Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 1
Measurable spaces

Families of Sets

Definition 1.1 (Order properties)**.** A countable[1] family $\{A_n\}_{n\in\mathbb{N}}$ of subsets of a non-empty set S is said to be

1. **increasing** if $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$,
2. **decreasing** if $A_n \supseteq A_{n+1}$ for all $n \in \mathbb{N}$,
3. **pairwise disjoint** if $A_n \cap A_m = \emptyset$ for $m \neq n$,
4. a **partition** of S if $\{A_n\}_{n\in\mathbb{N}}$ is pairwise disjoint and $\cup_n A_n = S$.

[1] To make sure we are all on the same page, let us fix some notation and terminology:

- $\subseteq$ denotes a subset (not necessarily proper).
- A set A is said to be **countable** if there exists an injection (one-to-one mapping) from A into $\mathbb{N}$. Note that finite sets are also countable. Sets which are not countable are called **uncountable**.
- For two functions $f : B \to C$, $g : A \to B$, the **composition** $f \circ g : A \to C$ of f and g is given by
$$(f \circ g)(x) = f(g(x)),$$
for all $x \in A$.
- $\{A_n\}_{n\in\mathbb{N}}$ denotes a sequence. More generally, $(A_\gamma)_{\gamma\in\Gamma}$ denotes a collection indexed by the set Γ.
- We use the notation $A_n \nearrow A$ to denote that the sequence $\{A_n\}_{n\in\mathbb{N}}$ is increasing and $A = \cup_n A_n$. Similarly, $A_n \searrow A$ means that $\{A_n\}_{n\in\mathbb{N}}$ is decreasing and $A = \cap_n A_n$.

Here is a list of some properties that a family $\mathcal{S}$ of subsets of a nonempty set S can have:

(A1) $\emptyset \in \mathcal{S}$,

(A2) $S \in \mathcal{S}$,

(A3) $A \in \mathcal{S} \Rightarrow A^c \in \mathcal{S}$,

(A4) $A, B \in \mathcal{S} \Rightarrow A \cup B \in \mathcal{S}$,

(A5) $A, B \in \mathcal{S}$, $A \subseteq B \Rightarrow B \setminus A \in \mathcal{S}$,

(A6) $A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$,

(A7) $A_n \in \mathcal{S}$ for all $n \in \mathbb{N} \Rightarrow \cup_n A_n \in \mathcal{S}$,

(A8) $A_n \in \mathcal{S}$, for all $n \in \mathbb{N}$ and $A_n \nearrow A$ implies $A \in \mathcal{S}$,

(A9) $A_n \in \mathcal{S}$, for all $n \in \mathbb{N}$ and $\{A_n\}_{n\in\mathbb{N}}$ is pairwise disjoint implies $\cup_n A_n \in \mathcal{S}$,

Definition 1.2 (Families of sets)**.** A family $\mathcal{S}$ of subsets of a non-empty set S is called an

1. **algebra** if it satisfies (A1),(A3) and (A4),

2. σ**-algebra** if it satisfies (A1), (A3) and (A7)
3. π**-system** if it satisfies (A6),
4. λ**-system** if it satisfies (A2), (A5) and (A8).

Problem 1.1. Show that:

1. Every σ-algebra is an algebra.
2. Each algebra is a π-system and each σ-algebra is an algebra and a λ-system.
3. A family $\mathcal{S}$ is a σ-algebra if and only if it satisfies (A1), (A3), (A6) and (A9).
4. A λ-system which is a π-system is also a σ-algebra.
5. There are π-systems which are not algebras.
6. There are algebras which are not σ-algebras
7. There are λ-systems which are not π-systems.

Hint: Pick all finite subsets of an infinite set. That is not an algebra yet, but sets can be added to it so as to become an algebra which is not a σ-algebra.

Definition 1.3 (Generated σ-algebras)**.** For a family $\mathcal{A}$ of subsets of a non-empty set S, the intersection of all σ-algebras on S that contain $\mathcal{A}$ is denoted by $\sigma(\mathcal{A})$ and is called the σ**-algebra generated** [2] **by** $\mathcal{A}$.

Problem 1.2. Show, by means of an example, that the *union* of a family of algebras (on the same S) does not need to be an algebra. Repeat for σ-algebras, π-systems and λ-systems.

Definition 1.4 (Topology)**.** A **topology**[3] on a set S is a family τ of subsets of S which contains $\emptyset$ and S and is closed under finite intersections and arbitrary (countable or uncountable!) unions. The elements of τ are often called the **open sets**. A set S on which a topology is chosen (i.e., a pair (S, τ) of a set and a topology on it) is called a **topological space**.

Definition 1.5 (Borel σ-algebras)**.** If (S, τ) is a topological space, then the σ-algebra $\sigma(\tau)$, generated by all open sets, is called the **Borel** σ-algebra on (S, τ) or, less precisely[4], the Borel σ-algebra on S.

Example 1.6. Some important σ-algebras. Let S be a non-empty set:

1. The set $\mathcal{S} = 2^S$ (also denoted by $\mathcal{P}(S)$) consisting of all subsets of $\mathcal{S}$ is a σ-algebra.
2. At the other extreme, the family $\mathcal{S} = \{\emptyset, S\}$ is the smallest σ-algebra on S. It is called the **trivial** σ-algebra on $\mathcal{S}$.

[2] Since the family 2^S of *all* subsets of S is a σ-algebra, the concept of a generated σ-algebra is well defined: there is always at least one σ-algebra containing $\mathcal{A}$ - namely 2^S. $\sigma(\mathcal{A})$ is itself a σ-algebra (why?) and it is the smallest (in the sense of set inclusion) σ-algebra that contains $\mathcal{A}$. In the same vein, one can define the algebra, the π-system and the λ-system generated by $\mathcal{A}$. The only important property is that intersections of σ-algebras, π-systems and λ-systems are themselves σ-algebras, π-systems and λ-systems.

[3] Almost all topologies in these notes will be generated by a metric, i.e., a set $A \subset S$ will be open if and only if for each $x \in A$ there exists $\varepsilon > 0$ such that $\{y \in S : d(x,y) < \varepsilon\} \subseteq A$. The prime example is $\mathbb{R}$ where a set is declared open if it can be represented as a union of open intervals.

[4] We often abuse terminology and call S itself a topological space, if the topology τ on it is clear from the context. In the same vein, we often speak of the Borel σ-algebra on a set S.

3. The set $\mathcal{S}$ of all subsets of S which are either countable or whose complements are countable is a σ-algebra. It is called the **countable-cocountable σ-algebra** and is the smallest σ-algebra on S which contains all singletons, i.e., for which $\{x\} \in \mathcal{S}$ for all $x \in S$.

4. The Borel σ-algebra on $\mathbb{R}$ (generated by all open sets as defined by the Euclidean metric on $\mathbb{R}$), is denoted by $\mathcal{B}(\mathbb{R})$.

Problem 1.3. Show that the $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{A})$, for any of the following choices of the family $\mathcal{A}$:

1. $\mathcal{A} = \{$all open subsets of $\mathbb{R}\}$,
2. $\mathcal{A} = \{$all closed subsets of $\mathbb{R}\}$,
3. $\mathcal{A} = \{$all open intervals in $\mathbb{R}\}$,
4. $\mathcal{A} = \{$all closed intervals in $\mathbb{R}\}$,
5. $\mathcal{A} = \{$all left-closed right-open intervals in $\mathbb{R}\}$,
6. $\mathcal{A} = \{$all left-open right-closed intervals in $\mathbb{R}\}$,
7. $\mathcal{A} = \{$all open intervals in $\mathbb{R}$ with rational end-points$\}$, and
8. $\mathcal{A} = \{$all intervals of the form $(-\infty, r]$, where r is rational$\}$.

Hint: An arbitrary open interval $I = (a, b)$ in $\mathbb{R}$ can be written as $I = \cup_{n\in\mathbb{N}}[a + n^{-1}, b - n^{-1}]$.

Measurable mappings

Definition 1.7 (Measurable spaces)**.** A pair $(S, \mathcal{S})$ consisting of a non-empty set S and a σ-algebra $\mathcal{S}$ of its subsets is called a **measurable space**[5].

[5] If $(S, \mathcal{S})$ is a measurable space, and $A \in \mathcal{S}$, we often say that A **is measurable in** $\mathcal{S}$.

Definition 1.8 (Pull-backs and push-forwards)**.** For a function $f : S \to T$ and subsets $A \subseteq S$, $B \subseteq T$, we define the

1. **push-forward** $f(A)$ of $A \subseteq S$ as
$$f(A) = \{f(x) : x \in A\} \subseteq T,$$
2. **pull-back**[6] $f^{-1}(B)$ of $B \subseteq T$ as
$$f^{-1}(B) = \{x \in S : f(x) \in B\} \subseteq S.$$

[6] It is often the case that the notation is abused and the pull-back of B under f is denoted simply by $\{f \in B\}$. This notation presupposes, however, that the domain of f is clear from the context.

Problem 1.4. Show that the pull-back operation preserves the elementary set operations, i.e., for $f : S \to T$, and $B, \{B_n\}_{n\in\mathbb{N}} \subseteq T$,

1. $f^{-1}(T) = S$, $f^{-1}(\emptyset) = \emptyset$,
2. $f^{-1}(\cup_n B_n) = \cup_n f^{-1}(B_n)$,
3. $f^{-1}(\cap_n B_n) = \cap_n f^{-1}(B_n)$, and
4. $f^{-1}(B^c) = [f^{-1}(B)]^c$.

Note: The assumption that the families in 2. and 3. above are countable is not necessary. Uncountable unions or intersections commute with the pull-back, too.

Give examples showing that the push-forward analogues of the statements 1., 3. and 4. above are <u>not true</u>.

Definition 1.9 (Measurability). A mapping $f : S \to T$, where $(S, \mathcal{S})$ and $(T, \mathcal{T})$ are measurable spaces, is said to be $(\mathcal{S}, \mathcal{T})$-**measurable**[7] if $f^{-1}(B) \in \mathcal{S}$ for each $B \in \mathcal{T}$.

[7] When $T = \mathbb{R}$, we tacitly assume that the Borel σ-algebra is defined on T, and we simply call f *measurable*. In particular, a function $f : \mathbb{R} \to \mathbb{R}$, which is measurable with respect to the pair of the Borel σ-algebras is often called a **Borel function**.

Proposition 1.10 (A measurability criterion). *Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be two measurable spaces, and let $\mathcal{C}$ be a subset of $\mathcal{T}$ such that $\mathcal{T} = \sigma(\mathcal{C})$. If $f : S \to T$ is a mapping with the property that $f^{-1}(C) \in \mathcal{S}$, for any $C \in \mathcal{C}$, then f is $(\mathcal{S}, \mathcal{T})$-measurable.*

Proof. Let $\mathcal{D}$ be the family of subsets of T defined by

$$\mathcal{D} = \{B \subset T : f^{-1}(B) \in \mathcal{S}\}.$$

By the assumptions of the proposition, we have $\mathcal{C} \subseteq \mathcal{D}$. On the other hand, by Problem 1.4, the family $\mathcal{D}$ has the structure of the σ-algebra, i.e., $\mathcal{D}$ is a σ-algebra that contains $\mathcal{C}$. Remembering that $\mathcal{T} = \sigma(\mathcal{C})$ is the *smallest* σ-algebra that contains $\mathcal{C}$, we conclude that $\mathcal{T} \subseteq \mathcal{D}$. Consequently, $f^{-1}(B) \in \mathcal{S}$ for all $B \in \mathcal{T}$. $\square$

Problem 1.5. Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be measurable spaces.

1. Suppose that S and T are topological spaces, and that $\mathcal{S}$ and $\mathcal{T}$ are the corresponding Borel σ-algebras. Show that each continuous function $f : S \to T$ is $(\mathcal{S}, \mathcal{T})$-measurable.

Hint: Remember that the function f is continuous if the pull-backs of open sets are open.

2. For a function $f : S \to \mathbb{R}$, show that f is measurable if and only if

$$\{x \in S : f(x) \leq q\} \in \mathcal{S}, \text{ for all rational } q.$$

3. Find an example of $(S, \mathcal{S})$, $(T, \mathcal{T})$ and a measurable function $f : S \to T$ such that $f(A) = \{f(x) : x \in A\} \notin \mathcal{T}$ for all nonempty $A \in \mathcal{S}$.

Proposition 1.11 (Compositions of measurable maps). *Let $(S, \mathcal{S})$, $(T, \mathcal{T})$ and $(U, \mathcal{U})$ be measurable spaces, and let $f : S \to T$ and $g : T \to U$ be measurable functions. Then the composition $h = g \circ f : S \to U$, given by $h(x) = g(f(x))$ is $(\mathcal{S}, \mathcal{U})$-measurable.*

Proof. It is enough to observe that $h^{-1}(B) = f^{-1}(g^{-1}(B))$, for any $B \subseteq U$. $\square$

Corollary 1.12 (Compositions with a continuous maps). *Let $(S, \mathcal{S})$ be a measurable space, T be a topological space and $\mathcal{T}$ the Borel σ-algebra on T. Let $g : T \to \mathbb{R}$ be a continuous function. Then the map $g \circ f : S \to \mathbb{R}$ is measurable for each measurable function $f : S \to T$.*

Definition 1.13 (Generation by several functions)**.** Let[8] $(f_\gamma)_{\gamma\in\Gamma}$ be a family of maps from a set S into a measurable space $(T,\mathcal{T})$. The **σ-algebra generated by** $(f_\gamma)_{\gamma\in\Gamma}$, denoted by $\sigma\Big((f_\gamma)_{\gamma\in\Gamma}\Big)$, is the intersection of all σ-algebras on S which make each f_γ, $\gamma\in\Gamma$, measurable.

[8] The letter Γ will typically be used to denote an abstract index set - we only assume that it is nonempty, but make no other assumptions about its cardinality.

Problem 1.6. In the setting of Definition 1.13, show that

$$\sigma\Big((f_\gamma)_{\gamma\in\Gamma}\Big)=\sigma\Big(\bigcup_{\gamma\in\Gamma}f_\gamma^{-1}(\mathcal{T})\Big),$$

where $f_\gamma^{-1}(\mathcal{T})=\{f_\gamma^{-1}(B)\,:\,B\in\mathcal{T}\}$.

Products of measurable spaces

Definition 1.14 (Products, choice functions)**.** Let $(S_\gamma)_{\gamma\in\Gamma}$ be a family of sets, parametrized by some (possibly uncountable) index set Γ. The **product**[9] $\prod_{\gamma\in\Gamma}S_\gamma$ is the set of all functions $\boldsymbol{s}:\Gamma\to\cup_\gamma S_\gamma$ (called **choice functions**[10]) with the property that $\boldsymbol{s}(\gamma)\in S_\gamma$.

[9] When Γ is finite, each function $\boldsymbol{s}:\Gamma\to\cup_{\gamma\in\Gamma}S_\gamma$ can be identified with an ordered "tuple" $(\boldsymbol{s}(\gamma_1),\dots,\boldsymbol{s}(\gamma_n))$, where n is the cardinality (number of elements) of Γ, and $\gamma_1,\dots,\gamma_n$ is some ordering of its elements. With this identification, it is clear that our definition of a product coincides with the well-known definition in the finite case.

[10] The celebrated *Axiom of Choice* in set theory postulates that no matter what the family $(S_\gamma)_{\gamma\in\Gamma}$ is, there exists at least one choice function. In other words, axiom of choice simply asserts that products of sets are non-empty.

Definition 1.15 (Natural projections)**.** For $\gamma_0\in\Gamma$, the function $\pi_{\gamma_0}:\prod_{\gamma\in\Gamma}S_\gamma\to S_{\gamma_0}$ defined by

$$\pi_{\gamma_0}(\boldsymbol{s})=\boldsymbol{s}(\gamma_0),\text{ for }\boldsymbol{s}\in\prod_{\gamma\in\Gamma}S_\gamma,$$

is called the **(natural) projection onto the coordinate** γ_0.

Definition 1.16 (Products of measurable spaces)**.** Let $\{(S_\gamma,\mathcal{S}_\gamma)\}_{\gamma\in\Gamma}$ be a family of measurable spaces. The **product** $\otimes_{\gamma\in\Gamma}(S_\gamma,\mathcal{S}_\gamma)$ is a measurable space $(\prod_{\gamma\in\Gamma}S_\gamma,\otimes_{\gamma\in\Gamma}\mathcal{S}_\gamma)$, where $\otimes_{\gamma\in\Gamma}\mathcal{S}_\gamma$ is the smallest σ-algebra that makes all natural projections $(\pi_\gamma)_{\gamma\in\Gamma}$ measurable.

Example 1.17. When Γ is finite, the above definition can be made more intuitive. Suppose, just for simplicity, that $\Gamma=\{1,2\}$, so that $(S_1,\mathcal{S}_1)\otimes(S_2,\mathcal{S}_2)$ is a measurable space of the form $(S_1\times S_2,\mathcal{S}_1\otimes\mathcal{S}_2)$, where $\mathcal{S}_1\otimes\mathcal{S}_2$ is the smallest σ-algebra on $S_1\times S_2$ which makes both π_1 and π_2 measurable. The pull-backs under π_1 of sets in $\mathcal{S}_1$ are given by

$$\pi^{-1}(B_1)=\{(x,y)\in S_1\times S_2\,:\,x\in B_1\}=B_1\times S_2,\text{ for }B_1\in\mathcal{S}_1.$$

Similarly

$$\pi^{-1}(B_2)=S_1\times B_2,\text{ for }B_2\in\mathcal{S}_2.$$

Therefore, by Problem 1.6,

$$\mathcal{S}_1\otimes\mathcal{S}_2=\sigma\Big(\{B_1\times S_2,S_1\times B_2\,:\,B_1\in\mathcal{S}_1,B_2\in\mathcal{S}_2\}\Big).$$

Equivalently (why?)

$$\mathcal{S}_1 \otimes \mathcal{S}_2 = \sigma\Big(\{B_1 \times B_2 \,:\, B_1 \in \mathcal{S}_1, B_2 \in \mathcal{S}_2\}\Big).$$

In a completely analogous fashion, we can show that, for finitely many measurable spaces $(S_1, \mathcal{S}_1), \dots, (S_n, \mathcal{S}_n)$, we have

$$\bigotimes_{i=1}^{n} \mathcal{S}_i = \sigma\Big(\{B_1 \times B_2 \times \cdots \times B_n \,:\, B_1 \in \mathcal{S}_1, B_2 \in \mathcal{S}_2, \dots, B_n \in \mathcal{S}_n\}\Big)$$

The same goes for countable products[11].

[11] Uncountable products, however, behave very differently.

Problem 1.7. We know that the Borel σ-algebra (based on the usual Euclidean topology) can be constructed on each $\mathbb{R}^n$. A σ-algebra on $\mathbb{R}^n$ (for $n > 1$), can also be constructed as a product σ-algebra $\otimes_{i=1}^{n} \mathcal{B}(\mathbb{R})$. A third possibility is to consider the mixed case where $1 < m < n$ is picked and the σ-algebra $\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^{n-m})$ is constructed on $\mathbb{R}^n$ (which is now interpreted as a product of $\mathbb{R}^m$ and $\mathbb{R}^{n-m}$). Show that we get the same σ-algebra in all three cases.

Problem 1.8. Let $(P, \mathcal{P})$, $\{(S_\gamma, \mathcal{S}_\gamma)\}_{\gamma\in\Gamma}$ be measurable spaces and set $S = \prod_{\gamma\in\Gamma} S_\gamma$, $\mathcal{S} = \otimes_{\gamma\in\Gamma}\mathcal{S}_\gamma$. Prove that a map $f : P \to S$ is $(\mathcal{P}, \mathcal{S})$-measurable if and only if the composition $\pi_\gamma \circ f : P \to S_\gamma$ is $(\mathcal{P}, \mathcal{S}_\gamma)$ measurable for each $\gamma \in \Gamma$.

Note: Loosely speaking, this result states that a "vector"-valued mapping is measurable if and only if all of its components are measurable.

Definition 1.18 (Cylinder sets)**.** Let $\{(S_\gamma, \mathcal{S}_\gamma)\}_{\gamma\in\Gamma}$ be a family of measurable spaces, and let $(\prod_{\gamma\in\Gamma} S_\gamma, \otimes_{\gamma\in\Gamma}\mathcal{S}_\gamma)$ be its product. A subset $C \subseteq \prod_{\gamma\in\Gamma} S_\gamma$ is called a **cylinder set** if there exist a finite subset $\{\gamma_1, \dots, \gamma_n\}$ of Γ, as well as a measurable set $B \in \mathcal{S}_{\gamma_1} \otimes \mathcal{S}_{\gamma_2} \otimes \cdots \otimes \mathcal{S}_{\gamma_n}$ such that

$$C = \{\boldsymbol{s} \in \prod_{\gamma\in\Gamma} S_\gamma \,:\, (\boldsymbol{s}(\gamma_1), \dots, \boldsymbol{s}(\gamma_n)) \in B\}.$$

A cylinder set for which the set B can be chosen of the form $B = B_1 \times \cdots \times B_n$, for some $B_1 \in \mathcal{S}_1, \dots, B_n \in \mathcal{S}_n$ is called a **product cylinder set**. In that case

$$C = \{\boldsymbol{s} \in \prod_{\gamma\in\Gamma} S_\gamma \,:\, (\boldsymbol{s}(\gamma_1) \in B_1, \boldsymbol{s}(\gamma_2) \in B_2, \dots, \boldsymbol{s}(\gamma_n) \in B_n\}.$$

Problem 1.9.

1. Show that the family of product cylinder sets generates the product σ-algebra.
2. Show that (not-necessarily-product) cylinders are measurable in the product σ-algebra.

3. Which of the 4 families of sets from Definition 1.2 does the collection of all product cylinders belong to in general? How about (not-necessarily-product) cylinders?

Example 1.19. The following example will play a major role in probability theory. Hence the name **coin-toss space**. Here $\Gamma = \mathbb{N}$ and for $i \in \mathbb{N}$, $(S_i, \mathcal{S}_i)$ is the discrete two-element space $S_i = \{-1, 1\}$, $\mathcal{S}_i = 2^{S_1}$. The product $\prod_{i\in\mathbb{N}} S_i = \{-1,1\}^{\mathbb{N}}$ can be identified with the set of all sequences $\boldsymbol{s} = (s_1, s_2, \dots)$, where $s_i \in \{-1,1\}$, $i \in \mathbb{N}$. For each cylinder set C, there exists (why?) $n \in \mathbb{N}$ and a subset B of $\{-1,1\}^n$ such that

$$C = \{\boldsymbol{s} = (s_1, \dots, s_n, s_{n+1}, \dots) \in \{-1,1\}^{\mathbb{N}} : (s_1, \dots, s_n) \in B\}.$$

The product cylinders are even simpler - they are always of the form $C = \{-1,1\}^{\mathbb{N}}$ or $C = C_{n_1,\dots,n_k;b_1,\dots,b_k}$, where

$$C_{n_1,\dots,n_k;b_1,\dots,b_k} = \Big\{\boldsymbol{s} = (s_1, s_2, \dots) \in \{-1,1\}^{\mathbb{N}} : s_{n_1} = b_1, \dots, s_{n_k} = b_k\Big\},$$

for some $k \in \mathbb{N}$, $1 \le n_1 < n_2 < \dots < n_k \in \mathbb{N}$ and $b_1, b_2, \dots, b_k \in \{-1,1\}$.

We know that the σ-algebra $\mathcal{S} = \otimes_{i\in\mathbb{N}} \mathcal{S}_i$ is generated by all projections $\pi_i : \{-1,1\}^{\mathbb{N}} \to \{-1,1\}$, $i \in \mathbb{N}$, where $\pi_i(\boldsymbol{s}) = s_i$. Equivalently, by Problem 1.9, $\mathcal{S}$ is generated by the collection of all cylinder sets.

Problem 1.10. One can obtain the product σ-algebra $\mathcal{S}$ on $\{-1,1\}^{\mathbb{N}}$ as the Borel σ-algebra corresponding to a particular topology which makes $\{-1,1\}^{\mathbb{N}}$ compact. Here is how. Start by defining a mapping $d : \{-1,1\}^{\mathbb{N}} \times \{-1,1\}^{\mathbb{N}} \to [0,\infty)$ by

$$d(\boldsymbol{s}^1, \boldsymbol{s}^2) = 2^{-i(\boldsymbol{s}^1,\boldsymbol{s}^2)}, \text{ where } i(\boldsymbol{s}^1, \boldsymbol{s}^2) = \inf\{i \in \mathbb{N} : s^1_i \ne s^2_i\}, \quad (1.1)$$

for $\boldsymbol{s}^j = (s^j_1, s^j_2, \dots)$, $j = 1, 2$.

1. Show that d is a metric on $\{-1,1\}^{\mathbb{N}}$.
2. Show that $\{-1,1\}^{\mathbb{N}}$ is compact under d.

Hint: Use the diagonal argument.

3. Show that each cylinder of $\{-1,1\}^{\mathbb{N}}$ is both open and closed under d.
4. Show that each open ball is a cylinder.
5. Show that $\{-1,1\}^{\mathbb{N}}$ is separable, i.e., it admits a countable dense subset.
6. Conclude that $\mathcal{S}$ coincides with the Borel σ-algebra on $\{-1,1\}^{\mathbb{N}}$ under the metric d.

Real-valued measurable functions

Let $\mathcal{L}^0(S, \mathcal{S}; \mathbb{R})$ (or, simply, $\mathcal{L}^0(S; \mathbb{R})$ or $\mathcal{L}^0(\mathbb{R})$ or $\mathcal{L}^0$ when the domain $(S, \mathcal{S})$ or the co-domain $\mathbb{R}$ are clear from the context) be the set of all $\mathcal{S}$-measurable functions $f : S \to \mathbb{R}$. The set of non-negative measurable functions is denoted by $\mathcal{L}^0_+$ or $\mathcal{L}^0([0, \infty))$.

Proposition 1.20 (Measurable functions form a vector space). *$\mathcal{L}^0$ is a vector space, i.e.*

$$\alpha f + \beta g \in \mathcal{L}^0, \text{ whenever } \alpha, \beta \in \mathbb{R},\ f, g \in \mathcal{L}^0.$$

Proof. Let us define a mapping $F : S \to \mathbb{R}^2$ by $F(x) = (f(x), g(x))$. By Problem 1.7, the Borel σ-algebra on $\mathbb{R}^2$ is the same as the product σ-algebra when we interpret $\mathbb{R}^2$ as a product of two copies of $\mathbb{R}$. Therefore, since its compositions with the coordinate projections are precisely the functions f and g, Problem 1.8 implies that F is $(\mathcal{S}, \mathcal{B}(\mathbb{R}^2))$-measurable.

Consider the function $\phi : \mathbb{R}^2 \to \mathbb{R}$ given by $\phi(x, y) = \alpha x + \beta y$. It is linear, and, therefore, continuous. By Corollary 1.12, the composition $\phi \circ F : \mathcal{S} \to \mathbb{R}$ is $(\mathcal{S}, \mathcal{B}(\mathbb{R}))$-measurable, and it only remains to note that

$$(\phi \circ F)(x) = \phi(F(x)) = \alpha f(x) + \beta g(x), \text{ i.e., } \phi \circ F = \alpha f + \beta g. \qquad \square$$

Proposition 1.21 (Products and maxima preserve measurability). *Let f, g be in $\mathcal{L}^0$. Then fg, $\max(f, g)$ and $\min(f, g)$ belong to $\mathcal{L}^0$.*

Proof. The functions $(x, y) \mapsto \max(x, y)$ and $(x, y) \to xy$ are continuous from $\mathbb{R}^2$ to $\mathbb{R}$. $\square$

Problem 1.11. Suppose that $f \in \mathcal{L}^0$ has the property that $f(x) \neq 0$ for all $x \in S$. Then the function $1/f$ is also in $\mathcal{L}^0$.

Hint: Find a measurable function g, defined on $\mathbb{R}$ such that $1/f(x) = g(f(x))$, for all $x \in S$.

Definition 1.22 (The indicator function of a set). For $A \subseteq S$, the **indicator function $\mathbf{1}_A$** is defined by

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

Despite their simplicity, indicators will be extremely useful throughout these notes.

Problem 1.12. Show that for $A \subset S$, we have $A \in \mathcal{S}$ if and only if $\mathbf{1}_A \in \mathcal{L}^0$.

Mathematical structures and measurability

A measurable structure on a set gives relates to various topological and algebraic structures encountered throughout mathematics.

Since it contains the function 1 and the products of pairs of its elements, the set $\mathcal{L}^0$ has the structure of an *unital algebra* (not to be confused with the algebra of sets defined above). It is true, however, that any algebra $\mathcal{A}$ of subsets of a non-empty set S, together with the operations of union, intersection and complement forms a *Boolean algebra*. Alternatively, it can be given the (algebraic) structure of a *commutative ring with a unit*. Indeed, under the operation $\triangle$ of symmetric difference, $\mathcal{A}$ is an Abelian group (prove that!). If, in addition, the operation of intersection is introduced in lieu of multiplication, the resulting structure is, indeed, the one of a commutative ring.

Additionally, a natural partial order given by $f \preceq g$ if $f(x) \leq g(x)$, for all $x \in S$, can be introduced on $\mathcal{L}^0$. This order is compatible with the operations of addition and multiplication and has the additional property that each pair $\{f, g\} \subseteq \mathcal{L}^0$ admits a *least upper bound*, i.e., the element $h \in \mathcal{L}^0$ such that $f \preceq h$, $g \preceq h$ and $h \preceq k$, for any other k with the property that $f, g \preceq k$. Indeed, we simply take $h(x) = \max(f(x), g(x))$. A similar statement can be made for a *greatest lower bound*. A vector space with a partial order which satisfies the above properties is called a *vector lattice*.

As we move through the theory, we will see how many other mathematical structures appear naturally in measure theory.

Extended real-valued measurable functions

Since a limit of a sequence of real numbers does not necessarily belong to $\mathbb{R}$, it is often necessary to consider functions which are allowed to take the values ∞ and $-\infty$. The set $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ is called the **extended** set of real numbers. Most (but not all) of the algebraic and topological structure from $\mathbb{R}$ can be lifted to $\bar{\mathbb{R}}$. In some cases there is no unique way to do that, so we choose one of them as a matter of convention.

1. **Arithmetic operations.** For $x, y \in \bar{\mathbb{R}}$, all the arithmetic operations are defined in the usual way when $x, y \in \mathbb{R}$. When one or both are in $\{\infty, -\infty\}$, we use the following convention, where $\oplus \in \{+, -, *, /\}$:

   We define $x \oplus y = z$ if all pairs of sequences $\{x_n\}_{n\in\mathbb{N}}$, $\{y_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}$ such that $x = \lim_n x_n$, $y = \lim_n y_n$ and $x_n \oplus y_n$ is well-defined for all $n \in \mathbb{N}$, we have

$$z = \lim_n (x_n \oplus y_n).$$

Otherwise, $x \oplus y$ is not defined. This basically means that all intuitively obvious conventions (such as $\infty + \infty = \infty$ and $\frac{a}{\infty} = 0$ for $a \in \mathbb{R}$ hold). In measure theory, however, we do make one important *exception* to the above rule. We set

$$0 \times \infty = \infty \times 0 = 0 \times (-\infty) = (-\infty) \times 0 = 0.$$

2. **Order.** $-\infty < x < \infty$, for all $x \in \mathbb{R}$. Also, *each* non-empty subset of $\bar{\mathbb{R}}$ admits a supremum and an infimum in $\bar{\mathbb{R}}$ in an obvious way.

3. **Convergence.** It is impossible to extend the usual (Euclidean) metric from $\mathbb{R}$ to $\bar{\mathbb{R}}$, but a metric $d' : \mathbb{R} \times \mathbb{R} \to [0, \pi)$ given by

$$d'(x, y) = |\arctan(y) - \arctan(x)|,$$

extends readily to a metric on $\bar{\mathbb{R}}$ if we set $\arctan(\infty) = \pi/2$ and $\arctan(-\infty) = -\pi/2$. We define convergence (and topology) on $\mathbb{R}$ using d'. For example, a sequence $\{x_n\}_{n\in\mathbb{N}}$ in $\bar{\mathbb{R}}$ converges to $+\infty$ if

(a) It contains only a finite number of terms equal to $-\infty$,

(b) Every subsequence of $\{x_n\}_{n\in\mathbb{N}}$ whose elements are in $\mathbb{R}$ converges to $+\infty$ (in the usual sense).

We define the notions of **limit superior** and **limit inferior** on $\bar{\mathbb{R}}$ for a sequence $\{x_n\}_{n\in\mathbb{N}}$ in the following manner:

$$\limsup_n x_n = \inf_n S_n, \text{ where } S_n = \sup_{k\geq n} x_k,$$

and

$$\liminf_n x_n = \sup_n I_n, \text{ where } I_n = \inf_{k\geq n} x_k.$$

If you have forgotten how to manipulate limits inferior and superior, here is an exercise to remind you:

Problem 1.13. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in $\bar{\mathbb{R}}$. Prove the following statements:

1. $a \in \bar{\mathbb{R}}$ satisfies $a \geq \limsup_n x_n$ if and only if for any $\varepsilon \in (0, \infty)$ there exists $n_\varepsilon \in \mathbb{N}$ such that $x_n \leq a + \varepsilon$ for $n \geq n_\varepsilon$.

2. $\liminf_n x_n \leq \limsup_n x_n$.

3. Define

$$A = \{\lim_k x_{n_k} : x_{n_k} \text{ is a convergent (in } \bar{\mathbb{R}}) \text{ subsequence of } \{x_n\}_{n\in\mathbb{N}}\}.$$

Show that

$$\{\liminf_n x_n, \limsup_n x_n\} \subseteq A \subseteq [\liminf_n x_n, \limsup_n x_n].$$

Give an example in which both inclusion above are strict.

Having introduced a topology on $\bar{\mathbb{R}}$ we immediately have the σ-algebra $\mathcal{B}(\bar{\mathbb{R}})$ of Borel sets there and the notion of measurability for functions mapping a measurable space $(S, \mathcal{S})$ into $\bar{\mathbb{R}}$.

Problem 1.14. Show that a subset $A \subseteq \bar{\mathbb{R}}$ is in $\mathcal{B}(\bar{\mathbb{R}})$ if and only if $A \setminus \{\infty, -\infty\}$ is Borel in $\mathbb{R}$. Show that a function $f : S \to \bar{\mathbb{R}}$ is measurable in the pair $(\mathcal{S}, \mathcal{B}(\bar{\mathbb{R}}))$ if and only if the sets $f^{-1}(\{\infty\})$, $f^{-1}(\{-\infty\})$ and $f^{-1}(A)$ are in $\mathcal{S}$ for all $A \in \mathcal{B}(\mathbb{R})$ (equivalently and more succinctly, $f \in \mathcal{L}^0(\bar{\mathbb{R}})$ iff $\{f = \infty\}, \{f = -\infty\} \in \mathcal{S}$ and $f\mathbf{1}_{\{f \in \mathbb{R}\}} \in \mathcal{L}^0$).

The set of all measurable functions $f : S \to \bar{\mathbb{R}}$ is denoted by $\mathcal{L}^0(S, \mathcal{S}; \bar{\mathbb{R}})$, and, as always we leave out S and $\mathcal{S}$ when no confusion can arise. The set of extended non-negative measurable functions often plays a role, so we denote it by $\mathcal{L}^0([0,\infty])$ or $\mathcal{L}^0_+(\bar{\mathbb{R}})$. Unlike $\mathcal{L}^0(\mathbb{R})$, $\mathcal{L}^0(\bar{\mathbb{R}})$ is not a vector space, but it retains all the order structure. Moreover, it is particularly useful because, unlike $\mathcal{L}^0(\mathbb{R})$, it is closed with respect to the limiting operations. More precisely, for a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^0(\bar{\mathbb{R}})$, we define the functions $\limsup_n f_n : S \to [-\infty, \infty]$ and $\liminf_n f_n : S \to [-\infty, \infty]$ by

$$(\limsup_n f_n)(x) = \limsup_n f_n(x) = \inf_n \left(\sup_{k \geq n} f_k(x) \right),$$

and

$$(\liminf_n f_n)(x) = \liminf_n f_n(x) = \sup_n \left(\inf_{k \geq n} f_k(x) \right).$$

Then, we have the following result, where the supremum and infimum of a sequence of functions are defined pointwise (just like the limits superior and inferior).

Proposition 1.23 (Limiting operations preserve measurability)**.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0(\bar{\mathbb{R}})$. Then*

1. $\sup_n f_n, \inf_n f_n \in \mathcal{L}^0(\bar{\mathbb{R}})$,
2. $\limsup_n f_n, \liminf_n f_n \in \mathcal{L}^0(\bar{\mathbb{R}})$,
3. *if $f(x) = \lim_n f_n(x)$ exists in $\bar{\mathbb{R}}$ for each $x \in S$, then $f \in \mathcal{L}^0(\bar{\mathbb{R}})$, and*
4. *the set $A = \{\lim_n f_n$ exists in $\bar{\mathbb{R}}\ \}$ is in $\mathcal{S}$.*

Proof.

1. We show only the statement for the supremum. It is clear that it is enough to show that the set $\{\sup_n f_n \leq a\}$ is in $\mathcal{S}$ for all $a \in (-\infty, \infty]$ (why?). This follows, however, directly from the simple identity

$$\{\sup_n f_n \leq a\} = \cap_n \{f_n \leq a\},$$

and the fact that σ-algebras are closed with respect to countable intersections.

2. Define $g_n = \sup_{k \geq n} f_k$ and use part 1. above to conclude that $g_n \in \mathcal{L}^0(\bar{\mathbb{R}})$ for each $n \in \mathbb{N}$. Another appeal to part 1. yields that $\limsup_n f_n = \inf_n g_n$ is in $\mathcal{L}^0(\bar{\mathbb{R}})$. The statement about the limit inferior follows in the same manner.

3. If the limit $f(x) = \lim_n f_n(x)$ exists for all $x \in S$, then $f = \liminf_n f_n$ which is measurable by part 2. above.

4. The statement follows from the fact that $A = f^{-1}(\{0\})$, where

$$f(x) = \arctan\left(\limsup_n f_n(s)\right) - \arctan\left(\liminf_n f_n(x)\right).$$

$\square$

Note: The unexpected use of the function arctan is really noting to be puzzled by. The only property needed is its measurability (it is continuous) and monotonicity+bijectivity from $[-\infty, \infty]$ to $[-\pi/2, \pi/2]$. We compose the limits superior and inferior with it so that we don't run into problems while trying to subtract $+\infty$ from itself.

Additional Problems

Problem 1.15. Which of the following are σ-algebras on $\mathbb{R}$?

1. $\mathcal{S} = \{A \subseteq \mathbb{R} : 0 \in A\}$.
2. $\mathcal{S} = \{A \subseteq \mathbb{R} : A \text{ is finite}\}$.
3. $\mathcal{S} = \{A \subseteq \mathbb{R} : A \text{ is finite, or } A^c \text{ is finite}\}$.
4. $\mathcal{S} = \{A \subseteq \mathbb{R} : A \text{ is countable or } A^c \text{ is countable}\}$.
5. $\mathcal{S} = \{A \subseteq \mathbb{R} : A \text{ is open}\}$.
6. $\mathcal{S} = \{A \subset \mathbb{R} : A \text{ is open or } A \text{ is closed}\}$.

Problem 1.16. A **partition** of a set S is a family $\mathcal{P}$ of non-empty subsets of S with the property that each $\omega \in S$ belongs to exactly one $A \in \mathcal{P}$.

1. Show that the number of different algebras on a finite set S is equal to the number of different partitions of S.
2. How many algebras are there on the set $S = \{1, 2, 3\}$?
3. Does there exist an algebra with 754 elements?
4. For $N \in \mathbb{N}$, let a_n be the number of different algebras on the set $\{1, 2, \ldots, n\}$. Show that $a_1 = 1$, $a_2 = 2$, $a_3 = 5$, and that the following recursion holds (where $a_0 = 1$ by definition),

$$a_{n+1} = \sum_{k=0}^{n} \binom{n}{k} a_k.$$

Note: This number for $S_n = \{1, 2, \ldots, n\}$ is called the n^{th} **Bell number** B_n, and no nice closed-form expression for it is known. See below, though.

5. Show that the exponential generating function for the sequence $\{a_n\}_{n\in\mathbb{N}}$ is $f(x) = e^{e^x-1}$, i.e., that

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!} = e^{e^x-1} \text{ or, equivalently, } a_n = \left(\tfrac{d^n}{dx^n} e^{e^x-1}\right)\Big|_{x=0}.$$

Problem 1.17. Let $(S, \mathcal{S})$ be a measurable space. For $f, g \in \mathcal{L}^0$ show that the sets $\{f = g\} = \{x \in S : f(x) = g(x)\}$, $\{f < g\} = \{x \in S : f(x) < g(x)\}$ are in $\mathcal{S}$.

Problem 1.18. Show that all

1. monotone,
2. convex

functions $f : \mathbb{R} \to \mathbb{R}$ are measurable.

Problem 1.19. Let $(S, \mathcal{S})$ be a measurable space and let $f : S \to \mathbb{R}$ be a Borel-measurable function. Show that the graph

$$G_f = \{(x, y) \in S \times \mathbb{R} : f(x) = y\},$$

of f is a measurable subset in the product space $(S \times \mathbb{R}, \mathcal{S} \otimes \mathcal{B}(\mathbb{R}))$.

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 2
MEASURES

Measure spaces

Definition 2.1 (Measure)**.** Let $(S, \mathcal{S})$ be a measurable space. A mapping $\mu : \mathcal{S} \to [0, \infty]$ is called a **(positive) measure** if

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

2. $\mu(\cup_n A_n) = \sum_{n\in\mathbb{N}} \mu(A_n)$, for all *pairwise disjoint* $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$.

A triple $(S, \mathcal{S}, \mu)$ consisting of a non-empty set, a σ-algebra $\mathcal{S}$ on it and a measure μ on $\mathcal{S}$ is called a **measure space**.

Remark 2.2.

1. A mapping whose domain is some nonempty set $\mathcal{A}$ of subsets of some set S is sometimes called a **set function**.

2. If the requirement 2. in the definition of the measure is weakened so that it is only required that $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$, for $n \in \mathbb{N}$, and pairwise disjoint $A_1, \dots, A_n$, we say that the mapping μ is a **finitely-additive measure**. If we want to stress that a mapping μ satisfies the original requirement 2. for *sequences* of sets, we say that μ is **σ-additive** (**countably additive**).

Definition 2.3 (Terminology)**.** A measure μ on the measurable space $(S, \mathcal{S})$ is called

1. a **probability measure**, if $\mu(S) = 1$,

2. a **finite measure**, if $\mu(S) < \infty$,

3. a **σ-finite measure**, if there exists a sequence $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$ such that $\cup_n A_n = S$ and $\mu(A_n) < \infty$,

4. **diffuse** or **atom-free**, if $\mu(\{x\}) = 0$, whenever $x \in S$ and $\{x\} \in \mathcal{S}$.

A set $N \in \mathcal{S}$ is said to be **null** if $\mu(N) = 0$.

Example 2.4 (Examples of measures)**.** Let S be a non-empty set, and let $\mathcal{S}$ be a σ-algebra on S.

1. **Measures on countable sets.** Suppose that S is a finite or countable set. Then each measure μ on $\mathcal{S} = 2^S$ is of the form

$$\mu(A) = \sum_{x \in A} p(x),$$

for some function $p : S \to [0, \infty]$ (why?). In particular, for a finite set S with N elements, if $p(x) = 1/N$ then μ is a probability measure called the **uniform measure**[1] on S.

2. **Dirac measure.** For $x \in S$, we define the set function δ_x on $\mathcal{S}$ by

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

It is easy to check that δ_x is indeed a measure on $\mathcal{S}$. Alternatively, δ_x is called the **point mass at** x (or an **atom on** x, or the **Dirac function**, even though it is not really a function). Moreover, δ_x is a probability measure and, therefore, a finite and a σ-finite measure. It is atom free only if $\{x\} \notin \mathcal{S}$.

3. **Counting Measure.** Define a set function $\mu : \mathcal{S} \to [0, \infty]$ by

$$\mu(A) = \begin{cases} \#A, & A \text{ is finite}, \\ \infty, & A \text{ is infinite}, \end{cases}$$

where, as above, $\#A$ denotes the number of elements in the set A. Again, it is not hard to check that μ is a measure - it is called the **counting measure**. Clearly, μ is a finite measure if and only is S is a finite set. μ could be σ-finite, though, even without S being finite. Simply take $S = \mathbb{N}$, $\mathcal{S} = 2^{\mathbb{N}}$. In that case $\mu(S) = \infty$, but for $A_n = \{n\}$, $n \in \mathbb{N}$, we have $\mu(A_n) = 1$, and $S = \cup_n A_n$. Finally, μ is never atom-free and it is a probability measure only if $\#S = 1$.

[1] In the finite case, it has the well-known property that $\mu(A) = \frac{\#A}{\#S}$, where # denotes the cardinality (number of elements).

Example 2.5 (A finitely-additive set function which is not a measure)**.** Let $S = \mathbb{N}$, and $\mathcal{S} = 2^S$. For $A \in S$ define $\mu(A) = 0$ if A is finite and $\mu(A) = \infty$, otherwise. For $A_1, \dots, A_n \subseteq S$, either

1. all A_i is finite, for $i = 1, \dots, n$. Then $\cup_{i=1}^n A_i$ is also finite and so

$$0 = \mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i), \text{ or}$$

2. at least one A_i is infinite. Then $\cup_{i=1}^n A_i$ is also infinite and so

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

On the other hand, take $A_i = \{i\}$, for $i \in \mathbb{N}$. Then $\mu(A_i) = 0$, for each $i \in \mathbb{N}$, and, so, $\sum_{i\in\mathbb{N}} \mu(A_i) = 0$, but $\mu(\cup_i A_i) = \mu(\mathbb{N}) = \infty$.

Note: It is possible to construct very simple-looking finite-additive measures which are not σ-additive. For example, there exist $\{0,1\}$-valued finitely-additive measures on all subsets of $\mathbb{N}$, which are not σ-additive. Such objects are called **ultrafilters** and their existence is equivalent to a certain version of the Axiom of Choice.

Proposition 2.6 (First properties of measures). *Let $(S, \mathcal{S}, \mu)$ be a measure space.*

1. *For $A_1, \dots, A_n \in \mathcal{S}$ with $A_i \cap A_j = \emptyset$, for $i \neq j$, we have*
$$\sum_{i=1}^{n} \mu(A_i) = \mu(\cup_{i=1}^{n} A_i) \qquad \textit{(Finite additivity)}$$
2. *If $A, B \in \mathcal{S}$, $A \subseteq B$, then*
$$\mu(A) \leq \mu(B) \qquad \textit{(Monotonicity of measures)}$$
3. *If $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$ is increasing, then*
$$\mu(\cup_n A_n) = \lim_n \mu(A_n) = \sup_n \mu(A_n).$$
(Continuity with respect to increasing sequences)
4. *If $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$ is decreasing and $\mu(A_1) < \infty$, then*
$$\mu(\cap_n A_n) = \lim_n \mu(A_n) = \inf_n \mu(A_n).$$
(Continuity with respect to decreasing sequences)
5. *For a sequence $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$, we have*
$$\mu(\cup_n A_n) \leq \sum_{n\in\mathbb{N}} \mu(A_n). \qquad \textit{(Subadditivity)}$$

Proof.

1. Note that the sequence $A_1, A_2, \dots, A_n, \emptyset, \emptyset, \dots$ is pairwise disjoint, and so, by σ-additivity,
$$\begin{aligned}\mu(\cup_{i=1}^{n} A_i) &= \mu(\cup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i) \\ &= \sum_{i=1}^{n} \mu(A_i) + \sum_{i=n+1}^{\infty} \mu(\emptyset) = \sum_{i=1}^{n} \mu(A_i).\end{aligned}$$
2. Write B as a disjoint union $A \cup (B \setminus A)$ of elements of $\mathcal{S}$. By (1) above, $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$.
3. Define $B_1 = A_1$, $B_n = A_n \setminus A_{n-1}$ for $n > 1$. Then $\{B_n\}_{n\in\mathbb{N}}$ is a pairwise disjoint sequence in $\mathcal{S}$ with $\cup_{k=1}^{n} B_k = A_n$ for each $n \in \mathbb{N}$ (why?). By σ-additivity we have
$$\begin{aligned}\mu(\cup_n A_n) &= \mu(\cup_n B_n) = \sum_{n\in\mathbb{N}} \mu(B_n) = \lim_n \sum_{k=1}^{n} \mu(B_k) \\ &= \lim_n \mu(\cup_{k=1}^{n} B_k) = \lim_n \mu(A_n).\end{aligned}$$

4. Consider the increasing sequence $\{B_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$ given by $B_n = A_1 \setminus A_n$. By De Morgan laws, finiteness of $\mu(A_1)$ and (3) above, we have

$$\begin{aligned}\mu(A_1) - \mu(\cap_n A_n) &= \mu(A_1 \setminus (\cap_n A_n)) = \mu(\cup_n B_n) = \lim_n \mu(B_n) \\ &= \lim_n \mu(A_1 \setminus A_n) = \mu(A_1) - \lim_n \mu(A_n).\end{aligned}$$

Subtracting both sides from $\mu(A_1) < \infty$ produces the statement.

5. We start from the observation that for $A_1, A_1 \in \mathcal{S}$ the set $A_1 \cup A_2$ can be written as a disjoint union

$$A_1 \cup A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1) \cup (A_1 \cap A_2),$$

so that

$$\mu(A_1 \cup A_2) = \mu(A_1 \setminus A_2) + \mu(A_2 \setminus A_1) + \mu(A_1 \cap A_2).$$

On the other hand,

$$\begin{aligned}\mu(A_1) + \mu(A_2) &= (\mu(A_1 \setminus A_2) + \mu(A_1 \cap A_2)) \\ &\quad + \Big(\mu(A_2 \setminus A_1) + \mu(A_1 \cap A_2)\Big) \\ &= \mu(A_1 \setminus A_2) + \mu(A_2 \setminus A_1) + 2\mu(A_1 \cap A_2),\end{aligned}$$

and so

$$\mu(A_1) + \mu(A_2) - \mu(A_1 \cup A_2) = \mu(A_1 \cap A_2) \geq 0.$$

Induction can be used to show that

$$\mu(A_1 \cup \dots \cup A_n) \leq \sum_{k=1}^{n} \mu(A_k).$$

Since all $\mu(A_n)$ are nonnegative, we now have

$$\mu(A_1 \cup \dots \cup A_n) \leq \alpha, \text{ for each } n \in \mathbb{N}, \text{ where } \alpha = \sum_{n\in\mathbb{N}} \mu(A_n).$$

The sequence $\{B_n\}_{n\in\mathbb{N}}$ given by $B_n = \cup_{k=1}^n A_k$ is increasing, so the continuity of measure with respect to increasing sequences implies that

$$\mu(\cup_n A_n) = \mu(\cup_n B_n) = \lim_n \mu(B_n) = \lim_n \mu(A_1 \cup \dots \cup A_n) \leq \alpha. \quad \square$$

Remark 2.7. The condition $\mu(A_1) < \infty$ in the part (4) of Proposition 2.6 cannot be significantly relaxed. Indeed, let μ be the counting measure on $\mathbb{N}$, and let $A_n = \{n, n+1, \dots\}$. Then $\mu(A_n) = \infty$ and, so $\lim_n \mu(A_n) = \infty$. On the other hand, $\cap A_n = \emptyset$, so $\mu(\cap_n A_n) = 0$.

In addition to unions and intersections, one can produce other important new sets from sequences of old ones. More specifically, let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of subsets of S. The subset $\liminf_n A_n$ of S, defined by

$$\liminf_n A_n = \cup_n B_n, \text{ where } B_n = \cap_{k\geq n} A_k,$$

is called the **limit inferior** of the sequence A_n. It is also denoted by $\underline{\lim}_n A_n$ or $\{A_n, \text{ ev.}\}$ (*ev.* stands for *eventually*[2]).

Similarly, the subset $\limsup_n A_n$ of S, defined by

$$\limsup_n A_n = \cap_n B_n, \text{ where } B_n = \cup_{k\geq n} A_k,$$

is called the **limit superior** of the sequence A_n. It is also denoted by $\overline{\lim}_n A_n$ or $\{A_n, \text{ i.o.}\}$ (*i.o.* stands for *infinitely often*[3]). Clearly, we have

$$\liminf_n A_n \subseteq \limsup_n A_n.$$

[2] the reason for the use of the word *eventually* is the following: $\liminf_n A_n$ is the set of all $x \in S$ which belong to A_n for *all but finitely many values* of the index n, i.e., from some value of the index n onwards.

[3] in words, $\limsup_n A_n$ is the set of all $x \in S$ which belong A_n for *infinitely many values* of n.

Problem 2.1. Let $(S, \mathcal{S}, \mu)$ be a *finite* measure space. Show that

$$\mu(\liminf A_n) \leq \liminf \mu(A_n) \leq \limsup \mu(A_n) \leq \mu(\limsup A_n),$$

for any sequence $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$. Give an example of a (single) sequence $\{A_n\}_{n\in\mathbb{N}}$ for which all inequalities above are strict.

Hint: For the second part, a measure space with finite (and small) S will do.

Proposition 2.8 (Borel-Cantelli Lemma I)**.** *Let $(S, \mathcal{S}, \mu)$ be a measure space, and let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of sets in $\mathcal{S}$ with the property that $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$. Then*

$$\mu(\limsup_n A_n) = 0.$$

Proof. Set $B_n = \cup_{k\geq n} A_k$, so that $\{B_n\}_{n\in\mathbb{N}}$ is a decreasing sequence of sets in $\mathcal{S}$ with $\limsup_n A_n = \cap_n B_n$, and so

$$\mu(\limsup A_n) \leq \mu(B_n), \text{ for each } n \in \mathbb{N}.$$

Using the subadditivity of measures of Proposition 2.6, part 5., we get

$$\mu(B_n) \leq \sum_{k=n}^{\infty} \mu(A_n). \tag{2.1}$$

Since $\sum_{n\in\mathbb{N}} \mu(A_n)$ converges, the right-hand side of (2.1) can be made arbitrarily small by choosing large enough $n \in \mathbb{N}$. $\square$

Extensions of measures and the coin-toss space

Example 1.19 has introduced the measurable space $(\{-1,1\}^{\mathbb{N}}, \mathcal{S})$, with $\mathcal{S} = \otimes_n 2^{\{-1,1\}}$ being the product σ-algebra on $\{-1,1\}^{\mathbb{N}}$. The purpose of the present section is to turn $(\{-1,1\}^{\mathbb{N}}, \mathcal{S})$ into a measure space, i.e., to define a suitable measure on it. It is easy to construct just any measure on $\{-1,1\}^{\mathbb{N}}$, but the one we are after is the one which will justify the name *coin-toss space*.

The intuition we have about tossing a fair coin infinitely many times should help us start with the definition of the coin-toss measure - denoted by μ_C - on cylinders. Since the coordinate spaces $\{-1,1\}$ are particularly simple, each product cylinder is of the form $C = \{-1,1\}^{\mathbb{N}}$ or $C = C_{n_1,\dots,n_k;b_1,\dots,b_k}$, where

$$C_{n_1,\dots,n_k;b_1,\dots,b_k} = \Big\{ \boldsymbol{s} = (s_1, s_2, \dots) \in \{-1,1\}^{\mathbb{N}} \,:\, s_{n_1} = b_1, \dots, s_{n_k} = b_k \Big\},$$

for some $k \in \mathbb{N}$, and a choice of $1 \leq n_1 < n_2 < \dots < n_k \in \mathbb{N}$ of coordinates and the corresponding values $b_1, b_2, \dots, b_k \in \{-1,1\}$.

In the language of elementary probability, each cylinder corresponds to the event when the outcome of the n_i-th coin is $b_i \in \{-1,1\}$, for $i = 1, \dots, n$. The measure (probability) of this event can only be given by

$$\mu_C(C_{n_1,\dots,n_k;b_1,\dots,b_k}) = \underbrace{\tfrac{1}{2} \times \tfrac{1}{2} \times \dots \times \tfrac{1}{2}}_{k \text{ times}} = 2^{-k}. \tag{2.2}$$

The hard part is to extend this definition to *all* elements of $\mathcal{S}$, and not only cylinders. For example, in order to state the law of large numbers later on, we will need to be able to compute the measure of the set

$$\Big\{ \boldsymbol{s} \in \{-1,1\}^{\mathbb{N}} \,:\, \lim_n \tfrac{1}{n} \sum_{k=1}^{n} s_k = 0 \Big\},$$

which is clearly not a cylinder.

Problem 1.9 states, however, that cylinders form an algebra and generate the σ-algebra $\mathcal{S}$. Luckily, this puts us close to the conditions of the following important theorem of Caratheodory.

Theorem 2.9 (Caratheodory's Extension Theorem)**.** *Let S be a non-empty set, let $\mathcal{A}$ be an algebra of its subsets and let $\mu : \mathcal{A} \to [0,\infty]$ be a set-function with the following properties:*

1. $\mu(\emptyset) = 0$, *and*
2. $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$, *if* $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ *is a partition of* A.

Then, there exists a measure $\tilde{\mu}$ on $\sigma(\mathcal{A})$ with the property that $\mu(A) = \tilde{\mu}(A)$ for $A \in \mathcal{A}$.

Note: In words, a σ-additive measure on an algebra $\mathcal{A}$ can be extended to a σ-additive measure on the σ-algebra generated by $\mathcal{A}$. It is clear that the σ-additivity requirement of Theorem 2.9 is necessary, but it is quite surprising that it is actually sufficient.

Of Theorem 2.9. **PART I.** We start by defining a "measure-like object", called an **outer measure**, $\mu^* : 2^S \to [0,\infty]$ in the following way:

$$\mu^*(B) = \inf \Big\{ \sum_{n=1}^{\infty} \mu(A_n) \,:\, B \subseteq \bigcup_{n=1}^{\infty} A_n,\ A_n \in \mathcal{A} \text{ for all } n \in \mathbb{N} \Big\}.$$

Note: Intuitively, we try all different countable covers of B with elements of $\mathcal{A}$ and minimize the total μ.

Even though we don't expect the infimum in the definition of μ^* to be attained, μ^* has the following properties:

1. $\mu^*(\emptyset) = 0$ *(nontriviality)*,

2. for $B \subseteq C$, $\mu^*(B) \leq \mu^*(C)$ *(monotonicity)*, and

3. $\mu^*(\cup_k B_k) \leq \sum_{k=1}^\infty \mu^*(B_k)$ *(subadditivity)*

Parts 1. and 2. are immediately clear, while, to show 3., we pick $\varepsilon > 0$ and $k \in \mathbb{N}$ and find a countable cover $\{A_n^k\}_{n\in\mathbb{N}}$ with elements of $\mathcal{A}$ such that

$$\mu^*(B_k) \geq \sum_{n=1}^\infty \mu(A_n^k) + \tfrac{\varepsilon}{2^k}.$$

Using $\{A_n^k\}_{k\in\mathbb{N},n\in\mathbb{N}}$ as a candidate cover for $\cup_k B_k$, we conclude that $\mu^*(\cup_k B_k) \leq \sum_{k=1}^\infty \mu^*(B_k) + \varepsilon$. This being true for each $\varepsilon > 0$ implies 3.

We remark at this point that μ^* and μ coincide on $\mathcal{A}$. By using $(A, \emptyset, \emptyset, \dots)$ as a candidate countable cover of $A \in \mathcal{A}$, we can conclude that $\mu^*(A) \leq \mu(A)$, for all $A \in \mathcal{A}$. Conversely, suppose that $A \subseteq \cup_{n\in\mathbb{N}} A_n$ with $A_n \in \mathcal{A}$. Given that the elements of $\mathcal{A}$ form an algebra, we can assume that $A_n \subseteq A$, for all $n \in \mathbb{N}$ and that $\{A_n\}_{n\in\mathbb{N}}$ are pairwise disjoint, as any sequence covering A can be transformed into such a sequence without increasing $\sum_n \mu(A_n)$. The assumed countable additivity of μ on $\mathcal{A}$ now comes into play since, for partition $\{A_n\}_{n\in\mathbb{N}}$ of A into elements in $\mathcal{A}$, we necessarily have $\sum_n \mu(A_n) = \mu(A)$, and, so $\mu^*(A) \geq \mu(A)$.

Note: It is, probably, interesting to note that this is the only place in the entire proof where the countable additivity of μ on $\mathcal{A}$ is used.

PART II. The set-function μ^* is, in general, not a measure, but comes with the advantage of being defined on *all* subsets of S. The central idea of the proof is to recover countable additivity by restricting its domain a little. We say that a subset $M \subseteq S$ is **Caratheodory-measurable** or **μ^*-measurable** if

$$\mu^*(B) = \mu^*(B \cap M) + \mu^*(B \cap M^c) \text{ for all } B \subseteq S, \tag{2.3}$$

with the family of all μ^*-measurable subsets of S denoted by $\mathcal{M}^*$. We note that, by subadditivity, the equality sign in the definition of the measurability can be replaced by $\geq$; this will be used below.

The first thing we need to establish about $\mathcal{M}^*$ is that it is an algebra and that μ^* is a finitely-additive measure on $\mathcal{M}^*$. Clearly $\emptyset \in \mathcal{A}^*$ and the complement axiom follows directly from the symmetry in (2.3). Only the closure under finite unions needs some discussion, and, by induction, we only need to consider two-element unions; for that, we pick $M, N \in \mathcal{M}^*$, and introduce the following notation

$$M_{00} = M^c \cap N^c,\ M_{01} = M^c \cap N,\ M_{10} = M \cap N^c\ M_{11} = M \cap N. \tag{2.4}$$

By the measurability of M and N, for any $B \subseteq S$, we have

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap N^c) + \mu^*(B \cap N) \\ &= \mu^*(M_{00} \cap B) + \mu^*(M_{10} \cap B) + \mu^*(M_{01} \cap B) + \mu^*(M_{11} \cap B) \end{aligned}$$

On the other hand, $M_{01} \cup M_{10} \cup M_{11} = M \cup N$, so that, by subadditivity and (2.4), we have

$$\begin{aligned} &\mu^*\big((M \cup N)^c \cap B\big) + \mu^*\big((M \cup N) \cap B\big) \\ &\quad = \mu^*(M_{00} \cap B) + \mu^*\Big((M_{01} \cap B) \cup (M_{10} \cap B) \cup (M_{11} \cap B)\Big) \\ &\quad \leq \mu^*(M_{00} \cap B) + \mu^*(M_{10} \cap B) + \mu^*(M_{01} \cap B) + \mu^*(M_{11} \cap B) \\ &\quad = \mu^*(B), \end{aligned}$$

which implies that that $M \cup N \in \mathcal{M}^*$. When $M \cap N = \emptyset$ an application of measurability of N to $B = M \cup N$ yields the finite additivity of μ^* on $\mathcal{M}^*$:

$$\mu^*(M \cup N) = \mu^*((M \cup N) \cap N) + \mu^*((M \cup N) \cap N^c) = \mu^*(N) + \mu^*(M).$$

PART III. We now turn to the closure of $\mathcal{M}^*$ under countable unions and the σ-additive property of μ^*. Since $\mathcal{M}^*$ already known to be an algebra, it will be enough to show that it is closed under pairwise-disjoint unions, i.e., that $M \in \mathcal{M}^*$ whenever $\{M_n\}_{n \in \mathbb{N}}$ are pairwise disjoint elements in $\mathcal{M}^*$ with $M = \cup_n M_n$.

For $n \in \mathbb{N}$, we set $L_n = \cup_{k=1}^n M_k$ so that, for $B \subseteq S$, we have

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap L_n) + \mu^*(B \cap L_n^c) \\ &\geq \textstyle\sum_{k=1}^n \mu^*(B \cap M_k) + \mu^*(B \cap M^c), \end{aligned}$$

so that

$$\begin{aligned} \mu^*(B) &\geq \mu^*(B \cap M^c) + \textstyle\sum_{k \in \mathbb{N}} \mu^*(B \cap M_k) \\ &\geq \mu^*(B \cap M^c) + \mu^*(\cup_k (B \cap M_k)) + \mu^*(B \cap M^c) \\ &= \mu^*(B \cap M) + \mu^*(B \cap M^c). \end{aligned}$$

Since all the inequalities above need to be equalities, we immediately conclude that, with $B = S$,

$$\mu^*(M) = \sum_k \mu^*(M_k),$$

i.e., that μ^* is a countably-additive measure on $\mathcal{M}^*$. Since $\mathcal{A} \subseteq \mathcal{M}^*$, we have $\mathcal{M}^* \supseteq \sigma(\mathcal{A})$ and $\tilde{\mu} = \mu^*|_{\sigma(\mathcal{A})}$ is the required σ-additive extension of μ. $\square$

Back to the coin-toss space. In order to apply Theorem 2.9 in our situation, we need to check that μ is indeed a countably-additive measure on the algebra $\mathcal{A}$ of all cylinders. The following problem will help pinpoint the hard part of the argument:

Problem 2.2. Let $\mathcal{A}$ be an algebra on a non-empty set S, and let $\mu : \mathcal{A} \to [0, \infty]$ be a finite ($\mu(S) < \infty$) and finitely-additive set function on $\mathcal{S}$ with the following, additional, property:

$$\lim_n \mu(A_n) = 0, \text{ whenever } A_n \searrow \emptyset. \tag{2.5}$$

Then μ satisfies the conditions of Theorem 2.9.

The part about finite additivity is easy (perhaps a bit messy) and we leave it to the reader:

Problem 2.3. Show that the set-function μ_C, defined by (2.2) on the product cyllinders and extended by additivity to the algebra $\mathcal{A}$ of cylinders, is finitely additive.

Lemma 2.10 (Conditions of Caratheodory's theorem)**.** *The set-function μ_C, defined by* (2.2)*, and extenede by additivity to the the algebra $\mathcal{A}$ of cylinders, has the property* (2.5)*.*

Proof. By Problem 1.10, cylinders are closed sets, and so $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of closed sets whose intersection is empty. The same problem states that $\{-1,1\}^{\mathbb{N}}$ is compact, so, by the finite-intersection property[4], we have $A_{n_1} \cap \dots A_{n_k} = \emptyset$, for some finite collection $n_1, \dots, n_k$ of indices. Since $\{A_n\}_{n\in\mathbb{N}}$ is decreasing, we must have $A_n = \emptyset$, for all $n \geq n_k$, and, consequently, $\lim_n \mu(A_n) = 0$. □

[4] The *finite-intersection property* refers to the following fact, familiar from real analysis: If a family of closed sets of a compact topological space has empty intersection, then it admits a *finite* subfamily with an empty intersection.

Proposition 2.11 (Existence of the coin-toss measure)**.** *There exists a measure μ_C on $(\{-1,1\}^{\mathbb{N}}, \mathcal{S})$ with the property that (2.2) holds for all cylinders.*

Proof. Thanks to Lemma 2.10, Theorem 2.9 can now be used. □

In order to prove uniqueness, we will need the celebrated π-λ Theorem of Eugene Dynkin:

Theorem 2.12 (Dynkin's "π-λ" Theorem)**.** *Let $\mathcal{P}$ be a π-system on a non-empty set S, and let Λ be a λ-system which contains $\mathcal{P}$. Then Λ also contains the σ-algebra $\sigma(\mathcal{P})$ generated by $\mathcal{P}$.*

Proof. Using the result of part 4. of Problem 1.1, we only need to prove that $\lambda(\mathcal{P})$ (where $\lambda(\mathcal{P})$ denotes the λ-system generated by $\mathcal{P}$) is a π-system. For $A \subseteq S$, let $\mathcal{G}_A$ denote the family of all subsets of S whose intersections with A are in $\lambda(\mathcal{P})$:

$$\mathcal{G}_A = \{C \subseteq S : C \cap A \in \lambda(\mathcal{P})\}.$$

Claim: $\mathcal{G}_A$ is a λ-system for $A \in \lambda(\mathcal{P})$.

- Since $A \in \lambda(\mathcal{P})$, clearly $S \in \mathcal{G}_A$.
- For an increasing family $\{C_n\}_{n\in\mathbb{N}}$ in $\mathcal{G}_A$ we have $(\cup_n C_n) \cap A = \cup_n (C_n \cap A)$. Each $C_n \cap A$ is in Λ, and the family $\{C_n \cap A\}_{n\in\mathbb{N}}$ is increasing, so $(\cup_n C_n) \cap A \in \Lambda$.
- Finally, for $C_1, C_2 \in \mathcal{G}$ with $C_1 \subseteq C_2$, we have

$$(C_2 \setminus C_1) \cap A = (C_2 \cap A) \setminus (C_1 \cap A) \in \Lambda,$$

  because $C_1 \cap A \subseteq C_2 \cap A$.

$\mathcal{P}$ is a π-system, so for any $A \in \mathcal{P}$, we have $\mathcal{P} \subseteq \mathcal{G}_A$. Therefore, $\lambda(\mathcal{P}) \subseteq \mathcal{G}_A$, because $\mathcal{G}_A$ is a λ-system. In other words, for $A \in \mathcal{P}$ and $B \in \lambda(\mathcal{P})$, we have $A \cap B \in \lambda(\mathcal{P})$.

That means, however, that $\mathcal{P} \subseteq \mathcal{G}_B$, for any $B \in \lambda(\mathcal{P})$. Using the fact that $\mathcal{G}_B$ is a λ-system we must also have $\lambda(\mathcal{P}) \subseteq \mathcal{G}_B$, for any $B \in \lambda(\mathcal{P})$, i.e., $A \cap B \in \lambda(\mathcal{P})$, for all $A, B \in \lambda(\mathcal{P})$, which shows that $\lambda(\mathcal{P})$ is π-system. □

Proposition 2.13 (Measures which agree on a π-system)**.** *Let $(S, \mathcal{S})$ be a measurable space, and let $\mathcal{P}$ be a π-system which generates $\mathcal{S}$. Suppose that μ_1 and μ_2 are two measures on $\mathcal{S}$ with the property that $\mu_1(S) = \mu_2(S) < \infty$ and*

$$\mu_1(A) = \mu_2(A), \text{ for all } A \in \mathcal{P}.$$

Then $\mu_1 = \mu_2$, i.e., $\mu_1(A) = \mu_2(A)$, for all $A \in \mathcal{S}$.

Proof. Let $\mathcal{L}$ be the family of all subsets A of $\mathcal{S}$ for which $\mu_1(A) = \mu_2(A)$. Clearly $\mathcal{P} \subseteq \mathcal{L}$, but $\mathcal{L}$ is, potentially, bigger. In fact, it follows easily from the elementary properties of measures (see Proposition 2.6) and the fact that $\mu_1(S) = \mu_2(S) < \infty$ that it necessarily has the structure of a λ-system[5]. By Theorem 2.12 (the π-λ Theorem), $\mathcal{L}$ contains the σ-algebra generated by $\mathcal{P}$, i.e., $\mathcal{S} \subseteq \mathcal{L}$. On the other hand, by definition, $\mathcal{L} \subseteq \mathcal{S}$ and so $\mu_1 = \mu_2$. □

[5] It seems that the structure of a λ-system is defined so that it would exactly describe the structure of the family of all sets on which two measures (with the same total mass) agree. The structure of the π-system corresponds to the minimal assumption that allows Proposition 2.13 to hold.

Proposition 2.14 (Uniqueness of the coin-toss measure)**.** *The measure μ_C is the unique measure on $(\{-1, 1\}^{\mathbb{N}}, \mathcal{S})$ with the property that (2.2) holds for all cylinders.*

Proof. The existence is the content of Proposition 2.11. To prove uniqueness, it suffices to note that algebras are π-systems and use Proposition 2.13. □

Problem 2.4. Define $D_1, D_2 \subseteq \{-1, 1\}^{\mathbb{N}}$ by

1. $D_1 = \{s \in \{-1,1\}^{\mathbb{N}} : \limsup_n s_n = 1\}$,
2. $D_2 = \{s \in \{-1,1\}^{\mathbb{N}} : \exists\, N \in \mathbb{N},\, s_N = s_{N+1} = s_{N+2}\}$.

Show that $D_1, D_2 \in \mathcal{S}$ and compute $\mu(D_1), \mu(D_2)$.

Our next task is to probe the structure of the σ-algebra $\mathcal{S}$ on $\{-1,1\}^{\mathbb{N}}$ a little bit more and show that $\mathcal{S} \neq 2^{\{-1,1\}^{\mathbb{N}}}$. It is interesting that such a result (which deals exclusively with the structure of $\mathcal{S}$) requires a use of a measure in its proof.

Example 2.15 (A non-measurable subset of $\{-1,1\}^{\mathbb{N}}$ (*))**.** Since σ-algebras are closed under countable set operations, and since the product σ-algebra $\mathcal{S}$ for the coin-toss space $\{-1,1\}^{\mathbb{N}}$ is generated by sets obtained by restricting finite collections of coordinates, one is tempted to think that $\mathcal{S}$ contains *all* subsets of $\{-1,1\}^{\mathbb{N}}$. That is not the case. We will use the axiom of choice, together with the fact that a measure μ_C can be defined on the whole of $\{-1,1\}^{\mathbb{N}}$, to show to "construct" an example of a non-measurable set.

Let us start by constructing a relation $\sim$ on $\{-1,1\}^{\mathbb{N}}$ in the following[6] way: we set $s^1 \sim s^2$ if and only if there exists $n \in \mathbb{N}$ such that $s^1_k = s^2_k$, for $k \geq n$ (here, as always, $s^i = (s^i_1, s^i_2, \dots)$, $i = 1,2$). It is easy to check that $\sim$ is an equivalence relation and that it splits $\{-1,1\}^{\mathbb{N}}$ into disjoint equivalence classes. One of the many equivalent forms of the axiom of choice states that there exists a subset N of $\{-1,1\}^{\mathbb{N}}$ which contains exactly one element from each of the equivalence classes.

[6] In words, s^1 and s^2 are related if they only differ in a finite number of coordinates.

Let us suppose that N is an element in $\mathcal{S}$ and see if we can reach a contradiction. For each nonempty $\boldsymbol{n} = \{n_1, \dots, n_k\} \in 2^{\mathbb{N}}_{fin}$, where $2^{\mathbb{N}}_{fin}$ denotes the family of all finite subsets of $\mathbb{N}$, let us define the mapping $T_{\boldsymbol{n}} : \{-1,1\}^{\mathbb{N}} \to \{-1,1\}^{\mathbb{N}}$ in the following[7] manner:

[7] $T_{\boldsymbol{n}}$ flips the signs of the elements of its argument on the positions corresponding to $\boldsymbol{n}$.

$$T_{\emptyset} = \mathrm{Id} \text{ and } (T_{\boldsymbol{n}}(s))_l = \begin{cases} -s_l, & l \in \boldsymbol{n}, \\ s_l, & l \notin \boldsymbol{n}, \end{cases} \text{ for } \boldsymbol{n} \in \mathbb{N}.$$

Since $\boldsymbol{n}$ is finite, $T_{\boldsymbol{n}}$ preserves the $\sim$-equivalence class of each element. Consequently (and using the fact that N contains exactly one element from each equivalence class) the sets N and $T_{\boldsymbol{n}}(N) = \{T_{\boldsymbol{n}}(s) : s \in N\}$ are disjoint. Similarly and more generally, the sets $T_{\boldsymbol{n}}(N)$ and $T_{\boldsymbol{n}'}(N)$ are also disjoint whenever $\boldsymbol{n} \neq \boldsymbol{n}'$. On the other hand, each $s \in \{-1,1\}^{\mathbb{N}}$ is equivalent to some $\hat{s} \in N$, i.e., it can be obtained from $\hat{s}$ by flipping a finite number of coordinates. Therefore, the family

$$\mathcal{N} = \{T_{\boldsymbol{n}}(N) : \boldsymbol{n} \in 2^{\mathbb{N}}_{fin}\}$$

forms a partition of $\{-1,1\}^{\mathbb{N}}$.

The mapping $T_{\boldsymbol{n}}$ has several other nice properties. First of all, it is immediate that it is involutory, i.e., $T_{\boldsymbol{n}} \circ T_{\boldsymbol{n}} = \mathrm{Id}$. To show that it is $(\mathcal{S}, \mathcal{S})$-measurable, we need to prove that its composition with each projection map $\pi_k : \mathcal{S} \to \{-1, 1\}$ is measurable. This follows immediately from the fact that for $k \in \mathbb{N}$

$$(\pi_k \circ T_{\boldsymbol{n}})^{-1}(\{1\}) = \begin{cases} C_{k;1}, & k \notin \boldsymbol{n}, \\ C_{k;-1}, & k \in \boldsymbol{n}, \end{cases}$$

where, for $b \in \{-1, 1\}$, we recall that $C_{k;b} = \{\boldsymbol{s} \in \{-1, 1\}^{\mathbb{N}} : s_k = b\}$ is a product cylinder. If we combine the involutivity and measurability of $T_{\boldsymbol{n}}$, we immediately conclude that $T_{\boldsymbol{n}}(A) \in \mathcal{S}$ for each $A \in \mathcal{S}$. In particular, $\mathcal{N} \subseteq \mathcal{S}$.

In addition to preserving measurability, the map $T_{\boldsymbol{n}}$ also preserves the measure[8] the in μ_C, i.e., $\mu_C(T_{\boldsymbol{n}}(A)) = \mu_C(A)$, for all $A \in \mathcal{S}$. To prove that, let us pick $\boldsymbol{n} \in F$ and consider the set-function $\mu_{\boldsymbol{n}} : \mathcal{S} \to [0, 1]$ given by

$$\mu_{\boldsymbol{n}}(A) = \mu_C(T_{\boldsymbol{n}}(A)).$$

It is a simple matter to show that $\mu_{\boldsymbol{n}}$ is, in fact, a measure on $(S, \mathcal{S})$ with $\mu_{\boldsymbol{n}}(S) = 1$. Moreover, thanks to the simple form (2.2) of the action of the measure μ_C on cylinders, it is clear that $\mu_{\boldsymbol{n}} = \mu_C$ on the algebra of all cylinders. It suffices to invoke Proposition 2.13 to conclude that $\mu_{\boldsymbol{n}} = \mu_C$ on the entire $\mathcal{S}$, i.e., that $T_{\boldsymbol{n}}$ preserves μ_C.

The above properties of the maps $T_{\boldsymbol{n}}$, $\boldsymbol{n} \in F$ can imply the following: $\mathcal{N}$ is a partition of $\mathcal{S}$ into countably many measurable subsets of equal measure. Such a partition $\{N_1, N_2, \dots\}$ cannot exist, however. Indeed if it did, one of the following two cases would occur:

1. $\mu(N_1) = 0$. In that case
$$\mu(S) = \mu(\cup_k N_k) = \sum_n \mu(N_k) = \sum_n 0 = 0 \neq 1 = \mu(S).$$
2. $\mu(N_1) = \alpha > 0$. In that case
$$\mu(S) = \mu(\cup_k N_k) = \sum_n \mu(N_k) = \sum_n \alpha = \infty \neq 1 = \mu(S).$$

Therefore, the set N cannot be measurable[9] in $\mathcal{S}$.

The Lebesgue measure

As we shall see, the coin-toss space can be used as a sort of a universal measure space in probability theory. We use it here to construct the Lebesgue measure on $[0, 1]$. We start with the notion somewhat dual to the already introduced notion of the pull-back in Definition 1.8. We leave it as an exercise for the reader to show that the set function $f_*\mu$ from Definition 2.16 is indeed a measure.

[8] Actually, we say that a map f from a measure space $(S, \mathcal{S}, \mu_S)$ to the measure space $(T, \mathcal{T}, \mu_T)$ is **measure preserving** if it is measurable and $\mu_S(f^{-1}(A)) = \mu_T(A)$, for all $A \in \mathcal{T}$. The involutivity of the map $T_{\boldsymbol{n}}$ implies that this general definition agrees with our usage in this example.

[9] Somewhat heavier set-theoretic machinery can be used to prove that most of the subsets of S are not in $\mathcal{S}$, in the sense that the cardinality of the set $\mathcal{S}$ is strictly smaller than the cardinality of the set 2^S of all subsets of S

Definition 2.16 (Push-forwards)**.** Let $(S, \mathcal{S}, \mu)$ be a measure space and let $(T, \mathcal{T})$ be a measurable space. The measure $f_*\mu$ on $(T, \mathcal{T})$, defined by

$$f_*\mu(B) = \mu(f^{-1}(B)), \text{ for } B \in \mathcal{T},$$

is called the **push-forward** of the measure μ by f.

Let $f : \{-1, 1\}^{\mathbb{N}} \to [0, 1]$ be the mapping given by

$$f(\boldsymbol{s}) = \sum_{k=1}^{\infty} \left(\tfrac{1+s_k}{2}\right) 2^{-k}, \ \boldsymbol{s} \in \{-1, 1\}^{\mathbb{N}}.$$

The idea is to use f to establish a correspondence between all real numbers in $[0, 1]$ and their expansions in the binary system, with the coding $-1 \mapsto 0$ and $1 \mapsto 1$. It is interesting to note that f is not one-to-one[10] as it, for example, maps $\boldsymbol{s}_1 = (1, -1, -1, \dots)$ and $\boldsymbol{s}_2 = (-1, 1, 1, \dots)$ into the same value - namely $\frac{1}{2}$. Let us show, first, that the map f is continuous in the metric d defined by part (1.2) of Problem 1.9. Indeed, we pick $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ in $\{-1, 1\}^{\mathbb{N}}$ and remember that for $d(\boldsymbol{s}_1, \boldsymbol{s}_2) \leq 2^{-n}$, the first $n - 1$ coordinates of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ coincide. Therefore,

[10] The reason for this is, poetically speaking, that $[0, 1]$ is not the Cantor set.

$$|f(\boldsymbol{s}_1) - f(\boldsymbol{s}_2)| \leq \sum_{k=n}^{\infty} 2^{-k} = 2^{-n+1} = 2d(\boldsymbol{s}_1, \boldsymbol{s}_2).$$

Hence, the map f is Lipschitz and, therefore, continuous.

The continuity of f (together with the fact that $\mathcal{S}$ is the Borel σ-algebra for the topology induced by the metric d) implies that $f : (\{-1, 1\}^{\mathbb{N}}, \mathcal{S}) \to ([0, 1], \mathcal{B}([0, 1]))$ is a measurable mapping. Therefore, the push-forward $\lambda = f_*(\mu)$ is well defined on $([0, 1], \mathcal{B}([0, 1]))$, and we call it the **Lebesgue measure** on $[0, 1]$.

Proposition 2.17 (Intuitive properties of the Lebesgue measure)**.** *The Lebesgue measure λ on $([0, 1], \mathcal{B}([0, 1]))$ satisfies*

$$\lambda([a, b)) = b - a, \ \lambda(\{a\}) = 0 \textit{ for } 0 \leq a < b \leq 1. \tag{2.6}$$

Proof.

1. Consider a, b of the form $b = \frac{k}{2^n}$ and $b = \frac{k+1}{2^n}$, for $n \in \mathbb{N}$ and $k < 2^n$. For such a, b we have $f^{-1}([a, b)) = C_{1,\dots,n;c_1,c_2,\dots,c_n}$, where $\overline{c_1 c_2 \dots c_n}$ is the base-2 expansion of k (after the "recoding" $-1 \mapsto 0$, $1 \mapsto 1$). By the very definition of λ and the form (2.2) of the action of the coin-toss measure μ_C on cylinders, we have

$$\lambda\Big([a, b)\Big) = \mu_C\Big(f^{-1}([a, b))\Big) = \mu_C(C_{1,\dots,n;c_1,c_2,\dots,c_n}) = 2^{-n} = \tfrac{k+1}{2^n} - \tfrac{k}{2^n}.$$

Therefore, (2.6) holds for a, b of the form $b = \frac{k}{2^n}$ and $b = \frac{l}{2^n}$, for $n \in \mathbb{N}$, $k < 2^n$ and $l = k + 1$. Using (finite) additivity of λ, we immediately conclude that (2.6) holds for all k, l, i.e., that it holds for all dyadic rationals. A general $a \in (0,1]$ can be approximated by an increasing sequence $\{q_n\}_{n\in\mathbb{N}}$ of dyadic rationals from the left, and the continuity of measures with respect to decreasing sequences implies that

$$\lambda\Big([a,p)\Big) = \lambda\Big(\cap_n [q_n,p)\Big) = \lim_n \lambda\Big([q_n,p)\Big) = \lim_n (p-q_n) = (p-a),$$

whenever $a \in (0,1]$ and p is a dyadic rational. In order to remove the dyadicity requirement from the right limit, we approximate it from the left by a sequence $\{p_n\}_{n\in\mathbb{N}}$ of dyadic rationals with $p_n > a$, and use the continuity with respect to increasing sequences to get, for $a < b \in (0,1)$,

$$\lambda\Big([a,b)\Big) = \lambda\Big(\cup_n [a,p_n)\Big) = \lim_n \lambda\Big([a,p_n)\Big) = \lim_n (p_n-a) = (b-a).$$

□

The Lebesgue measure has another important property:

Problem 2.5. Show that the Lebesgue measure is **translation invariant**. More precisely, for $B \in \mathcal{B}([0,1])$ and $x \in [0,1)$, we have

Hint: Use Proposition 2.13 for the second part.

1. $B +_1 x = \{b + x \pmod 1 \,:\, b \in B\}$ is in $\mathcal{B}([0,1])$ and
2. $\lambda(B +_1 x) = \lambda(B)$,

where, for $a \in [0,2)$, we define

Geometrically, the set $x +_1 B$ is obtained from B by translating it to the right by x and then shifting the part that is "sticking out" by 1 to the left.

$$a \pmod 1 = \begin{cases} a, & a \le 1, \\ a-1, & a > 1. \end{cases}$$

Finally, the notion of the Lebesgue measure is just as useful on the entire $\mathbb{R}$, as on its compact subset $[0,1]$. For a general $B \in \mathcal{B}(\mathbb{R})$, we can define the Lebesgue measure of B by measuring its intersections with all intervals of the form $[n, n+1)$, and adding them together, i.e.,

$$\lambda(B) = \sum_{n=-\infty}^{\infty} \lambda\Big(\big(B \cap [n,n+1)\big) - n\Big).$$

Note how we are overloading the notation and using the letter λ for both the Lebesgue measure on $[0,1]$ and the Lebesgue measure on $\mathbb{R}$.

It is a quite tedious, but does not require any new tools, to show that many of the properties of λ on $[0,1]$ transfer to λ on $\mathbb{R}$:

Problem 2.6. Let λ be the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Show that

1. $\lambda([a,b)) = b - a$, $\lambda(\{a\}) = 0$ for $a < b$,

2. λ is σ-finite but not finite,

3. $\lambda(B+x) = \lambda(B)$, for all $B \in \mathcal{B}(\mathbb{R})$ and $x \in \mathbb{R}$, where $B + x = \{b + x : b \in B\}$.

Additional Problems

Problem 2.7 (Local separation by constants)**.** Let $(S, \mathcal{S}, \mu)$ be a measure space and let $f, g \in \mathcal{L}^0(S, \mathcal{S}, \mu)$ satisfy $\mu\big(\{x \in S : f(x) < g(x)\}\big) > 0$. Prove or construct a counterexample for the following statement:

> "There exist constants $a, b \in \mathbb{R}$ such that
> $\mu\big(\{x \in S : f(x) \leq a < b \leq g(x)\}\big) > 0$."

Problem 2.8 (A pseudometric on sets)**.** Let $(S, \mathcal{S}, \mu)$ be a finite measure space. For $A, B \in \mathcal{S}$ define

$$d(A, B) = \mu(A \triangle B),$$

where $\triangle$ denotes the symmetric difference: $A \triangle B = (A \setminus B) \cup (B \setminus A)$. Show that d is a pseudometric on $\mathcal{S}$, and for $A \in \mathcal{S}$ describe the set of all $B \in \mathcal{S}$ with $d(A, B) = 0$.

Note: Let X be a nonempty set. A function $d : X \times X \to [0, \infty)$ is called a **pseudometric** if

1. $d(x, y) + d(y, x) \geq d(x, z)$, for all $x, y, z \in X$,
2. $d(x, y) = d(y, x)$, for all $x, y \in X$, and
3. $d(x, x) = 0$, for all $x \in X$.

Note how the only difference between a metric and a pseudometric is that for a metric $d(x, y) = 0$ implies $x = y$, while no such requirement is imposed on a pseudometric.

Problem 2.9 (Complete measure spaces)**.** A measure space $(S, \mathcal{S}, \mu)$ is called **complete** if all subsets of null sets are themselves in $\mathcal{S}$. For a (possibly incomplete) measure space $(S, \mathcal{S}, \mu)$ we define the **completion** $(S, \mathcal{S}^*, \mu^*)$ in the following way:

$$\mathcal{S}^* = \{A \cup N^* : A \in \mathcal{S} \text{ and } N^* \subseteq N \text{ for some } N \in \mathcal{S} \text{ with } \mu(N) = 0\}.$$

For $B \in \mathcal{S}^*$ with representation $B = A \cup N^*$ we set $\mu^*(B) = \mu(A)$.

Note: Unfortunately, the same notation μ^* is often used for the completion of the measure μ and the outer measure associated with μ as in the proof of Theorem 2.9. Fortunately, it can be shown that these two object coincide on the domain of the completion.

1. Show that $\mathcal{S}^*$ is a σ-algebra.

2. Show that the definition $\mu^*(B) = \mu(A)$ above does not depend on the choice of the decomposition $B = A \cup N^*$, i.e., that $\mu(\hat{A}) = \mu(A)$ if $B = \hat{A} \cup \hat{N}^*$ is another decomposition of B into a set $\hat{A}$ in $\mathcal{S}$ and a subset $\hat{N}$ of a null set in $\mathcal{S}$.

3. Show that μ^* is a measure on $(S, \mathcal{S}^*)$ and that $(S, \mathcal{S}^*, \mu^*)$ is a complete measure space with the property that $\mu^*(A) = \mu(A)$, for $A \in \mathcal{S}$.

Problem 2.10 (The Cantor set)**.** The **Cantor set** is defined as the collection of all real numbers x in $[0, 1]$ with the representation

$$x = \sum_{n=1}^{\infty} c_n 3^{-n}, \text{ where } c_n \in \{0, 2\}.$$

Show that it is Borel-measurable and compute its Lebesgue measure.

Problem 2.11 (The uniform measure on a circle)**.** Let S^1 be the unit circle, and let $f : [0,1) \to S^1$ be the "winding map"

$$f(x) = \Big(\cos(2\pi x), \sin(2\pi x)\Big), \; x \in [0,1).$$

1. Show that the map f is $(\mathcal{B}([0,1)), \mathcal{S}^1)$-measurable, where $\mathcal{S}^1$ denotes the Borel σ-algebra on S^1 (with the topology inherited from $\mathbb{R}^2$).

2. For $\alpha \in (0, 2\pi)$, let R_α denote the (counter-clockwise) rotation of $\mathbb{R}^2$ with center $(0,0)$ and angle α. Show that $R_\alpha(A) = \{R_\alpha(x) : x \in A\}$ is in $\mathcal{S}^1$ if and only if $A \in \mathcal{S}^1$.

3. Let μ^1 be the push-forward of the Lebesgue measure λ by the map f. Show that μ^1 is rotation-invariant, i.e., that $\mu^1(A) = \mu^1\big(R_\alpha(A)\big)$.

Note: The measure μ^1 is called the **uniform measure** (or the **uniform distribution**) on S^1.

Problem 2.12 (Asymptotic densities)**.** We say that the subset A of $\mathbb{N}$ **admits asymptotic density** if the limit

$$d(A) = \lim_n \frac{\#(A \cap \{1, 2, \dots, n\})}{n},$$

exists (remember that # denotes the number of elements of a set). Let $\mathcal{D}$ be the collection of all subsets of $\mathbb{N}$ which admit asymptotic density.

1. Is $\mathcal{D}$ an algebra? A σ-algebra?

2. Is the map $A \mapsto d(A)$ finitely-additive on $\mathcal{D}$? A measure?

Problem 2.13 (A subset of the coin-toss space)**.** An element in $\{-1,1\}^{\mathbb{N}}$ (i.e., a sequence $\boldsymbol{s} = (s_1, s_2, \dots)$ where $s_n \in \{-1,1\}$ for all $n \in \mathbb{N}$) is said to be **eventually periodic** if there exists $N_0, K \in \mathbb{N}$ such that $s_n = s_{n+K}$ for all $n \geq N_0$. Let $P \subseteq \{-1,1\}^{\mathbb{N}}$ be the collection of all eventually-period sequences. Show that P is measurable in the product σ-algebra $\mathcal{S}$ and compute $\mu_C(P)$.

Problem 2.14 (Regular measures)**.** The measure space $(S, \mathcal{S}, \mu)$, where (S, d) is a metric space and $\mathcal{S}$ is a σ-algebra on S which contains the Borel σ-algebra $\mathcal{B}(d)$ on S is called **regular** if for each $A \in \mathcal{S}$ and each $\varepsilon > 0$ there exist a closed set C and an open set O such that $C \subseteq A \subseteq O$ and $\mu(O \setminus C) < \varepsilon$.

1. Suppose that $(S, \mathcal{S}, \mu)$ is a regular measure space, and that the measure space $(S, \mathcal{B}(d), \mu|_{\mathcal{B}(d)})$ is obtained from $(S, \mathcal{S}, \mu)$ by restricting the measure μ onto the σ-algebra of Borel sets. Show that $\mathcal{S} \subseteq \mathcal{B}(d)^*$, where $\big(S, \mathcal{B}(d)^*, (\mu|_{\mathcal{B}(d)})^*\big)$ is the completion (in the sense of Problem 2.9) of $(S, \mathcal{B}(d), \mu|_{\mathcal{B}(d)})$

2. Suppose that (S, d) is a metric space and that μ is a finite measure on $\mathcal{B}(d)$. Show that $(S, \mathcal{B}(d), \mu)$ is a regular measure space.

Hint: Consider a collection $\mathcal{A}$ of subsets A of S such that for each $\varepsilon > 0$ there exists a closed set C and an open set O with $C \subseteq A \subseteq O$ and $\mu(O \setminus C) < \varepsilon$. Argue that $\mathcal{A}$ is a σ-algebra. Then show that each closed set can be written as an intersection of open sets; use (but prove, first) the fact that the map

$$x \mapsto d(x, C) = \inf\{d(x, y) : y \in C\},$$

is continuous on S for any nonempty $C \subseteq S$.

3. Show that $(S, \mathcal{B}(d), \mu)$ is regular if μ is not necessarily finite, but has the property that $\mu(A) < \infty$ whenever $A \in \mathcal{B}(d)$ is bounded, i.e., when $\sup\{d(x, y) : x, y \in A\} < \infty$.

   Hint: Pick a point $x_0 \in S$ and, for $n \in \mathbb{N}$, define the family $\{R_n\}_{n\in\mathbb{N}}$ of subsets of S as follows:

$$R_1 = \{x \in S : d(x, x_0) < 2\}, \text{ and}$$
$$R_n = \{x \in S : n - 1 < d(x, x_0) < n + 1\}, \text{ for } n > 1,$$

   as well as a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of set functions on $\mathcal{B}(d)$, given by $\mu_n(A) = \mu(A \cap R_n)$, for $A \in \mathcal{B}(d)$. Under the right circumstances, even countable unions of closed sets are closed.

4. Conclude that the Lebesgue measure on $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\right)$ is regular.

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 3
The Lebesgue Integral

The construction of the integral

Unless expressly specified otherwise, we pick and fix a measure space $(S, \mathcal{S}, \mu)$ and assume that all functions under consideration are defined there.

Definition 3.1 (Simple functions)**.** A function $f \in \mathcal{L}^0(S, \mathcal{S}, \mu)$ is said to be **simple** if it takes only a finite number of values.

The collection of all simple functions is denoted by $\mathcal{L}^{\text{Simp},0}$ (more precisely by $\mathcal{L}^{\text{Simp},0}(S, \mathcal{S}, \mu)$) and the family of non-negative simple functions by $\mathcal{L}_+^{\text{Simp},0}$. Clearly, a simple function $f : S \to \mathbb{R}$ admits a (not necessarily unique) representation

$$f = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k}, \tag{3.1}$$

for $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ and $A_1, \dots, A_n \in \mathcal{S}$. Such a representation is called the **simple-function representation** of f.

When the sets A_k, $k = 1, \dots, n$ are intervals in $\mathbb{R}$, the graph of the simple function f looks like a collection of steps (of heights $\alpha_1, \dots, \alpha_n$). For that reason, the simple functions are sometimes referred to as *step functions*. The Lebesgue integral is very easy to define for non-negative simple functions and this definition allows for further generalizations[1]:

[1] In fact, the progression of events you will see in this section is typical for measure theory: you start with indicator functions, move on to non-negative simple functions, then to general non-negative measurable functions, and finally to (not-necessarily-non-negative) measurable functions. This approach is so common, that it has a name - **the Standard Machine**.

Definition 3.2 (Lebesgue integration for simple functions)**.** For $f \in \mathcal{L}_+^{\text{Simp},0}$ we define the **(Lebesgue) integral** $\int f \, d\mu$ of f with respect to μ by

$$\int f \, d\mu = \sum_{k=1}^{n} \alpha_k \mu(A_k) \in [0, \infty],$$

where $f = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k}$ is a simple-function representation of f,

Problem 3.1. Show that the Lebesgue integral is well-defined for simple functions, i.e., that the value of the expression $\sum_{k=1}^{n} \alpha_k \mu(A_k)$ does not depend on the choice of the simple-function representation of f.

Remark 3.3.

1. It is important to note that $\int f\,d\mu$ can equal $+\infty$ even if f never takes the value $+\infty$. It is enough to pick $f = \mathbf{1}_A$ where $\mu(A) = +\infty$ - indeed, then $\int f\,d\mu = 1\mu(A) = \infty$, but f only takes values in the set $\{0, 1\}$. This is one of the reasons we start with *non-negative* functions. Otherwise, we would need to deal with the (unsolvable) problem of computing $\infty - \infty$. On the other hand, such examples cannot be constructed when μ is a finite measure. Indeed, it is easy to show that when $\mu(S) < \infty$, we have $\int f\,d\mu < \infty$ for all $f \in \mathcal{L}_+^{\text{Simp},0}$.

2. One can think of the (simple) Lebesgue integral as a generalization of the notion of (finite) additivity of measures. Indeed, if the simple-function representation of f is given by $f = \sum_{k=1}^n \mathbf{1}_{A_k}$, for pairwise disjoint $A_1, \dots, A_n$, then the equality of the values of the integrals for two representations $f = \mathbf{1}_{\cup_{k=1}^n A_k}$ and $f = \sum_{k=1}^n \mathbf{1}_{A_k}$ is a simple restatement of finite additivity. When $A_1, \dots, A_n$ are not disjoint, then the finite additivity gives way to finite subadditivity
$$\mu(\cup_{k=1}^n A_k) \leq \sum_{k=1}^n \mu(A_k),$$
but the integral $\int f\,d\mu$ "takes into account" those x which are covered by more than one A_k, $k = 1, \dots, n$. Take, for example, $n = 2$ and $A_1 \cap A_2 = C$. Then
$$f = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} = \mathbf{1}_{A_1 \setminus C} + 2\mathbf{1}_C + \mathbf{1}_{A_2 \setminus C},$$
and so
$$\int f\,d\mu = \mu(A_1 \setminus C) + \mu(A_2 \setminus C) + 2\mu(C) = \mu(A_1) + \mu(A_2) + \mu(C).$$

It is easy to see that $\mathcal{L}_+^{\text{Simp},0}$ is a **convex cone**, i.e., that it is closed under finite linear combinations with non-negative coefficients. The integral map $f \mapsto \int f\,d\mu$ preserves this structure:

Problem 3.2. For $f_1, f_2 \in \mathcal{L}_+^{\text{Simp},0}$ and $\alpha_1, \alpha_2 \geq 0$ we have

1. if $f_1(x) \leq f_2(x)$ for all $x \in S$ then $\int f_1\,d\mu \leq \int f_2\,d\mu$, and
2. $\int (\alpha_1 f_1 + \alpha_2 f_2)\,d\mu = \alpha_1 \int f_1\,d\mu + \alpha_2 \int f_2\,d\mu$.

Having defined the integral for $f \in \mathcal{L}_+^{\text{Simp},0}$, we turn to general[2] non-negative measurable functions. In fact, at no extra cost we can consider a slightly larger set consisting of all measurable $[0, \infty]$-valued functions which we denote by $\mathcal{L}_+^0([0, \infty])$.

[2] Even though there is no obvious advantage at this point of integrating a function which takes the value $+\infty$, it will become clear soon how convenient it really is.

Definition 3.4 (Lebesgue integral for nonnegative functions)**.** For a function $f \in \mathcal{L}^0_+([0,\infty])$, we define its **Lebesgue integral** $\int f\,d\mu$ by

$$\int f\,d\mu = \sup\left\{\int g\,d\mu \,:\, g \in \mathcal{L}_+^{\text{Simp},0},\ g(x) \le f(x),\ \forall\, x \in S\right\} \in [0,\infty].$$

Note: While there is no question that this definition produces a unique number $\int f\,d\mu$, one can wonder if it matches the previously given definition of the Lebesgue integral for simple functions. A simple argument based on the monotonicity property of part 1. of Problem 3.2 can be used to show that this is, indeed, the case.

Problem 3.3. Show that $\int f\,d\mu = \infty$ if there exists a measurable set A with $\mu(A) > 0$ such that $f(x) = \infty$ for $x \in A$. On the other hand, show that $\int f\,d\mu = 0$ for f of the form

$$f(x) = \infty \mathbf{1}_A(x) = \begin{cases} \infty, & x \in A, \\ 0, & x \neq A, \end{cases}$$

whenever $\mu(A) = 0$.

Note: Relate this to our convention that $\infty \times 0 = 0 \times \infty = 0$.

Finally, we are ready to define the integral for general measurable functions. Each $f \in \mathcal{L}^0$ can be written as a difference of two functions in $\mathcal{L}^0_+$ in many ways. There exists a decomposition which is, in a sense, minimal. We define

$$f^+ = \max(f,0),\ f^- = \max(-f,0),$$

so that $f = f^+ - f^-$ (and both f^+ and f^- are measurable). The minimality we mentioned above is reflected in the fact that for each $x \in S$, at most one of f^+ and f^- is non-zero.

Definition 3.5 (Integrable functions)**.** A function $f \in \mathcal{L}^0$ is said to be **integrable** if

$$\int f^+\,d\mu < \infty \text{ and } \int f^-\,d\mu < \infty.$$

The collection of all integrable functions in $\mathcal{L}^0$ is denoted by $\mathcal{L}^1$. The family of integrable functions is tailor-made for the following definition:

Definition 3.6 (The Lebesgue integral)**.** For $f \in \mathcal{L}^1$, we define the **Lebesgue integral** $\int f\,d\mu$ of f by

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

Remark 3.7.

1. We have seen so far two cases in which an integral for a function $f \in \mathcal{L}^0$ can be defined: when $f \ge 0$ or when $f \in \mathcal{L}^1$. It is possible to combine the two and define the Lebesgue integral for all functions $f \in \mathcal{L}^0$ with $f^- \in \mathcal{L}^1$. The set of all such functions is denoted by

$\mathcal{L}^{0-1}$ and we set

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu \in (-\infty,\infty], \text{ for } f \in \mathcal{L}^{0-1}.$$

Note that no problems of the form $\infty - \infty$ arise here, and also note that, like $\mathcal{L}^0_+$, $\mathcal{L}^{0-1}$ is only a convex cone, and not a vector space. While the notation $\mathcal{L}^0$ and $\mathcal{L}^1$ is quite standard, the one we use for $\mathcal{L}^{0-1}$ is not.

2. For $A \in \mathcal{S}$ and $f \in \mathcal{L}^{0-1}$ we usually write $\int_A f\,d\mu$ for $\int f \mathbf{1}_A\,d\mu$.

Problem 3.4. Show that the Lebesgue integral remains a monotone operation in $\mathcal{L}^{0-1}$. More precisely, show that if $f \in \mathcal{L}^{0-1}$ and $g \in \mathcal{L}^0$ are such that $g(x) \geq f(x)$, for all $x \in S$, then $g \in \mathcal{L}^{0-1}$ and $\int g\,d\mu \geq \int f\,d\mu$.

First properties of the integral

The wider the generality to which a definition applies, the harder it is to prove theorems about it. Linearity of the integral is a trivial matter for functions in $\mathcal{L}^{Simp,0}_+$, but you will see how much we need to work to get it for $\mathcal{L}^0_+$. In fact, it seems that the easiest route towards linearity is through two important results: an approximation theorem and a convergence theorem. Before that, we need to pick some low-hanging fruit:

Problem 3.5. Show that for $f_1, f_2 \in \mathcal{L}^0_+([0,\infty])$ and $\alpha \in [0,\infty]$ we have

1. if $f_1(x) \leq f_2(x)$ for all $x \in S$ then $\int f_1\,d\mu \leq \int f_2\,d\mu$.
2. $\int \alpha f\,d\mu = \alpha \int f\,d\mu$.

Theorem 3.8 (Monotone convergence theorem)**.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0_+([0,\infty])$ with the property that*

$$f_1(x) \leq f_2(x) \leq \dots \text{ for all } x \in S.$$

Then

$$\lim_n \int f_n\,d\mu = \int f\,d\mu,$$

where $f(x) = \lim_n f_n(x) \in \mathcal{L}^0_+([0,\infty])$, for $x \in S$.

Proof. The (monotonicity) property (1) of Problem 3.5 above implies that the sequence $\int f_n\,d\mu$ is non-decreasing and that $\int f_n\,d\mu \leq \int f\,d\mu$. Therefore, $\lim_n \int f_n\,d\mu \leq \int f\,d\mu$. To show the opposite inequality, we deal with the case $\int f\,d\mu < \infty$ and pick $\varepsilon > 0$ and $g \in \mathcal{L}^{Simp,0}_+$ with $g(x) \leq f(x)$, for all $x \in S$ and $\int g\,d\mu \geq \int f\,d\mu - \varepsilon$ (the case $\int f\,d\mu = \infty$

is similar and left to the reader). For $0 < c < 1$, define the (measurable) sets $\{A_n\}_{n\in\mathbb{N}}$ by

$$A_n = \{f_n \geq cg\}, n \in \mathbb{N}.$$

By the increase of the sequence $\{f_n\}_{n\in\mathbb{N}}$, the sets $\{A_n\}_{n\in\mathbb{N}}$ also increase. Moreover, since the function cg satisfies $cg(x) \leq g(x) \leq f(x)$ for all $x \in S$ and $cg(x) < f(x)$ when $f(x) > 0$, the increasing convergence $f_n \to f$ implies that $\cup_n A_n = S$. By non-negativity of f_n and monotonicity,

$$\int f_n \, d\mu \geq \int f_n \mathbf{1}_{A_n} \, d\mu \geq c \int g \mathbf{1}_{A_n} \, d\mu,$$

and so

$$\sup_n \int f_n \, d\mu \geq c \sup_n \int g \mathbf{1}_{A_n} \, d\mu.$$

Let $g = \sum_{i=1}^k \alpha_i \mathbf{1}_{B_i}$ be a simple-function representation of g. Then

$$\int g \mathbf{1}_{A_n} \, d\mu = \int \sum_{i=1}^k \alpha_i \mathbf{1}_{B_i \cap A_n} \, d\mu = \sum_{i=1}^k \alpha_i \mu(B_i \cap A_n).$$

Since $A_n \nearrow S$, we have $A_n \cap B_i \nearrow B_i$, $i = 1, \dots, k$, and the continuity of measure implies that $\mu(A_n \cap B_i) \nearrow \mu(B_i)$. Therefore,

$$\int g \mathbf{1}_{A_n} \, d\mu \nearrow \sum_{i=1}^k \alpha_i \mu(B_i) = \int g \, d\mu.$$

Consequently,

$$\lim_n \int f_n \, d\mu = \sup_n \int f_n \, d\mu \geq c \int g \, d\mu, \text{ for all } c \in (0, 1),$$

and the proof is completed when we let $c \to 1$. □

Remark 3.9.

