COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS



Numerical estimation of generalized Rényi entropies

DIPLOMA THESIS

2026 Bc. Máté Nagy

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

Numerical estimation of generalized Rényi entropies

DIPLOMA THESIS

Study programme:	Mathematical Economics, Finance and Modeling
Field of Study:	Mathematics
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Supervisor:	RNDr. Anna Krakovská, PhD.

Bratislava 2026 **Bc. Máté Nagy**





Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Máté Nagy
Study programme:	Mathematical Economics, Finance and Modelling (Single degree study, master II. deg., full time form)
Field of Study:	Mathematics
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Numerical estimation of generalized Rényi entropies

Annotation: Rényi's generalized entropies are a tool for estimating the degree of uncertainty or information complexity of the investigated processes. They are an extension of the classic Shannon entropies and provide a richer view of data and information. Estimating generalized entropies can be complex and often computationally demanding. The simplest way is to discretize the data and then use the corresponding histogram to estimate the Rényi entropy. There are a number of other approaches, but they are problematic and regularly criticized. Therefore, the number of methods for estimating entropies is constantly increasing. Our proposal is a generalization of a simple and fast technique, which is based on evaluating the distances of the nearest neighboring points in the state space of the system under study. It was published in connection with the estimation of the second-order fractal dimension, but the generalization to Rényi entropies is a natural next step that we would like to explore and compare with standard approaches. The methods will be tested on time series from multifractal processes. In the more distant horizon, the motivation is the effective application of entropies in the search for causal relationships in a number of recorded real time series, for example in a macroeconomic context.

Supervisor:	RNDr. Anna Krakovská, CSc.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	doc. Mgr. Igor Melicherčík, PhD.
Assigned:	09.01.2025
Approved:	15.01.2025

prof. RNDr. Daniel Ševčovič, DrSc.
Guarantor of Study Programme

.....
Student

.....
Supervisor





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Máté Nagy
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Numerical estimation of generalized Rényi entropies
Numerický odhad Rényiho zovšeobecnených entropií

Anotácia: Rényiho zovšeobecnené entropie sú nástrojom na odhad miery neistoty alebo informačnej zložitosti skúmaných procesov. Sú rozšírením klasických Shannonových entropií a poskytujú bohatší pohľad na dáta a informácie. Odhadovať zovšeobecnené entropie môže byť zložité a často aj výpočtovo náročné. Najjednoduchším spôsobom je diskretizovať dáta a príslušný histogram potom použiť na odhad Rényiho entropie. Existuje množstvo ďalších prístupov, ale všetky sú problematické a neustále kritizované. Preto metód na odhad entropií neustále pribúda. Naším návrhom je zovšeobecnenie jednoduchej a rýchlej techniky, ktorá je založená na vyhodnotení vzdialeností najbližších susedných bodov v stavovom priestore skúmaného systému. Publikovaná bola v súvislosti s odhadom fraktálnej dimenzie druhého rádu, zovšeobecnenie pre Rényiho entropie je ale prirodzeným ďalším krokom, ktorý by sme chceli preskúmať a porovnať so štandardnými prístupmi. Metódy budú testované na časových radoch z multifraktálnych procesov. Vo vzdialenejšom horizonte je motiváciou efektívna aplikácia entropií pri hľadaní príčinných vzťahov v množstve zaznamenávaných reálnych časových radov, napríklad v makroekonomickom kontexte.

Vedúci: RNDr. Anna Krakovská, CSc.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: doc. Mgr. Igor Melicherčík, PhD.

Dátum zadania: 09.01.2025

Dátum schválenia: 15.01.2025

prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Declaration I hereby declare that I have prepared the entire diploma thesis on the topic "Numerical estimation of generalized Rényi entropies", including all its appendices and figures, independently, using the literature listed in the attached bibliography and artificial intelligence tools. I declare that I have used artificial intelligence tools in accordance with the applicable legal regulations, academic rights and freedoms, as well as ethical and moral principles, while maintaining academic integrity, and that their use is appropriately indicated in the thesis.

Bratislava, 2026

. . . . . .

Bc. Máté Nagy

Acknowledgements : I would like to thank RNDr. Anna Krakovská, PhD. for her supervision, willingness, and patience during my work on this diploma thesis. I am also thankful to my family for their constant support throughout my studies.

Abstrakt

Táto diplomová práca sa zameriava na numerický odhad zovšeobecnených Rényiho entropií z konečných dátových množín. Rényiho entropia preberá klasickú Shannonovu entropiu a rozširuje ju zavedením nového parametra q. To umožňuje podrobnejšiu charakterizáciu rozdelenia pravdepodobnosti a zložitých systémov, najmä tých, ktoré vykazujú multi-fraktálne správanie.

Vyskúšali sme tri rôzne odhady entropie: histogramovú metódu, metódu korelačného súčtu a odhad z k najbližších susedov. Pre každú metódu sme študovali presnosť, výpočtovú zložitosť a robustnosť. Kľúčovými charakteristikami vzorky údajov sú veľkosť vzorky, dimenzionalita a typ distribúcie.

Výkonnosť jednotlivých metód je hodnotená na rozdeleniach so známymi analytickými hodnotami Rényiho entropie, vrátane binárneho, rovnomerného, exponenciálneho a gaussovského rozdelenia ako referenčných štandardov. Potom sme sa pokúsili použiť tieto zistenia na odhadnutie entropií systémov s neznámymi hodnotami entropie, ale so známymi fraktálnymi dimenziami, ako sú Cantorova množina, Hénonovo zobrazenie, Lorenzov a Rösslerov chaotický systém.

Zohľadnili sme aj problémy týkajúce sa rozdielov medzi odhadom entropie v diskrétnych a spojitých rozdeleniach, vrátane použitia korekcie v prípadoch spojitej entropie.

Výsledky ukazujú, že histogramová metóda je výpočtovo efektívna, ale citlivá na výber parametrov. Metóda korelačného súčtu je použiteľná alternatíva, ale dáva horšie výsledky pri distribúciách s ťažšími chvostami, zatiaľ čo metóda k-NN poskytuje najstabilnejší výkon pri viacdimenzionálnych spojitých systémoch.

Táto práca poskytuje porovnanie Rényiho estimátorov entropie, pričom zdôrazňuje použiteľnosť a obmedzenia každej metódy.

Kľúčové slová: Rényiho entropia, numerický odhad entropie, histogramová metóda, korelačný súčet, k najbližších susedov

Abstract

This thesis focuses on the numerical estimation of generalized Rényi entropies from finite data sets. Rényi entropy takes the classical Shannon entropy and extends it by introducing a new parameter q. This enables a more detailed characterization of probability distributions and complex systems, especially the ones exhibiting multi-fractal behavior.

We tried three different entropy estimators: the binning (histogram-based) method, the correlation sum method, and the k-nearest neighbors (k-NN) estimator. For each method, we studied the accuracy, computational complexity, and robustness. Key characteristics of the data sample are the sample size, dimensionality, and distribution type.

The performance of the methods is evaluated on distributions with known analytical Rényi entropies, including binary, uniform, exponential, and Gaussian distributions as a benchmark. Then, we applied these findings to estimate the entropies of systems with unknown entropy values but known fractal dimensions, such as the Cantor set, Hénon map, Lorenz, and Rössler chaotic systems.

We also took into consideration the issues surrounding the differences between entropy estimation in discrete and continuous distributions, including the use of correction in continuous entropy cases.

The results show that the binning method is computationally efficient but sensitive to parameter choice. The correlation sum method is a viable alternative but suffers in distributions with heavier tails, while the k-NN method provides the most stable performance in higher-dimensional continuous systems.

This thesis provides a comparison of Rényi entropy estimators, highlighting the applicability and limitations of each method.

Keywords: Rényi entropy, numerical entropy estimation, binning method, correlation sum, k-nearest neighbors

Contents

Introduction **12**

1 Theoretical foundations of generalized Rényi entropies **14**

1.1 Shannon entropy 14

1.2 Rényi generalized entropy 14

1.3 Special cases 14

1.4 Other Rényi measures 15

1.4.1 Rényi dimensions 15

1.4.2 Rényi Entropy rates 16

1.4.3 Rényi divergences 16

1.5 Rényi differential entropy 17

2 Numerical estimation methods for generalized Rényi entropy **18**

2.1 Binning (histogram) method 18

2.1.1 Binning method: Methodology 18

2.1.2 Binning method: Implementation 18

2.2 Correlation sum method 19

2.2.1 Correlation sum method: Methodology 19

2.2.2 Correlation sum method: Implementation 20

2.3 Correlation sum method: Triplet-counting variant 20

2.4 k-NN method 21

2.4.1 k-NN method: Methodology 21

2.4.2 k-NN method: Implementation 22

3 Description of datasets **23**

3.1 Distributions with known analytical Rényi entropy 23

3.1.1 Binary distribution 23

3.1.2 Discrete uniform distribution 24

3.1.3 Discrete distribution over 20 values 24

3.1.4 Exponential distribution 25

3.1.5 Gaussian univariate distribution 26

3.1.6 Gaussian bivariate distribution 26
3.2 Distributions with unknown analytical Rényi entropy 27
3.2.1 Cantor set 27
3.2.2 Hénon map 28
3.2.3 Lorenz system 29
3.2.4 Rössler system 29

4 Results and discussion **31**
4.1 Binary distribution results 31
4.1.1 Binary distribution: Binning method 31
4.1.2 Binary distribution: Correlation sum method 32
4.1.3 Binary distribution: k-NN method 34
4.1.4 Binary distribution: Comparison of methods 34
4.2 Uniform distribution results 36
4.2.1 Uniform distribution (100 values): Binning method 36
4.2.2 Uniform distribution (100 values): Correlation sum method 38
4.2.3 Uniform distribution (100 values): k-NN method 41
4.2.4 Uniform distribution (100 values): Comparison of methods 42
4.3 Discrete distribution (20 values) results 44
4.3.1 Discrete distribution (20 values): Binning method 44
4.3.2 Discrete distribution (20 values): Correlation sum method 46
4.3.3 Discrete distribution (20 values): Comparison of methods 49
4.4 Exponential distribution results 50
4.4.1 Exponential distribution: Binning method 51
4.4.2 Exponential distribution: Correlation sum method 52
4.4.3 Exponential distribution: k-NN method 54
4.4.4 Exponential distribution: Comparison of methods 55
4.5 Gaussian distribution results 57
4.5.1 Gaussian distribution: Binning method 57
4.5.2 Gaussian distribution: Correlation sum method 58
4.5.3 Gaussian distribution: k-NN method 60
4.5.4 Gaussian distribution: Comparison of methods 61

4.6 Bivariate gaussian distribution results 63
4.6.1 Bivariate gaussian distribution: Binning method 63
4.6.2 Bivariate gaussian distribution: Correlation sum method 64
4.6.3 Bivariate gaussian distribution: k-NN method 66
4.6.4 Bivariate gaussian distribution: Comparison of methods 67
4.7 Cantor set results 69
4.8 Hénon map results 72
4.9 Lorenz system results 75
4.10 Rössler system results 78

Conclusion **82**

References **84**

Introduction

Information theory began with the work of Claude Shannon, more specifically with his 1948 paper, "A Mathematical Theory of Communication" [22]. The idea established a system for measuring concepts such as uncertainty, information, and complexity. The key concept of this system is entropy.

Depending on the field we are in, we can come across different ways of defining what entropy means. From the perspective of mathematics, the most common definition is that entropy measures the uncertainty of a random variable. An easy way to imagine this is, for example, if we try to predict a coin toss with a regular coin or a biased coin. The second event has lower entropy, as there is less uncertainty. Shannon was able to establish a mathematical formula that calculated this uncertainty. This had a massive effect on fields such as telecommunications, computer science, neuroscience, and economics.

However, like many different things in the world, Shannon entropy has its limitations. It is a reliable measure of the average information content, but it fails to capture the rich scaling structure present in many real-world systems, such as turbulent fluids, financial markets, or neural networks, especially in systems known as multi-fractals. In such systems, uncertainty manifests differently across scales.

There was a growing need for generalized entropies that would extend Shannon entropy by introducing new parameters. This enables us to look at uncertainty in a system from different angles. One of the most common generalizations is the Rényi entropy. The name comes from its inventor, Alfréd Rényi, who introduced this generalization of Shannon entropy in 1961 [17].

Rényi generalized Shannon entropy by introducing the parameter q, often called the "order". We will dive deeper into the significance of this parameter later in the work. For now, it is enough to know that this property will be crucial for analyzing multi-fractal systems, where the complexity of the system quantified by Rényi entropy reflects contributions from both common and rare events.

Rényi introduced other concepts as well, one of them is the Rényi dimensions [18] (or generalized dimensions) D_q. This measure generalizes the notion of fractal dimension by providing a spectrum of dimensions that describes the rich scaling behavior of complex systems.

In the real world, we need to estimate Rényi entropies from a finite data set. This leads us to the main objective of this thesis. We would like to study and compare several methods for estimating Rényi entropies, namely binning method, the correlation sum method, and the k-NN method [12].

These methods should give us consistent estimates without requiring prior knowledge of the underlying data structure. First, we compare the accuracy and computational complexity of the methods on data for which the theoretical values are well known (for example, data with binary, uniform, or Gaussian distributions). Subsequently, we will apply the methods to more complex systems (such as the Cantor set, Lorenz system, or Rössler system). For these, we do not know the theoretical Rényi entropy values, but we seek to obtain reasonable estimates.

Estimating Rényi entropies from empirical data presents certain challenges. The simplest approach, known as the histogram method, faces its biggest challenge in estimating the probability mass function $p(x)$. This problem could cause significant bias in certain cases.

These challenges have motivated the development of alternative approaches to substituting the probability mass function $p(x)$. We employ the approach that substitutes the entire sum of probabilities with correlation sums and, finally, look at a method that uses nearest-neighbors.

However, these methods can prove to be computationally demanding for larger datasets or real-time applications. We also consider this factor when evaluating each method. All the calculations are made using the R programming language [16].

Another great property of the Rényi entropy is that it is possible to extend it to continuous random variables. But contrary to what we would expect, we do not obtain the formula for the continuous case by discretizing the variable into finer and finer bins.

As shown by Sanei Tabass M., Mohtashami Borzadaran, G. R., and Amini, M. in [20], in the case of continuous distributions, an additional logarithmic correction term related to the bin size must be added when estimating Rényi entropies using the binning method.

1 Theoretical foundations of generalized Rényi entropies

1.1 Shannon entropy

The definition of Shannon entropy [22] for a discrete random variable X with N possible outcomes and a probability mass function $p(x)$ is:

$$H_1 = -\sum_{i=1}^{N} p_i \log p_i. \tag{1}$$

We can choose the base of the logarithm. The most common options are either 2 or the natural logarithm. The first option gives us measurements called bits, while the second gives us nats. The reason this definition of entropy is regularly used is that it reaches its maximum value for equally likely outcomes and a minimum for one outcome being certain. Moreover, its logarithmic form ensures a natural physical interpretation, notably the additivity property for independent systems.

1.2 Rényi generalized entropy

As mentioned previously, Alfréd Rényi, a Hungarian mathematician, generalized Shannon entropy by adding a parameter q, while including the original Shannon entropy as a special case. He defined the new entropy for a discrete random variable X and order q ($q \neq 1$) as [17]:

$$H_q(X) = \frac{1}{1-q} \log \left(\sum_{i=1}^{N} p_i^q \right). \tag{2}$$

Parameter q is essentially a "focusing parameter" that enables us to focus on different parts of the probability distribution.

1.3 Special cases

The framework of generalized Rényi entropy includes many special cases, such as:

- $q \to 0$: Hartley entropy

$$H_0(X) = \log |N| \tag{3}$$

- $q \to 1$: Shannon entropy

$$\lim_{q \to 1} H_q(X) = H(X) = -\sum_{i=1}^{N} p_i \log p_i, \tag{4}$$

- $q = 2$: Correlation (collision) entropy

$$H_2(X) = -\log \sum_{i=1}^{N} p_i^2 \tag{5}$$

- $q \to \infty$: Min-entropy

$$H_\infty(X) = -\log \max_{i \in N} p_i \tag{6}$$

  Min-entropy has its use in theoretical computer science. It emphasizes the most predictable outcome.

1.4 Other Rényi measures

Alongside a measure for entropy, Alfréd Rényi has also defined several other measures that help us better understand the structure of probability distributions in complex systems. We briefly mention them for completeness, but their estimation is beyond the scope of this work.

1.4.1 Rényi dimensions

As a natural follow-up to Rényi entropy, the concept of Rényi dimensions (also known as generalized dimensions) D_q is defined as:

$$D_q = \frac{1}{q-1} \lim_{r \to 0} \frac{\log \sum_{i=1}^{N(r)} p_i^q}{\log r}.$$

A useful property of Rényi dimensions is that they are constant for mono-fractals; however, this is not the case for multi-fractals. It is also worth noting that D_1 corresponds to the information dimension, D_2 is the correlation dimension, and D_0 is the box-counting (capacity) dimension.

A relation between the Rényi dimension and the Rényi entropy is expressed via the scaling of the entropy with the resolution r:

$$D_q = -\lim_{r \to 0} \frac{H_q(r)}{\log r}, \qquad q \neq 1.$$

It means that we can approximate D_q from H_q as follows

$$D_q \approx -\frac{H_q(r_2) - H_q(r_1)}{\log r_2 - \log r_1} = -\frac{H_q(r_2) - H_q(r_1)}{\log\left(\frac{r_2}{r_1}\right)}, \qquad q \neq 1 \tag{7}$$

where $H_q(r_1)$ and $H_q(r_2)$ are computed for small resolutions r_1, r_2.

In the opposite direction, the asymptotic relation is

$$H_q(r) \approx -D_q \log r, \qquad r \to 0,\ q \neq 1.$$

Obviously, Rényi entropies measure the uncertainty of a probability distribution at a given scale, while Rényi dimensions quantify how that uncertainty scales with resolution, reflecting the effective degrees of freedom or the fractal structure of the underlying system.

1.4.2 Rényi Entropy rates

Another concept that has proven to be important for dynamical systems and time series is the Rényi entropy rate. The Rényi entropy rate of order q is defined as:

$$K_q = \lim_{n\to\infty} \lim_{r\to 0} \frac{1}{n} H_q(n, r),$$

where $H_q(n, r)$ is the Rényi entropy for blocks of length n at resolution r.

For the limit $q \to 1$, the Rényi entropy rate converges to the well-known Kolmogorov–Sinai entropy.

Entropy rates differ from standard entropies computed for a static distribution. K_q measures the average information produced per time step in a process, which makes it crucial for understanding how complexity unfolds over time.

1.4.3 Rényi divergences

As the final measure from the Rényi-based family, we mention the Rényi divergence. We can understand this concept as the difference between two probability distributions P_1 and P_2. For order $q \neq 1$, it is defined as:

$$D_q(P_1 \| P_2) = \frac{1}{q-1} \log \sum_{i=1}^{N} \frac{p_{1_i}^q}{p_{2_i}^{q-1}}.$$

The case $q \to 1$ is known as the classical Kullback–Leibler divergence.

1.5 Rényi differential entropy

While the Rényi entropy, defined in Equation (2), was formulated for discrete distributions, several of our test data follow continuous distributions. For a continuous random variable X with a probability density function $f(x)$, entropy is defined in terms of the so-called differential entropy:

$$H_q(X) = \frac{1}{1-q} \log \left(\int_{-\infty}^{\infty} f(x)^q \, \mathrm{d}x \right).$$

Estimates of the differential entropy diverge with increasing resolution of the distribution, but there is a way to compensate for this divergence.

Namely, the Rényi entropy of a continuous distribution can be approximated as

$$H_q(X) \approx H_q(X_{\text{discrete}}) + \log \Delta,$$

where Δ is the size of the intervals (bins) represented by the approximating discrete outcomes $\{x_1, x_2, \ldots, x_N\}$, and $H_q(X_{\text{discrete}})$ is the Rényi entropy of the corresponding discrete distribution [4, 23].

Extending into the d–dimensional situation is straightforward:

$$H_q^{(d)}(X) \approx H_q(X_{discrete}^{(d)}) + d \log \Delta. \tag{8}$$

To sum up, in contrast to the entropy of a discrete random variable, differential entropy lacks a straightforward interpretation and is not an invariant measure by itself. However, the diverging term $\log \Delta$ is purely a resolution effect. It depends only on the bin width. When taking differences of differential entropies (like in the estimation process of Rényi entropy rates), the $\log \Delta$ terms cancel, and the resulting measures become invariant under changes of measurement resolution.

2 Numerical estimation methods for generalized Rényi entropy

In this section, we discuss the three main approaches we chose to estimate Rényi entropy and the concrete algorithms we implemented for them. The methods are the binning (histogram) method, the correlation sum method, and the k-nearest neighbors method.

2.1 Binning (histogram) method

The binning method, also known as the histogram-based or plug-in estimator, is the simplest approach to estimating Rényi entropies. The main idea is to create a finite number of discrete bins and simply divide the data into them.

2.1.1 Binning method: Methodology

We take a data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of points in R^d and divide it into hypercubes (bins) of equal size. Then the probability p_i is simply the estimated relative frequency of a data point falling into the i-th bin:

$$p_i \approx \frac{\text{count}_i}{N}, \tag{9}$$

where count_i is the number of data points within the i-th bin. After plugging these probabilities into the discrete definition, we get the Rényi entropy estimate of order q:

$$H_q \approx \frac{1}{1-q} \log \left(\sum_{i=1}^{N^d} p_i^q \right). \tag{10}$$

Note that we are only including bins with $p_i > 0$.

2.1.2 Binning method: Implementation

We are using a custom R function to obtain the estimation for this method. It is fitted to work with data of up to 4 dimensions. The input argument for the function is the signal (vector or matrix), the number of bins per dimension n, and a sample size; in other words, the length of the time series we would like to work with, `Length`.

If the specified `Length` is smaller than the size of the signal, then the function takes a random subsample.

After that, the range of the data is divided into $n+1$ breakpoints, creating n bins. To obtain empirical probabilities, the bin counts are normalized, and all the bins with $p_i < 0$ are removed. The function automatically computes Rényi entropy for q values from -10 to 10.

The case of $q = 1$ (index 12 in the code) is calculated according to the Shannon entropy formula $-\sum p_i \log p_i$.

For continuous cases, we created a separate function by adding a correction term. This term is based on the article [5] and is:

$$\log(\Delta) = \log((max(signal) - min(signal))/n) \tag{11}$$

Where Δ is the length of the bins, and it is calculated using n, which is the number of bins.

To obtain the results and visualize them, the following R packages were used in this diploma thesis [25], [21], [7], [26], [27], [28].

For enhanced data visualization and faster processing, AI tool [2] were used in this thesis.

2.2 Correlation sum method

The second method we will take a look at is a variation of the so-called correlation sum method. The basic idea is to substitute the sum of the estimated probabilities with the correlation sum $C_q(r)$.

$$\sum_{i=1}^{N} p_i^q = C_q(r) \tag{12}$$

2.2.1 Correlation sum method: Methodology

The idea came from the case $q = 2$, known as correlation entropy H_2, as Grassberger and Procaccia showed in the work [9], that $C_2(r)$ can be approximated as:

$$C_2(r) \approx \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|), \tag{13}$$

where N represents the number of points on the attractor, and Θ is a function that gives 1 for a positive argument and 0 for a non-positive argument, called the Heaviside function.

This case for $q = 2$ can be generalized, as shown in [15]. Then we get the correlation sum of order q, as follows:

$$C_q(r) \approx \frac{1}{N(N-1)^{q-1}} \sum_{i=1}^{N} \left[\sum_{j=1, j\neq i}^{N} \Theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|) \right]^{q-1} \tag{14}$$

The Rényi entropy is then estimated as:

$$H_q(r) = \frac{\log C_q(r)}{1 - q} \tag{15}$$

It would be possible to avoid the generalization in Equation (14) and instead calculate the distances between pairs, triples (which we will later try for $C_3(r)$), quadruples,..., and q-tuples of points. However, such an approach computes $C_q(r)$ only for integer values of q and is computationally demanding, as we will demonstrate using triplets as an example.

2.2.2 Correlation sum method: Implementation

A slight improvement compared to the original algorithm is sorting the data at the beginning, as suggested by AI [1]. This enables the while-loops to break once the distances exceed r. The function searches for neighbors within a radius r for each point i, both forwards $(i + j)$ and also backwards $(i - j)$ in the already sorted list.

After this, for each q values from -10 to 10 (excluding $q = 1$), the function calculates the correlation sum $C_q(r)$ and estimates H_q.

Using the equations above, we are unable to calculate the Shannon entropy case. Instead, we simply replace H_1 this time with the average of $H_{0.999}$ and $H_{1.001}$.

For continuous distributions, the correction term is not well defined.

2.3 Correlation sum method: Triplet-counting variant

Directly using the third-order correlation sum $C_3(r)$ is an interesting but computationally demanding variant of the correlation sum method. The methodology and the implementa-

tion are practically identical to the previous version in all but one step. Namely, instead of calculating the distance of one pair compared to a parameter r, here the maximal distance between three points is compared to r:

$$\max\left(\|\mathbf{x}_i - \mathbf{x}_j\|, \|\mathbf{x}_i - \mathbf{x}_k\|, \|\mathbf{x}_j - \mathbf{x}_k\|\right) \leq r$$

2.4 k-NN method

The k-nearest neighbors method is a non-parametric approach that estimates the probability density function around each data point. It is considered more robust than the binning method, especially in higher dimensions.

2.4.1 k-NN method: Methodology

The k-NN Shannon entropy estimator, also known as the Kozachenko-Leonenko method [11], is one of the most commonly used estimators. It estimates the probability function using the distance from each point to its k-th nearest neighbor and the volume of unit (d-dimensional) balls V_d in R^d. The estimator is written as [11]:

$$H(X) \approx -\psi(k) + \psi(N) + \log V_d + \frac{d}{N}\sum_{i=1}^{N} \log \epsilon(i),$$

where ψ is the *digamma function*, k is the number of nearest neighbors, and ϵ_i is twice the distance from x_i to its k^{th} nearest neighbor.

This estimator was later generalized by Leonenko in [12] for Rényi entropy estimation. Based on his work, we can use the following equations:

$$H_{N,k,q} \approx \frac{\log \hat{I}_{N,k,q}}{1-q},$$

where

$$\hat{I}_{N,k,q} = \frac{1}{N}\sum_{i=1}^{N} \left(\zeta_{N,i,k}\right)^{1-q},$$

$$\zeta_{N,i,k} = (N-1)\, C_k\, V_d\, \left(\rho_{k,N-1}(i)\right)^n,$$

$$C_k = \left[\frac{\Gamma(k)}{\Gamma(k+1-q)}\right]^{1/(1-q)}.$$

$\rho_{k,N-1}(i)$ is the distance from point X_i to its k-th nearest neighbor, and the volume of unit balls V_d in R^d is defined as follows:

$$V_m = \frac{\pi^{m/2}}{\Gamma(m/2+1)}.$$

2.4.2 k-NN method: Implementation

During the implementation of the k-NN method, the "get.knn" function was used from the package "FNN" [3]. This method was also improved by AI for better computational performance [1], however this did not impact the precision of the results. The algorithm computes the full Euclidean distance matrix at once and then extracts the required k-th nearest neighbor distances. $H(X)$ as defined above is subsequently calculated with the requested parameters. It is common to use the parameter $k \leq \sqrt{N}$.

