Journal of Theoretical Probability, Vol. 5, No. 1, 1992

Uniformities for the Convergence in Law and in Probability

S. T. Rachev,[1] L. Rüschendorf,[2] and A. Schief[3]

Received August 6, 1990; revised April 24, 1991

The convergence $\delta(X_n, Y_n) \to 0$ is investigated and characterized for probability metrics δ which metrize convergence in distribution or in probability. Some related metrics are also considered.

KEY WORDS: Prohorov type metric; Ky-Fan-type metric; almost sure representation; uniformities.

1. NOTATION

Let (S, d) be a separable metric space with Borel σ-field $\mathfrak{B}$, and let $M(S)$ denote the set of Borel probability measures on S. Define for $\mu, v \in M(S)$, $0 < \lambda < \infty$, the Prohorov-type metrics:

$$\pi_\lambda(\mu, v) = \inf\{\varepsilon > 0; \mu(F) \leqslant v(F^{\lambda\varepsilon}) + \varepsilon \text{ for all } F \in \mathfrak{F}(S)\}$$

$$\pi_0(\mu, v) = \sup\{|\mu(F) - v(F)|; F \in \mathfrak{F}(S)\}$$

$$\pi_\infty(\mu, v) = \inf\{\varepsilon > 0; \mu(F) \leqslant v(F^\varepsilon) \text{ for all } F \in \mathfrak{F}(S)\}$$

[$\mathfrak{F}(S)$ is the collection of closed sets in S and $F^\varepsilon = \{x \in S;\ d(x, F) < \varepsilon\}$.] Note that π_0, π_∞ are limiting cases of π_λ in the sense that $\pi_\lambda(\mu, v) \xrightarrow[\lambda \to 0]{} \pi_0(\mu, v)$ and $\lambda\pi_\lambda(\mu, v) \xrightarrow[\lambda \to \infty]{} \pi_\infty(\mu, v)$, so the class of metrics π_λ connects the supremum metric π_0 and the π_∞ metric, thus changing the topological structure extremely in the limits.

[1] Statistics Department, University of California, Santa Barbara, California 93106.

[2] Institut für Mathematische Statistik, Universität Münster, Einsteinstrasse 62, D-4400 Münster, Germany.

[3] Mathematisches Institut, Universität München, Theresienstrasse 39, D-8000 München, Germany.

860/5/1-3

0894-9840/92/0100-0033$06.50/0 © 1992 Plenum Publishing Corporation

Define, furthermore, for S-valued random variables X, Y on $(\Omega, \mathfrak{U}, P)$ and $0 \leqslant \lambda < \infty$ the Ky-Fan type metrics:

$$K_\lambda(X, Y) = \inf\{\varepsilon > 0; P(d(X, Y) > \lambda\varepsilon) < \varepsilon\}$$

$$K_\infty(X, Y) = \text{ess sup } d(X, Y)$$

Similarly, for the compound metrics K_λ we obtain a limit version as $\lambda \to \infty$ by

$$\begin{aligned}\lim_{\lambda \to \infty} \lambda K_\lambda(X, Y) &= \lim_{\lambda \to \infty} \inf\{\varepsilon > 0\colon P(d(X, Y) > \varepsilon) < \varepsilon/\lambda\}\\ &= \text{ess sup } d(X, Y) = K_\infty(X, Y)\end{aligned}$$

For each $\lambda \in [0, \infty]$, π_λ is the minimal metric corresponding to K_λ, i.e.,

$$\pi_\lambda(\mu, \nu) = \inf\{K_\lambda(X, Y); X \text{ and } Y \text{ have distributions } \mu \text{ and } \nu \text{, respectively}\}$$

(see Refs. 1, 4, and 8). The representation of π_λ as minimal metric of K_λ is for $0 < \lambda < \infty$ a particular case of the Strassen–Dudley theorem (cf. Refs. 2 and 4) implying the representation for the cases $\lambda = 0$ and $\lambda = \infty$ by the limiting relations above (cf. Refs. 1 and 8). Alternatively, the case $\lambda = \infty$ can be inferred also from the following equivalence (cf. Ref. 2): For $\varepsilon > 0$ (1) there exists a probability Q on $S \times S$ with marginals μ, ν, and $Q\{(x, y); d(x, y) > \varepsilon\} = 0$ if and only if (2) For all closed sets F, $\mu(F) \leqslant \nu(F^\varepsilon)$. Note that $\pi = \pi_1$ is the Prohorov metric, $K = K_1$ is the Ky-Fan metric.