1. The monotone convergence theorem is really about the robustness of the Lebesgue integral. Its stability with respect to limiting operations is one of the reasons why it is a de-facto "industry standard".

2. The "monotonicity" condition in the monotone convergence theorem cannot be dropped. Take, for example $S = [0, 1]$, $\mathcal{S} = \mathcal{B}([0, 1])$, and $\mu = \lambda$ (the Lebesgue measure), and define

$$f_n = n \mathbf{1}_{(0, n^{-1}]}, \text{ for } n \in \mathbb{N}.$$

Then $f_n(0) = 0$ for all $n \in \mathbb{N}$ and $f_n(x) = 0$ for $n > \frac{1}{x}$ and $x > 0$. In either case $f_n(x) \to 0$. On the other hand

$$\int f_n \, d\lambda = n\lambda \left((0, \tfrac{1}{n}] \right) = 1,$$

so that

$$\lim_n \int f_n \, d\lambda = 1 > 0 = \int \lim_n f_n \, d\lambda.$$

We will see later that the while the equality of the limit of the integrals and the integral of the limit will not hold in general, they will always be ordered in a specific way, if the functions $\{f_n\}_{n\in\mathbb{N}}$ are non-negative (that will be the content of Fatou's lemma below).

Proposition 3.10 (Approximation by simple functions)**.** *For each $f \in \mathcal{L}^0_+([0,\infty])$ there exists a sequence $\{g_n\}_{n\in\mathbb{N}} \in \mathcal{L}^{\text{Simp},0}_+$ such that*

1. *$g_n(x) \le g_{n+1}(x)$, for all $n \in \mathbb{N}$ and all $x \in S$,*
2. *$g_n(x) \le f(x)$ for all $x \in S$,*
3. *$f(x) = \lim_n g_n(x)$, for all $x \in S$, and*
4. *the convergence $g_n \to f$ is uniform on each set of the form $\{f \le M\}$, $M > 0$, and, in particular, on the whole S if f is bounded.*

Proof. For $n \in \mathbb{N}$, let A^n_k, $k = 1, \dots, n2^n$ be a collection of subsets of S given by

$$A^n_k = \{\tfrac{k-1}{2^n} \le f < \tfrac{k}{2^n}\} = f^{-1}\Big([\tfrac{k-1}{2^n}, \tfrac{k}{2^n})\Big),\ k = 1, \dots, n2^n.$$

Note that the sets A^n_k, $k = 1, \dots, n2^n$ are disjoint and that the measurability of f implies that $A^n_k \in \mathcal{S}$ for $k = 1, \dots, n2^n$. Define the function $g_n \in \mathcal{L}^{\text{Simp},0}_+$ by

$$g_n = \sum_{k=1}^{n2^n} \tfrac{k-1}{2^n} \mathbf{1}_{A^n_k} + n\mathbf{1}_{\{f \ge n\}}.$$

The statements 2., 3., and 4. follow immediately from the following three simple observations:

- $g_n(x) \le f(x)$ for all $x \in S$,
- $g_n(x) = n$ if $f(x) = \infty$, and
- $g_n(x) > f(x) - 2^{-n}$ when $f(x) < n$.

Finally, we leave it to the reader to check the simple fact that $\{g_n\}_{n\in\mathbb{N}}$ is non-decreasing. □

Problem 3.6. Show, by means of an example, that the sequence $\{g_n\}_{n\in\mathbb{N}}$ would not necessarily be monotone if we defined it in the following way:

$$g_n = \sum_{k=1}^{n^2} \tfrac{k-1}{n} \mathbf{1}_{\{f \in [\frac{k-1}{n}, \frac{k}{n})\}} + n\mathbf{1}_{\{f \ge n\}}.$$

Proposition 3.11 (Linearity of the integral for non-negative functions)**.** *For $f_1, f_2 \in \mathcal{L}^0_+([0,\infty])$ and $\alpha_1, \alpha_2 \geq 0$ we have*

$$\int (\alpha_1 f_1 + \alpha_2 f_2)\, d\mu = \alpha_1 \int f_1\, d\mu + \alpha_2 \int f_2\, d\mu.$$

Proof. Thanks to Problem 3.5 it is enough to prove the statement for $\alpha_1 = \alpha_2 = 1$. Let $\{g^1_n\}_{n\in\mathbb{N}}$ and $\{g^2_n\}_{n\in\mathbb{N}}$ be sequences in $\mathcal{L}^{\mathrm{Simp},0}_+$ which approximate f^1 and f^2 in the sense of Proposition 3.10. The sequence $\{g_n\}_{n\in\mathbb{N}}$ given by $g_n = g^1_n + g^2_n$, $n \in \mathbb{N}$, has the following properties:

- $g_n \in \mathcal{L}^{\mathrm{Simp},0}_+$ for $n \in \mathbb{N}$,
- $g_n(x)$ is a nondecreasing sequence for each $x \in S$,
- $g_n(x) \to f_1(x) + f_2(x)$, for all $x \in \mathcal{S}$.

Therefore, we can apply the linearity of integration for the simple functions and the monotone convergence theorem (Theorem 3.8) to conclude that

$$\begin{aligned}\int (f_1 + f_2)\, d\mu &= \lim_n \int (g^1_n + g^2_n)\, d\mu = \lim_n \left(\int g^1_n\, d\mu + \int g^2_n\, d\mu \right) \\ &= \int f_1\, d\mu + \int f_2\, d\mu.\end{aligned}$$

□

Corollary 3.12 (Countable additivity of the integral)**.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0_+([0,\infty])$. Then*

$$\int \sum_{n\in\mathbb{N}} f_n\, d\mu = \sum_{n\in\mathbb{N}} \int f_n\, d\mu.$$

Proof. Apply the monotone convergence theorem to the partial sums $g_n = f_1 + \dots + f_n$, and use linearity of integration. □

Once we have established a battery of properties for non-negative functions, an extension to $\mathcal{L}^1$ is not hard. We leave it to the reader to prove all the statements in the following problem:

Problem 3.7. The family $\mathcal{L}^1$ of integrable functions has the following properties:

1. $f \in \mathcal{L}^1$ iff $\int |f|\, d\mu < \infty$,
2. $\mathcal{L}^1$ is a vector space,
3. $\left|\int f\, d\mu\right| \leq \int |f|\, d\mu$, for $f \in \mathcal{L}^1$.
4. $\int |f + g|\, d\mu \leq \int |f|\, d\mu + \int |g|\, d\mu$, for all $f, g \in \mathcal{L}^1$.

We conclude the present section with two results, which, together with the monotone convergence theorem, play the central role in the Lebesgue integration theory.

Theorem 3.13 (Fatou's lemma). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0_+([0,\infty])$. Then*

$$\int \liminf_n f_n \, d\mu \leq \liminf_n \int f_n \, d\mu.$$

Note: The strength of Fatou's lemma comes from the fact that, apart from non-negativity, it requires no special properties for the sequence $\{f_n\}_{n\in\mathbb{N}}$. Its conclusion is not as strong as that of the monotone convergence theorem, but it proves to be very useful in various settings because it gives an upper bound (namely $\liminf_n \int f_n \, d\mu$) on the integral of the non-negative function $\liminf f_n$.

Proof. Set $g_n(x) = \inf_{k\geq n} f_k(x)$, so that $g_n \in \mathcal{L}^0_+([0,\infty])$ and $g_n(x)$ is a non-decreasing sequence for each $x \in S$. The monotone convergence theorem and the fact that $\liminf f_n(x) = \sup_n g_n(x) = \lim_n g_n(x)$, for all $x \in S$, imply that

$$\int g_n \, d\mu \nearrow \int \liminf_n \, d\mu.$$

On the other hand, $g_n(x) \leq f_k(x)$ for all $k \geq n$, and so

$$\int g_n \, d\mu \leq \inf_{k\geq n} \int f_k \, d\mu.$$

Therefore,

$$\lim_n \int g_n \, d\mu \leq \lim_n \inf_{k\geq n} \int f_k \, d\mu = \liminf_n \int f_k \, d\mu.$$

□

Remark 3.14.

1. The inequality in the Fatou's lemma does not have to be equality, even if the limit $\lim_n f_n(x)$ exists for all $x \in S$. You can use the sequence $\{f_n\}_{n\in\mathbb{N}}$ of Remark 3.9 to see that.
2. Like the monotone convergence theorem, Fatou's lemma requires that all function $\{f_n\}_{n\in\mathbb{N}}$ be non-negative. This requirement is necessary - to see that, simply consider the sequence $\{-f_n\}_{n\in\mathbb{N}}$, where $\{f_n\}_{n\in\mathbb{N}}$ is the sequence of Remark 3.9 above.

Theorem 3.15 (Dominated convergence theorem). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0$ with the property that there exists $g \in \mathcal{L}^1$ such that $|f_n(x)| \leq g(x)$, for all $x \in X$ and all $n \in \mathbb{N}$. If $f(x) = \lim_n f_n(x)$ for all $x \in S$, then $f \in \mathcal{L}^1$ and*

$$\int f \, d\mu = \lim_n \int f_n \, d\mu.$$

Note: The dominated convergence theorem combines the lack of monotonicity requirements of Fatou's lemma and the strong conclusion of the monotone convergence theorem. The price to be paid is the uniform boundedness requirement. There is a way to relax this requirement a little bit (using the concept of *uniform integrability*), but not too much. Still, it is an unexpectedly useful theorem.

Proof. The condition $|f_n(x)| \leq g(x)$, for all $x \in X$ and all $n \in \mathbb{N}$ implies that $g(x) \geq 0$, for all $x \in S$. Since $f_n^+ \leq g$, $f_n^- \leq g$ and $g \in \mathcal{L}^1$, we immediately have $f_n \in \mathcal{L}^1$, for all $n \in \mathbb{N}$. The limiting function

f inherits the same properties $f^+ \leq g$ and $f^- \leq g$ from $\{f_n\}_{n\in\mathbb{N}}$ so $f \in \mathcal{L}^1$, too.

Clearly $g(x) + f_n(x) \geq 0$ for all $n \in \mathbb{N}$ and all $x \in S$, so we can apply Fatou's lemma to get

$$\int g\, d\mu + \liminf_n \int f_n\, d\mu = \liminf_n \int (g + f_n)\, d\mu \geq \int \liminf_n (g + f_n)\, d\mu$$
$$= \int (g + f)\, d\mu = \int g\, d\mu + \int f\, d\mu.$$

In the same way (since $g(x) - f_n(x) \geq 0$, for all $x \in S$, as well), we have

$$\int g\, d\mu - \limsup_n \int f_n\, d\mu = \liminf_n \int (g - f_n)\, d\mu \geq \int \liminf_n (g - f_n)\, d\mu$$
$$= \int (g - f)\, d\mu = \int g\, d\mu - \int f\, d\mu.$$

Therefore

$$\limsup_n \int f_n\, d\mu \leq \int f\, d\mu \leq \liminf_n \int f_n\, d\mu,$$

and, consequently, $\int f\, d\mu = \lim_n \int f_n\, d\mu$. □

Remark 3.16.

Null sets

An important property - inherited directly from the underlying measure - is that it is blind to sets of measure zero. To make this statement precise, we need to introduce some language:

Definition 3.17. Let $(S, \mathcal{S}, \mu)$ be a measure space.

1. $N \in \mathcal{S}$ is said to be a **null set** if $\mu(N) = 0$.
2. A function $f : S \to \bar{\mathbb{R}}$ is called a **null function** if there exists a null set N such that $f(x) = 0$ for $x \in N^c$.
3. Two functions f, g are said to be **equal almost everywhere** - denoted by $f = g$, a.e. - if $f - g$ is a null function, i.e., if there exists a null set N such that $f(x) = g(x)$ for all $x \in N^c$.

Remark 3.18.

1. In addition to almost-everywhere equality, one can talk about the almost-everywhere version of any relation between functions which can be defined on points. For example, we write $f \leq g$, a.e. if $f(x) \leq g(x)$ for all $x \in S$, except, maybe, for x in some null set N.

2. One can also define the a.e. equality of sets: we say that $A = B$, a.e., for $A, B \in \mathcal{S}$ if $\mathbf{1}_A = \mathbf{1}_B$, a.e. It is not hard to show (do it!) that $A = B$ a.e., if and only if $\mu(A \triangle B) = 0$ (Remember that $\triangle$ denotes the symmetric difference: $A \triangle B = (A \setminus B) \cup (B \setminus A)$).

3. When a property (equality of functions, e.g.) holds almost everywhere, the set where it fails to hold is not necessarily null. Indeed, there is no guarantee that it is measurable at all. What is true is that it is *contained* in a measurable (and null) set. Any such (measurable) null set is often referred to as the **exceptional set**.

Problem 3.8. Prove the following statements:

1. The almost-everywhere equality is an equivalence relation between functions.

2. The family $\{A \in \mathcal{S} : \mu(A) = 0 \text{ or } \mu(A^c) = 0\}$ is a σ-algebra (the so-called **μ-trivial σ-algebra**).

The "blindness" property of the Lebesgue integral we referred to above can now be stated formally:

Proposition 3.19. *Suppose that $f = g$, a.e,. for some $f, g \in \mathcal{L}^0_+$. Then*

$$\int f \, d\mu = \int g \, d\mu.$$

Proof. Let N be an exceptional set for $f = g$, a.e., i.e., $f = g$ on N^c and $\mu(N) = 0$. Then $f\mathbf{1}_{N^c} = g\mathbf{1}_{N^c}$, and so $\int f\mathbf{1}_{N^c} \, d\mu = \int g\mathbf{1}_{N^c} \, d\mu$. On the other hand $f\mathbf{1}_N \leq \infty\mathbf{1}_N$ and $\int \infty\mathbf{1}_N \, d\mu = 0$, so, by monotonicity, $\int f\mathbf{1}_N \, d\mu = 0$. Similarly $\int g\mathbf{1}_N \, d\mu = 0$. It remains to use the additivity of integration to conclude that

$$\begin{aligned} \int f \, d\mu &= \int f\mathbf{1}_{N^c} \, d\mu + \int f\mathbf{1}_N \, d\mu \\ &= \int g\mathbf{1}_{N^c} \, d\mu + \int g\mathbf{1}_N \, d\mu = \int g \, d\mu. \end{aligned}$$ □

A statement which can be seen as a converse of Proposition 3.19 also holds:

Problem 3.9. If $f \in \mathcal{L}^0_+$ and $\int f \, d\mu = 0$, show that $f = 0$, a.e.

Hint: What is the negation of the statement "$f = 0$. a.e." for $f \in \mathcal{L}^0_+$?

The monotone convergence theorem and the dominated convergence theorem both require the sequence $\{f_n\}_{n \in \mathbb{N}}$ functions to converge for each $x \in S$. A slightly weaker notion of convergence is required, though:

Definition 3.20. A sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ is said to **converge almost everywhere** to the function f, if there exists a null set N such that

$$f_n(x) \to f(x) \text{ for all } x \in N^c.$$

Remark 3.21. If we want to emphasize that $f_n(x) \to f(x)$ for *all* $x \in S$, we say that $\{f_n\}_{n\in\mathbb{N}}$ converges to f **everywhere**.

Proposition 3.22 (Monotone (almost-everywhere) convergence theorem)**.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0_+([0,\infty])$ with the property that*

$$f_n \leq f_{n+1} \text{ a.e., for all } n \in \mathbb{N}.$$

Then

$$\lim_n \int f_n \, d\mu = \int f \, d\mu,$$

if $f \in \mathcal{L}^0_+$ and $f_n \to f$, a.e.

Proof. There are "$\infty + 1$ a.e.-statements" we need to deal with: one for each $n \in \mathbb{N}$ in $f_n \leq f_{n+1}$, a.e., and an extra one when we assume that $f_n \to f$, a.e. Each of them comes with an exceptional set; more precisely, let $\{A_n\}_{n\in\mathbb{N}}$ be such that $f_n(x) \leq f_{n+1}(x)$ for $x \in A_n^c$ and let B be such that $f_n(x) \to f(x)$ for $x \in B^c$. Define $A \in \mathcal{S}$ by $A = (\cup_n A_n) \cup B$ and note that A is a null set. Moreover, consider the functions $\tilde{f}$, $\{\tilde{f}_n\}_{n\in\mathbb{N}}$ defined by $\tilde{f} = f\mathbf{1}_{A^c}$, $\tilde{f}_n = f_n\mathbf{1}_{A^c}$. Thanks to the definition of the set A, $\tilde{f}_n(x) \leq \tilde{f}_{n+1}(x)$, for all $n \in \mathbb{N}$ and $x \in S$; hence $\tilde{f}_n \to \tilde{f}$, everywhere. Therefore, the monotone convergence theorem (Theorem 3.8) can be used to conclude that $\int \tilde{f}_n \, d\mu \to \int \tilde{f} \, d\mu$. Finally, Proposition 3.19 implies that $\int \tilde{f}_n \, d\mu = \int f_n \, d\mu$ for $n \in \mathbb{N}$ and $\int \tilde{f} \, d\mu = \int f \, d\mu$. $\square$

Problem 3.10. State and prove a version of the dominated convergence theorem where the almost-everywhere convergence is used. Is it necessary for all $\{f_n\}_{n\in\mathbb{N}}$ to be dominated by g for all $x \in S$, or only almost everywhere?

Remark 3.23. There is a subtlety that needs to be pointed out. If a sequence $\{f_n\}_{n\in\mathbb{N}}$ of measurable functions converges to the function f *everywhere*, then f is necessarily a measurable function (see Proposition 1.23). However, if $f_n \to f$ only almost everywhere, there is no guarantee that f is measurable. There is, however, always a measurable function which is equal to f almost everywhere; you can take $\liminf_n f_n$, for example.

Additional Problems

Problem 3.11 (The monotone-class theorem)**.** Prove the following result, known as the *monotone-class theorem* (remember that $a_n \nearrow a$ means that a_n is a non-decreasing sequence and $a_n \to a$)

Hint: Use Theorems 3.10 and 2.12

Let $\mathcal{H}$ be a class of bounded functions from S into $\mathbb{R}$ satisfying the following conditions

1. *$\mathcal{H}$ is a vector space,*
2. *the constant function 1 is in $\mathcal{H}$, and*
3. *if $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of non-negative functions in $\mathcal{H}$ such that $f_n(x) \nearrow f(x)$, for all $x \in S$ and f is bounded, then $f \in \mathcal{H}$.*

Then, if $\mathcal{H}$ contains the indicator $\mathbf{1}_A$ of every set A in some π-system $\mathcal{P}$, then $\mathcal{H}$ necessarily contains every bounded $\sigma(\mathcal{P})$-measurable function on $\mathcal{S}$.

Problem 3.12 (A form of continuity for Lebesgue integration)**.** Let $(S, \mathcal{S}, \mu)$ be a measure space, and suppose that $f \in \mathcal{L}^1$. Show that for each $\varepsilon > 0$ there exists $\delta > 0$ such that if $A \in \mathcal{S}$ and $\mu(A) < \delta$, then $\left|\int_A f\, d\mu\right| < \varepsilon$.

Problem 3.13 (Sums as integrals)**.** In the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$, let μ be the counting measure.

1. For a function $f : \mathbb{N} \to [0, \infty]$, show that
$$\int f\, d\mu = \sum_{n=1}^{\infty} f(n).$$
2. Use the monotone convergence theorem to show the following special case of Fubini's theorem
$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{kn},$$
whenever $\{a_{kn} : k, n \in \mathbb{N}\}$ is a double sequence in $[0, \infty]$.
3. Show that $f : \mathbb{N} \to \mathbb{R}$ is in $\mathcal{L}^1$ if and only if the series
$$\sum_{n=1}^{\infty} f(n),$$
converges absolutely.

Problem 3.14 (A criterion for integrability)**.** Let $(S, \mathcal{S}, \mu)$ be a finite measure space. For $f \in \mathcal{L}^0_+$, show that $f \in \mathcal{L}^1$ if and only if

$$\sum_{n\in\mathbb{N}} \mu(\{f \geq n\}) < \infty.$$

Problem 3.15 (A limit of integrals)**.** Let $(S, \mathcal{S}, \mu)$ be a measure space, and suppose $f \in \mathcal{L}^1_+$ is such that $\int f\, d\mu = c > 0$. Show that the limit

$$\lim_n \int n \log\Big(1 + (f/n)^\alpha\Big)\, d\mu$$

exists in $[0, \infty]$ for each $\alpha > 0$ and compute its value.

Hint: Prove and use the inequality $\log(1 + x^\alpha) \leq \alpha x$, valid for $x \geq 0$ and $\alpha \geq 1$.

Problem 3.16 (Integrals converge but the functions don't ...)**.** Construct a sequence $\{f_n\}_{n\in\mathbb{N}}$ of continuous functions $f_n : [0,1] \to [0,1]$ such that $\int f_n\, d\mu \to 0$, but the sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ is divergent for each $x \in [0,1]$.

Problem 3.17 (... or they do, but are not dominated)**.** Construct an sequence $\{f_n\}_{n\in\mathbb{N}}$ of continuous functions $f_n : [0,1] \to [0,\infty)$ such that $\int f_n\, d\mu \to 0$, and $f_n(x) \to 0$ for all x, but $f \notin \mathcal{L}^1$, where $f(x) = \sup_n f_n(x)$.

Problem 3.18 (Functions measurable in the generated σ-algebra)**.** Let $S \neq \emptyset$ be a set and let $f : S \to \mathbb{R}$ be a function. Prove that a function $g : S \to \mathbb{R}$ is measurable with respect to the pair $(\sigma(f), \mathcal{B}(\mathbb{R}))$ if and only if there exists a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $g = h \circ f$.

Problem 3.19 (A change-of-variables formula)**.** Let $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \nu)$ be two measurable spaces, and let $F : S \to T$ be a measurable function with the property that $\nu = F_*\mu$ (i.e., ν is the push-forward of μ through F). Show that for every $f \in \mathcal{L}^0_+(T, \mathcal{T})$ or $f \in \mathcal{L}^1(T, \mathcal{T})$, we have

$$\int f\, d\nu = \int (f \circ F)\, d\mu.$$

Problem 3.20 (The Riemann Integral)**.** A finite collection $\Delta = \{t_0, \dots, t_n\}$, where $a = t_0 < t_1 < \dots < t_n = b$ and $n \in \mathbb{N}$, is called a **partition** of the interval $[a, b]$. The set of all partitions of $[a, b]$ is denoted by $P([a, b])$.

For a bounded function $f : [a, b] \to \mathbb{R}$ and $\Delta = \{t_0, \dots, t_n\} \in P([a, b])$, we define its **upper and lower Darboux sums** $U(f, \Delta)$ and $L(f, \Delta)$ by

$$U(f, \Delta) = \sum_{k=1}^{n} \left(\sup_{t\in(t_{k-1},t_k]} f(t) \right) (t_k - t_{k-1})$$

and
$$L(f,\Delta) = \sum_{k=1}^{n} \left(\inf_{t\in(t_{k-1},t_k]} f(t) \right) (t_k - t_{k-1}).$$

A function $f : [a,b] \to \mathbb{R}$ is said to be **Riemann integrable** if it is bounded and
$$\sup_{\Delta\in P([a,b])} L(f,\Delta) = \inf_{\Delta\in P([a,b])} U(f,\Delta).$$

In that case the common value of the supremum and the infimum above is called the **Riemann integral** of the function f - denoted by $(R)\int_a^b f(x)\,dx$.

1. Suppose that a bounded Borel-measurable function $f : [a,b] \to \mathbb{R}$ is Riemann-integrable. Show that
$$\int_{[a,b]} f\,d\lambda = (R)\int_a^b f(x)\,dx.$$

2. Find an example of a bounded an Borel-measurable function $f : [a,b] \to \mathbb{R}$ which is not Riemann-integrable.

3. Show that every continuous function is Riemann integrable.

4. It can be shown that for a bounded Borel-measurable function $f : [a,b] \to \mathbb{R}$ the following criterion holds (and you can use it without proof):

   f is Riemann integrable if and only if there exists a Borel set $D \subseteq [a,b]$ with $\lambda(D) = 0$ such that f is continuous at x, for each $x \in [a,b] \setminus D$.
   Show that

   - all monotone functions are Riemann-integrable,
   - $f \circ g$ is Riemann integrable if $f : [c,d] \to \mathbb{R}$ is Riemann integrable and $g : [a,b] \to [c,d]$ is continuous,
   - products of Riemann-integrable functions are Riemann-integrable.

5. Let $([a,b], \mathcal{B}([a,b])^*, \lambda^*)$ denote the completion of $([a,b], \mathcal{B}([a,b]), \lambda)$. Show that any Riemann-integrable function on $[a,b]$ is $\mathcal{B}([a,b])^*$-measurable.

   Hint: Pick a sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P([a,b])$ so that $\Delta_n \subseteq \Delta_{n+1}$ and $U(f,\Delta_n) - L(f,\Delta_n) \to 0$. Using those partitions and the function f, define two sequences of Borel-measurable functions $\{\overline{f}_n\}_{n\in\mathbb{N}}$ and $\{\underline{f}_n\}_{n\in\mathbb{N}}$ so that $\underline{f}_n \nearrow \underline{f}$, $\overline{f}_n \searrow \overline{f}$, $\underline{f} \leq f \leq \overline{f}$, and $\int(\overline{f} - \underline{f})\,d\lambda = 0$. Conclude that f agrees with a Borel measurable function on a complement of a subset of the set $\{\underline{f} \neq \overline{f}\}$ which has Lebesgue measure 0.

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 4

Lebesgue spaces and inequalities

Lebesgue spaces

We have seen how the family of all functions $f \in \mathcal{L}^1$ forms a vector space and how the map $f \mapsto ||f||_{\mathcal{L}^1}$, from $\mathcal{L}^1$ to $[0, \infty)$ defined by $||f||_{\mathcal{L}^1} = \int |f| \, d\mu$ has the following properties

1. $f = 0$ implies $||f||_{\mathcal{L}^1} = 0$, for $f \in \mathcal{L}^1$,
2. $||f + g||_{\mathcal{L}^1} \leq ||f||_{\mathcal{L}^1} + ||g||_{\mathcal{L}^1}$, for $f, g \in \mathcal{L}^1$,
3. $||\alpha f||_{\mathcal{L}^1} = |\alpha| \, ||f||_{\mathcal{L}^1}$, for $\alpha \in \mathbb{R}$ and $f \in \mathcal{L}^1$.

Any map from a vector space into $[0, \infty)$ with the properties 1., 2., and 3. above is called a **pseudo norm**. A pair $(V, || \cdot ||)$ where V is a vector space and $|| \cdot ||$ is a pseudo norm on V is called a **pseudo-normed space**.

If a pseudo norm happens to satisfy the (stronger) axiom

1.' $f = 0$ if and only if $||f|| = 0$,

instead of 1., it is called a **norm**, and the pair $(V, || \cdot ||)$ is called a **normed space**.

The pseudo-norm $|| \cdot ||_{\mathcal{L}^1}$ is, in general, not a norm. Indeed, by Problem 3.9, we have $||f||_{\mathcal{L}^1} = 0$ iff $f = 0$, a.e., and unless $\emptyset$ is the only null-set, there are functions different from the constant function 0 with this property.

Remark 4.1. There is a relatively simple procedure one can use to turn a pseudo-normed space $(V, || \cdot ||)$ into a normed one. Declare two elements x, y in V *equivalent* (denoted by $x \sim y$) if $||y - x|| = 0$, and let $\tilde{V}$ be the quotient space $V / \sim$ (the set of all equivalence classes). It is easy to show that $||x|| = ||y||$ whenever $x \sim y$, so the pseudo-norm $|| \cdot ||$ can be seen as defined on $\tilde{V}$. Moreover, it follows directly from the properties of the pseudo norm that $(\tilde{V}, || \cdot ||)$ is, in fact a normed space. Idea is, of course, bundle together the elements of V which differ by such a "small amount" that $|| \cdot ||$ cannot detect it.

This construction can be applied to the case of the pseudo-norm $||\cdot||_{\mathcal{L}^1}$ on $\mathcal{L}^1$, and the resulting normed space is denoted by $\mathbb{L}^1$. The normed space $\mathbb{L}^1$ has properties similar to those of $\mathcal{L}^1$, but its elements are not functions anymore - they are equivalence classes of measurable functions[1].

[1] Such a point of view is very useful in analysis, but it sometimes leads to confusion in probability (especially when one works with stochastic processes with infinite time-index sets). Therefore, we will stick to $\mathcal{L}^1$ and deal with the fact that it is only a pseudo-normed space.

A pseudo-norm $||\cdot||$ on a vector space can be used to define a **pseudo metric** (pseudo-distance function) on V by the following simple prescription:

$$d(x,y) = ||y-x||, \ x,y \in V.$$

Just like a pseudo norm, a pseudo metric has most of the properties of a metric

1. $d(x,y) \in [0,\infty)$, for $x,y \in V$,
2. $d(x,y) + d(y,z) \geq d(x,z)$, for $x,y,z \in V$,
3. $d(x,y) = d(y,x)$, $x,y \in V$,
4. $x = y$ implies $d(x,y) = 0$, for $x,y \in V$.

The missing axiom is the stronger version of 4. given by

4′. $x = y$ if and only if $d(x,y) = 0$, for $x,y \in V$.

Luckily, a pseudo metric is sufficiently structured to yield a topological notion of convergence as follows: we say that a sequence $\{x_n\}_{n\in\mathbb{N}}$ in V converges towards $x \in V$ if $d(x_n,x) \to 0$, as $n \to \infty$. If we apply it to our original example $(\mathcal{L}^1, ||\cdot||_{\mathcal{L}^1})$, we have the following definition:

Definition 4.2 (Convergence in $\mathcal{L}^1$). For a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^1$, we say that $\{f_n\}_{n\in\mathbb{N}}$ converges to f **in** $\mathcal{L}^1$ if

$$||f_n - f||_{\mathcal{L}^1} \to 0.$$

To get some intuition about convergence in $\mathcal{L}^1$, here is a problem:

Problem 4.1. Show that the conclusion of the dominated convergence theorem (Theorem 3.15) can be replaced by "$f_n \to f$ in $\mathcal{L}^1$". Does the original conclusion follow from the new one?

The only problem that arises when one defines convergence using a pseudo metric (as opposed to a bona-fide metric) is that limits are not unique. This is, however, merely an inconvenience and one gets used to it quite readily:

Problem 4.2. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ converges to f in $\mathcal{L}^1$. Show that $\{f_n\}_{n\in\mathbb{N}}$ also converges to $g \in \mathcal{L}^1$ if and only if $f = g$, a.e.

In addition to the space $\mathcal{L}^1$, one can introduce many other vector spaces of a similar flavor. For $p \in [1, \infty)$, let $\mathcal{L}^p$ denote the family of all functions $f \in \mathcal{L}^0$ such that $|f|^p \in \mathcal{L}^1$.

Problem 4.3. Show that there exists a constant $C > 0$ (depending on p, but independent of a, b) such that $(a+b)^p \leq C(a^p + b^p)$, $p \in (0, \infty)$ and for all $a, b \geq 0$. Deduce that $\mathcal{L}^p$ is a vector space for all $p \in (0, \infty)$.

We will see soon that the map $||\cdot||_{\mathcal{L}^p}$, defined by

$$||f||_{\mathcal{L}^p} = \left(\int |f|^p \, d\mu \right)^{1/p}, \ f \in \mathcal{L}^p,$$

is a pseudo norm on $\mathcal{L}^p$. The hard part of the proof - showing that $||f+g||_{\mathcal{L}^p} \leq ||f||_{\mathcal{L}^p} + ||g||_{\mathcal{L}^p}$ will be a direct consequence of an important inequality of Minkowski which will be proved below.

Finally, there is a nice way to extend the definition of $\mathcal{L}^p$ to $p = \infty$.

Definition 4.3 (Essential supremum)**.** A number[2] $a \in \bar{\mathbb{R}}$ is called an **essential supremum** of the function $f \in \mathcal{L}^0$ - and is denoted by $a = \operatorname{esssup} f$ - if

[2] Necessarily unique!

1. $\mu(\{f > a\}) = 0$
2. $\mu(\{f > b\}) > 0$ for any $b < a$.

A function $f \in \mathcal{L}^0$ with $\operatorname{esssup} f < \infty$ is said to be **essentially bounded from above**. When $\operatorname{esssup} |f| < \infty$, we say that f is **essentially bounded**.

Remark 4.4. Even though the function f may take values larger than a, it does so only on a null set. In fact, it can happen that a function is unbounded, but that its essential supremum exists in $\mathbb{R}$. Indeed, take $(S, \mathcal{S}, \mu) = ([0,1], \mathcal{B}([0,1]), \lambda)$, and define

$$f(x) = \begin{cases} n, & x = \frac{1}{n} \text{ for some } n \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\operatorname{esssup} f = 0$, since $\lambda(\{f > 0\}) = \lambda(\{1, 1/2, 1/3, \dots\}) = 0$, but $\sup_{x \in [0,1]} f(x) = \infty$.

Let $\mathcal{L}^\infty$ denote the family of all essentially bounded functions in $\mathcal{L}^0$. Define $||f||_{\mathcal{L}^\infty} = \operatorname{esssup} |f|$, for $f \in \mathcal{L}^\infty$.

Problem 4.4. Show that $\mathcal{L}^\infty$ is a vector space, and that $||\cdot||_{\mathcal{L}^\infty}$ is a pseudo-norm on $\mathcal{L}^\infty$.

The convergence in $\mathcal{L}^p$ for $p > 1$ is defined similarly to the $\mathcal{L}^1$-convergence:

Definition 4.5 (Convergence in $\mathcal{L}^p$)**.** Let $p \in [1, \infty]$. We say that a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^p$ **converges in** $\mathcal{L}^p$ to $f \in \mathcal{L}^p$ if

$$||f_n - f||_{\mathcal{L}^p} \to 0, \text{ as } n \to \infty.$$

Problem 4.5. Show that $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{L}^\infty$ converges to $f \in \mathcal{L}^\infty$ in $\mathcal{L}^\infty$ if and only if there exist functions $\{\tilde{f}_n\}_{n\in\mathbb{N}}$, $\tilde{f}$ in $\mathcal{L}^0$ such that

1. $\tilde{f}_n = f_n$, a.e., and $\tilde{f} = f$, a.e, and
2. $\tilde{f}_n \to \tilde{f}$ uniformly[3].

[3] we say that $g_n \to g$ uniformly if $\sup_x |g_n(x) - g(x)| \to 0$, as $n \to \infty$.

Inequalities

Definition 4.6 (Conjugate exponents)**.** We say that $p, q \in [1, \infty]$ are **conjugate exponents** if $\frac{1}{p} + \frac{1}{q} = 1$.

Lemma 4.7 (Young's inequality)**.** *For all $x, y \geq 0$ and conjugate exponents $p, q \in [1, \infty)$ we have*

$$\frac{x^p}{p} + \frac{y^q}{q} \geq xy. \tag{4.1}$$

The equality holds if and only if $x^p = y^q$.

Proof. If $x = 0$ or $y = 0$, the inequality trivially holds so we assume that $x > 0$ and $y > 0$. The function log is strictly concave on $(0, \infty)$ and $\frac{1}{p} + \frac{1}{q} = 1$, so

$$\log(\tfrac{1}{p}\xi + \tfrac{1}{q}\eta) \geq \tfrac{1}{p}\log(\xi) + \tfrac{1}{q}\log(\eta),$$

for all $\xi, \eta > 0$, with equality if and only if $\xi = \eta$. If we substitute $\xi = x^p$ and $\eta = y^q$, and exponentiate both sides, we get

$$\frac{x^p}{p} + \frac{y^q}{q} \geq \exp(\tfrac{1}{p}\log(x^p) + \tfrac{1}{q}\log(y^q)) = xy,$$

with equality if and only if $x^p = y^q$. $\square$

Remark 4.8. If you do not want to be fancy, you can prove Young's inequality by locating the maximum of the function $x \mapsto xy - \frac{1}{p}x^p$ using nothing more than elementary calculus.

Proposition 4.9 (Hölder's inequality)**.** *Let $p, q \in [1, \infty]$ be conjugate exponents. For $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^q$, we have*

$$\int |fg| \, d\mu \leq ||f||_{\mathcal{L}^p} ||g||_{\mathcal{L}^q}. \tag{4.2}$$

The equality holds if and only if there exist constants $\alpha, \beta \geq 0$ with $\alpha + \beta > 0$ such that $\alpha |f|^p = \beta |g|^q$, a.e.

Proof. We assume that $1 < p, g < \infty$ and leave the (easier) extreme cases to the reader. Clearly, we can also assume that $||f||_{\mathcal{L}^p} > 0$ and $||q||_{\mathcal{L}^q} > 0$ - otherwise, the inequality is trivially satisfied. We define $\tilde{f} = |f| / ||f||_{\mathcal{L}^p}$ and $\tilde{g} = |g| / ||g||_{\mathcal{L}^q}$, so that $||\tilde{f}||_{\mathcal{L}^p} = ||\tilde{g}||_{\mathcal{L}^q} = 1$.

Plugging $\tilde{f}$ for x and $\tilde{g}$ for y in Young's inequality (Lemma 4.7 above) and integrating, we get

$$\tfrac{1}{p} \int \tilde{f}^p \, d\mu + \tfrac{1}{q} \int \tilde{g}^q \, d\mu \geq \int \tilde{f}\tilde{g} \, d\mu, \tag{4.3}$$

and consequently,

$$\int \tilde{f}\tilde{g} \, d\mu \leq 1, \tag{4.4}$$

because $\int \tilde{f}^p \, d\mu = ||\tilde{f}||^p_{\mathcal{L}^p} = 1$, and $\int \tilde{g}^q \, d\mu = ||\tilde{g}||_{\mathcal{L}^p} = 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Hölder's inequality (4.2) now follows by multiplying both sides of (4.4) by $||f||_{\mathcal{L}^p} ||g||_{\mathcal{L}^q}$.

If the equality in (4.2) holds, then it also holds a.e. in the Young's inequality (4.3). Therefore, the equality will hold if and only if $||g||^q_{\mathcal{L}^q} |f|^p = ||f||^p_{\mathcal{L}^p} |g|^q$, a.e. The reader will check that if a pair of constants α, β as in the statement exists, then $(||g||^q_{\mathcal{L}^q}, ||f||^p_{\mathcal{L}^p})$ must be proportional to it. □

For $p = q = 2$ we get the following well-known special case:

Corollary 4.10 (Cauchy-Schwarz inequality)**.** *For $f, g \in \mathcal{L}^2$, we have*

$$\int |fg| \, d\mu \leq ||f||_{\mathcal{L}^2} ||g||_{\mathcal{L}^2}.$$

Corollary 4.11 (Minkowski's inequality)**.** *For $f, g \in \mathcal{L}^p$, $p \in [1, \infty]$, we have*

$$||f + g||_{\mathcal{L}^p} \leq ||f||_{\mathcal{L}^p} + ||g||_{\mathcal{L}^p}. \tag{4.5}$$

Proof. Like above, we assume $p < \infty$ and leave the case $p = \infty$ to the reader. Moreover, we assume that $||f + g||_{\mathcal{L}^p} > 0$ - otherwise, the inequality trivially holds. Note, first that for conjugate exponents p, q we have $q(p - 1) = p$. Therefore, Hölder's inequality implies that

$$\begin{aligned} \int |f| \, |f+g|^{p-1} \, d\mu &\leq ||f||_{\mathcal{L}^p} ||(f+g)^{p-1}||_{\mathcal{L}^q} \\ &= ||f||_{\mathcal{L}^p} \left(\int |f+g|^{q(p-1)} \, d\mu \right)^{1/q} \\ &= ||f||_{\mathcal{L}^p} ||f+g||^{p/q}_{\mathcal{L}^p}, \end{aligned}$$

and, similarly,

$$\int |g| \, |f+g|^{p-1} \, d\mu \leq ||g||_{\mathcal{L}^p} ||f+g||^{p/q}_{\mathcal{L}^p}.$$

Therefore,

$$||f+g||^p_{\mathcal{L}^p} = \int |f+g|^p \, d\mu \leq \int |f|\,|f+g|^{p-1} \, d\mu + \int |g|\,|f+g|^{p-1} \, d\mu$$
$$\leq \Big(||f||_{\mathcal{L}^p} + ||g||_{\mathcal{L}^p}\Big) ||f+g||^{p-1}_{\mathcal{L}^p},$$

and if we divide through by $||f+g||^{p-1}_{\mathcal{L}^p} > 0$, we get (4.5). □

Corollary 4.12 ($\mathcal{L}^p$ is pseudo-normed)**.** *($\mathcal{L}^p, ||\cdot||_{\mathcal{L}^p}$) is a pseudo-normed space for each $p \in [1,\infty]$.*

A pseudo-metric space (X,d) is said to be **complete** if each Cauchy sequence converges. A sequence $\{x_n\}_{n\in\mathbb{N}}$ is called a Cauchy sequence if

$$\forall\, \varepsilon > 0, \; \exists N \in \mathbb{N}, \; m,n \geq N \; \Rightarrow \; d(x_n, x_m) < \varepsilon.$$

A pseudo-normed space $(V, ||\cdot||)$ is called a **pseudo-Banach space** if is it complete for the metric induced by $||\cdot||$. If $||\cdot||$ is, additionally, a norm, $(V, ||\cdot||)$ is said to be a **Banach space**.

Problem 4.6. Let $\{x_n\}_{n\in\mathbb{N}}$ be Cauchy sequence in a pseudo-metric space (X,d), and let $\{x_{n_k}\}_{k\in\mathbb{N}}$ be a subsequence of $\{x_n\}_{n\in\mathbb{N}}$ which converges to $x \in X$. Show that $x_n \to x$.

Proposition 4.13 ($\mathcal{L}^p$ is pseudo-Banach)**.** *$\mathcal{L}^p$ is a pseudo-Banach space, for $p \in [1,\infty]$.*

Proof. We assume $p \in [1,\infty)$ and leave the case $p = \infty$ to the reader. Let $\{f_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{L}^p$. Thanks to the Cauchy property, there exists a subsequence of $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that

$$||f_{n_{k+1}} - f_{n_k}||_{\mathcal{L}^p} < 2^{-k}, \text{ for all } k \in \mathbb{N}.$$

We define the sequence $\{g_k\}_{k\in\mathbb{N}}$ in $\mathcal{L}^0_+$ by $g_k = |f_{n_1}| + \sum_{i=1}^{k-1} |f_{n_{i+1}} - f_{n_i}|$, as well as the function $g = \lim_k g_k \in \mathcal{L}^0([0,\infty])$. The monotone-convergence theorem implies that

$$\int g^p \, d\mu = \lim_n \int g_n^p \, d\mu,$$

and, by Minkowski's inequality, we have

$$\int g_k^p \, d\mu = ||g_k||^p_{\mathcal{L}^p} \leq \Big(||f_{n_1}||_{\mathcal{L}^p} + \sum_{i=1}^{k-1} ||f_{n_{k+1}} - f_{n_k}||_{\mathcal{L}^p} \Big)^p$$
$$\leq (||f_{n_1}||_{\mathcal{L}^p} + 1)^p, \; \forall\, k \in \mathbb{N}.$$

Therefore, $\int g^p \, d\mu \leq (1 + ||f_{n_1}||_{\mathcal{L}^p})^p < \infty$, and, in particular, $g \in \mathcal{L}^p$ and $g < \infty$, a.e. It follows immediately from the absolute convergence of the series $\sum_{i=1}^{\infty} \left| f_{n_{k+1}} - f_{n_k} \right|$ that

$$f_{n_k}(x) = f_{n_1}(x) + \sum_{i=1}^{k-1} (f_{n_{k+1}}(x) - f_{n_k}(x)),$$

converges in $\mathbb{R}$, for almost all x. Hence, the function $f = \liminf_k f_{n_k}$ is in $\mathcal{L}^p$ since $|f| \leq g$, a.e.

Since $|f| \leq g$ and $\left| f_{n_k} \right| \leq g$, for all $k \in \mathbb{N}$, we have $\left| f - f_{n_k} \right|^p \leq 2\,|g|^p \in \mathbb{L}^1$, so the dominated convergence theorem implies that

$$\int \left| f_{n_k} - f \right|^p \, d\mu \to 0, \text{ i.e., that } f_{n_k} \to f \text{ in } \mathcal{L}^p.$$

Finally, by the result of Problem 4.6 we have $f_n \to f$ in $\mathcal{L}^p$. □

The following result is a simple consequence of the (proof of) Proposition 4.13.

Corollary 4.14 ($\mathcal{L}^p$-convergent $\Rightarrow$ a.e.-convergent, after passing to subsequence)**.** *For $p \in [1, \infty]$, let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^p$ such that $f_n \to f$ in $\mathcal{L}^p$. Then there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ of $\{f_n\}_{n\in\mathbb{N}}$ such that $f_{n_k} \to f$, a.e.*

We have seen above how the concavity of the function log was used in the proof of Young's inequality (Lemma 4.7). A generalization of the definition of convexity, called *Jensen's inequality*, is one of the most powerful tools in measure theory. Recall that $\mathcal{L}^{0-1}$ denotes the set of all $f \in \mathcal{L}^0$ with $f^- \in \mathcal{L}^1$.

Proposition 4.15 (Jensen's inequality)**.** *Suppose that $\mu(S) = 1$ (μ is a probability measure) and that $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function. For a function $f \in \mathcal{L}^1$ we have $\varphi(f) \in \mathcal{L}^{0-1}$ and*

$$\int \varphi(f) \, d\mu \geq \varphi\Big(\int f \, d\mu\Big).$$

Before we give a proof, we need a lemma about convex functions:

Lemma 4.16 (Convex functions as suprema of sequences of affine functions)**.** *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. Then, there exists two sequences $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ of real numbers such that*

$$\varphi(x) = \sup_{n\in\mathbb{N}} (a_n x + b_n).$$

Proof (*). For $x \in \mathbb{R}$, we define the left and right derivative $\frac{\partial}{\partial x}\varphi^-$ and $\frac{\partial}{\partial x}\varphi^+$ of φ at x by

$$\frac{\partial}{\partial x}\varphi^-(x) = \sup_{\varepsilon>0} \tfrac{1}{\varepsilon}(\varphi(x) - \varphi(x - \varepsilon)),$$

and
$$\frac{\partial}{\partial x}\varphi^+(x) = \inf_{\varepsilon>0} \tfrac{1}{\varepsilon}(\varphi(x+\varepsilon)-\varphi(x)).$$
Convexity of the function φ implies that the difference quotient
$$\varepsilon \mapsto \tfrac{1}{\varepsilon}(\varphi(x+\varepsilon)-\varphi(x)),\ \varepsilon \neq 0,$$
is a non-decreasing function (why?), and so both $\frac{\partial}{\partial x}\varphi^-$ and $\frac{\partial}{\partial x}\varphi^+$ are, in fact, limits as $\varepsilon > 0$. Moreover, we always have
$$\begin{aligned} \tfrac{1}{\varepsilon'}(\varphi(x)-\varphi(x-\varepsilon')) \leq \frac{\partial}{\partial x}\varphi^-(x) &\leq \frac{\partial}{\partial x}\varphi^+(x) \\ &\leq \tfrac{1}{\varepsilon}(\varphi(x+\varepsilon)-\varphi(x)), \end{aligned} \tag{4.6}$$
for all x and all $\varepsilon, \varepsilon' > 0$.

Let $\{q_n\}_{n\in\mathbb{N}}$ be an enumeration of rational numbers in $\mathbb{R}$. For each $n \in \mathbb{N}$ we pick $a_n \in [\frac{\partial}{\partial x}\varphi^-(q_n), \frac{\partial}{\partial x}\varphi^+(q_n)]$ and set $b_n = \varphi(q_n) - a_n q_n$, so that the line $x \mapsto a_n x + b_n$ passes through $(q_n, \varphi(q_n))$ and has a slope which is between the left and the right derivative of φ at q_n.

Let us first show that $\varphi(x) \geq a_n x + b_n$ for all $x \in \mathbb{R}$ and all $n \in \mathbb{N}$. We pick $x \in \mathbb{R}$ and assume that $x \geq q_n$ (the case $x < q_n$ is analogous). If $x = q_n$ then $\varphi(x) = a_n x + b_n$ by construction, and we are done. When $\varepsilon = x - q_n > 0$, relation (4.6) implies that
$$a_n x + b_n = a_n(x-q_n) + \varphi(q_n) \leq \left(\tfrac{\varphi(x)-\varphi(q_n)}{x-q_n}\right)(x-q_n) + \varphi(q_n) = \varphi(x).$$
Conversely, suppose that $\varphi(x) > \sup_n(a_n x + b_n)$, for some $x \in \mathbb{R}$. Both functions φ and ψ, where $\psi(x) = \sup_n(a_n x + b_n)$ are convex (why?), and, therefore, continuous. It follows from $\varphi(x) > \psi(x)$, that there exists a rational number q_n such that $\varphi(q_n) > \psi(q_n)$. This is a contradiction, though, since $\psi(q_n) \geq a_n q_n + b_n = \varphi(q_n)$. □

Proof of Proposition 4.15. Let us first show that $(\varphi(f))^- \in \mathcal{L}^1$. By Lemma 4.16, there exists sequences $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ such that $\varphi(x) = \sup_{n\in\mathbb{N}}(a_n x + b_n)$. In particular, $\varphi(f(x)) \geq a_1 f(x) + b_1$, for all $x \in S$. Therefore,
$$\big(\varphi(f(x))\big)^- \leq (a_1 f(x) + b_1)^- \leq |a_1|\,|f(x)| + |b_1| \in \mathcal{L}^1.$$
Next, we have $\int \varphi(f)\, d\mu \geq \int a_n f + b_n\, d\mu = a_n \int f\, d\mu + b_n$, for all $n \in \mathbb{N}$. Hence
$$\int \varphi(f)\, d\mu \geq \sup_n (a_n \int f\, d\mu + b_n) = \varphi(\int f\, d\mu). \qquad \square$$

Problem 4.7. State and prove a generalization of Jensen's inequality when φ is defined only on an interval I of $\mathbb{R}$, but $\mu(\{f \notin I\}) = 0$.

Problem 4.8. Use Jensen's inequality on an appropriately chosen measure space to prove the **arithmetic-geometric inequality**

$$\tfrac{a_1+\dots+a_n}{n} \geq \sqrt[n]{a_1 \dots a_n}, \text{ for } a_1, \dots, a_n > 0.$$

Proposition 4.17 (Markov's inequality). *For $f \in \mathcal{L}^0_+$ and $\alpha > 0$ we have*

$$\mu(\{f \geq \alpha\}) \leq \tfrac{1}{\alpha} \int f \, d\mu.$$

Note: The name **Markov's inequality** in used in probability theory, but not much wider than that. In analysis, **Chebyshev's inequality** is much more common.

Proof. Consider the function $g \in \mathcal{L}^0_+$ defined by

$$g(x) = \alpha \mathbf{1}_{\{f \geq \alpha\}} = \begin{cases} \alpha, & f(x) \in [\alpha, \infty) \\ 0, & f(x) \in [0, \alpha). \end{cases}$$

Then $f(x) \geq g(x)$ for all $x \in S$, and so

$$\int f \, d\mu \geq \int g \, d\mu = \alpha \mu(\{f \geq \alpha\}). \qquad \square$$

Additional problems

Problem 4.9 (Projections onto a convex set)**.** A subset K of a vector space is said to be **convex** if $\alpha x + (1-\alpha) y \in K$, whenever $x, y \in K$ and $\alpha \in [0,1]$. Let K be a closed and convex subset of $\mathcal{L}^2$, and let g be an element of its complement $\mathcal{L}^2 \setminus K$. Prove that

Hint: Pick a sequence $\{f_n\}_{n\in\mathbb{N}}$ in K with $||f_n - g||_{\mathcal{L}^2} \to \inf_{f\in K} ||f-g||_{\mathcal{L}^2}$ and show that it is Cauchy. Use (but prove first) the **parallelogram identity** $2||h||^2_{\mathcal{L}^2} + 2||k||^2_{\mathcal{L}^2} = ||h+k||^2_{\mathcal{L}^2} + ||h-k||^2_{\mathcal{L}^2}$, for $h, k \in \mathcal{L}^2$.

1. There exists an element $f^* \in K$ such that $||g - f^*||_{\mathcal{L}^2} \leq ||g - f||_{\mathcal{L}^2}$, for all $f \in K$.
2. $\int (f - f^*)(g - f^*) \, d\mu \leq 0$, for all $f \in K$.

Problem 4.10 (Egorov's theorem)**.** Suppose that μ is a finite measure, and let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0$ which converges a.e. to $f \in \mathcal{L}^0$. Prove that for each $\varepsilon > 0$ there exists $E \in \mathcal{S}$ with $\mu(E^c) < \varepsilon$ such that

Hint: Define $A_{n,k} = \cup_{m\geq n}\{|f_m - f| \geq \frac{1}{k}\}$, show that for each $k \in \mathbb{N}$, there exists $n_k \in \mathbb{N}$ such that $\mu(A_{n_k,k}) < \varepsilon/2^k$, and set $E = \cap_k A^c_{n_k,k}$.

$$\lim_{n\to\infty} \operatorname{esssup} |f_n \mathbf{1}_E - f \mathbf{1}_E| = 0.$$

Problem 4.11 (Relationships between different $\mathcal{L}^p$ spaces)**.**

1. Show that for $p, q \in [1, \infty)$ and $t = 1/p - 1/p$, we have

$$||f||_{\mathcal{L}^p} \leq ||f||_{\mathcal{L}^q} \mu(S)^r.$$

   Conclude that $\mathcal{L}^q \subseteq \mathcal{L}^p$, for $p \leq q$ if $\mu(S) < \infty$.

2. For $p_0 \in [1, \infty)$, construct an example of a measure space $(S, \mathcal{S}, \mu)$ and a function $f \in \mathcal{L}^0$ such that $f \in \mathcal{L}^p$ if and only if $p = p_0$.

3. Suppose that $f \in \mathcal{L}^r \cap \mathcal{L}^\infty$, for some $r \in [1,\infty)$. Show that $f \in \mathcal{L}^p$ for all $p \in [r,\infty)$ and
$$||f||_{\mathcal{L}^\infty} = \lim_{p\to\infty} ||f||_{\mathcal{L}^p}.$$

Problem 4.12 (Convergence in measure). A sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^0$ is said to **converge in measure** toward $f \in \mathcal{L}^0$ if
$$\forall\, \varepsilon > 0,\ \mu\Big(\{\, |f_n - f| \geq \varepsilon\}\Big) \to 0 \text{ as } n \to \infty.$$
Assume that $\mu(S) < \infty$ (parts marked by (†) are true without this assumption).

1. Show that the mapping
$$d(f,g) = \int \tfrac{|f-g|}{1+|f-g|}\, d\mu,\ f,g \in \mathcal{L}^0,$$
defines a pseudo metric on $\mathcal{L}^0$ and that convergence in d is equivalent to convergence in measure.
2. Show that $f_n \to f$, a.e., implies that $f_n \to f$ in measure. Give an example which shows that the assumption $\mu(S) < \infty$ is necessary.
3. Give an example of a sequence which converges in measure, but not a.e.
4. (†) For $f \in \mathcal{L}^0$ and a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^0$, suppose that
$$\sum_{n\in\mathbb{N}} \mu\Big(\{\, |f_n - f| \geq \varepsilon\}\Big) < \infty, \text{ for all } \varepsilon > 0.$$
Show that $f_n \to f$, a.e.
5. (†) Show that each sequence convergent in measure has a subsequence which converges a.e.
6. (†) Show that each sequence convergent in $\mathcal{L}^p$, $p \in [1,\infty)$ converges in measure.
7. For $p \in [1,\infty)$, find an example of a sequence which converges in measure, but not in $\mathcal{L}^p$.
8. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0$ with the property that any of its subsequences admits a (further) subsequence which converges a.e. to $f \in \mathcal{L}^0$. Show that $f_n \to f$ in measure.
9. Let $\Phi : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function, and let $\{f_n\}_{n\in\mathbb{N}}$ and $\{g_n\}_{n\in\mathbb{N}}$ be two sequences in $\mathcal{L}^0$. If $f,g \in \mathcal{L}^0$ are such that $f_n \to f$ and $g_n \to g$ in measure, then
$$\Phi(f_n, g_n) \to \Phi(f,g) \text{ in measure.}$$
Note: The useful examples include $\Phi(x,y) = x+y$, $\Phi(x,y) = xy$, etc.

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 5
Theorems of Fubini-Tonelli and Radon-Nikodym

Products of measure spaces

We have seen that it is possible to define products of arbitrary collections of measurable spaces - one generates the σ-algebra on the product by all finite-dimensional cylinders. The purpose of the present section is to extend that construction to products of measure spaces, i.e., to define products of measures.

Let us first consider the case of two measure spaces $(S, \mathcal{S}, \mu_S)$ and $(T, \mathcal{T}, \mu_T)$. If the measures are stripped, the product $S \times T$ is endowed with the product σ-algebra $\mathcal{S} \otimes \mathcal{T} = \sigma(\{A \times B : A \in \mathcal{S}, B \in \mathcal{T}\})$. The family $\mathcal{P} = \{A \times B : A \in \mathcal{S}, B \in \mathcal{T}\}$ serves as a good starting point towards the creation of the product measure $\mu_S \otimes \mu_T$. Indeed, if we interpret of the elements in $\mathcal{P}$ as rectangles of sorts, it is natural to define

$$(\mu_S \otimes \mu_T)(A \times B) = \mu_S(A)\mu_T(B).$$

The family $\mathcal{P}$ is a π-system (why?), but not necessarily an algebra, so we cannot use Theorem 2.9 (Caratheodory's extension theorem) to define an extension of $\mu_S \otimes \mu_T$ to the whole $\mathcal{S} \otimes \mathcal{T}$. It it not hard, however, to enlarge $\mathcal{P}$ a little bit, so that the resulting set is an algebra, but that the measure $\mu_S \otimes \mu_T$ can still be defined there in a natural way. Indeed, consider the smallest algebra that contains $\mathcal{P}$. It is easy to see that it must contain the family $\mathcal{S}$ defined by

$$\mathcal{A} = \{\cup_{k=1}^n A_k \times B_k : n \in \mathbb{N},\, A_k \in \mathcal{S}, B_k \in \mathcal{T},\, k = 1, \dots, n\}.$$

Problem 5.1. Show that $\mathcal{A}$ is, in fact, an algebra and that each element $C \in \mathcal{A}$ can be written in the form

$$C = \cup_{k=1}^n A_k \times B_k,$$

for $n \in \mathbb{N}, A_k \in \mathcal{S}, B_k \in \mathcal{T}, k = 1, \dots, n$, such that $A_1 \times B_1, \dots, A_n \times B_n$ are *pairwise disjoint*.

The problem above allows us to extend the definition of the set function $\mu_S \otimes \mu_T$ to the entire $\mathcal{A}$ by

$$(\mu_S \otimes \mu_T)(C) = \sum_{k=1}^{n} \mu_S(A_k)\mu_T(B_k),$$

where $C = \cup_{k=1}^{n} A_k \times B_k$ for $n \in \mathbb{N}, A_k \in \mathcal{S}, B_k \in \mathcal{T}, k = 1, \dots, n$ is a representation of C with pairwise disjoint $A_1 \times B_1, \dots, A_n \times B_n$.

At this point, we could attempt to show that the so-defined set function is σ-additive on $\mathcal{A}$ and extend it using the Caratheodory extension theorem. This is indeed possible - under the additional assumption of σ-finiteness - but we will establish the existence of product measures as a side-effect in the proof of Fubini's theorem below.

Lemma 5.1 (Sections of measurable sets are measurable)**.** *Let C be an $\mathcal{S} \otimes \mathcal{T}$-measurable subset of $S \times T$. For each $x \in S$ the* ***section*** *$C_x = \{y \in T : (x, y) \in C\}$ is measurable in $\mathcal{T}$.*

Proof. In the spirit of most of the measurability arguments seen so far in these notes, let $\mathcal{C}$ denote the family of all $C \in \mathcal{S} \times \mathcal{T}$ such that C_x is $\mathcal{T}$-measurable for each $x \in S$. Clearly, the "rectangles" $A \times B$, $A \in \mathcal{S}$, $B \in \mathcal{T}$ are in $\mathcal{A}$ because their sections are either equal to $\emptyset$ or B, for each $x \in S$. Remember that the set of all rectangles generates $\mathcal{S} \otimes \mathcal{T}$. The proof of the theorem will, therefore, be complete once it is established that $\mathcal{C}$ is a σ-algebra. This easy exercise is left to the reader. □

Problem 5.2. Show that an analogous result holds for measurable functions, i.e., show that if $f : S \times T \to \bar{\mathbb{R}}$ is a $\mathcal{S} \otimes \mathcal{T}$-measurable function, then the function $x \mapsto f(x, y_0)$ is $\mathcal{S}$-measurable for each $y_0 \in T$, and the function $y \mapsto f(x_0, y)$ is $\mathcal{T}$-measurable for each $y_0 \in T$.

Proposition 5.2 (A simple Cavalieri's principle)**.** *Let μ_S and μ_T be finite measures. For $C \in \mathcal{S} \otimes \mathcal{T}$, define the functions $\varphi_C : T \to [0, \infty)$ and $\psi_C : S \to [0, \infty)$ by*

$$\varphi_C(y) = \mu_S(C_y), \ \psi_C(x) = \mu_T(C_x).$$

Then,

$$\varphi_C \in \mathcal{L}^0_+(\mathcal{T}), \ \psi_C \in \mathcal{L}^0_+(\mathcal{S}) \text{ and } \int \varphi_C \, d\mu_T = \int \psi_C \, d\mu_S. \tag{5.1}$$

Proof. Note that, by Problem 5.2, the function $x \mapsto \mathbf{1}_C(x, y)$ is $\mathcal{S}$-measurable for each $y \in T$. Therefore,

$$\int \mathbf{1}_C(\cdot, y) \, d\mu_S = \mu_S(C_y) = \varphi_C(y), \tag{5.2}$$

and the function φ_C is well-defined.

Let $\mathcal{C}$ denote the family of all sets in $\mathcal{S} \otimes \mathcal{T}$ such that (5.1) holds. First, observe that $\mathcal{C}$ contains all rectangles $A \times B$, $A \in \mathcal{S}$, $B \in \mathcal{T}$, i.e., it contains a π-system which generates $\mathcal{S} \otimes \mathcal{T}$. Therefore, by the π-λ Theorem (Theorem 2.12), it will be enough to show that $\mathcal{C}$ is a λ-system. We leave the details to the reader[1] □

[1] Use representation (5.2) and the monotone convergence theorem. Where is the finiteness of the measures used?

Proposition 5.3 (Simple Cavalieri holds for σ-finite measures)**.** *The conclusion of Proposition 5.2 continues to hold if we assume that μ_S and μ_T are only σ-finite.*

Proof (*). Thanks to σ-finiteness, there exists pairwise disjoint sequences $\{A_n\}_{n\in\mathbb{N}}$ and $\{B_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$ and $\mathcal{T}$, respectively, such that $\cup_n A_n = S$, $\cup_m B_m = T$ and $\mu_S(A_n) < \infty$ and $\mu_S(B_m) < \infty$, for all $m, n \in \mathbb{N}$.

For $m, n \in \mathbb{N}$, define the set-functions μ_S^n and μ_T^m on $\mathcal{S}$ and $\mathcal{T}$ respectively by

$$\mu_S^n(A) = \mu_S(A_n \cap A),\ \mu_T^m(B) = \mu_T(B_m \cap B).$$

It is easy to check that all μ_S^n and μ_T^m, $m, n \in \mathbb{N}$ are finite measures on $\mathcal{S}$ and $\mathcal{T}$, respectively. Moreover, $\mu_S(A) = \sum_{n=1}^\infty \mu_S^n(A)$, $\mu_T(B) = \sum_{m=1}^\infty \mu_T^m(B)$. In particular, if we set $\varphi_C^n(y) = \mu_S^n(C_y)$ and $\psi_C^m(x) = \mu_T^m(C_x)$, for all $x \in S$ and $y \in S$, we have

$$\varphi_C(y) = \mu_S(C_y) = \sum_{n=1}^\infty \mu_S^n(C_y) = \sum_{n=1}^\infty \varphi_C^n(y), \text{ and}$$
$$\psi_C(x) = \mu_T(C_x) = \sum_{m=1}^\infty \mu_T^m(C_x) = \sum_{m=1}^\infty \psi_C^m(x),$$

for all $x \in S$, $y \in T$.

We can apply the conclusion of Proposition 5.2 to all pairs $(S, \mathcal{S}, \mu_S^n)$ and $(T, \mathcal{T}, \mu_T^m)$, $m, n \in \mathbb{N}$, of finite measure spaces to conclude that all elements of the sums above are measurable functions and that so are φ_C and ψ_C.

Similarly, the sequences of non-negative functions $\sum_{i=1}^n \varphi_C^n(y)$ and $\sum_{i=1}^m \psi_C^i(x)$ are non-decreasing and converge to φ_C and ψ_C. Therefore, by the monotone convergence theorem,

$$\int \varphi_C \, d\mu_T = \lim_n \sum_{i=1}^n \int \varphi_C^i \, d\mu_T, \text{ and } \int \psi_C \, d\mu_S = \lim_n \sum_{i=1}^n \int \psi_C^i \, d\mu_S.$$

On the other hand, we have $\int \varphi_C^n d\mu_T^m = \int \psi_C^m \, d\mu_S^n$, by Proposition 5.2. Summing over all $n \in \mathbb{N}$ we have

$$\int \varphi_C \, d\mu_T^m = \sum_{n\in\mathbb{N}} \int \psi_C^m \, d\mu_S^n = \int \psi_C^m \, d\mu_S,$$

where the last equality follows from the fact (see Problem 5.3 below) that

$$\sum_{n\in\mathbb{N}} \int f\, d\mu_S^n = \int f\, d\mu_S,$$

for all $f \in \mathcal{L}_+^0$. Another summation - this time over $m \in \mathbb{N}$ - completes the proof [2]. □

[2] The argument of the proof above uncovers the fact that integration is a bilinear operation, i.e., that the mapping

$$(f, \mu) \to \int f\, d\mu,$$

is linear in both arguments.

Problem 5.3. Let $\{A_n\}_{n\in\mathbb{N}}$ be a measurable partition of S, and let the measure μ^n be defined by $\mu^n(A) = \mu(A \cap A_n)$ for all $A \in \mathcal{S}$. Show that for $f \in \mathcal{L}_+^0$, we have

$$\int f\, d\mu = \sum_{n\in\mathbb{N}} \int f\, d\mu^n.$$

Proposition 5.4 (Finite products of measure spaces)**.** *Let $(S_i, \mathcal{S}_i, \mu_i)$, $i = 1, \dots, n$ be σ-finite measure spaces. There exists a unique measure[3] - denoted by $\mu_1 \otimes \dots \otimes \mu_n$ - on the product space $(S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n)$ with the property that*

$$(\mu_1 \otimes \dots \otimes \mu_n)(A_1 \times \dots \times A_n) = \mu_1(A_1) \dots \mu_n(A_n),$$

for all $A_i \in \mathcal{S}_i$, $i = 1, \dots, n$. Such a measure is necessarily σ-finite.

[3] The measure $\mu_1 \otimes \dots \otimes \mu_n$ is called the **product measure**, and the measure space $(S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n, \mu_1 \otimes \dots \otimes \mu_n)$ the **product (measure space)** of measure spaces $(S_1, \mathcal{S}_1, \mu_1)$, $\dots$, $(S_n, \mathcal{S}_n, \mu_n)$.

Proof. To simplify the notation, we assume that $n = 2$ - the general case is very similar. For $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$, we define

$$(\mu_1 \otimes \mu_2)(C) = \int_{S_2} \varphi_C\, d\mu_2,$$

where $\varphi_C(y) = \mu_1(C_y)$ and $C_y = \{x \in S_1 : (x, y) \in C\}$. It follows from Proposition 5.3 that $\mu_1 \otimes \mu_2$ is well-defined as a map from $\mathcal{S}_1 \otimes \mathcal{S}_2$ to $[0, \infty]$. Also, it is clear that $(\mu_1 \otimes \mu_2)(A \times B) = \mu_1(A)\mu_2(B)$, for all $A \in \mathcal{S}_1$, $B \in \mathcal{S}_2$. It remains to show that $\mu_1 \otimes \mu_2$ is a measure. We start with a pairwise disjoint sequence $\{C_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}_1 \otimes \mathcal{S}_2$. For $y \in S_2$, the sequence $\{(C_n)_y\}_{n\in\mathbb{N}}$ is also pairwise disjoint, and so, with $C = \cup_n C_n$, we have

$$\varphi_C(y) = \mu_1(C_y) = \sum_{n\in\mathbb{N}} \mu_2\Big((C_n)_y\Big) = \sum_{n\in\mathbb{N}} \varphi_{C_n}(y),\ \forall\, y \in S_2.$$

Therefore, by the monotone convergence theorem (see Problem 3.13 for details) we have

$$(\mu_1 \otimes \mu_2)(C) = \int_{S_2} \varphi_C\, d\mu_2 = \sum_{n\in\mathbb{N}} \int_{S_2} \varphi_{C_n}\, d\mu = \sum_{n\in\mathbb{N}} (\mu_1 \otimes \mu_2)(C_n).$$

Finally, let $\{A_n\}_{n\in\mathbb{N}}$, $\{B_n\}_{n\in\mathbb{N}}$ be sequences in $\mathcal{S}_1$ and $\mathcal{S}_2$ (respectively) such that $\mu_1(A_n) < \infty$ and $\mu_2(B_n) < \infty$ for all $n \in \mathbb{N}$ and $\cup_n A_n = S_1$, $\cup_n B_n = S_2$. Define $\{C_n\}_{n\in\mathbb{N}}$ as an enumeration of the countable family $\{A_i \times B_j : i, j \in \mathbb{N}\}$ in $\mathcal{S}_1 \otimes \mathcal{S}_2$. Then $(\mu_1 \otimes \mu_2)(C_n) < \infty$ and all $n \in \mathbb{N}$ and $\cup_n C_n = S_1 \times S_2$. Therefore, $\mu_1 \otimes \mu_2$ is σ-finite. □

Now that we know that product measures exist, we can state and prove the important theorem which, when applied to integrable functions bears the name of Fubini, and when applied to non-negative functions, of Tonelli. We state it for both cases simultaneously (i.e., on $\mathcal{L}^{0-1}$) in the case of a product of two measure spaces. An analogous theorem for finite products can be readily derived from it. When the variable or the underlying measure space of integration needs to be specified, we write $\int_S f(x)\,\mu(dx)$ for the Lebesgue integral $\int f\,d\mu$.

Theorem 5.5 (Fubini, Tonelli)**.** *Let $(S, \mathcal{S}, \mu_S)$ and $(T, \mathcal{T}, \mu_T)$ be two σ-finite measure spaces. For $f \in \mathcal{L}^{0-1}(S \times T)$ we have*

$$\begin{aligned} \int f\,d(\mu_S \otimes \mu_T) &= \int_S \left(\int_T f(x,y)\,\mu_T(dy) \right) \mu_S(dx) \\ &= \int_T \left(\int_S f(x,y)\,\mu_S(dx) \right) \mu_T(dy) \end{aligned} \tag{5.3}$$

Proof. All the hard work has already been done. We simply need to crank the Standard Machine. Let $\mathcal{H}$ denote the family of all functions in $\mathcal{L}^0_+(S \times T)$ with the property that (5.3) holds. Proposition 5.3 implies that $\mathcal{H}$ contains the indicators of all elements of $\mathcal{S} \otimes \mathcal{T}$. Linearity of all components of (5.3) implies that $\mathcal{H}$ contains all simple functions in $\mathcal{L}^0_+$, and the approximation theorem 3.10 implies that the whole $\mathcal{L}^0_+$ is in $\mathcal{H}$. Finally, the extension to $\mathcal{L}^{0-1}$ follows by additivity. □

Since f^- is always in $\mathcal{L}^{0-1}$, we have the following corollary

Corollary 5.6. *For $f \in \mathcal{L}^0(S \times T)$, we have*

$$f \in \mathcal{L}^{0-1}(S \times T) \text{ if and only if } \int_S \left(\int_T f^-(x,y)\,\mu_T(dy) \right) \mu_S(dx) < \infty.$$

Example 5.7. The assumption of σ-finiteness cannot be left out of the statement of Theorem 5.5. Indeed, let $(S, \mathcal{S}, \mu) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and $(T, \mathcal{T}, \nu) = ([0,1], 2^{[0,1]}, \gamma)$, where γ is the counting measure on $2^{[0,1]}$, so that $(T, \mathcal{T}, \nu)$ fails the σ-finite property. Define $f \in \mathcal{L}^0(S \times T)$ (why it is product-measurable?) by

$$f(x,y) = \begin{cases} 1, & x = y, \\ 0, & x \neq y. \end{cases}$$

Then

$$\int_S f(x,y)\,\mu(dx) = \lambda(\{y\}) = 0,$$

and so

$$\int_T \int_S f(x,y)\,\mu(dx)\,\nu(dy) = \int_{[0,1]} 0\,\gamma(dy) = 0.$$

On the other hand,

$$\int_T f(x,y)\,\nu(dy) = \gamma(\{x\}) = 1,$$

and so

$$\int_S \int_T f(x,y)\,\nu(dy)\,\mu(dx) = \int_{[0,1]} 1\,\lambda(dx) = 1.$$

Example 5.8. The integrability of either f^+ or f^- for $f \in \mathcal{L}^0(S \times T)$ is (essentially) necessary for validity of Fubini's theorem, even if all iterated integrals exist. Here is what can go wrong. Let $(S, \mathcal{S}, \mu) = (T, \mathcal{T}, \nu) = (\mathbb{N}, 2^{\mathbb{N}}, \gamma)$, where γ is the counting measure. Define the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ by

$$f(n,m) = \begin{cases} 1, & m = n, \\ -1, & m = n+1, \\ 0, & \text{otherwise} \end{cases}$$

Then

$$\int_T f(n,m)\,\gamma(dm) = \sum_{m\in\mathbb{N}} f(n,m) = 0 + \dots + 0 + 1 + (-1) + 0 + \dots = 0,$$

and so

$$\int_S \int_T f(n,m)\,\gamma(dm)\,\gamma(dn) = 0.$$

On the other hand,

$$\int_S f(n,m)\,\gamma(dn) = \sum_{n\in\mathbb{N}} f(n,m) = \begin{cases} 1, & m = 1, \\ 0, & m > 1, \end{cases}$$

i.e.,

$$\int_S f(n,m)\,\gamma(dn) = \mathbf{1}_{\{m=1\}}.$$

Therefore,

$$\int_T \int_S f(n,m)\,\gamma(dn)\,\gamma(dm) = \int_T \mathbf{1}_{\{m=1\}}\,\gamma(dm) = 1.$$

If you think that using the counting measure is cheating, convince yourself that it is not hard to transfer this example to the setup where $(S, \mathcal{S}, \mu) = (T, \mathcal{T}, \nu) = ([0,1], \mathcal{B}([0,1]), \lambda)$.

The existence of the product measure gives us an easy access to the Lebesgue measure on higher-dimensional Euclidean spaces. Just as λ on $\mathbb{R}$ measures the "length" of sets, the Lebesgue measure on $\mathbb{R}^2$ will measure "area", the one on $\mathbb{R}^3$ "volume", etc. Its properties are collected in the following problem:

Problem 5.4. For $n \in \mathbb{N}$, show the following statements:

1. There exists a unique measure λ (note the notation overload) on $\mathcal{B}(\mathbb{R}^n)$ with the property that

$$\lambda\Big([a_1, b_1) \times \cdots \times [a_n, b_n)\Big) = (b_1 - a_1) \dots (b_n - a_n),$$

   for all $a_1 < b_1, \dots, a_n < b_n$ in $\mathbb{R}$.

2. The measure λ on $\mathbb{R}^n$ is invariant with respect to all isometries[4] of $\mathbb{R}^n$.

   Note: Feel free to use the following two facts without proof:

   (a) A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry if and only if there exists $x_0 \in \mathbb{R}^n$ and an orthogonal linear transformation $O : \mathbb{R}^n \to \mathbb{R}^n$ such that $f(x) = x_0 + Ox$.

   (b) Let O be an orthogonal linear transformation. Then R_1 and OR_1 have the same Lebesgue measure, where R_1 denotes the unit rectangle $[0, 1) \times \cdots \times [0, 1)$.

[4] An **isometry** of $\mathbb{R}^n$ is a map $f : \mathbb{R}^n \to \mathbb{R}^n$ with the property that $d(x, y) = d(f(x), f(y))$ for all $x, y \in \mathbb{R}^n$. It can be shown that the isometries of $\mathbb{R}^3$ are precisely translations, rotations, reflections and compositions thereof.

Note: probably the least painful way to prove this fact is by using the change-of-variable formula for the Riemann integral.

The Radon-Nikodym Theorem

We start the discussion of the Radon-Nikodym theorem with a simple observation:

Problem 5.5 (Integral as a measure)**.** For a function $f \in \mathcal{L}^0([0, \infty])$, we define the set-function $\nu : \mathcal{S} \to [0, \infty]$ by

$$\nu(A) = \int_A f \, d\mu. \tag{5.4}$$

1. Show that ν is a measure.

2. Show that $\mu(A) = 0$ implies $\nu(A) = 0$, for all $A \in \mathcal{S}$.

3. Show that the following two properties are equivalent

   - $\mu(A) = 0$ if and only if $\nu(A) = 0$, $A \in \mathcal{S}$, and
   - $f > 0$, a.e.

Definition 5.9 (Absolute continuity, etc.)**.** Let μ, ν be measures on the measurable space $(S, \mathcal{S})$. We say that

1. ν is **absolutely continuous** with respect to μ - denoted by $\nu \ll \mu$ - if $\nu(A) = 0$, whenever $\mu(A) = 0$, $A \in \mathcal{S}$.

2. μ and ν are **equivalent** if $\nu \ll \mu$ and $\mu \ll \nu$, i.e., if $\mu(A) = 0 \Leftrightarrow \nu(A) = 0$, for all $A \in \mathcal{S}$,

3. μ and ν are **(mutually) singular** - denoted by $\mu \perp \nu$ - if there exists $D \in \mathcal{S}$ such that $\mu(D) = 0$ and $\nu(D^c) = 0$.

Problem 5.6. Let μ and ν be measures with ν finite and $\nu \ll \mu$. Show that for each $\varepsilon > 0$ there exists $\delta > 0$ such that for each $A \in \mathcal{S}$, we have $\mu(A) \leq \delta \Rightarrow \nu(A) \leq \varepsilon$. Show that the assumption that ν is finite is necessary.

Problem 5.5 states that the prescription (5.4) defines a measure on $\mathcal{S}$ which is absolutely continuous with respect to μ. What is surprising is that the converse also holds under the assumption of σ-finiteness: all absolutely continuous measures on $\mathcal{S}$ are of that form. That statement (and more) is the topic of this section. Since there is more than one measure in circulation, we use the convention that *a.e.* always uses the notion of the null set as defined by the measure μ.

Theorem 5.10 (The Lebesgue decomposition)**.** *Let $(S, \mathcal{S})$ be a measurable space and let μ and ν be two σ-finite measures on $\mathcal{S}$. Then there exists a unique decomposition $\nu = \nu_a + \nu_s$, where*

1. *$\nu_a \ll \mu$,*
2. *$\nu_s \perp \mu$.*

Furthermore, there exists an a.e.-unique function $h \in \mathcal{L}^0_+$ such that

$$\nu_a(A) = \int_A h \, d\mu.$$

The proof is based on the following particular case of the Riesz representation theorem

Proposition 5.11. *Let $(S, \mathcal{S}, \mu)$ be a measure space, and let $L : \mathcal{L}^2(\mu) \to \mathbb{R}$ be a linear map with the property that*

$$|Lf| \leq C||f||_{\mathcal{L}^2} \text{ for all } f \in \mathcal{L}^2$$

for some constant $C \geq 0$. Then, there exists an a.e.-unique element $g \in \mathcal{L}^2$ such that

$$Lf = \int fg \, d\mu, \text{ for all } f \in \mathcal{L}^2.$$

Proof. If $L = 0$, the statement clearly holds with $g = 0$, so we assume $Lh \neq 0$ for some $h \in \mathcal{L}^2$. The set

$$\operatorname{Nul} L = \{f \in \mathcal{L}^2 : Lf = 0\}$$

is a nonempty and closed (why?) subspace of $\mathcal{L}^2$ with $h \notin \operatorname{Nul} L$. Therefore, there exists a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $\operatorname{Nul} L$ such that

$$a_n = ||f_n - h|| \to M = \inf_{f \in \operatorname{Nul} L} ||f - h|| > 0.$$

It is easy to check that

$$\begin{aligned}||f_n - f_m||^2 &= -||f_n + f_m - 2h||^2 + 2||f_n - h||^2 + 2||f_m - h||^2 \\ &\leq -4M^2 + 2a_n^2 + 2a_m^2 \to 0 \text{ as } m, n \to \infty,\end{aligned}$$

and, hence, that $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{L}^2$. By completeness (Proposition 4.13) there exists $\hat{f} \in \mathcal{L}^2$ such that $f_n \to \hat{f}$ in $\mathcal{L}^2$. It follows immediately that $||\hat{g}|| = M$, where $\hat{g} = \hat{f} - h$. Therefore,

$$||\hat{g}||^2 \leq ||\hat{g} + \varepsilon f||^2, \text{ for all } \varepsilon > 0, f \in \operatorname{Nul} L,$$

and so

$$\varepsilon||f||^2 + 2\int f\hat{g} \geq 0, \text{ for all } \varepsilon > 0, f \in \operatorname{Nul} L,$$

which, in turn, immediately implies that[5]

$$\int f\hat{g}\, d\mu = 0, \text{ for all } f \in \operatorname{Nul} L.$$

[5] Compare this argument to Problem 4.9

In words, $\hat{g}$ is orthogonal to $\operatorname{Nul} L$. Starting from the identity

$$Lf = L\alpha h, \text{ where } \alpha = Lf/Lh,$$

we deduce that $f - \alpha h \in \operatorname{Nul} L$ and, so,

$$\int (f - \alpha h)\hat{g}\, d\mu = 0, \text{ i.e.,} \int f\hat{g}\, d\mu = \tfrac{Lf}{Lh}\int h\hat{g}\, d\mu.$$

Therefore, noting that h cannot be orthogonal to $\hat{g}$, we get

$$Lf = \int fg\, d\mu \text{ where } g = \tfrac{Lh}{\int h\hat{g}\, d\mu}\hat{g}. \qquad \square$$

Proof of Theorem 5.10. Uniqueness: Suppose that $\nu_a^1 + \nu_s^1 = \nu = \nu_a^2 + \nu_s^2$ are two decompositions satisfying (1) and (2) in the statement. Let D^1 and D^2 be as in the definition of mutual singularity applied to the pairs μ, ν_s^1 and μ, ν_s^2, respectively. Set $D = D^1 \cup D^2$, and note that $\mu(D) = 0$ and $\nu_s^1(D^c) = \nu_s^2(D^c) = 0$. For any $A \in \mathcal{S}$, we have $\mu(A \cap D) = 0$ and so, thanks to absolute continuity,

$$\nu_a^1(A \cap D) = \nu_a^2(A \cap D) = 0$$

so that $\nu_s^1(A \cap D) = \nu_s^2(A \cap D) = \nu(A \cap D)$. By singularity, we have

$$\nu_s^1(A \cap D^c) = \nu_s^2(A \cap D^c) = 0,$$

and, consequently, $\nu_a^1(A \cap D^c) = \nu_a^2(A \cap D^c) = \nu(A \cap D^c)$. Finally,

$$\nu_a^1(A) = \nu_a^1(A \cap D) + \nu_a^1(A \cap D^c) = \nu_a^2(A \cap D) + \nu_a^2(A \cap D^c) = \nu^2(A),$$

and, similarly, $\nu_s^1 = \nu_s^2$.

To establish the uniqueness of the function f with the property that $\nu_a(A) = \int_A f\, d\mu$ for all $A \in \mathcal{S}$, we assume that there are two such functions, f_1 and f_2, say. Define the sequence $\{B_n\}_{n\in\mathbb{N}}$ by

$$B_n = \{f_1 \geq f_2\} \cap C_n,$$

where $\{C_n\}_{n\in\mathbb{N}}$ is a pairwise-disjoint sequence in $\mathcal{S}$ with the property that $\nu(C_n) < \infty$, for all $n \in \mathbb{N}$ and $\cup_n C_n = S$. Then, with $g_n = f_1 \mathbf{1}_{B_n} - f_2 \mathbf{1}_{B_n} \in \mathcal{L}^1_+$ we have

$$\int g_n\, d\mu = \int_{B_n} f_1\, d\mu - \int_{B_n} f_2\, d\mu = \nu_a(B_n) - \nu_a(B_n) = 0.$$

By Problem 3.9, we have $g_n = 0$, a.e., i.e., $f_1 = f_2$, a.e., on B_n, for all $n \in \mathbb{N}$, and so $f_1 = f_2$, a.e., on $\{f_1 \geq f_2\}$. A similar argument can be used to show that $f_1 = f_2$, a.e., on $\{f_1 < f_2\}$, as well.

Existence: Assume, first, that μ and ν are probability measures and set $\varphi = \mu + \nu$. We define

$$Lf = \int f\, d\nu \text{ for } f \in \mathcal{L}^2(\varphi),$$

and note that,

$$|Lf| \leq \int |f|\, d\nu \leq \int |f|\, d\varphi \leq ||f||_{\mathcal{L}^2(\varphi)},$$

Proposition 5.11 can be applied to yield the existence of $g \in \mathcal{L}^2(\varphi)$ such that $Lf = \int fg\, d\varphi$, for all $f \in \mathcal{L}^2(\varphi)$. Since $Lf \geq 0$, for $f \geq 0$, φ-a.e., we conclude (why?) that $g \geq 0$, φ-a.e. By approximation (via the monotone convergence theorem) we then have

$$\int F\, d\nu = \int Fg\, d\nu + \int Fg d\mu, \text{ for all } F \in \mathcal{L}^0_+(\varphi). \tag{5.5}$$

If we plug $F = \mathbf{1}_{\{g \geq 1+\varepsilon\}}$ in (5.5), we obtain

$$\nu(g \geq 1+\varepsilon) \geq (1+\varepsilon)\varphi(g \geq 1+\varepsilon),$$

and conclude that $g \leq 1$, φ-a.e. It also follows, by taking $F = \mathbf{1}_D$, where $D = \{g = 1\}$, that $\mu(D) = 0$. The function $h = \frac{g}{1-g}\mathbf{1}_{\{g<1\}}$ is nonnegative, measurable and finitely valued φ-a.e., so we can use $F = \frac{f}{1-g}\mathbf{1}_{D^c}$, for $f \in \mathcal{L}^0_+(\varphi)$ in (5.5) to obtain

$$\int f \mathbf{1}_{D^c}\, d\nu = \int fh\, d\mu.$$

It follows immediately that

$$\nu(A) = \int \mathbf{1}_A h\, d\mu + \nu(A \cap D),$$

which provides the required decomposition

$$\nu_a(A) = \int \mathbf{1}_A h\, d\mu,\ \nu_s(A) = \nu(A \cap D). \qquad \square$$

Corollary 5.12 (Radon-Nikodym)**.** *Let μ and ν be σ-finite measures on $(S, \mathcal{S})$ with $\nu \ll \mu$. Then there exists $f \in \mathcal{L}^0_+$ such that*

$$\nu(A) = \int_A f\, d\mu, \text{ for all } A \in \mathcal{S}. \tag{5.6}$$

For any other $g \in \mathcal{L}^0_+$ with the same property, we have $f = g$, a.e.

Any function f for which (5.6) holds is called the **Radon-Nikodym derivative** of ν with respect to μ and is denoted by $f = \frac{d\nu}{d\mu}$, a.e. [6]

[6] The Radon-Nikodym derivative $f = \frac{d\nu}{d\mu}$ is defined only up to a.e.-equivalence, and there is no canonical way of picking a representative defined for all $x \in S$. For that reason, we usually say that a function $f \in \mathcal{L}^0_+$ is a **version** of the Radon-Nikodym derivative of ν with respect to μ if (5.6) holds. Moreover, to stress the fact that we are talking about a whole class of functions instead of just one, we usually write

$$\frac{d\nu}{d\mu} \in \mathbb{L}^0_+ \text{ and not } \frac{d\nu}{d\mu} \in \mathcal{L}^0_+.$$

We often neglect this fact notationally, and write statements such as "*If $f \in \mathcal{L}^0_+$ and $f = \frac{d\nu}{d\mu}$ then ...*". What we really mean is that the statement holds regardless of the particular *representative* f of the Radon-Nikodym derivative we choose. Also, when we write $\frac{d\nu}{d\mu} = \frac{d\rho}{d\mu}$, we mean that they are equal as elements of $\mathbb{L}^0_+$, i.e., that there exists $f \in \mathcal{L}^0_+$, which is both a version of $\frac{d\nu}{d\mu}$ and a version of $\frac{d\rho}{d\mu}$.

Problem 5.7. Let μ, ν and ρ be σ-finite measures on $(S, \mathcal{S})$. Show that

1. If $\nu \ll \mu$ and $\rho \ll \mu$, then $\nu + \rho \ll \mu$ and
$$\frac{d\nu}{d\mu} + \frac{d\rho}{d\mu} = \frac{d(\nu+\rho)}{d\mu}.$$
2. If $\nu \ll \mu$ and $f \in \mathcal{L}^0_+$, then
$$\int f\, d\nu = \int g\, d\mu \text{ where } g = f\frac{d\nu}{d\mu}.$$
3. If $\nu \ll \mu$ and $\rho \ll \nu$, then $\rho \ll \mu$ and
$$\frac{d\nu}{d\mu}\frac{d\rho}{d\nu} = \frac{d\rho}{d\mu}.$$
Note: Make sure to pay attention to the fact that different measure give rise to different families of null sets, and, hence, to different notions of *almost everywhere*.
4. If $\mu \sim \nu$, then
$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$

Problem 5.8. Let $\mu_1, \mu_2, \nu_1, \nu_2$ be σ-finite measures with μ_1 and ν_1, as well as μ_2 and ν_2, defined on the same measurable space. If $\nu_1 \ll \mu_1$ and $\nu_2 \ll \mu_2$, show that $\nu_1 \otimes \nu_2 \ll \mu_1 \otimes \mu_2$.

Example 5.13. Just like in the statement of Fubini's theorem, the assumption of σ-finiteness cannot be omitted. Indeed, take $(S, \mathcal{S}) = ([0,1], \mathcal{B}([0,1]))$ and consider the Lebesgue measure λ and the counting measure γ on $(S, \mathcal{S})$. Clearly, $\lambda \ll \gamma$, but there is no $f \in \mathcal{L}^0_+$ such that $\lambda(A) = \int_A f\, d\gamma$. Indeed, suppose that such f exists and set $D_n = \{x \in S : f(x) > 1/n\}$, for $n \in \mathbb{N}$, so that $D_n \nearrow \{f > 0\} = \{f \neq 0\}$. Then

$$1 \geq \lambda(D_n) = \int_{D_n} f\, d\gamma \geq \int_{D_n} \tfrac{1}{n}\, d\gamma = \tfrac{1}{n}\#D_n,$$

and so $\#D_n \leq n$. Consequently, the set $\{f > 0\} = \cup_n D_n$ is countable. This leads to a contradiction since the Lebesgue measure does not "charge" countable sets, and so

$$1 = \lambda([0,1]) = \int f \, d\gamma = \int_{\{f>0\}} f \, d\gamma = \lambda(\{f > 0\}) = 0.$$

Additional Problems

Problem 5.9 (Area under the graph of a function)**.** For $f \in \mathcal{L}^0_+$, let $H = \{(x,r) \in S \times [0,\infty) : f(x) \geq r\}$ be the "region under the graph" of f. Show that $\int f \, d\mu = (\mu \otimes \lambda)(H)$.

Note: The equality in this problem is consistent with our intuition that the value of the integral $\int f \, d\mu$ corresponds to the "area under the graph of f".

Problem 5.10 (A layered representation)**.** Let ν be a measure on $\mathcal{B}([0,\infty))$ such that $N(u) = \nu([0,u)) < \infty$, for all $u \in \mathbb{R}$. Let $(S, \mathcal{S}, \mu)$ be a σ-finite measure space. For $f \in \mathcal{L}^0_+(S)$, show that

1. $\int N \circ f \, d\mu = \int_{[0,\infty)} \mu(\{f > u\}) \, \nu(du)$.
2. for $p > 0$, we have $\int f^p \, d\mu = p \int_{[0,\infty)} u^{p-1} \mu(\{f > u\}) \, \lambda(du)$, where λ is the Lebesgue measure.

Problem 5.11 (A useful integral)**.**

1. Show that $\int_0^\infty \left|\frac{\sin x}{x}\right| dx = \infty$.

Hint: Find a function laying below $\left|\frac{\sin x}{x}\right|$ which is easier to integrate.

2. For $a > 0$, let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} e^{-xy} \sin(x), & 0 \leq x \leq a, \ 0 \leq y, \\ 0, & \text{otherwise.} \end{cases}$$

Show that $f \in \mathcal{L}^1(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \lambda)$, where λ denotes the Lebesgue measure on $\mathbb{R}^2$.

3. Establish the equality

$$\int_0^a \frac{\sin x}{x} \, dx = \frac{\pi}{2} - \cos(a) \int_0^\infty \frac{e^{-ay}}{1+y^2} \, dy - \sin(a) \int_0^\infty \frac{ye^{-ay}}{1+y^2} \, dy.$$

4. Conclude that for $a > 0$, $\left| \int_0^a \frac{\sin(x)}{x} \, dx - \frac{\pi}{2} \right| \leq \frac{2}{a}$, so that

$$\lim_{a\to\infty} \int_0^a \frac{\sin(x)}{x} \, dx = \frac{\pi}{2}.$$

Problem 5.12 (The Cantor measure)**.** Let $(\{-1,1\}^{\mathbb{N}}, \mathcal{B}(\{-1,1\}^{\mathbb{N}}), \mu_C)$ be the coin-toss space. Define the mapping $f : \{-1,1\}^{\mathbb{N}} \to [0,1]$ by

$$f(\mathbf{s}) = \sum_{n\in\mathbb{N}} (1 + s_n) 3^{-n}, \text{ for } \mathbf{s} = (s_1, s_2, \dots).$$

Let δ be the push-forward of μ_C by the map f. It is called the **Cantor measure**.

1. Let d be the metric on $\{-1,1\}^{\mathbb{N}}$ (as given by the equation (1.1)). Show that for $\alpha = \log_3(2)$ and $\boldsymbol{s}^1, \boldsymbol{s}^2 \in \{-1,1\}^{\mathbb{N}}$, we have

$$d(\boldsymbol{s}^1, \boldsymbol{s}^2)^\alpha \leq \left| f(\boldsymbol{s}^2) - f(\boldsymbol{s}^1) \right| \leq 3d(\boldsymbol{s}^1, \boldsymbol{s}^2)^\alpha.$$

2. Show that δ is atom-free, i.e., that $\delta(\{x\}) = 0$, for all $x \in [0,1]$,

3. For a measure μ on the σ-algebra of Borel sets of a topological space X, the **support** of μ is collection of all $x \in X$ with the property that $\mu(O) > 0$ for each open set O with $x \in O$. Describe the support of δ. *Hint:* Guess what it is and prove that your guess is correct. Use the result in (1).

4. Prove that $\delta \perp \lambda$. *Note:* The Cantor measure is an example of a **singular measure**. It has no atoms, but is still singular with respect to the Lebesgue measure.

Problem 5.13 (Joint measurability)**.**

1. Give an example of a function $f : [0,1] \times [0,1] \to [0,1]$ such that $x \mapsto f(x,y)$ and $y \mapsto f(x,y)$ are $\mathcal{B}([0,1])$-measurable functions for each $y \in [0,1]$ and $x \in [0,1]$, respectively, but that f is not $\mathcal{B}([0,1] \times [0,1])$-measurable.

Hint: You can use the fact that there exists a subset of $[0,1]$ which is not Borel measurable.

2. Let $(S, \mathcal{S})$ be a measurable space. A function $f : S \times \mathbb{R} \to \mathbb{R}$ is called a **Caratheodory function** if
   - $x \mapsto f(x,y)$ is $\mathcal{S}$-measurable for each $y \in \mathbb{R}$, and
   - $y \mapsto f(x,y)$ is continuous for each $x \in \mathbb{R}$.

   Show that Caratheodory functions are $\mathcal{S} \otimes \mathcal{B}(\mathbb{R})$-measurable.

Hint: Express a Caratheodory function as limit of a sequence of the form $f_n = \sum_{k \in \mathbb{Z}} g_{n,k}(x) h_{n,k}(r)$, $n \in \mathbb{N}$.

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 6
Basic Probability

Probability spaces

A mathematical setup behind a probabilistic model consists of a **sample space** Ω, a family of **events** and a **probability** $\mathbb{P}$. One thinks of Ω as being the set of all possible outcomes of a given random phenomenon, and the occurrence of a particular **elementary outcome** $\omega \in \Omega$ as depending on factors whose behavior is not fully known to the modeler. The family $\mathcal{F}$ is taken to be some collection of subsets of Ω, and for each $A \in \mathcal{F}$, the number $\mathbb{P}[A]$ is interpreted as the likelihood that some $\omega \in A$ occurs. Using the basic intuition that $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$, whenever A and B are disjoint (**mutually exclusive**) events, we conclude $\mathbb{P}$ should have all the properties of a finitely-additive measure. Moreover, a natural choice of normalization dictates that the likelihood of the **certain event** Ω be equal to 1. A regularity assumption[1] is often made and $\mathbb{P}$ is required to be σ-additive. All in all, we can single out probability spaces as a sub-class of measure spaces:

[1] Whether it is harmless or not leads to a very interesting philosophical discussion, but you will not get to read about it in these notes

Definition 6.1. A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a non-empty set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

In many (but certainly not all) aspects, probability theory is a part of measure theory. For historical reasons and because of a different interpretation, some of the terminology/notation changes when one talks about measure-theoretic concepts in probability. Here is a list of what is different, and what stays the same:

1. We will always assume - often without explicit mention - that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given and fixed.
2. Continuity of measure is called **continuity of probability**, and, unlike the general case, does not require and additional assumptions in the case of a decreasing sequence (that is, of course, because $\mathbb{P}[\Omega] = 1 < \infty$.)

3. A measurable function from Ω to $\mathbb{R}$ is called a **random variable**[2]. Typically, the sample space Ω is too large for analysis, so we often focus our attention to families of real-valued functions[3] X on Ω. This way, $X^{-1}([a,b])$ is the set of all elementary outcomes $\omega \in \Omega$ with for which $X(\omega) \in [a,b]$. If we want to be able to compute the probability $\mathbb{P}[X^{-1}([a,b])]$, the set $X^{-1}([a,b])$ better be an event, i.e., $X^{-1}([a,b]) \in \mathcal{F}$. Hence the measurability requirement.

   Sometimes, it will be more convenient for random variables to take values in the extended set $\bar{\mathbb{R}}$ of real numbers. In that case we talk about **extended random variables** or **$\bar{\mathbb{R}}$-valued random variables**.

4. We use the measure-theoretic notation $\mathcal{L}^0$,$\mathcal{L}^0_+$, $\mathcal{L}^0(\bar{\mathbb{R}})$, etc. to denote the set of all random variables, non-negative random variables, extended random variables, etc.

5. Let $(S, \mathcal{S})$ be a measurable space. An $(\mathcal{F}, \mathcal{S})$-measurable map $X : \Omega \to S$ is called a **random element (of S)**.

   Random variables are random elements, but there are other important examples. If $(S, \mathcal{S}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, we talk about **random vectors**. More generally, if $S = \mathbb{R}^{\mathbb{N}}$ and $\mathcal{S} = \prod_n \mathcal{B}(\mathbb{R})$, the map $X : \Omega \to S$ is called a **discrete-time stochastic process**. Sometimes, the object of interest is a set (the area covered by a wildfire, e.g.) and then S is a collection of subsets of $\mathbb{R}^n$. There are many more examples.

6. The class of null-sets in $\mathcal{F}$ still plays the same role as it did in measure theory, but now we use the acronym **a.s.** (which stands for *almost surely*) instead of the measure-theoretic a.e.

7. The Lebesgue integral with respect to the probability $\mathbb{P}$ is now called **expectation** and is denoted by $\mathbb{E}$, so that we write

$$\mathbb{E}[X] \text{ instead of } \int X \, d\mathbb{P}, \text{ or } \int_\Omega X(\omega) \, \mathbb{P}[d\omega].$$

   For $p \in [1, \infty]$, the $\mathcal{L}^p$ spaces are defined just like before, and have the property that $\mathcal{L}^q \subseteq \mathcal{L}^p$, when $p \le q$.

8. The notion of a.e.-convergence is now re-baptized as **a.s. convergence**, while convergence in measure is now called **convergence in probability**. We write $X_n \overset{a.s.}{\to} X$ if the sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges to a random variable X, a.s. Similarly, $X_n \overset{\mathbb{P}}{\to} X$ refers to convergence in probability. The notion of convergence in $\mathcal{L}^p$, for $p \in [1, \infty]$ is exactly the same as before. We write $X_n \overset{\mathcal{L}^p}{\to} X$ if $\{X_n\}_{n\in\mathbb{N}}$ converges to X in $\mathcal{L}^p$.

9. Since the constant random variable $X(\omega) = M$, for $\omega \in \Omega$ is integrable, a special case of the dominated convergence theorem,

[2] Random variables are usually denoted by capital letters such as X, Y Z, etc.

[3] If we interpret the knowledge of $\omega \in \Omega$ as the information about the true state of all parts of the model, $X(\omega)$ will typically correspond to a single numerical aspect ofi it.

known as the **bounded convergence theorem** holds in probability spaces:

Theorem 6.2 (Bounded convergence)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that there exists $M \geq 0$ such that $|X_n| \leq M$, a.s., and $X_n \to X$, a.s., then*

$$\mathbb{E}[X_n] \to \mathbb{E}[X].$$

10. The relationship between various forms of convergence can now be represented diagramatically as

$$\begin{tikzcd}
 & \mathcal{L}^\infty \arrow[dl] \arrow[dd] \arrow[dr] \arrow[drr] & & \\
\text{a.s.} \arrow[dr] & & \mathcal{L}^p \arrow[dl] & \mathcal{L}^q \arrow[l] \arrow[dll] \\
 & \mathbb{P} & &
\end{tikzcd}$$

where $1 \leq p \leq q < \infty$ and an arrow $A \to B$ means that A implies B, but that B *does not* imply A in general.

Distributions of random variables, vectors and elements

As we have already mentioned, Ω typically too big to be of direct use. Luckily, if we are only interested in a single random variable, all the useful probabilistic information about it is contained in the probabilities of the form[4] $\mathbb{P}[X \in B]$, for $B \in \mathcal{B}(\mathbb{R})$.

4 It is standard to write $\mathbb{P}[X \in B]$ instead of the more precise $\mathbb{P}[\{X \in B\}]$ or $\mathbb{P}[\{\omega \in \Omega : X(\omega) \in B\}]$. Similarly, we will write $\mathbb{P}[X_n \in B_n$, i.o] instead of $\mathbb{P}[\{X_n \in B_n\}$ i.o.] and $\mathbb{P}[X_n \in B_n$, ev.] instead of $\mathbb{P}[\{X_n \in B_n\}$ ev.]

The map $B \mapsto \mathbb{P}[X \in B]$ is, however, nothing but the push-forward of the measure $\mathbb{P}$ by the map X onto $\mathcal{B}(\mathbb{R})$:

Definition 6.3. The **distribution** of the random variable X is the probability measure μ_X on $\mathcal{B}(\mathbb{R})$, defined by

$$\mu_X(B) = \mathbb{P}[X^{-1}(B)],$$

that is the push-forward of the measure $\mathbb{P}$ by the map X.

In addition to be able to recover the information about various probabilities related to X from μ_X, one can evaluate any possible integral involving a function of X by integrating that function against μ_X (compare the statement to Problem 5.10):

Problem 6.1. Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel function. Then $g \circ X \in \mathcal{L}^{0-1}(\Omega, \mathcal{F}, \mathbb{P})$ if and only if $g \in \mathcal{L}^{0-1}(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ and, in that case,

$$\mathbb{E}[g(X)] = \int g \, d\mu_X.$$

In particular,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \mu_X(dx).$$

Taken in isolation from everything else, two random variables X and Y for which $\mu_X = \mu_Y$ are the same from the probabilistic point of view. In that case we say that X and Y are **equally distributed** and write $X \overset{(d)}{=} Y$. On the other hand, if we are interested in their relationship with a third random variable Z, it can happen that X and Y have the same distribution, but that their relationship to Z is very different. It is the notion of **joint distribution** that sorts such things out. For a random vector $\boldsymbol{X} = (X_1, \dots, X_n)$, the measure $\mu_{\boldsymbol{X}}$ on $\mathcal{B}(\mathbb{R}^n)$ given by

$$\mu_{\boldsymbol{X}}(B) = \mathbb{P}[\boldsymbol{X} \in B],$$

is called the **distribution** of the random vector $\boldsymbol{X}$. Clearly, the measure $\mu_{\boldsymbol{X}}$ contains the information about the distributions of the individual components $X_1, \dots, X_n$, because

$$\begin{aligned}\mu_{X_1}(A) = \mathbb{P}[X_1 \in A] &= \mathbb{P}[X_1 \in A, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R}] \\ &= \mu_{\boldsymbol{X}}(A \times \mathbb{R} \times \dots \times \mathbb{R}).\end{aligned}$$

When $X_1, \dots, X_n$ are viewed as components in the random vector $\boldsymbol{X}$, their distributions are sometimes referred to as **marginal distributions**.

Example 6.4. Let $\Omega = \{1, 2, 3, 4\}$, $\mathcal{F} = 2^{\Omega}$, with $\mathbb{P}$ characterized by $\mathbb{P}[\{\omega\}] = \frac{1}{4}$, for $\omega = 1, \dots, 4$. The map $X : \Omega \to \mathbb{R}$, given by $X(1) = X(3) = 0$, $X(2) = X(4) = 1$, is a random variable and its distribution is the measure $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ on $\mathcal{B}(\mathbb{R})$ (check that formally!), where δ_a denotes the Dirac measure on $\mathcal{B}(\mathbb{R})$, concentrated on $\{a\}$.

