APPROXIMATIONS FOR THE ENTROPY FOR FUNCTIONS OF MARKOV CHAINS[1]

By John J. Birch

University of Nebraska

1. Summary. If $\{Y_n\}$ is a stationary ergodic Markov process taking on values in a finite set $\{1, 2, \cdots, A\}$, then its entropy can be calculated directly. If ϕ is a function defined on $1, 2, \cdots, A$, with values $1, 2, \cdots, D$, no comparable formula is available for the entropy of the process $\{Y_n = \phi(Y_n)\}$. However, the entropy of this functional process can be approximated by the monotonic functions $\bar{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1)$ and $\underline{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1, Y_0)$, the conditional entropies. Furthermore, if the underlying Markov process $\{Y_n\}$ has strictly positive transition probabilities, these two approximations converge exponentially to the entropy H, where the convergence is given by $0 \leqq \bar{G}_n - H \leqq B\rho^{n-1}$ and $0 \leqq H - \underline{G}_n \leqq B\rho^{n-1}$ with $0 < \rho < 1$, ρ being independent of the function ϕ.

2. Introduction. Let $\{Y_n, -\infty < n < \infty\}$ be a stationary ergodic stochastic process taking on values in a finite set $\{1, \cdots, D\}$. For any finite sequence $s = (i_1, \cdots, i_k), i_j \varepsilon \{1, \cdots, D\}, j = 1, \cdots, k$, let $p(s) = P\{(Y_1, \cdots, Y_k) = s\}$ and let $Z_k = p(Y_1, \cdots, Y_k)$. In extending Shannon's [1] pioneer work on information theory, McMillan [2] has shown that associated with the sequence $\{Z_k\}$ of random variables is a number H, called the entropy of the process, such that

$$(1/n) \log Z_n \to -H \text{ in } L^1 \text{ as } n \to \infty$$

(where the log has base 2). This result, which is fundamental in information theory, implies that for large n it is highly probable that the sequence of states of length n which actually occurs is one whose probability is about 2^{-nH}. McMillan further shows that, if (the r.v.) $U_i = P\{Y_1 = i \mid Y_0, Y_{-1}, \cdots\}$ and $V = U_i$ on $\{Y_1 = i\}$, then

$$H = -E \log V \tag{1}$$

thus giving an alternate definition of the entropy.

If $\{Y_n\}$ is a Markov process with transition probabilities $m(i, j) = P\{Y_{n+1} = j \mid Y_n = i, Y_{n-1}, \cdots\} i, j = 1, 2, \cdots, A$ and stationary probability distribution $\lambda_i = P\{Y_n = i\}$, then we have $U_i = m(Y_0, i)$ and $V = m(Y_0, Y_1)$, so that

$$H = -\sum_{i,j} \lambda_i m(i, j) \log m(i, j). \tag{2}$$

Received May 8, 1961; revised October 18, 1961.

[1] This research was performed while the author was at the University of California, Berkeley, and was partially supported by the Office of Naval Research under Contract Nonr 222(53).

This provides an easy way to calculate the entropy of a Markov process. On the other hand, if ϕ is a function defined on $1, 2, \cdots, A$ with values $1, 2, \cdots, D$, no comparable formula is available for the entropy of $\{X_n = \phi(Y_n)\}$. Moreover, Blackwell [3] has shown that, if $\{Y_n\}$ is a stationary ergodic finite-state Markov chain, the entropy of the process $\{X_n = \phi(Y_n)\}$ is given by

$$H = -\int \sum_{d=1}^{D} r_d(w) \log r_d(w)\, dQ(w).$$

Here, r_d is a function that is defined on the set W of all $w = (w_1, \cdots, w_A)$ such that $w_i \geqq 0$, $\sum_1^A w_i = 1$ by $r_d(w) = \sum_{i=1}^A w_i \sum_{j \varepsilon \phi^{-1}(d)} m(i, j)$ for those j's such that $\phi(j) = a$. Q is the distribution of the conditional distribution of Y_0 given $X_0, X_{-1}, \cdots$. Furthermore, Blackwell obtains an integral equation for Q and illustrates how Q may be concentrated on a finite or countable set or be continuous. In the latter case, he conjectures that Q is singular, thus suggesting that this entropy is intrinsically a complicated function of the transition matrix M and the function ϕ. Consequently, for application, it is desirable to find approximations for the entropy of this functional process which converge rapidly.

