Soft Computing Fusion with Applications

www.scfa.reapress.com

Soft. Comput. Fusion. Appl. Vol. 1, No. 2 (2024) 87–98.

Paper Type: Original Article

On the Rényi Entropy Functional, Tsallis Distributions and Lévy Stable Distributions with Entropic Applications to Machine Learning

Ismail A Mageed*

PhD, AIMMA, IEEE, IAENG, School of Computer Science, AI, and Electronics, Faculty of Engineering and Digital Technologies, University of Bradford, United Kingdom; drismail664@gmail.com.

Citation:

Received: 4 July 2024 *Revised: 16 September 2024* *Accepted: 10 November 2024*	Mageed, I. A. (2024). On the Rényi entropy functional, Tsallis distributions and lévy stable distributions with entropic applications to machine learning. *Soft computing fusion with applications, 1*(2), 87-98.

Abstract

The Rényi entropy function was optimized for the novel findings of this investigation under various conditions. Their research's findings suggest that generalized t-distributions cover the complete spectrum of Lévy stable distributions. Additionally, an exposition was undertaken to prove that the Lévy distribution generalizes the Tsallisian distribution rather than the reverse. The current study is a strong generalization of an existing research work in the literature. Notable entropic applications to machine learning are addressed. Concluding remarks are given, combined with some strong open problems and future research pathways.

Keywords: Shannon entropy functional, Rényi entropy functional, Tsallis non-extensive entropy functional, Lévy stable distributions, Machine learning.

1 | Introduction

The most justifiable version of information entropy is the Rényi entropy with a free Rényi non-extensivity parameter q, and the Tsallis entropy can be thought of as a linear approximation [1] to the Rényi entropy when $q \approx 1$. When $q \to 1$, the Boltzmann-Shannon entropy function replaces both other entropy functions. When the Rényi entropy functional is subjected to the Maximum Information Entropy (MEP) principle, the result is the microcanonical (homogenous) distribution for an isolated system. The Boltzmann entropy functional replaces the Rényi entropy functional in this situation, which supports the universality of Boltzmann's principle of statistical mechanics, regardless of the value of the Rényi parameter q.

The need for non-extensive statistics based on Tsallis' information entropy is critical, given its quick development. The one-parameter family of Rényi entropies (or just Rényi entropy) seems to be the most rational one [2]. The well-known Boltzmann-Shannon entropy functional replaces the Rényi entropy functional when the Rényi parameter q is equal to unity. The non-extensive Tsallis' entropy functional is produced by linearizing the extended Rényi entropy functional in the vicinity [2] of a point $q \approx 1$.

Corresponding Author: drismail664@gmail.com

When the principle of Maximum MEP is applied to the Rényi entropy functional of an isolated system. At this phase, the Rényian functional reduces to the Boltzmannian functional, thus enforcing the Boltzmann principle from which all thermodynamic properties of extensive and non-extensive Hamiltonian systems can be deduced.

The measure of information in this example of a system's incomplete statistical descriptor with the help of probabilistic distribution is called the information entropy functional, or just entropy $p = \{p_i\}, 0 \leq p_i \leq 1, i = 1, \ldots, n$.

Boltzmann-Shannon representation of the entropy functional is the most well-known to read as in *Eq. (1)*:

$$H_B = -K_B \sum_{i=1}^{n} p_i \ln p_i. \quad \textbf{(1)}$$

The entropy H_B correlates with the thermodynamic entropy functional in the situation given, where the distribution p_i is the system's macroscopic equilibrium state, and the subscripts i denote dynamic microstates in the Gibbs phase space.

This entropy function was justified by [2]–[4] based on a system of axioms presented in a theorem form. Their axioms were analyzed in [2]–[4] where it was shown that a uniquely determined Boltzmann-Shannon entropic form is provided by a quite artificial axiom related to a form of conditional entropy functional (that is, the entropy functional of a subsystem of a system being in a prescribed state). On another strong note, [3] examined several papers on this topic and discovered that the Shore and Johnson [6] axiom system, which results in the Rényi entropy function [4], follows *Eq. (2)*:

$$H_R^{(q)}(p) = \frac{K_B}{1-q} \ln \left(\sum_{i=1}^{n} p_i^q \right), \sum_{i=1}^{n} p_i = 1, \quad \textbf{(2)}$$

The Rényi entropy functional, which is a mathematical measure, was used to quantify the amount of information or disorder in a system. It mentions that q (c.f., *Eq. (2)*) must be positive and not less than zero. The properties and characteristics of Rényi entropy functional are further explored in related literature [4]–[6]. Among its basic properties, we may mention positivity ($H_R^{(q)} \geq 0$), concavity for $q \leq 1$, and in addition $\lim_{q\to 1} H_R^q = H_B$.

In the case of $|1 - \sum_i p_i^q| \ll 1$ (which, in view of the normalization of the distribution $\{p_i\}$, corresponds to the condition $|1 - q| \ll 1$), one can restrict oneself to the linear term of the logarithm in the expression for $H_R^{(q)}(p)$ over this difference and $H_R^{(q)}(p)$ changes to the defined form of *Eq. (3)*.

$$H_T^{(q)}(p) = -\frac{K_B}{1-q}(1 - \left(\sum_{i=1}^{n} p_i^q \right)). \quad \textbf{(3)}$$

Such a linearization of the Rényi entropy functional was proposed by [11], [12]; at this time, the Tsallisian case had come into existence by [12].