The arguments given above show that the consideration of the parametric versions π_λ, K_λ is useful in order to obtain information on the limiting cases π_0, π_∞. As an example for this interrelation, consider the question of whether on $\mathbb{R}^1$ for some nondecreasing function φ with $\varphi(0+) = \varphi(0) = 0$, $\pi \leqslant \varphi(L)$, where L is the Levy metric. Introduce the parametric version L_λ of L by

$$\begin{aligned}L_\lambda(X, Y) = L_\lambda(F_X, F_Y) = \inf\{\varepsilon > 0; F_X(x - \lambda\varepsilon) - \varepsilon\\ \leqslant F_Y(x) \leqslant F_X(x + \lambda\varepsilon) + \varepsilon, \forall x \in \mathbb{R}^1\}\end{aligned}$$

Assume that $\pi \leqslant \varphi(L)$ with φ as above, then

$$\pi_\lambda(X, Y) = \pi\left(\frac{1}{\lambda} X, \frac{1}{\lambda} Y\right) \leqslant \varphi\left(L\left(\frac{1}{\lambda} X, \frac{1}{\lambda} Y\right)\right) = \varphi(L_\lambda(X, Y))$$

For $\lambda \to 0$ the above inequality leads to the obvious contradiction $\pi_0 \leqslant \varphi(\rho)$, where $\rho = L_0 = \lim_{\lambda \to 0} L_\lambda$ is the uniform distance between d.f.'s, since π_0 is topologically stronger than ρ. Note however that, for $\lambda \to \infty$, λL_λ has the same limit as $\lambda \pi_\lambda$, namely,

$$\lim_{\lambda \to \infty} \lambda L_\lambda(X, Y) = \pi_\infty(X, Y) = \sup_{0 \leqslant t \leqslant 1} |F_X^{-1}(t) - F_Y^{-1}(t)|$$

where F_X^{-1}, F_Y^{-1} denote the generalized inverses of F_X, F_Y.

The aim of this paper is to investigate the uniformity structure of the metrics π_λ (Section 2) and the corresponding compound metrics K_λ. More precisely we consider on the one hand side a Skorohod type a.s. convergence result [in the noncompact case, i.e., $\pi_\lambda(\mu_n, \nu_n) \to 0$] and on the other hand the question of convergence uniformly on some Lipschitz classes of functions. Some basic results and methods concerning these questions were developed in Dudley.[2] Later extensions of these results were given in Refs. 3, 4, 10, and 11 (cf. also the references within these papers).

Finally for $c: S \times S \to \mathbb{R}_+$, a nonnegative product measurable cost function on S, define the Kantorovich functional:

$$\mu_c(\mu, \nu) = \inf\{Ec(X, Y); X \text{ and } Y \text{ have distributions } \mu \text{ and } \nu, \text{ respectively}\}$$

which is the basic functional in transportation problems.

2. CONVERGENCE IN DISTRIBUTION

In this section we investigate the convergence $\pi_\lambda(\mu_n, \nu_n) \to 0$. The classical Skorohod–Dudley–Wichura theorem states that $\pi_1(\mu_n, \mu) \to 0$ is equivalent to the existence of S-valued random variables X_n, X on some probability space $(\Omega, \mathfrak{U}, P)$ such that $d(X_n, X) \to 0$ P-a.s. and X_n, X have distributions μ_n, μ, respectively. For extensions of this result concerning the underlying assumptions, see Refs. 2–4 and 11. While this a.s. representation characterizes the topological structure (compact case), the following theorem concerns the corresponding a.s. characterization of the uniformity structure of the metrics π_λ (noncompact case). Part (a) of the theorem has been proved already before by Dudley[4] and, independently, by Rachev, Rüschendorf and Schief.[10]

Theorem 1. Let $\mu_n, \nu_n \in M(S)$, $\lambda \in [0, \infty]$. Then $\pi_\lambda(\mu_n, \nu_n) \to 0$ if and only if there exist S-valued r.v.'s X_n, Y_n on a probability space $(\Omega, \mathfrak{U}, P)$, such that X_n, Y_n have distributions μ_n, ν_n, respectively, and

(a) $d(X_n, Y_n) \to 0$ in P-probability, if $\lambda \in (0, \infty)$

(b) $1_{\{X_n \neq Y_n\}} \to 0$ in P-probability, if $\lambda = 0$

(c) $\operatorname{ess\,sup} d(X_n, Y_n) \to 0$ if $\lambda = \infty$

The equivalence is also valid if "in P-probability" is replaced by "P-almost surely."

The proof of Theorem 1 is based on the following general realization lemma.

