2. Martingales

2.1. Stochastic Processes; Filtrations; Stopping Times

This lecture is about stochastic processes, that is we are interested in "time" dependent random outcome. We denote the set of different times t by $\mathbf{T}$.

Definition 2.1. A *stochastic process* – or simply process – is a family $X = (X_t)_{t\in\mathbf{T}}$ of random variables $X_t : \Omega \to \mathbb{R}$ indexed by $\mathbf{T}$.

Intending to model the time, $\mathbf{T}$ should have a *"direction"*. Therefore, throughout this lecture, we always assume that $\mathbf{T}$ is a subset of the positive extended real line $[0,\infty]$.[35] We will assume that $0 \in \mathbf{T}$ and denote $T := \sup \mathbf{T}$ which might be ∞. If not otherwise specified, elements of $\mathbf{T}$ are designed by the letter $s, t, u, \ldots$.
For the first part of the lecture $\mathbf{T}$ will be discrete, that is $\mathbf{T} = \{0, 1, \ldots\}$. Later, as we construct the stochastic integral, we consider more general times set such as $\mathbf{T} = [0, T]$ where $T > 0$ is a fixed time horizon or $\mathbf{T} = \{2^k T/2^n : 0 \leq k \leq 2^n, n \in \mathbb{N}\}$ the dyadic times points between 0 and T.
The mappings $t \mapsto X_t(\omega)$ for $\omega \in \Omega$ are called the paths – or sample paths, trajectories – of the process. A stochastic process $X = (X_t)_{t=0,\ldots,T}$ may also be viewed as

- a single random variable

$$\begin{aligned} X : \Omega \times \{0,\ldots,T\} &\longrightarrow \mathbb{R} \\ (\omega, t) &\longmapsto X_t(\omega) \end{aligned}$$

  where the σ-algebra on $\Omega \times \{0,\ldots,T\}$ is given by the product σ-algebra $\mathcal{F} \otimes 2^{\{0,\ldots,T\}}$.

- a measurable function with values in the sample space

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \times \cdots \times \mathbb{R} = \mathbb{R}^{T+1} \\ \omega &\longmapsto (X_0(\omega), \ldots, X_T(\omega)) \end{aligned}$$

  where the σ-algebra on the sample space is the product Borel σ-algebra on $\mathbb{R}^{T+1}$.

Exercice 2.2. Show that the three definition of a stochastic process in finite discrete time are equivalent.$\diamondsuit$

Example 2.3. Consider now our example of coin tossing but infinitely many times. As seen, the state space is defined as follows

$$\Omega = \prod_{t\in\mathbb{N}} \{-1, 1\} = \{-1, 1\}^{\mathbb{N}} = \{\omega = (\omega_t) \colon \omega_t = \pm 1 \text{ for every } t\}$$

On each $\Omega_t = \{-1, 1\}$ we consider the σ-algebra $\mathcal{F}_t = \{\emptyset, \{-1\}, \{1\}, \{-1, 1\}\}$ and on Ω the product σ-algebra $\mathcal{F} = \otimes \mathcal{F}_t$. We saw that it is generated by the finite product cylinders:

$$C = \{\omega = (\omega_t) \in \Omega \colon \omega_{t_k} = e_k, k = 1, \ldots n\} \tag{2.1}$$

for a given set of values $e_k \in \{-1, 1\}$, and times $t_k \in \mathbb{N}$, $k = 1, \ldots, n$. Suppose that the probability of getting head, that is 1, for the coin toss is equal to $p \in [0, 1]$, we can define on the collection of these finite product cylinder $\mathcal{R}$, which is a semi-ring, a content $P : \mathcal{R} \to [0, 1]$ given by

$$P[C] = p^l (1-p)^{n-l}$$

[35] More generally, though, any directed set can be considered with or without an origin. For instance in statistical mechanics, indexing a process by subsets of a countable set ordered by inclusion.

for every $C \in \mathcal{R}$ of the form (2.1) where l is equal to the number of those $k = 1, \dots, n$ where $e_k = 1$. We will show later that this content fulfills the sub-additivity property required in Caratheodory's theorem and therefore extends to a probability measure on Ω.
Now that we have a probability space at hand, we can define the stochastic processes $X = (X_t)$ and $S = (S_t)$ by

$$X_0(\omega) = 0 \quad \text{and} \quad X_t(\omega) = \begin{cases} 1 & \text{if } \omega_t = 1 \\ -1 & \text{if } \omega_t = -1 \end{cases} = \omega_t, \quad t = 1, \dots, \quad \omega \in \Omega$$

and

$$S_t = x_0 + \sum_{s=0}^{t} X_s, \quad t = 0, 1, \dots$$

where $x_0 \in \mathbb{R}$ is the start value, or start price of S. The stochastic process S is called the *random walk* and the process X tells us what is the result of the coin toss at time t.
As an exercise in Ipython, make a plot of 5 sample paths of the random walk for

- an horizon of $T = 10, 100, 1.000, 100.000$;
- for $p = 1/3, 1/2, 2/3$. $\Diamond$

As such, a process is nothing else than an arbitrary family of random variables indexed by the time. However, our intuitive understanding of a process rather corresponds to X_s "having less, or knowing less" than X_t whenever $s \leq t$. To model this intuition we use an increasing set of information.

Definition 2.4. A *filtration* $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbf{T}}$ is a family of σ-algebras on Ω indexed by $\mathbf{T}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ whenever $s \leq t$ with $s, t \in \mathbf{T}$. A measurable space together with a filtration is called a *filtered* space. A stochastic process X is $\mathbb{F}$-*adapted* – or simply adapted – if X_t is $\mathcal{F}_t$-measurable for every $t \in \mathbf{T}$;

The σ-algebras in a filtration becomes finer and finer due the inclusion. It means that the considered events at time t provide more information than the ones at previous times. Filtration can be given, but also generated by stochastic processes.

Definition 2.5. Let X be a stochastic process. The family of σ-algebra

$$\mathcal{F}_t^X = \sigma(X_s \colon s \leq t) := \sigma\left(\left\{X_s^{-1}(A) \colon A \in \mathcal{B}(\mathbb{R}), s \leq t\right\}\right), \quad t \in \mathbf{T}$$

is a filtration called the filtration generated by X which we denote by $\mathbb{F}^X$.

The fact that the filtration generated by a stochastic process is indeed a filtration is easy to verify.

Example 2.6. In our random walk example, we did not specify a filtration, but we can consider the following sequences of σ-algebras for $t \in \mathbb{N}_0$

- $\mathcal{F}_t^X$;
- $\mathcal{F}_t^S$;
- $\mathcal{G}_t := \sigma(S_t)$;
- $\mathcal{H}_t := \sigma(X_t)$;

As an exercise, try to figure out which sequence of $sigma$-algebra is a filtration give an expression for their generators in the case where $x_0 = 0$. $\Diamond$

From now on, until we mention otherwise,

$$\mathbf{T} = \{0, 1, \ldots\}!!!!! \quad \text{and} \quad (\Omega, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbf{T}}) \text{ is a filtrated space.}$$

A further important notion in the theory of stochastic processes, are the so called stopping time. As an illustration of which, consider the following game. You pay 10 Kuai and a coin will be tossed every minute. If it is head you win one Kuai, if it is tail you loose one Kuai. So the evolution of your wealth as times goes by follows

$$S_t := 10 + \sum_{k=1}^{t} X_t$$

However you would like to leave the game before loosing too much money, that is, you stop the first time you reach let's say 3 kuai.

$$\tau = \inf \{t \colon S_t \leq 3\}$$

This time however is no longer known but random since it depends on the random outcomes $S_t(\omega)$. This is the same on financial markets, where investors wants to know the time until which a company might be bankrupt for instance, or the time until they reach a certain level of wealth in their strategic investment.

Exercice 2.7. In the case when the coin toss is fair, what is the probability that you exit the game before 100 minutes? $\Diamond$

Intuitively, a *random time* gives information about when a random event occurs.

Definition 2.8. On a measurable space, a *random time* is a measurable mapping $\tau : \Omega \to \mathbf{T} \cup T$. Given a filtration, a random time is a *stopping time* if $\{\tau \leq t\} \in \mathcal{F}_t$ for every $t \in \mathbf{T}$.

For a process X and a subset B of $\mathbb{R}$, we define the *hitting time* of X in B as

$$\tau_B(\omega) = \inf\{t \in \mathbf{T} : X_t(\omega) \in B\}.$$

This function is not necessarily random even if X is adapted, however we have the following.