An important note needs to be made at this point: the whole methodology is derived for continuous random variables, so this method does not require a bias correction to be added. On the other hand, it is not recommended to be applied to discrete data as it is.

3 Description of datasets

In this thesis, we evaluate three entropy estimation methods on various datasets. The data vary in many properties, including discrete and continuous distributions, low and high dimensionality, or mono-fractal and multi-fractal characteristics. We divide the datasets into two groups. Firstly, we discuss the datasets that we are able to benchmark using the known analytical values of entropy. After that, we move on to datasets for which there are no known analytical values.

3.1 Distributions with known analytical Rényi entropy

These datasets have the primary function of helping us understand the behavior, precision, and time complexity of our estimation methods.

3.1.1 Binary distribution

As the first dataset, we have used a simple discrete binary distribution. The dataset contains five one-dimensional series with different probabilities p = [0.1, 0.2, 0.3, 0.4, 0.5]. Each series contains 1,000,000 data points. This dataset was generated using the Matlab function `binornd`.

The analytical Rényi entropy for data with a binary distribution is calculated using:

$$H_q(X) = \frac{1}{1-q} \ln \left(p^q + (1-p)^q\right). \tag{16}$$

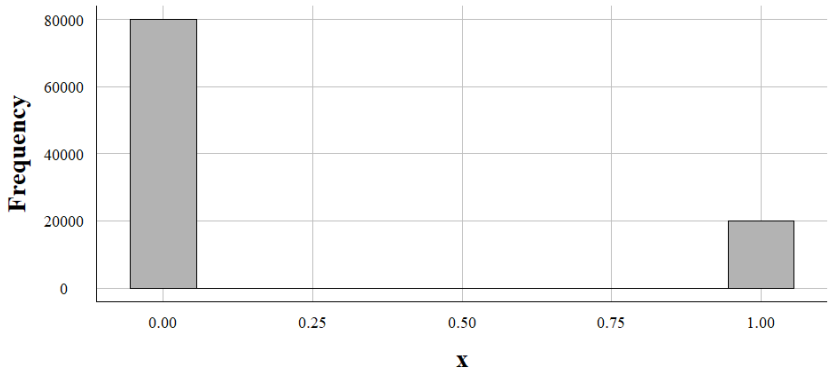


Figure 1: Histogram of binary data.

3.1.2 Discrete uniform distribution

The second dataset on which we tested our methods followed a uniform distribution. We generated four different one-dimensional time series using a custom R function `generate_discrete_ts` with differing means: 10, 100, 1000, 10000, corresponding to probabilities 0.1, 0.01, 0.001, 0.0001. Each series again contains 1,000,000 data points. For a uniform distribution with N outcomes, the analytical Rényi entropy is calculated using the equation $H_q(X) = \ln N$ for all q.

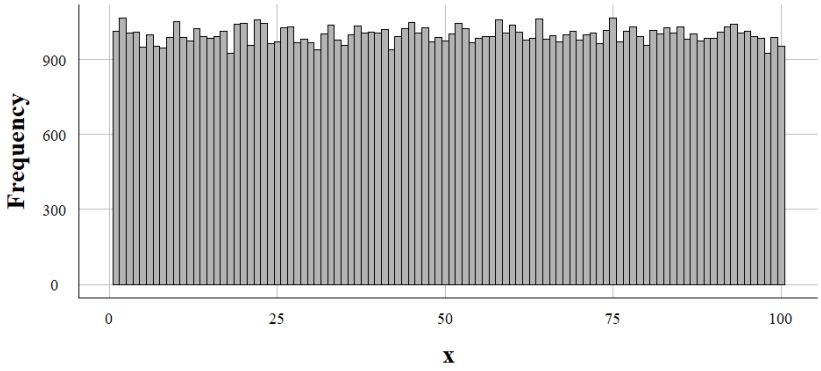


Figure 2: Histogram of a uniform distribution with 100 outcomes.

3.1.3 Discrete distribution over 20 values

Next, we test our methods on a dataset named 20rand, containing 1 million data points, each taking one of 20 possible values with different probabilities (see Table 1). The theoretical values can be calculated by adding the probabilities to the Rényi entropy equation:

$$H_q(X) = \frac{1}{1-q} \log \left(\sum_{i=1}^{20} p_i^q \right) \tag{17}$$

values	probability	values	probability	values	probability	values	probability
1	0.040915	6	0.017041	11	0.002249	16	0.020035
2	0.103857	7	0.006345	12	0.105954	17	0.033236
3	0.079964	8	0.094622	13	0.090937	18	0.057325
4	0.065398	9	0.065667	14	0.023196	19	0.047186
5	0.017044	10	0.077351	15	0.019863	20	0.031814

Table 1: 20 values and the corresponding probabilities in the 20rand dataset.

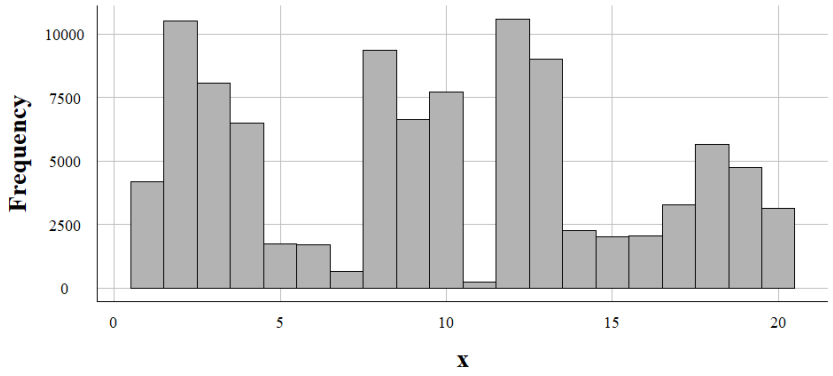


Figure 3: Histogram of a discrete distribution with 20 values of different probabilities.

3.1.4 Exponential distribution

The last of the distributions we use, for which the theoretical values are known, is the exponential distribution. Our time series was generated using the MATLAB function `exprnd` with exponent 1.0. Again, the dataset contains 1,000,000 data points, and the equation for calculating the analytical values is:

$$H_q(X) = -\ln(q) + \frac{\ln q}{(q-1)} \tag{18}$$

[5].

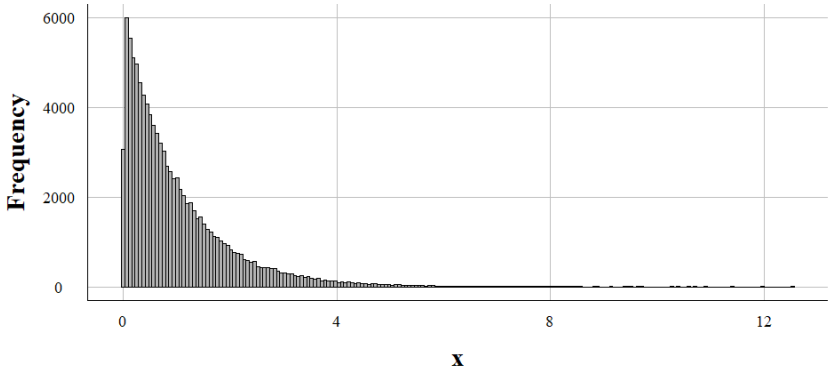


Figure 4: Histogram of exponential distribution.

3.1.5 Gaussian univariate distribution

The next dataset to use is one with a Gaussian distribution, containing 1,000,000 data points. First, a univariate version with parameters $\mu = 0$ and $\sigma = 1$. This time series was generated using the MATLAB function `randn`. The theoretical Rényi entropy can be calculated by

$$H_q(X) = \frac{1}{2}\ln(2\pi\sigma^2) - \frac{\ln q}{2(1-q)} \tag{19}$$

[8].

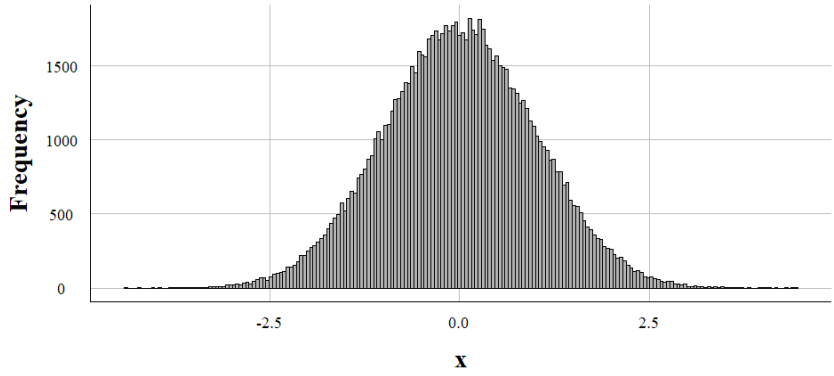


Figure 5: Histogram of the Gaussian distribution.

3.1.6 Gaussian bivariate distribution

We will also take a look at data from the bivariate version of the Gaussian distribution with a mean vector $\mu = [0, 0]$ and a covariance matrix $\sigma = [4, 1.5; 1.5, 4]$. As in the univariate case, the dataset contains 1,000,000 data points and was generated using the

same function. For the bivariate version, we can calculate the theoretical values using

$$H_q(X) = \frac{d}{2}\log 2\pi + \frac{d\log\alpha}{2(\alpha-1)} + \frac{1}{2}\log|\Sigma| \tag{20}$$

[14].

Under the assumption of any mean and a fixed variance, the Gaussian distribution maximizes the differential entropy [4].



Figure 6: Histograms of the x and y coordinates of the bivariate Gaussian distribution.

3.2 Distributions with unknown analytical Rényi entropy

We now introduce datasets with unknown entropy values, for which we aim to estimate the Rényi entropy using the already tested methods.

3.2.1 Cantor set

The Cantor set is a classic example of a fractal, infinitely detailed structure that is large in cardinality but small in measure. It is built by repeatedly removing the middle third of intervals. To approximate the Cantor set, we generated in R with an interval divided into 30 parts (length 1,000,000). For the Cantor set, we do not know the theoretical Rényi entropy values, but the theoretical Rényi dimensions are approximately $D_q = 0.63093$ for all q.

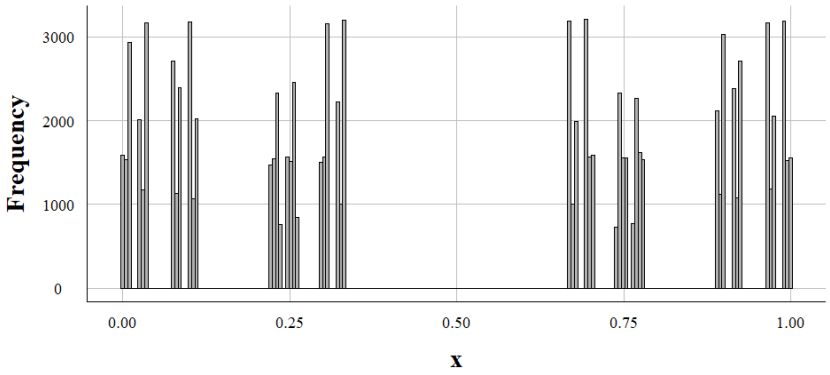


Figure 7: Histogram of Cantor distribution.

3.2.2 Hénon map

The Hénon map is a well-studied two-dimensional chaotic map known for its strange attractor. After starting with the initial condition $[0.7, 0]$, the first $1,000$ points were discarded, and the next $1,000,000$ two-dimensional state space vectors were saved. This Hénon map can be defined by this set of equations:

$$\begin{aligned} x(t+1) &= 1.4 - x^2(t) + 0.3y(t) \\ y(t+1) &= x(t) \end{aligned} \tag{21}$$

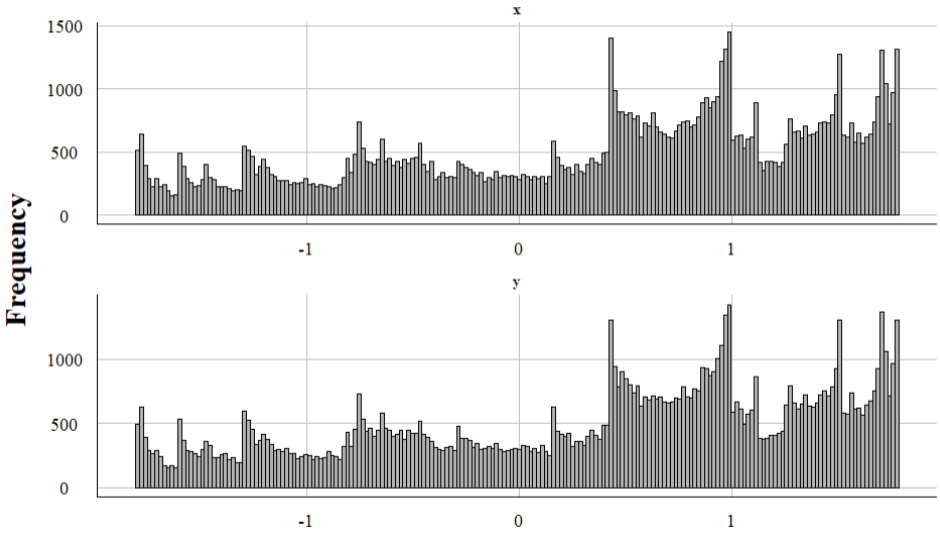


Figure 8: Histograms of the x and y coordinates of the Hénon map attractor.

3.2.3 Lorenz system

The Lorenz system is a three-dimensional system of ordinary differential equations, well known for chaotic behavior [13].

Specifically, we used the following equations:

$$\begin{aligned} \dot{x} &= 10(-x+y) \\ \dot{y} &= 28x - y - xz \\ \dot{z} &= xy - \frac{8}{3}z \end{aligned} \tag{22}$$

We obtained the data by numerical integration using the adaptive Runge-Kutta method in MATLAB. An integration step size of 0.1 and a starting point of $[0.3, 0.3, 0.3]$ were used. We dropped the first 1000 points and saved the next 1000000.

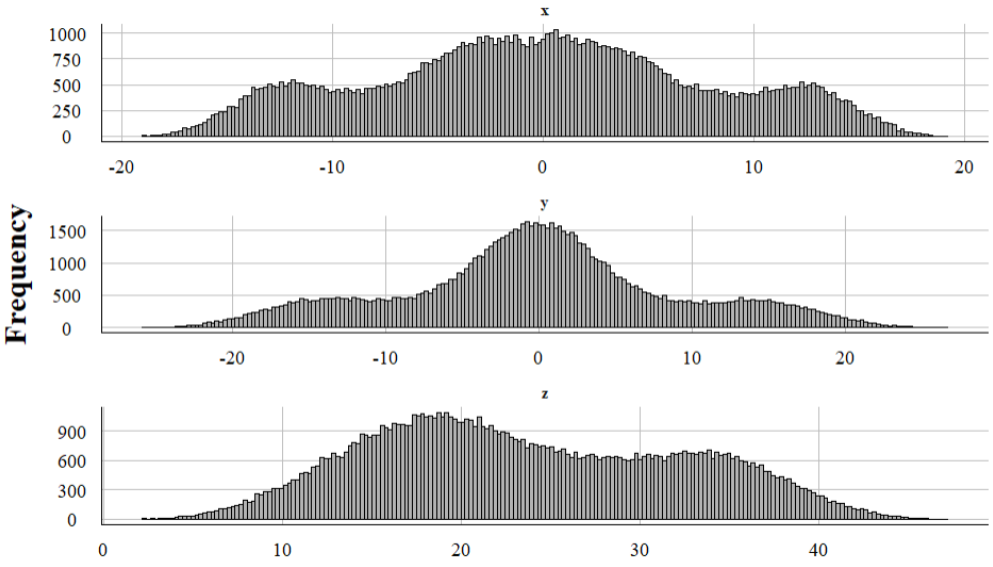


Figure 9: Histograms of the x, y, and z coordinates of the Lorenz attractor.

3.2.4 Rössler system

Another classic three-dimensional chaotic system we studied is the Rössler system [19].

The equations we used are as follows:

$$\begin{aligned} \dot{x} &= -6(y+z) \\ \dot{y} &= 6(x+0.2y) \\ \dot{z} &= 6(0.2+z(x-5.7)) \end{aligned} \tag{23}$$

We got our test time series as solutions to the equations derived using the adaptive fourth-order Runge-Kutta method in MATLAB. A starting point of $[0, 0, 0.4]$ and an integration step size of 0.1 were used. We dropped the first $1,000$ points and saved the next $1,000,000$.

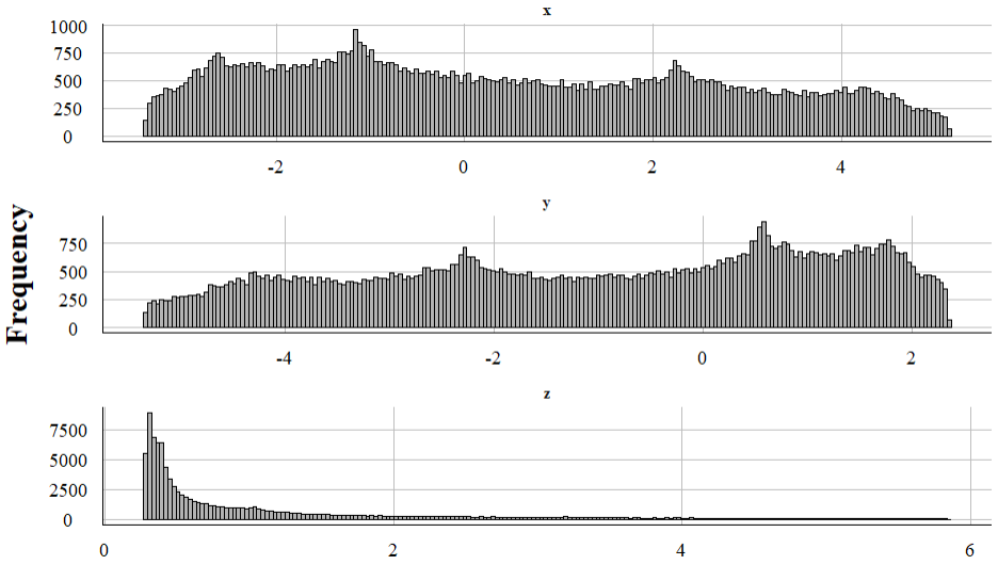


Figure 10: Histograms of the x, y, and z coordinates of the Rössler attractor.

4 Results and discussion

4.1 Binary distribution results

Firstly, we will take a look at the simplest of our data sets, which is data with a binary distribution. This will allow us to validate the method and understand what we can expect from our methods for data with discrete distributions.

4.1.1 Binary distribution: Binning method

In Figure 11, we can see that the results have the expected characteristics. For $p = 0.5$, we observe mono-fractal behavior, which is the expected shape for a distribution that we can also describe as uniform. For the rest of the curves, we see descending behavior with a larger slope near $q = 0$.

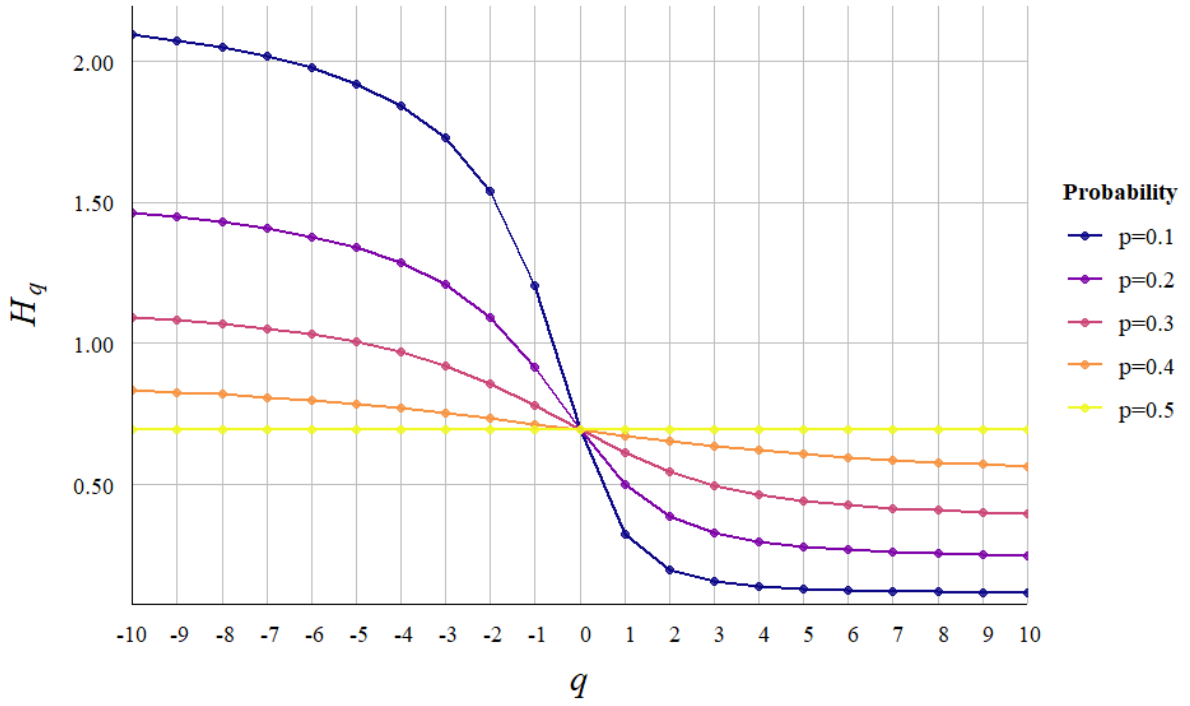


Figure 11: Estimated Rényi entropy for binary distributions with different probability parameters p and time series length 1,000,000 calculated with binning method and 2 bins.

This data set also has the smallest computational demand, as we can use the minimal number of bins $n = 2$ to compute the estimated Rényi entropy values. However, for other data sets, we may find ourselves in a situation where we cannot use as many data points

as we did in this run. The reasons can be computational or simply that we do not have enough data at our disposal. Because of this, it is worth taking a look at how precise our methods are with different time series lengths. In Figure 12, we see that the results for binary distribution are very precise, even for smaller data sets. The only visible difference is in the case of $q < 0$ for time series length 1,000. The reason for these inconsistencies is the significant variation of the sample data set. This can affect even the case where we have 1,000 data points.

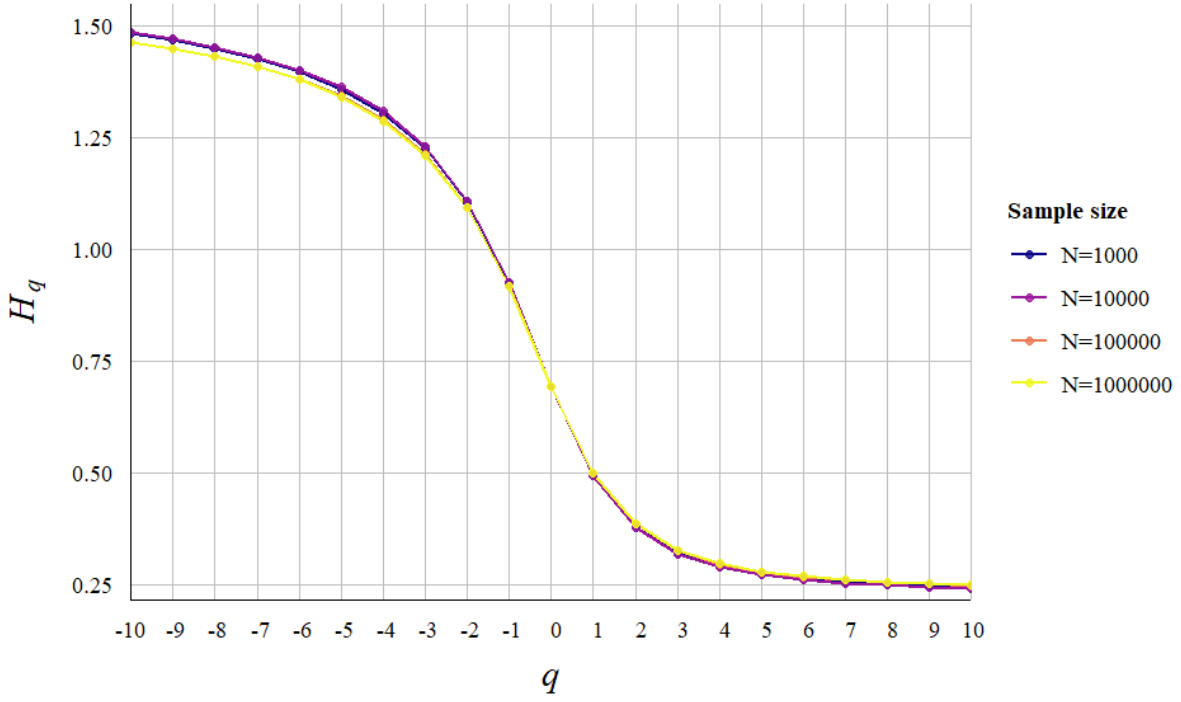


Figure 12: Rényi entropy estimates for binary distribution (p=0.2) with increasing time series length and 2 bins.

4.1.2 Binary distribution: Correlation sum method

The next method we will test binary data on is the correlation sum. The same shape is shown for this method in Figure 13, as it was for the binning method in Figure 11. Again, we see the s-shaped curves for all data except the one with $p = 0.5$. For $p = 0.5$, we have a uniform distribution, for which Rényi entropy is constant across all values of q, whereas heterogeneous distributions produce monotonically decreasing Rényi entropy curves.

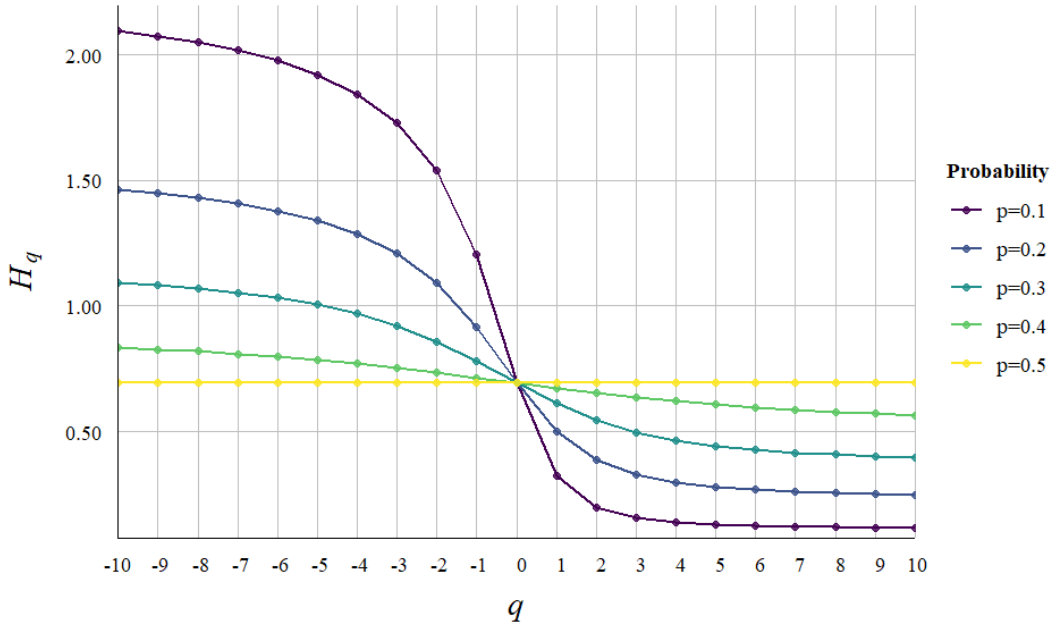


Figure 13: Rényi entropy for binary distributions with different probability parameters p and time series size 100,000 estimated using the correlation sum method with $r = 0.01$.

