Indag. Mathem., N.S., 7 (4), 503–519 December 16, 1996

The finitary coding of two Bernoulli schemes with unequal entropies has finite expectation

by J. Serafin

Faculty of Mathematics and Informatics, Technical University Delft, Mekelweg 4, 2628 CD Delft, the Netherlands (until January 1, 1996)
Institute of Mathematics, Technical University of Wrocław, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland (after January 1, 1996)

Communicated by Prof. M.S. Keane at the meeting of November 27, 1995

ABSTRACT

It is well known that for any two Bernoulli schemes with a finite number of states and unequal entropies, there exists a finitary homomorphism from the scheme with the larger entropy to the one with smaller entropy. We prove that the average number of coordinates in the larger entropy scheme needed to determine one coordinate in the image point is *finite*.

1. INTRODUCTION

It was shown by Ornstein ([2]) that if two Bernoulli schemes of unequal entropies are given, then the scheme of smaller entropy is a measurable factor of the one with the larger entropy. It seems natural to require that the factor map satisfy continuity property of some kind. We consider codes which are finitary. A coding from one sequence space to another is called finitary (or almost continuous) if after removing sets of measure zero from both spaces, the coding is continuous on the remaining sets. A more descriptive way of thinking of a finitary code is the following: in order to determine a coordinate in the image process, one needs to examine a finite (but perhaps very large) number of past and future coordinates in the original process.

In 1977 Keane and Smorodinsky ([1]) constructed a finitary coding between two Bernoulli schemes with unequal entropies. The last statement in their article is that the coding has finite expectation. It was communicated to the author by M. Keane that the proof of that statement had been missing. This paper supplies the proof.

We begin by recalling some of the definitions and notations as in [1]. For the convenience of the reader we sketch the basic idea of the finitary coding construction. The interested reader should consult [1] for the details, as well as the proofs of some of the statements which shall be made in this article.
Let

$$A = \{1, 2, \dots, a\}$$

be a finite alphabet, $a \geq 2$, and let $p = (p_1, p_2, \dots, p_a)$ be a strictly positive probability vector. The *Bernoulli scheme* $\mathcal{B}(p) = (X, \mathcal{A}, \mu, T)$ is a dynamical system defined as follows

$$X = A^{\mathbb{Z}}, \qquad \mathcal{A} = \text{product } \sigma\text{-algebra on } X, \qquad \mu = p^{\mathbb{Z}},$$

$$T = \text{left shift on } X.$$

If $F = f_1 \dots f_n$ is a finite concatenation of symbols from A, we denote by $\mu(F)$ the measure of the cylinder set determined by F. The entropy of T is given by the formula

$$h(T) = h = -\sum_{i=1}^{a} p_i \log p_i.$$

Denote by $\mathcal{B}(\bar{p}) = (\bar{X}, \bar{\mathcal{A}}, \bar{\mu}, \bar{T})$ a second Bernoulli scheme with alphabet $\bar{A}$, probability vector $\bar{p}$ and entropy $\bar{h}$.

A *homomorphism* (factor map) ϕ from $\mathcal{B}(p)$ to $\mathcal{B}(\bar{p})$ is a measurable map ϕ from a subset of measure one of X to $\bar{X}$ such that $\bar{\mu} = \mu\phi^{-1}$ and $\phi \circ T = \bar{T} \circ \phi$. The homomorphism ϕ is called *finitary* if for almost every $x \in X$ there exist integers $q = q(x)$, $r = r(x)$, $q \leq r$, such that if $y \in X$ and $[x_q, \dots, x_r] = [y_q, \dots, y_r]$ and if $\phi(y)$ is defined, then $(\phi(x))_0 = (\phi(y))_0$. In other words, in order to determine the 0 coordinate of the image of a point x, we only need to examine a finite number (depending upon x) of coordinates of x. The main result in [1] is the following.

Theorem 1. *If $\bar{h} < h$, then there exists a finitary homomorphism from $\mathcal{B}(p)$ to $\mathcal{B}(\bar{p})$.*

From now on we shall assume that ϕ is a finitary homomorphism from $\mathcal{B}(p)$ to $\mathcal{B}(\bar{p})$, constructed as in [1].

A natural way of defining the expected coding time for ϕ is the following. For $x \in X$, let $q = q(x)$ and $r = r(x)$ be minimal positive integers, determined by the finitary coding procedure, with the property that $(\phi(x))_0$ is determined by the cylinder $[x_{-q}, \dots, x_r]$.

Definition 1. The *coding time* for $x \in X$ is $C(x) = r(x) + q(x) + 1$.

The purpose of this paper is to prove that on average, $C(x)$ is finite.

Theorem 2. *If $\bar{h} < h$, ϕ is a finitary homomorphism as in* [1] *and C is the coding time function, then $\mathbf{E}_{\mu}(C) < \infty$.*

The rest of this paper is devoted to proving the above result.

3. THE NECESSARY TOOLS

In this chapter we recall some of the ideas leading to the construction of a finitary coding from $\mathcal{B}(p)$ to $\mathcal{B}(\bar{p})$. We also express the function C in a form that is most convenient for our purpose. Finally, we use some of the results from [1] to bound the expectation of C by a function whose expectation is more easily computable than that of C.

3.1. Markers

One of the crucial steps in constructing a desired coding in [1] is the following. For a given positive integer k_0, there exists an integer $k \geq k_0$ and a Bernoulli scheme $\mathcal{B}(\tilde{p})$ with entropy $\tilde{h}$ such that

(i) $\bar{h} < \tilde{h} < h$

(ii) $p_1^{k-1} p_2 = \tilde{p}_1^{k-1} \tilde{p}_2$

(iii) there exists a finitary homomorphism ψ from $\mathcal{B}(\tilde{p})$ to $\mathcal{B}(\bar{p})$, obtained by identifying certain two symbols in $\mathcal{B}(\tilde{p})$ and copying the remaining symbols; the coding time for ψ is equal to 1.

Consequently, constructing a finitary map from $\mathcal{B}(p)$ to $\mathcal{B}(\bar{p})$ boils down to finding a finitary map $\tilde{\phi}$ from $\mathcal{B}(p)$ to $\mathcal{B}(\tilde{p})$, with $\mathcal{B}(\tilde{p})$ as above. It is also clear that if $\tilde{\phi}$ has finite expectation, so does $\phi = \tilde{\phi} \circ \psi$.

With the above remark in mind, we may limit our attention to the case that the two given Bernoulli schemes $\mathcal{B}(p)$ and $\mathcal{B}(\bar{p})$ enjoy the property

$$(1) \qquad p_1^{k-1} p_2 = \bar{p}_1^{k-1} \bar{p}_2$$

where k may be chosen as large as we wish. It follows that there are two blocks of the same length k, one in each scheme, with the property that their occurrences in both schemes coincide.

