Lecture 3: Method of One Probability Space

Historical remarks. Convergence of random variables in distribution, in probability and almost sure. Skorokhod representation theorem. Applications to limit theorems for stochastic processes and their superpositions.

1. Historical remarks

Anatoliy V. Skorokhod, 1930-2011 (1956, 1957)

2. Convergence of random variables in distribution, in probability and with probability 1

2.1 Convergence in distribution. Let $\xi_\varepsilon, \varepsilon \geq 0$ be a family of real-valued random variables depending on a parameter $\varepsilon \geq 0$. We denote by $F_\varepsilon(x) = P\{\xi_\varepsilon \leq x\}, x \in \mathbb{R}_1$, the distribution function of the random variable $\xi_\varepsilon$.

The concept of convergence in distribution plays a central role in probability theory and its applications. It is enough to recall that the fundamental limit theorems such as the weak law of large numbers and the central limit
theorem are, actually, statements about convergence in distribution for random variables.

We say that random variables $\xi_\varepsilon$ converge in distribution to a random variable $\xi_0$ as $\varepsilon \to 0$ if $F_\varepsilon(x) \to F_0(x)$ as $\varepsilon \to 0$ for all points $x$ which are points of continuity for the limiting distribution function.

This convergence is denoted as $\xi_\varepsilon \overset{d}{\to} \xi_0$ as $\varepsilon \to 0$.

Convergence in distribution of random variables is, in fact, convergence of their distributions. That is why we can equivalently use term weak convergence of distribution functions and denote this as $F_\varepsilon(\cdot) \Rightarrow F_0(\cdot)$ as $\varepsilon \to 0$.

Also, symbol $\xi_\varepsilon \Rightarrow \xi_0$ as $\varepsilon \to 0$ is used and convergence in distribution can be referred as weak convergence of random variables.

It is also useful to note that random variables can be indexed in different ways. For example, a sequence of random variables $\xi_n$ that depends on the index $n = 1, 2, \ldots$ can be an object of consideration. The notation of convergence in distribution is modified in an obvious way as, $\xi_n \overset{d}{\to} \xi_0$ as $n \to \infty$.

Recall that the random variable $\xi_\varepsilon$ is defined on some probability space $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$. This means that $\xi_\varepsilon = \xi_\varepsilon(\omega)$ is a mapping $\Omega_\varepsilon \to \mathbb{R}_1$, measurable in the sense that $\{\omega : \xi_\varepsilon(\omega) \in A\} \in \mathcal{F}_\varepsilon$, for $A \in \mathcal{B}_1$.

Since, convergence in distribution (weak convergence) is, in fact, convergence of distribution functions, the probability space can be different for different $\varepsilon$.

-------------------------------

A typical example is the law of small numbers. In this model, $\nu_{p,n}$ is the number of successful trials in series of $n$ Bernoulli trials. As is known, $\nu_{p,n} \overset{d}{\to} \nu_\lambda$ as $n \to \infty, p \to 0$ such that $np \to \lambda > 0$, where $\nu_\lambda$ is a random variable, which has the Poisson distribution with parameter $\lambda$.

-------------------------------
The definition of convergence in distribution gives rise to the following question. Why is the point-wise convergence of distribution functions required only in points of continuity of the corresponding limiting distribution function?

The following standard example explains why points of discontinuity should be excluded from the set of point-wise convergence. Let us consider a sequence of numbers \( a_n, n = 1, 2, \ldots \), such that \( a_n \to a_0 \) as \( n \to \infty \), where \( a_0 \) is a finite real constant. The constant \( a_n \) can be considered as a random variable. It is natural to expect that convergence in distribution, \( a_n \overset{d}{\to} a_0 \) as \( n \to \infty \), would be equivalent to the usual convergence \( a_n \to a_0 \) as \( n \to \infty \). The distribution function of \( a_n \), considered as a random variable, is the indicator function \( F_n(x) = \chi_{[a_n, \infty)}(x) \). It is easy to check that \( a_n \to a_0 \) as \( n \to \infty \) if and only if \( F_n(x) \to F_0(x) \) as \( n \to \infty \) for all \( x \neq a_0 \), i.e., \( F_n(x) \to 0 \) as \( n \to \infty \) for \( x < a_0 \) and \( F_n(x) \to 1 \) as \( n \to \infty \) for \( x > a_0 \). Note that convergence of \( F_n(x) \) to \( F_0(x) \) in the point \( a_0 \) (which is the only point of discontinuity of the limiting distribution function \( F_0(x) \)) is not required to provide convergence of \( a_n \) to \( a_0 \). If, for example, \( a_n \) is a decreasing sequence, then \( F_n(a_0) = 0 \) for all \( n = 1, 2, \ldots \), but \( F_0(a_0) = 1 \). Therefore, convergence of \( a_n \) to \( a_0 \) does not imply convergence of \( F_n(a_0) \) to \( F_0(a_0) \).

It should also be noted that weak convergence of random variables is equivalent to the usual point-wise convergence of their distribution functions in all points \( x \in \mathbb{R}_1 \), if the limiting distribution function is continuous.

