B8.1 Probability, Measure and Martingales
by Zhongmin Qian

Contents

1 Measures and integration

In this part, we review the theory of integration that you have learned in A4 paper with general measures, so these notes should be read along with the lecture notes "A4 Integration (2017 HT)" posted at

https://courses.maths.ox.ac.uk/node/view_material/37658

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 $\infty+\infty=\infty$.

1.1 Basic definitions and theorems

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

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

1.2) $\mu(A)\le\mu(B)$ if $A,B\in\mathscr{R}$ and $A\subseteq B$, 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 $A_i\in\mathscr{R}$ $(i=1,2,\cdots)$ which are disjoint, such that $\bigcup_{i=1}^{\infty}A_i\in\mathscr{R}$.

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

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

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

3. Finite measures and σ-finite measures. A measure μ on $\mathscr{R}$ is finite if $\mu(E)<\infty$ for every $E\in\mathscr{R}$; μ is called σ-finite on $\mathscr{R}$ if there is a sequence of subsets $E_i\in\mathscr{R}$ 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* on $\mathscr{R}$.

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

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

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

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

5. Measure spaces and probability spaces. If $(\Omega,\mathscr{F})$ is a measurable space and μ is a measure on $\mathscr{F}$, then $(\Omega,\mathscr{F},\mu)$ is called a *measure space*. If $\mu(\Omega)=1$ then $(\Omega,\mathscr{F},\mu)$ is called a probability space. In this case Ω is called a sample space (of fundamental events), an element A in the σ-algebra $\mathscr{F}$ 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 $\mathscr{F}$, or simply called $\mathscr{F}$-measurable, if

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

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

7. Structure of measurable functions. A simple (measurable) function φ on $(\Omega,\mathscr{F})$ 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 $E_k\in\mathscr{F}$. A function $f:\Omega\to[0,\infty]$ is $\mathscr{F}$-measurable, if and only if there is an increasing sequence of non-negative, $\mathscr{F}$-measurable simple functions φ_n such that $\varphi_n\uparrow f$ everywhere on Ω.

8. Definition of Lebesgue's integrals. Let $(\Omega,\mathscr{F},\mu)$ be a measure space. Lebesgue's 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 a simple function, namely, if $\varphi=\sum_{j=1}^{m}c_j 1_{E_j}$ is a non-negative ($\mathscr{F}$-measurable) *simple function* on Ω, where $c_i\geq 0$ and $E_i\in\mathscr{F}$ 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 $\mathscr{F}$-measurable function, then

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

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 $f\in L^1(\Omega,\mathscr{F},\mu)$, $f\in L^1(\Omega,\mu)$, $L^1(\Omega)$ or simply by $f\in L^1$ if the measure space in question is clear. If $f:\Omega\to[-\infty,\infty]$ is $\mathscr{F}$-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 $f\in L^1(\Omega,\mathscr{F},\mu)$ 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, $\mathscr{F}$-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

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

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

The convergence theorems are applicable to a measure space $(\Omega,\mathscr{F},\mu)$, 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 *(*Reversed 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 $f \in L^1(\Omega, \mathscr{F}, \mu)$ 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 $A \in \mathscr{F}$, then $(A, A \cap \mathscr{F}, \mu)$ is a measure space too. In this case $\int_A f d\mu$ coincides 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^pspace for $p \in [1, \infty]$ can be defined over a measure space. When dealing with L^p-spaces, we identify an $\mathscr{F}$-measurable function f on $(\Omega, \mathscr{F}, \mu)$ with its equivalent class of all $\mathscr{F}$-measurable functions which are equal to f almost surely on Ω. Then $L^p(\Omega, \mathscr{F}, \mu)$ is the vector space of all $\mathscr{F}$-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

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

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

17. Convergence in L^p-spaces. $L^p(\Omega, \mathscr{F}, \mu)$ are Banach spaces. $f \to ||f||_p$ is a norm on $L^p(\Omega, \mathscr{F}, \mu)$, and $L^p(\Omega, \mathscr{F}, \mu)$ 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 $L^p(\Omega, \mathscr{F}, \mu)$ if f_n and f belong to $L^p(\Omega, \mathscr{F}, \mu)$ 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 $L^1(\Omega, \mathscr{F}, \mu)$ 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,

$$||f_n| - |f|| \le |f_n - f|$$

so that

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

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: $A_n = \{f_n f \ge 0\}$, $B_n = \{f_n f < 0\}$. Then

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

and, by the triangle inequality,

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

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 rewritten 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

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

so that

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

for all n. Since $|f|$ is integrable, and $(|f_n| - |f|)^- \to 0$ almost surely, by the DCT

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

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 $n \to \infty$. $|f| 1_{B_n}$ is controlled by the integrable function $|f|$, so by DCT, $\int_{B_n} |f| d\mu = \int_\Omega |f| 1_{B_n} d\mu \to 0$. Therefore, by Sandwich lemma, it follows from (1.1) that $\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0$. ■

1.2 Examples

Recall that we use the following notations. If Ω is a space, then $\mathscr{P}(\Omega)$ is the collection of *all* subsets of Ω, called the power set of Ω (a notion introduced in algebra). $\mathscr{P}(\Omega)$ is a σ-algebra, which is the largest σ-algebra on Ω. That is, if $\mathscr{F}$ is a σ-algebra on Ω, then $\mathscr{F} \subset \mathscr{P}(\Omega)$. A fact like this is so *obvious*, which requires no proof. Now suppose $\mathscr{C}$ is a (non-empty) collection of *some* subsets of Ω, we use $\sigma\{\mathscr{C}\}$ to denote the smallest (according to the semi order $\subset$ among sets) σ-algebra on Ω which contains $\mathscr{C}$. Firstly if $\mathscr{C}$ is already a σ-algebra, then of course $\sigma\{\mathscr{C}\} = \mathscr{C}$. Otherwise, we need to show the uniqueness and existence of $\sigma\{\mathscr{C}\}$, which can be sone in one go – just write down what is $\sigma\{\mathscr{C}\}$. In fact

$$\sigma\{\mathscr{C}\} = \bigcap \{\mathscr{F} : \mathscr{F} \text{ are } \sigma\text{-algebras on } \Omega \text{ and } \mathscr{C} \subset \mathscr{F}\}.$$

where the right-hand side means the intersection of all σ-algebras which contain $\mathscr{C}$. To prove this, we need to do three things. 1) Show that the intersection we are talking about on the right-hand side is no-empty, which is easy, because $\mathscr{P}(\Omega)$ is such element $\mathscr{F}$. This means $\sigma\{\mathscr{C}\}$ is well-defined as a collection of some subsets of Ω. Since each $\mathscr{F} \supset \mathscr{C}$ in the intersection, so that $\sigma\{\mathscr{C}\} \supset \mathscr{C}$. 2) Verify that $\sigma\{\mathscr{C}\}$ is a σ-algebra. In fact we can verify the following general fact – if $\mathscr{F}_\alpha$ (where $\alpha \in \Lambda$, Λ is an index set) is a family of σ-algebras on (the same sample space) Ω, then $\bigcap_\alpha \mathscr{F}_\alpha$ is σ-algebra too. This fact follows from a general statement that "if each element $\mathscr{F}_\alpha$ possesses a property P, then so does their intersection $\bigcap_\alpha \mathscr{F}_\alpha$". 3) Finally we need to show that $\sigma\{\mathscr{C}\}$ given in the display above is the smallest one: if $\mathscr{F} \supset \mathscr{C}$ is a σ-algebra, then $\sigma\{\mathscr{C}\} \subset \mathscr{F}$. This is due to the fact the intersection of a family of sets is a subset of each individual set in the intersection.

Consider measurable functions (random variables). Let $(\Omega, \mathscr{F})$ be a measurable space, so $\mathscr{F}$ is a σ-algebra on Ω. Suppose we have another measurable space (S, Σ), that is, S is a state space and Σ is a σ-algebra on S [Typical example is the following: $S = \mathbb{R}^n$ is the Euclidean space and $\Sigma = \mathscr{B}(\mathbb{R}^n)$ is the Borel σ-algebra — this is the most interesting case for this course]. A mapping $X : \Omega \to S$ (such a mapping is also called a function on Ω taking values in S, or called an S-valued function on Ω) is called a measurable mapping (or called a S-valued measurable function) from $(\Omega, \mathscr{F})$ to (S, Σ), if $X^{-1}(G) \in \mathscr{F}$ for every $G \in \Sigma$. For simplicity, if the σ-algebra Σ is clear in the discussion, we simply say X is measurable w.r.t. $\mathscr{F}$, or just say X is $\mathscr{F}$-measurable. This is an extension of the concept of measurable (real) functions. Here we use the following notations

$$X^{-1}(G) \equiv \{w \in \Omega : X(w) \in G\} \equiv \{X \in G\}$$

the pre-image of G under the mapping X, where $\equiv$ means that "denoted also by".

If $X : \Omega \to S$ is $\mathscr{F}$-measurable (or more precisely X is a measurable mapping between measurable spaces $(\Omega, \mathscr{F})$ and (S, Σ)), then we use $\sigma\{X\}$ to denote *the smallest* σ-algebra on Ω with respect to which X is measurable, i.e. X is a measurable mapping between $(\Omega, \sigma\{X\})$ and (Ω, Σ). By definition, $\mathscr{F} \supset \sigma\{X\}$. In fact

$$\sigma\{X\} = \left\{X^{-1}(G) : G \in \Sigma\right\}$$

that is, $\sigma\{X\}$ is the collection of all pre-images of the sets in Σ. Let us denote the collection on the right-hand side by $X^{-1}(\Sigma)$, that is, we define

$$X^{-1}(\Sigma) \equiv \left\{X^{-1}(G) : G \in \Sigma\right\}$$

We claim that $\sigma\{X\} = X^{-1}(\Sigma)$. To show this we need to 1) Verify that the collection on the right-hand side *is* a σ-algebra – this is a routine exercise, you check that the conditions of σ-algebra for $X^{-1}(\Sigma)$ are satisfied. 2) In order to ensure that X is $\mathscr{G}$-measurable, $X^{-1}(G) \in \mathscr{G}$ for every $G \in \Sigma$, which yields that $\mathscr{G} \supset X^{-1}(\Sigma)$. 3) Finally, by just the definition, X is $X^{-1}(\Sigma)$-measurable.

How about if you have several random variables? While the situation is similar. Thus, if $X_i : \Omega \to \Sigma$ are measurable mappings from $(\Omega, \mathscr{F})$ to (S, Σ), where $i = 1, 2, \cdots, n$. Then we use $\sigma\{X_1, \cdots, X_n\}$ to denote the smallest σ-algebra on Ω with respect to which all X_1, $\cdots$, X_n are measurable. Let us denote $\mathscr{G} \equiv \sigma\{X_1, \cdots, X_n\}$ for simplicity. Since $X_1, \cdots, X_n$ are $\mathscr{G}$-measurable, so that $X_i^{-1}(\Sigma)$ $(i = 1, \cdots, n)$ are sub-σ-algebras of $\mathscr{G}$, thus $\mathscr{G}$ must contain the union $\bigcup_{i=1}^n X_i^{-1}(\Sigma)$, which is unfortunately in general not a σ-algebra, so that

$$\mathscr{G} = \sigma\left\{\bigcup_{i=1}^n X_i^{-1}(\Sigma)\right\}$$

while

$$\begin{aligned}\bigcup_{i=1}^n X_i^{-1}(\Sigma) &= \bigcup_{i=1}^n \left\{X_i^{-1}(G) : G \in \Sigma\right\} \\ &= \left\{\bigcap_{i=1}^n X_i^{-1}(G_i) : G_i \in \Sigma \text{ for } i = 1, 2, \cdots, n\right\}\end{aligned}$$

[The second equality follows the fact that $X_i^{-1}(S) = \Omega$ and the fact that $S \in \Sigma$]. The last equality above shows that

$$\mathscr{C} \equiv \left\{\bigcap_{i=1}^n X_i^{-1}(G_i) : G_i \in \Sigma \text{ for } i = 1, 2, \cdots, n\right\}$$

which is a π-system which generates the σ-algebra $\mathscr{G} = \sigma\{\mathscr{C}\}$, so according to Dynkin's lemma (see Lemma 2.1 below), $\mathscr{G} = \sigma\{\mathscr{C}\} = \mathscr{M}(\mathscr{C})$, where $\mathscr{M}(\mathscr{C})$ the smallest monotone class containing the π-system $\mathscr{C}$.

Now we ask, what kind of functions on Ω taking values in $\mathbb{R}$ is $\sigma\{X_1, \cdots, X_n\}$-measurable? The answer is that only those functions which are (Borel measurable) functions of $X_1, \cdots, X_n$. How to prove this? We need to prove that if $f : \Omega \to \mathbb{R}$ is $\mathscr{G}$-measurable, then there is a function $F : S \times \cdots \times S \to \mathbb{R}$ which is $\Sigma \times \cdots \times \Sigma$-measurable, such that $f = F(X_1, \cdots, X_n)$, recall here that

$$\mathscr{G} = \sigma\{X_1, \cdots, X_n\} = \sigma\{\mathscr{C}\}$$

where

$$\mathscr{C} = \bigcup_{i=1}^n X_i^{-1}(\Sigma) = \left\{\bigcap_{i=1}^n X_i^{-1}(G_i) : G_i \in \Sigma \text{ for } i = 1, 2, \cdots, n\right\}$$

is a π-system on Ω. For simplicity, we consider the case that $n = 2$ [and $S = \mathbb{R}$, $\Sigma = \mathscr{B}(\mathbb{R})$ if you like – there is no lose of generality in fact – we only use the fact that Σ is a σ-algebra on S]. Note that the product σ-algebra on the product space $S \times S$ is defined by

$$\Sigma \times \Sigma = \sigma\{A \times B : \text{ where } A, B \in \Sigma\}.$$

Let $\mathscr{H}$ be the collection of all functions $Y = F(X_1, X_2)$ where $F : S \times S \to \mathbb{R}$ is $\Sigma \times \Sigma$-measurable (real valued function). We then apply the Dynkin lemma [*function version in Question 1 in Sheet 1*] to $\mathscr{H}$, where $\mathscr{C}$ the π-system as above, and $\sigma(\mathscr{C}) = \mathscr{G}$. You need to show that 1) If $E \in \mathscr{C}$, then you can construct a function F on $S \times S$-measurable such that $1_E = F(X_1, X_2)$ [There is an obvious one of course you should find out ...], 2) then you verify the other conditions for $\mathscr{H}$ in Q1 are satisfied .. then we conclude that $\mathscr{H}$ actually exhaust all $\mathscr{G}$-measurable real functions on Ω.

2 Carathéodory's extension theorem

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

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

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

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

Since a σ-algebra must be a monotone class, so that $\mathscr{M}(\mathscr{C}) \subseteq \sigma(\mathscr{C})$ by definition. To prove the other inclusion that $\sigma(\mathscr{C}) \subseteq \mathscr{M}(\mathscr{C})$, one only needs to verify that $\mathscr{M}(\mathscr{C})$ is a σ-algebra by using the fact that $\mathscr{C}$ is a π-system. The proof is routine, see for example page 193, D. Williams: 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 $(\Omega, \mathscr{F})$, and suppose $\mathscr{C} \subseteq \mathscr{F}$ is a π-system containing the sample space Ω such that $\sigma(\mathscr{C}) = \mathscr{F}$. If $\mu_1(E) = \mu_2(E)$ for every $E \in \mathscr{C}$, then $\mu_1 = \mu_2$ on $\mathscr{F}$.*

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

Proof. Let $\mathscr{G}$ be the collections of all $E \in \mathscr{F}$ such that $\mu_1(E) = \mu_2(E)$. Then $\mathscr{C} \subseteq \mathscr{G}$ by assumptions. We prove that $\mathscr{G}$ is a monotone class. In fact, it is assumed that $\Omega \in \mathscr{G}$. Since $\mu_1(\varnothing) = \mu_2(\varnothing) = 0$ so that $\varnothing \in \mathscr{G}$. If $A, B \in \mathscr{G}$ 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 $B \setminus A \in \mathscr{G}$. Suppose now $A_n \in \mathscr{G}$, 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 $\bigcup_{n=1}^{\infty} A_n \in \mathscr{G}$. Thus $\mathscr{G}$ is a monotone class containing $\mathscr{C}$. By Lemma 2.1, $\mathscr{G} \supseteq \mathscr{M}(\mathscr{C}) = \sigma\{\mathscr{C}\} = \mathscr{F}$, so that $\mu_1 = \mu_2$ on $\mathscr{F}$. ■

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

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

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 $E \in \mathscr{F}$. Letting $n \uparrow \infty$ to obtain that $\mu_1(E) = \mu_2(E)$ for every $E \in \mathscr{F}$. ■

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 $\mathscr{H}$ be a σ-algebra over a sample space Ω, and $\mu^* : \mathscr{H} \to [0,\infty]$ be an outer measure on $(\Omega, \mathscr{H})$, so that

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

3.2) $\mu^*(A) \leq \mu^*(B)$ for any $A \subseteq B, A, B \in \mathscr{H}$; 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 $E_n \in \mathscr{H}$ $(n = 1, 2, \cdots)$.

