IERG 6300: Theory of Probability
Spring 2022 - Lecture Notes

Chandra Nair

Abstract. These are class notes in probability theory. Most of these sections are taken from Varadhan's lecture notes as well as Durrett's book.

Contents

CHAPTER 1

Basic Ideas of Probability Theory

1.1. Preliminaries

1.1.1. Motivation. *Basic notion*: One wants to assign chances to outcomes or groups of outcomes of an experiment.

If the number of possible outcomes are finite, say $\omega_1, ..., \omega_n$, then a natural way to assign chances or probabilities is to assign for each outcome a number $\mathrm{P}(\omega_i) = p_i$, where $0 \leq p_i \leq 1$ and $\sum_{i=1}^n p_i = 1$. One can extend this line of thought reasonably to even experiments with *countable* number of outcomes.

However things become a bit complicated when the number of outcomes become uncountable. There is no reasonable way to assign positive probabilities to an uncountable number of outcomes, and still make their sum to be 1. To see this consider the following collection of events $\mathcal{E}_k = \{w : \mathrm{P}(w) \geq \frac{1}{k}\}$. Clearly $|\mathcal{E}_k| \leq k$. Let $\mathcal{E} = \cup_k \mathcal{E}_k = \{w : \mathrm{P}(w) > 0\}$. However, it is clear that the number of elements in $\mathcal{E}$ is countable.

A Remedy: Instead of defining probabilities for individual outcomes, one defines probabilities for collections of outcomes (or *events*).

Let $\mathcal{A}$ denote the collection of events for which probabilities are assigned. Then one would like $\mathcal{A}$ to have the following properties:

- if $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$ ("or" event), and $A \cap B \in \mathcal{A}$ ("and" event)
- $\emptyset, \Omega \in \mathcal{A}$
- if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$

Such a collection $\mathcal{A}$ is called an *algebra* or a *field*. We may also require the set of events for which we wish to assign probabilities to have the following additional property:

- If $A_i \in \mathcal{F}$ then $\cup_i A_i \in \mathcal{F}$.

An algebra or a field with (closed under countable unions) is called a *sigma*-algebra or a *sigma*-field.

Aside: Why can't we then assign probabilities to all subsets of outcomes, i.e. to the power set 2^Ω. Let us take $\Omega = \mathbb{R}^3$. Assume that we want the measure to be constructed to yield the "volume" of the set. There is a rather deep result that says that we must make one of the following concessions:

(1) The volume of a set might change when it is rotated.
(2) The volume of the union of two disjoint sets might be different from the sum of their volumes.
(3) Some sets might be tagged "non-measurable", and one would need to check whether a set is "measurable" before talking about its volume.
(4) The axioms of ZFC (Zermelo-Fraenkel set theory with the axiom of Choice) might have to be altered.

Most probabilists choose to accept (3), i.e. to tag certain sets as "unmeasurable". See `https://en.wikipedia.org/wiki/Non-measurable_set`

In any case, in this class we will talk about the standard axiomatic treatment under the assumption (3).

1.1.2. Definitions.

DEFINITION 1.1.1. A σ-algebra is a collection $\mathcal{F}$ of events $A \subset \Omega$ such that

- $\emptyset, \Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- $A_i \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a *measurable space*.

DEFINITION 1.1.2. A *measure* μ on the measurable space $(\Omega, \mathcal{F})$ is a mapping $\mu : \mathcal{F} \mapsto [0, \infty]$ satisfying:

(1) $\mu(\emptyset) = 0$.
(2) if $A_i \in \mathcal{F}$ is a countable collection of pairwise disjoint sets, then $\mu(\sqcup_i A_i) = \sum_i \mu(A_i)$.

DEFINITION 1.1.3. A *p*robability measure is a measure that satisfies $\mathrm{P}(\Omega) = 1$.

EXERCISE 1.1.1. Show that the following holds for a probability measure:

(1) $\mathrm{P}(\cup_i A_i) \le \sum_i \mathrm{P}(A_i)$ (sub-additivity)
(2) if $A_i \uparrow A$, then $\mathrm{P}(A_i) \uparrow \mathrm{P}(A)$ (monotone up-limits)
(3) if $A_i \downarrow A$, then $\mathrm{P}(A_i) \downarrow \mathrm{P}(A)$ (monotone down-limits)

EXERCISE 1.1.2. Show that $(1), (2)$ holds for general measures, while (3) may not.

(Hint: To see a counter example to (3), define $B_i = p_i \mathcal{N}$ (here p_i is the ith prime) and set $A_i = \mathcal{N} \setminus \cup_{j=1}^{i} B_i$. Use counting measure as the measure and show that $\mu(A_i) = \infty, \forall i$, while $A = \{1\}$, and hence $\mu(A) = 1$.)

REMARK 1.1.1. If A_i is a collection such that $\mathrm{P}(A_i) = 0, \forall i$ then $\mathrm{P}(\cup_i A_i) = 0$ (from subadditivity).

EXERCISE 1.1.3. Consider events $\{A_n\}$ in a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ that are almost pairwise disjoint, i.e. $\mathrm{P}(A_n \cap A_m) = 0$ whenever $n \neq m$. Show that

$$\mathrm{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

The triple $(\Omega, \mathcal{F}, \mathrm{P})$ is called a *probability space*.

DEFINITION 1.1.4. $(\Omega, \mathcal{F}, \mathrm{P})$ is said to be *non-atomic* if $\forall A$ s.t. $\mathrm{P}(A) > 0, \exists B \subset A, B \in \mathcal{F}$ s.t. $\mathrm{P}(A) > \mathrm{P}(B) > 0$.

EXERCISE 1.1.4. If $(\Omega, \mathcal{F}, \mathrm{P})$ is non-atomic and $\mathrm{P}(A) > 0$ then show that

(1) $\forall \epsilon > 0, \exists B \subset A$ s.t. $0 < \mathrm{P}(B) < \epsilon$.
(2) If $0 < a < \mathrm{P}(A)$ then there exists $B \subset A$ s.t. $\mathrm{P}(B) = a$.

Consider a collection of events $\mathcal{A}$. We wish to *extend* the collection to a sigma field. Or in other words is there a smallest σ-field that contains $\mathcal{A}$. The answer is Yes. To see this, the following exercise is useful.

Exercise 1.1.5. Let $\mathcal{F}_\alpha$ be an arbitrary collection of σ-fields. Then $\cap_\alpha \mathcal{F}_\alpha$ is also a σ-field.

From the above exercise the intersection of all σ-algebras that contain $\mathcal{A}$ is a σ-algebra and is clearly the smallest σ-algebra that contains $\mathcal{A}$. This is denoted as $\sigma(\mathcal{A})$, and called the σ-algebra generated by $\mathcal{A}$.

Supposed Ω is a topological space (i.e. equipped with the notion of open sets). Then the σ-algebra generated by the open sets is called the *Borel σ-algebra.*

Exercise 1.1.6. Given a collection of sets $\mathcal{A}$, let $\sigma(\mathcal{A})$ denote the smallest σ-field containing the elements of $\mathcal{A}$. Verify the following alternate definitions for Borel σ-field $\mathcal{B}_R$ of reals (i.e. show that all the following σ-fields are identical):

- $\sigma(\{(a,b) : a < b \in \mathbb{R}\})$
- $\sigma(\{[a,b] : a < b \in \mathbb{R}\})$
- $\sigma(\{(-\infty, b) : b \in \mathbb{R}\}$
- $\sigma(\{(-\infty, b) : b \in \mathbb{Q}\})$
- $\sigma(\{\mathcal{O} \subset \mathbb{R} \text{ is open}\})$.

1.1.3. Other classes of sets. Besides σ-algebra's and algebras, other collections of sets with certain properties also turn out to be useful in probability theory.

Definition 1.1.5. A collection $\mathcal{M}$, of subsets of Ω, is called a monotone class, if $\mathcal{M}$ is non-empty, and satisfies the following two conditions:

- If a countable increasing collection of sets, $M_1 \subseteq M_2 \subseteq \cdots$, belong to $\mathcal{M}$, then $\cup_i M_i \in \mathcal{M}$. In other words $\mathcal{M}$ is closed under monotone up limits.
- If a countable decreasing collection of sets, $M_1 \supseteq M_2 \supseteq \cdots$, belong to $\mathcal{M}$, then $\cap_i M_i \in \mathcal{M}$. In other words $\mathcal{M}$ is closed under monotone down limits.

Remark 1.1.2. An example of a monotone class on the reals is the collection $\{(0,1], (2,3]\}$. This is clearly not closed under unions or intersections. Note that increasing unions can only be formed by taking identical elements and hence it is a monotone class.

Remark 1.1.3. It is clear that arbitrary intersections of monotone classes in a monotone class. Therefore give a collection of sets $\mathcal{S}$, we can talk about the monotone class generated by $\mathcal{S}$ denoted by $\mathcal{M}(\mathcal{S})$, to be the smallest monotone class that contains $\mathcal{S}$.

Theorem 1.1.1 (Monotone Class Theorem). *Let $\mathcal{A}$ be an algebra. Let $\mathcal{M}(\mathcal{A})$ be the monotone class generated by $\mathcal{A}$. Then $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$.*

Proof. Since $\sigma(\mathcal{A})$ is closed under monotone limits, it is clear that $\sigma(\mathcal{A}) \supseteq \mathcal{M}(\mathcal{A})$. Therefore suffices to show that $\mathcal{M}(\mathcal{A})$ is a σ-algebra. Since $\emptyset \in \mathcal{A}$ we have $\emptyset \in \mathcal{M}(\mathcal{A})$.

Let $\mathcal{M}_0 = \{B : B \in \mathcal{M}(\mathcal{A}) \text{ such that } B^c \in \mathcal{M}(\mathcal{A})\}$. Clearly $\mathcal{A} \subseteq \mathcal{M}_0$. Since $\mathcal{M}(\mathcal{A})$ is closed under monotone limits, it is clear that $\mathcal{M}_0$ is also a monotone class, hence $\mathcal{M}_0 = \mathcal{M}(\mathcal{A})$. Thus $B \in \mathcal{M}(\mathcal{A})$ implies that $B^c \in \mathcal{M}(\mathcal{A})$.

To complete the proof, it suffices to show that $\mathcal{M}(\mathcal{A})$ is a field. Towards this, fix $A \in \mathcal{A}$, and define

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

Clearly $\mathcal{A} \subseteq \mathcal{M}_A$ and $\mathcal{M}_A$ is a monotone class. Hence $\mathcal{M}_A = \mathcal{M}(\mathcal{A})$.

Now fix $C \in \mathcal{M}(\mathcal{A})$ and define

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

Clearly $\mathcal{A} \subseteq \mathcal{M}_C$ and $\mathcal{M}_C$ is a monotone class. Hence $\mathcal{M}_C = \mathcal{M}(\mathcal{A})$.

Therefore $\mathcal{M}(\mathcal{A})$ is closed under finite unions and intersections, hence an algebra. Now closure under monotone limits follows from definition implying that $\mathcal{M}(\mathcal{A})$ is a σ-algebra. □

DEFINITION 1.1.6. A collection $\mathcal{H}$, of subsets of Ω, is called a π-system, if $\mathcal{H}$ is non-empty, and $H_1, H_2 \in \mathcal{H}$ implies that $H_1 \cap H_2$ belongs to $\mathcal{H}$.

DEFINITION 1.1.7. A collection $\mathcal{D}$, of subsets of Ω, is called a λ-system (or a d-system), if:

- $\Omega \in \mathcal{D}$,
- $D_1, D_2 \in \mathcal{D}$ and $D_1 \subseteq D_2$ implies $(D_2 \setminus D_1) \in \mathcal{D}$,
- If a countable increasing collection of sets, $D_1 \subseteq D_2 \subseteq \cdots$, belong to $\mathcal{D}$, then $\cup_i D_i \in \mathcal{D}$. In other words $\mathcal{D}$ is closed under monotone up limits.

REMARK 1.1.4. It is clear that arbitrary intersections of λ-systems is a λ-system. Therefore give a collection of sets $\mathcal{S}$, we can talk about the λ-system generated by $\mathcal{S}$ denoted by $\mathcal{D}(\mathcal{S})$, to be the smallest λ-system that contains $\mathcal{S}$.

REMARK 1.1.5. Since $\Omega \in \mathcal{D}$, if $D \in \mathcal{D}$ we have $\Omega \setminus D - D6c \in \mathcal{D}$. Therefore it is closed under taking complements, and consequently $\mathcal{D}$ is also closed under monotone down limits.

THEOREM 1.1.2 (Dynkin's π-λ Theorem). *Let $\mathcal{H}$ be a π-system. Let $\mathcal{D}(\mathcal{H})$ be the λ-system generated by $\mathcal{H}$. Then $\mathcal{D}(\mathcal{H}) = \sigma(\mathcal{H})$.*

PROOF. It is immediate that any σ-algebra is also a λ-system. Hence it suffices to show the non-trivial direction, i.e. to show that $\mathcal{D}(\mathcal{H})$ is a σ-algebra. If $\mathcal{D}(\mathcal{H})$ was additionally π-system, then for any $\{D_i\} \in \mathcal{D}(\mathcal{H})$, we have $\cup_{i=1}^k D_i = \left(\cap_{i=1}^k D_i^c\right)^c \in \mathcal{D}(\mathcal{H})$. Further as $\mathcal{D}(\mathcal{H})$ is closed under monotone up-limits, $\cup_i D_i = \cup_k \left(\cup_{i=1}^k D_i\right) \in \mathcal{D}(\mathcal{H})$. Thus, it suffices to show that $\mathcal{D}(\mathcal{H})$ was additionally π-system.

Define $\mathcal{D}_1 = \{D \in \mathcal{D}(\mathcal{H}) : D \cap H \in \mathcal{D}(\mathcal{H}) \ \forall H \in \mathcal{H}\}$. Observe that $\Omega \in \mathcal{D}_1$, and since $\mathcal{H}$ is a π-system, $\mathcal{H} \subseteq \mathcal{D}_1$. Further, if $D_2, D_1 \in \mathcal{D}_1$ and $D_2 \supseteq D_1$, as $(D_2 \setminus D_1) \cap H = (D_2 \cap H) \setminus (D_1 \cap H)$, we have that $(D_2 \setminus D_1) \in \mathcal{D}_1$. In the above we used that $\mathcal{D}(\mathcal{H})$ is a λ-system, and that $(D_2 \setminus H) \supseteq (D_1 \setminus H)$, as $D_2 \supseteq D_1$. Finally if a countable increasing collection of sets, $D_1 \subseteq D_2 \subseteq \cdots$, belong to $\mathcal{D}_1$, then as $(\cup_i D_i) \cap H = \cup_i (D_i \cap H) \in \mathcal{D}$, we have that $\cup_i D_i \in \mathcal{D}_1$. This implies that $\mathcal{D}_1 \subseteq \mathcal{D}(\mathcal{H})$ is a λ-system containing $\mathcal{H}$, implying $\mathcal{D}_1 = \mathcal{D}(\mathcal{H})$.

Define $\mathcal{D}_2 = \{\hat{D} \in \mathcal{D}(\mathcal{H}) : \hat{D} \cap D \in \mathcal{D}(\mathcal{H}) \ \forall D \in \mathcal{D}(\mathcal{H})\}$. Observe that $\Omega \in \mathcal{D}_2$, and from the previous part $\mathcal{H} \subseteq \mathcal{D}_2$. Now, if $D_2, D_1 \in \mathcal{D}_2$ and $D_2 \supseteq D_1$, as $(D_2 \setminus D_1) \cap D = (D_2 \cap D) \setminus (D_1 \cap D)$, we have that $(D_2 \setminus D_1) \in \mathcal{D}_2$. In the above we used that $\mathcal{D}(\mathcal{H})$ is a λ-system, and that $(D_2 \setminus D) \supseteq (D_1 \setminus D)$, as $D_2 \supseteq D_1$. Finally if a countable increasing collection of sets, $D_1 \subseteq D_2 \subseteq \cdots$, belong to $\mathcal{D}_1$, then as $(\cup_i D_i) \cap D = \cup_i (D_i \cap D) \in \mathcal{D}$, we have that $\cup_i D_i \in \mathcal{D}_2$. This implies that $\mathcal{D}_2 \subseteq \mathcal{D}(\mathcal{H})$ is a λ-system containing $\mathcal{H}$, implying $\mathcal{D}_2 = \mathcal{D}(\mathcal{H})$.

Therefore $\mathcal{D}(\mathcal{H})$ is a λ-system, completing the proof. □

1.1.4. Existance and construction of measures.

THEOREM 1.1.3. *(Caratheodory Extension Theorem)*
Any countably additive probability measure P *on a field* $\mathcal{A}$ *extends uniquely as a countably additive probability measure on* $\mathcal{F} = \sigma(\mathcal{A})$.

PROOF. The proof consists of various steps: Here we outline the main steps

1. Define a quantity $\mathrm{P}^*(B), B \in \mathcal{F}$ (which will turn out an outer measure) as follows:

$$\mathrm{P}^*(B) = \inf_{A_i \in \mathcal{A}, \cup_i A_i \supseteq B} \sum_i \mathrm{P}(A_i).$$

(Show that without loss of generality can assume A_i to be pairwise disjoint.)

Remark: We could also have extended the definition of $\mathrm{P}^*(B)$ to all $B \in 2^\Omega$ but it really does not buy us anything in the context of this proof.

Show that

(1) Subadditivity: $\mathrm{P}^*(\cup_i B_i) \leq \sum_i \mathrm{P}^*(B_i)$. (Hint: A collection of "good" covers of B_i is only one possible cover for $\cup_i B_i$.)
(2) if $A \in \mathcal{A}$ then $\mathrm{P}^*(A) \leq \mathrm{P}(A)$ (trivial)
(3) if $A \in \mathcal{A}$ then $\mathrm{P}^*(A) \geq \mathrm{P}(A)$ (Hint: take a good cover of A and use countable additivity of P on $\mathcal{A}$.)

2. Define a set $E \in \mathcal{F}$ to be measurable if for all $B \in \mathcal{F}$

$$\mathrm{P}^*(B) \geq \mathrm{P}^*(B \cap E) + \mathrm{P}^*(B \cap E^c)$$

(clearly from subadditivity, this forces an equality.) Let $\mathcal{E}$ be the set of all measurable sets. Show that $\mathcal{E}$ is a σ-field and that P^* is a countably additive probability measure on $\mathcal{E}$. Finally show that $\mathcal{E} \supseteq \mathcal{A}$ and hence $\mathcal{E} \supseteq \sigma(\mathcal{A})$.

Outline of proof: Clearly $\mathrm{P}^*(\emptyset) = 0$, hence $\emptyset \in \mathcal{E}$. It is also clear that if $E \in \mathcal{E}$ then $E^c \in \mathcal{E}$. Now suppose $E_1, E_2 \in \mathcal{E}$. Then observe that

$$\begin{aligned}
\mathrm{P}^*(B) &= \mathrm{P}^*(B \cap E_1) + \mathrm{P}^*(B \cap E_1^c) \\
&= \mathrm{P}^*(B \cap E_1) + \mathrm{P}^*(B \cap E_1^c \cap E_2) + \mathrm{P}^*(B \cap E_1^c \cap E_2^c) \\
&\geq \mathrm{P}^*(B \cap (E_1 \cup E_2)) + \mathrm{P}^*(B \cap E_1^c \cap E_2^c).
\end{aligned}$$

Here the equalities follow from the measurability of E_1, E_2, and the inequality follows from sub-additivity of $\mathrm{P}^*(\cdot)$. This implies finite unions are measurable. We can conclude that $\mathcal{E}$ is an algebra.

Let E_i be a pairwise disjoint collection. Clearly $G_n = \cup_{i=1}^n E_i$ is measurable. Let $G = \cup_i E_i$. Therefore

$$\begin{aligned}
\mathrm{P}^*(B) &= \mathrm{P}^*(B \cap G_n) + \mathrm{P}^*(B \cap G_n^c) \\
&\geq \mathrm{P}^*(B \cap G_n) + \mathrm{P}^*(B \cap G^c) \\
&= \mathrm{P}^*(B \cap G_n \cap E_n) + \mathrm{P}^*(B \cap G_n \cap E_n^c) + \mathrm{P}^*(B \cap G^c) \\
&= \mathrm{P}^*(B \cap E_n) + \mathrm{P}^*(B \cap G_{n-1}) + \mathrm{P}^*(B \cap G^c) \\
&= \sum_{i=1}^n \mathrm{P}^*(B \cap E_i) + \mathrm{P}^*(B \cap G^c).
\end{aligned}$$

Taking limits $\mathrm{P}^*(B) \geq \sum_{i=1}^\infty \mathrm{P}^*(B \cap E_i) + \mathrm{P}^*(B \cap G^c) \geq \mathrm{P}^*(B \cap G) + \mathrm{P}^*(B \cap G^c)$, where the last inequality follows from sub-additvity. Hence $G = \cup_i E_i$ is measurable or $\mathcal{E}$ is an σ-algebra.

To show that $\mathrm{P}^*(\cdot)$ is a countably additive probability measure on $\mathcal{E}$, note the following. Let E_i be a pairwise disjoint collection and let $G = \cup_i E_i$. Then

$$\begin{aligned}\mathrm{P}^*(E_1 \cup E_2) &= \mathrm{P}^*((E_1 \cup E_2) \cap E_1) + \mathrm{P}^*((E_1 \cup E_2) \cap E_1^c)\\ &= \mathrm{P}^*(E_1) + \mathrm{P}^*(E_2).\end{aligned}$$

Hence $\mathrm{P}^*(\cdot)$ is a finitely additive probability measure on $\mathcal{E}$. Countable additivity is a simple consequence of subadditivity and the following:

$$\mathrm{P}^*(G) \geq \mathrm{P}^*(G_n) = \sum_{i=1}^{n} \mathrm{P}^*(E_i).$$

Hence $\mathrm{P}^*(G) = \sum_{i=1}^{\infty} \mathrm{P}^*(E_i)$.

To show that $\mathcal{A} \subseteq \mathcal{E}$, for any B take a "good" cover of B, and show that this induces a cover on $B \cap A$ and $B \cap A^c$. Hence show that

$$\mathrm{P}^*(B) \geq \mathrm{P}^*(B \cap A) + \mathrm{P}^*(B \cap A^c).$$

Thus $\mathcal{A} \subseteq \mathcal{E}$ and since $\mathcal{E}$ is a sigma-algebra, we also have $\sigma(\mathcal{A}) \subseteq \mathcal{E}$.

3. To show uniqueness, show the following: Let $\mathcal{M} = \{B : p_1(B) = p_2(B)\}$ be the collection of all sets in which the two countably additive probability extensions agree. Then it is clear that $\mathcal{M}$ is a monotone class. The proof then follows Theorem 1.1.1. $\square$

Thus Caratheordory's extension theorem reduces the burden of construction of measures on σ-algebras to those on algebras.

1.1.4.1. *Constructing countably additive probability measures on algebras.* We will now see that there is a canonical way of constructing countably additive probability measure on $\mathcal{F}_B$, the borel σ-algebra on real line.

Consider the following collection of intervals: $\mathcal{I} = \{I_{a,b} : -\infty \leq a < b \leq \infty\}$, where $I_{a,b} = (a, b]$ when $b < \infty$ and $I_{a,\infty} = (a, \infty)$.

Exercise 1.1.7. Show that the class, $\mathcal{A}_B$, of finite disjoint union of members of $\mathcal{I}$ is an algebra.

Assume we are given a function $F(x)$ which is nondecreasing, right-continuous, and satisfies

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1.$$

Then we can define a finitely additve P by first defining $P(I_{a,b}) = F(b) - F(a)$ for intervals, and then extending it to $\mathcal{A}_B$ by defining it as sum for disjoint unions from $\mathcal{I}$.

We will now prove Lebesgue's theorem which shows when one can extend the finitely additive P to a countably additive P.

Theorem 1.1.4. *(Lebesgue)*
Let P be a finitely additive probability measure on $\mathcal{A}_B$. P is countably additive on $\mathcal{A}_B$ if and only if $F(x) = P((-\infty, x])$ is right continuous function of x.

Remark 1.1.6. Essentially this means (using Caratheodory's theorem) that for every right continuous function non-decreasing $F(x)$ that satisfies $F(-\infty) = 0$ and $F(\infty) = 1$, there is a unique countably additive probability measure on $\mathcal{F}_B$; and conversely every countably additive probability measure on $\mathcal{F}_B$ induces a right continuous function.

PROOF. Suppose P is countably additive on $\mathcal{A}$. Then for any x and $\{\epsilon_n\} \downarrow 0$, the collection of intervals $J_n = (x, x + \epsilon_n]$ decreases to $\emptyset$. This means that $F(x + \epsilon_n) - F(x) \downarrow 0$ and this suffices (why? hint: $F(x)$ is non-decreasing).

The tricky part is clearly the reverse direction, i.e. getting countable additivity from right continuous and non-decreasing $F(x)$ with $F(-\infty) = 0$ and $F(\infty) = 1$. Suppose $A_j \in \mathcal{A}, A_j \downarrow \emptyset$. Assume that $\mathrm{P}(A_j) \geq \delta > 0$. (We wish to show a contradiction). Now pick l large enough that $1 - F(l) + F(-l) < \frac{\delta}{2}$, and define $B_j = A_j \cap (-l, l]$. Clearly $\mathrm{P}(B_j) \geq \frac{\delta}{2}, \forall j$.

Since B_j is a finite disjoint union of left open right closed intervals, create $C_j \subset B_j$ by moving the left (open) end point of the intervals to the right (i.e. shortening each interval). Clearly this can be done so as to guarantee that

$$\mathrm{P}(B_j \setminus C_j) \leq \frac{\delta}{3 \cdot 2^j}, \ \forall j.$$

Define $D_j = closure(C_j)$. Clearly $D_j \subset B_j \subset A_j$.

