A DIRECT PROOF THAT VWB PROCESSES ARE CLOSED IN THE $\bar{d}$-METRIC

BY

JOHN C. KIEFFER[†]

ABSTRACT

As defined in the literature, a process is very weak Bernoulli if a certain property $P(\varepsilon)$ is satisfied for every $\varepsilon > 0$. By means of an easy proof, it is shown that given $\varepsilon > 0$, there exists $\delta > 0$ such that given any two stationary processes whose $\bar{d}$-distance is less than δ, if one of the processes is very weak Bernoulli then the other process is "almost" very weak Bernoulli in the sense that the property $P(\varepsilon)$ is satisfied. Using this result a direct proof can be given that the very weak Bernoulli processes are closed under the $\bar{d}$-distance, and also that a finitely determined process is very weak Bernoulli. Relativized versions of these results are also considered.

1. In this paper if S is a finite set, S^∞ denotes the set of all doubly infinite sequences $x = (x_i)_{i=-\infty}^{\infty}$ from S. We make S^∞ into a measurable space by adjoining to S^∞ the usual product σ-field of subsets of S^∞. By a process we will mean a measurable map X from some measurable space Ω to a sequence space S^∞, S finite. We say X has state space S. If $X : \Omega \to S^\infty$ is a process and i is an integer, X_i will denote the map from $\Omega \to S$ such that

$$X_i(\omega) = X(\omega)_i, \qquad \omega \in \Omega.$$

For integers m, n with $m \leqq n$, X_m^n will denote the function $(X_m, X_{m+1}, \cdots, X_n)$, $X_{-\infty}^n$ will denote the function $(\cdots, X_{n-1}, X_n)$, and if n is positive, X^n will denote $(X_1, \cdots, X_n)$. If $(\Omega, \mathcal{F})$ is a measurable space let $\mathcal{P}(\Omega)$ be the family of all probability measures on $\mathcal{F}$. If $X_1, \cdots, X_n$ are measurable maps from Ω to measurable spaces $S_1, \cdots, S_n$ respectively, and $P \in \mathcal{P}(\Omega)$, then $P(\cdot \mid X_1, \cdots, X_n)$ denotes a map from $\Omega \to \mathcal{P}(\Omega)$ such that for each $E \in \mathcal{F}$, the random variable $P(E \mid X_1, \cdots, X_n)$ serves as the conditional expectation under P of the characteristic function of E given the sub-σ-field of $\mathcal{F}$ generated by $X_1, \cdots, X_n$. If, in

[†] Supported by NSF Grant MCS-7821335.
Received January 29, 1980 and in revised form July 15, 1981

addition, X is a measurable map from Ω to a measurable space $(S, \mathcal{S})$, $P^X(\cdot \mid X_1, \cdots, X_n)$ denotes the $\mathcal{P}(S)$-valued map defined on Ω such that for each $E \in \mathcal{S}$,

$$P^X(E \mid X_1, \cdots, X_n) = P(\{X \in E\} \mid X_1, \cdots, X_n).$$

P^X denotes the distribution of X; i.e., the probability measure on $\mathcal{S}$ such that

$$P^X(E) = P(X \in E), \qquad E \in \mathcal{S}.$$

Let C, D be finite sets and $\sigma : C \times D \to [0, \infty)$ a non-negative function. Let U, V be the projections from $C \times D$ to C, D, respectively. Then if $\mu \in \mathcal{P}(C)$, $\nu \in \mathcal{P}(D)$, we define $\bar{\sigma}(\mu, \nu)$, the $\bar{\sigma}$-distance between μ and ν, as follows:

$$\bar{\sigma}(\mu, \nu) = \inf_{\lambda \in \mathcal{P}(\mu, \nu)} E_\lambda \sigma(U, V),$$

where $\mathcal{P}(\mu, \nu)$ denotes the set of all $\lambda \in \mathcal{P}(C \times D)$ such that $\lambda^U = \mu$, $\lambda^V = \nu$, and E_λ denotes expectation with respect to λ.

Fix finite sets A, B for the rest of the paper. Let T_A (T_B) denote the shift transformation on A^∞ (B^∞). Let $T_A \times T_B$ denote the map from $A^\infty \times B^\infty \to A^\infty \times B^\infty$ such that

$$(T_A \times T_B)(x, y) = (T_A x, T_B y).$$

Let $\mathcal{P}(T_A)$ $(\mathcal{P}(T_B))$ denote the set of measures in $\mathcal{P}(A^\infty)$ $(\mathcal{P}(B^\infty))$ stationary with respect to T_A (T_B). Let $\mathcal{P}(T_A \times T_B)$ denote the set of measures in $\mathcal{P}(A^\infty \times B^\infty)$ stationary with respect to $T_A \times T_B$.

Let $d : B \times B \to [0, 1]$ be the Hamming distance ($d(x, y) = 0$, $x = y$; $d(x, y) = 1$, $x \neq y$). For each $n = 1, 2, \cdots$, let $d_n : B^n \times B^n \to [0, \infty)$ denote the nth order Hamming distance, which is defined so that if $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$ are in B^n then

$$d_n(x, y) = n^{-1} \sum_{i=1}^{n} d(x_i, y_i).$$

Let X, Y be the processes which are the projections from $A^\infty \times B^\infty \to A^\infty, B^\infty$ respectively. Let $\bar{Y} : B^\infty \to B^\infty$ be the identity map. If $\lambda \in \mathcal{P}(T_A)$, let $\mathcal{P}(\lambda)$ be the set of all measures μ in $\mathcal{P}(T_A \times T_B)$ such that $\mu^X = \lambda$.

If $\nu \in \mathcal{P}(T_B)$, we say ν is very weak Bernoulli (VWB) [9, p. 92] if for any $\varepsilon > 0$, there exists a positive integer m such that

$$E_\nu \bar{d}_m(\nu^{\bar{Y}^m}, \nu^{\bar{Y}^m}(\cdot \mid \bar{Y}^0_{-\infty})) < \varepsilon.$$

If $\lambda \in \mathcal{P}(T_A)$ and $\mu \in \mathcal{P}(\lambda)$, we say μ is λ-conditionally VWB [11] if for any $\varepsilon > 0$ there exists a positive integer m such that

$$E_\mu \bar{d}_m(\mu^{Y^m}(\cdot \mid X), \mu^{Y^m}(\cdot \mid X, Y^0_{-\infty})) < \varepsilon.$$

If $\mu, \nu \in \mathcal{P}(T_B)$, then $\bar{d}(\mu, \nu)$, the $\bar{d}$-distance between μ and ν [3, p. 15], is $\inf_{(U,V)} Ed(U_0, V_0)$, where the infimum is over all pairs of jointly stationary processes (U, V) (defined on a common probability space) such that U, V have state space B, U has distribution μ, and V has distribution ν. If $\lambda \in \mathcal{P}(T_A)$ and $\mu, \nu \in \mathcal{P}(\lambda)$, then $\bar{d}_\lambda(\mu, \nu)$, the λ-relativized $\bar{d}$-distance between μ and ν [10], is $\inf_{(H,U,V)} Ed(U_0, V_0)$, where the infimum is over all triples of jointly stationary processes (H, U, V) such that H has state space A, U, V have state space B, the joint distribution of (H, U) is μ and the joint distribution of (H, V) is ν.

2. We state now the main result of the paper. Its simple proof will be given at the end of the paper.

THEOREM 1. *Let $\lambda \in \mathcal{P}(T_A)$, $\mu, \nu \in \mathcal{P}(\lambda)$. Let μ be λ-conditionally* VWB. *Then if $\bar{d}_\lambda(\mu, \nu) < \varepsilon$ there exists a positive integer m such that*

$$E_\nu \bar{d}_m(\nu^{Y^m}(\cdot \mid X), \nu^{Y^m}(\cdot \mid X, Y^0_{-\infty})) < 2\varepsilon.$$

Note that if $\lambda \in \mathcal{P}(T_A)$ is concentrated on a single element of A^∞, and $\mu, \nu \in \mathcal{P}(\lambda)$, then $\bar{d}_\lambda(\mu, \nu) = \bar{d}(\mu^Y, \nu^Y)$. Also, μ is λ-conditionally VWB if and only if μ^Y is VWB. Thus we obtain the following corollary.

COROLLARY 1. *Let $\mu, \nu \in \mathcal{P}(T_B)$. Let μ be* VWB. *If $\bar{d}(\mu, \nu) < \varepsilon$ then there exists m such that*

$$E_\nu \bar{d}_m(\nu^{\tilde{Y}^m}, \nu^{\tilde{Y}^m}(\cdot \mid \tilde{Y}^0_{-\infty})) < 2\varepsilon.$$

As an application, we present four results following from Theorem 1 and Corollary 1. The results are known, but the first published proofs of them were long and indirect. The first two results follow immediately from Theorem 1 and Corollary 1 and so require no proof.

I. Let $\lambda \in \mathcal{P}(T_A)$. If $\mu \in \mathcal{P}(\lambda)$ is the $\bar{d}_\lambda$-limit of a sequence of λ-conditionally VWB measures from $\mathcal{P}(\lambda)$, then μ is λ-conditionally VWB.

II. If $\mu \in \mathcal{P}(T_B)$ is the $\bar{d}$-limit of a sequence of VWB measures from $\mathcal{P}(T_B)$, then μ is VWB.

For the next two results we need a couple of definitions. We say $\mu \in \mathcal{P}(T_B)$ is finitely determined (FD) [9, p. 81] if convergence of any sequence $\{\mu_n\}$ from

$\mathcal{P}(T_B)$ to μ weakly and in entropy implies convergence to μ in $\bar{d}$-distance. If $\lambda \in \mathcal{P}(T_A)$ and $\mu \in \mathcal{P}(\lambda)$, μ is λ-conditionally FD [10] if convergence of any sequence $\{\mu_n\}$ from $\mathcal{P}(\lambda)$ to μ weakly and in entropy implies convergence to μ in $\bar{d}_\lambda$-distance.

III Let $\lambda \in \mathcal{P}(T_B)$ and let $\mu \in \mathcal{P}(\lambda)$ be λ-conditionally FD. Then μ is λ-conditionally VWB.

PROOF. Pick a sequence $\{\mu_n\}$ from $\mathcal{P}(\lambda)$ which converges to μ weakly and in entropy, such that for each n, there exists a positive integer m such that for μ_n-almost all $x \in A^\infty$, the μ_n-conditional distribution of Y given $X = x$ is mth order mixing Markov. Then each μ_n is λ-conditionally VWB [7] and $\bar{d}_\lambda(\mu_n, \mu) \to 0$. Hence, by Result I, μ is λ-conditionally VWB.

The last result can be proved in a similar fashion to III, or it can be observed that it follows from III by taking λ to be a trivial measure.

IV. If $\mu \in \mathcal{P}(T_B)$ is FD, it is VWB.

HISTORICAL REMARKS. IV was first shown by Ornstein and Weiss [4] using a long, indirect argument. A much simpler second proof was given by Feldman [1] using a property of measures called $\bar{d}$-extremality. III was shown by Rahe [6] using a generalization of the argument of Ornstein and Weiss. II had to be deduced previous to this paper in an indirect manner by using the fact that $\bar{d}$-limits of FD measures are FD [2, p. 66], and that a measure is FD if and only if it is VWB [4], [9, theorem 12.3]. (Or one could proceed in a similar manner using the concept of $\bar{d}$-extremality [1] in place of the FD concept in the previous sentence.) Similarly, I had to be deduced using the facts that a $\bar{d}_\lambda$-limit of λ-conditionally FD measures is λ-conditionally FD [10, prop. 7] and that a measure in $\mathcal{P}(\lambda)$ is λ-conditionally FD if and only if it is λ-conditionally VWB [6] [11, lemma 6].

3. We conclude the paper by giving a couple of lemmas and then the proof of Theorem 1.

LEMMA 1. *Let $C_1, \cdots, C_n, D_1, \cdots, D_n$ be finite sets. Let U_i $(i = 1, \cdots, n)$ be the projection from $C_1 \times \cdots \times C_n \to C_i$. For each i, let $\sigma_i : C_i \times D_i \to [0, \infty)$ be given. Let $\sigma : (C_1 \times \cdots \times C_n) \times (D_1 \times \cdots \times D_n) \to [0, \infty)$ be the function*

$$\sigma((x_1, \cdots, x_n), (y_1, \cdots, y_n)) = \sum_{i=1}^{n} \sigma_i(x_i, y_i).$$

Let $\mu \in \mathcal{P}(C_1 \times \cdots \times C_n)$, $\nu \in \mathcal{P}(D_1 \times \cdots \times D_n)$. Suppose there are measures $\nu_i \in \mathcal{P}(D_i)$ $(i = 1, \cdots, n)$ such that $\nu = \nu_1 \times \cdots \times \nu_n$. Then,

$$\bar{\sigma}(\mu, \nu) \leqq \bar{\sigma}_1(\mu^{U_1}, \nu_1) + E_\mu \left[\sum_{i=2}^{n} \bar{\sigma}_i(\mu^{U_i}(\cdot \mid U^{i-1}), \nu_i) \right].$$

PROOF. We can assume $n = 2$. (A simple induction argument can then be used for general n.) Let $\bar{U}_i, \bar{V}_i$ be the projections from $C_i \times D_i \to C_i, D_i$ respectively, $i = 1, 2$. Fix a probability measure π on $C_1 \times D_1$ such that the distribution of $\bar{U}_1$ under π is μ^{U_1}, the distribution of $\bar{V}_1$ under π is ν_1, and $E_\pi \sigma_1(\bar{U}_1, \bar{V}_1) = \bar{\sigma}_1(\mu^{U_1}, \nu_1)$. For each $x \in C_1$ fix a probability measure $\mu(\cdot \mid x)$ on C_2 such that $\mu(\cdot \mid U_1) = \mu^{U_2}(\cdot \mid U_1)$ a.e. $[\mu]$. For each $(x, y) \in C_1 \times D_1$ define a probability measure $\pi(\cdot \mid x, y)$ on $C_2 \times D_2$ such that the distribution of $\bar{U}_2$ under this measure is $\mu(\cdot \mid x)$, the distribution of $\bar{V}_2$ is ν_2 and the $\pi(\cdot \mid x, y)$ expectation of $\sigma_2(\bar{U}_2, \bar{V}_2)$ is $\bar{\sigma}_2(\mu(\cdot \mid x), \nu_2)$. Let $\hat{\pi} \in \mathcal{P}((C_1 \times C_2) \times (D_1 \times D_2))$ be the measure

$$\hat{\pi}((x_1, x_2), (y_1, y_2)) = \pi(x_1, y_1)\pi(x_2, y_2 \mid x_1, y_1).$$

Then the distribution of U under $\hat{\pi}$ is μ and the distribution of V under $\hat{\pi}$ is ν, where U, V are the projections from $(C_1 \times C_2) \times (D_1 \times D_2)$ to $C_1 \times C_2$, $D_1 \times D_2$ respectively. Hence,

$$\begin{aligned} \bar{\sigma}(\mu, \nu) &\leqq E_{\hat{\pi}} \sigma(U, V) \\ &= E_\pi \sigma_1(\bar{U}_1, \bar{U}_2) + \int_{C_1 \times D_1} \left[\int_{C_2 \times D_2} \sigma_2(\bar{U}_2, \bar{V}_2) d\pi(\cdot \mid x_1, y_1) \right] d\pi(x_1, y_1) \\ &= \bar{\sigma}_1(\mu^{U_1}, \nu_1) + \int_{C_1 \times D_1} \bar{\sigma}_2(\mu(\cdot \mid x_1), \nu_2) d\pi(x_1, y_1) \\ &= \bar{\sigma}_1(\mu^{U_1}, \nu_1) + E_\mu \bar{\sigma}_2(\mu^{U_2}(\cdot \mid U_1), \nu_2). \end{aligned}$$

The proof of the following lemma is so simple that we omit it.

LEMMA 2. *Let C, D be finite sets, and let $\sigma : C \times D \to [0, \infty)$ be a given function. Let $\mu_1, \cdots, \mu_n \in \mathcal{P}(C)$, $\nu_1, \cdots, \nu_n \in \mathcal{P}(D)$, and let $\alpha_1, \cdots, \alpha_n$ be non-negative numbers summing to one. Then,*

$$\bar{\sigma}\left(\sum_{i=1}^{n} \alpha_i \mu_i, \sum_{i=1}^{n} \alpha_i \nu_i \right) \leqq \sum_{i=1}^{n} \alpha_i \bar{\sigma}(\mu_i, \nu_i).$$

PROOF OF THEOREM 1. Fix $\lambda \in \mathcal{P}(T_A)$ and $\mu, \nu \in \mathcal{P}(\lambda)$ such that μ is λ-conditionally VWB and $\bar{d}_\lambda(\mu, \nu) < \varepsilon$. Fix ε' so that $\bar{d}_\lambda(\mu, \nu) < \varepsilon' < \varepsilon$, and also $\delta > 0$ so that $\delta + \varepsilon' < \varepsilon$. On some probability space $(\Omega, \mathcal{F}, P)$ we may find jointly stationary processes (H, U, V) such that the distribution of (H, U) is μ, the distribution of (H, V) is ν, and $P[U_0 \neq V_0] < \varepsilon'$. Since μ is λ-conditionally VWB find k such that

$$E\bar{d}_k(P^{U^k}(\cdot \mid H), P^{U^k}(\cdot \mid H, U^0_{-\infty})) < \delta.$$

For each positive integer j, let $\Pi_{i=0}^{j-1} P^{U_{ik+1}^{ik+k}}(\cdot \mid H)$ denote the $\mathcal{P}(B^{jk})$-valued map on Ω such that for each $\omega \in \Omega$, $[\Pi_{i=0}^{j-1} P^{U_{ik+1}^{ik+k}}(\cdot \mid H)](\omega)$ is the product measure

$$P^{U^k}(\cdot \mid H)(\omega) \times P^{U_{k+1}^{2k}}(\cdot \mid H)(\omega) \times \cdots \times P^{U_{(j-1)k+1}^{jk}}(\cdot \mid H)(\omega).$$

Then, by Lemma 1,

$$E\bar{d}_{jk}\left(P^{U^{jk}}(\cdot \mid H), \prod_{i=0}^{j-1} P^{U_{ik+1}^{ik+k}}(\cdot \mid H)\right) \leqq j^{-1} \sum_{i=1}^{j-1} E\bar{d}_k(P^{U_{ik+1}^{ik+k}}(\cdot \mid H, U^{ik}), P^{U_{ik+1}^{ik+k}}(\cdot \mid H))$$

$$= j^{-1} \sum_{i=1}^{j-1} E\bar{d}_k(P^{U^k}(\cdot \mid H, U^0_{1-ik}), P^{U^k}(\cdot \mid H)),$$

which approaches

$$E\bar{d}_k(P^{U^k}(\cdot \mid H, U^0_{-\infty}), P^{U^k}(\cdot \mid H)) \qquad \text{as } j \to \infty.$$

Similarly,

$$E\bar{d}_{jk}\left(P^{U^{jk}}(\cdot \mid H, U^0_{-\infty}, V^0_{-\infty}), \prod_{i=0}^{j-1} P^{U_{ik+1}^{ik+k}}(\cdot \mid H)\right)$$

(1)

$$\leqq j^{-1} \sum_{i=0}^{j-1} E\bar{d}_k(P^{U^k}(\cdot \mid H, U^0_{-\infty}, V^{-ik}_{-\infty}), P^{U^k}(\cdot \mid H)).$$

Fix i. For each $\omega \in \Omega$, the $\bar{d}_k$-distance between the measures $P^{U^k}(\cdot \mid H, U^0_{-\infty}, V^{-ik}_{-\infty})(\omega)$ and $P^{U^k}(\cdot \mid H, U^0_{-\infty})(\omega)$ is no greater than the total variation distance, which by [5, chap. 2] is upper bounded by

$$\phi\left[\sum_{b \in B^k} P^{U^k}(b \mid H, U^0_{-\infty}, V^{-ik}_{-\infty})(\omega) \log \frac{P^{U^k}(b \mid H, U^0_{-\infty}, V^{-ik}_{-\infty})(\omega)}{P^{U^k}(b \mid H, U^0_{-\infty})(\omega)}\right],$$

where $\phi : [0, \infty) \to [0, \infty)$ is the function $\phi(x) = 2x + 20\sqrt{x}$, and the logarithm is natural. Now ϕ is a concave function, so by Jensen's inequality

$$E\bar{d}_k(P^{U^k}(\cdot \mid H, U^0_{-\infty}, V^{-ik}_{-\infty}), P^{U^k}(\cdot \mid H, U^0_{-\infty})) \leqq \phi(I(U^k, V^{-ik}_{-\infty} \mid H, U^0_{-\infty})),$$

where the argument of ϕ denotes the conditional mutual information of U^k, $V^{-ik}_{-\infty}$ given H, $U^0_{-\infty}$. But, by [8, theorem 7.6], $\lim_{i \to \infty} I(U^k, V^{-ik}_{-\infty} \mid H, U^0_{-\infty}) = 0$. Therefore, the right hand side of inequality (1) converges to $E\bar{d}_k(P^{U^k}(\cdot \mid H), P^{U^k}(\cdot \mid H, U^0_{-\infty}))$ as $j \to \infty$. By the triangle inequality, we have then that for j sufficiently large,

$$E\bar{d}_{jk}(P^{U^{jk}}(\cdot \mid H), P^{U^{jk}}(\cdot \mid H, U^0_{-\infty}, V^0_{-\infty})) < 2\delta.$$

By definition of the $\bar{d}_{jk}$-distance,

$$\bar{d}_{jk}(P^{U^{jk}}(\cdot \mid H), P^{V^{jk}}(\cdot \mid H)) \leqq E[d_{jk}(U^{jk}, V^{jk}) \mid H].$$

Taking the expected value,

$$E\bar{d}_{jk}(P^{U^{jk}}(\cdot \mid H), P^{V^{jk}}(\cdot \mid H)) \leqq E[d_{jk}(U^{jk}, V^{jk})] = P[U_0 \neq V_0] < \varepsilon'.$$

Similarly,

$$E\bar{d}_{jk}(P^{U^{jk}}(\cdot \mid H, U^0_{-\infty}, V^0_{-\infty}), P^{V^{jk}}(\cdot \mid H, U^0_{-\infty}, V^0_{-\infty})) < \varepsilon',$$

and so by the triangle inequality and Lemma 2, for j sufficiently large,

$$\begin{aligned} E_\nu \bar{d}_{jk}(\nu^{Y^{jk}}(\cdot \mid X), \nu^{Y^{jk}}(\cdot \mid X, Y^0_{-\infty})) &= E\bar{d}_{jk}(P^{V^{jk}}(\cdot \mid H), P^{V^{jk}}(\cdot \mid H, V^0_{-\infty})) \\ &\leqq E\bar{d}_{jk}(P^{V^{jk}}(\cdot \mid H), P^{V^{jk}}(\cdot \mid H, V^0_{-\infty}, U^0_{-\infty})) \\ &< 2\delta + 2\varepsilon' < 2\varepsilon. \end{aligned}$$

References

1. J. Feldman, *Some remarks about finitely determined and finitely fixed processes*, preprint.

2. J. Moser, E. Phillips and S. Varadhan, *Ergodic Theory, A Seminar*, New York University, 1975.

3. D. S. Ornstein, *Ergodic Theory, Randomness, and Dynamical Systems*, Yale University Press, 1974.

4. D. S. Ornstein and B. Weiss, *Finitely determined implies very weak Bernoulli*, Israel J. Math. **17** (1974), 94–104.

5. M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, 1964.

6. M. Rahe, *Relatively finitely determined implies relatively VWB*, Canad. J. Math. **30** (1978), 531–548.

7. M. Rahe, *Finite coding factors of Markov generators*, Israel J. Math. **32** (1979), 349–355.

8. V. A. Rokhlin, *Lectures on the entropy theory of measure preserving transformations*, Russian Math. Surveys **22** (1967), 1–52.

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

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

11. J. P. Thouvenot, *Remarques sur les systèmes dynamiques donnes avec plusieurs facteurs*, Israel J. Math. **21** (1975), 215–232.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MISSOURI-ROLLA
ROLLA, MO 65401 USA