On variable length coding of asymptotically mean stationary measures*

Łukasz Dębowski
debowski@cwi.nl
Centrum voor Wiskunde en Informatica
Kruislaan 413, NL-1098 SJ Amsterdam
The Netherlands

Abstract

We discuss variable length coding of doubly infinite sequences. The attention is paid to the problem how the coding affects probability measures defined for the sequences. Particularly studied is the class of asymptotically mean stationary (AMS) measures, which is closed under the investigated coding. We discuss explicit formulae for the stationary means and conditions for: preserving stationarity, equivalence of shift invariant σ-fields, preserving finite energy property, and equality of mutual information.

Key words: asymptotically mean stationary processes, variable length coding, complete fix-free sets, shift invariant algebras, finite energy processes, mutual information

MSC 2000: 60G10, 28D99, 94A45, 94A17

Running head: On variable length coding of AMS measures

*The research reported in this work was supported in part by the Polish Ministry of Scientific Research and Information Technology, grant no 1/P03A/045/28, and the IST Programme of the European Community, under the PASCAL II Network of Excellence, IST-2002-506778. This publication reflects only the author's views.

1 Introduction

Let $\mathbb{X}$ and $\mathbb{Y}$ be a pair of countable sets, called alphabets, for which a concatenation operation is defined. As usual, denote the set of nonempty strings over an alphabet $\mathbb{X}$ as $\mathbb{X}^+ := \bigcup_{n\in\mathbb{N}} \mathbb{X}^n$ and the set of all strings as $\mathbb{X}^* := \mathbb{X}^+ \cup \{\lambda\}$, where λ is the empty string. For a function $f : \mathbb{X} \to \mathbb{Y}^*$ that maps single symbols into strings, define its extension $f^* : \mathbb{X}^* \to \mathbb{Y}^*$, where

$$f^* : x_1 x_2 ... x_m \mapsto f(x_1) f(x_2) ... f(x_m) \tag{1}$$

for $x_i \in \mathbb{X}$. This function, which transforms strings into strings of possibly variable albeit finite length, is a fundamental concept in coding theory.

In probabilistic analyses, however, it is often convenient to consider doubly infinite sequences $\mathbf{x} = (x_i)_{i\in\mathbb{Z}}$, $x_i \in \mathbb{X}$. Thus it becomes also necessary to generalize (1) as

$$f^*(\mathbf{x}) = \mathbf{y} = (y_i)_{i\in\mathbb{Z}}, \quad y_i \in \mathbb{Y}, \tag{2}$$

where where for every m there exist k and l such that $y_1 y_2 ... y_k = f^*(x_1 x_2 ... x_m)$ and $y_{-l} y_{-l+1} ... y_0 = f^*(x_{-m} x_{-m+1} ... x_0)$.

In this note, we collect several observations on how the mapping f^* transforms the distribution of a stochastic process $(X_i)_{i\in\mathbb{Z}}$, $X_i : \Omega \to \mathbb{X}$, cf. [14, 6]. Assume that f^* is an injection. It was noticed in [14, Example 6] that if the "shrunk" process $(X_i)_{i\in\mathbb{Z}}$ is stationary then the "expanded" process

$$(Y_i)_{i\in\mathbb{Z}} := f^*((X_i)_{i\in\mathbb{Z}}), \quad Y_i : \Omega \to \mathbb{Y}, \tag{3}$$

is asymptotically mean stationary (AMS). This observation will be complemented with the following new results:

(i) As a generalization, we will show that the image and preimage of an AMS process are AMS under quite general f^*. Thus the class of AMS processes is almost closed under (2) (Section 3).

(ii) The image of a stationary process under ${f^*}^{-1}$ is stationary if f is complete fix-free (Section 4).

(iii) The shift invariant algebras for the shrunk and the expanded processes, as well as for their stationary means, are equivalent if f^* is a synchronizable injection (Section 5).

(iv) The image of a finite energy process [18] under f^* is a finite energy process if f is prefix-free and the image of f is finite (Section 6).

(v) For a fixed length injection $f : \mathbb{X} \to \mathbb{Y}^K$, block entropies of the stationary means of $(X_i)_{i\in\mathbb{Z}}$ and $(Y_i)_{i\in\mathbb{Z}}$ can be related quite easily (Section 7).

Some basic properties of AMS processes are briefly recalled in Section 2.

The selection of topics is motivated by interest in constructing a finite alphabet analogue of a nonergodic stationary process $(X_i)_{i\in\mathbb{Z}}$ over an infinite alphabet $\mathbb{X} = \mathbb{N} \times \{0, 1\}$, where

$$X_i := (K_i, Z_{K_i}). \tag{4}$$

In this example, $(Z_k)_{k\in\mathbb{N}}$ and $(K_i)_{i\in\mathbb{Z}}$ are assumed independent. Binary sequence $(Z_k)_{k\in\mathbb{N}}$ is a sequence of independent uniformly distributed variables, $P(Z_k = 0) = P(Z_k = 1) = 1/2$. Variables K_i are also IID but satisfy $P(K_i = k) = k^{-1/\beta}/\zeta(1/\beta)$, $k \in \mathbb{N}$, $\beta \in (0, 1)$. Denote the blocks of variables $X_{k:l} := X_k X_{k+1} ... X_l$. At least for $\beta < 0.365$, it can be shown easily that the block mutual information for (4) is $I(X_{-n+1:0}; X_{1:n}) = \Omega(n^\beta)$, which stems from large mutual information between blocks $X_{k:l}$ and the shift invariant σ-field of the process being nonatomic, cf. [7].[1] The process $(X_i)_{i\in\mathbb{Z}}$ is also a finite energy process in the sense of Shields [18].

We have been interested whether a suitable choice of transformation f^* for (4) can produce a process (3) over a finite alphabet $\mathbb{Y} = \{0, 1, ..., D-1\}$ whose stationary mean $(\bar{Y}_i)_{i\in\mathbb{Z}}$ has properties similar to $(X_i)_{i\in\mathbb{Z}}$. If $(\bar{Y}_i)_{i\in\mathbb{Z}}$ actually had the nonatomic shift invariant σ-field, finite energy, and mutual information $I(\bar{Y}_{-n+1:0}; \bar{Y}_{1:n}) = \Omega(n^\beta)$ then it would be an interesting simplistic stochastic model of texts in natural language, of a new kind: It would enjoy both a power law growth of vocabulary for its shortest grammar based coding [8] and a simple formal semantic interpretation stemming from its particular form of nonergodicity [7]. Both properties seem relevant for probabilistic language modeling.

The general results collected in this article are too weak to yield a definitive answer to our original specific question but provide some initial insights.

