CHAPTER 1

Martingales, continued

Martingales are first and foremost a tool to establish the existence and properties of random limits. The basic limit theorems of probability (the law of large numbers and Glivenko-Cantelli) establish that averages of i.i.d. variables converge to their expected value. (The central limit theorem also falls into this class—it additionally establishes a rate of convergence.) The sequence of averages is a random sequence, but it *completely derandomizes in the limit.* For more complicated processes—typically, if the variables are stochastically dependent—the limit might be random. In general, random limits are very hard to handle mathematically, since we have to precisely quantify the effect of dependencies and cope with the randomness aggregating as we get further into the sequence.

It turns out that it is possible to control dependencies and randomness if a process has the martingale property (or is at least a sub- or supermartingale). Hence:

Martingales are tools for working with random limits.

They are not the only such tools, but there are very few others. This fact accounts for the almost overwhelming importance of martingales in probability.

1.1. Martingales indexed by partially ordered sets

Martingales have already been discussed in Probability Theory I, where they were indexed by values in $\mathbb{N}$ (so-called "discrete-time martingales"). We will briefly discuss martingales in "continuous time"—that is, martingales indexed by $\mathbb{R}_+$. However, martingales can much more generally be defined for index sets that need not be totally ordered, and we will proof two useful results for such general martingales.

Partially ordered index sets. Let $\mathbb{T}$ be a set. Recall that a binary relation $\preceq$ on $\mathbb{T}$ is called a **partial order** if it is

(1) reflexive, that is $s \preceq s$ for every $s \in \mathbb{T}$.
(2) antisymmetric, that is if $s \preceq t$ and $t \preceq s$, then $s = t$.
(3) transitive, that is if $s \preceq t$ and $t \preceq u$, then $s \preceq u$.

In general, a partially ordered set may contain elements that are not comparable, i.e. some s, t for which neither $s \preceq t$ nor $t \preceq s$ (hence "partial"). If all pairs of elements are comparable, the partial order is called a **total order**.

In various contexts—including martingales and the construction of stochastic processes—we need partially ordered index sets. We have to be careful, though: Using arbitrary partially ordered sets can lead to all kinds of pathologies. Roughly speaking, the problem is that a partially ordered set can decompose into subsets between which elements cannot be compared at all, as if we were indexing arbitrarily by picking indices from completely unrelated index sets. For instance, a partially ordered set could contain two sequences $s_1 \preceq s_2 \preceq s_3 \preceq \dots$ and $t_1 \preceq t_2 \preceq t_3 \preceq \dots$ of elements which both grow larger and larger in terms of the partial order, but whose elements are completely incomparable between the sequences. To avoid such pathologies, we impose an extra condition:

$$\text{If } s,t \in \mathbb{T}, \text{ there exists } u \in \mathbb{T} \text{ such that } s \preceq u \text{ and } t \preceq u \; . \tag{1.1}$$

A partially ordered set $(\mathbb{T}, \preceq)$ which satisfies (1.1) is called a **directed set**.

1.1 Example. Some examples of directed sets:

(a) The set of all subsets of an arbitrary set, ordered by inclusion.
(b) The set of all finite subsets of an infinite set, ordered by inclusion.
(c) The set of all positive definite $n \times n$ matrices over $\mathbb{R}$, in the Löwner partial order.
(d) Obviously, any totally ordered set (such as $\mathbb{N}$ or $\mathbb{R}$ in the standard order) is directed.

◁

Just as we can index a family of variables by $\mathbb{N}$ and obtain a sequence, we can more generally index it by a directed set; the generalization of a sequence so obtained is called a *net*. To make this notion precise, let $\mathcal{X}$ be a set. Recall that, formally, an **(infinite) sequence** in $\mathcal{X}$ is a mapping $\mathbb{N} \to \mathcal{X}$, that is, each index s is mapped to the sequence element x_i. We usually denote such a sequence as $(x_i)_{i \in \mathbb{N}}$, or more concisely as (x_i).

1.2 Definition. Let $(\mathbb{T}, \preceq)$ be a directed set. A **net** in a set $\mathcal{X}$ is a function $x : \mathbb{T} \to \mathcal{X}$, and we write $x_t := x(t)$ and denote the net as $(x_t)_{t \in \mathbb{T}}$. ◁

Clearly, the net is a sequence if $(\mathbb{T}, \preceq)$ is the totally ordered set $(\mathbb{N}, \leq)$. Just like sequences, nets may converge to a limit. The definition is analogous to that of a Cauchy sequence:

1.3 Definition. A net $(x_t)_{t \in I}$ is said to **converge** to a point x if, for every open neighborhood U of x, there exists an index $t_0 \in \mathbb{T}$ such that

$$x_t \in U \qquad \text{whenever } t_0 \preceq t \; . \tag{1.2}$$

◁

Nets play an important role in real and functional analysis: Too establish certain properties in spaces more general than $\mathbb{R}^d$, we have to demand e.g. that every net satisfying certain properties converges (not just every sequence). Sequences have much stronger properties than nets; for example, in any topological space, the set consisting of all elements of a convergent sequence and its limit is a compact set. The same needs not be true for for nets.

Filtrations and martingales. The discrete-time martingales discussed in Probability I are a special type of *random sequences*. Analogously, the more general martingales we discuss in the following are a form of *random nets*.

Let $(\mathbb{T}, \preceq)$ be a directed set. By a **filtration**, we mean a family $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ of σ-algebras $\mathcal{F}_i$, indexed by the elements of $\mathbb{T}$, which satisfy

$$s \preceq t \quad \Longrightarrow \quad \mathcal{F}_s \subset \mathcal{F}_t \; . \tag{1.3}$$

The case $\mathbb{T} = \mathbb{N}$, which you have already encountered in Probability Theory I, is often called the discrete-time case; similarly, $\mathbb{T} = \mathbb{R}_+$ is called the continuous-time case. The filtration property states that each σ-algebra $\mathcal{F}_s$ contains all preceding ones. For a filtration, there is also a uniquely determined, smallest σ-algebra which contains all σ-algebras in $\mathcal{F}$, namely

$$\mathcal{F}_\infty := \sigma\Big(\bigcup\nolimits_{s\in\mathbb{T}}\mathcal{F}_s\Big) . \tag{1.4}$$

Clearly, every random sequence or net $(X_s)_{s\in\mathbb{T}}$ is adapted to the filtration

$$\mathcal{F}_t := \sigma\Big(\bigcup\nolimits_{s\preceq t}\sigma(X_s)\Big) , \tag{1.5}$$

which is called the **canonical filtration** of (X_s).

Now, let $(\mathbb{T},\preceq)$ be a partially ordered set and $\mathcal{F}=(\mathcal{F}_s)_{s\in\mathbb{T}}$ a filtration. As in the discrete-time case, we call a family $(X_s)_{s\in\mathbb{T}}$ of random variables **adapted** to $\mathcal{F}$ if X_s is $\mathcal{F}_s$-measurable for every s. Then $(X_s,\mathcal{F}_s)_{s\in\mathbb{T}}$ is called a **martingale** if it satisfies

$$X_s =_{\text{a.s.}} \mathbb{E}[X_t|\mathcal{F}_s] \qquad \text{whenever } s\preceq t , \tag{1.6}$$

or equivalently,

$$\forall A\in\mathcal{F}_s : \qquad \int_A X_s d\mathbb{P} = \int_A X_t d\mathbb{P} \qquad \text{whenever } s\preceq t . \tag{1.7}$$

If it the conditions only with equality weakened to $\leq$ (that is, $X_s\leq\mathbb{E}(X_t|\mathcal{F}_s)$ or $\int X_s d\mathbb{P}\leq\int X_t d\mathbb{P}$), it is called a **submartingale**; for $\geq$, it is called a **supermartingale**.

