

Article

Not peer-reviewed version

---

# Definition of Optimal Time Intervals in the Queues' Analysis: The Use of Epsilon-Entropy and Epsilon-Capacity

---

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

Posted Date: 9 September 2024

doi: 10.20944/preprints202409.0583.v1

Keywords: Arrival rate; departure rate; epsilon-entropy; epsilon-capacity; epsilon-information; partition



Preprints.org is a free multidiscipline 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 is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Definition of Optimal Time Intervals in the Queues' Analysis: The Use of $\varepsilon$ -Entropy and $\varepsilon$ -Capacity

Eugene Kagan <sup>1,\*</sup> and Alexander Novoselsky <sup>2</sup>

<sup>1</sup> Ariel University, Kiryat ha-Mada, Ariel 4070000, Israel

<sup>2</sup> Log-On Software, 3 ha-Hilazon St., Ramat Gan 5252267, Israel

\* Correspondence: evganyk@ariel.ac.il

**Abstract:** In the paper, we suggest a method for calculating optimal time intervals in the queue analysis. The suggested method processes partitioning of the time interval and utilizes the  $\varepsilon$ -entropy and  $\varepsilon$ -capacity of the partition for finding an optimal partition. Optimality of the partition is specified based on its  $\varepsilon$ -information. The suggested method is illustrated by defining the intervals for histograms of differently distributed samples and demonstrated its effectiveness in comparison with the existing methods.

**Keywords:** Arrival rate; departure rate;  $\varepsilon$ -entropy;  $\varepsilon$ -capacity;  $\varepsilon$ -information; partition

**MSC:** 60K25; 90B22

## 1. Introduction

The use of queue models implies the knowledge of the arrival and departure rates, which in their turn require the well-defined time intervals [3].

For example, consider a clerk serving clients of some office during the day. Then, the arrival rate  $\lambda$  and the departure rate  $\mu$  per day completely 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, specification of the rates  $\lambda$  and  $\mu$ , for example, per minute is also useless since both the clients and the clerk do not act with such rates.

The incorrect specification of the time intervals leads to incorrect consideration of the processes with unsteady arrivals or departures. In many cases, such situations are resolved using the queues with time dependent rates [5,8], but even in such considerations certain time intervals per which the rates are defined have to be specified.

Similar problem appears in statistics while plotting histograms of the data sample  $X$  and it is required to define the lengths  $\delta$  of the bins. Since there is no strictly proven formula, which defines the bin length with respect to the number  $n$  of data counts or distribution over the sample, the heuristic formulas are used.

For example, the simplest heuristic defines the bin length as

$$\delta_1 = \frac{\max(X) - \min(X)}{\sqrt{n}}. \quad (1)$$

The Sturges rule [15] defines the number of bins as  $\lceil \log_2 n \rceil + 1$ . Then the bin length is

$$\delta_2 = \frac{\max(X) - \min(X)}{\lceil \log_2 n \rceil + 1}. \quad (2)$$

The Scott rule [12] defines the bin length

$$\delta_3 = \frac{3.49s}{\sqrt[3]{n}} \quad (3)$$

with respect to the standard deviation  $s$  of the sample. Finally, the Freedman-Diaconis formula [2] uses the interquartile range instead of the standard deviation and defines the bin length

$$\delta_4 = \frac{2(Q_3 - Q_1)}{\sqrt[3]{n}}, \quad (4)$$

where  $Q_3$  is the third quartile and  $Q_1$  is the first quartile of the sample.

In general, this problem can be considered in the terms of discretization of stochastic processes [4], where it is required to build a discretization scheme, which is a sequence  $t_i$ ,  $i = 0, 1, 2, \dots$ , of stopping times such that  $\Delta = t_{i+1} - t_i$  and  $t_i = i\Delta$ . But if  $t_i$  is a random variable, then the length  $\Delta$  of the time intervals is also random and depend on the distribution of the considered process. Similarly, if the discretization scheme is regular with constant interval lengths  $\Delta$ , then the increments of the process at the times  $t_i$  are random.

For example, let  $W_t$  be a Wiener process at the time interval  $T = [0, t_m]$  starting with  $W_0 = 0$ . In such a process, the increments  $dW_t$  are independent and for any  $t_i$  and  $t_j > t_i$  the differences  $W_{t_j} - W_{t_i}$  have normal distribution  $\mathcal{N}(0, \sigma_t^2)$  with the variance  $\sigma_t^2 = t_j - t_i$ . Assume that the interval  $T$  is divided to  $n$  sub-intervals with the length  $\Delta = t_m/n$ . Then, the stopping times are  $t_i = i\Delta$ ,  $i = 0, 1, 2, \dots$ , and the increments  $dW_t = W_{t_{i+1}} - W_{t_i}$  are normally distributed with  $\sigma_t^2 = \Delta$ .

In this paper, we seek the answer to the following question formulated by Yaakov Reis [9]. Given a total period, a sequence of clients arriving at the times  $t_0, t_1, t_2, \dots, t_m$  to a service point, what is an optimal length  $\Delta$  of the time interval, on which the arrival rate  $\lambda$  and the service rate  $\mu$  (which is a departure rate) have to be defined?

An immediate answer to this question follows the heuristics used for definition of the bin length  $\delta$  in histogram. However, such heuristics cannot be considered as the best method and their result is not strictly proved approximation.

To find an optimal length  $\Delta$  we follow the line of the Schwarz information criterion [11] and apply well-known concepts of  $\varepsilon$ -entropy and  $\varepsilon$ -capacity, which were introduced by Kolmogorov and Tikhomirov [6]. The calculations of the optimal interval are also based on the concept of the entropy of partition introduced by Rokhlin [10].

Initially  $\varepsilon$ -entropy and  $\varepsilon$ -capacity were used for analysis of functions and functional spaces and then, as well as the entropy of partition, were applied to the studies of dynamical systems. For many examples of application of these concepts and their relationship with the Shannon entropy [13] see the paper by Dinaburg [1] and the books by Vitushkin [16] and by Sinai [14].

