Article

Rényi Entropy Estimation via Affine Transformations of Generalized Means

[Alessandro Gecchele](#) *

Article

Rényi Entropy Estimation via Affine Transformations of Generalized Means

Eng. Alessandro Gecchele [ID]

Brescia (Italy); a.gecchele70@gmail.com

Abstract: Rényi entropy is a generalization of Shannon entropy that enables the analysis of different aspects of the informative structure of probability distributions. This family of entropy measures is particularly valuable in scientific and technological contexts requiring the classification, comparison, or characterization of probability distributions, discrete stochastic processes, or discretized dynamical systems. Despite their theoretical appeal, Rényi entropies are challenging to estimate from empirical data—especially in settings involving limited samples, high dimensionality, or highly irregular behaviors. To address these difficulties, we propose a general-purpose estimation method applicable across the entire Rényi family. The method is based on affine transformations that enhance generality, robustness, and accuracy. In particular, the estimation of collision entropy achieves optimality in a single step. To help visualize and understand the effect of the transformation, we also introduce a geometric framework that represents probability distributions from the point of view of their entropic state. This provides intuitive insight into how our estimator works and why it is effective in practice.

Keywords: estimation of Rényi entropies

1. Introduction

1.1. From Probability Theory to Information Theory

Probability theory plays a central role in modeling complex systems that operate under uncertainty. From Cardano's early reflections on dice games of chance (*De Ludo Aleae*, ~1564) [1] and the correspondence between Fermat and Pascal on the problem posed by Pacioli (1654) [2], to Kolmogorov's axiomatic formalization (1933) [3], the discipline of probability theory has gradually evolved from its heuristic origins into a mathematically rigorous and essential framework. Subsequently, information theory enabled a deeper exploration of the structure of probability distributions by using the concept of information entropy. Drawing inspiration from Boltzmann's thermodynamic formulation, Shannon in 1948 [4] defined the entropy of a discrete probability distribution $\boldsymbol{p}$ as

$$H_S(\boldsymbol{p}) = -\sum_{k=1}^{n} p_k \log p_k \tag{1}$$

where p_k denotes the probability of the kth event, and n is the number of possible outcomes. However, this fundamental quantity does not always distinguish between distributions with different internal structures. To address this limitation, Rényi in 1961 [5] introduced a parametric family of entropy measures, generalizing Shannon's definition:

$$H_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha} \log \sum_{k=1}^{n} p_k^\alpha \qquad \alpha > 0, \ \alpha \neq 1. \tag{2}$$

As $\alpha \to 1$, the Rényi entropy converges to the Shannon entropy, preserving theoretical consistency. Varying α, the measure emphasizes different parts of the distribution, enabling finer structural analysis of probability distributions.

1.2. The Challenge of Accurate Entropy Estimation

Rényi's axiomatic approach offered a flexible generalization with theoretical applications in communication theory [6,7], statistical mechanics [8,9], cryptography [10,11] and artificial intelligence [12,13]. However, estimating Rényi entropies from empirical data is difficult due to structural and statistical factors that compromise result reliability [14–16], including:

(a) *Zero-frequency problem:* rare events may be entirely absent from finite samples, leading to underestimated entropy—especially in heavy-tailed or sparse distributions;

(b) *Paradox of data sparsity:* as data become sparser in the sample space, the chance of observing all outcomes in correct relative proportions declines, paradoxically shrinking the empirical entropy even as statistical unreliability increases;

(c) *Amplification of errors:* Entropy's nonlinear dependence on probabilities means that small empirical errors can be magnified, yielding high variance and instability;

(d) *Curse of dimensionality:* as the dimensionality increases, the event space grows exponentially, making data sparse and many outcomes rare or unobserved;

(e) *Sensitivity to outliers and noise:* anomalous events or random fluctuations can distort frequencies and cause erratic entropy values.

These challenges are particularly acute in small-sample regimes, high-dimensional settings, and data derived from high-entropy processes—all common in modern data analysis. As a result, even well-established estimation techniques often produce biased or unstable results.

1.3. Related Works and Previous Contributions

The estimation of members of the family of Rényi entropies—including Shannon entropy ($\alpha \to 1$), collision entropy ($\alpha = 2$), and min-entropy ($\alpha \to \infty$)—has been widely investigated and has garnered increasing attention due to its relevance in various applicative fields. Immediately after the invention of these indices, it was soon recognized that their naive plug-in estimators suffer from high bias and variance in small-sample or high-dimensional regimes. Later works proposed corrections and non-parametric models [17,18], though these remained limited in effectiveness. In response to these challenges, more efforts have progressively been directed toward developing increasingly sophisticated estimation methods: [10,15,16,19–89]. Despite these improvements, an universally recognized method capable of expressing optimal performance in any situation has not yet been achieved. Thus, in the continued absence of a definitive and unique procedure, the pursuit of a robust general-purpose estimator remains an open and central challenge in the study of Rényi entropies. It is also worth noting that, due to these difficulties, many researchers have proposed alternative indices, equipped with their customized estimators, such as Onicescu's informational energy [90], approximate entropy [91], sample entropy [92], multiscale entropy [93], permutation entropy [94] and bubble entropy [95]. Beyond these isolated alternatives, one of the few comprehensive frameworks comparable in scope to the Rényi entropy family is the parametric family of Tsallis entropies [96], whose applicability is hindered by the lack of key properties such as additivity, scale invariance, and ease of normalization.

1.4. Motivation and Contribution of This Paper

This work is motivated by the persistent lack of a unified and widely accepted method for estimating Rényi entropies—a gap that has remained open for more than sixty years since their theoretical inception. We propose a novel method for the accurate estimation of Rényi entropies that requires minimal parameter tuning and can be used under diverse data conditions. The method effectively addresses the challenges mentioned in the previous paragraph. Empirical evidence confirms its performance in low-sample, high-dimensional, and high-entropy settings. At the core of the procedure lies an affine transformation[1] applied to the empirical frequency distribution, introducing a structured geometric regularization whose overall effect is a global linear stretching. This transformation regularizes the geometry of the input space prior to the entropy computation, leading to an improved approximation of the true Rényi entropy via the Rényi mean[2]. This approach stems from the idea that the empirical entropies reside in a context that is different from the context of theoretical entropies and a transformation is necessary for the "translation" of these measures from one context to the other. The proposed transformation reorients and rescales the data in a principled manner, considerably reducing the estimation error for any entropy index and directly reaching the optimality for the case of $\alpha = 2$ (collision entropy).

2. Overview on Materials and Methods of Information Theory

2.1. Transforming a Discrete-State Stochastic Process into a Probability Distribution

Consider a discrete-state stochastic process (DSP) whose sequence of infinite samples $s_{-\infty}, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_{\infty}$ presents values belonging to an alphabet X_q containing q ordered symbols. Let $\Omega(q, d)$ be a d-dimensional discrete sample space resulting from the Cartesian product d times of X_q.

$$\Omega(q, d) = \underbrace{X_q \times X_q \times \ldots \times X_q}_{d \ times}, \tag{3}$$

and let $n = q^d$ be the cardinality of $\Omega(q, d)$. Each elementary event e_k, with $k \in \{1, 2, \ldots, n\}$, is uniquely identified by a vector with d coordinates $(x1_k, x2_k, \ldots, xd_k)$, with $x1_k, x2_k, \ldots, xd_k \in X_q$. According to the procedure indicated by Shannon in [4] at pages 5 and 6, the infinite sequence of samples constituting the DSP can be transformed into occurrences $\#(e_k)$ of the elementary events of $\Omega(q, d)$ by progressively considering all the d-grams taken from the samples as if they were the vector coordinates of the events and counting the number of times each vector appears in the sequence. Then, according to the frequentist definition of probability, the final resulting probability distribution is expressible in set theory notation as

$$\boldsymbol{p}(\Omega(q, d))_{DSP} = \left\{ p(\boldsymbol{e}_k)_{DSP} = \frac{\#(\boldsymbol{e}_k)_{DSP}}{\sum_{k=1}^{n} \#(\boldsymbol{e}_k)_{DSP}} \,\middle|\, \boldsymbol{e}_k \in \Omega(q, d) \right\}. \tag{4}$$

In the following, in the absence of ambiguity, $\boldsymbol{p}(\Omega(q, d))_{DSP}$—that is, a DPD obtained by elaborating the data of a DSP— will be indicated with the bold symbol $\boldsymbol{p}$ and the probability of the kth event with p_k.

---

1.   *An affine transformation* is a mapping of the form $f(\boldsymbol{x}) = \boldsymbol{Ax} + \boldsymbol{B}$, where A is a linear operator (e.g., a matrix) and B is a fixed vector. It preserves points, straight lines, and planes, and includes operations such as scaling, rotation, translation, and shear.

2.   The *Rényi mean* of order α of a probability distribution is a generalized averaging operator defined as ${}^R \overline{\boldsymbol{p}}_\alpha = \left(\sum_{k=1}^{n} p_k^\alpha \right)^{1/(\alpha-1)}$. It forms the basis of Rényi entropy, which is defined as $\eta_\alpha(\boldsymbol{p}) = -\log_n \left({}^R \overline{\boldsymbol{p}}_\alpha \right)$.

2.2. Integer-Order Rényi α-Entropies as Synthetic Indices for the Characterization of DPDs

The family of integer-order Rényi α-entropies offers synthetic indices that characterize *DPD*s in terms of their informative content [5]. They are defined as:

$$
\begin{aligned}
\alpha = 1 \qquad & H_1(\boldsymbol{p}) \triangleq - \sum_{k=1}^{n} p_k \log p_k \\
\alpha \in \mathbb{N}^+ \qquad \alpha \neq 1 \qquad & H_\alpha(\boldsymbol{p}) \triangleq \frac{1}{1-\alpha} \log \Big(\sum_{k=1}^{n} p_k^\alpha \Big) \qquad\qquad 0 \leq H_\alpha(\boldsymbol{p}) \leq \log n \\
\alpha \to \infty \qquad & H_\infty(\boldsymbol{p}) \triangleq -\log max\{\boldsymbol{p}\}.
\end{aligned}
\tag{5}
$$

The corresponding specific entropies are then defined as

$$
\begin{aligned}
\alpha = 1 \qquad & \eta_1(\boldsymbol{p}) \triangleq \frac{H_1(\boldsymbol{p})}{\log n} = - \sum_{k=1}^{n} p_k \log_n p_k \\
\alpha \in \mathbb{N}^+ \qquad \alpha \neq 1 \qquad & \eta_\alpha(\boldsymbol{p}) \triangleq \frac{H_\alpha(\boldsymbol{p})}{\log n} = \frac{1}{1-\alpha} \log_n \Big(\sum_{k=1}^{n} p_k^\alpha \Big) \qquad 0 \leq \eta_\alpha(\boldsymbol{p}) \leq 1 \\
\alpha \to \infty \qquad & \eta_\infty(\boldsymbol{p}) \triangleq \frac{H_\infty(\boldsymbol{p})}{\log n} = -\log_n max\{\boldsymbol{p}\}.
\end{aligned}
\tag{6}
$$

Specific entropies are preferred over plain ones because:
(a) They range in $[0,1]$ via min-max normalization, which provides a unique bounded scale that greatly facilitates the interpretation of entropy values at any scale;
(b) They minimize the dependence on the sample space size n;
(c) They resolve doubts about the choice of the logarithm base, highlighting scale invariance;
(d) Entropies in base b are then given by $H_\alpha(\boldsymbol{p}, b, n) = \eta_\alpha(\boldsymbol{p}) \cdot \log_b n$.