Proposition 2.9. *If X is an adapted process and B is Borel, then τ_B is a stopping time.*

Proof. Let $t \in \mathbf{T}$. Since $\mathbf{T}$ is discrete, the infimum is in fact a minimum. Hence, it follows that

$$\{\tau_B \leq t\} = \bigcup_{s=0,\ldots,t} \{X_s \in B\}$$

Since X is adapted, it follows that $A_s = \{X_s \in B\} \in \mathcal{F}_s$ for every s. Furthermore, $\mathbf{F}$ being a filtration, it holds $\mathcal{F}_s \subseteq \mathcal{F}_t$. Hence, $A_s \in \mathcal{F}_s$ for every $s \leq t$. Finally, $\mathcal{F}_t$ being a σ-algebra, the finite union of A_s for $s \leq t$ is also in $\mathcal{F}_t$ showing that $\{\tau_B \leq t\}$ is a stopping time. $\square$

Let us collect some standard properties of stopping times.

Proposition 2.10. *The following assertions hold*

(a) $\tau + \sigma$, $\tau \vee \sigma$ *and* $\tau \wedge \sigma$ *are stopping times as soon as* τ, σ *are stopping times.*

(b) $\lim \tau^n$ *is a stopping time as soon as* (τ^n) *is an increasing sequence of stopping times.*

(c) If τ *is a stopping time, then the collection* $\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$ *is a* σ*-algebra and* τ *is* $\mathcal{F}_\tau$*-measurable.*

(d) *For any two stopping times, it holds $\mathcal{F}_\sigma \cap \mathcal{F}_\tau = \mathcal{F}_{\sigma\wedge\tau}$. In particular, $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$, if $\sigma \leq \tau$. For every integrable random variable X with respect to some probability on $\mathcal{F}$, it holds $E[E[X \mid \mathcal{F}_\sigma] \mid \mathcal{F}_\tau] = E[X \mid \mathcal{F}_{\sigma\wedge\tau}]$.*

Proof. *(a)* follows from

$$\{\tau + \sigma \leq t\} = \bigcup_{q=0}^{t} \{\sigma \leq t - q\} \cap \{\tau \leq q\} \in \mathcal{F}_t$$

$$\{\tau \vee \sigma \leq t\} = \{\tau \leq t\} \cap \{\sigma \leq t\} \in \mathcal{F}_t \quad \text{ and } \quad \{\tau \wedge \sigma \leq t\} = \{\tau \leq t\} \cup \{\sigma \leq t\} \in \mathcal{F}_t$$

(b) follows from $\{\lim \tau^n \leq t\} = \{\tau^n \leq t : \text{ for all } n\} = \cap_n \{\tau^n \leq t\} \in \mathcal{F}_t$.

(c) Clearly $\emptyset, \Omega \in \mathcal{F}_\tau$. For $A \in \mathcal{F}_\tau$ it holds $A^c \cap \{\tau \leq t\} = (A \cup \{\tau > t\})^c = [(A \cap \{\tau \leq t\}) \cup \{\tau \leq t\}^c]^c \in \mathcal{F}_t$. Finally, for $(A_n) \subseteq \mathcal{F}_\tau$ it holds $(\cup A_n) \cap \{\tau \leq t\} = \cup(A_n \cap \{\tau \leq t\}) \in \mathcal{F}_t$.

(d) Follows from $\{\tau \wedge \sigma \leq t\} = \{\tau \leq t\} \cup \{\sigma \leq t\}$. Indeed, let $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$, it follows that $A \cap \{\sigma \leq t\} \in \mathcal{F}_t$ and $A \cap \{\tau \leq t\} \in \mathcal{F}_t$ for every t. Hence $(A \cap \{\sigma \leq t\}) \cup (A \cap \{\tau \leq t\}) \in \mathcal{F}_t$ for every t, but $(A \cap \{\sigma \leq t\}) \cup (A \cap \{\tau \leq t\}) = A \cap (\{\sigma \leq t\} \cup \{\tau \leq t\}) = A \cap \{\sigma \wedge \tau \leq t\}$ showing that $A \in \mathcal{F}_{\sigma\wedge\tau}$ and therefore $\mathcal{F}_\sigma \cap \mathcal{F}_\tau \subseteq \mathcal{F}_{\sigma\wedge\tau}$. Reciprocally, let $A \in \mathcal{F}_{\tau\wedge\sigma}$, it follows that $A \cap (\{\sigma \leq t\} \cup \{\tau \leq t\}) = (A \cap \{\sigma \leq t\}) \cup (A \cap \{\tau \leq t\}) \in \mathcal{F}_t$ for every t. Since $\{\sigma \leq t\} \in \mathcal{F}_t$, it follows that $(A \cap \{\sigma \leq t\}) \cup (A \cap \{\tau \leq t\}) \cap \{\sigma \leq t\} = A \cap \{\sigma \leq t\}$ is also in $\mathcal{F}_t$ for every t. Hence $A \in \mathcal{F}_\sigma$. Symetrically, $A \in \mathcal{F}_\tau$ and therefore $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$ showing that $\mathcal{F}_{\sigma\wedge\tau} \subseteq \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Let us show that $E[E[X|\mathcal{F}_\sigma]|\mathcal{F}_\tau] = E[X|\mathcal{F}_{\sigma\wedge\tau}]$. Let $A \in \mathcal{F}_{\sigma\wedge\tau}$. Since $A \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$, it follows that

$$E\left[E[E[X|\mathcal{F}_\sigma]|\mathcal{F}_\tau]1_A\right] = E\left[E\left[E\left[X1_A|\mathcal{F}_\sigma\right]|\mathcal{F}_\tau\right]\right] = E[X1_A]$$

for every $A \in \mathcal{F}_{\sigma\wedge\tau}$. Thus, per definition of the conditional expectation, $E[E[X \mid \mathcal{F}_\sigma] \mid \mathcal{F}_\tau] = E[X \mid \mathcal{F}_{\sigma\wedge\tau}]$. □

Proposition 2.11. *Let X be an adapted process and τ a stopping time. If τ is finite, that is $\tau < \infty$, then $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$ is an $\mathcal{F}_\tau$-measurable random variable. Furthermore, for every stopping time τ, the process $X^\tau := (X_{\cdot\wedge\tau})$ is an adapted process.*

Proof. Let B be a Borel subset of $\mathbb{R}$ and τ be a finite stopping time. It holds

$$\{X_\tau \in B\} = \cup \{X_\tau \in B\} \cap \{\tau = t\} = \cup \left(\{X_t \in B\} \cap \{\tau = t\}\right) \in \mathcal{F}$$

hence X_τ is measurable. Let $A = \{X_\tau \in B\}$ and fix t. It holds

$$A \cap \{\tau \leq t\} = \cup_{s \leq t} \left(\{X_s \in B\} \cap \{\tau = s\}\right)$$

However $\{X_s \in B\} \cap \{\tau = s\} = \{X_s \in B\} \cap \{\tau \leq s\} \cap \{\tau \leq s-1\}^c \in \mathcal{F}_s \subseteq \mathcal{F}_t$ for every $s \leq t$. Hence, $A \cap \{\tau \leq t\} \in \mathcal{F}_t$ for every t showing that $A \in \mathcal{F}_\tau$ by definition. Thus X_τ is $\mathcal{F}_\tau$-measurable.
Let now τ be any stopping time, it follows that $t \wedge \tau$ is a finite stopping time smaller than t, and therefore $\mathcal{F}_{t\wedge\tau} \subseteq \mathcal{F}_t$. Since $X^\tau_t = X_{t\wedge\tau}$ is $\mathcal{F}_{t\wedge\tau}$-measurable, it is in particular $\mathcal{F}_t$ measurable so that X^τ is an adapted process too. □

Let us now define one of the most important object of stochastic analysis, namely, the *stochastic integral*. Given an adapted process X and a *predictable process* H, that is H_t is $\mathcal{F}_{t-1}$-measurable for every t, we denote by $H \bullet X$ the process

$$H \bullet X_t = H_0 X_0 + \sum_{s=1}^{t} H_s \left(X_s - X_{s-1}\right) = H_0 X_0 + \sum_{s=1}^{t} H_s \Delta X_s.$$

