Proof. For X in L^1, it defines two finite measures on $\mathcal{G}$ given by

$$Q^{\pm}(A) = E\left[1_A X^{\pm}\right], \quad A \in \mathcal{G}$$

which are by definition both absolutely continuous with respect to P.[30] It follows from Radon-Nikodym Theorem 1.55 that there exists two P-almost surely unique positive $\mathcal{G}$-measurable random variables $Z^{\pm} \in L^1(\mathcal{G})$ such that

$$Q^{\pm}(A) = E[1_A Z^{\pm}]$$

Defining $E\left[X|\mathcal{G}\right] = Z^+ - Z^- \in L^1(\mathcal{G})$ as the conditional expectation end the proof of the existence and uniqueness. The properties *(i)*–*(vii)* are left as an exercise, where the monotone or dominated convergence of Lebesgue as to be used for some. ☐

Exercice 1.57. Under the assumptions of the Theorem 1.47, show that for a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, if $\varphi(X)$ is integrable, then it holds

$$\varphi\left(E\left[X|\mathcal{G}\right]\right) \leq E\left[\varphi\left(X\right)|\mathcal{G}\right]$$ ♢

For your interest, here is the proof of the existence of conditional expectation using Hilbert projections.

Proof. Suppose first that $X \in L^2(\mathcal{F})$. Note that $L^2(\mathcal{F})$ is an Hilbert space for the norm $\|\cdot\|_2$ and $L^2(\mathcal{G})$ is a closed linear subspace of $L^2(\mathcal{F})$. Hence, by Hilbert's projection theorem, there exists a unique $Y \in L^2(\mathcal{G})$ such that $X - Y$ is orthogonal to $L^2(\mathcal{G})$. Since $1_A \in L^2(\mathcal{G})$ for every $A \in \mathcal{G}$ it follows that

$$E\left[(X-Y)1_A\right] = \langle X - Y, 1_A\rangle = 0, \quad A \in \mathcal{G}$$

showing the main assertion. The properties *(ii)*–*(vii)* are easy to verify in L^2 from the definition and therefore left as an exercise.
We show property *(i)*. For $X \in L^2$, let $A = \{E[X|\mathcal{F}] \geq 0\}$ which is an event in $\mathcal{G}$, it follows that

$$E\left[\left|E\left[X|\mathcal{F}\right]\right|\right] = E\left[E\left[X|\mathcal{F}\right]; A\right] - E\left[E\left[X|\mathcal{F}\right]; A^c\right] = E\left[X; A\right] - E\left[X; A^c\right] \leq E\left[|X|\right]$$

Hence

$$\sup\left\{E\left[\left|E\left[X|\mathcal{G}\right]\right|\right] : X \in L^2, \|X\|_1 = E\left[|X|\right] \leq 1\right\} \leq 1$$ ☐

showing that the linear functional $E[\cdot|\mathcal{F}]$ on L^2 is L^1-continuous. Since L^2 is dense in L^1 which is complete, it follows that this linear extension extends uniquely to a continuous one on L^1, and the properties *(i)*–*(vii)* extends as well to L^1 which ends the proof.

1.5. Uniform Integrability

Throughout the script, we may use the following notation

$$E[X; A] := E[1_A X] \qquad \text{as well as } P[A; B] := P[A \cap B]$$

We finish this subsection with some results about uniform integrability. Note that for $X \in L^1$, Lebesgues dominated convergence implies that $E[|X|; |X| \geq n] \to 0$. Uniform integrability is a similar requirement but on a whole set of random variables.

Definition 1.58. A set $H \subseteq L^1$ is called uniformly integrable if

$$\sup_{X \in H} E\left[|X|; |X| \geq n\right] \to 0$$

[30] Verify that these are indeed measures!

Proposition 1.59. *For $H \subseteq L^1$, the following assertions are equivalent*

(i) *H is uniformly integrable;*

(ii) *the following two assertions holds*

- *H is bounded in L^1, that is $\sup_{X \in H} E[|X|] < \infty$;*
- *For every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$E\left[|X|\,;A\right] \leq \varepsilon$$

for all $X \in H$ and $A \in \mathcal{F}$ such that $P[A] \leq \delta$.

