IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 2, June 2024, pp. 2247~2253
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i2.pp2247-2253

Kernel density estimation of Tsalli's entropy with applications in adaptive system training

Leena Chawla, Vijay Kumar, Arti Saxena
Department of Applied Sciences, School of Engineering and Technology, Manav Rachna International Institute of Research and Studies, Faridabad, India

Article Info

Article history:

Received May 31, 2023
Revised Nov 25, 2023
Accepted Jan 16, 2024

Keywords:

Entropy estimator
Information theoretic measure
Machine learning
Non-parametric estimator
Tsalli's entropy

ABSTRACT

Information theoretic learning plays a very important role in adaption learning systems. Many non-parametric entropy estimators have been proposed by the researchers. This work explores kernel density estimation based on Tsallis entropy. Firstly, it has been proved that for linearly independent samples and for equal samples, Tsallis-estimator is consistent for the PDF and minimum respectively. Also, it is investigated that Tsallis-estimator is smooth for differentiable, symmetric, and unimodal kernel function. Further, important properties of Tsallis-estimator such as scaling and invariance for both single and joint entropy estimation have been proved. The objective of the work is to understand the mathematics behind the underlying concept.

Corresponding Author:

Vijay Kumar
Department of Applied Sciences, School of Engineering and Technology
Manav Rachna International Institute of Research and Studies
Faridabad, Haryana, India
Email: vijaykumar.fet@mriu.edu.in

1. INTRODUCTION

In statistical learning, model generates sample data and similar data for a larger population to learn from the set of variables and patterns. Generally, statistical models are deployed for three purposes such as: prediction, investigation of stochastic conjectures, and information extraction. These aspects of the statistical learning attract the researchers across various domains of studies. Data is growing exponentially and therefore there is need for robust probability density estimator to process the information. In learning, statistics plays a very vital role to find the optimality criterion of any adaptive system required for training. For this purpose, information-theoretic criteria measures are required to lower the uncertainty of the system. Many nonparametric algorithms have been proposed by many researchers for Renyi's entropy. These algorithms have applications in many problems of machine learning in classification for feature reduction. The pioneering work of Wiener has been used in adaptive systems training using second-order statistics but in many situations, second-order statistics is not enough to propose an optimality criterion. In adaption learning systems, the main idea is to establish a simple entropy estimators that is computationally cheap and satisfy the underlying properties of continuity and differentiable in terms of samples. In such systems, the main objective is to estimate the quantity while optimizing the parameters. The notion of information as an outcome of a random event was introduced by Gacs [1], which was extended by Shannon [2] to establish a theory, known as information theory that has applications across domains. Information theory is a separate discipline that gives birth to many disciplines [3]–[8]. Many generalized measures of Shannon entropy has been proposed and applied in various fields of study. Therefore, there is a requirement for new information theoretic criteria that have been used to develop adaptive algorithms for training. In this paper, a nonparametric estimator has been proposed for Tsalli's entropy and the results have been discussed in section 5. The proposed estimator has applications in the development of learning algorithms. The concepts proposed are describes mathematical point of view only.

2. LITERATURE REVIEW

Some researchers extended the work of Renyi and proposed various measures that have practical approaches [9]–[13]. A team of researchers applied Renyi's entropy to the problems of dimensionality reduction, feature extraction, and blind source separation [14]–[18]. In [14], [18] was the first to conceptualize blink source separation and subspace projection using Renyi's entropy. Principe *et al.* [15] was the first one to coin the term information theoretic learning into adaptive systems. Entropy estimation has appeared in many domains of science and technology from biology [19] and physics [20] to engineering [21], [22]. Beirlant *et al.* [23] discussed the mathematical viewpoint to estimate the entropy of continuous random variables. In this approach, a PDF drawn from the samples belongs to the known parametric family of PDFs to estimate the parameters using maximum likelihood methods. Dmitriev and Tarasenko [24] proposed the approximate estimate for Shannon entropy, whereas Joe [25] used a kernel-based PDF estimate to propose an approximate integral estimate for Shannon entropy for multivariate cases but the evaluation of this estimation was found to be complicated due to the increased of number of samples with the dimension of the data. Along the same lines, Ahmad and Lin [26] presented a kernel-based estimate for Shannon entropy and proved mean square consistency.

Some researchers proposed entropy estimates for Shannon entropy that depends on the type of problem in hand and customized them according to the requirement of the algorithm for computation [27]–[31]. They use spectral estimation-based PDF estimates except for [31] and applied entropy estimates in various problems of electrical engineering. Another approach for entropy estimation has been proposed in which the sample set is divided into two parts and estimates using density estimation and sample mean in the first part and second part respectively [32]–[34]. Also, Ivanov and Rozhkova [35] used a cross-validation approach to estimate Shannon entropy using the leave-one-out principle and kernel-based PDF estimator. The estimates integral, re-substitution, splitting data, and cross-validation are known as the plug-in estimates. Some estimates are based on sample spacing in which density estimate is constructed on the basis of sample differences [36]–[38]. The PDF estimate for the entropy can be made as re-substitution estimates but these m-spacing estimates are weakly consistent. In multivariate cases, the generalization of these estimates is non-trivial. In the case of general multivariate densities, the nearest neighbor (NN) entropy estimate can be used to test the consistency.