2.3. Rényi Entropy Rates

Unlike Rényi entropies, whose utility is mainly related to the classification of *DPD*s, Rényi entropy rates are important theoretical quantities useful for the characterization of *DSP*s [61,72]; they are defined as

$$
H_\alpha'(DSP) \triangleq \lim_{d \to \infty} \frac{1}{d} H_\alpha(\boldsymbol{p}(\Omega(q,d))_{DSP}) \qquad 0 \leq H_\alpha'(DSP) \leq \log q.
\tag{7}
$$

Moreover, it is known that, for strongly stationary *DSP*, any Rényi entropy rate converges to the same limit of a sequence of Cesaro means of conditional entropies:

$$
H_\alpha'(DSP) = \lim_{d \to \infty} H_\alpha(\boldsymbol{p}(A_d)|\boldsymbol{p}(A_1 \times A_2 \times \cdots \times A_{d-1})).
\tag{8}
$$

and, as conditional Rényi entropies preserve the chain rule [97–99], they can also be calculated as

$$
H_\alpha'(DSP) = \lim_{d \to \infty} [H_\alpha(\boldsymbol{p}(\Omega(q,d))_{DSP}) - H_\alpha(\boldsymbol{p}(\Omega(q,d-1))_{DSP})].
\tag{9}
$$

Specific Rényi entropy rates are defined by the following min-max normalization:

$$
\begin{aligned}
\eta_\alpha'(DSP) = \frac{H_\alpha'(DSP)}{\log q} &= \lim_{d \to \infty} \frac{[H_\alpha(\boldsymbol{p}(\Omega(q,d))_{DSP}) - H_\alpha(\boldsymbol{p}(\Omega(q,d-1))_{DSP})]}{\log q} = \\
&= \lim_{d \to \infty} [d\, \eta_\alpha(\boldsymbol{p}(\Omega(q,d))_{DSP}) - (d-1)\, \eta_\alpha(\boldsymbol{p}(\Omega(q,d-1))_{DSP})]
\end{aligned}
\tag{10}
$$

with $0 \leq \eta_\alpha'(DSP) \leq 1$.

2.4. Relationship Between Specific Rényi Entropy Rate and Specific Rényi Entropy

The following relationship holds:

$$\eta_\alpha'(DSP) = \lim_{d \to \infty} \eta_\alpha(\boldsymbol{p}(\Omega(q,d))_{DSP}) \tag{11}$$

Accordingly, the specific Rényi entropy rate $\eta_\alpha'(DSP)$ can be computed via either (10) or (11), with the former being preferable due to faster convergence.

2.5. Rényi Generalized Dimensions

Rényi entropies and their associated rates provide a unified framework for analyzing probability distributions and classifying stochastic processes. A closely related concept in multifractal analysis is the *Rényi generalized dimension* [100], defined as

$$D_\alpha(\Theta(d,\epsilon)) = \frac{1}{\alpha-1} \cdot \frac{\log(\sum_i p_i(\Theta(d,\epsilon))^\alpha)}{\log \epsilon}, \tag{12}$$

where $\Theta(d,\epsilon)$ denotes a partition of the state space into d-dimensional cells of linear size ϵ, and $p_i(\Theta(d,\epsilon))$ is the probability that a trajectory of the dynamical system (DS) crosses the i-th cell:

$$p_i(\Theta(d,\epsilon)) = \frac{\mu(\text{cell}_i)}{\sum_{k=1}^n \mu(\text{cell}_k)}, \quad \sum_{i=1}^n p_i(\Theta(d,\epsilon)) = 1,$$

with $\mu(\text{cell}_i)$ denoting the number of visits in the i-th d-cell by the trajectory.

2.6. Rényi Generalized Dimension Rates

In analogy with entropy rates, we define the *Rényi generalized dimension rate* as

$$D_\alpha'(DS) = \lim_{\epsilon \to 0} \frac{1}{\alpha-1} \cdot \frac{\log(\sum_i p_i(\Theta(d,\epsilon))^\alpha)}{\log \epsilon} = \lim_{\epsilon \to 0} \frac{H_\alpha(\boldsymbol{p})}{\log(1/\epsilon)}, \tag{13}$$

where $H_\alpha(\boldsymbol{p})$ is the Rényi entropy of order α associated with the distribution $\boldsymbol{p} = \{p_i\}$ induced by the partition.

Letting $q = 1/\epsilon$, the number of partition elements scales as $n = q^d$. Since the specific Rényi entropy is defined as $\eta_\alpha(\boldsymbol{p}) = H_\alpha(\boldsymbol{p})/\log n$, the dimension rate becomes

$$D_\alpha'(DS) = \lim_{q \to \infty} \frac{\eta_\alpha(\boldsymbol{p}) \cdot \log n}{\log q} = d \cdot \lim_{q \to \infty} \eta_\alpha(\boldsymbol{p}), \qquad 0 \le D_\alpha'(DS) \le d. \tag{14}$$

Here, d is fixed, while q increases. To improve convergence and enable stable numerical estimation, we employ a differential formulation based on logarithmic increments:

$$D_\alpha'(DS) = d \cdot \lim_{q \to \infty} \frac{\eta_\alpha(\boldsymbol{p}(\Omega(q,d))) \cdot \log q - \eta_\alpha(\boldsymbol{p}(\Omega(q-1,d))) \cdot \log(q-1)}{\log q - \log(q-1)} \tag{15}$$

2.7. Rényi Mutual Information

Another key quantity derived from α-Rényi entropies is the α-Rényi mutual information [48], defined for three random variables X, Y, and $(X \times Y)$, with respective events $\{x_k\}$, $\{y_i\}$, and $\{(x_k, y_i)\}$, and corresponding probability distributions $\boldsymbol{p}$, $\boldsymbol{t}$, and $\boldsymbol{u}$:

$$I_\alpha(X;Y) \triangleq \sum_{k=1}^n \sum_{i=1}^m u_{ki} \log \frac{u_{ki}}{p_k t_i} = H_\alpha(\boldsymbol{p}(X)) + H_\alpha(\boldsymbol{t}(Y)) - H_\alpha(\boldsymbol{u}(X \times Y)) \tag{16}$$

The corresponding expression in terms of specific entropy is:

$$\iota_\alpha(X;Y) \triangleq \sum_{k=1}^{n} \sum_{i=1}^{m} u_{ki} \log_n \frac{u_{ki}}{p_k\, t_i} = \eta_\alpha(\boldsymbol{p}(X)) + \eta_\alpha(\boldsymbol{t}(Y)) - \eta_\alpha(\boldsymbol{u}(X \times Y)). \tag{17}$$

2.8. Measures of Entropic Distance Between Distributions

Dissimilarity between two distributions *A* and *B* is often measured using Rényi divergences [101], of which the Kullback–Leibler divergence [102] (also known as relative entropy) is a special case. Since these divergences are not directly derived from Rényi entropies, we instead propose to measure their geometric distance in entropy space using the following formula:

$$D_\eta(\overline{AB}) = \sqrt{\frac{\sum_\alpha (\eta_\alpha(A) - \eta_\alpha(B))^2}{\#\eta_\alpha}}, \qquad 0 \le D_\eta(\overline{AB}) \le 1. \tag{18}$$

where $\#\eta_\alpha$ represents the number of entropy indices used to determine the state. This method provides a symmetric and broader generalization with respect to Rényi divergences, also allowing the comparison of distributions with different event-space cardinalities.

2.9. Two Fundamental Families of Structured Probability Distributions

To support controlled and interpretable experimentation, we define two fundamental families of structured probability distributions that form the basis of our experimental analysis. Designed to span the entire range of distributions over a finite alphabet, these families exhibit distinct structural properties that represent entropy extremes within the distribution space (see Figures 1 and 2). They are:
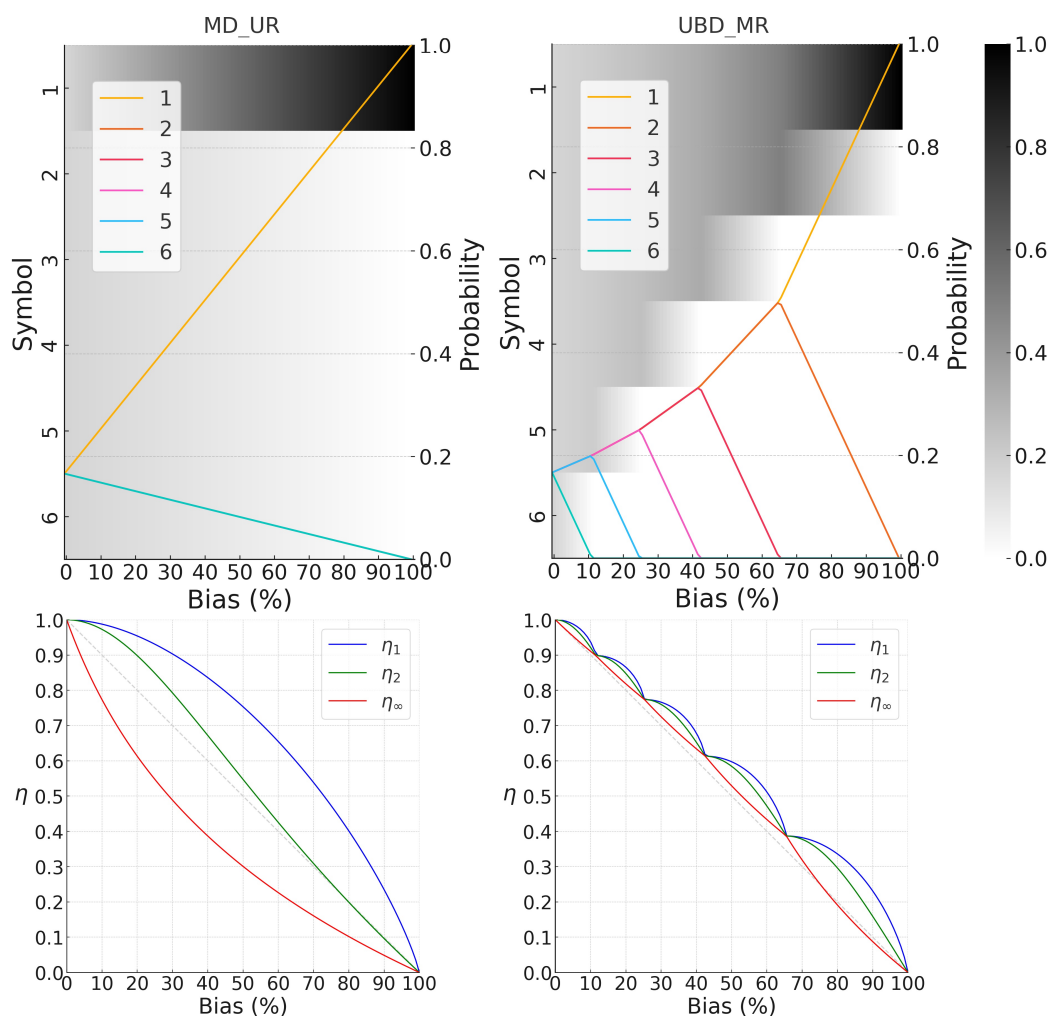
1. *Mono Dominant with Uniform Residues* (`MD_UR`): This family generates distributions in which, starting from the uniform case, a single symbol progressively absorbs more probability mass by subtracting it uniformly from the other symbols. The family can be defined as:

$$p_1 = \frac{1}{q} + \theta\left(1 - \frac{1}{q}\right)$$

$$p_i = \frac{1 - p_1}{q - 1} \quad \text{for } i = 2, \dots, q$$

   where the parameter $\theta \in [0, 1]$ controls the strength of the dominant symbol.

