Part I

Analysis and measure theory

1 Measure theory: introductory remarks

The aim of measure theory is to give a systematic method for thinking about concepts such as the size or content of subsets of a given set S. The theory for general sets will need some time to develop, but we can develop intuition for the type of definitions we need by studying the finite setting first.

Example 1.1 *Suppose $S = \{s_1, ..., s_n\}$ is a finite set.*

1. *The **counting measure** on S is the function which assigns to any subset $A \subseteq S$ the natural number $|A|$, the number of elements in A. This function is well-defined on the power set $\mathcal{P}(S)$, i.e. the set of all subsets of S.*
2. *If $\lambda_1, .., \lambda_n$ are real numbers in $[0, 1]$ which satisfy $\sum_{i=1}^{n} \lambda_i = 1$. To each singleton $\{s_i\}$ we can define $P(\{s_i\}) = \lambda_i$. There is then a unique extension of this function to $\mathcal{P}(S)$ such that for any subset $\{s_{i_1}, ..., s_{i_k}\} \subseteq S$ the function P is finitely additive, i.e.*

$$P\left(\{s_{i_1}, ..., s_{i_k}\}\right) = \sum_{l=1}^{k} \lambda_{i_l}. \tag{1}$$

In these two examples the finiteness of the underlying set S makes it easy to define a measure on the power set (which must also be finite). For general uncountably infinite sets, this may no longer be possible. The approach must instead be to begin with a measure defined on a subclass of sets having a high level of structure, and then to extend the measure to the greatest possible extent. The prototype we will refer to most often is the generalisation of length, area and volume etc. on subsets of $\mathbb{R}^n$. For example when $n = 1$, we understand that length for intervals can be defined by

$$\lambda((a, b)) = \lambda((a, b]) = \lambda([a, b)) = \lambda([a, b]) = b - a \in [0, \infty]$$

for $a \leq b$. We can similarly define the areas of Cartesian products of intervals (i.e. rectangles) in $\mathbb{R}^2$ and so on. Two immediate questions we need to ask are:

1. To which subsets of $\mathbb{R}$ is it possible to extend λ?
2. What properties does such an extension need to have?

These two questions are related. We will see that the key is to insist that λ is *countably additive* in the sense that for any countably infinite collection of disjoint subsets $(A_j)_{j=1}^{\infty}$ in $\mathbb{R}$ we need

$$\lambda\left(\cup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \lambda(A_j).$$

Of course, to make this precise we need to know that λ is defined on $\cup_{j=1}^{\infty} A_j$ which brings us back to question 1 above.

2 Classes of sets

It is important that the set of subsets on which we define our measures have the right properties. In this lecture we consider some different systems of subsets.

2.1 Topological spaces

Given a set X we recall that a topology on X is a collection $\mathcal{O}$ of subsets of X which satisfies

1. $\emptyset, X \in \mathcal{O}$.
2. For any indexing set I such that $O_i \in \mathcal{O}$ for all $i \in I$ we have $\cup_{i \in I} O_i \in \mathcal{O}$.
3. For any $O_1, ..., O_n \in \mathcal{O}$ we have $\cap_{i=1}^n O_i \in \mathcal{O}$.

Note that while topological spaces are stable under arbitrary unions, they are, in general, only stable under finite intersections. We call the pair $(X, \mathcal{O})$ a **topological space**.

Definition 2.1 *Given a topological space* $(X, \mathcal{O})$, *we say a subset* $A \subseteq X$ *is* ***open*** *if* $A \in \mathcal{O}$ *and* ***closed*** *if* $A^c := X \setminus A \in \mathcal{O}$.

Remark 2.2 *Note that a subset can be open, closed, neither or both.*

A typical example for us will be where (X, d) is a metric space. We can then define a topology (exercise: check that this really is a topology) by saying that $A \in \mathcal{O}$ if for any $a \in A$ there exists $\delta > 0$ such that

$$B(a, \delta) := \{x \in X : d(x, a) < \delta\} \subset A.$$

The following notions can be used for general topological spaces.

Definition 2.3 *If* $(X, \mathcal{O})$ *is a topological space and* $x \in X$ *we say that*

1. *A subset* $A \subseteq X$ *is a neighbourhood of* x *is there exists* $O \in \mathcal{O}$ *such that* $x \in O \subseteq A$.
2. x *is a limit point of* $A \subseteq X$ *if every neighbourhood of* x *has a non-empty intersection with* $A \setminus \{x\}$.

The following exercises are good to do to test your understanding of these basic definitions.

Exercise 2.4 *Prove that* $A \subseteq X$ *is open iff* A *is a neighbourhood of every element in* A. *Show that* A *is closed iff every limit point of* A *is an element of* A.

Topology is tailored to study continuous functions. Suppose that $(X_1, \mathcal{O}_1)$ and $(X_2, \mathcal{O}_2)$ are two topological spaces, recall that a function $f : X_1 \to X_2$ is continuous if for every $O_2 \in \mathcal{O}_2$ we have

$$f^{-1}(O_2) := \{x \in X_1 : f(x) \in O_2\} \in \mathcal{O}_1.$$

Topological spaces and continuous functions have some drawbacks from the point of view of measure theory. For example, the pointwise limit of continuous functions is not continuous (e.g. take $f_n : \mathbb{R} \to \mathbb{R}$ given by $f_n(x) = \min(nx, 1)\, 1_{(0,\infty)}(x)$) and a countable collection of open sets may not remain open (e.g. $\cap_{n=1}^{\infty}(-1/n, 1 + 1/n) = [0, 1]$).

2.2 Fields and σ-fields

Definition 2.5 *Given a set Ω, and a collection of subset $\mathcal{A}$ of Ω we say that $\mathcal{A}$ is a field if:*

1. $\emptyset, \Omega \in \mathcal{A}$
2. *If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$*
3. *If $A_1, ..., A_n \in \mathcal{A}$ then $\cup_{k=1}^{n} A_k \in \mathcal{A}$*

We say that $\mathcal{A}$ is a $\sigma-$field if, in addition to 3, we have that $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$ for any $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{A}$.

σ-fields have the important property of being stable under countably many set operations.

Example 2.6 *The trivial $\sigma-$field is given by $\{\emptyset, \Omega\}$*

Exercise 2.7 *Suppose for every $i \in I$ that $\mathcal{A}_i$ is a $\sigma-$field of subsets of Ω, prove that $\cap_{i \in I} \mathcal{A}_i$ is again a σ-field.*

Given a class (not necessarily a σ-field) $\mathcal{G}$ of subsets of Ω we define $\sigma(\mathcal{G})$ to be the smallest σ-field containing the class $\mathcal{G}$. It is an easy exercise to check that this is well-defined by using the previous exercise to show that

$$\sigma(\mathcal{G}) = \cap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field containing } \mathcal{G}\}. \tag{2}$$

Note that the power set of Ω is a σ-field containing $\mathcal{G}$, so the intersection in (2) is not over the empty set.

Definition 2.8 *If $(X, \mathcal{O})$ is a topological space we define the Borel σ-field on X to be $\sigma(\mathcal{O})$, i.e. it is the smallest σ-field containing all the open sets. We denote the Borel σ-field by $\mathcal{B}(X)$.*

Lemma 2.9 *Let $X = \mathbb{R}$ endowed with the Euclidean topology and let $\pi(\mathbb{R})$ denote the class of intervals*

$$\pi(\mathbb{R}) = \{(-\infty, x] : x \in \mathbb{R}\},$$

then $\sigma(\pi(\mathbb{R})) = \mathcal{B}(\mathbb{R})$.

Proof. Let $x \in \mathbb{R}$ and notice that

$$(-\infty, x] = \cap_{n=1}^{\infty} \underbrace{\left(-\infty, x + \frac{1}{n}\right)}_{\in \mathcal{O}} \in \mathcal{B}(\mathbb{R}).$$

It follows at once that $\pi(\mathbb{R}) \subseteq \mathcal{B}(\mathbb{R})$ and hence, since $\mathcal{B}(\mathbb{R})$ is a σ-field, that $\sigma(\pi(\mathbb{R})) \subseteq \mathcal{B}(\mathbb{R})$.

To prove the converse, we let $G \in \mathcal{O}$ and use the fact that G can be expressed as the countable union of disjoint open intervals: $G = \cup_{i=1}^{\infty} (a_i, b_i)$ where $a_i < b_i$. For any $a < b$ we then notice that

$$(a, b] = (-\infty, a]^c \cap (-\infty, b] \in \sigma(\pi(\mathbb{R}))$$

and hence

$$(a, b) = \cup_{n=1}^{\infty} \left(a, b - \frac{\epsilon}{n}\right] \in \sigma(\pi(\mathbb{R}))$$

where $\epsilon := 2^{-1}(b - a)$. It follows that $G \in \sigma(\pi(\mathbb{R}))$ and thus $\mathcal{O} \subseteq \sigma(\pi(\mathbb{R}))$ which implies $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\pi(\mathbb{R}))$, completing the proof. ∎

3 Measures

We want to provide a systematic study of the measurement of subsets belonging to a σ-field $\mathcal{A}$ of subset of Ω.

Definition 3.1 *Let Ω be a set and $\mathcal{A}$ and σ-field of subsets of Ω. We call the pair $(\Omega, \mathcal{A})$ a **measurable space.***

Definition 3.2 *If $(\Omega, \mathcal{A})$ a measurable space, a measure $\mu : \mathcal{A} \to [0, \infty]$ is a set function with the following properties:*

1. $\mu(\emptyset) = 0$;
2. *If $\{A_n : n \in \mathbb{N}\}$ is any collection of (pairwise) disjoint subsets in $\mathcal{A}$ then*

$$\mu\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{3}$$

Remark 3.3 *A set function satisfying (3) is said to be countably additive.*

Definition 3.4 *If $(\Omega, \mathcal{A})$ a measurable space, and μ is a measure on $(\Omega, \mathcal{A})$ we call the triple $(\Omega, \mathcal{A}, \mu)$ a **measure space.***

If $\mu(\Omega) < \infty$ then we call μ a finite measure. A measure is called σ-finite if there exists a sequence of subsets $(A_n)_{n=1}^{\infty}$ in $\mathcal{A}$ such that $A_n \subseteq A_{n+1}$, $\mu(A_n) < \infty$ for all n and $\Omega = \cup_{n=1}^{\infty} A_n$. All the measures we deal with will be σ-finite where intuition usually adapts well from the finite case. Non-σ-finite measures can exhibit strange properties.

3.1 Carathéodory's extension theorem

We will often not define a measure on a full σ-field initially, but on some smaller class of subsets where we understand the measure well. We will then try to extend the measure to the $\sigma-$field generated by this smaller class. If $\mathcal{A}$ is a class of subset of Ω which is *not* necessarily a σ-field, and $\mu : \mathcal{A} \to [0, \infty]$ is a function satisfying $\mu(\emptyset) = 0$ and (3) for any collection of disjoint sets $\{A_n : n \in \mathbb{N}\}$ for which $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$, then we call μ a **pre-measure on $\mathcal{A}$** .

Definition 3.5 *A class of subsets $\mathcal{S}$ is called a **semi-ring** if*

1. *$\emptyset \in \mathcal{S}$.*
2. *A, $B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$.*
3. *For all $A, B \in \mathcal{S}$ there exist disjoint $A_1, ..., A_n \in \mathcal{S}$ such that $A \setminus B = \cup_{i=1}^{n} A_i$.*

Exercise 3.6 *Prove that $\mathcal{S} := \{(a, b] : -\infty \leq a \leq b < \infty\}$ is a semi-ring of subsets of $\mathbb{R}$.*

Theorem 3.7 (Carathéodory's extension theorem) *Let $\mathcal{S}$ be a semi-ring of subsets of a set Ω and suppose $\mu : \mathcal{S} \to [0, \infty]$ is a pre-measure on $\mathcal{S}$. Then μ can be extended to a measure on $\sigma(\mathcal{S})$. If μ is $\sigma-$finite then this extension is unique.*

Proof (sketch). Define the set function μ^* on the power set $\mathcal{P}(\Omega)$ by

$$\mu^*(A) := \inf \left\{ \sum_{n=1}^{\infty} \mu(S_n) : A \subseteq \cup_{n=1}^{\infty} S_n, S_n \in \mathcal{S} \right\}.$$

Step 1: Prove that μ^* is an *outer measure*, i.e.

1. $\mu^*(\emptyset) = 0$;
2. $A \subseteq B$ then $\mu^*(A) \leq \mu^*(B)$;
3. μ^* is countably subadditive: $\mu^*(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$.

Step 2: $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{S}$.
Step 3: Let $\mathcal{A}^*$ denote the class of sets $A \subseteq \Omega$ such that

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A) \text{ for all } B \subseteq \Omega,$$

and prove that $\mathcal{S} \subset \mathcal{A}^*$.
Step 4: Show that $\mathcal{A}^*$ is a $\sigma-$field and μ^* is a measure on $(\Omega, \mathcal{A}^*)$.
Step 5: Show that μ^* is a measure on $\sigma(\mathcal{S}) \subseteq \mathcal{A}^*$ which extends μ on S.
Step 6: Uniqueness in the σ-finite case. ∎

3.2 Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

We have seen in Exercise 3.6 that $\mathcal{S} = \{(a, b] : -\infty \leq a \leq b < \infty\}$ is a semi-ring. We can define a pre-measure on $\mathcal{S}$ by setting

$$\mu((a, b]) = b - a.$$

To check that this is a pre-measure first observe that $\mu(\emptyset) = \mu((a, a]) = 0$. Second we need to show that if $(A_n)_{n=1}^{\infty}$ is a collection of disjoint sets in $\mathcal{S}$ such that $\cup_{n=1}^{\infty} A_n \in \mathcal{S}$ then

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

The proof is not trivial (see D. Williams A1.9). The key is is to first prove finite additivity and then use a compactness argument. Carathéodory's extension theorem allows us to extend μ to a measure on $\sigma(\mathcal{S})$, but in Lecture 2 we proved that $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$. Further μ is σ-finite since $\mathbb{R} = \cup_{n=1}^{\infty}(-n, n]$, $(-n, n] \in \mathcal{S} \subseteq \sigma(\mathcal{S})$ and $\mu((-n, n]) = 2n < \infty$, and hence the measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is uniquely defined. We call this measure the *Lebesgue measure* on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

3.3 Completions

If $(\Omega, \mathcal{A}, \mu)$ is a measure space we say $N \subseteq \Omega$ is *null* if $N \in \mathcal{A}$ and $\mu(N) = 0$. Morally, if $N \in \mathcal{A}$ is null then any $A \subseteq N$ ought also to be measurable and $\mu(A) = 0$. Let

$$\mathcal{N} = \{A \subseteq \Omega : \exists N \in \mathcal{A}, A \subseteq N \text{ and } \mu(N) = 0\}.$$

Definition 3.8 *A measure space* $(\Omega, \mathcal{A}, \mu)$ *is called **complete** if* $\mathcal{N} \subseteq \mathcal{A}$.

In other words, every subset of every null set is $\mathcal{A}$-measurable (and hence also null). A given measure space will not always be complete, but we can always make it complete. To so this we first extend the σ-field by defining

$$\mathcal{A}^* := \{A \cup N : A \in \mathcal{A} \text{ and } N \in \mathcal{N}\}.$$

It is not difficult to prove that $\mathcal{A}^*$ is another σ-field, in fact $\mathcal{A}^* = \mathcal{A} \vee \mathcal{N} := \sigma(\mathcal{A} \cup \mathcal{N})$ (Exercise: prove this!). We can then define a measure μ^* on $(\Omega, \mathcal{A}^*)$ by setting

$$\mu^*(A^*) := \mu(A),$$

where $A^* = A \cup N$. We call the measure space $(\Omega, \mathcal{A}^*, \mu^*)$ the *completion* of $(\Omega, \mathcal{A}, \mu)$.

Exercise 3.9 *Check that* μ^* *is well-defined, i.e. if* $A^* = A \cup N = B \cup M$ *for* $A, B \in \mathcal{A}$ *and* $N, M \in \mathcal{N}$ *then* $\mu(A) = \mu(B)$. *Hence deduce that* $\mu^*(A) = \mu(A)$ *for all* $A \in \mathcal{A}$*, and prove that* μ^* *is a measure.*

4 Measurable sets

We define the Lebesgue measurable subsets to be $\mathrm{Leb}(\mathbb{R}) = \mathcal{B}(\mathbb{R})^*$, i.e. the completion of the Borel-measurable sets. The extension μ^* of the Lebesgue measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ constructed in this way will still be called the Lebesgue measure (on $(\mathbb{R}, \mathrm{Leb}(\mathbb{R}))$).

Remark 4.1 *$Leb(\mathbb{R}) \setminus \mathcal{B}(\mathbb{R})$ is non-empty.*

4.1 Extended example: A non-Lebesgue measurable subset of $[0, 2\pi)$

Consider the measure space $([0, 2\pi), \mathrm{Leb}([0, 2\pi)), \lambda)$, where $\mathrm{Leb}([0, 2\pi)) = \mathcal{B}([0, 2\pi))^*$ and λ is the extension of the Lebesgue measure on $\mathcal{B}([0, 2\pi))$ to $\mathrm{Leb}([0, 2\pi))$. We will need the axiom of choice:

Axiom 4.2 (Axiom of choice) *Let X be a set. A choice function is a function f which associates to each non-empty subset $E \subseteq X$ an element of E, i.e. $f(E) \in E$. We assume that every set has a choice function.*

Let $S^1 = \left\{e^{i\theta} : \theta \in \mathbb{R}\right\}$ is it straight forward to identify S^1 with $[0, 2\pi)$ via a homeomorphism. $\mathcal{B}([0, 2\pi))$ can thus be identified with $\mathcal{B}\left(S^1\right)$ and $\mathcal{B}([0, 2\pi))^*$ with $\mathcal{B}\left(S^1\right)^*$. We will construct a subset of S^1 which is not in $\mathcal{B}\left(S^1\right)^*$. To do so, we define an equivalence relation $\sim$ on S^1 by declaring that

$$z \sim w \text{ if } z = e^{i\alpha} \text{ and } w = e^{i\beta} \text{ for some } \alpha, \beta \in \mathbb{R} \text{ such that } \alpha - \beta \in \mathbb{Q}.$$

This equivalence relation partitions S^1 into disjoint equivalence classes. Using the axiom of choice, there exists a set A which contains precisely one member of each equivalence class. We then define for every $q \in \mathbb{Q}$ the set

$$A_q := e^{iq} A := \left\{e^{iq} a : a \in A\right\}.$$

It is easy to see (from the definition of A) that $A_q \cap A_p = \emptyset$ if q, p are distinct rationals. Furthermore, we have that $S^1 = \cup_{q \in \mathbb{Q}} A_q$. If A_q were Lebesgue measurable we would have $\lambda(A_q) = \lambda(A)$ for every $q \in \mathbb{Q}$ and since λ is countably additive it would follow that

$$2\pi = \lambda\left(S^1\right) = \sum_{q \in \mathbb{Q}} \lambda(A_q) = \begin{cases} 0 & \text{if } \lambda(A) = 0 \\ \infty & \text{if } \lambda(A) \neq 0 \end{cases},$$

and in either case this is a contradiction. It follows that $A \notin \mathrm{Leb}([0, 2\pi))$.

Remark 4.3 *The proof is not constructive in that it relies on the choice function from the axiom of choice. This is quite typical of uses of the axiom of choice in mathematics. Solovay (1970) proved that it is* not *possible to prove the existence of a non-Lebesgue measurable set without using the axiom of choice. It is possible, although very difficult, to construct a subset of $\mathbb{R}$ which is not in $\mathcal{B}(\mathbb{R})$.*

4.2 D-systems and π-systems

Definition 4.4 (d-system) *A d-system (or Dynkin system) $\mathcal{D}$ is a class of subsets of Ω such that:*

1. $\Omega \in \mathcal{D}$.
2. *If $A, B \in \mathcal{D}$ with $A \subseteq B$ then $B \setminus A \in \mathcal{D}$.*
3. *If $(A_n)_{n=1}^{\infty}$ is a sequence of sets in $\mathcal{D}$ and $A_n \subseteq A_{n+1}$ for every n then $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$.*

The intersection of two d-systems is another d-system. For a given class of subsets $\mathcal{G}$ of Ω, we let $\mathcal{D}(\mathcal{G})$ denote the smallest d-system containing $\mathcal{G}$.

Exercise 4.5 *Prove that $\mathcal{D}(\mathcal{G})$ is well-defined and is given by*

$$\mathcal{D}(\mathcal{G}) = \cap \{\mathcal{F} : \mathcal{F} \text{ is a d-system}, \mathcal{F} \supseteq \mathcal{G}\}.$$

Definition 4.6 (π-system) *A π-system is a class of subsets which is closed under finite intersections.*

Exercise 4.7 *Prove the following:*

1. *$\mathcal{G}$ is a σ-field if and only if it is both a π-system and a d-system.*
2. *(Dynkin's π-system lemma) If $\mathcal{G}$ is a π-system then $\mathcal{D}(\mathcal{G}) = \sigma(\mathcal{G})$.*

Using the previous exercise we can prove the following useful result.

Proposition 4.8 *Let μ_1 and μ_2 be two finite measures on the measurable space $(\Omega, \mathcal{A})$. Suppose $\mathcal{G} \subseteq \mathcal{A}$ is a π-system. If μ_1 and μ_2 agree on $\mathcal{G}$ and if $\mu_1(\Omega) = \mu_2(\Omega)$ then μ_1 and μ_2 agree on $\sigma(\mathcal{G})$.*

Proof. Let $\mathcal{F}$ denote the class of subsets defined by

$$\mathcal{F} = \{A \in \sigma(\mathcal{G}) : \mu_1(A) = \mu_2(A)\}.$$

We first prove that $\mathcal{F}$ is a d-system. By assumption $\Omega \in \mathcal{F}$, so we assume that $A \subseteq B$ are sets in $\mathcal{F}$ and notice that $B = A \cup (B \setminus A)$ implies

$$\begin{aligned} \mu_1(B) &= \mu_1(A) + \mu_1(B \setminus A), \text{ and likewise} \\ \mu_2(B) &= \mu_2(A) + \mu_2(B \setminus A). \end{aligned}$$

Since $\mu_1(A) = \mu_2(A)$ and $\mu_1(B) = \mu_2(B)$, these last two equations imply $\mu_1(B \setminus A) = \mu_2(B \setminus A)$, i.e. $B \setminus A \in \mathcal{F}$. Finally, it we assume that $A_n \subseteq A_{n+1} \subseteq ...$ is a sequence of subsets in $\mathcal{F}$ then by defining $E_1 = A_1$ and

$$E_n := A_n \setminus A_{n-1} \text{ for } n = 2, 3, ...$$

we have a sequence of disjoint measurable sets $(E_n)_{n=1}^{\infty}$ satisfying $\cup_{n=1}^{\infty} E_n = \cup_{n=1}^{\infty} A_n$ and

$$\mu_1 (E_n) = \mu_1 (A_n) - \mu_1 (A_{n-1}) = \mu_2 (A_n) - \mu_2 (A_{n-1}) = \mu_2 (E_n) .$$

Using the countably additivity of μ_1 and μ_2 we obtain

$$\begin{aligned} \mu_1 (\cup_{n=1}^{\infty} A_n) &= \mu_1 (\cup_{n=1}^{\infty} E_n) \\ &= \sum_{n=1}^{\infty} \mu_1 (E_n) = \sum_{n=1}^{\infty} \mu_2 (E_n) \\ &= \mu_2 (\cup_{n=1}^{\infty} E_n) \\ &= \mu_2 (\cup_{n=1}^{\infty} A_n) , \end{aligned}$$

i.e. $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$. It follows that $\mathcal{F}$ is a d-system. Since $\mathcal{G} \subseteq \mathcal{F}$ and because, by assumption, $\mathcal{G}$ is a $\pi-$system we have that $\mathcal{F} \supseteq \mathcal{D}(\mathcal{G}) = \sigma(\mathcal{G})$. ■

5 Integration

One of the motivations behind measure theory is to develop a robust theory of integration. To appreciate better the advantages of the measure-theoretic approach, we first recall the Riemann integral.