Lemma 1. Let $f: S \to \mathbb{R}$ be a measurable function and ξ_n, $n \in \mathbb{N}$, a sequence of S-valued r.v.'s on $(\Omega, \mathfrak{U}, P)$ such that $f \circ \xi_n \to 0$ in P-probability. Then there exists a sequence ξ'_n on a probability space $(\Omega', \mathfrak{U}', P')$ such that $P^{\xi_n} = P'^{\xi'_n}$ for each $n \in \mathbb{N}$ and

$$f \circ \xi'_n \to 0 \ P'\text{-almost surely}$$

Proof. The method of proof is similar to some arguments used in Refs. 2–4 and 11. Choose $N_k \uparrow \infty$ such that for all $n \geqslant \mathbb{N}_k$

$$P\left(|f \circ \xi'_n| > \frac{1}{k} \right) \leqslant \frac{1}{k}$$

$(N_1 = 1)$. For $n \in \mathbb{N}$ let k_n be the largest integer such that $n \geqslant N_{k_n}$. This implies $k_n \uparrow \infty$ and

$$P(A_n) \leqslant \frac{1}{k_n} \qquad \text{for} \quad A_n = \left\{ |f \circ \xi'_n| > \frac{1}{k_n} \right\}$$

Define for $n \in \mathbb{N}$, $t \in]0, 1[$ a probability on $(\Omega, \mathfrak{U})$

$$P'_{n,t} := \begin{cases} P(\cdot \mid A_n^c) & \text{if} \quad t \leqslant P(A_n^c) \\ P(\cdot \mid A_n) & \text{if} \quad t > P(A_n^c) \end{cases}$$

On $\Omega' = \bigotimes_{n \in \mathbb{N}} \Omega$, $\mathfrak{U}' = \bigotimes_{n \in \mathbb{N}} \mathfrak{U}$ set

$$P'_t = \bigotimes_{n \in \mathbb{N}} P'_{n,t}, \qquad P' = \int P'_t \lambda(dt)$$

and

$$\xi'_n = \xi_n \circ \pi_n$$

where π_n denotes the projection on the nth coordinate. We get

$$P'^{\xi'_n} = \int P'^{\xi_n}_{n,t} \lambda(dt) = P^{\xi_n}$$

Furthermore for every $t \in]0, 1[$, $t \leqslant P(A_n^c)$ for n large enough, implying $P'_t(|f \circ \xi'_n| \leqslant 1/k_n) = 1$ eventually, which yields

$$P'_t(f \circ \xi'_n \to 0) = 1 \qquad \text{for all } t \in]0, 1[$$

and obviously

$$P'(f \circ \xi_n' \to 0) = 1 \qquad \square$$

Proof of Theorem 1.

a. Since for $0 < \lambda < \infty$, $\min(1, 1/\lambda)\pi \leqslant \pi_\lambda \leqslant \max(1, 1/\lambda)\pi$ the metrics π_1, π_λ are equivalent. Therefore, only the case $\lambda = 1$ has to be considered. As remarked above, a proof in this case was already given in Refs. 4 and 10. Based on Lemma 1, the proof is as follows. Since π is the minimal metric of K, there exist random variables X_n, Y_n on a suitable probability space $(\Omega, \mathfrak{U}, P)$ such that $P^{X_n} = \mu_n$, $P^{Y_n} = \nu_n$, and $K(X_n, Y_n) \to 0$. Apply Lemma 1 for $\xi_n = (X_n, Y_n)$, $f = d$.

b. Since π_0 is the minimal metric of K_0, there exist probability measures Q_n on $(S \times S, \mathfrak{B} \otimes \mathfrak{B})$ with marginals μ_n, ν_n such that $Q_n(\{(x, y) \in S^2; x \neq y\}) \to 0$. Let

$$(\Omega, \mathfrak{U}, Q) = \left((S^2)^{\mathbb{N}}, (\mathfrak{B} \otimes \mathfrak{B})^{\mathbb{N}}, \bigotimes_{n \in \mathbb{N}} Q_n\right)$$

and let (X_n, Y_n) be the projection on the nth component; then $Q(X_n \neq Y_n) \to 0$. Apply Lemma 1 with

$$\xi_n = (X_n, Y_n) \qquad \text{and} \qquad f(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

c. Since π_∞ is the minimal metric of K_∞, there exists a sequence $\alpha_n \downarrow 0$, $\alpha_n > \pi_\infty(\mu_n, \nu_n)$, and measures $Q_n \in M(S^2)$ with marginals μ_n, ν_n such that

$$Q_n(\{(x, y); d(x, y) > \alpha_n\}) = 0$$

With $(\Omega, \mathfrak{U}, Q)$ as in the proof of part b, we obtain

$$\operatorname*{ess\,sup}_Q d(X_n, Y_n) \leqslant \alpha_n \to 0$$

The other implication is obvious. $\square$

Remark 1.