(iii) *There exists a Borel measurable function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi(x)/x \to \infty$ as $x \to \infty$ for which holds*

$$\sup_{X \in H} E\left[\varphi(|X|)\right] < \infty.$$

Proof. Suppose that *(i)* holds. It follows that for n large enough we have $E[|X|\,;|X| \geq n] \leq 1$ for all $X \in H$. Hence $E[|X|] \leq n+1$ for all $X \in H$ showing that H is bounded in L^1. Let further $\varepsilon > 0$ and choose n large enough such that $E[|X|\,;|X| \geq n] \leq \varepsilon/2$. Seting $\delta = \varepsilon/(2n)$, for every $A \in \mathcal{F}$ such that $P[A] \leq \delta$, it follows that

$$E\left[|X|\,;A\right] = E\left[|X|\,;A \cap \{|X| \geq n\}\right] + E\left[|X|\,;A \cap \{|X| < n\}\right] \leq nP[A] + \varepsilon/2 \leq \varepsilon,$$

showing that *(i)* implies *(ii)*.
Reciprocally, suppose that *(i)* holds. Denote by $M = \sup_X E[|X|] < \infty$, and let $\varepsilon > 0$. There exists $\delta > 0$ such that $E[X; A] \leq \varepsilon$ for every $A \in \mathcal{F}$ with $P[A] \leq \delta$. Choose then n greater than M/δ. For $X \in H$, Markov inequality yields

$$P\left[|X| \geq n\right] \leq \frac{E[|X|]}{n} \leq \frac{M}{n} \leq \delta$$

Hence

$$\sup_{X \in H} E\left[|X|\,;|X| \geq n\right] \leq \varepsilon$$

showing the uniform integrability of H.
Suppose that *(iii)* holds and denote by $M = \sup_{X \in H} E[\varphi(X)]$. For $\varepsilon > 0$, there exists n_ε such that $\varphi(x) \geq Mx/\varepsilon$ for every $x \geq n_\varepsilon$. Hence

$$M \geq \sup_{X \in H} E\left[\varphi(|X|)\right] \geq \sup_{X \in H} E\left[\varphi(|X|); |X| \geq n_\varepsilon\right] \geq M \sup_{X \in H} E\left[|X|\,;|X| \geq n_\varepsilon\right]/\varepsilon$$

showing that $\sup_n \sup_{X \in H} E\left[|X|\,;|X| \geq n\right] \leq \sup_{X \in H} E\left[|X|\,;|X| \geq n_\varepsilon\right] \leq \varepsilon$ and so the uniform integrability of H.
Reciprocally assume *(i)* and choose a sequence (c_n) which can always be chosen increasing, such that $\sup_{X \in H} E[|X|\,;|X| \geq c_n] \leq 1/n^3$. Define the function $\varphi : \mathbb{R}_+$ as a piecewise linear, equal to 0 on $[0, c_1]$ and the derivative equal to n on $[c_n, c_{n+1}]$ which implies that $\varphi(x)/x \to \infty$ as $x \to \infty$. It follows that

$$E[\varphi(|X|)] = \sum E[\varphi(|X|); c_n \leq |X| \leq c_{n+1}] = \sum n\left(E[|X| \wedge c_{n+1}] - E[|X| \wedge c_n]\right)$$

However,

$$\begin{aligned} &E[|X| \wedge c_{n+1}] - E[|X| \wedge c_n] \\ &\qquad = E[|X|\,; c_n \leq |X| < c_{n+1}] + E[c_{n+1}; |X| \geq c_{n+1}] - E[c_n; |X| \geq c_n] \\ &\qquad\qquad \leq E[|X|\,;|X| \geq c_n] + E[|X|\,;|X| \geq c_{n+1}] \leq 2/n^3 \end{aligned}$$

which shows that $\sup_{X \in H} E[\varphi(|X|)] \leq \sum 2n/n^3 < \infty$. $\square$

Theorem 1.60. *Let $(X_n) \subseteq L^1$ be a sequence of random variables such that X_n converges in probability to a random variable X.[31] Then, the following assertions are equivalent*

(i) the sequence is uniformly integrable;[32]

(ii) X_n converges to X in L^1.

