J. Austral. Math. Soc. 23 (Series A) (1977), 481–498.

RANDOM SEMIGROUP ACTS ON A FINITE SET

GÖRAN HÖGNÄS

(Received 20 May 1975; revised 1 February 1977)

Communicated by T. E. Hall

Abstract

Let X be a finite set and S a semigroup of transformations of X. We investigate the trace on X of a random walk on S. We relate the structure of the trace process, which turns out to be a Markov chain, to that of the random walk. We show, for example, that all periods of the trace process divide the period of the random walk.

1.

Let X be a finite set with n elements. By $\mathscr{T}_X$ or $\mathscr{T}_n$ we denote the set of all transformations of X, that is, all the maps of X into itself. $\mathscr{T}_X$ is a semigroup under composition of mappings. We write xf for the image of $x \in X$ under the mapping $f \in \mathscr{T}_X$. For basic semigroup terminology the reader is referred to the fundamental Clifford and Preston (1961).

An *act* is a function from $X \times S$ to X, with S a semigroup, satisfying

$$(1) \qquad x(s_1 s_2) = (xs_1)s_2$$

for all $x \in X$ and all $s_1, s_2 \in S$, where xs denotes the image of the point (x, s), see Day and Wallace (1967). The act makes X a *right S-semimodule* (Deussen (1971)).

The object of this paper is the study of *random acts* and the induced *random process on* X. More specifically, we investigate the process on X induced by a *random walk* on S.

The notion of random walk on semigroups is a natural extension of the ordinary notion of random walk (on the real line). In that case we study the behaviour of sums

$$s_m = \sum_{i=1}^{m} X_i$$

of independent, identically distributed random variables X_i. In the semigroup case we study products

$$\prod_{i=1}^{m} X_i$$

of independent, identically distributed random elements of a semigroup. (Recall that the semigroup operation is denoted multiplicatively.)

The fundamental work on random walks on *discrete semigroups* is Martin-Löf (1965) and we shall frequently quote results from this paper. Larisse (1972) is another very thorough paper on discrete semigroups. Much of the theory of probabilities on *compact semigroups* was developed by Rosenblatt and can be found in the monograph Rosenblatt (1971). For later work in this area see, e.g., Mukherjea, Sun and Tserpes (1973), Sun, Mukherjea and Tserpes (1973), Högnäs (1974a, 1974b). For a survey over the field of probability theory on general algebraic structures the reader is referred to the book by Grenander (1963).

The referee has kindly done a lot of work in examining my paper. I am deeply indebted to him for his helpful suggestions resulting in numerous improvements. He also pointed out a number of useful references.

2.

In this section we introduce the basic algebraic definitions and results needed in subsequent sections. Most act definitions are quoted from Day and Wallace (1967). Note that a finite space is a compact Hausdorff space in its discrete topology; hence all functions defined on it are continuous. In particular, any theorem about compact semigroups (acts) is true for finite ones. The semigroup S is said to act *unitarily* if

$$x \in xS \text{ for each } x \in X.$$

The set xS, that is $\{xs \mid s \in S\}$, is called the *orbit* of the point $x \in X$. S acts *effectively* on X if $xs = xt$ for all $x \in X$ implies $s = t$. If S acts effectively on X then S can be viewed as a subset of $\mathcal{T}_X$.

Let ρ be the equivalence relation

$$\{(s,t) \mid xs = xt \text{ for all } x \in X\}.$$

It is readily verified that ρ is a congruence and that $S/_\rho$ acts effectively. Hence $S/_\rho$ is (isomorphic to) a *subsemigroup of* $\mathcal{T}_X$.

REMARK. The notion of act appears naturally in the *theory of automata*, cf. Deussen (1971). The *input* of an automaton is a semigroup S, usually a free semigroup over some alphabet. The input of an element of S (a **word**) causes a change in the internal state of the automaton. The transition from one state x to another state is a function of that state and the input s. Denote the function by $(x, s) \curvearrowright xs$, where $s \in S$ and x belongs to the *state space* X. We require the function to satisfy (1), in other words, to be an act. The semigroup $S/_\rho$ is called the *semigroup of state transition maps*. If X is finite then the semigroup of state transition maps is also finite ($\subseteq \mathcal{T}_X$).

A non-empty subset M of X is called an *act ideal* or S-*ideal* if $MS \subseteq M$, where

$$MS \equiv \{xs \mid x \in M,\ s \in S\}.$$

M is a *minimal act ideal* if no proper subset of M is an act ideal. The sets xS are act ideals since $xSS \subseteq xS$. The intersection of two act ideals is an act ideal (or empty). Hence the minimal act ideals are either equal or disjoint. When X is finite there is at least one minimal act ideal M. Clearly $M = xS$ for each $x \in M$; in particular, $x \in xS$ for $x \in M$. Thus, in the terminology of Stadtlander (1968), p. 483, a minimal act ideal is a δ-class and its elements are minimal in the natural quasi-order on X.

S is called *transitive* (*simply transitive*) if, for each couple $(x, y) \in X \times X$ there is at least (exactly) one $s \in S$ with $xs = y$. A subset Y of X is a *set of transitivity* (and S is said to act transitively on Y) if, given any $x, y \in Y$, $xs = y$ for some $s \in S$. For details see for example Clifford and Preston (1961) and Hall (1959). The sets of transitivity are disjoint.

In the theory of random walks on semigroups the *kernel* of the semigroup is of crucial importance. We shall therefore briefly look at the kernel of the semigroup $S \subseteq \mathcal{T}_X$. For the notations and terminology the reader is referred to Clifford and Preston (1961).