2 Preliminaries

Denote the product measurable space of doubly infinite sequences $(\mathbb{U}, \mathcal{U}) = \times_{k\in\mathbb{Z}}(\mathbb{X}, \mathcal{X})$ and the shift operation

$$T(\mathbf{x}) = (x_{i+1})_{i\in\mathbb{Z}}. \tag{5}$$

For a probability space $(\Omega, \mathcal{J}, P)$ that supports process $(X_i)_{i\in\mathbb{Z}}$, where $X_i : (\Omega, \mathcal{J}) \to (\mathbb{X}, \mathcal{X})$, let

$$\mu = P((X_i)_{i\in\mathbb{Z}} \in \cdot)$$

be its distribution on $(\mathbb{U}, \mathcal{U})$. An AMS measure μ can be equivalently characterized either as such (i) that the ergodic theorem is satisfied, i.e., for every bounded measurable function $g : (\mathbb{U}, \mathcal{U}) \to (\mathbb{R}, \mathcal{R})$ limit

$$\lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} g \circ T^{+i}$$

exists μ-almost everywhere (it need not be constant, though), or (ii) that limit

$$\bar{\mu}(A) = \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \mu \circ T^{-i}(A) \tag{6}$$

exists for all $A \in \mathcal{U}$ [14, Theorem 1]. Moreover, the limit $\bar{\mu}$, if it exists, forms a measure on $(\mathbb{U}, \mathcal{U})$ called the stationary mean. It is in fact stationary, i.e., $\bar{\mu} \circ T^{-1} = \bar{\mu}$.

Trivially, $\bar{\mu} = \mu$ for a stationary μ. In the general AMS case, equality

$$\bar{\mu}(A) = \mu(A) \tag{7}$$

[1] This citation will be replaced by a more appropriate one once its details are known.

is satisfied for each set A in the T-invariant algebra $\mathcal{I}_{\mathbb{U}} := \{A \in \mathcal{U} : T^{-1}A = A\}$. This follows directly from (6), cf. [14]. Extending the concept for stationary measures, an AMS measure μ is called *ergodic* if $\mu(A) \in \{0, 1\}$ for all $A \in \mathcal{I}_{\mathbb{U}}$.

Stationary means enjoy a simple but useful frequency interpretation. To write it down, let us introduce some notations. The Iverson bracket will be written as $\mathbf{1}_{\{\tau\}}$, i.e., $\mathbf{1}_{\{\tau\}} := 1$ if τ is true and $\mathbf{1}_{\{\tau\}} := 0$ else. Also, let $|u|$ be the length of string u and

$$[u] := \{(x_i)_{i\in\mathbb{Z}} : x_{1:|u|} = u\}$$

stand for a cylinder set spanned by u, where we switch to using notation $x_{k:l} := x_k x_{k+1} ... x_l$ for brevity. On the other hand, for a finite string v, let $\{v\}_{k:l}$ be a notation for such a substring that $v = v' \{v\}_{k:l} v''$, $|v'| = k-1$, and $|v''| = |v|-l$. We also use $\{v\}_k := \{v\}_{k:k}$.

Lemma 1 *We have equality*

$$\bar{\mu}(A) = \int \left[\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{T^i\mathbf{x}\in A\}} \right] d\mu(\mathbf{x}). \tag{8}$$

In particular for $A = [u]$, we have $\mathbf{1}_{\{T^i\mathbf{x}\in A\}} = \mathbf{1}_{\{x_{i+1:i+|w|}=u\}}$. Moreover, in the case of an ergodic process the integrated expression is μ-almost everywhere constant.

Proof: By the dominated convergence theorem,

$$\begin{aligned}\bar{\mu}(A) = \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \mu(T^{-i}A) &= \lim_{n\to\infty} n^{-1} \int \left[\sum_{i=0}^{n-1} \mathbf{1}_{\{\mathbf{x}\in T^{-i}A\}} \right] d\mu(\mathbf{x}) \\ &= \int \left[\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{T^i\mathbf{x}\in A\}} \right] d\mu(\mathbf{x}).\end{aligned}$$

□

It is said that a measure τ dominates μ, written $\tau \gg \mu$, if $\tau(A) = 0$ implies $\mu(A) = 0$ for all measurable A. The following lemma resumes two well known facts [17], [11, Theorem 0]:

Lemma 2 *The measure μ is AMS if and only if there exists a stationary measure $\tau \gg \mu$. In the latter case, $\tau \gg \bar{\mu} \gg \mu$.*

Proof: Firstly, if $\tau \gg \mu$ then the limits integrated in (8) exist not only τ-almost everywhere (by the ergodic theorem) but also μ-almost everywhere. Thus $\bar{\mu}$ exists.

On the other hand, assume that μ is an AMS measure. If there existed A such that $\mu(A) > \bar{\mu}(A) = 0$ then surely we would have $\mu(B) > \bar{\mu}(B) = 0$ for $B = \bigcup_{i\in\mathbb{Z}} T^i A$. But B is shift invariant so $\mu(B) = \bar{\mu}(B)$. Thus, our assumption was false and there is $\bar{\mu} \gg \mu$. Moreover, if $\tau \gg \mu$ then $\tau(A) = 0$ implies $\mu(T^{-i}A) = 0$ and hence $\bar{\mu}(A) = 0$ as well. □

Lemma 2 may inspire a slightly different justification of Lemma 1. Using directly the definition of conditional probability $\mu(A||\mathcal{I}_{\mathbb{U}})$, which is $\mathcal{I}_{\mathbb{U}}$-measurable, and identity (7) yields

$$\bar{\mu}(A) = \int \bar{\mu}(A||\mathcal{I}_{\mathbb{U}})d\bar{\mu} = \int \bar{\mu}(A||\mathcal{I}_{\mathbb{U}})d\mu = \int \left[\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{T^i\mathbf{x}\in A\}}\right] d\bar{\mu}$$

since, by the ergodic theorem, equality

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{T^i\mathbf{x}\in A\}} = \bar{\mu}(A||\mathcal{I}_{\mathbb{U}})$$

holds $\bar{\mu}$-almost everywhere and thus μ-almost everywhere.

3 Measures of expanded processes

Consider partial sums $S_+(\mathbf{x}, n) := \sum_{i=1}^{n} |f(x_i)|$ and $S_-(\mathbf{x}, n) := \sum_{i=-n+1}^{0} |f(x_i)|$. If $\lim_n S_\pm(\cdot, n) = \infty$ holds μ-almost everywhere then (3) defines a doubly infinite sequence of random variables with distribution

$$\nu = P((Y_i)_{i\in\mathbb{Z}} \in \cdot) = \mu \circ {f^*}^{-1}$$

on $(\mathbb{W}, \mathcal{W}) = \times_{k\in\mathbb{Z}}(\mathbb{Y}, \mathcal{Y})$.

