Probability Theory 2 (NMSA405)

Zbyněk Pawlas

January 13, 2016

1 Random sequence

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$. We consider a sequence of real random variables $\{X_n, n \in \mathbb{N}\}$, i.e. measurable mappings $X_n : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, where $\mathcal{B} = \mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

Definition 1.1. A sequence $X_1, X_2, \dots$ of random variables is called a *random sequence.*

Let $\mathbb{R}^{\mathbb{N}}$ be the space of all sequences of real numbers. The random sequence $X_1, X_2, \dots$ creates a mapping $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ defined by

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots), \quad \omega \in \Omega. \tag{1}$$

The question is whether this mapping is in some sense measurable.

Definition 1.2. If $(S_n, \mathcal{S}_n)$ are measurable spaces, we define a *product σ-algebra* $\bigotimes_{n=1}^{\infty} \mathcal{S}_n$ of subsets of the product space $\bigtimes_{n=1}^{\infty} S_n$ as

$$\bigotimes_{n=1}^{\infty} \mathcal{S}_n = \sigma\{A_1 \times A_2 \times \cdots \times A_n \times S_{n+1} \times S_{n+2} \times \cdots : A_k \in \mathcal{S}_k, n \in \mathbb{N}\}.$$

The set of the form $A_1 \times A_2 \times \cdots \times A_n \times S_{n+1} \times S_{n+2} \times \cdots$ is called a *finite-dimensional cylinder set.* When $S_n = S$ and $\mathcal{S}_n = \mathcal{S}$ for any $n \in \mathbb{N}$, we use the notation $S^{\mathbb{N}} = \bigtimes_{n=1}^{\infty} S_n$ and $\mathcal{S}^{\mathbb{N}} = \bigotimes_{n=1}^{\infty} \mathcal{S}_n$.

In our situation we have $S_n = \mathbb{R}$ and $\mathcal{S}_n = \mathcal{B}$.

Proposition 1.1. *If $X_1, X_2, \dots$ is a random sequence, then X defined in* (1) *is measurable with respect to $\mathcal{B}^{\mathbb{N}}$. We write $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$.*

Proof. Applying Definition 1.2 for $S_n = \mathbb{R}$ and $\mathcal{S}_n = \mathcal{B}$, it suffices to verify that $F = [X \in A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R} \times \mathbb{R} \times \cdots] \in \mathcal{F}$ for any $A_1, \dots, A_n \in \mathcal{B}$ and $n \in \mathbb{N}$. However, this is obvious because $F = \bigcap_{k=1}^{n} [X_k \in A_k] \in \mathcal{F}$. □

The space $\mathbb{R}^{\mathbb{N}}$ can be naturally turned into metric space.

Definition 1.3. Define the distance of two real sequences $x = (x_1, x_2, \dots)$ and $y = (y_1, y_2, \dots)$ as

$$d(x, y) = \sum_{n=1}^{\infty} \frac{|x_n - y_n| \wedge 1}{2^n},$$

where $a \wedge b = \min\{a, b\}$.

Remark: Note that the series in definition of d is always convergent and $d(x,y) \le 1$ for any $x, y \in \mathbb{R}^{\mathbb{N}}$.

Proposition 1.2. *(a) The function d is a metric on $\mathbb{R}^{\mathbb{N}}$.*

(b) The sequence $x^n = (x_1^n, x_2^n, \dots) \in \mathbb{R}^{\mathbb{N}}$ converges as $n \to \infty$ to the sequence $x = (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}}$ in metric d if and only if $x_j^n \underset{n\to\infty}{\longrightarrow} x_j$ for any $j \in \mathbb{N}$.

(c) The metric space $(\mathbb{R}^{\mathbb{N}}, d)$ is separable and complete.

(d) For $-\infty < a_n \le b_n < \infty$, the set $\bigtimes_{n=1}^{\infty} [a_n, b_n]$ is a compact subset of $\mathbb{R}^{\mathbb{N}}$.

Proof. (a), (b), (c) at exercise classes.

(d) Without loss of generality assume $[a_n, b_n] = [0,1]$. Let $x^n = (x_1^n, x_2^n, \dots)$ be a sequence of elements from $[0,1]^{\mathbb{N}}$. Since $x_1^1, x_1^2, \dots$ is a sequence of real numbers in compact interval $[0,1]$, there exists a subsequence $x_1^{n(1,1)}, x_1^{n(1,2)}, \dots, x_1^{n(1,k)}, \dots$, that converges to some $x_1 \in [0,1]$ as $k \to \infty$. Next we can find a subsequence $\{n(2,k), k \in \mathbb{N}\}$ of $\{n(1,k), k \in \mathbb{N}\}$ such that $x_2^{n(2,k)} \underset{k\to\infty}{\longrightarrow} x_2 \in [0,1]$. By induction we construct

$$\begin{gathered}
x_1^{n(1,1)}, x_1^{n(1,2)}, \dots, x_1^{n(1,k)}, \dots \underset{k\to\infty}{\longrightarrow} x_1 \in [0,1], \\
x_2^{n(2,1)}, x_2^{n(2,2)}, \dots, x_2^{n(2,k)}, \dots \underset{k\to\infty}{\longrightarrow} x_2 \in [0,1], \\
\vdots \\
x_\ell^{n(\ell,1)}, x_\ell^{n(\ell,2)}, \dots, x_\ell^{n(\ell,k)}, \dots \underset{k\to\infty}{\longrightarrow} x_\ell \in [0,1], \\
\vdots
\end{gathered}$$

so that $\{n(\ell+1,k), k \in \mathbb{N}\}$ is a subsequence of $\{n(\ell,k), k \in \mathbb{N}\}$. We use the diagonal selection principle and consider a sequence $\{n(k,k), k \in \mathbb{N}\}$. For any $\ell \in \mathbb{N}$ we have $\{n(k,k), k \ge \ell\} \subseteq \{n(\ell,k), k \in \mathbb{N}\}$ and hence

$$x_\ell^{n(k,k)} \underset{k\to\infty}{\longrightarrow} x_\ell.$$

Using part b) we get that $x^{n(k,k)} = (x_1^{n(k,k)}, x_2^{n(k,k)}, \dots)$ converges in metric d as $k \to \infty$ to the sequence $x = (x_1, x_2, \dots) \in [0,1]^{\mathbb{N}}$. This proves that $[0,1]^{\mathbb{N}}$ is compact. □

Now let us examine how the Borel σ-algebra $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ looks like.

Theorem 1.3. *The relation $\mathcal{B}^{\mathbb{N}} = \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ holds, i.e.*

$$\sigma\{A_1 \times \cdots \times A_n \times \mathbb{R} \times \cdots : A_1, \dots, A_n \in \mathcal{B}, n \in \mathbb{N}\} = \sigma\{U : U \subseteq \mathbb{R}^{\mathbb{N}} \text{ open set}\}.$$

Proof. The inclusion $\mathcal{B}^{\mathbb{N}} \subseteq \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is obvious. We have to show that every open set $U \subseteq \mathbb{R}^{\mathbb{N}}$ lies in σ-algebra $\mathcal{B}^{\mathbb{N}}$. For each $x \in U$ there exists $\delta_x > 0$ such that $U = \bigcup_{x \in U} B(x, \delta_x)$, where $B(x,\delta_x) = \{y : d(y,x) < \delta_x\}$ is an open ball in $\mathbb{R}^{\mathbb{N}}$ with centre x and of radius δ_x. The metric space $\mathbb{R}^{\mathbb{N}}$ is separable (Proposition 1.2). Therefore, from an open covering of U we can select a countable subcollection (Lindelöf's covering theorem) which also covers U. We get $U = \bigcup_{k=1}^{\infty} B(x_k, \delta_{x_k})$, where $x_k \in U$. In order to finish the proof it suffices to show that $B(x,\delta) \in \mathcal{B}^{\mathbb{N}}$ for any $x \in \mathbb{R}^{\mathbb{N}}$ and $\delta > 0$. For fixed $x \in \mathbb{R}^{\mathbb{N}}$ the map $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^+$ given by $T : y \mapsto \sum_{j=1}^{\infty} 2^{-j}(|x_j - y_j| \wedge 1)$ is measurable with respect to $\mathcal{B}^{\mathbb{N}}$. Hence, $B(x,\delta) = T^{-1}((0,\delta)) \in \mathcal{B}^{\mathbb{N}}$. □

Definition 1.4. Let E be a metric space. Then $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B}(E))$ is called a *random element* with values in E.

Corollary 1.4. *A random sequence $X = (X_1, X_2, \dots)$ is a random element with values in $\mathbb{R}^{\mathbb{N}}$.*

Proof. The assertion follows from Proposition 1.1 and Theorem 1.3. □

There exist several useful non-trivial sub-σ-algebras of σ-algebra $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$.

Definition 1.5. The map $p : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is called a *finite permutation (of order n)* if there exist $n \in \mathbb{N}$ and permutation $(k_1, \dots, k_n)$ of the set $\{1, \dots, n\}$ such that

$$p(x_1, \dots, x_n, x_{n+1}, \dots) = (x_{k_1}, \dots, x_{k_n}, x_{n+1}, \dots), \quad (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}}.$$

Definition 1.6. The map $s : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ given by

$$s(x_1, x_2, \dots) = (x_2, x_3, \dots), \quad (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}},$$

is called a *shift*.

Definition 1.7. The set $T \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is called *terminal* if the following implication holds:

$$x = (x_1, x_2, \dots) \in T, y = (y_1, y_2, \dots) \in \mathbb{R}^{\mathbb{N}} : y_k = x_k \text{ for all but finitely many } k \in \mathbb{N} \ \Rightarrow y \in T.$$

We say that $T \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is *n-terminal* if

$$x = (x_1, x_2, \dots) \in T, y = (y_1, y_2, \dots) \in \mathbb{R}^{\mathbb{N}} : y_k = x_k \text{ for } k > n \ \Rightarrow y \in T.$$

Definition 1.8. Denote the following collections of sets:

- *n-symmetric sets*: $\mathcal{S}_n = \{S \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) : p(S) = S \text{ for any finite permutation } p \text{ of order } n\}$,
- *symmetric sets*: $\mathcal{S} = \{S \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) : p(S) = S \text{ for any finite permutation } p\}$,
- *shift-invariant sets*: $\mathcal{I} = \{I \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) : s^{-1}I = I\}$,
- *n-terminal sets*: $\mathcal{T}_n = \{T \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) : T \ n\text{-terminal}\}$.
- *terminal sets*: $\mathcal{T} = \{T \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) : T \text{ terminal}\}$.

Proposition 1.5. *(a) Any finite permutation $p : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is homeomorphism.*

(b) Shift s is continuous map.

(c) The set $T \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is n-terminal if and only if there exists $T_n \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ such that $T = \mathbb{R}^n \times T_n$.

(d) The collections $\mathcal{I}$, $\mathcal{T}$ and $\mathcal{S}$ are σ-algebras such that $\mathcal{I} \subset \mathcal{T}_n \subset \mathcal{S}_n$ for any $n \in \mathbb{N}$. Consequently, $\mathcal{I} \subset \mathcal{T} \subset \mathcal{S}$. All inclusions are strict, i.e. $\mathcal{I} \neq \mathcal{T}_n \neq \mathcal{S}_n$ and $\mathcal{I} \neq \mathcal{T} \neq \mathcal{S}$.

Proof. Exercise class. □

By Corollary 1.4 the random sequence $X = (X_1, X_2, \dots)$ is a random element with values in $\mathbb{R}^{\mathbb{N}}$. Hence, it has a probability distribution.

Definition 1.9. Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B}(E))$ be a random element with values in a metric space E. Let $\mathcal{P}(E)$ denote the family of Borel probability measures on E. Define $P_X(B) = \mathbb{P}(X \in B)$ for $B \in \mathcal{B}(E)$. Then $P_X \in \mathcal{P}(E)$ is called a *probability distribution* of X.

It means that the probability distribution of a random sequence $X = (X_1, X_2, \dots)$ is a probability measure P_X on $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ defined as $P_X(B) = \mathbb{P}((X_1, X_2, \dots) \in B)$ for $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$. We are going to show that the distribution of X is determined by the set of finite-dimensional distributions.

Definition 1.10. We say that the set $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ is *finite-dimensional* if there exist $n \in \mathbb{N}$ and $B_n \in \mathcal{B}(\mathbb{R}^n)$ such that $B = B_n \times \mathbb{R}^{\mathbb{N}}$.

Proposition 1.6. *Let $\mathcal{A}$ be the family of finite-dimensional set. This system is an algebra that generates $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$, i.e. $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R}^{\mathbb{N}})$.*

Proof. Exercise class. □

Theorem 1.7. *The probability distribution of a random sequence $X = (X_1, X_2, \dots)$ is uniquely determined by the probability distribution of all random vectors $(X_1, X_2, \dots, X_n)$, $n \in \mathbb{N}$.*

Proof. Let $X = (X_1, X_2, \dots)$ and $Y = (Y_1, Y_2, \dots)$ be random sequences satisfying $P_{(X_1,\dots,X_n)} = P_{(Y_1,\dots,Y_n)}$ for any $n \in \mathbb{N}$. We have to show that $P_X = P_Y$. Let $B = B_n \times \mathbb{R}^{\mathbb{N}}$ be a finite-dimensional set. Then

$$P_X(B) = \mathbb{P}(X \in B) = \mathbb{P}((X_1, \dots, X_n) \in B_n) = \mathbb{P}((Y_1, \dots, Y_n) \in B) = \mathbb{P}(Y \in B) = P_Y(B).$$

The measures P_X and P_Y coincide on algebra $\mathcal{A}$ of finite-dimensional sets. From measure theory we know that if two finite measures coincide on some π-system (system closed under finite intersections) then they coincide on the σ-algebra generated by this π-system. Applying Proposition 1.6 we get that $P_X = P_Y$ on $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$. □

Fundamental problem: We prescribe finite-dimensional probability distributions $P_n \in \mathcal{P}(\mathbb{R}^n)$ for $n \in \mathbb{N}$. When does the random sequence $X = (X_1, X_2, \dots)$ exist so that $P_{(X_1,\dots,X_n)} = P_n$ for every $n \in \mathbb{N}$?

We easily find a necessary condition.

Definition 1.11. We say that a sequence $\{P_n \in \mathcal{P}(\mathbb{R}^n), n \in \mathbb{N}\}$ of probability distributions is *projective* if $P_{n+1}(B_n \times \mathbb{R}) = P_n(B_n)$, $B_n \in \mathcal{B}^n$, $n \in \mathbb{N}$, i.e. P_n is a marginal distribution of P_{n+1} for arbitrary $n \in \mathbb{N}$.

The distribution $P_n \in \mathcal{P}(\mathbb{R}^n)$ is uniquely determined by its distribution function

$$F_n(x_1, \dots, x_n) = P_n\big((-\infty, x_1] \times \cdots \times (-\infty, x_n]\big), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Therefore, we easily get the following result.

Proposition 1.8. *The system $\{P_n \in \mathcal{P}(\mathbb{R}^n), n \in \mathbb{N}\}$ is projective if and only if*

$$\lim_{x_{n+1} \to \infty} F_{n+1}(x_1, \dots, x_n, x_{n+1}) = F_n(x_1, \dots, x_n) \quad \textit{for } (x_1, \dots, x_n) \in \mathbb{R}^n \textit{ and } n \in \mathbb{N},$$

where F_n denotes the distribution function of P_n.

It is obvious that $\{P_n, n \in \mathbb{N}\}$ in our fundamental problem must be projective. A deep result is that this condition is also sufficient. The following two theorems are special cases of the Daniell-Kolmogorov extension theorem.

Theorem 1.9. (Daniell's extension theorem) *Let $\{P_n \in \mathcal{P}(\mathbb{R}^n), n \in \mathbb{N}\}$ be a projective family. Then there exists a random sequence $X = (X_1, X_2, \dots)$ such that $P_{(X_1,\dots,X_n)} = P_n$ for all $n \in \mathbb{N}$.*

Theorem 1.10. *Let $\{P_n \in \mathcal{P}(\mathbb{R}^n), n \in \mathbb{N}\}$ be a projective family. Then there exists a unique Borel probability measure P on $\mathbb{R}^{\mathbb{N}}$ such that*

$$P(B_n \times \mathbb{R}^{\mathbb{N}}) = P_n(B_n), \quad B_n \in \mathcal{B}^n, \ n \in \mathbb{N}. \tag{2}$$

We postpone the proof of Theorem 1.10 to the end of this section. Theorem 1.9 is a consequence of Theorem 1.10.

Proof. (of Theorem 1.9) By Theorem 1.10 there exists $P \in \mathcal{P}(\mathbb{R}^{\mathbb{N}})$ satisfying (2). We use a canonical construction. Take $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), P)$ and $X = \text{id}$. The projections $X_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, $n \in \mathbb{N}$, given by

$$X_n(x_1, \dots, x_n, x_{n+1}, \dots) = x_n \quad \text{pro } (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}},$$

are continuous and hence measurable in the sense $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \to (\mathbb{R}, \mathcal{B})$. The distribution of the random vector $(X_1, \dots, X_n)$ is

$$P_{(X_1,\dots,X_n)}(B_n) = \mathbb{P}((X_1, \dots, X_n) \in B_n) = \mathbb{P}(X \in B_n \times \mathbb{R}^{\mathbb{N}}) = P(B_n \times \mathbb{R}^{\mathbb{N}}) = P_n(B_n), \quad B_n \in \mathcal{B}^n.$$

The random sequence $X = (X_1, X_2, \dots)$ satisfies the required property. □

Recall the notion of product measure from measure theory.

Definition 1.12. Let $Q_1, \dots, Q_n \in \mathcal{P}(\mathbb{R})$. A *product measure* $Q = \bigotimes_{k=1}^{n} Q_k$ is a unique probability measure in $\mathcal{P}(\mathbb{R}^n)$ satisfying the property $Q(\bigtimes_{k=1}^{n} B_k) = Q_1(B_1) \cdots Q_n(B_n)$ for all $B_1, \dots, B_n \in \mathcal{B}$.

Proposition 1.11. *The random variables $X_1, \dots, X_n$ are independent if and only if $P_{(X_1,\dots,X_n)} = \bigotimes_{k=1}^{n} P_{X_k}$.*

Proof. See Probability Theory 1. □

If $\{Q_k \in \mathcal{P}(\mathbb{R}), k \in \mathbb{N}\}$ is the sequence of probability measures, then it is clear that $\{P_n = \bigotimes_{k=1}^{n} Q_k, n \in \mathbb{N}\}$ is a projective family. Then Theorem 1.10 has the following form.