Let $r \leqq |X|$ be the integer $\min\{\operatorname{rank}(s) \mid s \in S\}$. We wish to show that the set $K' \equiv \{s \in S \mid \operatorname{rank}(s) = r\}$ is the kernel of S. If S is the full transformation semigroup $\mathcal{T}_X$ then $r = 1$ and the kernel consists of all the constant functions.

For any $s, t \in S$, $|Xst| \leqq |Xs|$ and $|Xts| \leqq |Xs|$. ($|A|$ denotes the number of elements of the set A.) Hence K' is an ideal of S. To see that $K' = K$, the kernel of S, it suffices to show that $K' \subseteq K$. Let $f \in K'$ and $h \in K$. Then $fhf \in K' \cap K$. Now $|Xf|$ is minimal and $Xfhf \subseteq Xf$, so these sets are equal. Similarly, $Xfhf = Xe$, where e is the identity of the group of fhf in K (since $Xfhfe = Xfhf \subseteq Xe$ and $Xe = Xfhfg \subseteq Xfhf$, where g is an element of the group). Thus $Xf = Xe$, hence $xf = (xf)e = x(fe)$ for each $x \in X$, because e is the identity mapping on Xf. Hence $f = fe$, by the effectiveness of S. Therefore $f \in K$ and $K' \subseteq K$. I am grateful to the referee for this proof which is clearer than my original one.

We have thus proved the

THEOREM 1. *The kernel K of S is precisely the set of elements of S with minimal rank r.*

We shall henceforth refer to the minimal rank r as *the rank of the kernel K.*

A minimal right ideal of S is a set of the form

$$R_k \equiv \{s \in K \mid \pi_s = \pi_k\}$$

and a minimal left ideal can be written

$$L_k \equiv \{s \in K \mid Xs = Xk\}$$

where k is any element of K. This can be seen by the following argument. Consider a minimal right ideal R. R is a subset of R_k for some k; otherwise $R \cap R_k$ would be an ideal contained in R. Take an $s \in R$. The element sk has partition $\pi_s (= \pi_k)$ and range Xk. Its restriction to Xk is a bijection. Hence there is an n with $(sk)^n = e$, e having the property $ek = k$. Thus $k = (sk)^n k$, i.e. $k \in sS = R$ by the minimality of R. Hence $R_k = R$. There is thus a one-to-one correspondence between the set $\mathscr{R}$ of minimal right ideals and the set $\{R_k \mid k \in K\}$ as well as between the set $\mathscr{L}$ of minimal left ideals and the set $\{L_k \mid k \in K\}$.

The maximal subgroups of S are of the form

$$\{s \in K \mid \pi_s = \pi_k, Xs = Xk'\}$$

where k, k' are any elements of K.

Let π_0, X_0 be the partition and the range, respectively, of some element g in K. Let k be an element of K, with partition π and range Y. Then, $\pi_{kg} = \pi$, $Xkg = X_0$, $\pi_{gk} = \pi_0$, and $Xgk = Xk$. With the technique used above we find idempotents i, j with $\pi_i = \pi$, $X_i = X_0$, $\pi_j = \pi_0$, and $Xj = Xk$. Since k belongs to the same minimal right ideal as i and the same minimal left ideal as j, k may be written

$$k = ih = h'j$$

for some h, $h' \in K$. Again, for some power n of $h'j$, $k = k(h'j)^n = k(h'j)^{n-1}h'j = ih(h'j)^{n-1}h'j$.

The mapping $g_0 = h(h'j)^{n-1}h'$ permutes the elements of X_0. We have, then, proved the following

THEOREM 2. *Any element k of K is expressible as a product $k = ig_0 j$ where g_0 is an element of the group with fixed partition π_0 and fixed range X_0, i is the idempotent with partition π_k and range X_0, and j is the idempotent with partition π_0 and range Xk.*

REMARK. This theorem is well-known, see for instance Wallace (1957), Theorem 3 and Wallace (1963, 4.3).

In the terminology of the Rees–Susckhewitsch structure theorem the above theorem reads as follows:

COROLLARY. *K may be written as Rees product*

$$\mathcal{R} \times G \times \mathcal{L}$$

where G is a subgroup of the symmetric group $\mathcal{G}_r$, isomorphic to

$$\{k \in K \mid \pi_k = \pi_0, Xk = X_0\}$$

and $\mathcal{R}$ and $\mathcal{L}$ are the set of minimal right and left ideals, respectively. The function $\phi: \mathcal{L} \times \mathcal{R} \to G$ is given by

$$\phi(l, r) \equiv \phi(Xk, \pi) = j_{Xk} i_\pi$$

where the representatives j_{Xk} and i_π are the idempotents with partition π_0 and range Xk and partition π and range X_0, respectively.

Note that ϕ is non-trivial, in general. For future reference we include the following well-known result, see, for example, Schwarz (1964), p. 99:

LEMMA 1. *If $x, y \in K$ and G is a maximal group in K, then $xyK = xK$ and $xyG = xG$.*

3.

Let S be any discrete semigroup. If $s, t \in S$ define $s^{-1}t$ to be the set $\{u \in S \mid su = t\}$. Let ν be a probability measure on S. (We shall constantly write $\nu(s)$ for $\nu\{s\}$.) Then ν and the multiplication on S induce a *right random walk* $\{S_n\}$ on S in a natural way: Define the transition probability function by

$$\Pr\{S_n = t \mid S_{n-1} = s\} = \nu(s^{-1}t).$$

Intuitively, if $s^{-1}t$ is large, then s leads to t via many elements, and the probability of "t following s" will be large.

The random walk thus defined may also be looked upon from another point of view. Let $x_1, x_2, \cdots$ be independent random elements of S with identical distribution ν. Then the partial products $\{S_n\}$ where

$$S_n = x_1 x_2 \cdots x_n = S_{n-1} x_n$$

form a Markov chain with transition probability function $\nu(s^{-1}t)$. Obviously, we can define another Markov chain $\{T_n\}$, the *left random walk*, by multiplying from the left instead:

$$T_n = x_n x_{n-1} \cdots x_2 x_1 = x_n T_{n-1}.$$

In addition, *two-sided* and *mixed* random walk have been introduced as well, cf. Martin-Löf (1965), Högnäs (1974).