Lemma 2.12. *The collection $\mathcal{S}$ of all predictable processes is a vector space. Given an adapted process X, the stochastic integral with respect to X is a linear operator with values in the vector space of adapted processes.*
Furthermore, for every stopping time τ, predictable process H, the stopped process H^τ as well as the process $H1_{\{\cdot\le\tau\}}$ are predictable. It holds

$$H1_{\{\cdot\le\tau\}} \bullet X = H \bullet X^\tau = (H \bullet X)^\tau \,.$$

Proof. The fact that $\mathcal{S}$ is a vector space is direct, as well as the linearity of the stochastic integral. Let now H be a predictable process and τ be a stopping time. Let us show that according to the previous proposition, the adapted process H^τ is predictable. It holds $\Omega = \{\tau \le t-1\} \cup \{t \le \tau\} = (\cup_{s\le t-1}\{\tau = s\}) \cup \{t \le \tau\}$. For every $s = 0, \dots, t-1$, it holds $\{\tau = s\} \in \mathcal{F}_s \subseteq \mathcal{F}_{t-1}$. Also, $\{t \le \tau\} = \{\tau < t\}^c = \{\tau \le t-1\}^c \in \mathcal{F}_{t-1}$. Hence, we have a partition of Ω into $\mathcal{F}_{t-1}$-measurable sets. It holds

$$H_t^\tau = \sum_{s=0}^{t-1} H_{t\wedge\tau} 1_{\{\tau=s\}} + H_{t\wedge\tau} 1_{\{t\le\tau\}} = \sum_{s=0}^{t-1} H_s 1_{\{\tau=s\}} + H_t 1_{\{t\le\tau\}}$$

However $H_s 1_{\{\tau=s\}}$ is $\mathcal{F}_s$-measurable as product of a $\mathcal{F}_{s-1}$-measurable random variable H_s and $\mathcal{F}_s$ measurable random variable $1_{\{\tau=s\}}$ for every $s = 0, \dots, t-1$. So they are in particular $\mathcal{F}_{t-1}$-measurable since $\mathbb{F}$ is a filtration. Also $H_t 1_{\{t\le\tau\}}$ is $\mathcal{F}_{t-1}$-measurable random variable as product of the $\mathcal{F}_{t-1}$-measurable random variables H_t and $1_{\{t\le\tau\}}$. It follows that H_t^τ is $\mathcal{F}_{t-1}$-measurable as sum of $\mathcal{F}_{t-1}$-measurable random variables. As for the second case, since $\{t \le \tau\}$ is $\mathcal{F}_{t-1}$-measurable for every t, it follows that $1_{\{\cdot\le\tau\}}$ is predictable and, therefore, so is $H1_{\{\cdot\le\tau\}}$.
Let us show the equality per induction. The case $t = 0$ is trivial. Suppose that the equality holds up to time $t-1$, since $H \bullet X_t = H \bullet X_{t-1} + H_t(X_t - X_{t-1})$ we just have to show that $(H_t 1_{\{t\le\tau\}}(X_t - X_{t-1})) = (H_t(X_t^\tau - X_{t-1}^\tau)) = (H_t(X_t - X_{t-1}))^\tau$. This is however clear since $X_t^\tau - X_{t-1}^\tau = X_{t\wedge\tau} - X_{t-1\wedge\tau}$ is equal to 0 on $\{\tau \le t-1\}$ and $X_\tau - X_{t-1}$ on $\{t \le \tau\}$. ☐

In particular, since $1 \bullet X = X$, it follows that

$$1_{\{\cdot\le\tau\}} \bullet X = X^\tau.$$

Stochastic integrals do have particular properties when the integrator belongs to the class of martingales.

2.2. Martingales

Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbf{T}}, P)$ be a filtrated probability space.

Definition 2.13. A process X is called a *martingale* if

a) X is adapted;

b) X_t is integrable for every $t \in \mathbf{T}$;

c) $X_s = E[X_t \,|\, \mathcal{F}_s]$ whenever $s \le t$, $s, t \in \mathbf{T}$.

A process X is called a *super-martingale* if instead of c) we require

c') $X_s \ge E[X_t \,|\, \mathcal{F}_s]$ whenever $s \le t$, $s, t \in \mathbf{T}$.

A process X is called a *sub-martingale* if instead of c) we require

c") $X_s \le E[X_t \,|\, \mathcal{F}_s]$ whenever $s \le t$, $s, t \in \mathbf{T}$.

We say that a martingale, super-martingale or sub-martingale X is closed on the right if there exists $\xi \in L^1$ such that $E[\xi \mid \mathcal{F}_t] = X_t$, $E[\xi|\mathcal{F}_t] \leq X_t$ or $E[\xi|\mathcal{F}_t] \geq X_t$, respectively, for every $t \in \mathbf{T}$.

Remark 2.14. Note that a martingale is in particular a super- and a sub-martingale at the same time. Furthermore, given $\xi \in L^1$, the process given by $X_t = E[\xi|\mathcal{F}_t]$ for $t \in \mathbf{T}$ defines a martingale.[36] ♦

Example 2.15. Consider the random walk S of example 2.3 in its own filtration $\mathbb{F}^S$. If

- $p = 1/2$, then S is a martingale;
- $p \geq 1/2$, then S is a sub-martingale;
- $p \leq 1/2$, then S is a super-martingale. ♢

Proposition 2.16. *Let X be an adapted process and $\varphi : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\varphi(X_t)$ is integrable for every t.*

- *If X is a martingale and φ is convex, then $Y = (\varphi(X_t))$ is a sub-martingale;*
- *If X is a martingale and φ is concave, then $Y = (\varphi(X_t))$ is a super-martingale;*
- *If X is a sub-martingale and φ is convex and increasing, then $Y = (\varphi(X_t))$ is a sub-martingale.*

Proof. Since a process Y is a sub-martingale if and only if $-Y$ is a super-martingale and φ is convex if and only if $-\varphi$ is concave, we just show the first point to get the second. Clearly, Y is adapted. By assumption Y_t is integrable for every t. Finally, using Jensen's inequality for conditional expectation, and the martingale property $X_s = E[X_t|\mathcal{F}_s]$, it follows that $E[Y_t|\mathcal{F}_s] = E[\varphi(X_t)|\mathcal{F}_s] \geq \varphi(E[X_t|\mathcal{F}_s]) = \varphi(X_s) = Y_s$. If X is a sub-martingale and φ is convex and increasing, it holds $E[Y_t|\mathcal{F}_s] = E[\varphi(X_t)|\mathcal{F}_s] \geq \varphi(E[X_t|\mathcal{F}_s]) \geq \varphi(X_s) = Y_s$ showing the sub-martingale property and therefore the third point. □

Stochastic integration with respect to a martingale.

Proposition 2.17. *Let H be a predictable process. The following holds true:*

(i) If X is a martingale and $H \bullet X_t$ is integrable for every t, then $H \bullet X$ is a martingale.

(ii) If X is a super/sub-martingale, $H \geq 0$ and $H \bullet X_t$ is integrable for every t, then $H \bullet X$ is a super/sub-martingale.

Proof. Suppose that X is a martingale and H such that $H \bullet X$ is integrable. Adaptiveness is immediate. From H being predictable, that is H_{t+1} is $\mathcal{F}_t$-measurable, and X is a martingale, that is $E[X_{t+1} - X_t|\mathcal{F}_t] = E[X_{t+1}|\mathcal{F}_t] - X_t = 0$, it follows

$$E\left[H \bullet X_{t+1}|\mathcal{F}_t\right] = E\left[H \bullet X_t + H_t(X_{t+1} - X_t)|\mathcal{F}_t\right] = H \bullet X_t + H_{t+1}E\left[X_{t+1} - X_t|\mathcal{F}_t\right] = H \bullet X_t$$

The argumentation in the sub-martingale case is similar, using the fact that $H_{t+1} \geq 0$ and $E[X_{t+1} - X_t|\mathcal{F}_t] = E[X_{t+1}|\mathcal{F}_t] - X_t \geq 0$ to get

$$E\left[H \bullet X_{t+1}|\mathcal{F}_t\right] = H \bullet X_t + H_{t+1}E\left[X_{t+1} - X_t|\mathcal{F}_t\right] \geq H \bullet X_t$$

and similarly for the super-martingale case. □

[36] Why?

Remark 2.18. Note that in Proposition 2.17, if there exists a constant $C > 0$ such that $|H_t| < C$ for every t, then $H \bullet X_t$ is integrable for every t as soon as X is integrable. Indeed,

$$E\left[|H \bullet X_t|\right] \leq E\left[|H_0 X_0|\right] + \sum_{s=1}^{t} E\left[|H_t||X_t - X_{t-1}|\right] \leq 2C \sum_{s=0}^{t} E[|X_t|] < \infty$$

So that the assumption $|H \bullet X_t|$ integrable for every t can be replaced by H uniformly bounded in (i) and (ii) of Proposition 2.17. ♦

This remark allows to formulate the original Doob's sampling's theorem.