2. *Uniform Block Dominant with Mono Residue* (`UBD_MR`): This family defines structured, biased distributions starting from a uniform base. Probability mass is progressively removed from the last non-zero event and equally redistributed among all preceding ones. The resulting distribution exhibits a dominant block of equal, high-probability values at the front, followed by a single lower residual value, with all subsequent entries set to zero. This construction induces a non-trivial truncation of the support and does not admit a simple closed-form expression.

Figure 1. Differences between the "Mono Dominant with Uniform Residues" (`MD_UR`) family of theoretical probability distributions (left) and the "Uniform Block Dominant with Mono Residue" (`UBD_MR`) family of theoretical probability distributions (right). Above: evolution of the probability mass varying the bias for the two opposite cases. Below: the corresponding evolution of specific Shannon entropy (in blue), collision entropy (in green) and min-entropy (in red).
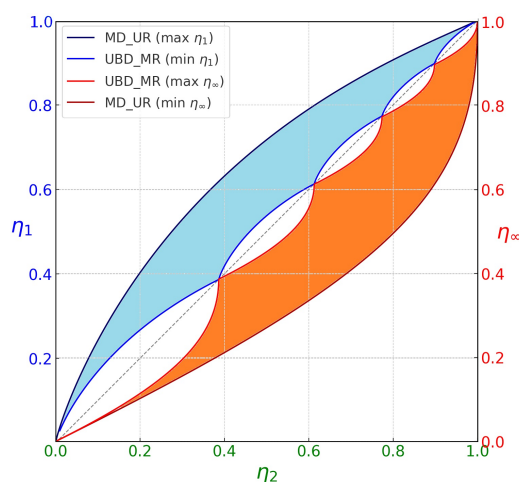
Figure 2. Localization of the possible state points (in terms of entropies) of any theoretical probability distribution associated with a stochastic process with $q = 6$ and arbitrary dimension d. In light blue, the region of state points (η_2, η_1); in orange, the region of state points (η_2, η_∞). The regions are bounded by the state points generated by the `MD_UR` and `UBD_MR` families of probability distributions.

For both families, explicit generation algorithms are provided to support reproducibility and controlled experimentation. These distributions define the entropic bounds that encompass all admissible shapes, from highly concentrated to nearly uniform. They are well suited to reveal how entropy estimators behave under structural constraints and will be used to benchmark performance, assess sensitivity, and identify systematic biases.

Algorithm 1: Generating a Mono Dominant with Uniform Residues (`MD_UR`) distribution based on a given distortion percentage

---

Input: *percentage* (distortion percentage), *q* (number of events)
Output: A vector *prob* representing the biased distribution

1   $base\_prob \leftarrow \frac{1}{q}$;
2   $delta \leftarrow \frac{percentage}{100} \times (1 - base\_prob)$;
3   $prob0 \leftarrow base\_prob + delta$;
4   $other\_prob \leftarrow \frac{1 - prob0}{q-1}$;
5   Initialize *prob* as a vector of size *q*;
6   $prob[0] \leftarrow prob0$;
7   **for** $i \leftarrow 1$ **to** $q-1$ **do**
8       $prob[i] \leftarrow other\_prob$;
9   **return** *prob*;

Algorithm 2: Generating a Uniform Block Dominant with Mono Residue (`UBD_MR`) distribution based on a given distortion percentage

---

Input: *percentage* (distortion percentage), *q* (number of events)
Output: A vector *prob* representing the biased distribution

1   Initialize *prob* as a vector of size *q* with all entries set to $\frac{1}{q}$;
2   $total\_prob\_moved \leftarrow 0$;
3   **for** $i \leftarrow 0$ **to** $q-2$ **do**
4       $total\_prob\_moved \leftarrow total\_prob\_moved + \frac{1}{q-i}$;
5   $total\_to\_redistribute \leftarrow \frac{percentage}{100} \times total\_prob\_moved$;
6   $remaining \leftarrow total\_to\_redistribute$;
7   **for** $i \leftarrow q-1$ **to** 1 **by** -1 **while** $remaining > 0$ **do**
8       $to\_remove \leftarrow \min(prob[i], remaining)$;
9       $prob[i] \leftarrow prob[i] - to\_remove$;
10     $share \leftarrow \frac{to\_remove}{i}$;
11   **for** $j \leftarrow 0$ **to** $i-1$ **do**
12         $prob[j] \leftarrow prob[j] + share$;
13     $remaining \leftarrow remaining - to\_remove$;
14   **return** *prob*;

2.10. Ambiguity Reduction in Entropy-Based Measures

While a single entropy index provides valuable information, it may be insufficient to distinguish between probability distributions that share the same entropy value but differ in structural characteristics. This limitation can be addressed by combining multiple independent and complementary entropy indices—such as different Rényi entropies (e.g., Shannon entropy and collision entropy)—which enhance the ability to differentiate among distributions. For instance, for two distributions with identical Shannon entropy, the further observation of their collision entropy reveals structural differences that would otherwise remain hidden (see Figure 3).
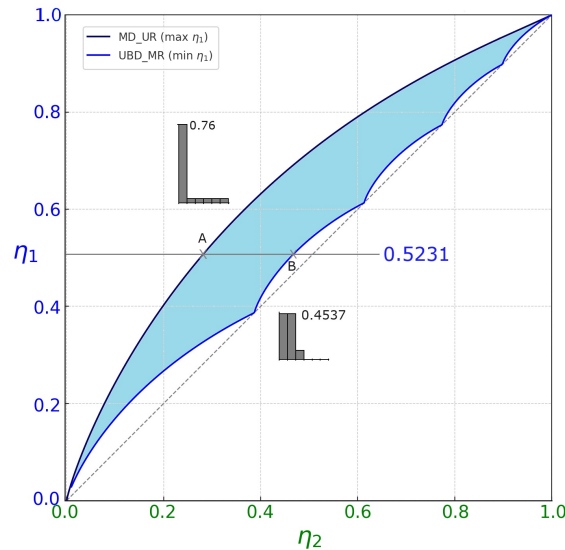
Figure 3. Example of two distinct probability distributions, *A* and *B*, both defined over an alphabet of size $q = 6$, that share the same specific Shannon entropy ($\eta_1(\boldsymbol{p}_A) = \eta_1(\boldsymbol{p}_B)$) but differ in specific collision entropy ($\eta_2(\boldsymbol{p}_A) \neq \eta_2(\boldsymbol{p}_B)$). A comparison based solely on Shannon entropy fails to distinguish them. However, their differing structures become evident when analyzing the entropy pairs ($\eta_2(\boldsymbol{p}_A), \eta_1(\boldsymbol{p}_A)$) and ($\eta_2(\boldsymbol{p}_B), \eta_1(\boldsymbol{p}_B)$), which occupy distinct positions in the entropy plane.

2.11. Interpreting Rényi Entropies as the Negative Logarithm of Generalized Means

The *generalized α-Hölder mean* [103] of a discrete probability distribution $\boldsymbol{p}$ over n elementary events is defined as:

$$^{H}\overline{\boldsymbol{p}}_{\alpha} \triangleq \left(\frac{1}{n} \sum_{i=1}^{n} p_i^{\alpha} \right)^{\frac{1}{\alpha}}, \qquad \frac{1}{n} \leq {}^{H}\overline{\boldsymbol{p}}_{\alpha} \leq 1 \tag{19}$$

This mean interpolates between classical means as α varies: in fact $^{H}\overline{\boldsymbol{p}}_{-1}$ is the harmonic mean, $^{H}\overline{\boldsymbol{p}}_{0}$ is the geometric mean, $^{H}\overline{\boldsymbol{p}}_{1}$ is the arithmetic mean, $^{H}\overline{\boldsymbol{p}}_{2}$ is the root mean square, $^{H}\overline{\boldsymbol{p}}_{\infty}$ is the mean considering the maximum value. Building on this idea, we define the *generalized α-Rényi mean* of a discrete probability distribution $\boldsymbol{p}$ as:

$$^{R}\overline{\boldsymbol{p}}_{\alpha} \triangleq \left(\sum_{k=1}^{n} p_k^{\alpha} \right)^{\frac{1}{\alpha-1}}, \qquad \frac{1}{n} \leq {}^{R}\overline{\boldsymbol{p}}_{\alpha} \leq 1 \tag{20}$$

Accordingly, the Rényi entropy of order α can be expressed as:

$$\eta_{\alpha}(\boldsymbol{p}) = -\log_n \left({}^{R}\overline{\boldsymbol{p}}_{\alpha} \right) \qquad \Longleftrightarrow \qquad {}^{R}\overline{\boldsymbol{p}}_{\alpha} = n^{-\eta_{\alpha}(\boldsymbol{p})} \tag{21}$$

Each order α corresponds to a distinct generalized averaging scheme, providing insight into the distribution's internal structure from complementary informational perspectives. This framework also recovers several well-known entropy measures as special cases:

Shannon entropy ($\alpha \to 1$) from the Shannon mean:

$$\eta_1(\boldsymbol{p}) = -\log_n \left({}^{R}\overline{\boldsymbol{p}}_{1} \right), \qquad {}^{R}\overline{\boldsymbol{p}}_{1} \triangleq \prod_{k=1}^{n} p_k^{p_k} \qquad \left(\text{if } p_k = 0 \Longrightarrow p_k^{p_k} = 1 \right) \tag{22}$$

Collision entropy ($\alpha = 2$) from the collision mean:

$$\eta_2(\boldsymbol{p}) = -\log_n \left({}^{R}\overline{\boldsymbol{p}}_{2} \right), \qquad {}^{R}\overline{\boldsymbol{p}}_{2} \triangleq \sum_{k=1}^{n} p_k^{2} \tag{23}$$

Min-entropy ($\alpha \to \infty$) from the max-mean, equal to the maximum probability:

$$\eta_\infty(\boldsymbol{p}) = -\log_n\left(^R\overline{\boldsymbol{p}}_\infty\right), \qquad {}^R\overline{\boldsymbol{p}}_\infty \triangleq p_1 \quad \text{with } p_1 \geq \cdots \geq p_n \tag{24}$$

From this perspective, each specific Rényi entropy can be interpreted as the negative logarithm (base n) applied to the corresponding generalized Rényi mean, resulting in a normalized, scale-invariant measure of an intrinsic property of the probability distribution.

3. Practical Implementation of Methods of Information Theory

This section addresses the transition from idealized models—based on exact knowledge of discrete probability distributions—to practical scenarios involving finite samples, where the true underlying distributions are unknown. By reformulating the theoretical framework in terms of observed relative frequencies, it enables inference and estimation of properties of the underlying generative process. This procedure is composed of three steps: (a) Determining relative frequencies from samples; (b) Calculating *apparent* Rényi means/entropies of the relative frequencies; (c) Inferring the *actual* Rényi means/entropies of the underlying probability distribution.