The definition of convergence in distribution (weak convergence) given above can easily be extended from random variables to random vectors, i.e., random variables that take values in the space \( \mathbb{R}_m \). In this case, one-dimensional distribution functions should be replaced by the corresponding multi-dimensional distribution functions. One can use the definition of convergence in distribution as point-wise convergence in points of continuity for the corresponding limiting multi-dimensional distribution function.
However, if the random variables take values in a metric space, the definition of convergence in distribution should be modified. It can happen that direct analogues of the distribution functions do not exist. In this case, the definition can be given with the use of convergence of values of the probability measures generated by random variables. Convergence should be required for values of these measures on sets of continuity for the corresponding limiting measure.

2.2 Convergence in probability and convergence with probability 1. In this case, it is assumed that random variables $\xi, \varepsilon \geq 0$ are defined on the same probability space $\langle \Omega, \mathcal{F}, P \rangle$.

Random variables $\xi \varepsilon$ converge in probability to $\xi_0$ as $\varepsilon \to 0$ if

$$P\{ |\xi \varepsilon - \xi_0| > \delta \} \to 0 \text{ as } \varepsilon \to 0, \text{ for } \delta > 0.$$

Symbol $\xi \varepsilon \overset{P}{\to} \xi_0$ as $\varepsilon \to 0$ is used to denote convergence in probability.

**Lemma 1.** If $\xi \varepsilon \overset{P}{\to} \xi_0$ as $\varepsilon \to 0$, then $\xi \varepsilon \overset{d}{\to} \xi_0$ as $\varepsilon \to 0$.

Random variables $\xi \varepsilon$ converge with probability 1 (almost sure) to $\xi_0$ as $\varepsilon \to 0$, if there exists a random event $A_0 \in \mathcal{F}$ such that $\xi \varepsilon(\omega) \to \xi_0(\omega)$ as $\varepsilon \to 0$ for every $\omega \in A_0$, and $P(A_0) = 1$.

Symbol, $\xi \varepsilon \overset{P_1}{\to} \xi_0$ as $\varepsilon \to 0$, or, $\xi \varepsilon \overset{a.s.}{\to} \xi_0$ as $\varepsilon \to 0$, is used to denote this convergence.

**Lemma 2.** If $\xi \varepsilon \overset{P_1}{\to} \xi_0$ as $\varepsilon \to 0$, then $\xi \varepsilon \overset{P}{\to} \xi_0$ as $\varepsilon \to 0$.

2.3 Skorokhod representation theorem - I. Let a real-valued random variable $\xi \varepsilon$ is defined on some probability space $\langle \Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon \rangle$, for $\varepsilon \geq 0$.

The following lemma is an elementary variant of the famous Skorokhod representation theorem.
Lemma 3. If $\xi_\varepsilon \xrightarrow{d} \xi_0$ as $\varepsilon \to 0$, then it is possible to construct a probability space $\langle \Omega, \mathcal{F}, P \rangle$ and random variables $\tilde{\xi}_\varepsilon, \varepsilon \geq 0$, defined on this space, such that (a) the random variables $\tilde{\xi}_\varepsilon \xrightarrow{d} \xi_\varepsilon$ (have the same distribution), for every $\varepsilon \geq 0$, and (b) the random variables $\tilde{\xi}_\varepsilon \xrightarrow{P} \tilde{\xi}_0$ as $\varepsilon \to 0$.

Let, for example, a random variable $\xi_\varepsilon$ have the exponential distribution with parameter $\lambda_\varepsilon > 0$, i.e., $P\{\xi_\varepsilon \leq x\} = F_\varepsilon(x) = 1 - \exp\{-\lambda_\varepsilon x\}$ for $x \geq 0$. Let also $\lambda_\varepsilon \to \lambda_0 > 0$ as $\varepsilon \to 0$. In this case, it is obvious that $\xi_\varepsilon \xrightarrow{d} \xi_0$ as $\varepsilon \to 0$.

Let us consider the function $F^{-1}_\varepsilon(y) = (-1/\lambda_\varepsilon) \ln (1 - y)$, which is the inverse of the exponential distribution function $F_\varepsilon(x)$ introduced above. Let also $\rho$ be a random variable uniformly distributed on $[0, 1]$. Let us now consider the random variables $\tilde{\xi}_\varepsilon = -(1/\lambda_\varepsilon) \ln (1 - \rho)$. It is easy to check that the random variable $\tilde{\xi}_\varepsilon$ has the exponential distribution with parameter $\lambda_\varepsilon > 0$. So, for every $\varepsilon \geq 0$, the random variables $\xi_\varepsilon$ and $\tilde{\xi}_\varepsilon$ have the same distribution.

Also, the random variables $\tilde{\xi}_\varepsilon \xrightarrow{P} \tilde{\xi}_0$ as $\varepsilon \to 0$. This is so, because $(-1/\lambda_\varepsilon) \ln (1 - y) \to (-1/\lambda_0) \ln (1 - y)$ as $\varepsilon \to 0$ for every $y \in [0, 1)$ and $P\{\rho \in [0, 1]\} = 1$.

In the case of real-valued random variables, the above construction can be realised in a similar way.

