OUTLINE OF THE THEORY OF STOCHASTIC PROCESSES

In all our considerations we shall assume that we are given a fixed probability space $(\Omega, \mathfrak{F}, P)$, and that all random variables in question are defined on it. Let us recall that a probability P is a normalised measure on $\mathfrak{F}$ having the properties:

1. For every $A \in \mathfrak{F}$, $P(A') = 1 - P(A)$.
2. For every $A, B \in \mathfrak{F}$, if $A \subset B$, then

$$P(A) \leqslant P(B) \qquad \text{and} \qquad P(B \setminus A) = P(B) - P(A).$$

3. For every pairwise disjoint $A_1, \ldots, A_m \in \mathfrak{F}$

$$P\Big(\bigcup_{k=1}^{m} A_k\Big) = \sum_{k=1}^{m} P(A_k) \text{ — finite additivity.}$$

4. For every $A_n \in \mathfrak{F}$, $n = 1, 2, \ldots$,

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leqslant \sum_{n=1}^{\infty} P(A_n) \text{ — countable subadditivity.}$$

5. For every $A_1, \ldots, A_m \in \mathfrak{F}$

$$P\Big(\bigcup_{k=1}^{m} A_k\Big) \leqslant \sum_{k=1}^{m} P(A_k) \text{ — finite subadditivity.}$$

6. (a) For every ascending sequence of events (A_n), $A_n \subset A_{n+1}$,

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} P(A_n) \text{ — “continuity” of probability,}$$

(b) For every descending sequence of events (B_n), $B_{n+1} \subset B_n$,

$$P\Big(\bigcap_{n=1}^{\infty} B_n\Big) = \lim_{n\to\infty} P(B_n) \text{ — “continuity” of probability.}$$

In the theory of stochastic processes, we usually assume that the probability space $(\Omega, \mathfrak{F}, P)$ is *complete*, i.e. that the σ-field $\mathfrak{F}$ contains all subsets of the sets of probability zero. This assumption is not restrictive in any way since it can be proved that every measure space can be completed (roughly speaking, this completion consists in adding to $\mathfrak{F}$ all subsets of sets of measure zero).

Note the following consequence of the properties of probability.

Lemma 1. *Let $A_n \in \mathfrak{F}$, $n = 1, 2, \ldots$, be such that $P(A_n) = 1$. Then*

$$P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = 1.$$

Proof. For the complementary event, we have

$$0 \leqslant P\Big(\big(\bigcap_{n=1}^{\infty} A_n\big)'\Big) = P\Big(\bigcup_{n=1}^{\infty} A_n'\Big) \leqslant \sum_{n=1}^{\infty} P(A_n') = 0,$$

which proves the lemma. □

Definition. A stochastic process is a family of random variables $(X_t : t \in T)$, where T is an arbitrary index set. A stochastic process is sometimes called also a *random function*.

Remark. Due to the arbitrariness of the set T, we have e.g. that a random vector $(X_1, \ldots, X_n)$ is a stochastic process ($T = \{1, \ldots, n\}$), or that a sequence of random variables $(X_1, X_2, \ldots)$ is a stochastic process ($T = \mathbb{N}$). From the point of view of the theory of stochastic processes these cases are not interesting and in general we shall assume that T **is an interval on the line** (usually $T = [a, b]$ or $T = [0, \infty)$, or $T = \mathbb{R}$).

Let us notice that the definition above gives rise to three different ways of looking at a stochastic process. First, a process is simply a family of random variables. Second, a process is a function of two variables $T \times \Omega \ni (t, \omega) \mapsto X_t(\omega)$ with real values such that for each fixed t, the function $\Omega \ni \omega \mapsto X_t(\omega)$ is measurable. Third (most "sofisticated") a process is a map which to each element $\omega \in \Omega$ assigns the function $T \ni t \mapsto X_t(\omega)$ defined on T with real values (thus a process is here an "infinite dimensional random variable" mapping Ω in $\mathbb{R}^T$ — the space of all functions defined on T with values in $\mathbb{R}$, analogously to the case $T = \{1, \ldots, n\}$ where a process is an n-dimensional random variable: $\Omega \ni \omega \mapsto (X_1(\omega), \ldots, X_n(\omega)) \in \mathbb{R}^n$). Because of this last situation a stochastic process is often denoted by $(X(t) : t \in T)$ (or a little more precisely $(X(t, \cdot) : t \in T)$) and then the functions $t \mapsto X_t(\omega)$ mentioned above have a simple notation $X(\cdot, \omega)$. These functions are called *samples* or *trajectories* or *realisations* or *paths* of the process.

If we consider a process as a function of two variables, then we also use the notation $X(\cdot, \cdot)$, and then $X_t(\omega) = X(t, \omega)$.

Definition. A stochastic process $(X_t : t \in T)$ is said to be *continuous with probability one*, if its samples are continuous with probability one, i.e. if the set $\{\omega : \text{ function } X(\cdot, \omega) \text{ is continuous}\}$ is an event and has probability one (in other words, if almost all samples are continuous).

In the definition above one important element can be seen. Namely, in order that one can speak about the continuity of samples, a topology in the set T (possibly nontrivial) must be assumed (roughly speaking, the relation $t \to t_0$ must be defined). This is why the case of finite T or $T = \mathbb{N}$ is not interesting because on such sets all functions are continuous (of course, under a natural assumption that these sets are given the discrete topologies).

Definition. A process $(X_t : t \in T)$ is said to be *continuous in probability at point $t_0 \in T$*, if

$$X_t \underset{t \to t_0}{\longrightarrow} X_{t_0} \quad \textit{in probability}.$$

A process is said to be *continuous in probability*, if it is continuous in probability at each point.

The comment about the previous definition can (and should) be repeated here, namely, in the set T there must be some topology in order that one can speak about convergence $t \to t_0$.

Let us recall two basic modes of convergence considered in probability theory: convergence with probability one and convergence in probability — for simplicity we restrict attention to sequences.

Definition. A sequence of random variables (X_n) is said to *converge with probability one* to a random variable X, if the set $\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}$ is an event and has probability one (in other words, for almost all $\omega \in \Omega$, we have $X_n(\omega) \underset{n \to \infty}{\longrightarrow} X(\omega)$).

Definition. A sequence of random variables (X_n) is said to *converge in probability* to a random variable X, if for any $\varepsilon > 0$

$$P(|X_n - X| \geqslant \varepsilon) \underset{n \to \infty}{\longrightarrow} 0, \quad \text{equivalently} \quad P(|X_n - X| < \varepsilon) \underset{n \to \infty}{\longrightarrow} 1.$$

Remark. Similarly as in a "customary" definition of a limit, the arbitrariness of $\varepsilon > 0$ allows us to restrict attention to ε of the form $\varepsilon = \frac{1}{k}$ with arbitrary $k \in \mathbb{N}$.

A relation between these modes of convergence is as follows.

Theorem 2. *If a sequence of random variables (X_n) converges to a random variable X with probability one, then it converges to X in probability.*

Proof. The proof hinges on the following representation which, in turn, is a consequence of the very definition of limit.

$$\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\} = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \Big\{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{k}\Big\}.$$

(For each ω belonging to the right-hand side we have that for every $k \in \mathbb{N}$ there is $m \in \mathbb{N}$ such that for every $n \geqslant m$ the inequality $|X_n(\omega) - X(\omega)| < \frac{1}{k}$ holds, which is an "ε-definition" of the relation $\lim_{n \to \infty} X_n(\omega) = X(\omega)$ with $\varepsilon = \frac{1}{k}$). Denote the set on the left-hand side of the equality above by A:

$$A = \{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}.$$

From the assumption, we have $P(A) = 1$, hence

$$P\Big(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\Big\{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{k}\Big\}\Big) = 1,$$

which means that for any $k \in \mathbb{N}$, we have

$$P\Big(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\Big\{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{k}\Big\}\Big) = 1.$$

Set

$$A_m = \bigcap_{n=m}^{\infty}\Big\{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{k}\Big\}.$$

The sequence of events $(A_m : m = 1, 2 \dots)$ is ascending, and we have

$$P\Big(\bigcup_{m=1}^{\infty} A_m\Big) = 1,$$

consequently, the "continuity" of probability yields

$$\lim_{m\to\infty} P(A_m) = P\Big(\bigcup_{m=1}^{\infty} A_m\Big) = 1.$$

Since obviously

$$\bigcap_{n=m}^{\infty}\Big\{\omega : |X_n(\omega) - X(\omega)| < \frac{1}{k}\Big\} \subset \Big\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\Big\},$$

we obtain

$$P\Big(\Big\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\Big\}\Big) \underset{m\to\infty}{\longrightarrow} 1$$

for every $k \in \mathbb{N}$, which shows the convergence in probability of the sequence (X_m) to X. □

Remark. The above theorem can not be reversed, i.e. convergence in probability does not imply convergence with probability one, still the following theorem holds true.

Theorem 3. *If a sequence of random variables (X_n) converges in probability to a random variable X, then there is a subsequence (X_{k_n}) converging to X with probability one.*

Before proving this, recall known from elementary probability theory

Borel-Cantelli Lemma. *Let (A_n) be a sequence of events such that $\sum_{n=1}^{\infty} P(A_n) < \infty$. Then*

$$P\Big(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\Big) = 0.$$

Proof. Put

$$B_m = \bigcup_{n=m}^{\infty} A_n.$$

(B_m) is a descending sequence of events, thus the "continuity" and countable subadditivity of probability yield

$$\begin{aligned} P\Big(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\Big) &= P\Big(\bigcap_{m=1}^{\infty} B_m\Big) = \lim_{m\to\infty} P(B_m) \\ &= \lim_{m\to\infty} P\Big(\bigcup_{n=m}^{\infty} A_n\Big) \leqslant \lim_{m\to\infty}\sum_{n=m}^{\infty} P(A_n) = 0, \end{aligned}$$

since on the right-hand side we have a reminder of a convergent series. □

Proof of Theorem 3. Let n be fixed. Since

$$\lim_{m\to\infty} P\Big(|X_m - X| \geqslant \frac{1}{n}\Big) = 0,$$

there exists k_n such that

$$P\Big(|X_{k_n} - X| \geqslant \frac{1}{n}\Big) < \frac{1}{2^n},$$

and certainly we may assume that (k_n) is an increasing sequence. Let

$$A_n = \Big\{\omega : |X_{k_n}(\omega) - X(\omega)| \geqslant \frac{1}{n}\Big\}.$$

The Borel-Cantelli Lemma yields

$$P\Big(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\Big) = 0,$$

thus

$$\begin{aligned} 1 = P\Big(\big(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\big)'\Big) &= P\Big(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty} A_n'\Big) \\ &= P\Big(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\Big\{\omega : |X_{k_n}(\omega) - X(\omega)| < \frac{1}{n}\Big\}\Big). \end{aligned}$$

Put

$$A = \bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\Big\{\omega : |X_{k_n}(\omega) - X(\omega)| < \frac{1}{n}\Big\}.$$

Then $P(A) = 1$, and for $\omega \in A$ there exists $m \in \mathbb{N}$ such that for all $n \geqslant m$ the inequality

$$|X_{k_n}(\omega) - X(\omega)| < \frac{1}{n}$$

holds, so $X_{k_n}(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega)$, hence $X_{k_n} \underset{n\to\infty}{\longrightarrow} X$ *with probability one* (on the set A). $\square$

In the problem that follows, we show that in some specific situations convergence in probability is equivalent to convergence with probability one.

Problem 1. Let $\Omega = \{\omega_1, \omega_2, \dots\}$ be a discrete space, let $\mathfrak{F} = 2^{\Omega}$, and let $P(\{\omega_i\}) > 0$, $\sum_i P(\{\omega_i\}) = 1$. Assume that a sequence (X_n) of random variables on Ω *converges in probability* to a random variable X. Then $X_n \underset{n\to\infty}{\longrightarrow} X$ *with probability one.*

Solution. We must show that $X_n(\omega_i) \underset{n\to\infty}{\longrightarrow} X(\omega_i)$ for each i. To the contrary, assume that $X_n(\omega_{i_0}) \nrightarrow X(\omega_{i_0})$ for some i_0. Then there is $\varepsilon_0 > 0$ and a subsequence (k_n) such that

$$|X_{k_n}(\omega_{i_0}) - X(\omega_{i_0})| \geqslant \varepsilon_0 \quad \text{for all } n.$$

The subsequence (X_{k_n}) also converges in probability to X, and putting

$$A_n = \{\omega : |X_{k_n}(\omega) - X(\omega)| < \varepsilon_0\},$$

we have $P(A_n) \to 1$ and $\omega_{i_0} \notin A_n$ for every n, thus

$$A_n \subset \Omega \setminus \{\omega_{i_0}\}.$$

Hence we get

$$P(A_n) \leqslant P(\Omega \setminus \{\omega_{i_0}\}) = 1 - P(\{\omega_{i_0}\}),$$

and passing to the limit

$$1 = \lim_{n\to\infty} P(A_n) \leqslant 1 - P(\{\omega_{i_0}\}) < 1,$$

a contradiction.

Next two problems deal with the above modes of convergence.