Theorem 1.12. *For an arbitrary sequence $\{Q_k \in \mathcal{P}(\mathbb{R}), k \in \mathbb{N}\}$ there exists a unique probability measure $P \in \mathcal{P}(\mathbb{R}^{\mathbb{N}})$ satisfying*

$$P(B_n \times \mathbb{R}^{\mathbb{N}}) = \left(\bigotimes_{k=1}^{n} Q_k\right)(B_n), \quad B_n \in \mathcal{B}^n, n \in \mathbb{N}.$$

Theorem 1.9 then can be stated in the following form.

Theorem 1.13. *For an arbitrary sequence $\{Q_k \in \mathcal{P}(\mathbb{R}), k \in \mathbb{N}\}$ there exists a sequence $X = (X_1, X_2, \dots)$ of independent random variables such that $P_{X_k} = Q_k$, $k \in \mathbb{N}$. Moreover, $P_X = \bigotimes_{k=1}^{\infty} Q_k$.*

Proof. Since the system $\{P_n = \bigotimes_{k=1}^{n} Q_k, n \in \mathbb{N}\}$ is projective, by Theorem 1.9 there exists a random sequence $X = (X_1, X_2, \dots)$ such that $P_{(X_1,\dots,X_n)} = P_n = \bigotimes_{k=1}^{n} Q_k$, $n \in \mathbb{N}$. Hence, $P_{X_k} = Q_k$ for any $k \in \mathbb{N}$ and random variables $X_1, \dots, X_n$ are independent for any $n \in \mathbb{N}$ by Proposition 1.11. This in turn implies that random variables $X_1, X_2, \dots$ are independent. The equality $P_X = \bigotimes_{k=1}^{\infty} Q_k$ follows from $P_{(X_1,\dots,X_n)} = \bigotimes_{k=1}^{n} Q_k$, $n \in \mathbb{N}$ because $\bigotimes_{k=1}^{\infty} Q_k$ is uniquely determined probability measure satisfying

$$\left(\bigotimes_{k=1}^{\infty} Q_k\right)(B_1 \times B_2 \times \cdots \times B_n \times \mathbb{R} \times \cdots) = Q_1(B_1) \cdots Q_n(B_n) = \left(\bigotimes_{k=1}^{n} Q_k\right)(B_1 \times \cdots \times B_n)$$

for any finite-dimensional cylinder set $B_1 \times B_2 \times \cdots \times B_n \times \mathbb{R} \times \cdots$. □

If we prescribe one-dimensional distributions Q_k, then there always exists a sequence of independent random variables X_k that have distribution Q_k, $k \in \mathbb{N}$. When the Q_k are Bernoulli distributions with parameter $p \in [0, 1]$, this gives a mathematical model for a sequence of Bernoulli trials with the probability of success p In the case $p = 1/2$ we can proceed directly without needing Daniell's extension theorem.

Definition 1.13. *Binary expansion* of a number $x \in (0, 1]$ is a sequence $x_1, x_2, \dots$ of zeros and ones that contains infinitely many ones and

$$x = \sum_{k=1}^{\infty} \frac{x_k}{2^k}.$$

Binary expansion of $x = 0$ is a sequence of zeros.

Proposition 1.14. *Let X be a random variable with uniform distribution on $[0, 1]$ and let*

$$X(\omega) = \sum_{k=1}^{\infty} \frac{X_k(\omega)}{2^k} \tag{3}$$

be its binary expansion. Then $X_1, X_2, \dots$ is a sequence of independent random variables having Bernoulli distribution with parameter $1/2$.

Conversely, if we consider a sequence of independent random variables having Bernoulli distribution with parameter $1/2$, then the random variable X defined by (3) *has a uniform distribution on the interval* $[0, 1]$.

Proof. Exercise class. □

We will deal with several important types of random sequences that describe the movement of the particle in times $n = 1, 2, \dots$

Definition 1.14. We say that a random sequence $X = (X_1, X_2, \dots)$ is

- *iid*, if random variables X_j, $j \in \mathbb{N}$, are independent and identically distributed,
- *n-symmetric*, if $(X_1, \dots, X_n, X_{n+1}, \dots)$ and $(X_{k_1}, \dots, X_{k_n}, X_{n+1}, \dots)$ have the same distributions for every finite permutation $(k_1, \dots, k_n)$ of order $n \in \mathbb{N}$,
- *symmetric*, if it is n-symmetric for all $n \in \mathbb{N}$,
- *stationary*, if the distributions of $(X_1, \dots, X_n, X_{n+1}, \dots)$ and $(X_{n+1}, X_{n+2}, \dots)$ coincide for all $n \in \mathbb{N}$.

Examples and relations between these types of sequences are left to exercise classes. Other important types are Markov chains (course Stochastic processes 1) and martingales, which we are going to study in more details in next sections.

We end this section with the promised proof of Theorem 1.10.

Proof. (of Theorem 1.10) Let $\{P_n \in \mathcal{P}(\mathbb{R}^n), n \in \mathbb{N}\}$ be a projective system. Relation (2) defines a finitely additive probability measure P on algebra $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ of finite-dimensional sets.

We first need to verify that the definition is correct. Let a finite-dimensional set be expressed in two ways as $B_n \times \mathbb{R}^{\mathbb{N}} = B_m \times \mathbb{R}^{\mathbb{N}}$, where $B_n \in \mathcal{B}^n$, $B_m \in \mathcal{B}^m$ and $m > n$. Then $B_m = B_n \times \mathbb{R}^{m-n}$ and the projectivity property implies $P_m(B_m) = P_m(B_n \times \mathbb{R}^{m-n}) = P_n(B_n)$.

Next we verify that P is finitely additive on $\mathcal{A}$. If A and B are finite-dimensional sets, then there exist $n \in \mathbb{N}$ and $A_n, B_n \in \mathcal{B}^n$ such that $A = A_n \times \mathbb{R}^{\mathbb{N}}$ and $B = B_n \times \mathbb{R}^{\mathbb{N}}$. For disjoint sets A and B it is obvious that A_n and B_n are disjoint and $A \cup B = (A_n \cup B_n) \times \mathbb{R}^{\mathbb{N}}$. Therefore,

$$P(A \cup B) = P_n(A_n \cup B_n) = P_n(A_n) + P_n(B_n) = P(A) + P(B).$$

It remains to show that P is σ-additive probability measure on algebra $\mathcal{A}$. Then by Hopf extension theorem, P can be extended to a unique probability measure $\bar{P}$ on σ-algebra $\sigma(\mathcal{A})$. Proposition 1.6 claims that algebra $\mathcal{A}$ generates $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$. Thus, the extension $\bar{P}$ is defined on $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ and has the desired properties.

Let $A^n = B^n_{k_n} \times \mathbb{R}^{\mathbb{N}}$ be sets from $\mathcal{A}$ such that $A^1 \supseteq A^2 \supseteq \cdots$ and $\cap_{n=1}^{\infty} A^n = \emptyset$ (we write shortly $A^n \searrow \emptyset$). Since A^n is a monotone sequence, we can assume that the sets $B^n_{k_n} \in \mathcal{B}^{k_n}$ are chosen so that $k_1 < k_2 < \cdots$. Choose arbitrary $\varepsilon > 0$. The measure $P_{k_n} \in \mathcal{P}(\mathbb{R}^{k_n})$ is tight for any $n \in \mathbb{N}$. It means that there exists a compact set $K^n \subseteq B^n_{k_n}$ satisfying $P_{k_n}(B^n_{k_n} \setminus K^n) < \varepsilon/2^n$. We construct finite-dimensional sets $C^n = K^n \times \mathbb{R}^{\mathbb{N}} \in \mathcal{A}$. From the construction it follows that $C^n \subseteq A^n$, hence $\cap_{n=1}^{\infty} C^n = \emptyset$.

We can find $m \in \mathbb{N}$ such that $\cap_{n=1}^{m} C^n = \emptyset$. By contradiction assume that $\cap_{n=1}^{m} C^n \neq \emptyset$ for all $m \in \mathbb{N}$. Then there exist sequences $x^m = (x^m_1, x^m_2, \dots) \in \mathbb{R}^{\mathbb{N}}$, $m \in \mathbb{N}$, such that $(x^m_1, \dots, x^m_{k_n}) \in K^n$ for all $n = 1, \dots, m$. So we obtained a sequence $\{x^m, m \in \mathbb{N}\}$ in $\mathbb{R}^{\mathbb{N}}$ such that every sequence $(x^1_\ell, x^2_\ell, \dots)$ is bounded in $\mathbb{R}$. By Proposition 1.2d, the sequence $\{x^m, m \in \mathbb{N}\}$ has a limit point $x \in \mathbb{R}^{\mathbb{N}}$. From the construction we see that $x \in \cap_{n=1}^{\infty} C^n = \emptyset$, which leads to the desired contradiction.

Let $m \in \mathbb{N}$ be such that $\cap_{n=1}^{m} C^n = \emptyset$. Then

$$P(A^m) = P(A^m \setminus \cap_{n=1}^{m} C^n) \leq \sum_{n=1}^{m} P(A^n \setminus C^n) = \sum_{n=1}^{m} P_{k_n}(B^n_{k_n} \setminus K^n) < \sum_{n=1}^{m} \frac{\varepsilon}{2^n} < \varepsilon.$$

We used the relation $A^m \setminus \cap_{n=1}^m C^n \subseteq \cup_{n=1}^m (A^n \setminus C^n)$. Hence, we have $P(A^n) \le P(A^m) < \varepsilon$ for $n \ge m$, leading to $P(A^n) \searrow 0$ for $n \to \infty$.

If $\tilde{A}^1, \tilde{A}^2, \dots \in \mathcal{A}$ are pairwise disjoint such that $\cup_{k=1}^\infty \tilde{A}^k \in \mathcal{A}$, then $A^n = \cup_{k=n}^\infty \tilde{A}^k \in \mathcal{A}$ for $n \in \mathbb{N}$ and $A^n \searrow \emptyset$. We have already shown that $P(A^n) \searrow 0$ for $n \to \infty$. From this fact and finite additivity of P we can deduce that

$$P\left(\bigcup_{k=1}^{\infty} \tilde{A}^k\right) = P\left(\bigcup_{k=1}^{n-1} \tilde{A}^k\right) + P(A^n) = \sum_{k=1}^{n-1} P(\tilde{A}^k) + P(A^n) \underset{n\to\infty}{\longrightarrow} \sum_{k=1}^{\infty} P(\tilde{A}^k).$$

We proved the desired σ-additivity of P, which completes the proof. □

2 Stopping times, filtration and martingales

Let $X = (X_1, X_2, \dots)$ be a random sequence. It models the random movement of the particle in times $t = 1, 2, \dots$ Then the events that the particle encounters until time n are collected in σ-algebra $\mathcal{F}_n = \sigma(X_1, \dots, X_n) = \{[(X_1, \dots, X_n) \in B_n], B_n \in \mathcal{B}^n\}$. All events are collected in σ-algebra $\mathcal{F}_\infty = \sigma(X) = \{[X \in B], B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})\}$.

Proposition 2.1. *The following relation holds:* $\sigma(X) = \sigma\left(\cup_{n=1}^\infty \sigma(X_1, \dots, X_n)\right)$.

Proof. Exercise class. □

Definition 2.1. Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}$ be a non-decreasing sequence of σ-algebras. We say that $(\mathcal{F}_n)$ is a *filtration*. We denote $\mathcal{F}_\infty = \sigma(\cup_{n=1}^\infty \mathcal{F}_n)$.

Let $X = (X_1, X_2, \dots)$ be a sequence of random variables defined on $(\Omega, \mathcal{F})$ and let $(\mathcal{F}_n)$ be a filtration satisfying $\sigma(X_1, \dots, X_n) \subseteq \mathcal{F}_n$ for all $n \in \mathbb{N}$. We say that the sequence (X_n) is $\mathcal{F}_n$*-adapted*.

If $\sigma(X_1, \dots, X_n) = \mathcal{F}_n$ for all $n \in \mathbb{N}$, we say that $(\mathcal{F}_n)$ is a *canonical filtration* of $X = (X_1, X_2, \dots)$.

Proposition 2.2. *Let $X = (X_1, X_2, \dots)$ be a random sequence and let $S = (S_1, S_2, \dots)$ be a sequence of its partial sums, i.e. $S_n = \sum_{k=1}^n X_k$. Then X and S have the same canonical filtration, i.e. $\sigma(X) = \sigma(S)$.*

Proof. Exercise class. □

The movement of a particle goes through some important events.

Definition 2.2. Let $X = (X_1, X_2, \dots)$ be a random sequence. For $B \in \mathcal{B}$ denote $T_B(\omega) = \min\{n : X_n(\omega) \in B\}$, where $\min \emptyset = \infty$. We say that T_B is a *first hitting time* of the set B by the sequence X.

Note that

$$[T_B \le n] = \bigcup_{k=1}^{n} [X_k \in B] \in \sigma(X_1, \dots, X_n) \subseteq \mathcal{F}, \quad n \in \mathbb{N},$$

i.e. T_B is a random variable.

Definition 2.3. The mapping $T : \Omega \to \mathbb{N} \cup \{\infty\}$ is called a *stopping time (or Markov time)* with respect to the filtration $(\mathcal{F}_n)$ provided that $[T \le n] \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. Shortly we speak about $\mathcal{F}_n$*-stopping time (or $\mathcal{F}_n$-Markov time)*.

Let $X = (X_1, X_2, \dots)$ be a random sequence. A stopping time T with respect to the canonical filtration is called a *stopping time of the sequence* X, i.e. $T : \Omega \to \mathbb{N} \cup \{\infty\}$ and $[T \le n] \in \sigma(X_1, \dots, X_n)$ for all $n \in \mathbb{N}$.

The first hitting time T_B is a stopping time of X because $[T_B \le n] \in \sigma(X_1, \dots, X_n)$ for every $n \in \mathbb{N}$. This stopping time is a random variable that gives no information about the behaviour of X after time T_B.

We are looking for a suitable definition of σ-algebra that represents our information about the random sequence X up to the stopping time T.

Definition 2.4. Let $(\mathcal{F}_n)$ be a filtration and T be an $\mathcal{F}_n$-stopping time. Define

$$\mathcal{F}_T = \{F \in \mathcal{F}_\infty : F \cap [T \le n] \in \mathcal{F}_n \ \forall n \in \mathbb{N}\}.$$

Then $\mathcal{F}_T$ is a σ-algebra that is called *stopping time σ-algebra*.

At the moment $T(\omega) < \infty$ the particle occurs in $X_{T(\omega)}(\omega)$. For $\omega \in \Omega$ we denote

$$X_T(\omega) = \begin{cases} X_{T(\omega)}(\omega) & \text{if } T(\omega) < \infty, \\ 0 & \text{if } T(\omega) = \infty. \end{cases}$$

If $T < \infty$ a.s. we write shortly $T \overset{a.s.}{<} \infty$. In that case, X_T is almost surely a value of the random sequence stopped at time T.

Proposition 2.3. *Let $(\mathcal{F}_n)$ be a filtration. Then T is an $\mathcal{F}_n$-stopping time if and only if $[T = n] \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. Furthermore,*

$$\mathcal{F}_T = \{F \in \mathcal{F}_\infty : F \cap [T = n] \in \mathcal{F}_n \ \forall n \in \mathbb{N}\}.$$

Proof. Exercise class. □

Proposition 2.4. (calculus for stopping times) *Let $(\mathcal{F}_n)$ be a filtration. If S and T are $\mathcal{F}_n$-stopping times and (X_n) is an $\mathcal{F}_n$-adapted random sequence, then*

a) T and X_T are $\mathcal{F}_T$-measurable random variables,

b) $S \wedge T$, $S \vee T$ and $S + T$ are $\mathcal{F}_n$-stopping times,

c) $T \wedge n$ is $\mathcal{F}_n$-measurable random variable for any $n \in \mathbb{N}$,

d) $F \in \mathcal{F}_S \Rightarrow F \cap [S \leq T] \in \mathcal{F}_T$,

e) $S \leq T \Rightarrow \mathcal{F}_S \subseteq \mathcal{F}_T$,

f) $[S \leq T], [S = T] \in \mathcal{F}_S \cap \mathcal{F}_T$,

g) $\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S \wedge T}$.

Proof. a), b), c) exercise.

d) According to Proposition 2.3 we have to show that $F \in \mathcal{F}_S \Rightarrow F \cap [S \leq T] \cap [T = n] \in \mathcal{F}_n$ for all $n \in \mathbb{N}$:

$$F \cap [S \leq T] \cap [T = n] = F \cap [S \leq n] \cap [T = n] \in \mathcal{F}_n \cap \mathcal{F}_n = \mathcal{F}_n.$$

e) By part d), $S \leq T$ and $F \in \mathcal{F}_S$ implies that $F = F \cap [S \leq T] \in \mathcal{F}_T$.

f) By part d), we have $[S \leq T] = \Omega \cap [S \leq T] \in \mathcal{F}_T$. If we put $\lambda = S \wedge T$, then λ is a stopping time by part b). Since $[\lambda = T] \cap [T = n] = [\lambda = n] \cap [T = n] \in \mathcal{F}_n$, we get $[\lambda = T] = [S \geq T] \in \mathcal{F}_T$ by Proposition 2.3. Altogether we have $[S \leq T], [S \geq T] \in \mathcal{F}_T$ and hence also $[S = T] = [S \leq T] \cap [S \geq T] \in \mathcal{F}_T$. The events $[S \leq T], [S \geq T], [S = T]$ belong to $\mathcal{F}_S$ from the symmety.

g) By part e) we get $\mathcal{F}_{S \wedge T} \subseteq \mathcal{F}_S \cap \mathcal{F}_T$. Let $F \in \mathcal{F}_S \cap \mathcal{F}_T$. Then

$$F \cap [S \wedge T \leq n] = (F \cap [T \leq S] \cap [T \leq n]) \cup (F \cap [S \leq T] \cap [S \leq n]).$$

From part f) we have $F \cap [T \leq S] \in \mathcal{F}_T$ and $F \cap [S \leq T] \in \mathcal{F}_S$. Consequently, $F \cap [T \leq S] \cap [T \leq n] \in \mathcal{F}_n$ and $F \cap [S \leq T] \cap [S \leq n] \in \mathcal{F}_n$, which leads to $F \cap [S \wedge T \leq n] \in \mathcal{F}_n$ for any $n \in \mathbb{N}$. It means that $F \in \mathcal{F}_{S \wedge T}$. □

Proposition 2.5. *a) Let T be an $\mathcal{F}_n$-stopping time. Let $\lambda : \Omega \to \mathbb{N} \cup \{\infty\}$ be an $\mathcal{F}_T$-measurable random variable such that $\lambda \geq T$. Then λ is an $\mathcal{F}_n$-stopping time.*

b) Let $X = (X_1, X_2, \dots)$ be a random sequence and T its stopping time. For $B \in \mathcal{B}$ define $\lambda = \min\{k > T : X_k \in B\}$, it is the first hitting time of B after time T. Then λ is a stopping time of X.

