

Article

Not peer-reviewed version

Discretization of Dynamical Systems Based on Observations

[Eugene Kagan](#) * and [Alexander Novoselsky](#)

Posted Date: 12 December 2024

doi: 10.20944/preprints202409.0583.v2

Keywords: dynamical system; ϵ -entropy; ϵ -capacity; ϵ -information



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Discretization of Dynamical Systems Based on Observations

Eugene Kagan ^{1,*} and Alexander Novoselsky ²

¹ Ariel University, Kiryat ha-Mada, Ariel 4070000, Israel; evganyk@ariel.ac.il

² Log-On Software, 3 ha-Hilazon St., Ramat Gan 5252267, Israel

* Correspondence: kagan.eugene@gmail.com

Abstract. Based on the observations of dynamical system, we define partition ξ_ε , which represents dynamical system as good as possible in the sense of its entropy. The suggested method utilizes ε -entropy, ε -capacity, and the introduced ε -information. The resulting algorithm is also useful for defining bin lengths of histograms, especially for multimodal distributions.

Keywords: Dynamical system; ε -entropy; ε -capacity; ε -information

MSC: 28D20; 37A99

1. Introduction

Let $\mathfrak{S} = (M, \mathcal{M}, \mu, \nu, T)$ be dynamical system on a compact subset M of metric space R with metric ν and additive Lebesgue measure μ , where \mathcal{M} is a σ -algebra of the subsets of R and $T: M \rightarrow M$ is an endomorphism.

Assume that in the σ -algebra \mathcal{M} are distinguished m sets $X_1, X_2, \dots, X_m \subset M$ which represent observations of the system and assume that these observations form a partition $\xi = \{X_i\}$ of the set M .

Then, using observations' partition ξ it is required to define a real number $\varepsilon > 0$ and partition $\xi_\varepsilon = \{X_i \mid i \in \mathbb{N}\}$ of the set M such that the diameters of the sets $X_i \in \xi_\varepsilon$ are $d(X_i) = \sup_{x, y \in X_i} \nu(x, y) \leq 2\varepsilon$, which represents the system "as good as possible".

Before formulating strict criterion, let us clarify the problem by the following examples.

Assume that $M = [x_0, x_m] \in \mathbb{R}$ is an interval of real numbers and assume that observations of the interval are represented by the partition $\xi = \{X_i\}$, which includes intervals $X_1 = [x_0, x_1]$, $X_2 = (x_1, x_2]$, $X_3 = (x_2, x_3]$, ..., $X_m = (x_{m-1}, x_m]$ such that $x_j \in M$, $j = 0, 1, 2, \dots, m$, are drawn with respect to certain unknown distribution. Using the numbers x_j , $j = 0, 1, 2, \dots, m$, it is needed to plot a histogram, which represents this distribution.

This problem requires definition of the number of bins or the bin length such that the histogram "as good as possible" represents the distribution of the observed values. Despite existence of several formulas for calculating bin lengths, there is no strictly proven criteria for estimating goodness of the histogram.

From the other point of view, the problem can be considered in the terms of discretization of stochastic processes [5]. In this case, given observed realization of the process it is required to define a sequence t_i , $i = 0, 1, 2, \dots$, of times, $2\varepsilon = t_{i+1} - t_i$, $t_i = i\Delta$, such that the intervals $\Delta = 2\varepsilon$ between the values $x_i = x(t_i)$ "as good as possible" represent the process.

To formulate the criterion of consistency between the observations partition ξ and the partition ξ_ε , denote $\xi_\varepsilon^n = \bigvee_{i=0}^n T^i \xi_\varepsilon$ and $\xi^n = \bigvee_{i=0}^n T^i \xi$, where \bigvee stands for multiplication of the partitions.

Given observations' partition ξ , we say that partition ξ_ε represents dynamical system \mathfrak{S} if a distance $dist(\xi_\varepsilon^n, \xi^n)$ between partitions ξ_ε^n and ξ^n as a function of ε is bounded that is

$$a \leq dist(\xi_\varepsilon^n, \xi^n) \leq b,$$

for some constant numbers a and b , $a \leq b$, and that partition ξ_ε represents dynamical system \mathfrak{S} "as good as possible" if

$$dist(\xi_\varepsilon^n, \xi^n) = const.$$

In the paper, we suggest a method for defining partition ξ_ε . We assume that M is a compact set with a metric $\nu: M \rightarrow \mathbb{R}$ and together with the Kolmogorov-Sinai entropy $h(T)$ of dynamical system we consider the Kolmogorov ε -entropy and ε -capacity [6]. Then, we define a diameter 2ε of the sets X_i that results in the required partition ξ_ε .

The suggested method is also useful for definition of the bin lengths of histograms; in the paper, we apply the method to different datasets and compare it with the bin lengths obtained by the known methods.

2. Methods

We follow the line of the Schwarz information criterion [10] and apply the Kolmogorov ε -entropy and ε -capacity [6] (see also [2, 15]). Here we recall the required definitions and facts which are used in the solution of the problem.

Let $\mathfrak{S} = (M, \mathcal{M}, \mu, \nu, T)$ be a dynamical system. We assume that M is a compact subset of metric space R with metric ν and additive Lebesgue measure μ such that $\mu(M) = 1$; \mathcal{M} is a σ -algebra of the subsets of R and $T: M \rightarrow M$ is an endomorphism.

Let $\beta = \{B_i\}$, $B_i \in \mathcal{M}$, $\cup_i B_i = M$, $B_i \cap B_j = \emptyset$ for $i \neq j$, $i, j \in \mathbb{N}$, be a partition of the set M .

Definition 1 [12, 13]. A value

$$\mathcal{H}_\mu(\beta) = -\sum_i \mu(B_i) \log_2 \mu(B_i)$$

is called entropy of the partition β .

Let $\beta = \{B_i | B_i \in \mathcal{M}\}$ and $\gamma = \{C_j | C_j \in \mathcal{M}\}$ be two partitions of the set M .

Definition 2 [12, 13]. A value

$$\mathcal{H}_\mu(\beta | \gamma) = -\sum_j \mu(C_j) \sum_i \log_2 \mu(B_i | C_j)$$

is called conditional entropy of the partition β relatively to the partition γ .

Definition 3 [9, 12, 13]. A value

$$dist(\beta, \gamma) = \mathcal{H}_\mu(\beta | \gamma) + \mathcal{H}_\mu(\gamma | \beta)$$

is called Rokhlin distance between the partitions β and γ .

For any endomorphism T and any partition β , denote by $T\beta$ a partition of M to the sets $\{TB_i\}$, that is $T\beta = \{TB_i\}$.

Let $\beta \vee \gamma = \{B_i \cap C_j | B_i \in \beta, C_j \in \gamma, i, j \in \mathbb{N}\}$ be multiplication of the partitions $\beta = \{B_i | B_i \in \mathcal{M}\}$ and $\gamma = \{C_j | C_j \in \mathcal{M}\}$.

Definition 4 [12, 13]. A value

$$h(\beta, T) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{H}_\mu(\beta \vee T\beta \vee T^2\beta \vee \dots \vee T^n\beta)$$

is called entropy per symbol of partition or entropy per unit of time.

Definition 5 [12, 13]. A value

$$h(T) = \sup_{\beta} h(\beta, T),$$

where supremum is taken over all finite or countable measurable partitions β of M such that $\mathcal{H}_\mu(\beta) < \infty$, is called entropy of dynamical system.