Problem 2. Assume that for a sequence of random variables (X_n) and a random variable X we have $\sum_{n=1}^{\infty} P(|X_n - X| \geqslant \varepsilon_n) < \infty$ for $\varepsilon_n \to 0$. Show that $X_n \to X$ *with probability one.*

Solution. From the Borel-Cantelli Lemma we get

$$P\Big(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| \geqslant \varepsilon_n\}\Big) = 0,$$

i.e.

$$P\Big(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon_n\}\Big) = 1.$$

For
$$\omega \in \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon_n\})$$
we have that there is $m \in \mathbb{N}$ such that for all $n \geqslant m$ the following inequality holds
$$|X_n(\omega) - X(\omega)| < \varepsilon_n,$$
which means that $X_n(\omega) \to X(\omega)$, hence $X_n \to X$ *with probability one*.

Problem 3. Let for random variables X i Y
$$d(X,Y) = \int_{\Omega} \frac{|X-Y|}{1+|X-Y|}\, dP.$$
Show that for a sequence (X_n) of random variables and a random variable X, $X_n \to X$ *in probability* if and only if $d(X_n, X) \to 0$.

Solution. Let Z be an arbitrary non-negative random variable, and let $\varepsilon > 0$ be arbitrary. The following estimate holds true
$$\begin{aligned}
\int_{\Omega} \frac{Z}{1+Z}\, dP = & \int_{\{Z<\varepsilon\}} \frac{Z}{1+Z}\, dP + \int_{\{Z\geqslant\varepsilon\}} \frac{Z}{1+Z}\, dP \\
\leqslant & \int_{\{Z<\varepsilon\}} Z\, dP + \int_{\{Z\geqslant\varepsilon\}} 1\, dP \leqslant \int_{\{Z<\varepsilon\}} \varepsilon\, dP + P(Z \geqslant \varepsilon) \\
= & \varepsilon P(Z<\varepsilon) + P(Z \geqslant \varepsilon) \leqslant \varepsilon + P(Z \geqslant \varepsilon).
\end{aligned} \tag{1}$$
Consider the function
$$f(t) = \frac{t}{1+t}, \quad t \in [\varepsilon, \infty).$$
This function is increasing, thus it takes its minimum in the leftmost point of the domain which leads to the inequality
$$\frac{t}{1+t} \geqslant \frac{\varepsilon}{1+\varepsilon} \quad \text{for} \quad t \geqslant \varepsilon.$$
The above yields
$$\begin{aligned}
\int_{\Omega} \frac{Z}{1+Z}\, dP = & \int_{\{Z<\varepsilon\}} \frac{Z}{1+Z}\, dP + \int_{\{Z\geqslant\varepsilon\}} \frac{Z}{1+Z}\, dP \\
\geqslant & \int_{\{Z\geqslant\varepsilon\}} \frac{Z}{1+Z}\, dP \geqslant \int_{\{Z\geqslant\varepsilon\}} \frac{\varepsilon}{1+\varepsilon}\, dP = \frac{\varepsilon}{1+\varepsilon} P(Z \geqslant \varepsilon)
\end{aligned} \tag{2}$$
Assume that $X_n \to X$ *in probability*. From the inequality (1), we obtain, putting $Z = |X_n - X|$,
$$d(X_n, X) \leqslant \varepsilon + P(|X_n - X| \geqslant \varepsilon)$$

and taking n_0 such that for $n \geqslant n_0$ we have $P(|X_n - X| \geqslant \varepsilon) < \varepsilon$, we get

$$d(X_n, X) \leqslant \varepsilon + P(|X_n - X| \geqslant \varepsilon) < 2\varepsilon \text{ for } n \geqslant n_0,$$

which proves that $d(X_n, X) \to 0$.

Now let $d(X_n, X) \to 0$. From the inequality (2), we obtain that for arbitrary $\varepsilon > 0$

$$\frac{\varepsilon}{1+\varepsilon} P(|X_n - X| \geqslant \varepsilon) \to 0,$$

thus $P(|X_n - X| \geqslant \varepsilon) \to 0$ for arbitrary $\varepsilon > 0$, which proves that $X_n \to X$ *in probability*.

From Theorem 2, we obtain the following corollary showing that the continuity with probability one of a stochastic process is a stronger property than the continuity in probability of this process.

Corollary 4. *If a stochastic process $(X_t : t \in T)$ is continuous with probability one, then it is continuous in probability.*

Indeed, let

$$A = \{\omega : \text{ function } X(\cdot, \omega) \text{ is continuous}\}.$$

We have $P(A) = 1$, and for $\omega \in A$ and arbitrary t_0 the continuity of the samples $X(\cdot, \omega)$ yields $X_t(\omega) \underset{t \to t_0}{\longrightarrow} X_{t_0}(\omega)$. Thus $X_t \underset{t \to t_0}{\longrightarrow} X_t$ *with probability one* (for all $\omega \in A$), so $X_t \underset{t \to t_0}{\longrightarrow} X_{t_0}$ *in probability*, which means the continuity in probability of the process at point t_0. The arbitrariness of t_0 yields the continuity in probability of the process.

The following problem has a solution analogous to the solution of Problem 1.

Problem 4. If Ω is a discrete space, then the continuity in probability of the process $(X_t : t \in T)$ is equivalent to the continuity of all samples.

Solution. Let $t_0 \in T$ be arbitrary. The continuity in probability of the process $(X_t : t \in T)$ yields

$$X_t \underset{t \to t_0}{\longrightarrow} X_{t_0} \quad \textit{in probability},$$

and since Ω is discrete, it follows that

$$X_t \underset{t \to t_0}{\longrightarrow} X_{t_0} \quad \textit{with probability one},$$

thus there is an event A such that $P(A) = 1$ and for every $\omega \in A$ the relation

$$X_t(\omega) \underset{t \to t_0}{\longrightarrow} X_{t_0}(\omega)$$

holds. We have $\Omega = \{\omega_1, \omega_2, \dots\}$, $P(\{\omega_i\}) > 0$ for every $i = 1, 2, \dots$, so the only random event having probability one is the whole of Ω (no ω_i can be left out of such a random event). Consequently, $A = \Omega$ which means that for every ω we have

$$X(t,\omega) \underset{t\to t_0}{\longrightarrow} X(t_0,\omega)$$

showing that each sample $X(\cdot,\omega)$ is continuous at t_0. Since t_0 was arbitrary the conclusion follows.

Problem 5. Let $\Omega = (0,1)$, $\mathfrak{F} = \mathcal{B}((0,1))$, P — Lebesgue measure. For $t \in [0,\infty)$ define on Ω functions X_t by the formula

$$X_t(\omega) = \begin{cases} 0, & \text{for } \omega > t \\ 1, & \text{for } \omega \leqslant t \end{cases}, \quad \omega \in (0,1).$$

Show that the process $(X_t : t \geqslant 0)$ is continuous in probability.

Solution. Since for each fixed t the function X_t takes only two values, it is measurable, which proves that X_t is a random variable, so $(X_t : t \geqslant 0)$ is a stochastic process. Let $t_0 \geqslant 0$ be arbitrary. For $0 < \varepsilon < 1$, we have

$$\begin{aligned} &P(|X_t - X_{t_0}| \geqslant \varepsilon) = P(X_t = 0, X_{t_0} = 1) + P(X_t = 1, X_{t_0} = 0) \\ =&P(\{\omega : \omega > t, \omega \leqslant t_0\}) + P(\{\omega : \omega \leqslant t, \omega > t_0\}). \end{aligned} \tag{3}$$

1. Let $t_0 \geqslant 1$. Then $\{\omega : \omega \leqslant t, \omega > t_0\} = \emptyset$, thus

$$\begin{aligned} &P(|X_t - X_{t_0}| \geqslant \varepsilon) = P(\{\omega : \omega > t, \omega \leqslant t_0\}) \\ =&P(\{\omega : \omega > t\}) = \begin{cases} 1-t, & \text{dla } t < 1 \\ 0, & \text{dla } t \geqslant 1 \end{cases} \end{aligned}$$

For $t_0 = 1$, we have, for the left-hand and right-hand limits, respectively,

$$\lim_{t\to 1-} P(|X_t - X_{t_0}| \geqslant \varepsilon) = \lim_{t\to 1-} (1-t) = 0$$

and

$$\lim_{t\to 1+} P(|X_t - X_{t_0}| \geqslant \varepsilon) = \lim_{t\to 1+} 0 = 0,$$

so $\lim_{t\to 1} P(|X_t - X_{t_0}| \geqslant \varepsilon) = 0$.

For $t_0 > 1$, we have, for t sufficiently close to t_0, the relation $t > 1$, hence $\lim_{t\to t_0} P(|X_t - X_{t_0}| \geqslant \varepsilon) = \lim_{t\to t_0} 0 = 0$. Consequently, for $t_0 \geqslant 1$, we have $\lim_{t\to t_0} P(|X_t - X_{t_0}| \geqslant \varepsilon) = 0$, which means that the process $(X_t : t \geqslant 0)$ is continuous in probability at all points of the interval $[1,\infty)$.

2. Let now $t_0 < 1$. For $t < t_0$, we get from the formula (3)

$$\begin{aligned} &P(|X_t - X_{t_0}| \geqslant \varepsilon) = P(\{\omega : \omega > t, \omega \leqslant t_0\}) \\ =&P(\{\omega : t < \omega \leqslant t_0\}) = t_0 - t. \end{aligned}$$

For $t > t_0$, we get from the same formula

$$P(|X_t - X_{t_0}| \geqslant \varepsilon) = P(\{\omega : \omega > t_0, \omega \leqslant t\})$$
$$= P(\{\omega : t_0 < \omega \leqslant t\}) = t - t_0,$$

thus

$$\lim_{t\to t_0-} P(|X_t - X_{t_0}| \geqslant \varepsilon) = \lim_{t\to t_0-}(t_0 - t) = 0$$

and

$$\lim_{t\to t_0+} P(|X_t - X_{t_0}| \geqslant \varepsilon) = \lim_{t\to t_0+}(t - t_0) = 0,$$

which means that $\lim\limits_{t\to t_0} P(|X_t - X_{t_0}| \geqslant \varepsilon) = 0$, so the process is continuous in probability at all points of the interval $[0, 1)$.

Let us note that all samples of the process above are discontinuous functions (for fixed ω, the sample $X(\cdot, \omega)$ has a jump equal to one at point $t = \omega$), so we have an example of a process continuous in probability which is not continuous with probability one.

Theorem 5. *Let $(X_t : t \in T)$ be a stochastic process continuous in probability, and let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous function. Then the process $(f(X_t) : t \in T)$ is continuous in probability.*

Proof. Assume to the contrary that the process $(f(X_t) : t \in T)$ is not continuous in probability. Then there exists $t_0 \in T$ such that $f(X_t) \underset{t\to t_0}{\not\to} f(X_{t_0})$ *in probability*, so there exists $\varepsilon_0 > 0$ such that

$$P(|f(X_t) - f(X_{t_0})| \geqslant \varepsilon_0) \underset{t\to t_0}{\not\to} 0.$$

It follows that there are $\delta > 0$ and a sequence $t_n \to t_0$, such that

$$P(|f(X_{t_n}) - f(X_{t_0})| \geqslant \varepsilon_0) > \delta \ \text{ for all } \ n \in \mathbb{N}. \tag{4}$$

On account of continuity in probability of the process $(X_t : t \in T)$ we obtain

$$X_{t_n} \to X_{t_0} \quad \textit{in probability},$$

so from Theorem 2 it follows that there is a subsequence (t_{k_n}) such that

$$X_{t_{k_n}} \to X_{t_0} \quad \textit{with probability one.}$$

This in turn means that there is an event A such that $P(A) = 1$ and for $\omega \in A$ we have

$$X_{t_{k_n}}(\omega) \to X_{t_0}(\omega).$$

Since f is continuous, we obtain that for $\omega \in A$

$$f\big(X_{t_{k_n}}(\omega)\big) \to f\big(X_{t_0}(\omega)\big),$$

thus

$$f\big(X_{t_{k_n}}\big) \to f\big(X_{t_0}\big) \quad \textit{with probability one.}$$

In particular,

$$f\left(X_{t_{k_n}}\right) \to f\left(X_{t_0}\right) \quad \textit{in probability},$$

which contradicts the relation (4) which should hold also for the subsequence (t_{k_n}) of the sequence (t_n). $\square$

The problem below shows that the assumptions of independence and the identity of distributions, which are natural for sequences of random variables, lead in the case of stochastic processes to a significant irregularity of a process.

Problem 6. Let $(X_t : t \in T)$ be a stochastic process such that the random variables X_t are pairwise independent and have the same nondegenerated (i.e. not concentrated at a point) distributions. Then this process is not continuous in probability at any point.

