AISM (2008) 60:345–365
DOI 10.1007/s10463-006-0090-4

On weak convergence of random fields

Youri Davydov · Ričardas Zitikis

Received: 7 September 2005 / Revised: 14 April 2006 / Published online: 16 December 2006

Abstract We suggest simple and easily verifiable, yet general, conditions under which multi-parameter stochastic processes converge weakly to a continuous stochastic process. Connections to, and extensions of, R. Dudley's results play an important role in our considerations, and we therefore discuss them in detail. As an illustration of general results, we consider multi-parameter stochastic processes that can be decomposed into differences of two coordinate-wise non-decreasing processes, in which case the aforementioned conditions become even simpler. To illustrate how the herein developed general approach can be used in specific situations, we present a detailed analysis of a two-parameter sequential empirical process.

Keywords Stochastic processes · Random fields · Empirical processes · Weak convergence · Skorokhod spaces

1 Introduction

There are numerous problems where researchers encounter stochastic processes and need to establish their weak convergence. Many textbooks and monographs consider the topic, and the long list includes (subjectively selected):

Y. Davydov
Laboratoire Paul Painlevé, UMR 8524, Statistique et Probabilités, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France
e-mail: Youri.Davydov@univ-lille1.fr

R. Zitikis (✉)
Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON, Canada N6A 5B7
e-mail: zitikis@stats.uwo.ca

Billingsley (1968), Hall and Heyde (1980), Pollard (1984), Shorack and Wellner (1986), Csörgő and Horváth (1993), van der Vaart and Wellner (1996), Dudley (1999), and Korolyuk et al. (2000). Recent developments in the theory and applications of empirical processes in the context of econometric time series and, more generally, dependent observations have demonstrated that verifying standard weak convergence assumptions can be a laborious task, as seen, for example, from Dehling et al. (2002), Koul (2002), Doukhan et al. (2003), and references therein.

When considering weak convergence of stochastic processes defined on the real line and, particularly, on the unit interval $[0,1]$, Davydov (1996) suggested a new set of simple, yet general, conditions that have proved to be particularly useful in applications, especially if processes are discontinuous but converge to a continuous process. Furthermore, Davydov (1996) presented arguments showing that there are even simpler conditions if the processes—whose weak convergence we want to establish—can be decomposed into differences of two non-decreasing processes. Such processes make up an important class of stochastic processes.

Indeed, in many problems stochastic processes are discontinuous but converge to continuous processes, and it also frequently happens that the initial (discontinuous or continuous) stochastic processes are differences of two non-decreasing processes. Without overloading the current discussion with complexities, we note as a convenient example that the uniform empirical process $e_n(t) := \sqrt{n}\,(E_n(t)-t)$ is indeed the difference of two non-decreasing processes, $e_n^{\circ}(t) := \sqrt{n}\; E_n(t)$ and $e_n^{*}(t) := \sqrt{n}\; t$, where E_n is the empirical distribution function based on uniformly on $[0,1]$ distributed random variables $U_1,\ldots,U_n$. If a more complex example is desired, then we refer to the Lorenz process that has been thoroughly investigated by Goldie (1977). For further examples and references, we refer to the survey paper by Davydov and Zitikis (2004). We shall next elaborate—in general terms—on applications which suggest examples of stochastic processes that can be decomposed into differences of two non-decreasing processes, which is an important subclass of processes, covered by the main results of the present paper.

We have already noted several references in the area of econometric time series where classical and sequential empirical processes appear with independent and dependent observations. Other examples of stochastic processes with the aforementioned decomposition property (i.e., difference of two non-decreasing processes) arise, for example, when estimating (convex) Lorenz curves, monotonic mean residual life and failure rate functions, other population functions of interest in Actuarial, Engineering, Medical, and Social Sciences. Indeed, when estimating monotonic functions, researchers usually aim at constructing empirical estimators that satisfy same monotonicity properties as the corresponding population functions. Hence, the appropriately normalized differences between the estimators and the corresponding population functions are stochastic processes satisfying the above decomposition property.

There are also many problems where researchers want to establish weak convergence of *multi*-parameter stochastic processes. One of the simplest examples is the sequential (uniform) empirical process

$$e_n(t,u) := \frac{1}{\sqrt{n}} \sum_{i=1}^{[nu]} \left(\mathbf{1}_{(-\infty,t]}(U_i) - t\right).$$

The process can be written as the difference of two coordinate-wise non-decreasing processes:

$$e_n^{\circ}(t,u) := \frac{1}{\sqrt{n}} \sum_{i=1}^{[nu]} \mathbf{1}_{(-\infty,t]}(U_i) \quad \text{and} \quad e_n^{*}(t,u) := \frac{1}{\sqrt{n}}[nu]t.$$

Results by Davydov (1996) are not applicable in the current situation, because the domain of definition of the sequential empirical process is *two*-dimensional. This and other numerous examples show that it is useful to have the aforementioned Davydov's (1996) results extended to *multi*-parameter stochastic processes. We shall see below that such extensions are far from trivial.

The paper is organized as follows. In Sect. 2 we formulate our main result, which is Theorem 1, for general multi-parameter stochastic processes. Then we present an important corollary to the theorem, cf. Corollary 1, covering the class of stochastic processes that can be decomposed into differences of two coordinate-wise non-decreasing stochastic processes. As an illustration, at the end of Sect. 2 we present a detailed analysis of the aforementioned sequential empirical process. The proofs of Theorem 1 and Corollary 1 are given in Sect. 3. We conclude the paper with Sect. 4, which contains further notes on results of the present paper and their relationship to R. Dudley's contributions.

2 Main results

Let $B([0,1]^d;\mathbf{R})$ be the set of bounded, real-valued, and measurable functions defined on the d-dimensional cube $[0,1]^d$. Let $C([0,1]^d;\mathbf{R})$ be the set of all continuous functions. Furthermore, let D be a subset of $B([0,1]^d;\mathbf{R})$ such that, first, it contains $C([0,1]^d;\mathbf{R})$ and, second, the supremum norm of every function in D is determined by the supremum of that function over a countable subset of $[0,1]^d$. Furthermore, assume that there is a metric that makes D a complete separable space whose topology is weaker than the topology of uniform convergence. Examples of the space D with various metrics/topologies that make it complete and separable can be found in a number of works (cf., e.g., Dudley, 1999, and references therein). For comments on weak convergence in non-separable spaces and how the results of the present paper can be adjusted to become valid beyond the Skorokhod space $D([0,1]^d;\mathbf{R})$, we refer to Sect. 4. At this moment we only note that the sequential empirical process e_n can be considered as an element of D. Finally, let ω_f denote the modulus of continuity of function $f : [0,1]^d \to \mathbf{R}$, that is,

$$\omega_f(a_n) := \sup_{\|\boldsymbol{s}-\boldsymbol{t}\|\le a_n} |f(\boldsymbol{t}) - f(\boldsymbol{s})|\,,$$

where the norm $\|\cdot\|$ can, in principle, be any but we shall find it technically more convenient to work with $\|\boldsymbol{u}\| := \max_{1\le i\le d} |u_i|$. Now we are ready to formulate the main result of the present paper.

Theorem 1 *Let $\xi_n, n \ge 1$, be stochastic processes defined on $[0,1]^d$, taking values in $\mathbf{R}$, and whose paths are in the space D almost surely. Furthermore, let all the finite dimensional distributions of ξ_n converge to the corresponding ones of a process ξ. Assume that there are constants $\alpha \ge \beta > d, c \in (0,\infty)$, and $a_n \downarrow 0$ such that, for all $n \ge 1$, we have $\mathbf{E}(|\xi_n(\mathbf{0})|^\alpha) \le c$ and*

$$\mathbf{E}\left(|\xi_n(\boldsymbol{t}) - \xi_n(\boldsymbol{s})|^\alpha\right) \le c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta \quad \textit{whenever} \ \ \|\boldsymbol{t}-\boldsymbol{s}\| \ge a_n. \tag{1}$$

Furthermore, assume that, when $n \to \infty$,

$$\omega_{\xi_n}(a_n) \to_{\mathbf{P}} 0. \tag{2}$$

Then the processes ξ_n converge weakly to ξ, and the limiting process ξ has continuous paths almost surely.

A few notes concerning the above theorem follow.

(1) Theorem 1 generalizes a result by Davydov (1996) that was proved in the case $d = 1$.
(2) Theorem 1 holds for processes whose paths are in the space $B([0,1]^d;\mathbf{R})$ equipped with the uniform topology. However, it should be kept in mind that serious issues might emerge in this context due to the fact that the space is not separable. This is, indeed, the reason why in the theorem above we have restricted ourselves to a subset of $B([0,1]^d;\mathbf{R})$ equipped with a topology that is weaker than the topology of uniform convergence and makes D a separable complete metric space. We refer to Sect. 4 for further notes on the subject.
(3) Theorem 1 also holds for processes with values in a complete separable metric space, with metric r. In this case, condition (1) becomes $\mathbf{E}(r(\xi_n(\boldsymbol{t}),\xi_n(\boldsymbol{s}))^\alpha) \le c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta$ whenever $\|\boldsymbol{t}-\boldsymbol{s}\| \ge a_n$.

We shall now formulate a corollary to Theorem 1 that we find particularly useful in applications. Let $k_n := [1/a_n]$, and let $\boldsymbol{j} := (j_1,\ldots,j_d)$, where each of the d coordinates can be any numbers $0,1,\ldots,k_n+1$. If a coordinate of the vector $\boldsymbol{j}$ takes on any of the values $l = 0,1,\ldots,k_n$, then, by definition, the corresponding coordinate of the vector $\boldsymbol{t}(\boldsymbol{j}) := (t_{j_1},\ldots,t_{j_d})$ equals la_n. If, however, a coordinate of $\boldsymbol{j}$ takes on the value $l = k_n+1$, then, by definition, the corresponding coordinate of $\boldsymbol{t}(\boldsymbol{j})$ equals 1. Given $\boldsymbol{j}$, let $\boldsymbol{j}^\triangle$ denote any d-dimensional element such that $\|\boldsymbol{j}^\triangle - \boldsymbol{j}\| = 1$ and all the coordinates of $\boldsymbol{j}^\triangle$ except only one coincide with those of $\boldsymbol{j}$.