Our objective in the following. The objective of martingale convergence results is ideally to establish, for a given martingale $(X_s,\mathcal{F}_s)$, the existence of a limit random variable X_∞ which satisfies

$$X_s =_{\text{a.s.}} \mathbb{E}[X_\infty|\mathcal{F}_s] \qquad \text{for all } s\in\mathbb{T} . \tag{1.8}$$

Note that a result of the form (1.8) is more than just a convergence result: (1.8) is a representation theorem.

1.2. Notions of convergence for martingales

You will recall from the martingale convergence results discussed in Probability I that two types of convergence are of interest: Convergence almost surely, and in $\mathbf{L}_1$. We should briefly review why so.

In terms of convergence of a random variable, the strongest result we can hope for is almost sure convergence, i.e.

$$X_s(\omega)\to X_\infty(\omega) \qquad \mathbb{P}\text{-a.s.} \tag{1.9}$$

However, if (X_s) is a martingale, we also know that

$$X_s =_{\text{a.s.}} \mathbb{E}[X_t|\mathcal{F}_s] \qquad \text{for all } s \text{ with } s\preceq t . \tag{1.10}$$

This holds no matter how large (in the partial order) we make t. We hence might be forgiven to hope that (1.8) holds, too—but that is not generally true. What do we need to establish (1.8)? Suppose (X_s) converges to X_∞ in $\mathbf{L}_1$, i.e.

$$\lim\int_\Omega |X_s(\omega)-X_\infty(\omega)|\mathbb{P}(d\omega) = \lim\mathbb{E}\big[|X_s-X_\infty|\big]\to 0 . \tag{1.11}$$

(Note: If $\mathbb{T}$ is a directed set, then $(\mathbb{E}[|X_s - X_\infty|])_{s\in\mathbb{T}}$ is a net in $\mathbb{R}$, so $\mathbf{L}_1$ convergence means this net converges to the point 0 in the sense of Definition 1.3.) It is easy to verify the following fact:

1.4 Fact. If (X_s) converges in $\mathbf{L}_1$ to X_∞, then $(X_s\mathbb{I}_A)_i$ converges in $\mathbf{L}_1$ to $X_\infty\mathbb{I}_A$ for every measurable set A. ◁

Hence, for every index $s \in \mathbb{T}$, $\mathbf{L}_1$ convergence implies

$$\lim \int_A X_s d\mathbb{P} = \lim \int_\Omega X_s \mathbb{I}_A d\mathbb{P} = \int_\Omega X_\infty \mathbb{I}_A d\mathbb{P} = \int_A X_\infty d\mathbb{P} \tag{1.12}$$

for every $A \in \mathcal{F}_s$. By the martingale property (1.6), the sequence or net of integrals is additionally constant, that is

$$\int_A X_s d\mathbb{P} = \int_A X_t d\mathbb{P} \quad \text{ for all pairs } s \le t \text{ and hence } \int_A X_s d\mathbb{P} = \int_A X_\infty d\mathbb{P} \ , \tag{1.13}$$

which is precisely (1.8).

Recall from your discussion of notions of convergence for random variables in [Probability I, Chapter 17] how the different types of convergence relate to each other:

$$\begin{tikzcd}
& \boxed{\text{weakly}} & & \\
\boxed{\text{almost surely}} \arrow[r, shift left] & \boxed{\text{in probability}} \arrow[u] \arrow[l, shift left, "\text{subsequence}"] & \boxed{\mathbf{L}_1} \arrow[l] & \boxed{\mathbf{L}_p} \arrow[l]
\end{tikzcd} \tag{1.14}$$

Neither does almost sure convergence imply $\mathbf{L}_1$ convergence, nor vice versa. This is why "strong" martingale convergence results are typically of the form, "If (...), the martingale converges to a limit almost surely and in $\mathbf{L}_1$".

1.3. Uniform integrability

Recall from e.g. Theorem 27.1 in Probability I that the martingale property and a very mild supremum condition suffice to establish almost sure convergence: If $(X_n, \mathcal{F}_n)$ is a submartingale satisfying $\sup_n \mathbb{E}[\max\{0, X_n\}] < \infty$, then $X_\infty = \lim X_n$ exists almost surely. Moreover, the limit is finite almost surely, and integrable. The result does not yield $\mathbf{L}_1$ convergence. The property we need for martingales to converge in $\mathbf{L}_1$, and hence to establish (1.8), is uniform integrability. As a reminder:

1.5 Definition. Let $\mathbb{T}$ be an index set and $\{f_s | s \in \mathbb{T}\}$ a family of real-valued functions. The family is called **uniformly integrable** with respect to a measure μ if, for every $\varepsilon > 0$, there exists a positive function $g \ge 0$ such that

$$\int_{\{|f_s| \ge g\}} |f_s| d\mu \le \varepsilon \qquad \text{for all } s \in \mathbb{T} \ . \tag{1.15}$$

◁

Clearly, any finite set of random variables is uniformly integrable. The definition is nontrivial only if the index set is infinite. Here is a primitive example: Suppose the functions f_s in (1.15) are the constant functions on $[0, 1]$ with values $1, 2, \ldots$. Each function by itself is integrable, but the set is obviously not uniformly integrable. If the functions are in particular random variables $X_s : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, Definition 1.5 reads: For every ε, there is a positive random variable Y such that

$$\mathbb{E}\big[|X_s|\big||X_s| \geq Y\big] \leq \varepsilon \qquad \text{for all } s \in \mathbb{T} . \tag{1.16}$$

The definition applies to martingales in the obvious way: $(X_s, \mathcal{F}_s)_{s\in\mathbb{T}}$ is called a **uniformly integrable martingale** if it is a martingale and the family $(X_s)_{s\in\mathbb{T}}$ of functions is uniformly integrable.

Verifying (1.15) for a given family of functions can be pretty cumbersome, but can be simplified using various useful criteria. We recall two of them from [Probability I, Theorem 27.2]:

1.6 Lemma. *A family $(X_s)_{s\in\mathbb{T}}$ of real-valued random variables with finite expectations is uniformly integrable if there is a random variable Y with $\mathbb{E}[Y] < \infty$ such that either of the following conditions holds:*

(1) *Each X_s satisfies*

$$\int_{\{|X_s|\geq\alpha\}} |X_s| d\mu \leq \int_{\{|X_s|\geq\alpha\}} Y d\mu \qquad \textit{for all } \alpha > 0 . \tag{1.17}$$

(2) *Each X_s satisfies $|X_s| \leq Y$.*

◁

We now come to our first result for martingales on general index sets, which shows two important facts:

(1) We can easily obtain uniformly integrable martingales by positing a limit variable X and "filtering" it.
(2) Uniform integrability is *necessary* to for (1.8) to hold.

