INTRINSIC MARKOV CHAINS(1)

BY

WILLIAM PARRY

0. **Introduction.** The object of this paper is to analyse the structure of stochastic processes with finitely many states which behave qualitatively like Markov chains, in that the 'possible' sequences of the processes are determined by a chain rule. Such processes are called *intrinsically Markovian*.

In §1 we establish some necessary and sufficient conditions in order that a stochastic process be intrinsically Markovian. In §2 we investigate an equivalence relation *compatibility* (weaker than probability equivalence—*equivalence*) between stochastic processes with finitely many states. Within the compatibility class of an intrinsic Markov chain there are stationary Markov chains and processes which we term *piecewise linear*. These latter processes are in turn equivalent to stationary Markov chains. In §3 we define the *absolute entropy* of a stochastic process with finitely many states and show that this is an invariant of compatibility which dominates all the (probability) entropies of stationary processes within the compatibility class of an intrinsic Markov chain.

However, there is a unique stationary probability whose entropy is equal to the absolute entropy. This probability makes the process a Markov chain. Moreover, this Markov chain is equivalent to a process which is not only piecewise linear, but *uniformly piecewise linear*. This result leads to the conclusion that for every positive number between zero and the absolute entropy, there is a compatible stationary Markov chain (equivalent to a piecewise linear process) with this number as its entropy. We also outline a simple procedure for determining the absolute entropy and the chain which has this maximal entropy. An incidental result states that a process which behaves information theoretically like a Markov chain must be a Markov chain. These results appear in §3 and §4.

DEFINITIONS.

1. A nonatomic stochastic process with a finite number of states (n.p.f.) is a system $(X, \mathscr{B}, m, T)$ where:

(i) For some integer $s \geqq 2$,

$$X \subset \{x = x_0, x_1, \cdots : x_i \in (0, 1, \cdots, s-1)\}.$$

(ii) $\mathscr{B}$ is the σ-algebra generated by cylinders $C_n(x)$ of X, where

Received by the editors February 2, 1963.

(1) This research was supported in part by NSF G-25222.

$$C_n(x) = \{y : y_0, \cdots, y_n = x_0, \cdots, x_n\}$$
$$= (x_0, \cdots, x_n)$$

(iii) m is a nonatomic probability on B,

$$m\{x\} = 0 \text{ for all } x \in X.$$

(iv) T is the shift transformation of X onto itself,

$$T(x_0, x_1, \cdots) = (x_1, x_2, \cdots)$$

and T is nonsingular on cylinders, i.e.,

$$mT^{-1}C_n(x) = 0 \text{ if and only if } mC_n(x) = 0.$$

(We may suppose without loss of generality that $mC_0(x) > 0$ for all $x \in X$.)

2. An n.p.f. is called *transitive of order* k if for each pair of cylinders $(x_1, \cdots, x_k)$, $(y_1, \cdots, y_k)$ of length k with positive probability there exists a finite sequence $z_1, \cdots, z_n$ such that

$$m(x_1, \cdots, x_k, z_1, \cdots, z_n, \; y_1, \cdots, y_k) > 0.$$

An n.p.f. which is transitive of order 1 is said to be, simply, *transitive.*

3. An n.p.f. is said to be *intrinsically Markovian of order* k if

$$m(x_0, \cdots, x_n) > 0 \text{ and } m(x_{n-k+1}, \cdots, x_{n+1}) > 0$$

implies $m(x_0, \cdots, x_n, x_{n+1}) > 0$.

An n.p.f. which is intrinsically Markovian of order 1 is said to be, simply, *intrinsically Markovian.*

4. An n.p.f. is said to be *reduced* if $m\,C_n(x) > 0$ for all $x \in X$, $n = 0, 1, \cdots$.

5. $(X, \mathscr{B}, m, T)$ and $(X, \mathscr{B}, p, T)$ are said to be *compatible* $(m \sim p)$ when $m\,C_n(x) > 0$ implies $p\,C_n(x) > 0$, and vice versa.

