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 VWBO $O(1/n)$ implies $\phi$-mixing.

0. Introduction

Let $X_i: (\Omega, \mathcal{A}, \mu) \rightarrow (S, \mathcal{B}), \ i \in \mathbb{Z}$, be a stationary sequence of random variables on a probability space $(\Omega, \mathcal{A}, \mu)$ with values in a Polish space $(S, \mathcal{B})$. In this setting we define very weak Bernoulli processes with rate $\varepsilon(n)$, denoted by $\text{VWB} \varepsilon(n)$, and strictly $\text{VWB} \varepsilon(n)$ 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 $(X_i)_{i \in \mathbb{Z}}$ is VWB $O(1/n)$ iff $(X_i)_{i \in \mathbb{Z}}$ is strictly VWB $O(1/n)$. This strengthens the result of Eberlein, that real-valued strictly VWBO $O(1/n)$ 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 $\mathcal{B}$ is generated by the discrete metric. Our main result in this case is that VWB $O(1/n)$ implies $\phi$-mixing, which improves an earlier result of [D.D.P]. We show that VWBO $O(1/n)$ gives no constraints on the $\phi$-mixing rate, and that VWBO $O(1/n)$ does not imply $\psi$-mixing. After that we give a new upper bound for the Wasserstein-distance, which implies that a $\phi$-mixing process with $\phi$-mixing rate $\phi(i)$ is strictly VWB with rate $\frac{1}{n} \sum_{i=1}^{n} \phi(i)$; in particular $\phi$-mixing processes with summable rates are VWBO $O(1/n)$.

1. VWBO $O(1/n)$ implies strictly VWBO $O(1/n)$

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

\[ \rho_n(\mu, \mu(\cdot/Z)) := \inf \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^{n} \sigma(x_i, y_i) \, d\lambda(x_1, \ldots, x_n, y_1, \ldots, y_n) \]

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

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

1. \( \varepsilon(n) \to 0, \ n \to \infty \),
2. \( \forall n \in \mathbb{N} \ \forall m \in \mathbb{Z}^+ \ \exists D = D(m, n) \in \mathcal{A}_{-m}^0 \) with
   \[ \mu(D) \geq 1 - \varepsilon(n), \]

1.1. \( \mathbb{P}(D) > 1 - \varepsilon(n) \),
1.2. \( \mathbb{P}(A) > 0, \mathbb{P}(\mathcal{A} \cap D(m, k)) < \varepsilon(n), \ A \in \mathcal{A}_{-m} \).

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

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 \( \liminf \varepsilon(n) = 0 \) 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 \((X_i)_{i \in \mathbb{Z}}\) be a stationary sequence of random variables with values in a Polish space. Let \( 0 \leq M < \infty \). Then:

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

We need

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

\[ n \rho_n: M_n \times M_n \to \mathbb{R} \]

\[ (\nu_1, \nu_2) \to n \rho_n(\nu_1, \nu_2) \]

is a lower semicontinuous function.

**Proof of Theorem 3.** Let \( 0 \leq M < \infty \). Let \((X_i)_{i \in \mathbb{Z}}\) be VWB with rate \( M/n \). Then for all \( n \in \mathbb{N}, m \in \mathbb{Z}^+ \) sets \( D(m, n) \in \mathcal{A}_{-m}^0 \) 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 \( n \in \mathbb{N}, m \in \mathbb{Z}^+ \) and choose a set \( A \in \mathcal{A}_{-m}^0 \) 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 \( A \cap D(m, k) \in \mathcal{A}_{-m}^0 \forall k \), (1.2) implies

\[ \rho_k(\mu, \mu(\cdot/A \cap D(m, k))) \leq \frac{M}{k} \ \forall k \geq k_0. \]
Since \( n \leq k_0 \) we have
\[
\rho_n(\mu, \mu(\cdot / A \cap D(m, k))) \leq \frac{M}{n} \quad \forall k \geq k_0.
\]

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

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

Let \((X_i)_{i \in \mathbb{Z}}\) be a Markov chain with state space \( \mathbb{Z}^+ \) and transition probabilities
\[
\begin{align*}
p_{ij} = \begin{cases} 1, & i = j + 1, j \geq 0, \\ c_j, & i = 0, j \geq 0, \\ 0, & \text{otherwise}, \end{cases}
\end{align*}
\]

where \((c_n)_{n \geq 1}\) is a sequence with \( c_j \geq 0 \ \forall j \), \( c_j > 0 \) infinitely often, \( \sum_{j=1}^{\infty} j c_j < \infty \). Then the \((p_{ij})_{i,j \in \mathbb{Z}^+}\) 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 \( n \in \mathbb{N}, \ i \in \mathbb{Z}^+ \) (by Theorem 6). This gives the possibility by choosing \((c_j)_{j \geq 1}\) 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 \((X_i)_{i \in \mathbb{Z}}\) 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 \( \sigma \). There exists a stationary VWB \( O(1/n) \) process with uncountable state space \( S \subseteq \mathbb{R} \), where \( \sigma \) is the Euclidean metric, which is not even \( \alpha \)-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 \( Z \in \mathcal{A}_{\infty}^0, \mu(Z) > 0 \) and \( 1 \leq i \leq n < \infty \) we define the distribution distance of names by
\[
| \text{dist}_{X^n_i} \mu - \text{dist}_{X^n_i} \mu(\cdot / Z) | = \frac{1}{2} \sum_{(y_0, \ldots, y_i) \in S^{n-i+1}} | \mu(X_n = y_n, \ldots, X_i = y_i) - \mu(X_n = y_n, \ldots, X_i = y_i / Z) |.
\]