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

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

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

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

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

where the inf runs over all countable cover $\left\{E_j\right\}$ of E and $E_j \in \mathscr{R}$. Then every set $E \in \mathscr{R}$ 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 we need to check is that every element E of $\mathscr{R}$, $\mu^*(E) = \mu(E)$ (which is direct but not trivial).

4. Null sets. A subset $E \in \mathscr{H}$ 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 $(\Omega, \mathscr{M}(\mathscr{H}, \mu^*))$.

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

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

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

3 Lebesgue-Stieltjes measures – outline of their construction

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

3.1 Construction of LS measures

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

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

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 $\mathscr{M}_\rho$ 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 $a < s \leq t < b$, 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 $\mathscr{M}_\rho$, and $m_\rho^* : \mathscr{M}_\rho \to [0,\infty]$ 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 $\mathscr{M}_\rho$ 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 that m_ρ^* restricted on the ring $\mathscr{R}(a,b)$ is finitely additive. 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

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

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 $\mathscr{C}(a,b) \subset \mathscr{R}(a,b) \subset \mathscr{M}_\rho$. Thus the Borel σ-algebra $\mathscr{B}(a,b) \subset \mathscr{M}_\rho$. 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 restriction of the outer measure m_ρ^* on $\mathscr{M}_\rho$ 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 $((a,b),\mathscr{M}_\rho,m_\rho)$, 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 increasing 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 $\mathscr{M}_\rho = \mathscr{M}_{\rho_+}$. 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.

3.2 Examples

Example 1. If ρ is a *right continuous and increasing function* on an open interval (a,b), then there is a unique measure m_ρ on the measurable space $((a,b),\mathscr{B}(a,b))$ [where $\mathscr{B}(a,b)$ denotes the Borel σ-algebra on (a,b)], such that $m_\rho((s,t]) = \rho(t) - \rho(s)$ for any $a < s \le t < b$. m_ρ is σ-finite, but may be not finite. In fact, since m_ρ is a measure, so that the total measure of (a,b) is given by

$$m_\rho\left((a,b)\right) = \lim_{n\to\infty} m_\rho\left((a+\frac{1}{n},b-\frac{1}{n}]\right) = \lim_{n\to\infty}\left[\rho\left(b-\frac{1}{n}\right) - \rho\left(a+\frac{1}{n}\right)\right] = \rho(b-) - \rho(a+) \tag{3.1}$$

where $\rho(a+)$ is the right limit of ρ at a, which exists (but may be $-\infty$), and

$$\rho(a+) = \inf_{t\in(a,b)} \rho(t) \quad \text{ and } \rho(b-) = \sup_{t\in(a,b)} \rho(t).$$

We maintain the convention that $\infty - (-\infty) = \infty$. Therefore, m_ρ is a finite measure if and only if ρ is bounded on (a,b).

Suppose $t \in (a,b)$, then the singleton $\{t\}$ is Borel measurable. Choose $\varepsilon > 0$ such that $(t-\varepsilon,t+\varepsilon] \subset (a,b)$. Then

$$m_\rho\left((t-\varepsilon,t+\varepsilon]\right) = \rho(t+\varepsilon) - \rho(t-\varepsilon) < \infty$$

and therefore, as m_ρ is a measure,

$$m_\rho\left(\{t\}\right) = \lim_{\varepsilon\to 0} m_\rho\left((t-\varepsilon,t+\varepsilon]\right) = \lim_{\varepsilon\to 0} \rho(t+\varepsilon) - \rho(t-\varepsilon) = \rho(t+) - \rho(t-).$$

which implies that $m_\rho(\{t\}) = 0$ if and only if ρ is continuous at t.

Here comes a warning. If ρ is increasing, then one can show that its derivative ρ' exists and is non-negative almost everywhere (with respect to the Lebesgue measure) and is Lebesgue measurable [This is another theorem of Lebesgue]. Hence

$$\mu_\rho(E) = \int_E \rho'(t)dt$$

where E is Lebesgue measurable, defines a measure. In general, $m_\rho \neq \mu_\rho$. Certainly $m_\rho \neq \mu_\rho$ if ρ has at least one discontinuous point. Even ρ is continuous, m_ρ may not coincide with μ_ρ.

Example 2. Let $p > 0$ be a constant. Consider $\rho(t) = t^p$ which is increasing and continuous on $(0,\infty)$. ρ is unbounded, so its associated Lebesgue-Stieltjes measure m_ρ is σ-finite, but not finite. Since the right-limit at zero $\rho(0+) = 0$, so that for every $t > 0$

$$\rho(t) = \rho(t) - \rho(0+) = m_\rho((0,t])$$

by using a similar argument as for (3.1). Actually this statement is true with obvious modification for any increasing function ρ on $(0,\infty)$ as long as $\rho(0+) = 0$. In fact, for any $t > 0$

$$\rho(t+) - \rho(0+) = m_\rho((0,t]) = \int_{(0,t]} dm_\rho \tag{3.2}$$

an elementary fact about the Lebesgue-Stieltjes measures which is very useful. Let us continue the example. While for $t > s > 0$, applying the fundamental theorem in calculus to the function x^p on $[s,t]$, we have

$$m_\rho((s,t]) = t^p - s^p = \int_s^t px^{p-1}dx = \mu_\rho((s,t])$$

where $\mu_\rho(E) = \int_E px^{p-1}dx$ defines a measure on $((0,\infty),\mathscr{B}(0,\infty))$. By applying the Uniqueness Lemma [The version for σ-finite measures] to m_ρ and μ_ρ and the π-system $\mathscr{C} = \{(s,t] : 0 < s < t\}$, we may conclude that $m_\rho = \mu_\rho$, that is $m_\rho \ll m$ [where m denotes the Lebesgue measure on $(0,\infty)$], and the Radon-Nikodym derivative

$$\frac{dm_\rho}{dm} = px^{p-1}$$

on $(0,\infty)$.

Example 3. Let $(\Omega,\mathscr{F},\mu)$ be a σ-finite measure space and let ρ be an right-continuous increasing function on $(0,\infty)$. Suppose $\rho(0+) = \inf_{t>0}\rho(t) > -\infty$. Then

$$\rho(t) = \rho(0+) + \int_{(0,t]} dm_\rho \tag{3.3}$$

for $t > 0$ according to (3.2). Suppose $X : \Omega \to [0,\infty)$ is $\mathscr{F}$-measurable, and we want to compute the expectation or integral of $\rho(X)$ on $\{X > 0\}$:

$$\int_{\{X>0\}} \rho(X)d\mu.$$

According to (3.3) we have

$$\begin{aligned}\int_{\{X>0\}} \rho(X)d\mu &= \int_{\{X>0\}} \rho(0+)d\mu + \int_{\{X>0\}} \left[\int_{(0,X]} dm_\rho\right] d\mu \\ &= \rho(0+)\int_{\{X>0\}} d\mu + \int_\Omega \left[1_{\{X>0\}}\int_{(0,\infty)} 1_{(0,X]}(t)m_\rho(dt)\right] d\mu \\ &= \rho(0+)\mu\left[\{X>0\}\right] + \int_\Omega \left[\int_{(0,\infty)} 1_{(0,X]}(t)m_\rho(dt)\right] d\mu.\end{aligned}$$

We apply the Fubini theorem to the second term on the right-hand side: changing the order of the iterated integral, to obtain

$$\begin{aligned}\int_{\{X>0\}} \rho(X)d\mu &= \rho(0+)\mu\left[\{X>0\}\right] + \int_{(0,\infty)} \left[\int_\Omega 1_{(0,X]}(t)d\mu\right] m_\rho(dt) \\ &= \rho(0+)\mu\left[\{X>0\}\right] + \int_{(0,\infty)} \mu\left[X \geq t\right] m_\rho(dt).\end{aligned}$$

Example 4. Therefore, if ρ is right-continuous and increasing, $\rho(0+) = 0$, and if X is *non-negative* real measurable function on a σ-finite measure space $(\Omega, \mathscr{F}, \mu)$, then

$$\int_{\{X>0\}} \rho(X)d\mu = \int_{(0,\infty)} \mu\left[X \geq t\right] m_\rho(dt). \tag{3.4}$$

Example 5. If ρ is *continuous and increasing* on $(0,\infty)$, then $m_\rho((s,t]) = m_\rho((s,t))$, so that

$$\rho(t) = \rho(0+) + \int_{(0,t)} dm_\rho \tag{3.5}$$

for $t > 0$, and therefore for such ρ we have

$$\int_{\{X>0\}} \rho(X)d\mu = \rho(0+)\mu\left[\{X>0\}\right] + \int_{(0,\infty)} \mu\left[X > t\right] m_\rho(dt). \tag{3.6}$$

Example 6. In particular, by applying this to $\rho(t) = t^p$ with $p > 0$, we have

$$\int_\Omega X^p d\mu = p\int_{(0,\infty)} \mu\left[X > t\right] t^{p-1} dt \tag{3.7}$$

where the right-hand side is a Lebesgue integral for a non-negative function. Here we can drop $\{X > 0\}$ in the integration on the left-hand side because $p > 0$, $X^p = 0$ on $\{X = 0\}$.

4 Generalized measures and Radon-Nikodym's derivative

4.1 Generalized measures

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

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

for $E \in \mathscr{F}$ defines a function (called a signed measure) from $\mathscr{F}$ 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 $(\Omega,\mathscr{F},\mu)$, then

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

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

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

2) μ is countably additive in the sense that

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

for any $A_i \in \mathscr{F}$ 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 $(-\infty,\infty]$.

Clearly, any signed measure $\mu = \mu_1 - \mu_2$, where μ_i are measures on $(\Omega,\mathscr{F})$ and $\mu_2(\Omega) < \infty$, is a generalized measure. The converse is also true.

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

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

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

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

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

Choose a sequence $G_n \in \mathscr{F}$ 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 $E \in \mathscr{F}$, 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 $E \in \mathscr{F}$.

We only have to show that $\mu(E \cap A^+) \geq 0$ for every $E \in \mathscr{F}$, 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 $E_0 \cap A^- = \varnothing$, 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 $E_1 \subseteq \mathscr{F}$ 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 $E \in \mathscr{F}$. Therefore it must 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.

Thus, if μ is a generalized measure over $(\Omega, \mathscr{F})$, 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 $E \in \mathscr{F}$) define two measures on $(\Omega, \mathscr{F})$. 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,

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

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

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

defines an increasing function, where the sup runs over all finite partitions $D_t : a < t_0 < t_1 < \cdots < t_n = t < b$, 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 $\left|m_\rho\right| = m_{\rho_{\mathrm{TV}}}$.

The usual concepts about measures may be applied to generalized measures via Jordan's decomposition. 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 $(\Omega, \mathscr{F})$ too. For example, an $\mathscr{F}$-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

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

4.2 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 $(\Omega, \mathscr{F})$, then we say ν is* absolutely continuous *with respect to μ, written as $\nu \ll \mu$, if $E \in \mathscr{F}$ 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 $(\Omega, \mathscr{F})$, such that $\nu \ll \mu$, then there is a* non-negative *$\mathscr{F}$-measurable function ρ such that*

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

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

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

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 $E_i^{(n)} \in \mathscr{F}$ which are disjoint, and $\rho_n = h_i$ on $E_i^{(n)}$ for $i = 1, \cdots, n$. Thus, for every $E \in \mathscr{F}$, 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 $E \in \mathscr{F}$. 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 $E \in \mathscr{F}$. To this end consider the generalized measure

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

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

The following theorem follows from a routine computation.

Theorem 4.4 *Suppose μ and ν are two σ-finite measures on $(\Omega, \mathscr{F})$, such that $\nu \ll \mu$. Let f be an $\mathscr{F}$-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*

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

4.3 Conditional expectations

This is perhaps the most important concept in probability theory. Let $(\Omega,\mathscr{F},\mu)$ be a measure space, and let $f:\Omega\to[0,\infty]$ be $\mathscr{F}$-measurable. For every $A\in\mathscr{F}$, define $\mu_f(A)=\int_\Omega f1_A d\mu=\int_A f d\mu$. Then μ_f is a measure defined on $\mathscr{F}$. In fact, if A_n is a sequence of disjoint $\mathscr{F}$-measurable subsets, then $f1_{\bigcup_{n=1}^\infty A_n}=\sum_{n=1}^\infty f1_{A_n}$, thus, by MCT (series version) we have

$$\mu_f\left(\bigcup_{n=1}^\infty A_n\right)=\int_\Omega f1_{\bigcup_{n=1}^\infty A_n}d\mu=\sum_{n=1}^\infty\int_\Omega f1_{A_n}d\mu=\sum_{n=1}^\infty\mu_f(A_n)$$

so μ_f is a measure on $(\Omega,\mathscr{F})$. μ_f possesses an important property – if $A\in\mathscr{F}$ 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 absolutely continuous with respect μ, that is, $\mu_f\ll\mu$. Conversely is also true, which is the context of Randon-Nikydom's theorem.

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

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

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

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

that is

$$\mathbb{E}[f1_A]=\mathbb{E}[1_A\mathbb{E}[f|\mathscr{G}]].$$

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

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

Conditional expectations, in most text books, are, unfortunately, only defined on probability spaces. The restriction that the total mass of the measure is finite is not necessary, and in fact conditional expectations are just projections in the Hilbert spaces L^2.

As we will see the notion of conditional expectations plays a central role in the modern probability and in analysis, we may give a formal definition as the following.

Definition. Let $(\Omega,\mathscr{F},\mu)$ be a measure space, and $\mathscr{G}\subset\mathscr{F}$ be a sub σ-algebra. Let X be $\mathscr{F}$-measurable. Then a random variable Y is called a conditional expectation of X given $\mathscr{G}$, if two conditions below hold:

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

(2) for every $A\in\mathscr{G}$, we have

$$\mathbb{E}^\mu[X:A]=\mathbb{E}^\mu[Y:A].$$

as long as both sides make sense!

1) If μ is a finite measure, any non-negative or/and integrable function X possesses a unique (up to a null set) conditional expectation given $\mathscr{G} \subset \mathscr{F}$, denoted by $\mathbb{E}[X|\mathscr{G}]$.

2) If μ is an σ-finite measure on $\mathscr{G}$, and X is σ-integrable on $\mathscr{G}$, then the conditional expectation of X given $\mathscr{G}$ exists and unique.

3) If $(\Omega, \mathscr{F}, \mu)$ is a measure space, and if $X \in L^2(\Omega, \mathscr{F}, \mu)$, i.e. X is square-integrable, then $Y = \mathbb{E}^{\mu}[X|\mathscr{G}]$ exists and is unique (up to almost everywhere). In fact $\mathbb{E}^{\mu}[X|\mathscr{G}]$ is the unique $Y \in L^2(\Omega, \mathscr{G}, \mu)$ which minimizes the L^2-distance from X to $L^2(\Omega, \mathscr{G}, \mu)$. That is

$$\mathbb{E}^{\mu}\left[|X-Y|^2\right] = \inf\left\{\mathbb{E}^{\mu}\left[|X-Z|^2\right] : Z \in L^2(\Omega, \mathscr{G}, \mu)\right\}.$$

Therefore the conditional expectation $\mathbb{E}^{\mu}[X|\mathscr{G}]$ of X given $\mathscr{G}$ is the best approximation of X in $L^2(\Omega, \mathscr{G}, \mu)$ under the mean square distance (i.e. the L^2-distance). This is why we say $\mathbb{E}^{\mu}[X|\mathscr{G}]$ is the "best guess" of X given information $\mathscr{G}$.

