BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 2, Number 2, December 1974

ON THE M_1 TOPOLOGY OF SKOROKHOD SPACES

BY

W. YANG

To Ky Fan, on the occasion of his sixtieth birthday

Abstract. We prove that on the space of paths without second kind discontinuities the topology M_1 of Skorokhod is Polish in nature.

0. **Introduction.** In his fundamental paper Skorokhod introduced four different (pseudo-) topologies for the space of paths without second kind of discontinuities. These are the J_1, J_2, M_1, M_2 topologies, in his notation. By far the most important is J_1-topology, the strongest. It was proved by Kolmogorov and Prokhorov that the J_1-topology is Polish in nature, and the limit theorems fall within the realm of Prokhorov theory. (For one of the salient applications to Markov processes, see Yang [2].) We will study the M_1-topology. The topology, not just the sequential convergence, is introduced through an intermediate "space of pawns". The Polish nature is proved, and some remarks very relevant to applications are added.

Notation. 1. (E, ρ) is a *Polish* metric space (i.e. a complete separable metric space). In case $(E, \| \ \|)$ is a separable Banach space,

$$\rho(x, y) = \|x - y\|.$$

2. I is a compact interval, which for convenience will usually be $[0, 1]$. The set of all order-automorphisms is denoted by $\mathcal{A}_0(I) = \mathcal{A}_0$.

3. $\overline{\overline{\mathcal{G}}} = \overline{\overline{\mathcal{G}}}(I, E)$ is the set of all functions $X : I \to E$, free from discontinuities of the second kind, and normalized to be right-continuous in the interior $(0, 1)$ of $I = [0, 1]$. The subset defined by $X(0+) = X(0)$ is denoted by $\overline{\mathcal{G}}$, and the usual Skorokhod space

Received by the editors May 28, 1974.

is $\mathscr{G}$ of all X in $\overline{\overline{\mathscr{G}}}$ continuous at the two ends of I. (We write $\mathscr{G}_L$ in case I is replaced by $[0, L]$. Similarly for $\overline{\overline{\mathscr{G}}}$.)

4. $d(X;\ \theta;\ Y) = \sup\{|\theta t - t| + \rho(X(t);\ Y(\theta t)) : t \in I\}$ for $\theta \in \mathscr{A}_0$; $X, Y \in \overline{\overline{\mathscr{G}}}$. $d(X;\ Y) = \inf\{d(X;\ \theta) : \theta \in \mathscr{A}_0\}$.

5. $\Pi_- : \overline{\mathscr{G}} \to \mathscr{G}$ and $\Pi_+ : \overline{\mathscr{G}} \to E$ are defined by:

$$\Pi_- X(t) = X(t)\ \ (0 \leq t < 1)\ \text{or}\ X(1-)\ \ (t = 1).$$

$$\Pi_+ X \quad = X(1)\,.$$

In general, Π will be used for projections.

6. $\Gamma_c(X)$ (respectively $\bar{\Gamma}_c(X)$) is the image (the trace) of $X \in \overline{\mathscr{G}}$ up to $c \in I$:

$$\Gamma_c(X) = \{X(t) : 0 \leq t \leq c\}\,,$$

$$\bar{\Gamma}_c(X) = \text{the closure of } \Gamma_c(X)\,,$$

$$\Gamma X = \Gamma_1(X)\,.$$

7. A *regularity modulus* is a function w on $[0, 1]$, nondecreasing with $w(0+) = 0$. For $X \in C(I;\ E)$, the continuity-modulus of X is $\omega(\cdot;\ X) : \omega(\delta;\ X) = \sup\{\rho(X(t_2);\ X(t_1)) : |t_1 - t_2| \leq \delta\}$.

For $X \in \overline{\mathscr{G}}(I; E)$, the Skorokhod modulus of X is $\bar{\omega}(\cdot; X) : \bar{\omega}(\delta; X)$ = the maximum of the following three quantities.

$$\sup_{\delta < t < 1-\delta} \min\Big[\sup_{t-\delta \leq t' \leq t} \rho(X(t');\ X(t));\ \sup_{t \leq t'' \leq t+\delta} \rho(X(t);\ X(t''))\Big],$$

$$\sup_{0 \leq t \leq \delta} \rho(X(t),\ X(0+));\quad \sup_{1-\delta \leq t < 1} \rho(X(t);\ X(1-))\,.$$

8. $J^\rho(t;\ X) \equiv \rho(X(t-);\ X(t+))$ is the *jumping distance* of $X \in \overline{\overline{\mathscr{G}}}$ at time t. The discontinuity set of X is then $D(X) = \{t : J^\rho(t;\ X) > 0\}$ while the ε-discontinuity set is $D_\varepsilon(X) = \{t : J^\rho(t;\ X) > \varepsilon\}$.

