On Asymptotic Proximity of Distributions

Youri DAVYDOV[1]   and   Vladimir ROTAR [2]

[1] Laboratoire Paul Painlevé - UMR 8524

Université de Lille I - Bat. M2

59655 Villeneuve d'Ascq, France

Email: youri.davydov@univ-lille1.fr

[2] Department of Mathematics and Statistics

of the San Diego State University, USA and

the Central Economics and Mathematics Institute

of the Russian Academy of Sciences, RF

Email: vrotar@math.ucsd.edu

Abstract. We consider some general facts concerning convergence

$$P_n - Q_n \to 0 \ \text{ as } \ n \to \infty,$$

where P_n and Q_n are probability measures in a complete separable metric space. The main point is that the sequences $\{P_n\}$ and $\{Q_n\}$ are not assumed to be tight. We compare different possible definitions of the above convergence, and establish some general properties.

1

1 Introduction and results

1.1 Background and Motivation

Usually, a limit theorem of Probability Theory is a theorem that concerns convergence of a sequence of distributions P_n to a distribution P. However, there is a number of works where the traditional setup is modified, and the object of study is two sequences of distributions, $\{P_n\}$ and $\{Q_n\}$. The goal in this case consists in establishing conditions for convergence

$$P_n - Q_n \to 0 \qquad\qquad (1.1.1)$$

in a proper sense. In particular problems, P_n and Q_n are, as a rule, the distributions of the r.v.'s $f_n(X_1,...,X_n)$ and $f_n(Y_1,...,Y_n)$, where $f_n(\cdot)$ is a function, and $X_1, X_2, ...$ and $Y_1, Y_2, ...$ are two sequences of r.v.'s. The aim here is rather to show that different random arguments $X_1,...,X_n$ may generate close distributions of $f_n(X_1,...,X_n)$, than to prove that the distribution of $f_n(X_1,...,X_n)$ is close to some fixed distribution (which, above else, may be not true).

Consider, for example, a quadratic form $< A_n \mathbf{X}_n, \mathbf{X}_n >$, where A_n is a $n \times n$-matrix and $\mathbf{X}_n = (X_1,...,X_n)$ is a vector with i.i.d. coordinates. In this case, the limiting distribution, if any, depends on the matrices A_n. For instance, with a proper normalization, the form $X_1 X_2 + X_2 X_3 + ... + X_{n-1} X_n$ is asymptotically normal (if $E\{X_i^2\} < \infty$), while the form $(X_1 + ... + X_n)^2$ has the χ_1^2 limiting distribution. Nevertheless, one can state the following unified limit theorem.

Denote by P_{nF} the distribution of $< A_n \mathbf{X}_n, \mathbf{X}_n >$ in the case when each X_i has a distribution F. Then, under rather mild conditions,

$$P_{nF} - P_{nG} \to 0 \qquad\qquad (1.1.2)$$

for any two distributions F and G with the same first two moments (see, [13], [7] for detail, and references therein). A class $\mathcal{F}$ such that (1.1.2) is true for any $F, G \in \mathcal{F}$ is called an *invariance class* ([13]).

Let us come back to (1.1.1). Clearly, such a framework is more general than the traditional one. First, as was mentioned, the distributions P_n and Q_n themselves do not have to converge. Secondly, the sequences $\{P_n\}$ and $\{Q_n\}$ are not assumed to be tight, and the convergence in (1.1.1) covers situations when a part of the probability distributions or the whole distributions "move away to infinity" while the distributions P_n and Q_n are approaching each other.

To our knowledge, the scheme above was first systematically used in Loéve [10, Chapter VIII, Section 28, The Central Asymptotic Problem], who considered

2

sums of dependent r.v.'s. The same approach is applied in some non-classical limit theorems for sums of r.v.'s; that is, theorems not involving the condition of asymptotic negligibility of separate terms (see, e.g., monograph [19] by Zolotarev, survey [13] by Rotar, and references in [19] and [13]; Liptser and Shiryaev [9], and Davydov and Rotar [6] on a non-classical invariance principle.)

Non-linear functions f_n have been also considered in the above framework. In particular, it concerns polynomials, polylinear forms of r.v.'s, and quasi-polynomial functions; see, e.g., Rotar [13] and [14] for limit theorems, Götze and Tikhomirov [7], and Mossel, O'Donnel and Oleszkiewicz [11] for the accuracy of the corresponding invariance principle. Other interesting schemes different from those above were explored in D'Aristotile, Diaconis, and Freedman [2] and in Chatterjee [1].

The present paper addresses general facts on convergence (1.1.1). A corresponding theory was built in Dudley [3], [4], [5, Chapter 11] and D'Aristotile, Diaconis, and Freedman [2]. The paper [2] concerns some possible definitions of convergence (1.1.1) in terms of uniformities, and establishes connections between these definitions (see also below). The theory in [3]-[5] (which is used in part in [2] also) is mainly based on a metric approach. We complement and, to a certain extent, develop the theory from [2] and [5], paying more attention to a functional approach. Throughout this paper, we repeatedly refer to and cite results from [2]-[5].

First, consider three definitions of convergence (1.1.1) explored in [2] (or, in the terminology of [2], "merging").

D1. $\pi(P_n, Q_n) \to 0$ where π is the Lévy-Prokhorov metric.

D2.
$$\int f(x) \left(P_n(dx) - Q_n(dx)\right) \to 0 \qquad (1.1.3)$$

for all bounded continuous functions f.

D3. $T(P_n) - T(Q_n) \to 0$ for all bounded and continuous (with respect to weak * topology) functions T on the space of probability measures.

In [2], it was shown that **D3** $\Rightarrow$ **D2** $\Rightarrow$ **D1**, and **D1, D2, D3** are equivalent iff the space on which measures are defined, is compact. As one can derive from [2], and as follows from results below, once we consider particular sequences $\{P_n\}$ and $\{Q_n\}$, the above definitions are equivalent if one of the sequences is tight.