Definition 2. A *marker* (for either scheme) is the block

$$M = 1^{k-1} 2 = \underbrace{1 \ldots 1}_{(k-1)-times}\ 2.$$

Put $\hat{X} = \{\hat{1}, \hat{2}\}^{\mathbb{Z}}$ and define the *marker process* $(\hat{X}, \hat{T}, \hat{\mu})$ for either scheme as follows. For $x \in X$ (resp. $\bar{x} \in \bar{X}$) replace all the markers M occurring in x (resp. $\bar{x}$) by $\hat{1}^k$, and all other symbols replace by $\hat{2}$. Let $\hat{T}$ be the left shift on $\hat{X}$. The replacement operation commutes with the shifts T $(\bar{T})$ and $\hat{T}$ and sends μ $(\bar{\mu})$ to a measure $\hat{\mu}$ on $\hat{X}$. Observe that almost all $\hat{x} \in \hat{X}$ can be written in a form

$$\hat{x} = \ldots \hat{2}^{l_{-1}} \hat{1}^{u_{-1} k} \hat{2}^{l_0} \hat{1}^{u_1 k} \hat{2}^{l_1} \ldots$$

where $\{u_n\}$ and $\{l_n\}$ are sequences of positive integers and $\hat{x}_0$ is covered by the block $\hat{2}^{l_0}\hat{1}^{u_1 k}$.

3.2. Fillers

Consider the projection $\Pi : X \to \hat{X}$ from $\mathcal{B}(p)$ to the marker process. For each $\hat{x} \in \hat{X}$, the *fiber* of X above $\hat{x}$ is defined as

$$X(\hat{x}) = \Pi^{-1}(\hat{x}).$$

Let A_0^l ($\bar{A}_0^l$) denote the set of sequences of length l over A ($\bar{A}$) which contain no markers. Elements of A_0^l are called *fillers* of length l. We identify the fiber $X(\hat{x})$ above $\hat{x}$ by setting

$$X(\hat{x}) = \Pi_{n\in\mathbb{Z}} A_0^{l_n}$$

for a.e. $\hat{x} \in \hat{X}$. For $x \in X(\hat{x})$ denote by $F_n(x) \in A_0^{l_n}$, $n \in \mathbb{Z}$, the sequence of fillers determining x. Define, for $l \geq 1$, and for $F \in A_0^l$,

$$\mu_0(F) = \frac{\mu(F)}{\sum_{G\in A_0^l} \mu(G)} = \frac{\mu(F)}{\mu(A_0^l)};$$

and denote by $\mu_{\hat{x}}$ a normalized measure on almost every fiber $X(\hat{x})$, obtained by taking the product measure of the μ_0 on the $A_0^{l_n}$, for all $n \in \mathbb{Z}$. We call $\mu_{\hat{x}}$ the *filler measure*. By [1, Lemma 5] we have a conditional version of the Shannon–McMillan–Breiman theorem for the filler measure:

$$\lim_{n\to\infty} \frac{-1}{\sum_{i=-n}^{n} l_i(x)} \log \prod_{i=-n}^{n} \mu_0(F_i(x)) = \frac{1}{1-k\eta}(h-e) =: g \tag{2}$$

where $\eta = \mu(M)$ is the marker measure and e is the entropy of the marker process (e depends upon k and can be made as small as necessary). We call g the *filler entropy* for $\mathcal{B}(p)$. Let $\bar{g} = (1/(1-k\eta))(\bar{h}-e)$ be the filler entropy for the $\mathcal{B}(\bar{p})$ scheme.

3.3. Skeletons

Let r be a positive integer.

Definition 3. The *skeleton* $s_r(x)$ of rank r at x is a non-indexed sequence of markers and spaces between the markers of the form:

$$M^{u_{-m}} \underset{l_{-(m-1)}}{\underline{\qquad}} M^{u_{-(m-1)}} \cdots \underset{l_{-1}}{\underline{\qquad}} M^{u_{-1}} \underset{l_0}{\underline{\qquad}} M^{u_1} \underset{l_1}{\underline{\qquad}} \cdots M^{u_{n-1}} \underset{l_{n-1}}{\underline{\qquad}} M^{u_n}$$

where

(s1) $m = m_r(x)$, $n = n_r(x) \geq 1$
(s2) $l_i = l_i(x) \geq 1$, for $i = -(m-1), \ldots, n-1$
(s3) $u_{-m}, u_n \geq r > u_{-(m-1)}, \ldots, u_{n-1}$
(s4) x_0 appears inside a block separating $M^{u_{-1}}$ and M^{u_1}.

Clearly the rank one skeletons consist of two marker blocks separated by one filler block. The skeleton of rank r is obtained by looking to the left and to the right for the first appearance of M^r.

It should be obvious that if x and x' belong to the same fiber $X(\hat{x})$, then their skeletons $s_r(x)$ and $s_r(x')$ are the same, for all r. It follows that the skeletons $s_r(x)$ at x depend only upon the fiber which x belongs to, or in other words upon the marker process. Consequently, we write $s_r(\hat{x})$ for the skeleton of rank r at x, for all $x \in X(\hat{x})$ and $r \geq 1$. It is also clear that the indices m and n in the skeleton definition are in fact random variables depending upon r and $\hat{x}$, i.e. $m = m_r(\hat{x})$ and $n = n_r(\hat{x})$. Similarly $l_i = l_i(\hat{x})$ and $u_i = u_i(\hat{x})$. We define

$$(3) \qquad L_r := L_r(\hat{x}) = \sum_{i=-(m-1)}^{n-1} l_i, \qquad K_r := K_r(\hat{x}) = \sum_{i=-(m-1), i \neq 0}^{n-1} k \cdot u_i$$

the *filler length* of s_r and the *marker length* of s_r, respectively. Further, we define

$$b_r = b_r(\hat{x}) = m_r(\hat{x}) + n_r(\hat{x}) - 1,$$

the number of filler blocks in $s_r(\hat{x})$.

Definition 4. Let s_r be a skeleton with the filler lengths $l_{-(m_r-1)}, \ldots, l_{n_r-1}$. A *filler* F for s_r is an element of the *filler set* $\mathcal{F}(s_r) := \prod_{i=-(m_r-1)}^{n_r-1} A_0^{l_i}$.

Define, for $r \geq 2$, d_r and d_r' to be independent copies of a random variable, equal to the waiting time for the occurrence of M^r, given the occurrence of M^{r-1}. With d_r and d_r' as above, let

$$c_1 = l_0 + 2k$$

and

$$c_r = c_{r-1} + d_r + d_r'$$

for $r \geq 2$.

Remark 1. Observe that the joint distribution of the sum $L_r + K_r + 2 \cdot r \cdot k$ is the same as the distribution of c_r ($r \geq 1$).

It is important to note that $c_r = c_r(\hat{x})$ is the minimal number of coordinates that one has to examine in order to be able to locate the skeleton of rank r at $\hat{x}$. This is the reason why the random variables c_r, and therefore L_r and K_r will be very useful in estimating the expectation of the coding time C.

3.4. The coding

For the rest of the argument fix ε such that $(g - \bar{g})/3 > \varepsilon > 0$. We need to distinguish between fillers which are *good* and those which are *bad*.