Similarly, the maps $Y : \Omega \to \mathbb{R}$ and $Z : \Omega \to \mathbb{R}$, given by $Y(1) = Y(2) = 0$, $Y(3) = Y(4) = 1$, and $Z(\omega) = 1 - X(\omega)$ are random variables with the same distribution as X. The joint distributions of the random vectors (X, Y) and (X, Z) are very different, though. The pair (X, Y) takes 4 different values $(0, 0), (0, 1), (1, 0), (1, 1)$, each with probability $\frac{1}{4}$, so that the distribution of (X, Y) is given by

$$\mu_{(X,Y)} = \tfrac{1}{4}\Big(\delta_{(0,0)} + \delta_{(0,1)} + \delta_{(1,0)} + \delta_{(1,1)}\Big).$$

On the other hand, it is impossible for X and Z to take the same value at the same time. In fact, there are only two values that the pair (X, Z) can take - $(0, 1)$ and $(1, 0)$. They happen with probability $\frac{1}{2}$ each, so

$$\mu_{(X,Z)} = \tfrac{1}{2}\Big(\delta_{(0,1)} + \delta_{(1,0)}\Big).$$

We will see later that the difference between (X, Y) and (X, Z) is best understood if we analyze the way the component random variables depend on each other. In the first case, even if the value of X is revealed, Y can still take the values 0 or 1 with equal probabilities. In the second case, as soon as we know X, we know Z.

More generally, if $X : \Omega \to S$, is a random element with values in the measurable space $(S, \mathcal{S})$, the **distribution of** X is the measure μ_X on $\mathcal{S}$, defined by $\mu_X(B) = \mathbb{P}[X \in B] = \mathbb{P}[X^{-1}(B)]$, for $B \in \mathcal{S}$.

Sometimes it is easier to work with a real-valued function F_X defined by

$$F_X(x) = \mathbb{P}[X \leq x],$$

which we call the **(cumulative) distribution function** (**cdf** for short), of the random variable[5] X. The following properties of F_X are easily derived by using continuity of probability from above and from below:

[5] A notion of a (cumulative) distribution function can be defined for random vectors, too, but it is not used as often as the single-component case, so we do not write about it here.

Proposition 6.5. *Let X be a random variable, and let F_X be its distribution function. Then,*

1. *F_X is non-decreasing and takes values in $[0, 1]$,*
2. *F_X is right continuous,*
3. $\lim_{x\to\infty} F_X(x) = 1$ *and* $\lim_{x\to-\infty} F_X(x) = 0$.

The case when μ_X is absolutely continuous with respect to the Lebesgue measure is especially important:

Definition 6.6. A random variable X with the property that $\mu_X \ll \lambda$, where λ is the Lebesgue measure on $\mathcal{B}(\mathbb{R})$, is said to be **absolutely continuous**.

In that case, any Radon-Nikodym derivative $\frac{d\mu_X}{d\lambda}$ is called the **probability density function (pdf)** of X, and is denoted by f_X. Similarly, a random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ is said to be **absolutely continuous** if $\mu_{\boldsymbol{X}} \ll \lambda$, where λ is the Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$, and the Radon-Nikodym derivative $\frac{d\mu_{\boldsymbol{X}}}{d\lambda}$, denoted by $f_{\boldsymbol{X}}$ is called the probability density function (pdf) of $\boldsymbol{X}$.

Problem 6.2.

1. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be an absolutely-continuous random vector. Show that X_k is absolutely continuous, and that its pdf is given by

$$f_{X_k}(x) = \underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{n-1 \text{ integrals}} f(\xi_1, \dots, \xi_{k-1}, x, \xi_{k+1}, \dots, \xi_n)\, d\xi_1 \dots d\xi_{k-1}\, d\xi_{k+1} \dots d\xi_n.$$

Note: As is should, $f_{X_k}(x)$ is defined only for almost all $x \in \mathbb{R}$; that is because $f_{\boldsymbol{X}}$ is defined only up to null sets in $\mathcal{B}(\mathbb{R}^n)$.

2. Let X be an absolutely-continuous random variable. Show that the random vector (X, X) is *not* absolutely continuous, even though both of its components are .

Problem 6.3. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be an absolutely-continuous random vector with density $f_{\boldsymbol{X}}$. For a Borel function $g : \mathbb{R}^n \to \mathbb{R}$ such that $g f_{\boldsymbol{X}} \in \mathcal{L}^{0-1}(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$, show that $g(\boldsymbol{X}) \in \mathcal{L}^{0-1}(\Omega, \mathcal{F}, \mathbb{P})$ and that

$$\mathbb{E}[g(\boldsymbol{X})] = \int g f_{\boldsymbol{X}} \, d\lambda = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} g(\xi_1, \dots, \xi_n) f_{\boldsymbol{X}}(\xi_1, \dots, \xi_n) \, d\xi_1 \dots d\xi_n.$$

Definition 6.7. A random variable X is said to be **discrete** if there exists a countable set $B \in \mathcal{B}(\mathbb{R})$ such that $\mu_X(B) = 1$.

Problem 6.4. Show that a sum of two discrete random variables is discrete, but that a sum of two absolutely-continuous random variables does not need to be absolutely continuous.

Definition 6.8. A distribution which has no atoms and is singular with respect to the Lebesgue measure is called **singular**.

Example 6.9. According to Problem 5.12, there exists a measure μ on $[0, 1]$, with the following properties

1. μ has no atoms, i.e., $\mu(\{x\}) = 0$, for all $x \in [0, 1]$,
2. μ and λ (the Lebesgue measure) are mutually singular
3. μ is supported by the Cantor set.

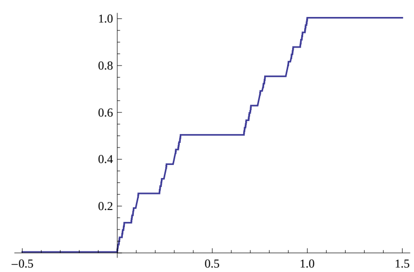


Figure 1: The CDF of the Cantor distribution.

We set $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \mu)$, and define the random variable $X : \Omega \to \mathbb{R}$, by $X(\omega) = \omega$. It is clear that the distribution μ_X of X has the property that

$$\mu_X(B) = \mu(B \cap [0, 1]),$$

Thus, X is a random variable with a singular distribution.

Independence

The point at which probability departs from measure theory is when independence is introduced. As seen in Example 6.4, two random variables can "depend" on each other in different ways. One extreme (the case of X and Y) corresponds to the case when the dependence is very weak - the distribution of Y stays the same when the value of X is revealed:

Definition 6.10. Two random variables X and Y are said to be **independent** if

$$\mathbb{P}[\{X \in A\} \cap \{Y \in B\}] = \mathbb{P}[X \in A] \times \mathbb{P}[Y \in B] \text{ for all } A, B \in \mathcal{B}(\mathbb{R}).$$

It turns out that independence of random variables is a special case of the more-general notion of independence between families of sets.

Definition 6.11. Families $\mathcal{A}_1, \dots, \mathcal{A}_n$ of elements in $\mathcal{F}$ are said to be

1. **independent** if

$$\mathbb{P}[A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}] = \mathbb{P}[A_{i_1}] \times \mathbb{P}[A_{i_2}] \times \dots \times \mathbb{P}[A_{i_k}], \tag{6.1}$$

   for all $k = 1, \dots, n$, $1 \leq i_1 < i_2 < \dots < i_k \leq n$, and all $A_{i_l} \in \mathcal{A}_{i_l}$, $l = 1, \dots, k$,

2. **pairwise independent** if

$$\mathbb{P}[A_{i_1} \cap A_{i_2}] = \mathbb{P}[A_{i_1}] \times \mathbb{P}[A_{i_2}],$$

   for all $1 \leq i_1 < i_2 \leq n$, and all $A_{i_1} \in \mathcal{A}_{i_1}$, $A_{i_2} \in \mathcal{A}_{i_2}$.

Problem 6.5.

1. Show, by means of an example, that the notion of independence would change if we asked for the product condition (6.1) to hold only for $k = n$ and $i_1 = 1, \dots, i_k = n$.

2. Show that, however, if $\Omega \in \mathcal{A}_i$, for all $i = 1, \dots, n$, it is enough to test (6.1) for $k = n$ and $i_1 = 1, \dots, i_k = n$ to conclude independence of $\mathcal{A}_i$, $i = 1, \dots, n$.

Problem 6.6. Show that random variables X and Y are independent if and only if the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent.

Definition 6.12. Random variables $X_1, \dots, X_n$ are said to be **independent** if the σ-algebras $\sigma(X_1), \dots, \sigma(X_n)$ are independent. Events $A_1, \dots, A_n$ are called **independent** if the families $\mathcal{A}_i = \{A_i\}$, $i = 1, \dots, n$, are independent.

When only two families of sets are compared, there is no difference between pairwise independence and independence. For 3 or more, the difference is non-trivial:

Example 6.13. Let X_1, X_2, X_3 be independent random variables, each with the **coin-toss** distribution, i.e., $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$, for

$i = 1, 2, 3$. It is not hard to construct a probability space where such random variables may be defined explicitly: let $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $\mathcal{F} = 2^\Omega$, and let $\mathbb{P}$ be characterized by $\mathbb{P}[\{\omega\}] = \frac{1}{8}$, for all $\omega \in \Omega$. Define

$$X_i(\omega) = \begin{cases} 1, & \omega \in \Omega_i \\ -1, & \text{otherwise} \end{cases}$$

where $\Omega_1 = \{1, 3, 5, 7\}$, $\Omega_2 = \{2, 3, 6, 7\}$ and $\Omega_3 = \{5, 6, 7, 8\}$. It is easy to check that X_1, X_2 and X_3 are independent (X_i is the "i-th bit" in the binary representation of ω).

With X_1, X_2 and X_3 defined, we set

$$Y_1 = X_2 X_3,\ Y_2 = X_1 X_3 \text{ and } Y_3 = X_1 X_2,$$

so that Y_i has a coin-toss distribution, for each $i = 1, 2, 3$. Let us show that Y_1 and Y_2 (and then, by symmetry, Y_1 and Y_3, as well as Y_2 and Y_3) are independent:

$$\begin{aligned} \mathbb{P}[Y_1 = 1, Y_2 = 1] &= \mathbb{P}[X_2 = X_3, X_1 = X_3] = \mathbb{P}[X_1 = X_2 = X_3] \\ &= \mathbb{P}[X_1 = X_2 = X_3 = 1] + \mathbb{P}[X_1 = X_2 = X_3 = -1] \\ &= \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{1}{4} = \mathbb{P}[Y_1 = 1] \times \mathbb{P}[Y_2 = 1]. \end{aligned}$$

We don't need to check the other possibilities, such as $Y_1 = 1$, $Y_2 = -1$, to conclude that Y_1 and Y_2 are independent (see Problem 6.7 below).

On the other hand, Y_1, Y_2 and Y_3 are not independent:

$$\begin{aligned} \mathbb{P}[Y_1 = 1, Y_2 = 1, Y_3 = 1] &= \mathbb{P}[X_2 = X_3, X_1 = X_3, X_1 = X_2] \\ &= \mathbb{P}[X_1 = X_2 = X_3] = \tfrac{1}{4} \\ &\neq \tfrac{1}{8} = \mathbb{P}[Y_1 = 1] \times \mathbb{P}[Y_2 = 1] \times \mathbb{P}[Y_3 = 1]. \end{aligned}$$

Problem 6.7. Show that if $A_1, \dots, A_n$ are independent, then so are the families $\mathcal{A}_i = \{A_i, A_i^c\}$, $i = 1, \dots, n$.

A more general statement is also true (and very useful):

Proposition 6.14. *Let $\mathcal{P}_i$, $i = 1, \dots, n$ be independent π-systems. Then, the σ-algebras $\sigma(\mathcal{P}_i)$, $i = 1, \dots, n$ are also independent.*

Proof. Let $\mathcal{F}_1$ denote the set of all $C \in \mathcal{F}$ such that

$$\mathbb{P}[C \cap A_{i_2} \cap \dots \cap A_{i_k}] = \mathbb{P}[C] \times \mathbb{P}[A_{i_2}] \times \dots \times \mathbb{P}[A_{i_k}],$$

for all $k = 2, \dots, n$, $1 < i_2 < \dots < i_k \leq n$, and all $A_{i_l} \in \mathcal{P}_{i_l}$, $l = 2, \dots, k$. It is easy to see that $\mathcal{F}_1$ is a λ-system which contains the π-system $\mathcal{P}_1$, and so, by the π-λ Theorem, it also contains $\sigma(\mathcal{P}_1)$. Consequently $\sigma(\mathcal{P}_1), \mathcal{P}_2, \dots \mathcal{P}_n$ are independent families.

A re-play of the whole procedure, but now with families $\mathcal{P}_2$, $\sigma(\mathcal{P}_1)$, $\mathcal{P}_3, \dots \mathcal{P}_n$, yields that the families $\sigma(\mathcal{P}_1), \sigma(\mathcal{P}_2), \mathcal{P}_3, \dots \mathcal{P}_n$ are also independent. Following the same pattern allows us to conclude after n steps that $\sigma(\mathcal{P}_1), \sigma(\mathcal{P}_2), \dots \sigma(\mathcal{P}_n)$ are independent. $\square$

Remark 6.15. All notions of independence above extend to infinite families of objects (random variables, families of sets) by requiring that every finite sub-family be independent.

The result of Proposition 6.14 can be used to help us check independence of random variables:

Problem 6.8. Let $X_1, \dots, X_n$ be random variables.

1. Show that $X_1, \dots, X_n$ are independent if and only if

$$\mu_{\boldsymbol{X}} = \mu_{X_1} \otimes \dots \otimes \mu_{X_n},$$

   where $\boldsymbol{X} = (X_1, \dots, X_n)$.

2. Show that $X_1, \dots, X_n$ are independent if and only if

$$\mathbb{P}[X_1 \leq x_1, \dots, X_n \leq x_n] = \mathbb{P}[X_1 \leq x_1] \times \dots \times \mathbb{P}[X_n \leq x_n],$$

   for all $x_1, \dots, x_n \in \mathbb{R}$.

Note: The family $\{\{X_i \leq x\} : x \in \mathbb{R}\}$ does not include Ω, so that part (2) of Problem 6.5 cannot be applied directly.

3. Suppose that the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ is absolutely continuous. Then $X_1, \dots, X_n$ are independent if and only if

$$f_{\boldsymbol{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n), \ \lambda\text{-a.e.},$$

   where λ denotes the Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$.

4. Suppose that $X_1, \dots X_n$ are discrete with $\mathbb{P}[X_k \in C_k] = 1$, for countable subsets $C_1, \dots, C_n$ of $\mathbb{R}$. Show that $X_1, \dots, X_n$ are independent if and only if

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n] = \mathbb{P}[X_1 = x_1] \times \dots \times \mathbb{P}[X_n = x_n],$$

   for all $x_i \in C_i$, $i = 1, \dots, n$.

Problem 6.9. Let $X_1, \dots, X_n$ be independent random variables. Show that the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ is absolutely continuous if and only if each X_i, $i = 1, \dots, n$ is an absolutely-continuous random variable.

The usefulness of Proposition 6.14 is not exhausted, yet.

Problem 6.10.

1. Let $\mathcal{F}_{ij}$ $i = 1, \dots, n$, $j = 1, \dots, m_i$, be an independent collection of σ-algebras on Ω. Show that the σ-algebras $\mathcal{G}_1, \dots, \mathcal{G}_n$, where $\mathcal{G}_i = \sigma(\mathcal{F}_{i1}, \dots, \mathcal{F}_{im_i})$, are independent.
2. Let X_{ij} $i = 1, \dots, n$, $j = 1, \dots, m_i$, be an independent random variables, and let $f_i : \mathbb{R}^{m_i} \to \mathbb{R}$, $i = 1, \dots, n$, be Borel functions. Then the random variables $Y_i = f_i(X_{i\,1}, \dots, X_{i\,m_i})$, $i = 1, \dots, n$ are independent.

Hint: $\cup_{j=1,\dots,m_i} \mathcal{F}_{ij}$ generates $\mathcal{G}_i$, but is not quite a π-system.

Problem 6.11.

1. Let $X_1, \dots, X_n$ be random variables. Show that $X_1, \dots, X_n$ are independent if and only if
$$\prod_{i=1}^{n} \mathbb{E}[f_i(X_i)] = \mathbb{E}[\prod_{i=1}^{n} f_i(X_i)],$$
for all n-tuples $(f_1, \dots, f_n)$ of bounded continuous real functions.
2. Let $\{X_n^i\}_{n\in\mathbb{N}}$, $i = 1, \dots, m$ be sequences of random variables such that $X_n^1, \dots, X_n^m$ are independent for each $n \in \mathbb{N}$. If $X_n^i \overset{a.s.}{\to} X^i$, $i = 1, \dots, m$, for some $X^1, \dots, X^m \in \mathcal{L}^0$, show that $X^1, \dots, X^m$ are independent.

Hint: Approximate!

The idea "independent means multiply" applies not only to probabilities, but also to random variables:

Proposition 6.16. *Let X, Y be independent random variables, and let $h : \mathbb{R}^2 \to [0, \infty)$ be a measurable function. Then*
$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x, y)\, \mu_X(dx) \right) \mu_Y(dy).$$

Proof. By independence and part 1. of Problem 6.8, the distribution of the random vector (X, Y) is given by $\mu_X \otimes \mu_Y$, where μ_X is the distribution of X and μ_Y is the distribution of μ_Y. Using Fubini's theorem, we get
$$\mathbb{E}[h(X, Y)] = \int h\, d\mu_{(X,Y)} = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x, y)\, \mu_X(dx) \right) \mu_Y(dy). \qquad \square$$

Proposition 6.17. *Let $X_1, X_2, \dots, X_n$ be independent random variables with $X_i \in \mathcal{L}^1$, for $i = 1, \dots, n$. Then*

1. $\prod_{i=1}^{n} X_i = X_1 \cdots X_n \in \mathcal{L}^1$, *and*

2. $\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \dots \mathbb{E}[X_n]$.

The product formula 2. remains true if we assume that $X_i \in \mathcal{L}^0_+$ (instead of $\mathcal{L}^1$), for $i = 1, \dots, n$.

Proof. Using the fact that X_1 and $X_2 \cdots X_n$ are independent random variables (use part 2. of Problem 6.10), we can assume without loss of generality that $n = 2$.

Focusing first on the case $X_1, X_2 \in \mathcal{L}^0_+$, we apply Proposition 6.16 with $h(x, y) = xy$ to conclude that

$$\begin{aligned} \mathbb{E}[X_1 X_2] &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x_1 x_2 \, \mu_{X_1}(dx_1) \right) \mu_{X_2}(dx_2) \\ &= \int_{\mathbb{R}} x_2 \mathbb{E}[X_1] \, \mu_{X_2}(dx_2) = \mathbb{E}[X_1]\mathbb{E}[X_2]. \end{aligned}$$

For the case $X_1, X_2 \in \mathcal{L}^1$, we split $X_1 X_2 = X_1^+ X_2^+ - X_1^+ X_2^- - X_1^- X_2^+ + X_1^- X_2^-$ and apply the above conclusion to the 4 pairs $X_1^+ X_2^+$, $X_1^+ X_2^-$, $X_1^- X_2^+$ and $X_1^- X_2^-$. □

Problem 6.12 (Conditions for "independent-means-multiply")**.** Proposition 6.17 states that for independent X and Y, we have

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \tag{6.2}$$

whenever *both* $X, Y \in \mathcal{L}^1$ or *both* $X, Y \in \mathcal{L}^0_+$. Give an example which shows that (6.2) is no longer necessarily true in general if $X \in \mathcal{L}^0_+$ and $Y \in \mathcal{L}^1$.

Hint: Build your example so that $\mathbb{E}[(XY)^+] = \mathbb{E}[(XY)^-] = \infty$. Use $([0,1], \mathcal{B}([0,1]), \lambda)$ and take $Y(\omega) = \mathbf{1}_{[0,1/2]}(\omega) - \mathbf{1}_{(1/2,0]}(\omega)$. Then show that any random variable X with the property that $X(\omega) = X(1-\omega)$ is independent of Y.

Problem 6.13. Two random variables X, Y are said to be uncorrelated, if $X, Y \in \mathcal{L}^2$ and $\mathrm{Cov}(X, Y) = 0$, where $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

1. Show that for $X, Y \in \mathcal{L}^2$, the expression for $\mathrm{Cov}(X, Y)$ is well defined.
2. Show that independent random variables in $\mathcal{L}^2$ are uncorrelated.
3. Show that there exist uncorrelated random variables which are not independent.

Sums of independent random variables and convolution

Proposition 6.18. *Let X and Y be independent random variables, and let $Z = X + Y$ be their sum. Then the distribution μ_Z of Z has the following representation:*

$$\mu_Z(B) = \int_{\mathbb{R}} \mu_X(B - y) \, \mu_Y(dy), \text{ for } B \in \mathcal{B}(\mathbb{R}),$$

where $B - y = \{b - y : b \in B\}$.

Proof. We can view Z as a function $f(x,y) = x + y$ applied to the random vector (X,Y), and so, we have $\mathbb{E}[g(Z)] = \mathbb{E}[h(X,Y)]$, where $h(x,y) = g(x+y)$. In particular, for $g(z) = \mathbf{1}_B(z)$, Proposition 6.16 implies that

$$\mu_Z(B) = \mathbb{E}[g(Z)] = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{\{x+y\in B\}} \, \mu_X(dx) \, \mu_Y(dy) =$$
$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{\{x\in B-y\}} \mu_X(dx) \right) \mu_Y(dy) = \int_{\mathbb{R}} \mu_X(B-y) \, \mu_Y(dy). \quad \square$$

One often sees the expression

$$\int_{\mathbb{R}} f(x) \, dF(x),$$

as notation for the integral $\int f \, d\mu$, where $F(x) = \mu((-\infty, x])$. The reason for this is that such integrals - called the **Lebesgue-Stieltjes** integrals - have a theory parallel to that of the Riemann integral and the correspondence between $dF(x)$ and $d\mu$ is parallel to the correspondence between dx and $d\lambda$.

Corollary 6.19. *Let X, Y be independent random variables, and let Z be their sum. Then*

$$F_Z(z) = \int_{\mathbb{R}} F_X(z-y) \, dF_Y(y).$$

Definition 6.20. Let μ_1 and μ_2 be two probability measures on $\mathcal{B}(\mathbb{R})$. The **convolution** of μ_1 and μ_2 is the probability measure $\mu_1 * \mu_2$ on $\mathcal{B}(\mathbb{R})$, given by

$$(\mu_1 * \mu_2)(B) = \int_{\mathbb{R}} \mu_1(B-\xi) \, \mu_2(d\xi), \text{ for } B \in \mathcal{B}(\mathbb{R}),$$

where $B - \xi = \{x - \xi : x \in B\} \in \mathcal{B}(\mathbb{R})$.

Problem 6.14. Show that $*$ is a commutative and associative operation on the set of all probability measures on $\mathcal{B}(\mathbb{R})$.

Hint: Use Proposition 6.18

It is interesting to see how convolution relates to absolute continuity. To simplify the notation, we write $\int_A f(x) \, dx$ instead of (the more precise) $\int_A f(x) \, \lambda(dx)$ for the (Lebesgue) integral with respect to the Lebesgue measure on $\mathbb{R}$. When $A = [a,b] \in \bar{\mathbb{R}}$, we write $\int_a^b f(x) \, dx$.

Proposition 6.21. *Let X and Y be independent random variables, and suppose that X is absolutely continuous. Then their sum $Z = X + Y$ is also absolutely continuous and its density f_Z is given by*

$$f_Z(z) = \int_{\mathbb{R}} f_X(z-y) \, \mu_Y(dy).$$

Proof. Define $f(z) = \int_{\mathbb{R}} f_X(z-y)\,\mu_Y(dy)$, for some density f_X of X (remember, the density function is defined only λ-a.e.). The function f is measurable (why?) so it will be enough (why?) to show that

$$\mathbb{P}\Big[Z \in [a,b]\Big] = \int_{[a,b]} f(z)\,dz, \text{ for all } -\infty < a < b < \infty. \tag{6.2}$$

We start with the right-hand side of (6.2) and use Fubini's theorem to obtain

$$\begin{aligned}\int_{[a,b]} f(z)\,dz &= \int_{\mathbb{R}} \mathbf{1}_{[a,b]}(z)\left(\int_{\mathbb{R}} f_X(z-y)\,\mu_Y(dy)\right)\,dz \\ &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} \mathbf{1}_{[a,b]}(z) f_X(z-y)\,dz\right)\,\mu_Y(dy)\end{aligned} \tag{6.3}$$

By the translation-invariance property of the Lebesgue measure, we have

$$\begin{aligned}\int_{\mathbb{R}} \mathbf{1}_{[a,b]}(z) f_X(z-y)\,dz &= \int_{\mathbb{R}} \mathbf{1}_{[a-y,b-y]}(z) f_X(z)\,dz \\ &= \mathbb{P}\Big[X \in [a-y,b-y]\Big] = \mu_X\Big([a,b]-y\Big).\end{aligned}$$

Therefore, by (6.3) and Proposition 6.18, we have

$$\begin{aligned}\int_{[a,b]} f(z)\,dz &= \int_{\mathbb{R}} \mu_X\Big([a,b]-y\Big)\,\mu_Y(dy) \\ &= \mu_Z\Big([a,b]\Big) = \mathbb{P}\Big[Z \in [a,b]\Big].\end{aligned}$$ □

Definition 6.22. The **convolution** of functions f and g in $\mathcal{L}^1(\mathbb{R})$ is the function $f * g \in \mathcal{L}^1(\mathbb{R})$ given by

$$(f*g)(z) = \int_{\mathbb{R}} f(z-x) g(x)\,dx.$$

Problem 6.15.

1. Use the reasoning from the proof of Proposition 6.21 to show that the convolution is well-defined operation on $\mathcal{L}^1(\mathbb{R})$.
2. Show that if X and Y are independent random variables and X is absolutely-continuous, then $X+Y$ is also absolutely continuous.

Do independent random variables exist?

We leave the most basic of the questions about independence for last: do independent random variable exist? We need a definition and two auxiliary results, first.

Definition 6.23. A random variable X is said to be **uniformly distributed on** (a,b), for $a < b \in \mathbb{R}$, if it is absolutely continuous with density

$$f_X(x) = \tfrac{1}{b-a}\mathbf{1}_{(a,b)}(x).$$

Our first result states a uniform random variable on $(0,1)$ can be transformed deterministically into any a random variable of prescribed distribution (cdf).

Proposition 6.24. *Let μ be a measure on $\mathcal{B}(\mathbb{R})$ with $\mu(\mathbb{R}) = 1$. Then, there exists a function $H_\mu : (0,1) \to \mathbb{R}$ such that the distribution of the random variable $X = H_\mu(U)$ is μ, whenever U is a uniform random variable on $(0,1)$.*

Note: this proposition is a basis for a technique used to simulate random variables. There are efficient algorithms for producing simulated values which resemble the uniform distribution in $(0,1)$ (so-called **pseudo-random numbers**). If a simulated value drawn with distribution μ is needed, one can simply apply the function H_μ to a pseudo-random number.

Proof. Let

F be the cdf corresponding to μ, i.e.,

$$F(x) = \mu((-\infty, x]).$$

The function F is non-decreasing, so it "almost" has an inverse: define

$$H_\mu(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}.$$

Since $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$, $H_\mu(y)$ is well-defined and finite for all $y \in (0,1)$. Moreover, thanks to right-continuity and non-decrease of F, we have

$$H_\mu(y) \leq x \Leftrightarrow y \leq F(x), \text{ for all } x \in \mathbb{R}, y \in (0,1).$$

Therefore

$$\mathbb{P}[H_\mu(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x), \text{ for all } x \in \mathbb{R},$$

and the statement of the Proposition follows. □

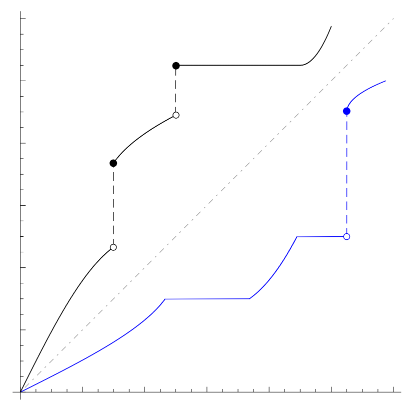


Figure 2: The right-continuous inverse H_μ (blue) of the CDF F (black)

Our next auxiliary result tells us how to construct a sequence of independent uniforms:

Proposition 6.25. *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and on it a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables such that*

1. *X_n has the uniform distribution on $(0,1)$, for each $n \in \mathbb{N}$, and*
2. *the sequence $\{X_n\}_{n\in\mathbb{N}}$ is independent.*

Proof. Set $(\Omega, \mathcal{F}, \mathbb{P}) = (\{-1,1\}^{\mathbb{N}}, \mathcal{S}, \mu_C)$ - the coin-toss space with the product σ-algebra and the coin-toss measure. Let $a : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be a bijection, i.e., $(a_{ij})_{i,j\in\mathbb{N}}$ is an arrangement of all natural numbers into a double array. For $i, j \in \mathbb{N}$, we define the map $\xi_{ij} : \Omega \to \{-1,1\}$, by

$$\xi_{ij}(s) = s_{a_{ij}},$$

i.e., ξ_{ij} is the natural projection onto the a_{ij}-th coordinate. It is straightforward to show that, under $\mathbb{P}$, the collection $(\xi_{ij})_{i,j\in\mathbb{N}}$ is independent; indeed, it is enough to check the equality

$$\mathbb{P}[\xi_{i_1 j_1} = 1, \dots, \xi_{i_n j_n} = 1] = \mathbb{P}[\xi_{i_1 j_1} = 1] \times \dots \times \mathbb{P}[\xi_{i_n j_n} = 1],$$

for all $n \in \mathbb{N}$ and all different $(i_1, j_1), \ldots, (i_n, j_n) \in \mathbb{N} \times \mathbb{N}$.

At this point, we recycle the idea we used to construct the Lebesgue measure to construct an independent copy of a uniformly-distributed random variable from each row of $(\xi_{ij})_{i,j \in \mathbb{N}}$. We set

$$X_i = \sum_{j=1}^{\infty} \left(\tfrac{1+\xi_{ij}}{2} \right) 2^{-j}, \; i \in \mathbb{N}. \tag{6.4}$$

By second parts of Problems 6.10 and 6.11, we conclude that the sequence $\{X_i\}_{i \in \mathbb{N}}$ is independent. Moreover, thanks to (6.4), X_i is uniform on $(0,1)$, for each $i \in \mathbb{N}$. □

Proposition 6.26. *Let $\{\mu_n\}_{n \in \mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(\mathbb{R})$. Then, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a sequence $\{X_n\}_{n \in \mathbb{N}}$ of random variables defined there such that*

1. *$\mu_{X_n} = \mu_n$, and*
2. *$\{X_n\}_{n \in \mathbb{N}}$ is independent.*

Proof. Start with the sequence of Proposition 6.25 and apply the function H_{μ_n} to X_n for each $n \in \mathbb{N}$, where H_{μ_n} is as in the proof of Proposition 6.24. □

An important special case covered by Proposition 6.26 is the following:

Definition 6.27. A sequence $\{X_n\}_{n \in \mathbb{N}}$ of random variables is said to be **independent and identically distributed (iid)** if $\{X_n\}_{n \in \mathbb{N}}$ is independent and all X_n have the same distribution.

Corollary 6.28. *Given a probability measure μ on $\mathbb{R}$, there exist a probability space supporting an iid sequence $\{X_n\}_{n \in \mathbb{N}}$ such that $\mu_{X_n} = \mu$.*

Additional Problems

Problem 6.16 (The standard normal distribution)**.** An absolutely continuous random variable X is said to have the **standard normal distribution** - denoted by $X \sim N(0,1)$ - if it admits a density of the form

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2), \; x \in \mathbb{R}$$

For a r.v. with such a distribution we write $X \sim N(0,1)$.

1. Show that $\int_{\mathbb{R}} f(x)\,dx = 1$.
2. For $X \sim N(0,1)$, show that $\mathbb{E}[|X|^n] < \infty$ for all $n \in \mathbb{N}$. Then compute the n^{th} moment $\mathbb{E}[X^n]$, for $n \in \mathbb{N}$.

Hint: Consider the double integral $\int_{\mathbb{R}^2} f(x)f(y)\,dx\,dy$ and pass to polar coordinates.

3. A random variable with the same distribution as X^2, where $X \sim N(0,1)$, is said to have the χ^2**-distribution**. Find an explicit expression for the density of the χ^2-distribution.

4. Let Y have the χ^2-distribution. Show that there exists a constant $\lambda_0 > 0$ such that $\mathbb{E}[\exp(\lambda Y)] < \infty$ for $\lambda < \lambda_0$ and $\mathbb{E}[\exp(\lambda Y)] = +\infty$ for $\lambda \geq \lambda_0$.

Note: For a random variable $Y \in \mathcal{L}^0_+$, the quantity $\mathbb{E}[\exp(\lambda Y)]$ is called the **exponential moment of order** λ.

5. Let $\alpha_0 > 0$ be a fixed, but arbitrary constant. Find an example of a random variable $X \geq 0$ with the property that $\mathbb{E}[X^\alpha] < \infty$ for $\alpha \leq \alpha_0$ and $\mathbb{E}[X^\alpha] = +\infty$ for $\alpha > \alpha_0$.

Hint: This is not the same situation as in 4. - this time the critical case α_0 is included in a different alternative. Try $X = \exp(Y)$, where $\mathbb{P}[Y \in \mathbb{N}] = 1$.

Problem 6.17 (The "memory-less" property of the exponential distribution)**.** A random variable is said to have **exponential distribution** with parameter $\lambda > 0$ - denoted by $X \sim \text{Exp}(\lambda)$ - if its distribution function F_X is given by

$$F_X(x) = 0 \text{ for } x < 0, \text{ and } F_X(x) = 1 - \exp(-\lambda x), \text{ for } x \geq 0.$$

1. Compute $\mathbb{E}[X^\alpha]$, for $\alpha \in (-1, \infty)$. Combine your result with the result of part 3. of Problem 6.16 to show that

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi},$$

where Γ is the Gamma function.

2. Remember that the conditional probability $\mathbb{P}[A|B]$ of A, given B, for $A, B \in \mathcal{F}$, $\mathbb{P}[B] > 0$ is given by

$$\mathbb{P}[A|B] = \mathbb{P}[A \cap B]/\mathbb{P}[B].$$

Compute $\mathbb{P}[X \geq x_2 | X \geq x_1]$, for $x_2 > x_1 > 0$ and compare it to $\mathbb{P}[X \geq (x_2 - x_1)]$.

conversely, suppose that y is a random variable with the property that $\mathbb{P}[Y > 0] = 1$ and $\mathbb{P}[Y > y] > 0$ for all $y > 0$. Assume further that

$$\mathbb{P}[Y \geq y_2 | Y \geq y_1] = \mathbb{P}[Y \geq y_2 - y_1], \text{ for all } y_2 > y_1 > 0. \quad (6.5)$$

Show that $Y \sim \text{Exp}(\lambda)$ for some $\lambda > 0$.

Note: This can be interpreted as follows: the knowledge that the bulb stayed functional until x_1 does not change the probability that it will not explode in the next $x_2 - x_1$ units of time; bulbs have no memory.

Hint: You can use the following fact: let $\phi : (0,\infty) \to \mathbb{R}$ be a Borel-measurable function such that $\phi(y) + \phi(z) = \phi(y+z)$ for all $y, z > 0$. Then there exists a constant $\mu \in \mathbb{R}$ such that $\phi(y) = \mu y$ for all $y > 0$.

Problem 6.18 (Some extensions of the Borel-Cantelli Lemma)**.**

1. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0_+$. Show that there exists a sequence of positive constants $\{c_n\}_{n\in\mathbb{N}}$ with the property that

$$\frac{X_n}{c_n} \to 0, \text{ a.s.}$$

Hint: Use the Borel-Cantelli lemma.

2. (The first) Borel-Cantelli lemma states that $\sum_{n\in\mathbb{N}} \mathbb{P}[A_n] < \infty$ implies $\mathbb{P}[A_n, \text{ i.o.}] = 0$. There are simple examples showing that the converse does not hold in general. Show that it *does* hold if the events $\{A_n\}_{n\in\mathbb{N}}$ are assumed to be independent. More precisely, show that, for an independent sequence $\{A_n\}_{n\in\mathbb{N}}$, we have

$$\sum_{n\in\mathbb{N}} \mathbb{P}[A_n] = \infty \text{ implies } \mathbb{P}[A_n, \text{ i.o.}] = 1.$$

This is often known as the **second Borel-Cantelli lemma**.

Hint: Use the inequality $(1-x) \leq e^{-x}$, $x \in \mathbb{R}$.

3. Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of events.
   (a) Show that $(\limsup A_n) \cap (\limsup A_n^c) \subseteq \limsup(A_n \cap A_{n+1}^c)$.
   (b) If $\liminf_{n\to\infty} \mathbb{P}[A_n] = 0$ and $\sum_n \mathbb{P}[A_n \cap A_{n+1}^c] < \infty$, show that

$$\mathbb{P}[\limsup_n A_n] = 0.$$

4. Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence in $\mathcal{L}^0$. Show that

$$\mathbb{E}[|X_1|] < \infty \text{ if and only if } \mathbb{P}[|X_n| \geq n, \text{ i.o.}] = 0.$$

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 7
Weak Convergence

The definition

In addition to the modes of convergence we introduced so far (a.s.-convergence, convergence in probability and $\mathcal{L}^p$-convergence), there is another one, called **weak convergence** or **convergence in distribution**. Unlike the other three, whether a sequence of random variables (elements) converges in distribution or not depends only on their *distributions*. In addition to its intrinsic mathematical interest, convergence in distribution (or, equivalently, the weak convergence) is precisely the kind of convergence we encounter in the central limit theorem.

We take the abstraction level up a notch and consider sequences of probability measures on $(S, \mathcal{S})$, where (S, d) is a metric space and $\mathcal{S} = \mathcal{B}(d)$ is the Borel σ-algebra there. In fact, it will always be assumed that S is a metric space and $\mathcal{S}$ is the Borel σ-algebra on it, throughout this chapter.

Definition 7.1. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $(S, \mathcal{S})$. We say that μ_n converges **weakly**[1] to a probability measure μ on $(S, \mathcal{S})$ - and write $\mu_n \stackrel{w}{\to} \mu$ - if

$$\int f\, d\mu_n \to \int f\, d\mu,$$

for all $f \in C_b(S)$, where $C_b(S)$ denotes the set of all continuous and bounded functions $f : S \to \mathbb{R}$.

[1] It would be more in tune with standard mathematical terminology to use the term *weak-$*$* convergence instead of weak convergence. For historical reasons, however, we omit the $*$.

Definition 7.2. A sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables (elements) is said to **converge in distribution** to the random variable (element) X, denoted by $X_n \stackrel{\mathcal{D}}{\to} X$, if $\mu_{X_n} \stackrel{w}{\to} \mu_X$.

For the uniqueness of limits, we need a simple approximation result:

Problem 7.1. Let F be a closed set in S. Show that for any $\varepsilon > 0$ there exists a Lipschitz and bounded function $f_{F;\varepsilon} : S \to \mathbb{R}$ such that

1. $0 \leq f_{F;\varepsilon}(x) \leq 1$, for all $x \in \mathbb{R}$,

2. $f_{F;\varepsilon}(x) = 1$ for $x \in F$, and

3. $f_{F;\varepsilon}(x) = 0$ for $d(x, F) \geq \varepsilon$, where $d(x, F) = \inf\{d(x, y) : y \in F\}$.

Hint: Show first that the function $x \mapsto d(x, F)$ is Lipschitz. Then argue that $f_{F;\varepsilon}(x) = h(d(x, F))$ has the required properties for a well-chosen function $h : [0, \infty) \to [0, 1]$.

Proposition 7.3. *Suppose that $\{\mu_n\}_{n\in\mathbb{N}}$ is a sequence of probability measures on $(S, \mathcal{S})$ such that $\mu_n \xrightarrow{w} \mu$ and $\mu_n \xrightarrow{w} \mu'$. Then $\mu = \mu'$.*

Proof. By the very definition of weak convergence, we have

$$\int f \, d\mu = \lim_n \int f \, d\mu_n = \int f \, d\mu', \tag{7.1}$$

for all $f \in C_b(S)$. Let F be a closed set, and let $\{f_k\}_{k\in\mathbb{N}}$ be as in Problem 7.1, with $f_k = f_{F;\varepsilon}$ corresponding to $\varepsilon = 1/k$. If we set $F_k = \{x \in S : d(x, F) \leq 1/k\}$, then F_k is a closed set (why?) and we have $\mathbf{1}_F \leq f_k \leq \mathbf{1}_{F_k}$. By (7.1), we have

$$\mu(F) \leq \int f_k \, d\mu = \int f_k d\mu' \leq \mu'(F_k),$$

and, similarly, $\mu'(F) \leq \mu(F_k)$, for all $k \in \mathbb{N}$. Since $F_k \searrow F$ (why?), we have $\mu(F_k) \searrow \mu(F)$ and $\mu'(F_k) \searrow \mu'(F)$, and it follows that $\mu(F) = \mu'(F)$.

It remains to note that the family of all closed sets is a π-system which generates the σ-algebra $\mathcal{S}$ to conclude that $\mu = \mu'$. □

Out next task if to give a useful operational characterization of weak convergence. Before we do that, we need a simple observation; remember that ∂A denotes the topological boundary $\partial A = \operatorname{Cl} A \setminus \operatorname{Int} A$ of a set $A \subseteq S$.

Problem 7.2. Let $(F_\gamma)_{\gamma\in\Gamma}$ be a partition of S into (possibly uncountably many) measurable subsets. Show that for any probability measure μ on $\mathcal{S}$, $\mu(F_\gamma) = 0$, for all but countably many $\gamma \in \Gamma$

Hint: For $n \in \mathbb{N}$, define $\Gamma_n = \{\gamma \in \Gamma : \mu(F_\gamma) \geq \frac{1}{n}\}$. Argue that Γ_n has at most n elements.

Definition 7.4. A set $A \in \mathcal{S}$ with the property that $\mu(\partial A) = 0$, is called a μ**-continuity set**

Theorem 7.5 (Portmanteau Theorem)**.** *Let μ, $\{\mu_n\}_{n\in\mathbb{N}}$ be probability measures on $\mathcal{S}$. Then, the following are equivalent:*

1. $\mu_n \xrightarrow{w} \mu$,

2. $\int f \, d\mu_n \to \int f \, d\mu$, *for all bounded,* Lipschitz *continuous* $f : S \to \mathbb{R}$,

3. $\limsup_n \mu_n(F) \leq \mu(F)$, *for all closed* $F \subseteq S$,
4. $\liminf_n \mu_n(G) \geq \mu(G)$, *for all open* $G \subseteq S$,
5. $\lim_n \mu_n(A) = \mu(A)$, *for all* μ*-continuity sets* $A \in \mathcal{S}$.

Note: Here is a way to remember whether closed sets go together with the lim inf or the lim sup: take a convergent sequence $\{x_n\}_{n\in\mathbb{N}}$ in S, with $x_n \to x$. If μ_n is the Dirac measure concentrated on x_n, and μ the Dirac measure concentrated on x, then clearly $\mu_n \xrightarrow{w} \mu$ (since $\int f\, d\mu_n = f(x_n) \to f(x) = \int f\, d\mu$). Let F be a closed set. It can happen that $x_n \notin F$ for all $n \in \mathbb{N}$, but $x \in F$ (think of x on the boundary of F). Then $\mu_n(F) = 0$, but $\mu(F) = 1$ and so $\limsup \mu_n(F) = 0 < 1 = \mu(F)$.

Proof. 1. $\Rightarrow$ 2.: trivial.

2. $\Rightarrow$ 3.: given a closed set F and let $F_k = \{x \in S : d(x, F) \leq 1/k\}$, $f_k = f_{F;1/k}$, $k \in \mathbb{N}$, be as in the proof of Proposition 7.3. Since $\mathbf{1}_F \leq f_k \leq \mathbf{1}_{F_k}$ and the functions f_k are Lipschitz continuous, we have

$$\limsup_n \mu_n(F) = \limsup_n \int \mathbf{1}_F\, d\mu_n \leq \lim_n \int f_k\, d\mu_n = \int f_k\, d\mu \leq \mu(F_k),$$

for all $k \in \mathbb{N}$. Letting $k \to \infty$ - as in the proof of Proposition 7.3 - yields 3.

3. $\Rightarrow$ 4.: follows directly by taking complements.

4. $\Rightarrow$ 1.: Pick $f \in C_b(S)$ and (possibly after applying a linear transformation to it) assume that $0 < f(x) < 1$, for all $x \in S$. Then, by Problem 5.10, we have $\int f\, d\nu = \int_0^1 \nu(f > t)\, dt$, for any probability measure on $\mathcal{B}(\mathbb{R})$. The set $\{f > t\} \subseteq S$ is open, so by 3., $\liminf_n \mu_n(f > t) \geq \mu(f > t)$, for all t. Therefore, by Fatou's lemma,

$$\begin{aligned}\liminf_n \int f\, d\mu_n &= \liminf_n \int_0^1 \mu_n(f > t)\, dt \geq \int_0^1 \liminf_n \mu_n(f > t)\, dt \\ &\geq \int_0^1 \mu(f > t)\, dt = \int f\, d\mu.\end{aligned}$$

We get the other inequality - $\int f\, d\mu \geq \limsup_n \int f\, d\mu_n$, by repeating the procedure with f replaced by $-f$.

3., 4. $\Rightarrow$ 5.: Let A be a μ-continuity set, let $\operatorname{Int} A$ be its interior and $\operatorname{Cl} A$ its closure. Then, since $\operatorname{Int} A$ is open and $\operatorname{Cl} A$ is closed, we have

$$\begin{aligned}\mu(\operatorname{Int} A) &\leq \liminf_n \mu_n(\operatorname{Int} A) \leq \liminf_n \mu_n(A) \leq \limsup_n \mu_n(A) \\ &\leq \limsup_n \mu_n(\operatorname{Cl} A) \leq \mu(\operatorname{Cl} A).\end{aligned}$$

Since $0 = \mu(\partial A) = \mu(\operatorname{Cl} A \setminus \operatorname{Int} A) = \mu(\operatorname{Cl} A) - \mu(\operatorname{Int} A)$, we conclude that all inequalities above are, in fact, equalities so that $\mu(A) = \lim_n \mu_n(A)$

5. $\Rightarrow$ 3.: For $x \in S$, consider the family $\{B_F(r) : r \geq 0\}$, where

$$B_F(r) = \{x \in S : d(x, F) \leq r\},$$

of closed sets.

Claim: There exists a countable subset R of $[0,\infty)$ such that $B_F(r)$ is a μ-continuity set for all $r \in [0,\infty) \setminus R$.

Proof. For $r \geq 0$ define $C_F(r) = \{x \in S : d(x,F) = r\}$, so that $\{C_F(r) : r \geq 0\}$ forms a measurable partition of S. Therefore, by Problem, 7.2, there exists a countable set $R \subseteq [0,\infty)$ such that $\mu(C_F(r)) = 0$ for $r \in [0,\infty) \setminus R$. It is not hard to see that $\partial B_F(r) \subseteq C_F(r)$ (btw, the inclusion may be strict), for each $r \geq 0$. Therefore, $\mu(\partial B_F(r)) = 0$, for all $r \in [0,\infty) \setminus R$. □

The above claim implies that there exists a sequence $r_k \in [0,\infty) \setminus R$ such that $r_k \searrow 0$. By 5. and the Claim above, we have $\mu_n(B_F(r_k)) \to \mu(B_F(r_k))$ for all $k \in \mathbb{N}$. Hence, for $k \in \mathbb{N}$,

$$\mu(B_F(r_k)) = \lim_n \mu_n(B_F(r_k)) \geq \limsup_n \mu_n(F).$$

By continuity of measure we have $\mu(B_F(r_k)) \searrow \mu(F)$, as $k \to \infty$, and so $\mu(F) \geq \limsup_n \mu_n(F)$. □

As we will soon see, it is sometimes easy to prove that $\mu_n(A) \to \mu(A)$ for all A in some subset of $\mathcal{B}(\mathbb{R})$. Our next result has something to say about cases when that is enough to establish weak convergence:

Proposition 7.6. *Let $\mathcal{I}$ be a collection of open subsets of S such that*

1. *$\mathcal{I}$ is a π-system,*
2. *Each open set in S can be represented as a finite or countable union of elements of $\mathcal{I}$.*

If $\mu_n(I) \to \mu(I)$, for each $I \in \mathcal{I}$, then $\mu_n \xrightarrow{w} \mu$.

Proof. For $I_1, I_2 \in \mathcal{I}$, we have $I_1 \cap I_2 \in \mathcal{I}$, and so

$$\begin{aligned}
\mu(I_1 \cup I_2) &= \mu(I_1) + \mu(I_2) - \mu(I_1 \cap I_2) \\
&= \lim_n \mu_n(I_1) + \lim_n \mu_n(I_2) - \lim_n \mu_n(I_1 \cap I_2) \\
&= \lim_n \Big(\mu_n(I_1) + \mu_n(I_2) - \mu_n(I_1 \cap I_2) \Big) = \lim_n \mu_n(I_1 \cup I_2).
\end{aligned}$$

Therefore, we can assume, without loss of generality that $\mathcal{I}$ is closed under finite unions.

For an open set G, let $G = \cup_k I_k$ be a representation of G as a union of a countable family in $\mathcal{I}$. By continuity of measure, for each $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $\mu(G) \leq \mu(\cup_{k=1}^K I_k) + \varepsilon$. Since $\cup_{k=1}^K I_K \in \mathcal{I}$, we have

$$\mu(G) - \varepsilon \leq \mu(\cup_{k=1}^K I_k) = \lim_n \mu_n(\cup_{k=1}^K I_k) \leq \liminf \mu_n(G).$$

Given that $\varepsilon > 0$ was arbitrary, we get $\mu(G) \leq \liminf_n \mu_n(G)$. □

Corollary 7.7. *Suppose that $S = \mathbb{R}$, and let μ_n be a family of probability measures on $\mathcal{B}(\mathbb{R})$. Let $F(x) = \mu((-\infty, x])$ and $F_n(x) = \mu_n((-\infty, x])$, $x \in \mathbb{R}$ be the corresponding cdfs. Then, the following two statements are equivalent:*

1. *$F_n(x) \to F(x)$ for all x such that F is continuous at x, and*
2. *$\mu_n \stackrel{w}{\to} \mu$.*

Proof. 2. $\Rightarrow$ 1.: Let C be the set of all x such that F is continuous at x; eqivalently, $C = \{x \in \mathbb{R} : \mu(\{x\}) = 0\}$. The sets $(-\infty, x]$ are μ-continuity sets for $x \in C$, so the Portmanteau theorem (Theorem 7.5) implies that $F_n(x) = \mu_n((-\infty, x]) \to \mu((-\infty, x]) = F(x)$, for all $x \in C$.

1. $\Rightarrow$ 2.: The set C^c is at most countable (why?) and so the family

$$\mathcal{I} = \{(a, b) : a < b, a, b \in C\},$$

satisfies the the conditions of Proposition 7.6. To show that $\mu_n \stackrel{w}{\to} \mu$, it will be enough to show that $\mu_n(I) \to \mu(I)$, for all $a, b \in \mathcal{I}$. Since $\mu((a, b)) = F(b-) - F(a)$, where $F(b-) = \lim_{x \nearrow b} F(x)$, it will be enough to show that

$$F_n(x-) \to F(x),$$

for all $x \in C$. Since $F_n(x-) \leq F_n(x)$, we have $\limsup_n F_n(x-) \leq \lim F_n(x) = F(x)$. To prove the other inequality, we pick $\varepsilon > 0$, and, using the continuity of F at x, find $\delta > 0$ such that $x - \delta \in C$ and $F(x - \delta) > F(x) - \varepsilon$. Since $F_n(x - \delta) \to F(x - \delta)$, there exists $n_0 \in \mathbb{N}$ such that $F_n(x - \delta) > F(x) - 2\varepsilon$ for $n \geq n_0$, and, by increase of F_n, $F_n(x-) > F(x) - 2\varepsilon$, for $n \geq n_0$. Consequently $\liminf_n F_n(x-) \geq F(x) - 2\varepsilon$ and the statement follows. □

One of the (many) reasons why weak convergence is so important, is the fact that it possesses nice compactness properties. The central result here is the theorem of Prohorov which is, in a sense, an analogue of the Arzelá-Ascoli compactness theorem for families of measures. The statement we give here is not the most general possible, but it will serve all our purposes.

Definition 7.8. A subset $\mathcal{M}$ of probability measures on $\mathcal{S}$ is said to be

1. **tight**, if for each $\varepsilon > 0$ there exists a compact set K such that

$$\sup_{\mu \in \mathcal{M}} \mu(K^c) \leq \varepsilon.$$

2. **relatively (sequentially) weakly compact** if any sequence $\{\mu_n\}_{n \in \mathbb{N}}$ in $\mathcal{M}$ admits a weakly-convergent subsequence $\{\mu_{n_k}\}_{k \in \mathbb{N}}$.

Theorem 7.9 (Prohorov). *Suppose that the metric space (S, d) is complete and separable, and let $\mathcal{M}$ be a set of probability measures on $\mathcal{S}$. Then $\mathcal{M}$ is relatively weakly compact if and only if it is tight.*

Proof. (Tight $\Rightarrow$ relatively weakly compact): Suppose that $\mathcal{M}$ is tight, and let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}$. Let Q be a countable and dense subset of $\mathbb{R}$, and let $\{q_k\}_{k\in\mathbb{N}}$ be an enumeration of Q. Since all $\{\mu_n\}_{n\in\mathbb{N}}$ are probability measures, the sequence $\{F_n(q_1)\}_{n\in\mathbb{N}}$, where $F_n(x) = \mu_n((-\infty, x])$ is bounded. Consequently, it admits a convergent subsequence; we denote its indices by $n_{1,k}$, $k \in \mathbb{N}$. The sequence $\{F_{n_{1,k}}(q_2)\}_{k\in\mathbb{N}}$ is also bounded, so we can extract a further subsequence - let's denote it by $n_{2,k}$, $k \in \mathbb{N}$, so that $F_{n_{2,k}}(q_2)$ converges as $k \to \infty$. Repeating this procedure for each element of Q, we arrive to a sequence of increasing sequences of integers $n_{i,k}$, $k \in \mathbb{N}$, $i \in \mathbb{N}$ with the property that $n_{i+1,k}$, $k \in \mathbb{N}$ is a subsequence of $n_{i,k}$, $k \in \mathbb{N}$ and that $F_{n_{i,k}}(q_j)$ converges for each $j \leq i$. Therefore, the diagonal sequence $m_k = n_{k,k}$, is a subsequence of each $n_{i,k}$, $k \in \mathbb{N}$, $i \in \mathbb{N}$, and can define a function $\tilde{F} : Q \to [0, 1]$ by

Note: In addition to the fact that the stated version of the theorem is not the most general available, we only give the proof for the so-called *Helly's selection theorem*, i.e., the special case $S = \mathbb{R}$. The general case is technically more involved, but the key ideas are similar.

$$\tilde{F}(q) = \lim_{k\to\infty} F_{m_k}(q).$$

Each F_n is non-decreasing and so if $\tilde{F}$. As a matter of fact the "right-continuous" version

$$F(x) = \inf_{q<x, q\in Q} \tilde{F}(q),$$

is non-decreasing and right-continuous (why?), with values in $[0, 1]$.

Our next task is to show that $F_{m_k}(x) \to F(x)$, for each $x \in C_F$, where C_F is the set of all points where F is continuous. We pick $x \in C_F$, $\varepsilon > 0$ and $q_1, q_2 \in Q$, $y \in \mathbb{R}$ such that $q_1 < q_2 < x < y$ and

$$F(x) - \varepsilon < F(q_1) \leq F(q_2) \leq F(x) \leq F(y) < F(x) + \varepsilon.$$

Since $F_{m_k}(q_2) \to \tilde{F}(q_2) \geq F(q_1)$ and $F_{m_k}(s) \to \tilde{F}(s) \leq F(s)$ (why is $\tilde{F}(s) \leq F(s)$?), we have, for large enough $k \in \mathbb{N}$

$$F(x) - \varepsilon < F_{m_k}(q_2) \leq F_{m_k}(x) \leq F_{m_k}(s) < F(x) + \varepsilon,$$

which implies that $F_{m_k}(x) \to F(x)$.

It remains to show - thanks to Corollary 7.7 - that $F(x) = \mu((-\infty, x])$, for some probability measure μ on $\mathcal{B}(\mathbb{R})$. For that, in turn, it will be enough to show that $F(x) \to 1$, as $x \to \infty$ and $F(x) \to 0$, as $x \to -\infty$. Indeed, in that case, we would have all the conditions needed to use Problem 6.24 to construct a probability space and a random variable X on it so that F is the cdf of X; the required measure μ would be the distribution $\mu = \mu_X$ of X.

To show that $F(x) \to 0, 1$ as $x \to \pm\infty$, we use tightness (note that this is the only place in the proof where it is used). For $\varepsilon > 0$, we pick

$M > 0$ such that $\mu_n([-M, M]) \geq 1 - \varepsilon$, for all $n \in \mathbb{N}$. In terms of corresponding cdfs, this implies that

$$F_n(-M) \leq \varepsilon \text{ and } F_n(M) \geq 1 - \varepsilon \text{ for all } n \in \mathbb{N}.$$

We can assume that $-M$ and M are continuity points of F (why?), so that

$$F(-M) = \lim_k F_{m_k}(-M) \leq \varepsilon \text{ and } F(M) = \lim_k F_{m_k}(M) \geq 1 - \varepsilon,$$

so that $\lim_{x\to\infty} F(x) \geq 1 - \varepsilon$ and $\lim_{x\to-\infty} F(x) \leq \varepsilon$. The claim follows from the arbitrariness of $\varepsilon > 0$.

(Relatively weakly compact $\Rightarrow$ tight): Suppose to the contrary, that $\mathcal{M}$ is relatively weakly compact, but not tight. Then, there exists $\varepsilon > 0$ such that for each $n \in \mathbb{N}$ there exists $\mu_n \in \mathcal{M}$ such that $\mu_n([-n, n]) < 1 - \varepsilon$, and, consequently,

$$\mu_n([-M, M]) < 1 - \varepsilon \text{ for } n \geq M. \tag{7.2}$$

The sequence $\{\mu_n\}_{n\in\mathbb{N}}$ admits a weakly-convergent subsequence, denoted by $\{\mu_{n_k}\}_{k\in\mathbb{N}}$. By (7.2), we have

$$\limsup_k \mu_{n_k}([-M, M]) \leq 1 - \varepsilon, \text{ for each } M > 0,$$

so that $\mu([-M, M]) \leq 1 - \varepsilon$ for all $M > 0$. Continuity of probability implies that $\mu(\mathbb{R}) \leq 1 - \varepsilon$ - a contradiction with the fact that μ is a *probability* measure on $\mathcal{B}(\mathbb{R})$. □

The following problem cases tightness in more operational terms:

Problem 7.3. Let $\mathcal{M}$ be a non-empty set of probability measures on $\mathbb{R}$. Show that $\mathcal{M}$ is tight if and only if there exists a non-decreasing function $\varphi : [0, \infty) \to [0, \infty)$ such that

1. $\varphi(x) \to \infty$ as $x \to \infty$, and
2. $\sup_{\mu\in\mathcal{M}} \int \varphi(|x|)\, \mu(dx) < \infty$.

Prohorov's theorem goes well with the following problem (it will be used soon):

Problem 7.4. Let μ be a probability measure on $\mathcal{B}(\mathbb{R})$ and let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(\mathbb{R})$ with the property that every subsequence $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mu_n\}_{n\in\mathbb{N}}$ has a (further) subsequence $\{\mu_{n_{k_l}}\}_{l\in\mathbb{N}}$ which converges towards μ. Show that $\{\mu_n\}_{n\in\mathbb{N}}$ is convergent.

Hint: If $\mu_n \stackrel{w}{\not\to} \mu$, then there exists $f \in C_b$ and a subsequence $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mu_n\}_{n\in\mathbb{N}}$ such that $\int f\, d\mu_{n_k}$ converges, but not to $\int f\, d\mu$.

We conclude with a comparison between convergence in distribution and convergence in probability.

Proposition 7.10 (Relation between $\stackrel{\mathbb{P}}{\to}$ and $\stackrel{\mathcal{D}}{\to}$)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Then $X_n \stackrel{\mathbb{P}}{\to} X$ implies $X_n \stackrel{\mathcal{D}}{\to} X$, for any random variable X. Conversely, $X_n \stackrel{\mathcal{D}}{\to} X$ implies $X_n \stackrel{\mathbb{P}}{\to} X$ if there exists $c \in \mathbb{R}$ such that $\mathbb{P}[X = c] = 1$.*

Proof. Assume that $X_n \stackrel{\mathbb{P}}{\to} X$. To show that $X_n \stackrel{\mathcal{D}}{\to} X$, the Portmanteau theorem guarantees that it will be enough to prove that $\limsup_n \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F]$, for all closed sets F. For $F \subseteq \mathbb{R}$, we define $F^{\varepsilon} = \{x \in \mathbb{R} : d(x, F) \leq \varepsilon\}$. Therefore, for a closed set F, we have

$$\begin{aligned}\mathbb{P}[X_n \in F] &= \mathbb{P}[X_n \in F, |X - X_n| > \varepsilon] + \mathbb{P}[X_n \in F, |X - X_n| \leq \varepsilon] \\ &\leq \mathbb{P}[|X - X_n| > \varepsilon] + \mathbb{P}[X \in F_{\varepsilon}].\end{aligned}$$

because $X \in F_{\varepsilon}$ if $X_n \in F$ and $|X - X_n| \leq \varepsilon$. Taking a lim sup of both sides yields

$$\limsup \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F_{\varepsilon}] + \limsup_n \mathbb{P}[|X - X_n| > \varepsilon] = \mathbb{P}[X \in F_{\varepsilon}].$$

Since $\cap_{\varepsilon>0} F^{\varepsilon} = F$, the statement follows.

For the second part, without loss of generality, we assume $c = 0$. Given $m \in \mathbb{N}$, let $f_m \in C_b(\mathbb{R})$ be a continuous function with values in $[0, 1]$ such that $f_m(0) = 1$ and $f_m(x) = 0$ for $|x| > 1/m$. Since $f_m(x) \leq \mathbf{1}_{[-1/m,1/m]}(x)$, we have

$$\mathbb{P}[|X_n| \leq 1/m] \geq \mathbb{E}[f_m(X_n)] \to f_m(0) = 1,$$

for each $m \in \mathbb{N}$. $\square$

Remark 7.11. It is not true that $X_n \stackrel{\mathcal{D}}{\to} X$ implies $X_n \stackrel{\mathbb{P}}{\to} X$ in general. Here is a simple example: take $\Omega = \{1, 2\}$ with uniform probability, and define $X_n(1) = 1$ and $X_n(2) = 2$, for n odd and $X_n(1) = 2$ and $X_n(2) = 1$, for n even. Then all X_n have the same distribution, so we have $X_n \stackrel{\mathcal{D}}{\to} X_1$. On the other hand $\mathbb{P}[|X_n - X_1| \geq \frac{1}{2}] = 1$, for n even. In fact, it is not hard to see that $X_n \stackrel{\mathbb{P}}{\not\to} X$ for any random variable X.

Additional Problems

Problem 7.5 (Total-variation convergence)**.** A sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of probability measures on $\mathcal{B}(\mathbb{R})$ is said to converge to the probability measure μ **in (total) variation** if

$$\sup_{A\in\mathcal{B}(\mathbb{R})} |\mu_n(A) - \mu(A)| \to 0 \text{ as } n \to \infty.$$

Compare convergence in variation to weak convergence: if one implies the other, prove it. Give counterexamples, if they are not equivalent.

Problem 7.6 (Scheffé's Theorem)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be absolutely continuous random variables with densities f_{X_n}, such that $f_{X_n}(x) \to f(x)$, λ-a.e., where f is the density of the absolutely-continuous random variable X. Show that X_n converges to X in total variation (defined in Problem 7.5), and, therefore, also in distribution.

Hint: Show that $\int_{\mathbb{R}} |f_{X_n} - f|\, d\lambda \to 0$ by writing the integrand in terms of $(f - f_{X_n})^+ \le f$.

Problem 7.7 (Convergence of moments)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ and X be random variables with a common uniform bound, i.e., such that

$$\exists\, M > 0,\ \forall\, n \in \mathbb{N},\ |X_n| \le M,\ |X| \le M, \text{ a.s.}$$

Show that the following two statements are equivalent:

1. $X_n \xrightarrow{\mathcal{D}} X$ (where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution), and
2. $\mathbb{E}[X_n^k] \to \mathbb{E}[X^k]$, as $n \to \infty$, for all $k \in \mathbb{N}$.

Hint: Use the Weierstrass approximation theorem: *Given $a < b \in \mathbb{R}$, a continuous function $f : [a,b] \to \mathbb{R}$ and $\varepsilon > 0$ there exists a polynomial P such that* $\sup_{x\in[a,b]} |f(x) - P(x)| \le \varepsilon$.

Problem 7.8 (Convergence of Maxima)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence of standard normal ($N(0,1)$) random variables. Define the sequence of *up-to-date-maxima* $\{M_n\}_{n\in\mathbb{N}}$ by

$$M_n = \max(X_1, \dots, X_n).$$

Show that

1. Show that $\lim_{x\to\infty} \frac{\mathbb{P}[X_1 > x]}{x^{-1}\exp(-\frac{1}{2}x^2)} = (2\pi)^{-\frac{1}{2}}$ by establishing the following inequality

   $$\frac{1}{x} \ge \frac{\mathbb{P}[X_1 > x]}{\phi(x)} \ge \frac{1}{x} - \frac{1}{x^3},\ x > 0, \tag{7.3}$$

   where, $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$ is the density of the standard normal.

   Hint: Integration by parts.
2. Prove that for any $\theta \in \mathbb{R}$, $\lim_{x\to\infty} \frac{\mathbb{P}[X_1 > x + \frac{\theta}{x}]}{\mathbb{P}[X_1 > x]} = \exp(-\theta)$.
3. Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence of real numbers with the property that $\mathbb{P}[X_1 > b_n] = 1/n$. Show that

   $$\mathbb{P}[b_n(M_n - b_n) \le x] \to \exp(-e^{-x}).$$
4. Show that $\lim_n \frac{b_n}{\sqrt{2\log n}} = 1$.
5. Show that $\frac{M_n}{\sqrt{2\log n}} \to 1$ in probability.

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 8
Characteristic Functions

First properties

A characteristic function is simply the Fourier transform, in probabilistic language. Since we will be integrating complex-valued functions, we define (both integrals on the right need to exist)

$$\int f\,d\mu = \int \Re f\,d\mu + i\int \Im f\,d\mu,$$

where $\Re f$ and $\Im f$ denote the real and the imaginary part of a function $f:\mathbb{R}\to\mathbb{C}$. The reader will easily figure out which properties of the integral transfer from the real case.

Definition 8.1. The **characteristic function** of a probability measure μ on $\mathcal{B}(\mathbb{R})$ is the function $\varphi_\mu:\mathbb{R}\to\mathbb{C}$ given by

$$\varphi_\mu(t) = \int e^{itx}\,\mu(dx)$$

When we speak of the characteristic function φ_X of a random variable X, we have the characteristic function φ_{μ_X} of its distribution μ_X in mind. Note, moreover, that

$$\varphi_X(t) = \mathbb{E}[e^{itX}].$$

While difficult to visualize, characteristic functions can be used to learn a lot about the random variables they correspond to. We start with some properties which follow directly from the definition:

Proposition 8.2. *Let X, Y and $\{X_n\}_{n\in\mathbb{N}}$ be a random variables.*

1. $\varphi_X(0)=1$ *and* $|\varphi_X(t)|\leq 1$*, for all* t.
2. $\varphi_{-X}(t)=\overline{\varphi_X(t)}$*, where bar denotes complex conjugation.*
3. φ_X *is uniformly continuous.*
4. *If* X *and* Y *are independent, then* $\varphi_{X+Y}=\varphi_X\varphi_Y$.

5. *For all $t_1 < t_2 < \dots < t_n$, the matrix $A = (a_{ij})_{1\leq i,j\leq n}$ given by*
$$a_{jk} = \varphi_X(t_j - t_k),$$
is Hermitian and positive semi-definite, i.e., $A^* = A$ and $\xi^T A \bar{\xi} \geq 0$, for any $\xi \in \mathbb{C}^n$,

6. *If $X_n \xrightarrow{\mathcal{D}} X$, then $\varphi_{X_n}(t) \to \varphi_X(t)$, for each $t \in \mathbb{R}$.*

Note: We do not prove (or use) it in these notes, but it can be shown that a function $\varphi : \mathbb{R} \to \mathbb{C}$, continuous at the origin with $\varphi(0) = 1$ is a characteristic function of some probability measure μ on $\mathcal{B}(\mathbb{R})$ if and only if it is **positive semidefinite**, i.e., if it satisfies part 5. of Proposition 8.2. This is known as **Bochner's theorem**.

Proof.

1. Immediate.

2. $\overline{e^{itx}} = e^{-itx}$.

3. We have $|\varphi_X(t) - \varphi_X(s)| = \left|\int (e^{itx} - e^{isx})\,\mu(dx)\right| \leq h(t-s)$, where $h(u) = \int \left|e^{iux} - 1\right|\,\mu(dx)$. Since $\left|e^{iux} - 1\right| \leq 2$, dominated convergence theorem implies that $\lim_{u\to 0} h(u) = 0$, and, so, φ_X is uniformly continuous.

4. Independence of X and Y implies the independence of $\exp(itX)$ and $\exp(itY)$. Therefore,
$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}] = \mathbb{E}[e^{itX}]\mathbb{E}[e^{itY}] = \varphi_X(t)\varphi_Y(t).$$

5. The matrix A is Hermitian by 2. above. To see that it is positive semidefinite, note that $a_{jk} = \mathbb{E}[e^{it_jX}e^{-it_kX}]$, and so
$$\begin{aligned}\sum_{j=1}^n\sum_{k=1}^n \xi_j\overline{\xi_k}a_{jk} &= \mathbb{E}\left[\left(\sum_{j=1}^n \xi_j e^{it_jX}\right)\overline{\left(\sum_{k=1}^n \xi_k e^{it_kX}\right)}\right]\\ &= \mathbb{E}[|\sum_{j=1}^n \xi_j e^{it_jX}|^2] \geq 0.\end{aligned}$$

6. The functions $x \mapsto \cos(tx)$ and $x \mapsto \sin(tx)$ and bounded and continuous so it suffices to apply the definition of weak convergence. □

Here is a simple problem you can use to test your understanding of the definitions:

Problem 8.1. Let μ and ν be two probability measures on $\mathcal{B}(\mathbb{R})$, and let φ_μ and φ_ν be their characteristic functions. Show that **Parseval's Identity** holds:
$$\int_{\mathbb{R}} e^{-its}\varphi_\mu(t)\,\nu(dt) = \int_{\mathbb{R}} \varphi_\nu(t-s)\,\mu(dt), \text{ for all } s \in \mathbb{R}.$$

Our next result shows μ can be recovered from its characteristic function φ_μ:

Theorem 8.3 (Inversion theorem)**.** *Let μ be a probability measure on $\mathcal{B}(\mathbb{R})$, and let $\varphi = \varphi_\mu$ be its characteristic function. Then, for $a < b \in \mathbb{R}$, we have*

$$\mu((a,b)) + \tfrac{1}{2}\mu(\{a,b\}) = \tfrac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt. \tag{8.1}$$

Proof. We start by picking $a < b$ and noting that

$$\frac{e^{-ita} - e^{-itb}}{it} = \int_a^b e^{-ity}\, dy,$$

so that, by Fubini's theorem, the integral in (8.1) is well-defined:

$$F(a,b,T) = \int_{[-T,T]\times[a,b]} \exp(-ity)\varphi(t)\, dy\, dt,$$

where

$$F(a,b,T) = \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt.$$

Another use of Fubini's theorem yields:

$$\begin{aligned} F(a,b,T) &= \int_{[-T,T]\times[a,b]\times\mathbb{R}} \exp(-ity)\exp(itx)\, dy\, dt\, \mu(dx) \\ &= \int_{\mathbb{R}} \left(\int_{[-T,T]\times[a,b]} \exp(-it(y-x))\, dy\, dt \right) \mu(dx) \\ &= \int_{\mathbb{R}} \left(\int_{[-T,T]} \tfrac{1}{it} \left(e^{-it(a-x)} - e^{-it(b-x)} \right) dt \right) \mu(dx). \end{aligned}$$

Set

$$f(a,b,T) = \int_{-T}^{T} \tfrac{1}{it}(e^{-it(a-x)} - e^{-it(b-x)})\, dt \text{ and } K(T,c) = \int_0^T \tfrac{\sin(ct)}{t}\, dt,$$

and note that, since cos is an even and sin an odd function, we have

$$\begin{aligned} f(a,b,T;x) &= 2\int_0^T \left(\tfrac{\sin((a-x)t)}{t} - \tfrac{\sin((b-x)t)}{t} \right) dt \\ &= 2K(T;a-x) - 2K(T;b-x). \end{aligned}$$

Note: The integral

$$\int_{-T}^{T} \tfrac{1}{it} \exp(-it(a-x))\, dt$$

is not defined; we *really need* to work with the full $f(a,b,T;x)$ to get the right cancellation.

Since

$$K(T;c) = \begin{cases} \int_0^T \frac{\sin(ct)}{ct} d(ct) = \int_0^{cT} \frac{\sin(s)}{s}\, ds = K(cT;1), & c > 0 \\ 0, & c = 0 \\ -K(|c|\, T;1), & c < 0, \end{cases} \tag{8.2}$$

Problem 5.11 implies that

$$\lim_{T\to\infty} K(T;c) = \begin{cases} \frac{\pi}{2}, & c > 0, \\ 0, & c = 0, \\ -\frac{\pi}{2}, & c < 0. \end{cases}$$

and so

$$\lim_{T\to\infty} f(a,b,T;x) = \begin{cases} 0, & x \in [a,b]^c, \\ \pi, & x = a \text{ or } x = b, \\ 2\pi, & a < x < b. \end{cases}$$

Observe first that the function $T \mapsto K(T;1)$ is continuous on $[0,\infty)$ and has a finite limit as $T \to \infty$ so that $\sup_{T\geq 0} |K(T;1)| < \infty$. Furthermore, (8.2) implies that $|K(T;c)| \leq \sup_{T\geq 0} K(T;1)$ for any $c \in \mathbb{R}$ and $T \geq 0$ so that

$$\sup\{|f(a,b,T;x)| : x \in \mathbb{R}, T \geq 0\} < \infty.$$

Therefore, we can use the dominated convergence theorem to get that

$$\begin{aligned} \lim_{T\to\infty} \tfrac{1}{2\pi} F(a,b,T;x) &= \lim_{T\to\infty} \tfrac{1}{2\pi} \int f(a,b,T;x)\, \mu(dx) \\ &= \tfrac{1}{2\pi} \int \lim_{T\to\infty} f(a,b,T;x)\, \mu(x) \\ &= \tfrac{1}{2}\mu(\{a\}) + \mu((a,b)) + \tfrac{1}{2}\mu(\{b\}). \qquad \square \end{aligned}$$

Corollary 8.4. *For probability measures μ_1 and μ_2 on $\mathcal{B}(\mathbb{R})$, the equality $\varphi_{\mu_1} = \varphi_{\mu_2}$ implies that $\mu_1 = \mu_2$.*

Proof. By Theorem 8.3, we have $\mu_1((a,b)) = \mu_2((a,b))$ for all $a, b \in C$ where C is the set of all $x \in \mathbb{R}$ such that $\mu_1(\{x\}) = \mu_2(\{x\}) = 0$. Since C^c is at most countable, it is straightforward to see that the family $\{(a,b) : a, b \in C\}$ of intervals is a π-system which generates $\mathcal{B}(\mathbb{R})$. $\square$

Corollary 8.5. *Suppose that $\int_{\mathbb{R}} |\varphi_\mu(t)|\, dt < \infty$. Then $\mu \ll \lambda$ and $\frac{d\mu}{d\lambda}$ is a bounded and continuous function given by*

$$\frac{d\mu}{d\lambda} = f, \text{ where } f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_\mu(t)\, dt \text{ for } x \in \mathbb{R}.$$

Proof. Since φ_μ is integrable and $|e^{-itx}| = 1$, f is well defined. For $a < b$ we have

$$\begin{aligned} \int_a^b f(x)\, dx &= \frac{1}{2\pi} \int_a^b \int_{\mathbb{R}} e^{-itx} \varphi_\mu(t)\, dt\, dx \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_\mu(t) \left(\int_a^b e^{-itx}\, dx \right) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt \qquad (8.2) \\ &= \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt \\ &= \mu((a,b)) + \tfrac{1}{2}\mu(\{a,b\}), \end{aligned}$$

by Theorem 8.3, where the use of Fubini's theorem above is justified by the fact that the function $(t,x) \mapsto e^{-itx}\varphi_\mu(t)$ is integrable on $[a,b] \times \mathbb{R}$, for all $a < b$. For a, b such that $\mu(\{a\}) = \mu(\{b\}) = 0$, the equation (8.2) implies that $\mu((a,b)) = \int_a^b f(x)\,dx$. The claim now follows by the $\pi - \lambda$-theorem. $\square$

Example 8.6. Here is a list of some common distributions and the corresponding characteristic functions:

1. *Continuous distributions.*

	Name	Parameters	Density $f_X(x)$	Ch. function $\varphi_X(t)$
1	Uniform	$a < b$	$\frac{1}{b-a}\,\mathbf{1}_{[a,b]}(x)$	$\frac{e^{-ita}-e^{-itb}}{it(b-a)}$
2	Symmetric Uniform	$a > 0$	$\frac{1}{2a}\,\mathbf{1}_{[-a,a]}(x)$	$\frac{sin(at)}{at}$
3	Normal	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}}\,\exp(-\,\frac{(x-\mu)^2}{2\sigma^2}\,)$	$\exp(i\mu t - \frac{1}{2}\sigma^2 t^2)$
4	Exponential	$\lambda > 0$	$\lambda\exp(-\lambda x)\mathbf{1}_{[0,\infty)}(x)$	$\frac{\lambda}{\lambda - it}$
5	Double Exponential	$\lambda > 0$	$\frac{1}{2}\,\lambda\exp(-\lambda\,\lvert x\rvert)$	$\frac{\lambda^2}{\lambda^2+t^2}$
6	Cauchy	$\mu \in \mathbb{R}, \gamma > 0$	$\frac{\gamma}{\pi(\gamma^2+(x-\mu)^2)}$	$\exp(i\mu t - \gamma\,\lvert t\rvert)$

2. *Discrete distributions.*

	Name	Parameters	Distribution μ_X,	Ch. function $\varphi_X(t)$
7	Dirac	$c \in \mathbb{R}$	δ_c	$\exp(itc)$
8	Biased Coin-toss	$p \in (0,1)$	$p\delta_1 + (1-p)\delta_{-1}$	$\cos(t) + (2p-1)i\sin(t)$
9	Geometric	$p \in (0,1)$	$\sum_{n\in\mathbb{N}_0} p^n(1-p)\delta_n$	$\frac{1-p}{1-e^{it}p}$
10	Poisson	$\lambda > 0$	$\sum_{n\in\mathbb{N}_0} e^{-\lambda}\,\frac{\lambda^n}{n!}\,\delta_n,\ n \in \mathbb{N}_0$	$\exp(\lambda(e^{it}-1))$

3. *A singular distribution.*

	Name	Ch. function $\varphi_X(t)$
11	Cantor	$e^{it/2}\prod_{k=1}^{\infty}\cos(\frac{t}{3^k})$

Tail behavior

We continue by describing several methods one can use to extract useful information about the tails of the underlying probability distribution from a characteristic function.

Proposition 8.7. *Let X be a random variable. If $\mathbb{E}[\lvert X\rvert^n] < \infty$, then $\frac{d^n}{(dt)^n}\varphi_X(t)$ exists for all t and*

$$\frac{d^n}{(dt)^n}\varphi_X(t) = \mathbb{E}[e^{itX}(iX)^n].$$

In particular

$$\mathbb{E}[X^n] = (-i)^n\,\frac{d^n}{(dt)^n}\varphi_X(0).$$

Proof. We give the proof in the case $n = 1$ and leave the general case to the reader:

$$\lim_{h\to 0} \tfrac{\varphi(h)-\varphi(0)}{h} = \lim_{h\to 0} \int_{\mathbb{R}} \tfrac{e^{ihx}-1}{h}\, \mu(dx) = \int_{\mathbb{R}} \lim_{h\to 0} \tfrac{e^{ihx}-1}{h}\, \mu(dx) = \int_{\mathbb{R}} ix\, \mu(dx),$$

where the passage of the limit under the integral sign is justified by the dominated convergence theorem which, in turn, can be used since

$$\left| \tfrac{e^{ihx}-1}{h} \right| \leq |x|\,, \text{ and } \int_{\mathbb{R}} |x|\ \mu(dx) = \mathbb{E}[|X|] < \infty. \qquad \square$$

Remark 8.8.

1. It can be shown that for *n even*, the existence of $\frac{d^n}{(dt)^n}\varphi_X(0)$ (in the appropriate sense) implies the finiteness of the n-th moment $\mathbb{E}[|X|^n]$.
2. When n is odd, it can happen that $\frac{d^n}{(dt)^n}\varphi_X(0)$ exists, but $\mathbb{E}[|X|^n] = \infty$ - see Problem 8.6.

Finer estimates of the tails of a probability distribution can be obtained by finer analysis of the behavior of φ around 0:

Proposition 8.9. *Let μ be a probability measure on $\mathcal{B}(\mathbb{R})$ and let $\varphi = \varphi_\mu$ be its characteristic function. Then, for $\varepsilon > 0$ we have*

$$\mu([-\tfrac{2}{\varepsilon}, \tfrac{2}{\varepsilon}]^c) \leq \tfrac{1}{\varepsilon} \int_{-\varepsilon}^{\varepsilon} (1 - \varphi(t))\, dt.$$

Proof. Let X be a random variable with distribution μ. We start by using Fubini's theorem to get

$$\begin{aligned} \tfrac{1}{2\varepsilon} \int_{\varepsilon}^{\varepsilon} (1-\varphi(t))\, dt &= \tfrac{1}{2\varepsilon}\mathbb{E}[\int_{-\varepsilon}^{\varepsilon} (1 - e^{itX})\, dt] \\ &= \tfrac{1}{\varepsilon}\mathbb{E}[\int_0^{\varepsilon} (1-\cos(tX))\, dt] = \mathbb{E}[1 - \tfrac{\sin(\varepsilon X)}{\varepsilon X}]. \end{aligned}$$

It remains to observe that $1 - \frac{\sin(x)}{x} \geq 0$ and $1 - \frac{\sin(x)}{x} \geq 1 - \frac{1}{|x|}$ for all x. Therefore, if we use the first inequality on $[-2, 2]$ and the second one on $[-2, 2]^c$, we get $1 - \frac{\sin(x)}{x} \geq \frac{1}{2}\mathbf{1}_{\{|x|>2\}}$ so that

$$\tfrac{1}{2\varepsilon} \int_{\varepsilon}^{\varepsilon} (1 - \varphi(t))\, dt \geq \tfrac{1}{2}\mathbb{P}[|\varepsilon X| > 2] = \tfrac{1}{2}\mu([-\tfrac{2}{\varepsilon}, \tfrac{2}{\varepsilon}]^c). \qquad \square$$

Problem 8.2. Use the inequality of Proposition 8.9 to show that if $\varphi(t) = 1 + O(|t|^\alpha)$ for some $\alpha > 0$, then $\int_{\mathbb{R}} |x|^\beta\ \mu(dx) < \infty$, for all $\beta < \alpha$. Give an example where $\int_{\mathbb{R}} |x|^\alpha\ \mu(dx) = \infty$.

Note: "$f(t) = g(t) + O(h(t))$" means that, for some $\delta > 0$, we have

$$\sup_{|t|\leq\delta} \tfrac{|f(t)-g(t)|}{h(t)} < \infty.$$

Problem 8.3 (Riemann-Lebesgue theorem)**.** Suppose that $\mu \ll \lambda$. Show that

$$\lim_{t\to\infty} \varphi_\mu(t) = \lim_{t\to -\infty} \varphi_\mu(t) = 0.$$

Hint: Use (and prove) the fact that $f \in \mathcal{L}^1_+(\mathbb{R})$ can be approximated in $\mathcal{L}^1(\mathbb{R})$ by a function of the form $\sum_{k=1}^n \alpha_k \mathbf{1}_{[a_k,b_k]}$.

The continuity theorem

Theorem 8.10 (Continuity theorem)**.** *Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability distributions on $\mathcal{B}(\mathbb{R})$, and let $\{\varphi_n\}_{n\in\mathbb{N}}$ be the sequence of their characteristic functions. Suppose that there exists a function $\varphi : \mathbb{R} \to \mathbb{C}$ such that*

1. *$\varphi_n(t) \to \varphi(t)$, for all $t \in \mathbb{R}$, and*
2. *φ is continuous at $t = 0$.*

Then, φ is the characteristic function of a probability measure μ on $\mathcal{B}(\mathbb{R})$ and $\mu_n \xrightarrow{w} \mu$.

Proof. We start by showing that the continuity of the limit φ implies tightness of $\{\mu_n\}_{n\in\mathbb{N}}$. Given $\varepsilon > 0$ there exists $\delta > 0$ such that $1 - \varphi(t) \leq \varepsilon/2$ for $|t| \leq \delta$. By the dominated convergence theorem we have

$$\begin{aligned}\limsup_{n\to\infty} \mu_n([-\tfrac{2}{\delta}, \tfrac{2}{\delta}]^c) &\leq \limsup_{n\to\infty} \tfrac{1}{\delta} \int_{\delta}^{\delta} (1 - \varphi_n(t))\, dt \\ &= \tfrac{1}{\delta} \int_{-\delta}^{\delta} (1 - \varphi(t))\, dt \leq \varepsilon.\end{aligned}$$

By taking an even smaller $\delta' > 0$, we can guarantee that

$$\sup_{n\in\mathbb{N}} \mu_n([-\tfrac{2}{\delta'}, \tfrac{2}{\delta'}]^c) \leq \varepsilon,$$

which, together with the arbitrariness of $\varepsilon > 0$ implies that $\{\mu_n\}_{n\in\mathbb{N}}$ is tight.

Let $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ be a convergent subsequence of $\{\mu_n\}_{n\in\mathbb{N}}$, and let μ be its limit. Since $\varphi_{n_k} \to \varphi$, we conclude that φ is the characteristic function of μ. It remains to show that the whole sequence converges to μ weakly. This follows, however, directly from Problem 7.4, since any convergent subsequence $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ has the same limit μ. □

Problem 8.4. Let φ be a characteristic function of some probability measure μ on $\mathcal{B}(\mathbb{R})$. Show that $\hat{\varphi}(t) = e^{\varphi(t)-1}$ is also a characteristic function of some probability measure $\hat{\mu}$ on $\mathcal{B}(\mathbb{R})$.

Additional Problems

Problem 8.5 (Atoms from the characteristic function)**.** Let μ be a probability measure on $\mathcal{B}(\mathbb{R})$, and let $\varphi = \varphi_\mu$ be its characteristic function.

1. Show that $\mu(\{a\}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-ita} \varphi(t)\, dt$.
2. Show that if $\lim_{t\to\infty} |\varphi(t)| = \lim_{t\to-\infty} |\varphi(t)| = 0$, then μ has no atoms.

3. Show that converse of (2) is false.

Hint: Prove that $|\varphi(t_n)| = 1$ along a suitably chosen sequence $t_n \to \infty$, where φ is the characteristic function of the Cantor distribution.

Problem 8.6 (Existence of $\varphi'_X(0)$ does not imply that $X \in \mathbb{L}^1$)**.** Let X be a random variable which takes values in $\mathbb{Z} \setminus \{-2, -1, 0, 1, 2\}$ with

$$\mathbb{P}[X = k] = \mathbb{P}[X = -k] = \tfrac{C}{k^2 \log(k)}, \text{ for } k = 3, 4, \dots,$$

where $C = \frac{1}{2}(\sum_{k \geq 3} \frac{1}{k^2 \log(k)})^{-1} \in (0, \infty)$. Show that $\varphi'_X(0) = 0$, but $X \notin \mathbb{L}^1$. *Hint:* Argue that, in order to establish that $\varphi'_X(0) = 0$, it is enough to show that

$$\lim_{h \to 0} \frac{1}{h} \sum_{k \geq 3} \frac{\cos(hk) - 1}{k^2 \log(k)} = 0.$$

Then split the sum at k close to $2/h$ and use (and prove) the inequality $|\cos(x) - 1| \leq \min(x^2/2, x)$. Bounding sums by integrals may help, too.

Problem 8.7 (Multivariate characteristic functions)**.** Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a random vector. The *characteristic function* $\varphi = \varphi_{\boldsymbol{X}} : \mathbb{R}^n \to \mathcal{C}$ is given by

$$\varphi(t_1, t_2, \dots, t_n) = \mathbb{E}[\exp(i \sum_{k=1}^{n} t_k X_k)].$$

We will also use the shortcut $\boldsymbol{t}$ for $(t_1, \dots, t_n)$ and $\boldsymbol{t} \cdot \boldsymbol{X}$ for the random variable $\sum_{k=1}^{n} t_k X_k$. Prove the following statements

Note: Take for granted the following statement (the proof of which is similar to the proof of the 1-dimensional case):

Suppose that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are random vectors with $\varphi_{\boldsymbol{X}_1}(\boldsymbol{t}) = \varphi_{\boldsymbol{X}_2}(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathbb{R}^n$. Then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same distribution, i.e. $\mu_{\boldsymbol{X}_1} = \mu_{\boldsymbol{X}_2}$.

1. Random variables X and Y are independent if and only if

$$\varphi_{(X,Y)}(t_1, t_2) = \varphi_X(t_1)\varphi_Y(t_2) \text{ for all } t_1, t_2 \in \mathbb{R}.$$

2. Random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same distribution if and only if random variables $\boldsymbol{t} \cdot \boldsymbol{X}_1$ and $\boldsymbol{t} \cdot \boldsymbol{X}_2$ have the same distribution for all $\boldsymbol{t} \in \mathbb{R}^n$. (This fact is known as *Wald's device*.)

An n-dimensional random vector $\boldsymbol{X}$ is said to be *Gaussian* (or, to have the *multivariate normal distribution*) if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and a symmetric positive semi-definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ such that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(i\, \boldsymbol{t} \cdot \boldsymbol{\mu} - \tfrac{1}{2} \boldsymbol{t}^\tau \boldsymbol{\Sigma} \boldsymbol{t}),$$

where $\boldsymbol{t}$ is interpreted as a column vector, and $()^\tau$ is transposition. This is denoted as $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. $\boldsymbol{X}$ is said to be *non-degenerate* if $\boldsymbol{\Sigma}$ is positive definite.

3. Show that a random vector $\boldsymbol{X}$ is Gaussian, if and only if the random vector $\boldsymbol{t} \cdot \boldsymbol{X}$ is normally distributed (with some mean and variance) for each $\boldsymbol{t} \in \mathbb{R}^n$.

Note: Be careful, nothing in the second statement tells you what the mean and variance of $\boldsymbol{t} \cdot \boldsymbol{X}$ are.

4. Let $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ be a Gaussian random vector. Show that X_k and X_l, $k \neq l$, are independent if and only if they are uncorrelated.

5. Construct a random vector (X, Y) such that both X and Y are normally distributed, but that $\boldsymbol{X} = (X, Y)$ is *not* Gaussian.
6. Let $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ be a random vector consisting of n independent random variables with $X_i \sim N(0, 1)$. Let $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ be a given positive semi-definite symmetric matrix, and $\boldsymbol{\mu} \in \mathbb{R}^n$ a given vector. Show that there exists an affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ such that the random vector $T(\boldsymbol{X})$ is Gaussian with $T(\boldsymbol{X}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
7. Find a necessary and sufficient condition on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ such that the converse of the previous problem holds true: For a Gaussian random vector $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, there exists an affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $T(\boldsymbol{X})$ has independent components with the $N(0, 1)$-distribution (i.e. $T(\boldsymbol{X}) \sim N(0, \boldsymbol{y}I)$, where $\boldsymbol{y}I$ is the identity matrix).

Problem 8.8 (Slutsky's Theorem)**.** Let X, Y, $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ be random variables defined on the same probability space, such that

$$X_n \stackrel{\mathcal{D}}{\to} X \text{ and } Y_n \stackrel{\mathcal{D}}{\to} Y. \tag{8.3}$$

Show that

1. It is not necessarily true that $X_n + Y_n \stackrel{\mathcal{D}}{\to} X + Y$. For that matter, we do not necessarily have $(X_n, Y_n) \stackrel{\mathcal{D}}{\to} (X, Y)$ (where the pairs are considered as random elements in the metric space $\mathbb{R}^2$).
2. If, in addition to (8.3), there exists a constant $c \in \mathbb{R}$ such that $\mathbb{P}[Y = c] = 1$, show that $g(X_n, Y_n) \stackrel{\mathcal{D}}{\to} g(X, c)$, for any continuous function $g : \mathbb{R}^2 \to \mathbb{R}$.

Hint: It is enough to show that $(X_n, Y_n) \stackrel{\mathcal{D}}{\to} (X_n, c)$. Use Problem 8.7).

Problem 8.9 (Convergence of a normal sequence)**.**

1. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of normally-distributed random variables converging weakly towards a random variable X. Show that X must be a normal random variable itself.
2. Let X_n be a sequence of normal random variables such that $X_n \stackrel{a.s.}{\to} X$. Show that $X_n \stackrel{\mathbb{L}^p}{\to} X$ for all $p \geq 1$.

Hint: Use this fact: *for a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of real numbers, the following two statements are equivalent*

(a) $\mu_n \to \mu \in \mathbb{R}$, *and*

(b) $\exp(it\mu_n) \to \exp(it\mu)$, *for all t.*

You don't need to prove it, but feel free to try.

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 9

The weak law of large numbers and the central limit theorem

The weak law of large numbers

We start with a definitive form of the weak law of large numbers. We need two lemmas, first[1]:

[1] Feel free to skip the proofs, but understand why the statement of Lemma 9.2 even needs one

Lemma 9.1. *Let $u_1, u_2, \dots, u_n$ and $w_1, w_2, \dots, w_n$ be complex numbers, all of modulus at most $M > 0$. Then*

$$\left|\prod_{k=1}^n u_k - \prod_{k=1}^n w_k\right| \leq M^{n-1} \sum_{k=1}^n |u_k - w_k|. \tag{9.1}$$

Proof. We proceed by induction. For $n = 1$, the claim is trivial. Suppose that (9.1) holds. Then

$$\begin{aligned}\left|\prod_{k=1}^{n+1} u_k - \prod_{k=1}^{n+1} w_k\right| &\leq \left|\prod_{k=1}^n u_k\right| |u_{n+1} - w_{n+1}| + |w_{n+1}| \left|\prod_{k=1}^n u_k - \prod_{k=1}^n w_k\right| \\ &\leq M^n |u_{n+1} - w_{n+1}| + M \times M^{n-1} \sum_{k=1}^n |u_k - w_k| \\ &= M^{(n+1)-1} \sum_{k=1}^{n+1} |u_k - w_k|. \qquad \square\end{aligned}$$

Lemma 9.2. *Let $\{z_n\}_{n\in\mathbb{N}}$ be a sequence of complex numbers with $z_n \to z \in \mathbb{C}$. Then $(1 + \frac{z_n}{n})^n \to e^z$.*

Proof. Using Lemma 9.1 with $u_k = 1 + \frac{z_n}{n}$ and $w_k = e^{z_n/n}$ for $k = 1, \dots, n$, we get

$$\left|(1 + \tfrac{z_n}{n})^n - e^{z_n}\right| \leq n M_n^{n-1} \left|1 + \tfrac{z_n}{n} - e^{z_n/n}\right|, \tag{9.2}$$

where $M_n = \max(\left|1 + \frac{z_n}{n}\right|, \left|e^{z_n/n}\right|)$. Let $K = \sup_{n\in\mathbb{N}} |z_n| < \infty$, so that $\left|e^{z_n/n}\right|^n \leq e^{|z_n|} \leq e^K$. Similarly, $\left|1 + \frac{z_n}{n}\right|^n \leq (1 + \frac{K}{n})^n \to e^K$. Therefore

$$L = \sup_{n\in\mathbb{N}} M_n^{n-1} < \infty. \tag{9.3}$$

To estimate the last term in (9.2), we start with the Taylor expansion $e^b = 1 + b + \sum_{k\geq 2} \frac{b^k}{k!}$, which converges absolutely for all $b \in \mathbb{C}$. Then, we use the fact that $\frac{1}{k!} \leq 2^{-k+1}$, to obtain

$$\left|e^b - 1 - b\right| \leq \sum_{k\geq 2} |b|^k \tfrac{1}{k!} \leq \left|b^2\right| \sum_{k\geq 2} 2^{-k+1} = |b|^2, \text{ for } |b| \leq 1. \tag{9.4}$$

Since $|z_n|/n \leq 1$ for large-enough n, it follows from (9.2), (9.4) and (9.3), that

$$\limsup_n \left|(1 + \tfrac{z_n}{n})^n - e^{z_n}\right| \leq \limsup_n nL \left|\tfrac{z_n}{n}\right|^2 = 0.$$

It remains to remember that $e^{z_n} \to e^z$ to finish the proof. □

Theorem 9.3 (Weak law of large numbers)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence of random variables with the (common) distribution μ and the characteristic function $\varphi = \varphi_\mu$ such that $\varphi'(0)$ exists. Then, $c = -i\varphi'(0)$ is a real number and*

$$\frac{1}{n}\sum_{k=1}^n X_k \to c \text{ in probability.}$$

Proof. Since $\varphi(-s) = \overline{\varphi(s)}$, we have

$$\varphi'(0) = \lim_{s\to 0} \tfrac{\varphi(-s)-1}{-s} = \lim_{s\to 0} \tfrac{\overline{\varphi(s)}-1}{-s} = -\lim_{s\to 0} \overline{\tfrac{\varphi(s)-1}{s}} = -\overline{\varphi'(0)}.$$

Therefore, $c = -i\varphi'(0) \in \mathbb{R}$.

Let $S_n = \sum_{k=1}^n X_k$. According to Proposition 7.10, it will be enough to show that $\frac{1}{n}S_n \xrightarrow{\mathcal{D}} c = -i\varphi'(0) \in \mathbb{R}$. Moreover, by Theorem 8.10, all we need to do is show that $\varphi_{\frac{1}{n}S_n}(t) \to e^{itc} = e^{t\varphi'(0)}$, for all $t \in \mathbb{R}$.

The iid property of $\{X_n\}_{n\in\mathbb{N}}$ and the fact that $\varphi_{\alpha X}(t) = \varphi_X(\alpha t)$ imply that

$$\varphi_{\frac{1}{n}S_n}(t) = (\varphi(\tfrac{t}{n}))^n = (1 + \tfrac{z_n}{n})^n,$$

where $z_n = n(\varphi(\frac{t}{n}) - 1)$. By the assumption, we have $\lim_{s\to 0} \frac{\varphi(s)-1}{s} = ic$, and so $z_n \to t\varphi'(0)$. Therefore, by Lemma 9.2 above, we have $\varphi_{\frac{1}{n}S_n}(t) \to e^{itc}$.

□

Remark 9.4.

1. It can be shown that the converse of Theorem 9.3 is true in the following sense: if $\frac{1}{n}S_n \xrightarrow{\mathbb{P}} c \in \mathbb{R}$, then $\varphi'(0)$ exists and $\varphi'(0) = ic$. That's why we call the result of Theorem 9.3 definitive.

2. $X_1 \in \mathbb{L}^1$ implies $\varphi'(0) = \mathbb{E}[X_1]$, so that Theorem 9.3 covers the classical case. As we have seen in Problem 8.6, there are cases when $\varphi'(0)$ exists but $\mathbb{E}[|X_1|] = \infty$.

Problem 9.1. Let $\{X_n\}_{n\in\mathbb{N}}$ be iid with the Cauchy distribution[2]. Show that φ_X is not differentiable at 0 and show that there is no constant c such that

$$\tfrac{1}{n}S_n \overset{\mathbb{P}}{\to} c,$$

where $S_n = \sum_{k=1}^n X_k$. *Hint:* What is the distribution of $\frac{1}{n}S_n$?

[2] Remember, the density of the Cauchy distribution is given by $f_X(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$.

An "iid"-central limit theorem

We continue with a central limit theorem for iid sequences. Unlike in the case of the (weak) law of large numbers, existence of the first moment will not be enough - we will need to assume that the second moment is finite, too. We will see how this assumption can be relaxed when we state and prove the Lindeberg-Feller theorem. We start with an estimate of the "error" term in the Taylor expansion of the exponential function of imaginary argument:

Lemma 9.5. *For $\xi \in \mathbb{R}$ we have*

$$\left| e^{i\xi} - \sum_{k=0}^n \tfrac{(i\xi)^k}{k!} \right| \leq \min\left(\tfrac{|\xi|^{n+1}}{(n+1)!}, 2\tfrac{|\xi|^n}{n!}\right)$$

Proof. If we write the remainder in the Taylor formula in the integral form (derived easily using integration by parts), we get

$$e^{i\xi} - \sum_{k=0}^n \tfrac{(i\xi)^k}{k!} = R_n(\xi), \text{ where } R_n(\xi) = i^{n+1} \int_0^\xi e^{iu} \tfrac{(\xi-u)^n}{n!}\, du.$$

The usual estimate of R_n gives:

$$|R_n(\xi)| \leq \tfrac{1}{n!} \int_0^{|\xi|} (|\xi| - u)^n\, du = \tfrac{|\xi|^{n+1}}{(n+1)!}.$$

We could also transform the expression for R_n by integrating it by parts:

$$\begin{aligned} R_n(\xi) &= \tfrac{i^{n+1}}{n!} \left(\tfrac{1}{i}\xi^n - \tfrac{n}{i} \int_0^\xi e^{iu} (\xi - u)^{n-1}\, du \right) \\ &= \tfrac{i^n}{(n-1)!} \left(\int_0^\xi (\xi - u)^{n-1}\, du - \int_0^\xi e^{iu} (\xi - u)^{n-1}\, du \right), \end{aligned}$$

since $\xi^n = n\int_0^\xi (\xi - u)^{n-1}\, du$. Therefore

$$\begin{aligned} |R_n(\xi)| &\leq \tfrac{1}{(n-1)!} \int_0^{|\xi|} (|\xi| - u)^{n-1} \left| e^{iu} - 1 \right| du \\ &\leq \tfrac{2}{n!} \int_0^{|\xi|} n(|\xi| - u)^{n-1}\, du = \tfrac{2|\xi|^n}{n!}. \end{aligned}$$

□

While the following result can be obtained as a direct consequence of twice-differentiability of the function φ at 0, we use the (otherwise useful) estimate based on Lemma 9.5 above:

Corollary 9.6. *Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\mathbb{E}[X^2] = \nu < \infty$, and let the function $r : [0, \infty) \to [0, \infty)$ be defined by*

$$r(t) = \mathbb{E}[X^2 \min(t\,|X|\,, 1)],\ t \geq 0. \tag{9.5}$$

Then

1. $\lim_{t \searrow 0} r(t) = 0$ *and*
2. $\left|\varphi_X(t) - 1 - it\mu + \frac{1}{2}\nu t^2\right| \leq t^2 r(|t|)$.

Proof. The inequality in 2. is a direct consequence of Lemma 9.5 (with the extra factor $\frac{1}{6}$ neglected). Part 1. follows from the dominated convergence theorem because

$$X^2 \min(1, t\,|X|) \leq X^2 \in \mathcal{L}^1 \text{ and } \lim_{t \to 0} X^2 \min(1, t\,|X|) = 0. \qquad \square$$

Theorem 9.7 (Central Limit Theorem - iid version)**.** *Let $\{X_n\}_{n \in \mathbb{N}}$ be an iid sequence of random variables with $0 < \mathrm{Var}[X_1] < \infty$. Then*

$$\tfrac{\sum_{k=1}^n (X_k - \mu)}{\sqrt{\sigma^2 n}} \xrightarrow{\mathcal{D}} \chi,$$

where $\chi \sim N(0,1)$, $\mu = \mathbb{E}[X_1]$ and $\sigma^2 = \mathrm{Var}[X_1]$.

Proof. By considering the sequence $\{(X_n - \mu)/\sqrt{\sigma^2}\}_{n \in \mathbb{N}}$, instead of $\{X_n\}_{n \in \mathbb{N}}$, we may assume that $\mu = 0$ and $\sigma = 1$. Let φ be the characteristic function of the common distribution of $\{X_n\}_{n \in \mathbb{N}}$ and set $S_n = \sum_{k=1}^n X_k$, so that

$$\varphi_{\frac{1}{\sqrt{n}} S_n}(t) = (\varphi(\tfrac{t}{\sqrt{n}}))^n.$$

By Theorem 8.10, the problem reduces to whether the following statement holds:

$$\lim_{n \to \infty} (\varphi(\tfrac{t}{\sqrt{n}}))^n = e^{-\frac{1}{2}t^2}, \text{ for each } t \in \mathbb{R}. \tag{9.6}$$

Corollary 9.6 guarantees that

$$\left|\varphi(t) - 1 + \tfrac{1}{2}t^2\right| \leq t^2 r(t) \text{ for all } t \in \mathbb{R},$$

where r is given by (9.5), i.e.,

$$\left|\varphi(\tfrac{t}{\sqrt{n}}) - 1 + \tfrac{t^2}{2n}\right| \leq \tfrac{t^2}{n} r(t/\sqrt{n}).$$

Lemma 9.1 with $u_1 = \dots = u_n = \varphi(\frac{t}{\sqrt{n}})$ and $w_1 = \dots = w_n = (1 - \frac{t^2}{2n})$ yields:

$$\left| (\varphi(\tfrac{t}{\sqrt{n}}))^n - (1 - \tfrac{t^2}{2n})^n \right| \leq t^2 r(t/\sqrt{n}),$$

for $n \geq \frac{2}{t^2}$ (so that $\max(|\varphi(\frac{t}{\sqrt{n}})|, |1 - \frac{t^2}{2n}|) \leq 1$). Since $\lim_n r(t/\sqrt{n}) = 0$, we have

$$\lim_{n\to\infty} \left| (\varphi(\tfrac{t}{\sqrt{n}}))^n - (1 - \tfrac{t^2}{2n})^n \right| = 0,$$

and (9.6) follows from the fact that $(1 - \frac{t^2}{2n})^n \to e^{-\frac{1}{2}t^2}$, for all t. $\square$

The Lindeberg-Feller Theorem

Unlike Theorem 9.7, the Lindeberg-Feller Theorem does not require summands to be equally distributed - it only prohibits any single term from dominating the sum. As usual, we start with a technical lemma:

Lemma 9.8. *Let $(c_{n,m})$, $n \in \mathbb{N}$, $m = 1, \dots, n$ be a (triangular) array of real numbers with*

1. *$\sum_{m=1}^n c_{n,m} \to c \in \mathbb{R}$, and $\sum_{m=1}^n |c_{n,m}|$ is a bounded sequence,*
2. *$m_n \to 0$, as $n \to \infty$, where $m_n = \max_{1\leq m\leq n} |c_{n,m}|$.*

Then

$$\prod_{m=1}^n (1 + c_{n,m}) \to e^c \text{ as } n \to \infty.$$

Proof. Without loss of generality we assume that $m_n < \frac{1}{2}$ for all n, and note that the statement is equivalent to $\sum_{m=1}^n \log(1 + c_{n,m}) \to c$, as $n \to \infty$. Since $\sum_{m=1}^n c_{n,m} \to c$, this is also equivalent to

$$\sum_{m=1}^n (\log(1 + c_{n,m}) - c_{n,m}) \to 0, \text{ as } n \to \infty. \tag{9.7}$$

Consider the function $f(x) = \log(1 + x) + x^2 - x$, $x > -1$. It is straightforward to check that $f(0) = 0$ and that the derivative $f'(x) = \frac{1}{1+x} + 2x - 1$ satisfies $f'(x) > 0$ for $x > 0$ and $f'(x) < 0$ for $x \in (-1/2, \infty)$. It follows that $f(x) \geq 0$ for $x \in [-1/2, \infty)$ so that (the absolute value can be inserted since $x \geq \log(1 + x)$)

$$|\log(1 + x) - x| \leq x^2 \text{ for } x \geq -\tfrac{1}{2}.$$

Since $m_n < \frac{1}{2}$, we have $|\log(1 + c_{n,m}) - c_{n,m}| \leq c_{n,m}^2$, and so

$$\begin{aligned} \left| \sum_{m=1}^n (\log(1 + c_{n,m}) - c_{n,m}) \right| &\leq \sum_{m=1}^n |\log(1 + c_{n,m}) - c_{n,m}| \leq \sum_{m=1}^n c_{n,m}^2 \\ &\leq m_n \sum_{m=1}^n |c_{n,m}| \to 0, \end{aligned}$$

because $\sum_{m=1}^n |c_{n,m}|$ is bounded and $m_n \to 0$. $\square$

Theorem 9.9 (Lindeberg-Feller)**.** *Let $X_{n,m}$, $n \in \mathbb{N}$, $m = 1, \dots, n$ be a (triangular) array of random variables such that*

1. $\mathbb{E}[X_{n,m}] = 0$, *for all* $n \in \mathbb{N}$, $m = 1, \dots, n$,
2. $X_{n,1}, \dots, X_{n,n}$ *are independent,*
3. $\sum_{m=1}^n \mathbb{E}[X_{n,m}^2] \to \sigma^2 > 0$, *as* $n \to \infty$,
4. *for each* $\varepsilon > 0$, $s_n(\varepsilon) \to 0$, *as* $n \to \infty$, *where* $s_n(\varepsilon) = \sum_{m=1}^n \mathbb{E}[X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| \geq \varepsilon\}}]$.

