Probability Limit Theorems

MATH-SHU 350

Yao Xiao

Fall 2023

Contents

8/28 Lecture

1 Measure Theory and Probability Spaces

1.1 Measure Theory

1.1.1 Basic Concepts

Definition 1.1. Let E be a set, and let $\mathscr{A}$ be a set of subsets of E. We say that

$$\begin{aligned}
&\text{if} \quad \varnothing \in \mathscr{A}, \quad A^C \in \mathscr{A},\ A \cup B \in \mathscr{A}, \quad \forall A, B \in \mathscr{A}, \quad \text{then } \mathscr{A} \text{ is an } \textbf{algebra}, && (1)\\
&\text{if} \quad \varnothing \in \mathscr{A}, \quad A \cap B \in \mathscr{A}, \quad \forall A, B \in \mathscr{A}, \quad \text{then } \mathscr{A} \text{ is a } \boldsymbol{\pi}\textbf{-system}, && (2)\\
&\text{if} \quad \varnothing \in \mathscr{A}, \quad A^C \in \mathscr{A},\ \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}, \quad \forall A, A_n \in \mathscr{A}, \quad \text{then } \mathscr{A} \text{ is a } \boldsymbol{\sigma}\textbf{-algebra}. && (3)
\end{aligned}$$

Definition 1.2. A pair $(E, \mathscr{E})$, where E is a set and $\mathscr{E}$ is a σ-algebra on E, is called a **measurable space**. In this case, each $A \in \mathscr{E}$ is called an **$\mathscr{E}$-measurable set**.

Definition 1.3. The intersection of all σ-algebras containing $\mathscr{C}$ (there is at least one) is called the **σ-algebra generated by $\mathscr{C}$**, denoted by $\sigma(\mathscr{C})$.

Remark 1.4. Note that $\mathscr{P}(E)$ (the set of all subsets of E) is a σ-algebra, and the intersection of any collection of σ-algebras is a σ-algebra. $\sigma(\mathscr{C})$ is the smallest σ-algebra containing $\mathscr{C}$.

Remark 1.5. It is commonly impossible to write down the typical element of a σ-algebra $\mathscr{E}$, so we try to find π-systems with simpler elements that generate our σ-algebra and work with them instead.

Definition 1.6. A **topological space** E with **topology** $\mathscr{T}$ is a space endowed with a set of open subsets. The **Borel σ-algebra** is then defined as $\mathscr{B}(E) = \sigma(\mathscr{T})$, which is the σ-algebra generated by the family of all open subsets of E. We use the standard abbreviation $\mathscr{B} := \mathscr{B}(\mathbb{R})$.

Example 1.7. Let $\pi(\mathbb{R})$ be the π-system $\pi(\mathbb{R}) := \{(-\infty, x];\ x \in \mathbb{R}\}$. We shall be able to check that $\sigma(\pi(\mathbb{R})) = \mathscr{B}(\mathbb{R})$, *i.e.*, $\sigma(\pi(\mathbb{R}))$ is the Borel σ-algebra of $\mathbb{R}$. See Homework 1.

Definition 1.8. Let E be a set and $\mathscr{A}$ be an algebra on E. A **set function** is any function $\mu : \mathscr{A} \to [0, \infty]$ with $\mu(\varnothing) = 0$.

Definition 1.9. Let E be a set, $\mathscr{A}$ be an algebra on E, and μ be a set function on $\mathscr{A}$. We say that

$$\begin{aligned}
&\mu \text{ is } \textbf{increasing} && \text{if} \quad \mu(A) \le \mu(B), && \forall A, B \in \mathscr{A} \text{ with } A \subseteq B, && (4)\\
&\mu \text{ is } \textbf{additive} && \text{if} \quad \mu(A \cup B) = \mu(A) + \mu(B), && \forall A, B \in \mathscr{A} \text{ disjoint}, && (5)\\
&\mu \text{ is } \boldsymbol{\sigma}\textbf{-additive} && \text{if} \quad \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n), && \forall \{A_n\}_{n\in\mathbb{N}} \subseteq \mathscr{A} \text{ disjoint, with } \bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}. && (6)
\end{aligned}$$

Remark 1.10. σ-additivity is also called countable additivity. Let $(E, \mathscr{E})$ be a measurable space, then if $\mu : \mathscr{E} \to [0, \infty]$ is countable additive, it is called a **measure** and the triple $(E, \mathscr{E}, \mu)$ is called a **measure space**. In particular if $\mu(E) = 1$, then μ is called a **probability measure** and $(E, \mathscr{E}, \mu)$ is called a **probability space**. The notation $(\Omega, \mathscr{F}, \mathbb{P})$ is often used instead for probability spaces.

Remark 1.11. Given a measure space $(E, \mathscr{E}, \mu)$, μ is said to be **finite** if $\mu(E) < \infty$, and **σ-finite** if there exists $\{E_n\}_{n\in\mathbb{N}} \subseteq \mathscr{E}$, such that $\mu(E_n) < \infty$ and $\bigcup_{n\in\mathbb{N}} E_n = E$. A set $F \in \mathscr{E}$ is called **μ-null** if $\mu(F) = 0$.

Remark 1.12. A statement ϕ about points $s \in E$ holds **almost everywhere** (*a.e.*) if $F = \{s \in E;\ \phi(s) \text{ is false}\} \in \mathscr{E}$ and F is μ-null, *i.e.*, $\mu(F) = 0$. For a probability space, almost everywhere is often called **almost surely** (*a.s.*).

Example 1.13 (Infinitely often)**.** A_n occurs infinitely often if $\forall p \in \mathbb{N}$, $\exists n \geq p$, such that A_n holds. Now look at the event that an element $\omega \in$ infinitely many A_n's. This is equivalent to $\forall p \in \mathbb{N}$, $\exists n \geq p$, such that $\omega \in A_n$. Then $\forall p \in \mathbb{N}$, $\omega \in \bigcup_{n \geq p} A_n$, *i.e.*, $\omega \in \bigcap_{p \in \mathbb{N}} \bigcup_{n \geq p} A_n$. Therefore, we can see that

$$\{\omega;\ \omega \in A_n \text{ infinitely often}\} = \bigcap_{p \in \mathbb{N}} \bigcup_{n \geq p} A_n. \tag{7}$$

Note that we can interpret unions and intersections as

$$\bigcup_{n \in \mathbb{N}} A_n = \{A_n \text{ occurs at least once}\}, \qquad \bigcap_{n \in \mathbb{N}} A_n = \{\text{all } A_n\text{'s occur}\}. \tag{8}$$

Example 1.14 (A set we cannot measure on $\mathbb{R}$)**.** Take the equivalence class $x \sim y \iff y - x \in \mathbb{Q}$. The axiom of choice enables us to define a set C, such that C as exactly one element in each equivalent class, and we assume that $C \subseteq [0,1]$ (applicable by translation). Assume $\mu(C) = 0$ and we define for some $r \in \mathbb{R}$ that

$$C_r = C + r = \{x + r;\ x \in C,\ r \in \mathbb{R}\}. \tag{9}$$

We can check that C_r, $r \in \mathbb{Q}$ are disjoint. Indeed, if we take $x \in C_r \cap C_s$, then we can write $x = c + r = c' + s$ for some $c, c' \in C$. This would imply that $c' = c + r - s$, meaning that $r = s$ since there is can be only one element in each equivalent class. This leads to a contradiction, so that $C_r \cap C_s = \varnothing$, $\forall r, s \in \mathbb{Q}$. Then we can see that

$$\mu(\mathbb{R}) = \mu\left(\bigcup_{r \in \mathbb{Q}} C_r\right) = \sum_{r \in \mathbb{Q}} \mu(C_r) = \underbrace{\sum_{r \in \mathbb{Q}} \mu(C)}_{\text{translation invariance}} = \sum_{r \in \mathbb{Q}} 0 = 0, \tag{10}$$

which leads to a contradiction. Therefore C is not measurable on $\mathbb{R}$. We shall also be able to check that if $\mu(C) > 0$, then $\mu((0,2)) = \infty$, and in general $\mu((a,b)) = \infty$, $\forall a < b$. Indeed, we have that

$$\mu((0,2)) \geq \mu\left(\bigcup_{r \in \mathbb{Q} \cap [0,1]} C_r\right) = \sum_{r \in \mathbb{Q} \cap [0,1]} \mu(C_r) = \sum_{r \in \mathbb{Q} \cap [0,1]} \mu(C) = \infty. \tag{11}$$

8/30 Lecture

Example 1.15 (An algebra that is not a σ-algebra)**.** Let $\mathbb{N}$ be the space and take

$$\mathscr{E} = \left\{A \subseteq \mathbb{N};\ A \text{ is finite or } A^C \text{ is finite}\right\}. \tag{12}$$

If $A \in \mathscr{E}$, then either A is finite (which means $(A^C)^C$ is finite) or A^C is finite, implying that $A^C \in \mathscr{E}$. Now if $A, B \in \mathscr{E}$, we discuss two cases. If A and B are both finite, then $A \cup B$ is finite. Otherwise A^C is finite, thus $(A \cup B)^C = A^C \cap B^C$ is finite. We have shown that $\mathscr{E}$ is indeed an algebra. However, if we take $A_n = \{2n\}$, then $A = \bigcup_{n \in \mathbb{N}} A_n$ is the set of all even numbers, which is not finite and does not have a finite complement either. It follows that $\mathscr{E}$ is not a σ-algebra.

Example 1.16. Given $A_1, \ldots, A_m$ subsets of a non-empty set E, and let $\mathscr{A} := \{A_1, \ldots, A_m\}$. The goal is to describe all elements of $\mathscr{E} = \sigma(\mathscr{A})$. To do this, we denote $A_i^1 := A_i$ and $A_i^0 := A_i^C$, then for some $\epsilon \in \{0,1\}^m$, define $A_\epsilon = \bigcap_{1 \leq i \leq m} A_i^{\epsilon_i}$. We shall first show that

$$A \in \mathscr{E} \iff A = \bigcup_{\epsilon \in I} A_\epsilon \quad \text{for some } I \subseteq \{0,1\}^m. \tag{13}$$

$\Longleftarrow$ This is trivial by definition. Each $A_i \in \mathscr{E}$, so each $A_i^C \in \mathscr{E}$, and thus each A_ϵ, being a finite union of A_i's and their complements, is in $\mathscr{E}$. This implies that any countable union of A_ϵ's is also in $\mathscr{E}$.

$\Longrightarrow$ Let

$$\mathscr{F} = \left\{ A \subseteq E;\ A = \bigcup_{\epsilon \in I} A_\epsilon \text{ for some } I \subseteq \{0,1\}^m \right\}. \tag{14}$$

Clearly $\varnothing \in \mathscr{F}$. Take an arbitrary $A = \bigcup_{\epsilon \in I} A_\epsilon \in \mathscr{F}$, then we have that

$$A^C = \bigcup_{\epsilon \in I^C} A_\epsilon \in \mathscr{F}. \tag{15}$$

This is because all A_ϵ's are disjoint, and their union is the entire space E. Now take $A_n \in \mathscr{F}$, $n \in \mathbb{N}$, then

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{\epsilon \in \bigcup_{n \in \mathbb{N}} I_n} A_\epsilon \in \mathscr{F}, \tag{16}$$

thus we can conclude that $\mathscr{F}$ is a σ-algebra. Note that $\mathscr{A} \subseteq \mathscr{F}$. Indeed, if we visualize on a Venn diagram, we will see that each A_ϵ represents a single and unique area in that diagram. Each set A_i would be some union of these single areas, so $\mathscr{A} \subseteq \mathscr{F}$. Moreover, $\mathscr{E} = \sigma(\mathscr{A})$ is the smallest σ-algebra containing $\mathscr{A}$, and since $\mathscr{F}$ is an σ-algebra containing $\mathscr{A}$, we can conclude that $\mathscr{E} \subseteq \mathscr{F}$, and the proof is complete.

Now we can see that each element of $\mathscr{E}$ can be written in the form of $A = \bigcup_{\epsilon \in I} A_\epsilon$ for some $I \subseteq \{0,1\}^m$. Note that for all $\epsilon \in \{0,1\}^m$ and $A \in \mathscr{E}$, either $A \cap A_\epsilon = \varnothing$ or $A_\epsilon \subseteq A$. This is the same meaning as "a single area in the Venn diagram" as in the proof above. The subsets A_ϵ are thus called the atoms of $\mathscr{E}$.

1.1.2 Construction and Characterization of Measures

As previously mentioned, it is difficult to directly work with a σ-algebra (for instance, defining a measure). Therefore, we try to specify the values on a π-system generating it, and extend the definition to the whole σ-algebra.

Lemma 1.17. Let E be a set, $\mathscr{I}$ be a π-system on E, and $\mathscr{E} = \sigma(\mathscr{I})$. Suppose that μ_1 and μ_2 are measures on $(E, \mathscr{E})$, such that $\mu_1|_{\mathscr{I}} = \mu_2|_{\mathscr{I}}$ and $\mu_1(E) = \mu_2(E) < \infty$, then $\mu_1 = \mu_2$ on $\mathscr{E}$.

Remark 1.18. Note that if $\mathscr{I}$ is not a π-system, the uniqueness (as specified in the lemma above) can fail. For instance, let $E = \{1,2,3,4\}$ and $\mathscr{I} = \{\{1,2\},\{1,3\}\}$. Let $r_i = \mu_1(\{i\})$ and $s_i = \mu_2(\{i\})$. Assuming that $\mu_1|_{\mathscr{I}} = \mu_2|_{\mathscr{I}}$, we have that $r_1 + r_2 = s_1 + s_2$ and $r_1 + r_3 = s_1 + s_3$. However, this does not necessarily mean that $r_1 + r_2 + r_3 + r_4 = s_1 + s_2 + s_3 + s_4$, so $\mu_1 = \mu_2$ does not necessarily hold on $\mathscr{E}$ (note that $\mathscr{E}$ necessarily contains E because it is a σ-algebra).

Theorem 1.19 (Carathéodory's extension theorem)**.** Let E be a set, $\mathscr{E}_0$ be an algebra on E, and $\mathscr{E} := \sigma(\mathscr{E}_0)$. Assume that $\mu_0 : \mathscr{E}_0 \to [0,\infty]$ is a countably additive set function on $\mathscr{E}_0$, then there exists a measure μ on $(E, \mathscr{E})$, such that $\mu = \mu_0$ on $\mathscr{E}_0$.

Remark 1.20. Lemma 1.17 and Theorem 1.19 are important tools for proving uniqueness and existence of measures (respectively), for instance, the Lebesgue measure.

Theorem 1.21. There exists a unique Borel measure μ on $\mathbb{R}$, such that $\forall a < b \in \mathbb{R}$, $\mu((a,b]) = b - a$. This measure μ is called the **Lebesgue measure on** $\mathbb{R}$.

1.1.3 Elementary Inequalities

Lemma 1.22. Let $(E, \mathscr{E}, \mu)$ be a measure space. Then,

$$\mu(A \cup B) \le \mu(A) + \mu(B), \qquad \mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{17}$$

Proof. We have that

$$\mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) \le \mu(A) + \mu(B), \tag{18}$$

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B \setminus A) + \mu(A \cap B) = \mu(A) + \mu(B), \tag{19}$$

so the proof is complete. The second equation in fact implies the first inequality. □

Remark 1.23. Both observations above can be generalized to a finite number of sets. The first inequality follows by the observation that $\bigcup_{n\in\mathbb{N}} F_n = \bigcup_{n\in\mathbb{N}} \left(F_n \setminus \bigcup_{i\leq n-1} F_i\right)$. The second equality follows by induction into the inclusion-exclusion formula.

9/4 Lecture

1.1.4 Monotone Convergence Properties

Theorem 1.24. Let $(E, \mathscr{E}, \mu)$ be a measure space.

(1) If $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathscr{E}$ and $F_n \uparrow F$, then $\mu(F_n) \uparrow \mu(F)$.

(2) If $\{G_n\}_{n\in\mathbb{N}} \subseteq \mathscr{E}$ and $G_n \downarrow G$ with $\mu(G_k) < \infty$ for some $k \in \mathbb{N}$, then $\mu(G_n) \downarrow \mu(G)$.

Note that $F_n \uparrow F$ means that $F_n \subseteq F_{n+1}$, $\forall n \in \mathbb{N}$, and $\bigcup_{n\in\mathbb{N}} F_n = F$.

Proof. (1) Let $G_1 = F_1$ and $G_n = F_n \setminus F_{n-1}$ for $n \geq 2$. Then the sets $\{G_n\}_{n\in\mathbb{N}}$ are disjoint, and that

$$\mu(F_n) = \mu\left(\bigcup_{k=1}^{n} G_k\right) = \sum_{k=1}^{n} \mu(G_k) \uparrow \sum_{k=1}^{\infty} \mu(G_k) = \mu\left(\bigcup_{k=1}^{\infty} G_k\right) = \mu\left(\bigcup_{k=1}^{\infty} F_k\right) = \mu(F). \tag{20}$$

(2) Take $F_n = G_k \setminus G_{k+n}$, then $F_n \uparrow G_k \setminus G$. Applying the previous part, we have that $\mu(F_n) \uparrow \mu(G_k \setminus G)$. Clearly $G_k \supseteq G_{k+n} \supseteq G$, so that $\mu(G_k) - \mu(G_{k+n}) = \mu(F_n) \uparrow \mu(G_k \setminus G) = \mu(G_k) - \mu(G)$, which implies that $\mu(G_{k+n}) \downarrow \mu(G)$. The proof is thus complete. □

Remark 1.25. The assumption in the second part of the previous theorem is necessary. We can take $H_n := (n, \infty)$, $n \in \mathbb{N}$ as a counterexample otherwise. Indeed, $H_n \downarrow \varnothing$, but $\mu(H_n) = \infty$, $\forall n \in \mathbb{N}$, where we take μ as the Lebesgue measure on $\mathbb{R}$.

1.2 Probability Spaces

We can think of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ as being a model for an experiment whose outcome is subject to chance. Ω is the set of all possible outcomes of the experiment; $\mathscr{F}$ is the set of observable outcomes (or events); and $\forall A \in \mathscr{F}$, $\mathbb{P}(A)$ is the probability of the event A, *i.e.*, the probability that the outcome $\omega \in \Omega$ of the experiment falls in the event A.

Example 1.26. (1) Let $(S, \mathscr{E}, \mu)$ be a measure space, and let $A \subseteq S$. Take $\mathscr{F} = \{A \cap C;\ C \in \mathscr{E}\}$. We shall be able to see that $\mathscr{F}$ is a σ-algebra on A, $\mu|_{\mathscr{F}}$ is a measure on $(A, \mathscr{F})$, and if $\mu(A) < \infty$, $\mathbb{P}(C) = \mu(C)/\mu(A)$ is a probability measure.

- $\varnothing \in \mathscr{F}$ because $\varnothing \in \mathscr{E}$. Take arbitrary $A \cap C \in \mathscr{F}$, then $(A \cap C)^C = A^C \cup C^C$. However, we are checking whether $\mathscr{F}$ is a σ-algebra on A, so this is just $C^C = A \cap C^C \in \mathscr{F}$ because $C^C \in \mathscr{E}$. Now take arbitrary $A \cap C_n \in \mathscr{F}$, then $\bigcup_{n\in\mathbb{N}}(A \cap C_n) = A \cap \left(\bigcup_{n\in\mathbb{N}} C_n\right) \in \mathscr{F}$ because $\bigcup_{n\in\mathbb{N}} C_n \in \mathscr{E}$. Therefore we have shown that $\mathscr{F}$ is a σ-algebra on $\mathscr{A}$.
- The properties of $\mu|_{\mathscr{F}}$ follow from that of μ.
- Clearly $\mathbb{P}(A) = 1$ and $\mathbb{P}(C) = \mu(C)/\mu(A) \leq \mu(C')/\mu(A) = \mathbb{P}(C')$ if $C \subseteq C'$. Therefore, it suffices to show that $\mathbb{P}$ is countably additive. Taking disjoint $C_n \in \mathscr{E}$, we have that

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} C_n\right) = \frac{\mu\left(\bigcup_{n\in\mathbb{N}} C_n\right)}{\mu(A)} = \frac{\sum_{n\in\mathbb{N}} \mu(C_n)}{\mu(A)} = \sum_{n\in\mathbb{N}} \frac{\mu(C_n)}{\mu(A)} = \sum_{n\in\mathbb{N}} \mathbb{P}(C_n). \tag{21}$$

(2) On the measurable space $(S, \mathscr{P}(S))$ where $\mathscr{P}(S)$ is the power set of S, the **Dirac delta measure** or **unit mass** at $x \in S$ is defined as

$$\delta_x(A) := \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases} \tag{22}$$

9/6 Lecture

Example 1.27. Consider an infinite sequence of coin tosses. In this case, we may choose $\Omega = \{H, T\}^{\mathbb{N}}$. A sample point $\omega \in \Omega$ is a sequence $\omega = (\omega_n)_{n\geq 1}$ where $\omega_n = 0, 1$ (0 stands for H and 1 stands for T). Note that Ω is a very large space, and any useful probability measure can only be defined on a smaller σ-algebra than that of all subsets of Ω. Now let $A_l = \{\omega_l = 0\}$ be the event that 0 is tossed at time l, $\forall l \geq 1$. We further denote $A_l^1 = A_l$ and $A_l^0 = A_l^C$, and let $A_I^\epsilon := \bigcap_{k\in I} A_k^\epsilon$, $\epsilon \in \{0, 1\}$. For any $\epsilon \in \{0, 1\}^m$, let $A_\epsilon := A_1^{\epsilon_1} \cap \ldots \cap A_m^{\epsilon_m}$. Now for all $k \in \mathbb{N}$, $j \in \mathbb{N}^k$, and $x \in \{H, T\}^k$, let

$$E_{k,j,x} = \{\omega;\ \omega_{j_1} = x_1, \ldots, \omega_{j_k} = x_k\}, \tag{23}$$

which is the event that x_1 is tossed at time j_1, etc., and x_k is tossed at time j_k. Let

$$\mathscr{C} = \left\{E_{k,j,x};\ k \in \mathbb{N},\ j \in \mathbb{N}^k,\ x \in \{H, T\}^k\right\}, \tag{24}$$

where $\mathscr{C}$ is sometimes called the **set of cylinders**.

- $E_{k,j,x}$ can be written as an intersection of subsets of the type A_I^ϵ. Indeed, we can take $I_1 = \{j_i;\ x_i = H\}$ and $I_2 = \{j_i;\ x_i = T\}$, then $E_{k,j,x} = A_{I_1}^0 \cap A_{I_2}^1$.
- $\mathscr{C}$ is a π-system. Indeed, $\varnothing \in \mathscr{C}$ if we take $k = 0$. Moreover, we can see that

$$E_{k,j,x} \cap E_{k',j',x'} = A_{I_1}^0 \cap A_{I_2}^1 \cap A_{I_1'}^0 \cap A_{I_2'}^1 = A_{I_1\cup I_1'}^0 \cap A_{I_2\cup I_2'}^1 \in \mathscr{C}. \tag{25}$$

Let $\mathscr{E}$ be the smallest algebra containing $\mathscr{C}$, defined as the intersection of all algebras containing $\mathscr{C}$.

- $A \in \mathscr{E}$ if and only if there exists $m \geq 1$ and $K \subseteq \{0, 1\}^m$ finite, such that $A = \bigcup_{\epsilon\in K} A_\epsilon$.

$\Longleftarrow$ This is trivial. Note that we are dealing with algebras instead of σ-algebras, so we can extend to finite unions but not countable unions. Now each A_ϵ is a finite intersection of elements in $\mathscr{E}$, so $A_\epsilon \in \mathscr{E}$. Therefore, any finite union $A = \bigcup_{\epsilon\in K} A_\epsilon$ should clearly be still in $\mathscr{E}$.

$\Longrightarrow$ Let

$$\mathscr{G} = \left\{A;\ A = \bigcup_{\epsilon\in K} A_\epsilon \text{ for some } K \subseteq \{0, 1\}^m \text{ finite, } m \geq 1\right\}. \tag{26}$$

Clearly $\varnothing \in \mathscr{G}$. Take an arbitrary $A = \bigcup_{\epsilon\in K} A_\epsilon \in \mathscr{G}$, then we have that

$$A^C = \bigcup_{\epsilon\in K^C} A_\epsilon \in \mathscr{G}. \tag{27}$$

Note that the complement of K is taken for a fixed m. This is because for a fixed m, all A_ϵ's are disjoint, and their union is the entire space. Now take $A_1, A_2 \in \mathscr{G}$, then

$$A_1 \cup A_2 = \bigcup_{\epsilon\in K_1\cup K_2} A_\epsilon \in \mathscr{G}, \tag{28}$$

because K_1 and K_2 are both finite and thus $K \subseteq \{0, 1\}^m$ is also finite. Therefore, we can conclude that $\mathscr{G}$ is an algebra. Note that $\mathscr{C} \subseteq \mathscr{G}$. Each $E_{k,j,x}$ is a fixed sequence of k coin tosses, and each A_ϵ represents a pattern for a subsequence of coin tosses. Their finite union is then multiple patterns for a subsequence of coin tosses. Clearly, each fixed sequence $E_{k,j,x}$ can find a suitable union to fall in. Now $\mathscr{G}$ is an algebra containing $\mathscr{C}$, while $\mathscr{E}$ is the smallest algebra containing $\mathscr{C}$. This means that $\mathscr{E} \subseteq \mathscr{G}$, and the proof is complete.

- $\mathbb{P}(E_{k,j,x}) := 2^{-k}$ (independent fair coin tosses) uniquely defines a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$, where $\mathscr{F} = \sigma(\mathscr{C})$ is the σ-algebra generated by $\mathscr{C}$. Uniqueness follows from Lemma 1.17, because $\mathscr{C}$ is a π-system as is previously shown. Now we check the existence. As we have previously shown, for each $A \in \mathscr{E}$, there exists $m \geq 1$ and $K \subseteq \{0, 1\}^m$ finite, such that $A = \bigcup_{\epsilon\in K} A_\epsilon$ and $\mathbb{P}(A) = |K|2^{-m}$. To see that this is self-consistent, suppose that $n \geq m$ and $A = \bigcup_{\epsilon'\in K\times\{0,1\}^{n-m}} A_{\epsilon'}$ is an alternative expression for A. Then we can check that

$$\mathbb{P}(A) = \frac{|K \times \{0, 1\}^{n-m}|}{2^n} = \frac{|K|2^{n-m}}{2^n} = \frac{|K|}{2^m}, \tag{29}$$

so $\mathbb{P}$ is well-defined. In addition, $\mathbb{P}$ is countably additive since any countable union of $A_n \in \mathscr{E}$ can be written as $\bigcup_{\epsilon\in K} A_\epsilon$ with K being finite, so $\mathbb{P}(\bigcup_n A_n) < \infty$. By Carathéodory's extension theorem (Theorem 1.19), $\mathbb{P}$ is well-defined on $\mathscr{F} = \sigma(\mathscr{C})$.

Let $T : \{0,1\}^{\mathbb{N}} \to [0,1]$, such that $\omega \mapsto 0.\omega_1\omega_2\ldots\omega_n\ldots = \sum_{n=1}^{\infty} \omega_n 2^{-n}$.

- Let $\mathscr{B}([0,1])$ be the Borel σ-algebra on $[0,1]$, then $T^{-1}(\mathscr{B}([0,1])) \subseteq \mathscr{F}$. First we note that T is not bijective. For instance, $(1,0,0,\ldots)$ and $(0,1,1,\ldots)$ are both mapped to $1/2$, indicating that T is not injective and thus not bijective. We split the interval $[0,1]$. For instance, $T^{-1}([0,1/2)) = \{\omega_1 = 0\} \setminus \{\omega_2 = \omega_3 = \ldots = 1\}$ (the excluded part would give $1/2$ because the sequence is infinite). As a result, we observe that

$$T^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right) \\ = E_{n,\{1,\ldots,n\},x} \setminus \{\omega_{n+1} = \omega_{n+2} = \ldots = 1\} \cup \{\omega_l = y_l \text{ for } 1 \le l \le n \text{ and } \omega_{n+1} = \omega_{n+2} = \ldots = 1\} \in \mathscr{F}, \tag{30}$$

where we set x and y such that

$$k = \sum_{i=1}^{n} 2^{n-i} x_i, \qquad k-1 = \sum_{i=1}^{n} 2^{n-i} y_i. \tag{31}$$

The reason is as follows. $E_{n,\{1,\ldots,n\},x}$ stands for the interval $[k2^{-n},(k+1)2^{-n})$, because it has already taken $k2^{-n}$ according to x, and the rest starts from $2^{-n-1}, 2^{-n-2}, \ldots$, cumulating to at most 2^{-n} so the sum will never exceed $(k+1)2^{-n}$. However, we need to exclude a case that sums to exactly $(k+1)2^{-n}$ from $E_{n,\{1,\ldots,n\},x}$ because $2^{-n-1}+2^{-n-2}+\ldots = 2^{-n}$. We also need an additional case that sums to exactly $k2^{-n}$ from $(k-1)2^{-n}$, which is the last term as above.

- Let μ be the measure defined on $\mathscr{B}([0,1])$ by $\mu(I) = \mathbb{P}(T^{-1}(I))$, $I \in \mathscr{B}([0,1])$. Then we shall be able to see that μ is the Borel measure on $[0,1]$. To do this, we want to show that any open subset of $[0,1]$ can be expressed as a countable union of $[k2^{-n},(k+1)2^{-n})$. Indeed, for any interval (a,b), we can write it as

$$(a,b) = \bigcup_{a<\frac{k}{2^n}<\frac{k+1}{2^n}\le b} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right). \tag{32}$$

Note that there always exists such $k \in \mathbb{Z}$. Indeed, let $a < k2^{-n} \le x < (k+1)2^{-n} \le b$, then there exists some integer k as long as $x - a > 2^{-n}$ and $b - x > 2^{-n}$. Now on the π-system of these $[k2^{-n},(k+1)2^{-n})$ intervals, we can see that μ is equal to the Borel measure on $[0,1]$. By the uniqueness lemma (Lemma 1.17), we can thus conclude that μ is exactly the Borel measure on $[0,1]$ on the whole of $\mathscr{B}([0,1])$.

1.2.1 Independence

Definition 1.28. Let I be a countable set. We say that the events $\{A_i \in \mathscr{F};\ i \in I\}$ are **independent** if, for all finite subsets $J \subseteq I$, we have that

$$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i). \tag{33}$$

We say that the σ-algebras $\{A_i \subseteq \mathscr{F};\ i \in I\}$ are **independent** if $\{A_i \in \mathscr{F};\ i \in I\}$ are independent whenever $A_i \in \mathscr{A}_i$ for all $i \in I$.

Example 1.29. Let $\Omega := [0,1]^2$, $\mathscr{F} = \mathscr{B}([0,1]^2)$, and μ be the Borel measure on $[0,1]$. Let $\mathbb{P} := \mathscr{L}([0,1]^2)$ be the Lebesgue measure on $\mathbb{R}^2$, *i.e.*, the unique measure $\mathbb{P}$ such that $\mathbb{P}(A_1 \times A_2) = \mu(A_1)\mu(A_2)$, $\forall A_1, A_2 \in \mathscr{B}([0,1])$. Now let $\mathscr{E}_1 := \{A \times [0,1];\ A \in \mathscr{B}([0,1])\}$ and $\mathscr{E}_2 := \{[0,1] \times B;\ B \in \mathscr{B}([0,1])\}$, then $\mathscr{E}_1$ and $\mathscr{E}_2$ are independent on $(\Omega, \mathscr{F}, \mathbb{P})$. Indeed, take arbitrary $G = A \times [0,1] \in \mathscr{E}_1$ and $H = [0,1] \times B \in \mathscr{E}_2$, we can see that

$$\mathbb{P}(G \cap H) = \mathbb{P}(A \times B) = \mu(A)\mu(B) = \mathbb{P}(A \times [0,1])\mathbb{P}([0,1] \times B) = \mathbb{P}(G)\mathbb{P}(H). \tag{34}$$

Example 1.30. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of events. Show that A_n, $n \in \mathbb{N}$ are independent if and only if the σ-algebras they generate, *i.e.*, $\sigma(A_n) = \{\varnothing, A_n, A_n^C, \Omega\}$, are independent.

$\Longleftarrow$ This is immediate by definition.

$\Longrightarrow$ It suffices to show that for all finite $J, K \subset \mathbb{N}$ disjoint, we have that

$$\mathbb{P}\left(\left(\bigcap_{n\in J} A_n\right) \cap \left(\bigcap_{n\in K} A_n^C\right)\right) = \prod_{n\in J} \mathbb{P}(A_n) \prod_{n\in K} \mathbb{P}(A_n^C). \tag{35}$$

We prove by mathematical induction on $|K|$. If $|K| = 0$, then this holds trivially by independence of $\{A_n\}_{n\in\mathbb{N}}$. Now assume that the above holds for all $K \subset \mathbb{N}$ with $|K| = N$, then for $|K| = N+1$, we can write $K = K_0 \cup \{j\}$. Using the inductive hypothesis, we can write that

$$\begin{aligned}
\mathbb{P}\left(\left(\bigcap_{n\in J} A_n\right) \cap \left(\bigcap_{n\in K} A_n^C\right)\right) &= \mathbb{P}\left(\left(\bigcap_{n\in J} A_n\right) \cap \left(\bigcap_{n\in K_0} A_n^C\right) \cap A_j^C\right) \\
&= \mathbb{P}\left(\left(\bigcap_{n\in J} A_n\right) \cap \left(\bigcap_{n\in K_0} A_n^C\right)\right) - \mathbb{P}\left(\left(\bigcap_{n\in J} A_n\right) \cap \left(\bigcap_{n\in K_0} A_n^C\right) \cap A_j\right) \\
&= \mathbb{P}\left(\left(\bigcap_{n\in J} A_n\right) \cap \left(\bigcap_{n\in K_0} A_n^C\right)\right) - \mathbb{P}\left(\left(\bigcap_{n\in J\cup\{j\}} A_n\right) \cap \left(\bigcap_{n\in K_0} A_n^C\right)\right) \\
&= \underbrace{\prod_{n\in J} \mathbb{P}(A_n) \prod_{n\in K_0} \mathbb{P}(A_n^C)}_{\text{inductive hypothesis}} - \underbrace{\prod_{n\in J\cup\{j\}} \mathbb{P}(A_n) \prod_{n\in K_0} \mathbb{P}(A_n^C)}_{\text{inductive hypothesis}} \\
&= \left(\prod_{n\in J} \mathbb{P}(A_n) \prod_{n\in K_0} \mathbb{P}(A_n^C)\right)(1 - \mathbb{P}(A_j)) = \left(\prod_{n\in J} \mathbb{P}(A_n) \prod_{n\in K_0} \mathbb{P}(A_n^C)\right)\mathbb{P}(A_j^C) \\
&= \prod_{n\in J} \mathbb{P}(A_n) \prod_{n\in K_0\cup\{j\}} \mathbb{P}(A_n^C) = \prod_{n\in J} \mathbb{P}(A_n) \prod_{n\in K} \mathbb{P}(A_n^C).
\end{aligned} \tag{36}$$

Therefore, we can conclude the proof by induction.

