Stochastic Processes
– Lecture Notes –

Marcin Pitera
Jagiellonian University

November 29, 2020

Contents

1 Introduction & Preliminaries

In this course we assume the knowledge of basic concepts, techniques, and theorems from Analysis, Measure theory, Probability Theory, and Statistics. Before we begin we quickly outline some basics related to conditional expectations and martingale theory that will be relevant for this course.

If not stated otherwise, we assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is the underlying probability space. Also, we assume that all random variables introduced throughout this course are integrable. If not stated otherwise, all the inequalities should be understood in the almost-sure sense.

This lecture notes are partly based on polish lecture notes prepared by prof. Szymon Peszat. I would like to thank prof. Peszat for his authorisation to share the materials.

1.1 Conditional expectation

Let $\mathcal{G}$ be a sub σ–algebra of $\mathcal{F}$, i.e a σ–algebra of subsets of Ω such that $\mathcal{G} \subseteq \mathcal{F}$ and let $X : \Omega \to \mathbb{R}$ be an integrable random variable.

Definition 1.1. A **conditional expectation of X with respect to $\mathcal{G}$** is a random variable Y such that

1) Y is $\mathcal{G}$-measurable;

2) For any $A \in \mathcal{G}$ we get $\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A Y]$.

While for integrable random variable the conditional expectation always exist, it might be non-unique, i.e. more than one random variable might satisfy properties from Definition 1.1. Nevertheless, it is defined uniquely up to a set of zero measure.[1] For transparency, we use symbol $\mathbb{E}[X|\mathcal{G}]$ to denote the conditional expectation (keeping in mind that we operate on the space $L_1(\Omega, \mathcal{F}, \mathbb{P})$ of equivalent classes of random variables, rather than on the space of all random variables, and all the inequalities are understood in almost-sure sense).

Next, we define a conditional expectation of X with respect to another random variable Y.

Definition 1.2. Let X and Y be a random variable. We denote by $\mathbb{E}[X|Y]$ the **conditional expectation of X with respect to Y** given by

$$\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)],$$

where $\sigma(Y)$ is a σ–algebra generated by Y.

From Definition 1.1 we know that $\mathbb{E}[X|Y]$ is $\sigma(Y)$ measurable. Thus, from Doob–Dynkin lemma it follows that

$$\mathbb{E}[X|Y] = g(Y),$$

for some Borel measurable function $g\colon \mathbb{R} \to \mathbb{R}$. From now on, we use the convention

$$\mathbb{E}[X|Y = y] := g(y), \quad y \in \mathbb{R}.$$

Before we outline basic properties of conditional expectation, let us present two examples of conditional expectations.

[1] That is, if Y and V satisfy all the properties from Definition 1.1, then $\mathbb{P}[Y = V] = 1$.

Example 1.3. Let $\mathcal{G} = \sigma(\{A_1, A_2, \ldots, A_n\})$, where $\{A_i\}_{i=1}^n$ is a finite partition of Ω.[2] Then, for any integrable random variable X we get the representation

$$\mathbb{E}[X|\mathcal{G}](\omega) = \begin{cases} \frac{1}{\mathbb{P}[A_i]}\mathbb{E}[\mathbb{1}_{A_i}X] & \text{if } \omega \in A_i \text{ for which } \mathbb{P}[A_i] > 0; \\ 0 & \text{if } \omega \in A_i \text{ for which } \mathbb{P}[A_i] = 0. \end{cases} \tag{1.1}$$

Note that in Equation (1.1) we can substitute 0 with any real number, for any $i \in \{1, \ldots, n\}$ such that $\mathbb{P}[A_i] = 0$. Also, using the convention $\frac{0}{0} = 0$ we can rewrite (1.1) as

$$\mathbb{E}[X|\mathcal{G}] = \sum_{i=1}^{n} \frac{\mathbb{E}[\mathbb{1}_{A_i}X]}{\mathbb{P}[A_i]}\mathbb{1}_{A_i}.$$

Proof. We divide the proof of (1.1) into a few steps.
1) First of all, let us show that

$$\mathcal{G} = \left\{ \bigcup_{i\in I} A_i : I \subseteq \{1, 2, \ldots, n\} \right\}. \tag{1.2}$$

Let $\tilde{\mathcal{G}}$ denote the RHS of (1.2). Because $\tilde{\mathcal{G}}$ contains $\{A_1, \ldots, A_n\}$ we get that $\mathcal{G} \subseteq \tilde{\mathcal{G}}$. On the other hand we know that for any $I \subseteq \{1, \ldots, n\}$ we get $\bigcup_{i\in I} A_i \in \mathcal{G}$, so $\bar{\mathcal{G}} \subseteq \mathcal{G}$. Thus, (1.2) is proved.
2) Next, we show that random variable Y is $\mathcal{G}$–measurable if and only if Y is constant on any A_i. On the contrary, let us assume that there exists Y and i_0 such that Y is $\mathcal{G}$–measurable and Y is not constant on A_{i_0}. Then, there exists $x \in \mathbb{R}$ such that for $B = (-\infty, x]$ we get

$$Y^{-1}(B) \cap A_{i_0} \neq \emptyset \quad \text{and} \quad Y^{-1}(B^c) \cap A_{i_0} \neq \emptyset.$$

This contradicts (1.2), because $Y^{-1}(B) \cap A_{i_0}$ is a (proper) subset of A_{i_o}, which belongs to $\mathcal{G}$.
3) Finally, we are now ready to prove (1.1). From the previous steps we know that

$$\mathbb{E}[X|\mathcal{G}] = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i},$$

for some sequence of real numbers $a_1, \ldots, a_n$. Now, using Definition 1.1 and noting that for $i \neq j$ we have $\mathbb{1}_{A_i}\mathbb{1}_{A_j} = \mathbb{1}_{A_i \cap A_j} = 0$, we get

$$\mathbb{E}[\mathbb{1}_{A_i}X] = \mathbb{E}\left[\mathbb{1}_{A_i}\left(\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}\right)\right] = \mathbb{E}\left[\mathbb{1}_{A_i}a_i\right] = \mathbb{P}[A_i]a_i,$$

which concludes the proof of (1.1). □

Example 1.4. Let Y be a random variable that takes finitely many values $\{y_1, \ldots, y_n\}$ – all with positive probability. Then, for any integrable random variable X we get

$$\mathbb{E}[X|Y] = \sum_{i=1}^{n} \frac{\mathbb{E}[\mathbb{1}_{\{Y=y_i\}}X]}{\mathbb{P}[Y = y_i]}\mathbb{1}_{\{Y=y_i\}}. \tag{1.3}$$

Alternatively, we can rewrite (1.3) as

$$\mathbb{E}[X|Y = y] = \begin{cases} \frac{1}{\mathbb{P}[Y=y_i]}\mathbb{E}[\mathbb{1}_{\{Y=y_i\}}X] & \text{if } y = y_i \text{ for some } i \in \{1, \ldots, n\}, \\ 0 & \text{if } y \notin \{y_1, \ldots, y_n\}. \end{cases}$$

[2] A collection of sets such that $A_i \cap A_j = \emptyset$ for $i \neq j$, and $\Omega = A_1 \cup \ldots \cup A_n$.

Proof. We know that

$$\sigma(Y) = \sigma(\{A_1, \dots, A_n\})$$

where $A_i := \{Y = y_i\}$ for $i = 1, \dots, n$. Moreover, it is easy to note that $\{A_1, \dots, A_n\}$ is a partition of Ω. Thus, the proof follows from the proof of Example 1.3. □

For brevity, we state all the properties/theorems in this section without proofs. We start with some basic properties of conditional expectation.

Theorem 1.5 (Basic properties)**.** *Let X and Y be integrable random variables and let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then,*

1) If X is $\mathcal{G}$–measurable, then $\mathbb{E}[X|\mathcal{G}] = X$;

2) If X is $\mathcal{G}$–measurable and XY is integrable, then $\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}]$;

3) If X is independent of $\mathcal{G}$,[3] then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$;

4) If X is independent of $\mathcal{G}$, and $\mathcal{H}$ is a sub σ–algebra of $\mathcal{F}$, then $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{H}]$;

5) If $\mathcal{G} = \{\Omega, \emptyset\}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$;

6) For any $a, b \in \mathbb{R}$ we get $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$;

7) If $\mathcal{H}$ is sub σ-field of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\,\mathbb{E}[X|\mathcal{G}]\,|\mathcal{H}]$;

8) If $X \geq 0$, then $\mathbb{E}[X|\mathcal{G}] \geq 0$;

9) If $X \geq Y$, then $\mathbb{E}[X|\mathcal{G}] \geq \mathbb{E}[Y|\mathcal{G}]$;

1.2 Useful classical theorems

Next, we present a couple of classical results transferred from the static to the conditional case. For brevity, we omit most of the proofs – they are very similar to the unconditional ones that could be found in any good differential and integral calculus book.

Theorem 1.6 (Conditional monotone convergence theorem)**.** *Let (X_n) be a non-negative non-decreasing sequence of random variables that converge (almost-surely) to X. Then, for any sub σ-algebra $\mathcal{G}$ (of $\mathcal{F}$) we get*

$$\mathbb{E}[X_n|\mathcal{G}] \xrightarrow{a.s.} E[X|\mathcal{G}], \qquad n \to \infty.$$

Theorem 1.7 (Conditional dominated convergence theorem)**.** *Let (X_n) be a sequence of random variables that converge (almost-surely) to X, and let Y be an integrable random variable, such that for any $n \in \mathbb{N}$ we get $|X_n| < Y$. Then, random variable X is integrable and for any sub σ-algebra $\mathcal{G}$ (of $\mathcal{F}$) we get*

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}].$$

Theorem 1.8 (Conditional Fatou lemma)**.** *Let (X_n) be a non-negative sequence of integrable random variables. Then, for any sub σ–algebra $\mathcal{G}$ (of $\mathcal{F}$) we get*

$$\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}].$$

[3] That is $\sigma(X)$ and $\mathcal{G}$ are independent: for any $A \in \sigma(X)$ and $B \in \mathcal{G}$, we get $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$.

Finally, let us note that if X is square integrable (i.e. $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$), then the conditional expectation wrt. sub σ–algebra $\mathcal{G}$ could be seen as the orthogonal projection from $L^2(\Omega, \mathcal{F}, \mathbb{P})$ to $L^2(\Omega, \mathcal{G}, \mathbb{P})$.[4] This implies that $E[X|\mathcal{G}]$ is the least-square best $\mathcal{G}$–measurable predictor (that is also square integrable). In particular, If $\mathcal{G}$ relates to knowledge about a system (e.g. stock market) and X is a random variable corresponding to some system-related process (e.g. stock price), then we know that $E[X|\mathcal{G}]$ might be treated as the best guess about X given information $\mathcal{G}$; see Theorem 1.9.

Theorem 1.9 (Conditional expectation as the best least-square predictor)**.** *Let X be a square integrable random variable and let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then, for any square integrable $\mathcal{G}$–measurable random variable Z we get*

$$\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2] \leq \mathbb{E}[(X - Z)^2]. \tag{1.4}$$

Proof. Let Z be a square integrable and $\mathcal{G}$-measurable. Using the tower property, we get

$$\begin{aligned} \mathbb{E}[(X - Z)^2] &= \mathbb{E}[((X - \mathbb{E}[X|\mathcal{G}]) + (\mathbb{E}[X|\mathcal{G}]) - Z))^2] \\ &\geq \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2] + 2\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])(\mathbb{E}[X|\mathcal{G}]) - Z)] \\ &= \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2] - 2\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])Z] \end{aligned}$$

Next, we have

$$\begin{aligned} \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])Z] &= \mathbb{E}\big[\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])Z \,|\, \mathcal{G}]\big] \\ &= \mathbb{E}[Z\,\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}]) \,|\, \mathcal{G}]] \\ &= \mathbb{E}[Z\,(\mathbb{E}[X|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}])] \\ &= 0, \end{aligned}$$

which concludes the proof. □

2 Introduction to stochastic processes

In this section we use $\mathbb{T}$ to denote time. In the discrete case $\mathbb{T}$ is typically associated with the set of days or years, e.g. $\mathbb{T} = \{1, 2, \ldots, T\}$ for some fixed $T \in \mathbb{N}$, or $\mathbb{T} = \mathbb{N}$, or $\mathbb{T} = \mathbb{Z}$. In the continuous case $\mathbb{T}$ is usually linked to some fixed interval of $\mathbb{R}$, e.g. $\mathbb{T} = [0, T]$ for some fixed $T \in \mathbb{R}_+$, or $\mathbb{T} = \mathbb{R}_+$, or $\mathbb{T} = \mathbb{R}$.

2.1 Basic definitions and properties

First, we introduce the definition of a stochastic process.

Definition 2.1. A collection of random variables $X = (X_t)_{t\in\mathbb{T}}$ defined on a same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and indexed by time is called a **stochastic process**.

It should be noted that stochastic process could be seen as a function $X \colon \mathbb{T} \times \Omega \to \mathbb{R}$. We call X a **vector stochastic process** if it is a collection od random vectors indexed by time, and when the output is also random vector. For brevity we will always use the term **stochastic process**, even if we talk about random vectors rather than random variables. For any fixed $\omega \in \Omega$, one can see $(X_t(\omega))_{t\in\mathbb{T}}$ as a function of time – a specific realisation of the stochastic process.

[4] It could be shown that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space (with inner product $\langle X, Y \rangle = \mathbb{E}XY$), where we identify random variables which agree almost-surely, and $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is it's linear subspace.

Definition 2.2. Let X be a stochastic process. A **sample path (trajectory, realisation)** of X corresponding to $\omega \in \Omega$ is a function $t \to X_t(\omega)$, where $t \in \mathbb{T}$.

Usually, we require additional properties from the stochastic processes like the continuity of sample paths. Some of them are summarised in Definition 2.3. For brevity we present the properties in the continuous time setting. Note that while some of them might be extended directly to the discrete time case (e.g. integrability), others cannot (e.g. continuity).

Definition 2.3. Let X be a continuous-time stochastic process. We say that X is

1) **measurable** if the map $X : \mathbb{T} \times \Omega \to \mathbb{R}$ is $\mathcal{F} \times \mathcal{B}(\mathbb{T})$–measurable.
2) **continuous** (resp. left continuous, right continuous) if all sample paths of X are continuous (resp. left continuous, right continous).
3) **continuous in probability** (or *stochastically continuous*) if for any $t \in \mathbb{T}$ we get $X_s \to X_t$ in probability whenever $s \to t$, i.e. when

$$\forall_{\epsilon>0}\exists_{\delta>0} : \quad \mathbb{P}[\,|X_t - X_s| \geq \epsilon\,] \leq \epsilon, \quad \text{for } s \in \mathbb{T} \text{ such that } |t-s| \leq \delta;$$

4) **cádlág** if X is right continuous and all sample paths have left limits for any $t \in \mathbb{T}$.[a]
5) **integrable** if $\mathbb{E}|X_t| < \infty$ for any $t \in \mathbb{T}$.
6) **square-integrable** if $\mathbb{E}|X_t|^2 < \infty$ for any $t \in \mathbb{T}$.
7) **p-power integrable** if $\mathbb{E}|X_t|^p < \infty$ for $p \in \mathbb{N}$ and any $t \in \mathbb{T}$.

[a] fr. *continu á droite at limites á gauche*

Now, we want to identify processes for which almost all trajectories are the same.

Definition 2.4. Let X, Y be two stochastic process defined on the same probability space. We say that

1) Y is **indistinguishable** from X if $\mathbb{P}[X_t = Y_t\,,\,\forall t \in \mathbb{T}] = 1$.
2) Y is a **modification** of X if for any $t \in \mathbb{T}$ we get $\mathbb{P}[X_t = Y_t] = 1$.
3) Y has the **same finite-dimensional distribution** as X if for any $n \in \mathbb{N}$, finite number of time points $(t_1, \ldots, t_n) \in \mathbb{T}^n$, and $A \in \mathcal{B}(\mathbb{R}^n)$, we get

$$\mathbb{P}[(X_{t_1}, \ldots, X_{t_n}) \in A] = \mathbb{P}[(Y_{t_1}, \ldots, Y_{t_n}) \in A].$$

Sometimes, instead of the term *modification* we say that X is a *version* of Y. Note that if X and Y are modifications of each other and are (a.s.) continuous, they are indistinguishable. Let us also show an example of two processes which are modifications of each other but are not indistinguishable.

Example 2.5. Let $\mathbb{T} = [0, 1]$ and let $\Omega = [0, 1]$ be a standard probability space. Let X be such that $X_t \equiv 0$ for all $t \in \mathbb{T}$ and let Y be given by

$$Y_t(\omega) = \begin{cases} 1 & \text{if } t = \omega, \\ 0 & \text{if } t \neq \omega. \end{cases}$$

Then, for any $t \in \mathbb{T}$ we get $\mathbb{P}[X_t = Y_t] = \mathbb{P}[\Omega \setminus \{t\}] = 1$, and $\mathbb{P}[X_t = Y_t\,,\, \forall t \in \mathbb{T}] = 0$.

In the end of this subsection let us define a three properties, that are often used to characterise certain classes of stochastic processes (most of the properties could be translated into discrete-time framework).

Definition 2.6. Let X be a continuous-time stochastic process. We say that X

1) is **stationary** if for any time-shift $h \in \mathbb{T}$ the processes $(X_t)_{t\in\mathbb{T}}$ and $(X_{t+h})_{t\in\mathbb{T}}$ have the same finite-dimensional distributions.