$X \to \mathbb{E}^{\mu}[X|\mathscr{G}]$ is called the *projection* from $L^2(\Omega, \mathscr{F}, \mu)$ onto its sub space $L^2(\Omega, \mathscr{G}, \mu)$, which possesses a very important property: the projection $X \to \mathbb{E}^{\mu}[X|\mathscr{G}]$ preserves the positivity, that is, if $X \geq 0$ almost everywhere, the so is $\mathbb{E}^{\mu}[X|\mathscr{G}]$.

Example 1. If X is $\mathscr{G}$-measurable, then $\mathbb{E}^{\mu}[X|\mathscr{G}] = X$, which is obvious and you can check that the conditions are satisfied.

Example 2. If Z is $\mathscr{G}$-measurable then

$$\mathbb{E}^{\mu}[ZX|\mathscr{G}] = Z\mathbb{E}^{\mu}[X|\mathscr{G}]$$

that is, you may take off from the conditional expectation operator $\mathbb{E}^{\mu}[\bullet|\mathscr{G}]$ those "you know already" given $\mathscr{G}$. This is called the "smoothing property" of the conditional expectations. This property can be verified by checking the two conditions for the conditional expectations.

Example 3. Suppose $\mathscr{G}_1 \subset \mathscr{G}_2$ are two sub σ-algebras of $\mathscr{F}$, then

$$\mathbb{E}^{\mu}[\mathbb{E}^{\mu}(X|\mathscr{G}_2)|\mathscr{G}_1] = \mathbb{E}^{\mu}[X|\mathscr{G}_1]$$

(as long as both sides are defined). This is called the tower property.

To prove this, we need to show that $Z \equiv \mathbb{E}^{\mu}[X|\mathscr{G}_1]$ is the conditional expectation of $Y \equiv \mathbb{E}^{\mu}(X|\mathscr{G}_2)$ given $\mathscr{G}_1$. To prove this, we check the two conditions in the definition. First Z is $\mathscr{G}_1$-measurable by definition. Now for every $A \in \mathscr{G}_1$, then $A \in \mathscr{G}_2$ too. Since $Y = \mathbb{E}^{\mu}(X|\mathscr{G}_2)$ so that

$$\mathbb{E}^{\mu}[X : A] = \mathbb{E}^{\mu}[Y : A]$$

and also $Z = \mathbb{E}^{\mu}[X|\mathscr{G}_1]$, thus

$$\mathbb{E}^{\mu}[X : A] = \mathbb{E}^{\mu}[Z : A]$$

hence

$$\mathbb{E}^{\mu}[Y : A] = \mathbb{E}^{\mu}[Z : A].$$

Thus two conditions for conditional expectations are satisfied. Therefore $Z = \mathbb{E}^{\mu}[Y|\mathscr{G}_1]$, which is what we want to prove.

Example 4. In probability theory, the conditional probability of B given A is defined by $\mathbb{P}[B|A] = \mathbb{P}(A\cap B)/\mathbb{P}(A)$ as long as $\mathbb{P}(A) > 0$. What we are going to demonstrate is that the concept of conditional expectations is the general form of conditional probabilities.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, that is $\mathbb{P}$ is a probability measure on $\mathscr{F}$. Let $\{A_n : n = 1, 2, \cdots\}$ be a measurable and finite or countable partition of the sample space Ω. That is, all $A_n \in \mathscr{F}$, A_n are disjoint, and $\Omega = \bigcup_{n=1}^{\infty} A_n$. Let $\mathscr{G} = \sigma\{A_1, A_2, \cdots\}$ be the σ-algebra generated by the partition. Suppose X is a random variable. We want to calculate $\mathbb{E}[X|\mathscr{G}]$, which is also denoted by $\mathbb{E}[X|\sigma\{A_1, A_2, \cdots\}]$, notations we have seen in the theory of Markov chains. It is a reasonable guess that the conditional expectation of X given $\mathscr{G}$ (i.e. given A_1, A_2, $\cdots$) is a linear combination of the character functions 1_{A_n}, so we may write

$$\mathbb{E}[X|\mathscr{G}] = \sum_{n=1}^{\infty} c_n 1_{A_n},$$

where c_n are constants, and we want to determine these coefficients. Using $\mathbb{E}[\mathbb{E}[X|\mathscr{G}] : A] = \mathbb{E}[X : A]$ for every $A \in \mathscr{G}$, for every A_j we have

$$\mathbb{E}\left[\sum_{n=1}^{\infty} c_n 1_{A_n} : A_j\right] = \mathbb{E}\left[X : A_j\right]$$

which yields

$$c_j \mathbb{P}\left(A_j\right) = \mathbb{E}\left[X : A_j\right].$$

Hence we must have

$$\mathbb{E}[X|\sigma\{A_1, A_2, \cdots\}] = \sum_{j=1}^{\infty} \frac{\mathbb{E}\left[X : A_j\right]}{\mathbb{P}\left(A_j\right)} 1_{A_j}$$

which results from our best guess of the conditional expectation of X given the partition, which is in fact correct. The reader may verify two conditions in the definition of conditional expectations are satisfied, as long as $\mathbb{E}\left[X : A_j\right]$ are all defined, which is the case if X is non-negative, or X is integrable, and all $\mathbb{P}\left(A_j\right) > 0$.

In particular, if $B \in \mathscr{F}$, by applying the above to $X = 1_B$ we have

$$\begin{aligned}\mathbb{E}[1_B|\sigma\{A_1, A_2, \cdots\}] &= \sum_{j=1}^{\infty} \frac{\mathbb{E}\left[1_B : A_j\right]}{\mathbb{P}\left(A_j\right)} 1_{A_j} = \sum_{j=1}^{\infty} \frac{\mathbb{P}\left[B \cap A_j\right]}{\mathbb{P}\left(A_j\right)} 1_{A_j} \\ &= \sum_{j=1}^{\infty} \mathbb{P}\left[B|A_j\right] 1_{A_j}\end{aligned}$$

which is what we have expected for condition expectation of B given $A_1, A_2, \cdots$.

In particular, if $A \in \mathscr{F}$ with $\mathbb{P}(A) > 0$ and $\mathbb{P}(A^c) > 0$, then $\{A, A^c\}$ is a partition of Ω . It is easy to see that

$$\sigma\{A\} = \sigma\{A, A^c\} = \{\Omega, \emptyset, A, A^c\}$$

and according to the previous formula

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

Recall, in elementary probability (Prelims Probability, or in Part A Probability), if X takes discrete values, then the conditional expectation of X given an event A is defined as the following

$$\mathbb{E}(X|A) = \sum_x x\mathbb{P}(X = x|A) = \frac{\sum_x x\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{E}[X : A]}{\mathbb{P}(A)}$$

so the previous formula may be written as

$$\mathbb{E}[X|\sigma\{A_1, A_2, \cdots\}] = \sum_{j=1}^{\infty} \mathbb{E}[X|A_j] 1_{A_j}.$$

Example 5. Suppose X is a real valued random variable on a probability space. $\sigma\{X\}$ is the smallest σ-algebra with which X is measurable. In fact that $\sigma\{X\} = X^{-1}(\mathscr{B}(\mathbb{R}))$. If Z is another random variable, the conditional expectation $\mathbb{E}[X|\sigma\{X\}]$ of Z given $\sigma\{X\}$ is naturally called the conditional expectation of Z given X, denoted also by $\mathbb{E}[Z|X]$. Then $\mathbb{E}[Z|X] = h(X)$ for some Borel measurable function h (h may be not unique as we only need that $h(X)$ is unique). As matter of fact, h depends only on the *joint distribution* of (Z, X).

Let us first consider the case that X takes only finite or countable many distinct values a_n ($n = 1, 2, \cdots$), such that $\mathbb{P}[X = a_n] > 0$. Then $A_n \equiv \{X = a_n\}$ form a partition of the sample space Ω, so that, according to the previous example,

$$\mathbb{E}[Z|X] = \sum_{j=1}^{\infty} \frac{\mathbb{E}[Z : X = a_j]}{\mathbb{P}[X = a_j]} 1_{\{X = a_j\}}.$$

In this case we take

$$h(x) = \sum_{j=1}^{\infty} \frac{\mathbb{E}[Z : X = a_j]}{\mathbb{P}[X = a_j]} 1_{\{a_j\}}(x)$$

for every $x \in \mathbb{R}$ which is of course Borel measurable, and it holds that $\mathbb{E}[Z|X] = h(X)$.

Recall that

$$\frac{\mathbb{E}[Z : X = x]}{\mathbb{P}[X = x]} = \mathbb{E}[Z|X = x]$$

is the expectation of Z conditional on $X = x$ (as defined in Prelims Probability), so that

$$\mathbb{E}[Z|X] = \sum_{j=1}^{\infty} \mathbb{E}[Z|X = a_j] 1_{\{X = a_j\}}.$$

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 $\mathscr{F}_1$ and $\mathscr{F}_2$ are algebras on spaces Ω_1 and Ω_2 respectively, then $\mathscr{F}_1 \times \mathscr{F}_2$ is in general not an algebra, but the collection of all finite unions $\bigcup_{j=1}^{k} A_j \times B_j$ (where $A_j \in \mathscr{F}_1$ and $B_j \in \mathscr{F}_2$ and k is a positive integer) is an algebra. If $\mathscr{F}_i$ are σ-algebras, $\mathscr{F}_1 \times \mathscr{F}_2$ is in general not a σ-algebra. Define $\mathscr{F}_1 \otimes \mathscr{F}_2$ to be the smallest σ-algebra containing $\mathscr{F}_1 \times \mathscr{F}_2$, that is, $\mathscr{F}_1 \otimes \mathscr{F}_2 = \sigma\{\mathscr{F}_1 \times \mathscr{F}_2\}$. The construction may be extended to the product space of finite many spaces. More precisely, if $(\Omega_i, \mathscr{F}_i)$ ($i = 1, \cdots, n$) are measurable spaces, then

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

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

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

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

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

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

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

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

Lemma 5.3 *Let $(\Omega_i, \mathscr{F}_i)$ ($i = 1, 2$) be measurable spaces. 1) If $A \in \mathscr{F}_1 \otimes \mathscr{F}_2$, 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. $A_{x_1} \in \mathscr{F}_2$. Similarly

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

belongs to $\mathscr{F}_1$.

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

Proof. Proof of 1). Let $\mathscr{E}$ be the collection of all $E \subseteq \Omega_1 \times \Omega_2$ such that its x_1-section is measurable. Then $\mathscr{E}$ is a σ-algebra containing all $A \times B$ where $A \in \mathscr{F}_1$ and $B \in \mathscr{F}_2$. Therefore $\mathscr{F}_1 \otimes \mathscr{F}_2 \subset \mathscr{E}$ 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 $\mathscr{F}_1 \otimes \mathscr{F}_2$-measurable), so its x_1-section is $\mathscr{F}_2$-measurable. Therefore f_{x_1} is $\mathscr{F}_2$-measurable. ■

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

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

Proof. Let $\mathscr{L}$ denote the collection of all subsets $A \in \mathscr{F}_1 \otimes \mathscr{F}_2$ such that both functions $\mu_2(A_{x_1})$ and $\mu_1(A^{x_2})$ are measurable and (5.1) holds. By a direct calculation, $\mathscr{F}_1 \times \mathscr{F}_2 \subset \mathscr{L}$. By linearlity of integration, we can see that $\mathscr{L}$ is an algebra. 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 $\mathscr{L} = \mathscr{F}_1 \otimes \mathscr{F}_2$. ■

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

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

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 $\mathscr{F}_1 \otimes \mathscr{F}_2$. 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 $E_k \in \mathscr{F}_1 \otimes \mathscr{F}_2$ 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

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

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

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

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 $(\Omega_i, \mathscr{F}_i, \mu_i)$ $(i = 1, \cdots, n)$ 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 $(\Omega_1 \times \cdots \times \Omega_n, \mathscr{F}_1 \otimes \cdots \otimes \mathscr{F}_n)$ such that*

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

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 for finite many i. This approach on the other hand works for probability spaces $(\Omega_i,\mathscr{F}_i,\mu_i)$ 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 sufficiently large i.

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

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

for any $A_i\in\mathscr{F}_i$ 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:

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

$A_k^{(j)}\in\mathscr{F}_k$ 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 $A^{(n)}\in\prod_{k=1}^{K}\mathscr{F}_k$, and therefore

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

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

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

Let $(\Omega_i, \mathscr{F}_i, \mu_i)$ $(i = 1,2)$ 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 $(\Omega_2, \mathscr{F}_2, \mu_2)$. Hence, there is a set $N_1 \in \Omega_1$ with $\mu_1(N_1) = 0$, and for any $x_1 \in \Omega_1 \setminus N_1$, $f_{x_1} \in L^1(\Omega_2, \mathscr{F}_2, \mu_2)$, 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 $\tilde{h} \in L^1(\Omega_1, \mathscr{F}_1, \mu_1)$, 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 $(\Omega_j, \mathscr{F}_j)$, where $j = 1,2$. Suppose $f : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ is a measurable function on the product measure space $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$.*

1) If $f \in L^1(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \times \mu_2)$, 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 $f \in L^1(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \times \mu_2)$.

Proof. By Theorem 5.5 and the definition of the product measure $\mu_1 \times \mu_2$, for every $E \in \mathscr{F}_1 \otimes \mathscr{F}_2$ 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 $\mathscr{F}_1 \otimes \mathscr{F}_2$-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, $\mathscr{F}_2$-measurable and $\Phi_n \uparrow$, thus by MCT applying to $\{\Phi_n\}$ on $(\Omega_2, \mathscr{F}_2, \mu_2)$ 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 $(\Omega_1, \mathscr{F}_1, \mu_1)$ 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 equations 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 $(\Omega, \mathscr{F}, \mu)$ is a σ-finite measure space, then $\mathscr{F}^\mu$ is the completed σ-algebra of $\mathscr{F}$ under the measure μ, that is, $\mathscr{N}$ denotes the collection of all subsets of Ω with outer measure zero, then $\mathscr{F}^\mu = \sigma\{\mathscr{F}, \mathscr{N}\}$. We have shown that μ can be uniquely extended to a σ-finite measure on $\mathscr{F}^\mu$, denoted again by μ. Complications may arise if we consider the completion of $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \times \mu_2)$. In general, the completion of $\mathscr{F}_1 \otimes \mathscr{F}_2$ under $\mu_1 \times \mu_2$ does not coincide with the product σ-algebra of the completions of $\mathscr{F}_i$ under μ_i, but we have

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

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

and

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

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

If $f : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ is measurable w.r.t $(\mathscr{F}_1 \otimes \mathscr{F}_2)^{\mu_1 \times \mu_2}$, 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. $\mathscr{F}_2^{\mu_2}$, however, according to definition, there is a function $\tilde{f} : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ which is measurable w.r.t. $\mathscr{F}_1 \otimes \mathscr{F}_2$ and $f = \tilde{f}$ $\mu_1 \times \mu_2$-almost surely, and $\tilde{f}_{x_1}$ is measurable w.r.t. $\mathscr{F}_2$ 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 $\mathscr{F}_2^{\mu_2}$-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 $f \in L^1(\Omega_1 \times \Omega_2, (\mathscr{F}_1 \otimes \mathscr{F}_2)^{\mu_1 \times \mu_2}, \mu_1 \times \mu_2)$, then we choose $\tilde{f}$ which is $\mathscr{F}_1 \otimes \mathscr{F}_2$-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 $(\Omega_i, \mathscr{F}_i, \mu_i)$ be two σ-finite measure spaces. Suppose $f : \Omega_1 \times \Omega_2 \to (-\infty, \infty)$ is $(\mathscr{F}_1 \otimes \mathscr{F}_2)^{\mu_1 \times \mu_2}$-measurable.*

1) If $f \in L^1(\Omega_1 \times \Omega_2, (\mathscr{F}_1 \otimes \mathscr{F}_2)^{\mu_1 \times \mu_2}, \mu_1 \times \mu_2)$, 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 $\mathscr{F}_1 \otimes \mathscr{F}_2$-measurable, then $f \in L^1(\Omega_1 \times \Omega_2, (\mathscr{F}_1 \otimes \mathscr{F}_2)^{\mu_1 \times \mu_2}, \mu_1 \times \mu_2)$.