Let S be a semigroup acting on X, and consider a random walk on S. S transforms X in the same way as $S/_\rho$ where ρ is the *effectiveness congruence* defined in the beginning of section 2. Proposition 1 below guarantees that the process induced on $S/_\rho$ by a random walk on S is also a random walk (to which known results can be applied). That is why the results in section 2 were stated mainly for finite semigroups, in particular subsemigroups of $\mathcal{T}_X$.

Consider a (discrete) semigroup S and a congruence η on S. Write $[s]$ for the η-equivalence class containing $s \in S$ and put as usual

$$u^{-1}[s] = \{w \in S \mid uw \in [s]\}.$$

We have then

$$u \,\eta\, v \Rightarrow ut \,\eta\, vt \text{ for all } t, u, v \in S,$$

and hence $u^{-1}[s] = v^{-1}[s]$ for all $s \in S$. Let ν generate a right random walk on S, i.e.

$$\Pr\{S_n = t \mid S_{n-1} = s\} = \nu(s^{-1}t).$$

Consider the probability measure ν' on $S/_\eta$ defined by

$$\nu'[s] = \nu[s], \qquad s \in S.$$

By the remarks made above

$$\nu'[s]^{-1}[t] = \nu(s^{-1}[t])$$

is well defined because the expressions are independent of the choice of representative of $[s]$. The probability

$$\Pr\{[S_n] = u \mid [S_{n-1}] = v\} \qquad (u, v \in S/_\eta)$$

is just

$$\nu'(v^{-1}u).$$

We have thus shown that $\{[S_n]\}$ is a right random walk on $S/_\eta$. For the left, two-sided and mixed random walk the reasoning is analogous. We have the

PROPOSITION 1. *If $\{S_n\}$ is a random walk on S then $\{[S_n]\}$ is a random walk on $S/_\eta$.*

REMARK. If $\{S_n\}$ is just a Markov chain on S then we cannot be sure that $\{[S_n]\}$ is a Markov chain on $S/_\eta$ at all.

We will limit our study to right random walks since they seem to be the most natural in the present context (successive random transformations). In the sequel we require that S act effectively, that is, *S is a subsemigroup of $\mathcal{T}_X$* and thus *finite*. This causes no loss of generality, in view of Proposition 1.

If x and y are independent random elements of S with probability distributions μ and ν, respectively, then the distribution λ of the *product* xy is

$$\lambda(s)=\sum_{t\in S}\mu(t)\nu(ts^{-1})=\sum_{t\in S}\mu(s^{-1}t)\nu(t)$$

$$=\sum\mu(r)\nu(t)$$

where the last summation is to be taken over all r and t such that $rt = s$. The probability measure λ is called the *convolution* of μ and ν (notation: $\mu * \nu$). For $\nu * \nu$ we write $\nu^{(2)}$ and $\nu^{(n)} = \nu * \nu^{(n-1)}$, $n = 2,3,\cdots$. The *support* $C(\nu)$ of ν is the set $\{s \in S \mid \nu(s) > 0\}$. One has $C(\mu * \nu) = C(\mu)C(\nu)$. The semigroup $\bigcup_{n\geqq 1} C(\nu^{(n)}) = \bigcup_{n\geqq 1}(C(\nu))^n$ is called the *semigroup generated by* (*the support of*) ν. Section 3 of Martin-Löf (1965) is a thorough discussion of the concept of convolution of measures on a (discrete) semigroup.

If S is a finite semigroup and ν induces a right random walk $\{S_n\}$ on S, then the probabilistically interesting part of S is actually $\bigcup_{n\geqq 1}(C(\nu))^n$, that is, the subsemigroup generated by the support of ν. Thus we do not lose in generality when we assume our semigroup S to be *generated by the support of* ν. This assumption means that, given $s \in S$, there is an n such that $\nu^{(n)}(s) > 0$, in other words, there are s_i $(i = 1,\cdots,n)$ with $\nu(s_i) > 0$ and $s_1s_2\cdots s_n = s$. We shall adhere to this assumption throughout the rest of the paper. We are now going to investigate the *trace* on X of the random walk $\{S_n\}$ on S, that is, the process $\{xS_n \mid n = 0, 1 \cdots\}$ on X (where x is some element of X and $xS_0 = x$).

LEMMA 2. *For every $x \in X$, $\{xS_n\}$ is a Markov chain on X, called the trace chain.*

PROOF. Immediate. The probability of going from $xS_n \in X$ to $y \in X$ is given by

$$\sum_{xS_ns=y}\nu(s). \qquad \square$$

REMARK. As we are going to see later, *all* Markov chains on X can be described as a process $\{xS_n\}$ where $\{S_n\}$ is a random walk on (a subset of) $\mathcal{T}_X$. It is only when we have *a priori* knowledge of $\{S_n\}$ that it is really worth-while to investigate the Markov chain in this very roundabout way.

Let $P = (P(u, y))$, $P(u,y) = \Pr\{xS_{n+1} = y \mid xS_n = u\}$, be the transition probability matrix for the trace chain on X. P^n, the n-step transition probability matrix, is then just the matrix product on n factors P. By an abuse of notation set $x^{-1}y \equiv \{s \in S \mid xs = y\}$, $x, y \in X$. Then $P(x, y) = \nu(x^{-1}y)$ by the proof of Lemma 2. We shall make use of both these notations in the sequel.