## 2. Problem Formulation

Let  $T = [t_0, t_m]$  be a time interval of the length  $t_m - t_0 > 0$  and assume that during this interval sequentially occur  $m + 1$  events  $a_0, a_1, a_2, \dots, a_m$ . The times of occurrences of these events are  $t_0 \leq t_1 \leq t_2 \leq \dots \leq t_m$ , respectively.

The problem is to define a length  $\Delta$  of the time interval or, that is the same, the stopping times  $t_i = i\Delta$ ,  $i = 0, 1, 2, \dots, n$ , such that  $n$  intervals  $T_i = [t_i, t_{i+1}]$  cover the interval  $T$  and such that they as better as possible represent the times  $t_j$ ,  $j = 0, 1, 2, \dots, m$ , when the considered events occurred.

To illustrate the problem, let us consider a simple example of a non-steady supply process. Assume that the mentioned above clerk serves the clients with the rate  $\mu = 5$  clients/hour. During the workday of 8 hours arrive 24 clients. Then, the arrival rate of the clients defined over a workday is  $\lambda = 24/8 = 3$  clients/hour and the transition rate  $\rho = \lambda/\mu = 3/5 < 1$  that should guarantee that at the end of the workday all clients will be served.

Additionally, assume that in the morning, during the first two hours of the day, arrive 12 clients. Then during the next four hours the clients do not arrive and then, in the evening, during the last two hours of the day, arrive the last 12 clients. Thus, in the morning and in the evening the arrival rate is  $\lambda = 12/2 = 6$  clients/hour, that means that the first 12 clients will wait in the queue and the last 12 clients will not be served until the end of the workday.

Certainly, such phenomena are well-known; in the queue theory they are solved using the state-dependent and time-dependent arrival rates [3,5,8], and in practice are overcome by adding the clerks in the morning and in the evening and by stopping the service in the midday. However, a prior definition of the appropriate time intervals can simplify further analysis and even decrease the expected number of varying rates.

Finally, note that the considered problem is essentially discrete problem, where it is required to split the discrete dataset. Together with that, since it is closely related to the discretization problems dealing with the continuous functions, below we will make some remarks on such problems as well.

### 3. Methods

The suggested solution of the problem is based on the concepts of  $\varepsilon$ -entropy and  $\varepsilon$ -capacity, which were introduced by Kolmogorov and Tikhomirov in the middle of 1950-s and presented in detail in their paper [6]. In addition, it uses the multiplication of partitions as it was implemented by Rokhlin [10] and by Sinai [14] in the studies of dynamical systems.

#### 3.1. $\varepsilon$ -Entropy and $\varepsilon$ -Capacity

Let  $U \subset R$  be a non-empty bounded set of a metric space  $R$  and let  $\varepsilon > 0$  be a real number.

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

The set  $U$  is said to be  $\varepsilon$ -distinguishable, if any two of its distinct points are located at distance greater than  $\varepsilon$ .

Given a bounded set  $U \subset R$ , for any  $\varepsilon > 0$  there exists a finite  $\varepsilon$ -covering of  $U$ , and for any  $\varepsilon > 0$  any  $\varepsilon$ -distinguishable set  $U \subset R$  is finite.

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

The value

$$\mathcal{H}_\varepsilon(U) = \log_2 \mathcal{N}_\varepsilon(U) \quad (5)$$

is called the  $\varepsilon$ -entropy of the set  $U$ , and the value

$$\mathcal{E}_\varepsilon(U) = \log_2 \mathcal{M}_\varepsilon(U) \quad (6)$$

is called the  $\varepsilon$ -capacity of the set  $U$ .

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

Among the properties of  $\varepsilon$ -entropy  $\mathcal{H}_\varepsilon(U)$  and  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(U)$  we will use the following fact [6]: given the bounded set  $U$ , both  $\varepsilon$ -entropy and  $\varepsilon$ -capacity as functions of  $\varepsilon$  are non-increasing with increasing  $\varepsilon$ .

Examples of calculation of the  $\varepsilon$ -entropy and  $\varepsilon$ -capacity of the sets in different metric spaces can be found in the paper by Kolmogorov and Tikhomirov [6] and in the book by Vitushkin [16].

#### 3.2. $\varepsilon$ -entropy of Partition

Let  $\beta = \{B: B \subset U\}$  be a partition of the set  $U \subset R$  that is  $U = \bigcup_{B \in \beta} B$  and for any two sets  $B', B'' \in \beta$  holds  $B' \cap B'' = \emptyset$ .

The entropy of partition is defined as follows [10,14]. Let  $\mu$  be a non-negative measure on the set  $U$  such that  $\mu(\emptyset) = 0$  and  $\mu(U) = 1$ . Then,  $\mu(B) \in [0, 1]$  for any  $B \in \beta$ . The value

$$\mathcal{H}_\mu(\beta) = -\sum_{B \in \beta} \mu(B) \log_2 \mu(B) \quad (7)$$

is called the *entropy of partition*. If  $\mu$  is the probability measure on  $U$ , then the sets  $B \in \beta$  can be interpreted as events and the entropy  $\mathcal{H}_\mu$  is equivalent to the Shannon entropy [13].

Assume that the partition  $\beta$  is finite and the number of the sets in  $\beta$  is  $N$ . Define the measure  $\mu$  on the set  $U$  as follows:

$$\mu(B) = \begin{cases} 0, B = \emptyset \\ \frac{1}{N}, B \neq \emptyset, B \neq U \\ 1, B = U \end{cases} \quad (8)$$

Then, the entropy of partition is reduced to the value

$$\mathcal{H}_\mu(\beta) = \log_2 N \quad (9)$$

Finally, if the diameter of any set  $B \in \beta$  is not greater than  $2\varepsilon$ , then the partition  $\beta$  is an  $\varepsilon$ -covering and called  $\varepsilon$ -partition. Then, the entropy  $\mathcal{H}_\mu(\beta)$  of partition is equivalent to the  $\varepsilon$ -entropy  $\mathcal{H}_\varepsilon(U)$  of the set  $U = \bigcup_{B \in \beta} B$  defined by equation (5).

