ISRAEL JOURNAL OF MATHEMATICS, Vol. 59, No. 2, 1987

# NORMAL SEQUENCES FOR MARKOV SHIFTS AND INTRINSICALLY ERGODIC SUBSHIFTS

BY

M. SMORODINSKY[a] AND B. WEISS[b]

[a]*School of Mathematical Sciences, The Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv, Israel; and*

[b]*Institute of Mathematics, The Hebrew University of Jerusalem, Jerusalem, Israel*

ABSTRACT

A construction of normal sequences, similar to Champernowne's one, is obtained for Markov shifts and intrinsically ergodic subshifts. For each $n$ a set $\Omega_n$ of $n$-blocks is selected. A normal sequence is constructed by first concatenating the blocks of $\Omega_n$ (in any order) and then concatenating the resultant finite sequences successively.

## 1. Introduction

A number $t \in (0, 1)$ is said to be *normal* to the base $b$ if in the $b$-ary expansion of $t$,

$$t = \sum_{1}^{\infty} \frac{d_j}{b^j} = .\, d_1 d_2 d_3 \cdots$$

each fixed finite block of digits of length $k$ appears with an asymptotic frequency of $b^{-k}$ along the sequence $\{d_j\}_1^{\infty}$. Borel proved that almost every number (in the sense of Lebesgue measure) is normal to the base 10, or any base for that matter, but several decades passed before the first explicit such number was written down by Champernowne, namely

$$.\,1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12\,13\cdots,$$

Received January 14, 1987

the number obtained by successively concatenating all of the natural numbers. Actually Champernowne first proved the normality of the number

$$.0\,1\,2\,3\,4\,5\,6\,7\,8\,9\,0\,1\,02\,03\cdots,$$

which is obtained by first writing all the single digits (blocks of length one), then all the pairs of digits (blocks of length two) and so on.

We can put Champernowne's result in the following way. Let $T$ be the transformation of the unit interval defined by

$$Tx = 10x(\mathrm{mod}\ 1).$$

Then $T$ preserves Lebesgue measure $\lambda$. Let $\boldsymbol{P} = (P_0, P_1, \ldots, P_9)$ be the partition of the unit interval defined by

$$P_i = \left(\frac{i}{10}, \frac{i+1}{10}\right).$$

Then the decimal expansion of $x$ is its ***P**-name* with respect to $T$, that is, the sequence of indices $d_j$ such that

$$T^{j-1}x \in P_{d_j}, \qquad j = 1, 2, \ldots .$$

We say, in general, that a sequence $\{d_j\}^\infty$ is *normal* for a stationary process $(T, P, \mu)$ if for every finite block $b_1 b_2 \cdots b_k$, its asymptotic frequency in the sequence $\{d_j\}^\infty$ equals

$$\mu(P_{b_1} \cap T^{-1}P_{b_2} \cap \cdots \cap T^{-k+1}P_{b_k}).$$

The individual ergodic theorem then says that for ergodic processes the $\boldsymbol{P}$-name of $\mu$-a.e. point is normal, but even though the process may be very easily described, it is in general quite difficult to give explicit constructions of normal sequences (cf. [1] for the case of the continued fraction expansion). We shall generalize Champernowne's construction to obtain explicit normal sequences for finite state ergodic Markov processes and for intrinsically ergodic subshifts (i.e. subshifts whose measure of maximal entropy is unique).As examples of the latter we have shifts of finite type and $\beta$-transformations [4]. For each $n \geqq 1, \Omega_n \subset S^n$, where $S$ is the state space of the process, will be given. Then $w_n$ will be formed by concatenating all the elements of $\Omega_n$ (in any order), and the sequence is formed by concatenating the $w_n$'s, $w_1 w_2 w_3 \cdots$. In Champernowne's construction $\Omega_n = S^n$. For Markov processes we will have to do a little work to get the appropriate $\Omega_n$, while for intrinsically ergodic subshifts the $\Omega_n$ will simply be all of the admissible $n$-blocks.

## 2. Definitions and notation

Let $T$ be a measure preserving transformation of a probability space $X$ with probability measure $\mu$. Let $\boldsymbol{P} = (P_i)$, $i \in S$ be a finite measurable partition of $X$. The pair $(T, \boldsymbol{P})$ is a *process* and $S$ is the state space. The $n - T - P$ name of $x \in X$ is the finite sequence $x^{(n)} = x_1, \ldots, x_n$, $x_i \in S$, $T^{i-1}x = P_{x_i}$, $1 \leqq i \leqq n$.