3.1. Converting a Realization into a Relative Frequency Distribution

In practical applications, the theoretical procedure described in § 2.1 can be adapted as follows: consider the L samples $s_1, s_2, \ldots, s_d, s_{d+1}, \ldots, s_L$ of a realization r, extracted from a *DSP* represented using q ordered symbols. Each d-gram, consisting of d consecutive samples from r, is treated as the occurrence of an elementary event in the d-dimensional sample space $\Omega(q,d)$, where each d-gram defines a unique vector in this space. For example, the first two d-grams taken from r_q, $(x_1, x_2, ..., x_d)$ and $(x_2, x_3, ..., x_{d+1})$ identify the first occurrences of two elementary events. The count of the occurrences of events is performed for all the d-grams progressively identified in the sequence of the samples of r. The absolute frequency of each elementary event $\#(e_k)$ is normalized by the total number of occurrences, ($N = \sum_{k=1}^n \#(e_k)_r = L - d + 1$), yielding the corresponding relative frequency $f(e_k)_r$. The final resulting *RFD* is expressible in set theory notation as

$$\boldsymbol{f}(\Omega(q,d))_r = \left\{ f(e_k)_r = \frac{\#(e_k)_r}{\sum_{k=1}^n \#(e_k)_r} \,\middle|\, e_k \in \Omega(q,d) \right\} \tag{25}$$

where $\sum_{k=1}^n \#(e_k)_r$ represents the total number of observed d-grams. For brevity, when there is no risk of ambiguity, an RFD $\boldsymbol{f}(\Omega(q,d))_r$ derived from a realization within a given sample space will be denoted simply as $\boldsymbol{f}$, and f_k represents the relative frequency attributed to the kth event.

3.2. Rényi Mean and Specific Rényi Entropy of a Relative Frequency Distribution

For a relative frequency distribution, the empirical Rényi mean is defined as:

$$^R\overline{\boldsymbol{f}}_\alpha = \left(\sum_{k=1}^n f_k^\alpha \right)^{\frac{1}{\alpha-1}} \qquad \left(\frac{1}{N} \leq {}^R\overline{\boldsymbol{f}}_\alpha \leq 1 \right) \tag{26}$$

Consequently, the empirical specific Rényi entropy results

$$\eta_\alpha(\boldsymbol{f}) = -\log_n\left(^R\overline{\boldsymbol{f}}_\alpha\right) \quad \Longleftrightarrow \quad {}^R\overline{\boldsymbol{f}}_\alpha = n^{-\eta_\alpha(\boldsymbol{f})} \tag{27}$$

3.3. Apparent vs. Actual Specific Entropies in Controlled Probabilistic Processes

Figures 4 and 5 show the result of the comparison of theoretical and empirical specific entropies for the two previously introduced limit families of discrete probability distributions, MD_UR and UBD_MR. Three theoretical indices, $\eta_1(\boldsymbol{p})$, $\eta_2(\boldsymbol{p})$, and $\eta_\infty(\boldsymbol{p})$, are placed in relationship each with its empirical counterpart, respectively $\overline{\eta_1}(\boldsymbol{f})$, $\overline{\eta_2}(\boldsymbol{f})$, and $\overline{\eta_\infty}(\boldsymbol{f})$, calculated from relative frequencies. Each family includes 100 synthetic processes, generated by progressively deforming a fair die distribution ($q = 6$, $p_1 = .. = p_6 = 1/6$) into a degenerate one ($p_1 = 1$), with a distortion percentage increasing from 0% to 100% along its entropic path. For each theoretical distribution, $r = 20,000$ realizations (each $L = 254$ samples long) are produced, and $N = 250$ event samples are drawn in a $d = 5$ dimensional space Ω ($n = 6^5 = 7776$), resulting in a sparse data regime (data density $\delta = N/n = 3.21\%$). As distortion grows, both true and empirical entropies decrease, but empirical estimates systematically underestimate true values, especially near uniformity. Saturation sets in beyond the threshold $\eta(\boldsymbol{f})_{\text{Lim}} = \log_n N \approx 0.616$, highlighted in the figures by a vertical dashed line, where the empirical entropy is restricted by the limited number of observed data. Despite construction differences, both families exhibit similar transitions from uniformity to concentration, highlighting common limitations of entropy estimation under finite sampling. Thus, these models provide a controlled framework for studying the interaction between distribution structure, sparsity, and estimation reliability, providing insights for high-dimensional inference.
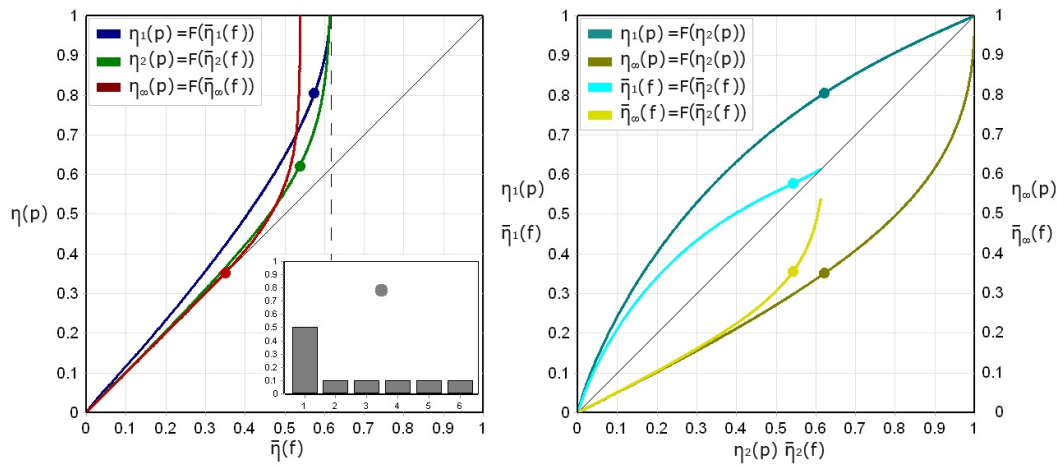

Figure 4. Apparent vs. actual specific entropies for the MD_UR family of probability distributions.
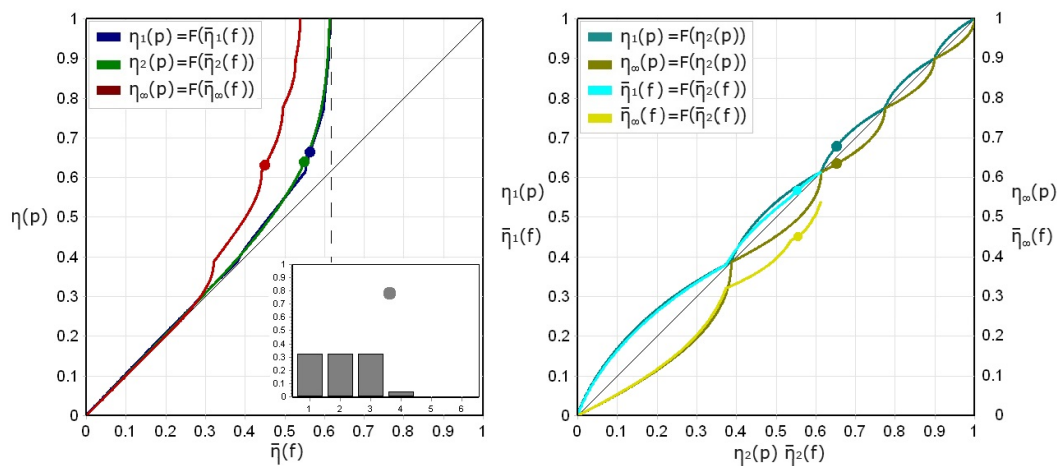

Figure 5. Apparent vs. actual specific entropies for the UBD_MR family of probability distributions.

3.4. Determination of the Relationship Between Actual and Apparent Rényi Mean

To enable a meaningful comparison between *actual* Rényi mean (based on the true distribution $\boldsymbol{p}$) and *apparent* Rényi mean (computed from empirical frequencies $\boldsymbol{f}$), it is convenient to apply a min-max normalization to both quantities using their respective minimal and maximal attainable values. These transformations give two comparable quantities, both ranging in $[0, 1]$ interval:

$$0 \leq \frac{{}^R\overline{\boldsymbol{p}}_\alpha - \frac{1}{n}}{1 - \frac{1}{n}} \leq 1 \qquad 0 \leq \frac{{}^R\overline{\boldsymbol{f}}_\alpha - \frac{1}{N}}{1 - \frac{1}{N}} \leq 1 \tag{28}$$

The second normalization assumes that when the true entropy is maximal—that is, when $\eta_\alpha(\boldsymbol{p}) = 1$—the apparent Rényi mean ${}^R\overline{\boldsymbol{f}}_\alpha$ reaches its minimum value of exactly $\frac{1}{N}$. However, *empirical evidence*—as shown in Figures 4, 5—suggests otherwise: the minimum value of the apparent Rényi mean ${}^R\overline{\boldsymbol{f}}_\alpha$ under maximum entropy conditions can deviate from $\frac{1}{N}$ and this deviation is maximal for $\alpha \to \infty$.

To account for this discrepancy, we introduce the term $\overline{m}_\alpha(N, q, d)$, which represents the empirical lower bound of the α-Rényi mean. It is computed as the average value of the Rényi mean over multiple realizations drawn from a perfectly uniform process, governed by a perfectly uniform distribution (U)—i.e., one with maximal specific entropy—under the same sampling conditions (N, q, d).

This empirically refined mean minimum, specific to each triplet (N, q, d), is defined as:

$$\overline{m}_\alpha(N, q, d) \triangleq \left\langle {}^R\overline{\boldsymbol{f}}_\alpha \right\rangle_{r(U)} \tag{29}$$

Then, as in general $\overline{m}_\alpha \leq {}^R\overline{\boldsymbol{f}}_\alpha \leq 1$, the relationship between the empirical Rényi mean ${}^R\overline{\boldsymbol{f}}_\alpha$ and the true Rényi mean ${}^R\overline{\boldsymbol{p}}_\alpha$ can be described as:

$$\frac{\widehat{{}^R\overline{\boldsymbol{p}}_\alpha} - \frac{1}{n}}{1 - \frac{1}{n}} = \frac{\left\langle {}^R\overline{\boldsymbol{f}}_\alpha \right\rangle_r - \overline{m}_\alpha}{1 - \overline{m}_\alpha} \tag{30}$$

Solving for $\widehat{{}^R\overline{\boldsymbol{p}}_\alpha}$, we obtain:

$$\widehat{{}^R\overline{\boldsymbol{p}}_\alpha} = \left(\frac{1 - \frac{1}{n}}{1 - \overline{m}_\alpha} \right) \cdot \left\langle {}^R\overline{\boldsymbol{f}}_\alpha \right\rangle_r + \left(\frac{1}{n} - \frac{1 - \frac{1}{n}}{1 - \overline{m}_\alpha} \cdot \overline{m}_\alpha \right) \tag{31}$$

This expression reveals that the first-order approximation of any Rényi mean can be obtained by applying a linear scaling and an offset to the empirical Rényi mean. Explicitly, this is a deterministic affine transformation of the form:

$$\widehat{{}^R\overline{\boldsymbol{p}}_\alpha} = A \cdot \left\langle {}^R\overline{\boldsymbol{f}}_\alpha \right\rangle_r + B \tag{32}$$

where:

- $A = \dfrac{1 - \frac{1}{n}}{1 - \overline{m}_\alpha}$ is the scaling factor;
- $B = \dfrac{1}{n} - A \cdot \overline{m}_\alpha$ is the translation offset;

3.5. Toward Accurate Estimation of Rényi Entropies via Affine Transformations

The affine structure of (32) suggests a possible interpretation of the estimator: it maps the apparent internal state of the distribution (observed empirically from relative frequencies) into its actual internal state (in the theoretical context of probabiity distributions). The Rényi entropy itself $(\eta_\alpha(\boldsymbol{p}) = -\log_n({}^R\overline{\boldsymbol{p}}_\alpha))$ can thus be interpreted as an observable readout of this internal state. Based on equation (31), the estimated specific Rényi entropy is:

$$
\widehat{\eta_\alpha}(\boldsymbol{p}) = -\log_n\left[\widehat{{}^R\overline{\boldsymbol{p}}_\alpha}\right] = -\log_n\left[\frac{1-\frac{1}{n}}{1-\overline{m}_\alpha}\left(\left\langle {}^R\overline{\boldsymbol{f}}_\alpha\right\rangle_{\boldsymbol{r}} - \overline{m}_\alpha\right) + \frac{1}{n}\right] =
$$
$$
= -\log_n\left[\frac{1-\frac{1}{n}}{1-\overline{m}_\alpha}\left(n^{-\widehat{\eta_\alpha}(f)} - \overline{m}_\alpha\right) + \frac{1}{n}\right]
$$

(33)

This expression holds for all $\alpha \in \mathbb{R}^+$, including limiting cases such as Shannon entropy ($\alpha \to 1$) and min-entropy ($\alpha \to \infty$). This approach yields a non-iterative, scalable, and theoretically grounded estimator. Its derivation is based on:

- Min-max normalization of actual and empirical Rényi means, to ensure comparability;
- A linear transformation between the two domains;
- The logarithmic relationship between Rényi entropy and Rényi mean.

