V. I. BOGACHEV AND A. F. MIFTAKHOV

ON WEAK CONVERGENCE OF FINITE-DIMENSIONAL AND INFINITE-DIMENSIONAL DISTRIBUTIONS OF RANDOM PROCESSES

We study conditions on metrics on spaces of measurable functions under which weak convergence of Borel probability measures on these spaces follows from weak convergence of finite-dimensional projections of the considered measures.

INTRODUCTION

In the theory of stochastic processes, one often uses weak convergence of finite-dimensional distributions of processes. In terms of distributions in path spaces this corresponds to weak convergence of measures on a path space equipped with the topology of pointwise convergence. However, in many applications this convergence turns out to be too weak, so it becomes necessary to complement it by various additional conditions in order to obtain convergence of distributions of processes in functional spaces with various norms and metrics. For instance, $C[0,1]$ is a natural path space for continuous processes on $[0,1]$; convergence of finite-dimensional distributions does not imply convergence of distributions in $C[0,1]$ equipped with the usual sup-norm, so it is necessary to require additionally the uniform tightness of distributions. These problems were thoroughly studied already 30-40 years ago (see [3], [7], [9], [10], [13], [14], [15], [16], [17], and [22]). A significant number of the known results is covered by the following scheme: if a certain path space X is equipped with a norm or a metric (in [9], the whole range of such metrics is considered that imply convergence in measure), then for weak convergence in X of the distributions P_n of processes ξ_n it is sufficient to have weak convergence of their finite-dimensional distributions and the uniform tightness of the measures P_n, which means that for each $\varepsilon > 0$ there is a compact set $K_\varepsilon \subset X$ such that $P_n(X \backslash K_\varepsilon) < \varepsilon$ for all n. Usually, the latter condition cannot be omitted. However, it is known (see [24] and [25]) that this condition can be omitted in the case where the path space is equipped with the metric of convergence in measure (in place of the usual uniform metric). The proof of this result in [24] was based on the paper [16]. We give a relatively short proof of a more general result (see Theorem 2.1) that does not employ any other results and actually yields also a more general assertion than in [16] (see Theorem 2.6). Corollary 2.4 partially improves the results from [14]–[16], [17] on weak convergence of measures on L^p: in place of convergence of moments as in the cited papers their uniform boundedness for some larger exponent is required.

Our second main result, which was originally our motivation, concerns the following question. The standard metric (or semimetric) of convergence in measure on the space of Borel functions or on the space of continuous functions on $[0,1]$ has the property that the pointwise convergence of a sequence yields convergence in this metric. In addition, this metric is translation invariant and monotone in the sense that $d(f,0) \leq d(g,0)$ if $|f| \leq |g|$. Do there exist other metrics, not equivalent to metrics of convergence in measure (for different measures on $[0,1]$), having these properties? We prove Theorem 3.1

that asserts that a metric with the aforementioned three properties is equivalent to the metric of convergence in measure for some Borel measure provided that it satisfies one additional technical condition. However, this additional condition is not necessary, so that the question above in full generality remains open.

Let us note at once that the established property to obtain weak convergence from convergence of finite-dimensional projections cannot hold for a norm in place of the metric of convergence in measure. Indeed, if we take non-random processes $x_n \in C[0, 1]$ with disjoint supports in the intervals $[2^{-n}, 2^{-n} + 8^{-n}]$, then their finite-dimensional distributions, defined by values at the points $t_1, \ldots, t_k$, converge weakly to the respective finite-dimensional distributions of the identically zero process. The same is true for the processes $C_n x_n$ for any constants C_n. If $C[0, 1]$ is equipped with a norm p, then we can take C_n such that $p(C_n x_n) \to +\infty$, and then there is no weak convergence of Dirac's measures at the functions $C_n x_n$. Therefore, in the wide range of metrics on path spaces considered in [9] and ensuring convergence in measure, such as the uniform metric and the L^p-metrics, only the metric of convergence in measure has the property that for this metric it is not necessary to require additionally the uniform tightness.

1. NOTATION AND AUXILIARY RESULTS

The image of a measure μ on a measurable space $(X, \mathcal{B})$ under a measurable mapping f with values in a measurable space $(Y, \mathcal{A})$ is denoted by the symbol $\mu \circ f^{-1}$ and defined by the formula
$$\mu \circ f^{-1}(A) = \mu(f^{-1}(A)), \quad A \in \mathcal{A}.$$

For a general metric space (X, d), the space $\mathcal{P}(X)$ of probability Borel measures on X is equipped with the weak topology defined on the whole space $\mathcal{M}(X)$ of all finite Borel measures (possibly, signed) by the seminorms of the form
$$q(\mu) = \left| \int_X f(x)\mu(dx) \right|,$$

where f is a bounded continuous function on X; the set of all such functions is denoted by $C_b(X)$. A sequence of Borel measures μ_n converges weakly to a Borel measure μ if for all $f \in C_b(X)$ we have

$$(1.1) \qquad \lim_{n \to \infty} \int_X f(x)\, \mu_n(dx) = \int_X f(x)\, \mu(dx).$$

