Stochastic Processes and their Applications 18 (1984) 33–45
North-Holland

A LIMITATION OF MARKOV REPRESENTATION FOR STATIONARY PROCESSES

H.C.P. BERBEE
Mathematical Centre, Kruislaan 413, 1098 SJ, Amsterdam, The Netherlands
R.C. BRADLEY*
Department of Mathematics, Indiana University, Bloomington, IN 47405, USA

Received 13 January 1983
Revised 22 June 1983

The existence of a representation of a stationary process as an instantaneous function of a real, irreducible Markov chain (Harris chain) imposes important restrictions on the distribution of the process. We construct a countably-valued stationary process with a very strong mixing property for which such a representation does not exist.

instantaneous function of irreducible Markov chain strictly stationary * ψ^*-mixing * entropy

1. Introduction

Suppose one is interested in a certain random stationary phenomenon. To study it, one makes a series of measurements and thus obtains a stationary sequence $\xi := (\xi_n)_{n\in\mathbb{Z}}$. Then one often models ξ as a functional on an underlying Markov chain (or perhaps as a Markov chain itself). This approach is of great value; it provides a nice probabilistic structure that can be used in the statistical analysis of the phenomenon. However we shall show below that there are quite reasonable situations where, in a certain sense, such an approach can never be entirely correct.

Throughout this article we restrict our attention to strictly stationary processes.

Let $\xi := (\xi_n)_{n\in\mathbb{Z}}$ be a stationary process and $Y := (Y_n)_{n\in\mathbb{Z}}$ a stationary Markov chain. The process ξ is *represented as an instantaneous function of* Y if

$$\xi_n = f(Y_n) \quad \text{for } n \in \mathbb{Z}, \tag{1.1}$$

where f is a measurable function on the state space of Y. We want to consider quite general Markov chains, though we have to impose restrictions to avoid a trivial representation like

$$Y_n := (\dots, \xi_{n-1}, \xi_n), \quad \xi_n = f(Y_n), \tag{1.2}$$

* Partially supported by N.S.F. grant M.C.S. 81–01583

with f denoting projection onto the last coordinate. Of course Y in (1.2) is a Markov chain. Nevertheless (1.2) does not describe a useful representation because the random variables Y_n retain all information about the past, which is impractical. So we want to investigate here representations with Markov chains having 'loss of memory'. For this reason we impose a well-known irreducibility assumption on Y, to be formulated later. Our aim is to construct examples of stationary processes ξ having some nice properties (e.g. very strong mixing properties) that cannot be represented as in (1.1) with such Y.

Let us now describe the probabilistic structure of our examples. First let $\mathcal{N}$ be the class of stationary processes which can be represented as an instantaneous function of a stationary, finite-state, irreducible, aperiodic Markov chain. Such processes are well known to have very nice asymptotic properties, including very strong mixing properties. We shall construct a stationary process $R := (R_n)_{n\in\mathbb{Z}}$ of the form

$$R_n := (X_n^{(1)}, X_n^{(2)}, \ldots), \quad n \in \mathbb{Z},$$

where for each k, $X^{(k)} := (X_n^{(k)})_{n\in\mathbb{Z}}$ is a process in the class $\mathcal{N}$. We may see R_n as describing 'reality' at time n, and the X-variables as giving the various aspects of 'reality'. The processes $X^{(k)}$ will be independent of each other. This simplifies the structure (but is perhaps unrealistic). The (infinite dimensional) random vector R_n consists of countably many random variables. A statistician is interested in much less information. Say he observes only

$$\xi_n = g(R_n),$$

where g is some countably-valued function. The process $\xi := (\xi_n)_{n\in\mathbb{Z}}$ is stationary. We shall show that there need not exist an irreducible-Markov representation for ξ. In our examples g will have a simple form, such that ξ_n depends only on finitely many components of R_n. Of course the number of components on which ξ_n depends will not be bounded (for otherwise we would have $\xi \in \mathcal{N}$).

The examples are presented to show that (nontrivial) Markov representations are not always correct in circumstances that seem quite reasonable. The few examples that we present do not seem to enable one to get a general picture of when Markov representation can or cannot be used. Nevertheless results as presented here lead us to emphasize the importance of a theory of statistical inference for stationary processes (e.g. central limit theory) where no Markov assumptions are present.

Already earlier studies were concerned with Markov representation. We can mention Johnson [4] who discussed representation from a very general and only slightly related point of view. Rosenblatt [7] surveys literature on representation in terms of finite-state Markov chains and mentions a necessary and sufficient condition for such a representation.