(iii) $\|X_n\|_1$ converges to $\|X\|_1$

Proof. We show that (i) implies (ii). By Proposition 1.52, there exists a subsequence (Y_n) of (X_n) that converges P-almost surely to X. In particular, (Y_n) is uniformly integrable. Using Fatou and the L^1 boundedness of the family (X_n), see Proposition 1.59, it follows that $E[|Y|] \leq \liminf E[|Y|_n] \leq \sup_n E[|Y|_n] < \infty$ showing that $X \in L^1$. It follows that the sequence $(X_n - X)$ is uniformly integrable and therefore without loss of generality we can assume that (X_n) is a uniform integrable family converging in probability to 0. For $\varepsilon > 0$ it holds

$$E\left[|X_n|\right] = E\left[|X_n|\,; |X_n| \leq \varepsilon/2\right] + E\left[|X_n|\,; |X_n| > \varepsilon/2\right] \leq \varepsilon/2 + E\left[|X_n|\,; |X_n| > \varepsilon/2\right]$$

By uniform integrability of the family (X_n), making use of Proposition 1.59, let $\delta > 0$ such that $\sup_n E[|X_n|\,; A] \leq \varepsilon/2$ for every $A \in \mathcal{F}$ with $P[A] \leq \delta$. Further, by convergence of (X_n) in probability to 0, there exists n_0 such that $P[|X_n| > \varepsilon/2] \leq \delta$ for every $n \geq n_0$. Thus, for every $n \geq n_0$, it holds $E[|X_n|] \leq \varepsilon/2 + \sup_{k \geq n_0} E[|X_n|\,; |X_n| > \varepsilon/2] \leq \varepsilon$ showing that X_n converges in L^1 to 0.
The fact that (ii) implies (iii) is trivial from $||x| - |y|| \leq |x - y|$, and therefore we finish the proof by showing that (iii) implies (i). For $M > 0$, define φ_M as being the identity on $[0, M-1]$, 0 on $[M, \infty[$ and linearly interpolated on the remaining part of the real line. Let $\varepsilon > 0$ and using the dominated convergence theorem, choose M such that $E[|X|] - E[\varphi_M(|X|)] \leq \varepsilon/2$ since $\varphi_M(|X|)$ converges to and is dominated by $|X| \in L^1$. By continuity of φ_M, it follows that $\varphi_M(|X_n|) \to \varphi_M(|X|)$ also in probability. Now, since $\varphi_M(|X_n|) \leq M$ for every n, the dominated convergence theorem in its convergence in probability fashion, see Proposition 1.52 yields $E[\varphi(|X_n|)] \to E[\varphi_M(|X|)]$. Hence, together with $E[|X_n|] \to E[|X|]$, there exists some integer n_0 such that

$$E[|X_n|] - E[|X|] \leq \varepsilon/4 \quad \text{and} \quad E[\varphi_M(|X|)] - E[\varphi(|X_n|)] \leq \varepsilon/4$$

for every $n \geq \varepsilon$. Henceforth

$$E[|X_n|\,; |X_n| \geq M] \leq E[|X_n|] - E[\varphi_M(|X_n|)] \leq \varepsilon/2 + E[|X|] - E[\varphi_M(|X|)] \leq \varepsilon$$

for every $n \geq n_0$. Increases the value of M so that this inequality remains true for the remaining $n \geq n_0$, to conclude the uniform integrability of (X_n). □

[31] That is $P[|X_n - X| \geq \varepsilon] \to 0$ for every ε.
[32] That is $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable.