Corollary 1 *Let all the assumptions of Theorem 1 be satisfied except (2), which is now replaced by the following more easily verifiable condition: assume that the process ξ_n can be written as the difference of two coordinate-wise non-decreasing processes ξ_n° and ξ_n^*, and let ξ_n^* be such that*

$$\sup_{\boldsymbol{j},\boldsymbol{j}^\triangle} \left| \xi_n^*(\boldsymbol{t}(\boldsymbol{j}^\triangle)) - \xi_n^*(\boldsymbol{t}(\boldsymbol{j})) \right| \to_{\mathbf{P}} 0 \tag{3}$$

when $n \to \infty$, where the supremum is taken over all $\boldsymbol{j} = (j_1, \dots, j_d)$ whose coordinates take on the values $0, 1, \dots, k_n$. Then the processes ξ_n converge weakly to ξ, and the limiting process ξ has continuous paths almost surely.

Corollary 1 generalizes a result by Davydov (1996) that was proved in the case $d = 1$.

To illustrate Corollary 1, we consider the earlier defined sequential empirical process. We have already noted that $e_n(t, u)$ can be written as the difference of $e_n^\circ(t, u)$ and $e_n^*(t, u)$, which are coordinate-wise non-decreasing processes. Writing $\boldsymbol{t}(\boldsymbol{j}^\triangle) = (t_1, t_2)$ and $\boldsymbol{t}(\boldsymbol{j}) = (s_1, s_2)$ for notational simplicity, we start verifying assumption (3) with the bound:

$$\left| e_n^*(t_1, t_2) - e_n^*(s_1, s_2) \right| \le \frac{1}{\sqrt{n}} \big(n|t_1 - s_1| + n|t_2 - s_2| + 1 \big). \tag{4}$$

The right-hand side of (4) does not exceed $c/\sqrt{n}$ since both $|t_1 - s_1|$ and $|t_2 - s_2|$ do not exceed a_n, which equals $1/n$ by definition. Next we verify assumption (1) with $\alpha = 6$. (Choosing $\alpha = 4$ is not sufficient since we would have $\beta = 2$ whereas $\beta > 2$ is needed in the two-parameter case.) We have

$$\begin{aligned} \mathbf{E}\left(|e_n(t_1, t_2) - e_n(t_1, s_2)|^6 \right) &= \frac{1}{n^3} \mathbf{E}\left(\left| \sum_{i=[ns_2]+1}^{[nt_2]} \big(\mathbf{1}_{(-\infty, t_1]}(U_i) - t_1 \big) \right|^6 \right) \\ &\le c|t_2 - s_2|^3 + \frac{1}{n^3}. \end{aligned}$$

Next, we obtain

$$\begin{aligned} &\mathbf{E}\left(|e_n(t_1, s_2) - e_n(s_1, s_2)|^6 \right) \\ &\quad = \frac{1}{n^3} \mathbf{E}\left(\left| \sum_{i=1}^{[ns_2]} \left\{ \big(\mathbf{1}_{(-\infty, t_1]}(U_i) - t_1 \big) - \big(\mathbf{1}_{(-\infty, s_1]}(U_i) - s_1 \big) \right\} \right|^6 \right) \\ &\quad \le c|t_1 - s_1|^3 + \frac{c}{n}|t_1 - s_1|^2 + \frac{c}{n^2}|t_1 - s_1|. \end{aligned}$$

The two bounds above and the assumption $\|\boldsymbol{t}-\boldsymbol{s}\| \geq a_n$ with $a_n = 1/n$ imply that the expectation $\mathbf{E}(|e_n(t_1,t_2) - e_n(s_1,s_2)|^6)$ does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^3$. This completes the verification of assumption (1) with $\alpha = 6$ and $\beta = 3$.

There are monographs and numerous articles that investigate (or use) weak convergence of *multi*-parameter stochastic processes. So far we have deliberately constrained ourselves to mainly referring to monographs, due to space consideration. On this occasion we shall also obey this rule and refer to the monographs by Ivanoff and Merzbach (2000), Khoshnevisan (2002), and to the list of references therein. We also emphasize that the present article is not about which assumptions are better to use (or not to use) when verifying weak convergence. This article is just a mere attempt to present a set of simple and readily in practice verifiable assumptions that imply weak convergence of multi-parameter stochastic processes.

3 Proofs of Theorem 1 and Corollary 1

3.1 Idea of the proof

Condition (1) is similar to a well known Kolmogorov's condition (cf. (5)) that assures relative compactness in the context of continuous stochastic processes. Among numerous results generalizing Kolmogorov's condition to random fields, in the monograph by Ibragimov and Has'minskii (1981, Theorem 20, p. 378) we find the following result (reformulated here in a less general form) particularly useful in the context of the present paper.

Theorem 2 (Ibragimov and Has'minskii, 1981) *Let η_n be continuous stochastic processes defined on the d-dimensional cube $[0,1]^d$. Furthermore, let all the finite dimensional distributions of η_n converge to the corresponding ones of a process η. Assume that there are constants $\alpha \geq \beta > d$ and $c \in (0,\infty)$ such that, for all $n \geq 1$, we have $\mathbf{E}(|\eta_n(\mathbf{0})|^\alpha) \leq c$ and*

$$\mathbf{E}\left(|\eta_n(\boldsymbol{t}) - \eta_n(\boldsymbol{s})|^\alpha\right) \leq c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta \quad \textit{for all } \boldsymbol{s},\boldsymbol{t} \in [0,1]^d. \tag{5}$$

Then the processes η_n converge weakly to η, and the limiting process η has continuous paths almost surely.

Since the processes ξ_n might not, in general, be continuous, Theorem 2 might not therefore be directly applicable to them. Nevertheless, Theorem 2 suggests the following route for proving weak convergence. Suppose that we have constructed continuous stochastic processes η_n such that they satisfy condition (5) and uniformly approximate (cf. the next paragraph for detail) the processes ξ_n when $n \to \infty$, that is,

$$\|\xi_n - \eta_n\|_\infty := \sup_{\boldsymbol{t}\in[0,1]^d} |\xi_n(\boldsymbol{t}) - \eta_n(\boldsymbol{t})| \to_{\mathbf{P}} 0. \tag{6}$$

By Theorem 2 the processes η_n converge weakly in $C([0,1]^d;\mathbf{R})$ to a process η, and so in view of (6) the processes ξ_n weakly converge to η in the above defined space D. This is exactly the result we aim at.

The supremum in (6) should, of course, be a random variable, so that convergence in probability would be well defined. This follows from these facts: the processes η_n are continuous; the processes ξ_n are elements of the space D; the supremums $\|\xi_n - \eta_n\|_\infty$ can be calculated as supremums over countable subsets of the cube $[0,1]^d$; and the latter supremums (over countable subsets) are random variables. Later in this paper (cf. Sect. 4) we shall discuss the σ-algebra $\mathcal{B}_0$ generated by open balls, and require that the processes ξ_n be measurable with respect to the σ-algebra. This will take care of the measurability problems related to the supremum norm $\|\cdot\|_\infty$ of the processes ξ_n.

3.2 Construction of η_n in d-dimensions

As we have noted above, the main idea of the proof relies on constructing continuous stochastic processes η_n that well approximate the processes ξ_n when n tends to infinity. We start constructing such processes η_n with an operator Δ_n defined on the set of rectangles $\Pi := \prod_{i=1}^d [u_i, v_i] \subseteq [0,1]^d$ by the formula

$$\begin{aligned}\Delta_n(\Pi) := {}& \xi_n(v_1,\ldots,v_d) - \sum_{1\le m\le d} \xi_n(v_1,\ldots,v_{m-1},u_m,v_{m+1},\ldots,v_d)\\ &+ \sum_{1\le p<q\le d} \xi_n(v_1,\ldots,v_{p-1},u_p,v_{p+1},\ldots,v_{q-1},u_q,v_{q+1},\ldots,v_d)\\ &- \cdots + (-1)^d \xi_n(u_1,\ldots,u_d).\end{aligned}$$

The quantity $\Delta_n(\Pi)$ is a dth-order difference. For example, when $d=1$, then it is the first order difference $\xi_n(v_1)-\xi_n(u_1)$, and when $d=2$, it is the difference between the first order differences $\xi_n(v_1,v_2)-\xi_n(u_1,v_2)$ and $\xi_n(v_1,u_2)-\xi_n(u_1,u_2)$.

The operator Δ_n associates each rectangle $R(\boldsymbol{j}) = \prod_{i=1}^d [t_{j_i}, t_{j_i+1}]$ with the random variable $\Delta_n(R(\boldsymbol{j}))$. Note that if, for an index set J, the rectangles $R(\boldsymbol{j}), \boldsymbol{j}\in J$, are such that their union $\bigcup_{\boldsymbol{j}\in J} R(\boldsymbol{j})$ is also a rectangle, say $\Pi = \prod_{i=1}^d [u_i, v_i]$, then the sum $\sum_{\boldsymbol{j}\in J} \Delta_n(R(\boldsymbol{j}))$ equals $\Delta_n(\Pi)$. While this property is obvious in two dimensions, in the general d-dimensional case it can be verified by induction.

