcup_i A_i) = \sum_i \mathbf{P}_1(A_i) = \sum_i \mathbf{P}_2(A_i) = \mathbf{P}_2(\bigcup_i A_i).$$

That is, $\bigcup_i A_i \in \mathcal{L}$. □

Theorem 2.2 (Set Induction) *Let $\mathbf{P}_1$, $\mathbf{P}_2$ be two probability measures on $\{\Omega, \mathcal{F}\}$. Let $\mathcal{P}$ be a π-system such that*

$$A \in \mathcal{P} \Rightarrow \mathbf{P}_1(A) = \mathbf{P}_2(A),$$

then

$$B \in \sigma(\mathcal{P}) \Rightarrow \mathbf{P}_1(B) = \mathbf{P}_2(B).$$

Proof. Define

$$\mathcal{L} = \{A \in \mathcal{F} : \mathbf{P}_1(A) = \mathbf{P}_2(A)\}.$$

By Proposition 2.2 $\mathcal{L}$ is a λ-system. But $\mathcal{P} \subset \mathcal{L}$, therefore, by Dynkin's Theorem $\sigma(\mathcal{P}) \subset \mathcal{L}$. $\square$

Exercise 2.4 Give an example that shows that if $\mathcal{P}$ is not a π-system then Set Induction Theorem does not hold.

Proposition 2.3 *Let* $\Omega = \mathbb{R}$. *Let* $\mathbf{P}_1$, $\mathbf{P}_2$ *be two probability measures on* $\{\mathbb{R}, \mathcal{B}(\mathbb{R})\}$ *such that their distribution function are equal, i.e., for any* $x \in \mathbb{R}$ $F_1(x) = F_2(x)$. *Then* $\mathbf{P}_1 \equiv \mathbf{P}_2$ *on* $\mathcal{B}(\mathbb{R})$.

Proof. Let

$$\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}.$$

Obviously, $\mathcal{P}$ is a π-system. As we know $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$. Now, $F_1(x) = F_2(x)$ implies that $\mathbf{P}_1$ and $\mathbf{P}_2$ agree on $\mathcal{P}$, therefore they agree on $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$. $\square$

2.4 Construction of Probability Spaces: Discrete Models

Let $\Omega = \{\omega_1, \omega_2, ...\}$ is countable. For each ω_i we assign the number p_i, where

$$i \geq 1, \quad p_i \geq 0 \text{ and } \sum_{i \geq 1} p_i = 1.$$

Let $\mathcal{F}$ be the set of all subsets of Ω, 2^{Ω}.

For $A \in \mathcal{F}$, we define

$$\mathbf{P} = \sum_{\omega_i \in A} p_i.$$

The set function $\mathbf{P}$ is a probability measure.

Exercise 2.5 Show that $\mathbf{P}$ is a probability measure.

The introduced probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$ is called a *discrete probability space.*

Example 2.1 Flipping a loaded coin N times.

$$\Omega = \{0,1\}^N = \{\omega = (\omega_1, ..., \omega_N) : \omega_i = 0 \text{ or } 1\}.$$

Probability $\mathbf{P}$ is determined by

$$p_\omega = p^{\sum_i \omega_i} q^{N - \sum_i \omega_i},$$

where $p \geq 0, q \geq 0, p + q = 1$. Check that $\mathbf{P}$ is a probability measure.

Example 2.2 Coincidences. Suppose the integers $1, 2, ...n$ are randomly permuted.

$$\Omega = \{\omega = (x_1, ..., x_n) : x_i \in \{1, ..., n\}; i = 1, ..., n; x_i \neq x_j\}.$$

Probability $\mathbf{P}$ is defined by

$$\mathbf{P}(\omega) = 1/n!$$

What is the probability that there is an integer left unchanged by the permutation? Let A_i is event when i left by the permutation on the ith position. By the inclusion-exclusion formula we obtain

$$
\begin{aligned}
\mathbf{P}(\bigcup_{i=1}^{n} A_i) &= \sum_i \mathbf{P}(A_i) - \sum_{i<j} \mathbf{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbf{P}(A_i \cap A_j \cap A_k) \\
&\quad + \cdots + (-1)^{n+1} \mathbf{P}(A_1 \cap \cdots \cap A_n). \\
&= C_1^n \frac{(n-1)!}{n!} - C_2^n \frac{(n-2)!}{n!} + \cdots + (-1)^{n+1} \frac{1}{n!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \frac{1}{n!} \\
&\approx 1 - e^{-1} \approx .632
\end{aligned}
$$

Note that the convergence is fast.

Example 2.3 Birthday paradox. Suppose that there are n students in class. Let us suppose that each student's birthday is on one of 365 days and that all days are equally probable. What is the probability, P_n, there are at least two students in the class whose birthdays coincide?

$$\Omega = \{\omega : \omega = (a_1, ..., a_n); a_i = 1, ..., M\}.$$

Probability $\mathbf{P}$ is determined by

$$\mathbf{P}(\omega) = \frac{1}{M^n}.$$

Let

$$A = \{\omega : \omega = (a_1, ..., a_n)\}; a_i \neq a_j; i \neq j\},$$

i.e., the event in which there is no repetition. It is easy to see that $|A| = (M)_n = M(M-1)...(M-n+1)$, therefore

$$P_n = 1 - \mathbf{P}(A) = 1 - \frac{(M)_n}{M^n}.$$

n	4	16	22	23	40	64
P_n	.016	.284	.476	.507	.891	.997

Exercise 2.6 The Chevalier de Mere problem. What event has a better chance to occur: (1) rolling a 6 in four tosses of a single fair die or (2) rolling "double-6" in twenty-four tosses of two fair dice?

2.5 Construction of Probability Spaces: Uncountable Spaces

Unfortunately, not every problem can be solved within framework of discrete probability space. Even if it is about a flipping a fair coin, sometimes we have questions that involve infinite number of flips. For instance, what is the expected waiting time till the first occurrence of head? Theoretically, we can have 100 or 1000 tails before the first head, therefore, an appropriate sample space would be:

$$\Omega = \{H, T\}^{\mathbb{N}}$$

which is an uncountable set.

Exercise 2.7 Show that $\Omega = \{H, T\}^{\mathbb{N}}$ is uncountable.

Earlier we saw that for every probability measure $\mathbf{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ one can construct a distribution function associated with $\mathbf{P}$. We now show that converse is also true. The workhorse that we are going to use in this construction is Caratheodory's (extension) theorem. The proof of the theorem is omitted—Caratheodory's theorem is important but it will be used only once.

Theorem 2.3 (Caratheodory's Theorem) *Let Ω be a space, $\mathcal{A}$ is a field of its subsets, $\mathcal{B}$ is $\sigma(\mathcal{A})$. Let μ_0 be a finite (i.e, $\mu_0(\Omega) < \infty$)) σ-additive measure on $(\Omega, \mathcal{A})$. There exists a measure μ on $(\Omega, \mathcal{B})$ such that*

$$\mu(A) = \mu_0(A), \text{ for any } A \in \mathcal{A},$$

and this extension is unique.

Theorem 2.4 (Lebesgue-Stieltjes Integral) *Let $F(x)$ be a cumulative distribution function on the real line $\mathbb{R}$. There exists a unique probability measure $\mathbf{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that*

$$\mathbf{P}((a, b]) = F(b) - F(a)$$

for all $a, b, -\infty \leq a < b < \infty$.

Proof. Let $\mathcal{A}$ be the system of finite sums of disjoint intervals of the form $(a, b]$, i.e.,

$$A \in \mathcal{A} \text{ iff} A = \sum_{i=1}^{n} (a_i, b_i].$$

On this set we define a set function $\mathbf{P}_0$ by the following equation:

$$\mathbf{P}_0(A) = \sum_{k=1}^{n} [F(b_k) - F(a_k)], \quad A \in \mathcal{A}.$$

It is easy to check that $\mathbf{P}_0$ is a well-define finitely additive set function on $\mathcal{A}$ with $\mathbf{P}_0(\mathbb{R}) = 1$. If we show that $\mathbf{P}_0$ is also σ-additive then by Caratheodory's theorem there exists a unique probability measure $\mathbf{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that extends $\mathbf{P}_0$ on $\mathcal{B}(\mathbb{R})$.

So let us show that $\mathbf{P}_0$ is σ-additive on $\mathcal{A}$. By Theorem 2.1 one needs to show that $\mathbf{P}_0$ continuous at $\emptyset$, that is, if $A_n \downarrow \emptyset$, $A_n \in \mathcal{A}$, then $\mathbf{P}_0(A_n) \downarrow 0$.

First let us suppose that the sets A_n belong to a closed interval $[-N, N]$, $N < \infty$. For any interval $(a, b]$ and $a' \downarrow a$ by the right-continuity of F we get that

$$\mathbf{P}_0((a', b]) = F(b) - F(a') \to F(b) - F(a) = \mathbf{P}_0((a, b]).$$

Thus for any $\epsilon > 0$ we can find $B_n \in \mathcal{A}$ such that

$$\text{closure } [B_n] \subset A_n, \text{ and } \mathbf{P}_0(A_n) - \mathbf{P}_0(B_n) \leq \epsilon 2^{-n}.$$

Since $\bigcap_n A_n = \emptyset$ we have $\bigcap_n [B_n] = \emptyset$. But the sets $[B_n]$ are closed, and therefore by Heine-Borel theorem[4] there exists $n_0 = n_0(\epsilon)$ such that

$$\bigcap_{n=1}^{n_0} [B_n] = \emptyset.$$

Just note that $[-N, N]$ is a closed interval, and $\{[B_n]^c\}_{n\geq 1}$ is an open cover of $[-N, N]$, because

$$\bigcup_{n\geq 1} [B_n]^c = \left(\bigcap_{n\geq 1} [B_n]\right)^c = \mathbb{R}.$$

By Heine-Borel theorem there exists n_0 such that

$$[-N, N] \subset \bigcup_{n=1}^{n_0} [B_n]^c = \left(\bigcap_{n=1}^{n_0} [B_n]\right)^c.$$

Since $\bigcap_{n=1}^{n_0} [B_n] \subset [-N, N]$ we get $\bigcap_{n=1}^{n_0} B_n \subset \bigcap_{n=1}^{n_0} [B_n] = \emptyset$. Now since $A_{n_0} \subset A_{n_0-1} \subset \cdots \subset A_1$ we get

$$\begin{aligned}
\mathbf{P}_0(A_{n_0}) &= \mathbf{P}_0\left(A_{n_0} \setminus \bigcap_{k=1}^{n_0} B_k\right) = \mathbf{P}_0\left(\bigcup_{k=1}^{n_0} (A_{n_0} \setminus B_k)\right) \\
&\leq \mathbf{P}_0\left(\bigcup_{k=1}^{n_0} (A_k \setminus B_k)\right) \\
&\leq \sum_{k=1}^{n_0} \mathbf{P}_0(A_k \setminus B_k) \leq \sum_{k=1}^{n_0} \epsilon 2^{-k} \leq \epsilon
\end{aligned}$$

That is, for any $\epsilon > 0$ there exists n_0 such that for all $n > n_0$ $\mathbf{P}_0(A_n) \leq \mathbf{P}_0(A_{n_0}) \leq \epsilon$. Therefore, $\mathbf{P}_0(A_n) \downarrow 0$.

[4] Heine-Borel Theorem: Any cover of a closed interval $[a, b]$ by a system of open intervals (or, more generally, open sets) has a finite subcover. An exercise: give an example of open cover of $(0, 1)$ that does not allow a finite subcover.

Now, let us treat the general case. For any $\epsilon > 0$ one can find (why?) N such that $\mathbf{P}_0([-N, N]) > 1 - \epsilon$. Therefore,

$$\begin{aligned}
\mathbf{P}_0(A_n) &= \mathbf{P}_0(A_n \cap [-N, N]) + \mathbf{P}_0(A_n \cap [-N, N]^c) \\
&\leq \mathbf{P}_0(A_n \cap [-N, N]) + \epsilon
\end{aligned}$$

But the first term is also small for all sufficiently large n by the first part of the proof. □

Exercise 2.8 Consider a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{A}$ be a field such that $\sigma(\mathcal{A}) = \mathcal{F}$. Show that for any $B \in \mathcal{F}$ and any $\epsilon > 0$ one can find $A \in \mathcal{A}$ such that

$$P(B \triangle A) < \epsilon.$$

2.6 Lebesgue Measure on $[0, 1]$

If the distribution function is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

then the corresponding probability measure λ is called *Lebesgue measure on* $([0, 1], \mathcal{B}([0, 1]))$. Obviously, Lebegues measure formalizes the concept of length. By Theorem 2.4 we have a probability measure defined only on Borel σ-field

$\mathcal{B}([0,1])$. There is a standard procedure that allows us extend this measure to a wider σ-field $\bar{\mathcal{B}}([0,1])$. We say that $\Lambda \subset [0,1]$ belongs to $\bar{\mathcal{B}}([0,1])$ if there are two Borel sets A and B such that $A \subset \Lambda \subset B$ and $\lambda(B \setminus A) = 0$.

Exercise 2.9 Show that $\bar{\mathcal{B}}([0,1])$ is a σ-field.

The extension of λ to a set function $\bar{\lambda}$ that is defined on $\bar{\mathcal{B}}([0,1])$ is done as follows. If $\Lambda \in \bar{\mathcal{B}}([0,1])$, then there are Borel sets A and B such that $A \subset \Lambda \subset B$ and $\lambda(B \setminus A) = 0$. We define $\bar{\lambda}(\Lambda) = \lambda(A)$.

Exercise 2.10 Show that $\bar{\lambda}$ is well-defined (that is, if there are Borel sets A_i and B_i, $i = 1, 2$ such that $A_i \subset \Lambda \subset B_i$ and $\lambda(B_i \setminus A_i) = 0$, then $\lambda(A_1) = \lambda(A_2)$), and it is a probability measure on $([0,1], \bar{\mathcal{B}}([0,1])$.

Definition 2.3 $\bar{\mathcal{B}}([0,1])$ *is called* Lebesgue σ-field*, and measure* $\bar{\lambda}$ *is called* Lebesgue measure on $([0,1], \bar{\mathcal{B}}([0,1]))$.

Chapter 3

Random Variables

3.1 Measurability

Definition 3.1 Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. A function

$$X : \Omega \mapsto \mathbb{R}$$

is called a $\mathcal{F}$*-measurable* or *random variable* if for any Borel set on $\mathbb{R}$

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

Example 3.1 Indicator of an event. Let $A \in \mathcal{F}$, and

$$X(\omega) = I_A(\omega).$$

X is a random variable.

Example 3.2 Simple function. Let $A_i \in \mathcal{F}, i = 1, 2, ..., n$, and

$$X(\omega) = \sum_{i=1}^{n} x_i I_{A_i}(\omega).$$

X is a random variable (prove it!).

Theorem 3.1 *Let $\mathcal{I}$ be a system of sets on $\mathbb{R}$ such that $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$. Then a function $X : \Omega \mapsto \mathbb{R}$ is a random variable if*

$$\{\omega : X(\omega) \in B\} \in \mathcal{F}$$

for all $B \in \mathcal{I}$.

Proof. First note that taking the inverse image preserve the set operations of union, intersection and complement, i.e.,

$$\begin{aligned} X^{-1}(\bigcup_t B_t) &= \bigcup_t X^{-1}(B_t) \\ X^{-1}(\bigcap_t B_t) &= \bigcap_t X^{-1}(B_t) \\ X^{-1}(B_t^c) &= (X^{-1}(B_t))^c \end{aligned}$$

Let

$$\mathcal{D} = \{D \in \mathcal{B}(\mathbb{R}) : X^{-1}(D) \in \mathcal{F}\}.$$

Since $\mathcal{D}$ is a σ-field (why?) we have

$$\mathcal{I} \subset \mathcal{D} \subset \mathcal{B}(\mathbb{R})$$

and

$$\sigma(\mathcal{I}) \subset \sigma(\mathcal{D}) = \mathcal{D} \subset \mathcal{B}(\mathbb{R}).$$

But $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$, therefore $\mathcal{D} = \mathcal{B}(\mathbb{R})$. $\square$

That is, the measurability of a map of Ω to $\mathbb{R}$ can be checked on much smaller collection of events. In particular,

Corollary 3.1 *A function $X : \Omega \mapsto \mathbb{R}$ is a random variable if*

$$\{\omega : X(\omega) < x\} \in \mathcal{F}$$

for all $x \in \mathbb{R}$, or

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F}$$

for all $x \in \mathbb{R}$.

3.2 Approximation by Simple Random Variables

Proposition 3.1 *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a Borel function*[1]*, and X be a random variable. Then $f(X)$ is a random variable.*

[1] i.e., $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for any $B \in \mathcal{B}(\mathbb{R})$

Proof. Since for any $B \in \mathcal{B}(\mathbb{R})$, $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ we have

$$\{\omega : f(X(\omega)) \in B\} = \{\omega : X(\omega) \in f^{-1}(B)\} \in \mathcal{F}.$$

$\square$

That is, if X is a random variable, then so are X^n, $X^+ = \max(X, 0)$, $X^- = -\min(X, 0)$, and $|X|$ because x^n, x^+, x^-, and $|x|$ are Borel functions.

Exercise 3.1 Show that x^n, x^+, x^-, and $|x|$ are Borel functions.

Definition 3.2 Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. A function

$$X : \Omega \mapsto \bar{\mathbb{R}} = [-\infty, \infty]$$

is called an *extended random variable* if $X^{-1}(-\infty) \in \mathcal{F}$, $X^{-1}(\infty) \in \mathcal{F}$, and for any Borel set on $\mathbb{R}$

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

Theorem 3.2 *The following two statements are true.*

(1) If (extended) random variable $X \geq 0$, there is a sequence of simple random variables $X_1, X_2, \ldots$ such that $X_n(\omega) \uparrow X(\omega)$ for all $\omega \in \Omega$ as $n \to \infty$.

(2) For every (extended) random variable X, there exists a sequence of simple random variables $X_1, X_2, \ldots$ such that $|X_n| \leq |X|$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$ as $n \to \infty$.

Proof. Let us prove the first part. The second one follows from the first because any X can be presented in the form $X^+ - X^-$. For $n \geq 1$ let us define

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{(k-1)/2^n \leq X < k/2^n}(\omega) + nI_{X \geq n}(\omega).$$

It is easy to see that $X_n(\omega) \uparrow X(\omega)$. $\square$

Exercise 3.2 Show that $X_n(\omega) \uparrow X(\omega)$.

Exercise 3.3 Let $\Omega = [-1, 1]$, $\mathcal{F} = \mathcal{B}([-1, 1])$, $P = \lambda/2$. Consider random variable

$$X(\omega) = |1 - \omega|.$$

Provide explicit formulas for simple random variables X_1 and X_2 from Theorem 3.2.

3.3 Limits and Measurability

Theorem 3.3 *Let $X_1, X_2, \ldots$ be a sequence of extended random variables. Then*

(1) $\sup X_n$, $\inf X_n$, $\limsup X_n$, *and* $\liminf X_n$ *are also extended variables.*

(2) If $X(\omega) = \lim X_n(\omega)$ exists for every $\omega \in \Omega$, then X is an extended random variable.

Proof. (1) Just note that

$$\{\omega : \sup X_n > x\} = \bigcup_n \{\omega : X_n > x\} \in \mathcal{F},$$

and

$$\{\omega : \inf X_n < x\} = \bigcup_n \{\omega : X_n < x\} \in \mathcal{F}.$$

This gives us measurability of $\sup X_n$ and $\inf X_n$. The measurability of upper and lower limits, $\limsup X_n$ and $\liminf X_n$, follows from the following observation:

$$\limsup X_n = \inf_n \sup_{m \geq n} X_n, \quad \liminf X_n = \sup_n \inf_{m \geq n} X_n.$$

(2) We have

$$\begin{aligned}
\{\omega : X(\omega) < x\} &= \{\omega : \lim X_n < x\} \\
&= \{\omega : \limsup X_n = \liminf X_n\} \cap \{\omega : \limsup X_n < x\} \\
&= \Omega \cap \{\omega : \limsup X_n < x\} = \{\omega : \limsup X_n < x\} \in \mathcal{F}.
\end{aligned}$$

□

Corollary 3.2 *If X and Y are random variables, then $X + Y$, $X - Y$, XY, and X/Y are also random variables (if they are well defined).*

Proof. By Theorem 3.2 there exist sequences of simple random variables X_n and Y_n such that

$$\lim X_n = X \text{ and } \lim Y_n = Y.$$

Then

$$X_n \pm Y_n \quad \to \quad X \pm Y,$$

$$\begin{aligned} X_n Y_n &\to XY, \\ \frac{X_n}{Y_n + 1/n I_{Y_n=0}} &\to \frac{X}{Y}. \end{aligned}$$

Since sum, difference, product and quotient of two simple random variables are random variables, therefore, by Theorem 3.3 $X+Y$, $X-Y$, XY, and X/Y are random variables as limits of random variables. $\square$

3.4 Composition and Measurability

Definition 3.3 *Let X be a random variable. The following collection of sets from $\mathcal{F}$*

$$\mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$$

is called the σ-field generated by X.

Exercise 3.4 Show that the collection $\mathcal{F}_X$ is a σ-field.

Exercise 3.5 Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbf{P} = \lambda$. Consider random variable

$$X(\omega) = \begin{cases} \omega, & \text{if } 0 \le \omega \le 1/2, \\ 1, & \text{if } 1/2 < \omega \le 1. \end{cases}$$

Describe $\mathcal{F}_X$.

Let us recall (Proposition 3.1) that if f is a Borel function then $f(X)$ is a random variable. The converse is also true.

Theorem 3.4 *Let Y be a $\mathcal{F}_X$-measurable random variable. Then there is a Borel function f such that $Y = f \circ X$.*

Proof. Let us introduce two classes:

Φ_1 is the class of $\mathcal{F}_X$-measurable functions,

Φ_2 is the class of $\mathcal{F}_X$-measurable functions that can be represented in form $f \circ X$ for some Borel function f.

It is obvious that $\Phi_2 \subset \Phi_1$, therefore all we need to show that $\Phi_1 \subset \Phi_2$.

First, let $Y = I_A$, where $A \in \mathcal{F}_X$. Let us show that $Y \in \Phi_2$. Since $A \in \mathcal{F}_X$ there exists $B \in \mathcal{B}(\mathbb{R})$ such that $A = \{\omega : X(\omega) \in B\}$. Let

$$f(x) = I_B(x).$$

Then $Y = I_A = I_B(X) = f(X)$, i.e. $Y \in \Phi_2$.

Second, if Y is a simple function (i.e., it is a sum of indicators), then it also belongs to Φ_2.

Finally, let Y be an arbitrary $\mathcal{F}_X$-measurable function. By Theorem 3.2 there exists a sequence of simple measurable functions Y_n such that $Y_n \to Y$. As we have shown there are Borel functions f_n such that $Y_n = f_n(X)$, and $f_n(X(\omega)) \to Y(\omega)$. Let

$$f(x) = \begin{cases} \lim f_n(x), & \text{if } \lim_n f_n(x) \text{ exists,} \\ 0, & \text{otherwise.} \end{cases}$$

One can show that f is a Borel function (see exercises) and

$$Y(\omega) = \lim_n f_n(X(\omega)) = f(X(\omega)).$$

$\square$

Exercise 3.6 Show that $\{x : \lim_n f_n(x) \text{ exists}\}$ is a Borel set if f_n are Borel functions.

Exercise 3.7 If f is a Borel function, and B is a Borel set then

$$g(x) = \begin{cases} f(x), & \text{if } x \in B, \\ 0, & \text{otherwise,} \end{cases}$$

is a Borel function.

3.5 Random Elements of Metric Spaces

Definition 3.4 *By a* metric space *is meant a pair* (S, d) *consisting of a set* S *and a* metric (distance)*, i.e., non-negative real function defined for all* $x, y \in S$ *which satisfies the following three properties:*

(1) $d(x, y) = 0$ *iff* $x = y$*;*

(2) Symmetry: $d(x, y) = d(y, x)$;

(3) Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

Definition 3.5 *A subset* O *of metric space* (S, d) *is called* open *if for any* $x \in A$ *there exists an open ball with center* x *and radius* $r > 0$ $B_r(x) = \{y \in S :$

$d(x, y) < r\}$ *that also is a subset of* O.

Definition 3.6 *Let* $\mathcal{O}$ *be the class of open subsets of* (S, d)*. We define* Borel σ-field $\mathcal{B}(S, d)$ *to be the smallest* σ*-field generated by open sets, i.e.* $\mathcal{B}(S) = \sigma(\mathcal{O})$.

Definition 3.7 *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space. A function*

$$X : \Omega \mapsto S$$

is called random element *if for any Borel set* B *from* $\mathcal{B}(S, d)$

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

Definition 3.8 If X is a random variable, then the probability measure $\mathbf{P}_X$ defined on $(\mathbb{R}, \mathcal{B})$ by $\mathbf{P}_X(B) = \mathbf{P}(X \in B)$, where $B \in \mathcal{B}$, is called *probability distribution of random variable* X. The cdf of $\mathbf{P}_X$, respectively, is called *cumulative distribution function of random variable* X.

Definition 3.9 If X is a random element, then the probability measure $\mathbf{P}_X$ defined on $(S, \mathcal{B}(S, d))$ by $\mathbf{P}_X(B) = \mathbf{P}(X \in B)$, where $B \in \mathcal{B}(S, d)$, is called *probability distribution of random element* X.

Note that there are no cdfs for random elements.

Exercise 3.8 Show that $\mathbf{P}_X$ is a probability measure (for random variables and elements).

Examples of metric spaces.

- Real line $\mathbb{R}$ with $d_0(x,y) = |x-y|$ is a metric space. Random element on $\mathbb{R}$ is called random variable.
- Real line $\mathbb{R}$ with
$$d_1(x,y) = \frac{|x-y|}{1+|x-y|}$$
is a metric space.
- $\mathbb{R}^n$ with
$$d(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$
is a metric space. This space is called *Euclidean n-space.* Random element on $\mathbb{R}^n$ is called *random vector.*
- The space of sequences, $\mathbb{R}^\infty$, with metric
$$d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{\infty} 2^{-i} \frac{|x_i - y_i|}{1+|x_i - y_i|}$$
is a metric space. Random elements on $\mathbb{R}^\infty$ is called *random sequence.*
- The space of continuous functions on $[0,1]$, $C[0,1]$, with metric
$$d(\mathbf{x},\mathbf{y}) = \max_{0 \le t \le 1} |x(t) - y(t)|$$
is a metric space. Random element on $C[0,1]$ is called *random function.*

Exercise 3.9 Show that all the metrics above *are* metrics.

Exercise 3.10 Show that $\mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}, d_0) = \mathcal{B}(\mathbb{R}, d_1)$.

Chapter 4

Independence

4.1 Definitions of Independence

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.

Definition 4.1 *Let $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be sub-σ-fields of $\mathcal{F}$. These σ-fields are called* independent *if for any finite collection $\{i_k\}_{k=1}^n$ and $A_{i_k} \in \mathcal{F}_{i_k}$ we have*

$$\mathbf{P}(\bigcap_{k=1}^{n} A_{i_k}) = \prod_{k=1}^{n} \mathbf{P}(A_{i_k}).$$

Definition 4.2 *Random variables $X_1, X_2, \ldots$ are called* independent *if the corresponding σ-fields $\sigma(X_1), \sigma(X_2), \ldots$ are independent.*

Definition 4.3 *Events $A_1, A_2, \ldots$ are called* independent *if $\sigma(I_{A_1}), \sigma(I_{A_2}), \ldots$ are independent.*

4.2 Basic Criterion of Independence

Example 4.1 Let A and B be two events such that $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. Then, for instance, we also have

$$\begin{aligned}
\mathbf{P}(A^c \cap B^c) &= 1 - \mathbf{P}(A) - \mathbf{P}(B) + \mathbf{P}(A \cap B) \\
&= 1 - \mathbf{P}(A) - \mathbf{P}(B) + \mathbf{P}(A)\mathbf{P}(B) \\
&= (1 - \mathbf{P}(A))(1 - \mathbf{P}(B)) \\
&= \mathbf{P}(A^c)\mathbf{P}(B^c).
\end{aligned}$$

The next result generalizes this idea. Independence of σ-fields can be checked on a smaller classes of events. More specifically, the following result is true.

Theorem 4.1 *Suppose that $\mathcal{I}$ and $\mathcal{J}$ are π-systems, and $\sigma(\mathcal{I})$ and $\sigma(\mathcal{J})$ are sub-σ-fields of $\mathcal{F}$. The σ-fields $\sigma(\mathcal{I})$ and $\sigma(\mathcal{J})$ are independent iff*

$$\mathbf{P}(I \cap J) = \mathbf{P}(I)\mathbf{P}(J) \text{ for any } I \in \mathcal{I}, J \in \mathcal{J}.$$

Proof. Fix $I \in \mathcal{I}$ such that $\mathbf{P}(I) > 0$. Let us define the probability measure (check that it is a probability measure) on $\sigma(\mathcal{J})$:

$$\mathbf{P}_I(B) = \frac{\mathbf{P}(I \cap B)}{\mathbf{P}(I)}, \quad B \in \sigma(\mathcal{J}).$$

The measures $\mathbf{P}$ and $\mathbf{P}_I$ agree on $\mathcal{J}$. Therefore, by the set induction theorem

(Theorem 2.2) $\mathbf{P}$ and $\mathbf{P}_I$ agree on $\sigma(\mathcal{J})$, i.e.

$$\mathbf{P}_I(B) = \frac{\mathbf{P}(I \cap B)}{\mathbf{P}(I)} = \mathbf{P}(B) \quad \text{for any } B \in \sigma(\mathcal{J}), I \in \mathcal{I}.$$

If $\mathbf{P}(I) = 0$, then $\mathbf{P}(I \cap B) = \mathbf{P}(I)\mathbf{P}(B)$ is obviously true.

Now, fix $B \in \sigma(\mathcal{J})$ such that $\mathbf{P}(B) > 0$. For any $A \in \sigma(\mathcal{I})$ we define

$$\mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}, \quad A \in \sigma(\mathcal{I}).$$

By the first part the measures $\mathbf{P}$ and $\mathbf{P}_B$ agree on $\mathcal{I}$. Therefore, by the set induction theorem $\mathbf{P}$ and $\mathbf{P}_B$ agree on $\sigma(\mathcal{I})$, i.e.

$$\mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \mathbf{P}(A) \quad \text{for any } A \in \sigma(\mathcal{I}), B \in \sigma(\mathcal{J}).$$

If $\mathbf{P}(B) = 0$, then $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ is obviously true. □

By induction this result can be extended to any finite number of π-system.

Corollary 4.1 *Two random variables X and Y are independent iff for any $x, y \in \mathbb{R}$*

$$\mathbf{P}(X \leq x \cap Y \leq y) = \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y).$$

Proof. Just note that $\{\omega : X(\omega) \leq x\}$ is a π-system that generates $\sigma(X)$. □

Corollary 4.2 *The finite collection of random variables $X_1, X_2, ..., X_n$ are in-*

dependent iff for any $x_1, x_2, ..., x_n \in \mathbb{R}$

$$\mathbf{P}(\bigcap_{i=1}^{n} X_i \leq x_i) = \prod_{i=1}^{n} \mathbf{P}(X_i \leq x_i).$$

Finally, let us state here one important result without a proof even though we will use it all the time.