Taking a look at how the correlation sum method performs on the smaller data set available in Figure 14, we see that, again, similar to the binning method, there are noticeable differences only in the case of $q < 0$, but this method seems to be less precise, as we already observe the differences for 1,000 data points.

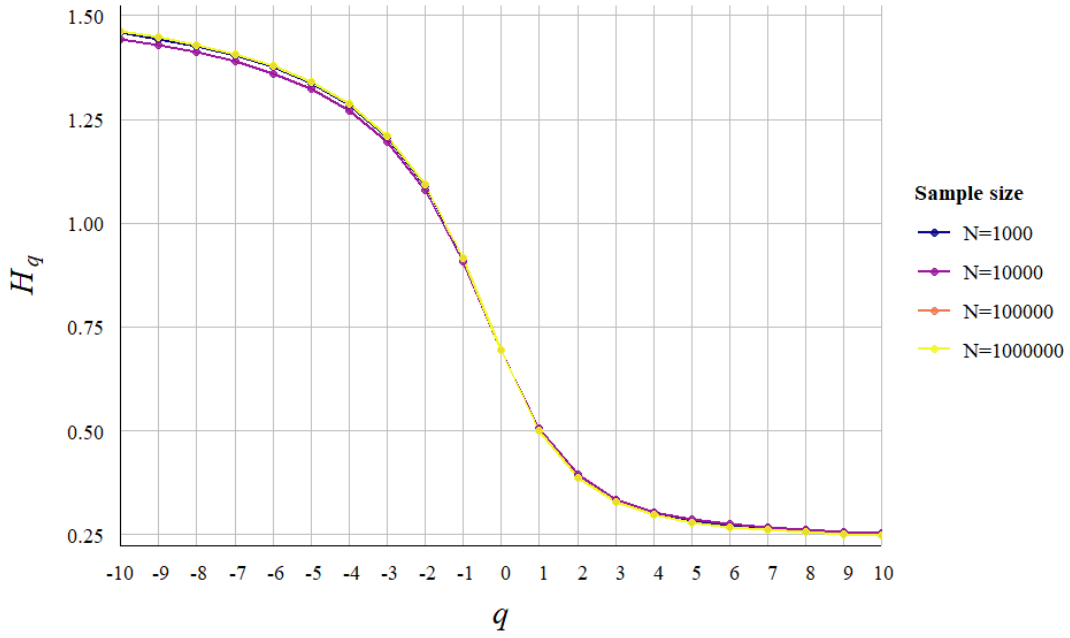


Figure 14: Rényi entropy estimates for binary distribution (p=0.2) with increasing time series length using the correlation sum method with $r = 0.01$.

4.1.3 Binary distribution: k-NN method

We seem to find the biggest weakness of the k-nearest neighbor method with our first data set. This method functions extremely inefficiently and even unusably for discrete data containing a large number of identical values. The reason is that the method calculates the distances between the nearest neighbors, and we would need to calculate for approximately $k = 500000$ nearest neighbors (for a data set of 1,000,000) to find the first non-zero distance. For this reason, we will not try this method on data with a binary distribution. However, we will showcase this weakness of the k-NN method on other data sets with a discrete distribution.

4.1.4 Binary distribution: Comparison of methods

We compare the estimated values with the theoretically calculated values. To assess the precision of our estimates, we use the Mean Absolute Error (MAE_q), defined as:

$$\text{MAE}_q = H_q^{\text{theoretical}} - H_q^{\text{estimated}}$$

where H_q is the number of entropy orders considered. And the average of mean absolute errors across all examined q values is:

$$\overline{MAE}_{q_{min}:q_{max}} = \frac{1}{q_n} \sum_{i=q_{min}}^{q_{max}} \text{MAE}_i,$$

where q_{min} is the smallest q value, q_{max} is the largest q value, and q_n is the number of q values.

We can see in Figure 15 as well as Table 2 that the correlation sum and binning methods are both very precise. However, the binning method is much faster for the same time series size, see Table 2. The triple-counting method is noticeably less precise, which may be due to the smaller time series length. We are unable to run the method with more data points because of its greater computational demand.

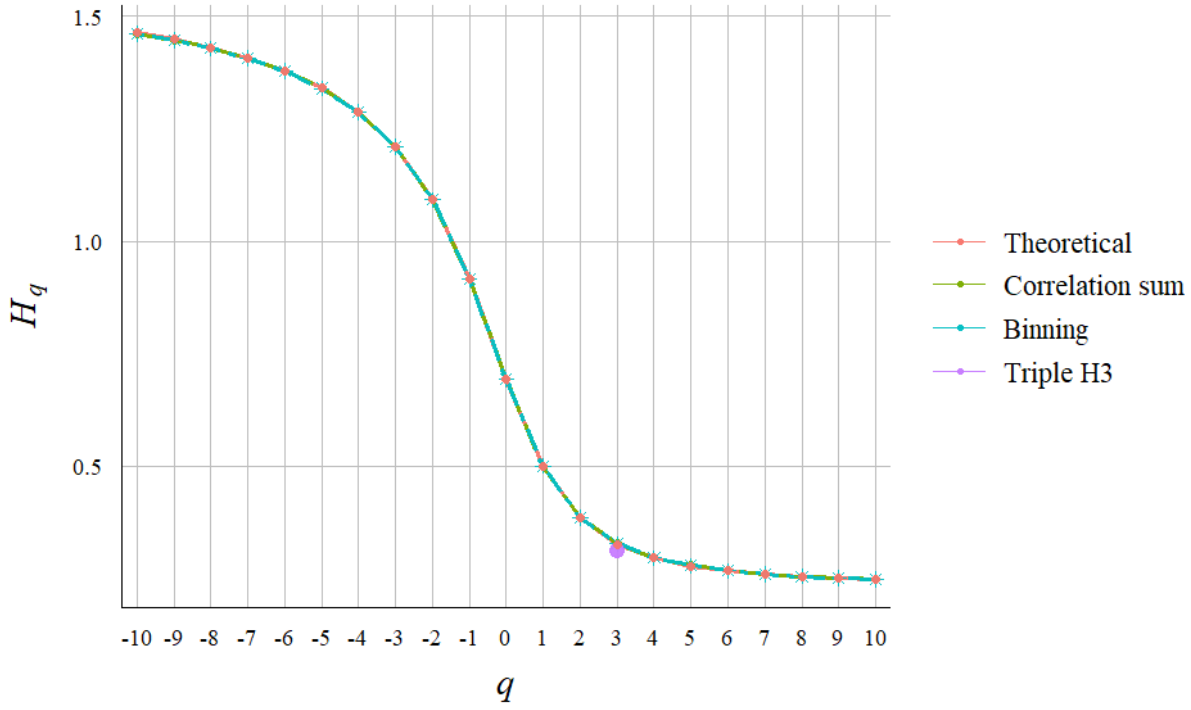


Figure 15: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 2, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000) and triplet counting method (r = 0.1, ts size 1,000) for binary data with probability 0.2.

Table 2: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 2, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000) and triplet counting method (r = 0.1, ts size 1,000) for binary data with probability 0.2.

Method	Runtime (seconds)	$\overline{MAE}_{-10:10}$
Correlation sum	51.1	0.000809
Binning (correction)	0.020	0.000825
Triple counting	18.2	0.0151

The runtime was recorded on a personal computer with an AMD Ryzen 5 7600 CPU, 32 gigabytes of RAM, and an AMD RX 6800 GPU.

4.2 Uniform distribution results

The next data set we will look at has a discrete uniform distribution with 100 unique values. This will allow us to observe the precision of each method for different q values. The theoretical values for this type of distribution are the same across all q values ($H_q = \ln 100 = 4.60517$).

4.2.1 Uniform distribution (100 values): Binning method

For the binning method, we see in Figure 16 that we obtain the expected result once we use 100 bins, and this result does not seem to change even if we increase the number of bins further. We observe another straight line for 50 bins, but between these two numbers, we see a different behavior than expected for another type of data. The reason is that in the case of 50 bins, the data is distributed in a way that resembles a uniform distribution with 50 unique values. Only once we reach the necessary number of bins, in our case 100, can the binning method fully capture the characteristics of the data set.

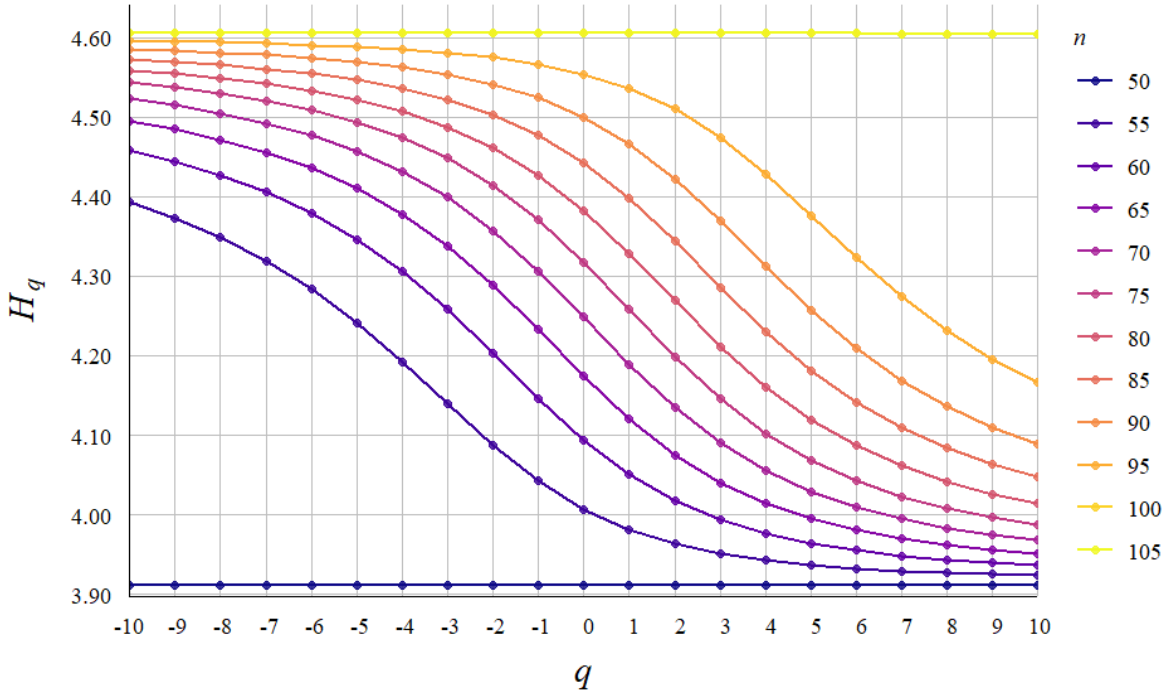


Figure 16: Rényi entropy estimates for uniform distribution (100 values) using binning method with time series size 1,000,000.

Once again, we see in Figure 17 that the results seem very precise once we reach a time series length of 10,000. For fewer data points, the function overestimates Rényi entropy for $q < 0$ and underestimates it for $q > 0$. For H_0, the results seem to be precise even for a data set of 500 values.

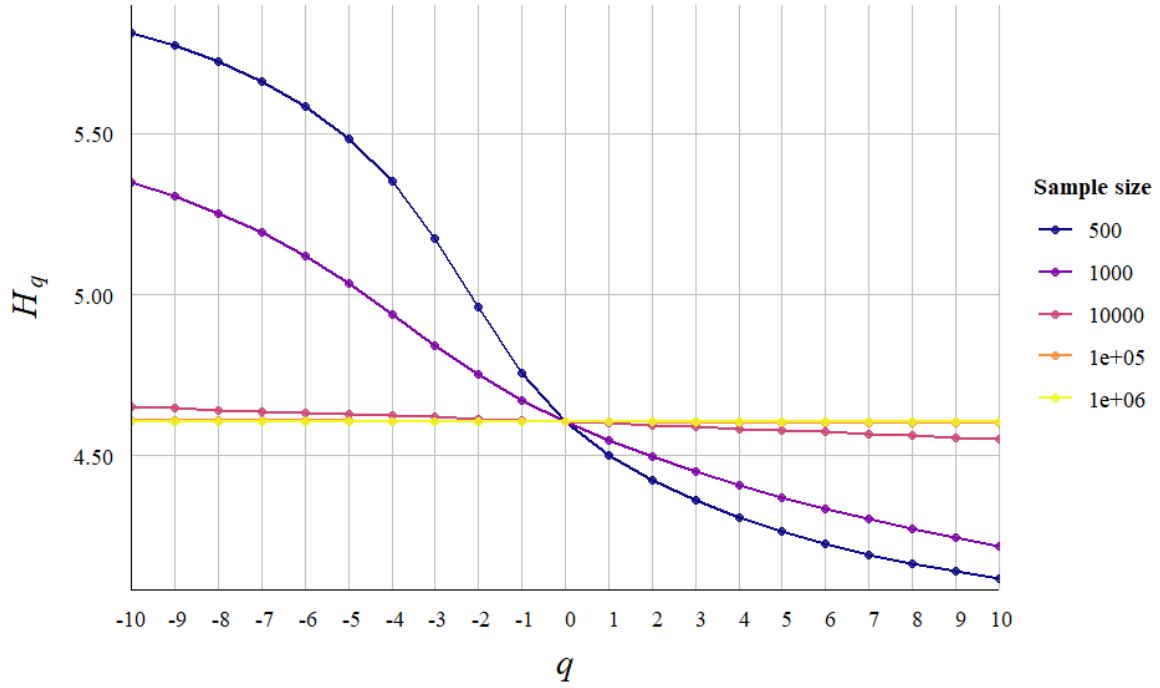


Figure 17: Rényi entropy estimates for uniform distribution (100 values) using binning method with number of bins 100 and different time series sizes.

4.2.2 Uniform distribution (100 values): Correlation sum method

In Figure 18, we see that the correlation sum method functions similarly to the binning method. Once we reach the needed precision, in our case $r = 1$, we are able to capture the data in a way that produces good results.

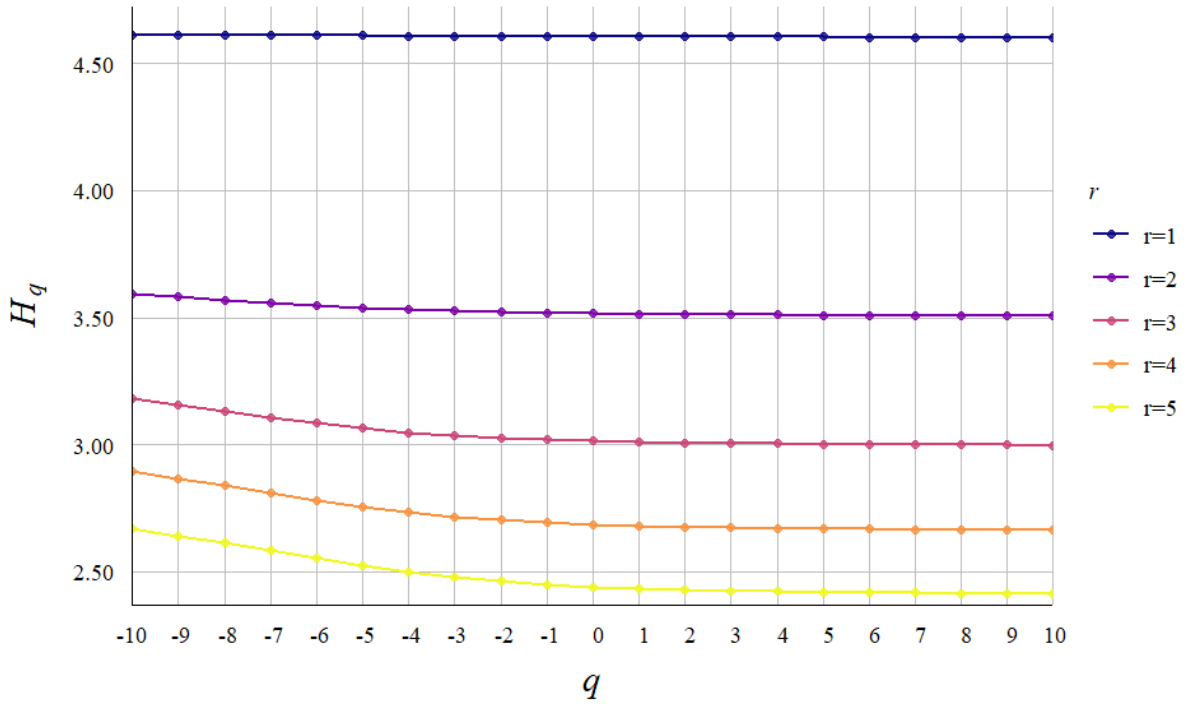


Figure 18: Rényi entropy estimates for uniform distribution (100 values) using correlation sum method with time series size 100,000.

A slight difference compared to the binning method, see Figure 19, is that the correlation sum method, even if the results become more consistent with 10,000 data points, has the division between overestimation and underestimation at the case $q = 2$. For this case, we still see more precise results. The reason is that the correlation sum method originates in second-order (pairwise) statistics, making it most precise at $q = 2$ while leading to overestimation at lower orders and underestimation at higher orders.

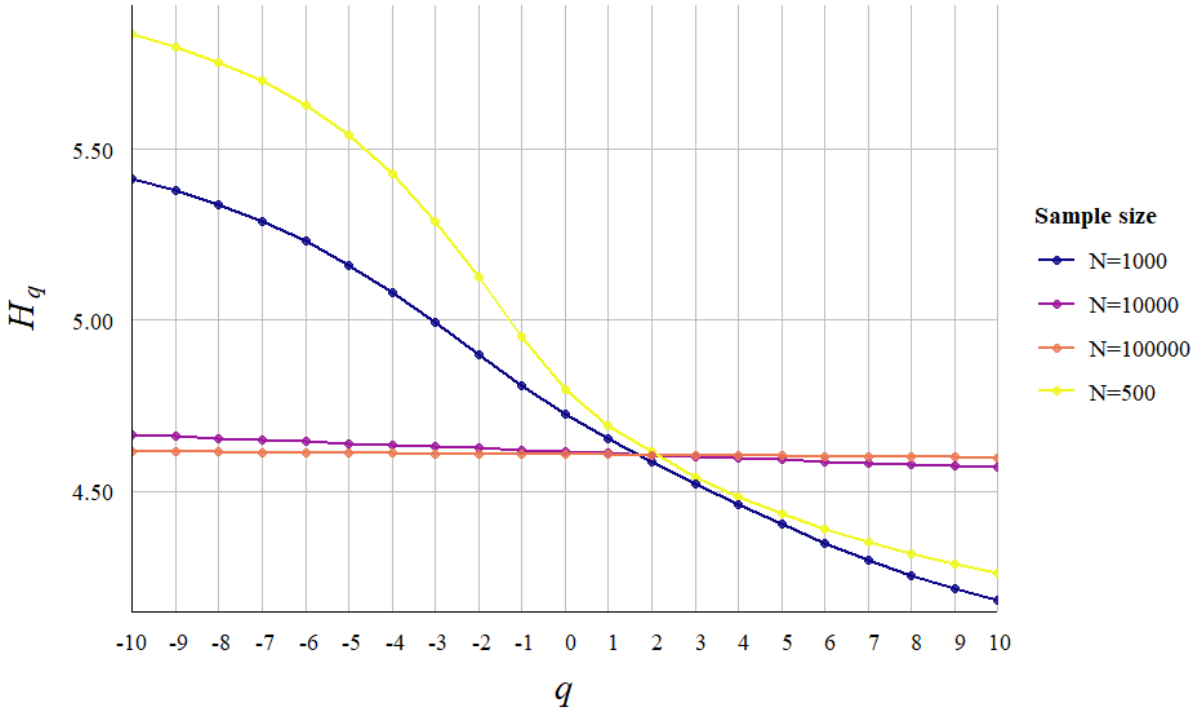


Figure 19: Rényi entropy estimates for uniform distribution (100 values) using correlation sum method with different time series sizes and radius 1.

We would expect the same result using the triplet-counting correlation sum method, but Figure 20 shows that we need to decrease the parameter r to obtain the most precise result.

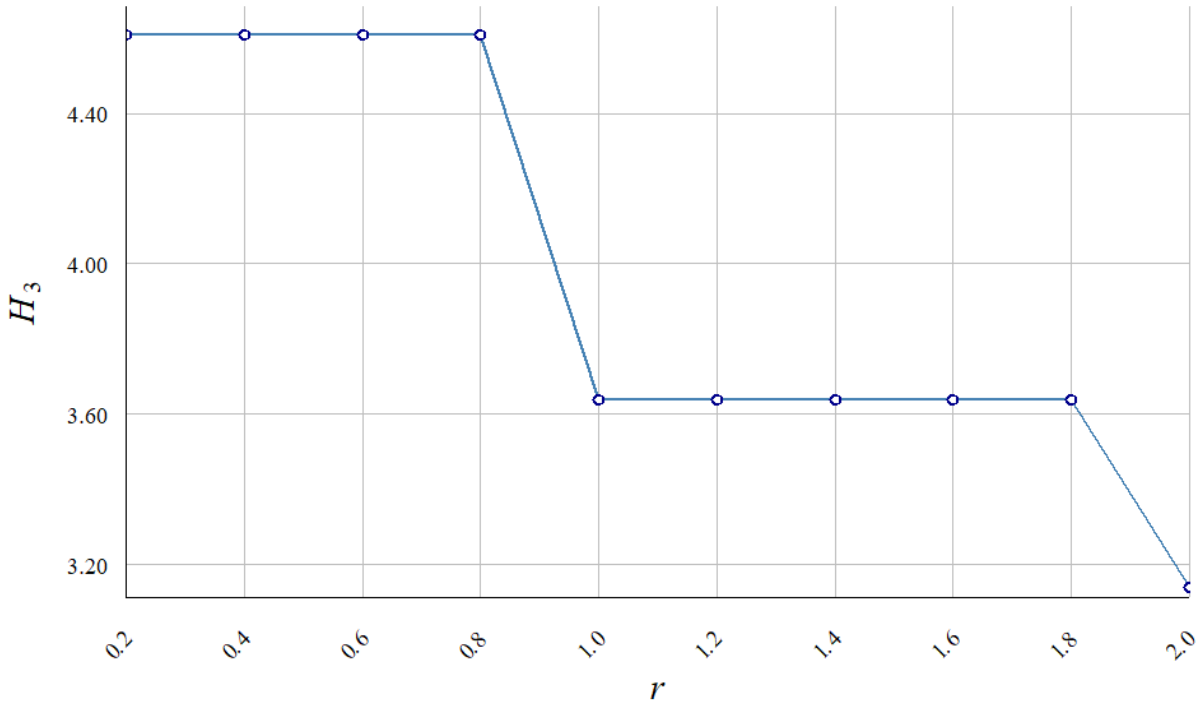


Figure 20: Rényi entropy estimates of order 3 for uniform distribution (100 values) using triplet-counting method with time series size 1,000.

4.2.3 Uniform distribution (100 values): k-NN method

We have previously mentioned that the k-NN method is not suited for calculating cases with discrete distributions. The method provides the most precise calculation for the case with $k = 1969$, after which the function exhibits strange behavior for estimates with $q < 0$, as demonstrated in Figure 21.

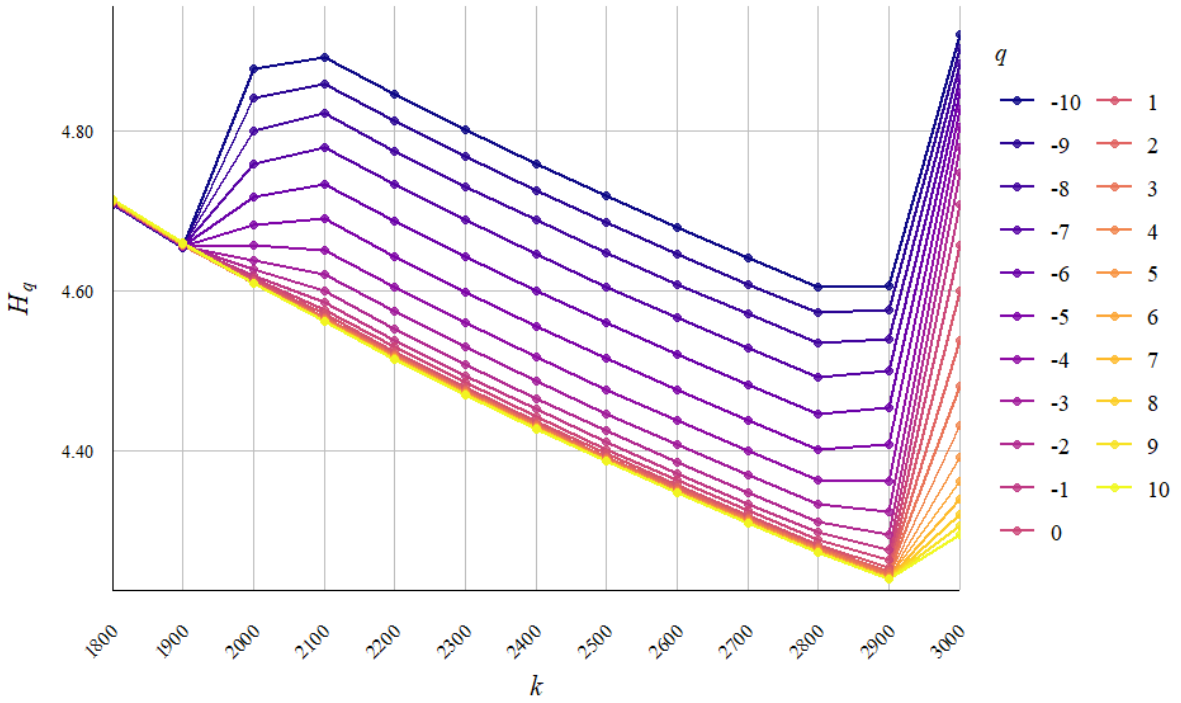


Figure 21: Rényi entropy estimates for uniform distribution (100 values) using k-NN method with time series size 100,000.

4.2.4 Uniform distribution (100 values): Comparison of methods

For the uniform distribution, we see that all of the methods give us relatively good results; however, the k-NN method is the least precise and the most difficult for parameter setting. We obtained the best possible result for $k = 1969$, which is much higher than the very common heuristic choice of $k = \sqrt{\textbf{ts size}}$. This is caused by the repeating identical values in the data set. This is not ideal for a method that relies on nearest neighbors distance calculation.