Proof. Exercise class. □

Definition 2.5. Let $X_1, X_2, \dots$ be an iid random sequence. The partial sum sequence $S_n = \sum_{k=1}^n X_k$ is called a *random walk*. If the random variables X_i take only values 1 and -1, then (S_n) is called a *simple random walk*.

Theorem 2.6. (strong Markov property of a random walk) *Let $S_n = \sum_{k=1}^{n} X_k$ be a random walk and let $T \overset{a.s.}{<} \infty$ be its stopping time. Denote $R_k = S_{T+k} - S_T$ for $k \in \mathbb{N}$. Then $(R_1, R_2, \dots) \overset{d}{=} (S_1, S_2, \dots)$ and the sequence $(R_1, R_2, \dots)$ is independent of σ-algebra $\mathcal{F}_T$.*

Proof. Consider $n \in \mathbb{N}$, $F \in \mathcal{F}_T$ and $B \in \mathcal{B}^n$. Then

$$\begin{aligned}
\mathbb{P}([(R_1, \dots, R_n) \in B] \cap F) &= \sum_{k=1}^{\infty} \mathbb{P}([(R_1, \dots, R_n) \in B, T = k] \cap F) \\
&= \sum_{k=1}^{\infty} \mathbb{P}([(S_{k+1} - S_k, \dots, S_{k+n} - S_k) \in B] \cap [T = k] \cap F) \\
&= \sum_{k=1}^{\infty} \mathbb{P}((S_{k+1} - S_k, \dots, S_{k+n} - S_k) \in B) \cdot \mathbb{P}([T = k] \cap F) \\
&= \mathbb{P}((S_1, \dots, S_n) \in B) \cdot \sum_{k=1}^{\infty} \mathbb{P}([T = k] \cap F) = \mathbb{P}((S_1, \dots, S_n) \in B) \cdot \mathbb{P}(F).
\end{aligned}$$

By choosing $F = \Omega$ we get $\mathbb{P}((R_1, \dots, R_n) \in B) = \mathbb{P}((S_1, \dots, S_n) \in B)$ for all $n \in \mathbb{N}$ and $B \in \mathcal{B}^n$. Applying Theorem 1.7 this in turn means that the distributions of $(R_1, R_2, \dots)$ and $(S_1, S_2, \dots)$ coincide. Furthermore, we have $\mathbb{P}([(R_1, \dots, R_n) \in B] \cap F) = \mathbb{P}((R_1, \dots, R_n) \in B) \cdot \mathbb{P}(F)$. Thus, $(R_1, \dots, R_n)$ and $\mathcal{F}_T$ are independent for any $n \in \mathbb{N}$. This is equivalent to the independence of $(R_1, R_2, \dots)$ and $\mathcal{F}_T$. □

Proposition 2.7. (stationarity with respect to a stopping time) *Let $(X_1, X_2, \dots)$ be an iid random sequence and let $T \overset{a.s.}{<} \infty$ be its stopping time. Then $(X_{T+1}, X_{T+2}, \dots) \overset{d}{=} (X_1, X_2, \dots)$ and the sequence $(X_{T+1}, X_{T+2}, \dots)$ is independent of σ-algebra $\mathcal{F}_T$.*

Proof. Analogously as in the proof of Theorem 2.6 consider arbitrary $n \in \mathbb{N}$, $F \in \mathcal{F}_T$ and $B \in \mathcal{B}^n$. Then

$$\begin{aligned}
\mathbb{P}([(X_{T+1}, \dots, X_{T+n}) \in B] \cap F) &= \sum_{k=1}^{\infty} \mathbb{P}([(X_{T+1}, \dots, X_{T+n}) \in B] \cap F \cap [T = k]) \\
&= \sum_{k=1}^{\infty} \mathbb{P}([(X_{k+1}, \dots, X_{k+n}) \in B]) \mathbb{P}(F \cap [T = k]) \\
&= \mathbb{P}([(X_1, \dots, X_n) \in B]) \sum_{k=1}^{\infty} \mathbb{P}(F \cap [T = k]) \\
&= \mathbb{P}([(X_1, \dots, X_n) \in B]) \mathbb{P}(F).
\end{aligned}$$

By choosing $F = \Omega$ we get $\mathbb{P}([(X_{T+1}, \dots, X_{T+n}) \in B]) = \mathbb{P}([(X_1, \dots, X_n) \in B])$. Hence,

$$\mathbb{P}([(X_{T+1}, \dots, X_{T+n}) \in B] \cap F) = \mathbb{P}([(X_{T+1}, \dots, X_{T+n}) \in B]) \mathbb{P}(F).$$

□

Definition 2.6. Let $X_1, X_2, \dots$ be an iid random sequence such that $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. The corresponding simple random walk (S_n) is called the *symmetric simple random walk.*

Proposition 2.8. *Consider a symmetric simple random walk (S_n) associated with an iid random sequence $(X_1, X_2, \dots)$. Let $T \overset{a.s.}{<} \infty$ be a stopping time of this sequence. Then $(X_1, \dots, X_T, -X_{T+1}, -X_{T+2}, \dots) \overset{d}{=} (X_1, X_2, \dots)$.*

Proof. The random sequence $(X_1, \dots, X_T, 0, \dots)$ is $\mathcal{F}_T$-measurable because for any $B \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ and $n \in \mathbb{N}$ we have

$$[(X_1, \dots, X_T, 0, \dots) \in B] \cap [T = n] = [(X_1, \dots, X_n, 0, \dots) \in B] \cap [T = n] \in \mathcal{F}_n,$$

which implies $[(X_1, \dots, X_T, 0, \dots) \in B] \in \mathcal{F}_T$ by Proposition 2.3. Random sequences

$$(0, \dots, 0, X_{T+1}, X_{T+2}, \dots) \quad \text{and} \quad (0, \dots, 0, -X_{T+1}, -X_{T+2}, \dots)$$

have the same distribution and they are independent of $\mathcal{F}_T$ by Proposition 2.7. Therefore, the random sequences

$$(X_1, X_2, \dots) = (X_1, \dots, X_T, 0, \dots) + (0, \dots, 0, X_{T+1}, X_{T+2}, \dots)$$

and

$$(X_1, \dots, X_T, -X_{T+1}, -X_{T+2}, \dots) = (X_1, \dots, X_T, 0, \dots) + (0, \dots, 0, -X_{T+1}, -X_{T+2}, \dots),$$

that are given as the sums of two independent sequences, have the same distribution. □

Proposition 2.9. (reflection principle) *Let (S_n) be a symmetric simple random walk. Let T be a first hitting time of the set $\{a\}$ (for some $a \in \mathbb{N}$) by this random walk. Denote $S_k^r = 2S_{k\wedge T} - S_k$, $k \in \mathbb{N}$. Then $(S_1^r, S_2^r, \dots)$ has the same distribution as $(S_1, S_2, \dots)$.*

Proof. Exercise class. □

Proposition 2.10. (maxima of symmetric simple random walk) *For symmetric simple random walk (S_n) denote $M_n = \max_{k=1,\dots,n} S_k$, $n \in \mathbb{N}$. Let T be a first hitting time of the set $\{a\}$ (for some $a \in \mathbb{N}$) by the random walk (S_n). Then*

$$\mathbb{P}(T \le n) = \mathbb{P}(M_n \ge a) = 2\mathbb{P}(S_n \ge a) - \mathbb{P}(S_n = a) \quad a \quad \lim_{n\to\infty} \mathbb{P}(M_n \ge a) = 1.$$

Proof. Exercise class. □

Definition 2.7. Let $H : (\Omega, \mathcal{F}) \to (E, \mathcal{B}(E))$ be a random element with values in E. We define a *σ-algebra generated by H* as $\sigma(H) = \{[H \in B], B \in \mathcal{B}(E)\}$. It is the smallest sub-σ-algebra $\mathcal{A} \subseteq \mathcal{F}$ such that $H : (\Omega, \mathcal{A}) \to (E, \mathcal{B}(E))$.

Definition 2.8. Consider $H : (\Omega, \mathcal{F}) \to (E, \mathcal{B}(E))$ and $T : \Omega \to \bar{\mathbb{R}}$. We say that T is *H-measurable random variable* if $T : (\Omega, \sigma(H)) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$, i.e. $\sigma(T) \subseteq \sigma(H)$.

Proposition 2.11. *A random variable T is H-measurable if and only if there exists $f : (E, \mathcal{B}(E)) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ such that $T = f(H)$.*

Proof. See Probability Theory 1. □

If T is a stopping time, then it follows that $\sigma(T) \subset \mathcal{F}_T$. For an example when this inclusion is sharp, consider $T = n$ for some $n \in \mathbb{N}$ such that $\mathcal{F}_n$ is non-trivial σ-algebra. Then $\sigma(T) = \{\emptyset, \Omega\} \subset \mathcal{F}_n = \mathcal{F}_T$.

Before we get to the definition of martingale, let us recall the definition and basic properties of conditional expectation. We write $X \in L_1(\mathcal{F})$ for a random variable X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and satisfying $\mathbb{E}|X| < \infty$. For a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ we denote by $Y \in L_1(\mathcal{G})$ a random variable Y on $(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$ satisfying $\mathbb{E}|Y| < \infty$.

Definition 2.9. Let $X \in L_1(\mathcal{F})$ and $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. A random variable $\mathbb{E}^{\mathcal{G}} X = \mathbb{E}[X|\mathcal{G}] \in L_1(\mathcal{G})$ is called *conditional expectation of X given $\mathcal{G}$* if for any $G \in \mathcal{G}$ we have

$$\int_G X \,\mathrm{d}\mathbb{P} = \int_G \mathbb{E}^{\mathcal{G}} X \,\mathrm{d}\mathbb{P}.$$

The conditional expectation $\mathbb{E}^{\mathcal{G}} X$ is $\mathbb{P}$-a.s. uniquely determined. We write $X \overset{a.s.}{=} Y$ if $\mathbb{P}(X = Y) = 1$ and $X \overset{a.s.}{\le} Y$ if $\mathbb{P}(X \le Y) = 1$.

Proposition 2.12. (calculus for conditional expectation) *For $X, Y \in L_1(\mathcal{F})$ and sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ the following relations hold:*

a) $\mathbb{E}^{\mathcal{G}}(aX + bY + c) \overset{a.s.}{=} a\mathbb{E}^{\mathcal{G}} X + b\mathbb{E}^{\mathcal{G}} Y + c$ *for* $a, b, c \in \mathbb{R}$,

b) $X \overset{a.s.}{\le} Y \Rightarrow \mathbb{E}^{\mathcal{G}} X \overset{a.s.}{\le} \mathbb{E}^{G} Y$,

c) $h : \mathbb{R} \to \mathbb{R}$ *convex,* $h(X) \in L_1(\mathcal{F}) \Rightarrow h(\mathbb{E}^{\mathcal{G}} X) \overset{a.s.}{\le} \mathbb{E}^{\mathcal{G}} h(X)$, *d)* Y *$\mathcal{G}$-measurable random variable,* $X \cdot Y \in L_1(\mathcal{F}) \Rightarrow \mathbb{E}^{\mathcal{G}} XY \overset{a.s.}{=} Y \cdot \mathbb{E}^{\mathcal{G}} X$ *(in particular,* X *$\mathcal{G}$-measurable* $\Rightarrow \mathbb{E}^{\mathcal{G}} X \overset{a.s.}{=} X$*),*

e) $\mathcal{D} \subseteq \mathcal{F}$ *sub-σ-algebra such that* $\mathcal{D}$ *and* $\sigma(X) \vee \mathcal{G}$ *are independent* $\Rightarrow \mathbb{E}[X|\mathcal{G} \vee \mathcal{D}] \overset{a.s.}{=} \mathbb{E}[X|\mathcal{G}]$ *(notation:* $\mathcal{A} \vee \mathcal{B} = \sigma(\mathcal{A} \cup \mathcal{B})$*),*

f) $\mathcal{D} \subseteq \mathcal{G}$ *sub-σ-algebra* $\Rightarrow \mathbb{E}^{\mathcal{D}} \mathbb{E}^{\mathcal{G}} X \overset{a.s.}{=} \mathbb{E}^{\mathcal{G}} \mathbb{E}^{\mathcal{D}} X \overset{a.s.}{=} \mathbb{E}^{\mathcal{D}} X$ *(in particular,* $\mathbb{E}(\mathbb{E}^{\mathcal{G}} X) = \mathbb{E} X$*),*

g) $(X, I_G) \overset{d}{=} (Y, I_G)\ \forall G \in \mathcal{G} \Rightarrow \mathbb{E}^{\mathcal{G}} X \overset{a.s.}{=} \mathbb{E}^{\mathcal{G}} Y$.

Proof. See Probability Theory 1. □

We may also consider conditioning given the random element H with values in a metric space E. In particular, $H = (H_1, H_2, \dots)$ can be a random sequence. If $X \in L_1(\mathcal{F})$ then $\mathbb{E}[X|H] = \mathbb{E}[X|\sigma(H)]$ denotes the conditional expectation of X given H. It is a.s. uniquely determined by the conditions that $\mathbb{E}[X|H]$ is integrable H-measurable random variable and

$$\int_{[H \in B]} X \,\mathrm{d}\mathbb{P} = \int_{[H \in B]} \mathbb{E}[X|H] \,\mathrm{d}\mathbb{P}$$

for all $B \in \mathcal{B}(E)$. According to Proposition 2.11 there exists a Borel measurable function $f : E \to \mathbb{R}$ such that $\mathbb{E}[X|H] = f(H)$. By $f(h) = \mathbb{E}[X|H = h]$ we denote the conditional expectation of X given $H = h$.

The following two properties will later play an important role.

Proposition 2.13. *Let* X *and* Y *be random elements with values in metric spaces* E_1 *and* E_2*, respectively. Let* $g : E_1 \times E_2 \to \mathbb{R}$ *be a Borel measurable function such that* $g(X, Y) \in L_1(\mathcal{F})$*.*

(i) If $Z \in L_1(\mathcal{F})$ *and* (Y, Z) *and* X *are independent, then* $\mathbb{E}[Z|X, Y] \overset{a.s.}{=} \mathbb{E}[Z|Y]$*.*

(ii) If X *and* Y *are independent, then* $\mathbb{E}[g(X, Y)|X = x] = \mathbb{E} g(x, Y)$ *for* P_X*-a.s.* $x \in E_1$*.*

Proof. See Probability Theory 1. □

When studying martingale differences we will need the following result.

Proposition 2.14. *For* $X \in L_2(\mathcal{F})$ *and σ-algebra* $\mathcal{G} \subset \mathcal{F}$ *we have*

(i) $\mathbb{E}^{\mathcal{G}} X \in L_2(\mathcal{G})$, $\mathbb{E}(X - \mathbb{E}^{\mathcal{G}} X)^2 = \mathbb{E} X^2 - \mathbb{E}(\mathbb{E}^{\mathcal{G}} X)^2$,

(ii) $\mathbb{E}(X - \mathbb{E}^{\mathcal{G}} X) Y = 0$ *for* $Y \in L_2(\mathcal{G})$,

(iii) $\mathbb{E}(X - \mathbb{E}^{\mathcal{G}} X)^2 = \min_{Y \in L_2(\mathcal{G})} \mathbb{E}(X - Y)^2$.

Proof. See Probability Theory 1. □

The mapping $\mathbb{E}^{\mathcal{G}} : L_2(\mathcal{F}) \to L_2(\mathcal{G})$ is a projection operator in the Hilbert space L_2. If we denote the L_2-norm $\|X\| = \sqrt{\mathbb{E} X^2}$, then $X - Y = (X - \mathbb{E}^{\mathcal{G}} X) + (\mathbb{E}^{\mathcal{G}} X - Y)$ is the decomposition into two perpendicular summands by part (ii). Moreover,

$$\|X - Y\|^2 = \|X - \mathbb{E}^{\mathcal{G}} X\|^2 + \|\mathbb{E}^{\mathcal{G}} X - Y\|^2 \geq \|X - \mathbb{E}^{\mathcal{G}} X\|^2,$$

and the equality holds for $Y = \mathbb{E}^{\mathcal{G}} X$.

We will substantially improve rule f) from Proposition 2.12 for itereated conditioning.

Proposition 2.15. *Let* $(\mathcal{F}_n)$ *be a filtration and let* S *and* T *be its stopping times. Assume that* $Z \in L_1(\mathcal{F})$*. Then*

(i) the implication $S \leq T \Rightarrow \mathbb{E}^{\mathcal{F}_S} Z = \mathbb{E}^{\mathcal{F}_{S \wedge T}} Z$ *holds a.s.,*

(ii) $\mathbb{E}^{\mathcal{F}_S} \mathbb{E}^{\mathcal{F}_T} Z \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_T} \mathbb{E}^{\mathcal{F}_S} Z \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_{S \wedge T}} Z$.

