# INFORMATION FLOW IN ONE-DIMENSIONAL MARKOV SYSTEMS ${ }^{1}$ 

D. A. DAWSON


#### Abstract

The information flow in discrete Markov systems provides a method for determining that such a system has a unique invariant measure. Estimates are obtained for the information flow and conditions under which there is a unique invariant measure for a one-dimensional Markov system are obtained.


1. Introduction. Markov chains on the infinite product of discrete spaces have been introduced by O. N. Stavskaya and I. I. PyatetskiiShapiro [4] to model certain parts of the central nervous system. One of the main problems in the study of such systems is the determination of the invariant probability measures. The concept of information flow in such a system has been introduced by the author in [2] and its relationship with the uniqueness of the invariant measure has been demonstrated.
2. Terminology. Let $A$ be a countable set, $S$ a finite set and $\Gamma=S^{A}$. $\Gamma$ will serve as the state space for the class of Markov chains to be discussed. Let $\mathscr{G}$ be the $\sigma$-algebra of subsets of $\Gamma$ generated by the coordinate functions and for $B \subset A$, let $\mathscr{G}_{B}$ be the $\sigma$-algebra generated by the coordinate functions in $B$. A mapping $P: \Gamma \otimes \mathscr{G} \rightarrow[0,1]$ is a probability transition kernel if for each $\gamma \in \Gamma, P(\gamma, \cdot)$ is a probability measure on $\mathscr{G}$ and for every $G \in \mathscr{G}, P(\cdot, G)$ is a $\mathscr{G}$-measurable function.

For each $\alpha \in A$, let $N(\alpha)$ be a finite subset of $A$ whose elements are called the nearest neighbours of $\alpha$. The kernel $P(\cdot, \cdot)$ is said to be synchronous if for every finite subset $B \subset A$ and $x_{\alpha} \in S, \alpha \in B$,

$$
\begin{equation*}
P\left(y,\left\{\gamma: \gamma(\alpha)=x_{\alpha}, \alpha \in B\right\}\right)=\prod_{\alpha \in B} P\left(y,\left\{\gamma: \gamma(\alpha)=x_{\alpha}\right\}\right), \tag{2.1}
\end{equation*}
$$

and local if for each fixed $\alpha$ and $x_{\alpha} \in S, P\left(\cdot,\left\{\gamma: \gamma(\alpha)=x_{\alpha}\right\}\right)$ is measurable with respect to $\mathscr{G}_{N^{\prime}(\alpha)}$ and hence

$$
\begin{equation*}
P\left(y,\left\{\gamma: \gamma(\alpha)=x_{\alpha}\right\}\right)=q_{\alpha}\left(\pi_{N(\alpha)} y ; x_{\alpha}\right) \tag{2.2}
\end{equation*}
$$

Received by the editors October 20, 1972 and, in revised form, May 24, 1973 and July 24, 1973.

AMS (MOS) subject classifications (1970). Primary 60J10; Secondary 60K35.
Key words and phrases. Discrete Markov system, invariant measure, information flow.
${ }^{1}$ This research was supported by the National Research Council of Canada.
where $q_{\alpha}(\cdot ; \cdot)$ is a probability transition function from $S^{N(\alpha)}$ to $S$ and $\pi_{N(\alpha)}$ is the projection from $\Gamma$ to $S^{N(\alpha)}$.

Let $\Omega \equiv \Gamma^{Z^{+}}$where $Z^{+} \equiv\{1,2,3, \cdots\}$, let $\mathscr{F}$ denote the product $\sigma$-algebra of subsets of $\Omega$ and for $n=1,2,3, \cdots$, let $X_{n}(\omega)$ denote the $n$th coordinate of $\omega$. Given an initial measure $m$ on $\Gamma$ and the kernel $P(\cdot, \cdot)$ we can construct a measure $P_{m}$ on $(\Omega, \mathscr{F})$ such that $\left\{X_{n}\right\}$ forms a Markov chain with transition probabilities given by $P(\cdot, \cdot)$.

Consider finite subalgebras $\mathscr{F}_{1}, \mathscr{F}_{2}, \mathscr{F}_{3}$ and $\mathscr{F}_{4}$ of $\mathscr{F}$ and let $\mathscr{A}\left(\mathscr{F}_{i}\right)$ denote the set of atoms of $\mathscr{F}_{i}$. The entropy of $\mathscr{F}_{i}$ is defined by

$$
\begin{equation*}
I\left(\mathscr{F}_{i}\right) \equiv-\sum_{A \in \mathscr{Q}\left(\mathscr{F}_{i}\right)} P_{m}(A) \log \left(P_{m}(A)\right) \tag{2.3}
\end{equation*}
$$

where all logarithms are taken to base two. The conditional entropy in $\mathscr{F}_{1}$ given $\mathscr{F}_{2}$ is defined by

$$
\begin{equation*}
I\left(\mathscr{F}_{1} \mid \mathscr{F}_{2}\right) \equiv-\sum_{B \in \mathscr{A}\left(\mathscr{F}_{2}\right)} \sum_{A \in \mathscr{A}\left(\mathscr{F}_{1}\right)} P_{m}(A \cap B) \log \left(P_{m}(A \mid B)\right) \tag{2.4}
\end{equation*}
$$