1. **The space** $(\mathscr{C}(J;\ E)/\sim;\ \bar{d})$. The set $\mathscr{C}(J;\ E)$ of all continuous functions from the compact interval J to the Polish metric space $(E,\ \rho)$ is equipped with the metric

$$d(f;\ g) = \sup_{t \in J} \rho(f(t);\ g(t))\,.$$

We consider the equivalence relation $\approx$ defined by: $f \approx g$ iff there is an automorphism θ of the ordered set J such that $f = g \circ \theta$. We define also

$$\bar{d}(f;\ g) = \inf\{d(f;\ g\circ\theta) : \theta\in\mathcal{A}_0(J)\},$$

where $\mathcal{A}_0(J)$ is the set of all automorphisms of J.

LEMMA 1. *$\bar{d}$ is a pseudo-metric on $\mathcal{C}(J;\ E)$ compatible with $\approx$; i.e., if $f\approx g$, then $\bar{d}(f;\ g)=0$.*

Proof. We need only check the triangle inequality. For f, g, h in $\mathcal{C}(J;\ E)$, and $\varepsilon>0$, we can find $\theta_1, \theta_2\in\mathcal{A}_0$ such that

$$\bar{d}(f;\ g)\geq d(f;\ g\circ\theta_1)-\varepsilon,$$

and

$$\bar{d}(f;\ h)\geq d(g;\ h\circ\theta_2)-\varepsilon.$$

Therefore

$$\begin{aligned}\bar{d}(f;\ g)+\bar{d}(f;\ h) &\geq d(f;\ g\circ\theta_1)+d(f;\ h\circ\theta_2)-2\varepsilon\\ &= d(f;\ g\circ\theta_1)+d(g\circ\theta_1;\ h\circ\theta_2\circ\theta_1)-2\varepsilon\\ &\geq d(f;\ h\circ\theta_2\circ\theta_1)-2\varepsilon\\ &\geq \bar{d}(f;\ h)-2\varepsilon.\quad \text{Q.E.D.}\end{aligned}$$

Unfortunately $\bar{d}(f;\ g)=0$ does not mean $f\approx g$. We will first clear up this situation.

LEMMA 2. *If θ is a surjective nondecreasing function from J onto J, then*

$$\bar{d}(f;\ f\circ\theta)=0.$$

Proof. Indeed we can always find a sequence $(\theta_n)\subset\mathcal{A}_0$ such that $\lim_n\theta_n(t)=\theta(t)$ uniformly.

We proceed to prove the converse (Lemma 6). Let us write $\mathcal{A}_1(J)$ for the set of all nondecreasing surjections of J onto itself, and also write $\mathcal{A}_2(J)$ for the set of all nondecreasing right-continuous functions from J into J. More generally, we consider functions from compact intervals J to $\tilde{J}$ and write $\mathcal{A}_i(J,\tilde{J})$ with $i=0, 1, 2$. We will also normalize $\phi\in\mathcal{A}_2(J,\tilde{J})$ so that ϕ maps the $\max J$ to $\max\tilde{J}$, and also so that symbolically $\phi(\min J-)\equiv\min\tilde{J}-$.

If $\phi\in\mathcal{A}_2(J,\tilde{J})$, we define the dual of ϕ as the function $\tilde{\phi}\in\mathcal{A}_2(\tilde{J};\ J) : \tilde{\phi}(v)=\inf\{u\in J : \phi(u)>v\}$, with the convention that empty-infimum will refer to the maximum. This name is justified because of the following elementary lemma, whose proof we omit.

LEMMA 3. $\tilde{\tilde{\phi}} = \phi$.

In case $\phi \in \mathcal{A}_0(J, \tilde{J})$, we see that $\tilde{\phi} = \phi^{-1} \in \mathcal{A}_0(\tilde{J}, J)$. In this connection it is of interest to observe that in some sense all $\phi \in \mathcal{A}_2$ can be approximated by such invertible functions.

Let $\phi \in \mathcal{A}_2(J, \tilde{J})$. The Stieltjes measure $d\phi$ on J is defined by

$$d\phi(x, x'] = \phi(x') - \phi(x) \qquad \text{for} \quad \min J < x < x',$$

and

$$d\phi\{\min J\} = \phi(\min J) - \min \tilde{J}.$$

LEMMA 4. *Let* $(\phi_n) \subset \mathcal{A}_2(J, \tilde{J})$. *If* $d\phi_n$ *converges weakly to* $d\phi_0$, *then* $d\tilde{\phi}_n$ *converges weakly to* $d\tilde{\phi}_0$.

Proof. The sequence $(d\tilde{\phi}_n)$ is certainly weakly tight; therefore we may assume that $(d\tilde{\phi}_n)$ converges to $d\psi$ with $\psi \in \mathcal{A}_2(\tilde{J}, J)$, and it suffices to show $\psi = \tilde{\phi}_0$.

Let v be a continuity point for ψ and $\tilde{\phi}$. We then have

$$\psi(v) = \lim_n \tilde{\phi}_n(v) = u,$$

say.