Proof. (i) We have to prove that there exists $N \in \mathcal{F}$ with property $\mathbb{P}(N) = 0$ so that $S(\omega) \leq T(\omega) \Rightarrow (\mathbb{E}^{\mathcal{F}_S} Z)(\omega) = (\mathbb{E}^{\mathcal{F}_{S\wedge T}} Z)(\omega)$ for $\omega \notin N$. In another words, we want to show that $\mathbf{1}_{[S\leq T]}\mathbb{E}^{\mathcal{F}_S} Z \overset{a.s.}{=} \mathbf{1}_{[S\leq T]}\mathbb{E}^{\mathcal{F}_{S\wedge T}} Z$. By Proposition 2.4 we have $\mathbf{1}_{[S\leq T]} \in \mathcal{F}_{S\wedge T} \subseteq \mathcal{F}_S$. Hence, taking into account Proposition 2.12d we have to verify $\mathbb{E}^{\mathcal{F}_S}\mathbf{1}_{[S\leq T]} Z \overset{a.s.}{=} \mathbf{1}_{[S\leq T]}\mathbb{E}^{\mathcal{F}_{S\wedge T}} Z$. This means that we have to show

$$\int_F \mathbf{1}_{[S\leq T]} Z \,\mathrm{d}\mathbb{P} = \int_F \mathbf{1}_{[S\leq T]}\mathbb{E}^{\mathcal{F}_{S\wedge T}} Z \,\mathrm{d}\mathbb{P}$$

for all $F \in \mathcal{F}_S$. The last relation can be rewritten as

$$\int_{F\cap[S\leq T]} Z \,\mathrm{d}\mathbb{P} = \int_{F\cap[S\leq T]} \mathbb{E}^{\mathcal{F}_{S\wedge T}} Z \,\mathrm{d}\mathbb{P}. \tag{4}$$

Now it suffices to note that $F \cap [S \leq T] \in \mathcal{F}_T \cap \mathcal{F}_S = \mathcal{F}_{S\wedge T}$ by Proposition 2.4. Therefore, (4) follows from the definition of conditional expectation.

(ii) We have to show that $\mathbb{E}^{\mathcal{F}_{S\wedge T}} Z \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_S}\left(\mathbb{E}^{\mathcal{F}_T} Z\right)$, i.e. $\int_F \mathbb{E}^{\mathcal{F}_{S\wedge T}} Z \,\mathrm{d}\mathbb{P} = \int_F \mathbb{E}^{\mathcal{F}_T} Z \,\mathrm{d}\mathbb{P}$ for all $F \in \mathcal{F}_S$. For arbitrary $F \in \mathcal{F}_S$ we get

$$\int_{F\cap[S\leq T]} \mathbb{E}^{\mathcal{F}_{S\wedge T}} Z \,\mathrm{d}\mathbb{P} = \int_{F\cap[S\leq T]} Z \,\mathrm{d}\mathbb{P} = \int_{F\cap[S\leq T]} \mathbb{E}^{\mathcal{F}_T} Z \,\mathrm{d}\mathbb{P}.$$

The first equality is (4) and the second equality follows from $F\cap[S \leq T] \in \mathcal{F}_T$ (see Proposition 2.4d) and the definition of conditional expectation. Similarly as in (i) we can show that $\mathbf{1}_{[T<S]}\mathbb{E}^{\mathcal{F}_{S\wedge T}} Z \overset{a.s.}{=} \mathbf{1}_{[T<S]}\mathbb{E}^{\mathcal{F}_T} Z$. Consequently,

$$\int_{F\cap[T<S]} \mathbb{E}^{\mathcal{F}_{S\wedge T}} Z \,\mathrm{d}\mathbb{P} = \int_{F\cap[T<S]} \mathbb{E}^{\mathcal{F}_T} Z \,\mathrm{d}\mathbb{P}.$$

□

The continuity of $\mathbb{E}[X \mid \mathcal{G}]$ in both arguments plays an important role.

Proposition 2.16. *Let $X_n, X \in L_1(\mathcal{F})$ and let $\mathcal{G}_n, \mathcal{G}$ be sub-σ-algebras of $\mathcal{F}$.*

a) (continuity in L_1): $\mathbb{E}|X_n - X| \underset{n\to\infty}{\longrightarrow} 0 \Rightarrow \mathbb{E}|\mathbb{E}^{\mathcal{G}} X_n - \mathbb{E}^{\mathcal{G}} X| \underset{n\to\infty}{\longrightarrow} 0$,

b) (continuity in L_2): $\mathbb{E}|X_n - X|^2 \underset{n\to\infty}{\longrightarrow} 0 \Rightarrow \mathbb{E}|\mathbb{E}^{\mathcal{G}} X_n - \mathbb{E}^{\mathcal{G}} X|^2 \underset{n\to\infty}{\longrightarrow} 0$,

c) (uniform integrability): if the sequence (X_n) is uniformly integrable, then also the sequence $(\mathbb{E}[X_n \mid \mathcal{G}_n])$ is uniformly integrable,

d) (monotone convergence theorem): $0 \leq X_n \nearrow X$ a.s. $\Rightarrow 0 \leq \mathbb{E}^{\mathcal{G}} X_n \nearrow \mathbb{E}^{\mathcal{G}} X$ a.s.,

e) (conditional Fatou's lemma): $X_n \geq 0$, $X = \liminf_{n\to\infty} X_n \in L_1 \Rightarrow 0 \overset{a.s.}{\leq} \mathbb{E}^{\mathcal{G}} X \overset{a.s.}{\leq} \liminf_{n\to\infty} \mathbb{E}^{\mathcal{G}} X_n$,

f) (dominated convergence theorem): $|X_n| \leq Y \in L_1(\mathcal{F})$, $X_n \overset{\text{a.s.}}{\underset{n\to\infty}{\longrightarrow}} X \Rightarrow \mathbb{E}|\mathbb{E}^{\mathcal{G}} X_n - \mathbb{E}^{\mathcal{G}} X| \underset{n\to\infty}{\longrightarrow} 0$ and $\mathbb{E}^{\mathcal{G}} X_n \overset{\text{a.s.}}{\underset{n\to\infty}{\longrightarrow}} \mathbb{E}^{\mathcal{G}} X$.

Proof. a) Jensen's inequality (Proposition 2.12c) implies $|\mathbb{E}^{G} X_n - \mathbb{E}^{\mathcal{G}} X| \leq \mathbb{E}^{\mathcal{G}}|X_n - X|$. Now it suffices to take expectation on both sides of the inequality.

b) Again by Jensen's inequality we have $|\mathbb{E}^{\mathcal{G}} X_n - \mathbb{E}^{G} X|^2 \leq \mathbb{E}^{\mathcal{G}}|X_n - X|^2$, and taking the expectation gives the continuity in L_2.

c) Denote $Y_n = \mathbb{E}[X_n \mid \mathcal{G}_n]$. Then Jensen's inequality provides $|Y_n| \leq \mathbb{E}^{\mathcal{G}_n}|X_n|$ and for the probability of a $\mathcal{G}_n$-measurable event $[|Y_n| \geq c]$ we get

$$\mathbb{P}(|Y_n| \geq c) \leq c^{-1}\int_{[|Y_n|\geq c]} |Y_n| \,\mathrm{d}\mathbb{P} \leq c^{-1}\int_{[|Y_n|\geq c]} \mathbb{E}^{\mathcal{G}_n}|X_n| \,\mathrm{d}\mathbb{P} \leq c^{-1}\sup_{n\in\mathbb{N}} \mathbb{E}|X_n|.$$

Hence, $\mathbb{P}(|Y_n| \geq c) \underset{c\to\infty}{\longrightarrow} 0$ uniformly in n. Furthermore,

$$\int_{[|Y_n|\geq c]} |Y_n| \,\mathrm{d}\mathbb{P} \leq \int_{[|Y_n|\geq c]} \mathbb{E}^{\mathcal{G}_n}|X_n| \,\mathrm{d}\mathbb{P} = \int_{[|Y_n|\geq c]} |X_n| \,\mathrm{d}\mathbb{P}.$$

Since the X_n have uniformly absolutely continuous integrals, the right-hand side goes to zero for $c \to \infty$ uniformly in n.

d) See Probability Theory 1.

e) By applying part d) for monotone sequence $0 \leq \inf_{k\geq n} X_k \nearrow X$ a.s. we get $0 \leq \mathbb{E}^{\mathcal{G}} \inf_{k\geq n} X_k \nearrow \mathbb{E}^{\mathcal{G}} X$ a.s. Since $\mathbb{E}^{\mathcal{G}} \inf_{k\geq n} X_k \overset{a.s.}{\leq} \inf_{k\geq n} \mathbb{E}^{\mathcal{G}} X_k$, we have $\mathbb{E}^{\mathcal{G}} X \overset{a.s.}{\leq} \liminf_{n\to\infty} \mathbb{E}^{\mathcal{G}} X_n$.

f) The sequences $(Y \pm X_n)$ are non-negative. By e) we know that

$$\mathbb{E}^{\mathcal{G}}(Y \pm X) \overset{a.s.}{\leq} \mathbb{E}^{\mathcal{G}} Y + \liminf_{n\to\infty} \mathbb{E}^{\mathcal{G}}(\pm X_n).$$

After subtraction of $\mathbb{E}^{\mathcal{G}} Y$ we get $\pm\mathbb{E}^{\mathcal{G}} X \overset{a.s.}{\leq} \liminf_{n\to\infty} \pm\mathbb{E}^{\mathcal{G}} X_n$, i.e. $\mathbb{E}^{\mathcal{G}} X \overset{a.s.}{\leq} \liminf_{n\to\infty} \mathbb{E}^{\mathcal{G}} X_n$ and $\mathbb{E}^{\mathcal{G}} X \overset{a.s.}{\geq} \limsup_{n\to\infty} \mathbb{E}^{\mathcal{G}} X_n$. Altogether,

$$\mathbb{E}^{\mathcal{G}} X \overset{a.s.}{\leq} \liminf_{n\to\infty} \mathbb{E}^{\mathcal{G}} X_n \leq \limsup_{n\to\infty} \mathbb{E}^{\mathcal{G}} X_n \overset{a.s.}{\leq} \mathbb{E}^{\mathcal{G}} X$$

and all the inequality signs $\leq$ are in fact equality signs $=$. So we have proved have $\mathbb{E}^{\mathcal{G}} X_n \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}^{\mathcal{G}} X$.

The uniform integrability of (X_n) and a.s. convergence imply convergence in L_1. So the L_1 convergence of conditional expectations follows from a).

□

Now we are ready to define martingales.

Definition 2.10. Let $(\mathcal{F}_n)$ be a filtration and let $X = (X_1, X_2, \dots)$ be a sequence of integrable random variables. We say that X is an $(\mathcal{F}_n)$-*martingale* if it is $\mathcal{F}_n$-adapted and

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \overset{a.s.}{=} X_n \quad \text{for all } n \in \mathbb{N}. \tag{5}$$

In a particular case of canonical filtration $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, X is simply called a *martingale*. It satisfies

$$\mathbb{E}[X_{n+1} \mid X_1, X_2, \dots, X_n] \overset{a.s.}{=} X_n \quad \text{for all } n \in \mathbb{N}. \tag{6}$$

If $\overset{a.s.}{=}$ in (5) and (6) is replaced by $\overset{a.s.}{\geq}$, we say that X is an $(\mathcal{F}_n)$-*submartingale* and *submartingale*, respectively.

If $\overset{a.s.}{=}$ in (5) and (6) is replaced by $\overset{a.s.}{\leq}$, we say that X is an $(\mathcal{F}_n)$-*supermartingale* and *supermartingale*, respectively.

From definition it is clear that every martingale has constant expectation. The sequence $(\mathbb{E}X_n)$ is non-decreasing for a submartingale while it is non-increasing for a supermartingale.

Note that

$$(5) \iff \mathbb{E}[X_n \mid \mathcal{F}_k] \overset{a.s.}{=} X_k \quad \text{for } k \leq n,$$
$$(6) \iff \mathbb{E}[X_n \mid X_1, \dots, X_k] \overset{a.s.}{=} X_k \quad \text{for } k \leq n.$$

It is enough to use Proposition 2.12f):

$$\mathbb{E}[X_n \mid \mathcal{F}_k] \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_k}\mathbb{E}^{\mathcal{F}_{k+1}} \cdots \mathbb{E}^{\mathcal{F}_{n-1}} X_n \overset{a.s.}{=} X_k.$$

Similar equivalences hold for submartingales and supermartingales.

Proposition 2.17. *(stability of martingale property)* *(i) If a random sequence $X_1, X_2, \dots$ is an $\mathcal{F}_n$-martingale, then it is also an $\mathcal{G}_n$-martingale for any filtration $(\mathcal{G}_n)$ satisfying $\sigma(X_1, \dots, X_n) \subseteq \mathcal{G}_n \subseteq \mathcal{F}_n$ for all $n \in \mathbb{N}$. In particular, each $\mathcal{F}_n$-martingale is a martingale.*

(ii) Let $X_1, X_2, \dots$ be an $\mathcal{F}_n$-martingale and let $\mathcal{D}$ be a σ-algebra that is independent with $\mathcal{F}_\infty$. Then $X_1, X_2, \dots$ is an $(\mathcal{F}_n \vee \mathcal{D})$-martingale.

Proof. (i) By Proposition 2.12f) we have $\mathbb{E}[X_{n+1} \mid \mathcal{G}_n] \overset{a.s.}{=} \mathbb{E}^{\mathcal{G}_n}\mathbb{E}^{\mathcal{F}_n} X_{n+1} \overset{a.s.}{=} \mathbb{E}^{\mathcal{G}_n} X_n \overset{a.s.}{=} X_n$.

(ii) By Proposition 2.12e) we have $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n \vee \mathcal{D}] \overset{a.s.}{=} \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \overset{a.s.}{=} X_n$.

□

Remark: Similar results are true for submartingales and supermartingales.

The fundamental examples of martingales are provided by sums or products of independent random variables.

Proposition 2.18. *Let $X_1, X_2, \dots$ be a sequence of independent integrable random variables. Denote $S_n = \sum_{j=1}^n X_j$ for $n \in \mathbb{N}$.*

a) If $\mathbb{E}X_n = 0$, then (S_n) is a martingale. If $\mathbb{E}X_n \geq 0$, then (S_n) is a submartingale. If $\mathbb{E}X_n \leq 0$, then (S_n) is a supermartingale.

b) If $X_n \in L_2$, $\mathbb{E}X_n = 0$ and $\mathbb{E}X_n^2 = \sigma^2$, then $M_n = S_n^2 - n\sigma^2$ is a martingale.

c) If $\mathbb{E}X_n = 1$, then $Z_n = \prod_{j=1}^n X_j$ is a martingale.

d) If $\mathbb{P}(X_n = -1) = q$ and $\mathbb{P}(X_n = 1) = p$, where $q = 1 - p$ and $p \in (0, 1)$, then $Y_n = (q/p)^{S_n}$ is a martingale.

Proof. a) Obviously, $S_n \in L_1$. Therefore, it suffices to realize that

$$\mathbb{E}[S_{n+1} \mid S_1, \dots, S_n] \overset{a.s.}{=} \mathbb{E}[S_{n+1} \mid X_1, \dots, X_n] \overset{a.s.}{=} \mathbb{E}[S_n + X_{n+1} \mid X_1, \dots, X_n] \overset{a.s.}{=} S_n + \mathbb{E}X_{n+1}.$$

b) The integrability of M_n follows from the assumption $X_n \in L_2$. Furthermore,

$$\mathbb{E}[S_{n+1}^2 \mid S_1, \dots, S_n] \overset{a.s.}{=} \mathbb{E}[(S_n + X_{n+1})^2 \mid X_1, \dots, X_n] \overset{a.s.}{=} S_n^2 + 2S_n\mathbb{E}X_{n+1} + \mathbb{E}X_{n+1}^2 = S_n^2 + \sigma^2,$$

which yields

$$\mathbb{E}[M_{n+1} \mid S_1, \dots, S_n] \overset{a.s.}{=} \mathbb{E}[S_{n+1}^2 \mid S_1, \dots, S_n] - (n+1)\sigma^2 \overset{a.s.}{=} S_n^2 - n\sigma^2 = M_n,$$

i.e. M_n is a $\sigma(S_1, \dots, S_n)$-martingale. Since

$$\sigma(M_1, \dots, M_n) = \sigma(S_1^2, \dots, S_n^2) \subseteq \sigma(S_1, \dots, S_n) = \sigma(X_1, \dots, X_n),$$

we conclude that (M_n) is also a martingale by Proposition 2.17.

c), d) Exercise class.

□

Now we prove the result about submartingales.

Proposition 2.19. *(i) Let $X_1, X_2, \dots$ be an $\mathcal{F}_n$-martingale and let $g : \mathbb{R} \to \mathbb{R}$ be a convex function such that $g(X_n) \in L_1$ for any $n \in \mathbb{N}$. Then $g(X_1), g(X_2), \dots$ is an $\mathcal{F}_n$-submartingale.*

(ii) If $X_1, X_2, \dots$ is an $\mathcal{F}_n$-submartingale and $g : \mathbb{R} \to \mathbb{R}$ is a convex and non-decreasing function such that $g(X_n) \in L_1$ for any $n \in \mathbb{N}$. Then $g(X_1), g(X_2), \dots$ is an $\mathcal{F}_n$-submartingale.

Proof. By our assumptions, $(g(X_n))$ is an $\mathcal{F}_n$-adapted sequence of integrable random variables. From Jensen's inequality we have

$$\mathbb{E}[g(X_{n+1}) \mid \mathcal{F}_n] \overset{a.s.}{\geq} g(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]).$$

The right-hand side is a.s. equal to $g(X_n)$ in case (i) due to the martingale property (5) and it is a.s. greater or equal to $g(X_n)$ in case (ii) due to the submartingale property and monotonicity of g. □

Remark: In particular, (X_n^+) is a submartingale if (X_n) is submartingale and $(|Y_n|^p)$ for $p \geq 1$ is a submartingale if (Y_n) is a martingale.

From Proposition 2.18a we know that a random walk (S_n) with centred steps is a martingale. Therefore, (S_n^2) is a submartingale and by Proposition 2.18b it can be decomposed into a martingale and an increasing sequence: $S_n^2 = M_n + n\sigma^2$. It is possible to make a similar decomposition for any submartingale.

Definition 2.11. Let $(\mathcal{F}_n)$ be a filtration. The random sequence $I_1, I_2, \ldots$ is $\mathcal{F}_n$*-predictable* if I_n is $\mathcal{F}_{n-1}$-measurable for all $n \in \mathbb{N}$, where we put $\mathcal{F}_0 = \{\emptyset, \Omega\}$, i.e. I_1 is constant.

Remark: Every $\mathcal{F}_n$-predictable $\mathcal{F}_n$-martingale (M_n) is constant a.s. because it must satisfy $M_n \overset{a.s.}{=} \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \overset{a.s.}{=} M_{n+1}$.