By integrating these elements, the estimator provides a principled way to recover the entropy of the underlying process from observed frequencies, based on the following properties:

1. Change of paradigm: data do not need *correction*, but they have to be *rescaled* from an empirical to a theoretical context.
2. Consistency: as $N \to \infty$, we have $\overline{m}_\alpha \to 1/n$ and $\boldsymbol{f} \to \boldsymbol{p}$, ensuring that $\widehat{\eta_\alpha}(\boldsymbol{p})$ converges to the true entropy $\eta_\alpha(\boldsymbol{p})$.
3. Bias and variance control: the estimator compensates for the known bias in empirical frequencies by incorporating a *prior* empirical mean $\overline{m}_\alpha$, resulting in a lower mean squared error compared to classical estimators.
4. Computational efficiency: the formula requires only a few arithmetic operations and, optionally, a numerically precomputed lookup table for $\overline{m}_\alpha(q, d, N)$, making it suitable for large-scale inference.

The left column of Figure 6 consolidates the results previously shown in the left panels of Figures 4 and 5, reorganizing the same data into separate plots by entropy type: specific Shannon entropy η_1 at the top, specific collision entropy η_2 in the middle, and specific min-entropy η_∞ at the bottom. For each entropy type, the results obtained from the `MD_UR` and `UBD_MR` distribution families define a region that encloses all points achievable by applying the same procedures to any probability distribution under identical sampling conditions ($N = L - d + 1$) and sample space structure ($n = q^d$). The right column of Figure 6 illustrates the effect of applying formula (33) to the data, which induces a geometric transformation that diagonally stretches the region representing the relationship between the entropy of theoretical and empirical distributions. In the specific case of collision entropy, this transformation causes the curves associated with all distribution families—from `MD_UR` to `UBD_MR`—to nearly coincide with the diagonal line (see central panel in Figure 6). As a result, this transformation alone completes the estimation procedure for specific collision entropy. In contrast, the estimation of other Rényi entropies, such as Shannon and min-entropy, requires an additional adjustment step to reach comparable accuracy.
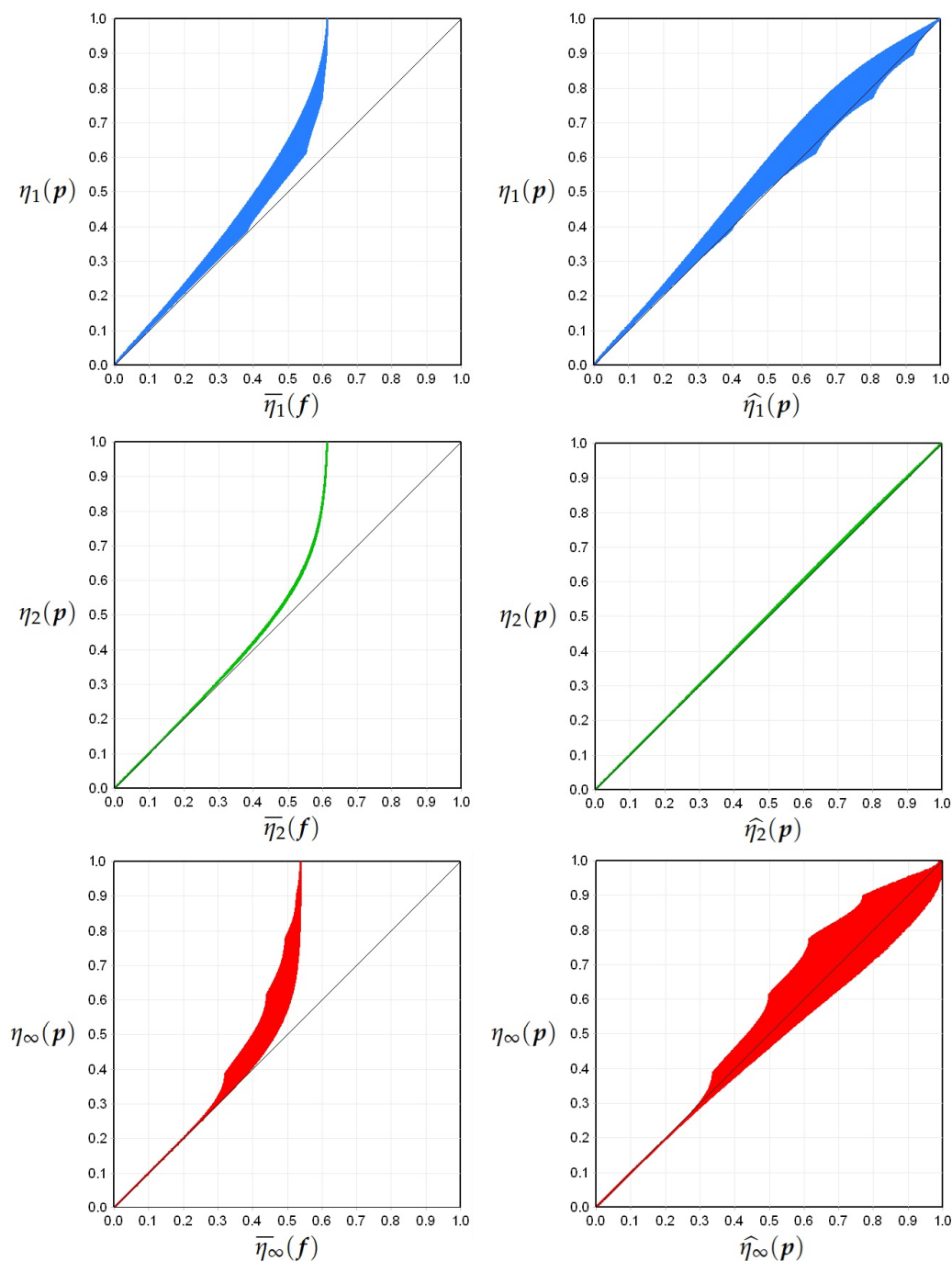
Figure 6. Left column: Relationship between apparent specific entropies $\overline{\eta_\alpha}(f)$ and true specific entropies $\eta_\alpha(p)$ for processes generated by distributions bounded by the `MD_UR` and `UBD_MR` families. Results from previous figures (4 and 5) are reorganized here by entropy type: η_1 in blue, η_2 in green, and η_∞ in red. **Right column:** Effect of the transformation described by Equation (33).

As a result, the specific collision entropy emerges as the most coherent index. Moreover, its optimal estimation can be performed with a single transformation. For this reason, it will serve as the horizontal reference axis in subsequent diagrams, facilitating comparative assessments of the estimators of other entropies. Using this approach, the data from Figure 6 can be recomposed by plotting the specific collision entropy η_2 on the X-axis and mapping η_1 on the left Y-axis and η_∞ on the right Y-axis of the diagram, respectively. The outcome of this new composition is shown in Figure 7.
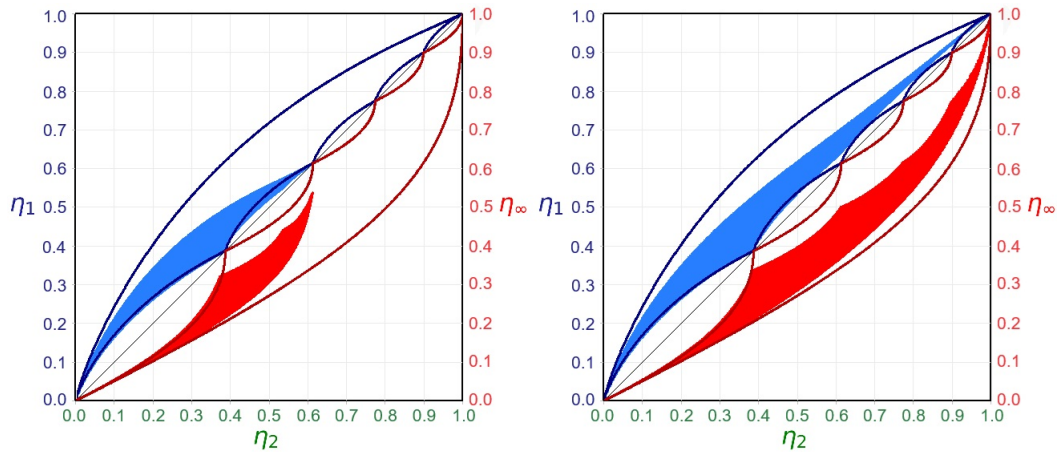
Figure 7. Left: representation of the regions associated to the relative frequency distributions generated using realizations with 254 samples and whose data are inserted in a 5-dimensional sample space (the same parameters of the previous figures): in blue $(\overline{\eta_2}(f), \overline{\eta_1}(f))$ and in red $(\overline{\eta_2}(f), \overline{\eta_\infty}(f))$. The observed contraction relative to the theoretical bounds with respect to the theoretical bounds arises because of the limited amount of data. **Right:** effect of the affine transformation on the reduced regions: the stretched regions $(\widehat{\eta_2}(p), \widehat{\eta_1}(p))$ and $(\widehat{\eta_2}(p), \widehat{\eta_\infty}(p))$ touch the maximum entropy vertex in $(1, 1)$, but yet cannot completely overlap the original theoretical region.

3.6. A Second Transformation for More Accurate Estimation of Entropies: Orthogonal Stretch

While the diagonal stretch is sufficient for accurately estimating the specific collision entropy, it proves inadequate for other Rényi entropies. As illustrated in Figure 7, this affine transformation alone fails to fully map the region derived from empirical data onto the corresponding region defined by theoretical probability distributions. To overcome this limitation, a second transformation—orthogonal to the direction associated with the collision mean—is required to complete the geometric alignment. A detailed analysis of this additional step, which represents a further advancement toward accurate Rényi entropy estimation, lies beyond the scope of the present work, which primarily serves as an introduction to this novel approach.

