KYBERNETIKA — VOLUME 49 (2013), NUMBER 6, PAGES 868–882

UNIVERSALLY TYPICAL SETS FOR ERGODIC SOURCES OF MULTIDIMENSIONAL DATA

Tyll Krüger, Guido Montúfar, Ruedi Seiler,
and Rainer Siegmund-Schultze

We lift important results about universally typical sets, typically sampled sets, and empirical entropy estimation in the theory of samplings of discrete ergodic information sources from the usual one-dimensional discrete-time setting to a multidimensional lattice setting. We use techniques of packings and coverings with multidimensional windows to construct sequences of multidimensional array sets which in the limit build the generated samples of any ergodic source of entropy rate below an h_0 with probability one and whose cardinality grows at most at exponential rate h_0.

Keywords: universal codes, typical sampling sets, entropy estimation, asymptotic equipartition property, ergodic theory

Classification: 94A24, 62D05, 94A08

1. INTRODUCTION

An entropy-typical set is defined as a set of nearly full measure consisting of output sequences the negative log-probability of which is close to the entropy of the source distribution. The scope of this definition is revealed by the asymptotic equipartition property (AEP), which was introduced by McMillan [6] as the convergence in probability of the sequence $-\frac{1}{n}\log\mu(x_1^n)$ to a constant h, namely, the Shannon entropy rate of the process μ [9]. Many processes have the AEP, as has been shown, e. g., in [1, 2, 6, 7]. In particular, for stationary discrete-time ergodic processes, this property is guaranteed by the Shannon–McMillan (SM) theorem [6] and in the stronger form of almost-sure convergence by the Shannon–McMillan–Breiman (SMB) theorem [2]. These two theorems have been extended from discrete-time to amenable groups, including $\mathbb{Z}^d$ as a special case, by Kieffer [3] and Ornstein–Weiss [7], respectively.

Roughly speaking, the AEP implies that the output sequences of a random process are typically confined to a 'small' set of events which have all approximately the same probability of being realized, in contrast to the much larger set of all possible output sequences. This means that individual outcomes with much higher or smaller probability than e^{-nh} will rarely be observed. By the AEP, the entropy-typical sets have total probability close to one and their cardinality is fairly minimal among all sets with this property. This way, entropy-typical sets provide an important theoretical framework for

communication theory. Lossless source coding is a type of algorithm which performs data compression while ensuring that the exact reconstruction of the original data is possible from the compressed data. Lossless data compression can be achieved by encoding the typical set of a stochastic source with fixed length block codes of length nh. By the AEP, this length nh is also the average length needed. Hence compression at an asymptotic rate equal to the entropy rate is possible. This rate is optimal, in view of Shannon's source coding theorem [9].

In universal source coding, the aim is to find codes which efficiently compress down to the theoretical limit, i. e., the entropy rate, for any ergodic source without a need to be adapted to the specific source. We emphasize here that codes of that type are optimal data compressors for any *stationary* source, since by the ergodic decomposition theorem (see, e. g., [10]) any stationary source is a convex mixture of ergodic sources. Moreover, any asymptotically optimal universal compression scheme defines sequences of universally typical sets: for given ε, the set of all n-blocks such that their compression needs at most $(h + \varepsilon)n$ bits, is universally typical for all sources with entropy rate h or less. Vice versa, any *constructive* solution to the problem of finding universally typical sets yields an universal compression scheme, since the index in the universally typical set is an optimal code for the block. As will turn out, our approach for multidimensional sources is constructive. But one has to admit that such an *ad hoc* algorithm is, generally speaking, not very useful in practice, because determining the index should be very time consuming.

Many formats for lossless data compression, like ZIP, are based on the implementation of the algorithms proposed by Lempel and Ziv (LZ) [13] and [14], or variants of them, like the Welch modification [12]. The LZ algorithms allow to construct universally typical libraries. However, they are designed for text compression, i. e., for compression of 1-dimensional data sources. Lempel and Ziv [4] showed that universal coding of images is possible by first transforming the image to a 1-dimensional stream (scanning the image with a Peano–Hilbert curve, a special type of Hamilton path) and then applying the 1-dimensional algorithm LZ78 from [14]. The idea behind that approach is that the Peano–Hilbert curve scans hierarchically complete blocks before leaving them, maintaining most local correlations that way. In contrast, a simple row-by-row scan only preserves horizontal correlations. But with the Peano curve approach, while preserving local correlations in all directions, these correlations are much encrypted due to the inevitably fractal nature of that space-filling curve.

We take the point of view that the techniques of packing and counting can be better exploited in data compression with unknown distributions if, instead of transforming the 'image' into a 1-dimensional stream by scanning it with a curve, the multidimensional block structure is left untouched. This will allow to take more advantage of multidimensional correlations between neighbouring parts of the data, speed up the convergence of the counting statistics, and in turn fasten estimation and compression. This approach will be carried out in a forthcoming paper. The idea of the present paper is to extend theoretical results about typical sampling sets and universally typical sets to a truly multidimensional sampling-window setting. The proofs of these extensions are guided by the discussion of the 1-dimensional situation in Shields' monograph [11].

2. SETTINGS

We consider the d-dimensional lattice $\mathbb{Z}^d$ and the quadrant $\mathbb{Z}^d_+$. Consider a finite alphabet $\mathcal{A}$, $|\mathcal{A}| < \infty$ and the set of arrays with that alphabet: $\Sigma = \mathcal{A}^{\mathbb{Z}^d}$, $\Sigma_+ = \mathcal{A}^{\mathbb{Z}^d_+}$. We define the set of n-words as the set of $n \times \cdots \times n$ arrays $\Sigma^n := \mathcal{A}^{\Lambda_n}$ for the n-box $\Lambda_n := \{(i_1, \ldots, i_d) \in \mathbb{Z}^d_+ : 0 \leq i_j \leq n-1, j \in \{1, \ldots, d\}\}$. An element $x^n \in \Sigma^n$ has elements $x^n(\mathbf{i}) \in \mathcal{A}$ for $\mathbf{i} \in \Lambda_n$.

Let $\mathfrak{A}^{\mathbb{Z}^d}$ denote the σ-algebra of subsets of Σ generated by cylinder sets, i. e., sets of the following kind:

$$[y] := \{x \in \Sigma : x(\mathbf{i}) = y(\mathbf{i}), \mathbf{i} \in \Lambda\}, \quad y \in \mathcal{A}^\Lambda, |\Lambda| < \infty.$$

If C is a subset of $\mathcal{A}^\Lambda$, we will use the notation $[C]$ for $\cup_{y \in C}[y]$.