It should be noted that if the functional process $\{X_n\}$ is Markovian[2], then the entropy can be computed directly by Eq. (2). Therefore, we shall consider only cases where $\{X_n\}$ is not Markovian. Moreover, whenever ϕ is a function that collapses only one class of states, Blackwell [3] has shown that the entropy can be expressed as a sum of converging elements; i.e., Q is concentrated on a countable set. Thus we shall consider the more general case where at least two classes of states of the Markov process collapse into single states of the functional process.

In this paper, we show that if $\{Y_n\}$ is a stationary ergodic finite-state Markov process, then the entropy of the process $\{X_n = \phi(Y_n)\}$ can be approximated by the monotonic functions

$$\bar{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1)$$

and

$$\underline{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1, Y_0).$$

Furthermore, if this Markov process $\{Y_n\}$ has strictly positive transition probabilities, these two approximations converge exponentially to the entropy H, where the convergence is given by

$$0 < \bar{G}_n - H \leqq B\rho^{n-1} \quad \text{and} \quad 0 \leqq H - \underline{G}_n \leqq B\rho^{n-1}$$

with $B = [N_D \log e]/[N_1 \min_{i,j} m(i, j)]$ and

$$0 < \rho = 1 - \min_{i,j,k,m,n} \left(\frac{N_1}{N_D^2}\right)\left(\frac{m(i, k)m(k, n)}{m(i, j)m(j, m)}\right)^2 < 1$$

[2] Necessary and sufficient conditions for $\{X_n\}$ to be Markovian are given by Burke and Rosenblatt [4].

where N_1 and N_D are the minimum and maximum number of states collapsed by ϕ, respectively.

Examples are presented in which these functions are used to approximate the rate of a two-state channel for several different input distributions.

3. Notations and definitions. Let Y be a random variable taking on values in the finite space $U = (u_1, \cdots, u_k)$. We define the entropy $h(Y)$ of Y as

$$h(Y) = -\sum_i P\{Y = u_i\} \log P\{Y = u_i\}.$$

We recall for future use some mathematical properties of the h function, given in [1]:

(i) $h(Y) \leqq \log k$,

(ii) $h(X, Y) = h(Y) + h(X \mid Y)$ where $h(X \mid Y) = -\sum_{i,j} P\{Y = u_i, X = v_j\} \log P\{X = v_j \mid Y = u_i\}$,

(iii) $h(X \mid Y, Z) \leqq h(X \mid Y) \leqq h(X)$ with equalities if X, Y, Z are independent,

(iv) for any function ϕ defined on the range of Y, $h(\phi(Y)) \leqq h(Y)$ and $h(X \mid Y) \leqq h(X \mid \phi(Y))$.

We also note that all of the above quantities are non-negative.

For any positive integer N, denote by $I(N)$ the set of integers $\{1, 2, \cdots, N\}$.

Throughout what follows we shall assume that we are given an $A \times A$ Markov matrix M with elements $m(i, j)$, $i, j = 1, 2, \cdots, A$, a function ϕ from $I(A)$ to $I(D)$ and an initial stationary probability distribution $\lambda = (\lambda_1, \cdots, \lambda_A)$ on $I(A)$. Further, $Y_0, Y_1, Y_2, \cdots$ will be a stationary ergodic Markov chain taking on values in $I(A)$ and distributed according to (λ, M). $X_1, X_2, \cdots$ will be the process defined by $X_k = \phi(Y_k)$ and taking on values in $I(D)$. It, of course, will be a stationary ergodic process.