3.7. A Practical Example of Applying the First Order Affine Transformation

In this example, we demonstrate how the entropic state of a *DPD* can be recovered from a set of short realizations sampled from the process defined by the original *DPD*. The recovery is carried out by applying the previously described first-order affine transformation to the generalized means of empirical data distributions. Its procedure consists of the following steps:

1. *Choice of the stochastic process*: In this case, without loss of generality, and following the foundational approach adopted by early pioneers of probability theory (e.g., Cardano, Galileo, Pascal, Fermat, Huygens, Bernoulli, de Moivre, Newton), we consider a memoryless stochastic process generated by repeated rolls of a six-sided die subject to a specific probability distribution. The key advantage of using memoryless processes is that their entropic state remains invariant under changes in the dimensionality of the sample space used for statistical analysis.

2. *Choice of the theoretical DPD*: We select a distribution whose entropic characteristics lie far from the regions typically covered by empirical distributions derived from small samples, thereby making the recovery task more challenging, for instance:

$$T = \{0.36, \; 0.2, \; 0.11, \; 0.11, \; 0.11, \; 0.11\}$$

3. *Choice of the composition of the entropic space*: We consider a three dimensional entropic space. The entropies are: specific Shannon entropy (η_1), specific collision entropy (η_2), and specific min-entropy (η_∞). A generic three-dimensional point $A(\eta_1, \eta_2, \eta_\infty)$ is projected onto two two-dimensional points: $A(\eta_2, \eta_1)$, located above the diagonal in the collision entropy-Shannon

entropy plane (blue zone), and $A(\eta_2, \eta_\infty)$, located beneath the diagonal in the collision entropy-min-entropy plane (red zone). The value of η_2 is shared between the two projections.

4. *Visualization of theoretical, empirical, and stretched contexts*: To clearly distinguish the regions concerning the states of theoretical probability distributions, of their derived empirical distributions, and of the states resulting after the application of the first-order affine transformation, the diagram includes the curves generated by the elaborations over `MD_UR` and `UBD_MR` families, along with the curves relative to the corresponding realizations and their stretched versions.

5. *Determination of the entropic state points of the theoretical distribution T*: These points describe the entropic state of the initial distribution and represent the target of the estimation process:

$$T(\eta_2(\boldsymbol{p}), \eta_1(\boldsymbol{p})) = (0.850,\ 0.927)$$

$$T(\eta_2(\boldsymbol{p}), \eta_\infty(\boldsymbol{p})) = (0.850,\ 0.570)$$

6. *Generation of realizations*: A set of $r = 1{,}000$ realizations is generated from the process by applying a random number generator to the previously defined probability distribution T. A large number of realizations helps reduce the deviation of the final estimation from the target value. Each realization consists of a sequence of $L = 254$ samples.

7. *Choice of the the parameters of the sample space* Ω: To significantly reduce the density of data, we select a sample space $\Omega(q, d)$ with the same alphabet of the process $q = 6$ and dimension $d = 5$. The total number of elementary events is $n = 7776$.

8. *Embedding data into the sample space to derive relative frequencies*: Consequently, the number of elementary events that occur in $\Omega(q, d)$ is $N = L - d + 1 = 254 - 5 + 1 = 250$. This leads to a very sparse data regime, with a density $\delta = N/n = 250/7776 = 3.21\%$.

9. *Calculation of their center of gravity S*: For each realization, we compute the Rényi means and their logarithmic mapping in entropy plane; the results are then averaged:

$$S(\overline{\eta_2}(\boldsymbol{f}), \overline{\eta_1}(\boldsymbol{f})) = (0.603,\ 0.601)$$

$$S(\overline{\eta_2}(\boldsymbol{f}), \overline{\eta_\infty}(\boldsymbol{f})) = (0.603,\ 0.494)$$

10. *Calculation of their center of gravity E*: The affine transformation is applied to the Rényi means of the empirical data and the entropic state points of the translated Rényi means are derived; the results are then averaged:

$$E(\widehat{\eta_2}(\boldsymbol{p}), \widehat{\eta_1}(\boldsymbol{p})) = (0.845,\ 0.880)$$

$$E(\widehat{\eta_2}(\boldsymbol{p}), \widehat{\eta_\infty}(\boldsymbol{p})) = (0.845,\ 0.615)$$

11. *Evaluation of the distance between E and T*: The estimation method is satisfying when point E results coincident or very near to point T

Figure 8 illustrates the effect of the previous steps.
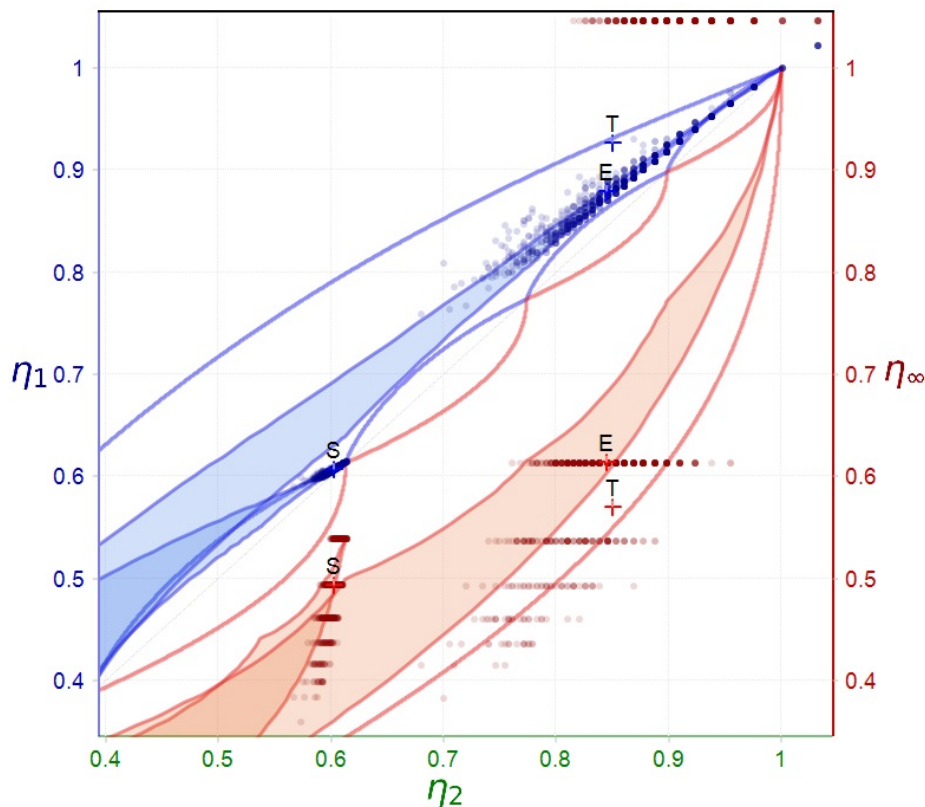
Figure 8. Final diagram of the example described in § 3.7. Specifically: (a) the light blue and light red regions represent the stretched domains obtained by applying the affine transformation to the compressed regions corresponding to empirical relative frequency distributions; (b) point *T*: the entropic state of the original probability distribution of the process; (c) point *S*: the average entropic state computed from 1,000 empirical relative frequency distributions; (d) point *E*: the center of gravity of the points resulting from the affine transformation applied to the Rényi means of the data.

4. Discussion

4.1. Considerations on the Example Presented in § 3.7

The example in § 3.7 illustrates a typical scenario where the entropic state of a theoretical distribution *T* significantly diverges from the empirical state *S*, derived from relative frequencies. Directly using values from *S* to estimate *T* would thus lead to substantial errors. Applying the affine transformation (31) to the Rényi generalized means computed from data yields entropy estimates defining a new state *E*, much closer to *T*. The collision entropy of *E* closely matches *T*, whereas the Shannon and min-entropy estimates show larger discrepancies, highlighting the need for further adjustments. Moreover, empirical points appearing clustered become more dispersed after transformation, indicating an increase in variance. Given that the value of m_α in this example is computed from a great number fo realizations, applying it to a smaller set (e.g., 1,000 realizations) might produce state points exceeding 1.0, particularly for min-entropy.

4.2. Considerations Concerning Algorithms for Generating `MD_UR` and `UBD_MR` Families of DPDs

To explore entropy trends in controlled distributions, we developed two algorithms: one for Mono Dominant with Uniform Residues (`MD_UR`) and another for Uniform Block Dominant with Mono Residue (`UBD_MR`), both derived from a uniform baseline by applying increasing distortion percentages. Algorithm 1 (`MD_UR`) is straightforward, with linear complexity $O(q)$, involving a single pass over events. Algorithm 2 (`UBD_MR`), however, redistributes probability from the tail to the head using nested loops and conditional logic, resulting in quadratic complexity $O(q^2)$. Thus, steering distributions toward Algorithm 1's path is computationally simpler than compressing them via Algorithm 2,

reflecting intrinsic asymmetries in entropy landscapes that could impact the modeling of natural probabilistic processes.

4.3. Data Contextualization vs. Data Correction in Entropy Estimation Methods

Another insight is a relativistic reinterpretation of relative frequencies in entropy estimation. Instead of viewing these frequencies as biased or erroneous, we suggest considering them as geometrically constrained observations. Limited dataset size naturally compresses empirical frequencies within the entropy state space. The affine transformation of Rényi generalized means thus acts not as traditional correction but as geometric re-contextualization, stretching empirical observations to align them with the theoretical entropy structure. This interpretation parallels general relativity's insight that mass shapes spacetime: data mass similarly shapes observable entropy space. Small datasets produce compressed entropy states, which expand toward theoretical configurations as data mass increases. Figure 7 clearly visualizes this phenomenon, showing initial empirical compression and subsequent geometric expansion via affine transformation and logarithmic mapping.

5. Conclusion

In a broad sense, this work aims to contribute to answer to Hilbert's Sixth Problem by addressing the longstanding challenge of rigorously bridging axiomatic foundations and empirical methodologies. It proposes a novel entropy estimator, grounded in a structured affine transformation of empirical frequency distributions, effectively translating formal definitions of Rényi entropy into practical, statistically robust and computationally efficient estimation methods. The paper introduces three core innovations: (1) the use of specific entropies η_α, enabling clear geometric interpretations; (2) the definition of two structured probability distribution families—Mono Dominant with Uniform Residues (`MD_UR`) and Uniform Block Dominant with Mono Residue (`UBD_MR`); and (3) a general estimation framework applicable across varying datasets, data types, and dimensions, capable of accurately estimating Shannon, collision, and min-entropy. Among these, collision entropy stands out because of its simplicity of estimation via affine stretching. However, relying solely on collision entropy can obscure critical nuances. Thus, the proposed approach naturally encourages the complementary use of multiple entropy indices (collision, Shannon, and min-entropy), enriching the characterization of uncertainty. Furthermore, this framework offers rigorous epistemological insights by precisely quantifying how empirical data contribute to our understanding of underlying realities.