a. Theorem 1 is not true for the Levy-metric L on $S = \mathbb{R}$. Let $\mu_n(\{2j\}) = \nu_n(\{2j+1\}) = 1/n$, $1 \leqslant j \leqslant n$. Then $L(\mu_n, \nu_n) = 1/n$ but $\pi(\mu_n, \nu_n) = 1$ for all n (cf. Ref. 2). Therefore, L and π are not equivalent in the noncompact case.

b. Theorem 1 holds for many relevant probability metrics. We give some examples: for $\sigma(\mu, \nu) = \sup\{|\mu(f) - \nu(f)|;\ f \in L(S)\}$, $\beta(\mu, \nu) = \sup\{|\mu(f) - \nu(f)|;\ f \in \tilde{L}(S)\}$, $(L(S) = \{f: S \to [0, 1];\ |f(x) - f(y)| \leqslant d(x, y)$ for all $x, y \in S\}$, $\tilde{L}(S) = \{f: S \to [0, 1];\ \sup_{x \neq y} |f(x) - f(y)|/[d(x, y)] + \sup_x |f(x)| \leqslant 1\})$; the inequalities $\frac{1}{2}\sigma \leqslant \pi \leqslant \sqrt{\sigma}$ and $\frac{2}{3}\pi^2 \leqslant \beta \leqslant 2\pi$ (Ref. 8) ensure that Theorem 1 is applicable. These inequalities imply uniformity of convergence on classes of bounded Lipschitz functions (cf. also Refs. 2 and 4).

c. One direction of Theorem 1 is valid for the minimal L_p-metric $\hat{L}_p(\mu, \nu) = \inf\{(Ed^p(X, Y))^{1/p};\ X, Y$ have distributions $\mu, \nu\}$. Since $\pi \leqslant (\hat{L}_p)^{p/(1+p)}$, $\hat{L}_p(\mu_n, \nu_n) \to 0$ implies the existence of a.s. constructions. If the metric d is bounded, both directions of Theorem 1 are true for $\hat{L}_p$. Some further examples are discussed in Ref. 10.

d. The above implication is a special case $[U = M(S) \times M(S),\ \mu = \pi,\ \nu = \hat{L}_p]$ of the following:

Proposition 1. Suppose μ and ν are two mappings of a space U into $[0, \infty]$. Then the following are equivalent:

i. $\nu(u_n) \to 0 \Rightarrow \mu(u_n) \to 0$.

ii. There exists a nondecreasing function $\psi: [0, \infty] \to [0, \infty]$, $\psi(0) = \psi(0+) = 0$ and such that $\mu \leqslant \psi(\nu)$.

Moreover, if $\inf\{\nu(u): u \in U\} = 0$, then (ii) is equivalent to the following:

iii. There exists a nondecreasing function $\varphi: (0, \infty] \to (0, \infty]$, $\varphi(0) = 0$ and such that $\varphi(\mu) \leqslant \nu$.

Proof. (i) $\Rightarrow$ (ii). Take $\psi(x) = \sup\{\mu(u): u \in U,\ \nu(u) \leqslant x\}$, $x \geqslant 0$.
(i) $\Rightarrow$ (iii). Take $\varphi(x) = \inf\{\nu(u): u \in U,\ \mu(u) \geqslant x\}$, $x \geqslant 0$. □

Note that if ν and μ are semimetrics on $M(S)$, then

$$\nu(P_n, Q_n) \to 0 \Rightarrow \mu(P_n, Q_n) \to 0$$

is equivalent to (ii) and (iii). (In case of μ and ν being metrics this was shown in Ref. 8, Theorem 1.)

e. In the "compact convergence" case holds (cf. Ref. 9): Let $p \geqslant 1$, $\mu_n, \mu \in M(S)$ such that $m_{a,p}(\mu) = \int d^p(x, a)\, \mu(dx)$ and $m_{a,p}(\mu_n)$, $n \in \mathbb{N}$ exist for some $a \in S$. Then: $\hat{L}_p(\mu_n, \mu) \to 0$ iff there exist r.v.'s X_n, X on $(\Omega, \mathfrak{U}, P)$ with distributions μ_n, μ such that $d(X_n, X) \to 0$ in P-probability and $m_{a,p}(\mu_n) \to m_{a,p}(\mu)$.