If the space X is separable, then on the set $\mathcal{P}(X)$ weak convergence is metrizable; different metrics are known that generate the weak topology on $\mathcal{P}(X)$; for example (see details in [4, Chapter 8] or [11]), it is possible to use the Prohorov metric or the Kantorovich–Rubinstein metric [18] (equivalent to the Fortet–Mourier metric [12]) defined by the formula
$$d_0(\mu, \nu) = \sup\left\{ \int_X \varphi(x)\mu(dx) - \int_X \varphi(x)\nu(dx), \ \varphi \in \mathrm{Lip}_1(X), \sup |\varphi(x)| \le 1 \right\},$$

where $\mathrm{Lip}_1(X)$ is the class of Lipschitz continuous functions φ on X with the Lipschitz constant 1, i.e., $|\varphi(y) - \varphi(x)| \le d(x, y)$. In the case of bounded X, one can also use the equivalent Kantorovich metric (see [4] and [5] about such metrics) defined by
$$d_K(\mu, \nu) = \sup\left\{ \int_X \varphi(x)\mu(dx) - \int_X \varphi(x)\nu(dx), \ \varphi \in \mathrm{Lip}_1(X) \right\}.$$

In a nonseparable case, the same is true if in place of the whole set $\mathcal{P}(X)$ we take only its part consisting of tight measures, i.e., measures μ such that, for each $\varepsilon > 0$, there is a compact set $K_\varepsilon \subset X$ such that $\mu(X \backslash K_\varepsilon) < \varepsilon$.

It is important to note that for weak convergence of probability measures μ_n to a probability measure μ it is enough to have convergence in (1.1) only for functions f of

class $\mathrm{Lip}_1(X)$; this is not true for signed measures. This fact implies the following known simple condition for weak convergence (cf. [26, 1.12.3]).

Lemma 1.1. *If X is separable, then for weak convergence of measures $\mu_n \in \mathcal{P}(X)$ to a measure $\mu \in \mathcal{P}(X)$ it is enough to have equality* (1.1) *for functions of the form*

$$f(x) = \max(c_0, c_1 - d(x, y_1), \ldots, c_n - d(x, y_n)),$$

where $c_0, \ldots, c_n$ are real numbers, $y_1, \ldots, y_n \in X$. If the space X is bounded, then it is enough to verify (1.1) *just for polynomials in the variables $d(x, y_1), \ldots, d(x, y_n)$, where $y_1, \ldots, y_n \in X$.*

In the case of tight measures μ_n and μ the separability of X is not needed.

Proof. Let us take a countable everywhere dense set $\{y_i\}_{i=1}^{\infty}$ in X. It is readily verified that for any bounded function $f \in \mathrm{Lip}_1(X)$ such that $f \geq c$ we have

$$f(x) = \sup_{j \in \mathbb{N}} \max(c, f(y_j) - d(x, y_j)).$$

In the case of a bounded space X one can take sup over the values $f(y_j) - d(x, y_j)$. Let us set

$$f_k(x) = \max_{j \leq k} \max(c, f(y_j) - d(x, y_j)).$$

By assumption, for every $k \in \mathbb{N}$, we have the equality

$$\lim_{n \to \infty} \int_X f_k(x)\mu_n(dx) = \int_X f_k(x)\mu(dx).$$

Since, as $k \to \infty$, the right-hand side tends to the integral of f with respect to the measure μ by the Lebesgue dominated convergence theorem and

$$\int_X f_k(x)\mu_n(dx) \leq \int_X f(x)\mu_n(dx),$$

we obtain the inequality

$$\int_X f(x)\mu(dx) \leq \liminf_{n \to \infty} \int_X f(x)\mu_n(dx).$$

By replacing f with $-f$ we obtain the opposite inequality, which completes the proof of the first assertion.

In the case of bounded X, any continuous function in the variables $d(x, y_1), \ldots, d(x, y_n)$ is uniformly approximated by polynomials in these variables, so it suffices to verify convergence of integrals only for these polynomials.

Finally, if all measures μ_n and μ are tight, then they are concentrated on a separable part of the space X. $\qquad\square$

2. Weak convergence of measures on functional spaces related to convergence in measure

Let $(T, \mathcal{B}, \lambda)$ be a measurable space with a finite nonnegative measure λ and let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $L^0(\lambda)$ denote the space of equivalence classes of measurable real functions on T with the metric $d(\cdot, \cdot)$ of convergence in measure λ on T defined by the formula

$$d(x, y) = \int_T \min\{|x(t) - y(t)|, 1\}\, \lambda(dt).$$

We recall that in nontrivial cases (for example, for Lebesgue measure) convergence in this metric cannot be defined by a norm, and the space of measurable functions with this metric is not a locally convex space (see [4, Exercise 4.7.61]). Note that the metric d is bounded.

We assume that the measure λ is separable, i.e., $L^0(\lambda)$ is separable; this is equivalent to the separability of $L^1(\lambda)$ or $L^2(\lambda)$.

Suppose we are given a certain space $\mathcal{F}$ of functions on T measurable with respect to the measure λ (not equivalence classes, like L^1, but individual functions) such that the equality almost everywhere of two functions in $\mathcal{F}$ implies their pointwise equality. Examples of such spaces $\mathcal{F}$ are the class of continuous functions $C(T)$ in case of a topological space T equipped with a Borel measure λ that does not vanish on nonempty open sets and the class of left-continuous functions on an interval with Lebesgue measure.

