RANDOM DYNAMICAL SYSTEMS ON ORDERED TOPOLOGICAL SPACES

Hans G. Kellerer

University of Munich

June 30, 2002

Let $(X_n,\ n \geq 0)$ be a random dynamical system and its state space be endowed with a reasonable topology. Instead of completing the structure as common by some linearity, this study stresses – motivated in particular by economic applications – order aspects. If the underlying random transformations are supposed to be order-preserving, this results in a fairly complete theory. First of all, the classical notions of and familiar criteria for recurrence and transience can be extended from discrete Markov chain theory. The most important fact is provided by existence and uniqueness of a locally finite invariant measure for recurrent systems. It allows to derive ergodic theorems as well as to introduce an attractor in a natural way. The classification is completed by distinguishing positive and null recurrence corresponding, respectively, to the case of a finite or infinite invariant measure; equivalently, this amounts to finite or infinite mean passage times. For positive recurrent systems, moreover, strengthened versions of weak convergence as well as generalized laws of large numbers are available.

Introduction. The random dynamical systems studied in this paper evolve in one-sided discrete time with independent and stationary increments. Therefore the formalism needed for more general models (see the monographs by Kifer [26] or Arnold [2]) is dispensable, and the process $(X_n,\ n \geq 0)$ can be introduced as an "iterated function system"

$$X_n = H_n(X_{n-1}) \quad \text{for} \quad n \in \mathbf{N}\,,$$

where $(H_n, n \in \mathbf{N})$ is a sequence of i. i. d. transformations of the state space E, independent of the initial E–valued variable X_0. If no specific structure in E has to be taken into account, this is just another way to introduce a homogeneous Markov chain on E (see von Weizsäcker [39]).

If, however, state space and mappings are supposed to have appropriate linearity (or smoothness) properties, the multiplicative ergodic theorem yields additional insight, provided the relevant integrability conditions are satisfied. Another common model – studied in particular by Barnsley and Elton (see e.g. [6, 5, 14]) – supposes E to be a complete metric space and the mappings H_n to satisfy an average contractivity. But in both cases research concerns primarily existence resp. uniqueness of an equilibrium and consequences on the long-term behaviour of the system. In both cases, too, the required moment conditions are by no means necessary.

To exemplify the problems left open, choose $E = \mathbf{R}_+$ and restrict H_n to affine maps, i.e. consider the "autoregressive model"

$$X_n = U_n X_{n-1} + V_n \quad \text{for} \quad n \in \mathbf{N}$$

2000 Mathematics Subject Classifications. Primary 60J05, 60H25; secondary 54H20, 54F05.

with a sequence of i. i. d. $\mathbf{R}_+^2$–valued variables (U_n, V_n). If $(S_n,\, n \geq 0)$ denotes the random walk with increments $\log U_n\, (\geq -\infty)$, then – without any moment conditions on U_n and under weak boundedness conditions on V_n – the following trichotomy is established in the preprint [24]:

(a) if $S_n \to -\infty$, then $(X_n,\, n \geq 0)$ is "positive recurrent",

(b) if $(S_n,\, n \geq 0)$ oscillates, then $(X_n,\, n \geq 0)$ is "null recurrent",

(c) if $S_n \to +\infty$, then $(X_n,\, n \geq 0)$ is "transient".

In the existing literature (see in particular the surveys by Vervaat [38] and by Embrechts/Goldie [15] and Goldie/Maller [20]) there is nearly no distinction between cases (b) and (c), though a common treatment of cases (a) and (b) is in fact much more adequate. But it is not the affine (and topological) structure that allows a natural extension of classical notions and central criteria from discrete Markov chain theory to an uncountable state space. While under a topological structure alone there is a variety of definitions for transience and (null or positive) recurrence (see e.g. the various notions in Tweedie [37] and Meyn/Tweedie [32]), the results obtained in the preprint [25] prove that order and topology combined provide an ideal framework for a fairly complete theory of random dynamical systems as considered in this paper.

To cover, however, autoregressive models of higher dimension or stochastic recursions of higher order, assuming E to be totally ordered, as is the case in [24] or [25], is too restrictive. In the sequel, therefore, these preprints are disregarded, and the state space is supposed to be any (partially) ordered topological space. In essence, there is only one restriction: E is supposed to be bounded from below, a condition that is largely in accordance with theory as well as with applications. Indeed, to give only two typical examples: most work on products of i. i. d. matrices is actually restricted to nonnegative matrices (see e.g. Högnäs/Mukherjea [22, Chapter 4]), while economic processes, modeling dams, insurance risk, queues, storage, traffic etc. (see e.g. Asmussen [3, Part C]) in general have 0 as natural lower bound. These and other examples, moreover, justify the restriction to mappings H_n that respect the structure of E, i.e. are order-preserving and continuous.

The precise assumptions on the state space E are collected at the beginning of Section 0. They are met not only in the classical case $E = \mathbf{R}_+^d$, but as well, for instance, by (rooted) tree models. The significance of the order is reflected by the singular role of order convex sets and functions of bounded variation, leading to the basic classes $\mathfrak{V}(E)$ and $\mathcal{V}(E)$, respectively. The extension of these notions from total to partial order is straightforward, apart from the fact that monotone sets or functions need not be measurable in the general case. But the arising problems, as some less common notions and facts concerning ordered topological spaces, are postponed to the final section. Random is introduced by specifying an element ν in the space $\mathsf{N}[E]$ of distributions on the space $\mathcal{H}[E]$ of order-preserving continuous transformations of E. This defines a (coupled) family of homogeneous Markov chains $(X_n^x, n \geq 0)$, depending on the initial state $x \in E$.

If the interest is not limited to stationary distributions and more generally recurrence properties are studied, the existence of not necessarily finite invariant measures for the underlying transition kernel is just as interesting. To assure uniqueness, being essential for ergodic theorems, restriction to suitably "irreducible" systems is inevitable. As it turns out, the adequate notion – introduced at the beginning of

Section 1 – employs the order instead of the topology of E. It has to be emphasized here that this, as all subsequent notions, is invariant under conjugation, i.e. under order-preserving homeomorphisms (not distinguishing, for instance, state spaces $\mathbf{R}_+$ and $[0,1[$). Since in general no natural metric is available in E, to answer the basic question of asymptotic equivalence of two processes $(X_n^x, n \geq 0)$ and $(X_n^y, n \geq 0)$ translations $f : E \to \mathbf{R}$ must be used. Again, topology has to be replaced by order: while the differences $|f(X_n^x) - f(X_n^y)|$ need not converge to zero for bounded continuous functions f, they even are summable, if f is of bounded variation. The fundamental inequality (1.4) and its corollaries are crucial in the sequel.

A first consequence is the zero-one law (2.1), applying to all sets in the algebra $\mathfrak{V}(E)$ and concerning the probability of an infinite number of visits by the process $(X_n,\, n \geq 0)$. Since this value is independent of the initial law, the definition of recurrence resp. transience in Section 2 is straightforward, if restricted to these sets. Then it is quite natural to introduce recurrent systems by employing the topological structure and requiring the existence of a compact recurrent set K. As can be shown only in (4.5), K may actually be replaced by an interval $[0, y]$ (where 0 stands for $\min E$). Transient systems are characterized in (2.6) in two different ways: by almost sure divergence of the process $(X_n,\, n \geq 0)$ to infinity (in the Alexandrov compactification) or, equivalently, by convergence of the associated potential kernel for all compact sets.

The main result of Section 3 is fundamental for all what follows: whenever the process $(X_n,\, n \geq 0)$ does not diverge to infinity, there is a unique invariant measure μ for the associated Markov kernel. The easy part is the existence proof, being based on the topological properties of the state space. To establish uniqueness, E is exhausted by an increasing sequence of compact sets, and the classical ergodic theorem is applied to the associated embedded systems. The main tool, however, are the summability results from (1.5). An example at the end of the section shows that the uniqueness result in fact requires a restriction to measures compatible with the topological structure, i.e. to Radon measures. The only results related to (3.4), known so far, are due to Babillot et al. [4], who study order-preserving affine systems of a special structure and under suitable moment conditions (see the remarks following (9.3)). In addition, they derive a limit theorem for occupation times as considered in the next section.

Without postulating the invariant measure to be finite, the ergodic theorems in Section 4 have to concern ratios, regarding bounded functions with compact support. While the version for means in (4.2) is restricted to functions in $\mathcal{K}(E)$, the pointwise version in (4.3) can be established for functions of bounded variation and their uniform limits, defining a class $\mathcal{R}(E)$ that in general is much larger than $\mathcal{K}(E)$. As a consequence recurrent sets can simply be described as those having positive invariant measure. More generally, (4.4) characterizes recurrence resp. transience of a given system through hitting probabilities as well as through the potential kernel – in complete analogy to the well-known criteria for discrete Markov chains. The section concludes by deriving a strong version of irreducibility and aperiodicity in the recurrent case.

From the above description of recurrent sets it is easily deduced that, with probability 1, the set of limit points of a recurrent system equals the support M of the invariant measure μ, independently of the initial law. By (5.2) it is justified to call M the "attractor" of the system, whether μ is finite or not. Since M need not be compact, to state its self-similarity in general involves a passage to the closure. Nevertheless it allows to characterize the attractor as the smallest nonempty closed set that is mapped into itself by the mappings H_n with probability 1. A more explicit characterization of the attractor is established at the end of Section 5: M can be identified with the set of constants in the closed semigroup generated by the mappings that define the system, i.e. that are contained in the support $\mathcal{N}$ of ν. If the state space is only partially ordered, actually "local boundedness" of E has to be supposed here (see the final section).

As mentioned above, the existing literature concerns almost exclusively positive recurrent systems, i.e. the case of a finite invariant measure, as treated in Section 6. There are numerous studies considering $E = \mathbf{R}_+$ with its total order and random transformations preserving this order as, for instance, Alpuim/Athayde [1], Goldie [19], Helland/Nilsen [21], Lund et al. [30], Yahav [40]. There is, moreover, a number of papers dealing also with only partially ordered state spaces as Bougerol/Picard [10], Diaconis/Freedman [12, Section 3], Glasserman/Yao [18], Jarner/Tweedie [23], Mairesse [31], Rachev/Samorodnitsky [34], Rachev/Todorovic [35], and in particular Brandt et al. [11, Section 1.3] (where, however, a metric compatible with the order is postulated). The full benefit from order and monotonicity turns out under time reversal, i.e. replacing left by right composition of the underlying random transformations. The resulting (non-Markovian) "dual process" $(Y_n^0,\, n \geq 0)$ increases pointwise, and its limit Y satisfies a zero-one law, distinguishing positive and null recurrence, where in the positive recurrent case (6.7) reveals the stationary distribution as the law of Y. Combined with the asymptotic behaviour of the original process $(X_n^0, n \geq 0)$ this yields the criterion (6.8) and the unified characterization of positive recurrence, null recurrence, and transience at the end of the section.

Another criterion for positive resp. null recurrence can be established by considering the hitting times T_B of sets B that are determined by the order. While for increasing intervals $[x, \cdot\,]$ it is shown already in (1.2) that $\mathbf{E}(T_{[x,\cdot]})$, regardless of the initial law, is always finite, (7.1) exhibits a different behaviour for decreasing intervals $[0, x]$. Now, with T_B^y denoting the hitting time of B for the process starting at y, for any $x \in M$ the mean passage time $\mathbf{E}(T_{[0,x]}^x)$ is finite if and only if the system is positive recurrent. The main tool in the proof is (a slight extension of) the recurrence theorem of Kac. Section 7 concludes by the topological analogue (7.3) of the classical result on mean passage times for discrete Markov chains (here again local boundedness enters).

In Section 8 the ratio ergodic theorems are considerably strengthened for positive recurrent systems. In this case the laws $\mu_n = \mathcal{L}(Y_n^0)$ of the dual process, hence also the laws $\mathcal{L}(X_n^0) = \mu_n$, are easily seen to converge weakly to the stationary distribution μ. By (1.6) this extends to arbitrary initial laws and from functions in $\mathcal{C}(E)$ to those in the class $\mathcal{R}(E)$. The resulting convergence for all decreasing sets suggests the introduction of a metric d for distributions on the space E (lacking any metric structure) such that metric convergence is strictly stronger than weak convergence. Employing an idea in Dubins/Freedman [13] (apparently the first treatment of iterated function systems, though restricted to the case $E = [0, 1]$), it is then possible to verify even $d(\mu_n, \mu) \to 0$ in (8.3) – with geometric convergence whenever the state space is bounded. The section concludes by proving the mixing property for the stationary version $(X_n,\, n \geq 0)$ and deriving from it in (8.5) a fairly general law of large numbers, again independent of the initial law.

Basic assumption throughout the preceding sections is – apart from the existence of $\min E$ – the irreducibility of the system. Under a total ordering this means in fact no real restriction: simply reduce the state space to those elements that can be reached from 0 in the sense of (1.1). But this procedure need not work in general, as is demonstrated in Section 9. The arising problems, however, can be settled as above, whenever the system is strictly order-preserving. This is a familiar hypothesis in affine models, for instance, where it concerns the occurrence of zero entries in the associated random matrix. For strictly order-preserving systems the reduction works as in the totally ordered case and leads to a state space that is in addition locally bounded (as required for one half of (5.7) and (7.3)).

The concluding Section 10 collects the necessary material on ordered topological spaces. The fundamental fact needed here is the existence of an open base consisting of order convex sets. This is a simple consequence of Nachbin's extension theorem, i.e. of Tietze's extension theorem carried over to ordered topological spaces. Results not to be found in the standard reference [33] concern the Alexandrov compactification of ordered topological spaces with a lower bound in (10.5), its consequences for the approximation by nonnegative decreasing functions from $\mathcal{K}(E)$ in (10.6) and (10.7), and finally the explicit representation of sets in the algebra $\mathfrak{V}(E)$ resp. of functions in the vector lattice $\mathcal{V}(E)$ seen from the viewpoint of universal measurability in (10.8).

0. Preliminaries. Throughout Sections 1 – 9 the state space E is an ordered topological space (the definition being recalled in Section 10), where regarding the topology it is sufficient to suppose that

(E1) *E is locally compact and second countable,*

while regarding the order it is necessary to suppose that

(E2) *E has a lower bound, denoted by* 0.

If $E^* = E \cup \{\infty\}$ is the Alexandrov compactification of E (∞ being an isolated point if E is compact), the order is extended from E to E^* in accordance with (E2) by letting ∞ be an upper bound of E^*. This makes E^* again an ordered topological space if and only if (see (10.5))

(E3) *the decreasing hull of a compact subset of E is compact,*

entailing compactness in particular for each interval $[0, x]$. In the sequel conditions (E1) – (E3) will be taken for granted, calling E *admissible* in this case.

The topological notations are as usual: $\mathfrak{G}(E)$ / $\mathfrak{F}(E)$ / $\mathfrak{K}(E)$ / $\mathfrak{B}(E)$ denote, respectively, the class of open / closed / compact / Borel subsets of E, while $\mathcal{C}(E)$ / $\mathcal{K}(E)$ / $\mathcal{B}(E)$ stand, respectively, for the class of bounded continuous / compactly supported continuous / Borel functions $f : E \to \mathbf{R}$.

The following notations refer to the order: If A is a subset of E, then $A^{\downarrow}$ and $A^{\uparrow}$ denote, respectively, its decreasing and increasing hull, i.e.

$$A^{\downarrow} = \bigcup\nolimits_{x \in A} [0, x] \quad \text{and} \quad A^{\uparrow} = \bigcup\nolimits_{x \in A} [x, \cdot\,] .$$

A is called (*order*) *convex* if $A \ni x \le y \in A$ implies $[x, y] \subset A$ or, equivalently, if $A = A^{\downarrow} \cap A^{\uparrow}$. Arrows are also used to denote, for instance, by $\mathfrak{G}^{\downarrow}(E)$ the class of decreasing sets in $\mathfrak{G}(E)$ or by $\mathcal{B}^{\uparrow}(E)$ the class of increasing functions in $\mathcal{B}(E)$.

To extend the notion of *bounded variation* from total to partial order, let $\mathcal{V}(E)$ be the class of functions $f \in \mathcal{B}(E)$ satisfying

$$\sup \left\{ \sum_{k \in \mathbf{N}} |f(x_{k+1}) - f(x_k)| : x_1 \leq x_2 \leq \ldots \right\} < \infty$$

and $\mathfrak{V}(E)$ the class of sets $B \in \mathfrak{B}(E)$ with $1_B \in \mathcal{V}(E)$. Universal null sets disregarded, $\mathcal{V}(E)$ is the linear space generated by the bounded monotone functions, while $\mathfrak{V}(E)$ is the algebra generated by the convex sets (see (10.8)). The closure of $\mathcal{V}(E)$ with respect to the uniform norm, yielding the class of *regular* functions (having finite limits from left and right everywhere on E^*) under a total ordering and including $\mathcal{K}(E)$ in the general case (see (10.6)), is denoted by $\mathcal{R}(E)$.