The entropy functional stops being exhaustive due to logarithm linearization. To examine a range of non-extensive systems, Tsallisian followers have extensively utilized this quality [7]–[15]. In doing so, the constraint $|1 - q| \ll 1$ mentioned above is typically ignored.

According to MEP, when describing a system statistically, its distribution function should accurately represent the average quantities observed in the system. If these quantities are not known, the distribution function should be as indeterminate as possible. This approach has been widely used in constructing equilibrium statistical thermodynamics for isolated or weakly interacting thermodynamic systems.

Gibbs ensembles [1]–[5] were widely used as a statistical approach to accurately represent average quantities observed in a system, as they are commonly used in constructing equilibrium statistical thermodynamics for isolated or weakly interacting thermodynamic systems. The information entropy functional, commonly referred to as the Boltzmann-Shannon entropy functional, is traditionally used to quantify the disorder or uncertainty in a system.

This theorem introduces as it reads a new physical interpretation: as Rényi entropy functional is more general than both Shannon and Tsallis, since Rényi entropy functional reduces to Shannon case as the parameter $q \to 1$, and that Rényian entropy functional's linearization about a point $q \approx 1$ is the Tsallisian entropy functional. This leads to a newer ground in information theory; we could represent it by the following *Fig. 1*:

$$H_T^{(q)} \xrightarrow{H_R^{(q\to 1)}} H_B \xleftarrow{H_T^{(q\to 1)}} \xrightarrow{\textbf{(linearization of } H_R^{(q)} \textbf{ in the neighbourhood of a point } q\,\approx 1)}$$

Fig.1. A new cornerstone to information theory.

In other words, this could be read as employing the Rényi entropy function to research any concept would generate the special case of Shannon case as the parameter $q \to 1$, and would reduce to the Tsallis case if we carry out the linearization of the Rényi entropy function in the neighbourhood of a point $q \approx 1$.

Shannon entropy, sometimes referred to as Gibbs entropy in statistical physics, is a measure of disorder in a system. As an alternative to Gibbs entropy, Tsallis developed a non-extensive entropy [8], [10], indexed by q, which results in an infinite family of Tsallis non-extensive entropies. While Tsallis's non-extensive entropy produces type II generalized Lévy stable distributions with heavy tails that obey power laws, Gibbs's entropy produces exponential distributions. It is important to remember that Tsallis entropy is equivalent to Havrda-Charvat's structural q-entropy [9], but the non-extensive mechanics community frequently ignores this relationship. Additionally, rather than the other way around, Tsallis distributions are derived from Lévy distributions.

The current paper contributes to:

I. The current study is a strong generalization of an existing research work in literature.

II. The provision of entropic applications to machine learning.

III. The exposition of several open problems.

The flowchart of the provided work reads as follows:

I. Introduction and literature review.

II. Results.

III. Entropic application to machine learning.

IV. Conclusion, open problems, and future research pathways.

2 | Results

We'll talk about how four different generalized t-distribution types can be derived from the definitions of Rényi entropy functional. The Lagrangian equation of the calculus of variations, a mathematical technique for optimizing functions subject to restrictions, is used to reach these conclusions. According to the author's analysis, these discoveries are brand-new, and they are presented.

Result 1

Rényi entropy functional can be optimized under *Eq. (4)*:

$$\sum_{i=1}^{n} p_i = 1. \tag{4}$$

Under the defined constraint of *Eq. (5)*:

$$E\left(1 + a_1|x|^b\right) = \text{constant}. \tag{5}$$

We have the Lagrangian function $\mathcal{L}$ to be as in *Eq. (6)*:

$$\mathcal{L} = \left[\frac{K_B}{(1-q)} \ln\left(\sum_{i=1}^{n} p_i^q\right) - \alpha_1\left(\sum_{i=1}^{n} p_i - 1\right) - \beta_1\left(E\left(1 + a_1|x|^b\right) - \text{constant}\right)\right]. \tag{6}$$

$\frac{\partial \mathcal{L}}{\partial p_i} = 0$ implies.

Hence by *Eq. (7)*,

$$\left(\frac{\frac{qK_B}{1-q}}{\left(\sum_{i=1}^{n} p_i^q\right)}\left(\sum_{i=1}^{n} p_i^{q-1}\right) - \alpha_1\left(\sum_{i=1}^{n} 1\right) - \beta_1 \sum_{i=1}^{n}\left(1 + a_1|x|^b\right)\right) = 0. \tag{7}$$

i.e., by *Eq. (8)*:

$$\frac{\frac{qK_B}{1-q} p_i^{q-1}}{\left(\sum_{i=1}^{n} p_i^q\right)} = \alpha_1 + \beta_1 + \beta_1 a_1 |x|^b = (\alpha_1 + \beta_1)\left(1 + \frac{\beta_1 a_1}{\alpha_1 + \beta_1}|x|^b\right). \tag{8}$$

Defining *Eq. (9)* to be:

$$\frac{\beta_1 a_1}{\alpha_1 + \beta_1} = a, \qquad C^{q-1} = \frac{(\alpha_1 + \beta_1)(1-q)}{qK_B}\left(\sum_{i=1}^{n} p_i^q\right), d = \frac{1}{1-q}. \tag{9}$$

Hence, clearly, it follows by *Eq. (10)* that

$$p_i = \frac{c}{(1 + a|x|^b)^d}, \tag{10}$$

where $-\infty < x < \infty, a, b, c, d > 0, bd > 0, 1 > q > 0.$

Result two

We have two constraints, by *Eq. (4)*.