6. Future Work

The findings of this study suggest several future research directions:

- *Improved transformation models:* Develop secondary orthogonal transformations to refine entropy estimation, especially min-entropy, through adaptive corrections.
- *Analytical characterization and simplification:* Investigating analytical approximations or tight bounds for empirical Rényi means m_α, reducing computational complexity.
- *Variance control and confidence estimation:* Enhancing robustness by exploring variance control methods, including smoothing kernels, shrinkage techniques, or deriving confidence intervals analytically.
- *Extension to real-world data:* Validation of the estimator on real datasets, particularly in cybersecurity, neuroscience, and computational biology, to test practical effectiveness.
- *Generalization to other information functionals:* Extending the affine estimation framework to broader information-theoretic measures, expanding its theoretical and practical scope.
- *Integration into statistical and machine learning workflows:* Exploring applications of entropy estimation within machine learning as feature transformations, loss functions, or regularizers to innovate data-driven modeling.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

X_q	Set of q ordered symbols (alphabet)
DSP	Stationary, infinite length, discrete-state stochastic process whose samples belong to X_q
r	Physical realization (data sequence of finite length) taken from a *DSP*
L	Number of samples constituting r
$\boldsymbol{r} = \{r_i\}$	Set of physical realizations
$\langle expr(r_i)\rangle_{\boldsymbol{r}}$	Arithmetic mean of $expr(r_i)$ calculated over $\boldsymbol{r}$
$\Omega(q,d)$	Sample space resulting from the Cartesian product d times of X_q
d	Dimension of Ω
$n = q^d$	Cardinality of Ω
$N = L - d + 1$	Number of occurrences of elementary events of Ω observed during the evolution of r
$\delta = N/n$	Data density in the sample space Ω
DPD	Discrete probability distribution
`MD_UR`	Mono Dominant (with Uniform Residues) family of *DPD*s
`UBD_MR`	Uniform block dominant (with Mono Residue) family of *DPD*s
RFD	Relative frequency distribution
$\boldsymbol{p}(\Omega)_{DSP}$	*DPD* obtained from a *DSP* whose infinite d-grams are considered coordinates of occurred elementary events of Ω
$\boldsymbol{f}(\Omega)_r$	*RFD* obtained from a realization r whose finite d-grams are considered coordinates of occurred elementary events of Ω
α	Order of a Rényi mean or of a Rényi entropy
$^R\overline{\boldsymbol{p}}_\alpha$	α-Rényi mean of a *DPD*
$^R\overline{\boldsymbol{f}}_\alpha$	α-Rényi mean of a *RFD*
$\widehat{^R\overline{\boldsymbol{p}}_\alpha}$	Estimated α-Rényi mean of a *DPD*
$H_\alpha(\boldsymbol{p})$	α-Rényi entropy of a *DPD*
$H_\alpha(\boldsymbol{f})$	α-Rényi entropy of a *RFD*
$\eta_\alpha(\boldsymbol{p})$	Specific α-Rényi entropy of a *DPD*
$\eta_\alpha(\boldsymbol{f})$	Specific α-Rényi entropy of an *RFD*
$\widehat{\eta}_\alpha(\boldsymbol{p})$	Estimated specific Rényi α-entropy of a *DPD*
$\eta'_\alpha(DSP)$	Specific Rényi α-entropy rate of a *DSP*
$\eta'_\alpha(r)$	Specific Rényi α-entropy rate of r
$\widehat{\eta'_\alpha}(DSP)$	Estimated specific Rényi α-entropy rate of a *DSP*
DS	Dynamical system
$\Theta(d,\epsilon)$	Partition of the state space into d-dimensional cells of linear size ϵ
$p_i(\Theta(d,\epsilon))$	Probability that a trajectory of the *DS* crosses the i-th cell
$D_\alpha(\Theta(d,\epsilon))$	Rényi generalized dimension of order α
$D'_\alpha(DS)$	Rényi Generalized Dimension Rate of order α of the *DS*
$I_\alpha(X;Y)$	α-Rényi mutual information
$\iota_\alpha(X;Y)$	Specific α-Rényi mutual information

References

1. Cardano, G. *Liber de Ludo Aleae (The Book on the Game of DiceGames of Chance)*; 1564. Published posthumously in 1663.
2. Pascal, B.; de Fermat, P. *Correspondence on the Problem of Points*; Original correspondence, Toulouse and Paris, 1654. Reprinted and translated in various historical anthologies on the foundations of probability theory.
3. Kolmogorov, A. *Foundations of the theory of probability*, 1950 ed.; Chelsea Publishing Co.: New York, 1933.

4. Shannon, C. A mathematical theory of communication. _The Bell System Technical Journal_ **1948**, _27_, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

5. Rényi, A. On measures of entropy and information. In Proceedings of the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, Berkeley, CA, USA, 1961; pp. 547–561.

6. Arikan, E. An inequality on guessing and its application to sequential decoding. In Proceedings of the Proceedings of 1995 IEEE International Symposium on Information Theory, 1995, pp. 322–. https://doi.org/10.1109/ISIT.1995.550309.

7. Csiszár, I. Generalized cutoff rates and Rényi's information measures. _IEEE Transactions on information theory_ **1995**, _41_, 26–34.

8. Beck, C. Generalised information and entropy measures in physics. _Contemporary Physics_ **2009**, _50_, 495–510.

9. Fuentes, J.; Gonçalves, J. Rényi Entropy in Statistical Mechanics. _Entropy_ **2022**, _24_. https://doi.org/10.3390/e24081080.

10. Cachin, C. Smooth entropy and Rényi entropy. In Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques. Springer, 1997, pp. 193–208.

11. Boztas, S. On Rényi entropies and their applications to guessing attacks in cryptography. _IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences_ **2014**, _97_, 2542–2548.

12. Badhe, S.S.; Shirbahadurkar, S.D.; Gulhane, S.R. Renyi entropy and deep learning-based approach for accent classification. _Multimedia Tools and Applications_ **2022**, pp. 1–33.

13. Sepúlveda-Fontaine, S.A.; Amigó, J.M. Applications of Entropy in Data Analysis and Machine Learning: A Review. _Entropy_ **2024**, _26_. https://doi.org/10.3390/e26121126.

14. Pál, D.; Póczos, B.; Szepesvári, C. Estimation of Rényi entropy and mutual information based on generalized nearest-neighbor graphs. _Advances in neural information processing systems_ **2010**, _23_.

15. Acharya, J.; Orlitsky, A.; Suresh, A.; Tyagi, H. The Complexity of Estimating Rényi Entropy. In Proceedings of the The twenty-sixth annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015. SIAM, SIAM, 2015, pp. 1855–1869. https://doi.org/10.1137/1.9781611973730.124.

16. Skorski, M. Practical Estimation of Renyi Entropy, 2020, [arXiv:cs.DS/2002.09264].

17. Miller, G.A. Note on the bias of information estimates. _Information Theory in Psychology_ **1955**, _II-B_, 95–100.

18. Vasicek, O.A. A Test for Normality Based on Sample Entropy. _Journal of the Royal Statistical Society: Series B (Methodological)_ **1976**, _38_, 54–59.

19. Basharin, G.P. On a statistical estimate for the entropy of a sequence of independent random variables. _Theory Probab. Appl._ **1959**, _4_, 333–336.

20. Good, I. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. _Ann. Math. Statist._ **1963**, _34_, 911–934.

21. Harris, B. The statistical estimation of entropy in the non-parametric case. _Mathematics Research Center – Technical Summary Report_ **1975**.

22. Grassberger, P.; Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. _Phys. Rev. A_ **1983**, _28_, 2591–2593. https://doi.org/10.1103/PhysRevA.28.2591.

23. Akaike, H. Prediction and entropy. In _Selected Papers of Hirotugu Akaike_; Springer, 1985; pp. 387–410.

24. Kozachenko, L.F.; Leonenko, N.N. Sample Estimate of the Entropy of a Random Vector. _Problemy Peredachi Informatsii_ **1987**, _23:2_, 95—101.

25. Grassberger, P. Finite sample corrections to entropy and dimension estimates. _Physics Letters A_ **1988**, _128_, 369–373. https://doi.org/10.1016/0375-9601(88)90193-4.

26. Joe, H. Estimation of Entropy and Other Functionals of a Multivariate Density. _Annals of the Institute of Statistical Mathematics_ **1989**, _41_, 683–697. https://doi.org/10.1007/BF00057735.

27. Hall, P.; Morton, S. On the estimation of entropy. _Annals of the Institute of Statistical Mathematics_ **1993**, _45_, 69–88.

28. Wolpert, D.; Wolf, D. Estimating functions of probability distributions from a finite set of samples. _Phys. Rev. E_ **1995**, _52_, 6841–6854. https://doi.org/10.1103/PhysRevE.52.6841.

29. Pöschel, T.; Ebeling, W.; Rosé, H. Guessing probability distributions from small samples. _Journal of Statistical Physics_ **1995**, _80_, 1443–1452. https://doi.org/10.1007/BF02179880.

30. Schürmann, T.; Grassberger, P. Entropy estimation of symbol sequences. _Chaos: An Interdisciplinary Journal of Nonlinear Science_ **1996**, _6_, 414–427.

31. Paluš, M. Coarse-grained entropy rates for characterization of complex time series. _Physica D: Nonlinear Phenomena_ **1996**, _93_, 64–77.

32. Panzeri, S.; Treves, A. Analytical estimates of limited sampling biases in different information measures. *Network: Computation in Neural Systems* **1996**, *7*, 87–107. https://doi.org/10.1088/0954-898X_7_1_006.

33. Beirlant, J.; Dudewicz, E.; Györfi, L.; Denes, I. Nonparametric entropy estimation. An overview. 1997, Vol. 6, pp. 17–39.

34. Schmitt, A.; Herzel, H. Estimating the Entropy of DNA Sequences. *Journal of theoretical biology* **1997**, *188*, 369–77. https://doi.org/10.1006/jtbi.1997.0493.

35. Strong, S.P.; Koberle, R.; de Ruyter van Steveninck, R.R.; Bialek, W. Entropy and Information in Neural Spike Trains. *Phys. Rev. Lett.* **1998**, *80*, 197–200. https://doi.org/10.1103/PhysRevLett.80.197.

36. Porta, A.; Baselli, G.; Liberati, D.; Montano, N.; Cogliati, C.; Gnecchi-Ruscone, T.; Malliani, A.; Cerutti, S. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow. *Biological cybernetics* **1998**, *78*, 71–8.

37. Holste, D.; Große, I.; Herzel, H. Bayes' estimators of generalized entropies. *Journal of Physics A: Mathematical and General* **1998**, *31*, 2551–2566. https://doi.org/10.1088/0305-4470/31/11/007.

38. Chen, S.; Goodman, J. An empirical study of smoothing techniques for language modeling. *Computer Speech & Language* **1999**, *13*, 359–394.

39. de Wit, T.D. When do finite sample effects significantly affect entropy estimates? *The European Physical Journal B - Condensed Matter and Complex Systems* **1999**, *11*, 513–516. https://doi.org/10.1007/s100510050963.

40. Rached, Z.; Alajaji, F.; Campbell, L.L. Rényi's entropy rate for discrete Markov sources. In Proceedings of the Proceedings of the CISS, 1999, Vol. 99, pp. 17–19.

41. Antos, A.; Kontoyiannis, I. Convergence properties of functional estimates for discrete distributions. *Random Structures & Algorithms* **2001**, *19*, 163–193. https://doi.org/10.1002/rsa.10019.

42. Nemenman, I.; Shafee, F.; Bialek, W. Entropy and inference, revisited. *In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, Advances in Neural Information Processing Systems* **2002**, *14*, 471–478.

43. Chao, A.; Shen, T.J. Non parametric estimation of Shannon's index of diversity when there are unseen species. *Environ. Ecol. Stat.* **2003**, *10*, 429–443. https://doi.org/10.1023/A:1026096204727.

44. Grassberger, P. Entropy estimates from insufficient samples. *arXiv2003, physics/0307138v2* **2003**.

45. Paninski, L. Estimation of entropy and mutual information. *Neural Computation* **2003**, *15*, 1191—-1253. https://doi.org/10.1162/089976603321780272.