Write $x \to y$ (x leads to y) if $P^n(x, y) > 0$ for some n. Here this condition interprets as $xs_1s_2 \cdots s_n = y$ for some $s_1, s_2, \cdots, s_n \in C(\nu)$ or simply $xs = y$ for some $s \in S$. So $x \to y$ means $y \in xS$. If $x \to y$ and $y \to x$ then x and y are said to *communicate* (notation: $x \sim y$). $\sim$ is an equivalence relation on the set $\{x \mid x \to x\}$ of "points of return". If S acts transitively then all of X is one equivalence class. Otherwise, the sets of transitivity are the $\sim$-classes. If $x \to y$ implies $y \to x$ then x is said to be *essential*, otherwise x is *inessential*. Here this means that $y \in xS \Rightarrow x \in yS$. The elements of a $\sim$-class are of the same type, either all essential or all inessential. Since X is finite the essential elements are precisely the recurrent elements and the inessential elements are the transient ones. Henceforth, we prefer the terms recurrent and transient. (x being recurrent means that

$$\Pr\{xS_n = x \text{ for infinitely many } n \mid xS_0 = x\} = 1.)$$

Note that

$$|\{y \mid x \to y\}|$$

is at most $|S|$. If S is simply transitive then $|S| = |X|$.

REMARK. The definitions and their interpretation are valid for the *right random walk* on S, too, provided $x, y, \cdots$ are taken to be elements of S. In that case the act $S \times S \to S$ is simply multiplication in S, cf. Martin-Löf (1965), section 4.

The next Lemma and Proposition give a connection to the notion of act ideal, defined in section 2. I am indebted to the referee who pointed out the connections to Stadtlander's work.

LEMMA 3. *For the trace chain on a finite space X these are equivalent*

1) *x is recurrent ($\Leftrightarrow$ x is essential)*
2) *$x \in xS$ and xS is a minimal act ideal.*

PROOF. 1) $\Rightarrow$ 2). Let $y \in xS$. Then $yS \subseteq xS^2 \subseteq xS$ and $x \in yS$ (since x is essential), hence $xS \subseteq yS$ also. Thus $xS = yS$ and $x \in xS$. Since this is valid for all $y \in xS$, xS is a minimal act ideal.

2) $\Rightarrow$ 1). If $y \in xS$ then $xS = yS$, because of the minimality of xS. Hence $x \in yS$ since $x \in xS$. □

PROPOSITION 2. (cf. Stadtlander (1968), Proposition 1). *x is recurrent if and only if $x \in XK$ where $XK = \{xk \mid x \in X, k \in K\}$.*

PROOF. By Lemma 3 we have to prove that x is an element of XK if and only if $x \in xS$ and xS is a minimal act ideal.

Let xS be a minimal act ideal containing x. Then $yS = xS$ for all $y \in xS$. In particular $xkS = xS$ for all $k \in K$. Hence $x \in xkS \subseteq xK \subseteq XK$. On the other hand, let $x \in XK$, that is, $x = zk$ for some $z \in X$, $k \in K$. There is an $e \in K$ with $ke = k$, whence $x = zK = zke = xe \in xS$, i.e., $x \in xS$. Since $x \in XK$, $x \in zR$ for some $z \in X$ and some minimal right ideal $R \subseteq K$. Then $zR = xS$. Let $y \in xS$; it remains to be shown that $yS = xS$ in order to prove that xS is minimal act ideal. Write $y = zr$ $(r \in R)$. Then $yS = zrS = zR = xS$. □

The observation that if xS is minimal act ideal then $xS = xK$ gives the following

COROLLARY. *x and y belong to the same recurrent class if and only if they are recurrent and $xK = yK$.*

THEOREM 3. *$xK = yK$ if and only if $xiG = yi'G$ for every maximal group G in K and some idempotents $i, i' \in K$.*

In other words, x and y belong to the same recurrent class if and only if

a) *x and y are recurrent elements*

b) *for any maximal group G in K,*

$$xiG = yi'G \text{ for some idempotents } i, i' \in K.$$

PROOF. Suppose that $xK = yK$, i.e. there are $k, k' \in K$ such that $xk = yk'$. Let i and i' be the idempotents of the groups of k and k', respectively. Then $xik = yi'k'$, whence $xikG = yi'k'G$. Therefore $xiG = yi'G$ for any maximal group G in K, by Lemma 1. Conversely, if $xiG = yi'G$ then $xK \cap yK \neq \varnothing$. Thus $xK = yK$. □

PROPOSITION 3. *The number of recurrent classes $\leqq$ the number of sets of transitivity of a maximal group $G \leqq$ the rank r of the kernel K.*

PROOF. Note first that the maximal groups in K are isomorphic, so the number of sets of transitivity is independent of the choice of G. Let x and y belong to a set of transitivity of G, which means that $y = xk$ for some $k \in G$. Hence $x \to y$. Analogously, $y \to x$. Obviously, x and y are also recurrent. Hence the number of transitivity sets of G is larger than the number of recurrent classes. The maximal number of sets of transitivity of G is r, since G is (isomorphic to) a subgroup of $\mathscr{G}_r$. □

REMARK. If G has m sets of transitivity with $r_1, r_2, \cdots, r_m$ elements $(r_1 + r_2 + \cdots + r_m = r)$ then G can be written as a *subdirect product* of groups

https://doi.org/10.1017/S1446788700019637 Published online by Cambridge University Press

G_{r_i}, which are subgroups of the symmetric groups $\mathscr{G}_{r_i}$ $(i = 1, 2, \cdots, m)$, see Hall (1959).

From now through section 5, we will limit ourselves to the recurrent elements, that is, we assume $X = XK$. Note that the recurrent classes will then be sets of transitivity of K. Let G be a maximal group. Define the equivalence relation

$$\sigma = \{(x, y) \mid xG = yG\},$$

that is, $x\sigma y$ if and only if x and y generate the same set of transitivity. σ is the equivalence relation on X induced by the sets of transitivity of G.