Solution. Assume to the contrary that the process $(X_t : t \in T)$ is continuous in probability at a point t_0. For arbitrary $\varepsilon > 0$, we have

$$P(|X_t - X_{t_0}| < \varepsilon) \underset{t \to t_0}{\longrightarrow} 1.$$

Let

$$\Delta_\varepsilon = \{(x, y) \in \mathbb{R}^2 : |x - y| < \varepsilon\}.$$

According to the definition of a joint distribution $\mu_{X_t, X_{t_0}}$, we have

$$\begin{aligned} \mu_{X_t, X_{t_0}}(\Delta_\varepsilon) &= P\left(\{\omega : (X_t(\omega), X_{t_0}(\omega)) \in \Delta_\varepsilon\}\right) \\ &= P\left(\{\omega : |X_t(\omega) - X_{t_0}(\omega)| < \varepsilon\}\right). \end{aligned} \tag{5}$$

Let μ be the common distribution of all random variables X_t. From the independence of X_t and X_{t_0}, we have

$$\mu_{X_t, X_{t_0}} = \mu_{X_t} \otimes \mu_{X_{t_0}} = \mu \otimes \mu,$$

where $\mu \otimes \mu$ stands for the product measure. The equality (5) yields

$$\mu \otimes \mu(\Delta_\varepsilon) = \mu_{X_t, X_{t_0}}(\Delta_\varepsilon) = P\left(\{\omega : |X_t(\omega) - X_{t_0}(\omega)| < \varepsilon\}\right) \underset{t \to t_0}{\longrightarrow} 1.$$

Since the left-hand side of the relation above is independent of t and of t_0, we infer that for arbitrary $\varepsilon > 0$

$$\mu \otimes \mu(\Delta_\varepsilon) = 1.$$

Let $\varepsilon = \frac{1}{n}$. The sequence of sets $\left(\Delta_{\frac{1}{n}}\right)$ is descending and

$$\bigcap_{n=1}^{\infty} \Delta_{\frac{1}{n}} = \{(x, y) \in \mathbb{R}^2 : x = y\}.$$

The "continuity" of the measure $\mu \otimes \mu$ yields

$$\begin{aligned}\mu \otimes \mu(\{(x,y) \in \mathbb{R}^2 : x = y\}) &= \mu \otimes \mu\Big(\bigcap_{n=1}^{\infty} \Delta_{\frac{1}{n}}\Big) \\ &= \lim_{n\to\infty} \mu \otimes \mu\Big(\Delta_{\frac{1}{n}})\Big) = 1\end{aligned}$$

By definition, μ is not concentrated at any single point which means that its distribution function F does not take only the values 0 and 1. Let $0 < c < 1$ be a value of F, and let t_0 be a point at which F takes the value c,

$$F(t_0) = c.$$

We have

$$\mu((-\infty, t_0]) = F(t_0) = c > 0$$

and

$$\mu((t_0, \infty)) = 1 - c > 0,$$

which means that

$$\mu \otimes \mu((-\infty, t_0] \times (t_0, \infty)) = \mu((-\infty, t_0])\mu((t_0, \infty)) = c(1-c) > 0.$$

Moreover,

$$((-\infty, t_0] \times (t_0, \infty)) \cap \{(x,y) : x = y\} = \emptyset,$$

so

$$\begin{aligned}1 &= \mu \otimes \mu(\mathbb{R} \times \mathbb{R}) \geqslant \mu \otimes \mu(((-\infty, t_0] \times (t_0, \infty)) \cup \{(x,y) : x = y\}) \\ &= \mu \otimes \mu((-\infty, t_0] \times (t_0, \infty)) + \mu \otimes \mu(\{(x,y) : x = y\}) \\ &= c(1-c) + 1 > 1,\end{aligned}$$

a contradiction.

Problem 7. Let X be a symmetric random variable such that $P(X = 0) = 0$, and let Y be an arbitrary random variable. Define a stochastic process $(Z_t : t \geqslant 0)$ by the formula

$$Z_t = t(X + t) + Y.$$

Find the probability that the samples of the process $(Z_t : t \geqslant 0)$ are increasing functions.

Solution. **1. approach (general).** We must find

$$\begin{aligned}&P(Z_{t_1} < Z_{t_2} \text{ for all } 0 \leqslant t_1 < t_2) \\ =&P(Z_{t_2} - Z_{t_1} > 0 \text{ for all } 0 \leqslant t_1 < t_2).\end{aligned}$$

For fixed $0 \leqslant t_1 < t_2$, we have

$$\begin{aligned}
&\{\omega : Z_{t_2}(\omega) - Z_{t_1}(\omega) > 0\}\\
=&\{\omega : (t_2(X(\omega) + t_2) + Y(\omega)) - (t_1(X(\omega) + t_1) + Y(\omega)) > 0\}\\
=&\{\omega : (t_2 - t_1)X(\omega) + (t_2^2 - t_1^2) > 0\}\\
=&\{\omega : (t_2 - t_1)(X(\omega) + (t_1 + t_2)) > 0\}\\
=&\{\omega : X(\omega) + (t_1 + t_2) > 0\} = \{\omega : X(\omega) > -(t_1 + t_2)\}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\{\omega : Z_{t_2}(\omega) - Z_{t_1}(\omega) > 0 \text{ for all } 0 \leqslant t_1 < t_2\}\\
=&\{\omega : X(\omega) > -(t_1 + t_2) \text{ for all } 0 \leqslant t_1 < t_2\}\\
=&\{\omega : X(\omega) \geqslant 0\}.
\end{aligned}$$

The condition that the random variable X is symmetric means, by definition, that for arbitrary $B \in \mathcal{B}(\mathbb{R})$ we have

$$P(X \in B) = P(X \in -B),$$

where $-B = \{-x : x \in B\}$. In particular, for $B = (0, \infty)$, we get

$$P(X > 0) = P(X \in (0, \infty)) = P(X \in (-\infty, 0)) = P(X < 0).$$

Further we have

$$\begin{aligned}
1 &= P(X < 0) + P(X = 0) + P(X > 0)\\
&= P(X < 0) + P(X > 0) = 2P(X > 0),
\end{aligned}$$

which yields

$$P(X > 0) = \frac{1}{2}.$$

Finally,

$$\begin{aligned}
&P(\{\omega : Z_{t_2}(\omega) - Z_{t_1}(\omega) > 0 \text{ for all } 0 \leqslant t_1 < t_2\})\\
=&P(\{\omega : X(\omega) \geqslant 0\})\\
=&P(\{\omega : X(\omega) > 0\}) + P(\{\omega : X(\omega) = 0\}) = \frac{1}{2}.
\end{aligned}$$

2. approach (concrete). For any $\omega \in \Omega$, the sample is the function

$$[0, \infty) \ni t \mapsto t(X(\omega) + t) + Y(\omega),$$

i.e. it is a *quadratic function*. Since a constant term can be neglected while considering monotonicity, it is enough to consider the function

$$[0, \infty) \ni t \mapsto t(X(\omega) + t).$$

This function has two zeros: 0 and $-X(\omega)$, and increases in the interval $[0, \infty)$ if and only if $-X(\omega) \leqslant 0$, i.e. if $X(\omega) \geqslant 0$. The probability of this event was calculated at point 1 and equals $\frac{1}{2}$.

Problem 8. Let $\Omega = [0,1]$, $\mathfrak{F} = \mathcal{B}([0,1])$, P — Lebesgue measure, $T = [0,1]$, and let

$$X_t(\omega) = \begin{cases} 1, & \text{if } \omega \in \left[0, \frac{1}{2}\right] \text{ and } t \in \left[\frac{1}{2}, 1\right] \\ 0, & \text{otherwise} \end{cases}.$$

Find the probability that the samples of the process (X_t) are continuous functions.

Solution. Fix $\omega \in \left[0, \frac{1}{2}\right]$. For the sample $X(\cdot, \omega)$ we have

$$X(t, \omega) = \begin{cases} 1, & \text{for } t \in \left[\frac{1}{2}, 1\right] \\ 0, & \text{for } t \in \left[0, \frac{1}{2}\right) \end{cases}$$

thus the sample $X(\cdot, \omega)$ is discontinuous at point $\frac{1}{2}$. If $\omega \in \left(\frac{1}{2}, 1\right]$, then $X(\cdot, \omega) \equiv 0$, i.e. this sample is a continuous function. Thus the samples are continuous for $\omega \in \left(\frac{1}{2}, 1\right]$ and discontinuous for $\omega \in \left[0, \frac{1}{2}\right]$, so the probability that the samples are continuous equals $P\left(\left(\frac{1}{2}, 1\right]\right) = \frac{1}{2}$.

Definition. Assume that in T there is a σ-field $\mathfrak{M}$ of subsets of T. The process $(X_t : t \in T)$ is said to be *measurable* if it is measurable as a function of two variables $T \times \Omega \ni (t, \omega) \mapsto X_t(\omega)$ with respect to the product σ-field $\mathfrak{M} \otimes \mathfrak{F}$ — σ-field generated by the sets of the form $B \times A$, $B \in \mathfrak{M}$, $A \in \mathfrak{F}$. (Since we consider here a process as a function of two variables, the most convenient notation is $X(\cdot, \cdot)$).

Proposition 6. *Let* $T = [0, \infty)$, $\mathfrak{M} = \mathcal{B}([0, \infty))$, *and let the process* $X(\cdot, \cdot)$ *be measurable. Let* $\tau \colon \Omega \to [0, \infty)$ *be a random variable. Define a function* $X_\tau \colon \Omega \to \mathbb{R}$ *by the formula*

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega) = X(\tau(\omega), \omega)$$

Then X_τ *is a random variable.*

Proof. Let a function $f \colon \Omega \to [0, \infty) \times \Omega$ be defined by the formula

$$f(\omega) = (\tau(\omega), \omega).$$

For arbitrary $A \in \mathfrak{F}$ and $B \in \mathcal{B}([0, \infty))$, we have

$$\begin{aligned} f^{-1}(B \times A) &= \{\omega : f(\omega) \in B \times A\} = \{\omega : (\tau(\omega), \omega) \in B \times A\} \\ &= \{\omega : \tau(\omega) \in B, \omega \in A\} = \{\omega : \tau(\omega) \in B\} \cap A \\ &= \tau^{-1}(B) \cap A \in \mathfrak{F}, \end{aligned}$$

since by virtue of measurability of τ, $\tau^{-1}(B) \in \mathfrak{F}$. Thus the function f is measurable. Since

$$X_\tau(\omega) = X(\tau(\omega), \omega) = (X \circ f)(\omega),$$

X_τ is measurable as a composition of measurable functions. $\square$

Remark. Random variables τ as above are called *stopping times* of the process $(X_t : t \in T)$. They play a significant role in martingale theory. If, for instance, we are given a sequence of random variables $(X_1, X_2, \dots)$ and $\tau\colon \Omega \to \mathbb{N}$, then $X_\tau = X_n$ on the set $\{\omega : \tau(\omega) = n\}$.

Problem 9. Prove that a stochastic process $(X_t : t \in (0,1])$ with all the samples right continuous is measurable.

Solution. Define functions $X_n\colon [0,1] \times \Omega \to \mathbb{R}$ by the formula

$$X_n(t,\omega) = X\Big(\frac{k}{n},\omega\Big) \quad \text{for } t \in \Big(\frac{k-1}{n}, \frac{k}{n}\Big], \; k = 1,\dots,n,$$

i.e.

$$X_n(t,\omega) = \sum_{k=1}^{n} X\Big(\frac{k}{n},\omega\Big)\chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}(t),$$

where χ_E is the indicator function of the set E:

$$\chi_E(t) = \begin{cases} 0, & \text{dla } t \notin E \\ 1, & \text{dla } t \in E \end{cases}.$$

For arbitrary $B \in \mathcal{B}(\mathbb{R})$, and arbitrary fixed $k = 1,\dots,n$, we have

$$\begin{aligned}
&\Big(X\Big(\frac{k}{n},\cdot\Big)\chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}\Big)^{-1}(B) = \Big\{(t,\omega) : X\Big(\frac{k}{n},\omega\Big)\chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}(t) \in B\Big\} \\
=&\big\{(t,\omega) : 0 \in B,\, \chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}(t) = 0\big\} \cup \\
\cup&\Big\{(t,\omega) : X\Big(\frac{k}{n},\omega\Big) \in B,\, \chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}(t) = 1\Big\} \\
=&\big\{(t,\omega) : 0 \in B,\, \chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}(t) = 0\big\} \cup \\
\cup&\Big(\Big(\frac{k-1}{n},\frac{k}{n}\Big] \times \Big\{\omega : X\Big(\frac{k}{n},\omega\Big) \in B\Big\}\Big) \\
=&\big\{(t,\omega) : 0 \in B,\, \chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}(t) = 0\big\} \cup \Big(\Big(\frac{k-1}{n},\frac{k}{n}\Big] \times X\Big(\frac{k}{n},\cdot\Big)^{-1}(B)\Big).
\end{aligned}$$

The second set in the union above belongs to the σ-field $\mathcal{B}((0,1]) \otimes \mathfrak{F}$, since certainly $\left(\frac{k-1}{n},\frac{k}{n}\right] \in \mathcal{B}((0,1])$ and $X\Big(\frac{k}{n},\cdot\Big)^{-1}(B) \in \mathfrak{F}$ because $X\Big(\frac{k}{n},\cdot\Big)$ is a random variable. For the first set, we have that it is empty if $0 \notin B$, but if $0 \in B$, it is equal $\big((0,1] \setminus \left(\frac{k-1}{n},\frac{k}{n}\right]\big) \times \Omega$, so in both the cases this sest belongs to the σ-field $\mathcal{B}((0,1]) \otimes \mathfrak{F}$. Consequently, the function $X\Big(\frac{k}{n},\cdot\Big)\chi_{\left(\frac{k-1}{n},\frac{k}{n}\right]}$ is measurable as a function of two variables (t,ω), thus X_n is measurable as a sum of measurable functions.

Let now t and ω be arbitrary and fixed, and let $\varepsilon > 0$ be arbitrary. The right-hand continuity of the sample $X(\cdot, \omega)$ at point t yields that there exists $\delta > 0$ such that for all t' satisfying the inequality $0 < t' - t < \delta$ we have

$$|X(t', \omega) - X(t, \omega)| < \varepsilon.$$

Let n_0 be such that $\frac{1}{n_0} < \delta$. For each $n \geqslant n_0$, there is $k_n \in \{1, \ldots, n\}$ such that $t \in \left(\frac{k_n - 1}{n}, \frac{k_n}{n}\right]$, so

$$|X_n(t, \omega) - X(t, \omega)| = \left| X\left(\frac{k_n}{n}, \omega\right) - X(t, \omega) \right| < \varepsilon,$$

since $0 < \frac{k_n}{n} - t < \frac{1}{n} < \delta$. This shows that

$$\lim_{n \to \infty} X_n(t, \omega) = X(t, \omega),$$

consequently, $X(\cdot, \cdot)$ is a measurable function as a limit of measurable functions.

Remark. The set $T = (0, 1]$ in the problem above was taken only for the sake of simple notation. In fact, the measurability of the process proved in this problem holds if T is an arbitrary interval $T \subset \mathbb{R}$ (a proof needs only a small change of notation).

Definition. By *finite dimensional distributions* of a stochastic process $(X_t : t \in T)$ are meant the distributions of all random vectors $(X_{t_1}, \ldots, X_{t_n})$ for arbitrary $t_1, \ldots, t_n \in T$, and arbitrary $n = 1, 2, \ldots$.

For the finite dimensional distributions of a process, the symbol $\mu_{t_1, \ldots, t_n}$ is sometimes used, so we have

$$\begin{aligned} \mu_{t_1, \ldots, t_n}(B) &= P((X_{t_1}, \ldots, X_{t_n})^{-1}(B)) \\ &= P(\{\omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}^n). \end{aligned}$$

Definition. Processes $(X_t : t \in T)$ and $(Y_t : t \in T)$ are said to be *equivalent* if for every $t \in T$

$$P(X_t = Y_t) = 1, \quad \text{equivalently} \quad P(X_t \neq Y_t) = 0.$$

In such a case we speak that the process (Y_t) is a *modification* of the process (X_t) (or that the process (X_t) is a modification of the process (Y_t)).

Remark. The definition above states that for each $t \in T$ there is an event $\Omega_t = \{\omega : X_t(\omega) = Y_t(\omega)\}$ such that $P(\Omega_t) = 1$, still we can have *distinct* Ω_t for *distinct* t's. It is seen in the example below.

Example. Let $\Omega = [0,1]$, $\mathfrak{F} = \mathcal{B}([0,1])$, P — Lebesgue measure, $T = [0,1]$. Define processes

$$X_t(\omega) = \begin{cases} 0, & \text{for } \omega \neq t \\ 1, & \text{for } \omega = t \end{cases}, \quad Y_t(\omega) = 0 \quad \text{for all } \omega, t \in [0,1].$$

For each $t \in [0,1]$, we have

$$\{\omega : X_t(\omega) \neq Y_t(\omega)\} = \{t\},$$

thus

$$\Omega_t = \{\omega : X_t(\omega) = Y_t(\omega)\} = [0,1] \setminus \{t\},$$

and since $P(\Omega_t) = P([0,1] \setminus \{t\}) = 1$, the processes $(X_t : t \in T)$ and $(Y_t : t \in T)$ are equivalent. At the same time, for distinct t's the events Ω_t are distinct.

Theorem 7. *Equivalent processes have the same finite dimensional distributions.*

Proof. Let the processes $(X_t : t \in T)$ and $(Y_t : t \in T)$ be equivalent. For arbitrary fixed $t_1, \dots, t_n \in T$, let

$$A_1 = \{\omega : X_{t_1}(\omega) = Y_{t_1}(\omega)\}, \dots, A_n = \{\omega : X_{t_n}(\omega) = Y_{t_n}(\omega)\},$$
$$A = A_1 \cap \dots \cap A_n.$$

From the equivalence of the processes $(X_t : t \in T)$ and $(Y_t : t \in T)$, it follows that $P(A_k) = 1$ for $k = 1, \dots, n$, so $P(A) = 1$. Moreover, for $\omega \in A$, we have

$$(X_{t_1}(\omega), \dots, X_{t_n}(\omega)) = (Y_{t_1}(\omega), \dots, Y_{t_n}(\omega)).$$

Note that for an arbitrary event C, we have

$$C = (A \cap C) \cup (A' \cap C),$$

so

$$P(C) = P(A \cap C) + P(A' \cap C) = P(A \cap C),$$

because

$$P(A' \cap C) \leqslant P(A') = 0.$$

For arbitrary $B \in \mathcal{B}(\mathbb{R}^n)$, the relations above yield

$$\begin{aligned} &P(\{\omega : (X_{t_1}(\omega), \dots, X_{t_n}(\omega) \in B\}) \\ =&P(A \cap \{\omega : (X_{t_1}(\omega), \dots, X_{t_n}(\omega)) \in B\}) \\ =&P(A \cap \{\omega : (Y_{t_1}(\omega), \dots, Y_{t_n}(\omega)) \in B\}) \\ =&P(\{\omega : (Y_{t_1}(\omega), \dots, Y_{t_n}(\omega)) \in B\}), \end{aligned}$$

which shows that the distributions of the random vectors $(X_{t_1}, \dots, X_{t_n})$ and $(Y_{t_1}, \dots, Y_{t_n})$ are the same. $\square$

Definition. A stochastic process $(X_t : t \in T)$ is said to be of the *second order* if for each $t \in T$, we have $\mathbb{E}X_t^2 < \infty$. The *correlation function* K of such a process is defined by the formula

$$\begin{aligned} K(s,t) &= \operatorname{cov}(X_s, X_t) = \mathbb{E}(X_s - \mathbb{E}X_s)(X_t - \mathbb{E}X_t) \\ &= \mathbb{E}X_s X_t - \mathbb{E}X_s \mathbb{E}X_t, \quad s,t \in T. \end{aligned}$$

For second order processes yet another mode of continuity may be introduced.

Definition. A stochastic process $(X_t : t \in T)$ of the second order is said to be *continuous in the mean* at point $t_0 \in T$ if

$$\mathbb{E}(X_t - X_{t_0})^2 \xrightarrow[t \to t_0]{} 0.$$

A process is said to be *continuous in the mean* if it is continuous in the mean at every point.

Remark. The definition above refers to the theory of Hilbert spaces. Namely, the space of real-valued square integrable functions on an arbitrary measure space (E, μ) is a real Hilbert space, denoted by $L^2_{\mathbb{R}}(E, \mu)$, with the inner product

$$\langle f | g \rangle = \int_E f(t) g(t)\, \mu(dt)$$

and norm

$$\|f\|_2 = \Big(\int_E f^2(t)\, \mu(dt) \Big)^{1/2}.$$

(More precisely, for a correct definition, instead of functions the equivalence classes of functions should be considered, i.e. functions which differ only on a set of measure zero should be identified). With such an approach, random variables are square integrable functions and the continuity in the mean denotes the continuity with respect to the norm $\|\cdot\|_2$, since we have

$$\mathbb{E}(X_t - X_{t_0})^2 = \int_\Omega \big(X_t(\omega) - X_{t_0}(\omega)\big)^2 P(d\omega) = \|X_t - X_{t_0}\|_2^2.$$

Recall now the fundamental Schwarz inequality.

Schwarz inequality. *Let X and Y be random variables with finite second moments, $\mathbb{E}X^2 < \infty$, $\mathbb{E}Y^2 < \infty$. Then*

$$|\mathbb{E}XY| \leqslant \big(\mathbb{E}X^2\big)^{1/2} \big(\mathbb{E}Y^2\big)^{1/2}. \tag{6}$$

Proof. Consider the function

$$f(t) = \mathbb{E}(tX - Y)^2, \quad t \in \mathbb{R}.$$

This function is non-negative as the expectation of a non-negative random variable, and we have

$$f(t) = (\mathbb{E}X^2)t^2 - 2(\mathbb{E}XY)t + \mathbb{E}Y^2.$$

Note that from the inequality
$$2|XY| \leqslant X^2 + Y^2$$
it follows that the function $|XY|$ is integrable since it is non-negative and bounded from above by a sum of two integrable functions thus there exists a finite $\mathbb{E}|XY|$ and because $|\mathbb{E}XY| \leqslant \mathbb{E}|XY|$, there exists a finite $\mathbb{E}XY$, which means that the function f takes finite values. Assume that $\mathbb{E}X^2 \neq 0$. Then f is a non-negative quadratic function having a positive coefficient at t^2, so we must have
$$0 \geqslant \Delta = \big(-2(\mathbb{E}XY)\big)^2 - 4\mathbb{E}X^2\mathbb{E}Y^2,$$
hence
$$\mathbb{E}X^2\mathbb{E}Y^2 \geqslant \big(\mathbb{E}XY\big)^2,$$
and the inequality (6) follows.

If $\mathbb{E}X^2 = 0$, then $X = 0$ (with probability one) and we have zero on both sides of the inequality (6). □

Remark. In the theory of Hilbert spaces, the Schwarz inequality has the form
$$|\langle x|y\rangle| \leqslant \|x\|\|y\|$$
where x and y are arbitrary elements of the Hilbert space in question, $\langle \cdot|\cdot \rangle$ is the inner product in this space, and the norm $\|\cdot\|$ is given by the formula
$$\|x\| = \big(\langle x|x\rangle\big)^{1/2}.$$
This general Schwarz inequality is proven by considering the inequality
$$0 \leqslant \langle \alpha x - y|\alpha x - y\rangle,$$
for arbitrary $\alpha \in \mathbb{C}$, using properties of the inner product, and putting $\alpha = \frac{\langle y|x\rangle}{\|x\|^2}$.

Observe the following simple consequence of the continuity in the mean of a process.

Lemma 8. *If a second order process $(X_t : t \in T)$ is continuous in the mean at point $t_0 \in T$, then $\big(\mathbb{E}X_t^2\big)^{1/2} \underset{t\to t_0}{\longrightarrow} \big(\mathbb{E}X_{t_0}^2\big)^{1/2}$.*

Proof. For arbitrary square integrable random variables X i Y we have on account of the Schwarz inequality
$$\mathbb{E}XY \leqslant \big(\mathbb{E}X^2\big)^{1/2}\big(\mathbb{E}Y^2\big)^{1/2},$$
thus
$$\mathbb{E}X^2 - 2\big(\mathbb{E}X^2\big)^{1/2}\big(\mathbb{E}Y^2\big)^{1/2} + \mathbb{E}Y^2 \leqslant \mathbb{E}X^2 - 2\mathbb{E}XY + \mathbb{E}Y^2,$$
i.e.
$$\Big(\big(\mathbb{E}X^2\big)^{1/2} - \big(\mathbb{E}Y^2\big)^{1/2}\Big)^2 \leqslant \mathbb{E}(X-Y)^2.$$

The inequality above yields, after taking $X = X_t$, $Y = X_{t_0}$, that if $\mathbb{E}\big(X_t - X_{t_0}\big)^2 \underset{t\to t_0}{\longrightarrow} 0$, then

$$\big(\mathbb{E}X_t^2\big)^{1/2} \underset{t\to t_0}{\longrightarrow} \big(\mathbb{E}X_{t_0}\big)^{1/2}. \qquad \square$$

A relation between continuity in the mean and continuity in probability is given in the theorem below. First recall the Chebyshev inequality known from elementary probability theory.

Chebyshev inequality (1. variant). *Let Z be a non-negative random variable. Then for arbitrary $\varepsilon > 0$ we have*

$$P(Z \geqslant \varepsilon) \leqslant \frac{\mathbb{E}Z}{\varepsilon}.$$

Proof. Of course, we may assume that $\mathbb{E}Z < \infty$. By virtue of the non-negativity of Z, we have

$$\mathbb{E}Z = \int_\Omega Z\,dP \geqslant \int\limits_{\{Z\geqslant\varepsilon\}} Z\,dP \geqslant \int\limits_{\{Z\geqslant\varepsilon\}} \varepsilon\,dP = \varepsilon P(Z \geqslant \varepsilon)$$

and dividing both sides by ε, we obtain the conclusion. $\square$

Chebyshew inequality (2. variant). *Let X be a random variable with finite expectation. Then for arbitrary $\varepsilon > 0$, we have*

$$P(|X - \mathbb{E}X| \geqslant \varepsilon) \leqslant \frac{\mathbb{D}^2X}{\varepsilon^2}.$$

Proof. Again, we may assume that the variance of the random variable X is finite. Then we have for $Z = (X - \mathbb{E}X)^2$ and ε^2 instead of ε

$$\begin{aligned} P(|X - \mathbb{E}X| \geqslant \varepsilon) &= P((X - \mathbb{E}X)^2 \geqslant \varepsilon^2) \\ &\leqslant \frac{\mathbb{E}(X - \mathbb{E}X)^2}{\varepsilon^2} = \frac{\mathbb{D}^2X}{\varepsilon^2}. \qquad \square \end{aligned}$$

Theorem 9. *If a process $(X_t : t \in T)$ is continuous in the mean at point $t_0 \in T$, then it is continuous in probability at this point.*

Proof. The proof follows from Chebyshev's inequality. Namely, for arbitrary $\varepsilon > 0$, we have

$$P(|X_t - X_{t_0}| \geqslant \varepsilon) = P(|X_t - X_{t_0}|^2 \geqslant \varepsilon^2) \leqslant \frac{\mathbb{E}(X_t - X_{t_0})^2}{\varepsilon^2},$$

and from the assumption, it follows that the right-hand side of the inequality above tends to zero. $\square$

Let $(X_t : t \in T)$ be a second order process and let

$$L(s,t) = \mathbb{E}X_sX_t, \quad s,t \in T.$$

We have the following characterisation of the continuity in the mean of a process.