For any $\varepsilon > 0$, we have $u + \varepsilon > \tilde{\phi}_n(v)$, for n sufficiently large; then $\phi_n(u + \varepsilon) > v$, and therefore

$$\phi(u + \varepsilon) \geq \overline{\lim}\, \phi_n(u + \varepsilon) \geq v.$$

It follows that

$$\phi(u) \geq v, \quad \phi(u) > v - \varepsilon, \quad \text{or} \quad u \geq \tilde{\phi}(v - \varepsilon)$$

and therefore $u \geq \tilde{\phi}(v)$ by the continuity of $\tilde{\phi}$ at v. We have $\psi \geq \tilde{\phi}$. Conversely, $u - \varepsilon < \tilde{\phi}_n(v)$ for large n; then

$$\phi_n(u - \varepsilon) \leq v, \quad \text{or} \quad \phi_n(u-) \leq v,$$

and

$$\phi(u-) \leq \underline{\lim}\, \phi_n(u-) \leq v,$$

thus

$$\phi(u - \delta) \leq v \quad \text{and} \quad \tilde{\phi}(v) \geq u - \delta, \quad \text{or} \quad \tilde{\phi} \geq \psi. \qquad \text{Q.E.D.}$$

Let us introduce the *diagram* $\hat{\Gamma}(\phi)$ of a pair $(\phi, \tilde{\phi})$ when $\phi \in \mathcal{A}_2(J, \tilde{J})$. This is the union of the graph of ϕ, with segments

$$\{(u, v) : \phi(u-) \leq v \leq \phi(u+)\} .$$

It is easy to see that the diagram $\hat{\Gamma}(\phi)$ and the diagram $\hat{\Gamma}(\tilde{\phi})$ are "dual" to each other by the permutation of two coordinates.

LEMMA 5. *There is a homeomorphism from I to $\hat{\Gamma}(\phi)$.*

Proof. This will be proved in a more general context. See Lemma 7.

Now let $\bar{d}(f; g)=0$ in $\mathcal{C}(J; E)$. We have a sequence $\theta_n \in \mathcal{A}_0(J)$ such that $\lim_n d(f; g\circ\theta_n)=0$. We may as well assume the weak convergence of $d\theta_n \to d\theta$ in $M_b(J)$ and therefore also that $d\theta_n^{-1} \to d\tilde{\phi}$ in $M_b(\tilde{J})$, where $J=\tilde{J}$. We have the validity of $f(u)=g\circ\phi(u)$, and $g(v)=f\circ\tilde{\phi}(v)$ for almost all u and almost all v, and hence for all u and all v because of right-continuity. It is also clear that for $\phi(u-) \leq v \leq \phi(u)$, we have $f(u)=g(v)$, and similarly $f(u)=g(v)$ for $\tilde{\phi}(v-) \leq u \leq \tilde{\phi}(v)$.

Let θ_0 be a homeomorphism from J to $\hat{\Gamma}(\phi)$, in the increasing direction, for the two components (θ_1, θ_2). Consider the function h defined by $h(r)=f\circ\theta_1(r)$. If $\theta_1(r)<\theta_2(s)$ for all $s>r$, then $\theta_2(r)=\phi(\theta_1(r))$ and therefore $g\circ\theta_2(r)=g\circ\phi(\theta_1(r))=f\circ\theta_1(r)$. Otherwise $\theta_2(r)$ is a continuity point of $\tilde{\phi}$ and $f\circ\tilde{\phi}(\theta_2(r))=g\circ(\theta_2(r))$ $=h(r)$, $\theta_1(r)=\tilde{\phi}(\theta_2(r))$. We have thus proved the "only if" part of the following lemma. The "if" part is just Lemma 2.

LEMMA 6. *$\bar{d}(f; g)=0$ in $\mathcal{C}(J; E)$ iff there exists*

$$\theta_1, \theta_2 \in \mathcal{A}_1(J) , \text{ and } h \in \mathcal{C}(J; E) ,$$

such that

$$h=f\circ\theta_1=g\circ\theta_2 .$$

If we define $f \sim g$ in $\mathcal{C}(J; E)$ to be the existence of a couple $\theta_1, \theta_2 \in \mathcal{A}_1(J)$ such that $f\circ\theta_1=g\circ\theta_2$, we have proved that this is an equivalence relation. This is just the saturation of the relation $\approx$ with respect to the pseudo-metric $\bar{d}$.

THEOREM 1. *The space $(\mathcal{C}(J, E)/\sim, \bar{d})$ is a Polish space.*

Proof. It remains to prove the completeness. With $(f_n \bmod \sim)$ a Cauchy sequence, it suffices to exhibit a convergent subsequence. We may therefore assume $\sum_1^\infty \bar{d}(f_n, f_{n+1}) < \infty$. It follows that we can choose $(\theta_n) \in \mathcal{A}_0$ such that $\sum_1^\infty d(f_n; f_{n+1} \circ \theta_n) < \infty$. Therefore, with

$$g_n \equiv f_n \circ \theta_{n-1} \circ \theta_{n-2} \circ \cdots \circ \theta_1 \approx f_n$$

we have

$$\sum_2^\infty d(g_n; g_{n+1}) = \sum_2^\infty d(f_n; f_{n+1} \circ \theta_n) < \infty .$$