Let  $\beta = \{B: B \subset U\}$  be  $\varepsilon$ -partition of the set  $U$  with  $\varepsilon = \varepsilon_B$  and  $\gamma = \{C: C \subset U\}$  be another  $\varepsilon$ -partition of the set  $U$  with  $\varepsilon = \varepsilon_C$ . Multiplication of the partitions  $\beta$  and  $\gamma$  is the partition

$$\beta \vee \gamma = \{D = B \cap C: B \in \beta, C \in \gamma\}. \quad (10)$$

Each set  $D \in \beta \vee \gamma$  is a subset of some set  $B \in \beta$  and of some set  $C \in \gamma$ . Then it is said that  $\beta \vee \gamma$  is a refinement of both  $\beta$  and  $\gamma$ ; this fact is denoted by  $\beta \preccurlyeq \beta \vee \gamma$  and  $\gamma \preccurlyeq \beta \vee \gamma$ . Hence, following the properties of the entropy of partition,

$$\mathcal{H}_\mu(\beta) \leq \mathcal{H}_\mu(\beta \vee \gamma) \text{ and } \mathcal{H}_\mu(\gamma) \leq \mathcal{H}_\mu(\beta \vee \gamma) \quad (11)$$

Moreover, the entropy  $\mathcal{H}_\mu(\beta \vee \gamma)$  of the multiplication  $\beta \vee \gamma$  of the partition  $\beta$  and  $\gamma$  is the  $\varepsilon$ -entropy  $\mathcal{H}_\varepsilon(U)$  of the set  $U$  with any  $\varepsilon \in [\min\{\varepsilon_B, \varepsilon_C\}, \max\{\varepsilon_B, \varepsilon_C\}]$ .

For the other properties of the entropy  $\mathcal{H}_\mu$  and its application for analysis of dynamical systems see the paper [10] and the book [14].

#### 4. Suggested Solution

Let  $T = [t_0, t_m]$ ,  $t_m > t_0$ , be a time interval and let  $\mathcal{T} = \{t_0, t_1, t_2, \dots, t_m\}$  be the set of moments in which certain events occur. We assume that the moments  $t_j$  have an increasing order such that  $t_j < t_{j+1}$ ,  $j = 0, 1, 2, \dots, m-1$ .

As it follows from the formulation of the problem, in the consideration below the interval  $T$  plays a role of the set  $U$  and the intervals  $[t_j, t_{j+1}]$ ,  $j = 0, 1, 2, \dots, m-1$ , are considered as the elements of  $\varepsilon$ -partitions of the interval  $T$ .

Given  $\varepsilon > 0$ , the minimal number of sets in the  $\varepsilon$ -covering  $\alpha$  of the set  $\mathcal{T}$

$$\mathcal{N}_\varepsilon(\mathcal{T}) = \left\lceil \frac{t_m - t_0}{2\varepsilon} \right\rceil. \quad (12)$$

Then, the  $\varepsilon$ -entropy of the set  $\tau$  is

$$\mathcal{H}_\varepsilon(\mathcal{T}) = \log_2 \mathcal{N}_\varepsilon(\mathcal{T}) = \log_2 \left\lceil \frac{t_m - t_0}{2\varepsilon} \right\rceil. \quad (13)$$

Let

$$\varepsilon_{\min} = \frac{1}{2} \frac{t_m - t_0}{m^2} \quad (14)$$

be a minimal value of  $\varepsilon$  for the set  $\mathcal{T}$ . Then, the value

$$\mathcal{H}_{\varepsilon_{\min}}(\mathcal{T}) = \log_2 \left\lceil \frac{t_m - t_0}{2\varepsilon_{\min}} \right\rceil = 2 \log_2 m \quad (15)$$

is maximal  $\varepsilon$ -entropy of the set  $\mathcal{T}$ .

Finally, assume that on the interval  $T$  two sets  $\mathcal{T}_1 = \{t_{1,0}, t_{1,1}, t_{1,2}, \dots, t_{1,m_1}\}$  and  $\mathcal{T}_2 = \{t_{2,0}, t_{2,1}, t_{2,2}, \dots, t_{2,m_2}\}$ ,  $t_{1,0} = t_{2,0}$  and  $t_{1,m_1} = t_{2,m_2}$ , of moments are defined. Denote by  $\tau_1 = \{[t_{1,0}, t_{1,1}], [t_{1,1}, t_{1,2}], \dots, [t_{1,m_1-1}, t_{1,m_1}]\}$  partition of the interval  $T$  corresponding to the set  $\mathcal{T}_1$  and by  $\tau_2 = \{[t_{2,0}, t_{2,1}], [t_{2,1}, t_{2,2}], \dots, [t_{2,m_2-1}, t_{2,m_2}]\}$  partition of the interval corresponding to the set  $\mathcal{T}_2$ . The number of intervals in the partition  $\tau_1$  is  $m_1$  and the number of intervals in the partition  $\tau_2$  is  $m_2$ .

Then, since the multiplication  $\tau_1 \vee \tau_2$  is a refinement of each of the partitions  $\tau_1$  and  $\tau_2$  of the sets, the size  $m_{1 \vee 2}$  of the partition  $\tau_1 \vee \tau_2$  is  $m_{1 \vee 2} \geq \max\{m_1, m_2\}$ , and the entropy  $\mathcal{H}_\mu(\tau_1 \vee \tau_2)$  of the multiplication  $\tau_1 \vee \tau_2$  is not smaller than the entropies  $\mathcal{H}_\mu(\tau_1)$  and  $\mathcal{H}_\mu(\tau_2)$  of the partitions  $\tau_1$  and  $\tau_2$ .

Hence, if

$$2\varepsilon \geq \max_{j=0,1,2,\dots,m_1-1} (t_{1,j+1} - t_{1,j}) \text{ and } 2\varepsilon \geq \max_{j=0,1,2,\dots,m_2-1} (t_{2,j+1} - t_{2,j}), \quad (16)$$

then, following equation (11),

$$\mathcal{H}_\varepsilon(\mathcal{T}_1) \leq \mathcal{H}_\varepsilon(\mathcal{T}_1 \vee \mathcal{T}_2) \text{ and } \mathcal{H}_\varepsilon(\mathcal{T}_2) \leq \mathcal{H}_\varepsilon(\mathcal{T}_1 \vee \mathcal{T}_2). \quad (17)$$

Following the line of the Schwarz information criterion [11], let us define  $\varepsilon$ -information of the set  $\mathcal{T}$ .