2) is **weakly stationary** if for any time-shift $h \in \mathbb{T}$ the processes $(X_t)_{t\in\mathbb{T}}$ and $(X_{t+h})_{t\in\mathbb{T}}$ have finite-dimensional distributions with the same (finite) first and second moments.

3) has **independent increments** if for any finite set of time-points $t_1 \leq t_2 \leq \ldots \leq t_n$ (from $\mathbb{T}$), the incremental random variables

$$X_{t_2} - X_{t_1}, \quad X_{t_3} - X_{t_2}, \quad \ldots, \quad X_{t_n} - X_{t_{n-1}}$$

are independent.

Moreover, if the distribution of the increment $X_t - X_s$ depends only on $t - s$, then we say that X has **independent and stationary increments**.

Note that while property 1) is more strict and we often require it for theoretical purposes, the property 2) is much easier to check and/or verify. It is especially important in signal processing. Note that weak stationarity implies that the mean value of the process must be the same for any time-point, and the auto-covariance function given by

$$C_X(t,s) := \operatorname{Cov}(X_t, X_s),$$

where $t, s \in \mathbb{T}$, depends only on the difference between time-points t and s.

2.2 Continuous modifications and Kolmogorov's theorems

We show that if a process has a continuous modification, then it must be continuous in probability.

Proposition 2.7. *Let X be a stochastic process for which there exists a continuous modification. Then X is continuous in probability.*

Proof. Let us fix $t \in \mathbb{T}$ and $\epsilon > 0$. Let

$$A_n := \{\omega \in \Omega : \text{ there exists } s \in \mathbb{T} \text{ such that } |t - s| \leq \tfrac{1}{n} \text{ and } |\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| \geq \epsilon\}.$$

From the continuity of $\tilde{X}$ we know that A_n is measurable since it could be expressed as

$$A_n := \{\omega \in \Omega : \text{ there exists } s \in \mathbb{T} \cap \mathbb{Q} \text{ such that } |t - s| \leq \tfrac{1}{n} \text{ and } |\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| \geq \epsilon\}.$$

Also, we know that $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence[5] such that $\mathbb{P}[\bigcap_{n\in\mathbb{N}} A_n] = 0$. Thus, there exists $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}[A_{n_0}] \leq \epsilon. \tag{2.1}$$

Next, because $\tilde{X}$ is the modification of $\tilde{X}$ and $\tilde{X}_t - \tilde{X}_s = (\tilde{X}_t - X_t) + (X_t - X_s) + (X_s - \tilde{X}_s)$ we get

$$\mathbb{P}[|\tilde{X}_t - \tilde{X}_s| \geq \epsilon] = \mathbb{P}[|X_t - X_s| \geq \epsilon]. \tag{2.2}$$

Combining (2.1) with (2.2) for any $\delta < \frac{1}{n_0}$ and any $s \in \mathbb{T}$ such that $|t - s| \leq \delta$ we get

$$\mathbb{P}[|X_t - X_s| \geq \epsilon] = \mathbb{P}[|\tilde{X}_t - \tilde{X}_s| \geq \epsilon] \leq \mathbb{P}[A_{n_0}] \leq \epsilon,$$

which concludes the proof. □

From Proposition 2.7 we know that any continuous process is continuous in probability. It should be also noted that any modification of a continuous in probability process is also continuous in probabiilty. Let us now show an example of process which is continuous in probability but does not have a continuous modification.

Example 2.8. Let $\mathbb{T} = [0, 1]$ and let $Z \sim U[0, 1]$. Let $X = (X_t)_{t\in\mathbb{T}}$ be given by

$$X_t(\omega) = \mathbb{1}_{[0,Z(\omega))}(t),$$

for $t \in [0, 1]$ and $\omega \in \Omega$. One can easily check that there exists no continuous modification of X but X is continuous in probability.

Before we state (without the proof) two simplified versions of the Kolmogorov's continuity theorem that is very useful in the theory of stochastic processes let us recall a concept of Hölder continuity.

Definition 2.9. We say that a function $f : I \to \mathbb{R}$ is **Hölder continuous** on $I = [a, b]$ with exponent $\alpha \in (0, 1)$ if there exists a positive constant $C \in \mathbb{R}$ such that for any $x, y \in I$ we get

$$|f(x) - f(y)| \leq C|x - y|^{\alpha}.$$

Of course any Hölder continuous function is continuous. Also, if f is Hölder continuous with exponent $\alpha \in (0, 1)$, then it is Hölder continuous with exponent γ, for any $\gamma \leq \alpha$.

Theorem 2.10 (Kolmogorov's continuity theorem 1)**.** *Let $\mathbb{T} = [a, b]$ and let X be a stochastic process. If there exist constants $p, K, \epsilon > 0$ such that for any $t, s \in \mathbb{T}$ we get*

$$\mathbb{E}|X_t - X_s|^p \leq K|t - s|^{1+\epsilon}, \tag{2.3}$$

then X has a modification, which is Hölder continuous for any exponent $\alpha \in (0, \epsilon/p)$.

While the condition $\mathbb{T} = [a, b]$ might look restrictive, it is often used to prove the existence of continuous modifications. Thoerem 2.11 is a direct consequence of Theorem 2.10.

Theorem 2.11 (Kolmogorov's continuity theorem 2)**.** *Let $\mathbb{T} = \mathbb{R}_+$ and let X be a stochastic process. Let us assume that for any $T \in \mathbb{T}$ there exist constants $p, K, \epsilon > 0$ such that (2.3) is satisfied for $s, t \leq T$. Then, there exists a continuous modification of X.*

[5] i.e. $A_n \supseteq A_{n+1}$ for any $n \in \mathbb{N}$

We are now ready to provide the second (type of) Kolmogorov's theorem - the extension theorem. Given, a finite-dimensional distributions for all finite collections of time-points we want to check if there exists a stochastic process, which admits those distributions. To formulate the result, we need to formalise the concept of finite-distributions.

Definition 2.12. Let X be a stochastic process. The mapping $\mathbb{P}_X : \mathcal{B}_{\mathbb{T}} \to [0,1]$, which characterises the **finite-dimensional distributions** of X, is given by

$$\mathbb{P}_X[A] := \mathbb{P}\left[\{\omega \in \Omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in \Gamma\}\right],$$

where $A = \{x \in \mathbb{R}^{\mathbb{T}} : (x_{t_1}, \ldots, x_{t_n}) \in \Gamma\}$ is a cylinder set[a], and $\mathcal{B}_{\mathbb{T}}$ is a family of all cylinder sets on $\mathbb{T}$.

[a]we call $A \subseteq \mathbb{R}^{\mathbb{T}}$ a **cylinder set** if $A = \{x \in \mathbb{R}^{\mathbb{T}} : (x_{t_1}, \ldots, x_{t_n}) \in \Gamma\}$, where $n \in \mathbb{N}$, the increasing sequence $(t_i)_{i=1}^n$ is a sequence of (finite) time-points of $\mathbb{T}$, and Γ is a Borel-measurable set on $\mathbb{R}^n$.

Note that while $\mathcal{B}_{\mathbb{T}}$ is always an algebra, it is not necessarily a σ-algebra. Given the auxiliary mapping $\mathbb{P}_X$ defined on cylinder sets $\mathcal{B}_{\mathbb{T}}$ we want to be able to (uniquely) extend it to a probability measure on $\mathcal{B}(\mathbb{R}^{\mathbb{T}})$ in order to get the distribution of the corresponding stochastic process. In other words, we want to check if the mapping $\mathbb{P}_X$ can be used to characterise X. This can be done if only the mapping preserves the *consistency* property. Before we state the main result, we introduce some additional notation.

Given $\mathbb{T}$, let $\mathcal{T}$ denote the set of all finite subsets of $\mathbb{T}$ with partial inclusion-order. Let us assume we are given a family $(\mathbb{P}_I)_{I\in\mathcal{T}}$ of probability measures $\mathbb{P}_I$ defined on $(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I))$. Then, for $I_1 \subset I_2 \subset \mathbb{T}$ and $A \in \mathcal{B}(\mathbb{R}^{I_1})$ we define

$$\begin{aligned} C_{I_2,I_1}(A) &:= \{\omega \in \mathbb{R}^{I_2} : \{\omega(t); t \in I_1\} \in A\}, \\ C_{I_1}(A) &:= \{\omega \in \mathbb{R}^{\mathbb{T}} : \{\omega(t); t \in I_1\} \in A\}. \end{aligned}$$

Note that $C_{I_2,I_1} : \mathcal{B}(\mathbb{R}^{I_1}) \to \mathcal{B}(\mathbb{R}^{I_2})$ and $C_{I_1} : \mathcal{B}(\mathbb{R}^{I_1}) \to \mathcal{B}(\mathbb{R}^{\mathbb{T}})$. We say that a family $(\mathbb{P}_I)_{I\in\mathcal{T}}$ is **consistent** if for any $I_1 \subset I_2 \subset \mathcal{T}$ we get

$$\mathbb{P}_{I_1}[A] = \mathbb{P}_{I_2}[C_{I_2,I_1}(A)]$$

for all $A \in \mathcal{B}(\mathbb{R}^{I_1})$. Now, we are ready to state the main Theorem.

Theorem 2.13 (Kolmogorov's extension theorem)**.** *Let $(\mathbb{P}_I)_{I\in\mathcal{T}}$ be a consistent family of measures. Then, there exists a unique probability measure $\mathbb{P}$ defined on $(\mathbb{R}^{\mathbb{T}}, \mathcal{B}(\mathbb{R}^{\mathbb{T}}))$ such that*

$$\mathbb{P}[C_I(A)] = \mathbb{P}_I[A],$$

for any $I \in \mathcal{T}$ and $A \in \mathcal{B}(\mathbb{R}^I)$.

Note that the consistency property from Theorem 2.13 is very natural. We simply want to preserve the distribution if we exclude some sets of time points from I_2 and restrict ourselves to I_1. The theorems states that given stochastic process X and the corresponding mapping $\mathbb{P}_X$, we can extend this mapping to the probability measure on $\mathbb{R}^{\mathbb{T}}$ (defined on σ-algebra of Borel sets on $\mathbb{R}^{\mathbb{T}}$).

2.3 Filtrations and adaptiveness

We start with basic definitions.

Definition 2.14. A **filtration** is a non-decreasing family of sub σ–algebras of $\mathcal{F}$ indexed by time, i.e. a family $\mathbb{F} := (\mathcal{F}_t)_{t\in\mathbb{T}}$ such that

$$\mathcal{F}_s \subseteq \mathcal{F}_t,$$

for $s \leq t$, where $t, s \in \mathbb{T}$.

Usually $\mathcal{F}_t$ corresponds to our knowledge about the system up to time t. For brevity, we call the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathbb{F}$ a *filtered probability space*. In continuous time, to guarantee smoothness, usually a right continuity axiom is imposed on filtration.

Definition 2.15. Let $\mathbb{F}$ be a (continuous time) filtration. We say that $\mathbb{F}$ is the **right-continuous filtration** if for any $t \in \mathbb{T}$ we get

$$\mathcal{F}_t = \mathcal{F}_{t_+},$$

where $\mathcal{F}_{t_+} := \bigcap_{s>t, s\in\mathbb{T}} \mathcal{F}_s$.

In stochastic process theory filtered probability spaces are often assumed to satisfy *usual conditions*. It simply means that the filtration is right-continuous and complete.[6]

As stated before, the filtration represents our overall knowledge about a system, while the stochastic process is the realisation of some specific system-related process. To formalise this intuition, we need to introduce the concept of adaptiveness.

Definition 2.16. A process X is said to be **adapted** to filtration $\mathbb{F}$ (or $\mathbb{F}$**-adapted**) if X_t is $\mathcal{F}_t$–measurable for any $t \in \mathbb{T}$.

Of course, our overall knowledge about the system might be embedded into the process, which is reflected in the next definition.

Definition 2.17. Let X be a stochastic process. We say that $\mathbb{F}^X := (\mathcal{F}^X_t)_{t\in\mathbb{T}}$, where

$$\mathcal{F}^X_t = \sigma(X_s, s \leq t, s \in \mathbb{T})$$

is a filtration **generated** by stochastic process X.

In other words, a filtration is generated by a stochastic process X if for any $t \in \mathbb{T}$, the σ-algebra $\mathcal{F}^X_t$ is the smallest σ–algebra such that X_s is $\mathcal{F}^X_t$–measurable, for any $s \in \mathbb{T}$, such that $s \leq t$. It should be noted that filtrations generated by right-continuous processes are in general not necessarily right-continuous.[7] The opposite implication is also not true, i.e. the filtration generated by the process which is not right-continuous could be right-continuous.[8] Nevertheless, under some additional assumptions (e.g. Feller property) it is the case.

Now, knowing the concept of filtration we can define a progressively measurable process.

[6] i.e. both $\mathcal{F}$ and $\mathcal{F}_t$, for $t \in \mathbb{T}$ contain all $\mathbb{P}$-null sets; recall that a probability space $(\Omega, \Sigma, \mathbb{P})$ is called complete if and only if for any $S \subset \Omega$ such that $S \subseteq N$ for some zero-measure set N ($N \in \Sigma$ and $\mathbb{P}[N] = 0$) we get $S \in \Sigma$.

[7] Let $X = (X_t)_{t\in[0,1]}$ be given by $X_t = tZ$, for some fixed random variable $Z \sim U[0, 1]$.

[8] Let Z be a strictly positive random variable and let $X = (X_t)_{t\in\mathbb{R}_+}$ be such that $X_t = tZ$ for $t \in [0, 1] \cup (2, +\infty)$ and $X_t = 0$ otherwise.

Definition 2.18. A process X defined on a filtered probability space is called **progressively measurable** if it is $\mathbb{F}$-adapted and for any $t \in \mathbb{T}$ the map $(\mathbb{T} \cap (-\infty, t]) \times \Omega \to \mathbb{R}$ defined by $(s, \omega) \to X_s(\omega)$ is $\mathcal{B}(\mathbb{T} \cap (-\infty, t]) \times \mathcal{F}_t$ measurable.

Intuitively speaking, we want the stochastic process terminated at time $t \in \mathbb{T}$ to be measurable with respect to information up to time t. As we show in the next example, the concepts of *measurability* and *progressive measurability* are not equivalent.

Example 2.19. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a standard probability space with $\Omega = [0, 1]$.[9] Let

$$\mathcal{A} := \sigma(N \subset [0, 1] : \#N < \infty)$$

denote the σ–algebra of countable sets (and their complements). For time horizon $\mathbb{T} = [0, +\infty)$ we define filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ by setting

$$\mathcal{F}_t := \begin{cases} \mathcal{A} & \text{for } t \in [0, 1); \\ \mathcal{F} & \text{for } t \in [1, \infty). \end{cases}$$

Next, we define a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ by setting

$$X_t(\omega) := \mathbb{1}_{\Delta}(t, \omega) = \begin{cases} 1 & \text{if } t = \omega \text{ and } t \leq 1/2 \\ 0 & \text{otherwise} \end{cases}, \qquad t \in \mathbb{T}, \omega \in \Omega.$$

where $\Delta := \{(t, t) : t \in [0, \frac{1}{2}]\}$ is a subset of $\mathbb{T} \times \Omega$.

It is easy to observe that X is a measurable process, since $\Delta \in \mathcal{B}([0, 1]) \times \mathcal{F}$. It is also $\mathbb{F}$-adapted since for any $t \geq 0$ and $A \in \mathcal{B}(\mathbb{R})$ we get

$$X_t^{-1}(A) = \begin{cases} \emptyset & \text{if } t \in [0, +\infty),\ 0 \notin A, \text{ and } 1 \notin A, \\ \{t\} & \text{if } t \in [0, 1/2],\ 0 \notin A, \text{ and } 1 \in A, \\ [0, 1] \setminus \{t\} & \text{if } t \in [0, 1/2],\ 0 \in A, \text{ and } 1 \notin A, \\ [0, 1] & \text{if } t \in [0, 1/2],\ 0 \in A, \text{ and } 1 \in A, \\ [0, 1] & \text{if } t \in [1/2, +\infty),\ 0 \in A. \end{cases}$$

and consequently $X_t^{-1}(A) \in \mathcal{F}_t$.

Now, we show that X is not progressively measurable. Let us fix $T = 1/2$ and show that

$$\Delta \notin \mathcal{B}([0, 1/2]) \times \mathcal{F}_{1/2}, \tag{2.4}$$

i.e. $\Delta \notin \mathcal{B}([0, 1/2]) \times \mathcal{A}$. Assuming (2.4) is not true then we would get

$$\Delta \in \sigma(A_n : n \in \mathbb{N}) \times \sigma(D_n : n \in \mathbb{N}),$$

for some sequences $(A_n)_{n \in \mathbb{N}}$ and $(D_n)_{n \in \mathbb{N}}$ of subsets of $\mathcal{B}([0, 1/2])$ and $\mathcal{A} \cap [0, 1/2]$, respectively. Thus, from Foubini theorem and the definition of Δ we immediately get that for any $t \in [0, 1/2]$ we get

$$\{t\} = \{\omega \in \Omega : (\omega, t) \in \Delta\} \in \sigma(D_n : n \in \mathbb{N}).$$[10]

[9] i.e. $([0, 1], \mathcal{B}([0, 1]), \mathcal{L})$, where $\mathcal{L}$ is the standard Lebesgue measure on $[0, 1]$.

[10] It might be easier to understand, if one draw Δ in $[0, 1]^2$ square.