For a process $(Y_i)_{i\in\mathbb{Z}}$, Kieffer and Gray [14, Example 6] considered variable length shift

$$T^* = f^* \circ T \circ {f^*}^{-1}, \tag{9}$$

which constitutes a function $f^*(\mathbb{U}) \to f^*(\mathbb{U})$ for an injection f^*. Assuming T^* being defined, $(Y_i)_{i\in\mathbb{Z}}$ is variable length stationary for a stationary $(X_i)_{i\in\mathbb{Z}}$, i.e., $\nu \circ {T^*}^{-1} = \nu$. The calculation of a stationary dominant $\rho \gg \nu$ given for this case in [14, Example 6] can be generalized to a noninjection f^* and an AMS μ.

Observe that for a general f we have a quasiperiodic identity

$$T^{|f(x_1)|} f^*(\mathbf{x}) = f^*(T\mathbf{x}). \tag{10}$$

Thus we may apply the generic idea of constructing a stationary measure by a randomized shift within the quasiperiod, cf. [6, 15].

Theorem 1 *Let $|f(x)| > 0$ for all $x \in \mathbb{X}$. If the stationary mean $\bar{\mu}$ exists and the expected expansion rate*

$$L := \int |f(x_1)|\, d\bar{\mu}(\mathbf{x}) \tag{11}$$

is in the range $(0, \infty)$ then there exists a stationary measure ρ on $(\mathbb{W}, \mathcal{W})$ satisfying

$$\rho(A) = \frac{1}{L} \int \sum_{k=1}^{|f(x_1)|} F(A, k, \mathbf{x}) d\bar{\mu}(\mathbf{x}), \tag{12}$$

where $F(A, k, \mathbf{x}) := \mathbf{1}_{\{f^*(\mathbf{x})\in T^{-k}A\}}$.

Proof: Stationarity of ρ was discussed in [14] for the injective f^*. Using the same trick, let us prove that ρ is a stationary measure in the general case. First of all, $\rho(\mathbb{W}) = 1$, whereas the countable additivity follows by the dominated convergence theorem. As for stationarity, we have

$$F(T^{-1}A, k, \mathbf{x}) = F(A, k+1, \mathbf{x}),$$
$$F(A, |f(x_1)| + 1, \mathbf{x}) = F(A, 1, T\mathbf{x})$$

by (10) so

$$\rho(T^{-1}A) = \frac{1}{L}\int \sum_{k=2}^{|f(x_1)|} F(A, k, \mathbf{x}) d\bar{\mu}(\mathbf{x}) + \frac{1}{L}\int F(A, 1, T\mathbf{x}) d\bar{\mu}(\mathbf{x}) = \rho(A)$$

in view of $\bar{\mu} \circ T^{-1} = \bar{\mu}$. □

Observe that $\rho(A) \geq L^{-1}\bar{\mu}(f^{*-1}A)$. Thus the corollaries of Theorem 1 and Lemma 2 are as follows:

(i) The process $(Y_i)_{i\in\mathbb{Z}}$ is AMS for an AMS $(X_i)_{i\in\mathbb{Z}}$ whenever $0 < L < \infty$.

(ii) In the AMS case, we have

$$\rho \gg \bar{\nu}, \bar{\mu} \circ f^{*-1} \gg \nu = \mu \circ f^{*-1}. \tag{13}$$

It is not obvious, however, that $\bar{\nu} = \rho$. Whereas $\bar{\nu}$ and ν take the same values on the T-invariant algebra, ρ may differ as it is indicated by the proof of the following theorem.

Theorem 2 *Let $|f(x)| > 0$ for all $x \in \mathbb{X}$. We have equality $\rho = \bar{\nu}$ if convergence*

$$\lim_{n\to\infty} S_+(\cdot, n)/n = L \in (0, \infty) \tag{14}$$

holds μ-almost everywhere for (11).

Proof: Consider a $w \in \mathbb{Y}^*$. For $m \geq |w|$ and $u \in \mathbb{X}^m$, we have $|f^*(u)| \geq |f(u_1)| + |w| - 1$. Hence

$$\begin{aligned}
\rho([w]) &= L^{-1} \sum_{u\in\mathbb{X}^m} \sum_{k=1}^{|f(u_1)|} \mathbf{1}_{\left\{\{f^*(u)\}_{k:k+|w|-1} = w\right\}} \bar{\mu}([u]) \\
&= L^{-1} \int \left[\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{S_+(\mathbf{x},n)} \mathbf{1}_{\left\{\{f^*(\mathbf{x})\}_{k:k+|w|-1} = w\right\}} \right] d\mu(\mathbf{x}) \\
&= \int \left[\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}_{\left\{\{f^*(\mathbf{x})\}_{k:k+|w|-1} = w\right\}} \right] d\mu(\mathbf{x}) \\
&= \int \left[\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}_{\left\{y_{k:k+|w|-1} = w\right\}} \right] d\nu(\mathbf{y}) = \bar{\nu}([w])
\end{aligned}$$

according to Lemma 1 and (14). Hence by the Kolmogorov process theorem and the π-λ theorem [4, Sections 2 and 36], $\rho = \bar{\nu}$ over their whole domain. □

Condition (14) is satisfied not only for an ergodic measure μ. Rewriting the expected expansion rate with the help of Lemma 1, we get

$$L = \sum_{x \in \mathbb{X}} |f(x)| \, \bar{\mu}([x]) = \sum_{x \in \mathbb{X}} |f(x)| \int \left[\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{x_i = x\}} \right] d\mu(\mathbf{x})$$
$$= \int \left[\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |f(x_i)| \right] d\mu(\mathbf{x}).$$

Thus (14) holds simply if $\lim_n S_+(\cdot, n)/n$ is a constant in the range $(0, \infty)$ μ-almost everywhere.

Example 1 *In particular, (14) holds for process (4) if one takes* $f(x) = g(k)z$ *for* $x = (k, z)$, $k \in \mathbb{N}$, $z \in \{0, 1\}$, $|g(k)| = O(\log k)$. *For example, we may take* $g(k)$ *to be one of several prefix codes considered by Elias [10].*

As accomplished in the following lemma, (14) implies also $\overline{\bar{\mu} \circ f^{*-1}} = \bar{\nu} := \overline{\mu \circ f^{*-1}}$ and consecutively $\bar{\nu} \gg \bar{\mu} \circ f^{*-1}$. The latest relation follows also from (13) and Theorem 2.

Lemma 3 *If (14) holds μ-almost everywhere then*

$$\bar{\tau} = \bar{\tau}' \implies \overline{\tau \circ f^{*-1}} = \overline{\tau' \circ f^{*-1}}$$

for any AMS measures $\tau, \tau' \ll \bar{\mu}$ *on* $\mathcal{U}$.