Theorem 2.20. (Doob decomposition theorem) *Let (S_n) be an $\mathcal{F}_n$-submartingale. Then there exists an $\mathcal{F}_n$-martingale (M_n) and a non-decreasing $\mathcal{F}_n$-predictable sequence (I_n) so that $S_n = M_n + I_n$, $n \in \mathbb{N}$. The summands M_n and I_n are a.s. uniquely determined under the additional condition $I_1 = 0$.*

Proof. Let (D_n) be a sequence of differences of (S_n), i.e. $D_1 = S_1$ and $D_{n+1} = S_{n+1} - S_n$ for $n \in \mathbb{N}$. The submartingale property immediately implies $\mathbb{E}^{\mathcal{F}_n} D_{n+1} \overset{a.s.}{\geq} 0$. Put $Z_1 = 0$ and $Z_{n+1} = (\mathbb{E}^{\mathcal{F}_n} D_{n+1})^+$ for $n \in \mathbb{N}$. Then $Z_{n+1} \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_n} D_{n+1}$ and (Z_n) is $\mathcal{F}_n$-predictable sequence. Furthermore, we define $Y_n = D_n - Z_n$, $n \in \mathbb{N}$. Now we proceed to the cumulative sum and introduce

$$M_n = \sum_{k=1}^n Y_k, \quad I_n = \sum_{k=1}^n Z_k, \quad S_n = \sum_{k=1}^n D_k = \sum_{k=1}^n Y_k + \sum_{k=1}^n Z_k = M_n + I_n, \quad n \in \mathbb{N}.$$

We know that $I_1 = Z_1 = 0$ and $I_{n+1} = \sum_{k=1}^{n+1} Z_k$ is $\mathcal{F}_n$-measurable for $n \in \mathbb{N}$. Hence, (I_n) is $\mathcal{F}_n$-predictable sequence that is moreover non-decreasing because $Z_n \geq 0$. The random sequence (M_n) is $\mathcal{F}_n$-adapted as it is a difference of $\mathcal{F}_n$-adapted sequence (S_n) and $\mathcal{F}_n$-predictable sequence (I_n). Clearly, both D_n and Z_n are integrable. Consequently, also Y_n and M_n are integrable. We verify that (M_n) satisfies the martingale property:

$$\mathbb{E}^{\mathcal{F}_n}(M_{n+1} - M_n) = \mathbb{E}^{\mathcal{F}_n} Y_{n+1} \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_n}(D_{n+1} - \mathbb{E}^{\mathcal{F}^n} D_{n+1}) \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_n} D_{n+1} - \mathbb{E}^{\mathcal{F}^n} D_{n+1} = 0,$$

i.e. $\mathbb{E}^{\mathcal{F}_n} M_{n+1} \overset{a.s.}{=} M_n$ and thus (M_n) is an $\mathcal{F}_n$-martingale. So we found the decompoisition $S_n = M_n + I_n$ into an $\mathcal{F}_n$-martingale and a non-decreasing $\mathcal{F}_n$-predictable sequence.

In order to show uniqueness assume that we have two decompositions $S_n = M_n + I_n = N_n + J_n$. Then $\bar{M}_n = M_n - N_n = J_n - I_n$ is both $\mathcal{F}_n$-martingale and $\mathcal{F}_n$-predictable sequence. According to Remark preceding Theorem the sequence $(\bar{M}_n)$ is constant a.s. This constant must be zero due to the condition $I_1 = J_1 = 0$. It means that $M_n \overset{a.s.}{=} N_n$ and $I_n \overset{a.s.}{=} J_n$. □

Definition 2.12. The sequence (I_n) from Doob decomposition theorem is called a *compensator* of a submartingale (S_n).

Proposition 2.21. (Martingale differences of L_2-martingale are orthogonal in L_2) *Let (M_n) be an $\mathcal{F}_n$-martingale such that $M_n \in L_2$ for all $n \in \mathbb{N}$. Denote $D_1 = M_1$ and $D_{n+1} = M_{n+1} - M_n$ for $n \in \mathbb{N}$. Then $\mathbb{E} D_n D_m = 0$ for $m \neq n$, and so $\mathbb{E} M_n^2 = \sum_{j=1}^n \mathbb{E} D_j^2$ and $\operatorname{var} M_n = \sum_{j=1}^n \operatorname{var} D_j$.*

Proof. Exercise class. □

3 Stopping theorems and maximal inequalities

Stopping problem: Let $X_1, X_2, \ldots$ be a martingale and $T_1 \leq T_2 \leq \cdots$ be a sequence of its stopping times. Consider a sequence $X_{T_1}, X_{T_2}, \ldots$ given by values of the martingale stopped at these stopping times. Is it again a martingale?

The answer is positive in the case $T_1 \leq T_2 \leq \cdots \leq K < \infty$. We formulate a corresponding version for two times.

Theorem 3.1. *Let $X_1, X_2, \dots$ be an $\mathcal{F}_n$-martingale and let S, T be $\mathcal{F}_n$-stopping times such that $S \leq T \leq K < \infty$ for some $K \in \mathbb{N}$. Then $X_S, X_T \in L_1$ and*

$$\mathbb{E}^{\mathcal{F}_S} X_T \overset{a.s.}{=} X_S.$$

In particular, $\mathbb{E}X_T = \mathbb{E}X_S$.

If $X_1, X_2, \dots$ is an $\mathcal{F}_n$-submartingale, then $X_S, X_T \in L_1$ and

$$\mathbb{E}^{\mathcal{F}_S} X_T \overset{a.s.}{\geq} X_S.$$

In particular, $\mathbb{E}X_T \geq \mathbb{E}X_S$.

Proof. The integrability of X_S and X_T follows from the simple bound

$$|X_T| \leq \max_{j=1,\dots,K} |X_j| \leq \sum_{j=1}^{K} |X_j| \in L_1.$$

First assume that $T - S \leq 1$. Then

$$\int_F (X_T - X_S) \, \mathrm{d}\mathbb{P} = \sum_{j=1}^{K-1} \int_{F \cap [S=j] \cap [T>j]} (X_{j+1} - X_j) \, \mathrm{d}\mathbb{P}.$$

Since $H_j = F \cap [S = j] \cap [T > j] \in \mathcal{F}_j$ for $F \in \mathcal{F}_S$, we get $\int_{H_j} (X_{j+1} - X_j) \, \mathrm{d}\mathbb{P} = 0$ in case of $\mathcal{F}_n$-martingale (X_n) and $\int_{H_j} (X_{j+1} - X_j) \, \mathrm{d}\mathbb{P} \geq 0$ in case of $\mathcal{F}_n$-submartingale (X_n). Therefore, $\int_F (X_T - X_S) \, \mathrm{d}\mathbb{P} = 0$ for martingale and $\int_F (X_T - X_S) \, \mathrm{d}\mathbb{P} \geq 0$ for submartingale.

In general case we join the times S and T by a finite chain of stopping times $V_j = (S + j) \wedge T$ satisfying $V_{j+1} - V_j \leq 1$ and $S = V_0 \leq V_1 \leq \dots \leq V_K = T$. Now we can iteratively use the already proved result for times that differ by at most 1. Then for submartingale it follows that

$$\mathbb{E}^{\mathcal{F}_S} X_T \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_{V_0}} \mathbb{E}^{\mathcal{F}_{V_1}} \dots \mathbb{E}^{\mathcal{F}_{V_{K-1}}} X_T \overset{a.s.}{\geq} \mathbb{E}^{\mathcal{F}_{V_0}} \mathbb{E}^{\mathcal{F}_{V_1}} \dots \mathbb{E}^{\mathcal{F}_{V_{K-2}}} X_{V_{K-1}} \overset{a.s.}{\geq} \dots \overset{a.s.}{\geq} \mathbb{E}^{\mathcal{F}_{V_0}} X_{V_1} \overset{a.s.}{\geq} X_S.$$

For martingale all inequalities $\overset{a.s.}{\geq}$ are equalities $\overset{a.s.}{=}$. □

Corollary 3.2. (optional stopping theorem) *Let $X = (X_1, X_2, \dots)$ be an $\mathcal{F}_n$-martingale (or $\mathcal{F}_n$-submartingale) and let T be its $\mathcal{F}_n$-stopping time. By stopping X at time T we obtain a random sequence $X^T = (X_{T\wedge 1}, X_{T\wedge 2}, \dots)$. This stopped sequence is an $\mathcal{F}_n$-martingale (or $\mathcal{F}_n$-submartingale).*

Proof. The random variables $X_{T\wedge k}$ are $\mathcal{F}_{T\wedge k}$-measurable (Proposition 2.4a), and so also $\mathcal{F}_k$-measurable ($\mathcal{F}_{T\wedge k} \subseteq \mathcal{F}_k$ by Proposition 2.4e). Therefore, X^T is $\mathcal{F}_n$-adapted sequence. Theorem 3.1 ensures that the $X_{T\wedge k}$ are integrable random variables. It remains to verify martingale (or submartingale) property. Assume that $X_1, X_2, \dots$ is a submartingale. Then by Theorem 3.1 and Proposition 2.15 we have

$$X_{T\wedge n} \overset{a.s.}{\leq} \mathbb{E}^{\mathcal{F}_{T\wedge n}} X_{T\wedge(n+1)} \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_n} \mathbb{E}^{\mathcal{F}_T} X_{T\wedge(n+1)} \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_n} X_{T\wedge(n+1)}.$$

In the last step we used that $X_{T\wedge(n+1)}$ is $\mathcal{F}_T$-measurable random variable by Proposition 2.4. □

Proposition 3.3. *Let $X_1, X_2, \dots$ be an $\mathcal{F}_n$-submartingale and let $S \leq T$ be $\mathcal{F}_n$-stopping times. Then*

$$\mathbb{E}^{\mathcal{F}_S} X_{T\wedge n} \overset{a.s.}{\geq} X_{S\wedge n} \quad \text{for any } n \in \mathbb{N}.$$

Proof. Similar as in the previous proof it follows by Proposition 2.4, Proposition 2.15 and Theorem 3.1 that

$$\mathbb{E}^{\mathcal{F}_S} X_{T\wedge n} \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_S} \mathbb{E}^{\mathcal{F}_n} X_{T\wedge n} \overset{a.s.}{=} \mathbb{E}^{\mathcal{F}_{S\wedge n}} X_{T\wedge n} \overset{a.s.}{\geq} X_{S\wedge n}.$$

□

Example: Let (S_n) be a symmetric simple random walk. Let T be its first hitting time of $\{a\}$, where $a \in \mathbb{N}$. We know that (S_n) is a martingale and T is its stopping time that is a.s. finite. In this case we have $\mathbb{E}S_T = a > 0 = \mathbb{E}S_1$. This example shows that we cannot expect that the result of Theorem 3.1 would be valid for general unbounded stopping time.

Theorem 3.4. *(i) Let $X_1, X_2, \dots$ be an $\mathcal{F}_n$-martingale and let $S \le T \overset{a.s.}{<} \infty$ be $\mathcal{F}_n$-stopping times such that*

$$X_T \in L_1 \quad \text{and} \quad \int_{[T>n]} |X_n| \,\mathrm{d}\mathbb{P} \underset{n\to\infty}{\longrightarrow} 0. \tag{7}$$

Then $\mathbb{E}^{\mathcal{F}_S} X_T \overset{a.s.}{=} X_S$ (and consequently $\mathbb{E}X_T = \mathbb{E}X_S$).

(ii) Let $X_1, X_2, \dots$ be an $\mathcal{F}_n$-submartingale and let $S \le T \overset{a.s.}{<} \infty$ be $\mathcal{F}_n$-stopping times such that

$$X_T^+ \in L_1 \quad \text{and} \quad \int_{[T>n]} X_n^+ \,\mathrm{d}\mathbb{P} \underset{n\to\infty}{\longrightarrow} 0. \tag{8}$$

Then $\mathbb{E}^{\mathcal{F}_S} X_T \overset{a.s.}{\ge} X_S$ (and consequently $\mathbb{E}X_T \ge \mathbb{E}X_S$).

Proof. We are going to prove only (ii), the proof for martingales proceeds analogously.

First we realize that (8) implies $X_T \in L_1$. To see this we show that $X_T^- \in L_1$. From Corollary 3.2 we know that $(X_{T\wedge n})$ is an $\mathcal{F}_n$-submartingale, so it must have non-decreasing expectation. In particular, $\mathbb{E}X_{T\wedge n} \ge \mathbb{E}X_1$. Thus, we obtain the following bound for negative part

$$\mathbb{E}X_{T\wedge n}^- = \mathbb{E}X_{T\wedge n}^+ - \mathbb{E}X_{T\wedge n} \le \mathbb{E}X_{T\wedge n}^+ - \mathbb{E}X_1 = \mathbb{E}X_n^+ \mathbf{1}_{[T>n]} + \mathbb{E}X_T^+ \mathbf{1}_{[T\le n]} - \mathbb{E}X_1 \le \mathbb{E}X_n^+ \mathbf{1}_{[T>n]} + \mathbb{E}X_T^+ - \mathbb{E}X_1.$$

By Fatou's lemma and (8) we have

$$\mathbb{E}X_T^- \le \liminf_{n\to\infty} \mathbb{E}X_{T\wedge n}^- \le \liminf_{n\to\infty} \mathbb{E}X_n^+ \mathbf{1}_{[T>n]} + \mathbb{E}X_T^+ - \mathbb{E}X_1 = \mathbb{E}X_T^+ - \mathbb{E}X_1 < \infty.$$

For a while assume that $X_S \in L_1$. We have to show that

$$\int_F X_T \,\mathrm{d}\mathbb{P} \ge \int_F X_S \,\mathrm{d}\mathbb{P} \quad \text{for} \quad F \in \mathcal{F}_S. \tag{9}$$

Since $F \cap [S \le n] = F \cap [S \le S \wedge n] \in \mathcal{F}_{S\wedge n}$ according to Proposition 2.4d, we get from Theorem 3.1 the following relation

$$\int_{F\cap[S\le n]} X_{T\wedge n} \,\mathrm{d}\mathbb{P} \ge \int_{F\cap[S\le n]} X_{S\wedge n} \,\mathrm{d}\mathbb{P} = \int_{F\cap[S\le n]} X_S \,\mathrm{d}\mathbb{P}.$$

The right-hand side goes to $\int_F X_S \,\mathrm{d}\mathbb{P}$ as $n \to \infty$ because $S \overset{a.s.}{<} \infty$ and $X_S \in L_1$. The left-hand side can be rewritten as

$$\int_{F\cap[S\le n]} X_{T\wedge n} \,\mathrm{d}\mathbb{P} = \int_{F\cap[S\le n]\cap[T\le n]} X_T \,\mathrm{d}\mathbb{P} + \int_{F\cap[S\le n]\cap[T>n]} X_n \,\mathrm{d}\mathbb{P}.$$

Since $S \le T \overset{a.s.}{<} \infty$ and $X_T \in L_1$, the first summand satisfies

$$\int_{F\cap[S\le n]\cap[T\le n]} X_T \,\mathrm{d}\mathbb{P} = \int_{F\cap[T\le n]} X_T \,\mathrm{d}\mathbb{P} \underset{n\to\infty}{\longrightarrow} \int_F X_T \,\mathrm{d}\mathbb{P}.$$

For the second summand we have by (8),

$$\liminf_{n\to\infty} \int_{F\cap[S\le n]\cap[T>n]} X_n \,\mathrm{d}\mathbb{P} \le \liminf_{n\to\infty} \int_{F\cap[S\le n]\cap[T>n]} X_n^+ \,\mathrm{d}\mathbb{P} \le \liminf_{n\to\infty} \int_{[T>n]} X_n^+ \,\mathrm{d}\mathbb{P} \underset{n\to\infty}{\longrightarrow} 0.$$

Altogether,

$$\liminf_{n\to\infty} \int_{F\cap[S\le n]} X_{T\wedge n} \,\mathrm{d}\mathbb{P} \le \int_F X_T \,\mathrm{d}\mathbb{P},$$

which implies

$$\int_F X_T \,\mathrm{d}\mathbb{P} \geq \liminf_{n\to\infty} \int_{F\cap[S\leq n]} X_{T\wedge n} \,\mathrm{d}\mathbb{P} \geq \liminf_{n\to\infty} \int_{F\cap[S\leq n]} X_{S\wedge n} \,\mathrm{d}\mathbb{P} = \int_F X_S \,\mathrm{d}\mathbb{P}.$$

In other words we have shown that $\mathbb{E}^{\mathcal{F}_S} X_T \overset{a.s.}{\geq} X_S$. This relation is true for any $\mathcal{F}_n$-stopping time $S \leq T$ such that $X_S \in L_1$. A particular example is $S = T \wedge k$ for arbitrary $k \in \mathbb{N}$ (integrability of $X_{T\wedge k}$ is assured by Theorem 3.1). Therefore, using Proposition 2.12c we obtain

$$X^+_{T\wedge k} \overset{a.s.}{\leq} \left(\mathbb{E}^{\mathcal{F}_{T\wedge k}} X_T\right)^+ \overset{a.s.}{\leq} \mathbb{E}^{\mathcal{F}_{T\wedge k}} X^+_T. \tag{10}$$

Now we can get rid of assumption $X_S \in L_1$. By noticing that $[S = k] = [S = k] \cap [T \geq k] \in \mathcal{F}_k \cap \mathcal{F}_T = \mathcal{F}_{T\wedge k}$ and applying (10) we can express

$$\begin{aligned}
\mathbb{E}X^+_S &= \sum_{k=1}^{\infty} \mathbb{E}X^+_k \mathbf{1}_{[S=k]} = \sum_{k=1}^{\infty} \mathbb{E}X^+_k \mathbf{1}_{[S=k\leq T]} = \sum_{k=1}^{\infty} \mathbb{E}X^+_{T\wedge k} \mathbf{1}_{[S=k\leq T]} \\
&\leq \sum_{k=1}^{\infty} \mathbb{E}X^+_T \mathbf{1}_{[S=k\leq T]} = \sum_{k=1}^{\infty} \mathbb{E}X^+_T \mathbf{1}_{[S=k]} = \mathbb{E}X^+_T < \infty.
\end{aligned}$$

By the same arguments as in the beginning of the proof we obtain $\mathbb{E}X^-_S < \infty$. Together it gives $X_S \in L_1$ and we have already shown that (9) holds. This completes the proof. □

Proposition 3.5. *The condition* (7) *is equivalent to the condition that the stopped sequence* $X^T = (X_{T\wedge n})$ *is uniformly integrable. Similarly, the condition* (8) *is equivalent to the uniform integrability of* $(X^+_{T\wedge n})$.

Proof. Exercise class. □

It is natural to ask how we can check condition (7) or equivalently the uniform integrability of the stopped sequence at time T. We give several sufficient conditions that are usually much easier to verify.