We know that B_j's are decreasing but C_j's may not be. Therefore define $E_j = \cap_{i=1}^{j} C_i$, and $F_j = \cap_{i=1}^{j} D_i$. Clearly $F_j \supset E_j$ (by construction), and observe that

$$\mathrm{P}(E_j) \geq \mathrm{P}(B_j) - \mathrm{P}(B_j \cap E_j^c) \geq \mathrm{P}(B_j) - \sum_{i=1}^{j} \mathrm{P}(B_j \cap C_i^c)$$

$$\geq \mathrm{P}(B_j) - \sum_{i=1}^{j} \mathrm{P}(B_i \cap C_i^c) \geq \frac{\delta}{2} - \frac{\delta}{3} > 0.$$

Therefore F_j is non-empty. Thus F_j is non-empty, closed, bounded and decreasing. Thus $\cap_j F_j$ cannot be the $\emptyset$ (finite intersection property). However $F_j \subset A_j$ and $A_j \downarrow \emptyset$, leading to a contradiction.(!) □

$F(\cdot)$ is $t$he distribution function corresponding to the probability measure P.

1.2. Random variables and integration

DEFINITION 1.2.1. A *random variable* or a *measurable function* is a map $f : (\Omega, \Sigma) \mapsto (\mathrm{R}, \mathcal{F}_B)$ (really a mapping $\Omega \to \mathrm{R}$ but with the sigma-algebras specified) such that $\forall B \in \mathcal{F}_B,\ f^{-1}(B) = \{w : f(w) \in B\} \in \Sigma$.

In the above definition $\mathcal{F}_B$ denotes the Borel σ-algebra generated by the open intervals on the real line.

EXERCISE 1.2.1. For a class of sets $\mathcal{A} \subset \mathcal{F}_B$, and a mapping $f : (\Omega, \Sigma) \mapsto (\mathrm{R}, \mathcal{F}_B)$, suppose it holds that $f^{-1}(A) \in \Sigma$ for all $A \in \mathcal{A}$, then show that $f^{-1}(C) \in \Sigma$ for all $C \in \sigma(\mathcal{A})$, where $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$.

REMARK 1.2.1. The above exercise show that to verify that a function is measurable, it suffices to consider the inverse images of any collection of sets that generate $\mathcal{F}_B$.

Some facts about random variables:

(1) If $A \in \Sigma$ then
$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$
is bounded and measurable.

(2) Sums, products, limits, etc of measurable functions are measurable.

EXERCISE 1.2.2. Show that if f_1, f_2 are measurable, then $f_1 f_2$ (their pointwise product) is measurable.

(3) if $\{A_j : 1 \le j \le n\}$ is a finite disjoint partition of Ω into measurable sets, then the function (called a *simple function*)
$$f(w) = \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}(\omega)$$
is bounded and measurable.

Lemma 1.2.1. *Any bounded measurable function is the uniform limit of simple functions.*

PROOF. Suppose $|f(\omega)| < M$, then divide the interval $(-M, M]$ into n disjoint, $\{I_i\}$, intervals of length $\frac{2M}{n}$. Let c_i denote the midpoint of the intervals. Then define $A_i = w : f(w) \in I_i$. Clealry $\{A_i\}$'s are measurable and disjoint. Consider the simple function
$$f_n(\omega) = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}(\omega).$$
Clearly $|f(\omega) - f_n(\omega)| \le \frac{M}{n}, \forall \omega$. Hence the convergence is uniform. □

1.2.1. Definition of integrals.

(1) For a simple function f defined on the probability space $(\Omega, \Sigma, \mathrm{P})$ we define the *integral* with respect to the probability measure as
$$\int f dP = \int \left(\sum_{i=1}^{n} c_i \mathbf{1}_{A_i} \right) dP = \sum_{i=1}^{n} c_i P(A_i).$$

(2) If f is a bounded, measurable function and f_n be any sequence of simple functions that converge to f uniformly, then we define

$$\int f dP = \lim_n \int f_n dP.$$

(Why does this limit exist, and why is it independent of the particular sequence f_n?)

The limit exists because the sequence $\int f_n dP$ is Cauchy. It is also independent of the particular sequence f_n because the difference of two such sequences $f_n - g_n$ is bounded and decreases to 0 pointwise (uniformly).

EXERCISE 1.2.3. Complete the details of the argument and show that $\int f dP$ is well defined, when f is bounded measurable function.

DEFINITION 1.2.2. A sequence of functions f_n is said to converge to f pointwise (or everywhere) if

$$\lim_{n\to\infty} f_n(\omega) = f(\omega),\ \forall \omega \in \Omega.$$

DEFINITION 1.2.3. A sequence of measurable functions f_n is said to converge to a measurable function f *almost everywhere* (or *almost surely*) if $\exists\ N \subset \Omega, \mathrm{P}(N) = 1$ such that

$$\lim_{n\to\infty} f_n(\omega) = f(\omega),\ \forall \omega \in N.$$

DEFINITION 1.2.4. A sequence of measurable functions f_n is said to converge to a measurable function f *in measure* (or *in probability*) if $\forall \epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}(\{w : |f_n(\omega) - f(\omega)| > \epsilon\}) = 0.$$

Convergence in measure is a weaker notion than almost sure convergence.

Lemma 1.2.2. *If a sequence of measurable functions f_n converge to f almost everywhere then the sequence of measurable functions also converge to f in measure.*

PROOF. For any $\epsilon > 0$, define the sets

$$A_n^\epsilon = \{\omega : |f_n(\omega) - f(\omega)| > \epsilon\}, \quad B_n^\epsilon = \cup_{m=n}^\infty A_n^\epsilon.$$

Let $B_n^\epsilon \downarrow B^\epsilon$. Since $f_n(\omega) \to f(\omega)$ for $\omega \in N$, then $B^\epsilon \subseteq N^c$, and hence $\mathrm{P}(B^\epsilon) = 0$ implying that $\mathrm{P}(B_n^\epsilon) \downarrow 0$; and since $\mathrm{P}(A_n^\epsilon) \leq \mathrm{P}(B_n^\epsilon)$ we have $\mathrm{P}(A_n^\epsilon) \to 0$ as desired, establishing convergence in measure. □

Let $\Omega = [0, 1]$ and the probability measure induced by the Lebesgue measure on this set. Consider the following sequence of real valued functions from $\mathbb{R}_+ \to [0, 1]$ define by

$$r_n(x) = \begin{cases} 1 & x \in (H_n, H_{n+1}] \\ 0 & \text{otherwise} \end{cases}.$$

Here $H_n = \sum_{i=1}^n \frac{1}{i}$ is the harmonic sum. Use the above sequence of functions to define measurable functions $f_n(\omega)$ according to

$$f_n(\omega) = \sum_{i=1}^\infty r_n(\omega + i).$$

EXERCISE 1.2.4. Show that $f_n(\omega) \to 0$ in measure, while $\lim_n f_n(\omega)$ does not exist almost surely, thus there is no convergence almost surely.

On the other hand, convergence in measure does imply almost sure convergence on a sub-sequence as demonstrated by the following lemma.

Lemma 1.2.3. *If a sequence of measurable functions f_n converge to f in measure, then there is a subsequence, n_i such that, the sequence of measurable functions f_{n_i} converge to f almost everywhere.*

PROOF. For any $k \in \mathbb{N}$, define the set

$$A_n^k = \{\omega : |f_n(\omega) - f(\omega)| > \frac{1}{k}\}.$$

From convergence in measure, we know that $\mathrm{P}(A_n^k) \to 0$ as $n \to \infty$. Let n_k be such that $\mathrm{P}(A_{n_k}^k) \leq \frac{1}{2^k}$. Define

$$B_k = \cup_{m=k}^{\infty} A_{n_m}^m.$$

Thus $\mathrm{P}(B_k) \leq \frac{1}{2^{k-1}}$ and $B_k \downarrow B$ with $\mathrm{P}(B) = 0$. Note that

$$B = \{\omega : \limsup_k |f_{n_k}(\omega) - f(\omega)| > 0\}$$

and this completes the argument. □

REMARK 1.2.2. We will only be dealing with measurable functions; so unless explicitly stated please assume that all functions are measurable.

THEOREM 1.2.4. *(Bounded Convergence Theorem) If a sequence $\{f_n(\omega)\}$ of uniformly bounded functions converge to a bounded function $f(\omega)$ in measure then*

$$\int f_n dP \to \int f dP.$$

PROOF. First note that (argue why using definition)

$$\int f_n dP - \int f dP = \int (f_n - f) dP.$$

Again argue that

$$\left| \int f_n dP - \int f dP \right| = \left| \int (f_n - f) dP \right| \leq \int |f_n - f| dP.$$

As before, define

$$A_n^{\epsilon} = \{\omega : |f_n(\omega) - f(\omega)| > \epsilon\}.$$

Using this we have

$$\left| \int f_n dP - \int f dP \right| \leq \int |f_n - f| dP \leq \epsilon(1 - \mathrm{P}(A_n^{\epsilon})) + 2M\mathrm{P}(A_n^{\epsilon}).$$

Since $\mathrm{P}(A_n^{\epsilon}) \to 0$ as $n \to \infty$ we get

$$\limsup \left| \int f_n dP - \int f dP \right| \leq \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we are done. □

DEFINITION 1.2.5. For a non-negative measurable function $f(\omega)$ we define

$$\int f(\omega) dP = \{\sup \int g dP : g \text{ is bounded}, 0 \leq g \leq f\}.$$

Lemma 1.2.5 (Fatou)**.** *Let $f_n \geq 0$ converge in measure to f (also assumed non-negative) as $n \to \infty$ then*

$$\int f dP \leq \liminf_{n\to\infty} \int f_n dP.$$

PROOF. Consider any $0 \leq g \leq f$ such that g is bounded. Define $h_n = \min\{f_n, g\}$. The observe that $\{h_n\}$ is uniformly bounded and $h_n \to g$ in measure. Thus from bounded convergence theorem, we have

$$\int g dP = \lim_{n\to\infty} \int h_n dP \leq \liminf_{n\to\infty} \int f_n dP.$$

Since g is an arbitrary bounded function such that $0 \leq g \leq f$, taking *sup* over the class of such g yields the desired result. □

An alternate version of Fatou's lemma that is often used is the following:

Lemma 1.2.6 (Fatou (alternate))**.** *Let $f_n \geq 0$ then*

$$\int \liminf_{n\to\infty} f_n dP \leq \liminf_{n\to\infty} \int f_n dP.$$

PROOF. First assume that $g = \liminf_{n\to\infty} f_n$ is finite *almost everywhere*. Define $g_n = \inf_{m\geq n} f_m$ and observe that $g_n \uparrow g = \liminf_{n\to\infty} f_n$ pointwise (and in measure (why?)). Thus from the former version we have that

$$\int \liminf_{n\to\infty} f_n dP = \int g dP \leq \liminf_{n\to\infty} \int g_n dP \leq \liminf_{n\to\infty} \int f_n dP,$$

where the last inequality is a consequence of $0 \leq g_n \leq f_n$.

If $g = +\infty$ on A with $P(A) > 0$, then for any $M > 0$ observe that the earlier part yields

$$MP(A) \leq \int \liminf_{n\to\infty} \{f_n \wedge M\} dP \leq \liminf_{n\to\infty} \int \{f_n \wedge M\} dP \leq \liminf_{n\to\infty} \int f_n dP.$$

Taking $M \to \infty$ implies that both integrals of interest tend to infinity. □

Corollary 1.2.7 (Monotone Convergence Theorem)**.** *If a sequence of non-negative functions $f_n \uparrow f$ then*

$$\lim_{n\to\infty} \int f_n dP \to \int f dP.$$

PROOF. $0 \leq f_n \leq f$ implies that (why?)

$$\int f_n dP \leq \int f dP$$

and taking lim sup yields

$$\limsup_{n\to\infty} \int f_n dP \leq \int f dP.$$

The other half follows from Fatou's lemma. □

DEFINITION 1.2.6. A non-negative measurable function $f(\omega)$ is said to be *integrable* if

$$\int f dP < \infty.$$

DEFINITION 1.2.7. A measurable function $f(\omega)$ is said to be *integrable* if

$$\int |f| dP < \infty.$$

For integrable functions f we define $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$. Thus $f = f_+ - f_-$, where f_+ and f_- are non-negative measurable functions. We now define

$$\int f dP = \int f_+ dP - \int f_- dP.$$

EXERCISE 1.2.5. Show that the integral satisfies the following properties.

a) If f, g are integrable, then for any $a, b \in \mathbb{R}$ the function $af + bg$ is also integrable.
b) If $f = 0$ almost everywhere, then f is integrable and $\int f dP = 0$. As a consequence, two integrable functions that agree almost everywhere has the same integral.

Lemma 1.2.8 (Jensen's inequality)**.** *If $\Phi(x)$ is a convex function, and $f(\omega)$ and $\Phi(f(\omega))$ are integrable, then*

$$\int \Phi(f(\omega)) dP \geq \Phi\left(\int f(\omega) dP\right).$$

PROOF. The proof of Jensen's inequality in this case, as well as in some other cases, will use the fact that any convex function can be written as the pointwise supremum of supporting hyperplanes, i.e.

$$\Phi(x) = \sup_{(a,b)\in\mathcal{E}} ax + b.$$

Hence for any $(a, b) \in \mathcal{E}$

$$\Phi(f(\omega)) \geq af(\omega) + b,$$

yielding

$$\int \Phi(f(\omega)) dP \geq a \int f(\omega) dP + b.$$

Taking supremum of $(a, b) \in \mathcal{E}$ yields the result. □

THEOREM 1.2.9 (Dominated Convergence Theorem). *If a sequence $\{f_n\}$ converge to f in measure and $|f_n| \leq g$, where g is an integrable function, then*

$$\lim_{n\to\infty} \int f_n dP \to \int f dP.$$

PROOF. A simple application of Fatou's lemma yields that f is integrable. The proof follows using further applications of Fatou's lemma. Observe that the two non-negative sequence of functions $g - f_n$ and $g + f_n$ converge in measure to $g - f$ and $g + f$ respectively (why?). Now argue that

$$\liminf_{n\to\infty} \int (g - f_n) dP = \int g dP - \limsup_{n\to\infty} \int f_n dP,$$

$$\liminf_{n\to\infty} \int (g + f_n) dP = \int g dP + \liminf_{n\to\infty} \int f_n dP.$$

Applying Fatou's Lemma yields (justify the second relations)

$$\int (g - f) dP \leq \liminf_{n\to\infty} \int (g - f_n) dP = \int g dP - \limsup_{n\to\infty} \int f_n dP$$

$$\implies \limsup_{n\to\infty} \int f_n dP \le \int f dP,$$

and

$$\int (g+f)dP \le \liminf_{n\to\infty} \int (g+f_n)dP = \int g dP + \liminf_{n\to\infty} \int f_n dP$$
$$\implies \liminf_{n\to\infty} \int f_n dP \ge \int f dP.$$

□

EXERCISE 1.2.6. Suppose an integrable f satisfies that

$$\int f(\omega)\mathbf{1}_A(\omega)dP = 0, \ \forall A \in \mathcal{F}.$$

Show that $f = 0$ *almost everywhere*.

1.2.2. Transformations.

DEFINITION 1.2.8. A *measurable transformation* $T : (\Omega_1, \mathcal{F}_1) \mapsto (\Omega_2, \mathcal{F}_2)$ is a mapping that satisfies

$$T^{-1}(B) = \{\omega_1 : T(\omega_1) \in B\} \in \mathcal{F}_1$$

for all $B \in \mathcal{F}_2$.

If the space $(\Omega_1, \mathcal{F}_1)$ was endowed with a probability measure P then a measurable mapping T induces a probability measure on $(\Omega_2, \mathcal{F}_2)$ according to

$$\mathrm{Q}(B) = \mathrm{P}(T^{-1}(B)), \forall B \in \mathcal{F}_2.$$

EXERCISE 1.2.7. Verify that the measure Q defined above is indeed a countably additive probability measure (assuming that P is a countable additive probability measure).

THEOREM 1.2.10. *Let $f(\omega_2)$ be a measurable mapping (random variable) on $(\Omega_2, \mathcal{F}_2)$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and T be a measurable transformation from $(\Omega_1, \mathcal{F}_1) \mapsto (\Omega_2, \mathcal{F}_2)$, then the mapping $g(\omega_1) := f(T(\omega_1))$ is measurable. Further $g(\omega_1)$ is integrable with respect to* P *if and only if $f(\omega_2)$ is integrable with respect to* $\mathrm{Q} = \mathrm{P}T^{-1}$ *and*

$$\int_{\Omega_2} f(\omega_2)dQ = \int_{\Omega_1} g(\omega_1)dP.$$

PROOF. For any $B \in \mathcal{B}_{\mathbb{R}}$ observe that

$$\{\omega_1 : g(\omega_1) \in B\} = T^{-1}(f^{-1}(B)).$$

Since $f^{-1}(B) \in \mathcal{F}_2$ (measurability of $f(\omega_2)$); by measurability of T we have $T^{-1}(f^{-1}(B)) \in \mathcal{F}_1$, establishing the measurability of $g(\omega_1)$. The second part follows by the *standard-machine argument*, i.e. verify it (using previous parts) when

1) $f(\omega_2)$ is indicator function (use definition of P and Q).
2) $f(\omega_2)$ is a simple function (use linearity)
3) $f(\omega_2)$ is a bounded non-negative function (use bounded convergence theorem)
4) $f(\omega_2)$ is a non-negative function (use monotone convergence theorem by considering $f_n = \min\{f, n\}$)
5) Finally, $f(\omega_2)$ is an integrable function (use positive and negative parts).

□

REMARK 1.2.3. One of the simplest applications of this result is to take $(\Omega_2, \mathcal{F}_2) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and the identity mapping $f(\omega_2) = \omega_2$. Thus T is a random variable (as $\Omega_2 = \mathbb{R}$) and let us denote $T(\omega_1) = X(\omega_1)$. Observe that

$$Q((-\infty, x]) = P(\{\omega_1 : X(\omega_1) \leq x\}) = F_X(x),$$

the distribution function of X. Further, let us $\omega_2 = x$, where $x \in \mathbb{R}$ and $dQ = dF_X$, (the last one being a notational convenience). Therefore, the statement of the theorem yields under this situation

$$\int_{\Omega_2=\mathbb{R}} f(x)dF = \int_{\Omega_1} g(\omega_1)dP.$$

Noting that $f(x) = x$ and $g(\omega_1) = f(X(\omega_1)) = X(\omega_1)$, we obtain

$$\int_{\mathbb{R}} xdF = \int_{\Omega_1} XdP.$$

1.2.3. Product spaces. Consider two probability spaces $(\Omega_1, \mathcal{F}_1, \mathrm{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathrm{P}_2)$. The goal of this section is to work with the product space $\Omega_1 \times \Omega_2$. A *natural* σ-field that one can define on the product space is the σ-field, $\mathcal{F}$, generated by the *measurable rectangles*, i.e. sets of the form $A_1 \times A_2$ with $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$. Our next goal is to define a countably additive probability measure on $(\Omega_1 \times \Omega_2, \mathcal{F})$ that naturally extends P_1 and P_2.

For a measurable rectangle, a natural candidate is

$$\mathrm{P}(A_1 \times A_2) = \mathrm{P}_1(A_1)\mathrm{P}_2(A_2).$$

For finite disjoint union of measurable rectangles one can define

$$\mathrm{P}(\sqcup_{i=1}^n A_{1i} \times A_{2i}) = \sum_{i=1}^n \mathrm{P}_1(A_{1i})\mathrm{P}_2(A_{2i}).$$

EXERCISE 1.2.8. Show that

i) Finite disjoint union of measurable rectangles is an algebra $\mathcal{A}$.
ii) P is well defined, i.e. if $\sqcup_{i=1}^n A_{1i} \times A_{2i} = \sqcup_{j=1}^m B_{1j} \times B_{2j}$ then

$$\sum_{i=1}^n \mathrm{P}_1(A_{1i})\mathrm{P}_2(A_{2i}) = \sum_{j=1}^m \mathrm{P}_1(B_{1j})\mathrm{P}_2(B_{2j}).$$

Thus P is a finitely additive probability measure on $\mathcal{A}$.

Lemma 1.2.11. P *is a countably additive probability measure on* $\mathcal{A}$.

PROOF. Let $E_n \downarrow \emptyset$, $E_n \in \mathcal{A}$. Define the set

$$E_{n,\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in E_n\}.$$

Now define $f_n(\omega_2) = \mathrm{P}_1(E_{n,\omega_2})$ (See (show) that $f_n(\omega_2)$ is a simple function, i.e. it takes only finitely many distinct values.) Note that

$$\mathrm{P}(E_n) = \int f_n(\omega_2)d\mathrm{P}_2.$$

Now $0 \leq f_n(\omega_2) \leq 1$ and since $E_{n,\omega_2} \downarrow \emptyset$ implies that $\mathrm{P}_1(E_{n,\omega_2}) \downarrow 0$ (by countable additivity of P_1). Hence $f_n(\omega_2) \downarrow 0$ and by using bounded convergence theorem we get that

$$\int f_n(\omega_2)d\mathrm{P}_2 \downarrow 0 \implies \mathrm{P}(E_n) \downarrow 0.$$

□

By Caratheodory's extension theorem, we can extend P to a countably additive probability measure on $\mathcal{F} = \sigma(\mathcal{A})$ on $\Omega_1 \times \Omega_2$ and this measure is called the *product measure.*

1.2.3.1. *Iterated integrals and Fubini's theorem.* In this section we establish an oft-invoked theorem for justifying exchange of integrals. The proofs are basically a consequence of the *standard-machine* argument. As in the earlier section we consider two probability spaces $(\Omega_1, \mathcal{F}_1, \mathrm{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathrm{P}_2)$ and its natural extension (as defined in the previous section) of $(\Omega_1 \times \Omega_2, \mathcal{F}, \mathrm{P})$. We begin by establishing the following lemma.

Lemma 1.2.12. *For any $A \in \mathcal{F}$ denote by A_{ω_1} and A_{ω_2} the sets (sections)*

$$A_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in A\}, \quad \text{and} \;\; A_{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in A\}.$$

Then

a) For each ω_1, $A_{\omega_1} \in \mathcal{F}_2$ and for each ω_2, $A_{\omega_2} \in \mathcal{F}_1$.
b) the functions $\mathrm{P}_2(A_{\omega_1})$ and $\mathrm{P}_1(A_{\omega_2})$ are measurable and

$$\mathrm{P}(A) = \int_{\Omega_1} \mathrm{P}_2(A_{\omega_1}) d\mathrm{P}_1 = \int_{\Omega_2} \mathrm{P}_1(A_{\omega_2}) d\mathrm{P}_2.$$

Outline of Proof. The first part claims that sections of sets in product sigma-algebra belong to the individual sigma-algebras. Observe that this part is immediate if the original set is a rectangle. Then observe that the collection of all sets for which the part holds is a sigma-algebra, thus contains the sigma-algebra generated by the rectangles.

The second assertion is again immediate if $A = A_1 \times A_2$, A is a measurable rectangle. From linearity, the assertion follows for finite disjoint union of rectangles (why?). Now consider the class, $\mathcal{C}$, of all sets for which the assertion is valid. Show that $\mathcal{C}$ is a monotone class and hence contains $\mathcal{F}$. □

Theorem 1.2.13 (Fubini). *Let $f(\omega) = f(\omega_1, \omega_2)$ be a measurable function on $(\Omega, \mathcal{B})$. For each fixed ω_1 consider $g_{\omega_1}(\omega_2) := f(\omega_1, \omega_2)$ as a mapping from $\Omega_2 \to \mathrm{R}$ and for each fixed ω_2 consider $h_{\omega_2}(\omega_1) := f(\omega_1, \omega_2)$ as a mapping from $\Omega_1 \to \mathrm{R}$. Then*

a) For each ω_1 the function $g_{\omega_1}(\omega_2)$ is measurable (similarly for each fixed ω_2, the function $h_{\omega_2}(\omega_1)$ is measurable).
b) If f is integrable, then for almost all ω_1, the function $g_{\omega_1}(\omega_2)$ is integrable. (Similarly for almost all ω_2, the function $h_{\omega_2}(\omega_1)$ is integrable) and further the functions

$$G(\omega_1) := \int_{\Omega_2} g_{\omega_1}(\omega_2) dP_2, \quad \text{and} \quad H(\omega_2) := \int_{\Omega_1} h_{\omega_2}(\omega_1) dP_2$$

are integrable. Finally, we also have

$$\int_{\Omega} f(\omega_1, \omega_2) = \int_{\Omega_1} G(\omega_1) dP_1 = \int_{\Omega_2} H(\omega_1) dP_2.$$

Proof. The first part is an immediate consequence of the first part in Lemma 1.2.12 that sections of sets in product sigma-algebra belong to the individual sigma-algebras.

The proof of the second part uses the *standard machine* approach. For indicator functions, the theorem reduces to second part of Lemma 1.2.12. Simple functions follows by linearity and uniform limits imply the result for bounded measurable functions. Now monotone convergence theorem implies the result for non-negative functions and by taking f_+ and f_- the result follows for integrable functions. $\square$

EXERCISE 1.2.9. Consider $\Omega = \mathbb{N}\times\mathbb{N}$ and $\mathrm{P}(i,j) = \frac{1}{2^{i+j}},\ i,j\geq 1$. For $i,j\geq 1$

$$f(i,j) = \begin{cases} 2^{i+j} & j = i+1 \\ -2^{i+j} & j = i-1, i\geq 2 \\ 0 & o.w. \end{cases}$$

Here $\mathrm{P}_1(i) = \mathrm{P}_2(i) = \frac{1}{2^i}, i\geq 1$. Compute the functions

$$G(i) := \sum_j f(i,j)\mathrm{P}_2(j) \quad \text{and} \quad H(j) := \sum_i f(i,j)\mathrm{P}_1(i).$$

What are the sums $\sum_{i\geq 1} G(i)\mathrm{P}_1(i)$ and $\sum_{j\geq 1} H(j)p_2(j)$. (Note that f is not integrable).

1.2.4. Borel-Cantelli Lemma 1.

Lemma 1.2.14. *Let $(\Omega,\mathcal{F},\mathrm{P})$ be a probability space. Consider a collection of sets $\{A_n\}, A_n\in\mathcal{F}$ such that $\sum_{n=1}^{\infty}\mathrm{P}(A_n)<\infty$. Define the set $\bar{A} := \cap_n\cup_{m\geq n} A_m$. Then $\mathrm{P}(\bar{A}) = 0$.*

PROOF. Define $B_n = \cup_{m\geq n}A_m$. Then $\mathrm{P}(B_n)\leq\sum_{m\geq n}\mathrm{P}(A_m)$; hence $\mathrm{P}(B_n)\to 0$ as $n\to\infty$, since $\sum_{n=1}^{\infty}\mathrm{P}(A_n)<\infty$. Now $B_n\downarrow\bar{A}$ and the result is immediate. $\square$

REMARK 1.2.4. This is an often used lemma to deduce sub-sequential almost-sure convergence. Let us use this to show that convergence in measure implies that there is a sub-sequential almost-sure convergence.