In the general setup, when the above sequences $\{P_n\}$ and $\{Q_n\}$ are not assumed to be tight, Definition **D2** (and hence **D3** also) looks too strong. As an example, consider that from [2]. Let us deal with distributions on the real line, and let P_n be concentrated at the point n, and Q_n – at the point $n + \frac{1}{n}$. Clearly, (1.1.3) is not true for, say, $f(x) = \sin(x^2)$. On the other hand, it would have been unnatural, if a definition had not covered such a trivial case of asymptotic proximity of distributions. (Clearly, in this case, $\pi(P_n, Q_n) \to 0$. To make the example simpler, one may consider $f(x) = \sin\left(\frac{\pi}{4}x^2\right)$. Certainly, the example above concerns the Euclidean metric in $\mathbb{R}$. For other metrics, points n and $n + \frac{1}{n}$ may be not close. Below, we cover the general case of a complete separable metric space.)

To fix the situation, one can consider (1.1.3) for functions only from the class of all bounded continuous functions vanishing at infinity, but such an approach would be too restrictive. In this case, the definition would not cover situations where parts of the distributions move away to infinity, continuing to approach each other (or, in the terminology from [2], "merge"). On the other hand, in accordance with the same definition, the distributions P_n and Q_n concentrated, for example, at points n and $2n$, would be viewed as merging, which also does not look reasonable.

In this paper, we suggest to define weak convergence as that with respect to all bounded *uniformly* continuous functions. (This type of convergence was not considered in [2] and [5].) We justify this definition proving that such a convergence is equivalent to convergence in the Lévy-Prokhorov metric that satisfies some natural, in our opinion, requirements. In the counterpart of Definition **D3**, we require the uniform continuity of T.

We establish also some facts concerning weak convergence uniform on certain classes of functions f (or linear functionals on the space of distributions); see Section 1.2.3 for detail. Proofs turn out to be, though not very difficult, but not absolutely trivial since the absence of tightness requires additional constructions. The point is that in the generalized setup, we cannot choose just one compact, not depending on n, on which all measures will be "almost concentrated".

On the other hand, as will follow from results below, if one of the sequences, P_n or Q_n, is tight, the definition of weak convergence suggested is equivalent to the classical definition, and we deal with the classical framework.

1.2 Main Results

1.2.1 Weak convergence and the Lévy-Prokhorov metric

Let $(\mathbb{H}, \rho)$ be a complete separable metric space, and $\mathcal{B}$ be the corresponding Borel σ-algebra.

The symbols P and Q, with or without indices, will denote probability distributions on $\mathcal{B}$. All functions f below, perhaps with indices, are continuous functions $f : \mathbb{H} \to \mathbb{R}$.

We denote by $\mathcal{C}$ the class of all bounded and continuous functions on $(\mathbb{H}, \rho)$, and by $\overline{\mathcal{C}}$ - the class of all bounded and uniformly continuous functions.

For two sequences of probability measures (distributions), $\{P_n\}$ and $\{Q_n\}$, we say that $P_n - Q_n \to 0$ *weakly with respect to (w.r.t.) a class of functions $\mathcal{K}$* if $\int f d(P_n - Q_n) \to 0$ for all $f \in \mathcal{K}$.

If we do not mention a particular class $\mathcal{K}$, *the term weak convergence (or more precisely, merging) will concern that w.r.t.* $\overline{\mathcal{C}}$. When it cannot cause misunderstanding, we will use the term "convergence" in the situation of merging also.

In the space of probability distributions on $\mathcal{B}$, we define - in a usual way - the *Lévy-Prokhorov metric*

$$\pi(P, Q) = \inf\{\varepsilon : P(A^\varepsilon) \le Q(A) + \varepsilon \text{ for all closed sets } A\}. \tag{1.2.1}$$

(One can restrict himself to only one inequality, not switching P and Q; see Dudley [4, Theorem 11.3.1].)

Our main result is

Theorem 1 *The difference*

$$P_n - Q_n \to 0 \quad \text{weakly (w.r.t. } \overline{\mathcal{C}}) \tag{1.2.2}$$

if and only if

$$\pi(P_n, Q_n) \to 0. \tag{1.2.3}$$

The choice of the Lévy-Prokhorov metric as a "good" metric that justifies the definition of weak convergence above, is connected, first of all, with the fact that the analog of the Skorokhod representation theorem ([16], see also, e.g., [5, Sec.11.7]) holds in the case of merging measures. More precisely, the following is true.

5

Let, the symbols X and Y, perhaps with indices, denote random variables defined on a probability space $(\Omega, \mathcal{A}, P)$, and assuming values in $\mathbb{H}$ (that is, these variables are $\mathcal{A} \to \mathcal{B}$ measurable.) The symbol P_X stands for the distribution of X.

Below, two metrics $r_1(x,y)$ and $r_2(x,y)$ in a space are said to be *uniformly equivalent*, if for any two sequences $\{x_n\}$ and $\{y_n\}$, the relations $r_1(x_n, y_n) \to 0$ and $r_2(x_n, y_n) \to 0$ are equivalent.

The first two (and main) assertions of the next theorem are stated and proved in Dudley [4]; see also the second edition [5, Theorem 11.7.1]. For the completeness of the picture, we present all facts regarding the metric π in one theorem.

Theorem 2 *Metric π is the only metric, to within uniform equivalence, that satisfies the following conditions.*

A. *If $\rho(X_n, Y_n) \overset{P}{\to} 0$, then $\pi(P_{X_n}, P_{Y_n}) \to 0$.*

B. *If $\pi(P_n, Q_n) \to 0$, then there exist a probability space and random elements X_n, Y_n on this space such that $P_{X_n} = P_n$, $P_{Y_n} = Q_n$, and $\rho(X_n, Y_n) \overset{P}{\to} 0$.*

C. *If $\pi(P_n, Q_n) \to 0$, then $\pi(P_n \circ f^{-1}, Q_n \circ f^{-1}) \to 0$ for any uniformly continuous function f.*

D. *If $Q_n = Q$, then the convergence $\pi(P_n, Q) \to 0$ is equivalent to weak convergence with respect to all bounded continuous functions.*

Remarks.