Let us now formulate the assumption we impose on the Markov chains Y that we consider in relation to representation (1.1). We assume

$$\begin{aligned}&\text{(i) } Y \text{ is a real, stationary Markov chain, and}\\&\text{(ii) } Y \text{ is irreducible (in the sense of [6]) with respect to the}\\&\quad\text{distribution } \pi \text{ of } Y_0.\end{aligned} \tag{1.3}$$

The assumption (1.3)(i) is not very restrictive. Because we are only interested in representation, if the state space of a Markov chain Y can be imbedded bimeasurably in the real line it is for our purpose real-valued. Note also that stationary, positive-recurrent Markov chains with a countable state space satisfy in essence our assumptions.

The assumption (1.3)(ii) means that for every real number x and every Borel set B with $\pi(B) > 0$ one has

$$P^n(x, B) > 0 \quad \text{for some } n \geq 1.$$

Here $P^n(\cdot, \cdot)$ denotes the nth iterate of a regular transition probability for Y. This assumption is certainly restrictive. For example in statistical mechanics Markov processes are used to describe time evolution of configurations and for these processes (1.3) is typically not satisfied. The assumptions in (1.3) are however quite natural in the following sense. The irreducibility in (1.3) generalizes the corresponding notion used for Markov chains with countable state space in such a way that the well-known limit theorems for transition probabilities carry over (see Orey [6]). These limit theorems correspond to a nice kind of 'loss of memory' and imply a mixing-type property for Y. Let us also note that (1.3)(ii) could be replaced by the seemingly stronger assumption of Harris recurrence i.e. if $x \in \mathbb{R}$ and B is a Borel set with $\pi(B) > 0$ then $P(Y_n \in B$ for infinitely many $n \geq 1 \mid Y_0 = x) = 1$; this can be deduced (with a little work) from Orey [6, p. 38, Theorem 8.1] and our assumption of stationarity.

Denote by $\mathcal{M}$ the class of processes ξ that have the same distribution as processes that can be represented as an instantaneous function of a Markov chain Y satisfying (1.3).

Though $\mathcal{M}$ is a large class it certainly does not contain all stationary processes. By our requirements (1.3) the processes in $\mathcal{M}$ satisfy a mixing property, as is well known. Our aim is to show that also assumptions of a different nature are implicit in the restriction to $\mathcal{M}$. But let us first describe this mixing-type property. Assume for the moment that Y satisfies (1.3) and is aperiodic. Then using Orey [6, p. 30, Theorem 7.1] it is easily seen that Y is strongly mixing, and if (1.1) holds then also ξ is strongly mixing. This argument can be used to show that in fact Y and ξ are absolutely regular. Absolute regularity is a lesser known, stronger mixing property, discussed e.g. in Volkonskii and Rozanov [10] and, under the name 'weak Bernoulli', in Shields [8]. We assumed Y is aperiodic; the argument above is however easily adapted to cover the periodic case too, and we leave the reader to formulate which restriction of a similar nature it implies for $\xi \in \mathcal{M}$ in general.

We use the notation P_Z for the distribution of a random vector Z. If a term like a_n is a subscript or superscript, it is usually written $a(n)$.

We want to develop necessary conditions for $\xi \in \mathcal{M}$. Markov chain theory leads easily to an interesting condition for $\xi \in \mathcal{M}$, as follows. Suppose Y satisfies (1.3). From irreducibility we have by Orey [6, p. 7, Theorem 2.1] that there exists a (positive) measure $\phi \neq 0$ (meaning $\phi(\mathbb{R}) > 0$) and an integer $k > 0$ such that

$$P_{Y(0),Y(k)} \geq \phi \times \phi. \tag{1.4}$$

If (1.1) holds then we also have

$$P_{\xi(0),\xi(k)} \geq \tilde{\phi} \times \tilde{\phi} \tag{1.5}$$

where $\tilde{\phi} := \phi \circ f^{-1}$. Thus in order that $\xi \in \mathcal{M}$ there must exist a measure $\tilde{\phi} \neq 0$ such that (1.5) holds for some $k > 0$. In Section 3 we discuss a process that violates this condition. The reader may verify easily that such processes are necessarily uncountably-valued. To remedy for this we derive in Section 4 a more restrictive necessary condition for $\xi \in \mathcal{M}$, to be used in our discussion of countably-valued processes.