With this notation we have the simple, but extremely useful

**Lemma 5.** Let \((n_i)_{i \in \mathbb{N}}\) be a strictly increasing sequence of natural numbers with \( n_0 := 0 \). Let \( Z \in \mathcal{A}_{\infty}^0 \) with \( \mu(Z) > 0 \). Then for all \( N \in \mathbb{N} \)
\[
n_{N+1} \rho_{n_{N+1}}(\mu, \mu(\cdot / Z)) \geq \sum_{i=0}^{N} | \text{dist}_{X^{n_{i+1}}_{n_i+1}} \mu - \text{dist}_{X^{n_{i+1}}_{n_i+1}} \mu(\cdot / Z) |.
\]

(If \( S \) is countable and the metric \( \sigma \) is bounded below by \( \varepsilon \), i.e., \( \sigma(x, y) \geq \varepsilon > 0 \ \forall x \neq y \), then this lemma holds with the RHS multiplied by \( \varepsilon \).) The next theorem gives a new upper bound for the Wasserstein-distance.
Theorem 6. Let \( n \in \mathbb{N}, \ Z \in \mathcal{A}_{\infty}^0 \) with \( \mu(Z) > 0 \). Then
\[
n_p(\mu, \mu(\cdot/Z)) \leq \sum_{i=1}^{n} |\text{dist } X_i^n \mu - \text{dist } X_i^n \mu(\cdot/Z)|.
\]

The proof of Theorem 6 is deferred to the Appendix. It depends on the construction of a joining \( \nu_n \) of \((X_1, \ldots, X_n)\mu\) and \((X_1, \ldots, X_n)\mu(\cdot/Z)\) such that
\[
\int _{S^n \times S^n} \sigma(x_i, y_i) d\nu_n((x_1, \ldots, x_n), (y_1, \ldots, y_n)) 
\leq |\text{dist } X_i^n \mu - \text{dist } X_i^n \mu(\cdot/Z)| \quad \forall 1 \leq i \leq n.
\]
The joining \( \nu_n \) is a generalisation of a construction in [F], and shows that for Markov chains
\[
n_p(\mu, \mu(\cdot/X_0 = x)) = \sum_{i=1}^{n} |\text{dist } X_i \mu - \text{dist } X_i \mu(\cdot/X_0 = x)|.
\]

We use the following mixing coefficients:
\[
\alpha(n) := \sup_{A \in \mathcal{A}_{\infty}^0} \sup_{B \in \mathcal{A}_{\infty}^0} |\mu(B \cap A) - \mu(B)\mu(A)|,
\]
\[
\text{WB}(n) := \sup_{m, k \geq 0} \sum_{B \in \mathcal{A}_{n+k}^0} \sum_{A \in \mathcal{A}_{-m}^0} \mu(A) \cdot |\mu(B|A) - \mu(B)|,
\]
where \( \mathcal{A}_{n+k}^0 \) (resp. \( \mathcal{A}_{-m}^0 \)) is the finest partition of \( \Omega \) into sets \( B \in \mathcal{A}_{n+k}^0 \) (resp. \( A \in \mathcal{A}_{-m}^0 \)).

\[
\phi(n) := \sup_{A \in \mathcal{A}_{\infty}^0} \sup_{B \in \mathcal{A}_{\infty}^0} |\mu(B|A) - \mu(B)|,
\]
\[
\psi(n) := \sup_{A \in \mathcal{A}_{\infty}^0} \sup_{B \in \mathcal{A}_{\infty}^0} |\mu(B|A)/\mu(B) - 1|
\]
(where as always \( \mu(A) > 0 \) is assumed, if necessary). \((X_i)_{i \in \mathbb{Z}}\) is said to be \( \alpha \)-mixing (weak Bernoulli (= WB), \( \phi \)-mixing or \( \psi \)-mixing) iff \( \alpha(n) \to 0 \) (\( \text{WB}(n) \to 0 \), \( \phi(n) \to 0 \), \( \psi(n) \to 0 \)), respectively. From the definitions of the mixing coefficients it is clear that
\[
\psi \text{-mixing } \Rightarrow \phi \text{-mixing } \Rightarrow \text{WB } \Rightarrow \alpha \text{-mixing}.
\]
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 \((X_i)_{i \in \mathbb{Z}}\) be a stationary process with at most countable state space \( S \) and discrete metric (or a metric bounded away from zero). Then \((X_i)_{i \in \mathbb{Z}}\) VWB \( O(1/n) \Rightarrow (X_i)_{i \in \mathbb{Z}} \phi\)-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 \((m \in \mathbb{N})\)
\[
M \geq (Nm)\rho_{Nm}(\mu, \mu(\cdot/D))
\geq \sum_{i=1}^{N} |\text{dist } X^\text{(im)}_{(i-1)m+1} \mu - \text{dist } X^\text{(im)}_{(i-1)m+1} \mu(\cdot/D)| \quad \text{by Lemma 5}.
\]
So that, given \( \varepsilon > 0 \), there is an \( N \in \mathbb{N} \) such that for any set
\[
D \in \mathcal{A}_0^\infty, \mu(D) > 0 \text{ there is an } i \leq N \text{ with }
\] (22.)
\[
\left| \text{dist } X_{(i-1)m+1} \right| \mu - \text{dist } X_{(i-1)m+1} \mu(\cdot/D) < \varepsilon.
\]

