TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 338, Number 1, July 1993

MIXING PROPERTIES OF A CLASS OF BERNOULLI-PROCESSES

DORIS FIEBIG

ABSTRACT. We prove that stationary very weak Bernoulli processes with rate $O(1/n)$ (VWB $O(1/n)$) are strictly very weak Bernoulli with rate $O(1/n)$. Furthermore we discuss the relation between VWB $O(1/n)$ and the classical mixing properties for countable state processes. In particular, we show that VWB $O(1/n)$ implies ϕ-mixing.

0. INTRODUCTION

Let $X_i: (\Omega, \mathscr{A}, \mu) \to (S, \mathscr{B})$, $i \in \mathbf{Z}$, be a stationary sequence of random variables on a probability space $(\Omega, \mathscr{A}, \mu)$ with values in a Polish space $(S, \mathscr{B})$. In this setting we define very weak Bernoulli processes with rate $\varepsilon(n)$, denoted by VWB $\varepsilon(n)$, and strictly VWB $\varepsilon(n)$ processes. It was shown by Dehling, Denker and Philipp [D.D.P] that $O(1/n)$ is the fastest rate for which nonindependent VWB-processes exist. We show that $(X_i)_{i\in\mathbf{Z}}$ is VWB $O(1/n)$ iff $(X_i)_{i\in\mathbf{Z}}$ is strictly VWB $O(1/n)$. This strengthens the result of Eberlein, that real-valued strictly VWB $O(1/n)$ processes with certain moment conditions satisfy an almost sure invariance principle [E]. Then we restrict ourselves to the discrete case, i.e., we assume S to be countable and that $\mathscr{B}$ is generated by the discrete metric. Our main result in this case is that VWB $O(1/n)$ implies ϕ-mixing, which improves an earlier result of [D.D.P]. We show that VWB $O(1/n)$ gives no constraints on the ϕ-mixing rate, and that VWB $O(1/n)$ does not imply ψ-mixing. After that we give a new upper bound for the Wasserstein-distance, which implies that a ϕ-mixing process with ϕ-mixing rate $\phi(i)$ is strictly VWB with rate $\frac{1}{n}\sum_{i=1}^{n}\phi(i)$; in particular ϕ-mixing processes with summable rates are VWB $O(1/n)$.

1. VWB $O(1/n)$ IMPLIES STRICTLY VWB $O(1/n)$

Let $X_i : (\Omega, \mathscr{A}, \mu) \to (S, \mathscr{B})$, $i \in \mathbf{Z}$, be a stationary sequence of random variables. Let $\sigma : S\times S \to \mathbf{R}$ be a metric, such that $\mathscr{B}$ is generated by σ and S is a Polish space. For $-\infty \leq m \leq n \leq \infty$ let $\mathscr{A}_m^n = \mathscr{A}(X_i,\ m \leq i \leq n)$ be the σ-algebra generated by X_i with indices between m and n. For two probability measures ν_1, ν_2 on (S^n, B^n) let $P_n(\nu_1, \nu_2) = \{\lambda: B^n \times B^n \to [0, 1]: \lambda$ is a probability measure with ith marginal ν_i, $i = 1, 2\}$. So $P_n(\nu_1, \nu_2)$ is the set of joinings of ν_1 and ν_2. Then, for $Z \in \mathscr{A}_{-\infty}^{0}$ with $\mu(Z) > 0$, define the

Received by the editors February 5, 1990 and, in revised form, April 13, 1991.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60G10; Secondary 60F99.

Wasserstein-distance

$$\rho_n(\mu, \mu(\cdot/Z)) := \inf \int_{S^n \times S^n} \frac{1}{n} \sum_{i=1}^{n} \sigma(x_i, y_i)\, d\lambda(x_1, \dots, x_n, y_1, \dots, y_n)$$

where the infimum is taken over $\lambda \in P_n((X_1, \dots, X_n)\mu, (X_1, \dots, X_n)\mu(\cdot/Z))$.

Definition 1 [E]. $(X_i)_{i\in\mathbf{Z}}$ is very weak Bernoulli with rate $\varepsilon(n)$ (VWB $\varepsilon(n)$) iff

(1) $\varepsilon(n) \to 0$, $n \to \infty$,

(2) $\forall n \in \mathbf{N}$ $\forall m \in \mathbf{Z}^+$ $\exists D = D(m, n) \in \mathscr{A}_{-m}^{0}$ with

$$\mu(D) \geq 1 - \varepsilon(n), \tag{1.1}$$

$$A \subset D,\ A \in \mathscr{A}_{-m}^{0},\quad \mu(A) > 0 \Rightarrow \rho_n(\mu, \mu(\cdot/A)) \leq \varepsilon(n). \tag{1.2}$$

Definition 2 [E]. $(X_i)_{i\in\mathbf{Z}}$ is strictly VWB $\varepsilon(n)$ iff $(X_i)_{i\in\mathbf{Z}}$ is VWB $\varepsilon(n)$ and all sets $D(m, n)$ can be chosen to be Ω, i.e., $\rho_n(\mu, \mu(\cdot/A)) \leq \varepsilon(n)$ $\forall A \in \mathscr{A}_{-\infty}^{0}$.

We shall tacitly assume $\mu(A) > 0$ when dealing with conditional probabilities as $\mu(\cdot/A)$.

In [D.D.P] it was shown that a VWB $\varepsilon(n)$ process with $\liminf n\varepsilon(n) = 0$ is already independent. This means that $\varepsilon(n) = O(1/n)$ is the fastest rate for which one can possibly have a nonindependent VWB $\varepsilon(n)$ process. Various classes of examples for VWB $O(1/n)$ processes are given in [F]. They include m-dependent processes, finite state mixing Markov chains and continuous factors of finite state mixing Markov chains. There it was shown that the VWB $O(1/n)$-property is not preserved under finitary factor maps, not even if the coding length of the factor map has moments of all orders [F].

Our main interest is the examination of VWB $O(1/n)$ processes. The fundamental observation is the following:

Theorem 3 [F]. *Let $(X_i)_{i\in\mathbf{Z}}$ be a stationary sequence of random variables with values in a Polish space. Let $0 \leq M < \infty$. Then:*

$(X_i)_{i\in\mathbf{Z}}$ is VWB with rate M/n iff $(X_i)_{i\in\mathbf{Z}}$ is strictly VWB with rate M/n.

We need

Lemma 4 [S]. *Let $M_n = \{\nu : (S^n, B^n) \to R : \nu$ probability measure$\}$ with weak topology. Then*

$$n\rho_n : M_n \times M_n \to \mathbf{R}$$
$$(\nu_1, \nu_2) \to n\rho_n(\nu_1, \nu_2)$$

is a lower semicontinuous function.

Proof of Theorem 3. Let $0 \leq M < \infty$. Let $(X_i)_{i\in\mathbf{Z}}$ be VWB with rate M/n. Then for all $n \in \mathbf{N}$, $m \in \mathbf{Z}^+$ sets $D(m, n) \in \mathscr{A}_{-m}^{0}$ can be chosen such that (1.1), (1.2) hold for $\varepsilon(n) = M/n$. We show that the process is strictly VWB M/n. Let $n \in \mathbf{N}$, $m \in \mathbf{Z}^+$ and choose a set $A \in \mathscr{A}_{-m}^{0}$ with $\mu(A) > 0$. Now pick $k_0 \geq n$ such that $\mu(A \cap D(m, k)) > \mu(A)/2 > 0$ $\forall k \geq k_0$. Then $\mu(\cdot/A \cap D(m, k)) \to \mu(\cdot/A)$ in weak topology. Because $A \cap D(m, k) \subset D(m, k)$ and $A \cap D(m, k) \in \mathscr{A}_{-m}^{0} \forall k$, (1.2) implies

$$\rho_k(\mu, \mu(\cdot/A \cap D(m, k))) \leq \frac{M}{k} \quad \forall k \geq k_0.$$