46. Wyner, A.; D., F. On the lower limits of entropy estimation. *IEEE Transactions on Information Theory - TIT* **2003**.

47. Amigó, J.M.; Szczepański, J.; Wajnryb, E.; Sanchez-Vives, M.V. Estimating the Entropy Rate of Spike Trains via Lempel-Ziv Complexity. *Neural Computation* **2004**, *16*, 717–736. https://doi.org/10.1162/089976604322860677.

48. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138. https://doi.org/10.1103/PhysRevE.69.066138.

49. Paninski, L. Estimating entropy on m bins given fewer than m samples. *IEEE Transactions on Information Theory* **2004**, *50*, 2200–2203. https://doi.org/10.1109/TIT.2004.833360.

50. Pöschel, T.; Ebeling, W.; Frömmel, C.; Ramirez, R. Correction algorithm for finite sample statistics. *The European physical journal. E, Soft matter* **2004**, *12*, 531–41.

51. Schürmann, T. Bias analysis in entropy estimation. *Journal of Physics A: Mathematical and General* **2004**, *37*, L295. https://doi.org/10.1088/0305-4470/37/27/L02.

52. Ciuperca, G.; Girardin, V. On the estimation of the entropy rate of finite Markov chains. In Proceedings of the Proceedings of the international symposium on applied stochastic models and data analysis, 2005, pp. 1109–1117.

53. Szczepański, J.; Wajnryb, E.; Amigó, J.M. Variance Estimators for the Lempel-Ziv Entropy Rate Estimator. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2006**, *16*, 043102. https://doi.org/10.1063/1.2347102.

54. Hlaváčková-Schindler, K.; Paluš, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* **2007**, *441*, 1–46.

55. Kybic, J. High-Dimensional Entropy Estimation for Finite Accuracy Data: R-NN Entropy Estimator. In Proceedings of the Information Processing in Medical Imaging; Karssemeijer, N.; Lelieveldt, B., Eds., Berlin, Heidelberg, 2007; pp. 569–580.

56. Vu, V.; Yu, B.; Kass, R. Coverage-adjusted entropy estimation. *Statistics in Medicine* **2007**, *26*, 4039–4060. https://doi.org/10.1002/sim.2942.

57. Bonachela, J.; Hinrichsen, H.; Muñoz, M. Entropy estimates of small data sets. *Journal of Physics A: Mathematical and Theoretical* **2008**, *41*, 9. https://doi.org/10.1088/1751-8113/41/20/202001.

58.   Grassberger, P.   Entropy Estimates from Insufficient Samplings, 2008, [arXiv:physics.data-an/physics/0307138].

59.   Hausser, J.; Strimmer, K. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *J. Mach. Learn. Res.* **2009**, *10*, 1469–1484.

60.   Lesne, A.; Blanc, J.; Pezard, L. Entropy estimation of very short symbolic sequences. *Physical Review E* **2009**, *79*, 046208. https://doi.org/10.1103/PhysRevE.79.046208.

61.   Golshani, L.; Pasha, E.; Yari, G. Some properties of Rényi entropy and Rényi entropy rate. *Information Sciences* **2009**, *179*, 2426–2433. https://doi.org/10.1016/j.ins.2009.03.002.

62.   Xu, D.; Erdogmuns, D., Renyi's Entropy, Divergence and Their Nonparametric Estimators. In *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer New York: New York, NY, 2010; pp. 47–102. https://doi.org/10.1007/978-1-4419-1570-2_2.

63.   Källberg, D.; Leonenko, N.; Seleznjev, O. Statistical Inference for Rényi Entropy Functionals. *arXiv preprint arXiv:1103.4977* **2011**.

64.   Nemenman, I. Coincidences and Estimation of Entropies of Random Variables with Large Cardinalities. *Entropy* **2011**, *13*, 2013–2023. https://doi.org/10.3390/e13122013.

65.   Vinck, M.; Battaglia, F.; Balakirsky, V.; Vinck, A.; Pennartz, C. Estimation of the entropy based on its polynomial representation. *Phys. Rev. E* **2012**, *85*, 051139. https://doi.org/10.1103/PhysRevE.85.051139.

66.   Boucheron, S.; Lugosi, G.; Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence*; Vol. 14, *Oxford Series in Probability and Statistics*, Oxford University Press: Oxford, UK, 2013. Focuses on concentration inequalities with applications in probability, statistics, and learning theory.

67.   Zhang, Z.; Grabchak, M. Bias Adjustment for a Nonparametric Entropy Estimator. *Entropy* **2013**, *15*, 1999–2011. https://doi.org/10.3390/e15061999.

68.   Valiant, G.; Valiant, P. Estimating the Unseen: Improved Estimators for Entropy and Other Properties. *J. ACM* **2017**, *64*. https://doi.org/10.1145/3125643.

69.   Li, L.; Titov, I.; Sporleder, C. Improved estimation of entropy for evaluation of word sense induction. *Computational Linguistics* **2014**, *40*, 671–685. https://doi.org/10.1162/COLI_a_00196.

70.   Archer, E.; Park, I.; Pillow, J. Bayesian entropy estimation for countable discrete distributions. *The Journal of Machine Learning Research* **2014**, *15*, 2833–2868.

71.   Schürmann, T. A Note on Entropy Estimation. *Neural Comput.* **2015**, *27*, 2097–2106. https://doi.org/10.1162/NECO_a_00775.

72.   Kamath, S.; Verdú, S. Estimation of entropy rate and Rényi entropy rate for Markov chains. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), 2016, pp. 685–689. https://doi.org/10.1109/ISIT.2016.7541386.

73.   Skorski, M. Improved estimation of collision entropy in high and low-entropy regimes and applications to anomaly detection. Cryptology ePrint Archive, Paper 2016/1035, 2016.

74.   Acharya, J.; Orlitsky, A.; Suresh, A.; Tyagi, H. Estimating Rényi entropy of discrete distributions. *IEEE Transactions on Information Theory* **2017**, *63*, 38–56. https://doi.org/10.1109/TIT.2016.2620435.

75.   de Oliveira, H.; Ospina, R. A Note on the Shannon Entropy of Short Sequences **2018**. https://doi.org/10.14209/sbrt.2018.8.

76.   Berrett, T.; Samworth, R.; Yuan, M. Efficient multivariate entropy estimation via *k*-nearest neighbour distances. *The Annals of Statistics* **2019**, *47*, 288–318. https://doi.org/10.1214/18-AOS1688.

77.   Verdú, S. Empirical estimation of information measures: a literature guide. *Entropy* **2019**, *21*, 720. https://doi.org/10.3390/e21080720.

78.   Goldfeld, Z.; Greenewald, K.; Niles-Weed, J.; Polyanskiy, Y. Convergence of smoothed empirical measures with applications to entropy estimation. *IEEE Transactions on Information Theory* **2020**, *66*, 4368–4391. https://doi.org/10.1109/TIT.2020.2975480.

79.   Kim, Y.; Guyot, C.; Kim, Y. On the efficient estimation of Min-entropy. *IEEE Transactions on Information Forensics and Security* **2021**, *16*, 3013—-3025. https://doi.org/10.1109/TIFS.2021.3070424.

80.   Contreras Rodríguez, L.; Madarro-Capó, E.; Legón-Pérez, C.; Rojas, O.; Sosa-Gómez, G. Selecting an effective entropy estimator for short sequences of bits and bytes with maximum entropy. *Entropy* **2021**, *23*. https://doi.org/10.3390/e23050561.

81.   Ribeiro, M.; Henriques, T.; Castro, L.; Souto, A.; Antunes, L.; Costa-Santos, C.; Teixeira, A. The entropy universe. *Entropy* **2021**, *23*. https://doi.org/10.3390/e23020222.

82.   Subramanian, S.; Hsieh, M.H. Quantum algorithm for estimating alpha-Rényi entropies of quantum states. *Physical Review A* **2021**, *104*, 022428. https://doi.org/10.1103/PhysRevA.104.022428.

83.   Grassberger, P. On Generalized Schürmann Entropy Estimators. *Entropy* **2022**, *24*. https://doi.org/10.3390/e24050680.

84.   Skorski, M. Towards More Efficient Rényi Entropy Estimation. *Entropy* **2023**, *25*, 185. https://doi.org/10.3390/e25020185.

85.   Gecchele, A. Collision Entropy Estimation in a One-Line Formula. *Cryptology ePrint Archive, Paper 2023/927* **2023**.

86.   Al-Labadi, L.; Chu, Z.; Xu, Y. Advancements in Rényi entropy and divergence estimation for model assessment. *Computational Statistics* **2024**, pp. 1–18.

87.   Álvarez Chaves, M.; Gupta, H.V.; Ehret, U.; Guthke, A. On the Accurate Estimation of Information-Theoretic Quantities from Multi-Dimensional Sample Data. *Entropy* **2024**, *26*, 387. https://doi.org/10.3390/e26050387.

88.   Pinchas, A.; Ben-Gal, I.; Painsky, A. A Comparative Analysis of Discrete Entropy Estimators for Large-Alphabet Problems. *Entropy* **2024**, *26*. https://doi.org/10.3390/e26050369.

89.   De Gregorio, J.; Sánchez, D.; Toral, R. Entropy Estimators for Markovian Sequences: A Comparative Analysis. *Entropy* **2024**, *26*. https://doi.org/10.3390/e26010079.

90.   Onicescu, O. Energie informationelle. *Comptes Rendues l'Academie des Sciences* **1966**, *263*.

91.   Pincus, S. Approximate entropy (ApEn) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **1995**, *5*, 110–117. https://doi.org/10.1063/1.166092.

92.   Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* **2000**, *278*, H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039.

93.   Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters* **2002**, *89*, 068102. https://doi.org/10.1103/PhysRevLett.89.068102.

94.   Bandt, C.; Pompe, B. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102.

95.   Manis, G.; Aktaruzzaman, M.; Sassi, R. Bubble entropy: An entropy almost free of parameters. *IEEE Transactions on Biomedical Engineering* **2017**, *64*, 2711–2718. https://doi.org/10.1109/TBME.2017.2664105.

96.   Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **1988**, *52*, 479–487.

97.   Golshani, L.; Pasha, E. Rényi entropy rate for Gaussian processes. *Information Sciences* **2010**, *180*, 1486–1491. https://doi.org/10.1016/j.ins.2009.12.012.

98.   Teixeira, A.; Matos, A.; Antunes, L. Conditional Rényi Entropies. *IEEE Transactions on Information Theory* **2012**, *58*, 4273–4277. https://doi.org/10.1109/TIT.2012.2192713.

99.   Fehr, S.; Berens, S. On the Conditional Rényi Entropy. *IEEE Transactions on Information Theory* **2014**, *60*, 6801–6810. https://doi.org/10.1109/TIT.2014.2357799.

100. Grassberger, P. Generalized dimensions of strange attractors. *Physics Letters A* **1983**, *97*, 227–230. https://doi.org/https://doi.org/10.1016/0375-9601(83)90753-3.

101. Van Erven, T.; Harremos, P. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory* **2014**, *60*, 3797–3820.

102. Kullback, S.; Leibler, R. On Information and Sufficiency. *The Annals of Mathematical Statistics* **1951**, *22*, 79–86.

103. Hölder, O. Ueber einen Mittelwertsatz. *Mathematische Annalen* **1889**, *34*, 511–518.