Subject to that defined by *Eq. (11)*:

$$\sum_{i=1}^{n} |x|^b p_i^q(x) = \text{constant}. \tag{11}$$

We have the Lagrangian function $\mathcal{L}$ to be as in *Eq. (12)*:

$$\mathcal{L} = \left[\frac{K_B}{(1-q)} \ln\left(\sum_{i=1}^{n} p_i^q\right) - \alpha_2\left(\sum_{i=1}^{n} p_i - 1\right) + \beta_2\left(\sum_{i=1}^{n} |x|^b p_i^q(x) - \text{constant}\right)\right]. \tag{12}$$

$\frac{\partial \mathcal{L}}{\partial p_i} = 0$ implies *Eq. (13)*:

$$(\frac{\frac{qK_B}{1-q}}{\left(\sum_{i=1}^{n} p_i^q\right)}\left(\sum_{i=1}^{n} p_i^{q-1}\right) - \alpha_2(\sum_{i=1}^{n} 1) + q\left(\beta_2 \sum_{i=1}^{n} |x|^b p_i^{q-1}\right) = 0. \quad (13)$$

Hence, *Eq. (14)* follows:

$$(\frac{\frac{qK_B}{1-q}}{\left(\sum_{i=1}^{n} p_i^q\right)}(1 + \frac{\beta_2(1-q)}{qK_B}\left(\sum_{i=1}^{n} p_i^q\right)|x|^b)p_i^{q-1} = \alpha_2. \quad (14)$$

Thus, one gets *Eq. (15)*:

$$\text{Define a } = \frac{\beta_2(1-q)}{qK_B}\left(\sum_{i=1}^{n} p_i^q\right),\ C^{q-1} = \frac{\alpha_2(1-q)}{qK_B}\left(\sum_{i=1}^{n} p_i^q\right), d = \frac{1}{q-1}. \quad (15)$$

Hence, we have *Eq. (16)* to take the form:

$$p_i = \frac{c}{(1 + a|x|^b)^d}, \quad (16)$$

where $-\infty < x < \infty, a, b, c, d > 0, bd > 0, \frac{2b+1}{b+1} > q > 1$. Because these are partially related through *Eqs. (4)*, *(11)* and *(16)* generate Tsallis distribution [8] for $b = 2$.

Result three

Following our approach as above is subject to the two constraints, as defined by *Eq. (4)* and, after prescribing *Eq. (17)*:

$$\sum_{i=1}^{n} |x|^b p_i\,(x) = \text{constant}. \quad (17)$$

The reader can check that after a few algebraic steps, the solution is in the closed-form representation defined by *Eq. (18)*:

$$p_i = \frac{c}{(1 + a|x|^b)^d}. \quad (18)$$

Such that *Eq. (19)* holds:

$$C = \frac{\alpha_3(1-q)}{qK_B}\left(\sum_{i=1}^{n} p_i^q\right),\ C^{q-1} = \frac{\beta_3}{\alpha_3}, d = \frac{1}{1-q}. \quad (19)$$

where $-\infty < x < \infty, a, b, c, d > 0, bd > 0, 1 > q > \frac{1}{b+1}$, α_3, β_3 are the Lagrangian multipliers. *Eq. (17)* defines the variance.

Result four

Following our approach as above is subject to more complex constraints, as by *Eq. (4)*. Combined with the recalled *Eq. (11)* and the communicated *Eq. (17)*.

After a few algebraic steps, the reader can check that by optimizing Rényi's entropy functional subject to *Constraints (4), (11)* and *(17)*, the reader can check that after few algebraic steps, the solution is in the closed form representation, as prescribed by *Eq. (20)*:

$$p_i = c(\frac{1 + a|x|^b}{1 + a'|x|^b})^d. \quad (20)$$

Exposing the parametric representation of *Eq. (21)*:

$$C^{q-1} = \frac{\alpha_4(1-q)}{qK_B}\left(\sum_{i=1}^{n} p_i^q\right), \qquad a = \frac{\gamma_4}{\alpha_4}, a' = \frac{\beta_4(1-q)}{qK_B}\left(\sum_{i=1}^{n} p_i^q\right) d = \frac{1}{q-1}. \tag{21}$$

where $-\infty < x < \infty, a, b, c, d > 0, bd > 0, \frac{2b+1}{b+1} > q > 1, \alpha_4, \beta_4, \gamma_4$ are the Lagrangian multipliers. The variance corresponds to $b = 2$, *Eq. (17)* defines the variance. We can see that for $0, b = 2$, *Eq. (20)* reduces to Tsallisian distribution [8] as a special case.

By *Eq. (20)*, we have for small values of $|x|$, *Eq. (22)*, follows:

$$p_i = c\left(\frac{1+a|x|^b}{1+a'|x|^b}\right)^d \sim c(1+(a-a')|x|^b)^d. \tag{22}$$

Carrying out the same analysis, we have for large values of $|x|$, as read by *Eq. (23)*:

$$p_i \sim c((a-a')|x|^b)^d. \tag{23}$$

This is summarized in the more compact form, as defined by *Eq. (24)*:

$$p_i(x) \sim \begin{cases} c(1+(a-a')|x|^b)^d = p_1(x), & \text{for small } |x|, \\ c((a-a')|x|^b)^d = p_2(x), & \text{for large } |x|. \end{cases} \tag{24}$$

The Probability Density Functions (PDFs), given by *Eq.s (10)*, *(16)*, *(18),* and *(20),* refer to generalized t-distributions, which exhibit polynomial tails for small values of x and power law tails for large values of x. These distributions encompass the entire range of Lévy stable distributions, which are commonly used to model extreme events and heavy-tailed phenomena. Specifically, *Eq. (10)* represents the PDFs for small values of x, while *Eq. (20)* represents the pdf for large values of x. This implies *Eq. (25)*

$$p(x) \sim \begin{cases} c\left(1-ad|x|^b\right) = p_1(x), & \text{for small } |x|, \\ c\left(a|x|^b\right)^{-d} = p_2(x), & \text{for large } |x|, \end{cases} \tag{25}$$

where $bd > 1$.