1. We have already mentioned above an example showing that convergence (1.1.3) for all $f \in C$ does not possess Property A. In other words, there exist r.v.'s X_n and Y_n such that $\rho(X_n, Y_n) \overset{P}{\to} 0$, while $P_{X_n} - P_{Y_n} \not\to 0$ weakly with respect to C. If $\mathbb{H}$ has bounded but non-compact sets, a similar example may be constructed even for continuous functions equal to zero out of a closed bounded set.

   Indeed, let $O_r(x) = \{y : \rho(x,y) \leq r\}$. Consider the case where for some x_0, the ball $O = O_1(x_0)$ is not compact. Then there exists a sequence $\{x_n\} \subset O$ which does not contain a converging subsequence. Furthermore, there exists a numerical sequence $\delta_n \to 0$, such that the balls $O_{\delta_n}(x_n)$ are disjoint, and each contains only one element from $\{x_n\}$, that is, the center x_n. We define

$$f(x) = \begin{cases} 1 - \frac{1}{\delta_n}\rho(x,x_n) & \text{if } x \in O_{\delta_n}(x_n), \\ 0 & \text{if } x \not\in \cup_k O_{\delta_k}(x_k). \end{cases}$$

The function f is, first, bounded, and secondly, due to the choice of $\{x_n\}$, f is continuous. On the other hand, one may set $X_n \equiv x_n$ and $Y_n \equiv y_n$, where y_n is a point from the boundary of $O_{\delta_n}(x_n)$. Clearly, $\rho(X_n, Y_n) = \delta_n \to 0$, while $\int f dP_{X_n} = 1$ and $\int f dP_{Y_n} = 0$.

2. As was mentioned in the introduction, if one of the sequences, say $\{Q_n\}$, is tight, and (1.2.2) is true, then the other sequence, $\{P_n\}$, is also tight. In this case, relation (1.1.3) is true for all $f \in C$, and we deal with the classical scheme. If $(\mathbb{H}, \rho)$ is a space in which any closed bounded set is compact, this fact is easy to prove directly. In the general case, it is easier to appeal to Theorem 1 and Prokhorov's theorem on relative compactness w.r.t. functions from C ([12]).

   More precisely, assume that (1.2.2) holds and $\{Q_n\}$ is tight. Then, by Prokhorov's theorem, $\{Q_n\}$ is relatively compact with respect to weak convergence for functions from C. Let a subsequence Q_{n_k} weakly converges to some Q w.r.t. C, and hence $\pi(Q_{n_k}, Q) \to 0$. By virtue of Theorem 1, $\pi(P_{n_k}, Q_{n_k}) \to 0$, and consequently, $\pi(P_{n_k}, Q) \to 0$. So, P_{n_k} converges to the same Q w.r.t. to C, and

   $$P_{n_k} - Q_{n_k} \to 0 \text{ w.r.t. to all functions from } C. \qquad (1.2.4)$$

   Thus, any subsequence of $\{P_n\}$ contains a subsubsequence convergent w.r.t. C. Hence, again by Prokhorov's theorem, $\{P_n\}$ is tight.

   Moreover, in this case $P_n - Q_n \to 0$ w.r.t. C. Indeed, otherwise we could select convergent subsequences P_{n_k} and Q_{n_k} and a bounded and continuous f such that $\int f d(P_{n_k} - Q_{n_k}) \not\to 0$, which would have contradicted (1.2.4).

3. Let us return to Definition **D3** from Section 1.1. To make it suitable to the setup of this paper, one should consider functions T on the space of probability measures on $\mathcal{B}$, *uniformly* continuous with respect to the the Lévy-Prokhorov metric (or, which is the same, w.r.t. the weak convergence regarding $\overline{C}$). So modified Definition **D3** will be equivalent to convergences (1.2.2)-(1.2.3).

1.2.2 Lipschitz functions

For $f \in C$, we set

$$\|f\|_L = \sup_{x \neq y} \left\{ \frac{|f(x) - f(y)|}{\rho(x,y)} \right\},$$

7

and $C_{BL} = \{f : \|f\|_L < \infty, \|f\|_\infty < \infty\}$.

In Dudley [5, Sec.11.3], it was shown that in the traditional setup, for the weak convergence $P_n \to P$ w.r.t. the whole class C, it suffices to have the convergence w.r.t. C_{BL}. A similar property is true for the generalized setup of merging probability measures. Namely, let $\{P_n\}$ and $\{Q_n\}$ be two fixed sequences of probability measures.

Theorem 3 *The weak convergence $P_n - Q_n \to 0$ with respect to C_{BL} implies the weak convergence with respect to $\overline{C}$.*

Next, we consider classes of functions on which weak convergence is uniform.

1.2.3 Uniform convergence

In Dudley [5], it was shown that in the traditional setup, weak convergence is uniform on any class of functions f with uniformly bounded norms $\|f\|_L$ and $\|f\|_\infty$. More precisely, consider the metric

$$\beta(P, Q) := \sup \left\{ \left| \int f d(P - Q) \right| : \ \|f\|_L + \|f\|_\infty \leq 1 \right\}.$$

In [5, Section 1.3], it was proved that the weak convergence $P_n \to P$ w. r. t. C, is equivalent to the convergence $\beta(P_n, P) \to 0$.

We establish a similar property in the generalized setup and for arbitrary classes of functions with a fixed order of their moduli of continuity.

For $f \in \overline{C}$, we define its modulus of continuity

$$\omega_f(h) = \sup\{|f(x) - f(y)| : \rho(x, y) \leq h\}. \tag{1.2.5}$$

Let $\omega(h)$ be a *fixed* non-decreasing function on $\mathbb{R}^+$, such that $\omega(h) \to 0$ as $h \to 0$. Set

$$C_\omega = \{f : \|f\|_\infty < \infty, \text{ and } \omega_f(h) = O(\omega(h) + h)\}.$$

(Usually, $h = O(\omega(h))$, and one can write just $\omega_f(h) = O(\omega(h))$. However, there are situations when $\omega(h)$ even equals zero for sufficiently small h's; for example, if $\mathbb{H} = \mathbb{N}$ with the usual metric.)