Let us equip the space $\mathcal{F}$ with the metric d of convergence in measure λ, i.e., by the metric from $L^0(\lambda)$ (although $L^0(\lambda)$ consists of equivalence classes). The obtained space will be denoted by $\mathcal{F}_d$. It is clear that $\mathcal{F}_d$ is separable.

In addition, suppose that for every $t \in T$ the evaluation function $x \mapsto x(t)$ on $\mathcal{F}$ is measurable with respect to each Borel measure on $\mathcal{F}_d$.

For example, this assumption is satisfied if λ is Lebesgue measure on the interval $[0,1]$ and $\mathcal{F} = C[0,1]$ (in this case the indicated evaluation functions are Borel measurable).

The image of a measure $\mu \in \mathcal{P}(\mathcal{F}_d)$ under the mapping from $\mathcal{F}_d$ to $\mathbb{R}^k$ defined by $x \mapsto (x(t_1), \ldots, x(t_k))$ is called a finite-dimensional distribution of the measure μ, i.e., its value on a Borel set $B \subset \mathbb{R}^k$ is given by

$$\mu\Big(x \in \mathcal{F}\colon \ (x(t_1), \ldots, x(t_k)) \in B\Big).$$

By our assumption of measurability of the evaluation functions, the image of the measure μ under this map is well-defined.

Theorem 2.1. *Suppose we are given a measure $\mu \in \mathcal{P}(\mathcal{F}_d)$ and a sequence of measures $\mu_n \in \mathcal{P}(\mathcal{F}_d)$. Suppose that the finite-dimensional distributions of the measures μ_n converge weakly to the respective finite-dimensional distributions of the measure μ. Then the measures μ_n converge weakly to the measure μ on the space $\mathcal{F}_d$.*

Proof. By assumption, the space $\mathcal{F}_d$ is separable. Let us take a countable everywhere dense set $\{y_i\}_{i=1}^\infty$ in it. By the lemma, we have to verify (1.1) for polynomials in the variables

$$d(x, y_i) = \int_T \min(1, |x(t) - y_i(t)|)\lambda(dt), \quad \text{where } y_1, \ldots, y_k \in L^0(\lambda).$$

Weak convergence of finite-dimensional distributions of the given measures means that, for all $t_1, \ldots, t_k \in T$ and all $\varphi \in C_b(\mathbb{R}^k)$, there holds the equality

$$\lim_{n \to \infty} \int_X \varphi(x(t_1), \ldots, x(t_k))\mu_n(dx) = \int_X \varphi(x(t_1), \ldots, x(t_k))\mu(dx).$$

By the Lebesgue dominated convergence theorem, for any bounded Borel function ψ on $\mathbb{R}^k \times \mathbb{R}^k$ that is continuous in the first k variables we obtain the equality

$$\lim_{n \to \infty} \int_X \int_{T^k} \psi(x(t_1), \ldots, x(t_k), t_1, \ldots, t_k)\lambda(dt_1) \cdots \lambda(dt_k)\mu_n(dx)$$
$$= \int_X \psi(x(t_1), \ldots, x(t_k), t_1, \ldots, t_k)\lambda(dt_1) \cdots \lambda(dt_k)\mu(dx).$$

In particular, this equality is true for all functions of the form

$$\varphi(s_1, \ldots, s_k, t_1, \ldots, t_k) = \prod_{i=1}^k \min(1, |s_i - y_i(t_i)|)$$

and their linear combinations. Thus, this equality holds for all polynomials in the variables $d(x, y_i)$. $\qquad\square$

In terms of stochastic processes we obtain the following result. We shall deal with measurable processes $(\xi_t)_{t \in T}$ on a probability space $(\Omega, \mathcal{A}, P)$; actually, we need a bit weaker measurability condition: it suffices to assume that the function $(t, \omega) \mapsto \xi_t(\omega)$ is $\lambda \otimes P$-measurable. If the σ-field $\mathcal{A}$ is countably generated and the process $(\xi_t)_{t \in T}$ has the property that the function

$$t \mapsto \int_A \arctan \xi_t(\omega) P(d\omega)$$

is λ-measurable for every $A \in \mathcal{A}$, then $(\xi_t)_{t \in T}$ has a measurable version, i.e., there exists a measurable process $(\eta_t)_{t \in T}$ such that, for each fixed $t \in T$, we have $\eta_t = \xi_t$ almost surely (see, e.g., [4, V. 2, p. 71] or [6, p. 54, Exercise 1.8.14]). Every measurable process $(\xi_t)_{t \in T}$ generates a probability measure on $L^0(\lambda)$, called its distribution and denoted by P_ξ. In the case of a bounded process (or a process with paths in $L^2(\lambda)$) the measure P_ξ can be defined on $L^2(\lambda)$ as the image of the measure P under the mapping

$$\omega \mapsto \sum_{n=1}^{\infty} \int_T \xi_t(\omega) e_n(t) \lambda(dt) e_n,$$

where $\{e_n\}$ is an orthonormal basis in $L^2(\lambda)$.

