B8.1: Martingales Through Measure Theory
by Zhongmin Qian

Contents

1 Measures and integration

Let us begin with a few complementary results in Lebesgue's theory of measures and integrations. These notes should be read along with my "Part A Integration" lecture notes if you haven't taken the Lebesgue Integration course yet. The latter provides the background material which I assume you already know.

The conventions about the extended real line $[-\infty, \infty]$ will be applied in these notes, where two symbols $-\infty$ and ∞ are added to $\mathbb{R}$, so that $[-\infty, \infty] = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$. For every $a \in \mathbb{R}$, $-\infty < a < \infty$,

$$a + \infty = \infty + a = \infty, \; a - \infty = -\infty + a = -\infty.$$

$$\frac{a}{-\infty} = \frac{a}{\infty} = 0,$$

but $\frac{\infty}{\infty}$, $\frac{a}{0}$, $\infty - \infty$, $\infty + (-\infty)$ and $(-\infty) + \infty$ are not defined, while $0 \cdot \infty = -\infty \cdot 0 = 0$, $-\infty + (-\infty) = -\infty$ and $\infty + \infty = \infty$.

Let us generalize the notions of measures and outer measures introduced in Part A Integration with modification, for our convenience for this course.

1. Measures. Let Ω be a (sample) space, and $\mathcal{R}$ be a collection of some subsets of Ω. Suppose $\mathcal{R}$ contains an empty set denoted by $\varnothing$. A function $\mu : \mathcal{R} \to [0, \infty]$ is called a *measure* on $\mathcal{R}$ if

1.1) $\mu(\varnothing) = 0$,

1.2) $\mu(A) \le \mu(B)$ for $A, B \in \mathcal{R}$ such thst $A \subseteq B$, and

1.3) μ is *countably additive*:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

for any $A_i \in \mathcal{R}$ $(i = 1, 2, \cdots)$ which are disjoint, such that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$.

2. Outer measures. If the condition of countable additivity 1.3) is replaced by countable sub-additivity, then we obtain the definition of outer measures. That is, μ is an *outer measure* on $\mathcal{R}$, if 1) and 2) hold, and μ is a countably sub-additive:

$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i)$$

for any $A_i, A \in \mathcal{R}$ $(i = 1, 2, \cdots)$ such that $A \subset \cup_{i=1}^{\infty} A_i$.

3. Finite measures and σ-finite measures. A measure μ on $\mathcal{R}$ is finite if $\mu(E) < \infty$ for every $E \in \mathcal{R}$. μ is called σ-finite on $\mathcal{R}$ if there is a sequence of subsets $E_i \in \mathcal{R}$ such that $\bigcup_{i=1}^{\infty} E_i = \Omega$ and $\mu(E_i) < \infty$ for every $i = 1, 2, \cdots$. If $\mu(\Omega) = 1$, then μ is called a *probability measure* on $\mathcal{R}$.

4. Ring, algebra, σ-algebras and measurable spaces. We haven't imposed any algebraic structures yet on $\mathcal{R}$. Several notions may be introduced via set-theoretic operations: $\cup$, $\cap$ and complementary operation $\setminus$. A collection $\mathcal{R}$ of subsets of Ω is called a *ring* over Ω if $E_1 \cup E_2 \in \mathcal{R}$ and $E_1 \setminus E_2 \in \mathcal{R}$ for any $E_1, E_2 \in \mathcal{R}$. A ring $\mathcal{R}$ is an *algebra* if the total space $\Omega \in \mathcal{R}$. An algebra $\mathcal{F}$ over Ω is called a σ-algebra (or called a σ-field) if $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$ for any $E_i \in \mathcal{F}$. If $\mathcal{F}$ is a σ-algebra over Ω, then $(\Omega, \mathcal{F})$ is called a measurable space.

If $\mathcal{A}$ is a non-empty collection of some subsets of Ω, then there is a unique σ-algebra over Ω, denoted by $\sigma\{\mathcal{A}\}$, which possesses the following properties: (1) $\mathcal{A} \subseteq \sigma\{\mathcal{A}\}$, and (2) if $\mathcal{F}$ is a σ-algebra over Ω containing $\mathcal{A}$, then $\sigma\{\mathcal{A}\} \subseteq \mathcal{F}$. In fact

$$\sigma\{\mathcal{A}\} = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra containing } \mathcal{A}\} .$$

$\sigma\{\mathcal{A}\}$ is the smallest σ-algebra containing $\mathcal{A}$, called the σ-algebra generated by $\mathcal{A}$.

5. Measure spaces and probability spaces. If $(\Omega, \mathcal{F})$ is a measurable space and μ is a measure on $\mathcal{F}$, then $(\Omega, \mathcal{F}, \mu)$ is called a *measure space.* If $\mu(\Omega) = 1$ then $(\Omega, \mathcal{F}, \mu)$ is called a probability space. In this case Ω is called a sample space (of fundamental events), an element A in the σ-algebra $\mathcal{F}$ is called an event, and $\mu(A)$ is called the probability that the event A occurs. A probability measure μ is usually denoted by a blackboard letter $\mathbb{P}$.

6. Measurable functions. $\mathscr{B}(\mathbb{R}^n)$ denotes the Borel σ-algebra on $\mathbb{R}^n$, which is the smallest σ-algebra containing open subsets. A function $f : \Omega \to [-\infty, \infty]$ is measurable with respect to a σ-field $\mathcal{F}$, or simply called $\mathcal{F}$-measurable, if

$$f^{-1}(G) = \{f \in G\} \equiv \{\omega \in \Omega : f(\omega) \in G\}$$

belongs to $\mathcal{F}$ for every $G \in \mathscr{B}(\mathbb{R})$, and both $f^{-1}(\infty)$ and $f^{-1}(-\infty)$ belong to $\mathcal{F}$ as well.

7. Structure of measurable functions. A simple (measurable) function φ on $(\Omega, \mathcal{F})$ is a (real valued) function on Ω which can be written as $\varphi = \sum_{k=1}^{n} c_k 1_{E_k}$ for some n, some constants c_k and some $E_k \in \mathcal{F}$. A function $f : \Omega \to [0, \infty]$ is $\mathcal{F}$-measurable, if and only if there is an increasing sequence of non-negative, $\mathcal{F}$-measurable simple functions φ_n such that $\varphi_n \uparrow f$ everywhere on Ω.

8. Definition of Lebesgue's integrals. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. The Lebesgue theory of integration, developed in Part A Integration, may be applied to the measure μ. Let us recall quickly the procedure of defining Lebesgue's integrals. First define integrals for simple functiona, namely, if $\varphi = \sum_{j=1}^{m} c_j 1_{E_j}$ is a non-negative ($\mathcal{F}$-measurable) *simple function* on Ω, where $c_i \geq 0$ and $E_i \in \mathcal{F}$ for $i = 1, \cdots, m$, then $\int_E \varphi d\mu = \sum_{i=1}^{m} c_i \mu(E_i)$. If $f : \Omega \to [0, \infty]$ is a non-negative $\mathcal{F}$-measurable function, then

$$\int_\Omega f d\mu = \sup \left\{ \int_E \varphi d\mu : \varphi \leq f \text{ where } \varphi = \sum_{i=1}^{m} c_i 1_{E_i} \text{ and } c_i \geq 0, E_i \in \mathcal{F} \right\} .$$

9. Integrable functions. If f is non-negative measurable and if $\int_\Omega f d\mu < \infty$, then we say f is (Lebesgue) integrable on Ω with respect to the measure μ, denoted by $f \in L^1(\Omega, \mathcal{F}, \mu)$, $f \in L^1(\Omega, \mu)$, $L^1(\Omega)$ or simply by $f \in L^1$ if the measure space in question is clear. If $f : \Omega \to [-\infty, \infty]$ is $\mathcal{F}$-measurable, so are $f^+ = f \vee 0$, $f^- = (-f) \vee 0$ and $|f| = f^+ - f^-$. If both f^+ and f^- are integrable, then we say f is integrable, denoted by $f \in L^1(\Omega, \mathcal{F}, \mu)$ etc., and define its (Lebesgue) integral by

$$\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu.$$

If $f : \Omega \to \mathbb{C}$ is a complex, $\mathcal{F}$-measurable function: $f = u + \sqrt{-1}v$, then f is integrable if both real part u and imaginary part v are integrable against the measure μ, and in this case, the Lebesgue integral of f is defined by

$$\int_\Omega f d\mu = \int_\Omega u d\mu + \sqrt{-1} \int_\Omega v d\mu.$$

$L^1(\Omega, \mathcal{F}, \mu)$ denotes the vector space of all $\mathcal{F}$-measuable (real or complex valued) integrable function on $(\Omega, \mathcal{F}, \mu)$.

The convergence theorems are applicable to a measure space $(\Omega, \mathcal{F}, \mu)$, and they may be stated as the following.

10. Monotone Convergence Theorem (MCT, due to Lebesgue and Levi). Suppose $f_n : \Omega \to [0, \infty]$ are non-negative, measurable, and suppose $f_{n+1} \geq f_n$ almost everywhere on Ω for all n, then

$$\int_\Omega \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu = \sup_n \int_\Omega f_n d\mu.$$

In particular, if $\left\{\int_\Omega f_n d\mu\right\}$ is bounded above, then $\lim_{n\to\infty} f_n$ is integrable.

11. Series version of MCT (due to Lebesgue and Levi). This is very useful and is handy in applications. If a_n are *non-negative* and *measurable*, then

$$\int_\Omega \sum_n a_n d\mu = \sum_n \int_\Omega a_n d\mu.$$

12. Fatou's Lemma. Suppose $f_n : \Omega \to [0, \infty]$ are non-negative and measurable, then

$$\int_\Omega \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

13. Lebesgue's Dominated Convergence Theorem (DCT). Suppose $f_n : \Omega \to [-\infty, \infty]$ (or $f_n : \Omega \to \mathbb{C}$) are measurable, $f_n \to f$ almost everywhere, and suppose there is an integrable (control) function g such that $|f_n| \leq g$ almost everywhere for all n, then f_n are integrable and

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu.$$

14. Reverse Fatou's Lemma. Suppose f_n and g are integrable, and $f_n \leq g$ almost surely for $n = 1, 2, \cdots$. Then $g - f_n$are non-negative, and $\liminf(g - f_n) = g - \limsup f_n$. Applying Fatou's lemma to $g - f_n$ we obtain

$$\begin{aligned}\int_\Omega \left[g - \limsup_{n\to\infty} f_n\right] d\mu &\leq \liminf_{n\to\infty} \left[\int_\Omega g - \int_\Omega f_n d\mu\right] \\ &= \int_\Omega g d\mu - \limsup_{n\to\infty} \int_\Omega f_n d\mu\end{aligned}$$

which in particular yields that

$$\int_\Omega g d\mu - \limsup_{n\to\infty} \int_\Omega f_n d\mu \geq 0$$

so that $\limsup_{n\to\infty} \int_\Omega f_n d\mu \leq \int_\Omega g d\mu$. If $\limsup_{n\to\infty} \int_\Omega f_n d\mu > -\infty$, then

$$\int_\Omega g d\mu - \limsup_{n\to\infty} \int_\Omega f_n d\mu < \infty$$

so that $g - \limsup_{n\to\infty} f_n$ is integrable, and $\limsup_{n\to\infty} \int_\Omega f_n d\mu \leq \int_\Omega \limsup_{n\to\infty} f_n d\mu$. Let us state what we have proved as the following.

Theorem 1.1 *(*Reverse Fatou's Lemma*) Suppose f_n and g are integrable, and $f_n \leq g$ almost surely for $n = 1, 2, \cdots$, and suppose* $\limsup_{n\to\infty} \int_\Omega f_n d\mu > -\infty$*, then* $\limsup_{n\to\infty} f_n$ *is integrable and*

$$\int_\Omega \limsup_{n\to\infty} f_n d\mu \geq \limsup_{n\to\infty} \int_\Omega f_n d\mu.$$

15. Notations. If $f \in L^1(\Omega, \mathcal{F}, \mu)$ or if f is non-negative and measurable, then we also use $\mathbb{E}^\mu(f)$, $\mu(f)$ or $\mathbb{E}(f)$ to denote Lebesgue integral $\int_\Omega f d\mu$. If $A \in \mathcal{F}$, then $(A, A \cap \mathcal{F}, \mu)$ is a measure space too. In this case $\int_A f d\mu$ concides with $\int_\Omega f 1_A d\mu$, which will be denoted by $\mathbb{E}^\mu [f : A]$ or by $\mathbb{E}[f : A]$ if the measure in question is clear.

16. The L^p space for $p \in [1, \infty]$ can be defined over a measure space. When dealing with L^p-spaces, we identify an $\mathcal{F}$-measurable function f on $(\Omega, \mathcal{F}, \mu)$ with its equivalent class of all $\mathcal{F}$-measurable functions which are equal to f almost surely on Ω. Then $L^p(\Omega, \mathcal{F}, \mu)$ is the vector space of all $\mathcal{F}$-measurable functions f such that $|f|^p$ is μ-integrable, equipped with the L^p-norm: if $p \in [1, \infty)$, then

$$\|f\|_p = \left(\int_\Omega |f|^p d\mu \right)^{\frac{1}{p}} = (\mathbb{E}|f|^p)^{\frac{1}{p}}.$$

If $p = \infty$, then

$$\|f\|_\infty = \inf \{K : |f| \leq K \text{ on } \Omega \setminus N \text{ for some } N \in \mathcal{F} \text{ such that } \mu(N) = 0\}$$

which is called the μ-essential supremum of $|f|$.

17. Convergence in L^p-spaces. $L^p(\Omega, \mathcal{F}, \mu)$ are Banach spaces. $f \to ||f||_p$ is a norm on $L^p(\Omega, \mathcal{F}, \mu)$, and $L^p(\Omega, \mathcal{F}, \mu)$ is a complete metric space under the induced distance $(f, g) \to ||f - g||_p$. We say a sequence f_n converges to f in $L^p(\Omega, \mathcal{F}, \mu)$ if f_n and f belong to $L^p(\Omega, \mathcal{F}, \mu)$ and $||f_n - f||_p \to 0$, which is equivalent to that $\int_\Omega |f_n - f|^p d\mu \to 0$.

Let us give a short discussion about the convergence in L^1-space, and we will come back to this topic by introducing the notion of uniform integrability. The following simple fact about L^1-convergence, it is quite useful though, and its proof is a good exercise about DCT.

Theorem 1.2 *(*Scheffe's Lemma*) Suppose f_n and f are integrable, and $f_n \to f$ almost surely. Then $f_n \to f$ in $L^1(\Omega, \mathcal{F}, \mu)$ if and only if $\mathbb{E}^\mu [|f_n|] \to \mathbb{E}^\mu [|f|]$.*

Proof. "Only if" part is easy. In fact, if $f_n \to f$ in L^1, then, by the triangle inequality,

$$||f_n| - |f|| \leq |f_n - f|$$

so that

$$0 \leq \left| \int_\Omega |f_n| d\mu - \int_\Omega |f| d\mu \right| = \left| \int_\Omega (|f_n| - |f|) \right| \leq \int_\Omega |f_n - f| d\mu \to 0$$

which implies that $\int_\Omega |f_n| d\mu \to \int_\Omega |f| d\mu$.

Proof of "If" part. Assume that $f_n \to f$ almost surely and $\int_\Omega |f_n| d\mu \to \int_\Omega |f| d\mu$. We want to show that $f_n \to f$ in L^1. To this end, we decompose the sample space Ω into two components for each n: $A_n = \{f_n f \geq 0\}$, $B_n = \{f_n f < 0\}$. Then

$$|f_n - f| = ||f_n| - |f|| \quad \text{on } A_n$$

and, by the triangle inequality,

$$|f_n - f| = ||f_n| + |f|| \leq ||f_n| - |f|| + 2|f| \quad \text{ on } B_n.$$

Hence

$$\begin{aligned}
\int_\Omega |f_n - f| d\mu &= \int_{A_n} |f_n - f| \, d\mu + \int_{B_n} |f_n - f| \, d\mu \\
&\leq \int_{A_n} ||f_n| - |f|| \, d\mu + \int_{B_n} [||f_n| - |f|| + 2|f|] \, d\mu \\
&= \int_\Omega ||f_n| - |f|| \, d\mu + 2 \int_{B_n} |f| d\mu \\
&= \int_\Omega ||f_n| - |f|| \, d\mu + 2 \int_\Omega 1_{B_n} |f| d\mu.
\end{aligned}$$

The first term on the right-hand side of the previous inequality many be rewrriten as the following

$$\begin{aligned}
\int_\Omega ||f_n| - |f|| \, d\mu &= \int_\Omega (|f_n| - |f|)^+ \, d\mu + \int_\Omega (|f_n| - |f|)^- \, d\mu \\
&= \int_\Omega (|f_n| - |f|) \, d\mu + 2 \int_\Omega (|f_n| - |f|)^- \, d\mu
\end{aligned}$$

where we have used the identity

$$|g| = g^+ + g^- = g^+ - g^- + 2g^- = g + 2g^-.$$

Putting together we obtain the following estimate for the L^1-norm of $f_n - f$:

$$\begin{aligned}
\int_\Omega |f_n - f| d\mu &\leq \int_\Omega ||f_n| - |f|| \, d\mu + 2 \int_\Omega 1_{B_n} |f| d\mu \\
&= \int_\Omega (|f_n| - |f|) \, d\mu + 2 \int_\Omega (|f_n| - |f|)^- \, d\mu + 2 \int_\Omega 1_{B_n} |f| d\mu. \qquad (1.1)
\end{aligned}$$

We next want to let $n \to \infty$ in the inequality above. The first term on the right-hand side tends to zero as $n \to \infty$ by assumption. In fact

$$\int_\Omega (|f_n| - |f|) \, d\mu = \int_\Omega |f_n| d\mu - \int_\Omega |f| d\mu \to 0$$

as $n \to \infty$. For the second term, we observe that

$$(|f_n| - |f|)^- = 0 \qquad \text{on } \{|f_n| \geq |f|\}$$

and

$$(|f_n| - |f|)^- = |f_n| - |f| \leq |f_n| \leq |f| \qquad \text{on } \{|f_n| < |f|\}$$

so that

$$(|f_n| - |f|)^- \leq |f|$$

for all n, $|f|$ is integrable, and $(|f_n| - |f|)^- \to 0$ almost surely, thus by the Dominated Convergence Theorem we conclude that

$$\int_\Omega (|f_n| - |f|)^- \, d\mu \to 0.$$

To show the last term on the right-hand side of (1.1) $\int_{B_n} |f| d\mu$ tends to zero, we notice that $|f| 1_{B_n} \to 0$. While it is clear that $|f| 1_{B_n} = 0$ on $\{|f| = 0\}$ for all n. If $|f(x)| > 0$, and $f_n(x) \to f(x)$, then there is N (depending on x in general) such that $|f_n(x) - f(x)| < \frac{1}{2} f(x)$ so that $f_n(x) f(x) > 0$ for all $n > N$, hence $x \notin B_n$ for $n > N$. Thus $1_{B_n}(x) = 0$ for all $n > N$. Hence $|f| 1_{B_n}(x) = 0$ for all $n > N$. Since $f_n \to f$ almost surely, we thus can conclude that $|f| 1_{B_n} \to 0$ almost everywhere as $n \to \infty$. a $|f| 1_{B_n}$ is controlled by the integral function $|f|$, so by DCT we have $\int_{B_n} |f| d\mu = \int_\Omega |f| 1_{B_n} d\mu \to 0$. Therefore, by Sandwich lemma, it follows from (1.1) that $\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0$. ■

2 Carathéodory's extension theorem

In this section we review the main tools for constructing measures.

1. π-system and montone class. Suppose $\mathcal{C}$ is a non-empty family of some subsets of Ω, then $\mathcal{C}$ is called a π-system if $\mathcal{C}$ is closed under the intersection, that is, $A \cap B \in \mathcal{C}$ whenever $A, B \in \mathcal{C}$. A collection$\mathcal{M}$ of some subsets of Ω is called a monotone class (or claaed d-class) if 1) $\Omega \in \mathcal{M}$, 2) if $A, B \in \mathcal{M}$ and $A \subseteq B$ then $B \setminus A \in \mathcal{M}$, 3) $\cup_{n=1}^\infty A_n \in \mathcal{M}$ whenever $A_n \in \mathcal{M}$ such that $A_n \uparrow$.

Given a non-empty family $\mathcal{H}$ of some subsets of Ω, $\mathcal{M}(\mathcal{H})$ denotes the smallest monotone class which contains $\mathcal{H}$, called the monotone class generated by $\mathcal{H}$. The existence and uniqueness of $\mathcal{M}(\mathcal{H})$ are left as an exercise for the reader.

Lemma 2.1 *(*Dynkin's lemma*) If $\mathcal{C}$ is a π-system over Ω, then $\mathcal{M}(\mathcal{C})$ coincides with the smallest σ-algebra $\sigma(\mathcal{C})$ containing $\mathcal{C}$, that is, $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.*

Since a σ-algebra must be a monotone class, so that $\mathcal{M}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$. To prove the other inclusion that $\sigma(\mathcal{C}) \subseteq \mathcal{M}(\mathcal{C})$, one only needs to verify that $\mathcal{M}(\mathcal{C})$ is a σ-algebra by using the fact that $\mathcal{C}$ is a π-system. The proof is routine, see for example page 193, D. Willams: Probability with martingales.

2. Uniqueness criterion. The following is a simple and useful uniqueness result.

Lemma 2.2 *(*Uniqueness lemma*) Suppose μ_j $(j = 1, 2)$ are two* finite *measures on a measurable space $(\Omega, \mathcal{F})$, and suppose $\mathcal{C} \subseteq \mathcal{F}$ is a π-system containing the sample space Ω such that $\sigma(\mathcal{C}) = \mathcal{F}$. If $\mu_1(E) = \mu_2(E)$ for every $E \in \mathcal{C}$, then $\mu_1 = \mu_2$ on $\mathcal{F}$.*

The proof of this lemma is an example how to use the Dynkin lemma.

Proof. Let $\mathcal{G}$ be the collections of all $E \in \mathcal{F}$ such that $\mu_1(E) = \mu_2(E)$. Then $\mathcal{C} \subseteq \mathcal{G}$ by assumptions. We prove that $\mathcal{G}$ is a monotone class. In fact, it is assumed that $\Omega \in \mathcal{G}$. Since $\mu_1(\varnothing) = \mu_2(\varnothing) = 0$ so that $\varnothing \in \mathcal{G}$. If $A, B \in \mathcal{G}$ and $A \subseteq B$, then, since $\mu_i(B) < \infty$, we have

$$\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A)$$

which yields that $B \setminus A \in \mathcal{G}$. Suppose now $A_n \in \mathcal{G}$, and $A_n \uparrow$, then

$$\mu_1\left(\bigcup_{n=1}^\infty A_n\right) = \lim_{n\to\infty} \mu_1(A_n) = \lim_{n\to\infty} \mu_2(A_n) = \mu_2\left(\bigcup_{n=1}^\infty A_n\right)$$

which implies that $\bigcup_{n=1}^\infty A_n \in \mathcal{G}$. Thus $\mathcal{G}$ is a monotone class containing $\mathcal{C}$. By Lemma 2.1, $\mathcal{G} \supseteq \mathcal{M}(\mathcal{C}) = \sigma\{\mathcal{C}\} = \mathcal{F}$, so that $\mu_1 = \mu_2$ on $\mathcal{F}$. ■

There is another version of the uniqueness for σ-finite measures.

Lemma 2.3 *Let μ_j $(j = 1, 2)$ be two measures on $(\Omega, \mathcal{F})$, and $\mathcal{R} \subseteq \mathcal{F}$ be a ring such that $\sigma(\mathcal{R}) = \mathcal{F}$. Suppose μ_1 and μ_2 are σ-finite on $\mathcal{R}$: there is a sequence of subsets $G_n \uparrow \Omega$, $G_n \in \mathcal{R}$ and $\mu_1(G_n) = \mu_2(G_n) < \infty$ for every n. Suppose $\mu_1(E) = \mu_2(E)$ for every $E \in \mathcal{R}$. Then $\mu_1 = \mu_2$ on $\mathcal{F}$.*

Proof. Apply Lemma 2.2 to finite measures $\mu_j(\cdot \cap G_n)$ for every n to conclude that $\mu_1(E \cap G_n) = \mu_2(E \cap G_n)$ for every n and $E \in \mathcal{F}$. Letting $n \uparrow \infty$ to obtain that $\mu_1(E) = \mu_2(E)$ for every $E \in \mathcal{F}$. ■

3. Measurable sets and Caratheodory's extension theorem. The construction of measures rely on the extension theorem of Carathéodory's, a theorem that tells us how to select measurable subsets for an outer measure. Let $\mathcal{H}$ be a σ-algebra over a sample space Ω, and $\mu^* : \mathcal{H} \to [0, \infty]$ be an outer measure on $(\Omega, \mathcal{H})$, so that

3.1) $\mu^*(\varnothing) = 0$;

3.2) $\mu^*(A) \leq \mu^*(B)$ for any $A \subseteq B$, $A, B \in \mathcal{H}$; and

3.3) μ^* is countably sub-additive:

$$\mu^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu^*(E_n)$$

for any sequence $E_n \in \mathcal{H}$ $(n = 1, 2, \cdots)$.

A subset $E \in \mathcal{H}$ is called μ^*-measurable, if E satisfies the Carathéodory condition that

$$\mu^*(F) = \mu^*(F \cap E) + \mu^*(F \cap E^c) \quad \text{for every } F \in \mathcal{H}. \tag{2.1}$$

The collection of all μ^*-measurable subsets is denoted by $\mathcal{M}$ or $\mathcal{M}(\mathcal{H}, \mu^*)$ (in order to indicate the dependence on the outer measure μ^* on $(\Omega, \mathcal{H})$.)

Theorem 2.4 *(*Caratheodory*) Let $(\Omega, \mathcal{H})$ be a measurable space and μ^* be an outer measure on $(\Omega, \mathcal{H})$. Then the collection $\mathcal{M}(\mathcal{H}, \mu^*)$ of all μ^*-measurable subsets forms a σ-algebra over Ω, and μ^* restricted on $\mathcal{M}(\mathcal{H}, \mu^*)$ is a measure.*

The proof of the previous theorem is exactly the same as that in Part A Integration.

Theorem 2.5 *(*Caratheodory*'s extension theorem) Let Ω be a space and $\mathcal{R}$ be a σ-algebra. If μ is a measure on the algebra $\mathcal{R}$, we can define the outer measure μ^* by*

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^{\infty} \mu(E_j) : \text{ where } E_j \in \mathcal{R} \text{ and } \bigcup_{j=1}^{\infty} E_j \supseteq E \right\}$$

where the inf runs over all countable cover $\{E_j\}$ of E and $E_j \in \mathcal{R}$. Then any set $E \in \mathcal{R}$ is μ^*-measurable, and $\mu^*(E) = \mu(E)$, so that μ^* restricted on the σ-algebra of all μ^*-measurable subsets is an extension of μ.

This is a consequence of Theorem 2.4, the only thing need to check is that every element E of $\mathcal{R}$, $\mu^*(E) = \mu(E)$ (which is direct but not trivial).

4. Null sets. A subset $E \in \mathcal{H}$ is μ^*-null set if $\mu^*(E) = 0$. If $\{E_i : i = 1, 2, \cdots\}$ is a sequence of μ^*-null sets, so is $\bigcup_{i=1}^{\infty} E_i$ by the countable sub-additivity. By definition, any μ^*-null set is μ^*-measurable. Therefore μ^* is a *complete* measure on $(\Omega, \mathcal{M}(\mathcal{H}, \mu^*))$.

5. Completion of a measure space. If $(\Omega, \mathcal{F}, \mu)$ is a measure space, so it is extended to an outer measure μ^* defined by

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^{\infty} \mu(E_j) : \text{ where } E_j \in \mathcal{F} \text{ such that } \bigcup_{n=1}^{\infty} E_j \supset E \right\}$$

and let $\mathcal{F}^*$ be the σ-field of all μ^*-measurable subsets. Then $(\Omega, \mathcal{F}^*, \mu)$ is a measure space, and $\mathcal{F} \subseteq \mathcal{F}^*$. Let $\mathcal{N}^\mu$ denotes the collection of all μ^*-null subsets, so that $\mathcal{N}^\mu \subseteq \mathcal{F}^*$ too. Hence $\mathcal{F}^\mu \equiv \sigma\{\mathcal{N}^\mu, \mathcal{F}\} \subseteq \mathcal{F}^*$. Thus $(\Omega, \mathcal{F}^\mu, \mu)$ is a complete measure space, called the completion of $(\Omega, \mathcal{F}, \mu)$.

3 Lebesgue-Stieltjes measures – outline of constructon

These are the most important examples of measures used in analysis.

1. Increasing functions. Let $\rho : (a, b) \to (-\infty, \infty)$ be an increasing function, where $(a, b) \subset (-\infty, \infty)$ is an open interval. Then the left limit $\rho(t-) = \lim_{s\uparrow t} \rho(s)$ and the right limit $\rho(t+) = \lim_{u\downarrow t} \rho(u)$ exist at every $t \in (a, b)$, and

$$\rho(s) \le \rho(t-) \le \rho(t) \le \rho(t+) \le \rho(u)$$

for any $a < s < t < u < b$. ρ is called right continuous (resp. left continuous) at $t \in (a, b)$ if $\rho(t) = \rho(t+)$ (resp. $\rho(t) = \rho(t-)$). For any increasing function ρ on (a, b), $\rho_+(t) \equiv \rho(t+)$ is right continuous at every $t \in (a, b)$. ρ_+ is called the right continuous modification of ρ. Similarly, $\rho_-(t) = \rho(t-)$ is left continuous at any $t \in (a, b)$, ρ_- is called the left continuous modification of ρ. Therefore, an increasing function ρ is right continuous on (a, b) if ρ_+ coincides with ρ by definition.

2. Constructing Lebesgue-Stieltjes measure. For every right continuous increasing function ρ on (a, b) we construct a measure m_ρ on a σ-algebra $\mathcal{M}_\rho$ consisting of m_ρ-measurable subsets of (a, b). The construction is divided into several steps.

2.1) Decide what we want. Let $\mathscr{C}(a, b)$ be the π-system of all intervals $(s, t]$, where $a < s \le t < b$, and we decide to assign a measure of such $(s, t]$ to be $m_\rho((s, t]) = \rho(t) - \rho(s)$.

2.2) Defining an outer measure. With m_ρ defined on the π-system $\mathscr{C}(a, b)$, we can assign an outer measure for any subset $E \subset (a, b)$, typically by

$$m_\rho^*(E) = \inf \left\{ \sum_{j=1}^{\infty} m_\rho(C_j) : \text{ where } C_j \in \mathscr{C}(a, b) \text{ such that } \bigcup_{j=1}^{\infty} C_j \supset E \right\}$$

where the inf runs over all possible *countable* covers of E through $\mathscr{C}$. m_ρ^* is an outer measure on $\mathscr{P}(a, b)$ which is the σ-algebra of all subsets of (a, b).

2.3) Apply Caratheodory's theorem. By Theorem 2.4, the collection of all m_ρ^*-measurable subsets E of (a, b) is a σ-algebra on (a, b), denoted by $\mathcal{M}_\rho$, and $m_\rho^* : \mathcal{M}_\rho \to [0, \infty]$ is a measure. m_ρ is called the Lebesgue-Stieltjes measure on (a, b) associated with a right continuous increasing function ρ on (a, b).

The above three steps of constructing measures from outer measures apply to general cases, not only for measures on intervals. The most important question is of course to identify the measurable sets, i.e. to identify the σ-algebra $\mathcal{M}_\rho$ of m_ρ^*-measurable subsets.

2.4) Identifying measurable sets. Let $\mathscr{R}(a,b)$ be the ring of all subsets $E \subset (a,b)$ which are finite unions of subsets in $\mathscr{C}(a,b)$. The main technical step is to prove the finite additivity of m_ρ^* restricted on the ring $\mathscr{R}(a,b)$. That is, if $E \in \mathscr{R}(a,b)$, so that $E = \cup_{j=1}^m C_j$ where $C_j = (s_j, t_j]$, $a < s_j \leq t_j < b$ $(j = 1, 2, \cdots, m)$ such that $(s_j, t_j]$ are disjoint, then

$$m_\rho^*(E) = \sum_{j=1}^m (\rho(t_j) - \rho(s_j)) .$$

Therefore, it follows that the outer measure m_ρ^* restricted on the ring $\mathscr{R}(a,b)$ is finitely additive.

We then can show that any set $E \in \mathscr{R}(a,b)$ is m_ρ^*-measurable, so that $\mathscr{C}(a,b) \subset \mathscr{R}(a,b) \subset \mathcal{M}_\rho$. Thus the Borel σ-algebra $\mathscr{B}(a,b) \subset \mathcal{M}_\rho$. It is easy to verify that

$$\begin{aligned}\mathscr{B}(a,b) &= (a,b) \bigcap \mathscr{B}(\mathbb{R}) = \left\{ (a,b) \bigcap G : \text{ where } G \in \mathscr{B}(\mathbb{R}) \right\} \\ &= \{ G : \ G \subset (a,b) \text{ and } G \in \mathscr{B}(\mathbb{R}) \} .\end{aligned}$$

Therefore any Borel subset of (a,b) is measurable with respect to the Lebesgue-Stieljes measure m_ρ. The restrictionof the outer measure m_ρ^* on $\mathcal{M}_\rho$ is denoted by m_ρ.

Thus for every right-continuous increasing function ρ on an open interval (a,b), we have constructed a measure space $((a,b), \mathcal{M}_\rho, m_\rho)$, which is σ-finite and complete. Also $((a,b), \mathscr{B}(a,b), m_\rho)$ is a measure space, σ-finite, which is not complete in general.

3. Notations. If ρ is an inceasing function on (a,b), then its right continuous modification $\rho_+(t) = \rho(t+)$ is right continuous, so that the Lebesgue-Stieljes measure m_{ρ_+} is defined, which is called the Lebesgue-Stieljes measure associated with ρ, denoted by m_ρ, that is, $m_\rho = m_{\rho_+}$ and $\mathcal{M}_\rho = \mathcal{M}_{\rho_+}$. In particular, m_ρ is the unique measure on $((a,b), \mathscr{B}(a,b))$ such that

$$m_\rho((s,t]) = \rho_+(t) - \rho_+(s) = \rho(t+) - \rho(s+)$$

for any $a < s < t < b$. In particular

$$\begin{aligned}m_\rho(\{t\}) &= \lim_{n\to\infty} m_\rho \left((t - \frac{1}{n}, t] \right) = \lim_{n\to\infty} \left[\rho(t+) - \rho(\left(t - \frac{1}{n} \right) +) \right] \\ &= \rho(t+) - \rho(t-)\end{aligned}$$

for every $t \in (a,b)$. In particular, $\{t\}$ (where $t \in (a,b)$) is an m_ρ-null set if and only if ρ is continuous at t.

4 Generalized measures and Radon-Nikodym's derivative

1. Generalized measures. Let $(\Omega, \mathcal{F})$ be a measurable space. If μ_1 and μ_2 are two measures on $\mathcal{F}$, and if one of them is finite so that their difference

$$\mu(E) = \mu_1(E) - \mu_2(E)$$

for $E \in \mathcal{F}$ defines a function (called a signed measure) from $\mathcal{F}$ to $[-\infty, \infty]$, which is, though not a positive measure, countably additive. Such "generalized measures" are interesting and are arisen naturally in Lebesgue's integration. For example, if f is integrable function on a measure space $(\Omega, \mathcal{F}, \mu)$, then

$$\mu_f(E) = \int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu, \text{ for } E \in \mathcal{F},$$

is an example of "generalised measures". We thertefore generalize the definition of measures to the so-called *generalized measures* as the following. A function $\mu : \mathcal{F} \to (-\infty, \infty]$ is called a *generalized measure* (which does not take value $-\infty$) if

1) $\mu(\emptyset) = 0$,

2) μ possesses the countable additivity:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

for any $A_i \in \mathcal{F}$ which are disjoint. While of course we can define generalized measures μ take values in $[-\infty, \infty)$ instead, but it is not necessary, as in this case $-\mu$ takes values in $(-\infty, \infty]$.

2. Hahn's decomposition for generalized measures. Clearly, any signed measure $\mu = \mu_1 - \mu_2$, where μ_i are measures on $(\Omega, \mathcal{F})$ and $\mu_2(\Omega) < \infty$, is a generalised measure. The converse is also true.

Theorem 4.1 *(*Hahn's decomposition*) If μ is a generalized measure on $(\Omega, \mathcal{F})$, then there is a decomposition $\Omega = A^+ \cup A^-$, where $A^+, A^- \in \mathcal{F}$ such that $A^+ \cap A^- = \emptyset$, and*

$$\mu(E \cap A^+) \geq 0, \ \mu(E \cap A^-) \leq 0$$

for every $E \in \mathcal{F}$. Moreover the positive and negative part A^+ and A^- are unique in the sense that if A_i^+ and A_i^- (where $i = 1, 2$) are two pairs satisfying the Hahn's decomposition, then

$$\mu(E \cap A_1^+) = \mu(E \cap A_2^+), \text{ and } \mu(E \cap A_1^-) = \mu(E \cap A_2^-)$$

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

Proof. [*The proof is not examinable.*] The unique sets A^+ and A^- (up to a "null set") are called the *positive* (resp. *negative*) set of the generalized measure μ. Let

$$\lambda = \inf \{\mu(G) : \text{ where } G \in \mathcal{F} \text{ such that } \mu(E \cap G) \leq 0 \text{ for all } E \in \mathcal{F}\}.$$

Choose a sequence $G_n \in \mathcal{F}$ such that $\mu(G_n) \to \lambda$ as $n \to \infty$. Then the candidate for A^- should be the largest possible negative set, that is

$$A^- = \bigcup_{n=1}^{\infty} \left(G_n \setminus \cup_{j=1}^{n-1} G_j\right).$$

In fact, A^- is still a negative set: $\mu(E \cap A^-) \leq 0$ for every $E \in \mathcal{F}$, and therefore $\mu(A^-) = \lambda$ (which yields also that $\lambda > -\infty$). We claim that the pair $A^+ = \Omega \setminus A^-$ and A^- is a decomposition satisfying that $\mu(E \cap A^+) \geq 0$ and $\mu(E \cap A^-) \leq 0$ for every $E \in \mathcal{F}$.

We only have to show that $\mu(E \cap A^+) \geq 0$ for every $E \in \mathcal{F}$, that is for any $E \subseteq A^+$, $\mu(E) \geq 0$. Let us argue by a contradiction. Suppose there is an $E_0 \subseteq A^+$ such that $\mu(E_0) < 0$. Then, since $E_0 \cap A^- = \emptyset$, so that

$$\mu(A^- \cup E_0) = \mu(A^-) + \mu(E_0) = \lambda + \mu(E_0) < \lambda$$

which is a contradiction to the definition of λ, and therefore $A^- \cup E_0$ can not be a negative set of μ, so there is a subset $A_1 \subseteq E_0$ such that $\mu(A_1) > 0$. Hence

$$k_1 = \min\left\{n \in \mathbb{N} : \text{ there is } A_1 \subseteq E_0, \mu(A_1) \geq \frac{1}{n}\right\}$$

exists, and we can find an $E_1 \subseteq \mathcal{F}$ such that $E_1 \subseteq E_0$ and $\frac{1}{k_1} \leq \mu(E_1) < \frac{1}{k_1 - 1}$. Clearly

$$\mu(E_0 \setminus E_1) = \mu(E_0) - \mu(E_1) < 0$$

so we can argue as above with $E_0 \setminus E_1$ in place of E_0 and choose $E_2 \subseteq E_0 \setminus E_1$ such that $\mu(E_2) > 0$ and $\frac{1}{k_2} \leq \mu(E_2) < \frac{1}{k_2 - 1}$, where

$$k_2 = \min\left\{n \in \mathbb{N} : \text{ there is } A_1 \subseteq E_0 \setminus E_1, \mu(A_1) \geq \frac{1}{n}\right\}.$$

Repeating the previous procedure we may construct a sequence of E_n inductively, such that $E_n \subseteq E_0 \setminus \cup_{j=1}^{n-1} E_j$ [in particular E_n are disjoint], k_n are non-decreasing, such that $\frac{1}{k_n} \leq \mu(E_n) < \frac{1}{k_n - 1}$, and

$$k_n = \min\left\{n \in \mathbb{N} : \text{ there is } A \subseteq E_0 \setminus \bigcup_{i=1}^{n-1} E_i \text{ such that } \mu(A) \geq \frac{1}{n}\right\}.$$

We claim that $\sum_n \frac{1}{k_n} < \infty$, since, otherwise, we would have

$$\sum_n \mu(E_n) \geq \sum_n \frac{1}{k_n} = \infty.$$

Since $\mu(E_0) < 0$ and

$$\mu(E_0) = \sum_n \mu(E_n) + \mu\left(E_0 \setminus \cup_{n=1}^{\infty} E_n\right)$$

we may deduce that

$$\mu\left(E_0 \setminus \bigcup_{n=1}^{\infty} E_n\right) = -\infty$$

which is a contradiction to the assumption that $\mu(E) > -\infty$ for every $E \in \mathcal{F}$. Therefore it must be hold that $k_n \to \infty$, so that $\mu(E_n) \to 0$, hence any subset of $E_0 \setminus \cup_{n=1}^{\infty} E_n$ has non-positive measure, and

$$\mu\left(E_0 \setminus \bigcup_{n=1}^{\infty} E_n\right) = \mu(E_0) - \sum_{n=1}^{\infty} \mu(E_n) < \lambda$$

which contradicts to the definition of λ. ∎

For a different approach, read W. Rudin: Real and Complex Analysis, Third Edition, pages 120-126.

3. Jordan's decomposition for generalized measures. Thus, if μ is a generalized measure over $(\Omega, \mathcal{F})$, and $\Omega = A^+ \cup A^-$ is an Hahn decomposition with respect to μ, then $\mu^+(E) = \mu(E \cap A^+)$ and $\mu^-(E) = -\mu(E \cap A^-)$ (where $E \in \mathcal{F}$) define two measures on $(\Omega, \mathcal{F})$. Moreover, μ^- is a finite measure. By definition, $\mu = \mu^+ - \mu^-$ is thus a *signed measure*, called the *Jordan decomposition* of the generalized measure μ. We may also define $|\mu| = \mu^+ + \mu^-$

which is also a measure on (Ω, μ), called the *total variation measure* of the generalized measure $\mu = \mu^+ - \mu^-$.

If ρ is a function defined on (a, b), which has finite total variation, that is,

$$\sup_D \sum_{j=1}^n |\rho(t_j) - \rho(t_{j-1})| < \infty$$

where the sup takes over all possible finite partitions $D : a < t_0 < t_1 < \cdots < t_n < b$. Then

$$\rho_{\mathrm{TV}}(t) \equiv \sup_{D_t} \sum_{j=1}^n |\rho(t_j) - \rho(t_{j-1})|$$

defines an increasing function, where the sup rns over all finite partions $D_t : a < t_0 < t_1 < \cdots < t_n = t$, for every $t \in (a, b)$. $\rho_N(t) \equiv \rho_{\mathrm{TV}}(t) - \rho(t)$ is also increasing. In particular, ρ is a difference of two increasing functions, so that ρ has left and right limits at every $t \in (a, b)$. Moreover, if ρ is right continuous at t, then so is ρ_{TV}. Therefore if ρ is right continuous and has finite total variation, then $\rho = \rho_1 - \rho_2$ a difference of two right continuous and increasing functions. $m_\rho \equiv m_{\rho_1} - m_{\rho_2}$ is a signed measure. In this case the total variation measure $|m_\rho| = m_{\rho_{\mathrm{TV}}}$.

4. Lebesgue's integrals w.r.t. a generalized measure. The usual concepts about measures may be applied to generalized measures via Jordan's dcompositions. For example, we say a generalized measure μ is σ-finite if $|\mu|$ is σ-finite, which is equivalent to say both μ^+ and μ^- are σ-finite. The theory of Lebesgue's integration may be applied to a generalized measure $\mu = \mu^+ - \mu^-$ on $(\Omega, \mathcal{F})$ too. For example, an $\mathcal{F}$-measurable function $f : \Omega \to [-\infty, \infty]$ is μ-integrable if and only if, by definition, f is integrable against the total variation measure $|\mu| = \mu^+ + \mu^-$ (which is equivalent to say f is integrable with respect both measures μ^+ and μ^-), and in this case

$$\int_\Omega f d\mu = \int_\Omega f d\mu^+ - \int_\Omega f d\mu^-.$$

5. Absolute continuity and Radon-Nikodym's theorem. Next we turn to an important concept about two generalized measures: the concept of absolute continuity.

Definition 4.2 *Let ν and μ be two measures on a measurable space $(\Omega, \mathcal{F})$, then we say ν is* absolutely continuous *with respect to μ, written as $\nu \ll \mu$, if $E \in \mathcal{F}$ and $\mu(E) = 0$ implies that $\nu(E) = 0$. That is, any μ-null set is also a ν-null set.*

Theorem 4.3 *(*Radon-Nikodym's derivative*) If μ and ν are two σ-finite measures on $(\Omega, \mathcal{F})$, such that $\nu \ll \mu$, then there is a* non-negative *$\mathcal{F}$-measurable function ρ such that*

$$\nu(E) = \int_E \rho d\mu \text{ for every } E \in \mathcal{F}.$$

Moreover ρ is unique up to μ-almost everywhere. ρ is called the Radon-Nikodym derivative of ν with respect to μ, denoted by $\frac{d\nu}{d\mu}$.

Proof. [*The proof is not examinable.*] Let us outline the proof of this important theorem for the case where ν and μ are two *finite measures*: $\mu(\Omega) < \infty$ and $\nu(\Omega) < \infty$. In this case, let $\mathscr{L}$ denote the collection of all non-negative measurable functions h such that

$$\mu[h : E] = \int_E h d\mu \leq \nu(E) \text{ for every } E \in \mathcal{F}.$$

Then, $\mathscr{L}$ is a non-empty class. Now consider $\lambda = \sup_{h\in\mathscr{L}} \int_\Omega h d\mu$. Then, clearly $\lambda \geq 0$ and $\lambda \leq \nu(\Omega) < \infty$. Choose a sequence of functions $h_n \in \mathscr{L}$ such that $\int_\Omega h_n d\mu \to \lambda$. Let $\rho = \sup_n h_n$. We claim that ρ is the Radon-Nikodym derivative. To this end, set $\rho_n = \max\{h_1, \cdots, h_n\}$ for every n. For every n, we may choose a decomposition $\Omega = \cup_{i=1}^n E_i^{(n)}$ where $E_i^{(n)} \in \mathcal{F}$ which are disjoint, and $\rho_n = h_i$ on $E_i^{(n)}$ for $i = 1, \cdots, n$. Thus, for every $E \in \mathcal{F}$, we have

$$\int_E \rho_n d\mu = \sum_{i=1}^n \int_{E_i \cap E} h_i d\mu \leq \sum_{i=1}^n \nu(E_i \cap E) = \nu(E)$$

that is, $\rho_n \in \mathscr{L}$. By definition, $\rho_n \uparrow \rho$, so by MCT, $\rho = \lim \rho_n \in L^1(\Omega, \mu)$, and by our construction, $\int_\Omega \rho d\mu = \lambda$ and $\rho \in \mathscr{L}$, i.e. $\int_E \rho d\mu \leq \nu(E)$ for every $E \in \mathcal{F}$. In particular, $\rho < \infty$ μ-almost everywhere, hence ν-almost everywhere as $\nu \ll \mu$. Therefore, we may assume that ρ is finite everywhere.

We next show that $\nu(E) = \int_E \rho d\mu$ for every $E \in \mathcal{F}$. To this end consider the generalized measure

$$m(E) = \nu(E) - \int_E \rho d\mu$$

where $E \in \mathcal{F}$. Since $\rho \in \mathscr{L}$, m is a measure, and we want to show that $m = 0$. Suppose there is $E_0 \in \mathcal{F}$ such that $m(E_0) > 0$, thus

$$\nu(E_0) > \int_{E_0} \rho d\mu.$$

Hence, there must exist $\varepsilon > 0$, such that $\nu(E_0) > \varepsilon\mu(E_0)$. Applying Hahn's decomposition to the generalized measure $\nu - \varepsilon\mu$, there is an positive set A^+ with respect to $\nu - \varepsilon\mu$, so that

$$\nu(A^+ \cap E) - \varepsilon\mu(A^+ \cap E) \geq 0$$

and

$$\nu(A^+) - \varepsilon\mu(A^+) > 0.$$

Since $\nu \ll \mu$, the last inequality yields that $\mu(A^+) > 0$. Now consider $\varphi = \rho + \varepsilon 1_{A^+}$. Then for every $E \in \mathcal{F}$, we have

$$\begin{aligned}
\int_E \varphi d\mu &= \int_{E\cap A^+} (\rho + \varepsilon 1_{A^+})\, d\mu + \int_{E\setminus A^+} \rho d\mu \\
&\leq (\nu - m)\,(E \cap A^+) + \varepsilon\mu(E \cap A^+) + \nu(E \setminus A^+) \\
&\leq \nu(E \cap A^+) + \nu(E \setminus A^+) \\
&= \nu(E)
\end{aligned}$$

so that $\varphi \in \mathscr{L}$. On the other hand

$$\int_\Omega \varphi d\mu = \int_\Omega \rho d\mu + \varepsilon \int_\Omega 1_A d\mu = \lambda + \varepsilon\mu(A) > \lambda$$

a contradiction to the definition of λ. ■

6. An integral formula. The following theorem follows from a routine computation.

Theorem 4.4 *Suppose μ and ν are two σ-finite measures on $(\Omega, \mathcal{F})$, such that $\nu \ll \mu$. Let f be an $\mathcal{F}$-measurable function. Then f is integrable with respect to ν if and only if $f\frac{d\nu}{d\mu}$ is integrable with respect to μ, and*

$$\int_\Omega f d\nu = \int_\Omega f \frac{d\nu}{d\mu} d\mu.$$

7. Conditional expectations. This is perhaps the most important concept in probability theory. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f : \Omega \to [0, \infty]$ be $\mathcal{F}$-measurable. For every $A \in \mathcal{F}$, define $\mu_f(A) = \int_\Omega f 1_A d\mu = \int_A f d\mu$. Then μ_f is a measure defined on $\mathcal{F}$. In fact, if A_n is a sequence of disjoint $\mathcal{F}$-measurable subsets, then $f 1_{\bigcup_{n=1}^\infty A_n} = \sum_{n=1}^\infty f 1_{A_n}$, thus, by MCT (series version) we have

$$\mu_f\left(\bigcup_{n=1}^\infty A_n\right) = \int_\Omega f 1_{\bigcup_{n=1}^\infty A_n} d\mu = \sum_{n=1}^\infty \int_\Omega f 1_{A_n} d\mu = \sum_{n=1}^\infty \mu_f(A_n)$$

so μ_f is a measure on $(\Omega, \mathcal{F})$.

μ_f possesses an important property – if $A \in \mathcal{F}$ is a μ-null set, i.e. $\mu(A) = 0$, then A is also a μ_f-null set: $\mu_f(A) = 0$ [which of course follows from that the integral of function on a null set is zero on any measure space]. That is to say the measure μ_f is abosoultely continuous with respect μ, that is, $\mu_f \ll \mu$. Conversely is also true, which is the context of Randon-Nikydom's theorem.

Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space, and $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$. Suppose μ is σ-finite on $\mathcal{G}$, so that there is a sequence $G_n \in \mathcal{G}$, $G_n \uparrow \Omega$ and $\mu(G_n) < \infty$ for every n. Let f be $\mathcal{F}$-measurable and non-negative such that f is σ-integrable on $\mathcal{G}$, that is, there are $G_n \in \mathcal{G}$ such that $G_n \uparrow \Omega$ and $\int_{G_n} f d\mu < \infty$ for every n. Then $\mu_f \ll \mu$ as measures on $(\Omega, \mathcal{G})$, and both μ_f and μ are σ-finite measure on $(\Omega, \mathcal{G})$, therefore, by applying Randon-Nikydom's theorem to μ and μ_f on $(\Omega, \mathcal{G})$, there is a $\mathcal{G}$-measurable and non-negative function ρ (unique up to μ-almost surely) such that $\mu_f(A) = \int_A \rho d\mu$ for every $A \in \mathcal{G}$ [that is, ρ is the Random-Nikydom's derivative of μ_f with respect to μ on $\mathcal{G}$, so denoted by $\rho = \left.\frac{d\mu_f}{d\mu}\right|_{\mathcal{G}}$]. $\left.\frac{d\mu_f}{d\mu}\right|_{\mathcal{G}}$ is called the *conditional expectation of f given $\mathcal{G}$*, denoted by $\mathbb{E}^\mu[f|\mathcal{G}]$ or simply by $\mathbb{E}[f|\mathcal{G}]$ if the measure μ involved is clear. The conditional expectation possesses the following properties:

1) $\mathbb{E}[f|\mathcal{G}]$ is $\mathcal{G}$-measurable,

2) for every $A \in \mathcal{G}$ we have

$$\mathbb{E}[f : A] = \mathbb{E}[\mathbb{E}(f|\mathcal{G}) : A]$$

that is

$$\mathbb{E}[f 1_A] = \mathbb{E}[1_A \mathbb{E}[f|\mathcal{G}]].$$

In particular, $\mathbb{E}[f] = \mathbb{E}[\mathbb{E}[f|\mathcal{G}]]$, so that, if f is integrable, so is its conditional expectation $\mathbb{E}[f|\mathcal{G}]$, which allows us to define the conditional expectation of an integrable function f by

$$\mathbb{E}[f|\mathcal{G}] = \mathbb{E}\left[f^+|\mathcal{G}\right] - \mathbb{E}\left[f^-|\mathcal{G}\right].$$

5 Product measures and Fubini's theorem

1. Product of several σ-algebras. Let A and B be two sets. Then $A \times B$ (the product set) is the set of all ordered pairs (x, y) where $x \in A$ and $y \in B$. Let Ω_1 and Ω_2 be two spaces. Then $\Omega_1 \times \Omega_2$ is also called the Cartesian product space. Suppose $\mathcal{F}_1$ and $\mathcal{F}_2$ are algebras on spaces Ω_1 and Ω_2 respectively, then $\mathcal{F}_1 \times \mathcal{F}_2$ is in general not an algebra, but the collection of all finite unions $\bigcup_{j=1}^k A_j \times B_j$ (where $A_j \in \mathcal{F}_1$ and $B_j \in \mathcal{F}_2$ and k is a positive integer) is an algebra. If $\mathcal{F}_i$ are σ-algebras, $\mathcal{F}_1 \times \mathcal{F}_2$ is in general not a σ-algebra, and we define $\mathcal{F}_1 \otimes \mathcal{F}_2$ to be the smallest σ-algebra containing $\mathcal{F}_1 \times \mathcal{F}_2$, that is, $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\{\mathcal{F}_1 \times \mathcal{F}_2\}$. The construction may be extended to the product space of finite many spaces. More precisely, if $(\Omega_i, \mathcal{F}_i)$ $(i = 1, \cdots, n)$ are measurable spaces, then

$$\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \sigma\{A_1 \times \cdots \times A_n : A_i \in \mathcal{F}_i\}$$

and $(\Omega_1 \otimes \cdots \otimes \Omega_n, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n)$ is called the product measurable space of $(\Omega_i, \mathcal{F}_i)$.

Exercise 5.1 *1) Suppose S_i $(i = 1, \cdots, n)$ are topological spaces, so that the product space $S_1 \times \cdots \times S_n$ carries the product topology. Show that*

$$\mathcal{B}(S_1 \times \cdots \times S_n) = \mathcal{B}(S_1) \otimes \cdots \otimes \mathcal{B}(S_n).$$

2) If $(\Omega_i, \mathcal{F}_i)$ $(i = 1, 2, \cdots)$ are measurable spaces, then

$$\begin{aligned} \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \mathcal{F}_3 &= \mathcal{F}_1 \otimes (\mathcal{F}_2 \otimes \mathcal{F}_3) \\ &= (\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3. \end{aligned}$$

2. Product σ-algebra of countable many σ-algebras. Let us now consider a sequence of measurable spaces $(\Omega_i, \mathcal{F}_i)$ $(i = 1, 2, \cdots)$. The Cartesain product $\prod_{i=1}^{\infty} \Omega_i$ is the space consisting of all sequences $(x_1, \cdots, x_i, \cdots)$ where $x_i \in \Omega_i$ for $i = 1, 2, \cdots$, and define $\prod_{i=1}^{\infty} \otimes \mathcal{F}_i$ to be the smallest σ-algebra containing all $\prod_{i=1}^{\infty} A_i$ where $A_i \in \mathcal{F}_i$ for all i and $A_i = \Omega_i$ except *for finite many $i \in \mathbb{N}$*. $(\prod_{i=1}^{\infty} \Omega_i, \prod_{i=1}^{\infty} \otimes \mathcal{F}_i)$ is called the product measurable space of $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2, \cdots$.

3. Measurable sections. Now let us come to the construction of product measures on product spaces. We need the following elementary fact.

Lemma 5.2 *If $\mathcal{F}_1$ and $\mathcal{F}_2$ are algebras on Ω_1 and Ω_2 respectively, then the collection $\mathcal{A}(\mathcal{F}_1, \mathcal{F}_2)$ of all finite disjoint unions $\bigcup_{i=1}^{k} A_i \times B_i$ for some $k \in \mathbb{N}$, where $A_i \in \mathcal{F}_1$, $B_i \in \mathcal{F}_2$ and all products $A_i \times B_i$ are disjoint, is an algebra on $\Omega_1 \times \Omega_2$. If $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ-algebras, then $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma \{\mathcal{A}(\mathcal{F}_1, \mathcal{F}_2)\}$.*

Lemma 5.3 *Let $(\Omega_i, \mathcal{F}_i)$ $(i = 1, 2)$ be measurable spaces. 1) If $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$, then for each $x_1 \in \Omega_1$ the section*

$$A_{x_1} = \{x_2 \in \Omega_2 : (x_1, x_2) \in A\}$$

is measurable, i.e. $A_{x_1} \in \mathcal{F}_2$. Similarly

$$A^{x_2} = \{x_1 \in \Omega_1 : (x_1, x_2) \in A\}$$

belongs to $\mathcal{F}_1$.

2) Suppose f is measurable on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, then for each $x_1 \in \Omega_1$, the function $f_{x_1}(x_2) = f(x_1, x_2)$ is $\mathcal{F}_2$-measurable.

Proof. Proof of 1). Let $\mathcal{E}$ be the collection of all $E \subseteq \Omega_1 \times \Omega_2$ such that its x_1-section is measurable. Then $\mathcal{E}$ is a σ-algebra containing all $A \times B$ where $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Therefore $\mathcal{F}_1 \otimes \mathcal{F}_2 \subset \mathcal{E}$ which proves 1). To show 2), we notice that

$$\{x_2 : f_{x_1}(x_2) > a\} = \{x_2 : f(x_1, x_2) > a\}$$

which is the x_1-section of $\{f > a\}$ (which is $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable), so its x_1-section is $\mathcal{F}_2$-measurable. Therefore f_{x_1} is $\mathcal{F}_2$-measurable. ■

In particular, if S_i are topological spaces with Borel σ-algebras, and if f is Borel measurable on $S_1 \times S_2$ with the product topology, then its section f_{x_1} (for each $x_1 \in S_1$) is Borel measurable on S_2.

4. Product measure of two measures. The following is the main technical fact in the construction of product measures.

Lemma 5.4 *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ $(i = 1, 2)$ be two finite measure spaces. Then for any $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$, $x_1 \to \mu_2(A_{x_1})$ (resp. $x_2 \to \mu_1(A^{x_2})$) is measurable on $(\Omega_1, \mathcal{F}_1)$ (resp. $(\Omega_2, \mathcal{F}_2)$) and*

$$\int_{\Omega_1} \mu_2(A_{x_1})\mu_1(dx_1) = \int_{\Omega_2} \mu_1(A^{x_2})\mu_2(dx_2) \tag{5.1}$$

the common value is denoted by $\mu_1 \times \mu_2(A)$, so that $\mu_1 \times \mu_2$ is defined on $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Proof. Let $\mathscr{L}$ denote the collection of all subsets $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ such that both functions $\mu_2(A_{x_1})$ and $\mu_1(A^{x_2})$ are measurable and (5.1) holds. By definition, $\mathcal{F}_1 \times \mathcal{F}_2 \subset \mathscr{L}$, and by linearlity of integration, we can see that $\mathscr{L}$ is a ring. On the other hand, by using MCT, we can show that $\mathscr{L}$ is a monotone class. Therefore $\mathscr{L}$ must be a σ-algebra, so that $\mathscr{L} = \mathcal{F}_1 \otimes \mathcal{F}_2$. ∎

Theorem 5.5 *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ $(i = 1, 2)$ be two σ-finite measure spaces. Choose a sequence $G_n = A_n \times B_n$, where $A_n \uparrow \Omega_1$, $A_n \in \mathcal{F}_1$, $\mu_1(A_n) < \infty$, and similarly, $B_n \uparrow \Omega_2$, $B_n \in \mathcal{F}_2$, $\mu_2(B_n) < \infty$, for every n. If $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$ then define*

$$m(E) = \lim_{n\to\infty} \mu_1 \times \mu_2(E \cap G_n)$$

where $\mu_1 \times \mu_2(E \cap G_n)$ is defined in Lemma 5.4. Then m is the unique σ-finite measure on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, such that

$$m(A \times B) = \mu_1(A)\mu_2(B) \qquad \forall A \in \mathcal{F}_1,\ B \in \mathcal{F}_2. \tag{5.2}$$

which will be denoted by $\mu_1 \times \mu_2$, called the product measure of μ_1 and μ_2.

Proof. Uniqueness follows from Lemma 2.3. Given a sequence $\{G_n\}$ satisfying the conditions in the theorem. Since $\mu_1 \times \mu_2(E \cap G_n)$ is non-negative and increasing, so that m is well defined on $\mathcal{F}_1 \otimes \mathcal{F}_2$. Clearly $m(\emptyset) = 0$, so we need to show that m is countably additive. We prove this in two steps.

Note that $\mu_1(\cdot \cap A_n)$ and $\mu_2(\cdot \cap B_n)$ are finite measures, so that $\mu_1 \times \mu_2(E \cap G_n)$ is well-defined via (5.1), and is non-negative, increasing in n. We want to show that m is countably additive. Suppose $E_k \in \mathcal{F}_1 \otimes \mathcal{F}_2$ are disjoint sequence, and $E = \cup_{k=1}^{\infty} E_k$. Then, for every n

$$\begin{aligned} m(E \cap G_n) &= \int_{\Omega_2} \mu_1((E \cap G_n)^{x_2})\mu_2(dx_2) = \int_{\Omega_2} \mu_1(\cup_k (E_k \cap G_n)^{x_2})\mu_2(dx_2) \\ &= \int_{\Omega_2} \sum_k \mu_1((E_k \cap G_n)^{x_2})\mu_2(dx_2) = \sum_k \int_{\Omega_2} \mu_1((E_k \cap G_n)^{x_2})\mu_2(dx_2) \\ &= \sum_k m(E_k \cap G_n). \end{aligned}$$

where the fourth equality follows from MCT (series version). It follows that

$$m(E \cap G_n) \le \sum_k m(E_k)$$

so that, by letting $n \to \infty$ we obtain $m(E) \le \sum_k m(E_k)$. On the other hand, for every N,

$$m(E \cap G_n) \ge \sum_{k=1}^{N} m(E_k \cap G_n).$$

Letting $n \to \infty$ we have $m(E) \geq \sum_{k=1}^{N} m(E_k)$, so that we also have $m(E) \geq \sum_k m(E_k)$. Therefore $m(E) = \sum_k m(E_k)$ which completes the proof. ■

5. Product measure of finite many σ-finite measures. Obviously, the same approach is applied to finite many σ-finite measure spaces, and we have

Theorem 5.6 *Suppose $(\Omega_i, \mathcal{F}_i, \mu_i)$ $(i = 1, \cdots, n)$ are σ-finite measure spaces, then there is a unique σ-finite measure $\mu_1 \times \cdots \times \mu_n$ called the product measure on $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n)$ such that*

$$\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n) \quad \forall A_i \in \mathcal{F}_i.$$

6. Product probability measure of countable many probability measures. However, there is obstruction for constructing product measures on the product space of countably many measure spaces, one can not, in general, use $\prod_{i=1}^{\infty} \mu_i(A_i)$ to define the measure of $\prod_{i=1}^{\infty} A_i$ even if $A_i = \Omega_i$ except finite many i. This approach on the other hand works for probability spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$ as in this case $\prod_{i=1}^{\infty} \mu_i(A_i)$ for $\prod_{i=1}^{\infty} A_i$, where $A_i = \Omega_i$ except finite many i, becomes a finite product as $\mu_i(\Omega_i) = 1$ for sufficient large i.

Theorem 5.7 *Suppose $(\Omega_i, \mathcal{F}_i, \mu_i)$ $(i = 1, 2, \cdots)$ are* probability *spaces, then there is a probability measure $\prod_{i=1}^{\infty} \mu_i$ (called the product probability measure) on $(\prod_{i=1}^{\infty} \Omega_i, \prod_{i=1}^{\infty} \otimes \mathcal{F}_i)$ such that*

$$\prod_{i=1}^{\infty} \mu_i(A_1 \times \cdots \times A_k \times \cdots) = \prod_{i=1}^{\infty} \mu_i(A_i).$$

for any $A_i \in \mathcal{F}_i$ for all i and $A_i = \Omega_i$ except for finite many i.

Proof. [*The proof is not examinable*] Let $\mathscr{R}$ denote the ring of all subsets $E \subset \prod_{i=1}^{\infty} \Omega_i$ which has the following form:

$$E = \bigcup_{j=1}^{n} A_j, \text{ where } A_j = A_1^{(j)} \times \cdots \times A_k^{(j)} \times \cdots$$

$A_k^{(j)} \in \mathcal{F}_k$ for $j = 1, \cdots, n$, and for every j, there is k_j, such that $A_k^{(j)} = \Omega_k$ for every $k > k_j$, for some $n \in \mathbb{N}$. If $E \in \mathscr{R}$ then we may choose a decomposition above such that A_j (for some n, $j = 1, \cdots, n$) are disjoint, and define

$$m(E) = \sum_{j=1}^{n} m(A_j) \text{ where } m(A_j) = \mu_1(A_1^{(j)}) \cdots \mu_k(A_k^{(j)}) \cdots$$

each $m(A_j)$ is in fact a finite product as all μ_k are probability measures. To see why m is well defined and is in fact a measure on $\mathscr{R}$, we make the following crucial observation. If $E_1, \cdots, E_N \in \mathscr{R}$, then, there is a common K, such that for all $n = 1, \cdots, N$ each $E_n = A^{(n)} \times \Omega_{K+1} \times \cdots$ for some $A^{(n)} \in \prod_{k=1}^{K} \mathcal{F}_k$, and therefore

$$E \equiv \bigcup_{n=1}^{N} E_n = A \times \Omega_{K+1} \times \cdots$$

for some $A \in \prod_{k=1}^{K} \mathcal{F}_k$. Since μ_k are probability measures, so by definition

$$m(E_n) = \mu_1 \times \cdots \times \mu_K(A^{(n)})$$

(the identity is no longer ensured if there are infinite many μ_k with total mass $\mu_k(\Omega_k) \neq 1$). Since $\mu_1 \times \cdots \times \mu_K$ is a measure, so that, if E_n $(n = 1, \cdots, N)$ are disjoint, then

$$m(E) = \mu_1 \times \cdots \times \mu_K(A) = \sum_{n=1}^{N} \mu_1 \times \cdots \times \mu_K(A^{(n)}) = \sum_{n=1}^{N} m(E_n)$$

which shows that m is well defined on the ring $\mathscr{R}$ and m is finitely additive. Next, the standard machinary may be applied to construct the product probability $\prod_{i=1}^{\infty} \mu_i$. Firstly, define outer measure

$$m^*(E) = \inf \left\{ \sum_{n=1}^{\infty} m(E_n) : \text{ where } E_n \in \mathscr{R} \text{ such that } \bigcup_{n=1}^{\infty} E_n \supset E \right\}$$

for every sunset $E \subset \prod_{i=1}^{\infty} \Omega_i$. Let $\mathcal{M}$ denote the σ-algebra of all m^*-measurable subsets. Then m^* is a measure on $\mathcal{M}$ (by the Carathodory extension theorem). Since $\mathscr{R}$ is a ring and m is finitely additive, we thus must have $\mathscr{R} \subset \mathcal{M}$. Since $\prod_{i=1}^{\infty} \otimes \mathcal{F}_i = \sigma(\mathscr{R}) \subset \mathcal{M}$, so that m^* restricted on $\prod_{i=1}^{\infty} \otimes \mathcal{F}_i$ is a probability measure. The construction is complete. ■

7. Fubini's theorem. Let us now turn to the Fubini theorem.

Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ $(i = 1, 2)$ be two σ-finite measure spaces. Suppose $f : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ is a measurable function, such that for almost all $x_1 \in \Omega_1$, f_{x_1} is integrable on $(\Omega_2, \mathcal{F}_2, \mu_2)$. Hence, there is a set $N_1 \in \Omega_1$ with $\mu_1(N_1) = 0$, and for any $x_1 \in \Omega_1 \backslash N_1$, $f_{x_1} \in L^1(\Omega_2, \mathcal{F}_2, \mu_2)$, so that we can define

$$h(x_1) = \int_{\Omega_2} f_{x_1}(x_2) \mu_2(dx_2) \quad \text{if } x_1 \in \Omega_1 \setminus N_1$$

otherwise $h(x_1) = 0$. If there is a $\tilde{h} \in L^1(\Omega_1, \mathcal{F}_1, \mu_1)$, such that $\tilde{h} = h$ almost surely w.r.t. μ_1, then we can form an integral

$$I_{1,2}(f) = \int_{\Omega_1} \tilde{h}(x_1) \mu_1(dx_1).$$

One can show that, if $I_{1,2}(f)$ exists (i.e. there is some N_1 and $\tilde{h}$ satisfying the above conditions), then $I_{1,2}(f)$ does not depend on N_1 and $\tilde{h}$, therefore $I_{1,2}(f)$ is called an iterated integral of f over $\Omega_1 \times \Omega_2$, denoted by

$$\int_{\Omega_1} \left(\int_{\Omega_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1).$$

Similarly we define the iterated integral

$$\int_{\Omega_2} \left(\int_{\Omega_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2).$$

Theorem 5.8 *(*Fubini's theorem*) Let μ_j be σ-finite measure on $(\Omega_j, \mathcal{F}_j)$, where $j = 1, 2$. Suppose $f : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ is a measurable function on the product measure space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.*

1) If $f \in L^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$, then both iterated integrals exist and equal to the integral $\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \times \mu_2)$.

2) Conversely, if one of the iterated integral of $|f|$ is finite, then $f \in L^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$.

Proof. By Theorem 5.5 and the definition of the product measure $\mu_1 \times \mu_2$, for every $E \in \mathcal{F}_1 \otimes \mathcal{F}_2$ we have

$$\int_{\Omega_1 \times \Omega_2} 1_E d\mu_1 \times \mu_2 = \int_{\Omega_2} \left[\int_{\Omega_1} 1_E d\mu_1 \right] d\mu_2 = \int_{\Omega_1} \left[\int_{\Omega_2} 1_E d\mu_2 \right] d\mu_1$$

which yields that Fubini's theorem holds for every non-negative simple measurable function.

Suppose f is non-negative and $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable, then we can choose a sequence of non-negative, measurable simple functions $\varphi_n : \Omega_1 \times \Omega_2 \to [0, \infty)$ such that $\varphi_n \uparrow f$. By MCT we have

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d\mu_1 \times \mu_2 &= \lim_{n \to \infty} \int_{\Omega_1 \times \Omega_2} \varphi_n d\mu_1 \times d\mu_2 \\ &= \lim_{n \to \infty} \int_{\Omega_2} \left[\int_{\Omega_1} \varphi_n d\mu_1 \right] d\mu_2 = \lim_{n \to \infty} \int_{\Omega_2} \Phi_n d\mu_2 \end{aligned}$$

where

$$\Phi_n = \int_{\Omega_1} \varphi_n d\mu_1$$

which are non-negative, $\mathcal{F}_2$-measurable and $\Phi_n \uparrow$, thus by MCT applying to $\{\Phi_n\}$ on $(\Omega_2, \mathcal{F}_2, \mu_2)$ to obtain

$$\lim_{n \to \infty} \int_{\Omega_2} \Phi_n d\mu_2 = \int_{\Omega_2} \lim_{n \to \infty} \Phi_n d\mu_2 = \int_{\Omega_2} \lim_{n \to \infty} \left[\int_{\Omega_1} \varphi_n d\mu_1 \right] d\mu_2.$$

Since for every x_2, $\varphi_n(\cdot, x_2) \uparrow f(\cdot, x_2)$ and non-negative, measurable, so by applying MCT on $(\Omega_1, \mathcal{F}_1, \mu_1)$ we thus have

$$\lim_{n \to \infty} \left[\int_{\Omega_1} \varphi_n d\mu_1 \right] = \int_{\Omega_1} f d\mu_1.$$

Putting the previous equalities together we obtain

$$\int_{\Omega_1 \times \Omega_2} f d\mu_1 \times \mu_2 = \int_{\Omega_2} \left[\int_{\Omega_1} f d\mu_1 \right] d\mu_2$$

and similarly

$$\int_{\Omega_1 \times \Omega_2} f d\mu_1 \times \mu_2 = \int_{\Omega_1} \left[\int_{\Omega_2} f d\mu_2 \right] d\mu_1$$

for any non-negative, measurable function f. The conclusions of the theorem follow immediately. ■

8. Completion of product measure spaces. Recall that, if $(\Omega, \mathcal{F}, \mu)$ is a σ-finite measure space, then $\mathcal{F}^\mu$ is the completed σ-algebra of $\mathcal{F}$ under the measure μ, that is, $\mathcal{N}$ denotes the collection of all subsets of Ω with outer measure zero, then $\mathcal{F}^\mu = \sigma\{\mathcal{F}, \mathcal{N}\}$. We have shown that μ can be uniquely extended to a σ-finite measure on $\mathcal{F}^\mu$, denoted again by μ. Complications may arise if we consider the completion of $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$. In general, the completion of $\mathcal{F}_1 \otimes \mathcal{F}_2$ under $\mu_1 \times \mu_2$ does not coincide with the product σ-algebra of the completions of $\mathcal{F}_i$ under μ_i, but we have

Lemma 5.9 *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ be two σ-finite measure spaces. Then*

$$\mathcal{F}_1^{\mu_1} \otimes \mathcal{F}_2^{\mu_2} \subset (\mathcal{F}_1 \otimes \mathcal{F}_2)^{\mu_1 \times \mu_2}$$

and

$$(\mathcal{F}_1 \otimes \mathcal{F}_2)^{\mu_1 \times \mu_2} = (\mathcal{F}_1^{\mu_1} \otimes \mathcal{F}_2^{\mu_2})^{\mu_1 \times \mu_2}.$$

Proof. The proof is routine, left as an exercise. ■

If $f : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ is measurable w.r.t $(\mathcal{F}_1 \otimes \mathcal{F}_2)^{\mu_1 \times \mu_2}$, then its section $f_{x_1} : \Omega_2 \to (-\infty, \infty)$ by sending x_2 to $f(x_1, x_2)$ is not necessary measurable w.r.t. $\mathcal{F}_2^{\mu_2}$, however, according to definition, there is a function $\tilde{f} : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ which is measurable w.r.t. $\mathcal{F}_1 \otimes \mathcal{F}_2$ and $f = \tilde{f}$ $\mu_1 \times \mu_2$-almost surely, and $\tilde{f}_{x_1}$ is measurable w.r.t. $\mathcal{F}_2$ for all $x_1 \in \Omega_1$. Moreover it is clear that $\tilde{f}_{x_1} = f_{x_1}$ for almost all $x_1 \in \Omega_1$ with respect to μ_1. Therefore f_{x_1} is $\mathcal{F}_2^{\mu_2}$-measurable for μ_1-almost all $x_1 \in \Omega_1$. The iterated integrals of f are defined to be those of $\tilde{f}$, and we can show that they are independent of the choice of a version $\tilde{f}$.

If $f \in L^1(\Omega_1 \times \Omega_2, (\mathcal{F}_1 \otimes \mathcal{F}_2)^{\mu_1 \times \mu_2}, \mu_1 \times \mu_2)$, then we choose $\tilde{f}$ which is $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable such that $f = \tilde{f}$ $\mu_1 \times \mu_2$-a.e., applying the Fubini theorem to $\tilde{f}$, we thus have the following refined version of Fubini's theorem.

Theorem 5.10 *(*Fubini's theorem*) Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ be two σ-finite measure spaces. Suppose $f : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ is $(\mathcal{F}_1 \otimes \mathcal{F}_2)^{\mu_1 \times \mu_2}$-measurable.*

1) If $f \in L^1(\Omega_1 \times \Omega_2, (\mathcal{F}_1 \otimes \mathcal{F}_2)^{\mu_1 \times \mu_2}, \mu_1 \times \mu_2)$, then the two iterated integrals of f exist and coincide with the integral $\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \times \mu_2)$.

2) Conversely, if one of the iterated integral of $|\tilde{f}|$ is finite, where $\tilde{f} = f$ $\mu_1 \times \mu_2$-a.e. and $\tilde{f}$ is $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable, then $f \in L^1(\Omega_1 \times \Omega_2, (\mathcal{F}_1 \otimes \mathcal{F}_2)^{\mu_1 \times \mu_2}, \mu_1 \times \mu_2)$.

6 Some concepts in probability

Let us now set up the probability setting by using the theory of measures developed in the previous sections.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. An $\mathcal{F}$-measurable function X (complex, or valued in $[-\infty, \infty]$) on Ω is called a random variable. The concept of random variables may be generalised to mappings, which may be useful in discussing probability models. In general, if $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are two measurable spaces, then a mapping $\Phi : \Omega_1 \to \Omega_2$ is measurable if $\Phi^{-1}(A) \in \mathcal{F}_1$ whenever $A \in \mathcal{F}_2$. Thus a real random variable $X : \Omega \to \mathbb{R}$ is just a measurable map from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

If X is integrable or non-negative random variable, then its integral $\int_\Omega X(\omega)\mathbb{P}(d\omega)$ is called the *expectation* of X, or the mean value of X, denoted by $\mathbb{E}[X]$. We say the expectation of X exists if X is integrable.

Exercise 6.1 *The* inclusion-exclusion formula *holds:*

$$\begin{aligned}\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) &= \sum_j \mathbb{P}(A_j) - \sum_{j_1<j_2} \mu\mathbb{P}(A_{j_1}A_{j_2}) + \sum_{j_1<j_2<j_3} \mathbb{P}(A_{j_1}A_{j_2}A_{j_3}) \\ &\quad - \cdots + (-1)^{k-1} \sum_{j_1<\cdots<j_k} \mathbb{P}(A_{j_1}\cdots A_{j_k}) + \cdots\end{aligned}$$

where $A_j \in \mathcal{F}$ for $j = 1, 2, \cdots$.

Exercise 6.2 *1) Let $f : \Omega \to S$. Show that $f^{-1}(\cup_\alpha A_\alpha) = \cup_\alpha f^{-1}(A_\alpha)$ and $f^{-1}(A^c) = (f^{-1}(A))^c$ for any sets A, A_α where α runs over an arbitrary index set.*

2) Let $f : \Omega \to S$ where $(\Omega, \mathcal{F})$ and (S, Σ) are two measurable spaces. Show that

$$f^{-1}(\Sigma) \equiv \left\{f^{-1}(A) : A \in \Sigma\right\}$$

is a σ-algebra on Ω, and f is measurable (with respect to $\mathcal{F}$) if and only if $f^{-1}(\Sigma) \subset \mathcal{F}$.

Exercise 6.3 *Let (S,Σ) be a measurable space, and $X_\alpha : \Omega \to S$ ($\alpha \in \Lambda$) be a family of functions on Ω taking values in S. Then we use $\sigma\{X_\alpha : \alpha \in \Lambda\}$ to denote the smallest σ-algebra such that each X_α is a measurable map from $(\Omega, \sigma\{(X_\alpha)_{\alpha\in\Lambda}\})$ to (S,Σ).*

1) Let $\Sigma_0 = \{X_\alpha^{-1}(A) : A \in \Sigma \text{ and } \alpha \in \Lambda\}$. Show that

$$\sigma\{X_\alpha : \alpha \in \Lambda\} = \sigma(\Sigma_0).$$

2) Let $\mathcal{F} \equiv \sigma\{X_\alpha : \alpha \in \Lambda\}$. Show that, if $\alpha_j \in \Lambda$ ($j = 1, 2, \cdots$) is a countable subset of Λ and $A \in \Sigma$, then

$$\left\{\omega : X_{\alpha_j}(\omega) \in A \text{ for all } j = 1, 2, \cdots\right\}$$

belongs to $\mathcal{F}$. The above event is often written as $\left\{X_{\alpha_j} \in A \text{ for } j = 1, 2, \cdots\right\}$.

6.1 Laws, distribution functions

These are basic concepts associated with random variables. Let us begin with the following

Proposition 6.4 *Let $(\Omega,\mathcal{F})$ and (S,Σ) be two measurable spaces, $\mathbb{P}$ a measure on $(\Omega,\mathcal{F})$, and $X : \Omega \to S$ be a measurable map. Define*

$$\begin{aligned}\mu(A) &\equiv \mathbb{P}\left(X^{-1}(A)\right) = \mathbb{P}[X \in A] \\ &= \mathbb{P}(\{\omega : X(\omega) \in A\})\end{aligned}$$

for every $A \in \Sigma$. Then μ is a measure on (S,Σ), denoted by $\mathbb{P}\circ X^{-1}$, which is called the distribution of X.

In particular, if X is a random variable on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ taking values in $\mathbb{R}^n$, then $\mathbb{P}\circ X^{-1}$ is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, called the *law* or called *the distribution* of the random variable X. Sometimes we also use μ_X to denote the distribution of X.

If $X : \Omega \to \mathbb{R}$ is a real-valued random variable, then its *distribution function*

$$\begin{aligned}F(x) &= \mathbb{P}(X \le x) \\ &= \mathbb{P}(\{\omega : X(\omega) \le x\}) \\ &= \mu_X((-\infty, x]),\end{aligned}$$

is a non-decreasing function on $\mathbb{R}$ with values in $[0,1]$. Then $0 \le F \le 1$; $F \uparrow$; $\lim_{x\to-\infty} F(x) = 0$; $\lim_{x\to\infty} F(x) = 1$; F is right-continuous:

$$\lim_{x\downarrow x_0} F(x) = F(x_0) \qquad \forall x_0 \in \mathbb{R}.$$

The Lebegue-Stieltjes measure m_F associated with the increasing and right-continuous function F is the unique measure such that

$$m_F((a,b]) = F(b) - F(a) = \mathbb{P}(a < X \le b) = \mu_X((a,b])$$

for all $a < b$. Since the collection $\mathcal{C}$ of all $(a,b]$ (where $a < b$ are reals) is a π-system, according to the Uniqueness Lemma 2.2, $m_F = \mu_X$, that is, the distribution (law) of a real random variable X is the Lebesgue-Stieltjes measure associated with the distribution function of X.

6.2 Independence

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

1. Independent events. Recall that, if $A, B \in \mathcal{F}$ be two events, then A and B are independent, if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)P(B). \tag{6.1}$$

Let

$$\begin{aligned} \mathcal{F}_A &= \sigma\{A\} = \{\Omega, A, A^c, \emptyset\}, \\ \mathcal{F}_B &= \sigma\{B\} = \{\Omega, B, B^c, \emptyset\}. \end{aligned}$$

Then (6.1) implies that

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F), \quad \forall E \in \mathcal{F}_A, F \in \mathcal{F}_B,$$

and therefore the σ-algebras $\mathcal{F}_A$ and $\mathcal{F}_B$ are *independent.*

Definition 6.5 *1) Let $\{\mathcal{F}_\alpha : \alpha \in \wedge\}$ be a collection of sub σ-algebras of $\mathcal{F}$. Then $\{\mathcal{F}_\alpha : \alpha \in \wedge\}$ are independent if for any $k \in \mathbb{N}$, and any $\alpha_1, \cdots, \alpha_k \in \wedge$ such that $\alpha_i \neq \alpha_j$ if $i \neq j$, we have*

$$\mathbb{P}(A_1 \cdots A_k) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_k), \quad \forall A_1 \in \mathcal{F}_{\alpha_1}, \cdots, A_k \in \mathcal{F}_{\alpha_k}.$$

2) Let $\{F_\alpha : \alpha \in \wedge\}$ be a family of events: $F_\alpha \in \mathcal{F}$. Then we say $\{F_\alpha : \alpha \in \wedge\}$ are independent if $\{\sigma(F_\alpha) : \alpha \in \wedge\}$ are independent.

3) Let $\{X_\alpha : \alpha \in \wedge\}$ be a family of random variables. Then $\{X_\alpha : \alpha \in \wedge\}$ are independent if the family of σ-algebras $\{\sigma(X_\alpha) : \alpha \in \wedge\}$ are independent.

2. Independence via π-system. In elementary probability theory, we already give a definition of independence for random variables. You should show that the above definition coincides with the one you have learned before. The following Lemma is very useful although it is very simple and follows a simple application of Lemma 2.2.

Lemma 6.6 *Let $\mathcal{F}_\alpha \equiv \sigma\{\mathcal{C}_\alpha\}$ where each $\mathcal{C}_\alpha$ is a π-system in the sense that*

$$A, B \in \mathcal{C}_\alpha \quad \textit{implies that } A \cap B \in \mathcal{C}_\alpha.$$

Then $\{\mathcal{F}_\alpha : \alpha \in \wedge\}$ are independent if and only if for any $k \in \mathbb{N}$, any $F_1 \in \mathcal{C}_{\alpha_1}, \cdots, F_k \in \mathcal{C}_{\alpha_k}$ where $\alpha_1, \cdots, \alpha_k$ are different, we have

$$\mathbb{P}[F_1 \cap \cdots \cap F_k] = \mathbb{P}[F_1] \cdots \mathbb{P}[F_k].$$

In fact, we can show the equality by induction on k. Consider two measures on $\mathcal{F}_{\alpha_k}$ defined by

$$\mu_1(E) = \mathbb{P}[F_1 \cap \cdots \cap F_{k-1} \cap E]$$

and

$$\mu_2(E) = \mathbb{P}[F_1 \cap \cdots \cap F_{k-1}]\,\mathbb{P}(E)$$

where F_i as in the lemma, but fixed, and $E \in \mathcal{F}_{\alpha_k}$. The induction assumption and the condition in the lemma implied that $\mu_1 = \mu_2$ on $\mathcal{C}_{\alpha_k}$, hence, by Lemma 2.2, $\mu_1 = \mu_2$ on $\mathcal{F}_{\alpha_k}$ and the proof is complete.

3. Independent random variables.

Theorem 6.7 *Let $X_1, \cdots, X_n, \cdots$ be a sequence of real random variables. Then $X_1, \cdots, X_n, \cdots$ are independent if and only if for any $k \in \mathbb{N}$, and any $x_1, \cdots, x_k \in \mathbb{R}$*

$$\mathbb{P}\left[X_1 \leq x_1, \cdots, X_k \leq x_k\right] = \mathbb{P}\left[X_1 \leq x_1\right] \cdots \mathbb{P}\left[X_k \leq x_k\right].$$

That is, the joint distribution of $X_1, \cdots, X_n$ is the product of the distribution functions of the random variables $X_k, 1 \leq k \leq n$.

This follows from the previous lemma, as $\mathcal{C}_k$ the collection of all subsets $\{X_k \leq x\}$ where x runs through all reals is a π-system, where $k = 1, 2, \cdots$.

Therefore, the *joint* law or distribution of a sequence of independent random variables $(X_1, X_2, \cdots, X_n, \cdots)$ is the product probability measure $\mu_1 \times \cdots \times \mu_n \times \cdots$, where μ_n is the distribution of X_n. In particular, if $\{X_n : n = 1, 2, \cdots\}$ is a sequence of independent real random variables, then its joint law (or called joint distribution) is the product probability measures of the Lebesgue-Stieltjes measure m_{F_n} where $F_n(x) = \mathbb{P}\left[X_n \leq x\right]$ is the distribution function of X_n, $n = 1, 2, \cdots$.

Theorem 6.8 *Let X be a random variable (valued in a measurable space) on some probability space. Then there is a sequence of independent identically distributed random variables $\{X_n : n \in \mathbb{N}\}$, each X_n has the same law as that of X.*

Proof. [*The proof is not examinable*] Let X be a random variable taking its values in a measurable space $(S, \mathcal{G})$, and let μ be the distribution of X. Then μ is a probability measure. Let $(S_n, \mathcal{G}_n, \mu_n) = (S, \mathcal{G}, \mu)$ $(n = 1, 2, \cdots)$ and let $\mathbb{P} = \mu_1 \times \cdots \times \mu_n \times \cdots$ be the product probability measure on $\Omega = \prod_{n=1}^{\infty} S_n$. Define $X_n : \Omega \to S$ by $X_n(w) = w_n$ if $w = (w_n) \in \Omega$ for $n = 1, 2, \cdots$. Then X_n are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ (where $\mathcal{F} = \prod_{n=1}^{\infty} \mathcal{G}_n$) and by construction, X_n have the common distribution μ, and (X_n) are independent. ■

6.3 Borel-Cantelli lemma

1. Limiting events, Borel-Cantelli's first and second lemma. Let $A_n \in \mathcal{F}$ for $n = 1, 2, \cdots$. The event that "A_n's occur infinitely often" (or "infinitely many A_n occur") is given by

$$\begin{aligned} \limsup_{n\to\infty} A_n &= \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n \\ &= \{\omega : \omega \text{ belongs to infinitely many } A_n\}. \end{aligned}$$

The event $\limsup_{n\to\infty} A_n$ is also denoted by $\{A_n : \text{ i.o.}\}$. Similarly, though less important in applications, the event that "A_n take place eventually" is

$$\begin{aligned} \liminf_{n\to\infty} A_n &= \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n \\ &= \{\omega : \exists N(\omega) \text{ s.t. } \omega \in A_n \text{ for all } n \geq N(\omega)\} \\ &= \{\omega : \omega \text{ eventually belongs to } A_n \text{ for large } n\}. \end{aligned}$$

This event is denoted sometimes by $\{A_n : \text{ev.}\}$. By definition, it is easy to see that

$$\limsup_{n\to\infty} A_n = \left\{ \sum_{n=\cdot}^{\infty} 1_{A_n} = \infty \right\} = \left\{ \limsup_{n\to\infty} 1_{A_n} = 1 \right\}$$

while

$$\liminf_{n\to\infty} A_n = \left\{ \lim_{n\to\infty} 1_{A_n} = 1 \right\}.$$

Theorem 6.9 *Let $A_n \in \mathcal{F}$ (where $n = 1, 2, \cdots$).*

*1) (*Borel-Cantelli Lemma*, first Borel-Cantelli lemma). If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then* $\mathbb{P}[\limsup_{n\to\infty} A_n] = 0$.

*2) (*Borel zero-one criterion*, second Borel Cantelli lemma). If the events $\{A_n\}$ are independent, then $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ if and only if $\mathbb{P}[\limsup_{n\to\infty} A_n] = 1$.*

Proof. 1) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\lim_{m\to\infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0$, and therefore

$$\mathbb{P}[A_n : \text{ i.o.}] = \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} A_n\right) \leq \lim_{m\to\infty} \sum_{n\geq m} \mathbb{P}(A_n) = 0.$$

2) If A_n are independent, and if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then

$$\begin{aligned}\mathbb{P}\left(\bigcap_{n=m}^{\infty} A_n^c\right) &= \lim_{N\to\infty} \prod_{n=m}^{N} \mathbb{P}(A_n^c) = \lim_{N\to\infty} \prod_{n=m}^{N} (1 - \mathbb{P}(A_n)) \\ &\leq \lim_{N\to\infty} \exp\left(-\sum_{n=m}^{N} \mathbb{P}(A_n)\right) \\ &= 0\end{aligned}$$

for every m, where we have used the elementary inequality: $1 - x \leq e^{-x}$ for $x \in [0, 1]$. Since

$$\{A_n : \text{ i.o}\}^c = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n^c$$

and every $\bigcap_{n=m}^{\infty} A_n^c$ has probability zero, so that their union $\{A_n : \text{ i.o}\}^c$ over $m = 1, 2, \cdots$ has zero probability too, hence $\mathbb{P}[A_n : \text{ i.o}] = 1$. ∎

2. Tail events and tail σ-algebra. The $\limsup A_n$ and $\liminf A_n$ are examples of so-called *tail events* – these events are determined by $\{A_{m+1}, A_{m+1}, \cdots, A_n, \cdots\}$ for every m. For example

$$\limsup_{n\to\infty} A_n = \left\{\sum_{n=m+1}^{\infty} 1_{A_n} = \infty\right\}$$

for any m. From Borel zero-one criterion above, we can deduce the limiting behavior of these tail events by combining with the concept of independence. *If* $X_1, X_2, \cdots, X_n, \cdots$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, then the σ-algebra $\mathcal{G}_{\infty} = \bigcap_{n=1}^{\infty} \sigma\{X_j : j > n\}$ is called the tail σ-algebra of $\{X_k\}_{k\geq 1}$. Any element in $\mathcal{G}_{\infty}$ is called a *tail event.*

Proposition 6.10 *(*A. Kolomogorov's 0-1 law*) If $\{X_n\}$ is a sequence of* independent *random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}_{\infty} = \bigcap_{n=1}^{\infty} \sigma\{X_j : j > n\}$. Then $\mathbb{P}(A) = 0$ or 1 for every $A \in \mathcal{G}_{\infty}$. In particular, if $\{A_n\}$ is a sequence of* independent *events, then $\mathbb{P}[\limsup_{n\to\infty} A_n] = 0$ or 1 .*

Proof of 0-1 law. Since $\sigma\{X_j : j \leq n\}$ and $\sigma\{X_j : j > n\}$ are independent for any $n = 1, 2, \cdots$, so that $\sigma\{X_j : j \leq n\}$ and $\mathcal{G}_{\infty}$ for every n are independent. It follows that $\bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$ and $\mathcal{G}_{\infty}$ are independent. If $B, C \in \bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$, then $B \cap C \in \bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$ as well, so $\bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}$ is a π-system, thus, by Lemma 6.6, the σ-algebra

$$\sigma\left[\bigcup_{n=1}^{\infty} \sigma\{X_j : j \leq n\}\right] = \sigma\{X_j : j \geq 1\}$$

and $\mathcal{G}_\infty$ are independent. Since $\mathcal{G}_\infty \subset \sigma\{X_j : j \geq 1\}$, $\mathcal{G}_\infty$ and itself are independent. Therefore, for every $A \in \mathcal{G}_\infty$, $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, which yields that $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. The last conclusion comes from the fact that $\limsup_{n\to\infty} A_n \in \mathcal{G}_\infty$, so that $\mathbb{P}[\limsup_{n\to\infty} A_n] = 0$ or 1.

3. Example. Suppose (X_n) is a sequence of independent random variables (real or complex), and $\mathcal{G}_\infty$ is its tail σ-algebra, and suppose $\{b_n\}$ be an increasing sequence of positive numbers such that $b_n \uparrow \infty$. Then the following events

$$\left\{\lim_{n\to\infty} X_n \text{ exists}\right\}, \left\{\sum_{n=1}^{\infty} X_n \text{ converges}\right\} \text{ and } \left\{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{b_n} \text{ exists}\right\}$$

are all tail events, i.e. belong to $\mathcal{G}_\infty$, and thus have probability one or zero.

7 Conditional expectations

1) *Definition of Conditional expectations.* Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space, $\mathcal{G} \subseteq \mathcal{F}$ is a sub-algebra and μ is σ-finite on $\mathcal{G}$. If X is a non-negative real random variable which is σ-integrable on $\mathcal{G}$, then there is a $\mathcal{G}$-measurable random variable $\mathbb{E}^\mu[X|\mathcal{G}]$, the conditional expectation of X, which is a unique (up to almost everywhere) function Y such that

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

2) $\mathbb{E}^\mu[Y 1_A] = \mathbb{E}^\mu[X 1_A]$ for every $A \in \mathcal{G}$.

A random variable Y (either non-negative or integrable) which satisfies conditions 1) and 2) above is called the conditional expectation of a random variable X , denoted by $\mathbb{E}^\mu[X|\mathcal{G}]$.

Therefore, if a random variable X is non-negative and σ-integrable on $\mathcal{G}$, then its conditional expectation $\mathbb{E}^\mu[X|\mathcal{G}]$ exists and unique up to almost everywhere.

In what follows, let us work with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subseteq \mathcal{F}$ is a sub σ-algebra. THe conditional expectation of X (if exists) will be denoted by $\mathbb{E}[X|\mathcal{G}]$.

2. Conditional expectations for integrable functions. Suppose X is integrable, thus X^+ and X^- are non-negative, $\mathcal{F}$-measurable and integrable, thus $\mathbb{E}[X^\pm|\mathcal{G}]$ are defined, $\mathcal{G}$-measurable, and integrable. Therefore both $\mathbb{E}[X^\pm|\mathcal{G}]$ are finite almost surely, so that

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}\left[X^+|\mathcal{G}\right] - \mathbb{E}\left[X^-|\mathcal{G}\right]$$

is integrable. $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable and $\mathbb{E}[X : A] = \mathbb{E}[\mathbb{E}(X|\mathcal{G}) : A]$ for every $A \in \mathcal{G}$, so that $\mathbb{E}[X|\mathcal{G}]$ is the conditional expectation of X.

If X is $\mathcal{F}$-measurable and non-negative, then for each n, $X \wedge n$ is bounded and $X \wedge n \uparrow f$. Thus $\mathbb{E}[X \wedge n|\mathcal{G}]$ is defined for each n, and $\mathbb{E}[X \wedge n|\mathcal{G}]$ is increasing, its limit Y exists. Y is $\mathcal{G}$-measurable, and for every $A \in \mathcal{G}$, according to MCT, we have $\mathbb{E}[X : A] = \mathbb{E}[Y : A]$, so that Y is the conditional expectation of X, denoted by $\mathbb{E}[X|\mathcal{G}]$.

3. Example. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $A \in \mathcal{F}$ with $0 < \mathbb{P}(A) < 1$. Let $\mathcal{G} = \sigma(A)$. If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then

$$\mathbb{E}[X|\mathcal{G}] = \frac{\mathbb{E}[X : A]}{\mathbb{P}(A)} \mathbf{1}_A + \frac{\mathbb{E}[X : A^c]}{\mathbb{P}(A^c)} \mathbf{1}_{A^c}.$$

In general, if $\{A_j\}$ is a countable partition of Ω, i.e. $\cup_j A_j = \Omega$, $\{A_j\}$ are disjoint and $\mathbb{P}(A_j) > 0$, then

$$\mathbb{E}[X|\mathcal{G}] = \sum_{j=1}^{\infty} \frac{\mathbb{E}[X : A_j]}{\mathbb{P}(A_j)} \mathbf{1}_{A_j}$$

where $\mathcal{G} = \sigma\{A_j : j = 1, 2, \cdots\}$.

4. Notations. The following convention on conditional expectations will be assumed. If Z is a random variable, then the conditional expectation of X given Z, denoted by $\mathbb{E}[X|Z]$, is defined to be the conditional expectation of X given $\sigma(Z)$. If $Z_1, \cdots, Z_n$ is a finite family of random variables, then we define

$$\mathbb{E}[X|Z_1, \cdots, Z_n] = \mathbb{E}[X|\sigma(Z_1, \cdots, Z_n)].$$

In general, if $\{Z_\alpha\}_{\alpha\in\Lambda}$ is a family of random variables, then

$$\mathbb{E}[X|Z_\alpha; \alpha \in \Lambda] = \mathbb{E}[X|\sigma(\{Z_\alpha\}_{\alpha\in\Lambda})] \ .$$

5. Example. Let X and Z be two random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with continuous joint probability density function $p(x, z)$, i.e.

$$\mathbb{P}\{(X, Z) \in D\} = \iint_D p(x, z)dxdz.$$

Then

$$\mathbb{E}[f(X)|Z] = \frac{\int_{\mathbb{R}} f(x)p(x, Z)dx}{\int_{\mathbb{R}} p(x, Z)dx}$$

where f is Borel measurable, non-negative or/and $f(X)$ is integrable. In fact, formally

$$\begin{aligned}\mathbb{P}[X = x|Z = z] &= \frac{\mathbb{P}(X = x, Z = z)}{\mathbb{P}(Z = z)} \\ &= \frac{p(x, z)}{\int_{\mathbb{R}} p(x, z)dx}.\end{aligned}$$

*6. Properties of the conditional expectation*s.

6.1) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}(X)$, i.e. the expectation of conditional expectation doesn't change. If X is integrable, and X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$. If Z is $\mathcal{G}$-measurable, then$\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$.

6.2) $X \to \mathbb{E}(X|\mathcal{G})$ is linear, additive and positive.

6.3) Convergence Theorems. 6.3.1) MCT for conditional expectations: If $0 \le X_n \uparrow X$ then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$. *8.3.2) Fatou's Lemma:* If $X_n \ge 0$, then $\mathbb{E}[\liminf X_n|\mathcal{G}] \le \liminf \mathbb{E}[X_n|\mathcal{G})]$. *8.3.3) Dominated Convergence*: If $|X_n| \le Z$ for some $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\lim X_n = X$, then $\mathbb{E}[X_n|\mathcal{G}] \Rightarrow \mathbb{E}[X|\mathcal{G}]$.

6.4) If $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{F}$, then $\mathbb{E}\{\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2\} = \mathbb{E}[X|\mathcal{G}_2]$ (this is called the power law for conditional expectations).

7. Jensen's inequality for conditional expectations. If φ is convex, and both X and $\varphi(X)$ are integrable, then

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \le \mathbb{E}[\varphi(X)|\mathcal{G}]$$

almost surely.

Let us prove the Jensen inequality. Recall that φ is convex on $\mathbb{R}$ if

$$\varphi(\lambda s + (1-\lambda)t) \le \lambda\varphi(s) + (1-\lambda)\varphi(t)$$

for all $s, t \in \mathbb{R}$ and $\lambda \in [0,1]$, which is equivalent to that

$$\frac{\varphi(u)-\varphi(s)}{u-s} \leq \frac{\varphi(t)-\varphi(u)}{t-u}$$

for any $s < u < t$ (with $u = \lambda s + (1-\lambda)t$). In particular, the right-derivative

$$\varphi'_{+}(s) = \lim_{t \downarrow s} \frac{\varphi(t)-\varphi(s)}{t-s} = \inf_{t>s} \frac{\varphi(t)-\varphi(s)}{t-s}$$

exists. Similarly

$$\varphi'_{-}(t) = \lim_{s \uparrow t} \frac{\varphi(t)-\varphi(s)}{t-s} = \sup_{s<t} \frac{\varphi(t)-\varphi(s)}{t-s}.$$

and both $t \to \varphi'_{\pm}(t)$ are increasing. By definition, for $s < t$ we have

$$\frac{\varphi(t)-\varphi(s)}{t-s} \leq \varphi'_{-}(t)$$

that is

$$\varphi(s) \geq \varphi(t) + \varphi'_{-}(t)(s-t)$$

for $s < t$. While if $s > t$, then

$$\frac{\varphi(s)-\varphi(t)}{s-t} \geq \varphi'_{+}(t) \geq \varphi'_{-}(t)$$

we thus also have

$$\varphi(s) \geq \varphi(t) + \varphi'_{-}(t)(s-t).$$

Therefore, for a convex function φ, we have

$$\varphi(s) \geq \varphi(t) + \varphi'_{-}(t)(s-t) \text{ for all } s. \tag{7.1}$$

Applying (7.1) $t = \mathbb{E}[X|\mathcal{G}]$ and $s = X$, to obtain

$$\varphi(X) \geq \varphi(\mathbb{E}[X|\mathcal{G}]) + \varphi'_{-}(\mathbb{E}[X|\mathcal{G}])(X - \mathbb{E}[X|\mathcal{G}]).$$

Now $t \to \varphi'_{-}(t)$ is increasing, so that it is Borel measurable, thus $\varphi'_{-}(\mathbb{E}[X|\mathcal{G}])$ is $\mathcal{G}$-measurable. Taking conditional expectation we deduce that

$$\begin{aligned}
\mathbb{E}[\varphi(X)|\mathcal{G}] &\geq \varphi(\mathbb{E}[X|\mathcal{G}]) + \mathbb{E}\left[\varphi'_{-}(\mathbb{E}[X|\mathcal{G}])(X - \mathbb{E}[X|\mathcal{G}])|\mathcal{G}\right] \\
&= \varphi(\mathbb{E}[X|\mathcal{G}]) + \varphi'_{-}(\mathbb{E}[X|\mathcal{G}])\,\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])|\mathcal{G}] \\
&= \varphi(\mathbb{E}[X|\mathcal{G}]).
\end{aligned}$$

8 Uniform integrability

1. Definition of uniform integrability. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The concept of uniform integrability for a family of integrable functions is used to handle the convergence in $L^1(\Omega)$. In spirit, it is very close to that of uniform convergence, uniform continuity etc. that you have learned in the analysis course. If f is integrable, then f is finite almost everywhere. Hence $|f|1_{\{|f|<N\}} \uparrow |f|$ almost everywhere as $N \uparrow \infty$, thus by the Monotone Convergence Theorem $\int_\Omega |f|1_{\{|f|<N\}} d\mathbb{P} \uparrow \int_\Omega |f| d\mathbb{P}$, so that $\lim_{N\to\infty} \int_{\{|f|\geq N\}} |f| \mathrm{d}\mathbb{P} = 0$.

Definition 8.1 *Let $\mathcal{A}$ be a family of integrable functions on $(\Omega, \mathcal{F}, \mu)$. $\mathcal{A}$ is uniformly integrable if*

$$\lim_{N\to\infty} \sup_{\xi\in\mathcal{A}} \int_{\{|\xi|\geq N\}} |\xi| d\mathbb{P} = 0 \ .$$

That is, $\mathbb{E}\left[|\xi| : |\xi| \geq N\right]$ tends to zero uniformly on $\mathcal{A}$ as $N \to \infty$.

2. Some simple properties.

2.1) Any finite family of integrable random variables is uniformly integrable.

2.2) Suppose $\mathcal{A} \subset L^1(\Omega)$ and there is $\eta \in L^1(\Omega)$ such that $|\xi| \leq \eta$ for every $\xi \in \mathcal{A}$, then $\mathcal{A}$ is uniformly integrable.

2.3) $\mathcal{A} \subset L^p(\Omega)$ such that $\sup_{\xi\in\mathcal{A}} \int_\Omega |\xi|^p d\mathbb{P} < \infty$ for some $p > 1$ [which is equivalent to that $\mathcal{A}$ is bounded in $L^p(\Omega)$], then $\mathcal{A}$ is uniformly integrable. In fact,

$$\begin{aligned} \sup_{\xi\in\mathcal{A}} \int_{\{|\xi|\geq N\}} |\xi| \mathrm{d}\mu &\leq \sup_{\xi\in\mathcal{A}} \int_{\{|\xi|\geq N\}} \frac{1}{N^{p-1}} |\xi|^p \mathrm{d}\mu \\ &\leq \frac{1}{N^{p-1}} \sup_{\xi\in\mathcal{A}} \mathbb{E}\left[|\xi|^p\right] \to 0. \end{aligned}$$

Theorem 8.2 *Let $\mathcal{A} \subset L^1(\Omega)$. Then $\mathcal{A}$ is uniformly integrable if and only if*

(a) $\mathcal{A}$ is a bounded subset of $L^1(\Omega)$, that is, $\sup_{\xi\in\mathcal{A}} \mathbb{E}\left[|\xi|\right] < \infty$.

(b) For any $\varepsilon > 0$ there is a $\delta > 0$ such that $\sup_{\xi\in\mathcal{A}} \mathbb{E}\left[|\xi| : E\right] \leq \varepsilon$ whenever $E \in \mathcal{F}$ with $\mu(E) \leq \delta$.

Proof. Suppose $\mathcal{A}$ is uniformly integrable. For any $E \in \mathcal{F}$ and $N > 0$

$$\begin{aligned} \int_E |\xi| \mathrm{d}\mathbb{P} &= \int_{E\cap\{|\xi|<N\}} |\xi| \mathrm{d}\mathbb{P} + \int_{E\cap\{|\xi|\geq N\}} |\xi| \mathrm{d}\mathbb{P} \\ &\leq N + \int_{\{|\xi|\geq N\}} |\xi| \mathrm{d}\mathbb{P} \ . \end{aligned}$$

Given $\varepsilon > 0$, choose $N > 0$ such that $\sup_{\xi\in\mathcal{A}} \mathbb{E}\left[|\xi| : |\xi| \geq N\right] \leq \varepsilon/2$. Then $\sup_{\xi\in\mathcal{A}} \mathbb{E}\left[|\xi| : E\right] \leq N + \varepsilon/2$ for any $E \in \mathcal{F}$. Thus$\delta = \varepsilon/(4N)$ will do.

Conversely, suppose 1) and 2) are satisfied. Let $\beta = \sup_{\xi\in\mathcal{A}} \mathbb{E}\left[|\xi|\right]$. Then, by the Markov inequality, $\mathbb{P}\left\{|\xi| \geq N\right\} \leq \beta/N$ for any $N > 0$. For any $\varepsilon > 0$, there is a $\delta > 0$ such that the inequality in 2) holds. Let $N = \beta/\delta$. Then $\mathbb{P}\left\{|\xi| \geq N\right\} \leq \delta$ so that $\mathbb{E}\left[|\xi| : |\xi| \geq N\right] \leq \varepsilon$ for any $\xi \in \mathcal{A}$. ■

Corollary 8.3 *Suppose $\mathcal{A} \subset L^1(\Omega)$ and $\eta \in L^1(\Omega)$ such that $\mathbb{E}\left[1_D|\xi|\right] \leq \mathbb{E}\left[1_D|\eta|\right]$ for any $D \in \mathcal{F}$ and $\xi \in \mathcal{A}$. Then $\mathcal{A}$ is uniformly integrable.*

3. L^1-convergence and uniform integrability. The following theorem demonstrates the importance of uniform integrability.

Theorem 8.4 *Let f_n be a sequence of integrable functions on $(\Omega, \mathcal{F}, \mathbb{P})$. Then $f_n \to f$ in $L^1(\Omega)$ as $n \to \infty$:*

$$||f_n - f||_{L^1(\Omega)} = \mathbb{E}\left[|f_n - f|\right] \to 0 \quad as \quad n \to \infty \ ,$$

if and only if $\{f_n\}$ is uniformly integrable and $f_n \to f$ in measure as $n \to \infty$.

Proof. *Necessity.* For any $\varepsilon > 0$ there is a natural number m such that $||f_n - f||_{L^1(\Omega)} < \varepsilon/2$ for all $n > m$. Therefore, for every measurable subset E,

$$\sup_n \int_E |f_n| \mathrm{d}\mathbb{P} \le \int_E |f| \mathrm{d}\mathbb{P} + \sup_{k \le m} \int_E |f_k| \mathrm{d}\mathbb{P} + \frac{\varepsilon}{2} \ .$$

In particular

$$\sup_n \mathbb{E}\left[|f_n|\right] \le \mathbb{E}\left[|f|\right] + \sup_{k \le m} \mathbb{E}\left[|f_k|\right] + \frac{\varepsilon}{2}$$

i.e. $\{f_n : n \ge 1\}$ is bounded in $L^1(\Omega)$. Moreover, since $f, f_1, \cdots, f_m$ belong to L^1, so that there is $\delta > 0$ such that, if $\mathbb{P}(E) \le \delta$, then

$$\int_E |f| \mathrm{d}\mathbb{P} + \sum_{k=1}^m \int_E |f_k| \mathrm{d}\mathbb{P} \le \frac{\varepsilon}{2}.$$

Therefore $\sup_n \int_E |f_n| \mathrm{d}\mathbb{P} \le \varepsilon$ as long as $\mu(E) \le \delta$.

Sufficiency. By Fatou's lemma $\int_\Omega |f| \mathrm{d}\mathbb{P} \le \sup_n \int_\Omega |f_n| \mathrm{d}\mathbb{P}$, so that $f \in L^1(\Omega)$. Therefore $\{f_n - f : n \ge 1\}$ is uniformly integrable, thus, by Theorem 8.2, for any $\varepsilon > 0$ there is $\delta > 0$ such that $\int_E |f_n - f| \mathrm{d}\mathbb{P} < \varepsilon$ for any $E \in \mathcal{F}$ satisfying that $\mathbb{P}(E) \le \delta$. Since $f_n \to f$ in probability, there is an $N > 0$ such that $\mathbb{P}\left(|X_n - X| \ge \varepsilon\right) \le \delta$ for any $n \ge N$. Therefore

$$\begin{aligned} \int_\Omega |f_n - f| \mathrm{d}\mathbb{P} &\le \int_{\{|X_n - X| \ge \varepsilon\}} |f_n - f| \mathrm{d}\mathbb{P} + \varepsilon \mathbb{P}\left\{f_n - f| < \varepsilon\right\} \\ &\le \varepsilon + \varepsilon \mathbb{P}\left\{|f_n - f| < \varepsilon\right\} \\ &\le 2\varepsilon \ . \end{aligned}$$

for $n \ge N$. By definition, $f_n \to f$ in $L^1(\Omega)$. ■

9 Martingales in discrete-time

In the 1950's, Doob wrote up a systemic account on the theory of martingales in his book "Stochastic Processes". Doob's book, although about 60 years old, remains very useful to researchers and still in print. The fundamental results in the martingale theory (in the restricted sense) include the optional stopping theorem, martingale inequalities and the martingale convergence theorem.

This chapter is devoted to the theory of martingales in discrete-time. We will only present the basic aspects of this subject with the emphasis on the use of filtrations (information flows), stopping times (random times) and sample paths of stochastic sequences.

In probability theory, we study probabilistic properties of random variables: properties determined by the distributions of random variables. It can be a very subtle problem to give a good description of laws of random variables taking values in infinite dimensional spaces. The classical probability deals with sequences of random variables, such as the law of large numbers, central limit theorems etc., typically starts with the assumption of independence among elements in the sequence. When we consider stochastic processes, that is, parametrized families of random variables, we will be interested in relationships between elements in the family and in particular properties determined by their (finite dimensional) joint distributions.

The basic concepts in the theory of martingales become natural and apparent as we will see, if we are allowed ourselves to use *a family of different σ algebras* on the same sample space instead one fixed collection of events, the technical used to prove deep limiting theorems, which were mastered only by few experts in the past, become systemic tools as long as we accept the notion of random times. It took some years for the probability society to digest these two fundamental ideas, and it took a generation to rewrite our textbooks on probability theory which introduce the basic theory of martingales from the very beginning.

Let us begin with the concept of *filtrations* (which model flows of information).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathbb{Z}_+ = \{0, 1, \cdots\}$ denote the ordered set of non-negative integers, and $\overline{\mathbb{Z}}_+ = \mathbb{Z}_+ \cup \{\infty\}$.

Definition 9.1 *A family $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ of sub σ algebras of $\mathcal{F}$ is called* a filtration*, if $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for every $n \in \mathbb{Z}_+$.*

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ together with a filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ is called a filtered probability space*, denoted by $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$.*

It is useful to consider $\mathcal{F}_n$ as the information available to us up to time n.

Given a sequence of random variables $X = (X_n)_{n\in\mathbb{Z}_+}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for every n, let $\mathcal{F}_n^X$ be the smallest σ-algebra with respect to which $X_0, \cdots, X_n$ are measurable, i.e. $\mathcal{F}_n^X = \sigma\{X_m : m \le n\}$. $\left(\mathcal{F}_n^X\right)$ is called the *filtration* generated by X. A sequence of random variables $X = (X_n)_{n\in\mathbb{Z}_+}$ can be considered as the state of some random process evolving in discrete time $n = 0, 1, 2, \cdots$. For example the value of the share price of a particular company at the end of each trading day. $\mathcal{F}_n^X$ is the information about this random evolution up to time n – that is, the history of the price process. In particular, each X_n is measurable with respect to $\mathcal{F}_n^X$, i.e. $X_n \in \mathcal{F}_n^X$, so that $X = (X_n)_{n\ge 0}$ is *adapted* to the filtration $\left(\mathcal{F}_n^X\right)$, which means that as long as we reach time n, then we know the value taken by the random variable X_n at that time. Here we abuse the system of notations: which doesn't mean X_n is an element of $\mathcal{F}_n^X$, but $\{X_n \in B\} \in \mathcal{F}_n^X$ for every Borel set B, as a convention, here $\{X_n \in B\}$ is the abbreviation of $\{\omega \in \Omega : X_n(\omega) \in B\}$, and the same convention applies to similar situations.

In stochastic analysis, a stochastic process is any parametrized family of random variables valued in an arbitrary (measurable) state space. In this book, however, by a *stochastic process* we will mean a sequence of random variables (X_n), on a filtered probability space. The name "stochastic process" (stochastic derives from the Greek for random) is used to underline the fact we are more concerned with the behavior of a random sequence evolving with time n, and we are not so interested in the properties of the individual random variables, although naturally the distribution of each random variable X_n will contribute to the global and limiting behavior of the whole sequence (X_n).

Definition 9.2 *1) A sequence $(X_n)_{n\in\mathbb{Z}_+}$ of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ is adapted to a filtration $(\mathcal{F}_n)$, if for every $n \in \mathbb{Z}_+$, X_n is $\mathcal{F}_n$-measurable. In this case we say $(X_n)_{n\in\mathbb{Z}_+}$ is an* adapted sequence*, or* adapted process *(with respect to $(\mathcal{F}_n)$).*

2) If $X_0 \in \mathcal{F}_0$ and if X_n is $\mathcal{F}_{n-1}$-measurable for any $n \in \mathbb{N}$, then we say (X_n) is predictable *or* previsible*.*

We may think that the sample point $\omega \in \Omega$ is chosen by the fates and over time the choice is revealed to us through the values taken by the process X_n. Thus at time n the σ-algebra $\mathcal{F}_n$ contains all those sets which can be resolved, i.e. we know if ω is in them or not. That is the meaning of adaptness

For a *predictable sequence* (X_n), you know X_n before the present time n, so it is *previsible* and you can certainly predict it!

Another important concept, *stopping times* [which are random times], allows us to articulate the idea of making a decision about when to stop a process based on the observations of its past behavior. However stopping times have far-reaching applications than its superficial definition. The concept of stopping times really synthesize many important technical like random partitions, localizations etc.

Definition 9.3 *Let $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. A measurable function $T : \Omega \to \{0, 1, 2, \cdots, \infty\}$ [thus it may take value ∞] is called a* stopping time *(with respect to $(\mathcal{F}_n)$; if one wishes to emphasize the underlying filtration in question), if $\{T = n\} \in \mathcal{F}_n$ for every n.*

A stopping time T is a random variable and$\{T = \infty\} \in \mathcal{F}$. Both finite constant time $T \equiv n$ and the infinity time $T \equiv \infty$ are stopping times.

Let $\mathcal{F}_\infty = \sigma\{\mathcal{F}_n : n \in \mathbb{Z}_+\} \subset \mathcal{F}$. If T is a stopping time, then

$$\{T = \infty\} = \Omega \setminus \bigcup_{n=0}^{\infty}\{T = n\} = \bigcap_{n=0}^{\infty}\{T > n\}$$

belongs to $\mathcal{F}_\infty$, and for every n

$$\{T \le n\} = \bigcap_{k=0}^{n}\{T = k\} \in \mathcal{F}_n$$

and

$$\{T > n\} = \{T \le n\}^c \in \mathcal{F}_n$$

for every $n \in \mathbb{Z}_+$.

In the literature prior to the French School establishing the general theory of stochastic processes, stopping times had been called Markov times (for example, see K. Ito and H. P. J. McKean: *Diffusion Processes and Their sample Paths.* Berlin, Springer-Verlag 1965).

Example 9.4 *Let $(X_n)_{n\in\mathbb{Z}_+}$ be an adapted process on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$, and $B \in \mathcal{B}(\mathbb{R})$. Then the first time T at which the process $(X_n)_{n\in\mathbb{Z}_+}$ hits B:*

$$T = \inf\{n \ge 0 : X_n \in B\}$$

is a stopping time with respect to $(\mathcal{F}_n)$. More precisely, T is a random variable defined by

$$T(\omega) = \inf\{n \ge 0 : X_n(\omega) \in B\} \quad \forall \omega \in \Omega$$

together with the convention that $\inf \emptyset = \infty$*. Hence*

$$\{T = n\} = \bigcap_{k=0}^{n-1}\{X_k \in B^c\} \cap \{X_n \in B\}.$$

Since (X_n) is adapted, therefore $\{X_k \in B^c\} \in \mathcal{F}_k$ and $\{X_n \in B\} \in \mathcal{F}_n$, so that $\{T = n\} \in \mathcal{F}_n$. T is a stopping time, called a hitting time.

Hitting times are essentially the only stopping times we are interested in.

Given a stopping time T on $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$, the σ-algebra $\mathcal{F}_T$ representing the information available up to the random time T is the following σ-algebra

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \text{ s.t. } A \cap \{T \le n\} \in \mathcal{F}_n \ \forall n = 0, 1, 2, \cdots\}.$$

Exercise 9.5 *If T is a stopping time on $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$, then $\mathcal{F}_T$ is a σ-algebra. If $T = n$ is a constant time, then $\mathcal{F}_T = \mathcal{F}_n$.*

Theorem 9.6 *Let $(X_n)_{n\in\mathbb{Z}_+}$ be an adapted random sequence on $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$, and T be a stopping time with respect to $(\mathcal{F}_n)$. Define*

$$X_T 1_{\{T<\infty\}}(\omega) = \begin{cases} X_{T(\omega)}(\omega), & \text{if } T(\omega) < \infty, \\ 0, & \text{if } T(\omega) = \infty. \end{cases}$$

Then $X_T 1_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable [In particular $X_T 1_{\{T<\infty\}}$ is a random variable.]

Proof. Since

$$\begin{aligned} \{X_T 1_{\{T<\infty\}} &\leq a\} = \{X_T \leq a, T < \infty\} \\ &= \bigcup_{k=0}^{\infty} \{X_k \leq a, T = k\} \cup \{0 \leq a, T = \infty\} \end{aligned}$$

so $\left\{X_T 1_{\{T<\infty\}}\right\} \in \mathcal{F}_T$. For any $n \in \mathbb{Z}$

$$\left\{X_T 1_{\{T<\infty\}} \leq a\right\} \cap \{T \leq n\} = \bigcup_{k=0}^{n} \{X_k \leq a, T = k\}$$

belongs to $\mathcal{F}_n$ as $\{X_k \leq a\} \cap \{T = k\} \in \mathcal{F}_k$, $k = 0, 1, \cdots, n$. Therefore $\left\{X_T 1_{\{T<\infty\}} \leq a\right\} \in \mathcal{F}_T$, which completes the proof. ∎

Exercise 9.7 *Let $(X_n)_{n\in\mathbb{Z}_+}$ be a sequence of independent random variables with identical distribution:*

$$\mathbb{P}(X_n = 1) = p \ , \quad \mathbb{P}(X_n = 0) = 1 - p$$

where $0 < p < 1$. Let $(\mathcal{F}_n)$ be the filtration generated by (X_n), and

$$\begin{aligned} T_1 &= \inf\{n \geq 1 : X_n = 1\} \ , \\ T_{n+1} &= \inf\{T > T_n : X_n = 1\} \quad \text{if } \ n \geq 1 \ . \end{aligned}$$

T_n is the time that the n-th time 1 occurs in the sequence. Then each T_n is a stopping time, and the sequence

$$T_1, T_2 - T_1, \cdots, T_n - T_{n-1}, \cdots$$

is a sequence of independent, identically distributed (with a geometric distribution).

We now introduce the definition of a martingale. The word *martingale* originated in gambling, describing the double-or-quits strategy or part of a horse's harness. Mathematically it encapsulates the idea of a fair game. That is, whatever information from the past history of the game you use in order to determine your betting strategy, your expected return from playing the game is the same as your current fortune.

Definition 9.8 *Let $X = (X_n)_{n\in\mathbb{Z}_+}$ be an adapted process on a filtered probability space $(\Omega, \mathcal{F}_n, \mathcal{F}, \mathbb{P})$. Suppose each X_n is integrable.*

1) X is a martingale, *if*

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] = X_n \quad a.s. \quad \forall n \in \mathbb{Z}_+.$$

2) X is a super-martingale *if*

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \leq X_n \quad a.s.\ \forall n \in \mathbb{Z}_+.$$

3) X is a sub-martingale *if*

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \geq X_n \quad a.s.\ \forall n \in \mathbb{Z}_+.$$

Exercise 9.9 *1) Prove that, an adapted, integrable random sequence (X_n) is a martingale if and only if*

$$\mathbb{E}\left[X_m|\mathcal{F}_n\right] = X_n \quad a.s.\ \forall m \geq n.$$

State a version of the statement for a super- or sub-martingale.

2) If (X_n) is a martingale, then $\mathbb{E}\left[X_n\right] = \mathbb{E}\left[X_0\right]$ for any n.

3) If (X_n) is a super-martingale, then $n \to \mathbb{E}\left[X_n\right]$ is decreasing, while $n \to \mathbb{E}\left[X_n\right]$ is increasing if (X_n) is a sub-martingale.

Example 9.10 *In these examples we are given a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$.*

1) Martingale by projection. Let $\xi \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ be an integrable random variable [i.e. $\mathbb{E}\left[|\xi|\right] < \infty$], and $X_n = \mathbb{E}\left[\xi|\mathcal{F}_n\right]$. Then (X_n) is a martingale.