In Figure 22 and Table 3 for the binning and correlation sum methods, we see similar results. The slight difference is that the binning method seems to be accurate at the H_0, and for the correlation sum, the transition between overestimation and underestimation is at H_2. Our expectation was that we would see the best result for H_3 using the triplet counting method, but that was not the case, even though we were able to increase the time series length for this method to 10,000 data points. For both the binning and correlation sum methods, the parameters are set easily as we reach a point after which the results do not change. For the binning method, the best possible result is for 100 bins when all of

the unique points are separated, and a similar concept is also valid for the correlation sum method. In this case, once we lower the radius to 1, we obtain the most precise result for the method, which does not change with further shrinking of the search radius. For the triplet counting method, this stability occurs after shrinking the search interval to < 1. When it comes to which method is the fastest, we once again see that the binning method is faster than any other method. The correlation sum method is, in this instance, noticeably faster than the k-NN method.

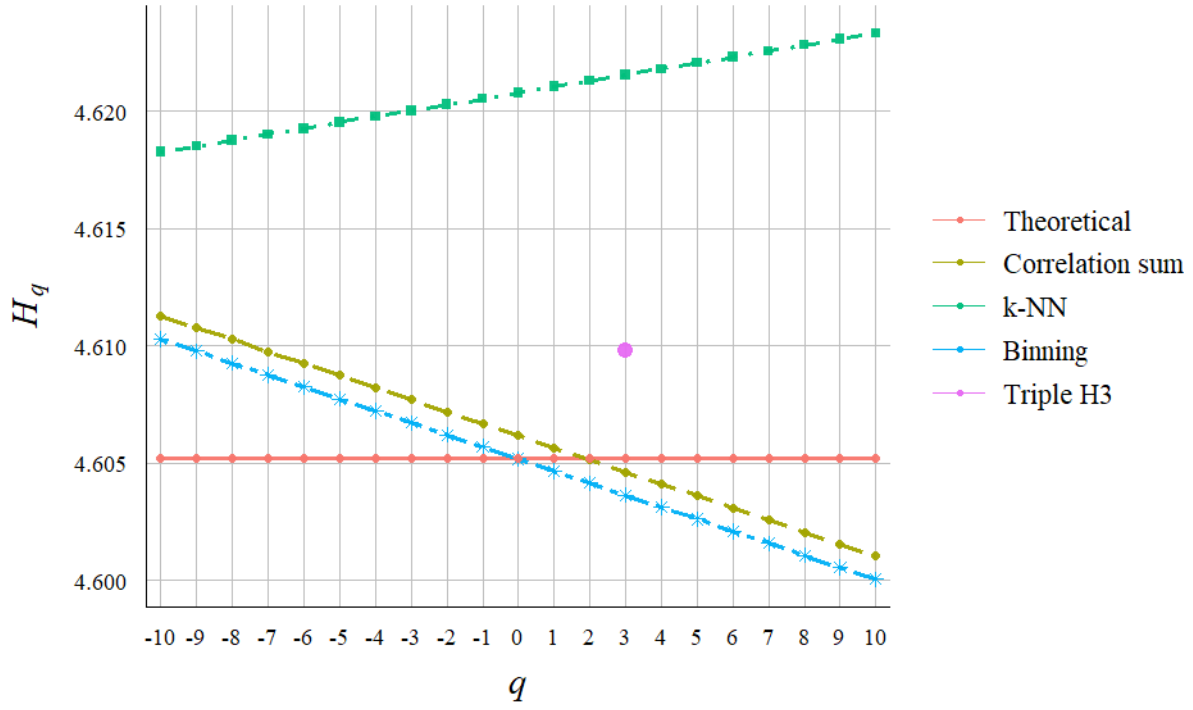


Figure 22: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 100, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000), k-NN (k = 1969, ts size = 100,000) and triplet counting method (r = 0.1, ts size 1,000) for uniform distribution (100 values) data.

Table 3: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 100, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000), k-NN (k = 1969, ts size = 100,000) and triplet counting method (r = 0.1, ts size 10,000) for uniform distribution (100 values) data.

Method	Runtime (seconds)	$\overline{MAE}_{-10:10}$
Correlation sum	1.27	0.00278
k-NN	5.82	0.0156
Binning (correction)	0.0200	0.00269
Triple counting	3.80	0.00462

4.3 Discrete distribution (20 values) results

The last of the discrete data sets we will take a look at has different characteristics compared to the uniform distribution. We expect to see the descending s-shaped behavior.

4.3.1 Discrete distribution (20 values): Binning method

Similar to the uniform distribution, we also see in Figure 23 that for this discrete distribution with 20 values, we once again need to reach the equivalent number of bins so that each value can fall into its own bin. The results after that do not change.

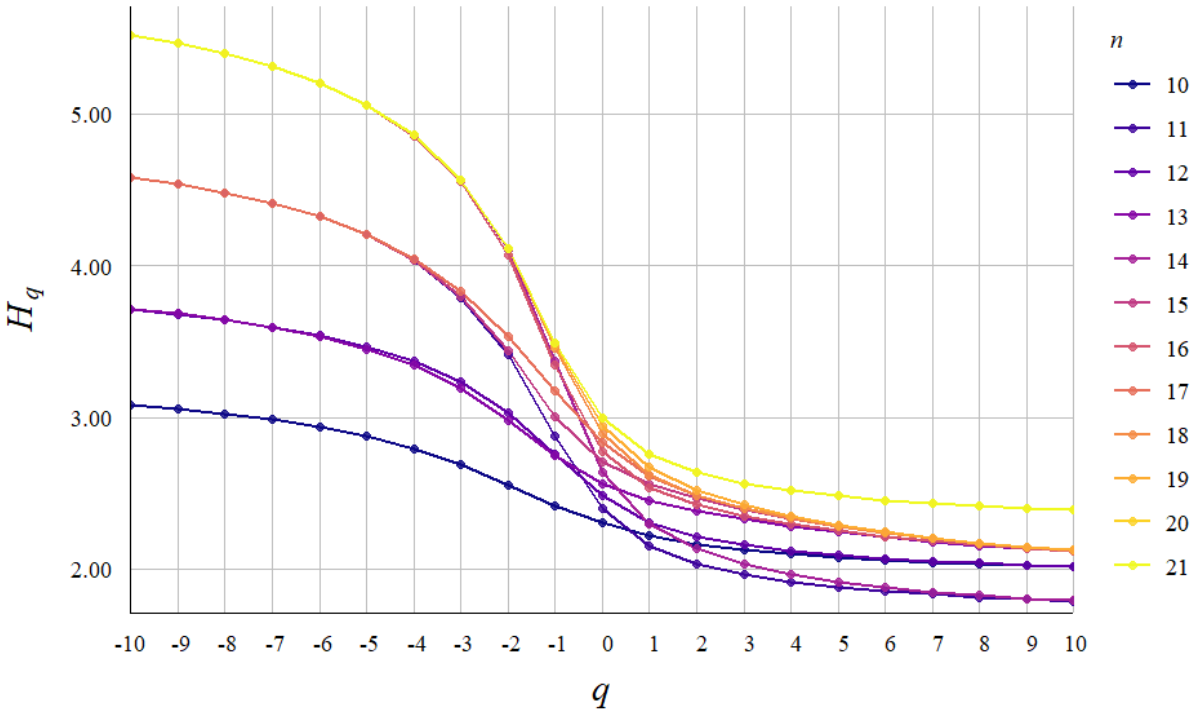


Figure 23: Rényi entropy estimates for discrete distribution (20 values) using binning method with time series size 1,000,000.

In Figure 24, we take a look at what happens when we have fewer data points at our disposal. We see that the method becomes significantly less accurate for $q < 0$. Even for a data set of 10,000 points, the result varies significantly from the case calculated for 100,000 data points.

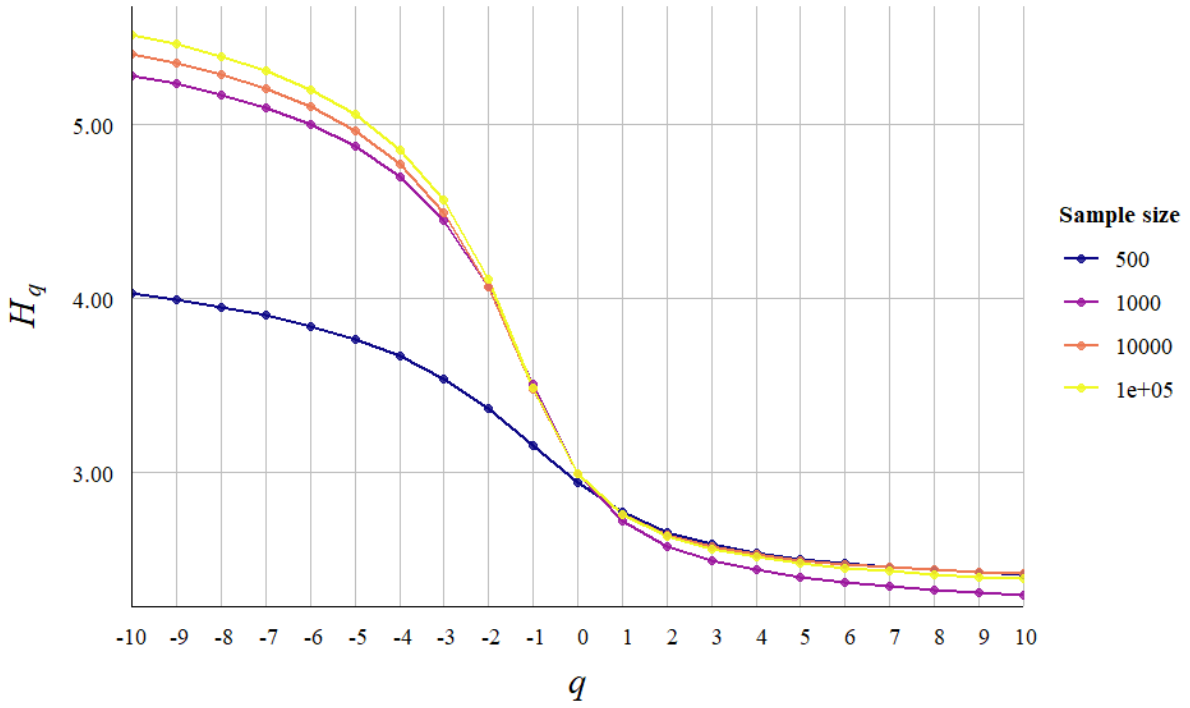


Figure 24: Rényi entropy estimates for discrete distribution (20 values) using binning method with different time series sizes and number of bins 20.

4.3.2 Discrete distribution (20 values): Correlation sum method

At this point, we do not find anything surprising in Figure 25. Once again, we need to decrease the r to separate the discrete values from each other, and we will get the best possible result from the correlation sum method. Note that the results for $r = 0.5$ and $r = 1$ are identical, as there is no difference between the calculated correlation sums.

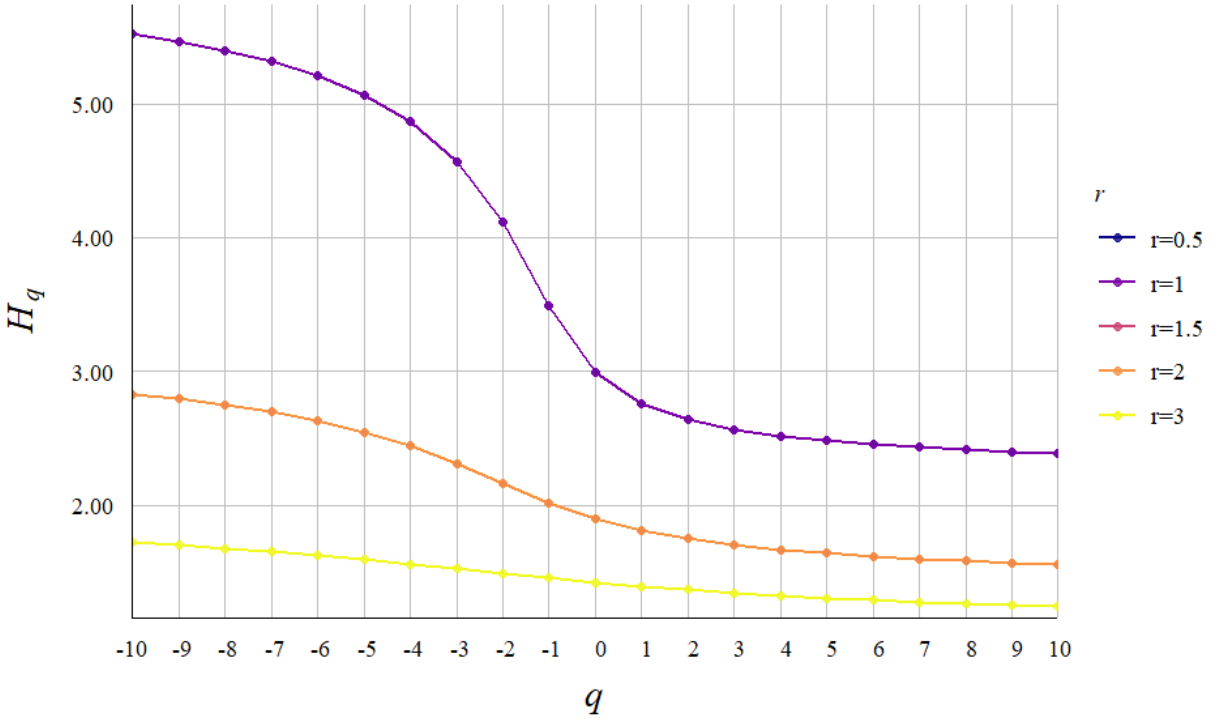


Figure 25: Rényi entropy estimates for discrete distribution (20 values) using correlation sum method with time series size 100,000 and different resolution r.

In Figure 26, we see bigger differences between different time series lengths, which can be caused by the fewer number of unique values in the data set compared to the observed uniform distribution. This can cause values to be overrepresented or underrepresented compared to their probability in the smaller data sample.

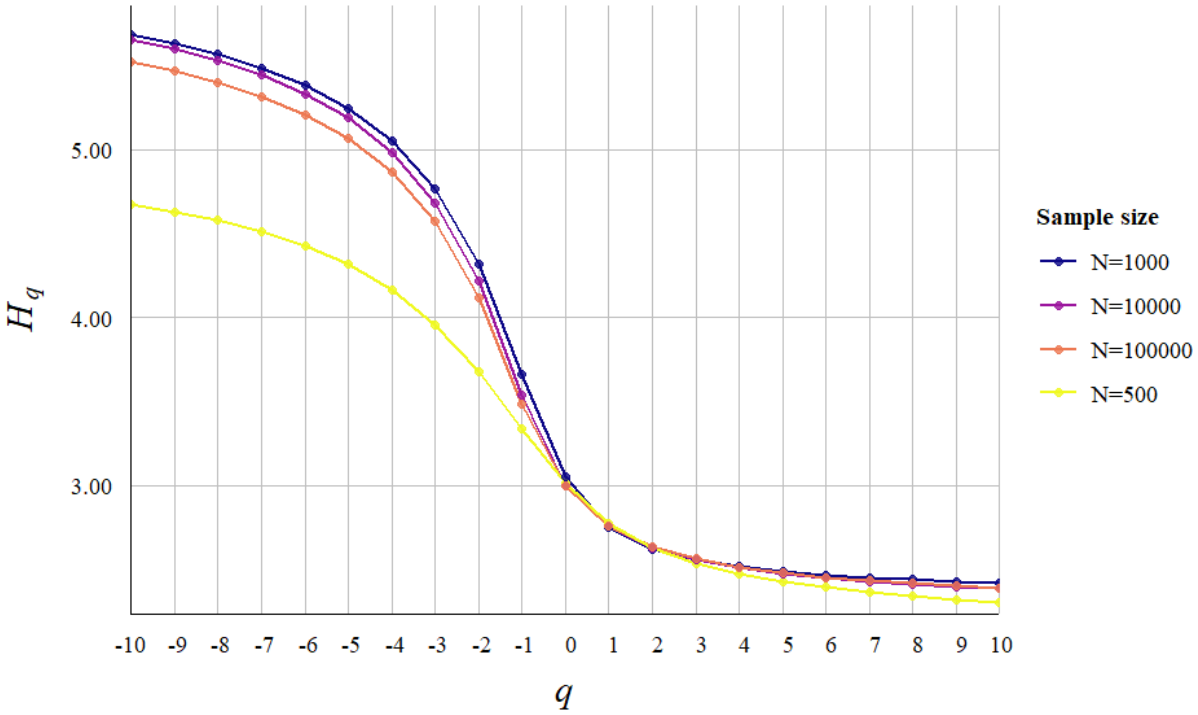


Figure 26: Rényi entropy estimates for discrete distribution (20 values) using correlation sum method with different time series sizes and radius 1.

For the triplet-counting method, we need to set the r parameter to be smaller than 1 for this data set, as we did for the uniform distribution. However, we see in Figure 27 that the estimations are less consistent than for the previous data set.

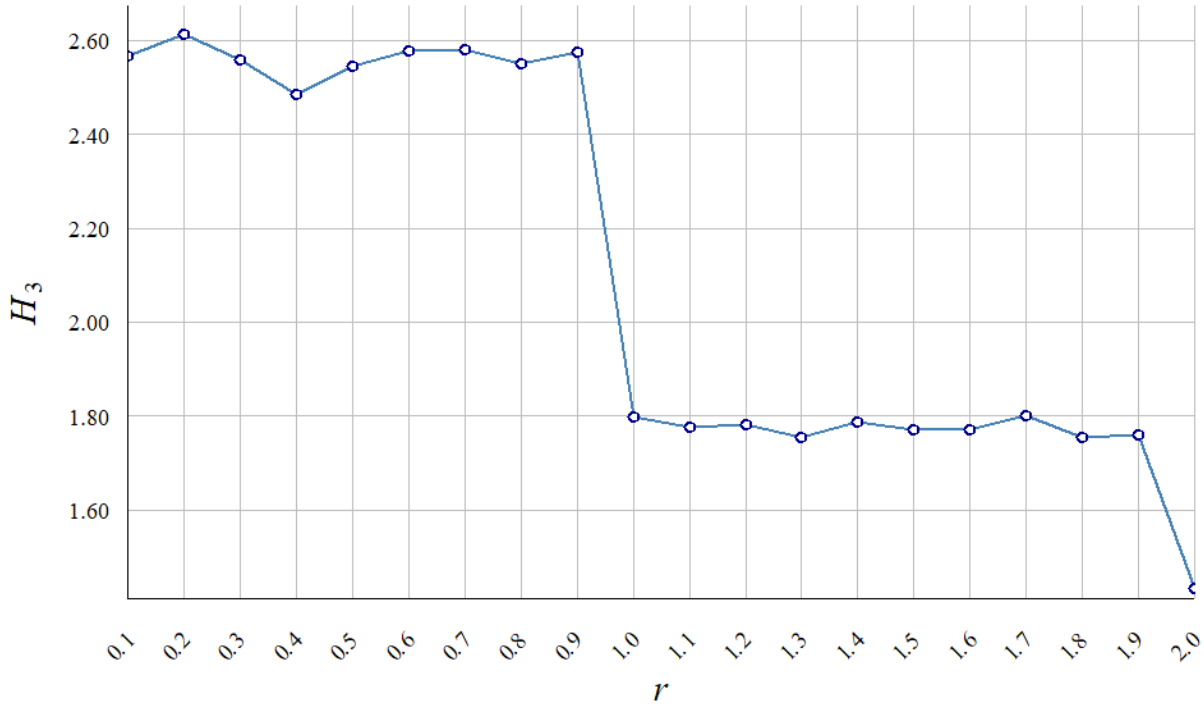


Figure 27: Rényi entropy estimates of order 3 for discrete distribution (20 values) using triplet-counting method with time series size 1,000.

4.3.3 Discrete distribution (20 values): Comparison of methods

For this discrete distribution, we see in Figure 28 and Table 4 similar results compared to the uniform distribution. The binning method is a bit less precise than the correlation sum method, but it is much faster. We are again able to determine the parameters for these methods intuitively with 20 bins and a radius of 1. For the triplet counting method, we get more precise results than before, however this can be caused by a fortunately selected sample, as we still need to use fewer data points for this method compared to the other two. The parameter selection for the triplet counting method is less stable for this data set, but we identified the best result for $r = 0, 2$.

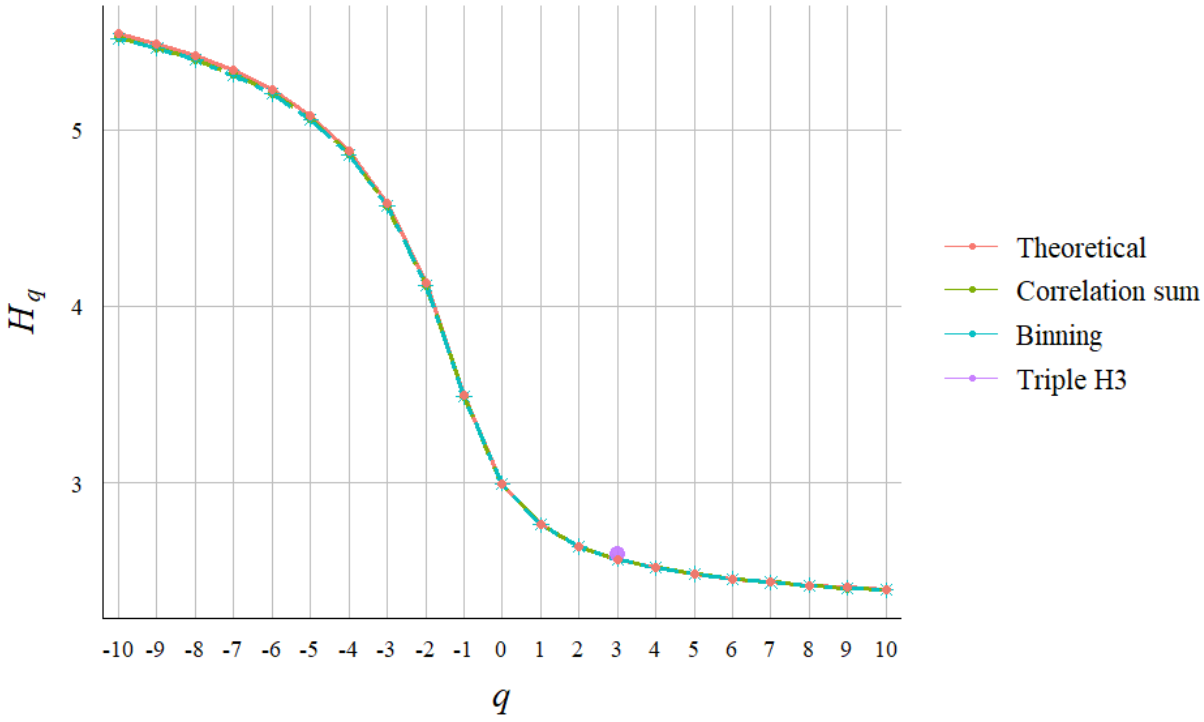


Figure 28: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 20, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000) and triplet counting method (r = 0.2, ts size 1,000) for discrete distribution (20 values) data.

Table 4: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 20, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000) and triplet counting method (r = 0.2, ts size 1,000) for discrete distribution (20 values) data.

Method	Runtime (seconds)	$\overline{MAE}_{-10:10}$
Correlation sum	9.15	0.00832
Binning (correction)	0.001	0.0102
Triple counting	0.220	0.0253

4.4 Exponential distribution results

This data set is our first example of data from a continuous distribution.

4.4.1 Exponential distribution: Binning method

We have applied the correction Formula 8 for the binning method and see a clear correction with increasing bin numbers. As we see in Figure 30, the estimation is consistent over a time series length of 1,000. We recommend choosing a parameter bigger than 20, as we see that the correction needs a certain number of bins to be effective. We cannot choose a number that is extremely high, as estimations still drift to a lower value. The number of bins with which we achieved the best possible result is 30, see Figure 29. This should not be taken as an exact number, rather an indication of the range where the method is the most precise.

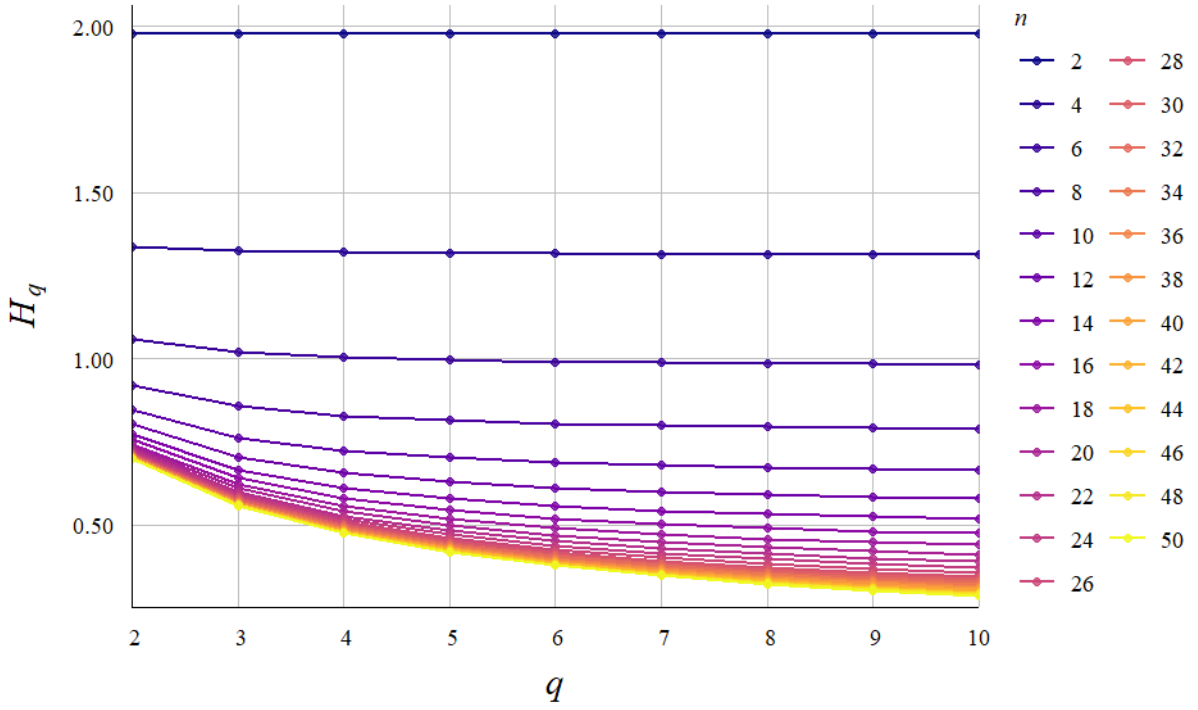


Figure 29: Rényi entropy estimates for exponential distribution using binning method with correction with time series size 1,000,000.