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

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.

6.1 Laws, distribution functions

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

Proposition 6.1 *Let $(\Omega, \mathscr{F})$ and (S, Σ) be two measurable spaces, $\mathbb{P}$ a measure on $(\Omega, \mathscr{F})$, 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 $(\Omega, \mathscr{F}, \mathbb{P})$ taking values in $\mathbb{R}^n$, then $\mathbb{P} \circ X^{-1}$ is a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, 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

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

for all $a < b$. Since the collection $\mathscr{C}$ 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 $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

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

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

Let

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

Then (6.1) implies that

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

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

Definition 6.2 *1) Let $\{\mathscr{F}_\alpha : \alpha \in \wedge\}$ be a collection of sub σ-algebras of $\mathscr{F}$. Then $\{\mathscr{F}_\alpha : \alpha \in \wedge\}$ 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*

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

2) Let $\{F_\alpha : \alpha \in \wedge\}$ be a family of events: $F_\alpha \in \mathscr{F}$. 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 definition we give here 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 6.3 *Let $\mathscr{F}_\alpha \equiv \sigma\{\mathscr{C}_\alpha\}$ where each $\mathscr{C}_\alpha$ is a π-system in the sense that*

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

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

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

See section Examples below for a proof.

3. Independent random variables.

Theorem 6.4 *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 $x_1, \cdots, x_k \in \mathbb{R}$*

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

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

This follows from the previous lemma, as $\mathscr{C}_k$ the collection of all subsets $\{X_k \le x\}$ 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 $F_n(x) = \mathbb{P}[X_n \le x]$ is the distribution function of X_n, $n = 1, 2, \cdots$.

Theorem 6.5 *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 $(S, \mathscr{G})$, and let μ be the distribution of X. Then μ is a probability measure. Let $(S_n, \mathscr{G}_n, \mu_n) = (S, \mathscr{G}, \mu)$ $(n = 1, 2, \cdots)$ 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 $(\Omega, \mathscr{F}, \mathbb{P})$ (where $\mathscr{F} = \prod_{n=1}^{\infty} \mathscr{G}_n$) 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 $A_n \in \mathscr{F}$ 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

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

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 6.6 *Let $A_n \in \mathscr{F}$ (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 $x \in [0,1]$, which can be seen from the following

$$e^{-x} = 1 - x + \left(\frac{x^2}{2!} - \frac{x^3}{3!} \right) + \cdots \geq 1 - x$$

for $|x| \leq 1$. 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 does that their countable union $\{A_n : \text{ i.o}\}^c$ over $m = 1, 2, \cdots$, 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-Cantelli's 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 $(\Omega, \mathscr{F}, \mathbb{P})$, then the σ-algebra $\mathscr{G}_\infty = \bigcap_{n=1}^{\infty} \sigma\left\{X_j : j > n\right\}$ is called the tail σ-algebra of $\{X_k\}_{k\geq 1}$. Any element in $\mathscr{G}_\infty$ is called a *tail event*.

Proposition 6.7 *(*A. Kolomogorov's 0-1 law*) If $\{X_n\}$ is a sequence of* independent *random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, and $\mathscr{G}_\infty = \bigcap_{n=1}^{\infty} \sigma\left\{X_j : j > n\right\}$. Then $\mathbb{P}(A) = 0$ or 1 for every $A \in \mathscr{G}_\infty$. 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 $\sigma\left\{X_j : j \leq n\right\}$ and $\sigma\left\{X_j : j > n\right\}$ are independent for any $n = 1, 2, \cdots$, so that $\sigma\left\{X_j : j \leq n\right\}$ and $\mathscr{G}_\infty$ for every n are independent. It follows that $\bigcup_{n=1}^{\infty} \sigma\left\{X_j : j \leq n\right\}$ and

$\mathscr{G}_\infty$ are independent. If $B, C \in \bigcup_{n=1}^\infty \sigma\left\{X_j : j \leq n\right\}$, then $B \cap C \in \bigcup_{n=1}^\infty \sigma\left\{X_j : j \leq n\right\}$ as well, so $\bigcup_{n=1}^\infty \sigma\left\{X_j : j \leq n\right\}$ is a π-system, thus, by Lemma 6.3, the σ-algebra

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

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

3. Example. Suppose (X_n) is a sequence of independent random variables (real or complex), and $\mathscr{G}_\infty$ 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 $\mathscr{G}_\infty$, and thus have probability one or zero.

6.4 Examples

Motivated by the notion of independent events in Prelims Probability, we have generalized the concept of independence to families of σ-algebras.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. If $\{\mathscr{F}_\alpha : \alpha \in \Lambda\}$ (where Λ is a non-empty index set) is a family of some sub σ-algebras on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, then $\{\mathscr{F}_\alpha : \alpha \in \Lambda\}$ are *independent* if

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n) \tag{6.2}$$

for any $A_i \in \mathscr{F}_{\alpha_i}$ where $\alpha_i \in \Lambda$ $(i = 1, \cdots, n)$ as long as $\alpha_1, \cdots, \alpha_n$ are different.

By definition, if $\{\mathscr{F}_\alpha : \alpha \in \Lambda\}$ are independent, then any its sub family of $\{\mathscr{F}_\alpha : \alpha \in \Lambda\}$ are independent.

Furthermore we don't need to test (6.2) for all $A_i \in \mathscr{F}_{\alpha_i}$, and very often we only need to verify (6.2) for those A_i in a π-system $\mathscr{C}_{\alpha_i}$ as long as it generates the σ-algebra $\mathscr{F}_{\alpha_i} = \sigma\{\mathscr{C}_{\alpha_i}\}$.

The proof of this lemma is a typical application of the *uniqueness lemma for finite measures*. Suppose (6.2) holds for $A_i \in \mathscr{C}_{\alpha_i}$ for $i = 1, \cdots, n$ and we want to show (6.2) holds well for $A_i \in \mathscr{F}_{\alpha_i} = \sigma\{\mathscr{C}_{\alpha_i}\}$. If $n = 1$ then there is nothing to prove, so assume $n > 1$, and for fixed $A_i \in \mathscr{C}_{\alpha_i}$ for $i = 1, \cdots, n-1$, consider two set-functions

$$\mu_1(A) = \mathbb{P}(A_1 \cap \cdots \cap A_{n-1} \cap A)$$

and

$$\mu_2(A) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_{n-1})\mathbb{P}(A)$$

for $A \in \mathscr{F}_{\alpha_n}$. Trivially μ_1 and μ_2 are two measures on $(\Omega, \mathscr{F}_{\alpha_n})$, and $\mu_1(\Omega) = \mu_2(\Omega)$ [due to assumption that (6.2) for any n where $A_i \in \mathscr{C}_{\alpha_i}$]. Since μ_1 coincides with μ_2 on $\mathscr{C}_{\alpha_n}$, so does on $\mathscr{F}_{\alpha_n}$. We therefore have

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$$

for any $A_i \in \mathscr{C}_{\alpha_i}$ where $\alpha_i \in \Lambda$ $(i = 1, \cdots, n-1)$ and $A_n \in \mathscr{F}_{\alpha_n}$, and *for any natural number n*. If $n > 2$, we consider, for any $A_i \in \mathscr{C}_{\alpha_i}$ where $\alpha_i \in \Lambda$ $(i = 1, \cdots, n-2)$ and $A_n \in \mathscr{F}_{\alpha_n}$, two functions

$$\mu_1(A) = \mathbb{P}(A_1 \cap \cdots \cap A_{n-2} \cap A \cap A_n)$$

and

$$\mu_2(A) = \mathbb{P}(A_1)\cdots\mathbb{P}(A_{n-2})\mathbb{P}(A)\mathbb{P}(A_n).$$

Then

$$\mu_1(\Omega) = \mathbb{P}(A_1\cap\cdots\cap A_{n-2}\cap A_n) = \mathbb{P}(A_1)\cdots\mathbb{P}(A_{n-2})\mathbb{P}(A_n) = \mu_2(\Omega)$$

where the second equality follows the first step (with $n-1$ in place of n), and by assumption and again the first step that $\mu_1(A) = \mu_2(A)$ for every $A \in \mathscr{C}_{\alpha_{n-1}}$ and therefore the same equality holds for any $A \in \mathscr{F}_{\alpha_{n-1}}$. By repeating the same procedure n times we may conclude that (6.2) holds for all $A_i \in \mathscr{F}_{\alpha_i}$ as long as α_i are different, which completes the proof.

From definition, we can see immediately that a family $\{\mathscr{F}_\alpha : \alpha \in \Lambda\}$ of σ-algebras are independent, *if and only if any finite subfamily* $\{\mathscr{F}_{\alpha_i} : i = 1,\cdots,n\}$ (where α_1, $\cdots$, α_n belong to Λ, for any n as long as it is not greater than the number of elements in the index set Λ) are independent. This is due to the required equality (6.2) involves only finite many indices, so only to do with finite many σ-algebras in the family.

Another direct consequence from the definition of independence is that, if $\{\mathscr{F}_n : n = 1,2,\cdots\}$ is a sequence of sub σ-algebras, then $\{\mathscr{F}_n\}$ are independent if and only if

$$\mathbb{P}(A_1\cap\cdots\cap A_n) = \mathbb{P}(A_1)\cdots\mathbb{P}(A_n) \tag{6.3}$$

for any n, and for any $A_1 \in \mathscr{F}_1$, $\cdots$, $A_n \in \mathscr{F}_n$. This consequence follows from the fact that Ω belongs to any σ-algebra, and $\mathbb{P}(\Omega) = 1$, so that we can insert as many as you want the term Ω in the intersection on the left-hand side, and as many as you wish $\mathbb{P}(\Omega)$ on the right-hand side of (6.3), which will not alter the equality.

If $\{A_\alpha : \alpha \in \Lambda\}$ is a family of events, i.e. all $A_\alpha \in \mathscr{F}$, then $\sigma\{A_\alpha\}$ (where $\alpha \in \Lambda$) are independent if and only if

$$\mathbb{P}(A_{\alpha_1}\cap\cdots\cap A_{\alpha_n}) = \mathbb{P}(A_{\alpha_1})\cdots\mathbb{P}(A_{\alpha_n})$$

for every finite subset $\{\alpha_1,\cdots,\alpha_n\} \subset \Lambda$ as long as α_i are different. That is, events A_α (where $\alpha \in \Lambda$) are independent as defined in the Prelim Probability.

The discussion can be extended to random variables. A family $\{X_\alpha : \alpha \in \Lambda\}$ of random variables on $(\Omega,\mathscr{F},\mathbb{P})$ are independent, by definition, if the family $\sigma\{X_\alpha\}$ (where $\alpha \in \Lambda$) of sub σ-algebras are independent, and thus if and only if any finite sub family X_{α_1}, $\cdots$, X_{α_n} are independent.

Since for any real random variable X,

$$\sigma\{X\} = X^{-1}(\mathscr{B}) = \left\{X^{-1}(B) : B \subset \mathbb{R} \text{ Borel measurable }\right\}$$

where

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \equiv \{X \in B\},$$

therefore random variables X_{α_1}, $\cdots$, X_{α_n} are independent, if and only if

$$\mathbb{P}[X_{\alpha_1} \in B_1,\cdots,X_{\alpha_n} \in B_n] = \mathbb{P}[X_{\alpha_1} \in B_1]\cdots\mathbb{P}[X_{\alpha_n} \in B_n] \tag{6.4}$$

for any Borel subsets B_1, $\cdots$, B_n.

Example 1. Real random variables X_1, $\cdots$, X_n are independent, if and only if *the joint distribution* μ of $(X_1,\cdots,X_n)$, defined to be the probability measure μ on $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ by

$$\mu(E) = \mathbb{P}[(X_1,\cdots,X_n) \in E] \text{ for } E \in \mathscr{B}(\mathbb{R}^n)$$

coincides with the product measure $\mu_1 \times \cdots \times \mu_n$ on $\mathscr{B}(\mathbb{R}^n)$, where μ_i is the distribution of X_i, that is $\mu_i(B) = \mathbb{P}[X_i \in B]$ for any $B \in \mathscr{B}(\mathbb{R})$.

Proof. If the joint distribution $\mu = \mu_1 \times \cdots \times \mu_n$, then by taking $E = B_1 \times \cdots \times B_n$, we obtain

$$\mu\left(B_1 \times \cdots \times B_n\right) = \mu_1(B_1) \cdots \mu_n(B_n) \tag{6.5}$$

for all $B_i \in \mathscr{B}(\mathbb{R})$, which is equivalent to

$$\mathbb{P}[X_1 \in B_1, \cdots, X_n \in B_n] = \mathbb{P}[X_1 \in B_1] \cdots \mathbb{P}[X_n \in B_n] \tag{6.6}$$

so $X_1, \cdots, X_n$ are independent.

Conversely, suppose $X_1, \cdots, X_n$ are independent, then (6.6) holds for all $B_i \in \mathscr{B}(\mathbb{R})$. Let $\mathscr{C}$ be the collection of all subsets of $\mathbb{R}^n$ which have a form $B_1 \times \cdots \times B_n$. Then $\mathscr{C}$ is a π-system on $\mathbb{R}^n$, and $\mathscr{B}(\mathbb{R}^n) = \sigma\{\mathscr{C}\}$. (6.6) says exactly that the joint distribution μ of $X_1, \cdots, X_n$ coincides with $\mu_1 \times \cdots \times \mu_n$ on the π-system $\mathscr{C}$, so they must equal on $\mathscr{B}(\mathbb{R}^n)$ according to the uniqueness lemma for measures. This completes the proof.

Example 2. Random variables (real valued) $X_{\alpha_1}, \cdots, X_{\alpha_n}$ are independent if and only if for any Borel measurable functions f_i $(i = 1, \cdots)$ such that

$$\mathbb{E}\left[f_1\left(X_{\alpha_1}\right) \cdots f_n\left(X_{\alpha_n}\right)\right] = \mathbb{E}\left[f_1\left(X_{\alpha_1}\right)\right] \cdots \mathbb{E}\left[f_n\left(X_{\alpha_n}\right)\right] \tag{6.7}$$

as long as integrals (expectations) exist.