Since $n \le k_0$ we have

$$\rho_n(\mu, \mu(\cdot/A \cap D(m, k))) \le \frac{M}{n} \quad \forall k \ge k_0.$$

By Lemma 4 this implies $\rho_n(\mu, \mu(\cdot/A)) \le M/n$, so the process is strictly VWB M/n. The converse is trivial. □

Theorem 3 does not hold for rates slower than $O(1/n)$. This was shown in [F] and we recall here the example:

Let $(X_i)_{i\in\mathbf{Z}}$ be a Markov chain with state space $\mathbf{Z}^+$ and transition probabilities

$$p_{ij} = \begin{cases} 1, & i = j+1,\ j \ge 0, \\ c_j, & i = 0,\ j \ge 0, \\ 0, & \text{otherwise}, \end{cases}$$

where $(c_n)_{n\ge 0}$ is a sequence with $c_j \ge 0\ \forall j$, $c_j > 0$ infinitely often, $\sum_{j=1}^{\infty} jc_j < \infty$. Then the $(p_{ij})_{i,j\in\mathbf{Z}^+}$ define a stationary Markov chain. For stationary Markov chains one can calculate the exact value of $\rho_n(\mu, \mu(\cdot/X_0 = i))$ for all $n \in \mathbf{N}$, $i \in \mathbf{Z}^+$ (by Theorem 6). This gives the possibility by choosing $(c_j)_{j\ge 0}$ to achieve a given VWB rate $\varepsilon(n)$ with $n\varepsilon(n) \to \infty$. In [F] it was shown that the Markov chain above is not strictly VWB, i.e., there is no rate $\varepsilon(n)$ for which $(X_i)_{i\in\mathbf{Z}}$ is strictly VWB $\varepsilon(n)$.

2. Relating VWB $O(1/n)$ to the classical mixing properties

Now we want to examine the mixing properties of VWB $O(1/n)$ processes. Because $\varepsilon(n) = O(1/n)$ is the fastest rate for which one can have nonindependent VWB $\varepsilon(n)$ processes, and because of Theorem 3, one expects that these processes have good mixing properties, but this depends strongly on the state space S and the metric σ. There exists a stationary VWB $O(1/n)$ process with uncountable state space $S \subset \mathbf{R}$, where σ is the Euclidean metric, which is not even α-mixing [B1], but on the other hand finite state VWB $O(1/n)$ processes are always weak Bernoulli [D.D.P]. From now on we restrict ourselves to stationary processes with at most countable state space, endowed with the discrete metric. For $Z \in \mathscr{A}_{-\infty}^{0}$, $\mu(Z) > 0$ and $1 \le i \le n < \infty$ we define the distribution distance of names by

$$|\operatorname{dist} X_i^n \mu - \operatorname{dist} X_i^n \mu(\cdot/Z)| := \frac{1}{2} \sum_{(y_i, \dots, y_n)\in S^{n-i+1}} |\mu(X_n = y_n, \dots, X_i = y_i) - \mu(X_n = y_n, \dots, X_i = y_i/Z)|.$$

With this notation we have the simple, but extremely useful

Lemma 5. *Let $(n_i)_{i\in\mathbf{N}}$ be a strictly increasing sequence of natural numbers with $n_0 := 0$. Let $Z \in \mathscr{A}_{-\infty}^{0}$ with $\mu(Z) > 0$. Then for all $N \in \mathbf{N}$*

$$n_{N+1}\rho_{n_{N+1}}(\mu, \mu(\cdot/Z)) \ge \sum_{i=0}^{N} |\operatorname{dist} X_{n_i+1}^{n_{i+1}} \mu - \operatorname{dist} X_{n_i+1}^{n_{i+1}} \mu(\cdot/Z)|.$$

(If S is countable and the metric σ is bounded below by ε, i.e., $\sigma(x, y) \ge \varepsilon > 0\ \forall x \ne y$, then this lemma holds with the RHS multiplied by ε). The next theorem gives a new upper bound for the Wasserstein-distance.

Theorem 6. *Let* $n \in \mathbf{N}$, $Z \in \mathscr{A}_{-\infty}^{0}$ *with* $\mu(Z) > 0$. *Then*

$$n\rho_n(\mu, \mu(\cdot/Z)) \le \sum_{i=1}^{n} |\operatorname{dist} X_i^n \mu - \operatorname{dist} X_i^n \mu(\cdot/Z)|.$$

The proof of Theorem 6 is deferred to the Appendix. It depends on the construction of a joining ν_n of $(X_1, \dots, X_n)\mu$ and $(X_1, \dots, X_n)\mu(\cdot/Z)$ such that

$$\begin{aligned}(2.1) \qquad &\int_{S^n \times S^n} \sigma(x_i, y_i)\, d\nu_n((x_1, \dots, x_n), (y_1, \dots, y_n)) \\ &\le |\operatorname{dist} X_i^n \mu - \operatorname{dist} X_i^n \mu(\cdot/Z)| \quad \forall 1 \le i \le n.\end{aligned}$$

The joining ν_n is a generalisation of a construction in [F], and shows that for Markov chains

$$n\rho_n(\mu, \mu(\cdot/X_0 = x)) = \sum_{i=1}^{n} |\operatorname{dist} X_i \mu - \operatorname{dist} X_i \mu(\cdot/X_0 = x)|.$$

We use the following mixing coefficients:

$$\alpha(n) := \sup_{A \in \mathscr{A}_{-\infty}^{0}} \sup_{B \in \mathscr{A}_n^{\infty}} |\mu(B \cap A) - \mu(B)\mu(A)|,$$

$$\mathrm{WB}(n) := \sup_{m, k \ge 0} \sum_{B \in \mathscr{P}_n^{n+k}} \sum_{A \in \mathscr{P}_{-m}^{0}} \mu(A) \cdot |\mu(B|A) - \mu(B)|,$$

where $\mathscr{P}_n^{n+k}$ (resp. $\mathscr{P}_{-m}^{0}$) is the finest partition of Ω into sets $B \in \mathscr{A}_n^{n+k}$ (resp. $A \in \mathscr{A}_{-m}^{0}$).

$$\phi(n) := \sup_{A \in \mathscr{A}_{-\infty}^{0}} \sup_{B \in \mathscr{A}_n^{\infty}} |\mu(B/A) - \mu(B)|,$$

$$\psi(n) := \sup_{A \in \mathscr{A}_{-\infty}^{0}} \sup_{B \in \mathscr{A}_n^{\infty}} |\mu(B/A)/\mu(B) - 1|$$

(where as always $\mu(A) > 0$ is assumed, if necessary). $(X_i)_{i \in \mathbf{Z}}$ is said to be α-mixing (weak Bernoulli (= WB), ϕ-mixing or ψ-mixing) iff $\alpha(n) \to 0$ ($\mathrm{WB}(n) \to 0$, $\phi(n) \to 0$, $\psi(n) \to 0$), respectively. From the definitions of the mixing coefficients it is clear that

$$\psi\text{-mixing} \Rightarrow \phi\text{-mixing} \Rightarrow \mathrm{WB} \Rightarrow \alpha\text{-mixing}.$$

The reverse implications do not hold. For general background on the properties of these mixing coefficients see [B3]. We first strengthen the result of [D.D.P] to

Theorem 7. *Let* $(X_i)_{i \in \mathbf{Z}}$ *be a stationary process with at most countable state space* S *and discrete metric (or a metric bounded away from zero). Then* $(X_i)_{i \in \mathbf{Z}}$ VWB $O(1/n) \Rightarrow (X_i)_{i \in \mathbf{Z}}$ ϕ*-mixing.*

For the proof we need the following Lemma 8, which is an easy consequence of the observation that VWB with rate M/n implies ($m \in N$)

$$\begin{aligned} M &\ge (Nm)\rho_{Nm}(\mu, \mu(\cdot/D)) \\ &\ge \sum_{i=1}^{N} |\operatorname{dist} X_{(i-1)m+1}^{im} \mu - \operatorname{dist} X_{(i-1)m+1}^{im} \mu(\cdot/D)| \quad \text{by Lemma 5.}\end{aligned}$$