The mutual information in the algebras $\mathscr{F}_{1}$ and $\mathscr{F}_{2}$ is defined by

$$
\begin{align*}
& I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2}\right) \\
& \quad \equiv I\left(\mathscr{F}_{1}\right)-I\left(\mathscr{F}_{1} \mid \mathscr{F}_{2}\right)  \tag{2.5}\\
& \\
& \quad=\sum_{B \in \mathscr{Q}\left(\mathscr{F}_{2}\right)} \sum_{A \in \mathscr{Q}\left(\mathscr{F}_{1}\right)} P_{m}(A \cap B) \log \left[P_{m}(A \cap B) / P_{m}(A) P_{m}(B)\right]
\end{align*}
$$

and the mutual information in $\mathscr{F}_{1}, \mathscr{F}_{2}$ and $\mathscr{F}_{3}$ is defined by

$$
\begin{equation*}
I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \wedge \mathscr{F}_{3}\right) \equiv I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2}\right)-I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \mid \mathscr{F}_{3}\right) . \tag{2.6}
\end{equation*}
$$

Note that $I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2}\right)=I\left(\mathscr{F}_{2} \wedge \mathscr{F}_{1}\right)$ and $I\left(\mathscr{F}_{i} \wedge \mathscr{F}_{j} \wedge \mathscr{F}_{k}\right)$ is symmetric in $i, j, k$ (refer to N. Abramson [1]).

We also define conditional mutual information as follows:

$$
\begin{align*}
I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \mid \mathscr{F}_{3}\right) & \equiv I\left(\mathscr{F}_{1} \mid \mathscr{F}_{3}\right)-I\left(\mathscr{F}_{1} \mid \mathscr{F}_{2} \vee \mathscr{F}_{3}\right) \\
& =\sum_{C \in \mathscr{A}\left(\mathscr{F}_{3}\right)} P_{m}(C) \sum_{B \in \mathscr{A}\left(\mathscr{F}_{2}\right)} \sum_{A \in \mathscr{P}\left(\mathscr{F}_{1}\right)} P_{m}(A \cap B \mid C)  \tag{2.7}\\
& \cdot \log \left[P_{m}(A \cap B \mid C) / P_{m}(A \mid C) P_{m}(B \mid C)\right]
\end{align*}
$$

$$
\begin{align*}
I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \wedge \mathscr{F}_{3} \mid \mathscr{F}_{4}\right) &  \tag{2.8}\\
& \equiv I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \mid \mathscr{F}_{4}\right)-I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \mid \mathscr{F}_{3} \vee \mathscr{F}_{4}\right)
\end{align*}
$$

where $\mathscr{F}_{2} \vee \mathscr{F}_{3}$ denotes the smallest algebra containing both $\mathscr{F}_{2}$ and $\mathscr{F}_{3}$. Note that $I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2}\right) \geqq 0$ and $I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \mid \mathscr{F}_{3}\right) \geqq 0$ but that

$$
I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \wedge \mathscr{F}_{3}\right) \quad \text { and } \quad I\left(\mathscr{F}_{1} \wedge \mathscr{F}_{2} \wedge \mathscr{F}_{3} \mid \mathscr{F}_{4}\right)
$$

can either be positive or negative.

Now let $\Omega^{\prime} \equiv\{0,1\} \otimes \Omega$, let $X_{0}\left(\omega^{\prime}\right)$ denote the first coordinate of $\omega^{\prime}$ and let $\mathscr{F}_{0} \equiv \sigma\left(X_{0}\right)$. Given a probability $(p, 1-p)$ on $\{0,1\}$ and two measures $m_{0}$ and $m_{1}$ on ( $\Gamma, \mathscr{G}$ ) we can construct a measure $P$ on $\left(\Omega^{\prime}, \mathscr{F}_{0} \otimes \mathscr{F}\right)$ such that $\left\{X_{n}\right\}$ is a Markov chain with

$$
P\left(X_{0}=0\right)=p
$$

and

$$
P\left(X_{1} \in G \mid X_{0}=0\right)=m_{0}(G), \quad P\left(X_{1} \in G \mid X_{0}=1\right)=m_{1}(G)
$$

and such that $X_{1}, X_{2}, X_{3}, \cdots$ is a time homogeneous Markov chain with transition probabilities given by the kernel $P(\cdot, \cdot)$.

Given $J \subset A$, let $\mathscr{F}_{J}^{n} \equiv \sigma\left\{X_{n, \alpha}: \alpha \in J\right\}$ where for $\omega \in \Omega^{\prime}, n \geqq 1, \alpha \in A$, $X_{n, \alpha}$ is defined to be the $\alpha$-coordinate of $X_{n}(\omega) \in S^{A}$. For fixed $\alpha$, consider $\mathscr{F}_{N(\alpha)}^{1}$ as the input and $\mathscr{F}_{\{\alpha\}}^{2}$ as the output of a communications channel. The transmission coefficient of the channel is defined to be

$$
\begin{equation*}
\rho(\alpha) \equiv \operatorname{Sup} I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{\alpha)}^{2}\right) / I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(\alpha)}^{1}\right) \tag{2.9}
\end{equation*}
$$

where the supremum is taken over all choices of $p, m_{0}$ and $m_{1}$. Because of the Markov property $I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{\alpha\}}^{2} \mid \mathscr{F}_{N(\alpha)}^{1}\right)=0$ and therefore

$$
\begin{aligned}
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{\alpha\}}^{2}\right) & =I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(\alpha)}^{1} \wedge \mathscr{F}_{\{\alpha\}}^{2}\right) \\
& =I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(\alpha)}^{1}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(\alpha)}^{1} \mid \mathscr{F}_{\{\alpha\}}^{2}\right) .
\end{aligned}
$$