5.1 Riemann integration

Recall that a function $f : [a, b] \to \mathbb{R}$ is a *step function* on $[a, b]$ if it can be written in the form

$$f(x) = \sum_{i=1}^{N} a_i 1_{(x_{i-1}, x_i)} + \sum_{i=1}^{N} b_i 1_{\{x_i\}}, \tag{4}$$

for some partition $a = x_0 < x_1 < .... < x_N = b$ of the interval $[a, b]$, and $a_i, b_i \in \mathbb{R}$ for $i = 0, 1, ..., N$. We denote the set of step function on $[a, b]$ by $\mathcal{S}[a, b]$. For a given element $f \in \mathcal{S}[a, b]$ of the form (4) we define the Riemann integral $I(f)$

$$I(f) = \sum_{i=1}^{N} a_i (x_i - x_{i-1}).$$

Exercise 5.1 *Prove that $I(f)$ is well-defined on step functions; i.e. prove that $I(f)$ is independent of the way in which it is represented in (4).*

For a general function $f : [a, b] \to \mathbb{R}$, we define

$$\begin{aligned} L(f) &= \sup \{I(g) : g \leq f \text{ and } g \in \mathcal{S}[a, b]\} \text{ and} \\ U(f) &= \inf \{I(g) : g \geq f \text{ and } g \in \mathcal{S}[a, b]\}. \end{aligned}$$

Definition 5.2 *We say a function $f : [a, b] \to \mathbb{R}$ is **Riemann integrable** if $L(f) = U(f)$ in which case we define the Riemann integral to be*

$$I(f) := \int_{[a,b]} f := L(f) = U(f).$$

Definition 5.3 (almost everywhere) *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. We say a property P holds almost everywhere (a.e.) if*

$$A := \{\omega : P(\omega) \text{ is true}\} \in \mathcal{A}$$

and $\mu(A^c) = 0.$.

Theorem 5.4 *A bounded $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if f is continuous a.e.*

There are problems with Riemann integration. To appreciate this, consider the function $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0,1] \\ 0 & \text{otherwise} \end{cases}.$$

Since $f = 0$ a.e. on $[0,1]$ we might expect that $\int_{[0,1]} f = 0$, this is not the case, however. In fact f is not even Riemann integrable. To see this recall that the rationals are dense in $\mathbb{R}$, and hence since step functions are piecewise constant on open intervals it follows that $0 = L(f) \neq U(f) = 1$. We will need an integral which is better behaved.

5.2 Measurable functions

Definition 5.5 *Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be two measurable spaces and $f : \Omega_1 \to \Omega_2$ be a function. We say f is **measurable** if for all $A \in \mathcal{A}_2$*

$$f^{-1}(A) = \{x \in \Omega_1 : f(x) \in A\} \in \mathcal{A}_1.$$

In other words a function between two measurable spaces $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ is measurable if the preimage of every $\mathcal{A}_2$-measurable set is $\mathcal{A}_1-$measurable. We will most often deal with the case where $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Lemma 5.6 *$f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable if and only if $f^{-1}((-\infty, a]) \in \mathcal{A}$ for all $a \in \mathbb{R}$.*

Proof. A full proof can be found in Schilling. The basic idea is to prove that $\mathcal{F} := \left\{B \in \mathcal{B}(\mathbb{R}) : f^{-1}(B) \in \mathcal{A}\right\} \subseteq \mathcal{B}(\mathbb{R})$. Because $\mathcal{F}$ contains $\pi(\mathbb{R})$ and (from lecture 2) we know that $\mathcal{B}(\mathbb{R}) = \sigma(\pi(\mathbb{R}))$, the result then follows. ■

The most basic functions for which we define the measure-theoretic integral is the set of simple functions.

Definition 5.7 *A function $f : \Omega \to \mathbb{R}$ is a **simple function** if there exists $c_i \in \mathbb{R}$ and $A_i \in \mathcal{A}$,$i = 1, \ldots.., n$ such that*

$$f(x) = \sum_{i=1}^{n} c_i 1_{A_i}(x)$$

Remark 5.8 *We will also consider cases where $f : \Omega \to [-\infty, \infty]$, i.e where f is allowed to take the values $\pm\infty$.*

Exercise 5.9 *Prove the following:*

1. *An simple function is measurable.*
2. *Any measurable function is simple iff it takes finitely many values.*
3. *If f and g are simple then $f \pm g, fg$ and λf are simple for any $\lambda \in \mathbb{R}$.*

4. *If f is simple then $f^+ = \max(f,0)$, $f^- = -\min(f,0)$ and $|f| = f^+ + f^-$ are simple.*

Theorem 5.10 *If $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable function, then there exists a sequence of simple functions $(f_n)_{n=1}^{\infty}$ satisfying*

1. *$|f_n(x)| \leq |f(x)|$ for all $x \in \Omega$;*
2. *$f_n(x) \to f(x)$ for all $x \in \Omega$.*

If $f \geq 0$ then we can take $f_n \geq 0$ for all n.

Proof. We prove the case $f \geq 0$, leaving the general case as an exercise. Fix $n \in \mathbb{N}$, and for $k = 0, 1, ..., n2^n - 1$ define the set

$$A_{k,n} := \left\{ x \in \Omega : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \right\} = f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n} \right) \right).$$

And also define

$$A_{n2^n,n} := \{ x \in \Omega : f(x) \geq n \} = f^{-1}([n, \infty)).$$

Since f is measurable $A_{k,n} \in \mathcal{A}$ for $k = 0, 1, ..., n2^n$; it follows that the function

$$f_n(x) := \sum_{k=0}^{n2^n} \frac{k}{2^n} 1_{A_{k,n}}(x)$$

is a simple function. By construction we know that $0 \leq f_n \leq f$, and also

$$|f_n(x) - f(x)| \leq \frac{1}{2^n} \text{ for all } x \text{ such that } f(x) \leq n.$$

It follows that $f_n(x) \uparrow f(x)$ as $n \to \infty$, for all $x \in \Omega$. ∎

Definition 5.11 *Let $(f_n)_{n=1}^{\infty} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a sequence of measurable functions. Define $\liminf f_n$ and $\limsup f_n$ by*

$$\begin{aligned}
\liminf f_n(x) &= \sup_{n\geq 1} \inf_{k \geq n} f_k(x) = \lim_{n\to\infty} \inf_{k\geq n} f_k(x) \in [-\infty, \infty], \\
\limsup f_n(x) &= \inf_{n\geq 1} \sup_{k \geq n} f_k(x) = \lim_{n\to\infty} \sup_{k\geq n} f_k(x) \in [-\infty, \infty]
\end{aligned}$$

for all $x \in \Omega$.

Lemma 5.12 *Let $(f_n)_{n=1}^{\infty} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a sequence of measurable functions, then the following are all measurable functions (from $(\Omega, \mathcal{A})$ to $([-\infty, \infty], \mathcal{B}([-\infty, \infty]))$*

1. $\inf f_n$;
2. $\liminf f_n$;

3. $\limsup f_n$.

Furthermore $\{x \in \Omega : \lim_{n\to\infty} f_n(x)$ *exists in* $\mathbb{R}\} \in \mathcal{A}$.

Proof. For the first we notice that

$$\{x : \inf f_n(x) \geq c\} = \cap_{n=1}^{\infty} \{x : f_n(x) \geq c\} \in \mathcal{A}.$$

For the second point, notice that $g_n(x) := \inf_{k\geq n} f_k(x)$ is measurable by the first part, and we have

$$\{x : \liminf f_n(x) \leq c\} = \cap_{n=1}^{\infty} \{x : g_n(x) \leq c\} \in \mathcal{A}.$$

The third part is similar. Finally we note that

$$\left\{x : \lim_{n\to\infty} f_n(x) \text{ exists in } \mathbb{R}\right\} = \{\liminf f_n > -\infty\} \cap \{\limsup < \infty\} \cap g^{-1}(\{0\}),$$

where $g := \limsup f_n - \liminf f_n$. ■

6 The Measure-theoretic integral

6.1 The basic definition of the integral

Definition 6.1 (Integration for simple functions) *If $f : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is simple, i.e. of the form $f = \sum_{i=1}^{n} c_i 1_{A_i}$ then we define*

$$\int f d\mu = \sum_{i=1}^{n} c_i \mu(A_i). \tag{5}$$

Exercise 6.2 *Check that if $f = \sum_{i=1}^{n} c_i 1_{A_i} = \sum_{j=1}^{m} d_j 1_{B_j}$ then*

$$\sum_{i=1}^{n} c_i \mu(A_i) = \sum_{j=1}^{m} d_i \mu(B_j),$$

so that (5) is well-defined.

Definition 6.3 (Integration for non-negative measurable functions) *If $f : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is non-negative and measurable then we define*

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu \in [0, \infty], \tag{6}$$

where $(f_n)_{n=1}^{\infty}$ is a sequence of simple functions with $0 \le f_n \uparrow f$.

Remark 6.4 *We proved in lecture 5 that such an approximating sequence of simple functions can always be found. There may, of course, be more than one such sequence but again the limit in (6) is independent of the choice of sequence.*

We notice that our definition allows for $\int f d\mu = \infty$. If $\int f d\mu < \infty$ however we say that f is **μ-integrable.**

Definition 6.5 (Integration for measurable functions) *If $f : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable function we say f is **μ-integrable** if $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ are both μ-integrable. We then define the integral to be*

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Definition 6.6 *For $p > 0$ we let $\mathcal{L}^p(\mu) = \mathcal{L}^p(\Omega, \mathcal{A}, \mu) := \{f : \Omega \to \mathbb{R} : |f|^p$ is μ-integrable$\}$.*

Exercise 6.7 (Properties of the integral) *Let $f, g \in \mathcal{L}^1(\mu)$.*

1. *If $A, B \in \mathcal{A}$ are disjoint and if $\int_A f d\mu := \int f 1_A d\mu$ then*

$$\int_{A\cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

2. *If* $A = \{x \in \Omega : |f(x)| < \infty\}$ *then* $A \in \mathcal{A}$ *and* $\mu(A^c) = 0$, *i.e.* $|f| < \infty$ *a.e.*

3. *The integral is linear:* $\forall a, b \in \mathbb{R}$, $\int (af + bg)\, d\mu = a \int f d\mu + b \int g d\mu$.

4. $\left|\int f d\mu\right| \leq \int |f|\, d\mu$.

5. *If* $f \geq 0$ *then* $\int f d\mu \geq 0$.

6. *If* $f \geq 0$ *and* $\int f d\mu = 0$ *then* $f = 0$ *a.e.*

7. *If* $f = g$ *a.e. then* $\int f d\mu = \int g d\mu$.

6.2 $L^p(\mu)$ spaces

If f and g are two functions in $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ which agree a.e. (i.e. $f = g$ a.e.) then we have $\int f d\mu = \int g d\mu$. From the point of view of measure-theoretic integration these functions should be treated as the same. For each $p > 0$ we define an equivalance relation $\sim$ on $\mathcal{L}^p(\mu)$ by

$$f \sim g \text{ if } f = g \text{ a.e.}$$

Definition 6.8 *We define* $L^p(\Omega, \mathcal{A}, \mu) = L^p(\mu)$ *to be the set of equivalence classes of* $\mathcal{L}^p(\mu)$ *under the equivalence relation* $\sim$.

It is important to be aware that elements of $L^p(\mu)$ are not, strictly speaking, functions any more but instead classes of functions which (pairwise) agree a.e. For most practically purposes it suffices to work with a representative function from the equivalence class. $L^p(\mu)$ has the structure of a vector space; for instance it is easily seen that

$$[\lambda f + \mu g] = \lambda[f] + \mu[g],$$

where $[f]$ is the equivalence class $\{g \in \mathcal{L}^p(\mu) : f = g \text{ a.e.}\}$.

Theorem 6.9 *If* $p \geq 1$ *then the function* $||\cdot||_p : L^p(\mu) \to [0, \infty)$ *given by*

$$||f||_p = \left(\int |f|^p\, d\mu\right)^{1/p}$$

satisfies Minkowski's inequality holds:

$$||f + g||_p \leq ||f||_p + ||g||_p \text{ for all } f, g \in L^p(\mu),$$

and furthermore $||\cdot||_p$ *is a complete norm, i.e.* $\left(L^p(\mu), ||\cdot||_p\right)$ *is a Banach space.*

6.3 The monotone convergence theorem

Theorem 6.10 (Monotone convergence theorem, MCT) *Let* $(f_n)_{n=1}^{\infty} : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be a sequence of non-negative measurable functions which are monotone in the sense that* $f_n(x) \le f_{n+1}(x)$ *for all* $x \in \Omega$ *and for all* $n \in \mathbb{N}$. *Let*

$$f(x) := \lim_{n\to\infty} f_n(x) \in [0,\infty],$$

then as $n \to \infty$

$$\int f_n d\mu \uparrow \int f d\mu \in [0,\infty].$$

Proof. For each natural number n let $\left(f_n^k\right)_{k=1}^{\infty}$ be a sequence of simple functions such that $0 \le f_n^k \le f_n^{k+1}$ for all k and $f_n^k \uparrow f_n$as $k \to \infty$. Let

$$g_k(x) := \max_{n\le k} f_n^k(x), \text{ for } k = 1,2,...$$

and notice that each g_k is a simple function. Define $g(x) = \lim_{k\to\infty} g_k(x) \in [0,\infty]$, observe that for all $n \le k$

$$f_n^k(x) \le g_k(x) \le f_k(x) \le f(x) \tag{7}$$

for all $x \in \Omega$, and let $k \to \infty$ and then $n \to \infty$ to see

$$f(x) \le g(x) \le f(x),$$

i.e. $0 \le g_k \le g_{k+1} \uparrow g = f$. By integrating (7) and using the order-preserving property of the integral we obtain for all $n \le k$

$$\int f_n^k d\mu \le \int g_k d\mu \le \int f_k d\mu.$$

Again taking limits in the order $k \to \infty$ and then $n \to \infty$ it follows that

$$\lim_{n\to\infty} \int f_n d\mu \le \lim_{k\to\infty} \int g_k d\mu = \int g d\mu = \int f d\mu \le \lim_{k\to\infty} \int f_k d\mu,$$

and the result follows. ■

7 Convergence properties of the integral

Two convergence results of the integral we use very often are Fatou's lemmas and the dominated convergence theorem.

Theorem 7.1 *Let $(f_n)_{n=1}^{\infty}$ be a sequence of non-negative measurable functions from a measure space $(\Omega, \mathcal{A}, \mu)$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then:*

1. *(Fatou's lemma) We have that*
$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu \in [0, \infty].$$
2. *(Reverse Fatou lemma) If, in addition, there exists a μ-integrable function $g : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $0 \leq f_n \leq g$ for all $n \in \mathbb{N}$, then*
$$\int \limsup_{n\to\infty} f_n d\mu \geq \limsup_{n\to\infty} \int f_n d\mu.$$

Proof. We first prove 1. For each $n \in \mathbb{N}$ we define the measurable function $g_n : \Omega \to \mathbb{R}$ by setting $g_n(x) := \inf_{k\geq n} f_k(x)$. We notice that $0 \leq g_n \leq g_{n+1}$ for every n, and g, the monotone limit of the g_n equals $\liminf f_n$ by definition. Using the monotone convergence theorem we obtain
$$\int g_n d\mu \to \int g d\mu = \int \liminf f_n d\mu \text{ as } n \to \infty.$$
On the other hand we have for every $k \geq n$
$$\int g_n d\mu \leq \int f_k d\mu,$$
and therefore
$$\int g_n d\mu \leq \liminf_{k\to\infty} \int f_k d\mu.$$
Letting $n \to \infty$ in this last inequality gives the result
$$\int g d\mu \leq \liminf_{k\to\infty} \int f_k d\mu.$$
2 follows from 1 by noting that the sequence $(g - f_n)_{n=1}^{\infty}$ is a non-negative sequence of measurable functions. Hence,
$$\int \liminf_{n\to\infty} (g - f_n) d\mu \leq \liminf_{n\to\infty} \int (g - f_n) d\mu = \int g d\mu + \liminf_{n\to\infty} \left(-\int f_n d\mu\right).$$
Using that $\liminf_{n\to\infty} (g - f_n) = g + \liminf_{n\to\infty} (-f_n) = g - \limsup_{n\to\infty} f_n$, and similarly that $\liminf_{n\to\infty} \left(-\int f_n d\mu\right) = -\limsup_{n\to\infty} \int f_n d\mu$ we obtain
$$\int g d\mu - \int \limsup_{n\to\infty} f_n d\mu \leq \int g d\mu - \limsup_{n\to\infty} \int f_n d\mu,$$
from which the result follows. ■

Remark 7.2 *The reverse Fatou lemma fails without the assumption that g is μ-integrable. As an exercise, try to think of a counterexample.*

Theorem 7.3 (Dominated convergence theorem) *Suppose $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions from $(\Omega, \mathcal{A}, \mu)$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that:*

1. *$f_n(x) \to f(x)$ for every $x \in \Omega$ as $n \to \infty$;*
2. *For some $\mu-$integrable g we have $|f_n(x)| \leq g(x)$ for every $x \in \Omega$ and every $n \in \mathbb{N}$.*

Then

$$\lim_{n\to\infty} \int f_n d\mu = \int \lim_{n\to\infty} f_n d\mu = \int f d\mu < \infty.$$

Proof. We first observe that $|f_n(x)| \leq g(x)$ implies $|f(x)| \leq g(x)$ for all $x \in \Omega$, and also $\int |f_n| d\mu \leq \int g d\mu$ and similarly $\int |f| d\mu \leq \int g d\mu$. For every $n \in \mathbb{N}$ we have $|f_n - f| \geq 0$ and $|f_n - f| \to 0$ as $n \to \infty$. Moreover $|f_n - f| \leq 2g$ and we may apply the reverse Fatou lemma to give

$$0 \leq \limsup_{n\to\infty} \int |f_n - f| d\mu \leq \int \limsup_{n\to\infty} |f_n - f| d\mu = 0.$$

It follows that

$$\limsup_{n\to\infty} \int |f_n - f| d\mu = \lim_{n\to\infty} \int |f_n - f| d\mu = 0,$$

and therefore

$$\lim_{n\to\infty} \left| \int (f_n - f) d\mu \right| \leq \lim_{n\to\infty} \int |f_n - f| d\mu = 0,$$

i.e. $\int f_n d\mu \to \int f d\mu$ as $n \to \infty$. ■

7.1 Stieljes integration

Suppose $F : \mathbb{R} \to \mathbb{R}$ is a function which is non-decreasing $(\forall x \leq y, F(x) \leq F(y))$ and right continuous $(\forall a \lim_{x\downarrow a} F(x) = F(a))$, then for all $-\infty < a \leq b < \infty$ we can define

$$\mu_F((a, b]) = F(b) - F(a).$$

Lemma 7.4 *There exists a unique measure (also called μ_F) defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu_F((a, b]) = F(b) - F(a)$ for all $a < b$.*

Proof (sketch). We have already seen in the construction of the Lebesgue measure that $\mathcal{S} = \{(a, b] : -\infty < a \leq b < \infty\}$ is a semi-ring of subsets of $\mathbb{R}$. It can be shown that μ_F is a pre-measure on $\mathcal{S}$ (exercise: adapt the proof in the construction of the Lebesgue measure). By Caratheodory's extension theorem

μ_F extends to a measure on $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$, which is unique since μ_F is σ-finite. ■

Conversely, given a measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu((-n, n]) < \infty$ for every $n \in \mathbb{N}$ we may define a non-decreasing right-continuous function by

$$F_\mu(x) = \begin{cases} \mu((0, x]) & x > 0 \\ 0 & x = 0 \\ -\mu((x, 0]) & x < 0 \end{cases}.$$

There is a one-to-one correspondence between the set of such functions and the measures μ_F.

Example 7.5 *The canonical examples of such functions arise naturally as the distribution functions of random variables, i.e.*

$$F(x) = \mathbb{P}(X \leq x),$$

we will often write

$$\int f dF := \int f(x) F(dx) := \int f(x) \mu_F(dx).$$

7.2 Signed measures

Definition 7.6 *A signed measure is a countably additive set function μ on a measurable space $(\Omega, \mathcal{A})$ such that $\mu(\emptyset) = 0$.*

Note that signed measures, unlike measures, can take negative values.

Let X be a topological space. The *support* of a measure μ on $(X, \mathcal{B}(X))$ is the smallest closed subset of X which contains the following set

$$\{x \in X : \ \mu(N_x) > 0 \text{ for some neighbourhood } N_x \text{ of } x \text{ in } \mathcal{B}(X) \text{ of } x\}$$

The Hahn-Jordan decomposition tells us that we can always decompose a signed measure μ uniquely as the difference of two measures μ^+ and μ^- having disjoint support, i.e. $\mu = \mu^+ - \mu^-$. If $F = F_1 - F_2$ is the difference of two non-decreasing, right-continuous functions then we define

$$\int f dF := \int f dF_1 - \int f dF_2.$$

Definition 7.7 *A function $F : [a, b] \to \mathbb{R}$ has bounded variation if*

$$Var(F; [a, b]) = \sup_{\mathcal{D} \in \mathcal{D}[a,b]} \sum_{t_i \in \mathcal{D}} |F(t_{i+1}) - F(t_i)| < \infty,$$

where $\mathcal{D}[a, b]$ is the set of partitions of $[a, b]$.