Then

$$X_{n,1} + \dots + X_{n,n} \xrightarrow{\mathcal{D}} \sigma \chi, \text{ as } n \to \infty,$$

where $\chi \sim N(0, 1)$.

Proof. Set $\varphi_{n,m} = \varphi_{X_{n,m}}$, $\sigma_{n,m}^2 = \mathbb{E}[X_{n,m}^2]$. Just like in the proofs of Theorems 9.3 and 9.7, it will be enough to show that

$$\prod_{m=1}^n \varphi_{n,m}(t) \to e^{-\frac{1}{2}\sigma^2 t^2}, \text{ for all } t \in \mathbb{R}.$$

We fix $t \neq 0$ and use Lemma 9.1 with $u_{n,m} = \varphi_{n,m}(t)$ and $w_{n,m} = 1 - \frac{1}{2}\sigma_{n,m}^2 t^2$ to conclude that

$$D_n(t) \leq M_n^{n-1} \sum_{m=1}^n \left| \varphi_{n,m}(t) - 1 + \tfrac{1}{2}\sigma_{n,m}^2 t^2 \right|,$$

where

$$D_n(t) = \left| \prod_{m=1}^n \varphi_{n,m}(t) - \prod_{m=1}^n (1 - \tfrac{1}{2}\sigma_{n,m}^2 t^2) \right|$$

and $M_n = 1 \vee \max_{1 \leq m \leq n}(\left|1 - \frac{1}{2}t^2\sigma_{n,m}^2\right|)$. Assumption (4) in the statement implies that

$$\begin{aligned}\sigma_{n,m}^2 &= \mathbb{E}[X_{n,m}^2 \mathbf{1}_{\{|X_{m,n}| \geq \varepsilon\}}] + \mathbb{E}[X_{n,m}^2 \mathbf{1}_{\{|X_{m,n}| < \varepsilon\}}] \leq \varepsilon^2 + \mathbb{E}[X_{n,m}^2 \mathbf{1}_{\{|X_{m,n}| < \varepsilon\}}] \\ &\leq \varepsilon^2 + s_n(\varepsilon),\end{aligned}$$

and so $\sup_{1 \leq m \leq n} \sigma_{n,m}^2 \to 0$, as $n \to \infty$. Therefore, for n large enough, we have $\frac{1}{2}t^2\sigma_{n,m}^2 \leq 2$ and $M_n = 1$.

According to Corollary 9.6 we now have (for large-enough n)

$$\begin{aligned}D_n(t) &\leq t^2 \sum_{m=1}^n \mathbb{E}[X_{n,m}^2 \min(t\,|X_{n,m}|, 1)] \\ &\leq t^2 \sum_{m=1}^n \Big(\mathbb{E}[X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| \geq \varepsilon\}}] + \mathbb{E}[t\,|X_{n,m}|^3 \mathbf{1}_{\{|X_{n,m}| < \varepsilon\}}] \Big) \\ &\leq t^2 s_n(\varepsilon) + t^3 \varepsilon \sum_{m=1}^n \mathbb{E}[X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| < \varepsilon\}}] \leq t^2 s_n(\varepsilon) + 2t^3 \varepsilon \sigma^2.\end{aligned}$$

Therefore, $\limsup_n D_n(t) \leq 2t^3 \varepsilon \sigma^2$, and so, since $\varepsilon > 0$ is arbitrary, we have $\lim_n D_n(t) = 0$.

Our last task is to remember that $\max_{1 \leq m \leq n} \sigma_{n,m}^2 \to 0$, note that $\sum_{m=1}^m \sigma_{n,m}^2 \to \sigma^2$ (why?), and use Lemma 9.8 to conclude that

$$\prod_{m=1}^{n} (1 - \tfrac{1}{2}\sigma_{n,m}^2 t^2) - e^{-\frac{1}{2}\sigma^2 t^2}. \qquad \square$$

Problem 9.2. Show how the iid central limit theorem follows from the Lindeberg-Feller theorem.

Example 9.10 (Cycles in a random permutation)**.** Let $\Pi : \Omega \to S_n$ be a random element taking values in the set S_n of all permutations of the set $\{1, \dots, n\}$, i.e., the set of all bijections $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$. One usually considers the probability measure on Ω such that Π is uniformly distributed over S_n, i.e. $\mathbb{P}[\Pi = \pi] = \frac{1}{n!}$, for each $\pi \in S_n$. A random element in S_n whose distribution is uniform over S_n is called a **random permutation**.

Remember that each permutation $\pi \in S_n$ be decomposed into cycles; a **cycle** is a collection $(i_1 i_2 \dots i_k)$ in $\{1, \dots, n\}$ such that $\pi(i_l) = i_{l+1}$ for $l = 1, \dots, k-1$ and $\pi(i_k) = i_1$. For example, the permutation $\pi : \{1,2,3,4\} \to \{1,2,3,4\}$, given by $\pi(1) = 3$, $\pi(2) = 1$, $\pi(3) = 2$, $\pi(4) = 4$ has two cycles: (132) and (4). More precisely, start from $i_1 = 1$ and follow the sequence $i_{k+1} = \pi(i_k)$, until the first time you return to $i_k = 1$. Write these number in order $(i_1 i_2 \dots i_k)$ and pick $j_1 \in \{1, 2, \dots, n\} \setminus \{i_1, \dots, i_k\}$. If no such j_1 exists, π consist of a single cycle. If it does, we repeat the same procedure starting from j_1 to obtain another cycle $(j_1 j_2 \dots j_l)$, etc. In the end, we arrive at the decomposition

$$(i_1 i_2 \dots i_k)(j_1 j_2 \dots j_l) \dots$$

of π into cycles.

Let us first answer the following, warm-up, question: what is the probability $p(n,m)$ that 1 is a member of a cycle of length m? Equivalently, we can ask for the number $c(n,m)$ of permutations in which 1 is a member of a cycle of length m. The easiest way to solve this is to note that each such permutation corresponds to a choice of $(m-1)$ district numbers of $\{2, 3, \dots\}$ - these will serve as the remaining elements of the cycle containing 1. This can be done in $\binom{n-1}{m-1}$ ways. Furthermore, the $m-1$ elements to be in the same cycle with 1 can be ordered in $(m-1)!$ ways. Also, the remaining $n-m$ elements give rise to $(n-m)!$ distinct permutations. Therefore,

$$c(n,m) = \binom{n-1}{m-1}(m-1)!(n-m)! = (n-1)!, \text{ and so } p(n,m) = \tfrac{1}{n}.$$

This is a remarkable result - all cycle lengths are equally likely. Note, also, that 1 is not special in any way.

Our next goal is to say something about the number of cycles - a more difficult task. We start by describing a procedure for producing a random permutation by building it from cycles. The reader will easily convince his-/herself that the outcome is uniformly distributed over all permutations. We start with $n-1$ independent random variables $\xi_2, \dots, \xi_n$ such that ξ_i is uniformly distributed over the set $\{0, 1, 2, \dots, n-i+1\}$. Let the first cycle start from $X_1 = 1$. If $\xi_2 = 0$, then we declare (1) to be a full cycle and start building the next cycle from 2. If $\xi_2 \neq 0$, we pick the ξ_2-th smallest element - let us call it X_2 - from the set of remaining $n-1$ numbers to be the second element in the first cycle. After that, we close the cycle if $\xi_3 = 0$, or append the ξ_3-th smallest element - let's call it X_3 - in $\{1, 2, \dots, n\} \setminus \{X_1, X_2\}$ to the cycle. Once the cycle $(X_1\ X_2\ \dots\ X_k)$ is closed, we pick the smallest element in $\{1, 2, \dots, n\} \setminus \{X_1, X_2, \dots, X_k\}$ - let's call it X_{k+1} - and repeat the procedure starting from X_{k+1} and using $\xi_{k+1}, \dots, \xi_n$ as "sources of randomness".

Let us now define the random variables (we stress the dependence on n here) $Y_{n,1}, \dots, Y_{n,n}$ by $Y_{n,k} = \mathbf{1}_{\{\xi_k = 0\}}$. In words, $Y_{n,k}$ is in indicator of the event when a cycle ends right after the position k. It is clear that $Y_{n,1}, \dots, Y_{n,k}$ are independent (they are functions of independent variables $\xi_1, \dots, \xi_n$). Also, $p(n,k) = \mathbb{P}[Y_{n,k} = 1] = \frac{1}{n-k+1}$. The number of cycles C_n is the same as the number of closing parenthesis, so $C_n = \sum_{k=1}^n Y_{k,n}$. (Btw, can you derive the identity $p(n,m) = \frac{1}{n}$ by using random variables $Y_{n,1}, \dots, Y_{n,n}$?)

It is easy to compute

$$
\begin{aligned}
\mathbb{E}[C_n] &= \sum_{k=1}^n \mathbb{E}[Y_{n,k} = 1] = \sum_{k=1}^n \tfrac{1}{n-k+1} = 1 + \tfrac{1}{2} + \dots + \tfrac{1}{n} \\
&= \log(n) + \gamma + o(1),
\end{aligned}
$$

where $\gamma \approx 0.58$ is the Euler-Mascheroni constant, and $a_n = b_n + o(n)$ means that $|b_n - a_n| \to 0$, as $n \to \infty$.

If we want to know more about the variability of C_n, we can also compute its variance:

$$
\begin{aligned}
\operatorname{Var}[C_n] &= \sum_{k=1}^n \operatorname{Var}[Y_{n,k}] = \sum_{k=1}^n \left(\tfrac{1}{n-k+1} - \tfrac{1}{(n-k+1)^2}\right) \\
&= \log(n) + \gamma - \tfrac{\pi^2}{6} + o(1).
\end{aligned}
$$

The Lindeberg-Feller theorem will give us the precise asymptotic behavior of C_n. For $m = 1, \dots, n$, we define

$$
X_{n,m} = \frac{Y_{n,m} - \mathbb{E}[Y_{n,m}]}{\sqrt{\log(n)}},
$$

so that $X_{n,m}$, $m = 1, \dots, n$ are independent and of mean 0. Furthermore, we have

$$\lim_n \sum_{m=1}^n \mathbb{E}[X_{n,m}^2] = \lim_n \frac{\log(n) + \gamma - \frac{\pi^2}{6} + o(1)}{\log(n)} = 1.$$

Finally, for $\varepsilon > 0$ and $\log(n) > 2/\varepsilon$, we have $\mathbb{P}[|X_{n,m}| > \varepsilon] = 0$, so

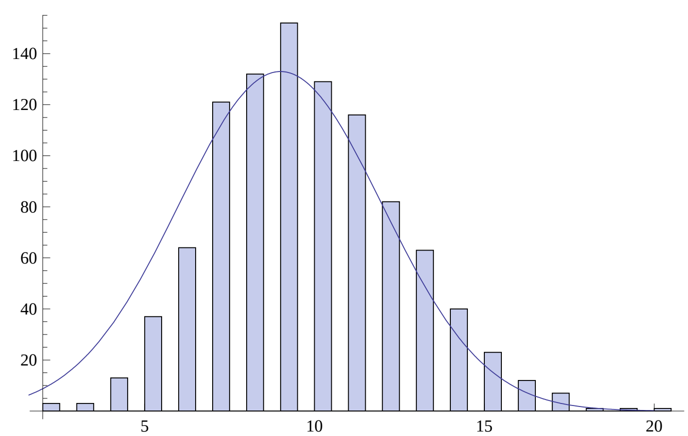
$$\sum_{m=1}^n \mathbb{E}[X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| \geq \varepsilon\}}] = 0.$$

Having checked that all the assumptions of the Lindeberg-Feller theorem are satisfied, we conclude that

$$\frac{C_n - \log(n)}{\sqrt{\log(n)}} \xrightarrow{\mathcal{D}} \chi, \text{ where } \chi \sim N(0,1).$$

It follows that (if we believe that the approximation is good) the number of cycles in a random permutation with $n = 8100$ is at most 18 with probability 99%.

How about variability? Here is histogram of the number of cycles from 1000 simulations for $n = 8100$, together with the appropriately-scaled density of the normal distribution with mean $\log(8100)$ and standard deviation $\sqrt{\log(8100)}$. The quality of approximation leaves something to be desired, but it seems to already work well in the tails: only 3 of 1000 had more than 17 cycles:



Additional Problems

Problem 9.3 (Lyapunov's theorem)**.** Let $\{X_n\}_{n \in \mathbb{N}}$ be an independent sequence, let $S_n = \sum_{m=1}^n X_m$, and let $\alpha_n = \sqrt{\text{Var}[S_n]}$. Suppose that $\alpha_n > 0$ for all $n \in \mathbb{N}$ and that there exists a constant $\delta > 0$ such that

$$\lim_n \alpha_n^{-(2+\delta)} \sum_{m=1}^n \mathbb{E}[|X_m - \mathbb{E}[X_m]|^{2+\delta}] = 0.$$

Show that

$$\frac{S_n - \mathbb{E}[S_n]}{\alpha_n} \xrightarrow{\mathcal{D}} \chi, \text{ where } \chi \sim N(0,1).$$

Problem 9.4 (Self-normalized sums)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be iid random variables with $\mathbb{E}[X_1] = 0$, $\sigma = \sqrt{\mathbb{E}[X_1^2]} > 0$ and $\mathbb{P}[X_1 = 0] = 0$. Show that the sequence $\{Y_n\}_{n\in\mathbb{N}}$ given by

$$Y_n = \frac{\sum_{k=1}^n X_k}{\sqrt{\sum_{k=1}^n X_k^2}},$$

converges in distribution, and identify its limit.

Hint: Use Slutsky's theorem (Problem 8.8)

Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 10
Conditional Expectation

The definition and existence of conditional expectation

For events A, B with $\mathbb{P}[B] > 0$, we recall the familiar object

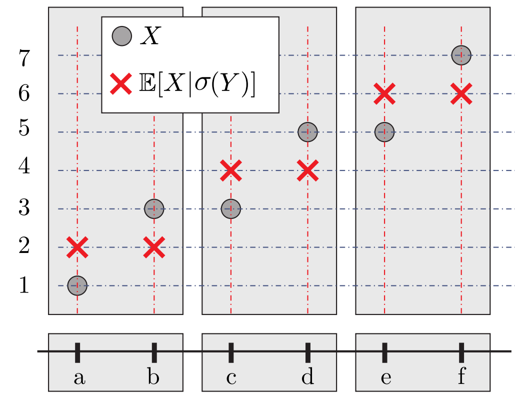
$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

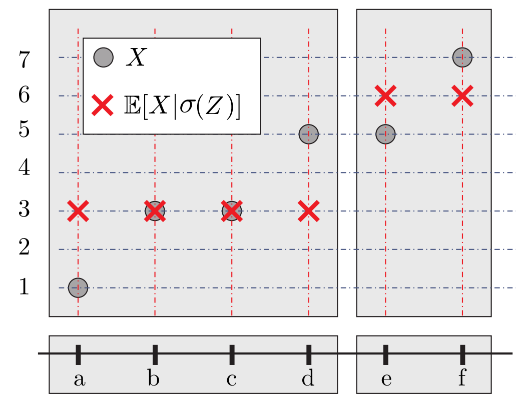
We say that $\mathbb{P}[A|B]$ the **conditional probability of** A**, given** B. It is important to note that the condition $\mathbb{P}[B] > 0$ is crucial. When X and Y are random variables defined on the same probability space, we often want to give a meaning to the expression $\mathbb{P}[X \in A | Y = y]$, even though it is usually the case that $\mathbb{P}[Y = y] = 0$. When the random vector (X, Y) admits a joint density $f_{X,Y}(x, y)$, and $f_Y(y) > 0$, the concept of conditional density $f_{X|Y=y}(x) = f_{X,Y}(x, y)/f_Y(y)$ is introduced and the quantity $\mathbb{P}[X \in A | Y = y]$ is given meaning via $\int_A f_{X|Y=y}(x, y)\, dx$. While this procedure works well in the restrictive case of absolutely continuous random vectors, we will see how it is encompassed by a general concept of a conditional expectation. Since probability is simply an expectation of an indicator, and expectations are linear, it will be easier to work with expectations and no generality will be lost.

Two main conceptual leaps here are: 1) we condition with respect to a σ-algebra, and 2) we view the conditional expectation itself as a random variable. Before we illustrate the concept in discrete time, here is the definition.

Definition 10.1. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $X \in \mathcal{L}^1$ be a random variable. We say that the random variable ξ is (a version of) the **conditional expectation of** X **with respect to** $\mathcal{G}$ - and denote it by $\mathbb{E}[X|\mathcal{G}]$ - if

1. $\xi \in \mathcal{L}^1$.
2. ξ is $\mathcal{G}$-measurable,
3. $\mathbb{E}[\xi \mathbf{1}_A] = \mathbb{E}[X \mathbf{1}_A]$, for all $A \in \mathcal{G}$.





Example 10.2. Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space where $\Omega = \{a, b, c, d, e, f\}$, $\mathcal{F} = 2^{\Omega}$ and $\mathbb{P}$ is uniform. Let X, Y and Z be random variables given by (in the obvious notation)

$$X \sim \begin{pmatrix} a & b & c & d & e & f \\ 1 & 3 & 3 & 5 & 5 & 7 \end{pmatrix},$$

$$Y \sim \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 1 & 1 & 7 & 7 \end{pmatrix} \text{ and } Z \sim \begin{pmatrix} a & b & c & d & e & f \\ 3 & 3 & 3 & 3 & 2 & 2 \end{pmatrix}$$

We would like to think about $\mathbb{E}[X|\mathcal{G}]$ as the average of $X(\omega)$ over all ω which are consistent with the our current information (which is $\mathcal{G}$). For example, if $\mathcal{G} = \sigma(Y)$, then the information contained in $\mathcal{G}$ is exactly the information about the exact value of Y. Knowledge of the fact that $Y = y$ does not necessarily reveal the "true" ω, but certainly rules out all those ω for which $Y(\omega) \neq y$.

In our specific case, if we know that $Y = 2$, then $\omega = a$ or $\omega = b$, and the expected value of X, given that $Y = 2$, is $\frac{1}{2}X(a) + \frac{1}{2}X(b) = 2$. Similarly, this average equals 4 for $Y = 1$, and 6 for $Y = 7$. Let us show that the random variable ξ defined by this average, i.e.,

$$\xi \sim \begin{pmatrix} a & b & c & d & e & f \\ 2 & 2 & 4 & 4 & 6 & 6 \end{pmatrix},$$

satisfies the definition of $\mathbb{E}[X|\sigma(Y)]$, as given above. The integrability is not an issue (we are on a finite probability space), and it is clear that ξ is measurable with respect to $\sigma(Y)$. Indeed, the atoms of $\sigma(Y)$ are $\{a, b\}$, $\{c, d\}$ and $\{e, f\}$, and ξ is constant over each one of those. Finally, we need to check that

$$\mathbb{E}[\xi \mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A], \text{ for all } A \in \sigma(Y),$$

which for an atom A translates into

$$\xi(\omega) = \tfrac{1}{\mathbb{P}[A]}\mathbb{E}[X\mathbf{1}_A] = \sum_{\omega' \in A} X(\omega')\mathbb{P}[\{\omega'\}|A], \text{ for all } \omega \in A.$$

The moral of the story is that when A is an atom, part 3. of Definition 10.1 translates into a requirement that ξ be constant on A with value equal to the expectation of X over A with respect to the conditional probability $\mathbb{P}[\cdot|A]$. In the general case, when there are no atoms, 3. still makes sense and conveys the same message.

Btw, since the atoms of $\sigma(Z)$ are $\{a, b, c, d\}$ and $\{e, f\}$, it is clear that

$$\mathbb{E}[X|\sigma(Z)](\omega) = \begin{cases} 3, & \omega \in \{a, b, c, d\}, \\ 6, & \omega \in \{e, f\}. \end{cases}$$

Look at the illustrations above and convince yourself that

$$\mathbb{E}[\mathbb{E}[X|\sigma(Y)]|\sigma(Z)] = \mathbb{E}[X|\sigma(Z)].$$

A general result along the same lines - called the *tower property of conditional expectation* - will be stated and proved below.

Our first task is to prove that conditional expectations always exist. When Ω is finite (as explained above) or countable, we can always construct them by averaging over atoms. In the general case, a different argument is needed. In fact, here are two:

Proposition 10.3. *Let $\mathcal{G}$ be a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$. Then*

1. *there exists a conditional expectation $\mathbb{E}[X|\mathcal{G}]$ for any $X \in \mathcal{L}^1$, and*
2. *any two conditional expectations of $X \in \mathcal{L}^1$ are equal $\mathbb{P}$-a.s.*

Proof. (*Uniqueness*): Suppose that ξ and ξ' both satisfy 1., 2. and 3. of Definition 10.1. Then

$$\mathbb{E}[\xi \mathbf{1}_A] = \mathbb{E}[\xi' \mathbf{1}_A], \text{ for all } A \in \mathcal{G}.$$

For $A_n = \{\xi' - \xi \geq \frac{1}{n}\}$, we have $A_n \in \mathcal{G}$ and so

$$\mathbb{E}[\xi \mathbf{1}_{A_n}] = \mathbb{E}[\xi' \mathbf{1}_{A_n}] \geq \mathbb{E}[(\xi + \tfrac{1}{n})\mathbf{1}_{A_n}] = \mathbb{E}[\xi \mathbf{1}_{A_n}] + \tfrac{1}{n}\mathbb{P}[A_n].$$

Consequently, $\mathbb{P}[A_n] = 0$, for all $n \in \mathbb{N}$, so that $\mathbb{P}[\xi' > \xi] = 0$. By a symmetric argument, we also have $\mathbb{P}[\xi' < \xi] = 0$.

(*Existence*): By linearity, it will be enough to prove that the conditional expectation exists for $X \in \mathcal{L}^1_+$.

1. *A Radon-Nikodym argument*. Suppose, first, that $X \geq 0$ and $\mathbb{E}[X] = 1$, as the general case follows by additivity and scaling. Then the prescription

$$\mathbb{Q}[A] = \mathbb{E}[X \mathbf{1}_A],$$

defines a probability measure on $(\Omega, \mathcal{F})$, which is absolutely continuous with respect to $\mathbb{P}$. Let $\mathbb{Q}^{\mathcal{G}}$ be the restriction of $\mathbb{Q}$ to $\mathcal{G}$; it is trivially absolutely continuous with respect to the restriction $\mathbb{P}^{\mathcal{G}}$ of $\mathbb{P}$ to $\mathcal{G}$. The Radon-Nikodym theorem - applied to the measure space $(\Omega, \mathcal{G}, \mathbb{P}^{\mathcal{G}})$ and the measure $\mathbb{Q}^{\mathcal{G}} \ll \mathbb{P}^{\mathcal{G}}$ - guarantees the existence of the Radon-Nikodym derivative

$$\xi = \frac{d\mathbb{Q}^{\mathcal{G}}}{d\mathbb{P}^{\mathcal{G}}} \in \mathbb{L}^1_+(\Omega, \mathcal{G}, \mathbb{P}^{\mathcal{G}}).$$

For $A \in \mathcal{G}$, we thus have

$$\mathbb{E}[X \mathbf{1}_A] = \mathbb{Q}[A] = \mathbb{Q}^{\mathcal{G}}[A] = \mathbb{E}^{\mathbb{P}^{\mathcal{G}}}[\xi \mathbf{1}_A] = \mathbb{E}[\xi \mathbf{1}_A].$$

where the last equality follows from the fact that $\xi \mathbf{1}_A$ is $\mathcal{G}$-measurable. Therefore, ξ is (a version of) the conditional expectation $\mathbb{E}[X|\mathcal{G}]$.

1. *An $\mathcal{L}^2$-argument*. Suppose, first, that $X \in \mathcal{L}^2$. Let H be the family of all $\mathcal{G}$-measurable elements in $\mathcal{L}^2$. Let $\bar{H}$ denote the closure of H in the topology induced by $\mathcal{L}^2$-convergence. Being a closed and convex (why?) subset of $\mathcal{L}^2$, $\bar{H}$ satisfies all the conditions of Problem **??** so that there exists $\xi \in \bar{H}$ at the minimal $\mathcal{L}^2$-distance from X (when $X \in \bar{H}$, we take $\xi = X$). The same problem states that ξ has the following property:

$$\mathbb{E}[(\eta - \xi)(X - \xi)] \geq 0 \text{ for all } \eta \in \bar{H},$$

and, since $\bar{H}$ is a linear space, we have

$$\mathbb{E}[(\eta - \xi)(X - \xi)] = 0, \text{for all } \eta \in \bar{H}.$$

It remains to pick η of the form $\eta = \xi + \mathbf{1}_A \in \bar{H}$, $A \in \mathcal{G}$, to conclude that

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\xi \mathbf{1}_A], \text{ for all } A \in \mathcal{G}.$$

Our next step is to show that ξ is $\mathcal{G}$-measurable (after a modification on a null set, perhaps). Since $\xi \in \bar{H}$, there exists a sequence $\{\xi_n\}_{n\in\mathbb{N}}$ such that $\xi_n \to \xi$ in $\mathcal{L}^2$. By Corollary **??**, $\xi_{n_k} \overset{a.s.}{\to} \xi$, for some subsequence $\{\xi_{n_k}\}_{k\in\mathbb{N}}$ of $\{\xi_n\}_{n\in\mathbb{N}}$. Set $\xi' = \liminf_{k\in\mathbb{N}} \xi_{n_k} \in \mathcal{L}^0([-\infty,\infty],\mathcal{G})$ and $\hat{\xi} = \xi' \mathbf{1}_{\{|\xi'|<\infty\}}$, so that $\hat{\xi} = \xi$, a.s., and $\hat{\xi}$ is $\mathcal{G}$-measurable.

We still need to remove the restriction $X \in \mathcal{L}^2_+$. We start with a general $X \in \mathcal{L}^1_+$ and define $X_n = \min(X, n) \in \mathcal{L}^\infty_+ \subseteq \mathcal{L}^2_+$. Let $\xi_n = \mathbb{E}[X_n|\mathcal{G}]$, and note that $\mathbb{E}[\xi_{n+1}\mathbf{1}_A] = \mathbb{E}[X_{n+1}\mathbf{1}_A] \geq \mathbb{E}[X_n \mathbf{1}_A] = \mathbb{E}[\xi_n \mathbf{1}_A]$. It follows (just like in the proof of uniqueness above) that $\xi_n \leq \xi_{n+1}$, a.s. We define $\xi = \sup_n \xi_n$, so that $\xi_n \nearrow \xi$, a.s. Then, for $A \in \mathcal{G}$, the monotone-convergence theorem implies that

$$\mathbb{E}[X\mathbf{1}_A] = \lim_n \mathbb{E}[X_n \mathbf{1}_A] = \lim_n \mathbb{E}[\xi_n \mathbf{1}_A] = \mathbb{E}[\xi \mathbf{1}_A],$$

and it is easy to check that $\xi \mathbf{1}_{\{\xi<\infty\}} \in \mathcal{L}^1(\mathcal{G})$ is a version of $\mathbb{E}[X|\mathcal{G}]$. □

Remark 10.4. There is no canonical way to choose "the version" of the conditional expectation. We follow the convention started with Radon-Nikodym derivatives, and interpret a statement such at $\xi \leq \mathbb{E}[X|\mathcal{G}]$, a.s., to mean that $\xi \leq \xi'$, a.s., for any version ξ' of the conditional expectation of X with respect to $\mathcal{G}$.

If we use the symbol $\mathbb{L}^1$ to denote the set of all a.s.-equivalence classes of random variables in $\mathcal{L}^1$, we can write:

$$\mathbb{E}[\cdot|\mathcal{G}] : \mathcal{L}^1(\mathcal{F}) \to \mathbb{L}^1(\mathcal{G}),$$

but $\mathbb{L}^1(\mathcal{G})$ cannot be replaced by $\mathcal{L}^1(\mathcal{G})$ in a natural way. Since $X = X'$, a.s., implies that $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X'|\mathcal{G}]$, a.s. (why?), we consider conditional expectation as a map from $\mathbb{L}^1(\mathcal{F})$ to $\mathbb{L}^1(\mathcal{G})$

$$\mathbb{E}[\cdot|\mathcal{G}] : \mathbb{L}^1(\mathcal{F}) \to \mathbb{L}^1(\mathcal{G}).$$

Properties

Conditional expectation inherits many of the properties from the "ordinary" expectation. Here are some familiar and some new ones:

Proposition 10.5. *Let $X, Y, \{X_n\}_{n\in\mathbb{N}}$ be random variables in $\mathcal{L}^1$, and let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. Then*

1. (linearity) $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}]$, *a.s.*

2. (monotonicity) $X \leq Y$, *a.s., implies* $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$, *a.s.*

3. (identity on $\mathbb{L}^1(\mathcal{G})$) *If X is $\mathcal{G}$-measurable, then $X = \mathbb{E}[X|\mathcal{G}]$, a.s. In particular, $c = \mathbb{E}[c|\mathcal{G}]$, for any constant $c \in \mathbb{R}$.*

4. (conditional Jensen's inequality) *If $\psi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}[|\psi(X)|] < \infty$ then*
$$\mathbb{E}[\psi(X)|\mathcal{G}] \geq \psi(\mathbb{E}[X|\mathcal{G}]), \text{ a.s.}$$

5. ($\mathcal{L}^p$-nonexpansivity) *If $X \in \mathcal{L}^p$, for $p \in [1, \infty]$, then $\mathbb{E}[X|\mathcal{G}] \in \mathbb{L}^p$ and*
$$||\mathbb{E}[X|\mathcal{G}]||_{\mathcal{L}^p} \leq ||X||_{\mathcal{L}^p}.$$
In particular,
$$\mathbb{E}[|X|\,|\mathcal{G}] \geq |\mathbb{E}[X|\mathcal{G}]| \text{ a.s.}$$

6. (pulling out what's known) *If Y is $\mathcal{G}$-measurable and $XY \in \mathcal{L}^1$, then*
$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}], \text{ a.s.}$$

7. ($\mathbb{L}^2$-projection) *If $X \in \mathcal{L}^2$, then $\xi^* = \mathbb{E}[X|\mathcal{G}]$ minimizes $\mathbb{E}[(X-\xi)^2]$ over all $\mathcal{G}$-measurable random variables $\xi \in \mathcal{L}^2$.*

8. (tower property) *If $\mathcal{H} \subseteq \mathcal{G}$, then*
$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}], \text{ a.s..}$$

9. (irrelevance of independent information) *If $\mathcal{H}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$ then*
$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}], \text{ a.s.}$$
In particular, if X is independent of $\mathcal{H}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$, a.s.

10. (conditional monotone-convergence theorem) *If* $0 \le X_n \le X_{n+1}$, *a.s., for all* $n \in \mathbb{N}$ *and* $X_n \to X \in \mathcal{L}^1$, *a.s., then*

$$\mathbb{E}[X_n|\mathcal{G}] \nearrow \mathbb{E}[X|\mathcal{G}], \text{ a.s.}$$

11. (conditional Fatou's lemma) *If* $X_n \ge 0$, *a.s., for all* $n \in \mathbb{N}$, *and* $\liminf_n X_n \in \mathcal{L}^1$, *then*

$$\mathbb{E}[\liminf_n X|\mathcal{G}] \le \liminf_n \mathbb{E}[X_n|\mathcal{G}], \text{ a.s.}$$

12. (conditional dominated-convergence theorem) *If* $|X_n| \le Z$, *for all* $n \in \mathbb{N}$ *and some* $Z \in \mathcal{L}^1$, *and if* $X_n \to X$, *a.s., then*

$$\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}], \text{ a.s. and in } \mathcal{L}^1.$$

Proof.

Note: Some of the properties are proved in detail. The others are only commented upon, since they are either similar to the other ones or otherwise not hard.

1. (linearity) $\mathbb{E}[(\alpha X + \beta Y)\mathbf{1}_A] = \mathbb{E}[(\alpha \mathbb{E}[X|\mathcal{G}] + \beta \mathbb{E}[Y|\mathcal{G}])\mathbf{1}_A]$, for $A \in \mathcal{G}$.

2. (monotonicity) Use $A = \{\mathbb{E}[X|\mathcal{G}] > \mathbb{E}[Y|\mathcal{G}]\} \in \mathcal{G}$ to obtain a contradiction if $\mathbb{P}[A] > 0$.

3. (identity on $\mathbb{L}^1(\mathcal{G})$) Check the definition.

4. (conditional Jensen's inequality) Use the result of Lemma **??** which states that $\psi(x) = \sup_{n\in\mathbb{N}}(a_n + b_n x)$, where $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ are sequences of real numbers.

5. ($\mathbb{L}^p$-nonexpansivity) For $p \in [1, \infty)$, apply conditional Jensen's inequality with $\psi(x) = |x|^p$. The case $p = \infty$ follows directly.

6. (pulling out what's known) For Y $\mathcal{G}$-measurable and $XY \in \mathcal{L}^1$, we need to show that

$$\mathbb{E}[XY\mathbf{1}_A] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A], \text{ for all } A \in \mathcal{G}. \tag{10.1}$$

Let us prove a seemingly less general statement:

$$\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]], \text{ for all } \mathcal{G}\text{-measurable } Z \text{ with } ZX \in \mathcal{L}^1. \tag{10.2}$$

The statement (10.1) will follow from it by taking $Z = Y\mathbf{1}_A$. For $Z = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$, (10.2) is a consequence of the definition of conditional expectation and linearity. Let us assume that both Z and X are nonnegative and $ZX \in \mathcal{L}^1$. In that case we can find a non-decreasing sequence $\{Z_n\}_{n\in\mathbb{N}}$ of non-negative simple random variables with $Z_n \nearrow Z$. Then $Z_n X \in \mathcal{L}^1$ for all $n \in \mathbb{N}$ and the monotone convergence theorem implies that

$$\mathbb{E}[ZX] = \lim_n \mathbb{E}[Z_n X] = \lim_n \mathbb{E}[Z_n \mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]].$$

Our next task is to relax the assumption $X \in \mathcal{L}^1_+$ to the original one $X \in \mathcal{L}^1$. In that case, the $\mathcal{L}^p$-nonexpansivity for $p = 1$ implies that

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

and so

$$|Z_n\mathbb{E}[X|\mathcal{G}]| \leq Z_n\mathbb{E}[|X|\,|\mathcal{G}] \leq Z\mathbb{E}[|X|\,|\mathcal{G}].$$

We know from the previous case that

$$\mathbb{E}[Z\mathbb{E}[|X|\,|\mathcal{G}]] = \mathbb{E}[Z\,|X|], \text{ so that } Z\mathbb{E}[|X|\,|\mathcal{G}] \in \mathcal{L}^1.$$

We can, therefore, use the dominated convergence theorem to conclude that

$$\mathbb{E}[Z\mathbb{E}[X|\mathcal{G}]] = \lim_n \mathbb{E}[Z_n\mathbb{E}[X|\mathcal{G}]] = \lim_n \mathbb{E}[Z_nX] = \mathbb{E}[ZX].$$

Finally, the case of a general Z follows by linearity.

7. ($\mathbb{L}^2$-projection) It is enough to show that $X - \mathbb{E}[X|\mathcal{G}]$ is orthogonal to all $\mathcal{G}$-measurable $\xi \in \mathcal{L}^2$. for that we simply note that for $\xi \in \mathcal{L}^2$, $x \in \mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])\xi] &= \mathbb{E}[\xi X] - \mathbb{E}[\xi\mathbb{E}[X|\mathcal{G}]] \\ &= \mathbb{E}[\xi X] - \mathbb{E}[\mathbb{E}[\xi X|\mathcal{G}]] = 0.\end{aligned}$$

8. (tower property) Use the definition.

9. (irrelevance of independent information) We assume $X \geq 0$ and show that

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A], \text{ a.s. for all } A \in \sigma(\mathcal{G}, \mathcal{H}). \tag{10.3}$$

Let $\mathcal{L}$ be the collection of all $A \in \sigma(\mathcal{G}, \mathcal{H})$ such that (10.3) holds. It is straightforward that $\mathcal{L}$ is a λ-system, so it will be enough to establish (10.3) for some π-system that generates $\sigma(\mathcal{G}, \mathcal{H})$. One possibility is $\mathcal{P} = \{G \cap H : G \in \mathcal{G}, H \in \mathcal{H}\}$, and for $G \cap H \in \mathcal{P}$ we use independence of $\mathbf{1}_H$ and $\mathbb{E}[X|\mathcal{G}]\mathbf{1}_G$, as well as the independence of $\mathbf{1}_H$ and $X\mathbf{1}_G$ to get

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_{G\cap H}] &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_G\mathbf{1}_H] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_G]\mathbb{E}[\mathbf{1}_H] \\ &= \mathbb{E}[X\mathbf{1}_G]\mathbb{E}[\mathbf{1}_H] = \mathbb{E}[X\mathbf{1}_{G\cap H}]\end{aligned} \tag{10.4}$$

10. (conditional monotone-convergence theorem) By monotonicity, we have $\mathbb{E}[X_n|\mathcal{G}] \nearrow \xi \in \mathcal{L}^0_+(\mathcal{G})$, a.s. The monotone convergence theorem implies that, for each $A \in \mathcal{G}$,

$$\mathbb{E}[\xi\mathbf{1}_A] = \lim_n \mathbb{E}[\mathbf{1}_A\mathbb{E}[X_n|\mathcal{G}]] = \lim_n \mathbb{E}[\mathbf{1}_AX_n] = \mathbb{E}[\mathbf{1}_AX].$$

11. (conditional Fatou's lemma) Set $Y_n = \inf_{k\geq n} X_k$, so that $Y_n \nearrow Y = \liminf_k X_k$. By monotonicity,

$$\mathbb{E}[Y_n|\mathcal{G}] \leq \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{G}], \text{ a.s.,}$$

and the conditional monotone-convergence theorem implies that

$$\mathbb{E}[Y|\mathcal{G}] = \lim_{n\in\mathbb{N}} \mathbb{E}[Y_n|\mathcal{G}] \leq \liminf_n \mathbb{E}[X_n|\mathcal{G}], \text{ a.s.}$$

12. (conditional dominated-convergence theorem) By the conditional Fatou's lemma, we have

$$\mathbb{E}[Z + X|\mathcal{G}] \leq \liminf_n \mathbb{E}[Z + X_n|\mathcal{G}],$$

as well as

$$\mathbb{E}[Z - X|\mathcal{G}] \leq \liminf_n \mathbb{E}[Z - X_n|\mathcal{G}], \text{ a.s.,}$$

and the a.s.-statement follows. □

Problem 10.1.

1. Show that the condition $\mathcal{H} \subseteq \mathcal{G}$ is necessary for the tower property to hold in general.

Hint: Take $\Omega = \{a, b, c\}$.

2. For $X, Y \in \mathcal{L}^2$ and a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, show that the following self-adjointness property holds

$$\mathbb{E}[X\,\mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,Y] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,\mathbb{E}[Y|\mathcal{G}]].$$

3. Let $\mathcal{H}$ and $\mathcal{G}$ be two sub-σ-algebras of $\mathcal{F}$. Is it true that

$$\mathcal{H} = \mathcal{G} \text{ if and only if } \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|\mathcal{H}], \text{ a.s., for all } X \in \mathcal{L}^1?$$

4. Construct two random variables X and Y in $\mathcal{L}^1$ such that $\mathbb{E}[X|\sigma(Y)] = \mathbb{E}[X]$, a.s., but X and Y are not independent.

Regular conditional distributions

Once we have a the notion of conditional expectation defined and analyzed, we can use it to define other, related, conditional quantities. The most important of those is the conditional probability:

Definition 10.6. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The **conditional probability** of $A \in \mathcal{F}$, given $\mathcal{G}$ - denoted by $\mathbb{P}[A|\mathcal{G}]$ - is defined by

$$\mathbb{P}[A|\mathcal{G}] = \mathbb{E}[\mathbf{1}_A|\mathcal{G}].$$

It is clear (from the conditional version of the monotone-convergence theorem) that

$$\mathbb{P}[\cup_{n\in\mathbb{N}} A_n|\mathcal{G}] = \sum_{n\in\mathbb{N}} \mathbb{P}[A_n|\mathcal{G}], \text{ a.s.} \tag{10.5}$$

We can, therefore, think of the conditional probability as a countably-additive map from events to (equivalence classes of) random variables $A \mapsto \mathbb{P}[A|\mathcal{G}]$. In fact, this map has the structure of a vector measure:

Definition 10.7. Let $(B, ||\cdot||)$ be a Banach space, and let $(S, \mathcal{S})$ be a measurable space. A map $\mu : \mathcal{S} \to B$ is called a **vector measure** if

1. $\mu(\emptyset) = 0$, and
2. for each pairwise-disjoint sequence $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$, we have

   $$\mu(\cup_n A_n) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

   (where the series in B converges absolutely).

Proposition 10.8. *The conditional probability $A \mapsto \mathbb{P}[A|\mathcal{G}] \in \mathbb{L}^1$ is a vector measure with values in $B = \mathbb{L}^1$.*

Proof. Clearly $\mathbb{P}[0|\mathcal{G}] = 0$, a.s. Let $\{A_n\}_{n\in\mathbb{N}}$ be a pairwise-disjoint sequence in $\mathcal{F}$. Then

$$\left|\left|\mathbb{P}[A_n|\mathcal{G}]\right|\right|_{\mathbb{L}^1} = \mathbb{E}[|\mathbb{E}[\mathbf{1}_{A_n}|\mathcal{G}]|] = \mathbb{E}[\mathbf{1}_{A_n}] = \mathbb{P}[A_n],$$

and so

$$\sum_{n\in\mathbb{N}} \left|\left|\mathbb{P}[A_n|\mathcal{G}]\right|\right|_{\mathbb{L}^1} = \sum_{n\in\mathbb{N}} \mathbb{P}[A_n] = \mathbb{P}[\cup_n A_n] \leq 1 < \infty,$$

which implies that $\sum_{n\in\mathbb{N}} \mathbb{P}[A_n|\mathcal{G}]$ converges absolutely in $\mathbb{L}^1$. Finally, for $A = \cup_{n\in\mathbb{N}} A_n$, we have

$$\begin{aligned}\left|\left|\mathbb{P}[A|\mathcal{G}] - \sum_{n=1}^{N} \mathbb{P}[A_n|\mathcal{G}]\right|\right|_{\mathbb{L}^1} &= \left|\left|\mathbb{E}[\sum_{n=N+1}^{\infty} \mathbf{1}_{A_n}|\mathcal{G}]\right|\right|_{\mathbb{L}^1} \\ &= \mathbb{P}[\cup_{n=N+1}^{\infty} A_n] \to 0 \text{ as } N \to \infty. \quad \square\end{aligned}$$

It is tempting to try to interpret the map $A \mapsto \mathbb{P}[A|\mathcal{G}](\omega)$ as a probability measure for a fixed $\omega \in \Omega$. It will not work in general; the reason is that $\mathbb{P}[A|\mathcal{G}]$ is defined only a.s., and the exceptional sets pile up when uncountable families of events A are considered. Even if we fixed versions $\mathbb{P}[A|\mathcal{G}] \in \mathcal{L}^0_+$, for each $A \in \mathcal{F}$, the countable additivity relation (10.5) holds only almost surely so there is no guarantee that, for a fixed $\omega \in \Omega$, $\mathbb{P}[\cup_{n\in\mathbb{N}} A_n|\mathcal{G}](\omega) = \sum_{n\in\mathbb{N}} \mathbb{P}[A_n|\mathcal{G}](\omega)$, for all pairwise disjoint sequences $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{F}$.

There is a way out of this predicament in certain situations, and we start with a description of an abstract object that corresponds to a well-behaved conditional probability:

Definition 10.9. Let $(R, \mathcal{R})$ and $(S, \mathcal{S})$ be measurable spaces. A map $\nu : R \times \mathcal{S} \to \mathbb{R}$ is called a **(measurable) kernel** if

1. $x \mapsto \nu(x, B)$ is $\mathcal{R}$-measurable for each $B \in \mathcal{S}$, and
2. $B \mapsto \nu(x, B)$ is a measure on $\mathcal{S}$ for each $x \in R$.

Definition 10.10. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, let $(S, \mathcal{S})$ be a measurable space, and let $e : \Omega \to S$ be a random element in S. A kernel $\mu_{e|\mathcal{G}} : \Omega \times \mathcal{S} \to [0, 1]$ is called the **regular conditional distribution of** e**, given** $\mathcal{G}$, if

$$\mu_{e|\mathcal{G}}(\omega, B) = \mathbb{P}[e \in B|\mathcal{G}](\omega), \text{ a.s., for all } B \in \mathcal{S}.$$

Remark 10.11.

1. When $(S, \mathcal{S}) = (\Omega, \mathcal{F})$, and $e(\omega) = \omega$, the regular conditional distribution of e (if it exists) is called the **regular conditional probability**. Indeed, in this case, $\mu_{e|\mathcal{G}}(\cdot, B) = \mathbb{P}[e \in B|\mathcal{G}] = \mathbb{P}[B|\mathcal{G}]$, a.s.
2. It can be shown that regular conditional distributions not need to exist in general if S is "too large".

When $(S, \mathcal{S})$ is "small enough", however, regular conditional distributions can be constructed. Here is what we mean by "small enough":

Definition 10.12. A measurable space $(S, \mathcal{S})$ is said to be a **Borel space** (or a **nice space**) if it is isomorphic to a Borel subset of $\mathbb{R}$, i.e., if there one-to-one map $\rho : S \to \mathbb{R}$ such that both ρ and ρ^{-1} are measurable.

Problem 10.2. Show that $\mathbb{R}^n$, $n \in \mathbb{N}$ (together with their Borel σ-algebras) are Borel spaces.

Hint: Show, first, that there is a measurable bijection $\rho : [0, 1] \to [0, 1] \times [0, 1]$ such that ρ^{-1} is also measurable. Use binary (or decimal, or ...) expansions.

Remark 10.13. It can be show that any Borel subset of any complete and separable metric space is a Borel space. In particular, the coin-toss space is a Borel space.

Proposition 10.14. *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $(S, \mathcal{S})$ be a Borel space. Any random element $e : \Omega \to S$ admits a regular conditional distribution.*

Proof. Let us, first, deal with the case $S = \mathbb{R}$, so that $e = X$ is a random variable. Let Q be a countable dense set in $\mathbb{R}$. For $q \in Q$, consider the random variable P^q, defined as an arbitrary version of

$$P^q = \mathbb{P}[X \leq q|\mathcal{G}].$$

By redefining each P^q on a null set (and aggregating the countably many null sets - one for each $q \in Q$), we may suppose that $P^q(\omega) \leq P^r(\omega)$, for $q \leq r$, $q, r \in Q$, for *all* $\omega \in \Omega$ and that $\lim_{q\to\infty} P^q(\omega) = 1$ and $\lim_{q\to-\infty} P^q(\omega) = 0$, for *all* $\omega \in \Omega$. For $x \in \mathbb{R}$, we set

$$F(\omega, x) = \inf_{q\in Q, q>x} P^q(\omega),$$

so that, for each $\omega \in \Omega$, $F(\omega, \cdot)$ is a right-continuous non-decreasing function from $\mathbb{R}$ to $[0, 1]$, which satisfies $\lim_{x\to\infty} F(\omega, x) = 1$ and $\lim_{x\to-\infty} F(\omega, x) = 0$, for all $\omega \in \Omega$. Moreover, as an infimum of countably many random variables, the map $\omega \mapsto F(\omega, x)$ is a random variable for each $x \in \mathbb{R}$.

By (the proof of) Proposition **??**, for each $\omega \in \Omega$, there exists a unique probability measure $\mu_{e|\mathcal{G}}(\omega, \cdot)$ on $\mathbb{R}$ such that $\mu_{e|\mathcal{G}}(\omega, (-\infty, x]) = F(\omega, x)$, for all $x \in \mathbb{R}$. Let $\mathcal{L}$ denote the set of all $B \in \mathcal{B}$ such that

1. $\omega \mapsto \mu_{e|\mathcal{G}}(\omega, B)$ is a random variable, and
2. $\mu_{e|\mathcal{G}}(\cdot, B)$ is a version of $\mathbb{P}[X \in B|\mathcal{G}]$.

It is not hard to check that $\mathcal{L}$ is a λ-system, so we need to prove that 1. and 2. hold for all B in some π-system which generates $\mathcal{B}(\mathbb{R})$. A convenient π-system to use is $\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}$. For $B = (-\infty, x] \in \mathcal{P}$, we have $\mu_{e|\mathcal{G}}(\omega, B) = F(\omega, x)$, so that 1. holds. To check 2., we need to show that $F(x, \omega) = \mathbb{P}[X \leq x|\mathcal{G}]$, a.s. This follows from the fact that

$$F(\cdot, x) = \inf_{q>x} P^q = \lim_{q\searrow x} P^q = \lim_{q\searrow x} \mathbb{P}[X \leq q|\mathcal{G}] = \mathbb{P}[X \leq x|\mathcal{G}], \text{ a.s.},$$

by the conditional dominated convergence theorem.

Turning to the case of a general random element e which takes values in a Borel space $(S, \mathcal{S})$, we pick a one-to-one measurable map $f : S \to \mathbb{R}$ whose inverse ρ^{-1} is also measurable. Then $X = \rho(e)$ is a random variable, and so, by the above, there exists a kernel $\mu_{X|\mathcal{G}} : \Omega \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ such that

$$\mu_{X|\mathcal{G}}(\cdot, A) = \mathbb{P}[\rho(e) \in A|\mathcal{G}], \text{ a.s.}$$

We define the kernel $\mu_{e|\mathcal{G}} : \Omega \times \mathcal{S} \to [0, 1]$ by

$$\mu_{e|\mathcal{G}}(\omega, B) = \mu_{X|\mathcal{G}}(\omega, \rho(B)).$$

Then, $\mu_{e|\mathcal{G}}(\cdot, B)$ is a random variable for each $B \in \mathcal{S}$ and for a pairwise disjoint sequence $\{B_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$, we have

$$\begin{aligned}\mu_{e|\mathcal{G}}(\omega, \cup_n B_n) &= \mu_{X|\mathcal{G}}(\omega, \rho(\cup_n B_n)) = \mu_{X|\mathcal{G}}(\omega, \cup_n \rho(B_n)) \\ &= \sum_{n\in\mathbb{N}} \mu_{X|\mathcal{G}}(\omega, \rho(B_n)) = \sum_{n\in\mathbb{N}} \mu_{e|\mathcal{G}}(\omega, B_n),\end{aligned}$$

which shows that $\mu_{e|\mathcal{G}}$ is a kernel; we used the measurability of ρ^{-1} to conclude that $\rho(B_n) \in \mathcal{B}(\mathbb{R})$ and the injectivity of ρ to ensure that $\{\rho(B_n)\}_{n\in\mathbb{N}}$ is pairwise disjoint. Finally, we need to show that $\mu_{e|\mathcal{G}}(\cdot, B)$ is a version of the conditional probability $\mathbb{P}[e \in B|\mathcal{G}]$. By injectivity of ρ, we have

$$\mathbb{P}[e \in B|\mathcal{G}] = \mathbb{P}[\rho(e) \in \rho(B)|\mathcal{G}] = \mu_{X|\mathcal{G}}(\cdot, \rho(B)) = \mu_{e|\mathcal{G}}(\cdot, B), \text{ a.s.} \quad \square$$

Remark 10.15. Note that the conditional distribution, even in its regular version, is not unique in general. Indeed, we can redefine it arbitrarily (as long as it remains a kernel) on a set of the form $N \times \mathcal{S} \subseteq \Omega \times \mathcal{S}$, where $\mathbb{P}[N] = 0$, without changing any of its defining properties. This will, in these notes, never be an issue.

One of the many reasons why regular conditional distributions are useful is that they sometimes allow non-conditional thinking to be transferred to the conditional case:

Proposition 10.16. *Let $\boldsymbol{X}$ be an $\mathbb{R}^n$-valued random vector, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $g : \mathbb{R}^n \to \mathbb{R}$ be a Borel function with the property $g(\boldsymbol{X}) \in \mathbb{L}^1$. Then $\int_{\mathbb{R}^n} g(\boldsymbol{x})\mu_{\boldsymbol{X}|\mathcal{G}}(\cdot, d\boldsymbol{x})$ is a $\mathcal{G}$-measurable random variable and*

$$\mathbb{E}[g(\boldsymbol{X})|\mathcal{G}] = \int_{\mathbb{R}^n} g(\boldsymbol{x})\mu_{\boldsymbol{X}|\mathcal{G}}(\cdot, d\boldsymbol{x}), \text{ a.s.}$$

Proof. When $g = \mathbf{1}_B$, for $B \in \mathbb{R}^n$, the statement follows by the very definition of the regular condition distribution. For the general case, we simply use the standard machine. $\square$

Just like we sometimes express the distribution of a random variable or a vector in terms of its density, cdf or characteristic function, we can talk about the conditional density, conditional cdf or the conditional characteristic function. All of those will correspond to the case covered in Proposition 10.14 and all conditional distributions will be assumed to be regular. For $\boldsymbol{x} = (x_1, \dots, x_n)$ and $\boldsymbol{y} = (y_1, \dots, y_n)$, $\boldsymbol{y} \leq_n \boldsymbol{x}$ means $y_1 \leq x_1, \dots, y_n \leq x_n$.

Definition 10.17. Let $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ be a random vector, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $\mu_{\boldsymbol{X}|\mathcal{G}} : \Omega \times \mathcal{B}(\mathbb{R}^n) \to [0,1]$ be the regular conditional distribution of $\boldsymbol{X}$ given $\mathcal{G}$.

1. The **(regular) conditional cdf of $\boldsymbol{X}$, given** $\mathcal{G}$ is the map $F : \Omega \times \mathbb{R}^n \to [0,1]$, given by

$$F(\omega, \boldsymbol{x}) = \mu_{\boldsymbol{X}|\mathcal{G}}(\omega, \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y} \leq_n \boldsymbol{x}\}), \text{ for } \boldsymbol{x} \in \mathbb{R}^n,$$

2. A map $f_{\boldsymbol{X}|\mathcal{G}} : \Omega \times \mathbb{R}^n \to [0, \infty)$ is called the **conditional density of $\boldsymbol{X}$ with respect to** $\mathcal{G}$ if

(a) $f_{X|\mathcal{G}}(\omega, \cdot)$ is Borel measurable for all $\omega \in \Omega$,

(b) $f_{X|\mathcal{G}}(\cdot, x)$ is $\mathcal{G}$-measurable for each $x \in \mathbb{R}^n$, and

(c) $\int_B f_{X|\mathcal{G}}(\omega, x)\, dx = \mu_{X|\mathcal{G}}(\omega, B)$, for all $\omega \in \Omega$ and all $B \in \mathcal{B}(\mathbb{R}^n)$,

3. The **conditional characteristic function of X, given** $\mathcal{G}$ is the map $\varphi_{X|\mathcal{G}} : \Omega \times \mathbb{R}^n \to \mathbb{C}$, given by

$$\varphi_{X|\mathcal{G}}(\omega, t) = \int_{\mathbb{R}^n} e^{it\cdot x}\, \mu_{X|\mathcal{G}}(\omega, dx), \text{ for } t \in \mathbb{R}^n \text{ and } \omega \in \Omega.$$

To illustrate the utility of the above concepts, here is a versatile result (see Example 10.20 below):

Proposition 10.18. *Let X be a random vector in $\mathbb{R}^n$, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. The following two statements are equivalent:*

1. *There exists a (deterministic) function $\varphi : \mathbb{R}^n \to \mathbb{C}$ such that for $\mathbb{P}$-almost all $\omega \in \Omega$,*
$$\varphi_{X|\mathcal{G}}(\omega, t) = \varphi(t), \text{ for all } t \in \mathbb{R}^n.$$

2. *$\sigma(X)$ is independent of $\mathcal{G}$.*

Moreover, whenever the two equivalent statements hold, φ is the characteristic function of X.

Proof. 1. $\to$ 2.. By Proposition 10.16, we have $\varphi_{X|\mathcal{G}}(\cdot, t) = \mathbb{E}[e^{it\cdot X}|\mathcal{G}]$, a.s. If we replace $\varphi_{X|\mathcal{G}}$ by φ, multiplying both sides by a bounded $\mathcal{G}$-measurable random variable Y and take expectations, we get

$$\varphi(t)\mathbb{E}[Y] = \mathbb{E}[Ye^{it\cdot X}].$$

In particular, for $Y = 1$ we get $\varphi(t) = \mathbb{E}[e^{it\cdot X}]$, so that

$$\mathbb{E}[Ye^{it\cdot X}] = \mathbb{E}[Y]\mathbb{E}[e^{it\cdot X}], \tag{10.5}$$

for all $\mathcal{G}$-measurable and bounded Y, and all $t \in \mathbb{R}^n$. For Y of the form $Y = e^{isZ}$, where Z is a $\mathcal{G}$-measurable random variable, relation (10.5) and (a minimal extension of) part 1. of Problem **??**, we conclude that X and Z are independent. Since Z is arbitrary and $\mathcal{G}$-measurable, X and $\mathcal{G}$ are independent.

2. $\to$ 1.. If $\sigma(X)$ is independent of $\mathcal{G}$, so is $e^{it\cdot X}$, and so, the "irrelevance of independent information" property of conditional expectation implies that

$$\varphi(t) = \mathbb{E}[e^{it\cdot X}] = \mathbb{E}[e^{it\cdot X}|\mathcal{G}] = \varphi_{X|\mathcal{G}}(\cdot, t), \text{ a.s.} \qquad \square$$

One of the most important cases used in practice is when a random vector $(X_1, \dots, X_n)$ admits a density and we condition on the σ-algebra generated by several of its components. To make the notation more intuitive, we denote the first d components $(X_1,\dots,X_d)$ by $\boldsymbol{X}^o$ (for *observed*) and the remaining $n-d$ components $(X_{d+1},\dots,X_n)$ by $\boldsymbol{X}^u$ (for *unobserved*).

Proposition 10.19. *Suppose that the random vector*

$$\boldsymbol{X} = (\boldsymbol{X}^o, \boldsymbol{X}^u) = (\underbrace{X_1,\dots,X_d}_{\boldsymbol{X}^o}, \underbrace{X_{d+1},\dots,X_n}_{\boldsymbol{X}^u})$$

admits a density $f_{\boldsymbol{X}}:\mathbb{R}^n\to[0,\infty)$ and that the σ-algebra $\mathcal{G}=\sigma(\boldsymbol{X}^o)$ is generated by the random vector $\boldsymbol{X}^o=(X_1,\dots,X_d)$, for some $d\in\{1,\dots,n-1\}$. Then, for $\boldsymbol{X}^u=(X_{d+1},\dots,X_n)$, there exists a conditional density $f_{\boldsymbol{X}^u|\mathcal{G}}:\Omega\times\mathbb{R}^{n-d}\to[0,\infty)$, of $\boldsymbol{X}^u$ given $\mathcal{G}$, and (a version of it) is given by

$$f_{\boldsymbol{X}^u|\mathcal{G}}(\omega,\boldsymbol{x}^u) = \begin{cases} \frac{f_{\boldsymbol{X}}(\boldsymbol{X}^o(\omega),\boldsymbol{x}^u)}{\int_{\mathbb{R}^{n-d}} f_{\boldsymbol{X}}(\boldsymbol{X}^o(\omega),\boldsymbol{y})\,d\boldsymbol{y}}, & \int_{\mathbb{R}^{n-d}} f(\boldsymbol{X}^o,\boldsymbol{y})\,d\boldsymbol{y}>0,\\ f_0(\boldsymbol{x}^u), & \text{otherwise,}\end{cases}$$

for $x\in\mathbb{R}^{n-d}$ and $\omega\in\Omega$, where $f_0:\mathbb{R}^{n-d}\to\mathbb{R}$ is an arbitrary density function.

Proof. First, we note that $f_{\boldsymbol{X}^u|\mathcal{G}}$ is constructed from the jointly Borel-measurable function $f_{\boldsymbol{X}}$ and the random vector $\boldsymbol{X}^o$ in an elementary way, and is, thus, jointly measurable in $\mathcal{G}\times\mathcal{B}(\mathbb{R}^{n-d})$. It remains to show that

$$\int_A f_{\boldsymbol{X}^u|\mathcal{G}}(\cdot,\boldsymbol{x}^u)\,d\boldsymbol{x}^u \text{ is a version of } \mathbb{P}[\boldsymbol{X}^u\in A|\mathcal{G}], \text{ for all } A\in\mathcal{B}(\mathbb{R}^{n-d}).$$

Equivalently, we need to show that

$$\mathbb{E}[\mathbf{1}_{\{\boldsymbol{X}^o\in A^o\}}\int_{A^u} f_{\boldsymbol{X}^u|\mathcal{G}}(\cdot,\boldsymbol{x}^u)\,d\boldsymbol{x}^u] = \mathbb{E}[\mathbf{1}_{\{\boldsymbol{X}^o\in A^o\}}\mathbf{1}_{\{\boldsymbol{X}^u\in A^u\}}],$$

for all $A^o\in\mathcal{B}(\mathbb{R}^d)$ and $A^u\in\mathcal{B}(\mathbb{R}^{n-d})$.

Fubini's theorem, and the fact that $f_{\boldsymbol{X}^o}(\boldsymbol{x}^o)=\int_{\mathbb{R}^{n-d}} f(\boldsymbol{x}^o,\boldsymbol{y})\,d\boldsymbol{y}$ is the density of $\boldsymbol{X}^o$ yield

$$\begin{aligned}\mathbb{E}[\mathbf{1}_{\{\boldsymbol{X}^o\in A^o\}}\int_{A^u} f_{\boldsymbol{X}^u|\mathcal{G}}(\cdot,\boldsymbol{x}^u)\,d\boldsymbol{x}^u] &= \int_{A^u}\mathbb{E}[\mathbf{1}_{\{\boldsymbol{X}^o\in A^o\}} f_{\boldsymbol{X}^u|\mathcal{G}}(\cdot,\boldsymbol{x}^u)]\,d\boldsymbol{x}^o\\ &= \int_{A^u}\int_{A^o} f_{\boldsymbol{X}^u|\mathcal{G}}(\boldsymbol{x}^o,\boldsymbol{x}^u) f_{\boldsymbol{X}^o}(\boldsymbol{x}^o)\,d\boldsymbol{x}^o\,d\boldsymbol{x}^u\\ &= \int_{A^u}\int_{A^o} f_{\boldsymbol{X}}(\boldsymbol{x}^o,\boldsymbol{x}^u)\,d\boldsymbol{x}^o\,d\boldsymbol{x}^u\\ &= \mathbb{P}[\boldsymbol{X}^o\in A^o,\boldsymbol{X}^u\in A^u].\end{aligned}$$

□

The above result expresses a conditional density, given $\mathcal{G}=\sigma(\boldsymbol{X}^o)$, as a (deterministic) function of $\boldsymbol{X}^o$. Such a representation is possible even when there is no joint density. The core of the argument is contained in the following problem:

Problem 10.3. Let X be a random vector in $\mathbb{R}^d$, and let $\mathcal{G} = \sigma(X)$ be the σ-algebra generated by X. Then, a random variable Z is $\mathcal{G}$-measurable if and only if there exists a Borel function $f : \mathbb{R}^d \to \mathbb{R}$ with the property that $Z = f(X)$.

Let $\boldsymbol{X}^o$ be a random vector in $\mathbb{R}^d$. For $X \in \mathcal{L}^1$ the conditional expectation $\mathbb{E}[X|\sigma(\boldsymbol{X}^o)]$ is $\sigma(\boldsymbol{X}^o)$-measurable, so there exists a Borel function $f : \mathbb{R}^d \to R$ such that $\mathbb{E}[X|\sigma(\boldsymbol{X}^o)] = f(\boldsymbol{X}^o)$, a.s. Note that f is uniquely defined only up to $\mu_{\boldsymbol{X}^o}$-null sets. The value $f(\boldsymbol{x}^o)$ at $\boldsymbol{x}^o \in \mathbb{R}^d$ is usually denoted by $\mathbb{E}[X|\boldsymbol{X}^o = \boldsymbol{x}^o]$.

Example 10.20 (Conditioning normals on their components)**.** Let $\boldsymbol{X} = (\boldsymbol{X}^o, \boldsymbol{X}^u) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ be a multivariate normal random vector with mean $\boldsymbol{\mu} = (\boldsymbol{\mu}^o, \boldsymbol{\mu}^u)$ and the variance-covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T]$, where $\tilde{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{\mu}$. A block form of the matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{ou} \\ \boldsymbol{\Sigma}_{uo} & \boldsymbol{\Sigma}_{uu} \end{pmatrix},$$

Where

$$\begin{aligned}
\boldsymbol{\Sigma}_{oo} &= \mathbb{E}[\tilde{\boldsymbol{X}}^o(\tilde{\boldsymbol{X}}^o)^T] \in \mathbb{R}^{d\times d} \\
\boldsymbol{\Sigma}_{ou} &= \mathbb{E}[\tilde{\boldsymbol{X}}^o(\tilde{\boldsymbol{X}}^u)^T] \in \mathbb{R}^{d\times (n-d)} \\
\boldsymbol{\Sigma}_{uo} &= \mathbb{E}[\tilde{\boldsymbol{X}}^u(\tilde{\boldsymbol{X}}^o)^T] \in \mathbb{R}^{(n-d)\times d} \\
\boldsymbol{\Sigma}_{uu} &= \mathbb{E}[\tilde{\boldsymbol{X}}^u(\tilde{\boldsymbol{X}}^u)^T] \in \mathbb{R}^{(n-d)\times (n-d)}.
\end{aligned}$$

We assume that $\boldsymbol{\Sigma}_{oo}$ is invertible. Otherwise, we can find a subset of components of $\boldsymbol{X}^o$ whose variance-covariance matrix in invertible and which generate the same σ-algebra (why?). The matrix $A = \boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}$ has the property that $\mathbb{E}[(\tilde{\boldsymbol{X}}^u - A\tilde{\boldsymbol{X}}^o)(\tilde{\boldsymbol{X}}^o)^T] = 0$, i.e., that the random vectors $\tilde{\boldsymbol{X}}^o - A\tilde{\boldsymbol{X}}^o$ and $\tilde{\boldsymbol{X}}^o$ are uncorrelated. We know, however, that $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{X}}^o, \tilde{\boldsymbol{X}}^u)$ is a Gaussian random vector, so, by Problem **??**, part 3., $\tilde{\boldsymbol{X}}^o - A\tilde{\boldsymbol{X}}^o$ is independent of $\tilde{\boldsymbol{X}}^o$. It follows from Proposition 10.18 that the conditional characteristic function of $\tilde{\boldsymbol{X}}^o - A\tilde{\boldsymbol{X}}^o$, given $\mathcal{G} = \sigma(\tilde{\boldsymbol{X}}^o)$ is deterministic and given by

$$\mathbb{E}[e^{i\boldsymbol{t}(\tilde{\boldsymbol{X}}^u - A\tilde{\boldsymbol{X}}^o)}|\mathcal{G}] = \varphi_{\tilde{\boldsymbol{X}}^u - A\tilde{\boldsymbol{X}}^o}(\boldsymbol{t}), \text{ for } \boldsymbol{t} \in \mathbb{R}^{n-d}.$$

Since $A\tilde{\boldsymbol{X}}^o$ is $\mathcal{G}$-measurable, we have

$$\mathbb{E}[e^{i\boldsymbol{t}\boldsymbol{X}^u}|\mathcal{G}] = e^{i\boldsymbol{t}\boldsymbol{\mu}^u} e^{i\boldsymbol{t}A\tilde{\boldsymbol{X}}^o} e^{-\frac{1}{2}\boldsymbol{t}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{t}}, \text{ for } \boldsymbol{t} \in \mathbb{R}^{n-d}.$$

where $\hat{\boldsymbol{\Sigma}} = \mathbb{E}[(\tilde{\boldsymbol{X}}^u - A\tilde{\boldsymbol{X}}^o)(\tilde{\boldsymbol{X}}^u - A\tilde{\boldsymbol{X}}^o)^T]$. A simple calculation yields that, conditionally on $\mathcal{G}$, $\boldsymbol{X}^u$ is multivariate normal with mean $\boldsymbol{\mu}_{\boldsymbol{X}^u|\mathcal{G}}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}^u|\mathcal{G}}$ given by

$$\boldsymbol{\mu}_{\boldsymbol{X}^u|\mathcal{G}} = \boldsymbol{\mu}^o + A(\boldsymbol{X}^o - \boldsymbol{\mu}^o), \ \boldsymbol{\Sigma}_{\boldsymbol{X}^u|\mathcal{G}} = \boldsymbol{\Sigma}_{uu} - \boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{ou}.$$

Note how the mean gets corrected by a multiple of the difference between the observed value $\boldsymbol{X}^o$ and its (unconditional) expected value. Similarly, the variance-covariance matrix gets corrected by $\boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{ou}$, but this quantity does not depend on the observation $\boldsymbol{X}^o$.

Problem 10.4. Let (X_1, X_2) be a bivariate normal vector with $\mathrm{Var}[X_1] > 0$. Work out the exact form of the conditional distribution of X_2, given X_1 in terms of $\mu_i = \mathbb{E}[X_i]$, $\sigma_i^2 = \mathrm{Var}[X_i]$, $i = 1, 2$ and the correlation coefficient $\rho = \mathrm{corr}(X_1, X_2)$.

Additional Problems

Problem 10.5 (Conditional expectation for non-negative random variables)**.** A parallel definition of conditional expectation can be given for random variables in $\mathcal{L}^0_+$. For $X \in \mathcal{L}^0_+$, we say that the random variable Y is **a conditional expectation of** X **with respect to** $\mathcal{G}$ - and denote it by $\mathbb{E}[X|\mathcal{G}]$ - if

(a) Y is $\mathcal{G}$-measurable and $[0, \infty]$-valued, and

(b) $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] \in [0, \infty]$, for $A \in \mathcal{G}$.

Show that

1. $\mathbb{E}[X|\mathcal{G}]$ exists for each $X \in \mathcal{L}^0_+$.
2. $\mathbb{E}[X|\mathcal{G}]$ is unique a.s.
3. $\mathbb{E}[X|\mathcal{G}]$ no longer necessarily exists for all $X \in \mathcal{L}^0_+$ if we insist that $\mathbb{E}[X|\mathcal{G}] < \infty$, a.s., instead of $\mathbb{E}[X|\mathcal{G}] \in [0, \infty]$, a.s.

Hint: The argument in the proof of Proposition 10.3 needs to be modified before it can be used.

Problem 10.6 (How to deal with the independent component)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a bounded Borel-measurable function, and let X and Y be independent random variables. Define the function $g : \mathbb{R} \to \mathbb{R}$ by

$$g(y) = \mathbb{E}[f(X, y)].$$

Show that the function g is Borel-measurable, and that

$$\mathbb{E}[f(X, Y)|Y = y] = g(y), \ \mu_Y - a.s.$$

Problem 10.7 (Some exercises in conditional probability)**.**

1. Let X, Y_1, Y_2 be random variables. Show that the random vectors (X, Y_1) and (X, Y_2) have the same distribution if and only if $\mathbb{P}[Y_1 \in B|\sigma(X)] = \mathbb{P}[Y_2 \in B|\sigma(X)]$, for all $B \in \mathcal{B}(\mathbb{R})$.
2. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative integrable random variables, and let $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ be sub-σ-algebras of $\mathcal{F}$. Show that $X_n \xrightarrow{\mathbb{P}} 0$ if $\mathbb{E}[X_n|\mathcal{F}_n] \xrightarrow{\mathbb{P}} 0$. Does the converse hold?

Hint: Prove that for $X_n \in \mathcal{L}^0_+$, we have $X_n \xrightarrow{\mathbb{P}} 0$ if and only if $\mathbb{E}[\min(X_n, 1)] \to 0$.

3. Let $\mathcal{G}$ be a complete sub-σ-algebra of $\mathcal{F}$. Suppose that for $X \in \mathcal{L}^1$, $\mathbb{E}[X|\mathcal{G}]$ and X have the same distribution. Show that X is $\mathcal{G}$-measurable.

Hint: Use the conditional Jensen's inequality.

Problem 10.8 (A characterization of $\mathcal{G}$-measurability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Show that for a random variable $X \in \mathcal{L}^1$ the following two statements are equivalent:

1. X is $\mathcal{G}$-measurable.
2. For all $\xi \in \mathcal{L}^\infty$, $\mathbb{E}[X\xi] = \mathbb{E}[X\mathbb{E}[\xi|\mathcal{G}]]$.

Problem 10.9 (Conditioning a part with respect to the sum)**.** Let $X_1, X_2, \dots$ be a sequence of iid r.v.'s with finite first moment, and let $S_n = X_1 + X_2 + \dots + X_n$. Define $\mathcal{G} = \sigma(S_n)$.

1. Compute $\mathbb{E}[X_1|\mathcal{G}]$.
2. Supposing, additionally, that X_1 is normally distributed, compute $\mathbb{E}[f(X_1)|\mathcal{G}]$, where $f : \mathbb{R} \to \mathbb{R}$ is a Borel function with $f(X_1) \in \mathbb{L}^1$.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 11
Discrete Martingales

Discrete-time filtrations and stochastic processes

One of the uses of σ-algebras is to single out the subsets of Ω to which probability can be assigned. This is the role of $\mathcal{F}$. Another use, as we have seen when discussing conditional expectation, is to encode information. The arrow of time, as we perceive it, points from less information to more information. A useful mathematical formalism is the one of a *filtration*.

Definition 11.1 (Filtration)**.** A **filtration** is a sequence $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, of sub-$\sigma$-algebras of $\mathcal{F}$ such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, for all $n \in \mathbb{N}_0$. A probability space with a filtration - $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}, \mathbb{P})$ - is called a **filtered probability space**.

We think of $n \in \mathbb{N}_0$ as the time-index and of $\mathcal{F}_n$ as the information available at time n.

Definition 11.2 (Stochastic Process)**.** A **(discrete-time) stochastic process** is simply a sequence $\{X_n\}_{n\in\mathbb{N}_0}$ of random variables.

A stochastic process is a generalization of a random vector; in fact, we can think of a stochastic processes as an infinite-dimensional random vector. More precisely, a stochastic process is a random element in the space $\mathbb{R}^{\mathbb{N}_0}$ of real sequences. In the context of stochastic processes, for each $\omega \in \Omega$, the sequence $(X_0(\omega), X_1(\omega), \dots)$ is called a **trajectory** of the stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$. This dual view of stochastic processes - as random trajectories (sequences) or as sequences of random variables - can be supplemented by another interpretation: a stochastic process is also a map from the product space $\Omega \times \mathbb{N}_0$ into $\mathbb{R}$.

Definition 11.3 (Adaptedness)**.** A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ is said to be **adapted** with respect to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ if X_n is $\mathcal{F}_n$-measurable for each $n \in \mathbb{N}_0$.

Intuitively, the process $\{X_n\}_{n\in\mathbb{N}_0}$ is adapted with respect to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ if its value X_n is fully known at time n (assuming that the time-n information is given by $\mathcal{F}_n$).

The most common way of producing filtrations is by generating them from stochastic processes. More precisely, for a stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$, the filtration $\{\mathcal{F}^X_n\}_{n\in\mathbb{N}_0}$, given by

$$\mathcal{F}^X_n = \sigma(X_0, X_1, \dots, X_n),\ n \in \mathbb{N}_0,$$

is called the **filtration generated by** $\{X_n\}_{n\in\mathbb{N}_0}$ or the **natural filtration** of $\{X_n\}_{n\in\mathbb{N}_0}$. Clearly, X is always adapted to its own natural filtration.

Martingales

Definition 11.4 (Sub-, Super-, Martingale)**.** Let $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a filtration. A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ is called an $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$-**supermartingale** if

1. $\{X_n\}_{n\in\mathbb{N}_0}$ is $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$-adapted,
2. $X_n \in \mathcal{L}^1$, for all $n \in \mathbb{N}_0$, and
3. $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$, a.s., for all $n \in \mathbb{N}_0$.

A process $\{X_n\}_{n\in\mathbb{N}_0}$ is called a **submartingale** if $\{-X_n\}_{n\in\mathbb{N}_0}$ is a supermartingale. A **martingale** is a process which is both a supermartingale and a submartingale at the same time, i.e., for which the equality holds in 3. above.

Remark 11.5*.* Very often, the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ is not explicitly mentioned. Then, it is often clear from the context what filtration should be used. Many times, the existence of an underlying filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ is assumed throughout. Alternatively, if no filtration is pre-specified, the filtration $\{\mathcal{F}^X_n\}_{n\in\mathbb{N}_0}$, generated by $\{X_n\}_{n\in\mathbb{N}_0}$ is used. It is important to remember, however, that the notion of a (super-, sub-) martingale only makes sense in relation to a filtration, and that different filtrations give rise to different families of martingales.

The fundamental examples of martingales are (additive or multiplicative) random walks:

Example 11.6.

1. *An additive random walk.* Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of iid random variables with $\xi_n \in \mathcal{L}^1$ and $\mathbb{E}[\xi_n] = 0$, for all $n \in \mathbb{N}$. We define

$$X_0 = 0,\ X_n = \sum_{k=1}^{n} \xi_k,\ \text{for } n \in \mathbb{N}.$$

The process $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale with respect to the filtration $\{\mathcal{F}_n^X\}_{n\in\mathbb{N}_0}$ generated by it (which is the same as $\sigma(\xi_1,\dots,\xi_n)$). Indeed, we clearly have $X_n \in \mathcal{L}^1(\mathcal{F}_n^X)$ for all $n\in\mathbb{N}$, and

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[\xi_{n+1}+X_n|\mathcal{F}_n] = \\ &= X_n + \mathbb{E}[\xi_{n+1}|\mathcal{F}_n] = X_n + \mathbb{E}[\xi_{n+1}] = X_n, \text{ a.s.,}\end{aligned}$$

where we used the "irrelevance of independent information"-property of conditional expectation (in this case ξ_{n+1} is independent of $\mathcal{F}_n^X = \sigma(X_0,\dots,X_n) = \sigma(\xi_1,\dots,\xi_n)$).

It is easy to see that if $\{\xi_n\}_{n\in\mathbb{N}}$ are still iid, but $\mathbb{E}[\xi_n] > 0$, then $\{X_n\}_{n\in\mathbb{N}_0}$ is a submartingale. When $\mathbb{E}[\xi_n] < 0$, we get a supermartingale.

2. *A multiplicative random walk.* Let $\{\xi_n\}_{n\in\mathbb{N}}$ be an iid sequence in $\mathcal{L}^1$ such that $\mathbb{E}[\xi_n] = 1$. We define

$$X_0 = 1,\ X_n = \prod_{k=1}^n \xi_k, \text{ for } n\in\mathbb{N}.$$

The process $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale with respect to the filtration $\{\mathcal{F}_n^X\}_{n\in\mathbb{N}_0}$ generated by it. Indeed, $X_n \in \mathcal{L}^1(\mathcal{F}_n^X)$ for all $n\in\mathbb{N}$ and

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[\xi_{n+1}X_n|\mathcal{F}_n] = \\ &= X_n\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] = X_n\mathbb{E}[\xi_{n+1}] = X_n, \text{ a.s.,}\end{aligned}$$

where, in addition to the "irrelevance of independent information" we also used "pulling out what's known".

Is it true that, if $\mathbb{E}[\xi_n] > 1$, we get a submartingale and that if $\mathbb{E}[\xi_n] < 1$, we get a supermartingale?

3. *Wald's martingales.* Let $\{\xi_n\}_{n\in\mathbb{N}}$ be an independent sequence, and let $\varphi_n(t)$ be the characteristic function of ξ_n. Assuming that[1] $\varphi_n(t) \neq 0$, for all $t\in\mathbb{R}$ and $n\in\mathbb{N}$, the previous example implies that the process $\{X_n\}_{n\in\mathbb{N}_0}$, defined by

$$X_0^t = 1,\ X_n^t = \prod_{k=1}^n \frac{e^{it\xi_k}}{\varphi_k(t)},\ n\in\mathbb{N},$$

is a martingale. Actually, it is complex-valued, so it would be better to say that its real and imaginary parts are both martingales. This martingale will be important in the study of hitting times of random walks.

4. *Lévy martingales.* For $X\in\mathcal{L}^1$, we define

$$X_n = \mathbb{E}[X|\mathcal{F}_n], \text{ for } n\in\mathbb{N}_0.$$

[1] Can you think of a useful sufficient conditions for this?

The tower property of conditional expectation implies that $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale.

5. *An urn scheme.* A nonempty urn contains b black and w white balls on day $n = 0$. On each subsequent day, a ball is chosen at random from the urn (each ball in the urn has the same probability of being picked) and then put back together with another ball of the same color. Therefore, at the end of day n, here are $n + b + w$ balls in the urn. Let B_n denote the number of black balls in the urn at day n, and let define the process $\{X_n\}_{n\in\mathbb{N}_0}$ by

$$X_n = \tfrac{B_n}{b+w+n},\ n \in \mathbb{N}_0,$$

to be the proportion of black balls in the urn at time n. Let $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ denote the filtration generated by $\{X_n\}_{n\in\mathbb{N}_0}$. The conditional probability - given $\mathcal{F}_n$ - of picking a black ball at time n is X_n, i.e.,

$$\mathbb{P}[B_{n+1} = B_n + 1|\mathcal{F}_n] = X_n \text{ and } \mathbb{P}[B_{n+1} = B_n|\mathcal{F}_n] = 1 - X_n.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_{n+1}\mathbf{1}_{\{B_{n+1}=B_n\}}|\mathcal{F}_n] + \mathbb{E}[X_{n+1}\mathbf{1}_{\{B_{n+1}=B_n+1\}}|\mathcal{F}_n] \\
&= \mathbb{E}[\tfrac{B_n}{b+w+n+1}\mathbf{1}_{\{B_{n+1}=B_n\}}|\mathcal{F}_n] + \mathbb{E}[\tfrac{B_n+1}{b+w+n+1}\mathbf{1}_{\{B_{n+1}=B_n+1\}}|\mathcal{F}_n] \\
&= \tfrac{B_n}{b+w+n+1}(1 - X_n) + \tfrac{B_n+1}{b+w+n+1}X_n \\
&= \tfrac{B_n(1-X_n)+(B_n+1)X_n}{b+w+n+1} = X_n
\end{aligned}$$

and $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale. How does this square with your intuition? Should not a high number of black balls translate into a high probability of picking a black ball? This will, in turn, only increase the number of black balls with high probability. In other words, why is $\{X_n\}_{n\in\mathbb{N}_0}$ not a submartingale (which is not a martingale), at least for large $\frac{b}{b+w}$?

To get some feeling for the definition, here is a simple exercise:

Problem 11.1.

1. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale. Show that $\mathbb{E}[X_n] = \mathbb{E}[X_0]$, for all $n \in \mathbb{N}_0$. Give an example of an adapted process $\{Y_n\}_{n\in\mathbb{N}_0}$ with $Y_n \in \mathcal{L}^1$, for all $n \in \mathbb{N}_0$ which is *not* a martingale, but $\mathbb{E}[Y_n] = \mathbb{E}[Y_0]$, for all $n \in \mathbb{N}_0$.

2. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale. Show that

$$\mathbb{E}[X_m|\mathcal{F}_n] = X_n, \text{ for all } m > n.$$

3. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a submartingale, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\varphi(X_n) \in \mathcal{L}^1$, for all $n \in \mathbb{N}_0$. Show that $\{\varphi(X_n)_n\}_{n\in\mathbb{N}_0} n \in \mathbb{N}_0$ is a submartingale, provided that either

(a) φ is nondecreasing, or

(b) $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale.

In particular, if $\{X_n\}_{n\in\mathbb{N}}$ is a submartingale, so are $\{X_n^+\}_{n\in\mathbb{N}_0}$ and $\{e^{X_n}\}_{n\in\mathbb{N}_0}$.

Hint: Use conditional Jensen's inequality.

Predictability and martingale transforms

Definition 11.7 (Predictability)**.** A process $\{H_n\}_{n\in\mathbb{N}}$ is said to be **predictable** with respect to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ if H_n is $\mathcal{F}_{n-1}$-measurable for $n \in \mathbb{N}$.

A process is predictable if you can predict its tomorrow's value today. We often think of predictable processes as strategies: let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of random variables which we interpret as gambles. At time n we can place a bet of H_n dollars, thus realizing a gain/loss of $H_n\xi_n$. Note that a negative H_n is allowed - the player wins money if $\xi_n < 0$ and loses if $\xi_n > 0$ in that case. If $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ is a filtration generated by the gambles, i.e., $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma\{\xi_1, \dots, \xi_n\}$, for $n \in \mathbb{N}$, then $H_n \in \mathcal{F}_{n-1}$, so that it does not use any information about ξ_n: we are allowed to adjust our bet according to the outcomes of previous gambles, but we don't know the outcome of ξ_n until after the bet is placed. Therefore, the sequence $\{H_n\}_{n\in\mathbb{N}}$ is a predictable sequence with respect to $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$.

Problem 11.2. Characterize predictable submartingales and predictable martingales.

Note: To comply with the setting in which the definition of predictability is given (processes defined on $\mathbb{N}$ and not on $\mathbb{N}_0$), simply discard the value at 0.

Definition 11.8 (Martingale transform)**.** Let $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a filtration and let $\{X_n\}_{n\in\mathbb{N}_0}$ be a process adapted to $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$. The stochastic process $\{(H\cdot X)_n\}_{n\in\mathbb{N}_0}$, defined by

$$(H\cdot X)_0 = 0,\ (H\cdot X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1}),\ \text{for } n \in \mathbb{N},$$

is called the **martingale transform of** X **by** H.

Remark 11.9.

1. The process $\{(H\cdot X)_n\}_{n\in\mathbb{N}_0}$ is called the martingale transform of X, even if neither H nor X is a martingale. It is most often applied to a martingale X, though - hence the name.
2. In terms of the gambling interpretation given above, X plays the role of the cumulative gain (loss) when a \$1-bet is placed each time:

$$X_0 = 0,\ X_n = \sum_{k=1}^{n} \xi_k,\ \text{where } \xi_k = X_n - X_{n-1},\ \text{for } n \in \mathbb{N}.$$

If we insist that $\{\xi_n\}_{n\in\mathbb{N}}$ is a sequence of *fair* bets, i.e., that there are no expected gains/losses in the n-th bet, even after we had the opportunity to learn from the previous $n-1$ bets, we arrive to the condition

$$\mathbb{E}[\xi_n|\mathcal{F}_{n-1}] = 0, \text{ i.e., that } \{X_n\}_{n\in\mathbb{N}_0} \text{ is a martingale.}$$

The following proposition states that no matter how well you choose your bets, you cannot make (or loose) money by betting on a sequence of fair games. A part of result is stated for submartingales; this is for convenience only. The reader should observe that almost any statement about submartingales can be turned into a statement about supermartingales by a simple change of sign.

Proposition 11.10 (Martingale transforms of (sub)martingales)**.** *Suppose that $\{X_n\}_{n\in\mathbb{N}_0}$ is adapted, and $\{H_n\}_{n\in\mathbb{N}}$ predictable. Then, the martingale transform $H\cdot X$ of X by H is*

1. *a martingale, provided that $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale and $H_n(X_n - X_{n-1}) \in \mathcal{L}^1$, for all $n\in\mathbb{N}$,*
2. *a submartingale, provided that $\{X_n\}_{n\in\mathbb{N}_0}$ is a submartingale, $H_n \geq 0$, a.s., and $H_n(X_n - X_{n-1}) \in \mathcal{L}^1$, for all $n\in\mathbb{N}$.*

Proof. Just check the definition and use properties of conditional expectation. □

Remark 11.11. The martingale transform is the discrete-time analogue of the *stochastic integral*. Note that it is crucial that H be predictable if we want a martingale transform of a martingale to be a martingale. Otherwise, we just take $H_n = \operatorname{sgn}(X_n - X_{n-1}) \in \mathcal{F}_n$ and obtain a process which is not a martingale unless X is constant. This corresponds to a player who knows the outcome of the game before the bet is placed and places the bet of $\$ \pm 1$ which is guaranteed to win.

Stopping times

Definition 11.12 (Stopping time)**.** A random variable T with values in $\mathbb{N}_0 \cup \{\infty\}$ is called a **random time**. A random time is said to be a **stopping time** with respect to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ if

$$\{T \leq n\} \in \mathcal{F}_n, \text{ for all } n\in\mathbb{N}.$$

Remark 11.13.

1. Stopping times are simply random instances with the property that at every instant you can answer the question "Has T already happened?" using only the currently-available information.

2. The additional element $+\infty$ is used as a placeholder for the case when T "does not happen".

Example 11.14.

1. Constant (deterministic) times $T = m$, $m \in \mathbb{N}_0 \cup \{\infty\}$ are obviously stopping times. The set of all stopping times can be thought of as an enlargement of the set of "time-instances". The meaning of "when Red Sox win the World Series again" is clear, but it does not correspond to a deterministic time.

2. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a stochastic process adapted to $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$. For a subset $B \in \mathcal{B}(\mathbb{R})$, we define the random time T_B by

$$T_B = \min\{n \in \mathbb{N}_0 \,:\, X_n \in B\}.$$

T_B is called the **hitting time** of the set B and is a stopping time. Indeed,

$$\{T_B \leq n\} = \{X_0 \in B\} \cup \{X_1 \in B\} \cup \cdots \cup \{X_n \in B\} \in \mathcal{F}_n.$$

3. Let $\{\xi_n\}_{n\in\mathbb{N}}$ be an iid sequence of coin tosses, i.e. $\mathbb{P}[\xi_i = 1] = \mathbb{P}[\xi_i = -1] = \frac{1}{2}$, and let $X_n = \sum_{k=1}^{n} \xi_k$ be the corresponding random walk. For $N \in \mathbb{N}$, let S be the random time defined by

$$S = \max\{n \leq N \,:\, X_n = 0\}.$$

S is called the **last visit time** to 0 before time $N \in \mathbb{N}$. Intuitively, S is not a stopping time since, in order to know whether S had already happened at time $m < N$, we need to know that $X_k \neq 0$, for $k = m+1, \dots, N$, and, for that, we need the information which is not contained in $\mathcal{F}_m$. We leave it to the reader to make this comment rigorous.

Stopping times have good stability properties, as the following proposition shows. All stopping times are with respect to an arbitrary, but fixed filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$.

Proposition 11.15 (Stability properties of stopping times)**.**

1. *A random time T is a stopping time if and only if the process $X_n = \mathbf{1}_{\{n \geq T\}}$ is $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$-adapted.*
2. *If S and T are stopping times, then so are $S+T$, $\max(S,T)$, $\min(S,T)$.*
3. *Let $\{T_n\}_{n\in\mathbb{N}}$ be a sequence of stopping times such that $T_1 \leq T_2 \leq \dots$, a.s. Then $T = \sup_n T_n$ is a stopping time.*
4. *Let $\{T_n\}_{n\in\mathbb{N}}$ be a sequence of stopping times such that $T_1 \geq T_2 \geq \dots$, a.s. Then $T = \inf_n T_n$ is a stopping time.*

Proof.

1. Immediate.
2. Let us show that $S+T$ is a stopping time and leave the other two to the reader:
$$\{S+T \leq n\} = \cup_{k=0}^{n} (\{S \leq k\} \cap \{T \leq n-k\}) \in \mathcal{F}_n.$$
3. For $m \in \mathbb{N}_0$, we have $\{T \leq m\} = \cap_{n\in\mathbb{N}} \{T_n \leq m\} \in \mathcal{F}_m$.
4. For $m \in \mathbb{N}_0$, we have $\{T_n \geq m\} = \{T_n < m\}^c = \{T_n \leq m-1\}^c \in \mathcal{F}_{m-1}$. Therefore,
$$\begin{aligned}\{T \leq m\} &= \{T < m+1\} = \{T \geq m+1\}^c \\ &= \cup_{n\in\mathbb{N}} \{T_n \geq m+1\}^c \in \mathcal{F}_m.\end{aligned}$$ □

Stopping times are often used to produce new processes from old ones. The most common construction runs the process X until time T and after that keeps it constant and equal to its value at time T. More precisely:

Definition 11.16 (Stopped process)**.** Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a stochastic process, and let T be a stopping time. The process $\{X_n\}_{n\in\mathbb{N}_0}$ **stopped at** T, denoted by $\{X_n^T\}_{n\in\mathbb{N}_0}$ is defined by
$$X_n^T(\omega) = X_{T(\omega)\wedge n}(\omega) = X_n(\omega)\mathbf{1}_{\{n \leq T(\omega)\}} + X_{T(\omega)}\mathbf{1}_{\{n > T(\omega)\}}.$$

Most of what is great about the martingale theory can be traced back to the simple observation that the (sub)martingale property is stable under stopping:

Proposition 11.17 (Stopped (sub)martingales are (sub)martingales)**.** *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a (sub)martingale, and let T be a stopping time. Then the stopped process $\{X_n^T\}_{n\in\mathbb{N}}$ is also a (sub)martingale.*

Proof. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a (sub)martingale. We note that the process $K_n = \mathbf{1}_{\{n\leq T\}}$, is predictable, non-negative and bounded, so its martingale transform $(K \cdot X)$ is a (sub)martingale. Moreover,

$$(K \cdot X)_n = X_{T\wedge n} - X_0 = X^T_n - X_0,$$

and so, X^T is a (sub)martingale, as well. □

Convergence of martingales

A judicious use of a predictable processes in a martingale transform yields the following important result:

Theorem 11.18. (Martingale convergence) *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale such that*

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}[|X_n|] < \infty.$$

Then, there exists a random variable $X \in \mathcal{L}^1(\mathcal{F})$ such that $X_n \to X$, a.s.

Proof. We pick two real numbers $a < b$ and define two sequences of stopping times as follows (see pictures on the next page):

$$\begin{aligned} &T_0 = 0, \\ &S_1 = \inf\{n \geq 0 : X_n \leq a\}, \quad T_1 = \inf\{n \geq S_1 : X_n \geq b\} \\ &S_2 = \inf\{n \geq T_1 : X_n \leq a\}, \quad T_2 = \inf\{n \geq S_2 : X_n \geq b\}, \text{ etc.} \end{aligned} \tag{11.0}$$

In words, let S_1 be the first time X falls under a. Then, T_1 is the first time after S_1 when X exceeds b, etc. We leave it to the reader to check that $\{T_n\}_{n\in\mathbb{N}}$ and $\{S_n\}_{n\in\mathbb{N}}$ are stopping times. These two sequences of stopping times allow us to construct a predictable process $\{H_n\}_{n\in\mathbb{N}}$ which takes values in $\{0, 1\}$. Simply, we "buy low and sell high":

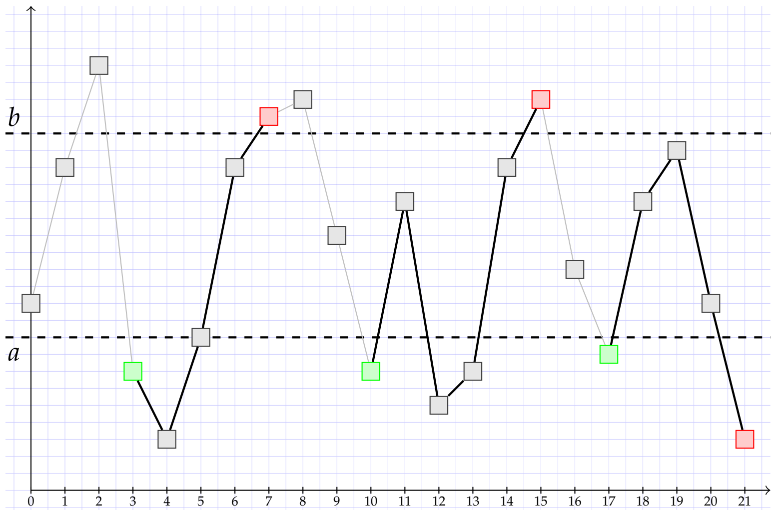
$$H_n = \sum_{k\in\mathbb{N}} \mathbf{1}_{\{S_k<n\leq T_k\}} = \begin{cases} 1, & S_k < n \leq T_k \text{ for some } k \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

Let $U^{a,b}_n$ be the number of " completed upcrossings by time n", i.e., the process defined by

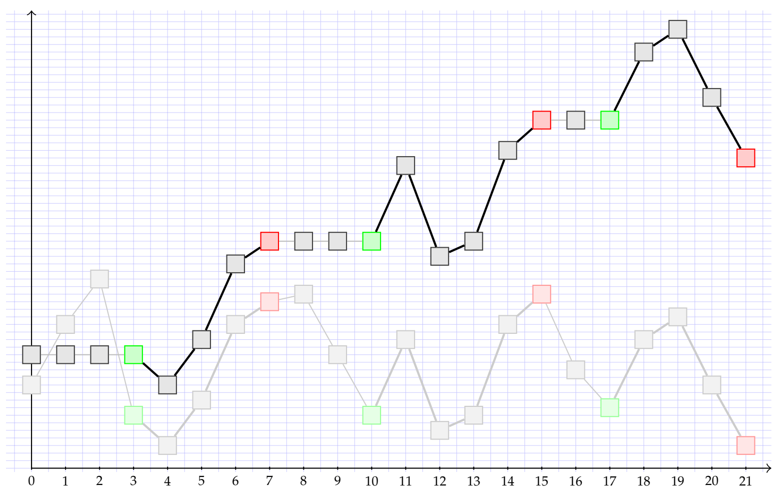
$$U^{a,b}_n = \inf\{k \in \mathbb{N} : T_k \leq n\}.$$

A bit of accounting yields:

$$(H \cdot X)_n \geq (b - a)U^{a,b}_n - (X_n - a)^-. \tag{11.1}$$



The green squares represent the stopping times $S_1, S_2, \dots$ and red squares represent $T_1, T_2, \dots$. Bold lines indicate that $H = 1$ and the thin ones that $H = 0$.



On the left is the graph of a typical trajectory of the process Y, obtained by "investing in" X when $H = 1$ and doing nothing when $H = 0$.

Indeed, the total gains from the strategy H can be split into two components. First, every time a passage from below a to above b is completed, we pocket at least $(b-a)$. After that, if X never falls below a again, H remains 0 and our total gains exceeds $(b-a)U_n^{a,b}$, which, in turn, trivially dominates $(b-a)U_n^{a,b} - (X_n - a)^-$. The other possibility is that after the last upcrossing, the process does reach the value below a at a certain point. The last upcrossing already happened, so the process never hits a value above b after that; it may very well happen that we "lose" on this transaction. The loss is overestimated by $(X_n - a)^-$.

Then the inequality (11.1) and fact that the martingale transform (by a bounded process) of a martingale is a martingale yield

$$\mathbb{E}[U_n^{a,b}] \leq \tfrac{1}{b-a}\mathbb{E}[(H \cdot X)_n] + \tfrac{1}{b-a}\mathbb{E}[(X_n - a)^-] \leq \tfrac{\mathbb{E}[|X_n|]+|a|}{b-a}.$$

Let $U_\infty^{a,b}$ be the total number of upcrossings, i.e., $U_\infty^{a,b} = \lim_n U_n^{a,b}$. Using the monotone convergence theorem, we get the so-called **upcrossings inequality**

$$\mathbb{E}[U_\infty^{a,b}] \leq \tfrac{|a|+\sup_{n\in\mathbb{N}_0}\mathbb{E}[|X_n|]}{b-a}$$

The assumption that $\sup_{n\in\mathbb{N}_0}\mathbb{E}[|X_n|] < \infty$, implies that $\mathbb{E}[U_\infty^{a,b}] < \infty$ and so $\mathbb{P}[U_\infty^{a,b} < \infty] = 1$. In words, the number of upcrossings is almost surely finite (otherwise, we would be able to make money by betting on an unfair game).

It remains to use the fact that $U_\infty^{a,b} < \infty$, a.s., to deduce that $\{X_n\}_{n\in\mathbb{N}_0}$ converges. First of all, by passing to rational numbers and taking countable intersections of probability-one sets, we can assert that

$$\mathbb{P}[U_\infty^{a,b} < \infty, \text{ for all } a < b \text{ rational}] = 1.$$

Then, we assume, contrary to the statement, that $\{X_n\}_{n\in\mathbb{N}_0}$ does not converge, so that

$$\mathbb{P}[\liminf_n X_n < \limsup_n X_n] > 0.$$

This can be strengthened to

$$\mathbb{P}[\liminf_n X_n < a < b < \limsup_n X_n, \text{ for some } a < b \text{ rational}] > 0,$$

which is, however, a contradiction since, on the event $\{\liminf_n X_n < a < b < \limsup_n X_n\}$, the process X completes infinitely many upcrossings.

We conclude that there exists an $[-\infty,\infty]$-valued random variable X_∞ such that $X_n \overset{a.s.}{\to} X_\infty$. In particular, we have $|X_n| \overset{a.s.}{\to} |X_\infty|$, and Fatou's lemma yields

$$\mathbb{E}[|X_\infty|] \leq \liminf_n \mathbb{E}[|X_n|] \leq \sup_n \mathbb{E}[|X_n|] < \infty. \qquad \square$$

Some, but certainly not all, results about martingales can be transferred to submartingales (supermartingales) using the following proposition:

Proposition 11.19 (Doob-Meyer decomposition). *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a submartingale. Then, there exists a martingale $\{M_n\}_{n\in\mathbb{N}_0}$ and a predictable process $\{A_n\}_{n\in\mathbb{N}}$ (with $A_0 = 0$ adjoined) such that $A_n \in \mathcal{L}^1$, $A_n \leq A_{n+1}$, a.s., for all $n \in \mathbb{N}_0$ and*

$$X_n = M_n + A_n, \text{ for all } n \in \mathbb{N}_0.$$

Proof. Define

$$A_n = \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1}|\mathcal{F}_{k-1}], \quad n \in \mathbb{N}.$$

Then $\{A_n\}_{n\in\mathbb{N}}$ is clearly predictable and $A_{n+1} \geq A_n$, a.s., thanks to the submartingale property of X. Finally, set $M_n = X_n - A_n$, so that

$$\begin{aligned}\mathbb{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}] &= \mathbb{E}[X_n - X_{n-1} - (A_n - A_{n-1})|\mathcal{F}_{n-1}] \\ &= \mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] - (A_n - A_{n-1}) = 0. \quad \square\end{aligned}$$

Corollary 11.20 (Submartingale convergence)**.** *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a submartingale such that*

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}[X_n^+] < \infty.$$

Then, there exists a random variable $X \in \mathcal{L}^1(\mathcal{F})$ such that $X_n \to X$, a.s.

Proof. Let $\{M_n\}_{n\in\mathbb{N}_0}$ and $\{A_n\}_{n\in\mathbb{N}}$ be as in Proposition 11.10. Since $M_n = X_n - A_n \leq X_n$, a.s., we have $\mathbb{E}[M_n^+] \leq \mathbb{E}[X_n^+] \leq \sup_n \mathbb{E}[X_n^+] < \infty$. Finally, since $\mathbb{E}[M_n^-] = \mathbb{E}[M_n^+] - \mathbb{E}[M_n] \leq \sup_n \mathbb{E}[X_n^+] - \mathbb{E}[M_0] < \infty$, we have

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] < \infty.$$

Therefore, $M_n \overset{a.s.}{\to} M_\infty$, for some $M_\infty \in \mathcal{L}^1$. Since $\{A_n\}_{n\in\mathbb{N}}$ is non-negative and non-decreasing, there exists $A_\infty \in \mathcal{L}^0_+$ such that $A_n \to A_\infty \geq 0$, a.s., and so

$$X_n \overset{a.s.}{\to} X_\infty = M_\infty + A_\infty.$$

It remains to show that $X_\infty \in \mathcal{L}^1$, and for that, it suffices to show that $\mathbb{E}[A_\infty] < \infty$. Since $\mathbb{E}[A_n] = \mathbb{E}[X_n] - \mathbb{E}[M_n] \leq C = \sup_n \mathbb{E}[X_n^+] + \mathbb{E}[|M_0|] < \infty$, monotone convergence theorem yields $\mathbb{E}[A_\infty] \leq C < \infty$. $\square$

Remark 11.21. Corollary 11.20 - or the simple observation that $\mathbb{E}[|X_n|] = 2\mathbb{E}[X_n^+] - \mathbb{E}[X_0]$ when $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ - implies that it is enough to assume $\sup_n \mathbb{E}[X_n^+] < \infty$ in the original martingale-convergence theorem (Theorem 11.18).

Corollary 11.22 (Convergence under positivity)**.** *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a non-negative supermartingale (or a non-positive submartingale). Then there exists a random variable $X \in \mathcal{L}^1(\mathcal{F})$ such that $X_n \to X$, a.s.*

To convince yourself that things can go wrong if the boundedness assumptions are not met, here is a problem:

Problem 11.3. Give an example of a submartingale $\{X_n\}_{n\in\mathbb{N}}$ with the property that $X_n \to -\infty$, a.s. and $\mathbb{E}[X_n] \to \infty$.

Hint: Use the Borel-Cantelli lemma.