Since $I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(\alpha)}^{1} \mid \mathscr{F}_{\{\alpha\}}^{2}\right) \geqq 0$,

$$
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{\alpha\}}^{2}\right) \leqq I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(\alpha)}^{1}\right)
$$

and hence $\rho(\alpha) \leqq 1$. We call the system strictly stochastic if $\rho(\alpha)<1$. It has been shown in [1] that the transmission coefficient for the binary symmetric channel is strictly less than one and it is reasonable to conjecture that $\rho<1$ for a large class of channels. The concept of transmission coefficient was first considered by A. Renyi in some of his unpublished work.
3. Information flow in Markov systems. Consider a local synchronous Markov system with $\rho(\alpha) \leqq \rho<1$ for all $\alpha \in A$. Let $J=\left\{\alpha_{0}, \cdots, \alpha_{m}\right\}$ be a finite subset of $A$. Then

$$
\begin{aligned}
I\left(\mathscr{F}_{J}^{n+1} \wedge \mathscr{F}_{0}\right)= & I\left(\mathscr{F}_{\left\{\alpha_{0}\right\}}^{n+1} \wedge \mathscr{F}_{0}\right)+I\left(\mathscr{F}_{\left\{\alpha_{1}\right\}}^{n+1} \wedge \mathscr{F}_{0} \mid \mathscr{F}_{\left\{\alpha_{0}\right\}}^{n+1}\right) \\
& +\cdots+I\left(\mathscr{F}_{\left\{\alpha_{m}\right\}}^{n+1} \wedge \mathscr{F}_{0} \mid \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{m-1}\right\}}^{n+1}\right) .
\end{aligned}
$$

Lemma 3.1. Let $B_{1}=\left\{\alpha_{1}, \cdots, \alpha_{r}\right\}$ and $B_{2}=\left\{\beta_{1}, \cdots, \beta_{k}\right\}$ be finite (possibly empty) subsets of $A, \alpha_{0} \in A, \alpha_{0} \notin B_{1}$ and $n \geqq 1$. Then

$$
\begin{equation*}
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0}\right\}}^{n+1} \mid \mathscr{F}_{B_{1}}^{n+1} \vee \mathscr{F}_{B_{2}}^{n}\right) \leqq \rho I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N\left(\alpha_{0}\right)}^{n} \mid \mathscr{F}_{B_{1}}^{n+1} \vee \mathscr{F}_{B_{2}}^{n}\right) \tag{3.1}
\end{equation*}
$$

where we adopt the convention that

$$
\begin{aligned}
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{B}^{m} \mid \mathscr{F}_{B_{1}}^{n+1} \vee \mathscr{F}_{B_{2}}^{n}\right) & =I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{B}^{m} \mid \mathscr{F}_{B_{1}}^{n+1}\right) & & \text { if } B_{2}=\varnothing \\
& =I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{B}^{m}\right) & & \text { if } B_{1}=B_{2}=\varnothing
\end{aligned}
$$

Proof. Given $x_{\alpha_{0}} \in S, x_{\alpha} \in S, \alpha \in B_{1}, y_{\beta} \in S, \beta \in B_{2} \cup N\left(\alpha_{0}\right)$ and $j \in\{0,1\}$, it can be verified by a straightforward calculation involving Equations (2.1) and (2.2) that

$$
\begin{gather*}
P\left(X_{n+1, \alpha_{0}}=x_{\alpha 0} \mid X_{0}=j ; X_{n+1, \alpha}=x_{\alpha}\right. \\
\left.\quad \alpha \in B_{1} ; X_{n, \beta}=y_{\beta}, \beta \in B_{2} \cup N\left(\alpha_{0}\right)\right)  \tag{3.2}\\
=q_{\alpha 0}\left(y_{\beta}, \beta \in N\left(\alpha_{0}\right) ; x_{\alpha 0}\right) .
\end{gather*}
$$

This implies that conditioned on $\mathscr{F}_{N\left(\alpha_{0}\right)}^{n}, \mathscr{F}_{\left\{\alpha_{0}\right\}}^{n+1}$ is independent of

$$
\mathscr{F}_{B_{1}}^{n+1} \vee \mathscr{F}_{B_{2}-N\left(\alpha_{0}\right)}^{n} \vee \mathscr{F}_{0}
$$