The outcomes for *Eq.s (16)*, *(18)* and *(20)* are comparable. Since the Lévy distributions can be represented by the generalized t-distributions, as obtained by *Eq.s (10)*, *(16)*, *(18),* and *(20)*, the density functions that can be obtained from Rényi entropy functional are even broader. It is clear from the analysis done and reported in *Eq. (24)*, for *Eq. (20)*, that the case is still relevant

3 | Entropic Applications to Machine Learning

Feature selection [16] in machine learning involves identifying the most relevant and predictive features while excluding irrelevant or redundant ones. Information theory, specifically mutual information measuring the correlation between features and labels, is commonly used for this purpose. The exploration of using Rényi min-entropy for feature selection, showing that in practical experiments with real datasets, the Rényi-based algorithm tends to outperform the traditional Shannon-entropy-based approach in terms of performance, was extensively undertaken by [16].

In a more detailed perspective, the authors of [16] have provided a method for feature selection based on Rényi min-entropy, which is a measure of uncertainty in information theory. Additionally, it was shown that selecting the optimal set of features using min-entropy is computationally challenging (NP-hard) and proposes an iterative strategy to approximate the best feature subset efficiently [16]. The study [16] compared this approach with one based on Shannon entropy and demonstrated through experiments that the Rényi-based algorithm performs better in practice across different datasets.

In this context, a dataset with 32 classes and 8 features is divided into two types [16], F and F', as showcased by *Fig. 2* (c.f., [16]). Each type consists of 4 features. At a specific step in the process, the algorithm based on

Shannon entropy selects a set of features, $S_1^3 = \{f_1, f_3, f_4\}$, as one of the possible outcomes [17]. This selection is a result of the algorithm's approach to feature selection based on the information content provided by Shannon entropy.

In the context of feature selection in machine learning [16], the text discusses the comparison between Rényi min-entropy and Shannon entropy algorithms. The explains that while the Rényi algorithm [16] may require additional steps to achieve complete accuracy in classification, the Shannon algorithm can stop when the classification is already completely accurate based on residual entropy and Bayes error. The comparison highlights that the choice between the two algorithms depends on the desired level of accuracy in the classification task [16].

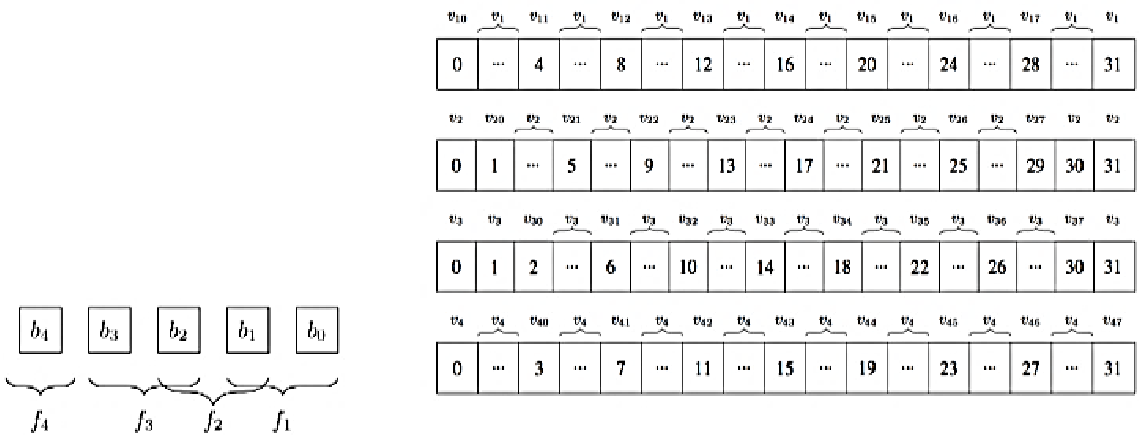


Fig. 2. Features F (left) and F' (right).

When dealing with Artificial Neural Networks (ANNs), the complexity increases due to numerous parameters that can impact the network's structure [18]. To address this, the authors utilized Bayesian optimization. This method combines extensive search with Gaussian processes to determine the best parameters for the ANN model, such as the number of hidden layers, neurons per layer [18], learning rate, batch size, and epochs. By incorporating the Spearmint Bayesian optimization codebase and optimizing these parameters, they aimed to create a model that generalizes well and maximizes accuracy through a 10-fold cross-validation process before testing on unseen data [18].

It is worth noting that, Iin [16] the study, a bootstrap procedure with 5 iterations was used to shuffle data and ensure the results were independent of the specific training, validation, and test set split. Each iteration involved a new experimental run with different training-test set splits. Features were selected based on the training set using Shannon and Rényi min-entropy methods, with each iteration adding one selected feature at a time for a total of 50 steps. The models were trained and tested on the test set after hyper-parameter tuning with 10-fold cross-validation, demonstrating the effectiveness of Rényi min-entropy in feature selection, particularly with the BASEHOCK dataset.