Setting $D := \bigcup_{n\in\mathbb{N}} D_n$ and noting that

$$\sigma(D_n : n \in \mathbb{N}) \subset \{A \in \Omega : A = \Gamma \text{ or } A = \Gamma \cup (\Omega \setminus D), \text{ where } \Gamma \subseteq D\},$$

we get that for any $t \in [0, 1/2]$ we have $\{t\} \subset \bigcup_{n\in\mathbb{N}} D_n$, which in turn implies

$$[0, 1/2] \subset D.$$

This contradicts the fact that the union of D_n's must be countable.

It is interesting to note that that the stochastic process $Y \equiv 0$ is a progressively measurable modification of X beacause

$$\mathbb{P}[X_t = 0] = \mathbb{P}[\{\omega \in \Omega : X_t(\omega) = 0] = \mathbb{P}[\Omega \setminus \{t\}] = 1.$$

In the end of this section we give define the property known as *predictability*.

Definition 2.20. Let $P_\mathbb{T}$ denote a σ-algebra on $\mathbb{T} \times \Omega$ that is generated by the left-continuous and adapted stochastic processes, i.e. the smallest σ-algebra that contain the sets

$$A \times (s, t], \qquad s, t \in \mathbb{T}, s < t, A \in \mathcal{F}_s.$$

Then,

- $P_\mathbb{T}$ is called the **predictable σ-algebra**.
- a stochastic process X is said to be **predictable** if it is $P_\mathbb{T}$–measurable.

Sometimes, it is also useful to define the *optional σ-algebra* denoted by $O_\mathbb{T}$, i.e. σ-algebra on $\mathbb{T} \times \Omega$ that is generated by the left-continuous and adapted stochastic processes. We call a stochastic process *optional* if it is $O_\mathbb{T}$–measurable.

2.4 Stopping times

Assuming that the stochastic process reflects the value of a game (or the price of the stock) we might be interested in the definition of stopping condition, e.g. when we win (or lose) a particular amount of money (the stock price pass a pre-specified threshold). This intuitive concept is embedded into the definition of stopping times (Markov moments). First, let us give a definition of a random time.

Definition 2.21. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. A random variable τ with values in time set $\mathbb{T} \cup \{+\infty\}$, i.e.

$$\tau : \Omega \to \mathbb{T} \cup \{+\infty\}$$

is called a **random time**. We call τ a **stopping time** (or $\mathbb{F}$–*stopping time*, or *Markov moment*) if

$$\{\tau \leq t\} \in \mathcal{F}_t.$$

for any $t \in \mathbb{T}$.

Intuitively speaking, a random time is a stopping time if we can decide if the event $\{\tau \leq t\}$ occurred (or not occured) basing only on the information available up to time $t \in \mathbb{T}$. This definition can be used to define a stopping rule which influences our control of the process (e.g. stop the game if your reward is bigger than a pre-specified number). Moreover, if not stated (explicitely) otherwise we we assume that all considered stopping times are *finite*, i.e. they satisfy the condition

$$\mathbb{P}[\{\tau = +\infty\}] = 0.$$

We do that mainly for technical reasons, and to present the statements of the results more transparently. Also, we use the convention

$$\inf \emptyset = +\infty,$$

i.e. if a stopping condition is not satisfied, then the value of the corresponding stopping time will be equal to $+\infty$. To better understand the concept of stopping times, let us provide a few examples.

Example 2.22 (First entry time)**.** Let $\mathbb{T} = \mathbb{N}$ and let X be an $\mathbb{F}$-adapted stochastic process. Then, for any $B \in \mathcal{B}(\mathbb{R})$ the random variable

$$\tau_B := \inf\{t \in \mathbb{T} : X_t \in B\}$$

is a stopping time. It could be interpreted as the first time of entry of X into B.

Proof. For any $t \in \mathbb{T}$ we get

$$\{\tau \leq t\} = \bigcup_{s\in\mathbb{T}:s\leq t} \{X_s \in B\}.$$

As X is $\mathbb{F}$–adapted we know that $\{X_s \in B\} \in \mathcal{F}_s$ and consequently $\{X_s \in B\} \in \mathcal{F}_t$. As the union is countable, the claim follows. □

One could also ask if the *last* hitting time is a stopping time, i.e. if

$$\rho = \sup\{t \in \mathbb{T} : X_t \in B\}.$$

Unfortunately it is **not** the case (despite some degenerate cases). Intuitively speaking, we would need to know the future in order to determine if X returns to B (proof is left as a simple exercise). Another natural question is if Example 2.22 could be translated to continuous time filtration. We show that it is indeed the case, under some additional technical conditions. Before we present the example, let us recall some basic facts related to stopping times.

First, we show that for stopping times there are additional types of $\mathcal{F}_t$-measurable events.

Lemma 2.23. *Let $\mathbb{F}$ be a filtration and let τ be a stopping time. Then, for any $t \in \mathbb{T}$ the events $\{\tau > t\}$, $\{\tau < t\}$, and $\{\tau > t\}$ belong to $\mathcal{F}_t$.*

Proof. For any $t \in \mathbb{T}$ we get $\{\tau \leq t\} \in \mathcal{F}_t$ and consequently $\{\tau \leq t\}^c \in \mathcal{F}_t$ (as $\mathcal{F}_t$ is a σ–algebra). Thus,

$$\{\tau > t\} = \{\tau \leq t\}^c \in \mathcal{F}_t.$$

Next, we get

$$\{\tau < t\} = \bigcup_{s<t, s\in\mathbb{Q}\cap\mathbb{T}} \{\tau \leq s\} \in \mathcal{F}_t,$$

where $\mathbb{Q}$ is a set of rational numbers. Finally,

$$\{\tau = t\} = \{\tau \leq t\} \setminus \{\tau < t\} \in \mathcal{F}_t.$$

□

Unfortunately, in continuous time usually we cannot replace condition $\{\tau \leq t\} \in \mathcal{F}_t$ with $\{\tau < t\} \in \mathcal{F}_t$ or $\{\tau = t\} \in \mathcal{F}_t$. Nevertheless, if we add some additional technical assumptions, then the definitions might be equivalent.

Lemma 2.24. *Let $\mathbb{T} = \mathbb{R}_+$ and let $\mathbb{F}$ be right continuous. Then, a random time τ is a stopping time if and only if for any $t \in \mathbb{T}$ we get*

$$\{\tau < t\} \in \mathcal{F}_t. \tag{2.5}$$

Proof. The first part of the proof (left implication) is given in Lemma 2.23. Let us assume that τ is such that (2.5) is satisfied. Then for any $t \in \mathbb{T}$ we get

$$\{\tau \leq t\} = \bigcap_{s>t, s\in\mathbb{Q}\cap\mathbb{T}} \{\tau < s\} \in \mathcal{F}_{t+}\,.$$

Since $\mathcal{F}_{t+} = \mathcal{F}_t$, the claim follows. □

Finally, w need to know that if $(\tau_n)_{n\in\mathbb{N}}$ is a sequence of stopping times, then so is the supremum of all τ_n's. Under continuity assumptions, same is true for infimum, limes inferior, and limes suprerior. Instead of writing $\sup_{n\in\mathbb{N}} \tau_n$ we use the notation $\sup \tau_n$ (noticing that supremum is ω-wise and taken with respect to n, rather than with respect to ω).

Lemma 2.25. *Let $(\tau_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{F}$–stopping times. Then, $\sup \tau_n$ is a stopping time. Moreover, if $\mathbb{F}$ is right-continuous, then $\inf \tau_n$, $\limsup \tau_n$, and $\liminf \tau_n$ are stopping times.*

Proof. It is easy to note that $\sup \tau_n$ is a random time. Moreover, for any $t \in \mathbb{T}$ we get

$$\{\sup \tau_n \leq t\} = \bigcap_{n\in\mathbb{N}} \{\tau_n \leq t\} \in \mathcal{F}_t.$$

Now, let us assume that $\mathbb{F}$ is right-continuous. Then, it is enough to show the strict inequalities; see Lemma 2.5. Taking the complement of $\{\inf \tau_n < t\}$, we get

$$\{\inf \tau_n \geq t\} = \bigcap_{n\in\mathbb{N}} \{\tau_n \geq t\} = \bigcap_{n\in\mathbb{N}} \{\tau_n < t\}^c \in \mathcal{F}_t,$$

so that $\{\inf \tau_n < t\} \in \mathcal{F}_t$. Next, we have

$$\{\limsup \tau_n < t\} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\tau_m < t - 1/k\} \in \mathcal{F}_t$$

$$\{\liminf \tau_n > t\} = \bigcup_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{\tau_m > t + 1/k\} \in \mathcal{F}_t$$

which concludes the proof. □

We are not finally ready to formulate the analog of Example 2.22 in continuous

Example 2.26. Let $\mathbb{T} = \mathbb{R}_+$ and let X be an $\mathbb{F}$-adapted stochastic process. Assume that X and $\mathbb{F}$ are right-continuous. Then, for any open (or closed) subset $B \in \mathcal{B}(\mathbb{R})$ the random variable

$$\tau_B := \inf\{t \in \mathbb{T} : X_t \in B\}$$

is a stopping time.

Proof. Let B be an open subset of $\mathbb{R}$. Due to Lemma 2.5 it is enough to operate on strict inequalities (and their complements). We know that for any $t \in \mathbb{T}$ we get

$$\{\tau_B \geq t\} = \{X_s \in B^c, s \in \mathbb{T}, s < t\}.$$

Now, because X is right-continuous we get

$$\{X_s \in B^c, s \in \mathbb{T}, s < t\} = \bigcap_{s<t, s\in\mathbb{T}\cap\mathbb{Q}} \{X_s \in B^c\} \in \mathcal{F}_t.$$

which concludes the proof for open subset. Now, let us assume that B is a closed subset of $\mathbb{R}$. For $\epsilon > 0$ let B_ϵ denote the ϵ-hull of B, i.e. the set of point for which the distance from D is strictly smaller than ϵ. Since B_ϵ is open, we know that τ_{D_ϵ} is a stopping time. Noting that

$$\tau_B = \lim_{\epsilon\to 0^+} \tau_{D_\epsilon}$$

and using Lemma 2.25 we get that τ_B is a stopping time, which concludes the proof. □

If the time is discrete, we can reformulate the definition of stopping time using equality instead of inequality.

Proposition 2.27. *Let $\mathbb{F}$ be a discrete-time filtration. Then τ is $\mathbb{F}$–stopping time if and only if*

$$\{\tau = t\} \in \mathcal{F}_t$$

for any $t \in \mathbb{T}$.

Proof. If τ is a stopping time, then

$$\{\tau = t\} = \{\tau \leq t\} \setminus \bigcup_{s\in\mathbb{T}, s<t} \{\tau \leq s\},$$

which proves our claim. On the other hand we get

$$\{\tau = t\} = \bigcup_{s\leq t, s\in\mathbb{T}} \{\tau = s\},$$

so the converse implication is also true. □

Finally, let us show that one can perform some basic operations on stopping times.

Proposition 2.28. *Let $\mathbb{F}$ be a filtration and let τ, ρ be two $\mathbb{F}$–stopping times. Then,*

1) If τ and ρ are non-negative, then $\tau + \rho$ is a stopping time.[11]

2) $\tau \wedge \rho := \min\{\tau, \rho\}$ is a stopping time.

3) $\tau \vee \rho := \max\{\tau, \rho\}$ is a stopping time.

[11] Assuming that $\rho + \tau$ has values in $\mathbb{T}$.

Proof. To prove 1) it is enough to note that

$$\{\tau+\rho>t\}=\Big[\{\tau=0\}\cap\{\rho>t\}\Big]\cup\Big[\bigcup_{s\in\mathbb{T}\cap\mathbb{Q}\cap[0,+\infty)}\{\tau>s\}\cap\{\rho>t-s\}\Big].$$

Next, to prove 2) and 3) it is enough to note that

$$\{\tau\wedge\rho\leq t\}=\{\tau\leq t\}\cap\{\rho\leq t\}$$

and

$$\{\tau\vee\rho\leq t\}=\{\tau\leq t\}\cup\{\rho\leq t\}.$$

□

Now, we outline some concepts and notation related to stopping times. For simplicity, from now on we assume that the considered stochastic processes are progressively measurable. As we have mentioned, stopping times might be used to stop a process or define a random sample from the process. Also, we can define the σ–algebra which provides the information up to a random time.

Definition 2.29. Let τ be a (finite) $\mathbb{F}$–stopping time and let X be a progressively $\mathbb{F}$-measurable stochastic process.

1) A **random time sample from stochastic process** X picked at τ, and denoted by X_τ, is given by

$$X_\tau(\omega):=X_{\tau(\omega)}(\omega),\qquad \text{for } \omega\in\Omega.$$

2) A **process stopped** at τ denoted $X^\tau=(X^\tau_t)_{t\in\mathbb{T}}$ is given by

$$X^\tau_t:=X_{t\wedge\tau},\qquad \text{for } t\in\mathbb{T}.$$

3) The **σ–algebra at a stopping time** τ, denoted by $\mathcal{F}_\tau$ is given by

$$\mathcal{F}_\tau:=\{A\in\mathcal{F}: A\cap\{\tau\leq t\}\in\mathcal{F}_t\,,\ \text{for any } t\in\mathbb{T}\}$$

Please note that in general the function X_τ might be not measurable at all (and consequently not $\mathcal{F}_\tau$–measurable). Nevertheless, under the progressive measurability it is the case. Also, we need to show that $\mathcal{F}_\tau$ is in fact a σ–algebra.

Proposition 2.30. *Let τ be a (finite) $\mathbb{F}$–stopping time. Then, the family of sets $\mathcal{F}_\tau$ is σ–algebra and τ is $\mathcal{F}_\tau$–measurable. Moreover, if the stochastic process X is progressively $\mathbb{F}$-measurable then X_τ is $\mathcal{F}_\tau$ measurable (and $X_{t\wedge\tau}$ is $\mathcal{F}_t$ measurable).*

Proof. From the definition of τ we get $\Omega\in\mathcal{F}_\tau$, as $\{\tau\leq t\}\in\mathcal{F}_t$ for any $t\in\mathbb{T}$. Now, let us assume that $A\in\mathcal{F}_\tau$. Then, for any $t\in\mathbb{T}$ we get

$$A^c\cap\{\tau\leq t\}=\{\tau\leq t\}\setminus(A\cap\{\tau\leq t\})\in\mathcal{F}_t,$$

and consequently $A^c\in\mathcal{F}_\tau$.

Next, we know that for a sequence $(A_n)_{n\in\mathbb{N}}$, where $A_n\in\mathcal{F}_\tau$, we get

$$(\bigcup_{n\in\mathbb{N}} A_n) \cap \{\tau \leq t\} = \bigcup_{n\in\mathbb{N}} (A_n \cap \{\tau \leq t\}) \in \mathcal{F}_t$$

for any $t \in \mathbb{T}$, which implies $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}_\tau$. This implies that $\mathcal{F}_\tau$ is indeed σ–algebra.

Now,to prove that τ in $\mathcal{F}_\tau$–measurable it is enough to show that for any $s \in \mathbb{R}$ the event $\{\tau \leq s\}$ belongs to $\mathcal{F}_\tau$. We can assume that $s \in \mathbb{T}$ (since τ has values in $\mathbb{T}$). Then, for $A_s = \{\tau \leq s\}$ we get

$$A_s \cap \{\tau \leq t\} = \{\tau \leq t \wedge s\} \in \mathcal{F}_{t\wedge s} \subseteq \mathcal{F}_t.$$

for any $t \in \mathbb{T}$, which concludes this part of the proof.

Finally, we need to show that for $\Gamma \in \mathcal{B}(\mathbb{R})$ and any $t \in \mathbb{T}$ we get

$$\{X_\tau \in \Gamma\} \cap \{\tau \leq t\} \in \mathcal{F}_t.$$

Let us fix $t \in \mathbb{T}$. Noting that $\tau \wedge t$ is a stopping time, we get that the map $Z : \Omega \to \Omega \times [0, t]$ given by $Z(\omega) = (\omega, \tau(\omega) \wedge t\}$, is $\mathcal{F}$–measurable (as its margins are measurable).[12] From the progressive measurability of X, we know that the map $W : \Omega \times [0, t] \to \mathbb{R}$ given by $W(\omega, s) = X_s(\omega)$, is $\mathcal{F}_t \times \mathcal{B}(\mathbb{R})$–measurable. Consequently, the map $V : \Omega \to \mathbb{R}$ given by

$$V(\omega) = W(Z(\omega)) = X_{\tau(\omega)\wedge t}(\omega),$$

is $\mathcal{F}_t$–measurable. Now, we note that

$$\{X_\tau \in \Gamma\} \cap \{\tau \leq t\} = \{X_{\tau\wedge t} \in \Gamma\} \cap \{\tau \leq t\} = V^{-1}(\Gamma) \cap \{\tau \leq t\} \in \mathcal{F}_t,$$

for any $t \in \mathbb{T}$ (as the union of two measurable events), which concludes the proof. □

If the time is discrete, then the progressive measurability is equivalent to measurability, so the claim follows immediately (the proof is left as a simple exercise).

2.5 Martingales

We begin this section with a general definition of the martingale property

Definition 2.31. Let X be $\mathbb{F}$-adapted integrable stochastic process. We say that

1) X is a **martingale** with respect to $\mathbb{F}$ if for any $s, t \in \mathbb{T}$, such that $s \leq t$ we get

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s;$$

2) X is a **submartingale** with respect to $\mathbb{F}$ if for any $s, t \in \mathbb{T}$, such that $s \leq t$ we get

$$\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s;$$

3) X is a **supermartingale** with respect to $\mathbb{F}$ if for any $s, t \in \mathbb{T}$, such that $s \leq t$ we get

$$\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s.$$

[12] note that the map does preserve $\mathcal{F}_t$–measurabilityas well.