Theorem 1.31. Suppose that $\mathscr{G}$ and $\mathscr{H}$ are sub-σ-algebras of $\mathscr{F}$, and that $\mathscr{G}_0$ and $\mathscr{H}_0$ are π-systems with $\sigma(\mathscr{G}_0) = \mathscr{G}$ and $\sigma(\mathscr{H}_0) = \mathscr{H}$. Then $\mathscr{G}$ and $\mathscr{H}$ are independent if and only if $\mathscr{G}_0$ and $\mathscr{H}_0$ are independent.

Proof. $\Longrightarrow$ This is trivial since $\mathscr{G}_0 \subseteq \mathscr{G}$ and $\mathscr{H}_0 \subseteq \mathscr{H}$.

$\Longleftarrow$ Fix $G \in \mathscr{G}_0$, then take $H \mapsto \mathbb{P}(G \cap H)$ and $H \mapsto \mathbb{P}(G)\mathbb{P}(H)$. They are clearly equal since $\mathscr{G}_0$ and $\mathscr{H}_0$ are independent, and they have the same total mass since $\mathbb{P}(G \cap \Omega) = \mathbb{P}(G)\mathbb{P}(\Omega) = \mathbb{P}(G)$. By the uniqueness (Lemma 1.17), they should be equal on $\mathscr{H}$. Now fix $H \in \mathscr{H}_0$ and analogously repeat the previous arguments. We can conclude that $\mathbb{P}(G \cap H) = \mathbb{P}(G)\mathbb{P}(H)$ for any $G \in \mathscr{G}$ and $H \in \mathscr{H}$, so $\mathscr{G}$ and $\mathscr{H}$ are independent. ☐

1.2.2 The Fatou and Borel-Cantelli Lemmas

For the rest of this section, let $(E, \mathscr{E}, \mu)$ be a measure space and let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathscr{E}$.

Definition 1.32. We define the **limit superior** and **limit inferior** of the sequence $\{A_n\}_{n\in\mathbb{N}}$ respectively as

$$\limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m = \{A_n \text{ occurs infinitely often } (i.o.)\}, \tag{37}$$

$$\liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_m = \{A_n \text{ occurs at large times}\}. \tag{38}$$

Remark 1.33. If $(E, \mathscr{E}, \mu)$ is a probability space, then we can interpret $\limsup_{n\to\infty} A_n$ as being the event on which A_n occurs for infinitely many n, and $\liminf_{n\to\infty} A_n$ as being the event on which there exists some (random) N such that A_n occurs for all $n \geq N$. Note that we always have $\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$.

Example 1.34. Check that

$$\mathbb{1}_{\limsup A_n} = \limsup \mathbb{1}_{A_n}, \qquad \mathbb{1}_{\liminf A_n} = \liminf \mathbb{1}_{A_n}. \tag{39}$$

9/11 Lecture

Lemma 1.35 (Reverse Fatou lemma for sets)**.** Assume that μ is finite, then $\mu(\limsup A_n) \geq \limsup \mu(A_n)$.

Proof. Let $G_m := \bigcup_{n\geq m} A_n$, then $G_m \downarrow \limsup A_n$ and therefore $\mu(G_m) \downarrow \mu(\limsup A_n)$ by the monotone convergence theorem (Theorem 1.24). Now note that $\mu(G_m) \geq \sup_{n\geq m} \mu(A_n)$ because G_m is larger than any of them. Therefore by taking $m \to \infty$ on both sides of the inequality, we can conclude that

$$\mu\left(\limsup_{n\to\infty} A_n\right) = \lim_{m\to\infty} \mu(G_m) \geq \lim_{m\to\infty} \sup_{n\geq m} \mu(A_n) = \limsup_{n\to\infty} \mu(A_n), \tag{40}$$

and the proof is complete. □

Lemma 1.36 (The first Borel-Cantelli lemma)**.** Assume $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$, then $\mu(\limsup A_n) = \mu(\{A_n \ i.o.\}) = 0$.

Proof. Let $G_m := \bigcup_{n\geq m} A_n$, then we have that $\mu(G_m) \leq \sum_{n\geq m} \mu(A_n) < \infty$. Note that $G_m \downarrow \limsup A_n$ and therefore $\mu(G_m) \downarrow \mu(\limsup A_n)$ by the monotone convergence theorem (Theorem 1.24), bring $m \to \infty$ so that

$$\mu(\limsup_{n\to\infty} A_n) \leq \lim_{m\to\infty} \sum_{n\geq m} \mu(A_n) = 0, \tag{41}$$

and the proof is complete. □

Lemma 1.37 (The second Borel-Cantelli lemma)**.** Assume that $\{A_n\}_{n\in\mathbb{N}}$ are independent events. If $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(\limsup A_n) = \mathbb{P}(\{A_n \ i.o.\}) = 1$.

Proof. By independence, for all m we have that

$$\mathbb{P}\left(\bigcap_{n\geq m} A_n^C\right) = \prod_{n\geq m} (1 - \mathbb{P}(A_n)) \leq \prod_{n\geq m} \exp\left(-\mathbb{P}(A_n)\right) = \exp\left(-\sum_{n\geq m} \mathbb{P}(A_n)\right) = 0. \tag{42}$$

Therefore, we can deduce that

$$1 \geq \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} A_n\right) = 1 - \mathbb{P}\left(\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_n^C\right) \geq 1 - \sum_{m\in\mathbb{N}} \mathbb{P}\left(\bigcap_{n\geq m} A_n^C\right) = 1, \tag{43}$$

which completes the proof. □

9/13 Lecture

Example 1.38 (Applying the first Borel-Cantelli lemma)**.** Assume you play some money at discrete times $2, 3, 4, \ldots$ with the following rule: if $A_n := \{\text{win } n^2 - 1 \text{ pounds at time } n\}$ and $B_n := \{\text{lose 1 pound at time } n\}$, then $\mathbb{P}(A_n) = n^{-2}$ and $\mathbb{P}(B_n) = 1 - n^{-2}$. Then the game is *"fair"* in the sense that, at each discrete time n, you make in expectation $(n^2 - 1)n^{-2} - (1 - n^{-2}) = 0$ pounds. However, since $\sum_{n\in\mathbb{N}} n^{-2} < \infty$, the first Borel-Cantelli lemma implies that you only win finitely often.

Example 1.39 (Applying the second Borel-Cantelli lemma)**.** A monkey is provided with a typewriter and, at each time step, it has probability $1/26$ to type any of the 26 letters, independently of other times. What is the probability to type `ABRACADABRA` at least once? What about infinitely often? We can consider the events

$$A_k := \{\texttt{ABRACADABRA} \text{ is typed between times } 11k + 1 \text{ and } 11(k + 1)\}. \tag{44}$$

The events A_k, $k \in \mathbb{N}$, are independent (because their time ranges are disjoint). Moreover, $\mathbb{P}(A_k) = (1/26)^{11}$, so that $\sum_k \mathbb{P}(A_k) = \infty$. By the second Borel-Cantelli lemma, A_k would occur infinitely often.

Lemma 1.40 (Fatou lemma for sets)**.** For any measure μ, we have that $\mu(\liminf A_n) \leq \liminf \mu(A_n)$.

Proof. Let $G_m := \bigcap_{n\geq m} A_n$, then $G_m \uparrow \liminf A_n$ and therefore $\mu(G_m) \uparrow \mu(\liminf A_n)$ by the monotone convergence theorem (Theorem 1.24). Now note that $\mu(G_m) \leq \inf_{n\geq m} \mu(A_n)$ because G_m is smaller than any of them. Therefore by taking $m \to \infty$ on both sides of the inequality, we can conclude that

$$\mu\left(\liminf_{n\to\infty} A_n\right) = \lim_{m\to\infty} \mu(G_m) \leq \lim_{m\to\infty} \inf_{n\geq m} \mu(A_n) = \liminf_{n\to\infty} \mu(A_n), \tag{45}$$

and the proof is complete. □

Remark 1.41. Note that Fatou lemma for sets does not require any condition on the measure μ, but the inverse Fatou lemma for sets require μ to be finite. This is because of the different requirements when applying the monotone convergence theorem on increasing or decreasing sequences of sets.

2 Measurable Functions and Random Variables

2.1 Measurable Functions

Definition 2.1. Let $(E, \mathscr{E})$ and $(G, \mathscr{G})$ be measurable spaces and let μ be a measure on $(E, \mathscr{E})$. A function $f : E \to G$ is said to be **measurable** if $f^{-1}(A) \in \mathscr{E}$ whenever $A \in \mathscr{G}$, where $f^{-1}(A)$ denotes the **inverse image** of A, such that $f^{-1}(A) = \{x \in E;\ f(x) \in A\}$.

Example 2.2. (1) Usually $G = \mathbb{R}$ or $G = [-\infty, \infty]$, and $\mathscr{G}$ is the Borel σ-algebra.

(2) If E and G are topological spaces, and $\mathscr{E} = \mathscr{B}(E)$ and $\mathscr{G} = \mathscr{B}(G)$ are Borel σ-algebras (*i.e.*, generated by the open subsets) on E and G, respectively. Then a measurable function on E taking values in G is called a **Borel function**. In particular, all continuous functions are measurable (which will be shown later). Note that any open set can be expressed as a (not necessarily countable) union of open balls.

(3) Let $(E, \mathscr{E})$ be a measurable space and $A \in \mathscr{E}$. Then the indicator function $\mathbb{1}_A : (E, \mathscr{E}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mathbb{1}_A(x) = 1$ if $x \in A$ and 0 otherwise, is measurable. Indeed, for any $B \in \mathscr{B}(\mathbb{R})$, we have that

$$\mathbb{1}_A^{-1}(B) = \begin{cases} E, & \text{if } 1 \in B \text{ and } 0 \in B, \\ A, & \text{if } 1 \in B \text{ and } 0 \notin B, \\ A^C, & \text{if } 1 \notin B \text{ and } 0 \in B, \\ \varnothing, & \text{otherwise,} \end{cases} \tag{46}$$

where clearly $E, A, A^C, \varnothing \in \mathscr{E}$.

(4) Given a function $f : E \to G$ and a σ-algebra $\mathscr{G}$ on G, let $\sigma(f) = f^{-1}(\mathscr{G}) := \left\{f^{-1}(G);\ G \in \mathscr{G}\right\}$. We can check that $\sigma(f)$ is a σ-algebra, called the **σ-algebra generated by** f. Indeed, $\varnothing \in \sigma(f)$ because $\varnothing \in \mathscr{G}$. For any $A \in \sigma(f)$, there exists $G \in \mathscr{G}$ such that $A = f^{-1}(G)$. Then $A^C = f^{-1}(G^C) \in \sigma(f)$ because $G^C \in \mathscr{G}$. Moreover, for any $\{A_n\}_{n\in\mathbb{N}} \subseteq \sigma(f)$, there exists $G_n \in \mathscr{G}$ such that $A_n = f^{-1}(G_n)$ for each n. Then $\bigcup_{n\in\mathbb{N}} A_n = f^{-1}\left(\bigcup_{n\in\mathbb{N}} G_n\right) \in \sigma(f)$ because $\bigcup_{n\in\mathbb{N}} G_n \in \mathscr{G}$. We have thus proved that $\sigma(f)$ is indeed a σ-algebra. We shall also be able to check that f is measurable if and only if $\sigma(f) \subseteq \mathscr{E}$ (which, in other words, means that $f^{-1} : \mathscr{G} \to \mathscr{E}$). Indeed, $\sigma(f) \in \mathscr{E}$ is equivalent to $f^{-1}(\mathscr{G}) \in \mathscr{E}$, which is equivalent to $f^{-1}(G) \in \mathscr{E}$ for all $G \in \mathscr{G}$, and is the definition of measurability of f. More generally, given a measurable space $(G, \mathscr{G})$, $I \subseteq \mathbb{N}$, and functions $f_i : E \to G$, $i \in I$, the σ-algebra $\mathscr{E} = \sigma\left(\left\{f_i^{-1}(A);\ A \in \mathscr{G}, i \in I\right\}\right)$ is called the **σ-algebra generated by** $\{f_i\}_{i\in I}$.

Proposition 2.3. Let $(E, \mathscr{E})$, $(G, \mathscr{G})$, and $(H, \mathscr{H})$ be measurable spaces. Take $\{A_i\}_{i\in\mathbb{N}} \in \mathscr{E}$ and let $f : E \to G$.

(1) The mapping f^{-1} preserves all set operations, *i.e.*, $f^{-1}\left(\bigcup_i A_i\right) = \bigcup_i f^{-1}(A_i)$ and $f^{-1}(A^C) = f^{-1}(A)^C$.

(2) If $\mathscr{G} = \sigma(\mathscr{A})$ for some $\mathscr{A} \subseteq \mathscr{G}$ and $f^{-1} : \mathscr{A} \to \mathscr{E}$, then $f^{-1} : \mathscr{G} \to \mathscr{E}$, *i.e.*, f is measurable.

(3) If E and G are topological spaces, $\mathscr{E} = \mathscr{B}(E)$, and $\mathscr{G} = \mathscr{B}(G)$, then f is measurable whenever it is continuous.

(4) If $G=\mathbb{R}$ and $\mathscr{G}=\mathscr{B}(\mathbb{R})$, then $f: E \to \mathbb{R}$ is measurable if and only if for all $c \in \mathbb{R}$, one has that

$$\{f \geq c\} := \{x \in E;\ f(x) \geq c\} \in \mathscr{E}. \tag{47}$$

(5) If $f: E \to G$ and $g: G \to H$ are measurable, then $g \circ f: E \to H$ is measurable.

Proof. (1)
- For any $x \in f^{-1}\left(\bigcup_i A_i\right)$, we have that $f(x) \in \bigcup_i A_i$. Then $f(x) \in A_i$ for some i, implying that $x \in f^{-1}(A_i)$ for some i. Hence $x \in \bigcup_i f^{-1}(A_i)$.
- For any $x \in \bigcup_i f^{-1}(A_i)$, we have that $x \in f^{-1}(A_i)$ for some i. Then $f(x) \in A_i$ for some i, implying that $f(x) \in \bigcup_i A_i$. Hence $x \in f^{-1}\left(\bigcup_i A_i\right)$.
- The proof for complements is the analogous in both inclusions.

(2) By the previous part, $\left\{A;\ f^{-1}(A) \in \mathscr{E}\right\}$ is a σ-algebra containing $\mathscr{A}$. But note that $\mathscr{G}$ is the smallest σ-algebra containing $\mathscr{A}$, so $\mathscr{G} \subseteq \left\{A;\ f^{-1}(A) \in \mathscr{E}\right\}$. This means that for any $A \in \mathscr{G}$, necessarily $f^{-1}(A) \in \mathscr{E}$, so we can conclude that f^{-1} is measurable.

(3) Since E and G are topological spaces, assuming f is continuous, we have that $f^{-1}(A)$ is open whenever $A \subseteq G$ is open. Since $\mathscr{G} = \mathscr{B}(G) = \sigma(\mathscr{T}_G)$ where $\mathscr{T}_G$ is the collection of all open sets in G, this degrades to the previous problem. Indeed, for any $A \in \mathscr{T}_G$, we have that $f^{-1}(A) \subseteq \sigma(\mathscr{T}_E) = \mathscr{B}(E) = \mathscr{E}$, so $f^{-1}: \mathscr{T}_G \to \mathscr{E}$, which corresponds to the condition in the previous part. Now we can conclude that f^{-1} is measurable.

(4) $\Longrightarrow$ If f is measurable, then for any $A \in \mathscr{B}(\mathbb{R})$, $f^{-1}(A) \in \mathscr{E}$. Since $[c,\infty) \in \mathscr{B}(\mathbb{R})$ for any $c \in \mathbb{R}$, we have that

$$f^{-1}([c,\infty]) = \{x \in E;\ f(x) \in [c,\infty]\} = \{f \geq c\} \in \mathscr{E}. \tag{48}$$

$\Longleftarrow$ Let $\mathscr{A} = \{[c,\infty);\ c \in \mathbb{R}\}$, then we have that $f^{-1}(A) \in \mathscr{E}$ for any $A \in \mathscr{A}$. Note that $\mathscr{B}(\mathbb{R}) = \sigma(\mathscr{A})$, so that f is measurable by the second part.

(5) This is trivial. Since f and g are measurable, we have that for any $A \in \mathscr{G}$ and $B \in \mathscr{H}$, $f^{-1}(A) \in \mathscr{E}$ and $f^{-1}(B) \in \mathscr{G}$. Then we have that

$$(g \circ f)^{-1}(B) = f^{-1}(\underbrace{g^{-1}(B)}_{\in \mathscr{G}}) \in \mathscr{E}, \tag{49}$$

so $g \circ f$ is measurable.

□

Lemma 2.4. Let $(E,\mathscr{E})$ be a measurable space, and let $f_n: E \to \mathbb{R}$, $n \in \mathbb{N}$ be measurable functions. Then,

(1) The function $g: E \to \mathbb{R}^2$ such that $x \mapsto (f_1(x), f_2(x))$ is measurable.

(2) Suppose that $F: \mathbb{R}^2 \to \mathbb{R}$ is a measurable function, then $h(x) := F(f_1(x), f_2(x))$ is measurable. In particular, $f_1 + f_2$ and $f_1 f_2$ are measurable.

(3) $\inf_{n\to\infty} f_n$, $\sup_{n\to\infty} f_n$, $\liminf_{n\to\infty} f_n$, and $\limsup_{n\to\infty} f_n$ are measurable.

(4) $\{s \in E;\ \lim_{n\to\infty} f_n(s) \text{ exists in } \mathbb{R}\} \in \mathscr{E}$.

Proof. (1) Note that $\mathscr{B}(\mathbb{R}^2) = \sigma(\{A \times B;\ A,\ B \text{ open subsets of } \mathbb{R}\})$, and so by the second part of Proposition 2.3, it suffices to check that for any $A, B \subseteq \mathbb{R}$, $g^{-1}(A \times B) = f_1^{-1}(A) \cap f_2^{-1}(B) \in \mathscr{E}$. Since f_1 and f_2 are both measurable, this clearly holds, and thus we can conclude that g is measurable.

(2) Note that $h = F \circ g$, and since F and g are both measurable, we can conclude that h is measurable by the fifth part of Proposition 2.3.

(3) We take the infimum as an example and the rest would be similar. For any $c \in \mathbb{R}$, we have that

$$\left(\inf_{n\to\infty} f_n\right)^{-1}((-\infty,c]) = \left\{\inf_{n\to\infty} f_n \leq c\right\} = \bigcap_{n\in\mathbb{N}} \{f_n \leq c\} = \bigcap_{n\in\mathbb{N}} f_n^{-1}((-\infty,c]). \tag{50}$$

Since each $f_n^{-1}((-\infty, c]) \in \mathscr{E}$ (because $(-\infty, c] \in \mathscr{B}(\mathbb{R})$ and f_n is measurable), we can see that the above is also in $\mathscr{E}$ because $\mathscr{E}$ is a σ-algebra. This means that $(\inf_{n\to\infty} f_n)^{-1} : \pi(\mathbb{R}) \to \mathscr{E}$, and since $\mathscr{B}(\mathbb{R}) = \sigma(\pi(\mathbb{R}))$, we can conclude that $(\inf_{n\to\infty} f_n)^{-1} : \mathscr{B}(\mathbb{R}) \to \mathscr{E}$, *i.e.*, $\inf_{n\to\infty} f_n$ is measurable. All the others can be done analogously. sup is taking union and using $[c, \infty)$'s. lim inf and lim sup are just sup of inf and inf of sup, respectively.

(4) For any $s \in E$, we observe that

$$\left\{\lim_{n\to\infty} f_n(s) \text{ exists in } \mathbb{R}\right\} = \left\{\limsup_{n\to\infty} f_n(s) < \infty\right\} \cap \left\{\liminf_{n\to\infty} f_n(s) > -\infty\right\} \cap \left\{\limsup_{n\to\infty} f_n(s) = \liminf_{n\to\infty} f_n(s)\right\}$$
$$= \underbrace{\left(\left(\limsup_{n\to\infty} f_n\right)^{-1} \underbrace{([-\infty, \infty))}_{\in\mathscr{B}(\mathbb{R})}\right)}_{\in\mathscr{E}} \cap \underbrace{\left(\left(\liminf_{n\to\infty} f_n\right)^{-1} \underbrace{((-\infty, \infty])}_{\in\mathscr{B}(\mathbb{R})}\right)}_{\in\mathscr{E}} \cap \underbrace{\left(\left(\limsup_{n\to\infty} f_n - \liminf_{n\to\infty} f_n\right)^{-1} \underbrace{(\{0\})}_{\in\mathscr{B}(\mathbb{R})}\right)}_{\in\mathscr{E}}, \tag{51}$$

since a sequence converges to a finite limit if and only if its lim inf and lim sup are equal, and by the measurability of lim inf and lim sup as shown in the previous part. Since $\mathscr{E}$ is a σ-algebra, the proof is already complete. ☐

Definition 2.5. Let $(E, \mathscr{E})$ and $(G, \mathscr{G})$ be measurable spaces, and let μ be a measure on $(E, \mathscr{E})$. Then a measurable function $f : E \to G$ induces an **image measure** $\nu = \mu \circ f^{-1}$ on $(G, \mathscr{G})$, given by

$$\nu(A) = \mu(f^{-1}(A)), \qquad \forall A \in \mathscr{G}. \tag{52}$$

9/18 Lecture

Definition 2.6. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and $(G, \mathscr{G})$ be a measure space. Then a measurable function $X : \Omega \to G$ is called a **random variable**. *In practice and unless otherwise stated, the random variables we consider in this course are taking values in* $G := \mathbb{R}$. The image measure $\mu_X := \mathbb{P} \circ X^{-1}$ is called the **law of distribution** of X. If X is real-valued, then its **distribution function** $F_X : \mathbb{R} \to \mathbb{R}$ is given by

$$F_X(x) = \mu_X((-\infty, x]) = \mathbb{P}(X \le x). \tag{53}$$

Remark 2.7. Note that since $\{(-\infty, x];\ x \in \mathbb{R}\}$ is a π-system generating $\mathscr{B}(\mathbb{R})$, the image measure of the random variable X, μ_X, is uniquely determined by its distribution function (Lemma 1.17).

Lemma 2.8. Let X be a random variable on some probability space. Then the distribution function of X, F_X, has the following properties.

(1) $F_X : \mathbb{R} \to [0, 1]$, and F_X is monotonically increasing.

(2) $\lim_{x\to\infty} F_X(x) = 1$, and $\lim_{x\to-\infty} F_X(x) = 0$.

(3) F_X is right-continuous, *i.e.*, $F_X(y) \to F_X(x)$ as $y \downarrow x$.

Proof. (1) Since μ_X is non-decreasing with image $[0, 1]$, this is trivial.

(2) This follows by the monotone convergence theorem (Theorem 1.24).

(3) If $x_n \downarrow x$, then $(-\infty, x_n] \downarrow (-\infty, x]$. Therefore $\mu_X((-\infty, x_n]) \downarrow \mu_X((-\infty, x])$ by the monotone convergence theorem, and hence $F_X(x_n) \downarrow F_X(x)$.

☐

Remark 2.9. F_X is *not* left-continuous. If $x_n \uparrow x$, then $(-\infty, x_n] \uparrow (-\infty, x)$. Then $\mu_X((-\infty, x_n]) \uparrow \mu_X((-\infty, x))$ by the monotone convergence theorem, but μ_X can thus have a "jump" at x, breaking then continuity.

Remark 2.10. Given a function F satisfying the three properties above, there exists a unique probability measure μ, such that for any $x \in \mathbb{R}$, we have that $\mu((-\infty, x]) = F(x)$. We call such a measure a **Lebesgue-Stieltjes measure**. This uniqueness of μ comes from the uniqueness on the π-system $\pi(\mathbb{R}) := \{(-\infty, x];\ x \in \mathbb{R}\}$. Similarly, we could prove the existence by Carathéodory's extension theorem, but this would require to show the countable additivity of μ on an algebra containing $\pi(\mathbb{R})$, as in the proof of the existence of Lebesgue measure (Homework 1).

Definition 2.11. We say that random variables X_n, $n \in \mathbb{N}$ are **independent** if the σ-algebras $\sigma(X_n)$, $n \in \mathbb{N}$ are independent.

Remark 2.12. For real-valued random variables, this is equivalent to the condition that

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \ldots \mathbb{P}(X_n \le x_n), \qquad \forall x_1, \ldots, x_n \in \mathbb{R}. \tag{54}$$

Indeed, we proved that the independence of two σ-algebras only needed to be checked on their generating π-systems, but the same holds true in the case of independence of a finite number of σ-algebras.

Remark 2.13. The random variables X_n, $n \in \mathbb{N}$ are called **independent identically distributed** if they are independent and, moreover, all of them have the same distribution (*i.e.*, the distribution functions μ_{X_n} are equal for all $n \in \mathbb{N}$).

Remark 2.14. A sequence of random variables $\{X_n;\ n \in \mathbb{N}\}$ is often regarded as a **process** evolving in time. The σ-algebra generated by $X_0, \ldots, X_n$ defined as

$$\mathscr{F}_n := \sigma(X_0, \ldots, X_n) = \sigma\left(\left\{X_i^{-1}(B);\ 1 \le i \le n,\ B \in \mathscr{B}(\mathbb{R})\right\}\right), \tag{55}$$

contains those events depending (measurably) on $X_0, \ldots, X_n$, and represents *what is known about the process by time* n.

2.1.1 Tail Events

Definition 2.15. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Define

$$\mathscr{T}_n := \sigma(X_{n+1}, X_{n+2}, \ldots), \qquad \mathscr{T} := \bigcap_{n\in\mathbb{N}} \mathscr{T}_n. \tag{56}$$

Then $\mathscr{T}$ is called the **tail σ-algebra** of the sequence $\{X_n\}_{n\in\mathbb{N}}$.

Remark 2.16. We can think of the tail σ-algebra as containing the events describing the limiting behavior of the sequence. For instance, Let $X_n \in \{H, T\}$, then $A = \{X_n = H\ i.o.\}$ is a tail event.

Theorem 2.17 (Kolmogorov's 0-1 law)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables. Then the tail σ-algebra $\mathscr{T}$ of $\{X_n\}_{n\in\mathbb{N}}$ contains only events of probability 0 or 1. Moreover, any $\mathscr{T}$-measurable random variable is almost surely constant.

Proof. Let $\mathscr{F}_n := \sigma(X_1, \ldots, X_n)$. Note that $\mathscr{F}_n$ is generated by the π-system of events

$$\mathscr{A} = \{\{X_1 \le x_1, \ldots, X_n \le x_n\};\ x_1, \ldots, x_n \in \mathbb{R}\}, \tag{57}$$

and that $\mathscr{T}_n$ is generated by the π-system of events

$$\mathscr{B} = \{X_{n+1} \le x_{n+1}, \ldots, X_{n+k} \le x_{n+k};\ x_{n+1}, \ldots, x_{n+k} \in \mathbb{R},\ k \in \mathbb{N}\}. \tag{58}$$

For any $A \in \mathscr{A}$ and $B \in \mathscr{B}$, we have that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ by independence of $\{X_n\}_{n\in\mathbb{N}}$. Then $\mathscr{A}$ and $\mathscr{B}$ are independent, and thus by Theorem 1.31, $\mathscr{F}_n$ and $\mathscr{T}_n$ are independent. Hence $\mathscr{F}_n$ and $\mathscr{T}$ are independent as well. Now $\bigcup_{n\in\mathbb{N}} \mathscr{F}_n$ is a π-system (because $\mathscr{F}_n$ is increasing) generating the σ-algebra $\mathscr{F}_\infty = \sigma(\{X_n\}_{n\in\mathbb{N}})$. By Theorem 1.31 again, we can see that $\mathscr{F}_\infty$ and $\mathscr{T}$ are independent. But $\mathscr{T} \subseteq \mathscr{F}_\infty$, so that for any $A \in \mathscr{T}$, we have that

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A) \implies \mathbb{P}(A) \in \{0, 1\}. \tag{59}$$

Now take an arbitrary $\mathscr{T}$-measurable random variable Y, then $F_Y(y) = \mathbb{P}(Y \le y)$ is right-continuous and only takes values in $\{0, 1\}$ because $\{Y \le y\} \in \mathscr{T}$ is a tail event. Let $c := \inf\{y;\ F_Y(y) = 1\}$, then for any $y < c$, we have that $F_y(y) \ne 1$ so $F_Y(y) = 0$. This means that $\mathbb{P}(Y < c) = 0$. But $\mathbb{P}(Y \le c) \ne 0$ because $F_Y(c) = 1$. This means that $\mathbb{P}(Y \le c) = 1$. As a consequence, $\mathbb{P}(Y = c) = 1$, and the proof is thus complete. ☐

Example 2.18. Let X_n, $n \in \mathbb{N}$ be independent identically distributed integrable random variables, and let $Z_n := \sum_{k=1}^{n} X_k$. Consider the random variable $L := \limsup_{n\to\infty} Z_n/n$, then we can check that L is measurable with respect to the tail σ-algebra $\mathscr{T} := \bigcap_{n\in\mathbb{N}} \sigma(X_{n+1}, X_{n+2}, \ldots)$. Indeed, for any $p \in \mathbb{N}$, we observe that

$$L = \limsup_{n\to\infty} \frac{Z_n}{n} = \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \limsup_{n\to\infty} \frac{1}{n}\sum_{k=p}^{n} X_k. \tag{60}$$

This is clearly $\sigma(X_{p+1}, X_{p+2}, \ldots)$-measurable. Therefore, L is $\mathscr{T}$-measurable as well. Now by Komolgorov's 0-1 law, L is constant almost surely. We will prove later in the strong law of large numbers that the limit of Z_n/n exists almost surely and is equal to the expectation of X_1. Now add the assumption that X_n's have mean 0 and variance 1. Consider the random variable $T := \limsup_{n\to\infty} Z_n/\sqrt{2n\log\log n}$. Then similarly, T is constant almost surely, and in fact T takes the value 1 almost surely. This is called the **Law of the Iterative Logarithm**.

2.2 Convergence in Measure and Convergence Almost Everywhere

Definition 2.19. Let f_n, $n \in \mathbb{N}$ and f be measurable functions. We say that a sequence of measurable functions $\{f_n\}_{n\in\mathbb{N}}$ converges to f **almost everywhere** (or **almost surely** if $\mu(E) = 1$) if

$$\mu(\{x;\ f_n(x) \not\to f(x) \text{ as } n \to \infty\}) = 0. \tag{61}$$

We say that $\{f_n\}_{n\in\mathbb{N}}$ converges to f **in measure** (or **in probability** if $\mu(E) = 1$) if

$$\mu(\{x;\ |f_n(x) - f(x)| > \epsilon\}) \to 0 \text{ as } n \to \infty, \qquad \forall \epsilon > 0. \tag{62}$$

Theorem 2.20. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions.

(1) Assume $\mu(E) < \infty$. If $f_n \to f$ almost surely, then $f_n \to f$ in measure.

(2) If $f_n \to f$ in measure, then for some subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$, $f_{n_k} \to f$ almost surely.

Proof. For all $n \in \mathbb{N}$ and $\epsilon > 0$, let $g_n = f_n - f$ and $A_{n,\epsilon} = \{|g_n| > \epsilon\}$. It follows from the definition of convergence that, for all $\omega \in E$, $f_n(\omega) \to f(\omega)$ as $n \to \infty$ if and only if for all $\epsilon > 0$, $\limsup A_{n,\epsilon}$ does not hold (*i.e.*, $A_{n,\epsilon}$ holds only finitely often).

(1) Since μ is finite by assumption, we can use the reverse Fatou Lemma (Lemma 1.35) to obtain that

$$\mu(A_{n,\epsilon}\ i.o.) = \mu\left(\limsup_{n\to\infty} A_{n,\epsilon}\right) \geq \limsup_{n\to\infty} \mu(A_{n,\epsilon}). \tag{63}$$

But since $f_n \to f$ almost surely, we have that $\mu(\limsup_{n\to\infty} A_{n,\epsilon}) = 0$. Therefore, $\lim_{n\to\infty} \mu(A_{n,\epsilon}) = 0$, *i.e.*, $f_n \to f$ in measure.