6. $(X, \mathscr{B}, m, T)$ and $(X, \mathscr{B}, p, T)$ are said to be *equivalent* $(m \approx p)$ when $m(E) > 0$ implies $p(E) > 0$ for all $E \in \mathscr{B}$ and vice versa.

7. The *structure matrix* of an intrinsic Markov chain $(X, \mathscr{B}, m, T)$ is defined as the $s \times s$ matrix

$$\Sigma = \| \sigma(i,j) \|, \quad i,j = 0, 1, \cdots, s-1$$

where

$$\sigma(i,j) = 0 \text{ if } m(i,j) = 0$$
$$= 1 \text{ otherwise, and } s \text{ is the number of states.}$$

Evidently Σ is an invariant of compatibility.

REMARK 1. It is not difficult to see that an n.p.f. can always be replaced by a reduced n.p.f. In fact if $\mathscr{N}$ is the set of cylinders $C(x)$ for which $m\,C(x) = 0$, then a reduced n.p.f. is obtained by considering the n.p.f. induced on the set $X^* = X - \bigcup_{C(x) \in \mathscr{N}} C(x)$. (Note that $\mathscr{N}$ is a countable collection of null cylinders.)

REMARK 2. If $(X, \mathscr{B}, m, T)$ is transitive of order k, so is its reduction.

REMARK 3. The reduction of an n.p.f. which is intrinsically Markovian of order k is also intrinsically Markovian of order k.

1. Intrinsic structure.

LEMMA 1. *An n.p.f. $(X, \mathscr{B}, m, T)$ is intrinsically Markovian of order k, if and only if*

$$m\,C_{k-1}(Tx) = m\,TC_k(x)$$

whenever $m\,C_k(x) > 0$.

We remark that $TC_k(x) = C_{k-1}(Tx) \cap TC_0(x)$. Consequently the condition

$$m\,C_{k-1}(Tx) = m\,TC_k(x)$$

is equivalent to

$$TC_0(x) \supset C_{k-1}(Tx) \quad [m], \text{ i.e., } m[C_{k-1}(Tx) - TC_0(x)] = 0.$$

Lemma 1 is a consequence of:

LEMMA 2. *A reduced n.p.f. is intrinsically Markovian of order k, if and only if*

$$C_{k-1}(Tx) = TC_k(x) \text{ or } TC_0(x) \supset C_{k-1}(Tx).$$

Proof. Suppose $m(x_0, \cdots, x_n) > 0$ and $m(x_{n-k+1}, \cdots, x_n, x_{n+1}) > 0$ implies, $m(x_0, \cdots, x_{n+1}) > 0$ and

$$TC_k(x) \subsetneqq C_{k-1}(Tx),$$

i.e., $m(x_0, \cdots, x_k) > 0$ and

$$T(x_0, \cdots, x_k) \subsetneqq (x_1, \cdots, x_k).$$

Then there exists $C = (x_1, \cdots, x_k, x_{k+1}, \cdots, x_{k+l}) \neq \varnothing$ such that $(x_0) \cap T^{-1}C = \varnothing$ but

$$(x_0) \cap T^{-1}(x_1, \cdots, x_{k+l-1}) \neq \varnothing.$$

Therefore $m(x_0, \cdots, x_{k+l-1}) > 0$ and $m(x_l, \cdots, x_{k+l}) > 0$ and by hypothesis $m(x_0, \cdots, x_{k+l}) > 0$ which is a contradiction.

Suppose $C_{k-1}(Tx) = TC_k(x)$. Suppose $m(x_0, \cdots, x_n) > 0$, $m(x_{n-k+1}, \cdots, x_{n+1}) > 0$, and $m(x_0, \cdots, x_{n+1}) = 0$. Then $(x_0, \cdots, x_{n+1}) = \varnothing$ and

$$\begin{aligned} \varnothing &= T(x_0, \cdots, x_{n+1}) = T[(x_0, \cdots, x_k) \cap T^{-(k+1)}(x_{k+1}, \cdots, x_{n+1})] \\ &= (x_1, \cdots, x_k) \cap T^{-k}(x_{k+1}, \cdots, x_{n+1})(^2) \\ &= (x_1, \cdots, x_{n+1}). \end{aligned}$$

(2) We have not assumed T to be one-one. Nevertheless, if $A \subset (x_i)$ and B is any set, then $T(A \cap T^{-1}B) = TA \cap B$.

