The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1563–1572

DISTANCES OF PROBABILITY MEASURES AND RANDOM VARIABLES

By R. M. Dudley[1]

Massachusetts Institute of Technology

1. Introduction. Let (S, d) be a separable metric space. Let $\mathcal{P}(S)$ be the set of Borel probability measures on S. $\mathcal{C}(S)$ denotes the Banach space of bounded continuous real-valued functions on S, with norm

$$\|f\|_\infty = \sup \{|f(x)| : x \varepsilon S\}.$$

On $\mathcal{P}(S)$ we put the usual weak-star topology TW^*, the weakest such that

$$P \to \int f\, dP, \quad P \varepsilon \mathcal{P}(S)$$

is continuous for each $f \varepsilon \mathcal{C}(S)$.

It is known ([8], [11], [1]) that TW^* on $\mathcal{P}(S)$ is metrizable. The main purpose of this paper is to discuss and compare various metrics and uniformities on $\mathcal{P}(S)$ which yield the topology TW^*.

For S complete, V. Strassen [10] proved the striking and important result that if $\mu, \nu \varepsilon \mathcal{P}(S)$, the Prokhorov distance $\rho(\mu, \nu)$ is exactly the minimum distance "in probability" between random variables distributed according to μ and ν. Theorems 1 and 2 of this paper extend Strassen's result to the case where S is measurable in its completion, and, with "minimum" replaced by "infimum", to an arbitrary separable metric space S. We use the finite combinatorial "marriage lemma" at the crucial step in the proof rather than the separation of convex sets (Hahn-Banach theorem) as in [10]. This offers the possibility of a constructive method of finding random variables as close as possible with the given distributions.

For S complete, V. Skorokhod ([9], Theorem 3.1.1, p. 281) proved the related result that if $\mu_n \to \mu_0$ for TW^* there exist random variables X_n with distributions μ_n such that $X_n \to X_0$ almost surely. This is proved in Section 3 below for a general separable S. Note that it is not sufficient to establish consistent finite-dimensional joint distributions for the X_n ; the Kolmogorov existence theorem for stochastic processes is not available in this generality. Instead we construct the joint distribution of $\{X_n\}_{n=0}^{\infty}$ out of suitable infinite Cartesian product measures.

When S is the real line R, various special constructions involving distribution and characteristic functions are known. In Section 4, we compare some of these uniformities on $\mathcal{P}(R)$.

2. Strassen's theorem. The metric of Prokhorov [8] is defined as follows. For any $x \varepsilon S$ and $T \subset S$ let

$$d(x, T) = \inf (d(x, y) : y \varepsilon T),$$

Received 11 January 1968.

[1] Fellow of the A. P. Sloan Foundation.

and for $\delta \geqq 0$ let

$$T^{\delta} = \{x \varepsilon S: d(x, T) < \delta\},$$
$$T^{\delta]} = \{x \varepsilon S: d(x, T) \leqq \delta\}.$$

Given P and Q in $\mathcal{P}(S)$ let

$$\begin{aligned} \sigma(P, Q) &= \inf (\epsilon > 0: P(F) \leqq Q(F^{\epsilon}) + \epsilon \quad \text{for all closed} \quad F \subset S), \\ \rho(P, Q) &= \max \quad (\sigma(P, Q), \sigma(Q, P)). \end{aligned} \tag{1}$$

Then ρ is a metric and metrizes TW^* on $\mathcal{P}(S)$ (this was proved in [8], Section 1.4, for S complete and is established for general separable S by results toward the end of this section).

We may replace F^{ϵ} by $F^{\epsilon]}$ in the definition of σ without changing its value. Also we may replace "all closed F" by "all Borel sets B" since if F is the closure of B, $F^{\epsilon} = B^{\epsilon}$ and $F^{\epsilon]} = B^{\epsilon]}$.

PROPOSITION 1 (known to Strassen [10]). *If P, $Q \varepsilon \mathcal{P}(S)$ and α, $\beta > 0$, then $P(F) \leqq Q(F^{\alpha}) + \beta$ for all closed F if and only if the same conditions hold with P and Q interchanged. Thus $\sigma(P, Q) = \sigma(Q, P) = \rho(P, Q)$.*

PROOF. Suppose $P(F) \leqq Q(F^{\alpha}) + \beta$ for all closed F and let T be closed. Then T^{α} is open,

$$T \subset (S \sim (S \sim T^{\alpha})^{\alpha}), \quad \text{and}$$
$$P(S \sim T^{\alpha}) \leqq Q((S \sim T^{\alpha})^{\alpha}) + \beta, \quad \text{so}$$
$$Q(T) \leqq Q(S \sim (S \sim T^{\alpha})^{\alpha}) \leqq P(T^{\alpha}) + \beta.$$

The conclusions follow.