(2) Since $f_n \to f$ in measure, there exists an increasing subsequence $\{n_k\}_{k\in\mathbb{N}}$, such that $\sum_{k\in\mathbb{N}} \mu\left(A_{n_k,1/k}\right) < \infty$. WHY IS THIS? Then we can use the first Borel-Cantelli lemma (Lemma 1.36) to see that

$$\mu\left(\limsup_{k\to\infty} A_{n_k,1/k}\right) = \mu\left(\{A_{n_k,1/k}\ i.o.\}\right) = 0. \tag{64}$$

This means that $\mu(\{|f_{n_k} - f| \geq 1/k\ i.o.\})$, so $f_{n_k} \to f$ almost surely and the proof is complete. □

9/20 Lecture

Example 2.21 (Convergence almost everywhere but not in measure)**.** On real numbers with lebesgue measure $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathscr{L}_{\mathbb{R}})$, we take the indicator functions $f_n := \mathbb{1}_{[n,\infty)}$. Clearly $f_n \to 0$ almost surely because $[n,\infty) \downarrow \varnothing$. However, $\mu(\{f_n > 1/2\}) = \mu([n,\infty)) = \infty$ for any $n \in \mathbb{N}$, which does not converge to 0. This counterexample is due to the breakage of $\mu(E) < \infty$ (here $E = \mathbb{R}$).

Example 2.22 (Convergence in measure but not almost everywhere)**.** On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, let A_n, $n \in \mathbb{N}$ be independent events such that $\mathbb{P}(A_n) \to 0$ but $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$. Take the indicator functions $f_n := \mathbb{1}_{A_n}$. Clearly $f_n \to 0$ in measure because $\mathbb{P}(\{f_n > \epsilon\}) = \mathbb{P}(A_n) \to 0$ as $n \to \infty$ if $0 < \epsilon < 1$, and $\mathbb{P}(\{f_n > \epsilon\}) = 0$ if $\epsilon \geq 1$. But by the second Borel-Cantelli lemma (Lemma 1.37), we have that $\mathbb{P}(A_n\ i.o.) = 1$. Then $f_n = \mathbb{1}_{A_n} \not\to 0$ almost surely.

3 Integration, Expectation, and L^p Spaces

Let $(E, \mathscr{E}, \mu)$ be a measure space. We want to define, when possible, for measurable functions $f : E \to [-\infty, \infty]$, the integral of f with respect to the measure μ, such that

$$\mu(f) = \int_E f d\mu = \int_{x\in E} f(x)\mu(dx). \tag{65}$$

For random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, the integral will be called **expectation** of X, denoted $\mathbb{E}(X)$.

Defining Integrals of Simple Functions

A **simple function** is a function of the form $f = \sum_{k=1}^m a_k \mathbb{1}_{A_k}$, where $0 < a_k \le \infty$ and each $A_k \in \mathscr{E}$. In this case, we define the integral of the simple function f to be

$$\mu(f) := \sum_{k=1}^m a_k \mu(A_k), \tag{66}$$

with the convention that $\infty \cdot 0 = 0$. We can check that $\mu(f)$ is well-defined. Indeed, if $f = \sum_{k=1}^m a_k \mathbb{1}_{A_k} = \sum_{l=1}^n b_l \mathbb{1}_{B_l}$, we observe that for any k and l such that $a_k \neq b_l$, we have that $A_k \cap B_l = \varnothing$, since otherwise we can take $x \in A_k \cap B_l$ but then $a_k = f(x) = b_l$. Using this observation and assuming that A_k's and B_l's are respectively disjoint, we can deduce that

$$\begin{aligned}\mu(f) &= \sum_{k=1}^m a_k\mu(A_k) = \sum_{k=1}^m a_k \mu\left(\bigcup_{l=1}^n (A_k\cap B_l)\right) = \sum_{k=1}^m a_k \sum_{l=1}^n \mu(A_k\cap B_l) = \sum_{k=1}^m\sum_{l=1}^n a_k\mu(A_k\cap B_l) \\ &= \underbrace{\sum_{k=1}^m\sum_{l=1}^n b_l\mu(A_k\cap B_l)}_{\text{either } a_k = b_l \text{ or } \mu(A_k\cap B_l)=0} = \sum_{l=1}^n b_l \sum_{k=1}^m \mu(A_k\cap B_l) = \sum_{l=1}^n b_l\mu\left(\bigcup_{k=1}^m (A_k\cap B_l)\right) = \sum_{l=1}^n b_l\mu(B_l) = \mu(f),\end{aligned} \tag{67}$$

so that $\mu(f)$ is independent of the representation of the simple function f. Moreover, we shall be able to check the following properties. They are all trivial applying similar observations as above, so the proofs will be neglected here.

(1) $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$.

(2) If $f \le g$, then $\mu(f) \le \mu(g)$.

(3) $f = 0$ almost everywhere if and only if $\mu(f) = 0$.

Extending to Integrals of General Non-Negative Measurable Functions

For general non-negative measurable functions f, we define the integral by

$$\mu(f) := \sup\{\mu(g);\ g \text{ simple},\ g \le f\}. \tag{68}$$

Note that the second property above implies that this definition is consistent with the definition of μ for simple functions.

Further Extending to Integrals of General Measurable Functions

Let $f^+ := f \vee 0$ and $f^- := (-f) \vee 0$. Let $f := f^+ - f^-$, then we have that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. We say that a measurable function f is **integrable** if $\mu(|f|) < \infty$, and define

$$\mu(f) := \mu(f^+) - \mu(f^-). \tag{69}$$

3.1 Convergence Results

Theorem 3.1 (Egorov's theorem)**.** Let $E \in \mathscr{E}$ be such that $\mu(E) < \infty$. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions, and assume that $f_n \to f$ μ-almost surely on E. Then for all $\epsilon > 0$, there exists $A_\epsilon \in \mathscr{E}$, such that $\mu(E \setminus A_\epsilon) < \epsilon$ and $f_n \to f$ uniformly on A_ϵ.

Proof. Fix an arbitrary $\epsilon > 0$, and let $A_{n,\epsilon} := \{x \in E;\ |f_n(x) - f(x)| \geq \epsilon\}$. Since $f_n \to f$ μ-almost surely on E, we know that $\mu(\{x;\ f_n(x) \not\to f(x) \text{ as } n \to \infty\})$. Note that $f_n(x) \not\to f(x)$ means that for any $N \in \mathbb{N}$, there exists $n \geq N$, such that $|f_n(x) - f(x)| \geq \epsilon$. In other words, we can see that

$$\mu\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} A_{n,\epsilon}\right) = 0. \tag{70}$$

Now let $B_{N,\epsilon} := \bigcup_{n\geq N} A_{n,\epsilon}$. Clearly, $B_{N,\epsilon} \downarrow \bigcap_{N\in\mathbb{N}} B_{N,\epsilon}$ as $N \to \infty$, and since $B_{N,\epsilon} \subseteq E$ which has finite measure, we can use the monotone convergence theorem (Theorem 1.24) to see that

$$\mu(B_{N,\epsilon}) \downarrow \mu\left(\bigcap_{N\in\mathbb{N}} B_{N,\epsilon}\right) = \mu\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} A_{n,\epsilon}\right) = 0, \qquad \text{as } N \to \infty. \tag{71}$$

Now choose $\{N_k\}_{k\in\mathbb{N}}$, such that $\mu(B_{N_k,1/k}) < 2^{-k}\epsilon$. This means that

$$\mu\left(\bigcup_{k\in\mathbb{N}} B_{N_k,1/k}\right) \leq \sum_{k\in\mathbb{N}} \mu(B_{N_k,1/k}) < \sum_{k\in\mathbb{N}} \epsilon 2^{-k} = \epsilon. \tag{72}$$

Take $A_\epsilon = \bigcap_{k\in\mathbb{N}} B^C_{N_k,1/k}$, then we have that $\mu(E \setminus A_\epsilon) = \mu(A_\epsilon^C) = \mu\left(\bigcup_{k\in\mathbb{N}} B_{N_k,1/k}\right) < \epsilon$. Note that for any $x \in A_\epsilon$, we have that for any $k \in \mathbb{N}$ and for any $n \geq N_k$, $|f_n(x) - f(x)| < \epsilon$. This necessarily implies uniform convergence on A_ϵ, and thus the proof is complete. □

Theorem 3.2 (Monotone convergence theorem for measurable functions)**.** Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions with $f_n \geq 0$ almost everywhere. If $f_n \uparrow f$ almost everywhere, then $\mu(f_n) \uparrow \mu(f)$.

Proof. Fix a simple function $g = \sum_{k=1}^m a_k \mathbb{1}_{A_k}$, such that $g \leq f$. Note that if $\mu(A_k) = \infty$ for some $1 \leq k \leq m$, then if $f_n \uparrow f$, we have that $\{f_n \geq a_k/2\} \cap A_k \uparrow A_k$, and thus by the monotone convergence theorem of sets (Theorem 1.24), we have that $\mu(\{f_n \geq a_k/2\} \cap A_k) \uparrow \mu(A_k) = \infty$ as $n \to \infty$. This implies that $\mu(f_n) \uparrow \infty = \mu(f)$ as $n \to \infty$. Therefore, it suffices to prove for g with finite support S, *i.e.*, $\mu(A_k) < \infty$ for all $1 \leq k \leq m$[1]. Fix an arbitrary $\epsilon > 0$. By Egorov's theorem (Theorem 3.1), there exists $A_\epsilon \subseteq S$, such that $\mu(S \setminus A_\epsilon) < \epsilon$ and $f_n \to f$ uniformly on A_ϵ. Now fix an arbitrary $\eta > 0$, then there exists $N \in \mathbb{N}$, such that $|f_n - f| < \eta$ on A_ϵ for all $n \geq N$. In other words, $f_n > f - \eta \geq g - \eta$ on A_ϵ. But $g > g - \eta$ as well, so we can deduce that

$$f_n \wedge g \geq \underbrace{f_n \mathbb{1}_{A_\epsilon} \wedge g \mathbb{1}_{A_\epsilon}}_{f_n, g \geq 0} > \underbrace{(g-\eta)}_{\text{observation}} \mathbb{1}_{A_\epsilon} \geq \underbrace{g\mathbb{1}_{A_\epsilon} - \eta}_{\eta<0} = g - \eta - g\mathbb{1}_{S\setminus A_\epsilon}. \tag{73}$$

Let $m = \max_{1\leq k\leq m} a_k$, then we can make use of the fact that $(g - \eta - g\mathbb{1}_{S\setminus A_\epsilon})^+$ is a simple function. Indeed, by taking $\eta = \delta/(\mu(S) + m)$, we can see that

$$\mu(f_n \wedge g) > \mu(g) - \eta\mu(S) - m\mu(S \setminus A_\epsilon) \geq \text{THEN WHAT?} \tag{74}$$

□

9/25 Lecture

Lemma 3.3 (Fatou's lemma for functions)**.** Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions, then $\mu(\liminf f_n) \leq \liminf \mu(f_n)$.

Proof. Define $g_n := \inf_{k\geq n} f_k$, then clearly $g_n \uparrow \lim_{n\to\infty} \inf_{k\geq n} f_k = \liminf f_n$, and thus $\mu(g_n) \uparrow \mu(\liminf f_n)$ by the monotone convergence theorem (Theorem 3.2). However, we have that $\mu(g_n) \leq \inf_{k\geq n} \mu(f_k)$ for all $n \in \mathbb{N}$, because $g_n \leq f_k$ for all $n \in \mathbb{N}$ and $k \geq n$. Therefore, we can conclude that

$$\mu(\liminf f_n) = \lim_{n\to\infty} \mu(g_n) \leq \lim_{n\to} \inf_{k\geq n} \mu(f_k) = \liminf \mu(f_n), \tag{75}$$

so the proof is complete. □

[1]A function with finite support means that the function vanishes outside a set with finite measure. Particularly in this case, we are only interested in functions such that all A_k are of finite measure. Since the number of A_k's is finite, g can take non-zero values only on a set of finite measure. Therefore we consider functions with finite support S.

Lemma 3.4 (Reverse Fatou lemma for functions)**.** Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions. Assume there exists a measurable function $g \geq 0$ such that $\mu(g) < \infty$ and $f_n \leq g$ for all $n \in \mathbb{N}$. Then $\mu(\limsup f_n) \geq \limsup \mu(f_n)$.

Proof. This is trivial by applying Fatou's lemma (Lemma 3.3) for functions to $\{g - f_n\}_{n\in\mathbb{N}}$. □

Lemma 3.5. For all integrable (*resp.* non-negative measurable) functions f, g and all constants $\alpha, \beta \in \mathbb{R}$ (*resp.* ≥ 0), the following properties hold.

(1) $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$.

(2) If $f \leq g$, then $\mu(f) \leq \mu(g)$.

(3) If $f = 0$ almost everywhere, then $\mu(f) = 0$.

If f is non-negative measurable, then the third property is an equivalence.

Proof. (1) Start with the case where $\alpha, \beta \geq 0$. Note that the property holds for all simple functions as have already been discussed. Here we employ a technique called the *Standard Machine*, also used in other contexts, in particular in exercises. Assume that $f, g \geq 0$, and define

$$f_n := \left(2^{-n}\lfloor 2^n f\rfloor\right) \wedge n, \qquad g_n := \left(2^{-n}\lfloor 2^n g\rfloor\right) \wedge n. \tag{76}$$

Then $f_n \uparrow f$ and $g_n \uparrow g$, and thus by the monotone convergence theorem (Theorem 3.2), we have that $\mu(f_n) \uparrow \mu(f)$ and $\mu(g_n) \uparrow \mu(g)$. This implies that $\alpha\mu(f_n) + \beta\mu(g_n) \uparrow \alpha\mu(f) + \beta\mu(g)$. But $\alpha f_n + \beta g_n \uparrow \alpha f + \beta g$, which means that $\alpha\mu(f_n) + \beta\mu(g_n) = \mu(\alpha f_n + \beta g_n) \uparrow \mu(\alpha f + \beta g)$, also by the monotone convergence theorem (Theorem 3.2). This proves that $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$. Now we extend this to general integrable functions. With $\alpha, \beta \geq 0$, we can observe that

$$\begin{aligned}\mu(\alpha f + \beta g) &= \mu((\alpha f + \beta g)^+ - (\alpha f + \beta g)^-) = \mu(\underbrace{(\alpha f^+ + \beta g^+)}_{\text{integrable}, \geq 0} - \underbrace{(\alpha f^- + \beta g^-)}_{\text{integrable}, \geq 0}) \\ &= \mu(\alpha f^+ + \beta g^+) - \mu(\alpha f^- + \beta g^-) = \alpha\mu(f^+) + \beta\mu(g^+) - \alpha\mu(f^-) - \beta\mu(g^-) \\ &= \alpha(\mu(f^+) - \mu(f^-)) + \beta(\mu(g^+) - \mu(g^-)) = \alpha\mu(f^+ - f^-) + \beta\mu(g^+ - g^-) = \alpha\mu(f) + \beta\mu(g).\end{aligned} \tag{77}$$

Now without the assumption that $\alpha, \beta \geq 0$, note that we can flip the sign of both α and f if $\alpha < 0$ (same for β and g). Then we observe that $\mu(-f) = -\mu(f)$ by the previous proof, so we can conclude this property now.

(2) To be done...

(3) To be done...

□

Remark 3.6. The second property in the above lemma implies that $|\mu(f)| \leq \mu(|f|)$.

Example 3.7 (Inclusion-exclusion formula)**.** First remark that, for all $A \in \mathscr{E}$, we have that $\mu(A) = \mu(\mathbb{1}_A)$. We apply this observation to prove the inclusion-exclusion formula without utilizing mathematical induction. We can write that

$$\mathbb{1}_{\bigcup_{j=1}^n A_j} = \underbrace{1 - \underbrace{\prod_{j=1}^n \underbrace{(1 - \mathbb{1}_{A_j})}_{0 \text{ on } A_j}}_{0 \text{ on any } A_j}}_{1 \text{ on any } A_j} = \sum_j \mathbb{1}_{A_j} - \sum_{j_1<j_2} \mathbb{1}_{A_{j_1}}\mathbb{1}_{A_{j_2}} + \ldots + (-1)^{n+1} \sum_{j_1<\ldots<j_n} \mathbb{1}_{A_{j_1}} \ldots \mathbb{1}_{A_{j_n}}. \tag{78}$$

Integrating both sides of the equation, we can thus see that

$$\begin{aligned}\mu\left(\bigcup_{j=1}^n A_j\right) &= \mu\left(\mathbb{1}_{\bigcup_{j=1}^n A_j}\right) = \sum_j \mu\left(\mathbb{1}_{A_j}\right) - \sum_{j_1<j_2} \mu\left(\mathbb{1}_{A_{j_1}}\mathbb{1}_{A_{j_2}}\right) + \ldots + (-1)^{n+1} \sum_{j_1<\ldots<j_n} \mu\left(\mathbb{1}_{A_{j_1}} \ldots \mathbb{1}_{A_{j_n}}\right) \\ &= \sum_j \mu(A_j) - \sum_{j_1<j_2} \mu(A_{j_1} \cap A_{j_2}) + \ldots + (-1)^{n+1} \sum_{j_1<\ldots<j_n} \mu(A_{j_1} \cap \ldots \cap A_{j_n}).\end{aligned} \tag{79}$$

9/27 Lecture

Theorem 3.8 (Dominated convergence theorem (DCT))**.** Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of integrable functions converging almost surely to a function f as $n\to\infty$. Suppose that $|f_n|\le g$ for all n for some integrable function g, then f is integrable and $\mu(f_n)\to\mu(f)$ as $n\to\infty$.

Proof. Let $h_n^{\pm} := g\pm f_n$ and note that both functions are non-negative. By Fatou's lemma (Lemma 3.3), we can see that $\mu(\liminf h_n^{\pm})\le\liminf\mu(h_n^{\pm})$. Therefore, we have that

$$\mu(g)+\mu(f)=\mu(g)+\mu(\liminf(f_n))=\mu(\liminf(g+f_n))\le\liminf\mu(g+f_n)=\mu(g)+\liminf\mu(f_n), \tag{80}$$

$$\mu(g)-\mu(f)=\mu(g)-\mu(\limsup(f_n))=\mu(\liminf(g-f_n))\le\liminf\mu(g-f_n)=\mu(g)-\liminf\mu(f_n). \tag{81}$$

The first inequality gives $\mu(f)\le\liminf\mu(f_n)$ and the second inequality gives $\mu(f)\ge\liminf\mu(f_n)$. Therefore, we can conclude that $\mu(f_n)\to\mu(f)$ as $n\to\infty$. Now since g is measurable, $|g|<\infty$ and thus $|f_n|<\infty$ and $|f|<\infty$, so clearly f is integrable. The proof is thus complete. □

Lemma 3.9 (Scheffé's lemma)**.** Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative integrable functions, and suppose that $f_n\to f$ almost everywhere where f is non-negative and integrable, then $\mu(|f_n-f|)\to 0$ if and only if $\mu(f_n)\to\mu(f)$.

Proof. $\Longrightarrow$ Assume that $\mu(|f_n-f|)\to 0$, then we have that $\mu(f_n)-\mu(f)=\mu(f_n-f)\le\mu(|f_n-f|)\to 0$. This means that $\mu(f_n)-\mu(f)\to 0$, so $\mu(f_n)\to\mu(f)$.

$\Longleftarrow$ Assume that $\mu(f_n)\to\mu(f)$. By the triangle inequality, we have that $|f_n-f|\le|f_n|+|f|=f_n+f$. Hence by Fatou's lemma (Lemma 3.3), we have that

$$\mu(\liminf(f_n+f-|f_n-f|))\le\liminf\mu(f_n+f-|f_n-f|)=\liminf\mu(f_n)+\mu(f)-\liminf\mu(|f_n-f|). \tag{82}$$

Since $f_n\to f$ almost everywhere, we can see that the left-hand side is equal to $2\mu(f)$. Also, we have assumed that $\mu(f_n)\to\mu(f)$, the right-hand side is equal to $2\mu(f)-\liminf\mu(|f_n-f|)$. Hence by the inequality, we can see that $\liminf\mu(|f_n-f|)\le 0$. Then necessarily $\mu(|f_n-f|)\to 0$ as $n\to\infty$. □

Theorem 3.10 (Series convergence theorem (SCT))**.** Let $\{g_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions. Then $\sum_{n=1}^{\infty}\mu(g_n)=\mu\left(\sum_{n=1}^{\infty}g_n\right)$.

Proof. Let $h_N:=\sum_{n=1}^{N}h_n$, then h_N is non-negative and measurable, with $h_N\uparrow h_\infty=\sum_{n=1}^{\infty}g_n$. By the monotone convergence theorem (Theorem 3.2), we have that $\mu(h_N)\uparrow\mu(h_\infty)$. This means that

$$\sum_{n=1}^{\infty}\mu(g_n)=\lim_{N\to\infty}\sum_{n=1}^{N}\mu(g_n)=\lim_{N\to\infty}\mu\left(\sum_{n=1}^{N}g_n\right)=\lim_{N\to\infty}\mu(h_N)=\mu(h_\infty)=\mu\left(\sum_{n=1}^{\infty}g_n\right), \tag{83}$$

so the proof is complete. □

Example 3.11. Consider the measurable space $(\mathbb{N},\mathscr{P}(\mathbb{N}))$.

(1) Let $\mu(A):=|A|$ (cardinality of the set A) be the counting measure on $\mathbb{N}$. Then for all $f:\mathbb{N}\to\mathbb{R}$, f is integrable if and only if $\mu(|f|)=\sum_{n=1}^{\infty}|f(n)|<\infty$ (absolute convergence of the series), and $\mu(f)=\sum_{n=1}^{\infty}f(n)$.

(2) In the case of a general measure μ and any $f:\mathbb{N}\to\mathbb{R}$, we alternatively have that f is integrable if and only if $\mu(|f|)=\sum_{n=1}^{\infty}\mu(\{n\})|f(n)|<\infty$, and $\mu(f)=\sum_{n=1}^{\infty}\mu(\{n\})f(n)$.

3.2 Image Measure and Probability Density Function

The following proposition provides a very useful property of the image measure. It can be applied to find the expected value of a function of a random variable X, when we know the distribution (or law) of X, *i.e.*, the image measure $\mu_X:=\mathbb{P}\circ X^{-1}$ of μ by X. More precisely, let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space, let $X:\Omega\to H$ be a random variable, where $(H,\mathscr{H})$ is a measurable space, and let $g:H\to\mathbb{R}_+$ be a measurable function. Then, the following proposition would imply that

$$\mathbb{E}(g(X))=\mu_X(g)=\int_{x\in E}g(x)\mu_X(dx). \tag{84}$$

Proposition 3.12. Let $(E, \mathscr{E}, \mu)$ and $(G, \mathscr{G}, \nu)$ be measure spaces. Suppose that $\nu = \mu \circ f^{-1}$ for some measurable function $f : E \to G$, *i.e.*, that ν is the image measure of μ by f. Then $\nu(g) = \mu(g \circ f)$ for any non-negative measurable or integrable function g on G.

Proof. Let us use again the *Standard Machine* technique.

- **Case 1.** Let $g := \mathbb{1}_A$ for some $A \in \mathscr{E}$. Then $\nu(g) = \nu(A) = \mu(f^{-1}(A)) = \mu(\mathbb{1}_{f^{-1}(A)}) = \mu(\mathbb{1}_A \circ f)$, because $(\mathbb{1}_A \circ f)(x) = \mathbb{1}_A(f(x)) = 1$ if and only if $f(x) \in A$, *i.e.*, $x \in f^{-1}(A)$.
- **Case 2.** By linearity and Case 1, $\nu(g) = \mu(g \circ f)$ for any simple function g.
- **Case 3.** Suppose now that g is a non-negative measurable function. Define a sequence of simple functions by $g_n := (2^{-n} \lfloor 2^n g \rfloor) \wedge n$, then clearly $g_n \uparrow g$ and $g_n \circ f \uparrow g \circ f$, so that $\nu(g_n) \uparrow \nu(g)$ and $\nu(g_n \circ f) \uparrow \nu(g \circ f)$ by the monotone convergence theorem (Theorem 3.2). Therefore, $\nu(g) = \mu(g \circ f)$ by convergence and Case 2.
- **Case 4.** Now suppose g is integrable and write $g = g^+ - g^-$. Then by Case 3, we have that $\nu(g^+) = \mu(g^+ \circ f) < \infty$ and $\nu(g^-) = \mu(g^- \circ f) < \infty$. Therefore, we can conclude by linearity that $\nu(g) = \mu(g \circ f)$. $\square$

Proposition 3.13. Let $(E, \mathscr{E}, \mu)$ be a measure space, and let f be a non-negative measurable function with $\mu(f) < \infty$. Define $(f\mu)(A) := \mu(f \mathbb{1}_A)$ for $A \in \mathscr{E}$, then ν is a finite measure on E and $(f\mu)(g) = \mu(fg)$ for all non-negative measurable functions g on E.

Proof. This proof will be done in Homework 3 with the same *Standard Machine* technique. $\square$

Remark 3.14. If λ denotes the measure $f\mu$ on $(E, \mathscr{E})$, we say that λ has **density** f relative to μ, and express this symbol via $d\lambda/d\mu = f$. In this case, for any $F \in \mathscr{E}$ such that $\mu(F) = 0$, we have that $\lambda(F) = 0$ so that only certain measures have density relative to μ. The next theorem will show a characterization of such measures.

Theorem 3.15 (Radon-Nikodým)**.** If λ and μ are σ-finite measures on $(E, \mathscr{E})$ such that $\lambda(F) = 0$ whenever $\mu(F) = 0$ for $F \in \mathscr{E}$, then $\lambda = f\mu$ for some measurable non-negative function f on E, which we denote $f = d\lambda/d\mu$ called the **Radon Nikodým derivative**.

Proof. See *Rudin, Real and Complex Analysis, McGraw-Hill, Chapter 6.* $\square$

Remark 3.16. A random variable X in $\mathbb{R}^n$ is said to have **probability density function** f if, for all $A \in \mathscr{B}(\mathbb{R}^n)$, we have that

$$\mathbb{P}(X \in A) = \int_{x \in A} f(x)dx. \tag{85}$$

In other words, the image measure (or law) of X, $\mu_X := \mathbb{P} \circ X^{-1}$, is $f\mathscr{L}$ where $\mathscr{L}$ is the Lebesgue measure on $\mathbb{R}^n$. Moreover by the previous propositions, for a random variable X in $\mathbb{R}^n$ with density f_X, we have that

$$\mathbb{E}(g(X)) = (f_X \mathscr{L})(g) = \int_{\mathbb{R}^n} g(x) f_X(x) dx, \tag{86}$$

where the first equality holds by Proposition 3.12 that $\mathbb{E}(g(X)) = \mathbb{P}(g \circ X) = \mu_X(g) = (f_X\mathscr{L})(g)$, and the second equality holds by Proposition 3.13 that $(f_X\mathscr{L})(g) = \mathscr{L}(f_X g) = \int_{\mathbb{R}^n} g(x) f_X(x) dx$. In general, random variables on $\mathbb{R}^n$ are indeed given by their density function if it exists, but in dimension 1 ($n = 1$), it can happen that the distribution function is provided instead (note that the distribution function is always defined then). Now here is how one can find the density function of a random variable with certain distribution function, if this density exists: if X is a random variable taking values in $\mathbb{R}$ with distribution function F_X and density f_X, then

$$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(t) dt, \tag{87}$$

and thus by a generalization of the fundamental theorem of Calculus, we have that $F_X'(x) = f_X(x)$ for almost all $x \in \mathbb{R}$.

Remark 3.17. Note that the differentiation under the integral sign can be carried out in the setting of general measures, as can be found in *Probability with Martingales, David Williams, Appendix A16.*

10/9 Lecture

3.3 Inequalities, L^p-Norms, and L^p-Spaces

3.3.1 Chebyshev's and Jensen's Inequalities

Theorem 3.18 (Chebyshev's inequality)**.** Let $(E, \mathscr{E}, \mu)$ be a measure space, $f \geq 0$ be measurable, and $\lambda > 0$, then

$$\mu(f \geq \lambda) \leq \frac{\mu(f)}{\lambda}. \tag{88}$$

Proof. Recall that $\{f \leq \lambda\} := \{x \in E;\ f(x) \geq \lambda\}$. Since $\lambda \mathbb{1}_{f \geq \lambda} \leq f$, we have that $\mu(\lambda \mathbb{1}_{f \geq \lambda}) \leq \mu(f)$. Therefore, we can conclude that $\mu(f \geq \lambda) = \mu(\mathbb{1}_{f \geq \lambda}) \leq \mu(f)/\lambda$, and the proof is thus complete. □

Remark 3.19. The Chebyshev's inequality is trivial, but has many applications. Indeed, let g be a measurable function (for instance, a random variable on a probability space), and let $\phi : \mathbb{R} \to [0, \infty]$ be a non-decreasing measurable function. Then for all $\lambda \geq 0$, we have that

$$\mu(g \geq \lambda) \leq \mu(\phi(g) \geq \phi(\lambda)) \leq (\phi(\lambda))^{-1} \mu(\phi(g)). \tag{89}$$

If $\mu(\phi(g)) < \infty$, this result enables us to obtain a "tail estimate", *i.e.*, an upper bound, of the measure of $\{g \geq \lambda\}$, by a term of the order $(\phi(\lambda))^{-1}$. These types of tail estimates will be required in some exercieses. For instance, given a sequence of independent identically distributed random variables $\{X_n\}_{n \in \mathbb{N}}$, in order to find the sequence a_n such that $\limsup_{n \to \infty} X_n / a_n = 1$, TO BE DONE... Also, given $\lambda \geq 0$ and a random variable Y, a way to find a "good" upper bound for $\mathbb{P}(Y \geq \lambda)$ is to choose the optimum θ for λ in $\mathbb{P}(Y \geq \lambda) \leq \exp(-\theta\lambda)\mathbb{E}(\exp(\theta Y)) = \exp(-\theta\lambda) M_Y(\theta)$. OPTIMIZE FOR Y A GAUSSIAN RANDOM VARIABLE.

Lemma 3.20. Let $c : \mathbb{R} \to \mathbb{R}$ be convex and $m \in \mathbb{R}$. Then there exists $a, b \in \mathbb{R}$, not necessarily unique, such that $c(x) \geq ax + b$ with equality at $x = m$.

Proof. By convexity, for $x < m < y$, we have that

$$\frac{c(m) - c(x)}{m - x} \leq \frac{c(y) - c(m)}{y - m}. \tag{90}$$

Choose $a \in \mathbb{R}$ such that

$$\sup_{x < m} \left(\frac{c(m) - c(x)}{m - x} \right) \leq a \leq \inf_{y > m} \left(\frac{c(y) - c(m)}{y - m} \right) = \inf_{x > m} \left(\frac{c(x) - c(m)}{x - m} \right). \tag{91}$$

From the first inequality we have that $c(x) \geq a(x - m) + c(m)$ for all $x < m$. From the second inequality we have that $c(x) \geq a(x - m) + c(m)$ for all $x > m$. Taking $b = c(m) - am$, we have that $c(x) \geq ax + b$ for all $x \neq m$, and note that $c(m) = am + b$ so that $c(x) = ax + b$ at $x = m$. The proof is thus complete. □

Theorem 3.21 (Jensen's inequality)**.** Let X be an integrable random variable and let $c : \mathbb{R} \to \mathbb{R}$ be a convex function. Then $c(X)^-$ is integrable, and $\mathbb{E}(c(X)) \geq c(\mathbb{E}(X))$.

Proof. Let $m = \mathbb{E}(X)$ and choose $a, b \in \mathbb{R}$ as in Lemma 3.20 above. Then we have that $c(X) \geq aX + b$ with equality at $X = \mathbb{E}(X)$. In particular, we have that

$$\mathbb{E}(c(X)^-) = \mathbb{E}(\max(-c(X), 0)) \leq \mathbb{E}(\max(-aX - b, 0)) \leq \mathbb{E}(|a||X| + |b|) = |a|\mathbb{E}(|X|) + |b| < \infty. \tag{92}$$

Hence $\mathbb{E}(c(X))$ is well-defined, though possibly infinite because $\mathbb{E}(c(X)^+)$ may be infinite. Moreover, we can see that

$$\mathbb{E}(c(X)) \geq \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = am + b = c(m) = c(\mathbb{E}(X)), \tag{93}$$

and the proof is thus complete. □

Remark 3.22. If $c : [0,\infty) \to \mathbb{R}$ is strictly increasing and $X \geq 0$, then $c(X)$ being integrable means that X is integrable and Jensen's inequality (Theorem 3.21) holds. Indeed, we can deduce from the proof of Lemma 3.20 that, for all $m > 0$, there exists $a > 0$ and $b \in \mathbb{R}$, such that $c(x) \geq ax + b$ for all $x \geq 0$ with equality at $x = m$. Indeed since c is strictly increasing, the leftmost term in (91) would be positive, and thus $a > 0$. In this case, $\mathbb{E}(X) < \infty$ if $\mathbb{E}(c(X)) < \infty$, because $\mathbb{E}(c(X)) \geq a\mathbb{E}(X) + b$ with $a > 0$.