So that, given $\varepsilon > 0$, there is an $N \in \mathbf{N}$ such that for any set

$$\text{(22.)} \qquad \begin{aligned} &D \in \mathscr{A}_{-\infty}^{0},\ \mu(D) > 0 \text{ there is an } i \leq N \text{ with} \\ &|\operatorname{dist} X_{(i-1)m+1}^{im}\mu - \operatorname{dist} X_{(i-1)m+1}^{im}\mu(\cdot/D)| < \varepsilon. \end{aligned}$$

Lemma 8. *Let* $(X_i)_{i\in\mathbf{Z}}$ *be VWB with rate* $\varepsilon(n) = M/n$, $0 \leq M < \infty$. *Let* $\sigma : S^{\mathbf{Z}} \to S^{\mathbf{Z}}$ *be the shift map, i.e.,* $\sigma((s_j)_{j\in\mathbf{Z}})_i = s_{i+1}\ \forall i \in \mathbf{Z}$. *Fix* $r \in \mathbf{N}$, $m \in \mathbf{N}$. *Choose* $A_1, \dots, A_r \in \mathscr{A}_1^m$ *with* $\mu(A_s) > 0\ \forall s$. *Fix* $\delta > 0$. *Then there is* $k = k(\min_{1\leq s\leq r}\mu(A_s), \delta) \in \mathbf{N}$ *such that*:

$$\begin{aligned} &\forall D \in \mathscr{A}_{-\infty}^{0},\ \mu(D) > 0\ \exists 0 \leq i \leq k\ (i \textit{ depends on } D) \textit{ with} \\ &|\mu(\sigma^{-im}A_s/D) - \mu(A_s)| < \delta\mu(A_s) \quad \forall s \leq r. \end{aligned}$$

Proof. Choose $\varepsilon < \delta \min_{1\leq s\leq r}\mu(A_s)$, and apply the observation (2.2) above, using the fact that $|\mu(\sigma^{-im}A_s/D) - \mu(A_s)| \leq |\operatorname{dist} X_{im+1}^{(i+1)m}\mu - \operatorname{dist} X_{im+1}^{(i+1)m}\mu(\cdot/D)|$. □

Remark. Lemma 8 remains valid for strictly VWB $\varepsilon(n)$ processes, for all rates $\varepsilon(n)$.

Proof of Theorem 7. Let $(X_i)_{i\in\mathbf{Z}}$ be VWB $\varepsilon(n)$, $\varepsilon(n) = M/n$, $M < \infty$. Assume $(X_i)_{i\in\mathbf{Z}}$ is not ϕ-mixing.

Claim 1. $\forall m \in \mathbf{N}\ \ \forall\varepsilon > 0\ \ \exists 1 = l_0 < l_1 < l_2 < \cdots < l_m < \infty\ \ \exists k \in \mathbf{N}$ $\exists B_i \in \mathscr{A}_{l_{i-1}}^{l_i-1}$, $1 \leq i \leq m$ and $\exists C \in \mathscr{A}_{-k}^{0}$, $\mu(C) > 0$, such that $\mu(B_i) > 1 - \varepsilon$, $\mu(B_i/C) < \varepsilon\ \ \forall i \in \{1, \dots, m\}$.

We prove this claim by induction on m. For $m = 1$, we apply Theorem 1 of [B2], so $(X_i)_{i\in\mathbf{Z}}$ not ϕ-mixing means $\phi(1) = 1$. This implies the claim for $m = 1$, because one can approximate sets in $\mathscr{A}_1^{\infty}$ (resp. $\mathscr{A}_{-\infty}^{0}$) arbitrarily well by sets in $\mathscr{A}_1^{l}$ (resp. $\mathscr{A}_{-k}^{0}$) for l (resp. k) large enough.

Let $\varepsilon > 0$ and pick $0 < \delta < \varepsilon/3$.

By hypothesis there are $1 = l_0 < l_1 < \cdots < l_m < \infty$ and sets $B_i \in \mathscr{A}_{l_{i-1}}^{l_i-1}$, $1 \leq i \leq m$, $C \in \mathscr{A}_{-s}^{0}$, $\mu(C) > 0$ with $\mu(B_i) > 1 - \delta$, $\mu(B_i/C) < \delta$. We shall show that there are sets B_{m+1} and E and that $B_1, \dots, B_m$, B_{m+1} and E satisfy the claim for $m + 1$ and ε. Let $I := \{i \in \{1, \dots, m\} : \mu(B_i/C) > 0\}$, $A_i := \sigma^{-s-1}(B_i \cap C)$, $1 \leq i \leq m$. Then $A_i \in \mathscr{A}_1^{l_m+s}\forall i$. We apply Lemma 8 to the $\{A_i,\ i \in I\} \cup \{\sigma^{-s-1}C\}$ with $m' := l_m + s$. So we get $k = k(\mu(C), (\mu(A_i))_{i\in I}; \frac{1}{2})$ such that for any set $D \in \mathscr{A}_{-\infty}^{0}\exists 0 \leq j \leq k$ such that

$$\text{(2.3)} \qquad \begin{aligned} &|\mu(\sigma^{-jm'}A_i/D) - \mu(A_i)| < \tfrac{1}{2}\mu(A_i) \quad \forall i \in I, \\ &|\mu(\sigma^{-jm'-s-1}C/D) - \mu(C)| < \tfrac{1}{2}\mu(C). \end{aligned}$$

Now, because not ϕ-mixing means in particular $\phi(n) = 1\ \ \forall n$, we find for $0 < \delta_1 < \delta$ with $2\delta_1/\mu(C) < \delta$ a number $L \geq (2k + 1)m' + 2$ and sets

$$\text{(2.4)} \qquad B \in \mathscr{A}_{(2k+1)m'}^{L}, \quad D \in \mathscr{A}_{-L}^{0} \quad \text{with } \mu(B) > 1 - \delta_1, \quad \mu(B/D) < \delta_1.$$

Let $E := C \cap \sigma^{jm'+s+1}D$ where j is according to (2.3). Then $E \in \mathscr{A}_{-L-jm'-s-1}^{0}$, and $\mu(E) = \mu(D)\mu(\sigma^{-jm'-s-1}C/D) > \frac{1}{2}\mu(D)\mu(C) > 0$ by (2.3). Let $B_{m+1} := \sigma^{jm'+s+1}B$, so $B_{m+1} \in \mathscr{A}_{(2k-j)m'}^{L-jm'-s-1}$ and $(2k - j)m' \geq km' \geq l_m$, so for $l_{m+1} := L - jm' - s - 1$ we have

$$\text{(2.5)} \qquad B_{m+1} \in \mathscr{A}_{l_m}^{l_{m+1}} \quad \text{and} \quad \mu(B_{m+1}) > 1 - \delta.$$

For $i \in \{1, \dots, m\} - I$ we have

$$\mu(B_i/E) \leq \frac{\mu(B_i \cap C)}{\mu(E)} = 0 < \delta.$$

For $i \in I$ we have

$$\begin{aligned}\mu(B_i/E) &= \frac{\mu(B_i \cap C \cap \sigma^{jm'+s+1}D)}{\mu(E)} = \frac{\mu(D)\mu(\sigma^{-jm'}A_i/D)}{\mu(E)} \\ &= \frac{\mu(\sigma^{-jm'}A_i/D)}{\mu(\sigma^{-jm'-s-1}C/D)} \\ &\leq \frac{\frac{3}{2}\mu(A_i)}{\frac{1}{2}\mu(C)} = 3\mu(B_i/C) < 3\delta \quad \text{(because of (2.3))}\end{aligned}$$

and

$$\begin{aligned}\mu(B_{m+1}/E) &= \frac{\mu(\sigma^{jm'+s+1}B \cap C \cap \sigma^{jm'+s+1}D)}{\mu(C \cap \sigma^{jm'+s+1}D)} \leq \frac{\mu(B/D)\mu(D)}{\mu(C \cap \sigma^{jm'+s+1}D)} \\ &= \frac{\mu(B/D)}{\mu(\sigma^{-jm'-s-1}C/D)} < \frac{\delta_1}{\frac{1}{2}\mu(D)} < \delta \quad \text{(because of (2.3), (2.4)).}\end{aligned}$$

Because $3\delta < \varepsilon$ we have sets $B_1, \dots, B_{m+1}$ and E which satisfy Claim 1 for $m+1$ and ε. This proves Claim 1.