Let $F_\varepsilon(x)$ be a distribution function of a random variable $\xi_\varepsilon$, and $F^{-1}_\varepsilon(y) = \inf \{x : F_\varepsilon(x) > y\}$ for $y \in [0, 1]$.

Let also $\rho$ be a random variable uniformly distributed on $[0, 1]$. For example, we can use the probability space with the space of outcomes $\Omega = [0, 1]$, the Borel $\sigma$-algebra of random events $\mathcal{B}_{[0,1]}$ as $\sigma$-algebra $\mathcal{F}$, and the Lebesgue measure $m(A)$ as the corresponding probability measure $P(A)$. Then we can define $\rho(\omega) = \omega$.

As is known, the random variable $\tilde{\xi}_\varepsilon = F^{-1}_\varepsilon(\rho)$ has the distribution function $F_\varepsilon(x)$.

It is not difficult to show that the point-wise convergence of $F_\varepsilon(x)$ to $F_0(x)$ as $\varepsilon \to 0$ (in all points of continuity of the limiting distribution func-
tion $F_0(x))$ implies that their inverses, $F^{-1}_\varepsilon(y)$, point-wise converge to $F^{-1}_0(y)$ as $\varepsilon \to 0$ (in all points of continuity of the limiting function $F^{-1}_0(y)$). Since this function is monotone, it has at most a countable set of discontinuity points. So, the set $C'_0$ of continuity points of this function has Lebesgue measure 1, i.e., $P\{\rho \in C'_0\} = 1$. Obviously, $\{\rho \in C'_0\} \subseteq \{\lim_{\varepsilon \to 0} \tilde{\xi}_\varepsilon = \xi_0\}$. Hence, $P\{\lim_{\varepsilon \to 0} \tilde{\xi}_\varepsilon = \xi_0\} = 1$.

The Skorokhod representation theorem generalises this result to random variables that take values in a complete, separable metric space. This generalisation is not trivial. The theorem allows to simplify proofs of some important limit theorems.

2.4 Convergence in distribution of transformed random variables. Let real-valued random variables $\xi_\varepsilon \xrightarrow{d} \xi_0$ as $\varepsilon \to 0$. Let also $f(x)$ be a measurable real-valued function (the inverse image of any Borel set is a Borel set) defined on the real line. In this case, $f(\xi_\varepsilon)$ is also a real-valued random variable. The question arises whether the random variables $f(\xi_\varepsilon) \xrightarrow{d} f(\xi_0)$ as $\varepsilon \to 0$?

It can be shown that this is true for all measurable functions $f(x)$ that are a.s. continuous with respect to the distribution of the limiting random variable $\xi_0$. This statement can be elegantly proved by using the Skorokhod representation theorem.

Indeed, let $\tilde{\xi}_\varepsilon, \varepsilon \geq 0$ be random variables constructed according Lemma 3, i.e, such that: (a) $\tilde{\xi}_\varepsilon \equiv \xi_\varepsilon$, for $\varepsilon \geq 0$ and (b) $\tilde{\xi}_\varepsilon \xrightarrow{p_1} \tilde{\xi}_0$ as $\varepsilon \to 0$.

Let, also, $C_f \in B_X$ be the set such that $f(x)$ is continuous at points $x \in C_f$ and $P\{\xi_0 \in C_f\} = 1$.

Denote $A = \{\omega : \xi_0(\omega) \in C_f\}$. Obviously $P\{\tilde{\xi}_0 \in C_f\} = P\{\omega : \tilde{\xi}_0(\omega) \in C_f\} = 1$.

The random variables $\tilde{\xi}_\varepsilon \xrightarrow{p_1} \tilde{\xi}_0$ as $\varepsilon \to 0$. This means that $P\{\omega : \tilde{\xi}_\varepsilon(\omega) \to \tilde{\xi}_0(\omega)\} = 1$. We denote $B = \{\omega : \tilde{\xi}_\varepsilon(\omega) \to \xi_0(\omega)\}$.

Obviously, $f(\tilde{\xi}_\varepsilon(\omega)) \to f(\xi_0(\omega))$ as $\varepsilon \to 0$, for $\omega \in A \cap B$. 

6
Since, $P(A) = 1$ and $P(B) = 1$, then $P(A \cap B) = 1$. Therefore, $f(\xi_\varepsilon) \xrightarrow{P^1} f(\xi_0)$ as $\varepsilon \to 0$. Since a.s. convergence implies convergence in distribution, the random variables $f(\xi_\varepsilon) \xrightarrow{d} f(\xi_0)$ as $\varepsilon \to 0$.

But, the random variables $f(\xi_\varepsilon) \xrightarrow{d} f(\xi_0)$, since random variables $\tilde{\xi}_\varepsilon \xrightarrow{d} \xi_\varepsilon$. Therefore, the random variables $f(\xi_\varepsilon) \xrightarrow{d} f(\xi_0)$ as $\varepsilon \to 0$.

This statement plays a very important role in the general theory of weak convergence of random variables in metric spaces. In the case of functional metric spaces, random variables that take values in such spaces are, actually, stochastic processes, while the corresponding transformed random variables are functionals defined on trajectories of these processes.