Theorem 7.8 *Suppose $F : [a, b] \to \mathbb{R}$, then F can be written as the difference of two non-decreasing, right-continuous functions if and only if F has bounded variation.*

8 Absolutely continuous measures

Suppose that $f \in L^1(\Omega, \mathcal{A}, \mu)$. For $A \in \mathcal{A}$ we continue to write

$$\int_A f d\mu = \int f 1_A d\mu$$

Theorem 8.1 *Suppose that $f \in L^1(\Omega, \mathcal{A}, \mu)$ then*

$$\int_A f d\mu \to 0 \text{ as } \mu(A) \to 0. \tag{8}$$

Remark 8.2 *(8) means explicitly that for every $\epsilon > 0$ there exists $\delta > 0$ such that for any $A \in \mathcal{A}$ with $\mu(A) \le \delta$ we have $\left|\int_A f d\mu\right| \le \epsilon$.*

Proof. Fix $\epsilon > 0$ and for every $n \in \mathbb{N}$ define

$$C_n = \{x \in \Omega : |f(x)| \le n\} \in \mathcal{A}.$$

We let $f_n := f 1_{C_n}$ and notice that $|f_n| \uparrow |f|$ monotonically as $n \to \infty$. From the monotone convergence theorem we have that

$$\int |f_n| \, d\mu \uparrow \int |f| \, d\mu < \infty \text{ as } n \to \infty.$$

We choose $N \in \mathbb{N}$ such that $\int |f_n| \, d\mu \ge \int |f| \, d\mu - \frac{\epsilon}{2}$ for all $n \ge N$. Let $\delta := \frac{\epsilon}{2N}$ and suppose $A \in \mathcal{A}$ is such that $\mu(A) \le \delta$ then

$$\begin{aligned} \left|\int_A f d\mu\right| &\le \int_A |f| \, d\mu \\ &\le \int_A |f| \, d\mu - \int_A |f_N| \, d\mu + \int_A |f_N| \, d\mu \\ &\le \frac{\epsilon}{2} + N \frac{\epsilon}{2N} \\ &= \epsilon. \end{aligned}$$

■

If $f \ge 0$ is a measurable function in $L^1(\mu)$, define a set function $\nu : \mathcal{A} \to [0, \infty)$ by

$$\nu(A) := \int_A f d\mu. \tag{9}$$

Lemma 8.3 *ν define by (9) is a measure on $(\Omega, \mathcal{A})$.*

Proof. Clearly $\nu(\emptyset) = 0$, so the only thing to check is that ν is countably additive. Let $(A_n)_{n=1}^{\infty}$ be a sequence of pairwise disjoint sets in $\mathcal{A}$. Define

$$g_n = 1_{\cup_{i=1}^n A_i} = \sum_{i=1}^n 1_{A_i}$$

and notice that $0 \leq g_n \leq g_{n+1}$ for every n, and that each g_n is measurable (in fact, simple). Using that $f \geq 0$ we also have that $fg_n \uparrow fg$ monotonically as $n \to \infty$ and hence by MCT we have

$$\int fg_n d\mu \uparrow \int fg d\mu \text{ as } n \to \infty.$$

On the other hand

$$\int fg_n d\mu = \sum_{i=1}^{n} \int f 1_{A_i} d\mu = \sum_{i=1}^{n} \nu(A_i),$$

and

$$\int fg d\mu = \int f 1_{\cup_{i=1}^{\infty} A_i} d\mu = \nu(\cup_{i=1}^{\infty} A_i).$$

It follows that

$$\nu(\cup_{i=1}^{\infty} A_i) = \lim_{n \to \infty} \sum_{i=1}^{n} \nu(A_i) = \sum_{i=1}^{\infty} \nu(A_i).$$

■

It follows from Theorem 8.1 that $\mu(A) = 0$ implies $\nu(A) = 0$, in other words the set of μ-null sets is contained in the set of ν-null sets. This characterises ν as being absolutely continuous with respect to μ,and we write $\nu << \mu$.

Theorem 8.4 (Radon-Nikodym theorem) *If μ and ν are σ-finite measures on $(\Omega, \mathcal{A})$ then $\nu << \mu$ if and only if there exists $f \in L^1(\mu)$ such that $f \geq 0$ and*

$$\nu(A) = \int_A f d\mu \text{ for all } A \in \mathcal{A}.$$

The function f is called the Radon-Nikodym derivative of ν with respect to μ and is sometimes written $\frac{d\nu}{d\mu}$. Notice that

$$\nu(A) = \int_A \frac{d\nu}{d\mu} d\mu = \int_A d\nu,$$

so that formally '$\frac{d\nu}{d\mu} d\mu = d\nu$'.

Exercise 8.5 *Show that if $\tau << \nu$ and $\nu << \mu$ then $\tau << \mu$ and*

$$\frac{d\tau}{d\mu} = \frac{d\tau}{d\nu} \frac{d\nu}{d\mu}.$$

Definition 8.6 *If μ and ν are measures on $(\Omega, \mathcal{A})$ such that $\mu << \nu$ and $\nu << \mu$ we say that μ and ν are equivalent measures.*

Remark 8.7 *Note that μ and ν are equivalent if and only if μ and ν have the same null sets, i.e. if $\mu(A) = 0$ if and only if $\nu(A) = 0$.*

Exercise 8.8 *If μ and ν are equivalent prove that*

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}$$

9 Miscellaneous results

Suppose f is a measurable function between two measurable spaces $(\Omega, \mathcal{A})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We may define the class of subsets

$$\sigma(f) := \left\{ f^{-1}(B) : B \in \mathcal{B}(\mathbb{R}) \right\} \subseteq \mathcal{A}.$$

It is a simple exercise to show that $\sigma(f)$ is a σ-field. We call $\sigma(f)$ the σ-field generated by f.

Proposition 9.1 (Doob's lemma) *Suppose f and g are measurable functions between the measurable spaces $(\Omega, \mathcal{A})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The following are equivalent:*

1. $\sigma(f) \subseteq \sigma(g)$;
2. *There exists a measurable function $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $f = h \circ g$*

Proof. Suppose first that 2 holds, then for any $B \in \mathcal{B}(\mathbb{R})$ we have that $A := h^{-1}(B) \in \mathcal{B}(\mathbb{R})$ and also

$$f^{-1}(B) = g^{-1}\left(h^{-1}(B)\right) = g^{-1}(A),$$

which shows that $\sigma(f) \subseteq \sigma(g)$. Now assume 1 and prove 2. We first assume $f = 1_A$ for some $A \in \mathcal{A}$. In this case since $\sigma(f) = \{\emptyset, A, A^c, \Omega\}$ and since $\sigma(f) \subseteq \sigma(g)$ we have that

$$A = g^{-1}(B) \text{ for some } B \in \mathcal{B}(\mathbb{R}).$$

It follows that $f = 1_A = h \circ g$ where $h = 1_B$. The same idea extends to the case where f is a simple fucntion, and then , by approximation, to the case where f is a non-negative measurable function. The case where f is a general non-negative function can then be dealt with by considering positive an negative parts. ■

It is useful to think about $\sigma-$field as describing information content. From this point of view $\sigma(f)$ represents the information described by the measurable function f.

Lemma 9.2 *Suppose f :$(\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable, $\sigma(f)$ is the smallest sub-σ-field of $\mathcal{A}$ such that f :$(\Omega, \sigma(f)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is still measurable.*

Proof. We have already seen that $\sigma(f)$ is a σ-field. By definition $f^{-1}(B) \in \sigma(f)$ for all $B \in \mathcal{B}(\mathbb{R})$ so f is certainly measurable with respect to $\sigma(f)$. On the other hand, if $\mathcal{F} \subseteq \mathcal{A}$ is any other σ-field for which $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable then $f^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}(\mathbb{R})$, i.e. $\mathcal{F}$ contains $\sigma(f)$. ■

9.1 Inequalities

We have already seen Minkowski's ineqaulity. which for any $f, g \in L^p(\mu)$ with $p \geq 1$ shows the triangle inequality for the L^p norm, i.e.

$$||f+g||_p \leq ||f||_p + ||g||_p .$$

Two further important inequalities (which you will be asked to prove on the problems sheet) are:

1. **Hölder's inequality**: Suppose that $f \in L^p(\mu)$ and $g \in L^q(\mu)$ where $p, q \geq 1$ satisfy $1/p + 1/q = 1$ then $fg \in L^1(\mu)$ and

   $$||fg||_1 \leq ||f||_p \, ||g||_q .$$

   When $p = q = 2$ this is called the Cauchy-Schwarz ineqaulity.

2. **Jensen's inequality:** $\phi : \mathbb{R} \to \mathbb{R}$ is convex if $\phi(\lambda x + (1-\lambda) y) \leq \lambda \phi(x) + (1-\lambda) \phi(y)$. If ϕ is convex and $f \in L^1(\mu) = L^1(\Omega, \mathcal{A}, \mu)$ is such that $\phi(\Omega) < \infty$ and $\phi(f) \in L^1(\mu)$ we have that

   $$\phi\left(\frac{1}{\mu(\Omega)} \int f d\mu\right) \leq \frac{1}{\mu(\Omega)} \int \phi(f) \, d\mu.$$

9.2 Modes of convergence

Definition 9.3 *Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions from $(\Omega, \mathcal{A}, \mu)$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We say that*

1. *$f_n \to f$ pointwise as $n \to \infty$ if for every $x \in \Omega$, $f_n(x) \to f(x)$ as $n \to \infty$;*

2. *$f_n \to f$ a.e. (almost everywhere) if $\mu\{x \in \Omega : f_n(x)$ does not converge to $f(x)$ as $n \to \infty\} = 0$;*

3. *$f_n \to f$ in L^p if $(f_n)_{n=1}^{\infty}, f \in L^p$ and $||f_n - f||_p \to 0$ as $n \to \infty$.*

Part II

Probability spaces, independence and conditional expectation

10 Probability spaces

From now on we work with a special class of measure space, namely probability spaces.

Definition 10.1 $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space if it is a measure space and* $\mathbb{P}(\Omega) = 1$. *We call* $(\Omega, \mathcal{F}, \mathbb{P})$ *a probability triple.*

We will make it a standing assumption that $(\Omega, \mathcal{F}, \mathbb{P})$ is complete. Recall that this means if $B \subseteq A$ and $\mathbb{P}(A) = 0$ then $B \in \mathcal{F}$ and $\mathbb{P}(B) = 0$. We call Ω the sample space. An element $\omega \in \Omega$ is called a sample point. The σ-field is called the family of events, in other words an event is an $\mathcal{F}$-measurable subset of Ω. We tend to think of the sample point ω as determining the outcome of some experiment. A statement S is said to be true almost surely (a.s.) or with probability 1 (w.p.1) if

$$F := \{\omega : S(\omega) \text{ is true}\} \in \mathcal{F} \text{ and } \mathbb{P}(F) = 1.$$

Lemma 10.2 *Suppose* $(F_n)_{n=1}^{\infty}$ *is a sequence of events in* $\mathcal{F}$ *such that* $\mathbb{P}(F_n) = 1$ *for every* n, *then* $\mathbb{P}(\cap_{n=1}^{\infty} F_n) = 1$.

Proof. Since $\Omega = F_n \cup F_n^c$ we must have $\mathbb{P}(F_n^c) = 0$ for every n. From Problem Sheet 1 we learn that $\mathbb{P}(\cup_{n=1}^{\infty} F_n^c) = 0$, and since $(\cup_{n=1}^{\infty} F_n^c)^c = \cap_{n=1}^{\infty} F_n$ it follows that $\mathbb{P}(\cap_{n=1}^{\infty} F_n) = 1$. ■

10.1 Random variables

Let $(\Omega, \mathcal{F})$ be a sample space together with a family of events $\mathcal{F}$ and a probability measure $\mathbb{P}$. A random variable is a measurable function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, i.e. a function $X : \Omega \to \mathbb{R}$ for which $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$.

Definition 10.3 *If* $\{Y_i : i \in I\}$ *is a collection of functions* $Y_i : \Omega \to \mathbb{R}$ *then we define* $\mathcal{Y} := \sigma(Y_i : i \in I)$ *to be the smallest* $\sigma-$*field such that* $Y_i : (\Omega, \mathcal{Y}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is measurable for every* $i \in I$.

Exercise 10.4 *Convince yourself that* $\mathcal{Y} = \sigma\left(\left\{Y_i^{-1}(B) : B \in \mathcal{B}(\mathbb{R}), i \in I\right\}\right)$.

Remark 10.5 $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is a random variable iff* $\sigma(X) \subseteq \mathcal{F}$.

Definition 10.6 *Suppose* X *is a random variable carried on* $(\Omega, \mathcal{F}, \mathbb{P})$.

1. *We define the law of X to be the set function $\mathcal{L}_X : \mathcal{B}(\mathbb{R}) \to [0,1]$ given by*

$$\mathcal{L}_X(A) = \mathbb{P}\left(X^{-1}(A)\right) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X \in A).$$

2. *We define the distribution function of X to be the function $F_X : \mathbb{R} \to [0,1]$ defined by*

$$F_X(x) = \mathcal{L}_X((-\infty, x]) = \mathbb{P}(X \leq x).$$

Remark 10.7 *$\mathcal{L}_X$ defined a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; it is just the Lebesgue-Stieltjes measure associated with the distribution function F.*

Lemma 10.8 *$F_X : \mathbb{R} \to [0,1]$ is non-decreasing and right-continuous. Furthermore*

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

Proposition 10.9 *Let $F : \mathbb{R} \to [0,1]$ be a non-decreasing and right-continuous function such that $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$. Then there exists a random variable, which has distribution function F, on the probability space:*

$$(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), Leb).$$

Proof. Let $\omega \in [0,1]$ and define

$$X(\omega) := \inf\{x : F(x) \geq \omega\} = \sup\{y : F(y) < \omega\}.$$

By definition $\omega \leq F(c)$ implies $X(\omega) \leq c$. On the other hand, if $z > X(\omega)$ then $F(z) \geq \omega$, hence right-continuity gives $F(X(\omega)) \geq \omega$ and consequently $X(\omega) \leq c$ implies $\omega \leq F(X(\omega)) \leq F(c)$. We have shown that

$$\{\omega : \omega \leq F(c)\} = \{\omega : X(\omega) \leq c\},$$

from which it follows that $\mathbb{P}(X \leq c) = \mathbb{P}(\{\omega : \omega \leq F(c)\}) =Leb(\{\omega : \omega \leq F(c)\}) = F(c)$. ■

11 Convergence of random variables

Let $(X_n)_{n=1}^{\infty}, X$ be random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. We have introduced the notion of almost sure convergence $X_n \overset{\text{a.s.}}{\to} X$ and convergence in probability $X_n \overset{\mathbb{P}.}{\to} X$. We now explore the relationship between different modes of convergence.

Lemma 11.1 *If $X_n \overset{a.s.}{\to} X$ as $n \to \infty$ then $X_n \overset{\mathbb{P}.}{\to} X$ as $n \to \infty$.*

Proof. Fix $\epsilon > 0$ and let $A_n := \{|X_n - X\ | > \epsilon\}$. By applying the reverse Fatou lemma to the bounded, non-negative $\mathcal{F}$-measurable functions 1_{A_n} we obtain

$$0 = \int \limsup_{n\to\infty} 1_{A_n} d\mathbb{P} \geq \limsup_{n\to\infty} \int 1_{A_n} d\mathbb{P} = \limsup_{n\to\infty} \mathbb{P}(A_n) \geq 0$$

where in the first equality we have used that $X_n \overset{\text{a.s.}}{\to} X$. It follows that $\mathbb{P}(A_n) \to 0$ as $n \to \infty$, i.e. $X_n \overset{\mathbb{P}.}{\to} X$ as $n \to \infty$. ■

The converse is not true. What is true, however, is that if $X_n \overset{\mathbb{P}.}{\to} X$ as $n \to \infty$ then there exists a subsequence $(X_{n_k})_{k=1}^{\infty}$ such that $X_{n_k} \overset{\text{a.s.}}{\to} X$ as $k \to \infty$.

Definition 11.2 *Let $p > 0$. If $(X_n)_{n=1}^{\infty}, X$ are in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ then we say $X_n \overset{L^p.}{\to} X$ as $n \to \infty$ if $||X_n - X||_p \to 0$ as $n \to \infty$.*

$X_n \overset{L^p}{\to} X$ implies $X_n \overset{\mathbb{P}.}{\to} X$ (we will prove this later). The converse again fails but, similarly to a.s.convergence, if $(X_n)_{n=1}^{\infty}, X$ are in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \overset{\mathbb{P}.}{\to} X$ then there exists a subsequence along which we have L^p-convergence.

11.1 Convergence in distribution

Definition 11.3 *Suppose $(X_n)_{n=1}^{\infty}, X$ are random variables, not necessarily defined on the same probability space, with distribution fucntions $(F_n)_{n=1}^{\infty}$, resp. F. We say X_n converges to X in distribution as $n \to \infty$, and write $X_n \overset{D.}{\to} X$ if for every $f \in C_b(\mathbb{R})$ (=continuous, bounded functions from $\mathbb{R}$ to $\mathbb{R}$) we have*

$$\int f dF_n \to \int f dF \text{ as } n \to \infty.$$

We sometimes write $F_n \overset{D.}{\to} F$.

Remark 11.4 *If $(X_n)_{n=1}^{\infty}, X$ are defined on the same probability space then $X_n \overset{D.}{\to} X$ if*

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \text{ as } n \to \infty.$$

Convergence in distribution is in some sense the weakest notion of convergence of random variables. It may be shown that $X_n \overset{\mathbb{P}}{\to} X$ implies $X_n \overset{D.}{\to} X$, but the converse is not true.

Proposition 11.5 *Suppose* $(X_n)_{n=1}^{\infty}$, X *and* $(F_n)_{n=1}^{\infty}$, F *are as above then the following are equivalent:*

1. $X_n \xrightarrow{D.} X$
2. $F_n(x) \rightarrow F(x)$ *for every point* x *at which* F *is continuous.*

Remark 11.6 *The restriction to continuity points of* F *is necessary, e.g. if* $X_n := c_n \downarrow c =: X$ *then* $F_n(x) \rightarrow F$

11.2 Characteristic functions

If X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ the characteristic function (CF) $\phi : \mathbb{R} \rightarrow \mathbb{C}$ of X is given by

$$\phi(t) = \mathbb{E}\left[e^{itX}\right] = \int e^{itx} dF(x) = \int e^{itx} \mu_X(dx),$$

where μ_X is the law of X. The characteristic function is well-defined since $|\phi(t)| = \left|\int e^{itx} dF(x)\right| \leq \int \left|e^{itx}\right| dF(x) = 1$. It is also continuous (exercise: prove it using DCT). ϕ determines F uniquely, and indeed there is an explicit inversion formula (due to Lévy).

Theorem 11.7 (Lévy's continuity theorem) *Suppose* $(X_n)_{n=1}^{\infty}$ *is a sequence of r.v.s with characteristic functions* $(\phi_n)_{n=1}^{\infty}$.

1. *If* X *is a r.v. such that* $X_n \xrightarrow{D.} X$ *as* $n \rightarrow \infty$ *and* ϕ *is the characteristic function of* X *then for every* t, $\phi_n(t) \rightarrow \phi(t)$ *as* $n \rightarrow \infty$.
2. *If, for every* t, $\phi_n(t) \rightarrow \phi(t)$ *as* $n \rightarrow \infty$ *and* ϕ *is continuous at* $t = 0$ *then there exists a random variable* X *with char. function* ϕ *such that* $X_n \xrightarrow{D.} X$ *as* $n \rightarrow \infty$

Proof. See Williams. ∎

12 Independence, the weak law of large numbers and the central limit theorem

The following definition is central to probability.

Definition 12.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{G}_i : i = 1, 2, ...\}$ a collection of sub σ-fields of $\mathcal{F}$. We say $\{\mathcal{G}_i : i = 1, 2, ...\}$ are **independent** if for every sequence of sets $(G_i)_{i=1}^{\infty}$ such that $G_i \in \mathcal{G}_i$ for every i, and every finite collection of distinct indices $i_1, ..., i_n \in \mathbb{N}$ the following equality holds:*

$$\mathbb{P}\left(G_{i_1} \cap ... \cap G_{i_n}\right) = \prod_{k=1}^{n} \mathbb{P}\left(G_{i_k}\right).$$

We then say a collection of random variables $(X_i)_{i=1}^{\infty}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if the σ-fields $(\sigma(X_i))_{i=1}^{\infty}$ are independent. Similarly, a collection of events $(A_i)_{i=1}^{\infty}$ in $\mathcal{F}$ are independent if the σ-fields $(\sigma(A_i))_{i=1}^{\infty} = (\{\emptyset, A_i, A_i^c, \Omega\})_{i=1}^{\infty}$ are independent.

Exercise 12.2 *If $(X_i)_{i=1}^{n}$ are independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $f_i : \Omega \to \mathbb{R}$ with $i = 1, ..., n$ are bounded measurable functions prove that*

$$\mathbb{E}\left[\prod_{k=1}^{n} f_i(X_i)\right] = \prod_{k=1}^{n} \mathbb{E}\left[f_i(X_i)\right]$$

12.1 The weak law of large numbers (WLLN) and central limit theorem (CLT)

We recall some basic facts. First, if X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with characteristic function $\phi(t) = \mathbb{E}\left[e^{itX}\right]$, then if $\mathbb{E}\left[X^k\right] < \infty$ for $k = 1, ..., n$ we have the Taylor expansion

$$\phi(t) = \sum_{k=0}^{n} \frac{(it)^k \mathbb{E}\left[X^k\right]}{k!} + R_n(t),$$

where $R_n(t)$ is $o(t^n)$ as $t \to 0$, i.e. $t^{-n} R_n(t) \to 0$ as $t \to 0$. Second, we recall the classical result that for any sequence of complex numbers $z_n \to z \in \mathbb{C}$ as $n \to \infty$ we have

$$\lim_{n\to\infty} \left(1 + \frac{z_n}{n}\right)^n = e^z.$$

Lemma 12.3 *Suppose $(X_n)_{n=1}^{\infty}$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \xrightarrow{D} X$ as $n \to \infty$. If X is constant a.s. then $X_n \xrightarrow{\mathbb{P}} X$ as $n \to \infty$.*

Proof. Suppose $X = c$ a.s. By considering $X_n - c$ we may assume that $c = 0$. Then, for $\epsilon > 0$ we let $f_\epsilon \in C_b(\mathbb{R})$ be the bounded continuous function defined by

$$f_\epsilon(x) = \frac{|x|}{\epsilon} 1_{[-\epsilon,\epsilon]}(x) + 1_{[-\epsilon,\epsilon]^c}(x).$$

Since $X_n \xrightarrow{D} X = 0$ as $n \to \infty$ we have

$$\lim_{n\to\infty} \mathbb{E}\left[f_\epsilon\left(X_n\right)\right] = 0.$$

On the other hand, $f_\epsilon\left(X_n\right) \geq 1_{[-\epsilon,\epsilon]^c}\left(X_n\right)$ and hence

$$0 \leq \mathbb{P}\left(\left|X_n - X\right| \geq \epsilon\right) = \mathbb{P}\left(\left|X_n\right| \geq \epsilon\right) = \mathbb{E}\left[1_{[-\epsilon,\epsilon]^c}\left(X_n\right)\right] \leq \mathbb{E}\left[f_\epsilon\left(X_n\right)\right] \to 0$$

as $n \to \infty$. ■

Definition 12.4 *We say a random variable X has finite mean if $\mu = \mathbb{E}[X] < \infty$ and, in this case X has finite variance if $\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right] < \infty$.*

Theorem 12.5 (WLLN) *Suppose $\left(X_n\right)_{n=1}^\infty$ is a sequence of independent and identically distributed (i.i.d) random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with finite mean μ, then*

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu \text{ as } n \to \infty.$$

Proof. Let $S_n := \sum_{i=1}^n X_i$. We compute the characteristic function $\phi_{\frac{S_n}{n}}(t)$ using the i.i.d assumption

$$\begin{aligned}
\phi_{\frac{S_n}{n}}(t) &= \mathbb{E}\left[e^{it\frac{S_n}{n}}\right] \\
&= \mathbb{E}\left[\prod_{j=1}^n e^{itX_j/n}\right] \\
&= \prod_{j=1}^n \mathbb{E}\left[e^{itX_j/n}\right] \quad \text{(independence)} \\
&= \prod_{j=1}^n \phi_{X_1}\left(\frac{t}{n}\right) \quad \text{(identically distributed)} \\
&= \phi_{X_1}\left(\frac{t}{n}\right)^n \\
&= \left(1 + \frac{i\mu t}{n} + o\left(\frac{1}{n}\right)\right)^n \\
&\to e^{i\mu t} \text{ as } n \to \infty.
\end{aligned}$$

$\phi(t) = e^{it\mu}$ is the characteristic function of the constant random variable $X = \mu$. Using Levy's continuity theorem we deduce that $\frac{1}{n}S_n \xrightarrow{D} \mu$ as $n \to \infty$ which, by the previous lemma, gives that $\frac{1}{n}S_n \xrightarrow{\mathbb{P}} \mu$ as $n \to \infty$. ■

Theorem 12.6 (CLT) *Suppose $\left(X_n\right)_{n=1}^\infty$ is a sequence of independent and identically distributed (i.i.d) random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with finite mean μ and finite variance σ^2. Let $S_n := \sum_{i=1}^n X_i$, then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z \ ,$$

where $Z \sim N(0,1)$ has the distribution of a standard normal random variable.