Now we choose $\varepsilon < \frac{1}{2}$ and $m \in \mathbf{N}$ such that $m(1-2\varepsilon) > M$. Then we choose sets B_i, C from Claim 1 to obtain by Lemma 5 the estimate

$$\begin{aligned}M &\geq l_m \rho_{l_m}(\mu, \mu(\cdot/C)) \geq \sum_{i=1}^{m} |\operatorname{dist} X_{l_{i-1}}^{l_i-1}\mu - \operatorname{dist}|X_{l_{i-1}}^{l_i-1}\mu(\cdot/C)| \\ &\geq \sum_{i=1}^{m} |\mu(B_i) - \mu(B_i/C)| \geq m(1-2\varepsilon) > M.\end{aligned}$$

This contradiction shows, $(X_i)_{i\in\mathbf{Z}}$ was, in fact, ϕ-mixing and proves the theorem. □

Remark. The key to the proof of Theorem 7 is Claim 1. In fact, one can prove Claim 1 for all strictly VWB $\varepsilon(n)$ processes, but of course, the fastest rate $\varepsilon(n) = O(1/n)$ was needed to produce a contradiction from Claim 1. We show in §3 that for each sequence $\varepsilon(n)$, $n\varepsilon(n) \to \infty$, $\varepsilon(n) \to 0$ there is a strictly VWB $\varepsilon(n)$ process which is not ϕ-mixing.

Theorem 7 is the strongest possible, since there exists VWB $O(1/n)$, a finite state process, which is not ϕ-mixing (see [F]).

Example 9. There exist a VWB $O(1/n)$ process with countable state space which is not ψ-mixing. Let $0 < p < 1$ and for i, $j \in \mathbf{Z}^+$ let

$$p_{ij} := \begin{cases} p, & \text{if } j = i+1, \ i \geq 0, \\ 1-p, & \text{if } i \geq 0, \ j = 0, \\ 0, & \text{otherwise.} \end{cases}$$

This stochastic matrix defines a stationary Markov chain $(X_i)_{i\in\mathbf{Z}}$ with state space $\mathbf{Z}^+$ and invariant measure μ, where $\mu(X_0 = i) = (1-p)\cdot p^i$, $i \geq 0$.

$(X_i)_{i\in\mathbf{Z}}$ is not ψ-mixing, because $\mu(X_n = n+1/X_0 = 0) = 0 \ \forall n$.

$(X_i)_{i\in\mathbf{Z}}$ is ϕ-mixing, as an easy calculation shows, so $(X_i)_{i\in\mathbf{Z}}$ is VWB $O(1/n)$, see Corollary 13.

The next theorem shows that VWB $O(1/n)$ has no constraints on the ϕ-mixing rate.

Theorem 10. *Let $(\lambda_n)_{n\geq 1}$ be a sequence with $\lambda_1 \leq 1$, $(\lambda_n)_{n\geq 1}$ nonincreasing, $\lambda_n \to 0$ as $n \to \infty$ and $-\log(1-\lambda_n)$ is convex on the set $\{k : \lambda_k < 1\}$. Then there exists a countable state process $(X_i)_{i\in\mathbf{Z}}$ which is* VWB $O(1/n)$ *and ϕ-mixing with $\frac{1}{2}\lambda_n \leq \phi(n) \leq \lambda_n$.*

Proof. Kesten and O'Brien have constructed an example in [K.O'B] (which we discuss in §3), where one easily checks that $\lambda_n = \mu(\bigcup_{k\geq n}\{U_k \geq k\})$. So $\lambda_n \to 0$ means $EU_0 < \infty$ in their construction. Apply Theorems 14 and 15. □

We do not expect the converse of Theorem 7 to be true, but we do have the following corollary from Theorem 6.

Corollary 11. *Let $(X_i)_{i\in\mathbf{Z}}$ be ϕ-mixing with ϕ-mixing rate $\phi(n)$, then $(X_i)_{i\in\mathbf{Z}}$ is strictly* VWB $\varepsilon(n)$ *for $\varepsilon(n) = \frac{1}{n}\sum_{i=1}^n \phi(i)$.*

Proof. Theorem 6 yields

$$\begin{aligned}\sup_{Z\in\mathscr{A}_{-\infty}^0} n\rho_n(\mu, \mu(\cdot/Z)) &\leq \sup_{Z\in\mathscr{A}_{-\infty}^0} \sum_{i=1}^n |\operatorname{dist} X_i^n\mu - \operatorname{dist} X_i^n\mu(\cdot/Z)| \\ &\leq \sup_{Z\in\mathscr{A}_{-\infty}^0} \frac{1}{2}\sum_{i=1}^n (|\mu(B_i^+) - \mu(B_i^+/Z)| + |\mu(B_i^-) - \mu(B_i^-/Z)|) \\ &\leq \sum_{i=1}^n \phi(i)\end{aligned}$$

where

$$B_i^+ := \{(y_i, \dots, y_n) : \mu(X_i = y_i, \dots, X_n = y_n) \geq \mu(X_i = y_i, \dots, X_n = y_n/Z)\}$$

and

$$\begin{aligned}B_i^- := \{(y_i, \dots, y_n) \colon \mu(X_i = y_i, \dots, X_n = y_n) \\ < \mu(X_i = y_i, \dots, X_n = y_n/Z)\}.\end{aligned} \quad \square$$

In particular, we have the following consequences.

Corollary 12. *If $(X_i)_{i\in\mathbf{Z}}$ is ϕ-mixing with $\sum_{i=1}^\infty \phi(i) < \infty$, then $(X_i)_{i\in\mathbf{Z}}$ is* VWB $O(1/n)$.

Corollary 13. *Let $(X_i)_{i\in\mathbf{Z}}$ be a stationary Markov chain with at most countable state space. Then $(X_i)_{i\in\mathbf{Z}}$ is* VWB $O(1/n)$ *iff $(X_i)_{i\in\mathbf{Z}}$ is ϕ-mixing.*

Proof. If $(X_i)_{i\in\mathbf{Z}}$ is ϕ-mixing then $\phi(n) = O(\lambda^n)$ for a $0 < \lambda < 1$ [R]. Apply Corollary 12. □

3. Some aspects of strictly VWB $\varepsilon(n)$ processes

We want to discuss a class of examples, which was given by Kesten and O'Brien in its original form. These examples will show that $\varepsilon(n) = O(1/n)$ is the only VWB rate which forces the process to be ϕ-mixing.

We use the notation of [K.O'B].

Let $(U_i)_{i\in\mathbf{Z}}$ be i.i.d. with values in $\mathbf{Z}^+$.

Let $(V_i)_{i\in\mathbf{Z}}$ be i.i.d. with values in $\{0, 1\}$, $\mu(V_0 = 0) = \frac{1}{2}$.

Let $(U_i)_{i\in\mathbf{Z}}$ be independent from $(V_i)_{i\in\mathbf{Z}}$.

The process which Kesten and O'Brien constructed is $X_n := (U_n, V_n, V_{n-U_n})$, $n \in \mathbf{Z}$. In this section $(X_i)_{i\in\mathbf{Z}}$ is always this process.