We denote by $\sigma_{\mathbf{r}}$ the natural lattice translation by the vector $\mathbf{r} \in \mathbb{Z}^d$ acting on Σ by $\sigma_{\mathbf{r}} x(\mathbf{i}) := x(\mathbf{i} + \mathbf{r})$. We use the same notation $\sigma_{\mathbf{r}}$ to denote the induced action on the set $\mathbb{P}$ of probability measures ν over $(\Sigma, \mathfrak{A}^{\mathbb{Z}^d})$: $\sigma_{\mathbf{r}}\nu(E) := \nu(\sigma_{\mathbf{r}}^{-1}E)$. The set of all stationary (translation-invariant) elements of $\mathbb{P}$ is denoted by $\mathbb{P}_{\text{stat}}$, i. e., $\nu \in \mathbb{P}_{\text{stat}}$ if $\sigma_{\mathbf{r}}\nu = \nu$ for each $\mathbf{r} \in \mathbb{Z}^d$. Those $\nu \in \mathbb{P}_{\text{stat}}$ which cannot be decomposed as a proper convex combination $\nu = \lambda_1\nu_1 + \lambda_2\nu_2$, with $\nu_1 \neq \nu \neq \nu_2$ and $\nu_1, \nu_2 \in \mathbb{P}_{\text{stat}}$ are called *ergodic*. The corresponding subset of $\mathbb{P}_{\text{stat}}$ is denoted by $\mathbb{P}_{\text{erg}}$. Throughout this paper μ will denote an ergodic $\mathcal{A}$-process on Σ. By ν^n we denote the restriction of the measure ν to the block Λ_n, obtained by the projection $\Pi_n : x \in \Sigma \to x^n \in \Sigma^n$ with $x^n(\mathbf{i}) = x(\mathbf{i}), \mathbf{i} \in \Lambda_n$. We use the same notation Π_k to denote the projections from Σ^n to $\Sigma^k, n \geq k$, defined in the same obvious way. The measurable map Π_n transforms the given probability measure ν to the probability measure denoted by ν^n.

The entropy rate of a stationary probability measure ν is defined as limit of the scaled n-word entropies:

$$H(\nu^n) := -\sum_{x \in \Sigma^n} \nu^n(\{x\}) \log \nu^n(\{x\})$$

$$h(\nu) := \lim_{n \to \infty} \frac{1}{n^d} H(\nu^n).$$

Here and in the following we write log for the dyadic logarithm $\log_2$.

For a *shift* $\mathbf{p} \in \Lambda_k$ we consider the following partition of $\mathbb{Z}^d$ into *k-blocks*:

$$\mathbb{Z}^d = \bigcup_{\mathbf{r} \in k \cdot \mathbb{Z}^d} (\Lambda_k + \mathbf{r} + \mathbf{p}),$$

and in general we use the following notation:

The *regular k-block partitions* of a subset $M \subset \mathbb{Z}^d$ are the families of sets defined by

$$\mathscr{R}_{M,k} := \{R_{M,k}(\mathbf{p}) : \mathbf{p} \in \Lambda_k\}, \quad R_{M,k}(\mathbf{p}) := \{(\Lambda_k + \mathbf{p} + \mathbf{r}) \cap M\}_{\mathbf{r} \in k \cdot \mathbb{Z}^d}.$$

Clearly, for any $\mathbf{p}$ the elements of $R_{M,k}(\mathbf{p})$ are disjoint and their union gives M.

In the case $M = \Lambda_n$, given a sample $x^n \in \Sigma^n$, such a partition yields a *parsing* of x^n in elements of $\mathcal{A}^{(\Lambda_k + \mathbf{r} + \mathbf{p}) \cap \Lambda_n}, \mathbf{r} \in k \cdot \mathbb{Z}^d$. We call those elements the *words* of the parsing of x^n induced by the partition $R_{\Lambda_n,k}(\mathbf{p})$. With exception of those $\mathbf{r}$, for which

$\Lambda_k + \mathbf{r} + \mathbf{p}$ crosses the boundary of Λ_n, these are cubic k-words. Forgetting about their $\mathbf{r}$-position, we may identify $\Pi_{\Lambda_k} x \sim \Pi_{\Lambda_k+\mathbf{r}} \sigma_{-\mathbf{r}} x \in \mathcal{A}^{\Lambda_k+\mathbf{r}} \cong \mathcal{A}^{\Lambda_k}$.

For $k, n \in \mathbb{N}$, $k < n$, any element $x \in \Sigma$ gives rise to a probability distribution defined by the relative frequency of the different k-words in a given parsing of x_n. Let us introduce the following expression for these frequency counts:

$$Z_x^{\mathbf{p},k,n}(a) := \sum_{\mathbf{r} \in \times_{i=1}^d \{0,\ldots,\lfloor (n-p_i)/k \rfloor - 1\}} \mathbf{1}_{[a]}(\sigma_{k\cdot\mathbf{r}+\mathbf{p}} x), \tag{1}$$
$$n \in \mathbb{N},\ k \leq n,\ a \in \mathcal{A}^{\Lambda_k},\ \mathbf{p} = (p_1, \ldots, p_d) \in \Lambda_k.$$

For regular k-block parsings, the *non-overlapping empirical k-block distribution* generated by $x \in \Sigma$ in the box Λ_n is the probability distribution on Σ^k given by:

$$\tilde{\mu}_x^{k,n}(\{a\}) := \frac{1}{\lfloor n/k \rfloor^d} Z_x^{\mathbf{0},k,n}(a) \quad \text{for } a \in \mathcal{A}^{\Lambda_k}. \tag{2}$$

Similarly, for any $\mathbf{p} = (p_1, \ldots, p_d) \in \Lambda_k$ the shifted regular k-block partition gives a non-overlapping empirical k-block distribution:

$$\tilde{\mu}_x^{\mathbf{p},k,n}(\{a\}) := \frac{1}{\prod_{i=1}^d \lfloor (n-p_i)/k \rfloor} Z_x^{\mathbf{p},k,n}(a). \tag{3}$$

We will also use the *overlapping empirical k-block distribution*, in which all k-words present in x are considered:

$$\tilde{\mu}_{x,overl}^{k,n}(\{a\}) := \frac{1}{(n-k+1)^d} \sum_{\mathbf{r} \in \Lambda_{n-k+1}} \mathbf{1}_{[a]}(\sigma_{\mathbf{r}} x) \quad \text{for } a \in \mathcal{A}^{\Lambda_k}. \tag{4}$$

3. RESULTS

The main contribution of this paper is the following:

Theorem 3.1. (Universally typical sets) For any given $0 < h_0 \leq \log |\mathcal{A}|$ there is a sequence of subsets $\{\mathscr{T}_n(h_0) \subset \Sigma^n\}_n$ such that for all $\mu \in \mathbb{P}_{\text{erg}}$ with $h(\mu) < h_0$ the following holds:

a) $\lim_{n\to\infty} \mu^n(\mathscr{T}_n(h_0)) = 1$ and, in fact, $x^n \in \mathscr{T}_n(h_0)$ eventually μ-almost surely.

b) $\lim_{n\to\infty} \dfrac{\log |\mathscr{T}_n(h_0)|}{n^d} = h_0.$

For each n, a possible choice of $\mathscr{T}_n(h_0)$ is the set of arrays with empirical k-block distributions of per-site entropies not larger than h_0, where $k = \left\lfloor \sqrt[d]{\frac{1}{2} \log_{|\mathcal{A}|} n^d} \right\rfloor$. Furthermore, for any sequence $\{\mathscr{U}_n \subset \Sigma^n\}_n$ with $\liminf_{n\to\infty} \frac{1}{n^d} \log |\mathscr{U}_n| < h_0$, there exists a $\mu \in \mathbb{P}_{\text{erg}}$ with $h(\mu) < h_0$ which satisfies:

c) $\liminf_{n\to\infty} \mu^n(\mathscr{U}_n) = 0.$

In fact, when $\limsup_{n\to\infty} \frac{1}{n^d} \log |\mathscr{U}_n| < h_0$, then $x^n \notin \mathscr{U}_n$ eventually μ-almost surely.