Proceeding in this way we get

$$(x_{n-k+1},\cdots,x_n,x_{n+1}) = (x_{n-k+1},\cdots,x_n) \cap T^{-k}(x_{n+1})$$
$$= \varnothing,$$

which is absurd since

$$m(x_{n-k+1},\cdots,x_{n+1}) > 0.$$

It is not difficult to see that an n.p.f. which is intrinsically Markovian of order k can be regarded as intrinsically Markovian of order 1, by enlarging and recoding the state space. In fact we need only assign non-negative integers to the distinct cylinders $C_{k-1}(x)$ and note that

$$(x) \equiv (x_0,x_1,\cdots) = C_{k-1}(x) \cap T^{-1}C_{k-1}(Tx) \cap T^{-2}C_{k-1}(T^2x) \cap \cdots$$

for if

$$y \in C_{k-1}(x) \cap T^{-1}C_{k-1}(Tx) \cap \cdots$$

then

$$\begin{aligned} y_0,\cdots,y_{k-1} &= x_0,\cdots,x_{k-1} \\ y_1,\cdots,y_k &= x_1,\cdots,x_k \\ &\cdots\cdots\cdots \\ y_k,\cdots,y_{2k-1} &= x_k,\cdots,x_{2k-1} \\ &\cdots\cdots\cdots \end{aligned}$$

i.e.,

$$y = x.$$

After recoding, the condition $C_{k-1}(Tx) = TC_k(x)$ for a reduced n.p.f. is then transformed into the condition $C_0(Tx) = TC_1(x)$ for the new cylinders and transitivity of order k becomes transitivity of order 1.

From here on we assume that $(X,\mathscr{B},m,T)$ is a fixed reduced process which is intrinsically Markovian and transitive of order 1.

2. **Compatibility of probabilities.** If $\Sigma = \|\sigma(i,j)\|$ is the structure matrix of $(X,\mathscr{B},m,T)$ we say that $\tau = \|\tau(i,j)\|$ is a stochastic transition matrix compatible with Σ, if τ is $s \times s$ (s is the number of states) and $\tau(i,j) > 0$ if $\sigma(i,j) = 1$, $\tau(i,j) = 0$ if $\sigma(i,j) = 0$ and if $\sum_j \tau(i,j) = 1$.

Theorem 1. *For every transition matrix τ compatible with Σ, $(X,\mathscr{B},m,T)$ is compatible with a stationary ergodic Markov chain $(X,\mathscr{B},p,T)$, with τ as its transition matrix.*

Proof. It is well known that there exist "stationary absolute positive probabilities" $p(0),\cdots,p(s-1)$ such that the probability defined by

$$p(x_0,\cdots,x_n) = p(x_0)\,\tau(x_0\,x_1)\cdots\tau(x_{n-1}\,x_n)$$

makes $(X,\mathscr{B},p,T)$ an ergodic stationary Markov chain. (The ergodicity of T and the positivity of p are consequences of the transitivity of T.) Moreover $p \sim m$.

$(X,\mathscr{B},p,T)$ is said to be a *piecewise linear process* if for each $i = 0,\cdots,s-1$ there exist constants $K(i)$ such that

$$K(i)\,p(i,x_1,\cdots,x_n) = p(x_1,\cdots,x_n)$$

for all cylinders $(i,x_1,\cdots,x_n)$ with

$$p(i,x_1,\cdots,x_n) > 0\,.$$

$(X,\mathscr{B},p,T)$ is said to be *uniformly piecewise linear* if it is linear and if the multiplicative factors $K(0),\cdots,K(s-1)$ are all equal.