Kesten and O'Brien proved

Theorem 14 [K.O'B]. *If $\phi(n)$ is the ϕ-mixing coefficient for $(X_i)_{i\in\mathbf{Z}}$ then*

$$\frac{1}{2}\mu\left(\bigcup_{k\geq n}\{U_k \geq k\}\right) \leq \phi(n) \leq \mu\left(\bigcup_{k\geq n}\{U_k \geq k\}\right).$$

In particular $(X_i)_{i\in\mathbf{Z}}$ is ϕ-mixing $\Leftrightarrow EU_0 < \infty$.

We prove an analogous estimate for the VWB-rate.

Theorem 15. *$(X_i)_{i\in\mathbf{Z}}$ is strictly* VWB $\varepsilon(n)$ *where*

$$\frac{1}{2n}\sum_{k=1}^{n}\mu(U_k \geq k) \leq \varepsilon(n) \leq \frac{1}{n}\sum_{k=1}^{n}\mu(U_k \geq k).$$

Proof. First we observe that

$$n\rho_n(\mu, \mu(\cdot/Z)) \geq \sum_{i=1}^{n}|\operatorname{dist} X_i\mu - \operatorname{dist} X_i\mu(\cdot/Z)|$$

by Lemma 5. Thus

$$\begin{aligned} n\varepsilon(n) &\geq \sup_{Z\in\mathscr{A}_{-\infty}^{0}}\sum_{i=1}^{n}|\operatorname{dist} X_i\mu - \operatorname{dist} X_i\mu(\cdot/Z)| \\ &\geq \sup_{Z\in\mathscr{A}_{-\infty}^{0}}\sum_{i=1}^{n}|\mu(U_i \geq i, V_{i-U_i} = 1) - \mu(U_i \geq i, V_{i-U_i} = 1/Z)| \\ &\geq \sum_{i=1}^{n}\frac{1}{2}\mu(U_i \geq i) \qquad \left(Z_N = \bigcap_{j=0}^{N}\{V_{-j} = 0\},\ \text{let } N \to \infty\right). \end{aligned}$$

So

$$\varepsilon(n) \geq \frac{1}{2n}\sum_{i=1}^{n}\mu(U_i \geq i).$$

For proving the upper bound we cannot apply Theorem 6, because for large n we have for $\mu(U_k = k) := 1/k(k+1)$, $k \geq 1$,

$$\sup_{Z\in\mathscr{A}_{-\infty}^{0}}\sum_{i=1}^{n}|\operatorname{dist} X_i^n\mu - \operatorname{dist} X_i^n\mu(\cdot/Z)| \geq \frac{1}{8}\sum_{k=1}^{n}\mu\left(\bigcup_{i\geq k} U_i \geq i\right) = \frac{1}{8}n.$$

Because $\frac{1}{n}\sum_{i=1}^{n}\mu(U_i \geq i) \to 0$ if $n \to \infty$, we have for large n

$$\frac{1}{n}\sum_{i=1}^{n}\mu(U_i \geq i) \leq \frac{1}{8} \leq \frac{1}{n}\sup_{Z\in\mathscr{A}_{-\infty}^{0}}\sum_{i=1}^{n}|\operatorname{dist} X_i^n\mu - \operatorname{dist} X_i^n\mu(\cdot/Z)|.$$

Thus Theorem 6 is not strong enough in this case, because it gives a trivial upper bound. So we have to construct a measure

$$\lambda \in P_n((X_1, \dots, X_n)\mu(\cdot/Z), (X_1, \dots, X_n)\mu).$$

Fix n, $m \in \mathbf{Z}^+$ and $Z \in \mathscr{A}_{-m}^0$ of the form $Z = \{U_0 = u_0, V_0 = v_0, V_{0-U_0} = w_0, \dots, U_{-m} = u_{-m}, V_{-m} = v_{-m}, V_{-m-U_{-m}} = w_{-m}\}$ such that $\mu(Z) > 0$. First we have, because $(U_i)_{i\in\mathbf{Z}}$, $(V_i)_{i\in\mathbf{Z}}$ are i.i.d.,

(3.1)

$$n\rho_n((U_1, \dots, U_n, V_1, \dots, V_n)\mu, (U_1, \dots, U_n, V_1, \dots, V_n)\mu(\cdot/Z)) = 0.$$

If $U_k < k$ or ($U_k > k+m$ and $k - U_k \neq -i - U_{-i}$ $\forall 0 \le i \le m$) then V_{k-U_k} is independent of $Z \in \mathscr{A}_{-m}^0$, and it is this property which helps us to find a good joining.

Let $C = \{U_1 = u_1, V_1 = v_1, \dots, U_n = u_n, V_n = v_n\}$. (3.1) implies $\mu(C) = \mu(C/Z)$.

Let $J(C)$ be the indices where C does not hit Z, so $J(C) := \{1 \le l \le n : u_l < l$ or ($u_l > l+m$ and $l - u_l \neq -i - u_{-i}$ $\forall 0 \le i \le m)\}$. Then $J(C) = \varnothing$ or $J(C) = \{j_1, \dots, j_r\}$ and $(X_{j_1}, \dots, X_{j_r})\mu(\cdot/C) = (X_{j_1}, \dots, X_{j_r})\mu(\cdot/C \cap Z)$.

If $J(C) = \{1, \dots, n\}$ then there is $\lambda_C : \{0, 1\}^n \times \{0, 1\}^n \to \mathbf{R}$ such that

(1) $\mathrm{pr}_1 \lambda_C((x_1, \dots, x_n)) = \mu(V_{1-U_1} = x_1, \dots, V_{n-U_n} = x_n/C \cap Z)$,
(2) $\mathrm{pr}_2 \lambda_C((y_1, \dots, y_n)) = \mu(V_{1-U_1} = y_1, \dots, V_{n-U_n} = y_n/C)$ and

$$\int \sum_{i=1}^n \sigma(x_i, y_i)\, d\lambda_C = 0.$$

If $J(C) \neq \{1, \dots, n\}$ then $\{1, \dots, n\} - J(C) = \{l_1, \dots, l_s\}$, $s \ge 1$. Then let

$$\overline{w_i} := \begin{cases} v_{l_i - u_{l_i}}, & \text{if } l_i - u_{l_i} \ge -m, \\ w_{-r}, & \text{if } l_i - u_{l_i} = -r - u_{-r} \text{ for } r \in \{0, \dots, m\}. \end{cases}$$

So we have $(\overline{w}_1, \dots, \overline{w}_s) \in \{0, 1\}^s$. Let $\lambda_C : \{0, 1\}^n \times \{0, 1\}^n \to \mathbf{R}$ be defined by

$$\begin{aligned} \lambda_C((x_1, \dots, x_n), (y_1, \dots, y_n)) &:= 0 \text{ if } (x_{l_1}, \dots, x_{l_s}) \neq (\overline{w}_1, \dots, \overline{w}_s) \text{ or} \\ &\qquad y_i \neq x_i \text{ for some } i \in J(C), \\ \lambda_C((x_1, \dots, x_n), (y_1, \dots, y_n)) &:= \mu(V_{1-U_1} = x_1, \dots, V_{n-U_n} = x_n/C \cap Z) \\ &\qquad \cdot \mu(V_{l_1 - U_{l_1}} = y_{l_1}, \dots, V_{l_s - U_{l_s}}/C) \text{ otherwise.} \end{aligned}$$