4. Approximations for the entropy of the functional process. First, let us define

$$H_n(M, \phi, \lambda) = (1/n) h(X_1, X_2, \cdots, X_n).$$

If λ is stationary (as we have assumed), then it is shown in [2] that $H_n(M, \phi, \lambda)$ converges monotonically downward to a limiting constant $H(M, \phi, \lambda)$, called the entropy of the $\{X_n\}$ process. In addition, the following result has been given in [5]

$$|H(M, \phi, \lambda) - H_n(M, \phi, \lambda)| \leqq (2/n) \log A.$$

Next, we let $\bar{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1)$, where $\{X_n = \phi(Y_n)\}$ is the functional process. From property (iii) we have that $\bar{G}_1 \geqq \bar{G}_2 \geqq \cdots$, and since $\bar{G}_n \geqq 0$, it follows that $\lim \bar{G}_n = H$ exists. Now using stationarity we have that $\lim (n + 1)^{-1} \sum_1^n \bar{G}_i = H(M, \phi, \lambda)$. Thus $\lim \bar{G}_n = H(M, \phi, \lambda)$.

If we let $\underline{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1, Y_0)$, then we have the following

LEMMA 3.1. *$\underline{G}_n$ converges monotonically upward to the entropy $H(M, \phi, \lambda)$.*

PROOF. To see that $\underline{G}_n$ is monotonic we note that

$$\underline{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1, Y_0, Y_{-1})$$
$$\leqq h(X_n \mid X_{n-1}, \cdots, X_1, \phi(Y_0), Y_{-1}) \qquad \text{by (iv)}$$
$$\underline{G}_n \leqq h(X_n \mid X_{n-1}, \cdots, X_1, X_0, Y_{-1}) = \underline{G}_{n+1} \qquad \text{by stationarity.}$$

Let $p(Y_0 \mid X_1, \cdots, X_n) = P\{Y_0 = i \mid X_1, \cdots\}$ on $\{Y_0 = i\}$. Now making use of (ii) we see that

$$\bar{G}_n - \underline{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1) - h(X_n \mid X_{n-1}, \cdots, X_1, Y_0)$$
$$= E \log \{p(Y_0 \mid X_1, \cdots, X_n)/p(Y_0 \mid X_1, \cdots, X_{n-1})\}.$$

However, for every fixed $(X_1, \cdots, X_n) = t_n$, $Y_0 = i$, the variable $p(Y_0 \mid X_1, \cdots, X_n)$ coincides with one of the $P\{Y_0 = i \mid X_1, \cdots, X_n\}$, where $i \in I(A)$. Thus, for any n and for any $(X_1, \cdots, X_n)$

$$|p(Y_0 \mid X_1, \cdots, X_n) - p(Y_0 \mid X_1, \cdots, X_{n-1})|$$
$$\leqq \sum_i |P\{Y_0 = i \mid X_1, \cdots, X_n\} - P\{Y_0 = i \mid X_1, \cdots, X_{n-1}\}|.$$

But $P\{Y_0 = i \mid X_1, \cdots, X_n\}$ is a martingale and therefore converges a.e. Consequently, $p(Y_0 \mid X_1, \cdots, X_n)$ converges.

Also, $\bar{G}_n \geqq \underline{G}_n$ follows from (iii).

5. Rate of convergence of $\bar{G}_n$ and $\underline{G}_n$. Our result will be for the case where the transition probabilities $m(i, j)$ are strictly positive for all i, j.

Let $Y_0, Y_1, \cdots,$ be a finite-state Markov chain with transition probabilities $m(i, j)$. Define a process $\{X_n\}$ by $X = i \Leftrightarrow Y \in \phi^{-1}(i)$. Now for fixed $X_1 = i_1, \cdots, X_{n-1} = i_{n-1}$, let

$$f_n(g, a) = P\{Y_n = a \mid Y_0 = g, Y_1 \in \phi^{-1}(i_1), \cdots, Y_{n-1} \in \phi^{-1}(i_{n-1})\}.$$

This is the probability of going from state g to a state a in n steps in a non-homogeneous Markov chain with transition probabilities for the kth step given by

$$p^{(k)}(j, l) = P\{Y_k = l \mid Y_{k-1} = j, Y_k \in \phi^{-1}(i_k), \cdots, Y_{n-1} \in \phi^{-1}(i_{n-1})\}$$
$$= \frac{m(j, l)P\{Y_k \in \phi^{-1}(i_k), \cdots, Y_{n-1} \in \phi^{-1}(i_{n-1}) \mid Y_k = l\}}{\sum_{l'} m(j, l')P\{Y_k \in \phi^{-1}(i_k), \cdots, Y_{n-1} \in \phi^{-1}(i_{n-1}) \mid Y_k = l'\}},$$

for $k = 1, \cdots, n-1$ and $p^{(n)}(j, l) = m(j, l)$. This is

$$f_n(g, a) = \sum p^{(1)}(g, a_1)p^{(2)}(a_1, a_2) \cdots p^{(n)}(a_{n-1}, a).$$

We shall make use of the following theorem concerning Markov chains. (It may be noted that a similar result is given by Harris [7].)