The following counterexample shows that this equivalence cannot be extended to the case of "noncompact convergence." Let $p=2$, $S=\mathbb{R}$, and

$$\mu_n=\frac{1}{n}\varepsilon_{2\sqrt{n}}+\frac{n-1}{n}\varepsilon_{n\sqrt{n/(n-1)}}, \qquad \nu_n=\frac{1}{n}\varepsilon_{\sqrt{n}}+\frac{n-1}{n}\varepsilon_{n\sqrt{n/(n-1)}}$$

Then $\pi(\mu_n, \nu_n)\leqslant 1/n$ implying the existence of r.v.'s X_n, Y_n with distributions μ_n, ν_n such that $d(X_n, Y_n)\to 0$ a.s. Furthermore, $m_{0,2}(\mu_n)-m_{0,2}(\nu_n)=\sqrt{4+n^2}-\sqrt{1+n^2}\to 0$ [even: $m_{a,2}(\mu_n)-m_{a,2}(\nu_n)\to 0$, $\forall a$ holds]. But

$$\begin{aligned}\hat{L}_2^2(\mu_n, \nu_n) &\geqslant \inf\{E(|X_n-\sqrt{n}|^2\, 1_{\{Y_n=\sqrt{n}\}});\, X_n, Y_n \text{ have distributions } \mu_n, \nu_n\}\\ &\geqslant \frac{1}{n}(\sqrt{n})^2=1\end{aligned}$$

Theorem 1 can partially be extended to Kantorovich-functionals μ_c for general nonnegative cost functions c.

Theorem 2. Let $\mu_n, \nu_n \in M(S)$ and $\mu_c(\mu_n, \nu_n)\to 0$; then there exist S-valued random variables X_n, Y_n with distributions μ_n, ν_n on a probability space $(\Omega, \mathfrak{U}, P)$, such that $c(X_n, Y_n)\to 0$ P-a.s.

Proof. Let ζ_n, η_n be random variables on a probability space $(\Omega', \mathfrak{U}', Q)$ with distributions μ_n, ν_n such that $Z_n=c(\zeta_n, \eta_n)\to 0$ in Q-probability. Apply now Lemma 1 for $\xi_n=(X_n, Y_n)$ and $f=c$. □

In the case of a universally measurable space S a proof of Theorem 2 can also be based on the a.s. construction method developed in Ref. 9.

3. CONVERGENCE IN PROBABILITY

The convergence in probability is metrized by $K=K_1$. The a.s. construction result of Theorem 1 can be sharpened in this case.

Theorem 3. Let X_n, Y_n be S-valued random variables. Then the following are equivalent:

a. $K(X_n, Y_n)\to 0$.
b. There exist S-valued random variables $\tilde{X}_n$, $\tilde{Y}_n$ on a probability space $(\Omega, \mathfrak{U}, P)$ such that (X_n, Y_n), $(\tilde{X}_n, \tilde{Y}_n)$ are identically distributed and $d(\tilde{X}_n, \tilde{Y}_n)\to 0$ P-a.s.

Proof. Take $f=d$, $\xi_n=(X_n, Y_n)$ in Lemma 1. □

The question of uniformity of convergence on bounded Lipschitz functions turns out to be more difficult w.r.t. K than w.r.t. π in Section 2. To formulate this question more precisely, define $BL(S, d) = \{f: S \to [-1, 1]; |f(x) - f(y)| \leqslant d(x, y)$ for all $x, y \in S\}$, and for S-valued r.v.'s X, Y, define the metric $d_{BL}(X, Y) = \sup\{E\,|f(X) - f(Y)|; f \in BL(S, d)\}$. The following lemma shows that $BL(S, d)$ is a class of functions determining the uniformity.

Lemma 2. Let $x_n, y_n \in S$. Then the following are equivalent:

1. $d(x_n, y_n) \to 0$
2. $f(x_n) - f(y_n) \to 0$ for all $f \in BL(S, d)$.

Proof. $1 \Rightarrow 2$ is obvious. $2 \Rightarrow 1$: If $d(x_n, y_n) \not\to 0$, we may assume $d(x_n, y_n) \geqslant 1$ for all $n \in \mathbb{N}$.

1. *Case*: There exists $a \in S$, such that $|\{n \in \mathbb{N}; x_n \in K_{1/4}(a)\}| = \infty$, $K_\varepsilon(a)$ the ball of radius ε with center a. We may assume that $x_n \in K_{1/4}(a)$ for all $n \in \mathbb{N}$. This implies that $y_n \notin K_{1/2}(a)$ and $d(x_n, y_m) \geqslant \frac{1}{4}$ for all $n, m \in \mathbb{N}$. Set $f(x_n) = 0$, $f(y_m) = \frac{1}{4}$, and apply the Kirszbraun–McShane Theorem (cf. Ref. 4, Theorem 6.1.1) to finish the proof of the first case.
2. *Case*: Exchange x and y.
3. *Case*: None of the above.