Let η be the smallest equivalence relation that relates x and y if $x\sigma y$ or $x\pi y$ for some $\pi \in \mathscr{R}$, in other words,

$$\eta = \sigma \vee \sup_{\pi \in \mathscr{R}} \pi.$$

For the definition of the supremum of a set of equivalence relations see, for instance, Arbib (1968).

The Corollary of Proposition 2 implies that x and y belong to the same recurrent class if and only if $xK = yK$. This is equivalent with $x \sim y$ since all points of X are assumed to be recurrent. The number of recurrent classes is therefore $|X/_\sim|$.

PROPOSITION 4. $\sim\, = \eta$.

PROOF. We verify first that $\eta \subseteq \sim$, i.e., if $x\pi y$ then $x \sim y$. $x\sigma y$ means by definition that $xG = yG$; $x(\sup_{\pi\in\mathscr{R}} \pi)y$ means that $xk = yk$ for some $k \in K$. In both cases $xK \cap yK \neq \varnothing$, hence $xK = yK$ or $x \sim y$. On the other hand, suppose that $x \sim y$. By Proposition 2, $xiG = yi'G$ for some idempotents $i, i' \in K$, that is, $xi\sigma yi'$. There is a $k \in K$ such that $ik = k$ whence $(xi)k = xk$, that is, there exists a $\pi \in \mathscr{R}$ with $x\pi xi$. Analogously, $yi'\pi' y$ for some $\pi' \in \mathscr{R}$. Hence $\sim\, \subseteq \pi \vee \sigma \vee \pi' \subseteq \eta$. $\square$

PROPOSITION 5. *If the trace chain has q recurrent classes then K can be written as a subdirect product of q transformation semigroups which are all kernel semigroups (that is, completely simple).*

PROOF. The definition of subdirect product of semigroups is a straightforward generalisation of the same notion for groups. $K \subseteq K_{C_1} \times K_{C_2} \times \cdots \times K_{C_q}$ where the C_i's are the sets of transitivity of K and K_{C_i} is the restriction of K to C_i (modulo the relation ρ defined in section 2). K is simple, that is, $KkK = K$ for all $k \in K$, whence $K_{C_i} k_{C_i} K_{C_i} = K_{C_i}$ for all $k_{C_i} \in K_{C_i}$. K is simple and thus, being finite, also completely simple. $\square$

4.

We will now look briefly at the periodicity of the trace chain in terms of the structure of S and the periodicity of the random walk on S. Let P be the transition probability matrix of a Markov chain on some discrete space Y. The *period* of an element $y \in Y$ is the greatest common divisor of $\{n \geqq 1 \mid P^n(y, y) > 0\}$. If $y \not\to y$ the period is undefined. The period is the same for all elements of a recurrent class. This number is called the period of the recurrent class.

If we consider the right random walk on the finite semigroup S, induced by the probability measure ν, the situation is still simpler: all the recurrent elements, that is, the elements of the kernel K, have the same period, to be called *the period of the kernel* K. As shown by Martin-Löf (1965), p. 87, the period of an $s \in S$ is the greatest common divisor of $\{n - m \mid \nu^{(n)}(s), \nu^{(m)}(s) > 0, n \neq m\}$.

If the random walk on S is periodic with period p, then there is a proper normal subgroup H of the group G in the Rees–Suschkewitsch decomposition (see the Corollary to Theorem 2) such that

$$(\mathscr{R} \times H \times \mathscr{L})C(\nu) = (\mathscr{R} \times gH \times \mathscr{L})$$

where $g \notin H$ and $\bigcup_{i=1}^{p} g^i H = G$, cf. Martin-Löf (1965), pp. 88, 94. Conversely, if $(\mathscr{R} \times H \times \mathscr{L})C(\nu) \subseteq \mathscr{R} \times gH \times \mathscr{L}$ for some $g \notin H$, H being a proper normal subgroup of G, then the random walk on S is obviously periodic. The period p is, of course, $\leqq$ the rank r of K, since $G \subseteq \mathscr{G}_r$.

PROPOSITION 6. *All periods of the trace chain divide* p.

PROOF. The recurrent classes of X may have different periods d. Let C_i be a recurrent class. For some sufficiently large N, $s \to s$ in np steps (that is, the probability of reaching s from s in np steps is positive) for all $n \geqq N$ and $s \in K$ (cf. Kemeny–Snell–Knapp (1966), p. 38). $x \in C_i$ can be written $x = ys$ for some $y \in X$ and $s \in K$. Hence $x \to x$ in np steps for all $n \geqq N$. Thus the period d of x ($=$ the period of C_i) divides p. $\square$

If the periods of the recurrent classes are d_i $(i = 1, \cdots, q)$ we have then $d_i \mid p$ and $r \geqq d_1 + d_2 + \cdots + d_q$.

5.

In this section, we are going to study the stationary probability distributions for the trace process on X. We have seen, in section 3, that the transition probabilities $P(x, y)$ may be written

$$P(x, y) = \nu(x^{-1}y) \equiv \nu\{s \mid xs = y\}.$$

https://doi.org/10.1017/S1446788700019637 Published online by Cambridge University Press

Then

$$P^2(x,y)=\sum_{z\in X}P(x,z)P(z,y)=\sum_{z}\sum_{zs=z}\nu(s)\sum_{zt=y}\nu(t)$$

$$=\sum_{xst=y}\nu(s)\nu(t)=\sum_{xs=y}\nu^{(2)}(s)=\nu^{(2)}(x^{-1}y)$$

and $P^n(x,y)=(\nu^{(n-1)}*\nu)(x^{-1}y)=\nu^{(n)}(x^{-1}y)$.

