ISRAEL JOURNAL OF MATHEMATICS, Vol. 28, Nos. 1–2, 1977

THE LIMITS IN $\bar{d}$ OF MULTI-STEP MARKOV CHAINS

BY

DANIEL J. RUDOLPH AND GIDEON SCHWARZ

ABSTRACT

The $\bar{d}$-closure of the class of multi-step Markov chains is shown to consist of all direct products of Bernoulli processes with processes of rational pure point spectrum. The class of processes that are approached in $\bar{d}$ by their canonical multi-step Markov approximations is also studied. It is found to be strictly smaller than the former class, dense in it, and characterized within it by a certain (noninvariant) property of its rotation factors.

1. Introduction

Everything about a (discrete-time, countable-state) stationary stochastic process is specified, if the joint distributions of arbitrarily many consecutive variables are given. If such distributions are given only for some fixed number, say $k + 1$, of consecutive variables, the process is not uniquely defined thereby, but among the class of processes compatible with such given joint distributions, there is one canonical representative, in some sense the most random of the class. It is the (unique) k-step Markov chain whose transition probabilities, from a block of k variables to the next variable, are as implied by the given joint distribution of $k + 1$ variables. We call this process the *canonical k-step Markov approximation* to the given process. The sequence of such approximations has a nice consistency property: for $j > k$, the k-step approximation to the j-step approximation is the k-step approximation to the given process; furthermore, the term "approximation" is justified by the fact that in the *vague topology* on processes, that defines convergence as convergence of finite joint distributions, every process is indeed the limit of its canonical k-step Markov approximations. For the study of long-range properties of processes the finer $\bar{d}$-topology, as introduced by Ornstein [3], is more appropriate, and the question of $\bar{d}$-

Received May 12, 1976

convergence of the canonical approximations arises naturally. Friedman and Ornstein obtained the following results about this question [2]:

(a) For a Bernoulli process (that is a process isomorphic to a sequence of independent variables) the canonical approximations do converge in $\bar{d}$ to the process.

(b) Any $\bar{d}$-limit of aperiodic k-step Markov chains is a Bernoulli process.

Since the canonical approximations possess at least the mixing properties of the process they approximate, total ergodicity of that process implies the same, and hence aperiodicity, of the approximations. Therefore, within the class of totally ergodic processes, (a) and (b) yield a characterization of Bernoulli processes as the processes whose canonical approximations converge $\bar{d}$, or alternatively, as the $\bar{d}$-closure of the class of aperiodic k-step Markov chains.

An attempt to obtain similar results beyond the totally ergodic class, runs into some difficulties. Schwarz's examples [4] show that without assuming aperiodicity,

(i) $\bar{d}$-convergence of the canonical approximations is no longer an isomorphism invariant;

(ii) the property of being isomorphic to a k-step Markov chain is not $\bar{d}$-closed;

(iii) A process may be a $\bar{d}$-limit of a sequence of k-step Markov chains, and yet its canonical approximations may diverge $\bar{d}$.

These facts demolish some otherwise plausible conjectures about modifications of (a) and (b) for the not necessarily totally ergodic case.

2. The results

The full answer to the questions raised above explicitly and implicitly, requires some preliminaries.

Following the result of Shields and Thouvenot [6], that the class of direct products of Bernoulli processes and processes of zero entropy is $\bar{d}$-closed, and the characterization by Adler, Shields and Smorodinsky [1] of the isomorphism class of Markov chains ("k-step" plays no role here) as direct products of Bernoulli processes and finite rotations, we can expect products of Bernoulli processes and entropy zero processes to be the only processes relevant to our topic.

In fact an even smaller class plays the leading role: those direct products whose entropy zero part has rational pure point spectrum. Since the transformation whose spectrum consists of the binary rationals on the circle is known as the von Neumann transformation, we call transformations with rational pure point spectrum *generalized von Neumann transformations.* Finally, a process that is isomorphic to the direct product of a Bernoulli process and a generalized von Neumann transformation shall be called a *BGVN process.*

THEOREM I. *The closure in $\bar{d}$ of the set of all k-step Markov chains is the set of all BGVN processes.*

THEOREM II. *The canonical k-step Markov approximations of a finite state process converge in $\bar{d}$ to the process, if and only if* (i) *the process is BGVN, and* (ii) *for every finite rotation factor of the transformation, there exists a finite block of variables of the process that spans that factor.*

REMARK. The family of transformations giving rise to BGVN processes has a very simple structure. Each member is uniquely determined by its entropy and its spectrum. Each rational spectrum is an additive subgroup of the rationals on the circle, and it determines the corresponding zero entropy transformation uniquely. The subgroups are in turn determined by their least common denominators, provided this concept is extended beyond the integers to include the G-numbers of Steinitz [7], defined as follows. An integer is determined by its prime power decomposition. A G-number is similarly defined as a formal product of prime powers, but now each power may be infinite, and infinitely many different primes may occur. If we adopt the obvious definition of divisibility for G-numbers, least common multiples can be defined for any sets of G-numbers. The G-numbers can be obtained from the positive integers by closure under such least common multiples.