Entropy $h(T)$ is an invariant of dynamical system, but direct calculation of the entropy $h(T)$ is possible only for certain endomorphisms; for examples see [12, chapter 15] and [13, chapter 7].

Theorem 1 [1, 12, 13]. If T is invertible and $\mathcal{M} = \bigvee_{n=-\infty}^{\infty} T^n \xi^*$, then $h(\xi^*, T) = h(T)$.

Corollary 1 [1]. If $\mathcal{M} = \bigvee_{n=0}^{\infty} T^n \xi^*$, then $h(\xi^*, T) = h(T)$.

Let $\varepsilon > 0$ be a real number.

Set $\alpha = \{A: A \subset R\}$ is called ε -covering of the set M , if $M \subseteq \bigcup_{A \in \alpha} A$ and diameter of any $A \in \alpha$ is not greater than 2ε .

Set M is said to be ε -distinguishable, if any two of its distinct points are located at distance greater than ε .

Lemma 1 [6]. Given a compact set M , for any $\varepsilon > 0$ there exists a finite ε -covering of M . Also, for any $\varepsilon > 0$ any ε -distinguishable set $M \subset R$ is finite.

Denote by $\mathcal{N}_\varepsilon(M)$ a minimal number of sets in ε -covering α of the set M , and by $\mathcal{M}_\varepsilon(M)$ a maximal number of points in an ε -distinguishable subset of the set M .

Definition 6 [6]. A value

$$\mathcal{H}_\varepsilon(M) = \log_2 \mathcal{N}_\varepsilon(M)$$

is called ε -entropy of the set M and a value

$$\mathcal{E}_\varepsilon(M) = \log_2 \mathcal{M}_\varepsilon(M)$$

is called ε -capacity of the set M .

These values are interpreted as follows: ε -entropy $\mathcal{H}_\varepsilon(M)$ is a minimal number of bits required to transmit the set M with the precision ε , and ε -capacity $\mathcal{E}_\varepsilon(M)$ is a maximal number of bits, which can be memorized by M with the precision ε .

We will use the following property of ε -entropy $\mathcal{H}_\varepsilon(M)$ and ε -capacity $\mathcal{E}_\varepsilon(M)$.

Theorem 2 [6]. Given a compact set M , both ε -entropy and ε -capacity are non-increasing functions of ε .

Finally, we will use the following property of the entropy of partitions.

Let $\beta = \{B: B \subset M\}$ and $\gamma = \{C: C \subset M\}$ be two partitions of the set M . If each set $C \in \gamma$ is a subset of some set $B \in \beta$, then it is said that partition γ is a refinement of the partition β , which is written as $\gamma \succcurlyeq \beta$.

Lemma 2 [12]. If $\beta \preccurlyeq \gamma$, then $\mathcal{H}_\mu(\beta) \leq \mathcal{H}_\mu(\gamma)$.

Let $\beta \vee \gamma$ be multiplication of the partitions β and γ . Since each set $D \in \beta \vee \gamma$ is a subset of some set $B \in \beta$ and of some set $C \in \gamma$, partition $\beta \vee \gamma$ is a refinement of both β and γ , that is $\beta \vee \gamma \succcurlyeq \beta$ and $\beta \vee \gamma \succcurlyeq \gamma$.

Thus, following lemma 2, $\mathcal{H}_\mu(\beta) \leq \mathcal{H}_\mu(\beta \vee \gamma)$ and $\mathcal{H}_\mu(\gamma) \leq \mathcal{H}_\mu(\beta \vee \gamma)$.

For other properties of the entropy \mathcal{H}_μ and its application for analysis of dynamical systems see the paper [9] and the books [7, 13, 15].

3. Suggested Solution

Let M is a compact subset of metric space R with metric ν and additive Lebesgue measure μ such that $\mu(M) = 1$; \mathcal{M} is a σ -algebra of the subsets of R .

Let $\beta = \{B_i | B_i \subset M\}$, $\cup_i B_i = M$, $B_i \cap B_j = \emptyset$ for $i \neq j$, $i, j \in \mathbb{N}$, be a partition of M . If diameter $d(B_i) = \sup_{x, y \in B_i} \nu(x, y)$ of each $B_i \in \beta$ is not greater than 2ε , then partition β is an ε -covering and is called ε -partition.

Lemma 3. Let β be ε -partition of M . If diameter $d(M) = \sup_{x, y \in M} \nu(x, y) = 1$ and measure $\mu(B) = \sup_{x, y \in B} \nu(x, y)$, $B \in \beta$, then $\mathcal{H}_\mu(\beta) \geq \mathcal{H}_\varepsilon(M)$.

Proof. Consider entropy

$$\mathcal{H}_\varepsilon(M) = \log_2 \mathcal{N}_\varepsilon(M) = -\sum_{\mathcal{N}_\varepsilon(M)} \frac{1}{\mathcal{N}_\varepsilon(M)} \log_2 \frac{1}{\mathcal{N}_\varepsilon(M)}.$$

Given ε , a minimal number $\mathcal{N}_\varepsilon(M)$ of sets in β is reached if diameters $d(B)$ of its elements $B \in \beta$ are maximal that is $d(B) = 2\varepsilon$. That is possible either if $\frac{d(M)}{d(B)} = \frac{1}{2\varepsilon}$ subsets from β completely cover M , or if $\left\lceil \frac{d(M)}{d(B)} \right\rceil$ subsets cover a part M' of M such that $d(M \setminus M') < 2\varepsilon$ and to cover this part it is required one additional subset, where $\lceil x \rceil$ denotes maximal integer number smaller than or equal to x . In this case, the number of elements in β is $\left\lceil \frac{d(M)}{d(B)} \right\rceil + 1$.

Then, if $\frac{1}{2\varepsilon}$ is integer,

$$\mathcal{N}_\varepsilon(M) = \frac{d(M)}{d(B)} = \frac{1}{2\varepsilon};$$

otherwise

$$\mathcal{N}_\varepsilon(M) = \left\lceil \frac{d(M)}{d(B)} \right\rceil + 1 = \left\lceil \frac{1}{2\varepsilon} \right\rceil + 1.$$

Substituting these values into the definition of entropy, in the first case one obtains

$$\mathcal{H}_\varepsilon(M) = -\sum_{i=1}^{\lfloor 1/(2\varepsilon) \rfloor} \frac{1}{1/(2\varepsilon)} \log_2 \frac{1}{1/(2\varepsilon)}$$

and in the second case –

$$\begin{aligned} \mathcal{H}_\varepsilon(M) &= -\sum_{i=1}^{\lfloor 1/(2\varepsilon) \rfloor + 1} \frac{1}{\lfloor 1/(2\varepsilon) \rfloor + 1} \log_2 \frac{1}{\lfloor 1/(2\varepsilon) \rfloor + 1} = \\ &= -\sum_{i=1}^{\lfloor 1/(2\varepsilon) \rfloor} \frac{1}{\lfloor 1/(2\varepsilon) \rfloor + 1} \log_2 \frac{1}{\lfloor 1/(2\varepsilon) \rfloor + 1} - \frac{1}{\lfloor 1/(2\varepsilon) \rfloor + 1} \log_2 \frac{1}{\lfloor 1/(2\varepsilon) \rfloor + 1}. \end{aligned}$$

Now consider entropy

$$\mathcal{H}_\mu(\beta) = -\sum_{B \in \beta} \mu(B) \log_2 \mu(B).$$

If number $\frac{1}{2\varepsilon}$ is integer, then

$$\mathcal{H}_\mu(\beta) = -\sum_{i=1}^{\lfloor 1/(2\varepsilon) \rfloor} 2\varepsilon \times \log_2(2\varepsilon) = -\sum_{i=1}^{\lfloor 1/(2\varepsilon) \rfloor} \frac{1}{1/(2\varepsilon)} \log_2 \frac{1}{1/(2\varepsilon)}$$

otherwise

$$\begin{aligned} \mathcal{H}_\mu(\beta) &= -\sum_{i=1}^{\lfloor 1/(2\varepsilon) \rfloor + 1} 2\varepsilon \times \log_2(2\varepsilon) = \\ &= -\sum_{i=1}^{\lfloor 1/(2\varepsilon) \rfloor} \frac{1}{1/(2\varepsilon)} \log_2 \frac{1}{1/(2\varepsilon)} - \frac{1}{1/(2\varepsilon)} \log_2 \frac{1}{1/(2\varepsilon)}. \end{aligned}$$

Thus, if number $\frac{1}{2\varepsilon}$ is integer, then $\mathcal{H}_\varepsilon(M) = \mathcal{H}_\mu(\beta)$.