In the general case the measure P_ξ can be obtained as a limit in variation of the distributions of bounded processes $\max(-k, \min(k, \xi_t))$; see also [4, v. 2, p. 171, Exercise 7.14.115]. If $\mathcal{F}_d$ is a full measure set with respect to P_ξ, then the distribution of the process can be regarded on $\mathcal{F}$; in this case we say that P_ξ is concentrated on $\mathcal{F}_d$. For example, this is possible for processes with continuous paths.

Corollary 2.2. *Let ξ and $\{\xi_n\}_{n=1}^{\infty}$ be stochastic processes on T with trajectories in the space $\mathcal{F}_d$ such that the finite-dimensional distributions of ξ_n converge to the respective finite-dimensional distributions of ξ. Then the measures P_{ξ_n} converge weakly to the measure P_ξ on the space $\mathcal{F}_d$.*

Remark 2.3. It is clear from the proof of the theorem that it suffices to have weak convergence of finite-dimensional distributions generated by points t_i in a set $T_0 \subset T$ of full λ-measure.

It is worth noting that, for a uniformly bounded sequence of functions, its convergence in measure λ is equivalent to convergence in $L^2(\lambda)$, and if the sequence is bounded in $L^p(\lambda)$, then its convergence in measure is equivalent to convergence in $L^r(\lambda)$ for each $r < p$. Therefore, although, as noted above, for the L^p-norm the analogue of the proven theorem is false, we obtain the following assertion.

Corollary 2.4. *Let ξ and $\{\xi_n\}_{n=1}^{\infty}$ be stochastic processes on T with trajectories in $\mathcal{F}_d$ such that the finite-dimensional distributions of ξ_n converge weakly to the respective finite-dimensional distributions of ξ. Suppose that for some $p \in (1, +\infty)$ we have*

$$\sup_n \mathrm{E} \|\xi_n\|_{L^p(\lambda)}^p < \infty.$$

Then, for any r in the interval $[1, p)$, the measures P_{ξ_n} converge weakly to the measure P_ξ on the space $L^r(\lambda)$.

Proof. By the Skorohod theorem (see [4, Chapter 8]) we have Borel mappings

$$\eta_n, \eta \colon [0, 1] \to L^0(\lambda)$$

such that the distributions of ξ_n and ξ coincide with $\lambda_1 \circ \eta_n^{-1}$ and $\lambda_1 \circ \eta^{-1}$, respectively, where λ_1 is Lebesgue measure on $[0, 1]$, and $\eta_n(s) \to \eta(s)$ a.e. in $L^0(\lambda)$. Hence the functions $|\eta_n(s)(t) - \eta(s)(t)|$ converge to zero in measure $\lambda \otimes \lambda_1$. Since the integral of $|\eta_n(s)(t)|^p$ with respect to $\lambda \otimes \lambda_1$ equals $\mathrm{E} \|\xi_n\|_{L^p(\lambda)}^p$, we obtain that the integrals of

$|\eta_n(s)(t) - \eta(s)(t)|^r$ converge to zero whenever $r < p$, i.e., the processes η_n converge to η in $L^r(\lambda \otimes \lambda_1)$, which yields weak convergence of their distributions in $L^r(\lambda)$.   □

We note that our condition of the uniform boundedness of moments is weaker than convergence of moments assumed in [14], [15], [16], [17], and also in [1] and [2] (where, however, much more general spaces are considered), but $r < p$.

For $r = p$ this conclusion can be false, but if the processes ξ_n are uniformly bounded, then it is true for any $r \in [1, +\infty)$.

Since by Prohorov's theorem weak convergence of measures on the separable (by our assumption) spaces $L^0(\lambda)$ and L^p implies the uniform tightness of these measures, in all aforementioned cases the uniform tightness follows from weak convergence of the finite-dimensional distributions. If the set $\mathcal{F}$ is measurable in $L^0(\lambda)$ with respect to each Borel measure (for example, is a Suslin space, as is the case for $C[0,1]$), then the measures μ_n are uniformly tight not only in $L^0(\lambda)$, but also in $\mathcal{F}_d$ (see LeCam's theorem in [19], [3] or [4, Theorem 8.6.4]).

Note also that by the aforementioned Skorohod theorem any weakly convergent sequence of Borel probability measures on a complete separable metric space X can be obtained as a sequence of distributions of almost everywhere convergent measurable maps from $[0,1]$ with Lebesgue measure to X. Therefore, in Corollary 2.2 we obtain stochastic processes η_n and η such that $P_{\xi_n} = P_{\eta_n}$, $P_\xi = P_\eta$ and $\eta_n(\cdot, \omega) \to \eta(\cdot, \omega)$ in measure λ for almost every ω. A similar assertion is true in the situation of Corollary 2.4, which has been used in its proof.

In some cases, it is useful to consider measures on path spaces not on the Borel σ-algebra, but on smaller σ-algebras. For example, it is known (see [8]) that the Wiener measure on the space of continuous paths on $[0, +\infty)$ with a finite norm of the form $\sup_t |x(t)/\alpha(t)|$ cannot be defined on the whole Borel σ-algebra generated by this norm, because the obtained space is not separable. In such cases, one can consider smaller σ-algebras, for example, the σ-algebra $\mathcal{B}_1(X)$ generated by all balls of our metric space. In the lemma, we have considered precisely convergence of integrals of functions measurable with respect to $\mathcal{B}_1(X)$. There are other possible solutions: for example, for the Wiener measure we can consider a separable path space with a norm of the form $\sup_t |x(t)/\beta(t)|$ consisting of functions such that $\lim_{t \to \infty} x(t)/\beta(t) = 0$. Of course, not always such a norm is suitable. An advantage of the metric of convergence in measure is its separability under broad assumptions.