Proof: Introduce signed measures $\gamma := \bar{\tau} - \bar{\tau}'$ and $\varsigma := \gamma \circ f^{*-1} = \bar{\tau} \circ f^{*-1} - \bar{\tau}' \circ f^{*-1}$. If $\bar{\gamma} = 0$ then for any $A \in \mathcal{U}$,

$$\int \left[\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_{\{T^i \mathbf{x} \in A\}} \right] d\gamma(\mathbf{x}) = 0.$$

Consider

$$A = \left\{ \mathbf{x} \in \mathbb{U} : \{f^*(\mathbf{x})\}_{k:k+|w|-1} = w, 1 \le k \le |f(x_1)| \right\}.$$

We obtain

$$\bar{\varsigma}([w]) = \int \left[\lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}_{\left\{ y_{k:k+|w|-1} = w \right\}} \right] d\varsigma(\mathbf{y})$$
$$= \int \left[\lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}_{\left\{ \{f^*(\mathbf{x})\}_{k:k+|w|-1} = w \right\}} \right] d\gamma(\mathbf{x})$$
$$= \int \left[\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{\left\{ \mathbf{1}_{\{T^i \mathbf{x} \in A\}} \right\}} \right] L^{-1} d\gamma(\mathbf{x}) = 0,$$

since variable $\lim_n S_+(\cdot, n)/n$ is $\mathcal{I}_{\mathbb{U}}$-measurable and thus (14) holds also $\bar{\mu}$-almost everywhere. Hence by the Kolmogorov process theorem and the π-λ theorem, $\bar{\varsigma}(B) = 0$ for all $B \in \mathcal{W}$. $\square$

Formula (12) can be also given an interpretation in terms of random variables. To avoid elaborate notation, let us assume that the distribution $\mu = P((X_i)_{i\in\mathbb{Z}} \in \cdot)$ is already stationary, i.e., $\bar{\mu} = \mu$. Then $\rho = P((\tilde{Y}_i)_{i\in\mathbb{Z}} \in \cdot)$ happens to be the distribution of the process

$$(\tilde{Y}_i)_{i\in\mathbb{Z}} = T^N f^*((\tilde{X}_i)_{i\in\mathbb{Z}}),$$

where the process $(\tilde{X}_i)_{i\in\mathbb{N}}$ and the random shift N satisfy:

$$\begin{aligned} P(\tilde{X}_1 = x_1) &= \frac{P(X_1 = x_1)\,|f(x_1)|}{L}, \\ P((\tilde{X}_i)_{i\in\mathbb{Z}} \in A|\tilde{X}_1 = x_1) &= P((X_i)_{i\in\mathbb{Z}} \in A|\tilde{X}_1 = x_1), \\ P(N = n|\tilde{X}_1 = x_1) &= \frac{\mathbf{1}_{\{0\le n\le |f(x_1)|-1\}}}{|f(x_1)|}, \end{aligned}$$

whereas N is conditionally independent from $(X_i)_{i\in\mathbb{Z}}$ given $\tilde{X}_1$. Alas, processes $(\tilde{X}_i)_{i\in\mathbb{N}}$ and $(X_i)_{i\in\mathbb{N}}$ may be identified only in the case of a fixed length mapping $f : \mathbb{X} \to \mathbb{Y}^K$.

4 Measures of shrunk processes

Assume that $f^* : \mathbb{U} \to \mathbb{W}$ is a bijection. Then we can define

$$(X_i)_{i\in\mathbb{Z}} := {f^*}^{-1}((Y_i)_{i\in\mathbb{Z}})$$

and construct its measure $\mu = \nu \circ f^*$ starting with an arbitrary distribution $\nu = P((Y_i)_{i\in\mathbb{Z}} \in \cdot)$.

An example of bijection f^* may be given thus. Let us recall that a set of strings $\mathcal{L} \subset \mathbb{Y}^*$ is called a *complete fix-free* set if

(i) it is *complete*, i.e., satisfies Kraft equality

$$\sum_{w\in\mathcal{L}} |\mathbb{Y}|^{-|w|} = 1,$$

where $|\mathbb{Y}|$ is the cardinality of $\mathbb{Y}$,

(ii) it is both *prefix-* and *suffix-free*, i.e., for any $w, w' \in \mathcal{L}$ there is no such $v \in \mathbb{Y}^+$ that $w = w'v$ (the condition for a *prefix-free* $\mathcal{L}$) or $w = vw'$ (the condition for a *suffix-free* $\mathcal{L}$).

Example 2 *The sets* $\mathbb{Y}^n$ *are some particular complete fix-free sets but there exist less trivial instances, e.g.*

$$\{01, 000, 100, 110, 111, 0010, 0011, 1010, 1011\}$$

for an alphabet $\mathbb{Y} = \{0, 1\}$ *[13, 1].*

Definition 1 *Function* $f : \mathbb{X} \to \mathbb{Y}^*$ *will be called* (complete) prefix/suffix/fix-free *if* f *is an injection and the image* $\mathcal{L} = f(\mathbb{X})$ *is respectively a (complete) prefix/suffix/fix-free set.*

We came across the following proposition, which seemingly has not been observed previously in its entirety [3]:

Theorem 3 *Consider a complete fix-free f with a finite preimage $\mathbb{X}$. Then:*

(i) $f^* : \mathbb{U} \to \mathbb{W}$ *is a bijection of doubly infinite sequences,*

(ii) the measure $\mu = \nu \circ f^*$ *is stationary if so is* ν.

The statement (i) may be false for some processes and infinite prefix-free image $f(\mathbb{X})$. For instance, for $f(\mathbb{X}) = \{0^{n-1}1 : n \in \mathbb{N}\}$ and $\mathbf{y} = (y_i)_{i\in\mathbb{Z}}$, where $y_i = 0$. We do not know any infinite set of strings that would be complete fix-free.

Proof: Let $\mathcal{L} = f(\mathbb{X})$.

(i) Clearly $|w| \geq 1$ for $w \in \mathcal{L}$ if f is a bijection. Thus $f^*(\mathbf{x})$ is doubly infinite for a doubly infinite $\mathbf{x}$. On the other hand, given $\mathbf{y} = f^*(\mathbf{x})$ and $\mathcal{L}$, we can reconstruct $\mathbf{x}$ by parsing infinite sequences $y_{-\infty:0}$ and $y_{1:\infty}$ in the opposite directions. For an arbitrary $\mathbf{y} \in \mathbb{W}$, the same parsing can be performed as well and is guaranteed never to stop by the following reasoning.

On the contrary, assume that there is an infinite sequence $y_{k:\infty}$ (the mirror-like argumentation applies for $y_{-\infty:k}$) such that no $w \in \mathcal{L}$ is a prefix of $y_{k:\infty}$. Let v be a prefix of $y_{k:\infty}$ that is longer than any $w \in \mathcal{L}$. Set $\{v\} \cup \mathcal{L}$ is prefix-free so by Kraft inequality $\sum_{w\in L} |\mathbb{Y}|^{-|w|} \leq 1 - |\mathbb{Y}|^{-|v|} < 1$. We arrived at a contradiction so the assumption was false.