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 $[0, \infty]$.[33] 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

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

[33] 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 $k = 1, \ldots, n$ 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

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

and

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

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 $p = 1/3, 1/2, 2/3$. ◊

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 $x_0 = 0$. ◊

From now on, until we mention otherwise,

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

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? $\lozenge$

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, **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

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

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

(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 $\{\tau \wedge \sigma \leq t\} = \{\tau \leq t\} \cap \{\sigma \leq t\}$. ☐

Proposition 2.11. *Let X be an adapted process and τ a stopping time. Then $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$ is an $\mathcal{F}_\tau$-measurable random variable. Furthermore, $X^\tau := (X_{\cdot\wedge\tau})$ is an adapted process.*

Proof. Let B be a Borel subset of $\mathbb{R}$. It holds

$$\{X_\tau \in B\} = \cup \{X_\tau \in B\} \cap \{\tau = t\} = \cup \{X_t \in B\} \cap \{\tau = t\} \in \mathcal{F}$$

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

$$A \cap \{\tau \leq t\} = (\cup_{s \leq t} \{X_s \in B\} \cap \{\tau = s\}) \cup (\{X_t \in B\} \cap \{\tau > t\})$$

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 $s \leq t$. Also, $\{X_t \in B\} \cap \{\tau > t\} = \{X_t \in B\} \cap \{\tau \leq t - 1\}^c \in \mathcal{F}_t$. 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. Furthermore $t \wedge \tau$ is also a stopping time smaller than t, and therefore $\mathcal{F}_{t\wedge\tau} \subseteq \mathcal{F}_t$. Since $X_t^\tau = X_{t\wedge\tau}$ 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

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

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 $H1_{\{\cdot\leq\tau\}}$ are predictable. It holds

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

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 $\Omega = \{\tau \leq t-1\} \cup \{t \leq \tau\} = (\cup_{s\leq t-1}\{\tau = s\}) \cup \{t \leq \tau\}$. For every $s = 0, \dots, t-1$, it holds $\{\tau = s\} \in \mathcal{F}_s \subseteq \mathcal{F}_{t-1}$. Also, $\{t \leq \tau\} = \{\tau < t\}^c = \{\tau \leq t-1\}^c \in \mathcal{F}_{t-1}$. Hence, we have a partition of Ω into $\mathcal{F}_{t-1}$-measurable sets. It holds

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

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 $H_t 1_{\{t\leq\tau\}}$ is $\mathcal{F}_{t-1}$-measurable random variable as product of the $\mathcal{F}_{t-1}$-measurable random variables H_t and $1_{\{t\leq\tau\}}$. 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 $\{t \leq \tau\}$ is $\mathcal{F}_{t-1}$-measurable for every t, it follows that $1_{\{\cdot\leq\tau\}}$ is predictable and, therefore, so is $H 1_{\{\cdot\leq\tau\}}$.
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 $(H_t 1_{\{t\leq\tau\}}(X_t - X_{t-1})) = (H_t(X_t^\tau - X_{t-1}^\tau)) = (H_t(X_t - X_{t-1}))^\tau$. 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 $\{\tau \leq t-1\}$ and $X_\tau - X_{t-1}$ on $\{t \leq \tau\}$. □

In particular

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

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 $s \leq t$, $s, t \in \mathbf{T}$.

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

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

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

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

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 $E[\xi \,|\, \mathcal{F}_t] = X_t$, $E[\xi|\mathcal{F}_t] \leq X_t$ 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.[34] ♦

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

[34] Why?

- $p = 1/2$, 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. $\Diamond$

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. $\square$

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. $\square$

Proposition 2.18. *If X is a martingale or super-martingale, then $E[X_\sigma \,|\, \mathcal{F}_\tau] = X_\tau$ or $E[X_\sigma \,|\, \mathcal{F}_\tau] \leq X_\tau$, 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. $\square$

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

Proof. The proof follows that argumentation of the previous proof for the integrability and then application of Proposition 2.17 and Lemma 2.12 for the predictable process $H = 1_{\{\cdot \leq \tau\}}$. □

The proof shows together with the previous proposition that if τ is a stopping time then X^τ is a (super/sub-)martingale whenever X is a (super/sub-)martingale.