Example 3.23. Let $\Omega = \mathbb{N}$, $\mathscr{F} = \mathscr{P}(\mathbb{N})$, and $\mathbb{P}(\{i\}) = \alpha_i$, $i \in \mathbb{N}$, with $\sum_i \alpha_i = 1$. Moreover, let $\{x_i\}_{i\in\mathbb{N}}$ be such that $\sum_i \alpha_i |x_i| < \infty$ and define $c : \mathbb{R} \to \mathbb{R}$ with $c(x) := \exp(x)$. Note that c is convex and so Jensen's inequality (Theorem 3.21) implies that

$$\prod_{j\in\mathbb{N}} \exp(\alpha_j x_j) = \exp\left(\sum_{j\in\mathbb{N}} \alpha_j x_j\right) = \exp(\mathbb{E}(X)) \leq \mathbb{E}(\exp(X)) = \sum_{j\in\mathbb{N}} \alpha_j \exp(x_j). \tag{94}$$

By substituting y_j for $\exp(x_j)$, we can obtain the following version of the **arithmetic-geometric mean inequality**, such that

$$\prod_{j\in\mathbb{N}} y_j^{\alpha_j} \leq \sum_{j\in\mathbb{N}} \alpha_j y_j, \tag{95}$$

where we define, for all $\{z_i\}_{i\in\mathbb{N}} \subseteq \mathbb{R}_+ \setminus \{0\}$, the infinite product of z_i by

$$\prod_{i\in\mathbb{N}} z_i = \exp\left(\sum_{i\in\mathbb{N}} \log z_i\right), \tag{96}$$

as long as $\sum_i (\log z_i)^- < \infty$ with x^- denoting the negative part of x. In particular, if we let $\alpha_i = 1/n$ for $1 \leq i \leq n$ and $\alpha_i = 0$ otherwise, we can obtain the classical inequality

$$\sqrt[n]{y_1 \ldots y_n} = \prod_{j=1}^{n} y_j^{1/n} \leq \sum_{j=1}^{n} y_j = y_1 + \ldots + y_n \tag{97}$$

10/11 Lecture

3.3.2 L^p-Norms, Hölder's Inequality, and Minkowski's Inequality

Let $(E, \mathscr{E}, \mu)$ be a measure space. For $1 \leq p < \infty$, let $L^p = L^p(E, \mathscr{E}, \mu)$ be the set of measurable functions f with finite L^p-norm, such that

$$\|f\|_p = \left(\int_E |f|^p d\mu\right)^{1/p} < \infty. \tag{98}$$

Let $L^\infty = L^\infty(E, \mathscr{E}, \mu)$ be the set of measurable functions with finite L^∞-norm, such that

$$\|f\|_\infty = \inf\{\lambda;\ |f| \leq \lambda \text{ almost everywhere}\} < \infty. \tag{99}$$

Note that if μ is a probability measure, then $\|f\|_p \leq \|f\|_\infty$ for all $1 \leq p < \infty$. Moreover, for $1 \leq p \leq \infty$, we say that f_n converges to f in L^p if $\|f_n - f\|_p \to 0$. We say that $p, q \in [1,\infty]$ are **conjugate indices** if $1/p + 1/q = 1$.

Theorem 3.24. Let p and q be conjugate indices, and let f and g be measurable functions. Then if $p > 1$, we have that $\mu(|fg|) \leq \|f\|_p \|g\|_q$, which is known as the **Hölder's inequality**, and implies in particular, that fg is integrable if $f \in L^p$ and $g \in L^q$. Also, $\|f+g\|_p \leq \|f\|_p + \|g\|_q$, which is known as the **Minkowski's inequality**, and implies in particular, that L^p is a vector space.

Proof. SEE *Williams p. 70*, TO BE DONE... ☐

Remark 3.25. Note that $\|\cdot\|_p$ is not a norm, since $\|f\|_p = 0$ does not necessarily imply that $f = 0$. We define the following equivalence relation on L^p, in order to quotient it into a normed vector space. Define $f \sim g$ if $f = g$ almost surely, and we write $\dot{f}$ for the equivalence class of f. Let $\mathscr{L}^p := \left\{\dot{f};\ f \in L^p\right\}$, then $\|\cdot\|_p$ is well-defined on $\mathscr{L}^p$ and is a norm.

Remark 3.26. Note that, for $f \in L^2$, we have $\|f\|_2^2 = \langle f, f \rangle$, where $\langle \cdot, \cdot \rangle$ is the symmetric bilinear form on L^2 given by

$$\langle f, g \rangle := \int_E fg d\mu, \qquad f, g \in L^2. \tag{100}$$

$\langle \cdot, \cdot \rangle$ is an inner product on $\mathscr{L}^2$, *i.e.*, $\left\langle \dot{f}, \dot{f} \right\rangle \geq 0$ if $\dot{f} \in \mathscr{L}^2$, with equality if and only if $\dot{f} = 0$. We then claim that $\langle \cdot, \cdot \rangle$ is well-defined on $\mathscr{L}^2$. Indeed, the integral makes sense because of Hölder's inequality (with $p = q = 2$, known as **Cauchy-Schwarz inequality**), and its value does not change over functions of the same equivalence class.

Remark 3.27. Observe that Jensen's inequality, using Remark 3.22, implies the monotonicity of the L^p-norms with respect to a probability measure. Indeed, letting $1 \leq p < q < \infty$ and setting $c(x) := x^{q/p}$, clearly c is convex and thus for any $X \in L^q(\Omega, \mathscr{F}, \mathbb{P})$, we have that

$$\|X\|_p = (\mathbb{E}(|X|^p))^{1/p} = (c(\mathbb{E}(|X|^p)))^{1/q} \leq (\mathbb{E}(c(|X|^p)))^{1/q} = (\mathbb{E}(|X|^q))^{1/q} = \|X\|_q, \tag{101}$$

and in particular, this implies that $L^p(\Omega, \mathscr{F}, \mathbb{P}) \supseteq L^q(\Omega, \mathscr{F}, \mathbb{P})$.

3.3.3 Completeness in $\mathscr{L}^p$ and the Orthogonal Projection in $\mathscr{L}^2$

Recall that a normed vector space V is complete if every Cauchy sequecence in V converges, *i.e.*, given any sequence $\{v_n\}_{n\in\mathbb{N}}$ in V such that $\|v_n - v_m\| \to 0$ as $m, n \to \infty$, there exists $v \in V$ such that $\|v_n - v\| \to 0$ as $n \to \infty$. A complete normed vector space is called a **Banach space**, and a complete inner product space is called a **Hilbert space**. The Banach spaces $\mathscr{L}^p$ are the spaces of interest here.

Theorem 3.28. We have that $\mathscr{L}^p$ is a Banach space for $1 \leq p \leq \infty$, and L^2 is moreover a Hilbert space.

Remark 3.29. Consider the Hilbert space $\mathscr{L}^2(E, \mathscr{E}, \mu)$. Let $f, g \in \mathscr{L}^2(E, \mathscr{E}, \mu)$, we say that f and g are **orthogonal**, denoted by $f \perp g$ if $\langle f, g \rangle = 0$. For any subset $V \subseteq \mathscr{L}^2$, we define

$$V^\perp := \left\{ f \in \mathscr{L}^2;\ \langle f, v \rangle = 0 \text{ for all } v \in V \right\}. \tag{102}$$

A subset $V \subseteq \mathscr{L}^2$ is said to be **closed** if, for every sequence $\{f_n\}_{n\in\mathbb{N}}$ in V with $f_n \to f$ in $\mathscr{L}^2$, we have $f = v$ almost everywhere for some $v \in V$. Let use observe the two following simple but important identities on $\mathscr{L}^2$, the **Pythagoras' rule**

$$\|f + g\|_2^2 = \|f\|_2^2 + 2\langle f, g \rangle + \|g\|_2^2, \tag{103}$$

and the **parallelogram law**

$$\|f + g\|_2^2 + \|f - g\|_2^2 = 2\left(\|f\|_2^2 + \|g\|_2^2\right). \tag{104}$$

Theorem 3.30 (Riesz, orthogonal projection)**.** Let V be a closed subspace of $\mathscr{L}^2$, then each $f \in \mathscr{L}$ has a decomposition $f = v + u$ with $v \in V$ and $u \in V^\perp$. Moreover, $\|f - v\|_2 \leq \|f - g\|_2$ for all $g \in V$, with equality if and only if $g = v$ almost everywhere. The function v is then called (a version of) the **orthogonal projection** of f on V.

Proof. Choose a sequence $g_n \in V$, such that $\|f - g_n\|_2 \to d(f, V) = \inf\{\|f - g\|_2;\ g \in V\}$. Then by the parallelogram law, we can see that

$$\left\| 2\left(f - \frac{g_n + g_m}{2}\right) \right\|_2^2 + \|g_n - g_m\|_2^2 = 2\left(\|f - g_n\|_2^2 + \|f - g_m\|_2^2\right). \tag{105}$$

But the first summand in the left-hand side is at least $4d(f, V)^2$, so we must have that $\|g_n - g_m\|_2 \to 0$ as $n, m \to \infty$ since the right-hand side approaches $4d(f, V)^2$ as $n, m \to \infty$. By completeness, $\|g_n - g\|_2 \to 0$ as $n \to \infty$ for some $g \in \mathscr{L}^2$. By closedness, $g = v$ almost everywhere for some $v \in V$. Hence, we now have that

$$\|f - v\|_2 = \lim_{n\to\infty} \|f - g_n\|_2 = d(f, V). \tag{106}$$

Now, for any $h \in V$ and $t \in \mathbb{R}$, we have that

$$d(f, V)^2 \leq \|f - (v + th)\|_2^2 = \|f - v\|_2^2 - 2t\langle f - v, h \rangle + t^2\|h\|_2^2 = d(f, V)^2 - 2t\langle f - v, h \rangle + t^2\|h\|_2^2, \tag{107}$$

and so $\langle f - v, h \rangle = 0$. Hence $u = f - v \in V^\perp$ as required. THe proof is now complete. ∎

Remark 3.31. The above theorem, which enables us to construct the orthogonal projection (modulo the equivalence class by almost sure equality defined in Remark 3.25) of an integrable function in $\mathscr{L}^2$ on a closed subspace V as a function in V that has minimal distance to it, will be a key tool to build up the notion of conditional expectation of a random variable. Namely, the conditional expectation of a random variable X on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with respect to a sub-σ-algebra $\mathscr{G}$ will be the orthogonal projection of X on the space of square integrable functions that are measurable with respect to $\mathscr{G}$.

10/16 Lecture

4 Conditional Expectation and Martingales

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability triple, *i.e.*, a measure space for which $\mathbb{P}(\Omega) = 1$.

4.1 Conditional Expectation

Suppose X and Z are random variables taking finitely many values, $x_1, \ldots, x_n$ and $z_1, \ldots, z_n$, respectively. In introductory courses to probability, we defined the conditional probability

$$\mathbb{P}(X = x_i \mid Z = z_j) := \frac{\mathbb{P}(X = x_i,\ Z = z_j)}{\mathbb{P}(Z = z_j)}, \tag{108}$$

and the conditional expectation was given by

$$\mathbb{E}(X \mid Z = z_j) := \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i \mid Z = z_j). \tag{109}$$

Using this, we can define a random variable, "the expectation of X given Z", which will map every ω in the probability space Ω to a number yielding the expectation of $X(\omega)$ given the information about the value of $Z(\omega)$, which we either call $\mathbb{E}(X \mid \sigma(Z))$ or $\mathbb{E}(X \mid Z)$, such that

$$\mathbb{E}(X \mid Z) := \sum_{j=1}^{m} \mathbb{E}(X \mid Z = z_j) \mathbb{1}_{\{Z = z_j\}}. \tag{110}$$

Now a more general question is, given a sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$, does it make sense to define $\mathbb{E}(X \mid \mathscr{G})$, *i.e.*, the expected value of X given the information $\mathscr{G}$? We want to define the random variable that maps every ω to the expected value $Y(\omega) = \mathbb{E}(X \mid \mathscr{G})(\omega)$ given this information. Given a single event $A \in \mathscr{F}$ with $\mathbb{P}(A) \neq 0$, we know that the conditional expectation of X knowing that the outcome $\omega \in A$ is

$$\mathbb{E}(X \mid \omega \in A) = \sum_{i=1}^{n} \frac{x_i \mathbb{P}(X = x_i,\ \omega \in A)}{\mathbb{P}(\omega \in A)} = \frac{\sum_{i=1}^{n} x_i \mathbb{P}(X \mathbb{1}_A = x_i)}{\mathbb{P}(A)} = \frac{\mathbb{E}(X \mathbb{1}_A)}{\mathbb{P}(A)}. \tag{111}$$

We would like our random variable to take into account all such information. Firstly, it has to be $\mathscr{G}$-measurable, since its value at a particular ω has to be a (possibly complicated) function of the values of (possibly infinitely many) $\mathscr{G}$-measurable functions at ω, which are the information you provide about ω. Secondly, conditionally on the information that ω belongs to a particular $G \in \mathscr{G}$ (with $\mathbb{P}(G) \neq 0$), the expectations of Y and X should not differ, *i.e.*, $\mathbb{E}(X \mid \omega \in G) = \mathbb{E}(Y \mid \omega \in G)$. This is equivalent to $\mathbb{E}(X \mathbb{1}_G) = \mathbb{E}(Y \mathbb{1}_G)$. The following theorem will imply that requiring the last equality to hold for all $G \in \mathscr{G}$ almost uniquely defines the random variable $Y = \mathbb{E}(X \mid \mathscr{G})$ if $X \in \mathscr{L}^1$.

Theorem 4.1. Let X be a random variable for which $\mathbb{E}(|X|) < \infty$, and let $\mathscr{G}$ be a sub-σ-algebra of $\mathscr{F}$. Then there exists a random variable Y, such that

(1) Y is $\mathscr{G}$-measurable,

(2) $\mathbb{E}(|Y|) < \infty$, and

(3) $\mathbb{E}(Y \mathbb{1}_G) = \mathbb{E}(X \mathbb{1}_G)$ for all $G \in \mathscr{G}$ (or equivalently, for all $G \in \mathscr{A}$ π-system with $\mathscr{G} = \pi(\mathscr{A})$).

Moreover, if $\tilde{Y}$ is another random variable with these properties, then $\tilde{Y} = Y$ almost surely. A random variable with such properties is called **a version of the conditional expectation of X given $\mathscr{G}$**, denoted as $\mathbb{E}(X \mid \mathscr{G})$, and we write $Y = \mathbb{E}(X \mid \mathscr{G})$ almost surely.

Remark 4.2. Given random variables $Z, Z_1, \ldots, Z_n$, by convention we often write $\mathbb{E}(X \mid Z)$ for $\mathbb{E}(X \mid \sigma(Z))$, and $\mathbb{E}(X \mid Z_1, \ldots, Z_n)$ for $\mathbb{E}(X \mid \sigma(Z_1), \ldots \sigma(Z_n))$.

Remark 4.3. Let us check that, if X and Z are random variables taking m and n values respectively, and if Y is a version of the conditional expectation $\mathbb{E}(X \mid Z)$, then the conditional expectation given after Theorem 4.1 provides almost surely the same random variable as the one we defined at the beginning of the section. Indeed, for all $j \in \{1, \ldots, n\}$, we have that

$$\begin{aligned}\mathbb{E}(Y\mathbb{1}_{\{Z=z_j\}}) &= \mathbb{E}(X \mid Z = z_j)\mathbb{P}(Z = z_j) \\ &= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i \mid Z = z_j)\mathbb{P}(Z = z_j) = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i,\ Z = z_j) = \mathbb{E}(X\mathbb{1}_{\{Z=z_j\}}). \qquad (112)\end{aligned}$$

Proof of Theorem 4.1. The proof is in four parts.

(1) The equivalence mentioned in the third property is true. We show that for any π-system $\mathscr{A}$, if $\mathbb{E}(Y\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A)$ holds for all $A \in \mathscr{A}$, then $\mathbb{E}(Y\mathbb{1}_G) = \mathbb{E}(X\mathbb{1}_G)$ holds for all $G \in \mathscr{G} = \sigma(\mathscr{A})$. Let $W := Y - X$, and define for each $G \in \mathscr{G}$ that $\mu^+(G) := \mathbb{E}(W^+\mathbb{1}_G)$ and $\mu^-(G) := \mathbb{E}(W^-\mathbb{1}_G)$. Then since X and Y are integrable, W must also be integrable, thus μ^+ and μ^- are necessarily finite measures which agree on a π system, according to Proposition 3.13. Therefore, we can conclude this part by the uniqueness lemma (Lemma 1.17).

(2) $\mathbb{E}(X \mid \mathscr{G})$ is almost surely unique. If $Y, \tilde{Y} \in \mathscr{L}^1(\Omega, \mathscr{G}, \mathbb{P})$ are two versions of $\mathbb{E}(X \mid \mathscr{G})$, then for all $G \in \mathscr{G}$, we have that $\mathbb{E}((Y - \tilde{Y})\mathbb{1}_G) = 0$. In particular, we have that

$$\mathbb{E}\left((Y - \tilde{Y})\mathbb{1}_{\{Y>\tilde{Y}\}}\right) = 0, \qquad (113)$$

which necessarily means that $(Y - \tilde{Y})\mathbb{1}_{\{Y>\tilde{Y}\}} = 0$ almost everywhere. Analogously, we are able to deduce that $(Y - \tilde{Y})\mathbb{1}_{\{Y\leq\tilde{Y}\}} = 0$ almost everywhere, so that $Y = \tilde{Y}$ almost everywhere.

(3) $\mathbb{E}(X \mid \mathscr{G})$ exists for all $X \in \mathscr{L}^2 \subseteq \mathscr{L}^1$. $V := \mathscr{L}^2(\Omega, \mathscr{G}, \mathbb{P})$ is complete, so that it is a closed subspace of $\mathscr{L}^2(\Omega, \mathscr{F}, \mathbb{P})$. If $X \in \mathscr{L}^2(\mathscr{F})$, let $Y \in \mathscr{L}^2(\mathscr{G})$ be the orthogonal projection of X on V, so that

$$\mathbb{E}((X - Y)^2) = \inf\left\{\mathbb{E}((X - W)^2);\ W \in \mathscr{L}^2(\mathscr{G})\right\}, \qquad (114)$$

$$\langle X - Y, Z\rangle = \mathbb{E}((X - Y)Z) = 0, \qquad \forall Z \in \mathscr{L}^2(\mathscr{G}). \qquad (115)$$

Clearly Y is measurable and $\mathbb{E}(|Y|) < \infty$ because $Y \in \mathscr{L}^2(\mathscr{G}) \subseteq \mathscr{L}^1(\mathscr{G})$, satisfying the first two requirements. On the other hand, taking $Z := \mathbb{1}_G \in \mathscr{G}$, we have that $\mathbb{E}(X\mathbb{1}_G) = \mathbb{E}(Y\mathbb{1}_G)$.

(4) $\mathbb{E}(X \mid \mathscr{G})$ exists for all $X \in \mathscr{L}^1$. Note that since $X = X^+ - X^-$, it is sufficient to prove existence for $X \in \mathscr{L}^1$, $X \geq 0$. We claim that if $U \in \mathscr{L}^2$ and $U \geq 0$ almost surely, then $\mathbb{E}(U \mid \mathscr{G}) \geq 0$ almost surely. Indeed, let W be a version of $\mathbb{E}(U \mid \mathscr{G})$, then $0 \geq \mathbb{E}(W\mathbb{1}_{\{W<0\}}) = \mathbb{E}(U\mathbb{1}_{\{W<0\}}) \geq 0$, which means that $\mathbb{E}(W\mathbb{1}_{\{W<0\}}) = 0$, i.e., $W\mathbb{1}_{\{W<0\}} = 0$ almost surely. This necessarily means that $W \geq 0$ almost surely, so the claim is proved. Now we approach X by an increasing sequence of non-negative simple (and thus square integrable) random variables X_n, i.e., $0 \leq X_n \uparrow X$. For each $n \in \mathbb{N}$, let Y_n be a version of $\mathbb{E}(X_n \mid \mathscr{G})$. Let $Y(\omega) := \limsup_{n\to\infty} Y_n(\omega)$ for each $\omega \in \Omega$, then $Y_n \uparrow Y$ almost surely by our previous claim, where Y is $\mathscr{G}$-measurable. WHY? By the monotone convergence theorem (Theorem 3.2), we thus have that $\mathbb{E}(X\mathbb{1}_G) = \mathbb{E}(Y\mathbb{1}_G)$ for all $G \in \mathscr{G}$. HOW? In particular, $\mathbb{E}(Y) < \infty$ and thus all requirements are satisfied by Y.

The proof is thus complete. □

10/25 Lecture

Theorem 4.4 (Properties of the conditional expectation)**.** Let X, Z, Z_n, $n \in \mathbb{N}$ be integrable random variables, let $\mathscr{G}$ and $\mathscr{H}$ be sub-σ-algebras of $\mathscr{F}$, and let $a_1, a_2 \in \mathbb{R}$.

(1) If Y is a version of $\mathbb{E}(X \mid \mathscr{G})$, then $\mathbb{E}(X) = \mathbb{E}(Y)$.

(2) If X is $\mathscr{G}$-measurable, then $\mathbb{E}(X \mid \mathscr{G}) = X$ almost surely.

(3) Linearity: $\mathbb{E}(a_1X_1 + a_2X_2 \mid \mathscr{G}) = a_1\mathbb{E}(X_1 \mid \mathscr{G}) + a_2\mathbb{E}(X_2 \mid \mathscr{G})$.

(4) Positivity: If $X \geq 0$, then $\mathbb{E}(X \mid \mathscr{G}) \geq 0$.

(5) Monotone convergence: If $0 \leq X_n \uparrow X$, then $\mathbb{E}(X_n \mid \mathscr{G}) \uparrow \mathbb{E}(X \mid \mathscr{G})$ almost surely.

(6) Fatou: If $X_n \geq 0$, then $\mathbb{E}(\liminf X_n \mid \mathscr{G}) \leq \liminf \mathbb{E}(X_n \mid \mathscr{G})$.

(7) Dominated convergence: If $|X_n| \leq V$ for some random variable V with $\mathbb{E}(|V|) < \infty$, and $X_n \to X$ almost surely, then $\mathbb{E}(X_n \mid \mathscr{G}) \to \mathbb{E}(X \mid \mathscr{G})$ almost surely.

(8) Jensen: If $c : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}(c(X)) < \infty$, then $\mathbb{E}(c(X) \mid \mathscr{G}) \geq c(\mathbb{E}(X \mid \mathscr{G}))$ almost surely.

(9) $\|\mathbb{E}(X \mid \mathscr{G})\|_p \leq \|X\|_p$, $p \geq 1$.

(10) Tower property: If $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$, then $\mathbb{E}(\mathbb{E}(X \mid \mathscr{G}) \mid \mathscr{H}) = \mathbb{E}(X \mid \mathscr{H})$ almost surely.

(11) If $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|ZX|) < \infty$, then $\mathbb{E}(ZX \mid \mathscr{G}) = Z\mathbb{E}(X \mid \mathscr{G})$.

(12) Rôle of independence: If $\mathscr{H}$ is independent of $\sigma(\sigma(X), \mathscr{G})$, then $\mathbb{E}(X \mid \sigma(\mathscr{G}, \mathscr{H})) = \mathbb{E}(X \mid \mathscr{G})$ almost surely. In particular, X being independent of $\mathscr{H}$ implies that $\mathbb{E}(X \mid \mathscr{H}) = \mathbb{E}(X)$.

Proof. (1) We have that $\mathbb{E}(Y) = \mathbb{E}(Y\mathbb{1}_\Omega) = \mathbb{E}(X\mathbb{1}_\Omega) = \mathbb{E}(X)$.

(2) If X is $\mathscr{G}$-measurable, then $Y := X$ satisfies all conditions of Theorem 4.1, thus being a version of $\mathbb{E}(X \mid \mathscr{G})$.

(3) Let $X := a_1X_1 + a_2X_2$ and $Y := a_1\mathbb{E}(X_1 \mid \mathscr{G}) + a_2\mathbb{E}(X_2 \mid \mathscr{G})$. Clearly Y is $\mathscr{G}$-measurable and $\mathbb{E}(|Y|) < \infty$. Moreover, for any $G \in \mathscr{G}$, we have that

$$\mathbb{E}(Y\mathbb{1}_G) = a_1\mathbb{E}(\mathbb{E}(X_1 \mid \mathscr{G})\mathbb{1}_G) + a_2\mathbb{E}(\mathbb{E}(X_2 \mid \mathscr{G})\mathbb{1}_G) = a_1\mathbb{E}(X_1\mathbb{1}_G) + a_2\mathbb{E}(X_2\mathbb{1}_G) = \mathbb{E}(X\mathbb{1}_G), \tag{116}$$

so Y satisfies all conditions of Theorem 4.1, thus being a version of $\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(a_1X_1 + a_2X_2 \mid \mathscr{G})$.

(4) To be done, similar to the last part of proof of Theorem 4.1...

(5) To be done, similar to the last part of proof of Theorem 4.1...

(6) To be done, based on monotone convergence...

(7) To be done, based on Fatou...

(8) Recall that the argument to prove Jensen's inequality was based on the existence of $a, b \in \mathbb{R}$, such that $c(x) \geq ax + b$, with equality at $x = \mathbb{E}(X)$. Here since $\mathbb{E}(X \mid \mathscr{G})$ is a function, we cannot treat it as a number to which we could attribute a and b. Instead we use countably many inequalities at the same time. Adapt Lemma 3.20 to show that there exist $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ real-valued sequences, such that $c(x) = \sup_{n\in\mathbb{N}}(a_nx+b_n)$. Hence $c(X) \geq a_nX + b_n$, and thus $\mathbb{E}(c(X) \mid \mathscr{G}) \geq a_n\mathbb{E}(X \mid \mathscr{G}) + b_n$ for all $n \in \mathbb{N}$. We can now conclude that

$$\mathbb{E}(c(X) \mid \mathscr{G}) \geq \sup_{n\in\mathbb{N}}(a_n\mathbb{E}(X \mid \mathscr{G}) + b_n) = c(\mathbb{E}(X \mid \mathscr{G})). \tag{117}$$

(9) Using Jensen's inequality with $c(x) := |x|^p$, we have that

$$\|\mathbb{E}(X \mid \mathscr{G})\|_p = (\mathbb{E}(|\mathbb{E}(X \mid \mathscr{G})|^p))^{1/p} \leq (\mathbb{E}(\mathbb{E}(|X|^p \mid \mathscr{G})))^{1/p} = (\mathbb{E}(|X|^p))^{1/p} = \|X\|_p\,, \tag{118}$$

where we have used that the expectation of a version of $\mathbb{E}(|X|^p \mid \mathscr{G})$ is the same as the expectation of $|X|^p$.

(10) Let Y be a version of $\mathbb{E}(X \mid \mathscr{G})$, then by definition $\mathbb{E}(Y\mathbb{1}_G) = \mathbb{E}(X\mathbb{1}_G)$ for any $G \in \mathscr{G}$. Moreover, let Z be a version of $\mathbb{E}(Y \mid \mathscr{H})$, then again by definition $\mathbb{E}(Z\mathbb{1}_H) = \mathbb{E}(Y\mathbb{1}_H)$ for any $H \in \mathscr{H}$. Hence $\mathbb{E}(Z\mathbb{1}_H) = \mathbb{E}(X\mathbb{1}_H)$ for all $H \in \mathscr{H}$. As long as Z is $\mathscr{H}$-measurable and $\mathbb{E}(|Z|) < \infty$, we can conclude that Z is a version of $\mathbb{E}(X \mid \mathscr{H})$, *i.e.*, $\mathbb{E}(\mathbb{E}(X \mid \mathscr{G}) \mid \mathscr{H}) = \mathbb{E}(X \mid \mathscr{H})$ almost surely.

(11) Let $Y := \mathbb{E}(X \mid \mathscr{G})$, and we use the *Standard Machine*.

- **Case 1.** Let $Z := \mathbb{1}_G$ for some $G \in \mathscr{G}$. Then YZ is integrable, and for all $H \in \mathscr{G}$, we have that

$$\mathbb{E}(YZ\mathbb{1}_H) = \mathbb{E}(Y\mathbb{1}_{G\cap H}) = \mathbb{E}(X\mathbb{1}_{G\cap H}) = \mathbb{E}(XZ\mathbb{1}_H). \tag{119}$$

This means that YZ is a version of $\mathbb{E}(XZ \mid \mathscr{G})$, so that $Z\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(XZ \mid \mathscr{G})$.

- **Case 2.** By linearity and Case 1, $Z\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(XZ \mid \mathscr{G})$. for any simple random variable Z.
- **Case 3.** Suppose now that Z is a non-negative random variable. Define a sequence of simple random variables by $W_n := (2^{-n} \lfloor 2^n Z \rfloor) \wedge n$, then clearly $W_n \uparrow Z$. For any $H \in \mathscr{G}$, we have that $YW_n \mathbb{1}_H \uparrow YZ\mathbb{1}_H$ and $XW_n\mathbb{1}_H \uparrow YZ\mathbb{1}_H$. Hence $\mathbb{E}(YW_n\mathbb{1}_H) \uparrow \mathbb{E}(YZ\mathbb{1}_H)$ and $\mathbb{E}(XW_n\mathbb{1}_H) \uparrow \mathbb{E}(YZ\mathbb{1}_H)$ by the monotone convergence theorem (Theorem 3.2). By Case 2, necessarily $\mathbb{E}(YW_n\mathbb{1}_H) = \mathbb{E}(XW_n\mathbb{1}_H)$ for all $n \in \mathbb{N}$, and thus $\mathbb{E}(YZ\mathbb{1}_H) = \mathbb{E}(XZ\mathbb{1}_H)$. This means that YZ is a version of $\mathbb{E}(XZ \mid \mathscr{G})$, so that $Z\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(XZ \mid \mathscr{G})$.
- **Case 4.** Now let Z be not necessarily non-negative and write $Z = Z^+ - Z^-$. Then by Case 3, we have that $Z^+\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(XZ^+ \mid \mathscr{G})$ and $Z^-\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(XZ^- \mid \mathscr{G})$. Therefore, we can conclude by linearity that $Z\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(XZ \mid \mathscr{G})$.

(12) The intuition here is that, $\mathscr{H}$ is independent of $\sigma(\sigma(X), \mathscr{G})$ if and only if given that $\omega \in G$ for some $G \in \mathscr{G}$, the fact that $\omega \in A$ for some $A \in \mathscr{H}$ does not provide any additional information on $X(\omega)$. Assume without loss of generality that $X \geq 0$ and $\mathbb{E}(X) < \infty$. Let $Y := \mathbb{E}(X \mid \mathscr{G})$, then we want to show that for all $I \in \sigma(\mathscr{G}, \mathscr{H})$, it holds that $\mathbb{E}(Y\mathbb{1}_I) = \mathbb{E}(X\mathbb{1}_I)$. First suppose $I := G \cap H$ for some $G \in \mathscr{G}$ and $H \in \mathscr{H}$, then

$$\mathbb{E}(X\mathbb{1}_I) = \mathbb{E}(X\mathbb{1}_G\mathbb{1}_H) = \mathbb{E}(X\mathbb{1}_G)\mathbb{E}(\mathbb{1}_H) = \mathbb{E}(X\mathbb{1}_G)\mathbb{P}(H), \tag{120}$$
$$\mathbb{E}(Y\mathbb{1}_I) = \mathbb{E}(Y\mathbb{1}_G\mathbb{1}_H) = \mathbb{E}(Y\mathbb{1}_G)\mathbb{E}(\mathbb{1}_H) = \mathbb{E}(Y\mathbb{1}_G)\mathbb{P}(H), \tag{121}$$

where we have used that $X\mathbb{1}_G$ and $\mathbb{E}(X \mid \mathscr{G})\mathbb{1}_G$ are both independent of $\mathbb{1}_H$ due to the assumption that $\mathscr{H}$ is independent of $\sigma(\sigma(X), \mathscr{G})$. Moreover, since Y is a version of $\mathbb{E}(X \mid \mathscr{G})$, we have that $\mathbb{E}(X)\mathbb{1}_G = \mathbb{E}(Y)\mathbb{1}_G$, so the finite measures $I \mapsto \mathbb{E}(X\mathbb{1}_I)$ and $I \mapsto \mathbb{E}(Y\mathbb{1}_I)$ agree on the π-system $\mathscr{I} := \{G \cap H;\ G \in \mathscr{G},\ H \in \mathscr{H}\}$, by arbitrariness of G and H. Hence they also agree on $\sigma(I) = \sigma(\mathscr{G}, \mathscr{H})$ by the uniqueness lemma (Lemma 1.17). In other words, $\mathbb{E}(X\mathbb{1}_I) = \mathbb{E}(Y\mathbb{1}_I)$ for all $I \in \sigma(\mathscr{G}, \mathscr{H})$, so that Y is a version of $\mathbb{E}(X \mid \sigma(\mathscr{G}, \mathscr{H}))$, which means that $\mathbb{E}(X \mid \mathscr{G}) = \mathbb{E}(X \mid \sigma(\mathscr{G}, \mathscr{H}))$. □

10/30 Lecture

Example 4.5. Let (X, Z) be a random variable with probability density function $f_{X,Z}(x, z)$. Set

$$f_Z(z) := \int_{\mathbb{R}} f_{X,Z}(x, z)dx, \tag{122}$$

which defines a probability density function of Z. Then we define the elementary conditional probability density function $f_{X|Z}$ by

$$f_{X|Z}(x, z) := \begin{cases} f_{X,Z}(x, z)/f_Z(z), & \text{if } f_Z(z) \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{123}$$

Let h be a Borel function such that $\mathbb{E}(|h(X)|) < \infty$. Then the function $g(z) := \int_{\mathbb{R}} h(x)f_{X|Z}(x, z)dx$ is defined almost everywhere for $z \in \mathbb{R}$, and $Y := g(Z) = \mathbb{E}(h(X) \mid \sigma(Z))$ almost surely. TO BE DONE, SEE *Williams p. 87...*

Example 4.6. Let $X_1, \ldots, X_r$ be independent random variables, $h : \mathbb{R}^r \to \mathbb{R}$ be bounded, and

$$\gamma^h(x_1) := \mathbb{E}(h(x_1, X_2, \ldots, X_r)). \tag{124}$$

Then $\gamma^h(X_1) = \mathbb{E}(h(X_1, \ldots, X_r) \mid X_1)$ almost surely. TO BE DONE, SEE *Williams p. 92...*

4.2 Martingales

4.2.1 Filtrations and Adapted Processes

Definition 4.7. An increasing family $\{\mathscr{F}_n;\ n \geq 0\}$ of sub-σ-algebras of $\mathscr{F}$, such that $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \ldots \subseteq \mathscr{F}_n \subseteq \ldots \subseteq \mathscr{F}$ is called a **filtration**. By convention, we denote

$$\mathscr{F}_\infty := \sigma\left(\bigcup_n \mathscr{F}_n\right) \subseteq \mathscr{F}. \tag{125}$$

Remark 4.8. $\mathscr{F}_n$ can be thought of as the information available to us at time n. Usually $\mathscr{F}_n = \sigma(W_0, \dots, W_n)$ for some stochastic process $\{W_n\}_{n\in\mathbb{N}}$, so that the information available to us from the outcome ω at time n are the values of $W_0(\omega), \dots, W_n(\omega)$. In the case of a general filtration, the information available to use are the values of $Z(\omega)$ for all $\mathscr{F}_n$-measurable functions Z.