The proof of Theorem 3.1 is based on other assertions following now. Although the 1-dimensional special case of the theorem can be inferred from the existence of universal codes for the class of ergodic processes on $\mathbb{Z}$ and the non-existence of too-good codes, to our knowledge it has not been formulated explicitly before. The strategy of our proof is guided by the discussion of 1-dimensional universal codes contained in [11, Theorem II.1.1, Theorem II.1.2, and Section II.3.d].

We start lifting the *packing lemma* [11, Lemma I.3.3]. We show that if a set of words $C \subset \Sigma^m$ is typical among all m-blocks present in a sample $x^k \in \Sigma^k$, $k \geq m$, i. e., C has large probability in the overlapping empirical m-block distribution, then the sample x^k can be parsed into non-overlapping blocks in such a way that nearly all words belong to C. The following lemma asserts that a parsing with many matchings and only few 'holes' can be realized by a regular partition; i. e., C receives large probability in the non-overlapping empirical distribution of some shift of x.

Lemma 3.2. (Packing lemma) For any $0 < \delta \leq 1$ let k and m be integers satisfying $k \geq d \cdot m/\delta$. Let $C \subset \Sigma^m$ and let $x \in \Sigma$ be such that $\tilde{\mu}^{m,k}_{x,overl}(C) \geq 1 - \delta$. Then there is a $\mathbf{p} \in \Lambda_m$ such that a) $\tilde{\mu}^{\mathbf{p},m,k}_x(C) \geq 1 - 2\delta$, and b) $|Z^{\mathbf{p},m,k}_x(C)| \geq (1 - 4\delta)(\lfloor \frac{k}{m} \rfloor + 2)^d$.

The condition on the array x means that $\sum_{\mathbf{r}\in\Lambda_{k-m+1}} \mathbf{1}_{[C]}(\sigma_{\mathbf{r}} x) \geq (1-\delta)(k-m+1)^d$. The first statement a) means that there exists a regular m-block partition $R_{\Lambda_k,m}(\mathbf{p}) \in \mathscr{R}_{\Lambda_k,m}$ that parses x^k in such a way that at least a $(1-2\delta)$-fraction of the m-words are elements of C. When $\delta = 0$ and $k \geq m$ this statement is trivial. The second statement b) implies that at least a $(1-4\delta)$-fraction of the total number of words are elements of C (this total number including non-cubical words at the boundary).

P r o o f . (Proof of Lemma 3.2) Denote by Ξ the set of vectors $\{\mathbf{r} \in \Lambda_{k-m+1} : \sigma_{\mathbf{r}} x \text{ is in } [C]\}$. For any $\mathbf{p} \in \Lambda_m$ denote by $\lambda(\mathbf{p})$ the number of those $\mathbf{r} \in \Xi$ satisfying $\mathbf{r} = \mathbf{p} \mathrm{mod}(m)$. Clearly, $\lambda(\mathbf{p}) = |Z^{\mathbf{p},m,k}_x(C)|$ is the number of cubic blocks in the $\mathbf{p}$-shifted regular m-block partition of Λ_k which belong to C. Then we have $\sum_{\mathbf{r}\in\Lambda_{k-m+1}} \mathbf{1}_{[C]}(\sigma_{\mathbf{r}} x) = \sum_{\mathbf{p}\in\Lambda_m} \lambda(\mathbf{p}) \geq (1-\delta)(k-m+1)^d$, by assumption. Hence, there is at least one $\mathbf{p}' \in \Lambda_m$ for which $\lambda(\mathbf{p}') \geq \frac{(1-\delta)(k-m+1)^d}{m^d}$. It is easy to see that $(1-\delta)\frac{(k-m+1)^d}{m^d} \geq (1-\delta)\frac{k^d - dmk^{d-1}}{m^d} \geq (1-\delta)^2 \frac{k^d}{m^d} \geq (1-2\delta)\frac{k^d}{m^d}$. Since the maximal number of m-blocks that can occur in $R_{\Lambda_k,m}(\mathbf{p}')$ is $(\frac{k}{m})^d$, this completes the proof of a). For b) observe that the total number of partition elements of the regular partition (including the non-cubic at the boundary) is upper bounded by $\left(\lfloor \frac{k}{m} \rfloor + 2\right)^d \leq \frac{1}{m^d}(k+2m)^d \leq \frac{1}{m^d}\left(k^d + (k+2m)^{d-1} 2dm\right) \leq \frac{1}{m^d}\sum_{j=0}^{d} k^{d-j}(2dm)^j \leq \frac{k^d}{m^d}\frac{1-(2\delta)^{d+1}}{1-2\delta}$. Here for the second inequality we used the estimate $1-(d-1)y \leq 1/(1+y)^{d-1}, y \geq 0$ and for the third one the estimate $\binom{d-1}{j} \leq d^j$. On the other hand, from the first part we have $\lambda(\mathbf{p}') = |Z^{\mathbf{p},m,k}_x(C)| \geq (1-2\delta)\frac{k^d}{m^d}$ and $1 - 2\delta \geq \frac{1-4\delta}{1-2\delta} \geq (1-4\delta)\frac{1-(2\delta)^{d+1}}{1-2\delta}$. This completes the proof. □

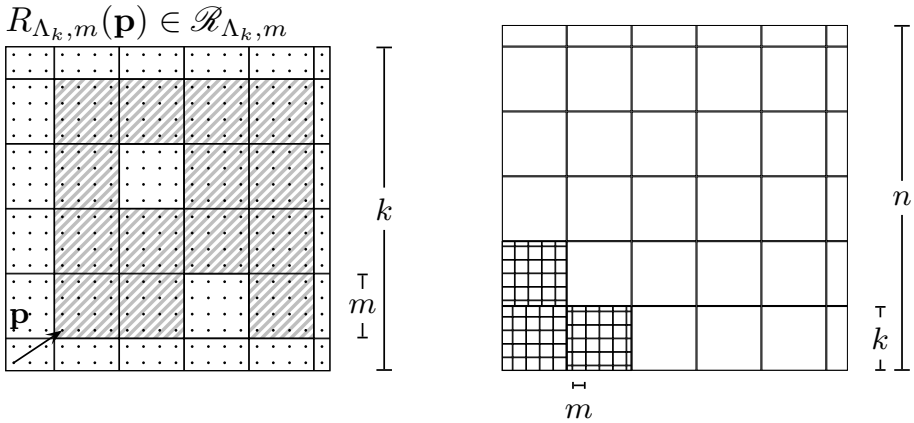

Fig. 1. Left: A $\mathbf{p}$-shifted regular m-block parsing of an array $x^k \in \mathcal{T}_k^\mu(\delta, m)$, for $d = 2$. The shaded blocks contain m-arrays from C_m^μ and fill at least a $(1-\delta)$-fraction of the total volume k^2. For $k \gg m$ the boundary blocks have a negligible volume. **Right:** A k-block parsing of an array x^n showing possible regular m-block parsings of the resulting k-blocks.