3. Weak convergence in metric spaces and Skorokhod representation theorem

3.1 Polish spaces. Let $X$ be a metric space with a distance $d(x, y)$. The space $X$ is complete if, for any fundamental sequence of points $x_n \in X$, i.e., a sequence such that $d(x_n, x_m) \to 0$ as $n, m \to \infty$, there exists a point $x \in X$ such that $d(x_n, x) \to 0$ as $n \to \infty$.

The space $X$ is separable if there exists a countable subset $Y = \{y_1, y_2, \ldots\} \subseteq X$ such that $\min_{k \leq n} d(y_k, x) \to 0$ as $n \to \infty$ for any point $x \in X$.

The term Polish space is used to indicate that $X$ is a complete separable metric space. Below, $X$ is always a Polish space.

Let $\mathcal{B}_X$ be the Borel $\sigma$-algebra of subsets of $X$ (the minimal $\sigma$-algebra containing any ball $B_r(x) = \{y \in X: d(x, y) \leq r\}$ in the space $X$).

The space $\mathbb{R}_m$ is a particular example of a Polish space. In this case,

$$d(\bar{x}, \bar{y}) = |\bar{x} - \bar{y}| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2},$$

for points $\bar{x} = (x_1, \ldots, x_m), \bar{y} = (y_1, \ldots, y_m) \in \mathbb{R}_m$.

The second important example is the functional space $C^m_{[0, T]}$ of continuous functions $\bar{x}(\cdot) = (\bar{x}(t), t \in [0, T])$ defined on interval $[0, T]$ and taking values
in space $\mathbb{R}_m$, with uniform distance,
\[
d_{U,T}(\bar{x}(\cdot), \bar{y}(\cdot)) = \sup_{t \in [0, T]} |\bar{x}(t) - \bar{y}(t)|,
\]
for points (functions) $\bar{x}(\cdot) = \langle \bar{x}(t), t \in [0, T] \rangle$, $\bar{y}(\cdot) = \langle \bar{y}(t), t \in [0, T] \rangle \in C^m_{[0,T]}$.

The third example is the functional space $D^m_{[0,T]}$ of càdlàg functions (continuous from the right and possessing limits from the left in every point of interval $[0,T]$) taking values in space $\mathbb{R}_m$, with some appropriate metrics, for example, those given by Billingsley (1968),
\[
d_{J,T}(\bar{x}(\cdot), \bar{y}(\cdot)) = \inf_{\lambda(\cdot) \in \Lambda_{[0,T]}} (\|\lambda\|_T + d_{U,T}(\bar{x}(\lambda(\cdot)), \bar{y}(\cdot))).
\]
where $\Lambda_{[0,T]}$ is the space of all continuous strictly monotonic mapping $\lambda(t)$ of the interval $(0,T]$ onto itself such that $\lambda(0) = 0$, $\lambda(T) = T$, and
\[
\|\lambda\|_T = \sup_{0 \leq t, s \leq T, t \neq s} \left| \ln \frac{\lambda(t) - \lambda(s)}{t - s} \right| \leq \infty.
\]

3.2 Convergence in distribution. Let $\xi_\varepsilon, \varepsilon \geq 0$ be random variables taking values in $\mathbb{X}$. We denote by $F_\varepsilon(A) = P\{\xi_\varepsilon \in A\}$, $A \in \mathcal{B}_\mathbb{X}$, the distribution of the random variable $\xi_\varepsilon$.

Let $\partial A$ denote the boundary of the set $A$, i.e., the set of points $x$ such that every ball $B_r(x)$, with centre in $x$ and a radius $r > 0$, has non-empty intersections with both sets $A$ and $\bar{A}$.

If $F_0(\partial A) = 0$, then $A$ is called a set of continuity for the distribution $F_0$. The family of such sets, $\mathcal{B}(F_0) \subseteq \mathcal{B}_\mathbb{X}$, is a $\sigma$-algebra of subsets of $\mathcal{B}_\mathbb{X}$.

Let $\xi_\varepsilon$ be a random variable defined on some probability spaces $\langle \Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon \rangle$ and taking values in some Polish space $\mathbb{X}$, for $\varepsilon \geq 0$.

Random variables $\xi_\varepsilon$ converge in distribution to $\xi_0$ as $\varepsilon \to 0$ if $F_\varepsilon(A) \to F_0(A)$ as $\varepsilon \to 0$ for all sets $A \in \mathcal{B}(F_0)$.
As above, symbols $\xi_\varepsilon \xrightarrow{d} \xi_0$ as $\varepsilon \to 0$ and $\xi_\varepsilon \xrightarrow{P_1} \xi_0$ as $\varepsilon \to 0$ are used to denote, respectively, convergence in distribution and convergence with probability 1.

Analogues of Lemmas 1 and 2 take place.

The following, so called the “portmanteau” theorem plays the basic role in the theory.

**Theorem 1.** The following tree propositions are equivalent:

(i) $\xi_\varepsilon \xrightarrow{d} \xi_0$ as $\varepsilon \to 0$.

(ii) $E f(\xi_\varepsilon) \to E f(\xi_0)$ as $\varepsilon \to 0$, for any measurable bounded real-valued function, a.s. continuous with respect to the distribution of the limiting random variable $\xi_0$.