Definition 5. Let s_r be a skeleton with filler length $L_r = \sum_{i=-(m_r-1)}^{n_r-1} l_i$. A filler $F \in \mathcal{F}(s_r)$ for $\mathcal{B}(p)$ is *bad* if

$$\mu_0(F) > 2^{-(g-\varepsilon)L_r}.$$

A filler $\bar{F} \in \bar{\mathcal{F}}(s_r)$ for $\mathcal{B}(\bar{p})$ is bad if $\bar{\mu}_0(\bar{F}) < 2^{-(\bar{g}+\varepsilon)L_r}$. Fillers which are not bad are called *good.*

We wish now to divide the elements of $\bar{\mathcal{F}}(s_r)$ into equivalence classes. If $\bar{F} \in \bar{\mathcal{F}}(s_r)$ is good then no other filler is equivalent to it. Suppose $\bar{F}$ is bad. It might happen that when we restrict $\bar{F}$ to a subskeleton s' of s_r, then this restriction is a good filler (for s'). The equivalence class of $\bar{F}$ consists now of all fillers $\bar{G}$ with the property that $\bar{G}$ is bad, $\bar{G}$ has the same collection of subskeletons on which its restrictions are good, and it agrees with $\bar{F}$ when restricted to each one of these subskeletons. We write $\bar{F} \sim \bar{G}$. Lemma 13 of [1] is telling us that there are no more than

$$(4) \qquad 2^{m_r + n_r - 1 + (\bar{g}+\varepsilon)L_r}$$

equivalence classes in $\bar{\mathcal{F}}$. The major step towards defining the coding is achieved by considering the *partial assignments.* For a skeleton s, a partial assignment P_s assigns each element $\bar{F} \in \bar{\mathcal{F}}(s)$ a subset $P_s(\bar{F})$ of $\mathcal{F}(s)$, in a way that

$$\bar{\mu}_0(\bar{F}) \leq \mu_0(P_s(\bar{F})).$$

This mapping is an example of a *society* (for a discussion of societies see [1]). The partial assignment is *good,* if it respects the equivalence classes of $\bar{F}(s)$, i.e. if

$$(5) \qquad \bar{F} \sim \bar{G} \Rightarrow P_s(\bar{F}) = P_s(\bar{G}).$$

Suppose $\bar{F} \sim \bar{G}$. By (5) each $F \in P_s(\bar{F})$ will be assigned for $\bar{F}$ or $\bar{G}$ or any other element of the equivalence class of $\bar{F}$. However, the Shannon–McMillan–Breiman theorem (2) implies that at some finite stage, $\bar{F}$ is a part of some much longer good filler $\bar{H} \in \bar{\mathcal{F}}(s_r)$ for a skeleton s_r of rank higher that the rank of s (s being a subskeleton of s_r). As $\bar{H}$ is good, no other element of $\bar{\mathcal{F}}(s_r)$ is equivalent to it, and each $F \in \mathcal{P}_{s_r}(\bar{H})$ is assigned to exactly one element of $\bar{\mathcal{F}}(s_r)$, this element being of course $\bar{H}$. Keane and Smorodinsky show in [1] how partial assignments can be unambiguously extended to global assignments, in such a way that if at some stage of the coding a filler $F \in \bar{F}(s)$ is uniquely assigned to a $\bar{F} \in \bar{\mathcal{F}}(s)$, then the assignments at the later stages, restricted to s, respect the $F \mapsto \bar{F}$ assignment. Finally, $\bar{F}$ is defined as the homomorphic image of F. The above procedure is shift-invariant and defines the required finitary homomorphism.

3.5. The distributions

In this section we compute the distributions of the random variables defined above. We begin by considering the u_1.

Distribution of u_1. Observe that $u_1 \geq 1$ and that the event $\{u_1 = t\}$ is the event that exactly t-many markers occur, given that at least one marker occurs. Consequently,

$$\mathbb{P}(u_1 = t) = \eta^{t-1}(1-\eta)$$

where $t \geq 1$ and $\eta = \mu(M) = p_1^{k-1} p_2$.

We note in passing that the random variables $\{u_i\}_{i \in \mathbb{Z}}$ are independent and identically distributed, distribution as above.

Distributions of m_r, n_r, b_r. Let $s \geq 1$. We then have

$$\{n_r = s\} = \{u_1 < r, \ldots, u_{s-1} < r, u_s \geq r\}.$$

For $r = 1$, $n_1 = m_1 = 1$ with probability one. For $r \geq 2$ we have

$$\mathbb{P}(n_r = s) = \mathbb{P}^{s-1}(u_1 < r) \cdot \mathbb{P}(u_s \geq r).$$

As the distribution of u_1 is given above, it is easy to check that n_r is geometrically distributed with parameter η^{r-1}, i.e.

$$(6) \qquad \mathbb{P}(n_r = s) = (1 - \eta^{r-1})^{s-1} \eta^{r-1}.$$

Note that if X and Y are two independent, geometrically distributed random variables, both with parameter q, then their sum has the following distribution: $\mathbb{P}(X + Y = t) = (t-1) q^2 (1-q)^{t-2}$, for $t \geq 2$. Since m_r and n_r are independent, we have

$$(7) \qquad \mathbb{P}(b_r = t) = \mathbb{P}(m_r + n_r - 1 = t) = t\eta^{2r-2}(1 - \eta^{r-1})^{t-1},$$

for $t \geq 1$.

Distribution of l_i, $i \neq 0$. It is not difficult to see that the random variables l_i, $i \neq 0$, are independent and identically distributed copies of a random variable $\mathcal{L}$, equal to the length of a run of $\hat{2}$'s which separate two runs of $\hat{1}$'s in the marker process. We wish to calculate the generating function of $\mathcal{L}$. It is convenient to start with the $\mathcal{L}'$, equal to the separation between any two markers. We have then $\mathbb{P}(\mathcal{L} = n) = \mathbb{P}(\mathcal{L}' = n)/(1 - \mathbb{P}(\mathcal{L}' = 0))$, for $n \geq 1$. The renewal sequence belonging to $\mathcal{L}' + k$ is the following:

$$q_0 = 1, q_1 = \cdots = q_{k-1} = 0. q_i = \eta,$$

for $i \geq k$. Indeed, if we just saw a marker ($q_0 = 1$) then we are not allowed to see any marker at time $1, \ldots, k-1$ (the reason is that the sequence $1^{k-1}2$ does not overlap itself). The renewal function $Q(s)$ is given by $Q(s) = \sum_{i \geq 0} q_i s^i = 1 + \eta(s^k/(1-s))$, and therefore the generating function $U(s)$ of $\mathcal{L}' + k$ satisfies:

$$U(s) = 1 - \frac{1}{Q(s)} = \frac{\eta s^k}{1 - s + \eta s^k}.$$

As a consequence we obtain the generating function of the $\mathcal{L}$ itself

$$(8) \qquad G(s) = G_{\mathcal{L}}(s) = \frac{s - \eta s^k}{1 - s + \eta s^k} \cdot \frac{\eta}{1 - \eta}.$$

One corollary of the above is that $\boldsymbol{E}(\mathcal{L}) = \boldsymbol{E}(l_1) = G'(1) = (1 - k\eta)/\eta(1 - \eta)$.

Distribution of l_0. Begin by observing that the distribution of l_0 differs from that of l_1, because we may no longer assume that a marker has been seen and we are waiting for another to occur. We must look now in both directions. However, if we suppose a marker appeared at coordinates $x_{-1}, \dots, x_{-k}$, then the waiting time to see a marker in the right-hand-side direction is $l_1 + 1$. By symmetry, we have to wait $l_{-1} + 1$ for a marker in the left-hand-side direction, if a marker occurred at $x_1, \dots, x_k$. Consequently, l_0 has the distribution equal to that of $l_{-1} + l_1 + 1$.