Associating with each subgroup of the rationals (on the circle) the least common denominator of its elements, we obtain a one-to-one correspondence between all subgroups and all G-numbers. Divisibility of G-numbers goes into inclusion of subgroups, which in turn goes into the factor relation between the transformations that have these subgroups for their spectra.

3. Limits of k-step Markov chains

Consider two (k-step) Markov processes X and $\bar{X}$. Each one is isomorphic to the direct product of a Bernoulli process with a finite rotation. Denote the respective rotation numbers by n and m, and assume $h(\bar{X}) \geq h(X)$. If Y is a direct product of a Bernoulli process of entropy $h(\bar{X})$ with a rotation corresponding to the least common multiple of n and m, both X and $\bar{X}$ occur as factors of Y. In fact, they can generally be embedded in Y in more than one way.

PROPOSITION. *There exists a function g, with $g(t)\to 0$ when $t\to 0$, such that, for any X, $\bar{X}$ and Y as above with finite $h(\bar{X})$, any embedding of $\bar{X}$ in Y admits an embedding of X in Y such that* $\mathrm{Prob}\,(X_0\neq \bar{X}_0)< g(\bar{d}(X,\bar{X}))$.

PROOF. The rotation factors of X and $\bar{X}$ generate together a factor of the ergodic joining of X and $\bar{X}$ that attains $\bar{d}\,(X,\bar{X})$. In fact, they generate exactly a rotation by the least common multiple of n and m, since this is the unique ergodic joining of two finite rotations. As a factor of zero entropy, this (n,m)-rotation is independent of each of the Bernoulli parts of X and of $\bar{X}$ in the joining. Denoting the rotations and their joining by R_n, R_m and $R_{(n,m)}$ respectively, we can reinterpret $\bar{d}\,(\bar{X},X)$ as the relative $\bar{d}$, given $R_{(n,m)}$, of the process generated by X and R_m and the process generated by $\bar{X}$ and R_n. The Proposition now follows from a result proved in the relative isomorphism theory by Thouvenot [8]:

LEMMA. *There exists a function g, with $g(t)\to 0$ when $t\to 0$, such that if Z and $\bar{Z}$ are two processes that share the same factor H, and H is Bernoulli-complemented in each of the processes, and $h(\bar{Z})\geqq h(Z)$, then Z can be embedded in $\bar{Z}$, so that* $\mathrm{Prob}\,(Z_0\neq \bar{Z}_0)< g\,(\bar{d}\,(Z,\bar{Z};H))$. *Here $\bar{d}\,(.,.;H)$ denotes the relative $\bar{d}$-distance given H.*

The Proposition is not quite symmetric in the two processes, since the order of their entropies is relevant. This could foil an attempt to obtain an appropriate embedding of a sequence of processes by applying the Proposition repeatedly, when the entropies do not form a monotone sequence. However, if $\{X^{(i)}\}$ is a $\bar{d}$-convergent sequence of processes, the entropies converge and therefore differ from a monotone sequence by a positive null sequence, say $\{\delta_i\}$. By replacing each process $X^{(i)}$ by its direct product with a Bernoulli process of entropy δ_i, we obtain a sequence with monotone entropies, and with the same $\bar{d}$-limit, if any, as the given $\{X^{(i)}\}$. Now we apply the proposition repeatedly, with $\varepsilon=\varepsilon_i$ where $\Sigma\,\varepsilon_i<\infty$, and obtain for a $\bar{d}$-convergent sequence of k-step Markov chains of periods n_i and entropies η_i, a $\bar{d}$-convergent embedding in the BGVN process whose entropy is $\sup\eta_i$ and whose G-number is the least common multiple of the n_i. The limit, being a factor of the BGVN process, is itself a BGVN process.