1.7 Theorem. *Let X be an integrable, real-valued random variable on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $(\mathbb{T}, \preceq)$ be a partially ordered set, and $\mathcal{F} = (\mathcal{F}_s)_{s\in\mathbb{T}}$ a filtration in $\mathcal{A}$. Then*

$$\big(\mathbb{E}[X|\mathcal{F}_s], \mathcal{F}_s\big)_{s\in\mathbb{T}} \tag{1.18}$$

is a uniformly integrable martingale. ◁

Proof. $(X_s, \mathcal{F}_s)$ is a martingale by construction: For any pair $s \preceq t$ of indices,

$$\mathbb{E}[X_t|\mathcal{F}_s] =_{\text{a.s.}} \mathbb{E}[\mathbb{E}[X|\mathcal{F}_t]|\mathcal{F}_s] \overset{\substack{\mathcal{F}_s \subseteq \mathcal{F}_t \\ \downarrow}}{=}_{\text{a.s.}} \mathbb{E}[X|\mathcal{F}_s] =_{\text{a.s.}} X_s . \tag{1.19}$$

To show that it is also uniformly integrable, we use Jensen's inequality [e.g. Probability I, Theorem 23.9]—recall: $\phi(\mathbb{E}[X|\mathcal{C}]) \leq \mathbb{E}[\phi(X)|\mathcal{C}]$ for any convex function ϕ—which implies

$$|X_s| =_{\text{a.s.}} \big|\mathbb{E}[X|\mathcal{F}_s]\big| \leq_{\text{a.s.}} \mathbb{E}\big[|X|\big|\mathcal{F}_s\big] . \tag{1.20}$$

Hence, for every $A \in \mathcal{F}_s$, we have

$$\int_A |X_s| dP \leq \int_A |X| dP , \tag{1.21}$$

which holds in particular for $A := \{|X_s| \geq \alpha\}$. By Lemma 1.6, the martingale is uniformly integrable. □

1.4. Convergence of martingales with directed index sets

You have already seen martingale convergence results for the discrete time case $\mathbb{T} = \mathbb{N}$ in the previous class. The next result generalizes the martingale convergence theorem [Probability I, Theorem 27.3] to the case of arbitrary directed index sets.

1.8 Martingale convergence theorem. *Let $(X_s, \mathcal{F}_s)_{s\in\mathbb{T}}$ be a martingale with directed index set $\mathbb{T}$. If the martingale is uniformly integrable, there exists a random variable X_∞ such that $(X_s, \mathcal{F}_s)_{s\in\mathbb{T}\cup\{\infty\}}$ is a martingale; that is,*

$$X_s =_{\text{a.s.}} \mathbb{E}[X_\infty|\mathcal{F}_s] \qquad \text{for all } s \in \mathbb{T} \,. \tag{1.22}$$

The variable X_∞ is integrable, and uniquely determined up to almost sure equivalence. ◁

Proof. The key idea is to use the directed structure of the index set to reduce to the martingale convergence theorem for the discrete-time case, Theorem 27.3 in [Probability I].

Step 1: The net satisfies the Cauchy criterion. We have to show that the random net $(X_s)_{s\in\mathbb{T}}$ converges in $\mathbf{L}_1$; in other words, that

$$\forall\varepsilon > 0 \; \exists s_0 \in \mathbb{T} : \quad \mathbb{E}[|X_t - X_u|] \leq \varepsilon \quad \text{ for all } t, u \text{ with } s_0 \preceq t \text{ and } s_0 \preceq u \,. \tag{1.23}$$

This follows by contradiction: Suppose (1.23) was not true. Then we could find an $\varepsilon > 0$ and a sequence $s_1 \preceq s_2 \preceq \ldots$ of indices such that $\mathbb{E}[|X_{s_{n+1}} - X_{s_n}|] \geq \varepsilon$ for all n. Since the $(X_s, \mathcal{F}_s)_{s\in\mathbb{T}}$ is a uniformly integrable martingale, so is the sub-family $(X_{s_n}, \mathcal{F}_{s_n})_{n\in\mathbb{N}}$—but we have just shown that is does not converge, which contradicts Theorem 27.3. Thus, (1.23) holds.

Step 2: Constructing the limit. Armed with (1.23), we can explicitly construct the limit, which we do using a specifically chosen subsequence: Choose ε in (1.23) consecutively as $1/1, 1/2, \ldots$. For each such $\varepsilon = 1/n$, choose an index s_n satisfying (1.23). Since $\mathbb{T}$ is directed, we can choose these indices increasingly in the partial order, $s_{1/1} \preceq s_{1/2} \preceq \ldots$. Again by [Probability I, Theorem 27.3], this makes $(X_{s_{1/n}})_n$ a convergent martingale, and there is a limit variable X. (We will still have to tweak X a little to obtain the variable X_∞ we are actually looking for.)

Step 3: The entire net converges to the limit X. For the sequence constructed above, if $n \leq m$, then $s_{1/n} \preceq s_{1/m}$. Substituting into (1.23) shows that

$$\mathbb{E}[|X_{s_{1/m}} - X_{s_{1/n}}|] \leq \frac{1}{n} \qquad \text{and hence} \qquad \mathbb{E}[|X - X_s|] \leq \frac{1}{n} \qquad \text{for all } s \text{ with } s_n \preceq s \,.$$

Hence, the entire net converges to X.

Step 4: X is determined almost everywhere. We have to show that the event

$$N := \{\omega \,|\, (X_{t_m}(\omega)) \text{ does not converge}\} \tag{1.24}$$

is a null set in $\mathcal{F}_\infty$. Recall from the convergence diagram (1.14) that $\mathbf{L}_1$ is pretty high up in the food chain: Convergence of a sequence of random variables in $\mathbf{L}_1$ implies convergence in probability, which in turn implies existence of an almost surely convergent subsequence. Since $(X_{s_{1/n}})$ by construction converges to X in $\mathbf{L}_1$, this means there is a subsequence $(t_1, t_2, \ldots)$ of $(s_{1/n})_n$ such that $(X_{t_m})_m$ converges almost surely to X.

Since X_s is a martingale, we also have $X_{s_{1/n}} =_{\text{a.s.}} \mathbb{E}[X_{t_m}|\mathcal{F}_{s_{1/n}}]$ for any index $s_{1/n}$ in the entire sequence which satisfies $s_{1/n} \preceq t_m$, and so the *entire* sequence $(X_{1/n})$ converges almost surely. Hence, N is a null set. Since all random variables X_s and X are $\mathcal{F}_\infty$-measurable, $N \in \mathcal{F}_\infty$.

Step 5: X does what we want. We have to convince ourselves that X indeed satisfies (1.22), and hence that

$$\int_A X_s d\mathbb{P} = \int X d\mathbb{P} \qquad \text{for all } A \in \mathcal{F}_s \; . \tag{1.25}$$

Since the entire net $(X_s)_{s\in\mathbb{T}}$ converges to X in $\mathbf{L}_1$, we can use Fact 1.4: $\mathbb{I}_A X_s$ also converges to $\mathbb{I}_A X$ in $\mathbf{L}_1$ for any $A \in \mathcal{F}_s$. Hence,

$$\int_A X_s d\mathbb{P} = \int_\Omega \mathbb{I}_A X_s d\mathbb{P} = \int_\Omega \mathbb{I}_A X d\mathbb{P} = \int_A X d\mathbb{P} \; . \tag{1.26}$$

Step 6: X is unique up to a.s. equivalence. Finally, suppose X' is another $\mathcal{F}_\infty$-measurable random variable satisfying (1.22). We have to show $X =_{\text{a.s.}} X'$. Since both variables are $\mathcal{F}_\infty$-measurable, we have to show that

$$\int_A X d\mathbb{P} = \int_A X' d\mathbb{P} \tag{1.27}$$

holds for all $A \in \mathcal{F}_\infty$. We will proof this using a standard proof technique, which I do not think you have encountered before. Since it is important and very useful, let me briefly summarize it in general terms before we continue:

1.9 Remark [Proof technique]. What we have to show in this step is that some property—in this case, (1.27)—is satisfied on all sets in a given σ-algebra $\mathcal{C}$ (here: $\mathcal{F}_\infty$). To solve problems of this type, we define two set systems:

(1) The set $\mathcal{D}$ of all sets $A \in \mathcal{C}$ which *do* satisfy the property. At this point, we do not know much about this system, but we know that $\mathcal{D} \subset \mathcal{C}$.