(ii) By the Kolmogorov process theorem and the π-λ theorem, stationarity of μ is equivalent to the set of equalities

$$\sum_{v\in\mathcal{L}} \nu([wv]) = \nu([w]) = \sum_{v\in\mathcal{L}} \nu([vw]), \quad w \in \mathcal{L}^*. \tag{15}$$

On the other hand, stationarity of ν is equivalent to

$$\sum_{a\in\mathbb{Y}} \nu([wa]) = \nu([w]) = \sum_{a\in\mathbb{Y}} \nu([aw]), \quad w \in \mathbb{Y}^*. \tag{16}$$

A useful fact to derive (15) from (16) is following: Let $l(\mathcal{M}) \geq 1$ denote the length of the longest string in set $\mathcal{M}$. Any finite complete prefix-free set $\mathcal{M} \subset \mathbb{Y}^*$ may be decomposed as $\mathcal{M} = \mathcal{M}_r \cup (\mathcal{M}_p \times \mathbb{Y})$, where $l(\mathcal{M}) = l(\mathcal{M}_p \times \mathbb{Y}) > l(\mathcal{M}_r)$ and $\mathcal{M}_m = \mathcal{M}_r \cup \mathcal{M}_p$ is a complete prefix-free set. The decomposition may be demonstrated by contradiction with Kraft inequality applied to $\mathcal{M}_m$.

Hence and from (16), it follows that for any complete prefix-free $\mathcal{M}$, $l(\mathcal{M}) \geq 1$, there exists a complete prefix-free $\mathcal{M}_m$ such that $l(\mathcal{M}) = l(\mathcal{M}_m) - 1$ and

$$\sum_{v\in\mathcal{M}} \nu([wv]) = \sum_{v\in\mathcal{M}_m} \nu([wv]).$$

Using this, the left equality in (15) may be proved by induction on $l(\mathcal{M})$ starting with $\mathcal{M}_m = \{\lambda\}$. The proof of the right equality is mirrorlike. □

As a corollary, under the hypothesis of Theorem 3, the process $(X_i)_{i\in\mathbb{Z}}$ is AMS for an AMS $(Y_i)_{i\in\mathbb{Z}}$. Indeed, we have $\mu = \nu \circ f^* \ll \bar{\nu} \circ f^*$ for the stationary measure $\bar{\nu} \circ f^*$.

5 Shift invariance

Let us recall again that $\bar{\mu}(A) = \mu(A)$ for $A \in \mathcal{I}_{\mathbb{U}}$. Analogical equality holds for $\nu = \mu \circ f^{*-1}$ and $\bar{\nu}$. We will demonstrate that the ergodic properties of measures μ, $\bar{\mu}$, ν, and $\bar{\nu}$ may be further related in some cases. Some apparent technical difficulty is that the set $f^*(\mathbb{U})$ usually does not belong to the strictly T-invariant algebra $\mathcal{I}_{\mathbb{W}} := \{A \in \mathcal{W} : T^{-1}A = A\}$. However, this can be overcome easily given certain care.

Lemma 4 *For an injection f^*, consider the T- and T^*-pseudo-invariant algebras*

$$\mathcal{Q}_{\mathbb{W}} := \{A \in \mathcal{W} : A = B \cap f^*(\mathbb{U}),\, T^{-1}B = B\},$$
$$\mathcal{Q}^*_{\mathbb{W}} := \left\{A \in \mathcal{W} : A = B \cap f^*(\mathbb{U}),\, T^{*-1}B = B\right\},$$

where T^* is defined by (9). We have

$$\mathcal{Q}_{\mathbb{W}} \subset \mathcal{Q}^*_{\mathbb{W}} = f^*(\mathcal{I}_{\mathbb{U}}).$$

Proof: The right equality is obvious. As for the left relation, observe that $\bigcup_{i\in\mathbb{Z}} T^iB \supset \bigcup_{i\in\mathbb{Z}}(T^*)^iB \supset B$. If $T^{-1}B = B$ then $\bigcup_{i\in\mathbb{Z}} T^iB = B$. Hence $B \cap f^*(\mathbb{U}) \in \mathcal{Q}^*_{\mathbb{W}}$ since formula $\bigcup_{i\in\mathbb{Z}}(T^*)^iB$ defines a T^*-invariant set. □

Definition 2 *We will say that $f^* : \mathbb{U} \to \mathbb{W}$ is a* synchronizable *injection if f^* is an injection and moreover $T^if^*(\mathbf{x}) = f^*(\mathbf{x}')$ for an $i \in \mathbb{Z}$ implies $T^j\mathbf{x} = \mathbf{x}'$ for a $j \in \mathbb{Z}$.*

Example 3 *In particular, f^* is a synchronizable injection for a comma separated code, $f(x) = g(x)c$, where $x \in \mathbb{X}$, $c \in \mathbb{Y}$, and g is an injection $\mathbb{X} \to (\mathbb{Y} \setminus \{c\})^*$. This f is also prefix-free.*

Obviously, injection f^* is not synchronizable for a complete fix-free f with $|f(x)| > 1$ for some $x \in \mathbb{X}$. For other concepts of synchronization, viz. [19, 5, 2].

Theorem 4 *For a synchronizable injection f^*,*

$$\mathcal{Q}_{\mathbb{W}} = \mathcal{Q}^*_{\mathbb{W}}.$$

Proof: By Lemma 4, $\mathcal{Q}_{\mathbb{W}} \subset \mathcal{Q}^*_{\mathbb{W}}$. Thus it suffices to show that $\mathcal{Q}^*_{\mathbb{W}} \subset \mathcal{Q}_{\mathbb{W}}$ or, equivalent, that $\mathcal{I}_{\mathbb{U}} \subset f^{*-1}(\mathcal{Q}_{\mathbb{W}})$. Now we will demonstrate the latter. Consider an $A \in \mathcal{I}_{\mathbb{U}}$ and construct set $E = f^*(\mathbb{U}) \cap \bigcup_{i\in\mathbb{Z}} T^if^*(A) \in \mathcal{Q}_{\mathbb{W}}$. Since f^* is synchronizable and A is T-invariant then $f^{*-1}(E) = A$. □

Theorem 5 *Consider a synchronizable injection f^*. For each $E \in \mathcal{I}_{\mathbb{W}}$ there exists such an $A \in \mathcal{I}_{\mathbb{U}}$ and for each $A \in \mathcal{I}_{\mathbb{U}}$ there exists such an $E \in \mathcal{I}_{\mathbb{W}}$ that*

$$\bar{\nu}(E) = \nu(E) = \nu(E \cap f^*(\mathbb{U})) = \mu(A) = \bar{\mu}(A). \tag{17}$$

Proof: For $E \in \mathcal{I}_{\mathbb{W}}$ take $A = f^{*-1}(E)$. For $A \in \mathcal{I}_{\mathbb{U}}$ take $E = \bigcup_{i\in\mathbb{Z}} T^if^*(A)$. Then the equalities follow immediately from Theorem 4 and (7). □

As a corollary, for a synchronizable injection, either all measures μ, $\bar{\mu}$, ν, and $\bar{\nu}$ are ergodic or none of them exhibits this property. Some oddness of (17) is buried in the fact that $\bar{\nu}(E)$ does not necessarily equal $\bar{\nu}(E \cap f^*(\mathbb{U}))$. It is only the support of ν that is confined to $f^*(\mathbb{U})$ and $f^*(\mathbb{U})$ need not be T-invariant, as it has been pointed out.