Proof. If (6.7) holds, then since $\sigma\{X_\alpha\} = X_\alpha^{-1}(\mathscr{B})$ (where $\mathscr{B}$ is the Borel σ-algebra), so that if $A_j \in \sigma\left\{X_{\alpha_j}\right\}$, then $A_j = X_{\alpha_j}^{-1}(B_j)$, where $B_j \in \mathscr{B}$. Thus

$$\begin{aligned}
\mathbb{P}\left(A_1 \cap \cdots \cap A_n\right) &= \mathbb{P}\left(X_{\alpha_1}^{-1}(B_1) \cap \cdots \cap X_{\alpha_n}^{-1}(B_n)\right) \\
&= \mathbb{P}\left(\{X_{\alpha_1} \in B_1\} \cap \cdots \cap \{X_{\alpha_n} \in B_n\}\right) \\
&= \mathbb{P}\left(\{X_{\alpha_1} \in B_1, \cdots, X_{\alpha_n} \in B_n\}\right) \\
&= \mathbb{E}\left(1_{\left\{X_{\alpha_1} \in B_1, \cdots, X_{\alpha_n} \in B_n\right\}}\right) = \mathbb{E}\left(1_{\left\{X_{\alpha_1} \in B_1\right\}} \cdots 1_{\left\{X_{\alpha_n} \in B_n\right\}}\right) \\
&= \mathbb{E}\left(1_{B_1}(X_{\alpha_1}) \cdots 1_{B_n}(X_{\alpha_n})\right) \\
&= \mathbb{E}\left(1_{B_1}(X_{\alpha_1})\right) \cdots \mathbb{E}\left(1_{B_n}(X_{\alpha_n})\right)
\end{aligned}$$

where the last equality follows from (6.7) applying to $f_i = 1_{B_i}$ which are Borel measurable as $B_i \in \mathscr{B}$. Hence

$$\mathbb{P}\left(A_1 \cap \cdots \cap A_n\right) = \mathbb{E}\left(1_{B_1}(X_{\alpha_1})\right) \cdots \mathbb{E}\left(1_{B_n}(X_{\alpha_n})\right) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n).$$

Now, we show that, if $X_{\alpha_1}, \cdots, X_{\alpha_n}$ are independent, then (6.7) holds. In fact, let μ_i denote the distribution of X_{α_i}, that is, $\mu_i(E) = \mathbb{P}[X_{\alpha_i} \in E]$ for $E \in \mathscr{B}$, then, by the previous example, the joint distribution μ of $(X_{\alpha_1}, \cdots, X_{\alpha_n})$ is exactly the product measure $\mu_1 \times \cdots \times \mu_n$, hence, by Fubini's theorem we have

$$\begin{aligned}
\mathbb{E}\left[f_1\left(X_{\alpha_1}\right) \cdots f_n\left(X_{\alpha_n}\right)\right] &= \int_{\Omega_1 \times \cdots \times \Omega_n} f_1(x_1) \cdots f_n(x_n) \mu(dx_1, \cdots, dx_n) \\
&= \int_{\Omega_1} \left[\cdots \left[\int_{\Omega_n} f_1(x_1) \cdots f_n(x_n) \mu_n(dx_n)\right] \cdots \right] \mu_1(dx_1) \\
&= \mathbb{E}\left[f_1\left(X_{\alpha_1}\right)\right] \cdots \mathbb{E}\left[f_n\left(X_{\alpha_n}\right)\right]
\end{aligned}$$

which proves (6.7).

Example 3. Suppose X, Y and Z are three independent real valued random variables, then, it should be clear that $X+Y$ and Z are independent, but how to prove this? While Dynkin's lemma and the uniqueness lemma for measures may help for this kind of questions.

Proof. According to definition, we want to show $\sigma\{X+Y\}$ and $\sigma\{Z\}$ are independent, that is, want to show that for any $D \in \sigma\{X+Y\}$ and $C \in \sigma\{Z\}$,

$$\mathbb{P}[D\cap C] = \mathbb{P}[D]\,\mathbb{P}[C]. \tag{6.8}$$

Since X, Y and Z are independent, so X and Y are independent too, hence

$$\mathbb{P}[A\cap B\cap C] = \mathbb{P}[A]\,\mathbb{P}[B]\,\mathbb{P}[C] = \mathbb{P}[A\cap B]\,\mathbb{P}[C]$$

that is, (6.8) holds for $D = A\cap B$ as long as $A \in \sigma\{X\}$ and $B \in \sigma\{Y\}$, all such sets consist of a π-system which generates the σ-algebra $\sigma\{X,Y\}$. Formally we let

$$\mathscr{C} = \{A\cap B : A \in \sigma\{X\}, B \in \sigma\{Y\}\}.$$

Then $\mathscr{C}$ is a π-system, and $\sigma\{X\} \subset \mathscr{C}$, $\sigma\{Y\} \subset \mathscr{C}$, so that $\sigma\{X,Y\} = \sigma\{\mathscr{C}\}$. Let $C \in \sigma\{Z\}$ be fixed but arbitrary. Considers two measures $\mu_1(D) = \mathbb{P}[D\cap C]$ and $\mu_2[D] = \mathbb{P}[D]\,\mathbb{P}[C]$ for $D \in \mathscr{F}$. Both μ_i are finite measures, $\mu_1(\Omega) = \mu_2(\Omega) = \mathbb{P}[C]$, and $\mu_1 = \mu_2$ on $\mathscr{C}$, hence by the Uniqueness Lemma for measures, $\mu_1 = \mu_2$ on $\sigma\{\mathscr{C}\} = \sigma\{X,Y\}$. It follows that (6.8) holds for any $D \in \sigma\{X,Y\}$. While we know that $X+Y$ is measurable with respect to $\sigma\{X,Y\}$, so that $\sigma\{X+Y\} \subset \sigma\{X,Y\}$, hence (6.8) holds for any $D \in \sigma\{X+Y\}$ and $C \in \sigma\{Z\}$, by definition, $X+Y$ and Z are independent.

From the proof, we can see that $f(X,Y)$ and Z are independent for any Borel measurable function f. For example $X^3 + \cos Y$ and Z are independent. You may extend this to any finite many independent random variables. For example, if $X_1,\cdots$, X_n, $Y_1,\cdots$, Y_m are independent (real) random variables on $(\Omega,\mathscr{F},\mathbb{P})$, then $f(X_1,\cdots,X_n)$ and $g(Y_1,\cdots,Y_m)$ are independent as long as f and g are Borel measurable.

Example 4. If X is a random variable, and $\mathscr{G}$ is a σ-algebra, then naturally we say X and $\mathscr{G}$ are independent (or say X is independent of the σ-algebra $\mathscr{G}$) if $\sigma\{X\}$ and $\mathscr{G}$ are independent. [This notion can be generalized to a family of random variables $\{X_\alpha\}$ and a family σ-algebras $\{\mathscr{F}_\beta\}$ in a natural way – leave for the reader as an exercise]. By definition, X and $\mathscr{G}$ are independent, if and only if X and 1_A are independent, and if and only if

$$\mathbb{E}[f(X):A] = \mathbb{E}[f(X)]\,\mathbb{P}(A)$$

for any $A \in \mathscr{G}$, and for any Borel measurable function f such that $f(X)$ is integrable.

Example 5. $\mathscr{G} \subset \mathscr{F}$ be a sub σ-algebra, and X be a random variable, non-negative or integrable, then $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$. In particular, if X and Z are independent, then $\mathbb{E}[X|Z] = \mathbb{E}[X]$.

Proof. Let $Y = \mathbb{E}[X]$. Then Y is a constant so is $\mathscr{G}$-measurable. For every $A \in \mathscr{G}$, we have

$$\mathbb{E}[X:A] = \mathbb{E}[X1_A] = \mathbb{E}[X]\,\mathbb{E}[1_A] = \mathbb{E}[\mathbb{E}[X]\,1_A]$$

so that $\mathbb{E}[X]$ is the conditional expectation of X given $\mathscr{G}$.

Example 6. Consider a sequence of independent Bernoulli trials $\{X_n : n = 1, 2, \cdots\}$, which is an i.i.d. sequence, independent identically distributed, with the same distribution

$$\mathbb{P}[X_n = 1] = p, \text{ and } \mathbb{P}[X_n = 0] = 1 - p$$

where $0 < p < 1$. Let T be the waiting time after the first time until the first success occurs

$$T = \inf\left\{ j \geq 0 : X_{j+1} = 1 \right\}.$$

Then

$$\mathbb{P}[T = k] = \mathbb{P}[X_i = 0 \text{ for } 1 \leq i \leq k \text{ and } X_{k+1} = 1] = (1-p)^k p$$

where $k = 0,\ 1,\ \cdots$. That is T has a geometric distribution. Similarly, for every $n = 1, 2, \cdots$, if L_n denotes the number of the longest success run starting from n, that is, $L_n = j$ if $X_i = 1$ for $i = n, \cdots, n + j - 1$ but $X_{n+j} = 0$, so that

$$\mathbb{P}[L_n = j] = p^j (1-p)$$

for $j = 0, 1, 2, \cdots$. Then

$$\mathbb{P}\left[\limsup_{n \to \infty} \frac{L_n}{\log_{\frac{1}{p}} n} = 1\right] = 1. \tag{6.9}$$

The proof of this result is a typical application of Borel-Cantelli lemma. First we recall from Prelims Analysis, if $\{x_n\}$ is a sequence of reals, and $l = \limsup x_n$ is real, then there is a sub-sequence $x_{n_k} \to l$. If there is no any sub-sequence such that $x_{n_l} > a$ then $\limsup x_n \leq a$.

Let $a > 1$ be arbitrary but fixed and set

$$A_n = \left\{ \frac{L_n}{\log_{\frac{1}{p}} n} > a \right\} \text{ for } n = 1, 2, \cdots .$$

Then

$$\mathbb{P}[A_n] = \mathbb{P}\left[L_n > a \log_{\frac{1}{p}} n\right] = \sum_{j > a \log_{\frac{1}{p}} n} \mathbb{P}[L_n = j]$$

$$= \sum_{j > a \log_{\frac{1}{p}} n} p^j (1-p) = (1-p) \frac{p^{\left[a \log_{\frac{1}{p}} n\right]}}{1-p} \leq p^{a \log_{\frac{1}{p}} n} = \frac{1}{n^a}$$

so that

$$\sum \mathbb{P}[A_n] = \sum_n \frac{1}{n^a} < \infty.$$

By Borel-Cantelli lemma, $\mathbb{P}[A_n \text{ i. o.}] = 0$, so that

$$\mathbb{P}\left[\limsup_{n \to \infty} \frac{L_n}{\log_{\frac{1}{p}} n} > a\right] = 0$$

which yields that

$$\mathbb{P}\left[\limsup_{n\to\infty}\frac{L_n}{\log_{\frac{1}{p}} n}\leq 1\right]=1.$$

Next we show that

$$\mathbb{P}\left[\limsup_{n\to\infty}\frac{L_n}{\log_{\frac{1}{p}} n}\geq 1\right]=1.$$

To this end, we apply the Borel-Cantelli lemma again. This time we assume that $0<a<1$ is arbitrary but fixed. Although X_n are independent, but events A_n may be not independent, so it is not a good idea to apply the Borel-Cantelli (part 2) to A_n. Instead, we apply Borel-Cantelli to a sub-sequence A_{n_k} where $n_k=\left[k\log_{\frac{1}{p}} k\right]$ for $k=1,2,\cdots$. Now, we show that

$$n_{k+1}-n_k>\left[a\log_{\frac{1}{p}} n_k\right].$$

for k large enough, so that $\{A_{n_k}\}$ are independent. To prove this, we estimate the gap

$$\begin{aligned}
n_{k+1}-\left(n_k+\left[a\log_{\frac{1}{p}} n_k\right]\right) &= \left[(k+1)\log_{\frac{1}{p}}(k+1)\right]-\left[k\log_{\frac{1}{p}} k\right]-\left[a\log_{\frac{1}{p}} n_k\right]\\
&\geq (k+1)\log_{\frac{1}{p}}(k+1)-1-k\log_{\frac{1}{p}} k-a\log_{\frac{1}{p}}\left(k\log_{\frac{1}{p}} k\right)\\
&= (k+1)\log_{\frac{1}{p}}(k+1)-1-k\log_{\frac{1}{p}} k-a\log_{\frac{1}{p}} k-a\log_{\frac{1}{p}}\log_{\frac{1}{p}} k\\
&\geq (1-a)\log_{\frac{1}{p}} k-a\log_{\frac{1}{p}}\log_{\frac{1}{p}} k\to\infty
\end{aligned}$$

as $k\to\infty$ (since $a<1$). Therefore there is k_0 such that for $k\geq k_0$

$$n_{k+1}-n_k>\left[a\log_{\frac{1}{p}} n_k\right]+1.$$

Hence $\{A_{n_k}\}$ are independent for $k\geq k_0$ and

$$\begin{aligned}
\mathbb{P}[A_{n_k}] &= \sum_{j>a\log_{\frac{1}{p}} n_k} p^j(1-p)=(1-p)\frac{p^{\left[a\log_{\frac{1}{p}} n_k\right]}}{1-p}\geq p^{a\log_{\frac{1}{p}} n_k+1}\\
&\geq \frac{p}{\left(k\log_{\frac{1}{p}} k\right)^a}
\end{aligned}$$

so that

$$\sum_{k\geq k_0}\mathbb{P}[A_{n_k}]=\infty$$

as $a<1$. Hence, by applying Borel-Cantelli to $\{A_{n_k}:k\geq k_0\}$, we conclude that $\mathbb{P}[A_{n_k}:\text{ i.o. }]=1$, so that

$$\mathbb{P}\left[\limsup_{n\to\infty}\frac{L_n}{\log_{\frac{1}{p}} n}\geq a\right]=1.$$

for every $a<1$, and therefore

$$\mathbb{P}\left[\limsup_{n\to\infty}\frac{L_n}{\log_{\frac{1}{p}} n}\geq 1\right]=1$$

which completes the proof.

7 Conditional expectations - revisited

Let us review the concept of conditional expectations in the setting of probability theory. The conditional expectation of a non-negative measurable function given $\mathscr{G} \subset \mathscr{F}$ on $(\Omega, \mathscr{F}, \mu)$, where μ is $\mathscr{G}$-finite, is defined in terms of Radon-Nikodym's theorem. Suppose X is measurable, non-negative, σ-integrable on $\mathscr{G}$ and μ is σ-finite on $\mathscr{G}$. That is, there is a sequence of $G_n \in \mathscr{G}$ such that $\mathbb{E}^{\mu}[X : G_n] < \infty$ and $\mu(G_n) < \infty$ for each n and $\bigcup_{n=1}^{\infty} G_n = \Omega$. Then $\mathbb{E}^{\mu}[X|\mathscr{G}]$ is well defined.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and $\mathscr{G} \subseteq \mathscr{F}$ be a sub σ-algebra.

If X is a non-negative real random variable on $(\Omega, \mathscr{F}, \mathbb{P})$, then there is a $\mathscr{G}$-measurable random variable $\mathbb{E}[X|\mathscr{G}]$, the conditional expectation of X, which is a unique (up to almost everywhere) function Y such that

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

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

If X is integrable, then both X^+ and X^- are non-negative, $\mathscr{F}$-measurable and integrable, hence $\mathbb{E}[X^{\pm}|\mathscr{G}]$ are defined, $\mathscr{G}$-measurable, and integrable. In particular, both $\mathbb{E}[X^{\pm}|\mathscr{G}]$ are finite almost surely. Define

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

Then $\mathbb{E}[X|\mathscr{G}]$ is $\mathscr{G}$-measurable and integrable, satisfying that $\mathbb{E}[X : A] = \mathbb{E}[\mathbb{E}(X|\mathscr{G}) : A]$ for every $A \in \mathscr{G}$.

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

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

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

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

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

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

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

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

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

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

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

Properties of the conditional expectations.

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

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

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

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

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

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

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

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

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

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

that is

$$\varphi(s) \ge \varphi(t) + \varphi'_-(t)(s-t)$$

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

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

we thus also have

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

Therefore, for a convex function φ, we have

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

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

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

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

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

8 Uniform integrability

1. Definition of uniform integrability. Let $(\Omega, \mathscr{F}, \mathbb{P})$ 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 $\mathscr{A}$ be a family of integrable functions on $(\Omega, \mathscr{F}, \mu)$. $\mathscr{A}$ is uniformly integrable if*

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

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

2. Some simple properties.

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

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

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

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

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

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

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

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

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

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

Conversely, suppose 1) and 2) are satisfied. Let $\beta = \sup_{\xi\in\mathscr{A}} \mathbb{E}[|\xi|]$. Then, by the Markov inequality, $\mathbb{P}\{|\xi| \geq N\} \leq \beta/N$ 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 $\mathbb{P}\{|\xi| \geq N\} \leq \delta$ so that $\mathbb{E}[|\xi| : |\xi| \geq N] \leq \varepsilon$ for any $\xi \in \mathscr{A}$. ■

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

Example. Let $\{\mathscr{G}_\alpha : \sigma \in \Lambda\}$ be a collection of sub σ-algebras of $\mathscr{F}$, $\eta \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and $\eta_\alpha = \mathbb{E}[\eta|\mathscr{G}_\alpha]$. Then $\mathscr{A} = \{\eta_\alpha : \alpha \in \Lambda\}$ is uniformly integrable.