In view of (2.7) it suffices to show that

$$
\begin{align*}
& I\left(\mathscr{F}_{\{\alpha 0\}}^{n+1} \wedge \mathscr{F}_{0} \mid X_{n+1, \alpha}=x_{\alpha}, \alpha \in B_{1} ; X_{n, \beta}=y_{\beta}, \beta \in B_{2}\right) \\
& \quad \leqq \rho\left(\alpha_{0}\right) I\left(\mathscr{F}_{N\left(\alpha_{0}\right)}^{n} \wedge \mathscr{F}_{0} \mid X_{n+1, \alpha}=x_{\alpha}, \alpha \in B_{1} ; X_{n, \beta}=y_{\beta}, \beta \in B_{2}\right) . \tag{3.3}
\end{align*}
$$

But (3.2) demonstrates that even conditioned on the event

$$
\left\{X_{n+1, \alpha}=x_{\alpha}, \alpha \in B_{1} ; X_{n, \beta}=y_{\beta}, \beta \in B_{2}\right\}
$$

the channel from $\mathscr{F}_{N\left(\alpha_{0}\right)}^{n}$ to $\mathscr{F}_{\left\{\alpha_{0}\right\}}^{n+1}$ is defined by $q_{\alpha_{0}}(\cdot ; \cdot)$ and hence is identical to the channel from $\mathscr{F}_{N\left(\alpha_{0}\right)}^{1}$ to $\mathscr{F}_{\left\{\alpha_{0}\right\}}^{2}$. If we then consider the definition (2.9) of $\rho\left(\alpha_{0}\right)$ with the roles of $m_{0}$ and $m_{1}$ in that definition being played by

$$
P\left(X_{n} \in \cdot \mid X_{n+1, \alpha}=x_{\alpha}, \alpha \in B_{1} ; X_{n, \beta}=y_{\beta}, \beta \in B_{2} ; X_{0}=j\right), \quad j=0,1
$$

then (3.3) follows and the proof is complete.
For future use we now derive some basic inequalities.

$$
\begin{align*}
& I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N\left(\alpha_{r}\right)}^{n} \mid \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{r-1}\right\}}^{n+1}\right) \\
&= I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N\left(\alpha_{r}\right)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N\left(\alpha_{r}\right)}^{n} \wedge \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{r-1}\right\}}^{n+1}\right) \\
&= I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N\left(\alpha_{r}\right)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{x_{0}, \cdots, \alpha_{r-1}\right\}}^{n+1}\right)  \tag{3.4}\\
&+I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{r-1}\right\}}^{n+1} \mid \mathscr{F}_{N\left(\alpha_{r}\right)}^{n}\right) .
\end{align*}
$$

But

$$
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha r-1\right\}}^{n+1} \mid \mathscr{F}_{N(\alpha r)}^{n} \vee \mathscr{F}_{\left(N\left(\alpha_{0}\right) \cup \cup \cup N\left(\alpha_{r-1}\right)\right)}^{n}\right)=0
$$

and therefore

$$
\begin{align*}
& I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{r-1}\right\}}^{n+1} \mid \mathscr{F}_{N\left(\alpha_{r}\right)}^{n}\right) \\
&=I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0} \cdots, \alpha_{r-1}\right\}}^{n+1} \wedge \mathscr{F}_{\left(N\left(\alpha_{0}\right) \cup \cdots \cup N\left(\alpha_{r-1}\right)\right)}^{n} \mid \mathscr{F}_{N\left(\alpha_{r}\right)}^{n}\right)  \tag{3.5}\\
& \leqq I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left(N\left(x_{0}\right) \cup \cdots \cup N\left(\alpha_{r-1}\right)\right)}^{n} \mid \mathscr{F}_{N(\alpha r)}^{n}\right) .
\end{align*}
$$

Therefore from (3.4) and (3.5),

$$
\begin{align*}
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N\left(\alpha_{r}\right)}^{n}\right. & \left.\mid \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{r-1}\right\}}^{n+1}\right) \\
& \leqq I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left.\left(N\left(\alpha_{0}\right) \cup \cdots \cup_{N\left(\alpha_{r}\right)}^{n}\right)\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{r-1}\right\}}^{n+1}\right)}\right. \tag{3.6}
\end{align*}
$$

In the remainder of this section we restrict our attention to the case in which $A=Z^{1}$, the set of integers. For simplicity we assume that the local transition probabilities are the same at each site and we restrict our attention to measures $m_{0}$ and $m_{1}$ which are invariant under the spatial shift on $A$. Under this assumption but without assuming that $m_{0}$ and $m_{1}$ are invariant under reflections we have the following result.

Lemma 3.2. If $m_{0}$ and $m_{1}$ are invariant under the spatial shift on $Z^{1}$, then

$$
I\left(\mathscr{F}_{\left\{\alpha_{m}, \cdots, \alpha_{m+r}\right\}}^{n} \wedge \mathscr{F}_{0} \mid \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{m-1}\right\}}^{n}\right)=I\left(\mathscr{F}_{\left\{\alpha-r, \cdots, \alpha_{0}\right\}}^{n} \wedge \mathscr{F}_{0} \mid \mathscr{F}_{\left\{\alpha_{1}, \cdots, \alpha_{m}\right\}}^{n}\right) .
$$

Proof.