We need two definitions before we continue formulating the results:

Definition 3.3. (Entropy-typical sets) Let $\delta \in (0, \frac{1}{2})$. For some μ with entropy rate $h(\mu)$ the *entropy-typical sets* are defined as:

$$C_m^\mu(\delta) := \left\{ x \in \Sigma^m : 2^{-m^d(h(\mu)+\delta)} \le \mu^m(\{x\}) \le 2^{-m^d(h(\mu)-\delta)} \right\}. \tag{5}$$

We use these sets to define the following *typical sampling sets*. See Figure 1.

Definition 3.4. (Typical sampling sets) For some μ, $\delta \in (0, \frac{1}{2})$, and $k \ge m$, we define a *typical sampling set* $\mathcal{T}_k^\mu(\delta, m)$ as the set of elements in Σ^k that have a regular m-block partition such that the resulting words belonging to the μ-entropy typical set $C_m^\mu = C_m^\mu(\delta)$ contribute at least a $(1-\delta)$-fraction to the (slightly modified) number of partition elements in that regular m-block partition.

$$\mathcal{T}_k^\mu(\delta, m) := \left\{ x \in \Sigma^k : \sum_{\substack{\mathbf{r} \in m \cdot \mathbb{Z}^d : \\ (\Lambda_m + \mathbf{r} + \mathbf{p}) \subseteq \Lambda_k}} \mathbf{1}_{[C_m^\mu]}(\sigma_{\mathbf{r}+\mathbf{p}} x) \ge (1-\delta) \left(\frac{k}{m}\right)^d \text{ for some } \mathbf{p} \in \Lambda_m \right\}.$$

We fix some $\alpha > 0$ and assume $\delta < \alpha/(\log|\mathcal{A}| + 1)$. In the following we will choose m depending on k such that $m \xrightarrow{k\to\infty} \infty$ and $\lim_{k\to\infty} \frac{m}{k} = 0$. As it turns out, a sequence

$\mathcal{T}_k^\mu(\delta, m)$ satisfying these conditions, denoted $\mathcal{T}_k(\alpha)$, is a sequence of 'small' libraries from which the realizations of the ergodic process μ can be constructed asymptotically almost surely. This is the statement of Theorem 3.5, which generalizes a previous result by Ornstein and Weiss [8, Section 2, Theorem 2] (see [11, Theorem II.3.1]).

Theorem 3.5. Let $\mu \in \mathbb{P}_{\text{erg}}$ and $\alpha \in (0, \frac{1}{2})$. Then:

a) For all k larger than some $k_0 = k_0(\alpha)$ there is a set $\mathcal{T}_k(\alpha) \subset \Sigma^k$ satisfying

$$\frac{\log |\mathcal{T}_k(\alpha)|}{k^d} \leq h(\mu) + \alpha,$$

and such that for μ-a.e. x the following holds:

$$\tilde{\mu}_x^{k,n}\left(\mathcal{T}_k(\alpha)\right) > 1 - \alpha,$$

for all n and k with $\frac{k}{n} < \varepsilon$ for some $\varepsilon = \varepsilon(\alpha) > 0$ and n larger than some $n_0(x)$.

b) Let $\{\tilde{\mathcal{T}}_{k,n}(x)\}_{k,n>0}$ be a family of double-sequences of subsets of Σ^k, depending measurably on $x \in \Sigma$, with cardinality $|\tilde{\mathcal{T}}_{k,n}(x)| \leq 2^{k^d(h(\mu)-\alpha)}$. Then there exists a $k_1(\alpha) \geq k_0(\alpha)$ and for μ-a.e. x there exists an $n_0(x)$ such that

$$\tilde{\mu}_x^{k,n}(\tilde{\mathcal{T}}_{k,n}(x)) \leq \alpha,$$

whenever $k > k_1(\alpha)$, $n > n_0(x)$, and $2^{k^d(h(\mu)+\alpha)} \leq n^d$.

Using Theorem 3.5 we will prove the following Theorem 3.6, which states that the entropy of the non-overlapping empirical distribution of a sample converges almost surely to the true entropy of the process as the size of the parsing blocks grows to infinity without exceeding a logarithmic bound with respect to the size of the sampled region. In particular, this result describes a procedure to estimate entropies from samples. In fact, the inspiring 1-dimensional result [11, Theorem II.3.5] is called *entropy-estimation theorem*. We will use the alternative name *empirical-entropy theorem*, referring to its resemblance to the SMB or entropy theorem. This result will be a central ingredient in proving the existence of small universally typical libraries (Theorem 3.1).

Theorem 3.6. (Empirical-entropy theorem) Let $\mu \in \mathbb{P}_{\text{erg}}$. Then for any sequence $\{k_n\}_n$ with $k_n \xrightarrow{n\to\infty} \infty$ and $k_n^d(h(\mu) + \alpha) \leq \log n^d$ (for some $\alpha > 0$) we have

$$\lim_{n\to\infty} \frac{1}{k_n^d} H(\tilde{\mu}_x^{k_n,n}) = h(\mu) \quad \mu\text{-almost surely.}$$

This concludes the section of results. Below we provide the proofs.

4. PROOFS

P r o o f . (Proof of Theorem 3.5 a)) We show that the claim holds choosing $\mathcal{T}_k(\alpha)$ as typical sampling sets $\mathcal{T}_k^\mu(\delta, m)$ from Definition 3.4 with $\delta < \frac{\alpha}{\log|\mathcal{A}|+1}$, $m \xrightarrow{k\to\infty} \infty$, and $\lim_{k\to\infty} \frac{m}{k} = 0$.