Additional problems

Problem 11.4 (Combining (super)martingales)**.** In 1., 2. and 3. below, $\{X_n\}_{n\in\mathbb{N}_0}$ and $\{Y_n\}_{n\in\mathbb{N}_0}$ are martingales. In 4., they are only supermartingales.

1. Show that the process $\{Z_n\}_{n\in\mathbb{N}_0}$ given by

$$Z_n = X_n \vee Y_n = \max\{X_n, Y_n\}$$

is a submartingale.

2. Give an example of $\{X_n\}_{n\in\mathbb{N}_0}$ and $\{Y_n\}_{n\in\mathbb{N}_0}$ such that $\{Z_n\}_{n\in\mathbb{N}_0}$ (defined above) is *not* a martingale.

3. Does the product $\{X_nY_n\}_{n\in\mathbb{N}_0}$ have to be a martingale? A submartingale? A supermartingale? (Provide proofs or counterexamples).

4. Let T be an $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$-stopping time (and remember that $\{X_n\}_{n\in\mathbb{N}_0}$ and $\{Y_n\}_{n\in\mathbb{N}_0}$ are supermartingales). Show that the process $\{Z_n\}_{n\in\mathbb{N}_0}$, given by

$$Z_n(\omega) = \begin{cases} X_n(\omega), & n < T(\omega) \\ Y_n(\omega), & n \geq T(\omega), \end{cases}$$

is a supermartingale, provided that $X_T \geq Y_T$, a.s.

Note: This result is sometimes called the *switching principle*. It says that if you switch from one supermartingale to a smaller one at a stopping time, the resulting process is still a supermartingale.

Problem 11.5 (An urn model)**.** A nonempty urn contains $B_0 \in \mathbb{N}$ black and $W_0 \in \mathbb{N}$ white balls at time 0. At each time we draw a ball (each ball in the urn has the same probability of being picked), throw it away, and replace it with $C \in \mathbb{N}$ balls of the same color. Let B_n denote the number of black balls at time n, and let X_n denote the proportion of black balls in the urn.

1. Show that there exists a random variable X such that

$$X_n \stackrel{\mathbb{L}^p}{\to} X, \text{ for all } p \in [1, \infty).$$

2. Find an expression for $\mathbb{P}[B_n = k]$, $k = 1, \dots, n+1$, $n \in \mathbb{N}_0$, when $B_0 = W_0 = 1, C = 2$, and use it to determine the distribution of X in that case.

Hint: Guess the form of the solution for $n = 1, 2, 3$, and prove that you are correct for all n.

Problem 11.6 (Stabilization of integer-valued submartingales)**.**

1. Let $\{M_n\}_{n\in\mathbb{N}_0}$ be an integer-valued (i.e., $\mathbb{P}[M_n \in \mathbb{Z}] = 1$, for $n \in \mathbb{N}_0$) submartingale bounded from above. Show that there exists an $\mathbb{N}_0$-valued random variable T with the property that

$$\forall\, n \in \mathbb{N}_0,\ M_n = M_T \text{ on } \{T \leq n\}, \text{ a.s.}$$

2. Can such T always be found in the class of stopping times?

Note: This is quite a bit harder than the other two parts.

3. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple biased random walk, i.e.,

$$X_0 = 0,\ X_n = \sum_{k=1}^{n} \xi_k,\ n \in \mathbb{N},$$

where $\{\xi_n\}_{n\in\mathbb{N}}$ are iid with $\mathbb{P}[\xi_1 = 1] = p$ and $\mathbb{P}[\xi_1 = -1] = 1-p$, for some $p \in (0,1)$. Under the assumption that $p \geq \frac{1}{2}$, show that X hits any nonnegative level with probability 1, i.e., that for $a \in \mathbb{N}$, we have $\mathbb{P}[\tau_a < \infty] = 1$, where $\tau_a = \inf\{n \in \mathbb{N} : X_n = a\}$.

Problem 11.7 (An application to gambling)**.** Let $\{\varepsilon_n\}_{n\in\mathbb{N}_0}$ be an iid sequence with $\mathbb{P}[\varepsilon_n = 1] = 1 - \mathbb{P}[\varepsilon_n = -1] = p \in (\frac{1}{2}, 1)$. We interpret $\{\varepsilon_n\}_{n\in\mathbb{N}_0}$ as outcomes of a series of gambles. A gambler starts with $Z_0 > 0$ dollars, and in each play wagers a certain portion of her wealth. More precisely, the wealth of the gambler at time $n \in \mathbb{N}$ is given by

$$Z_n = Z_0 + \sum_{k=1}^{n} C_k \varepsilon_k,$$

where $\{C_n\}_{n\in\mathbb{N}_0}$ is a predictable process such that $C_k \in [0, Z_{k-1})$, for $k \in \mathbb{N}$. The goal of the gambler is to maximize the "return" on her wealth, i.e., to choose a strategy $\{C_n\}_{n\in\mathbb{N}_0}$ such that the expectation $\frac{1}{T}\mathbb{E}[\log(Z_T/Z_0)]$, where $T \in \mathbb{N}$ is some fixed time horizon, is the maximal possible.

Note: It makes sense to call the random variable R such that $Z_T = Z_0 e^{RT}$ the *return*, and, consequently $\mathbb{E}[R] = \frac{1}{T}\mathbb{E}[\log(Z_T/Z_0)]$, the *expected return*. Indeed, if you put Z_0 dollars in a bank and accrue (a compound) interest with rate $R \in (0,\infty)$, you will get $Z_0 e^{RT}$ dollars after T years. In our case, R is not deterministic anymore, but the interpretation still holds.

1. Define $\alpha = H(\frac{1}{2}) - H(p)$, where $H(p) = -p\log p - (1-p)\log(1-p)$ and show that the process $\{W_n\}_{n\in\mathbb{N}_0}$ given by

$$W_n = \log(Z_n) - \alpha n,\ \text{for } n \in \mathbb{N}_0$$

is a supermartingale. Conclude that

$$\mathbb{E}[\log(Z_T)] \leq \log(Z_0) + \alpha T,$$

for any choice of $\{C_n\}_{n\in\mathbb{N}_0}$.

2. Show that the upper bound above is attained for some strategy $\{C_n\}_{n\in\mathbb{N}_0}$.

Note: The quantity $H(p)$ is called the **entropy** of the distribution of ξ_1. This problem shows how it appears naturally in a gambling-theoretic context: the optimal rate of return equals to "excess" entropy $H(\frac{1}{2}) - H(p)$.

Problem 11.8 (An application to analysis)**.** Let $\Omega = [0,1)$, $\mathcal{F} = \mathcal{B}[0,1)$, and $\mathbb{P} = \lambda$, where λ denotes the Lebesgue measure on $[0,1)$. For $n \in \mathbb{N}$ and $k \in \{0,1,\dots,2^n-1\}$, we define

$$I_{k,n} = [k2^{-n}, (k+1)2^{-n}),\ \mathcal{F}_n = \sigma(I_{0,n}, I_{1,n}, \dots, I_{2^n-1,n}).$$

In words, $\mathcal{F}_n$ is generated by the n-th dyadic partition of $[0,1)$. For $x \in [0,1)$, let $k_n(x)$ be the unique number in $\{0,1,\dots,2^n-1\}$ such that $x \in I_{k_n(x),n}$. For a function $f : [0,1) \to \mathbb{R}$ we define the process $\{X^f_n\}_{n\in\mathbb{N}_0}$ by

$$X^f_n(x) = 2^n\Big(f\big((k_n(x)+1)2^{-n}\big) - f\big(k_n(x)2^{-n}\big)\Big),\ x \in [0,1).$$

1. Show that $\{X^f_n\}_{n\in\mathbb{N}_0}$ is a martingale.
2. Assume that the function f is Lipschitz, i.e., that there exists $K > 0$ such that $|f(y) - f(x)| \leq K|y-x|$, for all $x, y \in [0,1)$. Show that the limit $X^f = \lim_n X^f_n$ exists a.s.
3. Show that, for f Lipschitz, X^f has the property that

$$f(y) - f(x) = \int_x^y X^f(\xi)\,d\xi, \text{ for all } 0 \leq x < y < 1.$$

Note: This problem gives an alternative proof of the fact that Lipschitz functions are absolutely continuous.

Problem 11.9 (A martingale with strange properties)**.** Let the stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ be constructed as follows: $X_0 = 0$, and for $n \in \mathbb{N}$, and conditionally on $\mathcal{F}_{n-1} = \sigma(X_0, \dots, X_{n-1})$, we set

$$X_n \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{2n} & 1-\frac{1}{n} & \frac{1}{2n} \end{pmatrix} \text{ when } X_{n-1} = 0, \text{ and}$$

$$X_n \sim \begin{pmatrix} nX_{n-1} & 0 \\ \frac{1}{n} & 1-\frac{1}{n} \end{pmatrix} \text{ when } X_{n-1} \neq 0.$$

Show that $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale which converges to 0 in probability but not a.s.

Problem 11.10 (An explicit Doob-Meyer decomposition)**.** Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple random walk, i.e., $X_0 = 0$, $X_n = \sum_{k=1}^n \xi_k$, for $n \in \mathbb{N}$, where $\{\xi_n\}_{n\in\mathbb{N}}$ is an iid sequence with $\mathbb{P}[\xi_1 = -1] = \mathbb{P}[\xi_1 = 1] = \frac{1}{2}$. Let the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be given by $\mathcal{F}_0 = 0$, $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, for $n \in \mathbb{N}$.

Let $|X| = M + A$ be the Doob-Meyer decomposition of the submartingale $|X|$, into a martingale M with $M_0 = 0$ and a non-decreasing predictable process A. Show that M admits the form

$$M = H \cdot X, \tag{11.1}$$

for some predictable process H and find an explicit expression for H.

Problem 11.11 (A sequence of martingales)**.** Let $\{\{M^{(k)}_n\}_{n\in\mathbb{N}_0}\}_{k\in\mathbb{N}}$, be a sequence of non-negative martingales defined on the same filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}, \mathbb{P})$. Under the assumption that the limit $M_n(\omega) = \lim_{k\to\infty} M^{(k)}_n(\omega)$ exists and is finite for all $\omega \in \Omega$, prove that the process $\{M_n\}_{n\in\mathbb{N}_0}$ is a supermartingale, but not necessarily a martingale.

Problem 11.12 (An a.s.-convergent, but not $\mathcal{L}^1$-covergent martingale in $\mathcal{L}^p$)**.** Given $p \geq 1$, show that there exists a martingale $\{X_n\}_{n\in\mathbb{N}_0}$ such that

1. $X_n \in \mathcal{L}^p$, for all $n \in \mathbb{N}$.
2. $\{X_n\}_{n\in\mathbb{N}}$ converges a.s., as $n \to \infty$,
3. $\{X_n\}_{n\in\mathbb{N}}$ is *not* Cauchy in $\mathcal{L}^1$.

Course: Theory of Probability II
Term: Fall 2015
Instructor: Gordan Zitkovic

Lecture 12
UNIFORM INTEGRABILITY

Uniform integrability is a compactness-type concept for families of random variables, not unlike that of tightness.

Definition 12.1 (Uniform integrability)**.** A non-empty family $\mathcal{X} \subseteq \mathcal{L}^0$ of random variables is said to be **uniformly integrable (UI)** if

$$\lim_{K\to\infty} \left(\sup_{X\in\mathcal{X}} \mathbb{E}[|X| \mathbf{1}_{\{|X|\geq K\}}] \right) = 0.$$

Remark 12.2. It follows from the dominated convergence theorem (prove it!) that

$$\lim_{K\to\infty} \mathbb{E}[|X| \mathbf{1}_{\{|X|\geq K\}}] = 0 \text{ if and only if } X \in \mathcal{L}^1,$$

i.e., that for integrable random variables, far tails contribute little to the expectation. Uniformly integrable families are simply those for which the size of this contribution can be controlled uniformly over all elements.

We start with a characterization and a few basic properties of uniform-integrable families:

Proposition 12.3 (UI = $\mathcal{L}^1$-bounded + uniformly absolutely continuous)**.** *A family $\mathcal{X} \subseteq \mathcal{L}^0$ of random variables is uniformly integrable if and only if*

1. *there exists $C \geq 0$ such that $\mathbb{E}[|X|] \leq C$, for all $X \in \mathcal{X}$, and*
2. *for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any $A \in \mathcal{F}$, we have*

$$\mathbb{P}[A] \leq \delta \to \sup_{X\in\mathcal{X}} \mathbb{E}[|X| \mathbf{1}_A] \leq \varepsilon.$$

Proof. UI $\to$ 1., 2. Assume $\mathcal{X}$ is UI and choose $K > 0$ such that

$$\sup_{X\in\mathcal{X}} \mathbb{E}[|X| \mathbf{1}_{\{|X|>K\}}] \leq 1.$$

Since

$$\mathbb{E}[|X|] = \mathbb{E}[|X| \mathbf{1}_{\{|X|\leq K\}}] + \mathbb{E}[|X| \mathbf{1}_{\{|X|>K\}}] \leq K + \mathbb{E}[|X| \mathbf{1}_{\{|X|>K\}}],$$

for any X, we have $\sup_{X\in\mathcal{X}} \mathbb{E}[|X|] \leq K+1$, and 1. follows.

For 2., we take $\varepsilon > 0$ and use the uniform integrability of $\mathcal{X}$ to find a constant $K > 0$ such that $\sup_{X\in\mathcal{X}} \mathbb{E}[|X| \mathbf{1}_{\{|X|>K\}}] < \varepsilon/2$. For $\delta = \frac{\varepsilon}{2K}$ and $A \in \mathcal{F}$, the condition $\mathbb{P}[A] \leq \delta$ implies that

$$\begin{aligned}\mathbb{E}[|X| \mathbf{1}_A] &= \mathbb{E}[|X| \mathbf{1}_A \mathbf{1}_{\{|X|\leq K\}}] + \mathbb{E}[|X| \mathbf{1}_A \mathbf{1}_{\{|X|>K\}}]\\ &\leq K\mathbb{P}[A] + \mathbb{E}[|X| \mathbf{1}_{\{|X|>K\}}] \leq \varepsilon.\end{aligned}$$

1., 2. $\to$ UI. Let $C > 0$ be the bound from 1., pick $\varepsilon > 0$ and let $\delta > 0$ be such that 2. holds. For $K = \frac{C}{\delta}$, Markov's inequality gives

$$\mathbb{P}[|X| \geq K] \leq \tfrac{1}{K}\mathbb{E}[|X|] \leq \delta,$$

for all $X \in \mathcal{X}$. Therefore, by 2., $\mathbb{E}[|X| \mathbf{1}_{\{|X|\geq K\}}] \leq \varepsilon$ for all $X \in \mathcal{X}$. □

Remark 12.4. Boundedness in $\mathcal{L}^1$ is not enough for uniform integrability. For a counterexample, take $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$, and define

$$X_n(\omega) = \begin{cases} n, & \omega \leq \frac{1}{n}, \\ 0, & \text{otherwise}. \end{cases}$$

Then $\mathbb{E}[X_n] = n\frac{1}{n} = 1$, but, for $K > 0$, $\mathbb{E}[|X_n| \mathbf{1}_{\{|X_n|\geq K\}}] = 1$, for all $n \geq K$, so $\{X_n\}_{n\in\mathbb{N}}$ is not UI.

Problem 12.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be two uniformly-integrable families (on the same probability space). Show that the following families are also uniformly integrable:

1. $\{Z \in \mathcal{L}^0 : |Z| \leq |X| \text{ for some } X \in \mathcal{X}\}$.
2. $\{X + Y : X \in \mathcal{X}, Y \in \mathcal{Y}\}$.

Another useful characterization of uniform integrability uses a class of functions which converge to infinity faster than any linear function:

Definition 12.5 (Test function of UI)**.** A Borel function $\varphi : [0,\infty) \to [0,\infty)$ is called a **test function of uniform integrability** if

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = \infty.$$

Proposition 12.6 (Characterization of UI via test functions)**.** *A nonempty family $\mathcal{X} \subseteq \mathcal{L}^0$ is uniformly integrable if and only if there exists a test function of uniform integrability φ such that*

$$\sup_{X\in\mathcal{X}} \mathbb{E}[\varphi(|X|)] < \infty. \qquad (12.1)$$

Moreover, if it exists, the function φ can be chosen in the class of non-decreasing convex functions.

The proof of necessity rests on a simple fact from analysis:

Lemma 12.7 (Functions arbitrary close to integrability)**.** *Let $f : [0,\infty) \to [0,\infty)$ be a non-increasing function with $f(x) \to 0$, as $x \to \infty$. Then, there exists a continuous function $g : [0,\infty) \to (0,\infty)$ such that*

$$\int_0^\infty g(x)\,dx = +\infty \text{ but } \int_0^\infty f(x)g(x)\,dx < \infty. \tag{12.2}$$

Moreover, g can be chosen so that the function $x \mapsto x\int_0^x g(\xi)\,d\xi$ is convex.

Proof. Let $\tilde f$ be a strictly positive continuously differentiable with $\tilde f(x) \geq f(x)$, for all $x \geq 0$, with $\tilde f(x) \to 0$, as $x \to \infty$. With such $\tilde f$ and $g = -\tilde f'/\tilde f$, we have

$$\int_0^\infty g(x)\,dx = \lim_{x\to\infty} \Big(\ln(\tilde f(0)) - \ln(\tilde f(x)) \Big) = \infty.$$

On the other hand $\int_0^\infty f(x)g(x)\,dx \leq \int_0^\infty \tilde f(x)g(x)\,dx = \lim_{x\to\infty} \Big(\tilde f(0) - \tilde f(x)\Big) = \tilde f(0) < \infty$.

We leave it to the reader to argue that a (perhaps even larger) $\tilde f$ can be constructed such that $x \mapsto -x\ln \tilde f(x)$ is convex. □

Proof of Proposition 12.6. Suppose, first, that (12.1) holds for some test function of uniform integrability and that the value of the supremum is $0 \leq M < \infty$. For $n > 0$, there exists $C_n \in \mathbb{R}$ such that $\varphi(x) \geq nMx$, for $x \geq C_n$. Therefore,

$$M \geq \mathbb{E}[\varphi(|X|)] \geq \mathbb{E}[\varphi(|X|)\mathbf{1}_{\{|X|\geq C_n\}}] \geq nM\mathbb{E}[|X|\,\mathbf{1}_{\{|X|\geq C_n\}}],$$

for all $X \in \mathcal{X}$. Hence, $\sup_{X\in\mathcal{X}} \mathbb{E}[|X|\,\mathbf{1}_{\{|X|\geq C_n\}}] \leq \frac{1}{n}$, and the uniform integrability of $\mathcal{X}$ follows.

Conversely, if $\mathcal{X}$ is uniformly integrable the function

$$f(K) = \sup_{X\in\mathcal{X}} \mathbb{E}[|X|\,\mathbf{1}_{\{|X|\geq K\}}],$$

satisfies the conditions of Lemma 12.7 and a function g for which (12.2) holds can be constructed. Consequently, the function $\varphi(x) = x\int_0^x g(\xi)\,d\xi$ is a test-function of uniform integrability. On the other hand, for $X \in \mathcal{X}$, we have

$$\begin{aligned}\mathbb{E}[\varphi(|X|)] &= \mathbb{E}[|X|\int_0^\infty \mathbf{1}_{\{K\leq|X|\}} g(K)\,dK] \\ &= \int_0^\infty g(K)\mathbb{E}[|X|\,\mathbf{1}_{\{|X|\leq K\}}]\,dK \\ &= \int_0^\infty g(K)f(K)\,dK < \infty.\end{aligned}$$

□

Corollary 12.8 ($\mathcal{L}^p$-boundedness implies UI for $p > 1$)**.** *For $p > 1$, let $\mathcal{X}$ be a nonempty family of random variables bounded in $\mathcal{L}^p$, i.e., such that $\sup_{X \in \mathcal{X}} ||X||_{\mathcal{L}^p} < \infty$. Then $\mathcal{X}$ is uniformly integrable.*

Problem 12.2. Let $\mathcal{X}$ be a nonempty uniformly-integrable family in $\mathcal{L}^0$. Show that $\operatorname{conv} \mathcal{X}$ is uniformly-integrable, where $\operatorname{conv} \mathcal{X}$ is the smallest convex set in $\mathcal{L}^0$ which contains $\mathcal{X}$, i.e., $\operatorname{conv} \mathcal{X}$ is the set of all random variables of the form $X = \alpha_1 X_1 + \cdots + \alpha_n X_n$, for $n \in \mathbb{N}$, $\alpha_k \geq 0$, $k = 1, \ldots, n$, $\sum_{k=1}^n \alpha_k = 1$ and $X_1, \ldots, X_n \in \mathcal{X}$.

Problem 12.3. Let $\mathcal{C}$ be a non-empty family of sub-σ-algebras of $\mathcal{F}$, and let X be a random variable in $\mathcal{L}^1$. The family

$$\mathcal{X} = \{\mathbb{E}[X|\mathcal{F}] : \mathcal{F} \in \mathcal{C}\},$$

is uniformly integrable.

Hint: Argue that it follows directly from Proposition 12.6 that $\mathbb{E}[\varphi(|X|)] < \infty$ for some test function of uniform integrability. Then, show that the same φ can be used to prove that $\mathcal{X}$ is UI.

First properties of uniformly-integrable martingales

When it is known that the martingale $\{X_n\}_{n \in \mathbb{N}}$ is uniformly integrable, a lot can be said about its structure. We start with a definitive version of the dominated convergence theorem:

Proposition 12.9 (Improved dominated-convergence theorem)**.** *Suppose that $\{X_n\}_{n \in \mathbb{N}}$ is a sequence of random variables in $\mathcal{L}^p$, where $p \geq 1$, which converges to $X \in \mathcal{L}^0$ in probability. Then, the following statements are equivalent:*

1. *the sequence $\{|X|_n^p\}_{n \in \mathbb{N}}$ is uniformly integrable,*

2. $X_n \stackrel{\mathcal{L}^p}{\to} X$*, and*

3. $||X_n||_{\mathcal{L}^p} \to ||X||_{\mathcal{L}^p} < \infty$.

Proof. 1. $\to$ 2.: Since there exists a subsequence $\{X_{n_k}\}_{k \in \mathbb{N}}$ such that $X_{n_k} \stackrel{a.s.}{\to} X$, Fatou's lemma implies that

$$\mathbb{E}[|X|^p] = \mathbb{E}[\liminf_k |X_{n_k}|^p] \leq \liminf_k \mathbb{E}[|X_{n_k}|^p] \leq \sup_{X \in \mathcal{X}} \mathbb{E}[|X|^p] < \infty,$$

where the last inequality follows from the fact that uniformly-integrable families are bounded in $\mathcal{L}^1$.

Now that we know that $X \in \mathcal{L}^p$, uniform integrability of $\{|X_n|^p\}_{n \in \mathbb{N}}$ implies that the family $\{|X_n - X|^p\}_{n \in \mathbb{N}}$ is UI (use Problem 12.1, 2.). Since $X_n \stackrel{\mathbb{P}}{\to} X$ if and only if $X_n - X \stackrel{\mathbb{P}}{\to} 0$, we can assume without loss of generality that $X = 0$ a.s., and, consequently, we need to show that

$\mathbb{E}[|X_n|^p] \to 0$. We fix an $\varepsilon > 0$, and start by the following estimate

$$\begin{aligned}\mathbb{E}[|X_n|^p] &= \mathbb{E}[|X_n|^p \mathbf{1}_{\{|X_n|^p \leq \varepsilon/2\}}] + \mathbb{E}[|X_n|^p \mathbf{1}_{\{|X_n|^p > \varepsilon/2\}}] \\ &\leq \varepsilon/2 + \mathbb{E}[|X_n|^p \mathbf{1}_{\{|X_n|^p > \varepsilon/2\}}].\end{aligned} \tag{12.3}$$

By uniform integrability there exists $\rho > 0$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p \mathbf{1}_A] < \varepsilon/2 \text{ whenever } \mathbb{P}[A] \leq \rho.$$

Convergence in probability now implies that there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, we have $\mathbb{P}[|X_n|^p > \varepsilon/2] \leq \rho$. It follows directly from (12.3) that for $n \geq n_0$, we have $\mathbb{E}[|X_n|^p] \leq \varepsilon$.

2\. $\to$ 3.: $\Big| ||X_n||_{\mathcal{L}^p} - ||X||_{\mathcal{L}^p} \Big| \leq ||X_n - X||_{\mathcal{L}^p} \to 0$

3\. $\to$ 1.: For $M \geq 0$, define the function $\psi_M : [0, \infty) \to [0, \infty)$ by

$$\psi_M(x) = \begin{cases} x, & x \in [0, M-1] \\ 0, & x \in [M, \infty) \\ \text{interpolated linearly}, & x \in (M-1, M). \end{cases}$$

For a given $\varepsilon > 0$, dominated convergence theorem guarantees the existence of a constant $M > 0$ (which we fix throughout) such that

$$\mathbb{E}[|X|^p] - \mathbb{E}[\psi_M(|X|^p)] < \frac{\varepsilon}{2}.$$

Convergence in probability, together with continuity of ψ_M, implies that $\psi_M(X_n) \to \psi_M(X)$ in probability, for all M, and it follows from boundedness of ψ_M and the bounded convergence theorem that

$$\mathbb{E}[\psi_M(|X_n|^p)] \to \mathbb{E}[\psi_M(|X|^p)]. \tag{12.4}$$

By the assumption 3. and (12.4), there exists $n_0 \in \mathbb{N}$ such that

$$\mathbb{E}[|X_n|^p] - \mathbb{E}[|X|^p] < \varepsilon/4 \text{ and } \mathbb{E}[\psi_M(|X|^p)] - \mathbb{E}[\psi_M(|X_n|^p)] < \varepsilon/4,$$

for $n \geq n_0$. Therefore, for $n \geq n_0$.

$$\begin{aligned}\mathbb{E}[|X_n|^p \mathbf{1}_{\{|X_n|^p > M\}}] &\leq \mathbb{E}[|X_n|^p] - \mathbb{E}[\psi_M(|X_n|^p)] \\ &\leq \varepsilon/2 + \mathbb{E}[|X|^p] - \mathbb{E}[\psi_M(|X|^p)] \leq \varepsilon.\end{aligned}$$

Finally, to get uniform integrability of the entire sequence, we choose an even larger value of M to get $\mathbb{E}[|X_n|^p \mathbf{1}_{\{|X_n|^p > M\}}] \leq \varepsilon$ for the remaining $n < n_0$. □

Problem 12.4. For $Y \in \mathcal{L}^1_+$, show that the family $\{X \in \mathcal{L}^0 : |X| \leq Y, \text{ a.s.}\}$ is uniformly integrable. Deduce the dominated convergence theorem from Proposition 12.9

Since convergence in $\mathcal{L}^p$ implies convergence in probability, we have:

Corollary 12.10 ($\mathcal{L}^p$-convergent $\to$ UI, for $p \geq 1$)**.** *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be an $\mathcal{L}^p$-convergent sequence, for $p \geq 1$. Then family $\{X_n : n \in \mathbb{N}_0\}$ is UI.*

Since UI (sub)martingales are bounded in $\mathcal{L}^1$, they converge by Theorem 11.18. Proposition 12.9 guarantees that, additionally, convergence holds in $\mathcal{L}^1$:

Corollary 12.11 (UI (sub)martingales converge)**.** *Uniformly-integrable (sub)martingales converge a.s., and in $\mathcal{L}^1$.*

For martingales, uniform integrability implies much more:

Proposition 12.12 (UI martingales are Lévy martingales)**.** *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale. Then, the following are equivalent:*

1. *$\{X_n\}_{n\in\mathbb{N}_0}$ is a Lévy martingale, i.e., it admits a representation of the form $X_n = \mathbb{E}[X|\mathcal{F}_n]$, a.s., for some $X \in \mathcal{L}^1(\mathcal{F})$,*
2. *$\{X_n\}_{n\in\mathbb{N}_0}$ is uniformly integrable.*
3. *$\{X_n\}_{n\in\mathbb{N}_0}$ converges in $\mathcal{L}^1$,*

In that case, convergence also holds a.s., and the limit is given by $\mathbb{E}[X|\mathcal{F}_\infty]$, where $\mathcal{F}_\infty = \sigma(\cup_{n\in\mathbb{N}_0}\mathcal{F}_n)$.

Proof. 1. $\to$ 2. The representation $X_n = \mathbb{E}[X|\mathcal{F}_n]$, a.s., and Problem 12.3 imply that $\{X_n\}_{n\in\mathbb{N}_0}$ is uniformly integrable.

2. $\to$ 3. Corollary 12.11.

3. $\to$ 2. Corollary 12.10.

2. $\to$ 1. Corollary 12.11 implies that there exists a random variable $Y \in \mathcal{L}^1(\mathcal{F})$ such that $X_n \to Y$ a.s., and in $\mathcal{L}^1$. For $m \in \mathbb{N}$ and $A \in \mathcal{F}_m$, we have $|\mathbb{E}[X_n\mathbf{1}_A - Y\mathbf{1}_A]| \leq \mathbb{E}[|X_n - Y|] \to 0$, so $\mathbb{E}[X_n\mathbf{1}_A] \to \mathbb{E}[Y\mathbf{1}_A]$. Since $\mathbb{E}[X_n\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_n]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$, for $n \geq m$, we have

$$\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A], \text{ for all } A \in \cup_n\mathcal{F}_n.$$

The family $\cup_n\mathcal{F}_n$ is a π-system which generated the sigma algebra $\mathcal{F}_\infty = \sigma(\cup_n\mathcal{F}_n)$, and the family of all $A \in \mathcal{F}$ such that $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$ is a λ-system. Therefore, by the $\pi - \lambda$ Theorem, we have

$$\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A], \text{ for all } A \in \mathcal{F}_\infty.$$

Therefore, since $Y \in \mathcal{F}_\infty$, we conclude that $Y = \mathbb{E}[X|\mathcal{F}_\infty]$. $\square$

Example 12.13. There exists a non-negative (and therefore a.s.-convergent) martingale which is not uniformly integrable (and therefore, not $\mathcal{L}^1$-convergent). Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple random walk starting from 1, i.e. $X_0 = 1$ and $X_n = 1 + \sum_{k=1}^n \xi_k$, where $\{\xi_n\}_{n\in\mathbb{N}}$ is an iid sequence with $\mathbb{P}[\xi_n = 1] = \mathbb{P}[\xi_n = -1] = \frac{1}{2}$, $n \in \mathbb{N}$. Clearly, $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale, and so is $\{Y_n\}_{n\in\mathbb{N}_0}$, where $Y_n = X_n^T$ and $T = \inf\{n \in \mathbb{N} : X_n = 0\}$. By convention, $\inf \emptyset = +\infty$. It is well known that a simple symmetric random walk hits any level eventually, with probability 1 (we will prove this rigorously later), so $\mathbb{P}[T < \infty] = 1$, and, since $Y_n = 0$, for $n \geq T$, we have $Y_n \to 0$, a.s., as $n \to \infty$. On the other hand, $\{Y_n\}_{n\in\mathbb{N}_0}$ is a martingale, so $\mathbb{E}[Y_n] = \mathbb{E}[Y_0] = 1$, for $n \in \mathbb{N}$. Therefore, $\mathbb{E}[Y_n] \not\to \mathbb{E}[X]$, which can happen only if $\{Y_n\}_{n\in\mathbb{N}_0}$ is *not* uniformly integrable.

Backward martingales

If, instead of $\mathbb{N}_0$, we use $-\mathbb{N}_0 = \{\dots, -2, -1, 0\}$ as the time set, the notion of a filtration is readily extended: it is still a family of sub-σ-algebras of $\mathcal{F}$, parametrized by $-\mathbb{N}_0$, such that $\mathcal{F}_{n-1} \subseteq \mathcal{F}_n$, for $n \in -\mathbb{N}_0$.

Definition 12.14 (Backward (sub)martingale)**.** We say that a stochastic process $\{X_n\}_{n\in-\mathbb{N}_0}$, is a **backward submartingale** with respect to the filtration $\{\mathcal{F}_n\}_{n\in-\mathbb{N}_0}$, if

1. $\{X_n\}_{n\in-\mathbb{N}_0}$ is $\{\mathcal{F}_n\}_{n\in-\mathbb{N}_0}$-adapted,
2. $X_n \in \mathcal{L}^1$, for all $n \in \mathbb{N}_0$, and
3. $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \geq X_{n-1}$, for all $n \in -\mathbb{N}_0$.

If, in addition to 1. and 2., the inequality in 3. is, in fact, an equality, we say that $\{X_n\}_{n\in-\mathbb{N}_0}$ is a **backward martingale**.

One of the most important facts about backward submartingales is that they (almost) always converge a.s., and in $\mathcal{L}^1$.

Proposition 12.15 (Backward submartingale convergence)**.** *Suppose that $\{X_n\}_{n\in-\mathbb{N}_0}$ is a backward submartingale such that*

$$\lim_{n\to-\infty} \mathbb{E}[X_n] > -\infty.$$

Then $\{X_n\}_{n\in-\mathbb{N}_0}$ is uniformly integrable and there exists a random variable $X_{-\infty} \in \mathcal{L}^1(\cap_n \mathcal{F}_n)$ such that

$$X_n \to X_{-\infty} \text{ a.s. and in } \mathcal{L}^1, \tag{12.5}$$

and

$$X_{-\infty} \leq \mathbb{E}[X_m | \cap_n \mathcal{F}_n], \text{ a.s., for all } m \in -\mathbb{N}_0. \quad (12.6)$$

Proof. We start by decomposing $\{X_n\}_{n \in -\mathbb{N}_0}$ in the manner of Doob and Meyer. For $n \in -\mathbb{N}_0$, set $\Delta A_n = \mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] \geq 0$, a.s., and $A_{-n} = \sum_{k=0}^{n} \Delta A_{-k}$, for $n \in \mathbb{N}_0$. The backward submartingale property of $\{X_n\}_{n \in \mathbb{N}_0}$ implies that $\mathbb{E}[X_n] \geq L = \lim_{n \to -\infty} \mathbb{E}[X_n] > -\infty$, so

$$\mathbb{E}[A_n] = \mathbb{E}[X_0 - X_n] \leq \mathbb{E}[X_0] - L, \text{ for all } n \in \mathbb{N}_0.$$

The monotone convergence theorem implies that $\mathbb{E}[A_{-\infty}] < \infty$, where $A_{-\infty} = \sum_{n=0}^{\infty} A_{-n}$. The process $\{M_n\}_{n \in -\mathbb{N}_0}$ defined by $M_n = X_n - A_n$ is a backward martingale. Indeed,

$$\mathbb{E}[M_n - M_{n-1} | \mathcal{F}_{n-1}] = \mathbb{E}[X_n - X_{n-1} - \Delta A_n | \mathcal{F}_{n-1}] = 0.$$

Since all backward martingales are uniformly integrable (why?) and the sequence $\{A_n\}_{n \in -\mathbb{N}_0}$ is uniformly dominated by $A_{-\infty} \in \mathcal{L}^1$ - and therefore uniformly integrable - we conclude that $\{X_n\}_{n \in -\mathbb{N}_0}$ is also uniformly integrable.

To prove convergence, we start by observing that the uniform integrability of $\{X_n\}_{n \in -\mathbb{N}_0}$ implies that $\sup_{n \in -\mathbb{N}_0} \mathbb{E}[X_n^+] < \infty$. A slight modification of the proof of the martingale convergence theorem (left to a very diligent reader) implies that $X_n \to X_{-\infty}$, a.s. for some random variable $X_\infty \in \cap_n \mathcal{F}_n$. Uniform integrability also ensures that the convergence holds in $\mathcal{L}^1$ and that $X_{-\infty} \in \mathcal{L}^1$.

In order to show (12.6), it is enough to show that

$$\mathbb{E}[X_{-\infty} \mathbf{1}_A] \leq \mathbb{E}[X_m \mathbf{1}_A], \quad (12.7)$$

for any $A \in \cap_n \mathcal{F}_n$, and any $m \in -\mathbb{N}_0$. We first note that since $X_n \leq \mathbb{E}[X_m | \mathcal{F}_n]$, for $n \leq m \leq 0$, we have

$$\mathbb{E}[X_n \mathbf{1}_A] \leq \mathbb{E}[\mathbb{E}[X_m | \mathcal{F}_n] \mathbf{1}_A] = \mathbb{E}[X_m \mathbf{1}_A],$$

for any $A \in \cap_n \mathcal{F}_n$. It remains to use the fact the $\mathcal{L}^1$-convergence of $\{X_n\}_{n \in -\mathbb{N}_0}$ implies that $\mathbb{E}[X_n \mathbf{1}_A] \to \mathbb{E}[X_{-\infty} \mathbf{1}_A]$, for all $A \in \mathcal{F}$. □

Remark 12.16. Even if $\lim \mathbb{E}[X_n] = -\infty$, the convergence $X_n \to X_{-\infty}$ still holds, but not in $\mathcal{L}^1$ and $X_{-\infty}$ may take the value $-\infty$ with positive probability.

Corollary 12.17 (Backward martingale convergence)**.** *If $\{X_n\}_{n \in -\mathbb{N}_0}$ is a backward martingale, then $X_n \to X_{-\infty} = \mathbb{E}[X_0 | \cap_n \mathcal{F}_n]$, a.s., and in $\mathcal{L}^1$.*

Applications of backward martingales

We can use the results about the convergence of backward martingales to give a non-classical proof of the strong law of large numbers. Before that, we need a useful classical result.

Proposition 12.18 (Kolmogorov's 0-1 law). *Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables, and let the **tail σ-algebra** $\mathcal{F}_{-\infty}$ be defined by*

$$\mathcal{F}_{-\infty} = \bigcap_{n\in\mathbb{N}} \mathcal{F}_{-n}, \text{ where } \mathcal{F}_{-n} = \sigma(\xi_n, \xi_{n+1}, \dots).$$

Then $\mathcal{F}_{-\infty}$ is $\mathbb{P}$-trivial, i.e., $\mathbb{P}[A] \in \{0,1\}$, for all $A \in \mathcal{F}_{-\infty}$.

Proof. Define $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, and note that $\mathcal{F}_{n-1}$ and $\mathcal{F}_{-n}$ are independent σ-algebras. Therefore, $\mathcal{F}_{-\infty} \subseteq \mathcal{F}_{-n}$ is also independent of $\mathcal{F}_n$, for each $n \in \mathbb{N}$. This, in turn, implies that $\mathcal{F}_{-\infty}$ is independent of the σ-algebra $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. On the other hand, $\mathcal{F}_{-\infty} \subseteq \mathcal{F}_\infty$, so $\mathcal{F}_{-\infty}$ is independent of itself. This implies that $\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]\mathbb{P}[A]$, for each $A \in \mathcal{F}_{-\infty}$, i.e., that $\mathbb{P}[A] \in \{0,1\}$. □

Theorem 12.19 (Strong law of large numbers). *Let $\{\xi_n\}_{n\in\mathbb{N}}$ be an iid sequence of random variables in $\mathcal{L}^1$. Then*

$$\tfrac{1}{n}(\xi_1 + \dots + \xi_n) \to \mathbb{E}[\xi_1], \text{ a.s. and in } \mathcal{L}^1.$$

Proof. For notational reasons, backward martingales are indexed by $-\mathbb{N}$ instead of $-\mathbb{N}_0$. For $n \in -\mathbb{N}$, let $S_n = \xi_1 + \dots + \xi_n$, and let $\mathcal{F}_n$ be the σ-algebra generated by $S_n, S_{n+1}, \dots$. The process $\{X_n\}_{n\in-\mathbb{N}}$ is given by

$$X_{-n} = \mathbb{E}[\xi_1|\mathcal{F}_n], \text{ for } n \in \mathbb{N}_0.$$

Since $\sigma(S_n, S_{n+1}, \dots) = \sigma(\sigma(S_n), \sigma(\xi_{n+1}, \xi_{n+2}, \dots))$, and the σ-algebra $\sigma(\xi_{n+1}, \xi_{n+2}, \dots)$ is independent of ξ_1, for $n \in \mathbb{N}$, we have

$$X_{-n} = \mathbb{E}[\xi_1|\mathcal{F}_n] = \mathbb{E}[\xi_1|\sigma(S_n)] = \tfrac{1}{n} S_n,$$

where the last equality follows from Problem 10.9 in Lecture 10. Backward martingales converge a.s., and in $\mathcal{L}^1$, so for the random variable $X_{-\infty} = \lim_n \frac{1}{n} S_n$ we have

$$\mathbb{E}[X_{-\infty}] = \lim_n \mathbb{E}[\tfrac{1}{n} S_n] = \mathbb{E}[\xi_1].$$

On the other hand, since $\lim_n \frac{1}{n} S_k = 0$, for all $k \in \mathbb{N}$, we have $X_{-\infty} = \lim_n \frac{1}{n}(\xi_{k+1} + \dots + \xi_n)$, for any $k \in \mathbb{N}$, and so $X_{-\infty} \in \sigma(\xi_{k+1}k, \xi_{k+2}, \dots)$. By Proposition 12.18, $X_{-\infty}$ is measurable in a $\mathbb{P}$-trivial σ-algebra, and is, thus, constant a.s. (why?). Since $\mathbb{E}[X_{-\infty}] = \mathbb{E}[\xi_1]$, we must have $X_{-\infty} = \mathbb{E}[\xi_1]$, a.s. □

Additional Problems

Problem 12.5 (A UI martingale not in H^1)**.** Set $\Omega = \mathbb{N}$, $\mathcal{F} = 2^{\mathbb{N}}$, and $\mathbb{P}$ is the probability measure on $\mathcal{F}$ characterized by $\mathbb{P}[\{k\}] = 2^{-k}$, for each $k \in \mathbb{N}$. Define the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ by

$$\mathcal{F}_n = \sigma\Big(\{1\}, \{2\}, \dots, \{n-1\}, \{n, n+1, \dots\}\Big), \text{ for } n \in \mathbb{N}.$$

Let $Y : \Omega \to [1, \infty)$ be a random variable such that $\mathbb{E}[Y] < \infty$ and $\mathbb{E}[YK] = \infty$, where $K(k) = k$, for $k \in \mathbb{N}$.

1. Find an explicit example of a random variable Y with the above properties.
2. Find an expression for $X_n = \mathbb{E}[Y|\mathcal{F}_n]$ in terms of the values $Y(k)$, $k \in \mathbb{N}$.
3. Using the fact that $X^*_\infty(k) := \sup_{n\in\mathbb{N}} |X_n(k)| \geq X_k(k)$ for $k \in \mathbb{N}$, show that $\{X_n\}_{n\in\mathbb{N}}$ is a uniformly integrable martingale which is not in H^1.

Note: A martingale $\{X_n\}_{n\in\mathbb{N}}$ is said to be in H^1 if $X^*_\infty \in \mathbb{L}^1$.

Problem 12.6 (Scheffé's lemma)**.** Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a sequence of random variables in $\mathcal{L}^1_+$ such that $X_n \to X$, a.s., for some $X \in \mathcal{L}^1_+$. Show that $\mathbb{E}[X_n] \to \mathbb{E}[X]$ if and only if the sequence $\{X_n\}_{n\in\mathbb{N}_0}$ is UI.

Problem 12.7 (Hunt's lemma)**.** Let $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a filtration, and let $\{X_n\}_{n\in\mathbb{N}_0}$ be a sequence in $\mathcal{L}^0$ such that $X_n \to X$, for some $X \in \mathcal{L}^0$, both in $\mathcal{L}^1$ and a.s.

1. *(Hunt's lemma)*. Assume that $|X_n| \leq Y$, a.s., for all $n \in \mathbb{N}$ and some $Y \in \mathcal{L}^1_+$ Prove that

$$\mathbb{E}[X_n|\mathcal{F}_n] \to \mathbb{E}[X|\sigma(\cup_n \mathcal{F}_n)], \text{ a.s.} \tag{12.8}$$

Hint: Define $Z_n = \sup_{m\geq n} |X_m - X|$, and show that $Z_n \to 0$, a.s., and in $\mathcal{L}^1$.

2. Find an example of a sequence $\{X_n\}_{n\in\mathbb{N}}$ in $\mathcal{L}^1$ such that $X_n \to 0$, a.s., and in $\mathcal{L}^1$, but $\mathbb{E}[X_n|\mathcal{G}]$ does not converge to 0, a.s., for some $\mathcal{G} \subseteq \mathcal{F}$. *Note:* The existence of such a sequence proves that (12.8) is not true without an additional assumption, such as the one of uniform domination in 1. It provides an example of a property which does not generalize from the unconditional to the conditional case.

Hint: Look for X_n of the form $X_n = \xi_n \frac{\mathbf{1}_{A_n}}{\mathbb{P}[A_n]}$ and $\mathcal{G} = \sigma(\xi_n; n \in \mathbb{N})$.

Problem 12.8 (Krickeberg's decomposition)**.** Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale. Show that the following two statements are equivalent:

Hint: Consider $\lim_n \mathbb{E}[X^+_{m+n}|\mathcal{F}_m]$, for $m \in \mathbb{N}_0$.

1. There exists martingales $\{X^+_n\}_{n\in\mathbb{N}_0}$ and $\{X^-_n\}_{n\in\mathbb{N}_0}$ such that $X^+_n \geq 0$, $X^-_n \geq 0$, a.s., for all $n \in \mathbb{N}_0$ and $X_n = X^+_n - X^-_n$, $n \in \mathbb{N}_0$.
2. $\sup_{n\in\mathbb{N}_0} \mathbb{E}[|X_n|] < \infty$.

Problem 12.9 (Branching processes)**.** Let ν be a probability measure on $\mathcal{B}(\mathbb{R})$ with $\nu(\mathbb{N}_0) = 1$, which we call the **offspring distribution**. A population starting from one individual ($Z_0 = 1$) evolves as follows. The initial member leaves a random number Z_1 of children and dies. After that, each of the Z_1 children of the initial member, produces a random number of children and dies. The total number of all children of the Z_1 members of the generation 1 is denoted by Z_2. Each of the Z_2 members of the generation 2 produces a random number of children, etc. Whenever an individual procreates, the number of children has the distribution ν, and is independent of the sizes of all the previous generations including the present one, as well as of the numbers of children of other members of the present generation.

1. Suppose that a probability space and iid sequence $\{\eta_n\}_{n\in\mathbb{N}}$ of random variables with the distribution μ is given. Show how you would construct a sequence $\{Z_n\}_{n\in\mathbb{N}_0}$ with the above properties.

   Hint: Z_{n+1} is a sum of iid random variables with the number of summands equal to Z_n.

2. For a distribution ρ on $\mathbb{N}_0$, we define the the **generating function** $P_\rho : [0,1] \to [0,1]$ of ρ by
$$P_\rho(x) = \sum_{k\in\mathbb{N}_0} \rho(\{k\}) x^k.$$
   Show that each P_ρ is continuous, non-decreasing and convex on $[0,1]$ and continuously differentiable on $(0,1)$.

3. Let $P = P_\nu$ be the generating function of the offspring distribution ν, and for $n \in \mathbb{N}_0$, we define $P_n(x)$ as the generating function of the distribution of Z_n, i.e., $P_n(x) = \sum_{k\in\mathbb{N}_0} \mathbb{P}[Z_n = k]x^k$. Show that $P_n(x) = P(P(\dots P(x)\dots))$ (there are n Ps).

   Hint: Note that $P(x) = \mathbb{E}[x^{Z_n}]$ for $x > 0$ and use the result of Problem 10.6

4. Define the **extinction probability** p_e by $p_e = \mathbb{P}[Z_n = 0,$ for some $n \in \mathbb{N}]$. Prove that p_e is a fixed point of the map P, i.e., that $P(p_e) = p_e$.

   Hint: Show that $p_e = \lim_n P^{(n)}(0)$, where $P^{(n)}$ is the n-fold composition of P with itself.

5. Let $\mu = \mathbb{E}[Z_1]$, be the expected number of offspring. Show that when $\mu \leq 1$ and $\nu(\{1\}) < 1$, we have $p_e = 1$, i.e., the population dies out with certainty if the expected number of offspring does not exceed 1.

   Hint: Draw a picture of the functions x and $P(x)$ and use (and prove) the fact that, as a consequence of the assumption $\mu \leq 1$, we have $P'(x) < 1$ for all $x < 1$.

6. Assuming that $0 < \mu < \infty$, show that the process $\{X_n\}_{n\in\mathbb{N}_0}$, given by $X_n = Z_n/\mu^n$, is a martingale (with respect to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$, where $\mathcal{F}_n = \sigma(Z_0, Z_1, \dots, Z_n)$).

7. Identify all probability measures ν with $\nu(\mathbb{N}_0) = 1$, and $\sum_{k\in\mathbb{N}_0} k\nu(\{k\}) = 1$ such that the branching process $\{Z_n\}_{n\in\mathbb{N}_0}$ with the offspring distribution ν is uniformly integrable.

Problem 12.10 (A bit of everything)**.** Given two independent simple symmetric random walks $\{\tilde{X}_n\}_{n\in\mathbb{N}_0}$ and $\{\tilde{Y}_n\}_{n\in\mathbb{N}_0}$, let $\{X_n\}_{n\in\mathbb{N}_0}$ denote $\{\tilde{X}_n\}_{n\in\mathbb{N}_0}$ stopped when it first hits the level 1, and let $\{Y_n\}_{n\in\mathbb{N}_0}$

be defined by

$$Y_0 = 0,\ Y_n = \sum_{k=1}^{n} 2^{-k}(\tilde{Y}_k - \tilde{Y}_{k-1}).$$

Identify the distribution of $\liminf_n (X_n + Y_n)$ and show that the sequence $\{X_n + Y_n\}_{n \in \mathbb{N}_0}$ is *not* uniformly integrable.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 13
Further Martingales

A bounded optional-sampling theorem

For a stopping time T and a stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$, we sometimes talk about the value X_T of $\{X_n\}_{n\in\mathbb{N}_0}$ **sampled at** T. When $T(\omega) \neq \infty$, for all $\omega \in \Omega$, then we can define

$$X_T(\omega) = X_{T(\omega)}(\omega).$$

When T takes the value $+\infty$ with positive probability, there is no canonical way of defining X_T. In that case, we often use the random variable $X_T \mathbf{1}_{\{T<\infty\}}$, which takes the value 0 on the set $T = \infty$. Alternatively, if $\lim_n X_n$ exists a.s. on $\{T = \infty\}$, we define

$$X_T = \lim_{n\to\infty} X_n \text{ on } \{T = \infty\}.$$

Our first result is an extension of Proposition 11.17 in Lecture 11, and shows that the (sub)martingale property extends from fixed to stopping times (under certain regularity assumptions).

Proposition 13.1. (Bounded Optional Sampling) *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a (sub)martingale, and let T be a stopping time. Then the stopped process $\{X_n^T\}_{n\in\mathbb{N}}$ is also a (sub)martingale. Moreover, we have*

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge m}] \leq \mathbb{E}[X_m],$$

for all $m \in \mathbb{N}_0$, and the inequalities become equalities when X is a martingale.

Proof. Assume, first, that $\{X_n\}_{n\in\mathbb{N}_0}$ is a submartingale. We note that for the processes $H_n = \mathbf{1}_{\{T<n\}}$, $K_n = 1 - H_n$, $n \in \mathbb{N}$, are both predictable, non-negative and bounded, so their martingale transforms $(H \cdot X)$ and $(K \cdot X)$ are submartingales. Moreover,

$$(H \cdot X)_n = X_n - X_{T\wedge n}, \quad (K \cdot X)_n = X_{T\wedge n} - X_0.$$

The submartingale property of $(K \cdot X)$ is equivalent to the submartingale property of X^T, and implies that $\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge m}]$, for $m \in \mathbb{N}_0$. To prove the second inequality, we use the submartingale property of $(H \cdot X)$ to conclude that $\mathbb{E}[X_m - X_{T\wedge m}] \geq 0$. □

Proposition 13.1 can be used to give a useful characterization of martingales which does not involve the conditional expectation:

Problem 13.1. Let $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ be a filtration, and let $\mathcal{S}$ denote the set of all bounded $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$-stopping times $T = T(\omega)$ (i.e., $T \in \mathcal{L}^1$ for $T \in \mathcal{S}$). Prove that (1) and (2) below are equivalent for an adapted process $\{X_n\}_{n\in\mathbb{N}_0}$:

1. $\{X_n\}_{n\in\mathbb{N}_0}$ is an $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$-martingale.
2. $X_T \in \mathcal{L}^1$ and $\mathbb{E}[X_T] = \mathbb{E}[X_0]$, for all $T \in \mathcal{S}$.

How about the equivalence of $1.'$ and $2.'$ below?

1.′ $\{X_n\}_{n\in\mathbb{N}_0}$ is an $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$-supermartingale.

2.′ $X_T \in \mathcal{L}^1$ and $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$, for all $T \in \mathcal{S}$.

The following simple consequence of Proposition 13.1 has far-reaching applications:

Corollary 13.2. (Doob's inequality) *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a submartingale. For $\lambda > 0$ and $n \in \mathbb{N}$ we have*

$$\lambda\mathbb{P}[\sup_{m\leq n} X_m \geq \lambda] \leq \mathbb{E}[X_n \mathbf{1}_{\{\sup_{m\leq n} X_m \geq \lambda\}}] \leq \mathbb{E}[X_n^+].$$

Proof. For $n \in \mathbb{N}_0$ and $\lambda > 0$, we define the stopping time T by

$$T = \inf\{m \in \mathbb{N}_0 : X_m \geq \lambda\} \wedge n$$

and the event $A = \{X_m \geq \lambda, \text{ for some } 1 \leq m \leq n\}$. Proposition 13.1 implies that $\mathbb{E}[X_n] \geq \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T]$, so

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_T \mathbf{1}_A] + \mathbb{E}[X_T \mathbf{1}_{A^c}] = \mathbb{E}[X_T \mathbf{1}_A] + \mathbb{E}[X_n \mathbf{1}_{A^c}] \geq \mathbb{E}[\lambda \mathbf{1}_A] + \mathbb{E}[X_n \mathbf{1}_{A^c}].$$

Therefore,

$$\lambda\mathbb{P}[A] \leq \mathbb{E}[X_n \mathbf{1}_A] \leq \mathbb{E}[X_n^+]. \qquad \square$$

For a stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$, the **maximal process** $\{X_n^*\}_{n\in\mathbb{N}_0}$ is defined by

$$X_n^* = \sup_{m\leq n} |X_m|.$$

We also write X_∞^* for the $\mathcal{L}_+^0([0,\infty])$-valued random variable $\sup_{n\in\mathbb{N}_0} |X_n|$.

Corollary 13.3. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale or a non-negative submartingale. Then*

$$\mathbb{P}[X_n^* \geq \lambda] \leq \tfrac{1}{\lambda}\mathbb{E}[|X_n| \mathbf{1}_{\{X_n^*\geq\lambda\}}] \leq \tfrac{1}{\lambda}\mathbb{E}[|X_n|],$$

for all $n \in \mathbb{N}_0$, $\lambda > 0$.

Problem 13.2. (Kolmogorov's inequality) Let $\{\xi_n\}_{n\in\mathbb{N}}$ be an iid sequence such that $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[\xi_1^2] = \sigma^2 < \infty$. If $\{X_n\}_{n\in\mathbb{N}_0}$ is a random walk with steps $\{\xi_n\}_{n\in\mathbb{N}}$, i.e., $X_0 = 0$, $X_n = \sum_{k=1}^n \xi_k$, $n \in \mathbb{N}$, show that, for $\lambda > 0$, we have

$$\mathbb{P}[\sup_{m\leq n} |X_m| \geq \lambda\sqrt{n}] \leq \tfrac{\sigma^2}{\lambda^2}.$$

Definition 13.4. A random variable Y is said to be in **weak** $\mathcal{L}^1$ - denoted by $w\mathcal{L}^1$ - if there exists a constant $C \geq 0$, such that

$$\lambda\mathbb{P}[|Y| > \lambda] \leq C, \text{ for all } \lambda > 0.$$

The smallest value of the constant C with the above property is denoted by $||Y||_{w\mathcal{L}^1}$.

Remark 13.5. Markov's inequality implies that the weak $\mathcal{L}^1$ is a superset of $\mathcal{L}^1$, and as the function $f(x) = \frac{1}{x}$ on $((0,1], \mathcal{B}((0,1]), \lambda)$ can be used to show, the two spaces are not equal.

Problem 13.3. Show that $||\cdot||_{w\mathcal{L}^1}$ does not satisfy the triangle inequality, so that it is *not* a (pseudo)-norm.

Hint: On the probability space $((0,1), \mathcal{B}(0,1), \lambda)$, define $X_1(\omega) = 1/\omega$, and $X_2(\omega) = 1/(1-\omega)$.

Corollary 13.3 states that the value X_n^* of the maximal process $\{X_n^*\}_{n\in\mathbb{N}_0}$ is in the weak $\mathcal{L}^1$ if X is a martingale or a non-negative submartingale, with the "$w\mathcal{L}^1$-norm" bounded from above by the $\mathcal{L}^1$ norm of X_n. More generally, we have

Proposition 13.6. ($w\mathcal{L}^1$-maximal inequality) *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale or a non-negative submartingale which is bounded in $\mathcal{L}^1$. Then*

$$||X_\infty^*||_{w\mathcal{L}^1} \leq \sup_{n\in\mathbb{N}_0} ||X_n||_{\mathcal{L}^1}.$$

Proof. For $\lambda > 0$, Corollary 13.3 implies that

$$\lambda\mathbb{P}[X_\infty^* > \lambda] = \lambda \sup_{n\in\mathbb{N}_0} \mathbb{P}[X_n^* > \lambda] \leq \sup_{n\in\mathbb{N}_0} \mathbb{E}[|X_n|]. \qquad \square$$

Example 13.7. It is interesting to note that the maximum of a non-negative (and therefore, $\mathcal{L}^1$-bounded) martingale does not have to be in $\mathcal{L}^1$. Here is an example: let $\{\xi_n\}_{n\in\mathbb{N}}$ be an iid sequence of coin-tosses, i.e., random variables with $\mathbb{P}[\xi_1 = 1] = \mathbb{P}[\xi = -1] = \frac{1}{2}$, and let $X_n = 1 + \sum_{k=1}^n \xi_k$, $n \in \mathbb{N}_0$ be a symmetric random walk starting from $X_0 = 1$. It will be shown a bit later that for the stopping times

$$\tau_M = \inf\{n \in \mathbb{N}_0 \,:\, X_n = M\}, \ M \in \mathbb{N}_0,$$

we have

$$\mathbb{P}[\tau_M \leq \tau_0] = \tfrac{1}{M}, \text{ for } M \geq 2.$$

Consider the non-negative martingale $Y_n = X_n^{\tau_0} = X_{\tau_0 \wedge n}$. The maximum Y_∞^* will be at least M if and only if X hits the level M before it hits the level 0. Therefore

$$\mathbb{P}[Y_\infty^* \geq M] = \tfrac{1}{M}, \text{ for } M \geq 1,$$

and so $||Y_\infty^*||_{\mathcal{L}^1} = \mathbb{E}[Y_\infty^*] = \sum_{M \geq 1} \mathbb{P}[X \geq M] = \sum_{M \geq 1} \frac{1}{M} = \infty$.

Corollary 13.3 uses Proposition 13.3 to relate the weak $\mathcal{L}^1$ norm of the maximal process to the $\mathcal{L}^1$-bound of the underlying martingale (or a non-negative submartingale). If one is willing to replace $\mathcal{L}^1$-estimates by $\mathcal{L}^p$-estimates for $p > 1$, one gets a more symmetric theory. We start with an easy but quite useless version:

Lemma 13.8. (An almost useless maximal inequality) *Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a martingale or a non-negative submartingale, and $p \in (1, \infty)$. Then, for $n \in \mathbb{N}_0$*

$$||X_n^*||_{\mathcal{L}^p} \leq (n+1)^{1/p} ||X_n||_{\mathcal{L}^p}.$$

Proof. Without loss of generality we assume that $\mathbb{E}[|X_n|^p] < \infty$. By Jensen's inequality $\{|X_n|^p\}_{n \in \mathbb{N}}$ is a submartingale and so $\mathbb{E}[|X_n|^p] \geq \mathbb{E}[|X_k|^p]$, for $k \leq n$. Therefore

$$\mathbb{E}[(X_n^*)^p] = \mathbb{E}[\sup_{k \leq n} |X_k|^p] \leq \sum_{k=0}^{n} \mathbb{E}[|X_k|^p] \leq (n+1)\mathbb{E}[|X_n|^p],$$

and the statement follows. □

The uselessness of the inequality in Lemma 13.8 stems from the fact that the constant $(n+1)^{1/p}$ depends on n and cannot be used to say anything about $||X_\infty^*||_{\mathcal{L}^p}$. The following (much deeper) result takes care of that issue.

Proposition 13.9. (Maximal inequalities) *Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a martingale or a non-negative submartingale, and $p \in (1, \infty)$. Then, for $n \in \mathbb{N}_0$*

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

Before we give a proof, here is a simple (and useful) result which, in the special case $Y = 1$, relates the tail of a distribution to its moments.

Lemma 13.10. *For non-negative random variables X and Y, and $p > 0$, we have*

$$\mathbb{E}[YX^p] = \int_0^\infty p\lambda^{p-1} \mathbb{E}[Y \mathbf{1}_{\{X \geq \lambda\}}] \, d\lambda.$$

Proof. Since $x^p = \int_0^x p\lambda^{p-1}\,d\lambda$, Fubini's theorem implies that

$$\begin{aligned}\mathbb{E}[YX^p] &= \mathbb{E}[Y\int_0^X p\lambda^{p-1}\,d\lambda] = \mathbb{E}[\int_0^\infty Y\mathbf{1}_{\{\lambda\leq X\}}p\lambda^{p-1}\,d\lambda] \\ &= \int_0^\infty p\lambda^{p-1}\mathbb{E}[Y\mathbf{1}_{\{X\geq\lambda\}}]\,d\lambda.\end{aligned}$$

□

Proof of Proposition 13.9. The process $\{|X_n|\}_{n\in\mathbb{N}_0}$ is a non-negative submartingale, so we can assume, without loss of generality, that $\{X_n\}_{n\in\mathbb{N}_0}$ is a non-negative submartingale and that $\mathbb{E}[(X_n)^p] < \infty$, for all $n \in \mathbb{N}_0$. Corollary 13.3 can be viewed as an (implicitly defined) estimate of the tail of X_n^* in terms X_n:

$$\mathbb{P}[X_n^* \geq \lambda] \leq \tfrac{1}{\lambda}\mathbb{E}[|X_n|\,\mathbf{1}_{\{X_n^*\geq\lambda\}}].$$

Lemma 13.10 turns this tail estimate into a moment estimate in the same spirit:

$$\begin{aligned}\mathbb{E}[(X_n^*)^p] &= \int_0^\infty p\lambda^{p-1}\mathbb{P}[X_n^*\geq\lambda]\,d\lambda \\ &\leq \tfrac{p}{p-1}\int_0^\infty (p-1)\lambda^{p-2}\mathbb{E}[X_n\mathbf{1}_{\{X_n^*\geq\lambda\}}]\,d\lambda = \tfrac{p}{p-1}\mathbb{E}[X_n(X_n^*)^{p-1}].\end{aligned}$$

Let $q = \frac{p}{p-1}$ be the conjugate exponent of p. Hölder's inequality applied to the right-hand side above yields

$$||X_n^*||_{\mathcal{L}^p}^p \leq q||X_n||_{\mathcal{L}^p}||(X_n^*)^{p-1}||_{\mathcal{L}^q} \leq q||X_n||_{\mathcal{L}^p}||X_n^*||_{\mathcal{L}^p}^{p/q}.$$

In order divide by $||X_n^*||_{\mathcal{L}^p}^{p/q}$ and finish the proof, we need to show that $\mathbb{E}[(X_n^*)^p] < \infty$. This estimate, however, follows directly from the (not so) useless maximal inequality of Lemma 13.8. □

Definition 13.11. A stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ is said to be **bounded in $\mathcal{L}^p$**, for $p \in [1,\infty]$ if the family $\{X_n\}_{n\in\mathbb{N}_0}$ of random variables is bounded in $\mathcal{L}^p$, i.e., if $\sup_{n\in\mathbb{N}_0}||X_n||_{\mathcal{L}^p} < \infty$.

Corollary 13.12. ($\mathcal{L}^p$-bounded martingales) *For $p \in (1,\infty)$, let $\{X_n\}_{n\in\mathbb{N}_0}$ be an $\mathcal{L}^p$-bounded process which is a martingale or a non-negative submartingale. Then,*

1. *the family $\{|X|_n^p\}_{n\in\mathbb{N}_0}$ is UI, and*
2. *there exists $X_\infty \in \mathcal{L}^p$ such that $X_n \to X_\infty$ a.s., and in $\mathcal{L}^p$.*

Proof. Since $\{X_n\}_{n\in\mathbb{N}_0}$ is bounded in $\mathcal{L}^p$, it is bounded in $\mathcal{L}^1$, so the martingale convergence theorem implies that there exists $X_\infty \in \mathcal{L}^1$ such that $X_n \to X_\infty$, a.s. Fatou's lemma can be used to conclude that $X_\infty \in \mathcal{L}^p$. Indeed, we have

$$\mathbb{E}[|X_\infty|^p] = \mathbb{E}[\lim_n |X_n|^p] \leq \liminf_n \mathbb{E}[|X_n|^p] \leq \sup_{n\in\mathbb{N}_0}||X_n||_{\mathcal{L}^p} < \infty.$$

In order to establish the uniform integrability of the sequence $\{|X_n|^p\}_{n\in\mathbb{N}_0}$, the full strength of the Maximal inequality needs to be used: Proposition 13.9 implies that for any $n \in \mathbb{N}$, we have

$$||X_n^*||_{\mathcal{L}^p} \leq \tfrac{p}{p-1}||X_n||_{\mathcal{L}^p} \leq M = \tfrac{p}{p-1}\sup_{n\in\mathbb{N}}||X_n||_{\mathcal{L}^p} < \infty.$$

Since $X_n^* \nearrow X_\infty^* = \sup_{n\in\mathbb{N}_0}|X_n|$, monotone convergence theorem implies that $||X_\infty^*||_{\mathcal{L}^p} \leq M < \infty$. Consequently,

$$\sup_{n\in\mathbb{N}}|X_n|^p \in \mathcal{L}^1,$$

and so the sequence $\{|X_n|^p\}_{n\in\mathbb{N}_0}$ is uniformly bounded by a random variable in $\mathcal{L}^1$, and, therefore, uniformly integrable. Moreover, it now follows directly from the dominated convergence theorem that $||X_n - X||_{\mathcal{L}^p} \to 0$. □

Remark 13.13. Note how, again, the case $p = 1$ differs greatly from the case $p > 1$. If a martingale (or a non-negative submartingale) is bounded in $\mathcal{L}^1$, it (i.e., the family $\{|X|_n^p\}_{n\in\mathbb{N}_0}$) does not have to be uniformly integrable and the convergence $X_n \to X_\infty$ does not have to hold in $\mathbb{L}^1$. When $p > 1$, mere boundedness in $\mathcal{L}^p$ is enough for uniform integrability of $\{|X_n|^p\}_{n\in\mathbb{N}_0}$ and the convergence $X_n \to X_\infty$ holds in $\mathcal{L}^p$. What's more, the family $\{|X_n|^p\}_{n\in\mathbb{N}_0}$ is uniformly integrable in a particularly strong way - it is uniformly dominated by the random variable $(X_\infty^*)^p \in \mathcal{L}^1$.

A general optional-sampling theorem

The (bounded) optional sampling theorem (Proposition 13.1) can be extended to a large class of *unbounded* stopping times. Some care is needed because not all stopping times will give reasonable results:

Example 13.14. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple symmetric random walk, i.e., $X_0 = 0$, $X_n = \sum_{k=1}^n \xi_k$, $n \in \mathbb{N}$, where $\{\xi_n\}_{n\in\mathbb{N}}$ is an iid sequence of coin tosses. Let T be the first hitting time of the level 1, i.e.,

$$T = \inf\{n \in \mathbb{N} : X_n = 0\}.$$

We will show soon that $\mathbb{P}[T < \infty] = 1$, but the optional sampling theorem fails dramatically when applied to $\{X_n\}_{n\in\mathbb{N}_0}$ and T:

$$\mathbb{E}[X_T] = 1 > 0 = \mathbb{E}[X_0].$$

Note that in Example 13.14, according to Proposition 13.1, we have $\mathbb{E}[X_{T\wedge n}] = 0$, but

$$1 = \mathbb{E}[X_T] = \mathbb{E}[\lim_n X_{T\wedge n}] \neq \lim_n \mathbb{E}[X_{T\wedge n}] = 0.$$

Therefore, the general optional sampling theorem cannot be derived from the bounded one because of the lack of convergence $X_{n\wedge T} \to X_T$ is $\mathcal{L}^1$. The conditions under which this happens are spelled out below. We remind the reader that while the value X_T is well defined for a stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ and a *finite-valued* stopping time T, problems arise when $\mathbb{P}[T = \infty] > 0$. In that case, we set $X_T = \lim_{n\to\infty} X_{T\wedge n}$ whenever the limit exists, a.s. Otherwise, we do not define X_T at all.

Proposition 13.15. (Optional Sampling under UI conditions) *For a martingale $\{X_n\}_{n\in\mathbb{N}_0}$, let T be a stopping time such that the stopped martingale $\{X_n^T\}_{n\in\mathbb{N}_0}$ is uniformly integrable. Then*

$$X_T = \lim_n X_{T\wedge n} \text{ exists a.s., } X_T \in \mathcal{L}^1 \text{ and } \mathbb{E}[X_T] = \mathbb{E}[X_0]. \tag{13.1}$$

If $\{X_n\}_{n\in\mathbb{N}_0}$ is a non-negative *martingale and T a stopping time for which (13.1) holds, then $\{X_n^T\}_{n\in\mathbb{N}_0}$ is uniformly integrable.*

Proof. Suppose, first, that $\{X_n^T\}_{n\in\mathbb{N}_0}$ is uniformly integrable. It follows immediately from the Proposition 12.12 in Lecture 12 that it converges a.s., and in $\mathcal{L}^1$ as $n \to \infty$. In particular, the random variable $X_T = \lim_{n\to\infty} X_{T\wedge n}$ is well defined. Moreover, by the same Proposition, we have $\mathbb{E}[X_0] = \mathbb{E}[X_0^T] = \mathbb{E}[\mathbb{E}[X_T|\mathcal{F}_0]] = \mathbb{E}[X_T]$.

Conversely, assume that (13.1) holds for a non-negative martingale $\{X_n\}_{n\in\mathbb{N}_0}$ and a stopping time T. For the sequence $\{Y_n\}_{n\in\mathbb{N}_0}$, defined by $Y_n = X_{T\wedge n}$, the bounded optional sampling theorem (Proposition 13.1) implies that $||Y_n||_{\mathcal{L}^1} = \mathbb{E}[Y_n] = \mathbb{E}[Y_0] = \mathbb{E}[X_0]$. Moreover, non-negativity of $\{X_n\}_{n\in\mathbb{N}_0}$ and (13.1) ensure that $Y_n \to X_T$, a.s., and $||Y_n||_{\mathcal{L}^1} \to ||X_T||_{\mathcal{L}^1}$. Proposition 12.9 in Lecture 12 now guarantees that $\{X_n^T\}_{n\in\mathbb{N}_0}$ is uniformly integrable. □

It is useful to have a practically-applicable criterion for the uniform integrability of the stopped martingale X^T. It boils down to good behavior of the process or the stopping time, or merely decent behavior of both.

Proposition 13.16. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a martingale, and let T be a stopping time. Then X^T is uniformly integrable in each of the following situations:*

1. *T is bounded*
2. *X is uniformly integrable*
3. *$\mathbb{E}[T] < \infty$ and $\mathbb{E}[|X_{n+1} - X_n|\,|\mathcal{F}_n] \leq C$, a.s., for all $n \in \mathbb{N}_0$, for some constant $C \geq 0$.*

Proof.

1. Suppose $T \leq N$, a.s. Then $|X_{T\wedge n}| \leq |X_1| + \cdots + |X_N| \in \mathcal{L}^1$, making the family $\{X_n^T\}_{n\in\mathbb{N}_0}$ uniformly dominated by a random variable in $\mathcal{L}^1$, and, therefore, uniformly integrable.

2. Since $\{X_n\}_{n\in\mathbb{N}_0}$ is a UI martingale, there exists $X \in \mathcal{L}^1$ such that $X_n = \mathbb{E}[X|\mathcal{F}_n]$, a.s., for each $n \in \mathbb{N}_0$. We multiply both sides by $\mathbf{1}_{\{T=k\}}$ and sum over $k \in \{0,\dots,n\}$ to get

$$\sum_{k=0}^{n} X_k \mathbf{1}_{\{T=k\}} = \sum_{k=0}^{n} \mathbb{E}[X|\mathcal{F}_k]\mathbf{1}_{\{T=k\}}.$$

If we add the term $X_n\mathbf{1}_{\{T>n\}}$, to both sides, we get

$$X_{T\wedge n} = \sum_{k=0}^{n} \mathbb{E}[X|\mathcal{F}_k]\mathbf{1}_{\{T=k\}} + \mathbb{E}[X|\mathcal{F}_n]\mathbf{1}_{\{T>n\}}. \tag{13.2}$$

Let $\varphi : [0,\infty) \to [0,\infty)$, $\varphi(0) = 0$ be a convex and increasing test function of uniform integrability for the one-element (and therefore UI) family $\{|X|\}$, so that $\mathbb{E}[\varphi(|X|)] < \infty$. Applying the convex function $x \mapsto \varphi(|x|)$ to both sides of (13.2) and using the conditional Jensen's inequality, we get

$$\begin{aligned}
\varphi(|X_m^T|) &= \sum_{k=0}^{n} \varphi(|\mathbb{E}[X|\mathcal{F}_k]|)\mathbf{1}_{\{T=k\}} + \varphi(|\mathbb{E}[X|\mathcal{F}_n]|)\mathbf{1}_{\{T>n\}} \\
&\leq \sum_{k=0}^{n} \mathbb{E}[\varphi(|X|)|\mathcal{F}_k]\mathbf{1}_{\{T=k\}} + \mathbb{E}[\varphi(|X|)|\mathcal{F}_n]\mathbf{1}_{\{T>n\}} \\
&= \sum_{k=0}^{n} \mathbb{E}[\varphi(|X|)\mathbf{1}_{\{T=k\}}|\mathcal{F}_k] + \mathbb{E}[\varphi(|X|)\mathbf{1}_{\{T>n\}}|\mathcal{F}_n].
\end{aligned}$$

Finally, it remains to take the expectation of both sides to obtain $\mathbb{E}[\varphi(\left|X_n^T\right|)] \leq \mathbb{E}[\varphi(|X|)] < \infty$, for all $n \in \mathbb{N}$.

3. The idea is to show that the family $\{X_n^T\}_{n\in\mathbb{N}_0}$ is, in fact, uniformly dominated by a random variable in $\mathcal{L}^1$, and, therefore, UI. We start with the identity

$$\begin{aligned}
X_n^T - X_0 &= \sum_{k=1}^{n}(X_k^T - X_{k-1}^T) = \sum_{k=1}^{\infty}(X_k^T - X_{k-1}^T)\mathbf{1}_{\{k\leq n\}} \\
&= \sum_{k=1}^{\infty}(X_k - X_{k-1})\mathbf{1}_{\{k\leq n\}}\mathbf{1}_{\{T\geq k\}},
\end{aligned}$$

which easily turns into the inequality

$$\left|X_n^T\right| \leq |X_0| + \sum_{k=1}^{\infty} |X_k - X_{k-1}| \mathbf{1}_{\{T\geq k\}},$$

the RHS of which does not depend on n. Consequently, it will be enough to show that $\sum_{k=1}^{\infty} \mathbb{E}[|X_k - X_{k-1}| \mathbf{1}_{\{T\geq k\}}] < \infty$. This follows, however, from fact that $\{T \geq k\} \in \mathcal{F}_{k-1}$, for $k \in \mathbb{N}$, so that

$$\sum_{k=1}^{\infty} \mathbb{E}[|X_k - X_{k-1}| \mathbf{1}_{\{T \geq k\}}] \leq \sum_{k=1}^{\infty} \mathbb{E}[\mathbb{E}[|X_k - X_{k-1}| \,|\mathcal{F}_{k-1}] \mathbf{1}_{\{T \geq k\}}]$$
$$\leq C \sum_{k=1}^{\infty} \mathbb{P}[T \geq k] = C\mathbb{E}[T] < \infty. \qquad \square$$

There is another set of conditions under which a form of the optional sampling theorem cabe established. This time, only an inequality can be obtained, but the assumptions are minimal.

Proposition 13.17. *Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a non-negative supermartingale, and let T be a stopping time. Then the limit $X_T = \lim_n X_{T \wedge n}$ is well-defined, $X_T \in \mathcal{L}^1$ and $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.*

Proof. The process X^T is also a non-negative supermartingale, so the martingale convergence theorem implies that the limit $X_T = \lim_n X_{T \wedge n}$ exists and that $X_T \in \mathcal{L}^1$. By Fatou's lemma, we have

$$\mathbb{E}[X_T] = \mathbb{E}[\lim_n X_{T \wedge n}] \leq \liminf_n \mathbb{E}[X_n^T] \leq \mathbb{E}[X_0]. \qquad \square$$

Propositions 13.15 and 13.17 can be used to establish the following, more general, optional-sampling theorem for submartingales. We leave the proof to the reader.

Problem 13.4. (Optional Sampling for Submartingales) Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a submartingale, and let T be a stopping time. Suppose that the family $\{X^+_{T \wedge n}\}_{n \in \mathbb{N}_0}$ is uniformly integrable. Then $X_T = \lim_n X_{T \wedge n}$ is well defined, $X_T \in \mathcal{L}^1$ and $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$.

In analogy with the σ-algebra $\mathcal{F}_n$, which models the information available at time $n \in \mathbb{N}$, we denote by $\mathcal{F}_T$ the σ-algebra which models the information available at the moment the stopping time T happens. More precisely,

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n, \ \forall \, n \in \mathbb{N}_0.\},$$

where $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$.

Proposition 13.18. *For a stopping time T, $\mathcal{F}_T$ is the σ-algebra generated by the family $\mathcal{S} = \{X_T : X \in \mathcal{X}\}$, where $\mathcal{X}$ denotes the set of all $\{\mathcal{F}_n\}_{n \in \mathbb{N}_0}$-adapted stochastic processes for which X_T is well-defined.*

Proof. Suppose, first, that $A = X_T^{-1}(B)$, for some $B \in \mathcal{B}(\mathbb{R})$ and $X \in \mathcal{X}$. Then,

$$A \cap \{T \leq n\} = \{X_T \in B, T \leq n\} = \cup_{k=0}^{n} \{X_k \in B, T = k\} \in \mathcal{F}_n.$$

Since the sets of the form $X_T^{-1}(B)$, $B \in \mathcal{B}(\mathbb{R})$ generate $\sigma(\mathcal{S})$, we conclude that $\sigma(\mathcal{S}) \subseteq \mathcal{F}_T$.

Conversely, for $A \in \mathcal{F}_T$, the process $X_n = \mathbf{1}_{A^c \cup \{n<T\}}$ is adapted and $X_T = \mathbf{1}_{A^c}$, so $A^c \in \sigma(\mathcal{S})$, and, consequently, $A \in \sigma(\mathcal{S})$. □

Proposition 13.19. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a uniformly integrable martingale, and let S and T be two stopping times with $T \leq S$, a.s. Then, $X_T, X_S \in \mathcal{L}^1$ and*

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

In particular,

$$\mathbb{E}[X_S] = \mathbb{E}[X_T].$$

Proof. We cover the case $S = \infty$ first. By Propositions 13.16 and 13.15 X_T is well-defined and in $\mathcal{L}^1$, and the family $X_{T\wedge n}$, $n \in \mathbb{N}_0$ is uniformly integrable. By Proposition 13.1, we have $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_n]$, and, upon letting $n \to \infty$, we get that

$$\mathbb{E}[X_T] = \mathbb{E}[X_\infty], \tag{13.3}$$

where $X_\infty = \lim_n X_n$, which exists a.s., thanks to the uniform integrability of $\{X_n\}_{n\in\mathbb{N}_0}$. We now take $A \in \mathcal{F}_T$, and define the stopping time T^A by $T^A = T\mathbf{1}_A + \infty\mathbf{1}_{A^c}$ (why is T^A a stopping time?). The equality (13.3), with T replaced by T^A implies that

$$\mathbb{E}[X_T\mathbf{1}_A] = \mathbb{E}[X_\infty\mathbf{1}_A],$$

which directly implies that $X_T = \mathbb{E}[X_\infty|\mathcal{F}_T]$.

To remove the assumption $S = \infty$, we consider the process $Y_n = X_n^S$ and use part (2) of Proposition 13.16. □

Square-integrable martingales

Definition 13.20. A martingale $\{X_n\}_{n\in\mathbb{N}_0}$ is said to be **square-integrable** (or an $\mathcal{L}^2$**-martingale**) if

$$X_n \in \mathcal{L}^2, \text{ for all } n \in \mathbb{N}_0.$$

Remark 13.21. The notion of a square-integrable martingale *does not coincide* with the notion of a martingale *bounded in* $\mathcal{L}^2$ of Definition 13.11. The simple random walk is an example of a square-integrable process which is not $\mathcal{L}^2$-bounded.

Square-integrable martingales have a number of interesting properties which fundamentally follow from the fact that the conditional expectation is a projection when restricted to $\mathcal{L}^2$:

Proposition 13.22. (Orthogonality of martingale increments) *For a square-integrable martingale $\{X_n\}_{n\in\mathbb{N}_0}$ we have*

$$\mathbb{E}[(X_{m_2} - X_{n_2})(X_{m_1} - X_{n_1})|\mathcal{F}_{n_1}] = 0, \tag{13.4}$$

and, in particular, $X_{m_2} - X_{n_2}$ and $X_{m_1} - X_{n_1}$ are orthogonal in $\mathbb{L}^2$. Furthermore,

$$\mathbb{E}[(X_{m_1} - X_{n_1})^2|\mathcal{F}_{n_1}] = \mathbb{E}[X_{m_1}^2 - X_{n_1}^2|\mathcal{F}_{n_1}], \tag{13.5}$$

for all $m_2 > n_2 \geq m_1 > n_1$.

Proof. Since $\mathbb{E}[X_{m_2}|\mathcal{F}_{m_1}] = X_{m_1} = \mathbb{E}[X_{n_2}|\mathcal{F}_{m_1}]$, we have $\mathbb{E}[(X_{m_2} - X_{n_2})|\mathcal{F}_{m_1}] = 0$, and (13.4) follows by multiplying both sides by $(X_{m_1} - X_{n_2}) \in \mathcal{F}_{m_1}$ and taking expectations. (13.5) is a direct consequence of (13.4). □

Remark 13.23. If one interprets a square-integrable martingale as a sequence of vectors (or a discrete path) in $\mathcal{L}^2$, the relation (13.4) implies that each new step of this path is orthogonal to all steps before it. In other words, each new step of a martingale enters a new dimension.

If $\{X_n\}_{n\in\mathbb{N}_0}$ is a square-integrable martingale, Jensen's inequality implies that the process $\{X_n^2\}_{n\in\mathbb{N}_0}$ is a submartingale, and, thus, admits a Doob-Meyer decomposition into a martingale $\{M_n\}_{n\in\mathbb{N}_0}$ and a predictable non-decreasing process $\{A_n\}_{n\in\mathbb{N}_0}$ with $A_0 = 0$. The process $\{A_n\}_{n\in\mathbb{N}_0}$ - often denoted by $\{\langle X\rangle_n\}_{n\in\mathbb{N}_0}$ - plays an important role in the study of square-integrable martingales and is called the **(predictable) quadratic variation** or the **increasing process associated with** X. It is easy to see that

$$\langle X\rangle_n - \langle X\rangle_{n-1} = \mathbb{E}[(X_n - X_{n-1})^2|\mathcal{F}_{n-1}], \text{ for } n \in \mathbb{N}.$$

Therefore, we can think of the increment of $\langle X\rangle$ as an estimate of the variance of the next increment of X, conditional on the present information. Aggregation of all increments leads to a (somewhat loose) interpretation of $\langle X\rangle$ as a path-by-path measure of the variation of the process $\{X_n\}_{n\in\mathbb{N}_0}$. To limit $\langle X\rangle_\infty = \lim_n \langle X\rangle_n \in [0,\infty]$ is especially important, as it measures the total "action" of the martingale $\{X_n\}_{n\in\mathbb{N}_0}$:

Proposition 13.24. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a square-integrable martingale. Then X is bounded in $\mathcal{L}^2$ if and only if $\mathbb{E}[\langle X\rangle_\infty] < \infty$. In that case, we have*

$$\mathbb{E}[(X_\infty^*)^2] \leq \mathbb{E}[X_0^2] + 4\mathbb{E}[\langle X\rangle_\infty]. \tag{13.6}$$

Proof. Since $M = X^2 - \langle X\rangle$ is a martingale, we have $\mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \mathbb{E}[\langle X\rangle_n]$, so $\sup_n \mathbb{E}[X_n^2] < \infty$ if and only if $\sup_n \mathbb{E}[\langle X\rangle_n] = \mathbb{E}[\langle X\rangle_\infty] < \infty$. The estimate (13.6) follows from the maximal inequality 13.9. □

The larger $\langle X \rangle$, the more we expect the martingale X to oscillate. In fact, many pathwise properties of a square-integrable martingale X can be deduced from the properties of $\langle X \rangle$. Here is an example:

Proposition 13.25. *Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a square-integrable martingale. The limit $\lim_n X_n$ exists a.s. on the set $\{\langle X \rangle_\infty < \infty\}$.*

Proof. For $N \in \mathbb{N}$, consider the random times

$$T_N = \min\{n \in \mathbb{N}_0 : \langle X \rangle_{n+1} \geq N\}.$$

Note that T_N is a stopping time because $\langle X \rangle$ is a predictable process. Using the fact that $X^2 - \langle X \rangle$ is a martingale, by the bounded optional sampling theorem, we have

$$\mathbb{E}[X^2_{T_N \wedge n}] = \mathbb{E}[X_0^2] + \mathbb{E}[\langle X \rangle_{T_N \wedge n}], \text{ for all } N, n \in \mathbb{N}..$$

On the other hand, due to the form of the definition of T_N, we have $\langle X \rangle_{T_N \wedge n} \leq N$, a.s., so

$$\mathbb{E}[X^2_{T_N \wedge n}] \leq \mathbb{E}[X_0^2] + N.$$

It follows that X^{T_N} is a martingale bounded in $\mathcal{L}^2$ for each $N \in \mathbb{N}$. In particular, the limit $\lim_{n \to \infty} X_n^{T_N}$ exists.

For almost all $\omega \in \{\langle X \rangle_\infty < \infty\}$, there exists $N \in \mathbb{N}$ such that $T_N(\omega) = \infty$. Therefore, for almost all $\omega \in \{\langle X \rangle_\infty < \infty\}$, the trajectories $X_n(\omega)$, $n \in \mathbb{N}$ and $X^{T_N}(\omega)$, $n \in \mathbb{N}$ coincide if large enough N is chosen. Consequently, the limit $\lim_n X_n(\omega)$ exists for almost all $\omega \in \{\langle X \rangle_\infty < \infty\}$. □

Problem 13.5. Construct an example of a square-integrable martingale $\{X_n\}_{n \in \mathbb{N}_0}$ with $\langle X \rangle_\infty = \infty$, a.s., such that $\lim_n X_n$ exists a.s.

Additional Problems

Problem 13.6 (Ruin in insurance)**.** The total yearly income stream (premium income minus claims paid) of an insurance company is modeled by a normal random variable with mean $\mu > 0$ and variance $\sigma^2 > 0$. If the company starts with (a deterministic) initial capital of $X_0 > 0$ and the income streams are assumed to be independent from year to year, show that

$$\mathbb{P}[R] \leq e^{-2X_0 \frac{\mu}{\sigma^2}},$$

where R denotes the event in which the total capital falls below 0 at some point in time.

Hint: The process $\exp(\theta X_n)$ is a martingale for some $\theta \in \mathbb{R}$.

Problem 13.7 (Optional Sampling for Submartingales)**.** Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a submartingale, and let T be a stopping time such that the family $\{X^+_{T\wedge n}\}_{n\in\mathbb{N}_0}$ is uniformly integrable. Show that $X_T = \lim_n X_{T\wedge n}$ is well defined, $X_T \in \mathcal{L}^1$ and $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$.

Problem 13.8 (Hitting times for the simple random walk)**.** Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a simple random walk starting from 1, i.e., $X_0 = 1, X_n = 1 + \sum_{k=1}^n \xi_k, n \geq 1$, where $\{\xi_n\}_{n\in\mathbb{N}}$ are iid with $\mathbb{P}[\xi_k = 1] = \mathbb{P}[\xi_k = -1] = \frac{1}{2}$.

1. For $\theta \in \mathbb{R}$ define $Y^\theta_n = e^{-\theta X_n - \alpha(\theta) n}$, $n \geq 0$, where $\alpha(\theta) = \log(\cosh(\theta))$. Show that $\{Y^\theta_n\}_{n\in\mathbb{N}_0}$ is a martingale for each $\theta \in \mathbb{R}$.
2. Define $T = \inf\{n \in \mathbb{N}_0 : X_n = 0\}$ be the first hitting time of the level 0. By using the optional sampling theorem applied to the process $\{Y^\theta_n\}_{n\in\mathbb{N}_0}$, $\theta > 0$ and the stopping time T, find an expression for the moment-generating function $M_T(\gamma) = \mathbb{E}[\exp(-\gamma T)]$, $\gamma \geq 0$, where we define $\exp(-\gamma T) = 0$ when $T = +\infty$, $\gamma > 0$.
3. Compute $\mathbb{P}[T < \infty]$ and $\mathbb{E}[T]$.

Problem 13.9 (Kolmogorov's "Three Series" Theorem)**.** Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of independent (not necessarily identically distributed) random variables, $K > 0$ and

$$\xi^K_n = \xi_n \mathbf{1}_{\{|\xi_n|\leq K\}}, \text{ for all } n \in \mathbb{N}.$$

The Three-Series Theorem states that

$\sum_n \xi_n$ converges a.s., if and only if the following three statements hold:

a. $\sum_n \mathbb{P}[|\xi_n| \geq K] < \infty$,

b. $\sum_n \mathbb{E}[\xi^K_n] < \infty$, *and*

c. $\sum_n \mathrm{Var}[\xi^K_n] < \infty$.

We only deal with the implication "Three series converge $\Rightarrow \sum_n \xi_n$ converges" in this problem:

1. Start by arguing that it is enough to show that $\sum_n(\xi^K_n - \mathbb{E}[\xi^K_n])$ converges.
2. Use the martingale properties of the process

$$X_n = \sum_{k=1}^n \hat{\xi}^K_n \text{ where } \hat{\xi}^K_n = \xi^K_n - \mathbb{E}[\xi^K_n]$$

   to complete the argument.

Problem 13.10 (Random signs)**.** Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of non-negative real numbers, and let $\{\xi_n\}_{n\in\mathbb{N}}$ be an iid sequence with $\mathbb{P}[\xi_n = 1] = \mathbb{P}[\xi_n = -1] = \frac{1}{2}$, for $n \in \mathbb{N}$ (we interpret $\{\xi_n\}_{n\in\mathbb{N}}$ as a sequence of random signs). The purpose of this problem is to analyze a.s.-convergence of the sequence $\{X_n\}_{n\in\mathbb{N}_0}$, where $X_0 = 0$ and

$$X_n = \sum_{k=1}^{n} \xi_k a_k, \ n \in \mathbb{N},$$

obtained from the original sequence by giving its elements random signs, and to prove the following result, called the **Khintchine's inequality**:

For each $p \in [1, \infty)$, there exists constants c_p and C_p such that

$$c_p||a||_{\ell^2} \leq || \textstyle\sum_{n=1}^{\infty} \xi_n a_n ||_{\mathcal{L}^p} \leq C_p||a||_{\ell^2}, \tag{13.7}$$

for all sequences $\{a_n\}_{n\in\mathbb{N}}$ with $||a||_{\ell^2} := \sqrt{\sum_{n=1}^{\infty} a_n^2} < \infty$.

1. Suppose that $\sum_n a_n^2 < \infty$. Show that $\sum_n \xi_n a_n$ converges a.s.
2. Prove that the condition $\sum a_n^2 < \infty$ is also necessary for a.s.-convergence of $\sum_n \xi_n a_n$.
3. Find an expression for the moment-generating function $\lambda \mapsto M(\lambda) = \mathbb{E}[e^{\lambda X}]$ of $X = \lim_n X_n$ and show that it satisfies the following bound: $M(\lambda) \leq \frac{1}{2} e^{c^2\lambda^2}$ for all $\lambda \in \mathbb{R}$, where $c = ||a||_{\ell^2}^2$.
4. Use the symmetry of X and the above bound to conclude that

$$\mathbb{P}[|X| \geq x] \leq 2e^{-\frac{x^2}{2c^2}} \text{ for all } x \geq 0,$$

   and deduce from there the right-hand side of (13.7).
5. Use the right-hand side of (13.7), together with Hölder's inequality do deduce the left-hand side of (13.7) (in the nontrivial case $p < 2$).

Hint: 1. Show that $\{X_n\}_{n\in\mathbb{N}_0}$ is a martingale bounded in $\mathcal{L}^2$. 2. Argue, first, that it can be assumed that $\sup_n a_n < \infty$. Then, prove that the stopping times $T_k = \inf\{n \in \mathbb{N} : |X_n| \geq k\}$, $k \in \mathbb{N}$, have the property that $\mathbb{E}[\langle X \rangle_{T_k}] < \infty$, for all $k \in \mathbb{N}$. Conclude that $\mathbb{P}[T_k < \infty] = 1$, for all $k \in \mathbb{N}$, unless $\sum_n a_n^2 < \infty$.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 14
Brownian Motion

The goal of this chapter is to introduce and study some of the simpler properties of, arguably, the most important continuous-time process - the Brownian motion. Before we do, however, we go over the definition and some of the most-important features of the multivariate normal distribution. We interpret the vector $\boldsymbol{t} \in \mathbb{R}^n$ as a column vector, i.e., we identify $\mathbb{R}^n$ with ahe space $\mathbb{R}^{n\times 1}$ of $n \times 1$-matrices.

The multivariate normal distribution

Several of the important properties of the multivariate distribution where already covered in Problems 6.16 and 8.7 in Lectures 6 and 8. For reader's convenience, we revisit them here. We also remind the reader that the characteristic function (Fourier transform) φ_γ of a measure γ on $\mathcal{B}(\mathbb{R}^n)$ is defined by $\varphi_\gamma(\boldsymbol{t}) = \int_{\mathbb{R}^n} \exp(i\boldsymbol{t}^\tau \boldsymbol{x})\gamma(d\boldsymbol{x})$, for $\boldsymbol{t} \in \mathbb{R}^n$, where $()^\tau$ denotes transposition.

Definition 14.1 (Gaussian measures)**.** A probability measure γ on $\mathbb{R}^n$ is said to be **Gaussian** with **mean (vector)** $\boldsymbol{\mu}$ and **variance-covariance (matrix)** $\boldsymbol{\Sigma}$ if its characteristic function φ_γ is of the form

$$\varphi_\gamma(\boldsymbol{t}) = \exp(i\boldsymbol{t}^\tau\boldsymbol{\mu} - \tfrac{1}{2}\boldsymbol{t}^\tau\boldsymbol{\Sigma}\boldsymbol{t}), \text{ for } \boldsymbol{t} \in \mathbb{R}^n. \tag{14.1}$$

A n-dimensional random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ is said to be **Gaussian** or to have the **(multivariate) normal distribution** if its distribution is a Gaussian measure on $\mathbb{R}^n$.

Remark 14.2.

1. "Normal", "Gaussian" and "multivariate normal" are all in use to describe the random vector with a Gaussian distribution. On the other hand, one rarely, if ever, talks about the "normal measure" on $\mathbb{R}^n$. Here, it seems, one has to say "Gaussian".

   When $\boldsymbol{\mu} = 0$, we say that $\boldsymbol{X}$ is **centered normal** or **centered Gaussian**. Similarly, we talk about **centered Gaussian measures**.

2. It follows directly from the finiteness of the Gaussian measure that $\mathbf{\Sigma}$ can be taken to be symmetric and is necessarily nonnegative definite. Even more directly, differentiation yields the expressions

$$\mu_i = \int_{R^n} x_i \, \gamma(d\boldsymbol{x}) \text{ and } \Sigma_{ij} = \int_{\mathbb{R}^n} (x_i - \mu_i)(x_j - \mu_j) \, \gamma(d\boldsymbol{x}),$$

for $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ and $\mathbf{\Sigma} = (\Sigma_{ij})_{1 \le i,j \le n}$.

3. The uniqueness theorem for the (multivariate) characteristic functions implies immediately that a Gaussian measure is uniquely determined by $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. The converse - which states that given any $\boldsymbol{\mu} \in \mathbb{R}^n$ and any *non-negative definite* matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$, the prescription (14.1) indeed defines a measure on $\mathbb{R}^n$ - will emerge as a consequence of Proposition 14.3 below.

4. A special, and well-known, case of the Gaussian measure on $\mathbb{R}^n$ for $n = 1$ is the **unit** or **standard Gaussian measure on** $\mathbb{R}$, which corresponds to $\mu = 0$ and $\Sigma = 1$. One readily checks that the measure γ on $\mathcal{B}(\mathbb{R})$, whose density with respect to the Lebesgue measure is given by $x \mapsto \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$ is, in fact, the unit Gaussian measure.

5. Note that, under this definition, the Dirac measures $\delta_{\boldsymbol{\mu}}$ are also Gaussian, for any $\boldsymbol{\mu} \in \mathbb{R}^n$.

Proposition 14.3 (The structure theorem for Gaussian random variables)**.** *The following are equivalent for a random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ on $\mathbb{R}^n$:*

1. *$\boldsymbol{X}$ is Gaussian.*
2. *For each $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$, the linear combination $\sum_{i=1}^n a_i X_i$ is normally distributed.*
3. *Either $\boldsymbol{X}$ is (deterministically) constant, or there exist*
   (a) *an integer $d \in \{1, \dots, n\}$,*
   (b) *a vector $\boldsymbol{\mu} \in \mathbb{R}^n$,*
   (c) *a rank-d matrix $\boldsymbol{A} \in \mathbb{R}^{d \times n}$ and*
   (d) *a random vector $\boldsymbol{Y} = (Y_1, \dots, Y_d)$, defined on the same probability space as $\boldsymbol{X}$, which consists of independent unit normals*

   such that
$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Y}.$$

Proof. 2. $\to$ 1. Normal random variables are in $\mathcal{L}^2$, so the mean vector $\boldsymbol{\mu}$ and the variance-covariance matrix $\mathbf{\Sigma}$, given by

$$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]),$$
$$\mathbf{\Sigma} = \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\tau] = (\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])])_{1 \le i,j \le n}$$

are well defined. The basic observation is that the characteristic function $\varphi_{\boldsymbol{X}}$ of $\boldsymbol{X}$ can be recovered from that of $Z = \boldsymbol{a}^\tau \boldsymbol{X} = \sum_{i=1}^n a_i X_i$ as follows:

$$\varphi_{\boldsymbol{X}}(\boldsymbol{a}) = \mathbb{E}[e^{i\boldsymbol{a}^\tau \boldsymbol{X}}] = \varphi_{\boldsymbol{a}^\tau \boldsymbol{X}}(1) = \varphi_Z(1) = e^{i\mathbb{E}[Z] - \frac{1}{2}\mathbb{E}[(Z-\mathbb{E}[Z])^2]}$$
$$= e^{i\boldsymbol{a}^\tau \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{a}^\tau \boldsymbol{\Sigma} \boldsymbol{a}}.$$

3. $\to$ 2. Follows by direct computation.

1. $\to$ 3. By replacing $\boldsymbol{X}$ by $\boldsymbol{X} - \boldsymbol{\mu}$, where $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}]$, we can assume, without loss of generality, that $\mathbb{E}[\boldsymbol{X}] = 0$. We also assume that $\boldsymbol{X}$ is not constant. The matrix $\boldsymbol{\Sigma}$ is symmetric and nonnegative definite, so $\mathbb{R}^n$ admits an orthonormal basis $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ consisting of its eigenvectors. We assume that the first d of then (where $1 \leq d \leq n$) correspond to strictly positive eigenvalues $\lambda_1, \dots, \lambda_d$, while the remaining span the null space of $\boldsymbol{\Sigma}$. Let $\boldsymbol{T} \in \mathbb{R}^{n \times d}$ denote a matrix of a linear operator which maps the canonical basis $\boldsymbol{f}_1, \dots, \boldsymbol{f}_d$ of $\mathbb{R}^d$ into $\boldsymbol{e}_1, \dots, \boldsymbol{e}_d$, respectively, so that

$$\boldsymbol{T}^\tau \boldsymbol{\Sigma} \boldsymbol{T} = \operatorname{diag}(\lambda_1, \dots, \lambda_d), \boldsymbol{T}^\tau \boldsymbol{T} = \boldsymbol{I}_{\mathbb{R}^d} \text{ and } \boldsymbol{T}\boldsymbol{T}^\tau = \boldsymbol{I}_{\operatorname{span}(\boldsymbol{e}_1, \dots, \boldsymbol{e}_d)}.$$

Since

$$\varphi_{\boldsymbol{T}^\tau \boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[e^{i\boldsymbol{t}^\tau \boldsymbol{T}^\tau \boldsymbol{X}}] = \varphi_{\boldsymbol{X}}(\boldsymbol{T}\boldsymbol{t}) = e^{-\frac{1}{2}\boldsymbol{t}^\tau \boldsymbol{T}^\tau \boldsymbol{\Sigma} \boldsymbol{T} \boldsymbol{t}} = e^{-\frac{1}{2}\boldsymbol{t}^\tau \operatorname{diag}(\lambda_1, \dots, \lambda_d)\boldsymbol{t}},$$

for $\boldsymbol{t} \in \mathbb{R}^d$, we conclude that $\boldsymbol{Y} = \operatorname{diag}(\lambda_1^{-1/2}, \dots, \lambda_d^{-1/2})\, \boldsymbol{T}^\tau \boldsymbol{X}$ is a vector of d independent unit normals and that $\boldsymbol{A}\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{T}^\tau \boldsymbol{X}$, for $\boldsymbol{A} = \boldsymbol{T} \operatorname{diag}(\lambda_1^{1/2}, \dots, \lambda_d^{1/2})$. It remains to observe that $\boldsymbol{T}\boldsymbol{T}^\tau \boldsymbol{X} = \boldsymbol{X}$, a.s. On one hand, we clearly have $||\boldsymbol{T}\boldsymbol{T}^\tau \boldsymbol{x}|| \leq ||\boldsymbol{x}||$ for all $\boldsymbol{x} \in \mathbb{R}^n$. On the other, the random vectors have the same characteristic function, and are, therefore, equally distributed. Consequently (why?) $\boldsymbol{X} = \boldsymbol{T}\boldsymbol{T}^\tau \boldsymbol{X}$, a.s. □

Corollary 14.4 (Support of a Gaussian measure)**.** *The support of a Gaussian measure on $\mathbb{R}^n$ with mean $\boldsymbol{\mu}$ and variance-covariance $\boldsymbol{\Sigma}$ is the following affine subspace of $\mathbb{R}^n$*

$$\boldsymbol{\mu} + \operatorname{Im} \boldsymbol{\Sigma} = \{\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{x} \,:\, \boldsymbol{x} \in \mathbb{R}^n\}.$$

Corollary 14.5 (Existence of Gaussian measures with prescribed parameters)**.** *For each $n \in \mathbb{N}$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and symmetric nonnegative-definite $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, there exists a Gaussian measure on $\mathbb{R}^n$ with mean $\boldsymbol{\mu}$ and variance-covariance $\boldsymbol{\Sigma}$.*

Remark 14.6. The smallest integer d from part 3. of Proposition 14.3 is called the **rank** of the normal distribution and corresponds to the dimension of the range of the variance-covariance matrix $\boldsymbol{\Sigma}$. It also corresponds to the dimension of the support of $\boldsymbol{X}$; when $d = n$, we say that the distribution of $\boldsymbol{X}$ is **non-degenerate**. Otherwise, we talk about a **degenerate normal distribution**.

Proposition 14.7 (Absolute continuity of Gaussian random vectors)**.** *A normally distributed random vector $\boldsymbol{X}$ on $\mathbb{R}^n$ is absolutely continuous if and only if it is nondegenerate. In that case, its density $f_{\boldsymbol{X}} : \mathbb{R}^n \to [0, \infty)$ is given by*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} \exp\Big(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\tau \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big),\ \boldsymbol{x} \in \mathbb{R}^n, \quad (14.2)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean and the variance-covariance matrix of $\boldsymbol{X}$.

Proof. In the degenerate case, $\boldsymbol{X}$ is not absolutely continuous because its support is contained in a proper affine subspace of $\mathbb{R}^n$ - a set of Lebesgue measure 0. In the non-degenerate case, one simply checks that the characteristic function corresponding to $f_{\boldsymbol{X}}$ given in (14.2) matches that in the definition of the normal distribution. □

The notion of conditional normality will appear several times later on in our treatment of Brownian motion.

Definition 14.8 (Conditional normal distribution)**.** An n-dimensional random vector $\boldsymbol{X}$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is said to be **conditionally normally distributed** with respect to the σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ if there exist

1. a random vector $\boldsymbol{\mu}_{X|\mathcal{G}}$
2. a random element $\boldsymbol{\Sigma}_{X|\mathcal{G}}$, with values in the space of symmetric nonnegative definite matrices,

such that $\boldsymbol{\mu}_{X|\mathcal{G}}$ and $\boldsymbol{\Sigma}_{X|\mathcal{G}}$ are $\mathcal{G}$-measurable and

$$\mathbb{E}[e^{i\boldsymbol{t}^\tau \boldsymbol{X}}|\mathcal{G}] = \exp\Big(i\boldsymbol{t}^\tau \boldsymbol{\mu}_{X|\mathcal{G}} - \tfrac{1}{2}\boldsymbol{t}^\tau \boldsymbol{\Sigma}_{X|\mathcal{G}} \boldsymbol{t}^\tau \Big), \text{ a.s., for all } \boldsymbol{t} \in \mathbb{R}^n.$$

Random elements $\boldsymbol{\mu}_{X|\mathcal{G}}$ and $\boldsymbol{\Sigma}_{X|\mathcal{G}}$ are called the **conditional mean** and the **conditional variance-covariance matrix** of $\boldsymbol{X}$, given $\mathcal{G}$.

Remark 14.9. Note that, for a conditionally normal random variable $\boldsymbol{X}$, with respect to $\mathcal{G}$, we necessarily have

$$\boldsymbol{\mu}_{X|\mathcal{G}} = \mathbb{E}[\boldsymbol{X}|\mathcal{G}] \text{ and } \boldsymbol{\Sigma}_{X|\mathcal{G}} = \mathbb{E}[(\boldsymbol{X}-\boldsymbol{\mu}_{X|\mathcal{G}})(\boldsymbol{X}-\boldsymbol{\mu}_{X|\mathcal{G}})^\tau|\mathcal{G}],$$

so that the terminology of Definition 14.8 matches the usual one. This observation makes it not too difficult to show that a random variable is conditionally normal in the sense of Definition 14.8 if and only if its $\mathcal{G}$-conditional distribution is normal, a.s.

The following simple, but important observation, follows directly from the definition of conditional normality.