THEOREM 2. *For every s positive numbers $l(0),\cdots,l(s-1)$ with $\sum_i l(i) = 1$, $(X,\mathscr{B},m,T)$ is compatible with a piecewise linear process $(X,\mathscr{B},p,T)$ such that $p(i) = l(i)$. (The piecewise linear process is completely determined by $p(i) = l(i)$, $i = 0,\cdots,s-1$.)*

Proof. Define $p(i) = l(i)$,

$$K(i) = \frac{1}{p(i)} \sum_j p(j)\sigma(i,j)$$

and

$$p(i,j) = \frac{1}{K(i)}\, p(j)\sigma(i,j);$$

then

$$\sum_j p(i,j) = p(i).$$

Define

$$p(x_0,\cdots,x_n) = \frac{p(x_n)}{K(x_0)K(x_1)\cdots K(x_{n-1})} \quad \text{if } m(x_0,\cdots,x_n) > 0,$$
$$= 0 \text{ otherwise}.$$

It will suffice to note that if $p(i,x_1,\cdots,x_n) > 0$ then

$$p(i,x_1,\cdots,x_n) = \frac{p(x_n)}{K(i)K(x_1)\cdots K(x_{n-1})}$$
$$= \frac{1}{K(i)}\, p(x_1,\cdots,x_n)\,.$$

Consequently $(X,\mathscr{B},p,T)$ is a piecewise linear process and $p \sim m$.

It is clear that there is only one piecewise linear process $(X,\mathscr{B},p,T)$ for which $p(i) = l(i)$, $i = 0,\cdots,s-1$.

THEOREM 3. *If $(X,\mathscr{B},m,T)$ is a piecewise linear process there exists a probability p, equivalent to m, such that $(X,\mathscr{B},p,T)$ is a stationary ergodic Markov chain.*

Proof. Define a transition matrix $\tau = \| \tau(i,j) \|$ by

$$\tau(i,j) = \frac{m(i,j)}{m(i)}.$$

There exist stationary absolute probabilities $p(0), \cdots, p(s-1)$ such that the probability

$$p(x_0, \cdots, x_n) = p(x_0)\tau(x_0, x_1) \cdots \tau(x_{n-1}, x_n)$$

makes $(X, \mathscr{B}, p, T)$ an ergodic stationary Markov chain.

We note that

$$p(x_0, \cdots, x_n) = p(x_0)\frac{m(x_0 x_1)}{m(x_0)} \cdots \frac{m(x_{n-1}x_n)}{m(x_{n-1})}$$

$$= p(x_0) \frac{m(x_n)}{m(x_0)\ K(x_0)\ \cdots\ K(x_{n-1})} \text{ (if } p(x_0, \cdots, x_n) > 0)$$

and

$$m(x_0, \cdots, x_n) = \frac{1}{K(x_0) \cdots K(x_{n-1})}\, m(x_n).$$

Consequently

$$\frac{p(x_0, \cdots, x_n)}{m(x_0, \cdots, x_n)} = \frac{p(x_0)}{m(x_0)}$$

and therefore,

$$\frac{dp}{dm} = h(x_0) \text{ on } (x_0)$$

where $h(x_0)$ is a finite positive number depending only on x_0.

3. **Absolute entropy.**

LEMMA 3. *For each $i,j \in (0, \cdots, s-1)$ there exists n such that the (i,j) entry of Σ^n is positive.*

Proof. We will show by induction that if $m(i, x_1, \cdots, x_{n-1}, j) > 0$ then the (i,j) entry of Σ^n is positive. (The lemma will then follow from the transitivity of $(X, \mathscr{B}, m, T)$.) This is certainly true for $n = 1$, by the definition of Σ. Suppose that it is true for $n = k$. If $m(i, x_1, \cdots, x_{k+1}, j) > 0$ then $m(i, \cdots, x_{k+1}) > 0$ and $m(x_{k+1}, j) > 0$. Consequently the (i, x_{k+1}) entry of Σ^k is positive, and the (x_{k+1}, j) entry of Σ^k is positive. But the (i,j) entry of Σ^{k+1} is $a(i,j) = \sum_{l=0}^{s-1} b(i,l) \cdot \sigma(l,j) > 0$, where $b(i,l)$ is the (i,l) entry of Σ^k.