Let $(\Omega, \mathcal{B}, \Pr)$ be a probability space and let $\mathcal{F}(\Omega, S)$ be the set of S-valued random variables over Ω, i.e. functions from Ω to S, measurable from $\mathcal{B}$ to the Borel σ-algebra in S, modulo functions vanishing with probability 1.

Then the natural topology of convergence in probability in $\mathcal{F}(\Omega, S)$ is metrized by the metric

$$d_{\Pr}(f, g) = \inf (\epsilon > 0: \Pr (d(f(\omega), g(\omega)) \geqq \epsilon) < \epsilon).$$

Now $f \times g: \omega \rightarrow (f(\omega), g(\omega))$ maps Ω measurably into $S \times S$ and defines an element $\Pr \circ (f \times g)^{-1}$ of $\mathcal{P}(S \times S)$. On $S \times S$ let π_1 and π_2 be the natural projections onto S:

$$\pi_1(x, y) \equiv x, \pi_2(x, y) \equiv y.$$

THEOREM 1. *Let S be a separable metric space, P, $Q \varepsilon \mathcal{P}(S)$, $\alpha \geqq 0$ and $\beta \geqq 0$. Then the following are equivalent:*

(I) *$P(T) \leqq Q(T^{\alpha]}) + \beta$ for all closed $T \subset S$;*

(II) *For any $\epsilon > 0$ there is a μ in $\mathcal{P}(S \times S)$ with $\mu \circ \pi_1^{-1} = P$, $\mu \circ \pi_2^{-1} = Q$, and $\mu(d(x, y) > \alpha + \epsilon) \leqq \beta + \epsilon$.*

PROOF. First assume (II). Then for any $\epsilon > 0$ and closed $T \subset S$, $P(T) \leqq Q(T^{\alpha+\epsilon}) + \beta + \epsilon$. Letting $\epsilon \downarrow 0$, this yields (I).

Conversely, assume (I). Given $\epsilon > 0$, let $\gamma = \epsilon/9$.

Let $\{x_n\}$ be a dense sequence in S. For any $x \varepsilon S$ let $f(x) = x_n$ for the least n such that $d(x, x_n) < \gamma$. Let $P_\gamma = P \circ f^{-1}$, $Q_\gamma = Q \circ f^{-1}$. Let $H_n = \{x_1, x_2, \cdots, x_n\}$ and choose n so that

$$\min (P_\gamma(H_{n-1}), Q_\gamma(H_{n-1})) > 1 - \gamma.$$

Then choose an integer m so that $n < m\gamma$. Let $P' \varepsilon \mathcal{P}(H_n)$ be such that for $i = 1, \cdots, n-1$, $mP'(x_i)$ is the largest integer $\leqq mP_\gamma(x_i)$. Likewise construct Q' from Q_γ. Then for any set $T \subset S$,

$$\max (|(Q' - Q_\gamma)(T)|, |(P' - P_\gamma)(T)|) \leqq 2\gamma,$$

$$\begin{aligned} P'(T) &\leqq P_\gamma(T) + 2\gamma \leqq P(T^\gamma) + 2\gamma \leqq Q(T^{\gamma+\alpha]}) + 2\gamma + \beta \\ &\leqq Q_\gamma(T^{2\gamma+\alpha}) + 2\gamma + \beta \leqq Q'(T^{2\gamma+\alpha}) + 4\gamma + \beta, \\ P'(T) &\leqq Q'(T^{2\gamma+\alpha}) + r/m, \end{aligned}$$

where r is the largest integer $\leqq m(4\gamma + \beta)$.

Let I be the unit interval [0, 1] with Lebesgue measure λ. On the Cartesian product $S \times I$ we form the product measures $P \times \lambda$ and $Q \times \lambda$. Let X be the natural projection of $S \times I$ on S.