(2) The set $\mathcal{E}$ of all $A \in \mathcal{C}$ for which we *already know* the property is satisfied.

Then clearly,

$$\mathcal{E} \subset \mathcal{D} \subset \mathcal{C} \; . \tag{1.28}$$

What we have to show is $\mathcal{D} = \mathcal{C}$.

The proof strategy is applicable if we can show that: (1) $\mathcal{E}$ is a generator of $\mathcal{C}$, i.e. $\sigma(\mathcal{E}) = \mathcal{C}$; (2) $\mathcal{E}$ is closed under finite intersections; and (3) $\mathcal{D}$ is closed under complements and increasing differences. If (2) and (3) are true, the monotone class theorem [Probability I, Theorem 6.2] tells us that $\sigma(\mathcal{E}) \subset \mathcal{D}$. In summary, (1.28) then becomes

$$\mathcal{C} = \sigma(\mathcal{E}) \subset \mathcal{D} \subset \mathcal{C} \; , \tag{1.29}$$

and we have indeed shown $\mathcal{D} = \mathcal{C}$, i.e. our property holds on all of $\mathcal{C}$. ◁

Now back to the proof at hand: In this case, we define $\mathcal{D}$ as the set of all $A \in \mathcal{F}_\infty$ which satisfy (1.27). We note that (1.27) is satisfied whenever $A \in \mathcal{F}_s$ for some index s, and so we choose $\mathcal{E}$ as

$$\mathcal{E} = \bigcup_{s\in\mathbb{T}} \mathcal{F}_s \; . \tag{1.30}$$

Then (1.28) holds (for $\mathcal{C} = \mathcal{F}_\infty$), and we have left to show $\mathcal{D} = \mathcal{F}_\infty$. Recall that $\sigma(\mathcal{E}) = \mathcal{F}_\infty$ by definition, so one requirement is already satisfied.

The system $\mathcal{D}$ is closed under complements and increasing limits: Suppose $A \in \mathcal{D}$, and let $\bar{A}$ be its complement. Since Ω is contained in every $\mathcal{F}_s$, it is also in $\mathcal{D}$. Since both A and Ω satisfy (1.27), so does $\bar{A} = \Omega \setminus A$, and $\mathcal{D}$ is closed under differences. Similarly, suppose $A_1 \subset A_2 \subset \ldots$ is a sequence of sets which are all in $\mathcal{D}$, and $A := \cup_n A_n$. By definition of the integral, $\int_A X d\mathbb{P} = \lim_n \int_{A_n} X d\mathbb{P}$. Applying the limit on both sides of (1.27) shows $A \in \mathcal{D}$.

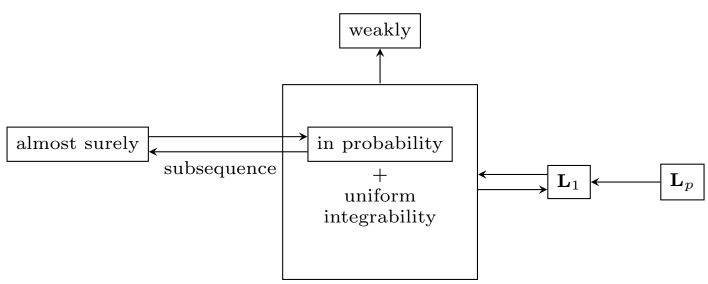
The set system $\mathcal{E}$ is closed under finite intersections: Suppose $A \in \mathcal{F}_s$ and $B \in \mathcal{F}_t$ for any two $s, t \in \mathbb{T}$. Since $\mathbb{T}$ is directed, there is some $u \in \mathbb{T}$ with $s, t \preceq u$, and hence $A, B \in \mathcal{F}_u$ and $A \cap B \in \mathcal{F}_u \subset \mathcal{E}$.

Hence, we have $\sigma(\mathcal{E}) = \mathcal{D}$ by the monotone class theorem, and

$$\mathcal{F}_\infty = \sigma(\mathcal{E}) = \mathcal{D} \subset \mathcal{F}_\infty \ , \tag{1.31}$$

so (1.27) indeed holds for all $A \in \mathcal{F}_\infty$. $\square$

Theorem 1.8 above shows that uniform integrability is not only necessary but also sufficient to establish (1.8). This is an example of a more general fact regarding the convergence of random variables (whether or not they form martingales): The diagram (1.14) shows that convergence in probability does not imply $\mathbf{L}_1$ convergence. To obtain the latter from the former, we need extra conditions. One possible condition is to assume that the random sequence or net in question (1) converges almost surely and (2) is dominated, i.e. bounded in absolute value by some random variable ($|X_s| \leq Y$ for some Y and all s). This fairly strong assumption can be weakened to: The net (1) converges in probability and (2) is uniformly integrable. We can hence augment diagram (1.8) as follows:



(1.32)

For martingales in particular, we can use the martingale property (plus mild regularity conditions) to ensure almost sure convergence to a limit (e.g. Theorem 27.1 in Probability I). Almost sure convergence implies convergence in probability; by also imposing an uniform integrability assumption, we then obtain the desired $\mathbf{L}_1$ convergence.

1.5. Application: The 0-1 law of Kolmogorov

Recall the 0-1 law from Probability I: If $X_1, X_2, \ldots$ is an infinite sequence of independent random variables, and a measurable event A does not depend on the value of the initial sequence $X_1, \ldots, X_n$ for any n, then A occurs with probability either 0 or 1. The prototypical example is the convergence of a series: If the random variables take values in, say, $\mathbb{R}^n$, the event

$$\left\{ \textstyle\sum_i X_i \text{ converges } \right\} \tag{1.33}$$

does not depend on the first n elements of the sequence for any finite n. Hence, the theorem states that the series either converges almost surely, or almost surely does not converge. However, the limit value it converges to *does* depend on every X_i. Thus, the theorem may tell us that the sequence converges, but not what it converges to.

In formal terms, the set of events which do not depend on values of the first n variables is the σ-algebra $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. The set of all events which do not depend on $(X_1, \ldots, X_n)$ for *any* n is $\mathcal{T} := \cap_n \mathcal{T}_n$, which is again a σ-algebra, and is called the **tail σ-algebra**, or simply the **tail field**.