Consider the case of non-integer $\frac{1}{2\varepsilon}$. For any ε such that number $\frac{1}{2\varepsilon}$ is not integer, holds

$$\frac{1}{\lfloor 1/(2\varepsilon) \rfloor + 1} < \frac{1}{1/(2\varepsilon)}.$$

Thus, because of monotonicity of the \log function, $\mathcal{H}_\mu(\beta) \geq \mathcal{H}_\varepsilon(M)$. ■

Note that lemma 3 is also true for a measure μ and a metric ν such that $\mu(M) = d(M)$.

From the proof of lemma 3 follows the next fact.

Corollary 2. $\mathcal{H}_\mu(\beta) = \mathcal{H}_\varepsilon(M) + c(\varepsilon)$, where $c(\varepsilon) = -2\varepsilon \log_2(2\varepsilon)$ monotonously increases with ε up to a single maximum and then monotonously decreases.

In addition, let us consider ε -entropy of multiplication of partitions.

Let $\beta = \{B_i \mid B_i \subset M\}$ and $\gamma = \{C_j \mid C_j \subset M\}$ be two ε -partitions of the set M and let $\beta \vee \gamma = \{D_k \mid D_k \subset M\}$ be multiplication of the partitions β and γ .

Lemma 4. Let $d(B_i) \leq 2\varepsilon_B$, $B_i \in \beta$, and $d(C_j) \leq 2\varepsilon_C$, $C_j \in \gamma$, and let $\varepsilon = \max\{\varepsilon_B, \varepsilon_C\}$. Then $\beta \vee \gamma$ is ε -partition with $d(D_k) \leq 2\varepsilon_D \leq \max\{\varepsilon_B, \varepsilon_C\}$, $D_k \in \beta \vee \gamma$, and

$$\mathcal{H}_{\varepsilon_D}(M) \geq \mathcal{H}_{\min\{\varepsilon_B, \varepsilon_C\}}(M).$$

Proof. Since $\beta \vee \gamma \succcurlyeq \beta$, holds $\mathcal{N}_{\varepsilon_D}(M) \geq \mathcal{N}_{\varepsilon_B}(M)$ and $\mathcal{H}_{\varepsilon_D}(M) \geq \mathcal{H}_{\varepsilon_B}(M)$; and since $\beta \vee \gamma \succcurlyeq \gamma$, holds $\mathcal{N}_{\varepsilon_D}(M) \geq \mathcal{N}_{\varepsilon_C}(M)$ and $\mathcal{H}_{\varepsilon_D}(M) \geq \mathcal{H}_{\varepsilon_C}(M)$.

Assume that $\varepsilon_B \geq \varepsilon_C$. Then, since $\mathcal{N}_{\varepsilon_B}(M) \leq \mathcal{N}_{\varepsilon_C}(M)$, holds $\mathcal{H}_{\varepsilon_B}(M) \leq \mathcal{H}_{\varepsilon_C}(M)$, and consequently $\mathcal{H}_{\varepsilon_D}(M) \geq \mathcal{H}_{\varepsilon_C}(M) \geq \mathcal{H}_{\varepsilon_B}(M)$.

On the other hand, if $\varepsilon_C \geq \varepsilon_B$, then $\mathcal{N}_{\varepsilon_C}(M) \leq \mathcal{N}_{\varepsilon_B}(M)$ and $\mathcal{H}_{\varepsilon_C}(M) \leq \mathcal{H}_{\varepsilon_B}(M)$. Thus $\mathcal{H}_{\varepsilon_D}(M) \geq \mathcal{H}_{\varepsilon_B}(M) \geq \mathcal{H}_{\varepsilon_C}(M)$. ■

The suggested method is based on the concept of ε -information which is defined as follows.

Let $\xi = \{X_i \mid i = 1, 2, \dots, m\}$ be a partition of the set M obtained by observations and let $\varepsilon_m > 0$ be a radius that provides maximal ε -entropy $\mathcal{H}_{\varepsilon_m}(M)$ with respect to ξ .

Radius ε_m depends on the diameter of the set M and the structure of the partition ξ and cannot be specified directly. Nevertheless, we can prove the following fact.

Lemma 5.

$$\mathcal{H}_{\varepsilon_m}(M) = \log_2 m.$$

Proof. Denote by $\xi_{\varepsilon_m} = \{Y_j \mid j = 1, 2, \dots, m_{\varepsilon_m}\}$ partition of M such that $d(Y_j) = \sup_{x, y \in Y} \nu(x, y) =$

$2\varepsilon_m$ for each $Y_j \in \xi_{\varepsilon_m}$. In other words, partition ξ_{ε_m} splits the set M to the subsets with equal diameters $2\varepsilon_m$.

Then ε_m is a radius such that an average number of bits in the description of the elements of ξ using the elements of ξ_{ε_m} is maximal. Formally it means that for the radius ε_m holds

$$\frac{1}{n} \mathcal{H}_\mu(\xi_{\varepsilon_m} \vee \xi) \rightarrow \max,$$

where $n = |\xi_{\varepsilon_m} \vee \xi|$ is a number of subsets in the partition $\xi_{\varepsilon_m} \vee \xi$.

Note that the number of elements in the partition $\xi_{\varepsilon_m} \vee \xi$ is $n \geq \max(m, m_{\varepsilon_m})$, and that the function $f(x) = \frac{1}{x} \log_2 x$ has a single maximum in the point $x = e$.

Hence, given partition ξ of the size m , maximum of the entropy $\frac{1}{n} \mathcal{H}_\mu(\xi_{\varepsilon_m} \vee \xi)$ is reached for partition ξ_{ε_m} of the size $m_{\varepsilon_m} = m$. Diameters of the elements $Y_j \in \xi_{\varepsilon_m}$ of this partition are

$$d(Y_j) = 2\varepsilon_m = \frac{d(M)}{m}, \quad j = 1, 2, \dots, m.$$

Recall that the measure of the set M is $\mu(M) = 1$. Since partition ξ_{ε_m} includes m elements with equal diameters, the measures of these elements are $\mu(Y_j) = \frac{1}{m}$. Thus, entropy of the partition ξ_{ε_m} is

$$\mathcal{H}_\mu(\xi_{\varepsilon_m}) = -\sum_{j=1}^m \mu(Y_j) \log_2 \mu(Y_j) = -\sum_{j=1}^m \frac{1}{m} \log_2 \frac{1}{m} = \log_2 m.$$

Finally, since partition ξ_{ε_m} includes the elements with equal diameters $2\varepsilon_m$, this partition is ε -partition of M and for given ε_m its ε -entropy is

$$\mathcal{H}_{\varepsilon_m}(M) = \log_2 m,$$

which is required. ■

Let $\beta = \{B: B \subset M\}$ be a partition of the set M . Similar to definition 6, denote by $\mathcal{N}_\varepsilon(\beta)$ a number of sets $B \in \beta$ such that $d(B) \leq 2\varepsilon$.