(iii) $f(\xi_\varepsilon) \xrightarrow{d} f(\xi_0)$ as $\varepsilon \to 0$, for any measurable real-valued function, a.s. continuous with respect to the distribution of the limiting random variable $\xi_0$.

### 3.3 Skorokhod representation theorem

The following Skorokhod representation theorem let one give an elegant proof of Theorem 1.

**Theorem 2.** If $\xi_\varepsilon \xrightarrow{d} \xi_0$ as $\varepsilon \to 0$, then it is possible to construct a probability space $\langle \Omega, \mathcal{F}, P \rangle$ and random variables $\tilde{\xi}_\varepsilon, \varepsilon \geq 0$, defined on this space, such that (a) the random variables $\tilde{\xi}_\varepsilon \xrightarrow{d} \xi_\varepsilon$ (i.e. $P\{\tilde{\xi}_\varepsilon \in A\} = P\{\xi_\varepsilon \in A\}$, for $A \in \mathcal{B}_\mathbb{X}$), for every $\varepsilon \geq 0$, and (b) the random variables $\tilde{\xi}_\varepsilon \xrightarrow{P_1} \tilde{\xi}_0$ as $\varepsilon \to 0$.

The proof given by Skorokhod is based on construction of measurable functions $f_\varepsilon(x)$ acting $[0, 1] \to \mathbb{X}$ such that: (a) $f_\varepsilon(\rho) \xrightarrow{d} \xi_\varepsilon$, for $\varepsilon \geq 0$, where $\rho$ is a random variable uniformly distributed in interval $[0, 1]$, (b) $f_\varepsilon(x) \to f_0(x)$ as $\varepsilon \to 0$, for all $x \in [0, 1]$ except at most a countable set of points.

Then, one can define the random variables $\tilde{\xi}_\varepsilon = f_\varepsilon(\rho)$, for $\varepsilon \geq 0$. 

---
A typical application of Theorem 2 relates to proofs of Theorem 1.

Let proposition (i) holds, i.e. \( \xi_\varepsilon \xrightarrow{d} \xi_0 \) as \( \varepsilon \to 0 \).

Let \( \xi_\varepsilon \) be the random variables constructed according to Theorem 2.

Then, for any real-valued function \( f(x) \) a.s. continuous with respect to the distribution of the limiting random variable \( \xi_0 \), the random variables \( f(\xi_\varepsilon) \xrightarrow{p} f(\xi_0) \) as \( \varepsilon \to 0 \).

But, \( f(\xi_\varepsilon) \xrightarrow{d} f(\xi_\varepsilon) \). Since, the a.s. convergence implies the convergence in distribution, the random variables \( f(\xi_\varepsilon) \xrightarrow{d} f(\xi_0) \) as \( \varepsilon \to 0 \).

Thus, proposition (iii) holds.

Also, in the case where the function \( f(x) \) is bounded, \( E f(\xi_\varepsilon) \to E f(\xi_0) \) as \( \varepsilon \to 0 \), by the Lebesgue theorem. Hence, \( E f(\xi_\varepsilon) \to E f(\xi_0) \) as \( \varepsilon \to 0 \), since \( E f(\xi_\varepsilon) = E f(\xi_\varepsilon) \).

Thus, proposition (ii) also holds.

Finally, let proposition (ii) holds. Let us take \( f(x) = \chi_A(x) \), where set \( A \in \mathcal{B}_\mathbb{X} \) such that \( P\{\xi_0 \in \partial A\} = 0 \). Since \( \partial A \) is the set of discontinuity points for function \( \chi_A(x) \), this function is a.s. continuous with respect to the distribution of the random variable \( \xi_0 \). Thus, by proposition (ii),

\[
E \chi_A(\xi_\varepsilon) = F_\varepsilon(A) \to E \chi_A(\xi_0) = F_0(A) \quad \text{as} \quad \varepsilon \to 0.
\]

Thus, proposition (i) holds.

### 3.4 Limit theorems for randomly stopped stochastic processes.

Let \( \bar{\xi}_\varepsilon(\cdot) = \langle \bar{\xi}_\varepsilon(t), t \in [0, T] \rangle, \varepsilon \geq 0 \) be continuous stochastic processes taking values in space \( \mathbb{R}_m \).

Processes \( \bar{\xi}_\varepsilon(\cdot) \) can be considered as random variables taking values in Polish space \( C^m_{[0, T]} \).

We shall use symbol \( \bar{\xi}_\varepsilon(\cdot) \xrightarrow{d} \bar{\xi}_0(\cdot) \) as \( \varepsilon \to 0 \) to denote convergence in distribution of random variables \( \bar{\xi}_\varepsilon(\cdot) \).

Also, let us introduce so-called modulus of U-comactness for continuous functions \( \bar{x}(\cdot) = \langle \bar{x}(t), t \in [0, T] \rangle \) from space \( C^m_{[0, T]} \),

\[
\Delta_{U,c}(\bar{x}(\cdot), c, T) = \sup_{0 \leq t' \leq t'' \leq t'+c \leq T} |\bar{x}(t') - \bar{x}(t'')|.
\]
The following theorem belonging to Prokhorov (1956) give necessary and sufficient conditions of convergence in distribution random variables $\bar{\xi}_\varepsilon(\cdot)$.