For $i = 1, \cdots, n - 1$ we select measurable subsets E_i and F_i of $(f \circ X)^{-1}(x_i)$ such that

$$(P \times \lambda)(E_i) = P'(x_i), \qquad (Q \times \lambda)(F_i) = Q'(x_i).$$

Let

$$\begin{aligned} E_n &= (S \times I) \sim (E_1 \cup \cdots \cup E_{n-1}), \\ F_n &= (S \times I) \sim (F_1 \cup \cdots \cup F_{n-1}). \end{aligned}$$

We divide each E_i into $mP'(x_i)$ sets E_{ij} with $(P \times \lambda)(E_{ij}) = 1/m$; likewise each F_i into $mQ'(x_i)$ sets F_{ij} with $(Q \times \lambda)(F_{ij}) = 1/m$. We call the E_{ij} "boys" and the F_{ij} "girls". For ω in E_{ij} let $b(\omega) = x_i$; we say the boy E_{ij} "lives at" x_i. Likewise on F_{ij} let $g(\omega) = x_i$. Let B (resp. G) be the set of boys (resp. girls) so far defined, m of each. Let U (resp V) be a new disjoint set of r elements called boys (resp. girls).

We say a boy b *knows* a girl g if they live at points less than $2\gamma + \alpha$ apart or if $b \varepsilon U$ or $g \varepsilon V$. Then for any set $A \subset B$, with k members, living on a set $T \subset H_n$,

$$k \leqq mP'(T) \leqq mQ'(T^{2\lambda+\alpha}) + r$$

$\leqq$ the numbers of girls known by the k boys.

Thus any set of boys in $B \cup U$ knows at least as many girls. Hence by the marriage lemma (Philip Hall [5]; cf. also [4], p. 60) there is a function M from $B \cup U$ onto $G \cup V$ such that b knows $M(b)$ for each b. Hence there is a function h from B onto G such that b knows $h(b)$ except for at most r boys in B.

Now for each boy $b = E_{ij}$, let

$$p_b(A) = (P \times \lambda)(A \cap b), \qquad q_b(A) = (Q \times \lambda)(A \cap h(b))$$

for any measurable set $A \subset S \times I$. Let μ_b be the product measure

$$m(p_b \circ X^{-1}) \times (q_b \circ X^{-1}) \quad \text{on} \quad S \times S, \quad \text{and} \quad \textstyle\sum_{b \varepsilon B} \mu_b = \mu.$$

Then since the $b \varepsilon B$ are disjoint with union $S \times I$, as are the $h(b)$, and $p_b(S \times I) = q_b(S \times I) = 1/m$, we have $\mu \circ \pi_1^{-1} = P$ and $\mu \circ \pi_2^{-1} = Q$.

All but at most $2m\gamma$ of the boys in B are subsets each of some $(f \circ X)^{-1}(x_i)$, $i = 1, \cdots, n-1$, all but at most r of them know $h(b)$, and likewise for the girls in G. Thus except for at most $4m\gamma + r$ of the boys b in B, the following three statements all hold:

$$b \subset (f \circ X)^{-1}(x_i) \quad \text{for some} \quad i,$$

$$h(b) \subset (f \circ X)^{-1}(x_j) \quad \text{for some} \quad j,$$

and $d(x_i, x_j) < 2\gamma + \alpha$.
Thus

$$\mu(d(x, y) > \alpha + \epsilon) = \textstyle\sum_b \mu_b(d(x, y) > \alpha + \epsilon) < 4\gamma + r/m < 8\gamma + \beta < \beta + \epsilon.$$

This completes the proof.

A separable metric space (S, d) is called *inner regular* if for every Borel probability measure ν on S and Borel set $A \subset S$,

$$\nu(A) = \sup\,(\nu(K) : K \subset A, K \quad \text{compact}).$$

Then S is inner regular if it is complete, or a Borel subset of its completion $\bar{S}$, or if (and only if) it is P-measurable for every $P \varepsilon \mathcal{P}(\bar{S})$ (Varadarajan [11], b, p. 224).