We may assume that $\eta \geq 0$ otherwise consider η^+ and η^- instead. Then all η_α are non-negative. Hence for every $C > 0$ and $N > 0$ we have

$$\begin{aligned}\mathbb{E}[\eta_\alpha : \{\eta_\alpha > C\}] &= \mathbb{E}[\eta : \{\eta_\alpha > C\}] \quad \text{(property of conditional expectations)} \\ &= \mathbb{E}[\eta : \{\eta_\alpha > C, \eta \leq N\}] + \mathbb{E}[\eta : \{\eta_\alpha > C, \eta > N\}] \\ &\leq N\mathbb{P}[\eta_\alpha > C] + \mathbb{E}[\eta : \{\eta > N\}] \\ &\leq \frac{N}{C}\mathbb{E}[\eta_\alpha] + \mathbb{E}[\eta : \{\eta > N\}] \quad \text{(Markov inequality)} \\ &= \frac{N}{C}\mathbb{E}[\eta] + \mathbb{E}[\eta : \{\eta > N\}] \quad \text{(property of conditional expectations).}\end{aligned}$$

Therefore

$$\sup_\alpha \mathbb{E}[\eta_\alpha : \{\eta_\alpha > C\}] \leq \frac{N}{C}\mathbb{E}[\eta] + \mathbb{E}[\eta : \{\eta > N\}]\,.$$

Letting $C \uparrow \infty$ first to obtain that

$$\lim_{C\to\infty} \sup_\alpha \mathbb{E}[\eta_\alpha : \{\eta_\alpha > C\}] \leq \mathbb{E}[\eta : \{\eta > N\}]$$

for every $N > 0$. However the left hand side is independent of N, while the right hand side tends to zero as $N \to \infty$, so that, by letting $N \to \infty$ we obtain that

$$\lim_{C\to\infty} \sup_\alpha \mathbb{E}[\eta_\alpha : \{\eta_\alpha > C\}] = 0$$

which completes the proof.

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 $(\Omega, \mathscr{F}, \mathbb{P})$. Then $f_n \to f$ in $L^1(\Omega)$ as $n \to \infty$:*

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

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,

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

In particular

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

i.e. $\{f_n : n \geq 1\}$ 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 $\mathbb{P}(E) \leq \delta$, then

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

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

Sufficiency. By Fatou's lemma $\int_\Omega |f| \mathrm{d}\mathbb{P} \leq \sup_n \int_\Omega |f_n| \mathrm{d}\mathbb{P}$, so that $f \in L^1(\Omega)$. Therefore $\{f_n - f : n \geq 1\}$ 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 $E \in \mathscr{F}$ satisfying that $\mathbb{P}(E) \leq \delta$. Since $f_n \to f$ in probability, there is an $N > 0$ such that $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \delta$ for any $n \geq N$. Therefore

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

for $n \geq N$. 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 σalgebras* 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 $(\Omega, \mathscr{F}, \mathbb{P})$ 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 $(\mathscr{F}_n)_{n\in\mathbb{Z}_+}$ of sub σ-algebras of $\mathscr{F}$ is called* a filtration, *if $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ for every $n \in \mathbb{Z}_+$.*

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

It is useful to consider $\mathscr{F}_n$ 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 $(\Omega, \mathscr{F}, \mathbb{P})$, for every n, let $\mathscr{F}_n^X$ be the smallest σ-algebra with respect to which X_0, $\cdots$, X_n are measurable, i.e. $\mathscr{F}_n^X = \sigma\{X_m : m \le n\}$. $\left(\mathscr{F}_n^X\right)$ 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. $\mathscr{F}_n^X$ 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 $\mathscr{F}_n^X$, i.e. $X_n \in \mathscr{F}_n^X$, so that $X = (X_n)_{n\ge 0}$ is *adapted* to the filtration $\left(\mathscr{F}_n^X\right)$, 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 $\mathscr{F}_n^X$, but $\{X_n \in B\} \in \mathscr{F}_n^X$ 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 parameterized 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 $(\Omega, \mathscr{F}, \mathbb{P})$ is adapted to a filtration $(\mathscr{F}_n)$, if for every $n \in \mathbb{Z}_+$, X_n is $\mathscr{F}_n$-measurable. In this case we say $(X_n)_{n\in\mathbb{Z}_+}$ is an* adapted sequence, *or* adapted process *(with respect to $(\mathscr{F}_n)$).*

2) If $X_0 \in \mathscr{F}_0$ and if X_n is $\mathscr{F}_{n-1}$-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 $\mathscr{F}_n$ 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 $(\mathscr{F}_n)_{n\in\mathbb{Z}_+}$ be a filtration on $(\Omega,\mathscr{F},\mathbb{P})$. A measurable function $T:\Omega\to\mathbb{Z}_+\cup\{\infty\}$ [thus it may take value ∞] is called a* stopping time *(with respect to $(\mathscr{F}_n)$; if one wishes to emphasize the underlying filtration in question), if $\{T=n\}\in\mathscr{F}_n$ for every n.*

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

Let $\mathscr{F}_\infty=\sigma\{\mathscr{F}_n:n\in\mathbb{Z}_+\}\subset\mathscr{F}$. 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 $\mathscr{F}_\infty$, and for every n

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

and

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

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

If S and T are two stopping times, then $S+T$, $S\vee T=\max\{S,T\}$ and $S\wedge T=\min\{S,T\}$ are stopping times too. In fact

$$\{S+T=n\}=\bigcup_{j=0}^{n}\{S=j\}\cap\{T=n-j\},$$

$$\{S\vee T=n\}=(\{S=n\}\cap\{T\le n\})\bigcup(\{T=n\}\cap\{S\le n\})$$

and

$$\{S\wedge T\le n\}=\{S\le n\}\cap\{T\le n\}$$

belong to $\mathscr{F}_n$ for every n.

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 $(\Omega,\mathscr{F},\mathscr{F}_n,\mathbb{P})$, and $B\in\mathscr{B}(\mathbb{R})$. 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 $(\mathscr{F}_n)$. 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 $\{X_k\in B^c\}\in\mathscr{F}_k$ and $\{X_n\in B\}\in\mathscr{F}_n$, so that $\{T=n\}\in\mathscr{F}_n$. 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 $(\Omega, \mathscr{F}_n, \mathscr{F}, \mathbb{P})$, the σ-algebra $\mathscr{F}_T$ representing the information available up to the random time T is the following σ-algebra

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

where $\mathscr{F}_\infty = \sigma\{\mathscr{F}_n : n \geq 0\}$.

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

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

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

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

Proof. In fact

$$\begin{aligned} \left\{X_T 1_{\{T<\infty\}} \in G\right\} \cap \{T = n\} &= \{X_n \in G, T = n\} \\ &= \{X_n \in G\} \cap \{T = n\} \end{aligned}$$

which belongs to $\mathscr{F}_n$ for and $G \in \mathscr{B}(\mathbb{R})$ and for every $n = 0, 1, 2, \cdots$. Therefore $\left\{X_T 1_{\{T<\infty\}}\right\} \in \mathscr{F}_T$, 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 $(\mathscr{F}_n)$ 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 $(\Omega, \mathscr{F}_n, \mathscr{F}, \mathbb{P})$. Suppose each X_n is integrable.*

1) X is a martingale*, if*

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

2) X is a super-martingale *if*

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

3) X is a sub-martingale *if*

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

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

$$\mathbb{E}[X_m|\mathscr{F}_n] = X_n \quad a.s.\ \forall m \ge n.$$

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

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

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

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

1) Martingale by projection. Let $\xi \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ be an integrable random variable [i.e. $\mathbb{E}[|\xi|] < \infty$], and $X_n = \mathbb{E}[\xi|\mathscr{F}_n]$. 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 $\mathscr{F}_n$ are independent [i.e. $\sigma\{\xi_{n+1}\}$ and $\mathscr{F}_n$ are independent]. An example is that $\{\xi_n\}$ is a sequence of independent random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ and $\mathscr{F}_n = \sigma\{\xi_m : m \le n\}$. Let $X_n = \sum_{k=0}^n \xi_k$ be the partial sum sequence. Then (X_n) is a martingale if $\mathbb{E}[\xi_n] = 0$ for any n, is a super-martingale if $\mathbb{E}[\xi_n] \le 0$, and a sub-martingale if $\mathbb{E}[\xi_n] \ge 0$ 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

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

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 super-martingales. In fact, since $Z_n = \min\{X_n, Y_n\}$ so that

$$\mathbb{E}[Z_{n+1}|\mathscr{F}_n] \le \mathbb{E}[X_{n+1}|\mathscr{F}_n] \le X_n$$

and also

$$\mathbb{E}[Z_{n+1}|\mathscr{F}_n] \le \mathbb{E}[Y_{n+1}|\mathscr{F}_n] \le Y_n$$

hence $\mathbb{E}[Z_{n+1}|\mathscr{F}_n] \le Z_n$, 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, $\varphi(\xi) \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, and $\mathscr{G}$ is a sub σ-field of $\mathscr{F}$, then

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

Functions $(t\ln t)\, 1_{(1,\infty)}(t)$, $t^+ = t1_{(0,\infty)}$ 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

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

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

$t^+ = \max\{t,0\} = t1_{(0,\infty)}$ 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, resp. sub-martingale), and $S \leq T$ two* bounded stopping times. *Then $\mathbb{E}[X_T|\mathscr{F}_S] = X_S$ (resp. $\mathbb{E}[X_T|\mathscr{F}_S] \leq X_S$, resp. $\mathbb{E}[X_T|\mathscr{F}_S] \geq X_S$).*

Proof. [*The proof is not examinable.*] We only need to prove the case for super-martingales. Thus $X = (X_n)$ is a super-martingale. Since S, T are bounded, so there is $N \in \mathbb{N}$ such that $S \leq T \leq N$. Since

$$\{X_S \in G\} \cap \{T = n\} = \{X_n \in G\} \cap \{T = n\} \in \mathscr{F}_n$$

for every n and any $G \in \mathscr{B}(\mathbb{R})$, so X_S is $\mathscr{F}_S$-measurable. Similarly X_T is $\mathscr{F}_T$-measurable. Moreover

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

so X_T is integrable. Similarly X_S is integrable too.

To show that $\mathbb{E}[X_T|\mathscr{F}_S] \leq X_S$, we only need to prove that

$$\mathbb{E}[X_T : A] \leq \mathbb{E}[X_S : A] \text{ for every } A \in \mathscr{F}_S$$

or equivalently we need to show that for each $A \in \mathscr{F}_S$, we have $\mathbb{E}[X_T - X_S : A] \leq 0$.

Let $A \in \mathscr{F}_S$. Since $S \le T \le N$ and $X_T - X_S = 0$ on $\{S = T\}$, we thus have

$$\mathbb{E}[X_T - X_S : A] = \mathbb{E}[X_T - X_S : A \cap \{S < T\}].$$

Now we use the typical technique via stopping times. Write $\{S < T\}$ as disjoint union according to the values. Since $S < T \le N$, S takes only possible values $0, \cdots, N-1$, so that

$$A \cap \{S < T\} = \bigcup_{j=0}^{N-1} A \cap \{S = j\} \cap \{T > j\}$$

is the disjoint union. Since $A \in \mathscr{F}_S$, so by definition $A \cap \{S = j\} \in \mathscr{F}_j$, and also $\{T > j\} = \{T \le j\}^c \in \mathscr{F}_j$ for $j = 0, \cdots, N-1$, so that

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

Hence, as $X_S = X_j$ on $\{S = j\}$ for $j = 0, \cdots, N-1$, we have

$$\mathbb{E}[X_T - X_S : A] = \mathbb{E}\left[X_T - X_S : \bigcup_{j=0}^{N-1} A_j\right] = \sum_{j=0}^{N-1} \mathbb{E}\left[X_T - X_j : A_j\right].$$

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

$$\mathbb{E}[X_T - X_S : A] = \sum_{j=0}^{N-1} \mathbb{E}\left[X_{j+1} - X_j : A_j\right].$$

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

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 $A \in \mathscr{F}_S$. Then $A \in \mathscr{F}_{R_j}$ 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 results hold for sub-martingales.

Proof. For 1) we have $\mathbb{E}[X_T | \mathscr{F}_S] \le \mathbb{E}[X_S]$, then taking expectations both sides we obtain $\mathbb{E}(X_T) \le \mathbb{E}(X_S)$. 2) follows 1) as $T \wedge m$ and $T \wedge n$ are both stopping times, and bounded by n. ■

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

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

In particular

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

Proof. Since (X_n) is a super-martingale, so its negative part (X_n^-) is a sub-martingale, hence $\mathbb{E}[X_{T\wedge n}] \le \mathbb{E}[X_0]$ and $\mathbb{E}\left[X_{T\wedge n}^-\right] \le \mathbb{E}[X_n^-]$. Since

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

we therefore 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

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

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. Since, for every Borel measurable subset G

$$\{X_{T\wedge n} \in G\} = \left(\cup_{k=0}^n \{X_k \in G\} \cap \{T = k\}\right) \cup \left(\{X_n \in G\} \cap \{T > n\}\right) \in \mathscr{F}_n$$

for every n, so that X^T is adapted to $(\mathscr{F}_n)$. Since $|X_{T\wedge n}| \le \sum_{j=0}^n |X_j|$, so $X_{T\wedge n}$ is integrable for every $n \in \mathbb{Z}$.

Now observe that for every n

$$X_{T\wedge(n+1)} - X_{T\wedge n} = 1_{\{T>n\}}(X_{n+1} - X_n).$$

Notice that $\{T > n\} = \{T \le n\}^c$ is $\mathscr{F}_n$-measurable, so that $1_{\{T>n\}}$ is $\mathscr{F}_n$-measurable and non-negative. Therefore, after taking conditional expectation of both sides of the above equality, we obtain

$$\begin{aligned} \mathbb{E}\left[X_{T\wedge(n+1)} - X_{T\wedge n}|\mathscr{F}_n\right] &= \mathbb{E}\left[1_{\{T>n\}}(X_{n+1} - X_n)|\mathscr{F}_n\right] \\ &= 1_{\{T>n\}}\mathbb{E}[X_{n+1} - X_n|\mathscr{F}_n] \\ &\le 0 \end{aligned}$$

here the last inequality follows from the super-martingale property:

$$\mathbb{E}[X_{n+1} - X_n|\mathscr{F}_n] = \mathbb{E}[X_{n+1}|\mathscr{F}_n] - X_n \le 0.$$

Therefore X^T is a super-martingale too. ∎

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 $\{X_{T\wedge n}\}$. In fact

$$|X_{T\wedge n}| = \sum_{j=0}^{n} |X_j| 1_{\{T=j\}} + |X_n| 1_{\{T>n\}} \leq \xi$$

so the inequality follows from the DCT. ■

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 exactly 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 filtered probability space $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathbb{P})$. Let $\mathscr{T}$ denote the collection of all* finite $(\mathscr{F}_n)$-stopping times. *Then we say $X = (X_n)$ is of* class D*, if the family $\{X_T : T \in \mathscr{T}\}$ 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 $(\Omega, \mathscr{F}, \mathbb{P})$, 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 $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathbb{P})$, then so is its running maximal.

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

$$\mathbb{P}\left[\sup_{k\leq n} Y_k \geq \lambda\right] \leq \frac{1}{\lambda}\mathbb{E}\left[Y_n : \sup_{k\leq n} Y_k \geq \lambda\right] \tag{10.2}$$

for any $\lambda>0$ and for every $n=0,1,2,\cdots$.