**Theorem 3.** Processes $\bar{\xi}_\varepsilon(\cdot) \xrightarrow{d_U} \bar{\xi}_0(\cdot)$ as $\varepsilon \to 0$ if and only if the following condition holds:

**U:** (a) $\bar{\xi}_\varepsilon(t), t \in \mathbb{U} \xrightarrow{d} \bar{\xi}_\varepsilon(t), t \in \mathbb{U}$ as $\varepsilon \to 0$, where $\mathbb{U}$ is some set dense in interval $[0, T]$ and containing points 0 and $T$,

(b) $\lim_{c \to 0} \lim_{\varepsilon \to 0} P\{\Delta_{U,c}(\bar{\xi}_\varepsilon(\cdot), c, T) > \delta\} = 0$, for $\delta > 0$.

Let $\xi_\varepsilon(\cdot) = \langle \xi_\varepsilon(t), t \in [0, T]\rangle$ be real-valued continuous stochastic processes and $\nu_\varepsilon$ a random variable taking values in interval $[0, T]$ defined, for every $\varepsilon \geq 0$, on some probability space $\langle \Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon \rangle$.

In this case, continuous process $\zeta_\varepsilon(\cdot) = \langle (\nu_\varepsilon, \xi_\varepsilon(t)), t \in [0, T]\rangle$ can be considered as random variable taking values in space $C^2_{[0,T]}$.

According Theorem 3, random variables $\bar{\zeta}_\varepsilon(\cdot) \xrightarrow{d_U} \bar{\zeta}_0(\cdot)$ as $\varepsilon \to 0$ if the following condition holds:

**V:** (a) $(\nu_\varepsilon, \xi_\varepsilon(t)), t \in \mathbb{U} \xrightarrow{d} (\nu_\varepsilon, \xi_\varepsilon(t)), t \in \mathbb{U}$ as $\varepsilon \to 0$, where $\mathbb{U}$ is some set dense in interval $[0, T]$ and containing points 0 and $T$,

(b) $\lim_{c \to 0} \lim_{\varepsilon \to 0} P\{\Delta_{U,c}(\xi_\varepsilon(\cdot), c, T) > \delta\} = 0$, for $\delta > 0$.

**Theorem 4.** Let condition **V** holds. Then the following relation holds,

$\xi_\varepsilon(\nu_\varepsilon) \xrightarrow{d} \xi_0(\nu_0)$ as $\varepsilon \to 0$.

According Skorokhod representation Theorem 2 one can construct some probability space $\langle \Omega, \mathcal{F}, P \rangle$ and random variables $\tilde{\zeta}_\varepsilon(\cdot), \varepsilon \geq 0$, defined on this space, such that (a) the random variables $\tilde{\zeta}_\varepsilon(\cdot), \varepsilon \geq 0$, and (b) the random variables $\tilde{\zeta}_\varepsilon(\cdot) \xrightarrow{p_1} \tilde{\zeta}_0(\cdot)$ as $\varepsilon \to 0$. 

11
Let us $A = \{ \omega : \tilde{\zeta}_\varepsilon(\cdot, \omega) \to \tilde{\zeta}_0(\cdot, \omega) \text{ as } \varepsilon \to 0 \}$. According the above a.s. convergence relation, $P(A) = 1$ and, for $\omega \in A$,

$$d_{U,T}(\tilde{\zeta}_\varepsilon(\cdot, \omega), \tilde{\zeta}_0(\cdot, \omega)) = \sup_{t \in [0,T]} \sqrt{|\nu_\varepsilon(\omega) - \nu_0(\omega)|^2 + |\tilde{\xi}_\varepsilon(t, \omega) - |\tilde{\xi}_0(t, \omega)|^2} \to 0 \text{ as } \varepsilon \to 0.$$ 

The above relation implies that following relation hold, for $\omega \in A$,

$$|\tilde{\xi}_\varepsilon(\tilde{\nu}_\varepsilon(\omega), \omega) - \tilde{\xi}_0(\tilde{\nu}_0(\omega), \omega)| \leq |\tilde{\xi}_\varepsilon(\tilde{\nu}_\varepsilon(\omega), \omega) - \tilde{\xi}_0(\tilde{\nu}_\varepsilon(\omega), \omega)| + |\tilde{\xi}_0(\tilde{\nu}_\varepsilon(\omega), \omega) - \tilde{\xi}_0(\tilde{\nu}_0(\omega), \omega)| \leq \sup_{t \in [0,T]} |\tilde{\xi}_\varepsilon(t, \omega) - |\tilde{\xi}_0(t, \omega)| + |\tilde{\xi}_0(\tilde{\nu}_\varepsilon(\omega), \omega) - \tilde{\xi}_0(\tilde{\nu}_0(\omega), \omega)| \to 0 \text{ as } \varepsilon \to 0.$$ 

Since $P(A) = 1$, the above relation implies that random variables $\tilde{\xi}_\varepsilon(\tilde{\nu}_\varepsilon) \overset{P_1}{\longrightarrow} \tilde{\xi}_0(\tilde{\nu}_0)$ as $\varepsilon \to 0$, and, thus, $\tilde{\xi}_\varepsilon(\tilde{\nu}_\varepsilon) \overset{d}{\longrightarrow} \tilde{\xi}_0(\tilde{\nu}_0)$ as $\varepsilon \to 0$.