Proof. By replacing X_i by $\sigma^{-1}(X_i - \mu)$ we may assume $\mu = 0$ and $\sigma = 1$. We compute the characteristic function $\phi_{\frac{S_n}{\sqrt{n}}}(t)$ to give

$$\begin{aligned}
\phi_{\frac{S_n}{\sqrt{n}}}(t) &= \mathbb{E}\left[e^{\frac{it}{\sqrt{n}}(X_1+X_2+\ldots+X_n)}\right] \\
&= \prod_{j=1}^{n} \mathbb{E}\left[e^{\frac{it}{\sqrt{n}}X_j}\right] \\
&= \phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n^{3/2}}\right)\right)^n \\
&\to e^{-t^2/2} = \phi(t).
\end{aligned}$$

ϕ is the characteristic function of the standard normal distribution. Since ϕ is continuous we may use by Levy's continuity theorem to deduce that $\frac{S_n}{\sqrt{n}} \xrightarrow{D} Z$ as $n \to \infty$. ■

13 Further consequences of independence

We recall that for a sequence of events $(A_n)_{n=1}^{\infty}$ we define

$$\begin{aligned}\limsup A_n &= \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k \\ &= \{A_n \text{ infinitely often}\}.\end{aligned}$$

Theorem 13.1 (Borel-Cantelli lemmas) *Let $(A_n)_{n=1}^{\infty}$ be a sequence of events on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

1. *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\limsup A_n) = 0$.*
2. *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and $(A_n)_{n=1}^{\infty}$ are independent then $\mathbb{P}(\limsup A_n) = 1$.*

Proof. For 1 let $B_n = \cup_{k=n}^{\infty} A_k$, then $B_n \supseteq B_{n+1}$ for all n and by the continuity properties of the probability measure we have

$$\mathbb{P}(B_n) \to \mathbb{P}(\cap_{n=1}^{\infty} B_n) = \mathbb{P}(\limsup A_n)$$

as $n \to \infty$. On the other hand,

$$\mathbb{P}(B_n) = \mathbb{P}(\cup_{k=n}^{\infty} A_k) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k) \to 0$$

as $n \to \infty$ because $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. It follows that $\mathbb{P}(\limsup A_n) = 0$.

For part 2 it suffices to show $\mathbb{P}((\limsup A_n)^c) = \mathbb{P}(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k^c) = 0$. Let $E_n := \cap_{k=n}^{\infty} A_k^c$ and notice by continuity that

$$\begin{aligned}\mathbb{P}(E_n) &= \lim_{N\to\infty} \mathbb{P}\left(\cap_{k=n}^{N} A_k^c\right) \\ &= \lim_{N\to\infty} \prod_{k=n}^{N} \mathbb{P}(A_k^c) \quad \text{(independence)} \\ &= \lim_{N\to\infty} \prod_{k=n}^{N} (1 - \mathbb{P}(A_k)) \\ &\leq \lim_{N\to\infty} \prod_{k=n}^{N} \exp(-\mathbb{P}(A_k)) \quad (\text{using } 1 - x \leq e^{-x} \text{ for } x \in (0,1)) \\ &= \lim_{N\to\infty} \exp\left(-\sum_{k=n}^{N} \mathbb{P}(A_k)\right) \\ &= 0\end{aligned}$$

using $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. It follows that $\mathbb{P}(\cup_{n=1}^{\infty} E_n) = \mathbb{P}((\limsup A_n)^c) = 0$ and the result is proved. ■

We illustrate the use of the Borel-Centelli lemmas with two examples.

Example 13.2 *Suppose* $X_n \xrightarrow{\mathbb{P}} X$ *as* $n \to \infty$ *then there exists a subsequence* $(X_{n_k})_{k=1}^{\infty}$ *such that* $X_{n_k} \xrightarrow{a.s} X$ *as* $k \to \infty$. *To see this for each* $k = 1, 2, 3....$*we choose* $n_k \in \mathbb{N}$ *such that for all* $n \geq n_k$

$$\mathbb{P}\left(|X_n - X| \geq \frac{1}{k}\right) \leq \frac{1}{k^2}.$$

This is possibly since $X_n \xrightarrow{\mathbb{P}} X$.*We then have*

$$\sum_{k=1}^{\infty} \mathbb{P}\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

By the Borel-Cantelli lemmas this implies $\mathbb{P}\left(|X_{n_k} - X| \geq \frac{1}{k} \text{ infinitely often }\right) = 0$, *which just says that* $X_{n_k} \xrightarrow{a.s} X$ *as* $k \to \infty$.

Example 13.3 *As another example, we record that the strong law of large numbers (SLLN) states that if* $(X_n)_{n=1}^{\infty}$ *are i.i.d and if* $\mathbb{E}[|X_1|] < \infty$ *then*

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{a.s} \mu = \mathbb{E}[X_1] \text{ as } n \to \infty.$$

(Note: above we have proved the weak law, which shows convergence in probability). We cannot relax the assumption $\mathbb{E}[|X_1|] < \infty$. *To see this suppose* $S_n := \sum_{i=1}^{n} X_i$ *and assume that* $\frac{S_n}{n}$ *converges a.s. as* $n \to \infty$, *then*

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \xrightarrow{a.s} 0.$$

This means that if $A_n := \{|X_n| \geq n\}$ *then* $\mathbb{P}(A_n \text{ infinitely often}) = 0$. *However, since* $(X_n)_{n=1}^{\infty}$ *are independent it follows that the events* $(A_n)_{n=1}^{\infty}$ *are independent and from the second part of the Borel-Cantelli lemmas we must have*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

(otherwise $\mathbb{P}(A_n \text{ infinitely often}) = 1$, *which we know to be false). Finally we*

use the fact that $(X_n)_{n=1}^{\infty}$ *are identically distributed to give that*

$$\begin{aligned}
\mathbb{E}[|X_1|] &\leq \sum_{k=0}^{\infty}(k+1)\,\mathbb{P}(k \leq |X_1| < k+1) \\
&\leq 1 + \sum_{k=1}^{\infty} k\mathbb{P}(k \leq |X_1| < k+1) \\
&= 1 + \sum_{k=1}^{\infty}\sum_{j=1}^{k}\mathbb{P}(k \leq |X_j| < k+1) \\
&= 1 + \sum_{j=1}^{\infty}\sum_{k=j}^{\infty}\mathbb{P}(k \leq |X_j| < k+1) \\
&= 1 + \sum_{j=1}^{\infty}\mathbb{P}(|X_j| \geq j) \\
&= 1 + \sum_{j=1}^{\infty}\mathbb{P}(A_j) < \infty.
\end{aligned}$$

14 Conditional expectation

Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Our aim is in this section is to make sense of the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ given a sub σ-field $\mathcal{G}$ of $\mathcal{F}$. To motivate the definition, we consider the simple example of two discrete random variables X and Z taking values in the sets $\{x_1, ..., x_n\}$ and $\{z_1, ..., z_m\}$ respectively. From Bayes' theorem, we understand that the conditional probability is given by

$$\mathbb{P}(X = x_i | Z = z_j) = \frac{\mathbb{P}(X = x_i, Z = z_j)}{\mathbb{P}(Z = z_j)}, \text{ if } \mathbb{P}(Z = z_j) \neq 0.$$

We can then define the conditional expectation of X given the event $\{Z = z_j\}$ by

$$\mathbb{E}[X | Z = z_j] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i | Z = z_j).$$

We want to understand these notions for general random variables. One problem is that the event $\{Z = z_j\}$may have zero probability. A route around this is to define the conditional expectation of X given Z as a new random variable$\mathbb{E}[X|Z]$. To see the features this random variable should have, we define

$$\mathbb{E}[X|Z] = Y = \sum_{j=1}^{m} \mathbb{E}[X | Z = z_j] 1_{\{Z=z_j\}} =: \sum_{j=1}^{m} y_j 1_{\{Z=z_j\}},$$

note that Y is constant on subsets where Z is constant, hence Y is measurable w.r.t. $\mathcal{G} := \sigma(Z)$. For any $j = 1, ..., m$ a simple calculation gives

$$\begin{aligned} \mathbb{E}\left[Y 1_{\{Z=z_j\}}\right] &= y_j \mathbb{P}(Z = z_j) \\ &= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i, Z = z_j) \\ &= \mathbb{E}\left[X 1_{\{Z=z_j\}}\right], \end{aligned}$$

and since $\mathcal{G} = \sigma(\{Z = z_1\}, ..., \{Z = z_m\})$ it is follows that $\mathbb{E}[Y 1_G] = \mathbb{E}[X 1_G]$ for all $G \in \mathcal{G}$.

14.1 Definition of conditional expectation

Theorem 14.1 (Kolmogorov) *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and suppose $\mathcal{G}$ is a sub $\sigma-$field of $\mathcal{F}$. There exists a random variable Y such that:*

1. *Y is $\mathcal{G}$-measurable;*
2. $\mathbb{E}[|Y|] < \infty$;
3. $\mathbb{E}[Y 1_G] = \mathbb{E}[X 1_G]$ *for all $G \in \mathcal{G}$.*

Moreover, if Y' is any other random variable which satisfies 1,2 and 3 then $Y = Y'$ almost surely.

Remark 14.2 *The random variable Y is called the conditional expectation of X given $\mathcal{G}$ and is written $\mathbb{E}[X|\, Z]$; it is unique up to a.s. equality. We use the notation $\mathbb{E}[X|\, \mathcal{G}]$. When $Z_1, ..., Z_n$ is a collection of random variables we denote by $\mathbb{E}[X|\, Z_1, ..., Z_n]$ the conditional expectation of X given $\sigma(Z_1, ..., Z_n)$.*

Proof. We first prove the uniqueness statement. Suppose that Y, $Y' \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ are two random variables satisfying properties 1,2 and 3 of the statement. Then for every $G \in \mathcal{G}$ we have

$$\mathbb{E}[Y 1_G] - \mathbb{E}[Y' 1_G] = \mathbb{E}[(Y - Y') 1_G] = \mathbb{E}[X 1_G] - \mathbb{E}[X 1_G] = 0.$$

For every n the set $G_n := \left\{Y - Y' > n^{-1}\right\} \in \mathcal{G}$ since Y, Y' are $\mathcal{G}$ −measurable. It follows that

$$\frac{1}{n}\mathbb{P}(G_n) \leq \mathbb{E}[(Y - Y') 1_{G_n}] = 0,$$

which implies $\mathbb{P}(G_n) = 0$ for every n. Hence

$$\mathbb{P}(\cup_{n=1}^{\infty} G_n) = \mathbb{P}(\{Y - Y' > 0\}) = 0,$$

by the same argument we have $\mathbb{P}(\{Y - Y' < 0\}) = 0$, i.e. $Y = Y'$ a.s.

We prove the existence of Y in two steps. First, we assume that $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and we identify X with an equivalence class (also called X) in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Since $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space under the inner product

$$\langle X_1, X_2 \rangle := \mathbb{E}[X_1 X_2]$$

and since $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed linear subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$, the orthogonal projection theorem for Hilbert spaces gives a unique $Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ such that

$$\begin{aligned} \|X - Y\|_2 &= \inf\left\{\|X - Z\|_2 : Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})\right\}, \text{ and} \\ \langle X - Y, Z \rangle &= 0 \text{ for all } Z \in L^2(\Omega, \mathcal{G}, \mathbb{P}). \end{aligned} \tag{10}$$

Using (19) we let $Z = 1_G$ for $G \in \mathcal{G}$ to obtain

$$\mathbb{E}[X 1_G] = \mathbb{E}[Y 1_G],$$

hence Y has properties 1,2 and 3 of the statement.

Finally, we prove the existence of Y for any given $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. By considering X^+ and X^- we may assume that X is non-negative. We then introduce the bounded approximations to X given by

$$X^N = X 1_{\{X \leq N\}}.$$

We have $0 \leq X^N \leq X^{N+1} \uparrow X$ as $N \to \infty$. Moreover since X^N in bounded it is in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and hence the conditional expectations $\mathbb{E}\left[X^N \middle|\, \mathcal{G}\right] = Y^N$ exist. An easy exercise shows that $0 \leq Y^N \leq Y^{N+1}$, we define the monotone limit to be $Y := \lim_{N\to\infty} Y^N$. This limit is a $\mathcal{G}$-measurable being the limit of such random variables. To prove that $\mathbb{E}\left[X \mid \mathcal{G}\right] = Y$ we notice by the monotone convergence theorem that for every $G \in \mathcal{G}$

$$\mathbb{E}\left[X 1_G\right] = \lim_{N\to\infty} \mathbb{E}\left[X^N 1_G\right] = \lim_{N\to\infty} \mathbb{E}\left[Y^N 1_G\right] = \mathbb{E}\left[Y 1_G\right],$$

which completes the proof. ■

15 Properties of conditional expectation

Lemma 15.1 (Elementary properties) *Let X, Y be in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$.*

1. *If X is $\mathcal{G}$-measurable then $\mathbb{E}[X|\mathcal{G}] = X$ a.s.*
2. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.
3. *Conditional expectation is linear: for all $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$ a.s.*

Proof. Exercise. ■

15.1 Conditional convergence theorems

We will use the following conditional versions of the classical limit theorems.

Lemma 15.2 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with $\mathcal{G}$ a sub σ-field of $\mathcal{F}$. We have:*

1. *Conditional MCT: If $0 \leq X_n \leq X_{n+1} \uparrow X$ a.s. as $n \to \infty$ then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$ a.s. as $n \to \infty$;*
2. *Conditional Fatou: For $X_n \geq 0$ we have $\mathbb{E}\left[\liminf_{n\to\infty} X_n \middle| \mathcal{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]$.*
3. *Conditional DCT: If $|X_n| \leq Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ for all n and $X_n \to X$ a.s. as $n \to \infty$ then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ a.s. as $n \to \infty$.*

Remark 15.3 *Note that conditional expectation makes sense for any non-negative random variable without the requirement of integrability (exercise: verify the details of this)*

Proof. We prove 1, leaving 2 and 3 as an exercise. First we define for every n the $\mathcal{G}$-measurable random variables $Y_n = \mathbb{E}[X_n|\mathcal{G}]$. Second we notice that $0 \leq Y_n \leq Y_{n+1}$, to see this define the $\mathcal{G}$-measurable sets G_k by

$$G_k := \left\{ Y_n - Y_{n+1} > \frac{1}{k} \right\},$$

then if $\mathbb{P}(G_k) > 0$ we would have

$$-k^{-1}\mathbb{P}(G_k) > \mathbb{E}[(Y_{n+1} - Y_n) 1_{G_k}] = \mathbb{E}[(X_{n+1} - X_n) 1_{G_k}] \geq 0,$$

a contradiction, hence $\mathbb{P}(G_k) = 0$ for every k and thus $\mathbb{P}(\cup_k G_k) = \mathbb{P}(Y_n > Y_{n+1}) = 0$. Let $Y := \lim_{n\to\infty} Y_n$. By the MCT we have

$$\mathbb{E}[X 1_G] = \lim_{n\to\infty} \mathbb{E}[X_n 1_G] = \lim_{n\to\infty} \mathbb{E}[Y_n 1_G] = \mathbb{E}[Y 1_G] \text{ for all } G \in \mathcal{G}$$

and hence $Y = \mathbb{E}[X|\mathcal{G}]$ a.s. ■

Lemma 15.4 (conditional Jensen inequality) *Let $c : \mathbb{R} \to \mathbb{R}$ be convex and let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $c(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$ we have:*

$$c(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[c(X)|\mathcal{G}] \text{ a.s.}$$

Proof. See Williams and the problem sheets. ■

Exercise 15.5 *Apply the conditional Jensen inequality to the function $c(x) = |x|^p$ to prove for $p \geq 1$ that*

$$||\mathbb{E}[X|\mathcal{G}]||_p \leq ||X||_p .$$

15.2 Further properties of conditional expectation

Theorem 15.6 *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$*

1. *(Tower property) If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$ a.s.*
2. *(Take out what is known) If Z is bounded and $\mathcal{G}$-measurable then $\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$ a.s.*
3. *(Independence) If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$ then $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}]$ a.s. In particular if $\mathcal{H}$ is independent of $\sigma(X)$ then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.*

Proof. For 1 we notice that for all $H \in \mathcal{H}$

$$\begin{aligned} \mathbb{E}[\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]1_H] &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_H] \\ &= \mathbb{E}[X1_H], \text{ since } H \in \mathcal{G} \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{H}]1_H], \text{since } H \in \mathcal{H}, \end{aligned}$$

we obtain the tower property by a.s. uniqueness of the conditional expectation.

For 2, we assume first that $Z = 1_G$ for some $G \in \mathcal{G}$. We then have for every $H \in \mathcal{G}$

$$\begin{aligned} \mathbb{E}[ZX1_H] &= \mathbb{E}[X1_{G\cap H}] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]1_{G\cap H}] \\ &= \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]1_H]. \end{aligned}$$

A.s. uniqueness then gives $\mathbb{E}[ZX|\mathcal{G}] = Z\mathbb{E}[X|\mathcal{G}]$. This extends to simple functions by linearity, then to non-negative bounded (and $\mathcal{G}$-measurable) Z by approximation. Finally, we extend to arbitrary bounded and $\mathcal{G}$-measurable Z by considering Z^+ and Z^- and again using linearity of conditional expectation.

For the final property we take $G \in \mathcal{G}$ and $H \in \mathcal{H}$ and notice that independence of $\mathcal{H}$ and $\sigma(\sigma(X), \mathcal{G})$ gives

$$\mathbb{E}[X1_{G\cap H}] = \mathbb{E}[X1_G 1_H] = \mathbb{E}[X1_G]\mathbb{P}(H).$$

On the other hand since $Y = \mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$-measurable we can use independence again to show

$$\mathbb{E}[Y1_{G\cap H}] = \mathbb{E}[Y1_G]\,\mathbb{P}(H) == \mathbb{E}[X1_G]\,\mathbb{P}(H)\,.$$

It follows that the two finite measures $\mu_1(F) = \mathbb{E}[X1_F]$ and $\mu_2(F) = \mathbb{E}[Y1_F]$ satisfy $\mu_1(\Omega) = \mu_2(\Omega)$ and agree on the π-system:

$$\{G \cap H : G \in \mathcal{G} \text{ and } H \in \mathcal{H}\}\,.$$

By the π-system lemma they agree on the σ-field generated by this $\pi-$system, which contains $\sigma(\mathcal{G}\ ,\mathcal{H})$; article 3 then follows immediately. ∎

Part III

Stochastic processes, martingales and stochastic integration

16 Martingales and discrete-time stochastic processes

Definition 16.1 (Filtrations) *A filtration on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a family* $\{\mathcal{F}_n\}_{n=0}^{\infty}$ *of sub* $\sigma-$*fields of* $\mathcal{F}$ *such that* $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 .... \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq ... \subseteq \mathcal{F}$.

A probability space a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ together with a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ is called a *filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$. A *stochastic process* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a family of $\mathcal{F}$-measurable random variables $X = (X_n)_{n=0}^{\infty}$.

Definition 16.2 *A stochastic process* $X = (X_n)_{n=0}^{\infty}$ *on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ *is said to be* ***adapted*** *(or* $\{\mathcal{F}_n\}$*-adapted) if* X_n *is* $\mathcal{F}_n$*-measurable for every* n.

Example 16.3 *If* $X = (X_n)_{n=0}^{\infty}$ *is a stochastic process on* $(\Omega, \mathcal{F}, \mathbb{P})$ *then we call the filtration defined by* $\mathcal{F}_n = \sigma(X_1, ..., X_n)$ *the natural filtration of* X. *By definition* X *is adapted to this natural filtration.*

Definition 16.4 (Martingales) *A stochastic process* $X = (X_n)_{n=0}^{\infty}$ *on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ *is said to be a martingale if*

1. X *is* $\{\mathcal{F}_n\}$*-adapted.*

2. $\mathbb{E}[|X_n|] < \infty$, *for every* n.

3. $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n$ *a.s. for every* n.

Remark 16.5 *A supermartingale is a stochastic process satisfying 1 and 2 above, with 3 replaced by* $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$ *for all* n. *A submartingale satisfies 1 and 2 and* $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n$ *for all* n.