Theorem 3.6. *Let* $X_1, X_2, \ldots$ *be an* $\mathcal{F}_n$*-martingale and let* $S \leq T \overset{a.s.}{<} \infty$ *be* $\mathcal{F}_n$*-stopping times. Consider the following conditions:*

$$\exists 0 < c < \infty : \quad T \geq n \Rightarrow |X_n| \leq c \quad a.s., \tag{11}$$

i.e. until time T *the trajectory* $X_1, X_2, \ldots$ *lies in the interval* $[-c, c]$ *a.s.;*

$$(\exists 0 < c < \infty : \quad T > n \Rightarrow |X_{n+1} - X_n| \leq c \quad a.s.) \quad and \quad \mathbb{E}T < \infty, \tag{12}$$

i.e. before time T *the increments* $|X_{n+1} - X_n|$ *are uniformly bounded a.s. and* T *is integrable;*

$$\left(\exists 0 < c < \infty : \quad T > n \Rightarrow \mathbb{E}^{\mathcal{F}_n}|X_{n+1} - X_n| \leq c \quad s.j.\right) \quad and \quad \mathbb{E}T < \infty, \tag{13}$$

i.e. before time T *the conditional increments are uniformly bounded a.s. and* T *is integrable. Then any of the conditions* (11)*,* (12) *and* (13) *implies that*

$$X_T \in L_1 \quad and \quad \mathbb{E}^{\mathcal{F}_S} X_T \overset{a.s.}{=} X_S.$$

Proof. The condition (12) implies the condition (13) because

$$\mathbf{1}_{[T>n]}\mathbb{E}^{\mathcal{F}_n}|X_{n+1} - X_n| = \mathbb{E}^{\mathcal{F}_n}\mathbf{1}_{[T>n]}|X_{n+1} - X_n| \leq c \quad \text{a.s.}$$

We are going to verify that both (11) and (13) are sufficient for (7). Then Theorem 3.4 finishes the proof.

Assume that (11) is satisfied. Then

$$\int_{[T>n]} |X_n| \,\mathrm{d}\mathbb{P} \leq c\mathbb{P}(T > n) \underset{n\to\infty}{\longrightarrow} c\mathbb{P}(T = \infty) = 0.$$

The stopped sequence $(X_{T\wedge n})$ is bounded (we have $|X_{T\wedge n}| \overset{a.s.}{\leq} c$) so its limit X_T must be integrable. This follows from Lebesgue's dominated convergence theorem:

$$\mathbb{E}|X_T| = \lim_{n\to\infty} \mathbb{E}|X_{T\wedge n}| \leq c < \infty.$$

Now assume that condition (13) is satisfied. Define $Y_n = |X_1| + \sum_{k=1}^{n-1} |X_{k+1} - X_k|$, $n \in \mathbb{N}$. Then $|X_n| \leq Y_n$ for any $n \in \mathbb{N}$ and

$$0 \leq |X_T| \leq Y_T \overset{a.s.}{=} |X_1| + \sum_{k=1}^{\infty} |X_{k+1} - X_k| \mathbf{1}_{[T>k]}.$$

If we realize that condition (13) implies

$$\mathbb{E}|X_{k+1} - X_k|\mathbf{1}_{[T>k]} = \mathbb{E}(\mathbb{E}^{\mathcal{F}_k}|X_{k+1} - X_k|)\mathbf{1}_{[T>k]} \leq c\mathbb{P}(T > k),$$

we get

$$\mathbb{E}|X_T| \leq \mathbb{E}Y_T \leq \mathbb{E}|X_1| + \sum_{k=1}^{\infty} \mathbb{E}|X_{k+1} - X_k|\mathbf{1}_{[T>k]} \leq \mathbb{E}|X_1| + c\sum_{k=1}^{\infty} \mathbb{P}(T > k) \leq \mathbb{E}|X_1| + c\mathbb{E}T < \infty,$$

that is $X_T \in L_1$. We have also shown that $Y_T \in L_1$, which helps us to verify the second part of (7):

$$\int_{[T>n]} |X_n| \, \mathrm{d}\mathbb{P} \leq \int_{[T>n]} Y_n \, \mathrm{d}\mathbb{P} \leq \int_{[T>n]} Y_T \, \mathrm{d}\mathbb{P} \underset{n\to\infty}{\longrightarrow} 0.$$

□

Remark: We can formulate similar sufficient conditions that ensure (8) for the case of submartingale (X_n).

Remark: In condition (11) we are not allowed to replace $T \geq n$ with $T > n$ (see exercise classes).

In applications we often consider that T is the first exit time from some bounded Borel set. Then (11) is automatically fulfilled.

Theorem 3.7. (optional sampling theorem)

(i) Let $X_1, X_2, \ldots$ be an $\mathcal{F}_n$-martingale and let $T_1 \leq T_2 \leq \cdots \overset{a.s.}{<} \infty$ be $\mathcal{F}_n$-stopping times. If

$$X_{T_k} \in L_1 \quad \textit{and} \quad \lim_{n\to\infty} \int_{[T_k>n]} |X_n| \, \mathrm{d}\mathbb{P} = 0$$

for all $k \in \mathbb{N}$, then $(X_{T_1}, X_{T_2}, \ldots)$ is an $\mathcal{F}_{T_n}$-martingale.

(ii) Let $X_1, X_2, \ldots$ be an $\mathcal{F}_n$-submartingale and let $T_1 \leq T_2 \leq \cdots \overset{a.s.}{<} \infty$ be $\mathcal{F}_n$-stopping times. If

$$X^+_{T_k} \in L_1 \quad \textit{and} \quad \lim_{n\to\infty} \int_{[T_k>n]} X^+_n \, \mathrm{d}\mathbb{P} = 0,$$

for all $k \in \mathbb{N}$, then $(X_{T_1}, X_{T_2}, \ldots)$ is an $\mathcal{F}_{T_n}$-submartingale.

Proof. The sequence (X_{T_n}) is $\mathcal{F}_{T_n}$-adapted by Proposition 2.4a. The integrability of X_{T_n} is either directly assumed in case (i) or it follows from the proof of Theorem 3.4 in case (ii). Martingale or submartingale property is obtained by applying Theorem 3.4 for any $k \in \mathbb{N}$. □

Remark: According to Proposition 3.5 we may equivalently rewrite the conditions in Theorem 3.7 using uniform integrability of the stopped sequences $(X_{T_k\wedge n}, n \in \mathbb{N})$ and $(X^+_{T_k\wedge n}, n \in \mathbb{N})$.

The following theorem provides an important application.

Theorem 3.8. *Let $S_n = \sum_{k=1}^{n} X_k$ be a random walk and let $T \in L_1$ be its stopping time. Then* *a)* $X_1 \in L_1 \Longrightarrow S_T \overset{a.s.}{=} \sum_{k=1}^{T} X_k \in L_1$ *and* $\mathbb{E}S_T = \mathbb{E}T \cdot \mathbb{E}X_1$,

b) if $X_1 \in L_2$, $\mathbb{E}X_1 = 0$ *and* $\exists c \in (0, \infty)$ *such that* $(T > n \Rightarrow |S_n| \leq c$ *a.s.)* $\forall n \in \mathbb{N}$, *then*

$$\operatorname{var} S_T = \mathbb{E}S_T^2 = \mathbb{E}T \cdot \mathbb{E}X_1^2 = \mathbb{E}T \cdot \operatorname{var} X_1.$$

Proof. a) Consider $Y_n = S_n - n\mathbb{E}X_1$, $n \in \mathbb{N}$. It is a martingale by Proposition 2.18a. Furthermore,

$$\mathbb{E}^{\mathcal{F}_n}|Y_{n+1} - Y_n| = \mathbb{E}^{\mathcal{F}_n}|X_{n+1} - \mathbb{E}X_1| = \mathbb{E}|X_1 - \mathbb{E}X_1| = c < \infty,$$

where $\mathcal{F}_n = \sigma(S_1, \ldots, S_n)$. It means that (Y_n) satisfies condition (13) in Theorem 3.6. Therefore,

$$\mathbb{E}Y_T = \mathbb{E}Y_1 = 0 \Longrightarrow \mathbb{E}(S_T - T\mathbb{E}X_1) = 0 \Longrightarrow \mathbb{E}S_T = \mathbb{E}T \cdot \mathbb{E}X_1.$$

b) Define $M_n = S_n^2 - n\mathbb{E}X_1^2$, $n \in \mathbb{N}$. It is a martingale by Proposition 2.18b. Again we verify condition (13):

$$\begin{aligned}\mathbb{E}^{\mathcal{F}_n}|M_{n+1} - M_n| &= \mathbb{E}^{\mathcal{F}_n}|2S_nX_{n+1} + X_{n+1}^2 - \mathbb{E}X_1^2| \\ &\overset{a.s.}{\leq} \mathbb{E}^{\mathcal{F}_n}2|S_n||X_{n+1}| + \mathbb{E}^{\mathcal{F}_n}X_{n+1}^2 + \mathbb{E}X_1^2 \overset{a.s.}{=} 2|S_n|\mathbb{E}|X_1| + 2\mathbb{E}X_1^2.\end{aligned}$$

Hence,

$$\mathbf{1}_{[T>n]}\mathbb{E}^{\mathcal{F}_n}|M_{n+1} - M_n| \overset{a.s.}{\leq} 2\mathbf{1}_{[T>n]}|S_n|\mathbb{E}|X_1| + 2\mathbb{E}X_1^2 \overset{a.s.}{\leq} 2c\mathbb{E}|X_1| + 2\mathbb{E}X_1^2 < \infty.$$

Theorem 3.6 gives

$$0 = \mathbb{E}M_1 = \mathbb{E}M_T = \mathbb{E}(S_T^2 - T\mathbb{E}X_1^2) = \mathbb{E}S_T^2 - \mathbb{E}T \cdot \mathbb{E}X_1^2,$$

which implies $\mathbb{E}S_T^2 = \mathbb{E}T \cdot \mathbb{E}X_1^2$.

□

We notice that Theorem 3.8 has simpler version.

Proposition 3.9. (Wald's equations) *Let* $S_n = \sum_{k=1}^{n} X_k$ *be a random walk and let* $T \in L_1$ *be its stopping time that is independent of* (S_n)*. Then*

a) $X_1 \in L_1 \Longrightarrow \mathbb{E}S_T = \mathbb{E}T \cdot \mathbb{E}X_1$,

b) $X_1 \in L_2$, $\mathbb{E}X_1 = 0 \Longrightarrow \operatorname{var} S_T = \mathbb{E}S_T^2 = \mathbb{E}T \cdot \mathbb{E}X_1^2 = \mathbb{E}T \cdot \operatorname{var} X_1$.

Proof. Denote $\mathcal{G}_n = \mathcal{F}_n \vee \sigma(T)$. Then T is $\mathcal{G}_n$-stopping time and $Y_n = S_n - n\mathbb{E}X_1$ is a $\mathcal{G}_n$-martingale (by Proposition 2.17). The proof of part a) follows in the same way as the proof of part a) in Theorem 3.8. Another possibility is to use independence and directly calculate $\mathbb{E}S_T$ (see Probability Theory 1).

Part b) could be shown directly:

$$\mathbb{E}S_T^2 = \mathbb{E}\sum_{k=1}^{\infty} \mathbf{1}_{[T=k]}S_k^2 = \sum_{k=1}^{\infty} \mathbb{P}(T = k)\mathbb{E}S_k^2 = \mathbb{E}X_1^2 \sum_{k=1} k\mathbb{P}(T = k) = \mathbb{E}T \cdot \mathbb{E}X_1^2.$$

□

The following theorem gives an application of the stopping theory.

Theorem 3.10. (bankruptcy is definitive) *Let* $X_1, X_2, \ldots$ *be a non-negative supermartingale. Consider* $T = \min\{n : X_n = 0\}$*, where* $\min \emptyset = \infty$*. Then the implication* $(T < \infty \Rightarrow X_{T+k} = 0$ *for* $k \in \mathbb{N})$ *holds a.s.*

Proof. If $T \overset{a.s.}{=} \infty$ there is nothing to prove. Let $\mathbb{P}(T < \infty) > 0$ and define $n_0 = \min\{n \in \mathbb{N} : \mathbb{P}(T \le n) > 0\}$. Take $n \ge n_0$ and $k \in \mathbb{N}$. Denote $T_n = T \wedge n$. Then $T_n \le T_n + k \le n + k$ are stopping times and the variant of Theorem 3.1 for supermartingale yields

$$\mathbb{E}^{\mathcal{F}_{T_n}} X_{T_n+k} \overset{a.s.}{\le} X_{T_n}.$$

Since $[T \le n] \in \mathcal{F}_T \cap \mathcal{F}_n = \mathcal{F}_{T_n}$ by Proposition 2.4, we have

$$0 \le \int_{[T\le n]} X_{T+k}\,\mathrm{d}\mathbb{P} = \int_{[T\le n]} X_{T_n+k}\,\mathrm{d}\mathbb{P} \le \int_{[T\le n]} X_{T_n}\,\mathrm{d}\mathbb{P} = \int_{[T\le n]} X_T\,\mathrm{d}\mathbb{P} = 0.$$

It means that $X_{T+k}\mathbf{1}_{[T\le n]} \overset{a.s.}{=} 0$. Taking the limit as $n \to \infty$ we obtain $X_{T+k}\mathbf{1}_{[T<\infty]} \overset{a.s.}{=} 0$. In other words, there exists $\mathbb{P}$-null set N_k (i.e. $\mathbb{P}(N_k) = 0$) such that $X_{T+k}(\omega)\mathbf{1}_{[T(\omega)<\infty]} = 0$ for $\omega \notin N_k$. From this it follows that the sequence $(X_{T+k}(\omega)\mathbf{1}_{[T(\omega)<\infty]}, k \in \mathbb{N})$ is equal to the null sequence for $\omega \notin N = \cup_{k=1}^{\infty} N_k$, where $\mathbb{P}(N) = 0$. □

The application of optional stopping theorem and optional sampling theorem to the simple random walk is left to exercise classes.

Definition 3.1. We say that the random walk (S_n) with steps (X_n) is *nontrivial* if $\mathbb{P}(X_1 \ne 0) > 0$.

A nontrivial random walk has one of the following properties (see exercise classes): 1. $S_n \xrightarrow[n\to\infty]{\text{a.s.}} \infty$, 2. $S_n \xrightarrow[n\to\infty]{\text{a.s.}} -\infty$, 3. $\limsup_{n\to\infty} S_n = \infty$ and $\liminf_{n\to\infty} S_n = -\infty$ a.s. In particular, if T^B is the first exit time of nontrivial random walk from a bounded Borel set $B \in \mathcal{B}(\mathbb{R})$, then $T^B \overset{a.s.}{<} \infty$. Moreover, it is true that T^B has all moments finite.

Theorem 3.11. *Let $T^B = \min\{n : X_n \notin B\}$ be the first exit time of a nontrivial random walk (S_n) from a bounded Borel set $B \in \mathcal{B}(\mathbb{R})$. Then $\mathbb{E}(T^B)^r < \infty$ for all $r \in \mathbb{N}$.*

Proof. Exercise class. □

Using Theorem 3.11 we get transparent variant of Theorem 3.8.

Theorem 3.12. *Let (S_n) be a nontrivial random walk and let T be its first exit time from some bounded Borel set. Then*

a) $X_1 \in L_1 \implies S_T \overset{a.s.}{=} \sum_{k=1}^{T} X_k \in L_1$ and $\mathbb{E}S_T = \mathbb{E}T \cdot \mathbb{E}X_1$,

b) $X_1 \in L_2$, $\mathbb{E}X_1 = 0 \implies S_T \in L_2$ and $\operatorname{var} S_T = \mathbb{E}S_T^2 = \mathbb{E}T \cdot \mathbb{E}X_1^2 = \mathbb{E}T \cdot \operatorname{var} X_1$.

Proof. It is a consequence of Theorem 3.8 as T is an integrable stopping time which satisfies $T > n \Rightarrow |S_n| \le c$ a.s. □

The theory of martingales was developed by a distinguished American mathematician J. L. Doob (1910–2004). We formulate two maximal inequalities that are named after him. However, first we prove the following lemma.

Lemma 3.13. *Let $X_1, X_2, \ldots$ be a non-negative submartingale. If we denote $M_n = \max_{k=1,\ldots,n} X_k$ for $n \in \mathbb{N}$, then*

$$\mathbb{P}(M_n \ge a) \le a^{-1} \int_{[M_n\ge a]} X_n\,\mathrm{d}\mathbb{P} \le a^{-1}\mathbb{E}X_n$$

for any $a > 0$.

Proof. Define $F_k = [X_1 < a, \ldots, X_{k-1} < a, X_k \ge a] \in \sigma(X_1, \ldots, X_k)$, $k = 1, \ldots, n$. Then

$$a\mathbb{P}(M_n \ge a) = a\mathbb{P}\left(\bigcup_{k=1}^{n} F_k\right) = a\sum_{k=1}^{n}\mathbb{P}(F_k) \le \sum_{k=1}^{n}\int_{F_k} X_k\,\mathrm{d}\mathbb{P} \le \sum_{k=1}^{n}\int_{F_k} X_n\,\mathrm{d}\mathbb{P} = \int_{[M_n\ge a]} X_n\,\mathrm{d}\mathbb{P} \le \mathbb{E}X_n.$$

□

Theorem 3.14. (Doob's maximal inequalities) *Let $X_1, X_2, \dots$ be a martingale or non-negative submartingale. Then for all $n \in \mathbb{N}$ we have*

1. $$\mathbb{P}\left(\max_{k=1,\dots,n} |X_k| \geq a\right) \leq a^{-p}\mathbb{E}|X_n|^p, \quad \textit{for } p \geq 1 \textit{ and } a > 0,$$

2. $$\mathbb{E}\left(\max_{k=1,\dots,n} |X_k|\right)^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|X_n|^p, \quad \textit{for } p > 1.$$

Proof. Let us fix $n \in \mathbb{N}$. Obviously we can assume that $X_n \in L_p$, otherwise the right-hand side is infinite and the inequalities hold trivially. If (X_n) is a martingale or non-negative submartingale, then $(|X_n|^p)$ with $p \geq 1$ is a non-negative submartingale by Proposition 2.19. Applying Lemma 3.13 to this submartingale gives the first inequality.