Let $g_n \to g$ in $(\mathcal{C}(J, E); d)$. Then

$$\bar{d}(f_n; g) \leq d(g_n; g) \to 0 . \quad \text{Q. E. D.}$$

2. **The Polish space** $(\bar{\bar{\mathcal{F}}}; \bar{d})$. Let J, I be two compact intervals and let $(E, \| \ \|)$ be a separable Banach space. If $f \in \mathcal{C}(J; I \times E)$ we write $f_I = \Pi_I \circ f$, $f_E = \Pi_E \circ f$, $f = f_I \oplus f_E$. The set of all $f \in \mathcal{C}(J; I \times E)$ such that $f_I \in \mathcal{A}_1(J; I)$ is written as $\mathcal{C}(J; I; E)$. Since it is obviously closed in $(\mathcal{C}(J, I \times E), \bar{d})$ we have a Polish metric space $\Lambda(I; E) = (\underline{\mathcal{C}(J, I; E)}; \bar{d})$ (the space of "pawns"). The metric in $I \times E$ will be written as $\bar{\rho}$. Note that J is not essential in the definition of Λ.

For each $t \in I$ and $f \in \mathcal{C}(J; I; E)$, $f(J) \cap (t \times E) = Af(t)$ is nonempty. There are two possibilities:

1. $f(J) \cap (t \times E)$ is a singleton, say $\{(t, X(t))\} = \{f(u)\} \equiv \{\hat{f}(u)\}$.
2. $f(J) \cap (t \times E)$ is the set $f(J_1)$ for J_1 a closed subinterval $[t_f^- \ t_f^+]$ of J.

(We then write $t \in Df$ and call t an *action time* of f.) We also write

$$X(t\pm) = f_E(t_f^\pm)$$

and

$$\hat{f}_E(u) = [\overline{f_E(t_f^-), f_E(t_f^+); (u - t_f^-)/(t_f^+ - t_f^-)}],$$

where we have employed the notation

$$[\overline{x, y; \alpha}] = (1 - \alpha)\, x + \alpha y \qquad \text{for } \alpha \in [0, 1].$$

In the following we also write

$$[x, y] \equiv \{[\overline{x, y, \alpha}] : \alpha \in [0, 1]\} .$$

The following observations are then clear:

OBSERVATIONS (1–5)

1. Af, Df and $X = \check{f}$ depend only on $f \bmod \sim$; hence DR, AR and $\check{R}$ are meaningful for $R \in \Lambda$.
2. $X = \check{R} \in \overline{\overline{\mathscr{G}}}$.
3. $f \to \hat{f} = f_I \oplus \hat{f}_E$ is homotopic to the identity on $(\mathscr{C}(J; I; E); d)$.

The natural homotopy here is $(f, \alpha) \to (1-\alpha)f + \alpha\hat{f}$. Unfortunately this is not a homotopy if we take the (weaker) pseudo-metric $\bar{d}$ instead of d. Therefore the homotopy does *not* pass down to $\Lambda(I; E)$. In any case we still have:

4. $(\hat{f} \bmod \sim)$ depends only on $(f \bmod \sim)$; hence $\hat{R}$ are meaningful for $R \in \Lambda$. Moreover, $R \to \hat{R}$ is a continuous retraction of Λ.
5. $\hat{R} \to \check{R} \in \overline{\overline{\mathscr{G}}}$ is injective.

We now show that $\{\check{R} : R \in \Lambda\} = \overline{\overline{\mathscr{G}}}$, so that $\overline{\overline{\mathscr{G}}}$ can be identified as a subset of Λ, and $\wedge$ or $\vee$ is a retraction of Λ onto $\overline{\overline{\mathscr{G}}}$. For $X \in \overline{\overline{\mathscr{G}}}$ we define its diagram to be

$$\hat{\Gamma}(X) = \{(t, x) : t \in I, x \in [X(t-), X(t+)]\} .$$

(Note: This is consistent with our notation for $\phi \in \mathscr{A}_2(J, \tilde{J})$.)

A pawn-representation for $X \in \overline{\overline{J}}$ is then a homeomorphism f from J onto $\hat{\Gamma}(X)$ such that f_I is nondecreasing.

LEMMA 7. *A pawn-representation exists.*

Proof. It is convenient to set $J = [0, 2]$, $I = [0, 1]$ and we consider the jumping times $\{\tau_1, \tau_2, \cdots\}$ of X. We will assume $D(X)$ to be countably infinite; the case when it is finite (or empty) is quite similar or simpler.

Put a discrete measure m on I supported by $D(X)$ and $m(\tau_k) = 2^{-k-1}$, and let $\phi(t) = t + m[0, t]$. Therefore $\phi \in \mathscr{A}_2(I, J)$. Define now for any $s \in J$,

$$f_E(s) = X(\hat{\phi}(s)) , \quad \text{when} \quad \hat{\phi}(s) \neq \tau_k , \quad \text{all } k , \quad \text{or}$$

$$[\overline{X(\tau_k-), X(\tau_k+), 2^{k+1}(s - \phi(\tau_k-))}] , \quad \text{when} \quad \tilde{\phi}(s) = \tau_k .$$

Then the mapping $f=\tilde{\phi}\oplus f_E$ is a pawn-representation for X. Q.E.D.