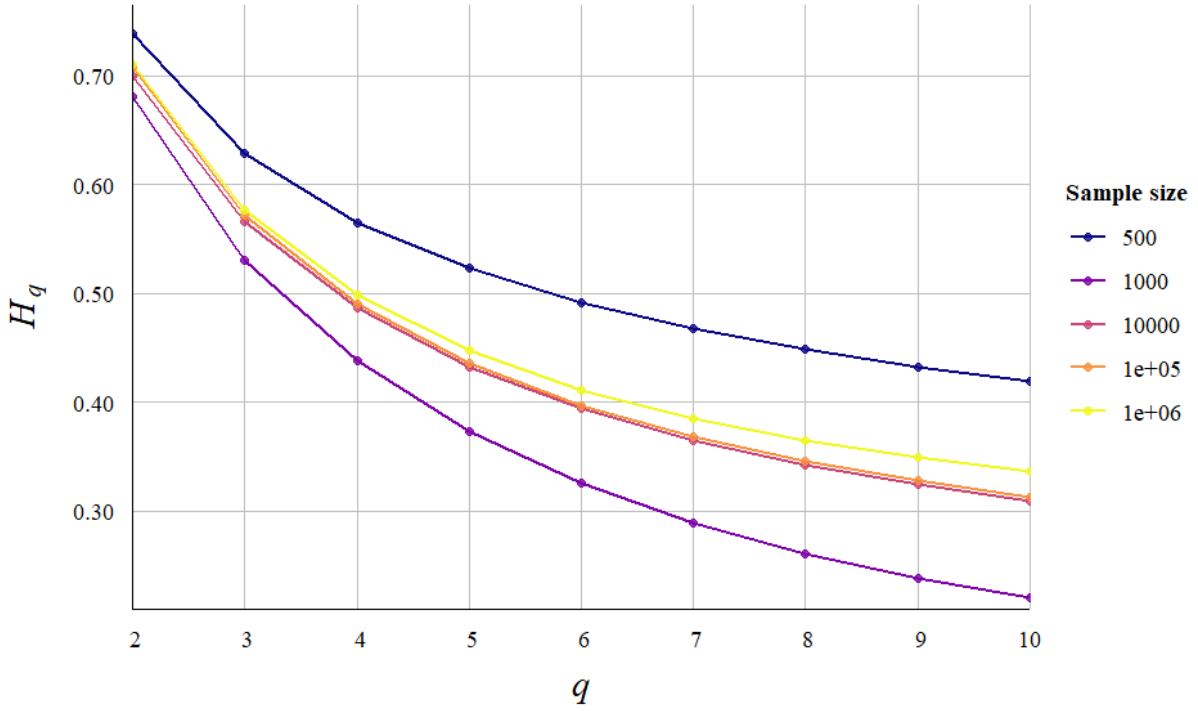


Figure 30: Rényi entropy estimates for exponential distribution using binning method with correction with different time series sizes and number of bins 30.

4.4.2 Exponential distribution: Correlation sum method

The correlation sum method is also able to provide reasonable but less precise estimates. However, a major flaw for all cases with continuous data is that the correction needed for these cases is not defined for the correlation sum method and triplet counting, see Figure 31 and 33. This brings the significant challenge of finding the correct radius, which is not the same across all distributions. In Figure 32 we see that there is bigger variance in the results of the correction sum method for different sample sizes.

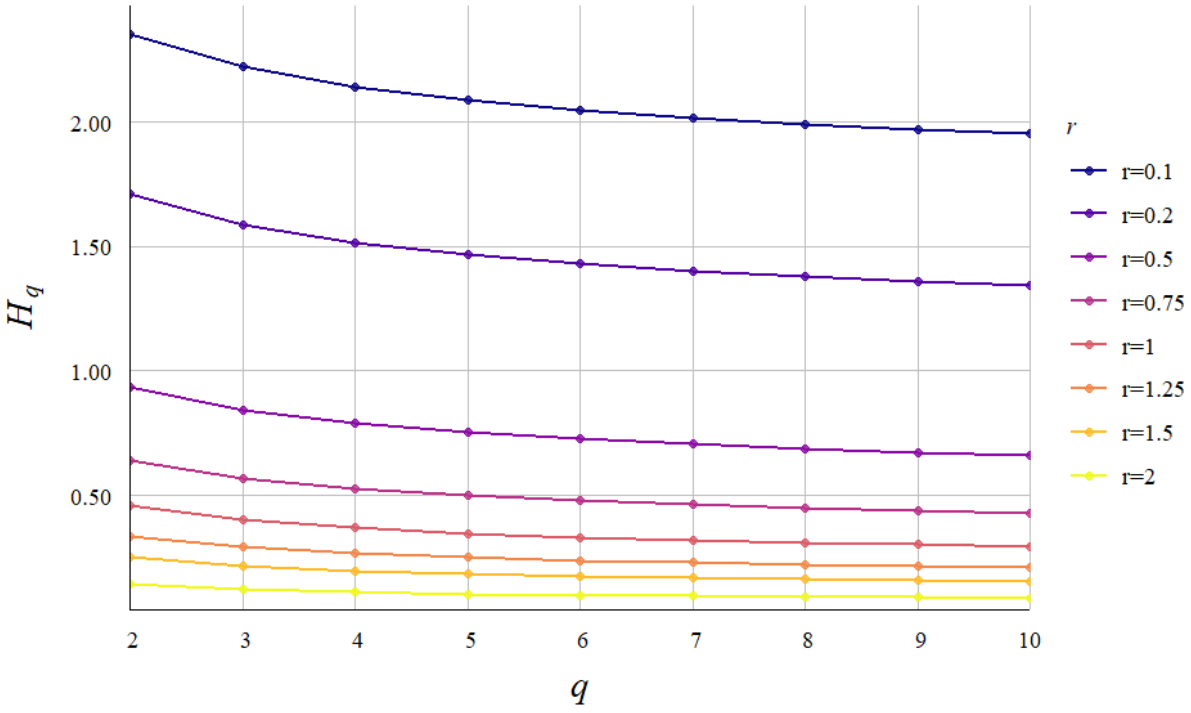


Figure 31: Rényi entropy estimates for exponential distribution using correlation sum method with time series size 100,000.

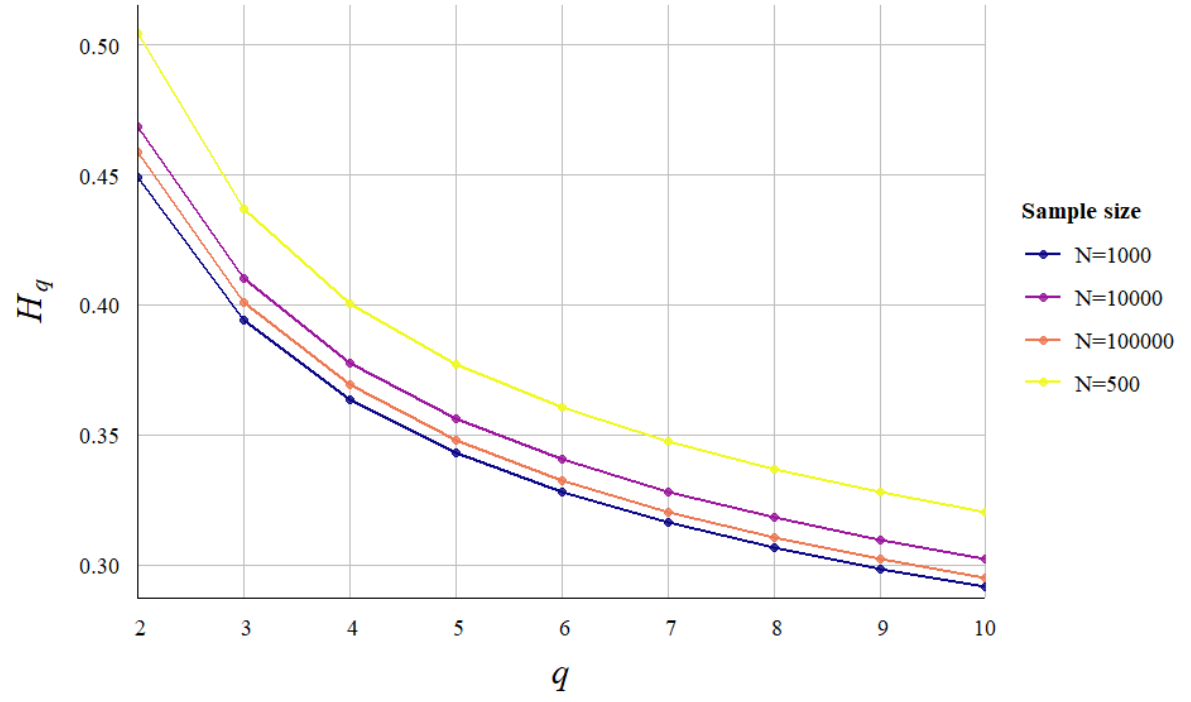


Figure 32: Rényi entropy estimates for exponential distribution using correlation sum method with different time series sizes and radius 1.

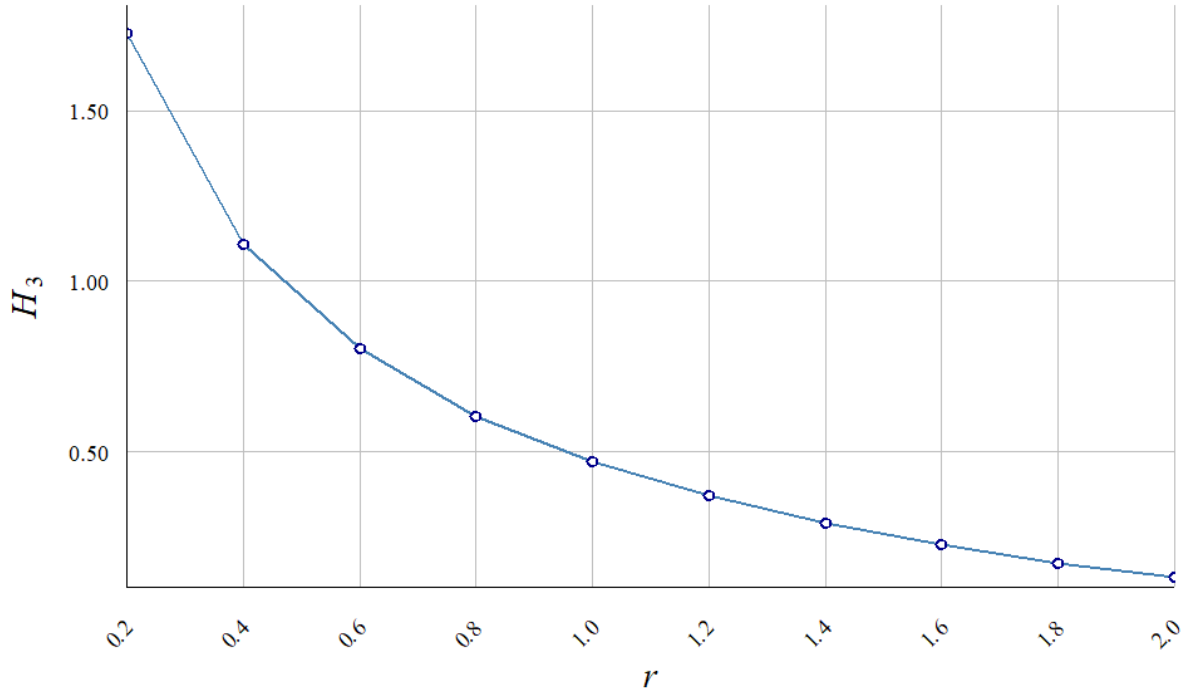


Figure 33: Rényi entropy estimates of order 3 for exponential distribution using triplet-counting method with time series size 1,000.

4.4.3 Exponential distribution: k-NN method

For the k-NN method, we have a very good result in Figure 34 as the parameter selection does not seem to be an issue for this method. We can apply the previously discussed rule of $k = \sqrt{\textbf{ts size}}$. But even if we choose a number of nearest neighbors that slightly differs from this value, we would receive a consistent result.

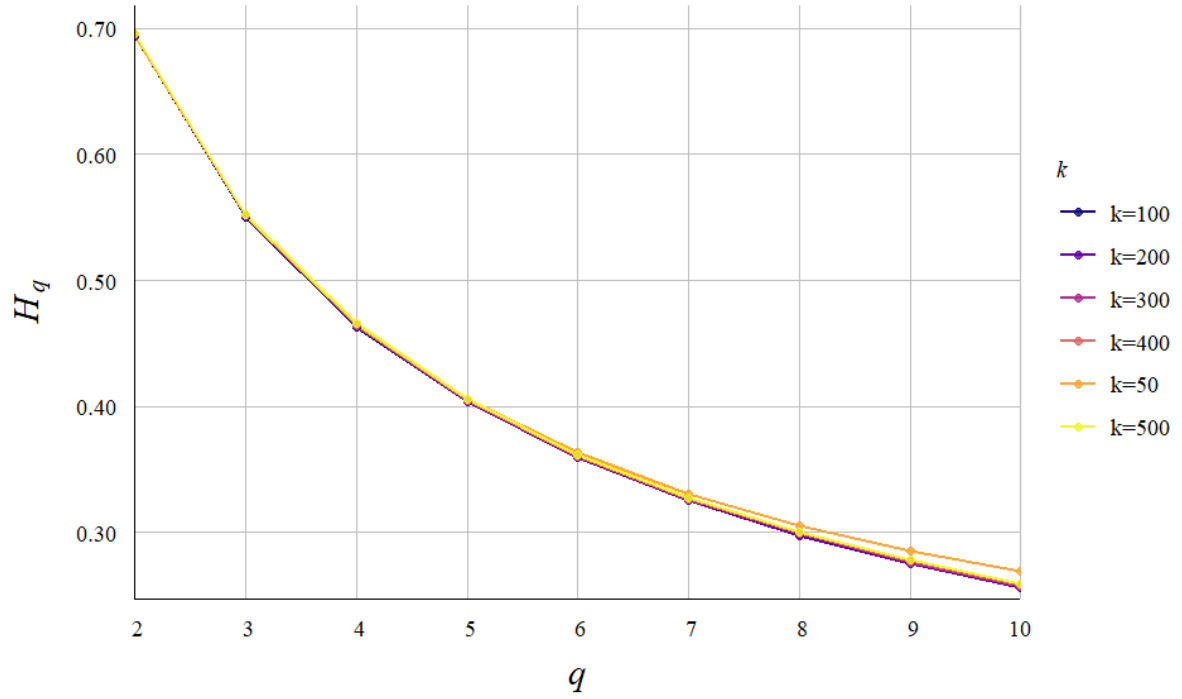


Figure 34: Rényi entropy estimates for exponential distribution using k-NN method with time series size 100,000.

4.4.4 Exponential distribution: Comparison of methods

We can clearly see in Figure 35 and Table 5 that the binning and k-NN methods are the most precise, with binning even beating the k-NN method by a small margin. The bigger difference is in the run time for these two methods. As the data with exponential distribution seems to be the most computationally challenging data set for both the correlation sum and triplet counting methods, we do not regard this as a viable option, also considering the imprecision and parameter selection challenges.

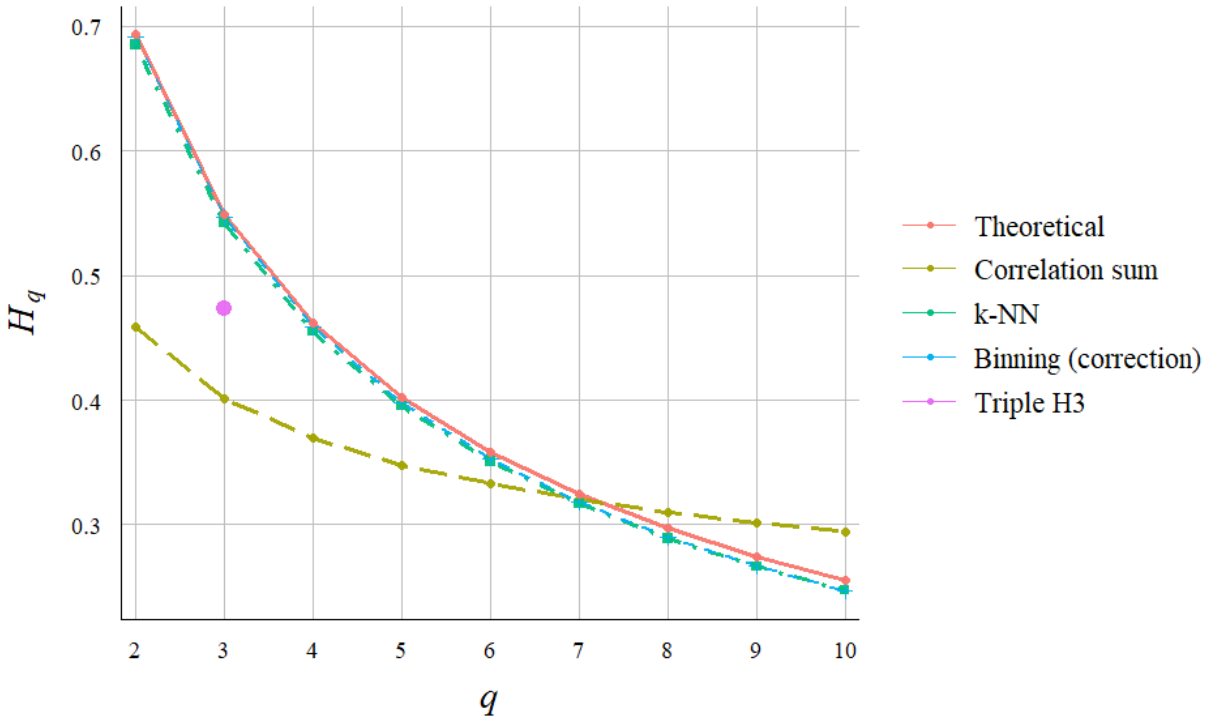


Figure 35: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 2,000, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000), k-NN (k = $\sqrt{100000}$, ts size = 100,000) and triplet counting method (r = 1, ts size 1,000) for exponential distribution data.

Table 5: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 2,000, ts size = 100,000), correlation sum method (r = 1, ts size = 100,000), k-NN (k = $\sqrt{100000}$, ts size = 100,000) and triplet counting method (r = 1, ts size 1,000) for exponential distribution data.

Method	Runtime (seconds)	$\overline{MAE}_{2:10}$
Correlation sum	44.3	0.0710
k-NN	2.52	0.00761
Binning (correction)	0.001	0.00531
Triple counting	152	0.0754

4.5 Gaussian distribution results

To grasp the performance of our method on cases with a continuous distribution further, we also took into consideration one of the most common distributions with these characteristics.

4.5.1 Gaussian distribution: Binning method

In Figure 36 we see a very encouraging result as estimations became very consistent after increasing the number of bins to over 20. The same consistency can be seen in Figure 37 for estimates using 10,000 data points and more.

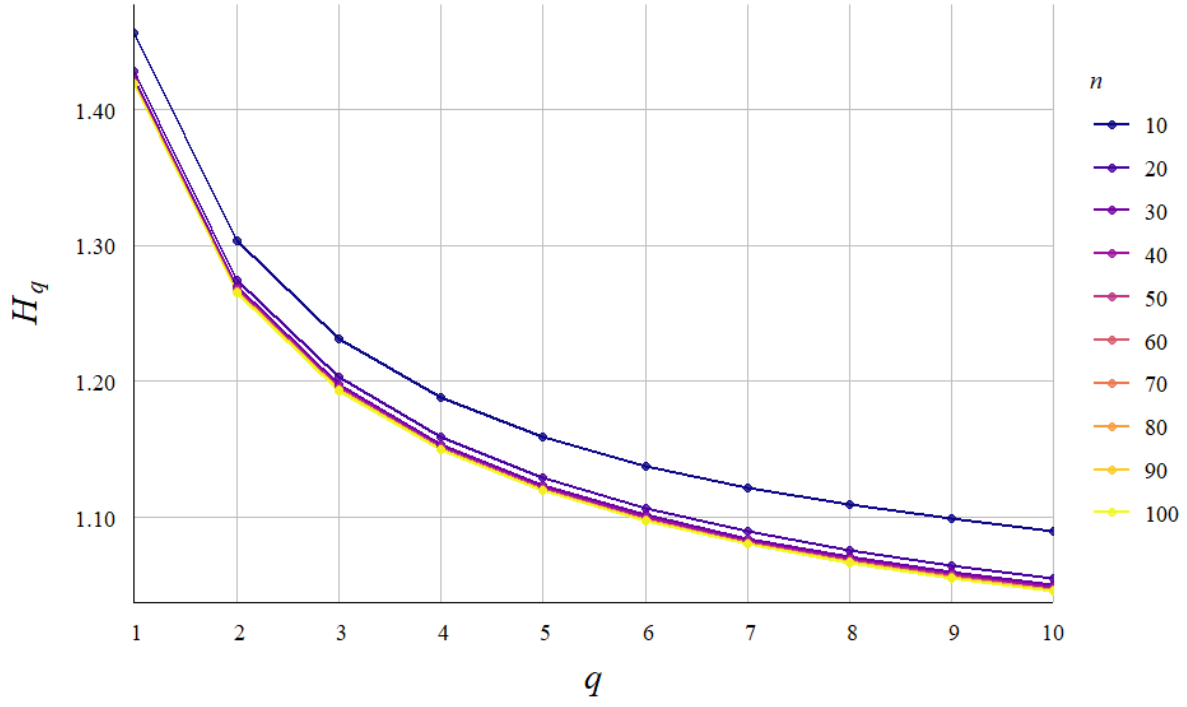


Figure 36: Rényi entropy estimates for Gaussian distribution using binning method with correction with time series size 1,000,000.

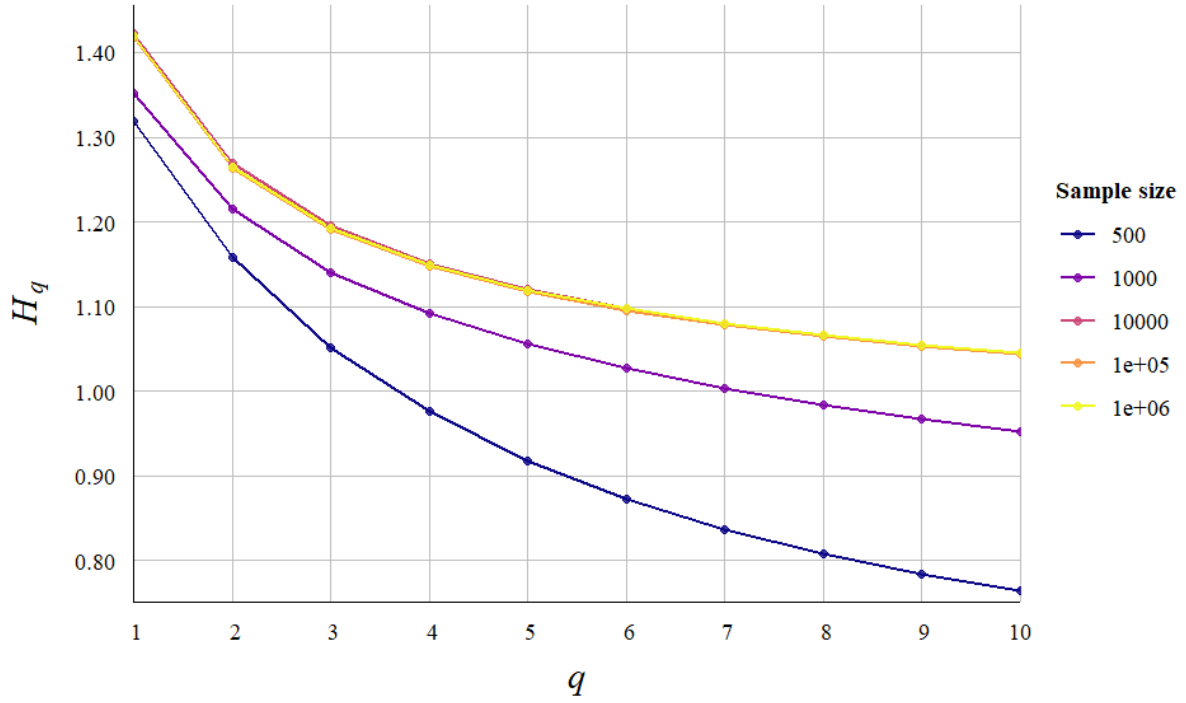


Figure 37: Rényi entropy estimates for Gaussian distribution using binning method with correction with different time series sizes and number of bins 100.

4.5.2 Gaussian distribution: Correlation sum method

Once again, we see in Figures 38 and 40 the issues with the correction sum and triplet counting methods for continuous datasets. This being the selection of the right search radius. It is still possible to find a good estimate of the Rényi entropies. However, we fail to find a general rule for why the specific values are the right ones. Another problem with the correlation sum method is that it is not that consistent when using smaller sample sizes, as shown in Figure 39.

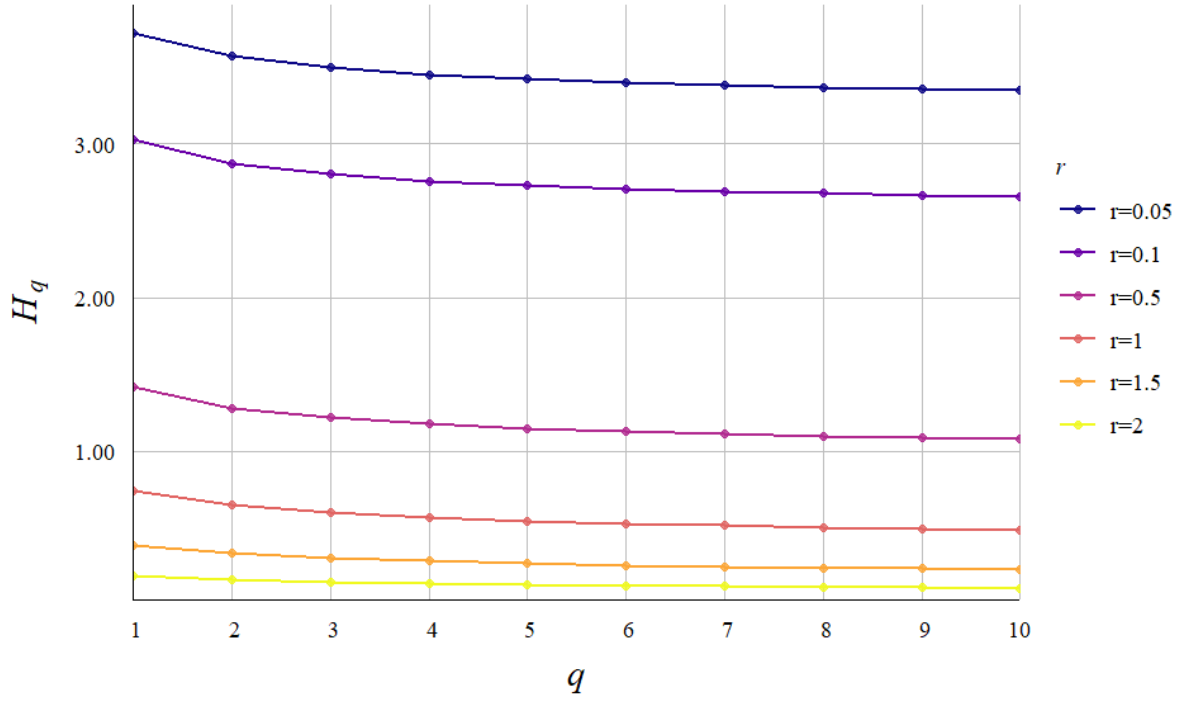


Figure 38: Rényi entropy estimates for Gaussian distribution using correlation sum method with time series size 100,000.