Let $Y = \max_{k=1,\dots,n} |X_k|$ and $p > 1$. Then

$$\mathbb{E}Y^p \leq \sum_{k=1}^{n} \mathbb{E}|X_k|^p \leq n\mathbb{E}|X_n|^p < \infty.$$

The expectation of a non-negative random variable is obtained by integrating its complementary distribution function. Therefore,

$$\mathbb{E}Y^p = \int_0^\infty \mathbb{P}(Y^p > t)\, \mathrm{d}t = \int_0^\infty \mathbb{P}(Y^p > a^p) p a^{p-1}\, \mathrm{d}a.$$

If we use Lemma 3.13 for $|X_1|, |X_2|, \dots$ and Fubini's theorem, we get the bound

$$\begin{aligned} \mathbb{E}Y^p &\leq \int_0^\infty p a^{p-1} a^{-1} \int_{[Y \geq a]} |X_n|\, \mathrm{d}\mathbb{P}\, \mathrm{d}a = \mathbb{E} \int_0^\infty p a^{p-2} |X_n| \mathbf{1}_{[Y \geq a]}\, \mathrm{d}a \\ &= p\mathbb{E}|X_n|\frac{Y^{p-1}}{p-1} = \frac{p}{p-1}\mathbb{E}|X_n|Y^{p-1}. \end{aligned}$$

By Hölder's inequality,

$$\mathbb{E}|X_n|Y^{p-1} \leq \left(\mathbb{E}|X_n|^p\right)^{\frac{1}{p}} \left(\mathbb{E}Y^p\right)^{\frac{p-1}{p}}.$$

Therefore,

$$\mathbb{E}Y^p \leq \frac{p}{p-1} \left(\mathbb{E}|X_n|^p\right)^{\frac{1}{p}} \left(\mathbb{E}Y^p\right)^{\frac{p-1}{p}}.$$

The second stated inequality is trivially satisfied for $\mathbb{E}Y^p = 0$. If $\mathbb{E}Y^p > 0$ we can divide both sides by $\left(\mathbb{E}Y^p\right)^{\frac{p-1}{p}}$ and obtain

$$\left(\mathbb{E}Y^p\right)^{1/p} \leq \frac{p}{p-1} \left(\mathbb{E}|X_n|^p\right)^{1/p},$$

which is equivalent to the stated inequality. ∎

Doob's maximal inequality implies classical maximal inequalities for independent random variables.

Theorem 3.15. (Kolmogorov's inequality) *Let $X_1, X_2, \dots$ be independent random variables with zero expectation and finite variance and let $S_n = X_1 + \cdots + X_n$. Then*

$$\mathbb{P}\left(\max_{k=1,\dots,n} |S_k| \geq a\right) \leq a^{-2}\mathbb{E}S_n^2 = a^{-2}\sum_{k=1}^{n} \mathbb{E}X_k^2, \quad a > 0.$$

Proof. Since (S_n) is a martingale, it is enough to use the first Doob's maximal inequality for $p = 2$. ∎

4 Submartingale convergence

Definition 4.1. Consider real numbers $a < b$ and a finite sequence $y^{(n)} = (y_1, \dots, y_n)$ of real numbers. We denote the *number of upcrossings of* (a, b) by

$$y^{(n)}\uparrow_a^b = \operatorname{card}\{(s,t) : 1 \le s < t \le n, \{y_{s+1}, \dots, y_{t-1}\} \subseteq (a,b) \subseteq [y_s, y_t]\},$$

where we let $[c, d] = \emptyset$ for $c > d$. Analogously, the *number of downcrossings of* (a, b) is

$$y^{(n)}\downarrow_a^b = \operatorname{card}\{(s,t) : 1 \le s < t \le n, \{y_{s+1}, \dots, y_{t-1}\} \subseteq (a,b) \subseteq [y_t, y_s]\}.$$

For an infinite sequence $y = (y_1, y_2, \dots)$ we put

$$y\uparrow_a^b = \lim_{n\to\infty} y^{(n)}\uparrow_a^b \quad \text{and} \quad y\downarrow_a^b = \lim_{n\to\infty} y^{(n)}\downarrow_a^b .$$

Remark: Clearly, $y\uparrow_a^b = (-y)\downarrow_{-b}^{-a}$ and $y\downarrow_a^b -1 \le y\uparrow_a^b \le y\downarrow_a^b +1$. If $X = (X_1, X_2, \dots)$ is a random sequence, then $X\uparrow_a^b$ is a random variable with values in $\mathbb{N} \cup \{0, \infty\}$.

Proposition 4.1. *Let $X = (X_1, X_2, \dots)$ be a random sequence.*

(i) There exists a random variable X^* (with values in $\mathbb{R} \cup \{-\infty, \infty\}$) such that $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X^*$ if and only if $\mathbb{P}(X\uparrow_a^b < \infty) = 1$ for each $a, b \in \mathbb{R} : a < b$.

(ii) There exists a random variable Y (with values in $\mathbb{R}$) such that $X_n \xrightarrow[n\to\infty]{\text{a.s.}} Y$ if and only if $\mathbb{P}(\sup_{n\in\mathbb{N}} |X_n| < \infty) = 1$ and $\mathbb{P}(X\uparrow_a^b < \infty) = 1$ for each $a, b \in \mathbb{R} : a < b$.

Proof. Both implications from left to right are obvious.

Assume that $\mathbb{P}(X\uparrow_a^b < \infty) = 1$ for each $a, b \in \mathbb{R} : a < b$ and consider random variables

$$X^* = \limsup_{n\to\infty} X_n \quad \text{and} \quad X_* = \liminf_{n\to\infty} X_n.$$

Then

$$\mathbb{P}(X_* < X^*) \le \sum_{a,b\in\mathbb{Q}:a<b} \mathbb{P}(X_* < a < b < X^*) \le \sum_{a,b\in\mathbb{Q}:a<b} \mathbb{P}(X\uparrow_a^b = \infty) = 0,$$

and so $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X^*$.

If we moreover assume that $\sup_{n\in\mathbb{N}} |X_n| \overset{a.s.}{<} \infty$, then $|X^*| \le \sup_{n\in\mathbb{N}} |X_n| \overset{a.s.}{<} \infty$ and we can take $Y = X^* \mathbf{1}_{[|X^*|<\infty]}$. □

Theorem 4.2. (Doob's upcrossing inequality) *Let (X_n) be an $\mathcal{F}_n$-submartingale. Then for $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$ such that $a < b$,*

$$\mathbb{E}X^{(n)}\uparrow_a^b \le \frac{\mathbb{E}(X_n - a)^+ - \mathbb{E}(X_1 - a)^+}{b - a} \le \frac{\mathbb{E}(X_n - a)^+}{b - a},$$

where $X^{(n)} = (X_1, \dots, X_n)$.

Proof. Consider a sequence $Z_n = (X_n - a)^+$, $n \in \mathbb{N}$. By Proposition 2.19 we know that (Z_n) is an $\mathcal{F}_n$-submartingale. Define $\mathcal{F}_n$-stopping times $\tau_0 = 1$, $\nu_j = \min\{k \ge \tau_{j-1} : Z_k = 0\} \wedge n$, $\tau_j = \min\{k \ge \nu_j : Z_k \ge b - a\} \wedge n$. From the definition we see that $\tau_0 \le \nu_1 \le \tau_1 \le \nu_2 \le \cdots$. Moreover, $\nu_j < n$ implies $\nu_j < \tau_j$. Similarly, $\tau_j < n$ implies $\tau_j < \nu_{j+1}$. Therefore, there exists $m \in \mathbb{N}$ such that $\tau_m = \nu_m = n$. We can write

$$Z_n - Z_1 = \sum_{j=1}^{m} (Z_{\nu_j} - Z_{\tau_{j-1}}) + \sum_{j=1}^{m} (Z_{\tau_j} - Z_{\nu_j}) \ge \sum_{j=1}^{m} (Z_{\nu_j} - Z_{\tau_{j-1}}) + (b - a) Z^{(n)}\uparrow_0^{b-a},$$

where $Z^{(n)} = (Z_1, \dots, Z_n)$. According to Theorem 3.1 for bounded times $\tau_{j-1} \le \nu_j$ we have $\mathbb{E}Z_{\nu_j} \ge \mathbb{E}Z_{\tau_{j-1}}$ for any j. Hence,

$$\mathbb{E}(Z_n - Z_1) \ge (b - a)\mathbb{E}Z^{(n)}\uparrow_0^{b-a} .$$

Finally, it suffices to use the definition of (Z_n) which yields

$$\mathbb{E}(X_n - a)^+ - \mathbb{E}(X_1 - a)^+ \geq (b-a)\mathbb{E}X^{(n)}\uparrow_a^b,$$

because $Z^{(n)}\uparrow_0^{b-a} = X^{(n)}\uparrow_a^b$. □

Remark: The statement for an $\mathcal{F}_n$-supermartingale (X_n) has the form

$$\mathbb{E}X^{(n)}\downarrow_a^b \leq \frac{\mathbb{E}(b - X_n)^+ - \mathbb{E}(b - X_1)^+}{b-a} \leq \frac{\mathbb{E}(b - X_n)^+}{b-a},$$

as it follows from the relation $y^{(n)}\uparrow_a^b = (-y)^{(n)}\downarrow_{-b}^{-a}$.

Theorem 4.3. (Doob's submartingale convergence theorem) *Let (X_n) be an $\mathcal{F}_n$-submartingale that satisfies $\sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ < \infty$. Then there exists a random variable $X_\infty \in L_1$ such that $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X_\infty$ and*

$$\mathbb{E}X_\infty^+ \leq \liminf_{n\to\infty} \mathbb{E}X_n^+ \leq \sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ < \infty$$
$$\mathbb{E}X_\infty^- \leq \liminf_{n\to\infty} \mathbb{E}X_n^- \leq \sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ - \mathbb{E}X_1 < \infty.$$

Proof. For $a < b$, $X\uparrow_a^b$ is the limit of a non-decreasing non-negative sequence $X^{(n)}\uparrow_a^b$, where $X^{(n)} = (X_1, \ldots, X_n)$. Hence, from Lévi's monotone convergence theorem and Theorem 4.2 we obtain

$$\begin{aligned}\mathbb{E}X\uparrow_a^b &= \lim_{n\to\infty} \mathbb{E}X^{(n)}\uparrow_a^b \leq \liminf_{n\to\infty} \frac{\mathbb{E}(X_n - a)^+}{b-a} \\ &\leq \liminf_{n\to\infty} \frac{\mathbb{E}X_n^+ + a^-}{b-a} \leq \frac{\sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ + a^-}{b-a} < \infty.\end{aligned}$$

By Proposition 4.1 there exists a random variable X_∞ such that $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X_\infty$. Positive and negative part are continuous functions, so also $X_n^+ \xrightarrow[n\to\infty]{\text{a.s.}} X_\infty^+$ and $X_n^- \xrightarrow[n\to\infty]{\text{a.s.}} X_\infty^-$. From Fatou's lemma we get

$$\begin{aligned}\mathbb{E}X_\infty^+ &= \mathbb{E}\liminf_{n\to\infty} X_n^+ \leq \liminf_{n\to\infty} \mathbb{E}X_n^+ \leq \sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ < \infty, \\ \mathbb{E}X_\infty^- &= \mathbb{E}\liminf_{n\to\infty} X_n^- \leq \liminf_{n\to\infty} (\mathbb{E}X_n^+ - \mathbb{E}X_n) \leq \sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ - \mathbb{E}X_1 < \infty.\end{aligned}$$

We used that the submartingale (X_n) satisfies $\mathbb{E}X_n \geq \mathbb{E}X_1$ for any $n \in \mathbb{N}$. Altogether $\mathbb{E}|X_\infty| = \mathbb{E}X_\infty^+ + \mathbb{E}X_\infty^- < \infty$. □

Remark: Similarly, every $\mathcal{F}_n$-supermartingale satisfying $\sup_{n\in\mathbb{N}} \mathbb{E}X_n^- < \infty$ has integrable a.s.-limit. As special cases we have the following statements:

1. A submartingale bounded from above has an integrable a.s.-limit.
2. A supermartingale bounded from below (e.g. non-negative supermartingale) has an integrable a.s.-limit.
3. Each martingale that is bounded either from above or from below has an integrable a.s.-limit.

Corollary 4.4. *Let (X_n) be a sequence of independent random variables such that $\mathbb{E}X_n = 0$ and*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left|\sum_{k=1}^n X_k\right| < \infty.$$

Then $\sum_{k=1}^\infty X_k$ is a.s.-summable and $\sum_{k=1}^\infty X_k \in L_1$.

Proof. Define $S_n = \sum_{k=1}^n X_k$, $n \in \mathbb{N}$. Then (S_n) is a martingale and we assume that $\sup_{n\in\mathbb{N}} \mathbb{E}|S_n| < \infty$. So it suffices to use Theorem 4.3. □

Remark: Recall that the condition $\sum_{n=1}^{\infty} \operatorname{var} X_n < \infty$ is sufficient for the summability of $\sum_{n=1}^{\infty}(X_n - \mathbb{E}X_n)$ a.s., in probability and in L_2 (see Probability Theory 1). In this case we have an improvement for a.s.-summability. The condition $\sum_{n=1}^{\infty} \operatorname{var} X_n < \infty$ implies $\sup_{n\in\mathbb{N}} \mathbb{E}|S_n| < \infty$ because

$$\mathbb{E}|S_n| \le \sqrt{\mathbb{E}S_n^2} = \sqrt{\sum_{k=1}^{n} \operatorname{var} X_k} \le \sqrt{\sum_{k=1}^{\infty} \operatorname{var} X_k}.$$

Definition 4.2. Let $(\dots, X_{-2}, X_{-1})$ be a random sequence indexed by negative integers. Let $\dots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1}$ be a non-decreasing sequence of σ-algebras (filtration). Assume that $X_{-n} \in L_1$ for any $n \in \mathbb{N}$ and $\sigma(\dots, X_{-n-1}, X_{-n}) \subseteq \mathcal{F}_{-n}$. We say that the sequence (X_{-n}) is an $\mathcal{F}_{-n}$*-martingale* if

$$\mathbb{E}[X_{-n} \mid \mathcal{F}_{-(n+1)}] \overset{a.s.}{=} X_{-(n+1)} \quad \text{for all } n \in \mathbb{N}.$$

If $\mathcal{F}_{-n} = \sigma(\dots, X_{-n-1}, X_{-n})$, then we speak about a *backwards martingale.* Analogously we define $\mathcal{F}_{-n}$*-submartingale* and $\mathcal{F}_{-n}$*-supermartingale.* We denote $\mathcal{F}_{-\infty} = \cap_{n=1}^{\infty} \mathcal{F}_{-n}$.

Theorem 4.5. (convergence of UI (sub)martingale)

a) Let (X_n) be a UI $\mathcal{F}_n$-submartingale (or $\mathcal{F}_n$-martingale), then there exists a random variable $X_\infty \in L_1$ such that $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X_\infty$ and $X_n \xrightarrow[n\to\infty]{L_1} X_\infty$. Furthermore, $\mathbb{E}[X_\infty \mid \mathcal{F}_n] \overset{a.s.}{\ge} X_n$ (or $\mathbb{E}[X_\infty \mid \mathcal{F}_n] \overset{a.s.}{=} X_n$) for all $n \in \mathbb{N}$.

b) Let (X_n) be a UI $\mathcal{F}_{-n}$-submartingale (or $\mathcal{F}_{-n}$-martingale), then there exists a random variable $X_{-\infty} \in L_1$ such that $X_{-n} \xrightarrow[n\to\infty]{\text{a.s.}} X_{-\infty}$ and $X_{-n} \xrightarrow[n\to\infty]{L_1} X_{-\infty}$. Furthermore, $\mathbb{E}[X_{-n} \mid \mathcal{F}_{-\infty}] \overset{a.s.}{\ge} X_{-\infty}$ (or $\mathbb{E}[X_{-n} \mid \mathcal{F}_{-\infty}] \overset{a.s.}{=} X_{-\infty}$) for all $n \in \mathbb{N}$.

Proof. a) The assumption of uniform integrability implies uniformly bounded moments and the assumption of Doob's submartingale convergence theorem (Theorem 4.3) is fulfilled: $\sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ \le \sup_{n\in\mathbb{N}} \mathbb{E}|X_n| < \infty$. Hence, there exists a random variable $X_\infty \in L_1$ such that $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X_\infty$. We obtain the L_1 convergence from UI property. For fixed integer numbers $n \le m$ we have $X_n \overset{a.s.}{\le} \mathbb{E}[X_m \mid \mathcal{F}_n]$. By the L_1 continuity of conditional expectation (Proposition 2.16a), $\mathbb{E}[X_m \mid \mathcal{F}_n]$ converges to $\mathbb{E}[X_\infty \mid \mathcal{F}_n]$ as $m \to \infty$. The L_1 convergence implies convergence in probability which in turn implies the existence of a subsequence that converges a.s. For this subsequence (m_k) we already know that $X_n \overset{a.s.}{\le} \mathbb{E}[X_{m_k} \mid \mathcal{F}_n]$. This inequality is preserved when passing to the limit. Hence, $X_n \overset{a.s.}{\le} \mathbb{E}[X_\infty \mid \mathcal{F}_n]$.

b) An analogy of Theorem 4.3 for $\mathcal{F}_{-n}$-submartingale and uniform integrability ensure the existence of a random variable $X_{-\infty} \in L_1$ such that $X_{-n} \xrightarrow[n\to\infty]{\text{a.s.}} X_{-\infty}$ and $X_n \xrightarrow[n\to\infty]{L_1} X_{-\infty}$. Then the L_1 continuity of conditional expectation (Proposition 2.16a) yields $\mathbb{E}[X_{-m} \mid \mathcal{F}_{-\infty}] \xrightarrow[m\to\infty]{L_1} \mathbb{E}[X_{-\infty} \mid \mathcal{F}_{-\infty}] = X_{-\infty}$. From $X_{-m} \overset{a.s.}{\le} \mathbb{E}[X_{-n} \mid \mathcal{F}_{-m}]$ for $m \ge n$ we get by conditioning with σ-algebra $\mathcal{F}_{-\infty}$ the relation $\mathbb{E}[X_{-m} \mid \mathcal{F}_{-\infty}] \overset{a.s.}{\le} \mathbb{E}[X_{-n} \mid \mathcal{F}_{-\infty}]$. The left-hand side goes in L_1 to $X_{-\infty}$ as $m \to \infty$ and the inequality is preserved when passing to the limit.