Example 16.6 *Suppose* $(X_n)_{n=0}^{\infty}$ *is a sequence of i.i.d. random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$ *such that* $\mathbb{E}[|X_n|] < \infty$ *and* $\mathbb{E}[X_n] = 0$ *for all* n. *If* $\mathcal{F}_0 = \{\emptyset, \Omega\}$ *and* $\mathcal{F}_n = \sigma(X_1, ..., X_n)$ *then the process* $S = (S_n)_{n=0}^{\infty}$ *defined by* $S_0 = 0$ *and*

$$S_n = \sum_{i=1}^{n} X_i \text{ for } n \geq 1$$

is a martingale. Adaptedness and integrability are easy to check here. To verify property 3 above we see that

$$\begin{aligned} \mathbb{E}\left[\left.S_{n+1}\right|\mathcal{F}_n\right] &= \mathbb{E}\left[\left.X_{n+1}+S_n\right|\mathcal{F}_n\right] \\ &= \mathbb{E}\left[X_{n+1}\right]+S_n \\ &= S_n. \end{aligned}$$

As a second example, if we are given a filtered probability space $\left(\Omega,\mathcal{F},\left\{\mathcal{F}_n\right\}_{n=0}^{\infty},\mathbb{P}\right)$ *and* $X\in\mathcal{L}^1\left(\Omega,\mathcal{F},\mathbb{P}\right)$ *then the stochastic process* $X_n:=\mathbb{E}\left[\left.X\right|\mathcal{F}_n\right]$ *defines a martingale. (Exercise : make sure you can check the details of this).*

17 Stopping times and the optional stopping theorem

Definition 17.1 *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ be a filtered probability space. A stopping time (or $\{\mathcal{F}_n\}_{n=0}^{\infty}$ −stopping time) is a function $T : \Omega \to \{0, 1, 2, ...\} \cup \{\infty\}$ such that for all $n \geq 0$ we have*

$$\{\omega : T(\omega) \leq n\} =: \{T \leq n\} \in \mathcal{F}_n. \tag{11}$$

Remark 17.2 *The condition (11) is equivalent (in discrete-time) to $\{T = n\} \in \mathcal{F}_n$ for every n.*

Theorem 17.3 *Suppose $X = (X_n)_{n=0}^{\infty}$ is a supermartingale on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ and T is an $\{\mathcal{F}_n\}_{n=0}^{\infty}$ −stopping time. Define the stopped process X^T by*

$$X_n^T = X_{T\wedge n} \text{ for } n = 0, 1, ...$$

Then X^T is again a supermartingale on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$. The same conclusion holds if 'supermartingale' is replaced by 'martingale'.

Proof. We observe that

$$X_{T\wedge n} = \sum_{k=0}^{n-1} X_k 1_{\{T=k\}} + X_n 1_{\{T\geq n\}}, \tag{12}$$

for each k the random variable $X_k 1_{\{T=k\}}$ is $\mathcal{F}_k$-measurable (from the adaptedness of X and the stopping time property), and $\{T \geq n\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$ implies that $X_{T\wedge n}$ is adapted. Integrability of $X_{T\wedge n}$ follows from

$$|X_{T\wedge n}| \leq \sum_{k=0}^{n} |X_k|$$

and the integrability of $X_1, ..., X_n$. Finally we notice that

$$\mathbb{E}\left[X_n 1_{\{T\geq n\}} \middle| \mathcal{F}_{n-1}\right] = \mathbb{E}\left[X_n | \mathcal{F}_{n-1}\right] 1_{\{T\geq n\}} \leq X_{n-1} 1_{\{T\geq n\}}.$$

Using these observations in (12) gives

$$\begin{aligned} \mathbb{E}\left[X_{T\wedge n} | \mathcal{F}_{n-1}\right] &\leq \sum_{k=0}^{n-1} X_k 1_{\{T=k\}} + X_{n-1} 1_{\{T\geq n\}} \\ &= \sum_{k=0}^{n-2} X_k 1_{\{T=k\}} + X_{n-1} 1_{\{T\geq n-1\}} \\ &= X_{T\wedge(n-1)}. \end{aligned}$$

The martingale case is similar and is left as an exercise. ■

Remark 17.4 *Note in particular that a supermartingale will satisfy* $\mathbb{E}\left[X_{T\wedge n}\right] \leq \mathbb{E}\left[X_0\right]$ *for all* n.

Theorem 17.5 (optional stopping theorem) *Suppose* $X = (X_n)_{n=0}^{\infty}$ *is a supermartingale on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n=0}^{\infty}, \mathbb{P})$ *and* T *is an* $\{\mathcal{F}_n\}_{n=0}^{\infty}$ *−stopping time. Under any of the following conditions we have that* X_T *is integrable and* $\mathbb{E}\left[X_T\right] \leq \mathbb{E}\left[X_0\right]$.

1. *T is bounded*
2. *X is bounded and $T < \infty$ a.s.*
3. *T is integrable and X has bounded increments.*

If X *is a martingale and any of the above holds then* X_T *is integrable and* $\mathbb{E}\left[X_T\right] = \mathbb{E}\left[X_0\right]$.

Proof. From the previous theorem we know that $\mathbb{E}\left[X_{T\wedge n}\right] \leq \mathbb{E}\left[X_0\right]$ for all n. If T is bounded then there exists N such that $T \leq N$ a.s. and therefore $\mathbb{E}\left[X_{T\wedge N}\right] = \mathbb{E}\left[X_T\right] \leq \mathbb{E}\left[X_0\right]$. If $T < \infty$ a.s. then $X_{T\wedge n} \to X_T$ a.s. as $n \to \infty$, the boundedness then allows us to deduce from the DCT that

$$\mathbb{E}\left[X_T\right] = \lim_{n\to\infty} \mathbb{E}\left[X_{T\wedge n}\right] \leq \mathbb{E}\left[X_0\right].$$

Finally, in the third case if the increments $|X_n - X_{n-1}| \leq K$ a.s. for every n then

$$|X_{T\wedge n}| \leq \sum_{k=1}^{T} |X_k - X_{k-1}| \leq KT \in L^1.$$

T is integrable gives $T < \infty$ a.s and hence $X_{T\wedge n} \to X_T$ a.s. as $n \to \infty$. An application of the DCT then shows the result. To recover the statement for a martingale, we notice that X being a martingale implies both X and $-X$ are supermartingales so we can apply the above result twice. ■

18 Doob's upcrossing lemma

We continue to work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0}^{\infty}, \mathbb{P})$.

Definition 18.1 *We say a stochastic process $C = (C_n)_{n=1}^{\infty}$ is previsible if C_n is $\mathcal{F}_{n-1}$-measurable for every n.*

Given a previsible process C and a submartingale (supermartinagle) $X = (X_n)_{n=0}^{\infty}$ we define a new stochastic process $C \cdot X$ by

$$(C \cdot X)_0 = 0 \text{ and } (C \cdot X)_n = \sum_{k=1}^{n} C_k \left(X_k - X_{k-1}\right),$$

and call $C \cdot X$ the submartingale (supermartingale) transform of C w.r.t. X.

Lemma 18.2 *Suppose C is a non-negative, bounded previsble process and X is a submartingale (supermartingale), then $C \cdot X$ is a submartingale (supermartingale). If X is a martingale and C is bounded previsble process then $C \cdot X$ is a martingale.*

Proof. Since X is adapted and C is previsble it follows that $(C \cdot X)_n$ is $\mathcal{F}_n-$ measurable for every n, i.e. $C \cdot X$ is adapted. Integrability of $(C \cdot X)_n$ follows from the boundedness of C and the integrability of X. For any n we have

$$\begin{aligned}
\mathbb{E}\left[(C \cdot X)_n - (C \cdot X)_{n-1} \middle| \mathcal{F}_{n-1}\right] &= \mathbb{E}\left[C_{n-1}\left(X_n - X_{n-1}\right) \middle| \mathcal{F}_{n-1}\right] \\
&= C_{n-1}\mathbb{E}\left[\left(X_n - X_{n-1}\right) \middle| \mathcal{F}_{n-1}\right].
\end{aligned}$$

If C is non-negative then this is ≥ 0 or ≤ 0 according to whether X is a sub- or super-martingale. If X is martingale then it $= 0$ without the non-negativity assumption on C. ■

For any interval $[a, b] \subset \mathbb{R}$ we define two sequences of stopping times as follows:

$$\sigma_1 = \min\{n \geq 0 : X_n \leq a\}, \ \tau_1 = \min\{n \geq \sigma_1 : X_n \geq b\}$$

and then inductively by

$$\sigma_{n+1} = \min\{n \geq \tau_n : X_n \leq a\}, \ \tau_{n+1} = \min\{n \geq \sigma_{n+1} : X_n \geq b\}.$$

We call an interval $[\sigma_n, \tau_n]$ an upcrossing of the interval $[a, b]$ by X, and we define $U_N([a, b]; X)$ to be the number of such upcrossings by time N.

Lemma 18.3 (Doob's upcrossing lemma) *Let X be a submartingale and $a < b$. Then*

$$\mathbb{E}\left[U_N([a, b]; X)\right] \leq \frac{\mathbb{E}\left[(X_N - a)_+\right]}{(b - a)}.$$

Proof. X is a submartingale implies $(X-a)_+$ is a submartingale (exercise: check) and, furthermore,

$$U_N([a,b];X) = U_N\left([0,b-a];(X-a)_+\right). \tag{13}$$

We can thus prove the statement with X replaced by $Y := (X-a)_+$, $a=0$ and b replaced by $b-a$. We define the previsble process $C=(C_n)_{n=1}^{\infty}$ by

$$C_n = \sum_{k=1}^{\infty} 1_{\{\sigma_k < n \leq \tau_k\}} \in \{0,1\}.$$

Since $(1-C_n)_{n=1}^{\infty} = 1-C$ is bounded, non-negative and previsble, we know from Lemma 18.2 that $(1-C)\cdot Y$ is a submartingale. We therefore have

$$\mathbb{E}\left[((1-C)\cdot Y)_N\right] \geq 0$$

which implies

$$\begin{aligned} \mathbb{E}[Y_N] &\geq \mathbb{E}[Y_N - Y_0] \\ &\geq \mathbb{E}\left[(C\cdot Y)_N\right] \\ &\geq (b-a)\,\mathbb{E}\left[U_N([0,b-a];Y)\right]. \end{aligned}$$

Using the definition of Y and (13) gives

$$\mathbb{E}\left[U_N([a,b];X)\right] \leq \frac{\mathbb{E}\left[(X_N-a)_+\right]}{(b-a)}.$$

■

19 The martingale convergence theorem

Definition 19.1 *We say a family of random variables $\{X_i : i \in I\}$ is bounded in L^1 if* $\sup_{i\in I} \mathbb{E}[|Y_i|] < \infty$

Theorem 19.2 (Submartingale convergence theorem) *Let $X = (X_n)_{n=0}^{\infty}$ be a submartingale which is bounded in L^1 then there exists a random variable X_∞ such that $X_n \stackrel{a.s.}{\to} X_\infty$ as $n \to \infty$.*

Proof. For each $a < b$ we have from the upcrossing lemma that

$$\mathbb{E}[U_N([a,b];X)] \leq \frac{\mathbb{E}\left[(X_N - a)_+\right]}{(b-a)} \leq \frac{\mathbb{E}[|X_N|] + |a|}{(b-a)}.$$

Boundedness in L^1 then gives that $U([a,b];X) := \lim_{N\to\infty} U_N([a,b];X)$ is integrable since, by MCT, we have

$$\begin{aligned} \mathbb{E}[U([a,b];X)] &= \lim_{N\to\infty} \mathbb{E}[U_N([a,b];X)] \\ &\leq \frac{\sup_N \mathbb{E}[|X_N|] + |a|}{(b-a)} < \infty. \end{aligned}$$

It follows that $U([a,b];X) < \infty$ a.s. Letting $X_* = \liminf_{n\to\infty} X_n$ and $X^* = \limsup_{n\to\infty} X_n$ we have

$$\{X_n \text{ does not converge}\} = \{X_* < X^*\} \cup \{X_* = X^* = \pm\infty\},$$

which we now prove is null. First, we have

$$\begin{aligned} \{X_* < X^*\} &\subseteq \cup_{a,b\in\mathbb{Q},a<b} \{X_* < a < b < X^*\} \\ &\subseteq \cup_{a,b\in\mathbb{Q},a<b} \{U([a,b];X) = \infty\}, \end{aligned}$$

which shows that $\mathbb{P}(X_* < X^*) = 0$. Using Fatou's lemma and boundedness in L^1 we obtain

$$\mathbb{E}\left[\lim\inf_{n\to\infty} |X_n|\right] \leq \lim\inf_{n\to\infty} \mathbb{E}[|X_n|] < \infty,$$

so that $X_* < \infty$ a.s. and hence $\mathbb{P}(X_* = X^* = \pm\infty) = 0$. ■

Corollary 19.3 *Suppose X is a non-negative supermartingale then there exists a random variable X_∞ such that $X_n \stackrel{a.s.}{\to} X_\infty$ as $n \to \infty$.*

Proof. Using non-negativity of X we have that $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \leq \mathbb{E}[X_n]$, which gives that X is bounded in L^1. $-X$ is thus a submartingale which is bounded in L^1. It follows form the previous theorem that there exists X_∞ such that $-X_n \stackrel{\text{a.s.}}{\to} -X_\infty$ as $n \to \infty$, i.e. $X_n \stackrel{\text{a.s.}}{\to} X_\infty$ as $n \to \infty$. ■

20 Uniform integrability

Definition 20.1 *We say a family of integrable random variables $\{X_i : i \in I\}$ is uniformly integrable (UI) if*

$$\lim_{K\to\infty} \sup_{i\in I} \mathbb{E}\left[|X_i| 1_{\{|X_i|\geq K\}}\right] = 0.$$

Remark 20.2 *If $\{X_i : i \in I\}$ is UI then it is bounded in L^1, but the converse is false (see the problem sheets).*

Theorem 20.3 *Suppose $X_n \overset{a.s.}{\to} X_\infty$ as $n \to \infty$. If $\{X_n : n \in \mathbb{N}\}$ and X_∞ are integrable and $\{X_n : n \in \mathbb{N}\}$ is UI then $X_n \overset{L^1}{\to} X_\infty$ as $n \to \infty$.*

Proof. Since $L^1(\Omega, \mathcal{F}, \mathbb{P})$ is complete it suffices to prove that the sequence (X_n) is Cauchy in L^1.Therefore fix $\epsilon > 0$ and for $a > 0$ define the function $\phi_a(x) = x1_{[-a,a]} + a1_{(a,\infty)} - a1_{(-\infty,-a)}$. Observe that

$$\begin{aligned} ||X_n - X_m||_1 &\leq ||X_n - \phi_a(X_n)||_1 + ||\phi_a(X_n) - \phi_a(X_m)||_1 \\ &\quad + ||\phi_a(X_m) - X_m||_1 . \end{aligned}$$

and note that for every n we have

$$||X_n - \phi_a(X_n)||_1 \leq \sup_n \mathbb{E}\left[|X_n| 1_{\{|X_n|\geq a\}}\right].$$

Using UI we can choose a large enough so that $\sup_n \mathbb{E}\left[|X_n| 1_{\{|X_n|\geq a\}}\right] < \epsilon/3$ and hence for every n and m

$$||X_n - X_m||_1 \leq \frac{2\epsilon}{3} + ||\phi_a(X_n) - \phi_a(X_m)||_1 . \tag{14}$$

Finally, $\phi_a(X_n) \overset{\text{a.s.}}{\to} \phi_a(X_\infty)$ as $n \to \infty$ because $X_n \overset{\text{a.s.}}{\to} X_\infty$ as $n \to \infty$ and ϕ_a is continuous. Since $|\phi_a(X_n)| \leq a$ we can use the dominated convergence theorem to deduce that $\phi_a(X_n) \overset{L^1}{\to} \phi_a(X_\infty)$ as $n \to \infty$. Again, since $L^1(\Omega, \mathcal{F}, \mathbb{P})$ is complete, this implies that it suffices to prove that the sequence $(\phi_a(X_n))$ is Cauchy in L^1 and there exists N such that

$$||\phi_a(X_n) - \phi_a(X_m)||_1 < \frac{\epsilon}{3} \text{ for all } n, m \geq N.$$

Hence from (14) we obtain

$$||X_n - X_m||_1 < \epsilon \text{ for all } n, m \geq N,$$

and since ϵ is arbitrary the proof is finished. ■

Lemma 20.4 *Suppose $\{X_n : n \in \mathbb{N}\}$ is a family of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the following are equivalent:*

1. *The family $\{X_n : n \in \mathbb{N}\}$ is UI.*

2. *$\{X_n : n \in \mathbb{N}\}$ is bounded in L^1 and for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$ we have $\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n| 1_A] < \epsilon$.*

Proof. Assume that 2 holds. Suppose $\epsilon > 0$ and let $L := \sup_n \mathbb{E}[|X_n|] < \infty$. Then for any $n \in \mathbb{N}$ and $K > 0$ we have

$$1_{\{|X_n|\geq K\}} \leq K^{-1} |X_n|,$$

which implies

$$\sup_n \mathbb{P}(|X_n| \geq K) \leq K^{-1} \sup_n \mathbb{E}[|X_n|] = K^{-1}L < \delta$$

for $K > L\delta^{-1}$. It follows from applying 2 that

$$\mathbb{E}\left[|X_n| 1_{\{|X_n|\geq K\}}\right] \leq \mathbb{E}\left[|X_n| 1_{\{|X_n|\geq K\}}\right] < \epsilon$$

for every n. Since ϵ was arbitrary, this implies 1.

On the other hand, if 1 holds then for any $A \in \mathcal{F}$ we have

$$\mathbb{E}[|X_n| 1_A] \leq K\mathbb{P}(A) + \mathbb{E}\left[|X_n| 1_{\{|X_n|\geq K\}}\right].$$

Therefore if $\epsilon > 0$ we can first choose K so that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[|X_n| 1_{\{|X_n|\geq K\}}\right] < \frac{\epsilon}{2},$$

and then for any A with $\mathbb{P}(A) < \frac{\epsilon}{2K}$ we will have

$$\mathbb{E}[|X_n| 1_A] < \epsilon$$

for all n. ∎

21 Uniform integrability continued

Proposition 21.1 *Assume $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ then the class of random variables*

$$\{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \text{ is a sub } \sigma - \text{field of } \mathcal{F}\}$$

is uniformly integrable.

Proof. Let $a > 0$ and notice by the conditional Jensen inequality that

$$|\mathbb{E}[X|\mathcal{G}]| 1_{\{|\mathbb{E}[X|\mathcal{G}]|>a\}} \leq \mathbb{E}[|X||\mathcal{G}] 1_{\{\mathbb{E}[|X||\mathcal{G}]>a\}}$$

and hence

$$\begin{aligned} \mathbb{E}\left[|\mathbb{E}[X|\mathcal{G}]| 1_{\{|\mathbb{E}[X|\mathcal{G}]|>a\}}\right] &\leq \mathbb{E}\left[\mathbb{E}[|X||\mathcal{G}] 1_{\{\mathbb{E}[|X||\mathcal{G}]>a\}}\right] \\ &= \mathbb{E}\left[|X| 1_{\{\mathbb{E}[|X||\mathcal{G}]>a\}}\right], \end{aligned} \tag{15}$$

where the last line using the fact that $\{\mathbb{E}[|X||\mathcal{G}] > a\} \in \mathcal{G}$ and the definition of conditional expectation. We then have for any $K > 0$

$$\begin{aligned} \mathbb{E}\left[|X| 1_{\{\mathbb{E}[|X||\mathcal{G}]>a\}}\right] &= \mathbb{E}\left[|X| 1_{\{\mathbb{E}[|X||\mathcal{G}]>a\}\cap\{|X|\leq K\}}\right] \\ &\quad +\mathbb{E}\left[|X| 1_{\{\mathbb{E}[|X||\mathcal{G}]>a\}\cap\{|X|\geq K\}}\right] \\ &\leq K\mathbb{P}(\mathbb{E}[|X||\mathcal{G}] > a) + \mathbb{E}\left[|X| 1_{\{|X|\geq K\}}\right]. \end{aligned}$$

Using Chebyshev's inequality we have $\mathbb{P}(\mathbb{E}[|X||\mathcal{G}] > a) \leq a^{-1}\mathbb{E}[\mathbb{E}[|X||\mathcal{G}]] = a^{-1}\mathbb{E}[|X|]$ so that for any $K > 0$

$$\mathbb{E}\left[|X| 1_{\{\mathbb{E}[|X||\mathcal{G}]>a\}}\right] \leq \frac{K}{a}\mathbb{E}[|X|] + \mathbb{E}\left[|X| 1_{\{|X|\geq K\}}\right].$$

Let $K = a^{1/2}$ then using this bound in (15) gives

$$\lim_{a\to\infty} \sup_{\mathcal{G}} \mathbb{E}\left[|\mathbb{E}[X|\mathcal{G}]| 1_{\{|\mathbb{E}[X|\mathcal{G}]|>a\}}\right] \leq \lim_{a\to\infty}\left(\frac{\mathbb{E}[|X|]}{a^{1/2}} + \mathbb{E}\left[|X| 1_{\{|X|\geq a^{1/2}\}}\right]\right) = 0, \tag{16}$$

where the last line use the DCT to see that $\mathbb{E}\left[|X| 1_{\{|X|\geq a^{1/2}\}}\right] \to 0$ as $a \to \infty$. ■

By strengthening the hypothesis of L^1-boundedness to uniform integrability, we can deduce L^1 convergence of a martingale to its terminal value. Before we discuss this result, we first prove the following lemma which can be viewed as a refinement of Fatou's lemma.

Lemma 21.2 *Suppose that $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable, then*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty}\mathbb{E}[X_n] \leq \limsup_{n\to\infty}\mathbb{E}[X_n] \leq \mathbb{E}\left[\limsup_{n\to\infty} X_n\right].$$

In particular, if $X_n \to X$ a.s. as $n \to \infty$ then $X \in L^1$ and $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$.

Proof. Let $\epsilon > 0$. For any $C > 0$, we have that

$$\mathbb{E}[X_n] = \mathbb{E}\left[X_n 1_{\{X_n < -C\}}\right] + \mathbb{E}\left[X_n 1_{\{X_n \geq -C\}}\right].$$

Using the uniform integrability hypothesis there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\left|\mathbb{E}\left[X_n 1_{\{X_n < -C\}}\right]\right| \leq \mathbb{E}\left[|X_n| 1_{\{|X_n| > C\}}\right] \leq \epsilon.$$

This implies that

$$\mathbb{E}[X_n] \geq -\epsilon + \mathbb{E}\left[X_n 1_{\{X_n \geq -C\}}\right], \text{ for all } n \geq N,$$

so that

$$\liminf_{n\to\infty} \mathbb{E}[X_n] \geq -\epsilon + \liminf_{n\to\infty} \mathbb{E}\left[X_n 1_{\{X_n \geq -C\}}\right]. \tag{17}$$

Since $X_n 1_{\{X_n \geq -C\}} + C \geq 0$ we may use Fatou's lemma to see that

$$\begin{aligned}
\mathbb{E}\left[\liminf_{n\to\infty} X_n 1_{\{X_n \geq -C\}}\right] + C &= \mathbb{E}\left[\liminf_{n\to\infty}\left(X_n 1_{\{X_n \geq -C\}} + C\right)\right] \\
&\leq \liminf_{n\to\infty} \mathbb{E}\left[\liminf_{n\to\infty} X_n 1_{\{X_n \geq -C\}}\right] + C.
\end{aligned}$$

Using this with the observation $X_n 1_{\{X_n < -C\}} \geq X_n$ shows that

$$\begin{aligned}
\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] &\leq \mathbb{E}\left[\liminf_{n\to\infty} X_n 1_{\{X_n \geq -C\}}\right] \\
&\leq \liminf_{n\to\infty} \mathbb{E}\left[\liminf_{n\to\infty} X_n 1_{\{X_n \geq -C\}}\right].
\end{aligned}$$

From (17) we then deduce that

$$\epsilon + \liminf_{n\to\infty} \mathbb{E}[X_n] \geq \mathbb{E}\left[\liminf_{n\to\infty} X_n\right],$$

and because $\epsilon > 0$ was arbitrary this gives the result. The proof for the $\limsup$ is similar and is left as an exercise. The conclusion that $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$ is them immediate once we know $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$. ■

22 Uniformly integrable martingales

We prove the following result.