Corollary 2.19 (Doob's optional sampling theorem). *Let X be a (super/sub-)martingale and τ a stopping time, then X^τ is a (super/sup-)martingale.*

Proof. Let τ be a stopping time. It holds $X^\tau = H \bullet X$ for the process $H = 1_{\{\cdot \leq \tau\}}$. However, H is predictable, uniformly bounded since $|H_t| \leq 1$ and positive. Hence, according to Proposition 2.17 together with Remark 2.18, it follows that X^τ is a (super/sup-)martingale. □

Proposition 2.20. *If X is a martingale or super-martingale, then $E[X_\tau \,|\, \mathcal{F}_\sigma] = X_\sigma$ or $E[X_\tau \,|\, \mathcal{F}_\sigma] \leq X_\sigma$, respectively, for every pair of bounded stopping times $\sigma \leq \tau$.*

Proof. Since $\tau \leq t$ for some t, it follows that $|X_\tau| \leq |X_0| + \cdots + |X_t|$ and thus X_τ is integrable. In particular, by means of Proposition 2.17 and Lemma 2.12 for the predictable process $H = 1_{\{\cdot \leq \tau\}}$, it follows that X^τ is a martingale. For $A \in \mathcal{F}_\sigma$, it holds $A \cap \{\sigma = s\} \in \mathcal{F}_s$. Hence

$$E\left[(X_t - X_\sigma)1_A\right] = \sum_{s \leq k} E\left[(X_t - X_s)1_{A \cap \{\sigma = s\}}\right] = \sum_{s \leq k} E\left[E\left[X_t - X_s \,|\, \mathcal{F}_s\right] 1_{A \cap \{\sigma = s\}}\right] = 0,$$

showing that $E[X_t \,|\, \mathcal{F}_\sigma] = X_\sigma$. Applying this to the stopped process X^τ yields the result. The proof in the super-martingale case follows the same argumentation. □

Proposition 2.21 (Doob's decomposition). *Let X be an adapted process such that X_t is integrable for every t. Then there exists a unique decomposition*

$$X = M - A$$

where M is a martingale and A is a predictable process with $A_0 = 0$ and A_t integrable for every t. This decomposition is called the Doob de

Proof. Define A by $A_0 := 0$ and $A_t := A_{t-1} - E[X_t - X_{t-1} | \mathcal{F}_{t-1}]$ for every $t \geq 1$. Then A is predictable, satisfies $A_0 = 0$ and A_t is integrable for every t. Further, $M := X + A$ is a martingale. Indeed, M is clearly adapted and M_t is integrable for every t. As for the martingale property it holds

$$E[M_{t+1} - M_t | \mathcal{F}_t] = E[X_{t+1} + A_{t+1} - X_t - A_t | \mathcal{F}_t] = E[X_{t+1} - X_t | \mathcal{F}_t] + A_{t+1} - A_t = 0.$$

The uniqueness follows, since a predictable martingale is constant. □

Proposition 2.22. *Let X be an adapted process such that X_t is integrable for every t with Doob's decomposition $X = M - A$.*

(i) The process X is a super-martingale if and only if A is increasing.

(ii) The process X is a sub-martingale if and only if A is decreasing.

Proof. Let X be a super-martingale, then $E[X_{t+1} - X_t | \mathcal{F}_t] \leq 0$ holds for every t. With the Doob decomposition $X = M - A$ we obtain that $E[M_{t+1} - A_{t+1} - M_t + A_t | \mathcal{F}_n] \leq 0$ holds for every t. Hence $A_t \leq A_{t+1}$ for every t, that is A is increasing. Reading this proof backwards yields the other implication. The sub-martingale case follows the same argumentation. □

2.3. Martingale Convergence

Recall that:

$$\mathbf{T} = \mathbb{N}_0 \quad \text{and} \quad T = \sup \mathbf{T} = \infty$$

Given a martingale X, this section treats the questions whether there exists X_T such that $X_t \to X_T$ and in which sense.

2.3.1. Almost Sure Convergence

The first building block for these questions is the so-called Doob's up-crossing's Lemma. Let X be a process, $x, y \in \mathbb{R}$ with $x < y$, and $F \subseteq \mathbf{T}$ finite. We set

$$\tau_0 = 0$$

and recursively

$$\begin{aligned}
\tau_1 &= \inf\{t \in F : t \geq \tau_0, X_t < x\} \\
\tau_2 &= \inf\{t \in F : t \geq \tau_1, X_t > y\} \\
&\vdots \\
\tau_{2k-1} &= \inf\{t \in F : t \geq \tau_{2k-2}, X_t < x\} \\
\tau_{2k} &= \inf\{t \in F : t \geq \tau_{2k-1}, X_t > y\}
\end{aligned}$$

with the convention that the infimum over the empty set is infinite. We define the random quantity

$$U_F\left(x, y, X(\omega)\right) = \sup\{k : \tau_{2k}(\omega) < \infty\}.$$

This corresponds to the strict positive number of up-crossing of $[x, y]$ by $t \mapsto X_t(\omega)$ on F. For an infinite set $I \subseteq \mathbf{T}$ we set

$$U_I(x, y, X(\omega)) = \sup\{U_F(x, y, X(\omega)) : F \subseteq I, F \text{ finite}\}.$$

Finally, we adopt the notation $[\![s, t]\!] := \{s, s+1, \ldots, t\}$ for every integers $s \leq t$. Doob's upcrossing's lemma reads as follows.

Lemma 2.23. *Let X be a sub-martingale. Then for every two reals $x < y$, the number $U_{[\![0,t]\!]}(x, y, X)$ of up-crossing of $[x, y]$ by $s \mapsto X_s$ up to time $t \in \mathbf{T}$, is a positive random variable and it holds*

$$(y - x)E\left[U_{[\![0,t]\!]}\left(x, y, X\right)\right] \leq E\left[\left(X_t - x\right)^+\right] - E\left[\left(X_0 - x\right)^+\right]. \tag{2.2}$$

Proof. First of all, the random times τ_k, $k = 0, 1, \ldots$ defining the up-crossing function are all stopping times. Since $[\![0, t]\!]$ is a discrete interval here, it follows that $U_{[\![0,t]\!]}(a, b, X)$ is a positive random variable. Define now the predictable gamble strategy, that is, the predictable process

$$H = \sum_{k \geq 1} 1_{]\tau_{2k-1}, \tau_{2k}]},$$

for which holds $H_0 = 0$. It is a predictable since it takes only values 0 and 1 and it holds

$$\{H_t = 1\} = \cup \{\tau_{2k-1} < t\} \cap \{\tau_{2k} < t\}^c \in \mathcal{F}_{t-1}$$

This gamble strategy H is a bet on upcrossings. Note that by the definition of τ_{2k} it follows that for every $\omega \in \Omega$, either $\tau_{2k}(\omega) \leq t$ or $\tau_{2k}(\omega) = \infty$. Further, by the definition of $U := U_{[\![0,t]\!]}\left(x, y, X\right)$ it holds that

$U(\omega) \leq t$, and therefore $\tau_{2U(\omega)} \leq t$ as well as $\tau_{2U(\omega)+2} = \infty$ for every ω. Finally, since U is a random variable, it follows that τ_{2U} is a random time.
We translate our problem at 0 by defining the process $Y = (X - x)^+$. Since $\varphi(z) = (z - x)^+$ is increasing and convex function, it follows from Proposition 2.16 that the process Y is a sub-martingale. It clearly holds that U also counts the number of up-crossings of $[0, y - x]$ up to time t by $s \mapsto Y_s$ and therefore

$$\begin{aligned} H \bullet Y_t(\omega) &= \sum_{s=1}^{t} H_s(\omega) \left(Y_s(\omega) - Y_{s-1}(\omega)\right) \\ &= \sum_{s=1}^{t} \sum_{k\geq 1} 1_{]\tau_{2k-1}(\omega), \tau_{2k}(\omega)]}(s) \left(Y_s(\omega) - Y_{s-1}(\omega)\right) \\ &= \sum_{k\geq 1} \sum_{s=(\tau_{2k-1}(\omega)+1)\wedge t}^{\tau_{2k}(\omega)\wedge t} \left(Y_s(\omega) - Y_{s-1}(\omega)\right) \end{aligned}$$

Two cases may occur:

- If $t = \tau_{2U(\omega)}$, then it holds

$$H \bullet Y_t(\omega) = \sum_{k=1}^{U(\omega)} \left(Y_{\tau_{2k}(\omega)}(\omega) - Y_{\tau_{2k-1}(\omega)}(\omega)\right)$$