Then one calculates
$\mathrm{pr}_1 \lambda_C((x_1, \dots, x_n)) = \mu(V_{1-U_1} = x_1, \dots, V_{n-U_n} = x_n/C \cap Z)$,
$\mathrm{pr}_2 \lambda_C((y_1, \dots, y_n)) = \mu(V_{1-U_1} = y_1, \dots, V_{n-U_n} = y_n/C)$, and

$$\begin{aligned}
&\int \sum_{i=1}^{n} \sigma(x_i, y_i)\, d\lambda_C \\
&= \sum_{\{(x_1, \ldots, x_n) : x_{l_i} = \bar{w}_i i \le s\}} \ \sum_{\{(y_1, \ldots, y_n) : y_i = x_i \text{ if } i \in J(C)\}} \ \sum_{i=1}^{n} \sigma(x_i, y_i) \\
&\quad \cdot \mu(V_{1-U_1} = x_1, \ldots, V_{n-U_n} = x_n / C \cap Z) \\
&\quad \cdot \mu(V_{l_1 - U_{l_1}} = y_{l_1}, \ldots, V_{l_s - U_{l_s}} = y_{l_s} / C) \\
&= \sum_{\{(x_1, \ldots, x_n) : x_{l_i} = \bar{w}_i i \le s\}} \ \sum_{\{(y_1, \ldots, y_n) : y_i = x_i \text{ if } i \in J(C)\}} \ \sum_{r=1}^{s} \sigma(x_{l_r}, y_{l_r}) \\
&\quad \cdot \mu(V_{1-U_1} = x_1, \ldots, V_{n-U_n} = x_n / C \cap Z) \\
&\quad \cdot \mu(V_{l_1 - U_{l_1}} = y_{l_1}, \ldots, V_{l_s - U_{l_s}} = y_{l_s} / C) \\
&\le \sum_{\{(x_1, \ldots, x_n) : x_{l_i} = \bar{w}_i i \le s\}} s\mu(V_{1-U_1} = x_1, \ldots, V_{n-U_n} = x_n / C \cap Z) \\
&\le s = \operatorname{card} J(C)^C.
\end{aligned}$$

So we get with λ: $(\mathbf{Z}^+ \times \{0, 1\} \times \{0, 1\})^n \times (\mathbf{Z}^+ \times \{0, 1\} \times \{0, 1\})^n \to \mathbf{R}$ defined by

$$\begin{aligned}
&\lambda(((u_1, v_1, w_1), \ldots, (u_n, v_n, w_n)) \times ((a_1, b_1, c_1), \ldots, (a_n, b_n, c_n))) \\
&\quad := \mu(C) \cdot \lambda_C((w_1, \ldots, w_n), (c_1, \ldots, c_n)),
\end{aligned}$$

C as above, if $u_i = a_i$, $v_i = b_i$ $\forall i$ and

$$\lambda(((u_1, v_1, w_1), \ldots, (u_n, v_n, w_n)) \times ((a_1, b_1, c_1), \ldots, (a_n, b_n, c_n))) := 0$$

otherwise, a probability measure $\lambda \in P_n((X_1, \ldots, X_n)\mu(\cdot/Z), (X_1, \ldots, X_n)\mu)$ by (3.1) and

$$\begin{aligned}
n\rho_n(\mu, \mu(\cdot/Z)) &\le \int \sum_{i=1}^{n} \sigma(x_i, y_i)\, d\lambda = \sum_{C} \mu(C) \int \sum_{i=1}^{n} \sigma(x_i, y_i)\, d\lambda_C \\
&\le \sum_{C} \mu(C) \cdot \operatorname{card} J(C)^c \le \sum_{C} \mu(C) \cdot \operatorname{card}(\{i \le n : U_i \ge i\} \cap C) \\
&= \sum_{i=1}^{n} \mu(U_i \ge i). \quad \square
\end{aligned}$$

Remark. One can actually strengthen this last construction and prove that if $(U_i)_{i \in \mathbf{Z}}$ is a stationary process with values in $\mathbf{Z}^+$ and $(V_i)_{i \in \mathbf{Z}}$ is a stationary process with values in $\{0, 1\}$ and $X_n := (U_n, V_n, V_{n-U_n})$ then

(1) $(X_i)_{i \in \mathbf{Z}}$ ϕ-mixing $\Leftrightarrow EU_0 < \infty$, $(U_i)_{i \in \mathbf{Z}}$ ϕ-mixing, $(V_i)_{i \in \mathbf{Z}}$ ϕ-mixing,
(2) $(X_i)_{i \in \mathbf{Z}}$ VWB $O(1/n) \Leftrightarrow EU_0 < \infty$, $(U_i)_{i \in \mathbf{Z}}$ VWB $O(1/n)$, $(V_i)_{i \in \mathbf{Z}}$ VWB $O(1/n)$.

For this one needs a Borel-Cantelli-Lemma for ϕ-mixing sequences.

We get as corollaries of Theorems 14 and 15:

Corollary 16. *$X_n := (U_n, V_n, V_{n-U_n})$ as above. Then*

$$(X_i)_{i\in\mathbf{Z}}\ \phi\text{-mixing} \Leftrightarrow (X_i)_{i\in\mathbf{Z}}\,\mathrm{VWB}\,O(1/n) \Leftrightarrow EU_0 < \infty.$$

Corollary 17. *For any rate $\varepsilon(n)$ with $(n+1)\varepsilon(n+1) - n\varepsilon(n) \le n\varepsilon(n) - (n-1)\varepsilon(n-1)$ $\forall n$, $\varepsilon(n) \to 0$, $n\varepsilon(n) \to \infty$ and $n\varepsilon(n) \le n$ $\forall n$ there is a process $(X_i)_{i\in\mathbf{Z}}$ which is strictly* VWB $\varepsilon(n)$ *and not ϕ-mixing.*

We would like to find an example of a process which is not VWB $O(1/n)$, but ϕ-mixing, but we have not yet been successful. We believe a good candidate is the following:

Let $(U_i)_{i\in\mathbf{Z}}$, $(V_i)_{i\in\mathbf{Z}}$ as above. Let $Y_n := (V_n, V_{n-U_n})$, $n \in \mathbf{Z}$. Then it is not hard to see that $EU_0 = \infty \Rightarrow (Y_i)_{i\in\mathbf{Z}}$ is not VWB $O(1/n)$. The conjecture is

$$EU_0 = \infty, \quad \sum_{k=1}^{\infty} \mu(U_k \ge k)^2 < \infty \Rightarrow (Y_i)_{i\in\mathbf{Z}} \text{ is } \phi\text{-mixing.}$$

Appendix

Proof of Theorem 6. Fix $n \in \mathbf{N}$, $Z \in \mathscr{A}_{-\infty}^{0}$, $\mu(Z) < 0$. We will need some elaborate notation. Let $\{X_1^n = s_1^n\} := \{X_1 = s_1, \dots, X_n = s_n\}$.

$$\begin{aligned}
I_1 &:= \{(s_1, \dots, s_n) \in S^n : \mu(X_1^n = s_1^n) > \mu(X_1^n = s^n/Z)\},\\
\overline{I_1} &:= \{(s_1, \dots, s_n) \in S^n : \mu(X_1^n = s_1^n) < \mu(X_1^n = s_1^n/Z)\},\\
\tau_1(s_1, \dots, s_n) &:= \mu(X_1^n = s_1^n), \quad \bar\tau_1(s_1, \dots, s_n) := \mu(X_1^n = s_1^n/Z),\\
\rho_1(s_1, \dots, s_n) &:= (\mu(X_1^n = s_1^n) - \mu(X_1^n = s_1^n/Z)) \cdot 1_{I_1}(s_1, \dots, s_n),\\
\bar\rho_1(s_1, \dots, s_n) &:= (\mu(X_1^n = s_1^n/Z) - \mu(X_1^n = s_1^n)) \cdot 1_{\overline{I}_1}(s_1, \dots, s_n).
\end{aligned}$$