The probability distribution π on X is said to be *stationary* if $\Sigma_x\,\pi(x)\nu(x^{-1}y)=\pi(y)$ for all $y\in X$.

LEMMA 4. *Let* $\mu=\lim_{n\to\infty}(1/n)\Sigma_{k=1}^{n}\nu^{(k)}$. *Then* π_x *defined by*

$$\pi_x(y)=\mu(x^{-1}y)$$

is a stationary distribution on X. *The support of* π_x *is the recurrent class containing* x.

PROOF.

$$\sum_{y}\pi_x(y)P(y,u)=\sum_{y}\sum_{xs=y}\mu(s)\sum_{yt=u}\nu(t)=\sum_{xs=u}(\mu*\nu)(s)$$

$$=\sum_{xs=u}\mu(s)=\pi_x(u)$$

since $\mu*\nu=\mu$, cf. Rosenblatt (1971). The support of μ being K the support of π_x must be xK, that is, the recurrent class containing x. Finally,

$$\sum_{y}\pi_x(y)=\sum_{y}\sum_{xs=y}\mu(s)=\sum_{s}\mu(s)=1$$

shows that π_x is a probability distribution.

Since the stationary probability distribution on a recurrent class is unique we have the

PROPOSITION 7. *The stationary probability distribution on a recurrent class* C_i *is given by*

$$\pi(y)=\sum_{xs=y}\mu(s)$$

where x *is any element of* C_i.

The general theory of Markov chains tells us that

$$\frac{1}{n}\sum_{k=1}^{n}P^k(x,y)\to\mu(x^{-1}y)=\sum_{xs=y}\mu(s)$$

as well as

https://doi.org/10.1017/S1446788700019637 Published online by Cambridge University Press

$$\frac{1}{d}\sum_{k=nd+1}^{(n+1)d} P^k(x,y) \to \mu(x^{-1}y)$$

where d is the period of (the class containing) x. The first of the above limits holds even if x is assumed transient. In that case $\mu(x^{-1}y)$ is a convex combination (a weighted average) of the stationary distributions on the recurrent classes, given in Proposition 7. The second assertion holds if d is chosen to be (a multiple of) the least common multiple of the periods $d_1, d_2, \cdots, d_q$ of the recurrent classes.

The kernel K of S will be a *group* if S is commutative. Actually, it suffices to require that the elements of $C(\nu)$ generate a group; then, of course, K will also be a group ($= S$). For a transitive group we have the following

COROLLARY. *If the kernel K is a transitive group acting on a subset Y of X then the stationary distribution is uniform on Y.*

PROOF. If K is a group then μ is the uniform (Haar) measure on K. The number of elements in $x^{-1}y = \{k \mid xk = y\}$ is independent of x and y. Suppose that $x^{-1}y$ has the maximal number (N) of elements, that is, there are N group elements k with $xk = y$. If $x'g = x$ then $x'gk = y$ for all $k \in x^{-1}y$. Thus $x'^{-1}y$ also has N elements. K being a transitive group $|x^{-1}y| = |y^{-1}x| = N$ for all x and $y \in Y$.

$$\pi_x(y) = \mu(x^{-1}y) = \frac{1}{|K|}\cdot N \text{ for all } y \in Y. \qquad \square$$

6.

We remarked in section 3 that *any* Markov chain on a finite set X may be viewed as the trace of a random walk on a subsemigroup S of $\mathcal{T}_X$. All we have to do is to show that any stochastic matrix can be written as a convex combination of *transformation matrices*, that is stochastic matrices with exactly one non-zero element in each row. The non-zero element is, of course, 1. The method of proof of the following well-known result will be used later.

LEMMA 5. *Any finite stochastic matrix is a convex combination of transformation matrices.*

PROOF. Let P be a $n \times n$ stochastic matrix, that is, a matrix (p_{ij}) with $p_{ij} \geqq 0$ for all i and j $(i, j = 1, \cdots, n)$ and

$$\sum_{j=1}^{n} p_{ij} = 1 \text{ for all } i = 1, \cdots, n.$$

Let a_1 be the minimal *positive* entry in P. For each i choose j such that

https://doi.org/10.1017/S1446788700019637 Published online by Cambridge University Press

$p_{ij} \geqq a_1$ is a minimal positive entry in the ith row. Let the transformation matrix $Q^{(1)}$ have its 1's in these places (i, j). Then

$$P = P^{(1)} + a_1 Q^{(1)}$$

where $P^{(1)}$ is a matrix with non-negative entries and row sum $1 - a_1$.

Repeating the described procedure on P^1 we obtain

$$P = P^{(2)} + a_1 Q^{(1)} + a_2 Q^{(2)}$$

where $Q^{(1)}$, $Q^{(2)}$ are transformation matrices, a_1, a_2 are positive numbers with $a_1 + a_2 \leqq 1$, and $P^{(2)}$ is a matrix with non-negative entries and row sum $1 - (a_1 + a_2)$.

Continuing until the row sum of the "residual" $P^{(m)}$ is zero we obtain

$$P = a_1 Q^{(1)} + a_2 Q^{(2)} + \cdots + a_m Q^{(m)}$$

where $m \leqq n^2$ and $\sum_{i=1}^m a_i = 1$. $\square$

Let $Q^{(i)}$ also denote the transformation of $\{1, 2, \cdots, n\}$ described by the matrix $Q^{(i)}$. Define ν by $\nu(Q^{(i)}) = a_i$ $(i = 1, \cdots, m)$. Then ν is a probability distribution on $\mathcal{T}_X$ with $X = \{1, 2, \cdots, n\}$. The support $Q^{(1)}, \cdots, Q^{(m)}$ of ν generates a subsemigroup S of $\mathcal{T}_X$. We then have

$$P(x, y) = \sum_{xs = y} \nu(s).$$

We have thus proved the

PROPOSITION 8. *Any Markov chain on a finite set X may be viewed as the trace of a random walk on a semigroup $S \subseteq \mathcal{T}_X$.*

The semigroup S and the probability ν are by no means unique. Consider, for example,

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix},$$

where S in the first case is a group and in the second case a semigroup of constant mappings. The transformation matrices are permutation matrices if there is exactly one 1 in every row and every column. If all Q's are permutation matrices then S will be a group.