The obtained result can be useful for the study of processes with trajectories in the Skorohod space D (see [3] and [21]); certainly, in this case the natural convergence in D should be replaced by convergence in measure (as was done in [20] and [23]).

We recall that the space D consists of right-continuous functions on $[0,1]$ having left limits. Convergence in D is determined by the metric d_0 defined as follows: given $x, y \in D$, $d_0(x, y)$ is the infimum of positive ε with the property that there is an increasing homeomorphism h of $[0,1]$ such that $\sup_t |h(t) - t| \le \varepsilon$ and $\sup_t |x(t) - y(h(t))| \le \varepsilon$. This metric makes D a separable space that is not complete, but there is another metric defining the same convergence and making D a complete separable space. Convergence in D is stronger than convergence in measure on $[0,1]$. For every measure $P \in \mathcal{P}(D)$, there is a set $T_P \subset [0,1]$ with an at most countable complement such that for each $t \in T_P$ the function $x \mapsto x(t)$ is continuous P-almost everywhere on D (see [3, Section 15]). For any t, this function is Borel measurable on D. Therefore, we obtain finite-dimensional distributions of P. Borel probability measures P_n on D converge weakly to a measure P provided that they are uniformly tight and their finite-dimensional distributions generated by points from T_P converge weakly to the respective finite-dimensional distributions of P. It follows from the results above that if we introduce on D the weaker metric of convergence in measure, then weak convergence is ensured by convergence of

finite-dimensional distributions generated by points in T_P. Certainly, the implied uniform tightness will also correspond to the metric d, not to the natural metric d_0.

Remark 2.5.   (i) The assumptions that $\mathcal{F}$ consists of individual functions and not of equivalence classes and that d is a metric on a $\mathcal{F}$ (not merely a semimetric) can be omitted if we modify the concept of finite-dimensional distributions. The usual concept adopted above assumes that the evaluation functions $t \mapsto x(t)$ are well-defined, which is the case when we deal with random processes. In case of measures μ_n and μ on $L^0(\lambda)$ a natural analog of weak convergence of finite-dimensional distributions is the relation

(2.1) $$\lim_{n \to \infty} \int \int_{T^k} \psi(x(t_1), \ldots, x(t_k), t_1, \ldots, t_k) \lambda(dt_1) \cdots \lambda(dt_k) \mu_n(dx)$$

$$= \int \int_{T^k} \psi(x(t_1), \ldots, x(t_k), t_1, \ldots, t_k) \lambda(dt_1) \cdots \lambda(dt_k) \mu(dx)$$

for every bounded function ψ on $\mathbb{R}^k \times T^k$ such that the function

$$(s_1, \ldots, s_k) \mapsto \psi(s_1, \ldots, s_k, t_1, \ldots, t_k)$$

is continuous for all fixed t_i and the function $(t_1, \ldots, t_k) \mapsto \psi(s_1, \ldots, s_k, t_1, \ldots, t_k)$ is λ^k-measurable for all fixed s_i. We shall call this property the integral convergence of finite-dimensional distributions. With this modification the previous theorem remains valid (with the same proof) also for measures on the space $L^0(\lambda)$ of equivalence classes or its subsets.

(ii) It is worth noting that there is another concept of a finite-dimensional projection of a measure on a topological vector space X as the image of this measure under a finite-dimensional operator of the form $x \mapsto (l_1(x), \ldots, l_k(x))$, where $l_1, \ldots, l_k$ are continuous linear functionals on X. However, in our specific situation the space $L^0(\lambda)$ typically has no nonzero continuous linear functionals, as has been noted above.

The hypothesis of the theorem is only a sufficient condition, but not necessary, which is seen from the trivial case of deterministic processes $\xi_n(t)$ that are continuous functions on $[0, 1]$ converging to zero in Lebesgue measure, but not pointwise. The corresponding finite-dimensional distributions are Dirac measures at the points $(\xi_n(t_1), \ldots, \xi_n(t_k))$, hence do not converge weakly. By the Riesz theorem this sequence contains a subsequence that converges almost everywhere. It turns out that in this sense the above theorem admits a partial converse.

Theorem 2.6. *Suppose that we are given a sequence of measurable random processes ξ_n and a measurable random process ξ on T such that there is the integral convergence of finite-dimensional distributions of ξ_n to the corresponding finite-dimensional distributions of ξ in the sense of (2.1). Then there is a subsequence in $\{\xi_n\}$ and a set $T_0 \subset T$ of full measure λ such that we shall have weak convergence of the usual finite-dimensional distributions generated by the points in T_0.*