Then inductively for $1 \le k \le n-1$

$$\begin{aligned}
\tau_{k+1}(s_{k+1}, \dots, s_n) &:= \sum_{(s_1, \dots, s_k)} \rho_k(s_1, \dots, s_k, s_{k+1}, \dots, s_n),\\
\bar\tau_{k+1}(s_{k+1}, \dots, s_n) &:= \sum_{(s_1, \dots, s_k)} \bar\rho_k(s_1, \dots, s_k, s_{k+1}, \dots, s_n),\\
I_{k+1} &= \{(s_{k+1}, \dots, s_n) \in S^{n-k} : \tau_{k+1}(s_{k+1}, \dots, s_n) > \bar\tau_{k+1}(s_{k+1}, \dots, s_n)\},\\
\overline{I}_{k+1} &= \{(s_{k+1}, \dots, s_n) \in S^{n-k} : \tau_{k+1}(s_{k+1}, \dots, s_n) < \bar\tau_{k+1}(s_{k+1}, \dots, s_n)\},\\
\rho_{k+1}(s_1, \dots, s_n) &:= \rho_k(s_1, \dots, s_n)\left(1 - \frac{\bar\tau_{k+1}(s_{k+1}, \dots, s_n)}{\tau_{k+1}(s_{k+1}, \dots, s_n)}\right) 1_{I_{k+1}}(s_{k+1}, \dots, s_n),\\
\bar\rho_{k+1}(s_1, \dots, s_n) &:= \bar\rho_k(s_1, \dots, s_n)\left(1 - \frac{\tau_{k+1}(s_{k+1}, \dots, s_n)}{\bar\tau_{k+1}(s_{k+1}, \dots, s_n)}\right) 1_{\overline{I}_{k+1}}(s_{k+1}, \dots, s_n),\\
\tau_{n+1} &:= \sum_{s\in S^n} \rho_n(s), \quad \bar\tau_{n+1} := \sum_{s\in S^n} \bar\rho_n(s).
\end{aligned}$$

We want to define a probability measure on $S^n \times S^n$, therefore we partition the set $S^n \times S^n = W_0 \cup W_1 \cup \cdots \cup W_n \cup R$ in disjoint sets, where

$$W_0 = \{(x, x) : x \in S^n\},$$

$$W_n = \{((x_1, \ldots, x_n), (y_1, \ldots, y_n)) : x_n \neq y_n,\ \rho_n(x_1, \ldots, x_n) > 0 \text{ and } \bar{\rho}_n(y_1, \ldots, y_n) > 0\},$$

$$W_i = \{((x_1, \ldots, x_n), (y_1, \ldots, y_n)) : x_i \neq y_i,\ x_r = y_r,\ i < r \leq n,\ \min(\tau_{i+1}(x_{i+1}, \ldots, x_n), \bar{\tau}_{i+1}(x_{i+1}, \ldots, x_n)) > 0\} \quad \text{for } 1 \leq i \leq n-1,$$

$$R := S^n \times S^n - \bigcup_{i=0}^{n} W_i.$$

Then we define $\nu_n : S^n \times S^n \to \mathbf{R}$ in the following way:

(1) $((s_1, \ldots, s_n), (s_1, \ldots, s_n)) \in W_0$:

$$\nu_n((s_1, \ldots, s_n), (s_1, \ldots, s_n)) := \min(\tau_1(s_1, \ldots, s_n), \bar{\tau}_1(s_1, \ldots, s_n)).$$

(2) $((a_1, \ldots, a_i, s_{i+1}, \ldots, s_n), (b_1, \ldots, b_i, s_{i+1}, \ldots, s_n)) \in W_i$, $1 \leq i \leq n-1$:

$$\begin{aligned}\nu_n&((a_1, \ldots, a_i, s_{i+1}, \ldots, s_n), (b_1, \ldots, b_i, s_{i+1}, \ldots, s_n)) \\ &:= \frac{\min(\tau_{i+1}(s_{i+1}, \ldots, s_n), \bar{\tau}_{i+1}(s_{i+1}, \ldots, s_n))}{\tau_{i+1}(s_{i+1}, \ldots, s_n)\bar{\tau}_{i+1}(s_{i+1}, \ldots, s_n)} \\ &\quad \cdot \rho_i(a_1, \ldots, a_i, s_{i+1}, \ldots, s_n)\bar{\rho}_i(b_1, \ldots, b_i, s_{i+1}, \ldots, s_n).\end{aligned}$$

(3) $((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \in W_n$:

$$\begin{aligned}\nu_n&((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \\ &:= \rho_n(a_1, \ldots, a_n)\bar{\rho}_n(b_1, \ldots, b_n)\frac{\min(\tau_{n+1}, \bar{\tau}_{n+1})}{\tau_{n+1}\bar{\tau}_{n+1}}.\end{aligned}$$

(4) $\nu_n((a_1, \ldots, a_n), (b_1, \ldots, b_n)) := 0$ if $((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \in R$. We use the abbreviated notation $s^{(i)} := (s_i, \ldots, s_n) \in S^{n-i+1}$, $1 \leq i \leq n$. First we want to prove that ν_n is a joining of

$$(X_1, \ldots, X_n)\mu \quad \text{and} \quad (X_1, \ldots, S_n)\mu(\cdot/Z).$$

One calculates

$$\begin{aligned}\alpha(s^{(1)}) &:= \sum_{t^{(1)} \in S^n} \nu_n(s^{(1)}, t^{(1)}) \\ &= \nu_n(s^{(1)}, s^{(1)}) + \sum_{i=1}^{n-1} \sum_{\{t^{(1)} \in S^n \,:\, t_i \neq s_i,\, t_{i+1} = s_{i+1}, \ldots, t_n = s_n\}} \nu_n(s^{(1)}, t^{(1)}) \\ &\quad + \sum_{\{t^{(1)} \in S^n \,:\, t_n \neq s_n\}} \nu_n(s^{(1)}, t^{(1)}) \\ &= \min(\tau_1(s^{(1)}), \bar{\tau}_1(s^{(1)})) + \sum_{i=1}^{n-1} \frac{\min(\tau_{i+1}(s^{(i+1)}), \bar{\tau}_{i+1}(s^{(i+1)}))}{\tau_{i+1}(s^{(i+1)})}\rho_i(s^{(1)}) \\ &\quad + \frac{\min(\tau_{n+1}, \bar{\tau}_{n+1})}{\tau_{n+1}}\rho_n(s^{(1)}).\end{aligned}$$

To calculate $\alpha(s^{(1)})$ we have to look for the set I_k that $s^{(k)}$ belongs to:

Case 1. $s^{(1)} \notin I_1$. Then $\rho_i(s^{(1)}) = 0 \;\; \forall i \geq 1$, so

$$\alpha(s^{(1)}) = \tau_1(s^{(1)}) = \mu(X_1 = s_1, \ldots, X_n = s_n).$$

Case 2. $s^{(1)} \in I_1$, $s^{(2)} \notin I_2$. Then $\rho_i(s^{(1)}) = 0 \;\; \forall i \geq 2$, and

$$\alpha(s^{(1)}) = \bar{\tau}_1(s^{(1)}) + \rho_1(s^{(1)}) = \mu(X_1 = s_1, \ldots, X_n = s_n).$$

General case. $s^{(1)} \in I_1, \ldots, s^{(k)} \in I_k$, $s^{(k+1)} \notin I_{k+1}$. Then the same argument as in Case 2 shows $\alpha(s^{(1)}) = \mu(X_1 = s_1, \ldots, X_n = s_n)$ and in the case $s^{(1)} \in I_1, \ldots, s^{(n)} \in I_n$ one uses the fact $\tau_{n+1} = \bar{\tau}_{n+1}$ to see $\alpha(s^{(1)}) = \mu(X_1 = s_1, \ldots, X_n = s_n)$. Similarly

$$\sum_{s^{(1)} \in S^n} \nu_n(s^{(1)}, t^{(1)}) = \mu(X_1 = t_1, \ldots, X_n = t_n / Z).$$

For proving (2.1) we need an equivalent definition of the sets I_k.