$\mathsf{M}(E)$ denotes the class of all locally finite measures on $\mathfrak{B}(E)$, which due to (E1) are Radon measures and can be identified with the positive linear functionals on $\mathcal{K}(E)$. If μf denotes the $\mu-$integral of a function f, then $\mathsf{M}(E)$ is endowed with the vague (weak*) topology, generated by the mappings $\mu \mapsto \mu f$, $f \in \mathcal{K}(E)$; the corresponding convergence is denoted by $\underset{v}{\to}$. On the subspace $\mathsf{M}_1(E)$ of probability measures this induces the weak (narrow) topology, generated by the mappings $\mu \mapsto \mu f$, $f \in \mathcal{C}(E)$; the corresponding convergence is denoted by $\underset{w}{\to}$.

The space $\mathcal{C}[E] := \mathcal{C}(E, E)$ of continuous mappings from E to E, endowed with the compact-open topology, is a Polish space due to (E1), and this carries over to the closed subspace $\mathcal{H}[E]$ of order-preserving continuous mappings $h : E \to E$, being a central object of this paper. Under composition $\mathcal{H}[E]$ is a subsemigroup of $\mathcal{C}[E]$, where composition is continuous by the local compactness of E and thus measurable by the second countability of $\mathcal{C}[E]$.

The main object of this paper is the space $\mathsf{N}[E]$ of distributions ν on $\mathcal{H}[E]$. The semigroup structure of $\mathcal{H}[E]$ induces a convolution in $\mathsf{N}[E]$, making this space a semigroup itself. Occuring powers are denoted by ν^n, hence

$$\int_{\mathcal{H}[E]} f(h(x))\, \nu^n(dh) = \int_{\mathcal{H}[E]} \cdots \int_{\mathcal{H}[E]} f(h_1 \circ \ldots \circ h_n\,(x))\, \nu(dh_1) \ldots \nu(dh_n)$$

for $x \in E$, $f \in \mathcal{C}(E)$ and $n \in \mathbf{N}$, while ν^0 is the unit measure ε_h with h being the identy map. Since $\mathcal{H}[E]$ is second countable, the support $\mathcal{N}$ is well-defined for $\nu \in \mathsf{N}[E]$.

Now the stochastic model can be formally introduced. Let be given an admissible space E and, on some probability space $(\Omega, \mathfrak{A}, \mathbf{P})$,

- a sequence of independent random variables $H_n : \Omega \to \mathcal{H}[E]$ with identical distribution $\nu \in \mathsf{N}[E]$,
- a random variable $X_0 : \Omega \to E$ that is independent of $(H_n, n \in \mathbf{N})$.

This defines an *order-preserving random dynamical system* by

$$X_n = H_n(X_{n-1}) \quad \text{for } \; n \in \mathbf{N}\,.$$

Therefore the distribution of $(X_n,\, n \geq 0)$ is completely determined by ν and the initial law $\mu_0 = \mathcal{L}(X_0)$. If in particular $X_0 = x$, this will be expressed by the notation $(X_n^x,\, n \geq 0)$, i.e.

$$X_n^x = H_n \circ \ldots \circ H_1\,(x) \quad \text{for } \; x \in E \; \text{and} \; n \geq 0\,.$$

Thus for general μ_0 conditional probabilities are given by

$$\mathbf{P}^x((X_n,\, n \geq 0) \in B) = \mathbf{P}((X_n^x,\, n \geq 0) \in B)$$

with an analogous equation for conditional expectations.

As usual, the initial law is largely of secondary importance, and the primary component is the distribution ν. Therefore, whenever possible, it will be briefly referred to the (*dynamical*) *system* (E,ν). All notions to be defined in Sections 1 – 9 will depend on ν, but this dependence will be suppressed in related notations (as for the support $\mathcal{N}$ above), because ν is supposed to be fixed.

Clearly, $(X_n,\, n \geq 0)$ is a homogeneous Markov chain. Its transition kernel P transforms (nonnegative) functions $f \in \mathcal{B}(E)$ into Pf given by

$$Pf(x) = \int_{\mathcal{H}[E]} f(h(x))\,\nu(dh) \qquad \text{for } \; x \in E$$

and a measure μ on E into μP given by

$$\mu P(B) = \int_E \nu(h(x) \in B)\,\mu(dx) \qquad \text{for } \; B \in \mathfrak{B}(E)\,,$$

which in the $\sigma-$finite case equals

$$\mu P(B) = \int_{\mathcal{H}[E]} \mu(h(x) \in B)\,\nu(dh) \qquad \text{for } \; B \in \mathfrak{B}(E)\,.$$

P is a Feller kernel, which in addition transforms increasing functions into functions of the same type. By (E2) this implies in particular:

(0.1) LEMMA *The sequence* $(\varepsilon_0\, P^n,\, n \geq 0)$ *is stochastically increasing, i.e.*

$$\mathbf{E}(f(X^0_{n-1})) \leq \mathbf{E}(f(X^0_n)) \qquad \textit{for } \; 0 \leq f \in \mathcal{B}^{\uparrow}(E)\,.$$

PROOF. This is immediate from the equations

$$\mathbf{E}(f(X^0_{n-1})) = \int_E P^{n-1} f\,(0)\,P(0;dx)\,,$$

$$\mathbf{E}(f(X^0_n)) = \int_E P^{n-1} f\,(x)\,P(0;dx)\,. \qquad \square$$

Finally, it has to be mentioned that the passage from the distribution ν to the kernel P in general is not injective.

1. Irreducible systems. Let (E,ν) be a dynamical system as introduced in the preceding section. Then, to classify it as recurrent or transient, requires some communication structure to prevent the state space from splitting into different classes. It turns out to be sufficient that, starting from the minimal state 0, each state x with positive probability can be reached or exceeded in the following sense:

(1.1) DEFINITION The system (E,ν) is called "(*upwards*) *irreducible*", if for any $x \in E$ there is some $n \in \mathbf{N}$ such that

$$P^n(0;[x,\cdot]) \;=\; \mathbf{P}(X^0_n \geq x) \;=\; \nu^n(h(0) \geq x) \;>\; 0\,.$$

It is immediate from (0.1) that the system (E,ν^n) is irreducible for all $n \in \mathbf{N}$, whenever this holds for one $n \in \mathbf{N}$.

The question, whether irreducibility can be accomplished by suitably reducing the state space, is postponed to Section 9. As an example consider the "Cantor system", assigning mass $\nu(\{h_i\}) = \frac{1}{2}$ to the two mappings $h_1 : x \mapsto x/3$ and $h_2 : x \mapsto x/3 + 2/3$: the adequate state space in the present setting is $E = [0, 1[$.

In spite of its weak appearance, the condition in (1.1) has strong recurrence implications for increasing intervals:

(1.2) PROPOSITION *If the system (E, ν) is irreducible, then for arbitrary initial law and all $x \in E$*

(a) $$\mathbf{P}(\limsup_{n\to\infty} \{X_n \geq x\}) = 1\,,$$

(b) $$\mathbf{E}(T_{[x,\cdot]}) < \infty\,,$$

where $T_B := \inf\{n \in \mathbf{N} : X_n \in B\}$ $(\leq \infty)$ for $B \in \mathfrak{B}(E)$.

PROOF. By assumption $\vartheta := \nu^n(h(0) \geq x) > 0$ for some n, which yields by monotonicity and independence

$$\begin{aligned}
\mathbf{P}(T_{[x,\cdot]} > kn) &\leq \mathbf{P}^0(T_{[x,\cdot]} > kn)\\
&\leq \textstyle\prod_{0\leq i<k} (1 - \mathbf{P}(H_{(i+1)n} \circ \ldots \circ H_{in+1}(0) \geq x))\\
&= (1-\vartheta)^k \quad \text{for all } \; k \in \mathbf{N}.
\end{aligned}$$

This settles (b), which in turn implies (a) by the Markov property. □

Under a total ordering the state space of an irreducible system, due to (E1), is a countable union of intervals $[0, y_n]$. Somewhat surprisingly, this result extends to the general case:

(1.3) PROPOSITION *If the system (E, ν) is irreducible, then there is a countable subset A of E such that $E = A^{\downarrow}$.*

PROOF. If $(Y_n^0,\, n \geq 0)$ is an independent copy of $(X_n^0,\, n \geq 0)$, then (1.2a) implies

$$\mathbf{P}\Big(\bigcap\nolimits_{m\in\mathbf{N}} \bigcup\nolimits_{n\in\mathbf{N}} \{x_m \leq Y_n^0\}\Big) = 1 \quad \text{for all } \; (x_n,\, n \in \mathbf{N}) \in E^{\mathbf{N}}.$$

Applying Fubini to the product of the distributions of $(X_n^0,\, n \geq 0)$ and $(Y_n^0,\, n \geq 0)$ twice, it follows that

$$\mathbf{P}\Big(\bigcap\nolimits_{m\in\mathbf{N}} \bigcup\nolimits_{n\in\mathbf{N}} \{X_m^0 \leq y_n\}\Big) = 1 \quad \text{for some } \; (y_n,\, n \in \mathbf{N}) \in E^{\mathbf{N}}.$$

For any $x \in E$ and $l \in \mathbf{N}$ satisfying $\mathbf{P}(X_l^0 \geq x) > 0$ therefore

$$\{X_l^0 \geq x\} \cap \Big(\bigcap\nolimits_{m\in\mathbf{N}} \bigcup\nolimits_{n\in\mathbf{N}} \{X_m^0 \leq y_n\}\Big) \neq \emptyset\,,$$

hence for $m = l$ in particular

$$\{X_l^0 \geq x\} \cap \Big(\bigcup\nolimits_{n\in\mathbf{N}} \{X_l^0 \leq y_n\}\Big) \neq \emptyset\,.$$

This implies $x \in A^{\downarrow}$ for $A := \{y_n : n \in \mathbf{N}\}$. □

The following inequality is the crucial tool in the sequel:

(1.4) THEOREM *Let the system (E,ν) be irreducible and suppose $0 \le f \in \mathcal{B}^{\uparrow}(E)$. Then for all $x \in E$*

$$\sum\nolimits_{n\ge 0} \mathbf{E}(f(X_n^x) - f(X_n^0)) \;\le\; \mathbf{E}(T^0_{[x,\cdot]}) \sup_{n\ge 0} \mathbf{E}(f(X_n^0))\,,$$

where $T_B^y := \inf\{n \in \mathbf{N} : X_n^y \in B\}$ $(\le \infty)$ for $y \in E$ and $B \in \mathfrak{B}(E)$.

PROOF. If f is replaced by $f \wedge k$, $k \in \mathbf{N}$, then the corresponding differences on the left-hand side increase for $k \to \infty$ to the difference in question, i.e. f may be assumed to be bounded. In addition, by (1.2b) the stopping time $T := T^0_{[x,\cdot]}$ may be assumed to be finite everywhere. Then, with the notation $A_l := \{T \le l\}$, for $n \ge 0$ fixed and $m \le n$ arbitrary

$$(*)\qquad \begin{aligned} \mathbf{E}(1_{A_{n-m}} f(X^0_{T+m})) &= \mathbf{E}(1_{A_{n-m}} f(H_{T+m}\circ\ldots\circ H_{T+1}(X_T^0)))\\ &\ge \mathbf{E}(1_{A_{n-m}} f(H_{T+m}\circ\ldots\circ H_{T+1}(x)))\\ &= \mathbf{P}(A_{n-m})\,\mathbf{E}(f(X_m^x))\,, \end{aligned}$$

where the inequality is a consequence of f being increasing, while the final equality uses the fact that $(H_{T+1},\ldots,H_{T+m})$ is independent of T and distributed as $(H_1,\ldots,H_m)$. But, f is nonnegative and thus

$$\sum\nolimits_{0\le m\le n} f(X_m^0) \;\ge\; \sum\nolimits_{0\le m,\,T+m\le n} f(X^0_{T+m}) \;=\; \sum\nolimits_{0\le m\le n} 1_{A_{n-m}} f(X^0_{T+m})\,.$$

This leads by integration and inserting from $(*)$ to

$$\begin{aligned} &\sum\nolimits_{0\le m\le n} \mathbf{E}(f(X_m^x) - f(X_m^0))\\ \le\;& \sum\nolimits_{0\le m\le n} \mathbf{E}(f(X_m^x)) \;-\; \sum\nolimits_{0\le m\le n} \mathbf{P}(A_{n-m})\,\mathbf{E}(f(X_m^x))\\ =\;& \sum\nolimits_{0\le m\le n} (1-\mathbf{P}(A_{n-m}))\,(\mathbf{E}(f(X_m^x) - f(X_m^0)) + \mathbf{E}(f(X_m^0)))\,, \end{aligned}$$

which by (0.1) implies

$$\begin{aligned} &\sum\nolimits_{0\le m\le n} \mathbf{P}(T \le n-m)\,\mathbf{E}(f(X_m^x) - f(X_m^0))\\ \le\;& \sum\nolimits_{0\le m\le n} \mathbf{P}(T > n-m)\,\mathbf{E}(f(X_m^0))\\ \le\;& \mathbf{E}(T)\,\mathbf{E}(f(X_n^0))\,. \end{aligned}$$

Since $\mathbf{P}(T \le n-m) \uparrow 1$ for $n \to \infty$, the assertion follows. □

Apart from arising measurability problems the preceding result can be easily extended to a larger class of functions:

(1.5) PROPOSITION *Let the system (E,ν) be irreducible and suppose $f \in \mathcal{V}(E)$. Then for all $x \in E$*

$$\text{(a)}\qquad \sum\nolimits_{n\ge 0} \mathbf{E}(|f(X_n^x) - f(X_n^0)|) < \infty\,,$$

and for arbitrary initial law

$$\text{(b)}\qquad \sum\nolimits_{n\ge 0} |f(X_n) - f(X_n^0)| < \infty \quad a.s.\,.$$

PROOF. According to (10.8) there are bounded increasing and universally measurable functions $f_i \geq 0$ such that $f = f_1 - f_2$. Since the variables $X_n^x : \Omega \to E$ are universally measurable as well, provided the underlying probability space $(\Omega, \mathfrak{A}, \mathbf{P})$ is assumed to be complete, the proof of (1.4) works for f_i, too. In view of (1.2b) this proves assertion (a), which by Fubini implies assertion (b). □

For some applications, where summability of the differences may be replaced by convergence to zero, the class of admissible functions can be enlarged once more:

(1.6) PROPOSITION *Let the system (E, ν) be irreducible and suppose $f \in \mathcal{R}(E)$. Then for arbitrary initial law*

$$f(X_n) - f(X_n^0) \to 0 \quad a.s..$$

PROOF. This is immediate from (1.5b) and the definition of $\mathcal{R}(E)$. □

It has to be emphasized that, even under a total ordering, the last result may fail for functions $f \in \mathcal{C}(E)$. To construct a nondegenerate (i.e. recurrent) counterexample consider the following system: on $E = \mathbf{R}_+$ let the support $\mathcal{N}$ of ν consist of the fractional linear mappings

$$h_1 : x \mapsto x + 1 \quad \text{and} \quad h_2 : x \mapsto x \,/\, (x+1)\,.$$

Starting with $n_0 := 0$ choose a sequence $(n_k, k \geq 0)$ such that

$$\limsup_{k\to\infty} \mathbf{P}(A_k) = 1 \quad \text{with} \quad A_k := \bigcup\nolimits_{n_{k-1} < n \leq n_k} \{X_n^0 \geq k\}\,,$$

as is possible in view of (1.2). Denote the finite support of $\mathcal{L}(X_n^x)$ by B_n^x and define the discrete sets

$$B^x := \bigcup\nolimits_{k\in\mathbf{N}} \Big(\Big(\bigcup\nolimits_{n_{k-1}<n\leq n_k} B_n^x\Big) \cap [k, \infty[\Big) \qquad \text{for} \;\; x \in E\,.$$

Then the sets B^0 and B^y are disjoint for fixed irrational y, due to $h_i^{-1}[\mathbf{Q}_+] = \mathbf{Q}_+$ for $i = 1, 2$. Thus there is a function $f \in \mathcal{C}(E)$ with $0 \leq f \leq 1$, satisfying $f(x) = 0$ for $x \in B^0$ and $f(x) = 1$ for $x \in B^y$. Then on A_k almost surely

$$\max_{n_{k-1}<n\leq n_k} (f(X_n^y) - f(X_n^0)) = 1\,,$$

hence in view of $\mathbf{P}(\limsup_{k\to\infty} A_k) = 1$ finally

$$\limsup_{n\to\infty} (f(X_n^y) - f(X_n^0)) = 1 \;\; \text{a.s.}.$$

2. Recurrence and transience. In general, the probability that the process $(X_n^x, n \geq 0)$ visits a set $B \in \mathfrak{B}(E)$ infinitely often neither obeys a zero-one law nor is independent of the starting point $x \in E$ – as can be seen, for instance, by the Cantor system even for appropriate countable sets. But an application of the final result of Section 1 yields:

(2.1) THEOREM *Let the system (E, ν) be irreducible and suppose $B \in \mathfrak{V}(E)$. Then*

$$\mathbf{P}\Big(\limsup_{n\to\infty} \{X_n \in B\}\Big) = 0 \;\; or \;\; 1\,,$$

where the value is independent of the initial law.