Also, relation $\tilde{\xi}_\varepsilon(\cdot) \overset{d}{=} \tilde{\xi}_\varepsilon(\cdot)$ readily implies that the random variable $\tilde{\xi}_\varepsilon(\tilde{\nu}_\varepsilon) \overset{d}{=} \xi_\varepsilon(\nu_\varepsilon)$, for every $\varepsilon \geq 0$.

The above remarks imply that $\xi_\varepsilon(\nu_\varepsilon) \overset{d}{\longrightarrow} \xi_0(\nu_0)$ as $\varepsilon \to 0$.

Theorem 4 can be generalised to the case of continuous processes $\xi_\varepsilon(t), t \geq 0$ defined on interval $[0, \infty)$ and unbounded stoping random variables $\nu_\varepsilon$.

Also, an analogous theorem can be proved for randomly stopped càdlàg processes.

The detailed presentation of the theory of limit theorem for randomly stopped càdlàg processes and superpositions of càdlàg processes can be found in the book by Silvestrov (2004),
3.5 Convergence of Lebesgue integrals in the series scheme. Let, for every \( \varepsilon \geq 0 \), \( f_\varepsilon(s) \) be real-valued bounded Borel functions defined on \( R_1 \).

We use the symbol \( f_\varepsilon(\cdot) \xrightarrow{u} f_0(\cdot) \) as \( \varepsilon \to 0 \) to indicate that functions \( f_\varepsilon(\cdot) \) converge to a function \( f_0(\cdot) \) \textit{locally uniformly} at point \( s \) as \( \varepsilon \to 0 \), if,

\[
\lim_{0<u \to 0} \overline{\lim}_{\varepsilon \to 0} \sup_{|v| \leq u} |f_\varepsilon(s + v) - f_0(s)| = 0.
\] (1)

**Lemma 4** Functions \( f_\varepsilon(\cdot) \xrightarrow{u} f_0(\cdot) \) as \( \varepsilon \to 0 \) if and only if \( f_\varepsilon(s_\varepsilon) \to f_0(s) \) as \( \varepsilon \to 0 \), for any \( s_\varepsilon \to s \) as \( \varepsilon \to 0 \).

Let \( \mathcal{B}_1 \) denote the Borel \( \sigma \)-algebra on \( \mathbb{R}_1 \) and let, for every \( \varepsilon \geq 0 \), \( \mu_\varepsilon(A) \) be a measure on \( \mathcal{B}_1 \) taking finite values on bounded Borel sets.

We use the symbol \( \mu_\varepsilon(A) \Rightarrow \mu_0(A) \) as \( \varepsilon \to 0 \) to indicate that the measures \( \mu_\varepsilon(A) \) weakly converge to a measure \( \mu_0(A) \) as \( \varepsilon \to 0 \). This means that, for all \(-\infty < u \leq v < \infty\) such that the limiting measure has not atoms in the points \( u \) and \( v \),

\[
\mu_\varepsilon((u, v]) \to \mu_0((u, v]) \quad \text{as} \quad \varepsilon \to 0.
\] (2)
Let now \( f_\varepsilon(s) \) be real-valued bounded Borel functions defined on \( \mathbb{R}_1 \) and \( \mu_\varepsilon(A) \) be finite measures on \( \mathcal{B}_1 \) for every \( \varepsilon \geq 0 \).

**Lemma 5** Let the following conditions hold: (\( \alpha \)) \( \mu_\varepsilon(A) \Rightarrow \mu_0(A) \) as \( \varepsilon \to 0 \); (\( \beta \)) \( \mu_\varepsilon(\mathbb{R}_1) \to \mu_0(\mathbb{R}_1) < \infty \) as \( \varepsilon \to 0 \); (\( \gamma \)) \( \lim_{\varepsilon \to 0} \sup_{s \in \mathbb{R}_1} |f_\varepsilon(s)| = f < \infty \); (\( \delta \)) functions \( f_\varepsilon(\cdot) \xrightarrow{\text{w}} f_0(\cdot) \) as \( \varepsilon \to 0 \), in every point \( s \in S \), where \( S \) is some subset of \( \mathcal{B}_1 \); (\( \epsilon \)) \( \mu_0(S) = 0 \). Then,

\[
\int_{\mathbb{R}_1} f\varepsilon(s)\mu_\varepsilon(ds) \to \int_{\mathbb{R}_1} f_0(s)\mu_0(ds) \text{ as } \varepsilon \to 0. \tag{3}
\]

If \( \mu_0(\mathbb{R}_1) = 0 \) then the statement of the lemma follows in an obvious way from conditions (\( \beta \)) and (\( \gamma \)), and the corresponding limit takes the value zero. If \( \mu_0(\mathbb{R}_1) > 0 \) then, condition (\( \beta \)) implies that \( \mu_\varepsilon(\mathbb{R}_1) > 0 \) for \( \varepsilon \) small enough, say \( \varepsilon \leq \varepsilon_0 \). Condition (\( \beta \)) permits to reduce the proof to the case where \( \mu_\varepsilon(\cdot) \) are probability measures. This follows from the identity

\[
\int_{\mathbb{R}_1} f\varepsilon(s)\mu_\varepsilon(ds) = \mu_\varepsilon(\mathbb{R}_1) \int_{\mathbb{R}_1} f\varepsilon(s)\mu_\varepsilon(ds),
\]

where \( \mu_\varepsilon(A) = \mu_\varepsilon(A)/\mu_\varepsilon(\mathbb{R}_1) \) is a probability measure and \( \varepsilon \leq \varepsilon_0 \).