- If $t > \tau_{2U(\omega)}$, since $\tau_{2U(\omega)+2} = \infty$ and therefore $\tau_{2U(\omega)+2} \wedge t = t$, then it holds

$$\begin{aligned} H \bullet Y_t(\omega) &= \sum_{k=1}^{U(\omega)} \left(Y_{\tau_{2k}(\omega)}(\omega) - Y_{\tau_{2k-1}(\omega)}(\omega)\right) + \sum_{s=(\tau_{2(U(\omega)+1)-1}(\omega)+1)\wedge t} \left(Y_s(\omega) - Y_{s-1}(\omega)\right) \\ &= \sum_{k=1}^{U(\omega)} \left(Y_{\tau_{2k}(\omega)}(\omega) - Y_{\tau_{2k-1}(\omega)}(\omega)\right) + Y_t(\omega) - Y_{\tau_{2U(\omega)+1}(\omega)\wedge t}(\omega) \end{aligned}$$

So as a random variable, it holds

$$H \bullet Y_t = \sum_{k=1}^{U} \left(Y_{\tau_{2k}} - Y_{\tau_{2k-1}}\right) + (Y_t - Y_{\tau_{2U+1}}))1_{\{t>\tau_{2U}\}}.$$

But if $\tau_{\tau_{2U(\omega)+1}(\omega)} = t$, for the last term it follows that $Y_t(\omega) - Y_{\tau_{2U(\omega)+1}(\omega)\wedge t}(\omega) = Y_t(\omega) - Y_t(\omega) = 0$. Hence, it holds

$$H \bullet Y_t = \sum_{k=1}^{U} \left(Y_{\tau_{2k}} - Y_{\tau_{2k-1}}\right) + (Y_t - Y_{\tau_{2U+1}}))1_{\{t>\tau_{2U+1}\}}.$$

One the one hand, per definition, it holds that $Y_{\tau_{2k-1}} = 0$ for every k, and therefore, from the positivity of Y it follows that

$$E\left[(Y_t - Y_{\tau_{2U+1}})1_{\{t>\tau_{2U}\}}\right] \geq 0.$$

On the other hand, $Y_{\tau_{2k}} - Y_{\tau_{2k-1}} \geq (y - x)$ showing that

$$E[H\bullet Y_t] = E\left[\sum_{k=1}^{U}(Y_{\tau_{2k}} - Y_{\tau_{2k-1}})\right] + E\left[(Y_t - Y_{\tau_{2U+1}})1_{\{t>\tau_{2U+1}\}}\right] \geq E\left[\sum_{k=1}^{U}(y - x)\right] = (y-x)E[U].$$

Defining $K_s = 1 - H_s$ for every $s \geq 1$ and $K_0 = 0$ which is a positive predictable process, hence by means of Proposition 2.17, it follows that $K \bullet Y$ is a submartingale and therefore $E[K \bullet Y_t] \geq E[K \bullet Y_0] = 0$. Since $K + H = 1_{\{1 \leq \cdot\}}$, it follows that

$$\begin{aligned}(y-x)E\left[U\right] \leq E\left[H \bullet Y_t\right] \leq E\left[H \bullet Y_t\right] + E[K \bullet Y_t] = E\left[\sum_{s=1}^{t} Y_s - Y_{s-1}\right] \\ = E\left[Y_t - Y_0\right] = E\left[(X_t - x)^+\right] - E\left[(X_0 - x)^+\right]\end{aligned}$$

which ends the proof. □

Remark 2.24. Given a process X, note that for a given $\omega \in \Omega$, the sample path $t \mapsto X_t(\omega)$ may have asymptotically only four kinds of behavior.

- $\liminf X_t(\omega) = \infty$, that is $X_t(\omega) \to \infty$;
- $\limsup X_t(\omega) = -\infty$, that is $X_t(\omega) \to -\infty$;
- $-\infty < \liminf X_t(\omega) = \limsup X_t(\omega) < \infty$, that is $\lim X_t(\omega)$ exists;
- $\liminf X_t(\omega) < \limsup X_t(\omega)$, also called oscillatory discontinuity.

Saying that the sample path $t \mapsto X_t(\omega)$ is oscillatory discontinuous, is equivalent to the fact there exists rationals q and r with $\liminf X_t(\omega) < q < r < \limsup X_t(\omega)$ such that the number of up-crossing of $[q, r]$ of $X_t(\omega)$ on $\mathbf{T}$ is infinite, that is

$$U_{\mathbf{T}}(q, r, X_t(\omega)) = \sup_{t \in \mathbf{T}} U_{[\![0,t]\!]}(q, r, X(\omega)) = \infty \qquad \blacklozenge$$

Theorem 2.25. *Let X be a sub-martingale such that $\sup E[X_t^+] < \infty$. Then $X_t \to X_T$ almost surely for some integrable random variable X_T.*

Proof. Note that if X is a sub-martingale, then $\sup E[|X|_t] < \infty$ is equivalent to $\sup E[X_t^+] < \infty$. Indeed, it follows from $|X|_t = 2X_t^+ - X_t$ and the sub-martingale property, that $E[X_t] \geq E[X_0] > -\infty$. Let

- A be the set of those $\omega \in \Omega$ such that $t \mapsto X_t(\omega)$ is oscillatory discontinuous, that is, according to Remark 2.24,

$$A = \bigcup_{q<r \text{ and } q,r\in\mathbb{Q}} \{U_{\mathbf{T}}\,(q, r, X) = \infty\} = \bigcup_{q<r \text{ and } q,r\in\mathbb{Q}} \left\{\sup_{t \in \mathbf{T}} U_{[\![0,t]\!]}(q, r, X) = \infty\right\}$$

- B be the set of those $\omega \in \Omega$ such that $t \mapsto X_t(\omega)$ has a real valued limit, that is

$$B = \{\infty < \liminf X_t = \limsup X_t < \infty\}$$

- C be the set of those $\omega \in \Omega$ such that $t \mapsto X_t(\omega)$ diverges to either ∞ or $-\infty$.

In other terms $t \mapsto X_t$ converges to some extended random variable X_T on $B \cup C$. As for A, it is a measurable set as a countable union of measurable sets. Furthermore, by means of Doob's up-crossing's Lemma 2.23, as well as monotone convergence, the assumptions of the theorem yields

$$E\left[\sup_{t \in \mathbf{T}} U_{[\![0,t]\!]}(q, r, X)\right] = \sup_{t \in \mathbf{T}} E\left[U_{[\![0,t]\!]}(q, r, X)\right] \leq \sup_{t} \left\{E\left[(X_t - q)^+\right] - E\left[(X_0 - q)^+\right]\right\} < \infty$$

It follows that $P[\sup_{t\in\mathbf{T}} U_{[\![0,t]\!]}(q,r,X)=\infty]=0$ from which follows

$$P[A] \leq \sum_{q<r \text{ and } q,r\in\mathbb{Q}} P\left[\sup_{t\in\mathbf{T}} U_{[\![0,t]\!]}(q,r,X)=\infty\right]=0.$$

Hence, $P[B\cup C]=1$, showing that $t\mapsto X_t$ converges almost surely to the extended real valued random variable X_T. Finally, by Fatou's Lemma, $E[|X_T|]\leq \liminf E[|X_t|]\leq \sup E[|X|_t]<\infty$ showing integrability of X_T and also that $P[X_T=\infty \text{ or } X_T=-\infty]=P[C]=0$. $\square$

Corollary 2.26. *Let X be a super-martingale such that $\sup_t E[X_t^-]<\infty$. Then $X_t\to X_T$ almost surely for some integrable random variable X_T.*

Proof. The process $Y=-X$ is a submartingale which satisfies $\sup E[Y_t^+]<\infty$ since $Y_t^+=X_t^-$. By Theorem 2.25 there exists Y_T integrable such that $Y_t\to Y_T$ almost surely. Defining $X_T=-Y_T$ yields the result. $\square$

Applications of Almost Convergence Theorem for Martingales

Theorem 2.27. *Let X be a martingale with $X_0=0$. Suppose that $|X_{t+1}-X_t|\leq c$ for every t and some constant $c>0$. Then it holds*

$$P\left[B\cup C\right]=1,$$

where

$$B=\{=\infty<\liminf X_t=\limsup X_t<\infty\} \quad \textit{and} \quad C=\{\liminf X_t=-\infty \textit{ and } \limsup X_t=\infty\}.$$