This explains that after conducting multiple iterations of feature selection using Rényi min-entropy and Shannon entropy methods, the average performance was computed and presented in *Figs. 3*, *4* and *5* (c.f., [16]). The results consistently showed that the feature selection approach based on Rényi min-entropy generally outperformed the Shannon entropy method, particularly when analyzing the BASEHOCK dataset.

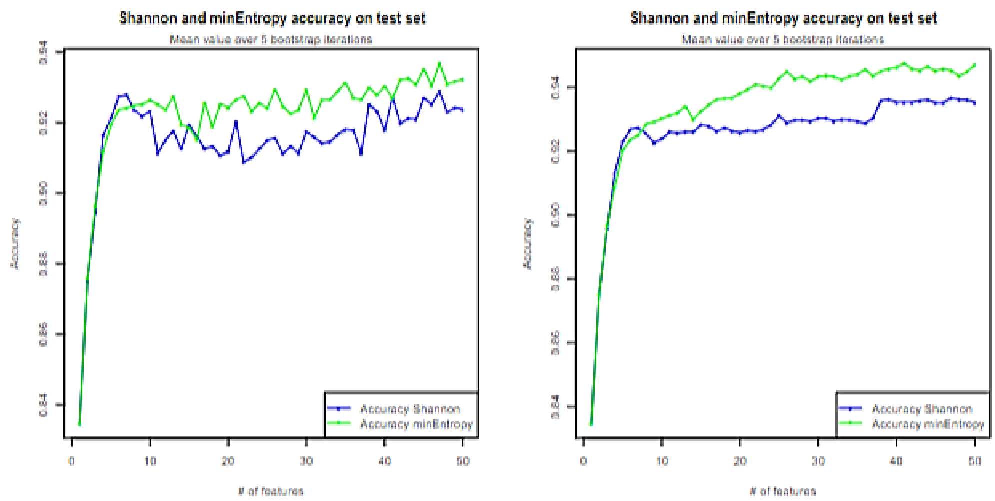


Fig. 3. Comparison of the accuracy of artificial neural networks and support vector machine classifiers on the BASEHOCK dataset using different feature selection methods, specifically Shannon entropy and Rényi min-entropy.

The results show that the feature selection method based on Rényi min-entropy generally yielded better performance compared to Shannon entropy, particularly with the BASEHOCK dataset, indicating the effectiveness of Rényi min-entropy in improving classification accuracy in machine learning tasks.

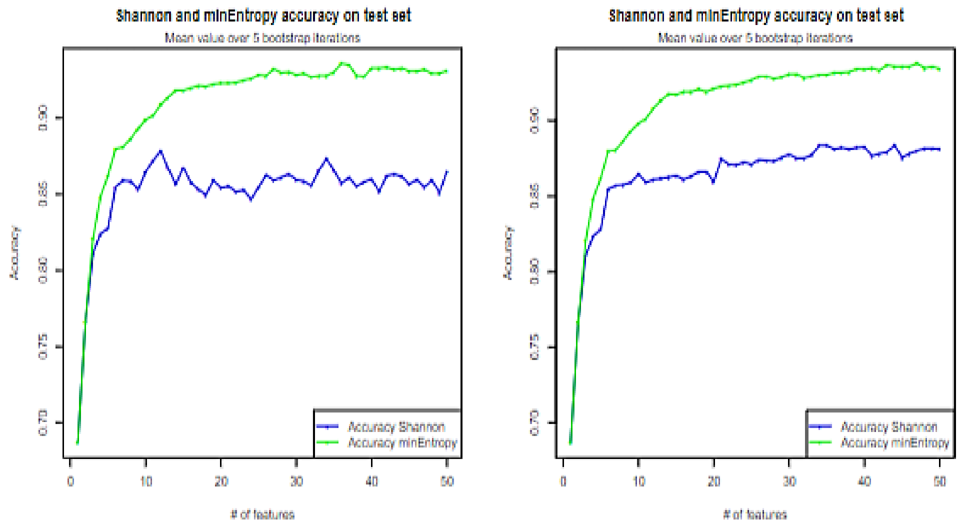


Fig. 4. The results show that the Rényi min-entropy feature selection method generally outperformed Shannon entropy, particularly evident with the BASEHOCK dataset.

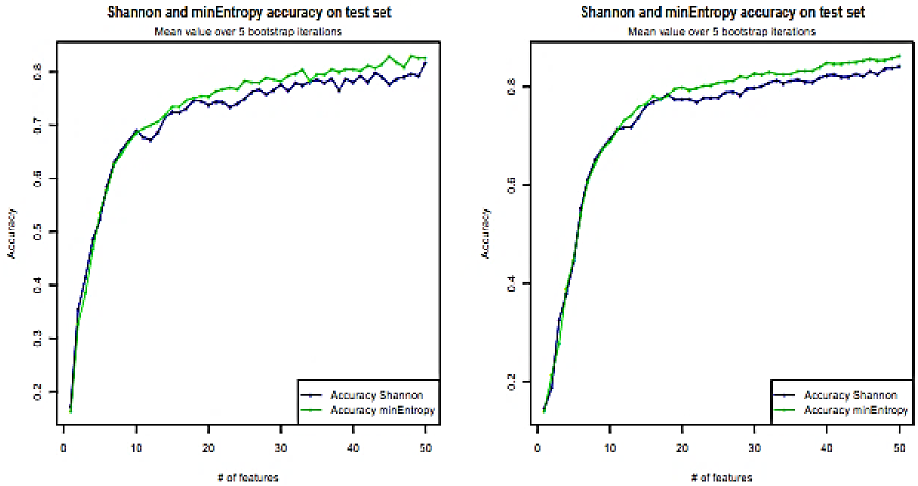


Fig. 5. The accuracy of artificial neural networks and support vector machine classifiers on the SEMEION dataset.