Proof. According to the previous remark, we have weak convergence of the measures μ_n on $L^0(\lambda)$ generated by our processes to the distribution μ of the process ξ. As we have mentioned above, by the Skorohod theorem, there are random elements η_n, η on $([0, 1], \lambda_0)$, where λ_0 is the usual Lebesgue measure, taking values in $L^0(\lambda)$ and such that $\lambda_0 \circ \eta_n^{-1} = \mu_n$, $\lambda_0 \circ \eta^{-1} = \mu$ and $\eta_n \to \eta$ λ-a.e. Let $\eta_n(s, t)$ and $\eta(s, t)$ be their jointly measurable versions as functions on $([0, 1] \times T, \lambda_0 \otimes \lambda)$. This follows from the fact that the functions η_n converge to η in measure $\lambda_0 \otimes \lambda$, since

$$\int_{[0,1] \times T} \min(1, |\eta_n(s, t) - \eta(s, t)|) ds \lambda(dt) = \int_0^1 d(\eta_n(s), \eta(s)) ds \to 0$$

by Fubini's theorem and the Lebesgue dominated convergence theorem (recall that d is bounded). By the Riesz theorem, there is a subsequence $\{\eta_{n_i}\}$ converging $\lambda_0 \otimes \lambda$-almost everywhere. Thus, there is a set $T_0 \subset T$ of full λ-measure such that for each $t \in T_0$ we have $\eta_{n_i}(s,t) \to \eta(s,t)$ for λ_0-almost all $s \in [0,1]$. Therefore, for all $t_1, \ldots, t_k \in T_0$ we have

$$\lim_{i \to \infty} \int_0^1 \psi(\eta_{n_i}(s,t_1), \ldots, \eta_{n_i}(s,t_k))ds = \int_0^1 \psi(\eta(s,t_1), \ldots, \eta(s,t_k))ds$$

for each bounded continuous function ψ on $\mathbb{R}^k$, as required. $\square$

3. Metric convergence implied by the pointwise convergence

Here we give a sufficient condition that guarantees that a metric or a semimetric on the space of measurable functions on a measurable space $(T, \mathcal{B})$ such that the pointwise convergence implies convergence in this metric is equivalent to the metric generated by convergence in measure for some probability measure on $(T, \mathcal{B})$. The equivalence of two metrics is understood as the coincidence of the collections of converging sequences in these metrics. Obviously, already in the case of the closed interval $[0,1]$ with its Borel σ-field there are incomparable metrics of convergence in measure for different Borel probability measures on $[0,1]$, but convergence in all such metrics follows from the pointwise convergence. We also observe that two probability measures μ and ν generate equivalent metrics of convergence in measure precisely when they are equivalent (have the same classes of zero sets).

If we do not impose any restrictions, then a semimetric on the space of Borel functions convergence in which follows from the pointwise convergence need not be associated with convergence in measure. This is seen from the following example:

$$d(f,g) = |f(0) - g(0) - f(1) + g(1)|.$$

Convergence in this semimetric follows from the pointwise convergence, but cannot be equivalent to convergence in measure for a probability Borel measure μ on $[0,1]$. Indeed, the sequence of constants n does not converge in measure μ, but $d(n,0) = 0$. This semimetric can be used to construct a true metric with the same property on the space $C[0,1]$. To this end, we take

$$d(f,g) = d_0(f,g) + |f(0) - g(0) - f(1) + g(1)|,$$

where d_0 is the usual metric of convergence in Lebesgue measure. Then there is no Borel probability measure behind convergence in the metric d. Indeed, let μ be such a measure. Then it is readily seen that its restriction to $(0,1)$ must be equivalent to Lebesgue measure, but it must also have atoms at 0 and 1, because convergence in d is not equivalent to convergence in Lebesgue measure (the sequence of functions $f_n(x) = nx^n$ converges to zero in Lebesgue measure, but $d(f_n, 0) \to 0$). This leads to a contradiction, since for the functions $g_n(x) = nx^n + n(x-1)^{2n}$ we have $d(g_n, 0) \to 0$, but $g_n(0) = g_n(1) = n$.

Theorem 3.1. *Let $(T, \mathcal{B})$ be a measurable space and let d be a semimetric on the space $\mathcal{F}$ of $\mathcal{B}$-measurable functions such that for all $f, g \in \mathcal{F}$ one has $d(f,g) = d(f - g, 0)$ and $d(f, 0) \leq d(g, 0)$ whenever $|f(t)| \leq |g(t)|$ for all $t \in T$, and, moreover, the pointwise convergence on T implies convergence in this semimetric. Let us define a set function μ on $\mathcal{B}$ by*

$$\mu(A) = \sup\Big\{\sum_i d(I_{A_i}, 0)\colon A_i \in \mathcal{B} \text{ are disjoint}, A = \bigcup_i A_i\Big\}.$$

If $\mu(T) < +\infty$, then μ is a measure on $\mathcal{B}$ and convergence of a sequence of $\mathcal{B}$-measurable functions on T in this measure is equivalent to convergence of this sequence in the metric d.