$$
\begin{aligned}
& I\left(\mathscr{F}_{\left\{\alpha-r, \cdots, \alpha_{m+r}\right\}}^{n} \wedge \mathscr{F}_{0}\right) \\
& \quad=I\left(\mathscr{F}_{\left\{\alpha-r, \cdots, \alpha_{m-1}\right\}}^{n} \wedge \mathscr{F}_{0}\right)+\left(\mathscr{F}_{\left\{\alpha_{m}, \cdots, \alpha_{m+r}\right\}}^{n} \wedge \mathscr{F}_{0} \mid \mathscr{F}_{\left\{\alpha-r, \cdots, \alpha_{m-1}\right\}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
& I\left(\mathscr{F}_{\left\{\alpha-r, \cdots, \alpha_{m+r}\right\}}^{n} \wedge \mathscr{F}_{0}\right) \\
& \quad=I\left(\mathscr{F}_{\left\{\alpha_{1}, \cdots, \alpha_{m+r}\right\}}^{n} \wedge \mathscr{F}_{0}\right)+I\left(\mathscr{F}_{\left\{\alpha-r, \cdots, \alpha_{0}\right\}}^{n} \wedge \mathscr{F}_{0} \mid \mathscr{F}_{\left\{\alpha_{1}, \cdots, \alpha_{m+r}\right\}}^{n}\right)
\end{aligned}
$$

Hence the result follows since by spatial invariance

$$
I\left(\mathscr{F}_{\left\{\alpha-r, \cdots, \alpha_{m-1}\right\}}^{n} \wedge \mathscr{F}_{0}\right)=I\left(\mathscr{F}_{\left\{\alpha_{1}, \cdots, \alpha_{m+r}\right\}}^{n} \wedge \mathscr{F}_{0}\right) .
$$

We now assume that for each $\alpha_{r}, N\left(\alpha_{r}\right)=\left\{\alpha_{r-s}, \cdots, \alpha_{r-1}, \alpha_{r}, \alpha_{r+1}, \cdots\right.$, $\left.\alpha_{r+s}\right\}$. For $k=0,1,2, \cdots$, let

$$
F_{n}(k) \equiv I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{k}\right\}}^{n}\right)
$$

A measure $m$ is a stationary probability for the Markov system if, for $G \in \mathscr{G}, m(G)=\int P(\gamma, G) m(d \gamma)$.

We now calculate the amount of information about the initial state which is contained in a finite subsystem at the $n$th epoch.

Proposition 3.1. Let $m_{0}$ and $m_{1}$ be two distinct spatially invariant probability measures. Then
$F_{n+1}(k) \leqq \rho F_{n}(k+2 s)+\rho(1-\rho) F_{n}(k+2 s-1)+\cdots+\rho(1-\rho)^{k} F_{n}(2 s)$.
Proof. By Lemma 3.1 we have

$$
F_{n+1}(0) \leqq \rho I\left(\mathscr{F}_{N(0)}^{n} \wedge \mathscr{F}_{0}\right)=\rho F_{n}(2 s)
$$

since $N(0)=\{-s, \cdots, 0, \cdots, s\}$ and we have assumed spatial invariance. Similarly, for $k \geqq 1$, Lemma 3.1 and (3.6) imply that

$$
\begin{aligned}
F_{n+1}(k)-F_{n+1}(k-1) & =I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{k\}}^{n+1} \mid \mathscr{F}_{\{0, \cdots, k-1\}}^{n+1}\right) \\
& \leqq \rho\left[I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left(N(0) \cup \cdots \cup_{N(k)}^{n}\right)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0, \cdots, k-1\}}^{n+1}\right)\right]
\end{aligned}
$$

Hence since $N(0) \cup \cdots \cup N(k)=\{-s, \cdots, 0, \cdots, s+k\}$,

$$
\begin{equation*}
F_{n+1}(k)-F_{n+1}(k-1)=\rho\left[F_{n}(k+2 s)-F_{n+1}(k-1)\right] . \tag{3.7}
\end{equation*}
$$

But Equation (3.7) can be rewritten as

$$
\begin{equation*}
F_{n+1}(k) \leqq \rho F_{n}(k+2 s)+(1-\rho) F_{n+1}(k-1) \tag{3.8}
\end{equation*}
$$

Then, by induction
$F_{n+1}(k) \leqq \rho F_{n}(k+2 s)+\rho(1-\rho) F_{n}(k+2 s-1)+\cdots+\rho(1-\rho)^{k} F_{n}(2 s)$.
Using the ideas of Proposition 3.1 we now obtain a stronger inequality for the information flow.