Define $n_1 = 1$, $n_{k+1} = \max\{n \in \mathbb{N}; x_n \in \bigcup_{l=1}^{k} K_{1/4}(y_{n_l})$ or $y_n \in \bigcup_{l=1}^{k} K_{1/4}(x_{n_l})\} + 1$. Then for any $k \neq l$ we get $d(x_{n_k}, y_{n_l}) \geqslant \frac{1}{4}$ and we may assume for any m, n, $d(x_n, y_m) \geqslant \frac{1}{4}$. The proof is completed as above. □

We define for a subset F of the set of continuous functions: F is a uniformly determining class (UD-class) if F fulfills Lemma 2, i.e., $d(x_n, y_n) \to 0$ iff $|f(x_n) - f(y_n)| \to 0$ for all $f \in F$.

Theorem 4. Let X_n, Y_n be r.v.'s with values in S.

a. If $F \subset BL(S, d)$, then $K(X_n, Y_n) \to 0$ implies $d_F(X_n, Y_n) = \sup_{f \in F} E\,|f(X_n) - f(Y_n)| \to 0$.
b. If F is a countable UD-class, then $d_F(X_n, Y_n) \to 0$ implies $K(X_n, Y_n) \to 0$.

Proof. Point a: This follows from the inequality $d_F(X_n, Y_n) \leqslant d_{BL}(X_n, Y_n) \leqslant E \min\{d(X_n, Y_n), 2\}$. *Point b*: For any subsequence $(n_k) \subset \mathbb{N}$ choose a subsequence $(n_l) \subset (n_k)$ such that

$$\sum_{l=1}^{\infty} l d_F(X_{n_l}, Y_{n_l}) < \infty$$

For $f \in F$ the inequality $P(|f(X_{n_l}) - f(Y_{n_l})| > 1/l) \leqslant l d_F(X_{n_l}, Y_{n_l})$ implies by Borel–Cantelli that $f(X_{n_l}) - f(Y_{n_l}) \to 0$ P-a.s. (cf. Ref. 5, 1.11.8). Therefore (using that F is countable), $d(X_{n_l}, Y_{n_l}) \to 0$ P-a.s. and, thus, $d(X_n, Y_n) \to 0$ in P-probability. □

Corollary 1. If (S, d) is a compact metric space, then the following are equivalent:

a. $K(X_n, Y_n) \to 0$

b. $d_{BL}(X_n, Y_n) \to 0$

c. $E|f(X_n) - f(Y_n)| \to 0$, $\forall f \in BL(S, d)$.

Proof. For compact metric spaces (S, d), $(BL(S, d), \| \|_\infty)$ is by Arzela–Ascoli separable. So any dense, countable set $F \subset BL(S, d)$ is UD and the equivalence of (a) and (b) follows from Theorem 4. It remains to show that (c) ⇒ (a). We have to prove $d(X_n, Y_n) \to 0$ in probability. Given any subsequence n_k, using the fact $f(X_n) - f(Y_n) \to 0$ in P-probability and a diagonal argument, we extract a subsubsequence $(n_l) \subset (n_k)$ such that

$$f(X_{n_l}) - f(Y_{n_l}) \to 0 \ P\text{-a.s.}$$

for all f belonging to a dense, countable set $F \subset BL(S, d)$. Clearly we get $d(X_{n_l}, Y_{n_l}) \to 0$ P-a.s., since F is UD. □

In $\mathbb{R}^k$ we can find a finite UD-class F of Lipschitz functions, namely, $F = \{f_1, ..., f_k\}$, $f_i(x) = x_i$, $1 \leqslant i \leqslant k$. The implication "$K(X_n, Y_n) \to 0$ implies $d_F(X_n, Y_n) \to 0$" needs in this case an additional integrability condition.

The following observation that (c) ⇒ (a) is true in the special case $S = \mathbb{N}$, $d(n, m) \geqslant 1$, if $n \neq m$ is due to the referee. Indeed, if $K(X, Y) > \varepsilon$ for an $\varepsilon > 0$, we get $P(X \neq Y) > \varepsilon$ or without loss of generality $P(X < Y) > \varepsilon/2$. The following recursion yields a set A in $\mathbb{N}$ such that

$$E(|1_A(X) - 1_A(Y)|) > \frac{\varepsilon}{4} \qquad (1_A \in BL(S, d) \text{ is obvious}):$$

$$1 \in A$$

$$2 \notin A$$

$$n \in A \Leftrightarrow P(X < n, Y = n, X \notin A) \geqslant P(X < n, Y = n) \cdot \tfrac{1}{2}$$

For general separable metric spaces we use a somewhat different approach than that suggested by Theorem 4.

Theorem 5. Let X_n, Y_n be S-valued r.v.'s and assume that P^{Y_n} have densities h_n w.r.t. a measure μ such that $|h_n| \leqslant h$ for some $h \in L^1(\mu)$. Then the equivalences of Corollary 1 hold.

Proof.

(a) $\Rightarrow$ (b). By Theorem 4.