We define processes σ_n on $[0,1]^d$, by the formula

$$\sigma_n(\boldsymbol{s}) := \sum_{\boldsymbol{j}} \Delta_n(R(\boldsymbol{j})) \frac{\mathbf{1}_{R(\boldsymbol{j})}(\boldsymbol{s})}{\lambda(R(\boldsymbol{j}))},$$

where the summation is taken over all $\boldsymbol{j}$ such that $R(\boldsymbol{j}) \subseteq [0,1]^d$, and $\mathbf{1}_{R(\boldsymbol{j})}$ and $\lambda(R(\boldsymbol{j}))$ denote, respectively, the indicator function and the d-dimensional Lebesgue measure of the rectangle $R(\boldsymbol{j})$. Using σ_n, we can now define the promised (continuous) stochastic processes η_n on $[0,1]^d$ by

$$\eta_n(\boldsymbol{t}) := \int_{[\boldsymbol{0},\boldsymbol{t}]} \sigma_n(\boldsymbol{s})\mathrm{d}\boldsymbol{s}, \tag{7}$$

where $[\boldsymbol{0},\boldsymbol{t}]$ denotes the rectangle $\prod_{i=1}^d [0,t_i]$. Since $\eta_n(\boldsymbol{0}) = \boldsymbol{0}$, the condition $\mathbf{E}(|\eta_n(\boldsymbol{0})|^\alpha) \le c$ of Theorem 2 is obviously satisfied. For every vertex $\boldsymbol{t}$ of every rectangle $R(\boldsymbol{k}) = \prod_{i=1}^d [t_{k_i}, t_{k_i+1}]$, we have that

$$\begin{aligned} \eta_n(\boldsymbol{t}) &= \sum_{\boldsymbol{j}} \Delta_n(R(\boldsymbol{j})) \frac{1}{\lambda(R(\boldsymbol{j}))} \int_{[\boldsymbol{0},\boldsymbol{t}]} \mathbf{1}_{R(\boldsymbol{j})}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \\ &= \sum_{\boldsymbol{j}:\, R(\boldsymbol{j})\subseteq[\boldsymbol{0},\boldsymbol{t}]} \Delta_n(R(\boldsymbol{j})) \\ &= \xi_n(\boldsymbol{t}), \end{aligned} \tag{8}$$

where the last equality in (8) holds due to the aforementioned addition property of the differences $\Delta_n(R(\boldsymbol{j}))$, provided that $\xi_n(\boldsymbol{t})$ vanish at every $\boldsymbol{t} \in [0,1]^d \setminus (0,1]^d$. Indeed, under the latter assumption, the difference $\Delta_n(\Pi)$ with $\Pi = [\boldsymbol{0},\boldsymbol{t}]$ equals $\xi_n(\boldsymbol{t})$. Hence, for some time in our considerations below, we assume that, for all n,

$$\xi_n(\boldsymbol{t}) = 0 \quad \text{for all } \boldsymbol{t} \in [0,1]^d \setminus (0,1]^d. \tag{9}$$

Later we shall remove property (9) and obtain Theorem 1 in its full generality.

3.3 Verification of (6) assuming (9)

Using the equality $\xi_n(\boldsymbol{t}(\boldsymbol{j})) = \eta_n(\boldsymbol{t}(\boldsymbol{j}))$ established in (8), we obtain the bound

$$\begin{aligned} \sup_{\boldsymbol{t}\in[0,1]^d} |\xi_n(\boldsymbol{t}) - \eta_n(\boldsymbol{t})| &= \sup_{\boldsymbol{j}} \sup_{\boldsymbol{t}\in R(\boldsymbol{j})} |\xi_n(\boldsymbol{t}) - \eta_n(\boldsymbol{t})| \\ &\le \sup_{\boldsymbol{j}} \sup_{\boldsymbol{t}\in R(\boldsymbol{j})} \left|\xi_n(\boldsymbol{t}) - \xi_n(\boldsymbol{t}(\boldsymbol{j}))\right| \\ &\quad + \sup_{\boldsymbol{j}} \sup_{\boldsymbol{t}\in R(\boldsymbol{j})} \left|\eta_n(\boldsymbol{t}(\boldsymbol{j})) - \eta_n(\boldsymbol{t})\right|. \end{aligned} \tag{10}$$

The first supremum on the right-hand side of (10) does not exceed $c\omega_{\xi_n}(a_n)$, which converges to 0 by assumption (2). We shall now show that the same statement is true for the second supremum on the right-hand side of (10).

For any $\boldsymbol{t} \in R(\boldsymbol{j})$, we write

$$\eta_n(\boldsymbol{t}) - \eta_n(\boldsymbol{t}(\boldsymbol{j})) = \int_{[\boldsymbol{0},\boldsymbol{t}]\setminus[\boldsymbol{0},\boldsymbol{t}(\boldsymbol{j})]} \sigma_n(\boldsymbol{s})\mathrm{d}\boldsymbol{s}. \tag{11}$$

Next, we express the set $[\mathbf{0}, \boldsymbol{t}] \setminus [\mathbf{0}, \boldsymbol{t}(\boldsymbol{j})]$ as the union of rectangles. For this, we first write $[\mathbf{0}, \boldsymbol{t}(\boldsymbol{j})]$ as the product $\prod_{i=1}^{d} T_i$ with $T_i := [0, t_{j_i}]$, and then write $[\mathbf{0}, \boldsymbol{t}]$ as the product $\prod_{i=1}^{d}(T_i \cup \Delta_i)$ with $\Delta_i := [t_{j_i}, t_i]$. Note that the product $\prod_{i=1}^{d}(T_i \cup \Delta_i)$ equals $\bigcup_{S \in \mathcal{F}_d}(\prod_{i \in S} T_i \times \prod_{j \notin S} \Delta_i)$ with $\mathcal{F}_d$ denoting the collection of all the subsets of $\{1, 2, \ldots, d\}$. Hence, we have the equality

$$[\mathbf{0}, \boldsymbol{t}] \setminus [\mathbf{0}, \boldsymbol{t}(\boldsymbol{j})] = \bigcup_{S \in \mathcal{F}_d^*} R_{\boldsymbol{t}}(S), \tag{12}$$

where $R_{\boldsymbol{t}}(S) := \prod_{i \in S} T_i \times \prod_{j \notin S} \Delta_i$ and $\mathcal{F}_d^*$ is the collection $\mathcal{F}_d$ without the largest set $\{1, 2, \ldots, d\}$. Using representation (12) on the right-hand side of (11), we obtain the equalities

$$\begin{aligned} \eta_n(\boldsymbol{t}) - \eta_n(\boldsymbol{t}(\boldsymbol{j})) &= \sum_{S \in \mathcal{F}_d^*} \int_{R_{\boldsymbol{t}}(S)} \sigma_n(\boldsymbol{s}) \mathrm{d}\boldsymbol{s} \\ &= \sum_{S \in \mathcal{F}_d^*} \sum_{\boldsymbol{l}} \Delta_n(R(\boldsymbol{l})) \frac{\lambda(R_{\boldsymbol{t}}(S) \cap R(\boldsymbol{l}))}{\lambda(R(\boldsymbol{l}))} \\ &= \sum_{S \in \mathcal{F}_d^*} \sum_{\boldsymbol{l}:\, R_{\boldsymbol{t}}(S) \cap R(\boldsymbol{l}) \neq \emptyset} \Delta_n(R(\boldsymbol{l})) \frac{\lambda(R_{\boldsymbol{t}}(S) \cap R(\boldsymbol{l}))}{\lambda(R(\boldsymbol{l}))}. \end{aligned} \tag{13}$$

For any fixed $S \in \mathcal{F}_d^*$, the ratios on the right-hand side of (13) are same for all indices $\boldsymbol{l}$ such that $R_{\boldsymbol{t}}(S) \cap R(\boldsymbol{l}) \neq \emptyset$, and we therefore denote the ratios by $c_{\boldsymbol{t}}(S)$, which is independent of $\boldsymbol{l}$. Hence, from (13) we obtain the equality

$$\eta_n(\boldsymbol{t}) - \eta_n(\boldsymbol{t}(\boldsymbol{j})) = \sum_{S \in \mathcal{F}_d^*} c_{\boldsymbol{t}}(S) \sum_{\boldsymbol{l}:\, R_{\boldsymbol{t}}(S) \cap R(\boldsymbol{l}) \neq \emptyset} \Delta_n(R(\boldsymbol{l})). \tag{14}$$

Using assumption (9) and the addition property of the operator Δ_n, we rewrite the sum $\sum_{\boldsymbol{l}: R_{\boldsymbol{t}}(S) \cap R(\boldsymbol{l}) \neq \emptyset} \Delta_n(R(\boldsymbol{l}))$ in the form of a sum $\sum(\xi_n(\boldsymbol{u}) - \xi_n(\boldsymbol{v}))$, where the summation is taken over neighbouring vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ of a rectangle in the $(d-1)$-dimensional space whose edges have lengths not exceeding a_n. The number of the summands in $\sum(\xi_n(\boldsymbol{u}) - \xi_n(\boldsymbol{v}))$ does not exceed a constant that depends only on the dimension d. Hence, we estimate the absolute value of the sum $\sum(\xi_n(\boldsymbol{u}) - \xi_n(\boldsymbol{v}))$ by $c\omega_{\xi_n}(a_n)$, where the constant c depends only on d. The quantities $c_{\boldsymbol{t}}(S)$ do not exceed 1, as it follows from their definition. Hence, the left-hand side of (14) does not exceed $c\omega_{\xi_n}(a_n)$, which implies that the second supremum on the right-hand side of (10) does not exceed $c\omega_{\xi_n}(a_n)$. This finishes the proof of statement (6) under assumption (9).