The next proposition, having its intrinsic value, plays an essential role in proving Theorem 1.

Theorem 4 *Let*

$$\int f d (P_n - Q_n) \to 0 \tag{1.2.6}$$

for all $f \in C_\omega$, and let

$$\mathcal{F}_\omega = \{f : \|f\|_\infty < 1, \ \text{and} \ \omega_f(h) \le \omega(h) \ \text{for all} \ h \ge 0\}.$$

Then

$$\sup_{f \in \mathcal{F}_\omega} \left| \int f d (P_n - Q_n) \right| \to 0. \tag{1.2.7}$$

Clearly, instead of the above class $\mathcal{F}_\omega$, one may consider a class of uniformly bounded (not necessarily by one) functions f such that $\omega_f(h) \le k_1 \omega(h) + k_2 h$, where k_1, k_2 are fixed constants. (Such a formal generalization follows from (1.2.7) just by replacing $\omega(h)$ by $k_1\omega(h) + k_2 h$.)

Corollary 5 *If (1.2.6) is true for all $f \in C_{BL}$, then*

$$\sup_{f \in \mathcal{F}} \left| \int f d (P_n - Q_n) \right| \to 0 \tag{1.2.8}$$

for any class $\mathcal{F}$ of uniformly bounded functions with uniformly bounded Lipschitz constants.

Corollary 6 *If (1.2.6) is true for all $f \in \overline{C}$, then (1.2.8) is true for any class $\mathcal{F}$ of uniformly bounded and uniformly equicontinuous functions.*

To derive the above corollaries from Theorem 4, one should set $\omega(h) = \omega_{\mathcal{F}}(h) := \sup_{f \in \mathcal{F}} \omega_f(h)$. (In the literature, there is no unity in definitions of equicontinuity: some authors define it pointwise; in other definitions, the word "uniformly" is redundant. When talking about uniform equicontinuity, we mean that the function $\omega_{\mathcal{F}}(h)$ is bounded and vanishing at the origin.)

2 Proofs

Proofs of Theorems 1 and 3 essentially use Corollary 5 from Theorem 4 and Theorem 2. We start with a proof of the latter theorem – the main assertions of this theorem, A and B, are known ([5, Sec.11.7]), and the rest of the proof is short. The proof of Theorem 4 is relegated to the last Section 2.2.

2.1 Proofs of Theorems 1 – 3

2.1.1 Proof of Theorem 2

To justify Property C, consider the r.v.'s X_n, Y_n defined in Property B. We have $|f(X_n) - f(Y_n)| \leq \omega_f(\rho(X_n, Y_n))$. Consequently, $f(X_n) - f(Y_n) \xrightarrow{P} 0$, and it suffices to appeal to Property A.

Property D is obvious. Now, let $r_1(\cdot, \cdot)$ and $r_2(\cdot, \cdot)$ be two metrics with Properties A-B. If $r_1(P_n, Q_n) \to 0$, then there exist X_n, Y_n for which $\rho(X_n, Y_n) \to 0$, and hence by Property A, $r_2(P_n, Q_n) \to 0$. ■

For proving Theorems 1 and 3, we need

2.1.2 Two lemmas

Lemma 7 *If* $\pi(P_n, Q_n) \to 0$, *then*

$$\sup_{f \in \mathcal{F}} \int f d(P_n - Q_n) \to 0 \tag{2.1.1}$$

for any class $\mathcal{F}$ of uniformly bounded and uniformly equicontinuous functions.

Proof. By Theorem 2, there exist X_n, Y_n such that $P_{X_n} = P_n$, $Q_n = P_{Y_n}$, and $\rho(X_n, Y_n) \xrightarrow{P} 0$. By conditions of the lemma, $M := \sup_{f \in \mathcal{F}} \|f\|_\infty < \infty$, and $\omega(h) := \sup_{f \in \mathcal{F}} \omega_f(h) \to 0$ as $h \to 0$. For any $\varepsilon > 0$,

$$
\begin{aligned}
\left| \int f d(P_n - Q_n) \right| &= |E\{f(X_n) - f(Y_n)\}| \leq E\{|f(X_n) - f(Y_n)|\} \\
&= E\{|f(X_n) - f(Y_n)| \,;\, \rho(X_n, Y_n) > \varepsilon\} \\
&+ E\{|f(X_n) - f(Y_n)| \,;\, \rho(X_n, Y_n) \leq \varepsilon\} \\
&\leq 2MP(\rho(X_n, Y_n) > \varepsilon) + \omega(\varepsilon).
\end{aligned}
$$

Hence,

$$\varlimsup_{n} \sup_{f \in \mathcal{F}} \left| \int f d(P_n - Q_n) \right| \leq \omega(\varepsilon) \to 0 \text{ as } \varepsilon \to 0. \quad ■$$

Lemma 8 *(Dudley [5, a part of Theorem 11.7.1]). Suppose (2.1.1) is true for* $\mathcal{F} = \mathcal{F}_1 := \{f : \|f\|_\infty \leq 1,\ \|f\|_L \leq 1\}$. *Then* $\pi(P_n, Q_n) \to 0$.

Proof. In [5], the proof of this fact is based on the relation $\pi \leq 2\sqrt{\beta}$, which is proved separately. For the completeness of the picture, we give a direct proof (which, in essence, is very close to the reasoning in [5, p.396]).

For a closed set K, and an $\varepsilon > 0$, set

$$
I_K^{\varepsilon}(x) = \begin{cases} 1 & \text{if } x \in K, \\ 1 - \frac{1}{\varepsilon}\rho(x, K) & \text{if } x \in K^{\varepsilon} \setminus K, \\ 0 & \text{otherwise.} \end{cases}
$$

(Here, K^{ε} is the ε-neighborhood of K.)