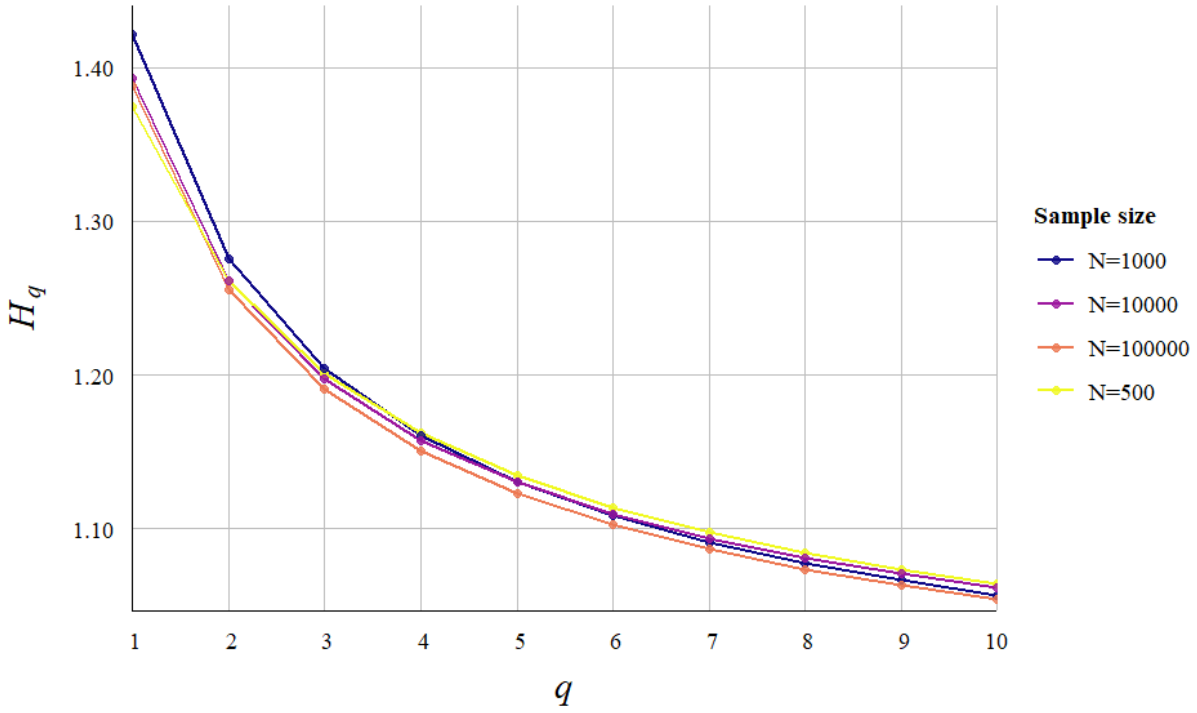


Figure 39: Rényi entropy estimates for Gaussian distribution using correlation sum method with different time series sizes and radius 0.515.

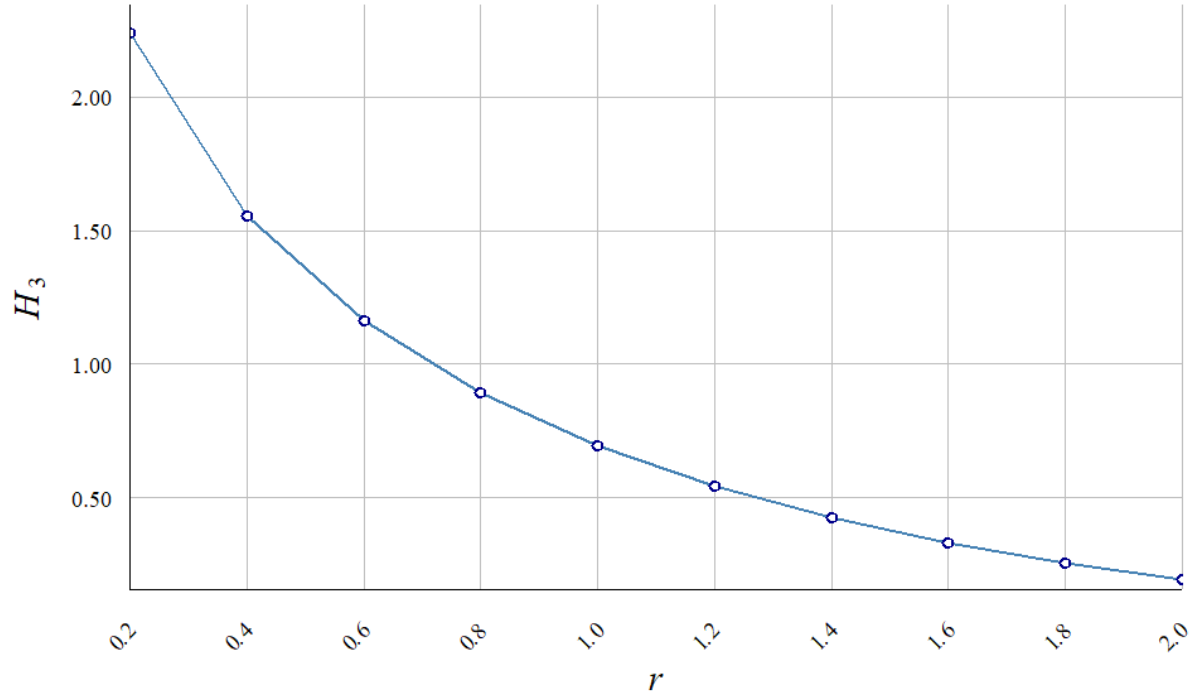


Figure 40: Rényi entropy estimates of order 3 for Gaussian distribution using triplet-counting method with time series size 1,000.

4.5.3 Gaussian distribution: k-NN method

The k-Nearest neighbors method excels in this category, as shown in Figure 41, where we see that the method provides us with consistent estimates across a wide range of parameter k. The best estimates will still remain the $k = \sqrt{\textbf{ts size}}$, which in our case is equal to $\sqrt{100,000} = 316,227766$.

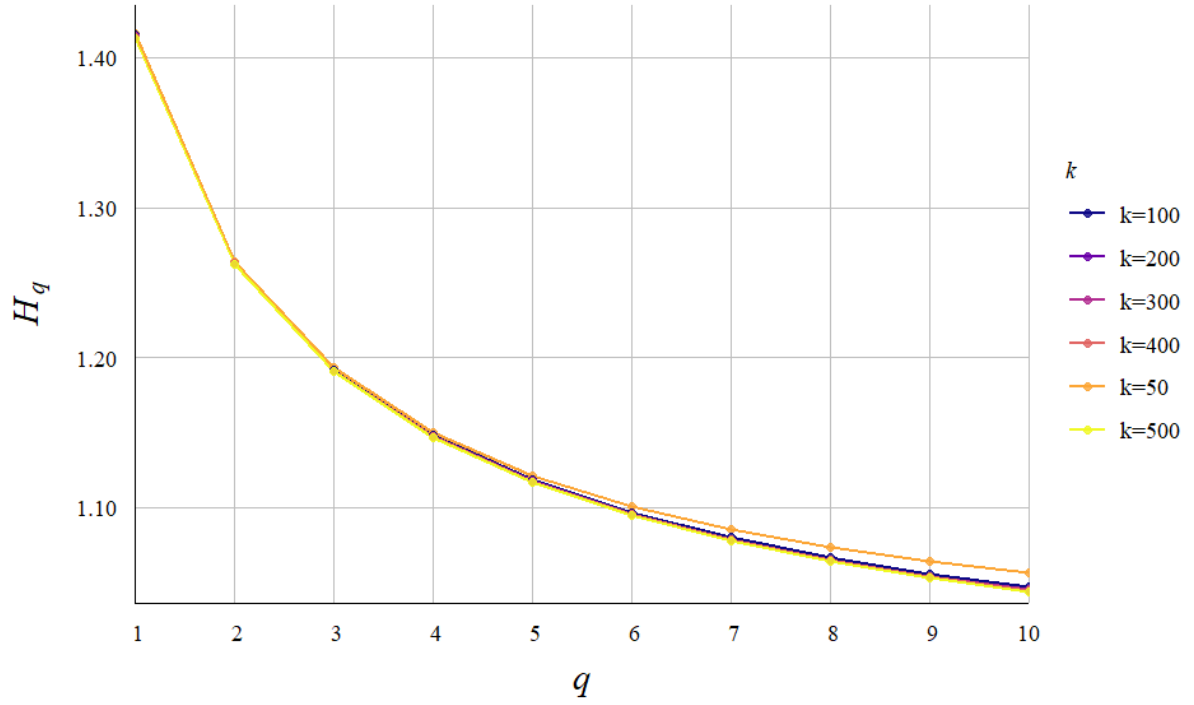


Figure 41: Rényi entropy estimates for Gaussian distribution using k-NN method with time series size 100,000.

4.5.4 Gaussian distribution: Comparison of methods

As we see in Figure 42 and also Table 6, the correction sum and related triplet counting methods provide us with much better estimates compared to the exponential distribution case, but they still remain the least precise methods. Out of the two better methods, once again the binning method is the faster option. However, as we are speaking regarding a difference of a few seconds in most cases, it is worth sacrificing computational power for a bit more precise results from the k-NN method.

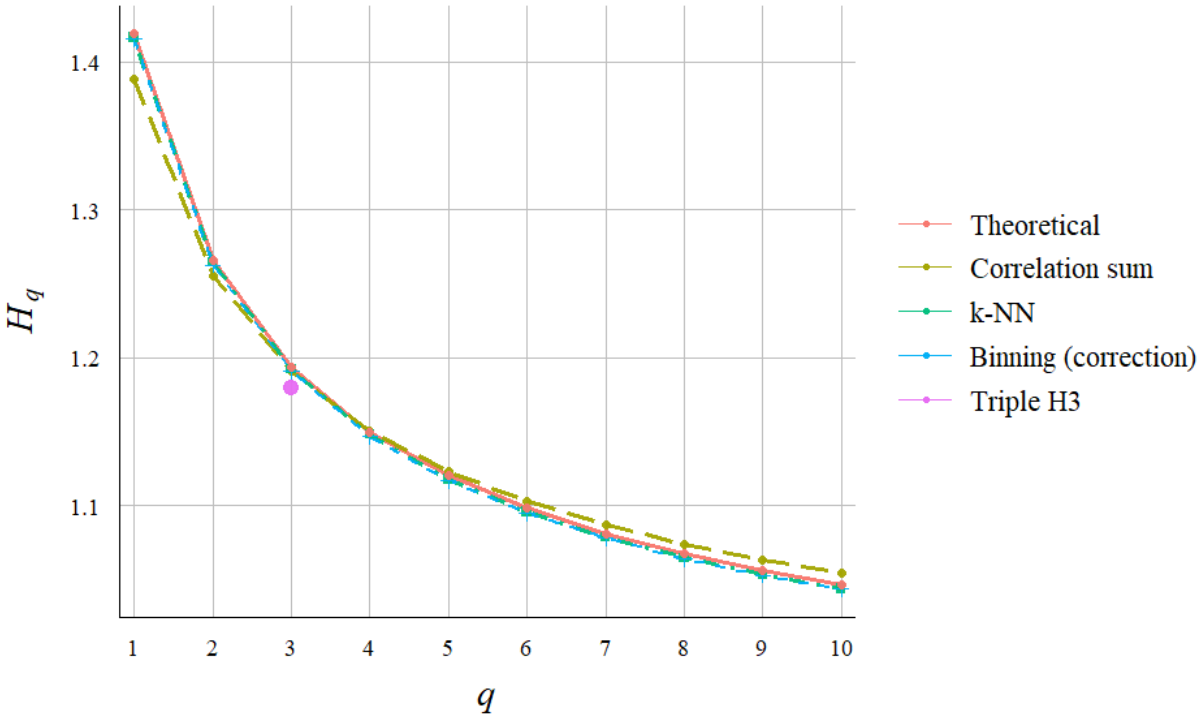


Figure 42: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 100, ts size = 100,000), correlation sum method (r = 0.515, ts size = 100,000), k-NN (k = $\sqrt{100000}$, ts size = 100,000) and triplet counting method (r = 0.59, ts size 1,000) for Gaussian distribution data.

Table 6: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 100, ts size = 100,000), correlation sum method (r = 0.515, ts size = 100,000), k-NN (k = $\sqrt{100000}$, ts size = 100,000) and triplet counting method (r = 0.59, ts size 1,000) for Gaussian distribution data.

Method	Runtime (seconds)	$\overline{MAE}_{1:10}$
Correlation sum	23.8	0.00783
k-NN	2.23	0.00229
Binning (correction)	0.001	0.00320
Triple counting	38	0.0144

4.6 Bivariate gaussian distribution results

Another challenge that can affect our methods is the case of multi-dimensional data. To test our methods, we select data with a two-dimensional gaussian distribution.

4.6.1 Bivariate gaussian distribution: Binning method

What we see in Figure 43 is the same image as we saw for the one-dimensional gaussian distribution case. The binning method provides consistent estimates after reaching 30 bins. A very different image is shown in Figure 44, as this method is much less consistent for multidimensional data in the case of a smaller amount of data points. Even in the case of 100,000 and 1,000,000 there is a visible difference in the provided estimates.

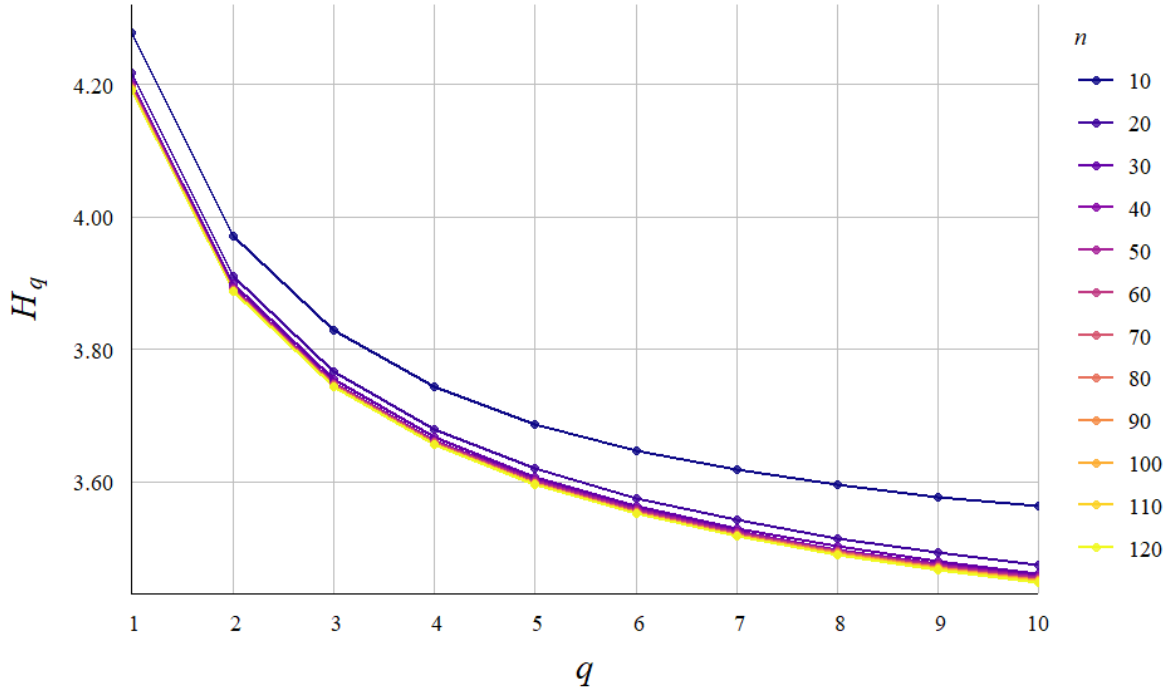


Figure 43: Rényi entropy estimates for bivariate Gaussian distribution using binning method with correction with time series size 1,000,000.

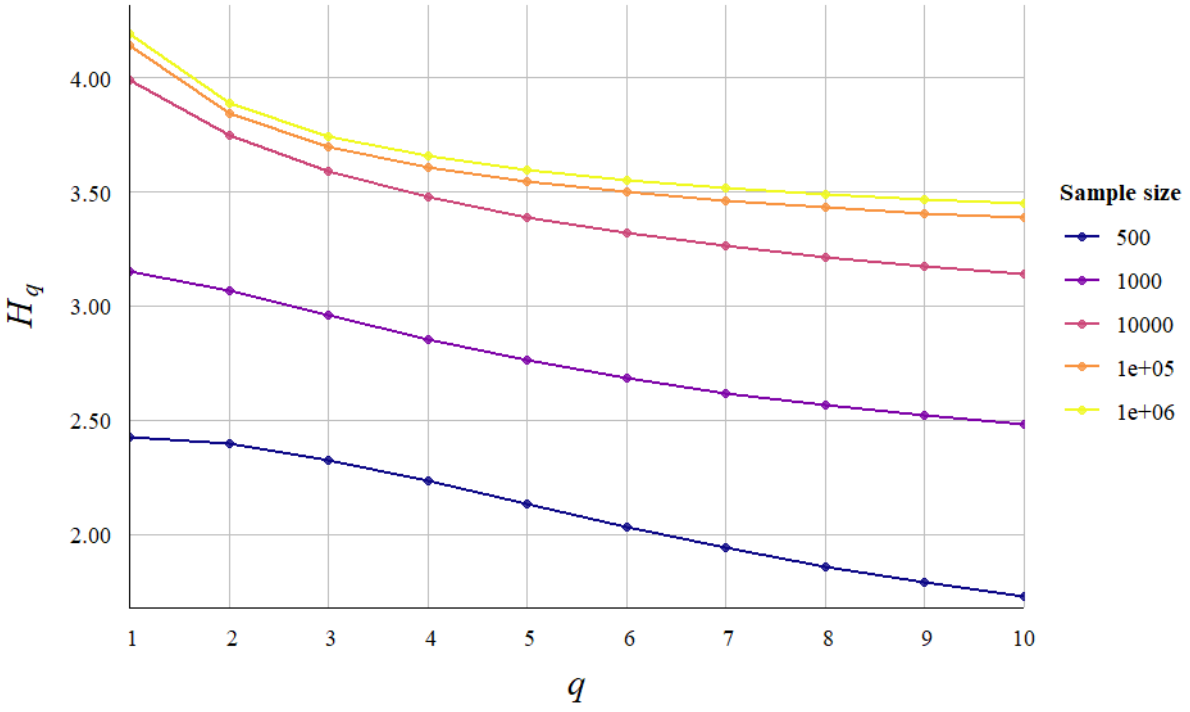


Figure 44: Rényi entropy estimates for bivariate Gaussian distribution using binning method with correction with different time series size and number of bins 90.

4.6.2 Bivariate gaussian distribution: Correlation sum method

We see nothing unexpected in Figures 45 and 47. The same problems arise for this data set as they did for the one-dimensional variant. When it comes to the estimation results with different time series lengths, we see in Figure 46 that for data with more than one dimension we need an increased number of data points to ensure consistent estimates.

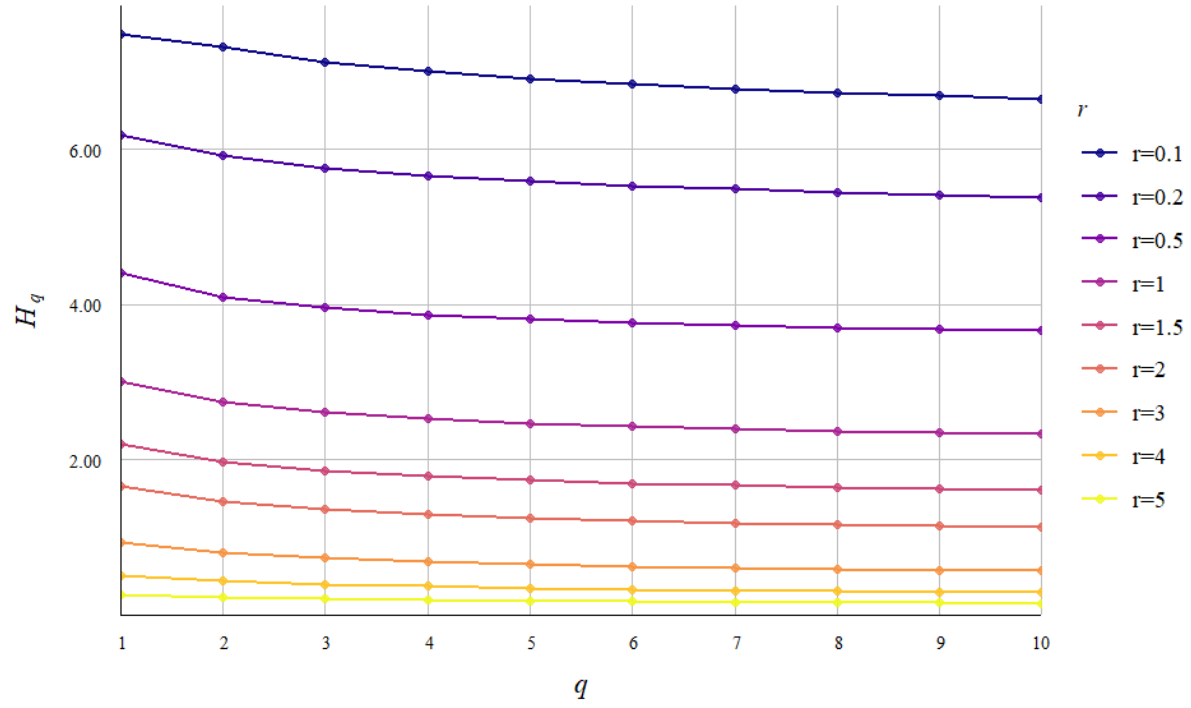


Figure 45: Rényi entropy estimates for bivariate Gaussian distribution using correlation sum method with time series size 100,000.

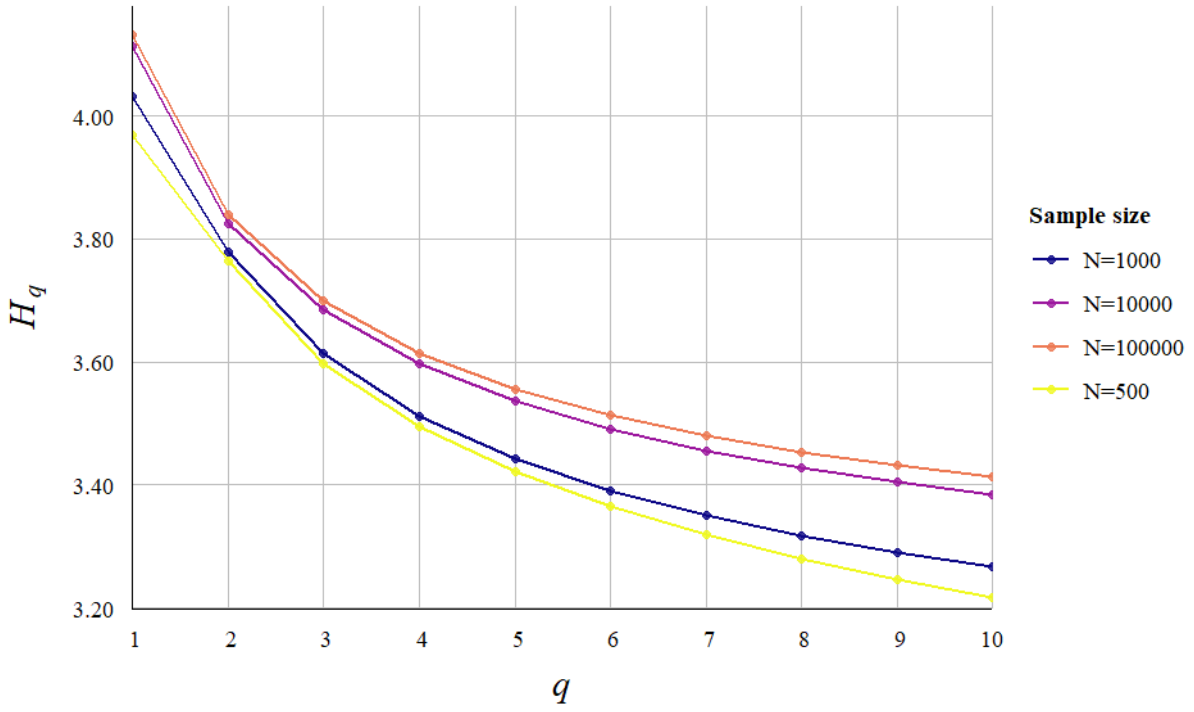


Figure 46: Rényi entropy estimates for bivariate Gaussian distribution using correlation sum method with different time series sizes and radius = 0.57.

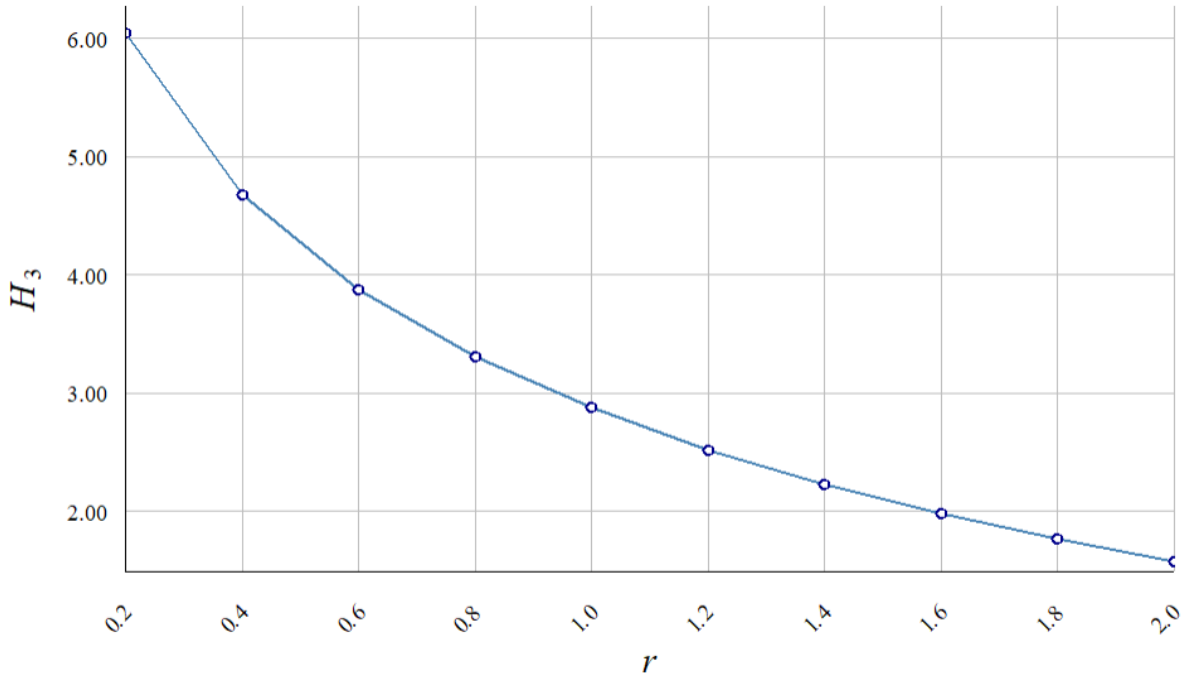


Figure 47: Rényi entropy estimates of order 3 for bivariate Gaussian distribution using triplet-counting method with time series size 1,000.