3.4 Proof that $\omega_{\xi_n}(a_n) \to_{\mathbf{P}} 0$ under the assumptions of Corollary 1

The statement $\omega_{\xi_n}(a_n) \to_{\mathbf{P}} 0$ follows from

$$\Omega_{\xi_n}(a_n) := \sup_{\boldsymbol{j}} \sup_{\boldsymbol{t}\in R(\boldsymbol{j})} |\xi_n(\boldsymbol{t}) - \xi_n(\boldsymbol{t}(\boldsymbol{j}))| \to_{\mathbf{P}} 0. \tag{15}$$

To prove statement (15), we proceed as follows. Since ξ_n can be expressed as the difference of the two coordinate-wise non-decreasing processes ξ_n° and ξ_n^*, we have that

$$\begin{aligned}\Omega_{\xi_n}(a_n) &\le \sup_{\boldsymbol{j}} \left|\xi_n^\circ(\boldsymbol{t}(\boldsymbol{j}+1)) - \xi_n^\circ(\boldsymbol{t}(\boldsymbol{j}))\right| + \sup_{\boldsymbol{j}} \left|\xi_n^*(\boldsymbol{t}(\boldsymbol{j}+1)) - \xi_n^*(\boldsymbol{t}(\boldsymbol{j}))\right| \\ &\le \sup_{\boldsymbol{j}} \left|\xi_n(\boldsymbol{t}(\boldsymbol{j}+1)) - \xi_n(\boldsymbol{t}(\boldsymbol{j}))\right| + 2\sup_{\boldsymbol{j}} \left|\xi_n^*(\boldsymbol{t}(\boldsymbol{j}+1)) - \xi_n^*(\boldsymbol{t}(\boldsymbol{j}))\right|, \end{aligned} \tag{16}$$

where $\boldsymbol{t}(\boldsymbol{j}+1) := (t_{j_1+1}, \ldots, t_{j_d+1})$. By assumption (2), the second supremum on the right-hand side of (16) converges to 0 in probability. To show that the first supremum on the right-hand side of (16) also converges to 0 in probability, for any fixed $\epsilon > 0$ we write the bounds:

$$\begin{aligned}\mathbf{P}\left\{\sup_{\boldsymbol{j}} \left|\xi_n(\boldsymbol{t}(\boldsymbol{j}+1)) - \xi_n(\boldsymbol{t}(\boldsymbol{j}))\right| > \epsilon\right\} &\le \frac{1}{\epsilon^\alpha}\sum_{\boldsymbol{j}} \mathbf{E}\left(\left|\xi_n(\boldsymbol{t}(\boldsymbol{j}+1)) - \xi_n(\boldsymbol{t}(\boldsymbol{j}))\right|^\alpha\right) \\ &\le \frac{c}{\epsilon^\alpha}\sum_{\boldsymbol{j}} \|\boldsymbol{t}(\boldsymbol{j}+1) - \boldsymbol{t}(\boldsymbol{j})\|^\beta, \end{aligned} \tag{17}$$

where the right-most bound in (17) follows from assumption (1). Note that the norm $\|\boldsymbol{t}(\boldsymbol{j}+1) - \boldsymbol{t}(\boldsymbol{j})\|$ does not exceed a_n. Furthermore, there are at most ca_n^{-d} indices $\boldsymbol{j}$ in the sum in (17). Hence, the right-hand side of (17) does not exceed $ca_n^{\beta-d}$ for a constant c that depends only on d. Since $\beta > d$ and $a_n \downarrow 0$ by assumption, the right-hand side of (17) converges to 0 when $n \to \infty$. Hence, we have proved statement (15) and thus, in turn, $\omega_{\xi_n}(a_n) \to_{\mathbf{P}} 0$. Notice that we have not used assumption (9) in this subsection.

3.5 Verification of (5) assuming (9)

In view of (6) we have reduced weak convergence of stochastic processes ξ_n to weak convergence of the *continuous* processes η_n, which are defined by Eq. (7). To establish weak convergence of the processes η_n, we apply Theorem 2. For this, we need to check assumption (5), which we do by separately considering the following three cases:

(1) Both points $\boldsymbol{s}$ and $\boldsymbol{t}$ are in *same* rectangle.
(2) The points $\boldsymbol{s}$ and $\boldsymbol{t}$ are in different rectangles that have *non-empty* intersection.

(3) The points $\boldsymbol{s}$ and $\boldsymbol{t}$ are in different rectangles that do *not* intersect.

We shall first show that the second and third cases follow from the first one, and we shall then verify assumption (5) in the first case. Hence, when considering the second and third cases, we assume the validity of (5) in the first case.

We start with the second case, that is, when both $\boldsymbol{s}$ and $\boldsymbol{t}$ are in different rectangles that have at least one common point. The straight line connecting the points $\boldsymbol{s}$ and $\boldsymbol{t}$ can be subdivided into a finite number of segments whose interiors lie in different rectangles $R(\cdot)$. Hence, we write

$$[\boldsymbol{s},\boldsymbol{t}] = \bigcup_{r=1}^{r_0}[\boldsymbol{t}_{r-1},\boldsymbol{t}_r], \tag{18}$$

where $\boldsymbol{t}_0 := \boldsymbol{s}$, $\boldsymbol{t}_{r_0} := \boldsymbol{t}$, and every interval $[\boldsymbol{t}_{r-1},\boldsymbol{t}_r]$ is of the form $[\boldsymbol{s},\boldsymbol{t}]\cap R(\boldsymbol{j})$ with a different $\boldsymbol{j}$. The number r_0 in (18) does not exceed a constant that depends only on the dimension d. We now rewrite the expectation $\mathbf{E}(|\eta_n(\boldsymbol{t}) - \eta_n(\boldsymbol{s})|^{\alpha})$ as $\mathbf{E}(|\eta_n(\boldsymbol{t}_{r_0}) - \eta_n(\boldsymbol{t}_0)|^{\alpha})$ and then note that the latter expectation does not exceed $c\sum_{r=1}^{r_0}\|\boldsymbol{t}_r - \boldsymbol{t}_{r-1}\|^{\beta}$ due to the assumed validity of (5) in case (1). The latter sum does not, in turn, exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^{\beta}$. This finishes the proof of (5) in case (2).

We next consider the third case, that is, when the two points $\boldsymbol{s}$ and $\boldsymbol{t}$ are in different rectangles that do not have any common point. Let the indices $\boldsymbol{j}$ and $\boldsymbol{k}$ be such that $\boldsymbol{s}\in R(\boldsymbol{j})$ and $\boldsymbol{t}\in R(\boldsymbol{k})$. We write the bound

$$\begin{aligned}\mathbf{E}(|\eta_n(\boldsymbol{t}) - \eta_n(\boldsymbol{s})|^{\alpha}) &\le c\mathbf{E}(|\eta_n(\boldsymbol{t}) - \eta_n(\boldsymbol{t}(\boldsymbol{k}))|^{\alpha})\\ &\quad + c\mathbf{E}(|\eta_n(\boldsymbol{t}(\boldsymbol{k}))-\eta_n(\boldsymbol{t}(\boldsymbol{j}))|^{\alpha}) + c\mathbf{E}(|\eta_n(\boldsymbol{t}(\boldsymbol{j}))-\eta_n(\boldsymbol{s})|^{\alpha}).\end{aligned} \tag{19}$$

Since we have assumed the validity of (5) in case (1), the first and the third expectations on the right-hand side of (19) do not exceed, respectively, $c\|\boldsymbol{t}-\boldsymbol{t}(\boldsymbol{k})\|^{\beta}$ and $c\|\boldsymbol{t}(\boldsymbol{j})-\boldsymbol{s}\|^{\beta}$. The two norms do not exceed ca_n^{β}, which in turn does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^{\beta}$, since $\|\boldsymbol{t}-\boldsymbol{s}\| \ge a_n$ due to the assumption that $\boldsymbol{s}$ and $\boldsymbol{t}$ are in rectangles that do not have any common point. The second expectation on the right-hand side of (19) equals $\mathbf{E}(|\xi_n(\boldsymbol{t}(\boldsymbol{k})) - \xi_n(\boldsymbol{t}(\boldsymbol{j}))|^{\alpha})$ and hence does not exceed $c\|\boldsymbol{t}(\boldsymbol{k}) - \boldsymbol{t}(\boldsymbol{j})\|^{\beta}$ due to (1). The norm $\|\boldsymbol{t}(\boldsymbol{k}) - \boldsymbol{t}(\boldsymbol{j})\|$ does not exceed $\|\boldsymbol{t}-\boldsymbol{s}\| + 2a_n$, which in turn does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|$ since $\|\boldsymbol{t}-\boldsymbol{s}\| \ge a_n$. Summarising the findings above, all the three expectations on the right-hand side of bound (19) do not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^{\beta}$. This concludes the proof of (5) in the third case.

We now consider the first case, that is, when both $\boldsymbol{s}$ and $\boldsymbol{t}$ are in same rectangle, say in $R(\boldsymbol{j})$. To simplify the proof we assume without loss of generality that all the coordinates of $\boldsymbol{s}$ except only one are equal to the corresponding coordinates of $\boldsymbol{t}$. To simplify the proof even further, we assume that only the last coordinate of $\boldsymbol{s}$ is different from the corresponding one of $\boldsymbol{t}$. (Other cases can be investigated similarly.) Thus, we write $\boldsymbol{s} = (s_1,\ldots,s_{d-1},s_d)$ and $\boldsymbol{t} = (s_1,\ldots,s_{d-1},t_d)$. Finally, without loss of generality we assume that $s_d \le t_d$. After all this preparatory work, we write the equalities:

$$\begin{aligned}\eta_n(\boldsymbol{t})-\eta_n(\boldsymbol{s}) &= \int\limits_{[\mathbf{0},\boldsymbol{t}]\setminus[\mathbf{0},\boldsymbol{s}]} \sigma_n(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\\ &= \int\limits_{\bigcup_{i=1}^{d-1}[0,t_i]\times[s_d,t_d]} \sigma_n(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\\ &= \int\limits_{\bigcup_{i=1}^{d-1}(T_i\cup\Delta_i)\times[s_d,t_d]} \sigma_n(\boldsymbol{s})\mathrm{d}\boldsymbol{s}, \end{aligned} \tag{20}$$

where $T_i=[0,t_{j_i}]$ and $\Delta_i=[t_{j_i},t_i]$. The product $\prod_{i=1}^{d-1}(T_i\cup\Delta_i)$ equals $\bigcup_{S\in\mathcal{F}_{d-1}}(\prod_{i\in S}T_i\times\prod_{j\notin S}\Delta_i)$, where $\mathcal{F}_{d-1}$ denotes the collection of all the subsets of $\{1,2,\ldots,d-1\}$. With the already introduced and used notation $R_{\boldsymbol{t}}(S)$ for $\prod_{i\in S}T_i\times\prod_{j\notin S}\Delta_i$, we continue the string of equalities (20) and have

$$\begin{aligned}\eta_n(\boldsymbol{t})-\eta_n(\boldsymbol{s}) &= \sum_{S\in\mathcal{F}_{d-1}}\int\limits_{R_{\boldsymbol{t}}(S)\times[s_d,t_d]}\sigma_n(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\\ &= \sum_{S\in\mathcal{F}_{d-1}}\sum_{\boldsymbol{l}}\Delta_n(R(\boldsymbol{l}))\frac{\lambda(R_{\boldsymbol{t}}(S)\times[s_d,t_d]\cap R(\boldsymbol{l}))}{\lambda(R(\boldsymbol{l}))}. \end{aligned} \tag{21}$$