Since $\omega_{I_K^{\varepsilon}}(h) \leq \frac{1}{\varepsilon}h$, the family $\{I_K^{\varepsilon}(x) : K \text{ is closed}\} \subset \mathcal{F}_{\varepsilon} := \{f : \|f\|_{\infty} \leq 1, \|f\|_L \leq 1/\varepsilon\}$. Clearly, if (2.1.1) holds for $\mathcal{F} = \mathcal{F}_1$, then it holds for $\mathcal{F} = \mathcal{F}_{\varepsilon}$ for any $\varepsilon > 0$. Therefore,

$$
\Delta_n(\varepsilon) := \sup_K \left| \int I_K^{\varepsilon}(x)d(P_n - Q_n) \right| \to 0 \tag{2.1.2}
$$

for any $\varepsilon > 0$ as $n \to \infty$.

On the other hand,

$$
P_n(K) \leq \int I_K^{\varepsilon}(x)dP_n \leq \Delta_n(\varepsilon) + Q_n(K^{\varepsilon}). \tag{2.1.3}
$$

From (2.1.2) and (2.1.3), it follows that for sufficiently large n,

$$
P_n(K) \leq \varepsilon + Q_n(K^{\varepsilon}),
$$

which implies that $\pi(P_n, Q_n) \leq \varepsilon$. Since ε is arbitrary small, this means that $\pi(P_n, Q_n) \to 0$. ∎

2.1.3 Proof of Theorem 1

The implication $\pi(P_n, Q_n) \to 0 \Rightarrow \int f d(P_n - Q_n) \to 0$ for any $f \in \overline{C}$, immediately follows from Lemma 7.

Assume $\int f d(P_n - Q_n) \to 0$ for any $f \in \overline{C}$. Then, the same is true for all $f \in C_{BL}$.

Hence, by Corollary 5 from Theorem 4, relation (2.1.1) holds for $\mathcal{F} = \mathcal{F}_1$ (defined in Lemma 8). This implies the convergence $\pi(P_n, Q_n) \to 0$ by virtue of Lemma 8. ∎

11

2.1.4 Proof of Theorem 3

Let $\int f d(P_n - Q_n) \to 0$ for any $f \in \mathcal{C}_{BL}$.

As was proved in Section 2.1.3 above, $\pi(P_n, Q_n) \to 0$. By virtue of Theorem 1, this implies that $\int f d(P_n - Q_n) \to 0$ for all $f \in \overline{\mathcal{C}}$. ∎

2.2 Proof of Theorem 4

2.2.1 Three more lemmas

Lemma 9 *For any two functions f and g,*

$$
\begin{array}{rcl}
\omega_{fg}(h) & \leq & ||g||_\infty \omega_f(h) + ||f||_\infty \omega_g(h), \qquad (2.2.1)\\
\omega_{f \vee g}(h) & \leq & \max\{\omega_f(h), \omega_g(h)\}, \qquad (2.2.2)\\
\omega_{f \wedge g}(h) & \leq & \max\{\omega_f(h), \omega_g(h)\}. \qquad (2.2.3)
\end{array}
$$

provided that the l.-h.sides are finite.

Proof is straightforward and very close to that in [5, Propositions 11.2.1-2] dealing with Lipschitz functions.

Next, for a function $f(x)$, a set K, and a number $t > 0$, we define the function

$$
f_K^{(t)}(x) = \begin{cases} f(x) & \text{if } x \in K, \\ f(x)\left(1 - \frac{\rho(x,K)}{t}\right) & \text{if } x \in K^t \setminus K, \\ 0 & \text{otherwise.} \end{cases} \qquad (2.2.4)
$$

Lemma 10 *Let f be a bounded uniformly continuous function. Then*

$$
\left\| f_K^{(t)} \right\|_\infty \leq \|f\|_\infty, \qquad (2.2.5)
$$

and for any $h \geq 0$,

$$
\omega_{f_K^{(t)}}(h) \leq \omega_f(h) + \|f\|_\infty \frac{h}{t}. \qquad (2.2.6)
$$

Proof. Bound (2.2.5) is obvious. Next, note that for any x and y,

$$
|\rho(x,K) - \rho(y,K)| \leq \rho(x,y). \qquad (2.2.7)
$$

(Indeed, for any $z \in K$,

$$
\rho(x,K) \leq \rho(x,z) \leq \rho(x,y) + \rho(y,z),
$$

12

which implies that $\rho(x,K) \le \rho(x,y) + \rho(y,K)$. We can also switch x and y.)

Thus, for $q_t(x) := 1 - \rho(x,K)/t$, we have

$$\omega_{q_t}(h) \le h/t.$$

Together with (2.2.1), this implies (2.2.6). ∎

Note that, in particular, from Lemma 10 it follows that if f is uniformly continuous, so does $f_K^{(t)}$.

Let the signed measure $\Psi_n = P_n - Q_n$.

Lemma 11 *Suppose*

$$\int f d\Psi_n \to 0 \qquad (2.2.8)$$

for any $f \in \overline{C}$. Let $\mathcal{F}$ be a class of uniformly bounded and uniformly equicontinuous functions. Set

$$\omega(h) := \sup_{f \in \mathcal{F}} \omega_f(h). \qquad (2.2.9)$$

Then for any compact K, and $t > 0$,

$$\varlimsup_n \sup_{f \in \mathcal{F}} \left| \int f_K^{(t)} d\Psi_n \right| \le 4\omega(t). \qquad (2.2.10)$$

Proof. By the Arzelà-Ascoli theorem, for any $\varepsilon > 0$, there is a finite family $\{f_1, ..., f_d\} \subset \mathcal{F}$ such that for any $f \in \mathcal{F}$, there exists $s = s(f) \in \{1, ..., d\}$, for which

$$\sup_{x \in K} |f(x) - f_s(x)| < \varepsilon. \qquad (2.2.11)$$

On the other hand, for any $z \in K$, and $x \in K^t$,

$$
\begin{aligned}
|f_K^{(t)}(x) - f_{s,K}^{(t)}(x)| &= |f(x) - f_s(x)|\left(1 - \frac{\rho(x,K)}{t}\right) \\
&\le |f(x) - f(z)| + |f_s(x) - f_s(z)| + |f(z) - f_s(z)| \\
&\le 2\omega(\rho(x,z)) + \varepsilon.
\end{aligned}
$$