Theorem 22.1 (Lévy) *Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0}^{\infty}, \mathbb{P})$ be a filtered probability space. Suppose $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then if $\mathcal{F}_\infty = \sigma(\cup_{n=0}^{\infty} \mathcal{F}_n)$*

$$\mathbb{E}[Y | \mathcal{F}_n] \to \mathbb{E}[Y | \mathcal{F}_\infty] \ \text{as } n \to \infty, \ \text{a.s. and in } L^1$$

Proof. Let $X_n := \mathbb{E}[Y | \mathcal{F}_n]$, then we have proved that $X = (X_n)_{n=0}^{\infty}$ is a martingale. It is UI because

$$\{X_n : n \in \mathbb{N} \cup \{0\}\} \subset \{\mathbb{E}[Y | \mathcal{G}] : \mathcal{G} \text{ is a sub } \sigma\text{-field of } \mathcal{F}\},$$

and this latter set is UI. By the martingale convergence theorem, there exists X_∞ such that $X_n \to X_\infty$ a.s. This convergence also takes place in L^1 because of uniform integrability. The only think left to check is that $X_\infty = \mathbb{E}[Y | \mathcal{F}_\infty]$.

By considering Y^+ and Y^- separately, we may assume that $Y \geq 0$ a.s. This will also imply that $X_n \geq 0$ a.s for every n and $X_\infty \geq 0$ a.s. We then define two measures on $(\Omega, \mathcal{F}_\infty)$ by

$$\mu_1(A) = \mathbb{E}[Y 1_A] \text{ and } \mu_2(A) = \mathbb{E}[X_\infty 1_A] \ ,$$

and prove that $\mu_1 = \mu_2$. To do so we first note that $\Pi := \cup_{n=0}^{\infty} \mathcal{F}_n$ is a π-system which generates $\mathcal{F}_\infty$; if we can show that $\mu_1 = \mu_2$ on Π then the proof will be complete. Let $A \in \Pi$ then $A \in \mathcal{F}_n$ for some n and therefore

$$\mu_1(A) = \mathbb{E}[Y 1_A] = \mathbb{E}[\mathbb{E}[Y | \mathcal{F}_n] 1_A] = \mathbb{E}[X_n 1_A],$$

whereas

$$\mu_2(A) = \mathbb{E}[X_\infty 1_A] = \lim_{k \to \infty} \mathbb{E}[X_k 1_A],$$

and because for all $k \geq n$ we have $\mathbb{E}[X_k 1_A] = \mathbb{E}[\mathbb{E}[X_k | \mathcal{F}_n] 1_A] = \mathbb{E}[X_n 1_A]$ it follows that

$$\mu_2(A) = \mathbb{E}[X_n 1_A] = \mu_1(A).$$

■

Theorem 22.2 (Martingale convergence theorem for UI martingales) *Let $X = (X_n)_{n=0}^{\infty}$ be a UI martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0}^{\infty}, \mathbb{P})$. There exists a random variable X_∞ such that $X_n \to X_\infty$ as $n \to \infty$, a.s. and in L^1,moreover $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ for all n.*

Proof. Since $(X_n)_{n=0}^{\infty}$ is UI it must be bounded in L^1. The martingale convergence theorem immediately gives the existence of the almost sure limit X_∞. The UI hypothesis gives the L^1-convergence. It remains to show that $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ a.s. To do so suppose $k \geq n$ and let $A \in \mathcal{F}_n$ then since X is a martingale we have

$$\mathbb{E}[X_k 1_A] = \mathbb{E}[\mathbb{E}[X_k | \mathcal{F}_n] 1_A] = \mathbb{E}[X_n 1_A]. \tag{18}$$

On the other hand

$$|\mathbb{E}[X_k 1_A] - \mathbb{E}[X_\infty 1_A]| \leq \mathbb{E}[|X_k - X_\infty|] \to 0 \text{ as } k \to \infty.$$

Letting $k \to \infty$ in (18) yields $\mathbb{E}[X_\infty 1_A] = \mathbb{E}[X_n 1_A]$ for all $A \in \mathcal{F}_n$, i.e. $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ a.s. ■

Definition 22.3 *Let $p > 0$. $X = (X_n)_{n=0}^\infty$ is called bounded in L^p if*

$$\sup_{n \geq 0} ||X_n||_p = \sup_{n \geq 0} \mathbb{E}[|X_n|^p]^{1/p} < \infty.$$

Theorem 22.4 (L^p martingale convergence theorem) *Let $X = (X_n)_{n=0}^\infty$ be a martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0}^\infty, \mathbb{P})$ which is bounded in L^p for some $p > 1$. There exists a random variable X_∞ such that $X_n \to X_\infty$ in L^p as $n \to \infty$.*

Proof. X being bounded in L^p for $p > 1$ implies that X is UI (see the problems sheet). Hence, by the UI martingale convergence theorem there exists X_∞ such that $X_n \to X_\infty$ as $n \to \infty$, a.s. and in L^1 and $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$. From Fatou's lemma we then have

$$\mathbb{E}[|X_\infty|^p] = \mathbb{E}\left[\liminf_{n\to\infty} |X_n|^p\right] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|^p],$$

which is finite by the L^p boundedness of X. Therefore $X_\infty \in L^p$. We finish the proof by considering two cases:

1. Assume X_∞ is bounded, i.e. there exists a finite K such that $|X_\infty| \leq K$ a.s Since $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ it then also follows that $|X_n| \leq K$ a.s for all n and the DCT immediately yields

$$\mathbb{E}[|X_n - X_\infty|^p] \to 0 \text{ as } n \to \infty.$$

2. In the unbounded case we first write

$$X_\infty = X_\infty 1_{\{|X_\infty| \leq K\}} + X_\infty 1_{\{|X_\infty| > K\}},$$

and then

$$\begin{aligned}
\mathbb{E}[|X_n - X_\infty|^p]^{1/p} &= \mathbb{E}[|\mathbb{E}[X_\infty | \mathcal{F}_n] - X_\infty|^p]^{1/p} \\
&\leq \mathbb{E}\left[\left|\mathbb{E}\left[X_\infty 1_{\{|X_\infty| \leq K\}} \middle| \mathcal{F}_n\right] - X_\infty 1_{\{|X_\infty| \leq K\}}\right|^p\right]^{1/p} \\
&\quad + \mathbb{E}\left[\left|\mathbb{E}\left[X_\infty 1_{\{|X_\infty| > K\}} \middle| \mathcal{F}_n\right] - X_\infty 1_{\{|X_\infty| > K\}}\right|^p\right]^{1/p} \\
&\leq \mathbb{E}\left[\left|\mathbb{E}\left[X_\infty 1_{\{|X_\infty| \leq K\}} \middle| \mathcal{F}_n\right] - X_\infty 1_{\{|X_\infty| \leq K\}}\right|^p\right]^{1/p} \\
&\quad + 2\mathbb{E}\left[\left|X_\infty 1_{\{|X_\infty| > K\}}\right|^p\right]^{1/p}.
\end{aligned}$$

From the bounded case we know that, for each K, as $n \to \infty$

$$\mathbb{E}\left[X_\infty 1_{\{|X_\infty| \leq K\}} \middle| \mathcal{F}_n\right] \xrightarrow{L^p} X_\infty 1_{\{|X_\infty| \leq K\}},$$

which deals with the first term above. We can handle the second term using the DCT which gives us that

$$\mathbb{E}\left[|X_\infty|^p \, 1_{\{|X_\infty|>K\}}\right] \to 0$$

as $K \to \infty$.

■

23 Continuous time stochastic processes

We now turn to the case of a real-valued stochastic process $X = (X_t)_{t \in T}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ where the indexing set T is uncountable (in typical cases we will have $T = [0, t]$ or $T = [0, \infty)$). We can again think of this stochastic process in two ways:

1. As a family of random variables $X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for every $t \in T$
2. As a measurable function $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \left(\mathbb{R}^T, \xi^T\right)$, where $\mathbb{R}^T$ denotes the set of all functions $f : T \to \mathbb{R}$ and ξ^T denotes the σ-field

$$\xi^T = \sigma\left(\pi_t : t \in T\right),$$

where $\pi_t : \mathbb{R}^T \to \mathbb{R}$ are the canonical evaluation maps given by $\pi_t(f) = f(t)$ for every $t \in T$.

Exercise 23.1 *Prove that $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \left(\mathbb{R}^T, \xi^T\right)$ is measurable iff $X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable for every $t \in T$*

Definition 23.2 *The law of a stochastic process $X = (X_t)_{t \in T}$ is the probability measure on $\left(\mathbb{R}^T, \xi^T\right)$ defined by*

$$\mu(A) = \mathbb{P}(X \in A) \text{ for } A \in \xi^T.$$

The family of finite dimensional distributions of X (or, equivalently, of μ) is the collection of probability measures indexed by all finite subsets of T, i.e.

$$\{\mu_S : S \subset T \text{ finite}\},$$

where μ_S is the probability measure on $\left(\mathbb{R}^S, \xi^S\right)$ given by

$$\mu_S(A) = \mathbb{P}\left(X|_S \in A\right) \text{ for } A \in \xi^S,$$

where $X|_S$ denotes the restriction of X to a function from S to $\mathbb{R}$.

One basic question is when a collection of finite dimensional distributions determine the law of a stochastic process. In order to do so the collection must satisfy some consistency properties. The basic requirement here is that they form a *projective family* in the sense that if $S \subseteq U$ and if $\phi_S^U : \mathbb{R}^U \to \mathbb{R}^S$ denote the canonical restriction map, then

$$\mu_S(A) = \mu_U\left(\left(\phi_S^U\right)^{-1}(A)\right) \text{ for all } A \in \xi^S. \tag{19}$$

This is made precise in the Daniell-Kolmogorov theorem.

Theorem 23.3 (Daniell-Kolmogorov) *Suppose* $\{\mu_S : S \subset T,\ S \text{ finite}\}$ *is a family of probability measures which satisfy the consistency relations (19). Then there exists a unique probability measure* μ *on* $\left(\mathbb{R}^T, \xi^T\right)$ *for which the given family is the collection of finite dimensional distributions of* μ*.*

23.1 Regularisation of continuous-time martingales

Definition 23.4 *A stochastic process* $X = (X_t)_{t\geq 0}$ *is said to be cadlag (or RCLL: right-continuous with left-limits) if a.s.*

$$\begin{aligned} \forall t &> 0 \lim_{s\uparrow t,} X_s \text{ exists in } \mathbb{R} \\ \text{and } \forall t &\geq 0 \lim_{s\downarrow t} X_s = X_t \end{aligned}$$

Definition 23.5 *Given two stochastic processes* X *and* Y *we say:*

1. X *is a version (or modification) of* Y *if,* $\forall t \geq 0, \mathbb{P}(X_t = Y_t) = 1$
2. X *and* Y *are indistinguishable if* $\mathbb{P}(X_t = Y_t, \forall t \geq 0) = 1$

Exercise 23.6 *If* X *and* Y *are both cadlag, prove that* X *and* Y *are indistinguishable if* X *is a modification of* Y*.*

Definition 23.7 *A process* $X = (X_t)_{t\geq 0}$ *is a super(sub) martingale with respected to the filtered probability space* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ *if* X *is adapted,* X_t *is integrable for every* t *and if*

$$\mathbb{E}[Y_t|\,\mathcal{F}_s] \leq Y_s \text{ a.s. (resp. } \mathbb{E}[Y_t|\,\mathcal{F}_s] \geq Y_s \text{ a.s.)}$$

for all $s \leq t$*. We call* X *a martingale if it is both a submartingale and a supermartingale.*

The following result is the key regularisation theorem for submartingales.

Theorem 23.8 *Suppose* $X = (X_t)_{t\geq 0}$ *is a submartingale on a filtered probability space* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$*. Then*

1. $\mathbb{P}$*almost surely the following statements hold*

$$\forall t > 0 \lim_{q\uparrow t, q\in\mathbb{Q}} X_q \text{ exists and } \forall t \geq 0 \lim_{q\downarrow t, q\in\mathbb{Q}} X_q \text{ exists.}$$

2. *Define a new process by setting* $Y_t := X_{t+} := \lim_{q\downarrow t, q\in\mathbb{Q}} X_q$ *for* $t \geq 0$*, then* $Y_t \in L^1(\Omega, \mathcal{F}, \mathbb{P})$*. The process* Y *is a submartingale with respect to* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_{t+}\}_{t\geq 0}, \mathbb{P}\right)$*, and is a martingale if* X *is. We have*

$$X_t \leq \mathbb{E}[Y_t|\,\mathcal{F}_t] \text{ a.s.}$$

with equality if $\mathbb{E}[X_t]$ *is a right-continuous function in* t

Proof. We do not work through the details – see Revuz and Yor Theorems 2.5 and Proposition 2.6 for full details. ■

Definition 23.9 *A filtered probability space* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ *satisfies the usual conditions if*

1. $\mathcal{F}$ *is* $\mathbb{P}-$*complete, i.e. if* $\mathcal{F} \supseteq \mathcal{N}$ *where* $\mathcal{N}$ *is the set of all subsets of all* $\mathbb{P}$*-null sets;*
2. $\mathcal{F}_t = \mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$ *(right-continuity of the filtration);*
3. $\mathcal{F}_0 \supseteq \mathcal{N}$.

As a corollary we can obtain the following.

Corollary 23.10 *Suppose* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ *is a filtered probability space, satisfying the usual conditions, with respect to which* X *is a submartingale. If* $\mathbb{E}[X_t]$ *is a right-continuous function in* t *then* X *has a cadlag modification* Y *which is again a submartingale on* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$.

From now on we make the standing assumption that $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ satisfies the usual conditions and that all (sub/super) martingales are cadlag. This, in particular, allows us to prove the martingale convergence theorem in continuous time by adapting the proof of the discrete-time analogues.

24 Quadratic variation

We have seen that stochastic integration w.r.t stochastic processes of finite variation can be understood using measure theory. The following result demonstrates that this theory is inadequate for a stochastic integration theory which hopes to include martingales as integrators.

Proposition 24.1 *Let $M = (M_t)_{t \geq 0}$ be a continuous martingale. If M has finite variation then M is constant.*

Proof. By considering $(M_t - M_0)_{t \geq 0}$ if necessary we may suppose that $M_0 = 0$. Let S_t denote the total variation of M on $[0, t]$ and let τ_n denote the stopping time

$$\tau_n = \inf \{u \geq 0 : S_u \geq n\},$$

then the stopped martingale M^{τ_n} has bounded total variation. It is therefore suffices to prove the result under the strengthened assumption of bounded total variation, i.e. $S_t \leq K$ a.s. for every t.

Now suppose that $\Delta = \{0 = t_0 < t_1 < ....t_r = t\}$ be an arbitrary partition of $[0, t]$. Since M is a martingale we have

$$\begin{aligned}
\mathbb{E}\left[M_t^2\right] &= \mathbb{E}\left[\sum_{i=0}^{k-1}\left(M_{t_{i+1}}^2 - M_{t_i}^2\right)\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{k-1}\left(M_{t_{i+1}} - M_{t_i}\right)^2\right] \\
&\leq \mathbb{E}\left[S_t \sup_i \left|M_{t_{i+1}} - M_{t_i}\right|\right] \\
&\leq K\mathbb{E}\left[\sup_i \left|M_{t_{i+1}} - M_{t_i}\right|\right].
\end{aligned}$$

The continuity of M implies that the RHS tends to zero when the mesh $|\Delta| := \max_i |t_{i+1} - t_i| \to 0$. Therefore $\mathbb{E}\left[M_t^2\right] = 0$, which implies $M_t = 0$ a.s. ■

The notion of finite quadratic variation is crucial to the stochastic calculus for semimartinagles. To introduce the definition for a process X we let

$$T_t^{\Delta}(X) := \sum_{i=0}^{k-1}\left(X_{t_{i+1}} - X_{t_i}\right)^2 + \left(X_t - X_{t_k}\right)^2 \text{ for } t \in [t_k, t_{k+1}]$$

where $\Delta = \{0 = t_0 < t_1 < ....\}$ is a countable partition of $\mathbb{R}_+$ such that $[0, s] \cap \Delta$ is a finite set for every $0 \leq s < \infty$.

Definition 24.2 *A process X has finite quadratic variation if, for each $t \geq 0$, $T_t^{\Delta}(X)$ converges in probability as $|\Delta| \to 0$. We denote the limit by $[X]_t$, and call it the quadratic variation of X at time t.*

We will make use of the following result, which we state without proof.

25 Quadratic variation continued

We will make use of the following result, which we state without proof.

Theorem 25.1 (Doob's L^p-inequality) *Let $p > 1$ and suppose $(M_t)_{t\geq 0}$ is a continuous martingale which is bounded in L^p. Then for any $t \geq 0$ we have*

$$\mathbb{E}\left[\sup_{s\leq t} |M_s|^p\right]^{1/p} \leq \frac{p}{p-1} \mathbb{E}\left[|M_t|^p\right]^{1/p}.$$

Theorem 25.2 (Existence of quadratic variation) *Suppose $M = (M_t)_{t\geq 0}$ is a continuous and bounded (i.e. $|M_t| \leq K$ a.s. for some K) martingale. There exists a unique process $[M]$ which is continuous, non-decreasing, adapted with $M_0 = 0$ such that $M^2 - [M] := \left(M_t^2 - [M]_t\right)_{t\geq 0}$ is a martingale.*

Proof. Without loss of generality we assume that $M_0 = 0$. The uniqueness assertion is clear, because if $[M]$ and $\widetilde{[M]}$ are two processes with the stated properties then the difference

$$[M] - \widetilde{[M]} = \left(M^2 - \widetilde{[M]}\right) - \left(M^2 - [M]\right),$$

is a continuous martingale of finite variation, and hence is the zero process.

To prove existence we let $n \in \mathbb{N}$ and define a sequence of stopping times by setting $T_0^n = 0$ and then

$$T_{k+1}^n = \inf\left\{t > T_k^n : \left|M_t - M_{T_k^n}\right| > \frac{1}{2^n}\right\}.$$

For $t > 0$ we let $t_k^n := t \wedge T_k^n$ and then notice that

$$\begin{aligned} M_t^2 &= \sum_{k=1}^{\infty} M_{t_k^n}^2 - M_{t_{k-1}^n}^2 \\ &= \sum_{k=1}^{\infty} \left(M_{t_k^n} - M_{t_{k-1}^n}\right)^2 + 2\sum_{k=1}^{\infty} M_{t_{k-1}^n}\left(M_{t_k^n} - M_{t_{k-1}^n}\right) \\ &=: T_t^{\Delta_n}(M) + 2\left(H^n \cdot M\right)_t \end{aligned} \tag{20}$$

where $\Delta_n = \{T_k^n : k = 0, 1, 2, ...\}$ and

$$H_u^n := \sum_{k=1}^{\infty} M_{T_{k-1}^n} 1_{(T_{k-1}^n, T_k^n]}(u)$$

(recall the definition of the discrete-time martingale transform). The process $t \mapsto (H^n \cdot M)_t$ is a continuous martingale which is bounded in L^2 and hence , for every n,

$$(H^n \cdot M)_t \to (H^n \cdot M)_\infty \text{ in } L^2 \text{ as } t \to \infty.$$

We note that

$$\sup_{t\geq 0}\left|H_t^{n+1}-H_t^n\right|\leq \frac{1}{2^{n+1}},\ \ \sup_{t\geq 0}\left|M_t-H_t^n\right|\leq \frac{1}{2^n}$$

and that $\Delta_n \subset \Delta_{n+1}$ while

$$T^{\Delta_n}_{T^n_k}(M)\leq T^{\Delta_n}_{T^n_k}(M)\ \text{ for all } k=0,1,2,... \tag{21}$$

A simple exercise shows that

$$\mathbb{E}\left[\left(\left(H^{n+1}\cdot M\right)_\infty-\left(H^n\cdot M\right)_\infty\right)^2\right]\leq \frac{1}{4^{n+1}}\mathbb{E}\left[M_\infty^2\right],$$

whereupon Doob's L^2 inequality gives

$$\mathbb{E}\left[\left|\sup_{t\geq 0}\left(\left(H^{n+1}\cdot M\right)_t-\left(H^n\cdot M\right)_t\right)\right|^2\right]\leq \frac{1}{4^n}\mathbb{E}\left[M_\infty^2\right].$$

It follows that the sequence $(H^n\cdot M)$ of L^2 bounded martingales converges uniformly in L^2 to a (necessarily) continuous process $(H\cdot M)$, which must also be am L^2 bounded martingales. From the identity (20) we see that $T^{\Delta_n}(M)$ must also converge uniformly in L^2 to a process $[M]$, which must again be continuous.

It remains to check that $[M]$ has the asserted properties. First, by defintion, we have that $M^2-[M]=H\cdot M$, which we have seen is a martingale. We also have for any $m\geq n$ the fact that

$$T^{\Delta_m}_{T^n_k}(M)\leq T^{\Delta_m}_{T^n_{k+1}}(M)\ \text{ for all } k=0,1,2,...$$

This follows from (21) together with $\Delta_n\subset\Delta_m$. By letting $m\to\infty$ we learn that $[M]$ is non-decreasing on $\Delta:=\cup_n\Delta_n$, and hence on $\bar{\Delta}$ by continuity of $[M]$. Finally, if I any interval in $[0,\infty)$ contained in the complement of Δ then M must be constant on I; it follows that $[M]$ must be constant on I. ■

Theorem 25.3 *Suppose $M=(M_t)_{t\geq 0}$ is a continuous and bounded martingale, then M has finite quadratic variation. Moreover, $[M]$ is the unique process identified in Theorem 25.2.*

Proof. We follow Revuz and Yor. The uniqueness is a consequence of the previous proposition because, if there were two such processes their difference would be a continuous martingale starting from 0 of finite variation, and hence it must be constant (and thus identically equal to 0).

To prove that $[M]$ exists we first notice that if $s\in[t_i,t_{i+1}]$ then

$$\mathbb{E}\left[\left(M_{t_{i+1}}-M_{t_i}\right)^2\middle|\mathcal{F}_s\right]=\mathbb{E}\left[\left(M_{t_{i+1}}-M_s\right)^2\middle|\mathcal{F}_s\right]+\left(M_s-M_{t_i}\right)^2,$$

from which it follows that

$$\begin{aligned}\mathbb{E}\left[T_t^{\Delta}(M) - T_s^{\Delta}(M)\middle|\mathcal{F}_s\right] &= \mathbb{E}\left[(M_t - M_s)^2\middle|\mathcal{F}_s\right] \\ &= \mathbb{E}\left[M_t^2 - M_s^2\middle|\mathcal{F}_s\right]. \end{aligned} \tag{22}$$

Hence, for every Δ, the process $M^2 - T^{\Delta}(M)$ is a continuous martingale.