Definition 7. A value

$$\mathcal{H}_\varepsilon(\beta) = \log_2 \mathcal{N}_\varepsilon(\beta)$$

is called ε -entropy of the partition β .

Let $\mathcal{H}_{\varepsilon_m}(M)$ be maximal ε -entropy of the set M with respect to the observations $\xi = \{X_i | i = 1, 2, \dots, m\}$ and let $\varepsilon > 0$ be a radius; recall that radius ε represents certain precision of coding the set M (see definition 6 and remarks after it). Then

$$\mathcal{H}_{\varepsilon_m}(M) - \mathcal{H}_\varepsilon(M)$$

is the difference between minimal number of bits required to transmit the set M with the precision ε_m provided by observations and minimal number of bits required to transmit the set M with the desired precision ε .

This difference can be negative which indicates that the observations do not allow transmitting the set M with the desired precision.

Similar to the proof of lemma 5, denote by $\xi_\varepsilon = \{Y | Y \subset M\}$ partition of M such that $d(Y) = \sup_{x,y \in Y} v(x,y) = 2\varepsilon$ for each $Y \in \xi_\varepsilon$. Then

$$\mathcal{H}_{\varepsilon_m}(M) - \mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$$

is the difference between minimal number of bits required to transmit the set M with the precision ε_m provided by observations and the number of bits required to transmit the observations ξ of the set M using the partition ξ_ε .

The sum of these differences results in the following definition.

Definition 8. A value

$$\mathcal{J}_\varepsilon(M) = 2\mathcal{H}_{\varepsilon_m}(M) - \mathcal{H}_\varepsilon(M) - \mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$$

is called ε -information of the set M .

In the formula of ε -information $\mathcal{J}_\varepsilon(M)$, the value $\mathcal{H}_{\varepsilon_m}(M)$ represents minimal number of bits required to transmit the set M with maximal precision allowed by the observations ξ , the second term $\mathcal{H}_\varepsilon(M)$ represents minimal number of bits required to transmit the set M with precision ε , and the last term $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$ represents the number of bits required to transmit the results of observations of the M measured using ξ_ε .

Thus, the value $\mathcal{J}_\varepsilon(M)$ is the number of bits remaining after the transmitting the observed part of the set M with the precision ε .

Theorem 3. Let ξ be the observations' partition of the size m . Then, ε -information $\mathcal{J}_\varepsilon(M)$ as a function of ε increases and its upper bound is $\log_2 m$.

Proof. In the definition of ε -information the first term $2\mathcal{H}_{\varepsilon_m}(M)$ is constant and the second term $\mathcal{H}_\varepsilon(M)$, by theorem 2, does not increase with increasing ε . Thus, it remains to prove that the last term $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$ decreases with increasing ε .

Since diameter of each $Z \in \xi \vee \xi_\varepsilon$ is $d(Z) \leq 2\varepsilon$, the number $\mathcal{N}_\varepsilon(\xi \vee \xi_\varepsilon)$ is equivalent to the number $n = |\xi \vee \xi_\varepsilon|$ of elements in the partition $\xi \vee \xi_\varepsilon$, and so ε -entropy is

$$\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon) = \log_2 n.$$

The number n of elements in the partition $\xi \vee \xi_\varepsilon$ decreases (not necessarily monotonously) with increasing ε down to the number $m = |\xi|$ of elements in the partition ξ . Hence, entropy $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$ also decreases down to the value $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon) = \log_2 m$, which results in increasing of ε -information $\mathcal{J}_\varepsilon(M)$ at maximum up to the value $2 \log_2 m - \log_2 m = \log_2 m$. ■

Let us return to the initial problem.

Let $\mathfrak{S} = (M, \mathcal{M}, \mu, \nu, T)$ be a dynamical system, let $\xi = \{X_i \mid X_i \subset \mathcal{M}, i = 1, 2, \dots, m\}$ be an observations' partition, and $\xi_\varepsilon^n = \bigvee_{i=0}^n T^i \xi_\varepsilon$ and $\xi^n = \bigvee_{i=0}^n T^i \xi$.

Theorem 4. *If T is invertible and*

$$|\mathcal{J}_\varepsilon(M) - \mathcal{E}_\varepsilon(M)| = 0,$$

then Rokhlin distance

$$\text{dist}(\xi_\varepsilon^n, \xi^n) = \mathcal{H}_\mu(\xi_\varepsilon^n | \xi^n) + \mathcal{H}_\mu(\xi^n | \xi_\varepsilon^n)$$

as a function of ε is bounded.

Before proving the theorem, let us recall the following facts. Let $\beta = \{B_i\}$ and $\gamma = \{C_j\}$ be partitions of the set M .

Lemma 6 [12, 13].

$$\mathcal{H}_\mu(\beta \vee \gamma) = \mathcal{H}_\mu(\beta | \gamma) + \mathcal{H}_\mu(\gamma).$$

Lemma 7 [12]. *If endomorphism T is invertible, then*

$$\mathcal{H}_\mu(T\beta | T\gamma) = \mathcal{H}_\mu(\beta | \gamma).$$

Proof of theorem 4. Following lemma 7,

$$\text{dist}(\xi_\varepsilon^n, \xi^n) = \text{dist}(\xi_\varepsilon, \xi) = \mathcal{H}_\mu(\xi_\varepsilon | \xi) + \mathcal{H}_\mu(\xi | \xi_\varepsilon).$$

Applying lemma 6 for each term, we obtain

$$\text{dist}(\xi_\varepsilon, \xi) = 2\mathcal{H}_\mu(\xi_\varepsilon \vee \xi) - \mathcal{H}_\mu(\xi_\varepsilon) - \mathcal{H}_\mu(\xi).$$

The last term $\mathcal{H}_\mu(\xi)$ does not depend on ε ; hence we are interested in the difference

$$2\mathcal{H}_\mu(\xi_\varepsilon \vee \xi) - \mathcal{H}_\mu(\xi_\varepsilon).$$

Consider ε -information $\mathcal{J}_\varepsilon(M)$. From corollary 2 it follows that

$$\mathcal{H}_{\varepsilon_m}(M) = \mathcal{H}_\mu(\xi_{\varepsilon_m}) + c_1,$$

$$\mathcal{H}_\varepsilon(M) = \mathcal{H}_\mu(\xi_\varepsilon) + c_2(\varepsilon),$$

$$\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon) = \mathcal{H}_\mu(\xi \vee \xi_\varepsilon) + c_3(\varepsilon),$$

where c_1 is constant and c_2 and c_3 monotonously increase with ε up to a single maximum and then monotonously decrease. Then

$$\mathcal{J}_\varepsilon(M) = 2\mathcal{H}_\mu(\xi_{\varepsilon_m}) - \mathcal{H}_\mu(\xi_\varepsilon) - \mathcal{H}_\mu(\xi \vee \xi_\varepsilon) + c,$$

where $c(\varepsilon) = 2c_1 - c_2(\varepsilon) - c_3(\varepsilon)$ monotonously decreases down to a single minimum and then monotonously increases.