Also, taking the definition of a limit in terms of convergent subsequences, one can always reduce the consideration to the case where the parameter \( \varepsilon \to \varepsilon_0 \) runs only over some subsequence of positive values, \( \varepsilon_n, n = 1, 2, \ldots \) such that \( \varepsilon_n \to 0 \) as \( n \to \infty \).

Skorokhod representation Theorem 3 implies that, under condition (\( \alpha \)), it is possible to construct a sequence of random variables \( \xi_{\varepsilon_n} \) defined on some probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) such that (\( \mathfrak{h} \)) the random variable \( \xi_{\varepsilon_n} \) has the distribution \( \mu_\varepsilon(\cdot) \) for every \( n = 0, 1, \ldots \); (\( \mathfrak{i} \)) \( \xi_{\varepsilon_n} \xrightarrow{\mathbb{P}} \xi_0 \) as \( n \to \infty \).

Let \( A \) be the set of elementary events \( \omega \in \Omega \) such that (\( \mathfrak{j} \)) \( \xi_{\varepsilon_n}(\omega) \to \xi_0(\omega) \) as \( n \to \infty \). Let also \( B \) be the set of elementary events \( \omega \) for which \( \xi_0(\omega) \in S \).

Relation (\( \mathfrak{i} \)) and condition (\( \mathfrak{e} \)) imply that (\( \mathfrak{k} \)) \( \mathbb{P}(A \cap B) = 1 \).

By Lemma 4, it follows from (\( \mathfrak{j} \)) and condition (\( \mathfrak{d} \)) that (\( \mathfrak{l} \)) \( f_{\varepsilon_n}(\xi_{\varepsilon_n}(\omega)) \to f_0(\xi_0(\omega)) \) as \( n \to \infty \) for all \( \omega \in A \cap B \).

Relations (\( \mathfrak{k} \)) and (\( \mathfrak{l} \)) mean that (\( \mathfrak{m} \)) \( f_{\varepsilon_n}(\xi_{\varepsilon_n}) \xrightarrow{\mathbb{P}} f_0(\xi_0) \) as \( n \to \infty \).

Note that condition (\( \mathfrak{g} \)) implies that (\( \mathfrak{n} \)) there exists \( N < \infty \) such that \( |f_{\varepsilon_n}(\xi_{\varepsilon_n})| \leq 2f < \infty \) for \( n \geq N \). Also conditions (\( \mathfrak{g} \)) and (\( \mathfrak{d} \)) imply that (\( \mathfrak{o} \)) \( \sup_{s \in S} |f_0(s)| \leq 2f < \infty \), and, therefore, by condition (\( \mathfrak{e} \)), (\( \mathfrak{p} \)) \( |f_0(\xi_0)| \leq
2f < ∞ with probability 1. Hence, by the Lebesgue theorem, it follows from (m), (n), and (p) that

$$\mathbb{E}_{f_{\varepsilon_n}}(\xi_{\varepsilon_n}) \rightarrow \mathbb{E}_{f_0}(\xi_0) \text{ as } n \rightarrow \infty.$$ 

By (h), this is equivalent to the assertion of the lemma.

4. Problems

3.1. Prove Lemma 1 and formulate and prove its analogue for random variables taking values in a Polish space.

3.2. Prove Lemma 2 and formulate and prove its analogue for random variables taking values in a Polish space.

3.3. Try to generalise Theorem 4 on the case of continuous processes \(\xi_{\varepsilon}(t), t \geq 0\) defined on interval \([0, \infty)\) and unbounded stoping random variables \(\nu_{\varepsilon} \).

3.4. Try to formulate and prove a theorem analogous to Theorem 4 for randomly stopped càdlàg processes.

3.5. Prove Lemma 4.

3.6. Let \(\xi_n, n = 0, 1, 2, \ldots\) be a sequence of real-valued random variables such that random variables \(\xi_n/n^\alpha \xrightarrow{d} \xi_0\) as \(n \rightarrow \infty\), where \(\alpha > 0\) and \(\nu_n, n = 0, 1, 2, \ldots\) be a sequence of positive integer-valued random variables such that random variables \(\nu_n/n^\beta \xrightarrow{d} \nu_0\) as \(n \rightarrow \infty\), where \(\beta > 0\). Let, also, the above two sequences of random variables are independent. Using the method of one probability space, prove that random variables \(\xi_{\nu_n}/n^{\alpha\beta} \xrightarrow{d} \xi_0\nu_0^\beta\) as \(n \rightarrow \infty\).