ISRAEL JOURNAL OF MATHEMATICS **126** (2001), 157–171

# LOOSE ALMOST BLOWING-UP AND RELATED PROPERTIES

BY

PAUL C. SHIELDS*

*Mathematics Department, The University of Toledo*
*Toledo, OH 43606, USA*
*e-mail: paul.shields@utoledo.edu*

ABSTRACT

It is shown that finitely-fixed processes have $\bar{f}$-metric versions of the almost blowing-up property, the approximate-match waiting-time property, and the strong form of extremality. The proofs make use of a new result that connects, in a local almost sure way, fixed-length blocks in a process with fixed-length blocks in an expansion of the process.

## 1. Introduction

Several results have recently been established for the class of finitely-determined processes, including a blowing-up characterization, [2], an empirical distribution result, [5, 4], and an approximate-match waiting-time theorem, [2]. Analogues of these results for the class of finitely-fixed processes and the $\bar{f}$-metric are established in this paper, along with a strong form of extremality.

Throughout the paper $A$ denotes a finite set, $x_m^n$ denotes the finite sequence $x_m, x_{m+1}, \ldots, x_n$, $A^n$ denotes the set of all $x_1^n$, $A^\infty$ denotes the set of infinite sequences $x = \{x_n: n \geq 1\}$, $A^Z$ denotes the set of all doubly-infinite sequences $x = \{x_n: n \in Z\}$, and $T$ denotes the shift on $A^\infty$ or its extension to $A^Z$. A stationary measure $\mu$ is a $T$-invariant Borel probability measure on $A^\infty$ or the stationary extension of such a measure to $A^Z$. The distribution on $A^n$ defined by $\mu_n(a_1^n) = \mu(\{x: x_1^n = a_1^n\})$, $a_1^n \in A^n$, is denoted by $\mu_n$, with the subscript omitted

* Supported in part by joint NSF–Hungarian Academy grant INT-9515485.
Received August 25, 1999

when it is understood. The entropy of the stationary measure $\mu$ is denoted by $H(\mu)$. The sequence $\{X_n\}$ of random variables defined by $X_n(x) = x_n, n \in Z$, is called the process defined by $\mu$. In a slight abuse of language, "stationary process" will mean either a stationary measure on $A^\infty$ or $A^Z$, or the process defined by such a measure. Unless stated otherwise such processes are taken to be ergodic.

The per-symbol Hamming distance $d_n(x_1^n, y_1^n)$ is the fraction of indices $i \leq n$ for which $x_i \neq y_i$. The $\bar{d}_n$-distance between probability distributions $\mu$ and $\nu$ on $A^n$ is defined by the formula

$$\bar{d}_n(\mu,\nu) = \min_\lambda E_\lambda(d_n(x_1^n, y_1^n)),$$

where $E_\lambda$ denotes expected value and the minimum is over all distributions $\lambda$ on $A^n \times A^n$ that have $\mu$ and $\nu$ as marginals. The extension to stationary measures is defined by the formula

$$\bar{d}(\mu,\nu) = \lim_{n\to\infty} \bar{d}_n(\mu_n, \nu_n).$$

The distributional distance between two probability measures on $A^n$ is defined by

$$|\mu - \nu|_k = \sum_{a_1^k} |\mu(a_1^k) - \nu(a_1^k)|.$$

A stationary measure $\mu$ is finitely determined (FD) if given $\epsilon > 0$ there exists $k$ and $\delta > 0$ such that $\bar{d}(\mu,\nu) \leq \epsilon$, for any ergodic measure $\nu$ for which $|\mu-\nu|_k \leq \delta$ and $|H(\mu) - H(\nu)| \leq \delta$.

The $(d_n, \delta)$-blowup of a set $C \subseteq A^n$ is the $\delta$-neighborhood of $C$ with respect to the $d_n$-metric, that is, the set $[C]_{d_n,\delta}$ defined by the formula

$$[C]_{d_n,\delta} = \{x_1^n\colon \min_{y_1^n \in C} d_n(x_1^n, y_1^n) \leq \delta\}.$$

A stationary measure $\mu$ has the **almost blowing-up property (ABUP)** if there is a sequence $B_n \subset A^n, n \geq 1$, such that $x_1^n \in B_n$, eventually almost surely, and such that given $\delta > 0$, there is a $\gamma > 0$ and an $N$, such that $\mu([C]_{d_n,\delta}) \geq 1-\delta$, for any $n \geq N$ and any set $C \subset B_n$ for which $\mu(C) \geq 2^{-\gamma n}$.

Given $\delta > 0$, the $d_n$-approximate-match waiting time is defined by

$$W_n(x, y, d_n, \delta) = \min\{m \geq 1\colon d_n(y_m^{m+n-1}, x_1^n) \leq \delta\}.$$

The following results were established in [2].

THEOREM 1:

1. *A process is finitely determined if and only if it has the almost blowing-up property.*
2. *If $\mu$ is finitely determined then*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log W_n(x,y,d_n,\delta)=H(\mu),$$

*almost surely with respect to the product measure $\mu\times\mu$.*

The per-symbol substitution/deletion distance $f_n(x_1^n, y_1^n)$ is the fraction of substitutions and deletions needed to change $x_1^n$ into $y_1^n$, that is, $f_n(x_1^n, y_1^n) = 1 - t/n$, where $t$ is the length of the longest subsequence of $x_1^n$ that agrees with a subsequence of $y_1^n$. The distances $\bar{f}_n(\mu,\nu)$ and $\bar{f}(\mu,\nu)$, the definitions of finitely fixed (FF), of the $(f_n,\delta)$-blowup, of the **loose almost blowing-up property (LABUP)**, and of the $f_n$-approximate-match waiting time function $W_n(x,y,f_n,\delta)$ are obtained by replacing $d_n$ by $f_n$ in the definitions of the preceding paragraphs.

The principal theorem to be established in this paper is

THEOREM 2:

1. *A process is finitely fixed if and only if it has the loose almost blowing-up property.*
2. *If $\mu$ is finitely fixed then*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log W_n(x,y,f_n,\delta)=H(\mu),$$

*almost surely with respect to the product measure $\mu\times\mu$.*

The proof that LABUP implies FF is obtained by simply replacing $d_n$ by $f_n$ throughout the proof that ABUP implies FD given in [2]. Thus the only real issue in proving the first half of the theorem is to show that FF implies LABUP. Two ideas will be used. The first is that a finitely-fixed process can always be expanded to a finitely-determined process by inserting spacers between symbols in an appropriate manner; this follows from the basic theory described in [3]. The second idea is a connection between fixed length blocks in an ergodic process and fixed-length blocks in an expansion of the process. The expansion concept and the key expansion lemma, Lemma 2, which is an almost sure local form of an unpublished observation of Jack Feldman, [1], are discussed in Section 2. Theorem 2 is proved in Section 3 by a direct application of the expansion lemma.

With more effort a somewhat stronger form of $\bar{f}$-extremality, which is of independent interest, can be also be established. This is given in Section 4. The stronger form of $\bar{f}$-extremality yields Theorem 2 by simply replacing $d_n$ by $f_n$ in the corresponding FD proofs. Likewise, it shows immediately that FF processes have the $\bar{f}$-analogue of the $\bar{d}$-recognition property of [5, 4].

In summary, two proofs of Theorem 2 are given, one utilizing strong $\bar{f}$-extremality to copy the proof of Theorem 1, the other a more direct argument utilizing the expansion lemma to connect $\bar{f}$-blowup for contractions with $\bar{d}$-blowup for expansions.

ACKNOWLEDGEMENT: The author benefitted from discussions with Don Ornstein, who suggested the idea used to prove Theorem 4, with Benjy Weiss, who suggested the proof of part of Theorem 3, and with Xuehong Li, whose criticisms led to many clarifications and simplifications.

## 2. The expansion lemma

The expansion and contraction language used in this paper is a special case of the towering and inducing/factoring language commonly used in ergodic theory. Informally, contraction is obtained by deleting all occurrences of a given symbol and expansion by inserting variable-length blocks of a new symbol between old symbols. The formal definition is given in the next paragraph.

Let $Y = \{Y_n\}$ be a stationary, ergodic $(A \cup \{s\})$-valued process, where $s \notin A$. For $Y_0 \in A$, put $r_0 = 0$ and let $r_{j+1}$ be the minimum $n > r_j$ such that $Y_n \in A$. The $A$-valued process defined by

$$X_n = Y_{r_n}, \quad n = 0, 1, 2, \ldots$$

is called the **contraction** of $\{Y_n\}$ to the times when $Y_n \in A$, while $\{Y_n\}$ is called an **expansion** of $\{X_n\}$. (See [7] for a discussion of these ideas.) For convenience we subtract 1 and focus on the number of steps before return, that is, the function defined for $y \in (A \cup \{s\})^Z, y_0 \in A$, by $g(y) = 0$, if $y_1 \in A$, and

$$g(y) = n, \quad \text{if } y_i \notin A,\ 1 \le i \le n, \text{ and } y_{n+1} \in A.$$

The expected length of the spacer blocks is denoted by $E(g)$.

In the sequel, the distribution of the contracted process $\{X_n\}$ is usually denoted by $\mu$, the distribution of an expanded process $\{Y_n\}$ by $\mu^g$, and $\mu^g$ is called an **expansion of $\mu$ with spacer symbol $s$ and spacer function $g(\cdot)$**.

For $y_1^m \in (A \cup \{s\})^m$, let $\mathcal{L}(y_1^m)$ denote the number of indices $i \in [1, m]$ for which $y_i \in A$, and let

$$t(1), t(2), \ldots, t(\mathcal{L}(y_1^m)) \tag{1}$$

be the subsequence of those indices where $y_i \in A$. If $\mathcal{L}(y_1^m) \geq n$, let $\pi_n(y_1^m) = x_1^n$, where $x_i = y_{t(i)}$, $1 \leq i \leq n$.

Fix $\delta > 0$, let

$$m(n) = n(1+\delta/2)(1+E(g)), \quad n \geq 1, \tag{2}$$

and put

$$\Gamma_{m(n)}(x_1^n) = \{y_1^{m(n)}\colon\ n \leq \mathcal{L}(y_1^{m(n)}) \leq (1+\delta)n, \text{ and } \pi_n(y_1^{m(n)}) = x_1^n\}, \tag{3}$$

and

$$\Gamma_{m(n)} = \bigcup_{x_1^n} \Gamma_{m(n)}(x_1^n) = \{y_1^{m(n)}\colon\ n \leq \mathcal{L}(y_1^{m(n)}) \leq (1+\delta)n\}.$$

The ergodic theorem implies that $\lim_{n\to\infty} \mu^g(\Gamma_{m(n)}) = 1$. The following is a stronger, local form of this fact.

LEMMA 1: $\lim_{n\to\infty} \frac{\mu^g(\Gamma_{m(n)}(x_1^n))}{\mu(x_1^n)} = 1$, *$\mu$-almost surely.*

*Proof:* For $n \geq 1$, put

$$G_{m(n)} = \{y\colon n' \leq \mathcal{L}(y_1^{m(n')}) \leq n'(1+\delta),\ \forall n' \geq n\},$$

and note that $G_{m(n)} \subseteq G_{m(n+1)}$, and $\mu^g(G_{m(n)}) \to 1$, by the ergodic theorem. Let $\Gamma^*_{m(n)}(x_1^n)$ denote the set of all $y \in (A \cup \{s\})^\infty$ for which $y_1^{m(n)} \in \Gamma_{m(n)}(x_1^n)$, and define the function

$$\phi_n(x) = \frac{\mu^g(\Gamma^*_{m(n)}(x_1^n) \cap G_{m(n)})}{\mu(x_1^n)}, \quad x \in A^\infty.$$

To prove the lemma, it is enough to show that $\phi_n(x) \to 1$, almost surely, since $\mu^g(G_{m(n)}) \to 1$. Towards this end, first note that

$$0 \leq \phi_n(x) \leq \frac{\mu^g(\Gamma_{m(n)}(x_1^n))}{\mu(x_1^n)} \leq 1, \quad n \geq 1,$$

since the set of $y$, for which $\mathcal{L}(y_1^m) \geq n$ and $\pi_n(y) = x_1^n$ for some $m$, contains the set $\Gamma^*_{m(n)}(x_1^n)$ and has $\mu^g$-measure equal to $\mu(x_1^n)$. Furthermore, the expected value of $\phi_n(x)$ goes to 1, as $n \to \infty$, since $\mu^g(\Gamma^*_{m(n)} \cap G_{m(n)}) \to 1$, where $\Gamma^*_{m(n)} = \bigcup \Gamma^*_{m(n)}(x_1^n)$. Thus, to complete the proof of the lemma, it is enough to show that $\{\phi_n\}$ is a supermartingale with respect to the filtration $\{\mathcal{F}_n\}$, where $\mathcal{F}_n$ denotes the $\sigma$-field generated by the $n$-dimensional cylinders, $\{x\colon x_1^n = a_1^n\}$, $a_1^n \in A^n$.

To establish the supermartingale property first note that the relation $G_{m(n)} \subseteq G_{m(n+1)}$ and the definition (3) of $\Gamma_{m(n)}(x_1^n)$ imply that for each $x_1^n$ the following holds:

$$\begin{aligned}\bigcup_{x_{n+1}} \Gamma^*_{m(n+1)}(x_1^{n+1}) \cap G_{m(n+1)} &\supseteq \bigcup_{x_{n+1}} \Gamma^*_{m(n+1)}(x_1^{n+1}) \cap G_{m(n)} \\ &= \Gamma^*_{m(n)}(x_1^n) \cap G_{m(n)}.\end{aligned}$$

This, in turn, gives

$$\begin{aligned}E(\phi_{n+1}|\mathcal{F}_n)(x) &= \sum_{x_{n+1}} \frac{\mu^g(\Gamma^*_{m(n+1)}(x_1^{n+1}) \cap G_{m(n+1)})}{\mu(x_1^{n+1})} \mu(x_{n+1}|x_1^n) \\ &\geq \frac{\mu^g(\Gamma^*_{m(n)}(x_1^n) \cap G_{m(n)})}{\mu(x_1^n)} = \phi_n(x),\end{aligned}$$

which completes the proof of the lemma. ∎

The proofs given in the next section will make use of the fact that the set $\Gamma_{m(n)}$ can be required to consist of "typical" sequences, as stated in the following lemma.

LEMMA 2 (The expansion lemma): *Suppose $\{B_m \subseteq (A \cup \{s\})^m\colon m \geq 1\}$ satisfies the condition that $y_1^m \in B_m$, $\mu^g$-eventually almost surely. Then*

$$\lim_{n\to\infty} \frac{\mu^g(\Gamma_{m(n)}(x_1^n) \cap B_{m(n)})}{\mu(x_1^n)} = 1, \quad \mu\text{-almost surely.}$$

*Proof:* This follows from the observation that the preceding proof applies with $G_{m(n)}$ replaced by $G'_{m(n)} = G_{m(n)} \cap_{k\geq n} \{y\colon y_1^{m(k)} \in B_{m(k)}\}$. ∎

## 3. Proof of Theorem 2

We make use of the fact that finitely-fixed processes have finitely-determined expansions. Only the existence of such expansions is needed in this section, but the stronger fact that expected spacer length can be made arbitrarily small will be needed in the next section, so we state the result in the following form.

THEOREM 3: *If $\mu$ is finitely fixed and $\epsilon > 0$, then it has a finitely-determined expansion whose spacer function has expected value at most $\epsilon$.*

*Proof:* First suppose $\mu$ has positive entropy. We can then apply Corollary 5.5 in [3] taking $\mu$ as $S$ and $T$ to be any FD process with smaller entropy, and labelling the complement of $A$ by $s$ to obtain a finitely-determined expansion. The desired result follows since we can assume the difference between the entropies of $S$ and $T$ is arbitrarily small. For zero entropy $\mu$, we first insert a single symbol $t \notin A$ with probability $p < \epsilon/2$, independently between each symbol of $A^Z$. This gives an FF expansion $\tilde{\mu}$ of $\mu$ of positive entropy whose spacer function has expected value $p$, to which the preceding argument can be applied, after which $t$ is replaced by $s$. ∎

By the preceding theorem, to show that FF implies LABUP it is enough to show that contractions of ABUP processes are LABUP processes. The expansion lemma shows that sets of "typical" $n$-sequences that are not exponentially too small lift to sets of "typical" $m(n)$-sequences that are also not exponentially too small, hence have large $\bar{d}$-blowups. This, combined with a simple bound on $f_n$-distance in terms of $d_{m(n)}$-distance, then implies that $\mu$ has the $\bar{f}$-almost blowing-up property. The simple bound, using the notation of (1), (2), and (3) from the preceding section, is stated as the following lemma.

LEMMA 3: *If $y_1^{m(n)} \in \Gamma_{m(n)}(x_1^n)$ and $\tilde{y}_1^{m(n)} \in \Gamma_{m(n)}(\tilde{x}_1^n)$, then*

$$f_n(x_1^n, \tilde{x}_1^n) \leq \delta + \frac{m(n)}{n} d_{m(n)}(y_1^{m(n)}, \tilde{y}_1^{m(n)}).$$

*Proof:* First note that if $t_{\min}$ is the minimum of $t(\mathcal{L}(y_1^{m(n)}))$ and $t(\mathcal{L}(\tilde{y}_1^{m(n)}))$, then

$$nf_n(x_1^n, \tilde{x}_1^n) \leq n - |\{i \leq t_{\min} : y_i = \tilde{y}_i \in A\}|.$$

On the other hand,

$$n - |\{i \leq t_{\min} : y_i = \tilde{y}_i \in A\}| \leq \delta n + m(n) d_{m(n)}(y_1^{m(n)}, \tilde{y}_1^{m(n)}),$$

by the definition of $\Gamma_{m(n)}(\cdot)$ and by considering the worst case when all the disagreements between $y_i$ and $\tilde{y}_i$ occur in the places where $i \leq t_{\min}$. The lemma follows. ∎

The following lemma states in precise form the fact that contractions of ABUP processes are LABUP processes.

LEMMA 4: *Let $\mu$ be a contraction of an ergodic $\mu^g$ that has the $\bar{d}$-almost blowing-up property. Given $0 < \epsilon < 1$, there is a $\gamma > 0$, an $N$, and sets $B_n \subseteq A^n$, $n \geq N$, such that $x_1^n \in B_n$, $\mu$-eventually almost surely, and such that if $n \geq N$ and $C \subseteq B_n$ has measure at least $2^{-\gamma n}$, then $\mu([C]_{f_n, \epsilon}) \geq 1 - \epsilon$.*

*Proof:* Fix $\epsilon > 0$ such that $\epsilon < 1$, and let $\alpha$ be a positive number to be determined later. Since $\mu^g$ has the $\bar{d}$-almost blowing-up property, there is a

$\beta > 0$, an $M$, and sets $B^g_m \subseteq (A \cup \{s\})^m$, $m \geq M$, such that $y_1^m \in B^g_m$, eventually almost surely, and such that the following holds:

(4) If $m \geq M, B \subseteq B^g_m$, and $\mu^g(B) \geq 2^{-\beta m}$, then $\mu^g([B]_{d_m,\alpha}) \geq 1-\alpha$.

Let $\delta = \epsilon/2$ and $K = (1+\delta/2)(1+E(g))$. For each $n$, put $m(n) = Kn$, and for each $x_1^n$, let $\Gamma_{m(n)}(x_1^n)$ be the set given by (3). Choose $N > (\beta K)^{-1}$ so that $m(n) \geq M$ whenever $n \geq N$.

For $n \geq N$ put

$$B_n = \left\{x_1^n \colon \mu^g(\Gamma_{m(n)}(x_1^n) \cap B^g_{m(n)}) \geq (1-\delta)\mu(x_1^n)\right\},$$

and note that $x_1^n \in B_n$, $\mu$-eventually almost surely, by Lemma 2, since $y_1^m \in B^g_m$, eventually almost surely.

Put $\gamma = \beta K - 1/N$, suppose $n \geq N$ and suppose $C \subseteq B_n$ has measure at least $2^{-\gamma n}$. The goal is to show that if $N$ was chosen large enough, the set $C$ has an $f_n$-blowup of large $\mu_n$-measure, for appropriate choice of $\alpha$. Towards this end, lift $C$ to the set

$$B = \bigcup_{x_1^n \in C} \Gamma_{m(n)}(x_1^n) \cap B^g_{m(n)}.$$

The definition of $B_n$ guarantees that

$$\mu^g(B) \geq (1-\delta) \sum_{x_1^n \in C} \mu(x_1^n) = (1-\delta)\mu(C) \geq (1-\delta)2^{-\gamma n}.$$

Since $\delta = \epsilon/2 < 1/2$ and $\gamma = \beta K - 1/N$, it follows that

$$\mu^g(B) \geq (1-\delta)2^{-\gamma n} \geq 2^{-\beta m(n)},$$

so property (4) yields

$$\mu^g([B]_{d_{m(n)},\alpha}) \geq 1-\alpha.$$

It remains to show how this implies that $\mu([C]_{f_n,\epsilon}) \geq 1-\epsilon$, for $N$ large enough and suitable choice of $\alpha$. Towards this end, let

$$G_{m(n)} = \bigcup_{x_1^n \in B_n} \Gamma_{m(n)}(x_1^n) \cap [B]_{d_{m(n)},\alpha}, \quad n \geq N.$$

Since $\mu(B_n) \to 1$, Lemma 2 implies that by making $N$ larger, if necessary, it can be supposed that $\mu^g(G_{m(n)}) \geq 1-2\alpha$. This means, however, that

$$\mu\Big(\pi_n(G_{m(n)})\Big) \geq 1-2\alpha, \quad n \geq N, \tag{5}$$

since

$$\begin{aligned}\mu\Big(\pi_n(G_{m(n)})\Big) &= \sum_{x_1^n \in \pi_n(G_{m(n)})} \mu^g(\{y\colon\ \mathcal{L}(y) \geq n, \text{ and } \pi_n(y) = x_1^n\}) \\ &\geq \sum_{x_1^n \in \pi_n(G_{m(n)})} \mu^g(\Gamma_{m(n)}(x_1^n)) \\ &\geq \sum_{x_1^n \in \pi_n(G_{m(n)})} \mu^g(\Gamma_{m(n)}(x_1^n) \cap [B]_{d_{m(n)},\alpha}) \\ &= \mu^g(G_{m(n)}).\end{aligned}$$

The final fact needed is that

$$\pi_n(G_{m(n)}) \subset [C]_{f_n,\delta+K\alpha}. \tag{6}$$

To prove this suppose $y_1^{m(n)} \in G_{m(n)}$. Since $y_1^{m(n)} \in [B]_{d_{m(n)},\alpha}$ there is a $\widetilde{y}_1^{m(n)} \in B$ such that

$$d_{m(n)}(y_1^{m(n)}, \widetilde{y}_1^{m(n)}) \leq \alpha. \tag{7}$$

Let $\pi_n(y_1^{m(n)}) = x_1^n$, which exists since $y_1^{m(n)} \in \bigcup_{x_1^n \in B_n} \Gamma_{m(n)}(x_1^n)$, and let $\pi_n(\widetilde{y}_1^{m(n)}) = \widetilde{x}_1^n$, which exists and belongs to $C$, by definition of $B$. Lemma 3 and the bound (7) then give $f_n(x_1^n, \widetilde{x}_1^n) \leq \delta + K\alpha$. This proves the claim (6).

The results (6) and (5) combine to give

$$\mu([C]_{f_n,\delta+K\alpha}) \geq 1 - 2\alpha, \quad n \geq N.$$

Lemma 4 follows from this, since $\delta = \epsilon/2$ and $\alpha > 0$ can be chosen to satisfy $K\alpha < \epsilon/2$ and $\alpha < \epsilon/2$. ∎

An exponential lower bound on the measure of the support of the empirical distribution of $k$-blocks, which is valid eventually almost surely for any ergodic process provided only that $k$ is not too large relative to path length $n$, is all that is needed to establish the second part of Theorem 2. Let $\hat{p}_k$ denote the empirical distribution of overlapping $k$-blocks in the sequence $x_1^n$, that is, the distribution defined by

$$\hat{p}_k(a_1^k) = \hat{p}_k(a_1^k|x_1^n) = \frac{|\{i \in [1, n-k]\colon\ x_i^{i+k-1} = a_1^k\}|}{n-k+1}, \tag{8}$$

and let $\hat{\mathcal{S}}_k$ denote the support of $\hat{p}_k$.

LEMMA 5: *If $\mu$ is an ergodic measure with entropy $H$ and $\epsilon > 0$, then there is a $K$ such that $\mu(\hat{\mathcal{S}}_k) \geq 2^{-2k\epsilon}$, for all $k \in [K, \frac{\log n}{H+\epsilon}]$, eventually almost surely as $n \to \infty$.*

*Proof:* Let $B_k = \{x_1^k \colon \mu(x_1^k) \geq 2^{-k(H+\epsilon)}\}$. The entropy theorem yields a $K$ such that $\hat{p}_k(B_k) \geq 1-\epsilon$, for all $k \in [K, n\epsilon/2]$, eventually almost surely as $n \to \infty$; see Theorem 7 in [7]. As shown in [5], it is also true that $|B_k \cap \hat{\mathcal{S}}_k| \geq 2^{k(H-\epsilon)}$, for all $k \in [K, \frac{\log n}{H+\epsilon}]$, eventually almost surely as $n \to \infty$. Thus

$$\mu(\hat{\mathcal{S}}_k) \geq \mu(B_k \cap \hat{\mathcal{S}}_k) \geq 2^{-k(H+\epsilon)} 2^{k(H-\epsilon)} \geq 2^{-k2\epsilon},$$

for all $k \in [K, \frac{\log n}{H+\epsilon}]$, eventually almost surely as $n \to \infty$. ∎

COROLLARY 1: *If $\mu$ is a finitely-fixed measure of entropy $H$ and $\epsilon > 0$ then $\mu([\hat{\mathcal{S}}_k]_{f_k,\epsilon}) \geq 1-\epsilon$, for $k \sim \frac{\log n}{H+\epsilon}$, eventually almost surely.*

*Proof:* This follows immediately from the preceding lemma and the fact that FF implies LABUP. ∎

The only properties used to prove the $\bar{d}$-form of the approximate-match waiting-time theorem were the almost blowing-up property and the $\bar{d}$-analogue of Corollary 1 (which clearly holds since FD implies ABUP); see [2]. Thus the second part of Theorem 2 now follows by replacing $d_k$ by $f_k$.

## 4. A finite form of finitely-fixed

The finitely-determined property has a finite form which is a stronger version of the extremality concept discussed in [6]. To state this, we use the following notation. If $\nu$ is a measure on $A^n$, and $k \leq n$, **the $\nu$-distribution of $k$-blocks in $n$-blocks** is the measure $\phi = \phi(k, \nu)$ on $A^k$ defined by the formula

$$\phi(a_1^k) = \sum_{x_1^n} \hat{p}_k(a_1^k | x_1^n) \nu(x_1^n) = E_\nu(\hat{p}_k(a_1^k | X_1^n)),$$

that is, the expected value with respect to $\nu$ of the empirical $k$-block distribution $\hat{p}_k$ defined by (8). A stationary measure $\mu$ is said to be **finitely finitely-determined (FFD)** if given $\epsilon > 0$, there is a $\delta > 0$ and positive integers $k$ and $N$ such that if $n \geq N$ then any measure $\nu$ on $A^n$ for which $|\mu - \phi(k,\nu)|_k \leq \delta$ and $|H(\mu_n) - H(\nu)| \leq n\delta$ must also satisfy $\bar{d}_n(\mu, \nu) \leq \epsilon$. Theorem IV.2.2 in [8] can be rephrased as follows.

THEOREM 4: *FD is equivalent to FFD.*

The definition of **finitely finitely-fixed (FFF)** is obtained by replacing $d_n$ by $f_n$ in the definition of FFD. Our goal in this section is to establish the following.

THEOREM 5: *FF is equivalent to FFF.*

*Proof:* It is obvious that FFF implies FF, so the only issue is to show that FF implies FFF. Let $\mu$ be a finitely-fixed measure with entropy $H(\mu)$, and fix $\epsilon > 0$. Let $k$ and $\ell$ be positive integers and $\alpha, \beta$, and $\delta$ be positive numbers, to be determined later. Theorem 3 provides a finitely-determined expansion $\mu^g$ for which $E(g) \leq \delta$.

The functions $\mathcal{L}(\cdot), \pi.(\cdot)$, and $m(\cdot)$ and the sets $\Gamma_{m(\cdot)}(\cdot)$ are defined as earlier; see (1), (2), and (3). By the expansion lemma, we can suppose $\ell$ is large enough so that, if $L = m(\ell)$, then

$$\mu^g(\Gamma_L) \geq 1 - \alpha/2 \quad \text{and} \quad \mu(\pi_\ell(\Gamma_L)) \geq 1 - \alpha/2.$$

For $K = m(k)$, let $V_K$ be the set of all $y_1^K \in \Gamma_K$ such that $\mu^g(y_1^K) > 0$ and $\mu(\pi_k(y_1^K)) > 0$ and such that the following conditions hold:

$$\text{(a)} \qquad x_1^k = \pi_k(y_1^K) \Longrightarrow \sum_{a_1^\ell \in \pi_\ell(\Gamma_L)} |\hat{p}(a_1^\ell|\, x_1^k) - \mu(a_1^\ell)| \leq \alpha.$$

$$\text{(b)} \qquad \sum_{b_1^L \in \Gamma_L} |\hat{p}(b_1^L|\, y_1^K) - \mu^g(b_1^L)| \leq \alpha.$$

By the expansion lemma and the ergodic and entropy theorems, we can suppose $k$ is large enough that

$$\text{(9)} \qquad \mu^g(V_K) \geq 1 - \alpha \quad \text{and} \quad \mu(\pi_k(V_K)) \geq 1 - \alpha.$$

Since $\mu^g$ is finitely determined we can suppose $\ell$ is large enough so there exists $\gamma > 0$ and $N_0$ such that $\bar{d}_N(\mu_N^g, \nu^*) \leq \delta^2$, for any $N \geq N_0$ and any measure $\nu^*$ on $(A \cup \{s\})^N$ that satisfies the two conditions

$$\text{(10)} \qquad \text{(i) } |\mu_L^g - \phi(L, \nu^*)|_L \leq \gamma, \quad \text{(ii) } |H(\mu_N^g) - H(\nu^*)| \leq N\gamma.$$

Given a positive integer $n$, let $\nu$ be a measure on $A^n$ that satisfies the two conditions

$$\text{(11)} \qquad \text{(i) } |\mu_k - \phi(\nu, k)|_k \leq \beta, \quad \text{(ii) } |H(\mu_n) - H(\nu)| \leq n\beta.$$

The goal is to show that if $\beta$ is small enough, then, by making $k$ and $\ell$ larger and $\alpha$ and $\delta$ smaller, if necessary, there will be an $n_0$ such that $\bar{f}_n(\mu,\nu) \le \epsilon$, for any $n \ge n_0$ and any measure $\nu$ on $A^n$ satisfying the conditions (11).

To achieve the goal, we let $N = m(n)$ and show how to construct a measure $\nu^*$ on $(A \cup \{s\})^N$ which is close to $\mu^g$ in $\ell$-th order distribution and entropy, in such a way that by removing the spacers and, possibly, a few other terms, $d_N$-close sequences project to $f_n$-close sequences. The key idea is to partition sample paths into disjoint blocks, most of which belong to $\pi_k(V_K)$, then add spacer to such $k$-blocks, independent of anything outside the $k$-blocks, using the conditional distribution determined by $\mu^g$. Standard arguments show that such partitions exist with large $\nu$-probability. The definition of $V_K$ guarantees closeness of $L$-block distributions, while the independence condition guarantees closeness in entropy.

In the construction it is sufficient to consider the conditional measure on "typical" $K$-blocks, given $k$-blocks, that is, the distribution defined by the formula

$$\mu^*(y_1^K|x_1^k) = \frac{\mu^g(y_1^K)}{\mu^g(V_K \cap \Gamma_K(x_1^k))}, \quad y_1^K \in V_K \cap \Gamma_K(x_1^k). \tag{12}$$

To carry out our strategy, let $W_n$ be the set of all sequences $x_1^n$ that can be expressed as a concatenation

$$x_1^n = w(1)w(2)\cdots w(J)$$

such that the total length of the $w(j)$ that do not belong to $\pi_k(V_K)$ is at most $2n\sqrt{\alpha+\beta}$.

LEMMA 6: *If $n$ is large enough then $\nu(W_n) \ge 1 - \sqrt{\alpha+\beta}$.*

*Proof:* First note that $\phi(\pi_k(V_K)) \ge 1 - (\alpha+\beta)$, since $\mu(\pi_k(V_K)) \ge 1-\alpha$ and $|\mu_k - \phi(\nu,k)|_k \le \beta$. The Markov inequality implies that the set of $x_1^n$ for which $\hat{p}_k(\pi_k(V_K)|x_1^n) \ge 1 - \sqrt{\alpha+\beta}$ has $\nu$-measure at least $1-\sqrt{\alpha+\beta}$. But if $\hat{p}_k(\pi_k(V_K)|x_1^n) \ge 1 - \sqrt{\alpha+\beta}$, the packing lemma, Lemma I.3.3 of [8], implies that $x_1^n \in W_n$, provided only that $k/n \le \sqrt{\alpha+\beta}$. ∎

The measure $\nu^*$ is defined as follows. If $x_1^n \notin W_n$, extend it to $x_1^n s^{N-n}$ and define $\nu^*(x_1^n s^{N-n}) = \nu(x_1^n)$. If $x_1^n \in W_n$ express it as the concatenation $x_1^n = w(1)w(2)\cdots w(J)$, where the total length of the $w(j)$ that do not belong to $\pi_k(V_K)$ is at most $2n\sqrt{\alpha+\beta}$. For each $j \le J$, let $\widetilde{w}(j) = w(j)$, if $w(j) \notin \pi_k(V_K)$; otherwise, choose $\widetilde{w}(j) \in V_K \cap \Gamma_K(w(j))$, form the concatenation

$$y_1^N = \widetilde{w}(1)\widetilde{w}(2)\cdots\widetilde{w}(J)\widetilde{w}(J+1), \tag{13}$$

where $\widetilde{w}(J+1)$ is a block of $s$'s with length chosen so that the total length is $N$, and define

$$\nu^*(y_1^N) = \nu(x_1^n) \prod_{w(j)\in\pi_k(V_K)} \mu^*(\widetilde{w}(j)|w(j)).$$

In particular, note that if $w(i) \in \pi_k(V_K)$ then $\widetilde{w}(i) \in V_K \cap \Gamma_K(w(j))$ is chosen according to the conditional distribution (12), independent of the values of $x_1^n$ that lie outside $w(j)$, and the values of $y_1^N$ that lie outside $\widetilde{w}(j)$. Let $W_N^*$ denote the set of those $y_1^N$ of the form (13) that come by extending in this way the members of $W_n$.

This completes the definition of $\nu^*$.

The next task is to show that if $\alpha$ and $\beta$ are small enough and $k$ large enough, then $\nu^*$ satisfies the distributional condition of (10). The distributional distance

$$\sum_{b_1^L} |\sum_{y_1^N} \hat{p}(b_1^L|y_1^N)\nu^*(y_1^N) - \mu^g(b_1^L)|$$

is upper bounded by

$$2\sqrt{\alpha+\beta} + \sum_{y_1^N\in W_N^*} \sum_{b_1^L} |\hat{p}(b_1^L|y_1^N) - \mu^g(b_1^L)|\nu^*(y_1^N), \tag{14}$$

by Lemma 6. On the other hand, if $|\widetilde{w}(j)|$ denotes sequence length, then for any $y_1^N$ of the form (13) the frequency $(N-L+1)\hat{p}(b_1^L|y_1^N)$ is upper bounded by

$$\sum_{\widetilde{w}(j)\in V_K} (K-L+1)\hat{p}(b_1^L|\widetilde{w}(j)) + \sum_{\widetilde{w}(j)\notin V_K} |\widetilde{w(j)}| + L\frac{N}{K},$$

where the second term bounds the number of $L$-blocks that could start in the $\widetilde{w}(j)$ that do not belong to $V_K$ and the final term bounds the number of $L$-blocks that could start in one of the last $L-1$ places of the $\widetilde{w}(j)$ that belong to $V_K$. Omitting the last term on the second sum yields the bound $2N\sqrt{\alpha+\beta}$, by the definition of $W_n$, while $|\widetilde{w}(J+1)| \leq 2N\sqrt{\alpha+\beta}$, also by the definition of $W_n$. Increasing $K$ if necessary it can also be assumed that the third term is bounded by $N\sqrt{\alpha+\beta}$. Combining these facts with (14), we conclude that if $K$ is large enough and $\alpha$ and $\beta$ are small enough, then

$$|\mu_L - \phi(L,\nu^*)|_L \leq \gamma,$$

provided $N$ is sufficiently large. Thus, part (i) of (10) can be forced to hold.

The measure $\nu^*$ was defined so as to guarantee closeness in entropy to $\mu_N^g$. To establish this, random vector notation will be used. Let $\widetilde{X}_1^k$ and $\widetilde{Y}_1^K$ denote random vectors with the respective conditional distributions $\mu_k(\cdot|\pi_k(V_K))$ and $\mu_K^g(\cdot|V_K)$. We can suppose $k$ is large enough and $\alpha$ small enough that

$$\left|\frac{1}{K}H(\widetilde{Y}_1^K|\widetilde{X}_1^k) - \frac{1}{N}H(\mu_N^g) + \frac{1}{n}H(\mu_n)\right| \leq \gamma/2. \tag{15}$$

Our construction guarantees that

$$\frac{1}{N}H(\nu^*) = \frac{1}{K}H(\widetilde{Y}_1^K|\widetilde{X}_1^k) + \frac{1}{n}H(\nu_n) + f(\alpha,\beta),$$

where $f(\alpha,\beta) \to 0$, as $\sqrt{\alpha+\beta} \to 0$. Combined with (15), this means that $|H(\mu_N^g) - H_N(\nu^*)| \leq N\gamma$ holds if $\alpha$ and $\beta$ are small enough.

In summary, if $k$ is chosen large enough and $\alpha$ and $\beta$ small enough, then the condition (10) for $\bar{d}$-closeness holds and hence $\bar{d}_N(\mu^g,\nu^*) \leq \delta^2$. Projecting downwards by removing the spacers and any extra symbols drawn from $A$ that are needed to create those $\widetilde{w}(j) \in V_K$ for which $\mathcal{L}(\widetilde{w}(j)) > k$ and using an obvious modification of Lemma 3, we arrive at the conclusion that $\bar{f}_n(\mu,\nu) \leq \epsilon$, for suitable choice of $\delta,\alpha,\beta,\ell$, and $k$, for all sufficiently large $n$ for any $\nu$ that satisfies the distribution and entropy conditions of (11). This completes the proof that FF implies FFF. ∎

*Remark 1:* Theorem 2 and the $\bar{f}$-recognition property for finitely-fixed processes follow immediately from Theorem 5, merely by replacing $d_n$ by $f_n$ in the corresponding $\bar{d}$ proofs, since those proofs only made use of the fact that FD implies FFD.

## References


[1] J. Feldman, *Some remarks about finitely determined and finitely fixed processes,* unpublished notes.

[2] K. Marton and P. Shields, *Almost sure waiting time results for weak and very weak Bernoulli processes,* Ergodic Theory and Dynamical Systems **15** (1995), 951–960.

[3] D. Ornstein, D. Rudolph and B. Weiss, *Equivalence of measure preserving transformations,* Memoirs of the American Mathematical Society **262** (1982).

[4] D. Ornstein and P. Shields, *The $\bar{d}$-recognition of processes,* Advances in Mathematics **104** (1994), 182–224.

[5] D. Ornstein and B. Weiss, *How sampling reveals a process,* Annals of Probability **18** (1990), 905–930.

[6] D. Ornstein and B. Weiss, *Entropy and isomorphism theorems for actions of amenable groups,* Journal d'Analyse Mathématique **48** (1987), 1–141.

[7] P. Shields, *The interactions between ergodic theory and information theory,* IEEE Transactions on Information Theory **IT-44** (1998), 2079–2093.

[8] P. Shields, *The Ergodic Theory of Discrete Sample Paths,* American Mathematical Society Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 1996.