Similarly, for ε -capacity holds

$$\mathcal{E}_\varepsilon(M) = \mathcal{H}_\mu(\xi_\varepsilon) + c_4(\varepsilon),$$

where $c_4(\varepsilon) = c_2(\varepsilon) + \text{const}$ monotonously increases with ε up to a single maximum and then monotonously decreases.

Zero difference between ε -information and ε -capacity leads to

$$2\mathcal{H}_\mu(\xi_{\varepsilon_m}) + c_5(\varepsilon) = \mathcal{H}_\mu(\xi \vee \xi_\varepsilon) + 2\mathcal{H}_\mu(\xi_\varepsilon),$$

where $c_5(\varepsilon) = c(\varepsilon) - c_2(\varepsilon) - \text{const}$ is bounded.

Since $2\mathcal{H}_\mu(\xi_{\varepsilon_m})$ does not depend on ε ,

$$\mathcal{H}_\mu(\xi \vee \xi_\varepsilon) + 2\mathcal{H}_\mu(\xi_\varepsilon) = c_5(\varepsilon) + \text{const}.$$

Consequently, the difference

$$2\mathcal{H}_\mu(\xi_\varepsilon \vee \xi) - \mathcal{H}_\mu(\xi_\varepsilon)$$

is bounded and theorem 4 is proven. ■

This theorem motivates formulation of the problem of finding a value $\Delta = 2\varepsilon$ in the terms of ε -information and ε -capacity of the set M .

Problem. Find radius ε such that

$$|\mathcal{J}_\varepsilon(M) - \mathcal{E}_\varepsilon(M)| \rightarrow \min.$$

Calculation of ε -information and ε -capacity is possible only for certain sets M . Example of such calculations for evenly distributed observations over an interval of real numbers are presented in section 4. However, an algorithmic solution of the problem is rather simple.

Below we present an algorithm of finding ε for the interval $M = [x_0, x_m] \subset R$ of real numbers, which can be directly extended to the subset of any lattice with appropriate σ -algebra.

Algorithm 1. Computing radius ε

Input: Set $M = [x_0, x_m] \subset R$;

observations' partition $\xi = \{X_1, X_2, \dots, X_m\}$, $X_{i+1} = [x_i, x_{i+1}]$, $x_i < x_{i+1}$,
 $i = 0, 1, 2, \dots, m - 1$;

step $s > 0$.

Output: Radius ε .

1. Calculate $\varepsilon_m = (x_m - x_0)/(2m)$.
2. Calculate $\mathcal{H}_{\varepsilon_m}(M) = \log_2 m$.
3. For $\varepsilon = \varepsilon_m$ to $(x_m - x_0)/2$ with step s do:
 4. Calculate $\mathcal{H}_\varepsilon(M) = \log_2[(x_m - x_0)/2\varepsilon]$.
 5. Create partition $\xi_\varepsilon = \{Y_1, Y_2, \dots, Y_{m_\varepsilon}\}$ of M such that
 $Y_{j+1} = [y_j, y_{j+1}]$, $y_j < y_{j+1}$, $y_0 = x_0$, $y_{m_\varepsilon} = x_m$, and $y_{j+1} - y_j = 2\varepsilon$, $j = 0, 1, 2, \dots, m_\varepsilon - 1$.
 6. Compute $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon) = \text{eps_entropy}(\xi, \xi_\varepsilon)$.
 7. Calculate $\mathcal{J}_\varepsilon(M) = 2\mathcal{H}_{\varepsilon_m}(M) - \mathcal{H}_\varepsilon(M) - \mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$.
 8. Compute $\mathcal{E}_\varepsilon(M) = \text{eps_capacity}(\xi, \varepsilon)$.
 9. If $\mathcal{J}_\varepsilon(M) > \mathcal{E}_\varepsilon(M)$ then
 10. Break
 11. End if.
 12. End for.
 13. Return ε .

The algorithm includes two functions, $\text{eps_entropy}(\xi, \xi_\varepsilon)$ and $\text{eps_capacity}(\xi, \varepsilon)$ which are defined as follows.

Function 1. $\text{eps_entropy}(\xi, \xi_\varepsilon)$

Input: Partition $\xi = \{X_1, X_2, \dots, X_m\}$, $X_{i+1} = [x_i, x_{i+1}]$, $x_i < x_{i+1}$,
 $i = 0, 1, 2, \dots, m - 1$.

Partition $\xi_\varepsilon = \{Y_1, Y_2, \dots, Y_{m_\varepsilon}\}$, $Y_{j+1} = [y_j, y_{j+1}]$, $y_j < y_{j+1}$,
 $j = 0, 1, 2, \dots, m_\varepsilon - 1$,

$y_0 = x_0$, $y_{m_\varepsilon} = x_m$, and $y_{j+1} - y_j = 2\varepsilon$, $j = 0, 1, 2, \dots, m_\varepsilon - 2$.

Output: ε -entropy $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$ of the partition $\xi \vee \xi_\varepsilon$.

1. From the partitions ξ and ξ_ε create, respectively, the sets of points
 $\mathcal{X} = \{x_0, x_1, x_2, \dots, x_m\}$ and $\mathcal{Y} = \{y_0, y_1, y_2, \dots, y_m\}$.
 2. Join the sets \mathcal{X} and \mathcal{Y} : $\mathcal{X}_{\text{joint}} = \mathcal{X} \cup \mathcal{Y}$.
 3. Find a number $N(\mathcal{X}_{\text{joint}})$ of elements in the set $\mathcal{X}_{\text{joint}}$.
 4. Set $\mathcal{N}_\varepsilon(\mathcal{X}_{\text{joint}}) = N(\mathcal{X}_{\text{joint}}) - 1$.
 5. Set $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon) = \log_2 \mathcal{N}_\varepsilon(\mathcal{X}_{\text{joint}})$.
 6. Return $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$.
-

Function eps_entropy can be implemented using concatenation of the sets with further removing of the doubling elements.

Function 2. $\text{eps_capacity}(\xi, \varepsilon)$

Input: Partition $\xi = \{X_1, X_2, \dots, X_m\}$, $X_{i+1} = [x_i, x_{i+1}]$, $x_i < x_{i+1}$,
 $i = 0, 1, 2, \dots, m - 1$.
 Radius $\varepsilon > 0$.

Output: ε -capacity $\mathcal{E}_\varepsilon(M)$.

1. From the partition ξ create the sets of points $\mathcal{X} = \{x_0, x_1, x_2, \dots, x_m\}$.
2. If $(x_m - x_0) \leq \varepsilon$ then
3. Set $\mathcal{M}_\varepsilon(M) = 1$.
4. Else
5. Set $\mathcal{M}_\varepsilon(M) = 2$.
6. Set $j = 0$.
7. For $i = 1$ to $m - 1$ do:
8. If $(x_i - x_j) \leq \varepsilon$ or $(x_m - x_i) \leq \varepsilon$ then
9. Continue.
10. Else
11. Set $\mathcal{M}_\varepsilon(M) = \mathcal{M}_\varepsilon(M) + 1$.
12. Set $j = i$.
13. End if.
14. End for.
15. End if.
16. Set $\mathcal{E}_\varepsilon(M) = \log_2 \mathcal{M}_\varepsilon(M)$.
17. Return $\mathcal{E}_\varepsilon(M)$.