Let  $\tau$  be a partition corresponding to the set  $\mathcal{T}$  and let  $\tau_\varepsilon$  be a partition corresponding to the set  $\mathcal{T}_\varepsilon = \{t_{\varepsilon,0}, t_{\varepsilon,1}, t_{\varepsilon,2}, \dots, t_{\varepsilon,m_\varepsilon}\}$ ,  $t_{\varepsilon,0} = t_0$  and  $t_{\varepsilon,m_\varepsilon} = t_m$ , in which  $t_{\varepsilon,j+1} - t_{\varepsilon,j} = 2\varepsilon$ ,  $j = 0, 1, 2, \dots, m - 2$ . In the partition  $\tau_\varepsilon$  all intervals except the last are of the length  $2\varepsilon$ .

Denote by  $\mathcal{T} \vee \mathcal{T}_\varepsilon$  the set of moments corresponding to the multiplication  $\tau \vee \tau_\varepsilon$  of the partitions  $\tau$  and  $\tau_\varepsilon$ . Then,  $\varepsilon$ -information of the set  $\mathcal{T}$  is defined as follows

$$\mathcal{J}_\varepsilon(\mathcal{T}) = \mathcal{H}_{\varepsilon_{\min}}(\mathcal{T}) - \mathcal{H}_\varepsilon(\mathcal{T}) - \mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon). \quad (18)$$

In this formula, the first term represents the number of bits required to transmit the set  $\mathcal{T}$  with maximal precision, the second term represents the number of bits required to transmit the set  $\mathcal{T}$  with precision  $\varepsilon$ , and the last term represents the number of bits required to transmit the set  $\mathcal{T}$  with precision  $\varepsilon$  using additional set  $\mathcal{T}_\varepsilon$  generated with precision  $\varepsilon$ . Thus, the value  $\mathcal{J}_\varepsilon(\mathcal{T})$  is the number of bits remained after the transmission of the set  $\mathcal{T}$  with the precision  $\varepsilon$ . In the other words,  $\varepsilon$ -information of the set  $\mathcal{T}$  characterizes the part of the set, which cannot be transmitted with the precision  $\varepsilon$ .

Using equations (13) and (15), formula (18) of  $\varepsilon$ -information can be simplified and written in the form

$$\mathcal{J}_\varepsilon(\mathcal{T}) = 2 \log_2 m - \log_2 \left\lceil \frac{t_m - t_0}{2\varepsilon} \right\rceil - \mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon). \quad (19)$$

The value of the entropy  $\mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon)$  depends on the distribution of time moments  $t_j \in \mathcal{T}$ ,  $j = 0, 1, 2, \dots, m$ , over the interval  $T$ . If the moments  $t_j$  are distributed evenly, then  $\mathcal{T} \vee \mathcal{T}_\varepsilon = \mathcal{T}$  and

$$\mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon) = \mathcal{H}_\varepsilon(\mathcal{T}) = \log_2 \left\lceil \frac{t_m - t_0}{2\varepsilon} \right\rceil. \quad (20)$$



Note that in general case equation (20) does not hold and calculation of the entropy of multiplication of partitions is processed according to the algorithm presented by Function 1 (see section 5).

Similarly, the value of  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(\mathcal{T})$  depends on the distribution of the time moments  $t_j$  over the interval  $T$ . If the moments  $t_j$  are distributed evenly such that  $t_{j+1} - t_j = t_{j+2} - t_{j+1}$  and  $t_{j+1} - t_j > \varepsilon$  for any  $j = 0, 1, 2, \dots, m-2$ , then

$$\mathcal{M}_\varepsilon(\mathcal{T}) = \left\lceil \frac{t_m - t_0}{\varepsilon} \right\rceil \quad (21)$$

and

$$\mathcal{E}_\varepsilon(\mathcal{T}) = \log_2 \mathcal{M}_\varepsilon(\mathcal{T}) = \log_2 \left\lceil \frac{t_m - t_0}{\varepsilon} \right\rceil. \quad (22)$$

If the distribution of the moments  $t_i$  is such that  $t_{m-1} - t_0 \leq \varepsilon$ , which means that all the moments except  $t_m$  are located between  $t_0$  and  $t_{m-1}$ , and  $t_m - t_0 > \varepsilon$ , then

$$\mathcal{M}_\varepsilon(\mathcal{T}) = 2 \quad (23)$$

and

$$\mathcal{E}_\varepsilon(\mathcal{T}) = \log_2 2 = 1. \quad (24)$$

Finally, if  $t_m - t_0 \leq \varepsilon$ , then the set  $T$  does not contain  $\varepsilon$ -distinguishable subset, and we assume that

$$\mathcal{M}_\varepsilon(\mathcal{T}) = 1 \quad (25)$$

and

$$\mathcal{E}_\varepsilon(\mathcal{T}) = \log_2 1 = 0. \quad (26)$$

Calculation of  $\varepsilon$ -capacity in general case follows the algorithm of Function 2 (see section 5).

The length  $\Delta$  of the time interval, which defines the stopping times  $t_i = i\Delta$ ,  $i = 0, 1, 2, \dots, n$ , is defined as

$$\Delta = 2\varepsilon, \quad (27)$$

where  $\varepsilon$  is such a value for which  $\varepsilon$ -information  $\mathcal{I}_\varepsilon(\mathcal{T})$  of the set  $\mathcal{T}$  is as close as possible to  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(\mathcal{T})$  of this set.

Note that given the set  $\mathcal{T}$ , the entropy  $\mathcal{H}_{\varepsilon_{min}}(\mathcal{T})$  is constant and both entropies  $\mathcal{H}_\varepsilon(\mathcal{T})$  and  $\mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon)$  as functions of  $\varepsilon$  are decreasing. Thus,  $\varepsilon$ -information  $\mathcal{I}_\varepsilon(\mathcal{T})$  increases with  $\varepsilon$ . Along with that,  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(\mathcal{T})$  as function of  $\varepsilon$  decreases.