Theorem 4.2 *Let* $\{F_n\}_{n\geq 1}$ *be a sequence of cdfs. Then there exist a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and a sequence of independent random variables* $\{X_n\}_{n\geq 1}$ *defined on the probability space which sdfs are* $\{F_n\}_{n\geq 1}$.

4.3 Borel-Cantelli Lemmas

Next two results are important tools in proving strong limit theorems (for instance, Kolmogorov's Law of Large Numbers).

Lemma 4.1 (First Borel-Cantelli Lemma) *Let* $\{A_n\}_{n\geq 1}$ *be a sequence of events with*

$$\sum_n \mathbf{P}(A_n) < \infty.$$

Then

$$\mathbf{P}(A_n \ i.o.) = \mathbf{P}(\limsup_{n\to\infty} A_n) = 0.$$

Proof. Observe that

$$\mathbf{P}(A_n \text{ i.o}) \quad = \quad \mathbf{P}(\lim_{n\to\infty} \bigcup_{j\geq n} A_j)$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \mathbf{P}(\bigcup_{j\geq n} A_j) \text{ (continuity of } \mathbf{P}) \\
&\leq \lim_{n\to\infty} \sum_{j\geq n} \mathbf{P}(A_j) \text{ (subadditivity of } \mathbf{P}) \\
&= 0,
\end{aligned}
$$

as a tail of the converging series $\sum_n \mathbf{P}(A_n)$. $\square$

Exercise 4.1 Let $\{X_n\}$ be a sequence of independent Bernoulli random variables (taking values 1 or -1) with $\mathbf{P}(X_i = 1) > \mathbf{P}(X_i = -1)$, and let $S_n = X_1 + ... + X_n$. Show that $\mathbf{P}(S_n = 0 \text{ i.o.}) = 0$.

Lemma 4.2 (Second Borel-Cantelli Lemma) *Let $\{A_n\}_{n\geq 1}$ be a sequence of* independent *events with*

$$\sum_n \mathbf{P}(A_n) = \infty.$$

Then

$$\mathbf{P}(A_n \ i.o.) = \mathbf{P}(\limsup_{n\to\infty} A_n) = 1.$$

Proof. First note that

$$(\limsup_{n\to\infty} A_n)^c = \liminf_{n\to\infty} A_n^c = \bigcup_n \bigcap_{j\geq n} A_j^c.$$

For any $i \geq j \geq n$ we have

$$\mathbf{P}(\bigcap_{i\geq j\geq n} A_j^c) = \prod_{i\geq j\geq n} (1 - \mathbf{P}(A_j)).$$

Taking $i \to \infty$ we get

$$\mathbf{P}(\bigcap_{j \geq n} A_j^c) = \prod_{j \geq n} (1 - \mathbf{P}(A_j)).$$

Using inequality $1 - x \leq \exp(-x)$ we obtain

$$\prod_{j \geq n} (1 - \mathbf{P}(A_j)) \leq \exp\left(- \sum_{j > n} \mathbf{P}(A_j) \right) = 0,$$

that is, $\mathbf{P}((\limsup_{n\to\infty} A_n)^c) = 0$. $\square$

Example 4.2 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent identically distributed (i.i.d.) random variables with exponential distribution:

$$\mathbf{P}(X_n > x) = \exp(-x), \quad x \geq 0.$$

Note that for $\alpha > 0$

$$\mathbf{P}(X_n > \alpha \log n) = n^{-\alpha},$$

therefore,

$$\mathbf{P}(X_n > \alpha \log n \text{ i.o.}) = \begin{cases} 0, & \text{if } \alpha > 1, \\ 1, & \text{if } \alpha \leq 1. \end{cases}$$

That is,

$$\mathbf{P}(\limsup_n (X_n / \log n)) \geq 1) \geq \mathbf{P}(X_n > \log n \text{ i.o.}) = 1.$$

On the other hand, for any $k \in \mathbb{N}$,

$$\mathbf{P}(\limsup_n (X_n/\log n) > 1 + k^{-1}) \leq \mathbf{P}(X_n > (1+k^{-1})\log n \text{ i.o.}) = 0.$$

Thus by monotonicity of $\mathbf{P}$ we get

$$\begin{aligned} \mathbf{P}(\limsup_n (X_n/\log n) > 1) &= \mathbf{P}(\bigcup_k \{\limsup_n (X_n/\log n) > 1 + k^{-1}\}) \\ &\leq \sum_k \mathbf{P}(X_n > (1+k^{-1})\log n \text{ i.o.}) = 0, \end{aligned}$$

and, as a consequence,

$$\limsup \frac{X_n}{\log n} = 1 \quad \text{a.s.}$$

In general, the following result is true.

Lemma 4.3 *Let* $\{X_n\}_{n\geq 1}$ *be random variables, and* $\{a_n\}_{n\geq 1}$ *be a sequence of positive numbers. Suppose that for any* $0 < \epsilon < 1$

$$\mathbf{P}(X_n \geq (1+\epsilon)a_n \ \ i.o.) = 0,$$

and

$$\mathbf{P}(X_n \geq (1-\epsilon)a_n \ \ i.o.) = 1.$$

Then

$$\limsup_n \frac{X_n}{a_n} = 1 \quad a.s.$$

Proof. Since

$$\{\omega : \limsup_n X_n(\omega)/a_n > 1+\epsilon\} \subset \{\omega : X_n \geq (1+\epsilon)a_n \text{ i.o.}\},$$

we get

$$\mathbf{P}(\limsup_n \frac{X_n}{a_n} > 1+\epsilon) = 0.$$

But we also have

$$\{\omega : \limsup_n X_n(\omega)/a_n \geq 1-\epsilon\} \supset \{\omega : X_n \geq (1-\epsilon)a_n \text{ i.o.}\}.$$

Therefore, it leads us to

$$\mathbf{P}(\limsup_n \frac{X_n}{a_n} \geq 1-\epsilon) = 1.$$

Thus, for $X = \limsup_n X_n(\omega)/a_n$ we obtain that for any $0 < \epsilon < 1$

$$\mathbf{P}(X > 1+\epsilon) = 0 \quad \text{and} \quad \mathbf{P}(X < 1-\epsilon) = 0.$$

Note that

$$\mathbf{P}(X > 1) = \mathbf{P}(\bigcup_{k\geq 1}\{X > 1+1/k\}) \leq \sum_k \mathbf{P}(X > 1+1/k) = 0,$$

and

$$\mathbf{P}(X < 1) = \mathbf{P}(\bigcup_{k\geq 1}\{X < 1 - 1/k\}) \leq \sum_k \mathbf{P}(X < 1 - 1/k) = 0.$$

Therefore, $\mathbf{P}(X = 1) = 1$. $\square$

Exercise 4.2 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with exponential distribution:

$$\mathbf{P}(X_n > x) = \exp(-x), \quad x \geq 0.$$

Show that

$$\limsup_n \frac{X_n}{\log n + \log\log n} = 1 \quad \text{a.s.}$$

Exercise 4.3 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with normal distribution $\mathcal{N}(0,1)$:

$$\mathbf{P}(X_n > x) = \int_x^\infty f(y)dy,$$

where

$$f(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}.$$

(a) Show that

$$\lim_{x\to\infty} \frac{\mathbf{P}(X_n > x)}{f(x)/x} = 1$$

(b) Show that

$$\limsup_n \frac{X_n}{\sqrt{2\log n}} = 1 \quad \text{a.s.}$$

Exercise 4.4 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with Poisson distribution with parameter λ:

$$\mathbf{P}(X_n = x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

(a) Show that

$$\frac{\lambda^n}{n!}e^{-\lambda} \leq \mathbf{P}(X_n \geq n) \leq \frac{\lambda^n}{n!}$$

(b) Show that

$$\limsup \frac{X_n}{\log n/\log(\log n)} = 1 \quad \text{a.s.}$$

4.4 Tail σ-field. Kolmogorov's 0-1 Theorem

Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables. For any $1 \leq m \leq k \leq \infty$ we denote

$$\mathcal{F}_m^k = \sigma(X_m, ..., X_k) = \sigma(\bigcup_{i=m}^{k} \sigma(X_i)).$$

Exercise 4.5 Show that $\mathcal{A} = \bigcup_{i=1}^{\infty} \mathcal{F}_1^i$ is a field, and $\mathcal{F}_1^\infty = \sigma(\mathcal{A})$.

Definition 4.4 *The* tail σ-field $\mathcal{T}$ *is defined as*

$$\mathcal{T} = \bigcap_n \mathcal{F}_n^\infty.$$

Example 4.3 The following events belongs to the tail σ-field:

- $A_1 = \{\sum_n X_n \text{ converges}\}$.
- $A_2 = \{X_n \in I_n \text{ i.o.}\}$ for $I_n \in \mathcal{B}(\mathbb{R})$. Note that if $\{X_n\}_{n\geq 1}$ are independent then

  $$\mathbf{P}(A_2) = 0 \Leftrightarrow \sum \mathbf{P}(X_n \in I_n) < \infty,$$

  $$\mathbf{P}(A_2) = 1 \Leftrightarrow \sum \mathbf{P}(X_n \in I_n) = \infty.$$

  That is, the probability of A_2 can assume only two values: 0 or 1.
- $A_3 = \{\limsup_n X_n < \infty\}$.
- $A_4 = \{\limsup_n (X_1 + ... + X_n)/n < \infty\}$.
- $A_5 = \{\limsup_n (X_1 + ... + X_n)/n < c\}$.
- $A_6 = \{(X_1 + ... + X_n)/n \text{ converges}\}$.
- $A_7 = \{\limsup_n (X_1 + ... + X_n)/b_n = 1\}$ for $b_n \uparrow \infty$

Example 4.4 The following event

$$B = \{\lim_n (X_1 + ... + X_n) \text{ exist and is less than } c\}$$

does not belong to the tail σ-field.

Theorem 4.3 (Kolmogorov's 0-1 Law) *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables, and let $A \in \mathcal{T}$. The probability of A can take only the values 0 or 1.*

Proof. Note that $A \in \mathcal{F}_1^\infty$, and there exist sets $A_n \in \mathcal{F}_1^n$ such that $\mathbf{P}(A \triangle A_n) \to 0$ as $n \to \infty$ (show it!). We have then

$$\mathbf{P}(A_n) \to \mathbf{P}(A), \quad \mathbf{P}(A_n \cap A) \to \mathbf{P}(A).$$

But $A_n \in \mathcal{F}_1^n$ and $A \in \mathcal{F}_{n+1}^\infty$, therefore, A and A_n are independent. Hence we get

$$\mathbf{P}(A_n \cap A) \to \mathbf{P}(A), \text{ and } \mathbf{P}(A_n \cap A) = \mathbf{P}(A_n)\mathbf{P}(A) \to \mathbf{P}(A)^2.$$

That is, $\mathbf{P}(A) = \mathbf{P}(A)^2$, and, as a consequence, $\mathbf{P}(A) = 0$ or 1. $\square$

Exercise 4.6 Show that for any $A \in \mathcal{F}_1^\infty$ there exist sets $A_n \in \mathcal{F}_1^n$ such that $\mathbf{P}(A \triangle A_n) \to 0$ as $n \to \infty$.

Exercise 4.7 Show that if X is measurable with respect to the tail σ-field which is generated by a sequence of independent random variables, then there is a constant c such that $\mathbf{P}(X = c) = 1$.

Chapter 5

Expectation

5.1 Expectation of Simple Random Variables

First we introduce the expectation for simple random variables. Recall that X is a simple random variable if it can be presented in the following form:

$$X = \sum_{i=1}^{n} a_i 1_{A_i},$$

where $|a_i| < \infty$, and $\sum_{i=1}^{n} A_i = \Omega$.

Definition 5.1 *Define for simple random variable X the* expectation *as*

$$\mathbf{E}(X)(= \int X d\mathbf{P}) = \sum_{i=1}^{n} a_i \mathbf{P}(A_i).$$

Properties of the expectation of simple random variables.

- $\mathbf{E}(1) = 1, \quad \mathbf{E}(1_A) = \mathbf{P}(A)$.

- If $X \geq 0$, then $\mathbf{E}(X) \geq 0$.

- (linearity) $\mathbf{E}(\alpha X + \beta Y) = \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y)$.

  Proof. Let $X = \sum_{i=1}^n a_i 1_{A_i}$ and $Y = \sum_{j=1}^m b_j 1_{B_j}$. Then

$$\alpha X + \beta Y = \sum_{i,j} (\alpha a_i + \beta b_j) 1_{A_i B_j}.$$

  Thus

$$\begin{aligned}
\mathbf{E}(\alpha X + \beta Y) &= \sum_{i,j} (\alpha a_i + \beta b_j) \mathbf{P}(A_i B_j) \\
&= \sum_{i,j} \alpha a_i \mathbf{P}(A_i B_j) + \sum_{i,j} \beta b_j \mathbf{P}(A_i B_j) \\
&= \alpha \sum_{i=1}^n a_i \sum_{j=1}^m \mathbf{P}(A_i B_j) + \beta \sum_{j=1}^m b_j \sum_{i=1}^n \mathbf{P}(A_i B_j) \\
&= \alpha \sum_{i=1}^n a_i \mathbf{P}(A_i) + \beta \sum_{j=1}^m b_j \mathbf{P}(B_j) \\
&= \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y).
\end{aligned}$$

  $\square$

- (monotonicity) If $X \leq Y$ then $\mathbf{E}(X) \leq \mathbf{E}(Y)$.

- (independence) If X and Y are independent, then $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ (prove it!).

- $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$

Proof. Just note that

$$|\mathbf{E}(X)| = |\sum_i a_i \mathbf{P}(A_i)| \leq \sum_i |a_i| \mathbf{P}(A_i) = \mathbf{E}(|X|).$$

$\square$

5.2 Expectation (Lebesgue Integral)

We know that any nonnegative random variable X from $(\Omega, \mathcal{F}, \mathbf{P})$ can be approximated by a sequence of nonnegative simple random variables X_n such that for every $\omega \in \Omega$ $X_n(\omega) \uparrow X(\omega)$. This allows us to define the expectation of X.

Definition 5.2 *Define for nonnegative random variable X the* expectation *as*

$$\mathbf{E}(X) = \lim_n \mathbf{E}(X_n),$$

where $\{X_n\}_{n \geq 1}$ is a sequence of nonnegative simple random variables X_n such that for every $\omega \in \Omega$ $X_n(\omega) \uparrow X(\omega)$.

First note that the limit exists (maybe $+\infty$) because $\mathbf{E}(X_n)$ is monotone increasing. But we also have to show that the expectation of X does not depend on the choice of the approximating sequence. That is, if $X_n(\omega) \uparrow X(\omega)$ and $X'_n(\omega) \uparrow X(\omega)$ then

$$\lim_n \mathbf{E}(X_n) = \lim_n \mathbf{E}(X'_n).$$

Lemma 5.1 *Let X_n and Y be nonnegative simple random variables, and*

$$X_n \uparrow X \geq Y.$$

Then

$$\lim_n \mathbf{E}(X_n) \geq \mathbf{E}(Y).$$

Proof. Let $\epsilon > 0$ and

$$A_n = \{\omega : X_n \geq Y - \epsilon\}.$$

Note that $A_n \uparrow \Omega$ and

$$X_n = X_n 1_{A_n} + X_n 1_{A_n^c} \geq X_n 1_{A_n} \geq (Y - \epsilon) 1_{A_n}.$$

Therefore,

$$\begin{aligned} \mathbf{E}(X_n) \geq \mathbf{E}((Y-\epsilon)1_{A_n}) &= \mathbf{E}(Y 1_{A_n}) - \epsilon \mathbf{P}(A_n) \\ &= \mathbf{E}(Y) - \mathbf{E}(Y 1_{A_n^c}) - \epsilon \mathbf{P}(A_n) \\ &\geq \mathbf{E}(Y) - C\mathbf{P}(A_n^c) - \epsilon, \end{aligned}$$

where $C = \max_\omega Y(\omega)$. Since ϵ is arbitrary and $\mathbf{P}(A_n^c) \to 0$, the proof is finished. $\square$

Exercise 5.1 Show that this lemma implies well definition of $\mathbf{E}(X)$, that is,

$$\lim_n \mathbf{E}(X_n) = \lim_n \mathbf{E}(X_n').$$

Exercise 5.2 Consider random variable

$$X = \sum_{k \geq 1} a_k 1_{A_k},$$

where $a_k \geq 0$ and $\sum_{k \geq 1} A_k = \Omega$. Show that

$$\mathbf{E}(X) = \sum_{k \geq 1} a_k \mathbf{P}(A_k).$$

Definition 5.3 *Define for random variable X the* expectation *as*

$$\mathbf{E}(X) = \mathbf{E}(X^+) - \mathbf{E}(X^-),$$

if at least one of $\mathbf{E}(X^+)$ and $\mathbf{E}(X^-)$ is finite. The expectation of X is said to be finite *if both $\mathbf{E}(X^+)$ and $\mathbf{E}(X^-)$ are finite (or $\mathbf{E}(|X|) < \infty$).*

5.3 Properties of Expectation

Let us give now some properties of the expectation of arbitrary random variable X.

- Let $c \neq 0$, and let $\mathbf{E}(X)$ exist. Then $\mathbf{E}(cX)$ also exists, and

$$\mathbf{E}(cX) = c\mathbf{E}(X).$$

Proof. We know that this true for simple random variable. Now let us assume first that $X \geq 0$ and $c > 0$. Let $X_n \uparrow X$, where X_n are simple random variables. Then $cX_n \uparrow cX$, and

$$\mathbf{E}(cX) = \lim_n \mathbf{E}(cX_n) = c \lim_n \mathbf{E}(X_n) = c\mathbf{E}(X).$$

In case of arbitrary X we need to employ the representation $X = X^+ - X^-$. (Do it!) $\square$

- (monotonicity) Let $X \leq Y$, then $\mathbf{E}(X) \leq \mathbf{E}(Y)$.

  Proof. Again we really need to show it only for nonnegative X and Y, that is, when $0 \leq X \leq Y$. Let $X_n \uparrow X$ and $Y_m \uparrow Y$. For any fix n by Lemma 5.1 we have

$$\mathbf{E}(Y) = \lim_m \mathbf{E}(Y_m) \geq \mathbf{E}(X_n),$$

  and therefore,

$$\mathbf{E}(Y) = \lim_n \mathbf{E}(Y) \geq \lim_n \mathbf{E}(X_n) = \mathbf{E}(X).$$

  Now, note that in the general case

$$X \leq Y \Rightarrow X^+ \leq Y^+, \quad \text{and } X^- \geq Y^-.$$

By the first part we have

$$\mathbf{E}(X^+) \leq \mathbf{E}(Y^+), \text{ and } \mathbf{E}(X^-) \geq \mathbf{E}(Y^-),$$

and by the definition of expectation we obtain

$$\mathbf{E}(X) \leq \mathbf{E}(Y).$$

□

- If $\mathbf{E}(X)$ exists then $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$

  Proof. Just note that $-|X| \leq X \leq |X|$, therefore, by the first two properties we get

  $$-\mathbf{E}(|X|) \leq \mathbf{E}(X) \leq \mathbf{E}(|X|).$$

  □

- If $\mathbf{E}(X)$ exists, then for any event A from $\mathcal{F}$ $\mathbf{E}(X1_A)$ exists. If $\mathbf{E}(X)$ is finite, then $\mathbf{E}(X1_A)$ is also finite.

  Proof. Note that $(X1_A)^+ = X^+1_A \leq X^+$ and $(X1_A)^- = X^-1_A \leq X^-$, so by the monotonicity we get the needed result. □

- (additivity) If X and Y are nonnegative or with finite expectations, then

  $$\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y).$$

Proof. Let us prove it for nonnegative random variables. Let X_n and Y_n be approximating sequences of simple random variables, then observe that

$$\mathbf{E}(X_n + Y_n) = \mathbf{E}(X_n) + \mathbf{E}(Y_n),$$

$$X_n + Y_n \uparrow X + Y,$$

$$\mathbf{E}(X_n + Y_n) \uparrow \mathbf{E}(X + Y), \mathbf{E}(X_n) \uparrow \mathbf{E}(X), \text{ and } \mathbf{E}(Y_n) \uparrow \mathbf{E}(Y).$$

□

Definition 5.4 *We say that a property holds* almost surely *(*$\mathbf{P}$*-almost surely,* $\mathbf{P}$*-a.s, a.s.) if there is a set* $\mathcal{N}$ *with* $\mathbf{P}(\mathcal{N}) = 0$ *such that the property holds for every point outside* $\mathcal{N}$*.*

Now we will present some properties of integrable random variables that holds "almost surely".

- If $X = 0$ a.s., then $\mathbf{E}(X) = 0$.

  Proof. Assume that $X \geq 0$. Let X_n be approximating sequence of simple random variables. If $X_n = \sum x_{kn} 1_{A_k} \leq X$, and $x_{kn} \neq 0$, then $\mathbf{P}(A_k) = 0$, and, as a result, $\mathbf{E}(X_n) = 0$, and $\mathbf{E}(X) = \lim \mathbf{E}(X_n) = 0$. □

- If $X = Y$ a.s. and their expectation are finite, then $\mathbf{E}(X) = \mathbf{E}(Y)$. (Prove it!)

- Let $X \geq 0$, and $\mathbf{E}(X) = 0$. Then $X = 0$ a.s.

Proof. Define

$$A = \{\omega : X(\omega) > 0\}, \quad A_n = \{\omega : X(\omega) > 1/n\}.$$

Obviously

$$0 \leq X1_{A_n} \leq X1_A \leq X,$$

and by monotonicity we get

$$0 \leq \mathbf{E}(X1_{A_n}) \leq \mathbf{E}(X) = 0.$$

But

$$0 = \mathbf{E}(X1_{A_n}) \geq \frac{1}{n}\mathbf{P}(A_n).$$

That is, $\mathbf{P}(A_n) = 0$ for all n. But $A_n \uparrow A$ and by continuity of the probability

$$\mathbf{P}(A) = \lim \mathbf{P}(A_n) = 0.$$

□

- Let X and Y be random variables with finite expectation. If for all $A \in \mathcal{F}$ $\mathbf{E}(X1_A) \leq \mathbf{E}(Y1_A)$, then $X \leq Y$ a.s.

  Proof. Define

$$A = \{\omega : X(\omega) > Y(\omega)\}.$$

For this event we have

$$\mathbf{E}(Y1_A) \leq \mathbf{E}(X1_A) \leq \mathbf{E}(Y1_A).$$

That is, $\mathbf{E}(X1_A) = \mathbf{E}(Y1_A)$, and by linearity $\mathbf{E}((X-Y)1_A) = 0$. By the previous property $(X-Y)1_A = 0$ a.s. Therefore, $0 = \mathbf{P}((X-Y)1_A \neq 0) = \mathbf{P}((X-Y)1_A > 0) = \mathbf{P}(A)$. $\square$

5.4 Taking Limits under Expectation Sign

Theorem 5.1 (Monotone Convergence Theorem) *Let $Y, X, X_1, X_2, \ldots$ be random variables.*

(1) If $X_n \geq Y$ for all n, $\mathbf{E}(Y) > -\infty$, and $X_n(\omega) \uparrow X(\omega)$ for every ω, then

$$\mathbf{E}(X_n) \uparrow \mathbf{E}(X).$$

(2) If $X_n \leq Y$ for all n, $\mathbf{E}(Y) < \infty$, and $X_n(\omega) \downarrow X(\omega)$ for every ω, then

$$\mathbf{E}(X_n) \downarrow \mathbf{E}(X).$$

Proof. Note that (2) follows from (1) if we substitute the random variables by their negatives, so we need to prove only (1).

First consider case when $Y \geq 0$. For every $k \geq 1$ define $\{X_{nk}\}$ as an approximating X_k sequence of simple random variables. Define $Z_n = \max_{1 \leq k \leq n} X_{nk}$.

Thus, we get the following diagram:

$$
\begin{array}{cccccccccccccc}
 & X_1 & \leq & X_2 & \leq & X_3 & \leq & \cdots & \leq & X_k & \leq & \cdots & \leq & X \\
 & \vee\!| & & \vee\!| & & \vee\!| & & & & \vee\!| & & & & \\
 & \vdots & & \vdots & & \vdots & & & & \vdots & & & & \\
 & \vee\!| & & \vee\!| & & \vee\!| & & & & \vee\!| & & & & \\
Z_3 & \mathbf{X_{31}} & & \mathbf{X_{32}} & & \mathbf{X_{33}} & & \cdots & & X_{3k} & & \cdots & & \\
 & \vee\!| & & \vee\!| & & \vee\!| & & & & \vee\!| & & & & \\
Z_2 & \mathbf{X_{21}} & & \mathbf{X_{22}} & & X_{23} & & \cdots & & X_{2k} & & \cdots & & \\
 & \vee\!| & & \vee\!| & & \vee\!| & & & & \vee\!| & & & & \\
Z_1 & \mathbf{X_{11}} & & X_{12} & & X_{13} & & \cdots & & X_{1k} & & \cdots & &
\end{array}
$$

Therefore, we have

$$Z_{n-1} \leq Z_n = \max_{1\leq k\leq n} X_{nk} \leq \max_{1\leq k\leq n} X_k = X_n.$$

Let $Z = \lim_n Z_n$. Since for $1 \leq k \leq n$

$$X_{nk} \leq Z_n \leq X_n,$$

by taking limits as $n \to \infty$ we get that

$$X_k \leq Z \leq X$$

which means that $Z = X$. Since Z_n is a "simple approximation" of Z, we get

$$\mathbf{E}(X) = \mathbf{E}(Z) = \lim_n \mathbf{E}(Z_n) \leq \lim_n \mathbf{E}(X_n) \leq \mathbf{E}(X).$$

(Note that the last inequality is obvious—why? Also why do we need here

$Y \geq 0$?)

Now let us look at the general case. If $\mathbf{E}(Y) = \infty$, then by monotonicity $\mathbf{E}(X) = \infty$ also. Now if $\mathbf{E}(Y) < \infty$, then $\mathbf{E}|Y| < \infty$. Then $X_n - Y \geq 0$, by the first part

$$\mathbf{E}(X_n - Y) \uparrow \mathbf{E}(X - Y),$$

then by linearity

$$\mathbf{E}(X_n) - \mathbf{E}(Y) \uparrow \mathbf{E}(X) - \mathbf{E}(Y),$$

and, since $\mathbf{E}(Y) < \infty$ we have

$$\mathbf{E}(X_n) \uparrow \mathbf{E}(X).$$

□

Exercise 5.3 Show that convergence for every ω in the theorem can be substituted by a.s. convergence.

Corollary 5.1 *Let $\{X_n\}_{n\geq 1}$ be a sequence of nonnegative random variables. Then*

$$\mathbf{E}(\sum_n X_n) = \sum_n \mathbf{E}(X_n).$$

Theorem 5.2 (Fatou's Lemma) *Let $Y, X_1, X_2, ...$ be random variables.*

(1) If $X_n \geq Y$ for all n, and $\mathbf{E}(Y) > -\infty$, then

$$\mathbf{E}(\liminf_n X_n) \leq \liminf_n \mathbf{E}(X_n).$$

(2) If $X_n \le Y$ for all n, and $\mathbf{E}(Y) < \infty$, then

$$\limsup_n \mathbf{E}(X_n) \le \mathbf{E}(\limsup_n X_n).$$

(3) If $|X_n| \le Y$ for all n, and $\mathbf{E}(Y) < \infty$, then

$$\mathbf{E}(\liminf_n X_n) \le \liminf_n \mathbf{E}(X_n) \le \limsup_n \mathbf{E}(X_n) \le \mathbf{E}(\limsup_n X_n).$$

Proof. We really need to prove only (1). Define $Z_n = \inf_{m \ge n} X_m$. Then $\liminf_n X_n = \lim_n \inf_{m \ge n} X_m = \lim_n Z_n$, and $Z_n \uparrow \liminf_n X_n$. By the monotone convergence theorem we obtain

$$\mathbf{E}(\liminf_n X_n) = \mathbf{E}(\lim_n Z_n) = \lim_n \mathbf{E}(Z_n) = \liminf_n \mathbf{E}(Z_n) \le \liminf_n \mathbf{E}(X_n),$$

because obviously $Z_n \le X_n$. □

Exercise 5.4 Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$. Let $X_n(\omega) = n^2 1_{\omega < 1/n}$. Check the Fatou's lemma for the random variables.

Theorem 5.3 (Dominated Convergence Theorem) *Let $Y, X_1, X_2, \ldots$ be random variables such that $|X_n| \le Y$ for all n, $\mathbf{E}(Y) < \infty$, and $X_n \to X$ a.s. Then $\mathbf{E}|X| < \infty$ and as $n \to \infty$*

$$\mathbf{E}(X_n) \to \mathbf{E}(X),$$

and

$$\mathbf{E}|X_n - X| \to 0.$$

Proof. Since $\liminf_n X_n = \limsup_n X_n = \lim_n X_n = X$ by Fatou's lemma we have (with $=$ instead of $\leq$)

$$\mathbf{E}(X) = \mathbf{E}(\liminf_n X_n) = \liminf_n \mathbf{E}(X_n) = \limsup_n \mathbf{E}(X_n) = \mathbf{E}(\limsup_n X_n) = \mathbf{E}(X),$$

that is, $\mathbf{E}(X_n) \to \mathbf{E}(X)$. The second equation can be obtained by considering $|X_n - X|$ and observing that it is dominated by $2Y$. $\square$

Corollary 5.2 *Let $Y, X_1, X_2, \ldots$ be random variables such that $|X_n| \leq Y$ for all n, $\mathbf{E}|Y|^p < \infty$ for $p > 0$, and $X_n \to X$ a.s.. Then $\mathbf{E}|X|^p < \infty$ and as $n \to \infty$*

$$\mathbf{E}|X_n - X|^p \to 0.$$

Theorem 5.4 *Let X and Y be independent random variables with finite expectations. Then $\mathbf{E}|XY| < \infty$ and*

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y).$$

Proof. Let us prove it for nonnegative X and Y (the general case then can be done via the standard representation $X = X^+ - X^-$). Let

$$X_n = \sum_{k \geq 1} \frac{k}{n} 1_{k/n \leq X < (k+1)/n},$$

and

$$Y_n = \sum_{k\geq 1} \frac{k}{n} 1_{k/n \leq Y < (k+1)/n}.$$

It is easy to see that by the dominated convergence theorem we have

$$\lim_n \mathbf{E}(X_n) = \mathbf{E}(X) \text{ and } \lim_n \mathbf{E}(Y_n) = \mathbf{E}(Y).$$

By independence we also have

$$\begin{aligned}
\mathbf{E} X_n Y_n &= \sum_{k,m\geq 1} \frac{km}{n^2} \mathbf{E}[1_{k/n \leq X < (k+1)/n} 1_{m/n \leq Y < (m+1)/n}] \\
&= \sum_{k,m\geq 1} \frac{km}{n^2} \mathbf{E}[1_{k/n \leq X < (k+1)/n}] \mathbf{E}[1_{m/n \leq Y < (m+1)/n}] \\
&= \mathbf{E}(X_n)\mathbf{E}(Y_n).
\end{aligned}$$

Finally, note

$$\begin{aligned}
|\mathbf{E}(XY) - \mathbf{E}(X_n Y_n)| &\leq \mathbf{E}|XY - X_n Y_n| \\
&\leq \mathbf{E}(X|Y - Y_n|) + \mathbf{E}(Y_n|X - X_n|) \\
&\leq \frac{1}{n}\mathbf{E}(X) + \frac{1}{n}\mathbf{E}(Y) \to 0.
\end{aligned}$$

That is, $\mathbf{E}(XY) = \lim_n \mathbf{E}(X_n Y_n) = \lim_n \mathbf{E}(X_n)\mathbf{E}(Y_n) = \mathbf{E}(X)\mathbf{E}(Y)$ $\square$

Exercise 5.5 This proof is a good illustration of a typical application of the dominated convergence theorem. However, the right way of doing it is via simple random variable approximations that are introduced in Theorem 3.2. Note that a "simple" approximation is less accurate, but it is monotone. Prove the theorem using the simple random variable approximation.