THEOREM 4.1. *Let g and h be two states of the Markov chain. If $m(i, j) > 0$ for $i, j = 1, 2, \cdots, A$, then the following holds:*

$$|f_n(g, a) - f_n(h, a)| \leqq \rho^{n-1}$$

where

$$\rho = 1 - \min_{i,j,l,m} \left(\frac{N_1}{N_D^2}\right)\left(\frac{m(i, l)m(l, n)}{m(i, j)m(j, m)}\right)^2.$$

The proof can be carried out using Doeblin's "two-particle" method [8].

Using this result we obtain

$$|f_n(g, \phi^{-1}(i)) - f_n(h, \phi^{-1}(i))| \leqq |\sum_{a\varepsilon\phi^{-1}(i)} f_n(g, a) - \sum_{a\varepsilon\phi^{-1}(i)} f_n(h, a)|$$
$$\leqq \sum_{a\varepsilon\phi^{-1}(i)} |f_n(g, a) - f_n(h, a)| = N_D\rho^{n-1}.$$

Furthermore $P\{X_n = i \mid Y_0 = g, X_1, \cdots, X_{n-1}\} - P\{X_n = i \mid X_1, \cdots, X_{n-1}\} \leqq N_D\rho^{n-1}$. To see this, it is sufficient to show that there exists an h (depending on $X_1, \cdots, X_{n-1}$) such that $P\{X_n = i \mid X_1, \cdots, X_{n-1}\} \geqq P\{X_n = i \mid Y_0 = h, X_1, \cdots, X_{n-1}\}$. But this follows from

$$P\{X_n = i \mid X_1, \cdots, X_{n-1}\} = \sum_h P\{X_n = i, Y_0 = h \mid X_1, \cdots, X_{n-1}\}$$
$$= \sum_h P\{X_n = i \mid Y_0 = h, X_1, \cdots, X_{n-1}\}P\{Y_0 = h \mid X_1, \cdots, X_{n-1}\}.$$

The above results allow us to show the following

THEOREM 4.2. *Let $\{Y_n\}$ be a stationary ergodic finite-state Markov chain with strictly positive transition probabilities. Then the entropy of the process $\{X_n = \phi(Y_n)\}$ can be approximated by $\bar{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1)$ and $\underline{G}_n = h(X_n \mid X_{n-1}, \cdots, X_1, Y_0)$. Furthermore, $0 \leqq \bar{G}_n - \underline{G}_n \leqq B\rho^{n-1}$, $0 < \rho < 1$,* where $B = N_D(\log e)/(N_1 \min_{i,j} m(i, j))$.

PROOF. There only remains to show that the approximations converge exponentially. We have

$$\bar{G}_n - \underline{G}_n = E \log \frac{p(X_n \mid Y_0, X_1, \cdots, X_{n-1})}{p(X_n \mid X_1, \cdots, X_{n-1})}.$$

Now $p(X_n \mid X_1, \cdots, X_{n-1}) \geqq N_1 \min_{i,j} m(i, j) = 1/r > 0$ and by the last result $p(X_n \mid Y_0, X_1, \cdots, X_{n-1}) - p(X_n \mid X_1, \cdots, X_{n-1}) \leqq N_D\rho^{n-1}$. Therefore $\bar{G}_n - \underline{G}_n \leqq \log (1 + rN_D\rho^{n-1}) \leqq (\log e)rN_D\rho^{n-1} = B\rho^{n-1}$ for n sufficiently large. Furthermore, since $\bar{G}_n \downarrow H$ and $\underline{G}_n \uparrow H$ we have that $\bar{G}_n - B\rho^{n-1} \leqq \underline{G}_n \leqq H \leqq \bar{G}_n \leqq \underline{G}_n + B\rho^{n-1}$.

It is interesting to note that the exponential bounds can be made independent of the function ϕ.