Theorem 10. *For a second order process, the following conditions are equivalent:*

(i) *the process is continuous in the mean,*
(ii) *the function L is continuous (as a function of two variables).*

Proof. Assume that the process is continuous in the mean. Using the Schwarz inequality, we have for arbitrary $s_0, t_0 \in T$

$$\begin{aligned}
&|L(s,t) - L(s_0,t_0)| \leqslant |L(s,t) - L(s_0,t)| + |L(s_0,t) - L(s_0,t_0)| \\
=&|\mathbb{E}X_sX_t - \mathbb{E}X_{s_0}X_t| + |\mathbb{E}X_{s_0}X_t - \mathbb{E}X_{s_0}X_{t_0}| \\
=&|\mathbb{E}(X_s - X_{s_0})X_t| + |\mathbb{E}X_{s_0}(X_t - X_{t_0})| \\
\leqslant&\big(\mathbb{E}(X_s - X_{s_0})^2\big)^{1/2}\big(\mathbb{E}X_t^2\big)^{1/2} + \big(\mathbb{E}X_{s_0}^2\big)^{1/2}\big(\mathbb{E}(X_t - X_{t_0})^2\big)^{1/2}.
\end{aligned}$$

From the assumption and Lemma 8, it follows that the right hand side of the inequality above tends to zero for $s \to s_0$ and $t \to t_0$, which means that the function L is continuous.

Assume now that the function L is continuous. For arbitrary $t_0 \in T$, we have

$$\begin{aligned}
\mathbb{E}\big(X_t - X_{t_0}\big)^2 &= \mathbb{E}X_t^2 - 2\mathbb{E}X_tX_{t_0} + \mathbb{E}X_{t_0}^2 \\
&= L(t,t) - 2L(t,t_0) + L(t_0,t_0)
\end{aligned}$$

and the right hand side of the inequality above tends to zero for $t \to t_0$, which means that the process is continuous in the mean at point t_0. Since t_0 is arbitrary, we obtain the continuity in the mean of the process at each point. □

For the process $(X_t : t \in T)$, denote

$$m(t) = \mathbb{E}X_t$$

under the assumption that the expectation is finite. Analogously to Theorem 10 we get

Theorem 11. *Let $(X_t : t \in T)$ be a second order process such that the function m is continuous. The following conditions are equivalent:*

(i) *the process is continuous in the mean,*
(ii) *the correlation function K is continuous (as a function of two variables).*

Proof. The proof follows from the relation

$$K(s,t) = L(s,t) - m(s)m(t)$$

and Theorem 10. □

In the next theorem, we shall prove important properties of the correlation function of a process.

Theorem 12. *Let K be the correlation function of a second order process $(X_t : t \in T)$. Then*

(i) $K(s,t) = K(t,s)$ *for arbitrary* $s, t \in T$,

(ii) K *is* positive definite, *i.e. for arbitrary* $t_1, \ldots, t_n \in T$ *and arbitrary complex numbers* $z_1, \ldots, z_n$ *we have*

$$\sum_{j,k=1}^{n} K(t_j, t_k) z_j \overline{z_k} \geqslant 0.$$

Proof. For simplicity denote

$$\widehat{X}_t = X_t - \mathbb{E}X_t.$$

Then

$$K(s,t) = \operatorname{cov}(X_s, X_t) = \mathbb{E}\widehat{X}_s \widehat{X}_t.$$

Point (i) is obvious since

$$K(t,s) = \mathbb{E}\widehat{X}_t \widehat{X}_s = \mathbb{E}\widehat{X}_s \widehat{X}_t = K(s,t).$$

As for point (ii), we shall show first that for arbitrary *real* $u_1, \ldots, u_n$ we have

$$\sum_{j,k=1}^{n} K(t_j, t_k) u_j u_k \geqslant 0.$$

Indeed, from the properties of expectation we obtain

$$\begin{aligned}
&\sum_{j,k=1}^{n} K(t_j, t_k) u_j u_k = \sum_{j,k=1}^{n} \big(\mathbb{E}\widehat{X}_{t_j} \widehat{X}_{t_k}\big) u_j u_k = \sum_{j,k=1}^{n} \mathbb{E}\big(u_j \widehat{X}_{t_j}\big)\big(u_k \widehat{X}_{t_k}\big) \\
=&\mathbb{E}\Big(\sum_{j,k=1}^{n} \big(u_j \widehat{X}_{t_j}\big)\big(u_k \widehat{X}_{t_k}\big)\Big) = \mathbb{E}\Big(\sum_{j=1}^{n} u_j \widehat{X}_{t_j}\Big)\Big(\sum_{k=1}^{n} u_k \widehat{X}_{t_k}\Big) \\
=&\mathbb{E}\Big(\sum_{j=1}^{n} u_j \widehat{X}_{t_j}\Big)^2 \geqslant 0.
\end{aligned}$$

For arbitrary complex numbers $z_1, \ldots, z_n$, we have

$$z_j = a_j + \boldsymbol{i} b_j, \quad \text{where } a_j, b_j \in \mathbb{R},\ \boldsymbol{i} = \sqrt{-1},$$

and

$$z_j \overline{z_k} = a_j a_k + b_j b_k + \boldsymbol{i}(a_k b_j - a_j b_k),$$

so

$$\begin{aligned}
\sum_{j,k=1}^{n} K(t_j, t_k) z_j \overline{z_k} &= \sum_{j,k=1}^{n} K(t_j, t_k) a_j a_k + \sum_{j,k=1}^{n} K(t_j, t_k) b_j b_k \\
&\quad + \boldsymbol{i}\Big(\sum_{j,k=1}^{n} K(t_j, t_k) a_k b_j - \sum_{j,k=1}^{n} K(t_j, t_k) a_j b_k\Big).
\end{aligned}$$

The first two sums on the right-hand side of the equality above are, on account of the first part of the proof, nonnegative. For the next two sums we have by virtue of point (i)

$$\sum_{j,k=1}^{n} K(t_j, t_k) a_k b_j = \sum_{j,k=1}^{n} K(t_k, t_j) a_k b_j = \sum_{l,r=1}^{n} K(t_l, t_r) a_l b_r$$

after substitution $k = l$, $j = r$, and

$$\sum_{j,k=1}^{n} K(t_j, t_k) a_j b_k = \sum_{l,r=1}^{n} K(t_l, t_r) a_l b_r$$

after substitution $k = r$, $j = l$. This proves that

$$\sum_{j,k=1}^{n} K(t_j, t_k) a_k b_j = \sum_{j,k=1}^{n} K(t_j, t_k) a_j b_k,$$

and thus

$$\sum_{j,k=1}^{n} K(t_j, t_k) z_j \overline{z_k} = \sum_{j,k=1}^{n} K(t_j, t_k) a_j a_k + \sum_{j,k=1}^{n} K(t_j, t_k) b_j b_k \geqslant 0,$$

which ends the proof. $\square$

Problem 10. Let $\Omega = (0,1)$, $\mathfrak{F} = \mathcal{B}((0,1))$, P — Lebesgue measure. For $t \in [0, \infty)$ define on Ω functions X_t by the formula

$$X_t(\omega) = \begin{cases} 0, & \text{for } \omega > t \\ 1, & \text{for } \omega \leqslant t \end{cases}, \quad \omega \in (0,1).$$

a) Show that the process $(X_t : t \geqslant 0)$ is continuous in the mean.
b) Find the correlation function of the process.

Solution. First check if the process is of the second order. We have for $t \leqslant 1$

$$\mathbb{E}X_t^2 = (\mathbb{E}X_t) = \int_0^t 1\, d\omega = t,$$

and for $t > 1$

$$\mathbb{E}X_t^2 = (\mathbb{E}X_t) = \int_0^1 1\, d\omega = 1,$$

thus $\mathbb{E}X_t^2 < +\infty$.

a) For $s < t \leqslant 1$ we have

$$\big(X_t(\omega) - X_s(\omega)\big)^2 = \begin{cases} 1, & \text{if } s < \omega \leqslant t \\ 0, & \text{otherwise} \end{cases},$$

and for $1 < s < t$

$$\big(X_t(\omega) - X_s(\omega)\big)^2 = 0.$$

Take an arbitrary t_0. If $t_0 < 1$, then

$$\mathbb{E}\big(X_t(\omega) - X_{t_0}(\omega)\big)^2 = \begin{cases} \int_{t_0}^{t} 1\, d\omega = t - t_0, & \text{if } t_0 < t \\ \int_{t}^{t_0} 1\, d\omega = t_0 - t, & \text{if } t_0 > t \end{cases}$$

thus in any case

$$\mathbb{E}\big(X_t(\omega) - X_{t_0}(\omega)\big)^2 \underset{t \to t_0}{\longrightarrow} 0.$$

If $t_0 > 1$, then taking t so close to t_0 that $t > 1$ we have

$$\mathbb{E}\big(X_t(\omega) - X_{t_0}(\omega)\big)^2 = 0.$$

Finally if $t_0 = 1$, then

$$\mathbb{E}\big(X_t(\omega) - X_1(\omega)\big)^2 = \begin{cases} 0, & \text{if } t > 1 \\ \int_{t}^{1} 1\, d\omega = 1 - t, & \text{if } t < 1 \end{cases}$$

so again

$$\mathbb{E}\big(X_t(\omega) - X_1(\omega)\big)^2 \underset{t \to 1}{\longrightarrow} 0.$$

b) Let $s \leqslant t$. We have for $s \leqslant 1$

$$X_s(\omega)X_t(\omega) = \begin{cases} 1, & \text{if } \omega \leqslant s \\ 0, & \text{otherwise} \end{cases},$$

and for $s > 1$

$$X_s(\omega)X_t(\omega) = 1,$$

hence for $s \leqslant t$

$$\mathbb{E}X_sX_t = \begin{cases} \int_0^s 1\, d\omega, & \text{if } s \leqslant 1 \\ 1, & \text{otherwise} \end{cases} = \begin{cases} s, & \text{if } s \leqslant 1 \\ 1, & \text{otherwise} \end{cases}.$$

Consequently, for $s \leqslant t$

$$K(s,t) = \mathbb{E}X_sX_t - \mathbb{E}X_s\mathbb{E}X_t = \begin{cases} s - st, & \text{if } s \leqslant 1, t \leqslant 1 \\ 0, & \text{if } s \leqslant 1, t \geqslant 1 \\ 0, & \text{if } s > 1 \end{cases},$$

and analogously for $s > t$

$$K(s,t) = \mathbb{E}X_sX_t - \mathbb{E}X_s\mathbb{E}X_t = \begin{cases} t - st, & \text{if } t \leqslant 1, s \leqslant 1 \\ 0, & \text{if } t \leqslant 1, s \geqslant 1 \\ 0, & \text{if } t > 1 \end{cases}.$$

Thus

$$K(s,t)=\begin{cases} s-st, & \text{if } s\leqslant 1, t\leqslant 1, s<t\\ t-st, & \text{if } t\leqslant 1, s\leqslant 1, s\geqslant t\\ 0, & \text{otherwise}\end{cases}$$
$$=\begin{cases}\min(s,t)-st, & \text{if } s\leqslant 1, t\leqslant 1\\ 0, & \text{otherwise}\end{cases}.$$

Problem 11. Let $\Omega=(0,1)$, $\mathfrak{F}=\mathcal{B}((0,1))$, P — Lebesgue measure, $T=[0,1]$, and let

$$X_t(\omega)=\begin{cases}0, & \text{for } \omega>t\\ \omega+t, & \text{for } \omega\leqslant t\end{cases},\quad \omega\in(0,1).$$

Find the correlation function of the process.

Solution. We have

$$\mathbb{E}X_t^2=\int_0^1 X_t^2(\omega)\,d\omega=\int_0^t(\omega+t)^2\,d\omega=\frac{7}{3}t^3<+\infty,$$

i.e. (X_t) is a second order process. Further

$$\mathbb{E}X_t=\int_0^1 X_t(\omega)\,d\omega=\int_0^t(\omega+t)\,d\omega=\frac{3}{2}t^2,$$

and for $s\leqslant t$

$$\mathbb{E}X_sX_t=\int_0^1 X_s(\omega)X_t(\omega)\,d\omega=\int_0^s(\omega+s)(\omega+t)\,d\omega=\frac{5}{6}s^3+\frac{3}{2}s^2t.$$

Analogously, for $s>t$ we obtain

$$\mathbb{E}X_sX_t=\frac{5}{6}t^3+\frac{3}{2}st^2.$$

Hence

$$\begin{aligned}K(s,t)&=\mathbb{E}X_sX_t-\mathbb{E}X_s\mathbb{E}X_t\\&=\begin{cases}\frac{5}{6}s^3+\frac{3}{2}s^2t-\frac{9}{4}s^2t^2, & \text{for } s\leqslant t\\ \frac{5}{6}t^3+\frac{3}{2}st^2-\frac{9}{4}s^2t^2, & \text{for } s>t\end{cases},\quad 0\leqslant s\leqslant 1, 0\leqslant t\leqslant 1\end{aligned}$$

Problem 12. Let K_X and K_Y be the correlation functions of the processes $(X_t : t\in T)$ and $(Y_t : t\in T)$, respectively. Show that $K=K_XK_Y$ is the correlation function of some stochastic process $(Z_t : t\in T)$.

Solution. First observe that the processes $(X_t : t\in T)$ and $(Y_t : t\in T)$ *need not be defined on the same probability space*. Thus let $(X_t : t\in T)$ be defined on a probability space $(\Omega_1,\mathfrak{F}_1,P_1)$, and let $(Y_t : t\in T)$ be defined on a probability space $(\Omega_2,\mathfrak{F}_2,P_2)$. (Of course, we can have $(\Omega_1,\mathfrak{F}_1,P_1)=(\Omega_2,\mathfrak{F}_2,P_2)$, but this does not change our construction). Define a new probability space $(\Omega, \mathfrak{F}, P)$ in the following way

$$\Omega = \Omega_1 \times \Omega_2, \quad \mathfrak{F} = \mathfrak{F}_1 \otimes \mathfrak{F}_2, \quad P = P_1 \otimes P_2.$$

Define on $(\Omega, \mathfrak{F}, P)$ processes $(\widehat{X}_t : t \in T)$ and $(\widehat{Y}_t : t \in T)$ by the formulae

$$\widehat{X}_t(\omega_1, \omega_2) = X_t(\omega_1), \quad \widehat{Y}_t(\omega_1, \omega_2) = Y_t(\omega_2).$$

For arbitrary $t_1, \ldots, t_n \in T$, and arbitrary Borel set $B \subset \mathbb{R}^n$, we have

$$\begin{aligned}
&P(\{(\omega_1, \omega_2) : (\widehat{X}_{t_1}(\omega_1, \omega_2), \ldots, \widehat{X}_{t_n}(\omega_1, \omega_2)) \in B\}) \\
=&P(\{(\omega_1, \omega_2) : (X_{t_1}(\omega_1), \ldots, X_{t_n}(\omega_1)) \in B\}) \\
=&P(\{\omega_1 : (X_{t_1}(\omega_1), \ldots, X_{t_n}(\omega_1) \in B\}) \times \Omega_2) \\
=&P_1(\{\omega_1 : (X_{t_1}(\omega_1), \ldots, X_{t_n}(\omega_1)) \in B\}) P_2(\Omega_2) \\
=&P_1(\{\omega_1 : (X_{t_1}(\omega_1), \ldots, X_{t_n}(\omega_1)) \in B\}),
\end{aligned} \tag{7}$$

which shows that the finite dimensional distributions of the processes $(X_t : t \in T)$ and $(\widehat{X}_t : t \in T)$ are the same. Analogously, the finite dimensional distributions of the processes $(Y_t : t \in T)$ and $(\widehat{Y}_t : t \in T)$ are the same, in particular, it follows that for the correlation functions we have

$$K_{\widehat{X}} = K_X, \quad K_{\widehat{Y}} = K_Y.$$

For arbitrary $s_1, \ldots, s_n \in T$ and $t_1, \ldots, t_m \in T$, n, m — arbitrary, and arbitrary Borel sets $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$, we have

$$\begin{aligned}
&P(\{(\omega_1, \omega_2) : (\widehat{X}_{s_1}(\omega_1, \omega_2), \ldots, \widehat{X}_{s_n}(\omega_1, \omega_2)) \in A, \\
&\qquad (\widehat{Y}_{t_1}(\omega_1, \omega_2), \ldots, \widehat{Y}_{t_m}(\omega_1, \omega_2)) \in B\}) \\
&= P(\{(\omega_1, \omega_2) : (X_{s_1}(\omega_1), \ldots, X_{s_n}(\omega_1)) \in A, \\
&\qquad (Y_{t_1}(\omega_2), \ldots, Y_{t_m}(\omega_2)) \in B\}) \\
&= P(\{\omega_1 : (X_{s_1}(\omega_1), \ldots, X_{s_n}(\omega_1)) \in A\} \times \\
&\qquad \times \{\omega_2 : (Y_{t_1}(\omega_2), \ldots, Y_{t_m}(\omega_2)) \in B\}) \\
&= P_1(\{\omega_1 : (X_{s_1}(\omega_1), \ldots, X_{s_n}(\omega_1)) \in A\} \\
&\qquad P_2(\{\omega_2 : (Y_{t_1}(\omega_2), \ldots, Y_{t_m}(\omega_2)) \in B\} \\
&= P(\{(\omega_1, \omega_2) : (\widehat{X}_{s_1}(\omega_1, \omega_2), \ldots, \widehat{X}_{s_n}(\omega_1, \omega_2)) \in A\} \times \\
&\qquad \times P(\{(\omega_1, \omega_2) : (\widehat{Y}_{t_1}(\omega_1, \omega_2), \ldots, \widehat{Y}_{t_m}(\omega_1, \omega_2)) \in B\},
\end{aligned}$$

where the last equality follows from (7). The relation above shows that the processes $(\widehat{X}_t : t \in T)$ and $(\widehat{Y}_t : t \in T)$ are independent.

Now define on $(\Omega, \mathfrak{F}, P)$ a process $(Z_t : t \in T)$ by the formula

$$Z_t = (\widehat{X}_t - \mathbb{E}\widehat{X}_t)(\widehat{Y}_t - \mathbb{E}\widehat{Y}_t).$$

The independence of $\widehat{X}_t$ and $\widehat{Y}_t$ yields

$$\mathbb{E}Z_t = \mathbb{E}(\widehat{X}_t - \mathbb{E}\widehat{X}_t)\mathbb{E}(\widehat{Y}_t - \mathbb{E}\widehat{Y}_t) = 0,$$

and

$$\mathbb{E}Z_t^2 = \mathbb{E}(\widehat{X}_t - \mathbb{E}\widehat{X}_t)^2(\widehat{Y}_t - \mathbb{E}\widehat{Y}_t)^2 = \mathbb{E}(\widehat{X}_t - \mathbb{E}\widehat{X}_t)^2\mathbb{E}(\widehat{Y}_t - \mathbb{E}\widehat{Y}_t)^2 < \infty,$$

hence $(Z_t : t \in T)$ is a process of the second order. Using again the independence of the processes $(\widehat{X}_t : t \in T)$ and $(\widehat{Y}_t : t \in T)$, we get for the correlation function K_Z of this process

$$\begin{aligned}
&K_Z(s,t) = \mathbb{E}Z_sZ_t = \mathbb{E}(\widehat{X}_s - \mathbb{E}\widehat{X}_s)(\widehat{Y}_s - \mathbb{E}\widehat{Y}_s)(\widehat{X}_t - \mathbb{E}\widehat{X}_t)(\widehat{Y}_t - \mathbb{E}\widehat{Y}_t)\\
=&\mathbb{E}(\widehat{X}_s - \mathbb{E}\widehat{X}_s)(\widehat{X}_t - \mathbb{E}\widehat{X}_t)(\widehat{Y}_s - \mathbb{E}\widehat{Y}_s)(\widehat{Y}_t - \mathbb{E}\widehat{Y}_t)\\
=&\mathbb{E}(\widehat{X}_s - \mathbb{E}\widehat{X}_s)(\widehat{X}_t - \mathbb{E}\widehat{X}_t)\mathbb{E}(\widehat{Y}_s - \mathbb{E}\widehat{Y}_s)(\widehat{Y}_t - \mathbb{E}\widehat{Y}_t)\\
=&K_{\widehat{X}}(s,t)K_{\widehat{Y}}(s,t) = K_X(s,t)K_Y(s,t).
\end{aligned}$$

Now we shall analyse the situation when a process can be considered as a map of the space Ω into the space of functions defined on the set T: $\Omega \ni \omega \mapsto X(\cdot,\omega) \in \mathbb{R}^T$. Our first step will be introducing in the space $\mathbb{R}^T$ an appropriate σ-field.

Definition. Let $t_1,\ldots,t_n \in T$, and let $B \in \mathcal{B}(\mathbb{R}^n)$. By a *cylindric set* $C_{t_1,\ldots,t_n}(B)$ with the base B and coordinates $t_1,\ldots,t_n$ we mean a subset of $\mathbb{R}^T$ defined as

$$C_{t_1,\ldots,t_n}(B) = \{x \in \mathbb{R}^T : (x(t_1),\ldots,x(t_n)) \in B\}.$$

(In the definition above, x denotes a real-valued function defined on the set T.)

Remark. As is seen, to a cylindric set belong, in general, plenty of functions; the only restriction put on these functions is a restriction on the values they take at the points $t_1,\ldots,t_n$. For instance, if B is a one-point set

$$B = \{(a_1,\ldots,a_n)\},$$

then to $C_{t_1,\ldots,t_n}(B)$ belong all functions that at the points $t_1,\ldots,t_n$ take the values $a_1,\ldots,a_n$, respectively, and are arbitrary otherwise.

Example. Let $T = \{1,2\}$. Then $\mathbb{R}^T = \mathbb{R}^2$, and we have for arbitrary $B, B_1, B_2 \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned}
&C_1(B) = \{(x_1,x_2) \in \mathbb{R}^2 : x_1 \in B\} = B \times \mathbb{R},\\
&C_2(B) = \{(x_1,x_2) \in \mathbb{R}^2 : x_2 \in B\} = \mathbb{R} \times B,\\
&C_{1,2}(B_1 \times B_2) = \{(x_1,x_2) \in \mathbb{R}^2 : (x_1,x_2) \in B_1 \times B_2\} = B_1 \times B_2.
\end{aligned}$$

Let us note that the representation of a cylindric set is not unique. We have e.g. for arbitrary $t' \in T$

$$\begin{aligned} C_{t_1,\ldots,t_n,t'}(B \times \mathbb{R}) &= \{x \in \mathbb{R}^T : (x(t_1), \ldots, x(t_n), x(t')) \in B \times \mathbb{R}\} \\ &= \{x \in \mathbb{R}^T : (x(t_1), \ldots, x(t_n)) \in B, x(t') \in \mathbb{R}\} \\ &= \{x \in \mathbb{R}^T : (x(t_1), \ldots, x(t_n)) \in B\} = C_{t_1,\ldots,t_n}(B), \end{aligned}$$

and similarly

$$C_{t_1,\ldots,t_n}(B) = C_{t',t_1,\ldots,t_n}(\mathbb{R} \times B).$$

It follows that arbitrary cylindric sets can be written on the same system of coordinates since we have for $C_{t_1,\ldots,t_n}(B_1)$ and $C_{s_1,\ldots,s_m}(B_2)$,

$$C_{t_1,\ldots,t_n}(B_1) = C_{t_1,\ldots,t_n,s_1,\ldots,s_m}(B_1 \times \mathbb{R}^m),$$

and

$$C_{s_1,\ldots,s_m}(B_2) = C_{t_1,\ldots,t_n,s_1,\ldots,s_m}(\mathbb{R}^n \times B_2).$$

Problem 13. Let $A, B \subset \mathbb{R}$, $t_1, t_2 \in T$. Represent the set $C_{t_1}(A) \cup C_{t_2}(B)$ as a cylindric set over the coordinates t_1, t_2.

Solution. We have

$$C_{t_1}(A) = C_{t_1,t_2}(A \times \mathbb{R}) \qquad \text{and} \qquad C_{t_2}(B) = C_{t_1,t_2}(\mathbb{R} \times B)$$

hence

$$\begin{aligned} C_{t_1}(A) \cup C_{t_2}(B) &= C_{t_1,t_2}(A \times \mathbb{R}) \cup C_{t_1,t_2}(\mathbb{R} \times B) \\ &= C_{t_1,t_2}((A \times \mathbb{R}) \cup (\mathbb{R} \times B)) \end{aligned}$$

Problem 14. Let $T = [0, +\infty)$, $f(t) = 2t + 1$. Find a set $B \subset \mathbb{R}^2$ such that $f \in C_{1,2}(B)$.

Solution. The condition $f \in C_{1,2}(B)$ is equivalent to the condition $(f(1), f(2)) \in B$, i.e. $(3, 5) \in B$. Thus for B we can take any subset of $\mathbb{R}^2$ containing the point $(3, 5)$.

Problem 15. Let $T = [0, +\infty)$. Find a linear function f such that $f \in C_{1,2}((0, 1) \times (2, 3))$.

Solution. The condition $f \in C_{1,2}((0, 1) \times (2, 3))$ is equivalent to the conditions $f(1) \in (0, 1)$ and $f(2) \in (2, 3)$, so we may take e.g. $f(1) = \frac{1}{2}$, $f(2) = \frac{5}{2}$. Since f should be linear we have $f(t) = 2t - \frac{3}{2}$.