Claim 2. $1 \leq k \leq n$. Then

$$I_k = \{s^{(k)} : \mu(X_k = s_k, \ldots, X_n = s_n) > \mu(X_k = s_k, \ldots, X_n = s_n / Z)\}.$$

Proof of the claim.

$$\begin{aligned}
\tau_k(s^{(k)}) &= \sum_{s_1, \ldots, s_{k-1}} \rho_{k-1}(s_1, \ldots, s_{k-1}, s_k, \ldots, s_n) \\
&= \sum_{s_1, \ldots, s_{k-2}} \sum_{s_{k-1}} \rho_{k-2}(s^{(1)}) \left(1 - \frac{\bar{\tau}_{k-1}(s^{(k-1)})}{\tau_{k-1}(s^{(k-1)})}\right) 1_{I_{k-1}}(s^{(k-1)}) \\
&= \sum_{\{s : (s, s_k, \ldots, s_n) \in I_{k-1}\}} (\tau_{k-1}(s, s_k, \ldots, s_n) - \bar{\tau}_{k-1}(s, s_k, \ldots, s_n)) \\
&= \sum_{\{s : (s, s^{(k)}) \in I_{k-1}\}} (\tau_{k-1}(s, s^{(k)}) - \bar{\tau}_{k-1}(s, s^{(k)})).
\end{aligned}$$

So we get

$$\begin{aligned}
&\tau_k(s^{(k)}) > \bar{\tau}_k(s^{(k)}) \\
&\quad \Leftrightarrow \sum_{\{s : (s, s^{(k)}) \in I_{k-1}\}} (\tau_{k-1}(s, s^{(k)}) - \bar{\tau}_{k-1}(s, s^{(k)})) \\
&\qquad > \sum_{\{s : (s, s^{(k)}) \in \bar{I}_{k-1}\}} (\bar{\tau}_{k-1}(s, s^{(k)}) - \tau_{k-1}(s, s^{(k)})) \\
&\quad \Leftrightarrow \sum_{s \in S} \tau_{k-1}(s, s^{(k)}) > \sum_{s \in S} \bar{\tau}_{k-1}(s, s^{(k)}) \\
&\quad \Leftrightarrow \sum_{s_1, \ldots, s_{k-1}} \tau_1(s_1, \ldots, s_{k-1}, s^{(k)}) > \sum_{s_1, \ldots, s_{k-1}} \bar{\tau}_1(s_1, \ldots, s_{k-1}, s^{(k)}) \\
&\hspace{20em} \text{by repeating the argument} \\
&\quad \Leftrightarrow \mu(X_k = s_k, \ldots, X_n = s_n) > \mu(X_k = s_k, \ldots, X_n = s_n / Z) \\
&\hspace{20em} \text{by definition of } \tau_1, \bar{\tau}_1.
\end{aligned}$$

This completes the proof of the claim.

Now we compute for $1 \le i \le n$

$$
\begin{aligned}
&\int_{S^n\times S^n} \sigma(x_i, y_i)\, d\nu_n((x_1, \dots, x_n), (y_1, \dots, y_n)) \\
&\quad = \nu_n(\{(s^{(1)}, t^{(1)}) : s_i \neq t_i\}) \\
&\quad \le 1 - \sum_{j=0}^{i-1} \nu_n(\{(s^{(1)}, t^{(1)}) : s_{j+1} = t_{j+1}, \dots, s_n = t_n, s_j \neq t_j\}) \\
&\quad = 1 - \sum_{j=0}^{i-1} \sum_{s^{(j+1)}} \min(\tau_{j+1}(s^{(j+1)}), \bar\tau_{j+1}(s^{(j+1)})) \\
&\quad = 1 - \sum_{j=1}^{i-1} \left(\sum_{s^{(j)}\in I_j} \bar\tau_j(s^{(j)}) + \sum_{s^{(j)}\notin I_j} \tau_j(s^{(j)}) \right) \\
&\qquad - \left(\sum_{s^{(i)}\in I_i} \sum_{s_1, \dots, s_{i-1}} \bar\rho_{i-1}(s^{(1)}) + \sum_{s^{(i)}\notin I_i} \sum_{s_1, \dots, s_{i-1}} \rho_{i-1}(s^{(1)}) \right) \\
&\quad = 1 - \sum_{j=1}^{i-2} \left(\sum_{s^{(j)}\in I_j} \bar\tau_j(s^{(j)}) + \sum_{s^{(j)}\notin I_j} \tau_j(s^{(j)}) \right) \\
&\qquad - \sum_{s^{(i-1)}\in I_{i-1}} \bar\tau_{i-1}(s^{(i-1)}) - \sum_{s^{(i-1)}\notin I_{i-1}} \tau_{i-1}(s^{(i-1)}) \\
&\qquad - \sum_{s^{(i)}\in I_i,\, s^{(i-1)}\in \bar I_{i-1}} (\bar\tau_{i-1}(s^{(i-1)}) - \tau_{i-1}(s^{(i-1)})) \\
&\qquad - \sum_{s^{(i)}\notin I_i,\, s^{(i-1)}\in I_{i-1}} (\tau_{i-1}(s^{(i-1)}) - \bar\tau_{i-1}(s^{(i-1)})) \\
&\quad = 1 - \sum_{j=1}^{i-2} \left(\sum_{s^{(j)}\in I_j} \bar\tau_j(s^{(j)}) + \sum_{s^{(j)}\notin I_j} \tau_j(s^{(j)}) \right) \\
&\qquad - \sum_{\{s^{(i-1)} : s^{(i)}\in I_i\}} \bar\tau_{i-1}(s^{(i-1)}) - \sum_{\{s^{(i-1)} : s^{(i)}\notin I_i\}} \tau_{i-1}(s^{(i-1)}) \\
&\quad = 1 - \sum_{\{s^{(1)} : s^{(i)}\in I_i\}} \bar\tau_1(s^{(1)}) - \sum_{\{s^{(1)} : s^{(i)}\notin I_i\}} \tau_1(s^{(1)}) \quad \text{(by repeating the argument)} \\
&\quad = 1 - \sum_{s^{(i)}\in I_i} \mu(X_i = s_i, \dots, X_n = s_n/Z) - \sum_{s^{(i)}\notin I_i} \mu(X_i = s_i, \dots, X_n = s_n) \\
&\quad = \sum_{s^{(i)}\in I_i} (\mu(X_i = s_i, \dots, X_n = s_n) - \mu(X_i = s_i, \dots, X_n = s_n/Z)) \\
&\quad = |\operatorname{dist} X_i^n \mu - \operatorname{dist} X_i^n \mu(\cdot/Z)| \quad \text{(by Claim 2).}
\end{aligned}
$$

So (2.1) is proved and therefore Theorem 6, also. □

REFERENCES

[B1] R. C. Bradley, *On a very weak Bernoulli condition*, Stochastic **13** (1984), 61–81.

[B2] ——, *On the ϕ-mixing condition for stationary random sequences*, Duke Math. J. **47** (1980), 421–433.

[B3] ——, *Basic properties of strong mixing conditions*: Dependence in Probability and Statistics (E. Eberlein and M. S. Taqqu, eds.), Birkhäuser, 1986, pp. 165–192.

[D.D.P] H. Dehling, M. Denker, and W. Philipp, *Versik processes and very weak Bernoulli processes with summable rates are independent*, Proc. Amer. Math. Soc. **91** (1984), 618–623.

[E] E. Eberlein, *Strong approximation of very weak Bernoulli processes*, Z. Wahrsch. Verw. Gebiete **62** (1983), 17–37.

[F] D. Fiebig, *Über very weak Bernoulli Prozesse mit schnellen Raten*, Dissertation, Göttingen, 1988.

[K.O'B] H. Kesten and G. L. O'Brien, *Examples of mixing sequences*, Duke Math. J. **43** (1976), 405–415.

[R] M. Rosenblatt, *Markov processes, structure and asymptotic behavior*, Springer-Verlag, Berlin, 1971.

[S.] W. Strittmatter, *Measures of dependence for processes in metric spaces*, Stochastics Stochastics Rep. **27** (1989), 33–50.

UNIVERSITÄT HEIDELBERG SFB 123, IM NEUENHEIMER FELD 294, D-6900 HEIDELBERG 1, WEST GERMANY

Current address: Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 6900 Heidelberg, Germany

E-mail address: bq6@dhdurzl.bitnet