Given a fixed set $S\in\mathcal{F}_{d-1}$, and for all indices $\boldsymbol{l}$ such that $R_{\boldsymbol{t}}(S)\cap R(\boldsymbol{l})\neq\emptyset$, we have that the ratios $\lambda(R_{\boldsymbol{t}}(S)\times[s_d,t_d]\cap R(\boldsymbol{l}))/\lambda(R(\boldsymbol{l}))$ are all equal. We therefore denote them by $c_{\boldsymbol{t}}(S)$, which is independent of $\boldsymbol{l}$. In turn, for any $S\in\mathcal{F}_{d-1}$, we write the second sum (with respect to $\boldsymbol{l}$) in (21) as

$$c_{\boldsymbol{t}}(S)\sum_{\boldsymbol{l}:R_{\boldsymbol{t}}(S)\times[s_d,t_d]\cap R(\boldsymbol{l})\neq\emptyset}\Delta_n(R(\boldsymbol{l})). \tag{22}$$

Since $c_{\boldsymbol{t}}(S)$ does not exceed $|t_d-s_d|/a_n$, we have the bound $c_{\boldsymbol{t}}(S)\le\|\boldsymbol{t}-\boldsymbol{s}\|/a_n$. Next we note that the sum in (22) can be written in the form $\sum(\xi_n(\boldsymbol{u})-\xi_n(\boldsymbol{v}))$, where the summation is taken over some neighbouring vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ of a $(d-1)$-dimensional rectangle. The number of summands in $\sum(\xi_n(\boldsymbol{u})-\xi_n(\boldsymbol{v}))$ does not exceed a constant that depends only on d. Furthermore, $\mathbf{E}(|\xi_n(\boldsymbol{u})-\xi_n(\boldsymbol{v})|^\alpha)$ does not exceed $c\|\boldsymbol{u}-\boldsymbol{v}\|^\beta$ and the latter quantity does not exceed ca_n^β since the vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ are neighbouring. In view of the notes above, we have from (21) that

$$\mathbf{E}\left(|\eta_n(\boldsymbol{t})-\eta_n(\boldsymbol{s})|^\alpha\right)\le c\|\boldsymbol{t}-\boldsymbol{s}\|^\alpha a_n^{\beta-\alpha}. \tag{23}$$

The right-hand side of (23) does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta$ because $\alpha\ge\beta$ and $\|\boldsymbol{t}-\boldsymbol{s}\|\le a_n$ (the latter bound holds because both $\boldsymbol{t}$ and $\boldsymbol{s}$ are in same rectangle). The verification of (5) in the first case is now finished. This also completes the proof of Theorem 1 under assumption (9).

3.6 Introduction to the proof of Theorem 1 *without* assumption (9)

The idea of removing assumption (9) is based on first shrinking the domain of definition $[0,1]^d$ of the process ξ_n to, say $[1/2,1]^d$, and then extending the "shrunken" process to $[0,1]^d$ in such a way that the extension, say $\widetilde{\xi}_n$, would satisfy the assumption $\widetilde{\xi}_n(\boldsymbol{t}) = 0$ for all $\boldsymbol{t} \in [0,1]^d \setminus (0,1]^d$. We shall now implement the idea rigorously.

First we define the process $\widehat{\xi}_n$ by the formula

$$\widehat{\xi}_n(\boldsymbol{u}) := \xi_n(2\boldsymbol{u} - 1)$$

for every $\boldsymbol{u} \in [1/2,1]^d$. Then we define $D_\theta := \{\boldsymbol{t} \in [0,1]^d : \|\boldsymbol{t} - [1/2,1]^d\| = \theta\}$ for every $0 < \theta \le \frac{1}{2}$. Finally, with $\boldsymbol{t}_* \in [1/2,1]^d$ denoting the closest to $\boldsymbol{t}$ point in $[1/2,1]^d$, we extend the process $\widehat{\xi}_n$ to the whole cube $[0,1]^d$ as follows:

$$\widetilde{\xi}_n(\boldsymbol{t}) := \begin{cases} \widehat{\xi}_n(\boldsymbol{t}) & \text{if } \boldsymbol{t} \in [1/2,1]^d, \\ (1-2\theta)\widehat{\xi}_n(\boldsymbol{t}_*) & \text{if } \boldsymbol{t} \in D_\theta. \end{cases}$$

By definition, $\widetilde{\xi}_n(\boldsymbol{t})$ vanishes for every $\boldsymbol{t} \in D_{1/2}$. Since $D_{1/2} = [0,1]^d \setminus (0,1]^d$, we have the desired property $\widetilde{\xi}_n(\boldsymbol{t}) = 0$ for every $\boldsymbol{t} \in [0,1]^d \setminus (0,1]^d$.

We need to verify that the processes $\widetilde{\xi}_n$ satisfy the conditions of Theorem 1. Hence, in the next two subsections we shall show that

$$\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s})\right|^\alpha\right) \le c\|\boldsymbol{t} - \boldsymbol{s}\|^\beta \quad \text{whenever } \|\boldsymbol{t} - \boldsymbol{s}\| \ge \frac{1}{2}a_n \tag{24}$$

and

$$\omega_{\widetilde{\xi}_n}(a_n) \to_{\mathbf{P}} 0. \tag{25}$$

When proving (24) and (25), we shall find it convenient to subdivide considerations into the following three cases:

(A) both points $\boldsymbol{s}$ and $\boldsymbol{t}$ are in $[1/2,1]^d$;
(B) both points $\boldsymbol{s}$ and $\boldsymbol{t}$ are in $[0,1]^d \setminus (1/2,1]^d$;
(C) one of the points $\boldsymbol{s}, \boldsymbol{t}$ is in $[1/2,1]^d$ and the other one is in $[0,1]^d \setminus (1/2,1]^d$.

3.7 Verification of (24)

When $\boldsymbol{s}, \boldsymbol{t} \in [1/2,1]^d$, then we have that

$$\begin{aligned} \mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s})\right|^\alpha\right) &= \mathbf{E}\left(\left|\xi_n(2\boldsymbol{t} - 1) - \xi_n(2\boldsymbol{s} - 1)\right|^\alpha\right) \\ &\le c\|\boldsymbol{t} - \boldsymbol{s}\|^\beta \end{aligned} \tag{26}$$

whenever $\|\boldsymbol{t} - \boldsymbol{s}\| \ge \frac{1}{2}a_n$. This proves bound (24) in case (A).

To start the proof of bound (24) when $\boldsymbol{s}, \boldsymbol{t} \in [0,1]^d \setminus (1/2,1]^d$, we assume without loss of generality that $\boldsymbol{s}$ is closer to $[1/2,1]^d$ than $\boldsymbol{t}$, that is, $\|\boldsymbol{s} - [1/2,1]^d\|$

does not exceed $\|\boldsymbol{t}-[1/2,1]^d\|$. With $\boldsymbol{s}_* \in [1/2,1]^d$ denoting the closest to $\boldsymbol{s}$ point in $[1/2,1]^d$, we draw a straight line going through the points $\boldsymbol{s}_*$ and $\boldsymbol{s}$. We then take the intersection of this line with the set $[0,1]^d \setminus (1/2,1]^d$ and denote it by $I_{\boldsymbol{s}}$. The intersection $I_{\boldsymbol{s}}$ is a closed interval. One endpoint of the interval is $\boldsymbol{s}_*$, and we denote the other one by $\boldsymbol{s}^*$. That is, $I_{\boldsymbol{s}}$ equals $[\boldsymbol{s}_*, \boldsymbol{s}^*]$. With $\theta := \|\boldsymbol{t} - [1/2,1]^d\|$, let $\boldsymbol{s}_\theta$ denote the unique point of the intersection $I_{\boldsymbol{s}} \cap D_\theta$. We now write the bound

$$\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s})\right|^\alpha\right) \le c\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s}_\theta)\right|^\alpha\right) + c\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{s}_\theta) - \widetilde{\xi}_n(\boldsymbol{s})\right|^\alpha\right) \tag{27}$$

and estimate the two expectation on the right-hand side separately.

Since both $\boldsymbol{t}$ and $\boldsymbol{s}_\theta$ are in the set D_θ, we estimate the first expectation on the right-hand side of (27) as follows:

$$\begin{aligned}\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s}_\theta)\right|^\alpha\right) &= c\mathbf{E}\left(\left|(1-2\theta)\widehat{\xi}_n(\boldsymbol{t}_*) - (1-2\theta)\widehat{\xi}_n(\boldsymbol{s}_*)\right|^\alpha\right) \\ &\le c\mathbf{E}\left(\left|\widehat{\xi}_n(\boldsymbol{t}_*) - \widehat{\xi}_n(\boldsymbol{s}_*)\right|^\alpha\right).\end{aligned} \tag{28}$$

Since the right-hand side of (28) equals $c\mathbf{E}(|\xi_n(2\boldsymbol{t}_*-1) - \xi_n(2\boldsymbol{s}_*-1)|^\alpha)$, it does not exceed $c\|\boldsymbol{t}_* - \boldsymbol{s}_*\|^\beta$ when $\|\boldsymbol{t}_* - \boldsymbol{s}_*\| \ge \frac{1}{2}a_n$. The norm $\|\boldsymbol{t}_* - \boldsymbol{s}_*\|$ does not exceed $\|\boldsymbol{t} - \boldsymbol{s}\|$. This gives the desired upper bound $c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta$ for the right-hand side of (28) under the assumption that $\|\boldsymbol{t}_* - \boldsymbol{s}_*\| \ge \frac{1}{2}a_n$. If, however, the opposite holds, that is, if $\|\boldsymbol{t}_* - \boldsymbol{s}_*\| < \frac{1}{2}a_n$, then we can find a point $\boldsymbol{v} \in [1/2,1]^d$ such that $\|\boldsymbol{t}_* - \boldsymbol{v}\| = \frac{1}{2}a_n$ and $\|\boldsymbol{s}_* - \boldsymbol{v}\| = \frac{1}{2}a_n$. Using this auxiliary point $\boldsymbol{v}$, we estimate the right-hand side of (28) as follows:

$$\begin{aligned}\mathbf{E}\left(\left|\widehat{\xi}_n(\boldsymbol{t}_*) - \widehat{\xi}_n(\boldsymbol{s}_*)\right|^\alpha\right) &\le c\mathbf{E}\left(\left|\widehat{\xi}_n(\boldsymbol{t}_*) - \widehat{\xi}_n(\boldsymbol{v})\right|^\alpha\right) + c\mathbf{E}\left(\left|\widehat{\xi}_n(\boldsymbol{v}) - \widehat{\xi}_n(\boldsymbol{s}_*)\right|^\alpha\right) \\ &\le c\|\boldsymbol{t}_* - \boldsymbol{v}\|^\beta + c\|\boldsymbol{v} - \boldsymbol{s}_*\|^\beta.\end{aligned} \tag{29}$$

The right-hand side of (29) equals ca_n^β, which does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta$ because $\|\boldsymbol{t}-\boldsymbol{s}\| \ge \frac{1}{2}a_n$. This completes the proof that the first summand on the right-hand side of (27) does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta$.