LEMMA 4. *Σ has a simple eigenvalue $\beta > 1$ such that the right and left eigenvectors corresponding to it have all their entries positive and $|\beta| > |z|$ for all eigenvalues z other than β.*

Proof. Cf. [1] and Lemma 3.

THEOREM 4. *$(X,\mathcal{B},m,T)$ is compatible with a uniformly piecewise linear process $(X,\mathcal{B},l,T)$.*

Proof. Let $\eta=(l(0),\cdots,l(s-1))$ be the right eigenvector corresponding to β (defined in Lemma 4) such that $|\eta|=l(0)+\cdots+l(s-1)=1$. Define

$$l(x_0,\cdots,x_n) = \frac{l(x_n)}{\beta^n} \text{ if } m(x_0,\cdots,x_n)>0,$$

$$= 0 \qquad \text{otherwise}.$$

It suffices to note that

$$K(i)= \beta = \frac{1}{l(i)}\cdot \sum_j l(j)\sigma(i,j),$$

by virtue of the proof of Theorem 2.

THEOREM 5. *The unique stationary probability p equivalent to l (defined in Theorem 4) is given by*

$$p(E) = \int_E h(x)\,dl,$$

where

$$h(x) = h(i) \text{ on } (i)$$

and

$$\xi = (h(0),\cdots,h(s-1))$$

is the left eigenvector of Σ corresponding to β for which

$$\sum_{i=0}^{s-1} h(i)\,l(i) = 1.$$

Proof. We know that $(X,\mathcal{B},l,T)$ is equivalent to an ergodic stationary Markov chain $(X,\mathcal{B},p,T)$ (by Theorem 3). Consequently p is unique. Moreover, the absolute probabilities $p(0),\cdots,p(s-1)$ are given by the left eigenvector of the transition matrix $\tau=\|\tau(i,j)\|$ where

$$\tau(i,j)=\frac{l(i,j)}{l(i)}=\frac{l(j)\,\sigma(i,j)}{\beta l(i)}.$$

Therefore

$$p(j) = \sum_i \frac{l(j)}{\beta l(i)}\cdot \sigma(i,j)\,p(i) \qquad (j=0,\cdots,s-1),$$

i.e,

$$\left(\frac{p(0)}{l(0)},\cdots,\frac{p(s-1)}{l(s-1)}\right)$$

is a left eigenvector of Σ corresponding to β and

$$h(x) = \frac{dp}{dl} = \frac{p(j)}{l(j)} = h(j) \text{ on } (j).$$

The *entropy* of a stationary process $(X, \mathscr{B}, q, T)$ is defined as

$$h_q(T) = \lim_{n\to\infty} \frac{-1}{n} \cdot \sum_{C_n(x)} qC_n(x) \log qC_n(x).$$

It is not difficult to compute the entropy of $(X, \mathscr{B}, p, T)$ where p is defined as in Theorem 5; in fact

$$h_p(T) = \log \beta.$$

We define an invariant of compatibility in the following way:

$$e(T) = \lim_{n\to\infty} \frac{1}{n} \log \theta_n$$

where θ_n is the number of cylinders $C_n(x)$ for which $mC_n(x) > 0$. $e(T)$ is called the *absolute entropy* of the process $(X, \mathscr{B}, m, T)$.

THEOREM 6. *The limit*

$$e(T) = \lim_{n\to\infty} \frac{1}{n} \log \theta_n$$

exists and $e(T) \geqq h_m(T)$ *or all compatible probabilities* m.

Proof. It is not difficult to see that

$$\theta_{nm} \leqq (\theta_n)^m.$$

Consequently

$$\frac{1}{nm} \log \theta_{nm} \leqq \frac{1}{n} \log \theta_n.$$

The proof now follows a well-known procedure [2]. Let $\alpha = \liminf n^{-1} \log \theta_n$, and choose $\varepsilon > 0$. Choose N so that

$$\alpha \leqq \frac{1}{N} \log \theta_N < \alpha + \varepsilon.$$

If $n > N$ then $n = mN + k$ where $0 \leqq k < N$, and

$$\frac{1}{n} \log \theta_n \leqq \frac{1}{n} \log \theta_{(m+1)N} \leqq \frac{1}{mN} \log \theta_{(m+1)N}$$

$$\leqq \frac{m+1}{m} \frac{1}{N} \log \theta_N < \frac{m+1}{m} (\alpha + \varepsilon).$$