2) Random walk. Let $(\xi_n)_{n\in\mathbb{Z}_+}$ be a sequence of adapted and integrable random variables. Suppose ξ_{n+1} and $\mathcal{F}_n$ are independent [i.e. $\sigma\left\{\xi_{n+1}\right\}$ and $\mathcal{F}_n$ are independent]. An example is that $\{\xi_n\}$ is a sequence of independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}_n = \sigma\{\xi_m : m \leq n\}$. Let $X_n = \sum_{k=0}^n \xi_k$ be the partial sum sequence. Then (X_n) is a martingale if $\mathbb{E}\left[\xi_n\right] = 0$ for any n, is a super-martingale if $\mathbb{E}\left[\xi_n\right] \leq 0$, and a sub-martingale if $\mathbb{E}\left[\xi_n\right] \geq 0$ for any n.

3) Likelihood ratios. Let f, g be two probability density functions, with support on the whole of $\mathbb{R}$. Let (X_n) be a sequence of independent, identically distributed random variables from the distribution with probability density function f. The likelihood ratio is given by

$$R_n = \frac{g(X_1)g(X_2)\ldots g(X_n)}{f(X_1)f(X_2)\ldots f(X_n)}$$

with $R_0 = 1$. Then (R_n) is a martingale with respect to the filtration generated by X.

4) Polya's Urn. At time $t = 0$ an urn contains 1 red and 1 black ball. At each time a ball is chosen randomly from the urn and it is then replaced along with another ball of the same color. Thus at the time of the n-th draw there are $n+2$ balls in the urn and we let B_n be the number of black balls. Then $M_n = B_n/(n+2)$ is a martingale with respect to the filtration generated by B_n.

Example 9.11 *[*Martingale transform, discrete stochastic integral*] If (H_n) is a* predictable *process and (X_n) is a martingale, then*

$$(H.X)_n = \sum_{k=1}^n H_k(X_k - X_{k-1}),\ (H.X)_0 = 0$$

is a martingale.

Exercise 9.12 *1) If (X_n) and (Y_n) are two martingales (resp. super-martingale), so is $(X_n + Y_n)$.*

2) Show that $(X_n \wedge Y_n)$ is a super-martingale, where (X_n) and (Y_n) are two martingales. In fact, since $Z_n = \min\left\{X_n, Y_n\right\}$ so that

$$\mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] \leq \mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \leq X_n$$

and also

$$\mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] \leq \mathbb{E}\left[Y_{n+1}|\mathcal{F}_n\right] \leq Y_n$$

hence $\mathbb{E}\left[Z_{n+1}|\mathcal{F}_n\right] \leq Z_n$*, thus* Z *is also a super-martingale.*

Recall Jensen's inequality for conditional expectation: if $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, ξ, $\varphi(\xi) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$, then

$$\varphi(\mathbb{E}\left[\xi|\mathcal{G}\right]) \leq \mathbb{E}\left[\varphi(\xi)|\mathcal{G}\right].$$

Functions $(t \ln t)\, 1_{(1,\infty)}(t)$, $t^+ = t 1_{(0,\infty)}$ and $|t|^p$ (for $p \geq 1$) are examples of convex functions.

Theorem 9.13 *1) Let* (X_n) *be a* martingale*, and* $\varphi : \mathbb{R} \to \mathbb{R}$ *be a convex function. Suppose* $\varphi(X_n)$ *are integrable for every* n*. Then* $\{\varphi(X_n)\}$ *is a sub-martingale.*

2) Let (X_n) *be a sub-martingale, and* $\varphi : \mathbb{R} \to \mathbb{R}$ *be* increasing *and convex. Suppose* $\varphi(X_n)$ *are integrable for every* n*, then* $\{\varphi(X_n)\}$ *is a sub-martingale.*

Proof. 1) In fact, applying Jensen's inequality

$$\begin{aligned} \varphi(X_n) &= \varphi\left(\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right]\right) \quad \text{(martingale property)} \\ &\leq \mathbb{E}\left[\varphi(X_{n+1})|\mathcal{F}_n\right] \quad \text{(Jensen's inequality).} \end{aligned}$$

which proved 1). The proof of 2) is similar. ■

$t^+ = \max\{t, 0\} = t 1_{(0,\infty)}$ is increasing and convex, thus, if (X_n) is a sub-martingale, so is $X_n^+ = \max\{X_n, 0\}$. If $X = (X_n)$ is a super-martingale, then $-X_n$ is a sub-martingale, so that $X_n^- = \max\{-X_n, 0\}$ is a sub-martingale. That is, the positive part of a sub-martingale is again a sub-martingale, while the *negative part of a super-martingale* is however a *sub-martingale*. Therefore, if X_n is a martingale, then both its positive part and its negative part are sub-martingales, so is its absolute value $|X_n| = X_n^+ + X_n^-$.

10 Martingale inequalities

In this section we prove the fundamental martingale inequalities.

We first establish Doob's optional sampling theorem which shows that the (super-, sub-)martingale property holds at bounded stopping times.

Theorem 10.1 *[*Doob's optional stopping theorem*] Let* (X_n) *be a martingale (resp. super-martingale), and* $S \leq T$ *two* bounded stopping times*. Then* $\mathbb{E}\left[X_T|\mathcal{F}_S\right] = X_S$ *(resp.* $\mathbb{E}\left[X_T|\mathcal{F}_S\right] \leq X_S$*).*

Proof. [*The proof is not examinable.*] By Theorem 9.6, $X_T \in \mathcal{F}_T$, $X_S \in \mathcal{F}_S$. Suppose S and T are bounded above by N, then

$$\mathbb{E}\left[|X_T|\right] = \sum_{j=0}^{N} \mathbb{E}\left[|X_j| 1_{\{T=j\}}\right] \leq \sum_{j=0}^{N} \mathbb{E}\left[|X_j|\right],$$

so X_T is integrable. Similarly X_S is integrable too.

We have to prove that $\mathbb{E}\left[X_T : A\right] \leq \mathbb{E}\left[X_S : A\right]$ for every $A \in \mathcal{F}_S$.

Let $A \in \mathcal{F}_S$. Since S and T are stopping times, $A \cap \{S = j\} \in \mathcal{F}_j$, $\{T > j\} \in \mathcal{F}_j$ for $j = 0, \cdots, N-1$, so that

$$A_j \equiv A \cap \{S = j\} \cap \{T > j\} \in \mathcal{F}_j$$

and $A \cap \{S < T\} = \bigcup_{j=0}^{N} A_j$ is a disjoint decomposition.

1) If $0 \le T - S \le 1$, then $X_T = X_{j+1}$ and $X_S = X_j$ on A_j for $j = 0, \cdots, N-1$, and therefore

$$\mathbb{E}[X_S - X_T : A] = \mathbb{E}[X_S - X_T : A \cap \{S < T\}] = \sum_{j=0}^{N-1} \mathbb{E}[X_j - X_{j+1} : A_j]$$

However, X is a super-martingale and $A_j \in \mathcal{F}_j$, so that $\mathbb{E}[X_{j+1} : A_j] \le \mathbb{E}[X_j : A_j]$. That is, $\mathbb{E}[X_j - X_{j+1} : A_j] \ge 0$ for $j = 0, \cdots, N-1$, and therefore $\mathbb{E}[X_S - X_T : A] \ge 0$, which is equivalent to that $\mathbb{E}[X_S : A] \ge \mathbb{E}[X_T : A]$ for every $A \in \mathcal{F}_S$.

2) In general, let $R_j = T \wedge (S + j)$, $j = 1, \cdots, n$. Then R_j are stopping times, and $S \le R_1 \le \cdots \le R_n = T$. Moreover $R_1 - S \le 1$ and $R_{j+1} - R_j \le 1$ for $1 \le j \le N-1$. Let $A \in \mathcal{F}_S$. Then $A \in \mathcal{F}_{R_j}$ as $S \le R_j$. Therefore by applying the first case to R_j we obtain

$$\mathbb{E}[X_S : A] \ge \mathbb{E}[X_{R_1} : A] \ge \cdots \ge \mathbb{E}[X_T : A]$$

so that $\mathbb{E}[1_A X_S] \ge \mathbb{E}[1_A X_T]$. The proof is complete. ■

Let us first deduce several easy but important consequences from Doob's optional stopping theorem.

Corollary 10.2 *Let $X = (X_n)$ be a* super-martingale*.*

1) If $T \ge S$ are two bounded stopping times*, then $\mathbb{E}[X_T] \le \mathbb{E}[X_S]$.*

2) If T is a stopping time, then $\mathbb{E}[X_{T\wedge n}] \le \mathbb{E}[X_{T\wedge m}]$ for any $n \ge m$, where $X_{T\wedge n} = X_T$ on $\{T \le n\}$ and $X_{T\wedge n} = X_n$ on $\{T > n\}$.

Similar conclusions hold for sub-martingales.

Corollary 10.3 *If $X = (X_n)$ is a super-martingale, and T is a stopping time, then*

$$\mathbb{E}[|X_{T\wedge n}|] \le \mathbb{E}[X_0] + 2\mathbb{E}\left[X_n^-\right] \quad \forall n \in \mathbb{Z}_+.$$

If in addition $\sup_n \mathbb{E}[|X_n|] < \infty$*, then*

$$\mathbb{E}\left[|X_T| 1_{\{T<\infty\}}\right] \le 3 \sup_n \mathbb{E}[|X_n|].$$

Proof. According to Theorem 10.1, since (X_n^-) is a sub-martingale, together the following equality

$$|X_{T\wedge n}| = X_{T\wedge n}^+ + X_{T\wedge n}^- = X_{T\wedge n} + 2X_{T\wedge n}^-$$

we have

$$\begin{aligned} \mathbb{E}[|X_{T\wedge n}|] &= \mathbb{E}[X_{T\wedge n}] + 2\mathbb{E}\left[X_{T\wedge n}^-\right] \\ &\le \mathbb{E}[X_0] + 2\mathbb{E}\left[X_n^-\right] \end{aligned}$$

which is the first inequality. It follows that

$$\mathbb{E}\left[|X_{T\wedge n}| 1_{\{T<\infty\}}\right] \le 3 \sup_n \mathbb{E}[|X_n|] \tag{10.1}$$

for every n. Since

$$|X_T| 1_{\{T<\infty\}} = \lim_{n\to\infty} |X_{T\wedge n}| 1_{\{T<\infty\}}$$

and applying Fatou's lemma to $|X_{T\wedge n}|1_{\{T<\infty\}}$, we obtain

By :

$$\mathbb{E}\left[|X_T|1_{\{T<\infty\}}\right] = \mathbb{E}\left[\lim_{n\to\infty}|X_T|1_{\{T<\infty\}}\right] \le \liminf_{n\to\infty}\mathbb{E}\left[|X_{T\wedge n}|1_{\{T<\infty\}}\right] \le 3\sup_n \mathbb{E}\left[|X_n|\right]$$

where the last inequality follows from (10.1). ■

Theorem 10.4 *(*Stopped super-martingales are super-martingales*) Suppose $X = (X_n)$ is a super-martingale, and suppose T is a stopping time, then the stopped process $X^T = (X_{T\wedge n})$ is again a super-martingale. A similar result holds for martingales and sub-martingales.*

Proof. According to the previous corollary, we know that $X_{T\wedge n}$ is integrable for every $n\in\mathbb{Z}$. For $n \ge m$ we have

$$\begin{aligned}
\mathbb{E}\left[X_{T\wedge n}|\mathcal{F}_m\right] &= \sum_{k=0}^{n}\mathbb{E}\left[X_k 1_{\{T=k\}}|\mathcal{F}_m\right] + \mathbb{E}\left[X_{T\wedge n}1_{\{T>m\}}|\mathcal{F}_m\right] \\
&= \sum_{k=0}^{m} X_k 1_{\{T=k\}} + \mathbb{E}\left[X_{T\wedge n}1_{\{T>m\}}|\mathcal{F}_m\right] \\
&= \sum_{k=0}^{m} X_k 1_{\{T=k\}} + 1_{\{T>m\}}\mathbb{E}\left[X_{T\wedge n}1_{\{T>m\}}|\mathcal{F}_m\right]. \qquad (10.2)
\end{aligned}$$

where we have used the fact that $\{T>m\}\in\mathcal{F}_m$. Let $S = T1_{\{T>m\}} + \infty 1_{\{T\le m\}}$. Then S is a stopping time. In fact, if $k\le m$, then $\{S=k\}=\emptyset$, and if $k>m$, then

$$\{S=k\} = \{T=k\}\cap\{T>m\}\in\mathcal{F}_k$$

as $\{T>m\}\in\mathcal{F}_m\subseteq\mathcal{F}_k$. By definition $S\wedge n\ge m$, and

$$X_{S\wedge n} = X_{T\wedge n}1_{\{T>m\}} + X_n 1_{\{T\le m\}}.$$

Hence, by applying Doob's stoping theorem to X and bounded stopping times $S\wedge n\ge m$ we obtain

$$\mathbb{E}\left[X_{S\wedge n}|\mathcal{F}_m\right] \le X_m$$

that is

$$\mathbb{E}\left[X_{T\wedge n}1_{\{T>m\}} + X_n 1_{\{T\le m\}}|\mathcal{F}_m\right] \le X_m.$$

Since $\{T\le m\}\in\mathcal{F}_m$, it follows that

$$\mathbb{E}\left[X_{T\wedge n}1_{\{T>m\}}|\mathcal{F}_m\right] + 1_{\{T\le m\}}\mathbb{E}\left[X_n|\mathcal{F}_m\right] \le X_m.$$

Thus, by multiplying both sides by $1_{\{T>m\}}$, we have

$$1_{\{T>m\}}\mathbb{E}\left[X_{T\wedge n}1_{\{T>m\}}|\mathcal{F}_m\right] \le X_m 1_{\{T>m\}}. \qquad (10.3)$$

Putting together (10.2) with (10.3) we obtain that

$$\mathbb{E}\left[X_{T\wedge n}|\mathcal{F}_m\right] \le \sum_{k=0}^{m} X_k 1_{\{T=k\}} + X_m 1_{\{T>m\}} = X_{T\wedge m}$$

which means that X^T is again a super-martingale. ■

Corollary 10.5 *Let T be a* finite *stopping time.*

1) If $X = (X_n)$ is a non-negative *super-martingale, then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.*

2) If $X = (X_n)$ is a super-martingale, and there is an integrable random variable ξ such that $|X_n| \leq \xi$ almost everywhere on Ω for all n, then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.

Proof. 1) In fact, since T is finite, $X_{T\wedge n} \to X_T$ as $n \to \infty$. By Fatou's lemma we have

$$\mathbb{E}[X_T] \leq \liminf_{n\to\infty} \mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_0]$$

which completes the proof.

2) This time we apply the Dominated Convergence Theorem to $\{X_{T\wedge n}\}$ to obtain $\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{T\wedge n}]$. ∎

Corollary 10.6 *Let T be a* finite *stopping time, and $X = (X_n)$ be a super-martingale. Let $\xi = \sup_{n=1,2,\cdots} |X_n - X_{n-1}|$. Suppose ξT is integrable, i.e. $\mathbb{E}[\xi T] < \infty$, then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$. In particular, if the sequence $|X_n - X_{n-1}| \leq L$ for every n, where L is a constant, and if T is an* integrable *stopping time, then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.*

Proof. For every n, we have

$$|X_{T\wedge n}| = \left| X_0 + \sum_{k=1}^{n\wedge T} (X_k - X_{k-1}) \right| \leq |X_0| + \xi T.$$

Since $|X_0| + \xi T$ is integrable, and $X_{T\wedge n} \to X_0$ almost everywhere, by Lebesgue's Dominated Convergence Theorem, $\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{T\wedge n}]$ which yields the conclusion. ∎

In order to establish a general result such as 2) in Corollary 10.5, the concept of uniform integrability may be useful. For example, we have the following

Corollary 10.7 *Let T be a* finite *stopping time, and $X = (X_n)$ be a super-martingale. Suppose $\{X_{T\wedge n} : n = 0, 1, 2, \cdots\}$ is uniformly integrable, then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.*

The proof is eacatly the same as that of 2), Corollary 10.5. In fact, since $X_{T\wedge n} \to X_T$ and $\{X_{T\wedge n}\}$ is uniformly integrable, by Theorem 8.4, $\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{T\wedge n}]$.

It is therefore useful to introduce the following definition.

Definition 10.8 *Let $X = (X_n)_{n\in\mathbb{Z}_+}$ be an adapted sequence of real random variables on a filtred probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$. Let $\mathcal{T}$ denote the collection of all* finite $(\mathcal{F}_n)$-stopping times. *Then we say $X = (X_n)$ is of* class D, *if the family $\{X_T : T \in \mathcal{T}\}$ is uniformly integrable.*

Next we derive the main martingale inequalities, as applications of Doob's optional stopping theorem. Let us introduce a notation first.

If $(X_n)_{n\in\mathbb{Z}_+}$ is a sequence of real random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, for each $n \in \mathbb{Z}_+$, set $X_n^*(\omega) = \max_{k\leq n} X_k(\omega)$ for $\omega \in \Omega$. Then (X_n^*) is called the sequence of running maximal of (X_n). It is obvious that each X_n^* is a random variable. If $(X_n)_{n\in\mathbb{Z}_+}$ is an adapted sequence on the filtered space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$, then so is its running maximal.

Theorem 10.9 [Doob's maximal inequality] *1) If $Y = (Y_n)$ is a sub-martingale, then*

$$\mathbb{P}\left[Y_n^* \geq \lambda\right] \leq \frac{1}{\lambda}\mathbb{E}\left[Y_n : Y_n^* \geq \lambda\right] \tag{10.4}$$

for any $\lambda > 0$. Since $|Y_n|$ is also a sub-martingale, so that

$$\mathbb{P}\left[\sup_{k\leq n}|Y_k| \geq \lambda\right] \leq \frac{1}{\lambda}\mathbb{E}\left[|Y_n| : \sup_{k\leq n}|Y_k| \geq \lambda\right].$$

2) If $X = (X_n)$ is a super-martingale*, then*

$$\mathbb{P}\left[X_n^* \geq \lambda\right] \leq \frac{1}{\lambda}\left(\mathbb{E}\left[X_0\right] - \mathbb{E}\left[X_n : X_n^* \leq \lambda\right]\right)$$

for any $\lambda > 0$, $n \in \mathbb{Z}_+$, and

$$\mathbb{P}\left[\sup_{k\leq n}|X_k| \geq \lambda\right] \leq \frac{1}{\lambda}\left(\mathbb{E}\left[X_0\right] + 2\mathbb{E}\left[X_n^-\right]\right) \tag{10.5}$$

for all $\lambda > 0$, where $X_n^- = \max\{0, -X_n\}$ which is a sub-martingale.

Proof. [*The proof is not examinable.*] We give two proofs of 1). Let $R = \inf\{k \geq 0 : Y_k \geq \lambda\}$ and $T = R \wedge n$. Then $T \leq n$, and we have the following facts: $Y_R \geq \lambda$ on $\{R < \infty\}$, $\{Y_n^* \geq \lambda\} \subseteq \{Y_T \geq \lambda\}$ and $\{Y_n^* < \lambda\} \subset \{T = n\}$. Apply Doob's stopping theorem to $n \geq T$ to obtain

$$\begin{aligned}\mathbb{E}\left[Y_n\right] \geq \mathbb{E}\left[Y_T\right] &= \mathbb{E}\left[Y_T : Y_n^* \geq \lambda\right] + \mathbb{E}\left[Y_T : Y_n^* < \lambda\right] \\ &\geq \lambda\mathbb{P}\left[Y_n^* \geq \lambda\right] + \mathbb{E}\left[Y_n : Y_n^* < \lambda\right].\end{aligned}$$

Rearrange the inequality to deduce that

$$\mathbb{P}\left[Y_n^* \geq \lambda\right] \leq \frac{1}{\lambda}\mathbb{E}\left[Y_n : Y_n^* \geq \lambda\right].$$

Here is a proof without using the notion of stopping times but partition techniques. Let $E_j = \{Y_k < \lambda \text{ for } k \leq j-1, Y_j \geq \lambda\}$ where $j = 0, 1, \cdots, n$. Then $\{Y_n^* \geq \lambda\} = \cup_j E_j$, E_j are disjoint, and $E_j \subseteq \mathcal{F}_j$. Therefore

$$\begin{aligned}\mathbb{P}\left[Y_n^* \geq \lambda\right] &= \sum_{j=0}^n \mathbb{P}\left[E_j\right] \leq \sum_{j=0}^n \frac{1}{\lambda}\mathbb{E}\left[Y_j : E_j\right] \\ &\leq \sum_{j=0}^n \frac{1}{\lambda}\mathbb{E}\left[Y_n : E_j\right] = \frac{1}{\lambda}\mathbb{E}\left[Y_n : Y_n^* \geq \lambda\right].\end{aligned}$$

2) Let $R = \inf\{k \geq 0 : X_k \geq \lambda\}$ and $T = R \wedge n$. Applying Doob's optional theorem to stopping times T and $S = 0$, one has

$$\begin{aligned}\mathbb{E}\left[X_0\right] \geq \mathbb{E}\left[X_T\right] &= \mathbb{E}\left[X_T : X_n^* \geq \lambda\right] + \mathbb{E}\left[X_T : X_n^* < \lambda\right] \\ &\geq \lambda\mathbb{P}\left[\sup_{k\leq n} X_k \geq \lambda\right] + \mathbb{E}\left[X_n : X_n^* < \lambda\right].\end{aligned}$$

To prove the last estimate, we combine 1) with the first inequality of 2). Since $X = (X_n)$ is a super-martingale, applying 1) to the sub-martingale $(-X_n)$ we have

$$\begin{aligned}\mathbb{P}\left[\inf_{k\le n} X_k \le -\lambda\right] &= \mathbb{P}\left[\sup_{k\le n}(-X_k) \ge \lambda\right] \\ &\le \frac{1}{\lambda}\mathbb{E}\left[-X_n : \inf_{k\le n} X_k \le -\lambda\right]\end{aligned}$$

together with the first inequality of 2) we deduce that

$$\begin{aligned}\mathbb{P}\left[\sup_{k\le n}|X_k| \ge \lambda\right] &= \mathbb{P}\left[\sup_{k\le n} X_k \ge \lambda, \text{ or } \inf_{k\le n} X_k \le -\lambda\right] \\ &\le \mathbb{P}\left[\sup_{k\le n} X_k \ge \lambda\right] + \mathbb{P}\left[\inf_{k\le n} X_k \le -\lambda\right] \\ &\le \frac{1}{\lambda}\mathbb{E}\left[X_0\right] - \frac{1}{\lambda}\mathbb{E}\left[X_n : X_n^* \le \lambda\right] + \frac{1}{\lambda}\mathbb{E}\left[-X_n : \inf_{k\le n} X_k \le -\lambda\right] \\ &\le \frac{1}{\lambda}\left(\mathbb{E}\left[X_0\right] + 2\mathbb{E}\left[X_n^-\right]\right)\end{aligned}$$

which is the last inequality. ∎

The following result plays a key role in proving the strong law of large numbers, which a strong version of the elementary Markov inequality.

Theorem 10.10 *[*Kolmogorov's inequality*] Let (X_n) be a martingale and $X_N \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ where N is a positive integer. Then for any $\lambda > 0$*

$$\mathbb{P}\left[\sup_{k\le N}|X_k| \ge \lambda\right] \le \frac{1}{\lambda^2}\mathbb{E}\left[X_N^2\right]. \tag{10.6}$$

Proof. By Jensen's inequality, for any $k \le N$

$$\mathbb{E}\left[X_k^2\right] = \mathbb{E}\left(\mathbb{E}\left[X_N|\mathcal{F}_k\right]\right)^2 \le \mathbb{E}\left[X_N^2\right] < \infty.$$

[That is (X_n) is a square-integrable martingale up to N]. Therefore (X_k^2) $(k = 0, 1, \cdots, N)$ is a sub-martingale (up to time N). Applying Doob's maximal inequality one obtains

$$\mathbb{P}\left[\sup_{k\le n} X_k^2 \ge \lambda^2\right] \le \frac{1}{\lambda^2}\mathbb{E}\left[X_n^2 : \sup_{k\le n} X_k^2 \ge \lambda^2\right] \le \frac{1}{\lambda^2}\mathbb{E}\left[X_n^2\right]$$

for all $n \le N$. ∎

Example 10.11 *Let (X_n) be independent and square integrable. Then $S_n = \sum_{k=0}^n (X_k - \mu_k)$ where $\mu_k = \mathbb{E}\left[X_k\right]$ is a martingale. Moreover*

$$\mathbb{E}\left[S_n^2\right] = \mathbb{E}\left[\sum_{k=0}^n (X_k - \mu_k)\right]^2 = \sum_{k=0}^n \sigma_k^2$$

where $\sigma_k^2 = var(X_k)$. According to Kolmogorov's inequality

$$\mathbb{P}\left[\sup_{k\le n}\left|\sum_{l=0}^k (X_l - \mu_l)\right| \ge \lambda\right] \le \frac{1}{\lambda^2}\sum_{k=0}^n \sigma_k^2$$

for any $\lambda > 0$.

Doob's maximal inequality is a tail estimate for the distribution of the running maximum of a martingale, thus can be used to estimate the L^p-norm, which is the context of Doob's L^p-inequality.