To prove that the second expectation on the right-hand side of (27) does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^\beta$, we start with the equalities

$$\begin{aligned}\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{s}_\theta) - \widetilde{\xi}_n(\boldsymbol{s})\right|^\alpha\right) &= \mathbf{E}\left(\left|(1-2\theta_1)\widehat{\xi}_n(\boldsymbol{s}_*) - (1-2\theta_2)\widehat{\xi}_n(\boldsymbol{s}_*)\right|^\alpha\right) \\ &= 2^\alpha|\theta_1 - \theta_2|^\alpha \mathbf{E}\left(\left|\widehat{\xi}_n(\boldsymbol{s}_*)\right|^\alpha\right),\end{aligned} \tag{30}$$

where θ_1 is the distance between $\boldsymbol{s}_\theta$ and $\boldsymbol{s}_*$, and θ_2 is the distance between $\boldsymbol{s}$ and $\boldsymbol{s}_*$. Since the three points $\boldsymbol{s}_\theta$, $\boldsymbol{s}$ and $\boldsymbol{s}_*$ are on same straight line, $|\theta_1 - \theta_2|$ is the distance between $\boldsymbol{s}_\theta$ and $\boldsymbol{s}$. The latter distance does not exceed the distance between $\boldsymbol{t}$ and $\boldsymbol{s}$. Hence, from (30) we have that

$$\begin{aligned}\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{s}_\theta) - \widetilde{\xi}_n(\boldsymbol{s})\right|^\alpha\right) &\le c\|\boldsymbol{t}-\boldsymbol{s}\|^\alpha \mathbf{E}\left(\left|\widehat{\xi}_n(\boldsymbol{s}_*)\right|^\alpha\right) \\ &\le c\|\boldsymbol{t}-\boldsymbol{s}\|^\alpha \Big(1 + \sup_n \mathbf{E}\left(\left|\xi_n(\mathbf{0})\right|^\alpha\right)\Big),\end{aligned} \tag{31}$$

where the right-most bound in (31) can be proved as follows. If $\|\boldsymbol{t}\| \geq a_n$, then

$$\begin{aligned}\sup_{\boldsymbol{s}\in[0,1]^d} \mathbf{E}\left(\left|\xi_n(\boldsymbol{t})\right|^{\alpha}\right) &\leq c\mathbf{E}\left(\left|\xi_n(\mathbf{0})\right|^{\alpha}\right) + \mathbf{E}\left(\left|\xi_n(\mathbf{0}) - \xi_n(\boldsymbol{t})\right|^{\alpha}\right)\\ &\leq c\sup_n \mathbf{E}\left(\left|\xi_n(\mathbf{0})\right|^{\alpha}\right) + c. \end{aligned} \tag{32}$$

If $\|\boldsymbol{t}\| \leq a_n$, then we find a point $\boldsymbol{v} \in [0,1]^d$ such that $\|\boldsymbol{v}\| \geq a_n$ and $\|\boldsymbol{t}-\boldsymbol{v}\| \geq a_n$, and then write the bounds:

$$\begin{aligned}\sup_{\boldsymbol{s}\in[0,1]^d} \mathbf{E}\left(\left|\xi_n(\boldsymbol{t})\right|^{\alpha}\right) &\leq c\mathbf{E}\left(\left|\xi_n(\mathbf{0})\right|^{\alpha}\right) + \mathbf{E}\left(\left|\xi_n(\mathbf{0}) - \xi_n(\boldsymbol{v})\right|^{\alpha}\right) + \mathbf{E}\left(\left|\xi_n(\boldsymbol{v}) - \xi_n(\boldsymbol{t})\right|^{\alpha}\right)\\ &\leq c\sup_n \mathbf{E}\left(\left|\xi_n(\mathbf{0})\right|^{\alpha}\right) + 2c. \end{aligned} \tag{33}$$

This completes the proof of the right-most bound in (31). The supremum on the right-hand side of (31) is finite by assumption. This implies that the second summand on the right-hand side of (27) does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^{\alpha}$ and thus, in turn, does not exceed $c\|\boldsymbol{t}-\boldsymbol{s}\|^{\beta}$ since both $\boldsymbol{t}$ and $\boldsymbol{s}$ are in the unit cube $[0,1]^d$. This completes the proof of (24) in case (B).

We shall now verify (24) in case (C), assuming without loss of generality that $\boldsymbol{s} \in [1/2,1]^d$ and $\boldsymbol{t} \in [0,1]^d \setminus (1/2,1]^d$. Recall that we want to verify bound (24) for $\boldsymbol{s}$ and $\boldsymbol{t}$ such that $\|\boldsymbol{t}-\boldsymbol{s}\| \geq \frac{1}{2}a_n$. Given all this, we find a point $\boldsymbol{w}$ in the intersection of the sets $[1/2,1]^d$ and $[0,1]^d \setminus (1/2,1]^d$ such that $\|\boldsymbol{t}-\boldsymbol{w}\| \geq \frac{1}{2}a_n$ and $\|\boldsymbol{w}-\boldsymbol{s}\| \geq \frac{1}{2}a_n$. Hence,

$$\begin{aligned}\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s})\right|^{\alpha}\right) &\leq c\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{w})\right|^{\alpha}\right) + c\mathbf{E}\left(\left|\widetilde{\xi}_n(\boldsymbol{w}) - \widetilde{\xi}_n(\boldsymbol{s})\right|^{\alpha}\right)\\ &\leq c\|\boldsymbol{t}-\boldsymbol{w}\|^{\beta} + c\|\boldsymbol{w}-\boldsymbol{s}\|^{\beta}. \end{aligned} \tag{34}$$

Note that in addition to the properties above, the point $\boldsymbol{w}$ can be chosen such that the bounds $\|\boldsymbol{t}-\boldsymbol{w}\| \leq c\|\boldsymbol{t}-\boldsymbol{s}\|$ and $\|\boldsymbol{w}-\boldsymbol{s}\| \leq c\|\boldsymbol{t}-\boldsymbol{s}\|$ hold with a constant c that does not depend on $\boldsymbol{s}$ and $\boldsymbol{t}$. This observation and bound (34) complete the proof of (24).

3.8 Verification of (25)

Consider the difference $\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s})$ when $\|\boldsymbol{s}-\boldsymbol{t}\| \leq \frac{1}{2}a_n$. When both $\boldsymbol{s}$ and $\boldsymbol{t}$ are in $[1/2,1]^d$, the difference equals $\xi_n(2\boldsymbol{t}-1) - \xi_n(2\boldsymbol{s}-1)$. Hence, we have the bound

$$\left|\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s})\right| \leq \omega_{\xi_n}(a_n). \tag{35}$$

The right-hand side of (35) converges to 0 in probability by assumption, which completes the verification of (25) in case (A).

Consider now the difference $\widetilde{\xi}_n(\boldsymbol{t}) - \widetilde{\xi}_n(\boldsymbol{s})$ when both $\boldsymbol{s}$ and $\boldsymbol{t}$ are in $[0,1]^d \setminus (1/2,1]^d$. Let $\|\boldsymbol{s}-[1/2,1]^d\| \leq \theta := \|\boldsymbol{t}-[1/2,1]^d\|$, and let $\boldsymbol{s}_\theta$ denote the unique point in the intersection between of the earlier defined sets $I_{\boldsymbol{s}}$ and D_θ. We have that $|\widetilde{\xi}_n(\boldsymbol{t})-\widetilde{\xi}_n(\boldsymbol{s})|$ does not exceed the sum of the quantities $|\widetilde{\xi}_n(\boldsymbol{t})-\widetilde{\xi}_n(\boldsymbol{s}_\theta)|$ and $|\widetilde{\xi}_n(\boldsymbol{s}_\theta)-\widetilde{\xi}_n(\boldsymbol{s})|$. We next estimate these two quantities, starting with the first one. Since both $\boldsymbol{t}$ and $\boldsymbol{s}_\theta$ are in the set D_θ, we have that

$$\begin{aligned}\left|\widetilde{\xi}_n(\boldsymbol{t})-\widetilde{\xi}_n(\boldsymbol{s}_\theta)\right| &= \left|(1-2\theta)\widehat{\xi}_n(\boldsymbol{t}_*)-(1-2\theta)\widehat{\xi}_n(\boldsymbol{s}_*)\right| \\ &\le \left|\xi_n(2\boldsymbol{t}_*-1)-\xi_n(2\boldsymbol{s}_*-1)\right| \\ &\le \omega_{\xi_n}(a_n), \end{aligned} \tag{36}$$

where the right-most bound in (36) holds because $\|\boldsymbol{t}_*-\boldsymbol{s}_*\|$ does not exceed $\|\boldsymbol{t}-\boldsymbol{s}\|$, and the latter does not exceed $\frac{1}{2}a_n$. To estimate the quantity $|\widetilde{\xi}_n(\boldsymbol{s}_\theta)-\widetilde{\xi}_n(\boldsymbol{s})|$, we note (cf. Eqs. (30) for a hint) that it equals $2|\theta_1-\theta_2||\widehat{\xi}_n(\boldsymbol{s}_*)|$, where θ_1 is the distance between $\boldsymbol{s}_\theta$ and $\boldsymbol{s}_*$, and θ_2 is the distance between $\boldsymbol{s}$ and $\boldsymbol{s}_*$. Since $|\theta_1-\theta_2|$ is the distance between $\boldsymbol{s}_\theta$ and $\boldsymbol{s}$, it does not exceed $\|\boldsymbol{t}-\boldsymbol{s}\|$ and thus, in turn, does not exceed $\frac{1}{2}a_n$. We therefore have that

$$\left|\widetilde{\xi}_n(\boldsymbol{s}_\theta)-\widetilde{\xi}_n(\boldsymbol{s})\right| \le a_n \sup_{\boldsymbol{s}\in[0,1]^d}\left|\xi_n(\boldsymbol{s})\right|. \tag{37}$$