Proposition 14.10 (A test for independence under conditional normality)**.** *Let $\boldsymbol{X}$ be conditionally normal with respect to $\mathcal{G}$. Then $\boldsymbol{X}$ is independent of $\mathcal{G}$ if and only if the random elements $\mathbb{E}[\boldsymbol{X}|\mathcal{G}]$ and $\mathbb{E}[(\boldsymbol{X}-\mathbb{E}[\boldsymbol{X}|\mathcal{G}])(\boldsymbol{X}-\mathbb{E}[\boldsymbol{X}|\mathcal{G}])^\tau|\mathcal{G}]$ are constant, a.s.*

A common example of conditional normality arises when one generates a σ-algebra by a subset of coordinates of a normal random vector.

Proposition 14.11 (Conditional distribution of normal components)**.** *Let $\boldsymbol{X}^o$ and $\boldsymbol{X}^u$ be random vectors in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, such that $\boldsymbol{X}=(\boldsymbol{X}^o,\boldsymbol{X}^u)\in\mathbb{R}^{m+n}$ is multivariate normal. Let $\boldsymbol{\mu}=(\boldsymbol{\mu}^o,\boldsymbol{\mu}^u)$ and $\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{ou}\\ \boldsymbol{\Sigma}_{uo} & \boldsymbol{\Sigma}_{uu}\end{pmatrix}$ be block-decompositions of the mean vector and the variance-covariance matrix into blocks with sizes m and n, and $m\times m$, $m\times n$, $n\times m$, and $n\times n$, respectively. Then*

1. *both $\boldsymbol{X}^o$ and $\boldsymbol{X}^u$ are normally distributed,*
2. *$\boldsymbol{X}^o$ and $\boldsymbol{X}^u$ are independent if and only if they are uncorrelated, i.e., if $\boldsymbol{\Sigma}_{ou}=0$.*
3. *$\boldsymbol{X}^u$ is conditionally normal with respect to $\mathcal{G}=\sigma(\boldsymbol{X}^o)$. When $\boldsymbol{\Sigma}_{oo}$ is invertible, the conditional mean and variance given by*

$$\boldsymbol{\mu}_{\boldsymbol{X}^u|\mathcal{G}}=\boldsymbol{\mu}^o+\boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}(\boldsymbol{X}^o-\boldsymbol{\mu}^o),\ \boldsymbol{\Sigma}_{\boldsymbol{X}^u|\mathcal{G}}=\boldsymbol{\Sigma}_{uu}-\boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{ou}. \quad (14.3)$$

Proof.

1. Follows immediately from Proposition 14.3, part 2.
2. When $\boldsymbol{\Sigma}_{uo}=0$, one readily checks that the characteristic function of $\boldsymbol{X}$ splits into a (tensor) product of the characteristic functions for $\boldsymbol{X}_o$ and $\boldsymbol{X}_u$. The converse direction is always true.
3. Set $\tilde{\boldsymbol{X}}^o=\boldsymbol{X}^o-\boldsymbol{\mu}_o$ and $\tilde{\boldsymbol{X}}^u=\boldsymbol{X}^u-\boldsymbol{\mu}_u$, so that the matrix $A=\boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}$ has the property that $\mathbb{E}[(\tilde{\boldsymbol{X}}^u-A\tilde{\boldsymbol{X}}^o)(\tilde{\boldsymbol{X}}^o)^T]=0$, i.e., that the random vectors $\tilde{\boldsymbol{X}}^o-A\tilde{\boldsymbol{X}}^o$ and $\tilde{\boldsymbol{X}}^o$ are uncorrelated. By 2. above, they are also independent. Therefore, the conditional characteristic function of $\tilde{\boldsymbol{X}}^o-A\tilde{\boldsymbol{X}}^o$, given $\mathcal{G}=\sigma(\tilde{\boldsymbol{X}}^o)$ is deterministic and of the form
$$\mathbb{E}[e^{i\boldsymbol{t}(\tilde{\boldsymbol{X}}^u-A\tilde{\boldsymbol{X}}^o)}|\mathcal{G}]=\varphi_{\tilde{\boldsymbol{X}}^u-A\tilde{\boldsymbol{X}}^o}(\boldsymbol{t}),\ \text{for } \boldsymbol{t}\in\mathbb{R}^n.$$
Since $A\tilde{\boldsymbol{X}}^o$ is $\mathcal{G}$-measurable, we have
$$\mathbb{E}[e^{i\boldsymbol{t}\boldsymbol{X}^u}|\mathcal{G}]=e^{i\boldsymbol{t}\boldsymbol{\mu}^u}e^{i\boldsymbol{t}A\tilde{\boldsymbol{X}}^o}e^{-\frac{1}{2}\boldsymbol{t}^T\hat{\boldsymbol{\Sigma}}\boldsymbol{t}},\ \text{for } \boldsymbol{t}\in\mathbb{R}^n.$$

where $\hat{\mathbf{\Sigma}} = \mathbb{E}[(\tilde{\boldsymbol{X}}^u - A\tilde{\boldsymbol{X}}^o)(\tilde{\boldsymbol{X}}^u - A\tilde{\boldsymbol{X}}^o)^T]$. The representations in (14.3) now follow. □

Remark 14.12. The assumption of invertibility of $\mathbf{\Sigma}_{00}$ in the part 3. above is quite harmless. Indeed, when $\mathbf{\Sigma}_{oo}$ is singular, we can find a subset of components of $\boldsymbol{X}^o$ whose variance-covariance matrix in invertible and which generate the same σ-algebra, up to null sets.

Gaussian processes

The notion of a Gaussian process is a natural generalization of a normal random vector.

Definition 14.13 (Continuous-time stochastic processes)**.** A **continuous-time stochastic process** is a family $\{X_t\}_{t\in[0,\infty)}$ of random variables defined on the same probability space and indexed by the set $[0,\infty)$. Each real-valued function $t \mapsto X_t(\omega)$, for $\omega \in \Omega$ is called the **trajectory** or the **path** of the stochastic process $\{X_t\}_{t\in[0,\infty)}$.

Remark 14.14.

1. The choice of $[0,\infty)$ for the time-set is prevalent, but not the only one we will use. For example, processes defined only up to a finite **time horizon** $T > 0$ are quite common. We leave to the reader to discern which results and concepts in the sequel need the time-set $[0,\infty)$, and which can be easily transfered to the finite-horizon setting $[0,T]$. Other time sets, such as $\mathcal{Q}_+$, the set of all nonnegative rationals, or $[t_0,\infty)$ for $t_0 > 0$ are also in use.

2. It is important to stress that each X_t is assumed to be $(-\infty,\infty)$-valued, and that in the continuous-time setting we never identify a.s.-equal random variables, unless such an idenfication is explicitely made. The reason is simple; we are dealing with uncountable families of random variables, and the exceptional sets can easily pile up. On a rare occasion, we will need our processes to take a value outside of $\mathbb{R}$; any such instance will be explicitely announced.

Definition 14.15 (Finite-dimensional distributions of a process). For $n \in \mathbb{N}$, the family of probability measures $\mu_{t_1,\dots,t_n}$ on $\mathcal{B}(\mathbb{R}^n)$ indexed by all n-tuples $(t_1,\dots,t_n) \in [0,\infty)^n$ is said to be the **(family of)** $\boldsymbol{n}$**-dimensional distributions** of the stochastic process $\{X_t\}_{t\in[0,\infty)}$ if

$$\mu_{t_1,\dots,t_n} \text{ is the distribution of the random vector } (X_{t_1},\dots,X_{t_n}),$$

for all $t_1,\dots,t_n \in [0,\infty)$. The collection of all n-dimensional distribution, with n ranging through $\mathbb{N}$ is called the **(family of) finite-dimensional distributions** of the process $\{X_t\}_{t\in[0,\infty)}$.

Definition 14.16 (Gaussian processes). A stochastic process $\{X_t\}_{t\in[0,\infty)}$ is said to be **Gaussian** if all of its finite-dimensional distributions are multivariate normal.

Since the normal distribution is uniquely determined by its mean and variance-covariance, the finite-dimensional distributions of a Gaussian process are fully determined by two functions - the **mean function** $\mu : [0,\infty) \to \mathbb{R}$ and the **covariance function** $\Sigma : [0,\infty) \times [0,\infty) \to \mathbb{R}$. These functions are given by

$$\mu(t) = \mathbb{E}[X_t] \text{ and } \Sigma(s,t) = \operatorname{Cov}[X_s, X_t].$$

A Gaussian process is said to be **centered** if $\mu(t) = 0$, for all $t \geq 0$.

The definition of the Brownian Motion

Definition 14.17 (Brownian motion)**.** A continuous-time stochastic process $\{B_t\}_{t\in[0,\infty)}$ is called a **Brownian motion** if

1. $B_0 = 0$ and $B_t - B_s \sim N(0, t-s)$, for all $0 \leq s < t$,
2. the increments $B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ are independent for all $0 \leq t_1 < t_2 < \dots < t_n$.
3. There exists $\Omega^* \in \mathcal{F}$ such that $\mathbb{P}[\Omega^*] = 1$ and the trajectory $t \mapsto B_t(\omega)$ is a continuous function for each $\omega \in \Omega^*$.

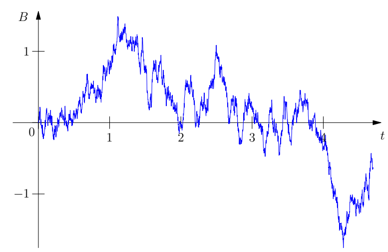


Figure 1:
A typical sample path of the Brownian Motion

The first two requirements specify the finite dimensional distributions (distributions of random vectors $(B_{t_1}, B_{t_2}, \dots, B_{t_n})$ for all $0 \leq t_1 < t_2 < \dots < t_n$) of the process, while the third one has to do with selection of the representatives in the a.s.-equivalence classes for each of the random variables $\{B_t\}_{t\in[0,\infty)}$; it is often abbreviated to "almost all trajectories of B are continuous".

Proposition 14.18 (Brownian motion as a Gaussian process)**.** *A stochastic process $\{X_t\}_{t\in[0,\infty)}$ is a Brownian motion if and only if*

1. *it is a centered Gaussian process with $\Sigma(s,t) = \min(s,t)$, for all $s,t \in [0,\infty)$, and*
2. *almost all of its trajectories are continuous.*

Proof. Exercise. □

The Lévy-Ciesielski Construction

A priori, Definition 14.17 could be entirely vacuous. Indeed, one can show that no stochastic process satisfies Definition 14.17, with the normal distribution replaced by, say, the Cauchy distribution. Fortunately, the Brownian motion does exists, and there are several, quite diverse, well-known constructions (existence proofs). We present one of them in detail, and only outline some of the others.

Before we start, we make a useful observations, namely, that one can easily construct the full Brownian motion by concatenation, if one knows how to do it on $[0,1]$. Here is a more precise statement, the proof of which we leave to the reader:

Problem 14.1 (A construction on [0,1] will suffice)**.** Let $\{X_t\}_{t\in[0,1]}$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the following properties:

1. X is centered Gaussian with $\Sigma(s,t) = s \wedge t$, for all $s,t \in [0,1]$, and
2. almost all of its trajectories are continuous.

Consider a countable product $\hat{\Omega} = \times_{i=1}^{\infty}\Omega$, endowed with the product measure $\hat{\mathbb{P}} = \otimes_{i=1}^{\infty}\mathbb{P}$, the product σ-algebra $\hat{\mathcal{G}} = \otimes_{i=1}^{\infty}\mathcal{F}$, and the stochastic process $\{B_t\}_{t\in[0,\infty)}$, defined on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ by the following prescription:

1. For $n \in \mathbb{N}$, and $\omega = (\omega_1, \omega_2, \dots) \in \hat{\Omega}$, let $X_t^{(n)}(\omega) = X_t(\omega_n)$.
2. Set $B_0 = 0$ and

$$X_n(\omega) = \sum_{i=1}^{n} X_1^{(i)}(\omega), \text{ for } n \in \mathbb{N}.$$

For t of the form $t = n + s$, with $n \in \mathbb{N}_0$ and $s \in (0,1)$, we set

$$B_t(\omega) = X_n(\omega) + X_s^{(n+1)}(\omega).$$

Show that $\{B_t\}_{t\in[0,\infty)}$ is a Brownian motion.

The construction we start with is due to Paul Lévy (later modified and simplified by Zbigniew Ciesielski) and produces the paths of the Brownian motion as uniform limits of piecewise linear functions. Thanks to Problem 14.1, it will be enough to focus on the time-interval $[0,1]$.

The idea is to start with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a countable collection of iid unit normals is given, and construct a sequence of "better and better" approximations. For convenience, we organize these unit normals into a double-indexed array $\{\zeta_k^{(n)}\}_{n\in\mathbb{N}_0, k\in\mathbb{N}}$.

The zero-th approximation, $B^{(0)}$ is simply the constant function $B_t^{(0)} = 0$, for $t \in [0,1]$. We pick the first approximation $B^{(1)}$ so that its distribution at $t = 0$ and $t = 1$ matches that of the Brownian motion, and interpolate linearly in between. More precisely, we set

$$B_t^{(1)} = t\zeta_1^{(0)} \text{ i.e., } B_t^{(1)} = B_t^{(0)} + \Delta_t^{(0)} \text{ where } \Delta_t^{(0)} = t\zeta_1^{(0)}.$$

The next step is to add in the correct distribution at $t = 1/2$, without touching the end-points. To this end, we set

$$B_t^{(2)} = B_t^{(1)} + \Delta_t^{(1)},$$

and try to choose $\Delta_t^{(1)}$ in such a way to keep the distribution intact at $t = 0$ and $t = 1$ and to improve the match at $t = \frac{1}{2}$. For this, we clearly require $\Delta_0^{(1)} = \Delta_1^{(1)} = 0$. Moreover, since the Brownian motion is a centered Gaussian process, it looks like a good idea to take the distribution of $\Delta_{1/2}^{(1)}$ to be centered Gaussian, and, perhaps, independent of $\zeta_1^{(0)}$; let us try $\Delta_{1/2}^{(1)} = \sigma_1\zeta_1^{(1)}$. In that case, we have

$$\mathbb{E}\left[B_1^{(2)}B_{1/2}^{(2)}\right] = \mathbb{E}\left[\zeta_1^{(0)}(\tfrac{1}{2}\zeta_1^{(0)} + \sigma_1\zeta_1^{(0)})\right] = \tfrac{1}{2}$$

and

$$\mathbb{E}\left[(B_{1/2}^{(2)})^2\right] = \mathbb{E}\left[(\tfrac{1}{2}\zeta_1^{(0)} + \sigma_1\zeta_1^{(1)})^2\right] = \tfrac{1}{4} + \sigma_1^2.$$

Clearly, the choice $\sigma_1 = \frac{1}{2}$ makes the distribution of the random vector $(B_0^{(2)}, B_{1/2}^{(2)}, B_1^{(2)})$ centered Gaussian, with the Brownian covariance function on the time-set $\{0, \frac{1}{2}, 1\}$. Linear interpolation between these values ensures the continuity of the paths of the process $B^{(2)}$.

The construction of $B^{(3)}$ proceeds in the similar fashion; we set

$$B_t^{(3)} = B_t^{(2)} + \Delta_t^{(2)},$$

where $\Delta_0^{(2)} = \Delta_{1/2}^{(2)} = \Delta_1^{(2)} = 0$ and

$$\Delta_{1/4}^{(2)} = \sigma_2\zeta_1^{(2)} \text{ and } \Delta_{3/4}^{(2)} = \sigma_2\zeta_2^{(2)},$$

where the choice $\sigma_2 = 1/\sqrt{8} = 2^{-3/2}$ turns out to be the right one. A simple calculation, similar to the one in the case of $\Delta^{(1)}$, shows that, on $0, 1/4, 1/2, 3/4, 1$, the distribution of $B^{(3)}$ matches that of the Brownian motion. As always, to achieve continuity, we interpolate linearly between those values.

In general, having constructed $B^{(n)}$, we build the process $\Delta^{(n)}$ as in

$$\Delta_t^{(n)} = \begin{cases} 0, & t = (2k)2^{-n}, 0 \leq k \leq 2^{n-1}, \\ \sigma_n\zeta_k^{(n)}, & t = (2k-1)2^{-n}, 1 \leq k \leq 2^{n-1}, \end{cases}$$

with $\sigma_n = 2^{-(1+n)/2}$, and interpolate linearly between these values.

Proposition 14.19 (Uniform convergence of the approximations)**.** *There exists an event $\Omega^* \in \mathcal{F}$ such that $\mathbb{P}[\Omega^*] = 1$ and, for $\omega \in \Omega^*$, the sequence of functions $\{B^{(n)}_\cdot(\omega)\}_{n\in\mathbb{N}}$ is uniformly convergent.*

Proof. First, we estimate the size of the sup-norm of $\Delta^{(n)}$. By piecewise linearity of $B^{(n)}$, for $n \geq 1$, we have

$$\begin{aligned} p_n(\delta) := \mathbb{P}\left[\max_{t\in[0,1]} \left|\Delta^{(n)}_t\right| \geq \delta\right] &= \mathbb{P}\left[\max_{1\leq k\leq 2^{n-1}} \left|\Delta^{(n)}_{(2k-1)2^{-n}}\right| \geq \delta\right] \\ &= \mathbb{P}\left[\max_{1\leq k\leq 2^{n-1}} \left|\zeta^{(n)}_k\right| \geq 2^{n/2}\delta\right] \leq 2^{n-1}\mathbb{P}\left[\left|\zeta^{(1)}_1\right| \geq 2^{n/2}\delta\right]. \end{aligned}$$

We can now use the standard estimate

$$\begin{aligned} \mathbb{P}[\zeta^{(1)}_1 \geq c] &= \int_c^\infty \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \leq \int_c^\infty \tfrac{x}{c} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \\ &= \tfrac{1}{\sqrt{2\pi}} \tfrac{1}{c} e^{-\frac{1}{2}c^2}, c > 0, \end{aligned}$$

to conclude that $p_n(\delta) \leq 2^n \exp\Big(-\tfrac{1}{2}(2^{n/2}\delta)^2 - \log(2^{n/2}\delta)\Big)$. We pick $\delta_n = \sqrt{Cn}2^{-n/2}$, with $C > \log(2)$, so that

$$p_n(\delta_n) \leq e^{-\frac{1}{2}n(C-\log(2))-\frac{1}{2}\log(C)\log(n)} \leq n^{-\alpha}\beta^n$$

with $\alpha = \tfrac{1}{2}\log(C)$ and $\beta = e^{-\frac{1}{2}(C-\log(2))} < 1$. Consequently $\sum_n p_n(\delta_n) < \infty$ and Borel-Cantelli's lemma implies that

$$\mathbb{P}[\Omega^*] = 1 \text{ where } \Omega^* = \Big\{ \max_{t\in[0,1]} \left|\Delta^{(n)}_t\right| < \delta_n, \text{ ev.}\Big\}. \tag{14.4}$$

Summability of $\sum_n \delta_n$ now implies that the sequence $B^{(n)}(\omega)$ is Cauchy for the uniform norm (and therefore uniformly convergent), for each $\omega \in \Omega^*$. □

Having established the a.s. convergence of the sequence $B^{(n)}$, we give a name to its limit, i.e., for $t \in [0,1]$, we set

$$B_t(\omega) = \lim_n B^{(n)}_t(\omega), \text{ for } \omega \in \Omega^* \text{ and } B_t(\omega) = 0 \text{ for } \omega \in \Omega \setminus \Omega^*.$$

Proposition 14.20 (The limit is a Brownian motion)**.** *B is a Brownian motion on $[0,1]$.*

Proof. For $\omega \in \Omega^*$, $B_t(\omega)$ is the uniform limit of a sequence of continuous functions, hence, it is a continuous function itself.

To deal with the distributional properties, we note first that, by construction, the distribution of $B^{(n)}_t$ stabilizes for large enough n, as soon as t is a dyadic rational in $[0,1]$, i.e., if $2^m t \in \mathbb{N}_0$ for some m. Since $B_t = \lim_n B^{(n)}_t$, we conclude that the distribution of B_t is centered normal, with variance t. Similarly, each random vector $(B_{t_1}, \dots, B_{t_m})$, for dyadic $0 \leq t_1 < t_2 < \dots < t_m < \infty$, is centered Gaussian with the (Brownian) covariance $\Sigma_{ij} = \mathbb{E}[B_{t_i} B_{t_j}] = \min(t_i, t_j)$.

It remains to argue that the same is true even if not all $t_1, \dots, t_m$ are dyadic. For that, we recall that the paths of the process B are continuous, so for (not-necessarily dyadic) $0 \leq t_1 < \dots < t_m$ we have

$$(B_{t_1}, \dots, B_{t_m}) = \lim_k (B_{t_1^k}, \dots, B_{t_m^k}), \text{ a.s.}$$

where, for $i = 1, \dots, m$, $\{t_i^k\}_{k\in\mathbb{N}}$ is a sequence of dyadic rationals which converges to t_i. The distribution of $(B_{t_1^k}, \dots, B_{t_m^k})$ is centered multivariate normal with the covariance matrix $\Sigma_{ij}^k = \min(t_i^k, t_j^k)$, and, so, its limit as $k \to \infty$ is centered normal with covariance $\Sigma_{ij} = \min(t_i, t_j)$. Finally, by the a.s.-convergence established above, we conclude that $(B_{t_1}, \dots, B_{t_m})$ is centered normal with variance-covariance matrix $\Sigma_{ij} = \min(t_i, t_j)$. □

The "Kolmogorov-Čentsov" construction

Unlike Lévy's construction which builds the Brownian motion from a sequence of better and better approximations, the Kolmogorov-Čentsov construction proceeds in two big steps. The first one produces a measure on a (very large) probability space, and the second one finely modifies the coordinate process so as to achieve continuity. Unlike in the previous construction, we only state the (two) fundamental results, and skip the proofs.

We start with a well-known theorem of Kolmogorov (its proof is based on Caratheodory's construction in measure theory and we skip it in these notes). Before we state it, we fix some notation and review some concepts on product spaces (introduced in Chapters **??** and **??**).

Let $\mathcal{T}$ be a nonempty set; for a family $\{\Omega_t\}_{t\in\mathcal{T}}$ of sets, their product $\Omega = \prod_{t\in\mathcal{T}} \Omega_t$, is defined as the set of all functions $\omega : \mathcal{T} \to \cup_{t\in\mathcal{T}} \Omega_t$, such that $\omega(t) \in \Omega_t$. For $t \in \mathcal{T}$, the map $x_t : \Omega \to \Omega_t$ given by $x_t(\omega) = \omega(t)$, for $\omega \in \Omega$ is called the **coordinate map**. For a family $\{(\Omega_t, \mathcal{F}_t)\}_{t\in\mathcal{T}}$ of measurable spaces, their **product** is the measurable space $(\prod_{t\in\mathcal{T}} \Omega_t, \prod_{t\in\mathcal{T}} \mathcal{F}_t)$, where $\prod_{t\in\mathcal{T}} \mathcal{F}_t = \sigma(x_t, t \in \mathcal{T})$ is the smallest σ-algebra on $\prod_{t\in\mathcal{T}} \Omega_t$ under which all coordinate mappings are measurable.

Theorem 14.21 (The Kolmogorov Extension Theorem)**.** *Let $\mathcal{T}$ be a nonempty set, and let*

$$\{\mu_{(t_1,t_2,\dots,t_n)} : n \in \mathbb{N}, t_1, \dots, t_n \in \mathcal{T}\} \tag{14.5}$$

be a family of probability measures on $\mathbb{R}^n$, $n \in \mathbb{N}$ with the property that

1. *for all* $n \in \mathbb{N}$, $(t_1, t_2, \dots, t_n) \in \mathcal{T}^n$, *a permutation* $\sigma \in S_n$ *and a family* $A_i \in \mathcal{B}(\mathbb{R})$, $i = 1, \dots, n$,

$$\mu_{(t_1,t_2,\dots,t_n)}(A_1 \times A_2 \times \dots \times A_n) = \\ = \mu_{(t_{\sigma(1)},t_{\sigma(2)},\dots,t_{\sigma(n)})}(A_{\sigma(1)} \times A_{\sigma(2)} \times \dots A_{\sigma(n)}),$$

2. *for all* $n \in \mathbb{N}$, $A \in \mathcal{B}(\mathbb{R}^n)$, *and* $(t_1, t_2, \dots, t_n, t_{n+1}) \in \mathcal{T}^{n+1}$,

$$\mu_{(t_1,t_2,\dots,t_n,t_{n+1})}(A \times \mathbb{R}) = \mu_{(t_1,t_2,\dots,t_n)}(A).$$

Then, there exists a probability measure $\mathbb{P}$ *on the product measurable space* $(\prod_{t\in\mathcal{T}} \mathbb{R}, \prod_{t\in\mathcal{T}} \mathcal{B}(\mathbb{R}))$ *such that that the finite-dimensional distributions of the coordinate process* $\{x_t\}_{t\in\mathcal{T}}$ *are given by* (14.5), *i.e., the distribution of the random vector* $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ *is* $\mu_{(t_1,\dots,t_n)}$, *for all* $n \in \mathbb{N}$, $(t_1, t_2, \dots, t_n) \in \mathcal{T}^n$.

Remark 14.22.

1. In spite of their formal appearance, the conditions of Theorem 14.21 are very natural. They simply state that the measures $\mu_{(t_1,\dots,t_n)}$ satisfy the immediate necessary conditions that a family of finite-dimensional distributions of a process with index set $\mathcal{T}$ would satisfy.

2. The meaning of the statement of the theorem is not hard to decipher, either. It basically says that if you define a family of finite-dimensional distributions (in a minimally consistent way), there will exist a probability space and a random process on it whose finite-dimensional distributions are exactly the ones you prescribed.

Corollary 14.23 (A "Brownian-motion-in-distribution" exists)**.** *There exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *supporting a stochastic process* $\{X_t\}_{t\in[0,\infty)}$ *which satisfies the conditions (1) and (2) of Definition 14.17.*

Proof. For $0 \le t_1 < t_2 < \dots < t_n < \infty$, let $\mu_{(t_1,\dots,t_n)}$ be the centered Gaussian measure on $\mathbb{R}^n$ with covariance $\Sigma_{ij} = \min(t_j, t_j)$. One routinely checks that the consistency conditions (1) and (2) from Theorem 14.21 are satisfied. □

The second ingredient in this construction is a theorem of Kolmogorov and Čentsov which allows us to produce a pathwise continuous modification of a process which possesses strong continuity properties in distribution. Here is what we mean by the word "modification":

Definition 14.24 (Modification and indistinguishability)**.** A stochastic process $\{X'_t\}_{t\in[0,\infty)}$ is said to be a **modification** of the process $\{X_t\}_{t\in[0,\infty)}$ if X' and X are defined on the same probability space and

$$X_t = X'_t, \text{ a.s., for all } t \geq 0.$$

The processes X and X' are said to be **indistinguishable** if there exists an event $\Omega^* \in \mathcal{F}$ with $\mathbb{P}[\Omega^*] = 1$ such that

$$X_t(\omega) = X'_t(\omega) \text{ for all } t \geq 0 \text{ and all } \omega \in \Omega^*.$$

Remark 14.25.

1. If X and X' are indistinguishable then they are modifications of each other. The converse does not hold. Take, for example, the processes X and X' defined on $([0,1], \mathcal{B}([0,1]), \lambda)$ by
$$X_t(\omega) = 0, \quad X'_t(\omega) = \mathbf{1}_{\{t=\omega\}}, \text{ for } (t,\omega) \in [0,1] \times [0,1].$$
2. Without further assumptions on the processes X and X', one cannot define indistinguishability by requiring that $\mathbb{P}[X_t = X'_t \text{ for all } t] = 1$. Indeed, there is no guarantee that the set $\{X_t = X'_t \text{ for all } t\}$ is measurable, as it is an intersection of uncountably many events of the form $\{X_t = X'_t\}$, $t \geq 0$.

We remind the reader of a particular form of the ubiquitous notion of Hölder continuity.

Definition 14.26 (Local Hölder continuity)**.** A real function $f : [0,\infty) \to \mathbb{R}$ is said to be **locally Hölder continuous** with exponent $\gamma \in (0,1]$ if there exist constants $\delta, K > 0$ such that

$$|f(t) - f(s)| \leq K\,|t-s|^\gamma \text{ for all } t,s \text{ with } |t-s| < \delta.$$

The following theorem turns out to be very useful beyond its role in the construction of the Brownian motion below. We give it without proof.

Theorem 14.27 (Kolmogorov-Čentsov)**.** *Suppose that there exists constants $\alpha, \beta, C > 0$ such that the stochastic process $\{X_t\}_{t\in[0,1]}$ satisfies*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq C\,|t-s|^{1+\beta}, \tag{14.6}$$

for all $t, s \in [0,1]$. Then, there exists a modification $\{\tilde{X}_t\}_{t\in[0,1]}$ of $\{X_t\}_{t\in[0,1]}$ and an event $\Omega^* \in \mathcal{F}$ such that $\mathbb{P}[\Omega^*] = 1$ and for every $\omega \in \Omega^*$ and every $0 < \gamma < \frac{\beta}{\alpha}$, the trajectory $t \mapsto \tilde{X}_t(\omega)$ is locally Hölder continuous with exponent γ.

Remark 14.28. Theorem 14.27 above takes one type of Hölder continuity and turns it into another. Indeed, the condition (14.6) requires that the map $t \mapsto X_t$ from $[0,1]$ into $\mathbb{L}^\alpha$ be Hölder continuous with order $\gamma' = (1+\beta)/\alpha$. The conclusion is that its paths can be made Hölder continuous with any exponent smaller than $\gamma' - 1/\alpha$.

Corollary 14.29 (Brownian existence and regularity of pathe via Kolmogorov-Čentsov)**.** *The Brownian motion exists. Moreover, almost all of its trajectories are locally Hölder continuous of order γ, for each $\gamma < \frac{1}{2}$.*

Proof. We start from the process $\{X_t\}_{t\in[0,1]}$, whose existence is guaranteed by Corollary 14.23. Since both $X_t - X_s$ and $\sqrt{(t-s)}X_1$ have the same distribution when $1 \geq t > s \geq 0$ - namely $N(0, t-s)$ - for each $n = 2, 3, \dots$ we have

$$\mathbb{E}[|X_t - X_s|^{2n}] = |t-s|^n \, \mathbb{E}[|X_1|^{2n}] = C_n \, |t-s|^n \, ,$$

where $C_n = \mathbb{E}[|X_1|^{2n}] = \frac{(2n)!}{2^n n!} < \infty$. Therefore, Theorem 14.27 applies and we can produce a continuous modification $\{\tilde{X}_t\}_{t\in[0,1]}$ of $\{X\}_{t\in[0,1]}$ - a Brownian motion on [0,1]. The same theorem implies that the trajectories of $\{\tilde{X}_t\}_{t\in[0,1]}$ are a.s. locally Hölder continuous of order $\gamma < \frac{n-1}{2n} = \frac{1}{2} - \frac{1}{2n}$, for each $n = 2, 3, \dots$. □

The "Donsker-Prohorov" construction

The canonical space $C[0,\infty)$. The set of all *continuous* functions $\omega : [0,\infty) \to \mathbb{R}$ is denoted by $C[0,\infty)$. It is typically endowed with topology of locally uniform convergence, i.e., the topology which is induced by the metric d, given by

$$d(\omega_1, \omega_2) = \sum_{k=1}^{\infty} 2^{-k} \min(1, \sup_{t\in[0,k]} |\omega_1(t) - \omega_2(t)|), \text{ for } \omega_1, \omega_2 \in C[0,\infty).$$

A sequence $\{\omega_n\}_{n\in\mathbb{N}} \subseteq C[0,\infty)$ converges to $\omega \in C[0,\infty)$ if and only if the restrictions $\omega_n|_{[0,k]}$ converge to $\omega|_{[0,k]}$ uniformly, for each $k \in \mathbb{N}$.

When we speak of the Borel sets on $C[0,\infty)$, we refer to the σ-algebra $\mathcal{B}(C[0,\infty))$ generated by the open sets in the metric d. The functions $x_t : C[0,\infty) \to \mathbb{R}$, $t \in [0,\infty)$, given by $x_t(\omega) = \omega(t)$, for $\omega \in C[0,\infty)$ and $t \geq 0$, are called the coordinate mappings on $C[0,\infty)$. The subsets of $C[0,\infty)$ of the form

$$\{\omega \in C[0,\infty) : (\omega(t_1), \dots, \omega(t_n)) \in A\}, \tag{14.7}$$

for some $n \in \mathbb{N}$, $(t_1, \dots, t_n) \in [0,\infty)^n$, and $A \in \mathcal{B}(\mathbb{R}^n)$, are called **finite-dimensional cylinders**. The collection of all finite-dimensional cylinders is denoted by $\mathcal{C}$.

Problem 14.2. Show that

1. d is a metric, and that $(C[0,\infty), d)$ is a separable and complete metric space.

2. The σ-algebra $\mathcal{B}(C[0,\infty))$ is generated by all finite-dimensional cylinders.

The set $C[0,\infty)$ is usually referred to as the **canonical space** for the Brownian motion, because of the following result:

Proposition 14.30 (The distribution of a continuous process)**.** *Suppose that almost all of trajectories of the stochastic process $\{X_t\}_{t\in[0,\infty)}$ are continuous. There exists a probability measure $\mathbb{P}_X$ on the Borel sets of $C[0,\infty)$ such that the coordinate process $\{x_t\}_{t\in[0,\infty)}$ under $\mathbb{P}_X$ has the same finite-dimensional distributions as $\{X_t\}_{t\in[0,\infty)}$.*

Proof. Since we are trying to produce the process which will match $\{X_t\}_{t\in[0,\infty)}$ only in finite-dimensional distributions, we can assume that *all* of its trajectories are continuous, without loss of generality. Therefore, we can interpret $\{X_t\}_{t\in[0,\infty)}$ as a mapping from $(\Omega,\mathcal{F})$ to $(C[0,\infty),\mathcal{B}(C[0,\infty)))$ and define the measure $\mathbb{P}_X$ by

$$\mathbb{P}_X[B] = \mathbb{P}[X \in B] \Big(= \mathbb{P}[X^{-1}(B)] = \mathbb{P}[\{\omega\in\Omega : X_\cdot(\omega)\in B\}]\Big),$$

provided that we show that $X^{-1}(B)\in\mathcal{F}$, i.e., that the mapping $X : (\Omega,\mathcal{F}) \to (C[0,\infty),\mathcal{B}(C[0,\infty)))$ is measurable. For that, we pick B of the form (14.7), and note that $X^{-1}(B) = \{(X_{t_1},\dots,X_{t_n})\in A\}$ which is an element of $\mathcal{F}$ by the assumption that X is a stochastic process. Since the sets of the form (14.7) generate $\mathcal{B}(C[0,\infty))$, we are done. □

Definition 14.31 (The law of a process; the Wiener measure)**.** For a continuous process X, the measure $\mathbb{P}_X$ as in Proposition 14.30 is called the **law** or the **distribution** of X. When X is a Brownian motion, then P_X is called the **Wiener measure**.

Weak convergence on $C[0,\infty)$ Once we understand how to identify continuous processes and probability measures on $C[0,\infty)$, we can begin to study convergence of processes in terms of the convergence of their laws. A review of some of the basic notions of weak convergence (covered in Lecture 7), specialized to the space $C[0,\infty)$, follows.

Definition 14.32 (Weak convergence)**.** Let $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $(C[0,\infty),\mathcal{B}(C[0,\infty)))$. We say that $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ **converges weakly** to a probability measure $\mathbb{P}$ on $(C[0,\infty),\mathcal{B}(C[0,\infty)))$, and denote it by $\mathbb{P}_n \stackrel{w}{\to} \mathbb{P}$, if

$$\mathbb{E}^{\mathbb{P}_n}[f(\omega.)] \to \mathbb{E}^{\mathbb{P}}[f(\omega.)],$$

i.e.,

$$\int_{C[0,\infty)} f(\omega)\, d\mathbb{P}_n(\omega) \to \int_{C[0,\infty)} f(\omega)\, d\mathbb{P}(\omega),$$

for all bounded and continuous functions $f : C[0,\infty) \to \mathbb{R}$.

The following theorem characterizes weak compactness in the set of all probability measures on $C[0,\infty)$. We give it without proof.

Theorem 14.33 (Prohorov)**.** *Let Π be a collection of probability measures on $C[0,\infty)$. Then, the following two conditions are equivalent*

1. *For each sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ in Π there exist a subsequence $\{\mathbb{P}_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ and a probability measure $\mathbb{P}$ on $C[0,\infty)$ such that $\mathbb{P}_{n_k} \stackrel{w}{\to} \mathbb{P}$.*
2. *The sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ is **tight**, i.e., for each $\varepsilon > 0$, there exists a compact set $K \subseteq C[0,\infty)$ such that $\sup_n \mathbb{P}_n[K^c] \leq \varepsilon$.*

Our next definition describes a mode of convergence weaker than the weak convergence introduced above and much easier to work with.

Definition 14.34 (Convergence of finite-dimensional distribution)**.** For each $m \in \mathbb{N}$ and $(t_1, \dots, t_m) \in [0,\infty)^m$, the mapping

$$C[0,\infty) \ni \omega \mapsto (\omega_{t_1}, \dots, \omega_{t_m}) \in \mathbb{R}^m,$$

is called the **natural projection**, and is denoted by $\pi_{(t_1,\dots,t_m)}$. For a probability measure $\mathbb{P}$ on $C[0,\infty)$, the probability measure $\mathbb{P}_{(t_1,\dots,t_m)}$, given by

$$\mathbb{P}_{(t_1,\dots,t_m)}[A] = \mathbb{P}[\pi^{-1}_{(t_1,\dots,t_m)}(A)], \text{ for } A \in \mathcal{B}(\mathbb{R}^m),$$

is called the $(t_1, \dots, t_m)$-**finite-dimensional distribution** of $\mathbb{P}$.

A sequence $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ of probability measures on $C[0,\infty)$ is said to converge to a probability measure $\mathbb{P}$ in the sense of **convergence of finite-dimensional distributions**, denoted by $\mathbb{P}^n \stackrel{f.d.d.}{\to} \mathbb{P}$ if $\mathbb{P}^n_{(t_1,\dots,t_m)} \stackrel{w}{\to} \mathbb{P}_{(t_1,\dots,t_m)}$, for all $m \in \mathbb{N}$ and all $(t_1, \dots, t_m) \in [0,\infty)^m$.

Remark 14.35*.* The reader will realize immediately that $\mathbb{P}_n \stackrel{f.d.d.}{\to} \mathbb{P}$ if and only if the distributions of random vectors $(\omega_{t_1}, \dots, \omega_{t_m})$, for all $m \in \mathbb{N}$, $(t_1, \dots, t_m) \in [0,\infty)^m$, converge weakly on $\mathbb{R}^m$. Equivalently (and, perhaps, more clearly), suppose that $\mathbb{P}_n$ is the law of process $\{X^n_t\}_{t\in[0,\infty)}$ and $\mathbb{P}$ is the law of $\{X_t\}_{t\in[0,\infty)}$. Then $\mathbb{P}_n \stackrel{f.d.d.}{\to} \mathbb{P}$ if and only if $(X^n_{t_1}, \dots, X^n_{t_m}) \stackrel{\mathcal{D}}{\to} (X_{t_1}, \dots, X_{t_m})$ for all $m \in \mathbb{N}$ and all m-tuples $(t_1, \dots, t_m)$ of non-negative reals.

Problem 14.3. Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability measures on $C[0,\infty)$ whose finite-dimensional distributions coincide, i.e., such that

$$\mathbb{P}_{(t_1,\dots,t_m)} = \mathbb{P}'_{(t_1,\dots,t_m)},$$

for all $m \in \mathbb{N}$, $(t_1,\dots,t_m) \in [0,\infty)^m$. Then $\mathbb{P} = \mathbb{P}'$.

It is not hard to see that $\mathbb{P}_n \stackrel{w}{\to} \mathbb{P}$ implies $\mathbb{P}_n \stackrel{f.d.d.}{\to} \mathbb{P}$, but the converse is only true if an additional property is satisfied:

Proposition 14.36 (Tightness + convergence in fdd = weak convergence)**.** *Let $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ be a tight sequence of probability measures on $C[0,\infty)$ such that $\mathbb{P}_n \stackrel{f.d.d.}{\to} \mathbb{P}$, for some probability measure $\mathbb{P}$ on $C[0,\infty)$. Then $\mathbb{P}_n \stackrel{w}{\to} \mathbb{P}$.*

Proof. Tightness of $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ and Theorem 14.33 imply that for some probability measure $\mathbb{P}'$ on $C[0,\infty)$ and some subsequence $\{\mathbb{P}_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ we have $\mathbb{P}_{n_k} \stackrel{w}{\to} \mathbb{P}'$. In particular, we have $\mathbb{P}_{n_k} \stackrel{f.d.d.}{\to} \mathbb{P}'$. On the other hand, $\mathbb{P}_n \stackrel{f.d.d.}{\to} \mathbb{P}$ and the same is true for all of its subsequences, so $\mathbb{P}$ and $\mathbb{P}'$ have the same finite-dimensional distributions. By Problem 14.3, $\mathbb{P} = \mathbb{P}'$. The same argument shows that every weakly convergent subsequence of $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ must have $\mathbb{P}$ as the limit.

Suppose that $\mathbb{P}_n \stackrel{w}{\not\to} \mathbb{P}$. Then, there exists a bounded continuous function $f : C[0,\infty) \to \mathbb{R}$ such that

$$\int_{C[0,\infty)} f(\omega)\, d\mathbb{P}_n(\omega) \not\to \int_{C[0,\infty)} f(\omega)\, d\mathbb{P}(\omega).$$

By passing to a subsequence (indexed by $\{n_k\}_{k\in\mathbb{N}}$) we can assume that the limit $\int_{C[0,\infty)} f(\omega)\, d\mathbb{P}_{n_k}(\omega)$ exists (remember, f is bounded) but is not equal to $\int_{C[0,\infty)} f(\omega)\, d\mathbb{P}(\omega)$. If we pass to a further subsequence - whose index set is denoted by $\{n_{k_l}\}_{l\in\mathbb{N}}$ - we can assume that $\mathbb{P}_{n_{k_l}}$ converges weakly, and, as we have shown above, its limit must be equal to $\mathbb{P}$. This is in contradiction with the fact that $\int_{C[0,\infty)} f(\omega)\, d\mathbb{P}_{n_{k_l}}(\omega) \not\to \int_{C[0,\infty)} f(\omega)\, d\mathbb{P}(\omega)$. Therefore, $\mathbb{P}_n \stackrel{w}{\to} \mathbb{P}$. $\square$

Problem 14.4. Give an explicit construction of a sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ of probability measures on $C[0,\infty)$ which converges in the sense of finite-dimensional distributions, but does not converge weakly.

Hint: Use Dirac measures.

Interpolated random walks. Our second construction of Brownian motion is due to Donsker, and starts with a sequence of random walks.

Definition 14.37 (Interpolated random walk)**.** Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of iid random variables with $\mathbb{E}[\xi_n] = 0$, $\text{Var}[\xi_n] = 1$, for all $n \in \mathbb{N}$. The **interpolated random walk with steps** $\{\xi_n\}_{n\in\mathbb{N}}$ is a stochastic process $\{X_t\}_{t\in[0,\infty)}$ defined by

$$X_t = \begin{cases} 0, & t = 0, \\ \sum_{k=1}^{t} \xi_k, & t \in \mathbb{N}, \\ \text{interpolated linearly}, & t \notin \mathbb{N}_0. \end{cases}$$

For $n \in \mathbb{N}$, the n**-scaled interpolated random walk with steps** $\{\xi_n\}_{n\in\mathbb{N}}$ is the stochastic process $\{X^n_t\}_{t\in[0,\infty)}$, given by

$$X^n_t = \tfrac{1}{\sqrt{n}} X_{nt}.$$

Theorem 14.38 (Donsker's invariance principle)**.** *Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of iid random variables with $\mathbb{E}[\xi_n] = 0$, $\mathrm{Var}[\xi_n] = 1$, for all $n \in \mathbb{N}$, and let $\{X^n\}_{n\in\mathbb{N}}$ be the sequence of corresponding scaled interpolated random walks. If $\mathbb{P}_n$ denotes the law of X^n on $C[0,\infty)$, then the sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ is weakly convergent and its limit is the Wiener measure.*

Proof. We only give the skeleton of the proof:

1. The first step is to show that the sequence $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ is tight. This is accomplished through some clever estimates and the use of Arzelá-Ascoli theorem.

2. Then, we need to show that the finite-dimensional distributions converge towards the finite-dimensional distributions of the Brownian motion. The main tool here is the Central Limit Theorem, and the procedure is simple, if a little tedious (you need to deal with the interpolated parts of X^n).

3. Finally, we conclude that the weak limit exists in $C[0,\infty)$ and that its finite-dimensional distributions are those of a Brownian motion. Therefore, the limit is the Wiener measure.

□

Additional Problems

Problem 14.5 (A characterization of the Gaussian)**.** Let $\mathcal{C}$ denote the subset of probability densities f on $\mathbb{R}$ with

$$\int x f(x)\, dx = 0 \text{ and } \int x^2 f(x)\, dx = 1.$$

Show that the standard Gaussian density φ maximizes the **entropy**

$$H(f) = -\int f(x)\, \log f(x)\, dx,$$

over all $f \in \mathcal{C}$.

Note: We set $0 \log 0 := 0$.

Problem 14.6 (Gaussian measures of balls)**.** Let B_n be the unit ball in $\mathbb{R}^n$, and let γ_n be the centered Gaussian measure on $\mathbb{R}^n$ whose variance-covariance matrix is the identity matrix I.

1. Show that
$$\lim_{n\to\infty} n\Gamma(\tfrac{n}{2})\gamma_n(B_n) = \tfrac{1}{2e},$$
where Γ is the Gamma function.

2. Prove that $\gamma_n(B_n) \to 0$ faster than any exponential.

3. Describe the asymptotic behavior of $\gamma_n(B_n)/\lambda_n(B_n)$, as $n \to \infty$, where λ_n denotes the n-dimensional Lebesgue measure.

Problem 14.7 (Brownian motion as a random expansion)**.** Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal set in $\mathbb{L}^2[0,1]$, i.e. $\int_0^1 e_n^2(u)\,du = 1$, as well as $\int_0^1 e_n(u)e_m(u)\,du = 0$, for $m \neq n \in \mathbb{N}$. If $\{\zeta_n\}_{n\in\mathbb{N}}$ is a sequence of iid unit normals, show that

1. The sequence $B_n = \sum_{k=1}^n \zeta_k e_k$, $n \in \mathbb{N}$, diverges in $\mathbb{L}^2[0,1]$, a.s.

2. On the other hand, the sequence $\langle B_n, h\rangle = \sum_{k=1}^n \zeta_k \int_0^1 e_k(u)h(u)\,du$, $n \in \mathbb{N}$, converges in $\mathbb{L}^2$ for each $h \in \mathbb{L}^2[0,1]$.

Hint: Show that $\sum_{n\in\mathbb{N}} \big(\int_0^1 e_n(u)h(u)\,du\big)^2 < \infty$.

3. With the random variable B_h defined as $B_h = \lim_n \langle B_n, h\rangle$, show that the stochastic process $(B_h)_{h\in\mathbb{L}^2[0,1]}$ is Gaussian and find its mean and covariance functions.

Note: This is an example of a stochastic process indexed by a something other than a subset of $\mathbb{R}$. Here, in fact, it is indexed by an infinite-dimensional Hilbert space.

4. Suppose, for this part only, that, for each pair $h, k \in \mathbb{L}^2[0,1]$ we have
$$\sum_n \Big(\int_0^1 e_n(u)h(u)\,du\Big)\Big(\int_0^1 e_n(u)k(u)\,du\Big) = \int_0^1 h(u)k(u)\,du. \quad (14.8)$$

Note: This will be the case if and only if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal *basis* of $\mathbb{L}^2[0,1]$.

Consider the stochastic process $\{W_t\}_{t\in[0,1]}$, where $W_t = B_{h_t}$, for $h_t = \mathbf{1}_{[0,t]}$, $t \in [0,1]$. Show that W is a "Brownian-motion-in-distribution", i.e., that it is centered Gaussian with the Brownian covariance structure.

Problem 14.8 (A process with no continuous modification)**.** Let η be a random variable with $\mathbb{P}[\eta > 0] = 1$, and let $\{X_t\}_{t\in[0,\infty)}$ be defined as follows:
$$X_t = \begin{cases} 0, & t < \eta, \\ 1, & \text{otherwise.} \end{cases}$$
Show that X is *not* a modification of any process with continuous paths.

Problem 14.9 (The Integrated Brownian Motion)**.** Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, with *all* trajectories $t \mapsto B_t(\omega)$, $\omega \in \Omega$, continuous, and define the process $\{X_t\}_{t\in[0,\infty)}$ by

$$X_t = \int_0^t B_u\,du \text{ for } t \in [0,\infty).$$

Show that X_t is a random variable for each t. Then, show that X is a Gaussian process and compute its mean and covariance functions.
On the other hand, let $Y_t = \int_0^t B_u^3\,du$. Compute $\mathrm{Var}[Y_t]$. (*) Is the process Y Gaussian?

Hint: Approximate the integrals that define X_t and Y_t and feel free to use software for integration.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 15
First Properties of the Brownian Motion

This lecture deals with some of the more immediate properties of the Brownian motion and its trajectories. Many other properties which require various tools from stochastic analysis will be scattered throughout the remainder of the notes.

Proposition 15.1 (Symmetries of the Brownian motion)**.** *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion. Then the following processes are also Brownian motions (only on $[0,1]$ in (5)):*

1. $\{-B_t\}_{t\in[0,\infty)}$, (reflection),
2. $\{\frac{1}{\sqrt{\alpha}}B_{\alpha t}\}_{t\in[0,\infty)}$, *for* $\alpha > 0$, (scaling),
3. $\{B_{t_0+t} - B_{t_0}\}_{t\in[0,\infty)}$, *for all* $t_0 \geq 0$, (shifting),
4. $\{X_t\}_{t\in[0,\infty)}$, *where* $X_0 = 0$ *and* $X_t = tB_{1/t}$, *for* $t > 0$, (inversion),
5. $\{B_1 - B_{1-t}\}_{t\in[0,1]}$, (time reversal).

Proof. It is easy to see that all of the above are centered Gaussian processes with the Brownian covariance strucure. Continuity of the paths is clear in 2., 3. and 5., and everywhere except at $t = 0$ in 4. To deal with that case, we may use the Kolmogorov-Čentsov theorem (how?), or use our next result. □

Proposition 15.2 (The Law of Large Numbers for Brownian Motion)**.** *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion. Then*

$$\lim_{t\to\infty} \tfrac{B_t}{t} = 0, \text{ a.s.}$$

Proof. Using the (discrete-time) Law of Large numbers, we have

$$\lim_{n\to\infty} \tfrac{B_n}{n} = \lim_n \tfrac{1}{n} \sum_{k=1}^n (B_k - B_{k-1}) = 0, \text{ a.s.}$$

The idea is to show that the trajectory of B cannot deviate too much from B_n on $[n, n+1]$. Indeed, if we can prove that

$$\sum_{n=0}^{\infty} \mathbb{P}\left[\sup_{t\in[n,n+1]} |B_t - B_n| \geq n^{2/3}\right] < \infty, \tag{15.1}$$

the Borel-Cantelli theorem will finish the job for us (why?). Therefore, all we need to do is estimate the probability in (15.1). We use Kolmogorov's inequality (a specialization of the maximal inequality for submartingales) applied to the discrete random walk $(B_{n+k2^{-m}} - B_n)_{k\in\mathbb{N}_0}$ to conclude that

$$\mathbb{P}[\sup_{0<k\leq 2^m} |B_{n+k2^{-m}} - B_n| \geq n^{2/3}] \leq \tfrac{1}{n^{4/3}}\mathbb{E}[(B_{n+1}-B_n)^2] = \tfrac{1}{n^{4/3}}.$$

However, by the continuity of trajectories of B, we have

$$\Big\{ \sup_{t\in[n,n+1]} |B_t - B_n| \geq n^{2/3} \Big\} = \bigcup_{m\in\mathbb{N}} \Big\{ \sup_{0<k\leq 2^m} |B_{n+k2^{-m}} - B_n| \geq n^{2/3} \Big\},$$

and the union is increasing. Therefore,

$$\mathbb{P}\left[\sup_{t\in[n,n+1]} |B_t - B_n| \geq n^{2/3} \right] \leq \tfrac{1}{n^{4/3}},$$

and (15.1) follows. □

Proposition 15.3 (Long-term behavior of trajectories)**.** *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion. Then,*

$$\limsup_{t\to\infty} \tfrac{B_t}{\sqrt{t}} = \infty, \textit{ and } \liminf_{t\to\infty} \tfrac{B_t}{\sqrt{t}} = -\infty, \textit{ a.s.}$$

Proof. For $K \in (0,\infty)$, we have

$$\begin{aligned} \mathbb{P}[B_n > K\sqrt{n} \text{ i.o.}] &= \mathbb{P}[\cap_{n\in\mathbb{N}} \cup_{m\geq n} \{B_m > K\sqrt{m}\}] \\ &= \lim_{n\to\infty} \mathbb{P}[\cup_{m\geq n}\{B_m > K\sqrt{m}\}] \\ &\geq \lim_{n\to\infty} \mathbb{P}[B_n > K\sqrt{n}] = \mathbb{P}[B_1 > K] > 0. \end{aligned} \tag{15.2}$$

We can view B_n as a sum $B_n = \sum_{k=1}^n \xi_k$, where $\xi_k = B_k - B_{k-1}$ of independent random variables and note that for any $n_0 \in \mathbb{N}$,

$$\limsup_n B_n/\sqrt{n} > K \text{ if and only if } \limsup_n (B_n - B_{n_0})/\sqrt{n} > K.$$

By Kolmogorov's 0-1 law[1] we have

$$\mathbb{P}[B_n > K\sqrt{n} \text{ i.o.}] = 1, \text{ so that } \limsup \tfrac{B_n}{\sqrt{n}} \geq K, \text{ a.s., for all } K > 0.$$

Therefore,

$$\limsup_{t\to\infty} \tfrac{B_t}{\sqrt{t}} \geq \limsup \tfrac{B_n}{\sqrt{n}} = +\infty, \text{ a.s.,}$$

The statement about the lim inf follows from the fact that $\{-B_t\}_{t\in[0,\infty)}$ is also a Brownian motion. □

A nice corollary of the above result is that the Brownian motion visits each point $a \in \mathbb{R}$ in each time-interval $[t,\infty)$, $t \geq 0$.

[1] that is the one that says that the σ-algebra $\cap_{n\in\mathbb{N}}\sigma(\xi_k; k\geq n)$ is trivial when $\{\xi_n\}_{n\in\mathbb{N}}$ are independent.

Corollary 15.4 (Recurrence of the Brownian motion). *The Brownian motion is recurrent, i.e., for each $a \in \mathbb{R}$,*

$$\text{the (random) set } \mathcal{L}^a(\omega) = \{t \in [0,\infty) : B_t(\omega) = a\} \text{ is unbounded, a.s.}$$

The quadratic variation of the Brownian motion

We start by introducing some space-saving notation related to partitions. Given $t > 0$, a sequence

$$0 = t_0 < t_1 < \cdots < t_k = t$$

is called a **partition of** $[0,t]$ and the set of all partitions of $[0,t]$ is denoted by $P_{[0,t]}$. The elements $t_0, t_1, \dots$ of a partition are referred to as its **nodes**. For $\Delta = \{t_0, \dots, t_k\} \in P_{[0,t]}$, the **mesh** $|\Delta|_{[0,t]}$ of Δ is defined by

$$|\Delta|_{[0,t]} = \sup_{i \in \{1,\dots,k\}} |t_i - t_{i-1}| .$$

A sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P_{[0,t]}$ is said to **converge to identity**, denoted by $\Delta_n \to \mathrm{Id}$, if $|\Delta_n|_{[0,t]} \to 0$, for each $t \geq 0$. Additionally, we say that the convergence is **fast**, denoted by $\Delta_n \overset{sum}{\to} \mathrm{Id}$, if $\sum_n |\Delta_n|_{[0,t]} < \infty$.

For $\Delta = \{t_0, \dots, t_k\} \in P_{[0,t]}$, a real function $f : [0,t] \to \mathbb{R}$ (or, more generally, $f : [0,\infty) \to \mathbb{R}$) and $p \geq 1$, we define the **p-variation** $\mathrm{Var}_p(f;\Delta)$ **of f along** Δ by

$$\mathrm{Var}_p(f;\Delta) = |f(0)|^p + \textstyle\sum_{i=1}^k |f(t_i) - f(t_{i-1})|^p .$$

The **total p-variation** $\mathrm{Var}_p(f;[0,t])$ **of f is given by**

$$\mathrm{Var}_p(f;[0,t]) = \sup_{\Delta \in P_{[0,t]}} \mathrm{Var}_p(f;\Delta),$$

and the function f is said to be **of finite p-variation** if $\mathrm{Var}_p(f;[0,t]) < \infty$, for all $t \geq 0$. When $p = 1$ we simplify the notation by writing Var for Var_1 and refer to the functions of finite 1-variation simply as **functions of finite variation** or **rectifiable functions**.

For a stochastic process $\{X\}_{t\in[0,\infty)}$, and $\Delta \in P_{[0,t]}$, the expression $\mathrm{Var}_p(X;\Delta)$ refers to the random variable, whose value on $\omega \in \Omega$ is $\mathrm{Var}_p(X_\cdot(\omega);\Delta)$. A similar interpretation can be applied for the total variation $\mathrm{Var}_p(X;[0,t])$.

Proposition 15.5 (The quadratic variation of the Brownian Motion). *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and, for $t \geq 0$, let $\{\Delta_n\}_{n\in\mathbb{N}}$ be a sequence in $P_{[0,t]}$ with $|\Delta_n|_{[0,t]} \to 0$. Then*

$$\lim_n \mathrm{Var}_2(B;\Delta_n) = t, \text{ in } \mathbb{L}^2. \tag{15.3}$$

If $\Delta_n \overset{sum}{\to} \mathrm{Id}$, then the convergence in (15.3) also holds in the a.s.-sense.

Proof. We start with two very simple identities, valid for all $0 \leq r \leq s \leq u \leq v$:

$$\mathbb{E}\left[\Big((B_s - B_r)^2 - (s-r)\Big)^2\right] = 2(s-r)^2,$$

and

$$\mathbb{E}\left[\Big((B_s - B_r)^2 - (s-r)\Big)\Big((B_v - B_u)^2 - (v-u)\Big)\right] = 0.$$

It follows that, for any partition $\Delta \in P_{[0,t]}$, we have

$$\mathbb{E}\left[\Big(\operatorname{Var}_2(B;\Delta) - t\Big)^2\right] = 2\operatorname{Var}_2(\mathrm{Id};\Delta),$$

where Id denotes the identity function $s \mapsto s$. The first claim is now a consequence of the estimate

$$\operatorname{Var}_2(\mathrm{Id},\Delta) = \sum_{i=1}^{k} |t_i - t_{i-1}|^2 \leq t\,|\Delta|_{[0,t]}\,.$$

The second one follows from the first one and an application of the Borel-Cantelli lemma. □

Remark 15.6. Note that Proposition 15.5 *does not* imply that the paths of Brownian motion have finite 2-variation, a.s. In fact, it can be proved that, for each $t > 0$, $\operatorname{Var}_2(B;[0,t]) = \infty$, a.s.

Corollary 15.7 (Non-rectifiability of Brownian paths)**.** *Paths of the Brownian Motion have infinite variation on $[0,t]$, for all $t \geq 0$, a.s., i.e.*

$$\forall\, t > 0,\ \operatorname{Var}_1(B;[0,t]) = \infty,\ \text{a.s.}$$

Proof. For a partition $\Delta = \{t_0,\dots,t_k\} \in P_{[0,t]}$, we clearly have

$$\operatorname{Var}_2(B;\Delta) \leq \delta(B;\Delta)\operatorname{Var}_1(B,\Delta) \text{ where } \delta(B,\Delta) = \sup_{1\leq i\leq k} \left|B_{t_i} - B_{t_{i-1}}\right|.$$

If $\{\Delta_n\}_{n\in\mathbb{N}}$ is a sequence of partitions in $P_{[0,t]}$ with $\Delta_n \overset{sum}{\to} \mathrm{Id}$, Proposition 15.5 implies that $\operatorname{Var}_2(B,\Delta_n) \to t$, a.s. On the other hand, the continuity (and therefore uniform continuity on compacts) of the paths of the Brownian motion allows us to conclude that $\delta(B,\Delta_n) \to 0$, a.s. It follows that, necessarily, $\operatorname{Var}_1(B,\Delta_n) \to \infty$, and, a fortiori, that

$$\operatorname{Var}_1(B,[0,t]) = \sup_{\Delta\in P_{[0,t]}} \operatorname{Var}_1(B;\Delta) \geq \sup_n \operatorname{Var}_1(B;\Delta_n) = \infty,\ \text{a.s.}$$ □

Logarithmic laws (*)

We finish with some fine regularity properties of Brownian paths.

Definition 15.8 (Uniform local modulus of continuity)**.** The map δ, defined in a right neighborhood of 0, is called the **uniform local modulus of continuity** for the function $f : [0,1] \to \mathbb{R}$, if there exists $h_0 > 0$ such that, for each $t \in [0,1]$,

$$|f(t+h) - f(t)| \leq \delta(h), \text{ for all } 0 < h < h_0 \text{ with } t + h \leq 1.$$

The estimates from the Lévy-Ciesielski construction above, lead relatively directly to the following result:

Proposition 15.9 (A uniform local modulus of continuity for the Brownian motion)**.** *There exists a constant C such that, for almost all* ω

$$\delta(h) = C\sqrt{h \log(1/h)}$$

is the local uniform modulus of continuity of the trajectory $t \mapsto B_t(\omega)$*.*

Proof. We use the notation from the Lévy-Ciesielski construction and recall that

$$B_t^{(n)} = \sum_{k=1}^{n} \Delta_t^{(n)} \to B_t, \text{ uniformly over } t \in [0,1], \text{ a.s.}$$

In the course of the proof of Proposition 14.19, we established that the following bound

$$\mathbb{P}\left[\sup_{t\in[0,1]} \left|\Delta_t^{(n)}\right| \leq C_1\sqrt{n2^{-n}} \text{ ev.}\right] = 1, \tag{15.4}$$

holds for all C_1 large enough, say $C_1 = \sqrt{2}$. Thanks to the piecewise-linear structure of $\Delta^{(n)}$, and the fact that the intervals of linearity are of the size 2^{-n}, we clearly have

$$\mathbb{P}\left[\sup_{t\in[0,1]} \left|\tfrac{d}{dt}\Delta_t^{(n)}\right| \leq C_1\sqrt{n2^{n}} \text{ ev.}\right] = 1. \tag{15.5}$$

Therefore, there exists a random variable L with values in $\mathbb{N}$ such that

$$\sup_{t\in[0,1]} \left|\Delta_t^{(n)}\right| \leq C_1\sqrt{n2^{-n}} \text{ and } \sup_{t\in[0,1]} \left|\tfrac{d}{dt}\Delta_t^{(n)}\right| \leq C_1\sqrt{n2^{n}}, \text{ for all } n \geq L, \text{ a.s.}$$

For any random variable $K \in \mathbb{N}$ with $K \geq L$, $t \in [0,1]$, and $h > 0$, we have

$$|B_{t+h} - B_t| \leq \sum_{n=0}^{\infty} \left|\Delta^{(n)}(t+h) - \Delta^{(n)}(t)\right| \leq (I) + (II) + (III),$$

where

$$(I) = h\sum_{n=0}^{L} \sup_{t\in[0,1]} \left|\tfrac{d}{dt}\Delta_t^{(n)}\right|, \; (II) = h\sum_{n=L+1}^{K} C_1\sqrt{n2^{n}},$$

and

$$(III) = 2\sum_{n=K+1}^{\infty} C_1\sqrt{n2^{-n}}.$$

First, we pick (a random) $h_0 > 0$ so small that

$$(I) \leq \sqrt{h\log(1/h)}, \text{ for } 0 < h \leq h_0.$$

Then, given $0 < h < h_0$, we construct an $\mathbb{N}$-valued random variable K' such that $2^{-K'-1} \leq h < 2^{-K'}$, and set $K = \max(K', L+1)$. Using the estimate

$$\begin{aligned}\sum_{n=N}^{\infty} \sqrt{n2^{-n}} &= N2^{-N}\sum_{n=N}^{\infty}\sqrt{\tfrac{n}{N}2^{-(n-N)}}\\ &= N2^{-N}\sum_{k=0}^{\infty}\sqrt{(1+\tfrac{k}{N})2^{-k}} \leq C_2 N2^{-N},\end{aligned}$$

where $C_2 = \sum_{k\geq 0}\sqrt{(1+k)2^{-k}} < \infty$, we conclude that

$$\begin{aligned}(III) &\leq 2C_1\sum_{n=K'+1}^{\infty}\sqrt{n2^{-n}} \leq 2C_1C_2\sqrt{(1+K')2^{-K'-1}}\\ &\leq 4C_1C_2\sqrt{h\log(1/h)}.\end{aligned}$$

Finally,

$$\begin{aligned}(II) &\leq hC_1\sum_{n=1}^{K'}\sqrt{n2^n} \leq C_1\sum_{n=1}^{K'}\sqrt{n2^{(n-2K')}} = C_1\sum_{k=K'}^{2K'}\sqrt{(k-K')2^{-k}}\\ &\leq C_1\sum_{k=K'}^{\infty}\sqrt{k2^{-k}} \leq C_1C_2\sqrt{K'2^{-K'}} \leq 2C_1C_2\sqrt{h\log(1/h)}.\end{aligned}$$

All in all, $|B_{t+h} - B_t| \leq (1+6C_1C_2)\sqrt{h\log(1/h)}$, for $h \leq h_0$. □

The full result of Lévy (whose proof we omit) is that the function δ of Proposition 15.9 is optimal, and that the the constant C can be chosen to be equal to $\sqrt{2}$.

Theorem 15.10 (Lévy's modulus of continuity)**.**

$$\sup_{t\in[0,1]}\limsup_{h\searrow 0}\frac{|B_{t+h}-B_t|}{\sqrt{2h\log(1/h)}} = 1, \ a.s.$$

If one focuses on the fluctuations of the Brownian motion around a single, fixed, point $t \geq 0$, one gets a slightly tighter estimate.

Theorem 15.11 (Law of iterated logarithm)**.** *For each $t \geq 0$, we have*

$$\limsup_{h\searrow 0}\frac{|B_{t+h}-B_t|}{\sqrt{2h\log(\log(1/h))}} = 1, \ a.s.$$

Additional Problems

Problem 15.1 (p-variations of functions)**.**

1. Show that, for a function $f : [0, t] \to \mathbb{R}$, $\mathrm{Var}_p(f; [0, t]) < \infty$ implies $\mathrm{Var}_q(f; [0, t]) < \infty$ for $q > p > 0$.

2. For each $q > 1$, find an example of a function $f : [0, 1] \to \mathbb{R}$ with the property that

$$\begin{cases} \mathrm{Var}_p(f; [0,1]) < \infty, & p > q, \\ \mathrm{Var}_p(f; [0,1]) = \infty, & p \le q, \end{cases}$$

3. For each $q \ge 1$, find an example of a function $f : [0, 1] \to \mathbb{R}$ with the property that

$$\begin{cases} \mathrm{Var}_p(f; [0,1]) < \infty, & p \ge q, \\ \mathrm{Var}_p(f; [0,1]) = \infty, & p < q. \end{cases}$$

4. For a function $f : [0, \infty) \to \mathbb{R}$ of finite variation (remember, that means that $\mathrm{Var}_1(f; [0, t]) < \infty$ for all $t > 0$), define $F : [0, \infty) \to \mathbb{R}$ by $F(t) = \mathrm{Var}(f; [0, t])$, $t \ge 0$. Show that $|f(t) - f(s)| \le |F(t) - F(s)|$, for all $t, s \in [0, \infty)$. Deduce that f can be written as a difference of two monotone functions, and, more generally, the f is of finite variation if and only if it can be written as a difference of two monotone functions.

Problem 15.2 (Quadratic covariation of independent Brownian motions)**.** Let the stochastic processes $\{X_t\}_{t\in[0,\infty)}$ and $\{Y_t\}_{t\in[0,\infty)}$ be defined on the same probability space. For a partition $\Delta \in P_{[0,t]}$, we define the **quadratic covariation of X and Y along Δ** by

$$\mathrm{Var}_2(X, Y; \Delta) = \sum_{i=1}^{k} (X_{t_i} - X_{t_{i-1}})(Y_{t_i} - Y_{t_{i-1}}),$$

where $\Delta = \{0 = t_0, \dots, t_k = t\}$. If X and Y are two independent Brownian motions, i.e., such that the σ-algebras $\sigma(X_t; t \ge 0)$ and $\sigma(Y_t; t \ge 0)$ are independent, show that $\mathrm{Var}_2(X, Y; \Delta^{(n)}) \to 0$ in $\mathbb{L}^2$, for each sequence $\{\Delta^{(n)}\}_{n\in\mathbb{N}}$ in $P_{[0,t]}$ with $\Delta^{(n)} \to \mathrm{Id}$.

Problem 15.3 (Higher-dimensional Brownian motion)**.** For $d \in \mathbb{N}$, a vector-valued stochastic process $(B_t^{(1)}, \dots, B_t^{(d)})_{t\ge0}$, with values in $\mathbb{R}^d$, is said to be a **d-dimensional Brownian motion** if its components $B^{(1)}$, $\dots$, $B^{(d)}$ are independent Brownian motions. Given $n \in \mathbb{N}$, find necessary and sufficient conditions on the $\mathbb{R}^{n\times d}$-matrix H such that the process

$$W_t = \begin{bmatrix} W_t^{(1)} \\ \vdots \\ W_t^{(n)} \end{bmatrix} \text{ where } W_t = HB_t \text{ and } B_t = \begin{bmatrix} B_t^{(1)} \\ \vdots \\ B_t^{(d)} \end{bmatrix} \quad t \geq 0$$

is an n-dimensional Brownian motion.

Problem 15.4 (Monotonicity and maxima of the Brownian path)**.** Prove the following statements for a Brownian motion B:

1. B is monotone on no interval of the form $[r, s]$, $0 \leq r < s$, a.s.

2. For each $p > 0$, the distribution of the random variable $M_t = \sup_{s \leq t} B_s$ is diffuse, i.e. $\mathbb{P}[M_t = a] = 0$, for all $a \in \mathbb{R}$. *Hint:* Argue, first, that it is enough to assume that $t = 1$. Let $\tilde{M}_1$ be an independent random variable with the same distribution as M_1. Show that $\sqrt{2}M_1$ and $\max(M_1, W_1 + \tilde{M}_1)$ have the same distribution. Deduce that the only possible atom for M_1 is 0. Then show that $\mathbb{P}[M_1 > 0] = 1$.

3. B attains different maxima on any two non-overlapping intervals (r_1, s_1) and (r_2, s_2), a.s.

4. Each local maximum of B is a strict local maximum, a.s.

5. B achieves its global maximum on $[0, 1]$ in exactly one point, a.s.

Problem 15.5 (Non-differentiability of Brownian paths)**.**

1. Show that if $f : [0, 1) \to \mathbb{R}$ is differentiable at $t \in [0, 1)$ (the right derivative is considered at the $t = 0$), then there exists $l, n_0 \in \mathbb{N}$ such that
$$\left| f(\tfrac{j}{n}) - f(\tfrac{j-1}{n}) \right| \leq \tfrac{l}{n},$$
for all $n \geq n_0$ and all $i < j \leq i + 3$, where $i = \lfloor nt \rfloor + 1$ and $\lfloor x \rfloor$ denotes the largest integer not larger than x.

2. Let $\{B_t\}_{t \in [0,\infty)}$ be a Brownian motion. For $s \in [0, 1)$ we define D_s as the set of all $\omega \in \Omega$ such that the trajectory $t \mapsto B_t(\omega)$ is differentiable at s. Show that
$$\bigcup_{s \in [0,1)} D_s \subseteq \Gamma, \text{ where } \Gamma = \bigcup_{l \geq 1} \liminf_{n \to \infty} \bigcup_{i=1}^{n+1} \bigcap_{j=i+1}^{i+3} \left\{ \left| B_{j/n} - B_{(j-1)/n} \right| \leq \tfrac{l}{n} \right\}.$$

3. (*) Show that $\mathbb{P}[\Gamma] = 0$.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 16
Abstract Nonsense

Brownian motion is just an example (albeit a particularly important one) of a whole "zoo" of interesting and useful continuous-time stochastic processes. This chapter deals with the minimum amount of the general theory of stochastic processes[1] and related topics necessary for the rest of these notes. To ease the understanding, we keep track of the amount of structure a certain notion depends on; e.g., if it depends on the presence of a filtration or it can be defined without it.

[1] widely known as the "Théorie générale" in homage to the French school that developed it

Properties without filtrations or probability

We remind the reader that a **stochastic process** (in continuous time) is a collection $\{X_t\}_{t\in[0,\infty)}$ of random variables. Without any further assumptions, little can be said about the regularity properties of its paths $t \mapsto X_t(\omega)$. Indeed, not even Borel-measurability can be guaranteed; one only need to define $X_t(\omega) = f(t)$, where $f(t)$ is a non-measurable function. The notion of **(joint) measurability**, defined below, helps:

Definition 16.1 (Measurability)**.** A stochastic process $\{X_t\}_{t\in[0,\infty)}$ is said to be **(jointly) measurable** if the map $X : [0,\infty) \times \Omega \to \mathbb{R}$, given by, $X(t,\omega) = X_t(\omega)$ is (jointly) measurable from $\mathcal{B}([0,\infty)) \otimes \mathcal{F}$ to $\mathcal{B}(\mathbb{R})$.

Even though non-measurable processes clearly exist (and abound), they are never really necessary. For that reason, make the following standing assumption:

Assumption 16.2. All stochastic processes from now on are assumed to be (jointly) measurable.

This is not a significant assumption. All processes in the sequel will be constructed from the already existing ones in a measurable manner and the (joint) measurability will be easy to check. We will see quite soon that the (everywhere-continuous version of the) Brownian motion is always measurable.

An immediate consequence of this assumption is that the trajectories $t \mapsto X_t(\omega)$ are automatically Borel-measurable functions. This follows from the fact that sections of jointly-measurable functions are themselves measurable.

Remark 16.3. The measurability problem arises only in continuous time. In discrete time, every stochastic process $\{X_n\}_{n\in\mathbb{N}}$ is automatically jointly measurable. The reason is that the measurable structure on $\mathbb{N}$ is much simpler than that on $[0,\infty)$.

Definition 16.4 (Path-regularity classes)**.** A stochastic process $\{X_t\}_{t\in[0,\infty)}$ is said to be

1. **Continuous**, if all of its trajectories $t \mapsto X_t(\omega)$ are continuous functions on $[0,\infty)$,
2. **Right-continuous**, if all of its trajectories $t \mapsto X_t(\omega)$ are right-continuous functions, i.e., if $X_t(\omega) = \lim_{s\searrow t} X_s(\omega)$, for all $t \in [0,\infty)$.
3. **Left-continuous**, if all of its trajectories $t \mapsto X_t(\omega)$ are left-continuous functions, i.e., if $X_t(\omega) = \lim_{s\nearrow t} X_s(\omega)$, for all $t \in (0,\infty)$.
4. **RCLL** if all of its trajectories $t \mapsto X_t(\omega)$ have the following two properties
   (a) $X_t(\omega) = \lim_{s\searrow t} X_s(\omega)$, for all $t \in [0,\infty)$,
   (b) $\lim_{s\nearrow t} X_s(\omega)$, exists for all $t \in (0,\infty)$.
5. **LCRL** if all of its trajectories $t \mapsto X_t(\omega)$ have the following two properties
   (a) $X_t(\omega) = \lim_{s\nearrow t} X_s(\omega)$, for all $t \in (0,\infty)$, and
   (b) $\lim_{s\searrow t} X_t(\omega)$, exists for all $t \in [0,\infty)$.
6. **of finite variation** if almost all of its trajectories have finite variation on all segments $[0,t]$.
7. **bounded** if there exists $K \geq 0$ such that all of its trajectories are bounded on $[0,\infty)$ by K in absolute value.

Remark 16.5.

1. The acronym **RCLL** (right-continuous with left limits) is sometimes replaced by "càdlàg", which stands for the French phrase "**c**ontinue **à d**roite, limitée **à g**auche". Similarly, **LCRL** (left-continuous with right limits) is replaced by "càglàd", which stands for "**c**ontinue **à g**auche, limitée **à d**roite".
2. For a RCLL process $\{X_t\}_{t\in[0,\infty)}$, it is customary to denote the left limit $\lim_{s\nearrow t} X_s$ by X_{t-} (by convention $X_{0-} = 0$). Similarly, $X_{t+} = \lim_{s\searrow t} X_s$, for a LCRL process. The random variable $\Delta X_t = X_t - X_{t-}$ (or $X_{t+} - X_t$ in the LCRL case) is called the **jump** of X at t.

Properties without probability

What really distinguishes the theory of stochastic processes from the theory of real-valued functions on product spaces, is the notion of adaptivity. The additional structural element comes in through the notion of a filtration and models the accretion of information as time progresses.

Filtrations and stopping times

Definition 16.6 (Filtrations and adaptedness)**.** A **filtration** is a family $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,\infty)}$ of sub-σ-algebras of $\mathcal{F}$ with the property that $\mathcal{F}_s \subseteq \mathcal{F}_t$, for $0 \leq s < t$. A stochastic process $X = \{X_t\}_{t\in[0,\infty)}$ is said to be **adapted** to $\mathbb{F}$ if $X_t \in \mathcal{F}_t$, for all t.

The **natural filtration** $\mathbb{F}^X$ of a stochastic process $\{X_t\}_{t\in[0,\infty)}$ is the smallest filtration with respect to which X is adapted, i.e.,

$$\mathcal{F}^X_t = \sigma(X_s; s \leq t),$$

and is related to the situation where the available information at time t is obtained by observing the values of the process (and nothing else) X up to time t.

Problem 16.1 (*)**.**

1. For a (possibly uncountable) family $(X_\alpha)_{\alpha\in A}$ of random variables, let $\mathcal{G}$ denote the σ-algebra generated by $(X_\alpha)_{\alpha\in A}$, i.e., $\mathcal{G} = \sigma(X_\alpha, \alpha \in A)$. Given a random variable $Y \in \mathcal{G}$, show that there exists a *countable* subset $C \subseteq A$, such that

$$Y \in \sigma(X_\alpha\ :\ \alpha \in C).$$

   Hint: Use the Monotone-Class Theorem.

2. Let $\mathbb{F}^X$ be the natural filtration of the stochastic process $\{X_t\}_{t\in[0,T]}$. Show that for each $\mathcal{F}_1$-measurable random variable Y there exists a countable subset S of $[0,1]$ such that $Y(\omega) = Y(\omega')$, as soon as $X_t(\omega) = X_t(\omega')$, for all $t \in S$.

A random variable τ taking values in $[0,\infty]$ is called a **random time**. The additional element $+\infty$ is used as a placeholder for the case when τ "does not happen".

Definition 16.7 (Stopping and optional times)**.** Given a filtration $\mathbb{F}$, a random time τ is said to be

- an $\mathbb{F}$-**stopping time** if

$$\{\tau \leq t\} \in \mathcal{F}_t, \text{ for all } t \geq 0.$$

- an $\mathbb{F}$-**optional time** if

$$\{\tau < t\} \in \mathcal{F}_t, \text{ for all } t > 0.$$

Both stopping and optional times play the role of the "stopping time" in the continuous case. In discrete time, they collapse into the same concept, but there is a subtle difference in the continuous time. Each stopping time is an optional time, as is easily seen as follows: for $t > 0$, Suppose that $t_n \nearrow t$ with $t_n < t$, for all n. Then

$$\{\tau < t\} = \bigcup_{n \in \mathbb{N}} \{\tau \leq t_n\} \in \mathcal{F}_t,$$

because $\{\tau \leq t_n\} \in \mathcal{F}_{t_n} \subseteq \mathcal{F}_t$, for each $n \in \mathbb{N}$. The following example shows that the two concepts are not the same:

Example 16.8 (Optional but not stopping times)**.** Let $\{B_t\}_{t\in[0,\infty)}$ be the standard Brownian motion, and let $\mathbb{F} = \mathbb{F}_B$ be its natural filtration. Consider the random time τ, defined by

$$\tau(\omega) = \inf\{t \geq 0 : B_t(\omega) > 1\} \subseteq [0, \infty),$$

under the convention that $\inf \emptyset = +\infty$. We will see shortly (in Proposition 16.12), τ is an optional time. It is, however, not a stopping time. The rigorous proof will have to wait a bit, but here is a heuristic argument. Having only the information available at time 1, it is always possible to decide whether $\tau \leq 1$ or not. If $B_u > 1$ for some $u \leq 1$, then, clearly, $\tau \leq 1$. Similarly, if $B_u \geq 1$ for all $u < 1$ and $B_1 > 1$, we can also easily conclude that $\tau > 1$. The case when $B_u \leq 1$, for all $u < 1$ and $B_1 = 1$ is problematic. To decide whether $\tau = 1$ or not, we need to know about the behavior of B in some right neighborhood of 1. The process could enter the set $(1, \infty)$ right after 1, in which case $\tau = 1$. Alternatively, it could "bounce back" and not enter the set $(1, \infty)$ for a while longer. The problem is that the time τ is defined using the *infimum* of a set it does not have to be an element of.

It you need to see a more rigorous example, here is one. We work on the discrete measurable space $(\Omega, \mathcal{F})$, where $\Omega = \{-1, 1\}$ and $\mathcal{F} = 2^{\{-1,1\}}$. Let $X_t(\omega) = \omega t$, for $t \geq 0$ and let $\mathbb{F} = \mathbb{F}^X$ be the natural filtration of X. Clearly, $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t = \mathcal{F}$, for $t > 0$. Consequently, the random time

$$\tau(\omega) = \inf\{t \geq 0 : X_t(\omega) > 0\} = \begin{cases} 0, & \omega = 1, \\ +\infty, & \omega = -1 \end{cases},$$

is not a stopping time, since $\{1\} = \{\tau \leq 0\} \notin \mathcal{F}_0$. On the other hand, it is an optional time. Indeed $\mathcal{F}_{t+} = \mathcal{F}$, for all $t \geq 0$.

The main difference between optional and stopping time - if Example 16.8 is to be taken as typical - is that no peaking is allowed for stopping times, while, for optional ones, a little bit of peaking is tolerated. What is exactly understood by "the future" is, in turn, dictated bt the filtration, and the difference between the "present" and the "immediate future" at time t is encoded in the difference between the σ-algebras: $\mathcal{F}_t$ and $\cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}$. If one is willing to add the extra information contained in the latter into one's information set, the difference between optional and stopping times would disappear. To simplify the discussion in the sequel, we introduce the following notation

$$\mathcal{F}_{t+} = \cap_{\varepsilon>0}\mathcal{F}_{t+\varepsilon}, \text{ and } \mathbb{F}_+ = \{\mathcal{F}_{t+}\}_{t\in[0,\infty)},$$

and call the filtration $\mathbb{F}_+$ the **right-continuous augmentation of** $\mathbb{F}$. A filtration $\mathbb{F}$ with $\mathbb{F} = \mathbb{F}_+$ is said to be **right continuous**. The notion of right continuity only makes (nontrivial) sense in continuous time. In the discrete case, it formally implies that $\mathcal{F}_n = \mathcal{F}_0$, for all $n \in \mathbb{N}$.

Proposition 16.9 ($\mathbb{F}$-optional = $\mathbb{F}_+$-stopping)**.** *A random time τ is an $\mathbb{F}$-optional time if and only if is an $\mathbb{F}_+$-stopping time.*

Proof. Let τ be a $\mathbb{F}$-optional time. For $t \geq 0$, we pick a sequence $\{t_n\}_{n\in\mathbb{N}}$ with $t_n \searrow t$ and $t_n > t$. For $m \in \mathbb{N}$, we have

$$\{\tau \leq t\} = \cap_{n\geq m} A_n, \text{ where } A_n = \{\tau < t_n\} \in \mathcal{F}_{t_n} \subseteq \mathcal{F}_{t_m}.$$

Therefore, $\{\tau \leq t\} \in \cap_m \mathcal{F}_{t_m} = \mathcal{F}_{t+}$, and τ is an $\mathbb{F}_+$-stopping time.

Conversely, if τ is a $\mathbb{F}_+$ stopping time, then, for $t > 0$ and a sequence $t_n \nearrow t$, with $t_n < t$ we have

$$\{\tau < t\} = \cup_n B_n, \text{ where } B_n = \{\tau \leq t_n\} \in \mathcal{F}_{t_n+} \subseteq \mathcal{F}_t.$$

Consequently, $\{\tau < t\} \in \mathcal{F}_t$, and τ is an $\mathbb{F}$-optional time. □

Corollary 16.10 (If $\mathbb{F} = \mathbb{F}_+$, then optional=stopping)**.** *If the filtration $\mathbb{F}$ is right continuous, the collections of optional and stopping times coincide.*

The above discussion is very useful when one tries to manufacture properties of optional times from the properties of stopping times. Here is an example:

Proposition 16.11 (Stability of optional and stopping times)**.** *If τ and σ are stopping (optional) times, then so are*

$$\sigma + \tau, \max(\sigma, \tau), \textit{ and } \min(\sigma, \tau).$$

Proof. By Proposition 16.9, it is enough to prove the statement in the case of stopping times. Indeed, to treat optional times, it will suffice to replace the filtration $\mathbb{F}$ with the right-continuous augmentation $\mathbb{F}_+$.

We focus on the case of the sum $\sigma + \tau$, and leave the other two to the reader. To show that $\sigma + \tau$ is a stopping time, we consider the event $\{\sigma + \tau > t\} = \{\sigma + \tau \leq t\}^c$ and note that

$$\{\sigma + \tau > t\} = \bigcup_{q \in \mathcal{Q}_+ \cap [0,t]} \Big(\{\sigma > q\} \cap \{\tau > t - q\} \Big).$$

It remains to observe that

$$\{\sigma > q\} \cap \{\tau > t - q\} \in \mathcal{F}_{\max(q, t-q)} \subseteq \mathcal{F}_t. \qquad \square$$

For a stochastic process $\{X_t\}_{t \in [0,\infty)}$ and a subset $A \subseteq \mathbb{R}$, we define the **hitting time** τ_A of A as

$$\tau_A = \inf\{t \geq 0 : X_t \in A\}.$$

Unlike in the discrete case, it is not immediately obvious that X is a stopping or an optional time. In fact, it is not even immediately clear that τ_A is even a random variable. Indeed, even if A is Borel-measurable, the set $\{\tau_A > t\}$ is a-priori defined as a combination of *uncountably* many restrictions. Under suitable regularity conditions, however, we do recover these intuitive properties (note, though, how the two cases below differ from each other):

Proposition 16.12 (Hitting times which are stopping times)**.** *If X is a* continuous *process, the map τ_A is*

1. *a stopping time when A is closed, and*
2. *an optional time when A is open or closed,*

Proof.

1. For $n \in \mathbb{N}$, define $A_n = \{x \in \mathbb{R} : d(x, A) < 1/n\}$. Since $d(\cdot, A)$ is a continuous function, A_n is an open set. The reader will easily check that

$$\{\tau_A \leq t\} = \{X_t \in A\} \quad \bigcup \quad \left(\bigcap_{n \in \mathbb{N}} \bigcup_{q \in \mathcal{Q}_+ \cap [0,t)} \{X_q \in A_n\} \right),$$

which implies directly that τ_A is a stopping time.

2. For $t > 0$, we clearly have

$$\{\tau_A < t\} = \bigcup_{q \in \mathcal{Q} \cap [0,t)} \{X_q \in A\},$$

and so $\{\tau_A < t\} \in \mathcal{F}_t^X$, for all $t > 0$. It follows that $\{\tau_A \le t\} = \cap_{n \ge m} \{\tau_A < t + 1/n\} \in \mathcal{F}^X_{t+1/m}$, for all $m \in \mathbb{N}$. Consequently, $\{\tau_A \le t\} \in \mathcal{F}_t^{X+}$. □

Remark 16.13. The results in Proposition 16.12 are just the tip of the iceberg. First of all, the proof readily generalized to the case when the process X takes values in a d-dimensional Euclidean space. Also, one can show that the continuity assumption is not necessary. In fact, one needs minimal regularity on A (say, Borel), and minimal regularity on X (much less than, say, RCLL) to conclude that τ_A is an optional time.

Progressive measurability In addition to the notions of measurability mentioned above, one often uses a related notion which is better suited for the situation when a filtration is present and appears as a separate concept only in continuous time.

Definition 16.14 (Progressive measurability)**.** We say that the stochastic process $\{X_t\}_{t\in[0,\infty)}$ is **progressively measurable** (or, simply, **progressive**) if, when seen as a mapping on the product space $[0,\infty) \times \Omega$, it is measurable with respect to the σ-algebra Prog, where

$$\text{Prog} = \Big\{ A \in \mathcal{B}([0,\infty)) \times \mathcal{F} \,:\, A \cap ([0,T] \times \Omega) \in \mathcal{B}([0,T]) \times \mathcal{F}_T, \text{ for all } T \ge 0. \Big\}$$

Problem 16.2. Show that an adapted process $\{X_t\}_{t\in[0,\infty)}$ is progressively measurable if and only if, for each $T > 0$, the stopped process X^T is measurable when understood as a process on $(\Omega, \mathcal{F}_T)$.

Problem 16.3. Let $\{H_t\}_{t\in[0,\infty)}$ be a bounded progressively-measurable process. Show that the process

$$\{\textstyle\int_0^t X_u \, du\}_{t\in[0,\infty)}$$

is well-defined, continuous and progressively measurable. Argue that the assumption of uniform boundedness can be relaxed to that of integrability.

Clearly, a progressively measurable process is adapted and measurable, but the converse implication is not true.

Example 16.15 (*)(Measurable and adapted, but not progressively measurable)**.** Let Ω denote the set of all lower semicontinuous functions[2] $\omega : (0,\infty) \to \mathbb{R}$ such that

$$\int_0^T |\omega(t)| \, dt < \infty, \text{ for all } T > 0.$$

[2] A function $f : (0,\infty) \to \mathbb{R}$ is said to be **lower semicontinuous** if $\{f > c\} = f^{-1}((c,\infty))$ is open for each $c \in \mathbb{R}$.

For each pair $0 < a \leq b < \infty$, we define the map $I_{a,b} : \Omega \to \mathbb{R}$ by

$$I_{a,b}(\omega) = \inf_{a \leq t \leq b} \omega(t) \in (-\infty, \infty),$$

and let the σ-algebra $\mathcal{F}$ be generated by all $I_{a,b}$, $0 < a \leq b < \infty$. The filtration $\mathbb{F}$ is the natural filtration of the coordinate process $\{H_t\}_{t \in [0,T]}$, where $H_t(\omega) = \omega(t)$, i.e., $\mathcal{F}_t = \sigma(H_s, s \leq t)$. Hence, H is trivially adapted. To show that it is measurable it suffices to observe that for each $c \in \mathbb{R}$, we have

$$\{(t,\omega) \in (0,\infty) \times \Omega : H_t(\omega) > c\} = \bigcup [p,q] \times \{I_{p,q} > c\},$$

where the union is taken over all pairs $0 < p \leq q < \infty$ of rational numbers.

To show that H is not progressive, it will be enough (see Problem 16.3) to show that the process $\{X_t\}_{t \in [0,T]}$, defined by

$$X_t = \int_0^t H_u \, du, \; t \geq 0,$$

is *not adapted*. In fact, we have $X_t \notin \mathcal{F}_t$, *for all* $t > 0$. Suppose, to the contrary, that for some $t > 0$, we have $X_t \in \mathcal{F}_t$. Then, by Problem 16.1, we could find a countable subset S of $[0,t]$ such that $X_t(\omega) = X_t(\omega')$ as soon as $\omega|_S = \omega'|_S$. That, however, leads to a contradiction. Indeed, we can always pick $\omega = 1$ and $\omega' = \mathbf{1}_O$, where O is an open subset of $(0,t]$ of Lebesgue measure strictly less than 1, which contains S.

Processes with mildly regular trajectories are progressive.

Proposition 16.16 (RCLL or LCRL + adapted $\to$ progressive)**.** *Suppose that the adapted stochastic process $\{X_t\}_{t \in [0,\infty)}$ has the property that all of its trajectories are right continuous or that all of its trajectories are left continuous. Then, X is progressively measurable.*

Proof. We assume that $\{X_t\}_{t \in [0,\infty)}$ is right continuous (the case of a left-continuous process is analogous). For $T \in (0,\infty)$ and $n \in \mathbb{N}$, $n \geq 1/T$, we define the process $\{X^n_t\}_{t \in [0,\infty)}$ by

$$X^n_t(\omega) = \begin{cases} X_T(\omega), & t \geq T, \\ \sum_{k=0}^{\lfloor nT \rfloor - 1} X_{\frac{k+1}{n}}(\omega) \mathbf{1}_{[\frac{k}{n}, \frac{k+1}{n})}(t), & t < T \end{cases}.$$

In words, we use the right-endpoint value of the process X throughout the interval $[\frac{k}{n}, \frac{k+1}{n})$. It is then easy to see that the restricted processes $\{X^n_t\}_{t \in [0,T]}$ are measurable with respect to $\mathcal{B}([0,T]) \times \mathcal{F}_T$, but note that it is not necessarily adapted.

It remains to show that X^n converges towards X. For $(t,\omega) \in (0,\infty) \times \Omega$ we have $X^n_t = X_{k_n(t)/n}$, where $k_n(t)$ is the smallest $k \in \mathbb{N}_0$ such that $k_n(t) \geq nt$. Since $k_n(t) - 1 < nt \leq k_n(t)$, we have $k_n(t)/n \searrow t$, as $n \to \infty$, and $X^n_t \to X_t$ by the RCLL property. □

Let $\{X_t\}_{t\in[0,\infty)}$ be a random process, and let τ be a random time. The **stopped process** $\{X^\tau_t\}_{t\in[0,\infty)}$ is defined by $X^\tau_t(\omega) = X_{t\wedge\tau(\omega)}(\omega)$, for $t \geq 0$.

Proposition 16.17 (A stopped progressive is still a progressive)**.** *Let $\{X_t\}_{t\in[0,\infty)}$ be a progressive process, and let τ be a stopping time. Then the stopped process X^τ is also progressive.*

Proof. We fix $T > 0$ and note that the map $(t,\omega) \to (\tau(\omega) \wedge t, \omega)$, from $([0,T] \times \Omega, \mathcal{B}([0,T]) \otimes \mathcal{F}_T)$ to itself is measurable. Then, we compose it with the map $(t,\omega) \to X_t(\omega)$, which, by the assumption, is jointly mesurable on $([0,T] \times \Omega, \mathcal{B}([0,T]) \otimes \mathcal{F}_T)$. The obtained mapping, namely the process $X^{T\wedge\tau} = (X^T)^\tau$, is thus measurable, and, since this holds for each $T > 0$, the process X^τ is progressive. □

In an analogy with the interpretation of $\mathcal{F}_t$ as the information available at time t, we define a σ-algebra which could be interpreted to contain all the information at a *stopping* time τ:

$$\mathcal{F}_\tau = \sigma\Big\{X^\tau_t \,:\, t \geq 0,\ X \text{ is a progressively-measurable process}\Big\}.$$

In words, $\mathcal{F}_\tau$ is generated by the values of all progressive processes stopped at τ.

Problem 16.4 (Properties of $\mathcal{F}_\tau$)**.** Let $\sigma, \tau, \{\tau_n\}_{n\in\mathbb{N}}$ be $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-stopping times. Show that

1. $\mathcal{F}_\tau = \{A \in \bigvee_{t\geq 0} \mathcal{F}_t \,:\, A \cap \{\tau \leq t\} \in \mathcal{F}_t$, for all $t \geq 0\}$, where $\bigvee_{t\geq 0} \mathcal{F}_t = \sigma(\cup_{t\geq 0} \mathcal{F}_t)$.
2. $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$, if $\sigma \leq \tau$,
3. $\mathcal{F}_\tau = \cap_{n\in\mathbb{N}} \mathcal{F}_{\tau_n}$ if $\tau_1 \geq \tau_2 \geq \dots \geq \tau$, $\tau = \lim_n \tau_n$, and the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ is right-continuous.
4. If there exists a countable set $Q = \{q_k \,:\, k \in \mathbb{N}\} \subseteq [0,\infty)$ such that $\tau(\omega) \in Q$, for all $k \in \mathbb{N}$, then
$$\mathcal{F}_\tau = \{A \in \bigvee_{t\geq 0} \mathcal{F}_t \,:\, A \cap \{\tau = q_k\} \in \mathcal{F}_{q_k},\ \text{for all } k \in \mathbb{N}\}.$$
5. $X_\tau \mathbf{1}_{\{\tau<\infty\}}$ is a random variable in $\mathcal{F}_\tau$, whenever X is a progressively measurable process.

If one only wants X_τ to be a random variable, but not necessarily in $\mathcal{F}_\tau$, progressive measurability is too much to require:

Proposition 16.18 (Sampling of a measurable process at a stopping time)**.** *Let $\{X_t\}_{t\in[0,\infty)}$ be a (measurable) process and let τ be a $[0,\infty]$-valued random variable. Then $X_\tau \mathbf{1}_{\{\tau<\infty\}}$ is a random variable.*

Proof. Simply pick the trivial filtration $\mathcal{F}_t = \mathcal{F}$, $t \geq 0$. In this case the notions of measurability and progressive measurability coincide. □

The progressive σ-algebra of Definition 16.14 has another nice property which will simplify our construction of the stochastic integral in the sequel.

Proposition 16.19 (Approximability of progressive processes)**.** *For a bounded progressive process X, there exists a sequence of continuous and adapted processes $\{X^{(n)}\}_{n\in\mathbb{N}}$ such that*

$$X_t^{(n)}(\omega) \to X_t(\omega), \lambda - \textit{a.e. in } t, \textit{ for each } \omega \in \Omega.$$

Proof. We consider the process Y defined by $Y_t(\omega) = \int_0^t X_u(\omega)\, du$. By Problem 16.3, the process Y is continuous and progressively measurable, and, therefore, so are the processes $\{X^{(n)}\}_{n\in\mathbb{N}}$, defined by

$$X_t^{(n)} = n\Big(Y_t - Y_{t-1/n}\Big), \text{ for } t \geq 0,$$

where we use the convention that $Y_s = 0$, for $s \leq 0$. It remains to use the Lebesgue Differentiation Theorem for Lipschitz functions. □

Continuous-time martingales

The definition of (super-, sub-) martingales in continuous time is the formally the same as the corresponding definition in discrete time:

Definition 16.20 (Continuous time (sub,super) martingales)**.** Given a filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$, a stochastic process $\{X_t\}_{t\in[0,\infty)}$ is said to be an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-**supermartingale** if

1. $X_t \in \mathcal{F}_t$, for all $t \geq 0$
2. $X_t \in \mathcal{L}^1$, for all $t \in [0,\infty)$, and
3. $\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s$, a.s., whenever $s \leq t$.

A process $\{X_t\}_{t\in[0,\infty)}$ is called a **submartingale** if $\{-X_t\}_{t\in[0,\infty)}$ is a supermartingale. A **martingale** is a process which is both a supermartingale and a submartingale at the same time, i.e., for which the equality holds in 3., above.

Problem 16.5 (Brownian motion is a martingale)**.** Show that the Brownian motion is a martingale with respect to its natural filtration $\mathbb{F}^B$, as well as with respect to its right-continuous augmentation $\mathbb{F}^B_+$.

Remark 16.21. An example of a continuous-time martingale that is not continuous is the **compensated Poisson process**, i.e., the process $N_t - \lambda t$, $t \geq 0$, where N_t is a Poisson process with parameter $\lambda > 0$ (see Problem 16.9). It is useful to keep $N_t - \lambda t$ in mind as a source of counterexamples.

The theory of continuous-time martingales (submartingales, supermartingales) largely parallels the discrete-time theory. There are two major tricks that help us transfer discrete-time results to the continuous time:

1. *Approximation*: a stopping time τ can be approximated from above by a sequence of stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ each of which takes values in a countable set.

2. *Use of continuity properties of the paths*: typically, the right-continuity of the trajectories will imply that $X_{\tau_n} \to X_\tau$, and an apropriate limit theorem is then used.

Instead of giving detailed proofs of all the continuous-time counterparts of the optional sampling theorems, martingale-inequalities, etc., we only state most of our results. We do give a detailed proof of our first theorem, though, and, before we do that, here is some notation and a useful concept:

1. for $t \geq 0$, $\mathcal{S}_t$ denotes the set of all stopping times τ such that $\tau(\omega) \leq t$, for all $\omega \in \Omega$,

2. $\mathcal{S}_b = \cup_{t\geq 0}\mathcal{S}_t$ is the set of all bounded stopping times, and

3. $\mathcal{S}$ denotes the set of all stopping times.

Definition 16.22 (Classes (DL) and (D))**.**

1. A measurable process $\{X_t\}_{t\in[0,\infty)}$ is said to be **of class (DL)** if the family

$$\{X_\tau \,:\, \tau \in \mathcal{S}_t\} \text{ is uniformly integrable for all } t \geq 0.$$

2. A measurable process $\{X_t\}_{t\in[0,\infty)}$ is said to be **of class (D)** if the family

$$\{X_\tau \mathbf{1}_{\{\tau<\infty\}} \,:\, \tau \in \mathcal{S}\} \text{ is uniformly integrable.}$$

Proposition 16.23 (Bounded Optional Sampling). *Let $\{M_t\}_{t\in[0,\infty)}$ be a right-continuous martingale. Then, M is of class (DL) and*

$$\mathbb{E}[M_t|\mathcal{F}_\tau] = M_\tau \text{ and } \mathbb{E}[M_\tau] = \mathbb{E}[M_0], \text{ for all } \tau \in \mathcal{S}_t,\ t \geq 0.. \tag{16.1}$$

Proof. Let $\mathcal{S}_t^f$ be the set of all $\tau \in \mathcal{S}_t$ such that τ takes only finitely many values. We pick $\tau \in \mathcal{S}_t^f$, and take $0 \leq t_1 < t_2 < \dots < t_n \leq t$ to be the set of all the values it can take. Then for $A \in \mathcal{F}_\tau$, we have

$$\begin{aligned}\mathbb{E}[M_\tau \mathbf{1}_A] &= \sum_{k=1}^n \mathbb{E}[M_\tau \mathbf{1}_A \mathbf{1}_{\{\tau=t_k\}}] = \sum_{k=1}^n \mathbb{E}[M_{t_k} \mathbf{1}_A \mathbf{1}_{\{\tau=t_k\}}] \\ &= \sum_{k=1}^n \mathbb{E}[\mathbb{E}[M_t \mathbf{1}_A \mathbf{1}_{\{\tau=t_k\}}|\mathcal{F}_{t_k}]] = \mathbb{E}[M_t \mathbf{1}_A],\end{aligned} \tag{16.2}$$

so that $M_\tau = \mathbb{E}[M_t|\mathcal{F}_\tau]$. Therefore, the family $\{M_\tau : \tau \in \mathcal{S}_t^f\}$ is uniformly integrable, as each of its members can be represented as the conditional expectation of M_t with respect to some sub-σ-algebra of $\mathcal{F}_t$. Consider now a general stopping time $\tau \in \mathcal{S}_t$, and approximate it by the sequence $\{\tau_n\}_{n\in\mathbb{N}}$, given by

$$\tau_n = 2^{-n}\lceil 2^n \tau \rceil \wedge t,$$

so that each τ_n is in $\mathcal{S}_t$. Since $\tau_n \searrow \tau$, right continuity implies that $M_{\tau_n} \to M_\tau$ and, so, by uniform integrability of $\{M_{\tau_n}\}_{n\in\mathbb{N}}$, we conclude that M_τ is in the $\mathbb{L}^1$-closure of the uniformly integrable set $\{M_\tau : \tau \in \mathcal{S}_t^f\}$. In other words,

$$\{M_\tau : \tau \in \mathcal{S}_t\} \subseteq \overline{\{M_\tau : \tau \in \mathcal{S}_t^f\}}^{\mathbb{L}^1},$$

where $\overline{(\cdot)}^{\mathbb{L}^1}$ denotes the closure in $\mathbb{L}^1$. Using the fact that the $\mathbb{L}^1$-closure of a uniformly integrable set is uniformly integrable (why?), we conclude that $\{M_t\}_{t\in[0,\infty)}$ is of class (DL). To show (16.1), we note that uniform integrability (via the backward martingale convergence theorem - see Corollary 12.17) implies that

$$M_\tau = \lim_n M_{\tau_n} = \lim_n \mathbb{E}[M_t|\mathcal{F}_{\tau_n}],$$

where all the limits are in $\mathbb{L}^1$. Taking the conditional expectation with respect to $\mathcal{F}_\tau$ yields that

$$\begin{aligned}M_\tau = \mathbb{E}[M_\tau|\mathcal{F}_\tau] &= \mathbb{E}[\lim_n \mathbb{E}[M_t|\mathcal{F}_{\tau_n}]|\mathcal{F}_\tau] \\ &= \lim_n \mathbb{E}[\mathbb{E}[M_t|\mathcal{F}_{\tau_n}]|\mathcal{F}_\tau] = \mathbb{E}[M_t|\mathcal{F}_\tau],\end{aligned}$$

and, in particular, that $\mathbb{E}[M_\tau] = \mathbb{E}[M_t] = \mathbb{E}[M_0]$. □

A partial converse and a useful martingality criterion is given in the following proposition:

Proposition 16.24 (A characterization of martingales). *Let $\{M_t\}_{t\in[0,\infty)}$ be an adapted and right-continuous process with the property that $M_\tau \in \mathbb{L}^1$ for all $\tau \in \mathcal{S}_b$. Then $\{M_t\}_{t\in[0,\infty)}$ is a martingale if and only if*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0], \text{ for all } \tau \in \mathcal{S}_b. \tag{16.3}$$

Proof. Only sufficiency needs a proof. We consider stopping times of the form $\tau = s\mathbf{1}_{A^c} + t\mathbf{1}_A$, for $0 \le s \le t < \infty$ and $A \in \mathcal{F}_s$, as for such a τ, the condition (16.3) implies that

$$\mathbb{E}[M_0] = \mathbb{E}[M_\tau] = \mathbb{E}[M_s\mathbf{1}_{A^c}] + \mathbb{E}[M_t\mathbf{1}_A] = \mathbb{E}[M_s] + \mathbb{E}[(M_t - M_s)\mathbf{1}_A].$$

Since $\mathbb{E}[M_s] = \mathbb{E}[M_0]$, we have $\mathbb{E}[M_s\mathbf{1}_A] = \mathbb{E}[M_t\mathbf{1}_A]$, for all $A \in \mathcal{F}_s$, which, by definition, means that $M_s = \mathbb{E}[M_t|\mathcal{F}_s]$. □

Corollary 16.25 (Stopped martingales are martingales). *If $\{M_t\}_{t\in[0,\infty)}$ is a right-continuous martingale, then so is the stopped process $\{M_t^\tau\}_{t\in[0,\infty)}$, for each stopping time τ.*

Before we state a general optional sampling theorem, we give a criterion for uniform integrability for continuous-time martingales. The the proof is based on the same ideas as the corresponding discrete-time result (Proposition 12.12), so we omit it.