Sometimes, for brevity we say that X is $\mathbb{F}$-martingale (resp. $\mathbb{F}$-submartingale, $\mathbb{F}$-supermartingale), or, if the underlying filtration is known, simply a martingale (resp. submartingale, supermartingale). Using the tower property, one can easily prove the following proposition.

Proposition 2.32. *Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space and let X be $\mathbb{F}$-adapted integrable stochastic process. Then*

1) If X is martingale, then $\mathbb{E}[X_t] = \mathbb{E}[X_s]$ for $t, s \in \mathbb{T}$;

2) If X is submartingale, then $\mathbb{E}[X_t] \geq \mathbb{E}[X_s]$ for $t, s \in \mathbb{T}$, such that $t \geq s$;

3) If X is supermartingale, then $\mathbb{E}[X_t] \leq \mathbb{E}[X_s]$ for $t, s \in \mathbb{T}$, such that $t \geq s$.

Given any random variable, it is easy to define the corresponding *regular martingale* by simply taking the conditional expectation.

Definition 2.33. Let X be a martingale. We call X a **regular martingale** If there exists a random variable η such that for any $t \in \mathbb{T}$ we get

$$X_t = \mathbb{E}[\eta|\mathcal{F}_t].$$

If the time is finite, then any martingale is regular. We leave the proof as an exercise.

Proposition 2.34. *Let $\mathbb{T} = [0, T]$ for some $T \in \mathbb{R}_+$*[13] *and let X be a martingale. Then $X_t = \mathbb{E}[X_T|\mathcal{F}_t]$ for any $t \in \mathbb{T}$.*

Of course, there exists martingales which are not regular. The simplest example is a random walk (with respect to its natural filtration). Finally, let us recall that Jensen's inequality might be useful to show the (sub)martingale property

Proposition 2.35. *Let X be a martingale and let $f\colon \mathbb{R} \to \mathbb{R}$ be a convex function. Then the process $(f(X_t))_{t \in \mathbb{T}}$ is a submartingale (assuming it is integrable).*

Proof. The proof follows easily from the Conditional Jensen's inequality. Indeed, for any $t, s \in \mathbb{T}$ such that $t > s$ we get

$$f(X_s) = f(\mathbb{E}[X_t|\mathcal{F}_s]) \geq \mathbb{E}[f(X_t)|\mathcal{F}_s].$$

□

We state the next result to show the interactions between stopping times and martingales. For simplicity, we present the proofs only in discrete time. The first result shows that under stopping, the martingale property is preserved.

Theorem 2.36. *Let $\mathbb{T}$ be discrete and let X be an $\mathbb{F}$–martingale (resp. $\mathbb{F}$-submartingale, $\mathbb{F}$-supermartingale). Then for any $\mathbb{F}$–stopping time τ the stopped process X^τ is an $\mathbb{F}$–martingale (resp. $\mathbb{F}$-submartingale, $\mathbb{F}$-supermartingale).*

Proof. For simplicity, let $\mathbb{T} = \mathbb{N}$. We only show the proof for submartingale process (from which other cases follow easily). As the time is discrete it is enough to check the submartingale condition

$$\mathbb{E}[X^\tau_{n+1}|\mathcal{F}_n] \geq X^\tau_n,$$

[13] or $\mathbb{T} = \{1, 2, \ldots, T\}$ for $T \in \mathbb{N}$

for all $n \in \mathbb{N}$. First, we know that for any $n \in \mathbb{N}$ we get

$$X_{n\wedge\tau} = \sum_{k=1}^{n-1} \mathbb{1}_{\{\tau=k\}} X_k + \mathbb{1}_{\{\tau\geq n\}} X_n,$$

Hence, the stopped process X^τ is integrable. Second, we get

$$X_{(n+1)\wedge\tau} - X_{n\wedge\tau} = \mathbb{1}_{\{\tau>n\}}(X_{n+1} - X_n), \tag{2.6}$$

and consequently

$$X_{n\wedge\tau} = X_1 + \sum_{k=1}^{n-1} (X_{(k+1)\wedge\tau} - X_{k\wedge\tau}) = X_1 + \sum_{k=1}^{n-1} \mathbb{1}_{\{\tau>k\}}(X_{k+1} - X_k).$$

Consequently, the process X^τ is $\mathbb{F}$–adapted. Finally, noting that $\{\tau > n\} = \{\tau \leq n\}^c \in \mathbb{F}_n$ and using (2.6) we get

$$\begin{aligned}
\mathbb{E}[X_{(n+1)\wedge\tau}|\mathcal{F}_n] &= \mathbb{E}[X_{n\wedge\tau} + X_{(n+1)\wedge\tau} - X_{n\wedge\tau}|\mathcal{F}_n] \\
&= \mathbb{E}[X_{n\wedge\tau} + \mathbb{1}_{\{\tau>n\}}(X_{n+1} - X_n)|\mathcal{F}_n] \\
&= X_{n\wedge\tau} + \mathbb{1}_{\{\tau>n\}}\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] \\
&\geq X_{n\wedge\tau}.
\end{aligned}$$

which concludes the proof. □

In fact, the martingale inequalities are true for stopping time, which is a consequence of the Doob's optional sampling theorem. Below, we present the simplified discrete-time version.

Theorem 2.37. *Let $\mathbb{T}$ be discrete and let X be an $\mathbb{F}$–martingale (resp. $\mathbb{F}$-submartingale or $\mathbb{F}$-supermartingale). Let $\tau_1 \leq \tau_2$ be two bounded $\mathbb{F}$–stopping times. Then, we get*

$$\mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}] = X_{\tau_1} \qquad (\textit{resp.} \ \leq \ \textit{or} \ \geq).$$

Proof. For simplicity let us assume that $\mathbb{T} = \mathbb{N}$ and show the proof for submartingale. From Proposition 2.30 we know that for $i = 1, 2$, the random variable X_{τ_i} is $\mathcal{F}_{\tau_i}$–measurable, and $\mathcal{F}_{\tau_i}$ is a proper σ–algebra.[14] Let $N \in \mathbb{N}$ be such that $\tau_2 \leq N$ (such number exists since τ_2 is bounded). We know that X_{τ_1} and X_{τ_2} are integrable since for $i = 1, 2$ we get

$$\mathbb{E}|X_{\tau_i}| \leq \sum_{n=1}^{N} E|X_n| < \infty.$$

We need to show that for any $A \in \mathcal{F}_{\tau_1}$ we get

$$\mathbb{E}[\mathbb{1}_A X_{\tau_1}] \leq \mathbb{E}[\mathbb{1}_A X_{\tau_2}].$$

Noting that

$$A = \bigcup_{k=1}^{N} A_k$$

[14] Note that time is discrete, so it is enough for X to be measurable and adapted.

where $A_k := \{\tau_1 = k\} \cap A$, and A_k's are $\mathcal{F}_{\tau_1}$–measurable, it is enough to show that

$$\mathbb{E}[\mathbb{1}_{A_k} X_{\tau_1}] \leq \mathbb{E}[\mathbb{1}_{A_k} X_{\tau_2}].$$

for any $k = 1, 2, \ldots, N$. Let us fix $k \in \mathbb{N}$ ($k \leq N$) and let

$$L(n) := \mathbb{E}[\mathbb{1}_{A_k} X_{n\wedge\tau_2}], \qquad \text{for } n = k, k+1, \ldots, N.$$

If we show that the function L is non-decreasing, then we would get (note that $\tau_1 \leq \tau_2$)

$$\mathbb{E}[\mathbb{1}_{B_k} X_{\tau_1}] = \mathbb{E}[\mathbb{1}_{B_k} X_{\tau_1\wedge\tau_2}] = \mathbb{E}[\mathbb{1}_{B_k} X_{k\wedge\tau_2}] = L(k) \leq L(N) = \mathbb{E}[\mathbb{1}_{B_k} X_{N\wedge\tau_2}] = \mathbb{E}[\mathbb{1}_{B_k} X_{\tau_2}]$$

which will conclude the proof. For any $n = k+1, \ldots, N$ we get

$$L(n+1) - L(n) = \mathbb{E}[\mathbb{1}_{A_k}(X_{(n+1)\wedge\tau_2} - X_{n\wedge\tau_2})] = \mathbb{E}[\mathbb{1}_{A_k\cap\{\tau_2>n\}}(X_{n+1} - X_n)].$$

Since X is a submartingale, it would be sufficient to show that for any n we get

$$A_k \cap \{\tau_2 > n\} \in \mathcal{F}_n.$$

From the definition of τ_2 we know that $\{\tau_2 > n\} \in \mathcal{F}_n$. Also, we know that

$$A_k = \{\tau_1 = k\} = A \cap \{\tau_1 \leq k\} \setminus A \cap \{\tau_1 \leq k-1\}.$$

As $A \in \mathcal{F}_{\tau_1}$,we know that $A \cap \{\tau_1 \leq k\} \in \mathcal{F}_k$ and $A \cap \{\tau_1 \leq k-1\} \in \mathcal{F}_{k-1}$. This concludes the proof as $k-1 < k \leq n$, and consequently $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k \subseteq \mathcal{F}_n$. □

Now, let us consider the continuous time setting. The analogues of Theorem 2.36 and Thoerem 2.37 are true for cádlág martingales. Before we explicitely state the results, let us provide a theorem, which states that for stochastically continuous martingales, the cádlág modification always exists. As mentioned, for brevity we state the results without the proofs.[15]

Theorem 2.38. *Let $\mathbb{T} = \mathbb{R}_+$ and let the stochastic process X be a stochastically continuous martingale (resp. supermartingale, submartingale). Then, there exists a cádlág modification of X.*

Theorem 2.39. *Let $\mathbb{T} = \mathbb{R}_+$ and let the stochastic process X be a cádlág martingale (resp. submartingale, supermartingale). Then, the stopped process X^τ is a martingale (resp. submartingale, supermartingale).*

Theorem 2.40. *Let $\mathbb{T} = \mathbb{R}_+$ and let the stochastic process X be a cádlág martingale (resp. submartingale, supermartingale). Let $\tau_1 \leq \tau_2$ be two bounded stopping times. Then, we get*

$$\mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}] = X_{\tau_1} \qquad (\textit{resp.} \ \leq, \geq).$$

2.5.1 Doob-Meyer decomposition

We start this section with the analogue of Doob-Meyer decomposition for discrete-time, where it can be presented in a much simpler form (so called Doob decomposition).

[15] see e.g. O. Kallenberg, *Foundations of modern probability*, Springer, New York 2002 and references therein for the proofs.

Theorem 2.41 (Doob decomposition)**.** *Let $\mathbb{T} = \mathbb{N}$ and let X be a submartingale (resp. supermartingale). Then we have a (unique) decomposition*

$$X_t = M_t + A_t,$$

for $t \geq 0$, where $M = (M_t)_{t\in\mathbb{T}}$ is a martingale, and $A = (A_t)_{t\in\mathbb{T}}$ is a predictable increasing (resp. decreasing) process starting from zero.[16]

Proof. Let X be a submartingale. Let A and M be given by

$$A_t = \sum_{k=1}^{t} (\mathbb{E}[X_k|\mathcal{F}_{k-1}] - X_{k-1}),$$

$$M_t = X_0 + \sum_{k=1}^{t} (X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}],$$

where $A_0 = 0$ and $M_0 = X_0$. One can easily see that for any $t \in \mathbb{N}$ we get

$$X_t = M_t + A_t.$$

Also, since X is a submartingale we get that A is increasing and starting from 0. Moreover, for any $t \in \mathbb{N}$ we get

$$\mathbb{E}[M_t - M_{t-1}|\mathcal{F}_t] = \mathbb{E}[X_t - \mathbb{E}[X_k|\mathcal{F}_{k-1}]] = 0.$$

so M is a martingale. The proof uniqueness is left as a simple excercise. □

Without the monotonicity assumptions imposed on A, the Doob decomposition could be in fact made for any integrable and adapted discrete-time stochastic process, and it is almost surely unique.

Now, we want to show that a similar theorem is true for continuous time. Before we state the main Theorem of this section, we need to recall the concept of uniform integrability.

Definition 2.42. Let $\mathbb{X} = \{X_\alpha\}_{\alpha\in I}$ be a family of random variables indexed by I. We say that a family $\mathbb{X}$ is **uniformly integrable** if

$$\forall_{\epsilon>0}\exists_{M>0} : \qquad \sup_{\alpha\in I} \mathbb{E}\left[\mathbb{1}_{\{|X_\alpha|>M\}}|X_\alpha|\right] < \epsilon.$$

Next, let us formally define the class of (D) and (DL) processes. To ease the notation for any $s \in \mathbb{T} \cup \{+\infty\}$ let Σ_s denote a family of all stopping times (with respect to the underlying filtration $\mathbb{F}$) with values smaller or equal to s, i.e.

$$\Sigma_s := \{\tau : \tau \leq s\}.$$

Definition 2.43. Let X be a right-continuous submartingale. We say that

- X is **of class (D)** if the family $\{X_\tau : \tau \in \Sigma_\infty\}$ is uniformly integrable,
- X is **of class (DL)** if for any $t \in \mathbb{T}$ the family $\{X_\tau : \tau \in \Sigma_t\}$ is uniformly integrable,

[16] In discrete time *predictability* is corresponding to the fact that X_t is $\mathcal{F}_{t-1}$ measurable.

One might look at (D) and (DL) class properties as generalisations of the concept of uniform integrability for stochastic processes. While Definition 2.43 might look strange, it is in fact generalisation of an intuitve property: From Theorem 2.40 we see that for a cádlág martingale process X, time index $t \in \mathbb{T}$, and bounded stopping time $\tau \leq t$, we get

$$\mathbb{E}[X_t|\mathcal{F}_\tau] = X_\tau. \tag{2.7}$$

As the sets of conditional expectations of a given random variable (with respect to various sub σ-fields) are uniformly integrable, the (DL) class property is satisfied. Indeed, we get the following result.

Proposition 2.44. *Let X be a martingale cádlág stochastic process. Then X is of class (DL).*

Proof. Let us fix $t \in \mathbb{T}$. As for any τ, such that $\tau \leq t$ we have (2.7) it is enough to prove that the family of conditional expectations

$$\{Y : Y = \mathbb{E}[X_t|\mathcal{B}] \text{ for some } \mathcal{B}, \text{ where } \mathcal{B} \text{ is a sub } \sigma\text{–algebra of } \mathcal{F}\}$$

is uniformly integrable. Let us fix $\epsilon > 0$. Since $|X_t|$ is integrable we know that there exists $\delta > 0$ such that for any $A \in \mathcal{F}$ satisfying $\mathbb{P}[A] \leq \delta$ we get

$$\mathbb{E}[\mathbb{1}_A|X_t|] < \epsilon.$$

By conditional Jensen's inequality for any $\mathcal{B}$ and $Y = \mathbb{E}[X_t|\mathcal{B}]$ we get

$$\mathbb{E}[|Y|] = \mathbb{E}[|\mathbb{E}[X_t|\mathcal{B}]|] \leq \mathbb{E}[\mathbb{E}[|X_t||\mathcal{B}]] = \mathbb{E}[|X_t|].$$

which implies (using Markov's inequality)

$$\mathbb{P}[|Y| > a] \leq \frac{1}{a}\mathbb{E}[|Y|] \leq \frac{1}{a}\mathbb{E}[|X_t|].$$

for any $a > 0$. Thus, setting $a := \frac{\mathbb{E}[|X_t|]}{\delta}$ we get

$$\mathbb{P}[A] \leq \delta,$$

for $A := \{|Y| > a\}$. Consequently, noting that $A \in \mathcal{B}$, we get

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_A|Y|\right] &\leq \mathbb{E}\left[\mathbb{1}_A|\mathbb{E}[X_t \mid \mathcal{B}]|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}_A\mathbb{E}[|X_t| \mid \mathcal{B}]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_A|X_t| \mid \mathcal{B}\right]\right] \\
&= \mathbb{E}[\mathbb{1}_A|X_t| \mid \mathcal{B}] \\
&\leq \epsilon.
\end{aligned}$$

As for the fixed $t \in \mathbb{T}$ the choice of a was independent of $\mathcal{B}$ we conclude the proof. ☐

Finally, we state (without the proof) the simplified version of Doob-Meyer decomposition theorem. Note that the statement is more general (i.e. for submartingale stochastic processes).

Theorem 2.45 (Doob-Meyer decomposition)**.** *Let X be a cádlág submartingale of class (DL). Then, we have a unique decomposition*

$$X_t = M_t + A_t,$$

for $t \geq 0$, where $M = (M_t)_{t\in\mathbb{T}}$ is a martingale, and $A = (A_t)_{t\in\mathbb{T}}$ is a predictable increasing process starting from zero.

The decomposition is often used to define the quadratic variation of X. We know that if X is a square-integrable cádlág martingale then X^2 is a submartingale of class (DL). The proof of this fact is left as as excercise (use Jensen's inequality!). Consequently, we can decompose X^2 using the Doob-Meyer decomposition, and the predictable part relates to the quadratic variation of X.[17]

Definition 2.46. Let X be a square-integrable cádlág martingale. The **quadratic variation of X** is the stochastic process $\langle X, X\rangle = (\langle X, X\rangle_t)_{t\in\mathbb{T}}$ defined by

$$\langle X, X\rangle_t := X_t^2 - M_t,$$

where $(M_t)_{t\in\mathbb{T}}$ is a martingale from Theorem 2.45.