Using bounds (36) and (37), we obtain that

$$\left|\widetilde{\xi}_n(\boldsymbol{t})-\widetilde{\xi}_n(\boldsymbol{s})\right| \le \omega_{\xi_n}(a_n)+a_n \sup_{\boldsymbol{s}\in[0,1]^d}\left|\xi_n(\boldsymbol{s})\right|. \tag{38}$$

The right-hand side of (38) converges to 0 in probability since a_n and $\omega_{\xi_n}(a_n)$ converge to zero, and we already know from (31)–(33) that the supremum $\sup_{\boldsymbol{s}\in[0,1]^d}|\xi_n(\boldsymbol{s})|$ is asymptotically bounded in probability.

Finally, we consider the difference $\widetilde{\xi}_n(\boldsymbol{t})-\widetilde{\xi}_n(\boldsymbol{s})$ in case (C), assuming without loss of generality that $\boldsymbol{s}\in[1/2,1]^d$ and $\boldsymbol{t}\in[0,1]^d\setminus(1/2,1]^d$. Using bounds (38) and (35), we have that

$$\begin{aligned}\left|\widetilde{\xi}_n(\boldsymbol{t})-\widetilde{\xi}_n(\boldsymbol{s})\right| &\le \left|\widetilde{\xi}_n(\boldsymbol{t})-\widetilde{\xi}_n(\boldsymbol{t}_*)\right|+\left|\widetilde{\xi}_n(\boldsymbol{t}_*)-\widetilde{\xi}_n(\boldsymbol{s})\right| \\ &\le \left(\omega_{\xi_n}(a_n)+a_n\sup_{\boldsymbol{s}\in[0,1]^d}\left|\xi_n(\boldsymbol{s})\right|\right)+\omega_{\xi_n}(a_n). \end{aligned} \tag{39}$$

The right-hand side of bound (39) converges to 0 in probability. This completes the verification of (25). The proofs of Theorem 1 and Corollary 1 are now finished.

4 Relationship of main results to Dudley's contributions

Let $\mathcal{X}$ be a complete metric space, and denote the metric by ρ. We do not assume that $\mathcal{X}$ is separable since the example we have in mind is the space $B([0,1]^d;\mathbf{R})$ equipped with the supremum norm. Next, let $\mathcal{Y}\subseteq\mathcal{X}$ be a complete separable metric space equipped with the same metric ρ. The example of $\mathcal{Y}$ we have in mind is the space $C([0,1]^d;\mathbf{R})$ equipped with the supremum norm.

Let $\xi_n, n \geq 1$, be random elements of $\mathcal{X}$, and we want to check if they weakly (in some sense) converge to an element η of $\mathcal{Y}$. The information we have consists of two facts. First, there are elements η_n of $\mathcal{Y}$ that weakly converge (in the usual sense) to η. Second, the distance $\rho(\xi_n, \eta_n)$ (assuming that it is a random variable) converges to 0 in probability. We should not worry about the distance $\rho(\xi_n, \eta_n)$ being a random variable since this problem is not usually difficult to check in practical situations. What we really need to find out, however, is how to define weak convergence of the processes ξ_n to η, since the classical definition

(A) $\lim_{n\to\infty} \mathbf{E}[f(\xi_n)] = \mathbf{E}[f(\eta)]$ holds for every bounded and continuous functional $f : \mathcal{X} \to \mathbf{R}$

does not work here, due to the fact that $f(\xi_n)$ might not be a random variable. Indeed, the element ξ_n might not be measurable with respect to the Borel σ-algebra $\mathcal{B}_{\mathcal{X}}$. Hence, the expectation $\mathbf{E}[f(\xi_n)]$ might not be defined. This situation occurs, for example, with the (classical) empirical processes e_n, since they are not measurable with respect to $\mathcal{B}_{\mathcal{X}}$. However, the processes are measurable with respect to the (smaller) σ-algebra $\mathcal{B}_0$ generated by all open balls of $\mathcal{X}$.

Dudley (1966) suggested a way to define weak convergence of stochastic processes ξ_n in possibly non-separable spaces $\mathcal{X}$ as follows:

(B) $\lim_{n\to\infty} \mathbf{E}^*[f(\xi_n)] = \lim_{n\to\infty} \mathbf{E}_*[f(\xi_n)] = \mathbf{E}[f(\eta)]$ for every bounded and continuous function $f : \mathcal{X} \to \mathbf{R}$, where $\mathbf{E}^*$ and $\mathbf{E}_*$ denote, respectively, the upper and lower expectations (cf. Dudley, 1966).

The latter definition of weak convergence is equivalent (cf. Dudley, 1967) to the statement that

(C) $\lim_{n\to\infty} \mathbf{E}[f(\xi_n)] = \mathbf{E}[f(\eta)]$ for every bounded, continuous, and $\mathcal{B}_0$-measurable function $f : \mathcal{X} \to \mathbf{R}$.

Formulation (C) of weak convergence is very close to what we really need for the purpose of the current project. Namely, we want to have the following statement:

(D) $\lim_{n\to\infty} \mathbf{E}[f(\xi_n)] = \mathbf{E}[f(\eta)]$ for every bounded, *uniformly* continuous, and $\mathcal{B}_0$-measurable function $f : \mathcal{X} \to \mathbf{R}$.

We shall later prove (cf. Proposition 1) that statements (C) and (D) are equivalent. Now we show how property (D) can be put into good use.

Suppose that we want to prove weak convergence of the processes ξ_n to continuous η. In view of (D), this means that we want to prove that the difference $\mathbf{E}[f(\xi_n)] - \mathbf{E}[f(\eta)]$ converges to 0 for every bounded and uniformly continuous function f, which is measurable with respect to the σ-algebra $\mathcal{B}_0$. We know from the classical weak convergence notion that $\mathbf{E}[f(\eta_n)] - \mathbf{E}[f(\eta)]$ converges to 0 for the aforementioned functions f. Thus, we are only left to verify that $\mathbf{E}[f(\xi_n)] - \mathbf{E}[f(\eta_n)]$ converges to 0 when n tends to infinity. Since the function f is *uniformly* continuous, we derive the desired weak convergence from the convergence of $\rho(\xi_n, \eta_n)$ to 0 in probability.

To see what can be obtained from the above discussed weak convergence of ξ_n to η, we have (cf. Dudley, 1966, 1999) that $f(\xi_n)$ converge to $f(\eta)$ in distribution, provided that the function f is $\mathcal{B}_0$-measurable and μ-almost everywhere continuous, where μ denotes the probability measure induced by the process η. Note that it is trivial to check the $\mathcal{B}_0$-measurability when the function f is the supremum norm $\|\cdot\|_\infty$. Verification of $\mathcal{B}_0$-measurability for functions other than the supremum norm might be less trivial. Keeping this potential problem in mind, in a number of specific applications we can reason as follows.

Assume that a specific application involves a functional, say f, and that we want to establish convergence in distribution of $f(\xi_n)$ to $f(\eta)$. It is worth noting that checking directly that the quantities $f(\xi_n)$ are random variables is usually an easier problem than doing so via specific measurability properties of f and ξ_n *separately*. Having thus resolved the measurability question of $f(\xi_n)$, we then check that $\rho(\xi_n, \eta_n)$ are random variables and that they converge to 0 in probability. From these facts and also using the weak convergence of the processes η_n to η, we derive (cf. Proposition 2) the convergence of $f(\xi_n)$ to $f(\eta)$ in distribution, provided that the function f is almost everywhere continuous with respect to the measure μ induced by η. Verification of the latter property is a task well discussed in the literature.

Proposition 1 *Statements* (C) *and* (D) *are equivalent.*

Proof Obviously, from statement (C) we have (D). Hence, we now assume the validity of statement (D) and prove (C). That is, let f be a continuous, bounded, and $\mathcal{B}_0$-measurable function. (In the proof below it will be convenient to work with functions f such that $0 \le f(x) \le 1$, which we can always assume without loss of generality.) We want to show that the convergence $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ holds with μ_n and μ denoting the probability measures induced by the processes ξ_n and η, respectively. This can, of course, be written as $\mathbf{E}[f(\xi_n)] \to \mathbf{E}[f(\eta)]$. Note that by assumption (D) the latter convergence holds when the function f is uniformly continuous.