Theorem 16.26 (UI martingales). *Let $\{M_t\}_{t\in[0,\infty)}$ be a right-continuous martingale. The following are equivalent:*

1. *$\{M_t\}_{t\in[0,\infty)}$ is UI,*
2. *$\lim_{t\to\infty} M_t$ exists in $\mathbb{L}^1$, and*
3. *$\{M_t\}_{t\in[0,\infty)}$ has a last element, i.e., there exists a random variable M_∞ such that*
$$M_t = \mathbb{E}[M_\infty|\mathcal{F}_t], \text{ for all } t \in [0,\infty).$$

The version of the optional sampling theorem presented here

Theorem 16.27 (Optional Sampling and Convergence). *Let $\{X_t\}_{t\in[0,\infty)}$ be a right-continuous submartingale with a last element, i.e., such that there exists $X \in \mathbb{L}^1(\mathcal{F})$ with the property that*

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

Then,

1. *$X_\infty = \lim_{t\to\infty} X_t$ exists a.s., and $X_\infty \in \mathbb{L}^1$ is the a.s.-minimal last element for $\{X_t\}_{t\in[0,\infty)}$, and*

2. *$X_\tau \le \mathbb{E}[X_\infty|\mathcal{F}_\tau]$, for all $\tau \in \mathcal{S}$.*

Corollary 16.28 (Optional sampling for nonnegative supermartingales)**.** *Let $\{M_t\}_{t\in[0,\infty)}$ be a nonnegative right-continuous supermartingale. Then $\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \le X_\sigma$ for all stopping times τ, σ, with $\tau \ge \sigma$.*

In complete analogy with discrete time (again), the (sub)martingale property stays stable under stopping:

Proposition 16.29 (Stability under stopping)**.** *Let $\{M_t\}_{t\in[0,\infty)}$ be a right-continuous martingale (submartingale), and let τ be a stopping time. Then the stopped process $\{M_t^\tau\}_{t\in[0,\infty)}$ is also a right-continuous martingale (submartingale).*

Most of the discrete-time results about martingales transfer to the continuous-time setting, provided that right-continuity of the paths is assumed. Such a condition is not as restrictive as it may seem at first glance, thanks to the following result. It is important to note, though, that it relies heavily on our standing assumption of right continuity for the filtration $\mathbb{F}$. It also needs a bit of completeness; more precisely, we assume that each probability-zero set in $\mathcal{F}$ belongs already to $\mathcal{F}_0$. These two assumption as so important that they even have a name:

Definition 16.30 (Usual conditions)**.** A filtration is said to satisfy the **usual conditions** if it is right continuous and complete in the sense that $A \in \mathcal{F}_0$, as soon as $A \in \mathcal{F}$ and $\mathbb{P}[A] = 0$.

Theorem 16.31 (Regularization of submartingales)**.** *Under usual conditions, a submartingale $\{X_t\}_{t\in[0,\infty)}$ has a RCLL modification if and only if the mapping $t \mapsto \mathbb{E}[X_t]$ is right continuous. In particular, each martingale has a RCLL modification.*

Various submartingale inequalities can also be extended to the case of continuous time. In order not to run into measurability problems, we associate the **maximal process**

$$X_t^* = \sup\{|X_q| : q = t \text{ or } q \text{ is a rational in } [0,t)\},$$

with a process X. Note that when X is RCLL or LCRL, $X_t = \sup_{s\in[0,t]} |X_s|$, a.s.

Theorem 16.32 (Doob's and Maximal Inequalities)**.** *Let $\{X_t\}_{t\in[0,\infty)}$ be a*

RCLL process which is either a martingale or a positive submartingale. Then,

$$\mathbb{P}[X^* \geq M] \leq \frac{1}{M^p} \sup_{t\geq 0} \mathbb{E}[|X_t|^p], \textit{ for } M > 0 \textit{ and } p \geq 1, \textit{ and}$$

$$||X^*||_{\mathbb{L}^p} \leq \frac{p}{p-1} \sup_{t\geq 0} ||X_t||_{\mathbb{L}^p}, \textit{ for } p > 1.$$

Proof. The main idea of the proof is to approximate $\sup_{t\geq 0} |X_t|$ by the random variables of the form $\sup_{t\in Q_n} |X_t|$, where $\{Q_n\}_{n\in\mathbb{N}}$ is an increasing sequence of finite sets whose union $Q = \cup_n Q_n$ is dense in $[0,\infty)$. By the right-continuity of the paths, $\sup_{t\geq 0} |X_t| = \sup_{t\in Q} |X_t|$, where Q is any countable dense set in $[0,\infty)$. To finish the proof, we can use the discrete-time inequalities in the pre-limit, and the monotone convergence theorem to pass to the limit. □

Additional Problems

Problem 16.6 (Predictable and optional processes)**.** A stochastic process $\{X_t\}_{t\in[0,\infty)}$ is said to be

- **optional**, if it is measurable with respect to the σ-algebra $\mathcal{O}$, where $\mathcal{O}$ is the smallest σ-algebra on $[0,\infty)\times\Omega$ with respect to which all RCLL and adapted adapted processes are measurable.
- **predictable**, if it is measurable with respect to the σ-algebra $\mathcal{P}$, where $\mathcal{P}$ is the smallest σ-algebra on $[0,\infty)\times\Omega$ with respect to which all LCRL adapted processes are measurable.

Show that

1. The predictable σ-algebra coincides with the σ-algebra generated by all *continuous* and piecewise linear adapted processes.
2. Show that $\mathcal{P} \subseteq \mathcal{O} \subseteq \text{Prog} \subseteq \mathcal{B}([0,\infty)) \otimes \mathcal{F}$.

Note: There are counterexamples which show that none of the implications above are equivalences. Some are very simple, and the others are quite involved.

Problem 16.7 (The total-variation process)**.** For each process $\{X_t\}_{t\in[0,\infty)}$ of finite variation, we define its **total-variation process** $\{|X|_t\}_{t\in[0,\infty)}$ as the process whose value at t is the total variation on $[0,t]$ of the path of $\{X_t\}_{t\in[0,\infty)}$.

If $\{X_t\}_{t\in[0,\infty)}$ is an RCLL adapted process of finite variation, show that $\{|X|_t\}_{t\in[0,\infty)}$ is

1. RCLL, adapted and of finite variation, and
2. continuous if $\{X_t\}_{t\in[0,\infty)}$ is continuous.

Problem 16.8 (The Poisson Point Process)**.** Let $(S,\mathcal{S},\mu)$ be a measurable space, i.e., S is a non-empty set, $\mathcal{S}$ is a σ-algebra, and μ is a positive measure on S. A mapping

$$\mathcal{N} : \Omega \times \mathcal{S} \to \mathbb{N} \cup \{\infty\},$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, is called a **Poisson Point Process** (PPP) with **mean measure** μ if

- the mapping $\mathcal{N}(B)$ (more precisely, the mapping $\omega \mapsto \mathcal{N}(\omega, B)$) is $\mathcal{F}$-measurable (a random variable) for each $B \in \mathcal{S}$ and has the Poisson distribution[3] with parameter $\mu(B)$ (denoted by $P(\mu(B))$), whenever $\mu(B) < \infty$, and
- for each $\omega \in \Omega$, the mapping $B \mapsto N(\omega, B)$ is an $\mathbb{N} \cup \{\infty\}$-valued measure, and
- random variables $(\mathcal{N}(B_1), \mathcal{N}(B_2), \dots, \mathcal{N}(B_d))$ are independent when the sets $B_1, B_2, \dots, B_d$ are (pairwise) disjoint.

[3] A r.v. X is said to have the **Poisson** distribution with parameter $c > 0$, if $\mathbb{P}[X = n] = e^{-c}\frac{c^n}{n!}$, for $n \in \mathbb{N}_0$. When $c = 0$, $\mathbb{P}[X = 0] = 1$.

The purpose of this problem is to show that, under mild conditions on $(S, \mathcal{S}, \mu)$, a PPP with mean measure μ exists.

1. We assume first that $0 < \mu(S) < \infty$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space which supports a random variable N and an iid sequence $\{X_k\}_{k\in\mathbb{N}}$, independent of N and such that

   - $N \sim P(\mu(S))$, and
   - for each k, X_k takes values in S and $\mathbb{P}[X_k \in B] = \mu(B)/\mu(S)$, for all $B \in \mathcal{S}$ (technically, a measurable mapping from Ω into a measurable space is called a **random element**).

   Show that such a probability space exists.

2. For $B \in \mathcal{S}$, define

   $$\mathcal{N}(\omega, B) = \sum_{k=1}^{N(\omega)} \mathbf{1}_{\{X_k(\omega) \in B\}},$$

   i.e., $\mathcal{N}(\omega, B)$ is the number of terms in $X_1(\omega), \dots, X_{N(\omega)}(\omega)$ that fall into B. Show that $\mathcal{N}(B)$ is a random variable for each $B \in \mathcal{S}$.

3. Pick (pairwise) disjoint $B_1, \dots, B_d$ in $\mathcal{S}$ and compute

   $$\mathbb{P}[\mathcal{N}(B_1) = n_1, \dots, \mathcal{N}(B_d) = n_d | N = m], \text{ for } m, n_1, \dots, n_d \in \mathbb{N}_0.$$

4. Show that $\mathcal{N}$ is a PPP with mean measure μ.

5. Show that a PPP with mean measure μ exists when $(S, \mathcal{S}, \mu)$ is merely a σ-**finite** measure space, i.e., there exists a sequence $\{B_n\}_{n\in\mathbb{N}}$ in $\mathcal{S}$ such that $S = \cup_n B_n$ and $\mu(B_n) < \infty$, for all $n \in \mathbb{N}_0$.

Problem 16.9 (The Poisson Process)**.** A stochastic process $\{N_t\}_{t\in[0,\infty)}$ is called a **Poisson process with parameter** $c > 0$ if

- it has independent increments, i.e., $N_{t_1} - N_{s_1}$, $N_{t_2} - N_{s_2}$, $\dots$, $N_{t_d} - N_{s_d}$ are independent random variables when $s_1 \le t_1 \le s_2 \le t_2 \le \dots \le s_d \le t_d$,
- $N_t - N_s$ has the Poisson distribution with parameter $c(t-s)$, and
- all paths $t \mapsto N_t(\omega)$ are RCLL

Let $\mathcal{N} : \Omega \times \mathcal{B}([0,\infty)) \to \mathbb{N} \cup \{\infty\}$ be a PPP on $([0,\infty), \mathcal{B}([0,\infty)), c\lambda)$, where $c > 0$ is a constant and λ is the Lebesgue measure on $[0,\infty)$. Define the stochastic process $\{N_t\}_{t\in[0,\infty)}$ by $N_t = \mathcal{N}([0,t])$, $t \ge 0$.

1. Show that $\{N_t\}_{t\in[0,\infty)}$ is a Poisson process.
2. Show that processes $N_t - ct$ and $(N_t - ct)^2 - ct$ are RCLL $\mathcal{F}_t$-martingales, where $\mathcal{F}_t = \sigma(N_s, s \le t)$, $t \ge 0$.
3. Let $N^{(2)}$ be a PPP on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \lambda^2)$, where λ^2 stands for the 2-dimensional Lebesgue measure. Is the process $\{M_t\}_{t\in[0,\infty)}$ a Poisson process, where
$$M_t = N^{(2)}\left(\{(x,y) \in \mathbb{R}^2 : \sqrt{x^2+y^2} \le t\}\right), t \ge 0\ ?$$

Problem 16.10 (Supermartingale Coalescence)**.** Let $\{X_t\}_{t\in[0,\infty)}$ be a nonnegative RCLL supermartingale, and let

$$\tau = \inf\{t \ge 0 : X_t = 0\}$$

be the first hitting time of level 0. Show that

$$X_t = 0 \text{ a.s., on } \{\tau \le t\}, \text{ for all } t \ge 0.$$

In words, once a nonnegative RCLL supermartingale hits 0, it stays there. Show, by means of an example, that the statement does not hold when X is a nonnegative RCLL *submartingale*.

Problem 16.11 (Hardy's inequality)**.** Using $([0,1], \mathcal{B}([0,1]), \lambda)$ ($\lambda =$ the Lebesgue measure) as the probability space, let $\mathcal{F}_t$ be the smallest sub sigma-field of $\mathcal{F}$ containing the Borel subsets of $[0,t]$ and all negligible sets of $[0,1]$, for each $t \in [0,1]$.

1. For $f \in L^1([0,1], \lambda)$, provide an explicit expression for the right-continuous version of the martingale
$$X_t = \mathbb{E}[f|\mathcal{F}_t]$$

2. Apply Doob's maximal inequality to the above martingale to conclude that if $g(t) = \frac{1}{1-t}\int_t^1 f(u)\,du$ then

$$||g||_{\mathbb{L}^p} \leq \tfrac{p}{p-1}||f||_{\mathbb{L}^p},$$

if $p > 1$. This inequality is know as **Hardy's inequality**.

3. Inspired by the construction above, give an example of a uniformly integrable martingale $\{X_t\}_{t\in[0,\infty)}$ for which $X^*_\infty \notin \mathbb{L}^1$.

Problem 16.12 (Polynomials that turn Brownian motion into a martingale)**.** For $c \in \mathbb{R}$ we define the function $F^c : [0,\infty) \times \mathbb{R} \to \mathbb{R}_+$ by

$$F^c(t,x) = \exp(cx - \tfrac{1}{2}c^2 t), \text{ for } (t,x) \in [0,\infty) \times \mathbb{R}.$$

1. For $n \in \mathbb{N}_0$, $(t,x) \in [0,\infty) \times \mathbb{R}$, define

$$F^{(n,c)}(t,x) = \tfrac{\partial^n}{(\partial c)^n} F^c(t,x),$$

where, by convention, $\frac{\partial^0}{(\partial c)^0} G = G$. Show that $P^{(n)}(t,x) = F^{(n,0)}(t,x)$ is a polynomial in t and x for each $n \in \mathbb{N}_0$, and write down expressions for $P^{(n)}(t,x)$, for $n = 0,1,2,3$.

2. Show that the process $\{Y^{(n)}_t\}_{t\in[0,\infty)}$, given by $Y^{(n)}_t = P^{(n)}(t,B_t)$, $t \in [0,\infty)$, is an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-martingale for each $n \in \mathbb{N}_0$.

Problem 16.13 (Hitting and exit from a strip for a Brownian motion)**.** Let $(B_t)_{0\leq t<\infty}$ be a standard Brownian motion. Set $\tau_x = \inf\{t \geq 0 : B_t = x\}$, for $x \in \mathbb{R}$, and compute

1. $\mathbb{P}[\tau_a < \tau_{-b}]$ and $\mathbb{E}[\tau_a \wedge \tau_{-b}]$ for $a,b > 0$.

2. $\mathbb{E}[e^{-\lambda\tau_x}]$, for $\lambda > 0$ and $x \in \mathbb{R}$.

Hint: Apply the optional sampling theorem to appopriately-chosen martingales.

Problem 16.14 (The maximum of a martingale that converges to 0)**.** Let $\{M_t\}_{t\in[0,\infty)}$ be a nonnegative continuous martingale with $M_0 = 1$ and $M_\infty = \lim_{t\to\infty} M_t = 0$, a.s. Find the distribution of the random variable $M^*_\infty = \sup_{t\geq 0} M_t$. *Hint:* For $x > 1$, set $\tau_x = \inf\{t \geq 0 : M^*_t \geq x\} = \inf\{t \geq 0 : M_t \geq x\}$, and note that $M^{\tau_x}_\infty \in \{0,x\}$, a.s.

Problem 16.15 (Convergence of paths of RCLL martingales)**.** Consider a sequence $\{M^n_t\}_{t\in[0,T]}$, $n \in \mathbb{N}$, of martingales with continuous paths (defined on the same filtered probability space). Suppose that $\{M^\infty_t\}_{t\in[0,T]}$ is an RCLL martingale such that $M^n_T \to M^\infty_T$ in $\mathcal{L}^1$, as $n \to \infty$. Show that M^∞ has continuous paths, a.s.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 17
Brownian motion as a Markov process

Brownian motion is one of the "universal" examples in probability. So far, it featured as a continuous version of the simple random walk and served as an example of a continuous-time martingale. It can also be considered as one of the fundamental Markov processes. We start by explaining what that means.

The Strong Markov Property of the Brownian Motion

Definition 17.1 (Markov property)**.** A stochastic process $\{X_t\}_{t\in[0,\infty)}$, defined on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, \mathcal{P})$ is said to be an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-**Markov process** if, for all $B \in \mathcal{B}(\mathbb{R})$, and $t, h \geq 0$, we have

$$\mathbb{P}[X_{t+h} \in B|\mathcal{F}_t] = \mathbb{P}[X_{t+h} \in B|\sigma(X_t)], \text{ a.s.}$$

Example 17.2 (Brownian motion is a Markov Process)**.** Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let $\mathcal{F}^B_t = \sigma(B_s, s \leq t)$ be its natural filtration. The independence of increments implies that $B_{t+h} - B_t$ is independent of $\mathcal{F}^B_t$, for $t, h \geq 0$. Therefore, for a bounded Borel function $f : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}[f(B_{t+h})|\mathcal{F}^B_t] = \mathbb{E}[f\big(B_t + (B_{t+h} - B_t)\big)|\mathcal{F}^B_t] = \hat{f}(B_t), \text{ a.s.}$$

where $\hat{f}(x) = \mathbb{E}[f(x + B_{t+h} - B_t)]$. In particular, we have

$$\mathbb{E}[f(B_{t+h})|\mathcal{F}^B_t] = \mathbb{E}[f(B_{t+h})|\sigma(B_t)], \text{ a.s.,}$$

and, by setting $f = \mathbf{1}_A$, we conclude that B is an $\mathcal{F}^B$-Markov process.

A similar statement, only when the deterministic time t is replaced by a stopping time τ is typically referred to as the **strong Markov property**. While the main goal of the section is to state and prove it, we make a quick detour and introduce another important notion, a slight generalization of what it means to be a Brownian motion.

Definition 17.3 ($\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian motion)**.** Let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be a filtration. An $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-adapted process $\{B_t\}_{t\in[0,\infty)}$ is said to be an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-**Brownian motion**, if

1. $B_t - B_s \sim N(0, t-s)$, for $0 \le s \le t < \infty$,
2. $B_t - B_s$ is independent of $\mathcal{F}_s$, for all $0 \le s \le t < \infty$, and
3. for all $\omega \in \Omega$, $t \mapsto B_t(\omega)$ is a continuous functions.

The following proposition gives an (at first glance unexpected) characterization of the $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian property. It is a special case of a very common theme in stochastics. It features a complex-valued martingale; that simply means that both its real and imaginary parts are martingales.

Proposition 17.4 (A characterization of the $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian motion)**.** *An $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-adapted process $\{X_t\}_{t\in[0,\infty)}$ with continuous paths is an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian motion if and only if the complex-valued process $\{Y^r_t\}_{t\in[0,\infty)}$, given by*

$$Y^r_t = e^{irX_t + \frac{1}{2}r^2 t} \text{ for } t \ge 0,\ r \in \mathbb{R}, \tag{17.1}$$

is an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-martingale, for each $r \in \mathbb{R}$. In particular, an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian motion is an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Markov process and also an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-martingale.

Proof. The martingality of Y^r (for each r) implies that the conditional characteristic function of the increment $X_{t+h} - X_t$, given $\mathcal{F}_t$, is centered normal with variance h, for all $t, h \ge 0$. By Proposition 10.18, $X_{t+h} - X_t$ is independent of $\mathcal{F}_t$, and we conclude that X is an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian motion.

The Markov property follows from the fact that the conditional distribution of X_{t+h}, given $\mathcal{F}_t$, is normal with mean X_t (and variance h), so it only depends on X_t. The martingale property, similarly, is the consequence of the fact that $X_{t+h} - X_h$ is centered, conditionally on $\mathcal{F}_t$. □

We know already that each Brownian motion is an $\{\mathcal{F}^B_t\}_{t\in[0,\infty)}$-Brownian motion. There are other filtrations, though, that share this property. A less interesting (but quite important) example is the natural filtration of a d-dimensional Brownian motion[1], for $d > 1$. Then, each of the components is an $\{\mathcal{F}^{(B^1,\dots,B^d)}\}_{t\in[0,\infty)}$-Brownian motion. A more unexpected example is the following:

[1] a d-dimensional Brownian motion $(B^1, \dots, B^d)$ is simply a process, taking values in $\mathbb{R}^d$, each of whose components is a Brownian motion in its own right, independent of all the others.

Proposition 17.5 (Brownian motion is an $\{\mathcal{F}^{B+}_t\}_{t\in[0,\infty)}$-Brownian motion)**.** *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let $\{\mathcal{F}^{B+}_t\}_{t\in[0,\infty)}$ be the right-continuous augmentation of its natural filtration $\{\mathcal{F}^B_t\}_{t\in[0,\infty)}$. Then $\{B_t\}_{t\in[0,\infty)}$ is an $\{\mathcal{F}^{B+}_t\}_{t\in[0,\infty)}$-Brownian motion, and an $\{\mathcal{F}^{B+}_t\}_{t\in[0,\infty)}$-Markov process.*

Proof. Thanks to Proposition 17.4, it suffices to show that (17.1) holds with $\mathcal{F}_t$ replaced by $\mathcal{F}_t^{B+}$. For that, we start with the fact that, for $\varepsilon < h$,

$$\mathbb{E}[e^{irX_{t+h}}|\mathcal{F}^B_{t+\varepsilon}] = e^{irX_{t+\varepsilon}-\frac{1}{2}r^2(h-\varepsilon)}, \text{ a.s., for all } r \in \mathbb{R}.$$

We condition both sides with respect to $\mathcal{F}^B_{t+}$ and use the tower property to conclude that

$$\mathbb{E}[e^{irX_{t+h}}|\mathcal{F}^B_{t+}] = \mathbb{E}[e^{irX_{t+\varepsilon}-\frac{1}{2}r^2(h-\varepsilon)}|\mathcal{F}^B_{t+}], \text{ a.s., for all } r \in \mathbb{R}.$$

We let $\varepsilon \searrow 0$ and use the dominated convergence theorem and the right continuity of X to get

$$\mathbb{E}[e^{irX_{t+h}}|\mathcal{F}^B_{t+}] = \mathbb{E}[e^{irX_t-\frac{1}{2}r^2h}|\mathcal{F}^B_{t+}] = e^{irX_t-\frac{1}{2}r^2h}, \text{ a.s., for all } r \in \mathbb{R}. \quad \square$$

Corollary 17.6 (Blumenthal's 0-1 law). *For $t \geq 0$, the σ-algebras $\mathcal{F}_t^{B+}$ and $\mathcal{F}_t^B$ are a.s.-equal, i.e.,*

for each $A \in \mathcal{F}_t^{B+}$ there exists $A' \in \mathcal{F}_t^B$ such that $\mathbb{P}[A\Delta A'] = 0$.

Proof. Thanks to the $\{\mathcal{F}_t^{B+}\}_{t\in[0,\infty)}$-Brownian property of the Brownian motion, given $t \geq 0$ and a nonnegative Borel function f, we have

$$\begin{aligned}\mathbb{E}[f(B_{s_1}, B_{s_2}, \dots, B_{s_n})|\mathcal{F}^B_{t+}] &= \mathbb{E}[f(B_{s_1}, B_{s_2}, \dots, B_{s_n})|\sigma(B_t)] \\ &= \mathbb{E}[f(B_{s_1}, B_{s_2}, \dots, B_{s_n})|\mathcal{F}^B_t], \text{ a.s.,}\end{aligned} \tag{17.2}$$

for all $s_n > s_{n-1} > \dots, s_1 > t$. Trivially,

$$\mathbb{E}[\mathbf{1}_A|\mathcal{F}^B_{t+}] = \mathbb{E}[\mathbf{1}_A|\mathcal{F}^B_t], \text{ a.s., for all } A \in \mathcal{F}_t. \tag{17.3}$$

Let $\mathcal{A}$ denote the set of all events A in $\mathcal{F}_\infty = \vee_{t\geq 0}\mathcal{F}^B_t$ such that $\mathbb{E}[\mathbf{1}_A|\mathcal{F}^B_t] = \mathbb{E}[\mathbf{1}_A|\mathcal{F}^B_{t+}]$, a.s.. By (17.2) and (17.3), $\mathcal{A}$ contains all sets of the form $A_1 \cap A_2$, for $A \in \mathcal{F}_t$ and $A_2 \in \sigma(B_s, s \geq t)$. The $\pi - \lambda$-theorem can now be used to conclude that $\mathbb{E}[\mathbf{1}_A|\mathcal{F}^B_{t+}] = \mathbb{E}[\mathbf{1}_A|\mathcal{F}^B_t]$, for *all* $A \in\ \mathcal{F}_\infty$, and, in particular, for all $A \in \mathcal{F}^B_{t+}$, i.e., that

$$\mathbf{1}_A = \mathbf{1}_{A'} \text{ a.s., where } \mathbf{1}_{A'} = \mathbb{E}[\mathbf{1}_A|\mathcal{F}^B_t]. \quad \square$$

Corollary 17.7. *With probability 1, the Brownian path takes both strictly positive and strictly negative values in each neighborhood of 0.*

Proof. Suppose, to the contrary, that $\mathbb{P}[A] > 0$, where

$$A = \{\exists\, \varepsilon > 0,\ B_t \geq 0 \text{ for } t \in [0, \varepsilon]\}.$$

Check that $A \in \mathcal{F}_0^{B+}$, so that $\mathbb{P}[A] = 1$. Since $\{-B_t\}_{t\in[0,\infty)}$ is also a Brownian motion, we have $\mathbb{P}[B] = 1$, where

$$B = \{\exists\, \varepsilon > 0,\ B_t \leq 0 \text{ for } t \in [0, \varepsilon]\}.$$

Consequently,

$$\mathbb{P}[\{\exists\, \varepsilon > 0,\ B_t = 0 \text{ for } t \in [0,\varepsilon]\} = 1,$$

and so, there exists $\varepsilon_0 > 0$ such that

$$\mathbb{P}[\{B_t = 0 \text{ for } t \in [0,\varepsilon_0]\} > 0.$$

In particular, $\mathbb{P}[B_{\varepsilon_0} = 0] > 0$, which is in contradiction with the fact that B_{ε_0} is normally distributed with variance $\varepsilon_0 > 0$.

We conclude that $\mathbb{P}[A] = 0$, i.e., with probability 1, for each $\varepsilon > 0$ there exists $t \in [0,\varepsilon)$ such that $B_t < 0$. Similarly, with probability 1, for each $\varepsilon > 0$ there exists $t \in [0,\varepsilon)$ such that $B_t > 0$, and the claim of the proposition follows. $\square$

Proposition 17.8 (The strong Markov property of the Brownian motion)**.** *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion and let $\tau < \infty$, a.s., be an $\{\mathcal{F}^{B+}_t\}_{t\in[0,\infty)}$-stopping time. Then the process $\{W_t\}_{t\in[0,\infty)}$, given by*

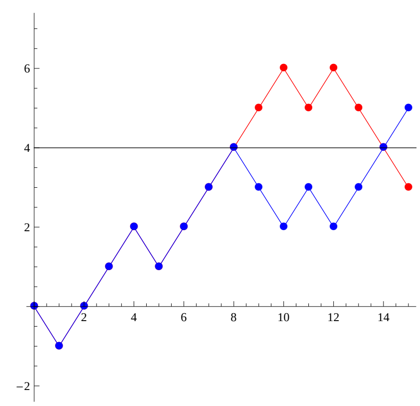
$$W_t = B_{\tau+t} - B_\tau,$$

is a Brownian motion, independent of $\mathcal{F}^{B+}_\tau$.

Proof. We do the proof only in the case of a bounded τ. We'll come back to the general case a bit later. Thanks to the characterization in Proposition 17.4, we need to show that the exponential process $\{e^{irX_{\tau+t}+\frac{1}{2}r^2(\tau+t)}\}_{t\in[0,\infty)}$ is a martingale with respect to $\{\mathcal{F}_{\tau+t}\}_{t\in[0,\infty)}$, for all $r \in \mathbb{R}$. Thanks to boundedness of τ, this is a direct consequence of the optional sampling theorem (specifically, Proposition 16.23). $\square$

The reflection principle

Before we deal with the continuous time, search for "reflection principle" and "ballot problem" for a neat simple idea in enumerative combinatorics. We are going to apply it in continuous time to the Brownian motion to say something about the distribution of its running maximum. As a preparation for the mathematics, do the following problem first:



The idea behing the reflection principle.

Problem 17.1. Let $\{X_t\}_{t\in[0,\infty)}$ be a stochastic process with continuous trajectories, and let τ be an $[0,\infty]$-valued random variable. Define the mapping $\kappa : \Omega \to C[0,\infty)$ by

$$\kappa(\omega) = \{X_{t\wedge\tau(\omega)}(\omega)\}_{t\in[0,\infty)}, \text{ for } \omega \in \Omega.$$

Show that κ is measurable from $(\Omega,\mathcal{F})$ into $(C[0,\infty),\mathcal{B}(C[0,\infty)))$.

Proposition 17.9 (Reflection principle). *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let τ be a stopping time with $\tau(\omega)<\infty$, for all $\omega\in\Omega$. Then the process $\{\tilde{B}\}_{t\in[0,\infty)}$, given by*

$$\tilde{B}_t(\omega)=\begin{cases} B_t, & t\leq\tau,\\ B_\tau-(B_t-B_\tau), & t>\tau,\end{cases} \tag{17.4}$$

is also a Brownian motion.

Proof. Consider the mapping $K:\Omega\to[0,\infty)\times C[0,\infty)\times C[0,\infty)$, given by

$$K(\omega)=\Big(\tau(\omega),\{B_{t\wedge\tau(\omega)}(\omega)\}_{t\in[0,\infty)},\{B_{\tau(\omega)+t}(\omega)-B_{\tau(\omega)}(\omega)\}_{t\in[0,\infty)}\Big).$$

It follows from Problem 17.1 (or a very similar argument) that K is $\mathcal{F}$–$\mathcal{B}([0,\infty))\times\mathcal{B}(C[0,\infty))\times\mathcal{B}(C[0,\infty))$-measurable, so it induces a measure $\mathbb{P}_K$ - a pushforward of $\mathbb{P}$ - on $[0,\infty)\times C[0,\infty)\times C[0,\infty)$. Moreover, by the strong Markov property of the Brownian motion, the process $\{B_{\tau+t}-B_\tau\}_{t\in[0,\infty)}$ is a Brownian motion, independent of $\mathcal{F}^{B+}_\tau$. Since both τ and $\{B_{\tau\wedge t}\}_{t\in[0,\infty)}$ are measurable with respect to $\mathcal{F}^{B+}_\tau$, the first two components $\tau(\omega),\{B_{t\wedge\tau(\omega)}(\omega)\}_{t\in[0,\infty)}$ of K are independent of the third one. Consequently, the measure $\mathbb{P}_K$ is a product of a probability measure $\mathbb{P}'$ on $[0,\infty)\times C[0,\infty)$ and a probability measure $\mathbb{P}_W$ on $C[0,\infty)$, where, as the notation suggests, the measure $\mathbb{P}_W$ is a Wiener measure on $C[0,\infty)$. The same argument shows that the pushforward of $\mathbb{P}$ under the mapping $L:\Omega\to[0,\infty)\times C[0,\infty)\times C[0,\infty)$, given by

$$L(\omega)=\Big(\tau(\omega),\{B_{t\wedge\tau(\omega)}(\omega)\}_{t\in[0,\infty)},\{-(B_{\tau(\omega)+t}(\omega)-B_{\tau(\omega)}(\omega))\}_{t\in[0,\infty)}\Big),$$

is also equal to $\mathbb{P}_K=\mathbb{P}'\times\mathbb{P}_W$, because $\{-W_t\}_{t\in[0,\infty)}$ is a Brownian motion, whenever $\{W_t\}_{t\in[0,\infty)}$ is one.

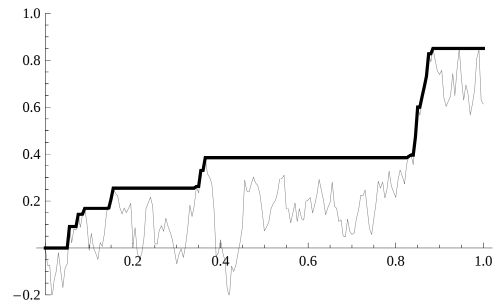
Define the mapping $S:[0,\infty)\times C[0,\infty)\times C[0,\infty)\to C[0,\infty)$, by $S(\tau,x_1,x_2)=x$, where

$$x(t)=\begin{cases} x_1(t), & t\leq\tau\\ x_1(\tau)+x_2(t), & t>\tau.\end{cases}$$

It is left to the reader to prove that S is continuous, and, therefore, measurable. Therefore, the compositions $S\circ K$ and $S\circ L$, both defined on Ω with values in $C[0,\infty)$ are measurable, and the pushforwards of $\mathbb{P}$ under them are the same (beacuse both K and L push $\mathbb{P}$ forward into the same measure on $\mathbb{R}\times C[0,\infty)\times C[0,\infty)$). The composition $S\circ L$ is easy to describe; in fact, we have

$$S(L(\omega))_t=B_t(\omega),$$

so the pushforward of $\mathbb{P}$ under $S \circ L$ equals to the Wiener measure $\mathbb{P}_W$ on $C[0,\infty)$. Therefore, the pushforward of $\mathbb{P}$ under $S \circ K$ is also a Wiener measure. In other words the process $S \circ K$ - which the reader will readily identify with $\{\tilde{B}_t\}_{t\in[0,\infty)}$ - is a Brownian motion. □



Definition 17.10. Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion. The process $\{M_t\}_{t\in[0,\infty)}$, where

$$M_t = \sup_{s\leq t} B_s, \ t \geq 0$$

is called the **running maximum** of the Brownian motion $\{B_t\}_{t\in[0,\infty)}$.

Proposition 17.11. *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let $\{M_t\}_{t\in[0,\infty)}$ be its running-maximum process. Then the random vector (M_1, B_1) is absolutely continuous with the density*

$$f_{(M_1,B_1)}(m,b) = 2\varphi'(2m-b)\mathbf{1}_{\{m>0,b<m\}},$$

where φ is the density of the unit normal.

Proof. By scaling we can assume that $t = 1$. We pick a level $m > 0$, and let $\tau_m = \inf\{s \geq 0 : B_s = m\}$ be the first hitting time of the level m. In order to be able to work with a finite stopping time, we set $\tau = \tau_m \wedge 1$, so that $\tau(\omega) < \infty$, for all $\omega \in \Omega$. Proposition 17.9 implies that the process $\{\tilde{B}_t\}_{t\in[0,\infty)}$, given by (17.4), is a Brownian motion. Therefore, the random vectors (B_1, M_1) and $(\tilde{B}_1, \tilde{M}_1)$, where $\tilde{M}_t = \sup_{s\leq t} \tilde{B}_s$, $t \geq 0$, are equally distributed. Moreover, $\{\tilde{M}_1 \geq m\} = \{M_1 \geq m\}$, so for $b < m$, $m > 0$ we have

$$\begin{aligned} \mathbb{P}[M_1 \geq m, B_1 \leq b] &= \mathbb{P}[\tilde{M}_1 \geq m, \tilde{B}_1 \geq 2m-b] \\ &= \mathbb{P}[M_1 \geq m, B_1 \geq 2m-b], \end{aligned}$$

i.e., since $\{M_1 \geq m\} = \{\tau_m \leq 1\}$ and $\{M_1 \geq m\} \supset \{B_1 \geq 2m-b\}$,

$$\mathbb{P}[M_1 \geq m, B_1 \leq b] = \mathbb{P}[B_1 \geq 2m-b]. \tag{17.5}$$

Therefore, the left-hand side is continuously differentiable in both m and b in the region $\{(m,b) \in \mathbb{R}^2 : m > 0, m < a\}$ and

$$\begin{aligned} f_{(M_1,B_1)}(m,b) &= \tfrac{\partial^2}{\partial m\, \partial b}\mathbb{P}[M_1 \geq m, B_1 < b] \\ &= \tfrac{\partial^2}{\partial m\, \partial b}\mathbb{P}[B_1 \geq 2m-b] = -2\varphi'(2m-b). \end{aligned}$$ □

Corollary 17.12. *The three random variables $M_1, M_1 - B_1$ and $|B_1|$ (defined in Proposition 17.11 above) have the same distribution.*

Proof. By (17.5), for $m > 0$, we have

$$\mathbb{P}[M_1 \geq m] = \mathbb{P}[M_1 \geq m, B_1 > m] + \mathbb{P}[M_1 \geq m, B_1 \leq m]$$
$$= \mathbb{P}[B_1 > m] + \mathbb{P}[B_1 \geq 2m - m] = \mathbb{P}[|B_1| \geq m].$$

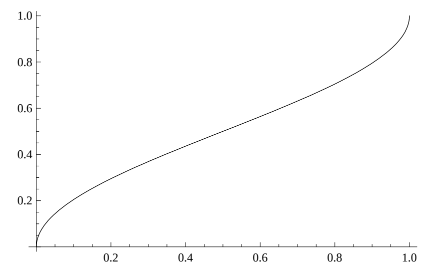
To obtain the second equality in distribution, we use the joint density function: for $a > 0$,

$$\mathbb{P}[B_1 \leq M_1 - a] = \int_0^\infty \int_{-\infty}^{m-a} -2\varphi'(2m - b)\, db\, dm =$$
$$= \int_0^\infty 2\varphi(m + a)\, dm = 2\int_a^\infty \varphi(m)\, dm = \mathbb{P}[|B_1| \geq a]. \square$$

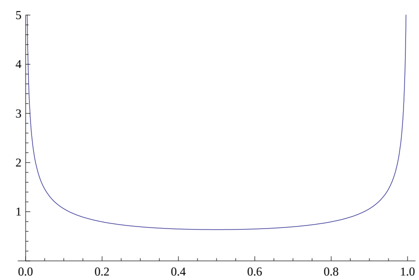
Arcsine laws

A random variable X is said to have the **arcsine distribution** if it is supported on $[0,1]$ with the cdf $F(x) = \frac{2}{\pi} \arcsin \sqrt{x}$, for $x \in [0,1]$. Its density computes to:

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}},\ x \in (0,1).$$



The distribution function of the arcsine distribution



The density function of the arcsine distribution

Problem 17.2. Show that

1. the random variable $X = \sin^2(\frac{\pi}{2}\alpha)$ has the arcsine distribution, if α is uniformly distributed on $[0,1]$.
2. the random variable $\frac{\xi^2}{\xi^2+\eta^2}$ has the arcsine distribution, if ξ and η are independent unit ($\mu = 0$, $\sigma = 1$) normals.

Proposition 17.13 (Arcsine laws)**.** *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let $\{M_t\}_{t\in[0,\infty)}$ be its running maximum. Then the following random variables both have the arcsine distribution:*

$$\tau_1 = \sup\{t \leq 1 : B_t = 0\} \text{ and } \tau_2 = \inf\{t \geq 0 : B_t = M_1\}.$$

Proof. We start with τ_1 and note that $\{\tau_1 < t\} = \{\sup_{t\leq s\leq 1} B_s < 0\} \cup \{\inf_{t\leq s\leq 1} B_s > 0\}$ and that the two sets are disjoint. Then, Proposition 17.11 and the independence of $B_1 - B_t$ and B_t imply that random vectors $(\sup_{t\leq s\leq 1} B_s - B_t, B_t)$ and $(|B_1 - B_t|, B_t)$ have the same joint distribution. Therefore, for $t \in (0,1]$, we have

$$\begin{aligned}\mathbb{P}[\tau_1 < t] &= \mathbb{P}\left[\sup_{t\leq s\leq 1} B_s < 0\right] + \mathbb{P}\left[\inf_{t\leq s\leq 1} B_s > 0\right] \\ &= 2\mathbb{P}\left[\sup_{t\leq s\leq 1} B_s < 0\right] = 2\mathbb{P}\left[\sup_{t\leq s\leq 1}(B_s - B_t) < -B_t\right] \\ &= 2\mathbb{P}[|B_1 - B_t| < -B_t] = \mathbb{P}[|B_1 - B_t| < |B_t|] \\ &= \mathbb{P}\left[(B_1 - B_t)^2 < B_t^2\right] = \mathbb{P}\left[(1-t)\eta^2 < t\xi^2\right] \\ &= \mathbb{P}\left[\tfrac{\xi^2}{\xi^2+\eta^2} < t\right],\end{aligned} \tag{17.6}$$

where $\xi = B_t/\sqrt{t}$ and $\eta = (B_1 - B_t)/\sqrt{1-t}$ are independent unit normals. Problem 17.2 implies that the τ_1, indeed, has the arcsine distribution. Moving on to τ_2, we note that

$$\begin{aligned}\mathbb{P}[\tau_2 \leq t] &= \mathbb{P}\left[\sup_{s\leq t}(B_s - B_t) \geq \sup_{s\geq t}(B_s - B_t)\right] \\ &= \mathbb{P}[|B_t| \geq |B_1 - B_t|] = \mathbb{P}\left[t\xi^2 \geq (1-t)\eta^2\right],\end{aligned} \tag{17.7}$$

and the conclusion follows just like above. □

The zero-set of the Brownian motion

Our final result show that the random set $\{t \geq 0 : B_t = 0\}$ - called the **zero set** of the Brownian motion - looks very much like the Cantor set.

Definition 17.14. A nonempty set $C \subseteq \mathbb{R}$ is said to be **perfect** if it is closed and has no isolated points, i.e., for each $x \in C$ and each $\varepsilon > 0$ there exists $y \in C$ such that $0 < |y - x| < \varepsilon$.

Problem 17.3. Show that perfect sets are uncountable. *Hint:* Assume that a perfect set C is countable, and construct a nested sequence of intervals which "miss" more and more of the elements of C.

Proposition 17.15 (The zero set of the Brownian path)**.** *For $\omega \in \Omega$, let $\mathcal{Z}(\omega) = \{t \geq 0 : B_t(\omega) = 0\}$, where $\{B_t\}_{t\in[0,\infty)}$ is a Brownian motion. Then, for almost all $\omega \in \Omega$,*

1. *$\mathcal{Z}(\omega)$ is perfect,*
2. *$\mathcal{Z}(\omega)$ is uncountable,*
3. *$\mathcal{Z}(\omega)$ is unbounded,*
4. *$\lambda(\mathcal{Z}(\omega)) = 0$, where λ is the Lebesgue measure on $[0,\infty)$.*

Proof.

1. $\mathcal{Z}(\omega)$ is closed because it is a level set of a continuous function. To show that it is perfect, we pick $M > 0$ and for a rational number $q \geq 0$ define the finite stopping time

$$\tau_q = \inf\{t \geq q : B_t = 0\} \wedge M.$$

   By the Strong Markov Property, $W_t = B_{\tau_q+t} - B_{\tau_q}$, $t \geq 0$ is a Brownian motion. In particular, there exists a set $A_q \in \mathcal{F}$ with $\mathbb{P}[A_q] = 1$ such that $W(\omega)_t$ takes the value 0 in any (right) neighborhood of 0. We pick $\omega \in A = \cap_q A_q$ and choose a zero t_0 of $B_t(\omega)$ in $[0, M)$ such that either $t_0 = 0$ or t_0 is isolated from the left. Then $t_0 = \tau_q(\omega)$ for q smaller than and close enough to t_0. In particular, $B_{\tau_q(\omega)}(\omega) = 0$ and, since $\omega \in A$, t_0 is not isolated in $\mathcal{Z}(\omega)$ from the right. Consequently, no $t \in \mathcal{Z}(\omega) \cap [0, M)$ can be isolated both from the left and from the right. The statement now follows from the fact that M is arbitrary.

2. Follows directly from (1) and Problem 17.3.

3. It is enough to note that Proposition 15.3 and continuity of the trajectories imply that a path of a Brownian motion will keep changing sign as $t \to \infty$.

4. Define $Z_t = \mathbf{1}_{\{B_t=0\}}$, for $t \geq 0$. Since the map $(t, \omega) \mapsto B_t(\omega)$ is jointly measurable, so is the map $(t, \omega) \mapsto Z_t(\omega)$. By Fubini's theorem, we have

$$\mathbb{E}[\lambda(\mathcal{Z})] = \mathbb{E}[\int_0^\infty Z_t \, dt] = \int_0^\infty \mathbb{E}[Z_t] \, dt = 0,$$

   since $\mathbb{E}[Z_t] = \mathbb{P}[B_t = 0] = 0$, for all $t \geq 0$. Thus $\lambda(\mathcal{Z}) = 0$, a.s. □

Additional Problems

Problem 17.4 (The family of Brownian passage times)**.** Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let the stopping times τ_c, $c \geq 0$, be defined as

$$\tau_c = \inf\{t \geq 0 : B_t > c\}.$$

Note: τ_c is also known as the (first) **passage time** of the level c.

1. Show that there exists a function $f : (0, \infty) \to (0, \infty)$ such that the random variables τ_c and $f(c)\tau_1$ have the same distribution.

2. We have derived the expression

$$\mathbb{E}[e^{-\lambda\tau_c}] = e^{-c\sqrt{\lambda}}, \text{ for } c, \lambda > 0,$$

for the Laplace transform of the distribution of τ_c in one of previous problems. Show how one can derive the density of τ_c by using the Brownian running-maximum process $\{M_t\}_{t\in[0,\infty)}$. *Note:* The distributions of τ_c form a family of so-called **Lévy distribution**, and provide one of the few examples of *stable* distributions with an explicit density function. These distributions are also special cases of the wider family of *inverse-Gaussian* distributions.

3. Find all $\alpha > 0$ such that $\mathbb{E}[\tau_c^\alpha] < \infty$. *Hint:* If you are going to use the density from 2. above, compute $\lim_{t\to\infty} \sqrt{t}\,\mathbb{P}[\tau_c > t]$.

4. Show that the process $(\tau_c)_{c\geq 0}$ has independent increments (just like in the definition of the BM) and that the increments over the intervals of the same size have the same distribution (again, just like the BM). Conclude that the distributions of τ_c, $c \geq 0$ form a **convolution semigroup**, i.e., that

$$\mu_{c_1} * \mu_{c_2} = \mu_{c_1+c_2} \text{ for } c_1, c_2 \geq 0,$$

where μ_c denotes the distribution of τ_c and $*$ denotes the convolution of measures. *Hint:* Use the Strong Markov Property or perform a tedious computation using the explicit expression from 2. above.

5. Show that the trajectories of $(\tau_c)_{c\geq 0}$ are a.s. RCLL (like the BM), but *a.s. not left continuous* (unlike the BM). *Hint:* What happens at levels c which correspond to the values of the local maxima of the Brownian path?

Note: $\{\tau_c\}_{c\geq 0}$ is an example of a *Lévy process*, or, more precisely, a *subordinator*.

Problem 17.5 (The planar Brownian motion does not hit points)**.** Let B be a two-dimensional Brownian motion (the coordinate processes are independent Brownian motions), and let $X_t = (0, a_0) + B_t$, for $t \geq 0$ and $a_0 \in \mathbb{R} \setminus \{0\}$. The process X is called the **planar Brownian motion** started at $(0, a_0)$.

The purpose of this problem is to prove that X will never hit $(0, 0)$ (with probability 1). In fact, by scaling and rotational invariance, this means that, given any point $(x, y) \neq (0, a_0)$ the planar Brownian motion will never hit (x, y) (with probability 1). In the language of stochastic analysis, we say that points are **polar** for the planar Brownian motion.

1. How can this be, given that each trajectory of X hits at least one point other than $(0, a_0)$? In fact, it hits uncountably many of them.

2. Show that the planar Brownian motion is rotationally invariant. More precisely, if $U : \mathbb{R}^2 \to \mathbb{R}^2$ is an orthogonal linear transformation, then the process $X'_t = UX_t$, $t \geq 0$, is also a planar Brownian motion (started at $U(0, a_0)$.)

3. We start by defining the stopping time

$$T = \inf\{t \geq 0 \,:\, X_t = (0,0)\},$$

so that what we need to prove is that $\mathbb{P}[T < \infty] = 0$. Before it gets the chance to hit $(0,0)$, X will have to hit the x-axis so that $\mathbb{P}[T \geq T_1] = 1$, where

$$T_1 = \inf\left\{t \geq 0 \,:\, X^{(2)}_t = 0\right\},$$

where $X = (X^{(1)}, X^{(2)})$. Let $a_1 = X^{(1)}_{T_1}$ be the x-coordinate of the position at which the x axis is first hit. Identify the distribution of a_1, and conclude that $\mathbb{P}[a_1 = 0] = 0$. *Hint:* Use the formula for the Laplace transform of T_1 and the fact that T_1 is independent of $\{X^{(1)}_t\}_{t\in[0,\infty)}$ (by the Strong Markov Property); also characteristic functions.

4. It follows that $T_1 < \infty$, a.s., and the point at which the process hits the x axis is different from $(0,0)$, with probability 1. Hence, $\mathbb{P}[T_2 > T_1] = 1$, where

$$T_2 = \inf\left\{t \geq T_1 \,:\, X^{(1)}_t = 0\right\}.$$

the first time after T_1 that the y axis is hit. Moreover, it is clear that $T \geq T_2$, a.s. Let $a_2 = X^{(2)}_{T_2}$ be the y-coordinate of the position of the process at time T_2.

The behavior of X from T_1 to T_2 has the same conditional distribution (given $\mathcal{F}_{T_1}$) as the planar Brownian motion started from $(a_1, 0)$ (instead of $(0, a_0)$). Use this idea to compute the conditional distributions of T_2 and a_2, given $\mathcal{F}_{T_1}$, and show that $\zeta_2 = (T_2 - T_1)/a_1^2$ and $\gamma_2 = a_2/a_1$ are independent of $\mathcal{F}_{T_1}$. *Hint:* Strong Markov Property and rotational invariance.

5. Continue the procedure outlined above (alternating the x and y axes) to define the stopping times $T_3, T_4, \dots$ and note that $T \geq T_n$, for each $n \in \mathbb{N}$. Consequently, our main statement will follow if we show that

$$\mathbb{P}[\sum_{n=1}^{\infty} \tau_n = +\infty] = 1,$$

where, $\tau_1 = T_1$ and $\tau_n = T_n - T_{n-1}$, for $n > 1$. Similarly, define the positions $a_3, a_4, \dots$. Show that there exist an iid sequence of random vectors $\{(\gamma_n, \zeta_n)\}_{n\in\mathbb{N}}$ such that

$$a_n = a_{n-1}\gamma_n \text{ and } \tau_n = \zeta_n a_{n-1}^2,$$

for $n \in \mathbb{N}$. What are the distributions of γ_n and ζ_n, for each $n \in \mathbb{N}$? *Note:* Even though it is not used in the proof of the main statement of the problem, the following question is a good test of your intuition: are γ_1 and ζ_1 independent?

6. At this point, forget all about the fact that the sequence $\{(\gamma_n, \zeta_n)\}_{n\in\mathbb{N}}$ comes from a planar Brownian motion, and show, using discrete-time methods, that

$$\sum_{n=1}^{\infty} \tau_n = \sum_{n=1}^{\infty} a_{n-1}^2 \zeta_n = \infty, \text{ a.s.} \tag{17.8}$$

7. Next, show that the event $A = \{\sum_{n=1}^{\infty} a_{n-1}^2 \zeta_n = \infty\}$ is trivial, i.e., that $\mathbb{P}[A] = 0$ or $\mathbb{P}[A] = 1$. Explain why it is enough to show that $\limsup_n \mathbb{P}[a_{n-1}^2 \zeta_n \geq a_0^2] > 0$ to prove that $\mathbb{P}[A] = 1$.

8. For the grand finale, show that the random variable $\ln|\gamma_1|$ is symmetric with respect to the origin and conclude that $\mathbb{P}[a_{n-1}^2 \geq a_0^2] = \frac{1}{2}$. Use that to show that (17.8) holds.

Problem 17.6 (An erroneous conclusion)**.** What is wrong with the following argument which seems to prove that no zero of the Brownian motion is isolated from the right (which would imply that $B_t = 0$ for all $t \geq 0$):

"Define

$$\tau = \inf\{t \geq 0 \,:\, B_t = 0 \textit{ and } \exists\, \varepsilon > 0,\ B_s \neq 0 \textit{ for } s \in [t, t+\varepsilon]\}$$

and note that τ is an $\{\mathcal{F}_t^{B+}\}_{t\in[0,\infty)}$-stopping time. Therefore the strong Markov property can be applied to the finite stopping time $\tau \wedge t_0$ to conclude that $W_t = B_{\tau\wedge t_0 + t} - B_{\tau\wedge t_0}$, $t \geq 0$, is a Brownian motion with the property that $W_t(\omega)$ does not take the value 0 on some ω-dependent neighborhood of 0 for $\omega \in \{\tau < t_0\}$. By Corollary from the notes, this happens only on a set of probability 0, so $\mathbb{P}[\tau < t_0] = 0$, and, since t_0 is arbitrary, $\tau = \infty$, a.s."

Problem 17.7 (The third arcsine law (*))**.** The **third arcsine law** states that the random variable

$$\tau_3 = \int_0^1 \mathbf{1}_{\{B_u > 0\}}\, du,$$

which is sometimes called the **occupation time of** $(0,\infty)$, has the arcsine distribution. The method of proof outlined in this problem is different from those for the first two arcsine laws. We start with an analogous result for the Brownian motion and use Donsker's invariance principle to apply it to the paths of the Brownian motion.

Let $(X_0, X_1, \dots, X_n)$ be the first n steps of a simple symmetric random walk, i.e., $X_0 = 0$ and the increments $\xi_1, \dots, \xi_n$, where $\xi_k = X_k - X_{k-1}$ are independent coin tosses. As in the proof of the reflection principle for random walks, we think of X as a random variable which takes values in the path space

$$C(0,\dots,n) = \{(x_0, x_1, \dots, x_n) \in \mathbb{Z}^{n+1} : x_0 = 0, x_i - x_{i-1} \in \{0,1\}, \text{ for } i = 1, \dots, n\}.$$

and define the functionals $I_n, J_n : C(0,\dots,n) \to \mathbb{N}_0$ by

$$I_n(x) = \#\{1 \le k \le n : x_k > 0\}$$

and

$$J_n(x) = \min\{0 \le k \le n : x_k = \max_{0 \le j \le n} x_j\},$$

for $x \in C(0,\dots,n)$. The first order of business is to show that $I_n(X)$ and $J_n(X)$ have the same distribution for $n \in \mathbb{N}$ when X is a simple random walk (this is known as **Richard's lemma**).

1. Let $P : \{-1,1\}^n$ be the partial sum mapping, i.e., for $c = (c_1, \dots, c_n) \in \{-1,1\}^n$, $P(c) = x$, where $x_0 = 0$, $x_k = \sum_{j=1}^k c_k$, $k = 1, \dots, n$. We define a mapping $T_c : \{-1,1\}^n \to \{-1,1\}^n$ by the following procedure:

- Place in *decreasing* order of k those c_k for which $P(c)_k > 0$.
- Then, place the remaining increments in the *increasing* order of k.

  For example,

$$T_c\Big((1,1,-1,-1,-1,1,1)\Big) = (1,-1,1,1,-1,-1,1).$$

  With T_c defined, let $T : C(0,\dots,n) \to C(0,\dots,n)$ be the transformation whose value on $x \in C(0,\dots,n)$ you obtain by applying T_c to the increments of x and returning the partial sums, i.e. $T = P^{-1} T_c P$. For example

$$T\Big((0,1,2,1,0,-1,0,1)\Big) = (0,1,0,1,2,1,0,1).$$

  Prove that T is a bijection and that if $(Y_0, \dots, Y_n) = T\Big((X_0, \dots, X_n)\Big)$, then $Y_0, \dots, Y_n$ is a simple symmetric random walk. *Hint:* Induction.

2. Show that $I_n(x) = J_n(T(x))$, for all $n \in \mathbb{N}$, $x \in C(0,\dots,n)$. Deduce Richard's lemma. *Hint:* Draw pictures and use induction.

3. Prove that functional

$$g(f) = \inf\{t \in [0,1] : f(t) = \sup_{s \in [0,1]} f(s)\}$$

   is not continuous on $C[0,1]$, but it is continuous at every $f \in C[0,1]$ which admits a unique maximum (hence a.s. with respect to the Wiener measure on $C[0,1]$.) *Note:* We are assuming the reader will easily adapt the notions and results from the canonical space $C[0,\infty)$ to its finite-horizon counterpart $C[0,1]$.

4. Show that the functional h, defined by

$$h(f) = \int_0^1 \mathbf{1}_{\{f(t)>0\}}\, dt,$$

is not continuous on $C[0,1]$, but is continuous at each $f \in C[0,1]$ which has the property that $\lim_{\varepsilon \searrow 0} \int_0^1 \mathbf{1}_{\{0 \leq f(t) \leq \varepsilon\}}\, dt = 0$, and, hence, a.s. with respect to the Wiener measure on $C[0,1]$.

5. Let $\{X_t^n\}_{t\in[0,1]}$ be the n-scaled interpolated random walk constructed from X as in Definition 14.37. Show that the difference

$$\tfrac{1}{n} I_n(X) - \int_0^1 \mathbf{1}_{\{X_t^n>0\}}\, dt$$

converges to 0 in probability.

6. Use Donsker's invariance principle (Theorem 14.38) and the Portmanteau theorem to identify the distribution of τ_3.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 19

$\mathbb{L}^2$-Stochastic integration

The stochastic integral for processes of finite variation

For $T > 0$, let $F : [0, T] \to \mathbb{R}$ be càdlàg function of finite variation, and let $|F| : [0, T] \to [0, \infty)$ be its total variation. For $0 \leq a < b \leq T$ we define

$$\mu((a, b]) = F(b) - F(a), \ \mu(\{0\}) = F(0).$$

A version of the Caratheodory extension theorem guarantees that the functional μ can be extended in a unique way to a signed (real) measure on $\mathcal{B}([0, T])$; this measure is called the **Stieltjes measure** induced by F, and it is typically denoted by dF. It is not hard to see (we leave the details to the reader) that there exist unique càdlàg functions $F_+, F_- : [0, T] \to \mathbb{R}$ of finite variation such that

1. dF_+ and dF_- are finite positive measures (do nottake negative values) with $dF = dF_+ - dF_-$, and
2. dF_+ and dF_- are absolutely continuous with respect to the measure $d|F|$ (the Stieltjes measure induced by the total variation $|F|$ of F) and $dF_+ + dF_- = d\,|F|$.

The decomposition $dF = dF_+ - dF_-$ is called the **Hahn-Jordan decomposition** of the Stieltjes measure F.

Definition 19.1 (Lebesgue-Stieltjes integral)**.** A function $h : [0, T] \to \mathbb{R}$ is said to be dF-integrable (denoted by $h \in \mathbb{L}^1(dF)$), if

$$\int_0^T |h(t)| \, d\,|F|\,(t) < \infty.$$

For $h \in \mathbb{L}^1(dF)$, its **Stieltjes integral** with respect to dF is defined by

$$\int_0^T h(u)\, dF(u) = \int_0^T h(u)\, dF_+(u) - \int_0^T h(u)\, dF_-(u).$$

When the function h is continuous, one can approximate the Lebesgue-Stieltjes integral by Riemann-like sums. More precisely, let $0 = t_0^n < t_1^n < \cdots < t_{d_n}^n < t_{d_n+1}^n = T$ be a sequence of partitions of $[0,T]$ such that

$$\lim_{n\to\infty} \sup_{0\leq k\leq d_n} \left| t_{k+1}^n - t_k^n \right| = 0,$$

and suppose that $h : [0,T] \to \mathbb{R}$ is a continuous function. We leave it to the reader to show that $h \in \mathbb{L}^1(dF)$ and that

$$\int_0^T h(x)\, dF(x) = \lim_{n\to\infty} \sum_{k=0}^{d_n} h(t_k^n)(F(t_{k+1}^n) - F(t_k^n)).$$

If, additionally, $F \in C^1([0,T])$, we have

$$\int_0^T h(t)\, dF(t) = \int_0^T h(t) F'(t)\, dt,$$

for all $h \in \mathbb{L}^1(dF)$.

We turn to processes, now. Let $\{X_t\}_{t\in[0,\infty)}$ be an adapted càdlàg process of finite variation, and let $\{H_t\}_{t\in[0,\infty)}$ be a progressively measurable process such that

$$\int_0^T |H_t(\omega)|\, d\,|X|_t\,(\omega) < \infty, \text{ for all } T \geq 0, \text{ for almost all } \omega,$$

where $d\,|X|_t\,(\omega)$ is the Stieltjes measure induced by the non-decreasing function $t \mapsto |X|_t\,(\omega)$ (the total-variation function of $X(\omega)$). Therefore, for almost all ω, and all $t \geq 0$, we can define the Stieltjes integral

$$Y_t(\omega) = \int_0^t H_t(\omega)\, dX_t(\omega),$$

where $dX_t(\omega)$ is the Stieltjes measure induced by the FV function $t \mapsto X_t(\omega)$. We set $Y_t(\omega) = 0$, for all $t \geq 0$ for ω in the exceptional set. It is not hard to show that the process $\{Y\}_{t\in[0,\infty)}$ is an adapted and càdlàg process of finite variation, and that it is continuous when X is continuous. The process $\{Y_t\}_{t\in[0,\infty)}$ is called the **stochastic integral** of $\{H_t\}_{t\in[0,\infty)}$ with respect to $\{X_t\}_{t\in[0,\infty)}$ and is sometimes denoted by $Y = (H \cdot X)$.

The importance of being of finite variation

Unfortunately, the Lebesgue-style integration stops with functions of finite variation. Here is why.

A sequence $0 = t_0 < t_1 < \cdots < t_k < t_{k+1} < \cdots < \infty$ of real numbers with the property that $\lim_{k\to\infty} t_k = +\infty$, is called a **partition of** $[0,\infty)$, and the set of all partitions of $[0,\infty)$ is denoted by $P_{[0,\infty)}$.

The elements $t_0, t_1, \dots$ of a partition are referred to as its **nodes**. For a partition $\Delta = \{t_k\}_{k\in\mathbb{N}} \in P_{[0,\infty)}$, and $t \geq 0$, we define

$$|\Delta|_{[0,t]} = \sup_{k\in\mathbb{N}_0} |t_{k+1} \wedge t - t_k \wedge t|$$

and note that $|\Delta|_{[0,t]} = \max(|t_1 - t_0|, |t_2 - t_1|, \dots, |t - t_{k(t)}|)$, where $k(t) = \sup\{k \in \mathbb{N} : t_k < t\}$. A sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P_{[0,\infty)}$ is said to **converge to identity**, denoted by $\Delta_n \to \mathrm{Id}$ if $|\Delta_n|_{[0,t]} \to 0$, for each $t \geq 0$.

Let $F : [0,\infty) \to \mathbb{R}$ be a continuous function, and let $\mathcal{D}_{simp}$ denote the set of all functions $h : [0,\infty) \to \mathbb{R}$ for which there exists a partition $\Delta = \{t_k\}_{k\in\mathbb{N}} \in P_{[0,\infty)}$, such that

$$h(t) = \alpha_0 \mathbf{1}_{\{0\}}(t) + \sum_{k=1}^{\infty} \alpha_k \mathbf{1}_{(t_k, t_{k+1}]}(t),\ n \in \mathbb{N},\ \alpha_k \in \mathbb{R},\ k \in \mathbb{N}. \quad (19.1)$$

For $h \in \mathcal{S}$ and $t > 0$ we naturally define

$$\int_0^t h(u)\, dF(u) = \alpha_0 F(0) + \sum_{k=1}^{\infty} \alpha_k (F(t \wedge t_{k+1}) - F(t \wedge t_k)). \quad (19.2)$$

Note that the sum is really finite (we could have stopped at the index of the first $t_k > t$). For a continuous function $h : [0,\infty) \to \mathbb{R}$ and a partition $\Delta \in P_{[0,\infty)}$, we define the approximation h^Δ to h in $\mathcal{D}_{simp}$ by

$$h^\Delta(t) = h(0)\mathbf{1}_{\{0\}}(t) + \sum_{k=1}^{\infty} h(t_k) \mathbf{1}_{(t_k, t_{k+1}]}(t),\ n \in \mathbb{N}, \quad (19.3)$$

and note that $h^\Delta \to h$ uniformly on compact sets. We ask the following question

Question 19.2. What properties does F need to have for the limit

$$\lim_{\Delta \to \mathrm{Id}} \int_0^t h^\Delta(u)\, dF(u),$$

to exist for each $t \geq 0$ and each continuous h?

Remark 19.3. The reader should note that the question we are asking above really deals with a weak continuity property for the eventual extension of the simple integral (19.2) to a larger class of integrands h. Indeed, if a functional $h \mapsto \int h\, dF$ is to be called an integral, it should be linear, and somewhat continuous. The requirement that the uniform convergence of integrands $h^\Delta \to h$ implies convergence of the integrals is a particularly weak form of the dominated convergence theorem. Nevertheless, as we will see, it restricts the class of functions F considerably.

The main analytic tool we are going to use is the celebrated Banach-Steinhaus (aka uniform-boundedness theorem). Remember that for a linear operator $T : X \to Y$ between two normed spaces $(X, ||\cdot||_X)$ and $(Y, ||\cdot||_Y)$, we define its **operator norm** $||A||$ by

$$||T|| = \sup\{||Tf||_Y : f \in X,\ ||f||_X \leq 1\}.$$

Theorem 19.4 (Banach, Steinhaus)**.** *Let X be a Banach space, let Y be a normed space, and let $(B_\alpha)_{\alpha\in A}$ be a family of continuous linear operators $B_\alpha : X \to Y$. Then*

$$\sup_{\alpha\in B} ||B_\alpha|| < \infty \ \textit{ if and only if } \ \forall f \in X,\ \sup_{\alpha\in B} ||B_\alpha f||_Y < \infty.$$

Proof. (*) One direction is trivial. For the other, suppose that

$$\sup_{\alpha\in B} ||B_\alpha f||_Y < \infty \text{ for each} f \in X.$$

Define the sequence $\{D_n\}_{n\in\mathbb{N}}$ of subsets of X by

$$D_n = \bigcap_{\alpha\in A} \{f \in X : ||B_\alpha f|| \leq n\},$$

which, as intersections of closed sets are themselves closed. By the assumption, $\cup_{n\in\mathbb{N}} D_n = X$. Since X is complete, Baire's theorem[1] implies that at least one of $\{D_n\}_{n\in\mathbb{N}}$ has nonempty interior. More explicitly, there exists $n \in \mathbb{N}$, $f_0 \in D_n$ and $\varepsilon > 0$ such that $f_0 + \varepsilon B_1(X) \subseteq D_n$, where $B_1(X) = \{f \in X : ||f||_X \leq 1\}$ denotes the unit ball of X. Consequently,

[1] Baire's (category) theorem states that the intersection of a countable collections of dense open sets in a complete metric space is dense itself.

$$||B_\alpha(\tfrac{1}{\varepsilon} f_0 + f)||_Y \leq \tfrac{1}{\varepsilon} n \text{ for all } f \in B_1(X) \text{ and all } \alpha \in A.$$

The triangle inequality now implies that

$$||B_\alpha f||_Y \leq \tfrac{1}{\varepsilon} n + ||B_\alpha \tfrac{1}{\varepsilon} f_0|| \leq \tfrac{n}{\varepsilon} + \tfrac{1}{\varepsilon} \sup_{\alpha\in A} ||B_\alpha f_0||, \text{ for all } \alpha \in A,\ f \in B_1(X),$$

and so $\sup_{\alpha\in A} ||B_\alpha|| \leq \frac{n}{\varepsilon} + \frac{1}{\varepsilon} \sup_{\alpha\in A} ||B_\alpha f_0|| < \infty$. □

Proposition 19.5. *Let F be a continuous function on $[0,1]$, such that for each continuous function $h : [0,1] \to \mathbb{R}$, the family*

$$\int_0^1 h^{\Delta_n}(u)\, dF(u), n \in \mathbb{N},$$

is bounded in $\mathbb{R}$ for each sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P_{[0,\infty)}$ such that $\Delta_n \to \mathrm{Id}$. Then F is a function of finite variation on $[0,1]$.

Proof. (*) Let $X = C[0,1]$ be the Banach space of all continuous functions on $[0,1]$ normed with $||h|| = \sup_{t\in[0,1]} |h(t)|$, for $h \in C[0,1]$, and let $Y = \mathbb{R}$. We fix a sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P_{[0,\infty)}$ with $\Delta_n \to \mathrm{Id}$, and assume, without loss of generality that $1 \in \Delta_n$, for all $n \in \mathbb{N}$. A sequence of operators can be defined by

$$B_n : X \to Y,\ B_n h = \int_0^1 h^{\Delta_n}(u)\, dF(u),\ n \in \mathbb{N}.$$

These operators are continuous (why?), and, by the assumption, the sequence $\{B_n h\}_{n\in\mathbb{N}}$ is bounded, for each $h \in X$. Therefore, by Theorem 19.4, there exists a constant $K \geq 0$ such that

$$\left| \int_0^1 h^{\Delta_n}(u)\, dF(u) \right| \leq K||h||,\ \forall\, h \in C[0,1],\ \forall\, n \in \mathbb{N}. \tag{19.4}$$

For each $n \in \mathbb{N}$, one can construct a function $h_n \in C[0,1]$ such that $h_n(0) = \mathrm{sgn}(F(0))$ and $h_n(t^n_k) = \mathrm{sgn}(F(t^n_{k+1}) - F(t^n_k))$, for all $k \leq k^{\Delta_n}(1)$, where $\Delta_n = \{t^n_k\}_{k\in\mathbb{N}}$. Moreover, such a function can be constructed with $||h_n|| = 1$. Using (19.4), we get

$$K \geq \left| \int_0^1 h_n^{\Delta_n}(u)\, dF(u) \right| = |F(0)| + \sum_{k=0}^{k^{\Delta_n}(1)} \left| F(t^n_{k+1}) - F(t^n_k) \right|.$$

We have proved that the sequence of variations of F along any sequence of partitions converging to identity remains bounded. It follows (why?) that F must be a function of finite variation. $\square$

We can repeat the above argument on each $[0,t]$ instead on $[0,1]$, and, since any convergent sequence is bounded, the answer to Question 19.2 must be *It is necessary that F be of finite variation*. On the other hand, when F is of finite variation, the dominated convergence theorem implies that the sequence in Question 19.2 indeed converges towards $\int_0^t h(u)\, dF(u)$.

Square-integrable martingales as integrators

In spite of the impossibility result of the previous subsection, we will still be able to construct a satisfactory integration theory when the integrand is not of finite variation. In that case, it will be important that some randomness is present and that the class of integrands be restricted to those that do not depend on the future. This way, the terms that would otherwise accumulate and lead to the explosion in the approximating sequence will cancel each other (in the law-of-large-numbers manner). Before we move on to the general case, here is an example in a simplified (but still fully-featured) setting:

Example 19.6. Without knowing it, we have constructed a very simple version of a stochastic integral when we proved the martingale convergence theorem. Indeed, let $\{\xi_n\}_{n\in\mathbb{N}}$ be an iid sequence of coin-tosses

($\mathbb{P}[\xi_1 = 1] = \mathbb{P}[\xi_1 = -1] = \frac{1}{2}$) and let $\{M_t\}_{t\in[0,1]}$ be defined as follows:

$$M_t = \sum_{k=1}^{\infty} S_k \mathbf{1}_{[1-2^{-k+1},1-2^{-k})}(t), \text{ where } S_k = \sum_{k=1}^{n} \tfrac{1}{k}\xi_k, \text{ for } t < 1.$$

The martingale convergence theorem guarantees that the limit $M_1 = \lim_{t\to 1} M_t$ exists a.s. The reader can check that the variation of the path $M_\cdot(\omega)$ on $[0,1]$ is given by $\sum_{k=1}^{\infty} \frac{1}{k}\,|\xi_k(\omega)| = \sum_{k=1}^{\infty} \frac{1}{k} = +\infty$, so that no trajectory of $\{M_t\}_{t\in[0,1]}$ is of finite variation[2]. On the other hand, let $\{H_t\}_{t\in[0,1)}$ be any bounded (say by $K \geq 0$) and $\{\mathcal{F}_t^M\}_{t\in[0,1]}$-adapted left-continuous stochastic process. On each segment $[0,t]$, $t < 1$, $\{M_t\}_{t\in[0,1]}$ is a process of finite variation and so we have the following expression

$$\int_0^t H_u\, dM_u = \sum_{k=1}^{k(t)} \tfrac{h_k}{k}\xi_k,$$

where $h_k = H_{1-2^{-k+1}}$ and $k(t)$ is such that $t \in [1-2^{-k+1}, 1-2^{-k})$. Moreover, $h_k \in \sigma(\xi_1, \xi_2, \dots, \xi_{k-1})$, so that the process

$$N_n = \sum_{k=1}^{n} \tfrac{h_k}{k}\xi_k,$$

is a martingale (it is, really, a martingale transform of $\{S_n\}_{n\in\mathbb{N}}$). Moreover, it is a martingale bounded in $\mathbb{L}^2$; indeed, by the orthogonality of martingale increments, we have

$$\mathbb{E}[N_n^2] = \sum_{k=1}^{n} \mathbb{E}[\tfrac{h_k^2}{k^2}\xi_k^2] = \sum_{k=1}^{n} \tfrac{1}{k^2} K^2 \leq K^2 \sum_{k=1}^{\infty} \tfrac{1}{k^2} < \infty,$$

so that $\sup_n \mathbb{E}[N_n^2] < \infty$. Consequently, the martingale convergence theorem implies that the limit $N_\infty = \lim_{n\in\mathbb{N}} N_n$ exists a.s., and in $\mathbb{L}^2$. Equivalently, the random variable

$$\int_0^1 H_u\, dM_u = \lim_{t\nearrow 1} \int_0^t H_u\, dM_u = \lim_n N_n = N_\infty,$$

is well defined, even though no path of M is of finite variation.

It is important to note that the sequence $\{N_n\}_{n\in\mathbb{N}}$ only converges a.s. There is no way to conclude that the limit exists for each ω. Indeed, take $h_n = 1$, for all $n \in \mathbb{N}$, and note that for those ω for which $\xi_k(\omega) = 1$, for all $k \in \mathbb{N}$, the sequence

$$\sum_{k=1}^{\infty} \tfrac{1}{k}\xi_k(\omega) = \sum_{k=1}^{\infty} \tfrac{1}{k} = +\infty,$$

does not converge. Moreover, the series $\sum_{k=1}^{\infty} \frac{1}{k}\xi_k(\omega)$ will *never* converge absolutely. It is the fact that the sequence $\{\xi_k\}_{k\in\mathbb{N}}$ fluctuates so much and the fact that h_k does not "see" ξ_k that guarantee convergence $N_n \to N_\infty$, and, equivalently, allow for the integral $\int_0^1 H_u\, dM_u$ to be well-defined.

[2] The reader will also note that $\{M_t\}_{t\in[0,1]}$ is not continuous; a similar example with $\{M_t\}_{t\in[0,1]}$ continuous can be concocted, but it would not be as transparent as the one we give here. Also, if one prefers to work with $[0,\infty)$ instead of $[0,1]$, one can define $M_t = \sum_{k=1}^{\infty} S_k \mathbf{1}_{[k-1,k)}(t)$.

Turning to the general case, we assume that a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,\infty)}, \mathbb{P})$, where $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ satisfies the usual conditions, is given and fixed throughout.

When the integrand is "simple" the definition of the stochastic integral is easy. More precisely, a stochastic process $\{H_t\}_{t\in[0,\infty)}$ is said to be **simple predictable** if there exists a partition $\Delta = \{t_n\}_{n\in\mathbb{N}} \in P_{[0,\infty)}$ and a sequence $\{K_n\}_{n\in\mathbb{N}_0}$ of random variables such that $K_n \in \mathcal{F}_{t_n}$, and

$$H_t = \sum_{n=0}^{\infty} K_n \mathbf{1}_{(t_n,t_{n+1}]}(t). \tag{19.5}$$

The set of all simple predictable processes is denoted by $\mathcal{H}_{simp}$. The subset $\mathcal{H}^{\infty}_{simp}$ of $\mathcal{H}_{simp}$ consists of all processes $\{H_t\}_{t\in[0,\infty)} \in \mathcal{H}_{simp}$ such that $||H||_{\mathcal{H}^{\infty}_{simp}} = \sup_{n\in\mathbb{N}_0} ||K_n||_{\mathbb{L}^\infty} < \infty$.

Definition 19.7 (Stochastic integration of simple processes)**.** Suppose that $\{M_t\}_{t\in[0,\infty)}$ is *an arbitrary* stochastic process, and let $\{H_t\}_{t\in[0,\infty)}$ be a simple predictable process with representation (19.5). The **(simple) stochastic integral** of H with respect to M is the stochastic process $\{(H\cdot M)_t\}_{t\in[0,\infty)}$, given by

$$(H\cdot M)_t = \sum_{n=0}^{\infty} K_n (M_{t\wedge t_{n+1}} - M_{t\wedge t_n}). \tag{19.6}$$

Remark 19.8.

1. Note that simple predictable processes are predictable (measurable in the σ-algebra generated by left-continuous adapted processes). Moreover, $\mathcal{H}_{simp}$ and $\mathcal{H}^{\infty}_{simp}$ are vector spaces and the functional $||\cdot||_{\mathcal{H}^{\infty}_{simp}}$ is a norm on $\mathcal{H}^{\infty}_{simp}$ (if indistinguishable processes are identified).

2. One of the best ways to visualize the stochastic integral is to remember the notion of the *martingale transform* from our discussion of discrete-time martingales. The gambling interpretation used there easily transfers to continuous time, at least when the integrands are simple predictable, as above. Note, though, that one has to think of "gambles" in an infinitesimal manner in continuous time.

3. The value at $t \geq 0$ of the stochastic integral is usually denoted by $(H\cdot M)_t$ or $\int_0^t H_u\, dM_u$. Sometimes the terminology is abused and the random variable $(H\cdot M)_t$ (as opposed to the whole process $\{(H\cdot M)_t\}_{t\in[0,\infty)}$ or, simply, $H\cdot M$) is called the stochastic integral.

The stochastic integral for square-integrable martingales

We turn now to a general construction. Let $\mathcal{M}^{2,c}_0$ denote the family of all continuous martingales M with $M_0 = 0$ such that $||M||_{\mathcal{M}^{2,c}_0} :=$

$\sqrt{\sup_{t\geq 0}\mathbb{E}[M_t^2]} < \infty$. Here is the fundamental property of the (simple) stochastic integral with respect to $M \in \mathcal{M}_0^{2,c}$:

Proposition 19.9 (Preservation of $\mathcal{M}_0^{2,c}$ - simple predictable integrands)**.** *For $M \in \mathcal{M}_0^{2,c}$ and $\{H_t\}_{t\in[0,\infty)} \in \mathcal{H}^\infty_{simp}$ we have $H\cdot M \in \mathcal{M}_0^{2,c}$ and*

$$\mathbb{E}[(H\cdot M)_t^2] = \sum_{n=0}^{\infty} \mathbb{E}[K_n^2(M_{t_{n+1}\wedge t} - M_{t_n\wedge t})^2]. \tag{19.7}$$

Proof. Continuity, value 0 at time 0, and the martingale property of $H\cdot M$ are evident from the definition. To establish (19.7), we assume that $t = t_N$ for some N (otherwise just add it to the partition), and use the fact that the increments of square-integrable martingales are orthogonal, i.e.,

$$\mathbb{E}[K_n(M_{t_{n+1}} - M_{t_n})K_m(M_{t_{m+1}} - M_{t_m})] = 0 \text{ for } n < m.$$

Finally, $(H\cdot M) \in \mathcal{M}_0^{2,c}$ because, for $t \geq 0$, we have

$$\begin{aligned}\mathbb{E}[(H\cdot M)_t^2] &= \sum_{n\in\mathbb{N}} \mathbb{E}[K_n^2(M_{t_{n+1}} - M_{t_n})^2]\\ &\leq ||H||^2_{\mathcal{H}^\infty_{simp}} \sum_n \mathbb{E}[(M_{t_{n+1}} - M_{t_n})^2]\\ &= ||H||^2_{\mathcal{H}^\infty_{simp}} ||M||^2_{\mathcal{M}_0^{2,c}}.\end{aligned}$$

□

We turn now to the notion of *quadratic variation* of a continuous martingale. The main theorem listing its properties is stated without proof, but the reader is instructed to draw the parallel with the case of the Brownian motion we covered in previous lectures. Before we start, we need to introduce some notation: let $\mathcal{A}_0^c$ denote the set of all continuous, adapted, nondecreasing processes $\{A_t\}_{t\in[0,\infty)}$ with $A_0 = 0$. We also remind the reader that, for a RCLL of RLCC process X, we define the **maximal process** X^* by $X_t^* = \sup_{s\leq t}|X_s|$. For a sequence $\{X^n\}_{n\in\mathbb{N}}$ of RCLL or RLCC processes, we say that X_n converges to X, **uniformly on compacts in probability (ucp)**, and write $X_n \overset{\text{ucp}}{\to} X$, if

$$(X^n - X)_t^* \to 0 \text{ in probability, for each } t \geq 0.$$

Theorem 19.10 (Quadratic variation of square-integrable martingales)**.** *For $M \in \mathcal{M}_0^{2,c}$, there exists a unique process $\langle M\rangle \in \mathcal{A}_0^c$ such that such that $M_t^2 - \langle M\rangle_t$ is a uniformly-integrable martingale. Furthermore, for each sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P_{[0,\infty)}$ with $\Delta_n \to \mathrm{Id}$, we have*

$$\langle M\rangle^{\Delta_n} \overset{\text{ucp}}{\to} \langle M\rangle,$$

where, for $\Delta = \{t_k\}_{k\in\mathbb{N}} \in P_{[0,\infty)}$, $\langle M\rangle_t^\Delta := \sum_{k=0}^\infty (M_{t_{k+1}\wedge t} - M_{t_k\wedge t})^2$.

The process $\langle M\rangle \in \mathcal{A}_0^c$ is called the **quadratic variation (process)** of M. It admits a natural interpretation as the increasing part in the (continuous-time) Doob-Meyer decomposition of M^2 and plays a central role in stochastic analysis. The reader should probably do Problem **??** at this point. Here is how quadratic variation can be used to construct the stochastic integral:

Given $M \in \mathcal{M}_0^{2,c}$, it follows from the fact that both M and $M^2 - \langle M\rangle$ are a martingales that

$$\begin{aligned}\mathbb{E}[(M_{t_{k+1}} - M_{t_k})^2|\mathcal{F}_{t_k}] &= \mathbb{E}[M^2_{t_{k+1}} - M^2_{t_k}|\mathcal{F}_{t_k}]\\ &= \mathbb{E}[\langle M\rangle_{t_k+1} - \langle M\rangle_{t_k}|\mathcal{F}_{t_k}].\end{aligned}$$

Therefore, the relation (19.7) can be written as

$$\begin{aligned}\mathbb{E}[(H\cdot M)^2_\infty] &= \mathbb{E}\Big[\sum_{n=0}^\infty K_k^2\, \mathbb{E}[\langle M\rangle_{t_{k+1}} - \langle M\rangle_{t_k}|\mathcal{F}_{t_k}]\Big] \qquad (19.8)\\ &= \mathbb{E}\Big[\sum_{n=0}^\infty K_k^2\, \big(\langle M\rangle_{t_{k+1}} - \langle M\rangle_{t_k}\big)\Big] = \mathbb{E}[\int_0^\infty H_t^2\, d\langle M\rangle_t].\end{aligned}$$

Inspired by the last expression in (19.8) above, for a progressively-measurable process H, we define

$$||H||_{\mathbb{L}^2(M)} := \sqrt{\mathbb{E}[\int_0^\infty H_t^2\, d\langle M\rangle_t]}.$$

It is not hard to check that the family $\mathbb{L}^2(M)$ of all progressively-measurable processes H for which $||H||_{\mathbb{L}^2(M)} < \infty$ forms a vector space, and that $||\cdot||_{\mathbb{L}^2(M)}$ is a norm there. We also note that $\mathcal{H}^\infty_{simp} \subseteq \mathbb{L}^2(M)$, for each $M \in \mathcal{M}_0^{2,c}$ (why?). In this, new, notation, the conclusion of (19.8) looks like this:

$$||H\cdot M||_{\mathcal{M}_0^{2,c}} = ||H||_{\mathbb{L}^2(M)},\ \forall\, H \in \mathcal{H}^\infty_{simp}, \qquad (19.9)$$

and is called **Itô's isometry**. We use it to extend the domain of the stochastic integral from $\mathcal{H}^\infty_{simp}$ to a much larger set. To do that, we need a simple, but very powerful, result from real analysis:

Proposition 19.11. *Let X be a metric space and Y a* complete *metric space. Moreover, let $f : A \to Y$, with $\emptyset \neq A \subseteq X$ be a* uniformly-continuous *function. Then, there exists a unique uniformly-continuous function $\tilde{f} : \operatorname{Cl} A \to Y$ such that $\tilde{f}(x) = f(x)$, for all $x \in A$.*

Proof. For $x \in \operatorname{Cl} A$, let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in A such that $x_n \to x$, and let $y_n = f(x_n)$. Since x_n is convergent in X and f is uniformly continuous, $\{y_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in Y. Since Y is complete it admits a unique limit y, and we set

$$\tilde{f}(x) = y.$$

To show that $\tilde{f}$ is well-defined, we need to argue that y does not depend on the choice of the sequence $\{x_n\}_{n\in\mathbb{N}}$. Indeed, let $\{x'_n\}_{n\in\mathbb{N}}$ be another sequence in A with $x'_n \to x$. Then $d(x_n, x'_n) \to 0$, and, so, by uniform continuity, $d(f(x_n), f(x'_n)) = 0$, for each y. It follows that $y = \lim_n f(x_n) = \lim_n f(x'_n)$.

Next, we argue that that $\tilde{f}$ is uniformly continuous. Given $\varepsilon > 0$, let $\delta > 0$ be such that $d(x, \hat{x}) < \delta$ implies $d(f(x), f(\hat{x})) < \varepsilon$, for all $x, \hat{x} \in A$. For any $x, x' \in \operatorname{Cl} A$ with $d(x, x') < \delta$, we can choose two sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{x'_n\}_{n\in\mathbb{N}}$ in A such that $x_n \to x$ and $x'_n \to x'$ and $d(x_n, x'_n) < \delta$, for all n. It follows that $d(f(x_n), f(x'_n)) < \varepsilon$, for all n, and, so, $d(\tilde{f}(x), \tilde{f}(x')) \leq \varepsilon$, estalishing the uniform continuity of $\tilde{f}$ on $\operatorname{Cl} A$.

Finally, given $x \in A$, we may take $x_n = x$, for all n, to conclude that $\tilde{f}(x) = f(x)$. □

The completeness of the space Y is crucial for the Proposition 19.11 to work. To check that it is applicable in our case, we need the following result:

Proposition 19.12. *$(\mathcal{M}_0^{2,c}, ||\cdot||_{\mathcal{M}_0^{2,c}})$ is a Banach space.*

Proof. Simple properties of square-integrable martingales suffice to show that $\mathcal{M}_0^{2,c}$ is a linear space, and that $||\cdot||_{\mathcal{M}_0^{2,c}}$ satisfies all the axioms of a norm. The more delicate matter is completeness. We start with a Cauchy sequence $\{M^n\}_{n\in\mathbb{N}}$ and conclude immediately that their last elements $\{M^n_\infty\}_{n\in\mathbb{N}}$ form a Cauchy sequence in $\mathbb{L}^2(\mathcal{F})$. Thus, $M^n_\infty \to M_\infty$ in $\mathbb{L}^2$, for some $M_\infty \in \mathbb{L}^2(\mathcal{F})$. We define $\{M\}_{t\in[0,\infty)}$ as the RCLL version of the Lévy martingale $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]$ and use Doob's inequality to get that

$$\Big|\Big| \sup_{t\geq 0} |M^n_t - M_t| \Big|\Big|_{\mathbb{L}^2} \leq 2||M^n_\infty - M_\infty||_{\mathbb{L}^2} \to 0,$$

Hence, possibly through a subsequence, we have

$$\sup_{t\geq 0} |M^n_t - M_t| \to 0, \text{ a.s.}$$

It follows that almost all trajectories of M are uniform limits of the corresponding trajectories of M^n. By assumption, all M^n are continuous martingales, and uniform limits of continuous functions are continuous, so M is a continuous martingale and $M = \lim_n M^n$ in $\mathcal{M}_0^{2,c}$. □

If we combine the Itô's isometry (19.9) (which yields the uniform continuity of the map $H \mapsto H \cdot M$ on $\mathcal{H}^\infty_{simp}$) with Proposition 19.12 (which asserts that the target space is complete) and use Proposition 19.11 we immediately conclude that the map

$$H \mapsto H \cdot M \in \mathcal{M}_0^{2,c}$$

admits a unique linear and continuous extension of the simple predictable integral from $\mathcal{H}^\infty_{simp}$ to its closure in $\mathbb{L}^2(M)$, with values in $\mathcal{M}^{2,c}_0$. The only question still remaining is to give a nice description of the closure of $\mathcal{H}^\infty_{simp}$ in $\mathbb{L}^2(M)$:

Proposition 19.13. *For any $M \in \mathcal{M}^{2,c}_0$, $\mathcal{H}^\infty_{simp}$ is dense in $\mathbb{L}^2(M)$.*

Proof. Without loss of generality, we assume that $\mathbb{E}[\langle M\rangle] = 1$. Let $\{\Delta_n\}_{n\in\mathbb{N}}$ be a sequence of nested partitions in $P_{[0,\infty)}$ such that $\Delta_n \to \mathrm{Id}$; for concreteness, we take $\Delta_n = \{k2^{-n} : 0 \leq k \leq n2^n\}$). For each n, let $\mathcal{P}_n$ be the σ-algebra on the product $[0,\infty) \times \Omega$ generated by all simple predictable processes H of the form

$$H_t = \sum_{k=0}^{n2^n} K_k \mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}(t),\ K_k \in \mathcal{F}_{k2^{-n}},\ \text{for all } k. \tag{19.10}$$

In fact, it is not hard to see that a process H is measurable with respect to $\mathcal{P}_n$, precisely when it is of the form (19.10). Also, since all left-continuous processes can be obtained as pointwise limits of the processes of the form (19.10), we have

$$\mathcal{P} = \sigma(\cup_n \mathcal{P}_n),$$

where $\mathcal{P}$ denotes the predictable σ-algebra. The product space $\Omega^* = [0,\infty) \times \Omega$, together with the σ-algebra $\mathcal{P}$, the filtration $\{\mathcal{P}_n\}_{n\in\mathbb{N}}$, and the probability measure $\mathbb{P}^*$, given by

$$\mathbb{P}^*[A] = \mathbb{E}[\int_0^\infty \mathbf{1}_A(t,\omega)\, d\langle M\rangle_t(\omega)],\ \text{for } A \in \mathcal{P},$$

forms a filtered probability space. Moreover, random variables on this space correspond precisely to the predictable processes on the original space, and their expectations to the expected $d\langle M\rangle$-integrals, i.e., if $\mathbb{E}^*$ denotes the expectation under $\mathbb{P}^*$, we have

$$\mathbb{E}^*[H] = \mathbb{E}[\int_0^T H_t\, d\langle M\rangle_t] \text{ where }.$$

In particular, we have

$$H_n \to H \text{ in } \mathbb{L}^2(\mathbb{P}^*) \text{ iff } H_n \to H \text{ in } \mathbb{L}^2(M).$$

Given $\tilde{H} \in \mathbb{L}^2(\mathbb{P}^*)$, define

$$\tilde{H}_n = \mathbb{E}^*[\tilde{H}|\mathcal{P}_n].$$

This is a square-integrable martingale and $\tilde{H} \in \mathcal{P}$, so $\tilde{H}^n \to \tilde{H}$ in $\mathbb{L}^2(\mathbb{P}^*)$. Therefore, by the above characterization of convergence in $\mathbb{L}^2(M)$, we have $\tilde{H}^n \to \tilde{H}$ in $\mathbb{L}^2(M)$, as required. □

We can now summarize all our findings in a compact statement:

Theorem 19.14. *Given $M \in \mathcal{M}_0^{2,c}$, there exist a unique linear isometry*

$$\mathbb{L}^2(M) \ni H \mapsto H \cdot M \in \mathcal{M}_0^{2,c}$$

which extends the simple predictable integral (19.6).

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 19
Semimartingales

Continuous local martingales

While tailor-made for the $\mathbb{L}^2$-theory of stochastic integration, martingales in $\mathcal{M}_0^{2,c}$ do not constitute a large enough class to be ultimately useful in stochastic analysis. It turns out that even the class of all martingales is too small. When we restrict ourselves to processes with continuous paths, a naturally stable family turns out to be the class of so-called local martingales.

Definition 19.1 (Continuous local martingales)**.** A continuous adapted stochastic process $\{M_t\}_{t\in[0,\infty)}$ is called a **continuous local martingale** if there exists a sequence $\{\tau_n\}_{n\in\mathbb{N}}$ of stopping times such that

1. $\tau_1 \leq \tau_2 \leq \dots$ and $\tau_n \to \infty$, a.s., and
2. $\{M_t^{\tau_n}\}_{t\in[0,\infty)}$ is a uniformly integrable martingale for each $n \in \mathbb{N}$.

In that case, the sequence $\{\tau_n\}_{n\in\mathbb{N}}$ is called the **localizing sequence** for (or is said to **reduce**) $\{M_t\}_{t\in[0,\infty)}$. The set of all continuous local martingales M with $M_0 = 0$ is denoted by $\mathcal{M}_0^{loc,c}$.

Remark 19.2.

1. There is a nice theory of local martingales which are not necessarily continuous (RCLL), but, in these notes, we will focus solely on the continuous case. In particular, a "martingale" or a "local martingale" will always be assumed to be continuous.
2. While we will only consider local martingales with $M_0 = 0$ in these notes, this is assumption is not standard, so we don't put it into the definition of a local martingale.
3. Quite often, instead for $\{M_t^{\tau_n}\}_{t\in[0,\infty)}$, the uniform integrability is required for $\{M_t^{\tau_n}\mathbf{1}_{\{\tau_n>0\}}\}_{t\in[0,\infty)}$, $n \in \mathbb{N}$. The difference is only significant when we allow $M_0 \neq 0$ and $\mathcal{F}_0$ contains non-trivial events,

and this modified definition leads to a very similar theory. It becomes important when one wants to study stochastic differential equations whose initial conditions are not integrable.

4. We do not have the right tools yet to (easily) construct a local martingale which is not a martingale, but many such examples exist.

Problem 19.1. Prove the following statements (remember, all processes with the word "martingale" in their name are assumed to be continuous),

1. Each martingale is a local martingale.
2. A local martingale is a martingale if and only if it is of class (DL).
3. A bounded local martingale is a martingale of class (D).
4. A local martingale bounded from below is a supermartingale.
5. For $M \in \mathcal{M}_0^{loc,c}$ and a stopping time τ, we have $M^\tau \in \mathcal{M}_0^{loc,c}$.
6. The set of all local martingales has the structure of a vector space.
   Note: Careful! The reducing sequence may differ from one local martingale to another.
7. For $M \in \mathcal{M}_0^{loc,c}$, the sequence

$$\tau_n = \inf\{t \geq 0 : |M_t| \geq n\},\ n \in \mathbb{N},$$

   reduces $\{M_t\}_{t\in[0,\infty)}$. *Note:* This is not true if M is not continuous.

The following result illustrates one of the most important techniques in stochastic analysis - namely, *localization*.

Proposition 19.3 (Continuous local martingales of finite variation are constant)**.** *Let M be a continuous local martingale of finite variation. Then $M_t = 0$, for all $t \geq 0$, a.s.*

Proof. Suppose that M is a continuous local martingale of finite variation and let $\{\tau_n\}_{n\in\mathbb{N}}$ be its reducing sequence. We can choose $\{\tau_n\}_{n\in\mathbb{N}}$ so that the stopped processes M^{τ_n}, $n \in \mathbb{N}$, are bounded, and, therefore, in $\mathcal{M}_0^{2,c}$. By Theorem 18.10, for each $n \in \mathbb{N}$, we have $\langle M^{\tau_n}\rangle^{\Delta_k} \overset{\text{ucp}}{\to} \langle M^{\tau_k}\rangle = \langle M\rangle^{\tau_k}$, for any sequence of partitions Δ_k with $\Delta_n \to \text{Id}$. On one hand,

$$\langle M^{\tau_n}\rangle_t^{\Delta_k} = \sum_{i=1}^{\infty}(M^{\tau_n}_{t^k_{i+1}\wedge t} - M^{\tau_n}_{t^k_i\wedge t})^2 \leq w_t^k \sum_{i=1}^{\infty} |M^{\tau_n}_{t^k_{i+1}\wedge t} - M^{\tau_n}_{t^k_i\wedge t}| \leq w_t^k \, |M|_t \,,$$

where $|M|_t$ denotes the total variation of M on $[0,t]$, and

$$w_t^k = \sup_{i \in \mathbb{N}} \left| M_{t_{i+1}^k \wedge t} - M_{t_i^k \wedge t} \right|.$$

Since the paths of M are continuous (and, thus, uniformly continuous on compacts), we have $w_t^k \to 0$, a.s., as $k \to \infty$, for each $t \geq 0$. Therefore, $\langle M \rangle^{\tau_n} = 0$. This implies that $\mathbb{E}[(M_t^{\tau_n})^2] = 0$, for all $t \geq 0$, i.e., $M^{\tau_n} = 0$. Since $\tau_n \to \infty$, a.s., we have $M = 0$. □

The process of localization allows us to extend the notion of quadratic variation from square-integrable martingales, to local martingales. We need a result which loosely states that the ucp convergence and localization commute:

Problem 19.2. Let $\{X^n\}_{n\in\mathbb{N}}$ be a sequence of RCLL or LCRL processes, and let $\{\tau_k\}_{k\in\mathbb{N}}$ be a sequence of stopping times such that $\tau_k \to \infty$, a.s. If X is such that $(X^n)^{\tau_k} \overset{\text{ucp}}{\to} X^{\tau_k}$, for each k, then $X^n \overset{\text{ucp}}{\to} X$.

Theorem 19.4 (Quadratic variation of continuous local martingales)**.** *Let $\{M_t\}_{t\in[0,\infty)}$ be a continuous local martingale. Then there exists a unique process $\langle M \rangle \in \mathcal{A}_0^c$ such that $M^2 - \langle M \rangle$ is a local martingale. Moreover,*

$$\langle M \rangle^{\Delta_n} \overset{\text{ucp}}{\to} \langle M \rangle, \tag{19.1}$$

for any sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P_{[0,\infty)}$ with $\Delta_n \to \mathrm{Id}$.

Proof. The main idea of the proof is to use a localizing sequence for M to reduce the problem to the situation dealt with in Theorem 18.10 in the previous lecture. Problem 19.1, part 7. above implies that we can use the hitting-time sequence $\tau_k = \inf\{t \geq 0 : |M_t| \geq k\}$, $k \in \mathbb{N}$ to reduce M, i.e., that for each n, the process $M^{(k)}$, given by $M_t^{(k)} = M_t^{\tau_k}$ is a uniformly integrable martingale. Moreover, thanks to the choice of τ_k, each $M^{(k)}$ is bounded, and, therefore, in $\mathcal{M}_0^{2,c}$. For each $k \in \mathbb{N}$, Theorem 18.10 states that there exists $\langle M^{(k)} \rangle \in \mathcal{A}_0^c$ such that $(M^{(k)})^2 - \langle M^{(k)} \rangle$ is a UI martingale. The uniqueness of such a process implies that $\langle M^{(k)} \rangle_t = \langle M^{(k+1)} \rangle_t$ on $\{t \leq \tau_n\}$ (otherwise, we could stop the process $\langle M^{(k+1)} \rangle$ at τ_n and use it in lieu of $\langle M^{(k)} \rangle$). In words, $\langle M^{(k+1)} \rangle$ is an "extension" of $\langle M^{(k)} \rangle$. Therefore, by using the process $\langle M^{(k)} \rangle$ on $t \in [\tau_{k-1}, \tau_k)$ (where $\tau_0 = 0$), we can construct a continuous, adapted and non-decreasing process $\langle M \rangle$ with the property that $\langle M \rangle_t = \langle M^{(k)} \rangle_t$ for $\{t \leq \tau_k\}$ and all $k \in \mathbb{N}$. Such a process clearly has the property that $N_t = M_t^2 - \langle M \rangle_t$ is a martingale on $[0, \tau_k]$. Equivalently, the stopped process N^{τ_k} is a martingale on the whole $[0,\infty)$, which means that N_t is a local martingale. The uniqueness of $\langle M \rangle$ follows directly from Proposition 19.3.

We still need to establish 19.1. This follows directly from Problem 19.2 above and Theorem 18.10, which implies that $(\langle M\rangle^{\tau_k})^{\Delta_n} \overset{\text{ucp}}{\to} \langle M\rangle^{\tau_k}$ for each k, as $n \to \infty$. □

Quadratic covariation of local martingales

The concept of the quadratic covariation between processes, already prescreened in Problem 18.1 for martingales in $\mathcal{M}_0^{2,c}$, rests on the following identity:

$$xy = \tfrac{1}{2}\big((x+y)^2 - x^2 - y^2\big),$$

is often refered to as the **polarization identity** and is used whenever one wants to produce a "bilinear" functional from a "quadratic" one. More precisely, we have the following definition:

Definition 19.5 (Quadratic covariation)**.** For $M, N \in \mathcal{M}_0^{loc,c}$, the finite-variation process $\{\langle M, N\rangle_t\}_{t\in[0,\infty)}$, given by

$$\langle M, N\rangle_t = \tfrac{1}{2}\Big(\langle M+N\rangle_t - \langle M\rangle_t - \langle N\rangle_t\Big),$$

is called the **quadratic covariation** (bracket) of M and N.

At this point, the reader should try his/her hand at the following problem:

Problem 19.3. Show that, for $M, N \in \mathcal{M}_0^{loc,c}$,

1. all conclusions of Problem 18.1, parts 1., 2., and 4. remain valid (now that M, N are assumed to be only local martingales).
2. $\langle M, N\rangle$ is the unique adapted and continuous process of finite variation such that $\langle M, N\rangle = 0$ amd $MN - \langle M, N\rangle$ is a local martingale.

We conclude the section on quadratic covariation with an imporant inequality (the proof is postponed for the Additional Problems section below). Since $\langle M, N\rangle$ is a continuous and adapted finite-variation process, its total-variation process $\{|\langle M, N\rangle|_t\}_{t\in[0,\infty)}$ is continuous, adapted and nondecreasing.

Theorem 19.6 (Kunita-Watanabe inequality)**.** *For $M, N \in \mathcal{M}_0^{loc,c}$ and any two measurable processes H and K, we have*

$$\int_0^\infty |H_t|\,|K_t|\; d\,|\langle M, N\rangle|_t \le \Big(\int_0^\infty H_t^2\, d\langle M\rangle_t\Big)^{1/2}\Big(\int_0^\infty K_t^2\, d\langle N\rangle_t\Big)^{1/2}, \text{ a.s.} \tag{19.2}$$

Moreover, for any $p, q \in [1,\infty]$ *with* $\frac{1}{p} + \frac{1}{q} = 1$, *we have*

$$\mathbb{E}\left[\int_0^\infty |H_t|\,|K_t|\; d\,|\langle M,N\rangle|_t\right] \leq$$
$$\leq \left\| \left(\int_0^\infty H_t^2\, d\langle M\rangle_t \right)^{1/2} \right\|_{\mathbb{L}^p} \left\| \left(\int_0^\infty K_t^2\, d\langle N\rangle_t \right)^{1/2} \right\|_{\mathbb{L}^q}. \quad (19.3)$$

Stochastic integration for local martingales

The restriction $H \in \mathbb{L}^2(M)$ on the integrand, and $M \in \mathcal{M}_0^{2,c}$ on the integrator in the definition of the stochastic integral $H \cdot M$ can be relaxed. For a continuous local martingale M, we define the class $L(M)$ which contains all predictable processes H with the property

$$\int_0^t H_u^2\, d\langle M\rangle_u < \infty, \text{ for all } t \geq 0, \text{ a.s.}$$

In comparison with the space $\mathbb{L}^2(M)$, the $d\langle M\rangle$-integrals are on compact intervals, and the finite-expectation assumption is replaced by "finite-a.s." requirement.

We define the stochastic integral $H \cdot M$ for $H \in L(M)$ as follows: let $\{\tau_n\}_{n\in\mathbb{N}}$ be a sequence of stopping times defined by

$$\tau_n = \inf\{t \geq 0 : |M_t| \geq n\} \wedge \inf\{t \geq 0 : \int_0^t H_u^2\, d\langle M\rangle_u \geq n\}.$$

This sequence clearly reduces[1] M. Moreover, we know that $M^{\tau_n} \in \mathcal{M}_0^{2,c}$ (because it is bounded) and $H\mathbf{1}_{[0,\tau_n]} \in \mathbb{L}^2(M^{\tau_n})$ for all $n \in \mathbb{N}$. Therefore, the integral $H\mathbf{1}_{[0,\tau_n]} \cdot M^{\tau_n}$ is well-defined in the $\mathbb{L}^2$-sense for each n, and the sequence has the property that

$$H\mathbf{1}_{[0,\tau_n]} \cdot M^{\tau_n} \text{ and } H\mathbf{1}_{[0,\tau_{n+1}]} \cdot M^{\tau_{n+1}} \text{ coincide on } [0,\tau_n].$$

[1] By reducing each process separately, and then taking the minimum of all stopping times over a (finite) family of processes, we can always find a single sequence that reduces each process in the family. We will continue doing this in the sequel, with minimal explanation.

As in the proof of Theorem 19.4 we can patch the "stopped integrals" together to obtain a stochastic process which we denote by $H \cdot M$. The process $H \cdot M$ is clearly continuous and adapted, and we have

$$(H \cdot M)^{\tau_n} = H\mathbf{1}_{[0,\tau_n]} \cdot M^{\tau_n} \in \mathcal{M}_0^{2,c}.$$

Hence, the sequence $\{\tau_n\}_{n\in\mathbb{N}}$ can be used as a reducing sequence for $H \cdot M$, yielding that $H \cdot M \in \mathcal{M}_0^{loc,c}$. Summarizing all of the above, we have the following result (the reader will easily supply the proofs of all the remaining statements):

Theorem 19.7. *For* $M \in \mathcal{M}_0^{loc,c}$ *and* $H \in L(M)$, *there exists a stochastic process* $H \cdot M$ *in* $\mathcal{M}_0^{loc,c}$ *with the following properties*

1. $H \cdot M$ *conincides with the* $\mathbb{L}^2$*-stochastic integral, for* $M \in \mathcal{M}_0^{2,c}$ *and* $H \in \mathbb{L}^2(M)$.

2. *For each stopping time τ, we have $(H \cdot M)^\tau = H\mathbf{1}_{[0,\tau]} \cdot M^\tau$.*

The two properties above are enough to characterize the $\mathcal{M}_0^{loc,c}$-integral, and to transfer many properties from the $\mathcal{M}_0^{2,c}$ to the $\mathcal{M}_0^{loc,c}$ setting (the reader will be asked to do this, in a more general framework in Problem ? below).