3 Important examples of stochastic processes

In this section we introduce some important examples of stochastic processes. If not stated otherwise, we assume that $\mathbb{T} = \mathbb{R}_+$, i.e. time is continuous and we have a starting point.

Recall that in Definition 2.6 we introduced properties related to process stationarity and independence of increments. Using concepts related to those definitions, we can define certain important families of stochastic processes. Before we focus of specific examples, let us provide a general definition of a Levy process.

Definition 3.1. Let $\mathbb{T} = \mathbb{R}_+$. We say that a stochastic process X is a **Lévy process** if

1) $X_0 = 0$;

2) X is stochastically continuous;

3) X has independent and stationary increments.

The class of Lévy processes is very important from both theoretical and practical point of view. It could be viewed as the continuous-time analog of a random walk. In particular, note that for any Lévy process the increments over different (disjoint) time intervals of the same length must be independent and must have the same distribution. This class of processes satisfy many useful properties, such as the Markov property (we will formally state this result later). Also, there always exists a cádlág modification of a Lévy process. Let us state this fact for future's reference (without the proof; see e.g. Kinney's Theorem).

Proposition 3.2. *Let X be a* **Lévy process***. Then, there exists a cádlág modification of X.*

[17] Why it is called *quadratic variation* and why it is interesting? We will need it later to integrate by parts and to get stochastic change of variables formula, known as *Ito's lemma*.

Usually, we will additionally require increments to have a particular distribution, e.g. Gaussian or Poisson with time-dependant parameters. To proof the existence of a certain stochastic process given it's finite-dimensional distributions, we utilise the Kolmogorov's existence theorem. Then, Kolmogorov's continuity theorem will be used to pick a proper cádlág modification. To illustrate this, we show how this could be done for a Brownian motion.

3.1 Brownian motion

In this section we define and show basic properties of the Brownian motion, sometimes referred to as the Wiener process. This process is the central part of stochastic analysis and will be later used in the definition of Ito's integral. We start with the definition.

Definition 3.3. Let $\mathbb{T} = \mathbb{R}_+$. We call $W = (W_t)_{t\in\mathbb{T}}$ a (standard) **Brownian motion** (or **Wiener process**) if W satisfies the following properties:

1. $W_0 = 0$;
2. W is (a.s.) continuous;
3. W has independent increments;[a]
4. $W_t - W_s \sim N(0, \sqrt{t-s})$ for any $0 \leq s \leq t$.

[a] i.e. for $0 \leq t_1 < \dots t_n < \infty$ the random variables W_{t_1}, $W_{t_2} - W_{t_1}$, ... , $W_{t_{n-1}} - W_{t_n}$, are independent.

Straight from the definition we see that Brownian motion is an example of the Lévy process, where the increments are defined in terms of the Gaussian distribution with constant (zero) mean and time-varying volatility. Note that we require increments to be stationary, not the process itself. Before we show that the Brownian motion exists let us outline it's basic properties– note that for gaussian vectors, the first two moments characterise the whole distribution. We leave the proof as an excercise.

Proposition 3.4. *Let W be a Brownian motion. Then*

1. *W is a gaussian process;*[18]
2. *$\mathbb{E}[W_t] = 0$, for $t \in \mathbb{T}$;*
3. *$C_W(t,s) = \mathrm{Cov}(W_t, W_s) = t \wedge s$, for $t, s \in \mathbb{T}$.*[19]

It should be noted that the Brownian motion could be alternatively defined using properties given in Proposition 3.4, i.e. any gaussian continuous process starting at zero for which the first two moments satisfy the above conditions is the Brownian motion.

Assuming the probability space is reach enough, we can show that Brownian motion indeed exists. This is the statement of the next Theorem.

Theorem 3.5. *Let $\mathbb{T} = \mathbb{R}_+$ and let $(\Omega, \mathcal{F}) = (\mathbb{R}^{\mathbb{T}}, \mathcal{B}(\mathbb{R}^{\mathbb{T}}))$. Then there exists a probability measure $\mathbb{P}$ defined on (Ω, Σ) and a stochastic process W, such that W is the Brownian motion.*

[18] i.e for any $0 \leq t_1 < \dots t_n < \infty$ the random vector $(W_{t_1}, W_{t_2}, \dots, W_{t_n})$ is a multivariate gaussian vector.

[19] where for random variables X and Y we set $\mathrm{Cov}[X,Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

Proof. For brevity, we show only the outline of the proof. We use $I = \{t_1, t_2 \ldots, t_n\} \in \mathcal{T}$ to denote (any) finite set of time-points from $\mathbb{T}$, such that $0 \leq t_1 < t_2 < \ldots < t_n$, where $n \in \mathbb{N}$. Moreover, let μ_I denote the Gaussian measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with zero mean vector, and $n \times n$ covariance matrix given by

$$M_I := [t_i \wedge t_j]_{i,j \in I}.$$ [20]

For any $I = \{t_1, t_2 \ldots, t_n\}$ and $\Gamma \in \mathcal{B}(\mathbb{R}^n)$ we set

$$C(I, \Gamma) := \{\omega \in \Omega : (\omega(t_1), \ldots, \omega(t_n) \in \Gamma\}.$$

We use $\mathcal{F}_I := \{C(I, \Gamma) : \Gamma \in \mathcal{B}(\mathbb{R}^n)\}$ to denote the σ–algebra of all cylinder sets with predefined time indices $I \in \mathcal{T}$ and we use $\mathbb{P}_I$ to denote the corresponding probability measure on $(\Omega, \mathcal{F}_I)$ which is given by

$$\mathbb{P}_I[C(I, \Gamma)] = \mu_I(\Gamma), \quad \Gamma \in \mathcal{B}(\mathbb{R}^n).$$

Then, it can be shown that the family of measures $(\mathbb{P}_I)_{i \in \mathcal{T}}$ is consistent (the proof is left as an exercise).

Thus, Using Kolmogorov's extension Theorem 2.13 we know that there exists a probability measure on $(\Omega, \mathcal{F})$, say $\mathbb{P}$, such that for any $I = \{t_1, \ldots, t_n\} \in \mathcal{T}$ and $\Gamma \in \mathcal{B}(\mathbb{R}^n)$ we get

$$\mathbb{P}[C(I, \Gamma)] = \mathbb{P}_I[C(I, \Gamma)].$$

On $(\Omega, \mathcal{F}, \mathbb{P})$ we define the stochastic process $W = (W_t)_{t \in \mathbb{T}}$ setting

$$W_t(\omega) := \omega, \qquad \text{for } t \in \mathbb{T} \text{ and } \omega \in \Omega.$$

It can be shown that the process W is Gaussian, starts from zero, and it's auto-covariance function coincides with auto-covariance function of the Wiener process (given in Proposition 3.4).

Consequently, to conclude the proof we only need to show that there exists a continuous modification of the stochastic process W. Noting that for any $t, s \in \mathbb{T}$ and $p \in \mathbb{N}$ we get

$$\mathbb{E}\left[\frac{|W_t - W_s|}{\sqrt{|t - s|}}\right]^{2p} = \mathbb{E}|X|^{2p},$$

where X is a standard Gaussian random variable, and setting $K_p := \mathbb{E}|X|^{2p}$, we get the property

$$\mathbb{E}|W_t - W_s|^{2p} = K_p |t - s|^p, \quad t, s \in \mathbb{T}. \tag{3.1}$$ [21]

Consequently, setting $\epsilon = p - 1$ and using Kolmogorov's continuity Theorem 2.11 we conclude that W has a continuous modification. The continuous modification is the standard Brownian motion, which concludes the proof. □

One can also provide a more constructive proof of the existence, e.g. by taking Haar functions on $[0, 1]$ and showing convergence for properly defined series of independent Gaussian random variables. Exemplary construction might be found in the literature under the name *Lévy-Ciesielski contruction*. It should be noted that the construction only require probability space to be reach enough to contain a sequence of independent standard Gaussian random variables. We are now ready to show some interesting properties of the Brownian motion.

Next, we outline some basic results about the Brownian motion. Let us start with basic transformations for which the translated process is still the Brownian motion.

[20] The proof that M_I is indeed a proper covariance matrix for any $I \in \mathcal{T}$ is left as an exercise.

[21] Note that all moments of the standard gaussian random variable exist so K_p is finite for any $p \in \mathbb{N}$.

Proposition 3.6 (Transformations of Brownian motion). *Let W be a Brownian motion. Then, the following symmetric transformations preserve the Brownian motion:*

- **Time-homogeneity**, *i.e. for any $s > 0$ the process $\tilde{W}_t = W_{t+s} - W_s$ is a Brownian motion;*
- **Positive-scaling**, *i.e. for any $c > 0$ the process $\tilde{W}_t = cW_{t/c^2}$ is a Brownian motion;*
- **Time-inversion**, *i.e. the process $\tilde{W}_0 = 0$ and $\tilde{W}_t = tW_{1/t}$ (for $t > 0$) is a Brownian motion;*
- **Path-inversion**, *i.e. the process $\tilde{W}_t = -W_t$ is a Brownian motion.*

The proof of Proposition 3.6 is very simple and it is left as an excercise. Next, we show the properties of the Brownian motion related to continuity of it's sample paths.

Proposition 3.7 (Continuity properties). *Let W be a Brownian motion. Then*

1. *W is Hölder continuous for any exponent $\alpha < 1/2$;*[22]
2. *W is not Hölder continuous with exponent $\alpha = 1/2$ on any (time) subinterval;*
3. *Almost all paths of W are nowhere differentiable;*
4. *W has infinite total variation on any (time) sub-interval.*

Proof. Let W be a Brownian motion.

1) Using Equation (3.1) from the proof of Theorem 3.5 we know that we have

$$\mathbb{E}|W_t - W_s|^{2p} = K_p|t-s|^p, \quad t, s \in \mathbb{T},$$

for any $p \in \mathbb{N}$. Noting that

$$\frac{\epsilon}{p} = \frac{p-1}{2p} \to \frac{1}{2} \quad \text{for } p \to \infty,$$

and using Theorem 2.10 we know that we can find a continuous modification of W which is Hölder continuous for any exponent $\alpha < 1/2$.

2) Let $B^C_{[a,b]}$ denote the set off all paths which are Hölder continuous with $\alpha = 1/2$ on time interval $[a, b]$ with constant C, i.e.

$$B^C_{[a,b]} := \{\omega \in \Omega \,|\, \forall_{s,t\in[a,b]} : |W_t(\omega) - W_s(\omega)| \leq C|t-s|^{1/2}\}.$$

Let us show that for any $C \in \mathbb{N}$ and $a, b \in \mathbb{Q}$ $(a < b)$ we get $\mathbb{P}\left[B^C_{[a,b]}\right] = 0$. Indeed, noting that W has stationary and independent increments, and we can shift subset $[a, b]$ to $[0, b-a]$, for any

[22] i.e. there exists such modification.

$n \in \mathbb{N}$ we get

$$\begin{aligned}\mathbb{P}\left[B^C_{[a,b]}\right] &\le \mathbb{P}\left[\left\{\left|W_{\frac{i}{n}(b-a)} - W_{\frac{i-1}{n}(b-a)}\right| \le C\sqrt{\frac{b-a}{n}}\,,\ \text{for } i = 1,2,\ldots,n\right\}\right]\\ &= \prod_{i=1}^{n}\mathbb{P}\left[\left\{\left|W_{\frac{i}{n}(b-a)} - W_{\frac{i-1}{n}(b-a)}\right| \le C\sqrt{\frac{b-a}{n}}\right\}\right]\\ &= \mathbb{P}\left[\left\{\left|W_{\frac{b-a}{n}}\right| \le C\sqrt{\frac{b-a}{n}}\right\}\right]^n\\ &= \mathbb{P}\left[\{|W_1| \le C\}\right]^n\\ &= a_n.\end{aligned}$$

Since $a_n \searrow 0$ (as $n \to \infty$) we conclude that $\mathbb{P}\left[B^C_{[a,b]}\right] = 0$. Noting that to prove the initial statement we only need to consider time-intervals with rational endpoints and

$$\mathbb{P}\left[\bigcup_{C\in\mathbb{N}}\bigcup_{\substack{a<b\\ a,b\in\mathbb{Q}}} B^C_{[a,b]}\right] \le \sum_{C\in\mathbb{N}}\sum_{\substack{a<b\\ a,b\in\mathbb{Q}}}\mathbb{P}\left[B^C_{[a,b]}\right] = 0,$$

we conclude this part of the proof.

3) For brevity, we only outline the proof.[23] Suppose that W is differentiable at some point $s \in \mathbb{T}$. Then, it can be shown that there exists $\epsilon > 0$ and $K \in \mathbb{N}$ such that for any $t \in [s, s+\epsilon)$ we get

$$|W_t - W_s| \le K(t-s). \tag{3.2}$$

Next, it can be shown that inequality (3.2) implies that for sufficiently big $n \in \mathbb{N}$ there exists $i \in \mathbb{N}$, such that for $k = 0, 1, 2$ we get

$$|W_{(i+k+1)/n} - W_{(i+k)/n}| \le 7\frac{K}{n}. \tag{3.3}$$

Now, let A_M denote the set of all paths, such that there exists at least one time-point in $[0, M)$ at which W is differentiable. Using property (3.3) we know that

$$A_M \subset \bigcup_{K=1}^{\infty}\bigcup_{n_0=1}^{\infty}\bigcap_{n=n_0}^{\infty}\bigcup_{i=0}^{Mn-4}\bigcap_{k=0}^{2}\left\{|W_{(i+k+1)/n} - W_{(i+k)/n}| \le 7\frac{K}{n}\right\}.$$

Now, noting that W has stationary and independent increments, and exploiting the fact that

[23] see e.g. H.P. McKean, *Stochastic integrals* for the full proof.

$\sqrt{n}W_{1/n} \sim N(0,1)$ we get

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{n=n_0}^{\infty}\bigcup_{i=0}^{Mn-4}\bigcap_{k=0}^{2}\left\{|W_{(i+k+1)/n} - W_{(i+k)/n}| \leq 7\frac{K}{n}\right\}\right] &\leq \liminf_{n\to\infty} Mn\mathbb{P}\left[|W_{1/n}| < 7\frac{K}{n}\right]^3 \\
&\leq \liminf_{n\to\infty} Mn\mathbb{P}\left[|W_1| < 7\frac{K}{\sqrt{n}}\right]^3 \\
&\leq \liminf_{n\to\infty} Mn\left(\frac{2}{\sqrt{2\pi}}\int_0^{7K/\sqrt{n}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x\right)^3 \\
&\leq \liminf_{n\to\infty} Mn\left(\frac{2}{\sqrt{2\pi}}\int_0^{7K/\sqrt{n}} 1\,\mathrm{d}x\right)^3 \\
&\leq \liminf_{n\to\infty} Mn\left(\frac{2}{\sqrt{2\pi}}\frac{7K}{\sqrt{n}}\right)^3 \\
&\leq \liminf_{n\to\infty} \frac{C}{\sqrt{n}}.
\end{aligned}$$

where the constant $C := M\left(\frac{14K}{\sqrt{2\pi}}\right)^3$ is independent of i and n. Noting that

$$\liminf_{n\to\infty} \frac{C}{\sqrt{n}} = 0,$$

and using the (countable) subadditivity property of the probability measure we conclude that $\mathbb{P}[A_M] = 0$.

4) The function $f\colon \mathbb{R} \to \mathbb{R}$ has a finite total-variation on the interval $[a,b]$ if

$$\sup_{\mathbf{t}\in\mathcal{P}[a,b]} \sum_{i=1}^{n} |f(t_{i+1}) - f(t_i)| < \infty,$$

where $\mathcal{P}[a,b]$ is the set of all finite partitions of $[a,b]$ and $\mathbf{t} = (t_1,\dots,t_n)$. Using similar arguments as before it is enough to show that for $a,b \in \mathbb{Q}_+$ $(a<b)$ and $C \in \mathbb{N}$ we get

$$\lim_{n\to\infty} \mathbb{P}\left[\left\{\sum_{i=1}^{n}\left|W_{\frac{i}{n}(b-a)} - W_{\frac{i-1}{n}(b-a)}\right| \leq C\right\}\right] = 0.$$

For any fixed $n \in \mathbb{N}$ we set $X_n := \sum_{i=1}^{n}\left|W_{\frac{i}{n}(b-a)} - W_{\frac{i-1}{n}(b-a)}\right|$. Recalling that for a standard normal random variable Z we get $\mathbb{E}[|Z|] = \sqrt{\frac{2}{\pi}}$ we get

$$\begin{aligned}
\mathbb{E}[X_n] &= \sum_{i=1}^{n}\sqrt{\frac{b-a}{n}}\mathbb{E}[|W_1|] = \sqrt{\frac{2(b-a)}{\pi}n}\,, \\
\mathrm{Var}\,[X_n] &= \sum_{i=1}^{n}\mathrm{Var}\left[\left|W_{\frac{i}{n}(b-a)} - W_{\frac{i-1}{n}(b-a)}\right|\right] = n\,\mathrm{Var}\left[\left|W_{\frac{b-a}{n}}\right|\right] = (b-a)\,\mathrm{Var}\,|W_1|\;.
\end{aligned}$$

The Chebyshev's inequality states that for any non-constant $Y \in L^2$ and any $r > 0$ we get

$$\mathbb{P}[|Y - \mathbb{E}[Y]| \geq r] \leq \frac{\mathrm{Var}\,[Y]}{r^2}.$$

Thus, setting $Y = -X_n$ and considering only positive part of $Y - \mathbb{E}[Y]$ we get

$$\mathbb{P}[X_n \leq \mathbb{E}[X_n] - r] \leq \frac{\mathrm{Var}\,[X_n]}{r^2}.$$

Next, noting that $\mathbb{E}[X_n] > 0$ and setting $r = \frac{\mathbb{E}[X_n]}{2}$ we get

$$\mathbb{P}\left[X_n \leq \frac{\mathbb{E}[X_n]}{2}\right] \leq \frac{4(b-a)\,\mathrm{Var}\,|W_1|}{E[X_n]^2} \leq \frac{2\pi\,\mathrm{Var}\,|W_1|}{n}.$$

Now, noting that $E[X_n] \to \infty$ (as $n \to \infty$) for any $C \in \mathbb{R}_+$ there exists $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ we get

$$\mathbb{P}[X_n \leq C] \leq \mathbb{P}\left[X_n \leq \frac{\mathbb{E}[X_n]}{2}\right] \leq \frac{2\pi\,\mathrm{Var}\,|W_1|}{n}.$$

Since $\frac{\pi\,\mathrm{Var}\,|W_1|}{n} \to 0$ (as $n \to \infty$) the proof is complete. □

W conclude this Section with basic characteristic of the Brownian motion related to it's martingale representations.