These classifiers are evaluated based on their performance in classifying data within the SEMEION dataset, with the accuracy metric used to measure their effectiveness in making correct predictions. The comparison of ANNs and Support Vector Machine (SVM) classifiers' accuracy provides insights into their respective capabilities in handling the dataset for classification tasks.

In deep reinforcement learning [19], a critical challenge is maintaining the agent's ability to explore effectively over the long term. To address this, a new method using Rényi entropy-based intrinsic rewards has been proposed to enhance exploration incentives and avoid the issue of vanishing rewards. This approach [19] improved exploration quality without the complexity of traditional methods, showing promising performance in simulations and offering potential applications in real-world scenarios like autonomous driving and smart manufacturing.

In the context of the provided findings [19], *Fig. 6* (c.f., [19]) illustrates how different objective functions behave when an agent learns from an environment with only three states. *Fig. 6* shows that the Shannon entropy tends to maintain higher values as state probabilities decrease, potentially limiting exploration, while the Rényi entropy aligns better with the agent's exploration needs by penalizing small probabilities more effectively. This comparison highlights how the Rényi entropy offers a more flexible approach to encourage exploration in the learning process compared to the Shannon entropy.

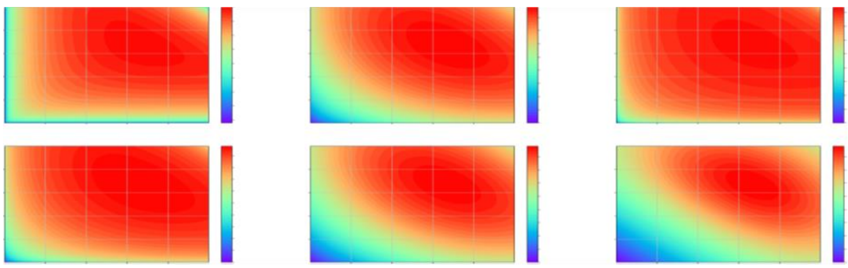
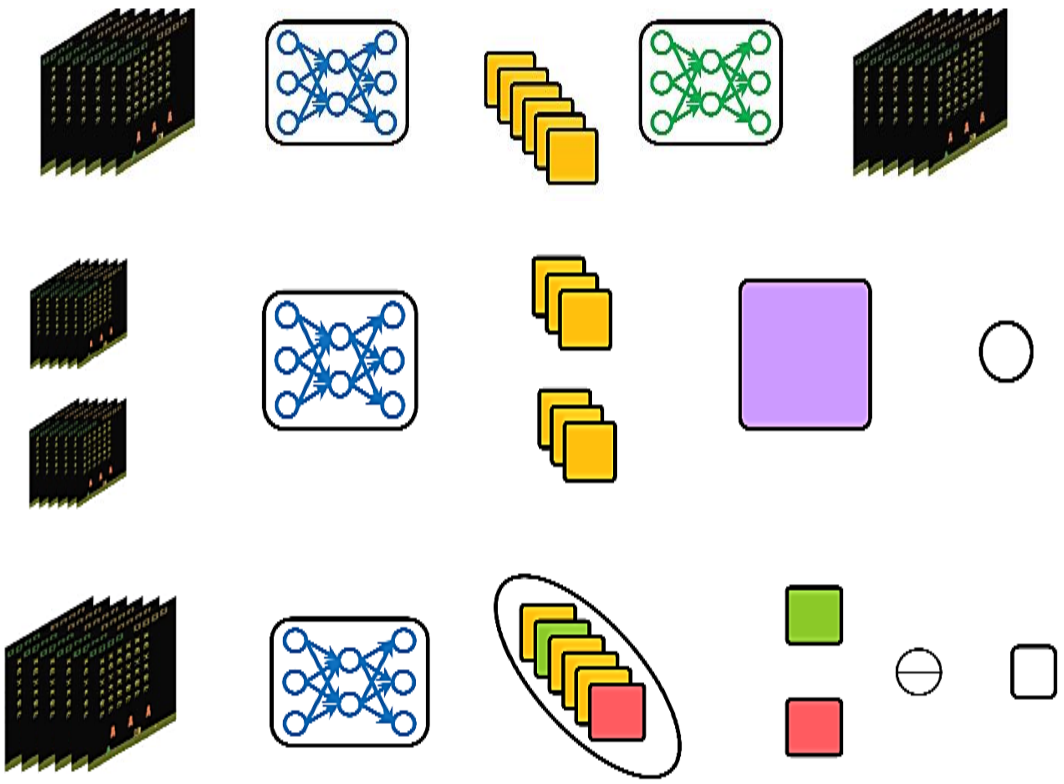


Fig. 6. Contours of different objective functions when |S| = 3.

Notably, in a standard Variational Autoencoder (VAE), there are two key components: a recognition model (encoder) denoted by $q_{\varphi}(z|s)$ and a generative model (decoder) denoted by $p\psi(s|z)$. The encoder processes input observations to encode them into latent variables while the decoder reconstructs the observations from these latent variables. The VAE is trained by minimizing a loss function that balances the reconstruction error and the Kullback–Liebler (KL) divergence between the encoder and decoder distributions, as illustrated by *Fig. 7* (c.f., [19]).



a.

b.

c.