Let $b = b_1 \cdots b_k$, $b_i \in S$ be a block of length $k$ and $\omega = \omega_1 \cdots \omega_n$, $\omega_i \in S$ a block of length $n$.

We say that $b$ occurs at the $i$th place in $\omega$ if $1 \leqq i \leqq n - k$ and

$$\omega_i = b_1 \cdots \qquad \omega_{i+k-1} = b_k.$$

Let $f_b(\omega)$ be the frequency of $b$ in $\omega$, i.e.

$$f_b(\omega) = \frac{1}{n-k} \, |\{i \mid b \text{ occurs at the } i\text{th place of } \omega\}|.$$

Here $|A|$ denotes the cardinality of the set $A$.

Put

$$\mu(b) = \mu\{x \mid x^{(k)} = b\}.$$

A sequence $(x_n)_{n=1}^{\infty}$ is normal for $(T, \boldsymbol{P})$ if for each $b$

$$\lim_{n \to \infty} f_b(x^{(n)}) = \mu(b).$$

($(x)_{n=1}^{\infty}$ is not necessarily a name of a point and $x^{(n)} = x_1, x_2, \ldots, x_n$.)

## 3. A set of normal sequences

Let $(T, P)$ be a process with state space $S$. Let $\Omega_n \subset S^n$ be a subset of blocks of length $n$.

DEFINITION. A sequence $\{\Omega_n\}_{n=2}^{\infty}$ is a LLN (Law of Large Numbers) sequence for $(T, P)$ if for any $b \in S^k$ and $\varepsilon > 0$ there exists $n(\varepsilon, b)$ such that if $n \geqq n(\varepsilon, b)$,

$$|\{\omega \in \Omega_n : |f_b(\omega) - \mu(b)| < \varepsilon\}| > (1 - \varepsilon)|\Omega_n|.$$

Denote by *concat* $(\Omega_n)$ the subset of $S^{n\cdot|\Omega_n|}$ which is obtained by concatenating all the blocks of $\Omega_n$ in all possible orders. And denote by *concat* $(\{\Omega_n\}_{n=2}^{\infty})$ the subset of $S^N$ which is all the infinite sequences obtained by concatenating successively for each $n$ one member of *concat* $(\Omega_n)$.

THEOREM 1. *Let $\{\Omega_n\}_{n=2}^{\infty}$ be a* LLN *sequence for $(T,P)$ and let $a>0$ be a constant such that for all $n$, $|\Omega_{n+1}|\leqq a|\Omega_n|$. Then each $W\in$ concat $(\{\Omega_n\}_{n=1}^{\infty})$ is normal for $(T,P)$.*

PROOF. Let $b$ of length $k$ and $\varepsilon>0$ be given. Let $W=W_1W_2\cdots$ where $W_n\in$ *concat* $\Omega_n$, consider

$$_nW=W_nW_{n+1}\cdots.$$

It is enough to prove that $b$ appears in $_nW$ in the correct frequency.

Choose $n>n(\varepsilon/3a)$ and so large that $k/n<\varepsilon/3$. Now, let $m$ be a natural number and $w_1\cdots w_m={}_nW^m$ the first $m$ terms of $_nW$ and put $N_n=|\Omega_n|$. Assume that

$$\begin{aligned} m&=nN_n+(n+1)N_{n+1}+\cdots+(n+p)N_{n+p}+(n+p+1)q+r\\ &=m_1+(n+p+1)q+r \end{aligned}$$

where $0\leqq q<N(n+p+1)$ and $0\leqq r<n+p+1$.

Obviously

$$|f_b({}_nW^{m_1})-\mu(b)|<\frac{\varepsilon}{3a}+\frac{k}{n}$$

(the $k/n$ comes from end effects of occurrence of $b$ between two $n$ blocks). Among the blocks of length $n+p+1$ at most $(\varepsilon/3a)N_{n+p+1}\leqq(\varepsilon/3)m$ do not have the proper frequency of $b$. Therefore

$$|f_b({}_nW^{m})-\mu(b)|<\frac{\varepsilon}{3a}+\frac{k}{n}+\frac{\varepsilon}{3}\leqq\varepsilon$$