Kozachenko and Leonenko [39] defined the NN entropy estimate as the logarithm of the sample average of the normalized NN distances with an Euler constant. Various types of consistency under certain conditions was explained in [39]–[41]. Hassan *et al.* [42] proposed a non-parametric model for power system security risk assessment. The authors used Parzen window density estimation and obtained probability density functions for power systems. Bakouch *et al.* [43] estimated the probability density function using kernel density estimation. Aruga and Tanaka [44] proposed a learning measure using maximizing principle of Tsalli's entropy. Li *et al.* [45] proposed minimum error entropy (MEE) criteria to improve sparse system identification in non-gaussian noises. Abhishek *et al.* [46] used a variable Parzen window to determine density function by considering ambient dimension, flatness range, and neighborhood size. The proposed technique increases the classification accuracy in graphs. Xiong *et al.* [47] proposed a hybrid technique of entropy and Parzen window that have applications in image analysis and computer vision. To consider this, new information-theoretic estimator has been proposed using the Tsallis measure of entropy with the parzen windowing function.

3. METHOD

Information-theoretic measures play a very vital role to understand the uncertainty of the system. The idea of introducing these measures in the system is to establish entropy estimators that optimize feature parameters. In this paper, the Tsallis entropy measure has been used with the Parzen window function to introduce kernel estimators. Some propositions and properties have been proposed, established from the results of three theorems (5.1-5.3). The methodology of the paper is theoretical in nature which can be observed in the coming sections.

4. PRELIMINARIES

The terms like entropy, measures of entropy and the generalized entropies of Shannon entropy are not covered as these are general in nature and discussed by many researchers in their literature. For detailed discussion over the generalized entropies, researchers can see ([2], [3], [7]). Some of the required preliminaries that are used in the research work are:

4.1. Window function

For a hypercube of unit length 1 and dimension centered at origin, the window function is defined as,

$$\Phi(x) = \left\{\begin{matrix} 1, & |u_j| \leq \frac{1}{2} \\ 0, & otherwise \end{matrix}\right\}; (\forall j = 1,2,\ldots.,d)$$

The generalization of the window function is given by [42], known as Parzen window which is a non-parametric density estimation technique to estimate density function and is defined as,

$$P_m(y) = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{h^d}\ \varphi\left(\frac{y_i - y}{h^m}\right) \text{ with } \varphi\left(\frac{y - y_i}{h_m}\right) = \kappa$$

where $m,\ h,\ \varphi,$ and $p(y)$ is the number of elements, dimension, window function, and probability density of x. Window width and kernel are the two critical parameters of Parzen window.

4.2. Kernel density estimator

Let $x_1, x_2, \ldots., x_n$ be independent and identically distributed(i.i.d) samples talen over from univariate distribution with an density function P_h at any point x. The kernel density estimator P_h is defined as,

$$\widehat{P_h}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_i - x) \ = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

where, $K > 0$ is the kernel function and h is the smoothing parameter. Moreover, the main concern is to estimate the shape of the function.

4.3. Tsalli's measure of entropy

Tsallis entropy is a generalization of Shannon entropy and is non-extensive in nature. The continuous version of Tsalli's measure of entropy is given as,

$$H_\alpha{}^T(X) = \frac{1}{1-\alpha}\left[\int_{-\infty}^{\infty} f_X{}^\alpha(x)\ dx - 1\right] \text{ where, } 0 < \alpha < 1;\ \alpha > 1$$

the entropy has applications in adaptive systems to estimate density. In this work, Tsalli's entropy is used to prove the results and are discussed in next section.

5. MAIN RESULTS

The results are presented in the form of theorems (5.1-5.3) for the Tsallis entropy estimator. Some properties (5.2.1, 5.3.1-5.3.2) are proved for the given entropy estimator followed by the prepositions (5.2.1-5.2.2) established from the results obtained from theorems and properties, are discussed as,

- Theorem 5.1: given that for consistent Parzen windowing and sample mean, the Tsalli's entropy estimator is consistent for the probability density function of linearly independent samples.
  Proof: Parzen [48] in the estimation of the probability density function, the sample mean converges to the population mean, which is the direct implication of the consistent Parzen window estimator.
- Theorem 5.2: for equal samples and maximum value of kernel κ_λ (0), the proposed entropy estimator (4.3) is minimum.