Proposition 2.20 (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 $A_t := A_{t-1} - \mathbb{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}]$ 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 2.21. *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. □

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 maximum process X^* by $X^*_t = \sup_{s \leq t} |X_s|$

Proposition 2.22. *The following assertions hold true.*

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

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

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

$$\left\|\sup_{s \leq t} X_s\right\|_p \leq q \left\|X_t\right\|_p$$

where $q = p/(p-1)$ is the conjugate of p.

Proof. (a) Define the stopping times $\tau = \inf\{s : X_s \geq \lambda\} \wedge t$ and $\sigma = \inf\{s : X_s \leq -\lambda\} \wedge t$. Observe that $E[X_t] - E[X_t^-] = E[X_t^+]$ and $E[X_t] - E[1_{\{\sup_{s\leq t} X_s < \lambda\}} X_t] = E\left[1_{\{\sup_{s\leq t} X_s \geq \lambda\}} X_t\right]$. By Doob's optional sampling theorem, and X being a sub-martingale it holds

$$E\left[X_0\right] \geq E\left[X_\tau\right] \geq \lambda P\left[\bar{X}_t \geq \lambda\right] + E\left[1_{\{\bar{X}_t < \lambda\}} X_t\right] \geq \lambda P\left[\bar{X}_t \geq \lambda\right] - E\left[X_t^-\right]$$

and

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

(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, applied to the super-martingale $-X$, 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 $\varphi(\lambda) = \lambda^p$, $p > 1$, it follows from Hölder's inequality that

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

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

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

as desired. If $\|Y\|_p^{p/q} = 0$ the inequality is trivial. If $\|Y\|_p^{p/q} = \infty$, 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 $\lim \tau^n \geq t$ P-almost surely. □

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

$$\|X_t^*\|_p \leq \left(\frac{p}{p-1}\right) \|X_t\|^p$$

for every $p > 1$.

We will consider the following important double sequence of random times. 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

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

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

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

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

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

And finally, the famous Doob's upcrossing's lemma reads as follows.

Theorem 2.23. *Let X be a sub-martingale. Then for every two reals $x<y$, the numbers of up-crossing of $[x,y]$ by X up to time t, $U_{[\![0,t]\!]}(x,y,X)$ where $[\![0,t]\!] := \{0,1,\cdots,t\}$ is a positive random variable and it holds*

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

There exists a similar relation for the down-crossings but we will not make use of it.

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

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

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

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

The function $y \mapsto (y-x)^+$ is increasing and convex function, and therefore, the process $Y = (X-x)^+$ defines a sub-martingale due to Proposition 2.16. Clearly, $U := U_{[\![0,t]\!]}\left(x,y,X\right)$ counts the number of up-crossings of $[0, y-x]$ up to time t and therefore

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

Since Y is submartingale, and $\{t > \tau_{2U}\} = \{t \leq \tau_{2U-1}\}^c \in \mathcal{F}_{\tau_{2U-1}}$, it follows that

$$E\left[(Y_t - Y_{\tau_{2U-1}})1_{\{t>\tau_{2U}\}}\right] = E\left[E\left[(Y_t - Y_{\tau_{2U-1}})|\mathcal{F}_{\tau_{2U-1}}\right]1_{\{t>\tau_{2U}\}}\right] \geq 0$$

Hence, since $Y_{\tau_{2k}} - Y_{\tau_{2k-1}} \geq (y-x)$, it follows that

$$E[H \bullet Y_t] \geq (y-x)E\left[U\right].$$

Defining $K = 1 - H$ which is a positive predictable process, hence by means of Proposition 2.17, it follows that $K \bullet Y$ is a submartingale and therefore $E[K \bullet Y_t] \geq E[K \bullet Y_0] \geq 0$. It follows that

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

which ends the proof. □