Proposition 3.2.

$$
\begin{equation*}
F_{n+1}(k) \leqq \rho H_{n}(k)+\rho(1-\rho) H_{n}(k-1)+\cdots+\rho(1-\rho)^{k} H_{n}(0) \tag{3.9}
\end{equation*}
$$

where

$$
\begin{align*}
H_{n}(k)= & (1-\rho)^{k+1} F_{n}(2 s)  \tag{3.10}\\
& +\left\{\rho(1-\rho)^{k} F_{n}(2 s)+\rho(1-\rho)^{k-1} F_{n}(2 s+1)+\cdots+\rho F_{n}(2 s+k)\right\}
\end{align*}
$$

Proof. From Lemma 3.1 and (3.4) we have for $k \geqq 1$,

$$
\begin{align*}
& F_{n+1}(k)-F_{n+1}(k-1)=I\left(\mathscr{F}_{\{k\}}^{n+1} \wedge \mathscr{F}_{0} \mid \mathscr{F}_{\{0, \cdots, k-1\}}^{n+1}\right)  \tag{3.11}\\
& \quad \leqq \rho\left[I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(k)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0, \cdots, k-1\}}^{n+1}\right)+I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0, \cdots, k-1\}}^{n+1} \mid \mathscr{F}_{N(k)}^{n}\right)\right]
\end{align*}
$$

For $r=0,1,2, \cdots$, let

$$
G_{n+1}(r) \equiv I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0, \cdots, r\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n}\right) .
$$

Then
(3.12) $\quad F_{n+1}(k) \leqq(1-\rho) F_{n+1}(k-1)+\rho\left[F_{n}(2 s)+G_{n+1}(k-1)\right]$.

By Lemma 3.2

$$
G_{n+1}(0)=I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0\}}^{n+1} \mid \mathscr{F}_{N(1)}^{n}\right)=I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{1\}}^{n+1} \mid \mathscr{F}_{N(0)}^{n}\right) .
$$

But by Lemma 3.1,

$$
\begin{gathered}
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{1\}}^{n+1} \mid \mathscr{F}_{N(0)}^{n}\right) \leqq \rho I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(1)}^{n} \mid \mathscr{F}_{N(0)}^{n}\right) \\
=\rho\left(F_{n}(2 s+1)-F_{n}(2 s)\right)
\end{gathered}
$$

and, hence,

$$
\begin{equation*}
G_{n+1}(0) \leqq \rho\left(F_{n}(2 s+1)-F_{n}(2 s)\right) \tag{3.13}
\end{equation*}
$$

For $r \geqq 1$,

$$
\begin{aligned}
& G_{n+1}(r)-G_{n+1}(r-1) \\
&=I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0, \cdots, r\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0, \cdots, r-1\}}^{n+1} \mid \mathscr{F}_{N(r)}^{n}\right) \\
&=I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0, \cdots, r\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{1, \cdots, r\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n}\right) \\
&=I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{0\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n} \vee \mathscr{F}_{\{1, \cdots, r\}}^{n+1}\right) .
\end{aligned}
$$

But Lemma 3.1 implies that

$$
\begin{equation*}
G_{n+1}(r)-G_{n+1}(r-1) \leqq \rho I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(0)}^{n} \mid \mathscr{F}_{N(r+1)}^{n} \vee \mathscr{F}_{\{1, \cdots, r\}}^{n+1}\right) \tag{3.14}
\end{equation*}
$$

But following the method of derivation of Equations (3.4) and (3.5),

$$
\begin{aligned}
& I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(0)}^{n} \mid \mathscr{F}_{N(r+1)}^{n} \vee \mathscr{F}_{\{1, \cdots, r\}}^{n+1}\right) \\
&= I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(0)}^{n} \mid \mathscr{F}_{N(r+1)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(0)}^{n} \wedge \mathscr{F}_{\{1, \cdots, r\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n}\right) \\
&= I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(0)}^{n} \mid \mathscr{F}_{N(r+1)}^{n}\right) \\
&+I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{1, \cdots, r\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n} \vee \mathscr{F}_{N(0)}^{n}\right)-I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\{1, \cdots, r\}}^{n+1} \mid \mathscr{F}_{N(r+1)}^{n}\right) \\
& \leqq I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{N(0)}^{n} \mid \mathscr{F}_{N(r+1)}^{n}\right) \\
& \quad+I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{(N(1) \cup \cdots \cup N(r))}^{n} \mid \mathscr{F}_{N(r+1)}^{n} \vee \mathscr{F}_{N(0)}^{n}\right)-G_{n+1}(r-1) \\
&= I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left(N(0) \cup \cdots \cup_{N(r))}^{n} \mid \mathscr{F}_{N(r+1)}^{n}\right)-G_{n+1}(r-1)}^{=}\right. \\
& F_{n}(r+2 s+1)-F_{n}(2 s)-G_{n+1}(r-1) .
\end{aligned}
$$

Therefore combining the last result with (3.14), we have

$$
\begin{equation*}
G_{n+1}(r) \leqq(1-\rho) G_{n+1}(r-1)+\rho\left(F_{n}(r+2 s+1)-F_{n}(2 s)\right) \tag{3.15}
\end{equation*}
$$

Hence by induction

$$
\begin{align*}
G_{n+1}(r) \leqq & \rho\left(F_{n}(2 s+r+1)-F_{n}(2 s)\right)  \tag{3.16}\\
& +\cdots+\rho(1-\rho)^{r}\left(F_{n}(2 s+1)-F_{n}(2 s)\right)
\end{align*}
$$

But then, from (3.12), we have for $k \geqq 1$

$$
\begin{align*}
& F_{n+1}(k) \leqq(1-\rho) F_{n+1}(k-1) \\
& \quad+\rho\left[F_{n}(2 s)+\rho\left(F_{n}(2 s+k)-F_{n}(2 s)\right)\right.  \tag{3.17}\\
&\left.\quad \quad \quad+\cdots+\rho(1-\rho)^{k-1}\left(F_{n}(2 s+1)-F_{n}(2 s)\right)\right]
\end{align*}
$$