If $\{X^{(i)}\}$ is a $\bar{d}$ convergent sequence of processes with unbounded entropies, by modifying successive partitions slightly we may assume the entropies are finite and monotone increasing to infinity. In an infinite entropy BVGN process with appropriate rational spectrum, suppose we have a copy of $X^{(i)}$. In this BVGN process there is an independent process $Z^{(i)}, \infty > h(Z^{(i)}) > h(X^{(i)}) - h(X^{(i+1)})$, independent of the copy of $X^{(i)}$. Add, as a direct product, a copy of $Z^{(i)}$ to the $\bar{d}$

joining of $X^{(i)}$ and $X^{(i+1)}$. Now apply the lemma to $(X^{(i)} \times Z^{(i)})$ and $X^{(i+1)}$ to copy $X^{(i+1)}$ into the BVGN process. The sequence of copies converge inside the BVGN process. This completes the proof of one direction of Theorem I: *the $\bar{d}$-closure of the class of k-step Markov chains is included in the class of BGVN processes.*

The proof of the other direction of Theorem I uses Theorem II, which we now proceed to prove.

4. When do the canonical approximations converge?

Consider now the sequence $\{M^{(k)}(X)\}$ of the canonical k-step Markov approximations to some fixed process X. Since, as was pointed out in the introduction, a canonical approximation shows at least the degree of mixing of the approximated process, the maximal rotation factor of $M^{(k)}(X)$ is nondecreasing in k. Now assume that $\{M^{(k)}(X)\}$ is $\bar{d}$-convergent. Its $\bar{d}$-limit cannot differ from its vague limit, which is X, and therefore X is BGVN. Let p^r be some finite prime-power factor of X. If the p^r-rotation factor of X were not spanned by some (finite) number k of variables of the process X, no $M^{(k)}(X)$ would have it as a factor. But then, for each $M^{(k)}(X)$, the process formed by joining the variables of $M^{(k)}(X)$ into blocks of p^r variables each, would be ergodic. The blocking preserves $\bar{d}$-convergence, and hence by a corollary of Shields and Thouvenot [6], the similarly blocked X would be ergodic as well, implying that X itself has no p^r-rotation factor, contrary to assumption. Thus we have one direction of Theorem II: *if the canonical approximations to X converge $\bar{d}$, X must be BGVN, and every finite rotation factor of X must be spanned by a finite block of variables.* Furthermore, the rotation factors of the $M^{(k)}(X)$ increase up to the full generalized von Neumann part of X.

For the other direction, we need once more a result from Thouvenot's "relative theory" [8].

LEMMA 2. *If $(T, P \vee H)$ is the direct product of a B process and (T, H) with finite state space, then given $\varepsilon > 0$, there are m, n, and δ so that if $(\bar{T}, \bar{P} \vee \bar{H})$ is any other ergodic process that satisfies*

$$(T, H) = (\bar{T}, \bar{H}) \tag{4.1}$$

$$\textit{for all but } \delta \textit{ of the atoms } E \subset \bigvee_{-m}^{m} T^i(H) = \bigvee_{-m}^{m} \bar{T}^i(\bar{H}), \tag{4.2}$$

$$\left| \operatorname{dist}\left(\bigvee_{-n}^{n} T^i(P)/E \right) - \operatorname{dist}\left(\bigvee_{-n}^{n} \bar{T}^i(\bar{P})/E \right) \right| < \delta \textit{ and}$$

(4.3) $$|h(T, P \vee H) - h(\bar{T}, \bar{P} \vee \bar{H})| < \delta, \textit{ then}$$

(4.4) $$\bar{d}(T, P; \bar{T}, \bar{P}) < \varepsilon.$$

Now let X be a BGVN process with finite state space all of whose prime power cycles are spanned by finite blocks. We denote its generalized von Neumann part by V, and apply the lemma as follows: Let (T, P) and (T_k, P_k) be the (shift, zero-partition) of X and $M^{(k)}(X)$ respectively. The role of H in the lemma is played by a generating partition for V. Also, let n_k be the period of $M^{(k)}(X)$, and B_k and R_{n_k} the Bernoulli and rotation parts of T_k.

If we can, in each (T_k, P_k), "complete" R_{n_k} to all V, with a generator H_k so that for any m, n and δ, for k large enough, we get 4.2 and 4.3 for $(T, P \vee H) = T_k, P_k \vee H_k)$, we will be done by 4.4. As $H \subset \bigcup_k R'_{n_k}$, in each R'_{n^k} there is an $\hat{H}'_k$ with $\mu(H \Delta \hat{H}'_k) \to 0$. Coded in (T_k, P_k) exactly as it is in (T, P), is an $\hat{H}_k$. We can think of $(T_k, P_k) = B_k \times R_{n_k}$ as a factor of $B_k \times V$, with this partition $\hat{H}_k$ in it. In V is also $\hat{H}'_k$. These both generate the same process, and so $\hat{H}_k = T^l_k(\hat{H}'_k)$. Let $H_k = T^l_k(H)$, H the generator of V. Now $\mu(H_k \Delta \hat{H}_k) = \mu(H \Delta \hat{H}'_k) \to 0$. This attaches to (T_k, P_k) a copy of the (T, H) process. Now suppose we are given n, m and δ. Let $K = \sup(m, n)$, and $k > 2K$. Then $\operatorname{dist}(\vee_{-K}^{K} T^i_k(P_k \vee \hat{H}_k)) = \operatorname{dist}(\vee_{-K}^{K} T^i(P \vee \hat{H}'_k))$ by the Markov property. If k is large enough we further get