5.5 Uniform Integrability

Now we provide a more delicate condition that allows us to take limits under the expectation sign.

Definition 5.5 *We say that a family of random variables* $\{X_t\}_{t\in T}$ *is* uniformly integrable *iff*

$$\sup_t \mathbf{E}(|X_t|1_{|X_t|>c}) \to 0, \ \textit{as } c \to \infty.$$

Exercise 5.6 Show that "$|X_n| < Y, \mathbf{E}(Y) < \infty$" makes the family $\{X_n\}_{n\geq 1}$ uniformly integrable.

The next theorem shows what we really need to have to obtain the implication of the Fatou's lemma.

Theorem 5.5 *Let* $\{X_n\}_{n\geq 1}$ *be uniformly integrable.*

(1) Then

$$\mathbf{E}(\liminf_n X_n) \leq \liminf_n \mathbf{E}(X_n) \leq \limsup_n \mathbf{E}(X_n) \leq \mathbf{E}(\limsup_n X_n).$$

(2) If $X_n \to X$ *a.s. then* $\mathbf{E}|X| < \infty$ *and as* $n \to \infty$

$$\mathbf{E}(X_n) \to \mathbf{E}(X),$$

and

$$\mathbf{E}|X_n - X| \to 0.$$

Proof. (1) For every $c > 0$

$$\mathbf{E}(X_n) = \mathbf{E}[X_n 1_{X_n \leq c}] + \mathbf{E}[X_n 1_{X_n > c}],$$

and for any $\epsilon > 0$ we can find c such that

$$\sup_n |\mathbf{E}[X_n 1_{X_n > c}]| < \epsilon.$$

Therefore, for such c

$$\limsup_n \mathbf{E}(X_n) \leq \limsup_n \mathbf{E}(X_n 1_{X_n \leq c}) + \epsilon.$$

By Fatou's lemma and by observing that $X_n 1_{X_n \leq c} \leq X_n$ we get

$$\begin{aligned} \limsup_n \mathbf{E}(X_n) - \epsilon &\leq \limsup_n \mathbf{E}(X_n 1_{X_n \leq c}) \\ &\leq \mathbf{E}(\limsup_n X_n 1_{X_n \leq c}) \\ &\leq \mathbf{E}(\limsup_n X_n). \end{aligned}$$

Since ϵ is arbitrary, we obtain that $\limsup_n \mathbf{E}(X_n) \leq \mathbf{E}(\limsup_n X_n)$. In similar way we can prove $\mathbf{E}(\liminf_n X_n) \leq \liminf_n \mathbf{E}(X_n)$.

(2) Take a look at the proof of the dominated convergence theorem. $\square$

Theorem 5.6 *Let $0 \leq X_n \to X$ a.s., and $\mathbf{E}(X_n) < \infty$. Then $\mathbf{E}(X_n) \to \mathbf{E}(X) < \infty$ if and only if the family $\{X_n\}_{n\geq 1}$ is uniformly integrable.*

Exercise 5.7 We already have established "if" part, prove the "only if" part.

Lemma 5.2 *If $\{X_n\}_{n\geq 1}$ be uniformly integrable, then*

$$\sup_n \mathbf{E}|X_n| < \infty.$$

Proof. Just note that

$$\sup_n \mathbf{E}|X_n| = \sup_n \left[\mathbf{E}[|X_n| 1_{|X_n|\leq c}] + \mathbf{E}[|X_n| 1_{|X_n|>c}]\right],$$

and let c be large enough to make $\sup_n \mathbf{E}[|X_n| 1_{|X_n|>c}] < 1$, then $\sup_n \mathbf{E}|X_n| < c + 1$. $\square$

But $\sup_n \mathbf{E}|X_n| < \infty$ is not enough for the uniform integrability. We need a bit more.

Lemma 5.3 *Let $\{X_n\}_{n\geq 1}$ be integrable, and let $G(\cdot)$ be a nonnegative increasing function such that $\lim_t G(t)/t = \infty$. If*

$$\sup_n \mathbf{E}[G(|X_n|)] = M < \infty,$$

then $\{X_n\}_{n\geq 1}$ is uniformly integrable.

Proof. For any (large) A we can find c large enough to guarantee $G(t)/t > A$ if

$t > c$. For that c we have

$$\begin{aligned} \sup_n \mathbf{E}[|X_n| 1_{|X_n|>c}] &= \sup_n \mathbf{E}[G(|X_n|) \frac{|X_n|}{G(|X_n|)} 1_{|X_n|>c}] \\ &\le \frac{1}{A} \sup_n \mathbf{E}[G(|X_n|) 1_{|X_n|>c}] \le \frac{M}{A} \end{aligned}$$

i.e. it can be made as small as we want. $\square$

Example 5.1 For instance, $G(t) = |t|^{1+\epsilon}$ for $\epsilon > 0$ fits the description.

Exercise 5.8 Give an example of $\{X_n\}_{n\ge1}$ for which $\sup_n \mathbf{E}|X_n| < \infty$, but the family is still not uniformly integrable.

5.6 Inequalities for Expectations

Proposition 5.1 (Chebyshev's Inequality) *Let X be nonnegative random variable. Then*

$$\mathbf{P}(X > \epsilon) \le \frac{\mathbf{E}(X)}{\epsilon},$$

in particular, for any random variable X we have

$$\mathbf{P}(|X| > \epsilon) \le \frac{\mathbf{E}(X^2)}{\epsilon^2},$$

and

$$\mathbf{P}(|X - \mathbf{E}(X)| > \epsilon) \le \frac{\mathbf{E}(X - \mathbf{E}X)^2}{\epsilon^2} = \frac{\mathbf{Var}(X)}{\epsilon^2}.$$

Proof. Just observe

$$\mathbf{E}(X) \geq \mathbf{E}(X 1_{X>\epsilon}) \geq \epsilon \mathbf{E}(1_{X>\epsilon}) \geq \epsilon \mathbf{P}(X > \epsilon).$$

$\square$

Proposition 5.2 (Cauchy-Shwarz's Inequality) *Let X and Y be random variables with $\mathbf{E}(X^2) < \infty$ and $\mathbf{E}(Y^2) < \infty$. Then*

$$(\mathbf{E}|XY|)^2 \leq \mathbf{E}(X^2)\mathbf{E}(Y^2).$$

Proof. When $\mathbf{E}(X^2) = 0$ or $\mathbf{E}(Y^2) = 0$ the inequality is obvious (still, show it!).

So suppose that both $\mathbf{E}(X^2) > 0$ and $\mathbf{E}(Y^2) > 0$. First note that $2|XY| \leq X^2 + Y^2$, therefore, $\mathbf{E}|XY| < \infty$. Now, let us look at the quadratic polynomial:

$$p(t) = \mathbf{E}(|X|t + |Y|)^2 = \mathbf{E}|X|^2 t^2 + 2\mathbf{E}|XY|t + \mathbf{E}|Y|^2 = At^2 + 2Bt + C.$$

Since $p(t) \geq 0$, we get $B^2 \leq AC$. $\square$

Definition 5.6 *A Borel function $f : \mathbb{R} \mapsto \mathbb{R}$ is said to be* convex *iff for any y there is a number a(y) such that*

$$f(x) \geq f(y) + (x - y)a(y)$$

for all $x \in \mathbb{R}$.

Proposition 5.3 (Jensen's Inequality) *Let f be a convex Borel function, and let X be random variables with a finite expectation. Then*

$$f(\mathbf{E}(X)) \leq \mathbf{E}(f(X)).$$

Proof. By convexity we have (assuming $x = X$, $y = \mathbf{E}(X)$)

$$f(X) \geq f(\mathbf{E}(X)) + (X - \mathbf{E}(X))a(\mathbf{E}(X)),$$

and after taking the expectation we obtain Jensen's inequality. $\square$

Proposition 5.4 (Lyapunov's Inequality) *If $0 < s < r$, then*

$$(\mathbf{E}|X|^s)^{1/s} \leq (\mathbf{E}|X|^r)^{1/r}.$$

Proof. Consider random variable $|X|^s$, and apply Jensen's inequality to convex function $f(\cdot) = |\cdot|^{r/s}$. $\square$

Proposition 5.5 (Hölder's Inequality) *Let X and Y be random variables with $\mathbf{E}|X|^p < \infty$ and $\mathbf{E}|Y|^q < \infty$ for some $p, q > 1$, $1/p + 1/q = 1$. Then*

$$\mathbf{E}(|XY|) \leq (\mathbf{E}|X|^p)^{1/p}(\mathbf{E}|Y|^q)^{1/q}.$$

Proof. If $\mathbf{E}|X|^p = 0$ or $\mathbf{E}|Y|^q = 0$, the inequality is trivial. So, assume that both $\mathbf{E}|X|^p$ and $\mathbf{E}|Y|^q$ are positive. Using the fact that e^x is convex one can show that for any $a, b \geq 0$ we have

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \tag{5.1}$$

Let

$$a = \frac{|X|}{(\mathbf{E}|X|^p)^{1/p}}, \text{ and } b = \frac{|Y|}{(\mathbf{E}|Y|^q)^{1/q}}.$$

That is, we have

$$\frac{|XY|}{(\mathbf{E}|X|^p)^{1/p}(\mathbf{E}|Y|^q)^{1/q}} \leq \frac{1}{p}\left|\frac{X}{(\mathbf{E}|X|^p)^{1/p}}\right|^p + \frac{1}{q}\left|\frac{Y}{(\mathbf{E}|Y|^q)^{1/q}}\right|^q.$$

Taking expectation we get

$$\frac{\mathbf{E}|XY|}{(\mathbf{E}|X|^p)^{1/p}(\mathbf{E}|Y|^q)^{1/q}} \leq 1.$$

□

Proposition 5.6 (Minkovski's Inequality) *Let X and Y be random variables with $\mathbf{E}|X|^p < \infty$ and $\mathbf{E}|Y|^p < \infty$ for some $p \geq 1$. Then $\mathbf{E}|X+Y|^p < \infty$ and*

$$(\mathbf{E}|X+Y|^p)^{1/p} \leq (\mathbf{E}|X|^p)^{1/p} + (\mathbf{E}|Y|^p)^{1/p}.$$

Proof. Using convexity of $|\cdot|^p$ one can show (do it!) that for any $a, b > 0$ and $p \geq 1$ we have

$$(a+b)^p \leq 2^{p-1}(a^p + b^p),$$

i.e., we have $\mathbf{E}|X+Y|^p < \infty$.

When $p=1$ the inequality is trivial. Consider the case when $p>1$. Since $p-1=p/q$, by Hölder's inequality we have

$$\begin{aligned}
\mathbf{E}|X+Y|^p &= \mathbf{E}|X+Y||X+Y|^{p/q} \\
&\leq \mathbf{E}|X||X+Y|^{p/q} + \mathbf{E}|Y||X+Y|^{p/q} \\
&\leq (\mathbf{E}|X|^p)^{1/p}(\mathbf{E}|X+Y|^p)^{1/q} + (\mathbf{E}|Y|^p)^{1/p}(\mathbf{E}|X+Y|^p)^{1/q} \\
&= [(\mathbf{E}|X|^p)^{1/p} + (\mathbf{E}|Y|^p)^{1/p}](\mathbf{E}|X+Y|^p)^{1/q}
\end{aligned}$$

By dividing the left and right sides of the inequality by $(\mathbf{E}|X+Y|^p)^{1/q}$ we get the result. $\square$

That is, if we ignore a.s. difference between random variables with finite p-moments, the function $d(X,Y) = (\mathbf{E}|X-Y|^p)^{1/p}$ is a metric.

5.7 Radon-Nikodým Theorem

Definition 5.7 *We say that a probability measure* $\mathbf{Q}$ *on* $(\Omega, \mathcal{F})$ *is* absolutely continuous *with respect to* $\mathbf{P}$ *if*

$$\mathbf{P}(A) = 0 \quad \Rightarrow \quad \mathbf{Q}(A) = 0.$$

Proposition 5.7 *Let* X *be a nonnegative random variable with* $\mathbf{E}(X) = 1$. *Then*

$$\mathbf{Q}(A) = \mathbf{E}(X 1_A)$$

is a probability measure and it is absolutely continuous with respect to $\mathbf{P}$.

Proof. Consider first a simple random variable $X = \sum_{k=1}^{n} x_k 1_{A_k}$. If $\mathbf{P}(A) = 0$ then

$$\mathbf{E}(X1_A) = \sum_{k=1}^{n} x_k \mathbf{P}(A_k \cap A) = 0.$$

Thus the proposition holds if X is a simple random variable. Now for an arbitrary X first we construct a sequence of simple random variables $X_n \uparrow X$ the by the monotone convergence theorem for A with $\mathbf{P}(A) = 0$ we get

$$\mathbf{Q}(A) = \mathbf{E}(X1_A) = \mathbf{E}(\lim_{n\to\infty} X_n 1_A) = \lim_{n\to\infty} \mathbf{E}(X_n 1_A) = 0.$$

□

The converse is also true (but much harder to prove) and it is called Radon-Nikodým Theorem.

Theorem 5.7 (Radon-Nikodým Theorem) *If a probability measure* $\mathbf{Q}$ *on* $(\Omega, \mathcal{F})$ *is* absolutely continuous *with respect to* $\mathbf{P}$ *then there exist a nonnegative random variable* X *(called* Radon-Nikodým *derivative) with* $\mathbf{E}(X) = 1$ *such that* $\mathbf{Q}(A) = \mathbf{E}(X1_A)$.

5.8 Change of Variables in a Lebesgue Integral

Theorem 5.8 (Change of Variables in a Lebesgue Integral) *Let* X *be a random variable with probability distribution* $\mathbf{P}_X$[1]. *If* h *is a Borel function and*

[1]The probability measure $\mathbf{P}_X$ is defined on $(\mathbb{R}, \mathcal{B})$ by $\mathbf{P}_X(B) = \mathbf{P}(X \in \mathcal{B})$, where $B \in \mathcal{B}$.

for $A \in \mathcal{B}$ either of the integrals,

$$\int_A h(x)\mathbf{P}_X(dx) = \int_R 1_A(x)h(x)\mathbf{P}_X(dx)$$

or

$$\int_{X^{-1}(A)} h(X(\omega))\mathbf{P}(d\omega) = \mathbf{E}\left[1_{X^{-1}(A)}h(X(\omega))\right],$$

exists then

$$\int_A h(x)\mathbf{P}_X(dx) = \int_{X^{-1}(A)} h(X(\omega))\mathbf{P}(d\omega). \tag{5.2}$$

Proof. First consider case when $h(x) = 1_B(x)$, where $B \in \mathcal{B}$. Observe that $X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B)$. Therefore,

$$\begin{aligned}\int_A 1_B(x)\mathbf{P}_X(dx) &= \int_R 1_A(x)1_B(x)\mathbf{P}_X(dx) = \int_R 1_{AB}(x)\mathbf{P}_X(dx) = \mathbf{P}_X(AB)\\ &= \mathbf{P}(X^{-1}(AB)) = \mathbf{E}\left[1_{X^{-1}(AB)}\right] = \mathbf{E}\left[1_{X^{-1}(A)}1_{X^{-1}(B)}\right]\\ &= \int_{X^{-1}(A)} 1_B(X(\omega))\mathbf{P}(d\omega).\end{aligned}$$

Now, if we have (5.2) for indicators, then we have it for nonnegative simple Borel functions, and, therefore, we also have (5.2) for nonnegative Borel functions. The general case is treated as usual via the representation $h = h^+ - h^-$. □

In particular,

$$\mathbf{E}(h(X)) = \int_\Omega h(X(\omega))\mathbf{P}(d\omega) = \int_R h(x)\mathbf{P}_X(dx) = \int_R h(x)F_X(dx),$$

i.e., to compute the expectation it is not necessary to know $\mathbf{P}$, knowing the distribution function F_X is enough.

Exercise 5.9 Suppose that

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy,$$

where f_X is a non-negative Riemann-integrable function. Show that the Lebesgue-Stieltjes integral

$$\int_R h(x)F_X(dx)$$

is equal to the Riemann integral

$$\int_{-\infty}^{+\infty} h(x)f_X(x)dx$$

for any non-negative Riemann-integrable Borel function h.

5.9 Product Spaces and Fubini's Theorem

Definition 5.8 *By the* Cartesian product *of two arbitrary sets A and B, denoted by $A \times B$, we mean the set of all ordered pairs (a, b), $a \in A$, $b \in B$.*

Let us consider two measurable spaces with two σ-finite[2] measures $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$, and a measurable space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, where $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\})$, that is, the smallest σ-field that contains the set of rectangles with measurable sides.

[2]A measure μ is said to be σ-finite if Ω can be partitioned into $\sum_i A_i$ in such way that $\mu(A_i) < \infty$ for all i.

Theorem 5.9 *There exists the unique measure* $\mu_1 \otimes \mu_2$ *defined on* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ *in such a way that*

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),$$

where $A_1 \in \mathcal{F}_1$ *and* $A_2 \in \mathcal{F}_2$.

The *product space* is the following triple $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$.

Theorem 5.10 (Fubini's Theorem) *Let* $f(\cdot, \cdot)$ *be a* $\mathcal{F}_1 \otimes \mathcal{F}_2$*-measurable function on* $\Omega_1 \times \Omega_2$ *such that*

$$\int_{\Omega_1 \times \Omega_2} |f(\omega_1, \omega_2)| d\mu_1 \otimes \mu_2 < \infty.$$

Then the integrals $\int_{\Omega_1} |f(\omega_1, \omega_2)| d\mu_1$ *and* $\int_{\Omega_2} |f(\omega_1, \omega_2)| d\mu_2$

- *are defined for all* ω_1 *and* ω_2,
- *are respectively* $\mathcal{F}_2$*- and* $\mathcal{F}_1$*-measurable functions with*

$$\mu_2\{\omega_2 : \int_{\Omega_1} |f(\omega_1, \omega_2)| d\mu_1 = \infty\} = 0,$$

and

$$\mu_1\{\omega_1 : \int_{\Omega_2} |f(\omega_1, \omega_2)| d\mu_2 = \infty\} = 0.$$

- *and*

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2)d\mu_1\otimes\mu_2 &= \int_{\Omega_2}\left[\int_{\Omega_1} f(\omega_1,\omega_2)d\mu_1\right]d\mu_2\\ &= \int_{\Omega_1}\left[\int_{\Omega_2} f(\omega_1,\omega_2)d\mu_2\right]d\mu_1.\end{aligned}$$

Corollary 5.3 *If*

$$\int_{\Omega_1}\left[\int_{\Omega_2} |f(\omega_1,\omega_2)|d\mu_2\right]d\mu_1 < \infty$$

the Fibini's theorem will hold.

Example 5.2 Let X be a non-negative random variable on $(\Omega,\mathcal{F},\mathbf{P})$. Consider $(\Omega\times\mathbb{R}_+,\mathcal{F}\otimes\mathcal{B}(\mathbb{R}_+),\mathbf{P}\otimes\lambda)$ and

$$f(\omega,x) = 1_A(\omega,x),\ \text{where } A = \{(\omega,x) : 0\le x\le X\}.$$

Note that

$$\int_{\mathbb{R}_+} f(\omega,x)d\lambda = X(\omega)$$

and

$$\int_\Omega f(\omega,x)d\mathbf{P} = \mathbf{P}(X\ge x).$$

Therefore, by Fubini's theorem we obtain

$$\begin{aligned}\mathbf{P}\otimes\lambda(A) &= \int_\Omega X(\omega)d\mathbf{P} = \mathbf{E}(X)\\ &= \int_{\mathbb{R}_+}\mathbf{P}(X\ge x)d\lambda = \int_0^{+\infty}\mathbf{P}(X\ge x)dx.\end{aligned}$$

Example 5.3 Consider a pair of random variables (X, Y) and suppose that their joint distribution has a density, i.e.

$$\mathbf{P}((X,Y) \in B) = \int_B f_{XY}(x,y)dxdy,$$

where $B \in \mathcal{B}(\mathbb{R}^2)$ and the integral is taken with respect two-dimensional Lebesgue measure.

Note that for $A \in \mathcal{B}(\mathbb{R})$ by Fibini's theorem we get

$$\mathbf{P}(X \in A) = \mathbf{P}((X,Y) \in A \times \mathbb{R}) = \int_{A\times\mathbb{R}} f_{XY}(x,y)dxdy = \int_A \left[\int_{\mathbb{R}} f_{XY}(x,y)dy\right] dx.$$

That is, the densities of X and Y exist and they are given by

$$f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)dy$$

and

$$f_Y(x) = \int_{\mathbb{R}} f_{XY}(x,y)dx.$$

According to Corollary 4.1 the random variables X and Y are independent iff for any $x, y \in \mathbb{R}$

$$\mathbf{P}(X \leq x \cap Y \leq y) = \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y).$$

Now one can show the following is true.

Corollary 5.4 *If a joint density $f_{XY}(x,y)$ exists then the random variables X*

and Y are independent iff

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

almost surely with respect two-dimensional Lebesgue measure.

Proof. Sufficiency. Just note that

$$\begin{aligned}
\mathbf{P}(X \le x \cap Y \le y) &= \int_{(-\infty,x]\times(-\infty,y]} f_{XY}(t,s)dtds \\
&= \int_{(-\infty,x]\times(-\infty,y]} f_X(t)f_Y(s)dtds \\
&= \int_{(-\infty,x]} f_X(t)dt \int_{(-\infty,y]} f_Y(s)ds \\
&= \mathbf{P}(X \le x)\mathbf{P}(Y \le y).
\end{aligned}$$

Necessity. If X and Y are independent then

$$\mathbf{P}(X \le x \cap Y \le y) = \mathbf{P}(X \le x)\mathbf{P}(Y \le y).$$

That is, by Fubini's theorem

$$\begin{aligned}
\int_{(-\infty,x]\times(-\infty,y]} f_{XY}(t,s)dtds &= \int_{(-\infty,x]} f_X(t)dt \int_{(-\infty,y]} f_Y(s)ds \\
&= \int_{(-\infty,x]\times(-\infty,y]} f_X(t)f_Y(s)dtds.
\end{aligned}$$

Therefore, for any $B \in \mathcal{B}(\mathbb{R}^2)$ we have that

$$\int_B f_{XY}(t,s)dtds = \int_B f_X(t)f_Y(s)dtds$$

which gives us (see page 67) that

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

almost surely with respect two-dimensional Lebesgue measure. $\square$

Exercise 5.10 Consider the following product space

$$([0,1] \times [0,1], \mathcal{B}([0,1]) \otimes \mathcal{B}([0,1]), \lambda \otimes \mu),$$

where λ is the Lebesgue measure, and μ just counts the number of elements in a set. Let A be a diagonal $\{(x,y) \in [0,1] \times [0,1] : x = y\}$.

a) Argue that $A \in \mathcal{B}([0,1]) \otimes \mathcal{B}([0,1])$.

b) Calculate

$$\int_{[0,1]} \left[\int_{[0,1]} 1_A(x,y) d\lambda \right] d\mu$$

and

$$\int_{[0,1]} \left[\int_{[0,1]} 1_A(x,y) d\mu \right] d\lambda.$$

c) Why do you think Fubini's Theorem does not hold for this example?

5.10 Variance, Covariance and Covariance Matrix

Definition 5.9 *Let X be a random variable with $\mathbf{E}X^2 < \infty$. The* variance *is defined by*

$$\mathbf{Var}(X) = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}(X^2) - (\mathbf{E}X)^2.$$

Definition 5.10 *Let X, Y be random variables with $\mathbf{E}X^2 < \infty$ and $\mathbf{E}Y^2 < \infty$. The* covariance *of X and Y is*

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

Exercise 5.11 Find the variance for uniform, normal, Poisson, and binomial distributions.

Properties of variance.

- $\mathbf{Var}(X) \geq 0$.
- $\mathbf{Var}(c) = 0$, where c is a constant.
- If $\mathbf{Var}(X) = 0$, then $\mathbf{P}(X = c) = 1$ and $c = \mathbf{E}(X)$.
- $\mathbf{Var}(cX) = c^2\mathbf{Var}(X)$.
- $\mathbf{Var}(aX + bY)^2 = a^2\mathbf{Var}(X) + b^2\mathbf{Var}(Y) + 2ab\mathbf{Cov}(X, Y)$.
- If X and Y are independent, then $\mathbf{Cov}(X, Y) = 0$ (because of Theorem 5.4), and $\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$.

Exercise 5.12 Give an example of two dependent random variables which covariance is equal to 0.

Definition 5.11 *Let* $\mathbf{X} = (X_1, \ldots, X_n)^\top$ *be a random (column) vector with* $\mathbf{E}X_i^2 < \infty$ *for* $i = 1, \ldots, n$*. Vector* $\mu = (\mathbf{E}X_1, \ldots, \mathbf{E}X_n)^\top$ *is called the* expectation *of* X*. The* covariance matrix *of random vector* $\mathbf{X}$ *is the following* $n \times n$ *matrix:*

$$\begin{aligned} \mathbf{\Sigma_X} &= \mathbf{E}(\mathbf{X}-\mu)(\mathbf{X}-\mu)^\top \\ &= \mathbf{E}(\mathbf{X}\mathbf{X}^\top) - \mu\mu^\top \\ &= \begin{bmatrix} \mathbf{Var}(X_1) & \mathbf{Cov}(X_1, X_2) & \ldots & \mathbf{Cov}(X_1, X_n) \\ \mathbf{Cov}(X_2, X_1) & \mathbf{Var}(X_2) & \ldots & \mathbf{Cov}(X_2, X_n) \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{Cov}(X_n, X_1) & \mathbf{Cov}(X_n, X_2) & \ldots & \mathbf{Var}(X_n) \end{bmatrix}. \end{aligned}$$

Properties of covariance matrix.

- If $\mathbf{a}$ is an n-dimensional (column) vector of real numbers, then $\mathbf{Var}(\mathbf{a}^\top\mathbf{X}) = \mathbf{a}^\top\mathbf{\Sigma}_X\mathbf{a}$.
- The covariance matrix $\mathbf{\Sigma_X}$ is positive-semidefinite and symmetric.

  Proof. Just observe that $\mathbf{Cov}(X_i, X_j) = \mathbf{Cov}(X_j, X_i)$ and for any $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a}^\top\mathbf{\Sigma}_X\mathbf{a} = \mathbf{Var}(\mathbf{a}^\top\mathbf{X}) \geq 0$. $\square$
- $\mathbf{\Sigma_{X+a}} = \mathbf{\Sigma_X}$.
- $\mathbf{\Sigma_{AX}} = \mathbf{A}\mathbf{\Sigma_X}\mathbf{A}^\top$, here $\mathbf{A}$ is an $m \times n$ matrix of real numbers.

- If $\text{rank}(\mathbf{\Sigma_X}) < n$, then there exists vector of real numbers $\mathbf{a} \neq \mathbf{0}$ such that with probability one $\mathbf{a}^\top \mathbf{X} = \mathbf{a}^\top \mu$, that is, with probability 1 the values of $\mathbf{X}$ belong to a hyperplane.

  Proof. $\text{rank}\,(\mathbf{\Sigma_X}) < n$

  $\Leftrightarrow \det(\mathbf{\Sigma_X}) = 0$

  $\Leftrightarrow \exists \mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq \mathbf{0}$ such that $\mathbf{\Sigma_X a} = \mathbf{0}$

  $\Leftrightarrow \exists \mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq \mathbf{0}$ such that $\mathbf{a}^\top \mathbf{\Sigma_X a} = 0$

  $\Leftrightarrow \exists \mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq \mathbf{0}$ such that $\mathbf{Var}(\mathbf{a}^\top \mathbf{X}) = 0$

  $\Leftrightarrow \exists \mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq \mathbf{0}$ such that $\mathbf{P}(\mathbf{a}^\top \mathbf{X} = \mathbf{a}^\top \mu) = 1$ □

Definition 5.12 *Let X, Y be random variables with $0 < \mathbf{Var}(X) < \infty$ and $0 < \mathbf{Var}(Y) < \infty$. The* correlation coefficient *of X and Y is*

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}(X)\mathbf{Var}(Y)}}.$$

Properties of correlation coefficient.

- $|\rho(X, Y)| \leq 1$

  Proof. By Cauchy-Shwarz's inequality (5.2) we have

$$[\mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y)]^2 \leq \mathbf{E}(X - \mathbf{E}X)^2 \mathbf{E}(Y - \mathbf{E}Y)^2.$$

  □

- $|\rho(X, Y)| = 1 \Leftrightarrow \exists a, b \in \mathbb{R}$ such that $\mathbf{P}(X = aY + b) = 1$

Proof. ($\Rightarrow$) Consider the following two-dimensional random vector $\xi = (X, Y)^\top$. Since $|\rho(X, Y)| = 1$, we get that $\det(\mathbf{\Sigma}_\xi) = 0$, that is, $\mathrm{rank}(\mathbf{\Sigma}_\xi) < 2$. Therefore, $\exists c \neq 0, d \neq 0 \in \mathbb{R}$ (why are both not equal to zero?) such that $\mathbf{P}(cX + dY = c\mathbf{E}X + d\mathbf{E}Y) = 1$, and $\mathbf{P}(X = -(d/c)Y + (c\mathbf{E}X + d\mathbf{E}Y)/c) = 1$. $\square$

Exercise 5.13 Let X, Y be random variables with $0 < \mathbf{Var}(X) < \infty$ and $0 < \mathbf{Var}(Y) < \infty$, and $\mathbf{P}(X = aY + b) = 1$. Show that $\rho(X, Y) = \mathrm{sign}(a)$.