**Lemma 8.** Let \((X_i)_{i \in \mathbb{Z}}\) be VVB with rate \( \varepsilon(n) = M/n, \ 0 \leq M < \infty \). Let \( \sigma : S^\mathbb{Z} \to S^\mathbb{Z} \) be the shift map, i.e., \( \sigma((s_i)_{i \in \mathbb{Z}})_i = s_{i+1} \forall i \in \mathbb{Z} \). Fix \( r \in \mathbb{N}, m \in \mathbb{N} \). Choose \( A_1, \ldots, A_r \in \mathcal{A}_m^r \) with \( \mu(A_i) > 0 \ \forall s \). Fix \( \delta > 0 \). Then there is \( k = k(\min_{1 \leq i \leq r} \mu(A_i), \delta) \in \mathbb{N} \) such that:
\[
\forall D \in \mathcal{A}_0^\infty, \mu(D) > 0 \ \exists 0 \leq i \leq k (i \text{ depends on } D) \text{ with }
\] \[
\left| \mu(\sigma^{-im}A_i/D) - \mu(A_i) \right| < \delta \mu(A_i) \ \forall s \leq r.
\]

**Proof.** Choose \( \varepsilon < \delta \min_{1 \leq i \leq r} \mu(A_i) \), and apply the observation (2.2) above, using the fact that \( \left| \mu(\sigma^{-im}A_i/D) - \mu(A_i) \right| \leq \left| \text{dist } X_{(i)mi+1} \mu - \text{dist } X_{(i)mi+1} \mu(\cdot/D) \right| \) \[\Box\]

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

**Proof of Theorem 7.** Let \((X_i)_{i \in \mathbb{Z}}\) be VVB \( \varepsilon(n), \varepsilon(n) = M/n, M < \infty \). Assume \((X_i)_{i \in \mathbb{Z}}\) is not \( \phi \)-mixing.

**Claim 1.** \( \forall m \in \mathbb{N} \ \forall \varepsilon > 0 \ \exists 1 = l_0 < l_1 < l_2 < \cdots < l_m < \infty \ \exists k \in \mathbb{N} \ \exists B_i \in \mathcal{A}_{l_i-1}^i, 1 \leq i \leq m \) and \( \exists C \in \mathcal{A}_0^0 \), \( \mu(C) > 0 \), such that \( \mu(B_i) > 1 - \varepsilon \), \( \mu(B_i/C) < \varepsilon \) \( \forall i \in \{1, \ldots, m\} \).

We prove this claim by induction on \( m \). For \( m = 1 \), we apply Theorem 1 of [B2], so \((X_i)_{i \in \mathbb{Z}}\) not \( \phi \)-mixing means \( \phi(1) = 1 \). This implies the claim for \( m = 1 \), because one can approximate sets in \( \mathcal{A}_1^\infty \) (resp. \( \mathcal{A}_0^\infty \)) arbitrarily well by sets in \( \mathcal{A}_1^l \) (resp. \( \mathcal{A}_0^l \)) 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 \( B_i \in \mathcal{A}_{l_i-1}^i, 1 \leq i \leq m, C \in \mathcal{A}_0^0, \mu(C) > 0 \) 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 \( B_1, \ldots, B_{m} \), \( B_{m+1} \) and \( E \) satisfy the claim for \( m + 1 \) and \( \varepsilon \). Let \( I := \{i \in \{1, \ldots, m\} : \mu(B_i/C) > 0\} \), \( A_i := \sigma^{s-1}(B_i \cap C), 1 \leq i \leq m \). Then \( A_i \in \mathcal{A}_{l_i-1+m}^i \). 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 \( D \in \mathcal{A}_0^\infty \exists 0 \leq j < k \) such that
\[
|\mu(\sigma^{-jm'}A_i/D) - \mu(A_i)| < \frac{1}{2} \mu(A_i) \ \forall i \in I,
\]
\[
|\mu(\sigma^{-jm'-s-1}C/D) - \mu(C)| < \frac{1}{2} \mu(C).
\]
Now, because not \( \phi \)-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
\[
(2.3) \ B \in \mathcal{A}_{(2k+1)m'}^L, \ D \in \mathcal{A}_0^{-L} \ \text{with } \mu(B) > 1 - \delta_1, \ \mu(B/D) < \delta_1.
\]

Let \( E := C \cap \sigma^{jm'-s-1}D \) where \( j \) is according to (2.3). Then \( E \in \mathcal{A}_L^{L-jm'-s-1} \), 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 \( B_{m+1} \in \mathcal{A}_{(2k-j)m'}^{L-jm'-s-1} \) and \( (2k-j)m' \geq km' \geq l_m \), so for \( l_{m+1} := L - jm' - s - 1 \) we have
\[
(2.5) \ B_{m+1} \in \mathcal{A}_{l_{m+1}}^{l_{m+1}} \ \text{and } \mu(B_{m+1}) > 1 - \delta.
\]
For $i \in \{1, \ldots, m\} - I$ we have
\[
\mu(B_i/E) \leq \frac{\mu(B_i \cap C)}{\mu(E)} = 0 < \delta.
\]