Remark 2. The filler length L_r of s_r is distributed as $N_{b_r+1} + 1$, where N_{b_r+1} is a sum of $(b_r + 1)$-many independent random variables, all with generating function $G = G_{\mathcal{L}}$.

Distribution of d_r. Let $r \geq 2$. The renewal sequence belonging to d_r is

$$\begin{aligned}
&q_0 = 1, q_1 = \cdots = q_{k-1} = 0 \\
&q_k = \eta, q_{k+1} = \cdots = q_{2k-1} = 0 \\
&\vdots \\
&q_{(r-1)k} = \eta^{r-1}, q_{(r-1)k+1} = \cdots = q_{rk-1} = 0 \\
&q_{rk} = \eta^r \quad \text{for } i \geq rk,
\end{aligned}$$

and the renewal function $Q_r(s) = (1 - s + (\eta s^k)^r s - (\eta s^k)^{r+1})/(1 - s)(1 - \eta s^k)$. The generating function $F_r(s)$ of d_r equals

$$F_r(s) = 1 - \frac{1}{Q_r(s)} = \eta s^k \frac{1 - s + (\eta s^k)^{r-1}(s - \eta s^k)}{1 - s + (\eta s^k)^r (s - \eta s^k)}.$$

We therefore have $E(d_r) = F_r'(1) = (1/\eta^r)$. Clearly, d_r' has the same distribution as d_r. It is convenient to think of d_r as waiting time for M^r, given the occurrence of M^{r-1}, to the right of $\hat{x}_0$, whereas d_r' might be considered the time spent waiting for M^r, given M^{r-1}, to the left of $\hat{x}_0$. We close this section with the following remark.

Remark 3. The d_{r+1} and d_{r+1}' are independent of L_r, for $r \geq 1$.

3.6. Estimating the expected coding time

Recall that $0 < 3\varepsilon < g - \bar{g}$.

The coding is performed in steps. At step r ($r \geq 1$) skeletons of rank r and their fillers are being examined. The coding procedure ensures that some of the fillers for $\mathcal{B}(p)$ are being assigned a code in $\bar{X}$. We now wish to determine the values of the coding time function C. It follows from the nature of the coding procedure that as we examine the fillers in $s_r(\hat{x})$, it is enough to detect the block

$$M^r \underset{l_{-(m_r-1)}}{\underline{\qquad}} M^{u_{-(m_r-1)}} \cdots \underset{l_{-1}}{\underline{\qquad}} M^{u_{-1}} \underset{l_0}{\underline{\qquad}} M^{u_1} \underset{l_1}{\underline{\qquad}} \cdots M^{u_{n_r-1}} \underset{l_{n_r-1}}{\underline{\qquad}} M^r$$

in order to be able to decide whether a filler $F \in \mathcal{F}(s_r)$ should be coded or not. Observe that the initial and terminal runs of markers might actually have been longer than r, but as soon as we see r-many consecutive markers we know that *all* the fillers contributing to s_r have already been seen. Note also that the length of the above block is $L_r + K_r + 2rk = c_r = c_r(\hat{x})$. In other words, if $F(x) \in \mathcal{F}(s_r(\hat{x}))$ is coded at step r, then $C(x) = c_r(\hat{x})$. By [1, Lemma 14] it is easy to conclude that

$$(9) \qquad \begin{cases} \mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) \text{ not coded}\} \\ \quad \leq 2^{b_r(\hat{x})-((g-\bar{g})/3)L_r(\hat{x})} + \mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) \text{ is bad}\} \\ \qquad + \bar{\mu}_0\{\bar{F}(\bar{x}) \in \bar{\mathcal{F}}(s_r(\hat{x})) \colon \bar{F}(\bar{x}) \text{ is bad}\}. \end{cases}$$

We now wish to write the expectation of C as a function of skeleton lengths and the above events (F and $\bar{F}$ being bad). It is shown in [1, Lemma 4] that the $\mu_{\hat{x}}$ are in fact the conditional probabilities of μ on the fibers $X(\hat{x})$. Consequently, we may disintegrate μ along the fibers and write

$$\boldsymbol{E}_\mu(C) = \int_{\hat{X}} \int_{X(\hat{x})} C(x)\, \mathrm{d}\mu_{\hat{x}}(x)\, \mathrm{d}\hat{\mu}(\hat{x}) = \boldsymbol{E}_{\hat{\mu}} \boldsymbol{E}_{\mu_{\hat{x}}}(C).$$

On the other hand, C is an integer valued random variable, and the only possible values of C are $c_r(\hat{x})$, $r = 1, 2, \ldots$, for $\hat{x} \in \hat{X}$, so we may write,

$$(10) \qquad \begin{cases} \boldsymbol{E}_{\mu_{\hat{x}}}(C) = \sum_{n=1}^{\infty} \mu_{\hat{x}}\{x \in X(\hat{x}) \colon C(x) \geq n\} \\ \qquad = \sum_{n=1}^{\infty} n\mu_{\hat{x}}\{x \in X(\hat{x}) \colon C(x) = n\} \\ \qquad = \sum_{r=1}^{\infty} c_r(\hat{x})\mu_{\hat{x}}\{x \in X(\hat{x}) \colon C(x) = c_r(\hat{x})\} \\ \qquad \leq \sum_{r=1}^{\infty} c_r(\hat{x})\mu_{\hat{x}}\{x \in X(\hat{x}) \colon C(x) \geq c_r(\hat{x})\}. \end{cases}$$

At this point note that

$$\{x \in X(\hat{x}) \colon C(x) \geq c_r(\hat{x})\} = \{x \in X(\hat{x}) \colon F(x) \in \mathcal{F}(s_{r-1}(\hat{x})) \text{ not coded}\},$$

the event that the filler $F(x)$ is not coded at step $r - 1$. Consequently,

$$\mu_{\hat{x}}\{x \in X(\hat{x}) \colon C(x) \geq c_r(\hat{x})\} = \mu_0\{F(x) \in \mathcal{F}(s_{r-1}(\hat{x})) \colon F(x) \text{ not coded}\}$$

and we have, by (9) and (10)

$$\begin{aligned} \boldsymbol{E}_\mu(C) &\leq \int_{\hat{X}} \sum_{r=1}^{\infty} c_{r+1}(\hat{x})\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) \text{ not coded}\}\, \mathrm{d}\hat{\mu}(\hat{x}) \\ &\leq \sum_{r=1}^{\infty} \int_{\hat{X}} c_{r+1}(\hat{x})(2^{b_r(\hat{x})-((g-\bar{g})/3)L_r(\hat{x})} + \mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) \text{ is bad}\} \\ &\quad + \bar{\mu}_0\{\bar{F}(\bar{x}) \in \bar{F}(s_r(\hat{x})) \colon \bar{F}(\bar{x}) \text{ is bad}\})\, \mathrm{d}\hat{\mu}(\hat{x}). \end{aligned}$$

Denote $c = (g - \bar{g})/3$. By symmetry of the definitions of $F \in \mathcal{F}$ and $\bar{F} \in \bar{\mathcal{F}}$ being bad, to prove that $\boldsymbol{E}_\mu(C)$ is finite, it suffices to show that

$$(11) \qquad \sum_{r=1}^{\infty} \boldsymbol{E}_{\hat{\mu}} c_{r+1}(2^{b_r - cL_r} + \mu_0\{F(x) \in \mathcal{F}(s_r)\colon F(x) \text{ is bad}\}) < \infty.$$

Let

$$S_1 = \sum_{r=1}^{\infty} \boldsymbol{E}_{\hat{\mu}}(c_{r+1} 2^{b_r - cL_r})$$

and

$$S_2 = \sum_{r=1}^{\infty} \boldsymbol{E}_{\hat{\mu}}(c_{r+1}\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x}))\colon F(x) \text{ is bad}\}).$$

The rest of the paper is devoted to proving that $S_1 + S_2 < \infty$.