Hence, the problem of finding the length  $\Delta$  is formulated as follows: given the set  $\mathcal{T}$ , find the value of  $\varepsilon$  such that

$$|\mathcal{I}_\varepsilon(\mathcal{T}) - \mathcal{E}_\varepsilon(\mathcal{T})| \rightarrow \min. \quad (28)$$

To illustrate the calculation of the length  $\Delta$ , let us consider a simple example. Assume that the considered time interval is  $T = [t_0, t_m]$  and the set  $\mathcal{T} = \{t_0, t_1, t_2, \dots, t_m\}$  consists of the evenly distributed moments  $t_j$  such that  $t_{j+1} - t_j = t_{j+2} - t_{j+1}$  for any  $j = 0, 1, 2, \dots, m-2$ . Then (here for simplicity we omit the notion of ceiling),

$$\mathcal{J}_\varepsilon(\mathcal{T}) - \mathcal{E}_\varepsilon(\mathcal{T}) = 2 \log_2 m - 2 \log_2 \frac{t_m - t_0}{2\varepsilon} - \log_2 \frac{t_m - t_0}{\varepsilon} = \log_2 m^2 - \log_2 \frac{(t_m - t_0)^3}{4\varepsilon^3}. \quad (29)$$

Hence, according to the criterion (28), it is required to specify the value  $\varepsilon$  such that

$$\frac{(t_m - t_0)^3}{4\varepsilon^3} = m^2, \quad (30)$$

which is

$$\varepsilon = \frac{t_m - t_0}{\sqrt[3]{4m^2}}. \quad (31)$$

and finally

$$\Delta = 2 \frac{t_m - t_0}{\sqrt[3]{4m^2}}. \quad (32)$$

For example, if  $\mathcal{T} = [0, 10]$  and  $A = \{0, 1, 2, \dots, 10\}$ , then

$$\Delta = 2 \frac{10-0}{\sqrt[3]{4 \times 11^2}} = 2.55.$$

For comparison, the indicated above methods (1)-(4) of specifying the bin length in the histograms result in the following values:

- the simplest rule:  $\delta_1 = \frac{\max A - \min A}{\sqrt{m}} = \frac{10-0}{\sqrt{11}} = 3.01,$
- the Sturges rule [15]:  $\delta_2 = \frac{\max A - \min A}{\log_2 m + 1} = \frac{10-0}{\log_2 11 + 1} = 2.24,$
- the Scott rule [12]:  $\delta_3 = \frac{3.49s}{\sqrt[3]{m}} = \frac{3.49 \times 3.32}{\sqrt[3]{11}} = 5.20,$
- the Freedman-Diaconis formula [2]:  $\delta_4 = \frac{2(Q_3 - Q_1)}{\sqrt[3]{m}} = \frac{2 \times (8-3)}{\sqrt[3]{11}} = 4.94.$

In the considered example, the interval length calculated using the suggested method is compatible with the lengths obtained using the methods of calculating the bin lengths, but for the other distributions the interval lengths can be strongly different.

Note again that in general case the interval lengths cannot be calculated using close formulas. In the next section we summarize the suggested methods in the form of an algorithm which is applicable to arbitrary data.

## 5. Algorithmic Implementation

We summarize the suggested solution in the form of an algorithm which can be directly implemented in any high-level programming language. In our trials we used the MATLAB® environment.

---

### Algorithm 1. Computing an optimal interval length

---

**Input:** Set  $\mathcal{T} = \{t_0, t_1, t_2, \dots, t_m\}$  of time moments,  $t_j < t_{j+1}$ ,  $j = 0, 1, 2, \dots, m-1$ ;  
step  $s > 0$ .

**Output:** Optimal interval length  $\Delta$ .

---

1. Calculate  $\varepsilon_{\min} = (t_m - t_0)/(2m^2)$  {minimal value of  $\varepsilon$ , equation (14)}.
2. Calculate  $\mathcal{H}_{\varepsilon_{\min}}(\mathcal{T}) = 2 \log_2 m$  {maximal  $\varepsilon$ -entropy, equation (15)}.
3. For  $\varepsilon = \varepsilon_{\min}$  to  $(t_m - t_0)/2$  with step  $s$  do:
4. Calculate  $\mathcal{H}_\varepsilon(A) = \log_2[(t_m - t_0)/2\varepsilon]$  { $\varepsilon$ -entropy, equation (13)}.



5. Create set  $\mathcal{T}_\varepsilon = \{t_{\varepsilon,0}, t_{\varepsilon,1}, t_{\varepsilon,2}, \dots, t_{\varepsilon,m_\varepsilon}\}$  such that  $t_{\varepsilon,j} < t_{\varepsilon,j+1}$ ,  $j = 0, 1, 2, \dots, m_\varepsilon - 1$ , and  $t_{\varepsilon,j+1} - t_{\varepsilon,j} = 2\varepsilon$ ,  $j = 0, 1, 2, \dots, m_\varepsilon - 2$ .
6. Compute  $\mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon) = \text{eps\_entropy}(\mathcal{T}, \mathcal{T}_\varepsilon)$  {entropy of  $\mathcal{T} \vee \mathcal{T}_\varepsilon$ , Function 1}
7. Calculate  $\mathcal{J}_\varepsilon(\mathcal{T}) = \mathcal{H}_{\varepsilon_{\min}}(\mathcal{T}) - \mathcal{H}_\varepsilon(\mathcal{T}) - \mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon)$  { $\varepsilon$ -information, equation (18)}.
8. Compute  $\mathcal{E}_\varepsilon(\mathcal{T}) = \text{eps\_capacity}(\mathcal{T}, \varepsilon)$  { $\varepsilon$ -capacity, Function 2}.
9. If  $\mathcal{J}_\varepsilon(\mathcal{T}) > \mathcal{E}_\varepsilon(\mathcal{T})$  then
10.       Break
11.   End if.
12. End for.
13. Return  $\Delta = 2\varepsilon$ .

The algorithm includes two functions,  $\text{eps\_entropy}(\mathcal{T}, \mathcal{T}_\varepsilon)$  and  $\text{eps\_capacity}(\mathcal{T}, \varepsilon)$  which are defined as follows.