Definition 4.9. $X = \{X_n;\ n \geq 0\}$ is called an **adapted process** if X_n is $\mathscr{F}_n$-measurable for all n.

Remark 4.10. Intuitively, X_n is a function of the information available at time n. Usually $\mathscr{F}_n = \sigma(W_0, \dots, W_n)$ and $X_n = f_n(W_0, \dots, W_n)$, where $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ is a measurable function.

4.2.2 Martingales

We now introduce the notion of martingales (respectively submartingales and supermartingales), which are stochastic processes "remaining stable (respectively increasing and decreasing) on average". To this end, we do not only look at the evolution of the expectation of the process, but rather at whether the increment is conditionally null (respectively positive and negative) in expectation.

Definition 4.11. Let $\{X_n;\ n \in \mathbb{N}\}$ be an $\mathbb{F} = \{\mathscr{F}_n;\ n \in \mathbb{N}\}$-adapted integrable process (*i.e.*, such that $\mathbb{E}(|X_n|) < \infty$ for all $n \in \mathbb{N}$). Then

$$\{X_n;\ n \in \mathbb{N}\} \text{ is a } \begin{cases} \textbf{martingale} \\ \textbf{submartingale} \\ \textbf{supermartingale} \end{cases} \quad \text{if } \begin{cases} \mathbb{E}(X_{n+1} \mid \mathscr{F}_n) = X_n \\ \mathbb{E}(X_{n+1} \mid \mathscr{F}_n) \geq X_n \\ \mathbb{E}(X_{n+1} \mid \mathscr{F}_n) \leq X_n \end{cases} \quad \text{almost surely for all } n \in \mathbb{N}. \tag{126}$$

Remark 4.12. $\{X_n;\ n \in \mathbb{N}\}$ is a martingale if and only if it is a submartingale and a supermartingale. Moreover, $\{X_n;\ n \in \mathbb{N}\}$ is a martingale (respectively submartingale and supermartingale) if and only if $(X_n - X_0)$ is a martingale (respectively submartingale and supermartingale). In practice, we focus our attention on processes null at 0.

Remark 4.13. If $\{X_n;\ n \in \mathbb{N}\}$ is a martingale (respectively submartingale and supermartingale), then $\mathbb{E}(X_n) = \mathbb{E}(X_0)$ (respectively $\mathbb{E}(X_n) \geq \mathbb{E}(X_0)$ and $\mathbb{E}(X_n) \leq \mathbb{E}(X_0)$) almost surely for all $n \in \mathbb{N}$. Indeed, if we look at supermartingales, we can see that $\mathbb{E}(X_0) \geq \mathbb{E}(\mathbb{E}(X_1 \mid \mathscr{F}_0)) = \mathbb{E}(X_1)$, and this is recursive. The other two types are analgous.

Remark 4.14. If $\{X_n;\ n \in \mathbb{N}\}$ is a martingale (respectively submartingale and supermartingale), $\mathbb{E}(X_{n+k} \mid \mathscr{F}_n) = X_n$ (respectively $\mathbb{E}(X_{n+k} \mid \mathscr{F}_n) \geq X_n$ and $\mathbb{E}(X_{n+k} \mid \mathscr{F}_n) \leq X_n$) almost surely for $n, k \in \mathbb{N}$. Indeed, if we look at supermartingales, we can see that

$$\mathbb{E}(X_{n+k} \mid \mathscr{F}_n) = \underbrace{\mathbb{E}(\mathbb{E}(X_{n+k} \mid \mathscr{F}_{n+k-1}) \mid \mathscr{F}_n)}_{\text{tower property}} \leq \mathbb{E}(X_{n+k-1} \mid \mathscr{F}_n) \leq \ldots \leq \mathbb{E}(X_{n+1} \mid \mathscr{F}_n) \leq X_n. \tag{127}$$

The other two types are analogous.

Example 4.15. Any deterministic non-decreasing (respectively non-increasing) sequence $\{a_n\}_{n\in\mathbb{N}}$ is a submartingale (respectively supermartingale).

Example 4.16. Let $\{X_n;\ n \in \mathbb{N}\}$ be a sequence of independent random variables, with $\mathbb{E}(|X_i|) < \infty$ and $\mathbb{E}(X_i) = 0$ (centered) for all $i \in \mathbb{N}$. Then $S_n := \sum_{i=1}^n X_i$ is a martingale with respect to the filtration $\mathscr{F}_n = \sigma(X_1, \dots, X_n)$, $n \geq 1$, and $\mathscr{F}_0 = \{\varnothing, \Omega\}$. Indeed, we can see that

$$\mathbb{E}(S_n \mid \mathscr{F}_{n-1}) = \underbrace{\mathbb{E}(S_{n-1} \mid \mathscr{F}_{n-1})}_{S_{n-1} \text{ is } \mathscr{F}_{n-1}\text{-measurable}} + \underbrace{\mathbb{E}(X_n \mid \mathscr{F}_{n-1})}_{X_n \text{ independent of } \mathscr{F}_{n-1}} = S_{n-1} + \mathbb{E}(X_n) = S_{n-1}. \tag{128}$$

Example 4.17. Let $\{X_n;\ n \in \mathbb{N}\}$ be a sequence of independent non-negative random variables with $\mathbb{E}(X_k) = 1$ for all k. Define $M_0 = 0$ and $M_n = X_1 \ldots X_n$, with $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $\mathscr{F}_n = \sigma(X_1, \dots, X_n)$, $n \geq 1$. Then $\{M_n;\ n \in \mathbb{N}\}$ is a martingale. Indeed, we can see that

$$\mathbb{E}(M_n \mid \mathscr{F}_{n-1}) = \mathbb{E}(X_n M_{n-1} \mid \mathscr{F}_{n-1}) = \mathbb{E}(X_n)\mathbb{E}(M_{n-1} \mid \mathscr{F}_{n-1}) = \mathbb{E}(X_n) M_{n-1} = M_{n-1}. \tag{129}$$

Example 4.18. Let $\{\mathscr{F}_n;\ n \in \mathbb{N}\}$ be a filtration, and let X be a random variable such that $\mathbb{E}(|X|) < \infty$. Let $X_n := \mathbb{E}(X \mid \mathscr{F}_n)$, then $\{X_n;\ n \in \mathbb{N}\}$ is a martingale. Indeed, by the tower property we can see that

$$\mathbb{E}(X_n \mid \mathscr{F}_{n-1}) = \mathbb{E}(\mathbb{E}(X \mid \mathscr{F}_n) \mid \mathscr{F}_{n-1}) = \mathbb{E}(X \mid \mathscr{F}_{n-1}) = X_{n-1}. \tag{130}$$

The question is, how general can this example be? We will prove later that a large class of martingales (called **uniformly integrable martingales**) can be written in this way, corresponding to the idea that X_n gives some "rough" information about an unknown variable X, and that this information is getting finer as n increases (and actually converging to X if X is $\mathscr{F}_\infty$-measurable, see later).

11/1 Lecture

The Discrete Stochastic Integral

Definition 4.19. Let $\{\mathscr{F}_n;\ n \in \mathbb{N}\}$ be a filtration on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We say that a process $C = \{C_n;\ n \geq 1\}$ is **previsible** if each C_n is $\mathscr{F}_{n-1}$-measurable (note that C_0 is not defined).

Remark 4.20. Let $\{X_n;\ n \in \mathbb{N}\}$ be $\mathbb{F} = \{\mathscr{F}_n;\ n \in \mathbb{N}\}$-adapted and $C = \{C_n;\ n \geq 1\}$ be $\mathbb{F}$-previsible. Define

$$Y_n := (C \circ X)_n := \sum_{k=1}^{n} C_k(X_k - X_{k-1}), \tag{131}$$

then $C \circ X$ can be seen as a discrete analogue of the stochastic integral $\int C dX$.

Theorem 4.21. (1) Let the process C be previsible and bounded, where boundedness means that there exists $K < \infty$, such that $|C_n(\omega)| < K$ for all $\omega \in \Omega$ and $n \geq 1$. If $\{X_n;\ n \in \mathbb{N}\}$ is a martingale (respectively submartingale and supermartingale), and $C_n \geq 0$, then $C \circ X$ is a martingale (respectively submartingale and supermartingale) that is null at 0.

(2) Replace the boundedness of C in the previous part by $C_n, X_n \in \mathscr{L}^2$ for all $n \in \mathbb{N}$, the same conclusion holds.

Proof. To be done... ☐

Theorem 4.22 (Doob's decomposition theorem)**.** Let $\{X_n;\ n \in \mathbb{N}\}$ be an adapted process with $X_n \in \mathscr{L}^1$, $n \in \mathbb{N}$. Then

(1) X has Doob decomposition $X = X_0 + M + A$, where M is a martingale null at 0, and A is a previsible process also null at 0. Moreover, if $X = X_0 + \tilde{M} + \tilde{A}$ is another Doob decomposition, then

$$\mathbb{P}\left(M_n = \tilde{M}_n,\ A_n = \tilde{A}_n,\ n \in \mathbb{N}\right) = 1. \tag{132}$$

(2) X is a submartingale (respectively supermartingale) if and only if A is increasing (respectively decreasing), *i.e.*, $\mathbb{P}(A_n \leq A_{n+1},\ n \in \mathbb{N}) = 1$ (respectively $\mathbb{P}(A_n \geq A_{n+1},\ n \in \mathbb{N}) = 1$).

Proof. (1) Define

$$A_n := \sum_{k=1}^{n} (\mathbb{E}(X_k \mid \mathscr{F}_{k-1}) - X_{k-1}), \qquad M_n := \sum_{k=1}^{n} (X_k - \mathbb{E}(X_k \mid \mathscr{F}_{k-1})), \tag{133}$$

where A_n adds up the expected increments of X, and M_n adds up the surprises, *i.e.*, the part of each X_k that is not known one step before. Intuitively and arithmetically, we have that $X_n = X_0 + M_n + A_n$. Moreover, M is a martingale because

$$\begin{aligned}\mathbb{E}(M_n \mid \mathscr{F}_{n-1}) &= \mathbb{E}(M_{n-1} \mid \mathscr{F}_{n-1}) + \mathbb{E}(X_n \mid \mathscr{F}_{n-1}) - \mathbb{E}(\mathbb{E}(X_n \mid \mathscr{F}_{n-1}) \mid \mathscr{F}_{n-1}) \\ &= \mathbb{E}(M_{n-1}) + \mathbb{E}(X_n \mid \mathscr{F}_{n-1}) - \mathbb{E}(X_n \mid \mathscr{F}_{n-1}) = \mathbb{E}(M_{n-1}),\end{aligned} \tag{134}$$

and A is a previsible because each A_n depends only on information given up till $\mathscr{F}_{n-1}$. The existence of the Doob composition is thus complete, and it suffices to prove its uniqueness. For a Doob decomposition $X = X_0 + M + A$, we can see that

$$\mathbb{E}(X_n - X_{n-1} \mid \mathscr{F}_{n-1}) = \underbrace{\mathbb{E}(M_n \mid \mathscr{F}_{n-1})}_{M \text{ martingale}} - \underbrace{\mathbb{E}(M_{n-1} \mid \mathscr{F}_{n-1})}_{\mathscr{F}_{n-1}\text{-measurable}} + \underbrace{\mathbb{E}(A_n \mid \mathscr{F}_{n-1})}_{\mathscr{F}_{n-1}\text{-measurable}} - \underbrace{\mathbb{E}(A_{n-1} \mid \mathscr{F}_{n-1})}_{\mathscr{F}_{n-1}\text{-measurable}}$$
$$= M_{n-1} - M_{n-1} + A_n - A_{n-1} = A_n - A_{n-1}, \tag{135}$$

so A_n can be uniquely determined by summing up these increments. Since X_0 is fixed, M_n can thus also be uniquely determined, so we can see that the Doob composition is unique.

(2) This is trivial from how we defined A_n. Indeed, if A_n is non-decreasing, then $\mathbb{E}(X_n \mid \mathscr{F}_{n-1}) - X_{n-1} \geq 0$ is non-negative for all n, and thus X_n is a submartingale by definition. The reverse is analogous, and the proof is complete. ☐

Stopping Times

Definition 4.23. A mapping $T : \Omega \to \mathbb{Z}_+ \cup \{\infty\}$ is called a **stopping time** if for all $n \in \mathbb{N}$, we have that

$$\{T \leq n\} = \{\omega \in \Omega;\ T(\omega) \leq n\} \in \mathscr{F}_n. \tag{136}$$

Remark 4.24. Equivalently, T is a stopping time if $\{T = n\} \in \mathscr{F}_n$ for all $n \in \mathbb{N}$. Check this...

Remark 4.25. Usually a stopping time is interpreted as a time when you take a certain decision, and the requirement in the definition corresponds to asking for this decision to depend only on the history up to (and including) time n. As an example, the decision might be the decision to stop playing a gambling game. As another example, for a meteorologist having the weather information up to the present time, the first day of 2009 when the temperature is above 15°C is a stopping time. The last day of 2009 when the temperature is above 15 ∘C is not a stopping time, because it may depend on future information.

Theorem 4.26. Let $\{X_n;\ n \in \mathbb{N}\}$ be a martingale (respectively submartingale and supermartingale), and let T be a stopping time. Then $\{X_{n\wedge T};\ n \in \mathbb{N}\}$ is a martingale (respectively submartingale and supermartingale).

Proof. Take $C_n := \mathbb{1}_{n \leq T}$, then we have the discrete stochastic integral

$$(C \circ X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}) = \sum_{k=1}^{n} \mathbb{1}_{k\leq T}(X_k - X_{k-1}) = \sum_{k=1}^{n\wedge T} (X_k - X_{k-1}) = X_{n\wedge T}. \tag{137}$$

Clearly C is previsible and bounded, so by Theorem 4.21, if $\{X_n;\ n \in \mathbb{N}\}$ is a martingale (respectively submartingale and supermartingale), then $\{X_{n\wedge T};\ n \in \mathbb{N}\} = \{C_n;\ n \in \mathbb{N}\}$ is a martingale (respectively submartingale and supermartingale) as well. The proof is thus complete. ☐

Theorem 4.27 (Doob's optional stopping theorem)**.** Let X be a supermartingale and T be a stopping time. Then X_T is integrable and $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ in each of the following situations.

(1) T is bounded, *i.e.*, there exists $K < \infty$, such that $T(\omega) \leq K$ for all $\omega \in \Omega$.

(2) T is almost surely finite, and X is bounded, *i.e.*, there exists $K < \infty$, such that $|X_n(\omega)| \leq K$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$.

(3) $\mathbb{E}(T) < \infty$, and the increments of X are bounded, *i.e.*, there exists $K < \infty$, such that $|X_n(\omega) - X_{n-1}(\omega)| \leq K$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$.

(4) T is almost surely finite, and X is non-negative, *i.e.*, $X_n(\omega) \geq 0$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$.

Proof. (1) By Theorem 4.26, we can see that $\{X_{n\wedge T};\ n \in \mathbb{N}\}$ is a supermartingale. This necessarily means that $\mathbb{E}(X_{n\wedge T}) \leq \mathbb{E}(X_{0\wedge T}) = \mathbb{E}(X_0)$ for all $n \in \mathbb{N}$. Assume that T is bounded by K, then taking $n \geq K$ concludes that $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$.

(2) Since T is almost surely finite, we have that $X_{n\wedge T} \to X_T$ as $n \to \infty$. Since X is bounded, by the dominated convergence theorem (Theorem 3.8), we can see that $\mathbb{E}(X_{n\wedge T}) \to \mathbb{E}(X_T)$. Same as the previous part, $\mathbb{E}(X_{n\wedge T}) \leq \mathbb{E}(X_0)$ for all $n \in \mathbb{N}$, so that $\mathbb{E}(X_T) \leq \mathbb{E}(X_0)$ as well.

(3) Same as in the previous part, $X_{n\wedge T} \to X_T$ as $n \to \infty$. We have that

$$|X_{n\wedge T} - X_0| = \left| \sum_{i=1}^{n\wedge T} (X_i - X_{i-1}) \right| \leq \sum_{i=1}^{n\wedge T} |X_i - X_{i-1}| \leq (n \wedge T)K \leq TK, \tag{138}$$

so we can apply the dominated convergence theorem (Theorem 3.8) to see that $\mathbb{E}(X_{n\wedge T}) \to \mathbb{E}(X_T)$. The rest of the proof is the same as in previous parts.

(4) Same as in the previous part, $X_{n\wedge T} \to X_T$ as $n \to \infty$. Moreover since X is non-negative, we can apply Fatou's lemma (Lemma 3.3) to see that

$$\mathbb{E}(X_T) = \mathbb{E}\left(\liminf_{n\to\infty} X_{n\wedge T}\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_{n\wedge T}) \leq \mathbb{E}(X_0), \tag{139}$$

where the last inequality is by arguments analogous to previous parts. □

Remark 4.28. Doob's optional stopping theorem (Theorem 4.27) would *not* hold without such boundedness assumptions on T or X. For instance, let $\{\Delta_i\}_{i\in\mathbb{N}}$ be independent and identically distributed with $\mathbb{P}(\Delta_i = 1) = \mathbb{P}(\Delta_i = -1) = 1/2$. Consider the filtration $\mathscr{F}_n := \sigma(\Delta_1, \ldots, \Delta_n)$ with $\mathscr{F}_0 = \{\varnothing, \Omega\}$. Let $X_n := \sum_{i=1}^n m_i \Delta_i$, where $\{m_i;\ i \in \mathbb{N}\}$ is some previsible process. Let $T := \inf\{k \geq 1;\ \Delta_k = 1\}$, and assume that $m_i := 2^{i-1} m_1 \mathbb{1}_{\{i \leq T\}}$. Then

$$X_T = \sum_{i=1}^T m_i \Delta_i = \sum_{i=1}^T 2^{i-1} m_1 \mathbb{1}_{\{i\leq T\}} \Delta_i = m_1 \sum_{i=1}^T 2^{i-1} \Delta_i = m_1 \left(- \sum_{i=1}^{T-1} 2^{i-1} + 2^{T-1} \right) = m_1. \tag{140}$$

Hence, $\mathbb{E}(X_T) = m_1 > 0$, even though $\{X_n;\ n \in \mathbb{N}\}$ is clearly an $\mathscr{F}_n$-adapted martingale.

Lemma 4.29. Assume that T is a stopping time and that there exists $N \in \mathbb{N}$ and $\epsilon > 0$, such that for all $n \in \mathbb{N}$, it holds that $\mathbb{P}(T < n + N \mid \mathscr{F}_n) \geq \epsilon$ almost surely. Then $\mathbb{E}(T) < \infty$.

Proof. EXERCISE, TO BE DONE... □

Remark 4.30. The above lemma provides a simple way to prove that $\mathbb{E}(T) < \infty$.

Proposition 4.31. Let D be a connected finite subset of $\mathbb{Z}^d$, $d \geq 1$ (connectedness means that for any $x, y \in D$, there exists $x_1, \ldots, x_k$, such that $x = x_1$, $y = x_k$, and $x_i \sim x_{i+1}$ for all i). Let

$$\partial D := \left\{ y \in \mathbb{Z}^d \setminus D;\ \text{there exists } x \in D \text{ such that } x \sim y \right\}. \tag{141}$$

Then for all $f : \partial D \to \mathbb{R}$, there exists a unique function $h : D \cup \partial D \to \mathbb{R}$ such that $h\mid_{\partial D} = f$ and h is harmonic, *i.e.*,

$$h(x) = \frac{1}{2d} \sum_{y:\ y\sim x} h(y). \tag{142}$$

Proof. HOMEWORK 4, TO BE DONE... □

11/6 Lecture

4.2.3 Forward Convergence Theorem for Martingales

Remark 4.32. Let $\{X_n;\ n \in \mathbb{N}\}$ be an integrable random process, for instance, modeling the value of the stock market. Consider the following strategy: (A) you do not invest until the value of X gets below a, in which case you buy a share; (B) you keep your share until X gets above b, in which case you sell your share and goes back to (A). There are three remarks.

(1) However clever this strategy may look, if X is a supermartingale and you stop playing at some bounded time, then your losses will always be greater than your earnings on average, according to Doob's optional stopping theorem (Theorem 4.27).

(2) Your earnings are bounded below by $(b-a)$ times the number of times the process went up from a to b.

(3) You may have some losses if you stop at a time n where you are still playing and the value is under the price at which you bought, in which case the loss is bounded above by $(X_n - a)^-$.

Combining the remarks, if X is a supermartingale, then the number of times the process went up from a to b must be bounded above by $\mathbb{E}((X_n - a)^-)/(b-a)$.

Definition 4.33. If $\{x_n\}_{n\in\mathbb{N}}$ is a real sequence and $a < b$ are two real numbers, we define two integer-valued sequences $S_k(x)$ and $T_k(x)$, $k \geq 1$ recursively as follows. Let $T_0(x) := 0$, and for $k \in \mathbb{N}$, let

$$S_{k+1}(x) := \inf\{n \geq T_k(x);\ x_n < a\}, \qquad T_{k+1}(x) := \inf\{n \geq S_{k+1}(x);\ x_n > b\}, \tag{143}$$

with the convention that $\inf(\varnothing) = \infty$. Let $N_n([a,b],x) := \sup\{k \in \mathbb{N};\ T_k(x) \leq n\}$ be the **number of upcrossings** of x between a and b before time n, which increases to the **total number of upcrossings** $N([a,b],x) := N_\infty([a,b],x) = \sup\{k \in \mathbb{N};\ T_k(x) < \infty\}$.

Remark 4.34. $T_0(x) = 0$, then $S_1(x)$ is the first time after 0 that x_n is below a, and $T_1(x)$ is the first time after $S_1(x)$ that x_n is above b. This completes one iteration of down and up, and so on.

Lemma 4.35 (Doob's upcrossing lemma)**.** Let X be a supermartingale, and let $a < b$ be two real numbers. Then for every $n \in \mathbb{N}$, we have that

$$(b-a)\mathbb{E}(N_n([a,b],X)) \leq \mathbb{E}((X_n - a)^-). \tag{144}$$

Proof. It is immediate by induction that $S_k = S_k(X)$ and $T_k = T_k(X)$ defined above are stopping times. Furthermore, define for all $n \geq 1$ that

$$C_n := \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k < n \leq T_k\}}. \tag{145}$$

We note that C_n only takes values 0 or 1 because $(S_k, T_k]$'s are non-intersecting by definition. Moreover, $\{C_n;\ n \geq 1\}$ is previsible because $\{S_k < n \leq T_k\} = \{S_k \leq n-1\} \cap \{T_k \leq n-1\}^C \in \mathscr{F}_{n-1}$. Letting $N_n := N_n([a,b],X)$, we have that the discrete stochastic integral

$$(C \circ X)_n = \sum_{i=1}^{n} C_i(X_i - X_{i-1}) = \sum_{k=1}^{N_n}(X_{T_k} - X_{S_k}) + (X_n - X_{S_{N_n+1}})\mathbb{1}_{\{S_{N_n+1} \leq n\}}. \tag{146}$$

To see this, we note that $C_i = 1$ if and only if $i \in (S_k, T_k]$ for some $k \in \mathbb{N}$ and otherwise $C_i = 0$. Then for each such interval $(S_k, T_k]$, we have $(X_{S_k+1} - X_{S_k}) + (X_{S_k+2} - X_{S_k+1}) + \ldots + (X_{T_k} - X_{T_k-1}) = X_{T_k} - X_{S_k}$, and outside of these intervals there are no contributions to the sum. Moreover, we need to limit T_k not to exceed n, because the summation is for i from 1 to n. By definition of N_n, we can see that N_n is the largest possible value of k such that $T_k \leq n$, so we sum for intervals only up to $(S_{N_n}, T_{N_n}]$. This however, leaves (possibly) remainder terms, because S_{N_n+1} may still be within n. Hence if $S_{N_n+1} \leq n$, we must further include the remainder term $(X_{S_{N_n+1}+1} - X_{S_{N_n+1}}) + (X_{S_{N_n+1}+2} - X_{S_{N_n+1}+1}) + \ldots + (X_n - X_{n-1}) = X_n - X_{S_{N_n+1}}$. Now we can further write that

$$(C \circ X)_n \geq (b-a)N_n + (X_n - a)\mathbb{1}_{\{X_n \leq a\}} = (b-a)N_n - (X_n - a)^-. \tag{147}$$

The first inequality holds because by definition of S_k and T_k, we know that X is below a for each S_k and above b for each T_k, so that each $X_{T_k} - X_{S_k} \geq b - a$. Moreover, CLEARLY $X_{S_{N_n+1}} \leq a$ BUT WHY $\{X_n \leq a\} \subseteq \{S_{N_n+1} \leq n\}$ OR THERE ARE SOME OTHER MAGIC FOR THE INDICATOR PART? For the second equality, it holds because

$$(X_n - a)^- = \max(a - X_n, 0) = (a - X_n)\mathbb{1}_{\{a \geq X_n\}} + 0\mathbb{1}_{\{a < X_n\}} = -(X_n - a)\mathbb{1}_{X_n \leq a}. \tag{148}$$

But since C only takes values 0 or 1, it is a non-negative bounded previsible process. Moreover, X is a supermartingale, so Theorem 4.21 gives that the discrete stochastic integral $C \circ X$ is also a supermartingale. This would mean that $\mathbb{E}((C \circ X)_n) \leq \mathbb{E}((C \circ X)_0) = 0$, so that

$$0 \geq \mathbb{E}((C \circ X)_n) \geq (b-a)\mathbb{E}(N_n) - \mathbb{E}((X_n - a)^-), \tag{149}$$

which finally implies that $(b-a)\mathbb{E}(N_n([a,b],X)) \leq \mathbb{E}((X_n - a)^-)$. The proof is now complete. □

Lemma 4.36. A real sequence x converges (in $\overline{\mathbb{R}} := [-\infty, \infty]$) if and only if $N([a,b],x) < \infty$ for all $a, b \in \mathbb{Q}$ with $a \neq b$.

Proof. We show that x does not converge if and only if there exists $a < b$ rationals, such that $N([a,b],x) = \infty$.

- $\Longleftarrow$ We have that $\liminf x_n \leq a < b \leq \limsup x_n$, because there are infinitely many crossings between a and b. This directly implies that x does not converge.
- $\Longrightarrow$ Since x does not converge, we have that $\liminf x_n < \limsup x_n$ and thus we can take two rationals $a < b$ in between.

The proof is thus complete by taking the contrapositive statement of above. □

Remark 4.37. The above lemma demonstrates how the notion of convergence is analytically related to the finiteness of the number of crossings.

Theorem 4.38 (Doob's forward convergence theorem)**.** Let X be a bounded supermartingale in $\mathscr{L}^1$, which means that $\sup_n \mathbb{E}(|X_n|) < \infty$. Then X_n converges almost surely towards an almost surely finite limit X_∞.

Proof. Fix rationals $a < b$, then by Doob's upcrossing lemma (Lemma 4.35), we have that

$$\mathbb{E}(N_n([a,b],X)) \leq \frac{(\mathbb{E}(X_n) - a)^-}{b-a} \leq \frac{|\mathbb{E}(X_n) - a|}{b-a} \leq \frac{|\mathbb{E}(X_n)| + |a|}{b-a} \leq \frac{\mathbb{E}(|X_n|) + |a|}{b-a}. \tag{150}$$

Since $N_n([a,b],X) \uparrow N([a,b],X)$ as $n \to \infty$ and $N_n([a,b],X)$ is non-negative, by the monotone convergence theorem (Theorem 3.2), we can see that $\mathbb{E}(N_n([a,b],X)) \uparrow \mathbb{E}(N([a,b],x))$ as well. Hence we have that

$$\mathbb{E}(N([a,b],X)) \leq \sup_n \frac{\mathbb{E}(|X_n|) + |a|}{b-a} = \frac{\sup_n \mathbb{E}(|X_n|) + |a|}{b-a} < \infty, \tag{151}$$

which necessarily implies that $N([a,b],X) < \infty$ almost surely. Therefore, we have that

$$\mathbb{P}\left(\bigcap_{\substack{a,b \in \mathbb{Q} \\ a<b}} \{N([a,b],X) < \infty\} \right) = 1, \tag{152}$$

and thus X_n converges almost surely to some X_∞, possibly infinite, according to Lemma 4.36. Now Fatou's lemma (Lemma 3.3) gives that $\mathbb{E}(|X_\infty|) \leq \liminf \mathbb{E}(|X_n|) < \infty$, so $|X_\infty| < \infty$ almost surely. The proof is thus complete. □

Corollary 4.39. If X is a non-negative supermartingale, then $X_\infty := \lim_n X_n$ exists almost surely.

Proof. Since X is non-negative supermartingale, we have that $\mathbb{E}(|X_n|) = \mathbb{E}(X_n) \leq \mathbb{E}(X_0) < \infty$ for all $n \in \mathbb{N}$. Hence by Doob's forward convergence theorem (Theorem 4.38), we can see that X_n necessarily converges almost surely to an almost surely finite X_∞, and the proof is thus complete. □

11/8 Lecture

Martingales Bounded in $\mathscr{L}^2$

The assumption that X is bounded in $\mathscr{L}^1$, *i.e.*, $\sup_n \mathbb{E}(|X_n|) < \infty$, as required in Doob's forward convergence theorem (Theorem 4.38), is not always easy to check. On the other hand, the advantage of working in $\mathscr{L}^2$ is that we have Pythagoras' rule, which will hold for martingales as it did for sums of independent random variables. However since $\mathscr{L}^2(\Omega, \mathscr{F}, \mathbb{P}) \subseteq \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ we are in a less general case here.

Definition 4.40. A process $M = \{M_n;\ n \in \mathbb{N}\}$ is bounded in $\mathscr{L}^p$, $p \geq 1$, if $\sup_n \|M_n\|_p < \infty$, or equivalently, $\sup_n \mathbb{E}(|M_n|^p) < \infty$.

Remark 4.41. If $M = \{M_n;\ n \in \mathbb{N}\}$ is a martingale bounded in $\mathscr{L}^2$, then for all $0 \leq r < s$, with the convention $M_{-1} := 0$, we can compute that

$$\begin{aligned}\langle M_s - M_{s-1}, M_r - M_{r-1}\rangle_{\mathscr{L}^2} &= \mathbb{E}((M_s - M_{s-1})(M_r - M_{r-1})) \\ &= \mathbb{E}(\mathbb{E}((M_s - M_{s-1})(M_r - M_{r-1})) \mid \mathscr{F}_{s-1}) \\ &= \mathbb{E}(\mathbb{E}(\underbrace{M_s}_{\text{martingale}} - \underbrace{M_{s-1}}_{\text{measurable}} \mid \mathscr{F}_{s-1}) \cdot \mathbb{E}(\underbrace{M_r}_{\text{measurable}} - \underbrace{M_{r-1}}_{\text{measurable}} \mid \mathscr{F}_{s-1})) \\ &= \mathbb{E}((M_{s-1} - M_{s-1})(M_r - M_{r-1})) = 0. \end{aligned} \tag{153}$$

Hence, still using the convention $M_{-1} := 0$, we can deduce that

$$\begin{aligned}\mathbb{E}(M_n^2) &= \mathbb{E}\left(\left(\sum_{k=0}^{n}(M_k - M_{k-1})\right)^2\right) \\ &= \mathbb{E}(M_0^2) + \sum_{k=1}^{n}\mathbb{E}\left((M_k - M_{k-1})^2\right) + 2\sum_{0\leq r<s\leq n}\mathbb{E}((M_s - M_{s-1})(M_r - M_{r-1})) \\ &= \mathbb{E}(M_0^2) + \sum_{k=1}^{n}\mathbb{E}\left((M_k - M_{k-1})^2\right). \end{aligned} \tag{154}$$

Theorem 4.42. Let M be a martingale for which $M_n \in \mathscr{L}^2$, $n \in \mathbb{N}$. Then M is bounded in $\mathscr{L}^2$ if and only if

$$\sum_{k=1}^{\infty}\mathbb{E}\left((M_k - M_{k-1})^2\right) < \infty, \tag{155}$$

and in this case, $M_n \to M_\infty$ almost surely and in $\mathscr{L}^2$.

Proof. It is obvious from the previous remark that the first conclusion holds. Based on this and using the fact that $(\mathbb{E}(X_n))^2 \leq \mathbb{E}(X_n^2)$ (by Jensen's inequality), we can use Doob's forward convergence theorem (Theorem 4.38) to ensure the almost sure existence of $M_\infty := \lim_n M_n$. Now we move on to prove the convergence in $\mathscr{L}^2$. Recall that that Pythagoras' theorem implies that

$$\mathbb{E}\left((M_{n+r} - M_n)^2\right) = \sum_{k=n+1}^{n+r}\mathbb{E}\left((M_k - M_{k-1})^2\right). \tag{156}$$

Hence by Fatou's lemma (Lemma 3.3), we can see that

$$\mathbb{E}\left((M_\infty - M_n)^2\right) = \mathbb{E}\left(\liminf_{r\to\infty}(M_{n+r} - M_n)^2\right) \leq \liminf_{r\to\infty}\mathbb{E}\left((M_{n+r} - M_n)^2\right) = \sum_{k=n+1}^{\infty}\mathbb{E}\left((M_k - M_{k-1})^2\right). \tag{157}$$

Bringing $n \to \infty$, necessarily $\mathbb{E}\left((M_\infty - M_n)^2\right) \to 0$, so we can conclude that $M_n \to M_\infty$ in $\mathscr{L}^2$. □

Remark 4.43. In fact, $\mathbb{E}\left((M_\infty - M_n)^2\right) = \sum_{k=n+1}^{\infty}\mathbb{E}\left((M_k - M_{k-1})^2\right)$, though in the proof above we only used that the left-hand side is less than or equal to the right-hand side. To conclude the equality, we are using that $f_r \to f$ in $\mathscr{L}^2$ as $r \to \infty$ implies that $\|f_r\|_2 \to \|f\|_2$, and in this case we take $f_r = M_{n+r} - M_n$ and $f = M_\infty - M_n$. Therefore, we can conclude that

$$\begin{aligned}\mathbb{E}\left((M_\infty - M_n)^2\right) &= \|(M_\infty - M_n)\|_2^2 \\ &= \lim_{r\to\infty}\|(M_{n+r} - M_n)\|_2^2 = \lim_{r\to\infty}\sum_{k=n+1}^{n+r}\mathbb{E}\left((M_k - M_{k-1})^2\right) = \sum_{k=n+1}^{\infty}\mathbb{E}\left((M_k - M_{k-1})^2\right). \end{aligned} \tag{158}$$

4.2.4 Uniform Integrability, Backwards Martingales, and the Strong Law of Large Numbers

Uniform Integrability

Uniform integrability is the requirement, on a class $\mathscr{C}$ of random variables, that the expectation of the absolute value of any of these random variables, restricted to taking large values, should be uniformly small. One could think of this as the uniform convergence to 0, over the class $\mathscr{C}$ and when $N \to \infty$, of the expectation of the modulus restricted to values larger than N. An important result on uniform integrability is that, a martingale $M = \{M_n;\ n \in \mathbb{N}\}$ can be written as $M_n = \mathbb{E}(Z \mid \mathscr{F}_n)$ for some $Z \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ if and only if it is uniformly integrable, which in turn occurs if and only if M_n converges almost surely and in $\mathscr{L}^1$. This last equivalence would be one of the tools to prove the strong law of large numbers, as we shall see later.