For $i \in I$ we have
\[
\begin{align*}
\mu(B_i/E) &= \frac{\mu(B_i \cap C \cap \sigma^{jm'+s+1+D})}{\mu(E)} = \frac{\mu(D)\mu(\sigma^{-jm'}A_i/D)}{\mu(E)} \\
&= \frac{\mu(\sigma^{-jm'}A_i/D)}{\mu(\sigma^{-jm'-s-1}C/D)} \\
&\leq \frac{3\delta}{2\mu(A_i)} = 3\mu(B_i/C) < 3\delta \quad \text{(because of (2.3))}
\end{align*}
\]
and
\[
\begin{align*}
\mu(B_{m+1}/E) &= \frac{\mu(\sigma^{jm'+s+1}B \cap C \cap \sigma^{jm'+s+1+D})}{\mu(\sigma^{jm'+s+1}C/D)} \\
&= \frac{\mu(B/D)\mu(D)}{\mu(\sigma^{jm'-s-1}C/D)} < \frac{\delta_1}{\frac{1}{2}\mu(D)} < \delta \quad \text{(because of (2.3), (2.4)).}
\end{align*}
\]
Because $3\delta < \epsilon$ we have sets $B_1, \ldots, B_{m+1}$ and $E$ which satisfy Claim 1 for $m + 1$ and $\epsilon$. This proves Claim 1.

Now we choose $\epsilon < \frac{1}{2}$ and $m \in \mathbb{N}$ such that $m(1 - 2\epsilon) > M$. Then we choose sets $B_i, C$ from Claim 1 to obtain by Lemma 5 the estimate
\[
M \geq l_m \rho_{l_m}(\mu, \mu(\cdot/C)) \geq \sum_{i=1}^{m} |\text{dist}\ X_{l_i-1}^{l_i-1} \mu - \text{dist}\ X_{l_i-1}^{l_i-1} \mu(\cdot/C)| \\
\geq \sum_{i=1}^{m} |\mu(B_i) - \mu(B_i/C)| \geq m(1 - 2\epsilon) > M.
\]

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

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

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

**Example 9.** There exist a VWB $O(1/n)$ process with countable state space which is not $\psi$-mixing. Let $0 < p < 1$ and for $i, j \in \mathbb{Z}^+$ 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 $(X_i)_{i \in \mathbb{Z}}$ with state space $\mathbb{Z}^+$ and invariant measure $\mu$, where $\mu(X_0 = i) = (1 - p) \cdot p^i$, $i \geq 0$.

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

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

The next theorem shows that VWB \(O(1/n)\) has no constraints on the \(\phi\)-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 \((X_i)_{i \in \mathbb{Z}}\) which is VWB \(O(1/n)\) and \(\phi\)-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. \(\square\)

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 \((X_i)_{i \in \mathbb{Z}}\) be \(\phi\)-mixing with \(\phi\)-mixing rate \(\phi(n)\), then \((X_i)_{i \in \mathbb{Z}}\) is strictly VWB \(\epsilon(n)\) for \(\epsilon(n) = \frac{1}{n} \sum_{i=1}^{n} \phi(i)\).

**Proof.** Theorem 6 yields

\[
\sup_{Z \in \mathcal{A}^0_{\infty}} n \rho_n(\mu, \mu(\cdot/Z)) \leq \sup_{Z \in \mathcal{A}^0_{\infty}} \sum_{i=1}^{n} |\operatorname{dist} X^n_i \mu - \operatorname{dist} X^n_i \mu(\cdot/Z)|
\leq \sup_{Z \in \mathcal{A}^0_{\infty}} \frac{1}{2} \sum_{i=1}^{n} (|\mu(B^+_i) - \mu(B^+_i/Z)| + |\mu(B^-_i) - \mu(B^-_i/Z)|)
\leq \sum_{i=1}^{n} \phi(i)
\]

where

\(B^+_i := \{(y_1, \ldots, y_n) : \mu(X_i = y_i, \ldots, X_n = y_n) \geq \mu(X_i = y_i, \ldots, X_n = y_n/Z)\}\)

and

\(B^-_i := \{(y_1, \ldots, y_n) : \mu(X_i = y_i, \ldots, X_n = y_n) < \mu(X_i = y_i, \ldots, X_n = y_n/Z)\}\). \(\square\)

In particular, we have the following consequences.

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

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

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

3. Some aspects of strictly VWB \(\epsilon(n)\) 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 \(\epsilon(n) = O(1/n)\) is the only VWB rate which forces the process to be \(\phi\)-mixing.
We use the notation of [K.O'B].

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

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

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

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

Kesten and O'Brien proved

**Theorem 14** [K.O'B]. If \(\phi(n)\) is the \(\phi\)-mixing coefficient for \((X_i)_{i\in\mathbb{Z}}\) 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 \((X_i)_{i\in\mathbb{Z}}\) is \(\phi\)-mixing \(\Leftrightarrow \mathbb{E}U_0 < \infty\).

We prove an analogous estimate for the VWB-rate.

**Theorem 15.** \((X_i)_{i\in\mathbb{Z}}\) is strictly \(\text{VWB} e(n)\) where

\[
\frac{1}{2n} \sum_{k=1}^{n} \mu(U_k \geq k) \leq e(n) \leq \frac{1}{n} \sum_{k=1}^{n} \mu(U_k \geq k).
\]

**Proof.** First we observe that

\[
n \rho_n(\mu, \mu(\cdot/Z)) \geq \sum_{i=1}^{n} |\text{dist}_{X_i,\mu} - \text{dist}_{X_i,\mu(\cdot/Z)}|
\]

by Lemma 5. Thus

\[
n e(n) \geq \sup_{Z \in \mathcal{A}_0} \sum_{i=1}^{n} |\text{dist}_{X_i,\mu} - \text{dist}_{X_i,\mu(\cdot/Z)}|
\]

\[
\geq \sup_{Z \in \mathcal{A}_0} \sum_{i=1}^{n} |\mu(U_i \geq i, V_{i-U_i} = 1) - \mu(U_i \geq i, V_{i-U_i} = 1/Z)|
\]

\[
\geq \sum_{i=1}^{n} \frac{1}{2} \mu(U_i \geq i) \left( Z_N = \bigcap_{j=0}^{N} \{V_j = 0\}, \text{ let } N \rightarrow \infty \right).
\]

So

\[
e(n) \geq \frac{1}{2n} \sum_{i=1}^{n} \mu(U_i \geq i).
\]

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

\[
\sup_{Z \in \mathcal{A}_0} \sum_{i=1}^{n} |\text{dist}_{X_i^n,\mu} - \text{dist}_{X_i^n,\mu(\cdot/Z)}| \geq \frac{1}{8} \sum_{k=1}^{n} \mu \left( \bigcup_{i \geq k} U_i \geq i \right) = \frac{1}{8} n.
\]

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

\[
\frac{1}{n} \sum_{i=1}^{n} \mu(U_i \geq i) \leq \frac{1}{8} \leq \frac{1}{n} \sup_{Z \in \mathcal{A}_0} \sum_{i=1}^{n} |\text{dist}_{X_i^n,\mu} - \text{dist}_{X_i^n,\mu(\cdot/Z)}|.
\]
Thus Theorem 6 is not strong enough in this case, because it gives a trivial upper bound. So we have to construct a measure
\[ \lambda \in P_n((X_1, \ldots, X_n)\mu(\cdot/Z), (X_1, \ldots, X_n)\mu). \]

Fix \( n, m \in \mathbb{Z}^+ \) and \( Z \in \mathcal{A}_m^0 \) of the form \( Z = \{U_0 = u_0, V_0 = v_0, V_0 - U_0 = w_0, \ldots, U_m = u_m, V_m = v_m, V_m - U_m = w_m\} \) such that
\[ \mu(Z) > 0. \]

First we have, because \((U_i)_{i \in \mathbb{Z}}, (V_i)_{i \in \mathbb{Z}}\) are i.i.d.,
\[ np_n((U_1, \ldots, U_n, V_1, \ldots, V_n)\mu, (U_1, \ldots, U_n, V_1, \ldots, V_n)\mu(\cdot/Z)) = 0. \]

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

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

Let \( J(C) \) be the indices where \( C \) does not hit \( Z \), so \( J(C) := \{1 \leq l \leq n : u_1 < l \) or \( (u_l > l + m \) and \( l - u_l \neq -i - u_{-i} \) \( \forall 0 \leq i \leq m \}). \) Then \( J(C) = \emptyset \) or \( J(C) = \{j_1, \ldots, j_r\} \) and \((X_{j_1}, \ldots, X_{j_r})\mu(\cdot/C) = (X_{j_1}, \ldots, X_{j_r})\mu(\cdot/C \cap Z). \)

If \( J(C) = \{1, \ldots, n\} \) then there is \( \lambda_C : \{0, 1\}^n \times \{0, 1\}^n \to \mathbb{R} \) such that
\[ \int \sum_{i=1}^{n} \sigma(x_i, y_i) d\lambda_C = 0. \]

If \( J(C) \neq \{1, \ldots, n\} \) then \( \{1, \ldots, n\} - J(C) = \{l_1, \ldots, l_s\}, s \geq 1. \) Then let
\[ \overline{w}_i := \begin{cases} w_{l_1 - u_{l_1} - r} & \text{if } l_i - u_{l_i} \geq -m, \\ w_r & \text{if } l_i - u_{l_i} = -r - u_{-r} \text{ for } r \in \{0, \ldots, m\}. \end{cases} \]

So we have \((\overline{w}_1, \ldots, \overline{w}_s) \in \{0, 1\}^s. \) Let \( \lambda_C : \{0, 1\}^n \times \{0, 1\}^n \to \mathbb{R} \) be defined by
\[ \lambda_C((x_1, \ldots, x_n), (y_1, \ldots, y_n)) := 0 \text{ if } (x_i, \ldots, x_i) \neq (\overline{w}_1, \ldots, \overline{w}_s) \text{ or } y_i \neq x_i \text{ for some } i \in J(C), \]
\[ \lambda_C((x_1, \ldots, x_n), (y_1, \ldots, y_n)) := \mu(V_{1-U_{l_1}} = x_1, \ldots, V_{n-U_n} = x_n/C \cap Z) \]
\[ \cdot \mu(V_{1-U_{l_1}} = y_{l_1}, \ldots, V_{l_1-U_{l_1}} = y_{l_1}/C) \text{ otherwise.} \]

Then one calculates
\[ \text{pr}_1 \lambda_C((x_1, \ldots, x_n)) = \mu(V_{1-U_{l_1}} = x_1, \ldots, V_{n-U_n} = x_n/C \cap Z), \]
\[ \text{pr}_2 \lambda_C((y_1, \ldots, y_n)) = \mu(V_{1-U_{l_1}} = y_1, \ldots, V_{n-U_n} = y_n/C), \text{ and} \]
\[
\int \sum_{i=1}^{n} \sigma(x_i, y_i) \, d\lambda_C
\]

\[
= \sum_{\{(x_1, \ldots, x_n) : x_i = w_i, i \leq s\}} \sum_{\{(y_1, \ldots, y_n) : y_i = x_i \text{ if } i \in J(C)\}} \sum_{i=1}^{n} \sigma(x_i, y_i)
\]

\[
\cdot \mu(V_{1-n} = x_1, \ldots, V_{n-n} = x_n/C \cap \mathbb{Z})
\]

\[
\cdot \mu(V_{1-n} = y_1, \ldots, V_{n-n} = y_n/C)
\]

\[
= \sum_{\{(x_1, \ldots, x_n) : x_i = w_i, i \leq s\}} \sum_{\{(y_1, \ldots, y_n) : y_i = x_i \text{ if } i \in J(C)\}} \sum_{r=1}^{s} \sigma(x_{i_r}, y_{i_r})
\]

\[
\cdot \mu(V_{1-n} = x_1, \ldots, V_{n-n} = x_n/C \cap \mathbb{Z})
\]

\[
\cdot \mu(V_{1-n} = y_1, \ldots, V_{n-n} = y_n/C)
\]

\[
\leq \sum_{\{(x_1, \ldots, x_n) : x_i = w_i, i \leq s\}} s \mu(V_{1-n} = x_1, \ldots, V_{n-n} = x_n/C \cap \mathbb{Z})
\]

\[
\leq s = \text{card } J(C)^C.
\]

So we get with \( \lambda : (\mathbb{Z}^+ \times \{0, 1\} \times \{0, 1\})^n \times (\mathbb{Z}^+ \times \{0, 1\} \times \{0, 1\})^n \rightarrow \mathbb{R} \) defined by

\[
\lambda(((u_1, v_1, w_1), \ldots, (u_n, v_n, w_n)) \times ((a_1, b_1, c_1), \ldots, (a_n, b_n, c_n)))
\]

\[
:= \mu(C) \cdot \lambda_C((w_1, \ldots, w_n), (c_1, \ldots, c_n)),
\]

\( C \) as above, if \( u_i = a_i, \ v_i = b_i \ \forall \ i \)

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

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

\[
n \rho_n(\mu, \mu(\cdot/Z)) \leq \int \sum_{i=1}^{n} \sigma(x_i, y_i) \, d\lambda = \sum_{C} \mu(C) \int \sum_{i=1}^{n} \sigma(x_i, y_i) \, d\lambda_C
\]

\[
\leq \sum_{C} \mu(C) \cdot \text{card } J(C)^C \leq \sum_{C} \mu(C) \cdot \text{card } (\{i \leq n : U_i \geq i\} \cap C)
\]

\[
= \sum_{i=1}^{n} \mu(U_i \geq i). \quad \square
\]

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

1. \( (X_i)_{i \in \mathbb{Z}} \phi\text{-mixing} \Leftrightarrow EU_0 < \infty, \ (U_i)_{i \in \mathbb{Z}} \phi\text{-mixing}, \ (V_i)_{i \in \mathbb{Z}} \phi\text{-mixing}, \) \( (V_i)_{i \in \mathbb{Z}} \VVBO(O(1/n)) \)

2. \( (X_i)_{i \in \mathbb{Z}} \VVBO(O(1/n)) \Rightarrow EU_0 < \infty, \ (U_i)_{i \in \mathbb{Z}} \VVBO(O(1/n)), \ (V_i)_{i \in \mathbb{Z}} \VVBO(O(1/n)). \)

For this one needs a Borel-Cantelli-Lemma for \( \phi\text{-mixing sequences.} \)

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 \mathbb{Z}} \text{ \(\phi\)-mixing} \iff (X_i)_{i \in \mathbb{Z}} \text{ VWB } O(1/n) \iff EU_0 < \infty.
\]

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

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

Let \((U_i)_{i \in \mathbb{Z}}, (V_i)_{i \in \mathbb{Z}}\) as above. Let \( Y_n := (V_n, V_{n-V_n}), n \in \mathbb{Z} \). Then it is not hard to see that \( EU_0 = \infty \Rightarrow (Y_i)_{i \in \mathbb{Z}} \) is not VWB \( O(1/n) \). The conjecture is
\[
EU_0 = \infty, \quad \sum_{k=1}^{\infty} \mu(U_k \geq k)^2 < \infty \Rightarrow (Y_i)_{i \in \mathbb{Z}} \text{ is } \phi\text{-mixing}.
\]

Appendix

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

\[
I_1 := \{(s_1, \ldots, s_n) \in S^n : \mu(X^n = s^n) > \mu(X^n = s^n/Z)\},
\]

\[
\overline{I}_1 := \{(s_1, \ldots, s_n) \in S^n : \mu(X^n = s^n) < \mu(X^n = s^n/Z)\},
\]

\[
\tau_1(s_1, \ldots, s_n) := \mu(X^n = s^n), \quad \tau_1(s_1, \ldots, s_n) := \mu(X^n = s^n/Z),
\]

\[
\rho_1(s_1, \ldots, s_n) := (\mu(X^n = s^n/Z) - \mu(X^n = s^n)) \cdot 1_{I_1}(s_1, \ldots, s_n),
\]

\[
\bar{\rho}_1(s_1, \ldots, s_n) := (\mu(X^n = s^n/Z) - \mu(X^n = s^n)) \cdot 1_{\overline{I}_1}(s_1, \ldots, s_n).
\]

Then inductively for \( 1 \leq k \leq n - 1 \)
\[
\tau_{k+1}(s_{k+1}, \ldots, s_n) := \sum_{(s_1, \ldots, s_k)} \rho_k(s_1, \ldots, s_k, s_{k+1}, \ldots, s_n),
\]

\[
\bar{\tau}_{k+1}(s_{k+1}, \ldots, s_n) := \sum_{(s_1, \ldots, s_k)} \bar{\rho}_k(s_1, \ldots, s_k, s_{k+1}, \ldots, s_n),
\]

\[
I_{k+1} := \{(s_{k+1}, \ldots, s_n) \in S^{n-k} : \tau_{k+1}(s_{k+1}, \ldots, s_n) > \tau_{k+1}(s_{k+1}, \ldots, s_n)\},
\]

\[
\overline{I}_{k+1} := \{(s_{k+1}, \ldots, s_n) \in S^{n-k} : \tau_{k+1}(s_{k+1}, \ldots, s_n) < \tau_{k+1}(s_{k+1}, \ldots, s_n)\},
\]

\[
\rho_{k+1}(s_1, \ldots, s_n) := \rho_k(s_1, \ldots, s_n) \left(1 - \frac{\tau_{k+1}(s_{k+1}, \ldots, s_n)}{\tau_{k+1}(s_{k+1}, \ldots, s_n)}\right) 1_{I_{k+1}}(s_{k+1}, \ldots, s_n),
\]

\[
\bar{\rho}_{k+1}(s_1, \ldots, s_n) := \bar{\rho}_k(s_1, \ldots, s_n) \left(1 - \frac{\tau_{k+1}(s_{k+1}, \ldots, s_n)}{\tau_{k+1}(s_{k+1}, \ldots, s_n)}\right) 1_{\overline{I}_{k+1}}(s_{k+1}, \ldots, s_n),
\]

\[
\tau_{n+1} := \sum_{s \in S^n} \rho_n(s), \quad \overline{\tau}_{n+1} := \sum_{s \in S^n} \bar{\rho}_n(s).
\]

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

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

\[
W_n = \{ ((x_1, \ldots, x_n), (y_1, \ldots, y_n)) : x_n \neq y_n, \rho_n(x_1, \ldots, x_n) > 0 \text{ and } \rho_n(y_1, \ldots, y_n) > 0 \},
\]

\[
W_i = \{ ((x_1, \ldots, x_n), (y_1, \ldots, y_n)) : x_i \neq y_i, x_r = y_r, i < r \leq n, \min(\tau_{i+1}(x_{i+1}, \ldots, x_n), \tau_{i+1}(x_{i+1}, \ldots, x_n)) > 0 \}
\]

for \( 1 \leq i \leq n - 1 \),

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

Then we define \( \nu_n : S^n \times S^n \to R \) in the following way:

1. \( ((s_1, \ldots, s_n), (s_1, \ldots, s_n)) \in W_0 : \nu_n((s_1, \ldots, s_n), (s_1, \ldots, s_n)) = \min(\tau_1(s_1, \ldots, s_n), \tau_1(s_1, \ldots, s_n)) \).

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 : \nu_n((a_1, \ldots, a_i, s_{i+1}, \ldots, s_n), (b_1, \ldots, b_i, s_{i+1}, \ldots, s_n))) = \min(\tau_{i+1}(s_{i+1}, \ldots, s_n), \tau_{i+1}(s_{i+1}, \ldots, s_n)) \cdot \rho_i(a_1, \ldots, a_i, s_{i+1}, \ldots, s_n)) \cdot \rho_i(b_1, \ldots, b_i, s_{i+1}, \ldots, s_n)).

3. \( ((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \in W_n : \nu_n((a_1, \ldots, a_n), (b_1, \ldots, b_n)) : = \rho_n(a_1, \ldots, a_n) \cdot \rho_n(b_1, \ldots, b_n) \cdot \frac{\min(\tau_{n+1}, \tau_{n+1})}{\tau_{n+1} \tau_{n+1}}.

4. \( \nu_n((a_1, \ldots, a_n), (b_1, \ldots, b_n)) : = 0 \text{ if } ((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \in R. \)

First we want to prove that \( \nu_n \) is a joining of

\( (X_1, \ldots, X_n) \mu \) and \( (X_1, \ldots, X_n) \mu(\cdot/Z) \).

One calculates

\[
\alpha(s^{(1)}) : = \sum_{t^{(1)} \in S^n} \nu_n(s^{(1)}, t^{(1)})\]

\[
= \nu_n(s^{(1)}, s^{(1)}) + \sum_{i=1}^{n-1} \sum_{\{t^{(1)} \in S^n : t_i \neq s_i, t_{i+1} = s_{i+1}, \ldots, t_n = s_n \}} \nu_n(s^{(1)}, t^{(1)}) + \sum_{\{t^{(1)} \in S^n : t_n \neq s_n \}} \nu_n(s^{(1)}, t^{(1)})
\]

\[
= \min(\tau_1(s^{(1)}), \tau_1(s^{(1)})) + \sum_{i=1}^{n-1} \frac{\min(\tau_{i+1}(s^{(1)}), \tau_{i+1}(s^{(1)})) \rho_i(s^{(1)}))}{\rho_n(s^{(1)})}.
\]

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
\[
\alpha(s^{(1)}) = \tilde{\tau}_1(s^{(1)}) + \rho_1(s^{(1)}) = \mu(X_1 = s_1, \ldots, X_n = s_n).
\]
General case. $s^{(1)} \in I_1$, $s^{(k)} \in I_k$, $s^{(k+1)} \notin I_{k+1}$. 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 $\tau_{n+1} = \tau_{n+1}$ 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.
\[
\tau_k(s^{(k)}) = \sum_{s_1, \ldots, s_{k-1}} \rho_{k-1}(s_1, \ldots, s_{k-1}, s_k, \ldots, s_n)
\]
\[
= \sum_{s_1, \ldots, s_{k-2}, s_{k-1}} \sum_{s_k} \rho_{k-2}(s^{(1)}) \left(1 - \frac{\tau_{k-1}(s^{(k-1)})}{\tau_{k-1}(s^{(k-1)})}\right) 1_{I_{k-1}}(s^{(k-1)})
\]
\[
= \sum_{s : (s, s_k, \ldots, s_n) \in I_{k-1}} (\tau_{k-1}(s, s_k, \ldots, s_n) - \tau_{k-1}(s, s_k, \ldots, s_n))
\]
\[
= \sum_{s : (s, s^{(k)}) \in I_{k-1}} (\tau_{k-1}(s, s^{(k)}) - \tilde{\tau}_{k-1}(s, s^{(k)}))
\]
So we get
\[
\tau_k(s^{(k)}) > \tilde{\tau}_k(s^{(k)})
\]
\[
\Leftrightarrow \sum_{s : (s, s^{(k)}) \in I_{k-1}} (\tau_{k-1}(s, s^{(k)}) - \tilde{\tau}_{k-1}(s, s^{(k)}))
\]
\[
> \sum_{s : (s, s^{(k)}) \in \tilde{I}_{k-1}} (\tilde{\tau}_{k-1}(s, s^{(k)}) - \tau_{k-1}(s, s^{(k)}))
\]
\[
\Leftrightarrow \sum_{s \in S} \tau_{k-1}(s, s^{(k)}) > \sum_{s \in S} \tilde{\tau}_{k-1}(s, s^{(k)})
\]
\[
\Leftrightarrow \sum_{s_1, \ldots, s_{k-1}} \tau_1(s_1, \ldots, s_{k-1}, s^{(k)}) > \sum_{s_1, \ldots, s_{k-1}} \tilde{\tau}_1(s_1, \ldots, s_{k-1}, s^{(k)})
\]
by repeating the argument
\[
\Leftrightarrow \mu(X_k = s_k, \ldots, X_n = s_n) > \mu(X_k = s_k, \ldots, X_n = s_n/Z)
\]
by definition of $\tau_1$, $\tilde{\tau}_1$.

This completes the proof of the claim.
Now we compute for $1 \leq i \leq n$

$$\int_{S^n \times S^n} \sigma(x_i, y_i) d\nu_n((x_1, \ldots, x_n), (y_1, \ldots, y_n))$$

$$= \nu_n(\{(s^{(1)}, t^{(1)}): s_i \neq t_i\})$$

$$\leq 1 - \sum_{j=0}^{i-1} \nu_n(\{(s^{(1)}, t^{(1)}): s_{j+1} = t_{j+1}, \ldots, s_n = t_n, s_j \neq t_j\})$$

$$= 1 - \sum_{j=0}^{i-1} \sum_{s^{(j+1)} = \bar{s}_{i+1}} \min(\tau_{j+1}(s^{(j+1)}), \bar{\tau}_{j+1}(s^{(j+1)}))$$

$$= 1 - \sum_{j=1}^{i-1} \left( \sum_{s^{(j)} \in I_j} \tau_j(s^{(j)}) + \sum_{s^{(j)} \notin I_j} \tau_j(s^{(j)}) \right)$$

$$- \left( \sum_{s^{(i-1)} \in I_{i-1}} \rho_{i-1}(s^{(1)}) + \sum_{s^{(i-1)} \notin I_{i-1}} \rho_{i-1}(s^{(1)}) \right)$$

$$= 1 - \sum_{j=1}^{i-2} \left( \sum_{s^{(j)} \in I_j} \tau_j(s^{(j)}) + \sum_{s^{(j)} \notin I_j} \tau_j(s^{(j)}) \right)$$

$$- \sum_{s^{(i-1)} \in I_{i-1}} \tau_{i-1}(s^{(i-1)}) - \sum_{s^{(i-1)} \notin I_{i-1}} \tau_{i-1}(s^{(i-1)})$$

$$- \sum_{s^{(i-1)} \in I_{i-1}, s^{(i-2)} \in \bar{I}_{i-2}} (\bar{\tau}_{i-1}(s^{(i-1)}) - \tau_{i-1}(s^{(i-1)}))$$

$$= 1 - \sum_{j=1}^{i-2} \left( \sum_{s^{(j)} \in I_j} \tau_j(s^{(j)}) + \sum_{s^{(j)} \notin I_j} \tau_j(s^{(j)}) \right)$$

$$- \sum_{s^{(i-1)} \in I_{i-1}} \tau_{i-1}(s^{(i-1)}) - \sum_{s^{(i-1)} \notin I_{i-1}} \tau_{i-1}(s^{(i-1)})$$

$$= 1 - \sum_{s^{(1)} \in I_1} \tau_1(s^{(1)}) - \sum_{s^{(1)} \notin I_1} \tau_1(s^{(1)})$$

(by repeating the argument)

$$= 1 - \sum_{s^{(1)} \in I_1} \mu(X_i = s_i, \ldots, X_n = s_n/Z) - \sum_{s^{(1)} \notin I_1} \mu(X_i = s_i, \ldots, X_n = s_n/Z)$$

$$= \sum_{s^{(1)} \in I_1} (\mu(X_i = s_i, \ldots, X_n = s_n) - \mu(X_i = s_i, \ldots, X_n = s_n/Z))$$

$$= |\text{dist} X_i^n \mu - \text{dist} X_i^n \mu(\cdot/Z)|$$

(by Claim 2).

So (2.1) is proved and therefore Theorem 6, also. $\square$
References


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