1.10 Kolmogorov's 0-1 law. *Let $X_1, X_2, \ldots$ be independent random variables, and let A be an event such that, for every $n \in \mathbb{N}$, A is independent of the outcomes of $X_1, \ldots, X_n$. Then $\mathbb{P}(A)$ is 0 or 1. That is, if $A \in \mathcal{T}$, then $\mathbb{P}(A) \in \{0, 1\}$.* $\triangleleft$

This result was also proven as Theorem 10.6 in [Probability I]. It can be proven very concisely using martingales. Arguably the key insight underlying the theorem is that every set in $\mathcal{T}$ is $\sigma(X_1, X_2, \ldots)$-measurable. The martingale proof shows this very nicely:

Proof. For any measurable set A, we have $\mathbb{P}(A) = \mathbb{E}[\mathbb{I}_A]$. Suppose $A \in \mathcal{T}$. Since A is independent of $X_1, \ldots, X_n$, we have

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{I}_A] =_{\text{a.s.}} \mathbb{E}[\mathbb{I}_A | X_{1:n}] \qquad \text{for all } n \in \mathbb{N} . \tag{1.34}$$

We use martingales because they let us determine the conditional expectation $\mathbb{E}[\mathbb{I}_A | X_{1:\infty}]$ given the entire sequence: The sequence $(\mathbb{E}[\mathbb{I}_A | X_{1:n}], \sigma(X_{1:n}))_n$ is an uniformly integrable martingale by Theorem 1.7, and by Theorem 1.8 converges almost surely to an a.s.-unique limit. Since

$$\mathbb{E}\big[\mathbb{E}[\mathbb{I}_A | X_{1:\infty}] \big| X_{1:n}\big] =_{\text{a.s.}} \mathbb{E}[\mathbb{I}_A | X_{1:n}] , \tag{1.35}$$

(1.22) shows that the limit is $\mathbb{E}[\mathbb{I}_A | X_{1:\infty}]$, and hence

$$\mathbb{E}[\mathbb{I}_A | X_{1:\infty}] =_{\text{a.s.}} \lim_n \mathbb{E}[\mathbb{I}_A | X_{1:n}] =_{\text{a.s.}} \lim_n \mathbb{P}(A) = \mathbb{P}(A) . \tag{1.36}$$

Since $\mathcal{T} \subset \sigma(X_{1:\infty})$, the function $\mathbb{I}_A$ is $\sigma(X_{1:\infty})$-measurable, and hence

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{I}_A | X_{1:\infty}] = \mathbb{I}_A \in \{0, 1\} \qquad \text{almost surely.} \tag{1.37}$$

$\square$

1.6. Continuous-time martingales

The so-called *continuous-time case* is the special case where the index set is chosen as $\mathbb{T} := \mathbb{R}_+$, so we can think of the martingale (X_t) as a time series, started at time $t = 0$. For any fixed $\omega \in \Omega$, we can the interpret the realization $(X_t(\omega))$ of the martingale as a random function $t \mapsto X_t(\omega)$. Each realization of this function is called a **sample path**. We can then ask whether this function is continuous, or at least piece-wise continuous—this is one of the aspects which distinguish the continuous-time case from discrete time. Rather than continuity, we will use a notion of piece-wise continuity:

1.11 Reminder [rcll functions]. Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a function. Recall that f is continuous at x if, for every sequence (x_n) with $x_n \to x$, we have $\lim_n f(x_n) = f(x)$. We can split this condition into two parts: For every sequence $x_n \to x$, (1) $\lim_n f(x_n)$ exists and (2) equals $f(x)$.

Now suppose that, instead of all sequence with limit x, we consider only those which **converge from above** to x, i.e. sequences with $x_n \to x$ and $x_n \geq x$ for all n; we denote convergence from above as $x_n \searrow x$. If condition (1) is satisfied for all sequences which converge to x from above, i.e. if $\lim_n f(x_n)$ exists for all $x_n \searrow x$, we say that f has a **right-hand limit** at x. If (2) is also satisfied, i.e. if $\lim_n f(x_n) = f(x)$ for all such sequence, we call f **right-continuous** at x. **Left-hand limits** and **left-continuity** are defined similarly, considering only sequence which converge to x from below.

We say that a function on $\mathbb{R}_+$ is **right-continuous with left-hand limits**, or **rcll** for short, if it is right-continuous at every point in $[0, \infty)$ and has a left-hand limit at every point in $(0, \infty]$. $\triangleleft$

Intuitively, rcll functions are functions that are piece-wise continuous functions which jump at an at most countable number of points (otherwise, they would not have right- and left-hand limits). If the function jumps at x, the function value $f(x)$ is part of the "right-hand branch" of the function (which is condition (2) in right-continuity).

Filtrations for continuous-time martingales. In this section, we look for conditions which ensure a martingale has rcll sample paths. To formulate such conditions, we have to impose additional requirements on filtrations. One requirement is that filtrations contain all negligible sets.

1.12 Reminder [Negligible sets and completions]. If $(\Omega, \mathcal{A})$ is a measurable space, the σ-algebra $\mathcal{A}$ does not usually contain all subsets of Ω. For a given probability measure, there may hence be a non-measurable set B which is contained in a $\mathbb{P}$-null set $A \in \mathcal{A}$. Sets which are contained in null sets are called **negligible sets**. (In other words, a null set is a negligible set which is also measurable.)

Even if a negligible set is not technically measurable, we might still argue that it is morally measurable, since we know what its measure would be if it happened to be in $\mathcal{A}$: $B \subset A$ and $\mathbb{P}(A) = 0$ implies the measure would have to be zero. With this rationale, we can simply regard all negligible sets as null sets, and add them to the σ-algebra. It is easy to check that the resulting set system is again a σ-algebra. It is called the **$\mathbb{P}$-completion** of $\mathcal{A}$, and denoted $\overline{\mathcal{A}}^{\mathbb{P}}$. Note that we cannot define a completion before specifying a measure on $(\Omega, \mathcal{A})$. $\triangleleft$

To work with rcll sample paths, we need a similar requirement for filtrations:

1.13 Definition. A filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called **complete** if it contains all $\mathbb{P}$-negligible sets, i.e. if $\mathcal{F}_t = \overline{\mathcal{F}}_t^{\mathbb{P}}$ for all t. $\triangleleft$

A second requirement is that the filtration itself is "smooth": Suppose for some index $s \in \mathbb{R}_+$, all filtrations $\mathcal{F}_t$ with $t > s$ suddenly contain much more information than $\mathcal{F}_s$—roughly speaking, the amount of information available "jumps up" at s. Such cases are excluded by the following definition:

1.14 Definition. A filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ is called **right-continuous** if

$$\mathcal{F}_t = \cup_{s>t} \mathcal{F}_s \qquad \text{for all } t \in \mathbb{R}_+ \,. \tag{1.38}$$

$\triangleleft$

Martingales with rcll paths. Theorem 1.16 below shows under which conditions a martingale, or even a submartingale, defined with respect to a complete and right-continuous filtration has rcll sample paths. To proof that result, we will need the following lemma. I will cheat a little and proof the theorem but not the lemma.

1.15 Lemma. *Let $\mathcal{F}$ be a filtration indexed by $\mathbb{R}_+$, and let $(X_t, \mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a submartingale. Then there is a null set N such that the following holds: For all $t \in \mathbb{R}_+$, there is a real-valued random variable X_{t+} such that*

$$X_{t+}(\omega) = \lim_{s\in\mathbb{Q}_+, s\searrow t} X_s(\omega) \qquad \textit{whenever } \omega \notin N \ . \tag{1.39}$$

Modify X_{t+} on the null set N by defining $X_{t+}(\omega) := 0$ for $\omega \in N$. If $\mathcal{F}$ is complete and right-continuous, then X_{t+} is integrable and

$$X_t \leq X_{t+} \qquad \textit{almost surely} \tag{1.40}$$

for each $t \in \mathbb{R}_+$, with equality almost surely if and only if the function $\mu(t) := \mathbb{E}[X_t]$ is right-continuous at t. ◁

Two remarks on Lemma 1.15:

(1) Note that the assertion in (1.39) is stronger than just almost sure convergence for each t: The latter would mean that, for each t, there is a null set N_t outside of which (1.39) holds. Since the index set is uncountable, $\cup_t N_t$ would *not* be guaranteed to be a null set. The lemma shows, however, that there is a single null set N outside of which (1.39) holds for all t.