Proposition 3.8. *Let W be a Brownian motion. Then,*

1. *The process (W_t) is a martingale (wrt. filtration generated by W).*
2. *The process $(W_t^2 - t)$ is a martingale (wrt. filtration generated by W).*
3. *Let $\mathbb{F}$ be a filtration and let Z be a process such that*
   - *$Z_0 = 0$;*
   - *Z is square-integrable and continuous;*
   - *(Z_t) is $\mathbb{F}$–martingale;*
   - *$(Z_t^2 - t)$ is $\mathbb{F}$–martingale;*

   Then, Z is a Brownian motion.

3.2 Poisson Process

Now, we introduce the second important stochastic process – the Poisson process. It is another example of a Lévy process which could be constructed given it's finite-dimensional distribution. Before we state the definition, let us recall what is a Poisson random variable. We say that random variable X has a Poisson distribution with intensity parameter $\lambda > 0$ if for $k = 0, 1, 2, \ldots$ we get

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}.$$

For brevity, we use the notation $X \sim \mathcal{P}(\lambda)$ to denote the Poisson random variable with intensity $\lambda > 0$. We are now ready to define the Poisson process.

Definition 3.9. Let $\mathbb{T} = \mathbb{R}_+$. We call $N = (N_t)_{t \in \mathbb{T}}$ a **Poisson process** with intensity $\lambda > 0$ if

1) $N_0 = 0$;

2) N is stochastically continuous;

3) N has independent and stationary increments;

4) $N_t - N_s \sim \mathcal{P}(\lambda(t-s))$, for any $0 \geq s \geq t$.

Directly from the definition we get that every Poisson process is a Lévy process. Using Kolmogorov's theorems one could show that such process exists. Nevertheless, we provide an alternative renewal theory based representation of the Poisson processes which is more constructive. Also, for simplicity we we work with the cádlág modicivation of the process. Let us start with the definition of the renewal process.

Definition 3.10. Let $\mathbb{T} = \mathbb{R}_+$. We call $N = (N_t)_{t\in\mathbb{T}}$ a **renewal process** if

$$N_t = \max\{n \in \mathbb{N} : S_n \leq t\},$$

where $S_0 = 0$, $S_n = X_1 + \ldots + X_n$ for $n \geq 1$, and $(X_n)_{n\in\mathbb{N}}$ is a sequence of non-negative independent and identically distributed random variables.

We usually associate S_n with the time of the n-th arrival of some predefined event (e.g. arrival of particle, malfunction of a machine, etc.), while X_n is associated with in-between (inter)arrival time. With that interpratation, for any $t \in \mathbb{T}$, the random variable N_t informs us how many times the even occurred in the time interval $[0, t]$. Note that we assume that we are dealing with a recurrent-event process where the inter-event times are independent and identically distributed.

The key property of the renewal process is the distribution of the inter-arrival times. To ensure something called Markov property, we would like our process to forget about the past - this is the property satisfied by exponential random variables. Those class of reneval processes coincide with the class of Poisson random processes, as explained in the next Theorem.

Theorem 3.11. *Let $\mathbb{T} = \mathbb{R}_+$. The stochastic process $N = (N_t)_{t\in\mathbb{T}}$ is a Poisson process with intensity $\lambda > 0$ if and only if it is a renewal process with inter-arrival times $(X_n)_{n\in\mathbb{N}}$ having exponential distribution with parameter $\lambda > 0$.*

Proof. First, assume that N is a Poisson process with intensity $\lambda > 0$. We define a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ be setting

$$\begin{aligned} X_1 &:= \inf\{t \in \mathbb{T} : N_t = 1\} \\ X_n &:= \inf\{t \in \mathbb{T} : N_t = n\} - \inf\{t \in \mathbb{T} : N_t = n-1\}. \end{aligned}$$

Setting $S_n := X_1 + \ldots + X_n$, noting that $S_n = \inf\{t \in \mathbb{T} : N_t = n\}$, and using the fact that N is almost-surely non-negative, for any $k \in \mathbb{N}$ we get

$$\{N_t = k\} = \{S_1 \leq t\} \cap \ldots \cap \{S_k \leq t\} \cap \{S_{k+1} > t\} = \left\{\max_{n\in\mathbb{N}}\{S_n \leq t\} = k\right\}.$$

Consequently, we get

$$N_t = \max\{n \in \mathbb{N} : S_n \leq t\}.$$

Now, we want to show that the sequence (X_n) is independent and exponentially distributed with parameter $\lambda > 0$. For any $t_1 \geq 0$ we get

$$\mathbb{P}[X_1 > t_1] = \mathbb{P}[N_{t_1} = 0] = e^{-\lambda t_1},$$

so X_1 is exponentially distributed with parameter $\lambda > 0$. Now, for any $t_1, t_2 \geq 0$ we get

$$\mathbb{P}[X_2 > t_2 \mid X_1 = t_1] = \mathbb{P}[S_2 > t_2+t_1 \mid X_1 = t_1] = \mathbb{P}[N_{t_1+t_2} = 1 \mid X_1 = t_1] = \mathbb{P}[N_{t_2+t_1}-N_{t_1} = 0 \mid X_1 = t_1]$$

Noting that $N_{t_2+t_1} - N_{t_1}$ is independent of $\sigma(N_s, s \leq t_1)$ and $\{X_1 = t_1\}$ is $\sigma(N_s, s \leq t_1)$–measurable we get

$$\mathbb{P}[X_2 > t_2 | X_1 = t_1] = \mathbb{P}[N_{t_2+t_1} - N_{t_1} = 0] = e^{-\lambda t_2}.$$

This shows that X_2 is in fact independent of X_1 and has exponential distribution with parameter λ. Same reasoning could be applied recursively to show that X_n is independent of $X_{n-1}, X_{n-2}, \ldots$, and X_1, which concludes this part of the proof.[24]

Now let us assume that N is a renewal process with exponential increments $(X_k)_{k\in\mathbb{N}}$, i.e.

$$N_t = \max\{n \in \mathbb{N} : S_n \leq t\},$$

where $S_n := X_1 + \ldots + X_n$. We use $(\mathcal{F}_t)_{t\in\mathbb{T}}$ to denote natural filtration generated by the process $(N_t)_{t\in\mathbb{T}}$. For brevity, we only outline the key steps of the proof:

1) For any $s \in \mathbb{T}$ we set

$$\tau_s := \inf\{t \geq 0 : N_{s+t} - N_s > 0\}.$$

Noting that for any $n \in \mathbb{N}$ we get $\{N_s = n\} = \{S_n \leq s\} \cap \{S_n + X_{n+1} > s\} \in \mathcal{F}_s$, for any $t \in \mathbb{T}$ we get

$$\begin{aligned}
\mathbb{P}[\tau_s \geq t \mid \mathcal{F}_s] &= \mathbb{P}\left[\bigcup_{n\in\mathbb{N}} \{\tau_s > t\} \cap \{N_s = n\} \mid \mathcal{F}_s\right] \\
&= \mathbb{P}\left[\bigcup_{n\in\mathbb{N}} \{N_{s+t} = n\} \cap \{N_s = n\} \mid \mathcal{F}_s\right] \\
&= \sum_{n\in\mathbb{N}} \mathbb{1}_{\{N_s=n\}} \mathbb{P}\left[S_n + X_{n+1} > t + s \mid \mathcal{F}_s\right] \\
&= \sum_{n\in\mathbb{N}} \mathbb{1}_{\{N_s=n\}} \mathbb{P}\left[X_{n+1} > (s - S_n) + t \mid \mathcal{F}_s\right] \\
&= \sum_{n\in\mathbb{N}} \mathbb{1}_{\{N_s=n\}} \mathbb{P}\left[X_{n+1} > t\right] \\
&= \sum_{n\in\mathbb{N}} \mathbb{1}_{\{N_s=n\}} e^{-\lambda t} \\
&= e^{-\lambda t},
\end{aligned}$$

where in the last equality we have used the memoryless property and the fact that the set $\{X_{n+1} > (s - S_n)\}$ conditioned on the event $\{N_s = n\}$ is $\mathcal{F}_s$-measurable. Consequently, we get that for any $t, s \in \mathbb{T}$ the random variable given by

$$\min\{N_t - N_s, 1\}$$

is independent of $\mathcal{F}_s$, and is the standard Bernoulli random variable with probability of success equal to $1 - e^{-\lambda(t-s)}$.

[24] See Billingsley, *Probability and Measure* for the full formal proof.

2) Next, for any $t, s \in \mathbb{T}$, and $n \in \mathbb{N}$, we consider equidistant partition of $[t, s]$ into $n+1$ separate time-points, say $t_1^n, \ldots, t_{n+1}^n$, and for $i = 1, 2, \ldots, n$ we set
$$\xi_i^n := \min\{N_{t_{i+1}^n} - N_{t_i^n}, 1\}.$$

3) We define the associated random variable Z_n given by
$$Z_n := \sum_{i=1}^n \xi_i^n,$$
and note that $Z_n \sim B(n, 1 - e^{-\lambda(t-s)/n})$, i.e. Z_n has Binomial distribution with n trials and probability of success equal to $1 - e^{-\lambda(t-s)/n}$.

4) We have
$$\lim_{n\to\infty} Z_n = N_t - N_s,$$
almost surely, as there cannot be infinitely many jump on a finite unit interval, and the size of jumps is equal to 1 almost surely. Consequently, as each Z_n is independent of $\mathcal{F}_s$, we get that $N_t - N_s$ in independent of $\mathcal{F}_s$. In particular, this implies that the increments of N are independent.

5) For $n \to \infty$ we get that $n(1 - e^{-\lambda(t-s)/n}) \to \lambda(t-s)$, i.e. the number of trials multiplied by intensity for the Bernoulli random variables Z_n goes to a constant. Thus, using the Poisson limit theorem we know that the sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ weakly converges to the Poisson random variable with intensity parameter $\lambda(t-s)$. This shows that increments are stationary, and have the Poisson distribution, which concludes the proof.

To calculate the distribution of N_t more directly one could note that for any $k \in \mathbb{N}$, the random variable S_k has the Gamma density $g_k \colon \mathbb{R} \to \mathbb{R}_+$ given by[25]
$$g_k(r) = \begin{cases} \lambda \frac{(\lambda r)^{k-1}}{(k-1)!} e^{-\lambda r}, & \text{if } r > 0. \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for $k \in \mathbb{N}$ we get
$$\begin{aligned}
\mathbb{P}[N_t = k] &= \mathbb{P}[\max\{n \in \mathbb{N} : S_n \leq t\} = k] \\
&= \mathbb{P}[S_k \leq t < S_k + X_{k+1}] \\
&= \int \int_{\{r \leq t \leq r+z\}} g_k(r) g_1(z) \, \mathrm{d}r \, \mathrm{d}z \\
&= \int_0^t \left[\int_{t-r}^\infty g_1(z) \, \mathrm{d}z \right] g_k(r) \, \mathrm{d}r \\
&= \int_0^t e^{-\lambda(t-r)} g_k(r) \, \mathrm{d}r \\
&= \frac{\lambda^n}{(k-1)!} e^{-\lambda t} \int_0^t r^{n-1} \, \mathrm{d}r \\
&= \frac{(\lambda t)^n}{k!} e^{-\lambda t},
\end{aligned}$$
which shows that $N_t \sim \mathcal{P}(\lambda t)$. □

[25]The proof of this fact is left as an excercise. It can be proved by induction. One can directly calculate the convolution of the increments.

As we have mentioned in the proof, the *jumps* for the (cádlág) Poisson process are almost surely finite and equal to one.

Definition 3.12. Let $X = (X_t)_{t\in\mathbb{T}}$ be a cadlag stochastic process. We use

$$\Delta X_t := X_t - X_{t_-},$$

to denote the *jump* of X at $t \in \mathbb{T}$.

Also, the intensity parameter $\lambda > 0$ is responsible for the mean number of signals. Those two facts are presented in the next proposition, the proof is left as an exercise.

Proposition 3.13. *Let $N = (N_t)_{t\in\mathbb{T}}$ be a Poisson process. Then*

1. $\mathbb{P}[\Delta N_t \in \{0, 1\}] = 1$*, for any* $t \in \mathbb{T}$*;*
2. $\lim_{t\to+\infty} \frac{N_t}{t} = \lambda$.

In fact, for any (cádlág) Levy process it is interesting to measure the size of jumps and their intensity. Let B_0 denote the family of Borel sets of $\mathbb{R}$ whose closures do not contain 0.

Definition 3.14. Let X be a (cádlág) Levy process. Then,

1) the **Poisson random measure** (or *jump measure*) of X is the map $\Pi_X : \mathbb{T} \times B_0 \times \Omega \to \mathbb{N}$ given by
$$\Pi_X(t, U)(\omega) = \sum_{s:0<s\le t} \mathbb{1}_{\{\Delta X_s \in U\}}(\omega);$$

2) the **Levy measure** of X is the map $\nu_X : B_0 \to \mathbb{R}$ given by
$$\nu_X(U) = \mathbb{E}[\Pi_X(1, U)].$$

The map Π_X measures how many times the jump of size from U occurred on time interval $[0, t]$ for a given path, while ν_X measures the average frequency of U jumps in the unit interval $[0, 1]$. It should be noted that Π_X is finite for all $U \in B_0$ and $t \in \mathbb{T}$ which follows from the cádlág property.

Proposition 3.15. *Let X be a (cádlág) Levy process. Then, Π_X is finite.*

Proof. Set $T_1 := \inf\{t > 0 : \Delta X_t \in U\}$. By the right-continuity of X we know that

$$\lim_{t\to 0^+} X_t = X_0 = 0.$$

Noting that we can find $\epsilon > 0$ such that for any $u \in U$ we get $u > \epsilon$, and exploiting right-continuity of X we get that $T_1 > 0$ (a.s.). Next, we inductively define

$$T_{n+1} = \inf\{t > 0 : \Delta X_t \in U \text{ and } t > T_n\}.$$

Using the above argument we know that $T_{n+1} > T_n$. To conclude the proof, it is enough to show that for $n \to \infty$ we get $T_n \to \infty$ (a.s.). On the contrary, let us assume that there exists a subset of positive measure, say D, such that for any $\omega \in D$ we get $T_n(\omega) \to T_\omega$, where $T_\omega < \infty$. Consequently, for all $\omega \in D$ we know that
$$\lim_{t\to T_\omega^-} X_t(\omega)$$
does not exist, which contradicts the cádlág property. ∎

Finally, we state the proposition which shows the connection between Poisson process, Poisson random measure, and the Levy measure.

Proposition 3.16. *Let X be a (cádlág) Levy process. Then, for any fixed $U \in B_0$ the process $N^{X,U} = (N_t^{X,U})_{t\in\mathbb{T}}$ given by*

$$N_t^{X,U}(\omega) := \Pi_X(t,U)(\omega),$$

is a Poisson process with intensity parameter $\lambda = \nu_X(U)$.

It should be noted that if X is itself a Poisson process with intensity $\lambda > 0$, and $1 \in U$, then we get $\Pi_X(t,U) = X$, and $\nu_X(U) = \lambda$.

3.3 Markov processes

In the previous Section we considered two examples of processes: Brownian motion and Poisson processes. Those processes were constructed in a way such that the most up-to-date value of the process was the only important information, when we wanted to say something about the future dynamics. This property is known in the literature as the *Markov property*. For brevity, we only present the results in continuous time (the discrete-time case should have been discussed in another course).

Definition 3.17. Let $\mathbb{T} = \mathbb{R}_+$ and let $X = (X_t)_{t\in\mathbb{T}}$ be an $\mathbb{F}$-adapted process. We say that X has the **Markov property** if for any $A \in \mathcal{B}(\mathbb{R})$, and $s,t \in \mathbb{T}$, such that $t > s$, we get

$$\mathbb{P}[X_t \in A \mid \mathcal{F}_s] = \mathbb{P}[X_t \in A \mid X_s].$$

We say that X is a Markov process with respect to filtration $\mathbb{F}$ if it satisfies the Markov property with respect to this filtration. We are now ready to formally define the Markov process.

Definition 3.18. Let $\mathbb{T} = \mathbb{R}_+$ and let $X = (X_t)_{t\in\mathbb{T}}$ be a stochastic process. We say that X is a **Markov process** if it satisfies the Markov property with respect to it's natural filtration.