THEOREM 2. *If in addition to the hypotheses of Theorem* 1 *S is inner regular, then* (I) *is equivalent to*

(II′) *There is a* μ *in* $\mathcal{P}(S \times S)$ *with*

$$\mu \circ \pi_1^{-1} = P, \qquad \mu \circ \pi_2^{-1} = Q, \qquad \textit{and} \qquad \mu(d(x, y) > \alpha) \leqq \beta.$$

PROOF. Clearly (II′) $\Rightarrow$ (II) $\Rightarrow$ (I). Assuming (I) let $\epsilon = \epsilon_k \downarrow 0$ in (II) and let μ_k be corresponding measures on $S \times S$. For any $\delta > 0$ there is a compact $K \subset S$ such that

$$P(S \sim K) < \delta/2, \qquad Q(S \sim K) < \delta/2,$$

so

$$\mu_k((S \times S) \sim (K \times K)) < \delta.$$

Thus the sequence $\{\mu_k\}$ is "tight" and has a TW^*-convergent sub-sequence (Varadarajan [11], Appendix, p. 223, Theorem 2; Part II, p. 202, Theorem 27). Thus we may assume $\mu_k \to \mu$ (TW^*) for some μ in $P(S)$. Then $\mu \circ \pi_1^{-1} = P$,

$\mu \circ \pi_2^{-1} = Q$, and

$$\mu(d(x, y) > \alpha) = \lim_{c \downarrow 0} \mu(d(x, y) > c + \alpha)$$
$$\leqq \lim_{c \downarrow 0} \liminf_{k \to \infty} \mu_k(d(x, y) > c + \alpha) \quad \text{([8], Theorem 1.2,)}$$
$$\leqq \liminf (\beta + \epsilon_k) = \beta. \qquad \text{q.e.d.}$$

We shall see in a moment that Theorem 2 cannot be proved under the hypotheses of Theorem 1 only. The following holds by definition of ρ, Proposition 1, and a passage to the limit:

COROLLARY 1. *Under the hypotheses of Theorems* 1 *or* 2, (I) *holds* (*hence* (II) *or* (II′) *respectively*) *when*

$$\alpha = \beta = \rho(P, Q).$$

In Theorems 1 and 2, μ depends on α and β. Now (I) will hold for different pairs (α, β) yet it may be impossible to obtain (II) for two different pairs simultaneously. For example let $S = R$, $P(0) = P(\frac{3}{2}) = \frac{1}{2} = Q(1) = Q(\frac{5}{2})$. Then (I) holds for $\alpha = \beta = \frac{1}{2}$ and for $\alpha = 1, \beta = 0$. If μ satisfied (II) for both these pairs then $\mu(x = \frac{3}{2}, y = 1) = \frac{1}{2}$ and $\mu(x = 0, y = 1) = \mu(x = \frac{3}{2}, y = \frac{5}{2}) = \frac{1}{2}$, a contradiction.

Note that Theorem 1 yields an independent proof of Proposition 1.

Now we give an example showing that the hypothesis of inner regularity cannot simply be dropped from Theorem 2. Let λ be Lebesgue measure on the real line. Then there is a subset A of the interval [0, 3] whose outer measure $\lambda^*(A)$ is 3, and such that A and $A + 1$ are disjoint (Halmos [6], Theorem E, p. 70). (Then A is not Lebesgue measurable and hence not inner regular.) Let $S = A$ and for any Borel set B in S let

$$P(B) = \lambda^*(B \cap [0, 2]/2),$$
$$Q(B) = \lambda^*(B \cap [1, 3])/2.$$

Then $P, Q \varepsilon \mathcal{P}(S)$ ([6], p. 75 Theorem A), and for any Bore lset B in S, $P(B) \leqq Q(B^1)$. Suppose

$$\mu \varepsilon \mathcal{P}(S \times S), \quad \mu \circ \pi_1^{-1} = P, \quad \mu \circ \pi_2^{-1} = Q, \quad \text{and} \quad \mu(|x - y| > 1) = 0.$$

Then $y \leqq x + 1$ almost surely, and

$$\int y \, d\mu = 2 = \int x \, d\mu + 1 = \int (x + 1) \, d\mu, \quad \text{so } y = x + 1$$

almost surely, contradicting disjointness of A and $A + 1$.

We shall use Theorem 1 to compare ρ with another metrization of TW^*[1]. Let $BL(S, d)$ denote the set of all bounded real-valued functions f on S which are Lipschitzian, i.e.

$$\|f\|_L \equiv \sup_{x \neq y} |f(x) - f(y)|/d(x, y) < \infty.$$

We let $\|f\|_{BL} = \|f\|_\infty + \|f\|_L$. (The use of Lipschitzian functions has been suggested in the excellent survey by Fortet [2], p. 191, and the boundedness assumption assures integrability for each probability measure. Cf. also Fortet and Mourier [2a].)