if $n$ is large enough so that $k/n<\varepsilon/3$. This completes the proof. □

## 4. Normal sequences for Markov shifts

Let $M=(q_{i,j})_{i,j\in S}$ be an irreducible Markov transition matrix on the state space $S$. Let $(p_i)$, $i\in S$ be the stationary probability vector of $M$. Let $(T,P)$ be the stationary Markov process obtained by $M$, and $\mu$ its probability measure. For each $n \geqq 2$ consider the set $\Omega_n \subset S^n$ such that each $(i,j) \in S^2$ occurs in $\omega \in \Omega_n$, $n_{i,j}$ times where $n_{ij}/(n-1) \to p_i p_{ij}$. And for each $i \in S$

$$(*) \qquad \sum_j n_{ij} = \sum_j n_{ji}.$$

Such a choice is possible since for each $i \in S$, $\Sigma_j p_i q_{i,j} = \Sigma_j p_j q_{ji}$ and one can choose $n_{i,j} = [(n-1)p_i q_{i,j}] + \theta_{ij}$ where $0 \leqq \theta_{ij} \leqq |S|$. Let $\nu_n$ be the uniform measure on $\Omega_n$. Let $X_m(\omega)$ be the $m$th coordinate of $\omega$, $m = 1, \ldots, n$.

LEMMA 1. *$\{X_m\}_{m=1}^n$ is stationary under $\nu_n$, and*

$$\nu_n[x_1 = i, x_2 = j] \underset{n\to\infty}{\to} p_i q_{ij} = \mu(i,j).$$

PROOF. Because of condition $(*)$ every sequence $w \in \Omega_n$ begins and ends with the same coordinate. So, identify first and last coordinates of $\omega$ to obtain a cyclic sequence. Now rotate by 1 and break the cycle by putting first and last coordinate the same. Obviously the new sequence $\omega'$ also belongs to $\Omega_n$ and such a correspondence is one to one. Since $\nu_n$ is the uniform measure, the stationarity follows.

Now, let $Y_m = 1_{[x_m = i, x_{m+1} = j]}$, $1 \leqq m \leqq n-1$, $\Sigma_{m=1}^{n-1} Y_m = n_{i,j}$ for all $w \in \Omega_n$ and therefore $n_{ij} = \Sigma_{m=1}^{n-1} E_{\nu_n}(Y_m) = (n-1)\ \nu_n[Y_1 = 1]$ (because of stationarity). And $\nu_n[Y_1 = 1] = \nu_n[X_1 = i, X_2 = j] = n_{ij}/(n-1)$. □

Next we need to estimate the growth of $|\Omega_n|$. We make use of a result in multigraphs. Let $G = (S, E)$ be the directed multigraph where $S$ are the nodes and for each $i, j \in S$ there are $n_i$ edges from $i$ to $j$. An Eulerian circuit is a circuit which uses each edge exactly once. According to a theorem of Aardenne–Ehrenfest, de Bruijn ([2], p. 240) there are $\Delta_1 \Pi_i (n_i - 1)!$ Eulerian circuits where $n_i = \Sigma_j n_{ij}$. And $\Delta_1$ is the number of arborescences subgraphs of $G$ rooted at the node 1 (see [2], p. 5). Each Eulerian circuit corresponds to a sequence $\omega \in \Omega_n$ in an obvious way. But since we do not distinguish the different $n_i$ edges leading from $i$ to $j$, we get that

$$(**) \qquad |\Omega_n| = n \cdot \Delta_1 \frac{\prod_i (n_i - 1)!}{\prod_{i,j} n_{i,j}!}.$$

LEMMA 2. $(\log|\Omega_n|)/n \to -\Sigma_i p_i \Sigma_j q_{ij} \log q_{ij}$.

PROOF. It is obvious that $\Delta$ grows polynomially with $n$ and is bounded by $(2n)^{|S|}$. Applying Stirling's formula and the asymptotic behaviour of $n_i$ and $n_{ij}$ will conclude the proof. □

LEMMA 3. *Let $\Omega_n \subset S^n$, $n \in N$ be such that the uniform measure $\nu_n$ on $\Omega_n$ converges weakly to an ergodic stationary measure $\mu$. Then, $\{\Omega_n\}$, $n \in N$ is* LLN *for $\mu$.*

PROOF. Let $b \in S^k$ be given and $\varepsilon > 0$. Let $m$ be large enough so that by the ergodic theorem the following holds:

$$B = \{\omega \in S^m : |f_b(\omega) - \mu(b)| < \varepsilon/4\},$$

$$\mu(B) > 1 - \varepsilon^2/8.$$

The weak convergence implies that $\nu_n(B) \to \mu(B)$. So for $n$ sufficiently large $\nu_n(B) > 1 - \varepsilon^2/4$ and $m/n < \varepsilon/2$.