Fig. 7. Overview architecture of RISE; a. Variational autoencoder model for embedding observations, b. K-value searching, c. Generation of intrinsic rewards, where k-NN is the k-nearest neighbour and _ denotes the Euclidean distance.

In the context of the conducted research in [20], the text refers to using a grid-based environment called Maze2D to showcase the effectiveness of Rényi state entropy-driven exploration. Maze2D is a simple example used to demonstrate how this exploration method works in navigating through a maze with the goal of finding the shortest path from the start point to the endpoint, emphasizing the efficiency of the approach in this scenario, as depicted by *Fig. 8* (c.f., [21]).

In the maze game scenario described, the agent can move in four directions- left, right, up, and down- aiming to find the shortest path from the start point to the endpoint. Additionally, the agent can teleport from one portal to another identical mark within the maze. The experimental setting involves using the Q-learning algorithm to navigate through mazes of varying sizes, updating the Q-table at each step to enhance training efficiency and achieve optimal exploration performance.

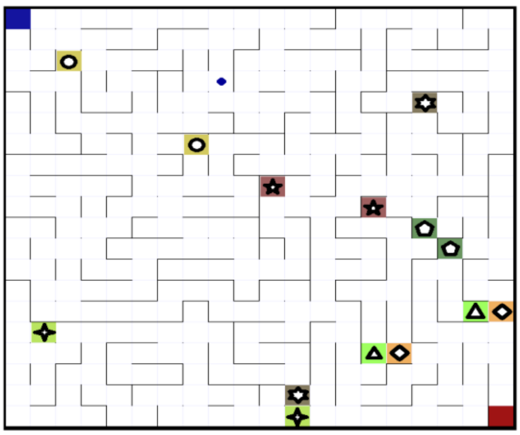


Fig. 8. A maze game with a grid size of 20 × 20.

As showcased by *Fig. 9* (c.f., [19]), Rényi State Entropy (RISE) outperformed other methods in terms of average return across six Atari games, with RISE achieving the highest performance consistently. Additionally, RISE demonstrated better training efficiency in terms of Frames Per Second (FPS) compared to other methods like RE3 and MaxRényi, showcasing its advantage in policy performance and learning efficiency.

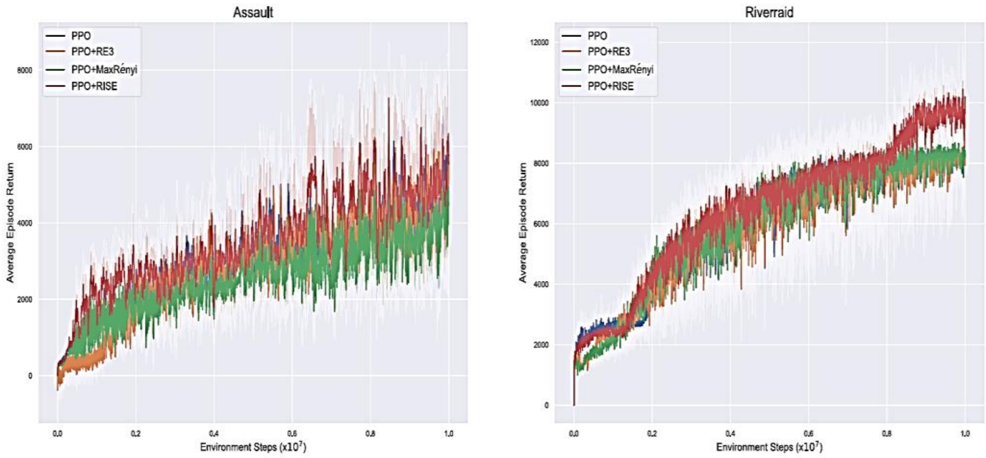


Fig. 9. Average episodes return versus the number of environment steps on Atari games.

4 | Conclusion, Open Problems, and Future Research Pathways

Using distinct sets of circumstances to optimize the Rényi entropy function, we revealed fresh results in this study. The outcomes produce generalized t-distributions for the whole family of Lévy stable distributions. It is discovered that the Lévy distribution generalizes the Tsallisian case, not the other way around. The work of [22], where Shannon and Havrda-Charvat entropy functional were utilized, is strongly generalized in this work.

The current paper flags several sophisticated open problems:

I. Is it possible to replace the undertaken mathematical mechanism by Ismail's entropy (IE) (c.f., [23]). If so, what will be the generated distributions?

II. It is expected that using IE would result in higher accuracies in feature selection in comparison to both Shannonian and Rényian cases. The mechanism could be quite sophisticated, but it is doable and would be highly promising as a new frontier to enhance machine learning.

Future research pathways include delving into unlocking the proposed open problems, as well as the exploration of novel information-theoretic applications to interdisciplinary fields of human knowledge.