4. PROOF OF THE MAIN THEOREM

4.1. S_1 is finite

In this section we prove that $S_1 < \infty$. We begin by observing that since $c_{r+1} = c_r + d_r + d_r'$ and d_r, d_r' are identically distributed, it suffices to define

$$S_{1,1} = \sum_{r=1}^{\infty} \boldsymbol{E}_{\hat{\mu}}(c_r 2^{b_r - cL_r}), \qquad S_{1,2} = \sum_{r=1}^{\infty} \boldsymbol{E}_{\hat{\mu}}(d_{r+1} 2^{b_r - cL_r}),$$

and prove that $S_{1,1} < \infty$ and $S_{1,2} < \infty$.

We start by computing $S_{1,1}$. Recall that c_r equals in distribution $L_r + K_r + 2kr$, and we shall show that each of the expectations $\boldsymbol{E}_{\hat{\mu}}(L_r 2^{b_r - cL_r})$, $\boldsymbol{E}_{\hat{\mu}}(K_r 2^{b_r - cL_r})$, $\boldsymbol{E}_{\hat{\mu}}(2kr 2^{b_r - cL_r})$ is summable with r. We begin by estimating the second term. Observe that

$$(12) \qquad K_r \le k \cdot (\text{number of markers inside } s_r) \le k \cdot b_r \cdot (r-1) < krb_r$$

and consequently

$$\begin{aligned}
\boldsymbol{E}_{\hat{\mu}}(K_r 2^{b_r - cL_r}) &< \boldsymbol{E}_{\hat{\mu}}(krb_r 2^{b_r - cL_r}) = \boldsymbol{E}(\boldsymbol{E}_{\hat{\mu}}(krb_r 2^{b_r - cL_r} \mid b_r)) \\
&= \sum_{l=1}^{\infty} \mathbb{P}(b_r = l)\boldsymbol{E}(krb_r 2^{b_r - cL_r} \mid b_r = l) \\
&= \sum_{l=1}^{\infty} l\eta^{2r-2}(1 - \eta^{r-1})^{l-1} krl 2^l G^{l+1}(2^{-c})2^{-c},
\end{aligned}$$

since L_r is a sum of $(b_r + 1)$-many i.i.d. random variables with the generating function G. Recall that if $0 < b < 1$, then $\sum_{l=1}^{\infty} l^2 b^l = (b(1+b))/(1-b)^3$. Consequently, if $2(1 - \eta^{r-1})G(2^{-c}) < 1$ then

$$\begin{aligned}
\boldsymbol{E}_{\hat{\mu}}(K_r 2^{b_r - cL_r}) &< \frac{kr\eta^{2r-2}G(2^{-c})2^{-c}}{1 - \eta^{r-1}} \sum_{l=1}^{\infty} l^2[2(1 - \eta^{r-1})G(2^{-c})]^l \\
&\le \frac{4G(2^{-c})}{[1 - 2G(2^{-c})]^3} kr\eta^{2r-2},
\end{aligned}$$

and since $(4G(2^{-c}))/[1 - 2G(2^{-c})]^3$ is a constant with respect to r, it is obvious that $\boldsymbol{E}(K_r 2^{b_r - cL_r})$ is summable with r, if only $2(1 - \eta^{r-1})G(2^{-c}) < 1$. In order for this condition to hold true, it is enough that $2G(2^{-c}) < 1$. From the fact that c is a given, positive constant and $G(s)$ is an increasing, continuous function on

$[0,1]$ with $G(0)=0$, $G(1)=1$ it follows that for a suitable choice of the parameter k (big enough, see (1)) we can ensure $G(2^{-c})<\frac{1}{2}$, as required.
It is an easy exercise to see that the same calculation as above shows that

$$\sum_{r=1}^{\infty} \boldsymbol{E}(2kr2^{b_r-cL_r})<\infty.$$

To show $S_{1,1}<\infty$ we only need to prove that $\sum_{r=1}^{\infty} \boldsymbol{E}(L_r 2^{b_r-cL_r})<\infty$.

Remark 4.

$$\lim_{r\to\infty} \frac{L_r(\hat{x})}{b_r(\hat{x})}=\bar{l}, \quad \hat{\mu}-\text{a.e.}$$

The $\bar{l}$ is the average filler length, and $\bar{l}=\bar{l}(k)=\boldsymbol{E}(\mathcal{L})=(1-k\eta)/\eta(1-\eta)$.

Assume k is chosen so big that $c\cdot\bar{l}>1$. Fix a positive constant δ. Denote, for every $r\geq 1$,

$$V_r=\left\{\hat{x}:\left|\frac{L_r(\hat{x})}{b_r(\hat{x})}-\bar{l}\right|>\delta\right\}.$$

We write $\hat{X}=V_r^c\cup V_{r,1}\cup V_{r,2}$, where $V_{r,1}=\{L_r>(\bar{l}+\delta)b_r\}$, $V_{r,2}=\{L_r<(\bar{l}-\delta)b_r\}$. Observe that on the set V_r^c we have $L_r<(\bar{l}+\delta)b_r$ and consequently

$$\int_{V_r^c} L_r 2^{b_r-cl_r}\,\mathrm{d}\hat{\mu}<\int_{V_r^c}(\bar{l}+\delta)b_r 2^{b_r-cl_r}\,\mathrm{d}\hat{\mu}\leq(\bar{l}+\delta)\boldsymbol{E}(b_r 2^{b_r-cl_r}). \tag{13}$$

From the considerations concerning K_r it is easy to deduce that $\boldsymbol{E}(b_r 2^{b_r-cl_r})$ is summable with r. In order to estimate the $V_{r,2}$ part of the expectation, note that

$$\int_{V_{r,2}} L_r 2^{b_r-cl_r}\,\mathrm{d}\hat{\mu}\leq(\bar{l}-\delta)\boldsymbol{E}(b_r 2^{b_r-cl_r})$$

and the latter is summable with r (see (13)). We now proceed to the computation of the integral over $V_{r,1}$. Note first that $L_r>(\bar{l}+\delta)b_r$ implies $b_r-cL_r<b_r(1-c\bar{l}-c\delta)$, and since $c\bar{l}>1$, there is a constant $d>c\delta>0$ such that $b_r-cL_r<-db_r$ on the set $V_{r,1}$. As a consequence we have

$$\begin{aligned}\int_{V_{r,1}} L_r 2^{b_r-cl_r}\,\mathrm{d}\hat{\mu}&\leq\int_{V_{r,1}} L_r 2^{-db_r}\,\mathrm{d}\hat{\mu}\leq\boldsymbol{E}(L_r 2^{-db_r})\\&=\sum_{l=1}^{\infty}\mathbb{P}(b_r=l)\boldsymbol{E}(L_r 2^{-db_r}\,|\,b_r=l)\\&=\sum_{l=1}^{\infty}\mathbb{P}(b_r=l)\boldsymbol{E}((N_{l+1}+1)2^{-dl}),\end{aligned}$$

with N_{l+1} as in Remark 2. It follows that

$$\begin{aligned}\int_{V_{r,1}} L_r 2^{b_r-cl_r}\,\mathrm{d}\hat{\mu}&\leq\sum_{l=1}^{\infty} l\eta^{2r-2}(1-\eta^{r-1})^{l-1}2^{-dl}\left[1+(l+1)\frac{1-k\eta}{\eta(1-\eta)}\right]\\&=\frac{1-k\eta}{\eta(1-\eta)}\,\frac{1}{1-\eta^{r-1}}\,\eta^{2r-2}\cdot C(d),\end{aligned}$$

where $C(d)$ is a constant, depending only upon d. It is clear that the above expression is summable with r. We finished proving that $S_{1,1}<\infty$.