Let

$$Y_i(\omega) = \begin{cases} 1 & \text{if } \omega_i \cdots \omega_{i+m-1} \in B, \\ 0 & \textit{otherwise}; \end{cases}$$

$$\nu_n(B) = \frac{1}{n-m} \sum_{i=1}^{n-m} E_{\nu_n}(Y_i) = \frac{1}{n-m} \sum_{i=1}^{n-m} \frac{1}{|\Omega_n|} \sum_{\omega \in \Omega_n} Y_i(\omega)$$

$$= \frac{1}{|\Omega_n|} \sum_{\omega \in \Omega_n} \frac{1}{n-m} \sum_{i=1}^{n-m} Y_i(\omega) > 1 - \frac{\varepsilon^2}{4}.$$

It follows that there is a set $B' \subset B$ of measure $\nu_n(B') > 1 - \varepsilon$ such that

$$\omega \in B' \Rightarrow \frac{1}{n-m} \sum_{i=1}^{n-m} Y_i(\omega) > 1 - \frac{\varepsilon}{4}.$$

For such an $\omega$

$$|f_b(\omega) - \mu(b)| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{m}{n} < \varepsilon.$$ □

THEOREM 2. *Let $\{\Omega_n\}_{n=2}^{\infty}$ be defined as above, then each $W \in concat$ $(\{\Omega\}_{n=2}^{\infty})$ is normal for $(T, P)$.*

PROOF. By Lemma 3, to check that $\{\Omega\}_{n=2}^{\infty}$ is LLN for $\mu$ it suffices to prove the weak convergence of the uniform measures $\nu_n$ to $\mu$. In Lemma 1 the convergence of $\nu_n$ on two-blocks to $\mu$ was proved. So

$$H_{\nu_n}[X_1 \mid X_2] \to H_\mu[X_1 \mid X_2] = h(T, P).$$

Now for $n \geqq k$

$$H_{\nu_n}[X_1 \mid X_2 \cdots X_k] \leqq H_{\nu_n}[X_1 \mid X_2].$$

By Lemma 2

$$\frac{1}{n} H_{\nu_n}[X_1 \cdots X_n] = \frac{1}{n} \sum_{i=1}^{n} H[X_1 \mid X_2 \cdots X_i] \to h(T, P).$$

Therefore, for fixed $k$

$$H_{\nu_n}[X_1 \mid X_2] - H_{\nu_n}[X_2 \cdots X_k] \underset{n \to \infty}{\to} 0.$$

So, $X_1$ and $X_3, \ldots, X_n$ are asymptotically independent given $X_2$, under $\nu_n$.

A result on the connection between conditional entropy and $\varepsilon$-independence ([5], p. 20) implies that for each $(i_1, \ldots, i_n) \in S^k$

$$\nu_n[X_1 = i_1 \mid X_2 = i_2 \cdots X_k = i_k] - \nu_n[x_1 = i_1 \mid X_1 = i_1] \underset{n \to \infty}{\to} 0.$$

Now,

$$\begin{aligned} &\nu_n[X_1 = i_1, X_2 = i_2, \ldots, X_k = i_k] \\ &\quad = \nu_n[X_i = i_2 \cdots X_k = i_a] \nu_n[X_1 = i_1 \mid X_2 = i_2 \cdots X_k = i_k] \end{aligned}$$

and an induction argument on $k$ completes the proof of the weak convergence.

Finally, the estimates of Lemma 2 show also the boundedness of $|\Omega_{n+1}| / |\Omega_n|$. □

## 5. Normal sequences for intrinsically ergodic processes

Let $X \subset S^Z$ be a closed shift invariant subset of the full $|S|$-shift such that there is a unique measure $\mu$ maximizing the entropy (achieving the topological entropy), i.e. $(X, \sigma)$ is intrinsically ergodic, where $\sigma$ is the shift transformation.

THEOREM 3. *Let $\Omega_n \subset S^n$ be the set of all the n-blocks that occur in $X$, i.e. $\omega \in \Omega_n$ if there is some $x \in X$ with $x_0 \cdots x_{n-1} = \omega$ where $x = (\ldots, x_{-1}x_0x_1, \ldots)$.*

*Then, for all $W \in$ concat $(\{\Omega_n\}_{n=1}^{\infty})$ $W$ is normal for $\mu$ the measure of maximal entropy.*