$$\left| \operatorname{dist}\left(\bigvee_{-K}^{K} T^i(H'_k)\right) - \operatorname{dist}\left(\bigvee_{-K}^{K} T^i(H)\right) \right| < \delta^2/4.$$

We conclude

(4.2′) $$\text{for all but } \delta \text{ of the atoms } E \subset \bigvee_{-K}^{K} T^i(H) = \bigvee_{-K}^{K} T^i_k(H_k),$$

$$\left| \operatorname{dist}\left(\bigvee_{-K}^{K} T^i_k(P_k \vee H_k)\right) - \operatorname{dist}\left(\bigvee_{-K}^{K} T^i(P \vee H)\right) \right| < \delta.$$

$$\text{As} \left| \operatorname{dist}\left(\bigvee_{-K}^{K} T^i_k(P_k \vee H_k)\right) - \operatorname{dist}\left(\bigvee_{-K}^{K} T^i(P \vee H)\right) \right| < \delta.$$

As the entropy of V and of R_{n_k} is zero, we also have

(4.3′) $$h(T_k, P_k \vee H_k) = h(T_k, P_k) \to h(T, P) = h(T, P \vee H).$$

Now Lemma 2 yields the $\bar{d}$-convergence of the $M^{(k)}(X)$, and thus completes the proof of Theorem II.

5. Processes with $\bar{d}$-convergent canonical approximations are dense in BGVN

This is the missing link for the other half of Theorem I to follow from Theorem II.

First, observe that since every G-number is the least common multiple of a sequence of finite integers, and finite state processes are dense in those with countably many states, processes of the form finite entropy Bernoulli $\times$ finite rotation are dense in the BGVN processes. So, let X be isomorphic to the direct product of a Bernoulli process and an n-rotation. Given $\varepsilon > 0$, there is a k such that the rotation factor of X is ε-spanned by $X_{-k}, \cdots, X_0, \cdots, X_k$. In particular, for R_1, one of the sets of the rotation factor, there is an R_1', differering from R_1 by ε, among the sets spanned by $X_{-k}, \cdots, X_k$. Defining the variable Y_0 as 1 on $R_1 - R_1'$, -1 on $R_1' - R_1$, and $Y_0 = X_0$ otherwise (assume that 1 and -1 do not occur among the states of X), we obtain a process Y, within ε of X in $\bar{d}$, for which the conditions of Theorem II hold. This completes the proof of the statement forming the title of this section, and also of Theorem I.

ACKNOWLEDGEMENT

Through repeated discussions with D. S. Ornstein and J.-P. Thouvenot, their part in these results considerably exceeds what might be concluded from our references to their publications.

REFERENCES

1. R. L. Adler, P. Shields and M. Smorodinsky, *Irreducible Markov shifts*, Ann. Math. Statist. **43** (1972), 1027–1029.

2. N. A. Friedman and D. S. Ornstein, *On isomorphism of weak Bernoulli transformations*, Advances in Math. **5** (1970), 365–394.

3. D. S. Ornstein, *Ergodic Theory, Randomness and Dynamical Systems*, Yale University Press, New Haven, 1973.

4. G. Schwarz, *Noninvariance of $\bar{d}$-convergence of k-step Markov approximations*, to appear in Ann. Probability.**4** (1976).

5. P. Shields, *The Theory of Bernoulli Shifts*, University of Chicago Press, Chicago, 1973.

6. P. Shields and J.-P. Thouvenot, *Entropy zero $\times$ Bernoulli processes are closed in the $\bar{d}$-metric*, Ann. Probability 3 (1975), 732–736.

7. E. Steinitz, *Theorie der Körper*, J. Reine Angew. Math. (1910) (reprinted 1950, Chelsea, New York).

8. J.-P. Thouvenot, *Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schéma de Bernoulli*, Israel J. Math. **21** (1975), 177–207.

THE INSTITUTE FOR ADVANCED STUDIES
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL