THEORY OF STOCHASTIC PROCESSES (PRELIMINARIES AND REPETITION)

ANDRZEJ ŁUCZAK

1. Prologue

The theory of stochastic processes is a more advanced part of Probability Theory. It deals with arbitrary families of random variables which often describe concrete situations. For example, if we want to describe the 'dynamics' of temperature in Istanbul on the first of June, then its value at some specific time t (e.g. t=12:30) is a value of a random variable X_t, and we can obtain information about the probability that this temperature equals e.g. $25°C$, $P(X_t = 25)$, on the basis of observations made during, say, last hundred years. In this way, we obtain a family of random variables $X_t : t \in [0,1]$ where 0 stands for the hour 0:00 of June 1, and 1 stands for the hour 24:00 of the same day. Observe that in this case we can say almost nothing about the probability space $(\Omega, \mathfrak{F}, P)$ on which these X_t's are defined since we don't know all factors which affect the temperature! On the other hand, we can estimate the distribution of X_t in an obvious way: if in the last hundred years the temperature between 25 and 30 degrees at time t happened in 60 years then we can say that $P(25 \leqslant X_t \leqslant 30) \approx \frac{60}{100} = \frac{3}{5}$. Observe also that if we fix ω, then the function $[0,1] \ni t \mapsto X_t(\omega)$ represents the temperature during the whole day June 1. The family $(X_t : t \in [0,1])$ is an example of a stochastic process. Thus we shall consider families $(X_t : t \in T)$ of random variables indexed by an arbitrary set T (but mainly T being an interval in $\mathbb{R}$), and investigate their various properties. It is assumed that the participants of the course have some knowledge of probability theory, in particular, the notions of a probability space, random variable and its distribution, random vector, expectation and variance of a random variable, though some of these notions will be reminded. Some knowledge of measure and integral theory would also be helpful.

2. Probability theory

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 Ω is an arbitrary non-empty set, $\mathfrak{F}$ is a σ-field of subsets of Ω, probability P is a normalised measure on $\mathfrak{F}$ which means that

(i) $P\colon \mathfrak{F} \to [0,1]$,
(ii) $P(\varnothing) = 0, \quad P(\Omega) = 1$,
(iii) for arbitrary pairwise disjoint $A_1, A_2, \ldots \in \mathfrak{F}$, we have

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

The probability P has 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
$$0 \leqslant P\Big(\big(\bigcap_{n=1}^{\infty} A_n\big)'\Big) = P\Big(\bigcup_{n=1}^{\infty} A_n'\Big) \leqslant \sum_{n=1}^{\infty} P(A_n') = 0,$$
hence
$$P\Big(\big(\bigcap_{n=1}^{\infty} A_n\big)'\Big) = 0,$$
i.e.
$$P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = 1,$$
which proves the lemma. □

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 $\varepsilon > 0$

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

A relation between these modes of convergence is as follows.

Theorem 1. *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.

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

$\big($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 $\varepsilon = \frac{1}{k}\big)$. 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

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

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 2. *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. 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

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

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.* □

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 $P(\{\omega_i\})>0$, $\sum\limits_i P(\{\omega_i\})=1$. 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 shall (and 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 $P(A_n)\to1$ 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

$$1=\lim_{n\to\infty}P(A_n)\leqslant1-P(\{\omega_{i_0}\})<1,$$

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 smallest 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

$$\int_{\Omega} \frac{Z}{1+Z}\,dP = \int_{\{Z<\varepsilon\}} \frac{Z}{1+Z}\,dP + \int_{\{Z\geqslant\varepsilon\}} \frac{Z}{1+Z}\,dP$$

(2)

$$\geqslant \int_{\{Z\geqslant\varepsilon\}} \frac{Z}{1+Z}\,dP \geqslant \int_{\{Z\geqslant\varepsilon\}} \frac{\varepsilon}{1+\varepsilon}\,dP = \frac{\varepsilon}{1+\varepsilon}P(Z \geqslant \varepsilon)$$

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.

Let $X\colon \Omega \to \mathbb{R}$ be a random variable. Its *distribution* μ_X is defined as a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by the formula

$$\mu_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

Accordingly, if $(X_1, \ldots, X_k)$ is a random vector, its *distribution* ($\equiv$ *joint distribution*) $\mu_{X_1,\ldots,X_k}$ is defined as a probability measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ by the formula

$$\begin{aligned}\mu_{X_1,\ldots,X_k}(B) &= P((X_1, \ldots, X_k)^{-1}(B)) \\ &= P(\{\omega : (X_1(\omega), \ldots, X_k(\omega)) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}^k).\end{aligned}$$

In the first equality above, $(X_1, \ldots, X_k)$ is treated as a map $(X_1, \ldots, X_k)\colon \Omega \to \mathbb{R}^k$ defined as

$$(X_1, \ldots, X_k)(\omega) = (X_1(\omega), \ldots, X_k(\omega))$$

Observe that having $\mu_{X_1,\ldots,X_k}$ we can find the distributions of each random variable X_i, $i = 1, \ldots, k$ ($\equiv$ *marginal distributions*), by the formula

$$\mu_{X_i}(B) = \mu_{X_1,\ldots,X_k}(\mathbb{R} \times \cdots \times \underset{i}{B} \times \cdots \times \mathbb{R}), \quad B \in \mathcal{B}(\mathbb{R}).$$

The inverse procedure is, in general, impossible, i.e. the knowledge of all marginal distributions μ_{X_i}, $i = 1, \ldots, k$, does not imply the knowledge of the joint distribution $\mu_{X_1,\ldots,X_k}$. However, if the random variables are *independent* then we do have the formula

(3) $$\mu_{X_1,\ldots,X_k} = \mu_{X_1} \otimes \cdots \otimes \mu_{X_k},$$

meaning that in this case the joint distribution is a product of the marginal distributions. (As a matter of fact, the formula (3) is *equivalent* to the independence of the random variables $X_1, \ldots, X_k$.)

Let X and Y be random variables, and let $g\colon \mathbb{R}^2 \to \mathbb{R}$ be a Borel function. Put $Z(\omega) = g(X(\omega), Y(\omega))$. Z is a random variable. For its distribution function F_Z we have

$$F_Z(t) = P(\{\omega : Z(\omega) \leqslant t\}) = P(\{\omega : g(X(\omega), Y(\omega)) \leqslant t\}).$$

Denote

$$B = \{(x, y) : g(x, y) \leqslant t\}.$$

Then for the map $(X, Y)\colon \Omega \to \mathbb{R}^2$ defined as

$$(X, Y)(\omega) = (X(\omega), Y(\omega))$$

we have

$$(X, Y)^{-1}(B) = \{\omega : (X(\omega), Y(\omega)) \in B\} = \{\omega : g(X(\omega), Y(\omega)) \leqslant t\}.$$

Hence

$$\begin{aligned} F_Z(x) &= P(\{\omega : g(X(\omega), Y(\omega)) \leqslant t\}) = P((X, Y)^{-1}(B)) \\ &= \mu_{X,Y}(B) = \mu_{X,Y}(\{(x, y) : g(x, y) \leqslant t\}), \end{aligned}$$

thus knowledge of the joint distribution $\mu_{X,Y}$ allows us to find the distribution of a function of the random variables X and Y.

Problem 4. Alice and Bob agree to have an appointment at a cafe between 17 and 18. The moments of their arrivals to the cafe are independent random variables with uniform distributions on the interval $[0, 1]$, where 0 stands for the hour 17 and 1 for the hour 18. It is agreed that each one waits 15 minutes ($= \frac{1}{4}$) for the other. Find the probability that they will meet.

Solution. Let X be the moment of arrival of Alice, and Y — the moment of arrival of Bob. The density functions of X and Y are the same and equal

$$f_X(x) = f_Y(x) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Since X and Y are independent, the density $f_{X,Y}$ of the random vector (X, Y) equals a product of the densities f_X and f_Y (more precisely, $f_{X,Y}(x, y) = f_X(x) f_Y(y)$), thus

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{for } 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

The probability that they meet equals $P\Big(|X - Y| < \frac{1}{4}\Big)$, hence

$$\begin{aligned} P\Big(|X - Y| < \frac{1}{4}\Big) &= \iint\limits_{\{(x,y)\in[0,1]\times[0,1]:|x-y|<\frac{1}{4}\}} f_{X,Y}(x, y)\, dxdy \\ &= \iint\limits_{\{(x,y)\in[0,1]\times[0,1]:|x-y|<\frac{1}{4}\}} dxdy \\ &= \text{Area}\Big(\Big\{(x, y) \in [0, 1] \times [0, 1] : |x - y| < \frac{1}{4}\Big\}\Big) = \frac{7}{16}. \end{aligned}$$

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

Lemma 2. *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 $f^{-1}(B) \in \mathfrak{M}$. □

Observe that this lemma is exploited in the definition of a random variable as a *measurable* function $X\colon \Omega \to \mathbb{R}$. Four equivalent definitions of measurability are as follows:

(i) for each $a \in \mathbb{R}$, $\{\omega : X(\omega) < a\} = X^{-1}((-\infty, a)) \in \mathfrak{F}$,
(ii) for each $a \in \mathbb{R}$, $\{\omega : X(\omega) \leqslant a\} = X^{-1}((-\infty, a]) \in \mathfrak{F}$,
(iii) for each $a \in \mathbb{R}$, $\{\omega : X(\omega) > a\} = X^{-1}((a, \infty)) \in \mathfrak{F}$,
(iv) for each $a \in \mathbb{R}$, $\{\omega : X(\omega) \geqslant a\} = X^{-1}([a, \infty)) \in \mathfrak{F}$.

However, there is a fifth equivalent (though more general) definition

(v) for each $B \in \mathcal{B}(\mathbb{R})$, $\{\omega : X(\omega) \in B\} = X^{-1}(B) \in \mathfrak{F}$

The equivalence of all definitions follows from the lemma. Namely, we have

$$\begin{aligned}\sigma(\{(-\infty, a) : a \in \mathbb{R}\}) &= \sigma(\{(-\infty, a] : a \in \mathbb{R}\})\\ =\sigma(\{(a, \infty) : a \in \mathbb{R}\}) &= \sigma(\{[a, \infty) : a \in \mathbb{R}\}) = \mathcal{B}(\mathbb{R}).\end{aligned}$$

For a random variable X, we define its *expectation* as

$$\mathbb{E}X \overset{\text{def}}{=} \int_\Omega X\,dP \overset{\text{basic formula}}{=} \int_{\mathbb{R}} x\,\mu_X(dx)$$

if X is *integrable*, i.e. $\int_\Omega |X|\,dP < +\infty$.

It follows that the expectation satisfies the conditions (as an integral)

(1) If a random variable X is constant with probability one, $P(X = c) = 1$ for some $c \in \mathbb{R}$, then $\mathbb{E}X = c$,
(2) for each $a \in \mathbb{R}$, $\mathbb{E}(aX) = a\mathbb{E}X$,
(3) for random variables X, Y, $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$,
(4) for random variables X, Y such that $X \leqslant Y$ with probability one, $\mathbb{E}X \leqslant \mathbb{E}Y$,

under the assumption that all expectations above exist.

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*

$$|\mathbb{E}XY| \leqslant \big(\mathbb{E}X^2\big)^{1/2}\big(\mathbb{E}Y^2\big)^{1/2}. \tag{4}$$

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 (4) follows.

If $\mathbb{E}X^2 = 0$, then $X = 0$ (with probability one) and we have zero on both sides of the inequality (4). □

For a random variable X such that $\mathbb{E}X$ exists, we define its *variance* as

$$\mathbb{D}^2X \overset{\text{def}}{=} \mathbb{E}(X - \mathbb{E}X)^2 \overset{\text{useful formula}}{=} \mathbb{E}X^2 - (\mathbb{E}X)^2$$

It follows that the variance satisfies the conditions:

(1) $\mathbb{D}^2X \geqslant 0$; moreover, $\mathbb{D}^2X = 0$ if and only if X is constant with probability one, $P(X = c) = 1$ for some $c \in \mathbb{R}$,
(2) for each $a \in \mathbb{R}$, $\mathbb{D}^2(aX) = a^2\mathbb{D}^2X$,
(3) for random variables X, Y,

$$\mathbb{D}^2(X + Y) = \mathbb{D}^2X + \mathbb{D}^2Y + 2(\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y),$$

under the assumption that all variances above exist.

Random variables X and Y are said to be *uncorrelated* if

$$\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y,$$

thus

$$\mathbb{D}^2(X + Y) = \mathbb{D}^2X + \mathbb{D}^2Y$$

if and only if X and Y are uncorrelated.

It is proven that if X and Y are independent, then they are uncorrelated (of course, under the assumption that $\mathbb{E}X$ and $\mathbb{E}Y$ exist). Hence, for independent random variables, variance of their sum equals a sum of their variances.

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. □

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

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

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

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

□

3. Problems to solve

Problem 5. The times of arrival of buses A and B are independent random variables having exponential distributions with parameters α and β, respectively. Find the distribution of the time of arrival of the first bus.

Problem 6. The setting is the same as in Problem 4 (Alice meets Bob) with the only difference that Alice does not wait, i.e. she comes to the cafe and if Bob is not present she leaves. Find the probability that they will meet.

Problem 7. The random vector (X, Y) has density

$$f(x, y) = \begin{cases} 1, & \text{for } 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Find the distribution of the random variable $Z = X + Y$.

Problem 8. Let X_n be random variables with the distribution functions

$$F_n(x) = \begin{cases} 0, & \text{for } x < 0 \\ nx, & \text{for } 0 \leqslant x < \frac{1}{n} \\ 1, & \text{for } x \geqslant \frac{1}{n} \end{cases}.$$

Prove that $X_n \to 0$ in probability.

Problem 9. Let $X_n \to X$ in probability, and $X_n \to X'$ in probability. Show that then $P(X = X') = 1$.

Hint. First prove that for random variables X and Y, and arbitrary $\varepsilon > 0$ the following inequality holds

$$P(|X+Y| \geqslant \varepsilon) \leqslant P\Big(|X| \geqslant \frac{\varepsilon}{2}\Big) + P\Big(|Y| \geqslant \frac{\varepsilon}{2}\Big).$$

Problem 10. Let X and Y be random variables, and define

$$\rho(X,Y) = \inf\{\varepsilon > 0 : P(|X-Y| \geqslant \varepsilon) \geqslant \varepsilon\}.$$

Show that $X_n \to X$ in probability if and only if $\rho(X_n, X) \to 0$.

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, ŁÓDŹ UNIVERSITY, ŁÓDŹ, POLAND
Email address: andrzej.luczak@wmii.uni.lodz.pl