Hence, for $x \in K^t$,

$$|f_K^{(t)}(x) - f_{s,K}^{(t)}(x)| \le 2\omega(t) + \varepsilon. \qquad (2.2.12)$$

13

Therefore,

$$\left| \int f_K^{(t)} d\Psi_n \right| \leq \left| \int (f_K^{(t)}(x) - f_{s,K}^{(t)}(x)) d\Psi_n \right| + \left| \int f_{s,K}^{(t)}(x) d\Psi_n \right|$$

$$\leq 2(2\omega(t) + \varepsilon) + \left| \int f_{s,K}^{(t)}(x) d\Psi_n \right|$$

$$\leq 4\omega(t) + 2\varepsilon + \max_{m \in \{1,\dots,d\}} \left| \int f_{m,K}^{(t)}(x) d\Psi_n \right|.$$

Since

$$\max_{m \in \{1,\dots,d\}} \left| \int f_{m,K}^{(t)}(x) d\Psi_n \right| \to 0 \text{ as } n \to \infty, \tag{2.2.13}$$

and ε is arbitrary small, this implies (2.2.10). ∎

We turn to

2.2.2 A direct proof of Theorem 4

Consider a fixed function $\omega(h)$, and the class $\mathcal{F} = \mathcal{F}_\omega$ from the statement of Theorem 4. Assume that there exist a sequence $\{f_1, f_2, \dots\} \subset \mathcal{F}$, a sequence $\{m_k\}$, and a $\delta > 0$, such that $|\int f_k d\Psi_{m_k}| \geq \delta$ for all $k = 1, 2, \dots$. Without loss of generality we can identify $\{m_k\}$ with $\mathbb{N}$, and write just

$$\left| \int f_k d\Psi_k \right| \geq \delta \text{ for all } k = 1, 2, \dots.$$

We may also assume that for all k's,

$$0 \leq f_k(x) \leq 1. \tag{2.2.14}$$

Now, let t and $\varepsilon \leq t$ be two *fixed* positive numbers to be chosen later. Let $n_1 = 1$, and K_1 be a compact such that

$$|\Psi_{n_1}|(K_1^{\complement}) \leq t.$$

(Here the measure $|\Psi|(dx)$ is the variation of $\Psi(dx)$, and $K^{\complement}$ is the complement of K.)

Let $n_2 \geq n_1 + 1$ be a number such that for all $n \geq n_2$

$$\left| \int f_{1,K_1}^{(\varepsilon/3)} d\Psi_n \right| \leq t, \tag{2.2.15}$$

14

and

$$\sup_j \left| \int f_{j,K_1}^{(t)} d\Psi_n \right| \leq 5\omega(t). \tag{2.2.16}$$

Inequality (2.2.15) is true for sufficiently large n because $f_{1,K_1}^{(\varepsilon/3)}$ is uniformly continuous due to Lemma 10, and (2.2.16) holds for large n by virtue of (2.2.10). Note also that in (2.2.16) we deal with the supremum over a class of functions, while (2.2.15) concerns *only one fixed* function.

Next, we consider a compact K_2 such that $K_2 \supset K_1$, and

$$|\Psi_{n_2}|(K_2^{\complement}) \leq t.$$

Let us set $\tilde{f}_1 = f_1$, and

$$\tilde{f}_2 = f_{n_2} - f_{n_2,K_1}^{(t)}.$$

By virtue of (2.2.16),

$$\int \tilde{f}_2 d\Psi_{n_2} = \int f_{n_2} d\Psi_{n_2} - \int f_{n_2,K_1}^{(t)} d\Psi_{n_2} \geq \delta - 5\omega(t).$$

Also,

$$\|\tilde{f}_2\|_\infty \leq 1, \quad \text{and} \quad \tilde{f}_2(x) = 0 \text{ for all } x \in K_1.$$

By Lemma 10,

$$\omega_{\tilde{f}_2}(h) \leq \omega_{f_{n_2}}(h) + \omega_{f_{n_2,K_1}^{(t)}}(h) \leq 2\omega(h) + \frac{h}{t}. \tag{2.2.17}$$

Now, we set $L_1 = K_1$, and $L_2 = K_2 \setminus K_1^\varepsilon$. Let

$$g_1(x) = \tilde{f}_{1,K_1}^{(\varepsilon/3)}(x), \quad \text{and} \quad g_2(x) = g_1(x) + \tilde{f}_{2,L_2}^{(\varepsilon/3)}(x). \tag{2.2.18}$$

By construction,

$$g_1(x) = 0 \text{ for } x \notin K_1^{\varepsilon/3}, \text{ and } g_2(x) = 0 \text{ for } x \notin K_2^{\varepsilon/3}.$$

By Lemma 10,

$$\omega_{g_1}(h) \leq \omega(h) + \frac{h}{\varepsilon/3} = \omega(h) + \frac{3h}{\varepsilon}. \tag{2.2.19}$$

15

Now, since the sets $L_1^{\varepsilon/3}$ and $L_2^{\varepsilon/3}$ are disjoint, in (2.2.18), either $g_1(x)$ or $\tilde{f}_{2,L_2}^{(\varepsilon/3)}(x)$ equals zero. So, we can also write that $g_2(x) = \max\{g_1(x),\, \tilde{f}_{2,L_2}^{(\varepsilon/3)}(x)\}$. Then, from Lemma 9, Lemma 10, (2.2.19), and (2.2.17), it follows that

$$\omega_{g_2}(h) \le \max\{\omega_{g_1}(x),\, \omega_{\tilde{f}_2}(h) + \frac{h}{\varepsilon/3}\} \le 2\omega(h) + \frac{h}{t} + \frac{3h}{\varepsilon}. \qquad (2.2.20)$$

In view of (2.2.19), bound (2.2.20) is true for both functions, g_1 and g_2.