We showed in section 4 that the rank r satisfies the condition $d_1 + d_2 +$

https://doi.org/10.1017/S1446788700019637 Published online by Cambridge University Press

$\cdots + d_q \leqq r$, where d_i is the period of the recurrent class C_i. We now show that there exists a "minimal semigroup".

PROPOSITION 9. *There is a semigroup S the kernel K of which has rank $d_1 + d_2 + \cdots + d_q$ and period p = least common multiple $\{d_1, \cdots, d_q\}$.*

PROOF. Let $P = (p_{ij})$ and $p_{kl} = \min\{p_{ij} \mid p_{ij} > 0\}$. There is a finite number $M \leqq n^n$ of transformation matrices consistent with P ($Q = (q_{ij})$ is *consistent* with P if $q_{ij} = 1 \Rightarrow p_{ij} > 0$, cf. Rosenblatt (1971), p. 173). Give each of these consistent transformation matrices the weight $p_{kl}/M > 0$. Then continue the procedure of writing P as a convex combination of transformation matrices, for example, by the method used in the proof of Lemma 5. Let $C(\nu)$ be the set of transformations thus obtained and let $S = \bigcup_{i \geqq 1} (C(\nu))^i$ be the semigroup generated by $C(\nu)$. We want to show that S has the desired properties.

If $1 \to y_1$ (transition from 1 to y_1), $2 \to y_2, \cdots, n \to y_n$ are possible in one step then, by construction, there is an $s \in C(\nu)$ with $1s = y_1, 2s = y_2, \cdots, ns = y_n$. If $1 \to y_1, 2 \to y_2, \cdots, n \to y_n$ in two steps then there is an $s \in (C(\nu))^2$ with $is = y_i$ $(i = 1, \cdots, n)$ because $i \to y_i$ $(i = 1, \cdots, n)$ in two steps implies the existence of a z_i with $i \to z_i$ in one step and $z_i \to y_i$ in one step. There are, then, $\alpha, \beta \in C(\nu)$ with $i_\alpha = z_i$, $z_i\beta = y_i$ $(i = 1, 2, \cdots, n)$ whence $is = y_i$ $(i = 1, \cdots, n)$ if we set $s = \alpha\beta \in (C(\nu))^2$.

Analogously, $i \to y_i$ $(i = 1, \cdots, n)$ in N steps implies that there is an $s \in (C(\nu))^N$ satisfying $is = y_i$.

Let Z be an element of a cyclically moving subset D of a recurrent class C_0 with period d. For a sufficeintly large N $x \to z$ in md steps for all $m > N$ and all $x \in D$; it is a consequence of the finiteness of the chain. Thus there is an $s \in S$ such that $Ds = \{x\}$. Furthermore, N can be chosen so large that any element of any cyclically moving subset of C_0 leads to a fixed representative z of that subset in md steps. This means that there is an $s \in (C(\nu))^{md}$ such that $|C_0 s| = d$, because there are d cyclically moving subclasses of C_0.

Continuing the same line of reasoning we find that there is an $s \in (C(\nu))^{md'}$, where m is large and d' is the least common multiple of the periods $d_1, \cdots, d_q$, such that s maps all recurrent elements into a set of representatives of all the cyclically moving subclasses. Since the transient elements of X eventually lead to some recurrent element we can choose s, by increasing m if necessary, such that the transient elements, too, are mapped into the set of representatives of the cyclically moving subclasses. Thus $|Xs| = d_1 + d_2 + \cdots + d_q$. Since that rank r of the kernel is at least $d_1 + d_2 + \cdots + d_q$ we have proved that there is equality in this particular case. Let d' = least common multiple of $d_1, \cdots, d_q$ and let Y be a set of representatives of the cyclically moving subclasses, $|Y| = d_1 + d_2 + \cdots + d_q$. Let $(y_1, \cdots, y_r)$ be a

https://doi.org/10.1017/S1446788700019637 Published online by Cambridge University Press

permutation of Y. Then $y_i \to y_i$ in $m_i d(y_i)$ steps where m_i is an integer and $d(y_i)$ is the period of y_i. Thus $y_1 \to y_1, y_2 \to y_2, \cdots, y_r \to y_r$ only in md' steps, where m is an integer. Using the same reasoning as in Kemeny–Snell–Knapp (1966), p. 38, we find that the period of the transformation mapping $(y_1, \cdots, y_r)$ into $(y_1, \cdots, y_r)$ is d'. Hence the random walk on S has period d' (for a discussion of the periodicity of the right and left walks on S see Martin-Löf (1965), p. 87). □

Proposition 10. *Let all elements of X be recurrent. If S is a minimal semigroup in the sense of Proposition 9 then its kernel K is the unique minimal right ideal of S and the map φ in the Rees–Suschkewitsch decomposition is trivial.*

Proof. Define π by $x \pi y \Leftrightarrow x$ and y belong to the same cyclically moving subclass. If there is a $k \in K$ such that $xk = yk$ then x and y must belong to the same subclass. Thus $\pi_k \subseteq \pi$. But π_k and π have both r equivalence classes whence $\pi_k = \pi$, that is, there is only one minimal right ideal. Consider $\varphi(Xk, \pi) = j_{Xk} i_\pi = j_{Xk} i_{\pi_0}$, cf. section 2. If $x \in X_0$ then $x \pi_0 x j_{Xk}$ whence $x = x j_{Xk} i_{\pi_0}$. In other words, $j_{Xk} i_\pi$ is the identity mapping on X_0. □