Let us begin with an elementary lemma which follows from Fubini's theorem directly.

Lemma 10.12 *Suppose ρ is right-continuous, increasing on $(0,\infty)$ and $\rho(0+)=0$, and ξ is a non-negative random variable on $(\Omega,\mathcal{F},\mathbb{P})$, then*

$$\rho(\xi)=\rho(\xi)-\rho(0+)=\int_{(0,\xi]} m_\rho(d\lambda) \quad \text{on } \{\xi>0\}$$

$$\begin{aligned}\mathbb{E}\left[\rho(\xi):\xi>0\right] &= \mathbb{E}\left[\int_{(0,\xi]} m_\rho(d\lambda):\xi>0\right]=\mathbb{E}\left[\int_{(0,\infty)} 1_{\{\lambda\le\xi\}} m_\rho(d\lambda)\right] \\ &= \int_{\Omega\times(0,\infty)} 1_{\{\xi\ge\lambda\}} m_\rho(d\lambda)d\mathbb{P}=\int_{(0,\infty)} \mathbb{P}\left[\xi\ge\lambda\right]\mu_\rho(d\lambda),\end{aligned}$$

where $m_\rho(d\lambda)$ is the Lebesgue-Stieltjes measure defined by ρ on $(0,\infty)$, so that $m_\rho((s,t])=\rho(t)-\rho(s)$ for any $t\ge s\ge 0$.

Theorem 10.13 *[*Doob's L^p-inequality*] 1) If (X_n) is a* non-negative sub-martingale*, then, for any $p>1$*

$$\mathbb{E}\left[|X_n^*|^p\right]\le\left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|X_n|^p\right]. \tag{10.7}$$

2) If (X_n) is a martingale, then, for any $p>1$,

$$\mathbb{E}\left[\max_{k\le n}|X_k|^p\right]\le\left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|X_n|^p\right] \tag{10.8}$$

The last inequality may be reformulated in terms of the L^p-norm as

$$\|X_n^*\|_p\le q\,\|X_n\|_p$$

where $\frac{1}{p}+\frac{1}{q}=1$, $\|\cdot\|_p$ denotes the L^p-norm, and $X_n^*=\max_{k\le n}X_k$ is the running maximum.

Proof. If (X_n) is a martingale, then $(|X_n|)$ is a sub-martingale, so (10.8) follows from (10.7). Let us prove the first conclusion. According to Doob's maximal inequality (10.4),

$$\mathbb{P}\left[X_n^*\ge\lambda\right]\le\frac{1}{\lambda}\mathbb{E}\left[X_n;X_n^*\ge\lambda\right].$$

If ρ is right continuous, increasing and $\rho(0+)=0$ on $(0,\infty)$, then, by Lemma 10.12

$$\begin{aligned}\mathbb{E}\left[\rho(X_n^*):X_n^*>0\right]&=\mathbb{E}\left[\int_{(0,X_n^*)} m_\rho(d\lambda):X_n^*>0\right]=\int_{(0,\infty)}\mathbb{P}\left[X_n^*\ge\lambda\right]m_\rho(d\lambda)\\ &\le\int_{(0,\infty)}\frac{1}{\lambda}\mathbb{E}\left[X_n:\,X_n^*\ge\lambda\right]m_\rho(d\lambda)\\ &=\int_{(0,\infty)}\left\{\frac{1}{\lambda}\int_{\{X_n^*\ge\lambda\}}X_n d\mathbb{P}\right\}m_\rho(d\lambda)\\ &=\mathbb{E}\left[X_n\left(\int_{(0,X_n^*]}\frac{1}{\lambda}m_\rho(d\lambda)\right):X_n^*>0\right].\end{aligned}$$

Choosing $\rho(\lambda) = \lambda^p$, then $\rho'(\lambda) = p\lambda^{p-1}$, we obtain

$$\begin{aligned}\mathbb{E}\left[|X_n^*|^p\right] = \mathbb{E}\left[|X_n^*|^p : X_n^* > 0\right] &\leq \mathbb{E}\left[X_n\left(\int_{(0,X_n^*]}\frac{1}{\lambda}m_\rho(d\lambda)\right) : X_n^* > 0\right] \\ &= \frac{p}{p-1}\mathbb{E}\left[X_n\left(X_n^*\right)^{p-1}\right] \\ &\leq q(\mathbb{E}|X_n|^p)^{\frac{1}{p}}(\mathbb{E}|X_n^*|^p)^{\frac{1}{q}}\end{aligned}$$

for $\frac{1}{p}+\frac{1}{q}=1$, here the last inequality we have used the Holder inequality

$$\int_\Omega |fg|d\mu \leq \|f\|_p \|g\|_q$$

if $p > 1$ and $\frac{1}{p}+\frac{1}{q}=1$. Rearranging the inequality above to obtain the last inequality in 2).
■

Doob's L^p-inequality does not apply to the case $p = 1$, as in this case $q = \infty$ which gives the infinity upper bound. That is to say, the L^1-norm of the terminal value of a martingale does not in general control the L^1-norm of its running maximal.

Exercise 10.14 *Prove that* $\log x \leq x/e$ *for all* $x > 0$, *hence prove that*

$$a\log^+ b \leq a\log^+ a + \frac{b}{e}. \tag{10.9}$$

Consider $h(t) = \log t - \frac{t}{e}$ *for* $t > 0$. *Then* $h(t) \to -\infty$ *as* $t \downarrow 0$ *or* $t \uparrow \infty$, *so* h *achieves its maximum in* $(0,\infty)$. *Since* $h'(t) = \frac{1}{t} - \frac{1}{e}$ *has unique root* $t = e$, e *is the maximum of* h. *Therefore* $h(t) \leq h(e) = 0$ *for all* $t > 0$, *that is,* $\log t \leq \frac{t}{e}$.

Now

$$\begin{aligned}\log^+(at) &= \max\{0, \log(at)\} = \max\{0, \log a + \log t\} \\ &\leq \max\left\{0, \log^+ a + \frac{t}{e}\right\} = \log^+ a + \frac{t}{e},\end{aligned}$$

Setting $t = \frac{b}{a}$ *we obtain the inequality (10.9).*

Theorem 10.15 *If* (X_n) *is a non-negative sub-martingale, then*

$$\mathbb{E}\left[\max_{k\leq n} X_k\right] \leq \frac{e}{e-1}\left(1 + \mathbb{E}\left[X_n \log^+ X_n\right]\right). \tag{10.10}$$

Proof. [*The proof is not examinable.*] We have seen from the proof of Doob's L^p-inequality

$$\mathbb{E}\left[\rho(X_n^*) : X_n^* > 0\right] \leq \mathbb{E}\left[X_n\int_{(0,X_n^*]}\frac{1}{\lambda}m_\rho(d\lambda) : X_n^* > 0\right].$$

where now $\rho(\lambda) = (\lambda - 1)^+$ which is a continuous increasing function with support on $[1,\infty)$. Therefore

$$\begin{aligned}\mathbb{E}\left[\rho(X_n^*)\right] = \mathbb{E}\left[\rho(X_n^*) : X_n^* > 0\right] &\leq \mathbb{E}\left[X_n\int_1^{X_n^*}\frac{1}{\lambda}d\lambda : X_n^* \geq 1\right] \\ &= \mathbb{E}\left[X_n \log^+ X_n^*\right] \\ &\leq \mathbb{E}\left[X_n \log^+ X_n\right] + \frac{1}{e}\mathbb{E}\left[X_n^*\right].\end{aligned}$$

where we have used the inequality

$$X_n \log X_n^* \leq X_n \log^+ X_n + \frac{X_n^*}{e}.$$

On the other hand

$$\begin{aligned}
\mathbb{E}\left[X_n^*\right] &= \mathbb{E}\left[X_n^* 1_{\{X_n^* \geq 1\}}\right] + \mathbb{E}\left[X_n^* 1_{\{X_n^* < 1\}}\right] \\
&\leq \mathbb{E}\left[\rho(X_n^*) 1_{\{X_n^* \geq 1\}}\right] + \mathbb{E}\left[1_{\{X_n^* > 1\}}\right] + \mathbb{E}\left[X_n^* 1_{\{X_n^* < 1\}}\right] \\
&\leq \mathbb{E}\left[\rho(X_n^*)\right] + 1.
\end{aligned}$$

Together with the previous estimate one thus deduces that

$$\mathbb{E}\left[X_n^*\right] \leq 1 + \mathbb{E}\left[X_n \log^+ X_n\right] + \frac{\mathbb{E}\left[X_n^*\right]}{e}$$

which yields the L^1-estimate. ■

11 The martingale convergence theorem

An important field in the probability theory is to study the asymptotic behavior of sequences of random variables. For example, we are interested in whether a sequence $\{X_n : n \geq 0\}$ converges or not as $n \to \infty$. One of the powerful tools to study the convergence of random sequences is the concept of up-crossing numbers through intervals by a random sequence.

Suppose (a_n) is a sequence of real numbers, then $\lim_{n\to\infty} a_n$ exists (as a real number), $\lim_{n\to\infty} a_n = \infty$ or $\lim_{n\to\infty} a_n = -\infty$, if and only if $\liminf a_n = \limsup a_n$. Therefore, if (X_n) is a random sequence of real randon variables, then

$$\lim_{n\to\infty} a_n \text{ exists in } [-\infty, \infty] \text{ if and only if } \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n.$$

Moreover, by definition, there are two sub-sequences n_k and m_k such that

$$\lim_{k\to\infty} a_{n_k} = \liminf_{n\to\infty} a_n \text{ and } \lim_{l\to\infty} a_{m_l} = \limsup_{n\to\infty} a_n$$

where we can choose two subsequences such that

$$n_0 < m_0 < n_1 < m_1 < \cdots < n_k < m_k < \cdots$$

In the case that $\liminf_{n\to\infty} a_n < \limsup_{n\to\infty} a_n$, then we can choose $a < b$ such that

$$\liminf_{n\to\infty} a_n < a < b < \limsup_{n\to\infty} a_n$$

(and we can demand that $a < b$ to be rational numbers). Then, by looking the sequence (a_n) along $a_{n_0}, a_{m_0}, \cdots, a_{n_k}, a_{m_k}, \cdots$, we can see that the sequence (a_n) must cross from low level a to upper level b infinitely many times. That is, the number of up-crossing (a, b) by (a_n) is infinite. Hencelim$_{n\to\infty} a_n$ exists in $[-\infty, \infty]$ if and only if the *up-crossing number* by (a_n) through any (a, b) (for every pair $a < b$ of rational numbers) is finte.

Apply this to a sequence (X_n) of real random variables,

$$\left\{\lim_{n\to\infty} X_n \text{ exists in } [-\infty, \infty]\right\} = \{\text{the up-corssing number of } (X_n) \text{ through } (a, b) < \infty \text{ for any rationals}$$

Let $X = (X_n)_{n\geq 0}$ be a sequence of real valued random variables, and $a < b$ be two numbers. An *up-crossing* is the event that the sequence X_n is below a at some n and then $X_m \geq b$ for some $m > n$, and similarly we may define a down-crossing. Let us concentrate on up-crossing events.

Define

$$\begin{aligned}
T_0 &= \inf\{n \geq 0 : X_n \leq a\}, \\
T_1 &= \inf\{n > T_0 : X_n \geq b\}, \\
&\cdots\cdots \\
T_{2j} &= \inf\{n > T_{2j-1} : X_n \leq a\}, \\
T_{2j+1} &= \inf\{n > T_{2j} : X_n \geq b\}, \\
&\cdots
\end{aligned}$$

T_0 is the first time that the sequence X goes to the level below a, and T_1 is the first time X goes back to the level b after reaching the level below a and so on. All T_k are random times but can take value ∞, and $\{T_k\}$ is increasing. Moreover

$$\begin{aligned}
X_{T_{2j}} &\leq a \quad \text{on} \quad \{T_{2j} < \infty\}, \\
X_{T_{2j+1}} &\geq b \quad \text{on} \quad \{T_{2j+1} < \infty\}.
\end{aligned}$$

If $T_{2j-1}(\omega) < \infty$ for some $j \in \mathbb{N}$, then the sequence

$$X_0(\omega), \cdots, X_{T_{2j-1}}(\omega)$$

up-crosses the interval $[a, b]$ exactly j times.

Let $U_a^b(X; n)$ denote the number of up-crossings of the interval $[a, b]$ by $\{X_0, \cdots, X_n\}$. Then

$$\left\{U_a^b(X; n) = j\right\} = \{T_{2j-1} \leq n < T_{2j+1}\} \tag{11.1}$$

and

$$\left\{U_a^b(X; n) \geq j\right\} = \{T_{2j-1} \leq n\} \tag{11.2}$$

for $j = 0, 1, \cdots$.

If $X = (X_n)_{n\geq 0}$ is adapted with respect to a filtration $\{\mathcal{F}_n : n \geq 0\}$, then T_k are stopping times. Hence $\left\{U_a^b(X; n) = j\right\} \in \mathcal{F}_n$ for any $n \in \mathbb{Z}_+$ and $j \in \bar{\mathbb{Z}}_+$.

Lemma 11.1 *For any $b > a$ and $n, k \in \mathbb{N}$ we have*

$$1_{\left\{U_a^b(X;n)\geq k\right\}} \leq -\frac{X_n - a}{b - a} 1_{\left\{T_{2(k-1)}\leq n<T_{2k-1}\right\}} + \frac{X_{T_{2k-1}\wedge n} - X_{T_{2(k-1)}\wedge n}}{b - a} \tag{11.3}$$

and

$$1_{\left\{U_a^b(X;n)\geq k\right\}} \leq \frac{X_n - a}{b - a} 1_{\left\{U_a^b(X;n)=k\right\}} + \frac{X_{T_{2k-1}\wedge n} - X_{T_{2k}\wedge n}}{b - a}. \tag{11.4}$$

Proof. [*The proof is not examinable*] For every $k = 1, 2, \cdots$, $T_{2(k-1)} < T_{2k-1} < T_{2k}$ on $\{T_{2k-1} < \infty\}$. Let us consider the increments of $X = (X_n)$ over $[T_{2(k-1)}, T_{2k-1}]$ and $[T_{2k-1}, T_{2k}]$ respectively, which must be greater than $b - a$ on $\{T_{2k-1} < \infty\}$ (resp. on $\{T_{2k} < \infty\}$).

It is elementary that

$$\begin{aligned}
X_{T_{2k-1}\wedge n} - X_{T_{2(k-1)}\wedge n} &= \left(X_{T_{2k-1}\wedge n} - X_{T_{2(k-1)}}\right) 1_{\left\{T_{2(k-1)}\le n\right\}} \\
&= \left(X_{T_{2k-1}} - X_{T_{2(k-1)}}\right) 1_{\left\{T_{2(k-1)}\le n\right\}} 1_{\{T_{2k-1}\le n\}} \\
&\quad + \left(X_n - X_{T_{2(k-1)}}\right) 1_{\left\{T_{2(k-1)}\le n\right\}} 1_{\{T_{2k-1}> n\}} \\
&= \left(X_{T_{2k-1}} - X_{T_{2(k-1)}}\right) 1_{\{T_{2k-1}\le n\}} \\
&\quad + \left(X_n - X_{T_{2(k-1)}}\right) 1_{\{T_{2(k-1)}\le n< T_{2k-1}\}}.
\end{aligned}$$

Since $X_{T_{2k-1}} - X_{T_{2(k-1)}} \ge b-a$ on $\{T_{2k-1}<\infty\}$, $X_{T_{2(k-1)}} \le a$ on $\left\{T_{2(k-1)}<\infty\right\}$, and since $\{T_{2k-1}\le n\} = \left\{U_a^b(X;n)\ge k\right\}$, we deduce frome the previous identity that

$$X_{T_{2k-1}\wedge n} - X_{T_{2(k-1)}\wedge n} \ge (b-a) 1_{\left\{U_a^b(X;n)\ge k\right\}} + (X_n - a)\, 1_{\left\{T_{2(k-1)}\le n< T_{2k-1}\right\}}$$

and (11.3) follows. Similarly, one may use the decomposition

$$\begin{aligned}
X_{T_{2k-1}\wedge n} - X_{T_{2k}\wedge n} &= \left(X_{T_{2k-1}} - X_{T_{2k}}\right) 1_{\{T_{2k}\le n\}} + \left(X_{T_{2k-1}} - X_n\right) 1_{\{T_{2k-1}\le n< T_{2k}\}} \\
&\ge (b-a)\, 1_{\{T_{2k}\le n\}} + (b - X_n)\, 1_{\{T_{2k-1}\le n< T_{2k}\}} \\
&= (b-a)\left(1_{\{T_{2k}\le n\}} + 1_{\{T_{2k-1}\le n< T_{2k}\}}\right) + (a - X_n)\, 1_{\{T_{2k-1}\le n< T_{2k}\}} \\
&= (b-a)\, 1_{\{T_{2k-1}\le n\}} + (a - X_n)\, 1_{\{T_{2k-1}\le n< T_{2k}\}}
\end{aligned}$$

where we have used the fact that $X_{T_{2k-1}} \ge b$ on $\{T_{2k-1}<\infty\}$, which yields that

$$1_{\{T_{2k-1}\le n\}} \le -\frac{a-X_n}{b-a} 1_{\{T_{2k-1}\le n< T_{2k}\}} + \frac{X_{T_{2k-1}\wedge n} - X_{T_{2k}\wedge n}}{b-a}.$$

■

Theorem 11.2 (Doob's up-crossing lemma) *1) If $X=(X_n)$ is a super-martingale, then for any $n\ge 1$, $k\ge 0$*

$$\mathbb{P}\left[U_a^b(X;n)\ge k\right] \le \mathbb{E}\left[\frac{(X_n-a)^-}{b-a} : U_a^b(X;n) = k\right]$$

and

$$\mathbb{E}\left[U_a^b(X;n)\right] \le \mathbb{E}\left[\frac{(X_n-a)^-}{b-a}\right].$$

[Note that $X_n - a$ is also a super-martingale for any constant a, so that $(X_n-a)^-$ is a sub-martingale.*]*

2) Similarly, if $X=(X_n)$ is a sub-martingale, then

$$\mathbb{P}\left[U_a^b(X;n)\ge k\right] \le \mathbb{E}\left[\frac{(X_n-a)^+}{b-a} : U_a^b(X;n) = k\right]$$

and

$$\mathbb{E}\left[U_a^b(X;n)\right] \le \mathbb{E}\left[\frac{(X_n-a)^+}{b-a}\right].$$

[For a sub-martingale, $(X_n-a)^+$ is again a sub-martingale for every constant a.]

Proof. [*The proof is not examinable*] 1) Suppose X is a super-martingale, according to Doob's optional stopping theorem

$$\mathbb{E}\left[X_{T_{2k-1}\wedge n} - X_{T_{2(k-1)}\wedge n}\right] \leq 0, \tag{11.5}$$

so that it follows from (11.3) that

$$\begin{aligned}\mathbb{P}\left\{U_a^b(X;n) \geq k\right\} &\leq -\mathbb{E}\left\{\frac{X_n - a}{b-a} 1_{\left\{U_a^b(X;n)=k\right\}}\right\} + \mathbb{E}\left(X_{T_{2k-1}\wedge n} - X_{T_{2(k-1)}\wedge n}\right) \\ &\leq -\mathbb{E}\left\{\frac{X_n - a}{b-a} 1_{\left\{U_a^b(X;n)=k\right\}}\right\}\end{aligned}$$

which proves the first inequality. Since $U_a^b(X,n)$ takes values in $\mathbb{Z}_+$ so that

$$\begin{aligned}\mathbb{E}U_a^b(X,n) &= \sum_{k=1}^{\infty} k\mathbb{P}\left\{U_a^b(X;n) = k\right\} = \sum_{k=1}^{\infty} \mathbb{P}\left\{U_a^b(X;n) \geq k\right\} \\ &= \sum_{k=1}^{\infty} \mathbb{E}\left\{\frac{(X_n - a)^-}{b-a} : U_a^b(X;n) = k\right\} \leq \mathbb{E}\left\{\frac{(X_n - a)^-}{b-a}\right\}.\end{aligned}$$

2) If X is a sub-martingale, then $\mathbb{E}\left(X_{T_{2k-1}\wedge n} - X_{T_{2k}\wedge n}\right) \leq 0$, so that, by (11.4) we obtain

$$\mathbb{P}\left\{U_a^b(X;n) \geq k\right\} \leq \mathbb{E}\left\{\frac{X_n - a}{b-a} 1_{\left\{U_a^b(X;n)=k\right\}}\right\}.$$

■

Theorem 11.3 *(*The martingale convergence theorem, *J. L. Doob) 1) Suppose $X = (X_n)$ is a super-martingale, such that* $\sup_n \mathbb{E}\left[|X_n|\right] < \infty$ *(i.e. (X_n) is bounded in $L^1(\Omega)$), then $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely and $X_\infty \in L^1(\Omega)$. If in addition $X = (X_n)$ is non-negative, then $\mathbb{E}\left[X_\infty|\mathcal{F}_n\right] \leq X_n$ for $n \geq 0$.*

2) If $X = (X_n)$ is uniformly integrable martingale*, then $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely and in $L^1(\Omega)$, and $X_n = \mathbb{E}\left[X_\infty|\mathcal{F}_n\right]$ for every n.*

Proof. [*The proof is not examinable*] For any pair of rationales $a, b \in \mathbb{Q}$ with $a < b$, $U_a^b(X) = \lim_{n\to\infty} U_a^b(X;n)$ is the total number of up-crossings of the interval (a,b) ever made by (X_n). By MCT and Doob's crossing lemma we have

$$\begin{aligned}\mathbb{E}\left[U_a^b(X)\right] &= \lim_{n\to\infty} \mathbb{E}\left[U_a^b(X;n)\right] \leq \sup_n \mathbb{E}\left[\frac{(X_n - a)^-}{b-a}\right] \\ &\leq \frac{|a|}{b-a} + \frac{1}{b-a}\sup_n \mathbb{E}\left[|X_n|\right] < \infty.\end{aligned}$$

That is, $U_a^b(X)$ is integrable, so is finite almost surely. Let

$$W_{(a,b)} = \left\{\text{liminf}_{n\to\infty} X_n < a,\ \text{limsup}_{n\to\infty} X_n > b\right\}$$

and $W = \cup_{(a,b)} W_{(a,b)}$, where the union runs through the countable set of rational pairs (a,b), $a < b$. It is easy to verify, by definition of limits, that $W_{(a,b)} \subset \left\{U_a^b(X) = \infty\right\}$, so that $\mathbb{P}\left[W_{(a,b)}\right] = 0$. Hence $\mathbb{P}(W) = 0$. However if $\omega \notin W$, then $\lim_{n\to\infty} X_n(\omega)$ exists, and we denote it by $X_\infty(\omega)$ and on W we set $X_\infty(\omega) = 0$. Then $X_n \to X_\infty$ on W^c. Thus

$X_n \to X_\infty$ almost surely. According to Fatou's lemma, $\mathbb{E}|X_\infty| \le \sup_n \mathbb{E}|X_n| < \infty$, so that $X_\infty \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

If in addition $\{X_n\}$ is non-negative, since $\mathbb{E}[X_m|\mathcal{F}_n] \le X_n$ for $m \ge n$, letting $m \to \infty$, Fatou's lemma then yields that $\mathbb{E}[X_\infty|\mathcal{F}_n] \le X_n$, the proof is thus complete.

If however $X = (X_n)$ is uniformly integrable martingale, then we also have $X_n \to X_\infty$ in L^1. Since for every $m > n$ we have $\mathbb{E}[X_m|\mathcal{F}_n] = X_n$, by letting $m \to \infty$ to obtain $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$. ■

Corollary 11.4 *(*Levy's "Upward" theorem*) Let $\xi \in L^1(\Omega)$ and $X_n = \mathbb{E}[\xi|\mathcal{F}_n]$. Then $X = (X_n)$ is a uniformly integrable martingale, and $\lim_{n\to\infty} X_n = \mathbb{E}[\xi|\mathcal{F}_\infty]$ almost everywhere, where $\mathcal{F}_\infty = \sigma\{\mathcal{F}_j : j \ge 0\}$.*

Proof. By considering ξ^+ and ξ^- instead, we may assume that ξ is non-negative without losing generality. Since $\xi \in L^1(\Omega)$, $X = (X_n)$ is a uniformly integrable martingale. Thus $\lim_{n\to\infty} X_n = X_\infty$ exists a.e. and X_∞ is non-negative. By definition X_∞ is $\mathcal{F}_\infty$-measurable. Let $\mathcal{C} = \cup_{j=0}^\infty \mathcal{F}_j$. Then, for every $A \in \mathcal{C}$ we have $\mathbb{E}(1_A X_\infty) = \mathbb{E}(1_A \xi)$, and $\mathcal{F}_\infty$ is the smallest σ-algebra containing $\mathcal{C}$. Consider the collection $\mathcal{G}$ of all subsets A in $\mathcal{F}$ such that $\mathbb{E}(1_A X_\infty) = \mathbb{E}(1_A \xi)$. Then, since $X_\infty$and ξ are non-negative and integrable, the monotone convergence theorem implies that $\mathcal{G}$ is a σ-algebra containing $\mathcal{C}$. Therefore $\mathcal{G} \supset \mathcal{F}_\infty$, thus we must have $X_\infty = \mathbb{E}[\xi|\mathcal{F}_\infty]$. ■

Corollary 11.5 *(*Kolmogorov's 0-1 law*) Let ξ_n $(n = 1, 2, \cdots)$ be a sequence of independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G}_n = \sigma\{\xi_j : j \ge n+1\}$ and $\mathcal{G}_\infty = \cap_{n=0}^\infty \mathcal{G}_n$. Any element in $\mathcal{G}_\infty$is called a tail event. If $A \in \mathcal{G}_\infty$, then $\mathbb{P}(A) = 0$ or 1. Thus any $\mathcal{G}_\infty$-measurable random variable is constant almost surely.*

Proof. Let $\mathcal{F}_n = \sigma\{\xi_j : j \le n\}$. Then, for every n, $\mathcal{F}_n$ and $\mathcal{G}_n$ are independent. Hence $\mathcal{F}_n$ and $\mathcal{G}_\infty$ are independent for any n. Let $\xi = 1_A$ where $A \in \mathcal{G}_\infty$, and let $X_n = \mathbb{E}[\xi|\mathcal{F}_n]$. Then $X = (X_n)$ is a uniformly integrable martingale, and $X_n \to \mathbb{E}[\xi|\mathcal{F}_\infty] = 1_A$. On the other hand, since ξ and $\mathcal{F}_n$ are independent, so that $X_n = \mathbb{E}[\xi|\mathcal{F}_n] = \mathbb{E}[\xi] = \mathbb{P}(A)$ almost everywhere. Therefore $\mathbb{P}(A) = 1_A$ almost surely, so that $\mathbb{P}(A) = 0$ or 1. ■

12 The strong law of large numbers

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, instead of a filtration, we are given a decreasing family of sub σ-algebras $(\mathcal{G}_n)_{n\ge 0}$, where $\mathcal{G}_{n+1} \subseteq \mathcal{G}_n$ for $n = 0, 1, 2, \cdots$, where the largest σ-algebra is the initial one $\mathcal{G}_0 \subset \mathcal{F}$. The final σ-algebra is $\mathcal{G}_\infty = \lim_{n\to\infty} \mathcal{G}_n = \bigcap_{j=0}^\infty \mathcal{G}_j$.