Now $(BL(S, d), \|\cdot\|_{BL})$ is a Banach space. If

$$\|\mu\|_{BL}^* = \sup \{|\int f\, d\mu| : \|f\|_{BL} \leqq 1\}$$

then the metric $\|\mu - \nu\|_{BL}^*$ metrizes TW^* on $\mathcal{P}(S)$ ([1], Theorems 6, 8, and 18).

PROPOSITION 2. *If the hypotheses of Theorem* 1 *and* (I) *hold then* $\|P - Q\|_{BL}^* \leqq 2 \max(\alpha, \beta)$.

PROOF. By (II), given $\epsilon > 0$ we take random variables X with distribution P and Y with distribution Q such that

$$P(d(X, Y) > \alpha + \epsilon) \leqq \beta + \epsilon.$$

Then for any f in $BL(S, d)$,

$$|\int f\, d(P - Q)| = |E(f(X) - f(Y))| \leqq (\alpha + \epsilon)\|f\|_L + 2(\beta + \epsilon)\|f\|_\infty.$$

Letting $\epsilon \downarrow 0$ we get the desired conclusion.

COROLLARY 2. *For S separable metric and $P, Q \varepsilon \mathcal{P}(S)$,*

$$\|P - Q\|_{BL}^* \leqq 2\rho(P, Q).$$

PROPOSITION 3. *If $P, Q \varepsilon \mathcal{P}(S)$, F is a closed set in the metric space S, $\alpha \geqq 0$, $\beta > 0$, and $P(F) > Q(F^\beta) + \alpha$, then*

$$\|P - Q\|_{BL}^* \geqq 2\alpha\beta/(2 + \beta).$$

PROOF. We deffine a function f in $BL(S)$ such that $f = 1$ on F, $f = -1$ on $S \sim F^\beta$, $\|f\|_\infty = 1$, and $\|f\|_{BL} \leqq 1 + 2/\beta$ ([1], Lemma 5)[2]. Then

$$(1 + 2/\beta)\|P - Q\|_{BL}^* \geqq \int f\, d(P - Q)$$
$$= \int (f + 1)\, d(P - Q) \geqq 2(P(F) - Q(F^\beta)) \geqq 2\alpha,$$

and the conclusion follows.

Now note that

$$f(x) \equiv 2x^2/(2 + x) = 2/[2/x^2 + 1/x]$$

is an increasing function of x for $x > 0$. Thus if $x \geqq 0$, $0 \leqq f(x) \leqq \frac{2}{3}$ if and only if $x \leqq 1$, and for $0 \leqq x \leqq 1$, $2x^2/3 \leqq f(x)$.

COROLLARY 3. *For S metric and $P, Q \varepsilon \mathcal{P}(S)$, $f(\rho(P, Q)) \leqq \|P - Q\|_{BL}^*$. Thus if $\rho(P, Q) \leqq 1$ or $\|P - Q\|_{BL}^* \leqq \frac{2}{3}$,*

$$\|P - Q\|_{BL}^* \geqq \tfrac{2}{3}\rho(P, Q)^2, \qquad \rho(P, Q) \leqq (\tfrac{3}{2}\|P - Q\|_{BL}^*)^{\frac{1}{2}}.$$

Corollaries 2 and 3 together imply that if S is separable (metric), $\|\cdot\|_{BL}^*$

---

[2] The extension of a Lipschitzian function f from $A \subset S$ to S without increasing $\| f \|_L$ was reportedly first shown by Banach (unpublished); cf. also McShane, E. J., "Extension of range of functions," *Bull. Amer. Math. Soc.* **40** (1934) 837–842.

and ρ define the same uniformity on $\mathcal{P}(S)$ (but this is not the weak-star uniformity, defined by all pseudo-metrics $|\int f\, d(P - Q)|$, $f \in \mathcal{C}(S)$, which indeed is not metrizable, unless S is compact ([1], Theorem 13)).

Here are examples showing that the inequalities in Corollaries 2 and 3 can be improved at most by a factor of 2. Let $d(p, q) = 1/n$, and let μ be a point mass 1 at p, and ν at q. Then

$$\rho(\mu, \nu) = 1/n, \qquad \|\mu - \nu\|_{BL}^{*} = 2/(2n + 1),$$

and the two distances are asymptotic as $n \to \infty$. On the other hand let

$$\sigma(p) = \sigma(q) = \tfrac{1}{2}, \qquad \tau(p) = \tfrac{1}{2} + 1/n, \qquad \tau(q) = \tfrac{1}{2} - 1/n.$$

Then $\rho(\sigma, \tau) = 1/n$, $\|\sigma - \tau\|_{BL}^{*} = \|\mu - \nu\|_{BL}^{*}/n$, asymptotic to $1/n^2$ as $n \to \infty$.

3. Almost sure convergence. A set A in a topological space S is called a *continuity set* of a measure $\mu \geqq 0$ if the boundary of A has μ-measure 0. If S is metrizable and $P_n \to P_0$ for TW^* in $\mathcal{P}(S)$, then $P_n(A) \to P_0(A)$ for every continuity set A of P_0([11], Theorem 2(IV), p. 182). The continuity sets of P_0 form an algebra ([8], Lemma 1.1) the proof does not use completeness of S.

Given $x \in S$, the balls

$$\{y \in S: d(x, y) < \epsilon\}$$

are continuity sets of P_0 except for at most countably many values of ϵ. Thus if S is separable, given $\delta > 0$ we can find finitely many disjoint continuity sets of P_0, each of diameter less than δ, and with total P_0-measure at least $1 - \delta$ (cf. [9], p. 281).

THEOREM 3. *Let S be a separable metric space, $P_n \in \mathcal{P}(S)$, $n = 0, 1, \cdots$, and $P_n \to P_0$ weak-star. Then there is a probability space $(\Omega, \mathcal{B}, \mu)$ with S-valued random variables X_n, $X_n \to X_0$ almost surely, and $\mu \circ X_n^{-1} \equiv P_n$.*

PROOF. For each $k = 1, 2, \cdots$, we take finitely many disjoint continuity sets of P_0, called $A(k, j)$, $j = 1, 2, \cdots, J_k$, each of diameter less than $1/k$, and satisfying

$$\textstyle\sum_j P_0(A(k, j)) \geqq 1 - 2^{-k}.$$

We may assume each term in the above sum is positive. Then for each k there is an n_k such that for all $n \geqq n_k$

$$\textstyle\sum_j |(P_n - P_0)(A(k, j))| < 2^{-k} \min_j P_0(A(k, j)).$$

We may assume $n_1 < n_2 < \cdots$.

Now for each n let S_n be a copy of S and I_n of the unit interval [0, 1] with Lebesgue measure λ_n. Let $\Omega_n = S_n \times I_n$ and let P_n' be the product measure $P_n \times \lambda_n$ on Ω_n. We define countable Cartesian products

$$\Omega_* = \textstyle\prod_{n=1}^{\infty} \Omega_n, \qquad \Omega = \Omega_0 \times \Omega_*.$$

Let X_n be the natural projection of Ω onto S_n and π the projection of Ω_n onto S_n for each n.

For each k, $j \leqq J_k$, and $n \geqq n_k$ we let

$$B(n, k, j) = A(k, j) \times [0, \delta(n, k, j)] \subset \Omega_n,$$
$$C(n, k, j) = A(k, j) \times [0, \gamma(n, k, j)] \subset \Omega_0,$$

choosing δ and γ so that

$$P_n'(B(n, k, j)) = P_0'(C(n, k, j)) = \min(P_n(A(k, j)), P_0(A(k, j))).$$

Then one of δ and γ is 1 and the other is at least $1 - 2^{-k}$. Let $B(n, k, 0) = \Omega_n \sim \bigcup_{j \geqq 1} B(n, k, j)$, $C(n, k, 0) = \Omega_0 \sim \bigcup_{j \geqq 1} C(n, k, j)$.

Let $n_0 = 1$ and for each n let $k(n)$ be the unique k such that $n_k \leqq n < n_{k+1}$.

For each n, Ω_0 is the disjoint union of finitely many sets $E(n, j) = C(n, k(n), j)$, $j = 0, 1, \cdots, J_{k(n)}$. For $j \geqq 1$ the $E(n, j)$ have diameters less than $1/k(n)$, and if also $n \geqq n_2$, $P_0(E(n, j)) > 0$. Likewise Ω_n is the disjoint union of finitely many sets

$D(n, j) = B(n, k(n), j), j = 0, 1, \cdots, J_{k(n)}$, with the same properties.