In the proof of Theorem 3 we shall make use of the following

LEMMA 4. *Let $h > 0$ and let $V_n \in S^n$ denote the set of n-blocks v such that their empirical k-block distribution has entropy $\leqq h$. Then for any $\varepsilon > 0$ we have*

$$|V_n| = O(2^{n(h+\varepsilon)}).$$

*(By the empirical k-block distribution of v we mean the probability vector $\{f_b(v) : b \in S^k\}$ and the entropy is $(1/k)H\{f_b(v) : b \in S^k\}$.)*

PROOF OF LEMMA 4. Let $\mu_1, \ldots, \mu_m$ be a finite set of stationary measures on $s^k$ such that

(i) $(1/k)H(\mu_i) \leqq h$, $1 \leqq i \leqq m$,

(ii) for any stationary measure $\mu$ on $S^k$ with $(1/k)H(\mu) \leqq h$ there is some $i$ and $|\mu - \mu_i| < \delta$ ($\delta$ small, to be chosen later).

Divide $V_n$ into sets $V_n^1, \ldots, V_n^m$ according to which $\mu_i$ is closest to the empirical distribution of $v \in V_n$. (Although the empirical distribution of $v$ is not necessarily stationary, its deviation from it is $O(1/n)$.)

We will count now each $V_n^i$ separately. $\mu_i$ being a stationary measure on $k$-blocks gives rise to a $(k-1)$-step stationary Markov measure. We assign to each $v \in V_n^i$ the probability of obtaining $v$ according to this Markov measure which we denote again by $\mu_i$. Put $\mu_i^*(b) = \mu_i(b_k \mid b_1 \cdots b_{k-1})$, then

$$\mu_i(v) = \mu_i(v^{(k-1)}) \prod_{b \in S^k} [\mu_i^*(b)]^{(n-k)f_b(v)}.$$

Since

$$\sum_{b \in S^k} |f_b(v) - \mu_i^*(b)| < \delta$$

we get that for $\delta$ sufficiently small

$$\mu_i(v) \geqq 2^{-n(h_i+\varepsilon)}$$

where $h_i$ is the entropy of $\mu_i$. Therefore the number of such $v$'s is at most $2^{n(h_i+\varepsilon)} \leqq 2^{n(h+\varepsilon)}$ as required. □

We return now to the

PROOF OF THEOREM 3. Let $W \in concat(\{\Omega_n\}_{n=1}^{\infty})$ and $W = v_1 \cdot v_2 \cdots v_n \cdots$ where $v_n \in concat(\Omega_n)$.

Since $|\Omega_{n+1}| \leqq |S| \cdot |\Omega_n|$ it suffices to show that for any fixed $k$ the empirical $k$-block distribution in $v_n$ tends to that given by $\mu$. Let $n_i$ be a subsequence such that for all $k$ the empirical $k$-block distribution in $v_{n_i}$ converges to some measure $\mu$. It suffices to show that it is always the case that $\mu = \nu$.

Obviously $\nu$ is a shift invariant measure. So, in order to verify $\mu = \nu$ we have to show that $h(\mu) = h(\nu)$.

If $h(\mu) = 0$ there is nothing to prove since always $h(\nu) \leqq h(\mu)$. Let $\varepsilon > 0$ be given, we shall now show that

$$(**) \qquad 1/kH_\nu[X_1, \ldots, X_k] \geqq h(\mu) - \varepsilon.$$

This will prove, by first letting $k \to \infty$ and then $\varepsilon \to \infty$, that $h(\nu) = h(\mu)$. From the definition of topological entropy it follows that if $n$ is large enough then

$$|\Omega_n| \geqq 2^{n(h(\mu) - \varepsilon/3)}.$$

That means by Lemma 4 that for any $\delta > 0$ all but $\delta |n|$ or $\omega \in \Omega_n$ will have entropy of the empirical distribution greater than $h(\mu) - \frac{2}{3}\varepsilon$ and therefore by the convexity of the entropy function, $(**)$ is proved. □

## REFERENCES


1. R. Adler, M. Keane and M. Smorodinsky, *A construction of a normal number for the continued fraction transformation*, J. Number Theory **13** (1981), 95–105.
2. C. Berge, *Graphs and Hypergraphs*, North-Holland, Amsterdam, 1976.
3. D. G. Champernowne, *The construction of decimal normal in the scale of ten*, J. London Math. Soc. **8** (1933), 254–260.
4. Franz Hofbauer, *β-Shifts have unique maximal measure*, Monatsh. Math. **85** (1978), 189–198.
5. M. Smorodinsky, *Ergodic Theory, Entropy*, Lecture Notes in Mathematics **214**, Springer-Verlag, Berlin, 1971.