Proof. Define the stopping time $\tau_k=\inf\{t\colon X_t>k\}$. According to Doob's sampling theorem, Corollary 2.19, it follows that X^{τ_k} is a martingale such that $\sup_t E[(X_t^{\tau_k})^+]\leq k+c<\infty$. Indeed, on $\{t<\tau_k\}$, it holds $X_t^{\tau_k}\leq k$ and on $\{\tau_k\leq t\}$, it holds $X_t^{\tau_k}=X_{\tau_k}\leq X_{\tau_k-1}+(X_{\tau_k}-X_{\tau_k-1})\leq k+c$. By Theorem 2.25, $\lim_{t\to\infty} X_t^{\tau_k}$ exists almost surely. On $\{\tau_k=\infty\}$ the processes X and X^{τ_k} coincide, so that $\lim X_t$ exists almost surely on $\{\tau_k=\infty\}$. In particular $\lim X_t$ exists almost surely on

$$\bigcup\{\tau_k=\infty\}=\{\limsup X_t<\infty\}.$$

A similar argumentation for $-X$ shows that $\lim X_t$ exists almost surely on $\{\liminf X_t>-\infty\}$. That is $\lim X_t$ exists almost surely on $\{\liminf X_t>-\infty\}\cup\{\limsup X_t<\infty\}=C^c$. It means that $P[C^c\setminus B]=P[C^c\cap B^c]=0$. Hence, taking complementation, it follows that $P[B\cup C]=P[(C^c\cap B^c)]=1$ which ends the proof. $\square$

Corollary 2.28. *We suppose that $\mathcal{F}_0=\{\emptyset,\Omega\}$. Let (A_t) be a sequence of elements in $\mathcal{F}$ such that $A_t\in\mathcal{F}_t$ for every t. Then*

$$\limsup A_t=\bigcap_t\bigcup_{s\geq t} A_s=\{\omega:\omega\in A_t \textit{ for infinitely many } t\}=\left\{\sum P(A_t|\mathcal{F}_{t-1})=\infty\right\}$$

holds almost surely, whereby $P[A_t|\mathcal{F}_{t-1}]=E[1_{A_t}|\mathcal{F}_{t-1}]$.

Proof. We define the process X as follows

$$X_0=0 \quad \text{and } X_t=\sum_{s=1}^t 1_{A_s}-P\left[A_s|\mathcal{F}_{s-1}\right], \text{ for } t\geq 1$$

Since $\mathcal{F}_0 = \{\emptyset, \Omega\}$, it follows that X is a martingale. Indeed, X is clearly adapted by definition, and $|X_t| \leq 2t$ so that X is integrable. Furthermore, $E[X_1 - X_0|\mathcal{F}_0] = E[X_1 - X_0] = P[A_1] - P[A_1] = 0$ and $E[X_t - X_{t-1}|\mathcal{F}_{t-1}] = E[1_{A_t} - E[1_{A_t}|\mathcal{F}_{t-1}]|\mathcal{F}_{t-1}] = E[E[1_{A_t} - 1_{A_t}|\mathcal{F}_{t-1}|\mathcal{F}_{t-1}] = 0$ for every $t \geq 2$. Since $|X_{t+1} - X_t| \leq 2$ holds for every t, we may apply Theorem 2.27. On $B = \{\liminf X_t = \limsup X_t \in \mathbb{R}\}$, it holds

$$\sum 1_{A_t} = \infty \quad \text{if, and only if,} \quad \sum P\left[A_n|\mathcal{F}_{t-1}\right] = \infty.$$

On $C = \{\liminf X_t = -\infty \text{ and } \limsup X_t = \infty\}$ it holds

$$\sum 1_{A_t} = \infty \quad \text{and} \quad \sum P\left[A_t|\mathcal{F}_{t-1}\right] = \infty.$$

Since $P[B \cup C] = 1$ we deduce

$$\sum 1_{A_t} = \infty \quad \text{if, and only if,} \quad \sum P[A_t|\mathcal{F}_{t-1}] = \infty$$

almost surely. Moreover, $\limsup A_t = \{\sum 1_{A_t} = \infty\}$, hence the claim follows. □

Corollary 2.29 (Borel-Cantelli). *Let (A_t) be a sequence of elements in $\mathcal{F}$.*

(i) If $\sum P[A_t] < \infty$, then it holds $P[\limsup A_t] = 0$.

(ii) If (A_t) is an independent sequence and $\sum P(A_t) = \infty$, then it holds $P[\limsup A_t] = 1$.

Proof. We consider the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbf{T}}$ given by $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t := \sigma(A_s\colon s \leq t)$ for $t \geq 1$. Define $\xi := \sum P[A_t|\mathcal{F}_{t-1}]$. The monotone convergence theorem as well as the tower property shows that

$$E[\xi] = E\left[\sum E[1_{A_t}|\mathcal{F}_{t-1}]\right] = \sum E[E[1_{A_t}|\mathcal{F}_{t-1}]] = \sum P[A_t].$$

(i) If $\sum P[A_t] < \infty$, then it holds $P[\xi = \infty] = 0$. Corollary 2.28 yields $P[\limsup A_t] = 0$.

(ii) Suppose that (A_t) is an independent sequence, therefore A_t is independent of $\mathcal{F}_{t-1}$ which implies $P[A_t|\mathcal{F}_{t-1}] = P[A_t]$ for all t. Hence $\sum P[A_t|\mathcal{F}_{t-1}] = \sum P[A_t] = \infty$ almost surely and by Corollary 2.28 it follows that $P[\limsup A_t] = 1$. □

2.3.2. L^p-Convergence

The building block for L^p convergence are the so called Doob's maximal inequalities. In the following, given a process X we define the

- running supremum process $\bar{X}$ by $\bar{X}_t = \sup_{s\leq t} X_s$;
- running infimum process $\underline{X}$ by $\bar{X}_t = \inf_{s\leq t} X_s$;
- running absolute supremum process X^* by $X^*_t = \sup_{s\leq t} |X_s|$

Proposition 2.30. *The following assertions hold true.*

(a) Let X be a sub-martingale and $\lambda > 0$. Then it holds

$$\begin{aligned}\lambda P\left[\bar{X}_t \geq \lambda\right] &\leq E\left[1_{\{\bar{X}_t\geq\lambda\}} X_t\right] \leq E\left[X_t^+\right];\\ \lambda P\left[\underline{X}_t \leq -\lambda\right] &\leq E\left[1_{\{\underline{X}_t>-\lambda\}} X_t\right] - E[X_0] \leq E\left[X_t^+\right] - E\left[X_0\right]\end{aligned}$$

(b) For X be a positive sub-martingale and $p>1$, it holds

$$\left\|\sup_{s\leq t} X_s\right\|_p \leq \frac{p}{p-1}\left\|X_t\right\|_p .$$

Proof. (a) For the stopping time $\tau=\inf\{s: X_s\geq \lambda\}$, observe that $\{\tau\leq t\}=\{\bar{X}_t\geq\lambda\}$. Also, on $\{\tau\leq t\}$, it holds $X_t^\tau=X_{\tau\wedge t}\geq\lambda$. Hence, $X_{\tau\wedge t}=X_\tau 1_{\{\bar{X}_t\geq\lambda\}}+X_t 1_{\{\tau>t\}}\geq \lambda 1_{\{\bar{X}_t\geq\lambda\}}+X_t 1_{\{\tau>t\}}$. It also holds, $X_t 1_{\{\tau>t\}}\geq -X_t^-$. All together, with Doob's optional sampling theorem, and X being a sub-martingale, we get

$$E\left[X_t\right]\geq E\left[X_{\tau\wedge t}\right]\geq \lambda P\left[\bar{X}_t\geq\lambda\right]+E\left[1_{\{\tau>t\}}X_t\right]\geq \lambda P\left[\bar{X}_t\geq\lambda\right]-E\left[X_t^-\right],$$

and conclude the first inequality by observing that $E[X_t^+]=E[X_t]+E[X_t^-]$ and $E[X_t]-E\left[1_{\{\tau>t\}}X_t\right]=E[(1-1_{\{\bar{X}_t<\lambda\}})X_t]=E[1_{\{\bar{X}_t\geq\lambda\}}X_t]$.
As for the second inequality, for the stopping time $\sigma=\inf\{s: X_s\leq-\lambda\}$, observe that $\{\sigma\leq t\}=\{\underline{X}_t\leq-\lambda\}$. Also, on $\{\sigma\leq t\}$, it holds $X_t^\sigma=X_{\sigma\wedge t}\leq-\lambda$. Hence, $X_{\sigma\wedge t}=X_\sigma 1_{\{\underline{X}_t\leq-\lambda\}}+X_t 1_{\{\sigma>t\}}\leq -\lambda 1_{\{\underline{X}_t\leq-\lambda\}}+X_t 1_{\{\sigma>t\}}$. All together, with Doob's optional sampling theorem, and X being a sub-martingale, we get

$$E\left[X_0\right]\leq E\left[X_{\sigma\wedge t}\right]\leq -\lambda P\left[\underline{X}_t\leq-\lambda\right]+E\left[1_{\{\sigma>t\}}X_t\right]\leq -\lambda P\left[\underline{X}_t\leq-\lambda\right]+E\left[X_t^+\right]$$

showing the second inequality by observing that $E[1_{\{\sigma>t\}}X_t]=E[1_{\{\underline{X}_t>-\lambda\}}X_t]$.