Typically, Markov processes are defined in terms of their's transition functions. Before introducing the formal definition let us define and discuss some underlying concepts. We start with the definition of a transition kernel. While the definition could be easily extended to stochastic processes taking values in any measurable space, for brevity we only consider the univariate real case $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 3.19. Let $t,s \in \mathbb{T}$ be such that $t > s$. A **probability kernel** (from s to t) is a map

$$P_{s,t}\colon \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to \mathbb{R}_+ \cup \{+\infty\}$$

such that

- for each $x \in \mathbb{R}$ the map $A \mapsto P_{s,t}(x,A)$ is a measure;
- for each $A \in \mathcal{B}(\mathbb{R})$ the map $x \mapsto P_{s,t}(x,A)$ is measurable.
- for all $x \in E$ we get $P_{s,t}(x,\mathbb{R}) = 1$.

The idea of a probability kernel is pretty straighforward. For any two time-points s and t we want to asses the probability of being at time t in the set A of the process which at time s was at state x. In other words, we want to check how the process which started at point x (at time s) behaves at time t. For completeness, for any $t \in \mathbb{T}$ we also define the probability kernel $P_{t,t}$ which is simply given by

$$P_{t,t}(x, A) := \delta_x(A).$$

In order to recover information about the evolution of the Markov process we need a whole collection of probability kernels for all possible time-points. For brevity we only consider the homogeneous case, i.e. any kernel $P_{t,s}$ (from a collection of probability kernels $(P_{t,s})_{t,s\in\mathbb{T}:t<s}$) should depend on time only though the increment $t - s$. In other words, for any $t > 0$ and $h \geq 0$ we want to have $P_{0,t} \equiv P_{h,t+h}$. In that case it is enough to consider the incremental family of kernels $(P_t)_{t\in\mathbb{T}}$ given e.g. by $P_t := P_{0,t}$. From now we restrict ourselves to the homogenous case and use the simplified notation.[26]

Before we define the transition function we need to introduce basic notation related to probability kernels. Let $\mathcal{C}(\mathbb{R})$ denote the space of all bounded and measurable functions $f : \mathbb{R} \to \mathbb{R}$. Given transition kernel P_t and $f \in \mathcal{C}(\mathbb{R})$ we define the associated integral for any $x \in \mathbb{R}$ by setting

$$P_t f(x) := \int_{\mathbb{R}} f(y) P_t(x, \mathrm{d}y).$$ [27]

Also, any two kernels, say P_t and P_s, could be combined. For any $f \in \mathcal{C}(\mathbb{R})$ and $x \in \mathbb{R}$ we write

$$P_t P_s f(x) := \int_{\mathbb{R}} \int_{\mathbb{R}} f(z) P_s(y, \mathrm{d}z) P_t(x, \mathrm{d}y),$$

and use $P_t P_s$ to denote the associated probability kernel. Also, note that $P_t f \in \mathcal{C}(\mathbb{R})$. We are now finally ready to define the transition function.

Definition 3.20. A homogeneous **transition function** P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a collection of probability kernels $(P_t)_{t\in\mathbb{T}}$ such that for any $t, s \in \mathbb{T}$ we get

$$P_t P_s = P_{t+s}. \tag{3.4}$$

The identity (3.4) is sometimes referred to as the *Chapman-Kolmorogov equation*. It is required in order to guarantee the time-consistency of process evolution, which could be expressed in terms of the conditional expectation via the tower rule property. Before we explain this in details let us show how the transition function P could be linked to a stochastic process.

Definition 3.21. Let X be an $\mathbb{F}$-adapted stochastic process. We say that X **is Markov with transition function** P if for any $f \in \mathcal{C}(\mathbb{R})$ and $t, s \in \mathbb{T}$, such that $t > s$, we get

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = P_{t-s} f(X_s). \tag{3.5}$$

[26] In fact, this assumption is not very restrictive - given time inhomogeneous Markov process we can simply look at the space-time process $((t, X_t))_{t\in\mathbb{T}}$ that will be homogeneous.

[27] Note that $P_t(x, \cdot)$ is a measure so this is well defined

Note that the Chapman-Kolmogorov property is indeed required in Definition 3.21. Using the tower property, for time-points $t, s, u \in \mathbb{T}$, we immediately get

$$\begin{aligned} P_{t+s}f(X_u) &= \mathbb{E}[f(X_{t+s+u})|\mathcal{F}_u] \\ &= \mathbb{E}[\mathbb{E}[f(X_{t+s+u})|\mathcal{F}_{t+u}]|\mathcal{F}_u] \\ &= \mathbb{E}[P_s f(X_{t+u})|\mathcal{F}_u] \\ &= P_t P_s f(X_u). \end{aligned}$$

It can be shown that both Brownian motion and Poisson process are Markov with appropriate transition functions. In fact, given any Levy process X we can define it's transition function P simply by setting

$$P_t(x, A) := \mathbb{P}[X_{t+s} \in A | X_s = x],$$

for all $t \in \mathbb{T}$, $x \in \mathbb{R}$, and $A \in \mathcal{B}(\mathbb{R})$ (and where $s > 0$). Note that RHS does not depend on s because X has stationary increments. Nevertheless, in practice it is usually not possible to write down transition functions explicitly. Instead, we define the processes in terms of solutions to differential equations, where the *infinitesimal generator* is provided instead of the transition function.

Next, it is important to know If we can construct the process with any pre-defined transition function. Such construction is indeed possible and the proof follows from the Kolmogorov's extension theorem. For simplicity, we assume that we know the initial state of the process (at time 0) and it is determined. It could be shown that similar results are true if we assume that the *initial distribution* of the process is known.

Theorem 3.22. *Let $\mathbb{T} = \mathbb{R}_+$ and let $\Omega = \mathbb{R}^{\mathbb{T}}$ be a space of functions $\omega : T \to \mathbb{R}$. Let $X = (X_t)_{t\in\mathbb{T}}$ denote the coordinate stochastic process, where $X_t : \Omega \to \mathbb{R}$ is given by*

$$X_t(\omega) = \omega(t),$$

for any function $\omega \in \Omega$. Let $\mathbb{F}$ denote a natural filtration of X, and let $\mathcal{F} := \sigma(X_t, t \in \mathbb{T})$. Then, for any transition function P and starting point $x \in \mathbb{R}$ there exists there is (a unique) probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ under which X is Markov with transition function P and initial state x.

Also, it should be noted that if we have two processes with the same initial state and the same transition functions, then they have same finite distributions. The proof of this fact is left as an exercise. Finally, we need to show that if a process is Markov with transition function, then it is indeed a Markov process.

Proposition 3.23. *Let X be Markov with transition function P. Then X is a Markov process.*

Proof. The proof of this fact is in fact straightforward. Let us fix $t, s \in \mathbb{T}$ such that $t > s$ and $A \in \mathcal{B}(\mathbb{R})$. Let $f : \mathbb{R} \to \mathbb{R}$ be the indicator function of set A, i.e. $f(x) = \mathbb{1}_{\{x\in A\}}$. We know that $f \in \mathcal{C}(\mathbb{R})$. Moreover, using property (3.5) we get

$$\mathbb{P}[X_t \in A \mid \mathcal{F}_s] = \mathbb{E}[1_{\{X_t\in A\}}|\mathcal{F}_s] = \mathbb{E}[f(X_t)|\mathcal{F}_s] = P_{t-s}f(X_s) = P_{t-s}(X_s, A).$$

Similarly,

$$\mathbb{P}[X_t \in A \mid X_s] = \mathbb{E}[\mathbb{E}[1_{\{X_t\in A\}}|\mathcal{F}_s]|X_s] = \mathbb{E}[P_{t-s}(X_s, A)|X_s] = P_{t-s}(X_s, A),$$

which conclude the proof. □

In fact, for Borel spaces, the reverse implication is also true, i.e. for any Markov process, there exists the associated transition probability kernel. However, the detailed explanation of this fact is out of scope of this lecture.

4 Introduction to stochastic Ito calculus

The main goal of this section is to define the stochastic integral from the Wiener process W, i.e. give a meaning to quantity

$$\int_0^t X_s \, \mathrm{d}W_s, \quad t \geq 0. \tag{4.1}$$

Initial thought might be to use the standard (non-stochastic) framework and define (4.1) for any ω using the Lebesgue-Stieltjes integral. Unfortunately, we have shown in Proposition 3.7 that Wiener process has infinite total variation on any time-interval, so that this cannot be done and special construction must be used. While the construction details in many places resemble the construction of the classical Lebesgue integral, there are a few key differences. The main one is connected to the fact that we cannot take any point from a predefined partition, and then take the limit (shrinking the size of all partitions). In fact, we will simply take the left end-point which will lead to good definition.[28] Let us start with the definition of simple process which might be seen as a main building block of the Ito integral.

In this section we assume that time is continuous, i.e. $\mathbb{T} = \mathbb{R}_+$, and $\mathbb{F}$ corresponds to filtration generated by a fixed Wiener process $W = (W_t)_{t\in\mathbb{T}}$. Also, for simplicity we focus on defining the integrals on time-intervals $[0, t)$. This could be easily extended to any sub-intervals.

4.1 Ito integral of an elementary process

The elementary processes are a simple extension of the elementary functions.

Definition 4.1. We call a stochastic process $\xi = (\xi_t)_{t\in\mathbb{T}}$ an **elementary (predictible) process** if ξ could be represented as

$$\xi_t = Z\mathbb{1}_{\{0\}}(t) + \sum_{k=0}^{n-1} Z_k \mathbb{1}_{(t_k, t_{k+1}]}(t), \quad t \geq 0, \tag{4.2}$$

where Z is $\mathcal{F}_0$–measurable, $n \in \mathbb{N}$, the sequence of finite time points $t_0, t_1, \ldots, t_n$ is strictly increasing with $t_0 = 0$, and Z_k is a square-integrable and $\mathcal{F}_{t_k}$–measurable random variable for any $k = 0, 1, \ldots, n-1$. We denote by $\mathcal{P}$ the class of all elementary processes.

Now, we define Ito integral of elementary process. Given $\xi \in \mathcal{P}$ we define first the infinite-time integral $I(\xi)$ by setting

$$I(\xi) := \int_0^\infty \xi_s \, \mathrm{d}W_s := \sum_{k=0}^{n-1} Z_k (W_{t_{k+1}} - W_{t_k}).$$

Next, we define the integral of ξ at time t by simply setting

$$I_t(\xi) := \int_0^\infty \xi_s \, \mathrm{d}W_s := I(\mathbb{1}_{[0,t]}\xi).$$ [29]

Note that the integral $(I_t)_{t\in\mathbb{T}}$ is a stochastic process itself. One could easily check that those definitions are well stated in the sense that they do not depend on representation 4.2 and the

[28] One could take e.g. middlepoint as well. Then, we will obtain different integral referred to as Stratonovich integral, the most common alternative to the Ito integral that is used frequently in physics.

[29] where $(\mathbb{1}_{[0,t]}\xi)_s = \mathbb{1}_{[0,t]}\xi_s$ for any $s \in \mathbb{T}$

linear combinations of linear processes is a linear process.[30] Let us now combine together the basic properties of the integral that we have just defined. They are outlined in Theorem 4.2.

Theorem 4.2. *Let $\xi, \nu \in \mathcal{P}$. Then*

1. *For any $\alpha, \beta \in \mathbb{R}$ and $t \geq 0$ we get $I_t(\alpha\xi + \beta\nu) = \alpha I_t(\xi) + \beta I_t(\nu)$.*
2. *For any $t > s \geq 0$ we get $I_t(\xi) = I_s(\xi) + I_{s,t}(\xi)$, where $I_{s,t}(\xi) = I(\mathbb{1}_{[s,t]}\xi)$.*
3. *The process $(I_t(\xi))_{t\in\mathbb{T}}$ has continuous paths.*
4. *The process $(I_t(\xi))_{t\in\mathbb{T}}$ is a square-integrable martingale (wrt. $\mathbb{F}$). Moreover, for any $t \geq 0$ we get*

$$\begin{aligned} \mathbb{E}[I_t(\xi)] &= 0, \\ \mathbb{E}[I_t(\xi)^2] &= \mathbb{E}\left[\int_0^t \xi_s^2 \, \mathrm{d}s\right]. \end{aligned} \tag{4.3}$$

Equality (4.3) is often referred to as Ito isometry.

Proof. The proof of 1) and 2) is left as an exercise. Let us now show 3). Noting that a (finite) sum of continuous stochastic processes is a continuous stochastic process it is enough to note that

$$Z_1(W_{t_2} - W_{t_1}) \tag{4.4}$$

is continuous, where Z_k is $\mathcal{F}_{t_1}$-measurable random variable and $t_2 > t_1 \geq 0$. This follow directly from the Wiener process path continuity property. Indeed, setting

$$\xi_t = Z_1 \mathbb{1}_{(t_1,t_2]}(t)$$

we simply get

$$I_t(\xi) = \begin{cases} 0 & t \in [0, t_1], \\ Z_1(W_t - W_{t_1}) & t \in (t_1, t_2), \\ Z_1(W_{t_2} - W_{t_1}) & t \in [t_2, \infty). \end{cases} \tag{4.5}$$

Now, we prove 4). Again, because the sum of square-integrable martingales is a square-integrable martingale we can consider the simplified process as in (4.4) and consider integral process $(I_t(\xi))_{t\in\mathbb{T}}$ given in (4.5). First, we note that the stochastic process $(I_t(\xi))_{t\in\mathbb{T}}$ is adapted since W is adapted and Z_1 is $\mathcal{F}_{t_1}$–measurable. Second, it is integrable since for any $t \in \mathbb{T}$ the process could be seen as a multiplication of two square-integrable random variables. Thus, we need to show the martingale property

$$\mathbb{E}[I_t(\xi) \mid \mathcal{F}_s] = I_s(\xi),$$

for all $t > s \geq 0$. Let us consider three cases: $0 \leq s \leq t_1$, $t_1 < s \leq t_2$, and $s > t_2$.

[30] the proof is left as an exercise. Simply note that for any two partitions there exists a sub-partition that is more granular compared to both of them.

1) Case 1: $0 \le s \le t_1$. In this case we know that $I_s(\xi) = 0$. If $t \le t_1$ the proof is straight forward as both numbers are equal to 0. For $t \in (t_1, t_2)$ using the fact that $\mathcal{F}_s \subseteq \mathcal{F}_{t_1}$ we get

$$\begin{aligned}\mathbb{E}[I_t(\xi) \mid \mathcal{F}_s] &= \mathbb{E}[Z_1(W_t - W_{t_1}) \mid \mathcal{F}_s]\\ &= \mathbb{E}[\mathbb{E}[Z_1(W_t - W_{t_1}) \mid \mathcal{F}_{t_1}] \mid \mathcal{F}_s]\\ &= \mathbb{E}[Z_1\mathbb{E}[W_t - W_{t_1} \mid \mathcal{F}_{t_1}] \mid \mathcal{F}_s]\\ &= \mathbb{E}[Z_1\, 0 \mid \mathcal{F}_s]\\ &= 0.\end{aligned}$$

The proof for $t \in [t_2, \infty)$ is similar.

2) Case 2: $t_1 < s \le t_2$. Let $t \in (s, t_2]$. Because Z_1 is $\mathcal{F}_s$–measurable we get

$$\begin{aligned}\mathbb{E}[I_t(\xi) \mid \mathcal{F}_s] &= \mathbb{E}[Z_1(W_t - W_{t_1}) \mid \mathcal{F}_s]\\ &= Z_1\mathbb{E}[W_t - W_{t_1} \mid \mathcal{F}_s]\\ &= Z_1\mathbb{E}[(W_t - W_s) + (W_s - W_{t_1}) \mid \mathcal{F}_s]\\ &= Z_1(W_s - W_{t_1})\\ &= I_s(\xi)\end{aligned}$$

The proof for $t \in [t_2, \infty)$ is similar.

3) Case 3: $s > t_2$. We get $I_t(\xi) = I_s(\xi)$.

This concludes the proof of the martingale property. Let us now show that for any $t \ge 0$ we get

$$\mathbb{E}[I_t(\xi)] = 0 \tag{4.6}$$

Again, we can utilise representation (4.4). For $t \le t_1$ we get $I_t(\xi) =$ which implies (4.6). On the other hand, if $t > t_1$ then we get

$$\mathbb{E}[I_t(\xi)] = \mathbb{E}[\mathbb{E}[I_t(\xi) \mid \mathcal{F}_{t_1}]] = \mathbb{E}[Z_1\mathbb{E}[W_{t\wedge t_2} - W_{t_1} \mid \mathcal{F}_{t_1}]] = 0.$$

Thus, to conclude the proof we need to show

$$\mathbb{E}[I_t(\xi)^2] = \mathbb{E}\left[\int_0^t \xi_s^2 \,\mathrm{d}s\right]. \tag{4.7}$$

Here, we cannot simply use (4.6) as the $(\cdot)^2$ transform is not linear. Let us assume that $\xi \in \mathcal{P}$ and it could be expressed as (4.2), i.e. we have

$$\xi_t = Z\mathbb{1}_{\{0\}}(t) + \sum_{k=0}^{n-1} Z_k \mathbb{1}_{(t_k, t_{k+1}]}(t).$$

Let $t \ge 0$ and let $\mathbb{M} := \{1, 2, \ldots, n-1\}$. We consider three cases: $t = 0$, $t \ge t_n$, and $t \in (0, t_n)$.

1) Case 1: $t = 0$. We get $I_t(\xi) = 0$ so (4.7) is satisfied.