PROOF. Choosing $f = 1_B \in \mathcal{V}(E)$ in (1.6) proves

$$\limsup_{n\to\infty} \{X_n \in B\} = \limsup_{n\to\infty} \{X_n^0 \in B\} =: A^0 \quad \text{a.s.}.$$

Applying this result to the distribution of X_1^0 as initial law and the shifted sequence $(H_n, n > 1)$ yields

$$A^0 = \limsup_{n\to\infty} \{H_n \circ \ldots \circ H_2\,(0) \in B\} \quad \text{a.s.}.$$

Repeating the argument shows A^0 to be contained in the completed tail σ–field of $(H_n, n \in \mathbf{N})$, and the assertion follows. □

The preceding result justifies:

(2.2) DEFINITION Let the system (E, ν) be irreducible. Then $B \in \mathfrak{V}(E)$ is called "*recurrent*", if

$$\mathbf{P}\Big(\limsup_{n\to\infty} \{X_n^x \in B\}\Big) = 1 \qquad \text{for one (or all) } x \in E\,,$$

and "*transient*" otherwise.

The following consequence of recurrence will be needed:

(2.3) LEMMA *Let the system (E, ν) be irreducible and $B \in \mathfrak{V}(E)$ be recurrent. Then for any finite subset A of E there exists some $n \in \mathbf{N}$ such that*

$$\mathbf{P}(X_n^x \in B \ \textit{for all}\ x \in A) > 0\,.$$

PROOF. According to (1.6)

$$1_B(X_n^x) - 1_B(X_n^0) \to 0 \ \text{a.s.} \qquad \text{for all } x \in A\,.$$

Since A is finite and B is recurrent, this implies

$$\limsup_{n\to\infty} \{X_n^x \in B \ \text{for all}\ x \in A\} = \limsup_{n\to\infty} \{X_n^0 \in B\} = \Omega \quad \text{a.s.}.$$

Therefore the probabilities in question even sum up to ∞. □

It is immediate from (1.2) that each increasing set $B \neq \emptyset$ is recurrent. For decreasing sets the following criterion is available:

(2.4) PROPOSITION *Let the system (E, ν) be irreducible. Then for arbitrary initial law a set $B \in \mathfrak{B}^{\downarrow}(E)$ is recurrent if and only if $\sum_{n\geq 0} \mathbf{P}(X_n \in B) = \infty$.*

PROOF. Due to (2.1), the condition is certainly necessary. To prove its sufficiency, let B be transient and assume without restriction $X_0 = 0$. Therefore the variable $Z := \sum_{n\in\mathbf{N}} 1_B(X_n^0)$ satisfies $\vartheta := \mathbf{P}(Z \geq l) < 1$ for suitable $l \in \mathbf{N}$. Using the stopping time

$$T := \inf\{n \in \mathbf{N} : \textstyle\sum_{1\leq m\leq n} 1_B(X_m^0) = l\}$$

with respect to $(H_n, n \in \mathbf{N})$, it follows as in the proof of (1.4) that

$$\begin{aligned}\mathbf{P}(Z \geq k\,l) &= \mathbf{P}(T < \infty\,,\ \sum\nolimits_{n\in\mathbf{N}} 1_B\,(H_{T+n} \circ \ldots \circ H_{T+1}\,(X_T^0)) \geq (k-1)\,l) \\ &\leq \mathbf{P}(T < \infty\,,\ \sum\nolimits_{n\in\mathbf{N}} 1_B\,(H_{T+n} \circ \ldots \circ H_{T+1}\,(0)) \geq (k-1)\,l) \\ &= \mathbf{P}(Z \geq l)\,\mathbf{P}(Z \geq (k-1)\,l)\,.\end{aligned}$$

Therefore

$$\mathbf{P}(Z \geq k\,l) \leq \vartheta^k \quad \text{for all } \ k \in \mathbf{N}$$

and thus indeed $\mathbf{E}(Z) < \infty$. □

The topological structure enters in the central classification:

(2.5) DEFINITION An irreducible system (E, ν) (or the process $(X_n,\, n \geq 0)$ or the kernel P) is called "*recurrent*", if $\mathfrak{V}(E)$ contains a compact recurrent set K, and "*transient*" otherwise.

In Section 4, using the invariant measure, it will be shown that the set K required in this definition may in fact be postulated to be an interval $[0, y]$.

The assumption (E3), connecting compactness and order in Section 0, is essential for the following criterion:

(2.6) THEOREM *Let the system (E, ν) be irreducible. Then for arbitrary initial law the following conditions are equivalent:*

(1) $$(E, \nu) \textit{ is transient}\,,$$

(2) $$X_n \to \infty \quad a.s.\,,$$

(3) $$\sum\nolimits_{n\geq 0} \mathbf{P}(X_n \in K) < \infty \quad \textit{for all } \ K \in \mathfrak{K}(E)\,.$$

PROOF. The equivalence (1) ⇔ (2) follows from the existence of a sequence of sets $K_l \in \mathfrak{K}(E)$ such that each $K \in \mathfrak{K}(E)$ is included in some K_l. The implication (1) ⇒ (3) is a consequence of (2.4), because the compact set K, due to (E3), may be assumed to be decreasing. The final implication (3) ⇒ (1) is straightforward. □

Restricting condition (3) to decreasing sets K, it follows by (0.1) that the system (E, ν^n) is recurrent for all $n \in \mathbf{N}$, whenever this holds for one $n \in \mathbf{N}$.

For an application consider an "exchange process" as has been studied in [21] and in its best accessible form is given by the recursion

$$X_n = (X_{n-1} - 1) \vee U_n \quad \text{for } \ n \in \mathbf{N}$$

with a sequence of i. i. d. variables $U_n \geq 0$. Here, X_{n-1} is the utility of some equipment in use at time $n-1$, losing one unit during period n, and U_n the utility of a new equipment available at time n. The corresponding mappings $h : x \mapsto (x-1) \vee u$, restricted to the state space

$$E = \{x \geq 0 : \mathbf{P}(U_n \geq x) > 0\}\,,$$

belong to $\mathcal{H}[E]$ for all $u \in E$. With $\nu \in \mathsf{N}[E]$ being the associated distribution, the system (E, ν) is obviously irreducible. The explicit representation

$$X_n^0 = (U_1 - (n-1)) \vee \ldots \vee (U_{n-1} - 1) \vee U_n \quad \text{for } \ n \in \mathbf{N}$$

implies by independence

$$\mathbf{P}(X_n^0 \le y) = \prod\nolimits_{0 \le m < n} F(y+m)\,,$$

where F denotes the common distribution function of U_n, $n \in \mathbf{N}$. Therefore (2.6) yields recurrence whenever E is bounded, because the summands $\mathbf{P}(X_n^0 \le y)$ are strictly positive and eventually constant for any $y \in E$ satisfying $\mathbf{P}(U_n \le y) > 0$. For $E = \mathbf{R}_+$ both cases are possible:

(1) if U_n, $n \in \mathbf{N}$, have the common density $f_1(x) = (x+1)^{-2}$, then $F(y) = \dfrac{y}{y+1}$ implies

$$\sum\nolimits_{n \ge 0} \mathbf{P}(X_n^0 \le y) = \sum\nolimits_{n \ge 0} \frac{y}{y+n} = \infty \qquad \text{for all } \; y > 0\,,$$

i.e. the process is recurrent;

(2) if the density is replaced by $f_2(x) = 2x(x+1)^{-3}$, then $F(y) = (\dfrac{y}{y+1})^2$, and it follows similarly that the process is transient.

It should be noted that – in spite of the contrasting asymptotic behaviour – the variables U_n behave similarly in both cases as far as it concerns the existence of moments, due to $f_1(x) \le f_2(x) \le 2f_1(x)$ for $x \ge 1$ (for a continuation see Sections 3, 4, and 6).

In general, as in discrete Markov chain theory, it may demand some effort to decide whether a system (E, ν) is recurrent or transient. For an example consider $E = [0, 1[$ with ν assigning mass $\frac{1}{2}$ to the mappings defined by $h_1(x) = (x+1)/2$ and $h_2(x) = x^2$ (solvable by conjugation).

3. Invariant measures. As in discrete Markov chain theory, a central question concerns the existence and uniqueness of invariant measures in the recurrent case. In accordance with the topological assumptions, this question will – and must – be treated within the class $\mathsf{M}(E)$ of locally finite measures. The easy task here is existence, where general results by Foguel [17] apply. Since the arguments can be simplified, due to monotonicity, the proof of the following assertion is outlined:

(3.1) PROPOSITION *Let the system (E, ν) be irreducible and suppose*

$$0 \le g \in \mathcal{K}^{\downarrow}(E) \quad \textit{with} \quad \sum\nolimits_{n \ge 0} \mathbf{E}(g(X_n^0)) = \infty\,.$$

Then the measures $\varrho_n \in \mathsf{M}(E)$ defined by

$$\varrho_n(f) := \sum\nolimits_{0 \le m < n} \mathbf{E}(f(X_m^0)) \Big/ \sum\nolimits_{0 \le m < n} \mathbf{E}(g(X_m^0)) \qquad \textit{for } \; f \in \mathcal{K}(E)$$

for all $n \ge n_0$ (say) satisfy

(a) *the sequence $(\varrho_n, n \ge n_0)$ is relatively compact,*

(b) *each limit point $\mu \in \mathsf{M}(E)$ is a nontrivial invariant measure for (E, ν) (i.e. for the kernel P).*

PROOF. (a) By assumption and (1.5a) the functions

$$s_n := \sum\nolimits_{0 \le m < n} P^m g \in \mathcal{C}(E) \qquad \text{for } \; n \in \mathbf{N}$$

increase to infinity for all $x \in E$. Therefore for fixed $f \in \mathcal{K}(E)$ there exists $l \in \mathbf{N}$ with $f \le s_l$, which by (0.1) implies

$$P^m f\,(0) \le P^m s_l\,(0) \le l\,P^m g\,(0) \quad \text{for all} \;\; m \ge 0\,.$$

Thus the sequence $(\varrho_n(f), n \ge n_0)$ is bounded for all $f \in \mathcal{K}(E)$, i.e. the sequence $(\varrho_n, n \ge n_0)$ is uniformly locally finite, which proves (a).

(b) If $\varrho_{n_k} \underset{v}{\to} \mu \in \mathsf{M}(E)$, then $\mu g = 1$ and thus μ is indeed nontrivial. To prove its invariance consider $0 \le f \in \mathcal{K}(E)$ and approximate $0 \le Pf \in \mathcal{C}(E)$ from below by functions from $\mathcal{K}(E)$. This yields

$$\begin{aligned}
\mu P f &\le \liminf_{k\to\infty} \varrho_{n_k} P f \\
&= \liminf_{k\to\infty} s_{n_k}(0)^{-1} \sum\nolimits_{0<m\le n_k} P^m f\,(0) \\
&= \liminf_{k\to\infty} s_{n_k}(0)^{-1} \sum\nolimits_{0\le m<n_k} P^m f\,(0) \\
&= \mu f\,,
\end{aligned}$$

because $s_{n_k}(0) \to \infty$ and $P^{n_k} f\,(0)$ is bounded by $\max f$. Applying the resulting inequality $\mu P \le \mu$ to $s_l - f \ge 0$, with f and l chosen as in the proof of (a), yields

$$\mu f - \mu P f \le \mu s_l - \mu P s_l \le \mu g =: \gamma\,.$$

Disregarding the middle term, this implies

$$\mu f - \mu P f \le \gamma \quad \text{for all} \;\; f \in \mathcal{K}(E)\,,$$

and – replacing f by multiples – proves the invariance of μ. □

Uniqueness of the invariant measure is established by a localization, which requires an elementary result from probabilistic potential theory:

(3.2) Lemma *Let the system (E, ν) be irreducible and $K \in \mathfrak{K}^{\downarrow}(E)$ be recurrent. With the almost surely finite hitting time T_K define the Markov kernel ${}^K\!P$ by*

$${}^K\!P(x; B) := \mathbf{P}^x(X_{T_K} \in B) \quad \textit{for} \;\; x \in K \;\; \textit{and} \;\; B \in \mathfrak{B}(K)\,.$$

If a measure $\mu \in \mathsf{M}(E)$ is invariant for P, then its restriction ${}^K\!\mu \in \mathsf{M}(K)$ is invariant for ${}^K\!P$.

Proof. $\mu P \le \mu$ implies ${}^K\!\mu\, {}^K\!P \le {}^K\!\mu$ (see e.g. Proposition 2.2.6 in [36]). Since ${}^K\!\mu$ is a finite measure and ${}^K\!P$ is a stochastic kernel, in fact equality has to hold in this inequality. □

For the main step towards uniqueness the results from Section 1 are crucial:

(3.3) Proposition *Let the system (E, ν) be irreducible and $K \in \mathfrak{K}^{\downarrow}(E)$ be recurrent. Moreover, let $\mu \in \mathsf{M}(E)$ be invariant for (E, ν) and satisfy $\mu(K) > 0$. Then for $0 \le f \in \mathcal{B}^{\downarrow}(E)$ with supp $f \subset K$ and all $x \in E$*

$$\sum\nolimits_{0\le m<n} f(X_m^x) \Big/ \sum\nolimits_{0\le m<n} 1_K(X_m^x) \to \mu f / \mu(K) \;\; \textit{a.s.}\,.$$

Proof. 1. The variables

$$\underline{Q}^x := \liminf_{n\to\infty} \Big(\sum\nolimits_{0\le m<n} f(X_m^x) \Big/ \sum\nolimits_{0\le m<n} 1_K(X_m^x) \Big) \quad \text{for} \;\; x \in E$$

are almost surely independent of x. Indeed:

(1) In the denominator x can be replaced by 0, because the quotient

$$\sum_{0\le m<n} 1_K(X_m^x) \Big/ \sum_{0\le m<n} 1_K(X_m^0)$$

$$= 1 + \Big(\sum_{0\le m<n} (1_K(X_m^x) - 1_K(X_m^0)) \Big/ \sum_{0\le m<n} 1_K(X_m^0)\Big)$$

tends to 1 almost surely by (1.5b) and the recurrence of K.

(2) In the nominator x can be replaced by 0, because the difference

$$\sum_{0\le m<n} f(X_m^x) \Big/ \sum_{0\le m<n} 1_K(X_m^0) \;-\; \sum_{0\le m<n} f(X_m^0) \Big/ \sum_{0\le m<n} 1_K(X_m^0)$$

$$= \sum_{0\le m<n} (f(X_m^x) - f(X_m^0)) \Big/ \sum_{0\le m<n} 1_K(X_m^0)$$

tends to 0 almost surely for the same reasons.

Neglecting single summands and applying Fubini, it follows as in the proof of (2.1) that in fact $\underline{Q}^x$ equals some constant $\underline{q}$ almost surely.

2. Because of $0 < \mu(K) < \infty$ the restriction ${}^K\mu$ of μ to K may be assumed to be normalized, hence by (3.2) to be a stationary distribution for the Markov kernel ${}^KP$. Let $({}^KX_n,\, n \ge 0)$ denote the corresponding "embedded process" of $(X_n,\, n \ge 0)$, where X_0 is distributed according to (the trivial extension of) ${}^K\mu$. Since f is bounded, the classical ergodic theorem applies to the stationary process $({}^KX_n,\, n \ge 0)$, i.e. the limit

$$Q := \lim_{n\to\infty} \frac{1}{n} \sum_{0\le m<n} f({}^KX_m)$$

exists almost surely and in view of $\operatorname{supp} f \subset K$ satisfies

$$\mathbf{E}(Q) = \mathbf{E}(f({}^KX_0)) = \mu f\,.$$

But on $\{X_0 = x\}$ the sequence of quotients defining $\underline{Q}^x$ arises from the successive means of $(f({}^KX_n),\, n \ge 0)$ through "extension to the right by constancy" in an evident sense. Therefore, applying Fubini once more, it turns out that the variable Q and the constant $\underline{q}$ agree almost surely and thus $\underline{q} = \mu f$. Since the argument carries over to the upper limit, the assertion is established. □

Now one of the main results of this paper can be stated:

(3.4) THEOREM *For any recurrent system (E,ν) there exists one and – up to multiples – only one nontrivial invariant measure $\mu \in \mathsf{M}(E)$.*

PROOF. 1. If $K \in \mathfrak{K}^{\downarrow}(E)$ is recurrent and $g_l \in \mathcal{K}^{\downarrow}(E)$ with $g_l \ge 1_K$ is chosen according to (10.7), then the condition in (3.1) is satisfied for $g = g_l$ by (2.4), and the existence of μ is established.