The examples that we construct are ϕ-mixing and have an even stronger mixing property. Define the *ψ^*-dependence* between two σ-fields of a probability space by

$$\psi^*(\mathcal{A}, \mathcal{B}) = \sup \frac{P(A \cap B)}{P(A)P(B)}, \quad A \in \mathcal{A}, B \in \mathcal{B}, P(A) > 0, P(B) > 0. \tag{1.6}$$

Obviously $\psi^*(\mathcal{A}, \mathcal{B}) \geq 1$ and equality holds if and only if $\mathcal{A}$ and $\mathcal{B}$ are independent σ-fields. A stationary process ξ will be called *ψ^*-mixing* if its past and future are asymptotically independent in the sense that

$$\psi_n^* := \psi^*(\mathcal{B}(\xi_k, k \leq 0), \mathcal{B}(\xi_k \geq n)) \to 1 \quad \text{as } n \to \infty.$$

Here the notation $\mathcal{B}(\xi_k, k \in K)$ means the Borel σ-field of events generated by the family of r.v.'s $(\xi_k, k \in K)$, K being any set of integers. To avoid ambiguity when other stationary sequences are present, we sometimes write $\psi_n^*(\xi)$ instead of ψ_n^*.

Our main result is stated as follows:

Theorem 1.1. *There exists a stationary countably-valued process ξ such that*

(i) *$\xi \notin \mathcal{M}$, so ξ cannot be represented as an instantaneous function of a Markov chain satisfying* (1.3)

(ii) *$\psi_n^* - 1 \to 0$ with exponential rate as $n \to \infty$.*

We shall discuss three examples of stationary ψ^*-mixing processes that do not belong to the class $\mathcal{M}$. The first and simplest one, which we shall call X, has the structure of the process R mentioned above. It has exponential mixing rate (as in Theorem 1.1(ii)) but is uncountably-valued. Its purpose is to help clarify the second and third examples. The process X will be constructed at the end of this section and is discussed in Sections 2 and 3.

The second example is the process ξ of Theorem 1.1. It will be constructed and studied in Secton 5. Section 4 develops a criterion that will be used to show $\xi \notin \mathcal{M}$.

The third example, discussed in Section 6, will also be countably-valued and will have finite entropy; its mixing rate will be slower than exponential.

Construction of the process X

As 'building blocks' we shall use a class of simple finite-state Markov chains. For each $m \geq 3$ let $\mathcal{S}(m)$ denote the distribution of a stationary Markov chain $W := (W_n)_{n\in\mathbb{Z}}$ with state space $\{1, 2, \ldots, m\}$, with invariant marginal distribution $(1/m, \ldots, 1/m)$, and with one-step transition probabilities given by

$$p_{ij} = 0 \qquad \text{if } i = j,$$
$$= \frac{1}{m-1} \text{otherwise.} \tag{1.7}$$

Such a process has small and rapidly decreasing ψ^*-mixing coefficients, especially if m is large (Lemma 2.1). Also note that $W_0 \neq W_1$ a.s.

Let us specify the integers

$$m_k := 9^k \quad \forall k \geq 1. \tag{1.8}$$

For each $k \geq 1$ let $X^{(k)} := (X_n^{(k)})_{n\in\mathbb{Z}}$ be a process such that the subsequence $(X_{nk}^{(k)})_{n\in\mathbb{Z}}$ has the distribution $\mathcal{S}(m_k)$ and is independent of the family of r.v.'s $(X_n^{(k)}: n \not\equiv 0 \bmod k)$ outside of this subsequence; we also require that $X^{(k)}$ be stationary, and thus its distribution is completely determined. Also we assume that $X^{(1)}, X^{(2)}, \ldots$ are independent processes.

The process $X := (X_n)_{n\in\mathbb{Z}}$ is defined by

$$X_n := (X_n^{(1)}, X_n^{(2)}, \ldots) \quad \forall n \in \mathbb{Z}. \tag{1.9}$$

Clearly X is not countably-valued. The random variable X_0 is of course dependent on the 'past', $(\ldots, X_{-2}, X_{-1})$. Note however that the kth component $X_0^{(k)}$ depends on the past only via $X_{-k}^{(k)}$ and in particular $X_0^{(k)} \neq X_{-k}^{(k)}$ a.s. One might say that the process is built such that it 'learns' not to attain certain values in certain situations. This viewpoint suggests a formulation of the process as a learning model as discussed in [3].

