Universally Typical Sets for Ergodic Sources of Multidimensional Data

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

http://arxiv.org/abs/1105.0393

Universal Lossless Encoding Algorithms

- data modeled by stationary/ergodic random process

Universal Lossless Encoding Algorithms

- data modeled by stationary/ergodic random process
- lossless: algorithm ensurs exact reconstruction

Universal Lossless Encoding Algorithms

- data modeled by stationary/ergodic random process
- lossless: algorithm ensurs exact reconstruction
- main idea (Shannon): encode typical but small set

Universal Lossless Encoding Algorithms

- data modeled by stationary/ergodic random process
- lossless: algorithm ensurs exact reconstruction
- main idea (Shannon): encode typical but small set
- *universal*: algorithm does not involve specific properties of random process.

Universal Lossless Encoding Algorithms

- data modeled by stationary/ergodic random process
- lossless: algorithm ensurs exact reconstruction
- main idea (Shannon): encode typical but small set
- *universal*: algorithm does not involve specific properties of random process.
- main idea: construction of *universally typical sets.*

Entropy Typical Set

$$(x_1^n) \quad \text{with} \quad -\frac{1}{n}\log\mu(x_1^n) \sim h(\mu)$$

.

have the *Asymptotic Equipartition Property*:

- all (x_1^n) have the same probability $e^{-nh(\mu)}$

Entropy Typical Set

$$(x_1^n) \quad \text{with} \quad -\frac{1}{n}\log \mu(x_1^n) \sim h(\mu)$$

.

have the *Asymptotic Equipartition Property*:

- all (x_1^n) have the same probability $e^{-nh(\mu)}$
- small size $e^{nh(\mu)}$ but still

Entropy Typical Set

$$(x_1^n) \quad \text{with} \quad -\frac{1}{n}\log\mu(x_1^n) \sim h(\mu)$$

.

have the *Asymptotic Equipartition Property*:

- all (x_1^n) have the same probability $e^{-nh(\mu)}$
- small size $e^{nh(\mu)}$ but still
- nearly full measure

Entropy Typical Set

$$(x_1^n) \quad \text{with} \quad -\frac{1}{n}\log \mu(x_1^n) \sim h(\mu)$$

.

have the *Asymptotic Equipartition Property*:

- all (x_1^n) have the same probability $e^{-nh(\mu)}$
- small size $e^{nh(\mu)}$ but still
- nearly full measure
- output sequences with higher or smaler probability than $e^{-nh(\mu)}$ will rarely be observed.

Shannon-Mcmillan-Briman

$\mathbb{Z}$ -ergodic processes:

- $-\frac{1}{n}\log \mu(x_1^n) \to h(\mu)$.
    - in probability (Shannon)

Shannon-Mcmillan-Briman

$\mathbb{Z}$ -ergodic processes:

- $-\frac{1}{n}\log\mu(x_1^n) \to h(\mu)$.
    - in probability (Shannon)
    - pointwise almost surely (Mcmillan, Briman)

Shannon-Mcmillan-Briman

$\mathbb{Z}$ -ergodic processes:

- $-\frac{1}{n} \log \mu(x_1^n) \to h(\mu)$.
    - in probability (Shannon)
    - pointwise almost surely (Mcmillan, Briman)
    - amenable groups, $\mathbb{Z}^d$ (Kiefer, Ornstein and Weiss)

Notation

d-dimensional:

- $\Lambda_n := \big\{(i_1, \ldots, i_d) \in \mathbb{Z}_+^d : 0 \leq i_j \leq n-1, j \in \{1, \ldots, d\}\big\}$

Notation

d-dimensional:

- $\Lambda_n := \big\{(i_1, \ldots, i_d) \in \mathbb{Z}_+^d : 0 \le i_j \le n-1, j \in \{1, \ldots, d\}\big\}$
- $\Sigma^n := \mathcal{A}^{\Lambda_n},\ \Sigma = \mathcal{A}^{\mathbb{Z}^d}$, $\mathcal{A}$ finite alphabet

Result

Theorem (Universally-typical-sets)

For any given h_0 *with* $0 < h_0 \leq \log(|\mathcal{A}|)$ *one can construct a sequence of subsets* $\{\mathcal{T}_n(h_0) \subset \Sigma^n\}_n$ *such that for all* $\mu \in \mathbb{P}_{erg}$ *with* $h(\mu) < h_0$ *the following holds:*

1. $\lim_{n\to\infty} \mu_n\left(\mathcal{T}_n(h_0)\right) = 1,$

Result

Theorem (Universally-typical-sets)

For any given h_0 *with* $0 < h_0 \leq \log(|\mathcal{A}|)$ *one can construct a sequence of subsets* $\{\mathcal{T}_n(h_0) \subset \Sigma^n\}_n$ *such that for all* $\mu \in \mathbb{P}_{erg}$ *with* $h(\mu) < h_0$ *the following holds:*

1. $\lim_{n\to\infty} \mu_n\left(\mathcal{T}_n(h_0)\right) = 1,$
2. $\lim_{n\to\infty} \frac{\log|\mathcal{T}_n(h_0)|}{n^d} = h_0.$

Result

Theorem (Universally-typical-sets)

For any given h_0 *with* $0 < h_0 \leq \log(|\mathcal{A}|)$ *one can construct a sequence of subsets* $\{\mathcal{T}_n(h_0) \subset \Sigma^n\}_n$ *such that for all* $\mu \in \mathbb{P}_{erg}$ *with* $h(\mu) < h_0$ *the following holds:*

1. $\lim_{n\to\infty} \mu_n\left(\mathcal{T}_n(h_0)\right) = 1$,
2. $\lim_{n\to\infty} \frac{\log|\mathcal{T}_n(h_0)|}{n^d} = h_0$.
3. *optimal*

Remarks about Proof:

- for each $x \in \Sigma$ *empirical measures* $\left\{\tilde{\mu}_x^{k,n}\right\}_{k \le n}$ on $\mathcal{A}^{\Lambda_k}$.