Chapter 6

Characteristic Functions

6.1 Definition and Basic Properties

Definition 6.1 *The* characteristic function *(CF) of a random variable X with distribution function F_X is the complex valued function of $t \in \mathbb{R}$ given by*

$$\begin{aligned}
\phi_X(t) &= \mathbf{E}e^{itX} \\
&= \mathbf{E}\cos(tX) + i\mathbf{E}\sin(tX) \\
&= \int_{\mathbb{R}} \cos(tx)F_X(dx) + i\int_{\mathbb{R}} \sin(tx)F_X(dx).
\end{aligned}$$

First, note that $\phi_X(t)$ always exists because $|e^{itX}| = 1$. Second, when we deal with discrete and absolutely continuous distributions we use more practical formulas:

- (discrete)

$$\phi_X(t) = \sum_{\text{all } x_k} e^{itx_k} p_k$$

- (absolutely continuous)

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$$

Example 6.1 (Standard Normal) Let X be a random variable with the standard normal distribution. Recall that the pdf of the standard normal distribution is given by

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

First let us show that for $n = 0, 1, \ldots$

$$\mathbf{E}X^{2n} = \int_{-\infty}^{\infty} x^{2n} f(x) dx = (2n-1)!!.$$

Indeed, note that via integration by parts one has

$$\begin{aligned}
\mathbf{E}X^{2n} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2n} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2n-1} d(-e^{-x^2/2}) \\
&= -\frac{1}{\sqrt{2\pi}} x^{2n-1} e^{-x^2/2} \Big|_{-\infty}^{+\infty} + (2n-1) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2n-2} e^{-x^2/2} dx.
\end{aligned}$$

Thus, $\mathbf{E}X^{2n} = (2n-1)\mathbf{E}X^{2n-2}$ and $\mathbf{E}X^0 = 1$. Therefore,

$$\mathbf{E}X^{2n} = (2n-1) \times \cdots \times 1 = (2n-1)!!.$$

Now, since $\mathbf{E}X^{2n+1} = 0$ we get

$$\begin{aligned}
\phi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixt} e^{-x^2/2} dx \\
&= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \int_{-\infty}^{\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!} (2n-1)!! = \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!!} \\
&= \sum_{n=0}^{\infty} (-1)^n \left(\frac{t^2}{2}\right)^n \frac{1}{n!} \\
&= e^{-t^2/2}.
\end{aligned}$$

Let us do it in a different way. Observe that again by integration by parts

$$\begin{aligned}
\phi'_X(t) &= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{ixt} e^{-x^2/2} dx \\
&= \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixt} d(-e^{-x^2/2}) \\
&= -t \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixt} e^{-x^2/2} dx.
\end{aligned}$$

That is, the characteristic function of the standard normal distribution satisfies the following differential equation:

$$\phi'_X(t) = -t\phi_X(t), \text{ with } \phi_X(0) = 1.$$

Solving the equation we get

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

Exercise 6.1 Find the characteristic functions of uniform, Poisson, and binomial distributions.

Properties of characteristic functions.

- $\phi(0) = 1$
- $|\phi(t)| \le 1$
- $\phi(-t) = \overline{\phi(t)}$.[1] If $\phi(\cdot)$ is a real function, then it is an even function.
- If X is symmetric (that is, X and $-X$ have the same distribution), then ϕ_X is an even real function.
- The characteristic function $\phi(t)$ is uniformly continuous on $\mathbb{R}$.

  Proof. For any $t, h \in \mathbb{R}$ we have

$$\begin{aligned} |\phi(t+h) - \phi(t)| &= |\mathbf{E}e^{itX}(e^{ihX} - 1)| \\ &\le \mathbf{E}|e^{itX}(e^{ihX} - 1)| \\ &= \mathbf{E}|(e^{ihX} - 1)|. \end{aligned}$$

Since $|e^{ihX} - 1| \to 0$ with probability 1 as $h \to 0$, $|e^{ihX} - 1| < 2$, and

[1] Recall the complex conjugate $\overline{a+ib} = a - ib$.

$|e^{ihX} - 1|$ does not depend on t, by the dominated convergence theorem $|\phi(t+h) - \phi(t)| \to 0$ as $h \to 0$ uniformly for all $t \in \mathbb{R}$. $\square$

- $\phi_{aX+b} = e^{itb}\phi_X(at)$. For instance, immediately we get that if X is a normal random variable with mean μ and standard deviation σ, then

$$\phi_X(t) = e^{it\mu - \sigma^2 t^2/2}. \tag{6.1}$$

- Let X and Y be independent random variables. Then

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

Proof. Because of independence

$$\phi_{X+Y}(t) = \mathbf{E}e^{it(X+Y)} = \mathbf{E}e^{itX}e^{itY} = \mathbf{E}e^{itX}\mathbf{E}e^{itY} = \phi_X(t)\phi_Y(t).$$

$\square$

The last property provides us with the main motivation for the introduction of characteristic functions. Random variables are measurable functions on a sample space. Adding two or more random variables or taking linear combinations are the most basic operations that we can apply to a collection of random variables. So it is not surprising that sums of random variables play an important role in applications (think about the sample mean or sample variance). But the distribution of sum of two independent random variables is given by the *convolution* of the distributions. Convolution is a relatively complex operation. The switch to characteristic functions allows us to substitute the operation of convolution of distributions by product of characteristic functions.

Lemma 6.1 *Let X and Y be independent random variable with distribution functions F_X and F_Y. Then the distribution function of $X+Y$ is the convolution of F_X and F_Y:*

$$F_{X+Y}(z) = \int_{\mathbb{R}} F_X(z-y) F_Y(dy).$$

Proof. By Theorem 5.8 (change of variables in a Lebesgue integral) and Fubini's theorem we have

$$\begin{aligned}
F_{X+Y}(z) &= \mathbf{P}(X+Y \leq z) \\
&= \mathbf{E} 1_{X+Y \leq z} \\
&= \int_{\mathbb{R}^2} 1_{X+Y \leq z} F_{X,Y}(dxdy) \\
&= \int_{\mathbb{R}} \left[\int_{-\infty}^{z-y} F_X(dx) \right] F_Y(dy) \\
&= \int_{\mathbb{R}} F_X(z-y) F_Y(dy).
\end{aligned}$$

□

Exercise 6.2 Let ϕ be a characteristic function, and $c > 0$. Show that $e^{c(\phi(t)-1)}$ is a CF as well. Hint: Let $\{X_k\}_{k \geq 1}$ be iid random variables, and N be an independent of $\{X_k\}_{k \geq 1}$ random variable with a Poisson distribution. Consider random variable $\sum_{k=1}^{N} X_k$.[2]

[2] This random variable has so-called compound Poisson distribution.

Theorem 6.1 (Bochner's Theorem) *An arbitrary complex valued function on* $\mathbb{R}$ $\phi(\cdot)$ *is the characteristic function of some random variable if and only if* $\phi(\cdot)$ *is positive definite*[3]*, continuous, and* $\phi(0) = 1$.

Exercise 6.3 Prove that any characteristic function is positive definite. Note that the "if" part of the theorem is more difficult.

The following sufficient condition is a bit more easy to check.

Theorem 6.2 (Pólya's Theorem) *If* $\phi(\cdot)$ *is a real-valued, even, continuous function which satisfies the conditions*

1) $\phi(0) = 1$,

2) $\phi(\cdot)$ *is convex on* $\mathbb{R}_+$,

3) $\phi(\infty) = 0$,

then $\phi(\cdot)$ *is the characteristic function of an absolutely continuous symmetric distribution.*

Exercise 6.4 Check that $e^{-|t|}$ is a characteristic function.

Exercise 6.5 Let $X_1, \ldots, X_n$ be independent identically distributed random variables with the following characteristic function:

$$\phi(t) = \exp(-|t|^\alpha),$$

where $0 < \alpha \leq 1$.

a) Using Polya's theorem verify that $\phi(\cdot)$ is a characteristic function.

[3] A function $\phi : \mathbb{R} \to \mathbb{C}$ is positive definite if any $t_1, \ldots, t_n \in \mathbb{R}$ and $z_1, \ldots, z_n \in \mathbf{C}$ $\sum_{i,j} \phi(t_i - t_j) z_i \overline{z}_j \geq 0$

b) Show that

$$\frac{1}{n^{1/\alpha}} \sum_{k=1}^{n} X_k$$

has the same characteristic function as X_1.

Definition 6.2 *We say that a discrete random variable has* lattice distribution *if every possible value of the random variable can be represented in the form* $a + kh$*, where* $k = 0, \pm 1, \pm 2, \ldots$

Theorem 6.3 (Characteristic Functions of Lattice Distributions) *A random variable X with characteristic function $\phi(\cdot)$ has a lattice distribution if and only if there exists a real number $t_0 \neq 0$ such that $|\phi(t_0)| = 1$.*

Proof. ($\Rightarrow$) Take $t_0 = 2\pi/h$. Then if $p_k = \mathbf{P}(X = a + kh)$ we have

$$\begin{aligned} |\phi(t_0)| &= |\sum_k e^{i(a+hk)t_0} p_k| = |e^{it_0 a} \sum_k e^{it_0 kh} p_k| \\ &= |e^{it_0 a}|| \sum_k e^{i2\pi k} p_k| = |\sum_k p_k| = 1. \end{aligned}$$

($\Leftarrow$) If for some $t_0 \neq 0$ we have $|\phi(t_0)| = 1$, then there exists a real number α such that $\phi(t_0) = e^{i\alpha}$. Therefore, $\mathbf{E} e^{i(t_0 X - \alpha)} = 1$. That is, $\mathbf{E} \cos(t_0 X - \alpha) = 1$ or

$$0 = \mathbf{E}[\cos(t_0 X - \alpha) - 1] = -2\mathbf{E} \sin^2 \frac{t_0 X - \alpha}{2}.$$

This means that with probability one $|\sin[(t_0 X - \alpha)/2]| = 0$, that is, the discrete distribution measure of X is concentrated on the set of zeros of function $\sin[(t_0 x - \alpha)/2]$. $\square$

6.2 Inversion Formula and Uniqueness

Theorem 6.4 (Inversion Formula) *Let $\phi(\cdot)$ be the characteristic function of distribution function $F(\cdot)$. Consider two points a and b $(a < b)$ at which F is continuous. Then*

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

First let us state few lemmas.

Lemma 6.2

$$\lim_{A\to\infty} \int_{-A}^{A} \frac{\sin(\alpha x)}{x} dx = \pi \mathrm{sign}(\alpha).$$

Proof. Exercise. Prove that

$$\lim_{A\to\infty} \int_{0}^{A} \frac{\sin(x)}{x} dx = \pi/2$$

using the fact that $\int_0^\infty e^{-xt} dt = 1/x$ and Fubini's theorem, but with some caution. $\square$

Lemma 6.3 *Let*

$$I(A, B, \alpha) = \int_{A}^{B} \frac{\sin(\alpha x)}{x} dx.$$

Then uniformly for all A, B, and α

$$|I(A, B, \alpha)| < C.$$

Proof. First note that

$$\begin{aligned}
|I(A,B,\alpha)| &\leq \left|\int_0^A \frac{\sin(\alpha x)}{x}dx\right| + \left|\int_0^B \frac{\sin(\alpha x)}{x}dx\right| \\
&= \left|\int_0^{A\alpha} \frac{\sin(x)}{x}dx\right| + \left|\int_0^{B\alpha} \frac{\sin(x)}{x}dx\right| \\
&\leq 2\sup_{z\geq 0}\left|\int_0^z \frac{\sin(x)}{x}dx\right|
\end{aligned}$$

Since

$$g(z) = \int_0^z \frac{\sin(x)}{x}dx$$

is a continuous function of z, and $\lim_{z\to\infty} g(z) = \pi/2$, we can take T large enough to guarantee that

$$\sup_{z>T} |g(z)| < \pi/2 + 1.$$

Because of continuity of g we also have

$$\sup_{0\leq z\leq T} |g(z)| < C.$$

Therefore,

$$|I(A,B,\alpha)| \leq 2\sup_{z\geq 0} |g(z)| < C_1,$$

where C_1 does not depend on A, B, and α. □

Lemma 6.4 *For $a < b$ consider*

$$\chi(x) = \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{e^{-itb} - e^{-ita}}{-it} e^{itx} dt.$$

Then

$$\chi(x) = \begin{cases} 0 & x < a \text{ or } x > b, \\ 1/2 & x = a \text{ or } x = b, \\ 1 & a < x < b. \end{cases}$$

Proof. Indeed, by Lemma 6.2 we have

$$\begin{aligned}
\chi(x) &= \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{e^{-itb} - e^{-ita}}{-it} e^{itx} dt \\
&= \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{e^{it(x-b)} - e^{it(x-a)}}{-it} dt \\
&= \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{\cos(t(x-b)) - \cos(t(x-a))}{-it} dt \\
&\quad + \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{\sin(t(x-b)) - \sin(t(x-a))}{-t} dt \\
&= \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt \\
&= \frac{1}{2}(\mathrm{sign}(x-a) - \mathrm{sign}(x-b)).
\end{aligned}$$

□

Theorem Proof. Since a, b are continuity points of F we get

$$\begin{aligned}
F(b) - F(a) &= \int_{\mathbb{R}} \chi(x) F(dx) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \lim_{A\to\infty} \int_{-A}^{A} \frac{e^{-itb} - e^{-ita}}{-it} e^{itx} dt F(dx)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi} \lim_{A\to\infty} \int_{\mathbb{R}} \int_{-A}^{A} \frac{e^{-itb} - e^{-ita}}{-it} e^{itx} dt F(dx) \\
&= \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \left(\int_{\mathbb{R}} e^{itx} F(dx) \right) \frac{e^{-itb} - e^{-ita}}{-it} dt \\
&= \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \phi(t) \frac{e^{-itb} - e^{-ita}}{-it} dt.
\end{aligned}
$$

We have two key steps here. First, because of Lemma 6.3 and the dominated convergence theorem we can change order of $\int_{\mathbb{R}}$ and $\lim_{A\to\infty}$. Then, by Fubini's theorem we can change order of integration $\int_{\mathbb{R}}$ and $\int_{-A}^{A}$. $\square$

Theorem 6.5 (Uniqueness Theorem) *Let F and G be distributions functions with the same characteristic function. Then $F(x) = G(x)$ for all x.*

Proof. Exercise. Hint: if distribution function F is not continuous at a, then we can find sequence $a_n \downarrow a$ such that F is continuous at every a_n. $\square$

Exercise 6.6 Let $f : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function.

a) Show that f has limits from the right and from the left.

b) The point $x \in \mathbb{R}$ is called the discontinuity point (of the first kind) of f if limits at x from the right and from the left do not coincide. Prove that f can have no more than countably many points of discontinuity.

Exercise 6.7 Let $X_i, i = 1, 2$ be independent normal random variables with mean μ_i and variance σ_i^2, then $X_1 + X_2$ has normal distribution with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

Exercise 6.8 Let $X_i, i = 1, 2$ be independent Poisson random variables with mean λ_i, then $X_1 + X_2$ has has Poisson distribution with mean $\lambda_1 + \lambda_2$.

Exercise 6.9 Let ϕ_1, ϕ_2, and ϕ_3 be characteristic functions, and $\phi_1(t)\phi_2(t) = \phi_1(t)\phi_3(t)$ for all $t \in \mathbb{R}$. Does it follow that $\phi_2(t) = \phi_3(t)$ for all $t \in \mathbb{R}$? Prove it or construct a counterexample.

Exercise 6.10 Let X_1, X_2, and X_3 be independent random variables such that $X_1 + X_2$ and $X_1 + X_3$ have the same distribution. Does it follow that X_2 and X_3 have the same distribution?

Proposition 6.1 (Inversion Formula for $\mathbb{Z}$-valued RVs) *Suppose random variable X takes values from $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, and $p_k = \mathbf{P}(X = k)$. If $\phi(\cdot)$ is a characteristic function of X, then*

$$p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \phi(t) dt.$$

Proof. Since

$$\phi(t) = \sum_{j=-\infty}^{\infty} e^{itj} p_j$$

we get

$$\int_{-\pi}^{\pi} e^{-itk} \phi(t) dt = \int_{-\pi}^{\pi} \sum_{j=-\infty}^{\infty} e^{-it(j-k)} p_j dt = \sum_{j=-\infty}^{\infty} \left(\int_{-\pi}^{\pi} e^{-it(j-k)} dt \right) p_j = 2\pi p_k,$$

because the integral in the brackets is 0 for $k \neq j$. □

Proposition 6.2 (Inversion Formula for Integrable CFs) *Let F be a distribution function, and ϕ is its characteristic function. If*

$$\int_{\mathbb{R}} |\phi(t)| dt < \infty,$$

then F is absolutely continuous, and its density is given by

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi(t) dt. \tag{6.2}$$

Proof. Let us take function given by formula (6.2). First note that it is a continuous function, because

$$|f(x+h) - f(x)| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |e^{-ith} - 1| |\phi(t)| dt \to 0, \text{ as } h \to 0,$$

by dominated convergence theorem. Therefore, it is integrable on $[a, b]$. Now by Fubini's theorem and the inversion formula we find

$$\begin{aligned}
\int_a^b f(x)dx &= \int_a^b \frac{1}{2\pi} \left(\int_{\mathbb{R}} e^{-itx} \phi(t) dt \right) dx \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \phi(t) \left(\int_a^b e^{-itx} dx \right) dt \\
&= \lim_{A \to \infty} \frac{1}{2\pi} \int_{-A}^{A} \phi(t) \left(\int_a^b e^{-itx} dx \right) dt \\
&= \frac{1}{2\pi} \lim_{A \to \infty} \int_{-A}^{A} \frac{e^{-itb} - e^{-ita}}{-it} \phi(t) dt \\
&= F(b) - F(a)
\end{aligned}$$

for all continuity points a and b. That is,

$$F(x) = \int_{-\infty}^{x} f(x) dx$$

for *all* $x \in \mathbb{R}$. Since one can show that $f(x) \geq 0$ for all x, it finishes the proof.

□

Exercise 6.11 Show that $f(x) \geq 0$ for all x. Hint: we already proved that it is continuous.

6.3 Characteristic Functions and Moments

Let X be a random variable. The k-th absolute moment, $\mathbf{E}|X|^k$, is denoted by β_k. If $\beta_k < \infty$ then the k-th moment, $\mathbf{E}X^k$, is denoted by α_k. Let us start with the following lemma.

Lemma 6.5 *For any $k = 0, 1, \ldots$ and $x \in \mathbb{R}$ we have*

$$\left| e^{ix} - \sum_{j=0}^{k} \frac{(ix)^j}{j!} \right| \leq \frac{|x|^{k+1}}{(k+1)!}.$$

Proof. We will prove it by induction. When $k = 0$ we obviously have

$$|e^{ix} - 1| = \left| \int_0^x e^{it} dt \right| \leq \int_0^{|x|} \left| e^{it} \right| dt = |x|.$$

Now if we introduce

$$R_k(x) = e^{ix} - \sum_{j=0}^{k} \frac{(ix)^j}{j!},$$

then one can show that

$$R_{k+1}(x) = i \int_0^x R_k(t) dt.$$

Therefore, for $x > 0$ we have

$$|R_{k+1}(x)| \leq \int_0^x |R_k(t)|dt \leq \int_0^x \frac{t^{k+1}}{(k+1)!}dt = \frac{x^{k+2}}{(k+2)!}.$$

□

Theorem 6.6 *Let $\phi(\cdot)$ be the characteristic function of random variable X with the cdf $F(\cdot)$ and suppose that for some $k > 0$ the absolute moment β_k is finite. Then*

1. *the k-th derivative $\phi^{(k)}(t)$ exists and for $r \leq k$*

$$\phi^{(r)}(t) = \int_{-\infty}^{\infty} (ix)^r e^{itx} F(dx),$$

2. $\phi^{(r)}(0) = i^r \alpha_r$,

3. *when $t \to 0$*

$$\phi(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!}\alpha_j + o(|t|^k).$$

Proof. In fact, we only need to prove 1. Again we will do it by induction. When $r = 0$ the statement is obvious. Assume that the formula is true for $r < k$. First note that

$$\begin{aligned}
\frac{\phi^{(r)}(t+h) - \phi^{(r)}(t)}{h} &= \int_{-\infty}^{\infty} (ix)^r \frac{e^{i(t+h)x} - e^{itx}}{h} F(dx) \\
&= \int_{-\infty}^{\infty} (ix)^r e^{itx} \frac{e^{ihx} - 1}{h} F(dx)
\end{aligned}$$

By Lemma 6.5 (the easy case when $k = 0$) we have $|e^{ihx} - 1| \leq |hx|$, and, as a consequence,

$$\left|(ix)^r e^{itx} \frac{e^{ihx} - 1}{h}\right| \leq |x|^{r+1}.$$

Since

$$\int_{\mathbb{R}} |x|^{r+1} F(dx) < \infty$$

the dominated convergence theorem tells us that

$$\begin{aligned}
\lim_{h \to 0} \frac{\phi^{(r)}(t+h) - \phi^{(r)}(t)}{h} &= \int_{-\infty}^{\infty} (ix)^r \lim_{h \to 0} \frac{e^{i(t+h)x} - e^{itx}}{h} F(dx) \\
&= \int_{-\infty}^{\infty} (ix)^{r+1} e^{itx} F(dx)
\end{aligned}$$

The second formula immediately follows from the first one, and Taylor's theorem gives 3.

But we also can prove the last statement directly. Indeed, note that for any real tX

$$e^{itX} = \cos(tX) + i\sin(tX) = \sum_{j=0}^{k-1} \frac{(itX)^j}{j!} + \frac{(itX)^k}{k!}[\cos(\theta_1 tX) + i\sin(\theta_2 tX)],$$

where $|\theta_1| < 1$ and $|\theta_2| < 1$. Taking expectation we get

$$\phi(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!} \mathbf{E}X^j + \frac{(it)^k}{k!} \mathbf{E}X^k[\cos(\theta_1 tX) + i\sin(\theta_2 tX) - 1].$$

Since by the dominated convergence theorem

$$\mathbf{E}X^k[\cos(\theta_1 tX) + i\sin(\theta_2 tX) - 1] \to 0, \text{ as } t \to 0,$$

this finishes the proof. $\square$

Theorem 6.7 *Let $\phi(\cdot)$ be the characteristic function of random variable X with the cdf $F(\cdot)$. Suppose that for some $k > 0$ derivative $\phi^{(2k)}(0)$ exists and finite. Then $\beta_{2k} < \infty$.*

First, let us start with case $k = 1$. By Fatou's lemma and l'Hôpital's rule we obtain

$$\begin{aligned}
\phi''(0) &= \lim_{h\to 0} \frac{1}{2}\left[\frac{\phi'(2h) - \phi'(0)}{2h} + \frac{\phi'(0) - \phi'(-2h)}{2h}\right] \\
&= \lim_{h\to 0} \left[\frac{\phi'(2h) - \phi'(-2h)}{4h}\right] \\
&= \lim_{h\to 0} \frac{1}{4h^2}\left[\phi(2h) - 2\phi(0) + \phi(-2h)\right] \\
&= \lim_{h\to 0} \int_{\mathbb{R}} \left[\frac{e^{ihx} - e^{-ihx}}{2h}\right]^2 F(dx) \\
&= -\lim_{h\to 0} \int_{\mathbb{R}} \left[\frac{\sin hx}{hx}\right]^2 x^2 F(dx) \\
&\leq -\int_{\mathbb{R}} \lim_{h\to 0} \left[\frac{\sin hx}{hx}\right]^2 x^2 F(dx) \\
&= -\int_{\mathbb{R}} x^2 F(dx)
\end{aligned}$$

That is, we have $\int_{\mathbb{R}} x^2 F(dx) \leq -\phi''(0) < \infty$.

Now, let us prove the general statement by induction. Assume that $\phi^{(2k+2)}(0)$ exists and finite, and $\beta_{2k} = \int_{\mathbb{R}} x^{2k} F(dx) < \infty$. Note that if $\beta_{2k} = 0$, then

$\beta_{2k+2} = 0$ as well, and we have nothing to prove. So assume that $\beta_{2k} > 0$. If we introduce the following (cumulative distribution) function

$$G(x) = \frac{1}{\beta_{2k}} \int_{-\infty}^{x} y^{2k} F(dy),$$

then using the Theorem 6.6 we get

$$\begin{aligned}
\int_{\mathbb{R}} e^{itx} G(dx) &= \frac{1}{\beta_{2k}} \int_{\mathbb{R}} e^{itx} x^{2k} F(dx) \\
&= \frac{i^{2k}}{\beta_{2k}} \int_{\mathbb{R}} e^{itx} (ix)^{2k} F(dx) \\
&= (-1)^k \phi^{(2k)}(t)/\beta_{2k}
\end{aligned}$$

That is, $(-1)^k \phi^{(2k)}(t)/\beta_{2k}$ is the characteristic function of $G(\cdot)$ and its second derivative exists and finite. Therefore, employing case $k = 1$ we get that

$$\int_{\mathbb{R}} x^2 G(dx) < \infty$$

Since $\beta_{2k+2}/\beta_{2k} = \int_{\mathbb{R}} x^2 G(dx)$, we are done.

Exercise 6.12 Show that e^{-t^4} is not a characteristic function.

Exercise 6.13 Show that if a characteristic function has a finite second derivative at 0 then it is differentiable everywhere.

6.4 Characteristic Function of Random Vectors

Definition 6.3 *Let* $\mathbf{X} = (X_1, \ldots, X_n)^\top$ *be a random vector (column) with distribution measure* $\mathbf{P}_X$*. Then the* characteristic function *of* $\mathbf{X}$ *is the complex valued function of* $\mathbf{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n$ *given by*

$$\begin{aligned}\phi_{\mathbf{X}}(\mathbf{t}) &= \mathbf{E}e^{i\mathbf{X}^\top \mathbf{t}} \\ &= \int_{\mathbb{R}^n} e^{i\mathbf{x}^\top \mathbf{t}} \mathbf{P}_{\mathbf{X}}(d\mathbf{x}).\end{aligned}$$

Properties of characteristic functions.

- $\phi(\mathbf{0}) = 1$
- $|\phi(\mathbf{t})| \leq 1$
- The characteristic function $\phi(\mathbf{t})$ is uniformly continuous on $\mathbb{R}^n$.
- $\phi_{\mathbf{AX+b}}(\mathbf{t}) = e^{i\mathbf{b}^\top \mathbf{t}} \phi_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t})$.

Definition 6.4 *Consider n-dimensional rectangle*

$$I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n].$$

We say that it is a rectangle of continuity *with respect to measure* $\mathbf{P}$ *(on* $\mathbb{R}^n$*) if*

$$\mathbf{P}(\partial I) = 0.$$

Theorem 6.8 (Inversion Formula for Random Vectors) *Let* $\phi(\cdot)$ *be the* *characteristic function of probability distribution measure* $\mathbf{P}$. *Consider rectangle of continuity* $I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$. *Then*

$$\mathbf{P}(I) = \frac{1}{(2\pi)^n} \lim_{A\to\infty} \int_{-A}^{A} \cdots \int_{-A}^{A} \prod_{k=1}^{n} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} \phi(\mathbf{t}) d\mathbf{t}.$$

Proof. Consider function (see Lemma 6.4)

$$\chi_{a,b}(x) = \frac{1}{2\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{e^{-itb} - e^{-ita}}{-it} e^{itx} dt.$$

Then by the dominated convergence theorem and Fubini's theorem we obtain

$$\begin{aligned}
\mathbf{P}(I) &= \int_{\mathbb{R}^n} \prod_{k=1}^{n} \chi_{a_k,b_k}(x_k) \mathbf{P}(d\mathbf{x}) \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \lim_{A\to\infty} \prod_{k=1}^{n} \int_{-A}^{A} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} e^{it_k x_k} dt_k \mathbf{P}(d\mathbf{x}) \\
&= \frac{1}{(2\pi)^n} \lim_{A\to\infty} \int_{\mathbb{R}^n} \int_{-A}^{A} \cdots \int_{-A}^{A} \prod_{k=1}^{n} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} e^{it_k x_k} d\mathbf{t}\mathbf{P}(d\mathbf{x}) \\
&= \frac{1}{(2\pi)^n} \lim_{A\to\infty} \int_{-A}^{A} \cdots \int_{-A}^{A} \prod_{k=1}^{n} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} \int_{\mathbb{R}^n} e^{i\mathbf{x}^\top \mathbf{t}} \mathbf{P}(d\mathbf{x}) d\mathbf{t} \\
&= \frac{1}{(2\pi)^n} \lim_{A\to\infty} \int_{-A}^{A} \cdots \int_{-A}^{A} \prod_{k=1}^{n} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} \phi(\mathbf{t}) d\mathbf{t}.
\end{aligned}$$

□

Theorem 6.9 *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be random vectors with the same characteristic function. Then the corresponding distributions are the same.*

Proof. Exercise. □

Proposition 6.3 (Inversion Formula for Integrable CFs) *Let* $\mathbf{X}$ *be a ran-*

dom vector, and ϕ is its characteristic function. If

$$\int_{\mathbb{R}^n} |\phi(\mathbf{t})| d\mathbf{t} < \infty,$$

then the distribution of $\mathbf{X}$ *is absolutely continuous, and its density is given by*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i\mathbf{x}^\top \mathbf{t}} \phi(\mathbf{t}) d\mathbf{t}. \tag{6.3}$$

Proof. Exercise. □

Theorem 6.10 *Let* $\mathbf{X} = (X_1, \ldots, X_n)^\top$ *be a random vector. The random variables* $X_1, ..., X_n$ *are independent if and only if for all* $\mathbf{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n$

$$\phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^{n} \phi_{X_k}(t_k).$$

Proof.

($\Rightarrow$) Just note that because of independence

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbf{E} e^{i\mathbf{X}^\top \mathbf{t}} = \mathbf{E} \prod_{k=1}^{n} e^{iX_k t_k} = \prod_{k=1}^{n} \mathbf{E} e^{iX_k t_k} = \prod_{k=1}^{n} \phi_{X_k}(t_k).$$

($\Leftarrow$) Random variables X_1, X_2,..., X_n are independent if for any Borel sets B_1, B_2,..., B_n we have that

$$\mathbf{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1)\mathbf{P}(X_2 \in B_2) \cdots \mathbf{P}(X_n \in B_n),$$

or

$$\mathbf{P}_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_n) = \mathbf{P}_{X_1}(B_1)\mathbf{P}_{X_2}(B_2) \cdots \mathbf{P}_{X_n}(B_n).$$

Let $\mathbf{Q} = \mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2} \otimes \cdots \otimes \mathbf{P}_{X_n}$ (a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$). Thus, we need to show that $\mathbf{P}_{\mathbf{X}} = \mathbf{Q}$. Let

$$\pi_1 = \{\emptyset,\ \text{all closed rectangles in } \mathbb{R}^n\},$$

$$\pi_2 = \{\emptyset,\ \text{all closed rectangles of continuity in } \mathbb{R}^n\},$$

and

$$\pi_3 = \{\emptyset,\ \text{all closed rectangles in } \mathbb{R}^n \text{ with } \mathbf{P}(X_k = a_k) = \mathbf{P}(X_k = b_k) = 0\}.$$

Collections π_1, π_2 and π_3 are π-systems, and $\pi_3 \subset \pi_2 \subset \pi_1$. Moreover, $\mathcal{B}(\mathbb{R}^n) = \sigma(\pi_1) = \sigma(\pi_2) = \sigma(\pi_3)$, because any closed rectangle can be approximated by rectangles from π_3. Therefore, by the set induction, it will be sufficient to check that for all rectangles from π_3,

$$I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n],$$

we have

$$\mathbf{P}_{\mathbf{X}}(I) = \prod_{k=1}^{n} \mathbf{P}_{X_k}([a_k, b_k]).$$

This is easy to verify with help of the inversion formulas for random variables and vectors:

$$\begin{aligned}
\mathbf{P}_{\mathbf{X}}(I) &= \frac{1}{(2\pi)^n} \lim_{A\to\infty} \int_{-A}^{A} \cdots \int_{-A}^{A} \prod_{k=1}^{n} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} \phi(\mathbf{t}) d\mathbf{t} \\
&= \frac{1}{(2\pi)^n} \lim_{A\to\infty} \int_{-A}^{A} \cdots \int_{-A}^{A} \prod_{k=1}^{n} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} \phi_{X_k}(t_k) d\mathbf{t} \\
&= \frac{1}{(2\pi)^n} \prod_{k=1}^{n} \lim_{A\to\infty} \int_{-A}^{A} \frac{e^{-it_k b_k} - e^{-it_k a_k}}{-it_k} \phi_{X_k}(t_k) dt_k \\
&= \prod_{k=1}^{n} \mathbf{P}_{X_k}([a_k, b_k]).
\end{aligned}$$

□

6.5 Multivariate Normal Distribution

Definition 6.5 *We say that random vector* $\mathbf{X}$ *has* multivariate normal distribution *if for any* $\mathbf{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n$ *random variable* $\mathbf{X}^\top \mathbf{t}$ *has a univariate normal distribution or a degenerate distribution.*

Exercise 6.14 Show that if random variable X is normal then $(X, X)^\top$ is a normal vector.