**Function 1.**  $\text{eps\_entropy}(\mathcal{T}, \mathcal{T}_\varepsilon)$

**Input:** Set  $\mathcal{T} = \{t_0, t_1, t_2, \dots, t_m\}$  of time moments,  $t_j < t_{j+1}$ ,  $j = 0, 1, 2, \dots, m - 1$ ; set  $\mathcal{T}_\varepsilon = \{t_{\varepsilon,0}, t_{\varepsilon,1}, t_{\varepsilon,2}, \dots, t_{\varepsilon,m_\varepsilon}\}$  of time moments,  $t_{\varepsilon,j} < t_{\varepsilon,j+1}$ ,  $j = 0, 1, 2, \dots, m_\varepsilon - 1$ .

**Output:**  $\varepsilon$ -entropy  $\mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon)$  of the set  $\mathcal{T} \vee \mathcal{T}_\varepsilon$ .

1. Join the sets  $\mathcal{T}$  and  $\mathcal{T}_\varepsilon$ :  $\mathcal{T}_{\text{joint}} = \mathcal{T} \cup \mathcal{T}_\varepsilon$ .
2. Find the number  $N(\mathcal{T}_{\text{joint}})$  of elements in the set  $\mathcal{T}_{\text{joint}}$ .
3. Set  $\mathcal{N}_\varepsilon(\mathcal{T}_{\text{joint}}) = N(\mathcal{T}_{\text{joint}}) - 1$ .
4. Set  $\mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon) = \log_2 \mathcal{N}_\varepsilon(\mathcal{T}_{\text{joint}})$ .
5. Return  $\mathcal{H}_\varepsilon(\mathcal{T} \vee \mathcal{T}_\varepsilon)$ .

The function  $\text{eps\_entropy}$  was implemented in MATLAB® by concatenation of the sets  $\mathcal{T}$  and  $\mathcal{T}_\varepsilon$  using the function `cat` with further removing of the doubling elements by the function `unique`.

**Function 2.**  $\text{eps\_capacity}(\mathcal{T}, \varepsilon)$

**Input:** Set  $\mathcal{T} = \{t_0, t_1, t_2, \dots, t_m\}$  of time moments,  $t_j < t_{j+1}$ ,  $j = 0, 1, 2, \dots, m - 1$ ; radius  $\varepsilon > 0$ .

**Output:**  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(\mathcal{T})$  of the set  $\mathcal{T}$ .

1. If  $(t_m - t_0) \leq \varepsilon$  then
2.   Set  $\mathcal{M}_\varepsilon(\mathcal{T}) = 1$ .
3. Else
4.   Set  $\mathcal{M}_\varepsilon(\mathcal{T}) = 2$ .
5.   Set  $j = 0$ .
6.   For  $i = 1$  to  $m - 1$  do:
7.       If  $(t_i - t_j) \leq \varepsilon$  or  $(t_m - t_i) \leq \varepsilon$  then
8.           Continue.
9.       Else
10.          Set  $\mathcal{M}_\varepsilon(\mathcal{T}) = \mathcal{M}_\varepsilon(\mathcal{T}) + 1$ .
11.          Set  $j = i$ .
12.       End if.
13.   End for.
14. End if.
15. Set  $\mathcal{E}_\varepsilon(\mathcal{T}) = \log_2 \mathcal{M}_\varepsilon(\mathcal{T})$ .

16. Return  $\mathcal{E}_\varepsilon(\mathcal{T})$ .

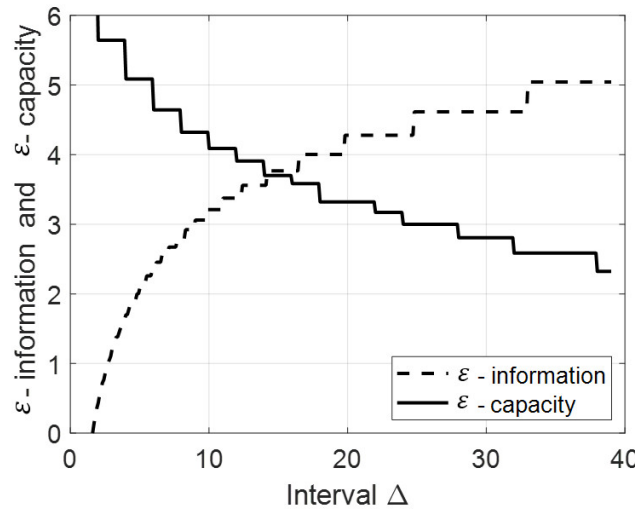
The function *eps\_capacity* computes the number  $\mathcal{M}_\varepsilon(\mathcal{T})$  of  $\varepsilon$ -distinguishable elements in the set  $\mathcal{T}$  for given  $\varepsilon$  and then computes  $\mathcal{E}_\varepsilon(\mathcal{T})$  as  $\log_2$  of this number.

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

$$\mathcal{C} = n \times \mathcal{O}(m \log m). \quad (33)$$

Convergence of Algorithm 1 is guaranteed by the indicated above fact that  $\varepsilon$ -information  $\mathcal{I}_\varepsilon(\mathcal{T})$  increases with increasing  $\varepsilon$  while  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(\mathcal{T})$  decreases with increasing  $\varepsilon$ . Since the interval  $T = [0, t_m]$  is bounded, the difference between increasing  $\varepsilon$ -information  $\mathcal{I}_\varepsilon(\mathcal{T})$  and decreasing  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(\mathcal{T})$  has its minimum in  $T$ , which is a terminating point of the algorithm.

Dependence of the functions  $\mathcal{I}_\varepsilon(\mathcal{T})$  and  $\mathcal{E}_\varepsilon(\mathcal{T})$  on the interval length  $\Delta = 2\varepsilon$  is illustrated in Figure 1.



**Figure 1.** Dependence of  $\varepsilon$ -information  $\mathcal{I}_\varepsilon(\mathcal{T})$  and  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(\mathcal{T})$  on the interval length  $\Delta = 2\varepsilon$  for the set  $\mathcal{T}$  of  $m = 100$  evenly distributed time moments;  $T = [0, t_m]$ ,  $t_m = 100$  and  $s = 1$ .