(b) Define the random variables $Y=\sup_{s\leq t}X_s$ and $Z=X_t=X_t^+$ since X is positive. For φ an increasing, right-continuous function with $\varphi(0)=0$, by Fubini's theorem and the previous inequalities, it holds

$$\begin{aligned}E\left[\varphi(Y)\right]=E\left[\int_0^\infty 1_{\{\lambda\leq Y\}}d\varphi(\lambda)\right]&=\int_0^\infty P\left[Y\geq\lambda\right]d\varphi(\lambda)\\ &\leq\int_0^\infty E\left[1_{\{Y\geq\lambda\}}Z\right]\frac{d\varphi(\lambda)}{\lambda}=E\left[Z\int_0^\infty 1_{\{Y\geq\lambda\}}\frac{d\varphi(\lambda)}{\lambda}\right].\end{aligned}$$

If we consider $\varphi(\lambda)=\lambda^p$, $p>1$, and define $q=p/(p-1)$ for which holds $1/p+1/q=1$, it follows from Hölder's inequality that

$$\left\|Y\right\|_p^p\leq pE\left[Z\int_0^\infty 1_{\{Y\geq\lambda\}}\lambda^{p-2}d\lambda\right]=\frac{p}{p-1}E\left[ZY^{p-1}\right]\leq q\left\|Z\right\|_p\left\|Y^{p-1}\right\|_q=q\left\|Z\right\|_p\left\|Y\right\|_p^{p/q}.$$

If $0<\left\|Y\right\|_p^{p/q}<\infty$, dividing the inequality by $\left\|Y\right\|_p^{p/q}$, noting that $p-p/q=1$, yields

$$\left\|\sup_{s\leq t}X_s\right\|_p=\left\|Y\right\|_p\leq q\left\|Z\right\|_p=q\left\|X_t\right\|_p,$$

as desired. If $\left\|Y\right\|_p^{p/q}=0$ the inequality is trivial. If $\left\|Y\right\|_p^{p/q}=\infty$, stop X at $\tau^n=\inf\{t: X_t\geq n\}$ for every n, use the inequality for X^{τ_n}, which is still a positive a sub-martingale, and then pass to the limit since $\lim\tau^n\geq t$ almost surely. □

Remark 2.31. In particular, if X is a martingale, and $p>1$, then by Proposition 2.16, $|X|^p$ is a positive sub-martingale, and so

$$\left\|X_t^*\right\|_p=E\left[\left(\sup_{s\leq t}|X_s|\right)^p\right]^{1/p}=E\left[\sup_{s\leq t}|X_s|^p\right]^{1/p}\leq\left(\frac{p}{p-1}\right)\left\|X_t\right\|^p \tag{2.3}$$

for every $p>1$. ♦

Theorem 2.32 (Martingale convergence theorem). *Let X be a martingale such that $\sup_t E[|X_t|^p] < \infty$ for some $p > 1$. Then, there exists a random variable $X_T \in L^p$ such that $X_t \to X_T$ almost surely and in L^p.*

Proof. Since Jensen's inequality yields $E[X_t^+] \le E[|X_t|] \le E[|X_t|^p]^{\frac{1}{p}}$, it follows that $\sup E[X_t^+] < \infty$. By the martingale convergence Theorem 2.25, there exists an integrable random variable X_T for which $X_t \to X_T$ almost surely. We are left to show that the sequence $|X_t - X_T|^p$ satisfies the assumptions of Lebesgue's dominated convergence. It holds

$$|X_t - X_T|^p \le c\left(|X_t|^p + |X_T|^p\right) \le c\left(\sup |X_t|^p + |X_T|^p\right).$$

On the one hand, by Fatou's lemma we have $E[|X_T|^p] \le \liminf E[|X_t|^p] < \infty$. On the other hand, by means of Remark 2.31, it holds $E[\sup_{s\le t} |X_s|^p] \le (p/(p-1))^p E[|X_t|^p]$ showing that $E[\sup |X_t|^p] = \sup_t E[\sup_{s\le t} |X_s|^p] \le (p/(p-1))^p \sup E[|X_t|^p] < \infty$. Thus, the dominated convergence theorem yields $X_t \to X_T$ in L^p. □

Application to the Law of Large Numbers We apply the L^p-convergence of martingales to show the law of large numbers that states that the sample average of independently distributed random variables with finite mean converges almost surely to its mean.

Theorem 2.33. *Let X be a square integrable martingale for which holds*

$$\sum E\left[\left(X_t - X_{t-1}\right)^2\right] < \infty.$$

Then, the sequence (X_t) converges almost surely and in L^2.

Beforehand, let us show the following lemma.

Lemma 2.34. *Let X be a martingale such that X_t is square integrable for every t. It follows that*

$$\begin{aligned} E\left[\left(X_u - X_t\right) X_s\right] &= 0 \\ E\left[\left(X_t - X_s\right)^2 |\mathcal{F}_s\right] &= E\left[X_t^2|\mathcal{F}_s\right] - X_s^2 \end{aligned}$$

for every $s \le t \le u$.

Proof. Since $s \le t \le u$ and X is a square integrable martingale, it follows from the properties of the conditional expectation

$$E\left[\left(X_u - X_t\right) X_s\right] = E\left[E\left[\left(X_u - X_t\right) X_s|\mathcal{F}_t\right]\right] = E\left[E\left[\left(X_u - X_t\right)|\mathcal{F}_t\right] X_s\right] = 0$$

showing the first equality. The same reasons yield

$$\begin{aligned} E\left[\left(X_t - X_s\right)^2 |\mathcal{F}_s\right] &= E\left[X_t^2|\mathcal{F}_s\right] - E\left[X_t X_s|\mathcal{F}_s\right] - E\left[(X_t - X_s)X_s|\mathcal{F}_s\right] \\ &= E\left[X_t^2|\mathcal{F}_s\right] - X_s E\left[X_t|\mathcal{F}_s\right] - X_s E\left[X_t - X_s|\mathcal{F}_s\right] = E\left[X_t^2|\mathcal{F}_s\right] - X_s^2, \end{aligned}$$

showing the second equality. □

Proof (of Theorem 2.33). For every t, by means of Lemma 2.34, it follows that

$$\begin{aligned} E\left[X_t^2\right] &= E\left[X_0^2\right] + \sum_{s=1}^{t} E\left[X_s^2 - X_{s-1}^2\right] = E\left[X_0^2\right] + \sum_{s=1}^{t} E\left[E\left[X_s^2 - X_{s-1}^2 | \mathcal{F}_{s-1}\right]\right] \\ &= E\left[X_0^2\right] + \sum_{s=1}^{t} E\left[E\left[(X_s - X_{s-1})^2 | \mathcal{F}_{s-1}\right]\right] = E\left[X_0^2\right] + \sum_{s=1}^{t} E\left[(X_s - X_{s-1})^2\right] \\ &\leq E\left[X_0^2\right] + \sum E\left[(X_s - X_{s-1})^2\right] \end{aligned}$$

It follows that $\sup_t E\left[X_t^2\right] < \infty$ and therefore, by means of Theorem 2.32 it follows that $X_t \to X_T$ almost surely and in L^2. □

Theorem 2.35. *Let X be a martingale and $a = (a_t)$ be an increasing sequence such that $a_t \to \infty$. If $\sum E[(X_t - X_{t-1})^2 / a_t^2] < \infty$, then it follows that*

$$\frac{X_t}{a_t} \longrightarrow 0$$

almost surely. In particular, if $\sup E[(X_t - X_{t-1})^2] < \infty$, then it holds

$$\frac{X_t}{t} \longrightarrow 0$$

almost surely.