Proof. Let $T=\inf\{k : Y_k \geq \lambda\}$. Then T is a stopping time, and

$$\{T=j\}=\left\{Y_0<\lambda,\cdots,Y_{j-1}<\lambda, Y_j\geq\lambda\right\}. \tag{10.3}$$

Therefore

$$\mathbb{P}\left[\sup_{k\leq n} Y_k \geq \lambda\right] = \mathbb{P}[T\leq n] = \sum_{j=0}^{n}\mathbb{P}[T=j].$$

By (10.3) for $j\leq n$, we have

$$\mathbb{P}[T=j]=\mathbb{P}\left[Y_j\geq\lambda, T=j\right]\leq\mathbb{E}\left[\frac{Y_j}{\lambda} : T=j\right]=\frac{1}{\lambda}\mathbb{E}\left[Y_j : T=j\right].$$

Since Y is a sub-martingale and $\{T=j\}\in\mathscr{F}_j$, so that $\mathbb{E}\left[Y_j : T=j\right]\leq\mathbb{E}[Y_n : T=j]$ for all $j\leq n$. Therefore

$$\begin{aligned}\mathbb{P}\left[\sup_{k\leq n} Y_k \geq \lambda\right] &= \sum_{j=0}^{n}\mathbb{P}[T=j]\leq\frac{1}{\lambda}\sum_{j=0}^{n}\mathbb{E}[Y_n : T=j]\\ &= \frac{1}{\lambda}\mathbb{E}[Y_n : T\leq n]=\frac{1}{\lambda}\mathbb{E}\left[Y_n : \sup_{k\leq n} Y_k\geq\lambda\right]\end{aligned}$$

which completes the proof. ■

As a consequence, we have the following important martingale inequality.

Corollary 10.10 *[*Doob's maximal inequality *for martingales] If $M=(M_n)$ is a martingale, then*

$$\mathbb{P}\left[\sup_{k\leq n}|M_k|\geq\lambda\right]\leq\frac{1}{\lambda}\mathbb{E}\left[|M_n| : \sup_{k\leq n}|M_k|\geq\lambda\right] \tag{10.4}$$

for any $\lambda>0$ and $n=0,1,\cdots$.

Proof. Since M is a martingale, so $(|M_n|)$ is a (non-negative) sub-martingale, and (10.4) follows from (10.3) immediately. ■

There is a slightly different version of the maximal inequality for super-martingales.

Theorem 10.11 *[*Doob's maximal inequality *for super-martingales] If $X=(X_n)$ is a* super-martingale*, then*

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

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

$$\mathbb{P}\left[\sup_{k\leq n}|X_k|\geq\lambda\right]\leq\frac{1}{\lambda}\left(\mathbb{E}[X_0]+2\mathbb{E}\left[X_n^-\right]\right) \tag{10.6}$$

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

Proof. Let $R = \inf\{k \geq 0 : X_k \geq \lambda\}$ and $T = R \wedge n$. Then T is a bounded stopping time. Since X is a super-martingale, so that, applying Doob's optional theorem to stopping times T and $S = 0$, one has $\mathbb{E}[X_0] \geq \mathbb{E}[X_T]$, hence

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

here for the second inequality we have used the fact that on $\left\{\sup_{k\leq n} X_k \geq \lambda\right\}$, $R \leq n$, so that $X_T = X_R \geq \lambda$, which is equivalent to (10.5).

Now we prove the second estimate. Since $X = (X_n)$ is a super-martingale, $(-X_n)$ is a sub-martingale, so that

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

together with (10.5) we deduce that

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

which is the last inequality. ■

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

Theorem 10.12 *[*Kolmogorov's inequality*] Let (X_n) be a martingale and $X_N \in L^2(\Omega, \mathscr{F}, \mathbb{P})$ where N is a positive integer. Then for any $\lambda > 0$*

$$\mathbb{P}\left[\sup_{k\leq N}|X_k| \geq \lambda\right] \leq \frac{1}{\lambda^2}\mathbb{E}\left[X_N^2\right]. \tag{10.7}$$

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

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

[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

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

for all $n \leq N$. ■

Example 10.13 *Let* (X_n) *be independent and square integrable. Then* $S_n = \sum_{k=0}^n (X_k - \mu_k)$ *where* $\mu_k = \mathbb{E}[X_k]$ *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 10.14 *Suppose* ρ *is right-continuous, increasing on* $(0,\infty)$ *and* $\rho(0+) = 0$*, and* ξ *is a non-negative random variable on* $(\Omega, \mathscr{F}, \mathbb{P})$*, then*

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

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

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

$$\mathbb{E}[(X_n^*)^p] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[(X_n)^p]. \tag{10.8}$$

2) Suppose (X_n) *is a martingale and* $p > 1$*,*

$$\mathbb{E}\left[\max_{k\le n}|X_k|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p]. \tag{10.9}$$

In particular, for $p > 1$*,*

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

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 (10.9) follows from (10.8). Thus we only need to consider non-negative sub-martingale X. Assume that $\mathbb{E}[(X_n)^p] < \infty$ otherwise the inequality is trivial. $X = (X_n)_{n\ge0}$ is a sub-martingale, and X_n are non-negative, by Doob's maximal inequality

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

for any $\lambda > 0$. If ρ is right continuous, increasing and $\rho(0+) = 0$ on $(0,\infty)$, by Lemma 10.14

$$\begin{aligned}
\mathbb{E}\left[\rho(X_n^*) : X_n^* > 0\right] &= \mathbb{E}\left[\int_{(0,X_n^*]} m_\rho(d\lambda) : X_n^* > 0\right] \\
&= \int_{(0,\infty)} \mathbb{P}\left[X_n^* \geq \lambda\right] m_\rho(d\lambda) \text{ [Fubini's Theorem]} \\
&\leq \int_{(0,\infty)} \frac{1}{\lambda}\mathbb{E}\left[X_n : X_n^* \geq \lambda\right] m_\rho(d\lambda) \\
&= \int_{(0,\infty)} \left\{\frac{1}{\lambda}\int_{\{X_n^* \geq \lambda\}} X_n d\mathbb{P}\right\} m_\rho(d\lambda) \\
&= \mathbb{E}\left[X_n\left(\int_{(0,X_n^*]} \frac{1}{\lambda} m_\rho(d\lambda)\right) : X_n^* > 0\right] \text{ Using Fubini's Theorem again].}
\end{aligned}$$

where the inequality above follows from the maximal inequality.

Let $p > 1$, and $\rho(\lambda) = \lambda^p$. Then $\rho'(\lambda) = p\lambda^{p-1}$ and $m_\rho(d\lambda) = p\lambda^{p-1}1_{(0,\infty)}d\lambda$. Applying the previous estimate to $\rho(\lambda) = \lambda^p$, we obtain that

$$\begin{aligned}
\mathbb{E}\left[(X_n^*)^p\right] &= \mathbb{E}\left[(X_n^*)^p : X_n^* > 0\right] \\
&\leq \mathbb{E}\left[X_n\left(\int_0^{X_n^*} \frac{1}{\lambda} p\lambda^{p-1} d\lambda\right) : X_n^* > 0\right] \\
&= \frac{p}{p-1}\mathbb{E}\left[X_n (X_n^*)^{p-1}\right].
\end{aligned}$$

For the term on the right-hand side, we apply the the Holder inequality

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

to $f = X_n$ and $g = (X_n^*)^{p-1}$, $\frac{1}{p} + \frac{1}{q} = 1$, so that

$$\mathbb{E}\left[(X_n^*)^p\right] \leq \frac{p}{p-1}\left(\mathbb{E}\left[(X_n)^p\right]\right)^{\frac{1}{p}}\left(\mathbb{E}\left[(X_n^*)^{(p-1)q}\right]\right)^{\frac{1}{q}}.$$

Since $\frac{1}{q} = \frac{p-1}{p}$, $(p-1)q = p$, so after simplification,

$$\mathbb{E}\left[(X_n^*)^p\right] \leq \frac{p}{p-1}\left(\mathbb{E}\left[(X_n)^p\right]\right)^{\frac{1}{p}}\left(\mathbb{E}\left[(X_n^*)^p\right]\right)^{1-\frac{1}{p}}. \tag{10.10}$$

Now $(X_k)^p$ is a sub-martingale, so that $k \to \mathbb{E}\left[(X_k)^p\right]$ is increasing, and therefore

$$\mathbb{E}\left[(X_n^*)^p\right] \leq \sum_{k=0}^n \mathbb{E}\left[(X_k)^p\right] \leq (n+1)\mathbb{E}\left[(X_n)^p\right] < \infty.$$

If $\mathbb{E}\left[(X_n^*)^p\right] = 0$, then $X_n^* = 0$ almost surely, so that $X_n = 0$ a.e. too, so Doob's inequality is true. If $\mathbb{E}\left[(X_n^*)^p\right] > 0$, then by dividing both sides of (10.10) by $\left(\mathbb{E}\left[(X_n^*)^p\right]\right)^{1-\frac{1}{p}}$ and taking p-th power both sides, to obtain

$$\mathbb{E}\left[(X_n^*)^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[(X_n)^p\right], \tag{10.11}$$

which completes the proof. ■

The Doob's inequality implies that, for a martingale $X = (X_n)$, the L^p-norm of X_n^* and the L^p-norm of X_n are equivalent as long as $p > 1$, and

$$\|X_n\|_{L^p} \leq \|X_n^*\|_{L^p} \leq \frac{p}{p-1}\|X_n\|_{L^p}.$$

In particular, for a martingale (X_n), X_n^* is p-th integrable if and only if X_n is p-th integrable for every $p > 1$.

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

$$a\log^+ b \leq a\log^+ a + \frac{b}{e}. \tag{10.12}$$

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 (10.12).*

Theorem 10.17 *If* (X_n) *is a non-negative sub-martingale, then*

$$\mathbb{E}\left[\max_{k\leq n} X_k\right] \leq \frac{e}{e-1}\left(1 + \mathbb{E}\left[X_n \log^+ X_n\right]\right). \tag{10.13}$$

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

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

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

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

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

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

Together with the previous estimate one thus deduces that

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

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 $n \to \infty$.

11.1 Doob's up-crossing lemma

Suppose (a_n) is a sequence of real numbers, and suppose $a^\star$ and $a_\star$ are the upper and lower limits of $\{a_n\}$ (which can be ∞ or $-\infty$), then $\lim_{n\to\infty} a_n$ exists (as a real number or $\pm\infty$), if and only if $a^\star = a_\star$. By definition, there are two sub-sequences n_k and m_k such that

$$\lim_{k\to\infty} a_{n_k} = a_\star \text{ and } \lim_{l\to\infty} a_{m_l} = a^\star$$

and we can choose two sub-sequences such that

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

If $a_\star < a^\star$, then we can choose two rationals $a < b$ such that

$$a_\star < a < b < a^\star$$

By looking at 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. Hence $\lim_{n\to\infty} a_n$ 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 finite.

Let us apply to this idea to the study of random sequences.

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

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

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

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

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 $\{X_0,\cdots,X_n\}$ through interval $[a,b]$. Then

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

and

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

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

If $X=(X_n)_{n\geq 0}$ is adapted with respect to a filtration $\{\mathscr{F}_n:n\geq 0\}$, then T_k are stopping times. Hence

$$\left\{U_a^b(X;n)=j\right\}=\left\{T_{2j-1}\leq n\right\}\bigcap\left\{T_{2j=1}>n\right\}\in\mathscr{F}_n$$

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

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

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 $[T_{2k-2},T_{2k-1}]$ 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

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

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

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

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 $X_{T_{2k-1}} \ge b$ and $X_{T_{2k}} \le a$ on $\{T_{2k} < \infty\}$, 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 $n \ge 1$, $k \ge 1$*

$$\mathbb{P}\left[U_a^b(X;n) \ge k\right] \le \mathbb{E}\left[\frac{(X_n - a)^-}{b-a} : U_a^b(X;n) = k-1\right]$$

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) Since X is a super-martingale, according to Doob's optional stopping theorem

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

so that it follows from (11.3) that

$$\begin{aligned}
\mathbb{P}\left[U_a^b(X;n) \ge k\right] &\le -\mathbb{E}\left[\frac{X_n - a}{b-a} : T_{2(k-1)} \le n < T_{2k-1}\right] + \mathbb{E}\left[X_{T_{2k-1}\wedge n} - X_{T_{2(k-1)}\wedge n}\right] \\
&\le \mathbb{E}\left[\frac{(X_n - a)^-}{b-a} : T_{2(k-1)} \le n < T_{2k-1}\right] \\
&\le \mathbb{E}\left[\frac{(X_n - a)^-}{b-a} : T_{2(k-1)-1} \le n < T_{2(k-1)+1}\right] \\
&= \mathbb{E}\left[\frac{(X_n - a)^-}{b-a} : U_a^b(X;n) = k-1\right]
\end{aligned}$$

which proves the first inequality. Since $U_a^b(X,n) \leq n$ and takes values in non-negative integers, so that

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

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

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

and therefore

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

which completes the proof. ■

11.2 Martingale convergence theorem

Theorem 11.3 *(*The martingale convergence theorem*, J. L. Doob) 1) Suppose $X=(X_n)_{n\geq 0}$ is a super-martingale (or a sub-martingale), bounded in $L^1(\Omega,\mathscr{F},\mathbb{P})$, that is,* $\sup_n \mathbb{E}[|X_n|] < \infty$*, then $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely and $X_\infty \in L^1(\Omega)$.*

2) If $Y=(Y_n)_{n\geq 0}$ is a non-negative super-martingale, bounded in L^1, then $Y_n \to Y_\infty$ exists, $Y_\infty \in L^1$, and $\mathbb{E}[Y_\infty|\mathscr{F}_n] \leq Y_n$ for $n\geq 0$.

3) If $M=(M_n)$ is uniformly integrable martingale*, that is, $M=(M_n)_{n\geq 1}$ is a martingale, and $\{M_n : n=0,1,\cdots\}$ is uniformly integrable, then $M_\infty = \lim_{n\to\infty} M_n$ exists almost surely, $M_n \to M_\infty$ in $L^1(\Omega)$, and $M_n = \mathbb{E}[M_\infty|\mathscr{F}_n]$ for every n.*

Proof. [*The proof is not examinable*] 1) For every pair of real numbers $a<b$, $U_a^b(X) = \lim_{n\to\infty} U_a^b(X;n)$ is the total number of up-crossings made by (X_n) through the interval (a,b). By MCT and Doob's crossing lemma we have

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

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

$$W_{(a,b)} = \{\liminf_{n\to\infty} X_n < a,\ \limsup_{n\to\infty} X_n > b\}$$

and

$$W = \bigcup \left\{ W_{(a,b)} : a < b \text{ and } a,b \text{ are rationals } \right\}.$$

where the union runs through the countable set of rational pairs (a,b), $a < b$. Then $W_{(a,b)} \subset \left\{U_a^b(X) = \infty\right\}$, so that $\mathbb{P}\left[W_{(a,b)}\right] = 0$. Hence $\mathbb{P}(W) = 0$. However, if $\omega \notin W$, then $\liminf X_n(\omega) = \limsup X_n(\omega)$, so that $\lim_{n\to\infty} X_n(\omega)$ exists (or equals $\pm\infty$), denoted it by $X_\infty(\omega)$, and we set $X_\infty(\omega) = 0$ for $\omega \in W$. Then $X_n \to X_\infty$ on W^c, so that $X_n \to X_\infty$ almost surely. According to Fatou's lemma

$$\mathbb{E}\left[|X_\infty|\right] = \mathbb{E}\left[\lim_{n\to\infty} |X_n|\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[|X_n|\right] \leq \sup_n \mathbb{E}|X_n| < \infty$$

so that $X_\infty \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. We have thus proved the first part of the theorem.

2) Since Y is a non-negative bounded super-martingale, then by 1) $Y_n \to Y_\infty$ a.e. Since $\mathbb{E}\left[Y_m|\mathscr{F}_n\right] \leq Y_n$ for $m \geq n$, letting $m \to \infty$, by Fatou's lemma (for conditional expectations),

$$\mathbb{E}\left[Y_\infty|\mathscr{F}_n\right] = \mathbb{E}\left[\lim_{m\to\infty} Y_m|\mathscr{F}_n\right] \leq \liminf_{m\to\infty} \mathbb{E}\left[Y_m|\mathscr{F}_n\right] \leq Y_n$$

the proof is therefore complete.