Hence $F_{n+1}(0) \leqq \rho F_{n}(2 s)$ and for $k \geqq 1$,

$$
\begin{equation*}
F_{n+1}(k) \leqq(1-\rho) F_{n+1}(k-1)+\rho H_{n}(k) \tag{3.18}
\end{equation*}
$$

where

$$
\begin{aligned}
H_{n}(k)= & {\left[\rho F_{n}(2 s+k)+\cdots+\rho(1-\rho)^{k-1} F_{n}(2 s+1)+\rho(1-\rho)^{k} F_{n}(2 s)\right] } \\
& +(1-\rho)^{k+1} F_{n}(2 s) .
\end{aligned}
$$

Using induction and (3.18) it can be shown that, for $k \geqq 0$,

$$
F_{n+1}(k) \leqq \rho H_{n}(k)+\rho(1-\rho) H_{n}(k-1)+\cdots+\rho(1-\rho)^{k} H_{n}(0)
$$

Corollary. If $L_{1}(k) \equiv F_{1}(k)$ for each $k \geqq 0$ and if $L_{n}(k)$ is defined recursively by $L_{n+1}(0)=\rho L_{n}(2 s)$, and, for $k \geqq 1$,

$$
\begin{align*}
L_{n+1}(k)= & (1-\rho) L_{n+1}(k-1) \\
& +\rho\left[(1-\rho)^{k+1} L_{n}(2 s)\right.  \tag{3.19}\\
& \left.\quad+\left\{\rho(1-\rho)^{k} L_{n}(2 s)+\cdots+\rho L_{n}(2 s+k)\right\}\right]
\end{align*}
$$

then for all $n$ and $k, F_{n}(k) \leqq L_{n}(k)$.
Proof. The proof follows by induction from equation (3.17).
Proposition 3.3. Consider a one-dimensional synchronous Markov system with $N\left(\alpha_{k}\right)=\left\{\alpha_{k-s}, \cdots, \alpha_{k}, \cdots, \alpha_{k+s}\right\}$. If $\rho \leqq 1 /(s+1)$, then for each $k$

$$
F_{n}(k)=I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{\left\{\alpha_{0}, \cdots, \alpha_{k}\right\}}^{n}\right) \rightarrow 0 \quad \text { as } n \rightarrow \infty .
$$

Proof. In view of the corollary to Proposition 3.2 it suffices to show that, for each $k, L_{n}(k) \rightarrow 0$ as $n \rightarrow \infty$. Since $\lim _{k \rightarrow \infty} L_{1}(k) \leqq 1$ and the recursive equations (3.19) are linear we can assume without loss of generality that $\lim _{k \rightarrow \infty} L_{1}(k)=1$ so that $L_{1}(\cdot)$ can be identified with the cumulative distribution function of a discrete random variable $W_{1}$. Let $Y_{1}, Y_{2}, Y_{3}, \cdots$, and $Z_{1}, Z_{2}, Z_{3}, \cdots$, be two sequences of independent discrete random variables with distribution given by $P\left(Y_{1}=k\right)=\rho(1-\rho)^{k}$ for $k=0,1$, $2, \cdots 8$, and define the sequence of random variables $\left\{W_{n}\right\}$ as follows:
$W_{1}$ is defined as above; if $W_{n} \leqq 2 s$, then $W_{n+1}=Y_{n}$; if $W_{n}=k>2 s$, then $W_{n+1}=k-2 s+Y_{n}+Z_{n}$. In other words,

$$
W_{n+1}=\chi_{\left\{W_{n} \leqq 2 s\right\}} Y_{n}+\chi_{\left\{W_{n}>2 s\right\}}\left(W_{n}-2 s+Y_{n}+Z_{n}\right),
$$

where $\chi_{G}$ denotes the indicator function of the event $G$. Now it is clear that $\left\{W_{n}\right\}$ is a Markov chain with state space $\{0,1,2, \cdots\}$. Let $E_{n}(k) \equiv$ $P\left(W_{n} \leqq k\right)$. Then

$$
\begin{aligned}
P\left(W_{n+1} \leqq k\right)= & P\left(W_{n} \leqq 2 s\right) P\left(Y_{n} \leqq k\right) \\
& +\sum_{l=2 s+1}^{2 s+k} P\left(2 s<W_{n} \leqq l\right) P\left(Y_{n}+Z_{n}=k+2 s-l\right)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
E_{n+1}(k)= & \rho \sum_{i=0}^{k}(1-\rho)^{i} E_{n}(2 s) \\
& +\sum_{l=2 s+1}^{2 s+k}\left(E_{n}(l)-E_{n}(2 s)\right) \rho^{2}(k+2 s-l+1)(1-\rho)^{k+2 s-l}
\end{aligned}
$$

since $P\left(Y_{n}+Z_{n}=r\right)=\rho^{2}(r+1)(1-\rho)^{r}$ for $r=0,1,2, \cdots$. But then

$$
\begin{aligned}
& E_{n+1}(k)-(1-\rho) E_{n+1}(k-1) \\
& =\rho E_{n}(2 s)+\rho^{2}\left(\left(E_{n}(2 s+k)-E_{n}(2 s)\right)\right. \\
& \left.\quad+\cdots+(1-\rho)^{k-1}\left(E_{n}(2 s+1)-E_{n}(2 s)\right)\right)
\end{aligned}
$$