Lemma 1.2.15 (subsequence convergence)**.** *Let f_n be a sequence of measurable functions that converge in measure to a measurable function f. Then, there is a subsequence $\{n_k\}$ such that $f_{n_k}\to f$ almost surely.*

PROOF. Given $k\in\mathbb{N}$, define

$$A_n := \{w : |f_n(\omega) - f(\omega)| > \frac{1}{k}\}.$$

Since $P(A_n)\to 0$, define n_k to be the smallest n such that $P(A_{n_k})\leq\frac{1}{2^k}$. Clearly $\sum_{k=1}^{\infty}P(A_{n_k})\leq 1<\infty$, hence (by the Borel-Cantelli lemma) $P(\bar{A}) = 0$ where $\bar{A} = \cap_k\cup_{m\geq k}A_{n_m}$. Note that $\bar{A}$ coincides with the set $\{\omega : \limsup_k|f_{n_k}(\omega)-f(\omega)|> 0\}$. $\square$

CHAPTER 2

Characteristic Functions

2.1. Preliminaries

Let $f : \Omega \to \mathbb{C}$ be a complex valued function. Let f_r, f_i denote its real and imaginary components. We say that f is measurable if f_r and f_i are measurable. Further if f_r and f_i are integrable, we define

$$\int f d\mathrm{P} = \int f_r d\mathrm{P} + i \int f_i d\mathrm{P}.$$

EXERCISE 2.1.1. Let $f : \Omega \to \mathbb{C}$ be a complex valued measurable function. Then show that

$$\left| \int f d\mathrm{P} \right| \leq \int |f| d\mathrm{P}.$$

For any random variable X we define the characteristic function according (see Remark 1.2.3) to

$$\phi(t) = E(\exp[itX]) = \int_{\Omega} e^{itX} dP - \int_{\mathbb{R}} e^{itx} dF.$$

Using above exercise, note that $|\phi(t)| \leq 1$.

THEOREM 2.1.1. *The characteristic function of any probability distribution is a uniformly continuous function of* t *that is positive definite, i.e. for any real numbers* $t_1, .., t_n$ *the matrix* $M \equiv [\phi(t_k - t_l)]$ *is non-negative semidefinite.*

PROOF.

$$\begin{aligned}
\vec{\xi} M \vec{\xi}^* &= \sum_{k,l} \xi_k \phi(t_k - t_l) \xi_l^* \\
&= \sum_{k,l} \xi_k E\left(\exp[i(t_k - t_l)X]\right) \xi_k^* \\
&= E\left(\sum_{k,l} \xi_k \exp[i(t_k - t_l)X] \xi_k^* \right) \\
&= E\left(\sum_k \xi_k \exp[it_k X] \sum_l \exp[-it_l X] \xi_l^* \right) \\
&= E\left(|\sum_k \xi_k \exp[it_k X]|^2 \right) \geq 0.
\end{aligned}$$

The equality holds if and only if $Y = \sum_i \xi_i \exp[it_i X] = 0$ almost surely (i.e. with probability 1).

To show uniform continuity observe that

$$\begin{aligned}|\phi(t)-\phi(s)| &= |E(e^{itX}-e^{isX})| \\ &\le E(|e^{itX}-e^{isX}|) \\ &= E(|e^{i(t-s)X}-1|).\end{aligned}$$

Thus it suffices to show that for every $\epsilon > 0$ we can pick a $\delta > 0$ such that whenever $|t| < \delta$ we have $E(|e^{itX}-1|) < \epsilon$. Assume otherwise, i.e. for a sequence $\delta_n \downarrow 0$ there exists points $t_n, |t_n| \le \delta_n$ such that $E(|e^{it_nX}-1|) \ge \epsilon$. Now $t_n \to 0$ and hence $Y_n = |e^{it_nX}-1| \to 0$ pointwise. Since Y_n is bounded we have from bounded convergence theorem that $E(Y_n) \to 0$, and this yields a contradiction. □

Lemma 2.1.2. *If $\int |X|dP < \infty$ then $\phi(t)$ is continuously differentiable and $\phi'(0) = i\int XdP$.*

PROOF.

$$\frac{1}{\delta}E(e^{itX}-e^{i(t-\delta)X}) = E(e^{itX}\frac{1-e^{-i\delta X}}{i\delta X}iX)$$

Set $Y_\delta = e^{itX}\frac{1-e^{-i\delta X}}{i\delta X}iX$ and $Y = iXe^{itX}$. Clearly $Y_\delta \to Y$ pointwise.

Further $|Y_\delta| \le c|X|$ when $c = \sup_x |\frac{1-e^{ix}}{x}| < \infty$ (see lemma below). Thus from dominated convergence theorem (since $E(|X|) < \infty$) we have that $E(Y_\delta) \to E(Y)$. Therefore $\phi'(t) = E(iXe^{itX})$ exists. The continuity of $\phi'(t)$ is left as an exercise. □

Lemma 2.1.3. $|\frac{1-e^{ix}}{x}| \le 1, \ \forall x \in \mathrm{R}$.

PROOF.

$$\left|\frac{1-e^{ix}}{x}\right| = \left|2\frac{\sin\left(\frac{x}{2}\right)}{x}\right| \le 1.$$

□

EXERCISE 2.1.2. Show that if $E(|X|^r) < \infty$ then $\phi(t)$ is r times continuously differentiable.
Extra credit: If r is even, show that the converse holds.

How do we get back the distribution function from the characteristic function?

THEOREM 2.1.4. *When a, b are continuity points of $F(x) := \mathrm{P}(X \le x)$, then*

$$F(b)-F(a) = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\phi(t)\frac{\exp[-itb]-\exp[-ita]}{-it}dt.$$

PROOF.

$$\begin{aligned}&\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\phi(t)\frac{\exp[-itb]-\exp[-ita]}{-it}dt \\ &= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\exp[-itb]-\exp[-ita]}{-it}\int e^{itX}dPdt \\ &\overset{Fub}{=} \lim_{T\to\infty}\frac{1}{2\pi}\int\int_{-T}^{T}\frac{\exp[it(X-b)]-\exp[it(X-a)]}{-it}dtdP \\ &= \lim_{T\to\infty}\frac{1}{2\pi}\int\int_{-T}^{T}\frac{\sin t(X-a)-\sin t(X-b)}{t}dtdP\end{aligned}$$

$$
\begin{aligned}
&= \lim_{T\to\infty} \frac{1}{\pi} \int \int_0^T \frac{\sin t(X-a) - \sin t(X-b)}{t} dt dP \\
&= \lim_{T\to\infty} \int u(T, X-a) - u(T, X-b) dP \\
&\overset{d.c}{=} \int \lim_{T\to\infty} \left(u(T, X-a) - u(T, X-b)\right) dP \\
&= \int \frac{1}{2} 1_{X>a} - \frac{1}{2} 1_{X<a} - \frac{1}{2} 1_{X>b} + \frac{1}{2} 1_{X<b} dP \\
&= \frac{1}{2} \left(P(X<b) - P(X>b) - P(X<a) + P(X>a)\right) \\
&= F(b) - F(a) - \frac{1}{2}(P(X=b) - P(X=a))
\end{aligned}
$$

Note that from (below) the definition of
Here

$$u(T,x) = \int_0^T \frac{\sin tx}{\pi t} dt = \int_0^{\frac{T}{\pi}} \frac{\sin \pi s x}{\pi s x} ds.$$

We know (see below) that $\sup_{T,x} |u(T,x)| \le C$ and

$$\lim_{T\to\infty} u(T,x) = \begin{cases} \frac{1}{2} & \text{if } x > 0 \\ -\frac{1}{2} & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Hence one can find the distribution function from the characteristic function. □

EXERCISE 2.1.3. Prove that: If two distribution functions agree on their points of continuity then they agree everywhere.
Hint: Show that the points of discontinuity are countable. Then use right continuity of the distribution functions.

Lemma 2.1.5. *Consider the Dirichlet integral defined according to*

$$u(T,x) = \int_0^T \frac{\sin tx}{\pi t} dt.$$

Then the following holds:

(*i*) $\sup_{T,x} |u(T,x)| \le C$, *($C = 2$ works).*

(*ii*)

$$\lim_{T\to\infty} u(T,x) = \begin{cases} \frac{1}{2} & \textit{if } x > 0 \\ -\frac{1}{2} & \textit{if } x < 0 \\ 0 & \textit{if } x = 0 \end{cases}.$$

PROOF. Proof of (i): Without loss of generality, let us assume $x > 0$. Further let k be such that $T \in (2k\frac{\pi}{x}, 2(k+1)\frac{\pi}{x}]$. If $k = 0$ then

$$\left| \int_0^T \frac{\sin tx}{\pi t} dt \right| \le \int_0^T \frac{x}{\pi} dt = \frac{Tx}{\pi} \le 2.$$

For $k \ge 1$, we express

$$u(T,x) = \sum_{j=0}^{k-1} \int_{2j\frac{\pi}{x}}^{(2j+1)\frac{\pi}{x}} \left(\frac{\sin tx}{\pi t} + \frac{\sin x(t+\frac{\pi}{x})}{\pi(t+\frac{\pi}{x})} \right) dt + \int_{2k\frac{\pi}{x}}^T \frac{\sin tx}{\pi t} dt$$

$$= \sum_{j=0}^{k-1} \int_{2j\frac{\pi}{x}}^{(2j+1)\frac{\pi}{x}} \frac{\sin tx}{t(tx+\pi)} dt + \int_{2k\frac{\pi}{x}}^{T} \frac{\sin tx}{\pi t} dt.$$

Thus we have (the second integral below only appears when $k \geq 2$)

$$\begin{aligned} |u(T,x)| &\leq \int_0^{\frac{\pi}{x}} \frac{\sin tx}{t(tx+\pi)} dt + \int_{2\frac{\pi}{x}}^{(2k-1)\frac{\pi}{x}} \frac{1}{t^2 x} dt + \int_{2k\frac{\pi}{x}}^{T} \frac{1}{\pi t} dt \\ &\leq \int_0^{\frac{\pi}{x}} \frac{x}{(tx+\pi)} dt + \frac{1}{2\pi} - \frac{1}{(2k-1)\pi} + \frac{1}{\pi} \ln\left(\frac{Tx}{2k\pi}\right) \\ &\leq \ln 2 + \frac{1}{2\pi} + \frac{1}{\pi} \ln\left(\frac{k+1}{k}\right) \leq \left(1 + \frac{1}{\pi}\right) \ln 2 + \frac{1}{2\pi}. \end{aligned}$$

This establishes part (i). We used $\sin(tx) \leq |tx|$ in the first inequality.

Proof of (ii): As before, w.l.o.g., let us assume $x > 0$. We first write the integral of interest as a complex line integral

$$u(T,x) = \frac{1}{2} \int_{L:(-T,0)\to(T,0)} \frac{e^{izx}}{i\pi z} dz.$$

For every $T > \epsilon > 0$, we consider almost semi-circular closed contour consisting of the following parts: a line from $(-T,0) \to (-\epsilon,0)$, a clockwise semicircle (center at origin and above the real axis) from $(-\epsilon,0) \to (\epsilon,0)$, a line from $(\epsilon,0) \to (T,0)$ and finally a counter-clockwise semi-circle from $(T,0)$ to $(-T,0)$. Since the closed contour does not have any poles in its interior, and the function $\frac{e^{izx}}{i\pi z}$ is analytic in the interior of the contour, we have

$$u(T,x) - \frac{1}{2} \int_{L:(-\epsilon,0)\to(\epsilon,0)} \frac{e^{izx}}{i\pi z} dz + \int_{\pi}^{0} \frac{1}{2\pi} e^{i\epsilon e^{i\theta} x} d\theta + \int_0^{\pi} \frac{1}{2\pi} e^{iTe^{i\theta} x} d\theta = 0.$$

We consider the three integrals separately. Note that

$$\begin{aligned} &\left| \frac{1}{2} \int_{L:(-\epsilon,0)\to(\epsilon,0)} \frac{e^{izx}}{i\pi z} dz \right| = \left| \frac{1}{2} \int_{L:(-\epsilon,0)\to(\epsilon,0)} \frac{\sin(zx)}{i\pi z} dz \right| \\ &\quad \leq \frac{1}{2} \int_{L:(-\epsilon,0)\to(\epsilon,0)} \frac{|x|}{\pi} dz = 2\epsilon \frac{|x|}{\pi}. \end{aligned}$$

Observe that

$$\int_{\pi}^{0} \frac{1}{2\pi} e^{i\epsilon e^{i\theta} x} d\theta = -\frac{1}{2} + \int_{\pi}^{0} \frac{1}{2\pi} (e^{i\epsilon e^{i\theta} x} - 1) d\theta.$$

Now

$$\begin{aligned} \left| \int_{\pi}^{0} \frac{1}{2\pi} (e^{i\epsilon e^{i\theta} x} - 1) d\theta \right| &= \left| \int_{\pi}^{0} \frac{1}{2\pi} (e^{-\epsilon \sin(\theta) x} e^{i\epsilon \cos(\theta) x} - 1) d\theta \right| \\ &= \left| \int_{\pi}^{0} \frac{1}{2\pi} (e^{-\epsilon \sin(\theta) x} (e^{i\epsilon \cos(\theta) x} - 1) + e^{-\epsilon \sin(\theta) x} - 1) d\theta \right| \\ &\leq \int_0^{\pi} \frac{1}{2\pi} e^{-\epsilon \sin(\theta) x} \left| (2 \sin(\epsilon \cos(\theta) x/2) \right| d\theta \\ &\quad + \int_0^{\pi} \frac{1}{2\pi} \left| 1 - e^{-\epsilon \sin(\theta) x} \right| d\theta \\ &\leq \frac{\epsilon x}{2\pi} \int_0^{\pi} |\cos\theta| d\theta + \frac{\epsilon x}{2\pi} \int_0^{\pi} \sin\theta d\theta = \frac{2\epsilon x}{\pi} \end{aligned}$$

The last inequality uses $|\sin(a)| \le |a|$ and $|1 - e^{-a}| \le a, a > 0$.

From these two estimates, setting $\epsilon \to 0$ we see that we have

$$u(T,x) - \frac{1}{2} + \int_0^\pi \frac{1}{2\pi} e^{iTe^{i\theta}x} d\theta = 0.$$

(Note: one can use this relation to get a better upper bound on part (i), if desired).

The remaining integral is dealt with as follows:

$$\begin{aligned}\int_0^\pi \frac{1}{2\pi} e^{iTe^{i\theta}x} d\theta &= \frac{1}{2\pi}\int_0^\pi e^{-T\sin\theta} e^{iT\cos\theta} d\theta \\ &= \frac{1}{2\pi}\int_0^{\sin^{-1}\frac{1}{\sqrt{T}}} e^{-T\sin\theta} e^{iT\cos\theta} d\theta + \frac{1}{2\pi}\int_{\sin^{-1}\frac{1}{\sqrt{T}}}^{\pi} e^{-T\sin\theta} e^{iT\cos\theta} d\theta.\end{aligned}$$

Bounding each integral separately (the first integrand by 1 and the second integrand by $e^{-\sqrt{T}}$) we obtain that

$$\left|\int_0^\pi \frac{1}{2\pi} e^{iTe^{i\theta}x} d\theta\right| \le \frac{1}{2\pi}\sin^{-1}\left(\frac{1}{\sqrt{T}}\right) + \frac{1}{2} e^{-\sqrt{T}}.$$

Thus when $x > 0$ we have

$$\lim_{T\to\infty} u(T,x) = \frac{1}{2}.$$

$\square$

2.2. Weak Convergence

Definition 2.2.1. A sequence P_n of probability distributions on $(R, \mathcal{F}_R)$ is said to converge *weakly* to a probability distribution P if

$$\lim_n P_n(I) = P(I),$$

where $I = [a,b]$ is any interval such that $P(\{a\}) = P(\{b\}) = 0$.

Exercise 2.2.1. Show that the following is an alternate definition of weak convergence: Let $F_n(x)$ be the distribution functions associated with P_n and $F(x)$ be the distribution function associated with P. Then $P_n \Rightarrow P$ if $\lim_n F_n(x) = F(x)$ at every continuity point of F.

Theorem 2.2.1. *(Levy-Cramer Continuity Theorem) The following are equivalent.*

(i) $P_n \Rightarrow P$ *or* $F_n \Rightarrow F$.

(ii) For every bounded continuous function $f(x)$ *on* R

$$\lim_n \int f(x) dF_n = \int f(x) dF.$$

(iii) Let $\phi_n(t)$ *be the characteristic function of* F_n *and* $\phi(t)$ *the characteristic function of* F. $\phi_n(t) \to \phi(t)$ *pointwise.*

Proof. We shall show the equivalence by showing that $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$.

- $(i) \Rightarrow (ii)$: Let $a < b$ be continuity points of F and $F(a) \le \epsilon, F(b) \ge 1 - \epsilon$. For large enough n, $F_n(a) \le 2\epsilon$ and $F_n(b) \ge 1 - 2\epsilon$.

  Pick a $\delta > 0$. Divide the interval $(a,b]$ to finite number N_δ of subintervals $\mathcal{X}_j := (a_j, a_{j+1}]$ $a = a_1 < a_2 < ... < a_{N_\delta+1} = b$ such that all end points are continuity points of F and the fluctuation of f in each $\mathcal{X}_j$ is less than δ (We can do this since any continuous function f is uniformly continuous in a compact interval).

Define $\hat{f} = \sum_{j=1}^{N_\delta} f(a_j)\mathbf{1}_{\mathcal{X}_j}$. Since $\lim_n F_n(a_j) = F(a_j)$ for all $1 \le j \le N$

$$\int \hat{f} dF_n = \sum_{j=1}^{N_\delta} f(a_i)(F_n(a_{j+1}) - F_n(a_j))$$

and taking $n \to \infty$ we obtain that

$$\begin{aligned} \lim_{n\to\infty} \int \hat{f} dF_n &= \lim_{n\to\infty} \sum_{j=1}^{N_\delta} f(a_i)(F_n(a_{j+1}) - F_n(a_j)) \\ &= \sum_{j=1}^{N_\delta} f(a_i)(F(a_{j+1}) - F(a_j)) = \int \hat{f} dF. \end{aligned}$$

Since f is bounded by M and $\hat{f} = 0$ on $(-\infty, a] \cup (b, \infty)$

$$\begin{aligned} \left| \int f dF_n - \int \hat{f} dF_n \right| &\le \left| \int_{[a,b]} f dF_n - \int_{[a,b]} \hat{f} dF_n \right| + 4M\epsilon \\ &\le \int_{[a,b]} \left| f - \hat{f} \right| dF_n + 4M\epsilon \le \delta + 4M\epsilon. \end{aligned}$$

Similarly

$$\left| \int f dF - \int \hat{f} dF \right| \le \delta + 2M\epsilon$$

and by triangle inequality we conclude

$$\limsup | \int f dF_n - \int f dF | \le 2\delta + 6M\epsilon.$$

Since $\epsilon, \delta > 0$ are arbitrary, we are done.

- $(ii) \Rightarrow (iii)$: Consider the bounded continuous function, $f = e^{itx}$.
- $(iii) \Rightarrow (i)$: This is the most interesting part of the Levy-Cramer Theorem.

First, we prove a stronger version with a lesser assumption on $\phi(t)$.

Let $\phi_n(t)$ be the characteristic function of F_n, for all $n \ge 1$. Assume $\phi_n(t) \to \phi(t)$ for all real t and $\phi(t)$ is continuous at $t = 0$. Then $\phi(t)$ is the characteristic function of some distribution F and $F_n \overset{W}{\Longrightarrow} F$.

Step 1: Let $r_1, r_2, ...$ be some enumeration of rationals and F_n is the distribution function corresponding to ϕ_n. Since $F_n(r_1)$ is a bounded sequence hence there exists a convergent subsequence $F_{n_k^{(1)}}(r_1)$. Again since $F_{n_k^{(1)}}(r_2)$ is a bounded sequence, there is a convergent subsequence (a subsubsequence of F_n), $F_{n_k^{(2)}}(r)$ that converges at both r_1 and r_2. By induction proceed to create subsequences of previously defined subsequences that also converge at the next rational point. Hence $F_{n_k^{(j)}}(r)$ will converge pointwise at all points $r_1, .., r_j$. Now define $G_k(x) = F_{n_k^{(k)}}(x)$. Observe that this sequence converges at all rational points r_i. (This is called the diagonalization argument.) Call this limit function on rationals to be $G_\infty(r)$.

Step 2: From $G_\infty(r)$, which is defined on rationals, define the function $G(x)$ on the real line as $G(x) = \inf\limits_{\substack{r>x \\ r\in\mathbb{Q}}} G_\infty(r)$. From definition, $G(x)$ is clearly non-decreasing.
If $x_n \downarrow x$ for any rational $r > x$, for large enough n, $r > x_n$ which allows to conclude $G_\infty(r) \geq \inf_n G(x_n) \geq G(x)$. Taking infimum over $r > x$ we get $G(x) \geq \inf_n G(x_n) \geq G(x)$, establishing right continuity.

Step 3: Here we show that at every continuity point of $G(x)$, $\lim_k G_k(x) = G(x)$. For any rational $r > x$ note that $G_k(r) \geq G_k(x)$; hence

$$G_\infty(r) = \lim_k G_k(r) \geq \limsup_k G_k(x).$$

Taking infimum over $r > x$, we obtain

$$G(x) \geq \limsup_k G_k(x).$$

On the other hand, for any $y < x$, take a rational r such that $y < r < x$. Then

$$\liminf_k G_k(x) \geq \lim_k G_k(r) = G_\infty(r) \geq G(y).$$

Since x is a point of continuity of G, letting $y \uparrow x$ yields that

$$G(x) \geq \limsup_k G_k(x) \geq \liminf_k G_k(x) \geq \lim_{y\uparrow x} G(y) = G(x).$$

Thus $F_{n_k^{(k)}}(x)$ converges pointwise to a right continuous, non-decreasing function, $G(x)$ at all continuity points of $G(x)$. Note that $0 \leq G(-\infty) \leq G(\infty) \leq 1$.

Remark 2.2.1. It is useful to encapsulate what we have obtained so far. Given any sequence of distributions F_n, we showed that there is a sub-sequence F_{n_k}, that converge pointwise to a right continuous, non-decreasing function, $G(x)$ at all continuity points of $G(x)$, and $0 \leq G(-\infty) \leq G(\infty) \leq 1$.

Step 4

$$\begin{aligned}
\frac{1}{2T}\int_{-T}^{T} \phi_n(t)dt &= \frac{1}{2T}\int_{-T}^{T}\left(\int e^{itx} dF_n(x)\right)dt \\
&= \int \left(\frac{1}{2T}\int_{-T}^{T} e^{itx}dt\right) dF_n(x) \qquad \text{(Fubini)} \\
&= \int \frac{\sin(Tx)}{Tx} dF_n(x).
\end{aligned}$$

Observe that

$$\begin{aligned}
\int\left|\frac{\sin(Tx)}{Tx}dF_n(x)\right| &\leq \int_{x\in(-l,l]}\left|\frac{\sin(Tx)}{Tx}dF_n(x)\right| + \int_{x\notin(-l,l]}\left|\frac{\sin(Tx)}{Tx}dF_n(x)\right| \\
&\leq \int_{x\in(-l,l]} dF_n(x) + \frac{1}{Tl}\int_{x\notin(-l,l]} dF_n(x) \\
&\leq F_n(l) - F_n(-l) + \frac{1}{Tl}(1 - F_n(l) + F_n(-l))
\end{aligned}$$

$$= (F_n(l) - F_n(-l))\left(1 - \frac{1}{Tl}\right) + \frac{1}{Tl}.$$

Thus

$$\left|\frac{1}{2T}\int_{-T}^{T} \phi_n(t)dt\right| \leq (F_n(l) - F_n(-l))\left(1 - \frac{1}{Tl}\right) + \frac{1}{Tl}.$$

In particular

$$\left|\frac{1}{2T}\int_{-T}^{T} \phi_{n_k^{(k)}}(t)dt\right| \leq \left(F_{n_k^{(k)}}(l) - F_{n_k^{(k)}}(-l)\right)\left(1 - \frac{1}{Tl}\right) + \frac{1}{Tl}.$$

Again, applying Bounded convergence theorem (to interchange limit and integration) and taking $k \to \infty$ and observing that $l \in \mathbb{N} \subseteq Q$, we get

$$\left|\frac{1}{2T}\int_{-T}^{T} \phi(t)dt\right| = (G_\infty(l) - G_\infty(-l))\left(1 - \frac{1}{Tl}\right) + \frac{1}{Tl}.$$

Let $T = \frac{1}{\sqrt{l}}$ and letting $l \to \infty$ we obtain (from the continuity of $\phi(t)$ at $t = 0$) and definition, non-decreasingness of $G_\infty(r), G(x)$

$$1 = G(\infty) - G(-\infty),$$

implying that $G(x)$ is a distribution function.

Thus $F_{n_k^{(k)}} \Rightarrow G$; however $\phi_{n_k^{(k)}}(t) \to \phi(t)$. Thus (as $(i) \implies (ii)$) $\phi(t)$ is the characteristic function of $G(x)$; and further G is uniquely determined by $\phi(t)$.

Step 5 To complete the argument, we need to show that $F_k \Rightarrow G$, i.e. the entire sequence converges pointwise at all continuity points of G. Assume note, then one can find a subsequence F_{k_n} and a continuity point x_0, of $G(x)$, such that $\lim_n |F_{k_n}(x_0) - G(x_0)| > \epsilon$, for some $\epsilon > 0$. Starting with this subsequence $F_{k_n}(x_0)$ we further find a sub-subsequence that converges to a distribution function; however since $\phi_{k_n}(t) \to \phi(t)$, that distribution function must be $G(x)$, yielding a contradiction.

□

DEFINITION 2.2.2. A sequence of distribution functions F_n is called *uniformly-tight* (or tight) if for every $\epsilon > 0$, there exists M_ϵ such that $F_n(M_\epsilon) \geq 1 - \epsilon, \forall n$ and $F_n(-M_\epsilon) \leq \epsilon, \forall n$.