In Sections 2 and 3 we show that X is exponentially ψ^*-mixing. A quite simple argument based on the fact (noted above) that

$$X_j^{(k)} \neq X_{j+k}^{(k)} \text{ a.s.} \quad \forall k \geq 1 \ \forall j \in \mathbb{Z} \tag{1.10}$$

will be used to disprove (1.5) and thus show that $X \notin \mathcal{M}$; this is done in Lemma 3.2.

The countably-valued process ξ of Section 5 (our second example) will be obtained from X as follows:

$$\xi_n = (\lambda_n, X_n^{(\lambda(n))}), \quad n \in \mathbb{Z}.$$

Here $\lambda := (\lambda_n)_{n\in\mathbb{Z}}$ is a certain i.i.d. sequence, independent of X and with values in the positive integers. With a little work the reader will be able to show that this process has the form

$$\xi_n = g(R_n), \quad n \in \mathbb{Z},$$

where (R_n) is a process having precisely the structure described earlier and $\xi_n = g(R_n)$ depends only on a finite (random) number of components of R_n. The third example, given in Section 6, will have a quite similar structure.

2. ψ^*-Mixing

First we study the mixing rates of the finite-state Markov chains with the distributions $\mathcal{S}(m)$, $m \geq 3$, for which the transition probabilities are given in (1.7).

Lemma 2.1. *If $m \geq 3$ then a Markov chain $W := (W_n)_{n\in\mathbb{Z}}$ with the distribution $\mathcal{S}(m)$ is exponentially ψ^*-mixing, such that*

$$\log \psi_n^*(W) \leq (2/m^{1/2})^n \quad \forall n \geq 1.$$

This inequality is crude but simple; in fact we shall use it only for $m \geq 9$ (the smallest m_k in (1.8)).

Proof. The transition probability matrix $\mathbb{P} := (p_{ij})$ in (1.7) can be written as

$$\mathbb{P} = [m/(m-1)]J_m - [1/(m-1)]I_m$$

where I_m is the $m \times m$ identity matrix and J_m is the $m \times m$ matrix with all entries equal to $1/m$. Using induction and the fact that $J_m^2 = J_m$, we have that

$$\mathbb{P}^n = J_m + [-1/(m-1)]^n (I_m - J_m) \quad \forall n \geq 1.$$

For each n the diagonal elements of $\mathbb{P}^n$ are equal to some common value d_n and the off-diagonal elements are equal to some value c_n. For each n one can show that

$$\psi_n^*(W) = \psi^*(\mathcal{B}(W_0), \mathcal{B}(W_n)) = \max_{1\leq i,j\leq m} \frac{P(W_0 = i,\ W_n = j)}{P(W_0 = i)\cdot P(W_n = j)}$$

$$= m \max(c_n, d_n).$$

The first equality here follows from the Markov property, the second can be proved with an elementary argument, and the third is trivial.

Since $m \geq 3$ (by assumption) we have that, for odd $n \geq 1$,

$$d_n < c_n = \frac{1}{m}(1 + [1/(m-1)]^n)$$

$$\leq \frac{1}{m}(1 + [1/m^{1/2}]^n)$$

and, for even $n \geq 2$,

$$c_n < d_n = \frac{1}{m}(1+(m-1)[1/(m-1)]^n)$$
$$\leq \frac{1}{m}(1+m[2/m]^n) \leq \frac{1}{m}(1+[2/m^{1/2}]^n).$$

Hence $\psi_n^*(W)-1 \to 0$ at the general rate $[1/(m-1)]^n$, and we also have $\psi_n^*(W) \leq (1+[2/m^{1/2}]^n)$ $\forall n \geq 1$, which implies $\log \psi_n^*(W) \leq (2/m^{1/2})^n$. □

The next step is to use Lemma 2.1 to get bounds on the mixing rate for each of the processes $X^{(k)}$, $k \geq 1$ (see (1.9)); this will be done in Section 3. Because these processes $X^{(k)}$, $k \geq 1$, are independent we have

$$\psi_n^*(X) = \prod_{k \geq 1} \psi_n^*(X^{(k)}) \quad \forall n \tag{2.1}$$

by Lemma 2.2 below, and (2.1) will be used in Section 3 to get an exponential bound on the mixing rate for the process X.

Lemma 2.2. *Suppose $\mathcal{A}_n$ and $\mathcal{B}_n$, $n=1, 2, \ldots$ are σ-fields. If the σ-fields $\mathcal{A}_n \vee \mathcal{B}_n$, $n = 1, 2, \ldots$ are independent then*

$$\psi^*\left(\bigvee_{n\geq 1} \mathcal{A}_n, \bigvee_{n \geq 1} \mathcal{B}_n\right) = \prod_{n \geq 1} \psi^*(\mathcal{A}_n, \mathcal{B}_n).$$

The proof is elementary and is sketched in Bradley [2, Lemma 1].