2) Case 2: $t \geq t_n$. We have

$$I_t(\xi) = I(\xi) = \sum_{k=0}^{n-1} Z_k (W_{t_{k+1}} - W_{t_k}),$$

Consequently,

$$\begin{aligned}
\mathbb{E}[I_t^2(\xi)] &= \sum_{j,k \in \mathbb{M}} \mathbb{E}\left[Z_j Z_k (W_{t_{j+1}} - W_{t_j})(W_{t_{k+1}} - W_{t_k})\right] \\
&= \sum_{k \in \mathbb{M}} \mathbb{E}\left[Z_k^2 (W_{t_{k+1}} - W_{t_k})^2\right] + 2 \sum_{\substack{j>k \\ j,k \in \mathbb{M}}} \mathbb{E}\left[Z_j Z_k (W_{t_{j+1}} - W_{t_j})(W_{t_{k+1}} - W_{t_k})\right]
\end{aligned}$$

Now, for any fixed $j, k \in \mathbb{M}$ such that $j > k$ we get

$$\begin{aligned}
\mathbb{E}\left[Z_j Z_k (W_{t_{j+1}} - W_{t_j})(W_{t_{k+1}} - W_{t_k})\right] &= \mathbb{E}\left[\mathbb{E}\left[Z_j Z_k (W_{t_{j+1}} - W_{t_j})(W_{t_{k+1}} - W_{t_k}) \mid \mathcal{F}_{t_j}\right]\right] \\
&= \mathbb{E}\left[Z_j Z_k (W_{t_{k+1}} - W_{t_k}) \mathbb{E}\left[W_{t_{j+1}} - W_{t_j} \mid \mathcal{F}_{t_j}\right]\right] \\
&= \mathbb{E}\left[Z_j Z_k (W_{t_{k+1}} - W_{t_k}) \mathbb{E}\left[W_{t_{j+1}} - W_{t_j}\right]\right] \\
&= 0. \qquad (4.8)
\end{aligned}$$

On the other hand, using Proposition 3.8 for any $t > s$ we get

$$\begin{aligned}
\mathbb{E}[(W_t - W_s)^2 | \mathcal{F}_s] &= \mathbb{E}[W_t^2 - 2W_t W_s + W_s^2 \mid \mathcal{F}_s] \\
&= \mathbb{E}[W_t^2 - W_s^2 \mid \mathcal{F}_s] \\
&= \mathbb{E}[W_t^2 - t \mid \mathcal{F}_s] + t - W_s^2 \\
&= W_s^2 - s + t - W_s^2 \\
&= t - s,
\end{aligned}$$

so that

$$\begin{aligned}
\sum_{k \in \mathbb{M}} \mathbb{E}\left[Z_k^2 (W_{t_{k+1}} - W_{t_k})^2\right] &= \sum_{k \in \mathbb{M}} \mathbb{E}\left[\mathbb{E}\left[Z_k^2 (W_{t_{k+1}} - W_{t_k})^2 \mid \mathcal{F}_{t_k}\right]\right] \\
&= \sum_{k \in \mathbb{M}} \mathbb{E}\left[\mathbb{E} Z_k^2 \left[(W_{t_{k+1}} - W_{t_k})^2 \mid \mathcal{F}_{t_k}\right]\right] \\
&= \sum_{k \in \mathbb{M}} \mathbb{E}\left[Z_k^2 (t_{k+1} - t_k)\right] \qquad (4.9)
\end{aligned}$$

Finally, noting that for any $t \in \mathbb{R}$ we get

$$\begin{aligned}
\int_0^t \xi_s^2 \, \mathrm{d}s &= \int_0^t \left[\sum_{k=0}^{n-1} Z_k^2 \mathbb{1}_{(t_k, t_{k+1}]}(s)\right] \mathrm{d}s \\
&= \sum_{k=0}^{n-1} Z_k^2 \int_0^t \mathbb{1}_{(t_k, t_{k+1}]}(s) \, \mathrm{d}s \\
&= \sum_{k=0}^{n-1} Z_k^2 (t_{k+1} - t_k). \qquad (4.10)
\end{aligned}$$

Combining (4.8), (4.9), and (4.10) we get

$$\mathbb{E}[I_t^2(\xi)] = \mathbb{E}\left[\int_0^t \xi_s^2 \, \mathrm{d}s\right],$$

which concludes this part of the proof.

3) Case 3: $t \in (0, t_n)$. We know that there exists $j \in \{0, \ldots, n-1\}$ such that $t \in (t_j, t_{j+1}]$. We get

$$\begin{aligned}\mathbb{E}[I_t^2(\xi)] &= \mathbb{E}\left[\left(I_{t_j}(\xi) + I_{t_j,t}(\xi)\right)^2\right] \\ &= \mathbb{E}\left[I_{t_j}^2(\xi)\right] + 2\,\mathbb{E}\left[I_{t_j}(\xi) I_{t_j,t}(\xi)\right] + \mathbb{E}\left[I_{t_j,t}^2(\xi)\right]. \end{aligned} \tag{4.11}$$

Following all the steps of the previous case we get

$$\mathbb{E}\left[I_{t_j}^2(\xi)\right] = \mathbb{E}\left[\int_0^{t_j} \xi_s^2 \,\mathrm{d}s\right]. \tag{4.12}$$

Next, using the fact that $(I_t(\xi))_{t\in\mathbb{T}}$ is a martingale we get

$$\begin{aligned}\mathbb{E}\left[I_{t_j}(\xi) I_{t_j,t}(\xi)\right] &= \mathbb{E}\left[\mathbb{E}\left[I_{t_j}(\xi) I_{t_j,t}(\xi) \mid \mathcal{F}_{t_j}\right]\right] \\ &= \mathbb{E}\left[I_{t_j}(\xi)\,\mathbb{E}\left[I_{t_j,t}(\xi) \mid \mathcal{F}_{t_j}\right]\right] \\ &= \mathbb{E}\left[I_{t_j}(\xi)\,\mathbb{E}\left[I_t(\xi) - I_{t_j}(\xi) \mid \mathcal{F}_{t_j}\right]\right] \\ &= \mathbb{E}\left[I_{t_j}(\xi)\,\mathbb{E}\left[I_{t_j}(\xi) - I_{t_j}(\xi) \mid \mathcal{F}_{t_j}\right]\right] \\ &= 0. \end{aligned} \tag{4.13}$$

Finally, using the fact that $t \in (t_j, t_{j+1}]$, as in the previous case, we get

$$\begin{aligned}\mathbb{E}\left[I_{t_j,t}^2(\xi)\right] &= \mathbb{E}\left[Z_j^2 (W_t - W_{t_j})^2\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[Z_j^2 (W_t - W_{t_j})^2 \mid \mathcal{F}_{t_j}\right]\right] \\ &= \mathbb{E}\left[Z_j^2\,\mathbb{E}\left[(W_t - W_{t_j})^2 \mid \mathcal{F}_{t_j}\right]\right] \\ &= \mathbb{E}\left[Z_j^2 (t - t_j)\right] \\ &= \mathbb{E}\left[\int_{t_j}^{t} \xi_s^2 \,\mathrm{d}s\right]. \end{aligned} \tag{4.14}$$

Combining (4.12), (4.13), and (4.14) with (4.11) we conclude the proof. □

4.2 Extending Ito integral to $\mathcal{L}^2$ space.

In this section we outline how to extend Ito integral from the space $\mathcal{P}$ to the space of square-integrable processes, i.e. processes of class $\mathcal{L}^2$. For brevity, we skip the proofs.

Definition 4.3. We say that the stochastic process X is of class $\mathcal{L}^2$ (i.e. $X \in \mathcal{L}^2$) if X is adapted, measurable, and for any $t > 0$ we get

$$\mathbb{E}\left[\int_0^t X_s^2 \,\mathrm{d}s\right] < \infty.$$

On $\mathcal{L}^2$, for any $t \in \mathbb{T}$ we define the seminorm

$$\|X\|_{\mathcal{L}^2,t} := \mathbb{E}\left[\int_0^t X_s^2 \,\mathrm{d}s\right].$$

The associated metric is given by

$$\rho_{\mathcal{L}^2}(X,Y) := \sum_{n=1}^{\infty} 2^{-n}(\|X-Y\|_{\mathcal{L}^2,n} \wedge 1), \quad X,Y \in \mathcal{L}^2.$$

It can be shown that $\rho_{\mathcal{L}^2}$ is indeed a metric, and that $(\mathcal{L}^2, \rho_{\mathcal{L}^2})$ is a Polish space. Nevertheless, the proof of this is out of scope of this lecture. Now, we need to show that we can approximate any process from $\mathcal{L}^2$ using processess from $\mathcal{P}$. This is the statement of Proposition 4.4.

Proposition 4.4. *The class of $\mathcal{P}$ processes is dense in $\mathcal{L}^2$, i.e. for any $X \in \mathcal{L}^2$ there exists a sequence $(\xi^n)_{n=1}^{\infty}$ of processes, such that $\xi_n \in \mathcal{P}$ (for $n \in \mathbb{N}$), and $\xi_n \to X$ (in $\rho_{\mathcal{L}^2}$).*

The proof of proposition (4.4) is typically split into several parts. First, we show that $\mathcal{P}$ is dense in the subspace of bounded processes with non-zero values on some finite time interval. Then, we extend the reasoning to continuous processes, progressively measurable processes, and finally to measurable and adapted processes (on space $\mathcal{L}^2$). The last extension is typically done using the Dellacherie-Meyer theorem, which states that any adapted and measurable process has progressively measurable modification.

Next, it can be shown the for any $X \in \mathcal{L}^2$ the Ito integrals for the approximating sequence from $\mathcal{P}$ do converge in the mean-square metric.

Proposition 4.5. *Let $X \in \mathcal{L}^2$ and let $(\xi^n)_{n=1}^{\infty}$ be a sequence of processes from $\mathcal{P}$ such that $\xi^n \to X$ (in $\rho_{\mathcal{L}^2}$). Then,*

1) For any $t \in \mathbb{T}$, the sequence of random variables $(I_t(\xi^n))_{n\in\mathbb{N}}$ converges in $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$.

2) If $(\nu^n)_{n=1}^{\infty}$ is sequence of processes from $\mathcal{P}$ such that $\nu^n \to X$ (in $\rho_{\mathcal{L}^2}$), then for any $t \in \mathbb{T}$ we get

$$\lim_{n\to\infty} I_t(\xi^n) = \lim_{n\to\infty} I_t(\nu^n), \quad \textit{in } L^2(\Omega, \mathcal{F}_t, \mathbb{P})$$

The proof od Proposition (4.5) is a simple consequence of Ito isometry combined with the fact that $(\mathcal{L}^2, \rho_{\mathcal{L}^2})$ is a Polish space. This proposition allow us to properly define Ito integral for any process from $\mathcal{L}^2$.

Definition 4.6. Let $X \in \mathcal{L}^2$. For any $t >$ we define the Ito integral of X at time $t \in \mathbb{T}$ as the L^2–limit of the sequence $(I_t(\xi^n))_n^{\infty}$, where $(\xi^n)_{n=1}^{\infty}$ is such that $\xi^n \in \mathcal{P}$ and

$$\xi^n \to X \text{ (in } \rho_{\mathcal{L}^2}).$$

We denote the limit by $I_t(X)$ or $\int_0^t X_s \, \mathrm{d}W_x$.

In the next Theorem we outline the basic properties of Ito integrals. Note that most of the properties are a direct extension of properties given in Theorem 4.2.

Theorem 4.7. *Let $X, Y \in \mathcal{L}^2$. Then*

1. *For any $\alpha, \beta \in \mathbb{R}$ and $t \geq 0$ we get $I_t(\alpha X + \beta Y) = \alpha I_t(X) + \beta I_t(T)$.*
2. *The process $(I_t(X))_{t\in\mathbb{T}}$ has continuous paths starting from 0 (i.e. $I_0(X) = 0$).*

3. *The process* $(I_t(X))_{t\in\mathbb{T}}$ *is a square-integrable martingale (wrt.* $\mathbb{F}$*). Moreover, for any* $\infty > t \geq k \geq 0$ *we get*

$$\mathbb{E}\left[\int_0^t X_s \, \mathrm{d}W_s\right] = 0,$$

$$\mathbb{E}\left[\left(\int_0^t X_s \, \mathrm{d}W_s\right)^2\right] = \mathbb{E}\left[\int_0^t X_s^2 \, \mathrm{d}s\right],$$

$$\mathbb{E}\left[\left(\int_0^t X_s \, \mathrm{d}W_s\right)\left(\int_0^t Y_s \, \mathrm{d}W_s\right)\right] = \mathbb{E}\left[\int_0^t X_s Y_s \, \mathrm{d}s\right],$$

$$\mathbb{E}\left[\left(\int_0^t X_s \, \mathrm{d}W_s - \int_0^k X_s \, \mathrm{d}W_s\right)^2 \mid \mathcal{F}_s\right] = \int_k^t X_s^2 \, \mathrm{d}s.$$

As before, we call the second equality in 3. the **Ito isometry**.

Final end remark: The typical next step in the Ito calculus is to extend the Ito integral from $\mathcal{L}^2$ space to $\mathcal{P}^2$ space, i.e. the space of measurable and adapted processes such that

$$\mathbb{P}\left[\int_0^t X_s^2 \, \mathrm{d}s < \infty, \text{ for any } t \in \mathbb{T}\right] = 1.$$

In the introduction to stochastic analysis we continue integra construction. We also define the class of Ito processes, and show why above definitions are useful, especially in mathematical finance (e.g. in Black-Scholes framework).

A Appendix

A.1 Exemplary list of questions (and scope) for the exam

1) Conditional expectation – definition, basic properties, examples.
2) Conditional expectation as the best least-square predictor.
3) Definition of the stochastic processes and basic (continuity) properties.
4) Stochastic process modifications, finite-dimensional distributions, and stationarity.
5) Properties of continuous modifications, Holder continuity, and Kolmogorov's continuity theorems.
6) Cylinder sets, consistency property, and Kolmogorov"s extension theorem.
7) Filtrations – basic definitions, properties, and examples.
8) Process measurability and progressive measurability. Examples.
9) Stopping times in continuous time – definition and basic properties.
10) First entry time, and different types of events constructed using stopping times.
11) Operations on stopping times (sup, +, etc.), and different characterisations of stopping times.
12) Stochastic process random time sample, stopped process, and stopping sigma algebra.
13) Martingales – definition, and basic properties.
14) Regular martingales, and concave transforms of martingales
15) Stopped processes and martingale property.
16) Optional sampling theorem in discrete-time.
17) Doob-Meyer decomposition (in discrete and continuous time).
18) Uniform integrability, (D) and (DL) property, (DL) property for cadlag martingales.
19) Levy processes – definition and examples.
20) Brownian Motion – definition, and basic properties
21) Existence of the Brownian motion.
22) Transformations of Brownian motion.
23) Continuity properties for Brownian motion (Holder continuity, Path differentiability, Total variation).
24) Martingale characterisation of Brownian motion.
25) Poisson Process - definition and basic properties.
26) Two definition of the Poisson process and their equivalence.
27) Process jumps, poisson random measure, and Levy measure.
28) Markov processes – definition, and examples.
29) Probability kernels, transition functions, and construction of Markov processes.
30) Elementary (predictible) processes and their's Ito integral.
31) Basic properties of Ito integral.
32) Ito integral on $\mathcal{L}^2$ space.

A.2 Notation

$(\Omega, \mathcal{F}, \mathbb{P})$	The probability space. If not stated otherwise, we assume that it will always be the underlying space
$L_0(\Omega, \mathcal{F}, \mathbb{P})$	The set of all (a.s. identified) random variables on $(\Omega, \mathcal{F}, \mathbb{P})$
$L_1(\Omega, \mathcal{F}, \mathbb{P})$	The set of all (a.s. identified) integrable RVs on $(\Omega, \mathcal{F}, \mathbb{P})$
$L_p(\Omega, \mathcal{F}, \mathbb{P})$	The set of all (a.s. identified) square-integrable RVs on $(\Omega, \mathcal{F}, \mathbb{P})$
$\sigma(X)$	σ-algebra generated by X
$\mathbb{1}_A(\cdot)$	characteristic function of set A
$\delta_x(\cdot)$	Dirac delta function for point x
A^c	Complement of set A, i.e. $\Omega \setminus A$.
$\mathcal{B}_{\mathbb{T}}$	The set of all cylinder sets on $\mathbb{T}$.
$\mathcal{T}$	The set of all finite subsets of $\mathbb{T}$.
$\langle X, X \rangle$	The quadratic variation of X.
Σ_s	The set of all stopping times with values up to time s, i.e. $\{\tau : \tau \leq s\}$.
C_X	Auto-covariance function of process X, i.e. $C_X(t,s) = \mathrm{Cov}(X_t, X_s)$
B_0	The family of Borel sets $U \subset \mathbb{R}$ whose closure does not contain 0.
$\mathcal{C}(E)$	The space of all bounded and measurable functions $f : E \to \mathbb{R}$.
P_t	Transition (time-homogenous) kernel at time t
$\mathcal{P}$	The family of all elementary (predictible) processes.
$\mathcal{L}^2$	The family of all adapted, measurable stochastic processes such that for any $t > 0$ we get $\mathbb{E}\left[\int_0^t X_s^2 \, \mathrm{d}s\right] < \infty$.
$I(X)$	$\int_0^\infty X_s \, \mathrm{d}W_s$, i.e. stochastic Ito integral of the process X on $[0, \infty)$.
$I_t(X)$	$\int_0^t X_s \, \mathrm{d}W_s$.
$I_{s,t}(X)$	$\int_s^t X_s \, \mathrm{d}W_s$.

RV	Random variable.
SP	Stochastic process.
BM	Brownian Motion (Wiener process).
PP	Poisson process.