Time complexity \mathcal{C} of Algorithm 1 includes the following terms: $\mathcal{O}(1)$ – complexity of the lines 1-4; $\mathcal{O}(m)$ – complexity of the line 5; $\mathcal{O}(m \log m)$ – complexity of the line 6; $\mathcal{O}(1)$ – complexity of the line 7; $\mathcal{O}(m)$ – complexity of the line 8 and $\mathcal{O}(1)$ – complexity of the lines 9-13. Then, time complexity of each iteration of the algorithm is $\mathcal{O}(m \log m)$. The maximal number of iterations is $n = (t_m - t_0)/2s$; hence complexity of Algorithm 1 is

$$\mathcal{C} = n \times \mathcal{O}(m \log m).$$

Convergence of Algorithm 1 is guaranteed by the indicated above fact that ε -information $J_\varepsilon(M)$ increases with increasing ε while ε -capacity $\mathcal{E}_\varepsilon(M)$ decreases with increasing ε . Since the interval $M = [x_0, x_m]$ is bounded, the difference between increasing ε -information $J_\varepsilon(M)$ and decreasing ε -capacity $\mathcal{E}_\varepsilon(M)$ has its minimum in M , which is a terminating point of the algorithm.

Algorithm 1 returns radius ε for which the difference $|J_\varepsilon(M) - \mathcal{E}_\varepsilon(M)|$ between ε -information and ε -capacity is minimal. Consequently, partition $\xi_\varepsilon = \{Y \mid Y \subset M\}$, in which diameters of the sets $Y \in \xi_\varepsilon$ are $d(Y) = 2\varepsilon$, is a required partition.

Dependence of the functions $J_\varepsilon(M)$ and $\mathcal{E}_\varepsilon(M)$ on the diameter 2ε is illustrated in Figure 1. In this example, $M = [x_0, x_m]$, $x_0 = 0$, $x_m = 100$, and in observations partition $\xi = \{X_1, X_2, \dots, X_m\}$, $X_{i+1} = [x_i, x_{i+1}]$, $x_i < x_{i+1}$, $i = 0, 1, 2, \dots, m - 1$, points x_i are evenly distributed on the interval M .

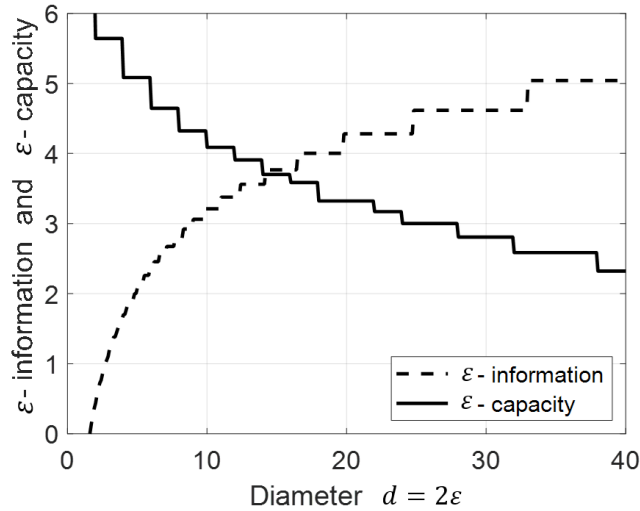


Figure 1. Dependence of ε -information $J_\varepsilon(M)$ and ε -capacity $\mathcal{E}_\varepsilon(M)$ on the diameter $d = 2\varepsilon$ of the sets $Y \in \xi_\varepsilon$ for the partition $\xi = \{X_1, X_2, \dots, X_m\}$, $X_{i+1} = [x_i, x_{i+1}]$, $x_i < x_{i+1}$, $i = 0, 1, 2, \dots, m-1$, where $m = 100$ points are evenly distributed on the interval $M = [x_0, x_m]$, $x_0 = 0$, $x_m = 100$. The step is $s = 1$.

The computed radius is $\varepsilon \approx 7.1$ and diameter is $d = 2\varepsilon \approx 14.2$. The values of ε -information and ε -capacity are $J_\varepsilon(M) = \mathcal{E}_\varepsilon(M) \approx 3.7$ bit.

Note that the accuracy of computing radius ε increases with decreasing the step s .

4. Numerical Examples

Let $M = [x_0, x_m] \subset \mathbb{R}$ be an interval of real numbers, $x_m > x_0$. Given $\varepsilon > 0$, a minimal number of sets in the ε -covering α of the interval M is

$$\mathcal{N}_\varepsilon(M) = \left\lceil \frac{x_m - x_0}{2\varepsilon} \right\rceil.$$

Then, the ε -entropy of the set M is

$$\mathcal{H}_\varepsilon(M) = \log_2 \mathcal{N}_\varepsilon(M) = \log_2 \left\lceil \frac{x_m - x_0}{2\varepsilon} \right\rceil.$$

Let ε_m be a minimal value of ε which provides maximal ε -entropy

$$\mathcal{H}_{\varepsilon_m}(M) = \log_2 \left\lceil \frac{x_m - x_0}{2\varepsilon_m} \right\rceil$$

of the interval M . Following lemma 5

$$\mathcal{H}_{\varepsilon_m}(M) = \log_2 m;$$

thus

$$\varepsilon_m = \frac{1}{2} \frac{x_m - x_0}{m}.$$

Let

$$\xi = \{X_1, X_2, \dots, X_m\}, \quad X_{i+1} = [x_i, x_{i+1}], \quad x_i < x_{i+1}, \quad i = 0, 1, 2, \dots, m-1,$$

be an observations partition of the interval M and

$$\xi_\varepsilon = \{Y_1, Y_2, \dots, Y_{m_\varepsilon}\}, \quad Y_{j+1} = [y_j, y_{j+1}], \quad y_j < y_{j+1}, \quad y_0 = x_0, \quad y_{m_\varepsilon} = x_m,$$

$$y_{j+1} - y_j = 2\varepsilon, \quad j = 0, 1, 2, \dots, m_\varepsilon - 1.$$

be a partition of the interval M such that all its elements are of the length 2ε .

Then, ε -information of the interval M is

$$J_\varepsilon(M) = 2 \log_2 m - \log_2 \left\lceil \frac{x_m - x_0}{2\varepsilon} \right\rceil - \mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon).$$

Entropy $\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon)$ depends on distribution of the values x_i , $i = 0, 1, 2, \dots, m$, over the interval M .

Assume that these values are distributed evenly and $\xi \approx \xi_\varepsilon$. Then $\xi \vee \xi_\varepsilon = \xi_\varepsilon$ and

$$\mathcal{H}_\varepsilon(\xi \vee \xi_\varepsilon) = \mathcal{H}_\varepsilon(M) = \log_2 \left\lceil \frac{x_m - x_0}{2\varepsilon} \right\rceil.$$

Note that in general this equivalence does not hold and calculation of the entropy of multiplication of partitions is processed according to function 1 (see section 3).

Similarly, ε -capacity $\mathcal{E}_\varepsilon(M)$ depends on the distribution of the values x_i , $i = 0, 1, 2, \dots, m$, over the interval M . If these values are distributed evenly such that $x_{j+1} - x_j = x_{j+2} - x_{j+1}$ and $x_{j+1} - x_j > \varepsilon$ for any $j = 0, 1, 2, \dots, m - 2$, then

$$\mathcal{M}_\varepsilon(M) = \left\lceil \frac{x_m - x_0}{\varepsilon} \right\rceil$$

and

$$\mathcal{E}_\varepsilon(M) = \log_2 \mathcal{M}_\varepsilon(M) = \log_2 \left\lceil \frac{x_m - x_0}{\varepsilon} \right\rceil.$$

If distribution of the values x_i is such that $x_{m-1} - x_0 \leq \varepsilon$, which means that all the values except x_m are located between x_0 and x_{m-1} , and $x_m - x_0 > \varepsilon$, then

$$\mathcal{M}_\varepsilon(M) = 2$$

and

$$\mathcal{E}_\varepsilon(M) = \log_2 2 = 1.$$

If $x_m - x_0 \leq \varepsilon$, then the interval M does not contain ε -distinguishable subsets, and we assume that

$$\mathcal{M}_\varepsilon(M) = 1$$

and

$$\mathcal{E}_\varepsilon(M) = \log_2 1 = 0.$$

In general case, ε -capacity is calculated using function 2 (see section 3).