Cardinality. We estimate the cardinality of the sets $\mathcal{T}_k^\mu(\delta, m)$. For a given m, there are m^d possible values of $\mathbf{p}$. There are at most $\left(\frac{k}{m}\right)^d$ cubic boxes in any m-block partition of Λ_k. Therefore, the number of choices for the contents of all blocks which belong to C_m^μ is at most $|C_m^\mu|^{\left(\frac{k}{m}\right)^d}$. By the definition of $\mathcal{T}_k^\mu(\delta, m)$, the number of lattice sites not belonging to the regular partition is at most δk^d. There are $|\mathcal{A}|^{\delta k^d}$ possible values for these sites. Let $K = \left\lfloor \frac{k}{m} \right\rfloor + 2$. The maximal number of blocks in the partition, including non-cubic ones, is K^d. For $\frac{m}{k}$ small enough, not more than a $2\delta \le \alpha < \frac{1}{2}$ fraction of all these blocks have contents not in C_m^μ. Taking into account that the binomial coefficients $\binom{K}{l}$ do not decrease in l while $l \le \frac{1}{2}K$, we get the following bound:

$$\begin{aligned} |\mathcal{T}_k^\mu(\delta, m)| &\le m^d \sum_{0\le l\le 2\delta K^d} \binom{K^d}{l} |\mathcal{A}|^{\delta k^d} |C_m^\mu|^{\left(\frac{k}{m}\right)^d} \\ &\le m^d K^d \binom{K^d}{\lfloor \frac{1}{2} K^d \rfloor} |\mathcal{A}|^{\delta k^d} |C_m^\mu|^{\left(\frac{k}{m}\right)^d}. \end{aligned}$$

We apply Stirling's formula $N! \simeq \sqrt{2\pi N}(\frac{N}{e})^N$, taking into account that the multiplicative error for positive N is uniformly bounded from below and above. A coarse bound will suffice. In the following estimate we make use of the relation $|C_m^\mu| \le 2^{m^d(h(\mu)+\delta)}$, following immediately from the definition of C_m^μ. For some positive constants c, c', and c'' we have

$$\begin{aligned} \log|\mathcal{T}_k^\mu(\delta, m)| &\le \log\ cm^d K^d \left(\frac{K^d}{\lfloor \frac{1}{2} K^d \rfloor}\right)^{K^d} \sqrt{\frac{K^d}{\lfloor \frac{1}{2} K^d \rfloor^2}} |\mathcal{A}|^{\delta k^d} |C_m^\mu|^{\left(\frac{k}{m}\right)^d} \\ &\le \log\ c' m^d 3^{K^d} K^{d/2} |\mathcal{A}|^{\delta k^d} |C_m^\mu|^{\left(\frac{k}{m}\right)^d} \\ &\le \log\ c'' k^d 3^{(\frac{k}{m}+2)^d} 2^{(h(\mu)+\delta+\delta\log|\mathcal{A}|)k^d} \\ &\le k^d \left(h(\mu) + \delta(\log|\mathcal{A}|+1) + \frac{2^d}{m^d}\log 3 + \frac{\log k^d + \log c''}{k^d} \right). \end{aligned}$$

In the last line we used $1/m + 2/k \le 2/m$, which holds when k/m is large enough. When $\delta < \frac{\alpha}{\log|\mathcal{A}|+1}$, and m as well as k are large enough (depending on α), this yields $\log|\mathcal{T}_k(\alpha)| \le k^d(h(\mu)+\alpha)$.

Probability bound. Ornstein and Weiss' extension [7] of the SMB theorem shows[1]:

$$\lim_{m\to\infty} -\frac{1}{m^d} \log \mu^m(\Pi_m x) = h(\mu) \quad \mu\text{-almost surely.}$$

[1]Here, in fact, we only need the convergence in probability [3], which ensures $\mu(C_m^\mu) \xrightarrow{m\to\infty} 1$.

Thus, by the definition of C^μ_m (Definition 3.3), there exists an $m_0(\delta)$ such that $\mu^m\left(C^\mu_m\right) \geq 1-\delta^2/5$ for all $m \geq m_0(\delta)$. We fix such an m. The individual ergodic theorem [5] asserts that the following limit exists for μ-almost every $x \in \Sigma$:

$$\lim_{n\to\infty} \frac{1}{n^d} \sum_{r\in\Lambda_n} \mathbf{1}_{[C^\mu_m]}\left(\sigma_r x\right) = \int \mathbf{1}_{[C^\mu_m]}(x)\,\mathrm{d}\mu(x) = \mu^m(C^\mu_m),$$

and therefore,

$$\sum_{r\in\Lambda_{n-m+1}} \mathbf{1}_{[C^\mu_m]}(\sigma_r x) \;\geq\; (1-\delta^2/4)(n-m+1)^d \;>\; (1-\delta^2/3)n^d \tag{6}$$

holds eventually almost surely, i. e., for μ-almost every x, choosing n large enough depending on x, $n \geq n_0(x)$.

Take an $x \in \Sigma$ and an $n \in \mathbb{Z}_+$ for which this is the case and eq. (6) is satisfied. Choose a k with $m < k < n$. Consider the unshifted regular k-block partition of the n-block Λ_n:

$$\Lambda_n = \bigcup_{\mathbf{r}\in k\cdot\mathbb{Z}^d} (\Lambda_k + \mathbf{r}) \cap \Lambda_n.$$

In the following we deduce from eq. (6) that if k/m and n/k are large enough, at least a $(1-2\delta)$-fraction of the k-blocks in this regular k-block parsing of $\Pi_n x$ (those which count for the empirical distribution $\tilde{\mu}^{k,n}_x$) satisfy

$$\frac{1}{(k-m+1)^d} \sum_{\mathbf{s}\in\Lambda_{k-m+1}} \mathbf{1}_{[C^\mu_m]}(\sigma_{\mathbf{s}+\mathbf{r}} x) \;\geq\; (1-\delta/4). \tag{7}$$

This is because if more than the specified 2δ-fraction of the k-blocks had more than a $\delta/4$-fraction of 'bad' m-blocks, then the total number of 'bad' m-blocks in $\Pi_n x$ would be larger than

$$2\delta \left\lfloor \frac{n}{k} \right\rfloor^d \cdot \frac{\delta}{4}(k-m+1)^d \;\geq\; \frac{\delta^2}{2}\left(\left(1-\frac{k}{n}\right)\left(1-\frac{m}{k}\right)\right)^d n^d \;>\; \frac{\delta^2}{3} n^d,$$

for $\frac{k}{n}$ and $\frac{m}{k}$ small enough, contradicting eq. (6). While n had to be chosen large enough depending on x, we see that k has to be chosen such that $\frac{k}{n}$ and $\frac{m}{k}$ are both small enough.

By Lemma 3.2, if $k \geq 4dm/\delta$, the k-blocks which satisfy eq. (7) have a regular m-block partition with at least a $(1-\delta)$-fraction of all partition members in C^μ_m. Hence, at least a $(1-2\delta)$-fraction of all k-blocks in Λ_n counting for the empirical distribution, belong to $\mathcal{T}^\mu_k(\delta, m)$. For $2\delta \leq \alpha$ we get the probability bound:

$$\tilde{\mu}^{k,n}_x\left(\mathcal{T}^\mu_k(\delta, m)\right) \geq 1-\alpha. \tag{8}$$

This completes the proof of Theorem 3.5 a). □

P r o o f. (Proof of Theorem 3.5 b)) The statement is trivial for $h(\mu) = 0$. Let $h(\mu) > 0$. For a fixed $\delta < \alpha$ consider the sets $E_n(\delta)$ of all x in Σ with

$$\tilde{\mu}^{k,n}_x(\mathcal{T}_k(\delta)) \geq 1-\delta \quad \text{for all } k \geq k_0(\delta),\ 2^{k^d(h(\mu)+\alpha)} \leq n^d,$$

where $k_0 = k_0(\delta)$ is chosen large enough as in the first part of the theorem. Consider the sets $D_n(\alpha, \delta)$ of all x in Σ with

$$\tilde{\mu}_x^{k,n}(\widetilde{\mathcal{T}}_{k,n}(x)) > \alpha \quad \text{for some } k \text{ with } k \geq k_0(\delta),\ 2^{k^d(h(\mu)+\alpha)} \leq n^d,$$

and let

$$F_n(\delta, \alpha) = [C_n^{\mu}(\delta)] \cap D_n(\alpha, \delta) \cap E_n(\delta).$$

The restriction $\Pi_n x$ of any $x \in D_n(\alpha, \delta) \cap E_n(\delta)$ can be described as follows.