Proof. We prove that μ is a measure. Let $A = \bigcup_{i=1}^{\infty} A_i$, where $A_i \in \mathcal{B}$ are disjoint. Let us fix $\varepsilon > 0$. Then there are sets $B_j \in \mathcal{B}$ and $C_{i,j} \in \mathcal{B}$ such that B_j are disjoint and $A = \bigcup_{j=1}^{\infty} B_j$, $C_{i,j}$ are mutually disjoint and $A_i = \bigcup_{j=1}^{\infty} C_{i,j}$,

$$\mu(A) < \sum_j d(I_{B_j}, 0) + \varepsilon, \quad \mu(A_i) < \sum_j d(I_{C_{i,j}}, 0) + \frac{\varepsilon}{2^i} \quad \text{for all } i.$$

Hence, on the account of the definition of μ and the equalities $B_j = \bigcup_{i=1}^{\infty} B_j \cap A_i$, $A_i = \bigcup_{j=1}^{\infty} A_i \cap B_j$, $A = \bigcup_{i,j=1}^{\infty} C_{i,j}$ we have

$$\mu(A) - \varepsilon < \sum_j d(I_{B_j}, 0) \leq \sum_j \sum_i d(I_{B_j \cap A_i}, 0) \leq \sum_i \mu(A_i)$$

$$< \sum_i \sum_j d(I_{C_{i,j}}, 0) + \varepsilon \leq \mu(A) + \varepsilon.$$

Therefore, $\mu(A) = \sum_{i=1}^{\infty} \mu(A_i)$. Note that $\mu(A) \geq d(I_A, 0)$ and $\mu(A) = 0$ if $d(I_A, 0) = 0$, since if $\mu(A) > 0$, then there is $A_i \subset A$ with $A_i \in \mathcal{B}$ and $d(I_{A_i}, 0) > 0$, whence it follows that $d(I_A, 0) \geq d(I_{A_i}, 0) > 0$.

Next we prove that for every function $f \in \mathcal{F}$ vanishing μ-almost everywhere we have $d(f, 0) = 0$. Replacing f by $|f|$, we can assume that $f \geq 0$. Let $A = \{t\colon f(t) \neq 0\}$. For every pair of nonnegative integer numbers (k, n) we consider the set

$$A_{k,n} = \left\{ t\colon \frac{k}{2^n} \leq f(t) < \frac{k+1}{2^n} \right\}.$$

Note that $A_{k,n} \subset A$ if $k \geq 1$ and that $d(I_A, 0) = 0$ by the supposed equality $\mu(A) = 0$ and the bound $\mu(A) \geq d(I_A, 0)$, which yields that $d\left(\frac{k}{2^n} I_{A_{k,n}}, 0\right) = 0$ for all k, n. Therefore, by the triangle inequality and the translation invariance of d, the functions $f_n = \sum_{k=1}^{\infty} \frac{k}{2^n} I_{A_{k,n}}$ satisfy the inequality

$$d(f_n, 0) \leq \sum_{k=1}^{\infty} d\left(\frac{k}{2^n} I_{A_{k,n}}, 0\right) = 0.$$

Since $f_n(t) \to f(t)$ pointwise, we have $d(f, 0) = 0$.

Now let $\{f_n\}$ converge to 0 in measure μ. Suppose that the numbers $d(f_n, 0)$ do not converge to zero. Passing to a subsequence, we can assume that $d(f_n, 0) \geq \varepsilon > 0$ for all n and that $\{f_n\}$ converges to zero μ-almost everywhere. Let

$$A = \left\{ t\colon \{f_n(t)\} \quad \text{does not converge to zero} \right\}.$$

Note that $A \in \mathcal{B}$. Redefining f_n by zero on A we obtain functions g_n pointwise converging to zero such that $d(f_n, g_n) = 0$. Hence $d(g_n, 0) \to 0$, which yields that $d(f_n, 0) \to 0$, which is a contradiction.

Conversely, suppose that $d(f_n, 0) \to 0$. Replacing f by $|f|$ we can assume that $f_n \geq 0$. Fix $N \in \mathbb{N}$ and set $A_n = \{t\colon f_n(t) \geq 1/N\}$. We have to show that $\mu(A_n) \to 0$. First we observe that by the triangle inequality, the translation invariance and monotonicity of d we have

$$d(I_{A_n}, 0) \leq N d\left(\frac{1}{N} I_{A_n}, 0\right) \leq N d(f_n, 0).$$

It follows that $d(I_{A_n}, 0) \to 0$. Suppose now that $\{\mu(A_n)\}$ does not converge to zero. Passing to a subsequence, we can assume that $\mu(A_n) \geq \varepsilon$ and $d(I_{A_n}, 0) \leq 2^{-n}$. Set $B_n = \bigcup_{k \geq n} A_k$ and $B = \bigcap_n B_n$. Then $\mu(B_n) \geq \mu(A_n) \geq \varepsilon$ and $d(I_{B_n}, 0) \leq 2^{1-n}$. It follows that $\mu(B) \geq \varepsilon$, and since $I_{B_n} \to I_B$ pointwise, we obtain $d(I_B, 0) = 0$. The equality $d(I_B, 0) = 0$ implies that $\mu(B) = 0$ (as noted above), which is a contradiction. $\qquad\square$

The condition in this theorem that μ is finite is not necessary. For example, the semimetric

$$d_2(f, g) = \left(\int_0^1 \min(|f(t) - g(t)|^2, 1) \, dt \right)^{1/2}$$

on the space of Borel functions on $[0, 1]$ is equivalent to the semimetric of convergence in Lebesgue measure, but $\mu([0, 1]) = +\infty$, since for every n we can divide $[0, 1]$ into intervals of length $1/n$, and for such an interval A_i we have $d(I_{A_i}, 0) = 1/\sqrt{n}$. Certainly, the hypothesis of the theorem could be replaced with the weaker one that there is an equivalent metric satisfying our assumption, which will be even a necessary condition as well (but such a condition is not interesting). However, we do not know whether the additional restrictions on d in this theorem enable one to replace d with an equivalent metric for which μ would be finite.