Exercise 6.15 Show that if X and Y are independent normal random variables then $(X, Y)^\top$ is a normal vector.

Lemma 6.6 *If* $\mathbf{X}$ *has normal distribution and* $\mathbf{A}$ *is* $n \times n$ *matrix then random variable* $\mathbf{AX}$ *is normal.*

Proof. Exercise. □

Theorem 6.11 (CFs of Normal Random Vector) *The following two statements hold.*

1. *Let* $\mathbf{X}$ *be a normal random vector with mean* μ *and covariance matrix* $\mathbf{\Sigma}$*, then*

$$\phi(\mathbf{t}) = e^{i\mu^\top \mathbf{t} - \mathbf{t}^\top \mathbf{\Sigma} \mathbf{t}/2}. \tag{6.4}$$

2. *For any vector* μ *and positive-semidefinite symmetric matrix* Σ *there exists normal vector* $\mathbf{X}$ *with the characteristic function (6.4).*

Proof.

1. Let us consider random variable $\mathbf{X}^\top \mathbf{t}$. It has normal distribution with mean $\mu^\top \mathbf{t}$ and variance $\mathbf{t}^\top \mathbf{\Sigma} \mathbf{t}$. By formula (6.1) we get

$$\begin{aligned}
\phi(\mathbf{t}) &= \mathbf{E} e^{i\mathbf{X}^\top \mathbf{t} \cdot s}\Big|_{s=1} \\
&= \phi_{\mathbf{X}^\top \mathbf{t}}(1) \\
&= \exp\left[i\mu^\top \mathbf{t} s - \mathbf{t}^\top \mathbf{\Sigma} \mathbf{t} s^2/2\right]\Big|_{s=1} \\
&= e^{i\mu^\top \mathbf{t} - \mathbf{t}^\top \mathbf{\Sigma} \mathbf{t}/2}.
\end{aligned}$$

2. Without loss of generality we assume that $\mu = \mathbf{0}$.

Case 1. If $\mathbf{\Sigma}$ is diagonal, that is, $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \dots, \sigma_n^2)$. Then formula (6.4) gives us

$$\phi(\mathbf{t}) = \exp\left[-\sum_{k=1}^n \sigma_k^2 t_k^2/2\right] = \prod_{k=1}^n \exp\left[-\sigma_k^2 t_k^2/2\right].$$

Now, if X_k, $k = 1, \dots, n$ are independent normal random variable with mean 0 and variance σ_k^2, then random vector $\mathbf{X} = (X_1, \dots, X_n)^\top$ has multivariate normal distribution. Moreover, by Theorem 6.10 its characteristic function is given by (6.4).

Case 2. If $\mathbf{\Sigma}$ is not diagonal (but positive-semidefinite and symmetric) then there exists an orthogonal[4] matrix $\mathbf{U}$ such that $\mathbf{U\Sigma U}^\top$ is diagonal. Let $\mathbf{Y}$ be the normal random vector that corresponds to matrix $\mathbf{U\Sigma U}^\top$. Let $\mathbf{X} = \mathbf{U}^\top\mathbf{Y}$. By Lemma 6.6 X is a normal vector, and its characteristic function is given by

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbf{E}e^{i\mathbf{X}^\top\mathbf{t}} = \mathbf{E}e^{i\mathbf{Y}^\top(\mathbf{Ut})} = \phi_{\mathbf{Y}}(\mathbf{Ut}) = e^{-\mathbf{t}^\top\mathbf{U}^\top\mathbf{U\Sigma U}^\top\mathbf{Ut}/2} = e^{-\mathbf{t}^\top\mathbf{\Sigma t}/2}.$$

□

Theorem 6.12 *Let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ be normal random vector with covariance matrix $\mathbf{\Sigma}$. Then the following three statements are equivalent.*

1. *Random variables $X_1, \ldots, X_n$ are independent.*
2. *Random variables $X_1, \ldots, X_n$ are uncorrelated.*
3. *Matrix $\mathbf{\Sigma}$ is diagonal.*

Proof.

(1. $\Rightarrow$ 2.) is obvious.

(2. $\Rightarrow$ 3.) is obvious.

(3. $\Rightarrow$ 1.) Let $\mathbf{\Sigma} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Since

$$\phi(\mathbf{t}) = e^{i\mu^\top\mathbf{t}-\mathbf{t}^\top\mathbf{\Sigma t}/2} = e^{i\mu^\top\mathbf{t}-\sum_{k=1}^n\sigma_k^2t_k^2/2} = \prod_{k=1}^n e^{i\mu_k t_k-\sigma_k^2t_k^2/2}.$$

[4]that is, $\mathbf{U}^{-1} = \mathbf{U}^\top$

by Theorem 6.10 we get 1. $\square$

Theorem 6.13 *Let* $\mathbf{X} = (X_1, \ldots, X_n)^\top$ *be a normal random vector with mean* μ *and covariance matrix* $\mathbf{\Sigma}$*. Assume that* $\det(\mathbf{\Sigma}) \neq 0$ *then the density of* $\mathbf{X}$ *is given by*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \det(\mathbf{\Sigma})^{1/2}} e^{-(\mathbf{x}-\mu)^\top \mathbf{\Sigma^{-1}} (\mathbf{x}-\mu)/\mathbf{2}}. \tag{6.5}$$

Proof. Without loss of generality assume that means are zeros.

Case 1. If $\mathbf{\Sigma}$ is diagonal, that is, $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and $\sigma_k^2 > 0$, then $X_1, \ldots, X_n$ are independent with means 0 and variances σ_k^2. As a result,

$$f(\mathbf{x}) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-x_k^2/2\sigma_k^2} = \frac{1}{(2\pi)^{n/2} \det(\mathbf{\Sigma})^{1/2}} e^{-(\mathbf{x}-\mu)^\top \mathbf{\Sigma^{-1}} (\mathbf{x}-\mu)/2}.$$

Case 2. If $\mathbf{\Sigma}$ is not diagonal (but positive-semidefinite, symmetric and invertable) then there exists an orthogonal matrix $\mathbf{U}$ such that $\mathbf{U\Sigma U}^\top$ is diagonal. Let $\mathbf{Y}$ be the normal random vector that corresponds to matrix $\mathbf{U\Sigma U}^\top$. Let $\mathbf{X} = \mathbf{U}^\top \mathbf{Y}$, then $\mathbf{X}$ is a normal vector with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. Now, for a Borel set $A \in \mathbb{R}^n$ with help of substitution $\mathbf{x} = \mathbf{U}^\top \mathbf{y}, \mathbf{y} = \mathbf{U}\mathbf{x}$ we obtain

$$\begin{aligned}
\mathbf{P}(\mathbf{X} \in A) &= \mathbf{P}(\mathbf{Y} \in \mathbf{U}A) \\
&= \int_{\mathbf{U}A} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \\
&= \frac{1}{(2\pi)^{n/2} \det(\mathbf{U\Sigma U}^\top)^{1/2}} \int_{\mathbf{U}A} e^{-\mathbf{y}^\top (\mathbf{U\Sigma U}^\top)^{-\mathbf{1}} \mathbf{y}/2} d\mathbf{y} \\
&= \frac{1}{(2\pi)^{n/2} \det(\mathbf{\Sigma})^{1/2}} \int_{A} e^{-\mathbf{x}^\top \mathbf{U}^\top (\mathbf{U\Sigma U}^\top)^{-\mathbf{1}} \mathbf{U}\mathbf{x}/2} |\det(\mathbf{U})| d\mathbf{x}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(2\pi)^{n/2}\det(\mathbf{\Sigma})^{1/2}} \int_A e^{-\mathbf{x}^\top \mathbf{U}^\top \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^\top \mathbf{U}\mathbf{x}/2} d\mathbf{x} \\
&= \frac{1}{(2\pi)^{n/2}\det(\mathbf{\Sigma})^{1/2}} \int_A e^{-\mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x}/2} d\mathbf{x}.
\end{aligned}
$$

Note that the absolute value of Jacobian determinant of $\mathbf{U}$ is 1, and $\det(\mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top) = \det(\mathbf{\Sigma})$.

$\square$

Exercise 6.16 Show that for any vector μ and positive-semidefinite symmetric matrix Σ with $\det(\mathbf{\Sigma}) \neq 0$ there exists normal vector $\mathbf{X}$ with pdf given by (6.5).

Exercise 6.17 Let $(X_1, X_2)^\top$ be two-dimensional normal vector with mean $(\mu_1, \mu_2)^\top$ and covariance matrix

$$
\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.
$$

Find the joint pdf.

Exercise 6.18 Let X be a normal random variable with mean 0 and variance 1, and ξ be a Bernoulli random variable with $\mathbf{P}(\xi = 1) = \mathbf{P}(\xi = 0) = 1/2$. Assume that X and ξ are independent. Let us define random variable Y by the following rule:

$$
Y = \begin{cases} -X, & \text{if } \xi = 1, \\ X, & \text{if } \xi = 0. \end{cases}
$$

a) Find characteristic function of Y to prove that it has the standard normal distribution.

b) Calculate $Cov(X, Y)$.

c) Find a linear combination $aX + bY$ that is not normally distributed to prove that $(X, Y)^\top$ is not a normal vector.

d) Find a Borel set B such that $\mathbf{P}(X \in B \cap Y \in B) \neq \mathbf{P}(X \in B)\mathbf{P}(Y \in B)$ to prove that X and Y are not independent.

Chapter 7

Convergence

7.1 Different Types of Convergence

Let $\{X, X_n\}_{n\geq 1}$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbf{P})$.

Definition 7.1 *We say that* $\{X_n\}$ *converges to random variable* X almost surely *(a.s. or with probability 1), written* $X_n \xrightarrow{a.s.} X$, *if*

$$\mathbf{P}[\omega : X_n(\omega) \to X(\omega)] = 1.$$

Definition 7.2 *We say that* $\{X_n\}$ *converges to random variable* X in probability, *written* $X_n \xrightarrow{\mathbf{P}} X$, *if for any* $\epsilon > 0$

$$\mathbf{P}[|X_n - X| > \epsilon] \to 0.$$

Definition 7.3 *We say that* $\{X_n\}$ *converges to random variable* X in L_p, $p > 0$, *written* $X_n \xrightarrow{L_p} X$, *if*

$$\mathbf{E}|X_n - X|^p \to 0.$$

Definition 7.4 *We say that* $\{X_n\}$ *with cdf* F_n *converges to random variable* X *with cdf* F in distribution, *written* $X_n \xrightarrow{d} X$, *if at any continuity point* x *of cdf* F

$$F_n(x) \to F(x).$$

Theorem 7.1 (An Iff Condition for a.s. Convergence) *Let* $\{X_n\}_{n\geq 1}$ *be a sequence of random variables. Then*

$$X_n \xrightarrow{a.s.} X$$

iff

$$\mathbf{P}\left[\sup_{k\geq n} |X_k - X| \geq \epsilon\right] \to 0, \ \textit{as } n \to \infty,$$

for every $\epsilon > 0$.

Proof. Note that

$X_n \xrightarrow{a.s.} X$

$\Leftrightarrow \mathbf{P}(\{|X_n - X| > \epsilon\} \text{ i.o.}) = 0$ for any $\epsilon > 0$

$\Leftrightarrow \mathbf{P}(\bigcap_{n\geq 1} \bigcup_{k\geq n} \{|X_k - X| > \epsilon\}) = 0$ for any $\epsilon > 0$

$\Leftrightarrow \lim_n \mathbf{P}(\bigcup_{k\geq n} \{|X_k - X| > \epsilon\}) = 0$ for any $\epsilon > 0$

$\Leftrightarrow \lim_n \mathbf{P}(\sup_{k\geq n} |X_k - X| > \epsilon) = 0$ for any $\epsilon > 0$. $\square$

Proposition 7.1 *Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables. If*

$$X_n \xrightarrow{\mathbf{P}} X,$$

then there exists subsequence n_k such that

$$X_{n_k} \xrightarrow{a.s.} X.$$

Proof. Let $n_1 = 1$, and for $k \geq 2$ define

$$n_k = \inf\{n : n > n_{k-1}, \mathbf{P}(|X_{n_k} - X| > 1/k) < 1/2^k\}.$$

It is possible, because $\mathbf{P}(|X_n - X| > \epsilon) \to 0$ for any $\epsilon > 0$. Since

$$\sum_k \mathbf{P}(|X_{n_k} - X| > 1/k) < \infty$$

by Borel-Cantelli Lemma we have that $\mathbf{P}(\{|X_{n_k} - X| > 1/k\} \text{ i.o.}) = 0$, that is, $X_{n_k} \xrightarrow{a.s.} X$. $\square$

Theorem 7.2 (Convergence Graph) *The following implications are true.*

$$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{\mathbf{P}} X, \tag{7.1}$$

$$X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{\mathbf{P}} X, \quad p > 0 \tag{7.2}$$

$$X_n \xrightarrow{\mathbf{P}} X \Rightarrow X_n \xrightarrow{d} X. \tag{7.3}$$

Proof. The statement (7.1) follows from Theorem 7.1. The L_p version of Chebyshev's inequality gives us (7.2). So, we really need to work out only the last implication. Let F_n denote the cdf of X_n, and let F denote the cdf of X. Consider x which is a continuity point of F. Note that for any $\epsilon > 0$

$$\begin{aligned}
\mathbf{P}(X_n \leq x) &= \mathbf{P}(X_n \leq x, |X_n - X| > \epsilon) + \mathbf{P}(X_n \leq x, |X_n - X| \leq \epsilon) \\
&\leq \mathbf{P}(|X_n - X| > \epsilon) + \mathbf{P}(X \leq x + \epsilon)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{P}(X \leq x - \epsilon) &= \mathbf{P}(X \leq x - \epsilon, |X_n - X| > \epsilon) + \mathbf{P}(X \leq x - \epsilon, |X_n - X| \leq \epsilon) \\
&\leq \mathbf{P}(|X_n - X| > \epsilon) + \mathbf{P}(X_n \leq x)
\end{aligned}$$

That is, for any $\epsilon > 0$ we have

$$F(x - \epsilon) - \mathbf{P}(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x + \epsilon) + \mathbf{P}(|X_n - X| > \epsilon).$$

Taking limit with respect to n gives us

$$F(x - \epsilon) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x + \epsilon).$$

Since x is a continuity point of F sending $\epsilon \to 0$ finishes the proof. $\square$

Theorem 7.3 *Let F_n denote the cdf of X_n, and let F denote the cdf of X. If $X_n \xrightarrow{d} X$ and F is continuous, then F_n converges to F uniformly.*

Proof. Fix $\epsilon > 0$. Since F is continuous we can find $-\infty < x_1 < \cdots < x_k < \infty$ such that

$$F(x_1) < \epsilon/2, \quad 1-F(x_k) < \epsilon/2, \quad \text{and } F(x_{i+1})-F(x_i) < \epsilon/2 \text{ for } i = 1, \ldots, k-1.$$

For $x_i \leq x \leq x_{i+1}$ $i = 1, \ldots, k-1$ and for sufficiently large n we have

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(x_{i+1}) - F(x_i) + F(x_{i+1}) - F(x_{i+1}) \\ &= F_n(x_{i+1}) - F(x_{i+1}) + F(x_{i+1}) - F(x_i) \\ &\leq \epsilon, \end{aligned}$$

and

$$\begin{aligned} F_n(x) - F(x) &\geq F_n(x_i) - F(x_{i+1}) + F(x_i) - F(x_i) \\ &= F_n(x_i) - F(x_i) + F(x_i) - F(x_{i+1}) \\ &\geq -\epsilon. \end{aligned}$$

That is, there exists N_i such that for all $n > N_i$

$$\sup_{x_i \leq x \leq x_{i+1}} |F_n(x) - F(x)| \leq \epsilon.$$

Similarly, if $x < x_1$

$$F_n(x) - F(x) \quad \leq \quad F_n(x_1) - F(x_1) + F(x_1)$$

$$\leq \epsilon,$$

and

$$\begin{aligned} F_n(x) - F(x) &\geq -F(x_1) \\ &\geq -\epsilon. \end{aligned}$$

Thus, there exists N_0 such that for all $n > N_0$

$$\sup_{x \leq x_1} |F_n(x) - F(x)| \leq \epsilon.$$

The same can be done for $x > x_k$. Therefore, for $n > \max\{N_0, N_1, \ldots, N_k\}$ we have

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \epsilon.$$

$\square$

Exercise 7.1 Give an example for each implication that shows that it is not invertible.

Exercise 7.2 Show that

(a) if $0 \leq X_n \leq Y_n$ and $Y_n \xrightarrow{\mathbf{P}} 0$, then $X_n \xrightarrow{\mathbf{P}} 0$,

(b) if $X_n \xrightarrow{\mathbf{P}} 0$ and $\mathbf{P}(|Y_n| > M) \to 0$ for some $M > 0$, then $X_n Y_n \xrightarrow{\mathbf{P}} 0$,

(c) if $X_n \xrightarrow{\mathbf{P}} X$ and $Y_n \xrightarrow{\mathbf{P}} Y$, then $X_n + Y_n \xrightarrow{\mathbf{P}} X + Y$,

(d) if $X_n \xrightarrow{\mathbf{P}} X$ and $Y_n \xrightarrow{\mathbf{P}} Y$, then $X_n Y_n \xrightarrow{\mathbf{P}} XY$.

Exercise 7.3 Show that if $X_n \xrightarrow{d} 0$, then $X_n \xrightarrow{\mathbf{P}} 0$.

7.2 Weak Convergence of Probability Measures

Definition 7.5 *Let* $\mathbf{P}, \mathbf{P}_n$ *be probability measures on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. The sequence of probability measures* $\{\mathbf{P}_n\}$ converges weakly *to probability measure* $\mathbf{P}$*, written* $\mathbf{P}_n \Longrightarrow \mathbf{P}$*, if*

$$\mathbf{P}_n(A) \to \mathbf{P}(A)$$

for every set $A = (-\infty, x]$ *with* $\mathbf{P}(\{x\}) = 0$*.*

For any probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we can introduce the corresponding cdf (by $F_n(x) = \mathbf{P}_n((-\infty, x])$ and $F(x) = \mathbf{P}((-\infty, x]))$, so we will also use notation $F_n \Longrightarrow F$. Thus convergence in distribution and weak convergence of probability measures are only different expression of the same fact. But the new definition is a bit more flexible, and it can be easily extended to probability measures on, say, metric spaces.

Exercise 7.4 Show that if $F_n \Longrightarrow F$ and $F_n \Longrightarrow G$, then $F = G$.

Exercise 7.5 Show that if $\lim_n F_n(x) = F(x)$ for x in a set D dense in $\mathbb{R}$, then and $F_n \Longrightarrow F$.

Lemma 7.1 (Quantile Function) *Let* F *be a cdf. For* $0 < \omega < 1$ *define set*

$$A(\omega) = \{x : \omega \leq F(x)\},$$

and

$$X(\omega) = \inf A(\omega).$$

Then

(a) $A(\omega)$ is a closed on the left interval stretching to ∞,

(b) $X(\omega)$ is non-decreasing,

(c) $\omega \leq F(x)$ if and only if $X(\omega) \leq x$ or, equivalently, by taking negation of the first statement we also have $\omega > F(x)$ if and only if $X(\omega) > x$.

Proof.

(a) Assume that $x \in A(\omega)$ and $x < x'$, then $\omega \leq F(x) \leq F(x')$, that is, $x' \in A(\omega)$. That is, $A(\omega)$ is an interval stretching to ∞ Now, let us show that $A(\omega)$ is a closed interval. If $x_n \to x$, $x_n > x$, and $x_n \in A(\omega)$, then

$$\omega \leq F(x_n) \downarrow F(x),$$

therefore, $\omega \leq F(x)$ and $x \in A(\omega)$. In particular, this means that $X(\omega) \in A(\omega)$ and $\omega \leq F(X(\omega))$.

(b) Since for $\omega \leq \omega'$ we have $A(\omega') \subset A(\omega)$ we get that $X(\omega)$ is non-decreasing.

(c) If $x < X(\omega) = \inf A(\omega)$, then $x \notin A(\omega)$, and $\omega > F(x)$. If $\inf A(\omega) = X(\omega) \leq x$, then $x \in A(\omega)$, and, therefore, $\omega \leq F(x)$. $\square$

Exercise 7.6 Consider strictly increasing continuous functions $f_n, f : [0,1] \mapsto [0,1]$, with $f_n(0) = f(0) = 0$ and $f_n(1) = f(1) = 1$. Assume that $f_n(x) \to f(x)$ for every $x \in [0,1]$. Show that inverse functions converge to the inverse of f, that is, $f_n^{-1}(y) \to f^{-1}(y)$ for every $0 \leq y \leq 1$.

Theorem 7.4 (Skorohod's Theorem) *Let $\mathbf{P}, \mathbf{P}_n$ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\mathbf{P}_n \Longrightarrow \mathbf{P}$. Then there exist random variables X_n and X on a common probability space $(\Omega, \mathcal{F}, \mathbf{Q})$ such that X_n has distribution $\mathbf{P}_n$, X has distribution $\mathbf{P}$, and for every $\omega \in \Omega$ $X_n(\omega) \to X(\omega)$.*

Proof. Let us take $\Omega = (0,1)$, $\mathcal{F} = \mathcal{B}((0,1))$, and $\mathbf{Q} = \lambda$, the Lebesgue measure. Consider the distribution functions F_n and F corresponding $\mathbf{P}_n$ and $\mathbf{P}$. For every $\omega \in (0,1)$ we define

$$X_n(\omega) = \inf\{x : \omega \leq F_n(x)\}$$

and

$$X(\omega) = \inf\{x : \omega \leq F(x)\}.$$

By Lemma 7.1, $\omega \leq F(x)$ if and only if $X(\omega) \leq x$. Therefore,

$$\mathbf{Q}[\omega : X(\omega) \leq x] = \mathbf{Q}[\omega : \omega \leq F(x)] = \lambda[(0, F(x))] = F(x).$$

That is, F is the cdf of X. In the same fashion, we can show that X_n has distribution F_n.

Let us take $\omega \in (0,1)$. For any $\epsilon > 0$ we can find x with $\mathbf{P}(\{x\}) = 0$ such that

$$X(\omega) - \epsilon < x < X(\omega).$$

Since $x < X(\omega) \Leftrightarrow F(x) < \omega$, and $F_n(x) \to F(x)$ we get that for all sufficiently large n $F_n(x) < \omega$ or $x < X_n(\omega)$. So, we have that

$$X(\omega) - \epsilon < x < X_n(\omega).$$

Taking $n \to \infty$, because of an arbitrary choice of ϵ we obtain

$$X(\omega) \leq \liminf_n X_n(\omega).$$

Consider now ω' such that $\omega < \omega'$. For any $\epsilon > 0$ we can find x' with $\mathbf{P}(\{x'\}) = 0$ such that

$$X(\omega') < x' < X(\omega') + \epsilon.$$

Since $X(\omega') \leq x' \Leftrightarrow \omega' \leq F(x')$, and $F_n(x') \to F(x')$ we get that for all sufficiently large n $\omega \leq F_n(x')$ or $X_n(\omega) \leq x'$. Note that in this direction we have the equivalency of non-strict inequalities; that is the reason why we need the extra gap.

Thus we get that

$$X_n(\omega) \leq x' < X(\omega') + \epsilon,$$

and, therefore,

$$\limsup_n X_n(\omega) \leq X(\omega').$$

Therefore, for any $0 < \omega < \omega' < 1$ we get that

$$X(\omega) \leq \liminf_n X_n(\omega) \leq \limsup_n X_n(\omega) \leq X(\omega').$$

So, if ω is a continuity point of X, then $X_n(\omega) \to X(\omega)$. Since X is nondecreasing, it has at most countable numbers of points of discontinuity (that is, the Lebesgue measure of this set is 0). Let us redefine X and X_n at those points by $X(\omega) = X_n(\omega) = 0$. This will not change the distributions of X and X_n, and therefore, the construction is finished. $\square$

Exercise 7.7 Let F be a cdf given by

$$F(x) = \begin{cases} 0, & x < 0 \\ x/3, & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 1.5 \\ x-1 & 1.5 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Construct

$$Q(p) = \inf\{x : p \leq F(x)\},$$

where $0 < p < 1$.

Theorem 7.5 (Mapping Theorem) *Suppose $h : \mathbb{R} \to \mathbb{R}$ is measurable, and the set D_h of its discontinuity is measurable as well. If $\mathbf{P}_n \Longrightarrow \mathbf{P}$ and $\mathbf{P}(D_h) = 0$, then $\mathbf{P}_n h^{-1} \Longrightarrow \mathbf{P}h^{-1}$.*[1]

Proof. Let us consider random variables X_n and X constructed in Theorem 7.4.

[1] $\mathbf{P}h^{-1}$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by $\mathbf{P}h^{-1}(A) = \mathbf{P}(h^{-1}A)$.

If $X(\omega) \notin D_h$, then for such ω we have $h(X_n(\omega)) \to h(X(\omega))$. Since

$$\mathbf{Q}[X(\omega) \in D_h] = \mathbf{P}[D_h] = 0,$$

we, in fact, have

$$h(X_n(\omega)) \xrightarrow{a.s.} h(X(\omega))$$

with respect to measure $\mathbf{Q}$. By Theorem 7.2 we get that $h(X_n) \xrightarrow{d} h(X)$. But $\mathbf{Q}[h(X) \in A] = \mathbf{Q}[X \in h^{-1}A] = \mathbf{P}[h^{-1}A]$, that is, $h(X)$ has distribution $\mathbf{P}h^{-1}$. Similarly, $h(X_n)$ has distribution $\mathbf{P}_n h^{-1}$ $\square$

Corollary 7.1 *(in terms of random variables) If* $X_n \xrightarrow{d} X$ *and* $\mathbf{P}[X \in D_h] = 0$*, then* $h(X_n) \xrightarrow{d} h(X)$*.*

Proposition 7.2 *If* $X_n \xrightarrow{d} X$ *and* X_n *are uniformly integrable, then* X *is integrable and* $\mathbf{E}X_n \to \mathbf{E}X$*.*

Proof. Exercise. Hint: use the Skorohod's theorem and note that the uniform integrability (see Definition 5.5) can be thought as a statement in terms of cdfs.

$\square$

Theorem 7.6 (Portmanteau Theorem) *The following four conditions are equivalent.*

1. $\mathbf{P}_n \Longrightarrow \mathbf{P}$.
2. $\int_{\mathbb{R}} f d\mathbf{P}_n \to \int_{\mathbb{R}} f d\mathbf{P}$ *for all bounded, continuous real* f *on* $\mathbb{R}$.
3. $\int_{\mathbb{R}} f d\mathbf{P}_n \to \int_{\mathbb{R}} f d\mathbf{P}$ *for all bounded, uniformly continuous real* f *on* $\mathbb{R}$.

4. $\mathbf{P}_n(A) \to \mathbf{P}(A)$ *for every Borel set* A *with* $\mathbf{P}(\partial A) = 0$.[2]

Proof.

(1. $\Rightarrow$ 2.) Let us again consider random variables X_n and X constructed in Theorem 7.4. Let f be a bounded, continuous function. Then $f(X_n) \to f(X)$ $\mathbf{Q}$-a.s. By change of variables theorem and the dominated convergence theorem we have

$$\int_{\mathbb{R}} f d\mathbf{P}_n = \mathbf{E}_{\mathbf{Q}}[f(X_n)] \to \mathbf{E}_{\mathbf{Q}}[f(X)] = \int_{\mathbb{R}} f d\mathbf{P}.$$

(2. $\Rightarrow$ 3.) Obvious.

(1. $\Rightarrow$ 4.) Let A be Borel set A with $\mathbf{P}(\partial A) = 0$. Consider function $f(x) = 1_A(x)$. It is a bounded function, and the set D_f of its discontinuities is equal to ∂A. Since $\mathbf{P}(\partial A) = 0$ we get that $f(X_n) \to f(X)$ $\mathbf{Q}$-a.s., therefore, again by change of variables and the dominated convergence theorem we get

$$\mathbf{P}_n(A) = \mathbf{E}_{\mathbf{Q}}[f(X_n)] \to \mathbf{E}_{\mathbf{Q}}[f(X)] = \mathbf{P}(A).$$

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

(3. $\Rightarrow$ 1.) Consider the distribution functions F_n and F that correspond to $\mathbf{P}_n$ and $\mathbf{P}$. For any $x < y$ let us introduce the following bounded, uniformly

[2] Here the boundary $\partial A = \text{closure}(A) \cap \text{closure}(A^c)$.

continuous function:

$$f(t) = \begin{cases} 1 & t \leq x \\ (y-t)/(y-x) & x < t < y \\ 0 & t \geq y \end{cases} .$$

Since

$$F_n(x) = \int_{(-\infty,x]} d\mathbf{P}_n \leq \int_{\mathbb{R}} f d\mathbf{P}_n$$

and

$$\int_{\mathbb{R}} f d\mathbf{P} \leq \int_{(-\infty,y]} d\mathbf{P} = F(y),$$

it follows from 3. that

$$\limsup_n F_n(x) \leq F(y),$$

and sending $y \downarrow x$ we get that (F is right-continuous)

$$\limsup_n F_n(x) \leq F(x).$$

In similar fashion for $u < x$ we can get that

$$F(u) \leq \liminf_n F_n(x)$$

and, as a consequence

$$F(x-) \leq \liminf_n F_n(x).$$

Thus, we have

$$F(x-) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x),$$

which means convergence at continuity points of F.

Using these five implications we can go from any statement to another one, so we are done! $\square$

Theorem 7.7 (Slutsky's Theorem) *Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} 0$, then $X_n + Y_n \xrightarrow{d} X$.*

Proof. Let f be a bounded, uniformly continuous function. By Portmanteau theorem, it is enough to show that $\mathbf{E}f(X_n + Y_n) \to \mathbf{E}f(X)$. For any $\delta > 0$ we get that

$$\begin{aligned}
|\mathbf{E}f(X_n + Y_n) - \mathbf{E}f(X)| &\leq |\mathbf{E}f(X_n + Y_n) - \mathbf{E}f(X_n)| + |\mathbf{E}f(X_n) - \mathbf{E}f(X)| \\
&\leq \mathbf{E}|f(X_n + Y_n) - f(X_n)|1_{|Y_n| \leq \delta} \\
&\quad + \mathbf{E}|f(X_n + Y_n) - f(X_n)|1_{|Y_n| > \delta} \\
&\quad + |\mathbf{E}f(X_n) - \mathbf{E}f(X)| \\
&\leq \sup_{|x-y| \leq \delta} |f(x) - f(y)| \\
&\quad + 2 \sup_x |f(x)| \mathbf{P}(|Y_n| > \delta) \\
&\quad + |\mathbf{E}f(X_n) - \mathbf{E}f(X)|.
\end{aligned}$$

The first term is small for small $\delta > 0$ because of the uniform continuity of f, the second one is small for large n because f is bounded and $Y_n \xrightarrow{\mathbf{P}} 0$, the last one is small because $X_n \xrightarrow{d} X$ and $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$.

□

7.3 Weak Convergence and Pointwise Convergence of CFs

Theorem 7.8 (Helly's selection Theorem) *For every sequence of cdfs $\{F_N\}_{N\geq 1}$ there exists a subsequence of cdfs, $\{F_n\}_{n\geq 1}$, and a nondecreasing, right-continuous function $F : \mathbb{R} \to [0,1]$ such that $F_n(x) \to F(x)$ at continuity points of F.*

Proof. Let $\mathbb{Q} = \{q_1, q_2, q_3, \dots\}$ be the ordered set of all rational numbers. Since $\{F_N(q_1)\}_{N\geq 1} \subseteq [0,1]$ there exists a sequence $F_{1n}(q_1)$ such that $F_{1n}(q_1)$ converges to a number, let us call it $Q(q_1)$. Since $\{F_{1n}(q_2)\}_{n\geq 1} \subseteq [0,1]$ there exists a *further subsequence* $F_{2n}(q_2)$ such that $F_{2n}(q_2)$ converges to a number that we denote $Q(q_2)$; and so on.

Let us consider sequence of cdfs $F_n = F_{nn}$.[3] By construction, for every $q \in \mathbb{Q}$ we have

$$F_n(q) \to Q(q).$$

Note that Q is non-decreasing function on $\mathbb{Q}$ with values in $[0,1]$.

Now, for all $x \in \mathbb{R}$ we define

$$F(x) = \inf\{Q(q) : x < q\}.$$

[3] It is so-called Cantor diagonal sequence.

It is easy to see that F is non-decreasing.

Next, we show that F is right-continuous. Let $x_k \downarrow x$ and $d = \lim_k F(x_k)$. Since for any x_k we have $F(x) \leq F(x_k)$ it follows that $F(x) \leq d$. Assume that $F(x) < d$. Then there exists $x < q \in \mathbb{Q}$ such that $Q(q) < d$. For large k we have $x < x_k < q$, and, therefore,

$$F(x_k) \leq Q(q) < d.$$

That is,

$$d = \lim_k F(x_k) \leq Q(q) < d,$$

a contradiction. Therefore, $F(x) = \lim_k F(x_k)$.

Finally, let x be a continuity point of F. Take any $y < x$. Consider two sequences $r_k, q_k \in \mathbb{Q}$ such that

$$y < r_k < x < q_k, \quad r_k \downarrow y, \quad \text{ and } q_k \downarrow x.$$

Since F_n is nondecreasing we have

$$F_n(r_k) \leq F_n(x) \leq F_n(q_k).$$

Taking $n \to \infty$ we get

$$Q(r_k) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq Q(q_k).$$

Since

$$\lim_k Q(r_k) = \inf_k Q(r_k) = \inf\{Q(q) : y < q\} = F(y),$$

and

$$\lim_k Q(q_k) = \inf_k Q(q_k) = \inf\{Q(q) : x < q\} = F(x),$$

we finally obtain that for any $y < x$

$$F(y) \le \liminf_n F_n(x) \le \limsup_n F_n(x) \le F(x).$$

Because of arbitrary choice of y we get that $F_n(x) \to F(x)$ at every point of continuity x of F. $\square$

Exercise 7.8 Find an example of a sequence of cdfs $\{F_n\}_{n\ge1}$ and a nondecreasing, right-continuous function F such that

(a) $F_n(x) \to F(x)$ at every continuity point x of F,

(b) $0 < F(+\infty) - F(-\infty) < 1$.

Definition 7.6 *A family of probability measures* $\{\mathbf{P}_t\}_{t\in T}$ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is said to be* relatively compact *if every sequence of measures from the family contains a subsequence that converges weakly to a probability measure. A family of cdfs* $\{F_t\}_{t\in T}$ *is* relatively compact *if the corresponding family of probability measures is relatively compact.*

We use word "relatively" because the limit need not belong to the original family.

Definition 7.7 *A family of probability measures* $\{\mathbf{P}_t\}_{t\in T}$ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is said to be* tight *if for each* $\epsilon > 0$ *there exists a compact subset* I *of* $\mathbb{R}$ *such that*

$$\inf_t \mathbf{P}_t(I) > 1 - \epsilon.$$

A family of cdfs $\{F_t\}_{t\in T}$ *is* tight *if the corresponding family of probability measures is tight.*

Note that in the definition (because we consider $\mathbb{R}$) the compact set I can be substituted by a finite interval.

Exercise 7.9 Consider a sequence of random variables $\{X_n\}_{n\geq 1}$ such that $X_n \geq 0$ and $\mathbf{E}X_n = 1$. Then $\mathbf{P}_n$ defined by

$$\mathbf{P}_n(A) = \mathbf{E}X_n 1_{X_n \in A}, A \in \mathcal{B}(\mathbb{R})$$

are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Show that the following two statements are equivalent.

(a) The sequence of random variables $\{X_n\}_{n\geq 1}$ is uniformly integrable.

(b) The sequence of probability measures $\{\mathbf{P}_n\}_{n\geq 1}$ is tight.

Theorem 7.9 (Prokhorov's theorem) *A family of probability measures* $\{\mathbf{P}_t\}_{t\in T}$ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is relatively compact if and only if it is tight.*

Proof. Necessity. Assume that the family $\{\mathbf{P}_t\}_{t\in T}$ is relatively compact but not tight. Since the family is not tight, there exists $\epsilon > 0$ such that for every interval $I_n = [-n, n]$ we can find $\mathbf{P}_n$ with $\mathbf{P}_n(I_n) \leq 1 - \epsilon$. But because the family is relatively compact we can find a subsequence $\mathbf{P}_{n_k}$ and a probability measure $\mathbf{P}$ such that $\mathbf{P}_{n_k} \Longrightarrow \mathbf{P}$. Let $[a, b]$ be a continuity set of $\mathbf{P}$ such that $\mathbf{P}([a, b]) > 1 - \epsilon$. Then by Portmanteau theorem we have

$$1-\epsilon < \mathbf{P}([a,b]) = \lim_k \mathbf{P}_{n_k}([a,b]) = \lim_k \mathbf{P}_{n_k}([a,b]\cap I_{n_k}) \leq \limsup_k \mathbf{P}_{n_k}(I_{n_k}) \leq 1-\epsilon,$$

a contradiction.

Sufficiency. Let $\{\mathbf{P}_n\}_{n\geq 1}$ be a tight sequence of probability measures from the family $\{\mathbf{P}_t\}_{t\in T}$. Let $\{F_n\}_{n\geq 1}$ be the corresponding sequence of cdfs. By Helly's selection theorem there exists subsequence F_{n_k} and a nondecreasing, right-continuous function F such that $F_{n_k}(x) \to F(x)$ at continuity points of F. Let us show that F is, in fact, a proper cdf. Fix $\epsilon > 0$. Since the family $\{\mathbf{P}_t\}_{t\in T}$ is tight we can find interval $[a, b]$ such that

$$\mathbf{P}_{n_k}([a,b]) \geq 1 - \epsilon.$$

Let $[a, b] \subset (a', b']$ and $\{a', b'\}$ is a continuity set of F. Then

$$1 - \epsilon \leq \mathbf{P}_{n_k}([a,b]) \leq \mathbf{P}_{n_k}((a',b']) = F_{n_k}(b') - F_{n_k}(a') \to F(b') - F(a'),$$

that is, $F(+\infty) - F(-\infty) = 1$, and, together with $0 \leq F(-\infty) \leq F(+\infty) \leq 1$, it gives us that $F_{n_k} \Longrightarrow F$, and, of course, $\mathbf{P}_{n_k} \Longrightarrow \mathbf{P}$. $\square$

Corollary 7.2 *Let $\{\mathbf{P}_n\}_{n\geq 1}$ be a tight sequence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Suppose that every weakly convergent subsequence converges to the* *same probability measure* $\mathbf{P}$*. Then* $\mathbf{P}_n \Longrightarrow \mathbf{P}$*.*

Proof. Suppose that $\mathbf{P}_n \Longrightarrow \mathbf{P}$ is not true. It means that there exists a point of continuity of $\mathbf{P}$, x, such that $\mathbf{P}_n[(-\infty, x]]$ do not converge to $\mathbf{P}[(-\infty, x]]$, that there exists $\epsilon > 0$ such that $|\mathbf{P}_{n_k}[(-\infty, x]] - \mathbf{P}[(-\infty, x]]| > \epsilon$ for some sequence $\{n_k\}$. By the Prokhorov's theorem, there is a further subsequence that converges weakly to some probability measure, and by the corollary assumption it must converge to $\mathbf{P}$, but no subsequence of $\mathbf{P}_{n_k}$ can converge weakly to $\mathbf{P}$. $\square$

Theorem 7.10 (Continuity Theorem) *Let* $\{\mathbf{P}, \mathbf{P}_n\}_{n\geq 1}$ *be probability measures on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *with characteristic functions* $\{\phi, \phi_n\}_{n\geq 1}$*. A necessary and sufficient condition for* $\mathbf{P}_n \Longrightarrow \mathbf{P}$ *is that* $\phi_n(t) \to \phi(t)$ *for every* t*.*

Let us first prove the following lemma.

Lemma 7.2 *Let* ϕ *be the characteristic function of a probability measure* $\mathbf{P}$*. Then for any* $A > 0$ *we have*

$$\mathbf{P}[(-A, A)] \geq 1 - \frac{\sqrt{A}}{2} \int_{|t|<1/\sqrt{A}} |1 - \phi(t)| dt - 1/\sqrt{A}.$$

Proof. For any $\epsilon > 0$ by Fubini's theorem we get

$$\begin{aligned} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} \phi(t) dt &= \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} \int_{\mathbb{R}} e^{itx} \mathbf{P}(dx) dt \\ &= \frac{1}{2\epsilon} \int_{\mathbb{R}} \int_{-\epsilon}^{\epsilon} e^{itx} dt \mathbf{P}(dx) \\ &= \frac{1}{2\epsilon} \int_{\mathbb{R}} \frac{2\sin(\epsilon x)}{x} \mathbf{P}(dx) \\ &= \int_{\mathbb{R}} \frac{\sin(\epsilon x)}{\epsilon x} \mathbf{P}(dx) \end{aligned}$$

$$\begin{aligned}
&\leq \int_{|x|<1/\epsilon^2} \mathbf{P}(dx) + \int_{|x|\geq 1/\epsilon^2} \frac{1}{\epsilon|x|}\mathbf{P}(dx) \\
&\leq \mathbf{P}[(-1/\epsilon^2, 1/\epsilon^2)] + \epsilon.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\mathbf{P}[(-1/\epsilon^2, 1/\epsilon^2)]dt &\geq 1 + \frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon} (\phi(t)-1)dt - \epsilon \\
&\geq 1 - \frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon} |\phi(t)-1)|dt - \epsilon.
\end{aligned}$$

Taking $\epsilon = 1/\sqrt{A}$ we get the result. $\square$

Theorem Proof.

Necessity. Since $\cos(tx)$ and $\sin(tx)$ are bounded continuous functions, the necessity immediately follows from the Portmanteau theorem.

Sufficiency. By Lemma 7.2 we have

$$\begin{aligned}
\mathbf{P}_n[(-A, A)] &\geq 1 - \frac{\sqrt{A}}{2}\int_{|t|<1/\sqrt{A}} |1-\phi_n(t)|dt - 1/\sqrt{A} \\
&\geq 1 - \frac{\sqrt{A}}{2}\int_{|t|<1/\sqrt{A}} |1-\phi(t)|dt \\
&\quad - \frac{\sqrt{A}}{2}\int_{|t|<1/\sqrt{A}} |\phi(t)-\phi_n(t)|dt - 1/\sqrt{A}. \\
&\geq 1 - \sup_{|t|<1/\sqrt{A}} |1-\phi(t)| \\
&\quad - \frac{\sqrt{A}}{2}\int_{|t|<1/\sqrt{A}} |\phi(t)-\phi_n(t)|dt - 1/\sqrt{A}.
\end{aligned}$$

By continuity of ϕ at 0 and the dominated convergence theorem we find that the $\{\mathbf{P}_n\}_{n\geq 1}$ is tight. If a subsequence $\mathbf{P}_{n_k}$ converge weakly to a probability measure $\mathbf{Q}$ with CF $q(t)$, then by the necessity part $\phi_{n_k}(t)$ converges to $q(t)$. Therefore, $\phi(t) = q(t)$, and, by the uniqueness theorem (Theorem 6.5), we get $\mathbf{P} = \mathbf{Q}$. Corollary 7.2 tells us that $\mathbf{P}_n \Longrightarrow \mathbf{P}$. $\square$

Exercise 7.10 Show that

$$\left|\frac{\sin(x)}{x}\right| \leq 1.$$

Chapter 8

Limit Theorems

8.1 Weak Law of Large Numbers

Theorem 8.1 (J. Bernoulli's Law of Large Numbers, 1713) *Let* $\{X_n\}_{n\geq 1}$ *be i.i.d. Bernoulli random variables (that is,* $\mathbf{P}(X_n = 1) = p$ *and* $\mathbf{P}(X_n = 0) = 1 - p$*, where* $0 < p < 1$*). Then for any* $\epsilon > 0$

$$\mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - p\right| > \epsilon\right) \to 0.$$

Theorem 8.2 (Chebyshev's Law of Large Numbers, 1867) *Let* $\{X_n\}_{n\geq 1}$ *be independent random variables such that* $\mathbf{Var}(X_n) < C$*. Then*

$$\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mathbf{E}X_1 + \cdots + \mathbf{E}X_n}{n}\right| \xrightarrow{\mathbf{P}} 0.$$

Proof. By Chebyshev's inequality we get for any $\epsilon > 0$

$$\begin{aligned}\mathbf{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mathbf{E}X_1 + \cdots + \mathbf{E}X_n}{n}\right| > \epsilon\right) &\leq \frac{\mathbf{Var}(X_1 + \cdots + X_n)}{n^2\epsilon^2} \\ &\leq \frac{nC}{n^2\epsilon^2} \to 0.\end{aligned}$$

□

Theorem 8.3 (Khinchin's Law of Large Numbers, 1929) *Let $\{X_n\}_{n\geq 1}$ be i.i.d. random variables such that $\mathbf{E}|X_n| < \infty$ and $\mathbf{E}X_n = \mu$. Then*

$$\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \xrightarrow{\mathbf{P}} 0.$$

Proof. Consider $Y_n = X_n - \mu$. We need to show that

$$\left|\frac{Y_1 + \cdots + Y_n}{n}\right| \xrightarrow{\mathbf{P}} 0.$$

Note if $\phi_X(t)$ is a characteristic function of X_n then the characteristic function of Y_n

$$\phi_Y(t) = e^{-i\mu t}\phi_X(t).$$

By Theorem 6.6 for any fixed t

$$\phi_{(Y_1+\cdots+Y_n)/n}(t) = [\phi_Y(t/n)]^n = \left[1 + i \cdot 0 \cdot \frac{t}{n} + o(1/n)\right]^n \to 1.$$

By continuity theorem $(Y_1 + \cdots + Y_n)/n \xrightarrow{d} 0$ and, therefore, we get that

$(Y_1 + \cdots + Y_n)/n \xrightarrow{\mathbf{P}} 0.$ $\square$

8.2 Central Limit Theorem

Let us start from a simple result for the binomial distribution.

Theorem 8.4 (Poisson's Theorem, 1837) *Let $\{X_{mn}\}_{1 \leq m \leq n}$ be i.i.d. Bernoulli random variables with probability of success $p_n = \lambda/n$ such that $0 < \lambda/n < 1$. Let Z be a random variable with Poisson distribution with mean λ. Then as $n \to \infty$*

$$X_{1n} + \cdots + X_{nn} \xrightarrow{d} Z.$$

Proof. Note that the characteristic function of X_{mn} is equal to $(1-\lambda/n)+e^{it}\lambda/n$. For any fixed t we have

$$\phi_{(X_{1n}+\cdots+X_{nn})}(t) = [1 + \frac{\lambda}{n}(e^{it} - 1)]^n \to \exp(\lambda(e^{it} - 1)),$$

as $n \to \infty$. The continuity theorem finishes the proof. $\square$

Central Limit Theorem (CLT) is a common name for limit theorems that provide conditions under which sum of random variables (appropriately centralized and normalized) weakly converges to the standard normal distribution. We will denote this convergence via $\xrightarrow{d} \mathcal{N}(0,1)$. The first CLT for sum of Bernoulli random variables with $p = 1/2$ was established by de Moivre in 1730. Laplace generalized it to the case of arbitrary p in 1812. The next theorem sometimes is called Levi's theorem

Theorem 8.5 (Central Limit Theorem) *Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with mean μ and variance $0 < \sigma^2 < \infty$. Then*

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. Consider $Y_n = (X_n - \mu)/\sigma$. Let $\phi_Y(t)$ be a characteristic function of Y_n. By Theorem 6.6 for any fixed t

$$\begin{aligned}
\phi_{(Y_1+\cdots+Y_n)/\sqrt{n}}(t) &= \left[\phi_Y(t/\sqrt{n})\right]^n \\
&= \left[1 + \frac{it}{\sqrt{n}}\mathbf{E}(Y_1) + \frac{i^2t^2}{2n}\mathbf{Var}(Y_1) + o(1/n)\right]^n \\
&= \left[1 - \frac{t^2}{2n} + o(1/n)\right]^n \\
&\to e^{-t^2/2}.
\end{aligned}$$

By continuity theorem we immediately get the result. $\square$

Now we prove a CLT for sums of independent (*but not necessarily identically distributed*) random variables. Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with means μ_n and variances $0 < \sigma_n^2 < \infty$. Denote

$$S_n = X_1 + \cdots + X_n,$$

$$B_n^2 = \mathbf{Var}(S_n) = \sum_{k=1}^{n} \sigma_k^2,$$

and for $\epsilon > 0$

$$L_n(\epsilon) = \frac{1}{B_n^2}\sum_{k=1}^{n} \mathbf{E}(X_k - \mu_k)^2 1_{|X_k - \mu_k| > \epsilon B_n}.$$

Theorem 8.6 (Lindeberg-Feller Theorem) *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with means μ_n and variances $0 < \sigma_n^2 < \infty$. If $L_n(\epsilon) \to 0$ for any $\epsilon > 0$ then*

1. *(uniform asymptotic negligibility)*

$$\max_{1\leq k\leq n} \frac{\sigma_k^2}{B_n^2} \to 0,$$

2. *(normality)*

$$\frac{S_n - \mathbf{E}S_n}{B_n} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof.

1. Note that for any $1 \leq k \leq n$ we have

$$\begin{aligned}
\frac{\sigma_k^2}{B_n^2} &= \frac{1}{B_n^2}\mathbf{E}(X_k - \mu_k)^2 \\
&= \frac{1}{B_n^2}\mathbf{E}(X_k - \mu_k)^2 1_{|X_k - \mu_k| \leq \epsilon B_n} + \frac{1}{B_n^2}\mathbf{E}(X_k - \mu_k)^2 1_{|X_k - \mu_k| > \epsilon B_n} \\
&\leq \epsilon^2 + L_n(\epsilon).
\end{aligned}$$

Therefore, we have a uniform bound

$$\max_{1\leq k\leq n} \frac{\sigma_k^2}{B_n^2} \leq \epsilon^2 + L_n(\epsilon),$$

and by choosing small ϵ, and then sufficiently large n we can make the right-hand side as small as we want.

2. Without loss of generality let us assume that $\mathbf{E}X_k = \mu_k = 0$. Let

$$\phi_k(t) = \mathbf{E}e^{itX_k},$$

and

$$\psi_n(t) = \mathbf{E}e^{itS_n/B_n} = \prod_{k=1}^{n} \phi_k(t/B_n).$$

We need to show that $\psi_n(t) \to e^{-t^2/2}$ for any $t \in \mathbb{R}$ as $n \to \infty$, or equivalently,

$$\ln \psi_n(t) + t^2/2 \to 0,$$

where ln denotes the principal branch of the complex logarithm.

First, let us show that for $k = 1, \ldots, n$

$$\ln \phi_k(t/B_n) = \phi_k(t/B_n) - 1 + r_k(t), \tag{8.1}$$

where $r_k(t)$ are such that $\sum_{k=1}^{n} |r_k(t)| \to 0$ as $n \to \infty$. Recall that (Lemma 6.5)

$$\left| e^{ix} - \sum_{j=0}^{k} \frac{(ix)^j}{j!} \right| \leq \frac{|x|^{k+1}}{(k+1)!}. \tag{8.2}$$

Therefore,

$$|\phi_k(t/B_n) - 1| = \left| \mathbf{E}\left[e^{itX_k/B_n} - 1 - \frac{itX_k}{B_n} \right] \right| \leq \mathbf{E}\frac{t^2 X_k^2}{2B_n^2} = \frac{t^2\sigma_k^2}{2B_n^2}.$$

Since

$$\frac{t^2\sigma_k^2}{2B_n^2} \le \frac{t^2}{2} \max_{1\le k\le n} \frac{\sigma_k^2}{B_n^2} \to 0,$$

as $n \to \infty$, we get that $|\phi_k(t/B_n) - 1| < 1/2$ for sufficiently large n, and

$$\begin{aligned}
\ln \phi_k(t/B_n) &= \ln(1 + \phi_k(t/B_n) - 1) \\
&= \phi_k(t/B_n) - 1 + \sum_{j=2}^{\infty} (-1)^{j+1} \frac{(\phi_k(t/B_n) - 1)^j}{j} \\
&= \phi_k(t/B_n) - 1 + r_k(t).
\end{aligned}$$

Now, taking into account that $|\phi_k(t/B_n) - 1| < 1/2$ first we get that

$$\begin{aligned}
|r_k(t)| &\le \sum_{j=2}^{\infty} |\phi_k(t/B_n) - 1|^j \\
&= |\phi_k(t/B_n) - 1|^2 \frac{1}{1 - |\phi_k(t/B_n) - 1|} \\
&\le 2|\phi_k(t/B_n) - 1|^2
\end{aligned}$$

Hence we then find that

$$\begin{aligned}
\sum_{k=1}^{n} |r_k(t)| &\le 2 \sum_{k=1}^{n} |\phi_k(t/B_n) - 1|^2 \\
&= 2 \sum_{k=1}^{n} \left| \mathbf{E}\left[e^{itX_k/B_n} - 1 - \frac{itX_k}{B_n} \right] \right|^2 \\
&\le 2 \sum_{k=1}^{n} \frac{t^4\sigma_k^4}{4B_n^4}
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{t^4}{2} \max_{1\le k\le n} \frac{\sigma_k^2}{B_n^2} \sum_{k=1}^n \frac{\sigma_k^2}{B_n^2} \\
&= \frac{t^4}{2} \max_{1\le k\le n} \frac{\sigma_k^2}{B_n^2} \to 0,
\end{aligned}
$$

as $n \to \infty$.

Using (8.1) and that $\sum_{k=1}^n \frac{\sigma_k^2}{B_n^2} = 1$ we get that

$$
\ln \psi_n(t) + t^2/2 = \sum_{k=1}^n \left[\phi_k(t/B_n) - 1 + \frac{t^2\sigma_k^2}{2B_n^2}\right] + \sum_{k=1}^n r_k(t).
$$

Next, note that because of (8.2) and the triangle inequality we obtain that

$$
\begin{aligned}
\left|\phi_k(t/B_n) - 1 + \frac{t^2\sigma_k^2}{2B_n^2}\right| &= \left|\mathbf{E}\left[e^{itX_k/B_n} - 1 - \frac{itX_k}{B_n} - \frac{i^2t^2X_k^2}{2B_n^2}\right]\right| \\
&\le \left|\mathbf{E}\left[e^{itX_k/B_n} - 1 - \frac{itX_k}{B_n} - \frac{i^2t^2X_k^2}{2B_n^2}\right] 1_{|X_k|\le \epsilon B_n}\right| \\
&+ \left|\mathbf{E}\left[e^{itX_k/B_n} - 1 - \frac{itX_k}{B_n} - \frac{i^2t^2X_k^2}{2B_n^2}\right] 1_{|X_k|> \epsilon B_n}\right| \\
&\le \mathbf{E}\frac{|t|^3|X_k|^3}{6B_n^3} 1_{|X_k|\le \epsilon B_n} + \mathbf{E}\frac{t^2X_k^2}{B_n^2} 1_{|X_k|> \epsilon B_n} \\
&\le \frac{|t|^3}{6}\epsilon \frac{\sigma_k^2}{B_n^2} + \frac{t^2}{B_n^2}\mathbf{E}X_k^2 1_{|X_k|> \epsilon B_n}.
\end{aligned}
$$

Therefore, finally we obtain

$$
\begin{aligned}
\ln \psi_n(t) + t^2/2 &\le \frac{|t|^3}{6}\epsilon \sum_{k=1}^n \frac{\sigma_k^2}{B_n^2} + \frac{t^2}{B_n^2}\sum_{k=1}^n \mathbf{E}X_k^2 1_{|X_k|> \epsilon B_n} + \sum_{k=1}^n |r_k(t)| \\
&= \frac{|t|^3}{6}\epsilon + t^2 L_n(\epsilon) + \sum_{k=1}^n |r_k(t)|,
\end{aligned}
$$

which can be made as small as we want by choosing ϵ, and then n. □

Note that the CLT for i.i.d. random variables follows from the Lindeberg-Feller Theorem. The next so-called Lyapunov's condition is easier to check and it is sufficient for $L_n(\epsilon) \to 0$. More specifically, we have the following theorem.

Theorem 8.7 (Lyapunov Theorem) *Assume that* $\mathbf{E}|X_k|^{2+\delta} < \infty$ *for some* $\delta > 0$ *and* $k = 1, 2, \ldots$. *If*

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} \mathbf{E}|X_k - \mu_k|^{2+\delta} \to 0$$

then

$$\frac{S_n - \mathbf{E}S_n}{B_n} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. Just note that

$$\begin{aligned}\frac{1}{B_n^2} \sum_{k=1}^{n} \mathbf{E}(X_k - \mu_k)^2 1_{|X_k - \mu_k| > \epsilon B_n} &= \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbf{E} \frac{|X_k - \mu_k|^{2+\delta}}{|X_k - \mu_k|^{\delta}} 1_{|X_k - \mu_k| > \epsilon B_n} \\ &\leq \frac{1}{\epsilon^{\delta} B_n^{2+\delta}} \sum_{k=1}^{n} \mathbf{E}|X_k - \mu_k|^{2+\delta} \to 0.\end{aligned}$$

□

Exercise 8.1 Let $\{X_n\}_{n \geq 1}$ be a sequence of independent random variables with variances $0 < \sigma_n^2 < \infty$. Show that $L_n(\epsilon) \to 0$ implies that $B_n \to \infty$.

Exercise 8.2 Let $\{X_n\}_{n \geq 1}$ be a sequence of independent random variables such that $0 < \inf_n \mathbf{Var}(X_n)$ and $\sup_n \mathbf{E}|X_n|^3 < \infty$. Show that $(S_n - \mathbf{E}S_n)/B_n \xrightarrow{d} \mathcal{N}(0, 1)$.

Exercise 8.3 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with means μ and variances $0 < \sigma^2 < \infty$. Show that $L_n(\epsilon) \to 0$.

8.3 Convergence of Series of Random Variables

Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{P}(X_n = 1) = P(X = -1) = 1/2$. Consider random variables $S_n = \sum_{k=1}^{n} X_k/k$. What can we say about the a.s. convergence of this sequence?

Theorem 8.8 (Cauchy Criterion for a.s. Convergence) *The following three conditions are equivalent.*

1. X_n *is convergent with probability 1,*
2. *for every* $\epsilon > 0$
$$\mathbf{P}\left[\sup_{k,l\geq n} |X_k - X_l| \geq \epsilon\right] \to 0, \text{ as } n \to \infty,$$
3. *for every* $\epsilon > 0$
$$\mathbf{P}\left[\sup_{k\geq 0} |X_{n+k} - X_n| \geq \epsilon\right] \to 0, \text{ as } n \to \infty.$$

Proof.

(1. $\Rightarrow$ 2.) Let $X_n \xrightarrow{a.s.} X$. Since

$$\sup_{k,l\geq n} |X_k - X_l| \leq \sup_{k\geq n} |X_k - X| + \sup_{l\geq n} |X_l - X|,$$

by Theorem 7.1 we get the result.

(2. $\Rightarrow$ 1.) Let $B = \{\omega : X_n(\omega) \text{ does not converge}\}$. Note that

$$\begin{aligned} B =& \{\omega : \liminf_n X_n(\omega) < \limsup_n X_n(\omega)\} \\ =& \{\omega : \exists N \geq 1\, \forall n \geq 1\, \exists k, l \geq n \text{ st } |X_l(\omega) - X_k(\omega)| > 1/N\}. \end{aligned}$$

So, we have

$$B = \bigcup_{N=1}^{\infty} \bigcap_{n=1}^{\infty} [\sup_{k,l \geq n} |X_k - X_l| > 1/N].$$

Note that $A_n = [\sup_{k,l\geq n} |X_k - X_l| > 1/N]$ is a monotone decreasing sequence of events. Therefore,

$$\mathbf{P}(B) \leq \sum_{N=1}^{\infty} \mathbf{P}\Big(\bigcap_{n=1}^{\infty} [\sup_{k,l \geq n} |X_k - X_l| > 1/N]\Big).$$

But for any fixed N we have

$$\begin{aligned} &\mathbf{P}\Big(\bigcap_{n=1}^{\infty} [\sup_{k,l \geq n} |X_k - X_l| > 1/N]\Big) \\ &\qquad = \lim_{n\to\infty} \mathbf{P}\Big([\sup_{k,l \geq n} |X_k - X_l| > 1/N]\Big) \\ &\qquad = 0, \end{aligned}$$

That is, we have that $\mathbf{P}(B) = 0$.