For each n, and x in Ω_0, let $j(n, x)$ be the j such that $x \in E(n, j)$. Let

$$A = \{x \in \Omega_0 : P_0'(E(n, j(n, x))) > 0 \quad \text{for all } n\}.$$

Then clearly $P_0'(\Omega_0 \sim A) = 0$. For x in A let $P(n, x)$ be the measure P_n' restricted to measurable subsets of $D(n, j(n, x))$ in Ω_n, then normalized to mass 1 (i.e. divided by $P_0'(E(n, j(n, x)))$). Let μ_x be the product measure

$$\textstyle\prod_{n=1}^{\infty} P(n, x) \text{ on } \Omega_*$$

(Halmos [6], Section 38, Theorem B, p. 157). Now I claim that for any measurable subset F of Ω_*, $x \to \mu_x(F)$ is a measurable function on Ω_0. In fact, for a given n, $P(n, x)$ has only finitely many possible values, each for x in a measurable set, and hence so does

$$\textstyle\prod_{n=1}^{N} P(n, x), \quad N \text{ finite}.$$

Thus the claim is true for sets $F = Y_N^{-1}(G)$ where Y_N is the projection of Ω_* on, and G is measurable in,

$$\textstyle\prod_{n=1}^{N} \Omega_n .$$

But the algebra of such sets generates the σ-algebra of measurable sets in Ω_*, and the class of sets F for which the claim holds is closed under countable monotone increasing and decreasing limits. Thus the claim holds for all measurable F ([6], Section 6, Theorem B, p. 27).

For any measurable $H \subset \Omega$, and $x \in \Omega_0$, let

$$H_x = \{y : (x, y) \in H\}, \quad \text{and} \quad \mu(H) = \int \mu_x(H_x)\, dP_0(x).$$

Note that $x \to \mu_x(H_x)$ is measurable if H is a finite union of measurable "rectangles" $A \times B$, $A \subset \Omega_0$, $B \subset \Omega_*$. Hence by monotone convergence it is measurable for any measurable $H \subset \Omega$, and μ is a countably additive probability measure on W. The distribution of X_n for μ is

$$[\sum_j P_0'(E(n, j))P(n, x)_{x\varepsilon E(n,j)}] \circ \pi^{-1} = P_n' \circ \pi^{-1} = P_n .$$

Since $\sum_k P_0(S \sim \bigcup_j A(k, j)) < \infty$, P_0-almost every point of S_0 belongs to $\bigcup_j A(k(n), j)$ for all large enough n. Also if $t \varepsilon I_0$ and $t < 1$, then $t < \gamma(n, k(n), j)$ for all j if n is large enough. Thus P_0'-almost all x belong to an $E(n, j)$ with $j \geqq 1$ for n large enough, and then

$$d(X_0, X_n) \leqq 1/k(n) \to 0$$

so $X_n \to X_0$. Thus $\mu(X_n \to X_0) = 1$, q.e.d.

4. The real line R. If $S = R$, the proof of Skorokhod ([9], Theorem 3.1) reduces naturally to the following. Let $P_n \varepsilon \mathcal{P}(R)$ and let F_n be their distribution functions

$$F_n(x) = P_n((-\infty, x]).$$

Let Ω be the unit interval [0, 1] with Lebesgue measure λ and for y in Ω let $X_n(y)$ be any x such that $F_n(x) = y$, or $F_n(x^-) \leqq y \leqq F_n(x)$. X_n is well-defined except for at most countably many values of y and hence is a well-defined random variable. If $P_n \to P_0$ for TW^*, then $F_n(x) \to F_0(x)$ whenever F_0 is continuous at x, and $X_n(y) \to X_0(y)$ except on the possible countable set of y where X_0 is not well-defined. Thus $X_n \to X$ almost surely. Clearly $\lambda \circ X_n^{-1} \equiv P_n$.

The above method seems unsuited to proving Theorem 1 on R. Let

$$P_n(j) = Q_n(j + 1) = 1/n, \quad j = 0, 1, \cdots, n - 1, P_n, Q_n \varepsilon \mathcal{P}(R).$$

$P_n - Q_n \to 0$ (even in total variation), but if X_n and Y_n are random variables on (Ω, λ) constructed from P_n and Q_n as above, then $X_n + 1 \equiv Y_n$.