(2) The lemma holds for a general submartingale. Recall that, if (X_t) is a martingale, then all X_t have identical mean $\mathbb{E}[X_t] = \mu_t = \mu$, so the function $t \mapsto \mathbb{E}[X_t]$ is constant and hence rcll. By the last assertion in the lemma, equality in (1.40) therefore holds automatically if (X_t) is a martingale.

1.16 Theorem [Submartingales with rcll sample paths]. *Let $(X_t, \mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a submartingale, where $\mathcal{F}$ is right-continuous and complete, and the function $\mu(t) := \mathbb{E}[X_t]$ is right-continuous. Then there exists a submartingale $(Y_t, \mathcal{F}_t)_{t\in\mathbb{R}_+}$ satisfying $X_t =_{\text{a.s.}} Y_t$ for all t whose paths $t \mapsto Y_t(\omega)$ are rcll almost surely.* ◁

Note that the result does not quite state that (X_t) is almost surely continuous, but rather that there is a martingale Y which is equivalent to X—in the sense that $X_t =_{\text{a.s.}} Y_t$, i.e. we are not able to distinguish Y from X by probabilistic means—and this equivalent martingale is rcll almost surely. The process Y is called a **version** or **modification** of X (since we modify the measurable function X on a null set to obtain Y; a precise definition will follow in a later chapter). Theorem 1.16 is our first example of a regularity result for a stochastic process, and we will see later on, in the chapter on stochastic processes, that most regularity results are stated in terms of the existence of almost surely regular versions.

Proof. Since (X_t) is a submartingale, Lemma 1.15 guarantees that the random variable X_{t+} defined in (1.39) exists for each t. Define $Y_t := X_{t+}$. Then the paths of Y_t are rcll by construction. Since $\mu(t)$ is right-continuous by hypothesis, Lemma 1.15 shows that $Y_t = X_t$ almost surely (equality holds in (1.40)). The only thing left to show is hence that $(Y_t, \mathcal{F}_t)$ is a submartingale, i.e. that $\int_A Y_s d\mathbb{P} \leq \int_A Y_t d\mathbb{P}$ for all $A \in \mathcal{F}_s$ and all $s < t$.

Let $s < t$. Then there are sequence $s_1 > s_2 > \ldots$ and $t_1 > t_2 > \ldots$ in $\mathbb{Q}_+$ such that $s_n \searrow s$ and $t_n \searrow t$. Since $s < t$, we can always choose the sequences such that $s_n < t_n$ for all n. As (X_n) is a submartingale, this implies

$$\int_A X_{s_n} d\mathbb{P} \leq \int_A X_{t_n} d\mathbb{P} \qquad \text{for all } A \in \mathcal{F}_{s_n} \ . \tag{1.41}$$

By (1.39), (X_{s_n}) converges to X_{s+} almost surely, and hence in probability. Since all $|X_{s_n}|$ are upper-bounded by $|X_{s_1}|$, convergence in probability implies $\mathbf{L}_1$-convergence. Hence, $\mathbb{E}[X_{s+}] = \lim_n \mathbb{E}[X_{s_n}]$, which in turn implies $\mathbb{E}[X_{s+}\mathbb{I}_A] = \lim_n \mathbb{E}[X_{s_n}\mathbb{I}A]$ for all $A \in \mathcal{F}_s$ (note that we can use $\mathcal{F}_s$ for X_{s+} by right-continuity). By the same device, $\mathbb{E}[X_{t+}\mathbb{I}_A] = \lim_n \mathbb{E}[X_{t_n}\mathbb{I}A]$, and we have

$$\int_A X_{s+} d\mathbb{P} = \lim \int_A X_{s_n}\mathbb{I}_A d\mathbb{P} \overset{(1.41)}{\leq} \lim \int_A X_{t_n}\mathbb{I}_A d\mathbb{P} = \int_A X_{t+} d\mathbb{P} \;, \tag{1.42}$$

so (X_{t+}) is indeed a submartingale. □

1.7. Tail bounds for martingales

A **tail bound** for a (real-valued) random variable X is an inequality of the form

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \;\leq\; f(\varepsilon) \;. \tag{1.43}$$

Almost all distributions—at least on unbounded sample spaces—concentrate most of their probability mass in some region around the mean, even if they are not unimodal. If we move sufficiently far away from the mean, the distribution decays. A tail bound quantifies how rapidly it does so; of interest is not so much how far out we have to move before the probability decays, but rather the shape of f in the region far away from the mean (in the "tails"). In particular, if f is of the form $f(\varepsilon) \leq ce^{-g(\varepsilon)}$ for some positive polynomial g, the distribution is said to exhibit **exponential decay**. If $f(\varepsilon) \sim \varepsilon^{-\alpha}$ for some $\alpha > 0$, it is called **heavy-tailed**.

1.17 Example [Hoeffding's inequality]. One of the most widely used tail bounds is the **Hoeffding bound**: Suppose $X_1, X_2, \ldots$ are independent, real-valued random variables (which need not be identically distributed), and each is bounded in the sense that $X_n \in [a_i, b_i]$ almost surely for some constants $a_i < b_i$. Then the empirical average $S_n = \frac{1}{n}(X_1 + \ldots + X_n)$ has tails bounded as

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq \varepsilon) \leq 2\exp\Big(-\frac{-2n\varepsilon^2}{\sum_i (b_i - a_i)^2}\Big) \;. \tag{1.44}$$

◁

If the X_n are dependent, we cannot generally hope for a bound of this form to hold. Remarkably, it still does if the X_n form a martingale. In this case, each X_n is required to be *conditionally* bounded given the previous value X_{n-1}, i.e. X_n must lie in the interval $[X_n - c_{n+1}, X_n + c_{n+1}]$.

1.18 Azuma-Hoeffding Inequality. *Let $(X_n, \mathcal{F}_n)_{i\in\mathbb{N}}$ be a martingale. Require that there exists a sequence of non-negative constants $c_n \geq 0$ such that*

$$|X_{n+1} - X_n| \leq c_{n+1} \qquad \textit{almost surely} \tag{1.45}$$

and $|X_1 - \mu| \leq c_1$ a.s. Then for all $\varepsilon > 0$

$$\mathbb{P}(|X_n - \mu| \geq \varepsilon) \leq 2\exp\Big(-\frac{\varepsilon^2}{2\sum_{j=0}^n c_n^2}\Big) \;. \tag{1.46}$$

◁

I can hardly stress enough how useful this result can be in applications: On the one hand, the only requirement on the martingales is the boundedness of increments, which makes the bound very widely applicable. On the other hand, when it is applicable, it is often also very sharp, i.e. we will not get a much better bound using more complicated methods. One application that has become very important over the past twenty or so years is the analysis of randomized algorithms:

1.19 Example [Method of bounded differences]. Suppose an iterative randomized algorithm computes some real-valued quantity X; since the algorithm is randomized, X is a random variable. At each iteration, the algorithm computes a candidate quantity X_n (which we usually hope to be a successively better approximation of some "true" value as n increases). If we can show that (1) the intermediate results X_n form a martingale and (2) the change of X_n from one step to the next is bounded, we can apply Theorem 1.18 to bound X. ◁

1.8. Application: The Pólya urn

Recall that an urn is a stochastic process defined by starting with a certain number of white and black balls, and repeatedly drawing a ball uniformly at random. You will be familiar with *sampling with replacement* (an urn in which the ball is replaced after having been drawn) and *sampling without replacement* (the ball is removed).