2. To prove uniqueness, let $\mu' \in \mathsf{M}(E)$ be another nontrivial invariant measure for (E,ν) and exhaust E by an increasing sequence of sets $K_l \in \mathfrak{K}^{\downarrow}(E)$, where $\mu(K_l) > 0$ and $\mu'(K_l) > 0$ and in addition recurrence of K_l may be assumed. Then an application of (3.3) yields constants γ_l such that

$$\mu'(1_{K_l} f) = \gamma_l\, \mu(1_{K_l} f) \quad \text{for } 0 \le f \in \mathcal{B}^{\downarrow}(E)\,.$$

Inserting $f = 1_{K_m}, m \neq l$, proves γ_l to be a constant γ independent of l and thus, with $\mathcal{K}_0$ as defined in (10.6a),

$$\mu'(1_{K_l} f) = \gamma\, \mu(1_{K_l} f) \quad \text{for } \; f \in \mathcal{K}_0 \,.$$

This equation extends to $f \in \mathcal{K}(E)$, because the uniform approximation of f by a sequence $(f_k,\, k \in \mathbf{N})$ from $\mathcal{K}_0$ according to (10.6a) can be carried out with all functions vanishing outside a fixed compact set (otherwise multiply by some g_l from (10.7)). The assertion follows for $l \to \infty$. □

As usual the actually one-dimensional family of invariant measures $\mu \in \mathsf{M}(E)$ will be briefly called *the invariant measure* in the sequel.

It has to be emphasized that, even under a total ordering, the uniqueness statement in (3.4) concerns locally finite measures only. To exhibit a counterexample, choose $E = [0, 1[$ and let ν assign mass $\frac{1}{2}$ to the piecewise affine maps defined by

$$h_1(x) = \frac{1}{3}\,(2x+1) \quad \text{and} \quad h_2(x) = \frac{1}{2}\, x \vee (2x-1)\,.$$

Then (E, ν) is clearly irreducible, and the uniform distribution μ on E is easily checked to be a finite invariant measure. With μ as initial law the series $\sum_{n \geq 0} \mathbf{P}(X_n \leq y)$ diverges for all $y > 0$, and thus (E, ν) is recurrent by (2.6). On the other hand, it is not hard to check that

$$\mu' := \sum\nolimits_{x \in D} x\, \varepsilon_x \quad \text{with} \quad D := E \cap \mathbf{Q}$$

defines another invariant measure, which, however, is $\sigma-$finite only. Incidentally, the existence of both a finite and an infinite, but $\sigma-$finite, invariant measure in this example proves the Doeblin-Harris theory of recurrent Markov chains on an abstract state space to be inadequate in the present setting.

Finally, it has to be mentioned that, again even under a total ordering, there is no converse of (3.4), i.e. there are transient systems with a unique nontrivial and locally finite invariant measure. For an example consider the transient case (2) of the exchange process from Section 2. It is easily seen that a measure $\mu \in \mathsf{M}(E)$ is invariant if and only if the function $\Phi(y) := \mu([0, y])$ satisfies

$$\Phi(y) = \Phi(y+1)\, F(y) = \Phi(y+1)\, (\frac{y}{y+1})^2 \qquad \text{for} \;\; y \geq 0\,.$$

Therefore $\Phi_0(y) = y^2$ defines a solution, while for any other solution Φ by iteration

$$\Phi(y) = \Phi(y+n)\, (\frac{y}{y+n})^2 \qquad \text{for} \;\; y \geq 0 \text{ and } n \in \mathbf{N}\,.$$

For $0 \leq y \leq m \in \mathbf{N}$ this provides by monotonicity the bounds

$$\Phi(y) = \Phi(y+n)\, (\frac{y}{y+n})^2 \geq \Phi(\,0\,+n)\, (\frac{y}{y+n})^2 = \Phi(1)\, y^2\, (\frac{n}{y+n}\,)^2\,,$$

$$\Phi(y) = \Phi(y+n)\, (\frac{y}{y+n})^2 \leq \Phi(m+n)\, (\frac{y}{y+n})^2 = \Phi(1)\, y^2\, (\frac{m+n}{y+n})^2\,,$$

which for $n \to \infty$ lead to $\Phi(y) = \Phi(1)\, y^2$ and thus to the asserted uniqueness.

4. Ergodic theorems. The first theorem in this section requires some preparation, concerning again asymptotic independence of the initial law:

(4.1) PROPOSITION *Let the system* (E,ν) *be recurrent with invariant measure* μ *and suppose* $0 \le f \in \mathcal{B}^{\downarrow}(E)$ *with* $\mu f > 0$. *Then for arbitrary initial law*

$$\sum\nolimits_{0\le m<n} \mathbf{E}(f(X_m)) \Big/ \sum\nolimits_{0\le m<n} \mathbf{E}(f(X_m^0)) \;\to\; 1\,.$$

PROOF. With the abbreviation

$$s_n^x := \sum\nolimits_{0\le m<n} \mathbf{E}(f(X_m^x)) \qquad \text{for } x \in E$$

an application of (1.4) to the increasing function $f' := f(0) - f \ge 0$ yields by (1.2b)

$$0 \le s_n^0 - s_n^x \;\le\; \mathbf{E}(T^0_{[x,\cdot]})\, f(0) < \infty\,.$$

If $K \in \mathfrak{K}^{\downarrow}(E)$ is a recurrent set with $\int_K f\,d\mu > 0$, then (3.3) implies

$$\sum\nolimits_{0\le m<n} 1_K f(X_n^0) \to \infty \;\; \text{a.s.}\,,$$

because the analogous result for 1_K replacing $1_K f$ follows from the recurrence of K. Therefore $s_n^0 \to \infty$ and thus the quotients s_n^x/s_n^0 converge to 1. Since this convergence is dominated by the constant 1, it continues to hold when integrated by the initial law. □

Now a mean ergodic theorem can be established:

(4.2) THEOREM *Let the system* (E,ν) *be recurrent with invariant measure* μ *and suppose* $f,\, g \in \mathcal{K}(E)$. *Then for arbitrary initial law*

$$\sum\nolimits_{0\le m<n} \mathbf{E}(f(X_m)) \Big/ \sum\nolimits_{0\le m<n} \mathbf{E}(g(X_m)) \;\to\; \mu f/\mu g\,,$$

whenever the quotient on the right-hand side makes sense.

PROOF. 1. First, the problem will be simplified in four steps:
(1) By comparing both the nominator and the denominator with the corresponding sum for a third function it is seen that $0 \le g \in \mathcal{K}^{\downarrow}(E)$ with $\mu g > 0$ may be assumed in the sequel.
(2) Since the uniform approximation of f by a sequence $(f_k,\, k \in \mathbf{N})$ from $\mathcal{K}_0$ according to (10.6a) can be carried out from below and above with all functions vanishing outside a fixed compact set (otherwise multiply by some g_l from (10.7)), f may be restricted to $\mathcal{K}_0$.
(3) Therefore it is sufficient to consider the case $0 \le f \in \mathcal{K}^{\downarrow}(E)$, where in addition $\mu f > 0$ may be assumed (otherwise replace f by $f + g$).
(4) Thus f and g both satisfy the assumptions of (4.1), which justifies finally the restriction to the initial law ε_0.

2. Now the relative compactness from (3.1a) and the uniqueness of the invariant measure yield a constant γ such that

$$\sum\nolimits_{0\le m<n} \mathbf{E}(f(X_m^0)) \Big/ \sum\nolimits_{0\le m<n} \mathbf{E}(g(X_m^0)) \;\to\; \gamma\, \mu f$$

with an analogous relation for g replacing f. This shows $\gamma = 1/\mu g$ and concludes the proof. □

It has to be mentioned that the question, whether the convergence in (4.2) extends from f, $g \in \mathcal{K}(E)$ to functions in $\mathcal{R}(E)$ with compact support, poses an open problem. An affirmative answer is possible under a total ordering (see Theorem 6.3 in [25]) as well as concerning the following pointwise ergodic theorem:

(4.3) THEOREM *Let the system (E,ν) be recurrent with invariant measure μ and the functions f, $g \in \mathcal{R}(E)$ have compact support. Then for arbitrary initial law*

$$\sum\nolimits_{0\le m<n} f(X_m) \Big/ \sum\nolimits_{0\le m<n} g(X_m) \;\to\; \mu f/\mu g \quad a.s.,$$

whenever the quotient on the right-hand side makes sense.

PROOF. Assume $X_0 = x$, as is justified by Fubini, and simplify the problem as in step (1) of the proof of (4.2) by considering $g = 1_K$ with $\operatorname{supp} f \subset K \in \mathfrak{K}^{\downarrow}(E)$ and $\mu(K) > 0$. Then proceed as follows:

(1) For $f \in \mathcal{V}(E)$ choose a representation $f = f_1 - f_2$ according to (10.8) and a constant $\gamma \ge f_1(0) \vee f_2(0)$, so that $f = f_1' - f_2'$ with universally measurable decreasing functions $f_1' := \gamma - f_2 \ge 0$ and $f_2' := \gamma - f_1 \ge 0$. If in addition $\operatorname{supp} f_i' \subset K$, as can be achieved by multipying with 1_K, then it follows as in the proof of (1.5a) that the arguments for (3.3) work as well for the functions f_i'.

(2) For $f \in \mathcal{R}(E)$ approximate uniformly by a sequence $(f_k,\, k \in \mathbf{N})$ from $\mathcal{V}(E)$, which can be carried out from below and above with $\operatorname{supp} f_k \subset K$ for all $k \in \mathbf{N}$ (otherwise multiply again by 1_K). □

As a first application of these ergodic theorems consider the recurrent case (1) of the exchange process from Section 2. Here $\mathbf{P}(X_n^0 \le y) = \dfrac{y}{y+n}$ implies

$$\mathbf{E}(f(X_n^0)) = \frac{1}{n}\int_0^\infty f(y)\left(\frac{y}{n}+1\right)^{-2} dy \qquad \text{for } n \in \mathbf{N}$$

and thus

$$\mathbf{E}(f(X_n^0)) \Big/ \mathbf{E}(g(X_n^0)) \;\to\; \int_0^\infty f(y)\,dy \Big/ \int_0^\infty g(y)\,dy \qquad \text{for } f,\, g \in \mathcal{K}(\mathbf{R}_+).$$

Since this convergence carries over to the corresponding quotients in (4.2), the invariant measure is simply the Lebesgue measure λ restricted to $\mathbf{R}_+$. By (4.3) this implies that, regardless of the initial law, the process $(X_n,\, n \ge 0)$ is "equidistributed" on $\mathbf{R}_+$, i.e.

$$\sum\nolimits_{0\le m<n} 1_{I_1}(X_m) \Big/ \sum\nolimits_{0\le m<n} 1_{I_2}(X_m) \;\to\; \lambda(I_1)/\lambda(I_2) \quad \text{a.s.}$$

for subintervals I_k of $\mathbf{R}_+$ of positive and finite length.

By means of (4.3) recurrence of a set $B \in \mathfrak{V}(E)$ can be seen to be simply equivalent to $\mu(B) > 0$. More generally, extending (2.1) and (2.4) and replacing sets by functions, the following dichotomy holds:

(4.4) THEOREM *Let the system (E,ν) be recurrent with invariant measure μ. Then for $0 \le f \in \mathcal{V}(E)$ and any $x \in E$*

(a) *$\mu f > 0$ implies*

$$\sum\nolimits_{n\ge 0} f(X_n^x) = \infty \quad a.s. \qquad and \qquad \sum\nolimits_{n\ge 0} \mathbf{E}(f(X_n^x)) = \infty,$$

(b) $\mu f = 0$ *implies*

$$\sum_{n\geq 0} f(X_n^x) < \infty \quad a.s. \qquad \text{and} \qquad \sum_{n\geq 0} \mathbf{E}(f(X_n^x)) < \infty\,.$$

PROOF. (a) Choose $K \in \mathfrak{K}^{\downarrow}(E)$ with $\int_K f\,d\mu > 0$ and apply (4.3) to the functions $1_K f$ and 1_K in $\mathcal{V}(E)$. This yields the first (and second) assertion.

(b) Use the invariance of μ to obtain

$$\int_E \Big(\sum_{n\geq 0} \mathbf{E}(f(X_n^x))\Big)\,\mu(dx) \;=\; \sum_{n\geq 0} \mu P^n f = 0\,.$$

Therefore the integrand vanishes $\mu-$almost everywhere and thus by (1.5a) is finite for all $x \in E$. This proves the second (and first) assertion. □

Together, (4.4) and (2.6) imply that the two familiar criteria for recurrence resp. transience from discrete Markov chain theory carry over to the present setting in the following form:

(1) If (E,ν) is recurrent, then for $x \in \operatorname{supp}\mu$ the assertion

$$\mathbf{P}^x(X_n \in G \text{ infinitely often}) = 1\,,$$

hence

$$\mathbf{E}^x(|\{n \geq 0 : X_n \in G\}|) = \infty\,,$$

holds, whenever G is an open neighborhood of x.

(2) If ν is transient, then for arbitrary $x \in E$ the assertion

$$\mathbf{E}^x(|\{n \geq 0 : X_n \in K\}|) < \infty\,,$$

hence

$$\mathbf{P}^x(X_n \in K \text{ infinitely often}) = 0\,,$$

holds, whenever K is a compact subset of E.

Clearly, under a total ordering the compact recurrent set in (2.5) can be required to be an interval. Now it is possible to extend this simplification to the general case:

(4.5) THEOREM *Let the system (E,ν) be recurrent. Then there exists $y \in E$ such that $K = [0,y]$ is recurrent.*

PROOF. Since (1.3) provides $y \in E$ with invariant measure $\mu([0,y]) > 0$, the assertion is an immediate consequence of (4.4a). □

Explicitly, this proves an irreducible system (E,ν) to be recurrent if and only if

$$\sum_{n\geq 0} \mathbf{P}(X_n^0 \leq y) = \infty \qquad \text{for some } \; y \in E\,.$$

As another consequence of (4.4), a recurrent system can be shown to be irreducible and aperiodic in a very strong sense. The essential step concerns decreasing intervals:

(4.6) LEMMA *Let the system (E,ν) be recurrent with invariant measure μ. Then any $y \in E$ with $\mu([0,y]) > 0$ satisfies*

$$\mathbf{P}(X_n^y \leq y) > 0 \qquad \textit{for almost all } \; n \geq 0\,.$$

PROOF. The Markov property and monotonicity imply that

$$D := \{n \geq 0 : \mathbf{P}(X_n^y \leq y) > 0\}$$

is an additive semigroup. Therefore it is sufficient to prove $d = 1$ for the greatest common divisor d of D. Now the assumption $d > 1$ yields $\mathbf{P}(X_{nd+1}^y \leq y) = 0$ for all $n \geq 0$ and thus by (1.5a)

$$\sum_{n \geq 0} \mathbf{P}(X_{nd+1}^0 \leq y) < \infty .$$

But this leads to a contradiction, because the probabilities $\mathbf{P}(X_n^0 \leq y)$ decrease by (0.1) and sum up to infinity by (4.4a). □

It has to be added that under a total ordering $\mu([0,y]) > 0$ in fact implies $\mathbf{P}(X_n^y \leq y) > 0$ for all $n \geq 0$. Indeed: the assumption $\nu(h(y) > y) = 1$ yields $\mathbf{P}(X_n^y > y) = 1$ for all $n \in \mathbf{N}$, while $\mu([0,y]) > 0$ implies $X_n^y \leq y$ infinitely often with probability 1.

Now (4.6) can be considerably extended:

(4.7) PROPOSITION *Let the system (E, ν) be recurrent with invariant measure μ. Then for convex sets $B \in \mathfrak{B}(E)$ with $\mu(B) > 0$ and arbitrary initial law*

$$\mathbf{P}(X_n \in B) > 0 \quad \textit{for almost all } n \geq 0 .$$

PROOF. Fix first $y \in E$ with $\mu([0,y]) > 0$, as is justified by (1.3). Then apply (4.4a) and (4.1) to get $k \in \mathbf{N}$ such that

$$\mathbf{P}(X_k \leq y) > 0 . \tag{1}$$

Next, use (4.6) to obtain $l \in \mathbf{N}$ such that

$$\mathbf{P}(X_n^y \leq y) > 0 \quad \text{for } n \geq l . \tag{2}$$

Finally, apply (4.4a) to $f = 1_B$ and (2.3) to $A = \{0, y\}$ to find $m \in \mathbf{N}$ such that

$$\mathbf{P}(X_m^z \in B \text{ for all } z \leq y) > 0 . \tag{3}$$

Combining (1) – (3) it follows easily that

$$\mathbf{P}(X_n^x \in B) > 0 \quad \text{for } n \geq k + l + m .$$

□

5. The attractor.

The following notation will be used in the sequel:

(5.1) DEFINITION Let the system (E, ν) be recurrent with invariant measure μ. Then the closed set $M := \operatorname{supp} \mu$ is called the "*attractor*" of (E, ν).