(b) $\Rightarrow$ (c). Obvious.

(c) $\Rightarrow$ (a). Let $S_0 = \{x_1, x_2, ...\}$ be dense in S and for $\varepsilon > 0$ define

$$f_i(x) = \begin{cases} 0 & \text{if } x \in K_{\varepsilon/4}(x_i) \\ \varepsilon/4 & \text{if } x \notin K_{3\varepsilon/4}(x_i) \end{cases}$$

Extend f_i by the Kirszbraun–McShane theorem (cf. Ref. 4, Theorem 6.1.1) to S [note that for the extension $0 \leqslant f_i(x) \leqslant \varepsilon/4$ and $f_i \in BL(S, d)$]. Define, furthermore,

$$A_1 = \{(x, y) \in S \times S;\ y \in K_{\varepsilon/4}(x_1)\}$$

$$A_i = \{(x, y) \in S \times S;\ y \in B_i\}$$

$$\text{where} \quad B_i = K_{\varepsilon/4}(x_i) \cap \bigcap_{j=1}^{i-1} (K_{\varepsilon/4}(x_j))^c$$

Then with $Q_n = P^{(X_n, Y_n)}$, $P(d(X_n, Y_n) > \varepsilon) = \sum_i Q_n(\{d(x, y) > \varepsilon\} \cap A_i) \leqslant \sum_i Q_n(\{d(x, y) > \frac{3}{4}\} \cap A_i) \leqslant (4/\varepsilon) \sum_i \int_{A_i} |f_i(x) - f_i(y)|\, dQ_n(x, y)$. Since $d(x, x_i) > \frac{3}{4}\varepsilon$ implies $f_i(x) = \varepsilon/4$ and $(x, y) \in A_i$ implies $f_i(y) = 0$. With $g_n(i) = (4/\varepsilon) \int_{A_i} |f_i(x) - f_i(y)|\, dQ_n(x, y)$ and $g(i) = \int_{B_i} h\, d\mu$ holds $g_n(i) \leqslant g(i)_i$, $\sum g(i) = \int h\, d\mu < \infty$, and $g_n(i) \leqslant E\, |f_i(X_n) - f_i(Y_n)| \xrightarrow[n \to \infty]{} 0$. Therefore, by dominated convergence $\lim_n P(d(X_n, Y_n) > \varepsilon) = 0$. □

Remark 2. The proof of Theorem 5 shows that there exists a countable class $F \subset BL(S, d)$ for all sequences (X_n, Y_n), where (Y_n) satisfies the boundedness assumption of Theorem 5. (Take $\varepsilon_n \to 0$ and the union of the classes of functions corresponding to ε_n in that proof.)

Corollary 2. If F is a uniformly bounded class of uniformly equicontinuous functions. Then under the assumptions of Theorem 5 or Corollary 1 the following holds: $K(X_n, Y_n) \to 0$ implies $d_F(X_n, Y_n) \to 0$.

Proof. On S define the pseudometric $\tilde{d}_F(x, y) = \sup_{f \in F} |f(x) - f(y)|$. Since F is uniformly equicontinuous. $(S, \tilde{d}_F)$ is separable and $K(X_n, Y_n) \to 0$ implies $K_F(X_n, Y_n) \to 0$, where K_F is the Ky-Fan metric w.r.t. $\tilde{d}_F$. Furthermore, if $f \in F$, then $\|f\|_{BL} \leqslant 1 + \sup_{f \in F} \|f\|_\infty < \infty$ (w.r.t. $\tilde{d}_F$). Therefore, from Theorem 5 resp. Corollary 1 we obtain $d_F(X_n, Y_n) \leqslant d_{BL}(X_n, Y_n) \to 0$. □

From Corollary 2 and Proposition 1 it follows that there exist functions φ and ψ defined as in Proposition 1 such that $\varphi(d_F) \leqslant K$ and $d_F \leqslant \psi(K)$.

The following counterexample for the implication (c) $\Rightarrow$ (a) is due to the referee.

Example 1. Corollary 1 does not hold in arbitrary separable metric spaces, Theorem 4b does not hold for arbitrary UD-classes.