More generally, an urn is a process where, each time we draw a ball, we may or may not replace it, and may or may not add additional balls to the urn. It can be parametrized as

$$\begin{pmatrix} w & a \\ d & b \end{pmatrix} \qquad \text{where} \qquad \begin{array}{r l} w = & \#\text{ initial white balls} \\ b = & \#\text{ initial black balls} \end{array} . \tag{1.47}$$

Each time a ball is drawn, we replace it by a balls of the same color and d balls of the opposite color. Important examples are:

$a = 0$	$d = 0$	Sampling with replacement
$a = -1$	$d = 0$	Sampling without replacement
$a > 0$	$d = 0$	Pólya urn
$a = -1$	$d = 1$	Ehrenfest urn (or Ehrenfest heat transfer model)

In particular, a **Pólya urn** with parameters (w_0, b_0, a) is a stochastic process defined by an urn initially containing w_0 white and b_0 black balls. At each step, draw a ball from the urn at random; then replace the ball, and add an additional a balls of the same color. We define X_n as the fraction of white balls after n steps,

$$X_n = \frac{\#\text{ white balls after } n \text{ steps}}{(\#\text{ white balls } + \#\text{ black balls}) \text{ after } n \text{ steps}} . \tag{1.48}$$

1.20 Proposition. *The proportions X_n converge almost surely: There exists a random variable X_∞ such that* $\lim_{n\to\infty} X_n(\omega) = X_\infty(\omega)$ *almost surely.* ◁

I want to complement this existence result with a result on the form of the limit, which we will not proof (since it does not involve a martingale argument):

1.21 Fact. The limiting distribution of proportions, i.e. the law of X_∞, is the beta distribution on $[0, 1]$ with density

$$p(x) = B(w/a, b/a)x^{w/a-1}(1-x)^{b/a-1} . \tag{1.49}$$

where B denotes the beta function. ◁

Before we proof the existence of the limit, reconsider for a moment what Proposition 1.20 tells us:

- We know from basic calculus that a sequence need not converge—the proportions could fluctuate perpetually. Proposition 1.20 shows that this is not the case here: Even though the sequence is generated at random by the urn, it *always* converges to a limit. Roughly speaking, if we would run the process for an infinite amount of time to obtain the proportions X_∞, and then restart it with those proportions, they would never change again (which of course can only be true since the urn has swollen to contain an infinite number of balls).
- On the other hand, the limit is random—if we start the process from the same initial values twice, we will typically obtain two distinct limiting proportions. (In fact, since the limiting distribution is continuous, they differ with probability 1.)

Proof of Proposition 1.20. We will show that (X_n) is a martingale, and then apply the martingale convergence theorem to verify existence of the limit. Let W_n and B_n respectively denote the number of white and black balls after n draws. The probability of observing a white ball in the $(n+1)$st draw is, conditionally on (W_n, B_n),

$$p_{n+1} = \frac{W_n}{W_n + B_n} \; . \tag{1.50}$$

Hence,

$$X_{n+1}|W_n, B_n = \begin{cases} \frac{W_n+a}{W_n+B_n+a} & \text{with probability } p_n \\ \frac{W_n}{W_n+B_n+a} & \text{with probability } 1-p_n \end{cases} \; . \tag{1.51}$$

The history of the process, up to step n, is given by the nth σ-algebra in the filtration $\mathcal{F}_n$. The conditional expectation of X_{n+1} given the history of the process is hence

$$\begin{aligned} \mathbb{E}[X_{n+1}|\mathcal{F}_n] =& \frac{W_n+a}{W_n+B_n+a} p_{n+1} + \frac{W_n}{W_n+B_n+a}(1-p_{n+1}) \\ =& \ldots = \frac{W_n}{W_n+B_n} = X_n \; . \end{aligned} \tag{1.52}$$

Since X_n is also clearly integrable, it is hence a martingale. We now use the martingale convergence theorem, in the form of Corollary 27.1 in [Probability I]—recall: If (X_n) is a non-negative supermartingale, then $\lim_{n\to\infty} X_n$ exists almost surely—which completes the proof. □

1.22 Remark [Preferential attachment networks]. The Pólya urn may seem primitive, but it has many important applications. One example are random graphs used as models for certain social networks: A **preferential attachment graph** is generated as follows. Fix an integer $m \geq 1$. Start with a graph consisting of a single vertex. At each step, insert a new vertex, and connect it to m randomly selected vertices in the current graph. These vertices are selected by **degree-biased sampling**, i.e. each vertex is selected with probability proportional to the number of edges currently attached to it. You will notice that (1) the placement of the next edge depends only on the vertex degrees (not on which vertex is connected to which), and (2) the model is basically a Pólya urn (where each vertex represents a color, and the degrees are the number of balls per color). It is hence not surprising that most proofs on asymptotic properties of this model involve martingales. This, in turn, is one of the reasons why this model is as well-studied as it is in the applied probability literature—the applicability of martingales makes it tractable, so we study it because we can. ◁

1.9. Application: The Radon-Nikodym theorem

Let P be probability measure on a measurable space $(\mathcal{X}, \mathcal{A})$, and let μ be a finite measure on the same space (that is, $\mu(\mathcal{X}) < \infty$). Recall that a **density** of μ with respect to P is an integrable function $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$ satisfying

$$\mu(dx) =_{\text{a.e.}} f(x)P(dx) . \tag{1.53}$$

When does a density exist for a given pair μ and P?

Equation (1.53) says that f transforms the set function P into μ by reweighting it point-wise. Since it we cannot transform 0 into a positive number by multiplication with any value, this clearly requires that μ vanishes wherever P vanishes, that is,

$$P(A) = 0 \qquad \Rightarrow \qquad \mu(A) = 0 \tag{1.54}$$

for all measurable sets A in $\mathcal{X}$. Recall that μ is called **absolutely continuous** with respect to P if μ and P satisfy (1.54)—in symbols, $\mu \ll P$. The term “absolute continuity” derives from the following:

1.23 Fact. $\nu \ll \mu$ holds if and only if

$$\text{for all } \varepsilon > 0 \text{ exists } \delta > 0 \text{ such that } \mu(A) \leq \delta \Rightarrow \nu(A) \leq \varepsilon \tag{1.55}$$

holds for all measurable sets A. ◁

That absolute continuity is a necessary condition for (1.53) to hold is obvious. Remarkably, it is also the only condition required:

1.24 Radon-Nikodym theorem (for probability measures). *Let P be a probability measure and μ a finite measure on a measurable space $\mathcal{X}$. Then μ has a density with respect to P if and only if $\mu \ll P$. Any two such densities differ only on a P null set.* ◁

Proof of the theorem. The idea of the proof is to subdivide the space $\mathcal{X}$ into a partition of n disjoint sets A_j, and define

$$Y_{(A_1,\ldots,A_n)}(x) := \sum_{j=1}^{n} f(A_j)\mathbb{I}_{A_j}(x) \qquad \text{where } f(A_j) := \begin{cases} \frac{\mu(A_j)}{P(A_j)} & P(A_j) > 0 \\ 0 & P(A_j) = 0 \end{cases} . \tag{1.56}$$

Think of Y as a “discretization” of the density f whose existence we wish to establish. Roughly speaking, we will make the partition finer and finer (by making the sets A_j smaller and increasing n), and obtain f as the limit of Y. Since Y is a measurable function on the space $\mathcal{X}$, which forms a probability space with P, we can regard the collection of Y we obtain for different partitions as a martingale.