This is sufficient to establish the convergence of $n^{-1}\log\theta_n$. We note that $-\sum_{C_n(x)} mC_n(x)\log mC_n(x)$ is maximised, by the distribution of equal weights to each $C_n(x)$ for which $mC_n(x) > 0$, i.e.,

$$-\sum_{C_n(x)} mC_n(x)\log mC_n(x) \leqq \sum_{C_n(x)} \frac{1}{\theta_n}\log\theta_n = \log\theta_n.$$

Consequently, $h_m(T) \leqq e(T)$.

THEOREM 7.

$$e(T) = \log\beta$$

where β is the maximum eigenvector of Σ.

Proof. Let ϕ_n^i be the number of $C_n(x)$ for which $x_n = i$ and $mC_n(x) > 0$. Then $\theta_n = \sum_i \phi_n^i = |\phi_n|$ where ϕ_n is the row vector $(\phi_n^0, \cdots, \phi_n^{s-1})$. Evidently $\phi_{n+1}^j = \phi_n^0\sigma(0,j) + \cdots + \phi_n^{s-1}\sigma(s-1,j)$. Consequently

$$\phi_n = \phi_0\Sigma^n = (1,\cdots,1)\,\Sigma^n.$$

But $\xi\,\Sigma = \beta\xi$ where ξ is the positive eigenvector defined in Theorem 5. If ε, δ are the least and greatest entries of ξ, respectively

$$\frac{1}{\varepsilon}\,\xi\,\Sigma \geqq (1,\cdots,1)\,\Sigma \geqq \frac{1}{\delta}\,\xi\,\Sigma,$$

i.e.,

$$\frac{1}{\varepsilon}\,\xi\,\Sigma^n \geqq (1,\cdots,1)\,\Sigma^n \geqq \frac{1}{\delta}\,\xi\,\Sigma^n.$$

Therefore

$$\frac{1}{\varepsilon}|\beta^n\xi| \geqq |\phi_n| = \theta_n \geqq \frac{1}{\delta}|\beta^n\xi|,$$

i.e.,

$$\frac{1}{\varepsilon}|\xi|\beta^n \geqq \theta_n \geqq \frac{1}{\delta}|\xi|\beta^n,$$

and $\lim_{n\to\infty} n^{-1}\log\theta_n = \log\beta$.

4. **The uniqueness of a maximal process.** We have shown that $\log\beta$ is the maximal entropy for all compatible stationary probabilities. Moreover this entropy is assumed by one stationary ergodic Markov chain. We devote this section to proving that this is the only stationary process within the compatibility class of probabilities which assumes this entropy.

Our first result states that a process which behaves like a Markov chain, from the point of view of information theory, is in fact a Markov chain.

For the notation we adopt cf. [3]. Let $\mathscr{A}$ denote the partition of X into sets A_i, $i = 0,\cdots,s-1$, where $A_i = (i) = \{x : x_0 = i\}$. Let $(X,\mathscr{B},q,T)$ be a stationary process. It is well known that if $(X,\mathscr{B},q,T)$ is a Markov chain then

$$H(T^{-1}\mathscr{A}/\mathscr{A}) = H(T^{-(n+1)}\mathscr{A}/T^{-n}\mathscr{A})$$
$$= H\left(T^{-(n+1)}\mathscr{A}\Big/\bigvee_{i=0}^{n} T^{-i}\mathscr{A}\right).$$

THEOREM 8. *If* $H(T^{-1}\mathscr{A}/\mathscr{A}) = H(T^{-(n+1)}\mathscr{A}/\bigvee_{i=0}^{n}T^{-i}\mathscr{A}$ *for* $n = 0,1,\cdots,$ *then* $(X,\mathscr{B},q,T)$ *is a Markov chain.*