Fix $a > 0$ we will prove that the limit in L^2, $\lim_{|\Delta|\to 0} T_a^{\Delta}(M)$ exists. To do so we use the notation $\Delta\Delta' := \Delta \cup \Delta'$, when Δ and Δ' are two partitions of $\mathbb{R}_+$. We have

$$X := T^{\Delta}(M) - T^{\Delta'}(M) = \left(M^2 - T^{\Delta'}(M)\right) - \left(M^2 - T^{\Delta'}(M)\right)$$

is a martingale (being the difference of two martingales). By applying (22) with M replaced by X and Δ by $\Delta\Delta'$ we learn that

$$\mathbb{E}\left[X_a^2\right] = \mathbb{E}\left[\left(T_a^{\Delta}(M) - T_a^{\Delta'}(M)\right)^2\right] = \mathbb{E}\left[T_a^{\Delta\Delta'}(X)\right].$$

Using the elementary inequality $(u+v)^2 \leq 2\left(u^2+v^2\right)$ we have

$$T_a^{\Delta\Delta'}(X) \leq 2\left[T_a^{\Delta\Delta'}\left(T^{\Delta}(M)\right) + T_a^{\Delta\Delta'}\left(T^{\Delta'}(M)\right)\right],$$

and therefore it is sufficient to prove that $\mathbb{E}\left[T_a^{\Delta\Delta'}\left(T^{\Delta}(M)\right)\right] \to 0$ as $\max\left(|\Delta|, |\Delta'|\right) \to 0$.

Let $s_k \in \Delta\Delta'$ be such that $[s_k, s_{k+1}] \subseteq [t_l, t_{l+1}]$ then

$$\begin{aligned} T_{s_{k+1}}^{\Delta}(M) - T_{s_k}^{\Delta}(M) &= \left(M_{s_{k+1}} - M_{t_l}\right)^2 - \left(M_{s_k} - M_{t_l}\right)^2 \\ &= \left(M_{s_{k+1}} - M_{s_k}\right)\left(M_{s_{k+1}} + M_{s_k} - 2M_{t_l}\right), \end{aligned}$$

so that

$$T_a^{\Delta\Delta'}\left(T^{\Delta}(M)\right) \leq \sup_k \left|\left(M_{s_{k+1}} + M_{s_k} - 2M_{t_l}\right)\right|^2 T_a^{\Delta\Delta'}(M).$$

Using the Cauchy-Schwarz inequality we have

$$\mathbb{E}\left[T_a^{\Delta\Delta'}\left(T^{\Delta}(M)\right)\right] \leq \underbrace{\mathbb{E}\left[\sup_k \left|\left(M_{s_{k+1}} + M_{s_k} - 2M_{t_l}\right)\right|^4\right]^{1/2}}_{\to 0 \text{ as } \max(|\Delta|,|\Delta'|)\to 0 \text{ by continuity and boundedness of } M} \mathbb{E}\left[T_a^{\Delta\Delta'}(M)^2\right]^{1/2}$$

and we need to show that the second term in the above product is bounded uniformly over Δ and Δ'.

To do this we assume $a = t_n \in \Delta$, and notice that

$$\begin{aligned}
T_a^\Delta(M)^2 &= \left[\sum_{k=1}^n \left(M_{t_k} - M_{t_{k-1}}\right)^2\right]^2 \\
&= \sum_{k=1}^n \left(M_{t_k} - M_{t_{k-1}}\right)^4 + 2\sum_{k_1=1}^n \sum_{k_2=k_1+1}^n \left(M_{t_{k_2}} - M_{t_{k_2-1}}\right)^2 \left(M_{t_{k_1}} - M_{t_{k_1-1}}\right)^2 \\
&= \sum_{k=1}^n \left(M_{t_k} - M_{t_{k-1}}\right)^4 + 2\sum_{k=1}^n \left(T_a^\Delta(M) - T_{t_k}^\Delta(M)\right)\left(T_{t_k}^\Delta(M) - T_{t_{k-1}}^\Delta(M)\right).
\end{aligned}$$

From (22) we have that $\mathbb{E}\left[T_a^\Delta(M) - T_{t_k}^\Delta(M)\middle|\mathcal{F}_{t_k}\right] = \mathbb{E}\left[\left(M_a - M_{t_k}\right)^2\middle|\mathcal{F}_{t_k}\right]$ which results in

$$\mathbb{E}\left[T_a^\Delta(M)^2\right] \le \mathbb{E}\left[\left\{\sup_k \left(M_{t_k} - M_{t_{k-1}}\right)^2 + 2\sup_k \left(M_a - M_{t_k}\right)^2\right\} T_a^\Delta(M)\right].$$

Since $|M_t|$ is bounded by K a.s. we may deduce that

$$\begin{aligned}
\mathbb{E}\left[T_a^\Delta(M)^2\right] &\le 12K^2\mathbb{E}\left[T_a^\Delta(M)\right] \\
&= 12K^2\mathbb{E}\left[\left(M_a - M_0\right)^2\right] \\
&\le 48K^4,
\end{aligned}$$

which proves that $\mathbb{E}\left[T_a^\Delta(M)^2\right]$ is bounded uniformly over all partitions Δ. We have thus proved that $\left(T_a^{\Delta_n}(M)\right)_{n=1}^\infty$ converges in L^2 (and hence in probability) for any sequence of partitions $(\Delta_n)_{n=1}^\infty$ such that $|\Delta_n| \to 0$; we denote this limit by $[M]_a$.

We now verify that $[M]_a$ has the properties claimed in the statement. To do so, we first apply Doob's L^2 inequality to the martingale $T^{\Delta_n}(M) - T^{\Delta_m}(M)$ over $[0,a]$. This gives

$$\mathbb{E}\left[\sup_{t\le a}\left|T_t^{\Delta_n}(M) - T_t^{\Delta_m}(M)\right|^2\right] \le 4\mathbb{E}\left[\left|T_a^{\Delta_n}(M) - T_a^{\Delta_m}(M)\right|^2\right],$$

and then, since $\left(T_a^{\Delta_n}(M)\right)_{n=1}^\infty$ converges in L^2 there exists a subsequence $(\Delta_{n_k})_{k=1}^\infty$ such that $T^{\Delta_{n_k}}(M)$ converges uniformly (over $[0,a]$) a.s. to the limit $[M]$, i.e.

$$\sup_{t\le a}\left|T^{\Delta_{n_k}}(M) - [M]_t\right| \to 0 \text{ as } k \to \infty \text{ a.s.}$$

Since the uniform limit of a continuous function is itself continuous, it follows that $[M]$ is continuous. Choosing the particular partition

$$\Delta_n = \left\{\frac{ka}{2^n} : k = 0, 1, 2, ...\right\},$$

so that $\Delta_n \subset \Delta_{n+1}$ and $\cup_n \Delta_n$ is dense in $[0, a]$, we see that for any $s < t$ both in $\cup_n \Delta_n$ there must exist N such that s and t are both in Δ_n for $n \geq N$. Thus,

$$T_s^{\Delta_n}(M) \leq T_t^{\Delta_n}(M) \text{ for all } n \geq N$$

and by taking limits we see that $[M]$ is increasing on $\cup_n \Delta_n$. Since $\cup_n \Delta_n$ is dense in $[0, a]$ and M is continuous it must be increasing everywhere. Finally, to verify the martingale property for $M^2 - [M]$ we recall from (22) that for all $s < t$

$$\mathbb{E}\left[M_t^2 - T_t^{\Delta}(M) \middle| \mathcal{F}_s \right] = M_s^2 - T_s^{\Delta}(M) \text{ a.s.}$$

By taking limits and using appropriate conditional limit theorem (exercise: work through the details) we deduce that

$$\mathbb{E}\left[M_t^2 - [M]_t \middle| \mathcal{F}_s \right] = M_s^2 - [M]_s \text{ a.s.}$$

as required. ■

26 Continuous local martingales

We will need to extended the definition of quadratic variation from continuous bounded martingales to continuous local martingales. We first prove a result which illustrates the difference between the set of martingales and local martingales. The result will rely on the following theorem, which we will state without proof.

Theorem 26.1 (Optional stopping theorem) *The following are equivalent:*

1. *X is a martingale*
2. *For every stopping time T, the stopped process $X^T := (X_{t\wedge T})_{t\geq 0}$ is a martingale.*
3. *$\mathbb{E}[X_T | \mathcal{F}_S] = X_{T\wedge S}$ for all bounded stopping times S and T.*
4. *$\mathbb{E}[X_T] = X_0$ for all bounded stopping times T.*

With this result in hand, we may prove

Proposition 26.2 *The following are equivalent:*

1. *X is a martingale.*
2. *X is a local martingale and for every $t > 0$ the family of random variables*

$$\mathcal{R}_t = \{X_T : T \text{ is a stopping time}, T \leq t\}$$

is uniformly integrable.

Proof. If we assume 1, then X is trivially a local martingale (martingale implies local martingale). Let $t > 0$ and suppose $T \leq t$ is a stopping time, then by the optional stopping theorem (OST) we have

$$X_T = \mathbb{E}[X_t | \mathcal{F}_T] \text{ a.s.}$$

therefore

$$\mathcal{R}_t \subseteq \{\mathbb{E}[X_t | \mathcal{G}] : \mathcal{G} \text{ is a sub } \sigma\text{-field of } \mathcal{F}\}.$$

Since we have shown that the latter is UI, and subsets of UI families are themselves UI we have that 2 holds.

Now we start by assuming 2. From the optional stopping theorem it is enough to prove that $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ for every bounded stopping time T. Thus, suppose $T \leq t$ and let $(T_n)_{n=1}^{\infty}$ be a sequence of stopping times such that $T_n \to \infty$ a.s. and such that each X^{T_n} is a martingale. Then $X_T^{T_n} = X_{T_n \wedge T} \to X_T$ a.s. as $n \to \infty$, and since

$$\left\{X_T^{T_n} : n = 1, 2....\right\} \subseteq \mathcal{R}_t$$

and $\mathcal{R}_t$ is assumed UI we have $\mathbb{E}\left[X_T^{T_n}\right] \to \mathbb{E}\left[X_T\right]$ a.s. as $n \to \infty$. On the other hand the OST gives for every n that

$$\mathbb{E}\left[X_T^{T_n}\right] = \mathbb{E}\left[X_0^{T_n}\right] = \mathbb{E}\left[X_0\right],$$

and hence $\mathbb{E}\left[X_T\right] = \mathbb{E}\left[X_0\right]$. ■

It is useful to be able to "reduce" a local martingale to a martingale along a sequence of stopping times which simultaneously make the stopped processes bounded. The following proposition ensures we can do this.

Proposition 26.3 *Let M be a continuous local martingale, and for $n = 1, 2, ...$ define*

$$S_n := \inf\{t \geq 0 : |M_t| \geq n\}. \tag{23}$$

Then $(S_n)_{n=1}^{\infty}$ is a sequence of stopping times such that (i) $S_n \to \infty$ a.s. as $n \to \infty$, and (ii) for each n, the stopped process M^{S_n} is a martingale.

Proof. That S_n is a stopping times is easily deduce from that fact that M is continuous and that $\{x : |x| \geq n\}$ is a closed subset.

1. To see that $S_n \to \infty$ a.s. as $n \to \infty$, we observe from the continuity of M that $\sup_{s \leq t} |M_s| < \infty$. Hence $S_n > t$ for every $n > \sup_{s \leq t} |M_s|$; that is, we have shown a.s. for every $t > 0$ there exist n such that $S_n > t$, which is what we needed to show.

2. Finally we prove that M^{S_n} is a martingale we let (T_n) be a sequence of stopping times with $T_n \to \infty$ a.s. such that M^{T_n} is a martingale (note: such a sequence exists, since M is assumed to be a local martingale). We then observe that because M^{T_n} is a martingale

$$\left(M^{S_n}\right)^{T_n} = M^{S_n \wedge T_n} = \left(M^{T_n}\right)^{S_n}$$

is also a martingale by the OST. Thus, (T_n) also reduces M^{S_n} to a martingale and it follows that M^{S_n} is a local martingale. To prove that M^{S_n} is a martingale we let $t > 0$ and observe that for any stopping time $T \leq t$ the random variable $\left|M_T^{S_n}\right| = |M_{T \wedge S_n}| \leq n$ it follows that

$$\mathcal{R}_t = \left\{M_T^{S_n} : T \text{ is a stopping time}, T \leq t\right\}$$

is a UI family. By the previous proposition, M^{S_n} is a martingale.

■

27 Quadratic variation of continuous local martingales

The results in the last lecture allow us to extend the definition of $[M]$ to the case where M is a continuous local martingale. We first need the following elementary lemma.

Lemma 27.1 *Let M be a bounded continuous martingale and let T be a stopping time then $\left[M^T\right] = [M]^T$.*

Proof. Note by OST M^T is again a bounded continuous martingale, hence $\left[M^T\right]$ is defined. By definition $M^2 - [M]$ is martingale, thus, again by OST, the stopped process $\left(M^2 - [M]\right)^T = \left(M^T\right)^2 - [M]^T$ is a martingale. From the uniqueness of the quadratic variation process we have $\left[M^T\right] = [M]^T$ ■

The previous lemma and proposition allow us to define the quadratic variation process for a given continuous local martingale.

Definition 27.2 *Let M be a continuous local martingale. The quadratic variation of M is defined by*

$$[M]_t := \left[M^{S_n}\right]_t \text{ for } t < S_n,$$

where S_n is the stopping time defined in (23).

This is well-defined because if $t < S_n \leq S_m$ then using lemma 27.1 we have

$$\left[M^{S_m}\right]_t = \left[M^{S_m}\right]_{t\wedge S_n} = \left[M^{S_m}\right]_t^{S_n} = \left[M^{S_m \wedge S_n}\right]_t = \left[M^{S_n}\right]_t$$

for all $t \geq 0$. As for the bounded case $[M]$ is the limit of the approximations $T_t^{\Delta}(M)$.

Proposition 27.3 *Let M be a continuous local martingale, then $[M]$ is the unique continuous adapted non-decreasing process such that $[M]_0 = 0$ and $M^2 - [M]$ is a continuous local martingale. For every $t > 0$ we have that*

$$\sup_{s\leq t} \left|T_s^{\Delta}(M) - [M]_s\right| \xrightarrow{\mathbb{P}} 0 \tag{24}$$

as $|\Delta| \to 0$.

Proof. The only non-trivial thing to check is (24), everything else follows from the definition of $[M]$. Let $S_n \leq S_{n+1} \uparrow \infty$ a.s. be the reducing for M defined by (23). Given $\epsilon, \delta > 0$ we need to show that

$$\mathbb{P}\left(\sup_{s\leq t} \left|T_s^{\Delta}(M) - [M]_s\right| > \epsilon\right) \leq \delta$$

for all $|\Delta|$ sufficiently small. We first select N such that for all $k \geq N$ we have $\mathbb{P}(S_k \leq t) \leq \delta/2$. Then

$$\begin{aligned}
& \mathbb{P}\left(\sup_{s\leq t}\left|T_s^{\Delta}(M) - [M]_s\right| > \epsilon\right) \\
\leq\ & \mathbb{P}(S_N \leq t) + \mathbb{P}\left(\sup_{s\leq t}\left|T_s^{\Delta}(M) - [M]_s\right| > \epsilon, S_N > t\right) \\
\leq\ & \delta/2 + \mathbb{P}\left(\sup_{s\leq t}\left|T_s^{\Delta}\left(M^{S_N}\right) - \left[M^{S_N}\right]_s\right| > \epsilon\right).
\end{aligned}$$

By definition M^{S_N} is a bounded martingale and the second term can be bounded by $\delta/2$ for all partitions Δ such that $|\Delta|$ is sufficiently small. ■

28 Itô's theory of stochastic integration

Let $M \in \mathcal{M}_{\mathrm{c,loc}}$ be a continuous local martingale. Our aim now is to understand the stochastic integral

$$\int H_s dM_s$$

for a wide class of stochastic processes H. As usual, we start by defining the stochastic integral for processes for which the integral is easy to interpret.

28.1 Stochastic integration of elementary processes

A process $H = (H_t)_{t\geq 0}$ is called a bounded elementary process if it has the form

$$H_t = \sum_{k=1}^{n} Z_k 1_{(S_k, T_k]}(t),$$

where: (i) $0 \leq S_1 \leq T_1 \leq S_2 \leq T_2 \leq .... \leq S_n \leq T_n$, (ii) Z_k is $\mathcal{F}_{S_k}$-measurable for every k , and (iii) for some $M < \infty$ we have $\forall k$, $|Z_k| \leq M$ a.s.

Notation 28.1 *We denote the class of bounded elementary processes by* $b\xi$.

Note, in particular, that any $H \in \xi$ a.s has left continuous sample paths $t \mapsto H_t(\omega)$ which have (finite) left-limits and that H is a bounded process ($\sup_{t\geq 0} |H_t| \leq M$). We let

$$||H||_\infty = \inf\left\{M > 0 : \sup_{t\geq 0} |H_t| \leq M\right\}$$

which, by definition, is finite whenever H is bounded elementary. The stochastic integral of H w.r.t. M is defined to be the stochastic process

$$(H \cdot M)_t = \sum_{k=1}^{n} Z_k (M_{t\wedge T_k} - M_{t\wedge S_k}), \tag{25}$$

where the conditions ensure the RHS of (25) is a finite sum for $t \in \mathbb{R}_+$. We have seen the a similar process $H \cdot M$ before in the form of the martingale transform of a discrete-time previsible process w.r.t. a (sub/super) martingale. The following lemma characterises the basic properties of $H \cdot M$.

Lemma 28.2 *Suppose* $M \in \mathcal{M}_{c,loc}, H \in b\xi$ *and let* T *be a stopping time. The following properties hold:*

1. $(H \cdot M)^T = H \cdot M^T$
2. $H \cdot M \in \mathcal{M}_{c,loc}$
3. *If* $M \in \mathcal{M}_c^2$ *then* $H \cdot M \in \mathcal{M}_c^2$ *and* $||H \cdot M||_2 \leq ||H||_\infty ||M||_2$

Proof. Exercise. ■

28.2 Stochastic integration for $\mathcal{M}_{\mathbf{c}}^2$

Suppose now that $M \in \mathcal{M}_{\mathrm{c}}^2$. We will extend the definition of the stochastic integral presented in the last section. We first introduce the following σ-field on the set $(0,\infty) \times \Omega$

Definition 28.3 *The previsible σ-field $\mathcal{P}$ is the smallest σ-field on $(0,\infty)\times\Omega$ such that every adapted caglad (left-continuous with right limits) process is $\mathcal{P}$-measurable.*

Recall that a stochastic process $H = (H_t)_{t>0}$ may be viewed as a function $H : (0,\infty)\times\Omega \to \mathbb{R}$. A stochastic process $H = (H_t)_{t>0}$ is called previsible if it is a $\mathcal{P}$-measurable function when viewed in this way. Every bounded elementary process is previsible. Fruthermore, it can be shown that $\sigma(\mathrm{b}\xi) = \mathcal{P}$.

Example 28.4 *Let $T < \infty$ be a stopping time then $(t,\omega) \mapsto 1_{[0,T(\omega)]}(t)$ is previsble. To see this note that for every ω and t*

$$1_{[0,T(\omega)]}(t) = \lim_{n\to\infty} 1_{[0,T^n(\omega)]}(t)$$

where

$$T^n(\omega) = \frac{1}{2^n}\lceil 2^n T(\omega)\rceil \wedge n.$$

Because T^n takes values in the finite set $\{k2^{-n} : k = 1, ..., n2^n\}$ with

$$\begin{aligned} A_{k,n} &= \left\{T^n = kn2^{-n}\right\} = \left\{T \in [(k-1)\,n2^{-n}, kn2^{-n})\right\} \\ A_{2^n,n} &= \{T \geq n\} \end{aligned}$$

It follows that $\left\{1_{[0,T^n]}(t) = 0\right\} = \{T^n > t\} = \{T > jn2^{-n}\}$ where $j2^{-n} \leq t < (j+1)\,2^{-n}$. Since $\{T > jn2^{-n}\} \in \mathcal{F}_t$ it follows that $1_{[0,T^n]}$ is an adapted process. It is also plainly left-continuous and therefore $\mathcal{P}$-measurable. It follows that $1_{[0,T]}$ is also $\mathcal{P}$-measurable (as the limits of functions which are themselves so).

29 Itô's isometry

Definition 29.1 ($L^2(M)$) *Let $M \in \mathcal{M}_c^2$ we define the space $L^2(M)$ to the be the class of previsible processes H for which*

$$||H||_{L^2(M)} := \mathbb{E}\left[\int_0^\infty H_s^2 d[M]_s\right]^{1/2} < \infty$$

The space $L^2(M)$ plays an important role in the definition of the Itô integral. It is a Hilbert space when we equip it with the inner product

$$\langle H, G\rangle_{L^2(M)} = \mathbb{E}\left[\int_0^\infty H_s G_s d[M]_s\right],$$

in fact it can be shown that it is the L^2 space $L^2((0,\infty)\times\Omega, \mathcal{P}, \mu)$ where μ is the measure obtained by extending

$$\mu((s,t]\times A) = \mathbb{E}[1_A([M]_t - [M]_s)], \quad A \in \mathcal{F}_s,\ s \le t$$

using the Carathéodory extension theorem. The key facts about $L^2(M)$ are the following:

1. the space $\mathcal{U} = \text{b}\xi \cap L^2(M)$ is a dense linear subspace of $L^2(M)$. i.e. the closure of $\mathcal{U}$ is $L^2(M)$ itself.
2. the map $H \mapsto H \cdot M$ is an isometry from $\mathcal{U}$ into $\mathcal{M}_c^2$.

The proof of the first of these facts can be found, e.g., Rogers and Williams vol 2. For the second fact we have:

Theorem 29.2 (Itô's isometry) *Let $M \in \mathcal{M}_c^2$. There exists a unique Hilbert space isometry $I : L^2(M) \to \mathcal{M}_c^2$ such that $I(H) = H \cdot M$ for all $H \in b\xi$.*

Proof. Let $H \in \text{b}\xi$ be of the form

$$H_t = \sum_{k=1}^n Z_k 1_{(S_k, T_k]}(t).$$

We will show that

$$||H \cdot M||_2 = \mathbb{E}\left[(H \cdot M)_\infty^2\right]^{1/2} = ||H||_{L^2(M)} \tag{26}$$

because $L^2(M)$ and $\mathcal{M}_c^2$ are both Hilbert spaces, I then extends uniquely to an isometry defined on the closure of ξ w.r.t. $||\cdot||_{L^2(M)}$. From the previous lemma, this closure is $L^2(M)$.