Finally, using the formulated criterion (see the problem in section 3), for evenly distributed values x_i we obtain (for simplicity we omit the notion of ceiling)

$$\mathcal{J}_\varepsilon(M) - \mathcal{E}_\varepsilon(M) = 2 \log_2 m - 2 \log_2 \frac{x_m - x_0}{2\varepsilon} - \log_2 \frac{x_m - x_0}{\varepsilon} = \log_2 m^2 - \log_2 \frac{(x_m - x_0)^3}{4\varepsilon^3}.$$

Thus, it is required to find a value ε such that

$$\frac{(x_m - x_0)^3}{4\varepsilon^3} = m^2,$$

which is

$$\varepsilon = \frac{x_m - x_0}{\sqrt[3]{4m^2}}.$$

Consequently, diameter $d(Y_j)$ of the intervals Y_j , $j = 1, 2, \dots, m_\varepsilon$, in the required partition $\xi_\varepsilon = \{Y_1, Y_2, \dots, Y_{m_\varepsilon}\}$ is

$$d(Y_j) = 2 \frac{x_m - x_0}{\sqrt[3]{4m^2}}.$$

To illustrate the suggested method, let us calculate diameters $d(Y_j)$, $j = 1, 2, \dots, m_\varepsilon$, of intervals in the partition ξ_ε and compare the obtained results with the bin lengths provided by the known methods appearing in the literature.

Let $M = [0, 10]$ and $\xi = \{X_1, X_2, \dots, X_{10}\}$, $X_1 = [0, 1]$, $X_2 = (1, 2], \dots$, $X_{10} = (9, 10]$. Then,

$$d(Y_j) = 2 \frac{10-0}{\sqrt[3]{4 \times 11^2}} = 2.55,$$

for all $j = 1, \dots, m_\varepsilon = 4$. For this interval M , bin lengths calculated by the known methods used for plotting histograms are the following:

- the simplest rule: $d_1 = \frac{\max A - \min A}{\sqrt{m}} = \frac{10-0}{\sqrt{11}} = 3.01,$
- the Sturges rule [14]: $d_2 = \frac{\max A - \min A}{\log_2 m + 1} = \frac{10-0}{\log_2 11 + 1} = 2.24,$
- the Scott rule [11]: $d_3 = \frac{3.49s}{\sqrt[3]{m}} = \frac{3.49 \times 3.32}{\sqrt[3]{11}} = 5.20,$
- the Freedman-Diaconis rule [3]: $d_4 = \frac{2(Q_3 - Q_1)}{\sqrt[3]{m}} = \frac{2 \times (8-3)}{\sqrt[3]{11}} = 4.94.$

It is seen that in this example with evenly distributed observations, diameter calculated using the suggested method is compatible with the bin lengths calculated using the known rules.

Now let us consider examples with different distributions of observed values x_i , $i = 0, 1, 2, \dots, m$, and so – different observations partitions $\xi = \{X_i | X_i \subset M\}$. In all examples we assume that $M = [x_0, x_m]$, $x_0 = 0$, $x_m = 100$ and $m = 100$.

In these examples, we used uniform distribution with $a = 0$ and $b = x_m$, normal distribution with $\mu = x_m/2$ and $\sigma = x_m/6$, and exponential distribution with $\mu = 2$. Diameters d calculated by the suggested method and bin lengths d_1, \dots, d_4 calculated by the known rules are the following:

- evenly distributed data:

$$d = 14.21, \quad d_1 = 9.90, \quad d_2 = 12.95, \quad d_3 = 21.81 \quad \text{and} \quad d_4 = 21.54,$$

- uniform distribution with $a = 0$ and $b = x_m$:

$$d = 14.11, \quad d_1 = 9.83, \quad d_2 = 12.87, \quad d_3 = 22.69 \quad \text{and} \quad d_4 = 22.65,$$

- normal distribution with $\mu = x_m/2$ and $\sigma = x_m/6$:

$$d = 11.91, \quad d_1 = 8.27, \quad d_2 = 10.82, \quad d_3 = 13.11 \quad \text{and} \quad d_4 = 10.38,$$

- exponential distribution with $\mu = 2$:

$$d = 1.20, \quad d_1 = 1.07, \quad d_2 = 1.41, \quad d_3 = 1.35 \quad \text{and} \quad d_4 = 0.91.$$

It is seen that the suggested method results in the diameters d that are close to the bin lengths d_1, \dots, d_4 provided by the conventional methods with respect to the distribution of the data. For evenly and uniformly distributed data diameter d is close to the bin length d_2 resulted by the Sturges method, for normal distribution $d_2 < d < d_3$ and for exponential distribution $d_1 < d < d_2$.

Histograms of these data with the bin lengths d_3 calculated using the Scott rule and with the bin lengths d calculated by the suggested method are shown in Figure 2.