References

1. A. N. Baushev, *On the weak convergence of probability measures in Orlicz spaces*, Theory Probab. Appl. **40** (1995), no. 3, 420–429.
2. A. N. Baushev, *On the weak convergence of probability measures in a Banach space*, J. Math. Sci. **109** (2002), no. 6, 2037–2046.
3. P. Billingsley, *Convergence of probability measures*, Wiley, New York, 1968.
4. V. I. Bogachev, *Measure theory. V. 1, 2*, Springer, New York, 2007.
5. V. I. Bogachev and A. V. Kolesnikov, *The Monge–Kantorovich problem: achievements, connections, and perspectives*, Uspehi Matem. Nauk. **67** (2012), no. 5, 3–110 (in Russian); English transl.: Russian Math. Surveys. **67** (2012), no. 5, 785–890.
6. V. I. Bogachev, N. V. Krylov, M. Röckner and S. V. Shaposhnikov, *Fokker–Planck–Kolmogorov equations*, Amer. Math. Soc., Providence, Rhode Island, 2015.
7. A. A. Borovkov, *Convergence of measures and random processes*, Uspehi Mat. Nauk. **31** (1976), no. 2, 3–68 (in Russian); English transl.: Russian Math. Surveys. **31** (1976), no. 2, 1–69.
8. A. A. Borovkov and A. I. Sakhanenko, *Remarks on convergence of random processes in non-separable metric spaces and on the non-existence of a Borel measure for processes in $C(0, \infty)$*, Teor. Veroyatn. i Primen. **18** (1973), no. 4, 812–815 (in Russian); English transl.: Theory Probab. **18** (1973), no. 4.
9. H. Cremers and D. Kadelka, *On weak convergence of stochastic processes with Lusin path spaces*, Manuscripta Math. **45** (1984), 115–125.
10. H. Cremers and D. Kadelka, *On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in L_p^E*, Stoch. Processes Appl. **21** (1986), 305–317.
11. R. M. Dudley, *Real analysis and probability*, Wadsworth & Brooks, Pacific Grove, California, 1989.
12. R. Fortet and E. Mourier, *Convergence de la répartition empirique vers la répartition théorique*, Ann. Sci. l'École Norm. Sup. **70** (1953), no. 3, 267–285.
13. I. I. Gikhman and A. V. Skorokhod, *The theory of stochastic processes. V. 1. (Translated from the Russian)*, Springer-Verlag, New York – Heidelberg, 1974 (Russian ed.: Moscow, 1971).
14. L. Š. Grinblat, *A limit theorem for measurable random processes and its applications*, Proc. Amer. Math. Soc. **61** (1976), 371–376.
15. L. Š. Grinblat, *Convergence of probability measures on separable Banach spaces*, Proc. Amer. Math. Soc. **67** (1977), 321–323.
16. L. Š. Grinblat, *Convergence of measurable random functions*, Proc. Amer. Math. Soc. **74** (1979), 322–325.
17. A. V. Ivanov, *Convergence of distributions of functionals of measurable random fields*, Ukrainskii Matem. Zhurn. **32** (1980), no. 1, 27–34 (in Russian); English tranl.: Ukrainian Math. J. **32** (1980), no. 1, 19–25.
18. L. V. Kantorovich and G. Sh. Rubinstein, *On a space of completely additive functions*, Vestnik Leningrad. Univ. **13** (1958), no. 7, 52–59 (in Russian).
19. L. Le Cam, *Convergence in distribution of stochastic processes*, Univ. Calif. Publ. Statist. **2** (1957), 207–236.
20. P. A. Meyer and W. A. Zheng, *Tightness criteria for laws of semi-martingales*, Annales l'Inst. H. Poincaré **20** (1984), no. 4., 353–372.

21. R. P. Pakshirajan, *A note on the weak convergence of probability measures in the D*[0, 1] *space*, Statist. Probab. Lett. **78** (2008), no. 6., 716–719.
22. D. Pollard, *Convergence of stochastic processes*, Springer, Berlin – New York, 1984.
23. R. Rebolledo, *Topologie faible et méta-stabilité*, Séminaire de probabilités XXI, Lecture Notes in Math. **1247** (1987), 545–562.
24. H. Sadi, *Une condition nécessaire et suffisante pour la convergence en pseudo-loi des processus*, Séminaire de probabilités XXII, Lecture Notes in Math. **1321** (1988), 434–437.
25. H. Tsukahara *On the convergence of measurable processes and prediction processes*, Illinois J. Math. **51** (2007), no. 4, 1231–1242.
26. A. W. van der Vaart and J. A. Wellner, *Weak convergence and empirical processes. With applications to statistics.*, Springer-Verlag, New York, 1996.

Department of Mechanics and Mathematics, Moscow State University, Moscow, Russia

St.-Tikhon's Orthodox Humanitarian University, Moscow, Russia

National Research University Higher School of Economics, Moscow, Russia