1. First we specify a k with $k \geq k_0(\delta), 2^{k^d(h(\mu)+\alpha)} \leq n^d$ as in the definition of $D_n(\alpha, \delta)$.
2. Next, for each of the $\left\lfloor \frac{n}{k} \right\rfloor^d$ blocks counting for the empirical distribution, we specify whether this block belongs to $\widetilde{\mathcal{T}}_{k,n}(x)$, to $\mathcal{T}_k(\delta) \setminus \widetilde{\mathcal{T}}_{k,n}(x)$ or to $\Sigma^k \setminus (\mathcal{T}_k(\delta) \cup \widetilde{\mathcal{T}}_{k,n}(x))$.
3. Then we specify for each such block its contents, pointing either to a list containing all elements of $\widetilde{\mathcal{T}}_{k,n}(x)$, or to a list containing $\mathcal{T}_k(\delta) \setminus \widetilde{\mathcal{T}}_{k,n}(x)$ or, in the last case, listing all elements of that block.
4. Finally, we list all boundary elements not covered by the empirical distribution.

In order to specify k we need at most $\log n$ bits (in fact, much less, due to the bound on k). We need at most $2\left\lfloor \frac{n}{k} \right\rfloor^d$ bits to specify which of the cases under 2. is valid for each of the blocks. For 3. we need the two lists for the given k. This needs at most $\left(2^{k^d(h(\mu)+\delta)} + 2^{k^d(h(\mu)-\alpha)}\right) k^d(\log|\mathcal{A}| + 1)$ bits. According to the definitions of $D_n(\alpha, \delta)$ and $E_n(\delta)$, to specify the contents of all k-blocks, we need at most

$$\left(\frac{n}{k} + 1\right)^d k^d \left(\alpha(h(\mu) - \alpha) + (1 - \alpha)(h(\mu) + \delta) + \delta(\log|\mathcal{A}| + 1)\right)$$

bits. For 4. we need at most $(n^d - \left\lfloor \frac{n}{k} \right\rfloor^d k^d)(\log|\mathcal{A}| + 1)$ bits. Hence the cardinality of $\Pi_n F_n(\delta, \alpha)$ can be estimated by

$$\begin{aligned}
& \log|\Pi_n F_n(\delta, \alpha)| \\
\leq\ & \log n + 2\frac{n^d}{k_1^d(\alpha)} \\
& + n^d \left(n^{-d\left(1 - \frac{h(\mu)+\delta}{h(\mu)+\alpha}\right)} + n^{-d\left(1 - \frac{h(\mu)-\alpha}{h(\mu)+\alpha}\right)}\right) \frac{d \log n}{h(\mu) + \alpha} (\log|\mathcal{A}| + 1) \\
& + n^d \left(1 + \frac{1}{n} \sqrt[d]{\frac{d \log n}{(h(\mu) + \alpha)}}\right)^d \left(h(\mu) - \alpha^2 + \delta(\log|\mathcal{A}| + 2)\right) \\
& + n^d \left(1 - \left(1 - \frac{1}{n} \sqrt[d]{\frac{d \log n}{h(\mu) + \alpha}}\right)^d\right) (\log|\mathcal{A}| + 1) \\
\leq\ & n^d(h(\mu) - \alpha^2/2 + \delta(\log|\mathcal{A}| + 2))
\end{aligned}$$

bits, supposed n is large enough and $k_1(\alpha)$ is chosen sufficiently large. Now, since $\Pi_n F_n(\delta,\alpha) \subset C_n^\mu(\delta)$, we get

$$\mu(F_n(\delta,\alpha)) = \mu^n(\Pi_n F_n(\delta,\alpha)) \leq 2^{-n^d(\alpha^2/2-\delta(\log|\mathcal{A}|+3))}.$$

Making δ small enough from the beginning, the exponent here is negative. Hence, by the Borel–Cantelli-lemma, only finitely many of the events $x \in F_n(\delta,\alpha)$ may occur, almost surely. But we know from the first part of the theorem that $x \in E_n(\delta)$ eventually almost surely (observe that the condition $2^{k^d(h(\mu)+\alpha)} \leq n^d$ implies $\frac{k}{n} < \varepsilon(\delta)$ as supposed there, for n large enough). And we know from the Ornstein–Weiss-theorem that $\Pi_n x \in C_n^\mu(\delta)$ eventually almost surely. Hence $x \in (\Sigma \setminus F_n(\delta,\alpha)) \cap E_n(\delta) \cap [C_n^\mu(\delta)] \subset \Sigma \setminus D_n(\delta,\alpha)$ eventually almost surely. This is the assertion b) of the theorem. □

Proof. (Proof of Theorem 3.6) The proof follows the ideas of the proof of the 1-dimensional statement [11, Theorem II.3.5].

Let $\alpha < \frac{1}{4}$ and consider the sets $\mathcal{T}_k(\alpha)$ from Theorem 3.5. Consider the sets $U_{k,n}(x) := \{a \in \mathcal{T}_k(\alpha) : \tilde{\mu}_x^{k,n}(a) < 2^{-k^d(h(\mu)+2\alpha)}\}$. Since $|\mathcal{T}_k(\alpha)| \leq 2^{k^d(h(\mu)+\alpha)}$, also $\tilde{\mu}_x^{k,n}(U_{k,n}(x)) \leq 2^{-k^d\alpha}$, for any x.

Consider also the sets $V_{k,n}(x) := \{a \in \mathcal{T}_k(\alpha) : \tilde{\mu}_x^{k,n}(a) > 2^{-k^d(h(\mu)-2\alpha)}\}$. Obviously $|V_{k,n}(x)| \leq 2^{k^d(h(\mu)-2\alpha)}$. Now, by the second part of Theorem 3.5, for μ-almost every x there exists an $n_0(x)$ with $\tilde{\mu}_x^{k,n}(V_{k,n}(x)) \leq 2\alpha$ whenever $n > n_0(x)$, $k > k_1(2\alpha)$, and $2^{k^d(h(\mu)+2\alpha)} \leq n^d$.

We conclude that, for μ-a.e. x, the sets $M_{k,n}(x) := \mathcal{T}_k(\alpha) \setminus (U_{k,n}(x) \cup V_{k,n}(x))$ satisfy

$$\tilde{\mu}_x^{k,n}(M_{k,n}(x)) \;\geq\; 1 - 4\alpha,$$

where we assume that $n > n_0(x)$, $k > k_2(2\alpha)$, $2^{k^d(h(\mu)+2\alpha)} \leq n^d$, and $k_2(\alpha) \geq k_1(\alpha)$ is chosen such that $2^{-k_2(\alpha)^d\alpha} < \alpha$.