Lemma 4.44 (Absolute continuity)**.** Suppose $X \subseteq \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$. Then for all $\epsilon > 0$, there exists $\delta > 0$, such that for all $F \in \mathscr{F}$, if $\mathbb{P}(F) < \delta$ then $\mathbb{E}(|X|\mathbb{1}_F) < \epsilon$.

Proof. Assume for contradiction that there exists $\epsilon > 0$ and a sequence $F_n \in \mathscr{F}$, $n \in \mathbb{N}$, with $\mathbb{P}(F_n) < 2^{-n}$ (so $\mathbb{P}(F_n)$ is smaller than any $\delta > 0$) but $\mathbb{E}(|X|\mathbb{1}_{F_n}) \geq \epsilon$. Let $F := \limsup F_n = \{F_n \text{ occurs infinitely often}\}$. Clearly by construction, $\sum_{n\in\mathbb{N}} \mathbb{P}(F_n) < \infty$, so by the first Borel-Cantelli lemma (Lemma 1.36), we have that $\mathbb{P}(F) = 0$. On the other hand, X is bounded in $\mathscr{L}^1$, so the reverse Fatou's lemma (Lemma 3.4) implies that

$$\mathbb{E}(|X|\mathbb{1}_F) = \mathbb{E}\left(\limsup_{n\to\infty}(|X|\mathbb{1}_{F_n})\right) \geq \limsup_{n\to\infty} \mathbb{E}(|X|\mathbb{1}_{F_n}) \geq \epsilon. \tag{159}$$

This is impossible given $\mathbb{P}(F) = 0$, leading to a contradiction. Hence we can conclude that for all $\epsilon > 0$, there exists $\delta > 0$, such that for all $F \in \mathscr{F}$, $\mathbb{P}(F) < \delta$ necessarily implies that $\mathbb{E}(|X|\mathbb{1}_F) < \epsilon$. The proof is thus complete. ☐

Corollary 4.45. Suppose $X \in \mathscr{L}^1$. Then for all $\epsilon > 0$, there exists $K \geq 0$, such that $\mathbb{E}(|X|\mathbb{1}_{|X|>K}) < \epsilon$.

Proof. Fix an arbitrary $\epsilon > 0$. By absolute continuity (Lemma 4.44), there exists $\delta > 0$, such that for all $F \in \mathscr{F}$, if $\mathbb{P}(F) < \delta$ then $\mathbb{E}(|X|\mathbb{1}_F) < \epsilon$. By Chebyshev's inequality (Theorem 3.18), we have that $\mathbb{P}(|X| > K) \leq \mathbb{E}(|X|)/K$ for any $K > 0$, and since $\mathbb{E}(|X|)$ is finite, there exists $K > 0$ such that $\mathbb{P}(|X| > K) < \delta$. Hence $\mathbb{E}(|X|\mathbb{1}_{|X|>K}) < \epsilon$, and the proof is complete by arbitrariness of $\epsilon > 0$. ☐

Remark 4.46. An alternative proof is to use the monotone convergence theorem (Theorem 3.2).

Definition 4.47. A class $\mathscr{C}$ of random variables is said to be **uniformly integrable** if, given $\epsilon > 0$, there exists $K \geq 0$, such that $\mathbb{E}(|X|\mathbb{1}_{|X|>K}) < \epsilon$ for all $X \in \mathscr{C}$.

Remark 4.48. A uniformly integrable family is always bounded in $\mathscr{L}^1$. Indeed, by definition there exists $K_1 \geq 0$ such that $\mathbb{E}(|X|\mathbb{1}_{|X|>K_1}) \leq 1$ for all $X \in \mathscr{C}$, so that $\mathbb{E}(|X|) \leq 1 + K_1\mathbb{P}(|X| \leq K_1) \leq 1 + K_1$ for all $X \in \mathscr{C}$. However, the converse is *not* true. As a counterexample, take $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}([0,1]), \mathscr{L}([0,1]))$ and $X_n := n\mathbb{1}_{(0,1/n)}$. Clearly $\mathbb{E}(|X_n|) = 1$, and so each X_n is bounded in $\mathscr{L}^1$. However, we note that for $K \leq n$, $|X_n(\omega)| > K$ if and only if $\omega \in (0, 1/n)$, so that $\mathbb{E}(|X_n|\mathbb{1}_{|X_n|>K}) = \mathbb{E}(|X_n|\mathbb{1}_{(0,1/n)}) = \mathbb{E}(|X_n|) = 1$, which can clearly not be arbitrarily small. Hence, boundedness in $\mathscr{L}^1$ does not necessarily imply that the class of random variables is uniformly integrable.

Remark 4.49. Though boundedness in $\mathscr{L}^1$ does not guarantee uniform integrability, a class $\mathscr{C}$ of random variables is uniformly integrable if and only if the following conditions hold.

(1) $\mathscr{C}$ is bounded in $\mathscr{L}^1$, so that $A := \sup\{\mathbb{E}(|X|);\ X \in \mathscr{C}\} < \infty$.

(2) For every $\epsilon > 0$, there exists $\delta > 0$, such that if $F \in \mathscr{F}$, $\mathbb{P}(F) < \delta$, and $X \in \mathscr{C}$, then $\mathbb{E}(|X|\mathbb{1}_F) < \epsilon$.

The "only if" direction is simple. The first holds given uniform integrability. The second condition also holds because

$$\mathbb{E}(|X|\mathbb{1}_F) \leq \mathbb{E}(|X|\mathbb{1}_{F\cap\{|X|>K\}}) + \mathbb{E}(|X|\mathbb{1}_{F\cap\{|X|\leq K\}}) \leq \mathbb{E}(|X|\mathbb{1}_{|X|>K}) + K\mathbb{P}(F) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \tag{160}$$

where we have used uniform integrability for the first part of the last inequality, and picked $\delta = \epsilon/(2K)$ for the second part of the last inequality. The "if" direction is also simple, since by Chebyshev's inequality, we have that

$$\mathbb{P}(|X| > K) \leq \frac{\mathbb{E}(|X|)}{K} \leq \frac{A}{K} < \delta, \tag{161}$$

by picking $K = 2A/\delta$. Thus by taking $F = \{|X| > K\}$ in the second condition, we have that $\mathbb{E}(|X|\mathbb{1}_{|X|>K}) < \epsilon$ for all $X \in \mathscr{C}$, and the proof is thus complete.

11/13 Leccture

Lemma 4.50 (Sufficient conditions for uniform integrability)**.** A class $\mathscr{C}$ of random variables is uniformly integrable if any of the following conditions holds.

(1) (Uniform boundedness in $\mathscr{L}^p$) There exists $p > 1$ and $A \geq 0$, such that $\mathbb{E}(|X|^p) < A$ for all $X \in \mathscr{C}$.

(2) (Uniform domination by integrable random variable) There exists $Y \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ with $Y \geq 0$, such that $|X| \leq Y$ for all $X \in \mathscr{C}$.

Proof. (1) If $|X| > K$, then $|X|^{1-p} < K^{1-p}$ because $p > 1$, and thus $|X| < K^{1-p}|X|^p$. Then we can compute that

$$\mathbb{E}(|X|\mathbb{1}_{|X|>K}) < K^{1-p}\mathbb{E}(|X|^p\mathbb{1}_{|X|>K}) \leq K^{1-p}\mathbb{E}(|X|^p). \tag{162}$$

This is smaller than arbitrary $\epsilon > 0$ as long as K is sufficiently large.

(2) We know that $|X|\mathbb{1}_{|X|>K} \leq |Y|\mathbb{1}_{|X|>K} \leq |Y|\mathbb{1}_{|Y|>K}$, and thus $\mathbb{E}(|X|\mathbb{1}_{|X|>K}) \leq \mathbb{E}(|Y|\mathbb{1}_{|Y|>K})$. Moreover, since Y is integrable in $\mathscr{L}^1$, necessarily $\mathbb{E}(|Y|) < \infty$, and thus this is smaller than arbitrary $\epsilon > 0$ as long as K is sufficient large. The proof is thus complete. ☐

Lemma 4.51. Let $X \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$. Then the class $\{\mathbb{E}(X \mid \mathscr{G});\ \mathscr{G}$ a sub-σ-algebra of $\mathscr{F}\}$ is uniformly integrable.

Proof. Let $Y = \mathbb{E}(X \mid \mathscr{G})$, where $\mathscr{G}$ is a sub-σ-algebra of $\mathscr{F}$. By Jensen's inequality (Theorem 3.21), we have that $|Y| = |\mathbb{E}(X \mid \mathscr{G})| \leq \mathbb{E}(|X| \mid \mathscr{G})$ almost surely. Therefore, for any $K > 0$, we have that

$$\mathbb{E}(|Y|\mathbb{1}_{|Y|\geq K}) \leq \mathbb{E}(\mathbb{E}(|X| \mid \mathscr{G})\mathbb{1}_{|Y|\geq K}) = \underbrace{\mathbb{E}(\mathbb{E}(|X|\mathbb{1}_{|Y|\geq K} \mid \mathscr{G}))}_{\{|Y|\geq K\} \text{ is } \mathscr{G}\text{-measurable}} = \underbrace{\mathbb{E}(|X|\mathbb{1}_{|Y|\geq K})}_{\text{tower property}}. \tag{163}$$

We want to show that the above is smaller than ϵ for some $K > 0$. To do this, by the absolute continuity lemma (Lemma 4.44), we know that for all $\epsilon > 0$, there exists $\delta > 0$, such that if $\mathbb{P}(F) < \delta$ then $\mathbb{E}(|X|\mathbb{1}_F) < \epsilon$, so it suffices to show that $\mathbb{P}(|Y| \geq K) < \delta$ for some $K > 0$ and take $F = \{|Y| \geq k\}$. Indeed, if we take $K = \mathbb{E}(|X|)/\delta$, by Chebyshev's inequality we can see that $\mathbb{P}(|Y| \geq K) \leq \mathbb{E}(|Y|)/K \leq \mathbb{E}(|X|)/K = \delta$, and thus the proof is complete. ☐

Lemma 4.52. Let $X = \{X_n;\ n \in \mathbb{N}\}$ be a martingale. The following statements are equivalent.

(1) $\{X_n;\ n \in \mathbb{N}\}$ is uniformly integrable.

(2) X_n converges almost surely and in $\mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ to a limit X_∞.

(3) There exists $Z \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$, such that $X_n = \mathbb{E}(Z \mid \mathscr{F}_n)$ for all $n \in \mathbb{N}$.

Proof.

- **(1) ⇒ (2).** Suppose that X is uniformly integrable, then it is bounded in $\mathscr{L}^1$. By Doob's forward convergence theorem (Theorem 4.38), it converges almost surely to an almost surely finite random variable X_∞. We now use uniform integrability to deduce convergence in $\mathscr{L}^1$. For all $K > 0$ and $x \in \mathbb{R}$, let $\phi_K(x) := (x \wedge K) \vee (-K)$, then we have that

$$\begin{aligned}\mathbb{E}(|X_n - X_\infty|) &\leq \mathbb{E}(|\phi_K(X_\infty) - \phi_K(X_n)|) + \mathbb{E}(|\phi_K(X_\infty) - X_\infty|) + \mathbb{E}(|\phi_K(X_n) - X_n|) \\ &\leq \mathbb{E}(|\phi_K(X_\infty) - \phi_K(X_n)|) + \mathbb{E}(|X_\infty|\mathbb{1}_{|X_\infty|\geq K}) + \mathbb{E}(|X_n|\mathbb{1}_{|X_n|\geq K}).\end{aligned} \tag{164}$$

The reason is that $|\phi_K(x) - x| \leq |x|\mathbb{1}_{|x|\geq K}$. Indeed, if $|x| < K$ then $\phi_K(x) = x$. If $x \geq K$, then $\phi_K(x) = K$ and thus $|\phi_K(x) - x| = x - K < x = |x|$. If $x \leq -K$, then $\phi_K(x) = -K$ and thus $|\phi_K(x) - x| = (-x) - K < (-x) = |x|$. Hence the aforementioned inequality holds. Now for any $\epsilon > 0$, uniform integrability implies that we can choose K large enough so that $\mathbb{E}(|X_\infty|\mathbb{1}_{|X_\infty|\geq K}) + \mathbb{E}(|X_n|\mathbb{1}_{|X_n|\geq K}) < \epsilon/2$. Given this choice of K, since $\phi_K(X_n)$ also converges to $\phi_K(X_\infty)$ almost surely and is bounded by K, by the dominated convergence theorem (Theorem 3.8) we have that $\mathbb{E}(|\phi_K(X_\infty) - \phi_K(X_n)|) < \epsilon/2$ for sufficiently large n. Therefore, $\mathbb{E}(|X_n - X_\infty|) < \epsilon/2 + \epsilon/2 = \epsilon$, and we can conclude that $X_n \to X_\infty$ in $\mathscr{L}^1$.

- **(2) ⇒ (3).** Suppose that $X_n \to X_\infty$ in $\mathscr{L}^1$, and we will choose $Z := X_\infty$. Indeed, for all $F \in \mathscr{F}_n$ and $m, n \in \mathbb{N}$ such that $m \geq n$, we have that $\mathbb{E}(X_m \mathbb{1}_F) = \mathbb{E}(X_n \mathbb{1}_F)$ since X is a martingale. Bringing $m \to \infty$, then the $\mathscr{L}^1$ convergence implies that $|\mathbb{E}((X_\infty - X_n)\mathbb{1}_F)| \leq \mathbb{E}(|X_\infty - X_n|\mathbb{1}_F) \leq \mathbb{E}(|X_\infty - X_n|) \to 0$ as $n \to \infty$. Therefore, $\mathbb{E}(X_\infty \mathbb{1}_F) = \mathbb{E}(X_n \mathbb{1}_F)$, implying that $X_n = \mathbb{E}(X_\infty \mid \mathscr{F}_n)$ by definition of conditional expectation.
- **(3) ⇒ (1).** This is an immediate consequence of Lemma 4.51.

□

11/15 Lecture

Backwards Martingales

Backwards martingales are martingales whose time-set is $\mathbb{Z}^-$. More precisely, given a collection of sub-σ-algebras $\{\mathscr{G}_{-n};\ n \in \mathbb{N}\}$ such that

$$\mathscr{G}_{-\infty} := \bigcap_{k\in\mathbb{N}} \mathscr{G}_{-k} \subseteq \ldots \subseteq \mathscr{G}_{-(N+1)} \subseteq \mathscr{G}_{-N} \subseteq \ldots \subseteq \mathscr{G}_{-1} \subseteq \mathscr{G}_0, \tag{165}$$

a $\{\mathscr{G}_n\}_{n\in-\mathbb{N}}$-adapted process $\{M_n;\ n \leq 0\}$ is a backwards martingale if $\mathbb{E}(M_{n+1} \mid \mathscr{G}_n) = M_n$ for all $n \leq 0$. Backwards martingales are automatically uniformly integrable by Lemma 4.51, since $M_0 \in \mathscr{L}^1$ and $M_n = \mathbb{E}(M_0 \mid \mathscr{G}_n)$ for all $n \leq 0$. Now we can adapt Doob's upcrossing lemma (Theorem 4.35) to prove that, if $N_m([a,b], M)$ is the number of upcrossings of M from a to b between times $-m$ and 0, we have that $(b-a)\mathbb{E}(N_m([a,b],M)) \leq \mathbb{E}((M_0-a)^-)$ by considering the (forward) supermartingale $\tilde{M} := \{M_{-m+k};\ k \leq 0 \leq m\}$. As $m \to -\infty$, by the monotone convergence theorem (Theorem 3.2) we can see that $\mathbb{E}(N_m([a,b],M)) \uparrow \mathbb{E}(N([a,b],M))$, so $(b-a)\mathbb{E}(N([a,b],M)) \leq \mathbb{E}((M_0-a)^-)$. *Backwards martingales have this nice property that can directly lead to convergence, while martingales need extra assumptions according to Doob's forward convergence theorem (Theorem 4.38). This is because the upper bound in the case of backward martingales is always $(M_0 - a)^-$ that does not change with m.* By Lemma 4.36 we can thus conclude that M_n converges almost surely towards a $\mathscr{G}_{-\infty}$-measurable random variable $M_{-\infty}$ as $n \to \infty$. It is also easy to further check this convergence in $\mathscr{L}^1$ similar to Lemma 4.52. We conclude this part by stating the following theorem.

Theorem 4.53. Let M be a backwards martingale, then M_n converges almost surely and in $\mathscr{L}^1$ as $n \to -\infty$ to the random variable $M_{-\infty} = \mathbb{E}(M_0 \mid \mathscr{G}_{-\infty})$.

Proof. The explicit proof would be ignored here and should follow the discussion above. □

Strong Law of Large Numbers

Theorem 4.54 (Strong law of large numbers)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent identically distributed random variables, with $\mathbb{E}(|X_k|) < \infty$ for all $k \in \mathbb{N}$. Let μ be the common value of $\mathbb{E}(X_n)$ and write $S_n := \sum_{k=1}^n X_k$. Then $n^{-1}S_n \to \mu$ almost surely and in $\mathscr{L}^1$ as $n \to \infty$.

Proof. For all $n \in \mathbb{N}$, define $\mathscr{G}_{-n} := \sigma(S_n, S_{n+1}, S_{n+2}, \ldots) = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$ and $\mathscr{G}_\infty := \bigcap_{k\in\mathbb{N}} \mathscr{G}_{-k}$. Let us first calculate $\mathbb{E}(X_1 \mid \mathscr{G}_{-n})$, where we would use the Rôle of independence (Theorem 4.4). Since $\sigma(X_{n+1}, X_{n+2}, \ldots)$ is independent of $\sigma(\sigma(X_1), S_n)$ (which is a sub-σ-algebra of $\sigma(X_1, \ldots, X_n)$), we have that

$$\mathbb{E}(X_1 \mid \mathscr{G}_{-n}) = \underbrace{\mathbb{E}(X_1 \mid \sigma(S_n, \sigma(X_{n+1}, X_{n+2}, \ldots)))}_{\text{by definition of } \mathscr{G}_{-n}} = \underbrace{\mathbb{E}(X_1 \mid S_n)}_{\text{Rôle's}}. \tag{166}$$

Moreover, by symmetry (i.e., X_n's are independent and identically distributed), we have that $\mathbb{E}(X_k \mid S_n)$ are all equal for $0 \leq k \leq n$. Since clearly $\mathbb{E}\left(\sum_{k=1}^n X_k \mid S_n\right) = S_n$, we can write that $\mathbb{E}(X_k \mid S_n) = n^{-1}S_n$ for all $0 \leq k \leq n$. Hence by the above formula, $\mathbb{E}(X_k \mid \mathscr{G}_{-n}) = n^{-1}S_n$ for all $0 \leq k \leq n$ as well. Therefore, for all $n \in \mathbb{N}$, we have

$$\mathbb{E}(S_n \mid \mathscr{G}_{-(n+1)}) = \mathbb{E}(S_{n+1} \mid \mathscr{G}_{-(n+1)}) - \mathbb{E}(X_{n+1} \mid \mathscr{G}_{-(n+1)}) = S_{n+1} - \frac{S_{n+1}}{n+1} = \frac{nS_{n+1}}{n+1}. \tag{167}$$

Hence, letting $M_n := n^{-1}S_{-n}$, $n \leq 0$, we can see that $\{M_n;\ n \leq 0\}$ is a backwards martingale with respect to its natural filtration $\{\mathscr{G}_{-n};\ n \leq 0\}$, because $\mathbb{E}(n^{-1}S_n \mid \mathscr{G}_{-(n+1)}) = (n+1)^{-1}S_{n+1}$. Now by the convergencec property of backwards martingales (Theorem 4.53), M_n converges almost surely and in $\mathscr{L}^1$ to the random variable $M_{-\infty} = \mathbb{E}(M_0 \mid \mathscr{G}_{-\infty})$ as $n \to -\infty$. In other words, $n^{-1}S_n \to \mathbb{E}(-M_0 \mid \mathscr{G}_{-\infty})$ almost surely and in $\mathscr{L}^1$ as $n \to \infty$. Now we need to show that this limit is almost surely constant. Indeed, we observe that

$$\liminf_{n\to\infty} n^{-1}S_n = \liminf_{n\to\infty} \left(\frac{S_m}{n} + \frac{X_{m+1} + \ldots + X_n}{n} \right) = \liminf_{n\to\infty} \frac{X_{m+1} + \ldots + X_n}{n}, \tag{168}$$

so that $\liminf n^{-1}S_n$ is $\mathscr{T}_m$-measurable where $\mathscr{T}_m := \sigma(X_{m+1}, X_{m+2}, \ldots)$. By arbitrariness of $m \in \mathbb{N}$, $\liminf n^{-1}S_n$ is moreover $\mathscr{T}$-measurable, where $\mathscr{T} = \bigcap_{m\in\mathbb{N}} \mathscr{T}_m$ is the tail σ-algebra. By Kolmogorov's 0-1 law (Theorem 2.17), we thus have that $\liminf n^{-1}S_n$ is almost surely constant. Similarly, we can show that $\limsup n^{-1}S_n$ is almost surely constant, i.e., $n^{-1}S_n \to \mathbb{E}(-M_0 \mid \mathscr{G}_{-\infty})$ which is almost surely constant, so it must be equal to its mean value

$$\mathbb{E}\left(\lim_{n\to\infty} n^{-1}S_n\right) = \underbrace{\mathbb{E}\left(\lim_{n\to\infty} \mathbb{E}(X_1 \mid \mathscr{G}_{-n})\right)}_{\mathbb{E}(X_k|S_n)=n^{-1}S_n,\ \forall 0\le k\le n} = \underbrace{\lim_{n\to\infty} \mathbb{E}(\mathbb{E}(X_1 \mid \mathscr{G}_{-n}))}_{\mathscr{L}^1 \text{ convergence}} = \underbrace{\lim_{n\to\infty} \mathbb{E}(X_1)}_{\text{tower property}} = \mathbb{E}(X_1) = \mu. \tag{169}$$

The proof is thus complete. ☐

Doob's Submartingale Inequality and Doob's $\mathscr{L}^p$ Inequality

Theorem 4.55 (Doob's submartingale inequality)**.** Let $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}_{n\in\mathbb{N}}, \mathbb{P})$ be a filtered space and $X = \{X_n\}_{n\in\mathbb{N}}$ be a submartingale. For all $N \in \mathbb{N}$ and $x > 0$, we have that

$$\mathbb{P}\left(\max_{0\le i\le N} X_k \ge x\right) \le \frac{\mathbb{E}(X_N^+ \mathbb{1}_{\{T\le N\}})}{x} \le \frac{\mathbb{E}(X_N^+)}{x}, \tag{170}$$

where $T := \inf\{m \in \mathbb{N};\ X_m \ge x\}$.

Proof. Clearly $X_n^+ \in \mathscr{L}^1$, and $\{X_n^+\}_{n\in\mathbb{N}}$ is a submartingale because $\mathbb{E}(X_{n+1}^+ \mid \mathscr{F}_n) \ge (\mathbb{E}(X_{n+1} \mid \mathscr{F}_n))^+ = X_n^+$ by Jensen's inequality (Theorem 3.21). Now we note that

$$\mathbb{E}(X_{N\wedge T}^+ \mathbb{1}_{\{T\le N\}}) = \sum_{k=0}^{N} \mathbb{E}(X_k^+ \mathbb{1}_{\{T=k\}}) \le \sum_{k=0}^{N} \mathbb{E}(\mathbb{E}(X_N^+ \mid \mathscr{F}_k)\mathbb{1}_{\{T=k\}}) = \sum_{k=0}^{N} \mathbb{E}(X_N^+ \mathbb{1}_{\{T=k\}}) = \mathbb{E}(X_N^+ \mathbb{1}_{\{T\le N\}}). \tag{171}$$

Note that by definition of T, we have $\mathbb{P}(T \le N) = \mathbb{P}(\max_{0\le k\le n} X_k \ge x)$, so that

$$x\mathbb{P}\left(\max_{0\le i\le N} X_k \ge x\right) = x\mathbb{P}(T \le N) \le \underbrace{\mathbb{E}(X_{N\wedge T}^+ \mathbb{1}_{\{T\le N\}})}_{X_{N\wedge T}\le x} \le \mathbb{E}(X_N^+ \mathbb{1}_{\{T\le N\}}), \tag{172}$$

which is exactly as required by the first inequality in the conclusion. The second inequality is trivial since X_N^+ is always non-negative. The proof is thus complete. ☐

11/20 Lecture

Theorem 4.56 (Doob's $\mathscr{L}^p$ inequality)**.** Let $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}_{n\in\mathbb{N}}, \mathbb{P})$ be a filtered space and $X = \{X_n\}_{n\in\mathbb{N}}$ be a non-negative submartingale. Fix $N \in \mathbb{N}$, then for all conjugate exponents $p, q > 1$, we have

$$\left(\mathbb{E}\left(\max_{0\le i\le N} X_i^p\right)\right)^{1/p} \le q\left(\mathbb{E}(X_N^p)\right)^{1/p}. \tag{173}$$

Proof. Let $Y := \sup_{i\le N} X_i$, and note that

$$\mathbb{E}(Y^p) = \int_{\mathbb{R}} |Y|^p d\mathbb{P} = \underbrace{p\int_{\mathbb{R}} t^{p-1}\mathbb{P}(Y \ge t)dt}_{\text{Layer cake representation}}. \tag{174}$$

For any $t > 0$, define $T_t := \inf\{m \in \mathbb{N};\ X_m \geq t\}$, so by Doob's submartingale inequality (Theorem 4.55), we have that $\mathbb{P}(Y \geq t) \leq \mathbb{E}(X_N \mathbb{1}_{\{T_t \leq N\}})/t$ (we ignore the plus sign in X_N^+ because X is non-negative). Now we can see that

$$\mathbb{E}(Y^p) \leq p \int_{\mathbb{R}} t^{p-2} \mathbb{E}(X_N \mathbb{1}_{\{T_t \leq N\}}) dt = \underbrace{p\mathbb{E}\left(X_N \int_{\mathbb{R}} t^{p-2} \mathbb{1}_{\{T_t \leq N\}} dt\right)}_{\text{Fubini's theorem}} = \underbrace{p\mathbb{E}\left(X_N \int_{\mathbb{R}} t^{p-2} \mathbb{1}_{\{Y \geq t\}} dt\right)}_{T_t \leq N \iff Y \geq t}$$

$$= p\mathbb{E}\left(X_N \int_0^Y t^{p-2} dt\right) = p\mathbb{E}\left(X_N \frac{Y^{p-1}}{p-1}\right) = \frac{p}{p-1}\mathbb{E}(X_N Y^{p-1}). \tag{175}$$

Since p and q are conjugate exponents, by Hölder's inequality (Theorem 3.24) we can deduce that

$$\mathbb{E}(Y^p) \leq \frac{p}{p-1}\mathbb{E}(X_N Y^{p-1}) = q\mathbb{E}(X_N Y^{p-1}) \leq q\left(\mathbb{E}(X_N^p)\right)^{1/p}\left(\mathbb{E}(Y^{(p-1)q})\right)^{1/q} = q\left(\mathbb{E}(X_N^p)\right)^{1/p}\left(\mathbb{E}(Y^p)\right)^{1/q}, \tag{176}$$

so by moving some terms in the inequality above we can conclude that

$$\left(\mathbb{E}\left(\max_{0 \leq i \leq N} X_i^p\right)\right)^{1/p} = \left(\mathbb{E}\left(\left(\max_{0 \leq i \leq N} X_i\right)^p\right)\right)^{1/p} = \left(\mathbb{E}(Y^p)\right)^{1/p} = \left(\mathbb{E}(Y^p)\right)^{1-1/q} \leq q\left(\mathbb{E}(X_N^p)\right)^{1/p}, \tag{177}$$

and thus the proof is complete. □

5 Characteristic Function

5.1 Introduction and Elementary Properties

Definition 5.1. Given a random variable X, the **characteristic function** $\phi : \mathbb{R} \to \mathbb{C}$ of X is defined by

$$\phi(\theta) = \phi_X(\theta) = \mathbb{E}(\exp(i\theta X)) = \int \exp(i\theta x)\mu(dx) = \mathbb{E}(\cos\theta X) + i\mathbb{E}(\sin\theta X). \tag{178}$$

Theorem 5.2 (Riemann-Stieltjes integral)**.** Let $F = F_X$ be a distribution function of X. Define the integral

$$\int_{-\infty}^{\infty} g(x)dF(x) := \lim_{N\to\infty} \sum_{j=0}^{N} g(x_j)\left(F(a_{N+1}^N) - F(a_j^N)\right), \tag{179}$$

where $-\infty < a_0^N < \ldots < a_N^N < a_{N+1}^N < \infty$ is a partition of the interval $[a_0^N, a_{N+1}^N]$ with $a_0^N \to -\infty$, $a_{N+1}^N \to \infty$, and $|a_{j+1}^N - a_j^N| \to 0$ uniformly in j as $n \to \infty$. Then

$$\int_{-\infty}^{\infty} g(x)dF(x) = \int_{\mathbb{R}} g(x)\mu(dx), \tag{180}$$

where μ is the law of X.

Proof. The proof is left as an exercise and can be done using the *Standard Machine*. See also *Rudin, Principles of Mathematical Analysis, pp. 104–114*. □

Theorem 5.3. The characteristic function $\phi(\theta)$ of any probability distribution is a uniformly continuous function of θ, such that for any $\xi_1, \ldots, \xi_n \in \mathbb{C}^n$ and $\theta_1, \ldots, \theta_n \in \mathbb{R}^n$, it satisfies that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \phi(\theta_i - \theta_j)\xi_i\overline{\xi_j} \geq 0. \tag{181}$$

Proof. Note that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \phi(\theta_i - \theta_j)\xi_i\overline{\xi_j} = \sum_{i=1}^{n}\sum_{j=1}^{n} \xi_i\overline{\xi_j} \int \exp(i(\theta_i - \theta_j)x)\mu(dx) = \int \left|\sum_{j=1}^{n} \xi_j \exp(i\theta_j x)\right|^2 \mu(dx) \geq 0. \tag{182}$$

Therefore, it suffices to prove that ϕ is uniformly continuous. Indeed, we have that

$$|\phi(\theta) - \phi(\tilde{\theta})| = \left| \int \left(\exp(i\theta x) - \exp(i\tilde{\theta} x) \right) \mu(dx) \right| \le \int \underbrace{|\exp(i\tilde{\theta} x)|}_{\le 1} \underbrace{\left| \exp(i(\theta - \tilde{\theta})x) - 1 \right|}_{\le 2, \to 0} \mu(dx) \to 0, \tag{183}$$

as $|\theta - \tilde{\theta}| \to 0$, by the dominated convergence theorem (Theorem 3.8). The proof is thus complete. □

Remark 5.4. Let $\phi = \phi_X$ for a random variable X, then the following properties hold.

(1) $\phi(0) = 1$.

(2) Triangle inequality: $|\phi(\theta)| \le 1$ for all $\theta \in \mathbb{R}$.

(3) Dominated convergence theorem: $\theta \mapsto \phi(\theta)$ is continuous on $\mathbb{R}$.

(4) $\phi_{-X}(\theta) = \overline{\phi_X(\theta)}$ for all $\theta \in \mathbb{R}$.

(5) $\phi_{aX+b}(\theta) = \exp(ib\theta)\phi_X(a\theta)$ for all $a, b \in \mathbb{R}$ and $\theta \in \mathbb{R}$.

(6) If $\mathbb{E}(|X|^n) < \infty$ for some $n \in \mathbb{N}$, then ϕ is n times differentiable and $\phi^{(n)}(\theta) = \mathbb{E}((iX)^n \exp(i\theta X))$. In particular, we have that $\phi^{(n)}(0) = i^n \mathbb{E}(X^n)$.

Lemma 5.5. Let X and Y be independent random variables. Then $\phi_{X+Y}(\theta) = \phi_X(\theta)\phi_Y(\theta)$ for all $\theta \in \mathbb{R}$.

Proof. For fixed independent random variables X and Y, we have that

$$\mathbb{E}(\exp(i\theta(X+Y))) = \mathbb{E}(\exp(i\theta X)\exp(i\theta Y)) = \mathbb{E}(\exp(i\theta X))\mathbb{E}(\exp(i\theta Y)), \tag{184}$$

so we can conclude that $\phi_{X+Y}(\theta) = \phi_X(\theta)\phi_Y(\theta)$ as desired. □

5.2 Lévy Inversion Formula

A natural question is whether we can retrieve the distribution function $F(x) = \mu((-\infty, x])$ from ϕ. The answer is positive and will be shown right afterwards. However, a difficulty arises from the presense of **atoms** $a \in \mathbb{R}$, such that $\mu(\{a\}) = F(a) - F(a^-) \neq 0$. Note that there are only countably many (since there are at most n of size $\ge 1/n$).