We may define martingales, sub-martingales and super-martingales with respect to the decreasing flow $(\mathcal{G}_n)$. Namely, a $(\mathcal{G}_n)$-adapted and integrable random sequence $X = (X_n)_{n\ge 0}$ is a martingale (resp. super-martingale, and resp. sub-martingale) if $\mathbb{E}[X_n|\mathcal{G}_{n+1}] = X_{n+1}$ (resp. $\mathbb{E}[X_n|\mathcal{G}_{n+1}] \le X_{n+1}$, and resp. $\mathbb{E}[X_n|\mathcal{G}_{n+1}] \ge X_{n+1}$).

If we set $\mathcal{F}_n = \mathcal{G}_{-n}$ where $n = \cdots, -2, -1, 0$ (with the natural order in $\mathbb{Z}_-$), then $(\mathcal{F}_n)$ (where $n = \cdots, -2, -1, 0$) is a filtration, i.e. an increasing flow of σ-algebras. Then $M_n = X_{-n}$ (where $n = \cdots, -2, -1, 0$) is martingale (resp. super-martingale, resp. sub-martingale) if $\mathbb{E}[M_n|\mathcal{F}_{n-1}] = M_{n-1}$ (resp. $\mathbb{E}[M_n|\mathcal{F}_{n-1}] \le M_{n-1}$, resp. $\mathbb{E}[M_n|\mathcal{F}_{n-1}] \ge M_{n-1}$) for $n = \cdots, -2, -1, 0$. The following technical lemma allows us apply the results we have established in the previous sections to martingales with respect to a decreasing flow of σalgebras, which follows directly from the definition.

Lemma 12.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space together with a decreasing family $(\mathcal{G}_n)_{n\geq 0}$ of sub σ-algebras of $\mathcal{F}$. Let $X = (X_n)_{n\geq 0}$, where $X_n \in L^1(\Omega, \mathcal{G}_n, \mathbb{P})$ for $n = 0, 1, 2, \cdots$. Then, X is martingale (resp. super-martingale, resp. sub-martingale), if and only if for every $N = 1, 2, \cdots$, the* time-reversed *random sequence $Y_n = X_{N-n}$ (where $n = 0, \cdots, N$) is a martingale (resp. super-martingale, resp. sub-martingale) up to time N (with terminal value X_0), with respect to the filtration $\mathcal{G}_{N-n}$.*

As a sample of applications of the previous lemma, we prove the following very useful convergence result.

Theorem 12.2 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space together with a decreasing family $(\mathcal{G}_n)_{n\geq 0}$ of sub σ-algebras of $\mathcal{F}$. If $X = (X_n)_{n\geq 0}$ is a super-martingale with respect to $(\mathcal{G}_n)$, then $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely. If in addition $\lim_{n\to\infty} \mathbb{E}[X_n] < \infty$ then $\{X_n : n \geq 0\}$ is uniformly integrable, and $X_n \to X_\infty$ in $L^1(\Omega)$.*

Proof. [*The proof is not examinable.*] For every $N = 1, 2, \cdots$, the time-reversed sequence $\{X_N, X_{N-1}, \cdots, X_0\}$ is a super-martingale (up to time N) with respect to $\mathcal{G}_{N-n}$, its up-crossing number through $[a, b]$ where $a < b$ is denoted by $U_a^b(X, -N)$. The label $-$ is used to indicate the reversed up-crossing, rather than $U_a^b(X, N)$ which is the up-crossing of $\{X_0, X_1, \cdots, X_N\}$, but they are equally useful in determining the convergence. Let $U_a^b(X) = \lim_{N\to\infty} U_a^b(X, -N)$ which represents the number of up-crossings for the time-reversed sequence $\{\cdots, X_N, X_{N-1}, \cdots, X_0\}$. According to Doob's up-crossing lemma, for every N,

$$\mathbb{E}\left[U_a^b(X; -N)\right] \leq \mathbb{E}\left[\frac{(X_0 - a)^-}{b - a}\right].$$

The right-hand side is independent of N, so by applying the Monotone Convergence Theorem, we have

$$\mathbb{E}\left[U_a^b(X)\right] \leq \mathbb{E}\left[\frac{(X_0 - a)^-}{b - a}\right].$$

Therefore $U_a^b(X)$ is integrable, so that $U_a^b(X) < \infty$ almost everywhere. A similar argument as the proof of the Martingale Convergence Theorem may apply to conclude that $X_\infty = \lim_{n\to\infty} X_n$ exists almost everywhere, and X_∞ is $\bigcap_{j=1}^\infty \mathcal{G}_j$-measurable.

Since $n \to \mathbb{E}[X_n]$ is increasing (note that not decreasing, as it is a time-reversed super-martingale), so that $\sup_n \mathbb{E}[X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$. Suppose that $\lim_{n\to\infty} \mathbb{E}[X_n] < \infty$. Then $\sup_n \mathbb{E}[X_n] < \infty$. Since X_0 is integrable, $\xi_n = \mathbb{E}[X_0|\mathcal{G}_n]$ is uniformly integrable (time-reversed) martingale, and $Q_n = X_n - \xi_n$ is (time-reversed) super-martingale. Since

$$Q_n = \mathbb{E}[Q_n|\mathcal{G}_n] = \mathbb{E}[X_n - X_0|\mathcal{G}_n] \geq 0$$

which implies that Q_n is non-negative, and $X_n = Q_n + \xi_n$. Therefore, to show that X is uniformly integrable, we only need to show that $Q = (Q_n)$ is uniformly integrable. Thus, without losing generality, we may assume that $X = (X_n)$ is a non-negative (time-reversed) super-martingale, and $\sup_n \mathbb{E}[X_n] = \lim_{n\to\infty} \mathbb{E}[X_n] < \infty$.

According to the time-reversed super-martingale property, for any $n > m \geq 0$ and $L > 0$, since $\{X_n \leq L\} \in \mathcal{G}_n$ we have

$$\begin{aligned}\mathbb{E}[X_n : X_n > L] &= \mathbb{E}[X_n] - \mathbb{E}[X_n : X_n \leq L] \leq \mathbb{E}X_n - \mathbb{E}[X_m : X_n \leq L] \\ &\leq \mathbb{E}[X_n] - \mathbb{E}[X_m] + \mathbb{E}[X_m : X_n > L].\end{aligned}$$

Since $\lim_{n\uparrow\infty}\mathbb{E}[X_n]$ exists, so for every $\varepsilon>0$, there is N_1 such that $0\le\mathbb{E}[X_n]-\mathbb{E}[X_m]<\frac{\varepsilon}{2}$ for all $n,m\ge N_1$. Since the finite family of integrable random variables$\{X_0,\cdots,X_{N_1}\}$ is uniformly integrable, so there is $\delta>0$ such that $\mathbb{E}[X_m:A]<\varepsilon/2$ for any A with $\mathbb{P}(A)<\delta$, for all $m\le N_1$. On the other hand, using Markov inequality $\mathbb{P}[X_n>L]\le\frac{\sup_n\mathbb{E}X_n}{L}$. Choose $L_0=\frac{\sup_n\mathbb{E}X_n}{\delta}$. Then $\mathbb{P}[X_n>L]<\delta$ for all $L\ge L_0$ and for all n. Therefore $\mathbb{E}[X_m:X_n>L]<\frac{\varepsilon}{2}$ for all $m\le N_1$ and $L\ge L_1$, and

$$\mathbb{E}[X_n:X_n>L]\le\mathbb{E}X_n-\mathbb{E}X_{N_1}+\mathbb{E}[X_{N_1}:X_n>L]<\varepsilon$$

for all $L\ge L_0$ and $n\ge N_1$. Putting all these estimates together we deduce that

$$\mathbb{E}[X_n:X_n>L]<\varepsilon$$

for all n and $L\ge L_0$, which proves that (X_n) is uniformly integrable. Hence $X_n\to X_\infty$ in $L^1(\Omega)$ as well. ■

Corollary 12.3 *(*Levy's "Downward" theorem*) Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space together with a decreasing family $(\mathcal{G}_n)_{n\ge0}$ of sub σalgebras of $\mathcal{F}$. Let $\xi\in L^1(\Omega)$ and $X=(X_n)_{n\ge0}$, where $X_n=\mathbb{E}[\xi|\mathcal{G}_n]$ for $n=0,1,2,\cdots$. Then, $X_n\to X_\infty=\mathbb{E}\left[\xi|\bigcap_{j=0}^\infty\mathcal{G}_j\right]$.*

This follows from the previous theorem, as $X=(X_n)$ is a uniformly integrable (time-reversed) martingale.

We are now in a position to prove the *strong law of large numbers* for i.i.d. sequences. We collect a few elementary facts about independent sequences in the following examples.

Example 12.4 *If $\{\xi_k\}_{k\ge1}$ is a sequence of independent random variables on $(\Omega,\mathcal{F},\mathbb{P})$, then $\mathcal{G}_\infty=\bigcap_{n=1}^\infty\sigma\{\xi_j:j>n\}$ is the tail σ-algebra of the independent random sequence $\{\xi_k\}_{k\ge1}$. Any element in $\mathcal{G}_\infty$ is called a tail event. Suppose $A\in\mathcal{G}_\infty$ we prove that $\mathbb{P}(A)=0$ or 1, which is called Kolmogorov's 0-1 law. It follows that $\mathcal{G}_\infty$-measurable function Z must be constant almost everywhere, so that $Z=\mathbb{E}[Z]$ a.e.*

Theorem 12.5 *(A. Kolmogorov,* The Strong Law of Large Numbers*) Let $\{\xi_k\}_{k\ge1}$ be a sequence of independent integrable random variables on $(\Omega,\mathcal{F},\mathbb{P})$ with the same distribution. Then $\frac{1}{n}\sum_{k=1}^n\xi_k\to\mathbb{E}[\xi_1]$ almost everywhere.*

Proof. Let $(\mathcal{G}_n)_{n\ge0}$ be the decreasing family of σ-algebras generated by the sequence (X_n), where $X_n=\sum_{k=1}^n\xi_k$. That is

$$\mathcal{G}_n=\sigma\{X_m:m\ge n\}=\sigma\{X_n,\xi_j\ge n+1\}.$$

Let $M_n=\frac{1}{n}X_n$. We show that $M=(M_n)$ is a (time-reversed) martingale with respect to $(\mathcal{G}_n)$. First we observe that, since $\xi_1,\cdots,\xi_n$ are independent, with the same distribution, we thus have

$$\mathbb{E}[\xi_i|\xi_1+\cdots+\xi_n]=\mathbb{E}[\xi_1|\xi_1+\cdots+\xi_n].$$

In fact, for every i, $X_n-\xi_i$ and ξ_i are independent, their distributions are independent of i. Thus (ξ_i,X_n) has the same distribution as that of (ξ_1,X_n). Since in general the conditional expectation $\mathbb{E}[\xi|\zeta]$ of ξ given ζ is a function of ζ depending only on their joint distribution, thus $\mathbb{E}[\xi_i|X_n]$ is independent of $i=1,\cdots,n$. Hence $\mathbb{E}[\xi_i|X_n]=\mathbb{E}[\xi_1|X_n]$ for all $i\le n$ (here the assumption that ξ_j has the same distribution is essential). On the other hand, since ξ_i are independent, so that

$$\mathbb{E}\left[\xi_1|X_n\right] = \mathbb{E}\left[\xi_1|X_n, \xi_j : j \geq n+1\right] = \mathbb{E}\left[\xi_1|\mathcal{G}_n\right]$$

and therefore

$$\begin{aligned} M_n &= \frac{1}{n}\mathbb{E}\left[X_n|X_n\right] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\xi_j|X_n\right] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\xi_1|X_n\right] \\ &= \mathbb{E}\left[\xi_1|X_n\right] = \mathbb{E}\left[\xi_1|\mathcal{G}_n\right]. \end{aligned}$$

Therefore, M is a time-reversed martingale, and $M = (M_n)$ is uniformly integrable, thus, according to Corollary 12.3, $M_n \to M_\infty = \mathbb{E}\left[\xi_1 | \bigcap_{j=1}^{\infty} \mathcal{G}_j\right]$ almost everywhere. Since

$$M_\infty = \lim_{n\to\infty} \frac{\xi_{m+1} + \cdots + \xi_n}{n}$$

so M_∞ is $\sigma\{\xi_j : j \geq m+1\}$-measurable for any m, thus M_∞ is measurable with respect to the tail σ-algebra, according to Kolmogorov's 0-1 law, M_∞ must be a constant (almost surely), so that $M_\infty = \mathbb{E}\left[M_\infty\right] = \mathbb{E}\left[\xi_1\right]$. ■

We should point out that the strong law of large numbers for i.i.d. sequences is still a special case of *Birkhoff's ergodic theorem* for strictly stationary sequences. Birkhoff's ergodic theorem however rquies a different approach and thus provides a different proof for the strong law of large numbers.

13 Doob's decomposition for super-martingales

We introduce an important tool for the study of martingales, Doob's decomposition for square-integrable super-martingales. The extension to the continuous time case is much more difficult, called Doob-Meyer's decomposition, which is the key in order to define stochastic integrals with respect to martingales.

Suppose $X = (X_n)$ is a super-martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$. Thus $\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \leq X_n$, so roughly speaking on average, $n \to X_n$ is decreasing. Doob's decomposition is an explicit statement about this fact. The idea is to seek for a martingale M_n and an increasing sequence A_n such that $X_n = M_n - A_n$. Let $A_0 = 0$ and $M_0 = X_0$. Since

$$X_{n+1} - X_n = M_{n+1} - M_n - (A_{n+1} - A_n)$$

and conditional on $\mathcal{F}_n$, to obtain

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] - X_n = -\mathbb{E}\left[A_{n+1} - A_n|\mathcal{F}_n\right].$$

If we impose the condition that A_n is $\mathcal{F}_{n-1}$-measurable for every $n \geq 1$ (such a sequence is called predictable). Then

$$A_{n+1} = A_n + X_n - \mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] = \sum_{j=0}^{n}\left(X_j - \mathbb{E}\left[X_{j+1}|\mathcal{F}_j\right]\right) = \sum_{j=0}^{n}\mathbb{E}\left[X_j - X_{j+1}|\mathcal{F}_j\right]$$

for $n \geq 0$. We note that, since X_n is a super-martingale, thus (A_n) is increasing and predictable, with $A_0 = 0$, and it is direct to verify that $M_n = X_n + A_n$ is a martingale.

Theorem 13.1 *(*Doob's decomposition for super-martingales*) Let $X = (X_n)$ be a super-martingale over a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$. Then there is a unique increasing predictable random sequence (A_n) with $A_0 = 0$, such that $M_n = X_n + A_n$ is a martingale. More precisely*

$$A_n = \sum_{j=0}^{n-1} (X_j - \mathbb{E}[X_{j+1}|\mathcal{F}_j])$$

and

$$M_n = X_n + \sum_{j=0}^{n-1} (X_j - \mathbb{E}[X_{j+1}|\mathcal{F}_j])$$

for $n = 1, 2, \cdots$, and $A_0 = 0$, $M_0 = X_0$. The decomposition $X_n = M_n - A_n$ is called Doob's decomposition for the super-martingale $X = (X_n)$.

Let us apply Doob's decomposition to square integrable martingales.

Suppose that $M = (M_n)$ is a martingale such that $\mathbb{E}[M_n^2] < \infty$ for each n. Then M_n^2 is a sub-martingale, so $-M_n^2$ is a super-martingale. Therefore there is a unique increasing predictable random sequence A_n such that $-M_n^2 + A_n$ is again a martingale, where

$$\begin{aligned} A_n &= \sum_{j=0}^{n-1} \left(-M_j^2 + \mathbb{E}\left[M_{j+1}^2|\mathcal{F}_j\right]\right) = \sum_{j=0}^{n-1} \mathbb{E}\left[M_{j+1}^2 - M_j^2|\mathcal{F}_j\right] \\ &= \sum_{j=0}^{n-1} \mathbb{E}\left[(M_{j+1} - M_j)^2 |\mathcal{F}_j\right] \end{aligned}$$

which is called the *bracket process* associated with M. The bracket process will play an important role in the study of martingales, so let us give a definition.

Definition 13.2 *1) Let $M = (M_n)$ be a* martingale *with $M_n \in L^2(\Omega)$ for every n. Then the bracket process $\langle M \rangle$ associated with M is the unique predictable, increasing sequence with $\langle M \rangle_0 = 0$ such that $M_n^2 - \langle M \rangle_n$ is a martingale. Explicitly $\langle M \rangle$ is given by*

$$\langle M \rangle_n = \sum_{j=0}^{n-1} \mathbb{E}\left[(M_{j+1} - M_j)^2|\mathcal{F}_j\right]$$

for $n \geq 1$, $\langle M \rangle_0 = 0$. That is, $\langle M \rangle$ is the conditional quadratic variational process associated with M. In particular, for any bounded *stopping time T, $\mathbb{E}[M_T^2 - M_0^2] = \mathbb{E}[\langle M \rangle_T]$, and*

$$\sup_n \mathbb{E}\left[M_n^2 - M_0^2\right] = \sup_n \mathbb{E}[\langle M \rangle_n] = \lim_{n\to\infty} \mathbb{E}[\langle M \rangle_n] = \mathbb{E}[\langle M \rangle_\infty]$$

where $\langle M \rangle_\infty = \lim_{n\to\infty} \langle M \rangle_n$ (which may be infinity), and the last equality follows from MCT applying to $\langle M \rangle_n \uparrow \langle M \rangle_\infty$.

2)The quadratic variational process $[M]_n$ associated with M is defined by $[M]_0 = 0$ and

$$[M]_n = \sum_{j=0}^{n-1} (M_{j+1} - M_j)^2$$

for $n \geq 1$.

3) A martingale $M = (M_n)$ is called a squared integrable martingale if $\sup_n \mathbb{E}[M_n^2] < \infty$ *(i.e. $\{M_n : n \geq 0\}$ is bounded in $L^2(\Omega)$.)*

By a direct computation we have

Lemma 13.3 *1) Let $M = (M_n)$ be a martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$ such that $M_n \in L^2(\Omega)$. Then $[M]_n - \langle M \rangle_n$ is a martingale, while $\langle M \rangle$ is predictable, and $[M]$ is an adapted increasing sequence.*

2) Suppose that M and N are two martingales such that $M_n, N_n \in L^2(\Omega)$, then $M_n N_n - \langle M, N \rangle_n$ is a martingale, where the mutual bracket

$$\begin{aligned}\langle M, N \rangle_n &= \frac{1}{4} \left(\langle M + N \rangle - \langle M - N \rangle \right) \\ &= \sum_{j=0}^{n-1} \mathbb{E}\left[(M_{j+1} - M_j)(N_{j+1} - N_j) | \mathcal{F}_j \right].\end{aligned}$$

for $n \geq 1$, $\langle M, N \rangle_0 = 0$, which is a predictable process.

Suppose that $M = (M_n)$ is a martingale, and $H = (H_n)$ is a predictable process, the martingale transform $H.M$ (which corresponds the Ito integral of H against M, so called discrete stochastic integral of H against M) is defined by $(H.M)_0 = 0$ and

$$(H.M)_n = \sum_{j=1}^{n} H_j (M_j - M_{j-1})$$

for $n \geq 1$. Then

$$\begin{aligned}\langle H.M \rangle &= \sum_{j=0}^{n-1} \mathbb{E}\left[\left(H_{j+1}(M_{j+1} - M_j) \right)^2 | \mathcal{F}_j \right] = \sum_{j=0}^{n-1} H_{j+1}^2 \mathbb{E}\left[(M_{j+1} - M_j)^2 | \mathcal{F}_j \right] \\ &= \sum_{j=1}^{n} H_j^2 \left(\langle M \rangle_j - \langle M \rangle_{j-1} \right)\end{aligned}$$

which is $H^2 . \langle M \rangle$, the stochastic integral of H^2 with respect to the increasing process $\langle M \rangle$.

The bracket processes play a fundamental role in Ito's stochastic integration theory. Here we only give an elementary application of the bracket process.

Theorem 13.4 *Let $M = (M_n)$ be a martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$ such that $M_n \in L^2(\Omega)$. Then $M_\infty = \lim_{n\to\infty} M_n$ exists on $\{\langle M \rangle_\infty < \infty\}$.*

Proof. [*The proof is not examinable*]. Since

$$\{\langle M \rangle_\infty < \infty\} = \bigcup_{l=1}^{\infty} \{\langle M \rangle_\infty \leq l\}$$

so we only need to show that $M_\infty = \lim_{n\to\infty} M_n$ exists on each $\{\langle M \rangle_\infty \leq l\}$. Let $l > 0$, and $T = \inf \left\{ k \geq 0 : \langle M \rangle_{k+1} > l \right\}$. Then T is a stopping time as $\langle M \rangle$ is predictable, so that by Theorem 10.4, $M_{T\wedge n}^2 - \langle M \rangle_{T\wedge n}$ is martingale, thus $\mathbb{E}\left[M_{T\wedge n}^2 \right] = \mathbb{E}\left[\langle M \rangle_{T\wedge n} \right] \leq l$ for all n. Therefore $\{M_{T\wedge n}\}$ is a uniformly integrable martingale, so that $\lim_{n\to\infty} M_{T\wedge n}$ exists. In particular, $\lim_{n\to\infty} M_n$ exists on $\{T = \infty\}$, so does on $\{\langle M \rangle_\infty \leq l\}$ for any $l > 0$. ■

Recall that if $X = (X_n)$ is a SMartingale which is uniformly integrable, then $X_n \to X_\infty$ almost surely and in L^1. For the L^p-bounded martingale, we have the following

Theorem 13.5 *Suppose $X = (X_n)_{n\geq 1}$ is a martingale which is bounded in L^p-space for some $p>1$, that is, $\sup_n \mathbb{E}[|X_n|^p] < \infty$, then $(X_n)_{n\geq 0}$ is uniformly integrable, and $X_n \to X_\infty$ almost surely, and in L^p-space. Moreover*

$$\mathbb{E}[|X_\infty|^p] = \sup_n \mathbb{E}[|X_n|^p].$$

Proof. [*The proof is not examinable.*] It is known that $\sup_n \mathbb{E}[|X_n|^p] < \infty$ for some $p>1$ implies that (X_n) is uniformly integrable, so that $X_n \to X_\infty$ almost surely and in L^1. Let $g = \lim_{n\to\infty} \sup_{k\leq n} |X_k|$. Applying Doob's L^p-inequality to the sub-martingale $|X_n|$ we have

$$\mathbb{E}\left[\left|\sup_{k\leq n} |X_k|\right|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p] \leq \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}[|X_n|^p].$$

Thus, by MCT we conclude that

$$\mathbb{E}[|g|^p] \leq \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}[|X_n|^p] < \infty$$

that is $|g|^p$ is integrable. Now $|X_n - X_\infty|^p \to 0$ almost surely, and $|X_n - X_\infty|^p \leq 2^p|g|^p$ for all n, so by Lebesgue's dominated convergence theorem, we have

$$\mathbb{E}[|X_n - X_\infty|^p] \to 0$$

as $n \to \infty$. Since $|X_n|^p$ is a sub-martingale, so that $n \to \mathbb{E}[|X_n|^p]$ is increasing, and therefore

$$\mathbb{E}[|X_\infty|^p] = \lim_{n\to\infty} \mathbb{E}[|X_n|^p] = \sup_n \mathbb{E}[|X_n|^p].$$

■