Consider now the Shannon entropy of the empirical distribution $\tilde{\mu}_x^{k,n}$,

$$\begin{aligned} H(\tilde{\mu}_x^{k,n}) &= -\sum_{a\in\Sigma^k} \tilde{\mu}_x^{k,n}(a)\log\tilde{\mu}_x^{k,n}(a) \\ &= \underbrace{-\sum_{\Sigma^k\setminus M_{k,n}} \dots}_{\Xi_{k,n}} \underbrace{-\sum_{M_{k,n}} \dots}_{\chi_{k,n}} . \end{aligned} \tag{9}$$

Let $B_{k,n}(x) := \Sigma^k \setminus M_{k,n}(x)$. For the first sum in eq. (9) an upper bound is given by[2]

$$\Xi_{k,n} \leq \tilde{\mu}_x^{k,n}(B_{k,n}(x))k^d \log|\mathcal{A}| - \tilde{\mu}_x^{k,n}(B_{k,n}(x))\log\tilde{\mu}_x^{k,n}(B_{k,n}(x)),$$

and hence $\limsup_{n\to\infty} \frac{1}{k_n^d}\Xi_{k(n),n} \leq 4\alpha\log|\mathcal{A}|$ holds μ-a.s. under the theorem's assumptions.

[2]Note that $\sum_{a\in B} p(a)\log p(a) \leq p(B)\log|B| - p(B)\log p(B)$.

For the second sum in eq. (9), note that the elements a from $M_{k,n}(x)$ satisfy

$$k^d(h(\mu)-2\alpha) \;\le\; -\log\tilde{\mu}_x^{k,n}(a) \;\le\; k^d(h(\mu)+2\alpha),$$

and thus

$$\frac{1}{k_n^d}\chi_{k,n} \;\ge \sum_{a\in M_{k,n}(x)} \tilde{\mu}_x^{k,n}(a)(h(\mu)-2\alpha) \ge (1-4\alpha)(h(\mu)-2\alpha)$$
$$\frac{1}{k_n^d}\chi_{k,n} \;\le \sum_{a\in M_{k,n}(x)} \tilde{\mu}_x^{k,n}(a)(h(\mu)+2\alpha) \le h(\mu)+2\alpha.$$

Therefore we have the following holding μ-a.s.:

$$\begin{aligned}(1-4\alpha)(h(\mu)-2\alpha) &\le \liminf_{n\to\infty}\frac{1}{k_n^d}H(\tilde{\mu}_x^{k(n),n})\\ &\le \limsup_{n\to\infty}\frac{1}{k_n^d}H(\tilde{\mu}_x^{k(n),n})\\ &\le h(\mu)+\alpha(2+4\log|\mathcal{A}|).\end{aligned}$$

Finally, note that a sequence k_n satisfying the two assumptions of the theorem for some $\alpha>0$ in fact satisfies them for any smaller α too. This completes the proof. □

P r o o f . (Proof of Theorem 3.1) When $h_0=\log|\mathcal{A}|$, the first two items are proven by choosing $\mathscr{T}_n(h_0)=\Sigma^n$. In the following we assume $h_0<\log|\mathcal{A}|$.

1. Each $x\in\Sigma$ gives rise to a family of empirical distributions $\{\tilde{\mu}_x^{k,n}\}_{k\le n}$. For each n we define the set $\mathscr{T}_n(h_0)$ as the set of elements in Σ^n having empirical k-block entropy per symbol not larger than h_0:

$$\mathscr{T}_n(h_0) := \Pi_n\left\{x\in\Sigma : H\left(\tilde{\mu}_x^{k,n}\right)\le k^d h_0\right\}. \tag{10}$$

Here we have to choose k depending on n (how exactly will be specified later).

The number of all non-overlapping empirical k-block distributions in Σ^n is upper bounded by $\left(\left(\frac{n}{k}\right)^d\right)^{|\mathcal{A}|^{k^d}}$, since $\left\lfloor\frac{n}{k}\right\rfloor^d$ is the maximal count of any particular k-block in the parsing of an element of Σ^n and $|\mathcal{A}|^{k^d}$ is the number of elements in Σ^k.

For the number of elements $x^n\in\Sigma^n$ with the same empirical distribution $(\tilde{\mu}_x^{k,n})$ we find an upper bound which depends only on the entropy of that empirical distribution: For a given n with $\lfloor n/k\rfloor = n/k$, we consider the product measure $P=(\tilde{\mu}_x^{k,n})^{\otimes(n/k)^d}$ on Σ^n: $P(y^n)=\prod_{\substack{\mathbf{r}\in k\cdot\mathbb{Z}^d\\ \Lambda_k+\mathbf{r}\subset\Lambda_n}}\tilde{\mu}_x^{k,n}(\Pi_k(\sigma_{\mathbf{r}}y))$, which yields

$$P(y^n)=\prod_{a\in\Sigma^k}\left(\tilde{\mu}_x^{k,n}(a)\right)^{(n/k)^d\tilde{\mu}_x^{k,n}(a)} = 2^{-(n/k)^d H\left(\tilde{\mu}_x^{k,n}\right)},\quad \forall y:\tilde{\mu}_y^{k,n}=\tilde{\mu}_x^{k,n}, \tag{11}$$

and thus $|\{y\in\Sigma^n:\tilde{\mu}_y^{k,n}=\tilde{\mu}_x^{k,n}\}|\le 2^{(n/k)H(\tilde{\mu}_x^{k,n})}$.

For a general $n : \lfloor n/k \rfloor \neq n/k$, the entries in the positions $\Lambda_n \backslash \Lambda_{k\cdot\lfloor n/k \rfloor}$ may be occupied arbitrarily, giving the following bound:

$$|\{y \in \Sigma^n : \tilde{\mu}_y^{k,n} = \tilde{\mu}_x^{k,n}\}| \leq 2^{\lfloor n/k \rfloor^d H(\tilde{\mu}_x^{k,n})} \cdot |\mathcal{A}|^{n^d-(n-k)^d}. \tag{12}$$

Now we are able to give an upper estimate for the number $|\mathscr{T}_n(h_0)|$ of all configurations in Λ_n which produce an empirical distribution with entropy at most $k^d h_0$:

$$\begin{aligned} |\mathscr{T}_n(h_0)| &\leq 2^{h_0 k^d \left(\frac{n}{k}\right)^d} |\mathcal{A}|^{n^d-(n-k)^d} \left(\left(\frac{n}{k}\right)^d\right)^{|\mathcal{A}|^{k^d}}, \\ \log|\mathscr{T}_n(h_0)| &\leq n^d h_0 + (n^d-(n-k)^d)\log|\mathcal{A}| + |\mathcal{A}|^{k^d} d \log\frac{n}{k}. \end{aligned}$$

Introducing the restriction $k^d \leq \frac{1}{1+\varepsilon}\log_{|\mathcal{A}|} n^d = \frac{\log n^d}{(1+\varepsilon)\log|\mathcal{A}|}$, with $\varepsilon > 0$ arbitrary, we conclude that $|\mathscr{T}_n(h_0)| \leq 2^{n^d h_0 + o(n^d)}$ (uniformly in k under the restriction). This yields $\limsup_{n\to\infty} \frac{\log|\mathscr{T}_n(h_0)|}{n^d} \leq h_0$.