The computed interval is  $\Delta = 14.21$ . For this interval and  $\varepsilon = \Delta/2 = 7.10$ , the values of  $\varepsilon$ -information and  $\varepsilon$ -capacity are  $\mathcal{I}_\varepsilon(\mathcal{T}) = \mathcal{E}_\varepsilon(\mathcal{T}) \approx 3.7$  bit. Note that the accuracy of computing the interval  $\Delta$  increases with decreasing the step  $s$ .

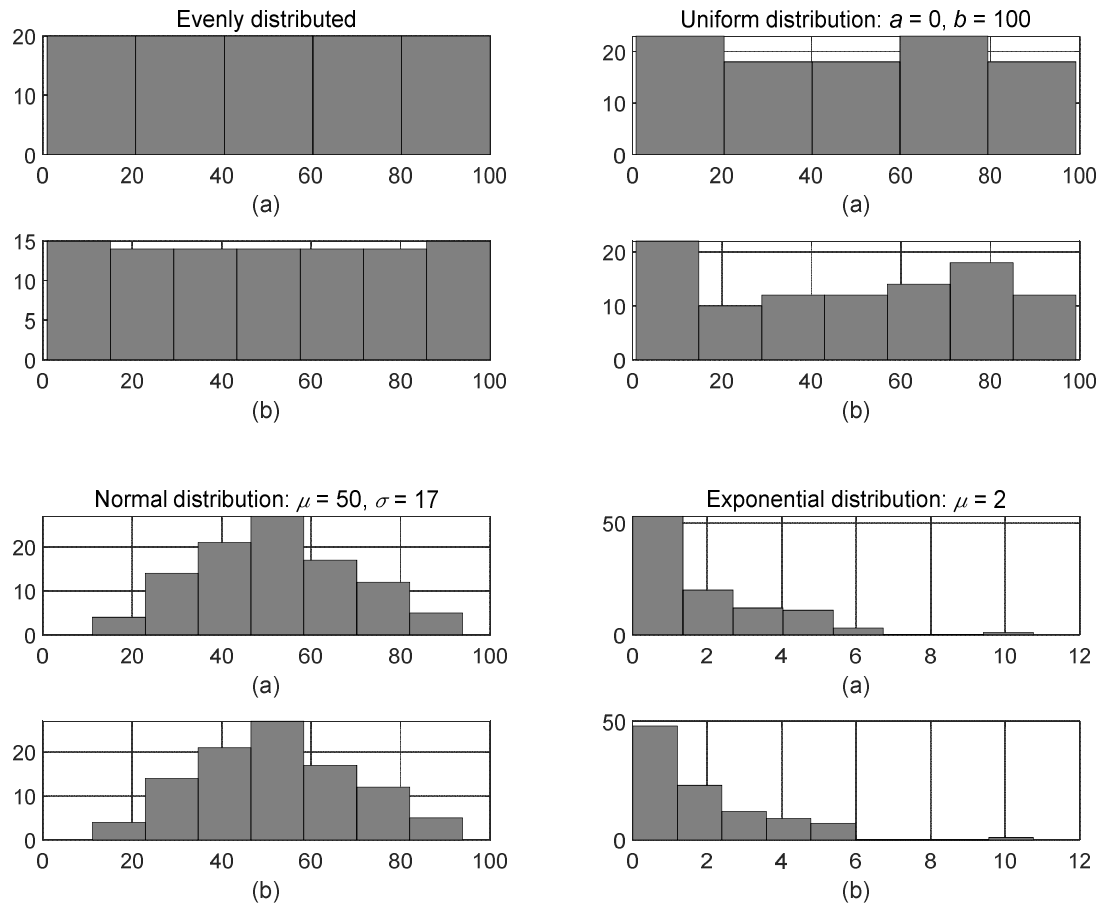
## 6. Examples

First, let us consider the examples of computing the interval lengths for different distributions of time intervals. In all considered cases we assume that the length of time interval  $T = [0, t_m]$  is  $t_m = 100$  and  $m = 100$ .

The data were generated by the MATLAB® function *random* with respect to the distribution created by the MATLAB® function *makedist*. In the examples, we used uniform distribution with  $a = 0$  and  $b = t_m$ , normal distribution with  $\mu = t_m/2$  and  $\sigma = t_m/6$ , and exponential distribution with  $\mu = 2$ .

The obtained interval lengths  $\Delta$  were used as bin lengths  $\delta$  in the histograms. For comparison, we present the histograms plotted with the bin lengths  $\delta$  calculated using the Scott rule (see

equation (3)), which is also used as a basis for a default method in MATLAB®. The resulting histograms are shown in Figure 2.



**Figure 2.** Histograms of the data plotted using the bin lengths computed by the Scott rule (figures (a) for each distribution) and using the bin lengths  $\delta = \Delta$  computed by the suggested algorithm (figures (b) for each distribution).

The values of the interval lengths  $\Delta$  are:

- evenly distributed data:

$$\Delta = 14.21, \delta_1 = 9.90, \delta_2 = 12.95, \delta_3 = 21.81 \text{ and } \delta_4 = 21.54,$$

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

$$\Delta = 14.11, \delta_1 = 9.83, \delta_2 = 12.87, \delta_3 = 22.69 \text{ and } \delta_4 = 22.65,$$

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

$$\Delta = 11.91, \delta_1 = 8.27, \delta_2 = 10.82, \delta_3 = 13.11 \text{ and } \delta_4 = 10.38,$$

- exponential distribution with  $\mu = 2$ :

$$\Delta = 1.20, \delta_1 = 1.07, \delta_2 = 1.41, \delta_3 = 1.35 \text{ and } \delta_4 = 0.91.$$