□

Corollary 4.6. *If (X_n) is an $\mathcal{F}_n$-adapted random sequence, then (X_n) is UI $\mathcal{F}_n$-martingale if and only if there exists $X_\infty \in L_1$ such that $X_n \overset{a.s.}{=} \mathbb{E}[X_\infty \mid \mathcal{F}_n]$.*

Proof. The implication from left to right follows from Theorem 4.5. Conversely, the sequence given by $X_n \overset{a.s.}{=} \mathbb{E}[X_\infty \mid \mathcal{F}_n]$ is an $\mathcal{F}_n$-martingale (see exercise class) that is UI (Proposition 2.16c). □

Convergence theorems imply that conditional expectations are continuous in the condition.

Proposition 4.7. *Let $Y \in L_1$ and $\dots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}$, $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}$ be non-decreasing sequences of σ-algebras. Then* *a)* $\mathbb{E}[Y \mid \mathcal{F}_n] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[Y \mid \mathcal{F}_\infty]$ *both a.s. and in* L_1*,*

b) $\mathbb{E}[Y \mid \mathcal{F}_{-n}] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[Y \mid \mathcal{F}_{-\infty}]$ *both a.s. and in* L_1*.*

Proof. a) We know that $Y_n = \mathbb{E}[Y \mid \mathcal{F}_n]$ is UI $\mathcal{F}_n$-martingale (Corollary 4.6). By Theorem 4.5 there exists $Y_\infty \in L_1$ so that $Y_n \underset{n\to\infty}{\longrightarrow} Y_\infty$ both a.s. and in L_1 and $Y_n \overset{a.s.}{=} \mathbb{E}[Y_\infty \mid \mathcal{F}_n]$. We show that $\mathbb{E}[Y \mid \mathcal{F}_\infty] \overset{a.s.}{=} Y_\infty$. For $F \in \mathcal{F}_n$ we have

$$\int_F Y \,\mathrm{d}\mathbb{P} = \int_F Y_n \,\mathrm{d}\mathbb{P} = \int_F Y_\infty \,\mathrm{d}\mathbb{P} = \int_F \limsup_{n\to\infty} \mathbb{E}[Y \mid \mathcal{F}_n] \,\mathrm{d}\mathbb{P}.$$

The first equality follows from $Y_n \overset{a.s.}{=} \mathbb{E}[Y \mid \mathcal{F}_n]$; the second equality from $Y_n \overset{a.s.}{=} \mathbb{E}[Y_\infty \mid \mathcal{F}_n]$; while the last equality from $Y_\infty \overset{a.s.}{=} \limsup_{n\to\infty} \mathbb{E}[Y \mid \mathcal{F}_n]$. We have verified the relation

$$\int_F Y \,\mathrm{d}\mathbb{P} = \int_F \limsup_{n\to\infty} \mathbb{E}[Y \mid \mathcal{F}_n] \,\mathrm{d}\mathbb{P}$$

for any $F \in \cup_{n=1}^\infty \mathcal{F}_n$, where $\cup_{n=1}^\infty \mathcal{F}_n$ is an algebra that generates $\mathcal{F}_\infty$. Moreover, $\limsup_{n\to\infty} \mathbb{E}[Y \mid \mathcal{F}_n]$ is $\mathcal{F}_\infty$-measurable, which yields $\mathbb{E}[Y \mid \mathcal{F}_\infty] \overset{a.s.}{=} \limsup_{n\to\infty} \mathbb{E}[Y \mid \mathcal{F}_n]$.

b) A random sequence $Y_{-n} = \mathbb{E}[Y \mid \mathcal{F}_{-n}]$ is UI backwards martingale (exercise class). By Theorem 4.5 there exists a random variable $Y_{-\infty} \in L_1$ such that $Y_{-n} \underset{n\to\infty}{\longrightarrow} Y_{-\infty}$ both a.s. and in L_1 and $Y_{-\infty} \overset{a.s.}{=} \mathbb{E}[Y_{-n} \mid \mathcal{F}_{-\infty}]$. Then for arbitrary $F \in \mathcal{F}_{-\infty}$ we have

$$\int_F Y_{-\infty} \,\mathrm{d}\mathbb{P} = \int_F Y_{-n} \,\mathrm{d}\mathbb{P} = \int_F Y \,\mathrm{d}\mathbb{P}.$$

In other words, $Y_{-\infty} \overset{a.s.}{=} \mathbb{E}[Y \mid \mathcal{F}_{-\infty}]$.

□

Theorem 4.8. (submartingale converges or explodes) *Let* (X_n) *be a submartingale. Denote* $Y_k = X_{k+1} - X_k$ *for* $k \in \mathbb{N}$*. If* $(\sup_{n\in\mathbb{N}} Y_n)^+ \in L_1$*, then there exists a random variable* X_∞ *such that* $X_n(\omega) \underset{n\to\infty}{\longrightarrow} X_\infty(\omega)$ *for a.a.* $\omega \in \Omega$ *with property* $\sup_{n\in\mathbb{N}} X_n(\omega) < \infty$*.*

Proof. For $k \in \mathbb{N}$ denote a stopping time $\tau_k = \min\{n \in \mathbb{N} : X_n \geq k\}$. We fix $k \in \mathbb{N}$. Optional stopping theorem (Corollary 3.2) states that $(X_{n\wedge\tau_k})$ is a submartingale. We distinguish the following three cases:

1. $\tau_k = 1 \Rightarrow X_{n\wedge\tau_k} = X_1$,
2. $1 < \tau_k \leq n \Rightarrow X_{n\wedge\tau_k} = X_{\tau_k} = X_{\tau_k - 1} + Y_{\tau_k - 1} \leq k + \sup_{n\in\mathbb{N}} Y_n$,
3. $\tau_k > n \Rightarrow X_{n\wedge\tau_k} = X_n < k$.

Combining all three cases we have

$$X^+_{n\wedge\tau_k} \leq X^+_1 + k \; + \left(\sup_{n\in\mathbb{N}} Y_n\right)^+ \in L_1.$$

It means that $\sup_{n\in\mathbb{N}} \mathbb{E}X^+_{n\wedge\tau_k} < \infty$ and we can apply Doob's submartingale convergence theorem (Theorem 4.3). Therefore, there exists a random variable $X^{(k)} \in L_1$ such that $X_{n\wedge\tau_k} \xrightarrow[n\to\infty]{\text{a.s.}} X^{(k)}$. Then for $A_k = [\tau_k = \infty] = [\sup_{n\in\mathbb{N}} X_n < k]$,

$$X_n \mathbf{1}_{A_k} = X_{n\wedge\tau_k} \mathbf{1}_{A_k} \xrightarrow[n\to\infty]{\text{a.s.}} X^{(k)} \mathbf{1}_{A_k}.$$

The events A_k make a non-decreasing sequence and their limit for $k \to \infty$ is $A = [\sup_{n\in\mathbb{N}} X_n < \infty]$. Furthermore, $X^{(k)}\mathbf{1}_{A_k} \overset{a.s.}{=} X^{(l)}\mathbf{1}_{A_k}$ for $l \geq k$. If we put $X_\infty = X^{(1)}\mathbf{1}_{A_1} + X^{(2)}\mathbf{1}_{A_2\setminus A_1} + \cdots$, then $X_n \mathbf{1}_A \xrightarrow[n\to\infty]{\text{a.s.}} X_\infty$. □

5 Limit theorems for martingale differences

Definition 5.1. Let (M_n) be a martingale. Put $M_0 = \mathbb{E}M_1$ and define $D_n = M_n - M_{n-1}$ for $n \in \mathbb{N}$. Then (D_n) is called a *martingale difference sequence* of the martingale (M_n). If (M_n) is an $\mathcal{F}_n$-martingale, then we speak about *$\mathcal{F}_n$-martingale difference sequence.*

Remark: Equivalently we can define $\mathcal{F}_n$-martingale difference sequence as the sequence satisfying $\mathbb{E}(D_n \mid \mathcal{F}_{n-1}) = 0$ for $n \in \mathbb{N}$, where we let $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Theorem 5.1. (summability of martingale differences) *Let (D_n) be a martingale difference sequence of the martingale (M_n) that satisfies $M_n \in L_2$ for each $n \in \mathbb{N}$. If $\sum_{n=1}^\infty \operatorname{var} D_n < \infty$, then the series $\sum_{n=1}^\infty D_n$ is summable both a.s. and in L_2, i.e. the martingale $M_n - \mathbb{E}M_1$ converges both a.s. and in L_2.*

Proof. According to Proposition 2.21 the random variables D_n are uncorrelated. Recall that from Probability Theory 1 we know that centred uncorrelated random variables are summable in L_2 if and only if the sum of their variances is finite. In order to get a.s.-summability we verify the assumption of Theorem 4.3:

$$\mathbb{E}|M_n - \mathbb{E}M_1| \le \sqrt{\mathbb{E}(M_n - \mathbb{E}M_1)^2} = \sqrt{\mathbb{E}\left(\sum_{k=1}^n D_k\right)^2} = \sqrt{\sum_{k=1}^n \mathbb{E}D_k^2} = \sqrt{\sum_{k=1}^n \operatorname{var} D_k} \le \sqrt{\sum_{k=1}^\infty \operatorname{var} D_k} < \infty,$$

and so $\sup_{n\in\mathbb{N}} \mathbb{E}|M_n - \mathbb{E}M_1| < \infty$. □

Theorem 5.2. (strong law of large numbers for martingale differences) *Let (D_n) be a martingale difference sequence of the martingale (M_n) that satisfies $M_n \in L_2$ for each $n \in \mathbb{N}$. Let $0 < b_n \nearrow \infty$ be a real sequence. If $\sum_{n=1}^\infty b_n^{-2} \operatorname{var} D_n < \infty$, then*

$$\frac{1}{b_n}\sum_{k=1}^n D_k = \frac{M_n - \mathbb{E}M_1}{b_n} \xrightarrow[n\to\infty]{} 0 \quad \textit{both a.s. and in } L_2.$$

Proof. The sequence $\left(\frac{D_n}{b_n}\right)$ is also a martingale difference sequence and it satisfies the assumption of Theorem 5.1. Therefore, the series $\sum_{n=1}^\infty b_n^{-1} D_n$ is a.s.-summable. The a.s.-convergence of $b_n^{-1}\sum_{k=1}^n D_k$ follows from Kronecker's lemma which says that if $\sum_{n=1}^\infty a_n < \infty$ and $0 < b_n \nearrow \infty$, then $\frac{1}{b_n}\sum_{k=1}^n a_k b_k \xrightarrow[n\to\infty]{} 0$. To show convergence in L_2 we apply Kronecker's lemma as well:

$$\mathbb{E}\left(\frac{1}{b_n}\sum_{k=1}^n D_k\right)^2 = \frac{1}{b_n^2}\sum_{k=1}^n \mathbb{E}D_k^2 = \frac{1}{b_n^2}\sum_{k=1}^n \operatorname{var} D_k \xrightarrow[n\to\infty]{} 0.$$

□

Theorem 5.3. (central limit theorem for martingale differences) *Let (D_n) be an $\mathcal{F}_n$-martingale difference sequence of the martingale (M_n). Assume that for each $n \in \mathbb{N}$ we have*

- *1.* $\mathbb{E}(D_n^2 \mid \mathcal{F}_{n-1}) = 1$,
- *2.* $\mathbb{E}(|D_n|^3 \mid \mathcal{F}_{n-1}) \le K < \infty$,

where $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n D_k = \frac{1}{\sqrt{n}}(M_n - \mathbb{E}M_1) \xrightarrow[n\to\infty]{\mathrm{d}} N(0,1).$$

Proof. Define

$$\varphi_{n,k}(t) = \mathbb{E}\left(\exp\left\{\mathrm{i}t\frac{D_k}{\sqrt{n}}\right\} \Big| \mathcal{F}_{k-1}\right), \quad k = 1, \dots, n,\ n \in \mathbb{N}.$$

From Taylor's expansion we get

$$\exp\left\{\mathrm{i}t\frac{D_k}{\sqrt{n}}\right\} = 1 + \mathrm{i}t\frac{D_k}{\sqrt{n}} - \frac{t^2 D_k^2}{2n} - \mathrm{i}t^3\frac{\Delta_k^3}{6n^{3/2}},$$

where Δ_k is a random variable such that $0 \le \Delta_k \le D_k$. Applying conditional expectation on both sides we obtain

$$\varphi_{n,k}(t) = 1 + \frac{\mathrm{i}t}{\sqrt{n}}\mathbb{E}(D_k \mid \mathcal{F}_{k-1}) - \frac{t^2}{2n}\mathbb{E}(D_k^2 \mid \mathcal{F}_{k-1}) - \frac{\mathrm{i}t^3}{6n^{3/2}}\mathbb{E}(\Delta_k^3 \mid \mathcal{F}_{k-1}),$$

which by our assumptions can be simplified to

$$\varphi_{n,k}(t) = 1 - \frac{t^2}{2n} - \frac{\mathrm{i}t^3}{6n^{3/2}}\mathbb{E}(\Delta_k^3 \mid \mathcal{F}_{k-1}).$$

For $p = 1, \dots, n$ we have

$$\begin{aligned}\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_p}{\sqrt{n}}\right\} &= \mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\exp\left\{\mathrm{i}t\frac{D_p}{\sqrt{n}}\right\}\right] = \mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\mathbb{E}\left(\exp\left\{\mathrm{i}t\frac{D_p}{\sqrt{n}}\right\}\Big|\mathcal{F}_{p-1}\right)\right]\\ &= \mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\varphi_{n,p}(t)\right] = \mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\left(1 - \frac{t^2}{2n} - \frac{\mathrm{i}t^3}{6n^{3/2}}\mathbb{E}(\Delta_p^3 \mid \mathcal{F}_{p-1})\right)\right].\end{aligned}$$

Consequently,

$$\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_p}{\sqrt{n}}\right\} - \left(1 - \frac{t^2}{2n}\right)\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\} = -\frac{\mathrm{i}t^3}{6n^{3/2}}\mathbb{E}\left[\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\mathbb{E}(\Delta_p^3 \mid \mathcal{F}_{p-1})\right].$$

Since $|\Delta_p| \le |D_p|$ and the conditional absolute third moments are bounded, it follows that

$$\left|\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_p}{\sqrt{n}}\right\} - \left(1 - \frac{t^2}{2n}\right)\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\right| \le K\,\frac{|t|^3}{6n^{3/2}}. \tag{14}$$

Let us fix $t \in \mathbb{R}$. For sufficiently large n (namely, $n \ge t^2/2$) we have $0 \le 1 - \frac{t^2}{2n} \le 1$. Therefore, the left-hand side of (14) is not going to increase by multiplication with $\left(1 - \frac{t^2}{2n}\right)^{n-p}$. It means that

$$\left|\left(1 - \frac{t^2}{2n}\right)^{n-p}\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_p}{\sqrt{n}}\right\} - \left(1 - \frac{t^2}{2n}\right)^{n-p+1}\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\right| \le K\,\frac{|t|^3}{6n^{3/2}}.$$

We apply the triangle inequality on the identity

$$\begin{aligned}&\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_n - \mathbb{E}M_1}{\sqrt{n}}\right\} - \left(1 - \frac{t^2}{2n}\right)^n\\ &\quad= \exp\left\{-\mathrm{i}t\frac{\mathbb{E}M_1}{\sqrt{n}}\right\}\sum_{p=1}^{n}\left[\left(1 - \frac{t^2}{2n}\right)^{n-p}\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_p}{\sqrt{n}}\right\} - \left(1 - \frac{t^2}{2n}\right)^{n-p+1}\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_{p-1}}{\sqrt{n}}\right\}\right]\end{aligned}$$

and obtain (for $n \ge t^2/2$)

$$\left|\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_n - \mathbb{E}M_1}{\sqrt{n}}\right\} - \left(1 - \frac{t^2}{2n}\right)^n\right| \le nK\frac{|t|^3}{6n^{3/2}} = K\,\frac{|t|^3}{6\sqrt{n}}.$$

As the right-hand side tends to zero and $\left(1 - \frac{t^2}{2n}\right)^n$ tends to $\mathrm{e}^{-t^2/2}$ for $n \to \infty$, we have

$$\lim_{n\to\infty}\mathbb{E}\exp\left\{\mathrm{i}t\frac{M_n - \mathbb{E}M_1}{\sqrt{n}}\right\} = \exp\left\{-\frac{t^2}{2}\right\}.$$

We showed the pointwise convergence of characteristic functions to the characteristic function of standard normal distibution. This proves the desired convergence in distribution. □

At the end we present (without proof) a generalization of the Feller-Lindeberg central limit theorem for triangular array of martingale differences.

Theorem 5.4. (Brown's central limit theorem for martingale differences) *Consider a triangular array* $(D_{k,n}, k = 1, \dots, k_n, n \in \mathbb{N})$ *such that for each row* $n \in \mathbb{N}$ *there are* σ*-algebras* $\mathcal{F}_{0,n} = \{\emptyset, \Omega\} \subseteq \mathcal{F}_{1,n} \subseteq \mathcal{F}_{2,n} \subseteq \cdots \subseteq \mathcal{F}_{k_n,n}$ *and* $D_{1,n}, \dots, D_{k_n,n}$ *is* $(\mathcal{F}_{k,n}, k = 1, \dots, k_n)$*-martingale difference sequence. Assume that*

(i) the conditional Feller-Lindeberg condition is satisfied, that is,

$$\sum_{k=1}^{k_n} \mathbb{E}\left(D_{k,n}^2 \mathbf{1}_{[|D_{k,n}| \geq \varepsilon]} \mid \mathcal{F}_{k-1,n}\right) \xrightarrow[n\to\infty]{\mathbb{P}} 0 \quad \text{for every } \varepsilon > 0,$$

(ii) $\sum_{k=1}^{k_n} \mathbb{E}(D_{k,n}^2 \mid \mathcal{F}_{k-1,n}) \xrightarrow[n\to\infty]{\mathbb{P}} 1.$

Then

$$\sum_{k=1}^{k_n} D_{k,n} \xrightarrow[n\to\infty]{\mathrm{d}} N(0,1).$$

Note that Theorem 5.3 is a special case of Theorem 5.4. It suffice to take $D_{k,n} = \frac{D_k}{\sqrt{n}}$, $k = 1, \dots, k_n = n$. Conditional Feller-Lindeberg condition follows from the assumption that conditional absolute third moments are bounded.