2. Next we have to prove that such a sequence of sets, with $k = k(n)$ suitably specified, is asymptotically typical for all $\mu \in \mathbb{P}_{\mathrm{erg}}$ with $h(\mu) < h_0$. Given any μ with $h(\mu) < h_0$, Theorem 3.6 states that for μ-a.e. x the k-block empirical entropy $\frac{1}{k}H(\tilde{\mu}_x^{k,n})$ converges to $h(\mu)$, provided $k = k(n)$ is a sequence with $k(n) \to \infty$ and $k^d(n) \leq \frac{\log n^d}{h(\mu)+\alpha}$, where $\alpha > 0$ can be chosen arbitrarily. Since any μ satisfies $h(\mu) \leq \log|\mathcal{A}|$, choosing $k^d(n) \leq \frac{\log n^d}{(1+\varepsilon)\log|\mathcal{A}|}$ with $\varepsilon > 0$ yields assertion a) by the definition of $\mathscr{T}_n(h_0)$, eq. (10).

3. Consider a sequence $\{\mathscr{U}_n \subset \Sigma^n\}_n$ with $\liminf_{n\to\infty} \frac{1}{n^d}\log|\mathscr{U}_n| = h_1 < h_0$. One can find an ergodic μ with $h(\mu) = h_2$ and $h_1 < h_2 < h_0$. We know that μ^n is asymptotically confined to the entropy typical subsets

$$C_n^\mu(\delta) = \left\{a \in \Sigma^n : 2^{-n^d(h_2+\delta)} \leq \mu^n(\{a\}) \leq 2^{-n^d(h_2-\delta)}\right\},$$

and therefore

$$\liminf_{n\to\infty}\mu(\mathscr{U}_n) = \liminf_{n\to\infty}\mu(\mathscr{U}_n \cap C_n^\mu(\delta)) \leq \liminf_{n\to\infty}|\mathscr{U}_n| 2^{-n^d(h_2-\delta)} = \lim_{n\to\infty} 2^{n^d(h_1-h_2+\delta)}.$$

Choosing δ small enough this limit is zero. The previous analysis, together with the Borel–Cantelli-lemma, shows that on any subsequence with $\limsup_{n'\to\infty}\frac{1}{n'^d}\log|\mathscr{U}_{n'}| < h_0$, only finitely many of the events $x^{n'} \in \mathscr{U}_{n'}$ may occur, almost surely. This proves c). Combining c) and a), we get $\liminf_{n\to\infty}\frac{1}{n^d}\log|\mathscr{T}_n(h_0)| \geq h_0$. In the first part of the proof we showed $\limsup_{n\to\infty}\frac{1}{n^d}\log|\mathscr{T}_n(h_0)| \leq h_0$. Thus b) is verified as well. □

5. CONCLUSIONS

We prove multidimensional extensions of theoretical results about samplings of ergodic sources which are important in the design of universal source coding schemes. Our results provide a truly multidimensional mathematical framework for the optimal compression

of multidimensional data. We show that the set of $n \times \cdots \times n$ arrays with empirical k-block distributions of per-site entropy not larger than h_0, defined in eq. (10), is asymptotically typical for all ergodic $\mathcal{A}$-processes of entropy rate smaller than h_0, where $k = \left\lfloor \sqrt[d]{c \log_{|\mathcal{A}|} n^d} \right\rfloor$, $0 < c < 1$. In other words, for all $\mathcal{A}$-processes of entropy rate smaller than h_0, the probability of the corresponding cylinder set tends to 1 as $n \to \infty$. These sets have a log cardinality of order $n^d h_0$.

ACKNOWLEDGEMENT

We are grateful to an anonymous referee for detailed comments and valuable suggestions.

(Received November 14, 2012)

REFERENCES

[1] I. Bjelaković, T. Krüger, R. Siegmund-Schultze, and A. Szkoła: The Shannon–McMillan theorem for ergodic quantum lattice systems. Invent. Math. *155* (2004) (1), 203–222.

[2] L. Breiman: The individual ergodic theorem of information theory. Ann. Math. Statist. *28* (1957), 809–811.

[3] J. C. Kieffer: A generalized Shannon–McMillan theorem for the action of an amenable group on a probability space. Ann. Probab. *3* (1975), 6, 1031–1037.

[4] A. Lempel and J. Ziv: Compression of two-dimensional data. IEEE Trans. Inform. Theory *32* (1986), 1, 2–8.

[5] E. Lindenstrauss: Pointwise theorems for amenable groups. Invent. Math. *146* (2001), 2, 259–295.

[6] B. McMillan: The basic theorems of information theory. Ann. Math. Statist. *24* (1953), 2, 196–219.

[7] D. S. Ornstein and B. Weiss: The Shannon–McMillan–Breiman theorem for a class of amenable groups. Israel J. Math. *44* (1983), 1, 53–60.

[8] D. S. Ornstein and B. Weiss: How sampling reveals a process. Ann. Probab. *18* (1990), 3, 905–930.

[9] C. E. Shannon: A mathematical theory of communication. Bell Syst. Techn. J. *27* (1948), 1, 379–423, 623–656.

[10] K. Schmidt: A probabilistic proof of ergodic decomposition. Sankhya: Indian J. Statist, Ser. A *40* (1978), 1, 10–18.

[11] P. Shields: The Ergodic Theory of Discrete Sample Paths. Amer. Math. Soc., Graduate Stud. Math. *13* (1996).

[12] T. A. Welch: A technique for high-performance data compression. Computer *17* (1984), 6, 8–19.

[13] J. Ziv and A. Lempel: A universal algorithm for sequential data compression. IEEE Trans. Inform. Theory *23* (1977), 3, 337–343.

[14] J. Ziv and A. Lempel: Compression of individual sequences via variable-rate coding. IEEE Trans. Inform. Theory *24* (1978), 5, 530–536.

Tyll Krüger, Universität Bielefeld, Fakultät für Physik, Universitätsstraße 25, 33501 Bielefeld. Germany.
e-mail: tkrueger@physik.uni-bielefeld.de

Guido Montúfar, Pennsylvania State University, Department of Mathematics, 218 McAllister Building, University Park, PA 16802. U. S. A.
Current address: Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig. Germany.
e-mail: gfm10@psu.edu

Ruedi Seiler, Technische Universität Berlin, Institut für Mathematik MA 7-2, Straße des 17. Juni 136, 10623 Berlin. Germany.
e-mail: seiler@math.tu-berlin.de

Rainer Siegmund-Schultze, Universität Bielefeld, Fakultät für Physik, Universitätsstraße 25, 33501 Bielefeld. Germany.
e-mail: siegmund@math.tu-berlin.de