In order to compute $S_{1,2}$, we first remark that:

$$\begin{aligned} E(d_{r+1}2^{b_r - cL_r}) &= \frac{1}{\eta^{r+1}} E(E(2^{b_r - cL_r} \mid b_r)) \\ &= \frac{1}{\eta^{r+1}} \sum_{l=1}^{\infty} \mathbb{P}(b_r = l) E(2^{b_r - cL_r} \mid b_r) \\ &= \frac{1}{\eta^{r+1}} \sum_{l=1}^{\infty} l\eta^{2r-2}(1-\eta^{r-1})^{l-1} 2^l 2^{-c} G^{l+1}(2^{-c}) \\ &= \eta^r \frac{2^{1-c}G^l(2^{-c})}{\eta^3(1-\eta^{r-1})} \sum_{l=1}^{\infty} l[2(1-\eta^{r-1})G(2^{-c})]^l. \end{aligned}$$

However, since $2(1-\eta^{r-1})G(2^{-c}) < 1$, the sum $\sum_{l=1}^{\infty} l[2(1-\eta^{r-1})G(2^{-c})]^l$ is bounded by a finite constant, independent of r. Consequently $\sum_{r=1}^{\infty} E(d_{r+1}2^{b_r - cL_r}) < \infty$, completing the proof of $S_1 < \infty$.

4.2. S_2 is finite

Recall that $S_2 = \sum_{r=1}^{\infty} E_{\hat{\mu}}(c_{r+1}\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x}))\colon F(x) \text{ is bad}\})$, filler $F \in \mathcal{F}(s_r)$ for $\mathcal{B}(p)$ is bad if $\mu_0(F) > 2^{-(g-\varepsilon)L_r}$ and our task is to show $S_2 < \infty$. We proceed as follows. First we find a large integer m. A filler is called too-long if its length exceeds m, otherwise we call it typical. We then choose a parameter $\delta > 0$ such that:

(i) in n is big enough then the probability that in a concatenation of n fillers, more than δ-fraction is covered by too-long fillers, is exponentially small

(ii) the contribution of the filler measure on the piece covered by too-long fillers is exponentially smaller than $\varepsilon/2$

(iii) if n is big enough, then the typical fillers appear with empirical frequencies δ-close to the limiting frequencies, outside a set of exponentially small measure (level-2 Large Deviation theory)

(iv) for a typical length $i \le m$ an appropriate filler entropy g_i is defined; g is shown to be a mixture of g_i's, exponential deviation of g (F being bad) is expressed in terms of exponential deviations of the g_i's; the probability of these deviations is shown to be exponentially small.

The computation is performed in two steps. First fix a positive integer N, to be determined later. Define $U_r = \{\hat{x}\colon b_r(\hat{x}) \le N\}$. It can easily be shown that

$$\sum_{r=1}^{\infty} \int_{U_r} (c_{r+1}\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x}))\colon F(x) \text{ is bad}\})\,\mathrm{d}\hat{\mu} < \infty. \tag{14}$$

Note first that

$$\begin{aligned} &\int_{U_r} (c_{r+1}\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x}))\colon F(x) \text{ is bad}\})\,\mathrm{d}\hat{\mu} \\ &\quad < \int_{U_r} c_{r+1}\,\mathrm{d}\hat{\mu} = E(d_{r+1}\mathbf{1}_{U_r}) + E(c_r\mathbf{1}_{U_r}). \end{aligned}$$

The d_{r+1} and $\mathbf{1}_{u_r}$ are clearly independent, so $\sum_{r=1}^{\infty} E(d_{r+1}\mathbf{1}_{U_r}) \le \sum_{r=1}^{\infty}(1/\eta^{r+1})N^2\eta^{2r-2} < \infty$, since $\hat{\mu}(U_r) = 1 - (1+N\eta^{r-1})(1-\eta^{r-1})^N \le N^2\eta^{2r-2}$.

For the computation of $\boldsymbol{E}(c_r \mathbf{1}_{U_r})$ recall that $c_r = L_r + K_r + 2kr$ and $K_r < krb_r$. We have now

$$\begin{aligned}\boldsymbol{E}(c_r \mathbf{1}_{U_r}) &= \sum_{l=1}^{\infty} \mathbb{P}(b_r = l)\boldsymbol{E}(c_r \mathbf{1}_{U_r} \mid b_r = l) \leq \sum_{l=1}^{N} \mathbb{P}(b_r = l)\boldsymbol{E}(c_r \mid b_r = l) \\ &= \eta^{2r-2} \sum_{l=1}^{N} l(1 - \eta^{r-1})^{l-1}[(l+1)\boldsymbol{E}(l_1) + 1 + lkr + 2kr] \\ &= \eta^{2r-2} C_1(N) + r\eta^{2r-2} C_2(N),\end{aligned}$$

where $C_1(N)$ and $C_2(N)$ are finite constants, independent of r.
Clearly $\sum_{r=1}^{\infty} \boldsymbol{E}(c_r \mathbf{1}_{U_r}) < \infty$, therefore showing (14).

To complete the proof of the main theorem, we proceed now according to the sketch given at the beginning of this section, we define $V_r = U_r^c$ and prove that

$$(15) \qquad \sum_{r=1}^{\infty} \int_{V_r} (c_{r+1}\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) \text{ is bad}\})\, \mathrm{d}\hat{\mu} < \infty.$$

Recall that each fiber $X(\hat{x})$ is identified with the product $\prod_{j \in \mathbb{Z}} A_0^{l_j}$, $l_j = l_j(\hat{x})$. Fix an integer m, to be determined later. For a fiber $X(\hat{x})$, let

$$I_i(\hat{x}) = \{j \in \mathbb{Z} \colon l_j(\hat{x}) = i\}, \quad i = 1, \ldots, m.$$

Define also $I'(\hat{x}) = \mathbb{Z} \setminus \bigcup_{i=1}^{m} I_i(\hat{x})$. Let $\rho_i = \mathbb{P}(\mathcal{L} = i)$ for $i \geq 1$. Clearly

$$\rho_i = \lim_{t \to \infty} \frac{1}{2t+1} \operatorname{card}\{-t \leq j \leq t \colon l_j(\hat{x}) = i\}, \quad \hat{\mu} - \text{a.e.}$$

i.e. ρ_i is the fraction of the fillers of length i among all fillers. Let us define

$$a_i := \frac{i\rho_i}{\sum_{i=1}^{\infty} i\rho_i} = \frac{1}{\bar{l}(k)} i\rho_i,$$

the limiting fraction of coordinates covered in almost every fiber by fillers of length $i \geq 1$. Let $\delta > 0$ be a parameter which we shall determine later and m be chosen as the minimal integer such that

$$\delta_m := 1 - \sum_{i=1}^{m} a_i \leq \delta.$$

Fillers with lengths greater than m we call *too-long*. Otherwise call a filler *typical*.