Proof. In general, for partitions $\mathscr{C},\mathscr{A}_i$ $(i = 0,1,\cdots)$

$$H\left(\mathscr{C}\Big/\bigvee_{i=0}^{n}\mathscr{A}_i\right) \leqq H\left(\mathscr{C}\Big/\bigvee_{i=1}^{n}\mathscr{A}_i\right)$$

and equality holds if and only if $\mathscr{C}$ and $\mathscr{A}_0$ are independently distributed when $\mathscr{A}_1,\cdots,\mathscr{A}_n$ are fixed, i.e.,

$$q\left(C\cap A\Big/\bigvee_{i=1}^{n}\mathscr{A}_i\right) = q\left(C\Big/\bigvee_{i=1}^{n}\mathscr{A}_i\right)q\left(A\Big/\bigvee_{i=1}^{n}\mathscr{A}_i\right)$$

for all $C\in\mathscr{C}$, $A\in\mathscr{A}_0$; cf. [4].

If $H(T^{-1}\mathscr{A}/\mathscr{A}) = H(T^{-(n+1)}\mathscr{A}/\bigvee_{i=0}^{n}T^{-1}\mathscr{A})$ for $n = 0,1,\cdots$, then (cf. [4])

$$H(T^{-(n+1)}\mathscr{A}/T^{-n}\mathscr{A}) = H(T^{-(n+1)}\mathscr{A}/T^{-(n-1)}\mathscr{A} \vee T^{-n}\mathscr{A})$$
$$= \cdots\cdots\cdots\cdots$$
$$= H(T^{-(n+1)}\mathscr{A}/T^{-1}\mathscr{A} \vee\cdots\vee T^{-n}\mathscr{A})$$
$$= H(T^{-(n+1)}\mathscr{A}/\mathscr{A} \vee\cdots\vee T^{-n}\mathscr{A}).$$

Consequently, for all $A,B\in\mathscr{A}$

$$q\left(T^{-(n+1)}A\cap B\Big/\bigvee_{i=1}^{n}T^{-i}\mathscr{A}\right) = q\left(T^{-(n+1)}A\Big/\bigvee_{i=1}^{n}T^{-i}\mathscr{A}) \times q(B\Big/\bigvee_{i=1}^{n}T^{-i}\mathscr{A}\right),$$

i.e.,

$$\frac{q(x_0,x_1,\cdots,x_n,x_{n+1})}{q(x_1,\cdots,x_n)} = \frac{q(x_1,\cdots,x_{n+1})}{q(x_1,\cdots,x_n)}\cdot\frac{q(x_0,\cdots,x_n)}{q(x_1,\cdots,x_n)}$$

when $q(x_1,\cdots,x_n) > 0$.

This relation clearly defines a Markov chain.

THEOREM 9. *If* $h_q(T) = \log\beta$ *where* $q\sim m$, *then* $(X,\mathscr{B},q,T)$ *is a Markov chain.*

Proof. It is not difficult to show that

$$h_q(T) = \lim_{n\to\infty}\frac{1}{n}H_n^1(\mathscr{A})$$

where $H_n^1(\mathscr{A}) = H(\bigvee_{i=1}^n T^{-i}\mathscr{A}/\mathscr{A})$. Moreover, $H_{nk}^1(\mathscr{A}) \leqq kH_n^1(\mathscr{A})$, and if $H_k^1(\mathscr{A}) < kH_1^1(\mathscr{A})$ for any k, then $h_q(T) < H_1^1(\mathscr{A})$. But $H_1^1(\mathscr{A})$ is the entropy of the Markov chain with transition matrix $\|\tau(i,j)\|$