3. The properties of the example X

Two properties of the (uncountably-valued) process X defined by (1.9) are given here.

Lemma 3.1. $\psi_n^*(X)-1 \to 0$ *exponentially as* $n \to \infty$.

Proof. For each fixed $k \geq 1$ the process $X^{(k)}$ can be split up into subsequences $(X_{jk+i}^{(k)})_{j\in\mathbb{Z}}$ for $i=1, 2, \ldots, k$. These subsequences are independent and have the distribution $\mathcal{S}(m_k)$. Let W be any process with the distribution $\mathcal{S}(m_k)$. If n is any positive integer, then it can be written as $n = jk+i$ where $1 \leq i \leq k$ and $j \geq 0$, and we have

$$\log \psi_n^*(X^{(k)}) \leq \log \psi_{jk+1}^*(X^{(k)}) = k \log \psi_{j+1}^*(W) \leq k[2/m_k^{1/2}]^{j+1}$$
$$= k(2/3^k)^{j+1} \leq (2/3)^{\max(n,k)}. \tag{3.1}$$

Here the first inequality is trivial. The first equality follows from Lemma 2.2 and the structure of $X^{(k)}$. The second inequality holds by Lemma 2.1 (see also (1.8)). The last inequality holds by the definition of j in terms of n.

By (2.1) we may conclude

$$\log \psi_n^*(X) \leq \sum_{k\geq 1} (2/3)^{\max(n,k)} = (n+2)(2/3)^n = o((3/4)^n) \quad \text{as } n\to\infty$$

and Lemma 3.1 follows. □

Lemma 3.2. $X \notin \mathcal{M}$.

Proof. The process X has its values in a space Γ of sequences of integers. As we mentioned in the first section, it suffices to show that there *cannot* exist a positive integer k and measure $\tilde{\phi} \neq 0$ on Γ such that (1.5) holds for X.

Suppose such a k and $\tilde{\phi}$ exist. Partition Γ as $\Gamma = \bigcup_i \Gamma_i$ where Γ_i consists of all sequences in Γ with i as their kth coordinate. Then $\tilde{\phi}(\Gamma_i) > 0$ for some i, and so by (1.5),

$$P(X_0^{(k)} = X_k^{(k)} = i) = P(X_0 \in \Gamma_i, X_k \in \Gamma_i) \geq \tilde{\phi} \times \tilde{\phi}(\Gamma_i \times \Gamma_i) > 0.$$

This contradicts the fact $X_0^{(k)} \neq X_k^{(k)}$ a.s. which holds by (1.10). Hence Lemma 3.2 holds. □

We have verified that X satisfies (i) and (ii) of Theorem 1.1. To prove $\xi \notin \mathcal{M}$, where ξ is the countably-valued process to be constructed in Section 5 (for Theorem 1.1), the argument in Lemma 3.2 cannot be used, as we noted earlier, because the existence of such a k and $\tilde{\phi}$ is automatic in the countable-state case. So in the next section we give another criterion which is similar to but stronger than (1.5).

4. Markov chains

Suppose Y satisfies (1.3). Because of (1.3(ii) there exists by Orey [6, p. 7, Theorem 2.1] a C-set, i.e. a Borel set C with $\pi(C) > 0$, an integer $m > 0$ and a number $c > 0$ such that

$$P^m(x, A) \geq c\pi(A) \quad \forall x \in C \ \ \forall A \subset C. \tag{4.1}$$

Here $P^m(\cdot,\cdot)$ denotes the m-step transition probability of Y as before, and it is understood that A is restricted to the class of Borel sets. Obviously (4.1) implies (1.4). The existence of a C-set has strong consequences for the distribution of a Markov process. In Orey [6] and also Nummelin [5] such sets play a central role in the study of the limit behavior of Y. We shall use another consequence of the existence of a C-set.