More formally, we construct a directed index set $\mathbb{T}$ as follows: A **finite measurable partition** $H = (A_1, \ldots, A_n)$ of $\mathcal{X}$ is a subdivision of $\mathcal{X}$ into a finite number of disjoint measurable sets A_i whose union is $\mathcal{X}$. Let $\mathbb{T}$ be the set of all finite measurable partitions of $\mathcal{X}$. Now we have to define a partial order: We say that a partition $H_2 = (B_1, \ldots, B_m)$ is a **refinement** of another partition $H = (A_1, \ldots, A_n)$ if every set B_j in H_2 is a subset of some set A_i in H_1; in words, H_2 can be obtaine from H_1

by splitting sets in H_1 further, without changing any of the existing set boundaries in H_1. We then define a partial order on $\mathbb{T}$ as

$$H_1 \preceq H_2 \qquad \Leftrightarrow \qquad H_2 \text{ is a refinement of } H_1 . \tag{1.57}$$

Since each index $s \in \mathbb{T}$ is now a measurable partition, we can define $\mathcal{F}_s$ as the σ-algebra generated by the sets in s,

$$\mathcal{F}_s := \sigma(A_1, \dots, A_n) \qquad \text{if } s = (A_1, \dots, A_n) . \tag{1.58}$$

1.25 Lemma. $(Y_s, \mathcal{F}_s)_{s \in \mathbb{T}}$ *is a uniformly integrable martingale.* ◁

Proof. It is easy to check the martingale property; we will show uniform integrability. Let $\alpha > 0$ and choose some index $s = (A_1, \dots, A_n)$. (Recall the definition of uniform integrability in (1.15); we choose g as the constant function with value α.) Then

$$\begin{aligned} \int_{\{|Y_s| \geq \alpha\}} |Y_s(x)| P(dx) &\overset{Y_s \geq 0}{=} \int_{\{Y_s \geq \alpha\}} Y_s(x) P(dx) \\ &= \int_{\mathcal{X}} \sum_{j=1}^{n} \frac{\mu(A_j)}{P(A_i)} \mathbb{I}\{x \in A_i \text{ and } Y_s(x) \geq \alpha\} P(dx) \\ &= \mu\{Y_s \geq \alpha\} . \end{aligned} \tag{1.59}$$

Since Y_s is a positive random variable, Markov's inequality for Y_s reads

$$P\{Y_s \geq \alpha\} \leq \frac{1}{\alpha} \mathbb{E}[Y_s] = \frac{1}{\alpha} \mu(\mathcal{X}) . \tag{1.60}$$

Now we use (1.55): For a given $\varepsilon > 0$, choose some δ which satisfies (1.55), and set $\alpha > \frac{\mu(\mathcal{X})}{\delta}$. Then (1.60) implies $P\{Y_s \geq \alpha\} \leq \delta$, and hence

$$\int_{\{|Y_s| \geq \alpha\}} |Y_s(x)| P(dx) \overset{(1.59)}{=} \mu\{Y_s \geq \alpha\} \overset{(1.55)}{\leq} \varepsilon . \tag{1.61}$$

The choice of ε and δ is independent of the index s (since the rightmost term in (1.60) does not depend on s). Hence, $(Y_s, \mathcal{F}_s)$ is uniformly integrable. □

The proof of uniform integrability is the only real leg work in the proof of the Radon-Nikodym theorem. The rest is easy:

Proof of Theorem 1.24. Since $(Y_s, \mathcal{F}_s)$ is a uniformly integrable martingale, Theorem 1.8 shows that an integrable random variable Y_∞ with $\mathbb{E}[Y_\infty | \mathcal{F}_s] =_{\text{a.s.}} Y_s$ exists and is uniquely determined, up to almost sure equivalence. To verify that Y_∞ is a density, we have to show that $\mu(A) = \int Y_\infty(x) P(dx)$, and that Y_∞ is nonnegative almost surely. The identity $\mathbb{E}[Y_\infty | \mathcal{F}_s] =_{\text{a.s.}} Y_s$ means

$$\int_A Y_\infty(x) P(dx) = \int_A Y_s(x) P(dx) \qquad \text{for all } A \in \mathcal{F}_s . \tag{1.62}$$

For each A, the index set $\mathbb{T}$ contains in particular the partition $s = (A, \bar{A})$ consisting only of A and its complement $\bar{A}$. For this s, the previous equation becomes

$$\begin{aligned} \int_A Y_\infty(x) P(dx) &= \int_A Y_s(x) P(dx) \\ &= \int_A \Big(\frac{\mu(A)}{P(A)} \mathbb{T}_A(x) + \frac{\mu(\bar{A})}{P(\bar{A})} \mathbb{T}_{\bar{A}}(x) \Big) P(dx) = \mu(A) . \end{aligned} \tag{1.63}$$

This also implies that $Y_\infty \geq 0$ almost everywhere—otherwise, there would be a non-null set A (i.e. $P(A) > 0$) on which Y_∞ takes only negative values, and by the previous equation, that would yield $\mu(A) < 0$. $\square$

FYI: The general case. The existence of densities is of course not limited to the case where P is a probability measure, or even finite; I have stated the result in the form above in order to prove it using martingales (and because the case where P is not normalized is not particularly relevant in the following). Nonetheless, I should stress that Theorem 1.24 still holds in precisely this form if μ and P are both σ-finite measures:

1.26 Radon-Nikodym theorem. *Let μ and ν be σ-finite measures on a measurable space $(\mathcal{X}, \mathcal{A})$. Then there exists a measurable function $f : \Omega \to [0, \infty)$ with $\mu(A) = \int_A f d\nu$ for all $A \in \mathcal{A}$ if and only if $\mu \ll \nu$.* $\triangleleft$

Indeed, there is a generalization beyond even the σ-finite case: ν need not be σ-finite, and μ need not even be a measure. I have found this result very useful at times, and since it is not widely known and almost impossible to find in textbooks, I state it here without proof (which you can read up in [4, 232E] if you feel so inclined):

1.27 Generalized Radon-Nikodym theorem. *Let ν be a measure on a measurable space $(\mathcal{X}, \mathcal{A})$, and let $\mu : \mathcal{A} \to \mathbb{R}_{\geq 0}$ be a finitely additive set function. Then there is a measurable function $f : \mathcal{X} \to \mathbb{R}_{\geq 0}$ satisfying $\mu(A) = \int_A f d\nu$ for all $A \in \mathcal{A}$ if and only if:*

(i) *μ is absolutely continuous with respect to ν.*
(ii) *For each $A \in \mathcal{A}$ with $\mu(A) > 0$, there exists a set $B \in \mathcal{A}$ such that $\nu(B) < \infty$ and $\mu(A \cap B) > 0$.*

If so, f is uniquely determined ν-a.e. $\triangleleft$

Bibliography

Main references

[A&B] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis.* 3rd. Springer, 2006.

[Cinlar] E. Cinlar. *Probability and Stochastics.* Springer, 2011.

[Probability I] J. Jacod and P. Protter. *Probability Essentials.* Springer, 2004.

[Kallenberg] O. Kallenberg. *Foundations of Modern Probability.* 2nd. Springer, 2001.

Other references

[1] H. Bauer. *Probability Theory.* W. de Gruyter, 1996.

[2] P. Billingsley. *Convergence of Probability Measures.* J. Wiley & Sons, 1999.

[3] P. Billingsley. *Probability and Measure.* J. Wiley & Sons, 1995.

[4] D. H. Fremlin. *Measure Theory.* Vol. I-IV. University of Essex, 2003–2006.

[5] D. Pollard. *A User's Guide to Measure Theoretic Probability.* Cambridge University Press, 2002.