This terminology is readily justified:

(5.2) THEOREM *Let the system (E, ν) be recurrent with invariant measure μ. Then for arbitrary initial law the random set*

$$L(\omega) := \{x \in E : x \textit{ is limit point of } (X_n(\omega),\, n \geq 0)\}$$

equals the attractor M with probability 1.

PROOF. According to (10.4) the second countable space E has a base consisting of convex open sets G_k, $k \in \mathbf{N}$, which in addition may be assumed to be relatively compact (otherwise choose functions g_l, $l \in \mathbf{N}$, according to (10.7) and take intersections with the sets $\{g_l > 0\}$). Then the inclusion $M \subset L(\omega)$ is obviously equivalent to

$$\sum_{n\geq 0} 1_{G_k}(X_n(\omega)) = \infty \quad \text{whenever} \quad G_k \cap M \neq \emptyset\,, \tag{1}$$

and it is easily checked that the inclusion $L(\omega) \subset M$ is equivalent to

$$\sum_{n\geq 0} 1_{G_k}(X_n(\omega)) < \infty \quad \text{whenever} \quad \overline{G}_k \cap M = \emptyset\,, \tag{2}$$

where $\overline{A}$ denotes the closure of a subset A of E. Now it follows from (1.5b) that X_n in (1) and (2) can be replaced by X_n^0 without changing probabilities. Then (4.4) applies: the series in (1) diverges almost surely because of $\mu(G_k) > 0$ for $G_k \cap M \neq \emptyset$, while the series in (2) converges almost surely because of $\mu(G_k) = 0$ otherwise. □

The following properties of μ and M will be needed in the sequel:

(5.3) PROPOSITION *Let the system (E, ν) be recurrent with invariant measure μ and attractor M. Then for each $x \in E$ there exists $y \geq x$ such that $\mu([x, y]) > 0$ and $M \cap [x, y] \neq \emptyset$.*

PROOF. The interval $[x, \cdot]$ is recurrent by (1.2a), hence satisfies $\mu([x, \cdot]) > 0$ by (4.4b). Now (1.3) provides $y \in E$ with $\mu([x, y]) > 0$, which implies the second assertion due to $\mu(E \setminus M) = 0$. □

The self-similarity of the attractor is a direct consequence of the topological assumptions. Since M need not be compact, however, it involves a passage to the closure:

(5.4) PROPOSITION *Let the system (E, ν) be recurrent with invariant measure μ. Then its attractor M satisfies*

$$M = \overline{\bigcup_{h\in\mathcal{N}} h[M]}\,.$$

PROOF. Since the spaces E and $\mathcal{H}[E]$ both are second countable, the product $\mu \otimes \nu$ is again a Radon mesure, and its support is given by $M \times \mathcal{N}$. In view of the continuity of the mapping $\varphi : (x, h) \mapsto h(x)$ the image of $\mu \otimes \nu$ under φ is easily checked to have as its support the closure of $\varphi[M \times \mathcal{N}]$. But this image equals μ, and thus the assertion is established. □

Now M can be characterized as is familiar from elementary models:

(5.5) PROPOSITION *Let the system (E, ν) be recurrent with invariant measure μ. Then its attractor M is the smallest nonempty set $F \in \mathfrak{F}(E)$ satisfying the condition*

$$\nu(h[F] \subset F) = 1\,, \tag{1}$$

or, equivalently, the condition

$$h[F] \subset F \quad \text{for all} \quad h \in \mathcal{N}\,. \tag{2}$$

PROOF. Since the mappings $h \mapsto h(x), x \in F$, are continuous, the set of mappings $h \in \mathcal{H}[E]$ with $h[F] \subset F$ is closed, and thus both conditions are clearly equivalent. Since, moreover, (5.4) shows that (2) is satisfied for $F = M$, it remains only to prove $F \supset M$ for any nonempty set $F \in \mathfrak{F}(E)$ satisfying (1). But this condition clearly continues to be satisfied, if ν is replaced by ν^n. For any $x \in F$ this implies $\mathbf{P}(X_n^x \in F) = 1$ for all $n \geq 0$, and thus the process $(X_n^x,\, n \geq 0)$ has with probability 1 all its limit points in F. Therefore the inclusion $M \subset F$ is a consequence of (5.2). $\square$

It is immediate from version (2) that the attractor depends on the distribution ν only through its support $\mathcal{N}$.

Another application concerns the special case of a totally ordered state space. Here $z := \min M$ exists and by (5.5) satisfies $h(z) \in M$, hence $h(z) \geq z$ for all $h \in \mathcal{N}$, while on the other hand by (5.2) any $x \in E$ with $h(x) \geq x$ for all $h \in \mathcal{N}$ yields $M \subset [x, \cdot]$, hence $z \geq x$. Put together, this means

$$\min M = \max\{x \in E : h(x) \geq x \text{ for all } h \in \mathcal{N}\}.$$

The term "attractor" does not mean that, regardless of the initial law, the event "$X_n \in M$ eventually" has probability 1 – as can be seen, for instance, by the Cantor system (E, ν), where $h_i\,[E \setminus M] \subset E \setminus M$ for $i = 1, 2$. In the present setting there is a substitute:

(5.6) PROPOSITION *If the system (E, ν) is recurrent with attractor M, then for arbitrary initial law*

$$\mathbf{P}(X_n \in M^{\uparrow} \textit{ for almost all } n \geq 0) = 1.$$

PROOF. To settle first the necessary measurability, note that M is σ–compact and thus $M^{\uparrow} = \bigcup_{l \in \mathbf{N}} K_l^{\uparrow}$ with $K_l \subset \mathfrak{K}(E)$, hence $K_l^{\uparrow} \in \mathfrak{F}(E)$ as follows by sequential compactness of E. Therefore $M^{\uparrow}$ is in fact of type F_{σ}. Since (5.5) implies

$$h[M^{\uparrow}] \subset (h[M])^{\uparrow} \subset M^{\uparrow} \quad \text{for all } h \in \mathcal{N},$$

the process $(X_n,\, n \geq 0)$ stays in $M^{\uparrow}$ almost surely whenever entering this set. But by (1.2a) this entrance occurs with probability 1. $\square$

For a final characterization of M let $\mathcal{N}^{\circ}$ $(\overline{\mathcal{N}^{\circ}})$ denote the (closed) semigroup generated by $\mathcal{N}$. Then the following criterion holds:

(5.7) PROPOSITION *Let the system (E, ν) be recurrent with invariant measure μ and denote by j the canonical injection of E into $\mathcal{H}[E]$. Then the attractor M satisfies*

(a) $j(x) \in \overline{\mathcal{N}^{\circ}} \;\Rightarrow\; x \in M$,

(b) $j(x) \in \overline{\mathcal{N}^{\circ}} \;\Leftarrow\; x \in M$, *provided E is locally bounded (see* (10.2)).

PROOF. (a) By (5.5) the inclusion $h[M] \subset M$ holds for $h \in \mathcal{N}$, hence for $h \in \overline{\mathcal{N}^{\circ}}$. Thus the assertion follows from $h[M] = \{x\}$ for $h = j(x)$.

(b) By definition of the topology in $\mathcal{H}[E]$ it has to be shown that

$$\{h \in \mathcal{N}^{\circ} : h[K] \subset G\} \neq \emptyset$$

for $G \in \mathfrak{G}(E)$ containing x, hence satisfying $\mu(G) > 0$, and arbitrary $K \in \mathfrak{K}(E)$. Since E is locally bounded, K can be covered by a finite number of bounded open sets and thus $K \subset A^{\downarrow}$ for some finite subset A of E. Since, moreover, 0 can be included in A and G can be supposed to be convex by (10.4), it is sufficient to prove

$$\{h \in \mathcal{N}^{\circ} : h(y) \in G \text{ for all } y \in A\} \neq \emptyset .$$

But translated into probabilities this is implied by

$$\mathbf{P}(X_n^y \in G \text{ for all } y \in A) > 0 \quad \text{for some } n \in \mathbf{N},$$

which follows indeed from (4.4a) and (2.3). □

Since a totally ordered space is locally bounded, in this case by (5.7) simply $M = j^{-1}[\overline{\mathcal{N}^{\circ}}]$. Whether this equation extends to the general case, remains an open problem (see, however, (9.5)). It has to be mentioned here that the assumption $j^{-1}[\overline{\mathcal{N}^{\circ}}] \neq \emptyset$ is the starting point of the PhD thesis [7], which tries to carry over methods and results from [24] and [25] to random dynamical systems defined by contractions of a complete metric space.

6. Positive recurrence and null recurrence. As usual two kinds of recurrence have to be distinguished:

(6.1) DEFINITION A recurrent system (E, ν) with invariant measure μ is called "*positive recurrent*" if μ is finite and "*null recurrent*" otherwise.

By (2.6) recurrence resp. transience of the system (E, ν^n) was seen to be independent of $n \in \mathbf{N}$; this extends to the present classification, because the invariant measures for these systems agree.

Nontrivial and locally finite invariant measures may exist also in the transient case, as is shown at the end of Section 3. Therefore the following fact has to be confirmed explicitly:

(6.2) PROPOSITION *If the system (E, ν) is irreducible, positive recurrence is equivalent to the existence of a stationary distribution μ.*

PROOF. It is sufficient to deduce recurrence from the existence of μ. But choosing μ as initial law and $K \in \mathfrak{K}(E)$ with $\mu(K) > 0$, this is immediate from (2.6). □

Clearly, compactness of the state space always implies positive recurrence; more generally:

(6.3) PROPOSITION *An irreducible system (E, ν) is positive recurrent under each of the following conditions:*

(a) $\nu(h[E]$ *is relatively compact*$) > 0$,

(b) *E contains a maximal element y.*

PROOF. (a) Choose a sequence of sets $K_l \in \mathfrak{K}(E)$ such that each $K \in \mathfrak{K}(E)$ is included in some K_l. Then the assumption concerns the union of the sets $\{h \in \mathcal{H}[E] : h[E] \subset K_l\}$, which due to E and K_l being of type K_σ and G_δ, respectively, are again of type G_δ. This settles the question of measurability and ensures $\gamma := \nu(h[E] \subset K_m) > 0$ for some m, which in view of the estimate

$$\mathbf{P}(X_n \in K_m) \geq \mathbf{P}(H_n(x) \in K_m \text{ for all } x \in E) = \gamma \quad \text{for all } n \in \mathbf{N}$$

proves the recurrence of (E, ν) by (2.6). Invariance of μ, moreover, implies

$$\mu(K_m) = \int_E \nu(h(x) \in K_m)\, \mu(dx) \geq \gamma\, \mu(E)$$

and thus $\mu(E) < \infty$ as asserted.

(b) The maximality of x and the monotonicity of $h \in \mathcal{H}[E]$ prove $h(0) \geq x$ and $h[E] = \{x\}$ to be equivalent statements. Since $\nu^n(h(0) \geq x) > 0$ for some $n \in \mathbf{N}$, by (a) the system (E, ν^n) – and thus also (E, ν) – is positive recurrent. □

It is a consequence of (6.3a) and (6.2) that the examples following (1.6) and (3.4) both are even positive recurrent counterexamples; on the other hand the latter one proves neither condition in (6.3) to be necessary for positive recurrence.

To obtain a necessary and sufficient condition ideas from queuing theory will be taken up that can be traced back to [28, 29] and lead to the following notion:

(6.4) Definition Let $(H_n, n \in \mathbf{N})$ be the generating sequence of the system (E, ν). Then the random variables

$$Y_n^x := H_1 \circ \ldots \circ H_n\,(x) \quad \text{for } n \geq 0$$

define the "*dual process*" belonging to $x \in E$.

Clearly, the distributions of X_n^x and Y_n^x agree for fixed n, but $(Y_n^x,\, n \geq 0)$ need not be a Markov chain. Moreover, while the joint distribution of the process $(X_n^x,\, n \geq 0)$ depends only on the kernel P and not on the underlying distribution ν, this fails in general for the process $(Y_n^x,\, n \geq 0)$.

Considered in the compactification E^* of E all dual processes converge to a common limit:

(6.5) Proposition *If the system (E, ν) is irreducible, there is a random variable $Y : \Omega \to E^*$ such that*

(a) $$Y_n^0 \uparrow Y \quad \textit{pointwise},$$

(b) $$Y_n^x \to Y \quad \textit{a.s.} \quad \textit{for all } x \in E\,.$$

Proof. (a) Since the ordered topological space E^* is sequentially compact and the sequence $(Y_n^0(\omega),\, n \geq 0)$ is increasing, it converges in view of (10.4) to a limit $Y(\omega)$ for each ω. Since, moreover, E^* is metrizable, the mapping $Y : \Omega \to E^*$ is measurable.

(b) Whenever $Y(\omega) = \infty$, monotonicity yields $Y_n^x(\omega) \to \infty$ for all x, too. Otherwise choose a countable base of E , which may be assumed to consist of convex sets G_k, $k \in \mathbf{N}$, by (10.4). Since (1.5a) – and thus (1.5b) – carries over from $(X_n^x,\, n \geq 0)$ to $(Y_n^x,\, n \geq 0)$, due to $\mathcal{L}(X_n^x) = \mathcal{L}(Y_n^x)$, this implies

$$1_{G_k}(Y_n^x) - 1_{G_k}(Y_n^0) \to 0 \text{ a.s.} \quad \text{for all } k \in \mathbf{N}\,.$$

Therefore $Y_n^x(\omega) \to Y(\omega)$ for almost all ω satisfying $Y(\omega) \neq \infty$. □

The variable Y obeys a zero-one law characterizing the type of recurrence:

(6.6) PROPOSITION *Let the system (E, ν) be irreducible. Then, with Y being defined by* (6.5),

(a) $\mathbf{P}(Y = \infty) = 0$ *if (E, ν) is positive recurrent,*

(b) $\mathbf{P}(Y = \infty) = 1$ *otherwise.*

PROOF. (a) Let μ be the stationary distribution and choose functions g_l, $l \in \mathbf{N}$, according to (10.7). Identifying g_l with its continuous extension to E^* yields by (6.5b) and the invariance of μ

$$\begin{aligned} \mathbf{E}(g_l(Y)) &= \lim_{n\to\infty} \int_E \mathbf{E}(g_l(Y_n^x))\,\mu(dx) \\ &= \lim_{n\to\infty} \int_E \mathbf{E}(g_l(X_n^x))\,\mu(dx) \\ &= \mu\, g_l \quad \text{for all } \ l \in \mathbf{N}\,. \end{aligned}$$

The assertion follows by the passage $l \to \infty$.

(b1) If (E, ν) is null recurrent and μ is the invariant measure, it follows similarly, now applying Fatou,

$$\begin{aligned} \mathbf{E}(g_l(Y))\mu(E) &\le \liminf_{n\to\infty} \int_E \mathbf{E}(g_l(Y_n^x))\,\mu(dx) \\ &= \liminf_{n\to\infty} \int_E \mathbf{E}(g_l(X_n^x))\,\mu(dx) \\ &= \mu\, g_l < \infty \quad \text{for all } \ l \in \mathbf{N}\,. \end{aligned}$$

In view of $\mu(E) = \infty$ therefore $\mathbf{E}(g_l(Y)) = 0$ for all $l \in \mathbf{N}$, and the assertion follows again for $l \to \infty$.

(b2) If finally (E, ν) is transient, then (6.5a) and (2.6) imply

$$\mathbf{P}(Y \in K) = \lim_{n\to\infty} \mathbf{P}(Y_n^0 \in K) = \lim_{n\to\infty} \mathbf{P}(X_n^0 \in K) = 0 \quad \text{for } \ K \in \mathfrak{K}^{\downarrow}(E)\,,$$

hence for all $K \in \mathfrak{K}(E)$. □

For an application of this criterion consider once more the exchange process from Section 2, denoting again by F the relevant distribution function. The explicit representation

$$Y = \sup_{n\in\mathbf{N}} (U_n - (n-1))$$

shows that

$$\mathbf{P}(Y \le y) = \prod\nolimits_{n\ge 0} F(y+n) > 0$$

if and only if $F(y) > 0$ and the series $\sum_{n\ge 0} (1 - F(y+n))$ converges. Therefore the system (E, ν) is positive recurrent if and only if the variables U_n, $n \in \mathbf{N}$, have a finite expectation (for extensions see [22]).

The following result is related to a "contraction principle" in [27]:

(6.7) THEOREM *If the system (E, ν) is positive recurrent, its stationary distribution is given by $\mathcal{L}(Y)$, with Y being defined by* (6.5).