Thus, both functions, g_1 and g_2, are bounded and uniformly continuous (since $\varepsilon, t > 0$ are fixed).

Next, we choose $n_3 \ge n_2 + 1$ such that for all $n \ge n_3$,

$$\left| \int g_2(x) d\Psi_n \right| \le t,$$

and

$$\sup_j \left| \int f_{j,K_2}^{(t)} d\Psi_n \right| \le 5\omega(t).$$

Let K_3 be a compact such that $K_3 \supset K_2$, and

$$|\Psi_{n_3}|(K_3^{\complement}) \le t.$$

We define a function

$$\tilde{f}_3 = f_{n_3} - f_{n_3,K_2}^{(t)}$$

which has properties similar to those of $\tilde{f}_2$, and we define the set

$$L_3 = K_3 \setminus K_2^{\varepsilon}.$$

The sets $L_1^{\varepsilon/3}$, $L_2^{\varepsilon/3}$, and $L_3^{\varepsilon/3}$ are mutually disjoint. We set

$$g_3(x) = g_2(x) + \tilde{f}_{3,L_3}^{(\varepsilon/3)}(x), \qquad (2.2.21)$$

and again note that in (2.2.21), either $g_2(x)$ or $\tilde{f}_{2,L_3}^{(\varepsilon/3)}(x)$ equals zero. Similarly to what we did above, we conclude that (2.2.20) is true for g_3 also. So, $g_3(x)$ is fixed, uniformly continuous, and $g_3(x) = 0$ for $x \notin K_3^{\varepsilon/3}$.

Continuing the recurrence procedure in the same fashion, we come to the following objects.

(a) The sequence $n_m \to \infty$.

16

(b) The sequence of compacts K_m such that for all m

$$|\Psi_{n_m}|(K_m^\complement) \le t. \tag{2.2.22}$$

(c) The sequence of compact sets $L_m \subset K_m$ such that the sets $L_m^{\varepsilon/3}$ are disjoint.

(d) The sequence of functions $\tilde{f}_m$ such that for all m

$$\tilde{f}_m(x) = 0 \text{ for all } x \in K_{m-1}, \tag{2.2.23}$$

$$\omega_{\tilde{f}_m}(h) \le 2\omega(h) + \frac{h}{t}, \tag{2.2.24}$$

and

$$\int \tilde{f}_m d\Psi_{n_m} \ge \delta - 5\omega(t). \tag{2.2.25}$$

(e) The non-decreasing sequence $\{g_1(x) \le g_2(x) \le ...\}$ such that $g_m(x) = g_{m-1}(x) + \tilde{f}_{m,L_m}^{(\varepsilon/3)}(x)$,

$$\left| \int g_{m-1}(x) d\Psi_{n_m} \right| \le t, \tag{2.2.26}$$

$g_m(x) = 0$ for $x \notin K_m^{\varepsilon/3}$, and

$$\omega_{g_m}(h) \le 2\omega(h) + \frac{h}{t} + \frac{3h}{\varepsilon}.$$

Let

$$g(x) = \lim_{m\to\infty} g_m(x).$$

Clearly, $0 \le g(x) \le 1$, and

$$\omega_g(h) \le 2\omega(h) + \frac{h}{t} + \frac{3h}{\varepsilon}. \tag{2.2.27}$$

Since the numbers ε and t are fixed, the function $g \in C_\omega$. We show that, nevertheless, one can choose ε and t such that $I_n := \int g d\Psi_n \not\to 0$.

To make exposition simpler, we replace the sequence $\{n_m\}$ by $\mathbb{N}$, write Ψ_n instead of Ψ_{n_m}, and remove $\tilde{\ }$ from $\tilde{f}$'s. All of this cannot cause misunderstanding.

By virtue of (2.2.25),

$$I_n = \int f_n d\Psi_n + \int (g - f_n) d\Psi_n \ge \delta - 5\omega(t) + \int (g - f_n) d\Psi_n.$$

17

For $J_n := \int (g - f_n) d\Psi_n$, we write

$$|J_n| \leq \left| \int_{K_n} (g - f_n) d\Psi_n \right| + |\Psi_n|(K_n^{\complement}) \leq \left| \int_{K_n} (g - f_n) d\Psi_n \right| + t.$$

Now,

$$\int_{K_n} (g - f_n) d\Psi_n = \int_{L_n} + \int_{K_n \cap K_{n-1}^{\varepsilon/3}} + \int_{K_n \setminus (L_n \cup K_{n-1}^{\varepsilon/3})} := J_{n1} + J_{n2} + J_{n3}.$$

By construction, $J_{1n} = 0$. For the second integral, we have

$$
\begin{aligned}
|J_{n2}| &= \left| \int_{K_{n-1}^{\varepsilon/3}} (g - f_n) d\Psi_n - \int_{K_n^{\complement} \cap K_{n-1}^{\varepsilon/3}} (g - f_n) d\Psi_n \right| \\
&\leq \left| \int_{K_{n-1}^{\varepsilon/3}} (g - f_n) d\Psi_n \right| + \int_{K_n^{\complement}} |\Psi_n|(dx) \\
&\leq \left| \int_{K_{n-1}^{\varepsilon/3}} g \, d\Psi_n \right| + \left| \int_{K_{n-1}^{\varepsilon/3}} f_n \, d\Psi_n \right| + t,
\end{aligned}
\tag{2.2.28}
$$

in view of (2.2.22).

By construction, $g(x) = g_{n-1}(x)$ for $x \in K_{n-1}^{\varepsilon/3}$. So,

$$\left| \int_{K_{n-1}^{\varepsilon/3}} g \, d\Psi_n \right| = \left| \int_{K_{n-1}^{\varepsilon/3}} g_{n-1} d\Psi_n \right| = \left| \int g_{n-1} d\Psi_n \right| \leq t, \tag{2.2.29}$$

by virtue of (2.2.26).