Explain empirical measure $\left\{\tilde{\mu}_x^{k,n}\right\}$ *by a drawing (black bord)*

Remarks about Proof:

- for each $x \in \Sigma$ *empirical measures* $\left\{\tilde{\mu}_x^{k,n}\right\}_{k \leq n}$ on $\mathcal{A}^{\Lambda_k}$.

- $\mathcal{T}_n(h_0) := \Pi_n\{x \in \Sigma : \frac{1}{k^d} H(\tilde{\mu}_x^{k,n}) \leq h_0\}$

Explain empirical measure $\left\{\tilde{\mu}_x^{k,n}\right\}$ *by a drawing (black bord)*

Remarks about Proof:

- for each $x \in \Sigma$ *empirical measures* $\left\{\tilde{\mu}_x^{k,n}\right\}_{k \leq n}$ on $\mathcal{A}^{\Lambda_k}$.
- $\mathcal{T}_n(h_0) := \Pi_n\{x \in \Sigma : \frac{1}{k^d} H(\tilde{\mu}_x^{k,n}) \leq h_0\}$
- $k^d \leq \frac{1}{1+\epsilon} \log_{|\mathcal{A}|} n^d, \quad \epsilon > 0$

Explain empirical measure $\left\{\tilde{\mu}_x^{k,n}\right\}$ *by a drawing (black bord)*

Remarks about Proof:

- for each $x \in \Sigma$ *empirical measures* $\left\{\tilde{\mu}_x^{k,n}\right\}_{k \leq n}$ on $\mathcal{A}^{\Lambda_k}$.
- $\mathcal{T}_n(h_0) := \Pi_n\{x \in \Sigma : \frac{1}{k^d} H(\tilde{\mu}_x^{k,n}) \leq h_0\}$
- $k^d \leq \frac{1}{1+\epsilon} \log_{|\mathcal{A}|} n^d, \quad \epsilon > 0$
- $\limsup \frac{\log |\mathcal{T}_n(h_0)|}{n^d} \leq h_0$

Explain empirical measure $\left\{\tilde{\mu}_x^{k,n}\right\}$ *by a drawing (black bord)*

Theorem (Empirical-Entropy Theorem)

Let $\mu \in \mathbb{P}_{erg}$. *For any sequence* $\{k_n\}$ *satisfying* $k_n \xrightarrow{n\to\infty} \infty$ *and* $k_n^d = o(n^d)$ *we have*

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

Main references

- Paul C. Shields: The Ergodic Theory of Discrete Sample Paths, Graduate Studies in Mathematics, Vol.13 AMS.
- Article to appear in KYBERNETICA

Background Material

Lemma (Packing Lemma)

Consider for any fixed $0 < \delta \leq 1$ *integers* k *and* m *related through* $k \geq d \cdot m/\delta$. *Let* $C \subset \Sigma^m$ *and* $x \in \Sigma$ *with the property that* $\tilde{\mu}^{m,k}_{x,overl}(C) \geq 1 - \delta$. *Then, there exists* $\mathbf{p} \in \Lambda_m$ *such that:*
a) $\tilde{\mu}^{\mathbf{p},m,k}_x(C) \geq 1 - 2\delta$, *and also b)*
$\left\lfloor \frac{k}{m} \right\rfloor^d \tilde{\mu}^{\mathbf{p},m,k}_x(C) \geq (1-4)\delta(\left\lfloor \frac{k}{m} \right\rfloor + 2)^d$.

Theorem

Given any $\mu \in \mathbb{P}_{erg}$ *and any* $\alpha \in (0, \frac{1}{2})$ *we have the following:*

- *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 *such that* $\frac{k}{n} < \varepsilon$ *for some* $\varepsilon = \varepsilon(\alpha) > 0$ *and* n *larger than some* $n_0(x)$*.*
- *(optimality)*

Definition (Entropy-typical-sets)

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

$$C_m(\delta) := \left\{ x \in \Sigma^m : 2^{-m^d(h(\mu)+\delta)} \leq \mu^m(\{x\}) \leq 2^{-m^d(h(\mu)-\delta)} \right\}. \tag{1}$$

We will use these sets as basic sets for the typical-sampling-sets defined below.

Definition (Typical-sampling-sets)

Consider some μ. Consider some $\delta < \frac{1}{2}$. For $k \geq m$, we define a *typical-sampling-set* $\mathcal{T}_k(\delta, m)$ 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 $C_m = C_m(\delta)$ contribute at least a $(1-\delta)$-fraction to the (slightly modified) number of partition elements in that regular m-block partition, more precisely, they occupy at least a $(1-\delta)$-fraction of all sites in Λ_k

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