ISRAEL JOURNAL OF MATHEMATICS, Vol. 43, No. 2, 1982

FINITELY FIXED IMPLIES LOOSELY BERNOULLI, A DIRECT PROOF

BY

JOHN C. KIEFFER[†] AND MAURICE RAHE

ABSTRACT

As defined in the literature, a process is loosely Bernoulli if a certain property $P(\varepsilon)$ is satisfied for every $\varepsilon > 0$. Using only facts about stationary joinings of processes, it is shown that given $\varepsilon > 0$ there exists $\delta > 0$ such that whenever two processes are separated by less than δ in the $\bar{f}$-metric and one of them is loosely Bernoulli, the other is "almost" loosely Bernoulli in the sense that $P(\varepsilon)$ is satisfied. As easy corollaries, one has that loosely Bernoulli processes are closed in the $\bar{f}$-metric and that finitely fixed processes are loosely Bernoulli.

1. Introduction

It was shown in [3] that one can give a very short proof of the fact that Ornstein's condition "finitely determined" implies "very weak Bernoulli" based on the following theorem.

THEOREM [3; corollary 1]. *Let B be a finite set, and let $U = \{U_i\}_{i=-\infty}^{\infty}$ and $V = \{V_i\}_{i=-\infty}^{\infty}$ be stationary stochastic processes with values in the space B^{∞} of doubly infinite sequences from B. Suppose, moreover, that U is very weak Bernoulli. Then if the processes U and V are within ε in the $\bar{d}$-metric, there exists a positive integer m such that*

$$E\bar{d}_m(\text{dist } V_1^m, \text{dist}(V_1^m \mid V_{-\infty}^0)) < 2\varepsilon$$

where E indicates expectation, $\text{dist } V_1^m$ *denotes the distribution of* $V_1^m = (V_1, \cdots, V_m)$ *and* $\text{dist}(V_1^m \mid V_{-\infty}^0)$ *denotes the conditional distribution of V_1^m given the past $V_{-\infty}^0 = (\cdots, V_{-1}, V_0)$, considered as a function of $V_{-\infty}^0$.*

In this article, we derive the corresponding result for the $\bar{f}$-metric and apply it to show that finitely fixed implies loosely Bernoulli. First, for the benefit of those

[†] Research supported by NSF Grant MCS-78-21335 and the Joint Services Electronics Program under Contract N00014-79-C-0424.

Received August 5, 1980 and in revised form May 12, 1982

readers who are familiar with the relevant $\bar{d}$-concepts but who may not be familiar with [3], we provide a thumbnail sketch of the proof of the above theorem as a reference point in following the rather more involved $\bar{f}$-arguments which follow. The necessary $\bar{f}$-definitions and notation are presented in the next section, followed by the proof that FF $\Rightarrow$ LB in §3, and finally in §4 the proof of the main result: the $\bar{f}$-version of the above theorem.

Sketch of Proof. The result depends on two facts. First, we can assume U and V arise from a jointly stationary process (U, V) such that $Ed_1(U_0, V_0) = \bar{d}(\text{dist}\, U, \text{dist}\, V)$. Second, by a result of Rohlin [6, p. 66] in the jointly stationary process, no additional information is gained about U_0 from the remote past $V_{-\infty}^{-j}$, provided we know the total past $U_{-\infty}^{-1}: \lim_{j\to\infty} h(U_0 \mid U_{-\infty}^{-1} V_{-\infty}^{-j}) = h(U_0 \mid U_{-\infty}^{-1})$, where h represents entropy. From this it follows that

$$\bar{d}_1(\text{dist}(U_0 \mid U_{-\infty}^{-1} V_{-\infty}^{-j}), \text{dist}(U_0 \mid U_{-\infty}^{-1})) \to 0 \qquad \text{as } j \to \infty.$$

Since U is VWB, we choose k so that $E\bar{d}_k(\text{dist}\, U_1^k, \text{dist}(U_1^k \mid U_{-\infty}^0)) < \varepsilon$. Let π represent the j-fold product of the distribution U_1^k with itself. By using the sub-additivity of the $\bar{d}$-metric over the disjoint blocks of length k and then applying stationarity [3], we conclude that

$$\begin{aligned} E\bar{d}_{jk}(\text{dist}\, U_1^{jk}, \pi) &\leqq j^{-1} \sum_{i=1}^{j-1} E\bar{d}_k(\text{dist}(U_{ik+1}^{ik+k} \mid U_1^{ik}), \text{dist}\, U_1^k) \\ &= j^{-1} \sum_{i=1}^{j-1} E\bar{d}_k(\text{dist}(U_1^k \mid U_{1-ik}^0), \text{dist}\, U_1^k) \end{aligned}$$

which approaches $E\bar{d}_k(\text{dist}(U_1^k \mid U_{-\infty}^0), \text{dist}\, U_1^k)$ as $j \to \infty$.

Similarly,

$$E\bar{d}_{jk}(\text{dist}(U_1^{jk} \mid U_{-\infty}^0 V_{-\infty}^0), \pi) \leqq j^{-1} \sum_{i=0}^{j-1} E\bar{d}_k(\text{dist}(U_1^k \mid U_{-\infty}^0 V_{-\infty}^{-ik}), \text{dist}\, U_1^k),$$

which, by our earlier remarks, approaches $E\bar{d}_k(\text{dist}(U_1^k \mid U_{-\infty}^0), \text{dist}\, U_1^k)$ as $j \to \infty$. Hence by the triangle inequality, for j sufficiently large,

$$(0) \qquad E\bar{d}_{jk}(\text{dist}\, U_1^{jk}, \text{dist}(U_1^{jk} \mid U_{-\infty}^0 V_{-\infty}^0)) < 2\varepsilon.$$

On the other hand, by definition of the $\bar{d}$-distance we conclude that

$$E\bar{d}_{jk}(\text{dist}\, U_1^{jk}, \text{dist}\, V_1^{jk}) \leqq Ed_{jk}(U_1^{jk}, V_1^{jk}) = Ed_1(U_0, V_0) < \varepsilon.$$

Similarly, we get that

$$E\bar{d}_{jk}(\text{dist}(U_1^{jk} \mid U_{-\infty}^0 V_{-\infty}^0), \text{dist}(V_1^{jk} \mid U_{-\infty}^0 V_{-\infty}^0)) < \varepsilon.$$

By the triangle inequality, in (0) we can replace U_1^{jk} by V_1^{jk}, provided we replace "2ε" by "4ε". Dropping the condition on $U_{-\infty}^0$, we get our result.

For the $\bar{f}$-case, the approach is similar but more involved, since in the stationary joining which realizes the $\bar{f}$-distance between U and V, the processes U and V are not themselves jointly stationary. Instead, U and V must be recovered from the jointly stationary processes $\tilde{U}$, $\tilde{V}$ by the non-stationary procedure of arbitrarily assigning a time-zero and concatenating $\tilde{U}$, $\tilde{V}$ outputs. Conversely, the transition from U, V to $\tilde{U}$, $\tilde{V}$ must be effected by means of marker processes M_U, M_V which delineate the beginnings of blocks of U, V outputs which are paired in the $\bar{f}$-match.

2. Preliminaries

For S countable, let S^∞ denote the set of all infinite sequences $x = (x_i)_{i=-\infty}^{\infty}$ from S. We make S^∞ a measurable space by adjoining the usual product σ-algebra generated by the partition of S into discrete points. By a process we mean a measurable map X from some measurable space Ω to S^∞. If $X : \Omega \to S^\infty$ is a process and i is an integer, X_i denotes the map from Ω to S such that $X_i(\omega) = X(\omega)_i$, $\omega \in \Omega$. For integers m, n with $m \leqq n$, X_m^n denotes the function $(X_m, X_{m+1}, \cdots, X_n)$; $X_{-\infty}^n$, the function $(\cdots, X_{n-1}, X_n)$; and if $n > 0$, X^n denotes $(X_1, \cdots, X_n)$. 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 if $P \in \mathcal{P}(\Omega)$, then $P(\cdot \mid X_1, \cdots, X_n)$ denotes a map from Ω to $\mathcal{P}(\Omega)$ such that for each set $E \in \mathcal{F}$, the random variable $P(E \mid X_1, \cdots, X_n)$ serves as a conditional expectation under P of the characteristic function of E given the sub-σ-algebra of $\mathcal{F}$ generated by $X_1, \cdots, X_n$. If in addition X is a measurable map from Ω to a measurable space $(S, \mathcal{S})$, then $P^X(\cdot \mid X_1, \cdots, X_n)$ denotes the $\mathcal{P}(S)$-valued map defined on Ω such that for each $E \in \mathcal{S}$ we have

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

The symbol P^X denotes the distribution of X, i.e., the probability measure on $\mathcal{S}$ such that $P^X(E) = P(X \in E)$, $E \in \mathcal{S}$.

For a finite set R, let $\hat{R} = \bigcup_{n=1}^\infty R^n$, the collection of all finite sequences from R. Since R is countable, we may apply the comments of the paragraph above and consider the measurable space $\hat{R}^\infty$, processes $\hat{X} : \Omega \to \hat{R}^\infty$, etc.

For the rest of the paper, fix a finite set B. Let T_B denote the shift on B^∞; $T_{\hat{B}}$, the shift on $\hat{B}^\infty$. We shall deal with distributions on B^∞ and $\hat{B}^\infty$ which are ergodic and have finite entropy.

We shall consider two versions of the $\bar{f}$-metric, which is described in detail in [1, 2, 7]. For $b \in \hat{B}$, let $l(b)$ denote its length. Define $f : \hat{B} \times \hat{B} \to [0, \infty)$ so that

$$f(b,c) = [l(b) + l(c) - 2l \text{ (longest monotone match in } b, c)]/2.$$

It is easily shown that f is a metric. For $n = 1, 2, \cdots$, define $\hat{f}_n : \hat{B}^n \times \hat{B}^n \to [0, \infty)$ so that $\hat{f}_n(b,c) = f(\hat{b}, \hat{c})/n$, where $\hat{b}$, $\hat{c}$ are the elements in $\hat{B}$ obtained from b, c by concatenating. Let $\bar{f}_n$ be the restriction of $\hat{f}_n$ to $B^n \times B^n$. The sequence $\{\hat{f}_n\}$ is subadditive in the following sense: let $(x_1, \cdots, x_{im})$, $(y_1, \cdots, y_{im}) \in \hat{B}^{im} \times \hat{B}^{im}$, then

$$\hat{f}_{im}((x_1, \cdots, x_{im}), (y_1, \cdots, y_{im})) \leqq \sum_{j=0}^{i-1} \hat{f}_m((x_{jm+1}, \cdots, x_{jm+m}), (y_{jm+1}, \cdots, y_{jm+m})).$$

Similarly for the sequence $\{\bar{f}_n\}$. If A is countable and $\rho : A^n \times A^n \to [0, \infty)$ is given, then if μ, ν are probability measures on A^n, the symbol $\rho_n(\mu, \nu)$ denotes $\inf_{(X,Y)} E\rho_n(x, y)$, where E indicates the expectation and the infimum is over all random variables X, Y which are A^n-valued with dist $X = \mu$, dist $Y = \nu$. If ν, μ are stationary probability measures on B^∞, $\bar{f}(\mu, \nu)$, the $\bar{f}$-distance between them, is defined to be $\limsup_{n\to\infty} \bar{f}_n(\mu_n, \nu_n)$, where μ_n, ν_n are the nth order marginal distributions of μ, ν, respectively.

Finally, if V is an ergodic process with state space B and distribution ν, we say that ν is loosely Bernoulli (LB) [1, 2] if for every $\varepsilon > 0$ there exists an integer m such that $E_\nu \bar{f}(\nu^{V^m}, \nu^{V^m}(\cdot \mid V^0_{-\infty})) < \varepsilon$.

3. Results

THEOREM. *Let Y, X be ergodic processes with state space B and distributions μ, ν. Let μ be LB. Then if $\bar{f}(\mu, \nu) < \varepsilon^2$, for all k sufficiently large we have $E_\nu \bar{f}_k(\nu^{X^k}, \nu^{X^k}(\cdot \mid X^0_{-\infty})) < 84\varepsilon/(1-\varepsilon)$.*

COROLLARY. *If $\{\mu_i\}_{i=1}^\infty$, μ are T_B-ergodic and if each μ_i, $i = 1, 2, \cdots$ is LB and if the μ_i converge to μ in the $\bar{f}$-metric, then μ is LB.*

DEFINITION. We say that a T_B-ergodic measure μ is finitely fixed (FF) [7] if convergence of any sequence $\{\mu_i\}$ of T_B-ergodic measures to μ both weakly and in entropy implies convergence of the μ_i to μ in the $\bar{f}$-metric.

COROLLARY. *If a T_B-ergodic measure μ is FF, then μ is LB.*

PROOF. The mth order Markov approximants of μ converge to μ weakly and in entropy and are LB. Hence μ is LB.

REMARK. The conclusions of the two corollaries are already known. However, as Weiss points out [7, p. 0.2, p. 6.5], the proof of these facts is unsatisfactory from an aesthetic point of view, since it involves a detour through the equivalence theorem and the $\bar{d}$ result of Ornstein and Weiss that finitely determined processes are very weak Bernoulli [5]. Our proof uses only properties of the stationary joining discussed in [7, prop. 2.6] and is a modification of a recent short proof of the Ornstein–Weiss result [3].

4. In this section we present the proof of the main theorem.

In the following if μ, ν are probability measures on a common space, $|\mu - \nu|$ denotes the total variation distance between them.

LEMMA 1. *Let P, Q be probability measures on a measurable space $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is countably generated. Let $\mathcal{G}$ be any sub-σ-field of $\mathcal{F}$. Then*

$$E_P |P(\cdot \mid \mathcal{G}) - Q(\cdot \mid \mathcal{G})| \leqq 4 |P - Q|.$$

PROOF. Routine.

As a corollary we get

LEMMA 2. *Let P, Q, $(\Omega, \mathcal{F})$, and $\mathcal{G}$ be as in Lemma 1. Let $X_1, \cdots, X_k$ be measurable functions from Ω to B. Then*

$$|E_P \bar{f}_k (P^{X^k}, P^{X^k}(\cdot \mid \mathcal{G})) - E_Q \bar{f}_k (Q^{X^k}, Q^{X^k}(\cdot \mid \mathcal{G}))| \leqq 6 |P - Q|.$$

PROOF. Clearly

$$\begin{aligned} &|\bar{f}_k (P^{X^k}, P^{X^k}(\cdot \mid \mathcal{G})) - \bar{f}_k (Q^{X^k}, Q^{X^k}(\cdot \mid \mathcal{G}))| \\ &\quad \leqq \bar{f}_k (P^{X^k}, Q^{X^k}) + \bar{f}_k (P^{X^k}(\cdot \mid \mathcal{G}), Q^{X^k}(\cdot \mid \mathcal{G})). \end{aligned}$$

From Lemma 1 we get that

$$\begin{aligned} &|E_P \bar{f}_k (P^{X^k}, P^{X^k}(\cdot \mid \mathcal{G})) - E_Q \bar{f}_k (Q^{X^k}, Q^{X^k}(\cdot \mid \mathcal{G}))| \\ &\leqq E_P |\bar{f}_k (P^{X^k}, P^{X^k}(\cdot \mid \mathcal{G})) - \bar{f}_k (Q^{X^k}, Q^{X^k}(\cdot \mid \mathcal{G}))| \\ &\quad + |E_P \bar{f}_k (Q^{X^k}, Q^{X^k}(\cdot \mid \mathcal{G})) - E_Q \bar{f}_k (Q^{X^k}, Q^{X^k}(\cdot \mid \mathcal{G}))| \\ &\leqq |P - Q| + 4|P - Q| + |P - Q| = 6 |P - Q|. \end{aligned}$$

Here we used the fact that the total variation distance upper bounds the $\bar{f}$-distance (since it upper bounds the $\bar{d}$-distance).

LEMMA 3. *Let A be a countable set. For each $n = 1, 2, \cdots$, let a metric $\rho_n : A^n \times A^n \to [0, \infty)$ be given. Suppose the sequence $\{\rho_n\}$ is subadditive. Let X, Y*

be jointly stationary processes defined on $(\Omega, \mathcal{F}, P)$, *where* X *has state space* A. *Then for any* $m = 1, 2, \cdots$ *we have*

$$\begin{aligned}\limsup_{i\to\infty} E\rho_{im}(P^{X^{im}}, P^{X^{im}}(\cdot \mid X^0_{-\infty}, Y^0_{-\infty})) &\leqq 2E\rho_m(P^{X^m}, P^{X^m}(\cdot \mid X^0_{-\infty})) \\ &+ \limsup_{i\to\infty} E\rho_m(P^{X^m}(\cdot \mid X^0_{-\infty}), P^{X^m}(\cdot \mid X^0_{-\infty}, Y^{-i}_{-\infty})).\end{aligned} \tag{1}$$

Moreover, if A *is finite and* $\rho_n(x, x) = 0$ *for every* $x \in A^n$, *then the second term on the right-hand side of* (1) *is zero.*

Proof. See [3, proof of theorem 1].

The following result is easily proved using the fact that $\bar{f}$ and $\hat{f}$ are metrics.

Lemma 4. *Let* $\tilde{U}$ *be a stationary process on* $(\Omega, \mathcal{F}, P)$ *with state space* $\hat{B}$. *Let* U *be the nonstationary process with state space* B *obtained by "concatenating" the* $\tilde{U}$ *output, i.e.,* U_1^∞ *is obtained by concatenating* $\tilde{U}_1, \tilde{U}_2, \cdots$, *while* $U^0_{-\infty}$ *comes from* $\cdots, \tilde{U}_{-1}, \tilde{U}_0$. *Then for any* $n = 1, 2, \cdots$ *and any sub-σ-fields* $\mathcal{G}_1$, $\mathcal{G}_2$ *we have*

$$|E\hat{f}_m(P^{\tilde{U}^m}(\cdot \mid \mathcal{G}_1), P^{\tilde{U}^m}(\cdot \mid \mathcal{G}_2)) - E\bar{f}_m(P^{U^m}(\cdot \mid \mathcal{G}_1), P^{U^m}(\cdot \mid \mathcal{G}_2))| \leqq El(\tilde{U}_0) - 1.$$

Finally, we come to the main result. Let $\{X_i\}$ denote the projections from B^∞ to B.

Theorem. *Let* μ, ν *be ergodic on* B^∞. *Let* $\bar{f}(\mu, \nu) < \varepsilon^2$. *Let* μ *be loosely Bernoulli. Then for all* k *sufficiently large we have* $E_\nu \bar{f}_k(\nu^{X^k}, \nu^{X^k}(\cdot \mid X^0_{-\infty})) <$ $84\varepsilon/(1-\varepsilon)$.

Proof. By inequality (1), with Y a trivial process, all we need to is find some k for which $E_\nu \bar{f}_k(\nu^{X^k}, \nu^{X^k}(\cdot \mid X^0_{-\infty})) < 42\varepsilon/(1-\varepsilon)$.

By [7, proposition 2.6] there exists a probability space $(\Omega, \mathcal{F}, P)$ and processes $\tilde{U}$, $\tilde{V}$, U, V, M_U, M_V defined on it such that the following conditions hold:

(a) $\tilde{U}$, $\tilde{V}$ have state space $\hat{B}$; U, V have state space B and are obtained from $\tilde{U}$, $\tilde{V}$ respectively, by concatenation as in Lemma 4; M_U, M_V have state space $\{0, 1\}$ and $(M_U)_i = 1$ if and only if U_i is the left-most entry of the $\tilde{U}_j$ in which it lies; similarly for M_V, vis-a-vis V, $\tilde{V}$,

(b) $\tilde{U}$, $\tilde{V}$ are jointly stationary; $|P^V - \nu| < 2\varepsilon$; and $|P^{(U, M_U)} - \lambda| < 2\varepsilon$, for some stationary λ on $B^\infty \times \{0, 1\}^\infty$ with B^∞-marginal μ,

(c) $El(\tilde{U}_0) = El(\tilde{V}_0) < 1/(1-\varepsilon)$, and

(d) with probability one, for each i, the outputs $\tilde{U}_i$, $\tilde{V}_i$ begin with the same element of B.

Let $\bar{X}$, $\bar{Y}$ be the processes which are the projections from $B^\infty \times \{0, 1\}^\infty \to B^\infty$,

$\{0,1\}^{\infty}$ respectively. Let τ represent the joint distribution $\tau = P^{(U,M_U)}$. By Lemma 2, for any m,

$$E\bar{f}_m(P^{U^m}, P^{U^m}(\cdot \mid U^0_{-\infty}, (M_U)^0_{-\infty})) = E\bar{f}_m(\tau^{\bar{X}^m}, \tau^{\bar{X}^m}(\cdot \mid \bar{X}^0_{-\infty}, \bar{Y}^0_{-\infty}))$$
$$\leqq E\bar{f}_m(\lambda^{\bar{X}^m}, \lambda^{\bar{X}^m}(\cdot \mid \bar{X}^0_{-\infty}, \bar{Y}^0_{-\infty})) + 6|\tau - \lambda|.$$

From condition (b) above, $|\tau - \lambda| < 2\varepsilon$, so the second term on the right is bounded by 12ε. As m gets large, the first term goes to zero by Lemma 3. (To see this, note that since B is finite, the second term on the right side of (1) is identically zero. The first term on the right side of (1) goes to zero since μ is LB.) Hence we see that we may fix m so that

$$E\bar{f}_m(P^{U^m}, P^{U^m}(\cdot \mid U^0_{-\infty}, (M_U)^0_{-\infty})) < 12\varepsilon. \tag{2}$$

Now Lemma 4 yields that

$$E\hat{f}_m(P^{\tilde{U}^m}, P^{\tilde{U}^m}(\cdot \mid \tilde{U}^0_{-\infty})) \leqq E\bar{f}_m(P^{U^m}, P^{U^m}(\cdot \mid \tilde{U}^0_{-\infty})) + \varepsilon/(1-\varepsilon),$$

since by hypothesis (c) $El(\tilde{U}) < 1/(1-\varepsilon)$. Since the first term is bounded by 12ε from (2), we conclude that

$$E\hat{f}_m(P^{\tilde{U}^m}, P^{\tilde{U}^m}(\cdot \mid \tilde{U}^0_{-\infty})) < 13\varepsilon/(1-\varepsilon). \tag{3}$$

Since the entropies $H(\tilde{U}_0)$ and $H(\tilde{V}_0)$ are finite [4, theorem 4] and since $(\tilde{U}, \tilde{V})$ are jointly stationary, theorem 6.2 [6, p. 66] assures us that the conditional mutual information $I(U^m, \tilde{V}^{-i}_{-\infty} \mid \tilde{U}^0_{-\infty}) \to 0$ as i gets large. Hence

$$\limsup_{i\to 0} E\bar{f}_m(P^{U^m}(\cdot \mid \tilde{U}^0_{-\infty}), P^{U^m}(\cdot \mid \tilde{U}^0_{-\infty}, \tilde{V}^{\,i}_{-\infty})) = 0$$

by exactly the same argument given in the proof of theorem 1 of [3]. Thus since $El(\tilde{U}_0) - 1 < \varepsilon/(1-\varepsilon)$ by condition (c) above, it follows immediately from Lemma 4 that

$$\limsup_{i\to\infty} E\hat{f}_m(P^{\tilde{U}^m}(\cdot \mid \tilde{U}^0_{-\infty}), P^{\tilde{U}^m}(\cdot \mid \tilde{U}^0_{-\infty}, \tilde{V}^{-i}_{-\infty})) < \varepsilon/(1-\varepsilon). \tag{4}$$

Therefore, by reapplying Lemma 3 for k any sufficiently large multiple of m, we get that

$$E\hat{f}_k(P^{\tilde{U}^k}, P^{\tilde{U}^k}(\cdot \mid \tilde{U}^0_{-\infty}, \tilde{V}^0_{-\infty})) < 27\varepsilon/(1-\varepsilon). \tag{5}$$

(The terms on the right-hand side of Equation (1) are bounded above in (3) and (4).)

On the other hand, both $E\hat{f}_k(P^{\tilde{U}^k}(\cdot \mid \tilde{U}^0_{-\infty}, \tilde{V}^0_{-\infty}), P^{\tilde{V}^k}(\cdot \mid \tilde{U}^0_{-\infty}, \tilde{V}^0_{-\infty}))$ and

$E\hat{f}_k(P^{\tilde{U}^k}, P^{\tilde{V}^k})$ are dominated by $E\hat{f}_k(\tilde{U}^k, \tilde{V}^k) \leqq E[(l(\tilde{U}^k) + l(\tilde{V}^k) - 2k)/(2k)] < \varepsilon/(1-\varepsilon)$. From these two bounds and (5), the triangle inequality yields that

$$E\hat{f}_k(P^{\tilde{V}^k}, P^{\tilde{V}^k}(\cdot \mid \tilde{U}^0_{-\infty}, \tilde{V}^0_{-\infty})) < 29\varepsilon/(1-\varepsilon). \tag{6}$$

Since $E\bar{f}_k(P^{V^k}, P^{V^k}(\cdot \mid V^0_{-\infty})) \leqq E\bar{f}_k(P^{V^k}, P^{V^k}(\cdot \mid \tilde{V}^0_{-\infty}))$ and since Lemma 4 yields $E\bar{f}_k(P^{V^k}, P^{V^k}(\cdot \mid \tilde{V}^0_{-\infty})) \leqq E\hat{f}_k(P^{\tilde{V}^k}, P^{\tilde{V}^k}(\cdot \mid \tilde{V}^0_{-\infty})) + \varepsilon/(1-\varepsilon)$, we apply (6) to conclude that

$$E\bar{f}_k(P^{V^k}, P^{V^k}(\cdot \mid V^0_{-\infty})) < 30\varepsilon/(1-\varepsilon). \tag{7}$$

Finally, by Lemma 2,

$$\begin{aligned} E_\nu \bar{f}_k(\nu^{X^k}, \nu^{X^k}(\cdot \mid X^0_{-\infty})) &\leqq E\bar{f}_k(P^{V^k}, P^{V^k}(\cdot \mid V^0_{-\infty})) + 6|P^V - \nu| \\ &\leqq 42\varepsilon/(1-\varepsilon), \end{aligned}$$

since the term $|P^V - \nu|$ is less than 2ε by condition (b) above and the previous term is bounded in (7).

References

1. J. Feldman, *Non-Bernoulli K-automorphisms and a problem of Kakutani*, Isr. J. Math. **24** (1976), 16–38.

2. A. Katok, *Change of time, monotone equivalence, and standard dynamical systems*, Doklady Akad. Nauk **223** (1975), 789–792.

3. John C. Kieffer, *A direct proof that VWB processes are closed in the d-metric*, Isr. J. Math. **41** (1982), 154–160.

4. Roman A. Murieka, *The maximization of entropy of discrete denumerably valued random variables with known mean*, Ann. Math. Stat. **43** (1972), 541–552.

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

6. W. Parry, *Entropy and Generators in Ergodic Theory*, W. A. Benjamin, New York, 1969.

7. B. Weiss, *Equivalence of measure preserving transformations*, Lecture Notes, The Institute for Advanced Studies, The Hebrew University of Jerusalem, 1975.

Department of Mathematics and Statistics
University of Missouri-Rolla
Rolla, MO 65401 USA

Department of Mathematics
Texas A & M University
College Station, TX 77843 USA