Lemma 4.1. *Suppose Y is a Markov chain satisfying* (1.3), *and let p denote its period. Then there exists a number $\gamma > 0$ and integers $m > 0$ and $n_0 > 0$ such that $p \mid m$ and for all $n \geq n_0$ with $p \mid n$ there exists a measure ϕ_n on $\mathbb{R}^{n+1}$ with $\phi_n(\mathbb{R}^{n+1}) = \gamma$ such that*

$$P_{Y(-n),Y(-n+1),\ldots,Y(0),Y(m),Y(m+1),\ldots,Y(m+n)} \geq \phi_n \times \phi_n. \tag{4.2}$$

Of course the restriction $p \mid n$ is superfluous for aperiodic Markov chains. The existence of a period p ($=1$ if Y is aperiodic) is a well-known property of stationary irreducible Markov chains; see Orey [6, p. 13, Theorem 3.1].

Proof. Let C be a C-set, and let m and c be as in (4.1). Also define the measure $\pi_C(\cdot) := \pi(\cdot \cap C)$. We have $p \mid m$ because p is the period of Y. If $n > 2m$ and if A and B are Borel sets then

$$\begin{aligned} P_{Y(0),Y(n)}(A \times B) &\geq \int_A \int_C \int_C P^m(z, B) P^{n-2m}(y, \mathrm{d}z) P^m(x, \mathrm{d}y) \pi(\mathrm{d}x) \\ &\geq \int_A \int_C \int_C c\pi_C(B) P^{n-2m}(y, \mathrm{d}z) c\pi(\mathrm{d}y) \pi_C(\mathrm{d}x) \\ &= c^2 \pi_C \times \pi_C(A \times B) P(Y_m \in C, Y_{n-m} \in C). \end{aligned}$$

Orey [6, p. 30, Theorem 7.1] implies that $P(Y_m \in C, Y_{n-m} \in C) \to p[\pi(C)]^2$ as $n \to \infty$ under the restriction $p \mid n$. Hence there exists $c' > 0$ and $n_0 > 0$ such that if $n \geq n_0$ and $p \mid n$ then $P_{Y(0),Y(n)} \geq c' \pi_C \times \pi_C$. Using this fact twice (with stationarity) a similar argument will show that there exists $c'' > 0$ such that if $n \geq n_0$ and $p \mid n$ then

$$P_{Y(-n),Y(0),Y(m),Y(m+n)} \geq c'' \pi_C \times \pi_C \times \pi_C \times \pi_C. \tag{4.3}$$

For each n define the measure ϕ_n on $\mathbb{R}^{n+1}$ by

$$\phi_n(B) := (c'')^{1/2} \int_{\mathbb{R}^2} P((Y_0, \ldots, Y_n) \in B \mid Y_0 = x, Y_n = y) \, \mathrm{d}\pi_C \times \pi_C(x, y)$$

for Borel sets $B \subset \mathbb{R}^{n+1}$. Then for $\gamma := (c'')^{1/2}[\pi(C)]^2$ we have that $\phi_n(\mathbb{R}^{n+1}) = \gamma \; \forall n$, and using the Markov property and (4.3) one proves (4.2). □

Remark 4.2. Suppose ξ is a stationary process satisfying (1.1). We noted earlier that (1.4) for Y implies (1.5) for ξ. Similarly the property of Y in Lemma 4.1 transfers to ξ, with the measures ϕ_n replaced by the obvious related measures $\tilde{\phi}_n$. (Thus a process ξ which *fails* to have this property cannot be in $\mathcal{M}$.)

Remark 4.3. ψ-Mixing is a property stronger than ψ^*-mixing. A stationary ψ-mixing process has the properties referred to in the above remark, and it is an open question whether there are such processes outside $\mathcal{M}$. Even for 1-dependent processes this question is open. (For the definition of ψ-mixing see Bradley [2]. The first sentence in [2] contains a minor error; the '*-mixing' condition defined by Blum, Hanson, and Koopmans [1] is closely related to ψ-mixing but it is not quite the same.)

5. Proof of Theorem 1.1

To construct the process ξ for Theorem 1.1 we consider again the process X defined by (1.9). Suppose $\lambda := (\lambda_n)_{n\in\mathbb{Z}}$ is an i.i.d. sequence which takes its values on the set of positive integers and which is independent of X and so independent of the processes $X^{(1)}, X^{(2)}, \ldots$. The stationary process $\xi := (\xi_n)_{n\in\mathbb{Z}}$ defined by

$$\xi_n := (\lambda_n, X_n^{(\lambda(n))}) \quad \forall n \in \mathbb{Z} \tag{5.1}$$

is countably-valued. By (1.10) we have

$$P(\xi_n = \xi_{n-k} \text{ and } \lambda_n = \lambda_{n-k} = k) = 0 \quad \forall k \geqslant 1 \ \forall n \in \mathbb{Z}. \tag{5.2}$$

Below we shall specify the distribution of λ_0. Property (5.2) will then be used along with Lemma 4.1 and Remark 4.2 in order to show that ξ cannot belong to $\mathcal{M}$.