Proof. Define the process Y by $Y_0 = 0$ and $Y_t = \sum_{s=1}^{t} (X_s - X_{s-1})/a_s$ for $t \geq 1$. It follows that Y is a martingale. Indeed, adaptiveness and integrability are immediate since X is a martingale. As for the martingale property, it holds

$$E\left[Y_t - Y_{t-1} | \mathcal{F}_t\right] = \frac{1}{a_t} E\left[X_t - X_{t-1} | \mathcal{F}_t\right] = 0.$$

Furthermore, it holds

$$\sum E\left[(Y_t - Y_{t-1})^2\right] = \sum \frac{1}{a_t^2} E\left[(X_t - X_{t-1})^2\right] < \infty$$

which by means of Theorem 2.33 implies that

$$Y_t = \sum_{s=1}^{t} \frac{X_s - X_{s-1}}{a_s} \longrightarrow Y_T = \sum \frac{X_t - X_{t-1}}{a_t}$$

almost surely and in L^2. The Kornecker's lemma states that if $\sum b_t/a_t < \infty$ for two sequences (a_t) and (b_t) whereby (a_t) is an increasing sequence of strictly positive numbers, it follows that $(\sum b_t)/a_t = 0$.[37] Hence, applying Kronecker's lemma, it follows that

$$\frac{X_t}{a_t} = \frac{1}{a_t} \sum_{s=1}^{t} X_s - X_{s-1} \to 0$$

almost surely. In particular, if $\sup E[(X_t - X_{t-1})^2] < \infty$ it follows that $\sum E[(X_t - X_{t-1})^2 / t^2] \leq \sup E[(X_t - X_{t-1})^2] \sum 1/t^2 < \infty$ and the second assertion of the Theorem follows. □

[37] See in exercise.

Corollary 2.36. *Let (X_t) be a sequence of integrable independent random variables such that $E[X_t] = 0$ for every t and such that $\sum E[X_t^2]/a_t^2 < \infty$ for some increasing sequence (a_t) of strictly positive real numbers such that $a_t \to \infty$. Then it holds*

$$\frac{1}{a_t}\sum_{s=1}^{t} X_s \longrightarrow 0$$

almost surely.

Proof. Define $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t = \sigma(X_s \colon s \leq t)$ and the process S by $S_0 = 0$ and $S_t = \sum_{s=1}^{t} X_s$. It follows that S is a martingale. Indeed, it is integrable by assumption. It is furthermore adapted since $\mathbb{F}$ is the filtration generated by X. Finally due to the independence, it follows that

$$E\left[S_t - S_{t-1} | \mathcal{F}_{t-1}\right] = E\left[X_t | \mathcal{F}_{t-1}\right] = E\left[X_t\right] = 0.$$

Furthermore, since

$$\sum \frac{1}{a_t^2} E\left[(S_t - S_{t-1})^2\right] = \sum \frac{1}{a_t^2} E\left[X_t^2\right] < \infty,$$

we can apply Theorem 2.35 to get

$$\frac{S_t}{a_t} = \frac{1}{a_t}\sum_{s=1}^{t} X_s \longrightarrow 0$$

almost surely. □

Theorem 2.37 (Strong Law of Large Numbers). *Let (X_t) be a sequence of integrable, independent, and identically distributed random variables. Then it holds*

$$\frac{1}{t}\sum_{s=1}^{t} X_s \to E\left[X_1\right]$$

almost surely.

Proof. Step 1: Define first the countable family (A_t) as $A_t = \{|X_t| > t\}$ of elements in $\mathcal{F}$. Using the fact that $X_t \sim X_1$ for every t and Fubini's Theorem, it holds

$$\sum P\left[A_t\right] = \sum P\left[|X_t| > t\right] = \sum P\left[|X_1| > t\right] \leq \int_0^\infty P\left[|X_1| > \lambda\right] d\lambda = E\left[|X_1|\right] < \infty$$

By Borel-Cantelli, it follows that $P[\limsup A_t] = 0$. Defining $Y_t = X_t 1_{A_t^c}$, it follows that for almost all $\omega \in \Omega$, there exists $t_0(\omega)$ such that that for every $t \geq t_0(\omega)$ it holds $Y_t(\omega) = X_t(\omega)$. Hence

$$\liminf \frac{1}{t}\sum_{s \leq t} X_s = \liminf \frac{1}{t}\sum_{s\leq t} Y_s \quad \text{as well as} \quad \limsup \frac{1}{t}\sum_{s\leq t} X_s = \limsup \frac{1}{t}\sum_{s\leq t} Y_s,$$

and so we just have to show that

$$\frac{1}{t}\sum_{s=1}^{t} Y_s \to E\left[Y_1\right] = E\left[X_1\right]$$

almost surely since $Y_1 = X_1$.

Step 2: Let $Z_t = Y_t - E[Y_t]$ for every t which is an independent sequence of random variables due to that fact that (X_t) and therefore (Y_t) is an independent sequence of random variables. Furthermore, note that

$$\sum \frac{E[Z_t^2]}{t^2} = \sum \frac{E\left[(Y_t - E[Y_t])^2\right]}{t^2} = \sum \frac{E\left[Y_t^2\right] - E\left[Y_t\right]^2}{t^2} \leq \sum \frac{E[Y_t^2]}{t^2}.$$

By Fubini's theorem, and the fact that $P[|Y_t| > s] = P[|Y_t| > t] = 0$ for every $s \geq t$ as well as $P[Y_t > \lambda] \leq P[X_t > \lambda] = P[X_1 > \lambda]$ for every t, it holds

$$E\left[Y_t^2\right] = E\left[\int_0^\infty 1_{\{|Y_t| > \lambda\}} 2\lambda d\lambda\right] = \int_0^\infty P\left[|Y_t| > \lambda\right] 2\lambda d\lambda \leq \int_0^t P\left[|X_1| > \lambda\right] 2\lambda d\lambda$$

The monotone convergence of Lebesgue yields

$$\sum \frac{1}{t^2} \int_0^t P\left[|X_1| > \lambda\right] 2\lambda d\lambda = \int_0^\infty \sum \frac{1_{\{t \geq \lambda\}}}{t^2} P\left[|X_1| > \lambda\right] 2\lambda d\lambda.$$

For $\lambda < 1$, it holds

$$2\lambda \sum \frac{1}{t^2} = 2\lambda \frac{\pi^2}{6} \leq 4\lambda \leq 4,$$

and for $\lambda \geq 1$, it holds

$$2\lambda \sum_{t \geq \lambda} \frac{1}{t^2} \leq \frac{2}{\lambda} + 2\lambda \int_\lambda^\infty \frac{1}{x^2} dx \leq 2 + 2\lambda \frac{1}{\lambda} = 4.$$

Hence

$$\sum \frac{E\left[Z_t^2\right]}{t^2} \leq 4 \int_0^\infty P\left[|X_1| > \lambda\right] d\lambda = 4E\left[|X_1|\right] < \infty.$$

According to 2.36, it follows that $(\sum_{s \leq t} Z_s)/t \to 0$ almost surely.

Step 3: We finally show that $(\sum_{s \leq t} Y_s)/t \to E[X_1]$ almost surely. It holds

$$\begin{aligned}
\left|\frac{1}{t}\sum_{s \leq t} Y_s - E[X_1]\right| &\leq \left|\frac{1}{t}\sum_{s \leq t} Z_s\right| + \frac{1}{t}\sum_{s \leq t} \left|E\left[Y_s\right] - E\left[X_1\right]\right| \\
&\leq \left|\frac{1}{t}\sum_{s \leq t} Z_s\right| + \frac{1}{t}\sum_{s \leq t} \left|E\left[X_s\right] - E\left[X_1\right]\right| \leq \left|\frac{1}{t}\sum_{s \leq t} Z_s\right| + \frac{1}{t}\sum_{s \leq t} \left|E\left[|X_1| 1_{|X_1| > s}\right]\right| \\
&= \left|\frac{1}{t}\sum_{s \leq t} Z_s\right| + E\left[|X_1| \frac{1}{t}\sum_{s \leq t} 1_{\{|X_1| > s\}}\right] = I_t + E[J_t].
\end{aligned}$$

We already shown in the previous step that $I_t \to 0$ almost surely. On the other hand, it holds $J_t \leq |X_1|$ with $E[|X_1|] < \infty$ and since $(\sum_{s \leq t} 1_{\{|X_1| > s\}})/t \to 0$ almost surely, it follows that $J_t \to 0$ almost surely. Hence, by Lebesgue's dominated convergence theorem, it follows that $E[J_t] \to 0$ which ends the proof. $\square$