THEOREM 2.2.2 (Prokhorov). *If F_n is a sequence of uniformly-tight probability distributions then there is a subsequence, F_{n_k}, that converge weakly to a distribution G.*

PROOF. The proof is rather immediate from Remark 2.2.1. We know, from Remark 2.2.1, that there is a subsequence F_{n_k} that converges pointwise to a right continuous, non-decreasing function, $G(x)$ at all continuity points of $G(x)$, and $0 \leq G(-\infty) \leq G(\infty) \leq 1$. Given $\epsilon > 0$, let $a > M_\epsilon$ and $b < M_\epsilon$ be continuity points of G. Then we know that $G(a) = \lim_k F_{n_k}(a) \geq 1 - \epsilon$, and $G(b) = \lim_k F_{n_k}(b) \leq \epsilon$. Since G is non-decreasing and ϵ is arbitrary, we obtain that $\lim_{x\to\infty} G(x) = 1$ and $\lim_{x\to-\infty} G(x) = 0$, establishing that G is a probability distribution. □

THEOREM 2.2.3 (Portmanteau). *Let (Ω, d) be a complete metric space and $\mathcal{B}$ be the corresponding Borel σ-algebra. Let P_n be a sequence of probability distributions on $(\Omega, \mathcal{B})$. If $P_n \Rightarrow P$, then*

1. *for all closed C, $\limsup_n P_n(C) \to P(C)$.*
2. *for all open sets C, $\liminf_n P_n(C) \geq P(C)$.*
3. *for all continuity sets of P, i.e. open sets C such that $P(closure(C) \setminus C) = 0$, $\lim_n P_n(C) = P(C)$.*

PROOF. (1) Let C be closed. Define the non-negative random variable

$$X(\omega, C) = \inf_{y \in C} d(y, \omega).$$

When $\omega \in C$, note that $X(\omega, C) = 0$. Let $f_k(X) = \left(\frac{1}{1+X(\omega,C)}\right)^k$. For every $k \geq 1$, note that

$$1_C \leq f_k(X), \quad \Longrightarrow P_n(C) \leq \int f_k(X) dP_n = \int f_k(x) dF_n.$$

Since $f_k(x) = \frac{1}{(1+x)^k}$ is bounded continuous function (on $x \geq 0$),

$$\limsup_n P_n(C) \leq \lim_n \int f_k dF_n = \int f_k dF.$$

Since f_k is bounded and decreases point wise to 1_C which is bounded, monotone convergence theorem (as k goes to infinity) yields

$$\limsup_n P_n(C) \leq \lim_k \int f_k dF \overset{(MCT)}{=} \int f dF = P(C).$$

(2) Taking complements yields this part.
(3) Combining the two yields the third part.

□

CHAPTER 3

Independence, Law of Large Numbers, and Limit Theorems

3.1. Independence

We assume an underlying probability space $(\Omega, \mathcal{F}, P)$ and that the random variables being considered are defined on this space.

DEFINITION 3.1.1. Events A and B are independent if $P(A \cap B) = P(A)P(B)$.

DEFINITION 3.1.2. Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$. We say that $\mathcal{A}$ and $\mathcal{B}$ are independent if $P(A \cap B) = P(A)P(B)\ \forall A \in \mathcal{A}, A \in \mathcal{B}$.

Lemma 3.1.1. *Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$. Further assume that $\mathcal{B}$ is a π-system. Then $\mathcal{A}$ and $\sigma(\mathcal{B})$ are independent.*

PROOF. Define $\mathcal{I}(\mathcal{A}) = \{F \in \mathcal{F} : P(F \cap A) = P(F)P(A)\ \forall A \in \mathcal{A}\}$. We will show that $\mathcal{I}(\mathcal{A})$ is a λ-system, and further since we are given that it contains that π-system $\mathcal{H}$, it must contain $\mathcal{D}(\mathcal{H})$, the λ-system generated by $\mathcal{H}$, and by the $\pi - \lambda$ theorem, Theorem 1.1.2, we see that $\mathcal{I}(\mathcal{A}) \supseteq \sigma(\mathcal{H})$.

To see that $\mathcal{I}(\mathcal{A})$ is a λ-system, it is immediate that $\Omega \in \mathcal{I}(\mathcal{A})$. If $F_2 \supseteq F_1 \in \mathcal{I}(\mathcal{A})$, then as

$$
\begin{aligned}
P((F_2 \setminus F_1) \cap A) &= P((F_2 \cap A) \setminus (F_1 \cap A)) = P(F_2 \cap A) - P(F_1 \cap A) \\
&= (P(F_2) - P(F_1))P(A) = P(F_2 \setminus F_1)P(A)
\end{aligned}
$$

for any $A \in \mathcal{A}$, we have $F_2 \setminus F_1 \in \mathcal{I}(\mathcal{A})$. Finally, if $\{F_n\} \in \mathcal{I}(\mathcal{A})$ is an increasing sequence of sets, and $F_\infty := \cup_n F_n$, then from the monotone limits property off probability measures (Exercise 1.1.1), note that

$$
P(F_\infty \cap A) = \lim_n P(F_n \cap A) = \lim_n P(F_n)P(A) = P(A)\left(\lim_n P(F_n)\right) = P(F_\infty)P(A).
$$

This implies that $F_\infty \in \mathcal{I}(\mathcal{A})$, completing the proof that $\mathcal{I}(\mathcal{A})$ is a λ-system. $\square$

We can extend this definition of independence to more than two collections as follows.

DEFINITION 3.1.3. Let $\mathcal{A}_1, \mathcal{A}_2, ...$ be subsets of $\mathcal{F}$. We say that the collection $\mathcal{A}_1, \mathcal{A}_2, ..$ are independent (or mutually independent) of each other if $P(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_l}) = \prod_{j=1}^{l} P(A_{i_j})\ \forall A_{i_j} \in \mathcal{A}_{i_j}, 1 \leq j \leq l, 1 \leq l \leq k$.

We now turn to independence of random variables.

DEFINITION 3.1.4. Random variables X and Y are independent if $\forall A, B \in \mathcal{B}_R$,

$$
P\{\omega : X(\omega) \in A, Y(\omega) \in B\} = P\{\omega : X(\omega) \in A\}P\{\omega : Y(\omega) \in B\}.
$$

REMARK 3.1.1. Equivalently, random variables X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.

The definition of independence extends to a finite number of random variables. $X_1, ...X_n$ are independent if $\forall A_i \in \mathcal{B}_R$,

$$P(X_1 \in A_1, ..., X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

REMARK 3.1.2. Let us take $\Omega = (0,1], \mathcal{F} = \mathcal{B}_{(0,1]}$, and the probability measure to the Lebesgue measure. Suppose you want to generate two independent and identically distributed Bernoulli random variables X and Y, such that $P(X=0) = P(Y=0) = \frac{1}{2}$ and $P(X=1) = P(Y=1) = \frac{1}{2}$.

One way to generate this is the following: $\omega \in (0, \frac{1}{4}] \implies X(\omega) = Y(\omega) = 0$, $\omega \in (\frac{1}{4}, \frac{1}{2}] \implies X(\omega) = 0, Y(\omega) = 1$, $\omega \in (\frac{1}{2}, \frac{3}{4}] \implies X(\omega) = 1, Y(\omega) = 0$, $\omega \in (\frac{3}{4}, 1] \implies X(\omega) = 1, Y(\omega) = 1$. Observe that X and Y are independent with the right distribution as desired.

Now, let us do something more interesting. Again we desire two independent and identically distributed random variables X and Y that are both uniformly distributed in the interval $(0,1]$. Clearly if we set $X(\omega) = \omega$, then X has the right distribution. But it is hard to describe an independent Y easily.

So how does one generate two independent and identically distributed random variables X and Y that are both uniformly distributed in the interval $(0,1]$. This is where product spaces come to the rescue. Consider $(0,1] \times (0,1]$, and consider the Lebesgue measure to be the probability measure. Set $X(\omega_1, \omega_2) = \omega_1$ and $Y(\omega_1, \omega_2) = \omega_2$. Now, observe (see exercise below) that we have two independent and identically distributed random variables X and Y that are both uniformly distributed in the interval $(0,1]$. Thus independent random variables and product spaces go hand-in-hand.

More generally, we have the following exercise.

EXERCISE 3.1.1. Two random variables X and Y defined on the same probability space are independent iff the measure induced by the joint mapping $(\omega_1, \omega_2) \to (X(\omega_1), Y(\omega_2))$, is the *product measure*.

EXERCISE 3.1.2. Show that for independent random variables X and Y and measurable functions f and g where $E[|f(X)|]$ and $E[|g(Y)|]$ are finite

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

REMARK 3.1.3. This implies that $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$. Note that the reverse does not hold, meaning that

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) \not\Rightarrow X \text{ and } Y \text{ are independent}$$

On the other hand note that

$$E[e^{i(t_1 X + t_2 Y)}] = E[e^{it_1 X}]E[e^{it_2 Y}] \quad \forall t_1, t_2 \iff X \text{ and } Y \text{ are independent}$$

3.2. Weak Law of Large Numbers

THEOREM 3.2.1. *$X_1, ..., X_n$ are pairwise independent random variables, satisfying $E(X_i) = 0$ and $E(X_i^2) = \sigma_i^2 < B$. Then $\frac{S_n}{n} \to 0$ in measure, where $S_n = X_1 + ... + X_n$.*

PROOF. Note that

$$E\left(\frac{S_n}{n}\right)^2 = \frac{1}{n^2}E(X_i^2) \le \frac{B}{n}.$$

Hence

$$\epsilon^2 P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) \le E\left(\frac{S_n}{n}\right)^2 \le \frac{B}{n}.$$

□

DEFINITION 3.2.1. A collection of random variables $\{X_\alpha\}$ is called *uniformly integrable* (U-I) if

$$\lim_{M\to\infty} \sup_\alpha E(|X_\alpha| 1_{|X_\alpha|>M}) = 0.$$

THEOREM 3.2.2. *If $X_1, ..., X_n$ are pairwise independent and uniformly integrable random variables with mean zero, then $\frac{S_n}{n} \to 0$ in measure.*

PROOF. Given $\delta > 0$, take M (it exists by uniform integrability) such that

$$E\left(|X_i| 1_{|X_i|>M}\right) < \delta, \quad \forall i.$$

Truncate X_i to two parts X_i^T and Y_i^T as follows

$$\begin{array}{ll} X_i^T = X_i \mathbf{1}_{\{|X_i| \le M\}} & E[X_i^T] = a_i \\ Y_i^T = X_i \mathbf{1}_{\{|X_i| > M\}} & E[Y_i^T] = -a_i \\ G_n^T = X_1^T + ... + X_n^T - \sum_{i=1}^n a_i & B_n^T = Y_1^T + ... + Y_n^T + \sum_{i=1}^n a_i. \end{array}$$

Note that $|a_i| < \delta$ for all i. Clearly $X_i = X_i^T + Y_i^T$. Now consider the quantity

$$\begin{aligned} E[|\frac{S_n}{n}|] &= E[|\frac{G_n^T + B_n^T}{n}|] \le E[|\frac{G_n^T}{n}|] + E[|\frac{B_n^T}{n}|] \\ &\le E[|\frac{G_n^T}{n}|] + 2\delta. \end{aligned} \tag{3.1}$$

Where the last inequality follows from

$$E[|\frac{B_n^T}{n}|] = E[|\frac{\sum Y_i^T + a_i}{n}|] \le \sum \frac{E[|Y_i^T|] + |a_i|}{n} \overset{U-I}{\le} 2\delta.$$

Note that

$$E((X_i^T - a_i)^2) \le (M + |a_i|)^2 \le (M + \delta)^2.$$

Therefore (from Cauchy-Schwartz)

$$E[|\frac{G_n^T}{n}|] \le \sqrt{E[(\frac{G_n^T}{n})^2]} \le \sqrt{\frac{1}{n}(M+\delta)^2}.$$

By Markov's inequality

$$P(|\frac{S_n}{n}| > \epsilon) \le \frac{E[|\frac{S_n}{n}|]}{\epsilon} \le \frac{\frac{1}{\sqrt{n}}(M+\delta) + 2\delta}{\epsilon}.$$

Taking limsup we get $P(|\frac{S_n}{n}| > \epsilon)$ is upper bounded by $\frac{2\delta}{\epsilon}$; but since $\delta > 0$ is arbitrary, we are done. □

3.3. Strong Law of Large Numbers

So far we considered convergence in measure of a sequence of random variables. We could get away without showing that one can actually define an infinite sequence of independent random variables. But one can formally talk about such a collection. To do so, let us start with a lemma.

Lemma 3.3.1. *Given any Borel set $A \subset \mathbb{R}^n$ and a probability measure P, for any $\epsilon > 0$ there exists a closed and bounded set $K_\epsilon \subset A$ such that $P(A \setminus K_\epsilon) < \epsilon$.*

Proof. Boundedness of K_ϵ is simple. Just truncate the space to within a closed ball, whose outside probability is at most, say $\frac{\epsilon}{3}$. Let us try to prove it for $n = 1$. Consider $A = (a, b]$. By right continuity the probability of $[a + \frac{1}{n}, b]$ goes to $(a, b]$. Hence the statement is true for such intervals, as well as finite disjoint union of such intervals. The set of all such A is a monotone class (why?) (Hint: truncate the countable union to a finite union with a small probability loss; for each set now approximate by closed set, and take their finite union. Quantifiers can be chosen show that total probability loss is small.)

For larger n start with finite disjoint union of rectangles and proceed similarly. $\square$

A family of probability measures P_n on $\mathbb{R}^n$ is said to be *consistent* if $P_{n+1}((A, \mathbb{R})) = P_n(A)$ for all A, a Borel set in $\mathbb{R}^n$.

Let $\Omega = \{x_1, ..., x_n, ...\} = \mathbb{R}^\infty$ be the space of all real sequences. Consider the natural sigma field, Σ, generated by the field of all finite dimensional cylinder sets, i.e., sets of the form $A \times \mathbb{R} \times \mathbb{R} \cdots$, where A is Borel set in $\mathbb{R}^n$ for some n.

Theorem 3.3.2 (Kolmogorov's consistency theorem). *Given a consistent family of finite-dimensional distributions P_n, there is a unique P on (Ω, Σ) such that for every n, under the natural projection Π_n, the induced measure $P\Pi_n^{-1}(A) = P_n(A)$ for all $A \in \mathcal{B}_n$.*

Proof. Consider the field of all finite dimensional cylinder sets. For A in this field define $P(A) = P_n(A)$ (defined uniquely because of the finite-dimensional consistency of the family). Suffices to show that this P is countably additive on the field (Caratheodory's extension theorem does the rest). Assume not. Hence there is $A_n \downarrow \emptyset$ but $P(A_n) \geq \delta > 0$. Since A_n is a finite dimensional cylinder set of $B_{k_n} \in \mathbb{R}^{k_n}$, from lemma, we can find a closed and bounded (hence compact) set $K_{k_n} \subset B_{k_n}$ with $P_n(B_{k_n} \setminus K_{k_n}) \leq \frac{\delta}{2^{n+2}}$. Let $C_n \subset A_n$ denote the closed cylinder generated by K_{k_n}. Let $D_n = \cap_{m \leq n} C_m$. Note that

$$A_n \setminus D_n = \cup_{m \leq n}(A_n \setminus C_m) \subset \cup_{m \leq n}(A_m \setminus C_m).$$

Hence

$$P(A_n \setminus D_n) \leq \sum_{m=1}^{n} P(A_m \setminus C_m) = \sum_{m=1}^{n} P_n(B_{k_m} \setminus K_{k_m}) \leq \sum_{m=1}^{n} \frac{\delta}{2^{m+2}} \leq \frac{\delta}{4}.$$

Therefore $P(D_n) \geq \frac{3\delta}{4}$, hence D_n is non-empty closet set of Ω. Further $D_n \downarrow \emptyset$.

Define $E_{n,k} = \Pi_k(D_n)$. Clearly from construction $E_{n,k}$ is a decreasing sequence (in n) of compact sets, and since D_n is non-empty, each $E_{n,k}$ is non-empty. Therefore let $\Pi_k(D_n) \downarrow F_k$, and as $\Pi_k(D_n)$ are a decreasing sequence of non-empty compact sets of $\mathbb{R}^k$, we get that $F_k = \cap_n \Pi_k(D_n)$ is non-empty subset of $\mathbb{R}^k$.

For $k \geq 2$ observe that $\Pi_{k-1}(E_{n,k}) = \Pi_{k-1}(D_n) = E_{n,k-1}$. Consequently, $\Pi_{k-1}(F_k) = \Pi_{k-1}(\cap_n \Pi_k(D_n)) = \cap_n(\Pi_{k-1}(\Pi_k(D_n))) = \cap_n \Pi_{k-1}(D_n) = F_{k-1}$.

Now construct a vector (sequence) $\mathbf{x} = (x_1, x_2, ...) \in \mathbb{R}^\infty$ inductively as follows. Let $x_1 \in F_1$. Since $\Pi_1(F_2) = F_1$, there exists x_2 such that $(x_1, x_2) \in F_2$. Proceeding inductively, let $(x_1, ..., x_{k-1}) \in F_{k-1}$. Since $\Pi_{k-1}(F_k) = F_{k-1}$, we can find x_k such that $(x_1, .., x_k) \in F_k$. Note that this limit point belongs to D_m for every m, contradicting $D_m \downarrow \emptyset$. □

We first prove a vanilla version of the main theorem by imposing a constraint on the fourth moment of the random variables.

THEOREM 3.3.3. *$X_1, ..., X_n$ are fourwise-independent random variables such that $E(X_i) = 0$ and $E[X_i^4] = c < \infty$. Then $\frac{S_n}{n} \to 0$ almost surely.*

PROOF. Consider

$$E\left(\left(\frac{S_n}{n}\right)^4\right) = \frac{1}{n^4}\left(\sum_i E(X_i^4) + 6\sum_{i<j} E(X_i^2)(X_j^2)\right). \tag{3.2}$$

Note that all the other cross terms are zero. (why?). Further $E(X_i^2) \leq \sqrt{c}$. Therefore

$$E\left(\frac{S_n}{n}^4\right) \leq \frac{1}{n^4}(n + 3n(n-1))c \leq \frac{3c}{n^2}.$$

Therefore, by Markov's inequality,

$$P(|\frac{S_n}{n}| > \epsilon) \leq \frac{1}{\epsilon^4} E\left(\frac{S_n}{n}^4\right) \leq \frac{3c}{\epsilon^4 n^2}.$$

Let $A_n = \{\omega : |\frac{S_n}{n}| > \epsilon\}$. Since $\sum_n P(A_n) < \infty$, from Borel-Cantelli Lemma we know that $P(A_n\ i.o.) = 0$.

Note that the set $B_\epsilon := A_n\ i.o$ is same as the set of omega's for which $\limsup_n |\frac{S_n}{n}| > \epsilon$. We have $P(B_\epsilon) = 0$ for any $\epsilon > 0$; hence $P(\cup_m B_{1/m}) = 0$. Therefore the set of omega's for which $\limsup_n |\frac{S_n}{n}| > 0$ has probability 0, establishing almost sure convergence.

□

THEOREM 3.3.4. *Let $X_1, \ldots, X_n, \ldots$ be pairwise independent and identically distributed non-negative random variables (say same as X) with $E(X) = \mu < \infty$ and $S_n = \sum_{i=1}^n X_i$. Then $\frac{S_n}{n} \to \mu$ a.s.*

PROOF. *Step 1*: (Truncation) Let $Y_k = X_k 1_{(X_k \leq k)}$. Let $T_n = \sum_{k=1}^n Y_k$. Define $Z_k = X_k - Y_k$. Note that

$$\begin{aligned}
\sum_{k=1}^\infty P(Z_k > 0) &= \sum_{k=1}^\infty P(X > k) \\
&\leq \int_0^\infty P(X > t)dt = \int_0^\infty \int 1_{X>t} dP dt \\
&= \int \int_0^X 1_{X>t} dt dP = \int X dP = \mu < \infty.
\end{aligned}$$

Hence (by Borel-Cantelli) $P(Z_k > 0\ i.o.) = 0$. Let $A = \{\omega : Z_k(w) = 0 \text{ eventually}\}$. Thus $P(A) = 1$.

Let $A_m = \{\omega : Z_k(\omega) = 0 \ \forall k \geq m\}$. We know that $A_m \uparrow A$. For all $\omega \in A_m$ we have $\frac{\sum_{k=1}^n Z_k}{n} = \frac{\sum_{k=1}^{\min\{n,m\}} Z_k}{n}$. Thus

$$0 \leq \liminf_n \frac{\sum_{k=1}^n Z_k}{n} \leq \limsup_n \frac{\sum_{k=1}^n Z_k}{n} \leq \limsup_n \frac{\sum_{k=1}^m Z_k}{n} = 0.$$

Taking union over m we obtain that

$$0 = \lim_n \frac{\sum_{k=1}^n Z_k}{n} = \lim_n \frac{S_n - T_n}{n}, \forall \omega \in A.$$

As $T_n \leq S_n$ implies that $\limsup_n \frac{E(T_n)}{n} \leq \mu$.

Note that $E(Z_n) = E(X1_{X>n})) \to 0$, hence (by Cesaro-sum)

$$\lim_n \frac{1}{n} E(S_n - T_n) = 0.$$

Therefore the theorem is proved if we establish that

$$\lim_n \frac{T_n - E(T_n)}{n} \to 0 \ a.s. \tag{3.3}$$

Step 2: (Subsequence convergence). Let $\alpha > 1$ be arbitrary and let $k_n = \lfloor \alpha^n \rfloor$. Then note that

$$\begin{aligned}
\sum_{n=1}^{\infty} P(|T_{k_n} - E(T_{k_n})| \geq \epsilon k_n) &\leq \sum_{n=1}^{\infty} \frac{\operatorname{var}(T_{k_n})}{\epsilon^2 k_n^2} \\
&= \sum_{n=1}^{\infty} \frac{1}{\epsilon^2 k_n^2} \sum_{m=1}^{k_n} \operatorname{var}(Y_m) \\
&= \sum_{m=1}^{\infty} \operatorname{var}(Y_m) \left(\sum_{n: k_n \geq m} \frac{1}{\epsilon^2 \lfloor \alpha^n \rfloor^2} \right) \\
&\leq \sum_{m=1}^{\infty} \frac{\operatorname{var}(Y_m)}{\epsilon^2 m^2} \left(\sum_{n=0}^{\infty} \frac{4}{\alpha^{2n}} \right) \\
&= 4 \frac{\alpha^2}{\epsilon^2(\alpha^2 - 1)} \sum_{m=1}^{\infty} \frac{\operatorname{var}(Y_m)}{m^2} \\
&\leq 16 \frac{\alpha^2}{\epsilon^2(\alpha^2 - 1)} E(X) < \infty.
\end{aligned}$$

The first inequality follows from Chebychev's inequality. To obtain the second inequality let n_0 be the smallest n such that $k_n \geq m$. Then note that $\forall n \geq n_0$ we have

$$\lfloor \alpha^n \rfloor = \lfloor \alpha^{n_0} \alpha^{n - n_0} \rfloor \geq \lfloor \alpha^{n_0} \rfloor \lfloor \alpha^{n - n_0} \rfloor \geq m \frac{\alpha^{n - n_0}}{2}.$$

The third inequality follows from Lemma 3.3.5 below.

Hence (by Borel-Cantelli)

$$\lim_n \frac{T_{k_n} - E(T_{k_n})}{k_n} = 0 \ a.s. \implies \lim_n \frac{T_{k_n}}{k_n} = \mu \ a.s.$$

since $\frac{E(T_{k_n})}{k_n} \to \mu$.

Step 3: (Sandwich) Note that

$$\frac{\alpha^{n+1}-1}{\alpha^n} \leq \frac{k_{n+1}}{k_n} \leq \frac{\alpha^{n+1}}{\alpha^n - 1}, \implies \lim_n \frac{k_{n+1}}{k_n} = \alpha.$$

For every m let n_m be such that $k_{n_m} \leq m < k_{n_m+1}$. We know that

$$\frac{k_{n_m}}{k_{n_m+1}} \frac{T_{k_{n_m}}}{k_{n_m}} \leq \frac{T_m}{m} \leq \frac{k_{n_m+1}}{k_{n_m}} \frac{T_{k_{n_m+1}}}{k_{n_m+1}}.$$

Thus on the set where $\lim_n \frac{T_{k_n}}{k_n} = \mu$ we have

$$\frac{\mu}{\alpha} \leq \liminf_m \frac{T_m}{m} \leq \limsup_m \frac{T_m}{m} \leq \alpha\mu.$$

Since $\alpha > 1$ is arbitrary, we have $\lim_m \frac{T_m}{m} = \mu$ *a.s.* □

Lemma 3.3.5. *Let $X_1, \ldots, X_n$ be pairwise independent and identically distributed non-negative random variables (say same as X) with $E(X) = \mu < \infty$. Let $Y_k = X_k 1_{(X_k \leq k)}$. Then*

$$\sum_{k=1}^{\infty} \frac{\operatorname{var}(Y_k)}{k^2} \leq 4E(X).$$

Proof. Recall that we have $Y_k \geq 0$. Note that

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\operatorname{var}(Y_k)}{k^2} &\leq \sum_{k=1}^{\infty} \frac{E(Y_k^2)}{k^2} \\
&= \sum_{k=1}^{\infty} \frac{\int \left(\int_0^{Y_k} 2t dt\right) dP}{k^2} \\
&= \sum_{k=1}^{\infty} \frac{\int \int_0^{\infty} 2t 1_{Y_k > t} dt dP}{k^2} \\
&\overset{(a)}{=} \sum_{k=1}^{\infty} \frac{\int_0^{\infty} 2t P(Y_k > t) dt}{k^2} \\
&\leq \sum_{k=1}^{\infty} \frac{\int_0^{k} 2t P(X > t) dt}{k^2} \\
&\overset{(b)}{=} \int_0^{\infty} 2t P(X > t) \left(\sum_{k : k \geq \max\{t,1\}} \frac{1}{k^2} \right) dt \\
&\leq \int_0^{\infty} P(X > t) \left(2t \frac{\pi^2}{6} 1_{t \leq 1} + 2t \left(\frac{\pi^2}{6} - 1 \right) 1_{1 < t \leq 2} + \frac{2t}{t-1} 1_{t > 2} \right) dt \\
&\leq \int_0^{\infty} 4P(X > t) dt \\
&= 4E(X).
\end{aligned}$$

Here $(a), (b)$ use Fubini, the second inequality uses that $X_k \geq Y_k$ implying $P(Y_k > t) \leq P(X \geq t)$, and the third inequality is just a case decomposition of t into three intervals. □

THEOREM 3.3.6. *(Kolmogorov's zero-one law) Suppose $X_1, ..., X_n, ...$ are independent random variables. Let $\mathcal{F}_n = \sigma(X_n, ...)$ and define $\mathcal{F}_\infty = \underset{n}{\cap}\mathcal{F}_n$. Then*

$$\forall A \in \mathcal{F}_\infty, P(A) \in \{0, 1\}.$$

PROOF. We will see that $A \in \mathcal{F}_\infty$ is independent of itself. W.l.o.g. let $P(A) > 0$.