But first let us quickly show that Theorem 1.1(ii) holds (regardless of the distribution of λ_0). Defining the process $Z := (Z_n)_{n\in\mathbb{Z}}$ by $Z_n := (\lambda_n, X_n)$ $\forall n \in \mathbb{Z}$, we have

$$\psi_n^*(\xi) \leqslant \psi_n^*(Z) = \psi_n^*(\lambda) \cdot \psi_n^*(X). \tag{5.3}$$

The first inequality holds because ξ_n is Z_n-measurable (for each fixed n), and the latter equality holds by Lemma 2.2. Because λ is i.i.d., $\psi_n^*(\lambda) = 1$ and using Lemma 3.1 we obtain Theorem 1.1(ii).

To prove Theorem 1.1(i) we impose the following restrictions on the r.v. λ_0 and on a set K:

(i) $P(\lambda_0 = k) = p_k \quad \forall k \in K, \qquad P(\lambda_0 \notin K) = 0. \qquad \left(\sum_{k\in K} p_k = 1.\right)$

(ii) $kp_k \to \infty$ as $k \to \infty$ along K. (5.4)

(iii) K is a set of positive integers such that for each integer $p > 0$ the set K contains arbitrarily large multiples of p.

For example one could take

$$K := \{1^2, 2^2, 3^2, \ldots\}, \qquad p_k := n^{-1/2} - (n+1)^{-1/2} \quad \text{for } n^2 = k \in K.$$

With ξ defined by (5.1) and λ such that (5.4) holds we have:

Lemma 5.1. $\xi \notin \mathcal{M}$.

Proof. The sequence ξ has its values in $\mathbb{Z}^2$. Suppose ξ has the form (1.1) with Y a Markov chain satisfying (1.3) with period p. By Lemma 4.1 and Remark 4.2, there exist $\gamma > 0$ and integers m, $n_0 > 0$ such that $p \mid m$ and for all $n \geqslant n_0$ with $p \mid n$ there exists a measure $\tilde{\phi}_n$ on $(\mathbb{Z}^2)^{n+1}$ with total mass γ such that

$$P_{\xi(-n),\ldots,\xi(0),\xi(m),\ldots,\xi(m+n)} \geqslant \tilde{\phi}_n \times \tilde{\phi}_n. \tag{5.5}$$

Let $k=n+m$ where $n \geq n_0$, $k \in K$, k and n are both multiples of p, and k is sufficiently large, such that

$$P(\lambda_j = k \text{ for some } -n \leq j \leq 0) = 1-(1-P(\lambda_0 = k))^{n+1} > 1-\gamma^2. \tag{5.6}$$

This is possible because by (5.4),

$$(n+1)P(\lambda_0 = k) = (k-m+1)p_k \to \infty$$

as $k \to \infty$ along $k \in K$.

With k and n fixed as above, let Λ be a subset of $(\mathbb{Z}^2)^{n+1}$ such that the following equality of events holds:

$$\{\lambda_j = k \text{ for some } -n \leq j \leq 0\} = \{(\xi_{-n}, \ldots, \xi_0) \in \Lambda\}. \tag{5.7}$$

Because by (5.6) and (5.5)

$$\gamma^2 > P((\xi_{-n}, \ldots, \xi_0) \notin \Lambda) \geq \tilde{\phi}_n \times \tilde{\phi}_n(\Lambda^c \times (\mathbb{Z}^2)^{n+1}) = \gamma\tilde{\phi}_n(\Lambda^c)$$

we have $\tilde{\phi}_n(\Lambda^c) < \gamma$, and so $\tilde{\phi}_n$ has positive mass on the (countable) set Λ. Take $y \in \Lambda$ with $\tilde{\phi}_n(\{y\}) > 0$. By (5.5) the event

$$\{(\xi_{-n}, \ldots, \xi_0) = y, (\xi_m, \ldots, \xi_{m+n}) = y\}$$

has probability at least $[\tilde{\phi}_n(\{y\})]^2 > 0$. On this event $\xi_j = \xi_{j+k}$ for all $-n \leq j \leq 0$ and moreover because $y \in \Lambda$ there is by (5.7) such a j with $k = \lambda_j$ $(=\lambda_{j+k})$. Hence for this j with positive probability

$$\xi_j = \xi_{j+k} \quad \text{and} \quad \lambda_j = \lambda_{j+k} = k$$

which contradicts (5.2). So $\xi \notin \mathcal{M}$. □

Thus we have proved Theorem 1.1.