References

[1] Mageed, I. A. (2023). The consistency axioms of the stable M/G/1 queue's z a, b non-extensive maximum entropy formalism with M/G/1 theory applications to 6G networks and multimedia applications. *2023 international conference on computer and applications (ICCA)* (pp. 1–6). IEEE. DOI: 10.1109/ICCA59364.2023.10401411

[2] Mageed, I. A., Zhang, Q., Kouvatsos, D. D., & Shah, N. (2022). M/G/1 queue with balking shannonian maximum entropy closed form expression with some potential queueing applications to energy. *IEEE global energy conference, GEC 2022* (pp. 105–110). IEEE. DOI: 10.1109/GEC55014.2022.9987144

[3] Mageed, I. A., & Zhang, Q. (2022). Inductive inferences of z-entropy formalism (ZEF) stable M/G/1 queue with heavy tails. *2022 27th international conference on automation and computing (ICAC)* (pp. 1–6). IEEE. DOI: 10.1109/ICAC55051.2022.9911090

[4] Mageed, I. A., & FRSS, I. (2023). *Where the mighty trio meet: information theory (IT), pathway model theory (PMT) and queueing theory (QT)* [presentation]. 39th annual uk performance engineering workshop.

[5] Mageed, I. A. (2023). The entropian threshold theorems for the steady state probabilities of the stable M/G/1 Queue with Heavy Tails with Applications of Probability Density Functions to 6G networks.

Electronic journal of computer science and information technology, *9*(1), 24–30. https://doi.org/10.52650/ejcsit.v9i1.138

[6] Begum, A., & Choudhury, G. (2023). Analysis of an M/(G1G2)/1 queue with bernoulli vacation and server breakdown. *International journal of applied and computational mathematics*, *9*(1), 9. DOI: 10.1007/s40819-022-01481-4

[7] Kloska, S., Pałczyński, K., Marciniak, T., Talaśka, T., Nitz, M., Wysocki, B. J., Davis, P., & Wysocki, T. A. (2021). Queueing theory model of Krebs cycle. *Bioinformatics*, *37*(18), 2912–2919. DOI: 10.1093/bioinformatics/btab177

[8] Mollaei, S., Darooneh, A. H., & Karimi, S. (2019). Multi-scale entropy analysis and Hurst exponent. *Physica a: statistical mechanics and its applications*, *528*, 121292. DOI: 10.1016/j.physa.2019.121292

[9] Kuaban, G. S., Soodan, B. S., Kumar, R., & Czekalski, P. (2022). A queueing-theoretic analysis of the performance of a cloud computing infrastructure: accounting for task reneging or dropping. *2022 international conference on electrical, computer, communications and mechatronics engineering (ICECCME)* (pp. 1–7). IEEE. DOI: 10.1109/ICECCME55909.2022.9988250

[10] Kłopotek, M. A., & Kłopotek, R. A. (2023). Towards continuous consistency axiom. *Applied intelligence*, *53*(5), 5635–5663. DOI: 10.1007/s10489-022-03710-1

[11] Guo, C., Fang, S., & He, Y. (2023). A generalized stochastic process: fractional G-Brownian motion. *Methodology and computing in applied probability*, *25*(1), 22. https://doi.org/10.1007/s11009-023-10010-9

[12] Scully, Z., & Harchol-Balter, M. (2021). The gittins policy in the M/G/1 queue. *2021 19th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks, wiopt 2021* (pp. 1–8). IEEE. DOI: 10.23919/WiOpt52861.2021.9589051

[13] Jessup II, M. A., & others. (2012). Using hybrid simulation/analytical queueing networks to capacitate usaf air mobility command passenger terminals. https://scholar.afit.edu/etd/1214

[14] Nielsen, F. (2020). An elementary introduction to information geometry. *Entropy*, *22*(10), 1100. https://doi.org/10.3390/e22101100

[15] Ouadfeul, S. A. (2024). *Fractal analysis-applications and updates*. BoD-books on demand.

[16] Palamidessi, C., & Romanelli, M. (2020). Feature selection in machine learning: R'enyi min-entropy vs Shannon entropy. *ArXiv preprint arxiv:2001.09654*. https://doi.org/10.48550/arXiv.2001.09654

[17] Palamidessi, C., & Romanelli, M. (2018). Feature selection with rényi min-entropy. *Artificial neural networks in pattern recognition: 8th IAPR TC3 workshop, ANNPR 2018, Siena, Italy, September 19-21, 2018, proceedings 8* (pp. 226–239). Springer, Cham. https://doi.org/10.1007/978-3-319-99978-4_18

[18] Will-Cole, A. R., Kusne, A. G., Tonner, P., Dong, C., Liang, X., Chen, H., & Sun, N. X. (2022). Application of Bayesian optimization and regression analysis to ferromagnetic materials development. *IEEE transactions on magnetics*, *58*(1), 1–8. DOI: 10.1109/TMAG.2021.3125250

[19] Yuan, M., Pun, M. O., & Wang, D. (2023). Rényi state entropy maximization for exploration acceleration in reinforcement learning. *IEEE transactions on artificial intelligence*, *4*(5), 1154–1164. DOI:10.1109/TAI.2022.3185180

[20] Borelli, R., Dovier, A., & Fogolari, F. (2022). Data structures and algorithms for k-th nearest neighbours conformational entropy estimation. *Biophysica*, *2*(4), 340–352. DOI: 10.3390/biophysica2040031

[21] Lim, H.-D., & Lee, D. (2022). *Regularized Q-learning*. ArXiv Preprint ArXiv:2202.05404. https://doi.org/10.48550/arXiv.2202.05404

[22] Rathie, P. N., & Da Silva, S. (2008). Shannon, Lévy, and Tsallis: A note. *Applied mathematical sciences*, *2*(28), 1359–1363.

[23] Mageed, I. A., & Zhang, Q. (2022). An information theoretic unified global theory for a stable M/G/1 queue with potential maximum entropy applications to energy works. *2022 global energy conference (GEC)* (pp. 300–305). IEEE. DOI: 10.1109/GEC55014.2022.9986719