We start proving the convergence of $\int f \mathrm{d}\mu_n$ to $\int f \mathrm{d}\mu$ with several notes. First, since the space $\mathcal{Y}$ is separable and complete, for any $\epsilon > 0$ we can find a compact set K such that $\mu(K^c) < \epsilon$, where K^c denotes the complement of K. Second, given $\epsilon > 0$, we can find $\delta > 0$ such that $|f(x) - f(y)| \le \epsilon$ whenever $x \in K, y \in \mathcal{X}$, and $\rho(x, y) < \delta$. We next construct a $\mathcal{B}_0$-measurable function f^+ on $\mathcal{X}$ such the bounds

$$f^+(x) - 2\epsilon \le f(x) \le f^+(x) \tag{40}$$

hold for every $x \in K(\delta)$, where $K(\delta)$ denotes the open δ-neighbourhood of K. Let $\{x_n\} \subset K$ be a countable and everywhere dense subset, and let

$$f^+(x) := \sup_n \left(f(x_n) + \epsilon\right) \varphi(x, x_n),$$

where

$$\varphi(x, x_n) := \mathbf{1}\left\{d(x, x_n) \le \frac{\delta}{2}\right\} + 2\left(1 - \frac{1}{\delta} d(x, x_n)\right)\mathbf{1}\left\{\frac{\delta}{2} < d(x, x_n) \le \delta\right\}.$$

It is easy to see that in this way constructed functions $x \mapsto \varphi(x, x_n)$ are $\mathcal{B}_0$-measurable, and so f^+ is $\mathcal{B}_0$-measurable as well.

Equipped now with f^+, we continue checking the convergence of the integrals $\int f \mathrm{d}\mu_n$ to $\int f \mathrm{d}\mu$ with the bounds

$$\limsup_{n\to\infty} \int f \mathrm{d}\mu_n \le \limsup_{n\to\infty} \int_{K(\delta)} f \mathrm{d}\mu_n + \limsup_{n\to\infty} \mu_n(K(\delta)^c) \le \limsup_{n\to\infty} \int_{K(\delta)} f^+ \mathrm{d}\mu_n + \epsilon \tag{41}$$

that hold because of $f \le f^+$ on $K(\delta)$, and also because of the following arguments. Let ϕ be a function on the real line such that $\phi(t) = 1$ when $t \le 0, \phi(t) = 1 - t$ when $0 \le t \le 1$, and $\phi(t) = 0$ when $t \ge 1$. Define

$$F(x) := 1 - \phi\left(\frac{1}{\delta}\, d(x, K)\right),$$

where $d(x, K)$ is the distance between the point x and the compact set K. Note that the function F is bounded, uniformly continuous, and $\mathcal{B}_0$-measurable. Hence, we have the bounds

$$\begin{aligned}\limsup_{n\to\infty} \mu_n(K(\delta)^c) &\le \limsup_{n\to\infty} \int \mathbf{1}_{K(\delta)^c} \mathrm{d}\mu_n \\ &\le \limsup_{n\to\infty} \int F \mathrm{d}\mu_n = \int F \mathrm{d}\mu \le \mu(K^c) < \epsilon.\end{aligned} \tag{42}$$

This completes the proof of the right-most bound in (41). We still need to further estimate the integral on the right-hand side of (41). For this, we use bounds (40) and obtain the following string of inequalities:

$$\limsup_{n\to\infty} \int_{K(\delta)} f^+ \mathrm{d}\mu_n \le \limsup_{n\to\infty} \int f^+ \mathrm{d}\mu_n = \int f^+ \mathrm{d}\mu \le \int_K f^+ \mathrm{d}\mu + \epsilon \le \int_K f \mathrm{d}\mu + 3\epsilon. \tag{43}$$

From (41) and (43) we obtain that

$$\limsup_{n\to\infty} \int f \mathrm{d}\mu_n \le \int_K f \mathrm{d}\mu. \tag{44}$$

Note that bound (44) also holds with the function $1 - f$. This proves that the lower limit of $\int f \mathrm{d}\mu_n$ has the same lower bound as the upper bound in (44). Hence, the limit of $\int f \mathrm{d}\mu_n$ exists and is equal to $\int_K f \mathrm{d}\mu$. The equivalence of (C) and (D) has been established.

Proposition 2 *Assume that $\rho(\xi_n, \eta_n)$ are random variables and that they converge to* 0 *in probability. Furthermore, let the stochastic processes $\eta_n \in \mathcal{Y}$ weakly converge to $\eta \in \mathcal{Y}$. Assume that a function f defined on $\mathcal{X}$ is such that its set of discontinuity points has μ-measure zero, where μ denotes the measure induced by the process η. Given the above, the quantities $f(\xi_n)$ (which we assume to be random variables) converge to $f(\eta)$ in distribution.*

Proof Since the processes η_n converge weakly to η in the complete separable space $\mathcal{Y}$, for any $\epsilon > 0$ we can find a compact set K such that $\mu(K^c) := \mathbf{P}\{\eta \in K^c\} < \epsilon$ and $\mu_n(K^c) := \mathbf{P}\{\eta_n \in K^c\} < \epsilon$ for all n. Furthermore, the compact set K can be chosen in such a way that the function f would be (uniformly) continuous on K. Now, for any $\nu > 0$ and $\delta > 0$, we write the bound

$$\begin{aligned}\mathbf{P}\{|f(\xi_n) - f(\eta_n)| > \nu\} \leq \mathbf{P}\{&|f(\xi_n) - f(\eta_n)| > \nu,\ \rho(\xi_n, \eta_n) \leq \delta,\ \eta_n \in K\} \\ &+ \mathbf{P}\{\rho(\xi_n, \eta_n) > \delta\} + \mathbf{P}\{\eta_n \in K^c\}.\end{aligned} \tag{45}$$

For all sufficiently large n, the second and the third probabilities on the right-hand side of (45) do not exceed ϵ. A similar statement holds for the first probability. To show this, we start with the note that for any $\nu > 0$ we can find $\delta > 0$ such that $|f(x) - f(y)| \leq \nu$ whenever $x \in K$ and $y \in \mathcal{X}$ are such that $d(x, y) < \delta$. Indeed, this immediately follows from the continuity of the function f at every point of K, including every boundary point of K. Hence, for every point $x \in K$, there is an open ball of radius $\delta(x) > 0$ such that the values of f on the ball are at the distance from $f(x)$ not larger than the given ϵ. Recall now that from every *open* covering of a compact set we can select a *finite* covering. Taking the minimum of the (finite number of) radii of the just selected balls, and then denoting the minimum by δ, we obtain the desired property $|f(x) - f(y)| \leq \nu$ whenever $x \in K$ and $y \in \mathcal{X}$ are such that $d(x, y) < \delta$. But this contradicts the statement $|f(\xi_n) - f(\eta_n)| > \nu$ inside the first probability on the right-hand side of (45). Hence, the probability must be zero. This, in turn, implies that, for every $\nu > 0$ and $\epsilon > 0$,

$$\limsup_{n\to\infty} \mathbf{P}\{|f(\xi_n) - f(\eta_n)| > \nu\} \leq 2\epsilon. \tag{46}$$

Consequently, the difference $f(\xi_n) - f(\eta_n)$ converges to zero in probability. This implies that $f(\xi_n)$ converges to $f(\eta)$ in distribution, which finishes the entire proof.

Acknowledgements Our sincere thanks are due to a referee, an Associate Editor, and the Editor for careful reading of the manuscript and constructive suggestions that helped us to revise the paper.

The impetus for writing this paper was given by a number of discussions with Lajos Horváth and Piotr Kokoszka on econometric time series, Edna Schechtman and Shlomo Yitzhaki on measures of economic inequality, Mark Bebbington and Chin-Diew Lai on reliability theory. Charles M. Goldie kindly agreed to read a preliminary draft of the paper and sent us a number of comments and suggestions that resulted in a substantial revision of the original manuscript. We are grateful to Charles M. Goldie and all the above noted individuals for their influence on the paper.

The research of the second author was partially supported by a Visiting Professorship at the University of Science and Technology of Lille, as well as by a Discovery Research Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.

Csörgő, M., Horváth, L. (1993). *Weighted approximations in probability and statistics*. Chichester: Wiley.

Davydov, Y. (1996). Weak convergence of discontinuous processes to continuous ones. In: *Probability theory and mathematical statistics* (pp. 15–18). Amsterdam: Gordon and Breach.

Davydov, Y., Zitikis, R. (2004). *Convex rearrangements of random elements*. In: Asymptotic methods in stochastics. Fields Institute Communications (pp. 141–171), Vol. 44. Providence, RI: American Mathematical Society.

Dehling, H., Mikosch, T., Sørensen, M. (2002). *Empirical process techniques for dependent data*. Boston, MA: Birkhäuser.

Doukhan, P., Oppenheim, G., Taqqu, M. S. (2003). *Theory and applications of long-range dependence*. Boston, MA: Birkhäuser.

Dudley, R. M. (1966). Weak convergences of probabilities on nonseperable metric spaces and empirical measures on Euclidean spaces. *Illinois Journal of Mathematics*, *10*, 109–126.

Dudley, R. M. (1967). Measures on non-separable metric spaces. *Illinois Journal of Mathematics*, *11*, 449–453.

Dudley, R. M. (1999). *Uniform central limit theorems*. Cambridge: Cambridge University Press.

Goldie, C. M. (1977). Convergence theorems for empirical Lorenz curves and their inverses. *Advances in Applied Probability*, *9*, 765–791.

Hall, P., Heyde, C. C. (1980). *Martingale limit theory and its application*. Berlin Heidelberg New York: Academic Press.

Ibragimov, I. A., Has'minskii, R. Z. (1981). Statistical estimation. *Asymptotic theory*. Berlin Heidelberg New York: Springer.

Ivanoff, G., Merzbach, E. (2000). *Set-indexed martingales*. Boca Raton, FL: Chapman and Hall.

Khoshnevisan, D. (2002). *Multiparameter processes. An introduction to random fields*. Berlin Heidelberg New York: Springer.

Korolyuk, V., Portenko, N., Syta, H. (2000). *Skorokhod's ideas in probability theory*. Kyiv: National Academy of Sciences of Ukraine.

Koul, H. L. (2002). *Weighted empirical processes in dynamic nonlinear models*. Berlin Heidelberg New York: Springer.

Pollard, D. (1984). *Convergence of stochastic processes*. Berlin Heidelberg New York: Springer.

Shorack, G. R., Wellner, J. A. (1986). *Empirical processes with applications to statistics*. New York: Wiley.

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes*. Berlin Heidelberg New York: Springer.