Therefore $E_{n}(k)$ satisfies (3.19) and hence $E_{n}(k)=L_{n}(k)$ for all $n$ and $k$. This means that the Markov chain $\left\{W_{n}\right\}$ with state space $\{0,1,2, \cdots\}$ has no absorbing states and, as long as $W_{n} \in\{2 s+1,2 s+2, \cdots\}$, it is identical to a random walk. But according to a basic result of renewal theory (refer to W. Feller [3, Chapter 2]),
$P\left(W_{n}=k\right) \rightarrow 0$ as $n \rightarrow \infty$ for all $k$ if and only if

$$
E\left(Y_{n}+Z_{n}-2 s\right)=2(1-\rho) / \rho-2 s \geqq 0
$$

Therefore $L_{n}(k) \rightarrow 0$ for all $k$ as $n \rightarrow \infty$ if and only if $\rho \leqq 1 /(s+1)$ and the proof is complete.

Now let $\mu$ be any spatially invariant initial measure for the Markov system and let

$$
\mu_{n}(G) \equiv \int P\left(X_{n} \in G \mid X_{0}=y\right) \mu(d y)
$$

Proposition 3.4. Under the above assumptions, the Markov system has a unique spatially invariant stationary measure if $\rho \leqq 1 /(s+1)$. Moreover for any finite set $B \subset A$ and any spatially invariant measure $\nu$,

$$
v_{n}\left|\mathscr{G}_{B} \rightarrow \mu_{\infty}\right| \mathscr{G}_{B} \quad \text { as } n \rightarrow \infty
$$

where $\mu_{\infty}$ is the spatially invariant stationary measure.
Proof. It has been shown in [2, Proposition 3.1] that at least one such stationary measure $\mu_{\infty}$ exists. Now let $\nu$ be another such stationary
measure. Applying Proposition 3.3 in the case $m_{1}=\mu_{\infty}$ and $m_{2}=\nu$ with the probability $\left(\frac{1}{2}, \frac{1}{2}\right)$ on $\mathscr{F}_{0}$ it follows that, for any finite set $B, I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{B}^{n}\right)$ $\rightarrow 0$ as $n \rightarrow \infty$. But since $\mu_{n}=\mu_{\infty}$ and $v_{n}=v$ this implies that $I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{B}^{1}\right)=0$ and hence $\mathscr{F}_{0}$ and $\mathscr{F}_{B}^{1}$ are independent (cf. N. Abramson [1, p. 107]). But this implies that $v\left|\mathscr{G}_{B}=\mu_{\infty}\right|_{\mathscr{G}_{B}}$ for every finite set $B$ which in turn implies that $\nu=\mu_{\infty}$.

Finally, let $\nu$ be any spatially invariant measure. Applying Proposition (3.3) in the case $m_{1}=\mu_{\infty}$ and $m_{2}=v$ with the probability ( $\frac{1}{2}, \frac{1}{2}$ ) on $\mathscr{F}_{0}$, it follows that, for any finite set $B \subset A$,

$$
\begin{equation*}
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{B}^{n}\right) \rightarrow 0 \quad \text { as } n \rightarrow \infty \tag{3.20}
\end{equation*}
$$

But

$$
\begin{equation*}
I\left(\mathscr{F}_{0} \wedge \mathscr{F}_{B}^{n}\right)=H\left(\frac{1}{2} \mu_{\infty}+\frac{1}{2} v_{n}\right)-\frac{1}{2} H\left(\mu_{\infty}\right)-\frac{1}{2} H\left(v_{n}\right) \tag{3.21}
\end{equation*}
$$

where $H(\mu)=-\sum \mu_{i} \log \mu_{i}$ and on the right-hand side of (3.21) we are considering the restrictions of the measures to $\mathscr{G}_{1 B}$. But $H\left(\frac{1}{2} \mu_{\infty}+\frac{1}{2} v\right)-$ $\frac{1}{2} H\left(\mu_{\infty}\right)-\frac{1}{2} H(\nu)=0$ if and only if $\nu=\mu_{\infty}$, and $H(v)$ is continuous as a


Acknowledgement. The author would like to thank the referee for a number of suggestions which greatly improved the exposition of the paper.

## References

1. N. Abramson, Information theory and coding, McGraw-Hill, New York, 1963. MR 32 \#7308.
2. D. A. Dawson, Information flow in discrete Markov systems, J. Appl. Probability 10 (1973), 63-83.
3. W. Feller, An introduction to probability theory and its applications. Vol. II, Wiley, New York, 1966. MR 35 \#1048.
4. O. N. Stavskaja and I. I. Pjatetskiĭ-Šapiro, Homogeneous networks of spontaneously active elements, Problemy Kibernet. No. 20 (1968), 91-106. (Russian) MR 44 \#3720.

Department of Mathematics, Carleton University, Ottawa, Ontario K1S 5B6 Canada