Corollary. *If S is minimal then φ is trivial.*

Proof. The proof above works in the general case, too, since X_0 is a subset of XK = the set of recurrent elements of X. □

Proposition 11. *If S is minimal then G in the Rees–Suschkewitsch decomposition of K is a cyclic group with p elements. G is a subdirect product of q cyclic groups with $d_1, d_2, \cdots, d_q$ elements, respectively, where d_i is the period of the ith recurent class.*

Proof. G maps the cyclically moving subclasses into one another; there is a $g \in G$ taking D_{ij} (the j^{th} subclass of the i^{th} recurrent class) into D_{ij+1}, where the addition is understood to be addition (mod d_i). Take, for example, $g = ese$, where e is the identity of G and $s \in C(\nu)$. g^p is the least power of g taking D_{ij} into itself. Let h be any element of G. Then h may be written $es^{(m)}e$ for some $s^{(m)} \in C(\nu)^m$. But $es^{(m)}e = es^m e = (ese)^m$ since ese is independent of the choice of $s \in C(\nu)$. Thus G is cyclic. Trivially, G is a subdirect product of the cyclic group on the different recurrent classes, which are precisely the sets of transitivity of G. □

Notice that the preceding Proposition implies that the group G is uniquely determined, that is, the minimal semigroup is essentially unique.

We sum up the last Propositions in a

THEOREM 4. *A Markov chain on a finite set X can always be written as the trace process of a random walk on a subsemigroup S of the full transformation semigroup $\mathcal{T}_X$. S can be chosen such that the rank of its kernel K is equal to the sum of the periods of the different recurrent classes of the Markov chain. The period p of K is the least common multiple of the periods. Furthermore, the group component G of K in the Rees–Suschkewitsch decomposition is cyclic of order p and may be written as the subdirect product of the simply transitive cyclic groups of order d_i where d_i is the period of the i^{th} recurrent class. The function φ in the decomposition and the right ideal structure of K are trivial.*
In other words, the group G completely describes the periodic structure of the Markov chain on X.

7.

Consider the infinite convex combination

$$P = \sum a_i Q^{(i)}$$

of transformation matrices. The largest weight a_i has to be $\leqq$ maximal row element in every row of P. Hence the stochastic matrix defined by

$$P_{nj} = \begin{cases} n^{-1} & \text{for} \quad j \leqq n \\ 0 & \text{for} \quad j > n\,;\, n = 1, 2, \cdots \end{cases}$$

cannot be written as a convex combination of transformation matrices.

This example shows that the results in section 8 do not extend to the infinite case. Simple examples also show that Markov chains on a compact space X are not, in general, trace processes of a random walk on a compact transformation semigroup. We hope to discuss these more general cases in subsequent papers.

The author is indebted to Magnus Ehrnrooths Stiftelse and Svenska Vetenskapliga Centralrådet for their financial support.

References

M. Arbib (ed.) (1968), *Algebraic Theory of Machines, Languages, and Semigroups* (Academic Press, 1968).

A. H. Clifford and G. B. Preston (1961), *The Algebraic Theory of Semigroups*, Volume I (Amer. Math. Soc., 1961).

J. M. Day and A. D. Wallace (1967), 'Multiplication induced in the state space of an act', *Math. Sys. Theory* **1**, 305–314.

P. Deussen (1971), *Halbgruppen und Automaten* (Springer, 1971).

U. Grenander (1963), *Probabilities on Algebraic Structures* (Wiley, 1963).

M. Hall (1959), *The Theory of Groups* (MacMillan, 1959).

G. Högnäs (1974a), 'Marches aléatoires sur un demi-groupe compact', *Ann. Inst. Henri Poincaré*, Sect. B **10**, 115–154.

G. Högnäs (1974b), 'Remarques sur les marches aléatoires dans un demi-groupe avec un idéal compact ayant une probabilité positive', *Ann. Inst. Henri Poincaré*, Sect. B **10**, 345–354.

J. K. Kemeny, J. L. Snell, and A. W. Knapp (1966), *Denumerable Markov Chains* (Van Nostrand, 1966).

J. Larisse (1972), 'Marches au hasard sur les demi-groupes discrets', *Ann. Inst. Henri Poincaré*, Sect. B **8**, 107–125, 127–173, 229–240.

P. Martin-Löf (1965), 'Probability theory on discrete semigroups', *Z. Wahrscheinlichkeitstheorie* **4**, 78–102.

A. Mukherjea, T. C. Sun and N. A. Tserpes (1973), 'Random walks on compact semigroups', *Proc. Amer. Math. Soc.* **39**, 599–609.

M. Rosenblatt (1971), *Markov Processes. Structure and Asymptotic Behavior* (Springer, 1971).

S. Schwarz (1964), 'Convolution semigroup of measures on compact non-commutative semi-groups', *Czechosl. Math. J.* **14** (89), 95–115.

D. Stadtlander (1968), 'Thread actions', *Duke Math. J.* **35**, 483–490.

T. C. Sun, A. Mukherjea and N. A. Tserpes (1973), 'On recurrent random walks on semigroups', *Trans. Amer. Math. Soc.* **185**, 213–227.

A. D. Wallace (1957), 'Retractions in semigroups', *Pac. J. Math.* **7**, 1513–1517.

A. D. Wallace (1963), 'Relative ideals in semigroups. II', *Acta Math. Acad. Sci. Hung.* **14**, 137–148.

Department of Mathematics,
Åbo Akademi,
Henriksgatan 9,
SF–20500 Åbo 50, Finland,