Example 4 *Process (4) has a nonatomic shift invariant σ-field [7]. In the case of a comma separated code $f(x) = g(x)c$, the same property also holds for the expanded process (3) and its stationary mean.*

6 Finite energy

As defined by Shields [18], a measure μ has *finite energy* if the conditional probabilities of all cylinder sets are uniformly exponentially damped, i.e., if

$$\mu([vu]) \le Kc^{|u|}\mu([v]) \tag{18}$$

for all $v, u \in \mathbb{X}^*$ and certain constants $c < 1$ and $K > 0$. Obviously, there must be $c \ge |\mathbb{X}|^{-1}$. In the case of a finite alphabet $\mathbb{X}$, the finite energy property implies an almost sure $O(\log n)$ bound for the length of the longest nonoverlapping repeat in block $X_{1:n}$ [18], which is used in [8].

In this section we shall investigate conditions under which mappings f^* and f^{*-1} conserve the finite energy property. The case we are particularly interested in is the process (4) over an infinite alphabet $\mathbb{X}$ and its image (3) over a finite alphabet $\mathbb{Y}$. Because of the bound for the repeat length, one may suppose that the mapping f^* preserves a generalization of (18) under mild conditions. After a while, we have realized that it is rather f^{*-1} that is so nice in our case.

There are two simple general results.

Theorem 6 *If μ has finite energy then so does $\bar{\mu}$ if it exists.*

Proof: Assume (18) and observe

$$\begin{aligned}\bar{\mu}([vu]) = \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \sum_{s\in\mathbb{X}^i} \mu([svu]) &\le Kc^{|u|} \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \sum_{s\in\mathbb{X}^i} \mu([sv]) \\ &\le Kc^{|u|}\bar{\mu}([v]).\end{aligned}$$

□

Theorem 7 *If ν has finite energy and f^* is an injection then the measure $\mu = \nu \circ f^*$ also has finite energy.*

Proof: Let $\nu([zw]) \le Kc^{|w|}\nu([z])$. Observe that $f^*([zw]) = [f^*(z)f^*(w)]$. If f^* is an injection then $|f^*(u)| \ge |u|$. Hence

$$\begin{aligned}\mu([vu]) = \nu(f^*([vu])) = \nu([f^*(v)f^*(u)]) &\le Kc^{|f^*(u)|}\nu([f^*(v)]) \\ &\le Kc^{|u|}\nu([f^*(v)]) = Kc^{|u|}\mu([v]).\end{aligned}$$

□

The converse of Theorem 7 does not seem to be true in general. There are several sources of obstacles on the way to prove it in special cases. First of all, the preimage ${f^*}^{-1}([w])$ of a cylinder set $[w]$ is not necessarily a cylinder set.

Lemma 5 *For $w \in \mathbb{Y}^*$ define set*

$$\mathbf{U}_f(w) = \left\{ u \in \mathbb{X}^* : \begin{array}{l} \exists_{z\in\mathbb{Y}^*} f^*(u) = wz, \\ \forall_{s,t\in\mathbb{X}^*} [u = st \implies |f^*(s)| < |w|] \end{array} \right\}.$$

Set $\mathbf{U}_f(w)$ is prefix-free and ${f^*}^{-1}([w]) = \{[u] : u \in \mathbf{U}_f(w)\}$.

Proof: That $\mathbf{U}_f(w)$ is prefix-free follows directly from its definition. As for the second property, clearly $f^*([u]) \subset [w]$ for $u \in \mathbf{U}_f(w)$. It remains to show that for all $\mathbf{x} \in \mathbb{U}_f$ such that $f^*(\mathbf{x}) \in [w]$ there exists a $u \in \mathbf{U}_f(w)$ such that $\mathbf{x} \in [u]$. The suitable u is the shortest string $x_{1:k}$ such that $f^*(x_{1:k}) = wz$ for some $z \in \mathbb{Y}^*$. □

Consider now a prefix-free f. In this case there are fewer difficulties. For $w \in \mathbb{Y}^*$ define the prefix-free sets

$$\mathbf{C}_f(w) = \left\{ z \in \mathbb{Y}^* : \begin{array}{l} \exists_{u\in\mathbb{X}^*} f^*(u) = wz, \\ \forall_{u\in\mathbb{X}^*} [f^*(u) = ws \implies z \neq st] \end{array} \right\}.$$

These are sets of completions of w to the strings in $f(\mathbb{X}^*)$. For $\nu = \mu \circ {f^*}^{-1}$ and $z \in f^*(\mathbb{X}^*)$ we have equalities $\mathbf{C}_f(zw) = \mathbf{C}_f(w)$ and

$$\nu([zw]) = \sum_{s\in\mathbf{C}_f(w)} \nu([zws]).$$

Next, consider functions

$$M_f(c) := \sup_{w\in\mathbf{P}_f} \sum_{s\in\mathbf{C}_f(w)} c^{|s|},$$

$$N_{f,\mu}(c) := \sup_{w\in\mathbf{P}_f} \frac{\sum_{s\in\mathbf{C}_f(w)} c^{-|s|}\mu([{f^*}^{-1}(ws)])}{\sum_{s\in\mathbf{C}_f(w)} \mu([{f^*}^{-1}(ws)])},$$

where $\mathbf{P}_f = \{w \in \mathbb{Y}^* : \exists_{u\in\mathbb{X}}\exists_{z\in\mathbb{Y}^*} f(u) = wz\}$. It is straightforward that $M_f(c) \geq 1$ and $N_{f,\mu}(c) \geq 1$.

Theorem 8 *Assume that f is prefix-free and $M_f(c) < \infty$. If $N_{f,\mu}(c_2) < \infty$ and*

$$\mu([vu]) \leq Kc^{|f^*(u)|}\mu([v]) \tag{19}$$

for all $v, u \in \mathbb{X}^*$ and certain constants $c \leq c_2 < 1$ and $K > 0$ then $\nu = \mu \circ {f^*}^{-1}$ has finite energy.