Given any $n \geq 1$, we know that $A \in \mathcal{F}_{n+1}$. Since $A \in \mathcal{F}^\infty \subset \mathcal{F}^{n+1}$ it is independent of $\mathcal{F}'_n := \sigma(X_1, ..., X_n)$. Therefore A is also independent of sets in the field $\mathcal{F}' = \cup_n \mathcal{F}'_n$.

Consider the class of sets, $\mathcal{A}$, that are independent of A. Define two countably probability additive probability measures on $\sigma(\mathcal{F}')$, one according to $(i)\frac{P(A\cap B)}{P(A)}$ and the other according to $(ii)P(B)$. Note that these two probability measures agree on the field $\mathcal{F}'$; hence must agree on $\sigma(\mathcal{F}')$.

Note that $\mathcal{F}_n$ is generated by finite dimensional cylinder sets; but each of these finite dimensional cylinder sets will be in some $\mathcal{F}'_n$. Hence $\mathcal{F}_n \subset \sigma(\mathcal{F}')$. Thus

We have

$$\underset{n}{\cap}\mathcal{F}_n = \mathcal{F}_\infty \subset \sigma(\mathcal{F}').$$

and since $A \in \mathcal{F}^\infty$, A is independent of itself hence $P(A) = P^2(A)$. $\square$

3.4. Central limit theorem

3.4.1. Some useful lemmas regarding real and complex numbers.

Lemma 3.4.1.

$$\left|e^{-z} - 1 + z\right| \leq |z|^2,\ 0 \leq |z| \leq 1.$$

PROOF. From Taylor expansion, suffices to show that

$$\sum_{n=2}^{\infty} \frac{|z|^n}{n!} \leq |z|^2.$$

This follows immediately since $\frac{|z|^n}{n!} \leq \frac{|z|^2}{2^{n-1}}$ for $|z| \leq 1, n \geq 2$. $\square$

Lemma 3.4.2.

$$\left|e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!}\right| \leq \min\left\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right\}.$$

PROOF. Using integration by parts, verify that

$$e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds.$$

Therefore, we will bound the right-hand-side in two ways. Observe that

$$\left|e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!}\right| = \left|\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds\right| \leq \left|\frac{1}{n!}\int_0^{|x|} s^n ds\right| = \frac{|x|^{n+1}}{(n+1)!}.$$

On the other hand applying the same bound to $n-1$ yields

$$\left|e^{ix} - \sum_{m=0}^{n-1} \frac{(ix)^m}{m!}\right| \leq \frac{|x|^n}{n!}.$$

From triangle inequality, the second upper bound follows. $\square$

Lemma 3.4.3. *Let $z_1, .., z_n$ and $\omega_1, .., \omega_n$ be complex numbers whose absolute value is bounded by θ. Then*

$$\left|\prod_{m=1}^{n} z_m - \prod_{m=1}^{n} w_m\right| \leq \theta^{n-1} \sum_{m=1}^{n} |z_m - \omega_m|.$$

PROOF. Note that

$$\begin{aligned}\left|\prod_{m=1}^{n} z_m - \prod_{m=1}^{n} w_m\right| &\leq \left|z_1 \left(\prod_{m=2}^{n} z_m - \prod_{m=2}^{n} w_m\right)\right| + \left|(z_1 - \omega_1) \prod_{m=2}^{n} w_m\right| \\ &\leq \theta \left|\prod_{m=2}^{n} z_m - \prod_{m=2}^{n} w_m\right| + \theta^{n-1}|z_1 - \omega_1|.\end{aligned}$$

Bounding the first term (induction) finishes the proof.

□

Lemma 3.4.4. *Let X be a random variable such that $E[X] = 0$, $E[X^2] = \sigma^2$. Then*

$$\lim_{n\to\infty} n \left(\phi_X\left(\frac{t}{\sqrt{n}}\right) - 1 + \frac{E[X^2]}{2}\frac{t^2}{n}\right) = 0.$$

PROOF.

$$\begin{aligned}&n\left(\phi_X\left(\frac{t}{\sqrt{n}}\right) - 1 + \frac{E[X^2]}{2}\frac{t^2}{n}\right) \\ &\quad = nE\left(e^{iX\frac{t}{\sqrt{n}}} - 1 - iX\frac{t}{\sqrt{n}} + \frac{X^2}{2}\frac{t^2}{n}\right).\end{aligned}$$

Let us split the integral into two parts $|X| > \frac{\epsilon\sqrt{n}}{2|t|}$ and $|X| \leq \frac{\epsilon\sqrt{n}}{2|t|}$. Consider

$$\begin{aligned}&E\left(\left|n(e^{iX\frac{t}{\sqrt{n}}} - 1) - iX\sqrt{n}t + \frac{X^2}{2}t^2\right| 1_{|X|>\frac{\epsilon\sqrt{n}}{2|t|}}\right) \\ &\quad \leq 2nP\left(|X| > \frac{\epsilon\sqrt{n}}{2|t|}\right) + |t|\sqrt{n}E\left(|X|1_{|X|>\frac{\epsilon\sqrt{n}}{2|t|}}\right) + \frac{t^2}{2}E\left(|X|^2 1_{|X|>\frac{\epsilon\sqrt{n}}{2|t|}}\right).\end{aligned}$$

Each of the term goes to 0 as n goes to infinity. The last term clearly goes to zero (by dominated convergence theorem); and the first two terms are upper bounded by a constant times last term. For instance,

$$n1_{|X|>\frac{\epsilon\sqrt{n}}{2|t|}} \leq 4\frac{t^2}{\epsilon^2}|X|^2 1_{|X|>\frac{\sqrt{n}}{2|t|}}.$$

Now consider

$$E\left(\left|n(e^{iX\frac{t}{\sqrt{n}}} - 1) - iX\sqrt{n}t + \frac{X^2}{2}t^2\right| 1_{|X|\leq\frac{\epsilon\sqrt{n}}{2|t|}}\right).$$

From Lemma 3.4.2 we have $|e^{iz} - 1 - iz + \frac{z^2}{2}| \leq \frac{|z|^3}{6}, |z|^2$. Hence (for $\epsilon < 1$)

$$\left|n(e^{iX\frac{t}{\sqrt{n}}} - 1) - iX\sqrt{n}t + \frac{X^2}{2}t^2\right| 1_{|X|\leq\frac{\epsilon\sqrt{n}}{2|t|}} \leq \frac{n}{6}\left(\frac{|Xt|}{\sqrt{n}}\right)^3 1_{|X|\leq\frac{\epsilon\sqrt{n}}{2|t|}} \leq |Xt|^2.$$

Hence by dominated convergence theorem, the expectation goes to zero. □

THEOREM 3.4.5. *(Central Limit Thoerem, Vanilla Version) if $X_1, ..., X_n$ are i.i.d random variables with $E[X_i] = 0$, $E[X_i^2] = \sigma^2$. Then*

$$\frac{X_1 + ... + X_n}{\sqrt{n}} \stackrel{w}{\Rightarrow} \mathcal{N}(0, \sigma^2).$$

PROOF.

$$\begin{aligned} E[e^{it\frac{X_1+...+X_n}{\sqrt{n}}}] &= \prod_{i=1}^{n} E[e^{it\frac{X_i}{\sqrt{n}}}] \\ &= \phi_X(\frac{t}{\sqrt{n}})^n. \end{aligned}$$

Using previous lemma we obtain

$$\phi_X(\frac{t}{\sqrt{n}})^n \to e^{-\sigma^2 t^2/2}$$

which is the characteristic function of $\mathcal{N}(0, \sigma^2)$. □

THEOREM 3.4.6. *(Lindberg-Feller Theorem) For each n, let $X_{n,m}, 1 \leq m \leq n$, be independent random variables with $E(X_{n,m}) = 0$. Suppose*

(1) $\lim_n \sum_{m=1}^n E(X_{n,m}^2) = \sigma^2 > 0$

(2) For all $\epsilon > 0$, $\lim_n \sum_{m=1}^n E\left(X_{n,m}^2 1_{|X_{n,m}| > \epsilon}\right) = 0$.

Then $S_n = X_{n,1} + \cdots + X_{n,n} \stackrel{w}{\Rightarrow} \mathcal{N}(0, \sigma^2)$.

PROOF. Let $\phi_{n,m}(t) = E\left(e^{itX_{n,m}}\right)$ and $\sigma_{n,m}^2 = E(X_{n,m}^2)$. Suffices (by Levy's continuity theorem) to show that

$$\lim_{n\to\infty} \prod_{m=1}^{n} \phi_{n,m}(t) = e^{-t^2\sigma^2/2}.$$

Let $a_n = \sup_{1\leq m\leq n} \sigma_{n,m}^2$. Note that

$$a_n = sup_{1\leq m\leq n} E(X_{n,m}^2) \leq \epsilon^2 + \sum_{m=1}^{n} E\left(X_{n,m}^2 1_{|X_{n,m}|>\epsilon}\right).$$

Therefore $\limsup_n a_n \leq \epsilon^2$, but since $\epsilon > 0$ is arbitrary $a_n \to 0$.

Hence for any $t > 0$, there exists n_0 large enough so that for all $n \geq n_0$, $|1 - \frac{t^2 a_n}{2}| \leq 1$. Applying Lemma 3.4.3 (with $\theta = 1$, $n \geq n_0$) we obtain

$$\begin{aligned} &\left| \prod_{m=1}^{n} \phi_{n,m}(t) - \prod_{m=1}^{n} (1 - \frac{t^2\sigma_{n,m}^2}{2}) \right| \\ &\leq \sum_{m=1}^{n} |\phi_{n,m}(t) - (1 - \frac{t^2\sigma_{n,m}^2}{2})| \\ &= \sum_{m=1}^{n} \left| E\left(e^{itX_{n,m}} - 1 - itX_{n,m} + \frac{t^2 X_{n,m}^2}{2} \right) \right| \\ &\stackrel{(a)}{\leq} \sum_{m=1}^{n} E\left(\min\left\{ \frac{|tX_{n,m}|^3}{6}, |tX_{n,m}|^2 \right\} \right) \\ &\leq \sum_{m=1}^{n} E\left(\frac{|t|\epsilon}{6} |tX_{n,m}|^2 1_{|X_{n,m}|\leq\epsilon} + |tX_{n,m}|^2 1_{|X_{n,m}|>\epsilon} \right). \end{aligned}$$

In the above (a) follows from Lemma 3.4.2. Hence

$$\limsup_n \left| \prod_{m=1}^n \phi_{n,m}(t) - \prod_{m=1}^n \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) \right| \le \epsilon \frac{|t|^3\sigma^2}{6},$$

however since $\epsilon > 0$ is arbitrary, the limit is zero.

To complete the proof, it suffices to show that

$$\limsup_n \left| \prod_{m=1}^n \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) - \prod_{m=1}^n e^{-t^2\sigma_{n,m}^2/2} \right| = 0.$$

Again applying Lemma 3.4.3 (with $\theta = 1$, $n \ge n_0$) we obtain

$$\begin{aligned} &\left| \prod_{m=1}^n \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) - \prod_{m=1}^n e^{-t^2\sigma_{n,m}^2/2} \right| \\ &\quad \le \sum_{m=1}^n \left| e^{-t^2\sigma_{n,m}^2/2} - \left(1 - \frac{t^2\sigma_{n,m}^2}{2}\right) \right| \\ &\quad \overset{(a)}{\le} \sum_{m=1}^n \frac{t^4\sigma_{n,m}^4}{2} \le \frac{t^4 a_n}{2} \sum_{m=1}^n \sigma_{n,m}^2. \end{aligned}$$

In the above (a) follows from Lemma 3.4.1. Since $a_n \to 0$ and $\sum_{m=1}^n \sigma_{n,m}^2 \to \sigma^2$, we are done.

□

Lemma 3.4.7 (Borel-Cantelli 2). *If an infinite sequence of mutually independent events A_i satisfy $\sum_i P(A_i) = \infty$, then $P(A_n \text{ i.o.}) = 1$.*

Proof. Define

$$B_k = \cup_{m \ge k} A_m.$$

Then, by independence,

$$P(B_k) = 1 - \prod_{m \ge k}(1 - P(A_m)) \ge 1 - \prod_{m \ge k} e^{-P(A_m)} = 1.$$

where the inequality follows since $(1-x) \le e^{-x}, 0 \le x \le 1$. Now $B_k \downarrow A_n$ *i.o.* and we are done. □

3.4.2. Convergence to a Poisson random variable. A Poisson(λ) random variable, Z, takes values in $\mathbb{N}$ and satisfies $P(Z = n) = \frac{\lambda^n e^{-\lambda}}{n!}$.

Theorem 3.4.8. *Let $X_{n,m}, 1 \le m \le n$ be independent non-negative integer valued random variables with $P(X_{n,m} = 1) = p_{n,m}$, and $P(X_{n,m} \ge 2) = \epsilon_{n,m}$. If*

(1) $\sum_{n=1}^m p_{n,m} \to \lambda \in (0, \infty)$
(2) $\max_{1 \le m \le n} p_{n,m} \to 0$, *and*
(3) $\sum_{m=1}^n \epsilon_{n,m} \to 0$,

then $S_n := \sum_{m=1}^n X_{n,m} \overset{w}{\Rightarrow} Z$, where Z is Poisson(λ).

Proof. Note that $E\left(e^{itS_n}\right) = \prod_{m=1}^n E\left(e^{itX_{n,m}}\right)$. From Lemma 3.4.3 (with $\theta = 1$) we have

$$\left| \prod_{m=1}^n E\left(e^{itX_{n,m}}\right) - \prod_{m=1}^n \left(1 - p_{n,m} + p_{n,m}e^{it}\right) \right)$$

$$\leq \sum_{m=1}^{n} \left|E\left(e^{itX_{n,m}}\right) - \left(1 - p_{n,m} + p_{n,m}e^{it}\right)\right|$$
$$= \sum_{m=1}^{n} \left|E\left(-\epsilon_{n,m} + E\left(e^{itX_{n,m}} 1_{X_{n,m}\geq 2}\right)\right)\right| \leq 2\sum_{m=1}^{n} \epsilon_{n,m} \to 0.$$

From Lemma 3.4.1 we know that as long as $p_{n,m} \leq \frac{1}{2}$

$$\left|e^{-p_{n,m}(1-e^{it})} - \left(1 - p_{n,m} + p_{n,m}e^{it}\right)\right| \leq 4p_{n,m}^2.$$

Again, from Lemma 3.4.3 (with $\theta = 1$) we have

$$\left|\prod_{m=1}^{n} e^{-p_{n,m}(1-e^{it})} - \prod_{m=1}^{n}\left(1 - p_{n,m} + p_{n,m}e^{it}\right)\right)$$
$$\leq \sum_{m=1}^{n} \left|e^{-p_{n,m}(1-e^{it})} - \left(1 - p_{n,m} + p_{n,m}e^{it}\right)\right|$$
$$\leq 4\sum_{m=1}^{n} p_{n,m}^2 \leq 4\left(\max_{1\leq m\leq n} p_{n,m}\right)\sum_{m=1}^{n} p_{n,m} \to 0.$$

Finally as

$$\left|e^{-\sum_{m=1}^{n} p_{n,m}(1-e^{it})} - e^{-\lambda(1-e^{it})}\right| \to 0,$$

we are done. $\square$

CHAPTER 4

Signed measures and Conditional Expectation

4.1. Signed Measure

Definition 4.1.1. Let $(\Omega, \mathcal{F})$ be a measurable space. A *finite signed-measure* is a mapping of $\mathcal{F}$ to $\mathbb{R}$ satisfying: if $\{A_i\}$ are pairwise disjoint; then $\lambda(\cup_i A_i) = \sum_i \lambda(A_i)$.

Note that $\lambda(\emptyset) = 0$.

Lemma 4.1.1. *Let λ be a finite signed measure, then $\sup_{A \in \mathcal{F}} |\lambda(A)| < \infty$.*

Proof. For any set A, let $\lambda_+(A) := \sup_{B \subseteq A} |\lambda(B)|$. If there exists any set A such that $\lambda_+(A)$ and $\lambda_+(A^c)$ are both finite, we are done. This is because, for any $B \in \mathcal{F}$,

$$|\lambda(B)| = |\lambda(B \cap A) + \lambda(B \cap A^c)| \leq \lambda_+(A) + \lambda_+(A^c).$$

Assume otherwise, i.e. for every set A, at least one of $\lambda_+(A)$ or $\lambda_+(A^c)$ is infinite. Wl.o.g. A such that $\lambda_+(A) = \infty$. Let $A_1 \subset A$ such that $|\lambda(A_1)| > n(1 + |\lambda(A)|)$ and $\lambda_+(A_1) = \infty$. Such A_1 exists because: for any $B \subseteq A$, if $|\lambda(B)| \geq k(1 + |\lambda(A)|)$, then

$$|\lambda(A \setminus B)| \geq |\lambda(B)| - |\lambda(A)| \geq k + (k-1)|\lambda(A)| \geq (k-1)(1 + |\lambda(A)|).$$

Further $\lambda_+(A) \leq \lambda_+(B) + \lambda_+(A \setminus B)$.

Repeat. Take $A_2 \subset A_1$ such that $|\lambda(A_2)| > n(1 + |\lambda(A_1)|) > n^2(1 + |\lambda(A)|)$ and $\lambda_+(A_2) = \infty$.

Therefore by induction, we have a decreasing sequence of sets $A_n \downarrow A_*$. Hence by countable additivity $\lambda(A_n) \to \lambda(A_*)$; however $|\lambda(A_n)| \to \infty$, contradicting finiteness of $\lambda(A_*)$. $\square$

Definition 4.1.2. A set A is called *totally positive*, if for every $B \subseteq A$ we have $\lambda(B) \geq 0$.

Lemma 4.1.2. *If $\lambda(A) = l \geq 0$, then these exists a totally positive set $A_+ \subseteq A$ such that $\lambda(A_+) \geq l$.*

Proof. Let $m = \inf_{B \subseteq A} \lambda(B)$. w.l.o.g. $m < 0$ (else $m = 0$ and set $A_+ = A$). Let B_1 satisfy $\lambda(B_1) < \frac{m}{2}$. Set $A_1 = A \setminus B_1$. Clearly $\lambda(A_1) \geq l$ and $m_1 = \inf_{B \subseteq A_1} \lambda(B) > \frac{m}{2}$.

repeat. Take $B_2 \subseteq A_1$ such that $\lambda(B_1) < \frac{m_1}{2}$, and set $A_2 = A_1 \setminus B_2$. Note that $m_2 = \inf_{B \subseteq A_2} \lambda(B) > \frac{m_1}{2} > \frac{m}{4}$. Further $\lambda(A_2) \geq l$.

Proceeding similarly we have that $A_n \downarrow A_*$ such that $m_k = \inf_{B \subseteq A_k} \lambda(B) > \frac{m}{2^k}$ $\forall k$ and $\lambda(A_k) \geq l$. Hence $\lambda(A_*) \geq l$ and $m_* = \inf_{B \subseteq A_*} \lambda(B) \geq 0$. This A_* is the required totally positive subset of A. $\square$

Note that the class of totally positive sets is closed under countable unions. Further any subset of a totally positive set is also totally positive.

Lemma 4.1.3. *There is a partition of Ω into a totally positive set Ω_+ and a totally negative set Ω_-.*

Proof. Let $m = \lambda_p(\Omega) := \sup_{A \subseteq \Omega} \lambda(A)$. Assume $m > 0$ (else set $\Omega_+ = \emptyset$). Then pick a totally positive A_1 such that $\lambda(A_1) > \frac{m}{2}$, and set $B_1 = \Omega \backslash A_1$. It is clear that $m_1 = \lambda_p(B_1) \leq \frac{m}{2}$. Now pick a totally positive $A_2 \subset B_1$ such that $\lambda(A_2) \geq \frac{m_1}{2}$, and set $B_2 = B_1 \setminus A_2 = \Omega \setminus (A_1 \cup A_2)$. Note that $m_2 = \lambda_p(B_2) \leq \frac{m_1}{2} \leq \frac{m}{4}$.

Define $\Omega_+ = \cup_i A_i$, and let $B_i \downarrow \Omega_-$. Note that Ω_+ and Ω_- partition Ω and Ω_+ is a totally positive set and Ω_- is a totally negative set. □

Remark: Note that this partitioning is not necessarily unique; however any two partitions differ by a set that is both totally positive and totally negative (hence it and all its subsets have measure 0). Let us call the sets that are totally positive and totally negative to be totally zero sets.

Therefore every finite signed measure λ induces two non-negative finite measures defined by: $\mu_+(A) = \lambda(A \cap \Omega_+)$, and $\mu_-(A) = -\lambda(A \cap \Omega_-)$. Further $\lambda(A) = \mu_+(A) - \mu_-(A)$.

Let $f(\omega)$ be an integrable function. Then define

$$\lambda(A) := \int f 1_A d\mu.$$

Observe that λ is a signed measure. (Countable additivity follows from dominated convergence theorem.) Further if $\mu(A) = 0$ then $\lambda(A) = 0$. To see this, let $\mu(A) = 0$ and note that

$$0 = n\mu(A) \geq \int (|f| \wedge n) 1_A d\mu \uparrow \int |f| 1_A d\mu \geq |\lambda(A)|.$$

Definition 4.1.3. Let $(\Omega, \mathcal{F})$ be a measurable space and let λ be a finite signed-measure on this space and μ be a non-negative measure on this space. Then λ is said to be *absolutely continuous* with respect to μ, denoted by $\lambda \ll \mu$, if $\mu(A) = 0$ implies $\lambda(A) = 0$.

Theorem 4.1.4 (Radon-Nikodym). *Let λ be a finite-signed-measure on $(\Omega, \mathcal{F})$ and μ be a non-negative measure on $(\Omega, \mathcal{G}), \mathcal{G} \supseteq \mathcal{F}$, such that $\mu(\Omega) < \infty$. If $\lambda \ll \mu$, then there exists a integrable function f, measurable w.r.t. $\mathcal{F}$ such that*

$$\lambda(A) = \int f 1_A d\mu, \quad \forall A \in \mathcal{F}.$$

Proof. For every $q \in \mathbb{Q}$, define the finite-signed-measure on $(\Omega, \mathcal{F})$

$$\lambda_q(A) = \lambda(A) - q\mu(A).$$

Let Ω_+^q be a totally positive partition of Ω induced by λ_q. Further by suitably discarding totally zero sets, we can have Ω_+^q to be a decreasing sequence of sets in q. (Argue why?). Once we have this collection, define

$$f(\omega) = \sup\{q : \omega \in \Omega_+^q\}.$$

To show the measurability of f, note that

$$\{\omega : f(\omega) > x\} = \cup_{q > x} \Omega_+^q.$$

Nest step is to show finiteness of $f(\omega)$ almost everywhere. Let $A = \cap_q \Omega_+^q$. Since $\lambda_q(A) \geq 0 \ \forall q$, we have $\lambda(A) \geq q\mu(A) \ \forall q$, which can happen only if $\mu(A) = 0$ (by finiteness of λ), and by absolute continuity $\lambda(A) = 0$ as well. Thus $A_\infty := \{\omega : f(\omega) = \infty\}$, satisfies $\mu(A_\infty) = 0$ and $\lambda(A_\infty) = 0$.

Suppose there is a set A such that $A \cap \Omega_q^+ = \emptyset$ for all $q \in Q$. This implies that $A \subseteq (\cup_q \Omega_+^q)^c$ which is essentially same as $\cap_q \Omega_q^-$. Hence $\lambda(A) - q\mu(A) \leq 0$ for all q. This can only happen again only if $\mu(A) = 0$ and hence $\lambda(A) = 0$. Thus $A_{-\infty} := \{\omega : f(\omega) = -\infty\}$, satisfies $\mu(A_{-\infty}) = 0$ and $\lambda(A_{-\infty}) = 0$. Thus $f(\omega)$ is finite on a set of measure 1.

For any two real number $a < b$, consider the set

$$I_{(a,b]} \subseteq \{\omega : f(\omega) \in (a, b]\}.$$

Therefore $I_{(a,b]} \subseteq \Omega_a^+ \ \forall q \leq a$ and $I_{[a,b]} \subseteq (\Omega_q)^- \ \forall q > b$. This implies that $\lambda(I_{[a,b]}) - q\mu(I_{[a,b]}) \geq 0 \ \forall q \leq a$ and $\lambda(I_{[a,b]}) - q\mu(I_{[a,b]}) \leq 0 \ \forall q \geq b$. Hence $a\mu(I_{(a,b]}) \leq \lambda(I_{(a,b]}) \leq b\mu(I_{(a,b]})$.

Lat $h > 0$, consider a grid and set

$$I_n = \{\omega : nh < f(\omega) \leq (n+1)h\}, \ -\infty \leq n \leq \infty.$$

Now, note that for all $A \in \mathcal{F}$ and every n

$$\begin{aligned} \lambda(A \cap I_n) - h\mu(A \cap I_n) \leq nh\mu(A \cap I_n) &\leq \int_{A \cap I_n} f d\mu \\ \leq (n+1)h\mu(A \cap I_n) &\leq \lambda(A \cap I_n) + h\mu(A \cap I_n). \end{aligned} \tag{4.1}$$

Take $A_+ = \{\omega : f(\omega) > 0\}$.

Summing over n (and using countable additivity and monotone convergence theorem) we obtain

$$\lambda(A_+) - h\mu(A_+) \leq \int_{A_+} f d\mu = \int_\Omega f_+ d\mu \leq \lambda(A_+) + h\mu(A_+).$$

This shows that $\int f_+ < \infty$. Similarly $\int f_- < \infty$; thus f is integrable.

Now take a generic A and sum (4.1) over n and (and using countable additivity and dominated convergence theorem) we get

$$\lambda(A) - h\mu(A) \leq \int_A f d\mu \leq \lambda(A) + h\mu(A).$$

Taking $h \to 0$ completes the proof. □

Remark 4.1.1. Note that the Radon-Nikodym derivative is essentially unique. To see this, suppose f, d were two derivatives, let $A_\epsilon = \{\omega : f(\omega) - g(\omega) \geq \epsilon\}$. Then since $\mu(A_\epsilon) = \int_{A_\epsilon} f d\mu = \int_{A_\epsilon} g d\mu$, which on the other hand $\int_{A_\epsilon} (f - g) d\mu \geq \epsilon\mu(A_\epsilon)$ we must have $\mu(A_\epsilon) = 0$. This shows that the functions match almost surely.

4.2. Conditional Expectation

Proposition 4.2.1. *Let f be an integrable function defined on $(\Omega, \mathcal{F}, \mu)$. Let $\mathcal{G} \subset \mathcal{F}$. Then there is a $\mathcal{G}$-measurable and integrable function g such that*

$$\int_A f d\mu = \int_A g d\mu \quad \forall g \in \mathcal{G}.$$

Further if g_1 and g_2 are two such functions, then $g_1 = g_2$ almost surely.

PROOF. Define a finite-signed-measure on $(\Omega, \mathcal{G})$ by

$$\lambda(A) = \int_A f d\mu \quad \forall A \in \mathcal{G}.$$

Note that $\lambda \ll \mu$. Hence by Radon-Nikodym theorem, there exists an integrable function $\hat{g} = \frac{d\lambda}{d\mu}$ that is $\mathcal{G}$-measurable such that

$$\lambda(A) = \int_A \hat{g} d\mu \quad \forall A \in \mathcal{G}.$$

Setting $g = \hat{g}$ completes the first part.

Now let $A_\epsilon = \{\omega : g_1(\omega) - g_2(\omega) > \epsilon\}$. Since A_ϵ is $\mathcal{G}$ measurable for every $\epsilon > 0$, we must have

$$\int_{A_\epsilon} g_1 d\mu = \int_{A_\epsilon} g_2 d\mu \quad \implies \mu(A_\epsilon) = 0.$$

Now $A_+ = \{\omega : g_1(\omega) - g_2(\omega) > 0\} = \cup_n A_{1/n}$ and hence $\mu(A_+) = 0$. Similarly, by symmetry, we see that $\mu(A_-) = 0$ where $A_- = \{\omega : g_2(\omega) - g_1(\omega) > 0\} = \cup_n A'_{1/n}$. □

Thus we denote any $g(\omega)$ that satisfies the proposition as the *conditional expectation* $E(f|\mathcal{G})$.

THEOREM 4.2.2 (Properties of Conditional Expectation). *Let f be integrable and $(\Omega, \mathcal{F})$-measurable. Let $\mathcal{G} \subset \mathcal{F}$. The following properties hold:*

(1) $\mathrm{E}(f) = \mathrm{E}(\mathrm{E}(f|\mathcal{G}))$.
(2) $f \geq 0$ *implies that* $\mathrm{E}(f|\mathcal{G}) \geq 0$ *a.s.*
(3) If f_1, f_2 be integrable and $(\Omega, \mathcal{F})$-measurable then

$$\mathrm{E}(af_1 + bf_2|\mathcal{G}) = a\,\mathrm{E}(f_1|\mathcal{G}) + b\,\mathrm{E}(f_2|\mathcal{G}) \text{ a.s.}$$

(4) $\mathrm{E}(|f|) \geq \mathrm{E}(|\,\mathrm{E}(f|\mathcal{G})|)$
(5) If h is bounded and $\mathcal{G}$-measurable, then

$$\mathrm{E}(hf|\mathcal{G}) = h\,\mathrm{E}(f|\mathcal{G}) \text{ a.s.}$$

(6) If $\mathcal{G}_1 \subseteq \mathcal{G}$ then

$$\mathrm{E}(f|\mathcal{G}_1) = \mathrm{E}(\mathrm{E}(f|\mathcal{G})|\mathcal{G}_1) \text{ a.s.}$$

(7) (Jensen's inequality) If Φ is a convex function then

$$\mathrm{E}(\Phi(f)|\mathcal{G}) \geq \Phi(\mathrm{E}(f|\mathcal{G})) \text{ a.s.}$$

PROOF. The proofs are rather straightforward

(1) Take $A = \Omega$ and apply definition.
(2) For $\epsilon > 0$, let $A_\epsilon^- = \{\omega : \mathrm{E}(f|\mathcal{G}) \leq -\epsilon\}$. Then argue that $\mu(A_\epsilon^-) = 0$. Take $\epsilon_n = \frac{1}{n} \downarrow 0$.
(3) Follows from definition and linearity of expectation.
(4) Let $A_+ = \{\omega : \mathrm{E}(f|\mathcal{G}) \geq 0\}$. Since $A_+ \in \mathcal{G}$, by definition,

$$\int_{A_+} \mathrm{E}(f|\mathcal{G}) d\mu = \int_{A_+} f d\mu \leq \int_{A_+} |f| d\mu.$$

Similarly we can consider $A_- = \{\omega : \mathrm{E}(f|\mathcal{G}) < 0\}$. Again

$$-\int_{A_-} \mathrm{E}(f|\mathcal{G}) d\mu = -\int_{A_-} f d\mu \leq \int_{A_-} |f| d\mu.$$

Adding them yields the desired result.

(5) If $h = 1_A$ for $A \in \mathcal{G}$, this is immediate from definition. Linearity extends it to simple functions. Let h_n be simple and $h_n \to h$ uniformly. Let $g_n = \mathrm{E}(h_n f|\mathcal{G})$ and $g = \mathrm{E}(hf|\mathcal{G})$. Note that from 4)

$$\int |g_n - g| dP \leq \int (|h_n - h||f|) dP \to 0.$$

Therefore for any fixed $A \in \mathcal{G}$

$$\int_A g dP \leftarrow \int_A g_n dP = \int_A h_n f dP \to \int_A h f dP.$$

(6) This follows from definitions easily. (Check!)

(7) The key is again to write a convex function as the pointwise supremum of a set of affine functions. For each such affine function we have

$$\Phi(f) \geq a_\alpha f + b_\alpha \quad \implies \mathrm{E}(\Phi(f)|\mathcal{G}) \geq a_\alpha \mathrm{E}(f|\mathcal{G}) + h_\alpha.$$

Now taking supremum over the class of affine functions that yields Φ implies the result.

$\square$

4.2.1. Conditional Probability. The goal of this section is to define a *conditional probability* distribution which is defined below.

DEFINITION 4.2.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$. A mapping $\mu : \Omega \times \mathcal{F} \to [0, 1]$ is called a *regular conditional probability* if it satisfies the following:

(1) For every $A \in \mathcal{F}$, $\mu(\omega, A) = \mathrm{E}(1_A|\mathcal{G})$
(2) For almost every ω, $\mu(\omega, A)$ is a probability measure on $(\Omega, \mathcal{F})$.

It turns out that these regular conditional probabilities need not always exist. However, if the space is "nice", then they do exist. When is a space "nice": If Ω is a Polish space, $\mathcal{F}$ is its Borel σ-algebra, and $\mathcal{G}$ is a countably generated sub σ-algebra of $\mathcal{F}$, then the space is "nice" enough.

Clearly, an initial approach would be to define for every $A \in \mathcal{F}$, $\mu(\omega, A) := \mathrm{E}(1_A|\mathcal{G})$. This would meet condition (1) above, but it is not necessary that for every ω, and for $A \subset B \in \mathcal{F}$ we would have $\mu(\omega, A) \leq \mu(\omega, B)$. Of course, we can throw away a set of measure zero and make the previous inequality hold. However of there are uncountable such pairs of sets, then this would cause issues. On the other hand, if the space is nice and if the σ-algebra is countably generated then we can try to avoid the above issue. We will work with $\Omega = [0, 1]$ and $\mathcal{F}$ to be the Borel σ-algebra.

We consider the collection $A_q = (-\infty, q]$ for $q \in Q$, and we define

$$\mu(\omega, A_q) = \mathrm{E}(1_{A_q}|\mathcal{G}) \quad q \in Q.$$

Since this collection is countable, we know that we can obtain a set S with $P(S) = 1$ such that for all $\omega \in S$, $\mu(\omega, A_q)$ is non-decreasing in q, and $\mu(\omega, A_q) = 1$ for all $q \geq 1$ and $\mu(\omega, A_q) = 0$ for all $q < 0$.

Consider the set $A_y = (-\infty, y], y \in \mathbb{R}$ and define

$$\mu(\omega, A_y) = \inf_{q > y} \mu(\omega, A_q).$$

Clearly, the above $F_\omega(y) := \mu(\omega, A_y)$ is a distribution function for every $\omega \in S$. Since this is in natural 1-1 correspondence with a probability measure (see Lebesgue's theorem) on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ we set $\mu(\omega, A)$ to be that probability measure. Hence we are left to show that this choice also satisfies the first condition.

Observe that, by construction, if $A = (-\infty, q], q \in Q$ then the first condition is satisfied. Consider the interval $A_y = (-\infty, y]$ for y in reals. Then for any $q > y$ we have

$$\int 1_{A_q} dP = \int \mu(\omega, A_q) dP.$$

Let $q \downarrow y$, monotone convergence theorem (on both sides) implies that

$$\int 1_{A_y} dP = \int \mu(\omega, A_y) dP.$$

Hence from linearity of conditional expectation, equality holds on the algebra comprised of finite disjoint unions of intervals of the form $(a, b]$. Now consider the collection, $\mathcal{S}$, of all sets for which

$$\int 1_A dP = \int \mu(\omega, A) dP.$$

This is a monotone class, since the limit on the left hand side is argued using monotone convergence theorem. On the left hand side, for every $\omega \in S$ the limit exists and now dominated convergence theorem implies that $\mathcal{S}$ is a monotone class. Hence $\mathcal{S}$ contained the σ-filed generated by the algebra of finite disjoint union of left-open right-closed intervals, completing the proof.

CHAPTER 5

Martingales

5.1. Martingales

In this section we will study an important class of sequences of random variables that arise naturally in many settings, and unnaturally in other settings as a tool to prove certain results.

Definition 5.1.1. Given a measurable space $(\Omega, \mathcal{F})$ a *filtration* is an increasing sequence of σ-algebra's $\mathcal{F}_i \subset \mathcal{F}$, i.e. if $i \leq j$ then $\mathcal{F}_i \subseteq \mathcal{F}_j$.

Definition 5.1.2. A sequence of random variables $\{X_i\}$ is said to be *adapted* to the filtration $\{\mathcal{F}_i\}$ of sigma-algebra's if X_i is $\mathcal{F}_i$-measurable for every i.

Definition 5.1.3. A sequence $\{X_i\}, i \geq 1$ of integrable random variables, adapted to a filtration $\{\mathcal{F}_i\}, i \geq 1$, is said to be a *martingale* if, for all $i \geq 1$,

$$\mathrm{E}(X_{i+1}|\mathcal{F}_i) = X_i \quad a.s.$$

A sequence $\{X_i\}, i \geq 1$ of integrable random variables, adapted to a filtration $\{\mathcal{F}_i\}, i \geq 1$, is said to be a *sub-martingale* if, for all $i \geq 1$,

$$\mathrm{E}(X_{i+1}|\mathcal{F}_i) \geq X_i \quad a.s.$$

A sequence $\{X_i\}, i \geq 1$ of integrable random variables, adapted to a filtration $\{\mathcal{F}_i\}, i \geq 1$, is said to be a *super-martingale* if, for all $i \geq 1$,

$$\mathrm{E}(X_{i+1}|\mathcal{F}_i) \leq X_i \quad a.s.$$

Remark 5.1.1. Usually, given a sequence of random variables $\{X_n\}$, we can define a natural filtration according to $\mathcal{F}_n = \sigma(X_1, .., X_n)$. Hence, unless otherwise specified, this will be the underlying filtration.

The following statements are easy to verify:

(1) If $\{X_i\}$ is a sequence of independent integrable random variables. Define $S_n = X_1 + \cdots + X_n$. Then $\{\alpha(S_n - \mathrm{E}(S_n)) + \beta\}$ is a martingale.
(2) If $\{X_i\}$ is a martingale and Φ is a concave function such that $\Phi(X_i)$ is integrable, then $\{\Phi(X_i)\}$ is a super-martingale.

    Proof. From Jensen's inequality $\mathrm{E}(\Phi(X_{i+1})|\mathcal{F}_i) \leq \Phi(\mathrm{E}(X_{i+1}|\mathcal{F}_i)) = \Phi(X_i)$ *a.s.*. $\square$

(3) Similarly, if $\{X_i\}$ is a martingale and Φ is a non-decreasing convex function such that $\Phi(X_i)$ is integrable, then $\{\Phi(X_i)\}$ is a sub-martingale.
(4) Given a filtration $\mathcal{F}_i$, if X is integrable then $X_i := \mathrm{E}(X|\mathcal{F}_i)$ is a martingale sequence.
(5) If $\{X_i\}$ is a martingale, then for $i \leq j$, $\mathrm{E}(X_j|\mathcal{F}_i) = X_i$ *a.s.*. The equality becomes the appropriate inequality for *sub*-martingales and *super*-martingales.

Definition 5.1.4. Given a filtration $\{\mathcal{F}_i\}$, a sequence $\{X_i\}, i \geq 1$ of integrable random variables is said to be a *predictable* if, for all $i \geq 1$, X_{i+1} is $\mathcal{F}_i$-measurable.

Theorem 5.1.1 (Doob's inequality). *Suppose X_n is a martingale (or a non-negative sub-martingale) sequence of length n. Define*

$$A_l := \{\omega : \sup_{1 \leq j \leq n} |X_j(\omega)| \geq l\}.$$

Then

$$P(A_l) \leq \frac{1}{l} \int_{A_l} |X_n| dP \leq \frac{1}{l} \mathrm{E}(|X_n|).$$

Proof. First note that $\{|X_n|\}$ is a sub-martingale. Then we partition A_n as follows. Let

$$B_j = \{\omega : |X_1(\omega)| < l, \ldots, |X_{j-1}(\omega)| < l, |X_j(\omega)| \geq l\},\ 1 \leq j \leq n.$$

Note that $B_j \in \mathcal{F}_j$. Further since $\{|X_n|\}$ is a sub-martingale, we have that $\mathrm{E}(|X_n||\mathcal{F}_j) \geq |X_j|$ *a.s.*, implying that

$$\int_{B_j} |X_n| dP = \int_{B_j} \mathrm{E}(|X_n||\mathcal{F}_j) dP \geq \int_{B_j} |X_j| dP \geq l P(B_j).$$

Observing that $A = \sqcup_j B_j$ completes the proof. □

Lemma 5.1.2. *If X, Y are two non-negative random variables on the same probability space such that*

$$\mathrm{P}(Y \geq l) \leq \frac{1}{l} \int_{Y \geq l} X dP$$

then for every $p > 1$

$$\int Y^p dP \leq \left(\frac{p}{p-1}\right)^p \int X^p dP.$$

Proof. Let $F(x) = \mathrm{P}(Y \leq x)$ be the distribution function. Let $T(x) = 1 - F(x) = P(Y > x)$. Then

$$\begin{aligned}
\int Y^p dP &= \int_0^\infty y^p dF(y) \\
&= -\int_0^\infty y^p dT(y) \\
&= p \int_0^\infty y^{p-1} T(y) dy && \text{(integration by parts)} \\
&\leq p \int_0^\infty y^{p-2} \left(\int_{Y>y} X dP\right) dy \\
&= p \int X \left(\int_0^Y y^{p-2} dy\right) dP \\
&= \frac{p}{p-1} \int X Y^{p-1} dP \\
&\leq \frac{p}{p-1} \left(\int X^p dP\right)^{\frac{1}{p}} \left(\int Y^p dP\right)^{\frac{p-1}{p}} && \text{(Hölder).}
\end{aligned}$$

If $\mathrm{E}(Y^p) < \infty$, we are done. Otherwise, obtain the result for $Y_m = \min\{Y, m\}$ and pass to the limit to get the desired inequality. □

Corollary 5.1.3 (Doob)**.** *Let X_n is a sub-martingale sequence of length n. Define $S = \sup_{1 \leq j \leq n} |X_j(\omega)|$. Then*

$$\int S^p dP \leq \left(\frac{p}{p-1}\right)^p \int |X_n|^p dP.$$

Definition 5.1.5 (Martingale Transform). Given a *predictable* process C_n and a martingale sequence $\{X_n\}$, the martingale transform is given by

$$(C \circ X)_n = \sum_{m=1}^{n} C_m (X_m - X_{m-1}).$$

Remark 5.1.2. This is the discrete analogue of a stochastic integral. Note that if C_m is bounded, then $(C \circ X)_n$ is a Martingale(why?).

Definition 5.1.6 (Stopping Time). A mapping $T : \Omega \mapsto \{0, 1, 2, 3, ..., \infty\}$ is called a *stopping time* with respect to a filtration $\{\mathcal{F}_n\}_{n \geq 0}$ if

$$\{\omega : T(\omega) \leq n\} \in \mathcal{F}_n, \quad \forall n.$$

Theorem 5.1.4. *If X_n is a super-martingale sequence and T is a stopping time, then the stopped process*

$$X_n^T(\omega) := X_{T(\omega) \wedge n}(\omega)$$

is a super-martingale. Consequently,

$$\mathrm{E}\left(X_{T \wedge n}\right) \leq \mathrm{E}(X_0).$$

If X_n is a martingale, then $\{X_n^T\}$ is a martingale and further equality holds above.

Proof. Observe that X_n^T is $\mathcal{F}_n$ measurable from the definitions. Now

$$\begin{aligned}
\mathrm{E}(X_{T \wedge (n+1)} | \mathcal{F}_n) &= \mathrm{E}(X_{T \wedge (n+1)} 1_{T \leq n} + X_{T \wedge (n+1)} 1_{T > n} | \mathcal{F}_n) \\
&= \mathrm{E}(X_{T \wedge n} 1_{T \leq n} + X_{n+1} 1_{T > n} | \mathcal{F}_n) \\
&= X_{T \wedge n} 1_{T \leq n} + 1_{T > n} \mathrm{E}(X_{n+1} | \mathcal{F}_n) \ a.s. \\
&\leq X_{T \wedge n} 1_{T \leq n} + 1_{T > n} X_n \ a.s. \\
&= X_{T \wedge n}.
\end{aligned}$$

The proof for the martingale case follows with the inequality in the above step replaced with equality (from definition). □

Theorem 5.1.5 (Doob's Optional Stopping Time Theorem). *The following results hold:*

1. *Let X_n be a supermartingale and T is a stopping time. If any of the following conditions is satisfied:*
   - *(a) T is bounded*
   - *(b) X_n is uniformly bounded and T is finite almost surely*
   - *(c) $E(T) < \infty$ and $|X_n - X_{n-1}|$ is uniformly bounded (say by K)*

   then $E(X_T) \leq E(X_0)$.
2. *If X_n is a martingale and any of the above three conditions hold, then $\mathrm{E}(X_T) = \mathrm{E}(X_0)$.*
3. *Let $\{C_n\}$ be a bounded predictable sequence and $\{X_n\}$ be a martingale such that $|X_n - X_{n-1}|$ is uniformly bounded, and T is a stopping time such that $\mathrm{E}(T) < \infty$. Then $\mathrm{E}\left((C \circ X)_T\right) = 0$.*

(4) If X_n is a non-negative super-martingale and T is almost surely finite, then $\mathrm{E}(X_T) \leq \mathrm{E}(X_0)$.

Proof. We will establish each part in sequence.

(1) Here X_n is a supermartingale.
   (a) Let $T \leq N$, then $X_{T\wedge N} = X_T$. The inequality follows from the previous theorem.
   (b) If T is finite almost surely, then $X_{T\wedge n} \to X_T$ almost surely. Since X_n is uniformly bounded (hence $X_{T\wedge n}, X_T$ are also bounded by the same constant), bounded convergence theorem implies that $E(X_{T\wedge n}) \to \mathrm{E}(X_T)$. Applying previous theorem completes the proof.
   (c) Note that

$$\begin{aligned}|X_{T\wedge n} - X_0| &= \left|\sum_{k=1}^{T\wedge n} X_k - X_{k-1}\right| \\ &\leq \sum_{k=1}^{T\wedge n} |X_k - X_{k-1} \leq K(T\wedge n) \leq KT.\end{aligned}$$

   Since K is a constant and T is integrable KT is integrable. Hence dominated convergence theorem says that

$$\mathrm{E}(X_T - X_0) = \mathrm{E}(\lim_n (X_{T\wedge n} - X_0)) = \lim_n \mathrm{E}(X_{T\wedge n} - X_0) \leq 0.$$

(2) Apply the earlier part to X_n and $-X_n$ to get the desired equality.
(3) We know that $(C \circ X)_n$ is a zero-mean martingale. Further the given conditions imply that $|(C\circ X)_n - (C\circ X)_{n-1}|$ in uniformly bounded. Hence, from previous parts we are done.
(4) This is a direct consequence of Fatou's Lemma and that $\mathrm{E}(X_{T\wedge n}) \leq \mathrm{E}(X_0)$.

□

Lemma 5.1.6. *If T is a stopping time and N be a number such that $P(T \leq n + N|\mathcal{F}_n) > \epsilon$ a.s. then $E(T) < \infty$.*

Proof. Observe that

$$\begin{aligned}P(kN < T) &= \mathrm{E}(1_{T>(k-1)N} 1_{T>(k-1)N+N}) \\ &= \mathrm{E}(\mathrm{E}(1_{T>(k-1)N} 1_{T>(k-1)N+N}|\mathcal{F}_{(k-1)N})) \\ &= \mathrm{E}(1_{T>(k-1)N}\,\mathrm{E}(1_{T>(k-1)N+N}|\mathcal{F}_{(k-1)N}) \\ &\leq (1-\epsilon)\,\mathrm{E}(1_{T>(k-1)N}) = (1-\epsilon)P(T > (k-1)N).\end{aligned}$$

Hence by induction $P(T > kN) \leq (1-\epsilon)^k$.

Now note that

$$\begin{aligned}\mathrm{E}(T) &= \sum_{n\geq 0} \mathrm{P}(T > n) = \sum_{l\geq 0}\sum_{m=0}^{N-1} \mathrm{P}(T > lN + m) \\ &\leq N \sum_{l\geq 0} \mathrm{P}(T > lN) \leq N \sum_{l\geq 0} (1-\epsilon)^l = \frac{N}{\epsilon} < \infty.\end{aligned}$$

□

Example 5.1.7. Consider a Monkey typing characters on a 26-key keyboard. Assume that the characters typed are i.i.d. and each character is uniformly chosen. What is the expected time before the Monkey types "WOWOWOW"?

Solution. The expected time is $26^7 + 26^5 + 26^3 + 26$. A beautiful argument using Martingales for such problems was developed by Li(1980).

Let $X_n, n \geq 1$ denote the characters typed by the monkey. Define a stopping time according to

$$T = \inf\{n : (X_{n-6}, .., X_n) = (W, O, W, O, W, O, W)\}.$$

Clearly this stopping time satisfies the previous lemma with $N = 7$ and $\epsilon = (26)^{-7}$. Hence $\mathrm{E}(T) < \infty$. For every $m \geq 1$ define the process $Y_n^{(m)}, n \geq m-1$ according to the following: $Y_{n-1}^{(m)} = 1$, $Y_{m+l-1}^{(m)} = 26^{l \wedge 7}$ for $l \geq 1$ if $(X_m, .., X_{m+(l \wedge 7)-1})$ matches the first $(l \wedge 7)$ characters of "WOWOWOW", else set $Y_n^{(m)} = 0$. Clearly $Y_n^{(m)}$ is $\mathcal{F}_n$ measurable. Further $\mathrm{E}(Y_{n+1}^{(m)} | \mathcal{F}_{n+1}) = Y_n^{(m)}$ for all $n \geq m-1$. Further if $n \geq m+6$ then $Y_{n+1}^{(m)} = Y_n^{(m)}$, and $Y_n^{(m)} \leq (26)^7$.

Observe that $M_n = \sum_{m=1}^{n+1} Y_n^{(m)} - n - 1$ is a zero-mean martingale. Further observe that

$$\begin{aligned} M_n - M_{n-1} &= \sum_{m=1}^{n} Y_n^{(m)} - Y_{n-1}^{(m)} \\ &= \sum_{m=1 \vee (n-6)}^{n} Y_n^{(m)} - Y_{n-1}^{(m)}. \end{aligned}$$

Hence $|M_n - M_{n-1}| \leq 2 \times 7 \times (26)^7$ (can improve this easily). Therefore, we can apply part $c)$ of Doob's optional stopping time theorem and obtain

$$\mathrm{E}(M_T) = 0 \implies \mathrm{E}\left(\sum_{m=1}^{T+1} Y_T^{(m)} - T - 1\right) = 0.$$

From the definition of the stopping time and $Y_{n,m}$, we see that $Y_T^{(T-6)} = (26)^7, Y_T^{(T-4)} = (26)^5, Y_T^{(T-2)} = (26)^3, Y_T^{(T)} = 26, Y_T^{(T+1)} = 1$, and the rest are zeros. Therefore $E(T) = 26^7 + 26^5 + 26^3 + 26$. $\square$

The following elegant combinatorial lemma is often regarded as an example of the reflection principle.

Lemma 5.1.7. *Consider the set of sequences, $\mathcal{S}$, $s_n \in \mathbb{Z}$, $n \geq 0$ such that $|s_{n+1} - s_n| = 1$. Let k, l be arbitrary integers, and let $m \leq \min\{k-1, l\}$. Let $\mathcal{S}_{k,l,m}^n \subset \mathcal{S}$ be the set of sequences such that $s_0 = k$ and $s_n = l$, and there exists $q \in [1:n]$ such that $s_q = m$. Then*

$$|\mathcal{S}_{k,l,m}^n| = \begin{cases} 0 & (n-k+l) \equiv 1 (\mathrm{mod} 2) \\ 0 & |l-k| > n \\ \binom{n}{\frac{n+k+l-2m}{2}} & \textit{otherwise} \end{cases}$$

Proof. Define $\hat{\mathcal{S}}_{k,l,m}^n \subset \mathcal{S}$ be the set of sequences such that $s_0 = 2m - k$ and $s_n = l$. Since $2m - k < m \leq l$, every such sequence must have $q \in [1:n]$ such that $s_q = m$. Observe that we can construct a bijection between the sequences in

$\hat{\mathcal{S}}^n_{k,l,m}$ and $\mathcal{S}^n_{k,l,m}$, by reflecting the sequence about m until the first time q such that $s_q = m$. The result then follows immediately from the cardinality of $\hat{\mathcal{S}}^n_{k,l}$. □

Here is a simple example that demonstrates the following: X_n is a martingale with bounded increments, i.e $|X_n - X_{n-1}| \leq K$. T is a stopping time that is almost surely finite. Yet $\mathrm{E}(X_T) \neq \mathrm{E}(X_0)$.

Consider a symmetric random walk in 1-dimension with $X_0 = 0$. Let X_n denote the location of the random walk at time n. Let $T = \inf\{n : X_n = -1\}$. Then clearly, if T is finite almost surely, then $\mathrm{E}(X_T) = -1 \neq \mathrm{E}(X_0) = 0$. To show that T is finite almost surely, we use Lemma 5.1.7. Consider the event $\mathcal{T}_n := \{\omega : T(\omega) > n\}$ and consider the following collection of events, for $0 \leq k \leq n$:

$$\mathcal{E}^n_k = \{\omega : X_0(\omega) = 0, .., X_n(\omega) = k\}$$
$$\mathcal{H}^n_k = \{\omega : X_0(\omega) = 0, .., X_n(\omega) = k, \exists m \in [1:n] \ s.t. \ X_m(\omega) = -1\}$$

Observe that we can partition as $\mathcal{T}_n = \cup_{k=0}^n (\mathcal{E}^n_k \setminus \mathcal{H}^n_k)$. Therefore we get

$$\begin{aligned} P(T > 2n) &= \sum_{k=0}^n \mathrm{P}(\mathcal{E}^{2n}_{2k} \setminus \mathcal{H}^{2n}_{2k}) \\ &= \sum_{k=0}^n \frac{1}{2^{2n}} \binom{2n}{n+k} \left(\binom{2n}{n+k} - \binom{2n}{n+k+1} \right) \\ &= \frac{1}{2^{2n}} \binom{2n}{n} \\ &\leq \frac{1}{\sqrt{\pi n}}. \end{aligned}$$

Thus T is finite almost surely.

5.1.1. Convergence theorems.

Definition 5.1.8 (Upcrossing). Given a sequence of random variables $\{X_n\}$ and two numbers $a < b$, we define a non-negative integer-valued non-decreasing sequence of random variables $U_n[a,b]$ according to

$$U_n[a,b] = \max\{k : \exists\ 0 \leq s_1 < t_1 < \cdots < s_k < t_k \leq n,\ s/t\ X_{s_i} \leq a, X_{t_i} > b\}.$$

If $\{X_n\}$ is adapted to the filtration $\{\mathcal{F}_n\}$ then $\{U_n[a,b]\}$ is also adapted to the filtration $\{\mathcal{F}_n\}$.

Lemma 5.1.8 (Doob's Upcrossing inequality)**.** *The following hold:*

(1) Let $\{X_n\}$ be a supermartingale. Then for any $n \geq 1$ and $a < b$

$$(b-a)\,\mathrm{E}(U_n[a,b]) \leq \mathrm{E}((X_n - a)_-) \leq \mathrm{E}(|X_n|) + |a|.$$

(2) Let $\{X_n\}$ be a submartingale. Then for any $n \geq 1$ and $a < b$

$$(b-a)\,\mathrm{E}(U_n[a,b]) \leq \mathrm{E}((X_n - a)_+) - \mathrm{E}((X_0 - a)_+) \leq \mathrm{E}(|X_n|) + |a|.$$

Proof. (1) Define a predictable process inductively as follows

$$C_1 = 1_{X_0 \leq a}$$
$$C_n = 1_{C_{n-1}=1} 1_{X_{n-1} \leq b} + 1_{C_{n-1}=0} 1_{X_{n-1} \leq a}.$$

In words, C_n is a sequence that takes a value 1, starting from an instance the process goes below a, till the instance the process goes above b for the first time. Then C_n becomes zero, and turns to 1 only when the process goes below a again.

Define $Y_0 = 0$ and $Y_n = (C \circ X)_n, n \geq 1$ and note that this is a supermartingale. Clearly,

$$Y_n \geq (b-a)U_n[a,b] - \max\{0, a - X_n\}(=: (X_n - a)_-).$$

Taking expectations and noting that $\mathrm{E}(Y_n) \leq \mathrm{E}(Y_0) = 0$ we obtain the result.

(2) Let $Z_n = (X_n - a)_+ + a$. Observe that Z_n is a sub-martingale and further it has the same number of up-crossings as X_n. Define $Y_n = (C \circ Z)_n, n \geq 1$ as before. Clearly

$$Y_n \geq (b-a)U_n[a,b].$$

Define similarly $\tilde{Y}_n = ((1-C) \circ Z)_n, n \geq 1$. Clearly $\tilde{Y}_n$ is also a sub-martingale and $\mathrm{E}(\tilde{Y}_n) \geq 0$ (verify). Now $Y_n + \tilde{Y}_n = Z_n - Z_0$, hence $\mathrm{E}(Y_n) \leq \mathrm{E}(Z_n - Z_0) = \mathrm{E}((X_n - a)_+) - \mathrm{E}((X_0 - a)_+)$.

□

THEOREM 5.1.9 (Martingale Convergence Theorem). *Let X_n be a super(or sub)-martingale with $\sup_n \mathrm{E}(|X_n|) < \infty$, then X_n will converge almost surely to a finite limit.*

PROOF. Clearly we have

$$\begin{aligned}
&\{\omega : \liminf_n X_n(\omega) < \limsup_n X_n(\omega)\} \\
&\quad = \cup_{a,b \in Q, a<b}\{\omega : \liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega)\} \\
&\quad = \cup_{a,b \in Q, a<b}\{\omega : U_\infty[a,b] = \infty\}
\end{aligned}$$

From Doob's upcrossing inequality and monotone convergence theorem we have

$$(b-a)\,\mathrm{E}(U_\infty[a,b]) \leq \sup_n \mathrm{E}(|X_n|) + |a| < \infty.$$

Hence for every $a, b \in Q, a < b$ we have $P(\{\omega : U_\infty[a,b] = \infty\}) = 0$. This implies that either X_n converges to a finite limit of $|X_n| \to \infty$.

We now only need to worry about $|X_n(\omega)| \to \infty$. Let $B_{M,k} = \{\omega : |X_n(\omega)| \geq M,\ \forall n \geq k\}$. From Markov's inequality, we have $\mathrm{P}(B_{M,k}) \leq \frac{\sup_n \mathrm{E}(X_n)}{M}$. Taking $M \to \infty$ we see that the limit of X_n is finite almost surely. □

Corollary 5.1.10. *If $\{X_n\}$ is a non-negative supermartingale, then $E(|X_n|) = \mathrm{E}(X_n) \leq \mathrm{E}(X_0) < \infty$. Hence it always converges.*

THEOREM 5.1.11. *Let $X_n, n \geq 0$ be a martingale with bounded increments, i.e. $|X_n - X_{n-1}| \leq K$ for all n, ω. Define*

$$\begin{aligned}
C &= \{\omega : \lim_n X_n(\omega) \text{ exists and is finite}\} \\
D &= \{\omega : \liminf_n X_n(\omega) = -\infty \text{ and } \limsup_n X_n(\omega) = \infty\}
\end{aligned}$$

Then $P(C \cup D) = 1$.

PROOF. Since $X_n - X_0$ is also a martingale with bounded increments, we assume w.l.o.g that $X_0 = 0$. For any $M > 0$, define the stopping time $T_M = \inf\{n : X_n < -M\}$. Then $X_{n \wedge T_M} + K + M$ is a martingale, and further $X_{n \wedge T_M} + M + K \geq 0$.

Applying Corollary 5.1.10 we see that $X_{n\wedge T_M} + K + M$ has a finite limit almost surely. This implies that on

$$\{\omega : T_M = \infty\},$$

the sequence X_n has a finite limit almost surely. Taking $M \to \infty$ along the integers we see that the sequence X_n has a finite limit almost surely on $\{\omega : \inf_n X_n(\omega) > -\infty\}$. Similarly by taking $-X_n$ we can argue that the sequence X_n has a finite limit almost surely on $\{\omega : \sup_n X_n(\omega) < \infty\}$. On the other hand, if $\inf_n X_n = -\infty$ and $\sup_n X_n = +\infty$ then clearly $\omega \in D$ completing the proof. □

THEOREM 5.1.12. *Let $\{X_n\}$ be a Martingale adapted to the filtration $\{\mathcal{F}_n\}$. If $\sup_n \mathrm{E}(X_n^2) < \infty$ then $X_n \to X_\infty$ a.s. and $\mathrm{E}((X_n - X_\infty)^2) \to 0$. Further $\mathrm{E}(X_\infty^2) < \infty$ and $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ a.s..*

PROOF. Since $\sup_n \mathrm{E}(X_n^2) < \infty$ we have $\sup_n \mathrm{E}(|X_n|) < \infty$; and Doob's convergence theorem yields the almost sure convergence. Fatou yields $\mathrm{E}(X_\infty^2) < \infty$.

W.l.o.g. let us center the Martingale and assume $X_0 = 0$. Further define $Y_n = X_n - X_{n-1}, n \geq 1$. Observe that

$$\mathrm{E}(Y_n X_{n-1}) = \mathrm{E}(X_{n-1}(X_n - X_{n-1})) = \mathrm{E}(\mathrm{E}(X_{n-1}(X_n - X_{n-1})|\mathcal{F}_{n-1})) = 0$$

In the last step we used that $\mathrm{E}(X_{n-1}X_n|\mathcal{F}_n) = X_{n-1}\mathrm{E}(X_n|\mathcal{F}_n) = X_{n-1}^2$ *a.s.* with the justification that tower property holds as $X_{n-1}X_n$ is integrable (Cauchy-Schwatrz). Hence by induction

$$\mathrm{E}(X_n^2) = \sum_{m=1}^{n} \mathrm{E}(Y_m^2).$$

Note that (Fatou yields)

$$\mathrm{E}((X_\infty - X_n)^2) \leq \lim_{m\to\infty} \mathrm{E}((X_m - X_n)^2) = \sum_{m=n+1}^{\infty} \mathrm{E}(Y_m^2) \to 0$$

as $n \to \infty$ as $\sum_{n\geq 1} \mathrm{E}(Y_n^2)$ is finite.

Note that for $m > n$

$$\mathrm{E}((X_\infty - X_m)|\mathcal{F}_n) = \mathrm{E}(X_\infty|\mathcal{F}_n) - X_n \ a.s.$$

Since $\mathrm{E}((X_\infty - X_m)^2) \to 0$ we have $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ *a.s.* □

Lemma 5.1.13. *Let X_n be integrable and X be integrable. Then $E(|X_n - X|) \to 0$ if (i) $X_n \to X$ in probability and (ii) $\{X_n\}$ is uniformly integrable.*

PROOF. Let $Y_n^K = X_n 1_{|X_n|\leq K} + \mathrm{sgn}(X_n)K$ and $Y^K = X 1_{|X|\leq K} + \mathrm{sgn}(X)K$. Since $|Y_n^K - Y^K| \leq |X_n - X|$ we see that $Y_n^K \to Y^K$ in probability and hence from bounded convergence theorem $\mathrm{E}(|Y_n^K - Y^K|) \to 0$.
Note that UI implies $\sup_n E(|X_n - Y_n^K|) \leq \sup_n \mathrm{E}(|X_n|1_{|X_n|>K}) \to 0$ and $K \to \infty$.
Now

$$\begin{aligned}\limsup_n \mathrm{E}(|X_n - X|) &\leq \limsup_n \mathrm{E}(|X_n - Y_n^K|) + \limsup_n \mathrm{E}(|Y_n^K - Y^K|) + \mathrm{E}(|X - Y^K|) \\ &\leq \sup_n \mathrm{E}(|X_n|1_{|X_n|>K}) + \mathrm{E}(|X|1_{|X|>K}),\end{aligned}$$

and the right hand side tends to 0 as $K \to \infty$. □

THEOREM 5.1.14. *Let $\{X_n\}$ be an uniformly-integrable Martingale adapted to the filtration $\{\mathcal{F}_n\}$. Then $X_n \to X_\infty$ a.s. and $\mathrm{E}(|X_n - X_\infty|) \to 0$. Further $\mathrm{E}(|X_\infty|) < \infty$ and $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ a.s..*

PROOF. $\{X_n\}$ is U-I implies $\sup_n \mathrm{E}(|X_n|) < \infty$. Hence, from convergence theorem, $X_n \to X_\infty$ a.s. and Fatou yields $\mathrm{E}(|X_\infty|) < \infty$. From previous lemma, we also have $\mathrm{E}(|X_n - X_\infty|) \to 0$. Note that for $m > n$

$$\mathrm{E}((X_\infty - X_m)|\mathcal{F}_n) = \mathrm{E}(X_\infty|\mathcal{F}_n) - X_n \ a.s.$$

Since $\mathrm{E}(|X_\infty - X_m|) \to 0$ we have

$$\mathrm{E}(|\,\mathrm{E}(X_\infty|\mathcal{F}_n) - X_n|) \le \mathrm{E}(\mathrm{E}(|X_\infty - X_m||\mathcal{F}_n)) = \mathrm{E}(|X_\infty - X_m|) \to 0.$$

Hence $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ *a.s.* □

THEOREM 5.1.15. *Fix $p > 1$. Let $\{X_n\}$ be a Martingale adapted to the filtration $\{\mathcal{F}_n\}$. If $\sup_n \mathrm{E}(|X_n|^p) < \infty$ then $X_n \to X_\infty$ a.s. and $\mathrm{E}(|X_n - X_\infty|^p) \to 0$. Further $\mathrm{E}(|X_\infty|^p) < \infty$ and $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ a.s..*

PROOF. $\sup_n \mathrm{E}(|X_n|^p) < \infty$ implies uniform integrability. Further Martingale convergence theorem implies almost sure convergence to a random variable X_∞ and Fatou implies X_∞ satisfies $\mathrm{E}(|X_\infty|^p) < \infty$. Previous theorem implies convergence in L_1.

Further a similar argument as before yields that

$$\mathrm{E}((X_\infty - X_m)|\mathcal{F}_n) = \mathrm{E}(X_\infty|\mathcal{F}_n) - X_n \ a.s.$$

Since $\mathrm{E}(|X_\infty - X_m|) \to 0$ we have $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ *a.s.*

Define Y_∞^K as before (from X_∞) and let $\hat{Y}_n^K = \mathrm{E}(Y_\infty^K|\mathcal{F}_n)$. Note that (Jensen for conditional expectation) $\|X_n - \hat{Y}_n^K\|_p \le \|X_\infty - Y_\infty^K\|_p$. Hence

$$\|X_n - X_\infty\|_p \le \|X_n - \hat{Y}_n^K\|_p + \|X_\infty - Y_\infty^K\|_p + \|Y_\infty^K - \hat{Y}_n^K\|_p \le 2\|X_\infty - Y_\infty^K\|_p + \|Y_\infty^K - \hat{Y}_n^K\|_p.$$

By construction $\hat{Y}_n^K$ is a bounded martingale sequence. Hence $\hat{Y}_n^K$ converges to $\hat{Y}_\infty^K$ a.s. and in L_2. Further $\mathrm{E}(\hat{Y}_\infty^K|\mathcal{F}_n) = Y_n$ a.s.

What is left is to show that $\hat{Y}_\infty^K = Y_\infty^K$ almost surely. By construction, $X_\infty, Y_\infty^K, \hat{Y}_\infty^K$ are $\sigma(\cup_n \mathcal{F}_n)$-measurable. Further since, for every $\mathcal{F}_n$

$$\hat{Y}_n^K = \mathrm{E}(\hat{Y}_\infty^K|\mathcal{F}_n) = \mathrm{E}(Y_\infty^K|\mathcal{F}_n) a.s$$

we have that, for all n and for all $A \in \mathcal{F}_n$

$$\int_A Y_\infty^K dP = \int_A Y^K dP.$$

The collection of all A for which the above equality holds is a monotone class (here use integrability of Y_∞^K and Y^K) and we are done. □

THEOREM 5.1.16 (Doob's decomposition). *Let $\{X_n\}, n \geq 0$ be a process adapted to $\{\mathcal{F}_n\}$. Then we can express*

$$X_n = X_0 + M_n + A_n,$$

where M_n is a martingale null at zero and A_n is a predictable. Further this decomposition is unique almost surely. Finally $\{X_n\}$ is a sub-martingale if and only if A_n is a non-decreasing sequence, almost surely.

PROOF. Note that if such a decomposition exists, then

$$\mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = A_n - A_{n-1}.$$

Therefore, by telescopic sum,

$$A_n = \sum_{k=1}^{n} \mathrm{E}(X_k - X_{k-1}|\mathcal{F}_{k-1})$$

since $A_0 = 0$. What remains to be verified is that $\{X_n - X_0 - A_n\}$ is a Martingale, but this is immediate. The uniqueness comes from the construction and essential uniqueness of the conditional expectation. The sub-martingale consequence is also immediate from the construction. □

DEFINITION 5.1.9 (Angle-bracket process). Let $\{X_n\}$ be a square-integrable martingale which is null at zero. Then $\{X_n^2\}$ is a sub-martingale and it has a Doob's decomposition leading to a predictable process $\{A_n\}$. Then we define $A_n = \langle X_n \rangle$.

Note that if $\mathrm{E}(A_\infty) < \infty$ then $\sup_n \mathrm{E}(X_n^2) \leq \mathrm{E}(A_\infty) < \infty$ and X_n converges almost-surely and in L_2.

THEOREM 5.1.17. *Let $\{X_n\}$ be a square-integrable martingale which is null at zero. Let $A_n = \langle X_n \rangle$. Then*

(*i*) *If $A_\infty(\omega) < \infty$ then $\lim_n X_n(\omega)$ exists (except possibly on a null set)*
(*ii*) *If X_n has uniformly bounded increments, the converse is true: i.e. when $\lim_n X_n(\omega)$ exists and is finite, then $A_\infty(\omega) < \infty$ (except possibly on a null set)*

PROOF. (*i*) Define a stopping time

$$S(k) = \inf\{n : A_{n+1} > k\}.$$

Then note that

$$A_{n\wedge S(k)} = A_{S(k)} 1_{S(k)\leq n-1} + A_n 1_{S(k)>n-1},$$

hence $A_{n\wedge S(k)}$ is predictable.

Observe that

$$\mathrm{E}(X^2_{n\wedge S(k)} - A_{n\wedge S(k)}|\mathcal{F}_{n-1}) = \left(X^2_{(n-1)\wedge S(k)} - A_{(n-1)\wedge S(k)}\right).$$

Thus $A_{n\wedge S(k)} = \langle X_{n\wedge S(k)} \rangle$. Since $A_{n\wedge S(k)}$ is bounded, hence $\mathrm{E}(A_{\infty\wedge S(k)}) \leq k$ implying $X_{n\wedge S(k)}$ converges almost surely. This implies X_n converges almost surely on $\{\omega : S(k) = \infty\}$. Taking $k \to \infty$ we have that If $A_\infty(\omega) < \infty$ then $\lim_n X_n(\omega)$ exists.

(ii) Define a new stopping time

$$T(k) = \inf\{n : |X_n| > k\}.$$

Hence, similar to before,

$$\mathrm{E}(X^2_{n\wedge T(k)}) = \mathrm{E}(A_{n\wedge T(k)}) \le (k+K)^2.$$

There on $\{\omega : T(k) = \infty\}$ we know that A_∞ is finite. Take $k \to \infty$. $\square$

THEOREM 5.1.18 (Levy's extension of Borel-Cantelli Lemmas). *Let $\{\mathcal{F}_n\}$ be a filtration and let $G_n \in \mathcal{F}_n$. Define*

$$Z_n = \sum_{k=1}^n 1_{G_k}.$$

Let $B_n = \mathrm{E}(1_{G_n}|\mathcal{F}_{n-1})$ and let

$$Y_n = \sum_{k=1}^n B_k.$$

Then, almost surely,

- $Y_\infty < \infty$ *implies* $Z_\infty < \infty$
- $Y_\infty = \infty$ *implies* $\frac{Z_n}{Y_n} \to 1$.

PROOF. Observe that

$$\mathrm{E}(Z_n - Y_n|\mathcal{F}_{n-1}) = Z_{n-1} - Y_{n-1},$$

hence $M_n := Z_n - Y_n$ is a Martingale. Let $A_n = \langle M_n \rangle$. Note that

$$\begin{aligned} A_n &= \sum_{k=1}^n \mathrm{E}(M_k^2 - M_{k-1}^2|\mathcal{F}_{k-1}) = \sum_{k=1}^n \mathrm{E}((M_k - M_{k-1})^2|\mathcal{F}_{k-1}) \\ &= \sum_{k=1}^n \mathrm{E}((1_{G_k} - B_k)^2|\mathcal{F}_{k-1}) = \sum_{k=1}^n B_k(1-B_k) \le Y_n. \end{aligned}$$

If $Y_\infty < \infty$ then $A_\infty < \infty$ and from previous theorem, $\lim_n M_n$ exists and hence $Z_\infty < \infty$.

To get the second part, note that if $A_\infty < \infty$ then $\lim_n M_n$ exists. Hence if $Y_\infty = \infty$ then $\frac{Z_n}{Y_n} \to 1$ (trivially). The more interesting case is when $A_\infty = \infty$.

We will show that the Martingale transform

$$U_n = \sum_{k=1}^n \frac{M_k - M_{k-1}}{1 + A_k}$$

is a square integrable martingale, where $U_0 = 0$. To observe this note that (please fill in the skipped steps),

$$\begin{aligned} \mathrm{E}(U_n^2) &= \mathrm{E}\left(\sum_{k=1}^n \frac{M_k - M_{k-1}}{1+A_k}\right)^2 \\ &= \mathrm{E}\left(\sum_{k=1}^n \left(\frac{M_k - M_{k-1}}{1+A_k}\right)^2 + 2\sum_{k<l} \frac{(M_k - M_{k-1})}{(1+A_k)}\frac{(M_l - M_{l-1})}{(1+A_l)}\right) \\ &= \mathrm{E}\left(\sum_{k=1}^n \left(\frac{M_k - M_{k-1}}{1+A_k}\right)^2\right) \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\left(\sum_{k=1}^{n}\left(\frac{1_{G_k}-B_k}{1+A_k}\right)^2\right)\\
&= \mathrm{E}\left(\sum_{k=1}^{n}\frac{B_k(1-B_k)}{(1+A_k)^2}\right)\\
&= \mathrm{E}\left(\sum_{k=1}^{n}\frac{A_k-A_{k-1}}{(1+A_k)^2}\right)\\
&\le \mathrm{E}\left(\sum_{k=1}^{n}\left(\frac{A_k}{1+A_k}-\frac{A_{k-1}}{1+A_{k-1}}\right)\right)\\
&= \mathrm{E}\left(\frac{A_n}{1+A_n}\right)\le 1.
\end{aligned}
$$

Now observe that, similarly,

$$
\begin{aligned}
\langle U_n\rangle &= \sum_{k=1}^{n}\mathrm{E}((U_k-U_{k-1})^2|\mathcal{F}_{k-1})\\
&= \sum_{k=1}^{n}\mathrm{E}\left(\frac{M_k^2-M_{k-1}^2}{(1+A_k)^2}|\mathcal{F}_{k-1}\right)\\
&\overset{a.s.}{=} \sum_{k=1}^{n}\frac{A_k-A_{k-1}}{(1+A_k)^2}\\
&\le \sum_{k=1}^{n}\left(\frac{A_k}{1+A_k}-\frac{A_{k-1}}{1+A_{k-1}}\right)=\frac{A_n}{1+A_n}\le 1.
\end{aligned}
$$

Since $\langle U_n\rangle \le 1$, hence U_n converges to a finite limit. Hence, Kronecker's lemma says that $\frac{M_n}{1+A_n}\to 0$. (Take $x_k=\frac{(M_k-M_{k-1})}{1+A_k}$ and $b_k=1+A_k$.) □

Lemma 5.1.19 (Kronecker)**.** *Let $\sum_{m=1}^n x_m$ be a convergent. It a sequence of positive numbers $b_n\uparrow\infty$ then*

$$\frac{\sum_{k=1}^n b_k x_k}{b_n}\to 0.$$

Proof. Let $s_n=\sum_{m=1}^n x_m$, and $s_n\to s$, and let $s_0=0$. The key is to express

$$\sum_{k=1}^{n}\frac{b_k x_k}{b_n}=\sum_{k=1}^{n}\frac{b_k((s_k-s)(s_{k-1}-s))}{b_n}=s_n-s-\sum_{k=1}^{n-1}\frac{(b_{k+1}-b_k)(s_k-s)}{b_n}.$$

Given $\epsilon>0$ we know that $\exists n_0$ such that $|s_n-s|\le\epsilon,\forall n\ge n_0$. Hence for $n>n_0$ we have

$$
\begin{aligned}
\left|\sum_{k=1}^{n}\frac{b_k x_k}{b_n}\right| &\le |s_n-s|+\sum_{k=1}^{n-1}\frac{(b_{k+1}-b_k)|s_k-s|}{b_n}\\
&\le \epsilon+\frac{(b_n-b_{n_0})}{b_n}\epsilon+\frac{1}{b_n}\sum_{k=1}^{n_0}(b_{k+1}-b_k)|s_k-s|.
\end{aligned}
$$

Taking $\limsup_n$ we obtain an upper bound of 2ϵ and since $\epsilon>0$ is arbitrary we are done. □

THEOREM 5.1.20 (Kakutani's theorem for Product Martingales). *Let $X_1, X_2, \cdots$ be independent non-negative random variables, each of mean 1. Define $M_0 = 1$ and let*

$$M_n := X_1 \cdots X_n.$$

Then $\{M_n\}$ is a non-negative martingale; hence $M_\infty := \lim_n M_n$ exists almost surely.

Secondly, the following five statements are equivalent:

(*a*) $\sum_n (1 - a_n) < \infty$.
(*b*) $\prod_n a_n > 0$, *where* $a_n = \mathrm{E}(\sqrt{X_n})$.
(*c*) $\{M_n\}$ *is UI.*
(*d*) $M_n \to M_\infty$ *in* L^1.
(*e*) $\mathrm{E}(M_\infty) = 1$.

If one of the statements fail to hold then $\mathrm{P}(M_\infty = 0) = 1$.

PROOF. Note that $0 < a_n \le 1$ (the second inequality follows by Jensen). The proof follows by showing that $(a) \iff (b)$. Then we will show that $(b) \implies (c) \implies (d) \implies (e) \implies (b)$. Finally we will show that if $\prod_n a_n = 0$ then $\mathrm{P}(M_\infty = 0) = 1$.

$(a) \iff (b)$: The equivalence between (a) and (b) is standard. Since $e^{-x} \ge 1-x$, we have

$$\prod_n a_n \le e^{-\sum_n (1-a_n)},$$

therefore (b) implies (a). Note that $e^{-2x} \le 1 - x, 0 \le x \le \frac{1}{2}$. Let $\sum_n (1 - a_n) = B < \infty$. Let $\mathcal{I} = \{n : a_n \le \frac{1}{2}\}$. Observe that $|\mathcal{I}| \le 2B$. Let $a_* = \min_{n \in \mathcal{I}} a_n$. Then observe that $\prod_{n \in \mathcal{I}} a_n \ge a_*^{|\mathcal{I}|} > 0$ and $\prod_{n \notin \mathcal{I}} a_n \ge e^{-2\sum_{n \notin \mathcal{I}}(1-a_n)} > 0$. Hence (a) implies (b).

$(b) \implies (c)$: Define

$$Y_n = \frac{\sqrt{X_1 X_2 \cdots X_n}}{a_1 a_2 \cdots a_n} = \frac{\sqrt{M_n}}{b_n},$$

where $b_n = \prod_{k=1}^n a_k$. Since Y_n is a non-negative Martingale we know that Y_n converges almost-surely to a finite limit. If $b_\infty = \prod_n a_n > 0$, then $\sup_n \mathrm{E}(Y_n^2) = \frac{1}{b_\infty^2} < \infty$. Now

$$\mathrm{E}(\sup_{1 \le k \le n} M_k) \le \mathrm{E}(\sup_{1 \le k \le n} Y_k^2) \le 4\,\mathrm{E}(Y_n^2) \le 4 \sup_n \mathrm{E}(Y_n^2) < \infty.$$

The second inequality is Doob's inequality for submartingales (Corollary 5.1.3), specialized to $p = 2$. Hence monotone convergence theorem yields

$$\mathrm{E}(\sup_n M_n) < \infty.$$

Defining $M_* = \sup_n M_n$, since $\mathrm{E}(M_*) < \infty$. Since, for every n, $\mathrm{E}(|M_n| 1_{|M_n|>K}) \le \mathrm{E}(|M_*| 1_{|M_*|>K})$ we have

$$\lim_{K \to \infty} \sup_n \mathrm{E}(|M_n| 1_{|M_n|>K}) \le \lim_{K \to \infty} \mathrm{E}(|M_*| 1_{|M_*|>K}) = 0,$$

where the last equality being due to the integrability of M_*. Thus $\{M_n\}$ is UI. Hence (b) implies (c).

$(c) \implies (d)$: As $M_n \to M_\infty$ almost surely and $\{M_n\}$ is uniformly integrable, Lemma 5.1.13 implies (d).

$(d) \implies (e)$: Note that $\mathrm{E}(M_\infty) \geq \mathrm{E}(M_n) - \mathrm{E}(|M_n - M_\infty|) = 1 - \mathrm{E}(|M_n - M_\infty|)$. Therefore (d) implies, by taking $n \to \infty$ that $\mathrm{E}(M_\infty) \geq 1$. On the other hand, from Fatou $E(M_\infty) \leq \liminf_n \mathrm{E}(M_n) = 1$. This establishes (e).

$(e) \implies (b)$: Consider Y_n defined as earlier. We knew that as Y_n is a non-negative Martingale it converges almost-surely to a finite limit. Now assume that $b_n \downarrow 0$. Then it must be that $M_\infty = 0$ almost surely. This contradicts $\mathrm{E}(M_\infty) = 1$.

Finally, if one of the conditions fail, say (b), then we have $b_\infty = 0$. Then it is clear from the almost-sure convergence of Y_n that $M_\infty = 0$ almost surely. $\square$

THEOREM 5.1.21 (Law of iterated Logarithm for Gaussians). *Let $\{X_n\}$ be i.i.d. random variables, each distributed as $\mathcal{N}(0,1)$. Let $S_n = \sum_{i=1}^n X_i$. Then*

$$\limsup_n \frac{S_n}{\sqrt{2n \log\log n}} = 1 \ a.s.$$

PROOF. Let $h_n = \sqrt{2n \log\log n}, n \geq 3$. Further it is easy to see that

$$\mathrm{E}(e^{\theta S_n}) = e^{\frac{1}{2}\theta^2 n}.$$

Now $\{e^{\theta S_n}\}$ is a sub-martingale, hence

$$P(\sup_{1 \leq k \leq n} S_k \geq l) \leq e^{-\theta l} e^{\frac{1}{2}\theta^2 n}.$$

Optimizing over θ yields

$$P(\sup_{1 \leq k \leq n} S_k \geq l) \leq e^{-\frac{l^2}{2n}}.$$

Take $K > 1$ and $l_n = Kh(K^{n-1})$. Now observe that

$$P\left(\sup_{1 \leq k \leq K^n} S_k \geq l_n\right) \leq e^{-\frac{l_n^2}{2K^n}} = e^{-K \log\log K^{n-1}} = \frac{1}{((n-1)\log K)^K}.$$

From B-C 1, we have, almost surely, for all $K^{n-1} \leq k \leq K^n$, and n large-enough $S_k \leq \sup_{1 \leq k \leq K^n} S_k \leq l_n = Kh(K^{n-1}) \leq Kh(k)$. Hence $\limsup_k \frac{S(k)}{h(k)} \leq K$ almost surely. Since $K > 1$ is arbitrary, we are done with the upper bound. By symmetry, we also have that $\liminf_k \frac{S(k)}{h(k)} \geq -1$ almost surely.

Let N be an integer larger than 1. Note that $\frac{S(N^{n+1}) - S(N^n)}{\sqrt{N^{n+1} - N^n}}$ is a Gaussian. Now from Lemma below we have

$$P(F_n) := P\left(\frac{S(N^{n+1}) - S(N^n)}{\sqrt{N^{n+1} - N^n}} \geq (1-\epsilon)\frac{h(N^{n+1} - N^n)}{\sqrt{N^{n+1} - N^n}}\right) \geq c(n \log N)^{-(1-\epsilon)^2}.$$

From B-C 2, infinitely often, we have

$$S(N^{n+1}) - S(N^n) \geq (1-\epsilon)h(N^{n+1} - N^n)$$

infinitely often. However from earlier part $S(N^n) \geq -Kh(N^n)$ for $K > 1$ eventually (in n), so infinitely often we have

$$S(N^{n+1}) \geq (1-\epsilon)h(N^{n+1} - N^n) - Kh(N^n).$$

Therefore

$$\limsup_k \frac{S(k)}{h(k)} \geq \limsup_n \frac{S(N^{n+1})}{h(N^{n+1})} \geq (1-\epsilon) + \theta(1/\sqrt{N}).$$

Letting $N \to \infty$ completes the proof. $\square$

Lemma 5.1.22. *Let X be standard normal. Then*

$$P(X > x) \geq (x + x^{-1}) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Proof. Let $\phi(x)$ be the density. Then $(x^{-1}\phi(x))' = -(1 + x^{-2})\phi(x)$. Hence

$$x^{-1}\phi(x) = -\int_x^\infty (y^{-1}\phi(y))' dy = \int_x^\infty (1 + y^{-2})\phi(y) dy \leq (1 + x^{-2}) P(X > x).$$

□

5.1.2. Backwards Martingales.

Definition 5.1.10. A *backwards martingale* is a sequence of random variables $\{X_n\}, n \leq 0$, adapted to a filtration, $\{\mathcal{F}_n\}_{n \leq 0}$ defined by

$$X_n := \mathrm{E}(X_0 | \mathcal{F}_n), n \leq 0,$$

where X_0 is an integrable random variable.

Lemma 5.1.23. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X be an integrable random variable. Consider the collection*

$$\{Y : Y = \mathrm{E}(X|\mathcal{G}),\ \textit{for some } \mathcal{G} \subset \mathcal{F}\}.$$

Then the collection of random variables is uniformly integrable.

Remark: Formally the collection contains versions of conditional expectation.

Proof. Let $\epsilon > 0$ be given. Let $c_\delta = \sup_{A \in \mathcal{F} : P(A) \leq \delta} \mathrm{E}(|X| 1_A)$. We know from dominated convergence theorem that $c_\delta \downarrow 0$ and $\delta \downarrow 0$. Choose δ_0 such that $c_{\delta_0} \leq \epsilon$. Choose M such that $\frac{1}{M} \mathrm{E}(|X|) \leq \delta_0$.

Jensen's inequality says that $|Y| \leq \mathrm{E}(|X||\mathcal{G})$ *a.s.* In particular $E(|Y|) \leq \mathrm{E}(|X|)$. Hence $P(|Y| > M) \leq \frac{1}{M} \mathrm{E}(|Y|) \leq \delta_0$. From the definition of conditional expectation

$$\mathrm{E}(|Y| 1_{|Y|>M}) \leq \mathrm{E}(|X| 1_{|Y|>M}) \leq \epsilon.$$

□

Theorem 5.1.24. *The limit $X_{-\infty} = \lim_{n \to -\infty} X_n$ exists a.s. and in L^1.*

Proof. Doob's upcrossing inequality for the number of upcrossings, U_n, between $[a, b]$ made by $X_{-n}, \ldots, X_0$ yields $(b - a)\mathrm{E}(U_n) \leq \mathrm{E}(X_0 - a)_+ < \infty$. Hence the limit exists almost surely. Since the collection is uniformly integrable (above lemma), we have convergence in L^1 (Lemma 5.1.13). □

5.1.3. Exchangeable Processes and de Finetti's theorem. Given a collection of random variables $X_1, X_2, ...$, take any measurable function $f(X_1, .., X_n, ...)$ and generate

$$A_n(f) = \frac{1}{{}^nP_k} \sum_{\pi \in S_n} f(X_{\pi(1)}, X_{\pi(2)}, .., X_{\pi(n)}, X_{n+1}, X_{n+2}, ...).$$

Define $\mathcal{E}_n$ to be the σ-algebra generated by $(\{A_n(f)\})$. In other words, let $\mathfrak{F}_n$ be the class of all measurable functions that are symmetric in the subset $X_1, .., X_n$. Hence $\mathcal{E}_n$ is generated by all random variables in $\mathfrak{F}_n$. It is immediate that $\mathfrak{F}_{n+1} \subseteq \mathfrak{F}_n$, hence $\mathcal{E}_{n+1} \subseteq \mathcal{E}_n$. Define $\mathcal{E} = \cap_n \mathcal{E}_n$ to be the *exchangeable* σ-algebra.

DEFINITION 5.1.11. A sequence $X_1, X_2, ..$ is said to be *exchangeable* if for every n and for every permutation $\pi \in S_n$,the distributions of $(X_1, .., X_n)$ and $X_{\pi(1)}, .., X_{\pi(n)}$ are the same.

THEOREM 5.1.25 (de Finetti). *If $X_1, X_2, ..$ are exchangeable, then conditioned on $\mathcal{E}$, the random variables $X_1, ...$ are independent and identically distributed.*

PROOF. Let f be a bounded measurable function. Define

$$A_n(f) := \frac{1}{{}^nP_k} \sum_{i \subset [n]} f(X_{i_1}, .., X_{i_k}).$$

Since A_n is exchangeable i.e. $\{\omega : A_n(f) \leq x\} \in \mathcal{E}_n$, we have

$$A_n(f) = \mathrm{E}(A_n(f)|\mathcal{E}_n) = \frac{1}{{}^nP_k} \sum_{I \subset [n], |I|=k} \mathrm{E}(f(X_{i_1}, .., X_{i_k})|\mathcal{E}_n) = \mathrm{E}(f(X_1, .., X_k)|\mathcal{E}_n).$$

From backwards martingale theorem $A_n(f) \to A_\infty(f) = \mathrm{E}(f(X_1, .., X_k)|\mathcal{E})$.

Let f, g be a bounded measurable functions on $\mathbb{R}^{k-1}$ and $\mathbb{R}$ respectively. Consider $\phi(x_1, .., x_k) = f(x_1, .., x_{k-1})g(x_k)$ and define $\phi_j(x_1, .., x_{k-1}) = f(x_1, .., x_{k-1})g(x_j)$ for $1 \leq j \leq k-1$. Then observe that

$${}^nP_{k-1} A_n(f) n A_n(g) = {}^nP_k A_n(\phi) + {}^nP_{k-1} \sum_{j=1}^{k-1} A_n(\phi_j) \qquad \Longleftrightarrow$$

$$A_n(f)A_n(g) = \frac{n-k+1}{n} A_n(\phi) + \frac{1}{n} \sum_{j=1}^{k-1} A_n(\phi_j).$$

Taking $n \to \infty$ we obtain $A_\infty(f)A_\infty(g) = A_\infty(\phi)$, or

$$\mathrm{E}(f(X_1, .., X_{k-1})|\mathcal{E})\, \mathrm{E}(g(X_k)|\mathcal{E}) = \mathrm{E}(f(X_1, .., X_{k-1})g(X_k)|\mathcal{E}).$$

The rest is routine. $\square$

CHAPTER 6

Ergodic Theorem

Definition 6.0.1. A sequence of random variables $\{X_n\}$ is said to be *stationary* if the distribution of $\{X_1, ..., X_n\}$ is identical to that of $\{X_{1+k}, ..., X_{n+k}\}$ for all $n \geq 1$ and $k \geq 1$.

Definition 6.0.2. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A measurable mapping $T : \Omega \to \Omega$ is said to be *measure-preserving* if $\mathrm{P}(T^{-1}(A)) = P(A)$ for all $A \in \mathcal{F}$.

A measure-preserving map naturally gives rise to a stationary sequence as follows: let $X(\omega)$ be a random variable; define $X_n(\omega) = X(T^n(\omega)), n \geq 0$, where $T^0(\omega) := \omega$.
To see that the above process is stationary, define $B := \{\omega : (X_1, ..., X_n) \in A\}$. Then note that $T^{-k}(B) = \{\omega : (X_{1+k}, ..., X_{n+k}) \in A\}$. Since T is measure-preserving, we are done.
This part of the notes will focus on measure-preserving transformations and hence T will always be assumed to be measure-preserving.

Definition 6.0.3. A set A is said to be *strictly-invariant* if $T^{-1}(A) = A$, while a set A is said to be *invariant* if $P(T^{-1}(A)\Delta A) = 0$.

Exercise 6.0.1. Show the following basic properties of mappings and sets.
1. $T^{-1}(\cup_i A_i) = \cup_i T^{-1}(A_i)$.
2. $T(T^{-1}(A)) = A$, $T^{-1}(T(A)) \supseteq A$.
3. $A\Delta B = A^c \Delta B^c$.
4. $\cup_i(A_i \Delta B_i) \supseteq (\cup_i A_i)\Delta(\cup_i B_i)$.
5. $T^{-1}(A^c) = (T^{-1}(A))^c$
6. $T^{-1}(A\Delta B) = T^{-1}(A)\Delta T^{-1}(B)$.
7. $A\Delta(\cup_{i=1}^{\infty} B_i) \subseteq (A\Delta B_1) \cup (\cup_{i=1}^{\infty}(B_i \Delta B_{i+1}))$

From the above properties, it is immediate that $\mathcal{I}$ - the collection of invariant sets - is a σ-algebra. we will call this to be the *invariant-σ-algebra*. Similarly, we can also define the *strictly-invariant-σ-algebra*.
Given an invariant set A, let $B := \cup_{n=0}^{\infty} T^{-n}(A)$. Note that $A \subseteq B$ and $T^{-1}(B) = \cup_{n=1}^{\infty} T^{-n}(A) \subseteq B$. Define $C := \cap_{n=0}^{\infty} T^{-n}(B)$. Note that $T^{-1}(C) = \cap_{n=1}^{\infty} T^{-n}(B)$; however since $B \cap T^{-1}(B) = T^{-1}(B)$ we have $T^{-1}(C) = C$. Argue that $P(A\Delta C) = 0$.

Lemma 6.0.1. *If X is $\mathcal{I}$-measurable then $X(T(\omega)) = X(\omega)$ almost surely.*

Proof. If A is invariant then $T^{-1}(A)$ is also invariant (use (6) in Exercise along with measure-preserving property). Therefore $X(T(\omega))$ is also $\mathcal{I}$-measurable. Given two rational numbers $p < q$ let $A_{p,q} = \{\omega : X(\omega) < p, X(T(\omega)) > q\}$, and let $B_p = \{\omega : X(\omega) < p\}$. It is clear that $A_{p,q} \subseteq B_p \Delta T^{-1}(B_p)$ and hence $P(A_{p,q}) = 0$. Now the lemma follows immediately. $\square$

Definition 6.0.4. A measure-preserving transformation associated with a stationary process is called *ergodic* if $A \in \mathcal{I}$, the invariant σ-algebra, implies that $P(A) = 0$ or $P(A) = 1$.

Lemma 6.0.2 (Maximal Ergodic Lemma)**.** *Let* $X_j(\omega) = X(T^k(\omega))$*,* $S_k = \sum_{i=0}^{k-1} X_i(\omega)$*, and* $M_k(\omega) = \max(0, S_1(\omega), .., S_k(\omega))$*. Then* $\mathrm{E}(X 1_{M_k > 0}) \geq 0$.

Proof. If $j \leq k$ then $M_k(T(\omega)) \geq S_j(T(\omega))$, implying

$$X(\omega) \geq S_{j+1}(\omega) - M_k(T(\omega)),\ j = 0, .., k$$

Therefore

$$\begin{aligned} \mathrm{E}(X(\omega) 1_{M_k > 0}) &\geq \int_{M_k > 0} \max\{S_1(\omega), ..., S_k(\omega)\} - M_k(T(\omega)) dP \\ &= \int_{M_k > 0} M_k(\omega) - M_k(T(\omega)) dP \geq 0. \end{aligned}$$

The last inequality is due to the following observation: $M_k(\omega) = 0$ we have $M_k(T(\omega)) \geq 0$. However since integrals of $M_k(\omega)$ and $M_k(T(\omega))$ are same (measure-preserving property of T), the inequality follows. $\square$

Theorem 6.0.3 (Birkhoff's Ergodic Theorem). *For any* $X \in L^1$*,*

$$\frac{1}{n} \sum_{m=0}^{n-1} X(T^m(\omega)) \to \mathrm{E}(X|\mathcal{I})\ a.s.$$

and in L^1*.*

Proof. Since $\mathrm{E}(X|\mathcal{I})$ is invariant under T (see Lemma 6.0.1 above) w.l.o.g. we can center X and assume $\mathrm{E}(X|\mathcal{I}) = 0$. Let $\bar{X} = \limsup \frac{S_n}{n}$ and let $D = \{\omega : \bar{X} > \epsilon\}$. Since $\bar{X}(T(\omega)) = \bar{X}(\omega)$, we have that $D \in \mathcal{I}$.

Define a new sequence of random variables $Y(\omega) = (X(\omega) - \epsilon)) 1_D$ and let $U_n = Y_0 + \cdots + Y_{n-1}$. Let $M_n(\omega) = \max(0, U_1(\omega), .., U_n(\omega))$. Observe that $M_n \uparrow$ and $\lim_n M_n > 0$ on D. Let $E_n = \{\omega : M_n > 0\}$. Hence $E_n \uparrow D$. Since $|Y| \leq |X| + \epsilon$ we have

$$0 \leq \mathrm{E}(Y 1_{E_n}) \to \mathrm{E}(Y 1_D).$$

where the inequality comes from Maximal ergodic lemma. Hence

$$\mathrm{E}((X(\omega) - \epsilon)) 1_D) \geq 0 \quad \Longrightarrow\ \mathrm{E}(\mathrm{E}(((X(\omega) - \epsilon) 1_D | \mathcal{I}))) = \mathrm{E}(1_D \mathrm{E}(X|\mathcal{I}) - \epsilon 1_D) \geq 0.$$

Since we have X centered, $\mathrm{E}(X|\mathcal{I}) = 0$ almost surely, implying that $\mathrm{P}(D) = 0$. Similarly working with $-X$ (and hence the $\liminf$) completes the almost sure convergence.

To show convergence in L_1, let $X_M = X 1_{|X| < M}$. Almost sure convergence above and bounded convergence theorem says that

$$E \left| \frac{1}{n} \sum_{m=0}^{n-1} X_M(T^m(\omega)) - \mathrm{E}(X_M|\mathcal{I}) \right| \to 0.$$

Let $\hat{X}_M = X - X_M$. By triangle inequality and measure-preserving property of T we have

$$E \left| \frac{1}{n} \sum_{m=0}^{n-1} \hat{X}_M(T^m(\omega)) \right| \leq \mathrm{E}(|\hat{X}_M|).$$

Note also that $|\operatorname{E}(\operatorname{E}(\hat{X}_M|\mathcal{I}))| \leq \operatorname{E}(|\hat{X}_M|)$. Combining

$$\begin{aligned}
&E\left|\frac{1}{n}\sum_{m=0}^{n-1} X(T^m(\omega)) - \operatorname{E}(X|\mathcal{I})\right| \\
&\leq E\left|\frac{1}{n}\sum_{m=0}^{n-1} X_M(T^m(\omega)) - \operatorname{E}(X_M|\mathcal{I})\right| + E\left|\frac{1}{n}\sum_{m=0}^{n-1} \hat{X}_M(T^m(\omega)) - \operatorname{E}(\hat{X}_M|\mathcal{I})\right| \\
&\leq E\left|\frac{1}{n}\sum_{m=0}^{n-1} X_M(T^m(\omega)) - \operatorname{E}(X_M|\mathcal{I})\right| + E\left|\frac{1}{n}\sum_{m=0}^{n-1} \hat{X}_M(T^m(\omega))\right| + \operatorname{E}\left|\operatorname{E}(\hat{X}_M|\mathcal{I})\right| \\
&\leq E\left|\frac{1}{n}\sum_{m=0}^{n-1} X_M(T^m(\omega)) - \operatorname{E}(X_M|\mathcal{I})\right| + 2\operatorname{E}(|\hat{X}_M|).
\end{aligned}$$

Since the first term of the right hand side goes to zero as $n \to \infty$, we have that the $\limsup_n$ of the term on the left-hand-side is upper bounded by $2\operatorname{E}(|X|1_{|X|\geq M})$. Since X is integrable and M is arbitrary, we are done. $\square$

Given a measurable transformation T, let $\mathcal{M}$ denote the convex set of all probability measures that is T-invariant (this could be empty).

THEOREM 6.0.4. *A probability measure $P \in \mathcal{M}$ is ergodic if and only if it is an extreme point of $\mathcal{M}$.*

PROOF. Assume that P is ergodic and yet $P = aP_1 + (1-a)P_2$ for $0 < a < 1$. Since P is ergodic, it implies that $P_1 = P_2$ on $\mathcal{I}$, hence P_1 and P_2 are also ergodic. Let f be any bounded measurable function on $(\omega, \mathcal{F})$. Define

$$h(\omega) = \lim_{n\to\infty} \frac{1}{n}\left(f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)\right)$$

when it exists. From Ergodic theorem, we know that the limit exists on a set E with $P_1(E) = P_2(E) = 1$. Further, from bounded convergence theorem we also know that

$$\operatorname{E}_{P_i}(h) = \int f dP_i \quad i = 1, 2.$$

However since h is $\mathcal{I}$ measurable and $P_1 = P_2$ on $\mathcal{I}$, we see that $\int f dP_1 = \int f dP_2$ for any bounded measurable function (in particular indicator functions). Hence $P_1 = P_2$ on $\mathcal{F}$.

If P is an extreme point of $\mathcal{M}$ and P is not ergodic, then there exists $A \in \mathcal{I}$ with $0 < P(A) < 1$. Define $P_1(E) = \frac{P(E\cap A)}{P(A)}$ and $P_2(E) = \frac{P(E\cap A^c)}{P(A^c)}$. Note that P_1, P_2 belong to $\mathcal{M}$ and P is a convex combination of P_1 and P_2. To show that $P_1 \neq P_2$, observe that we cannot have $\frac{P(E\cap A)}{P(A)} = \frac{P(E\cap A^c)}{P(A^c)} \forall E \in \mathcal{F}$ as it does not hold for $E = A$. Hence P (which can be written as a non-trivial convex combination of P_1, P_2) is not extremal in $\mathcal{M}$. $\square$

Lemma 6.0.5. *For any stationary measure P, the regular conditional probability of P given $\mathcal{I}$, denoted by $Q(\omega, :)$, is stationary and ergodic.*

PROOF. We know that almost surely

$$Q(\omega, A) = \operatorname{E}(1_A|\mathcal{I}).$$

We need to show that $Q(\omega, A) = Q(\omega, TA)$. Suffices to show that for all $I \in \mathcal{I}$

$$\int_I 1_A dP = \int_I 1_{TA} dP$$

or in other words $P(A \cap I) = P(TA \cap I)$ which is immediate due to invariance of I.

To show ergodicity, we need to show that $Q(\omega, I) = 0$ or 1, for $I \in \mathcal{I}$ for almost all ω. This is again immediate. (note that the issue of throwing away too many null sets was covered during definition of regular conditional probabilities). □

THEOREM 6.0.6. *Any invariant measure $P \in \mathcal{M}$ can be written as a convex combination of ergodic measures, i.e.*

$$P = \int_{\mathcal{M}_e} Q \mu_P(dQ).$$

PROOF. By regular conditional probabilities

$$P = \int Q(w, :) dP.$$

By previous lemma $Q(\omega, :) \in \mathcal{M}_e$ and hence we have a induced measure μ on measures in $\mathcal{M}_e$. By changing the integration with respect to that measure, we are done. □