Theorem 5.6 (Lévy inversion formula)**.** Let ϕ be the characteristic function of a random variable X with law μ and distribution function F. Then for $a < b$, we have that

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta} \phi(\theta) d\theta = \frac{\mu(\{a\})}{2} + \mu((a,b)) + \frac{\mu(\{b\})}{2} = \frac{F(b) + F(b^-)}{2} - \frac{F(a) + F(a^-)}{2}. \tag{185}$$

Proof. Let $a, b \in \mathbb{R}$ with $a < b$ and let $T > 0$. Define

$$F(a,b,T) = \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta} \phi(\theta) d\theta. \tag{186}$$

By Fubini's theorem, we can compute that

$$\begin{aligned} F(a,b,T) &= \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta} \left(\int_{\mathbb{R}} \exp(i\theta x)\mu(dx) \right) d\theta \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_{-T}^{T} \frac{\exp(i\theta(x-a)) - \exp(i\theta(x-b))}{i\theta} d\theta \right) \mu(dx) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_{-T}^{T} (\chi(\theta, x-a) - \chi(\theta, x-b)) d\theta \right) \mu(dx), \end{aligned} \tag{187}$$

where χ is the function defined by $\chi(\theta, x) = \exp(i\theta x)/(i\theta)$. Note that for all $y < z$, we have that

$$|\chi(\theta, y) - \chi(\theta, z)| \le \int_y^z \left| \frac{\partial}{\partial s} \chi(\theta, s) \right| ds = \int_y^z \underbrace{|\exp(i\theta s)|}_{\le 1} ds \le |y - z|, \tag{188}$$

so the integrand of $F(a, b, T)$ as expressed above is upper bounded by

$$\frac{1}{2\pi} \int_{-T}^{T} (\chi(\theta, x-a) - \chi(\theta, x-b)) d\theta \le \frac{1}{2\pi} \int_{-T}^{T} |a-b| d\theta = \frac{|a-b|T}{\pi}, \tag{189}$$

meaning that $F(a, b, T)$ as expressed above is indeed integrable. On the other hand, for all $y \in \mathbb{R}$, we know that $\phi \mapsto \mathrm{Re}(\chi(\theta, y))$ and $\phi \mapsto \mathrm{Im}(\chi(\theta, y))$ are respectively even and odd with respect to θ, so that

$$\int_{-T}^{T} \mathrm{Im}(\chi(\theta, y)) d\theta = 0, \qquad \text{so the imaginary part vanishes}, \tag{190}$$

$$\int_{-T}^{T} \mathrm{Re}(\chi(\theta, y)) d\theta = 2 \int_0^T \frac{\sin(\theta y)}{\theta} d\theta = 2\mathrm{sgn}(y) \int_0^{T|y|} \frac{\sin(\theta)}{\theta} d\theta = 2S(Ty), \tag{191}$$

where $\mathrm{sgn}(x) = 0, \pm 1$ depending on the sign of x, and S is defined as

$$S(u) = \int_0^u \frac{\sin(x)}{x} dx = \mathrm{sgn}(u) \int_0^{|u|} \frac{\sin(x)}{x} dx, \qquad u \in \mathbb{R}. \tag{192}$$

Note that $\int_0^\infty \sin(x)/x dx = \pi/2$ (which follows from a very beautiful mathematical proof but will be ignored here), so that as $T \to \infty$, we can see that

$$\begin{aligned} S(T(x-a)) - S(T(x-b)) &= \mathrm{sgn}(x-a) \int_0^{T|x-a|} \frac{\sin(x)}{x} dx - \mathrm{sgn}(x-b) \int_0^{T|x-b|} \frac{\sin(x)}{x} dx \\ &\to \frac{\pi}{2}(\mathrm{sgn}(x-a) - \mathrm{sgn}(x-b)) = \begin{cases} 0, & \text{if } x \in (-\infty, a) \cup (b, \infty), \\ \pi/2, & \text{if } x = a \text{ or } x = b, \\ \pi, & \text{if } x \in (a, b). \end{cases} \end{aligned} \tag{193}$$

Therefore, by the dominated convergence theorem (Theorem 3.8), we can see as $T \to \infty$ that

$$F(a, b, T) = \frac{1}{\pi} \int_{\mathbb{R}} (S(T(x-a)) - S(T(x-b))) \mu(dx) \to \frac{\mu(\{a, b\})}{2} + \mu((a, b)) = \frac{\mu(\{a\})}{2} + \mu((a, b)) + \frac{\mu(\{b\})}{2}, \tag{194}$$

because only on $\{a, b\}$ does the integrand take $\pi/2$ and only on (a, b) does the integrand take π. Now we can conclude the last equality in the theorem by observing that

$$\frac{\mu(\{a\})}{2} + \mu((a, b)) + \frac{\mu(\{b\})}{2} = \frac{F(a) - F(a^-)}{2} + (F(b^-) - F(a)) + \frac{F(b) - F(b^-)}{2} = \frac{F(b) + F(b^-)}{2} - \frac{F(a) + F(a^-)}{2}. \tag{195}$$

The proof is thus complete. ☐

Remark 5.7. Taking the limit when $a \to \infty$, note that $F(a), F(a^-) \to 0$, we can retrieve $F(b)$ from ϕ at any $b \in \mathbb{R}$ which is not an atom (*i.e.*, at which $F(b) = F(b^-)$), enabling us to retrieve F from ϕ. In particular, we can deduce that if F and G are distribution functions such that $\phi_F = \phi_G$, then necessarily $F = G$.

Corollary 5.8. Under the assumptions of the Lévy inversion formula (Theorem 5.6), and assuming furthermore that $\int_{\mathbb{R}} |\phi(\theta)| d\theta < \infty$, then X has continuous probability density function f, with

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-i\theta x) \phi(\theta) d\theta. \tag{196}$$

Proof. Let $a < b$ be such that F is continuous at a and b, then by Lévy inversion formula (Theorem 5.6), we have

$$F(b) - F(a) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta}\phi(\theta)d\theta =: G(a,b), \tag{197}$$

where we can directly write $\int_{-\infty}^{\infty}$ instead of $\lim_{T\to\infty}\int_{-T}^{T}$ because we can bound the integrand from above by

$$\left|\frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta}\phi(\theta)\right| = \left|\int_b^a \exp(-i\theta)d\theta\right| |\phi(\theta)| \leq |a-b||\phi(\theta)| = (b-a)|\phi(\theta)|. \tag{198}$$

By the dominated convergence theorem (Theorem 3.8), we can see that $G(a, b_n) \to G(a,b)$ for any sequence $b_n \to b$, so that $G(a,b)$ is continuous in b (and similarly in a), and thus $F(b) - F(a)$ is continuous in both a and b as well. Hence F is continuous. Now for all $a, b \in \mathbb{R}$ with $a < b$, we have that

$$\frac{F(b) - F(a)}{b-a} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta(b-a)}\phi(\theta)d\theta, \tag{199}$$

and again we have the upper bound

$$\left|\frac{\exp(-i\theta a) - \exp(-i\theta b)}{i\theta(b-a)}\phi(\theta)\right| = \left|\frac{1}{b-a}\int_b^a \exp(-i\theta)d\theta\right| |\phi(\theta)| \leq \frac{|a-b|}{|b-a|}|\phi(\theta)| = |\phi(\theta)|, \tag{200}$$

so that the dominated convergence theorem (Theorem 3.8) implies that

$$\begin{aligned}
\lim_{b_n\to a}\frac{F(b_n) - F(a)}{b_n - a} &= \lim_{b_n\to a}\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp(-i\theta a) - \exp(-i\theta b_n)}{i\theta(b_n-a)}\phi(\theta)d\theta \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\lim_{b_n\to a}\frac{\exp(-i\theta a) - \exp(-i\theta b_n)}{i\theta(b_n-a)}\right)\phi(\theta)d\theta \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\underbrace{\left(\lim_{b_n\to a}\frac{i\theta\exp(-i\theta b_n)}{i\theta}\right)}_{\text{L'Hospital}}\phi(\theta)d\theta = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-i\theta a)\phi(\theta)d\theta.
\end{aligned} \tag{201}$$

Therefore, F is differentiable at a with the derivative as we computed above, and is itself continuous by the dominated convergence theorem (Theorem 3.8). This is exactly the probability density function f of X, so the proof is complete. ◻

11/22 Lecture

5.3 Weak Convergence

Definition 5.9. Let $\mathscr{C}_b(\mathbb{R})$ be the space of bounded continuous functions on $\mathbb{R}$ and $\mathrm{Prob}(\mathbb{R})$ be the space of probability measures on $\mathbb{R}$. We say that a sequence of probability measures μ_n **converges weakly** to μ, denoted by $\mu_n \xrightarrow{w} \mu$, if and only if $\mu_n(h) \to \mu(h)$ for all $h \in \mathscr{C}_b(\mathbb{R})$.

Remark 5.10. Since $\{(-\infty, x];\ x \in \mathbb{R}\}$ is a π-system, we know that there is a bijection between $\mathrm{Prob}(\mathbb{R})$ and the distribution functions F through the correspondence $F(x) = \mu((-\infty, x]) = F_I(x)$ where I is the identity mapping. Therefore, we write $F_n \xrightarrow{w} F$ if and only if $\mu_n \xrightarrow{w} \mu$, where $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = \mu((-\infty, x])$.

Definition 5.11. Given a random variable X, we denote by $\mathscr{L}(X)$ the **law** of X, *i.e.*, $\mathscr{L}(X) = \mathbb{P}\circ X^{-1}$. We say that a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ **converges weakly** (or **in law**, or **in distribution**) to X, and write $X_n \xrightarrow{\mathscr{L}} X$ or $X_n \xrightarrow{w} X$, if and only if $\mathscr{L}(X_n) \xrightarrow{w} \mathscr{L}(X)$.

Lemma 5.12. If $X_n \to X$ in probability, then $\mathscr{L}(X_n) \xrightarrow{w} \mathscr{L}(X)$.

Proof. Take arbitrary $h \in \mathscr{C}_b(\mathbb{R})$, and let $M = \sup_{x\in\mathbb{R}} |h(x)|$. Then for any $K > 0$, we have that

$$\begin{aligned}\mathbb{E}(|h(X_n) - h(X)|) &= \mathbb{E}(|h(X_n) - h(X)|\mathbb{1}_{\{|X_n - X|\le\delta\}\cap\{|X|\le K\}}) + \mathbb{E}(|h(X_n) - h(X)|\mathbb{1}_{\{|X_n - X|>\delta\}\cup\{|X|>K\}}) \\ &\le \underbrace{\mathbb{E}(|h(X_n) - h(X)|\mathbb{1}_{\{|X_n - X|\le\delta\}\cap\{|X|\le K\}})}_{\mathbf{I}} + 2M(\underbrace{\mathbb{P}(|X_n - X| > \delta)}_{\mathbf{II}} + \underbrace{\mathbb{P}(|X| > K)}_{\mathbf{III}}). \end{aligned} \tag{202}$$

Firstly, we have that $\mathbb{P}(|X| \ge K) \le \mathbb{E}(|X|)/K \to 0$ as $K \to \infty$ by Chebyshev's inequality (Theorem 3.18). Secondly, h is uniformly continuous on $[-K-1, K+1]$, and thus for any $\epsilon > 0$, there exists $\delta > 0$, such that $|x - y| < \delta$ implies that $|h(x) - h(y)| < \epsilon$. Fix arbitrary $\epsilon > 0$ and choose this δ, so that

$$\mathbb{E}(|h(X_n) - h(X)|\mathbb{1}_{\{|X_n - X|\le\delta\}\cap\{|X|\le K\}}) < \epsilon. \tag{203}$$

Now assuming that $X_n \to X$ in probability, we have that $\mathbb{P}(|X_n - X| > \delta) \to 0$ as $n \to \infty$. To conclude, we have that **I** is at smaller that arbitrary $\epsilon > 0$, **II** converges to 0 as $K \to \infty$, and **III** converges to 0 as $n \to \infty$, so that $\mathbb{E}(|h(X_n) - h(X)|) \to 0$ as $n \to \infty$. Therefore, $\mathscr{L}(X_n) \xrightarrow{w} \mathscr{L}(X)$. ☐

Remark 5.13. By Theorem 2.20, the above lemma implies that if $X_n \to X$ almost surely, then $\mathscr{L}(X_n) \xrightarrow{w} \mathscr{L}(X)$.

Remark 5.14. $X_n \xrightarrow{w} X$ does not imply that $F_{X_n}(x) \to F_X(x)$ for all $x \in \mathbb{R}$. For instance, $X_n = 1/n$ converges weakly to $X = 0$, but $0 = F_{X_n}(0) \not\to F_X(0) = 1$.

Lemma 5.15. Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions on $\mathbb{R}$ and F be a distribution function on $\mathbb{R}$. Then $f_n \xrightarrow{w} f$ if and only if $F_n(x) \to F(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$ at which F is continuous.

Proof. To be done... ☐

Theorem 5.16. Suppose that $\{F_n\}_{n\in\mathbb{N}}$ is a sequence of distribution functions in $\mathbb{R}$, and that F is a distribution function on $\mathbb{R}$ such that $F_n(x) \to F(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$ at which F is continuous. Then there exists a probability triple $(\Omega, \mathscr{F}, \mathbb{P})$ carrying a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables and a random variable X, such that $F_n = F_{X_n}$ and $F = F_X$, and $X_n \to X$ almost surely.

Proof. To be done... ☐

Lemma 5.17 (Helly-Bray)**.** Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions on $\mathbb{R}$, then there exists a subsequence $\{n_i\}_{i\in\mathbb{N}}$ and a right-continuous non-decreasing function F on $\mathbb{R}$ with $0 \le F \le 1$, such that $F_{n_i}(x) \to F(x)$ as $i \to \infty$ for all $x \in \mathbb{R}$ at which F is continuous.

Proof. To be done... ☐

Remark 5.18. Note that F is not necessarily a distribution function in the lemma above, since $F(x)$ is not required to converge to 0 and 1 when $x \to -\infty$ and $x \to \infty$, respectively. Therefore, we cannot write the convergence in the lemma above as a weak convergence in $\text{Prob}(\mathbb{R})$, but rather in $\text{Prob}(\overline{\mathbb{R}})$ where $\overline{\mathbb{R}} = [-\infty, \infty]$.

Definition 5.19. A sequence $\{F_n\}_{n\in\mathbb{N}}$ of distribution functions is called **tight** if, for all $\epsilon > 0$, there exists $K > 0$, such that for all $n \in \mathbb{N}$, we have that

$$\mu_n([-K, K]) = F_n(K) - F_n(-K^-) > 1 - \epsilon. \tag{204}$$

Lemma 5.20. Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions.

(1) If $F_n \xrightarrow{w} F$ for some distribution function F, then $\{F_n\}_{n\in\mathbb{N}}$ is tight.

(2) If $\{F_n\}_{n\in\mathbb{N}}$ is tight, then there exists a subsequence $\{n_i\}_{i\in\mathbb{N}}$, such that $F_{n_i} \xrightarrow{w} F$ as $i \to \infty$ for some distribution function F.

Proof. To be done... ☐

5.4 Lévy's Convergence Theorem

Theorem 5.21 (Lévy's convergence theorem)**.** Let $\{F_n\}_{n\in\mathbb{N}}$ be a sequecence of distribution functions, with corresponding characteristic functions $\{\phi_n\}_{n\in\mathbb{N}}$. Assume that $g(\theta) = \lim_{n\to\infty} \phi_n(\theta)$ exists for all $\theta \in \mathbb{R}$ and that g is continuous at 0. Then $g = \phi_F$ for some distribution function F, and $F_n \xrightarrow{w} F$.

Proof. TO BE DONE... □

5.5 Proof of the Central Limit Theorem (CLT)

Theorem 5.22 (Central limit theorem)**.** Let $X_1, \ldots, X_n$ be independent identically distributed random variables with mean 0 and variance 1. Define $S_n := \sum_{k=1}^n X_k$, then

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1), \tag{205}$$

where $\mathscr{N}(0,1)$ denotes the normal distribution centered 0 of variance 1.

Proof. TO BE DONE... □

A Recitations

9/1 Recitation

Example A.1. Let Ω be a non-empty set. If $\mathscr{F}_i$ is a σ-algebra for each $i \in I$, where I is a non-empty set of indices. Show that $\mathscr{F} = \bigcap_{i\in I} \mathscr{F}_i$ is a σ-algebra.

Proof. We have that $\varnothing, \Omega \in \mathscr{F}_i$, $\forall i \in I$, thus $\varnothing, \Omega \in \mathscr{F}$. Take an arbitrary $A \in \mathscr{F}$, we have that $A \in \mathscr{F}_i$, $\forall i \in I$, and thus $A^C \in \mathscr{F}_i$, $\forall i \in I$, so that $A^C \in \mathscr{F}$. Now take arbitrary $\{A_n\}_{n\geq 1} \subseteq \mathscr{F}$, then $A_n \in \mathscr{F}_i$, $\forall i \in I$, $\forall n \geq 1$. Therefore, $\bigcup_{n\geq 1} A_n \in \mathscr{F}_i$, $\forall i \in I$. This means that $\bigcup_{n\geq 1} A_n \in \mathscr{F}$, and the proof is complete. □

Example A.2. Let Ω be a non-empty set.

(1) Show that if $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \dots$ are σ-algebras, then $\mathscr{F} = \bigcup_{n\geq 1} \mathscr{F}_n$ is an algebra.

(2) Show that $\mathscr{F}$ is not a σ-algebra.

Proof. (1) We have that $\varnothing, \Omega \in \mathscr{F}_n$, $\forall n$. Therefore, $\varnothing, \Omega \in \mathscr{F}$. Take an arbitrary $A \in \bigcup_{n\geq 1} \mathscr{F}_n$, then there exists $n_0 \geq 1$, such that $A \in \mathscr{F}_{n_0}$. This means that $A^C \in \mathscr{F}_{n_0}$, and thus $A^C \in \mathscr{F}$. Now take arbitrary $A, B \in \mathscr{F}$, then there exists $n_1, n_2 \geq 1$, such that $A \in \mathscr{F}_{n_1}$ and $B \in \mathscr{F}_{n_2}$. We have that $A \cup B \in \mathscr{F}_{\max(n_1,n_2)}$, so that $A \cup B \in \mathscr{F}$. This completes the proof.

(2) Take a counterexample $\mathscr{F}_n = \mathscr{P}(\{1, \dots, n\})$, the set of all subsets of $\{1, \dots, n\}$. We can see that $\mathscr{F}$ is then the set of subsets of $\mathbb{N}$ with *finite* cardinality. Now consider $A_n = \{n\} \in \mathscr{F}_n$, $n \geq 1$, then the countable union $\bigcup_{n\geq 1} A_n = \mathbb{N} \notin \mathscr{F}$. Therefore $\mathscr{F}$ is not a σ-algebra in this case. □

Example A.3. Let μ be a finitely additive measure in an algebra $\mathscr{S}$. Consider $(A_n)_{n\geq 1}$ disjoint sets in $\mathscr{S}$ such that $A = \bigcup_{n\geq 1} A_n \in \mathscr{S}$. Show that

$$\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n). \tag{206}$$

Proof. Note that $\mathscr{S}$ is not necessarily a σ-algebra. Decompose

$$A = \bigcup_{n=1}^{\infty} A_n = \underbrace{\left(\bigcup_{n=1}^{k} A_n\right)}_{\in\mathscr{S}} \cup \underbrace{\left(\bigcup_{n=k+1}^{\infty} A_n\right)}_{\in\mathscr{S}}. \tag{207}$$

Therefore, we can see that

$$\mu(A) = \mu\left(\bigcup_{n=1}^{k} A_n\right) + \underbrace{\mu\left(\bigcup_{n=k+1}^{\infty} A_n\right)}_{\geq 0} \geq \mu\left(\bigcup_{n=1}^{k} A_n\right). \tag{208}$$

If $\mu(A) = \infty$, we are done, so we assume that $\mu(A) < \infty$. Taking $k \to \infty$, we can conclude that

$$\mu(A) \geq \mu\left(\bigcup_{n=1}^{\infty} A_n\right). \tag{209}$$

□

Example A.4. Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a family of measurable sets $\{A_n\}_{n\geq 1}$ such that $\mathbb{P}(A_n) = 1$. Show that

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1. \tag{210}$$

Proof. We can see that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^C\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n^C) = \sum_{n=1}^{\infty} 0 = 0, \tag{211}$$

so the proof is complete by taking the complement. ☐

Example A.5. For a sequence of events $(A_n)_{n\geq 1}$ with $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$ and $0 < c < 1$, show that there exists a subsequence $(n_k)_{k\geq 1}$ with $n_k \to \infty$, such that

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} A_{n_k}\right) > c. \tag{212}$$

Proof. Fix $\epsilon_0 = 1 - c > 0$, and take $\epsilon_i = \epsilon_0/2^i$, $\forall i \geq 1$. From the given limit, we can see that $\forall \epsilon > 0$, $\exists n_0 \geq 1$, such that $\forall n \geq n_0$, we have that $|1 - \mathbb{P}(A_n)| < \epsilon$. Therefore, we can take n_k such that $|1 - \mathbb{P}(A_{n_k})| < \epsilon_k$. Then, we have that

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_{n_k}^C\right) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_{n_k}^C) < \sum_{k=1}^{\infty} \frac{\epsilon_0}{2^k} = \epsilon_0 = 1 - c. \tag{213}$$

Then we can conclude that

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_{n_k}\right) = 1 - \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_{n_k}^C\right) > c \tag{214}$$

so the proof is complete. ☐

9/8 Recitation

Example A.6. Let $\Omega = \mathbb{R}$ and $\mathscr{F}$ be a set of all $A \in \Omega$ such that either A or A^C is countable. Define

$$\mathbb{P}(A) = \begin{cases} 0, & \text{if } A \text{ is countable,} \\ 1, & \text{if } A^C \text{ is countable.} \end{cases} \tag{215}$$

Show that $(\Omega, \mathscr{F}, \mathbb{P})$ forms a probability space.

Proof. We first need to show that $\mathscr{F}$ is a σ-algebra on Ω. Clearly, $\varnothing \in \Omega$ since $\varnothing$ is countable. Take arbitrary $A, A_n \in \mathscr{F}$, then $A^C \in \mathscr{F}$ because if A is countable then A^C is countable, and otherwise $(A^C)^C = A$ is countable. Moreover, $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{F}$ because if all A_n's are countable then their countable union is still countable, and otherwise we can see that $\left(\bigcup_{n\in\mathbb{N}} A_n\right)^C = \bigcap_{n\in\mathbb{N}} (A_n)^C$ is countable since some $(A_n)^C$ would then be countable. Up till now we have shown that $\mathscr{F}$ is a σ-algebra on Ω. Now we check that $\mathbb{P}$ is a countably additive measure. Take arbitrary $A_n \in \mathscr{F}$ disjoint, $n \in \mathbb{N}$, then there are two cases.

- If all A_n's are countable, then $\bigcup_{n\in\mathbb{N}} A_n$ is countable, so $0 = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = 0$.
- If there exists A_{n_0} uncountable, then $\bigcup_{n\in\mathbb{N}} A_n$ is uncountable. Since the other A_n's are disjoint with A_{n_0}, each of them belongs to $A_{n_0}^C$, but note that $A_{n_0}^C$ is countable. Therefore, all other A_n's must be countable, so $1 = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\neq n_0} \mathbb{P}(A_n) + \mathbb{P}(A_{n_0}) = 1$.

This means that $\mathbb{P}$ is countably additive. Furthermore, $\Omega^C = \varnothing$ is countable, so that $\mathbb{P}(\Omega) = 1$, meaning that $\mathbb{P}$ is a probability measure, so the proof is complete. ☐

Example A.7. Let $\Omega = \{1, 2, 3, 4\}$, $\mathscr{F} = \mathscr{P}(\Omega)$, and $\mathbb{P}$ the probability measure such that

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}). \tag{216}$$

(1) Show that the two classes of events $\mathscr{C}_1 = \{\{1, 2\}\}$ and $\mathscr{C}_2 = \{\{2, 3\}, \{2, 4\}\}$ are independent by $\sigma(\mathscr{C}_1)$ and $\sigma(\mathscr{C}_2)$ are not.

(2) We have learned a theorem that stablish an extra condition in which we can conclude $\sigma(\mathscr{C}_1)$ and $\sigma(\mathscr{C}_2)$ are independent. Identify the condition that is missing in this example $\mathscr{C}_1$ and $\mathscr{C}_2$.

Proof. Computing the probabilities of the intersections, we have that

$$\mathbb{P}(\{1,2\}\cap\{2,3\}) = \mathbb{P}(\{2\}) = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = \mathbb{P}(\{1,2\})\cdot\mathbb{P}(\{2,3\}), \tag{217}$$

$$\mathbb{P}(\{1,2\}\cap\{2,4\}) = \mathbb{P}(\{2\}) = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = \mathbb{P}(\{1,2\})\cdot\mathbb{P}(\{2,4\}). \tag{218}$$

Therefore, $\mathscr{C}_1$ and $\mathscr{C}_2$ are independent. Now note that $\sigma(\mathscr{C}_1) = \{\varnothing, \{1,2\}, \{3,4\}, \Omega\}$, and $\sigma(\mathscr{C}_2) = \mathscr{P}(\Omega)$. Then, we can see that

$$\mathbb{P}(\underbrace{\{1,2\}}_{\in\sigma(\mathscr{C}_1)}\cap\underbrace{\{1,2\}}_{\in\sigma(\mathscr{C}_2)}) = \mathbb{P}(\{1,2\}) = \frac{1}{2} \neq \frac{1}{2}\cdot\frac{1}{2} = \mathbb{P}(\underbrace{\{1,2\}}_{\in\sigma(\mathscr{C}_1)})\cdot\mathbb{P}(\underbrace{\{1,2\}}_{\in\sigma(\mathscr{C}_2)}), \tag{219}$$

which means that $\sigma(\mathscr{C}_1)$ and $\sigma(\mathscr{C}_2)$ are not independent. To answer the second part of this question, this is because $\mathscr{C}_2$ is not a π-system (clearly $\{2,3\}\cap\{2,4\} = \{2\} \notin \mathscr{C}_2$). □

Example A.8. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and X be a measurable function. Show that μ defined in $(\mathbb{R}, \mathscr{B})$ given by $\mu(B) = \mathbb{P}(X^{-1}(B))$ is a probability measure.

Proof. To see that μ is a probability measure, we need to check that it is countably additive. Take arbitrary $B_n \in \mathscr{B}$ disjoint, $n \in \mathbb{N}$, then we have that

$$\mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right)\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} X^{-1}(B_n)\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(X^{-1}(B_n)) = \sum_{n\in\mathbb{N}} \mu(B_n), \tag{220}$$

so μ is indeed countably additive. Clearly $\mu(\mathscr{B}) = \mathbb{P}(X^{-1}(\mathscr{B})) = \mathbb{P}(\Omega) = 1$, so μ is a probability measure and the proof is complete. □

Remark A.9. In the above proof, we have used that $f^{-1}(A\cup B) = f^{-1}(A)\cup f^{-1}(B)$. This is indeed true.

- $\subseteq$ Take an arbitrary $x \in f^{-1}(A\cup B)$, then $f(x) \in A\cup B$. This means that either $f(x) \in A$ or $f(x) \in B$. Then either $x \in f^{-1}(A)$ or $x \in f^{-1}(B)$, implying that $x \in f^{-1}(A)\cup f^{-1}(B)$.
- $\supseteq$ Take an arbitrary $x \in f^{-1}(A)\cup f^{-1}(B)$, then either $x \in f^{-1}(A)$ or $x \in f^{-1}(B)$. This means that either $f(x) \in A$ or $f(x) \in B$, implying that $f(x) \in A\cup B$. Thus $x \in f^{-1}(A\cup B)$.

Example A.10. Prove that, if $\{A_n\}_{n\geq 1}$ are independent events, then

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \prod_{n=1}^{\infty} \mu(A_n). \tag{221}$$

Proof. Let $B_k = \bigcap_{n=1}^{k} A_n$, then by independence, we have that

$$\mathbb{P}(B_k) = \mathbb{P}\left(\bigcap_{n=1}^{k} A_n\right) = \prod_{n=1}^{k} \mathbb{P}(A_n). \tag{222}$$

Note that B_n is a decreasing event, such that $B_n \downarrow \bigcap_{n=1}^{\infty} A_n$. By monotone convergence, we thus have that

$$\lim_{k\to\infty} \mathbb{P}(B_k) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right). \tag{223}$$

Therefore, we can conclude that

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{k\to\infty} \prod_{n=1}^{k} \mathbb{P}(A_n) = \prod_{n=1}^{\infty} \mathbb{P}(A_n), \tag{224}$$

and the proof is complete. □

Example A.11. Prove that, if $\{A_n\}_{n\geq 1}$ are independent events, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = 1 - \prod_{n=1}^{\infty}(1-\mu(A_n)). \tag{225}$$

Proof. We know that if $\{A_n\}_{n\geq 1}$ are independent, then $\{A_n^C\}_{n\geq 1}$ are independent as well. Therefore, using the conclusion of Example A.10, we can deduce that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = 1-\mu\left(\bigcap_{n=1}^{\infty} A_n^C\right) = 1-\prod_{n=1}^{\infty}\mu(A_n^C) = 1-\prod_{n=1}^{\infty}(1-\mu(A_n)), \tag{226}$$

so the proof is complete. □

Example A.12. Prove that

$$\limsup_{n\to\infty} A_n^C = \left(\liminf_{n\to\infty} A_n\right)^C, \qquad \limsup_{n\to\infty}\left(\liminf_{k\to\infty}(A_n\cap A_k^C)\right) = \varnothing. \tag{227}$$

Proof. By definition, we have that

$$\left(\liminf_{n\to\infty} A_n\right)^C = \left(\bigcup_{n\geq 1}\bigcap_{m\geq n} A_m\right)^C = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m^C = \limsup_{n\to\infty} A_n^C, \tag{228}$$

$$\begin{aligned}\limsup_{n\to\infty}\left(\liminf_{k\to\infty}(A_n\cap A_k^C)\right) &= \limsup_{n\to\infty}\left(\bigcup_{k\geq 1}\bigcap_{p\geq k}(A_n\cap A_p^C)\right) = \limsup_{n\to\infty}\left(A_n\cap\left(\bigcup_{k\geq 1}\bigcap_{p\geq k} A_p^C\right)\right)\\ &= \limsup_{n\to\infty}\left(A_n\cap\liminf_{k\to\infty} A_k^C\right) = \left(\limsup_{n\to\infty} A_n\right)\cap\left(\liminf_{k\to\infty} A_k^C\right) = \varnothing,\end{aligned} \tag{229}$$

so the proof is complete. □

9/15 Recitation

Example A.13. In a sequence of independent Bernoulli random variables $\{X_n\}_{n\geq 1}$ with

$$\mathbb{P}(X_n = 1) = p = 1-\mathbb{P}(X_n = 0). \tag{230}$$

Let A_n be the event that a run of n consecutive 1's occur between the (2^n-1)th and 2^{n+1}th trials. If $p\geq 1/2$, show that there is probability 1 that infinitely many A_n occurs.

Proof. Let B_n be the event that there exists a block of all 1's, where we split the interval $[2^n-1, 2^{n+1}]$ into disjoint blocks of length n (from left to right). Then we have that

$$\mathbb{P}(A_n) \geq \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{k=1}^{\lfloor 2^n/n\rfloor} C_k\right) = \sum_{k=1}^{\lfloor 2^n/n\rfloor} p^n = \left\lfloor\frac{2^n}{n}\right\rfloor p^n \geq \frac{2^n}{2n}p^n \geq \frac{1}{2n}. \tag{231}$$

This implies that $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \infty$. Furthermore, $\{A_n\}_{n\geq 1}$ are clearly independent, because the desired intervals, with different values of n, are disjoint. By the second Borel-Cantelli lemma (Lemma 1.37), we can conclude that A_n occurs infinitely often, which completes the proof. □

Remark A.14. In problems that require the use of the second Borel-Cantelli lemma, it is common to use the strategy of "making blocks" as in the example above.

Example A.15. Consider $E = \{0,1\}$ with $\mathbb{P}(X=1) = p$ and $0<p<1$. Let $\Omega = E^{\mathbb{N}}$ with Borel σ-algebra. For a fixed finite sequence $\omega = \{\omega_k\}_{k=1}^{t}$ with $\omega_k\in\{0,1\}$ and $t<\infty$, we define

$$A_n = \left\{\{X_k\}_{k\geq 1}\in\Omega;\ X_n = \omega_1,\ldots,X_{n+t-1} = \omega_t\right\}. \tag{232}$$

Show that $\mathbb{P}(A_n\ i.o.) = 1$.

Proof. Let $B_m = \left\{\{X_k\}_{k\geq 1} \in \Omega;\ X_{(m-1)t+1} = \omega_1, \ldots, X_{mt} = \omega_t\right\}$. Note that B_m are independent, because the blocks are non-intersecting. Moreover, we have that

$$\sum_{m=1}^{\infty} \mathbb{P}(B_m) = \sum_{m=1}^{\infty} \mathbb{P}(B_1) = \infty. \tag{233}$$

By the second Borel-Cantelli lemma (Lemma 1.37), we can see that B_n occurs infinitely often. Now since $\{B_n\ i.o.\} \subseteq \{A_n\ i.o.\}$m we can conclude that A_n occurs infinitely often as well. ☐

Example A.16. Let $0 < \alpha \leq 1$, μ be a probability measure, and $\{A_n\}_{n\geq 1}$ be a sequence of events in a σ-algebra $\mathscr{F}$. If $\mu(A_n\ i.o.) \geq \alpha$ and $B \in \mathscr{F}$ such that $\sum_{n=1}^{\infty} \mu(A_n \cap B) < \infty$, prove that $\mu(B) \leq 1 - \alpha$.