EXAMPLE. Consider the following transition matrix:

$$M = \begin{Vmatrix} .25 & .24 & .26 & .25 \\ .23 & .25 & .24 & .28 \\ .27 & .25 & .26 & .22 \\ .25 & .26 & .24 & .25 \end{Vmatrix} \qquad \begin{aligned} \text{with } \phi(1) &= \phi(2) = 1 \\ \phi(3) &= \phi(4) = 2 \end{aligned}$$

Approximations for the entropy of this functional process are shown in Table 1.

TABLE 1

n	$\bar{G}_n$	$\underline{G}_n$	Difference
1	1.000 0000	0.999 2784	0.000 7216
2	0.999 3508	0.999 3505	0.000 0003
3	0.999 3508		

6. Applications. A finite-state channel is specified by

(i) An input set A.

(ii) An output set B.

(iii) A finite set of states S.

(iv) A probability function $p(b, s' \mid a, s)$ which represents the probability that, if the channel is currently in state s and receives the input a, it will move to state s' and produce output b. The capacity of such an indecomposable finite-state channel can be defined in terms of the entropy of finitary processes, as follows. Let the input process $X_1, X_2, \cdots$ be a function of a stationary ergodic Markov chain and let $\{X_n, Y_n\}$ be the join input-output process. The capacity is the upper bound over all such input processes of the rate $R(X, Y) = H(X) - H(X \mid Y)$. For these channels, the capacity cannot easily be evaluated. For example, the capacity of a simple two-state channel is not known. However, using the estimates for entropies derived in the earlier sections, we can approximate the rate of information of a finite-state channel for a given input distribution, thus getting a lower bound for the capacity. This will be illustrated for the case of a simple two-state channel. The computations[3] were performed on an IBM 704 digital computer using recursion formulas for the conditional problems.

We have for a two-state channel: $A = \{1, 2\}$, $B = \{1, 2\}$, $S = \{1, 2\}$ and

$$\begin{aligned} p(b, s' \mid a, s) &= \tfrac{1}{2} \quad \text{if} \quad a \neq s, b = a, s' = s \\ &= \tfrac{1}{2} \quad \text{if} \quad a \neq s, b = s, s' = a \\ &= 1 \quad \text{if} \quad b = s' = a = s \\ &= 0 \quad \text{otherwise} \end{aligned}$$

Now we take an input distribution which is a function of a finite-state Markov chain. Then the joint input-output process will be a function of a finite-state Markov chain (the combined input state-output-channel state process). For this input-output process $\{X_n, Y_n\}$, we can estimate $H(X)$, $H(Y)$, and $H(X, Y)$; hence also

$$R(X, Y) = H(X) + H(Y) - H(X, Y).$$

[3] Supported by the Information Systems Branch of the Office of Naval Research under Contract Nonr 222(53).

In this example, we take a two-state Markov input

$$\left\| \begin{matrix} p & q \\ q & p \end{matrix} \right\|.$$

For computing purposes, we can represent the input, output and input-output variables as functions of super variables, W of a Markov chain with transition matrix:

X	Y	W	1	2	3	4	5	6
1	1	1	p	0	0	$q/2$	$q/2$	0
1	1	2	0	$p/2$	$p/2$	0	0	q
1	2	3	p	0	0	$q/2$	$q/2$	0
2	1	4	0	$q/2$	$q/2$	0	0	p
2	2	5	q	0	0	$p/2$	$p/2$	0
2	2	6	0	$q/2$	$q/2$	0	0	p

where the stationary probabilities are given by

$$\lambda_1 = \lambda_6 = 1/(2 + 4q), \qquad \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = q(2 + 4q).$$

The entropy associated with the $\{X\}$ process is directly calculated $H(X) = -(p \log p + q \log q)$. Thus to approximate the rate of the channel with this input it is necessary to approximate the entropy only for the $\{Y\}$ and $\{X, Y\}$ processes. Therefore, we have

$$\bar{R}_n(X, Y) = H(X) + \bar{G}_n(Y) - \underline{G}_n(X, Y)$$

and

$$\underline{R}_n(X, Y) = H(X) + \underline{G}_n(Y) - \bar{G}_n(X, Y).$$

The results shown in Table 2 were obtained for different values of p of the Markov input.