4.6.3 Bivariate gaussian distribution: k-NN method

Based on Figure 48, we can say that, compared to the other methods, the k-NN method seems to handle the multidimensional data set much better. We do not see the consistency issues, as the results are nearly identical across the selected parameter range.

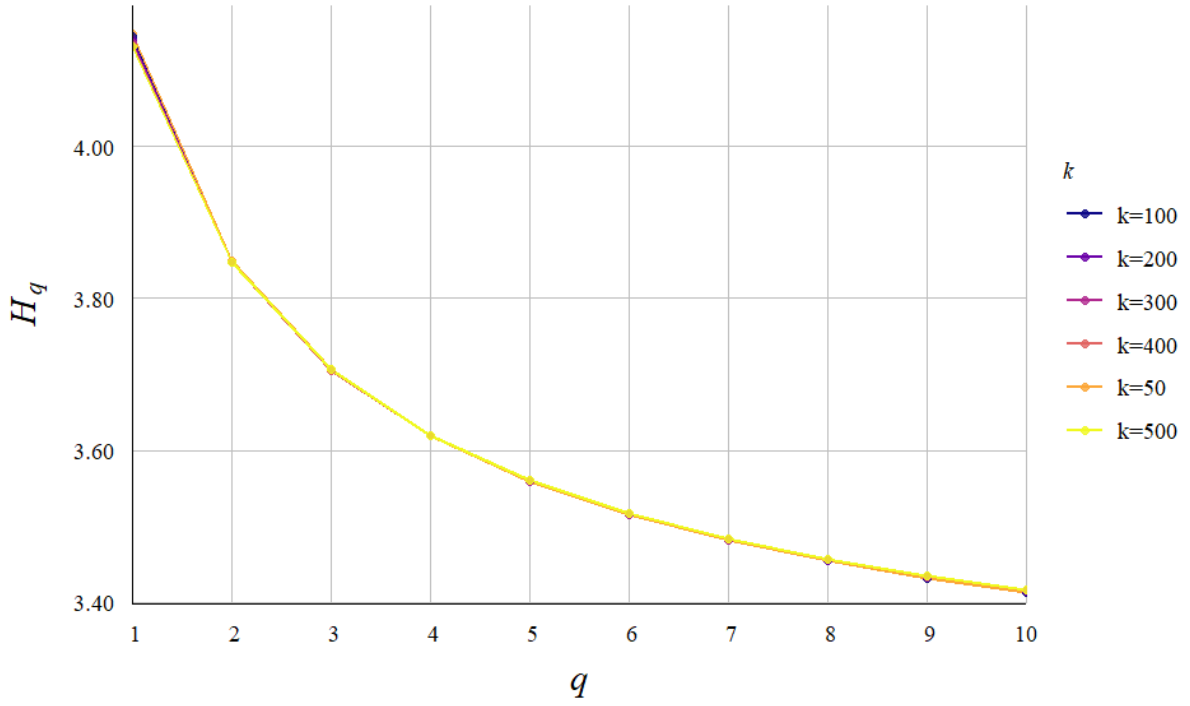


Figure 48: Rényi entropy estimates for bivariate Gaussian distribution using k-NN method with time series size 100,000.

4.6.4 Bivariate gaussian distribution: Comparison of methods

In Figure 49 and Table 7, we see a surprising result. The estimates provided by the binning method are much less precise compared to the cases with one-dimensional data, but on the other hand, the correlation sum and triplet counting methods are able to better estimate the Rényi entropy for this two-dimensional variant. When it comes to the k-nearest neighbors method, we see the same consistently good result for both the one- and two dimensional cases. If we take into account the lingering issue of the correlation sum method, namely the unreliable parameter selection. We can state that the k-NN method is the best choice for multidimensional data sets.

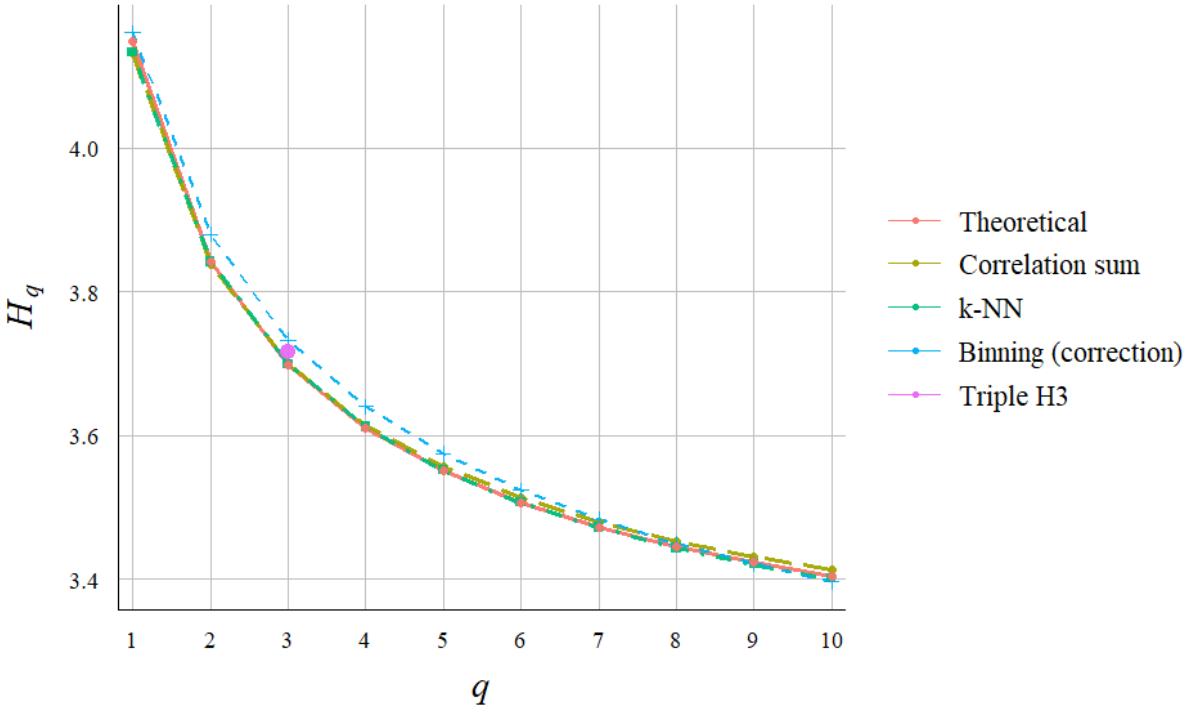


Figure 49: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 150, ts size = 100,000), correlation sum method (r = 0.57, ts size = 100,000), k-NN (k = $\sqrt{100000}$, ts size = 100,000) and triplet counting method (r = 0.65, ts size 1,000) for bivariate Gaussian distribution data.

Table 7: Comparison of Rényi entropy theoretical values to estimation by binning method (n = 150, ts size = 100,000), correlation sum method (r = 0.57, ts size = 100,000), k-NN (k = $\sqrt{100000}$, ts size = 100,000) and triplet counting method (r = 0.65, ts size 1,000) for bivariate Gaussian distribution data.

Method	Runtime (seconds)	$\overline{MAE}_{1:10}$
Correlation sum	61.9	0.00682
k-NN	5.61	0.00279
Binning (correction)	0.0300	0.0180
Triple counting	2.44	0.0188

4.7 Cantor set results

The estimation of Rényi entropy can be more difficult for complex data sets. The first of these data sets we look at is the Cantor set. Based on the histogram (Figure 7), we see that this data set is closest to our discrete data sets. Thanks to our previous calculations, we know that the k-NN method wont be a suitable candidate for this case. However, the binning method and correlation sum method performed well on similar data. In Figure 50 and 51, we see that the estimates do not stabilize with changing parameters. Validating results is more difficult due to the absence of theoretical Rényi entropy values, but what we have are the theoretical Rényi dimensions. Thanks to Equation 7, we can estimate the Rényi dimensions using our calculated entropy estimates and validate the results this way.

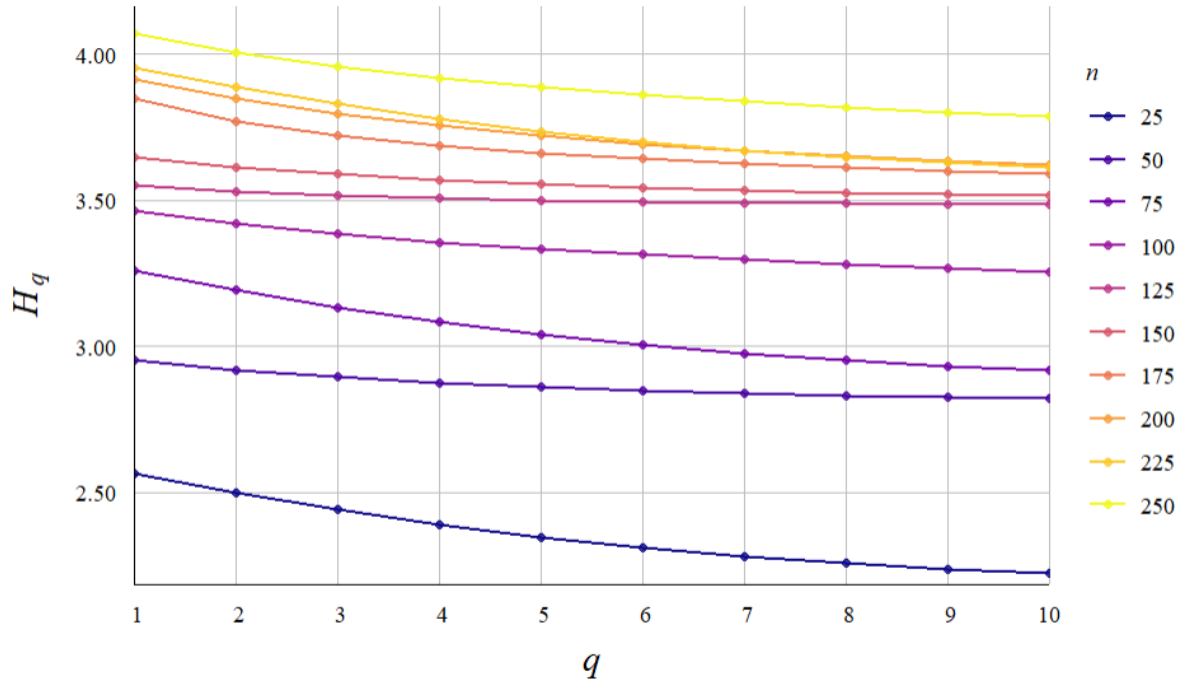


Figure 50: Rényi entropy estimates for Cantor set using binning method without correction with time series size 1,000,000.

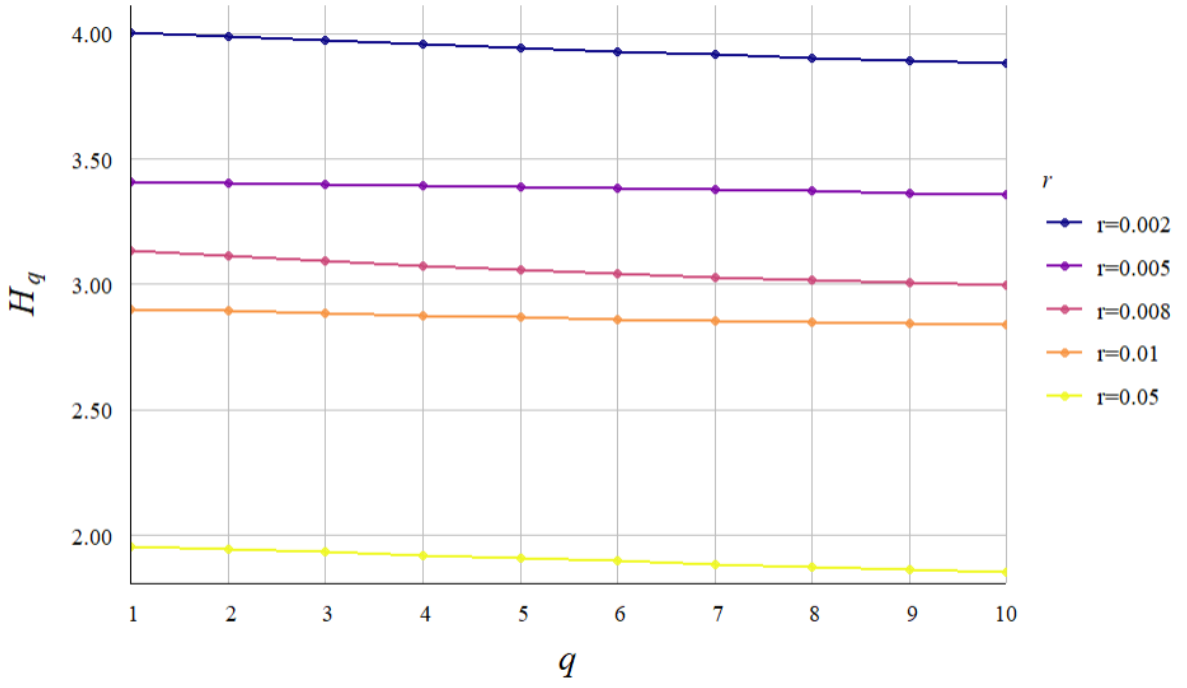


Figure 51: Rényi entropy estimates for Cantor set using correlation sum method with time series size 100,000.

We know that the Rényi dimensions for the Cantor set should be approximately $0,631$ for all q values. In Table 8 and 9, we find the estimated dimensions, for 100,000 data points, that are closest to the expected value, if using sufficiently small bins with the binning method or small resolutions with the correlation sum-based method. The best results were obtained when taking a resolution of 1% and 0.5% from the range of the data. These results, however, do not mean that we found general Rényi entropy values; rather, we have estimates corresponding to the parameters n and r.

q	n_1	n_2	Scale 1	Scale 2	$H_q(n_1)$	$H_q(n_2)$	D_q
0	100	200	0.01	0.005	3.526361	3.988984	0.6674247
2	100	200	0.01	0.005	3.419346	3.849082	0.6199785
3	100	200	0.01	0.005	3.384196	3.797468	0.5962257
7	100	200	0.01	0.005	3.295886	3.669214	0.5385983
10	100	200	0.01	0.005	3.254780	3.620523	0.5276564

Table 8: Comparison of Rényi entropy estimates for different scales for Cantor set by binning method for 100,000 data points.

q	r_1	r_2	Scale 1	Scale 2	$H_q(r_1)$	$H_q(r_2)$	D_q
0	0.01	0.005	0.01	0.005	2.910900	3.415243	0.7276125
2	0.01	0.005	0.01	0.005	2.892995	3.405420	0.7392734
3	0.01	0.005	0.01	0.005	2.884337	3.400035	0.7439955
7	0.01	0.005	0.01	0.005	2.854079	3.376264	0.7533548
10	0.01	0.005	0.01	0.005	2.836989	3.357829	0.7514141

Table 9: Comparison of Rényi entropy estimates for different scales for Cantor set by correlation sum method for 100,000 data points.

In Figure 52, we see that the estimations differ considerably across different parameters and methods. Due to this, our recommendation is to relatively compare entropies of different systems only using the same method and the same parameters.

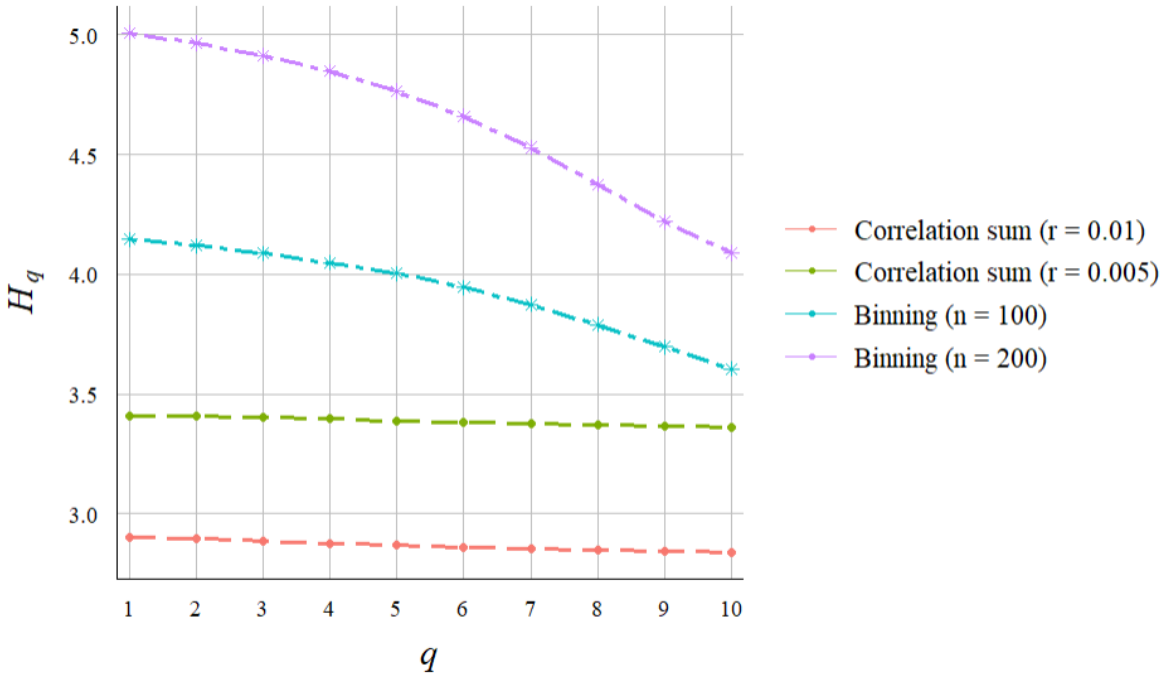


Figure 52: Comparison of Rényi entropy estimation by binning method (n = 100 and 200, ts size = 100,000, without correction) and correlation sum method (r = 0.01 and 0.005, ts size = 100,000) for Cantor distribution data.

4.8 Hénon map results

Similar to the Cantor set data, the Hénon map is also generated in discrete steps, but it is a two dimensional map. We once again use the binning and correlation sum methods, and the resulting estimates can be seen in Figures 53 and 54. In the case of the binning method, we can see more consistent estimates for values corresponding to $q < 0$, but for both cases, we cannot be sure of the correctness of the results.

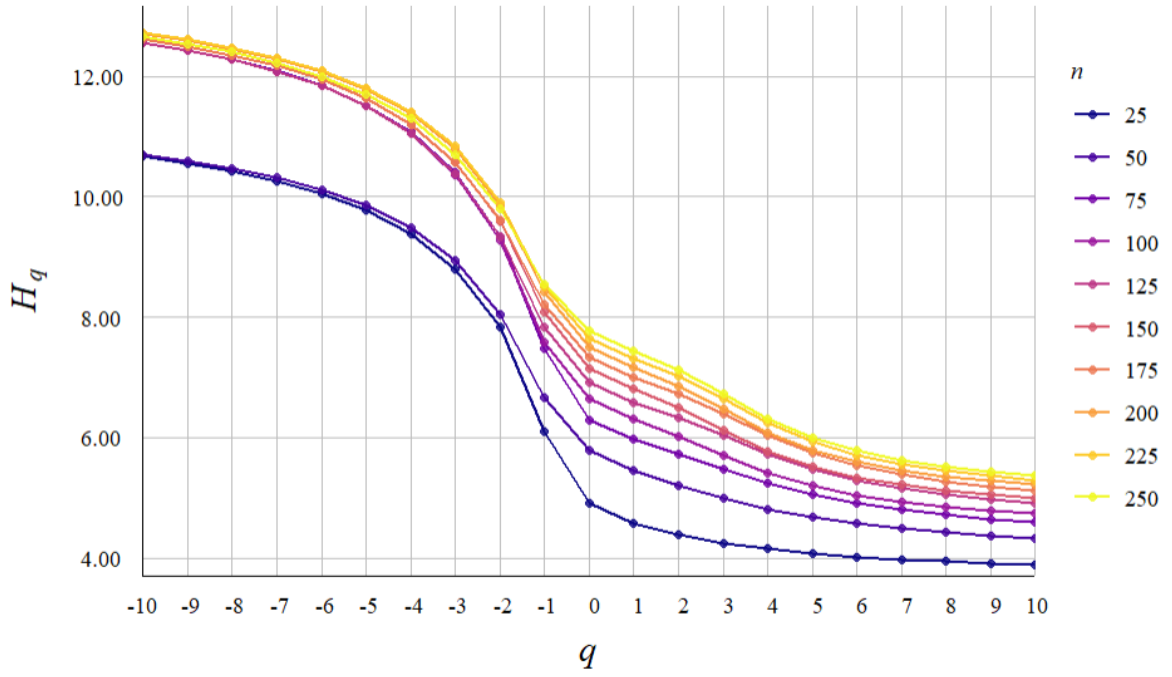


Figure 53: Rényi entropy estimates for Hénon map using binning method without correction with time series size 1,000,000.

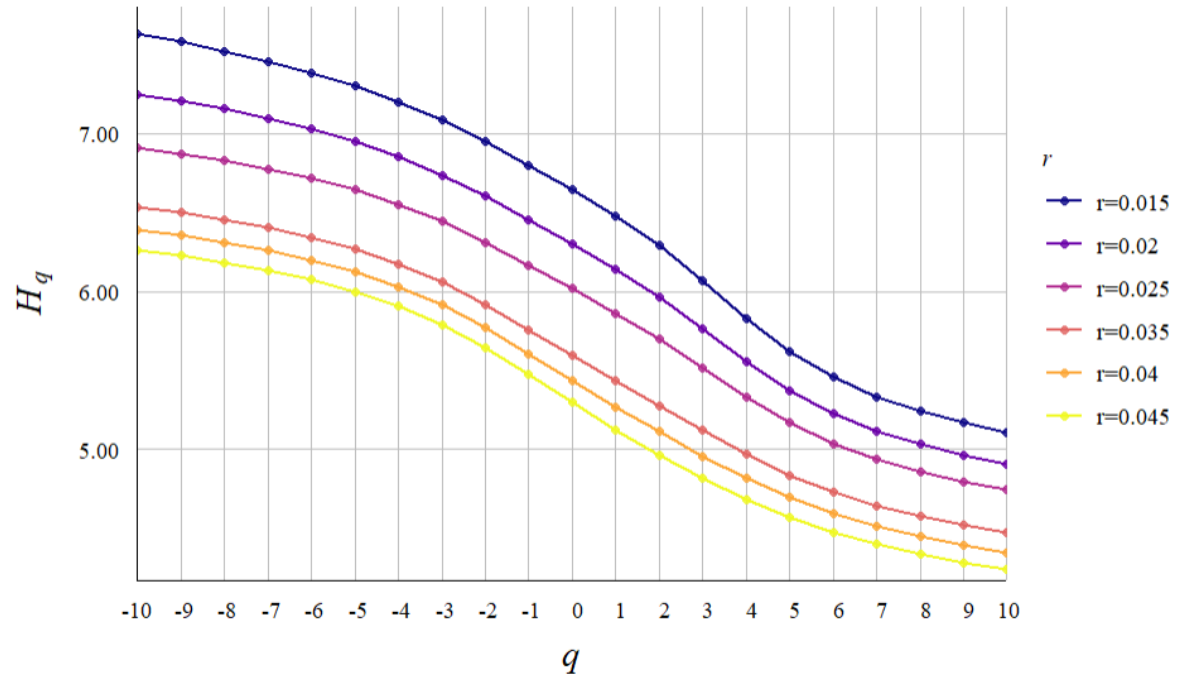


Figure 54: Rényi entropy estimates for Hénon map using correlation sum method with time series size 100,000.

We need to estimate the Rényi dimensions again, but in this case, we have the the-

oretical value for only the case $q = 2$, which is approximately 1.26. In Tables 10 and 11, we see results that are close to the expected value, with the correlation sum method producing more precise estimates.

q	n_1	n_2	Scale 1	Scale 2	$H_2(n_1)$	$H_2(n_2)$	D_2
2	90	190	0.03978547	0.01884575	5.946907	6.796684	1.137260
2	100	200	0.03580692	0.01790346	6.026007	6.857308	1.199314
2	110	210	0.03255175	0.01705092	6.159503	6.926748	1.186533
2	190	110	0.01884575	0.03255175	6.796684	6.159503	1.165837
2	200	90	0.01790346	0.03978547	6.857308	5.946907	1.140128

Table 10: Comparison of Rényi entropy estimates for 100,00 data points by binning method for Hénon map data.

q	r_1	r_2	Scale 1	Scale 2	$H_2(r_1)$	$H_2(r_2)$	D_2
2	0.020	0.045	0.020	0.045	5.955185	4.953144	1.235669
2	0.025	0.035	0.025	0.035	5.686426	5.270471	1.236226
2	0.035	0.015	0.035	0.015	5.270471	6.282357	1.194251
2	0.040	0.020	0.040	0.020	5.100981	5.955185	1.232357
2	0.045	0.025	0.045	0.025	4.953144	5.686426	1.247532

Table 11: Comparison of Rényi entropy estimates for 100,00 data points by correlation sum method for Hénon map data.

In Figure 55, we find the estimates to be more consistent for $q > 0$, but our recommendation is still to relatively compare entropies of different systems using only the same method and the same parameters.