Proof. By the first Borel-Cantelli lemma (Lemma 1.36), we can see that $\mu(\limsup A_n \cap B) = 0$. This means that

$$\begin{aligned} 0 = \mu\left(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} (A_n \cap B)\right) = \mu\left(B \cap \left(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_n\right)\right) &= \mu(B) + \underbrace{\mu\left(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_n\right)}_{\geq \alpha} - \underbrace{\mu\left(B \cup \left(\bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_n\right)\right)}_{\leq 1} \\ &\geq \mu(B) + \alpha - 1. \end{aligned} \tag{234}$$

This necessarily implies that $\mu(B) \leq 1 - \alpha$, so the proof is complete. ☐

Example A.17. Assume that $\{A_n\}_{n\geq 1}$ are independent with $\mathbb{P}(A_n) \geq 1$ for all n.

(1) Show that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = 1 - \prod_{n=1}^{\infty} (1 - \mathbb{P}(A_n)). \tag{235}$$

(2) Show that if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then

$$\prod_{n=1}^{\infty} (1 - \mathbb{P}(A_n)) = 0. \tag{236}$$

(3) Show that if $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) < 1$, then $\mathbb{P}(A_n\ i.o.) = 0$.

Proof. (1) See Example A.11.

(2) We have that

$$0 \leq \prod_{n=1}^{\infty} (1 - \mathbb{P}(A_n)) \leq \prod_{n=1}^{\infty} \exp(-\mathbb{P}(A_n)) = \exp\left(-\sum_{n=1}^{\infty} \mathbb{P}(A_n)\right) = 0. \tag{237}$$

(3) By the first part, we can see that $\prod_{n=1}^{\infty}(1 - \mathbb{P}(A_n)) > 0$. Using the contrapositive of the second part, clearly $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Then by the first Borel-Cantelli lemma (Lemma 1.36), we can conclude that $\mathbb{P}(A_n\ i.o.) = 0$, so the proof is complete. ☐

Example A.18. Assume that $\{A_n\}_{n\geq 1}$ are independent with $\mathbb{P}(A_n) < 1$ for all n. Show that, if $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = 1$, then $\mathbb{P}(A_n\ i.o.) = 1$.

Proof. We assume for contradiction that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Then by definition, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$, such that $\sum_{n=N}^{\infty} \mathbb{P}(A_n) < \epsilon$. Take $\epsilon = 1$, we have that

$$0 < 1 - \sum_{n=N}^{\infty} \mathbb{P}(A_n) \leq \prod_{n=N}^{\infty} (1 - \mathbb{P}(A_n)) \leq \prod_{n=1}^{\infty} (1 - \mathbb{P}(A_n)) = \underbrace{1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)}_{\text{1st part of Example A.17}} = 0, \tag{238}$$

which is clearly a contradiction. Therefore, $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. By the second Borel-Cantelli lemma (Lemma 1.37), we can thus conclude that $\mathbb{P}(A_n\ i.o.) = 1$, and the proof is complete. ☐

9/22 Recitation

Example A.19. Let $\Omega = \{1,2,3\}$ and $\mathscr{F} = \{\varnothing, \Omega, \{1,2\}, \{3\}\}$. Define $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, such that $X(i) = i$ for $i = 1,2,3$. Is $(\Omega, \mathscr{F})$ a measurable space? Is X a random variable?

Proof. Since $\mathscr{F} = \sigma(\{1,2\})$ which means that it is a σ-algebra of Ω, we can conclude that $(\Omega, \mathscr{F})$ is a measurable space. Now note that a random variable must be measurable. However, we have many connterexamples, for instance, $(3/2, 4) \in \mathscr{B}(\mathbb{R})$ but $X^{-1}((3/2,4)) = \{2,3\} \notin \mathscr{F}$. Therefore, X is not a random variable. □

Example A.20. Show that any distribution function has at most a countable number of discontinuity points.

Proof. Let F be a distribution function and A be the set of all discontinuity points of F. Note that F would be right-continuous and monotonically increasing (Lemma 2.8). The discontinuity points should thus satisfy that

$$\lim_{x\uparrow x_0} F(x) < \lim_{x\downarrow x_0} F(x) = F(x_0). \tag{239}$$

Now let A_n be the set of discontinuity points x_0 such that $\lim_{x\downarrow x_0} F(x) - \lim_{x\uparrow x_0} F(x) \geq 1/n$. Since we know that $0 \leq F(x) \leq 1$ (Lemma 2.8), partition it into $I_j = [j/n, (j+1)/n)$ for $0 \leq j \leq n-2$ and $I_{n-1} = [(n-1)/n, 1]$. Now assume that some interval I_j contains two values $F(x_1)$ and $F(x_2)$ with $x_1 < x_2$ in A_n. It is clear that each partition I_j can contain the values of at most one $x_0 \in A_n$. Therefore, $|A_n| \leq n$. Note that the set of all discontinuities $A = \bigcup_{n\in\mathbb{N}} A_n$, so that A is countable and the proof is complete. □

Example A.21. Show that, if X and Y are random variables, then $X+Y$ is a random variable.

Proof. We need to show that $X+Y$ is measurable, so it suffices to check that $\{X+Y \leq x\} \in \mathscr{F}$ for any $x \in \mathbb{R}$. Since X and Y are random variables, $\{X < x\} \in \mathscr{F}$ and $\{Y < x\} \in \mathscr{F}$ for any $x \in \mathbb{R}$. Then we have that

$$\{X+Y < x\} = \bigcup_{q\in\mathbb{Q}} (\{X < q\} \cap \{Y < x-q\}) \in \mathscr{F}, \tag{240}$$

since $\mathscr{F}$ is a σ-algebra. This completes the proof. Alternatively, see Lemma 2.4. □

Example A.22. Let X and Y be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ and take $A \in \mathscr{F}$. Show that

$$Z(\omega) = \begin{cases} X(\omega), & \text{if } \omega \in A, \\ Y(\omega), & \text{if } \omega \in A^C, \end{cases} \tag{241}$$

is a random variable.

Proof. For a fixed $B \in \mathscr{B}(\mathbb{R})$, we have that

$$Z^{-1}(B) = \{\omega;\ X(\omega) \in B \text{ and } \omega \in A\} \cup \{\omega;\ Y(\omega) \in B \text{ and } \omega \notin A\} = (X^{-1}(B) \cap A) \cup (Y^{-1}(B) \cap A^C). \tag{242}$$

Since X and Y are random variables, we have that $X^{-1}(B) \in \mathscr{F}$ and $Y^{-1}(B) \in \mathscr{F}$. Since $\mathscr{F}$ is a σ-algebra, we thus have that $Z^{-1}(B) \in \mathscr{F}$. By arbitrariness of $B \in \mathscr{B}(\mathbb{R})$, we can conclude that Z is measurable, and thus a random variable. The proof is now complete. □

Example A.23. Assume that $\mathscr{A}$ is a non-empty set and $\mathscr{S}$ is a σ-algebra. Show that if $\mathscr{S} = \sigma(\mathscr{A})$, then

$$\sigma(X) = \sigma(X^{-1}(\mathscr{A})) := \left\{X^{-1}(A);\ A \in \mathscr{A}\right\}. \tag{243}$$

Proof. Recall that $\sigma(X) := X^{-1}(\mathscr{S}) := \left\{X^{-1}(B);\ B \in \mathscr{S}\right\}$.

- $\supseteq$ Take an arbitrary $C \in X^{-1}(\mathscr{A})$, then there exists $A \in \mathscr{A}$, such that $C = X^{-1}(A)$. Since $A \in \mathscr{A} \subseteq \mathscr{S}$, we have that $C \in \sigma(X)$. Therefore, $X^{-1}(A) \subseteq \sigma(X)$, and thus $\sigma(X)$ is a σ-algebra containing $X^{-1}(A)$. But note that $\sigma(X^{-1}(\mathscr{A}))$ is the smallest σ-algebra containing $X^{-1}(\mathscr{A})$, so necessarily $\sigma(X) \supseteq \sigma(X^{-1}(\mathscr{A}))$.
- $\subseteq$ Take an arbitrary $C \in X^{-1}(\mathscr{S})$, then there exists $B \in \mathscr{S}$, such that $C = X^{-1}(B)$. Since $\mathscr{S} = \sigma(\mathscr{A})$, B can necessarily be written as a countable combination of unions and complements of elements of $\mathscr{A}$. Since X is a random variable and thus measurable, these set operations are preserved by X^{-1}. Therefore, C can be written as a countable combination of unions and complements of elements of $X^{-1}(\mathscr{A})$. In other words, $C \in \sigma(X^{-1}(\mathscr{A}))$, and thus $\sigma(X^{-1}(\mathscr{A}))$ is a σ-algebra containing $X^{-1}(\mathscr{S})$. But note that $\sigma(X)$ is the smallest σ-algebra containing $X^{-1}(\mathscr{S})$, so necessarily $\sigma(X) \subseteq \sigma(X^{-1}(\mathscr{A}))$.

The proof is thus complete. □

9/24 Recitation

Example A.24. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with $\Omega = \{-1,1\}^{\mathbb{N}}$. For each sequence $X = \{X_n\}_{n\geq 1} \in \Omega$, let $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$ for all $n \geq 1$. Now let $\{a_n\}_{n\geq 1}$ be a fixed sequence of real numbers. Show that

$$\mathbb{P}\left(\sum_{n=1}^{\infty} a_n X_n \text{ converges}\right) \in \{0,1\}. \tag{244}$$

Proof. It suffices to prove that $A = \{\sum_{n=1}^{\infty} a_n X_n \text{ converges}\}$ is a tail event. Indeed, $\sum_{n=1}^{\infty} a_n X_n$ converges if and only if $\sum_{n=k}^{\infty} a_n X_n$ converges for all $k \geq 1$. Therefore, $A \in \mathscr{T}_k$ for all $k \geq 1$ and thus A is a tail event. Now by Kolmogorov's 0-1 law (Theorem 2.17, we have our desired result. □

Example A.25. Let $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$ with $\Omega = \{-1,0,1\}$. Consider $\{X_n\}_{n\geq 1}$ to be a sequence of independent identically distributed random variables with $X_1 = 0$. Define $S_n = \sum_{k=1}^{n} X_k$ and consider the following events.

i) $\lim_{n\to\infty} S_n = \infty$,

ii) $\lim_{n\to\infty} S_n = -\infty$,

iii) $-\infty = \liminf_{n\to\infty} S_n < \limsup_{n\to\infty} S_n = \infty$.

iv) $-\infty < \liminf_{n\to\infty} S_n < \limsup_{n\to\infty} S_n = \infty$.

v) $-\infty = \liminf_{n\to\infty} S_n < \limsup_{n\to\infty} S_n < \infty$.

vi) $-\infty < \liminf_{n\to\infty} S_n \leq \limsup_{n\to\infty} S_n < \infty$.

vii) $S_n = 0$ almost surely for all $n \geq 1$.

The problem is as follows.

(1) Events i) through vi) are mutually exclusive. Prove that precisely one of them occurs with probability 1.

(2) Replace vi) with vii) and prove the same as the previous part.

Proof. (1) Note that events i) through vi) are all tail events. Indeed, there are all related only to the limiting behavior of X_n, and in particular, changing finitely many values of X_k does not change the values of the limits, limit superiors, and limit inferiors. Now since $\{X_n\}_{n\geq 1}$ are independent, we can apply Kolmogorov's 0-1 law (Theorem 2.17) to see that each of them has probability 0 or 1. But since the events are mutually exclusive, precisely one of them occurs with probability 1, so the proof is complete.

(2) To be done... □

Example A.26. Show that if f is continuous, X is finite almost surely, and $X_n \to X$ almost surely, then $f(X_n) \to f(X)$ almost surely.

Proof. Since $X_n \to X$ almost surely, we have that $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$. Let ω be an event in $\{\lim_{n\to\infty} X_n = X\}$, then $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. Since f is continuous, we can see that $\lim_{n\to\infty} f(X_n(\omega)) = f(X(\omega))$. Therefore, we can duduce that

$$1 = \mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) \leq \underbrace{\mathbb{P}\left(\lim_{n\to\infty} f(X_n) = f(X)\right)}_{\text{contains all events } \omega} \leq 1, \tag{245}$$

which necessarily means that $\mathbb{P}(\lim_{n\to\infty} f(X_n) = f(X)) = 1$, *i.e.*, $f(X_n) \to f(X)$ almost surely. The proof is thus complete. □

Example A.27. Let $\{X_n\}_{n\geq 1}$ and X be random variables, such that for every $\epsilon > 0$, we have that

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \epsilon) < \infty. \tag{246}$$

Show that $X_n \to X$ almost surely.

Proof. By the first Borel-Cantelli lemma (Lemma 1.36), we have that $\mathbb{P}(|X_n - X| \geq \epsilon \ i.o.) = 0$. Note that if for some ω we have that $X_n(\omega) \not\to X(\omega)$, this means that there exists $\epsilon > 0$, such that for any $N \in \mathbb{N}$, there exists $n \geq N$, such that $|X_n(\omega) - X(\omega)| \geq \epsilon$. Therefore, we can write that

$$\mathbb{P}\left(\{\omega;\ X_n(\omega) \not\to X(\omega)\}\right) = \mathbb{P}\left(\bigcup_{\epsilon>0}\bigcap_{N\in\mathbb{N}}\bigcup_{n\geq N}\{\omega;\ |X_n(\omega) - X(\omega)| \geq \epsilon\}\right) = \mathbb{P}\left(\bigcup_{\epsilon>0}\{|X_n - X| \geq \epsilon\ i.o.\}\right) = 0. \quad (247)$$

Therefore, we can conclude that $X_n \to X$ almost surely, and the proof is complete. ☐

Example A.28. Let $\{X_n\}_{n\geq 1}$ be independent identically distributed random variables with uniform distribution on $[0, 1]$, *i.e.*, $\mathbb{P}(X_n < t) = t$ for all $t \in [0, 1]$. Show that $X_n^{1/n} \to 1$ almost surely.

Proof. Fix an arbitrary $0 < \epsilon < 1$. Since each X_n is uniformly distributed on $[0, 1]$, we have that

$$\mathbb{P}(|X_n^{1/n} - 1| \geq \epsilon) = \mathbb{P}(X_n^{1/n} \leq 1 - \epsilon) = \mathbb{P}(X_n \leq (1 - \epsilon)^n) = (1 - \epsilon)^n. \quad (248)$$

Then we can check that

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n^{1/n} - 1| \geq \epsilon) = \sum_{n=1}^{\infty}(1 - \epsilon)^n = \frac{1 - \epsilon}{\epsilon} < \infty. \quad (249)$$

By the first Borel-Cantelli lemma (Lemma 1.36), we can thus see that $\mathbb{P}(\{|X_n^{1/n} - 1| \geq \epsilon\ i.o.\}) = 0$. By arbitrariness of $\epsilon > 0$, we can thus conclude that $\mathbb{P}(\{X_n^{1/n} \to 1\}) = 1$, *i.e.*, $X_n^{1/n} \to 1$ almost surely. The proof is thus complete. ☐

Example A.29 (Simplified version of Slutsky's theorem)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. If $\{X_n\}_{n\geq 1}$ is a sequence of random variables such that X_n converges to X in probability, where X is finite almost surely and ϕ is a continuous function, show that $\phi(X_n) \to \phi(X)$ in probability.

Proof. Fix an arbitrary $\epsilon > 0$. Note that for any $\eta > 0$, we can write that

$$\begin{aligned}\mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) &= \mathbb{P}(\{|f(X_n) - f(X)| \geq \epsilon\} \cap \{|X_n - X| \geq \eta\}) + \mathbb{P}(\{|f(X_n) - f(X)| \geq \epsilon\} \cap \{|X_n - X| < \eta\}) \\ &\leq \mathbb{P}(\{|X_n - X| \geq \eta\}) + \mathbb{P}(\{|f(X_n) - f(X)| \geq \epsilon\} \cap \{|X_n - X| < \eta\}). \quad (250)\end{aligned}$$

We argue that X_n being far away from X is a rare event (the first part), and when X_n is close to X, we exploit the continuity of ϕ (the second part). Since X_n converges to X in probability, we have that $\lim_{n\to 0}\mathbb{P}(|X_n - X| \geq \eta) = 0$ for any $\eta > 0$, so the first part converges to 0. For the second part, since X is finite almost surely, X_n must take values in some compact set $[-c, c]$. Since ϕ is continuous, it is uniformly continuous on $[-c, c]$. This means that there exists $\eta > 0$, such that for any $|\phi(x) - \phi(y)| < \epsilon$ whenever $|x - y| < \eta$. Take this specific $\eta > 0$, so the second part would be 0. By the previous arguments, we can see that $\mathbb{P}(|f(X_n) - f(X)| \geq \epsilon)$ converges to 0 as $n \to \infty$, which means that $f(X_n)$ converges to $f(X)$ in probability. The proof is thus complete. ☐

10/13 Recitation

Example A.30. Let $(\Omega, \mathscr{F}, \mu)$ be a probability measure. Assume that $f_n \geq 0$ is integrable and $f_n \to f$ pointwise, that is, $f_n(x) \to f(x)$ for all $x \in \Omega$. Assume that $\mu(f_n) \to \mu(f) < \infty$. Show that $\mu(f_n\mathbb{1}_A) \to \mu(f\mathbb{1}_A)$ for every $A \in \mathscr{F}$ using Fatou's lemma.

Proof. Take an arbitrary $A \in \mathscr{F}$, then by Fatou's lemma (Lemma 3.3), we can see that

$$\mu(f\mathbb{1}_A) = \mu\left(\liminf_{n\to\infty} f_n\mathbb{1}_A\right) \leq \liminf_{n\to\infty}\mu(f_n\mathbb{1}_A), \qquad \mu(f\mathbb{1}_{A^C}) = \mu\left(\liminf_{n\to\infty} f_n\mathbb{1}_{A^C}\right) \leq \liminf_{n\to\infty}\mu(f_n\mathbb{1}_{A^C}). \quad (251)$$

The second inequality moreover implies that

$$\limsup_{n\to\infty}\mu(f_n\mathbb{1}_A) = \limsup_{n\to\infty}\mu(f_n - f_n\mathbb{1}_{A^C}) = \limsup_{n\to\infty}\mu(f_n) - \liminf_{n\to\infty}\mu(f_n\mathbb{1}_{A^C}) \leq \mu(f) - \mu(f\mathbb{1}_{A^C}) = \mu(f\mathbb{1}_A). \quad (252)$$

But $\liminf \mu(f_n\mathbb{1}_A) \leq \limsup \mu(f_n\mathbb{1}_A)$, so clearly we have that $\liminf \mu(f_n\mathbb{1}_A) = \limsup \mu(f_n\mathbb{1}_A) = \mu(f\mathbb{1}_A)$. Therefore, we can conclude that $\mu(f_n\mathbb{1}_A) \to \mu(f\mathbb{1}_A)$ for any $A \in \mathscr{F}$, and the proof is thus complete. ☐

Example A.31. Let $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu)$ be a probability space and $f \geq 0$ be integrable on $\mathbb{R}$. Show that, for every $\epsilon > 0$, there exists $E \in \mathscr{B}(\mathbb{R})$ with $E \neq \mathbb{R}$, such that

$$\int_E f d\mu > \int_{\mathbb{R}} f d\mu - \epsilon. \tag{253}$$

Proof. Let $E_n := \{x \in \mathbb{R};\ f(x) > 1/n\}$ and let $f_n := f\mathbb{1}_{E_n}$. We can see that f_n is non-negative and clearly $f_n \uparrow f$. By monotone convergence theorem (Theorem 3.2), we can conclude that

$$\int_{E_n} f d\mu = \int_{\mathbb{R}} f_n d\mu \uparrow \int_{\mathbb{R}} f d\mu, \tag{254}$$

and thus the proof is complete by the characterization of limits. □

Example A.32. Let $\{f_n\}_{n\geq 1}$ be an increasing sequence functions such that $f_n \uparrow f$ and $\int f_1^- d\mu < \infty$. Show that

$$\int f_n d\mu \uparrow \int f d\mu. \tag{255}$$

Proof. Since $\{f_n\}_{n\geq 1}$ is increasing, we have that $f_n^- = (-f_n) \vee 0 \geq (-f_{n+1}) \vee 0 = f_{n+1}^-$, *i.e.*, $\{f_n^-\}_{n\geq 1}$ is decreasing. In particular, $f_n^- \leq f_1^-$ and $f^- \leq f_1^-$, so that $\int f_n^- d\mu \leq \int f_1^- d\mu < \infty$ and $\int f^- d\mu \leq \int f_1^- d\mu < \infty$. By the monotone convergence theorem (Theorem 3.2), since f_n^+ and $f_1^- - f_n^-$ are both non-negative and increasing, we have that

$$\int f_n^+ d\mu \uparrow \int f^+ d\mu, \qquad \int (f_1^- - f_n^-) d\mu \uparrow \int (f_1^- - f^-) d\mu. \tag{256}$$

Putting them together, we have that $\int (f_1^- + f_n) d\mu \uparrow \int (f_1^- + f) d\mu$, and the desired result can be concluded by noting that $\int f_1^- d\mu < \infty$. The proof is thus complete. □

Example A.33. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\{X_t\}_{t\in\mathbb{R}}$ be a family of random variables, such that (1) $X_t(\omega) : \mathbb{R} \to \mathbb{R}$ is continuous in t for each $\omega \in \Omega$, and (2) there exists an $\mathscr{L}^1$ random variable Y such that $|X_t| \leq Y$ for all $t \in \mathbb{R}$. Show that the function $F : \mathbb{R} \to \mathbb{R}$ given by $F(t) = \mathbb{E}(X_t)$ is well-defined and continuous.

Proof. For ant $t \in \mathbb{R}$, note that

$$\int_\Omega |X_t| d\mu \leq \int_\Omega Y d\mu < \infty. \tag{257}$$

Hence, F is well-defined. Now take $t_0 \in \mathbb{R}$ and $t_n \to t_0$. Since $X_t(\omega)$ is continuous in t for each $\omega \in \Omega$, we have that $X_{t_n}(\omega) \to X_{t_0}(\omega)$ for each $\omega \in \Omega$. Since $|X_{t_n}| \leq Y$ for all $n \in \mathbb{N}$, by dominated convergence theorem (Theorem 3.8), we can see that

$$F(t_n) = \mathbb{E}(X_{t_n}) = \int_\Omega X_{t_n} d\mu \to \int_\Omega X_{t_0} d\mu = \mathbb{E}(X_{t_0}) = F(t_0), \tag{258}$$

implying that F is continuous. The proof is thus complete. □

Example A.34. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $I \subseteq \mathbb{R}$ be an open interval, and $t_0 \in I$. Moreover, let $\{X_t\}_{t\in I}$ be a family of random variables, such that (1) $X_t(\omega) : I \to \mathbb{R}$ is differentiable in t on I, $\mathbb{P}$-almost surely, and (2) there exists an $\mathscr{L}^1$ random variable Y such that $|X_t| + |dX_t/dt| \leq Y$ $\mathbb{P}$-almost surely for all $t \in I$. Show that the function $F : I \to \mathbb{R}$ given by $F(t) = \mathbb{E}(X_t)$ is well-defined and differentiable at t_0 with

$$F'(t_0) = \mathbb{E}\left(\frac{d}{dt} X_t \bigg|_{t=t_0} \right). \tag{259}$$

Proof. Since $|X_t| \leq Y$ $\mathbb{P}$-almost surely for all $t \in I$, we have that F is well-defined. Take $t_0 \in \mathbb{R}$ and $t_n \to t_0$. By definition, we have that

$$F'(t_0) = \lim_{n\to\infty} \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim_{n\to\infty} \frac{\mathbb{E}(X_{t_n}) - \mathbb{E}(X_{t_0})}{t_n - t_0} = \lim_{n\to\infty} \mathbb{E}\left(\frac{X_{t_n} - X_{t_0}}{t_n - t_0} \right). \tag{260}$$

By the mean value theorem, there exists σ_n between t_0 and t_n, such that

$$\frac{X_{t_n}-X_{t_0}}{t_n-t_0}=\frac{d}{dt}X_t\bigg|_{t=\sigma_n}. \tag{261}$$

Moreover since $|dX_t/dt|\le Y$ $\mathbb{P}$-almost surely, we have that $|X_{t_n}-X_{t_0}|/|t_n-t_0|\le Y$ $\mathbb{P}$-almost surely as well, for all $n\ge 1$. Hence by the dominated convergence theorem (Theorem 3.8), we can see that

$$\begin{aligned}F'(t_0)=\lim_{n\to\infty}\mathbb{E}\left(\frac{X_{t_n}-X_{t_0}}{t_n-t_0}\right)=\lim_{n\to\infty}\int\left|\frac{X_{t_n}-X_{t_0}}{t_n-t_0}\right|d\mu=\int\lim_{n\to\infty}\left|\frac{X_{t_n}-X_{t_0}}{t_n-t_0}\right|d\mu=\int\left(\frac{d}{dt}X_t\bigg|_{t=t_0}\right)d\mu\\ =\mathbb{E}\left(\frac{d}{dt}X_t\bigg|_{t=t_0}\right).\end{aligned} \tag{262}$$

The proof is thus complete. □

Example A.35. Given an example of a strictly inequality in Fatou's lemma (Lemma 3.3).

Proof. For $n\ge 1$, define $f_n:=\mathbb{1}_{[n,n+1)}$, then clearly $f_n\to 0$. Let μ be the uniform measure in $\mathbb{R}$, then

$$\mu\left(\liminf_{n\to\infty}f_n\right)=\int_{-\infty}^{\infty}0d\mu=0. \tag{263}$$

On the other hand, we have that

$$\liminf_{n\to\infty}\mu(f_n)=\liminf_{n\to\infty}\int_{-\infty}^{\infty}\mathbb{1}_{[n,n+1)}d\mu=\liminf_{n\to\infty}\int_{n}^{n+1}d\mu=\liminf_{n\to\infty}1=1>\mu\left(\liminf_{n\to\infty}f_n\right), \tag{264}$$

so this is a strictly inequality in Fatou's lemma (Lemma 3.3). □

Example A.36. Suppose that $(\Omega,\mathscr{F},\mathbb{P})$ is a probability space and $\{A_n\}_{n\ge1}$ is a collection of events such that $\mathbb{P}(A_i\cap A_j)=0$ for all $i\ne j$. Show that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}\mathbb{P}(A_n). \tag{265}$$

Proof. By the series convergence theorem (Theorem 3.10), we can see that

$$\sum_{n=1}^{\infty}\mathbb{P}(A_n)=\sum_{n=1}^{\infty}\int_\Omega\mathbb{1}_{A_n}d\mathbb{P}=\int_\Omega\sum_{n=1}^{\infty}\mathbb{1}_{A_n}d\mathbb{P}=\underbrace{\int_\Omega\mathbb{1}_{\bigcup_{n=1}^\infty A_n}d\mathbb{P}}_{A_n\text{'s disjoint}}=\mathbb{P}\left(\bigcup_{n=1}^{\infty}A_n\right), \tag{266}$$

so the proof is complete. □

10/18 Midterm Exercises

Example A.37. Let $\Omega=\mathbb{R}$ and $\mathscr{F}$ be a set of subsets of $\mathbb{R}$ containing $\varnothing$ and all $A\in\mathscr{P}(\mathbb{R})$ with the property that, if $x\in A$, then $x\pm1,x\pm2,\dots$ belong to A. Show that $\mathscr{F}$ is a σ-algebra.

Proof. We already have $\varnothing\in\mathscr{F}$. Fix an arbitrary $A\in\mathscr{F}$, then for any $x\in A^C$, $x+n\notin A$ for all $n\in\mathbb{Z}$ because otherwise there would be a contradiction. In other words, any $x\in A^C$ satisfies that $x+n\in A^C$ for all $n\in\mathbb{Z}$, so that $A^C\in\mathscr{F}$. Now fix arbitrary $\{A_m\}_{m\in\mathbb{N}}\subseteq\mathscr{F}$. For any $x\in\bigcup_{m\in\mathbb{N}}A_m$, there must exist $m\in\mathbb{N}$ such that $x\in A_m$. But since $A_m\in\mathscr{F}$, necessarily $x+n\in A_m$ for all $n\in\mathbb{Z}$, and thus $x+n\in\bigcup_{m\in\mathbb{N}}A_m$ for all $n\in\mathbb{Z}$. Hence $\bigcup_{m\in\mathbb{N}}A_m\in\mathscr{F}$ and the proof is thus complete. □

Example A.38. Let Ω be an infinite countable set and $\mathscr{F}=\mathscr{P}(\Omega)$ be the collection of all subsets of Ω. Define, for each $A\in\mathscr{F}$, that

$$\mu(A)=\begin{cases}0, & \text{if } A \text{ is finite},\\ \infty, & \text{if } A \text{ is infinite}.\end{cases} \tag{267}$$

Show that the set function μ is additive but not σ-additive.

Proof. Take arbitrary $A, B \in \mathscr{F}$. If A and B are both finite, then $A \cup B$ is also finite, and thus $\mu(A \cup B) = 0 = \mu(A) + \mu(B)$. If A and B are both infinite, then $A \cup B$ is also infinite, and thus $\mu(A \cup B) = \infty = \mu(A) + \mu(B)$. Finally if one of them is finite, without loss of generality we assume that A is finite and B is infinite, then $A \cup B$ is infinite, and thus $\mu(A \cup B) = \infty = \mu(A) + \mu(B)$. The above argument has shown that μ is additive. However, it is not countably additive. As a counterexample, take $A_n = \{n\}$ for each $n \in \mathbb{N}$, then we can see that

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} \{n\}\right) = \mu(\mathbb{N}) = \infty \neq 0 = \sum_{n\in\mathbb{N}} 0 = \sum_{n\in\mathbb{N}} \mu(\{n\}) = \sum_{n\in\mathbb{N}} \mu(A_n). \tag{268}$$

Hence, μ is additive but not countably additive and the proof is thus complete. □

Example A.39. Show that if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then the series $\sum_{n=1}^{\infty} \mathbb{1}_{A_n}$ is finite almost everywhere.

Proof. By the first Borel-Cantelli lemma (Lemma 1.36), we have that $\mathbb{P}(A_n \ i.o.) = 0$. In other words, almost everywhere there exists $N \in \mathbb{N}$, such that for all $n \geq N$, we have that A_n does not hold, *i.e.*, $\mathbb{1}_{A_n} = 0$. Hence, we can conclude that $\sum_{n=1}^{\infty} \mathbb{1}_{A_n}$ is finite almost everywhere, and the proof is complete. □

Example A.40. Find an example of a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and function $X : \Omega \to \mathbb{R}$, such that $|X|$ is a random variable but X is not.

Proof. Let $\Omega = \mathbb{R}$ and $\mathscr{F} = \{\varnothing, \mathbb{R}\}$. Take some $A \neq \varnothing, \mathbb{R}$ and define $X(\omega) = 1$ for any $\omega \in A$ and $X(\omega) = -1$ otherwise. Clearly $|X(\omega)| = 1$ for all $\omega \in \mathbb{R}$. Hence, $|X|^{-1}(B)$ is either $\mathbb{R}$ or $\varnothing$ depending on whether $1 \in B$ or not, *i.e.*, $|X|^{-1}(B) \in \mathscr{F}$. This means that $|X|$ is measurable, and thus a random variable. However, we can see that $X^{-1}(\{1\}) = A \notin \mathscr{F}$, so X is not measurable. and thus not a random variable. □

Example A.41. Let X be a random variable with probability density function given by

$$f_X(x) = \begin{cases} cx^2, & \text{if } |x| \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{269}$$

Find c, $\mathbb{E}(X)$, and $\mathbb{P}(X \geq 1/2)$.

Proof. Necessarily $\int_{-\infty}^{\infty} f_X(x) = 1$, from which we can obtain that $c = 3/2$ and detailed computations will be ignored here. Consequently, $\mathbb{E}(X)$ can be computed as $\mathbb{E}(X) = \int_{-1}^{1} x f_X(x)dx$ and the answer is 0. Finally, $\mathbb{P}(X \geq 1/2)$ can be computed as $\mathbb{P}(X \geq 1/2) = \int_{1/2}^{1} f_X(x)dx$ and the answer is 7/16. □

Example A.42. Let X be a random variable with mean $m = \mathbb{E}(X)$ and variance $\sigma^2 = \mathbb{E}((X - \mathbb{E}(X))^2)$.

(1) Let $u \in \mathbb{R}$. Show that for every $\alpha \geq 0$, we have that

$$\mathbb{P}(X - m \geq \alpha) \leq \frac{\sigma^2 + u^2}{(\alpha + u)^2}. \tag{270}$$

(2) Find u that minimizes the right-hand side. Conclude Cantelli's inequality that

$$\mathbb{P}(X - m \geq \alpha) \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}. \tag{271}$$

Proof. (1) By Chebyshev's inequality (Theorem 3.18), we have that

$$\mathbb{P}(X - m \geq \alpha) = \mathbb{P}((X - m + u)^2 \geq (\alpha + u)^2) \leq \frac{\mathbb{E}((X - m + u)^2)}{(\alpha + u)^2} = \frac{\sigma^2 + 2u\mathbb{E}(X - m) + u^2}{(\alpha + u)^2} = \frac{\sigma^2 + u^2}{(\alpha + u)^2}. \tag{272}$$

(2) $u = \sigma^2/\alpha$ is the critical number (derivative is zero) of the right-hand side, and we can check the derivatives to see that this u minimizes the right-hand side. By arbitrariness of u in the previous part, we can conclude that

$$\mathbb{P}(X - m \geq \alpha) \leq \frac{\sigma^2 + \frac{\sigma^4}{\alpha^2}}{\left(\alpha + \frac{\sigma^2}{\alpha}\right)^2} = \frac{\sigma^2(\alpha^2 + \sigma^2)}{(\alpha^2 + \sigma^2)^2} = \frac{\sigma^2}{\sigma^2 + \alpha^2}, \tag{273}$$

and the proof is thus complete. □