For P in $\mathcal{P}(R)$ we introduce the usual characteristic function

$$\hat{P}(t) = \int_{-\infty}^{\infty} e^{ixt}\, dP(X).$$

On $\mathcal{P}(R)$ let UC be the uniformity of uniform convergence of $\hat{P}$ on compact sets, with a base given by the vicinities $\{(P, Q): |\hat{P}(t) - \hat{Q}(t)| \leqq 1/n$ whenever $|t| \leqq n\}$. Clearly the identity on $\mathcal{P}(R)$ is uniformly continuous from the BL^*(= Prokhorov) uniformity to UC. We do not have uniform continuity in the converse direction, as is shown by the following stronger result:

PROPOSITION 4. *For any $\delta > 0$ there exist $P, Q \varepsilon \mathcal{P}(R)$ with $\|P - Q\|_{BL}^* \geqq 1$ (in fact P concentrated in $x \geqq 1$ and Q in $x \leqq -1$) and $|\hat{P}(t) - \hat{Q}(t)| < \delta$ for all t.*

PROOF. For each $n = 1, 2, \cdots$, let

$$C_n = \sum_{k=1}^{n} 1/k$$

and let P_n have mass $1/kC_n$ at $k = 1, \cdots, n$, with $Q_n(A) \equiv P_n(-A)$. Then clearly $\|P_n - Q_n\|_{BL}^* \geqq 1$. Also $C_n|\hat{P}_n(t) - \hat{Q}_n(t)|$ is bounded uniformly in n and t (see e.g. Zygmund [12], volume 1, II, 9, p. 61), so $\hat{P}_n(t) - \hat{Q}_n(t) \to 0$ uniformly in t as $n \to \infty$, q.e.d.

The central limit theorem is generally proved using characteristic functions, and as long as one considers convergence $P_n \to P$ for a specific limit P, it is a question of topology rather than uniformity on $\mathcal{P}(R)$. But it is notable, and not surprising given Proposition 4, that in order to prove uniform closeness of n-fold convolutions $P*P* \cdots *P$, $P \varepsilon \mathcal{P}(R)$, to infinitely divisible distributions, one does not use characteristic functions (Kolmogorov [7]).

For P, $Q \varepsilon \mathcal{P}(R)$, Paul Lévy's metric $\rho_L(P, Q)$ may be defined by replacing, in the definition of Prokhorov's metric ρ, closed sets F by semi-infinite intervals $(-\infty, x]$. Now let P_n, $Q_n \varepsilon \mathcal{P}(R)$ where

$$P_n(2j) = Q_n(2j+1) = 1/n, \qquad j = 1, \cdots, n.$$

Then $\rho_L(P_n, Q_n) \equiv 1/n$, while $\|P_n - Q_n\|_{BL}^* \geqq \frac{1}{2}$, so the uniformity of Lévy's metric is strictly weaker than that of $\|\cdot\|_{BC}^*$ and ρ. ρ_L metrizes TW^* on $\mathcal{P}(R)$ ([3], p. 33, Theorem 1).

REFERENCES

[1] Dudley, R. M. (1966). Convergence of Baire measures. *Studia Math.* **27** 251–268.
[2] Fortet, Robert (1958). Recent advances in probability theory. *Some Aspects of Analysis and Probability.* Wiley, New York.
[2a] Fortet, Robert and E. Mourier (1953). Convergence de la répartition empirique vers la répartition théorique. *Ann. Sci. Ecole Norm. Sup.* **70** 266–285.
[3] Gnedenko, B. V., and A. N. Kolmogorov (1954). *Limit Distributions for Sums of Independent Random Variables* (transl. by K. L. Chung). Addison-Wesley, Reading, Mass.
[4] Hall, M. (1958). A survey of combinatorial analysis. *Some Aspects of Analysis and Probability.* Wiley, New York.
[5] Hall, P. (1935). On representatives of subsets. *J. London Math. Soc.* **10** 26–30.
[6] Halmos, P. (1950). *Measure Theory.* Van Nostrand, Princeton.
[7] Kolmogorov, M. N. (1963). Approximation to distributions of sums of independent terms by means of infinitely divisible distributions. *Trans. Moscow Math. Soc.* **12** 492–509.
[8] Prokhorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Prob. Appl.* **1** 157–214.
[9] Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theor. Prob. Appl.* **1** 261–290.
[10] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.
[11] Varadarajan, V. S. (1961). Measures on topological spaces. *Amer. Math. Soc. Transl.* Series 2, **48** 161–228.
[12] Zygmund, Antoni (1959). *Trigonometrical Series.* Cambridge Univ. Press.