Definition 6. If s_r is of rank r, the filler lengths are $l_{-(m_r-1)}, \ldots, l_{n_r-1}$ and $F \in \mathcal{F}(s_r)$, then we write

$$F = G_1 \times \cdots \times G_m \times G'$$

where G_i is the concatenation of F_j's with length i, and G' is the concatenation of too-long F_j's, i.e. $G_i = \prod_{j \in I_i(\hat{x})} F_j$ and $-(m_r - 1) \leq j \leq n_r - 1$.

In this manner each filler $F \in \mathcal{F}(s_r)$ is split into $(m+1)$ parts. We are going to show that in order for F to be bad, we have to deviate the measure of at least one of the $(m+1)$ parts of F, and that event will be shown to have exponentially small probability. We start by defining g_i, the entropy of fillers of length i, as the μ-a.e. limit of the form:

$$g_i := \lim_{n\to\infty} \frac{-1}{\sum_{j=-n, j\in I_i(\hat{x})}^{n} l_j(\hat{x})} \log \prod_{j\in I_i(\hat{x})} \mu_0(F_j(x)),$$

where $F_j(x) \in A_0^{l_j(\hat{x})}$. The entropy g' of too-long fillers is defined in the similar way. The filler entropy g (recall (2)) can be expressed in terms of the g_i's as follows:

$$\begin{aligned}
g &= \lim_{n\to\infty} \frac{-1}{\sum_{i=-n}^{n} l_i(\hat{x})} \log \prod_{i=-n}^{n} \mu_0(F_i(x)) \\
&= \lim_{n\to\infty} \sum_{i=1}^{m} \frac{\sum_{j\in I_i(\hat{x})} l_j(\hat{x})}{\sum_{j=-n}^{n} l_j(\hat{x})} \frac{-1}{\sum_{j\in I_i(\hat{x})} l_j(\hat{x})} \log \prod_{j\in I_i(\hat{x})} \mu_0(F_i(x)) \\
&\quad + \lim_{n\to\infty} \frac{\sum_{j\in I'(\hat{x})} l_j(\hat{x})}{\sum_{j=-n}^{n} l_j(\hat{x})} \frac{-1}{\sum_{j\in I'(\hat{x})} l_j(\hat{x})} \log \prod_{j\in I'(\hat{x})} \mu_0(F_i(x)) \\
&= \sum_{i=1}^{m} a_i g_i + g'\delta_m,
\end{aligned}$$

where $\log(1/(\max_i p_i)) < g' < \log(1/(\min_i p_i))$.

For $i = 1, \ldots, m$, find N_i, a positive integer and $d_i > 0$ such that if $\varepsilon_i = \varepsilon/i^3$, $n \geq N_i$ and $G_i = G_{i,1} \times \cdots \times G_{i,n}$ is a concatenation of n-many fillers of length i, then

$$\begin{aligned}
&\mu_0\{F \in \mathcal{F}(s_r) \colon F = G_1 \times \cdots \times G_i \times \cdots \times G_m \times G' \\
&\qquad \text{and } \mu_0(G_i) > 2^{-(g_i-\varepsilon_i)l(G_i)}\} < 2^{-d_i n}.
\end{aligned}$$

The above is a consequence of Large Deviation theory, as $G_{i,j}$ are independent random variables and g_i is the logarithm of their expected value.

Further, define $\rho_i(s_r(\hat{x}))$ to be the empirical frequency of length i among the lengths $l_{-(m_r-1)}(\hat{x}), \ldots, l_{n_r-1}(\hat{x})$; and $\delta_m(s_r(\hat{x}))$ to be the fraction of filler coordinates of $s_r(\hat{x})$, covered by too-long fillers. Recall that $V_r = \{\hat{x}\colon b_r(\hat{x}) > N\}$, where N is a fixed finite integer, not specified yet. Fix $\Delta > 0$, and define

$$\begin{aligned}
Z_r &= \{\hat{x}\colon |\rho_i - \rho_i(s_r(\hat{x}))| > \Delta, \text{ for some } i = 1, \ldots, m\} \\
&\quad \cup \{\hat{x}\colon |\delta_m - \delta_m(s_r(\hat{x}))| > \Delta\} \\
&\quad \cup \left\{\hat{x}\colon \left|\bar{l}(k) - \frac{1}{b_r(\hat{x})} \sum_{j=-(m_r-1)}^{n_r-1} l_j(\hat{x})\right| > \Delta\right\}.
\end{aligned}$$

We want to split the integration over V_r into two integrals: over $V_r \cap Z_r$ and $V_r \cap Z_r^c$. The right choice of parameters δ, m and N is going to make both integrals summable with r.

First of all, using level-2 Large Deviation theory, we find N_0 and $b > 0$ such that for $n \geq N_0$ we have

$$\hat{\mu}(Z_r \,|\, b_r = n) < 2^{-bn}. \tag{16}$$

We are ready to specify N now. Let $N = \max(N_0, \max_{1\le i\le m}(N_i/(\rho_i - \Delta)))$. The reason for choosing N as above is the following:

(i) if $b_r(\hat{x}) > N_0$ then we are able to estimate the measure of Z_r

(ii) if $b_r(\hat{x}) > \max_{1\le i\le m}(N_i/(\rho_i - \Delta))$ and $\hat{x} \in V_r \cap Z_r^c$, then there are at least N_i-many fillers of length i inside $s_r(\hat{x})$, and for $n \ge N_i$ we have Large Deviation estimation for the probability of 'bad-sized' concatenation of n-many fillers of length i.

For the purpose of showing that

$$(17) \qquad \sum_{r=1}^{\infty} \int_{V_r \cap Z_r^c} (c_{r+1}\,\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) \text{ is bad}\})\, \mathrm{d}\hat{\mu} < \infty$$

we need the following lemma.

Lemma 1. *If* $\hat{x} \in Z_r^c$, $n \ge N$, $b_r(\hat{x}) = n$, $F = G_1 \times \cdots \times G_m \times G' \in \mathcal{F}(s_r(\hat{x}))$, $\varepsilon_i = \varepsilon/i^3$ *and* $\mu_0(F) > 2^{-(g-\varepsilon)L_r}$, *then for some* $1 \le i \le m$ *we have* $\mu_0(G_i) > 2^{-(g_i - \varepsilon_i)l(G_i)}$.

Proof. Remark that by taking Δ small and m big enough (therefore δ_m small) we can make the fraction of F occupied by its too-long part G' so small that the contribution to the total measure on this part is exponentially smaller than $\varepsilon/2$, and consequently we may not consider G' at all, by taking $\varepsilon/2$ instead of ε in the statement of the lemma. With that remark in mind, we assume that F consists of only typical lengths and work with ε.