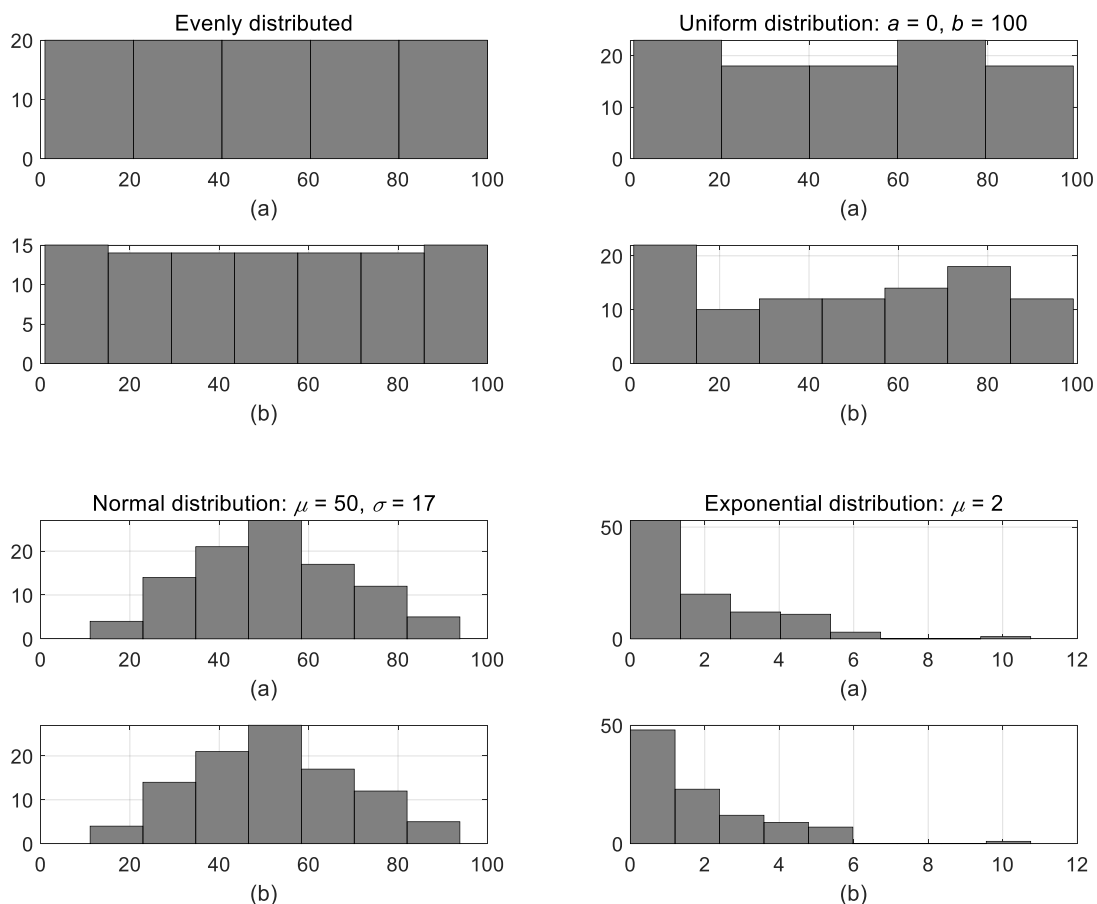


Figure 2. Histograms plotted using the bin lengths d_3 calculated using the Scott rule (figures (a) for each distribution) and using the bin lengths d calculated using the suggested method (figures (b) for each distribution).

Finally, let us consider specific distribution, which appears in the analysis of queues [4].

Consider a clerk serving the clients of some office during a day. Then, arrival λ and departure μ rates measured per day describe the state of the system at the end of the day, but do not provide any information about the system during the day.

On other hand, the rates λ and μ measured per minute are also useless since the clients do not arrive and the clerk does not serve with such velocities.

Thus, it is required to define correct time units per which arrival and service rates should be specified. In such a form the problem was formulated by Yaakov Reis [8].

Assume that the office, where the mentioned above clerk works, serves 480 clients 8 hours during a day. Also, assume that the clients arrive by three “waves” – in the morning, in the midday and in the evening.

For this data, diameter d calculated by the suggested method and bin lengths d_1, \dots, d_4 calculated by the known rules are the following:

$$d = 22.0, \quad d_1 = 21.91, \quad d_2 = 48.45, \quad d_3 = 67.99 \quad \text{and} \quad d_4 = 66.41.$$

Histogram for this data with the bin lengths d_3 calculated using the Scott rule is shown in Figure 3.a, and histogram with the bin lengths d calculated by the suggested method is shown in Figure 3.b.

Dependence of the functions $J_\varepsilon(M)$ and $\mathcal{E}_\varepsilon(M)$ on ε for this distribution is shown in Figure 3.c.

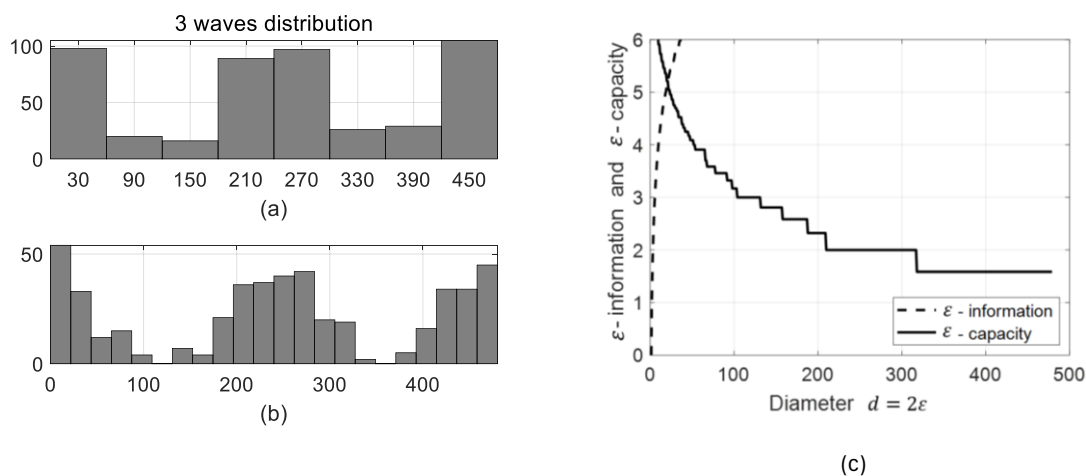


Figure 3. Arrivals of the clients during a day: (a) histogram with the bin length computed by the Scott rule; (b) histogram with the bin length computed by the suggested algorithm; (c) dependences of ε -information $J_\varepsilon(M)$ and ε -capacity $\mathcal{E}_\varepsilon(M)$ on the diameter $d = 2\varepsilon$.

It is seen that for this data the suggested method results in the diameter d which is close to the bin length d_1 provided by the simplest rule. Following this result, arrival rates during a day should be calculated each 22 minutes, while the Sturges rule they should be calculated each 48 minutes, by the Scott rule – each 68 minutes, and by the Freedman-Diaconis rule – each 66 minutes.

5. Conclusion

In the paper, we suggested a method of discretization of dynamical system basing on its observations. The method results in the partition ξ_ε , which represents the system, and is as close as possible to the observations partition ξ .

The method utilizes the Kolmogorov and Tikhomirov ε -entropy and ε -capacity and the Rokhlin distance between partitions.

Along with theoretical results, the method is presented in the form of ready-to-use algorithm.

The suggested method is also useful for definition of the bin lengths of histograms, especially for the datasets with multimodal distributions.

Funding: This research has not received any grant from funding agencies in the public, commercial, or non-profit sectors.

Conflicts of interest: The authors declare no conflict of interest.

References

1. Billingsley P. *Ergodic Theory and Information*. John Wiley and Sons: New York, 1965.
2. Dinaburg E. I. On the relations among various entropy characteristics of dynamical systems. *Math. USSR Izvestija*, 1971, 5(2), 337-378.
3. Freedman D., Diaconis P. On the histogram as a density estimator: L_2 theory. *Zeit. Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 1981, 57, 453-476.
4. Gross D., Shortle J. F., Thompson J. M., Harris C.M. *Fundamentals of Queueing Theory*, 4th ed. John Wiley & Sons: Hoboken, NJ, 2008.
5. Jacod J., Protte P. *Discretization of Processes*. Springer: Berlin, 2012.
6. Kolmogorov A. N., Tikhomirov V. M. ε -entropy and ε -capacity of sets in functional spaces. *Amer. Mathematical Society Translations, Ser. 2*, 1961, 17, 277-364.
7. Martin N., England J. *Mathematical Theory of Entropy*. Cambridge University Press: Cambridge, UK, 1984.
8. Reis Y. Private conversation. Ariel University, Ariel, March 2024.
9. Rokhlin V. A. New progress in the theory of transformations with invariant measure. *Russian Mathematical Surveys*, 1960, 15(4), 1–22.
10. Schwarz G. Estimating the dimension of a model. *Annals of Statistics*, 1978, 6(2), 461-464.
11. Scott D. W. On optimal and data-based histograms. *Biometrika*, 1979, 66, 605-610.
12. Sinai Y. G. *Introduction to Ergodic Theory*. Princeton University Press: Princeton, 1976.
13. Sinai Y. G. *Topics in Ergodic Theory*. Princeton University Press: Princeton, 1993.
14. Sturges H. The choice of a class-interval. *J. Amer. Statistics Association*, 1926, 21, 65-66.
15. Vitushkin A. G. *Theory of Transmission and Processing of Information*, Pergamon Press: New York, 1961.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.