(2. $\Leftrightarrow$ 3.) Just observe that

$$\sup_{k\geq 0}|X_{n+k}-X_n| \leq \sup_{k,l\geq 0}|X_{n+k}-X_{n+l}| \leq 2\sup_{k\geq 0}|X_{n+k}-X_n|.$$

This completes the proof. $\square$

Exercise 8.4 Show that if $0 \leq X_n \leq Y_n$ and $Y_n \xrightarrow{\mathbf{P}} 0$, then $X_n \xrightarrow{\mathbf{P}} 0$.

Theorem 8.9 (Kolmogorov's Inequality) *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with means μ_n and variances $\sigma_n^2 < \infty$. Then*

$$\mathbf{P}\left(\max_{1\leq k\leq n}\Big|\sum_{j=1}^{k}(X_j-\mu_j)\Big| \geq \epsilon\right) \leq \sum_{j=1}^{n}\frac{\sigma_j^2}{\epsilon^2}.$$

Proof. Without loss of generality we assume that $\mu_n = 0$. Denote $S_n = X_1 + \cdots + X_n$. Consider the following events:

$$\begin{aligned}
A &= \{\omega : \max_{1\leq k\leq n}|S_k(\omega)| \geq \epsilon\},\\
A_1 &= \{\omega : |S_1(\omega)| \geq \epsilon\},\\
A_2 &= \{\omega : |S_1(\omega)| < \epsilon, |S_2(\omega)| \geq \epsilon\},\\
&\cdots\\
A_n &= \{\omega : |S_1(\omega)| < \epsilon, \ldots, |S_{n-1}(\omega)| < \epsilon, |S_n(\omega)| \geq \epsilon\}.
\end{aligned}$$

It is obvious that $A = \cup_{k=1}^n A_k$, and $A_k A_l = \emptyset$ if $k \neq l$. First, note that

$$\mathbf{Var}(S_n) = \mathbf{E}S_n^2 \geq \mathbf{E}S_n^2 1_A = \sum_{k=1}^{n}\mathbf{E}S_n^2 1_{A_k}.$$

Now, we find that

$$\begin{aligned}
\mathbf{E}S_n^2 1_{A_k} &= \mathbf{E}[S_k + X_{k+1} + \cdots + X_n]^2 1_{A_k} \\
&= \mathbf{E}[S_k]^2 1_{A_k} + \mathbf{E}[X_{k+1} + \cdots + X_n]^2 1_{A_k} + 2\mathbf{E}S_k(X_{k+1} + \cdots + X_n)1_{A_k} \\
&= \mathbf{E}[S_k]^2 1_{A_k} + \mathbf{E}[X_{k+1} + \cdots + X_n]^2 1_{A_k} + 2\mathbf{E}S_k 1_{A_k}\mathbf{E}(X_{k+1} + \cdots + X_n) \\
&= \mathbf{E}[S_k]^2 1_{A_k} + \mathbf{E}[X_{k+1} + \cdots + X_n]^2 1_{A_k} \\
&\geq \mathbf{E}[S_k]^2 1_{A_k} \\
&\geq \epsilon^2 \mathbf{P}(A_k).
\end{aligned}$$

Thus we finally get that

$$\mathbf{Var}(S_n) \geq \sum_{k=1}^{n} \epsilon^2 \mathbf{P}(A_k) = \epsilon^2 \mathbf{P}(A).$$

□

Exercise 8.5 (symmetrization trick) Let X and X' be i.i.d. with mean μ, variance σ^2 and CF $\phi(t)$. Show that $X - X'$ is symmetrically distributed random variable with $\mathbf{E}(X - X') = 0$, $\mathbf{Var}(X - X') = 2\sigma^2$, and $\phi_{X-X'}(t) = |\phi(t)|^2$.

Theorem 8.10 (Two-Series Theorem) *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with means μ_n and variances $\sigma_n^2 < \infty$.*

1. *If both series $\sum_n \mu_n$ and $\sum_n \sigma_n^2$ converge, then $\sum_n X_n$ converges with probability 1.*

2. *If $\sum_n X_n$ converges with probability 1 and there is C such that $\mathbf{P}(|X_n| \leq$*

$C) = 1$ for all n, then both $\sum_n \mu_n$ and $\sum_n \sigma_n^2$ converge.

Proof.

1. Without loss of generality we assume that $\mu_n = 0$. Denote $S_n = X_1 + \cdots + X_n$. By Cauchy criterion (Theorem 8.8) we need to show that for every $\epsilon > 0$

$$\mathbf{P}\left[\sup_{k\geq 0} |S_{n+k} - S_n| \geq \epsilon\right] \to 0, \text{ as } n \to \infty.$$

By Kolmogorov's inequality we have

$$\begin{aligned}
\mathbf{P}\left[\sup_{k\geq 0} |S_{n+k} - S_n| \geq \epsilon\right] &= \lim_{N\to\infty} \mathbf{P}\left[\sup_{1\leq k\leq N} |S_{n+k} - S_n| \geq \epsilon\right] \\
&\leq \lim_{N\to\infty} \sum_{k=n+1}^{n+N} \sigma_k^2/\epsilon^2 \\
&= \sum_{k=n+1}^{\infty} \sigma_k^2/\epsilon^2 \to 0,
\end{aligned}$$

as $n \to \infty$.

2. Since $S_n \xrightarrow{a.s.} S$, at any $t \in \mathbb{R}$ the CFs of S_n converge to the CF of random variable S. But

$$\phi_{S_n}(t) = \prod_{j=1}^{n} \phi_{X_j}(t),$$

therefore,

$$\prod_{j=1}^{\infty} \phi_{X_j}(t) = \phi_S(t),$$

and

$$\prod_{j=1}^{\infty} |\phi_{X_j}(t)|^2 = |\phi_S(t)|^2 > 0$$

for all t sufficiently close to 0, because $|\phi_S(t)|^2$ is a CF. By a standard result from analysis we get

$$\begin{aligned}\infty &> \sum_{j=1}^{\infty}(1-|\phi_{X_j}(t)|^2)\\ &= \sum_{j=1}^{\infty}\int_{-2C}^{2C}(1-\cos(tx))F_j(dx)\\ &= 2\sum_{j=1}^{\infty}\int_{-2C}^{2C}\sin^2(tx/2)F_j(dx),\end{aligned}$$

where F_j is a cdf that corresponds to CF $|\phi_{X_j}(t)|^2$. Since $|\sin(y)| > |y|/2$ for small y, we can pick up t close to 0 such that

$$\begin{aligned}\infty &> 2\sum_{j=1}^{\infty}\int_{-2C}^{2C}\sin^2(tx/2)F_j(dx)\\ &> 2\sum_{j=1}^{\infty}\int_{-2C}^{2C}\frac{t^2x^2}{16}F_j(dx)\\ &= \frac{t^2}{8}\sum_{j=1}^{\infty}(2\sigma_j^2)\\ &= \frac{t^2}{4}\sum_{j=1}^{\infty}\sigma_j^2.\end{aligned}$$

That is, $\sum_n \sigma_n^2$ converges. By the first part of the theorem $\sum_n(X_n - \mu_n)$ converges with probability 1. Taking into account that $\sum_n X_n$ converges with probability 1 as well, we also get that $\sum_n \mu_n$ converges. $\square$

Exercise 8.6 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{P}(X_n = 1) = P(X = -1) = 1/2$. Show that $\sum_k X_k/k$ converges with probability 1.

Theorem 8.11 (Three-Series Theorem) *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables. For any $C > 0$ denote $X_j^C = X_j 1_{|X_j|\leq C}$.*

1. *If there is $C > 0$ such that $\sum_n \mathbf{E}X_n^C$, $\sum_n \mathbf{Var}(X_n^C)$ and $\sum_n \mathbf{P}(|X_n| > C)$ converge, then $\sum_n X_n$ converges with probability 1.*
2. *If $\sum_n X_n$ converges with probability 1, then $\sum_n \mathbf{E}X_n^C$, $\sum_n \mathbf{Var}(X_n^C)$ and $\sum_n \mathbf{P}(|X_n| > C)$ converge for every $C > 0$.*

Proof.

1. By the two-series theorem $\sum_n X_n^C$ converges with probability 1. Because $\sum_n \mathbf{P}(|X_n| > C) < \infty$, by the Borel-Cantelli lemma we have that $\mathbf{P}(|X_n| > C \text{ i.o.}) = 0$, that is, $X_n = X_n^C$ for all n with at most finitely many exceptions. Hence $\sum_n X_n$ converges with probability 1 as well.

2. The a.s. convergence of $\sum_n X_n$ implies that $X_n \to 0$ with probability 1. Therefore, $\mathbf{P}(|X_n| > C \text{ i.o.}) = 0$ for any $C > 0$, and by the Borel-Cantelli lemma (X_n are independent!) we obtain that $\sum_n \mathbf{P}(|X_n| > C) < \infty$. Moreover, the a.s. convergence of $\sum_n X_n$ together with $\mathbf{P}(|X_n| > C \text{ i.o.}) = 0$ implies that $\sum_n X_n^C$ converges with probability 1. Therefore, by the two-series theorem we get that $\sum_n \mathbf{E}X_n^C$ and $\sum_n \mathbf{Var}(X_n^C)$ converge. $\square$

Exercise 8.7 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent non-negative random variables. Show that $\sum_n X_n$ converges with probability 1 if and only if

$$\sum_n [\mathbf{P}(X_n > 1) + \mathbf{E}(X_n 1_{X_n \leq 1})] < \infty.$$

Exercise 8.8 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent non-negative random variables. Show that $\sum_n X_n$ converges with probability 1 if and only if

$$\sum_n \mathbf{E}\frac{X_n}{1+X_n} < \infty.$$

Exercise 8.9 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables random variables with $\mathbf{E}X_n = 0$. Show that if

$$\sum_n \mathbf{E}\frac{X_n^2}{1+|X_n|} < \infty,$$

then $\sum_n X_n$ converges with probability 1.

8.4 Strong Law of Large Numbers

We will begin with two auxiliary results.

Lemma 8.1 (Toeplitz Lemma) *Let $\{a_n\}_{n\geq 1}$ be a sequence of positive numbers, $b_n = \sum_{k=1}^n a_k$, and suppose that $b_n \uparrow \infty$. If $\{x_n\}_{n\geq 1}$ is such that $x_n \to x$ then*

$$\frac{1}{b_n}\sum_{k=1}^n a_k x_k \to x.$$

Proof. Fix $\epsilon > 0$. First, choose n_0 such that $|x_n - x| < \epsilon/2$ for all $n > n_0$. Then choose $n_1 > n_0$ such that

$$\frac{1}{b_{n_1}} \sum_{k=1}^{n_0} a_k |x_k - x| < \epsilon/2.$$

Now, for all $n > n_1$ we have

$$\begin{aligned}
\left| \frac{1}{b_n} \sum_{k=1}^{n} a_k x_k - x \right| &\leq \frac{1}{b_n} \sum_{k=1}^{n} a_k \left| x_k - x \right| \\
&\leq \frac{1}{b_n} \sum_{k=1}^{n_0} a_k \left| x_k - x \right| + \frac{1}{b_n} \sum_{k=n_0+1}^{n} a_k \left| x_k - x \right| \\
&\leq \frac{1}{b_{n_1}} \sum_{k=1}^{n_0} a_k \left| x_k - x \right| + \frac{1}{b_n} \sum_{k=n_0+1}^{n} a_k \left| x_k - x \right| \\
&\leq \frac{\epsilon}{2} + \frac{b_n - b_{n_0}}{b_n} \frac{\epsilon}{2} \\
&\leq \epsilon.
\end{aligned}$$

□

Exercise 8.10 Suppose that $\{x_n\}_{n \geq 1}$ is such that $x_n \to x$. Show that

$$\frac{x_1 + \cdots + x_n}{n} \to x.$$

Lemma 8.2 (Kronecker's Lemma) *Let $\{b_n\}_{n \geq 1}$ be a sequence of positive increasing numbers, and $b_n \uparrow \infty$. If $\{x_n\}_{n \geq 1}$ is such that $\sum_n x_n$ converges then*

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k x_k \to 0.$$

Proof. Denote $b_0 = 0$, $S_0 = 0$, $S_n = \sum_{k=1}^{n} x_k$, and $S = \lim_{n\to\infty} S_n$. Then

$$\sum_{k=1}^{n} b_k x_k = \sum_{k=1}^{n} b_k (S_k - S_{k-1}) = b_n S_n - \sum_{k=1}^{n} (b_k - b_{k-1}) S_{k-1},$$

that is,

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = S_n - \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) S_{k-1}.$$

Since $S_n \to S$, then by Toeplitz lemma

$$\frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) S_{k-1} \to S$$

as well. Therefore,

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k x_k \to 0.$$

□

Exercise 8.11 Suppose that $\sum_n x_n/n$ converges show that

$$\frac{x_1 + \cdots + x_n}{n} \to 0.$$

Theorem 8.12 (Strong LLN) *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with finite second moments. Suppose that $\{b_n\}_{n\geq 1}$ is a sequence of positive increasing numbers such that $b_n \uparrow \infty$, and*

$$\sum_n \frac{\mathbf{Var} X_n}{b_n^2} < \infty.$$

Then with probability 1

$$\frac{S_n - \mathbf{E}S_n}{b_n} \to 0.$$

Proof. By the two-series theorem with probability 1

$$\sum_n \frac{X_n - EX_n}{b_n}$$

converges. Therefore, by Kronecker's lemma with probability 1

$$\frac{S_n - \mathbf{E}S_n}{b_n} = \frac{1}{b_n}\sum_{k=1}^{n} b_k \frac{X_k - EX_k}{b_k} \to 0$$

as $n \to \infty$. $\square$

Exercise 8.12 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{P}(X_n = 1) = P(X = -1) = 1/2$. Show that with probability 1

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}\ln n} \to 0.$$

Lemma 8.3 *Let X be nonnegative random variable with $\mathbf{E}X < \infty$. Show that*

$$\sum_{n=1}^{\infty} \mathbf{P}(X \geq n) \leq \mathbf{E}X \leq 1 + \sum_{n=1}^{\infty} \mathbf{P}(X \geq n).$$

Proof. Exercise. Use $\mathbf{E}X = \int_0^\infty \mathbf{P}(X \geq x)dx$ (see Example 5.2). $\square$

Theorem 8.13 (Strong LLN for I.I.D. Random Variables) *Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{E}|X_1| < \infty$. Then with probability*

1 we have

$$\frac{S_n}{n} \to \mathbf{E}X_1.$$

Proof. Without loss of generality assume that $\mathbf{E}X_1 = 0$. Because of Lemma 8.3 and Borel-Cantelli we get that

$$\mathbf{P}(|X_n| \geq n \text{ i.o.}) = 0.$$

Let $Y_n = X_n 1_{|X_n|<n}$. Since $Y_n = X_n$ for all n with at most finitely many exceptions,

$$\frac{X_1 + \cdots + X_n}{n} \to 0, \quad \text{a.s.}$$

if and only if

$$\frac{Y_1 + \cdots + Y_n}{n} \to 0 \quad \text{a.s.}$$

Next, note that by the dominated convergence theorem $\mathbf{E}Y_n = \mathbf{E}X_1 1_{|X_1|<n} \to 0$ as $n \to \infty$, therefore, by Toeplitz lemma

$$\frac{\mathbf{E}Y_1 + \cdots + \mathbf{E}Y_n}{n} \to 0.$$

So, if we denote $Z_n = Y_n - \mathbf{E}Y_n$ we get that

$$\frac{Y_1 + \cdots + Y_n}{n} \to 0 \quad \text{a.s.}$$

if and only if

$$\frac{Z_1 + \cdots + Z_n}{n} \to 0 \quad \text{a.s.}$$

The two-series theorem together with Kronecker's Lemma tell us that all we need to prove is that

$$\sum_n \frac{\mathbf{Var}(Z_n)}{n^2} < \infty.$$

We have for

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\mathbf{Var}(Z_n)}{n^2} &\leq \sum_{n=1}^{\infty} \frac{\mathbf{E}Y_n^2}{n^2} \\
&= \sum_{n=1}^{\infty} \frac{\mathbf{E}X_n^2 1_{|X_n|<n}}{n^2} \\
&= \sum_{n=1}^{\infty} \frac{\mathbf{E}X_1^2 1_{|X_1|<n}}{n^2} \\
&= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} \mathbf{E}X_1^2 1_{k-1\leq|X_1|<k} \\
&= \sum_{k=1}^{\infty} \mathbf{E}X_1^2 1_{k-1\leq|X_1|<k} \sum_{n=k}^{\infty} \frac{1}{n^2} \\
&\leq \sum_{k=1}^{\infty} \frac{2}{k} \mathbf{E}X_1^2 1_{k-1\leq|X_1|<k} \\
&\leq 2 \sum_{k=1}^{\infty} \mathbf{E}|X_1| 1_{k-1\leq|X_1|<k} \\
&= 2\mathbf{E}|X_1| < \infty
\end{aligned}$$

This finishes the proof. $\square$

Exercise 8.13 Show that for every $k \geq 1$ we have

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{2}{k}.$$

Exercise 8.14 Show that $\mathbf{E}X^2 < \infty$ if and only if

$$\sum_{n=1}^{\infty} n\mathbf{P}(|X| > n) < \infty.$$

Finally, let us note that a converse (in a sense) is also true.

Proposition 8.1 *Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables such that with probability 1 we have*

$$\frac{S_n}{n} \to C < \infty.$$

Then $\mathbf{E}|X_1| < \infty$, and $C = \mathbf{E}X_1$.

Proof. Observe that with probability 1

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \to 0.$$

Therefore, $\mathbf{P}(|X_n| \geq n \text{ i.o.}) = 0$, and by Borel-Cantelli lemma

$$\sum_n \mathbf{P}(|X_1| \geq n) < \infty.$$

Lemma 8.3 gives us $\mathbf{E}|X_1| < \infty$, and by the strong LLN for i.i.d. sequences we get that $C = \mathbf{E}X_1$. $\square$

Example 8.1 (The Monte Carlo Method) Let $f : [0,1] \to [0,1]$ be continuous function. Let $X_1, Y_1, X_2, Y_2, \ldots$ be a sequence of i.i.d. random variables uniformly distributed on $[0,1]$. Let $Z_n = 1_{f(X_n) > Y_n}$. Since

$$\mathbf{E}Z_n = \int_0^1 f(x)dx < \infty,$$

by the strong LLN we get that with probability 1

$$\frac{Z_1 + \cdots + Z_n}{n} \to \int_0^1 f(x)dx.$$

Chapter 9

Martingales

9.1 Conditional Expectation: Definition

Definition 9.1 *Given are a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, a* σ*-field* $\mathcal{G} \subset \mathcal{F}$*, and a random variable* X *with* $\mathbf{E}|X| < \infty$*. We define* conditional expectation of X given $\mathcal{G}$, $\mathbf{E}(X|\mathcal{G})$*, to be any random variable* Y *that satisfies the following two conditions:*

a) Y *is* $\mathcal{G}$*-measurable,*

b) for all $A \in \mathcal{G}$ $\mathbf{E}X1_A = \mathbf{E}Y1_A$*.*

Theorem 9.1 (Existence and Uniqueness of Conditional Expectation) *The conditional expectation exists, a.s. unique, and integrable.*

Proof. First, let us prove that the conditional expectation is integrable. Let

$A = \{\omega : Y > 0\}$, then

$$\mathbf{E}Y1_A = \mathbf{E}X1_A \leq \mathbf{E}|X|1_A,$$

and

$$\mathbf{E}(-Y)1_{A^c} = \mathbf{E}(-X)1_{A^c} \leq \mathbf{E}|X|1_{A^c}.$$

Thus $\mathbf{E}|Y| \leq \mathbf{E}|X|$.

The existence follows from the Radon-Nikodým Theorem. Indeed, first assume that $X \geq 0$. Then

$$\nu(A) = \mathbf{E}X1_A, \quad A \in \mathcal{G}$$

is a measure on $(\Omega, \mathcal{G})$, and it is absolutely continuous with respect to $\mathbf{P}$. Therefore, there exists an integrable $\mathcal{G}$-measurable random variable $Y \geq 0$ such that

$$\nu(A) = \mathbf{E}Y1_A,$$

for all $A \in \mathcal{G}$. The general case is treated with help of Hahn's decomposition $X = X^+ - X^-$.

Assume that Y' is another conditional expectation of X given $\mathcal{G}$. Consider $A_\epsilon = \{\omega : Y - Y' \geq \epsilon\}$, where $\epsilon > 0$. Then

$$0 = \mathbf{E}(X - X)1_{A_\epsilon} = \mathbf{E}(Y - Y')1_{A_\epsilon} \geq \epsilon\mathbf{P}(A_\epsilon),$$

that is, $P(A_\epsilon) = 0$ for any $\epsilon > 0$. Therefore, $Y \leq Y'$ a.s. But by the same argument we get that $Y' \leq Y$ a.s. $\square$

Definition 9.2 Conditional probability of event A given σ-field $\mathcal{G}$ *is defined by*

$$\mathbf{P}(A|\mathcal{G}) = \mathbf{E}(1_A|\mathcal{G}).$$

Exercise 9.1 Suppose that two integrable random variables X_1 and X_2 coincide on $B \in \mathcal{G}$. Show that $\mathbf{E}(X_1|\mathcal{G}) = \mathbf{E}(X_2|\mathcal{G})$ a.s. on B.

Exercise 9.2 Let A and B be two events, and $\mathcal{G} = \sigma(1_B)$. Find $\mathbf{P}(A|\mathcal{G})$.

9.2 Properties of Conditional Expectation

Here all random variables on $(\Omega, \mathcal{F}, \mathbf{P})$ have finite absolute first moment, all σ-fields are subfileds of $\mathcal{F}$, and all equations/inequalities that involve random variables hold a.s.

- $\mathbf{E}(\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(X)$

  Proof. Since $\Omega \in \mathcal{G}$

  $$\mathbf{E}(\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(\mathbf{E}(X|\mathcal{G})1_\Omega) = \mathbf{E}(X1_\Omega) = \mathbf{E}(X).$$

  $\square$

- If X is $\mathcal{G}$-measurable, then $\mathbf{E}(X|\mathcal{G}) = X$.

  Proof. It is immediate from the definition. $\square$

- (*linearity*) $\mathbf{E}(aX + bY|\mathcal{G}) = a\mathbf{E}(X|\mathcal{G}) + b\mathbf{E}(Y|\mathcal{G})$.

  Proof. For any $A \in \mathcal{G}$ we have

$$\begin{aligned}\mathbf{E}[(a\mathbf{E}(X|\mathcal{G}) + b\mathbf{E}(Y|\mathcal{G}))1_A] &= a\mathbf{E}[\mathbf{E}(X|\mathcal{G})1_A] + b\mathbf{E}[\mathbf{E}(Y|\mathcal{G})1_A] \\ &= a\mathbf{E}(X1_A) + b\mathbf{E}(Y1_A) \\ &= \mathbf{E}[(aX + bY)1_A].\end{aligned}$$

  $\square$

- (*positivity*) If $X \geq 0$ then $\mathbf{E}(X|\mathcal{G}) \geq 0$.

  Proof. For any $A \in \mathcal{G}$ we have that $\mathcal{G}$-measurable random variable $\mathbf{E}(X|\mathcal{G})$ satisfies

$$\mathbf{E}(\mathbf{E}(X|\mathcal{G})1_A) = \mathbf{E}(X1_A) \geq 0,$$

  therefore, (see page 67) we get $\mathbf{E}(X|\mathcal{G}) \geq 0$. $\square$

- (*monotonicity*) If $X \geq Y$ then $\mathbf{E}(X|\mathcal{G}) \geq \mathbf{E}(Y|\mathcal{G})$.

  Proof. It immediately follows from positivity and linearity. $\square$

- (*conditional monotone convergence theorem*) If $0 \leq X_n \uparrow X$ with $\mathbf{E}X < \infty$, then $\mathbf{E}(X_n|\mathcal{G}) \uparrow \mathbf{E}(X|\mathcal{G})$ a.s.

  Proof. By positivity and monotonicity $0 \leq \mathbf{E}(X_n|\mathcal{G}) \uparrow$. Let $Y = \limsup_n \mathbf{E}(X_n|\mathcal{G})$, then Y is $\mathcal{G}$-measurable, and for any $A \in \mathcal{G}$ $\mathbf{E}(X_n|\mathcal{G})1_A \uparrow Y1_A$ a.s. Since for any $A \in \mathcal{G}$ we have

$$\mathbf{E}(X_n 1_A) = \mathbf{E}(\mathbf{E}(X_n|\mathcal{G})1_A),$$

therefore, by the monotone convergence theorem

$$\mathbf{E}(X1_A) = \mathbf{E}(Y1_A),$$

that is, $Y = \mathbf{E}(X|\mathcal{G})$. $\square$

- (*conditional Fatou's Lemma*)

  (1) If $X_n \geq Y$ for all n, and $\mathbf{E}(Y) > -\infty$, then

  $$\mathbf{E}(\liminf_n X_n|\mathcal{G}) \leq \liminf_n \mathbf{E}(X_n|\mathcal{G}).$$

  (2) If $X_n \leq Y$ for all n, and $\mathbf{E}(Y) < \infty$, then

  $$\limsup_n \mathbf{E}(X_n|\mathcal{G}) \leq \mathbf{E}(\limsup_n X_n|\mathcal{G}).$$

  (3) If $|X_n| \leq Y$ for all n, and $\mathbf{E}(Y) < \infty$, then

  $$\mathbf{E}(\liminf_n X_n|\mathcal{G}) \leq \liminf_n \mathbf{E}(X_n|\mathcal{G}) \leq \limsup_n \mathbf{E}(X_n|\mathcal{G}) \leq \mathbf{E}(\limsup_n X_n|\mathcal{G}).$$

  Proof. Exercise. $\square$

- (*conditional dominated convergence theorem*) Let $Y, X_1, X_2, \ldots$ be random variables such that $|X_n| \leq Y$ for all n, $\mathbf{E}(Y) < \infty$, and $X_n \to X$ a.s. Then as $n \to \infty$ with probability 1

  $$\mathbf{E}(X_n|\mathcal{G}) \to \mathbf{E}(X|\mathcal{G}).$$

Proof. Exercise. $\square$

- (*tower property*) If $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, then

$$\mathbf{E}(\mathbf{E}(X|\mathcal{G})|\mathcal{H}) = \mathbf{E}(X|\mathcal{H}) = \mathbf{E}(\mathbf{E}(X|\mathcal{H})|\mathcal{G}).$$

Proof. The second equation is trivial. Now, let $Y = \mathbf{E}(X|\mathcal{G})$ and $Z = \mathbf{E}(X|\mathcal{H})$. Just note that for every $A \in \mathcal{H} \subset \mathcal{G}$ using definitions of Y and Z we get

$$\mathbf{E}(Y1_A) = \mathbf{E}(X1_A) = \mathbf{E}(Z1_A).$$

$\square$

- (*non-anticipating multiplier property*) Suppose that both $\mathbf{E}|YX|$ and $\mathbf{E}|X|$ are both finite, and Y is $\mathcal{G}$-measurable, then

$$\mathbf{E}(YX|\mathcal{G}) = Y\mathbf{E}(X|\mathcal{G}).$$

Proof. Consider first $Y = 1_B$ where $B \in \mathcal{G}$. Then for any $A \in \mathcal{G}$ we have

$$\begin{aligned}\mathbf{E}(1_B\mathbf{E}(X|\mathcal{G})1_A) &= \mathbf{E}(\mathbf{E}(X|\mathcal{G})1_{AB})\\ &= \mathbf{E}(X1_{AB})\\ &= \mathbf{E}1_BX1_A,\end{aligned}$$

that is, $1_B\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(1_B X|\mathcal{G})$. By the linearity we get that it is true if Y is a non-negative simple random variable. The monotone convergence theorem gives us the formula for $Y, X \geq 0$. Splitting X and Y into positive and negative parts finishes the proof. $\square$

- (*conditional Jensen's inequality*) Let f be a convex Borel function, and let X be random variables with $\mathbf{E}|X|, \mathbf{E}|f(X)| < \infty$. Then

$$f(\mathbf{E}(X|\mathcal{G})) \leq \mathbf{E}(f(X)|\mathcal{G}).$$

In particular, for $p \geq 1$

$$|\mathbf{E}(X|\mathcal{G})|^p \leq \mathbf{E}(|X|^p|\mathcal{G}).$$

Proof. First recall that a Borel function $f : \mathbb{R} \mapsto \mathbb{R}$ is said to be convex iff for any y there is a number a(y) such that

$$f(x) \geq f(y) + (x - y)a(y)$$

for all $x \in \mathbb{R}$.

By convexity we have (assuming $x = X$, $y = \mathbf{E}(X|\mathcal{G})$)

$$f(X) \geq f(\mathbf{E}(X|\mathcal{G})) + (X - \mathbf{E}(X|\mathcal{G}))a(\mathbf{E}(X|\mathcal{G})),$$

and after taking the conditional expectation with respect to $\mathcal{G}$ we obtain

Jensen's inequality.

Exercise: Provide the details. Note that we do not know if the expectation of the RHS, in fact, exists. □

- (*independence*) If X is independent of $\mathcal{G}$, then $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X)$.

Proof. Note that $\mathbf{E}(X)$ is $\mathcal{G}$-measurable and for every $A \in \mathcal{G}$ we have

$$\mathbf{E}(\mathbf{E}(X)1_A) = \mathbf{E}(X)\mathbf{E}(1_A) = \mathbf{E}(X1_A).$$

□

- (*geometrical interpretation*) Suppose $\mathbf{E}X^2, \mathbf{E}Y^2 < \infty$ and Y is $\mathcal{G}$-measurable, then

$$\mathbf{E}(X - \mathbf{E}(X|\mathcal{G}))^2 \leq \mathbf{E}(X - Y)^2.$$

Proof. Denote $Z = \mathbf{E}(X|\mathcal{G})$. We have

$$\begin{aligned}
\mathbf{E}(X - Y)^2 &= \mathbf{E}(X - Z + Z - Y)^2 \\
&= \mathbf{E}(X - Z)^2 + \mathbf{E}(Z - Y)^2 + 2\mathbf{E}[(X - Z)(Z - Y)] \\
&\geq \mathbf{E}(X - Z)^2 + 2\mathbf{E}[(X - Z)(Z - Y)]
\end{aligned}$$

But since $Z - Y$ is $\mathcal{G}$-measurable we also get

$$\mathbf{E}[(X - Z)(Z - Y)] = \mathbf{E}[\mathbf{E}((X - Z)(Z - Y)|\mathcal{G})]$$

$$= \mathbf{E}[(Z-Y)\mathbf{E}(X-Z|\mathcal{G})]$$

$$= \mathbf{E}[(Z-Y)\cdot 0] = 0.$$

Note that the Cauchy-Schwarz inequality was used in the proof. □

Exercise 9.3 Suppose X and Y are independent. Let $f : \mathbb{R}^2 \mapsto \mathbb{R}$ be a Borel function such that $\mathbf{E}|f(X,Y)| < \infty$, and $g(\cdot) = \mathbf{E}(f(\cdot, Y))$. Then

$$\mathbf{E}(f(X,Y)|\sigma(X)) = g(X).$$

Exercise 9.4 Let $\mathbf{Var}(X|\mathcal{G}) = \mathbf{E}(X^2|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})^2$. Show that

$$\mathbf{Var}(X) = \mathbf{E}(\mathbf{Var}(X|\mathcal{G})) + \mathbf{Var}(\mathbf{E}(X|\mathcal{G})).$$

Exercise 9.5 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with means μ and variances $\sigma^2 < \infty$, and N an independent positive integer-valued random variable with mean n and variance ν^2. Show that

$$\mathbf{Var}(X_1 + \cdots + X_N) = \sigma^2 n + \mu^2\nu^2.$$

Exercise 9.6 Let $X_1, \ldots, X_n$ be i.i.d. random variables, $S_n = X_1 + \cdots + X_n$, and $\mathcal{G} = \sigma(S_n)$. Find $\mathbf{E}(X_1|\mathcal{G})$.