Suppose that $\mu_0(G_i) < 2^{-(g_i - \varepsilon_i)l(G_i)}$, for $i = 1, \ldots, m$. It suffices now to prove that

$$\prod_{i=1}^{m} \mu_0(G_i) < 2^{-(g-\varepsilon)l}$$

holds, with the appropriate choice of parameters. Here $l = l(F) = L_r$. The above can be rewritten in a form

$$\sum_{i=1}^{m} (g_i - \varepsilon_i) l_i \ge (g - \varepsilon) l,$$

where $l_i = l(G_i)$. From the fact that $\hat{x} \in Z_r^c$ it follows that

(i) $in(\rho_i + \Delta) \ge l_i \ge in(\rho_i - \Delta)$

(ii) $n\bar{l}(k) \ge l - n\Delta$,

and these inequalities imply

$$\sum_{i=1}^{m} (g_i - \varepsilon_i) l_i \ge lg - \delta_m l - \Delta ng + \Delta\delta_m n - \sum_{i=1}^{m} in(g_i\Delta + \rho_i\varepsilon_i + \Delta\varepsilon_i).$$

It is sufficient to show that the RHS of the above exceeds $(g - \varepsilon)l$, or equivalently that

$$\delta_m l + \Delta ng - \Delta\delta_m n + \sum_{i=1}^{m} in(g_i\Delta + \rho_i\varepsilon_i + \Delta\varepsilon_i) < \varepsilon l.$$

We now omit the term $-\Delta\delta_m n$ on the LHS, substitute n by $(1/\bar{l}(k))l(1 + \Delta)$

(observe that $n < (1/\bar{l}(k))(l + n\Delta) \le (1/\bar{l}(k))l(1+\Delta))$ therefore making the LHS bigger. To prove the lemma it is enough now to show

$$\delta_m + \Delta g \frac{1}{\bar{l}(k)}(1+\Delta) + \frac{1}{\bar{l}(k)}(1+\Delta)\Delta \sum_{i=1}^m ig_i + \frac{1}{\bar{l}(k)}(1+\Delta)\Delta \sum_{i=1}^m i\varepsilon_i + \frac{1}{\bar{l}(k)}(1+\Delta)\sum_{i=1}^m i\rho_i\varepsilon_i < \varepsilon.$$

Call the five terms on the LHS of the above inequality $t_1, \dots, t_5$. We shall choose the parameters in such a way that $t_i < \varepsilon/5$ for $i = 1, \dots, 5$.

Observe that $t_1 = \delta_m < \varepsilon/5$ if δ, so also δ_m is small enough (therefore m must be chosen sufficiently big). Further, $t_2 = \Delta g(1/\bar{l}(k))(1+\Delta) < 2g(1/\bar{l}(k)) < \varepsilon/5$, if k is big enough (recall that $\bar{l}(k) = (1/k\eta) - k$).

In order to make $t_3 = (1/\bar{l}(k))(1+\Delta)\Delta\sum_{i=1}^m ig_i$ smaller than $\varepsilon/5$, choose Δ so small that $\sqrt{\Delta} < \rho_m$. We then have

$$t_3 < (1+\Delta)\sqrt{\Delta}\sum_{i=1}^m \frac{1}{\bar{l}(k)} i\rho_m g_i < (1+\Delta)\sqrt{\Delta}\sum_{i=1}^m \frac{1}{\bar{l}(k)} i\rho_i g_i < (1+\Delta)\sqrt{\Delta} g$$

so we must take Δ so small that $(1+\Delta)\sqrt{\Delta} g < \varepsilon/5$. For the remaining two terms we have:

$$t_4 = \frac{1}{\bar{l}(k)}(1+\Delta)\Delta\sum_{i=1}^m i\varepsilon_i = \frac{1}{\bar{l}(k)}(1+\Delta)\Delta\sum_{i=1}^m i^{-2} < \varepsilon/5$$

if k is big enough, and finally

$$t_5 = \frac{1}{\bar{l}(k)}(1+\Delta)\sum_{i=1}^m i\rho_i\varepsilon_i \le \frac{1}{\bar{l}(k)}(1+\Delta)\varepsilon\sum_{i=1}^m \frac{\rho_i}{i^2} < \varepsilon/5$$

if $\bar{l}(k) > 10$, say, or else if k is big enough. The proof of the lemma is completed.

With the choice of parameters as above, the lemma implies the following

$$\begin{aligned}
&\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon \mu_0(F(x)) > 2^{-(g-\varepsilon)L_r}\} \\
&\le \sum_{i=1}^m \mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) = G_1 \times \cdots \times G_m \times G' \\
&\qquad \text{and } \mu_0(G_i) \ge 2^{-(g_i-\varepsilon_i)l_i}\},
\end{aligned}$$

and as m is a finite number, to conclude the proof of (17), it suffices to show that for $1 \le i \le m$

$$\sum_{r=1}^\infty \int_{V_r \cap Z_r^c} (c_{r+1}\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon \mu_0(G_i) \ge 2^{-(g_i-\varepsilon_i)l_i}\})\,\mathrm{d}\hat{\mu} < \infty.$$

Observe, however, that the above integral is bounded by

$$\sum_{r=1}^\infty \sum_{n=N}^\infty \mathbb{P}(b_r = n)\boldsymbol{E}[(d_{r+1} + c_r)2^{-d_i(\rho_i-\Delta)n}\mathbf{1}_{V_r \cap Z_r^c} \mid b_r = n],$$

and the same type of calculation as in Section 4.1 shows that the above is finite. To complete the proof we still have to show that

$$\sum_{r=1}^{\infty} \int_{V_r \cap Z_r} (c_{r+1}\mu_0\{F(x) \in \mathcal{F}(s_r(\hat{x})) \colon F(x) \text{ is bad}\})\, \mathrm{d}\hat{\mu} < \infty.$$

The above sum is bounded by $\sum_{r=1}^{\infty} \boldsymbol{E}(d_{r+1}\mathbf{1}_{V_r \cap Z_r}) + \sum_{r=1}^{\infty} \boldsymbol{E}[(L_r + K_r + 2kr)\mathbf{1}_{V_r \cap Z_r}]$. For the first of the two terms we have, using (16)

$$\begin{aligned}\sum_{r=1}^{\infty} \boldsymbol{E}(d_{r+1}\mathbf{1}_{V_r \cap Z_r}) &= \sum_{r=1}^{\infty} \frac{1}{\eta^{r+1}} \boldsymbol{E}(\boldsymbol{E}(\mathbf{1}_{V_r \cap Z_r} \mid b_r)) \\ &\leq \sum_{r=1}^{\infty} \frac{1}{\eta^{r+1}} \sum_{l=N}^{\infty} l\eta^{2r-2}(1-\eta^{r-1})^{l-1} 2^{-bl} < \infty.\end{aligned}$$

The calculation of the second sum is very similar to one of the computations from Section 4.1, and will be left to the reader. The proof is completed. □

I would like to express my thanks to Toshihiro Hamachi and Mike Keane for helpful conversations and pointing out an error in an earlier version of this paper.

REFERENCES

[1] Keane, M. and M. Smorodinsky – A class of finitary codes. Israel J. Math. **26**, 352–371 (1977).

[2] Ornstein, D. – Factors of Bernoulli shifts are Bernoulli. Advances in Mathematics **5**, 349–364 (1971).