3) If $M = (M_n)$ is uniformly integrable martingale, then $\{M_n : n = 0,1,\cdots\}$ is bounded, so that by 1), $M_n \to M_\infty$ almost surely, and hence $M_n \to M_\infty$ in L^1 (Theorem 8.4). Since for every $m > n$ we have $\mathbb{E}\left[M_m|\mathscr{F}_n\right] = M_n$, by letting $m \to \infty$ to obtain $M_n = \mathbb{E}\left[M_\infty|\mathscr{F}_n\right]$. ∎

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 11.4 *Suppose $X = (X_n)_{n\geq 1}$ is a martingale which is bounded in L^p-space for some $p > 1$, that is,* $\sup_n \mathbb{E}\left[|X_n|^p\right] < \infty$*, then $(X_n)_{n\geq 0}$ is uniformly integrable, and $X_n \to X_\infty$ almost surely, and in L^p-space. Moreover*

$$\mathbb{E}\left[|X_\infty|^p\right] = \sup_n \mathbb{E}\left[|X_n|^p\right].$$

Proof. [*The proof is not examinable.*] It is known that $\sup_n \mathbb{E}\left[|X_n|^p\right] < \infty$ 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 $g = \lim_{n\to\infty} \sup_{k\leq n} |X_k|^p$. Applying Doob's L^p-inequality to the sub-martingale $|X_n|^p$ we have

$$\mathbb{E}\left[\sup_{k\leq n} |X_k|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|X_n|^p\right] \leq \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}\left[|X_n|^p\right].$$

Thus, by MCT we conclude that

$$\mathbb{E}\left[g\right] = \lim_{n\to\infty} \mathbb{E}\left[\sup_{k\leq n} |X_k|^p\right] \leq \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}\left[|X_n|^p\right] < \infty$$

that is g is integrable. Now $|X_n - X_\infty|^p \to 0$ almost surely, and $|X_n - X_\infty|^p \leq 2^p g$ for all n, so by Lebesgue's dominated convergence theorem, we have

$$\mathbb{E}\left[|X_n - X_\infty|^p\right] \to 0$$

as $n \to \infty$. It follows that $\|X_n\|_p \to \|X_\infty\|_p$. Since $|X_n|^p$ is a sub-martingale, so that $n \to \mathbb{E}\left[|X_n|^p\right]$ is increasing, and therefore

$$\mathbb{E}\left[|X_\infty|^p\right] = \lim_{n\to\infty} \mathbb{E}\left[|X_n|^p\right] = \sup_n \mathbb{E}\left[|X_n|^p\right].$$

∎

11.3 Downward martingale convergence theorem

Let us now consider backward martingale convergence theorem.

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

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

Let $\mathscr{F}_n=\mathscr{G}_{-n}$ where $n=\cdots,-2,-1,0$ (with the natural order in $\mathbb{Z}_-$). Then $(\mathscr{F}_n)$ (where $n=\cdots,-2,-1,0$) is a filtration, i.e. an increasing flow of σ-algebras. Then $X=(X_n)_{n\geq 0}$ is a martingale (resp. super-martingale, and resp. sub-martingale), by definition, if and only if $M_n=X_{-n}$ (where $n=\cdots,-2,-1,0$) is martingale (resp. super-martingale, resp. sub-martingale), that is, if and only if $\mathbb{E}\left[M_n|\mathscr{F}_{n-1}\right]=M_{n-1}$ (resp. $\mathbb{E}\left[M_n|\mathscr{F}_{n-1}\right]\leq M_{n-1}$, resp. $\mathbb{E}\left[M_n|\mathscr{F}_{n-1}\right]\geq M_{n-1}$) 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.

Lemma 11.5 *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space together with a decreasing family $(\mathscr{G}_n)_{n\geq 0}$ of sub σ-algebras of $\mathscr{F}$. Let $X=(X_n)_{n\geq 0}$, where $X_n\in L^1(\Omega,\mathscr{G}_n,\mathbb{P})$ 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 $\mathscr{G}_{N-n}$.*

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

Theorem 11.6 *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space together with a decreasing family $(\mathscr{G}_n)_{n\geq 0}$ of sub σ-algebras of $\mathscr{F}$. If $X=(X_n)_{n\geq 0}$ is a super-martingale with respect to $(\mathscr{G}_n)$, then $X_\infty=\lim_{n\to\infty}X_n$ exists almost surely. If in addition $\lim_{n\to\infty}\mathbb{E}\left[X_n\right]<\infty$ 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 $\mathscr{G}_{N-n}$, 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 $\bigcap_{j=1}^\infty\mathscr{G}_j$-measurable.

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

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

which implies that Q_n is non-negative, and $X_n = Q_n + \xi_n$. Therefore, in order 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 $\{X_n \leq L\} \in \mathscr{G}_n$ 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 $0 \leq \mathbb{E}[X_n] - \mathbb{E}[X_m] < \frac{\varepsilon}{2}$ for all $n, m \geq N_1$. 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 $m \leq N_1$. On the other hand, by using Markov inequality, $\mathbb{P}[X_n > L] \leq \frac{\sup_n \mathbb{E}X_n}{L}$. Let $L_0 = \frac{\sup_n \mathbb{E}X_n}{\delta}$. Then $\mathbb{P}[X_n > L] < \delta$ for all $L \geq L_0$ and for all n. Therefore $\mathbb{E}[X_m : X_n > L] < \frac{\varepsilon}{2}$ for all $m \leq N_1$. For $L \geq L_1$, then

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

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

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

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

12 Probability limit theorem

The martingale convergence theorem is a powerful tool to show the limit theorems for random sequences.

12.1 Levy's upward and downward

Corollary 12.1 *(*Levy's "Upward" theorem*) Let $\xi \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ and $(\mathscr{F}_n)_{n\geq 0}$ be an increasing family of sub σ-algebras of $\mathscr{F}$. Then*

$$\lim_{n\to\infty} \mathbb{E}[\xi|\mathscr{F}_n] = \mathbb{E}[\xi|\mathscr{F}_\infty] \textit{ almost surely and in } L^1(\Omega)$$

where $\mathscr{F}_\infty = \sigma\{\mathscr{F}_j : j \geq 0\}$.

Proof. Let $X_n = \mathbb{E}[\xi|\mathscr{F}_n]$ for $n \geq 0$. Since $\xi \in L^1(\Omega)$, $X = (X_n)$ is a uniformly integrable martingale, so by Doob's martingale convergence theorem, $X_n \to X_\infty$, for some $X_\infty \in L^1(\Omega)$, almost everywhere and in $L^1(\Omega)$. We need to show that $X_\infty = \mathbb{E}[\xi|\mathscr{F}_\infty]$. By considering ξ^+ and ξ^- instead, we may assume that ξ is non-negative. Thus $X_\infty \geq 0$ a.e. and X_∞ is $\mathscr{F}_\infty$-measurable. To show that $X_\infty = \mathbb{E}[\xi|\mathscr{F}_\infty]$ we only need to prove that for every $A \in \mathscr{F}_\infty$, $\mathbb{E}[X_\infty : A] = \mathbb{E}[\xi : A]$. To this end we consider two finite measures $\mu_1(B) = \mathbb{E}[X_\infty : B]$ and $\mu_2(B) = \mathbb{E}[\xi : B]$ for $B \in \mathscr{F}$. What we need to prove is that $\mu_1 = \mu_2$ on $\mathscr{F}_\infty$. To this end we can utilize the Uniqueness Lemma for finite measures, Lemma 2.2 or Dynkin's lemma.

Let $\mathscr{C} = \cup_{j=0}^\infty \mathscr{F}_j$ which is a π-system, and

$$\begin{aligned}\mathscr{G} &= \{B \in \mathscr{F} : \mu_1(B) = \mu_2(B)\} \\ &= \{B \in \mathscr{F} : \mathbb{E}[X_\infty : B] = \mathbb{E}[\xi : B]\}.\end{aligned}$$

If $A \in \mathscr{C}$, then there is n, such that $A \in \mathscr{F}_n \subseteq \mathscr{F}_m$ for all $m \geq n$. Hence

$$\mathbb{E}[X_m : A] = \mathbb{E}[\xi : A]$$

and by letting $m \uparrow \infty$ we obtain that$\mathbb{E}(1_A X_\infty) = \mathbb{E}(1_A \xi)$. Hence $\mathscr{C} \subseteq \mathscr{G}$. Suppose $A_n \in \mathscr{G}$ and $A_n \uparrow A$, then $X_\infty 1_{A_n} \uparrow X_\infty 1_A$ and $\xi 1_{A_n} \uparrow \xi 1_A$ as $n \uparrow \infty$. By MCT and the assumption that $\mathbb{E}[X_\infty : A_n] = \mathbb{E}[\xi : A_n]$ for every n, we conclude that

$$\mathbb{E}[X_\infty : A] = \lim_{n\to\infty} \mathbb{E}[X_\infty : A_n] = \lim_{n\to\infty} \mathbb{E}[\xi : A_n] = \mathbb{E}[\xi : A]$$

so that $A \in \mathscr{G}$ too. Thus $\mathscr{G}$ is a monotone class, containing the π-system $\mathscr{C}$. Hence, by Dynkin's lemma, $\mathscr{G} \supseteq \sigma\{\mathscr{C}\} = \mathscr{F}_\infty$, which implies that $\mu_1 = \mu_2$ on $\mathscr{F}_\infty$. Therefore $X_\infty = \mathbb{E}[\xi|\mathscr{F}_\infty]$. The proof is complete. ■

Corollary 12.2 (Levy's "Downward" theorem) *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space together with a decreasing family $(\mathscr{G}_n)_{n\geq 0}$ of sub σalgebras of $\mathscr{F}$: $\mathscr{G}_{n+1} \subseteq \mathscr{G}_n$ for every n. Let $\xi \in L^1(\Omega)$. Then*

$$\lim_{n\to\infty} \mathbb{E}[\xi|\mathscr{G}_n] = \mathbb{E}\left[\xi | \bigcap_{j=0}^{\infty} \mathscr{G}_j\right] \text{ almost surely and in } L^1(\Omega).$$

This follows from the downward martingale convergence theorem.

12.2 Kolmogorov's 0-1 law

Corollary 12.3 (Kolmogorov's 0-1 law) *Let ξ_n $(n = 1, 2, \cdots)$ be a sequence of independent random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, $\mathscr{G}_n = \sigma\{\xi_j : j \geq n+1\}$ and $\mathscr{G}_\infty = \cap_{n=0}^\infty \mathscr{G}_n$. An element in $\mathscr{G}_\infty$is called a tail event. If Z is $\mathscr{G}_\infty$-measurable and integrable, then $Z = \mathbb{E}(Z)$ almost surely. Thus any $\mathscr{G}_\infty$-measurable random variable is constant almost surely.*

Proof. Let $\mathscr{F}_n = \sigma\{\xi_j : j \leq n\}$. Then, for every n, $\mathscr{F}_n$ and $\mathscr{G}_n$ are independent. Hence $\mathscr{F}_n$ and $\mathscr{G}_\infty$ are independent for any n. Let $X_n = \mathbb{E}[Z|\mathscr{F}_n]$. Then $X = (X_n)$ is a uniformly integrable martingale, so that $X_n \to \mathbb{E}[Z|\mathscr{F}_\infty]$ almost surely and in $L^1(\Omega)$. While $\mathscr{F}_\infty \supset \mathscr{G}_\infty$, so that $\mathbb{E}[Z|\mathscr{F}_\infty] = Z$ a.e. and therefore $X_n \to Z$ almost surely and in L^1. On the other hand, since Z and $\mathscr{F}_n$ are independent, so that $X_n = \mathbb{E}[Z|\mathscr{F}_n] = \mathbb{E}[Z]$ almost everywhere. Therefore $Z = \mathbb{E}(Z)$ almost surely. ■

12.3 The strong law of large numbers

Let us prove the *strong law of large numbers* for i.i.d. sequences.

Theorem 12.4 *(A. Kolmogorov,* The Strong Law of Large Numbers*) Let $\{\xi_k\}_{k\geq 1}$ be a sequence of independent integrable random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ with the same distribution. Then $\frac{1}{n}\sum_{k=1}^{n}\xi_k \to \mathbb{E}[\xi_1]$ almost everywhere.*

Proof. Let $(\mathscr{G}_n)_{n\geq 0}$ be the decreasing family of σ-algebras generated by the sequence (X_n), where $X_n = \sum_{k=1}^{n}\xi_k$. That is

$$\mathscr{G}_n = \sigma\{X_m : m \geq n\} = \sigma\left\{X_n, \xi_j \geq n+1\right\}.$$

Since $\xi_1, \cdots, \xi_n$ are independent with the same distribution, so that

$$X_n = \sum_{i=1}^{n}\mathbb{E}[\xi_i|\xi_1 + \cdots + \xi_n] = n\mathbb{E}[\xi_1|\xi_1 + \cdots + \xi_n],$$

which implies that

$$\begin{aligned}\frac{X_n}{n} &= \mathbb{E}[\xi_1|X_n] = \mathbb{E}\left[\xi_1|X_n, \xi_j \text{ for } j \geq n+1\right] \\ &= \mathbb{E}[\xi_1|\mathscr{G}_n]\end{aligned}$$

for every n. Thus according to Levy's downward theorem

$$\frac{X_n}{n} \to \mathbb{E}\left[\xi_1|\bigcap_{n=1}^{\infty}\mathscr{G}_n\right] = \mathbb{E}[\xi_1|\mathscr{G}_\infty]$$

almost surely and in L^1, where $\mathscr{G}_\infty$ is the tail σ-algebra. According to Kolmogorov's 0-1 law, since $\mathbb{E}[\xi_1|\mathscr{G}_\infty]$ is $\mathscr{G}_\infty$-measurable, $\mathbb{E}[\xi_1|\mathscr{G}_\infty] = \mathbb{E}[\xi_1]$ almost surely. Therefore

$$\frac{X_n}{n} \to \mathbb{E}[\xi_1]$$

almost surely and in $L^1(\Omega)$. ■

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.

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 $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathbb{P})$, so that $\mathbb{E}[X_{n+1}|\mathscr{F}_n] \leq X_n$, hence 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 $\mathscr{F}_n$, to obtain

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

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

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

for $n \geq 0$. Since X_n is a super-martingale, (A_n) is increasing and predictable, with $A_0 = 0$, and $M_n = X_n + A_n$ 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 $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathbb{P})$. 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*

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

for $n \geq 1$, and

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

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 $\mathbb{E}\left[M_n^2\right] < \infty$ 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

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

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*

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

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

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

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 variation process $[M]_n$ associated with M is defined by $[M]_0 = 0$ and

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

for $n \geq 1$.

3) A martingale $M = (M_n)$ is called a squared integrable martingale *if* $\sup_n \mathbb{E}\left[M_n^2\right] < \infty$ *(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 $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathbb{P})$ 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

$$\begin{aligned}\langle M, N \rangle_n &= \frac{1}{4}\left(\langle M+N \rangle - \langle M-N \rangle\right) \\ &= \sum_{j=0}^{n-1} \mathbb{E}\left[(M_{j+1} - M_j)(N_{j+1} - N_j) | \mathscr{F}_j\right].\end{aligned}$$

for $n \geq 1$, $\langle M, N \rangle_0 = 0$, which is a predictable process.

Suppose $M = (M_n)$ 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

$$\begin{aligned}\langle H.M \rangle_n &= \sum_{j=0}^{n-1} \mathbb{E}\left[\left(H_{j+1}(M_{j+1} - M_j)\right)^2 | \mathscr{F}_j\right] = \sum_{j=0}^{n-1} H_{j+1}^2 \mathbb{E}\left[(M_{j+1} - M_j)^2 | \mathscr{F}_j\right] \\ &= \sum_{j=1}^{n} H_j^2 \left(\langle M \rangle_j - \langle M \rangle_{j-1}\right)\end{aligned}$$

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 $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathbb{P})$ 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 $\mathbb{E}\left[M_{T\wedge n}^2 - M_0^2\right] = \mathbb{E}\left[\langle M \rangle_{T\wedge n}\right] \leq l$ 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 $l > 0$. ■