Theorem 13. *The cylindric sets constitute a field.*

Proof. We have $\emptyset = C_t(\emptyset)$ and $\mathbb{R}^T = C_t(\mathbb{R})$ for arbitrary fixed t, thus $\emptyset$ and $\mathbb{R}^T$ are cylindric sets.

For arbitrary cylindric set $C_{t_1,\ldots,t_n}(B)$, we have

$$\begin{aligned} C_{t_1,\ldots,t_n}(B)' &= \mathbb{R}^T \setminus \{x \in \mathbb{R}^T : (x(t_1), \ldots, x(t_n)) \in B\} \\ &= \{x \in \mathbb{R}^T : (x(t_1), \ldots, x(t_n)) \in B'\} = C_{t_1,\ldots,t_n}(B'), \end{aligned}$$

thus the complement of a cylindric set is also a cylindric set.

Let $C_{t_1,\dots,t_n}(B_1)$ and $C_{t_1,\dots,t_n}(B_2)$ be arbitrary cylindric sets (as we already know, they can be written on the same system of coordinates). We have

$$\begin{aligned} C_{t_1,\dots,t_n}(B_1) \cup C_{t_1,\dots,t_n}(B_2) &= \{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_n)) \in B\} \\ &\quad \cup \{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_n)) \in B\} \\ &= \{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_n)) \in B_1 \cup B_2\} \\ &= C_{t_1,\dots,t_n}(B_1 \cup B_2), \end{aligned}$$

thus a union of two cylindric sets is a cylindric set. □

The field of cylindric sets will be denoted by $\mathfrak{C}$. In the space $\mathbb{R}^T$, we shall consider the σ-field $\sigma(\mathfrak{C})$ generated by the cylindric sets. In this manner, we obtain a measurable space $(\mathbb{R}^T, \sigma(\mathfrak{C}))$.

Remark. The σ-field $\sigma(\mathfrak{C})$ is sometimes called the Borel σ-field of the space $\mathbb{R}^T$, and is denoted by $\mathcal{B}(\mathbb{R}^T)$. This name is justified by the fact that for finite T we have $\mathbb{R}^T = \mathbb{R}^n$, and the σ-field generated by the cylindric sets is just the Borel σ-field of subsets of $\mathbb{R}^n$.

The lemma that follows gives an important tool for proving measurability in most cases.

Lemma 14. *Let E be an arbitrary space with a σ-field $\mathfrak{M}$ of its subsets, let F be a space with a σ-field $\mathfrak{N}$ of its subsets, and let $\mathfrak{A} \subset \mathfrak{N}$ be such that the σ-field generated by $\mathfrak{A}$ equals $\mathfrak{N}$, $\sigma(\mathfrak{A}) = \mathfrak{N}$. Let $f\colon E \to F$ be a map such that for every $A \in \mathfrak{A}$ we have $f^{-1}(A) \in \mathfrak{M}$. Then for every $B \in \mathfrak{N}$ we have $f^{-1}(B) \in \mathfrak{M}$ (in other words: f is measurable).*

Proof. Let
$$\mathfrak{C} = \{C \subset F : f^{-1}(C) \in \mathfrak{M}\}.$$
We shall first prove that $\mathfrak{C}$ is a σ-field. Indeed, $f^{-1}(\emptyset) = \emptyset \in \mathfrak{M}$, showing that $\emptyset \in \mathfrak{C}$, and $f^{-1}(F) = E \in \mathfrak{M}$, showing that $F \in \mathfrak{C}$.

For $C \in \mathfrak{C}$, we have $f^{-1}(C') = f^{-1}(C)' \in \mathfrak{M}$, since $f^{-1}(C) \in \mathfrak{M}$, showing that $C' \in \mathfrak{C}$.

For $C_n \in \mathfrak{C}$, we have

$$f^{-1}\Big(\bigcup_{n=1}^{\infty} C_n\Big) = \bigcup_{n=1}^{\infty} f^{-1}(C_n) \in \mathfrak{M},$$

since $f^{-1}(C_n) \in \mathfrak{M}$, showing that $\bigcup_{n=1}^{\infty} C_n \in \mathfrak{C}$.

By assumption, we have $\mathfrak{A} \subset \mathfrak{C}$, thus $\mathfrak{N} = \sigma(\mathfrak{A}) \subset \mathfrak{C}$ (because $\mathfrak{C}$ is a σ-field containing $\mathfrak{A}$ and $\mathfrak{N} = \sigma(\mathfrak{A})$ is the *smallest* σ-field containing $\mathfrak{A}$). Consequently, for every $B \in \mathfrak{N}$, we have $B \in \mathfrak{C}$ which means that $f^{-1}(B) \in \mathfrak{M}$. □

Remark. The lemma above is helpful, for example, while proving that for the measurability of a random variable X it is enough to show the condition:

for every $a \in \mathbb{R}$, we have $\{\omega : X(\omega) < a\} = X^{-1}((-\infty, a)) \in \mathfrak{F}$,

since if we put

$$\mathfrak{A} = \{(-\infty, a) : a \in \mathbb{R}\},$$

then $\sigma(\mathfrak{A}) = \mathcal{B}(\mathbb{R})$. (The same holds if we take $\mathfrak{A} = \{(-\infty, a] : a \in \mathbb{R}\}$ or $\mathfrak{A} = \{(a, +\infty) : a \in \mathbb{R}\}$ or $\mathfrak{A} = \{[a, +\infty) : a \in \mathbb{R}\}$ giving another three equivalent conditions of measurability.)

Theorem 15. *Let $(X_t : t \in T)$ be a stochastic process. Define a map $\mathbb{X}\colon \Omega \to \mathbb{R}^T$ by the formula*

$$(*) \qquad \mathbb{X}(\omega)(t) = X_t(\omega), \quad \omega \in \Omega,\ t \in T.$$

Then $\mathbb{X}$ is measurable.

Conversely, for every measurable map $\mathbb{X}\colon \Omega \to \mathbb{R}^T$ define a function X_t on Ω by the formula

$$(**) \qquad X_t(\omega) = \mathbb{X}(\omega)(t), \quad \omega \in \Omega.$$

Then $(X_t : t \in T)$ is a stochastic process.
(The notation $\mathbb{X}(\omega)(t)$ as above follows from the fact that $\mathbb{X}(\omega)$ is a function on T.)

Proof. For an arbitrary cylindric set $C_{t_1,\ldots,t_n}(B)$, we have

$$\begin{aligned}\mathbb{X}^{-1}(C_{t_1,\ldots,t_n}(B)) &= \{\omega : \mathbb{X}(\omega) \in C_{t_1,\ldots,t_n}(B)\}\\ &= \{\omega : (\mathbb{X}(\omega)(t_1), \ldots, \mathbb{X}(\omega)(t_n)) \in B\}\\ &= \{\omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in B\}\\ &= (X_{t_1}, \ldots, X_{t_n})^{-1}(B) \in \mathfrak{F},\end{aligned}$$

and Lemma 14 yields the measurability of $\mathbb{X}$.

Assume now that that the map $\mathbb{X}\colon \Omega \to \mathbb{R}^T$ is measurable. For the function X_t defined by the formula $(**)$, and for arbitrary $B \in \mathcal{B}(\mathbb{R})$, the measurability of $\mathbb{X}$ yields

$$\begin{aligned}X_t^{-1}(B) &= \{\omega : X_t(\omega) \in B\} = \{\omega : \mathbb{X}(\omega)(t) \in B\}\\ &= \{\omega : \mathbb{X}(\omega) \in C_t(B)\} = \mathbb{X}^{-1}(C_t(B)) \in \mathfrak{F},\end{aligned}$$

since $C_t(B) \in \sigma(\mathfrak{C})$, which proves the measurability of X_t, thus $(X_t : t \in T)$ is a stochastic process. □

Let $\mathbb{X}$ be a measurable map from Ω to $\mathbb{R}^T$. The distribution $\mu_{\mathbb{X}}$ of $\mathbb{X}$ is defined as a probability measure on $(\mathbb{R}^T, \sigma(\mathfrak{C}))$ by the formula

$$\mu_{\mathbb{X}}(E) = P(\mathbb{X}^{-1}(E)), \quad E \in \sigma(\mathfrak{C}).$$

For the stochastic process $(X_t : t \in T)$ and the map $\mathbb{X}$ defined by the formula $(*)$, define on the probability space $(\widetilde{\Omega}, \widetilde{\mathfrak{F}}, \widetilde{P}) = (\mathbb{R}^T, \sigma(\mathfrak{C}), \mu_{\mathbb{X}})$ a function $\widetilde{X}_t$, $t \in T$ by

$$\widetilde{X}_t(\widetilde{\omega}) = \widetilde{\omega}(t), \quad \widetilde{\omega} \in \widetilde{\Omega} = \mathbb{R}^T.$$

For arbitrary $B \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned} \widetilde{X}_t^{-1}(B) &= \{\widetilde{\omega} \in \mathbb{R}^T : \widetilde{X}_t(\widetilde{\omega}) \in B\} \\ &= \{\widetilde{\omega} \in \mathbb{R}^T : \widetilde{\omega}(t) \in B\} = C_t(B) \in \sigma(\mathfrak{C}), \end{aligned}$$

thus $\widetilde{X}_t$ are measurable functions, hence $(\widetilde{X}_t : t \in T)$ is a stochastic process. This process is called the *canonical process* for the process $(X_t : t \in T)$.

Theorem 16. *The processes $(X_t : t \in T)$ and $(\widetilde{X}_t : t \in T)$ have the same finite dimensional distributions.*

Proof. For arbitrary $t_1, \ldots, t_n \in T$ and arbitrary $B \in \mathcal{B}(\mathbb{R}^n)$, we have

$$\begin{aligned} \mu_{\widetilde{X}_{t_1},\ldots,\widetilde{X}_{t_n}}(B) =& \widetilde{P}(\{\widetilde{\omega} \in \mathbb{R}^T : (\widetilde{X}_{t_1}(\widetilde{\omega}), \ldots, \widetilde{X}_{t_n}(\widetilde{\omega})) \in B\}) \\ =& \widetilde{P}(\{\widetilde{\omega} \in \mathbb{R}^T : (\widetilde{\omega}(t_1), \ldots, \widetilde{\omega}(t_n)) \in B\}) \\ =& \widetilde{P}(C_{t_1,\ldots,t_n}(B)) = \mu_{\mathbb{X}}(C_{t_1,\ldots,t_n}(B)) = P(\mathbb{X}^{-1}(C_{t_1,\ldots,t_n}(B))) \\ =& P(\{\omega : \mathbb{X}(\omega) \in C_{t_1,\ldots,t_n}(B)\}) \\ =& P(\{\omega : (\mathbb{X}(\omega)(t_1), \ldots, \mathbb{X}(\omega)(t_n)) \in B\}) \\ =& P(\{\omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in B\}) = \mu_{X_{t_1},\ldots,X_{t_n}}(B). \quad \square \end{aligned}$$

Remark. Despite its natural definition, it turns out that the σ-field $\sigma(\mathfrak{C})$ has some deficiencies. Namely, it can be shown that for T being an interval, to $\sigma(\mathfrak{C})$ do not belong the following classes of functions: continuous functions, linear functions, polynomials, non-decreasing functions, functions continuous at a fixed point.

Now we are going to discuss fundamental Kolmogorov's theorem about the existence of a stochastic process with the finite dimensional distributions given. Let $(X_t : t \in T)$ be a stochastic process, and let for arbitrary $t_1, \ldots, t_n \in T$, $\mu_{t_1,\ldots,t_n}$ be the distribution of the random vector $(X_{t_1}, \ldots, X_{t_n})$, i.e.

$$\mu_{t_1,\ldots,t_n}(B) = P((X_{t_1}, \ldots, X_{t_n}) \in B), \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Hence we have a family of distributions

$$\{\mu_{t_1,\ldots,t_n} : t_1, \ldots, t_n \in T, n = 1, 2, \ldots\},$$

such that $\mu_{t_1,\ldots,t_n}$ is a probability distribution on the space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Observe that this family fulfils the conditions:

1. For arbitrary $t \in T$

$$\mu_{t_1,\ldots,t_n,t}(B \times \mathbb{R}) = \mu_{t_1,\ldots,t_n}(B),$$

2. For an arbitrary permutation σ of the set $\{1, \dots, n\}$, and arbitrary $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$,

$$\mu_{t_1,\dots,t_n}(B_1 \times \dots \times B_n) = \mu_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(B_{\sigma(1)} \times \dots \times B_{\sigma(n)}).$$

Indeed, we have

$$\begin{aligned} \mu_{t_1,\dots,t_n,t}(B \times \mathbb{R}) &= P((X_{t_1}, \dots, X_{t_n}, X_t) \in B \times \mathbb{R}) \\ &= P((X_{t_1}, \dots, X_{t_n}) \in B) = \mu_{t_1,\dots,t_n}(B), \end{aligned}$$

and

$$\begin{aligned} \mu_{t_1,\dots,t_n}(B_1 \times \dots B_n) &= P((X_{t_1}, \dots, X_{t_n}) \in B_1 \times \dots \times B_n) \\ &= P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) \\ &= P(X_{t_{\sigma(1)}} \in B_{\sigma(1)}, \dots, X_{t_{\sigma(n)}} \in B_{\sigma(n)}) \\ &= P((X_{t_{\sigma(1)}}, \dots, X_{t_{\sigma(n)}}) \in B_{\sigma(1)} \times \dots \times B_{\sigma(n)}) \\ &= \mu_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(B_{\sigma(1)} \times \dots \times B_{\sigma(n)}). \end{aligned}$$

Conditions 1 and 2 are called the *consistency conditions*, and as is seen from the reasoning above, they are necessary in order that the distributions $\mu_{t_1,\dots,t_n}$ be the finite dimensional distributions of some stochastic process. It turns out that these conditions are also sufficient.