To check (26) we have

$$
\begin{aligned}
& \mathbb{E}\left[(H\cdot M)_\infty^2\right] \\
= \ & \mathbb{E}\left[\left(\sum_{k=1}^n Z_k\left(M_{T_k}-M_{S_k}\right)\right)^2\right] \\
= \ & \mathbb{E}\left[\sum_{k=1}^n Z_k^2\left(M_{T_k}-M_{S_k}\right)^2\right]+2\mathbb{E}\left[\sum_{j<k}^n Z_k Z_j\left(M_{T_k}-M_{S_k}\right)\left(M_{T_j}-M_{S_j}\right)\right] \\
= \ & \sum_{k=1}^n \mathbb{E}\left[Z_k^2\left(M_{T_k}-M_{S_k}\right)^2\right]+2\sum_{j<k}^n \mathbb{E}\left[Z_k Z_j\left(M_{T_k}-M_{S_k}\right)\left(M_{T_j}-M_{S_j}\right)\right],
\end{aligned} \tag{27}
$$

We further simplify by

$$
\begin{aligned}
\mathbb{E}\left[Z_k^2\left(M_{T_k}-M_{S_k}\right)^2\right] &= \mathbb{E}\left[\mathbb{E}\left[Z_k^2\left(M_{T_k}-M_{S_k}\right)^2\middle|\mathcal{F}_{S_k}\right]\right] \\
&= \mathbb{E}\left[Z_k^2\mathbb{E}\left[\left(M_{T_k}-M_{S_k}\right)^2\middle|\mathcal{F}_{S_k}\right]\right] \\
&= \mathbb{E}\left[Z_k^2\mathbb{E}\left[M_{T_k}^2-M_{S_k}^2\middle|\mathcal{F}_{S_k}\right]\right] \\
&= \mathbb{E}\left[Z_k^2\left([M]_{T_k}-[M]_{S_k}\right)\right],
\end{aligned}
$$

and for $j<k$

$$
\begin{aligned}
& \mathbb{E}\left[Z_k Z_j\left(M_{T_k}-M_{S_k}\right)\left(M_{T_j}-M_{S_j}\right)\right] \\
= \ & \mathbb{E}\left[\mathbb{E}\left[Z_k Z_j\left(M_{T_k}-M_{S_k}\right)\left(M_{T_j}-M_{S_j}\right)\middle|\mathcal{F}_{S_k}\right]\right] \\
= \ & \mathbb{E}\left[Z_k Z_j\left(M_{T_j}-M_{S_j}\right)\mathbb{E}\left[\left(M_{T_k}-M_{S_k}\right)\middle|\mathcal{F}_{S_k}\right]\right] \\
= \ & 0.
\end{aligned}
$$

Making use of this in (27) we obtain

$$
\begin{aligned}
||H\cdot M||_2 &= \mathbb{E}\left[(H\cdot M)_\infty^2\right]=\sum_{k=1}^n \mathbb{E}\left[Z_k^2\left([M]_{T_k}-[M]_{S_k}\right)\right] \\
&= \mathbb{E}\left[\sum_{k=1}^n Z_k^2\left([M]_{T_k}-[M]_{S_k}\right)\right] \\
&= \mathbb{E}\left[\sum_{k=1}^n Z_k^2\left([M]_{T_k}-[M]_{S_k}\right)\right] \\
&= \mathbb{E}\left[\int_0^\infty H_s^2 d\,[M]_s\right] \\
&= ||H||_{L^2(M)}^2 \ .
\end{aligned}
$$

∎

Definition 29.3 *For a previsible process $H \in L^2(M)$ we define the Itô stochastic integral of H w.r.t. M by*

$$(H \cdot M)_t := \int_0^t H_u dM_u = I(H)_t$$

30 Localisation of stochastic integrals

We will use the following lemma without proof.

Lemma 30.1 *Let M be in $\mathcal{M}_c^2$, $H \in L^2(M)$ and T a stopping time. Then we have $\left(H1_{[0,T]} \cdot M\right) = H \cdot M^T = (H \cdot M)^T$.*

Definition 30.2 *We say a previsible process is locally bounded if there exists a sequence of stopping times $T_n \leq T_{n+1} \uparrow \infty$ a.s. such that $H1_{[0,T_n]}$ is bounded and previsible.*

We have already seen that any continuous local martingale M can be reduced to a continuous bounded martingale along a reducing sequence $(S_n)_{n=1}^{\infty}$. By taking R_n to the stopping time $S_n \wedge T_n \uparrow \infty$ we can simultaneously reduce H to a bounded previsible process and M to a bounded continuous martingale. This allows us to generalise the definition of the stochastic integral by setting

$$\int H_u dM_u := (H \cdot M)_t := \left(H1_{[0,R_n]} \cdot M^{R_n}\right)_t \text{ for } t < R_n.$$

It is not difficult to check this is well-defined by verifying that

$$\left(H1_{[0,R_n]} \cdot M^{R_n}\right)_t = \left(H1_{[0,R_m]} \cdot M^{R_m}\right)_t \text{ whenever } t < R_n \leq R_m.$$

Proposition 30.3 *Let M be a continuous local martingale, H a locally bounded previsible process and T a stopping time . Then we have*

1. $\left(H1_{[0,T]} \cdot M\right) = H \cdot M^T = (H \cdot M)^T$;
2. *$H \cdot M$ is a continuous local martingale;*
3. $[H \cdot M] = H^2 \cdot [M] := \int_0^{\cdot} H_u^2 d\,[M]_u$;
4. *If K is a locally bounded previsible process then $H \cdot (K \cdot M) = (HK) \cdot M$.*

Proof. Assertion 1 follows from Lemma 30.1 by a simple localisation argument. 2 is similarly straight-forward using the definition of $H \cdot M$. For 3 we first assume by localisation in addition that H is bounded previsible and M is bounded. We need to prove that $H \cdot M - H^2 \cdot [M]$ is a martingale; for the uniqueness of the quadratic variation process then gives $[H \cdot M] = H^2 \cdot [M]$. From the optional stopping theorem this will follow if we can show that $\mathbb{E}\left[(H \cdot M)_S - \left(H^2 \cdot [M]\right)_S\right] = 0$ for every bounded stopping time S.To do we observe

$$(H \cdot M)_S = (H \cdot M)_\infty^S = \left(H \cdot M^S\right)_\infty$$

and, similarly,

$$\left(H^2 \cdot [M]\right)_S = \left(H^2 \cdot [M]\right)_\infty^S = \left(H^2 \cdot [M]^S\right)_\infty = \left(H^2 \cdot \left[M^S\right]\right)_\infty.$$

From the Itô isometry applied to the (bounded) martingale M and the bounded previsible process $H \in L^2(M)$ we obtain

$$\begin{aligned} \mathbb{E}\left[\left(H \cdot M^S\right)_\infty^2\right] &= \|H\|_{L^2(M^S)} \\ &= \mathbb{E}\left[\int_0^\infty H^2 d\left[M^S\right]\right] \\ &= \mathbb{E}\left[\left(H^2 \cdot \left[M^S\right]\right)_\infty\right], \end{aligned}$$

after which the proof is complete.

For the final claim, we again localise and assume H, K are bounded previsible and M is bounded. We take two sequences (H_n) and (K_n) in ξ which approximate H and K respectively in $L^2(M)$ and do so boundedly, i.e.

$$\|H - H_n\|_{L^2(M)} \to 0 \text{ and } \|K - K_n\|_{L^2(M)} \to 0 \text{ as } n \to \infty$$

and $\forall n$ $\|H_n\|_\infty \leq 2\|H\|_\infty, \|K_n\|_\infty \leq 2\|K\|_\infty$. It is easy to check for elementary processes that 4 holds, thus

$$H_n \cdot (K_n \cdot M) = (H_n K_n) \cdot M$$

holds for all n. To finish we notice that

$$\begin{aligned} & \|H_n \cdot (K_n \cdot M) - H \cdot (K \cdot M)\|_2 \\ \leq\ & \|(H_n - H) \cdot (K_n \cdot M)\|_2 + \|H \cdot ((K_n - K) \cdot M)\|_2 \\ =\ & \mathbb{E}\left[\int_0^\infty (H_n(s) - H(s))^2 d[K_n \cdot M]_s\right]^{1/2} + \mathbb{E}\left[\int_0^\infty H(s)^2 d[(K_n - K) \cdot M]_s\right]^{1/2} \\ =\ & \mathbb{E}\left[\int_0^\infty (H_n(s) - H(s))^2 K_n(s)^2 d[M]_s\right]^{1/2} + \mathbb{E}\left[\int_0^\infty H(s)^2 (K_n(s) - K(s))^2 d[M]_s\right]^{1/2} \\ \leq\ & 2\|K\|_\infty \|H - H_n\|_{L^2(M)} + \|H\|_\infty \|K - K_n\|_{L^2(M)} \to 0 \text{ as } n \to \infty, \end{aligned}$$

where the fourth line uses assertion 3 of the proposition. A similar argument shows that $\|(H_n K_n) \cdot M - (HK) \cdot M\|_2 \to 0$ as $n \to \infty$, whereupon it follows that $H \cdot (K \cdot M) = (HK) \cdot M$ ■

31 Integration-by-parts for stochastic integrals

The aim of this lecture is to prove an integration-by-parts formula for the Itô's integrals. We begin with the following result.

Lemma 31.1 *Suppose M is a continuous local martingale, and H is a locally bounded previsible process which is also left-continuous. Then*

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} H_{\frac{k-1}{2^n}} \left(M_{\frac{k}{2^n}} - M_{\frac{k-1}{2^n}} \right) \xrightarrow{\mathbb{P}} (H \cdot M)_t \tag{28}$$

uniformly on compact subsets of $\mathbb{R}_+$.

Proof. By localisation assume H and M are also bounded. For each $n \in \mathbb{N}$ define a new process $H_s^n(\omega) = H_{\frac{\lfloor 2^n s \rfloor}{2^n}}(\omega)$ then left-continuity ensures that, for every s and ω, $H_s^n(\omega) \to H_s(\omega)$ as $n \to \infty$. The summation on the LHS of (28) equals $(H^n \cdot M)_t$. By using the dominated convergence theorem (in t) we have that

$$\int_0^\infty (H_t^n - H_t)^2 \, d[M]_t \to 0$$

a.s. as $n \to \infty$. The dominated convergence theorem in ω then gives

$$||H - H^n||_{L^2(M)}^2 = \mathbb{E}\left[\int_0^\infty (H_t^n - H_t)^2 \, d[M]_t \right] \to 0$$

as $n \to \infty$. By exploiting Doob's L^2-inequality and Itô's isometry we hence learn that

$$\begin{aligned}
\mathbb{E}\left[\sup_{s \le t} \left((H^n \cdot M)_s - (H \cdot M)_s \right)^2 \right] &\le \mathbb{E}\left[\sup_{s < \infty} \left((H^n \cdot M)_s - (H \cdot M)_s \right)^2 \right] \\
&\le 4\mathbb{E}\left[\left((H^n \cdot M)_\infty - (H \cdot M)_\infty \right)^2 \right] \\
&= 4 \, ||H^n \cdot M - H \cdot M||_2^2 \\
&= 4 \, ||H - H^n||_{L^2(M)}^2 \to 0,
\end{aligned}$$

so that in particular $\sup_{s \le t} \left((H^n \cdot M)_s - (H \cdot M)_s \right) \xrightarrow{\mathbb{P}} 0$, as required. ■

Recall that a adapted process X is a continuous semimartingale if it may be written as $X_t = X_0 + M_t + A_t$ where X_0 is $\mathcal{F}_0$measurable, M is a continuous local martingale $M_0 = 0$ and A is an adapted continuous finite variation process with $A_0 = 0$. Given a locally bounded previsible process H we can now define

$$(H \cdot X)_t := \int_0^t H_s dX_s := \int_0^t H_s dM_s + \int_0^t H_s dA_s,$$

the latter integrals making sense using Itô integration and measure theory respectively. We can also generalise the covaration to the class of continuous semimartigales; given X as above and Y and continuous semimartingale with the decomposition $Y_t = Y_0 + N_t + B_t$ we define the covariation of X and Y to be

$$[X,Y]_t := [M,N]_t := \frac{1}{4}\left([M+N]_t - [M-N]_t\right)$$

(since continuous finite variation processes have zero quadratic variation). It is an easy exercise using what we already know to prove that $[X,Y]$ is again the limit in probability of

$$\sum_{k=1}^{\lfloor 2^n t \rfloor} \left(X_{k2^{-n}} - X_{(k-1)2^{-n}}\right)\left(Y_{k2^{-n}} - Y_{(k-1)2^{-n}}\right)$$

uniformly on compact sets in t. We will use the following identities without proof.

Theorem 31.2 (Kunita-Watanabe) *For any continuous local martingales M and N and any H which is locally bounded and previsible, we have*

$$[H \cdot M, N] = H \cdot [M,N] = [M, H \cdot N].$$

.

The next result is an almost immediate consequence of the existence of $[X,Y]$.

Proposition 31.3 (Integration-by-parts) *Suppose X and Y are continuous semimartigales then*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X,Y]_t \tag{29}$$

Proof. In the case where $X = Y$ the result reads

$$X_t^2 = X_0^2 + 2\int_0^t X_s dX_s + [X]_t\,. \tag{30}$$

It is suffiient only to prove this case since then we would have

$$\begin{aligned}
(X_t + Y_t)^2 &= (X_0 + Y_0)^2 + 2\int_0^t (X_s + Y_s)\, d(X_s + Y_s) + [X+Y]_t\,, \text{ and} && (31)\\
(X_t - Y_t)^2 &= (X_0 - Y_0)^2 + 2\int_0^t (X_s - Y_s)\, d(X_s - Y_s) + [X-Y]_t && (32)
\end{aligned}$$

and the result follows by polarisation by subtracting (32) from (31) and dividing by 4.

To prove (30) we notice by simple algebra that

$$
\begin{aligned}
& \sum_{k=1}^{\lfloor 2^n t \rfloor} \left(X_{k2^{-n}} - X_{(k-1)2^{-n}}\right)^2 \\
= \quad & \sum_{k=1}^{\lfloor 2^n t \rfloor} X_{k2^{-n}} \left(X_{k2^{-n}} - X_{(k-1)2^{-n}}\right) - \sum_{k=1}^{\lfloor 2^n t \rfloor} X_{(k-1)2^{-n}} \left(X_{k2^{-n}} - X_{(k-1)2^{-n}}\right) \\
= \quad & \sum_{k=1}^{\lfloor 2^n t \rfloor} \left(X_{k2^{-n}}^2 - X_{(k-1)2^{-n}}^2\right) - 2 \sum_{k=1}^{\lfloor 2^n t \rfloor} X_{(k-1)2^{-n}} \left(X_{k2^{-n}} - X_{(k-1)2^{-n}}\right) \\
= \quad & X_{\lfloor 2^n t \rfloor 2^{-n}}^2 - X_0^2 - 2 \sum_{k=1}^{\lfloor 2^n t \rfloor} X_{(k-1)2^{-n}} \left(X_{k2^{-n}} - X_{(k-1)2^{-n}}\right).
\end{aligned}
$$

By taking limits on both sides of this equation and using lemma 31.1 we obtain

$$[X]_t = X_t^2 - X_0^2 - 2 \int_0^t X_s dX_s$$

as desired. ■

Remark 31.4 *If X is any continuous semimartingale and Y is a continuous finite variation process then $[X, Y] = 0$ and the integration-by-parts formula from classical calculus holds:*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s.$$

32 Itô's formula

The centrepiece of stochastic integration theory is Itô's formula; this is the key change-of-variable formula for Itô integrals.

Theorem 32.1 (Itô's formula) *Let $X = \left(X^1, ..., X^d\right)$ be a $\mathbb{R}^d$−valued stochastic process such that each component X^i is a continuous semimartingale. Suppose f is a twice continuously differentiable function from $\mathbb{R}^d$ to $\mathbb{R}^d$ i.e. is in $C^2\left(\mathbb{R}^d\right)$, then $f(X)$ is a continuous semimartingale and*

$$f(X_t) = f(X_0) + \sum_{i=1}^{d} \int_0^t \frac{\partial f(X_s)}{\partial x_i} dX_s^i + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 f(X_s)}{\partial x_i \partial x_j} d\left[X^i, X^j\right]_s . \quad (33)$$

Proof. (**Sketch**) The case $f(x) = x^i$ is trivial because the second-order derivatives vanish. Assuming that we have proved (33) for a function $g \in C^2\left(\mathbb{R}^d\right)$ we prove that it continues to hold for the function $f(x) = x^k g(x)$. To do this we first use integration-by-parts to see

$$X_t^k g(X_t) = X_0^k g(X_0) + \int_0^t g(X_s)\, dX_s^k + \int_0^t X_s^k dg(X_s) + \left[X^k, g(X)\right]_t . \quad (34)$$

Then since by assumption (33) holds for $g(X)$ we can compute $\left[X^k, g(X)\right]$ using Kunita-Watanabe

$$\left[X^k, g(X)\right]_t = \left[X^k, \int \sum_{i=1}^{d} \frac{\partial g(X_s)}{\partial x_i} dX_s^i\right]_t = \int_0^t \sum_{i=1}^{d} \frac{\partial g(X_s)}{\partial x_i} d\left[X^k, X^i\right]_s \quad (35)$$

and also obtain that

$$\int_0^t X_s^k dg(X_s) = \sum_{i=1}^{d} \int X_s^k \frac{\partial g(X_s)}{\partial x_i} dX_s^i + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t X_s^k \frac{\partial^2 g(X_s)}{\partial x_i \partial x_j} d\left[X^i, X^j\right]_s . \quad (36)$$

Substituting (35) and (36) into (34) gives (33) for $f(x) = x^k g(x)$.
Proof. By induction (33) holds for any multi-variable polynomial in $\left(x^1, ..., x^d\right)$. By letting K be any compact subset of $\mathbb{R}^d$ containing X_0 and by stopping X if necessary we may assume that X stays inside K. If f is any function in $C^2(K)$ then there exists a sequence of multivariable polynomial functions (f^n) such that $f^n, Df^n, D^2 f^n$ approximate $f, Df, D^2 f$ uniformly on K (Stone-Weierstrass theorem). By taking limits in Itô's formula for the approximations and using the uniform convergence of the stochastic integrals in probability we obtain the stated result for f. ■ ■

If any of the components of $X = \left(X^1, ..., X^d\right)$ has finite variation, then we can relax the assumption on f to be only C^1 in these components.

Example 32.2 *Let M be a continuous local martingale. Suppose $f \in C^{1,2}(\mathbb{R} \times \mathbb{R})$ satisfies*

$$\frac{\partial f}{\partial x}(x, y) + \frac{1}{2} \frac{\partial^2 f}{\partial y^2}(x, y) = 0$$

then $f\left([M]_t, M_t\right)$*is a continuous local martingale. This follows from Itô's formula since*

$$\begin{aligned} f\left([M]_t, M_t\right) &= f\left(0, M_0\right)+\int_0^t \frac{\partial f}{\partial x}\left([M]_s, M_s\right) d\,[M]_s+\int_0^t \frac{\partial f}{\partial y}\left([M]_s, M_s\right) dM_s \\ &\quad +\frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial y^2}\left([M]_s, M_s\right) d\,[M]_s \\ &= f\left(0, M_0\right)+\int_0^t \frac{\partial f}{\partial y}\left([M]_s, M_s\right) dM_s . \end{aligned}$$

The most useful instance of this is in showing that $\exp\left(\theta M_t-\frac{\theta^2}{2}[M]_t\right)$ *is a continuous local martingale.*

33 Exponential martingales

Theorem 33.1 *Suppose* $X = \left(X^1, ..., X^d\right)$ *is a continuous local martingale such that* $X_0 = 0$ *and with* $\left[X^i, X^j\right]_t = t$ *if* $i = j$ *and* $\left[X^i, X^j\right]_t = 0$ *if* $i \neq j$*, then* X *is a* $d-$*dimensional Brownian motion.*

Proof. We need to prove for $s < t$ that $X_t - X_s$ is independent of $\mathcal{F}_s = \sigma\left(X_u : u \leq s\right)$, and has the distribution of a multi-variate normal random variable with mean the zero vector and covariance matrix $(t-s)\, I_d$. By the uniqueness of the characteristic function these properties will follow if we can show that

$$\mathbb{E}\left[\exp\left(i\left\langle \theta, X_t - X_s\right\rangle\right)\middle| \mathcal{F}_s\right] = \exp\left(-\frac{1}{2}\left|\theta\right|^2 (t-s)\right),$$

or, equivalently, if

$$\mathbb{E}\left[\exp\left(i\left\langle \theta, X_t\right\rangle - \frac{1}{2}\left|\theta\right|^2 t\right)\middle| \mathcal{F}_s\right] = \exp\left(i\left\langle \theta, X_s\right\rangle - \frac{1}{2}\left|\theta\right|^2 s\right),$$

which is just the condition for $Y_u := \exp\left(i\left\langle \theta, X_u\right\rangle - \frac{1}{2}\left|\theta\right|^2 u\right)$ to be a martingale on $[0,t]$. Using Itô's formula and the condition on $\left[X^i, X^j\right]$ we have that

$$Y_t = 1 + i\int_0^t Y_u \theta^j dX_u^j,$$

which is a continuous local martingale. But Y is bounded on $[0,t]$ hence it must be a martingale (bounded continuous local martingales are martingales). ■

If M is a continuous local martingale then for any real θ the process $Z_t = \exp\left(\theta M_t - \frac{1}{2}\theta^2 \left[M\right]_t\right)$ is a continuous local martingale which satisfies the stochastic differential equation $dZ_t = \theta Z_t dM_t$. We call Z the stochastic exponential of θM; it is often denoted as $\xi\left(\theta M\right)$.

Theorem 33.2 *Suppose* M *is a continuous local martingale with* $M_0 = 0$, *then for any* $\epsilon > 0$ *and any* $\delta > 0$

$$\mathbb{P}\left(\sup_{t\geq 0} M_t \geq \epsilon, \left[M\right]_\infty \leq \delta\right) \leq \exp\left(-\frac{\epsilon^2}{2\delta}\right).$$

Proof. For every $\theta \geq 0$ the process $\exp\left(\theta M_t - \frac{1}{2}\theta^2 \left[M\right]_t\right)$ is a continuous local martingale. Define $T = \inf\left\{t \geq 0 : M_t > \epsilon\right\}$, a stopping time and let

$$Z_t = \exp\left(\theta M_t - \frac{1}{2}\theta^2 \left[M\right]_t\right)^T = \exp\left(\theta M_t^T - \frac{1}{2}\theta^2 \left[M^T\right]_t\right).$$

Z is once again in $\mathcal{M}_{\mathrm{c,loc}}$ but it is also bounded and therefore it is a martingale. In particular, $\mathbb{E}[Z_\infty] = \mathbb{E}[Z_0] = 1$. It follows that

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\geq 0} M_t \geq \epsilon, [M]_\infty \leq \delta\right) &\leq \mathbb{P}\left(Z_\infty \geq \exp\left(\theta\epsilon - \frac{\theta^2\delta}{2}\right)\right) \\
&\leq \exp\left(\theta\epsilon - \frac{\theta^2\delta}{2}\right) \\
&=: \psi(\theta).
\end{aligned}$$

Differentiating we find that ψ is minimised over $\theta \geq 0$ for $\theta = \epsilon/\delta$. ■