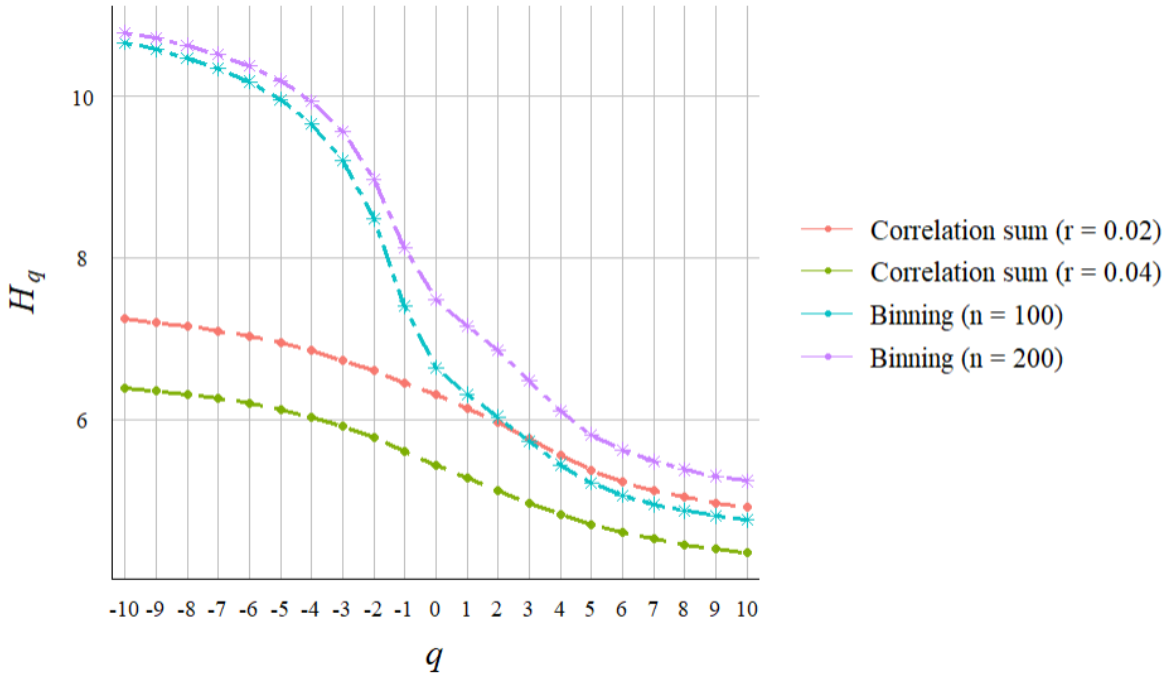


Figure 55: Comparison of Rényi entropy estimation by binning method (n = 100 and 200, ts size = 100,000, without correction) and correlation sum method (r = 0.02 and 0.04, ts size = 100,000) for Hénon map data.

4.9 Lorenz system results

The Lorenz chaotic system is a continuous system with 3 dimensions. We can use the k-NN and binning methods with correction; however, in the case of the binning method, we need to be cautious, as for multidimensional data, it produced less precise results than the k-NN method.

In Figure 56, we see a similar, not so consistent result as for the Hénon map and Cantor set. Thankfully, we can again estimate the Rényi dimensions and validate the entropy estimations. For the Lorenz system $D_2 \approx 2.06$ and in Table 12, we see that the estimations are close to the expected value.

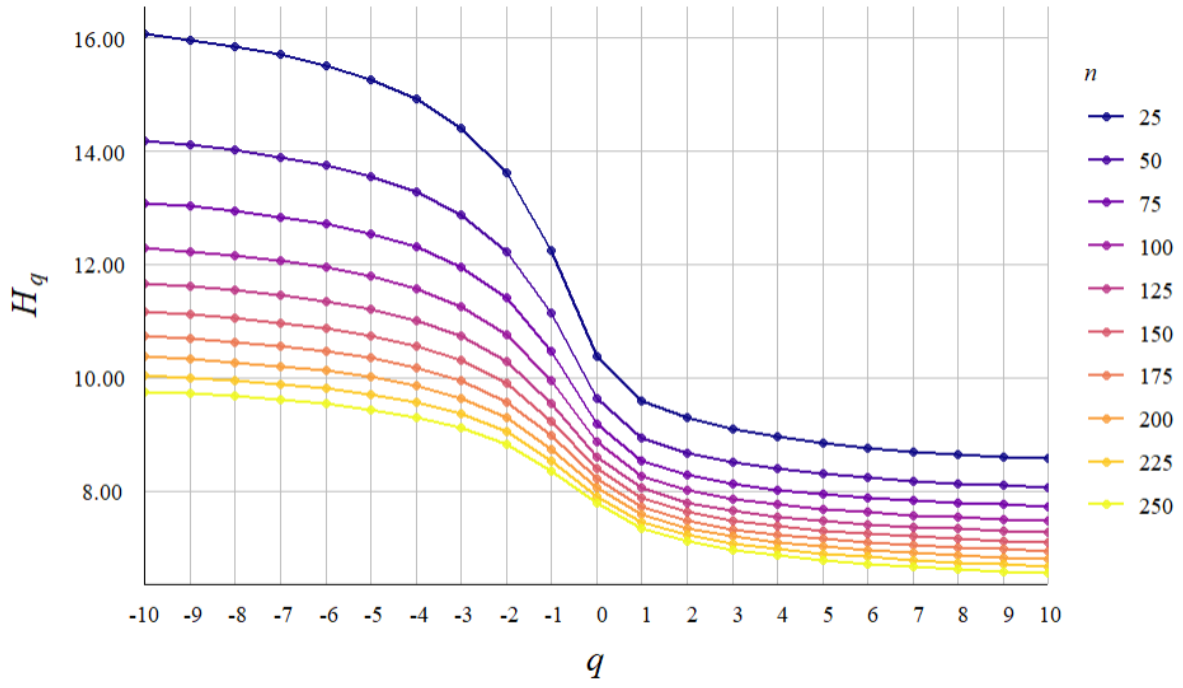


Figure 56: Rényi entropy estimates for Lorenz system using binning method with correction with time series size 1,000,000.

q	n_1	n_2	Scale 1	Scale 2	$H_2(n_1)$	$H_2(n_2)$	D_2
2	90	190	0.8147969	0.3859564	8.720215	10.244974	2.040591
2	100	200	0.7333172	0.3666586	8.940271	10.348102	2.031071
2	110	210	0.6666520	0.3491987	9.137013	10.443915	2.021106
2	190	110	0.3859564	0.6666520	10.244974	9.137013	2.027214
2	200	90	0.3666586	0.8147969	10.348102	8.720215	2.038662

Table 12: Comparison of Rényi entropy estimates for 100,000 data points by binning method for Lorenz system data.

Using the k-NN method, we obtain much better results for the Lorenz system; see Figure 57. In the case of applying the rule of $k = \sqrt{\textbf{ts size}} \approx 316$, we get a clear estimation.

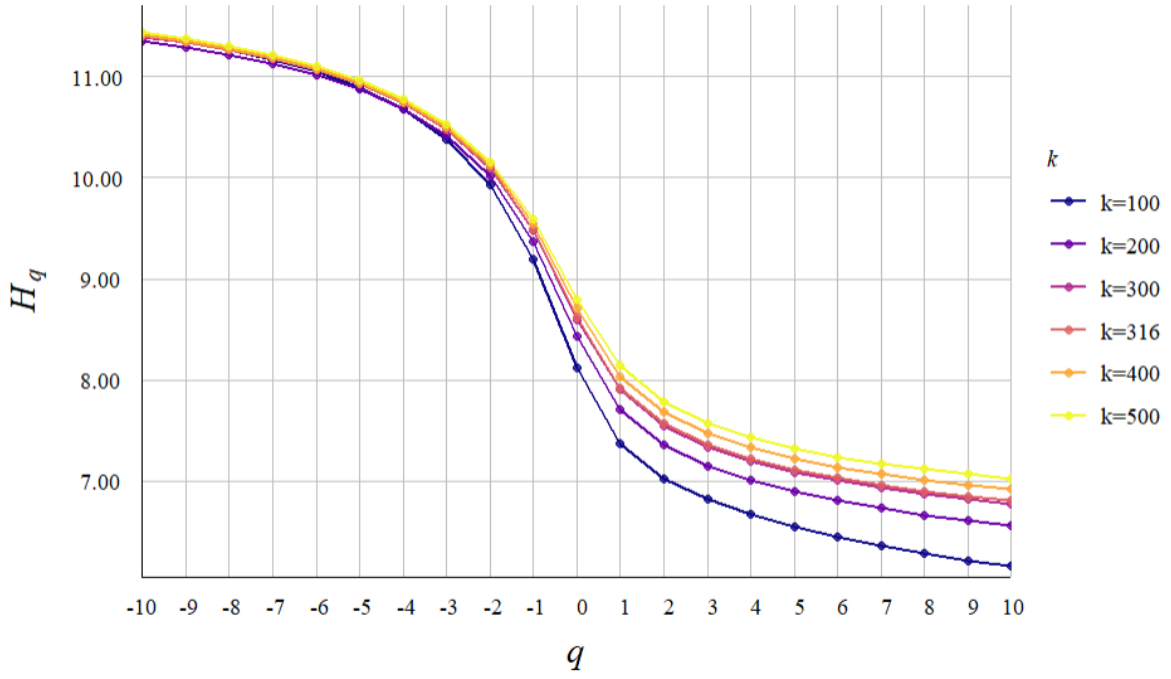


Figure 57: Rényi entropy estimates for Lorenz system using k-NN method with time series size 100,000.

In Figure 58 we can see the comparison of the results obtained by using the binning and k-NN methods. There is a clear difference for the cases of $q < 0$, but for positive q values, the estimations are more similar. Even though we have a rule for choosing the k parameter, we would still consider this an estimate corresponding to the parameter and would only compare the estimated Rényi entropy of this system to estimations of other systems using the same method and parameter.

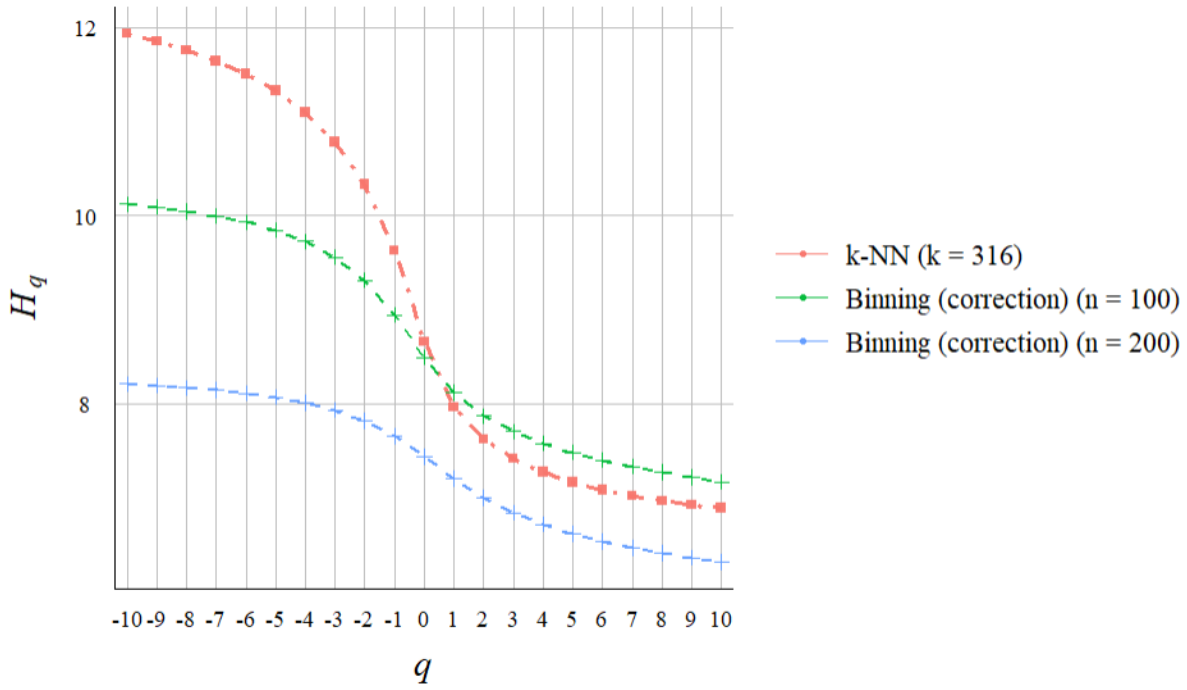


Figure 58: Comparison of Rényi entropy estimation by binning method (n = 100 and 200, ts size = 100,000, with correction) and k-NN (k = $\sqrt{100000}$, ts size = 100,000) for Lorenz system data.

4.10 Rössler system results

The last data set we subject to our methods is the Rössler system data. This is very similar to the Lorenz system; however, one difference is visible in Figure 10. Namely, it is the distribution on the z axis. This exponential-like distribution can cause problems for our methods. In the case of exponential distribution data, we found that the correlation sum method provided by far the worst performance compared to other datasets. Combining this with the continuous nature of the data set we will use the binning and k-NN method as we did for Lorenz system.

At this point, we were expecting the results shown in both Figure 59 and Table 13. The Rényi dimension (of order 2) for the Rössler system is $D_2 \approx 2$ and this is what we see as well.

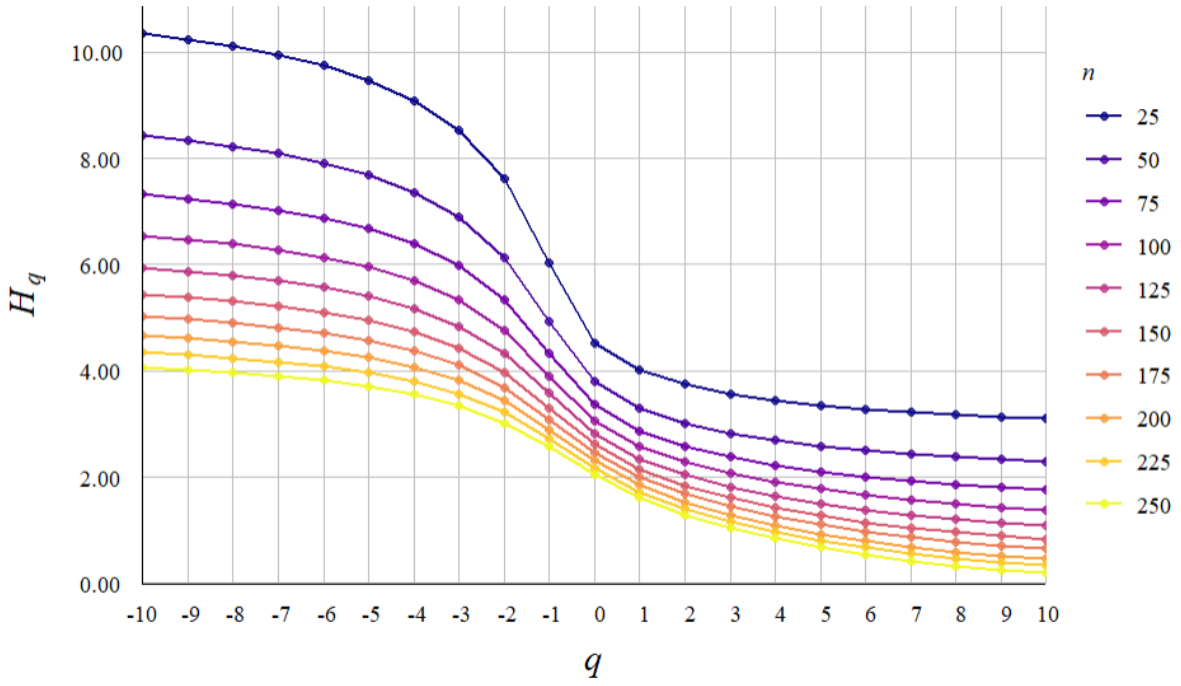


Figure 59: Rényi entropy estimates for Rössler system using binning method with correction with time series size 1,000,000.

q	n_1	n_2	Scale 1	Scale 2	$H_2(n_1)$	$H_2(n_2)$	D_2
2	90	190	0.12443325	0.05894207	8.640604	10.087708	1.936665
2	100	200	0.11198993	0.05599496	8.847588	10.180958	1.923647
2	110	210	0.10180902	0.05332854	9.028965	10.274654	1.926440
2	190	110	0.05894207	0.10180902	10.087708	9.028965	1.937160
2	200	90	0.05599496	0.12443325	10.180958	8.640604	1.929040

Table 13: Comparison of Rényi entropy estimates for 100,000 by binning method for Rössler system data.

The k-NN method provides less consistent results compared to the case of the Lorenz system, see Figure 60. The validation of these results using the Rényi dimension estimates is not possible for this method because the k-NN method does not rely on distance calculation.

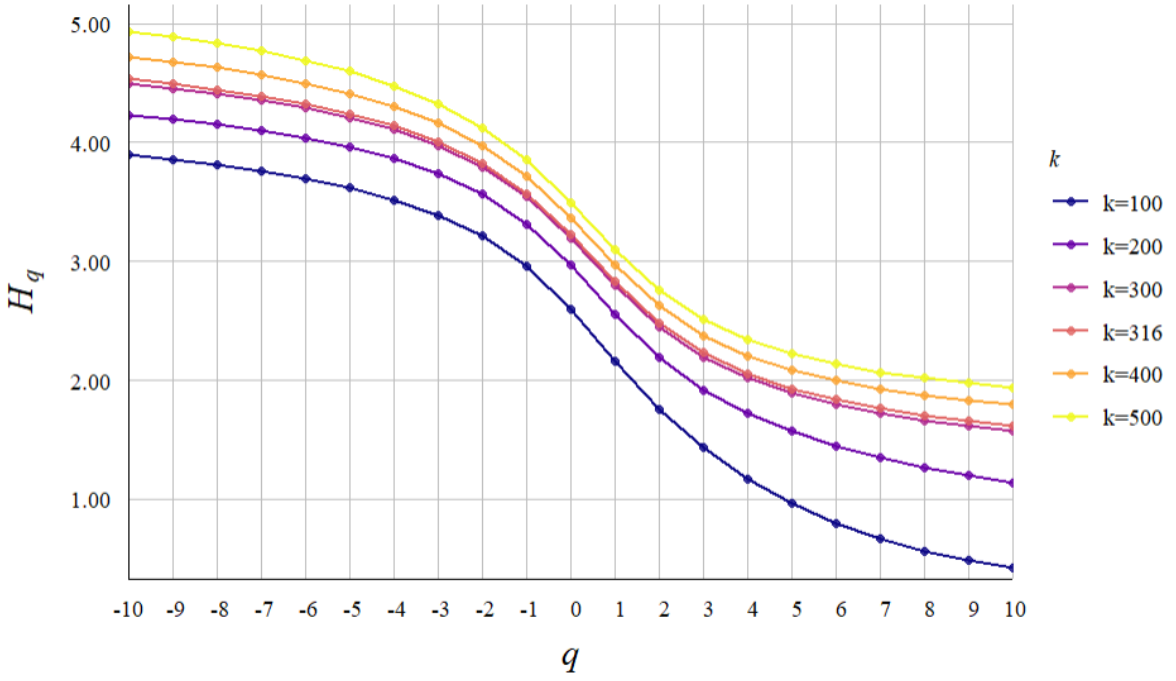


Figure 60: Rényi entropy estimates for Rössler system using k-NN method with time series size 100,000.

Comparing the binning method estimates to the ones obtained by the k-NN method, in Figure 61, we see a fairly good consistency.

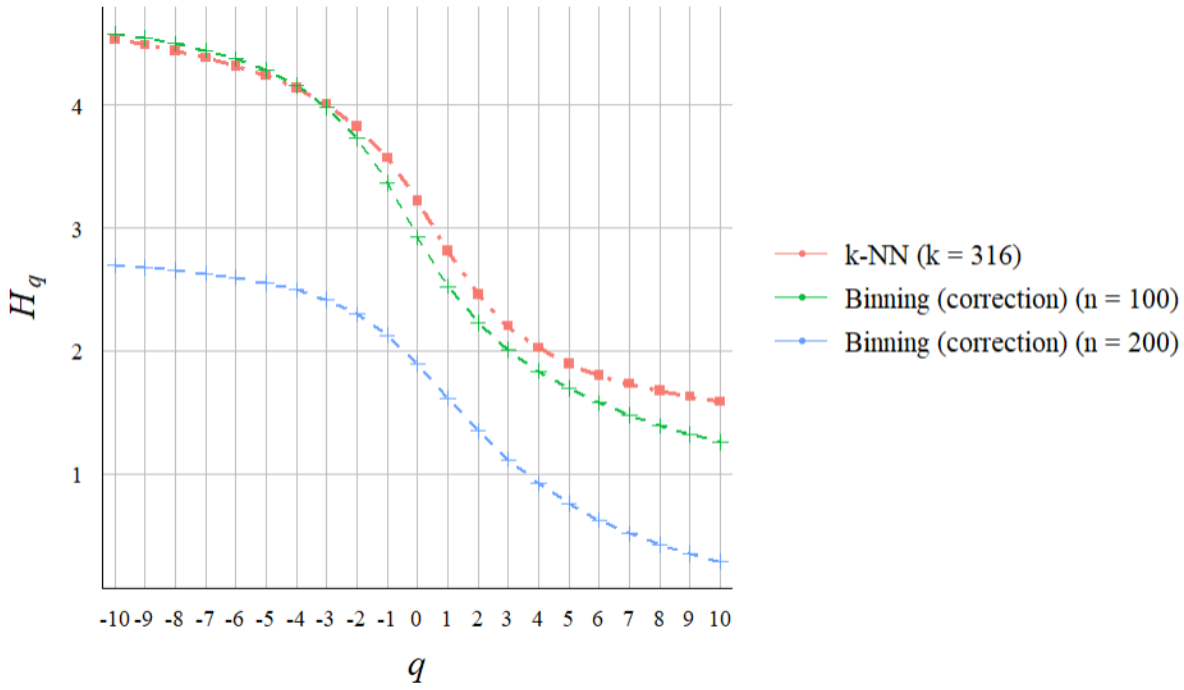


Figure 61: Comparison of Rényi entropy estimation by binning method (n = 100 and 200, ts size = 100,000, with correction) and k-NN (k = 316, ts size = 100,000) for Rössler system data.

Conclusion

In this thesis, we aimed to study and compare numerical estimation methods with the purpose of finding the most suitable generalized Rényi entropy estimator across different types of distributions for finite data sets.

We have evaluated three estimation methods: the binning method, the correlation sum method, and the k-nearest neighbors method. For this purpose, we used a range of synthetic data, including discrete and continuous distributions with known analytical entropy values, and complex systems exhibiting fractal and chaotic structure.

Both the binning and correlation sum methods showed high accuracy for discrete distributions, especially when the parameter selection was evident. For the binning method, this meant that the number of bins matched the number of possible states in the data. The correlation sum and triplet-based methods were dependent on the choice of the radius parameter and struggled with data that had heavy-tailed distributions.

The k-NN estimator was not usable on discrete distributions due to problems with repetitions of nearest-neighbor distances, but it performed well for continuous distributions. In comparison to the other methods, the k-NN-based estimator is relatively less sensitive to parameter selection and does not require correction in continuous distributions.

For continuous distributions, the binning method required a correction term to counterbalance the discretization effects. After this correction, the estimator provided more accurate estimates, although they proved to be less reliable in higher-dimensional settings.

None of the estimators were able to provide direct, unquestionable values for complex systems such as the Lorenz and Rössler attractors, and for fractal structures such as the Cantor set and Hénon map. We observed consistency with theoretical Rényi dimensions for smaller resolutions for the binning and correlation sum method. However, the estimates show a strong dependence on the parameters. Using the general rule of k parameter selection (k should not exceed the square root of the number of data points), the k-NN method proved to be the most consistent and reliable method for the Lorenz and Rössler systems.

Overall, the k-NN method proves to be the best approach for continuous and multidimensional data, while the binning method remains the simplest and fastest option for discrete or low-dimensional cases. The correlation sum method can be useful for discrete

cases but is less reliable for continuous distributions.

Our work revealed a surprising discrepancy between the different Rényi entropy estimators, as well as the lack of a clear universal option. Instead, the choice between these methods needs to be made with consideration of the characteristics of the studied data set.

Future research could focus on further refining the parameter selection in the Rényi entropy estimators, defining the correction needed to use the correlation sum method for continuous cases, and improving the correction term for the binning method.

References

[1] [AI Assistance with R code efficiency], (2026). Online. In: ChatGPT. Version: GPT-5.5 Thinking. Available from: OpenAI, https://chat.openai.com/. Accessed: 2026-05-03. Prompt/task: Assistance with R function for Correlation sum and k-NN methods, improving run time.

[2] [AI Assistance with R graphs and LaTeX formatting], (2026). Online. In: ChatGPT. Version: GPT-5.5 Thinking. Available from: OpenAI, https://chat.openai.com/. Accessed: 2026-05-03. Prompt/task: Assistance with R graphing functions, insertion of graphs and data.

[3] Beygelzimer, A., Kakade, S., Langford, J.: *FNN: Fast Nearest Neighbor Search Algorithms and Applications*. R package version 1.1.4.1, 2024.

[4] Blahut, R. E.: *Principles and Practice of Information Theory*. Addison–Wesley Longman Publishing Co., Inc., 1987.

[5] Bodnarchuk, I., Mishura, Y., Ralchenko, K.: Properties of the Shannon, Rényi and other entropies: dependence in parameters, robustness in distributions and extremes. Preprint, 2024.

[6] Cover, T. M., Thomas, J. A.: *Elements of Information Theory*. John Wiley & Sons, New York, 1999.

[7] Dowle, M., Srinivasan, A.: *data.table: Extension of Data.frame*. R package version 1.17.0, 2025.

[8] Golshani, L., Pasha, E.: Rényi entropy rate for Gaussian processes. *Information Sciences* 180.8 (2010): 1486-1491.

[9] Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. *Physica D* 9 (1983), 189–208.

[10] Jizba, P., Arimitsu, T.: The world according to Rényi: thermodynamics of multifractal systems. *Annals of Physics* 312 (2004), 17–59.

[11] Kozachenko, L. F., Leonenko, N. N.: Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii* 23 (1987), 95–100.

[12] Leonenko, N., Pronzato, L., Savani, V.: A class of Rényi information estimators for multidimensional densities. *The Annals of Statistics* 36 (2008), 2153–2182.

[13] Lorenz, E. N.: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20 (1963), 130–141.

[14] Nadarajah, S., Kebe, M.: Explicit expressions for most common entropies. *Entropy* 25.3 (2023): 534.

[15] Pawelzik, K.,Schuster, H. G.: Generalized dimensions and entropies from a measured time series. *Physical Review A*, 35(1) (1987), 481.

[16] R Core Team.: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. Available at: https://www.R-project.org/

[17] Rényi, A.: On measures of entropy and information. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley (1961), pp. 547–561.

[18] Rényi, A.: *Probability Theory*. Akadémiai Kiadó, Budapest, 1970.

[19] Rössler, O. E.: An equation for continuous chaos. *Physics Letters A* 57 (1976), 397–398.

[20] Sanei Tabass, M., Mohtashami Borzadaran, G. R., Amini, M.: Rényi entropy in continuous case is not the limit of discrete case. *Mathematical Sciences and Applications E-Notes* 4 (2016), 113–117.

[21] Schauberger, P., Walker, A.: *openxlsx: Read, Write and Edit XLSX Files*. R package version 4.2.8, 2025.

[22] Shannon, C. E.: A mathematical theory of communication. *The Bell System Technical Journal* 27 (1948), 379–423.

[23] Tabass, M. S., Borzadaran, G. R. M., Amini, M.: Rényi entropy in continuous case is not the limit of discrete case. *Mathematical Sciences and Applications E-Notes* 4 (2016), 113–117.

[24] Tsallis, C.: Possible generalization of Boltzmann–Gibbs statistics. *Journal of Statistical Physics* 52 (1988), 479–487.

[25] Wickham, H., Bryan, J.: *readxl: Read Excel Files*. R package version 1.4.5, 2024.

[26] Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, 2016.

[27] Wickham, H.: *reshape2: Flexibly Reshape Data*. R package version 1.4.4, 2020.

[28] Wickham, H., Vaughan, D., Girlich, M.: *tidyr: Tidy Messy Data*. R package version 1.3.1, 2024.