In view of (2.2.23) and (2.2.24),

$$
\begin{aligned}
\left| \int_{K_{n-1}^{\varepsilon/3}} f_n \, d\Psi_n \right| &= \left| \int_{K_{n-1}^{\varepsilon/3} \setminus K_{n-1}} f_n \, d\Psi_n \right| \leq \omega_{f_n}(\varepsilon/3) \int |\Psi_n| (dx) \\
&\leq \left(2\omega(\varepsilon/3) + \frac{\varepsilon/3}{t} \right) 2 = 4\omega(\varepsilon/3) + \frac{2\varepsilon}{3t}.
\end{aligned}
$$

Thus,

$$|J_{n2}| \leq 4\omega(\varepsilon/3) + \frac{2\varepsilon}{3t} + 2t. \tag{2.2.30}$$

To evaluate J_{n3}, first note that if $x \in K_n \setminus (L_n \cup K_{n-1}^{\varepsilon/3})$, then $x \in K_{n-1}^{\varepsilon}$, and hence

$$\left| \int_{K_n \setminus (L_n \cup K_{n-1}^{\varepsilon/3})} f_n(x) d\Psi_n \right| \leq \left| \int_{K_n \setminus (L_n \cup K_{n-1}^{\varepsilon/3})} \omega_{f_n}(\varepsilon) |\Psi_n|(dx) \right|$$

$$\leq \left(2\omega(\varepsilon) + \frac{\varepsilon}{t} \right) 2 = 4\omega(\varepsilon) + \frac{2\varepsilon}{t}. \quad (2.2.31)$$

Now, let us observe that if $x \in K_n \setminus (L_n \cup K_{n-1}^{\varepsilon/3})$ and $x \notin L_n^{\varepsilon/3}$, then $g(x) = 0$. So,

$$\int_{K_n \setminus (L_n \cup K_{n-1}^{\varepsilon/3})} g(x) d\Psi_n = \int_{(K_n \setminus L_n) \cap L_n^{\varepsilon/3}} g(x) d\Psi_n.$$

On the other hand, $(K_n \setminus L_n) \cap L_n^{\varepsilon/3} \subseteq K_{n-1}^{\varepsilon}$, and $g(x) \leq f_n(x)$ on $(K_n \setminus L_n) \cap L_n^{\varepsilon/3}$. Thus, for the function g, we have the bound similar to (2.2.31), and

$$|J_{n3}| \leq 8\omega(\varepsilon) + \frac{4\varepsilon}{t}.$$

Combining the bounds above, we have

$$|J_n| \leq 4\omega(\varepsilon/3) + 8\omega(\varepsilon) + \frac{14\varepsilon}{3t} + 3t \leq 12\omega(\varepsilon) + 5\frac{\varepsilon}{t} + 3t,$$

and

$$|I_n| \geq \delta - 5\omega(t) - 12\omega(\varepsilon) - 5\frac{\varepsilon}{t} - 3t \geq \delta - 17\omega(t) - 5\frac{\varepsilon}{t} - 3t,$$

because we choose $\varepsilon \leq t$. Without loss of generality we can take $t < 1$. Let $\varepsilon = t^2$. Then

$$|I_n| \geq \delta - 17\omega(t) - 8t.$$

Clearly, one can choose t for which $|I_n| \geq \frac{\delta}{2}$ for all n. ∎

References

[1] Chatterjee, S., (2006). *A generalization of the Lindeberg principle,* Ann. Probab., 34 no. 6, 2061-2076.

[2] D'Aristotile, A., Diaconis, P., and Freedman, D. (1988), *On merging of probabilities*, Sankhyǎ: The Indian Journal of Statistics, v.58, Series A, Pt.3, pp. 363-380.

[3] Dudley, R. M., (1968). *Distances of probability measures and random variables*, Ann. Math. Statist. 39, 1563-1572.

[4] Dudley, R. M., (1989). *Real Analysis and Probability*, 1st edition, Wadsworth, Inc.

[5] Dudley, R. M., (2002). *Real Analysis and Probability*, 2nd edition, Cambridge University Press.

[6] Davydov, Yu.A., and Rotar, V.I., (2007). *On a non-classical invariance principle*, http://arxiv.org/math.PR/0702085.

[7] Götze, F., and Tikhomirov, A. N. (1999). *Asymptotic distribution of quadratic forms*, Ann. Probab. 27 , no. 2, 1072–1098.

[8] Götze, F., and Tikhomirov, A. (2002). *Asymptotic distribution of quadratic forms and applications,* J. Theoret. Probab. 15, no. 2, 423–475.

[9] Liptser, R.Sh., and Shiryaev, A.N. (1983), On the invariance principle for semi-martingales: the "non-classical" case, Theory Probabl. Appl., XXVIII, 1, 1-34.

[10] Loève, M., (1963). *Probability Theory*, 3rd edition, Princeton, N.J., Van Nostrand.

[11] Mossel, E., O'Donnel, R., and Oleszkiewicz, K., (2005). *Noise stability of functions with low influences: invariance and optimality*, to appear, http://arxiv.org/math.PR/0503503.

[12] Prokhorov, Yu.V., (1956). *Convergence of random processes and limit theorems in Probability Theory*, Theory Probabl. Appl., I , 2, 157-214.

[13] Rotar, V.I., (1975). *Limit theorems for multilinear forms and quasi-polynomial functions*, Theory Probabl. Appl., XX, 3, 512-532.

[14] Rotar, V.I., (1979). *Limit theorems for polylinear forms*, Journal of Multivariate analysis, 9, 4, 511-530.

[15] Rotar, V.I., (1982). *On summation of independent variables in the nonclassical situation*; Russian Mathematical Surveys, 37:6, 151-175.

[16] Skorohod, A.V., (1956). *Limit theorems for stochastic processes*, Theory Probabl. Appl., I , 3, 261-290.

[17] Strassen, V., (1965). *The existence of probability measures with given marginal*s, Ann. Math. Stat., 36, pp. 423–439.

[18] Szulga, A, (1982). *On minimal metrics in the space of random variables*, Teor. Probab. Appl., 27, pp. 424–430.

[19] Zolotarev, V.M., (1997). *Modern Theory of Summation of Random Variables*, V.S.P. Intern. Science Publishers.