LEMMA 8. *$\mathcal{C}(I;\ E)\subset\overline{\overline{\mathcal{G}}}(I;\ E)$ is dense in $A(I,\ E)$.*

Proof. If $f\in\mathcal{C}(J;\ I,\ E)$, then $(f \bmod \sim)\in\mathcal{C}$ if $\Pi_1\circ f=f_I\in\mathcal{A}_0(J;\ I)$. But we can always approximate f_I, which in general belongs to $\mathcal{A}_1(J,\ I)$, uniformly by a sequence (θ_n) in $\mathcal{A}_0(J,\ I)$. Q.E.D.

If $R\in A$, the diagram of R is $\hat{\Gamma}R=f(J)\subset I\times E$ for any f representing R. The notion of "between" can be defined by: $f(u_0)$ is between $f(u_1)$ and $f(u_2)$ iff $u_0\in[u_1,\ u_2]$, and we denote this by $f(u_0)\in R[f(u_1);\ f(u_2)]$. Again this does not depend on the representation f of R. We define then the *error function* eR on I by:

$$eR(t)=\max\{\bar{\rho}(\bar{x}_0;\ [\bar{x}_1;\ \bar{x}_2]):\bar{x}_0\in R[\bar{x}_1;\ \bar{x}_2]\subset AR(t)\},$$

where

$\bar{\rho}(\bar{x}_0,\ [\bar{x}_1;\ \bar{x}_2])$ is the distance from $\bar{x}_0$ to the segment $[\bar{x}_1;\ \bar{x}_2]$.

The *gauge of error* of R is $E(R)=\max\{eR(t):t\}$. It is trivial to see

LEMMA 9. *The subset $\overline{\overline{\mathcal{G}}}$ of A is characterized by $ER=0$.*

We can prove that $\overline{\overline{\mathcal{G}}}$ is a $\mathfrak{G}_\delta$ set in the Polish space A, once we have proved the following

THEOREM 2. *The function $R\to ER$ is upper-semicontinuous on A.*

Assuming this, we have proved

COROLLARY. *The space $\overline{\overline{\mathcal{G}}}$ with the metric-topology of $\bar{d}$ is Polish.*

We return to the proof of the Theorem 2. First we observe

LEMMA 10. *If $\bar{\rho}(\bar{x}_i,\ \bar{y}_i)<\delta$ for $i=0,\ 1,\ 2,$ then $\bar{\rho}(\bar{y}_0,\ [\bar{y}_1,\ \bar{y}_2])\leq\bar{\rho}(\bar{x}_0;\ [\bar{x}_1;\ \bar{x}_2])+3\delta$.*

Proof. $\bar{\rho}(\bar{y}_0;[\bar{x}_1;\bar{x}_2])<\bar{\rho}(\bar{x}_0;[\bar{x}_1;\bar{x}_2])+\delta$, and also $\bar{\rho}(\bar{x}_0;[\bar{x}_1;\bar{y}_2])\leq\bar{\rho}(\bar{x}_0;\ [\bar{x}_1;\ \bar{x}_2])+\delta$. Q.E.D.

LEMMA 11. *For any $\varepsilon'>0$ and $R\in A$, there exists $\delta>0$, such that whenever $\bar{x}_0\in R[\bar{x}_1,\ \bar{x}_2]$ with $|\Pi_I\bar{x}_2-\Pi_I\bar{x}_1|<\delta$, then*

$$\bar{\rho}(\bar{x}_0;\ [\bar{x}_1,\ \bar{x}_2])<ER+\varepsilon'.$$

Proof. Given $\varepsilon>0$ and $R\in\Lambda$, we can find a number $\delta_1>0$ such that

$$\bar{\omega}(\delta_1;\ X)<\varepsilon \quad \text{with} \quad X=\check{R}\in\bar{\bar{J}}.$$

Let $\bar{D}_\varepsilon R=(D_\varepsilon X)\cup\{t : eR(t)>\varepsilon\}=\{\tau_1<\tau_2\cdots<\tau_l\}$, and let $\min_i\{|\tau_{i+1}-\tau_i|\}=\delta_2$. Choose $\delta_3>0$ such that if $\tau_i-\delta_3\leq t<\tau_i$, then $\|X(t)-X(\tau_i-)\|<\varepsilon$, and if $\tau_i<t\leq\tau_i+\delta_3$, then $\|X(t)-X(\tau_i+)\|<\varepsilon$. Let $\delta=\min\{\delta_1,\delta_2,\delta_3,\varepsilon\}$. Consider $\bar{x}_0,\bar{x}_1,\bar{x}_2\in\hat{\Gamma}R$ with

$$\bar{x}_0\in R[\bar{x}_1;\ \bar{x}_2] \quad \text{and} \quad t_2-t_1<\delta\,, \quad \text{where} \quad t_i=\Pi_I\,\bar{x}_i\,.$$

There are several possibilities for the positions of t_1, t_2 relative to the partition $\bar{D}_\varepsilon R$.

Case 1. t_1, t_2 are in the same open subinterval of $\bar{D}_\varepsilon R$.