$$\|\tau(i,j)\| = \frac{q(i,j)}{q(i)} \quad [2].$$

Consequently $h_q(T) < \log\beta$ if

$$H_k^1(\mathscr{A}) < kH_1^1(\mathscr{A}).$$

Hence by hypothesis,

$$H_k^1(\mathscr{A}) = kH_1^1(\mathscr{A}), \qquad k = 1,2,\cdots.$$

However,

$$\begin{aligned} kH_1^1(\mathscr{A}) &= H_k^1(\mathscr{A}) = H\left(\bigvee_{i=1}^k T^{-i}\mathscr{A}/\mathscr{A}\right) \\ &= H\left(\bigvee_{i=1}^{k-1} T^{-i}\mathscr{A}/\mathscr{A}\right) + H\left(T^{-k}\mathscr{A}/\mathscr{A} \vee \bigvee_{i=1}^{k-1} T^{-i}\mathscr{A}\right) \\ &\leqq (k-1)H_1^1(\mathscr{A}) + H\left(T^{-k}\mathscr{A}\Big/\bigvee_{i=0}^{k-1} T^{-i}\mathscr{A}\right), \end{aligned}$$

i.e.,

$$H\left(T^{-k}\mathscr{A}\Big/\bigvee_{i=0}^{k-1} T^{-i}\mathscr{A}\right) = H(T^{-1}\mathscr{A}/\mathscr{A}).$$

Therefore by Theorem 8, $(X,\mathscr{B},q,T)$ is a Markov chain.

THEOREM 10. *There is one and only one stationary process* $(X,\mathscr{B},p,T)$ *for which* $p \sim m$ *and* $h_p(T) = \log\beta$.

Proof. We have proved that there is one stationary process $(X,\mathscr{B},p,T)$ for which $p \sim m$ and $h_p(T) = \log\beta$. Suppose that $q \sim m$ and $h_q(T) = \log\beta$; then $(X,\mathscr{B},q,T)$ is a stationary Markov chain and since every transitive Markov chain is ergodic, $(X,\mathscr{B},q,T)$ is ergodic. But $(p+q)/2 \sim m$ and $(p+q)/2$ is stationary. Moreover

$$\begin{aligned} h_{(p+q)/2}(T) &= \lim_{n\to\infty} -\frac{1}{n}\sum_{C_n(x)} \frac{p+q}{2}C_n(x)\log\frac{p+q}{2}C_n(x) \\ &\geqq \lim_{n\to\infty} -\frac{1}{2n}\sum_{C_n(x)} (pC_n(x)\log pC_n(x) + qC_n(x)\log qC_n(x)) \\ &= \log\beta \text{ (by the convexity of } x\log x). \end{aligned}$$

Consequently $h_{(p+q)/2}(T) = \log\beta$ and $(X, \mathscr{B}, (p+q)/2, T)$ is a stationary Markov chain. Hence T is ergodic with respect to $(p+q)/2$. Either $p = q$, and there is nothing to prove, or p is singular with respect to q. In the latter case there exist invariant sets E, F such that $E \cup F = X$ and $E \cap F = \varnothing$ and $p(E) = 1$, $q(E) = 0$, $p(F) = 0$, $q(F) = 1$. Consequently $(p+q)(E)/2 = \frac{1}{2}$ and $(p+q)(F)/2 = \frac{1}{2}$ and this contradicts the ergodicity of T with respect to $(p+q)/2$.

REMARK. We have shown that every piecewise linear process is equivalent to a Markov chain, and to every set of absolute probabilities $p(0), \cdots, p(s-1)$ there corresponds a unique piecewise linear process. By setting $p(i)$, say, close to one and the remaining probabilities close to zero, one can show that piecewise linear processes with entropies arbitrarily close to zero can be constructed. By continuity considerations, therefore, we can conclude that for every real number $0 < h \leqq \log\beta$, there exists a stationary ergodic Markov chain (equivalent to a piecewise linear process) with h as its entropy.

REFERENCES

1. C. R. Putnam, *On bounded matrices with non-negative elements*, Canad. J. Math. **10** (1958), 587–591.

2. A. I. Khinchin, *Mathematical foundations of information theory*, Dover, New York, 1957.

3. P. R. Halmos, *Entropy in ergodic theory*, University of Chicago, Lecture Notes (mimeographed), 1959.

4. J. Wolfowitz, *Coding theorems of information theory*, Ergebnisse der Math. Grenz., Hefte 31, Springer-Verlag, Berlin, 1961.

YALE UNIVERSITY,
NEW HAVEN, CONNECTICUT
UNIVERSITY OF BIRMINGHAM,
BIRMINGHAM, ENGLAND