Let us modify the preceding example by feeding each input into the channel twice. Then $A = \{(1, 1), (2, 2)\}$, $B = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$, $S = \{1, 2\}$

TABLE 2

	$p = 0.50$	$p = 0.49$	$p = 0.51$	$p = 0.40$
n	$\underline{R}(X, Y)$	$\underline{R}(X, Y)$	$\underline{R}(X, Y)$	$\underline{R}(X, Y)$
1	0.09816	0.09716	0.09903	0.08064
2	0.29116	0.28498	0.29719	0.22477
3	0.37951	0.37372	0.38514	0.31706
4	0.42435	0.41887	0.42969	0.36496
5	0.45000	0.44470	0.45519	0.39310

TABLE 3

	$p = 0.50$	$p = 0.49$	$p = 0.51$	$p = 0.40$
n	$\underline{R}(XX, YY)$	$\underline{R}(XX, YY)$	$\underline{R}(XX, YY)$	$\underline{R}(XX, YY)$
1	0.7595	0.7587	0.7597	0.7279
2	0.8884	0.8875	0.8888	0.8558
3	0.9480	0.9474	0.9481	0.9164
4	0.9754	0.9749	0.9753	0.9448

and

$$
\begin{aligned}
p(b_1, b_2, s' \mid a, a, s) &= \tfrac{1}{4} \quad \text{if} \quad a \neq s,\ b_1 = a,\ b_2 = a,\ s' = s \\
&= \tfrac{1}{4} \quad \text{if} \quad a \neq s,\ b_1 = a,\ b_2 = s,\ s' = a \\
&= \tfrac{1}{2} \quad \text{if} \quad a \neq s,\ b_1 = s,\ b_2 = a,\ s' = a \\
&= 1 \quad \text{if} \quad b_1 = b_2 = s' = a = s \\
&= 0 \quad \text{otherwise.}
\end{aligned}
$$

Again we can represent the variables as functions of a Markov chain with the following transition matrix:

X,X	Y,Y	W	1	2	3	4	5	6	7	8
1,1	1,1	1	p	0	0	0	$q/4$	$q/4$	$q/2$	0
1,1	2,1	2	p	0	0	0	$q/4$	$q/4$	$q/2$	0
1,1	1,2	3	p	0	0	0	$q/4$	$q/4$	$q/2$	0
1,1	1,1	4	0	$p/2$	$p/4$	$p/4$	0	0	0	q
2,2	2,2	5	q	0	0	0	$p/4$	$p/4$	$p/2$	0
2,2	2,1	6	0	$q/2$	$q/4$	$q/4$	0	0	0	p
2,2	1,2	7	0	$q/2$	$q/4$	$q/4$	0	0	0	p
2,2	2,2	8	0	$q/2$	$q/4$	$q/4$	0	0	0	p

where the stationary probabilities are given by $\lambda_1 = \lambda_8 = (2p + 1)/(6 + 4q)$, $\lambda_2 = \lambda_7 = q/(3 + 2q)$ and $\lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = q/(6 + 4q)$.

The results shown in Table 3 were obtained for different values of p of the Markov input.

7. Acknowledgments. The author is indebted to Professor David Blackwell for suggesting the problem and for his encouragement and guidance during its solution. My thanks also go to the referee for his valuable comments.

REFERENCES

[1] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–423 and 623–656.

[2] MCMILLAN, M. (1953). The basic theorems of information theory. *Ann. Math. Statist.* **24** 196–219.

[3] BLACKWELL, D. (1957). The entropy of functions of finite-state Markov chains, *Transac-*

tions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague. Czechoslovak Academy, Prague.

[4] Burke, C. J. and Rosenblatt, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Statist.* **29** 1112–1122.

[5] Breiman, L. (1960). Finite-state channels, *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Liblice, Prague, 1959.* Czechoslovak Academy, Prague.

[6] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[7] Harris, T. E. (1955). On chains of infinite order. *Pacific J. Math.* **5** suppl. 1 707–725.

[8] Doeblin, W. (1938). Exposé de la théorie des chaînes simples constantes de Markoff à un nombre fini d'etats. *Rev. Math. de l'Union Interbalkanique.* **2** 77–105.