Here is another characterization of the $\mathcal{M}_0^{loc,c}$-integral, this time in terms of the quadratic-covariation processes. It also tells us how to compute quadratic (co)variations of stochastic integrals:

Proposition 19.8. *For $M, L \in \mathcal{M}_0^{loc,c}$ and $H \in L(M)$, we have*

$$L = H \cdot M \text{ iff } \langle L, N\rangle = \int_0^\cdot H_u \, d\langle M, N\rangle, \text{ for all } N \in \mathcal{M}_0^{loc,c}. \tag{19.4}$$

Proof. First, we note that, since $\int_0^t H_u^2 \, \langle M\rangle_u < \infty$, a.s., we can use the Kunita-Watanabe inequality (19.2) (with $K = \mathbf{1}_{[0,t]}$) to conclude that

$$\int_0^t |H_u| \; d\,|\langle M, N\rangle|_u < \infty, \text{ a.s.,}$$

and that the integral $\int_0^\cdot H_u \, d\langle M, N\rangle_u$ is well-defined.

Assuming that $M, N \in \mathcal{M}_0^{2,c}$, $H \in \mathcal{H}^\infty_{simp}$, the elementary manipulations show that

$$\mathbb{E}[(H \cdot M)_\infty N_\infty] = \mathbb{E}[\int_0^\infty H_u \, d\langle M, N\rangle_u]. \tag{19.5}$$

The "in-expectation" form (19.3) of the Kunita-Watanabe inequality with $p = q = \frac{1}{2}$, implies immediately that both sides of (19.5) are linear continuous functions of $H \in \mathbb{L}^2(M)$. Since they coincide on a dense subset $\mathcal{H}^\infty_{simp}$ of $\mathbb{L}^2$, they must conincide everywhere on $\mathbb{L}^2(M)$. Finally, for $M, N \in \mathcal{M}_0^{loc,c}$, we first localize into $\mathcal{M}_0^{2,c}$, and then apply the dominated-convergence theorem (with its use justified by the Kunita-Watanabe inequality) to "remove the localization" and conclude that the left-hand side of (19.4) implies the right-hand side.

For the other implication, we assume that $L \in \mathcal{M}_0^{loc,c}$ satisfies the equality on the right-hand side of (19.4). Let τ_n be a localizing sequence for L; without loss of generality, we can also suppose that, stopped at each τ_n, the processes L, M and $\int_0^\cdot H_u^2 \, d\langle M\rangle_u$ are bounded. For $N \in \mathcal{M}_0^{2,c}$, the process $L^{\tau_n} N - \langle L^{\tau_n}, N\rangle$ is a UI martingale, so

$$\mathbb{E}[L^{\tau_n}_\infty N_\infty] = \mathbb{E}[\int_0^\infty H_t \mathbf{1}_{[0,\tau_n]}(t) \, d\langle M^{\tau_n}, N\rangle_t],$$

for all $N \in \mathcal{M}_0^{2,c}$. On the other hand, we also have

$$\mathbb{E}[(H \cdot M)^{\tau_n} N_\infty] = \mathbb{E}[\int_0^\infty H_t \mathbf{1}_{[0,\tau_n]}(t) \, d\langle M^{\tau_n}, N\rangle_t],$$

for all n, so

$$\mathbb{E}[L_\infty^{\tau_n} N_\infty] = \mathbb{E}[(H \cdot M)^{\tau_n} N_\infty],$$

for all $N \in \mathcal{M}_0^{2,c}$. In particular, the choice $N = L^{\tau_n} - (H \cdot M)^{\tau_n}$ implies that $||L^{\tau_n} - (H \cdot M)^{\tau_n}||_{\mathcal{M}_0^{2,c}} = 0$, i.e., that $L^{\tau_n} = (H \cdot M)^{\tau_n}$. Letting $n \to \infty$ gives the left-hand side of (19.4). □

The semimartingale integral

At this point we know how to use 1) processes of finite variation, and 2) local martingales as integrators. One can show (it is beyond the scope of these notes, though) that linear combinations of those are, in a sense, all processes that can be used as integrators. For this reason, we give them a name, but before we do, we introduce one more piece of notation: $\mathcal{V}_0^c$ denotes the family of all continuous and adapted processes with paths of finite variation which vanish at $t = 0$.

Definition 19.9. A stochastic process X is called a **continuous semimartingale** if there exist processes $A \in \mathcal{V}_0^c$ and $M \in \mathcal{M}_0^{loc,c}$ such that

$$X_t = X_0 + M_t + A_t, \text{ for all } t \geq 0, \text{ a.s.}$$

Using the fact that there are no non-trivial continuous local martingales of finite variation one can show that that for a continuous semimartingale the decomposition $X = X_0 + M + A$ into the initial value, a continuous local martingale and a continuous adapted process of finite variation is unique. This decomposition is called the **semimartingale decomposition** of X.

Problem 19.4. Show that, for a continuous semimartingale X with the decomposition $X = X_0 + M + A$, we have

$$\langle X \rangle^{\Delta_n} \overset{\text{ucp}}{\to} \langle M \rangle,$$

for any sequence $\{\Delta_n\}_{n\in\mathbb{N}} \in P_{[0,\infty)}$, with $\Delta_n \to \mathrm{Id}$.

Problem 19.4 above makes it natural to define the quadratic-variation process $\langle X \rangle$ of the semimartingale $X = X_0 + M + A$ by

$$\langle X \rangle_t = X_0^2 + \langle M \rangle_t.$$

For a continuous semimartingale X with the semimartingale decomposition $X = X_0 + A + M$, let $L(X)$ denote the set of all predictable processes with the property that

$$\int_0^t |H_u| \; d\,|A|_u + \int_0^t H_u^2 \, d\langle M \rangle_u < \infty \text{ for all } t \geq 0, \text{ a.s.}$$

For $H \in L(X)$ we can define both the Lebesgue-Stieltjes integral $\int_0^t H_u \, dA_u$ (which we also denoted by $H \cdot A$) and the stochastic integral $H \cdot M$; thus, we define the stochastic integral $H \cdot X$ of H with respect to X by

$$(H \cdot X)_t = (H \cdot A)_t + (H \cdot M)_t, \text{ for all } t \geq 0.$$

It is immediately clear that $H \cdot A$ is an adapted process of finite variation and that $H \cdot M$ is a local martingale, so that $H \cdot X$ is a continuous semimartingale and $H \cdot X = H \cdot A + H \cdot M$ is its semimartingale decomposition. As the reader can check, the stochastic integral for semimartingales has the following properties:

Problem 19.5. Let X be a continuous semimartingale with the semimartingale decomposition $X = A + M$. Then

1. Both maps $H \mapsto (H \cdot X)$ and $X \mapsto (H \cdot X)$ are linear on their natural domains.
2. For $H \in L(X)$ and $K \in L(H \cdot X)$, we have $KH \in L(X)$ and
$$KH \cdot X = K \cdot (H \cdot X).$$
3. For $H \in L(X)$ and a stopping time T, we have $H^T \in L(X)$ and
$$(H \cdot X)^T = (H\mathbf{1}_{[0,T]} \cdot X) = H \cdot X^T.$$
4. For $H \in \mathcal{H}_{simp}$ with representation $H_t = \sum_{n=0}^\infty K_n \mathbf{1}_{(\tau_n, t_{n+1}]}(t)$, we have $H \in L(X)$ and
$$(H \cdot X)_t = \sum_{n=0}^\infty K_n (X_{t \wedge t_{n+1}} - X_{t \wedge \tau_n}).$$

We conclude the lecture on the semimartingale integration with a a very useful version of the dominated convergence theorem for stochastic integration. A stochastic process H is said to be **locally bounded** if it can be reduced to a (uniformly) bounded process, i.e., if there exists a nondecreasing sequence $\{\tau_n\}_{n \in \mathbb{N}}$ of stopping times with $\tau_n \to \infty$, a.s., such that H^{τ_n} is a (uniformly) bounded process.

Problem 19.6. Let X be a continuous semimartingale. Show that

1. Show that each adapted LCRL process is locally bounded, and that
2. each predictable and locally bounded process H is in $L(X)$.
3. Construct an example of an adapted RCLL process which is not locally bounded.

Proposition 19.10 (A Stochastic Dominated Convergence Theorem). *Let $X = M + A$ be a continuous semimartingale, and let $\{H^n\}_{n\in\mathbb{N}}$ be a sequence of progressively-measurable processes with the property that there exists a càglàd process H such that $|(H^n)_t| \leq H_t$ for all $t \geq 0$, a.s., for all $n \in \mathbb{N}$. Then $H^n \in L(X)$ for all $n \in \mathbb{N}$ and*

$$\lim_n H_t^n \to 0 \text{ for all } t \geq 0, \text{ a.s., implies } H^n \cdot X \overset{\text{ucp}}{\to} 0, \text{ as } n \to \infty.$$

Proof. Combine Problems 19.2 and 19.9. □

Additional Problems

Problem 19.7 (Square integrability and convergence via quadratic variation). For $M \in \mathcal{M}_0^{loc,c}$, the following statements hold:

1. M is in $\mathcal{M}_0^{2,c}$ iff $\mathbb{E}[\langle M\rangle_\infty] < \infty$, where $\langle M\rangle_\infty = \lim_{t\to\infty}\langle M\rangle_t$. *Hint:* Use Fatou's lemma to show that $\mathbb{E}[M_t^2] \leq \mathbb{E}[\langle M\rangle_\infty]$, for each t.

2. The limit $\lim_{t\to\infty} M_t$ exists a.s. on $\{\langle M\rangle_\infty < \infty\}$. *Hint:* Define $\tau_n = \inf\{t \geq 0 : \langle M\rangle_t \geq n\}$ and observe that M^{τ_n} converges a.s. Think about the behavior of the sequence $\{\tau_n\}_{n\in\mathbb{N}}$ on $\{\langle M\rangle_\infty < \infty\}$.

Problem 19.8 (The Kunita-Watanabe inequality).

1. Let $F, G : [0,\infty) \to [0,\infty)$ be two non-decreasing continuous functions with $F(0) = G(0) = 0$, and let $L : [0,\infty) \to \mathbb{R}$ be a continuous function of finite variation with the property that $L(0) = 0$ and

$$|L(t) - L(s)| \leq \sqrt{F(t) - F(s)}\sqrt{G(t) - G(s)},$$

for all $0 \leq s \leq t < \infty$. Show that, for any two measurable functions $h, k : [0,\infty) \to \mathbb{R}$, we have

$$\int_0^\infty |h(t)|\,|k(t)|\,d\,|L|\,(t) \leq$$
$$\leq \left(\int_0^\infty h^2(t)\,dF(t)\right)^{1/2}\left(\int_0^\infty k^2(t)\,dG(t)\right)^{1/2},$$

where $|L| : [0,\infty) \to [0,\infty)$ denotes the total-variation of L. *Hint:* Approximate.

2. Prove that, for all $M, N \in \mathcal{M}_0^{loc,c}$, we have $|\langle M,N\rangle_t - \langle M,N\rangle_s| \leq \sqrt{\langle M\rangle_t - \langle M\rangle_s}\sqrt{\langle N\rangle_t - \langle N\rangle_s}$, for all $s \leq t$, a.s., and use it to establish the Kunita-Watanabe inequality. *Hint:* $\langle M + rN\rangle_t \geq \langle M + rN\rangle_s$, a.s., for all rational r.

Problem 19.9 (Continuity of stochastic integration).

1. Let $\{M^n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}_0^{loc,c}$ with the property that

$$\langle M^n\rangle_\infty \to 0 \text{ in probability.}$$

Show that $(M^n)^*_\infty \to 0$ in probability, where $(M^n)^*_\infty = \sup_{t\geq 0} |M^n_t|$.
Hint: Define $\tau_n^\varepsilon = \inf\{t \geq 0 : \langle M^n_t\rangle \geq \varepsilon\}$ and use the fact that $(M^n)^{\tau_n^\varepsilon}$ is in $\mathcal{M}_0^{2,c}$, with the norm bounded by $\sqrt{\varepsilon}$. Use Doob's inequality.

2. Let X be a continuous semimartingale with the semimartingale decomposition $X = X_0 + A + M$. For $H \in L(X)$, we define the $[0,\infty]$-valued random variable

$$[H]_{L(X)} = \int_0^\infty |H_u|\, d\,|A|_u + \int_0^\infty (H_u)^2\, d\langle M\rangle_u.$$

Show the following continuity property of stochastic integration: let $\{H^n\}_{n\in\mathbb{N}}$ be a sequence in $L(X)$ such that $[H^n]_{L(X)} \stackrel{\mathbb{P}}{\to} 0$. Then $(H^n \cdot X)^*_\infty \stackrel{\mathbb{P}}{\to} 0$.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 20
Itô's formula

Itô's formula

Itô's formula is for stochastic calculus what the Newton-Leibnitz formula is for (the classical) calculus. Not only does it relate differentiation and integration, it also provides a practical method for computation of stochastic integrals. There is an added benefit in the stochastic case. It shows that the class of continuous semimartingales is closed under composition with C^2 functions. We start with the simplest version:

Theorem 20.1. *Let X be a continous semimartingale taking values in a segment $[a,b] \subseteq \mathbb{R}$, and let $f : [a,b] \to \mathbb{R}$ be a twice continuously differentiable function ($f \in C^2[a,b]$). Then the process $f(X)$ is a continuous semimartingale and*

$$f(X_t) - f(X_0) = \int_0^t f'(X_u)\, dM_u + \int_0^t f'(X_u)\, dA_u + \tfrac{1}{2}\int_0^t f''(X_u)\, d\langle M\rangle_u. \tag{20.1}$$

Remark 20.2. The Leibnitz notation $\int_0^t f'(X_u)\, dM_u$ (as opposed to the "semimartingale notation" $f'(X)\cdot M$) is more common in the context of the Itô formula. We'll continue to use both.

Before we proceed with the proof, let us state and prove two useful related results. In the first one we compute a simple stochastic integral explicitly. You will immediately see how it differs from the classical Stieltjes integral of the same form.

Lemma 20.3. *For $M \in \mathcal{M}_0^{2,c}$, we have*

$$M \cdot M = \tfrac{1}{2}M^2 - \tfrac{1}{2}\langle M\rangle.$$

Proof. By stopping, we may assume that M and $\langle M\rangle$ are bounded. For a partition $\Delta = \{0 = t_0 < t_1 < \dots\}$, let $\bar{M}^\Delta$ denote the "left-continuous simple approximation"

$$\bar{M}^\Delta_t = M_{t_{k^\Delta(t)}}, \text{ where } k^\Delta(t) = \sup_{k\in\mathbb{N}_0} t_k \le t.$$

Nothing but rearrangement of terms yields that

$$(\bar{M}^\Delta \cdot M)_t = \tfrac{1}{2}(M_t^2 - \langle M \rangle_t^\Delta) \text{ for all } t \geq 0, \tag{20.2}$$

and all partitions $\Delta \in P_{[0,\infty)}$. Indeed, assuming for notational simplicity that Δ is such that $t_n = t$, we have

$$\begin{aligned}(\bar{M}^\Delta \cdot M)_t &= \sum_{k=1}^n M_{t_{k-1}}(M_{t_k} - M_{t_{k-1}}) = \sum_{k=1}^n \Big(\tfrac{1}{2}(M_{t_k} + M_{t_{k-1}}) - \tfrac{1}{2}(M_{t_k} - M_{t_{k-1}})\Big)(M_{t_k} - M_{t_{k-1}}) \\ &= \sum_{k=1}^n \tfrac{1}{2}(M_{t_k}^2 - M_{t_{k-1}}^2) - \sum_{k=1}^n \tfrac{1}{2}(M_{t_k} - M_{t_{k-1}})^2 = \tfrac{1}{2}M_t^2 - \tfrac{1}{2}\langle M \rangle_t^\Delta.\end{aligned} \tag{20.3}$$

By Theorem 18.10, the right-hand side of (20.3) converges to $\frac{1}{2}(M_t^2 - \langle M \rangle_t)$ in $\mathbb{L}^2$, as soon as $\Delta \to \mathrm{Id}$. To show that the limit of the left-hand side converges in $M \cdot M$, it is enough to use the stochastic dominated-convergence theorem (Proposition 19.10). Indeed, the integrands are all, unifromly, bounded by a constant. □

The second preparatory result is the stochastic analogue of the **integration-by-parts formula**. We remind the reader that for two semimartingales $X = M + A$ and $Y = N + C$, we have $\langle X, Y \rangle = \langle M, N \rangle$.

Proposition 20.4. *Let $X = M + A, Y = N + C$ be semimartingale decompositions of two continuous semimartingales. Then XY is a continuous semimartingale and*

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_u \, dX_u + \int_0^t X_u \, dY_u + \langle X, Y \rangle_t. \tag{20.4}$$

Proof. By stopping, we can assume that the processes M, N, A and C are bounded (say, by $K \geq 0$). Moreover, we assume that $X_0 = Y_0 = 0$ - otherwise, just consider $X - X_0$ and $Y - Y_0$. We write XY as $(M + A)(N + C)$ and analyze each term. Using the polarization identity, the product MN can be written as $MN = \frac{1}{2}\left((M+N)^2 - M^2 - N^2\right)$, and Lemma 20.3 implies that

$$M_t N_t = \int_0^t M_u \, dN_u + \int_0^t N_u dM_u + \langle M, N \rangle_t, \tag{20.5}$$

holds for all $t \geq 0$. As far as the FV terms A and C are concerned, the equality

$$A_t C_t = \int_0^t A_u \, dC_u + \int_0^t C_u \, dA_u \tag{20.6}$$

follows by a representation of both sides as a limit of Riemann-Stieltjes sums. Alternatively, you can view the left-hand side of the above equality as the area (under the product measure $dA \times dC$) of the square $[0, t] \times [0, t] \subseteq \mathbb{R}^2$. The right-hand side can also be interpreted as the area of $[0,t] \times [0,t]$ - the two terms corresponding to the areas above and below the diagonal $\{(s,s) \in \mathbb{R}^2 : s \in [0,t]\}$ (we leave it up to the reader to supply the details).

Let us focus now on the mixed term MC. Take a sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ in $P_{[0,\infty)}$ with $\Delta_n = \{t^n_k\}_{k\in\mathbb{N}}$ and $\Delta_n \to \mathrm{Id}$ and write

$$
\begin{aligned}
M_t C_t &= \sum_{k=0}^{\infty} (M_{t\wedge t^n_{k+1}} C_{t\wedge t^n_{k+1}} - M_{t\wedge t^n_k} C_{t\wedge t^n_k}) = I^1_n + I^2_n + I^3_n, \text{ where} \\
I^1_n &= \sum_{k=0}^{\infty} M_{t\wedge t^n_k} (C_{t\wedge t^n_{k+1}} - C_{t\wedge t^n_k}) \\
I^2_n &= \sum_{k=0}^{\infty} C_{t\wedge t^n_k} (M_{t\wedge t^n_{k+1}} - M_{t\wedge t^n_k}) \\
I^3_n &= \sum_{k=0}^{\infty} (M_{t\wedge t^n_{k+1}} - M_{t\wedge t^n_k})(C_{t\wedge t^n_{k+1}} - C_{t\wedge t^n_k})
\end{aligned}
$$

By the properties of the Stieltjes integral (basically, the dominated convergence theorem), we have $I^1_n \to \int_0^t M_u \, dC_u$, a.s. Also, by the uniform continuity of the paths of M on compact intervals, we have $I^3_n \to 0$, a.s. Finally, we note that

$$I^2_n = \int_0^t \bar{C}^{\Delta_n}_u \, dM_u,$$

where $\bar{C}^{\Delta_n}_u = C_{t^{\Delta_n}(u)}$. By uniform continuity of the paths of C on $[0,t]$, $\bar{C}^{\Delta_n} \to C$, uniformly on $[0,t]$, a.s., and $\left|\bar{C}^{\Delta_n}_u - C_u\right| \leq 2C^*_u$, which is an adapted and continuous process. Therefore, we can use the stochastic dominated convergence theorem (Proposition 19.10) to conclude that $I^2_n \to \int_0^t C_u \, dM_u$ in probability, so that

$$M_t C_t = \int_0^t M_u \, dC_u + \int_0^t C_u \, dM_u \tag{20.7}$$

and, in the same way,

$$A_t N_t = \int_0^t A_u \, dN_u + \int_0^t N_u \, dA_u \tag{20.8}$$

The required equality (20.4) is nothing but the sum of (20.5), (20.6), (20.7) and (20.8). □

Remark 20.5. By formally differentiating (20.4) with respect to t, we write

$$d(XY)_t = X_t \, dY_t + Y_t \, dX_t + dX_t dY_t.$$

While meaningless in the strict sense, the "differential" representation above serves as a good mnemonic device for various formulas in stochastic analysis. One simply has to multiply out all the terms, disregard all terms of order larger than 2 (such as $(dX_t)^3$ or $(dX_t)^2 dY_t$), and use the following multiplication table:

	dM	dA
dN	$d\langle M,N\rangle$	0
dC	0	0

where $X = M + A$, $Y = N + C$ are semimartingale decompositions of X and Y. When $M = N = B$, where B is the Brownian motion, the multiplication table simplifies to

	dB	dA
dB	dt	0
dC	0	0

Proof of Theorem 20.1. Let $\mathcal{A}$ be the family of all functions $f \in C^2[a,b]$ such that the formula (20.1) holds for all $t \geq 0$ and all continuous semimartingales X which take values in $[a,b]$. It is clear that $\mathcal{A}$ is a linear space which contains all constant functions. Moreover, it is also closed under multiplication, i.e., $fg \in \mathcal{A}$ if $f,g \in \mathcal{A}$. Indeed, we need to use the integration-by-parts formula (20.4) and the associativity of stochastic integration (Problem 19.5, (2)) applied to $f(X)$ and $g(X)$. Next, the identity is clearly in $\mathcal{A}$, and so, $P \in \mathcal{A}$ for each polynomial P. It remains to show that $\mathcal{A} = C^2[a,b]$. For $f \in C^2[a,b]$, let $\{P''_n\}_{n\in\mathbb{N}}$ be a sequence of polynomials with the property that $P''_n \to f''$, uniformly on $[a,b]$. This can be achieved by the Weierstrass-Stone theorem, because polynomials are dense in $C[a,b]$. It is easy to show that the polynomials $\{P''_n\}_{n\in\mathbb{N}}$ can be taken to be second derivatives of a sequence $\{P_n\}_{n\in\mathbb{N}}$ of polynomials with the property that $P_n \to f$, $P'_n \to f'$ and $P''_n \to f''$, uniformly on $[a,b]$. Indeed, just use $P'_n(x) = f'(x) + \int_a^x P''_n(\xi)\, d\xi$ and $P_n(x) = f(x) + \int_a^x P'_n(\xi)\, d\xi$. Then, as the reader will easily check, the stochastic dominated convergence theorem 19.10 will imply that all terms in (20.1) for P_n converge in probability to the corresponding terms for f. This shows that $C^2[a,b] = \mathcal{A}$. $\square$

Remark 20.6. The proof of the Itô's formula given above is slick, but it does not help much in terms of intuition. One of the best ways of understanding the Itô formula is the following non-rigorous, heuristic, derivation, where $0 = t_0 < t_1 < \dots < t_n < t_{n+1} = t$ is a partiton of $[0,t]$. The main insight is that the second-order term in the Taylor's expansion $f(t) - f(s) = f'(s)(t-s) + \frac{1}{2}f''(s)(t-s)^2 + o((t-s)^2)$ has to be kept and cannot be discarded because - in contast to the classical case - the second (quadratic) variation does not vanish:

$$\begin{aligned} f(X_t) - f(X_0) &= \sum_{k=0}^{n} \left(f(X_{t_{k+1}}) - f(X_{t_k}) \right) \\ &\approx \sum_{k=0}^{n} \Big(f'(X_{t_k})(X_{t_{k+1}} - X_{t_k}) + \tfrac{1}{2} f''(X_{t_k})(X_{t_{k+1}} - X_{t_{k-1}})^2 \Big) \\ &= \sum_{k=0}^{n} f'(X_{t_k})(M_{t_{k+1}} - M_{t_k}) + \sum_{k=0}^{n} f'(X_{t_k})(A_{t_{k+1}} - A_{t_k}) \\ &\quad + \tfrac{1}{2} \sum_{k=0}^{n} f''(X_{t_k})(X_{t_{k+1}} - X_{t_k})^2 \\ &\approx \int_0^t f'(X_u)\, dM_u + \int_0^t f'(X_u)\, dA_u + \tfrac{1}{2} \int_0^t f''(X_u)\, d\langle X \rangle_u. \end{aligned}$$

Finally, one simply needs to remember that $\langle X \rangle = \langle M \rangle$.

The requirements that the semimartingale is one-dimensional, or that it takes values in a compact set $[a, b]$, are not necessary in general. Here is a general version of the Itô formula for continuous semimartingales. We do not give a proof as it is very similar to the one-dimensional case. The requirement that the process takes values in a compact set is relaxed by stopping.

Theorem 20.7. *Let $X^1, X^2, \dots, X^d$ be d continuous semimartingales and let $D \subseteq \mathbb{R}^d$ be an open set, such that $X_t = (X_t^1, \dots, X_t^d) \in D$, for all $t \geq 0$, a.s. Moreover, let $f : D \to \mathbb{R}$ be a twice continuously differentiable function ($f \in C^2(D)$). Then the process $\{f(X_t)\}_{t \in [0,\infty)}$ is a continuous semimartingale and*

$$f(X_t) - f(X_0) = \sum_{k=1}^{d} \int_0^t \frac{\partial}{\partial x_k} f(X_u)\, dX_u^k + \tfrac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(X_u)\, d\langle X^i, X^j \rangle_u. \tag{20.9}$$

Remark 20.8.

1. When the local martingale parts of some of the components of the process $X = (X^1, \dots, X^d)$ vanish, one does not need the full C^2 differentiability in those coordinates; C^1 will be enough.
2. Using the differential notation (and the multiplication table) of Remark 20.5, we can write the Itô formula in the more compact form:

$$df(X_t) = \sum_{k=1}^{n} \frac{\partial}{\partial x_k} f(X_u)\, dX_u^k + \tfrac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} f(X_u)\, dX_u^i dX_u^j.$$

Lévy's characterization of the Brownian motion

We start with a deep (but easy to prove) result which loosely states that if a continuous local martingale has the speed of the Brownian motion, then it is a Brownian motion.

Problem 20.1. Let $f : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be a function in $C^{1,2}([0,\infty) \times \mathbb{R})$ (the space functions continuously differentiable in t and twice continuously differentiable in x). Such a function f is said to be **space-time harmonic** if it satisfies $f_t + \frac{1}{2} f_{xx} = 0$. For $f \in C^{1,2}([0,\infty) \times \mathbb{R})$ and a Brownian motion B, show that the process $f(t, B_t)$ is a local martingale if and only if f is space-time harmonic. Show, additionally, that $f(t, B_t)$ is a martingale if f is space-time harmonic and f_x is bounded on each domain of the form $[0,t] \times \mathbb{R}$, $t \geq 0$.

Theorem 20.9 (Lévy's Characterization of Brownian Motion)**.** *Suppose that $M \in \mathcal{M}_0^{loc,c}$ has $\langle M \rangle_t = t$, for all t, a.s. Then M is an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-Brownian motion.*

Proof. Consider the complex-valued process

$$Z_t^u = \exp(iuM_t + \tfrac{1}{2}u^2 t),\ t \geq 0,$$

for $u \in \mathbb{R}$. We can use Itô's formula (applied separately to its real and imaginary parts) to show that Z^u is a local martingale for each $u \in \mathbb{R}$. Indeed, $Z_t^u = f(t, M_t)$, where $f(t,x) = e^{iux + \frac{1}{2}u^2 t}$, so that, as you can easily check, $f_t(t,x) + \frac{1}{2} f_{xx}(t,x) = 0$. Therefore,

$$\begin{aligned} f(t, M_t) &= f(0, M_0) + \int_0^t f_t(u, M_u)\, du + \int_0^t f_x(u, M_u)\, dM_u + \tfrac{1}{2} \int_0^t f_{xx}(u, M_u)\, d\langle M \rangle_u \\ &= f(0,0) + \int_0^t \left(f_t(u, M_u) + \tfrac{1}{2} f_{xx}(u, M_u)\right) du + \int_0^t f_x(u, M_u)\, dM_u \\ &= 1 + \int_0^t f_x(u, M_u)\, dM_u. \end{aligned}$$

We can, actually, assert that Z^u is a (true) martingale because it is bounded by $\exp(\frac{1}{2}u^2 t)$ on $[0,t]$, and, therefore, of class (DL). Consequently, we have $\mathbb{E}[Z_t^u | \mathcal{F}_s] = Z_s^u$, for all $u \in \mathbb{R}$, $0 \leq s \leq t < \infty$. With a bit of rearranging, we get

$$\mathbb{E}[e^{iu(M_t - M_s)} | \mathcal{F}_s] = e^{-\frac{1}{2}u^2 (t-s)}.$$

In other words, $e^{-\frac{1}{2}u^2(t-s)}$ is a regular conditional characteristic function of $M_t - M_s$, given $\mathcal{F}_s$. That implies that $M_t - M_s$ is independent of $\mathcal{F}_s$ and that its characteristic function is given by $e^{-\frac{1}{2}u^2(t-s)}$, i.e., that $M_t - M_s$ is normal with mean 0 and variance $t - s$. □

Remark 20.10*.* Continuity is very important in the Lévy's characterization. Note that the process $(N_t - t)^2 - t$ is a martingale (prove it!), so that the deterministic process t can be interpreted as a quadratic variation of the martingale $N_t - t$. Clearly, $N_t - t$ is not a Brownian motion.

Problem 20.2. Remember that a stochastic process $\{X_t\}_{t\in[0,\infty)}$ is said to be **Gaussian** if all of its finite-dimensional distributions are multivariate normal. Let $\{M_t\}_{t\in[0,\infty)}$ be a continuous martingale with $\langle M\rangle_t = f(t)$, for some continuous deterministic non-decreasing function $f : [0,\infty) \to [0,\infty)$. Show that $\{M_t\}_{t\in[0,\infty)}$ is Gaussian. Is it true that that each continuous, Gaussian martingale has deterministic quadratic variation?

Itô processes

Definition 20.11. A continuous semimartingale $\{X_t\}_{t\in[0,\infty)}$ is said to be an **Itô process** if there exist progressively measurable processes $\{\alpha_t\}_{t\in[0,\infty)}$ and $\{\beta_t\}_{t\in[0,\infty)}$ such that $\int_0^t \left(|\alpha_s| + \beta_s^2\right) ds < \infty$, a.s., and

$$X_t = X_0 + \int_0^t \alpha_s\, ds + \int_0^t \beta_s\, dB_s. \tag{20.10}$$

Note: This is usually written in differential notation as $dX_t = \alpha_t\, dt + \beta_t\, dB_t$.

Problem 20.3 (Stability of Itô processes)**.** Let X and Y be two Itô processes, and let H be a progressively measurable process in $L(X)$.

1. Show that $H \cdot X$ is an Itô process.
2. Let $F : [0,\infty) \times \mathbb{R}^2 \to \mathbb{R}$ be in the class $C^{1,2,2}$. Show that $Z_t = F(t, X_t, Y_t)$ is also an Itô process.

Definition 20.12. Let X be an Itô process with decomposition (20.10) such that the relation

$$\alpha_t = \mu(t, X_t),\ \beta_t = \sigma(t, X_t),$$

holds for some measurable μ, σ for all $t \geq 0$, a.s. Then, X is said to be an **inhomogeneous diffusion**. If μ and σ do not depend on t, X is said to be a **(homogeneous) diffusion**.

Example 20.13.

1. The Brownian motion $X = B$, the **scaled Brownian motion** $X = \sigma B$ and the **Brownian motion with drift** $X_t = B_t + ct$ (for $c \in \mathbb{R}$) are Itô processes (they are diffusions, as well).
2. The process $|B_t|$ is not an Itô process. Can you prove that? *Hint:* Assume that it is apply Itô's formula on $B_t^2 = |B_t|^2$. Derive an expression for $|B|$ and show that it cannot be non-negative all the time.

   Note: It can be shown, though, that $|B_t|$ is a continuous semimartingale.

3. Let B be a Brownian motion, and let M_t be the process defined by

$$M_t = \begin{cases} 0, & t \leq 1, \\ \mathbf{1}_{\{B_1 > 0\}}, & t > 1. \end{cases}$$

Consider the process X given by

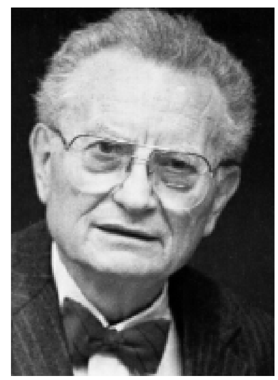
$$X_t = \int_0^t M_u \, du + B_t.$$

In words, X follows a Brownian motion until time 1 and then it gives it an extra drift of speed 1 if and only if $B_1 > 0$. It is clear that X is an Itô process. It is not a diffusion (homogeneous or not). Indeed, by the uniqueness of the semimartingale decomposition, it will be enough to show that M_u cannot be written as a deterministic function $\mu(\cdot, \cdot)$ of t and X_t. If it were possible, there would exist a function $f(x) = \mu(2, x)$ such that $\mathbf{1}_{\{B_1 > 0\}} = f(B_2 + \mathbf{1}_{\{B_1 > 0\}})$.

4. **(Geometric Brownian Motion)** Paul Samuelson proposed the Itô process

$$S_t = \exp\left(\sigma B_t + (\mu - \tfrac{1}{2}\sigma^2)t\right), \; S_0 = s_0 > 0, \tag{20.11}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are parameters, as the model for the time-evolution of the price of a common stock. This process is sometimes referred to as the **geometric Brownian motion (with drift)**. Itô formula yields the following expression for the process S in the differential notation:

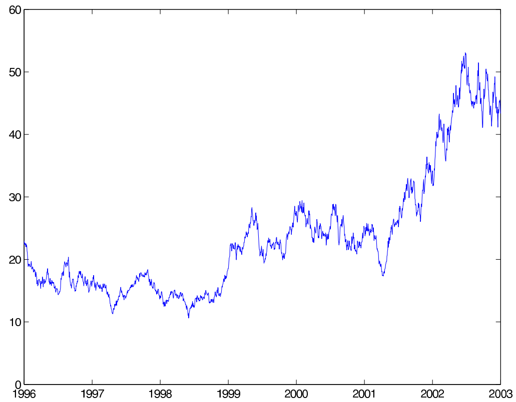


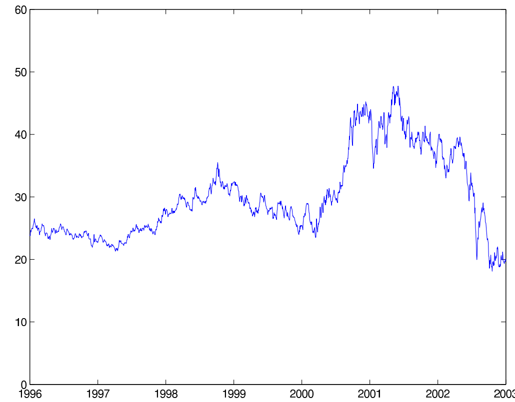
Paul Samuelson

$$\begin{aligned} dS_t &= \exp\left(\sigma B_t + (\mu - \tfrac{1}{2}\sigma^2)t\right)(\mu - \tfrac{1}{2}\sigma^2 + \tfrac{1}{2}\sigma^2)\, dt + \exp\left(\sigma B_t + (\mu - \tfrac{1}{2}\sigma^2)t\right)\sigma \, dB_t \\ &= S_t \mu \, dt + S_t \sigma dB_t. \end{aligned}$$

It follows that S is a diffusion with $\mu(x) = \mu x$ and $\sigma(x) = \sigma x$ (where we use the Greek letters μ and σ to denote both the functions appearing the definition of the diffusion, and the numerical coefficients in (20.12)).

One of the graphs below shows a simulated path of a geometric Brownian motion (Samuelson's model) and the other shows the actual evolution of a price of a certain stock. Can you guess which is which?





To understand the model a bit better, let us start with the case $\sigma = 0$, first. The relationship

$$dS_t = S_t \mu \, dt,$$

is then an example of a linear homogeneous ODE with constant coefficients and its solution describes the exponential growth:

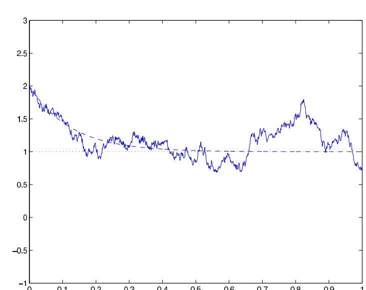
$$S_t = s_0 \exp(\mu t).$$

If we view μ as an interest-rate, S will model the amount of money we will have in the bank after t units of time, starting from s_0 and with the continuously compounded interest at the constant rate of μ.

When $\sigma \neq 0$, we can think of an increment of S as composed of two parts. The first on $S_t \mu \, dt$ models (deterministic) growth, and the second one $S_t \sigma \, dB_t$ the effect of a random fluctuation of size σ. Discretized, it looks like this:

$$\frac{\Delta S_{t_i}}{S_{t_i}} = \frac{S_{t_{i+1}} - S_{t_i}}{S_{t_i}} \approx \mu \Delta t + \sigma \Delta B_t \approx N(\mu dt, (\sigma \sqrt{dt})^2).$$

Therefore, might say that the stock-prices, according to Samuelson, are bank accounts + noise. It makes sense, then, to call the parameter μ the **mean rate of return** and σ the **volatility**.



Ornstein-Uhlenbeck process (solid) and its noise-free (dashed) version, with parameters $m = 1$, $\alpha = 8$, $\sigma = 1$, started at $X_0 = 2$.

5. **(The Ornstein-Uhlenbeck process)** The Ornstein-Uhlenbeck process (often abbreviated to OU-process) is an Itô process with a long history and a number of applications (in physics, related to the **Langevin equation** or in finance, where it drives the **Vasiček model** of interest rates). The OU-process can be defined as the unique process with the property that $X_0 = x_0$ and

$$dX_t = \alpha(m - X_t) \, dt + \sigma dB_t, \tag{20.12}$$

for some constants $m, \alpha, \sigma \in \mathbb{R}$, $\alpha, \sigma > 0$ (we'll show shortly that such a process exists).

Intuitively, the OU-process models a motion of a particle which drifts toward the level m, perturbed with the Brownian noise of intensity σ. Notice how the drift coefficient is negative when $X_t > m$ and positive when $X_t < m$.

Because of that feature, OU-process is sometimes also called a **mean-reverting process**. Another one of its nice properties is that the OU process is a Gaussian process, making it amenable to empirical and theoretical study. In the remainder of this example, we will show how this process can be constructed by using some methods similar to the ones you might have seen in your ODE class. The solution we get will not be completely explicit, since it will contain a stochastic integral, but as equations of this form go, even that is usually more than we can hope for.

First of all, let us simplify the notation (without any real loss of generality) by setting $m = 0$, $\sigma = 1$. Supposing that a process satisfying (20.12) exists, let us start by defining the process $Y_t = R_t X_t$, with $R_t = \exp(\alpha t)$, inspired by the technique one would use to solve the corresponding noiseless equation

$$d\hat{X}_t = -\alpha \hat{X}_t \, dt.$$

Clearly, the process R_t is a deterministic Itô process with $dR_t = \alpha R_t \, dt$, so the integration-by-parts formula (note that $dR_t dB_t = 0$) gives

$$\begin{aligned} dY_t &= R_t \, dX_t + X_t dR_t + dX_t dR_t = -\alpha X_t R_t \, dt + R_t \, dB_t + X_t \alpha R_t \, dt \\ &= R_t \, dB_t = e^{\alpha t} \, dB_t. \end{aligned}$$

It follows now that

$$X_t = x_0 e^{-\alpha t} + e^{-\alpha t} \int_0^t e^{\alpha s} \, dB_s.$$

Problem 20.4. Let $\{X_t\}_{t \in [0,\infty)}$ be an OU process with $m = 0$, $\sigma = 1$ with $\alpha \in \mathbb{R}$, started at $X_0 = x \in \mathbb{R}$. Show that X_t converges in distribution when $t \to \infty$ and find the limiting distribution.

Additional Problems

Problem 20.5 (Bessel processes)**.** For $d \in \mathbb{N}$, $d \geq 2$, let

$$B_t = (B_t^{(1)}, B_t^{(2)}, \dots, B_t^{(d)}), \; t \geq 0,$$

be a d-dimensional Brownian motion, and let the process X, given by

$$X_t = (X_t^{(1)}, X_t^{(2)}, \dots, X_t^{(d)}) = (1, 0, \dots, 0) + (B_t^{(1)}, B_t^{(2)}, \dots, B_t^{(d)}),$$

be a translation of the d-dimensional Brownian motion, started from the point $(1, 0, \dots, 0) \in \mathbb{R}^d$. Finally, let R be given by

$$R_t = ||X_t||, \; t \geq 0,$$

where $||\cdot||$ denotes the usual Euclidean norm in $\mathbb{R}^d$.

1. Use Itô's formula to show that R is a continuous semimartingale and compute its semimartingale decomposition (the filtration is the usual completion of the filtration generated by B.)
2. Supposing that $d = 3$, show that the process $\{L_t\}_{t\in[0,\infty)}$, given by $L_t = R_t^{-1}$, is well-defined and a local martingale.
3. (Quite computational and, therefore, *optional*) Write $\mathbb{E}[L_t]$ as a triple integral and show that the expression inside the integral is dominated by an integrable function for $t \geq 0$. Conclude from there that $\mathbb{E}[L_t] \to 0$, as $t \to \infty$.
4. Given the result in 3. above, can $\{L_t\}_{t\in[0,\infty)}$ be a (true) martingale?

Problem 20.6 (Stochastic exponentials)**.**

1. Show that, for any continuous semimartingale X with $X_0 = 0$, there exists a non-negative continuous semimartingale $\mathcal{E}(X)$ with the property that

   (20.13) $$\mathcal{E}(X)_t = 1 + \int_0^t \mathcal{E}(X)_u \, dX_u, \text{ for all } t \geq 0, \text{ a.s.}$$

   $\mathcal{E}(X)$ is called the **stochastic exponential of** X. *Hint:* Look for $\mathcal{E}(X)$ of the form $\exp(X - A)$, where A is a process of finite variation.
2. If X and Y are two continuous semimartingales starting at 0, relate the processes $\mathcal{E}(X)\mathcal{E}(Y)$ and $\mathcal{E}(X+Y)$. When are they equal? *Note:* You can use, without proof, the fact that process satisfying (20.13) is necessarily unique.
3. Let M be a continuous local martingale with $M_0 = 0$. Show that

   $$\{\mathcal{E}(M)_\infty = 0\} = \{\langle M \rangle_\infty = \infty\}, \text{ a.s.}$$

   Hint: $\mathcal{E}(M) = \mathcal{E}(\tfrac{1}{2}M)^2 e^{-\frac{1}{4}\langle M \rangle}$

Problem 20.7 (Linear Stochastic Differential Equations)**.** Let Y and Z be two continuous semimartingales. Solve (in closed form) the following linear stochastic differential equation for X where $x \in \mathbb{R}$:

$$\begin{cases} dX_t = X_t \, dY_t + dZ_t \\ X_0 = x, \end{cases}$$

i.e., find a continuous semimartingale X with $X_0 = x$ such that $X_t = x + Z_t + \int_0^t X_u \, dY_u$, for all $t \geq 0$, a.s.

Problem 20.8 (Preservation of the local martingale property)**.** Let $\{\mathcal{G}_t\}_{t\in[0,\infty)}$ be a right-continuous and complete filtration such that $\mathcal{G}_t \subseteq \mathcal{F}_t$, for all $t \geq 0$, and let $\{M_t\}_{t\in[0,\infty)}$ be a continuous $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-local martingale which is adapted to $\{\mathcal{G}_t\}_{t\in[0,\infty)}$. Show that $\{M_t\}_{t\in[0,\infty)}$ is also a continuous local martingale with respect to $\{\mathcal{G}_t\}_{t\in[0,\infty)}$.

Problem 20.9 (The distribution of a Brownian functional)**.** Let B, where

$$B_t = (B^{(1)}_t, B^{(2)})_t), t \geq 0,$$

be a 2-dimensional Brownian motion, and let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be the right-continuous augmentation of the natural filtration $\sigma(B^{(1)}_u, B^{(2)}_u; u \leq t)$. Show that $B^{(1)} \in L(B^{(2)})$ and $B^{(2)} \in L(B^{(1)})$, and identify the distribution of the random variable

$$F = \int_0^1 B^{(1)}_u \, dB^{(2)}_u + \int_0^1 B^{(2)}_u \, dB^{(1)}_u.$$

Hint: Compute the characteristic function of F, first. Don't try to build exponential martingales; there is an easier way.

Problem 20.10 (Stochastic area)**.** Let $r = (x, y) : [0, 1] \to \mathbb{R}^2$ be a smooth curve such that

- $r(0)$ is on the positive x-axis and $r(1)$ on the positive y-axis,
- r takes values in the first quadrant and never hits 0.
- r moves counterclockwise, i.e., $x'(t) < 0$, for all t.

Let A denote the region in $[0, \infty) \times [0, \infty)$ bounded by the coordinate axes from below, and by the image of r from above. First heuristically, and then formally, show that the area $|A|$ of A is given by

$$|A| = \tfrac{1}{2} \int_0^1 x(t) \, dy(t) \, dt - \tfrac{1}{2} \int_0^1 y(t) \, dx(t).$$

With the setup as in Problem 20.8, we define the random variable A (note the negative sign)

$$A = \int_0^1 B^{(1)}_u \, dB^{(2)}_u - \int_0^1 B^{(2)}_u \, dB^{(1)}_u.$$

Thanks to 1. above, we can interpret $A(\omega)$ as the double of the (signed) area that the chord with endpoints $(0, 0)$ and $(B^{(1)}_t(\omega), B^{(2)}_t(\omega))$ sweeps during the time interval $[0, 1]$. Use Itô's formula to show that

$$\mathbb{E}[e^{itA}] = \tfrac{1}{\cosh(t)}. \tag{20.14}$$

Note: The formula (20.14) if often known as **Lévy's stochastic area formula**. The distribution with the characteristic function apearing in it can be shown to admit a density of the form $f_A(x) = \frac{1}{2}(\cosh(\pi x/2))^{-1}$.

Problem 20.11 (Speed measures)**.** A continuous and strictly increasing function $s : \mathbb{R} \to \mathbb{R}$ is said to be a **scale function** for the diffusion $dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dB_t$ if $s(X_t)$ is a local martingale.

1. Suppose that the function $x \mapsto \frac{\mu(x)}{\sigma^2(x)}$ is locally integrable, i.e., that $\int_a^b \frac{|\mu(x)|}{\sigma^2(x)}\,dx < \infty$, for all $[a,b] \subseteq \mathbb{R}$. Show that a scale function exists.

2. Let $[u,d]$ be an interval with $X_0 \in [u,d]$ and $\mathbb{P}[\tau_u \wedge \tau_d = \infty] = 0$, where $\tau_y = \inf\{t \geq 0 : X_t = y\}$, for $y \in \mathbb{R}$. If a scale function s exists and $X_0 = x \in \mathbb{R}$, show that

$$\mathbb{P}[\tau_u < \tau_d] = \frac{s(x) - s(d)}{s(u) - s(d)}.$$

3. Let $X_t = B_t + ct$ be a Brownian motion with drift. For $u > 0$ an $d = 0$, show that $\mathbb{P}[\tau_u \wedge \tau_d = \infty] = 0$ and compute $\mathbb{P}[\tau_d < \tau_u]$.

4. For $c \in \mathbb{R}$, find the probability that $X_t = B_t + ct$ will never hit the level 1.

Problem 20.12 (One unit of activity)**.** Let $\{B_t\}_{t \in [0,\infty)}$ be a standard Brownian Motion and let $\{H_t\}_{t \in [0,\infty)}$ be a predictable process for the standard augmentation of the natural filtration of B. Given that

$$\forall\, t \geq 0,\ \int_0^t H_u^2\,du < \infty, \text{ and } \int_0^\infty H_u^2\,du = \infty.$$

identify the distribution of $\int_0^\tau H_u dB_u$, where

$$\tau = \inf\{t \geq 0 : \int_0^t H_u^2\,du = 1\}.$$

Problem 20.13 (The Lévy transform)**.** Let $(B_t)_{t \in [0,\infty)}$ be a Brownian motion, and let $(X_t)_{t \geq 0}$ be its *Lévy transform*, i.e., the process defined by

$$X_t = \int_0^t \operatorname{sign}(B_u)\,dB_u, \text{ where } \operatorname{sign}(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases}$$

1. Show that X is a Brownian motion,

2. show that the random variables B_t and X_t are uncorrelated, for each $t \geq 0$, and

3. show that B_t and X_t are not independent for $t > 0$. *Hint:* Compute $\mathbb{E}[X_t B_t^2]$.

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

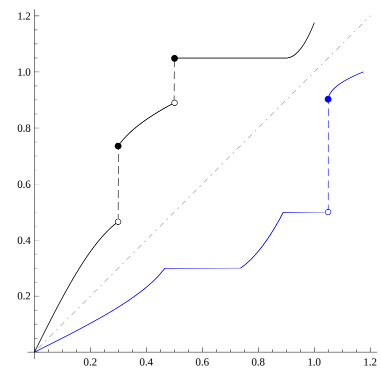
Lecture 21
Representations of Martingales

Right-continuous inverses

Let A_0 denote the set of all càdlàg and non-decreasing function defined on $[0,\infty)$ and taking values in $[0,\infty]$, with $f(0) = 0$. For $f \in A_0$, we also set $f(\infty) = \lim_{t\to\infty} f(t)$.

Even it does not need to be invertible, a function f in A_0 admits an a **right-continuous inverse**, namely the function $g : [0,\infty) \to [0,\infty]$, denoted by $g = f^{-1}$, and given by

$$g(s) = \inf\{t \geq 0 : f(t) > s\}.$$



Right-continuous inverses of each other.

The picture on the right shows a (portion of) the graph of a typical function in A_0 (black), as well as its right-continuous inverse (blue). Note how jumps correspond to flat stretches and vice versa, and how the non-injectivity is resolved to yield right continuity. As always $h(t_-) = \lim_{t'\to t, t'<t} h(t')$, for $t > 0$ and $h(0-) = 0$, for any càdlàg function h on $[0,\infty)$.

Here is a precise statement of some of the properties of right-continuous inverses. This is best understood by looking at the picture above, so no proof is given (to practice real analysis, supply the proof yourself).

Proposition 21.1. *For $f \in A_0$ and its right-continuous inverse $g = f^{-1}$, we have*

1. $g \in A_0$,
2. *f is the right-continuous inverse of g,*
3. $g(f(t)) = \sup \mathrm{Flat}_f(t)$, *for each $t \in [0,\infty)$, where* $\mathrm{Flat}_f(t) = \{t' \geq 0 : f(t) = f(t')\}$.
4. *Similarly,* $f(g(s)) = \sup \mathrm{Vert}_f(s)$, *for each $s \in f([0,\infty))$, where* $\mathrm{Vert}_f(s)$ *is the set of all values in the (unique) interval of the form $[f(t-), f(t)]$ which contains s. In particular, if f is continuous, then $f(g(s)) = s$, for all $s \geq 0$.*

Even though we will have no further use for it in these notes, the result of following problem (a change-of-variable formula) comes in handy from time to time. It is a mild generalization of the formula we use to compute expectations of functions of random variables (how?).

Problem 21.1. Show that, for $f \in A_0$, and a non-negative Borel function φ on $[0,\infty)$, we have

$$\int_0^\infty \varphi(t)\, df(t) = \int_0^\infty \varphi(g(s)) \mathbf{1}_{\{g(s)<\infty\}} ds,$$

where $g = f^{-1}$ is the right-continuous inverse of f. Deduce that, $0 \le a < b < \infty$, and a non-decreasing and continuous function $u : [a,b] \to [0,\infty)$, we have

$$\int_a^b \varphi(u(s))\, df(u(s)) = \int_{u(a)}^{u(b)} \varphi(t)\, df(t).$$

Time-changes

Definition 21.2. A (not-necessarily adapted) stochastic process $\{\tau_s\}_{s\in[0,\infty)}$ with trajectories in A_0 is called a **change of time** or **time change** if the random variable τ_s is a stopping time, for each $s \ge 0$.

Given a filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ and a predictablely measurable process $\{X_t\}_{t\in[0,\infty)}$, the composition X_{τ_s} defines a random variable for each $s \ge 0$; the stochastic process $\{X_{\tau_s}\}_{s\in[0,\infty)}$ is called the **time change of $\{X_t\}_{t\in[0,\infty)}$ by $\{\tau_s\}_{s\in[0,\infty)}$**. Similarly, we may define the **time-changed** filtration $\{\mathcal{G}_s\}_{s\in[0,\infty)}$ by $\mathcal{G}_s = \mathcal{F}_{\tau_s}$. By Problem 16.4, we have the following:

- $\{\mathcal{G}_s\}_{s\in[0,\infty)}$ is right-continuous, since $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ is (we always assume that, remember). Same for completeness.
- the time-changed process $\{X_{\tau_s}\}_{s\in[0,\infty)}$ is $\{\mathcal{G}_s\}_{s\in[0,\infty)}$ adapted.
- If $\{X_t\}_{t\in[0,\infty)}$ happens to be càdlàg then so is $\{X_{\tau_s}\}_{s\in[0,\infty)}$. Is the same true for continuous (or càglàd) processes?

Even though we are asking only for τ_s to be a stopping time for deterministic $s \ge 0$, this property extends to the class of all $\{\mathcal{G}_s\}_{s\in[0,\infty)}$-stopping times:

Proposition 21.3. *Then the random variable τ_σ is an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-stopping time, as soon as σ is a $\{\mathcal{G}_s\}_{s\in[0,\infty)}$-stopping time.*

Proof. It is enough to deal with countable-valued stopping times. Indeed, a general stopping time σ can be approximated from the right by a countable-valued sequence $\{\sigma_n\}_{n\in\mathbb{N}}$ of stopping times such that $\sigma_n \searrow \sigma$, so that, by right contiuity of τ, we have $\tau_{\sigma_n} \searrow \tau_\sigma$, a.s. The right continuity of the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ takes care of the rest.

We turn to a countable-valued stopping time σ of the form $\sigma = \sum_{k=1}^\infty s_k \mathbf{1}_{A_k}$, where $A_k \in \mathcal{G}_{s_k} = \mathcal{F}_{\tau_{s_k}}$, for $k \in \mathbb{N}$ and $A_1, A_2, \dots$ form a partition of Ω. For $t \geq 0$, we have

$$\{\tau_\sigma \leq t\} = \bigcup_{k=1}^\infty \Big(\{\tau_{s_k} \leq t\} \cap A_k \Big).$$

Since $A_k \in \mathcal{F}_{\tau_{s_k}}$, we have $A_k \cap \{\tau_{s_k} \leq t\} \in \mathcal{F}_t$, for each $k \in \mathbb{N}$, and so $\{\tau_\sigma \leq t\} \in \mathcal{F}_t$. □

Definition 21.4. Let $\{\tau_s\}_{s\in[0,\infty)}$ be a time change. A process $\{X_t\}_{t\in[0,\infty)}$ is said to be τ**-continuous** if it is continuous and X is constant on $[\tau_{s-}, \tau_s]$, for all $s \geq 0$, a.s.

It is clear that $\{X_{\tau_s}\}_{s\in[0,\infty)}$ is a continuous process if $\{\tau_s\}_{s\in[0,\infty)}$ is a time change and $\{X_t\}_{t\in[0,\infty)}$ is a τ-continuous process. A deeper property of τ-continuity is that it preserves martingality. Before we state the precise result, let us show what can go wrong:

Example 21.5. Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be the right-continuous augmentation of its natural filtration. We define the family $\{\tau_s\}_{s\in[0,\infty)}$ of stopping times by

$$\tau_s = \inf\{t \geq 0 : B_t > s\}, \; s \geq 0.$$

To show that $\{\tau_s\}_{s\in[0,\infty)}$ is a time change, we simply express its paths as right-continuous inverses of the continuous and non-decreasing process $\{M_t\}_{t\in[0,\infty)}$, where $M_t = \sup_{s\leq t} B_s$. By the continuity of M and the part (6) of Proposition 21.1, we have $B_{\tau_s} = s$, for all $s \geq 0$, a.s. Therefore, we have managed to time-change a martingale B into a constant, finite-variation process s. Note that the time change τ is by no means continuous, as all flat stretches of M correspond to jumps in τ. In fact, it can be shown that, in some sense, τ grows by jumps only.

Proposition 21.6. *Let $\{\tau_s\}_{s\in[0,\infty)}$ be a time change, and let $M \in \mathcal{M}_0^{loc,c}$ be a τ-continuous local martingale. Then the time-changed process $\{M_{\tau_s}\}_{s\in[0,\infty)}$ is a continuous local martingale for the time-changed filtration $\{\mathcal{G}_s\}_{s\in[0,\infty)} = \{\mathcal{F}_{\tau_s}\}_{s\in[0,\infty)}$.*

Proof. By τ-continuity, the process $\{N_s\}_{s\in[0,\infty)}$, given by $N_s = M_{\tau_s}$ is a continuous process adapted to the filtration $\{\mathcal{G}_s\}_{s\in[0,\infty)}$. Given an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-stopping time T such that M^T is bounded, we set

$$S = \inf\{s \geq 0 : \tau_s \geq T\},$$

and show that that N^S is a $\{\mathcal{G}_s\}_{s\in[0,\infty)}$-martingale. By τ-continuity of M, we have

$$N_{s\wedge S} = M_{\tau_{s\wedge S}} = M_{\tau_{(s\wedge S)-}},$$

and, since $\tau_{(s\wedge S)-} \leq T$, we conclude that N^S is bounded. Next, we pick a $\{\mathcal{G}_s\}_{t\in[0,\infty)}$-stopping time σ and note that the random variable $\tau_{(\sigma\wedge S)}$ is, according to Proposition 21.3, an $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-stopping time. Also, even though $\tau_{\sigma\wedge S}$ is not necessarily bounded from above by T, the martingale $M^{\tau_{\sigma\wedge S}}$ remains bounded by the same constant as M^T, so we can use the optional sampling theorem to get

$$\mathbb{E}[N^S_\sigma] = \mathbb{E}[N_{\sigma\wedge S}] = \mathbb{E}[M_{\tau_{(\sigma\wedge S)}}] = \mathbb{E}[M_{\tau_0}] = \mathbb{E}[N_0] = \mathbb{E}[N^S_0],$$

which implies that N^S is a martingale.

We can repeat the procedure for a reducing sequence $\{T_n\}_{n\in\mathbb{N}}$ with M^{T_n} bounded, to conclude that N is, indeed, a local martingale. □

We have seen that the class of local martingales is closed under time changes only if additional conditions are fulfulled. The situation is quite diferent with semimartingales. We state two important results the proofs of which are outside the scope of these notes. Both results deal with càdlàg semimartingales, i.e., with processes which can be decomposed into sums of a càdlàg local martingale and a càdlàg and adapted process of finite variation.

Theorem 21.7. *Let $\{X_t\}_{t\in[0,\infty)}$ be a càdlàg semimartingale, and $\{\tau_s\}_{s\in[0,\infty)}$ a time-change. Then the time-changed process $\{X_{\tau_s}\}_{s\in[0,\infty)}$ is a càdlàg semimartingale (not necessarily continuous even if X is).*

Theorem 21.8 (Monroe)**.** *Let $\{X_s\}_{s\in[0,\infty)}$ be a càdlàg semimartingale. Then there exist a filtered probability space $(\Omega', \mathcal{F}', \{\mathcal{F}'_t\}_{t\geq 0}, \mathbb{P}')$ which supports an $\{\mathcal{F}'_t\}_{t\in[0,\infty)}$-Brownian motion $\{B_t\}_{t\in[0,\infty)}$ and a time-change $\{\tau_s\}_{s\in[0,\infty)}$ such that $\{X_s\}_{s\in[0,\infty)}$ and $\{B_{\tau_s}\}_{s\in[0,\infty)}$ have the same (finite-dimensional) distributions.*

A theorem of Dambis, Dubins and Schwarz

As we have seen in Example 21.5, a right-continuous inverse (computed ω-wise) of an adapted, càdlàg and non-decreasing process is clearly a time change. The most important example of such a time-change for our purposes is obtained by taking a right-continuous inverse of the quadratic variation process of a continuous local martingale $\{M_t\}_{t\in[0,\infty)}$:

$$\tau_s = \inf\{t \geq 0 : \langle M\rangle_t > s\}.$$

We note that the process τ will not take the value ∞ if the local martingale $\{M_t\}_{t\in[0,\infty)}$ is **divergent**, i.e., if $\langle M\rangle_\infty = \infty$, a.s.

Lemma 21.9. *M is τ-continuous.*

Proof. By stopping, we may assume that both M and $\langle M\rangle$ are bounded. We pick a rational number $r \geq 0$, and define the process $N_t = M_{r+t} - M_r$, $t \geq 0$, which is clearly a martingale with $\langle N\rangle_t = \langle M\rangle_{r+t} - \langle M\rangle_t$. The random variable $T_r = \inf\{t \geq 0 : \langle N\rangle_t > 0\}$ is a stopping time, and therefore, the stopped martingale N^{T_r} is in $\mathcal{M}_0^{2,c}$ with $\langle N^{T_r}\rangle_t = 0$, for all $t \geq 0$. Therefore, $N_t^{T_r} = 0$, for all, $t \geq 0$, and so, M is constant on $[q, T_q]$. It is not hard to see that any interval of constancy of M is a closure of a countable union of intervals of the form $[r, t + T_r]$, and our claim follows. □

Theorem 21.10 (Dambis, Dubins and Schwarz)**.** *For a divergent $M \in \mathcal{M}_0^{loc,c}$, we define*

$$\tau_s = \inf\{t \geq 0 : \langle M\rangle_t > s\}, \text{ and } \mathcal{G}_s = \mathcal{F}_{\tau_s}, \text{ for } s \geq 0.$$

Then, the time-changed process $\{B_s\}_{s\in[0,\infty)}$, given by

$$B_s = M_{\tau_s},\ s \geq 0,$$

is a $\mathcal{G}$-Brownian motion and the local martingale M is a time-change of B, i.e.

$$M_t = B_{\langle M\rangle_t}, \text{ for } t \geq 0.$$

Proof. By Lemma 21.9 and Proposition 21.6, B is a continuous local martingale. To compute its quadratic variation, we start from the fact that $M_t^2 - \langle M\rangle_t$ is a τ-continous local martingale, and, therefore, time-change $M_{\tau_s}^2 - \langle M\rangle_{\tau_s}$ is a continuous process. Since $\langle M\rangle_{\tau_s} = s$, the process $B_s^2 - s$ is a $\{\mathcal{G}_s\}_{s\in[0,\infty)}$-local martingale, and so, $\langle B\rangle_s = s$. Lévy's characterization (Theorem implies that $\{B_s\}_{t\in[0,\infty)}$ is a Brownian motion.

The fact that $M_t = B_{\langle M\rangle_t}$ follows quite directly from Proposition 21.1, part 3. Indeed, $B_{\langle M\rangle_t} = M_{\tau_{\langle M\rangle_t}}$ and $\tau_{\langle M\rangle_t}$ is the the right edge of $\text{Flat}_{\langle M\rangle}(t)$. Since M is $\tau_{\langle M\rangle}$-continuous, we have $M_t = M_{\tau_{\langle M\rangle_t}} = B_{\langle M\rangle_t}$. □

Remark 21.11. The restriction that M be divergent is here mostly for convenience. It can be shown (try to prove it), that in that case M can still be written as $M_t = B_{\langle M\rangle_t}$, $t \geq 0$, where B is a Brownian motion. This time, however, B may have to be defined on an extension of the original probability space. The reason is that we "do not have enough of M" to construct the whole path of B from.

Martingals as stochastic integrals

Let $\{M_t\}_{t\in[0,\infty)}$ be a continuous local martingale. A continuous local martingale $\{Z\}_{t\in[0,\infty)}$ is called the **stochastic (Doléans-Dade) exponential** of M, denoted by $Z = \mathcal{E}(M)$, if

$$Z_t = 1 + \int_0^t Z_u \, dM_u, \text{ for all } t \geq 0, \text{ a.s.} \tag{21.1}$$

Itô's formula implies readily that the following prescription

$$\mathcal{E}(M)_t = \exp(M_t - \tfrac{1}{2}\langle M\rangle_t), t \geq 0,$$

defines a stochastic exponential of M. It can be shown that the process $\mathcal{E}(M)$ is the only such process, i.e., that the **stochastic integral equation** (21.1) has a unique solution (in a specific sense).

Exponential martingales, together with the following simple lemma (whose easy proof we leave to the reader), and the proposition that follows it, play a central role in the proof of the celebrated martingale-representation theorem below. We remind the reader that subset E of $\mathbb{L}^2$ is said to be **total** if its linear hull

$$\Big\{ \sum_{k=1}^n \alpha_k X_k \, : \, n \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{R}, X_1, \dots, X_k \in E \Big\}$$

is dense in $\mathbb{L}^2(\mathcal{F})$. Also, for $E \subseteq \mathbb{L}^2$, its **orthogonal complement** $E^\perp$ is given by

$$E^\perp = \{X \in \mathbb{L}^2(\mathcal{F}) \, : \, \mathbb{E}[XY] = 0 \text{ for all } Y \in E\} = \{0\}.$$

Lemma 21.12. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let E be a subset of $\mathbb{L}^2(\mathcal{F})$ such that $E^\perp = \{0\}$. Then, E is total in $\mathbb{L}^2(\mathcal{F})$.*

Proposition 21.13. *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, and let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be the usual augmentation of its natural filtration. With $\mathcal{I}$ denoting the set of all (deterministic!) functions $f : [0,\infty) \to \mathbb{R}$ of the form*

$$f(t) = \sum_{k=1}^n \lambda_k \mathbf{1}_{(t_{k-1}, t_k]}(t), \tag{21.2}$$

for some $n \in \mathbb{N}$, $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ and $0 = t_0 < t_1 < \dots < t_n < \infty$, the set

$$E = \Big\{ \mathcal{E}(\textstyle\int_0^\cdot f(u)\, dB_u)_\infty \, : \, f \in \mathcal{I} \Big\}$$

is total in $\mathbb{L}^2(\mathcal{F}_\infty)$.

Proof. Suppose that $Y \in \mathbb{L}^2(\mathcal{F}_\infty)$ is such that $\mathbb{E}[YX] = 0$, for all $X \in E$, i.e., Then, given a finite partition $0 = t_0 < t_1 < \dots < t_n < \infty$ and f as in (21.2), we have

$$0 = \mathbb{E}[\mathcal{E}(\int_0^\cdot f(u)\, dB_u)_\infty Y]$$
$$= \mathbb{E}\left[\exp\left(\sum_{k=1}^n \lambda_k (B_{t_k} - B_{t_{k-1}}) - \tfrac{1}{2}\sum_{k=1}^n \lambda_k^2 (t_k - t_{k-1})\right) Y\right].$$

By rearrangement and conditioning, it follows that

$$\mathbb{E}\left[\exp\left(\sum_{k=1}^n \alpha_k B_{t_k}\right) Z^+\right] = \mathbb{E}\left[\exp\left(\sum_{k=1}^n \alpha_k B_{t_k}\right) Z^-\right],$$

for all $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, where

$$Z = \mathbb{E}[Y|\sigma(B_{t_1}, \dots, B_{t_n})].$$

Given that Z is $\sigma(B_{t_1}, \dots, B_{t_n})$-measurable, there exists a Borel function ζ such that $Z = \zeta(B_{t_1}, \dots, B_{t_n})$. Consequently

$$\int_{\mathbb{R}^n} e^{\sum_{k=1}^n \alpha_k x_k}\, \zeta^+(x_1, \dots, x_n)\varphi(x_1, \dots, x_n) dx_1, \dots, dx_n =$$
$$\int_{\mathbb{R}^n} e^{\sum_{k=1}^n \alpha_k x_k}\, \zeta^-(x_1, \dots, x_n)\varphi(x_1, \dots, x_n) dx_1, \dots dx_n,$$

for all $\alpha_1, \dots, \alpha_n$, where φ is the density of $(B_{t_1}, \dots, B_{t_n})$. Both integrals above are Laplace transforms, one of $\zeta^-\varphi$ and the other of $\zeta^+\varphi$, and, since they agree, so must the transformed functions, i.e., $\zeta = 0$, a.e., with respect to the Lebesgue measure on $\mathbb{R}^n$. It follows that $\mathbb{E}[Y|\mathcal{G}] = 0$, i.e., that

$$\mathbb{E}[Y^+ \mathbf{1}_A] = \mathbb{E}[Y^- \mathbf{1}_A],$$

for each $A \in \sigma(B_{t_1}, \dots, B_{t_n})$. Since the union of all such "finite-dimensional" σ-algebras is a π-system which generates $\mathcal{F}_T$, we can conclude that $Y = 0$, a.s., and Lemma 21.12 applies. □

Proposition 21.14. *Let $\{B_t\}_{t\in[0,\infty)}$ be a Brownian motion, let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be the usual augmentation of its natural filtration, and let $\mathcal{F}_\infty = \sigma(\mathcal{F}_t; t \geq 0)$. For any random variable $X \in \mathbb{L}^2(\mathcal{F}_\infty)$ there exists a $\lambda \times \mathbb{P}$-a.e.-unique predictable process $\{H_t\}_{t\in[0,\infty)}$, such that $\mathbb{E}[\int_0^\infty H_u^2\, du] < \infty$ and*

$$X = \mathbb{E}[X] + \int_0^\infty H_u\, dB_u, \text{ a.s.} \tag{21.3}$$

Proof. Let us deal with uniqueness, first. Suppose that (21.3) holds for two predictable processes H and K. Then

$$0 = \mathbb{E}\left[\left(\int_0^\infty H_u\, dB_u - \int_0^\infty K_u\, dB_u\right)^2\right] = \mathbb{E}[\int_0^\infty (H_u - K_u)^2\, du],$$

and it follows that $H = K$, $\lambda \otimes \mathbb{P}$-a.e., Next, we deal with existence and start by defining the subset $\mathcal{H}$ of $\mathbb{L}^2(\mathcal{F}_\infty)$ by

$$\mathcal{H} = \{\int_0^\infty H_u \, dB_u \,:\, H \text{ is predictable and } \mathbb{E}[\int_0^\infty H_u^2 \, du] < \infty\}.$$

Let $\{Y_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ - with $Y_n = \int_0^\infty H_u^n \, dB_u$ - which converges to some $Y \in \mathbb{L}^2(\mathcal{F}_\infty)$. By completenes of $\mathbb{L}^2(\mathcal{F}_\infty)$, $\{Y_n\}_{n\in\mathbb{N}}$ is Cauchy, thus, thanks to Itô's isometry, the sequence $\{H^n\}_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{L}^2([0,\infty) \times \Omega) = \mathbb{L}^2([0,\infty) \times \Omega, \mathcal{P}, \lambda \times \mathbb{P})$, where $\mathcal{P}$ denotes the predictable σ-algebra. By completeness of $\mathbb{L}^2([0,\infty) \times \Omega)$, $H_n \to H$, for some $H \in \mathbb{L}^2([0,\infty) \times \Omega)$. Consequently, the

$$\int_0^\infty H_u \, dB_u = \lim_n \int_0^\infty H_u^n \, dB_u = \lim_n Y_n = Y, \text{ a.s.}$$

Therefore, $\mathcal{H}$ is a closed subspace of $\mathbb{L}^2(\mathcal{F}_\infty)$. Moreover, the representation (21.1) of the random variables $\mathcal{E}(\int_0^\cdot f(u) \, dB_u)_\infty$, $f \in \mathcal{I}$, (defined in the statement of Proposition 21.13) tells us that

$$\mathcal{E}(\int_0^\cdot f(u) \, dB_u)_\infty - 1 \in \mathcal{H}, \text{ for all } f \in \mathcal{I}.$$

It remains now to use Proposition 21.13 to conclude that the linear span

$$\operatorname{Span}(\mathcal{H}, 1) = \Big\{ x + Y \,:\, x \in \mathbb{R}, \; Y \in \mathcal{H} \Big\}$$

in $\mathbb{L}^2(\lambda \times \mathbb{P})$, generated by $\mathcal{H}$ and the constant 1, is dense in $\mathbb{L}^2(\lambda \times \mathbb{P})$, because it contains a total set. On the other hand, by the first part of the proof, $\operatorname{Span}(\mathcal{H}, 1)$ is closed in $\mathbb{L}^2(\lambda \times \mathbb{P})$, and, so, $\operatorname{Span}(\mathcal{H}, 1) = \mathbb{L}^2(\lambda \times \mathbb{P})$. It follows that each $X \in \mathbb{L}^2(\mathcal{F}_\infty)$ can be written as

$$X = x_0 + \int_0^\infty H_u \, dB_u,$$

for some predictable H with $\mathbb{E}[\int_0^\infty H_u^2 \, du] < \infty$, and some constant $x_0 \in \mathbb{R}$. Taking expectations of both sides yields that $x_0 = \mathbb{E}[X]$. $\square$

Even though, we have almost no use of non-continuous (RCLL) martingales in these notes, sometimes the show up naturally, even in the continuous-path framework. They are defined just like the continuous martingales - the only difference is that their trajectories are assumed to be RCLL. A deeper difference comes with the possible choices of the reducing sequence. In the continuous-path case this sequence can always be constructed to so that the stopped processed are bounded. That is no longer the case in the RCLL setting, and only uniform integrability, as required in the definition, can be guaranteed.

Corollary 21.15. *Let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be the usual agumentation of the Brownian filtration, and let M be an $\mathbb{L}^2$-bounded, càdlàg $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-martingale. Then, there exists a predictable process H with $\mathbb{E}[\int_0^\infty H_u^2\,du] < \infty$ such that*

$$M_t = M_0 + \int_0^t H_u\,dB_u, \text{ for all } t \geq 0, \text{ a.s.}$$

Proof. By Proposition 21.14, we can express the last element M_∞ of M as

$$M_\infty = M_0 + \int_0^\infty H_u\,dB_u,$$

for some predictable H such that $\mathbb{E}[\int_0^\infty H_u^2\,du] < \infty$. Let the continuous martingale N be defined as

$$N_t = M_0 + \int_0^t H_u\,dB_u,$$

so that $N_\infty = M_\infty$. It follows that $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t] = \mathbb{E}[N_\infty|\mathcal{F}_t] = N_t$, a.s., for all $t \geq 0$, and right continuity implies that M and N are indistinguishable. □

Proposition 21.16. *Let $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ be the usual augmentation of the Brownian filtration, and let $\{M_t\}_{t\in[0,\infty)}$ be a càdlàg $\{\mathcal{F}_t\}_{t\in[0,\infty)}$-local martingale. Then, there exists a predictable process H in $L(B)$ such that*

$$M_t = M_0 + \int_0^t H_u\,dB_u, \text{ for all } t \geq 0, \text{ a.s.}$$

In particular, each local martingale in the augmented Brownian filtration is continuous.

Proof. By stopping, we can reduce the statement to the case of an uniformly-integrable martingale (note that, as explained above, we cannot necessarily assume that M is a bounded martingale). Being UI, M admits the last element M_∞ and we can approximate it in $\mathbb{L}^2$. More precisely, given $\varepsilon > 0$, there exists $M_\infty^\varepsilon \in \mathbb{L}^2(\mathcal{F}_\infty)$ such that

$$||M_\infty^\varepsilon - M_\infty||_{\mathbb{L}^1} < \varepsilon.$$

Using M_∞^ε as the last element, We define the square-integrable martingale M^ε by

$$M_t^\varepsilon = \mathbb{E}[M_\infty^\varepsilon|\mathcal{F}_t], \text{ for } t \geq 0,$$

and take a càdlàg modification (remember, we can do that for any martingale on a filtration satisfying the usual conditions). By the maximal inequality for martingales, we have

$$\mathbb{P}\left[\sup_{t\geq 0} |M_t^n - M_t^\varepsilon| \geq \delta\right] \leq \tfrac{1}{\delta}\mathbb{E}[|M_\infty^n - M_\infty^\varepsilon|] \leq \tfrac{1}{\delta}\varepsilon. \tag{21.4}$$

Corollary 21.15 implies that M^{ε} is continuous, and we can interpret the equation (21.4) in the following way: with probability at least $1-\varepsilon/\delta$, the trajectory of M is in a δ-neighborhood of a continuous function. That means, in particular, that

$$\mathbb{P}[(\Delta M)^*_\infty \geq 2\delta] \leq \tfrac{1}{\delta}\varepsilon, \tag{21.5}$$

where the jump process ΔM is defined by $\Delta M_t = M_t - M_{t-}$. The left-hand side of (21.5) does not depend on ε, so, we conclude that $(\Delta M)^*_\infty \leq 2\delta$, a.s., for all $\delta > 0$, making M continuous.

Now that we know that M is continuous, we can reduce it by stopping to a square-integrable martingale and finish the proof. □

An explicit example

Martingale representations of random variables (and martingales) are seldom explicit. Here is an example of a nontrivial one where everything can be worked out closed-form.

Example 21.17. Let B be a Brownian motion and $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ its (augmented) filtration. With $S_t = \sup_{u\leq t} B_u$ denoting the running-maximum process, the random variable S_1 is in $\mathbb{L}^2(\mathcal{F}_1)$ so Proposition 21.14 guarantees that there exists a predictable process $\{H_t\}_{t\in[0,\infty)}$ in $\mathbb{L}^2(B)$ such that (note that, necessarily, $H_t = 0$ for $t > 1$)

$$S_1 = \mathbb{E}[S_1] + \int_0^1 H_u\, dB_u, \text{ a.s.}$$

Proposition 21.14 itself is silent about the exact form of H. We start by defining a martingale $\{M_t\}_{t\in[0,1]}$, in its continuous modification, by

$$M_t = \mathbb{E}[S_1|\mathcal{F}_t], \text{ a.s.},$$

so that $M_t = \int_0^t H_u\, dB_u$, for all $t \in [0,1]$, a.s. On the other hand, we have

$$\mathbb{E}[S_1|\mathcal{F}_t] = \mathbb{E}\Big[\max\Big(S_t, B_t + \sup_{u\in[t,1]}(B_u - B_t)\Big)\Big|\mathcal{F}_t\Big], \text{ a.s.}$$

The Strong Markov Property implies that $\sup_{u\in[t,1]}(B_u - B_t)$ is independent of $\mathcal{F}_t$ and distributed as S_{1-t}, so we have

$$\mathbb{E}[S_1|\mathcal{F}_t] = F(t, S_t, B_t), \text{ a.s.},$$

where, for $s \geq b$, we define

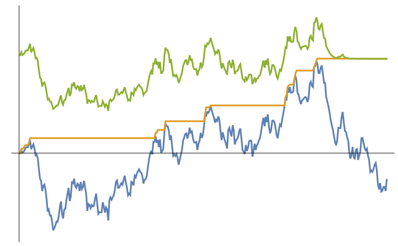
$$F(t,s,b) = \mathbb{E}[\max(s, b + S_{1-t})] \text{ a.s.},$$

By the reflection principle, we have

$$F(t,s,b) = \int_{-\infty}^{\infty} \max(s, b + |x|\sqrt{1-t})\, \varphi(x)\, dx,$$

with φ denoting the density of the unit normal. The function F is twice differentiable in b for $t < 1$ (this can be checked directly) martingale on $[0,1)$, and the the process $F(t, S_t, B_t)$ is a martingale. Therefore, Itô's formula implies that (why?)

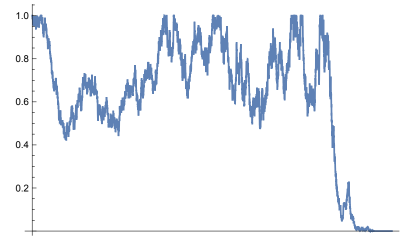
$$\mathbb{E}[S_1|\mathcal{F}_t] = \mathbb{E}[S_1] + \int_0^t \frac{\partial}{\partial b} F(u, S_u, B_u)\, dB_u \text{ for } t < 1, \tag{21.6}$$



A simulated path of a Brownian motion B (blue), its running max S (orange) and the martingale M (green)

Since all local martingales in the Brownian filtration are continuous, we can let $t \to 1$, and conclude that the equality in (21.6) holds a.s., even for $t = 1$. It remains to compute the derivative of F; with $\Phi(x) = \int_{-\infty}^{x} \varphi(\xi)\, d\xi$, we have

$$H_t = \int_{-\infty}^{\infty} \mathbf{1}_{\{B_t + |x|\sqrt{1-t} \geq S_t\}}\, \varphi(x)\, dx = 2\Big(1 - \Phi(\tfrac{S_t - B_t}{\sqrt{1-t}})\Big),$$



Values of the process H corresponding to the simulated path of the Brownian motion B above.

so that

$$S_1 = \sqrt{\tfrac{2}{\pi}} + \int_0^1 2\Big(1 - \Phi(\tfrac{S_u - B_u}{\sqrt{1-u}})\Big)\, dB_u$$

Course: Theory of Probability II
Term: Spring 2015
Instructor: Gordan Zitkovic

Lecture 22
Girsanov's Theorem

An example

Consider a finite Gaussian random walk

$$X_n = \sum_{k=1}^{n} \xi_k,\ n = 0, \dots, N,$$

where ξ_k are independent $N(0,1)$ random variables. The random vector $(X_1, \dots, X_N)$ is then, itself, Gaussian, and admits the density

$$f(x_1, \dots, x_N) = C_N e^{-\frac{1}{2}\left(x_1^2+(x_2-x_1)^2+\dots+(x_N-x_{N-1})^2\right)}$$

with respect to the Lebesgue measure on $\mathbb{R}^N$, for some $C_N > 0$.

Let us now repeat the whole construction, with the n-th step having the $N(\mu_n, 1)$-distribution, for some $\mu_1, \dots, \mu_N \in \mathbb{R}$. The resulting, Gaussian, distribution still admits a density with respect to the Lebesgue measure, and it is given by

$$\tilde{f}(x_1, \dots, x_N) = C_N e^{-\frac{1}{2}\left((x_1-\mu_1)^2+(x_2-x_1-\mu_2)^2+\dots+(x_N-x_{N-1}-\mu_N)^2\right)}.$$

The two densities are everywhere positive, so the two Gaussian measures are equivalent to each other and the Radon-Nikodym derivative turns out to be

$$\begin{aligned}
\frac{d\mathbb{Q}}{d\mathbb{P}} &= \frac{\tilde{f}(X_1,\dots,X_N)}{f(X_1,\dots,X_N)} \\
&= e^{-\left(\mu_1 X_1-\mu_2(X_2-X_1)-\dots-\mu_N(X_N-X_{N-1})\right)+\frac{1}{2}\left(\mu_1^2+\mu^2+\dots+\mu_N^2\right)} \\
&= e^{\sum_{k=1}^N \mu_k(X_k-X_{k-1})-\frac{1}{2}\sum_{k=1}^N \mu_k^2}.
\end{aligned}$$

Equivalent measure changes

Let $\mathbb{Q}$ be a probability measure on $\mathcal{F}$, equivalent to $\mathbb{P}$, i.e., $\forall A \in \mathcal{F}$, $\mathbb{P}[A] = 0$ if and only if $\mathbb{Q}[A] = 0$. Its Radon-Nikodym derivative

$Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$ is a non-negative random variable in $\mathbb{L}^1$ with $\mathbb{E}[Z] = 1$. The uniformly-integrable martingale

$$Z_t = \mathbb{E}[Z|\mathcal{F}_t],\ t \geq 0,$$

is called the **density** of $\mathbb{Q}$ with respect to $\mathbb{P}$ (note that we can - and do - assume that $\{Z_t\}_{t\in[0,\infty)}$ is càdlàg). We will often use the shortcut **$\mathbb{Q}$-(local, semi-, etc.) martingale** for a process which is a (local, semi-, etc.) martingale for $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q})$.

Proposition 22.1. *Let $\{X_t\}_{t\in[0,\infty)}$ be a càdlàg and adapted process. Then X is a $\mathbb{Q}$-local martingale if and only if the product $\{Z_t X_t\}_{t\in[0,\infty)}$ is a càdlàg $\mathbb{P}$-local martingale.*

Before we give a proof, here is a simple and useful lemma. Since the measures involved are equivalent, we are free to use the phrase "almost surely" without explicit mention of the probability.

Lemma 22.2. *Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space, and let $\mathcal{G} \subseteq \mathcal{H}$ be a sub-σ-algebra of $\mathcal{H}$. Givan a probability measure $\mathbb{Q}$ on $\mathcal{H}$, equivalent to $\mathbb{P}$, let $Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$ be its Radon-Nikodym derivative with respect to $\mathbb{P}$. For a random variable $X \in \mathbb{L}^1(\mathcal{F}, \mathbb{Q})$ we have $XZ \in \mathbb{L}^1(\mathbb{P})$ and*

$$\mathbb{E}^{\mathbb{Q}}[X|\mathcal{G}] = \tfrac{1}{\mathbb{E}[Z|\mathcal{G}]}\mathbb{E}[XZ|\mathcal{G}],\ a.s.$$

where $\mathbb{E}^{\mathbb{Q}}[\cdot|\mathcal{G}]$ denotes the conditional expectation on $(\Omega, \mathcal{H}, \mathbb{Q})$.

Proof. First of all, note that the Radon-Nikodym theorem implies that $XZ \in \mathbb{L}^1(\mathbb{P})$ and that the set $\{\mathbb{E}[Z|\mathcal{G}] = 0\}$ has $\mathbb{Q}$-probability (and, therefore $\mathbb{P}$-probability) 0. Indeed,

$$\begin{aligned}\mathbb{Q}[\mathbb{E}[Z|\mathcal{G}] = 0] &= \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{\mathbb{E}[Z|\mathcal{G}]=0\}}] = \mathbb{E}[Z\mathbf{1}_{\{\mathbb{E}[Z|\mathcal{G}]=0\}}] \\ &= \mathbb{E}[\mathbb{E}[Z\mathbf{1}_{\{\mathbb{E}[Z|\mathcal{G}]=0\}}|\mathcal{G}]]\end{aligned}$$

Therefore, the expression on the right-hand side is well-defined almost surely, and is clearly $\mathcal{G}$-measurable. Next, we pick $A \in \mathcal{G}$, observe that

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_A \tfrac{1}{\mathbb{E}[Z|\mathcal{G}]}\mathbb{E}[XZ|\mathcal{G}]] &= \mathbb{E}[Z\mathbf{1}_A \tfrac{1}{\mathbb{E}[Z|\mathcal{G}]}\mathbb{E}[XZ|\mathcal{G}]] \\ &= \mathbb{E}[\mathbb{E}[Z|\mathcal{G}]\mathbf{1}_A \tfrac{1}{\mathbb{E}[Z|\mathcal{G}]}\mathbb{E}[XZ|\mathcal{G}]] \\ &= \mathbb{E}[\mathbb{E}[ZX\mathbf{1}_A|\mathcal{G}]] = \mathbb{E}^{\mathbb{Q}}[X\mathbf{1}_A],\end{aligned}$$

and remember the definition of conditional expectation. □

Proof of Proposition 22.1. Suppose, first, that X is a $\mathbb{Q}$-martingale. Then $\mathbb{E}^{\mathbb{Q}}[X_t|\mathcal{F}_s] = X_s$, $\mathbb{Q}$-a.s. By the tower property of conditional expectation, the random variable Z_t is the Radon-Nikodym derivative of (the restriction of) $\mathbb{Q}$ with respect to (the restriction of) $\mathbb{P}$ on the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ (prove this yourself!). Therefore, we can use Lemma 22.2 with $\mathcal{F}_t$ playing the role of $\mathcal{H}$ and $\mathcal{F}_s$ the role $\mathcal{G}$, and rewrite the $\mathbb{Q}$-martingale property of X as

(22.1) $\frac{1}{Z_s}\mathbb{E}[X_t Z_t|\mathcal{F}_s] = X_s, \ \mathbb{Q} - \text{a.s., i.e. } \mathbb{E}[X_t Z_t|\mathcal{F}_s] = Z_s X_s, \ \mathbb{P} - \text{a.s.}$

We leave the other direction, as well as the case of a local martingale to the reader. □

Proposition 22.3. *Suppose that the density process $\{Z_t\}_{t\in[0,\infty)}$ is continuous. Let X be a continuous semimartingale under $\mathbb{P}$ with decomposition $X = X_0 + M + A$. Then X is also a $\mathbb{Q}$-semimartingale, and its $\mathbb{Q}$-semimartingale decomposition is given by $X = X_0 + N + B$, where*

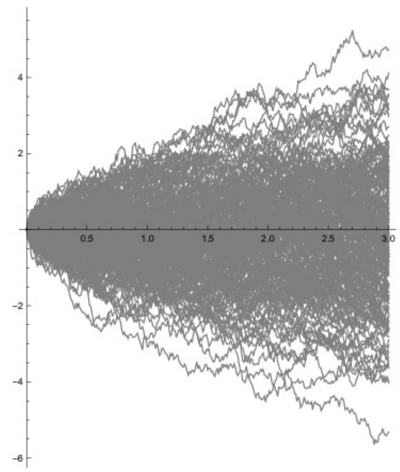
$$N = M - F, \ B = A + F \text{ where } F_t = \int_0^t \tfrac{1}{Z_t}\, d\langle M, Z\rangle.$$

Proof. The process F is clearly well-defined, continuous, adapted and of finite variation, so it will be enough to show that $M - F$ is a $\mathbb{Q}$-local martingale. Using Proposition 22.1, we only need to show that $Y = Z(M - F)$ is a $\mathbb{P}$-local martingale. By Itô's formula (integration-by-parts), the finite-variation part of Y is given by

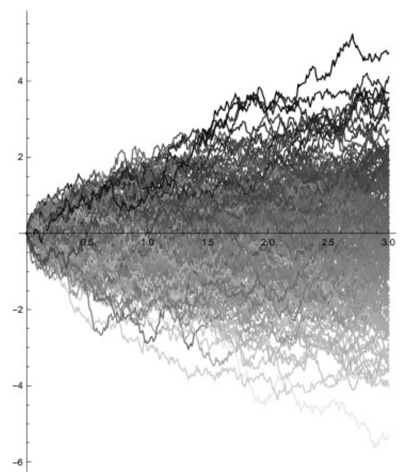
$$-\int_0^t Z_u\, dF_u + \langle Z, M\rangle_t,$$

and it is easily seen to vanish using the associative property of Stieltjes integration. □

One of the most important applications of the above result is to the case of a Brownian motion.



A cloud of simulated Brownian paths on $[0,3]$



The same cloud with darker-colored paths corresponding to higher values of the Radon-Nikodym derivative Z_3.

Theorem 22.4 (Girsanov; Cameron and Martin)**.** *Suppose that the filtration $\{\mathcal{F}_t\}_{t\in[0,\infty)}$ is the usual augmentation of the natural filtration generated by a Brownian motion $\{B_t\}_{t\in[0,\infty)}$.*

1. *Let $\mathbb{Q} \sim \mathbb{P}$ be a probability measure on $\mathcal{F}$ and let $\{Z_t\}_{t\in[0,\infty)}$ be the corresponding density process, i.e., $Z_t = \mathbb{E}[\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_t]$. Then, here exists a predictable process $\{\theta_t\}_{t\in[0,\infty)}$ in $L(B)$ such that $Z = \mathcal{E}(\int_0^\cdot \theta_u\, dB_u)$ and*

$$B_t - \int_0^t \theta_u\, du \text{ is a } \mathbb{Q}\text{-Brownian motion.}$$

2. *Conversely, let $\{\theta_t\}_{t\in[0,\infty)} \in L(B)$ have the property that the process $Z = \mathcal{E}(\int_0^\cdot \theta_u\, dB_u)$ is a uniformly-integrable martingale with $Z_\infty > 0$, a.s. For any probability measure $\mathbb{Q} \sim \mathbb{P}$ such that $\mathbb{E}[\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_\infty] = Z_\infty$,*

$$B_t - \int_0^t \theta_u\, du, \ t \geq 0,$$

is a $\mathbb{Q}$-Brownian motion.

Proof.

1. We start with an application of the martingale representation theorem (Proposition 21.16). It implies that there exists a process $\rho \in L(B)$ such that
$$Z_t = 1 + \int_0^t \rho_u \, dB_u.$$
Since Z is continuous and bounded away from zero on each segment, the process $\{\theta_t\}_{t\in[0,\infty)}$, given by $\theta_t = \rho_t / Z_t$ is in $L(B)$ and we have
$$Z_t = 1 + \int_0^t Z_u \theta_u dB_u.$$
Hence, $Z = \mathcal{E}(\int_0^\cdot \theta_u \, dB_u)$. Proposition 22.3 states that B is a $\mathbb{Q}$-semimartingale with decomposition $B = (B - F) + F$, where the continuous FV-process F is given by
$$F_t = \int_0^t \tfrac{1}{Z_u} \langle B, Z \rangle_u = \int_0^t \tfrac{1}{Z_u} Z_u \theta_u \, du = \int_0^t \theta_u \, du.$$
In particular, $B - F$ is a $\mathbb{Q}$-local martingale. On the other hand, its quadratic variation (as a limit in $\mathbb{P}$-, and therefore in $\mathbb{Q}$-probability) is that of B, so, by Lévy's characterization, $B - F$ is a $\mathbb{Q}$-Brownian motion.

2. We only need to realize that any measure $\mathbb{Q} \sim \mathbb{P}$ with $\mathbb{E}[\frac{d\mathbb{Q}}{d\mathbb{P}} | \mathcal{F}_\infty] = Z_\infty$ will have Z as its density process. The rest follows from (1). □

Even though we stated it on $[0, \infty)$, most of applications of the Girsanov's theorem are on finite intervals $[0, T]$, with $T > 0$. The reason is that the condition that $\mathcal{E}(\int_0^\cdot \theta_u \, dB_u)$ be uniformly integrable *on the entire* $[0, \infty)$ is either hard to check or even not satisfied for most practically relevant θ. The simplest conceivable example $\theta_t = \mu$, for all $t \geq 0$ and $\mu \in \mathbb{R} \setminus \{0\}$ gives rise to the exponential martingale $Z_t = e^{\mu B_t - \frac{1}{2}\mu^2 t}$, which is not uniformly integrable on $[0, \infty)$ (why?). On any finite horizon $[0, T]$, the (deterministic) process $\mu \mathbf{1}_{\{t \leq T\}}$ satisfies the conditions of Girsanov's theorem, and there exists a probability measure $\mathbb{P}^{\mu,T}$ on $\mathcal{F}_T$ with the property that $\hat{B}_t = B_t - \mu t$ is a $\mathbb{P}^{\mu,T}$ Brownian motion on $[0, T]$. It is clear, furthermore, that for $T_1 < T_2$, $\mathbb{P}^{\mu,T_1}$ coincides with the restriction of $\mathbb{P}^{\mu,T_2}$ onto $\mathcal{F}_{T_1}$. Our life would be easier if this consistency property could be extended all the way up to $\mathcal{F}_\infty$. It can be shown that this can, indeed, be done in the canonical setting, but not in same equivalence class. Indeed, suppose that there exists a probability measure $\mathbb{P}^\mu$ on $\mathcal{F}_\infty$, equivalent to $\mathbb{P}$, such that $\mathbb{P}^\mu$, restricted to $\mathcal{F}_T$, coincides with $\mathbb{P}^{\mu,T}$, for each $T > 0$. Let $\{Z_t\}_{t\in[0,\infty)}$ be the density process of $\mathbb{P}^\mu$ with respect to $\mathbb{P}$. It follows that
$$Z_T = \exp(\mu B_T - \tfrac{1}{2}\mu^2 T), \text{ for all } T > 0.$$

Since $\mu \neq 0$, we have $B_t - \frac{1}{2}\mu T \to -\infty$, a.s., as $T \to \infty$ and, so, $Z_\infty = \lim_{T\to\infty} Z_T = 0$, a.s. On the other hand, Z_∞ is the Radon-Nikodym derivative of $\mathbb{P}^\mu$ with respect to $\mathbb{P}$ on $\mathcal{F}_\infty$, and we conclude that $\mathbb{P}^\mu$ must be singular with respect to $\mathbb{P}$. Here is slightly different perspective on the fact that $\mathbb{P}$ and $\mathbb{P}^\mu$ must be mutually singular: for the event $A \in \mathcal{F}_\infty$, given by

$$A = \Big\{ \lim_{t\to\infty} \tfrac{B_t}{t} = 0 \Big\},$$

we have $\mathbb{P}[A] = 1$, by the Law of Large Numbers for the Brownian motion. On the other hand, with $\hat{B}_t$ being a $\mathbb{P}^\mu$ Brownian motion, we have

$$\mathbb{P}^\mu[A] = \mathbb{P}^\mu[\lim_{t\to\infty} \tfrac{B_t}{t} = 0] = \mathbb{P}^\mu[\lim_{t\to\infty} \tfrac{\hat{B}_t}{t} = -\mu] = 0,$$

because $\hat{B}_t/t \to 0$, $\mathbb{P}^\mu$-a.s. Not everything is lost, though, as we can still use employ Girsanov's theorem in many practical situations. Here is one (where we take $\mathbb{P}[X \in dx] = f(x)\,dx$ to mean that $f(x)$ is the density of the distribution of X.)

Example 22.5 (Hitting times of the Brownian motion with drift)**.** Define $\tau_a = \inf\{t \geq 0 : B_t = a\}$ for $a > 0$. By the formula derived in a homework, we have

$$\mathbb{P}[\tau_a \in dt] = \tfrac{|a|}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}}\, dt.$$

For $T \geq 0$, (the restriction of) $\mathbb{P}$ and $\mathbb{P}^{\mu,T}$ are equivalent on $\mathcal{F}_T$ with the Radon-Nikodym derivative

$$\tfrac{d\mathbb{P}^{\mu,T}}{d\mathbb{P}} = \exp(\mu B_T - \tfrac{1}{2}\mu^2 T).$$

The optional sampling theorem (justified by the uniform integrability of the martingale $\exp(\mu B_t - \frac{1}{2}\mu^2 t)$ on $[0,T]$) and the fact that $\{\tau_a \leq T\} \in \mathcal{F}_{\tau_a \wedge T} \subseteq \mathcal{F}_T$ imply that

$$\mathbb{E}[\exp(\mu B_T - \tfrac{1}{2}\mu^2 T)|\mathcal{F}_{\tau_a\wedge T}] = \exp(\mu B_{\tau_a \wedge T} - \tfrac{1}{2}\mu^2(\tau_a \wedge T)).$$

Therefore,

$$\begin{aligned}
\mathbb{P}^{\mu,T}[\tau_a \leq T] &= \mathbb{E}^{\mu,T}[\mathbf{1}_{\{\tau_a \leq T\}}] = \mathbb{E}[\exp(\mu B_T - \tfrac{1}{2}\mu^2 T)\mathbf{1}_{\{\tau_a \leq T\}}] \\
&= \mathbb{E}[\exp(\mu B_{\tau_a\wedge T} - \tfrac{1}{2}\mu^2(\tau_a \wedge T))\mathbf{1}_{\{\tau_a \leq T\}}] \\
&= \mathbb{E}[\exp(\mu B_{\tau_a} - \tfrac{1}{2}\mu^2 \tau_a)\mathbf{1}_{\{\tau_a \leq T\}}] \\
&= \mathbb{E}[\exp(\mu a - \tfrac{1}{2}\mu^2 \tau_a)\mathbf{1}_{\{\tau_a \leq T\}}] = \int_0^T e^{\mu a - \frac{1}{2}\mu^2 t}\,\mathbb{P}[\tau_a \in dt] \\
&= \int_0^T e^{\mu a - \frac{1}{2}\mu^2 t} \tfrac{|a|}{\sqrt{2\pi t^3}} e^{-\frac{a^2}{2t}}\, dt.
\end{aligned}$$

On the other hand, $\{B_t - \mu t\}_{t\in[0,T]}$ is a Brownian motion under $\mathbb{P}^{\mu,T}$, so

$$\mathbb{P}^{\mu,T}[\tau_a \leq T] = \mathbb{P}[\hat{\tau}_a \leq T],$$

where $\hat{\tau}_a$ is the first hitting time of the level a of the Brownian motion with drift μ. It follows immediately that the "density" of $\hat{\tau}_a$ is given by

$$\mathbb{P}[\hat{\tau}_a \in dt] = \tfrac{|a|}{\sqrt{2\pi t^3}} e^{-\frac{(a-\mu t)^2}{2t}}\, dt.$$

We quote the word "density" because, if one tries to integrate it over all $t \geq 0$, one gets

$$\mathbb{P}[\hat{\tau}_a < \infty] = \int_0^\infty \tfrac{|a|}{\sqrt{2\pi t^3}} e^{-\frac{(a-\mu t)^2}{2t}}\, dt = \exp(\mu a - |\mu a|).$$

In words, if μ and a have the same sign, the Brownian motion with drift μ will hit a sooner or later. On the other hand, if they differ in sign, the probability that it will never get there is strictly positive and equal to $e^{2\mu a}$.

Kazamaki's and Novikov's criteria

The message of the second part of Theorem 22.4 is that, given a "drift" process $\{\theta_t\}_{t\in[0,\infty)}$, we can turn a Brownian motion into a Brownian motion with drift θ, provided, essentially, that a certain exponential martingale is a UI martingale. Even though useful sufficient conditions for martingality of stochastic integrals are known, the situation is much less pleasant in the case of stochastic exponentials. The most well-known criterion is the one of Novikov. Novikov's criterion is, in turn, implied by a slightly stronger criterion of Kazamaki. We start with an auxiliary integrability result. In addition to the role it plays in the proof of Kazamaki's criterion, it is useful when one needs $\mathcal{E}(M)$ to be a little more than just a martingale.

Lemma 22.6. *Let $\mathcal{E}(M)$ be the stochastic exponential of $M \in \mathcal{M}_0^{loc,c}$. If*

$$\sup_{\tau\in\mathcal{S}_b} \mathbb{E}[e^{aM_\tau}] < \infty,$$

for some constant $a > \frac{1}{2}$, where the supremum is taken over the set $\mathcal{S}_b$ of all finite-valued stopping times τ, then $\mathcal{E}(M)$ is an $\mathbb{L}^p$-bounded martingale for $p = \frac{4a^2}{4a-1} \in (1,\infty)$.

Proof. We pick a finite stopping time τ and start from the following identity, which is valid for all constants $p, s > 0$:

$$\mathcal{E}(M)_\tau^p = \mathcal{E}(\sqrt{p/s}M)_\tau^s e^{(p-\sqrt{ps})M_\tau}.$$

For $1 > s > 0$, we can use Hölder's inequality (note that $1/s$ and $1/(1-s)$ are conjugate exponents), to obtain

$$\mathbb{E}[\mathcal{E}(M)^p] \leq (\mathbb{E}[\mathcal{E}(\sqrt{p/s}M_\tau)])^s(\mathbb{E}[\exp(\tfrac{p-\sqrt{ps}}{1-s}M_\tau)])^{1-s}. \tag{22.2}$$

The first term of the product is the s-th power of the expectation a positive local martingale (and, therefore, supermartingale) sampled at a finite stopping time. By the optional sampling theorem it is always finite (actually, it is less then 1). As for the second term, one can easily check that the expression $\frac{p-\sqrt{ps}}{1-s}$ attains its minimum in s over $(0,1)$ for $s = 2p - 1 - 2\sqrt{p^2 - p}$, and that this minimum value equals to $f(p)$, where $f(p) = \frac{1}{2}\frac{\sqrt{p}}{\sqrt{2p-1-2\sqrt{p^2-p}}}$. If we pick $p = \frac{4a^2}{4a-1}$, then $f(p) = a$ and both terms on the right hand side of (22.2) are bounded, uniformly in τ, so that $\mathcal{E}(M)$ is in fact a martingale and bounded in $\mathbb{L}^p$ (why did we have to consider all stopping times τ, and only deterministic times?). □

Proposition 22.7 (Kazamaki's criterion). *Suppose that for $M \in \mathcal{M}_0^{loc,c}$ we have*

$$\sup_{\tau\in\mathcal{S}_b} \mathbb{E}[e^{\frac{1}{2}M_\tau}] < \infty,$$

where the supremum is taken over the set $\mathcal{S}_b$ of all finite-valued stopping times, then $\mathcal{E}(M)$ is a uniformly integrable martingale.

Proof. Note, first, that the function $x \mapsto \exp(\frac{1}{2}x)$ is a test function of uniform integrability, so that the local martingale M is a uniformly integrable martingale and admits the last element M_∞. For the continuous martingale cM, where $0 < c < 1$ is an arbitrary constant, Lemma 22.6 and the assumption imply that the local martingale $\mathcal{E}(cM)$ is, in fact, a martingale bounded in $\mathbb{L}^p$, for $p = \frac{1}{2c-c^2}$. In particular, it is uniformly integrable. Therefore,

$$\mathcal{E}(cM)_t = \exp(cM_t - \tfrac{1}{2}c^2\langle M\rangle_t) = \mathcal{E}(M)_t^{c^2}e^{c(1-c)M_t}. \tag{22.3}$$

By letting $t \to \infty$ in (22.3), we conclude that $\mathcal{E}(M)$ has the last element $\mathcal{E}(M)_\infty$, and that the equality in (22.3) holds at $t = \infty$, as well. By Hölder's inequality with conjugate exponents $1/c^2$ and $1/(1-c^2)$, we have

$$1 = \mathbb{E}[\mathcal{E}(cM)_\infty] \leq \mathbb{E}[\mathcal{E}(M)_\infty]^{c^2}\mathbb{E}[\exp(\tfrac{c}{1+c}M_\infty)]^{1-c^2}.$$

Jensen's inequality implies that $\mathbb{E}[\exp(\frac{c}{1+c}M_\infty)] \leq \mathbb{E}[\exp(\frac{1}{2}M_\infty)]^{\frac{2c}{1+c}}$, and so

$$1 \leq \mathbb{E}[\mathcal{E}(M)_\infty]^{c^2}\mathbb{E}[\exp(\tfrac{1}{2}M_\infty)]^{2c(1-c)}.$$

We let $c \to 1$ to get $\mathbb{E}[\mathcal{E}(M)_\infty] \geq 1$, which, together with the nonnegative supermartingale property of $\mathcal{E}(M)$ implies that $\mathcal{E}(M)$ is a uniformly-integrable martingale. □

Theorem 22.8 (Novikov's criterion). *If $M \in \mathcal{M}_0^{loc,c}$ is such that*

$$\mathbb{E}[e^{\frac{1}{2}\langle M\rangle_\infty}] < \infty,$$

then $\mathcal{E}(M)$ is a uniformly integrable martingale.

Proof. Since $e^{\frac{1}{2}M_\tau} = \mathcal{E}(M)^{\frac{1}{2}} e^{\frac{1}{4}\langle M\rangle_\tau}$, the Cauchy-Schwarz inequality implies that

$$\mathbb{E}[e^{\frac{1}{2}M_\tau}] = \mathbb{E}[\mathcal{E}(M)_\tau]^{1/2}\mathbb{E}[e^{\frac{1}{2}\langle M\rangle_\tau}]^{1/2} \leq \mathbb{E}[e^{\frac{1}{2}\langle M\rangle_\infty}],$$

and Kazamaki's criterion can be applied. □