Kolmogorov Theorem. *Let for arbitrary $t_1, \dots, t_n \in T$, and arbitrary $n = 1, 2, \dots$, $\mu_{t_1,\dots,t_n}$ be probability distributions on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ satisfying the consistency conditions 1 and 2. Then there exists a stochastic process $(X_t : t \in T)$ such that $\mu_{t_1,\dots,t_n}$ are its finite dimensional distributions.*

The idea of the proof of this theorem is as follows. On the field $\mathfrak{C}$ of cylindric sets in the space $\mathbb{R}^T$ we define a set function μ by the formula

$$\mu(C_{t_1,\dots,t_n}(B)) = \mu_{t_1,\dots,t_n}(B). \tag{8}$$

Because of the non-uniqueness of the representation of a cylindric set it must be shown that this function is well-defined. This follows from the consistency conditions. For a cylindric set $C_{t_1,\dots,t_n}(B_1 \times \dots \times B_n)$, we have e.g. two distinct representations

$$\begin{aligned} C_{t_1,\dots,t_n}(B_1 \times \dots \times B_n) &= C_{t_1,\dots,t_n,t}(B_1 \times \dots \times B_n \times \mathbb{R})) \\ &= C_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(B_{\sigma(1)} \times \dots \times B_{\sigma(n)}), \end{aligned}$$

for arbitrary $t \in T$ and an arbitrary permutation σ of the set $\{1, \dots, n\}$, thus according to the formula (8), we should have

$$\begin{aligned} \mu(C_{t_1,\dots,t_n}(B_1 \times \dots \times B_n)) &= \mu_{t_1,\dots,t_n}(B_1 \times \dots \times B_n) \\ &= \mu_{t_1,\dots,t_n,t}(B_1 \times \dots \times B_n \times \mathbb{R}), \end{aligned}$$

and

$$\mu(C_{t_1,\dots,t_n}(B_1 \times \cdots \times B_n)) = \mu_{t_1,\dots,t_n}(B_1 \times \cdots \times B_n)$$
$$= \mu_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(B_{\sigma(1)} \times \cdots \times B_{\sigma(n)})),$$

but the equalities above hold true by virtue of the consistency conditions. A similar situation occurs for an arbitrary cylindric set $C_{t_1,\dots,t_n}(B)$. Note that μ is non-negative, and we have

$$\mu(\emptyset) = \mu(C_t(\emptyset)) = \mu_t(\emptyset) = 0, \quad \mu(\mathbb{R}^T) = \mu(C_t(\mathbb{R})) = \mu_t(\mathbb{R}) = 1.$$

Let $C_{t_1,\dots,t_n}(B_1)$ and $C_{t_1,\dots,t_n}(B_2)$ be disjoint. Then we have

$$\emptyset = C_{t_1,\dots,t_n}(B_1) \cap C_{t_1,\dots,t_n}(B_2) = C_{t_1,\dots,t_n}(B_1 \cap B_2),$$

thus $B_1 \cap B_2 = \emptyset$. It follows that

$$\mu(C_{t_1,\dots,t_n}(B_1) \cup C_{t_1,\dots,t_n}(B_2)) = \mu(C_{t_1,\dots,t_n}(B_1 \cup B_2)) = \mu_{t_1,\dots,t_n}(B_1 \cup B_2)$$
$$= \mu_{t_1,\dots,t_n}(B_1) + \mu_{t_1,\dots,t_n}(B_2)$$
$$= \mu(C_{t_1,\dots,t_n}(B_1)) + \mu(C_{t_1,\dots,t_n}(B_2)),$$

thus μ is additive.

Next it is proven that μ satisfies the conditions of the extension of measure theorem (the most difficult part), thus there exists a measure $\overline{\mu}$ on $\sigma(\mathfrak{C})$ such that for an arbitrary cylindric set $C_{t_1,\dots,t_n}(B)$ we have

$$\overline{\mu}(C_{t_1,\dots,t_n}(B)) = \mu(C_{t_1,\dots,t_n}(B)).$$

Now we define

$$\Omega = \mathbb{R}^T, \quad \mathfrak{F} = \sigma(\mathfrak{C}), \quad P = \overline{\mu}$$

and for every $t \in T$

$$X_t(\omega) = \omega(t), \quad \omega \in \Omega.$$

For arbitrary $B \in \mathcal{B}(\mathbb{R})$, we have

$$X_t^{-1}(B) = \{\omega \in \mathbb{R}^T : X_t(\omega) \in B\}$$
$$= \{\omega \in \mathbb{R}^T : \omega(t) \in B\} = C_t(B) \in \mathfrak{F},$$

hence the X_t's are random variables, thus $(X_t : t \in T)$ is a stochastic process. Moreover, for arbitrary $B \in \mathcal{B}(\mathbb{R}^n)$, we have

$$P((X_{t_1}, \dots, X_{t_n}) \in B) = P(\{\omega : (X_{t_1}(\omega), \dots, X_{t_n}(\omega)) \in B\})$$
$$= P(\{\omega : (\omega(t_1), \dots, \omega(t_n)) \in B\})$$
$$= P(C_{t_1,\dots,t_n}(B)) = \overline{\mu}(C_{t_1,\dots,t_n}(B))$$
$$= \mu(C_{t_1,\dots,t_n}(B)) = \mu_{t_1,\dots,t_n}(B),$$

which shows that $\mu_{t_1,\dots,t_n}$ are the finite dimensional distributions of the process $(X_t : t \in T)$.

Observe that the X_t's above are defined exactly in the same way as the $\widetilde{X}_t$'s for the canonical process, and the proof of measurability of the X_t's is a repetition of the proof of measurability of the $\widetilde{X}_t$'s. The basic difference consists in the fact that when defining the canonical process, we had the measure $\mu_{\mathbb{X}}$ on $(\mathbb{R}^T, \sigma(\mathfrak{C}))$ at our disposal (the distribution of the "infinite dimensional random variable" $\mathbb{X}\colon \Omega \to \mathbb{R}^T$ defined by the initial process), while in the Kolmogorov theorem this measure had to be constructed.

In many aspects of probability theory, for instance, in laws of large numbers or limit theorems, we assume independence of random variables with given distributions. However, *a priori* it is not clear at all if it is possible. Kolmogorov's theorem shows that this is the case. Namely, let μ_n, $n = 1, 2, \dots$, be arbitrary distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For arbitrary $t_1, \dots, t_n \in \mathbb{N}$, define distributions $\mu_{t_1,\dots,t_n}$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ by the formula

$$\mu_{t_1,\dots,t_n} = \mu_{t_1} \otimes \dots \otimes \mu_{t_n}.$$

For arbitrary $B \in \mathcal{B}(\mathbb{R}^n)$, we have

$$\begin{aligned}\mu_{t_1,\dots,t_n,t}(B \times \mathbb{R}) &= \mu_{t_1} \otimes \dots \otimes \mu_{t_n} \otimes \mu_t(B \times \mathbb{R})\\ &= \mu_{t_1} \otimes \dots \otimes \mu_{t_n}(B)\mu_t(\mathbb{R})\\ &= \mu_{t_1} \otimes \dots \otimes \mu_{t_n}(B) = \mu_{t_1,\dots,t_n}(B)\end{aligned}$$

and for arbitrary $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$ and a permutation σ of the set $\{1, \dots, n\}$

$$\begin{aligned}&\mu_{t_{\sigma(1)},\dots,t_{\sigma(n)}}(B_{\sigma(1)} \times \dots \times B_{\sigma(n)})\\ =&\mu_{t_{\sigma(1)}} \otimes \dots \otimes \mu_{t_{\sigma(n)}}(B_{\sigma(1)} \times \dots \times B_{\sigma(n)})\\ =&\mu_{t_{\sigma(1)}}(B_{\sigma(1)}) \dots \mu_{t_{\sigma(n)}}(B_{\sigma(n)}) = \mu_{t_1}(B_1) \dots \mu_{t_n}(B_n)\\ =&\mu_{t_1} \otimes \dots \otimes \mu_{t_n}(B_1 \times \dots \times B_n) = \mu_{t_1,\dots,t_n}(B_1 \times \dots \times B_n),\end{aligned}$$

thus the family of distributions $\{\mu_{t_1,\dots,t_n} : t_1, \dots, t_n \in \mathbb{N},\ n = 1, 2, \dots\}$ satisfies the consistency conditions. By virtue of Kolmogorov's theorem, we infer that there exists a process (since $T = \mathbb{N}$, it is a sequence) $(X_n : n = 1, 2, \dots)$, such that its finite dimensional distributions are equal to $\mu_{t_1,\dots,t_n}$; in particular, the distribution of the random variable X_n equals μ_n.

For arbitrary $t_1, \dots, t_n \in \mathbb{N}$, and arbitrary $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned}&P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) = P((X_{t_1}, \dots, X_{t_n}) \in B_1 \times \dots \times B_n)\\ &\mu_{t_1,\dots,t_n}(B_1 \times \dots \times B_n) = \mu_{t_1} \otimes \dots \otimes \mu_{t_n}(B_1 \times \dots \times B_n)\\ =&\mu_{t_1}(B_1) \dots \mu_{t_n}(B_n) = P(X_{t_1} \in B_1) \dots P(X_{t_n} \in B_n),\end{aligned}$$

which proves independence of the random variables $X_{t_1}, \dots, X_{t_n}$, thus the X_n's are independent.

The Kolmogorov theorem, basic from the point of view of the existence of a process with given finite dimensional distributions, says nothing about possible properties of the samples of such a process. For example, if we wanted this process to be continuous with probability one, then, since the samples of this process in the construction above are *all* functions on T, it would mean that the set of continuous functions has measure one while, as we saw before, this set does not belong to $\sigma(\mathfrak{C})$, consequently, it can not have any measure! In the examples of two basic processes: Poisson's and Wiener's (Brownian motion) that follow, properties of the samples are a consequence of special constructions of these processes.

Now we are going to define an important class of stochastic processes, namely, processes with independent increments.

Definition. A stochastic process $(X_t : t \geqslant 0)$ is said to have *independent increments*, if for arbitrary $0 \leqslant t_0 < t_1 < \cdots < t_n$ the random variables

$$X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}} \quad \text{(increments of the process)}$$

are independent.

We finish our considerations with definitions of two extremely important stochastic processes.

Definition. A stochastic process $(N_t : t \geqslant 0)$ (traditional notation) is a *Poisson process*, if

(1) $N_0 = 0$ with probability one,
(2) the process has independent increments,
(3) for $s < t$ the increments of the process have Poisson's distribution with parameter $\lambda(t-s)$:

$$P(N_t - N_s = n) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^n}{n!}, \quad n = 0, 1, \ldots,$$

(4) the samples of the process are with probability one non-decreasing functions.

If in the definition above we require only the first three conditions, then the existence of a Poisson process follows from Kolmogorov's theorem: using the independence of the increments we can find the finite dimensional distributions, and then show that they satisfy the consistency conditions. However, such an approach does not give samples non-decreasing with probability one since the set of non-decreasing functions does not belong to the σ-field $\sigma(\mathfrak{C})$. To obtain a Poisson process satisfying *all* conditions of the definition above a special construction is employed.

Definition. A stochastic process $(W_t : t \geqslant 0)$ (traditional notation) is a *Wiener process* or *Brownian motion*, if

(1) $W_0 = 0$ with probability one,
(2) the process has independent increments,

(3) for $s < t$ the increments of the process have normal distribution $N(0, t-s)$ with density:

$$f(x) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{x^2}{2(t-s)}},$$

(4) the samples of the process are with probability one continuous functions.

A comment made for a Poisson process can be repeated here almost word for word: if in the definition above we require only the first three conditions, then the existence of a Wiener process follows from Kolmogorov's theorem: using the independence of the increments we can find the finite dimensional distributions, and then show that they satisfy the consistency conditions. However, such an approach does not give samples continuous with probability one since the set of continuous functions does not belong to the σ-field $\sigma(\mathfrak{C})$. To obtain a Wiener process satisfying *all* conditions of the definition above a special construction is employed.

A surprising property of a Wiener process is presented in the theorem below which ends our considerations.

Theorem 17. *The samples of a Wiener process are with probability one functions not differentiable at any point (despite the fact that they are continuous functions!).*

Problem 16. Find the distributions of the random variables in a Poisson process and a Wiener process.

Solution. For Poisson and Wiener processes we have

$$N_t = N_t - N_0, \qquad W_t = W_t - W_0,$$

thus the distributions of the random variables N_t and W_t are the same as the distributions of the increments $N_t - N_0$ and $W_t - W_0$, respectively, thus N_t has the Poisson distribution

$$P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \ldots$$

and W_t has the normal distribution $N(0, t)$ with the density

$$f(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

References

[1] P. Billingsley, *Probability and Measure*, Wiley & Sons, New York, 1979.