Proof. Let S^n be the Euclidean sphere S^n equipped with the geodesic Euclidean distance d_n, and let $S = \bigcup_{n \in \mathbb{N}} S^n$ equipped with

$$d(x, y) = \begin{cases} \infty & \text{if} \quad x \in S_n,\ y \in S_m,\ n \neq m \\ d_n(x, y) & \text{if} \quad x, y \in S_n \end{cases}$$

Let $\hat{\mu}_n$ be the rotation-invariant probability on S^n and μ_n its trivial extension to S. Consider r.v.'s X_n with $P^{X_n} = \mu_n$ and $Y_n = -X_n$. Levy's isoperimetric inequality (cf. Ref. 7, Corollary 1.2, or Ref. 6, p. 221) yields for $\varepsilon > 0$

$$\sup_{f \in BL(S,d)} \mu_n(|f(x) - \text{med}_n f| > \varepsilon) < \left(\frac{\pi}{2}\right)^{1/2} \exp\left(-\frac{\varepsilon^2 n}{2}\right)$$

where $\text{med}_n f$ denotes the median of f with respect to μ_n. It follows that

$$\begin{aligned} E\,|f(X_n) - f(Y_n)| &\leqslant 2E\,|f(X_n) - \text{med}_n f| \\ &= 2\int_0^\infty P(|f(X_n) - \text{med}_n f| > \varepsilon)\, d\varepsilon \\ &\leqslant 2\int_0^\infty \left(\frac{\pi}{2}\right)^{1/2} \exp\left\{-\frac{\varepsilon^2 n}{2}\right\} d\varepsilon = \frac{\pi}{\sqrt{n}} \end{aligned}$$

and, therefore, $d_{BL}(X_n, Y_n) \to 0$. $K(X_n, Y_n) \not\to 0$ is obvious. □

While the Ky-Fan metric is related to the Prohorov metric ($\pi = \hat{K}$), the d_{BL} metric is related to the Kantorovich metric $\varkappa$. In fact, for S bounded ($d \leqslant 1$), d_{BL} coincides with $d_L := d_F$, where $F = \{f\colon S \to \mathbb{R}^1;\ |f(x) - f(y)| \leqslant d(x, y) \text{ for } x, y \in S\}$. In other words, d_{BL} may be viewed as d_L on $(S, d/(1+d))$. Then the Kantorovich metric $\varkappa(\mu, \nu) = \sup_{f \in F} |\int_S f\, d(\mu - \nu)|$ is the minimal metric w.r.t. d_L, $\hat{d}_L = \varkappa$. Here we simply use that $\varkappa \leqslant \hat{d}_L \leqslant \hat{L}_1$ and by the Kantorovich theorem $\hat{L}_1 = \varkappa$.

Moreover, $\varkappa$ appears as a "minimal norm" w.r.t. d_{BL}—a property which π and K do not enjoy—

$$\mathring{d}_L(\mu, \nu) := \inf \left\{ d_L(b) := \sup_{f \in F} \int_{S \times S} |f(x) - f(y)| \times b(dx, dy); b(A \times S) - b(S \times A) = (\mu - \nu)(A), A \in \mathfrak{B}(S) \right\} = \varkappa(\mu, \nu)$$

(See Ref. 4.) These relations seem to justify the consideration and analysis of the d_{BL} metric and uniformity.

ACKNOWLEDGMENT

We thank the referee for several critical and helpful remarks. In particular the example at the end of the paper is due to him. It clarified to a great extent the question of uniform convergence on Lipschitz functions.

REFERENCES

1. Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theory Prob. Appl.* **15**, 458–486.
2. Dudley, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39**, 1563–1572.
3. Dudley, R. M. (1985). *An Extended Wichura Theorem, Definitions of Donsker Class and Weighted Empirical Distributions.* In Lecture Notes in Mathematics, Vol. 1153, pp. 141–178, Springer-Verlag, Berlin.
4. Dudley, R. M. (1989). *Real Analysis and Probability.* Wadsworth & Brooks/Cole, Pacific Grove, CA.
5. Gänssler, P., and Stute, W. (1977). *Wahrscheinlichkeitstheorie.* Springer-Verlag, Berlin, Heidelberg, New York.
6. Levy, P. (1951). *Problèmes Concrets d'Analyse Fonctionelle.* Gauthier-Villers, Paris.
7. Milman, V. D. (1988). The heritage of P. Levy in geometrical functional analysis. *Astérisque* 157–158.
8. Rachev, S. T. (1984). Hausdorff metric construction in the probability measures space. *Pliska, Stud. Math. Bulgarica* **7**, 152–162.
9. Rachev, S. T., and Rüschendorf, L. (1990). A transformation property of minimal metrics. *Theor. Prob. Appl.* **35**, 131–137.
10. Rachev, S. T., Rüschendorf, L., and Schief, A. (1988). On the construction of almost surely convergent random variables. *Angew. Math. Inform.* (University of Münster), No. 10. Technical Report.
11. Schief, A. (1989). Almost surely convergent random variables with given laws. *Prob. Theor. Rel. Fields* **81**, 559–567.
12. Zolotarev, V. M. Linearly invariant probability metrics. In *Proceedings: Stability Problems for Stochastic Models.* VNIISI, Moscow (in Russian). English translation, *J. Soviet Math.* (to appear).