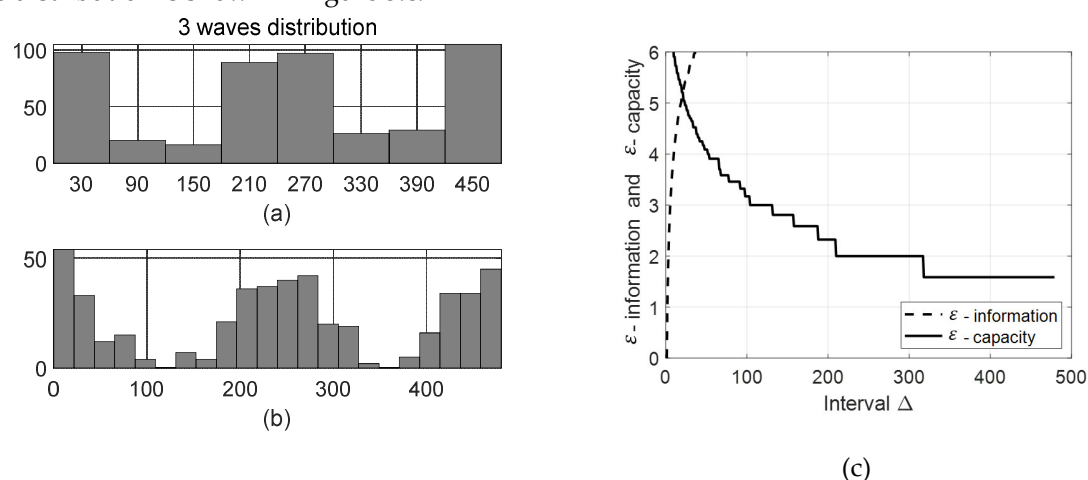
The suggested method results in the interval lengths  $\Delta$  that are close to the interval lengths  $\delta$  provided by the conventional methods with respect to the distribution of the data. In fact, for evenly and uniformly distributed data interval length  $\Delta$  is close to the interval length  $\delta_2$  resulted by the Sturges method, for normal distribution  $\delta_2 < \Delta < \delta_3$  and for exponential distribution  $\delta_1 < \Delta < \delta_2$ .

Now let us consider the use of the suggested algorithm for specification of the arrival rates  $\lambda$  and corresponding service rates  $\mu$ . Assume that the office, where the mentioned above clerk works, serves 480 clients 8 hours during the day that is  $T = 8 \times 60 = 480$  minutes. Also, assume that the clients arrive by three “waves” – in the morning, in the midday and in the evening. The histogram of the number of clients during the day is shown in Figure 3.a. In this histogram the bin length is computed by the Scott rule (value  $\delta_3$  below).

The values of the interval lengths  $\Delta$  for this distribution are:

$$\Delta = 22.0, \delta_1 = 21.91, \delta_2 = 48.45, \delta_3 = 67.99 \text{ and } \delta_4 = 66.41.$$

Histogram of the number of clients during the day with the bin length  $\delta = \Delta$  computed by the suggested algorithm is shown in Figure 3.b. Dependence of the functions  $\mathcal{I}_\varepsilon(A)$  and  $\mathcal{E}_\varepsilon(A)$  on  $\varepsilon$  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  $\delta = \Delta$  computed by the suggested algorithm; (c) dependences of  $\varepsilon$ -information  $\mathcal{I}_\varepsilon(A)$  and  $\varepsilon$ -capacity  $\mathcal{E}_\varepsilon(A)$  on  $\varepsilon$  for the set of arrival times.

From the results of computations of the interval length  $\Delta$  and the bin lengths  $\delta$  it follows that by the suggested algorithm the arrival rates during a day should be calculated each 22 minutes, while by the Scott they should be calculated each 68 minutes. Thus, for multimodal distribution the suggested algorithm results in shorter intervals that provides more exact representation of the data.

## 7. Conclusion

In the paper, we suggested the method of calculating optimal time intervals required for definition of arrival and departure rates. The method is useful for specification of the bin lengths in histograms, especially for the data with multimodal distributions.

The method utilizes the Kolmogorov and Tikhomirov  $\varepsilon$ -entropy and  $\varepsilon$ -capacity and the Rokhlin entropy of partition. Optimality of the partition is defined basing on the  $\varepsilon$ -information.

The procedure is presented in the form of a ready-to-use algorithm, which was compared with the known methods used for calculation of the interval lengths in histograms and demonstrated its robustness and correct sensitivity to the data.

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

**Competing interests:** The authors declare no competing interests.

## References

1. Dinaburg E. I. On the relations among various entropy characteristics of dynamical systems. *Math. USSR Izvestija*, 1971, 5(2), 337-378.

2. Freedman D., Diaconis P. On the histogram as a density estimator:  $L_2$  theory. *Zeit. Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 1981, 57, 453-476.
3. Gross D., Shortle J. F., Thompson J. M., Harris C.M. *Fundamentals of Queueing Theory*, 4th ed. John Wiley & Sons: Hoboken, NJ, 2008.
4. Jacod J., Protte P. *Discretization of Processes*. Springer: Berlin, 2012.
5. Keller J. B. Time-dependent queues. *SIAM Review*, 1982, 24(4), 401-412.
6. Kolmogorov A. N., Tikhomirov V. M.  $\varepsilon$ -entropy and  $\varepsilon$ -capacity of sets in functional spaces. *Amer. Mathematical Society Translations, Ser. 2*, 1961, 17, 277-364.
7. Lawler G. F. *Introduction to Stochastic Processes*. Chapman & Hall: New York, 1995.
8. Newell G. F. Queues with time-dependent arrival rates (I-III). *J. Applied Probability*, 1968, 5(2), 436-451 (I); 5(3), 579-590 (II); 5(3), 591-606 (III).
9. Reis Y. Private conversation. Ariel University, Ariel, March 2024.
10. Rokhlin V. A. New progress in the theory of transformations with invariant measure. *Russian Mathematical Surveys*, 1960, 15(4), 1-22.
11. Schwarz G. Estimating the dimension of a model. *Annals of Statistics*, 1978, 6(2), 461-464.
12. Scott D. W. On optimal and data-based histograms. *Biometrika*, 1979, 66, 605-610.
13. Shannon C. A mathematical theory of communication. *The Bell System Technical Journal*, 1948, 27(3), 379-423.
14. Sinai Y. G. *Topics in Ergodic Theory*. Princeton University Press: Princeton, 1993.
15. Sturges H. The choice of a class-interval. *J. Amer. Statistics Association*, 1926, 21, 65-66.
16. 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.