Proof: For a prefix-free f, the cardinality of ${f^*}^{-1}(\{z\})$ is at most one. If both z and w belong to $f^*(\mathbb{X}^*)$ then ${f^*}^{-1}([zw]) = [{f^*}^{-1}(z){f^*}^{-1}(w)]$. Hence

$$\begin{aligned} \nu([zw]) = \mu({f^*}^{-1}([zw])) &= \mu([{f^*}^{-1}(z){f^*}^{-1}(w)]) \\ &\leq Kc^{|w|}\mu([{f^*}^{-1}(z)]) = Kc^{|w|}\nu([z]). \end{aligned}$$

Now assume that only $z \in f^*(\mathbb{X}^*)$. Then

$$\nu([zw]) = \sum_{s \in \mathbf{C}_f(w)} \nu([zws]) \le \sum_{s \in \mathbf{C}_f(w)} K c^{|ws|} \nu([z]) \le M_f(c) K c^{|w|} \nu([z]).$$

Eventually, consider a $z \notin f^*(\mathbb{X}^*)$. Let

$$\mathbf{R}(w, s) = \begin{cases} \{a \in \mathbb{Y}^* : w = sa\}, & w \in s\mathbb{Y}^*, \\ \{\lambda\}, & s \in w\mathbb{Y}^*, \\ \emptyset, & \text{else.} \end{cases}$$

Set $\mathbf{R}(w, s)$ has at most one element. Moreover, we have $c^{|a|} \le c^{|w|-|s|}$ for $a \in \mathbf{R}(w, s)$. Let $\bar{K} := \max\{K, M_f(c)K\}$. Then

$$\begin{aligned} \nu([zw]) &= \sum_{s \in \mathbf{C}_f(z)} \sum_{a \in \mathbf{R}(w,s)} \nu([zsa]) \le \sum_{s \in \mathbf{C}_f(z)} \sum_{a \in \mathbf{R}(w,s)} \bar{K} c^{|a|} \nu([zs]) \\ &\le \sum_{s \in \mathbf{C}_f(z)} \sum_{a \in \mathbf{R}(w,s)} \bar{K} c_2^{|a|} \nu([zs]) \le \bar{K} c_2^{|w|} \sum_{s \in \mathbf{C}_f(z)} c_2^{-|s|} \nu([zs]) \\ &= N_{f,\mu}(c_2) \bar{K} c_2^{|w|} \nu([z]). \end{aligned}$$

□

It is easy to check that inequalities $M_f(c) < \infty$, $N_{f,\mu}(c_2) < \infty$, and (19) hold if μ has finite energy and f is prefix-free with a finite image $f(\mathbb{X})$. Thus ν has finite energy in the latter case. Unfortunately quite a negative result can be established for an f that has an infinite image $f(\mathbb{X})$:

Example 5 *Consider $\mathbb{X} = \mathbb{N}$ and $\mathbb{Y} = \{0, 1\}$. Let f be the Elias prefix-free representation γ [10], i.e.,*

$$f(n) = b_1 0 b_2 0 ... b_n 1, \tag{20}$$

where $1b_1b_2...b_n$ is the binary expansion of n. If μ is a measure of an IID process, $\mu([vu]) = \mu([v])\mu([u])$, where

$$L n^{-\beta} \le \mu([n]) \le U n^{-\beta}, \quad \beta < 1, \quad 0 < L, U < c^2,$$

for $c = 2^{-\beta/2}$ then (19), $M_f(c) = \infty$, and $N_{f,\mu}(c_2) = \infty$ all $c_2 \in [c, 1)$ are satisfied.

Proof: We have $|f(n)| = 2\lceil \log_2 n \rceil$. Let $u = n_1 n_2 ... n_k$, $n_i \in \mathbb{N}$. Since $n_i^{-\beta} = c^{2 \log_2 n_i}$ for $\beta = -2 \log_2 c$, we obtain

$$\mu([vu]) \le \mu([v]) U^k \prod_{i=1}^{k} c^{2 \log_2 n_i} \le \mu([v]) \prod_{i=1}^{k} c^{2 \log_2 n_i + 2} \le c^{|f^*(u)|} \mu([v]).$$

Thus inequality (19) is true. Next we have

$$\mathbf{C}_f(w) = \begin{cases} (\mathbb{Y}0)^* \mathbb{Y} 1, & w \in (\mathbb{Y}0)^*, \\ (0\mathbb{Y})^* 1, & w \in (\mathbb{Y}0)^* \mathbb{Y}, \\ \lambda, & w \in f(\mathbb{X}). \end{cases}$$

Hence

$$M_f(c) = \max\left\{1, \sum_{k=1}^{\infty} 2^k c^{2k}\right\} = \infty,$$

since $c \ge 2^{-1/2}$. In the following, we ignore the little discrepancy between $\log_2 n$ and $\lceil \log_2 n \rceil$, which can be overcome easily by rescaling constants L and U. Thus,

$$\begin{aligned} N_{f,\mu}(c_2) \ge \sup_{w\in\mathbf{P}_f} \frac{\sum_{s\in\mathbf{C}_f(w)} c_2^{-|s|} L c^{|w|+|s|}}{\sum_{s\in\mathbf{C}_f(w)} U c^{|w|+|s|}} &= \frac{L}{U} \sup_{w\in\mathbf{P}_f} \frac{\sum_{s\in\mathbf{C}_f(w)} (c/c_2)^{|s|}}{\sum_{s\in\mathbf{C}_f(w)} c^{|s|}} \\ &= \frac{L}{U} \max\left\{1, \frac{\sum_{k=1}^{\infty} 2^k (c/c_2)^{2k}}{\sum_{k=1}^{\infty} 2^k c^{2k}}\right\}, \end{aligned}$$

which is infinite since $c/c_2 > 2^{-1/2}$ for $c_2 \in [c, 1)$. □

Code (20) is very easy to analyze because of the simple form of sets $\mathbf{C}_f(w)$.

7 Information measures

Denote the expectation of variable U as $\mathbf{E}\, U$. Recall the definition of entropy

$$H(U) = -\mathbf{E} \log P(U = \cdot)$$

and mutual information $I(U;V) = H(U) + H(V) - H(U,V)$ of discrete variables U and V. The last problem that we are interested in is how the block mutual information $I(X_{-n+1:0}; X_{1:n})$ of (4) relates to $I(\bar{Y}_{-n+1:0}; \bar{Y}_{1:n})$ of the process $(\bar{Y}_i)_{i\in\mathbb{Z}}$ distributed according to the stationary mean

$$P((\bar{Y}_i)_{i\in\mathbb{Z}} \in \cdot) = \bar{\nu}.$$

The case of an infinite alphabet $\mathbb{X}$ and a finite alphabet $\mathbb{Y}$ is hard to analyze so we will only make easier observations here.