We have therefore, $\|X(t_i)-X(t_j)\|<3\varepsilon$, $i,j=0,1,2$. Also if $x_i=\Pi_E\,\bar{x}_i$, then $\|x_i-X(t_i)\|<2\varepsilon$, because $\|X(t_i+)-X(t_i-)\|<\varepsilon$, and $\varepsilon>\rho(x_i,[X(t_i-);\ X(t_i+)])$, where ρ has the same meaning for E as $\bar{\rho}$ for $I\times E$. We have therefore $\|x_0-x_1\|=\rho(x_1,x_2)<7\varepsilon$, and $\bar{\rho}(\bar{x}_0,[\bar{x}_1,\bar{x}_2])<8\varepsilon$.

Case 2. t_1, t_2 are separated by one $\tau\in\bar{D}_\varepsilon R$,

$(t_1<\tau<t_2)$ and $\rho(X(t_1);\ X(\tau-))<\varepsilon$, $\rho(X(t_2);\ X(\tau+))<\varepsilon$.
If (a) $t_0<\tau$, then also $\rho(X(t_0);\ X(t_1))<\varepsilon$. If (b) $t_0>\tau$, then also $\rho(X(t_0);\ X(t_2))<\varepsilon$. In these two circumstances, we will have the same reasoning as in case 1, so that $\bar{\rho}(\bar{x}_0;[\bar{x}_1,\bar{x}_2])<8\varepsilon$. Therefore we assume (c) $t_0=\tau$. We have now $\rho(x_0;[X(\tau-),X(\tau+)])\leq ER$. Since $\rho(x_i;\ X(t_i))\leq 2\varepsilon$ for $i=1,2$, we have $\rho(x_1;\ X(\tau-))\leq 3\varepsilon$ and $\rho(x_2;\ X(\tau+))<3\varepsilon$, and thus

$$\rho(x_0;\ [x_1,\ x_2])<ER+6\varepsilon\,, \quad \text{or}$$

$$\rho(\bar{x}_0;\ [\bar{x}_1,\ \bar{x}_2])<ER+8\varepsilon\,.$$

Case 3. At least one of $\{t_1,t_2\}$ is in $\bar{D}_\varepsilon R$. The proof is as easy as in Case 2. Therefore we have proved Lemma 2, with $\varepsilon'=8\varepsilon$.

Proof of Theorem 2. If $\bar{d}(S,R)<\delta/2$ then with any triple $\bar{y}_0,\bar{y}_1,\bar{y}_2\in AS(t)$ and $\bar{y}_0\in S[\bar{y}_1,\bar{y}_2]$, we have $\bar{x}_0,\bar{x}_1,\bar{x}_2\in\hat{\Gamma}R$ with $\bar{\rho}(\bar{x}_i;\ \bar{y}_i)<\delta/2$. In particular $|\Pi_I\,\bar{x}_2-\Pi_I\,\bar{x}_1|<\delta$, and therefore

$$\bar{\rho}(\bar{x}_0;\ [\bar{x}_1;\ \bar{x}_2]) < ER + 8\varepsilon\,, \quad \text{or}$$

$$\bar{\rho}(\bar{y}_0;\ [\bar{y}_1,\ \bar{y}_2]) < ER + 11\varepsilon\,.$$

REMARK. E cannot be lower-semicontinuous because $ER=0$ for all $R \in \bar{\bar{\mathcal{G}}}$, while $\bar{\bar{\mathcal{G}}}$ is dense in Λ.

3. **Further remarks.** From now on we will take the $\bar{d}$-topology for $\bar{\bar{\mathcal{G}}}$. It is a Polish *topological* space though $\bar{d}$ is *not* a *Polish metric* for $\bar{\bar{\mathcal{G}}}$.

LEMMA 12. *The Borel tribe of* $\bar{\bar{\mathcal{G}}}([0, 1];\ E)$ *is generated by coordinates* $X \mapsto X(t)$ *for* $t \in [0-,\ 1+]$.

Proof. The canonical mapping $E \times \mathcal{G}([0, 1];\ E) \times E \to \bar{\bar{\mathcal{G}}}$ is a continuous bijection.

We turn to elementary properties of the $\bar{d}$-topology of $\bar{\bar{\mathcal{G}}}$ which is so essential in Yu. V. Prokhorov's theory. It is ironical that A. V. Skorokhod did not regard the Prokhorov approach as very natural because of the difficulty of checking the Polish property of the space. Actually we now see that the $\bar{d}$-topology for $\bar{\bar{\mathcal{G}}}$ is Polish, and this topology is nothing but the M_1-topology invented also by Skorokhod, except that we extend it to the space $\bar{\bar{\mathcal{G}}}$ instead of $\mathcal{G}$. We now introduce the *Skorokhod error modulus* of $X \in \bar{\bar{\mathcal{G}}}$ as the function $\hat{\omega}(\cdot;\ X)$ defined by

$$\hat{\omega}(\delta;\ X) = \sup\{\rho(X(t_0);\ [X(t_1);\ X(t_2)]) = t_0 - \delta < t_1 < t_0 < t_2 < t_0 + \delta\}\,.$$

It is readily observed that

LEMMA 13. $\lim_{\delta \downarrow 0} \hat{\omega}(\delta;\ X) = 0$.

LEMMA 14. *If* $\bar{x}_0,\ \bar{x}_1,\ \bar{x}_2 \in \hat{\Gamma} X$ *and* $\bar{x}_0 \in X[\bar{x}_1,\ \bar{x}_2]$, *with*

$$\Pi_I\,\bar{x}_1 - \delta < \Pi_I\,\bar{x}_1 < \Pi_I\,\bar{x}_0 < \Pi_I\,\bar{x}_2 < \Pi_I\,\bar{x}_0 + \delta\,,$$

then

$$\rho(\Pi_E\,\bar{x}_0;\ [\Pi_E\,x_1;\ \Pi_E\,\bar{x}_2]) \le \hat{\omega}(\delta;\ X)\,.$$

This means precisely that $\hat{\omega}(\delta;\ X)$ could be defined by the supremum of the left side of the last inequality. This observation and Lemma 10 lead to

LEMMA 15. *Given* $0<\delta_1<\delta_2$ *and* $\varepsilon>0$, *there exists a* $\delta>0$ *such that*: *if* $\bar{d}(X;\ Y)<\delta$, *then*

$$\hat{\omega}(\delta_1;\ Y)\le\hat{\omega}(\delta_2;\ X)+\varepsilon .$$

(Simply take $\delta=\min(\varepsilon/3,\ (\delta_2-\delta_1)/2)$.)

LEMMA 16. *If* t *is a continuity point of* X^0 *or* $t=0-,\ 1+$ *and* $X^n\to X^0$ *in* $(\overline{\overline{\mathscr{J}}})$, *then* $X^n(t)\to X^0(t)$.

THEOREM 3. *In* $\overline{\overline{\mathscr{J}}}$, *a sequence* $X^n\to X^0$ *iff*

a. *for a dense set of* t, *including* $0-,\ 1+$, *we have* $X^n(t)\to X^0(t)$, *and*

b. $\lim_{\delta\downarrow 0}\sup_n\hat{\omega}(\delta;\ X^n)=0$.

COROLLARY. *A set* $K\subset\overline{\overline{\mathscr{J}}}$ *is relatively compact iff*

a. $\{X(t): X\in K,\ t\in I\}$ *is relatively compact, and*

b. $\lim_{\delta\downarrow 0}\sup_{X\in K}\hat{\omega}(\delta,\ X)=0$

REMARK. This theorem and the corollary are proved by Skorokhod in [1] for $\mathscr{J}$. We note that the necessity follows from Lemmas 15 and 16. The sufficiency is easy and can be copied verbatim for the case $\overline{\overline{\mathscr{J}}}$. We should say that the compactness criterion in [1] is more awkward simply because $\mathscr{J}$ is *not* a closed set in $\overline{\overline{\mathscr{J}}}$. In order that $K\subset\mathscr{J}$ be relative compact in $\overline{\overline{\mathscr{J}}}$, with closure $\bar{K}\subset J$, a condition of equicontinuity at 0 and 1 has also to be added.

The following theorem follows from the compactness criterion and the Prokhorov theorem.

THEOREM 4. *In order that a set* F *of probability measures on* $\overline{\overline{\mathscr{J}}}$ *be weakly relatively compact, it is necessary and sufficient that, for any* $\varepsilon>0$, *there is a* $\delta>0$ *and compact set* K *of* E, *such that*

$$\sup_{\mathfrak{p}\in F}\mathfrak{p}\,\{X\in\overline{\overline{\mathscr{J}}}: X(t)\notin K \text{ for some } t\}<\varepsilon ,$$

and

$$\sup_{\mathfrak{p}\in F}\mathfrak{p}\,\{X\in\overline{\overline{\mathscr{J}}}: \hat{\omega}(\delta,\ X)>\varepsilon\}<\varepsilon .$$

We mentioned that $\overline{\overline{\mathscr{J}}}$ is well-behaved under the restriction map

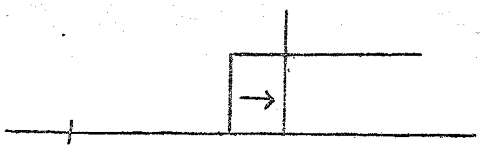
$$X\in\overline{\overline{\mathscr{J}}}([a,\ b],\ E)\mapsto X|_{[c-,\,d+]}\in\overline{\overline{\mathscr{J}}}([c,\ d]),\quad([c,\ d]\subset[a,\ b])$$

with obvious interpretation of notations. (Note that for

$X \in \mathscr{G}([a, b], E)$, the restriction $X|_{[c,d]}$ is not necessarily in $\mathscr{G}([c, d])$. By the compactness criterion we see that:

LEMMA 17. *If K is relatively compact in $\overline{\overline{\mathscr{G}}}[a, b]$, then $K|_{[c-,d+]} = \{X|_{[c-,d+)} : X \in K\}$ is relatively compact in $\overline{\overline{\mathscr{G}}}([c, d], E)$ for $[c, d] \subset [a, b]$.*

REMARK. But the restriction map is not continuous. See the following example:



On the other direction of "juxtaposing", we have the following. Again it is an easy consequence of the compactness criterion.

LEMMA 18. *Let $K \subset \overline{\overline{\mathscr{G}}}([a, b])$ and $c \in (a, b)$. If $K|_{[a-,c+]}$ is relatively compact in $\overline{\overline{\mathscr{G}}}([a, c])$, and $K|_{[c+,b]}$ is relatively compact and equicontinuous at $c+$, then K is relatively compact in $\overline{\overline{\mathscr{G}}}([a, b])$.*

We turn to another aspect with the well-behavior of $\overline{\overline{\mathscr{G}}}$ versus $\mathscr{G}$.

LEMMA 19. *Let $[c, d] \subset (a, b)$ and let $X \in \overline{\overline{\mathscr{G}}}([a, b])$. Then*

$$\lim_{\alpha \downarrow 0^-} X(\alpha + \cdot)|_{[c-,d+]} = X|_{[c+,d+]},$$

and

$$\lim_{\alpha \uparrow 0} X(\alpha + \cdot)|_{[c-,d+]} = X|_{[c-,d]}.$$

4. **The space of paths of memory.** We now introduce a complicated object, named in the title of this section. For definiteness we choose $I_0 = [-1, 0]$, $I = [0, 1]$, $\bar{I} = [-1, 1]$, and use the normalization that functions in $\overline{\overline{\mathscr{G}}}(\bar{I})$ are right-continuous in $(-1, 1)$.

For $c \in I$, $X \in \overline{\overline{\mathscr{G}}}(\bar{I})$ consider $X_c : t \in I_0 \mapsto X(c + t_0)$.

Now $c \mapsto X_c$ is of $\bar{J}$ type by the above lemma, i.e. this is a function in $\overline{\overline{\mathscr{G}}}(I, \overline{\overline{\mathscr{G}}}(I_0, E))$. This is written as HX and called the *memory displaying* of X.

LEMMA 20. *$X \mapsto HX$ is measurable.*

Proof. Because for every t, $X \mapsto HX(t) = X_t$ is measurable.

LEMMA 21. *If K is relatively compact in $\overline{\overline{\mathcal{G}}}(I;\ E)$, then $HKI = \{HX(t) : X \in K,\ t \in I\}$ is a relatively compact set in $\overline{\overline{\mathcal{G}}}(I_0;\ E)$.*

(Trivial by the compactness criterion.)

Let us estimate the Skorokhod modulus of HX. Let $0 \leq a < b \leq 1$ and consider $\bar{d}(Xa;\ Xb)$.

Let $b - a \leq \delta$, $\sup_{a \leq t \leq a+\delta} \|X(t) - X(a+0)\| \leq \varepsilon_2$, and $\sup_{a-1 \leq t \leq a-1+\delta} \|X(t) - X(a-1)\| \leq \varepsilon_1$.

It is easy to see the $\bar{d}(Xa;\ Xb) \leq \delta + \varepsilon_1 \vee \varepsilon_2$. But this is not a satisfactory estimate!

We could not assert the compactness of HK from the compactness of K.

If $0 \leq c - b \leq \delta$, with $(\varepsilon_3, \varepsilon_4)$ the analogue for (b, c) of $(\varepsilon_1, \varepsilon_2)$ above, then $\bar{d}(Xb;\ Xc) \leq \delta + \varepsilon_3 \vee \varepsilon_4$. The majorization

$$\min \{\bar{d}(Xa;\ Xb);\ \bar{d}(Xb;\ Xc)\} \leq \delta + \min\{\max\{\varepsilon_1', \varepsilon_2\}, \max\{\varepsilon_3, \varepsilon_4\}\}$$

is *not* useful, because we usually have only the majorization of $\varepsilon_1 \wedge \varepsilon_3$ and $\varepsilon_2 \wedge \varepsilon_4$. This is the tricky property of the Skorokhod modulus: it is the infimum of right-side and left-side deviation, and thus almost always it defies our control in any kind of composing.

But we could conceive of a situation when we have the majorization of $\varepsilon_1 \wedge \varepsilon_3$ and $\varepsilon_2 \vee \varepsilon_4$; then it is all right.

This has obviously its stochastic analogue:

THEOREM 5. *Let M be an equitight family of distributions on $(\mathcal{G}(\bar{I});\ \bar{d})$. If*

$$\lim_{\delta \downarrow 0} \sup_{P \in M} P\{(\varepsilon_1 \wedge \varepsilon_3) \vee (\varepsilon_3 \vee \varepsilon_4) > \varepsilon\} = 0, \qquad \forall \varepsilon > 0,$$

then by memory displaying we have an equitight family in $\mathcal{G}(I;\ \overline{\overline{\mathcal{G}}}(\Gamma_0))$.

Proof. By the above calculation and Lemma 14, the Skorokhod criterion is fulfilled.

REFERENCES

1. A. V. Skorokhod, *Limit theorems for stochastic processes,* Theor. Probability Appl. **1** (1956), 261–290.

2. W. Yang, *On the uniqueness of diffusions,* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **24** (1972), 247–261.

NATIONAL TAIWAN UNIVERSITY, TAIPEI, TAIWAN