6. A finite entropy example

The *entropy* $H(Z)$ of a countably-valued random variable Z is defined as

$$H(Z) := \sum q_i \log_2(1/q_i)$$

where $q_i = P(Z=i)$ and i runs over the values in the range of Z with $q_i > 0$.

We construct a ψ^*-mixing stationary process ξ with $H(\xi_0) < \infty$ that does not belong to $\mathcal{M}$. This process has the form (5.1) except that we choose integers m_k, $k \geq 1$, different from (1.8). The distribution of λ_0 satisfying (5.4) will also be chosen more carefully.

Because λ_0 is ξ_0-measurable we have by a familiar rule for entropy (see Smorodinsky [9, Theorem 4.12a]) that

$$H(\xi_0) = H(\lambda_0) + \sum_{k \in K} H(\xi_0 | \lambda_0 = k)P(\lambda_0 = k)$$

where $H(\xi_0 | \lambda_0 = k)$ denotes the entropy of ξ_0 under the conditional probability $P(\cdot | \lambda_0 = k)$. By (5.1) and because $X_0^{(k)}$ is independent of the event $\{\lambda_0 = k\}$ and attains m_k values, each with the same probability, we have $H(\xi_0 | \lambda_0 = k) = \log_2 m_k$ and so

$$H(\xi_0) = H(\lambda_0) + E(\log_2 m_{\lambda(0)}). \tag{6.1}$$

Obviously an exponential choice for m_k (as in (1.8)) would make $H(\xi_0) = \infty$ by (5.4)(ii). But let us choose

$$K := \{n^n, n \geq 3\}, \qquad p_k := c/n^{n-1} \quad \text{for } n^n = k \in K$$

for some normalizing constant $c > 0$. Then we have (5.4) and we can take m_k quite large such that as $k \to \infty$ along K,

$$\log_2 m_k \approx n^{n-3} \quad \text{for } n^n = k \in K$$

and then one concludes easily that $H(\xi_0) < \infty$ using (6.1). By Lemma 5.1 (whose proof holds verbatim in this new context) we have $\xi \notin \mathcal{M}$. Using (5.3) and an argument like Lemma 3.1 one can prove that ξ is ψ^*-mixing, i.e. $\psi_n^*(\xi) - 1 \to 0$, but with a rate that is slightly slower than exponential.

Remark 6.1. It seems clear that one can construct a two-state stationary process $\xi \notin \mathcal{M}$ that still satisfies the absolute regularity condition. A binary coding of an example like the one above, of course with entropy less than 1, might achieve this. Because of the technical complications this will not be investigated here. A stronger mixing property like ϕ-mixing or ψ^*-mixing might be attainable. However this is complicated by the fact that the coding of a single ξ-value may affect a long stretch of time extending far into both the negative and positive indices.

Acknowledgement

The authors would like to thank the referee for his useful comments.

References

[1] J.R. Blum, D.L. Hanson and L.H. Koopmans, On the strong law of large numbers for a class of stochastic processes, Z. Wahrsch. Verw. Gebiete 2 (1963) 1–11.

[2] R.C. Bradley, On the ψ-mixing condition for stationary random sequences, Trans. Amer. Math. Soc. **276** (1983) 55–66.

[3] M. Iosifescu and R. Theodorescu, Random processes and learning (Springer, Berlin, 1969).

[4] D.P. Johnson, Representations and classifications of stochastic processes, Trans. Amer. Math. Soc. 188 (1974) 179–197.

[5] E. Nummelin, A splitting technique for Harris recurrent Markov chains, Z. Wahrsch. Verw. Gebiete 43 (1978) 309–318.

[6] S. Orey, Lecture notes on limit theorems for Markov chain transition probabilities (Mathematical studies 34, Van Nostrand Reinhold, New York, 1971).
[7] M. Rosenblatt, Markov processes, structure and asymptotic behavior (Springer, Berlin, 1971).
[8] P. Shields, The theory of Bernoulli shifts (University of Chicago Press, 1973).
[9] M. Smorodinsky, Ergodic theory, entropy (Lecture notes in Mathematics 214, Springer, Berlin, 1971).
[10] V.A. Volkonskii and J.A. Rozanov, Some limit theorems for random functions I, Theory Prob. Appl. 4 (1959) 178–197.