As an elementary property, entropy and mutual information are invariants of bijective measurable mappings of discrete variables. In this article, such mappings have appeared in several cases. First of all, for a prefix-free f, we have $Y_{1:S_n} = f^*(X_{1:n})$ and $X_{1:n} = f^{*-1}(Y_{1:S_n})$, where $S_n = \sum_{i=1}^{n} |f(X_i)|$. Thus

$$H(Y_{1:S_n}) = H(X_{1:n}). \tag{21}$$

It is possible to extend the definition of entropy and mutual information in such way that, e.g., $I((X_i)_{i\le 0}; (X_i)_{i\ge 1}) = \lim_n I(X_{-n+1:0}; X_{1:n})$ [12, 9, 16]. The generalized entropy and mutual information are also invariants of bijective measurable mappings of random variables. Thus, if f is complete fix-free with a finite preimage $\mathbb{X}$ then we have

$$I((\bar{Y}_i)_{i\le 0}; (\bar{Y}_i)_{i\ge 1}) = I((Y_i)_{i\le 0}; (Y_i)_{i\ge 1}) = I((X_i)_{i\le 0}; (X_i)_{i\ge 1})$$

for a stationary $(X_i)_{i\in\mathbb{Z}}$ in view of Theorem 3.

In a simple case, we can also compare the entropies of stationary means for finite blocks. Assume that the distribution $\mu = P((X_i)_{i\in\mathbb{Z}} \in \cdot)$ is stationary and $f : \mathbb{X} \to \mathbb{Y}^K$ is a fixed length injection. Let $\nu = \mu \circ f^{*-1}$. According to Theorems 1 and 2, the stationary mean

$$\bar{\nu} = P((\bar{Y}_i)_{i\in\mathbb{Z}} \in \cdot) = \rho = P((\tilde{Y}_i)_{i\in\mathbb{Z}} \in \cdot)$$

is also the distribution of process

$$(\tilde{Y}_i)_{i\in\mathbb{Z}} = T^N f^*((X_i)_{i\in\mathbb{Z}}) \quad (22)$$

where $N : \Omega \to \{0, 1, ..., K-1\}$ is a uniformly distributed random variable probabilistically independent from $(X_i)_{i\in\mathbb{Z}}$.

If $\mathbb{X}$ is finite, it is immediate that

$$|H(\bar{Y}_{1:nK}) - H(X_{1:n})| = |H(\tilde{Y}_{1:nK}) - H(X_{1:n})| \le C \quad (23)$$

for a certain constant C. Indeed, by (22), we have

$$\begin{aligned} H(\tilde{Y}_{1:nK}) &\le H(X_{1:n+1}, N) \le H(X_{1:n}) + H(X_{n+1}) + H(N) \\ &\le H(X_{1:n}) + \log \operatorname{card} \mathbb{X} + \log K, \end{aligned}$$

where $\operatorname{card} \mathbb{X}$ is the cardinality of $\mathbb{X}$. On the other hand, since f is an injection, we have $X_{2:n+1} = g(\bar{Y}_{1:(n+1)K}, N)$ for a certain function g. Hence

$$\begin{aligned} H(X_{1:n}) &= H(X_{2:n+1}) \\ &\le H(\tilde{Y}_{1:(n+2)K}, N) \le H(\tilde{Y}_{1:n}) + H(\tilde{Y}_{nK+1:(n+1)K}) + H(N) \\ &\le H(\tilde{Y}_{1:nK}) + H(\tilde{Y}_{1:K}) + \log K \le H(\tilde{Y}_{1:nK}) + 2\log \operatorname{card} \mathbb{X} + 2\log K. \end{aligned}$$

Thus, we obtain (23) for $C = 2\log \operatorname{card} \mathbb{X} + 2\log K$.

Acknowledgement

The author thanks Peter Harremoës for several remarks.

References

[1] R. Ahlswede, B. Balkenhol, and L. H. Khachatrian. Some properties of fix-free codes. In *Proceedings of the First INTAS International Seminar on Coding Theory and Combinatorics, 1996, Thahkadzor, Armenia*, pages 20–33. 1996.

[2] R. Ahlswede, B. Balkenhol, C. Deppe, H. Mashurian, and T. Partner. *T*-shift synchronization codes. *Electr. Not. Disc. Math.*, 21:119–123, 2005.

[3] D. Bajic, C. Stefanovic, and D. Vukobratovic. Search process and probabilistic bifix approach. In *Proceedings of the International Symposium on Information Theory, 2005*, pages 19–22. 2005.

[4] P. Billingsley. *Probability and Measure*. Wiley, 1979.

[5] R. Capocelli, A. D. Santis, L. Gargano, and U. Vaccaro. The construction of statistically synchronizable codes. *IEEE Trans. Inform. Theor.*, 38:407–414, 1992.

[6] G. Cariolaro and G. Pierobon. Stationary symbol sequences from variable-length word sequences. *IEEE Trans. Inform. Theor.*, 23:243–253, 1977.

[7] Ł. Dębowski. Ergodic decomposition of excess entropy and conditional mutual information. IPI PAN Reports, Nr 993. Institute of Computer Science, Polish Academy of Sciences, 2006.

[8] Ł. Dębowski. On vocabulary size of grammar-based codes. In *2007 IEEE International Symposium on Information Theory. Nice, France, July 25-29*, pages 91–95. 2007.

[9] R. L. Dobrushin. A general formulation of the fundamental Shannon theorems in information theory. *Usp. Matem. Nauk*, 14(6):3–104, 1959. In Russian.

[10] P. Elias. Universal codeword sets and representations for the integers. *IEEE Trans. Inform. Theor.*, 21:194–203, 1975.

[11] R. Fontana, R. Gray, and J. Kieffer. Asymptotically mean stationary channels. *IEEE Trans. Inform. Theor.*, 27:308–316, 1981.

[12] I. M. Gelfand, A. N. Kolmogorov, and A. M. Yaglom. Towards the general definition of the amount of information. *Dokl. Akad. Nauk SSSR*, 111: 745–748, 1956. In Russian.

[13] D. Gillman and R. L. Rivest. Complete variable-length "fix-free" codes. *Desig. Cod. Cryptogr.*, 5:109–114, 1995.

[14] R. M. Gray and J. C. Kieffer. Asymptotically mean stationary measures. *Ann. Probab.*, 8:962–973, 1980.

[15] H. L. Hurd. Stationarizing properties of random shifts. *SIAM J. Appl. Math.*, 26:203–212, 1974.

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

[17] O. W. Rechard. Invariant measures for many-one transformations. *Duke Math. J.*, 23:477–488, 1956.

[18] P. C. Shields. String matching bounds via coding. *Ann. Probab.*, 25:329–336, 1997.

[19] J. Stiffler. *Theory of Synchronous Communications*. Prentice Hall, 1971.