PROOF. Application of (6.5) and (6.6) to the sequence $(H_n, n > 1)$ yields a variable Y' such that

$$H_2 \circ \ldots \circ H_n\,(0) \uparrow Y' \quad \text{and} \quad \mathbf{P}(Y' = \infty) = 0\,.$$

Therefore, by the continuity of the mappings $H_1(\omega)$, the variables Y and $H_1(Y')$ agree almost surely. Since Y' is independent of H_1 and distributed as Y, this common law is indeed a stationary distribution. □

Finally, the criterion (2.6) for recurrence resp. transience can be completed:

(6.8) THEOREM *Let the system (E, ν) be recurrent. Then for arbitrary initial law the following conditions are equivalent:*

(1) $$(E,\nu) \textit{ is null recurrent},$$

(2) $$\mathbf{P}(X_n \in K) \to 0 \quad \textit{for all } K \in \mathfrak{K}(E)\,.$$

PROOF. 1. In establishing that (1) implies (2) clearly $K \in \mathfrak{K}^{\downarrow}(E)$ may be assumed. By (6.5) this yields

$$\mathbf{P}(X_n \in K) \le \mathbf{P}(X_n^0 \in K) = \mathbf{P}(Y_n^0 \in K) \to \mathbf{P}(Y \in K)$$

with $\mathbf{P}(Y \in K) = 0$ by (6.6).

2. Conversely, let condition (2) be satisfied. Then the estimate

$$\mathbf{P}(X_n \in K) \ge \mathbf{P}(X_0 \le x)\,\mathbf{P}(X_n^x \in K) \quad \text{for } K \in \mathfrak{K}^{\downarrow}(E)\,,$$

with $\mathbf{P}(X_0 \le x) > 0$ for some $x \in E$ by (1.3), yields

$$\mathbf{P}(Y_n^x \in K) = \mathbf{P}(X_n^x \in K) \to 0 \quad \text{for } K \in \mathfrak{K}^{\downarrow}(E)\,,$$

hence for all $K \in \mathfrak{K}(E)$. Again by (6.5) and (6.6), this verifies (1). □

Now the results of (2.6) and (6.8) can be summarized as follows:

(E,ν)	positive recurrent	$\Leftrightarrow$	$\mathbf{P}(X_n^0 \to \infty) = 0$ and $\mathbf{P}(Y_n^0 \to \infty) = 0$,
(E,ν)	null recurrent	$\Leftrightarrow$	$\mathbf{P}(X_n^0 \to \infty) = 0$ and $\mathbf{P}(Y_n^0 \to \infty) = 1$,
(E,ν)	transient	$\Leftrightarrow$	$\mathbf{P}(X_n^0 \to \infty) = 1$ and $\mathbf{P}(Y_n^0 \to \infty) = 1$.

7. Mean passage times.

As stated in (1.2), the expected time to enter an increasing interval is finite in any case. In contrast, decreasing intervals can serve to distinguish positive from null recurrence. This relies on the well-known recurrence theorem of Kac, which is mostly stated under unnecessary restrictions. In the present setting the following dichotomy can be established:

(7.1) THEOREM *Let the system (E, ν) be recurrent with invariant measure μ. Then, with the notations T_B and T_B^x as defined in Section 1,*

(a) $\mathbf{E}(T^x_{[0,x]}) < \infty$, *if (E,ν) is positive recurrent and $\mu([0,x]) > 0$,*

(b) $\mathbf{E}(T^x_{[0,x]}) = \infty$, *if (E,ν) is null recurrent (and $x \in E$ arbitrary).*

PROOF. (a) Consider first the case that x is a maximal element of E. Then by (1.2b)

$$\mathbf{E}(T^x_{[0,x]}) \leq \mathbf{E}(T^x_{\{x\}}) = \mathbf{E}(T^x_{[x,\cdot]}) < \infty .$$

Otherwise choose $y \in [x, \cdot] \setminus \{x\}$ and $m \in \mathbf{N}$ such that $\mathbf{P}(X^0_m \geq y) > 0$. If $(X_n, n \geq 0)$ is the stationary process belonging to (E, ν), then it follows from $\mu([0,x]) > 0$ by the Markov property and monotonicity that

$$\mathbf{P}(X_0 \leq x, X_m \geq y) > 0 .$$

If m is chosen minimal with respect to this inequality, then

$$\mathbf{P}(X_0 \leq x, X_l \leq x, X_m \geq y) = 0 \quad \text{for } 0 < l < m ,$$

because otherwise, due to the stationarity, m could be replaced by $m - l$. Therefore $\mathbf{P}(A) > 0$ for the event

$$A := \{X_0 \leq x, X_l \notin [0,x] \text{ for } 0 < l < m, X_m \geq y\} \subset \{T_{[0,x]} > m\} .$$

With the increasing function

$$g(z) := \mathbf{E}(T^z_{[0,x]}) \quad \text{for } z \in E$$

the recurrence theorem of Kac in its familiar version and the Markov property imply

$$\begin{aligned} \mathbf{P}(T_{[0,x]} < \infty) &= \int_{\{X_0 \leq x\}} T_{[0,x]} \, d\mathbf{P} \\ &\geq \int_A T_{[0,x]} \, d\mathbf{P} \\ &= \int_A (m + g(X_m)) \, d\mathbf{P} \\ &\geq \mathbf{P}(A) \, (m + g(x)) . \end{aligned}$$

Therefore $g(x) < \infty$, as had to be shown.

(b) The extended version of the recurrence theorem of Kac needed here concerns the infinite measure $\varrho := \mu \otimes P \otimes P \otimes \ldots$ on $\bigotimes_{n \geq 0} \mathfrak{B}(E)$, which is shift invariant by the invariance of μ. Now the standard proof in the case of probability measures is easily checked to work as well for ϱ, resulting in

$$\int_B \mathbf{E}(T^y_B) \, \mu(dy) = \int_E \mathbf{P}(T^y_B < \infty) \, \mu(dy) \quad \text{for all } B \in \mathfrak{B}(E) .$$

In applying this equation to $B = [0,x]$ it means no restriction to assume again $\mu([0,x]) > 0$, because otherwise an application of (4.4b) to $f = 1_{[0,x]}$ and the Markov property yield $\mathbf{P}(T^x_{[0,x]} = \infty) > 0$. Under this assumption, now in view of (4.4a), $\mathbf{P}(T^y_{[0,x]} < \infty) = 1$ for all $y \in E$ and thus by monotonicity

$$\begin{aligned} \mathbf{E}(T^x_{[0,x]}) \, \mu([0,x]) &\geq \int_{[0,x]} \mathbf{E}(T^y_{[0,x]}) \, \mu(dy) \\ &= \int_E \mathbf{P}(T^y_{[0,x]} < \infty) \, \mu(dy) \\ &= \mu(E) \end{aligned}$$

with $\mu([0,x]) < \infty = \mu(E)$. □

It is trivial that assertion (b) extends to transient systems, because in this case again $\mathbf{P}(T^x_{[0,x]} = \infty) > 0$.

The final result of this section requires a uniform version of (7.1a), not restricted to decreasing intervals:

(7.2) PROPOSITION *Let the system (E,ν) be positive recurrent with stationary distribution μ. Let moreover $B \in \mathfrak{B}(E)$ be convex with $\mu(B) > 0$ and $y \in E$ be fixed. Then the stopping time*

$$T := \inf\{n \in \mathbf{N} : X^x_n \in B \textit{ for } 0 \le x \le y\}$$

has a finite expectation.

PROOF. Since B is recurrent by (4.4a), an application of (2.3) with $A = \{0, y\}$ yields $m \in \mathbf{N}$ such that

$$\vartheta := \mathbf{P}(X^x_m \in B \text{ for } 0 \le x \le y) > 0\,.$$

Since the system (E, ν') with $\nu' = \nu^m$ meets the assumptions as well and the corresponding stopping time T' satisfies $T \le m\,T'$, the notation can be simplified by supposing $m = 1$. In view of (5.3), moreover, $\mu([0, y]) > 0$ may be assumed in the sequel. Then the recursion

$$S_0 := 0 \quad \text{and} \quad S_{k+1} := \inf\{n > S_k : H_n \circ \ldots \circ H_{S_k+1}(y) \le y\}$$

defines a sequence of stopping times with respect to $(H_n, n \in \mathbf{N})$, which by (4.4a) may be assumed to be finite. By (7.1a) it follows as in the proof of (1.4) that

$$\text{(1)} \qquad \mathbf{E}(S_k - S_{k-1}) = \mathbf{E}(T^y_{[0,y]}) < \infty \quad \text{for } k \in \mathbf{N}\,.$$

Moreover, the events

$$A_k := \{H_{S_k+1}(x) \in B \text{ for } 0 \le x \le y\}$$

by the convention $m = 1$ satisfy

$$\text{(2)} \qquad \mathbf{P}(A_k) = \vartheta \quad \text{for } k \ge 0\,.$$

By construction the variables $1_{A_0}, \ldots, 1_{A_{k-1}}, S_{k+1} - S_k$ are independent for fixed k. Finally, the estimate

$$T \;\le\; \sum\nolimits_{k\ge 0}\; \prod\nolimits_{0 \le i \le k}\, (1 - 1_{A_i})\, (S_{k+1} - S_k) \;+\; 1$$

holds, because for fixed ω the right-hand side equals $S_k(\omega) + 1$, if k is the first index with $\omega \in A_k$, and is infinite, if there is no such index. By cancelling for each k the factor with $i = k$ the bound for T is increased, and the summands are composed of independent factors. By (1) and (2) this yields

$$\mathbf{E}(T) \le \vartheta^{-1}\, \mathbf{E}(T^y_{[0,y]}) + 1 < \infty\,. \qquad \square$$

Now the familiar criterion for positive resp. null recurrence by mean passage times carries over from discrete Markov chain theory to the present setting in the following form:

(7.3) THEOREM *Let the system (E,ν) be recurrent with attractor M.*

(a) *If (E,ν) is positive recurrent and $x \in M$, then*

$$\mathbf{E}(T_G^x) < \infty \quad \textit{for all } G \in \mathfrak{G}(E) \textit{ with } x \in G\,;$$

(b) *if (E,ν) is null recurrent and $x \in E$ arbitrary, then*

$$\mathbf{E}(T_G^x) = \infty \quad \textit{for some } G \in \mathfrak{G}(E) \textit{ with } x \in G\,,$$

provided E is locally bounded (see (10.2)).

PROOF. (a) Since G by (10.4) may be assumed to be convex, this is a special case of (7.2).

(b) By the local boundedness there exist $G_0 \in \mathfrak{G}(E)$ and $y \in E$ such that $x \in G_0 \subset [0,y]$, where $y \neq x$ may be assumed, because otherwise by (7.1b)

$$\mathbf{E}(T_{G_0}^x) \geq \mathbf{E}(T_{[0.x]}^x) = \infty\,.$$

But then $G_0 \cap [y,\,\cdot\,] = \emptyset$ may be assumed as well, because otherwise G_0 can be decreased to $G_0 \setminus [y,\,\cdot\,]$. Now let $m \in \mathbf{N}$ satisfy $\vartheta := \mathbf{P}(X_m^x \geq y) > 0$. If m is chosen minimal, then

$$\mathbf{P}(X_l^x = x, X_m^x \geq y) \leq \mathbf{P}(X_{m-l}^x \geq y) = 0 \quad \text{for } 0 < l < m\,.$$

Therefore outer regularity of $\mathcal{L}(X_l^x)$ provides $G_l \in \mathfrak{G}(E)$ with $x \in G_l$ and

$$\sum\nolimits_{0<l<m} \mathbf{P}(X_l^x \in G_l, X_m^x \geq y) < \vartheta\,.$$

Then $x \in G := \bigcap_{0 \leq l < m} G_l \in \mathfrak{G}(E)$ and $\mathbf{P}(A) > 0$ for the event

$$A := \{X_l^x \notin G \text{ for } 0 < l < m, X_m^x \geq y\}\,.$$

Since $G_0 \cap [y,\,\cdot\,] = \emptyset$ implies $T_G^x > m$ on A, the Markov property and monotonicity, again by (7.1b), yield

$$\mathbf{E}(T_G^x) \geq \mathbf{P}(A)\,(m + \mathbf{E}(T_{[0,y]}^y)) = \infty\,. \qquad \square$$

As in the context of (5.7) it is an open problem, whether local boundedness is essential for assertion (b) (see, however, (9.5)). Moreover, assertion (b) extends again to transient systems, because in this case $\mathbf{P}(T_G^x = \infty) > 0$ whenever G is relatively compact.

8. Further limit theorems. From the results of Section 6 it is easily derived that the distributions of X_n, $n \geq 0$, converge in the positive recurrent case weakly to the stationary distribution (and otherwise vaguely to 0). Actually, the class of functions, for which convergence holds, is considerably larger:

(8.1) PROPOSITION *Let the system (E,ν) be positive recurrent with stationary distribution μ. Then, for arbitrary initial law μ_0 and with $\mu_n := \mu_0 P^n$,*

$$\mu_n f \to \mu f \quad \textit{for all } f \in \mathcal{R}(E)\,.$$

PROOF. Since f is bounded, application of (1.6) with initial variable $X_0 = x_0$ resp. $X_0 = x$ yields

$$\mu_n f - \mu f = \int_E \int_E \mathbf{E}(f(X_n^{x_0}) - f(X_n^x))\, \mu_0(dx_0)\, \mu(dx) \to 0\,. \qquad \square$$

This result implies in particular that

$$\mathbf{P}(X_n = x) \to \mu(\{x\}) \quad \text{for all } \ x \in E\,.$$

Under a total ordering this fact is easily seen to imply uniform convergence of the distribution functions $F_n(y) := \mathbf{P}(X_n \le y)$ to the limit $F(y) := \mu([0,y])$. To extend this result to the general setting, an appropriate metric on $\mathsf{M}_1(E)$ is needed:

(8.2) PROPOSITION *The definition*

$$d(\mu_1, \mu_2) := \sup\{|\mu_1(B) - \mu_2(B)| : B \in \mathfrak{B}^{\downarrow}(E)\}$$

yields a metric on $\mathsf{M}_1(E)$ *with the following properties:*

(a) $$d(\mu_1,\mu_2) \ = \ \sup\{|\mu_1 f - \mu_2 f| : f \in \mathcal{B}^{\downarrow}(E) \ \ \textit{with} \ \ 0 \le f \le 1\}\,,$$

(b) $$d(\mu_n, \mu) \to 0 \quad \textit{implies} \quad \mu_n \underset{w}{\to} \mu\,,$$

(c) *the mapping* $\mu \mapsto \mu P$ *is a contraction.*

PROOF. (a) This follows from $\mu_i f = \int_{[0,1]} \mu_i(f(x) > y)\, dy$.

(b) By (a)

$$\mu_n f \to \mu f \quad \text{ for } \ f \in \mathcal{B}^{\downarrow}(E) \ \text{ with } \ 0 \le f \le 1\,,$$

which implies convergence for $f \in \mathcal{K}_0$ and thus for $f \in \mathcal{K}(E)$ according to (10.6a). Besides $\mu_n \underset{w}{\to} \mu$ this proves d to be indeed a metric.

(c) By (a) this follows from $\mu P f = \int_{\mathcal{H}[E]} \mu(f \circ h)\, \nu(dh)$. $\square$

An application of (10.3) to E^* yields $d(\varepsilon_x, \varepsilon_y) = 1$ for $x \ne y$, showing metric convergence in general to be much stronger than weak convergence. Using the notion of a "splitting point" (see z in the following proof), as introduced by Dubins/Freedman [13] for $E = [0,1]$ and extended by Bhattacharya/Majumdar [8, 9] to higher dimension, the weak convergence from (8.1) can be strengthened to metric convergence:

(8.3) THEOREM *If the system* (E, ν) *is positive recurrent with stationary distribution* μ*, then*

$$d(\mu_0 P^n, \mu) \to 0 \quad \textit{for all} \ \ \mu_0 \in \mathsf{M}_1(E)\,.$$

PROOF. 1. Applying (1.3), choose z with $\mu([0,z]) > 0$ and, applying (5.3), choose $\delta > 0$ such that

$$\mu([0,z]) \wedge \mu([z,\cdot\,]) > \delta\,.$$

Then, for given $\varepsilon > 0$ and with $\mu_n := \mu_0 P^n$, there exists a finite subset A of E satisfying

$$\mu(E \setminus A^{\downarrow}) \le \delta\,\varepsilon \quad \text{ and } \quad \mu_n(E \setminus A^{\downarrow}) \le \delta\,\varepsilon \quad \text{for all } \ \ n \in \mathbf{N}\,,$$

as follows from (1.3) and (8.1), applied to $f = 1_{A^\downarrow}$. Moreover, again by (8.1), there exists $k \in \mathbf{N}$ such that

$$\mathbf{P}(X_k^x \le z \text{ for all } x \in A^\downarrow) \ge \delta \quad \text{and} \quad \mathbf{P}(X_k^0 \ge z) \ge \delta\,,$$

where the first inequality uses the fact that by (1.5a)

$$\mathbf{P}\Big(\{X_n^0 \le z\} \setminus \big(\bigcap\nolimits_{x \in A^\downarrow} \{X_n^x \le z\}\big)\Big) \le \sum\nolimits_{x \in A} \left(\mathbf{P}(X_n^0 \le z) - \mathbf{P}(X_n^x \le z)\right) \to 0\,.$$

Since the values $d(\mu_n, \mu) = d(\mu_{n-1}P, \mu P)$ by (8.2c) form a decreasing sequence, it is sufficient to prove $d(\mu_{kn}, \mu) \to 0$. Therefore, passing from ν to ν^k, the notation can be simplified by assuming $k = 1$ in the sequel.

2. For arbitrary $B \in \mathfrak{B}^\downarrow(E)$ there are now two possibilities:

(1) In the case $z \in B$, i.e. $[0, z] \subset B$, consider

$$\mathcal{H}_* := \{h \in \mathcal{H}[E] : h[A^\downarrow] \subset [0, z]\}\,,$$

satisfying $A^\downarrow \subset h^{-1}[B] \in \mathfrak{B}^\downarrow(E)$ for $h \in \mathcal{H}_*$. Therefore

$$\begin{aligned} |\mu_n(B) - \mu(B)| &\le \int_{\mathcal{H}[E]} |\mu_{n-1}(h^{-1}[B]) - \mu(h^{-1}[B])|\, \nu(dh) \\ &\le (1-\delta)\, d(\mu_{n-1}, \mu) + \delta\, \varepsilon \quad \text{for all } n \in \mathbf{N}\,, \end{aligned}$$

because by the first part $\nu(\mathcal{H}[E] \setminus \mathcal{H}_*) \le 1 - \delta$ and on $\mathcal{H}_*$ the integrand is bounded by $\delta\, \varepsilon$.

(2) In the case $z \notin B$, i.e. $[z, \cdot\,] \subset E \setminus B$, replace $\mathcal{H}_*$ by

$$\mathcal{H}^* := \{h \in \mathcal{H}[E] : h(0) \ge z\}\,,$$

satisfying $h^{-1}[B] = \emptyset$ for $h \in \mathcal{H}^*$. Then it follows similarly, in fact somewhat simpler, that

$$|\mu_n(B) - \mu(B)| \le (1-\delta)\, d(\mu_{n-1}, \mu) \quad \text{for all } n \in \mathbf{N}\,.$$

Combining both cases and putting $\gamma := 1 - \delta$ this yields

$$d(\mu_n, \mu) \le \gamma\, d(\mu_{n-1}, \mu) + \delta\, \varepsilon$$

and thus by recursion

$$d(\mu_n, \mu) < \gamma^n\, d(\mu_0, \mu) + \varepsilon \quad \text{for all } n \in \mathbf{N}\,.$$

In view of $\gamma < 1$ and $d \le 1$ this proves the assertion. □

In case the state space is bounded (or at least of the form $E = A^\downarrow$ with A finite), the proof is easily checked to work with $\varepsilon = 0$, i.e. there is geometric convergence, which in addition is uniform in μ_0.

To conclude this section by a law of large numbers, ergodicity of the stationary version of the process $(X_n,\, n \ge 0)$ is essential. Actually, by means of the dual process from Section 6, a stronger result can be established:

(8.4) PROPOSITION *Let the system (E, ν) be positive recurrent with stationary distribution μ. Then the process $(X_n,\, n \geq 0)$ with initial law μ is mixing.*

PROOF. Extending H_n, $n \in \mathbf{N}$, let $H_n, n \in \mathbf{Z}$, be independent random variables with distribution ν. Then by (6.5) and (6.6)

$$X'_n := \lim_{n \geq m \to -\infty} H_n \circ \ldots \circ H_m\,(0) \in E \;\text{ a.s.}\,.$$

The continuity of the mappings $H_n(\omega)$ yields

$$X'_n = H_n(X'_{n-1}) \;\text{ a.s.} \quad \text{for } \; n \in \mathbf{N}\,.$$

Since X'_0 is independent of $(H_n, n \in \mathbf{N})$ and $\mathcal{L}(X'_0) = \mu$ by (6.7), the processes $(X_n,\, n \geq 0)$ and $(X'_n,\, n \geq 0)$ have the same distribution, and thus it is sufficient to prove the assertion for $(X'_n,\, n \geq 0)$. To this end denote by σ and σ' the shift in $W := \prod_{n \in \mathbf{Z}} \mathcal{H}[E]$ and $W' := \prod_{n \geq 0} E^*$, respectively, and consider the mapping

$$\tau : (h_n, n \in \mathbf{Z}) \mapsto \Big(\lim_{n \geq m \to -\infty} h_n \circ \ldots \circ h_m\,(0),\, n \geq 0 \Big)$$

from W to W', which is easily checked to be measurable, due to the topological properties of E^*. Then the mappings σ and σ' are conjugate under τ, i.e. satisfy $\tau \circ \sigma = \sigma' \circ \tau$. Therefore the mixing property of σ with respect to the product measure $\bigotimes_{n \in \mathbf{Z}} \nu$ carries over to σ' with respect to its image by τ. Since this obviously is the distribution of $(X'_n,\, n \geq 0)$, the assertion follows. □

To be complete, it has to be mentioned that in general the tail $\sigma-$field of $(X_n,\, n \geq 0)$, even under stationarity, need not be trivial. A counterexample is provided by the Cantor system, where X_{n-1} can be reconstructed from X_n with probability 1, and thus the tail $\sigma-$field of $(X_n,\, n \geq 0)$ coincides with the full $\sigma-$field generated by the process up to sets of probability 0.

Now a fairly general law of large numbers can be derived:

(8.5) THEOREM *Let the system (E, ν) be positive recurrent with stationary distribution μ. Then for arbitrary initial law*

$$\frac{1}{n} \sum\nolimits_{0 \leq m < n} f(X_m) \to \mu f \;\; a.s.\,,$$

whenever $f \in \mathcal{R}(E)$ or $f \in \mathcal{C}(E)$.

PROOF. If X_0 is replaced by X'_0 with $\mathcal{L}(X'_0) = \mu$, the resulting copy $(X'_n,\, n \geq 0)$ of $(X_n,\, n \geq 0)$ is ergodic by (8.4), hence satisfies

$$\frac{1}{n} \sum\nolimits_{0 \leq m < n} f(X'_m) \to \mu f \;\text{ a.s.} \quad \text{for all bounded } \; f \in \mathfrak{B}(E)\,.$$

(1) In the case $f \in \mathcal{R}(E)$ this convergence carries over to the process $(X_n,\, n \geq 0)$, because (1.6) implies $f(X_n) - f(X'_n) \to 0$ almost surely.

(2) In the case $f \in \mathcal{C}(E)$ assume without restriction $0 \leq f \leq 1$ and approximate f from below by functions from $\mathcal{K}(E)$. Since these functions by (10.6b) belong to $\mathcal{R}(E)$, they satisfy the assertion and thus

$$\liminf_{n \to \infty} \frac{1}{n} \sum\nolimits_{0 \leq m < n} f(X_m) \geq \mu f \;\text{ a.s.}\,.$$

Repeating the argument for $1 - f$ proves the assertion. □

A further extension of (8.5) from continuous to semicontinuous functions fails in general, as can be seen, for instance, by the Cantor system (E,ν). Here, for fixed $x \in E$, there exists a countable subset B of E such that $\mathbf{P}(X_n^x \in B) = 1$ for all $n \geq 0$. Since the stationary distribution μ is nonatomic, this yields a set $F \in \mathfrak{F}(E)$ with $B \cap F = \emptyset$ and $\mu(F) > 0$. Then the initial law ε_x and the function $f = 1_F$ are suited for a counterexample.

9. Strictly order-preserving systems. Throughout Sections 1 – 8 the system (E,ν) has always been supposed to be irreducible. To see the crucial role of this assumption consider the example following (3.4). If in this case the state space is enlarged from $[0,1[$ to $[0,1]$ (extending the mappings h_i continuously), the definition $\mu = \varepsilon_1$ yields another invariant measure. This observation suggests a restriction of the original state space:

(9.1) DEFINITION For an arbitrary system (E,ν) the "*reduced state space*" is given by the subspace

$$\widehat{E} := \{x \in E : \mathbf{P}(X_n^0 \geq x) > 0 \text{ for some } n \in \mathbf{N}\}.$$

If E is totally ordered, $\widehat{E}$ is obviously an open or closed decreasing subset of E and thus again admissible in the sense of Section 0. Moreover, it makes sense to restrict the mappings h to $\widehat{E}$, because

$$(*) \qquad h[\widehat{E}] \subset \widehat{E} \quad \text{for } \nu\text{-almost all } h \in \mathcal{H}[E].$$

Indeed, fix $x_0 \in \widehat{E}$ with $\mathbf{P}(X_n^0 \geq x_0) > 0$ and consider $h_0 \in \mathcal{H}[E]$ with $h_0(x_0) \notin \widehat{E}$. Then necessarily $\nu(h(x_0) \geq h_0(x_0)) = 0$, because otherwise by independence

$$\mathbf{P}(X_{n+1}^0 \geq h_0(x_0)) \geq \mathbf{P}(X_n^0 \geq x_0)\,\mathbf{P}(H_{n+1}(x_0) \geq h_0(x_0)) > 0.$$

Therefore, with the notation

$$\mathcal{H}(y) := \{h \in \mathcal{H}[E] : h(x_0) \geq y\} \qquad \text{for } y \in E,$$

each $h_0 \in \mathcal{H}[E]$ with $h_0(x_0) \notin \widehat{E}$ is contained in the union of all sets $\mathcal{H}(y)$ with $\nu(\mathcal{H}(y)) = 0$. Since the sets $\mathcal{H}(y)$ decrease for increasing y, this union can be replaced by a countable one, due to (E1), and thus is a ν-null set itself. Therefore $h_0(x_0) \in \widehat{E}$ for ν-almost all $h_0 \in \mathcal{H}[E]$ and, replacing x_0 by a sequence $(x_k,\, k \in \mathbf{N})$ with $[0,x_k] \uparrow \widehat{E}$, this settles $(*)$.

As this consideration shows, to treat only irreducible systems means no real restriction under a total ordering. This, however, does not hold for a general state space, as the following examples show. Choose $E = \mathbf{R}_+^2$ and let ν be supported by constant mappings (resulting in independent variables X_n, $n \geq 0$), taking their values in the totally disordered subset $D := \{x \in E : x_1 + x_2 = 1\}$ only. If ν assigns positive mass to each constant in a dense subset of D, the reduced state space $\widehat{E}$ is easily seen to be no longer locally compact. If on the other hand ν has no point masses at all, obviously

$$h[\widehat{E}] \cap \widehat{E} = \emptyset \quad \text{for } \nu\text{-almost all } h \in \mathcal{H}[E].$$

This observation is the motivation to investigate systems (E,ν) that are not necessarily irreducible but strictly order-preserving. Before introducing this notion the necessary measurability has to be settled:

(9.2) PROPOSITION *The subspace $\mathcal{J}[E]$ consisting of all mappings $h \in \mathcal{H}[E]$ such that*

$$h(x_1) < h(x_2) \quad \text{whenever} \quad x_1 < x_2$$

is of type G_δ (i.e. a Polish space).

PROOF. Denote by S the open set of all pairs (x_1, x_2) with $x_1 < x_2$. Since E is locally compact and second countable, it follows that

$$S = \bigcup_{l \in \mathbf{N}} (K_1^l \times K_2^l) \quad \text{with} \quad K_i^l \in \mathfrak{K}(E)\,.$$

Therefore it is sufficient to prove

$$\mathcal{H}_l := \{h \in \mathcal{H}[E] : h[K_1^l] \times h[K_2^l] \subset S\}$$

to be open for all $l \in \mathbf{N}$. In view of $h[K_i^l] \in \mathfrak{K}(E)$ a theorem of Wallace (see e.g. [16]) applies, i.e. $h \in \mathcal{H}_l$ is equivalent to the existence of sets $G_i \in \mathfrak{G}(E)$ satisfying $h[K_i^l] \subset G_i$ and $G_1 \times G_2 \subset S$. Since the sets $\{h \in \mathcal{H}[E] : h[K_i^l] \subset G_i\}$ are open, the assertion is established, □

Now the central assumption for this section can be made precise:

(9.3) DEFINITION The system (E,ν) is called "*strictly order-preserving*", if the following two conditions are satisfied:

(a) $$\nu(\mathcal{J}[E]) = 1\,,$$

(b) $$\nu^n(h(0) > 0) > 0 \quad \text{for some} \quad n \in \mathbf{N}\,.$$

Clearly, condition (a) is of relevance only in conjunction with condition (b).

To discuss briefly an important special case, consider generalized autoregressive models on $E = \mathbf{R}_+^d$, where ν is supported by affine maps $h : x \mapsto Ax + b$. If A and b are composed of the (nonnegative) variables a_{ik} and b_i, respectively, then conditions (a) and (b) are satisfied as soon as

(a′) $$\mathbf{P}(a_{i1} + \ldots + a_{id} > 0) = 1 \quad \text{for} \quad 1 \le i \le d\,,$$

(b′) $$\mathbf{P}(b_1 \ldots b_d > 0) > 0\,.$$

This example should be compared with the model in [4]. While the state space there is enlarged to $E = \mathbf{R}^d$, only mappings $h : x \mapsto ax + b$ with strictly positive scalar factors a are admitted, imposing in addition strong moment conditions on the variables a and b.

The notions in (9.1) and (9.3) are related by the following facts:

(9.4) LEMMA *If the system (E,ν) is strictly order-preserving, then any $x \in \widehat{E}$ satisfies*

(a) $$\mathbf{P}(X_n^0 > x) > 0 \quad \textit{for some} \quad n \in \mathbf{N}\,,$$

(b) $$x < y \quad \textit{for some} \;\; y \in \widehat{E}\,,$$

(c) $$h(x) \in \widehat{E} \quad \textit{for all} \;\; h \in \mathcal{N} \cap \mathcal{J}[E]\,.$$

PROOF. Since $\mathcal{J}[E]$ is stable under composition, $\nu(\mathcal{J}[E]) = 1$ implies $\nu^k(\mathcal{J}[E]) = 1$ for all $k \in \mathbf{N}$. In the sequel denote the support of ν^k by $\mathcal{N}^k$ and its elements by h_k.

(a) By the assumptions on x and (E, ν) there are $l \in \mathbf{N}$ and $m \in \mathbf{N}$ such that

$$\nu^l(h_l(0) \geq x) > 0 \quad \text{and} \quad \nu^m(h_m(0) > 0) > 0\,.$$

With $n = l + m$ and

$$\mathcal{J}_a := \{(h_l, h_m) : h_l \circ h_m\,(0) > x\}$$

this implies

$$\nu^n(h_n(0) > x) = \nu^l \otimes \nu^m\,(\mathcal{J}_a) \geq \nu^l \otimes \nu^m\,(h_l(0) \geq x,\, h_m(0) > 0) \;>\; 0\,.$$

(b) By (a) there are $h_m^0 \in \mathcal{N}^m$ with $h_m^0(0) > 0$ and $h_n^0 \in \mathcal{N}^n \cap \mathcal{J}[E]$ with $y := h_n^0(0) > x$. Then $h_n^0 \circ h_m^0\,(0) > y$ and thus

$$\mathcal{J}_b := \{(h_n, h_m) : h_n \circ h_m\,(0) > y\}$$

defines an open subset of $\mathcal{H}[E] \times \mathcal{H}[E]$ intersecting the support of $\nu^n \otimes \nu^m$. Therefore

$$\nu^{n+m}(h_{n+m}(0) > y) = \nu^n \otimes \nu^m\,(\mathcal{J}_b) > 0\,,$$

hence in particular $y \in \widehat{E}$.

(c) Choose $h_1^0 = h$ and h_n^0 as in (b). Then $h_1^0 \circ h_n^0\,(0) > h(x)$, and it follows as above, considering now

$$\mathcal{J}_c := \{(h_1, h_n) : h_1 \circ h_n\,(0) > h(x)\}\,,$$

that $h(x) \in \widehat{E}$. □

The crucial properties of the reduced state space follow readily:

(9.5) PROPOSITION *If the system (E, ν) is strictly order-preserving, then*

(a) $$\widehat{E} \;\; \textit{is locally compact}\,,$$

(b) $$h[\widehat{E}] \subset \widehat{E} \quad \textit{for } \nu\textit{-almost all} \;\; h \in \mathcal{H}[E]\,,$$

(c) $$\widehat{E} \;\; \textit{is locally bounded}\,.$$

PROOF. (a) $\widehat{E}$ is a decreasing subset of E, hence by (9.4b) open in E and thus again locally compact.

(b) Since $\widehat{E}$ is second countable and by (9.4b) covered by the family of open sets $\{x \in \widehat{E} : x < y\}$, $y \in \widehat{E}$, there are $y_k \in \widehat{E}$ such that $\widehat{E} = \bigcup_{k \in \mathbf{N}} [0, y_k]$. Therefore $h[\widehat{E}] \subset \widehat{E}$ if and only if $h(y_k) \in \widehat{E}$ for all $k \in \mathbf{N}$, as holds indeed by (9.4c) for all $h \in \mathcal{N} \cap \mathcal{J}[E]$.

(c) Local boundedness is immediate from (9.4b). □

Since the assumptions (E2) and (E3) clearly carry over from E to $\widehat{E}$, too, the reduced state space is again admissible in the sense of Section 0. Therefore the reduction of a strictly order-preserving system (E,ν) can be summarized as follows: Disregarding the $\nu-$null set of mappings with $h[\widehat{E}] \not\subset \widehat{E}$, let $\widehat{\nu}$ be the image of ν under the (continuous) mapping that assigns to $h \in \mathcal{H}[E]$ its restriction $\widehat{h} \in \mathcal{H}[\widehat{E}]$. Then the system $(\widehat{E},\widehat{\nu})$ is irreducible, and an associated process $(\widehat{X}_n,\ n \geq 0)$ behaves as $(X_n,\ n \geq 0)$, whenever the initial law is supported by $\widehat{E}$. By (9.5c), moreover, the results (5.7) and (7.3) on attractor and mean passage time simplify for the reduced system.

10. Order and topology. A set is called an *ordered topological space* (OTS), if its topology and (partial) order are compatible, i.e.

$$R := \{(x_1,x_2) : x_1 \leq x_2\} \ \in \mathfrak{F}(E \times E)\,.$$

By symmetry this holds as well for the inverse ordering, hence the diagonal is a closed subset of $E \times E$, and thus each OTS is a Hausdorff space. Moreover, each subspace with the induced ordering and each product space with the product ordering yield again an OTS.

The simplest example of an OTS is provided by a totally ordered set E, the topology being generated by the "open intervals" $E \setminus [x,\,\cdot\,]$ and $E \setminus [\,\cdot\,,x]$. While the meaning of a strict inequality "$x_1 < x_2$" is clear in this case, more care is necessary in the general case:

(10.1) Definition Let E be an arbitrary OTS. Then "$x_1 < x_2$" means existence of disjoint neighborhoods G_i of x_i such that

(a) $$G_1 \in \mathfrak{G}^{\downarrow}(E) \quad \text{and} \quad G_2 \in \mathfrak{G}^{\uparrow}(E)\,,$$

(b) $$y_1 \leq y_2 \quad \text{for} \quad y_i \in G_i\,.$$

This implies in particular

$$S := \{(x_1,x_2) : x_1 < x_2\} \ \in \mathfrak{G}(E \times E)$$

and the transitivity law

$$x_1 < x_3 \quad \text{whenever} \quad x_1 < x_2 \leq x_3 \quad \text{or} \quad x_1 \leq x_2 < x_3\,.$$

Another notion combining order and topology appears in Sections 5 and 7:

(10.2) Definition Let E be an arbitrary OTS. Then E is "*locally bounded*", if any $x \in E$ has a bounded neighborhood.

While this condition is clearly satisfied under a total ordering, it may well be violated in the general case: consider, for instance, the subspace

$$E = \{(x_1,x_2) \in \mathbf{R}_+^2 : x_1 + x_2 \leq 1\}\,.$$

The deepest result on order and topology used here is the analogue of Tietze's extension theorem, due to Nachbin:

(10.3) PROPOSITION *Let E be a compact OTS and E_0 be a closed subspace. Then any function $f_0 \in \mathcal{C}^{\uparrow}(E_0)$ can be extended to a function $f \in \mathcal{C}^{\uparrow}(E)$.*

PROOF. See [33, Corollary 3.4 and Theorem 3.6]. □

The following fact is an immediate consequence of (10.3) and therefore supplied with its simple proof:

(10.4) PROPOSITION *Let E be (a subspace of) a compact OTS. Then the class of convex open sets is a base of the topology.*

PROOF. Assume E to be compact and apply (10.3) to $E_0 = \{x_1, x_2\}$ with $x_1 \neq x_2$ to see that $\mathcal{C}^{\uparrow}(E)$ separates the points of E. Therefore $\mathcal{C}^{\uparrow}(E)$ induces a Hausdorff topology in E coarser than the underlying compact topology, and thus both topologies agree. Since sets $\{x \in E : a < f(x) < b\}$ are convex for $f \in \mathcal{C}^{\uparrow}(E)$ and $a, b \in \mathbf{R}$ and convexity is stable under intersection, the assertion follows. □

The next result is crucial for the introduction of admissible state spaces E in Section 0:

(10.5) PROPOSITION *Let E be a locally compact OTS and $E^* = E \cup \{\infty\}$ its Alexandrov compactification. Then, defining $x \leq \infty$ for all $x \in E$ makes E^* again an OTS if and only if*

$$(*) \qquad K^{\downarrow} \in \mathfrak{K}(E) \quad \textit{for all } \ K \in \mathfrak{K}(E)\,.$$

PROOF. 1. Let first E^* be an OTS and consider a set $K \in \mathfrak{K}(E) \subset \mathfrak{K}(E^*)$. Then denoting by R^* the order graph of E^* and by p_1 the projection of $E^* \times E^*$ onto its first factor leads to

$$K^{\downarrow} = p_1[\,(E^* \times K) \cap R^*\,] \in \mathfrak{K}(E^*)\,,$$

where $\infty \notin K^{\downarrow}$ and thus indeed $K^{\downarrow} \in \mathfrak{K}(E)$.

2. Let conversely condition $(*)$ be satisfied and R^* be defined as above. Then, to prove $(E^* \times E^*) \setminus R^*$ to be open, choose any (x_1, x_2) in this set.

(1) If $\{x_1, x_2\} \subset E$, then there are $G_i \in \mathfrak{G}(E) \subset \mathfrak{G}(E^*)$ such that

$$x_i \in G_i \quad \text{and} \quad G_1 \times G_2 \subset (E \times E) \setminus R \subset (E^* \times E^*) \setminus R^*\,.$$

(2) If $\{x_1, x_2\} \not\subset E$, then $x_1 = \infty$ and $x_2 \in E$. Since E is locally compact, there is a neighborhood $G_2 \in \mathfrak{G}(E) \subset \mathfrak{G}(E^*)$ of x_2 with closure $K \in \mathfrak{K}(E)$. By condition $(*)$, therefore, $K^{\downarrow} \in \mathfrak{K}(E)$ as well, and thus $G_1 := E^* \setminus K^{\downarrow} \in \mathfrak{G}(E^*)$ is a neighborhood of x_1. Then it is easily checked that $(G_1 \times G_2) \cap R^* = \emptyset$. □

Now an appropriate version of Stone's theorem can be established:

(10.6) PROPOSITION *Let E be a locally compact OTS satisfying condition $(*)$ of (10.5). Then*

$$\text{(a)} \qquad \mathcal{K}_0 := \{f_1 - f_2 : 0 \leq f_i \in \mathcal{K}^{\downarrow}(E)\}$$

is a dense subspace of $\mathcal{K}(E)$ with respect to the uniform norm,

$$\text{(b)} \qquad \mathcal{K}(E) \subset \mathcal{R}(E)\,.$$

Proof. Since E^* according to (10.5) is a compact OTS, it follows as in the proof of (10.4) that

$$\mathcal{C}_0^* := \{f_1^* - f_2^* : 0 \le f_i^* \in \mathcal{C}^\downarrow(E^*)\}$$

separates the points of E^*. All further conditions in Stone's theorem are obviously satisfied, and thus $\mathcal{C}_0^*$ is dense in $\mathcal{C}(E^*)$. Let now $f \in \mathcal{K}(E)$ be given and f^* be its trivial extension to E^*. Then there are nonnegative functions $f_{ik}^* \in \mathcal{C}^\downarrow(E^*)$ with

$$\|f_k^* - f^*\| \to 0 \quad \text{for} \quad f_k^* := f_{1k}^* - f_{2k}^*,$$

where by adding suitable constants $f_{ik}^*(\infty) = 0$ can be achieved. The restriction of $(f_{ik}^* - 1/k)^+$ to E yields nonnegative functions $f_{ik} \in \mathcal{C}(E)$ such that

$$\|f_k - f\| \to 0 \quad \text{for} \quad f_k := f_{1k} - f_{2k}.$$

Since the support of f_{ik} is a subset of $\{f_{ik}^* \ge 1/k\}$ and this set is compact in view of $f_{ik}^*(\infty) = 0$, the functions f_{ik} are in fact contained in $\mathcal{K}(E)$. This proves (a) and, due to $\mathcal{K}_0 \subset \mathcal{V}(E)$, also (b). □

The next result requires some countability:

(10.7) Proposition *Let E be a locally compact and second countable OTS satisfying condition $(*)$ of (10.5). Then there are functions $g_l \in \mathcal{K}^\downarrow(E)$ such that*

(1) $$0 \le g_1 \le g_2 \le \ldots \to 1,$$

(2) *each $K \in \mathcal{K}(E)$ is included in $\{g_l = 1\}$ for some l.*

Proof. By the topological properties of E there is a sequence of sets $K_l \in \mathfrak{K}^\downarrow(E)$ such that each $K \in \mathfrak{K}(E)$ is included in some K_l. Since E^* according to (10.5) is a compact OTS, (10.3) applies and yields functions $f_l^* \in \mathcal{C}^\downarrow(E^*)$ with $f_l^*|K_l^\downarrow = 1$ and $f_l^*(\infty) = -1$. If f_l denotes the restriction of f_l^* to E, the functions

$$g_l := (f_1^+ \vee \ldots \vee f_l^+) \wedge 1 \quad \text{for } l \in \mathbf{N}$$

meet all requirements. □

The final result concerns the classes $\mathcal{V}(E)$ and $\mathfrak{V}(E)$ introduced in Section 0. Clearly, $\mathcal{V}(E)$ includes the linear space of all differences $f_1 - f_2$ with bounded functions $f_i \in \mathcal{B}^\uparrow(E)$, and thus $\mathfrak{V}(E)$ includes the algebra of all finite unions of sets $B_1 \setminus B_2$ with $B_i \in \mathfrak{B}^\uparrow(E)$. Conversely:

(10.8) Proposition *Let E be a locally compact and second countable OTS with lower bound 0. If $\mathcal{U}(E)$ and $\mathfrak{U}(E)$ denote the class of universally measurable functions on and subsets of E, respectively, then*

(a) *each $f \in \mathcal{V}(E)$ is a difference $f_1 - f_2$ with bounded functions $0 \le f_i \in \mathcal{U}^\uparrow(E)$,*

(b) *each $B \in \mathfrak{V}(E)$ is a finite union of sets $B_1 \setminus B_2$ with $B_i \in \mathfrak{U}^\uparrow(E)$.*

Proof. (a) Assume without restriction $f(0) = 0$. As under a total ordering f can then be represented as $f = f^+ - f^-$ with bounded increasing functions $f^\sigma \ge 0$, $\sigma \in \{-,+\}$, defined by

$$f^\sigma(x) := \sup\left\{ \sum_{k \in \mathbf{N}} (f(x_{k+1}) - f(x_k))^\sigma : x_1 \le x_2 \le \ldots \le x \right\}.$$

To verify $f^\sigma \in \mathcal{U}(E)$, consider the Borel measurable function

$$g^\sigma(x_1, x_2, \ldots; x) := 1_{\{x_1 \leq x_2 \leq \ldots \leq x\}} \sum_{k \in \mathbf{N}} \left(f(x_{k+1}) - f(x_k)\right)^\sigma$$

on the Suslin space $E^{\mathbf{N}} \times E$. Then the set of $x \in E$ satisfying an inequality $f^\sigma(x) > \gamma$ equals the projection of the set of $(x_1, x_2, \ldots; x) \in E^{\mathbf{N}} \times E$ satisfying $g^\sigma(x_1, x_2, \ldots; x) > \gamma$ onto the second factor and is therefore a Suslin, hence universally measurable, set for all $\gamma \geq 0$.

(b) In the special case $f = 1_B$ the functions f^σ are integer-valued. This yields the representation

$$B = \bigcup_{k \in \mathbf{N}} \left(\{f^+ \geq k\} \setminus \{f^- \geq k\}\right),$$

where the right-hand side is in fact a finite union. □

It has to be mentioned that no null sets intervene under a total ordering, because in this case increasing functions are Borel measurable.

REFERENCES

[1] Alpuim, M., Athayde, E. (1990). On the stationary distribution of some extremal Markovian sequences. J. Appl. Prob. **27**, 291–302.

[2] Arnold, L. (1998). Random dynamical systems. Springer, New York.

[3] Asmussen, S. (1987). Applied probability and queues. Wiley, New York.

[4] Babillot, M., Bougerol, P., Elie, L. (1997). The random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. Ann. Probab. **25**, 478–493.

[5] Barnsley, M. (1993). Fractals everywhere, 2nd edition. Academic Press, London.

[6] Barnsley, M., Elton, J. (1988). A new class of Markov processes for image encoding. Adv. Appl. Prob. **20**, 14–32.

[7] Benda, M. (1998). Schwach kontraktive stochastische dynamische Systeme. PhD thesis, University of Munich.

[8] Bhattacharya, R., Majumdar, M. (1999). On a theorem of Dubins and Freedman. J. Theor. Prob. 1067–1087.

[9] Bhattacharya, R., Majumdar, M. (1999). Convergence to equilibrium of random dynamical systems generated by i. i. d. monotone maps, with applications to economics. Statist. Textbooks Monogr. **158**, 713–741. Dekker, New York.

[10] Bougerol, P., Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. J. Econ. **52**, 115–127.

[11] Brandt, A., Franken, P., Lisek, B. (1990). Stationary stochastic models. Wiley, New York.

[12] Diaconis, P., Freedman, D. (1999). Iterated random functions. SIAM Rev. **41**, 45–76.

[13] Dubins, L., Freedman, D. (1966). Invariant probabilities for certain Markov processes. Ann. Math. Stat. **37**, 837–848.

[14] Elton, J. (1987). An ergodic theorem for iterated maps. Ergodic Theory Dyn. Syst. **7**, 481–488.

[15] Embrechts, P., Goldie, C. (1994). Perpetuities and random equations. Asymptotic statistics (Prague 1993), 75–86. Physica, Heidelberg.

[16] Engelking, R. (1989). General topology, 2nd edition. Polish Scientific Publishers, Warzawa.

[17] Foguel, S. (1973). The ergodic theory of positive operators on continuous functions. Ann. Sc. Norm. Super. Pisa **27**, 19–51.

[18] Glasserman, P., Yao, D. (1995). Stochastic vector difference equations with stationary coefficients. J. Appl. Prob. **32**, 851–866.

[19] Goldie, C. (1991). Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. **1**, 126–166.

[20] Goldie, C., Maller, R. (2000). Stability of perpetuities. Ann. Probab. **28**, 1195–1218.

[21] Helland, I., Nilsen, T. (1976). On a general random exchange model. J. Appl. Prob. **13**, 781–790.

[22] Högnäs, G., Muhkerjea, A. (1995). Probability measures on semigroups. Plenum Press, London.

[23] Jarner, S., Tweedie, R. (2001). Locally contracting iterated functions and stability of Markov chains. J. Appl. Prob. **38**, 494–507.

[24] Kellerer, H. (1992). Ergodic behaviour of affine recursions I, II, III. Preprints, University of Munich (http://www.mathematik.uni-muenchen.de/~kellerer).

[25] Kellerer, H. (1995). Order-preserving random dynamical systems. Preprint, University of Munich (http://www.mathematik.uni-muenchen.de/~kellerer).

[26] Kifer, Y. (1986). Ergodic theory of random transformations. Birkhäuser, Basel.

[27] Letac, G. (1986). A contraction principle for certain Markov chains and its applications. Contemp. Math. **50**, 263–273.

[28] Lindley, D. (1952). The theory of a queue with a single server. Proc. Camb. Philos. Soc. **48**, 277–289.

[29] Loynes, R. (1962). The stability of a queue with non-independent inter-arrival and service times. Proc. Camb. Philos. Soc. **58**, 497–520.

[30] Lund, B., Meyn, P., Tweedie, R. (1996). Computable exponential convergence rates for stochastically ordered Markov processes. Ann. Appl. Probab. **6**, 218–237.

[31] Mairesse, J. (1997). Products of irreducible random matrices in the (max,+) algebra. Adv. Appl. Prob. **29**, 444–477.

[32] Meyn, P., Tweedie, R. (1993). Markov chains and stochastic stability. Springer, New York.

[33] Nachbin, L. (1965). Topology and order. Van Nostrand, Princeton.

[34] Rachev, S., Samorodnitsky, G. (1995). Limit laws for a stochastic process and random recursion arising in probability modelling. Adv. Appl. Prob. **27**, 185–202.

[35] Rachev, S., Todorovic, P. (1990). On the rate of convergence of some functionals of a stochastic process. J. Appl. Prob. **27**, 805–814.

[36] Revuz, D. (1984). Markov chains, 2nd edition. North-Holland, Amsterdam.

[37] Tweedie, R. (1976). Criteria for classifying general Markov chains. Adv. Appl. Prob. **8**, 737–771.

[38] Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. Adv. Appl. Prob. **11**, 750–783.

[39] von Weizsäcker, H. (1974). Zur Gleichwertigkeit zweier Arten der Randomisierung. Manuscr. Math. **11**, 91–94.

[40] Yahav, J. (1975). On a fixed point theorem and its stochastic equivalent. J. Appl. Prob. **12**, 605–611.