9.3 Martingale: Definition

Definition 9.3 *Consider probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*. A sequence of* σ*-fields* $\{\mathcal{F}_n\}_{n\geq 1}$ *is called* filtration *if*

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots \subset \mathcal{F}.$$

We also define $\mathcal{F}_\infty \subset \mathcal{F}$ *as* $\sigma(\cup_n \mathcal{F}_n)$*.*

Definition 9.4 *We say that a sequence of random variables* $\{X_n\}_{n\geq 1}$ *is* adapted *to a filtration* $\{\mathcal{F}_n\}_{n\geq 1}$ *if* X_n *is* $\mathcal{F}_n$*-measurable for every* n*.*

Definition 9.5 *A sequence of random variables* $\{X_n\}_{n\geq 1}$ *is called a* martingale *with respect to filtration* $\{\mathcal{F}_n\}_{n\geq 1}$ *if*

1. $\{X_n\}_{n\geq 1}$ *is adapted,*
2. $\mathbf{E}|X_n| < \infty$ *for all* n*,*
3. $\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n$ *for* $n \geq 1$*.*

If in the last condition we substitute $=$ *by* $\geq$ *then sequence is called a* submartingale*. If* $=$ *is replaced by* $\leq$ *then we say sequence forms a* supermartingale*.*

Exercise 9.7 Prove the following statements.

1. If $\{X_n\}_{n\geq 1}$ forms a submartingale, then $\{-X_n\}_{n\geq 1}$ is a supermartingale.
2. $\{X_n\}_{n\geq 1}$ is a martingale iff it is both a submartingale and supermartingale.

3. If $\{X_n\}_{n\geq 1}$ forms a martingale, then for any $1 \leq m \leq n$ we have $\mathbf{E}(X_n|\mathcal{F}_m) = X_m$.

Example 9.1 Here are some examples of martingales.

- Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{E}X_n = 0$ and $\mathbf{E}X_n^2 = 1$, $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, and $S_n = X_1 + \cdots + X_n$. Then both S_n and $S_n^2 - n$ are martingales.

- Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. non-negative random variables with $\mathbf{E}X_n = 1$, and $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then $\prod_{k=1}^n X_k$ is a martingale.

- Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{P}(X = 1) = p$, $\mathbf{P}(X = -1) = q = 1 - p$, where $0 < p < 1$, and $S_n = X_1 + \cdots + X_n$. Then $\left(\frac{q}{p}\right)^{S_n}$ is a martingale.

- Consider a random variable X with finite absolute first moment, and some filtration $\{\mathcal{F}_n\}_{n\geq 1}$. Define $X_n = \mathbf{E}(X|\mathcal{F}_n)$, $n \geq 1$, then X_n forms a martingale with respect to $\{\mathcal{F}_n\}_{n\geq 1}$.

Exercise 9.8 Assume that $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ are submartingales with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 1}$. Show that $\{\max(X_n, Y_n)\}_{n\geq 1}$ is a submartingale.

Exercise 9.9 Assume that $\{X_n\}_{n\geq 1}$ is a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 1}$, and f is a convex Borel function such that $\mathbf{E}|f(X_n)| < \infty$. Show that $\{f(X_n)\}_{n\geq 1}$ is a submartingale.

Exercise 9.10 Assume that $\{X_n\}_{n\geq 1}$ is a submartingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 1}$, and f is a convex, nondecreasing Borel function such that $\mathbf{E}|f(X_n)| < \infty$. Show that $\{f(X_n)\}_{n\geq 1}$ is a submartingale.

Exercise 9.11 Let $\mathcal{X} = \{x_1, x_2, ..., x_n\}$ be a fixed set of real numbers, and let $X_1, X_2, ..., X_n$ be the successive values of a sample of size n that is drawn sequentially without replacement from the set $\mathcal{X}$. Consider the sequence of sigma-fields $\sigma(X_1, ..., X_k)$, $1 \leq k \leq n$ and the partial sums $S_k = X_1 + \cdots + X_k$, $\leq k \leq n$. Show that $(nS_k - S_n k)/(n-k)$, $1 \leq k \leq n-1$ is a martingale.

9.4 Optional Stopping Theorem

Definition 9.6 *We say that a sequence of random variables $\{C_n\}_{n\geq 1}$ is* pre-dictable *with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$ if C_n is $\mathcal{F}_{n-1}$-measurable for every $n \geq 1$.*

Definition 9.7 *Let $\{X_n\}_{n\geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Let $\{C_n\}_{n\geq 1}$ be a predictable sequence. The* martingale transform *of $\{X_n\}_{n\geq 0}$ by $\{C_n\}_{n\geq 1}$, $\{(C \circ X)_n\}_{n\geq 1}$, is define by*

$$(C \circ X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}).$$

Theorem 9.2 (Gambling Theorem) *Let $\{X_n\}_{n\geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$, and let $\{C_n\}_{n\geq 1}$ be a predictable sequence such that $\mathbf{E}|C_n(X_n - X_{n-1})| < \infty$. Then $\{(C \circ X)_n\}_{n\geq 1}$ forms a martingale.*

If $\{X_n\}_{n\geq 0}$ is a supermartingale (submartingale), and $\{C_n\}_{n\geq 1}$ is also non-

negative, then $\{(C \circ X)_n\}_{n\geq 1}$ *is a supermartingale (submartingale).*

Proof. It is obvious that $(C \circ X)_n$ is $\mathcal{F}_n$-measurable and integrable for every $n \geq 1$. Since

$$\mathbf{E}(C_n(X_n - X_{n-1})|\mathcal{F}_{n-1}) = C_n\mathbf{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = 0$$

we get that $\{(C \circ X)_n\}_{n\geq 1}$ is a martingale (that starts at 0).

In a similar way we can prove the second statement.

□

Definition 9.8 *A non-negative integer-valued random* τ *is called a* stopping time *if for every* $n \geq 1$ *event* $\{\tau \leq n\} \in \mathcal{F}_n$.

Example 9.2 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{P}(X = 1) = p$, $\mathbf{P}(X = -1) = q = 1 - p$, where $0 < p < 1$, and $S_n = X_1 + \cdots + X_n$, $S_0 = 0$. Take integers $A, B \geq 1$. Define

$$\tau = \inf\{k : S_k = A \text{ or } S_k = -B\}.$$

The random variable τ is a stopping time.

Example 9.3 If τ_1 and τ_2 are stoping times, then $\tau_1 \wedge \tau_2 = \min(\tau_1, \tau_2)$ is a stopping time.

Theorem 9.3 (Stopped Martingale Theorem) *If* $\{X_n\}_{n\geq 0}$ *is a martingale (super- or sub-) and* τ *is a stopping time, then* $\{X_{n\wedge\tau}\}_{n\geq 0}$ *is a martingale (super- or sub-).*

Proof. Define $C_n = 1_{n \le \tau}$. The random variable C_n takes only two values, 0 or 1, and $\{C_n = 0\} = \{n > \tau\} = \{\tau \le n-1\} \in \mathcal{F}_{n-1}$. Therefore, C_n is $\mathcal{F}_{n-1}$-measurable. But since

$$X_{n\wedge\tau} = X_0 + (C \circ X)_n,$$

we get the results by the gambling theorem. □

Theorem 9.4 (Doob's Optional Stopping Theorem) *Let $\{X_n\}_{n\ge 0}$ be a martingale (super- or sub-) and τ be a stopping time. Then*

$$\mathbf{E}(X_0) = \mathbf{E}(X_\tau), \quad (\ge \ \ \textit{or} \ \ \le)$$

in each of the following situations:

1. *τ is bounded by integer N with probability 1;*
2. *$\sup_n |X_{n\wedge\tau}|$ is bounded with probability 1;*
3. *$\mathbf{E}\tau < \infty$ and $\sup_n |X_n - X_{n-1}|$ is bounded by K with probability 1.*

Proof. Let us prove the theorem for martingales, and leave the rest as an exercise. By the stopped martingale theorem $\{X_{n\wedge\tau}\}_{n\ge 1}$ is a martingale,

$$\mathbf{E}X_{n\wedge\tau} = \mathbf{E}X_0, \tag{9.1}$$

and $X_{n\wedge\tau} \to X_\tau$ with probability 1 as $n \to \infty$. Then for the first situation we take $n = N$ in (9.1). The second and third cases are proved by the dominated convergence theorem. The second situation is trivial. For the third case note that

$$|X_{n\wedge\tau} - X_0| = |\sum_{k=1}^{n\wedge\tau}(X_k - X_{k-1})| < K\tau,$$

and $\mathbf{E}K\tau < \infty$. □

Exercise 9.12 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{P}(X = 1) = p$, $\mathbf{P}(X = -1) = q = 1 - p$, where $0 < p < 1$, and $S_n = X_1 + \cdots + X_n$, $S_0 = 0$. Take integers $A, B \geq 1$. Define

$$\tau = \inf\{k : S_k = A \text{ or } S_k = -B\}.$$

Find $\mathbf{P}(S_\tau = A)$.

Exercise 9.13 (Wald's Identities) Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{E}X_1^2 < \infty$, and τ is a stopping time with respect to $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Assume that τ is bounded by integer N with probability 1. Show that

$$\mathbf{E}(X_1 + \cdots + X_\tau) = \mathbf{E}X_1 \cdot \mathbf{E}\tau,$$

and

$$\mathbf{E}[(X_1 + \cdots + X_\tau) - \tau\mathbf{E}X_1]^2 = \mathbf{Var}(X_1) \cdot \mathbf{E}\tau.$$

Exercise 9.14 Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. random variables with $\mathbf{P}(X = 1) = 1/2$, $\mathbf{P}(X = -1) = 1/2$, and $S_n = X_1 + \cdots + X_n$, $S_0 = 0$. Let $\tau = \inf\{k : S_k = 1\}$. Prove that $\mathbf{E}\tau = \infty$.

9.5 Doob's Convergence Theorem

Definition 9.9 *Let $\{X_n\}_{n\geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. The* number of upcrossings of $[a, b]$ made by $\{X_n\}_{n\geq 0}$ by time N, $U_N[a, b]$, *is defined to be the largest k such that one can find*

$$0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq N$$

such that

$$X_{s_i} < a, \quad X_{t_i} > b, \quad 1 \leq i \leq k.$$

Let us introduce the following predictable process $\{C_n\}_{n\geq 1}$. Define

$$C_1 = 1_{X_0 < a},$$

and for $n \geq 2$

$$C_n = 1_{C_{n-1}=1} 1_{X_{n-1}\leq b} + 1_{C_{n-1}=0} 1_{X_{n-1}<a}.$$

Let Y_n will be martingale transform of X_n, that is, $Y_n = (C \circ X)_n$. Let us explain the meaning of Y_n. Suppose that $X_n - X_{n-1}$ represent our winnings per unit stake in round n. Our gambling strategy is as follows. First we wait till $\{X_n\}$ is below a, then we play unit stakes until $\{X_n\}$ gets above b. Once it is above we stop and wait till it is below a, and then start playing again, and so on. The process $\{Y_n\}$ represent our total winnings. Figure 9.5 provides an illustration. It is easy to see that

$$Y_N \geq (b-a)U_N[a,b] - (X_N - a)^-. \tag{9.2}$$

Lemma 9.1 (Doob's Upcrossing Lemma) *For any $N \geq 0$ we have*

$$(b-a)\mathbf{E}U_N[a,b] \leq \mathbf{E}(X_N - a)^-.$$

Proof. Just observe that $\{Y_n\}$ is a martingale with $\mathbf{E}Y_N = 0$. $\square$

Theorem 9.5 (Doob's Convergence Theorem) *Let $\{X_n\}_{n\geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$ such that*

$$\sup_n \mathbf{E}|X_n| < \infty.$$

Then $\lim_n X_n$ exists and is finite with probability 1.

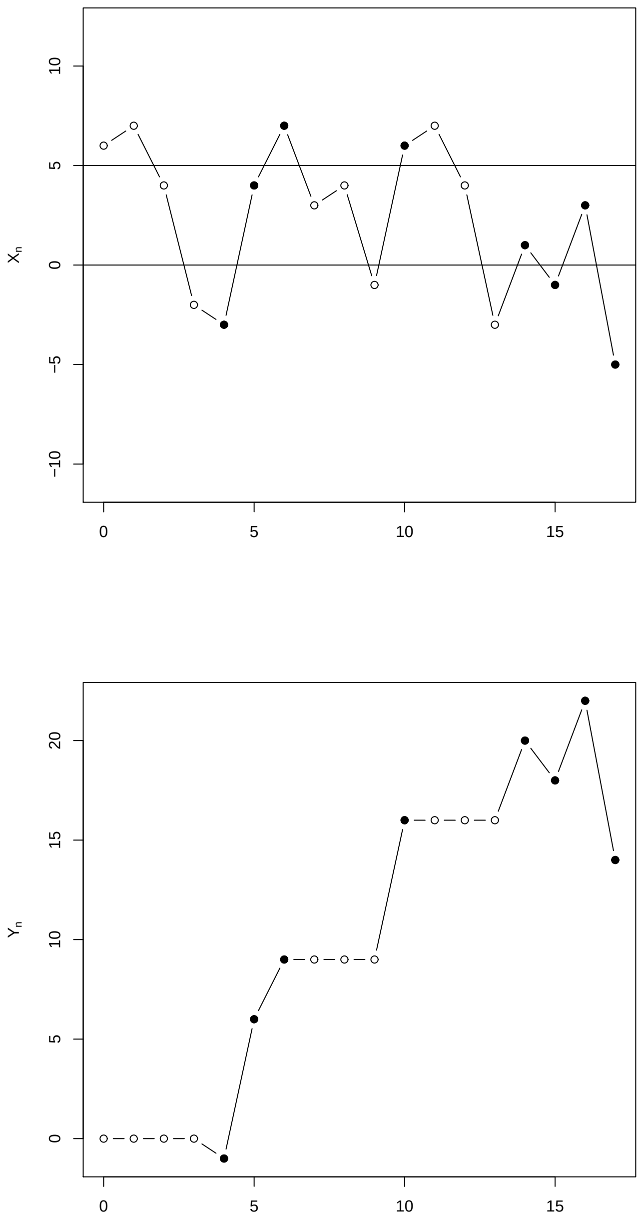
Proof. Let $a < b$. Define $U_\infty[a,b] = \lim_n U_n[a,b]$. By Lemma 9.1 we have

$$(b-a)\mathbf{E}U_N[a,b] \leq |a| + \mathbf{E}|X_N| \leq |a| + \sup_n \mathbf{E}|X_n|.$$

By monotone convergence theorem we get that

$$\mathbf{E}U_\infty[a,b] < \infty,$$

Figure 9.1: Upcrossings picture.



and, therefore,

$$\mathbf{P}(U_\infty[a,b] = \infty) = 0.$$

Now, note that

$$\begin{aligned}
A &= \{\omega : X_n(\omega) \text{ does not converge to a limit in } [-\infty, +\infty]\} \\
&= \{\omega : \liminf_n X_n(\omega) < \limsup_n X_n(\omega)\} \\
&= \bigcup_{a<b,\, a,b\in\mathbb{Q}} \{\omega : \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\}.
\end{aligned}$$

Since

$$\{\omega : \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\} \subset \{\omega : U_\infty[a,b](\omega) = \infty\},$$

by the countability we get that $\mathbf{P}(A) = 0$. Hence the limit of X_n exists a.s. in $[-\infty, +\infty]$. Let $X_\infty(\omega) = \lim_n X_n(\omega)$. Finally, by Fatou's lemma we obtain

$$\mathbf{E}|X_\infty| = \mathbf{E}\liminf_n |X_n| \le \liminf_n \mathbf{E}|X_n| \le \sup_n \mathbf{E}|X_n| < \infty,$$

so that

$$\mathbf{P}(X_\infty \text{ is finite}) = 1.$$

□

Note here that we *do not* have convergence in L_1.

Exercise 9.15 Prove Doob's convergence theorem for supermartingales.

Exercise 9.16 Let $\{X_n\}_{n\geq 0}$ be a non-negative martingale, then $\lim_n X_n$ exists and finite with probability 1.

9.6 L_2-Bounded Martingales

Let $\{X_n\}_{n\geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$, and suppose that $\mathbf{E}X_n^2 < \infty$ for every n.

Lemma 9.2 (Pythagoras's Formula)

$$\mathbf{E}X_n^2 = \mathbf{E}X_0^2 + \sum_{k=1}^{n} \mathbf{E}(X_k - X_{k-1})^2.$$

Proof. Let $s \leq t \leq u \leq v$. Then

$$\begin{aligned}
\mathbf{E}(X_v - X_u)(X_t - X_s) &= \mathbf{E}[\mathbf{E}\left((X_v - X_u)(X_t - X_s)|\mathcal{F}_u\right)] \\
&= \mathbf{E}[(X_t - X_s)\mathbf{E}\left((X_v - X_u)|\mathcal{F}_u\right)] \\
&= \mathbf{E}[(X_t - X_s)(X_u - X_u)] \\
&= 0.
\end{aligned}$$

That is, the formula

$$X_n = X_0 + \sum_{k=1}^{n}(X_k - X_{k-1})$$

expresses X_n as the sum of orthogonal terms. $\square$

Theorem 9.6 (L_2-Bounded Martingale Convergence) *Let $\{X_n\}_{n\geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 0}$ with $\mathbf{E}X_n^2 < \infty$ for every n.*

Then

$$\sup_n \mathbf{E} X_n^2 < \infty \tag{9.3}$$

if and only if

$$\sum_k \mathbf{E}(X_k - X_{k-1})^2 < \infty.$$

Moreover, if (9.3) holds then X_n converges a.s. and in L_2.

Proof. The first statement follows from Pythagoras's formula. If (9.3) holds then

$$\sup_n \mathbf{E}|X_n| < \infty,$$

and by Doob's convergence theorem we get that $\lim_n X_n$ exists a.s. Let $X_\infty = \lim_n X_n$. By Pythagoras's formula we have that

$$\mathbf{E}(X_{n+k} - X_n)^2 = \sum_{i=n+1}^{n+k} \mathbf{E}(X_i - X_{i-1})^2.$$

Therefore, by Fatou's Lemma we obtain

$$\mathbf{E}(X_\infty - X_n)^2 \leq \sum_{i=n+1}^{\infty} \mathbf{E}(X_i - X_{i-1})^2,$$

and, as a consequence,

$$\lim_n \mathbf{E}(X_\infty - X_n)^2 = 0.$$

□

Theorem 9.7 (Doob's Decomposition) *Any submartingale $\{X_n\}_{n\geq 0}$ can be* *written in a unique way as*

$$X_n = M_n + A_n,$$

where M_n is a martingale and A_n is an is a predictable increasing sequence with $A_0 = 0$.

Proof. Let

$$A_n - A_{n-1} = \mathbf{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1}$$

and

$$M_n = X_n - A_n.$$

It is clear that A_n is $\mathcal{F}_{n-1}$-measurable, and it is increasing because X_n is submartingale. Now, note that

$$\begin{aligned}
\mathbf{E}(M_n|\mathcal{F}_{n-1}) &= \mathbf{E}(X_n|\mathcal{F}_{n-1}) - A_n \\
&= X_{n-1} + A_n - A_{n-1} - A_n \\
&= X_{n-1} - A_{n-1} \\
&= M_{n-1},
\end{aligned}$$

that is M_n is a martingale.

To prove uniqueness, assume that there is another decomposition $X_n = M'_n + A'_n$. Then

$$\mathbf{E}(X_n|\mathcal{F}_{n-1}) = M'_{n-1} + A'_n = X_{n-1} - A'_{n-1} + A'_n.$$

This means that

$$A'_n - A'_{n-1} = \mathbf{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} = A_n - A_{n-1},$$

and $M'_n = M_n$. $\square$

If $\{X_n\}_{n\geq 0}$ is a L_2-bounded martingale with $X_0 = 0$, then by Jensen's inequality $\{X_n^2\}_{n\geq 0}$ is a submartingale. In this case

$$A_n - A_{n-1} = \mathbf{E}[(X_n^2 - X_{n-1}^2)|\mathcal{F}_{n-1}] = \mathbf{E}[(X_n - X_{n-1})^2|\mathcal{F}_{n-1}].$$

Since $\mathbf{E}X_n^2 = \mathbf{E}A_n$ we also get that

$$\sup_n \mathbf{E}X_n^2 < \infty$$

if and only if

$$\mathbf{E}(\sup_n A_n) < \infty.$$

Exercise 9.17 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with $\mathbf{E}X_n = 0$ and $\mathbf{E}X_n^2 < \infty$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, and $S_n = X_1 + \cdots + X_n$. What is A_n for the submartingale S_n^2?

9.7 UI Martingales

Theorem 9.8 (UI Martingale Convergence) *Let $\{X_n\}_{n\geq 1}$ be a martingale with respect to a filtration $\{\mathcal{F}_n\}_{n\geq 1}$, and assume that $\{X_n\}_{n\geq 1}$ is uniformly* *integrable. Then* X_n *converges a.s. and in* L_1*. Moreover, if we denote* $X_\infty = \lim_n X_n$*, then*

$$X_n = \mathbf{E}(X_\infty|\mathcal{F}_n).$$

Proof. The uniform integrability implies that

$$\sup_n \mathbf{E}|X_n| < \infty.$$

Therefore, by Doob's convergence theorem we get that $\lim_n X_n$ exists with probability 1. Because of uniform integrability we also have $X_n \xrightarrow{L_1} X_\infty$.

Now let us show that

$$X_n = \mathbf{E}(X_\infty|\mathcal{F}_n).$$

Note that for any $A \in \mathcal{F}_n$ and $k \geq n$ we have

$$\mathbf{E}X_k 1_A = \mathbf{E}[\mathbf{E}(X_k 1_A|\mathcal{F}_n)] = \mathbf{E}[\mathbf{E}(X_k|\mathcal{F}_n)1_A] = \mathbf{E}X_n 1_A.$$

But since $X_k 1_A \xrightarrow{L_1} X_\infty 1_A$ as well we get that

$$\mathbf{E}X_\infty 1_A = \mathbf{E}X_n 1_A.$$

□

Lemma 9.3 *Given are a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, a filtration* $\{\mathcal{F}_n\}_{n\geq 0}$*, and a random variable* X *with* $\mathbf{E}|X| < \infty$*. Let* $X_n = \mathbf{E}(X|\mathcal{F}_n)$*. Then* $\{X_n\}_{n\geq 0}$ *is a UI martingale.*

Proof. We only need to check that $\{X_n\}_{n\geq 0}$ is UI. Note that by Jensen's inequality we have $|X_n| \leq \mathbf{E}(|X||\mathcal{F}_n)$ a.s. and $\mathbf{E}|X_n| \leq \mathbf{E}|X|$, hence for any $a, b > 0$ we have

$$
\begin{aligned}
\mathbf{E}|X_n|1_{|X_n|\geq a} &\leq \mathbf{E}[1_{|X_n|\geq a}\mathbf{E}(|X||\mathcal{F}_n)] \\
&= \mathbf{E}[\mathbf{E}(|X|1_{|X_n|\geq a}|\mathcal{F}_n)] \\
&= \mathbf{E}|X|1_{|X_n|\geq a} \\
&= \mathbf{E}|X|1_{|X_n|\geq a}1_{|X|\leq b} + \mathbf{E}|X|1_{|X_n|\geq a}1_{|X|>b} \\
&\leq b\mathbf{P}(|X_n| \geq a) + \mathbf{E}|X|1_{|X|>b} \\
&\leq \frac{b}{a}\mathbf{E}|X_n| + \mathbf{E}|X|1_{|X|>b} \\
&\leq \frac{b}{a}\mathbf{E}|X| + \mathbf{E}|X|1_{|X|>b}.
\end{aligned}
$$

Choosing large b first, and then large a, we can make the RHS (that bounds the LHS uniformly w.r.t. n) as small as we want. □

Theorem 9.9 (Levy's Convergence Theorem) *Given are a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, a filtration $\{\mathcal{F}_n\}_{n\geq 0}$, and a random variable X with $\mathbf{E}|X| < \infty$. Let $X_n = \mathbf{E}(X|\mathcal{F}_n)$. Then $X_n \to \mathbf{E}(X|\mathcal{F}_\infty)$ a.s and L_1.*

Proof. By the UI martingale convergence theorem we have that X_n converge a.s. and in L_1. Let $Z = \lim_n X_n$ and $Y = \mathbf{E}(X|\mathcal{F}_\infty)$. All we need to show is that two $\mathcal{F}_\infty$-measurable Z and Y are a.s. equal. Without loss of generality, we assume that $X \geq 0$. Consider two measures on $(\Omega, \mathcal{F}_\infty)$:

$$\mu_Z(A) = \mathbf{E}Z1_A, \quad \text{and } \mu_Y(A) = \mathbf{E}Y1_A.$$

By the tower property and the UI martingale convergence theorem

$$\mathbf{E}(Y|\mathcal{F}_n) = \mathbf{E}(X|\mathcal{F}_n) = X_n = \mathbf{E}(Z|\mathcal{F}_n).$$

Thus for any $A \in \mathcal{F}_n$ we have

$$\mathbf{E}Y1_A = \mathbf{E}X_n1_A = \mathbf{E}Z1_A.$$

This means that two measure μ_Z and μ_Y coincide on field $\cup_n \mathcal{F}_n$, hence, by the set induction they must coincide on $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. Thus $Z = Y$ a.s. $\square$

9.8 Martingale Inequalities

Here is a submartingale version of Kolmogorov's inequality.

Theorem 9.10 (Doob's Submartingale Inequality) *Let $\{X_n\}_{n\geq 1}$ be a non-negative submartingale, and $X_n^* = \max_{1\leq k\leq n} X_k$. Then*

$$c\mathbf{P}\left(X_n^* \geq c\right) \leq \mathbf{E}\left(X_n 1_{X_n^* \geq c}\right) \leq \mathbf{E}\left(X_n\right).$$

Proof. Consider the following events:

$$\begin{aligned}
A \ &= \{\omega : X_n^*(\omega) \geq c\}, \\
A_1 &= \{\omega : X_1(\omega) \geq c\}, \\
A_2 &= \{\omega : X_1(\omega) < c, X_2(\omega) \geq c\}, \\
&\ \ \ldots \\
A_n &= \{\omega : X_1(\omega) < c, \ldots, X_{n-1}(\omega) < c, X_n(\omega) \geq c\}.
\end{aligned}$$

It is obvious that $A = \cup_{k=1}^n A_k$, and $A_k A_l = \emptyset$ if $k \neq l$. Now, $A_k \in \mathcal{F}_k$, and $X_k \geq c$ on A_k. Therefore,

$$\mathbf{E} X_n 1_{A_k} = \mathbf{E}[\mathbf{E}(X_n 1_{A_k} | \mathcal{F}_k)] = \mathbf{E}[1_{A_k} \mathbf{E}(X_n | \mathcal{F}_k)] \geq \mathbf{E}[1_{A_k} X_k] \geq c\mathbf{P}(A_k).$$

Summing over k finishes the proof. $\square$

Lemma 9.4 *For any non-negative X and $p > 0$ we have*

$$\mathbf{E} X^p = p \int_0^\infty t^{p-1} \mathbf{P}(X \geq t) dt.$$

Proof. Exercise. $\square$

Theorem 9.11 (Doob's L_p Maximal Inequality) *Let $\{X_n\}_{n \geq 1}$ be a non-negative submartingale, $X_n^* = \max_{1 \leq k \leq n} X_k$, and $p > 1$. Denote $|| \cdot ||_p = [\mathbf{E} | \cdot |^p]^{1/p}$. Then*

$$||X_n^*||_p \leq \frac{p}{p-1} ||X_n||_p.$$

Proof. Assume first that $||X_n^*||_p < \infty$. By Doob's inequality and Fubini's theorem we get

$$\begin{aligned}
\mathbf{E}(X_n^*)^p &= p\int_0^\infty t^{p-1}\mathbf{P}(X_n^* \geq t)dt \\
&\leq p\int_0^\infty t^{p-2}\mathbf{E}X_n 1_{X_n^*\geq t}dt \\
&\leq p\int_0^\infty t^{p-2}\int_\Omega X_n 1_{X_n^*\geq t}d\mathbf{P}dt \\
&= p\int_\Omega X_n\int_0^\infty t^{p-2}1_{X_n^*\geq t}dtd\mathbf{P} \\
&= p\int_\Omega X_n\int_0^{X_n^*} t^{p-2}dtd\mathbf{P} \\
&= \frac{p}{p-1}\mathbf{E}[X_n(X_n^*)^{p-1}].
\end{aligned}$$

Hence, by Hölder's inequality we obtain (here $q = p/(p-1)$)

$$\mathbf{E}(X_n^*)^p \leq q\mathbf{E}[X_n(X_n^*)^{p-1}] \leq q||X_n||_p||(X_n^*)^{p-1}||_q = q||X_n||_p[\mathbf{E}(X_n^*)^p]^{1/q}.$$

Dividing both sides of the inequality by $[\mathbf{E}(X_n^*)^p]^{1/q}$ gives us the result.

Finally, note that if $||X_n||_p = \infty$, then the inequality is obvious. However, if $\mathbf{E}||X_n||_p < \infty$, then $||X_n^*||_p < \infty$ as well (prove it!). $\square$

THE END

Bibliography

1. Billingsley, P. (1995) *Probability and measure*, Wiley-Interscience, New York, 3rd Edition.

2. Durrett, R. (2005) *Probability: Theory and Examples*, Thomson Learning, New York, 3rd Edition.

3. Feller, W. (1968) *An Introduction to probability theory and its application*, Wiley & Sons, New York, 3rd Edition.

4. Jacod J. and Protter P. (2003) *Probability Essentials*, Springer; New York, 2nd Edition.

5. Kolmogorov, A.N. and Fomin, S.V. (1970) *Introductory real analysis*, Dover Publications, New York.

6. Resnick, S.I. (1998) *A probability path*, Birkhauser, Boston.

7. Shiryayev, A.N. (1995) *Probability*, Springer-Verlag, New York, 2nd Edition.

8. Steele, J.M. (2004) *The Cauchy-Schwarz Master Class*, Cambridge University Press, Cambridge.

9. Williams, D. (1991) *Probability with martingales*, Cambridge University Press, Cambridge.

Index