Proof: $\widehat{H_\alpha}(x) = \frac{1}{1-\alpha}.\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N} \kappa_\lambda(x_j - x_i)\right) - \frac{1}{1-\alpha}$

For equal samples, $\widehat{H_\alpha}(x) = \frac{1}{1-\alpha}.\frac{1}{N^\alpha}\sum_{j=1}^{N}(\sum_{i=1}^{N} \kappa_\lambda(0))^{\alpha-1} - \frac{1}{1-\alpha} = \frac{1}{1-\alpha}\kappa_\lambda(0) - \frac{1}{1-\alpha}.$

Taking $\alpha = 2,\ \widehat{H_\alpha}(x) = -\kappa_\lambda(0) + 1$. To prove that the proposed entropy estimator is minimum, we need to show that,

$$\frac{1}{1-\alpha}\cdot\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}-\frac{1}{1-\alpha}\geq-\kappa_\lambda(0)+1$$

$$\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}\geq N^\alpha(-\kappa_\lambda(0)+1)(1-\alpha)+N^\alpha$$

For $\alpha>1$, $\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}\leq\kappa_\lambda(0)$
Replacing the left-hand side with the upper bound, we have,

$$\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}\leq N\max_j\left[\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}\right]\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}\leq$$

$$\frac{1}{N^{\alpha-1}}\max_j\left[\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}\right]\qquad\leq\max_{i,j}x\kappa_\lambda^{\alpha-1}(x_j-x_i)\quad;\forall x_j=x_i$$

Based on the theorems (5.1) and (5.2), following properties and propositions are proposed as:
Property 5.2.1: Erdogmus and Principe [49] proposed entropy estimator is invariant to the mean of the given density of the samples with respect to the actual entropy.
Proof: let us consider that X and $\overline{X}$ be two random variables in which $\overline{X}=X+m$ with m being a real constant.
Consider that, $H_\alpha{}^T(\overline{X})=\frac{1}{1-\alpha}\left[\int f_{\overline{X}}{}^\alpha(\overline{x})d\overline{x}-1\right]=\frac{1}{1-\alpha}\left[\int f_X{}^\alpha(x+m)dx-1\right]\quad=H_\alpha^T(X)$
Let $\{x_1,\ldots.,x_N\}$ be the samples of r.v X and $\{x_1+m,\ \ldots.,x_N+m\}$ are the samples of r.v variable $\bar{X}$.

$$\widehat{H_\alpha}{}^T(\overline{X})=\frac{1}{1-\alpha}\left[\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(\overline{x_j}-\overline{x_i})\right)^{\alpha-1}-1\right]$$

$$=\frac{1}{1-\alpha}\left[\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j+m-x_i-m)\right)^{\alpha-1}-1\right]$$

$$=\frac{1}{1-\alpha}\left[\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}-1\right]=\widehat{H_\alpha}{}^T(X)$$

Remarks:
Using Parzen windowing with sample mean approximation for expectation, following has been obtained as,
a) As $\alpha\to1$, $H_\alpha{}^T(X)\to H_S(X)$; where $H_S(X)$ is Shannon measure of entropy
b) For $\alpha=1$, $H_\alpha{}^T(X)$ is discontinuous
c) The derivative of the Shannon entropy is same as proposed entropy.
Proposition 5.2.1: using a kernel of size λ, if entropy can be estimated for samples $\{x_1,\ldots.,x_N\}$of a r.v X, then to estimate the samples $\{cx_1,\ldots.,cx_N\}$of a r.v cX, a kernel of size $|c|\lambda$ must be employed.
Proof: consider Tsalli's entropy of random variable $CX=\{cx_1,\ldots\ldots\ldots.,cx_N\}$,

$$H_\alpha(X)=\frac{1}{1-\alpha}\left[\int_{-\infty}^{\infty}f_X{}^\alpha(x)\,dx-1\right]$$

$$H_\alpha(CX)=\frac{1}{1-\alpha}\left[\int_{-\infty}^{\infty}\frac{1}{|C|^\alpha}f_X{}^\alpha\left(\frac{x}{c}\right)dx-1\right]=\frac{1}{1-\alpha}\left[\int_{-\infty}^{\infty}f_X{}^\alpha(t)\,dt-1\right]=\quad H_\alpha(X)$$

Proposition 5.2.2: let us consider the following,
a. Component of X: $X=\{x_1,\ldots.,x_N\}$ with $X^0\to\ 0^n$
b. Single dimensional kernel: $\kappa_\lambda{}^0(.)$
c. Multidimensional kernel for joint PDF: $\kappa_\Sigma{}^0(.)$
d. Parzen estimate for joint PDF: $f_X{}^n(x)=\frac{1}{N}\sum_{i=1}^{N}\kappa_\Sigma\,(x-x_i)$
e. Parzen estimate for marginal density of X^0: $f_X{}^n(x)=\frac{1}{N}\sum_{i=1}^{N}\kappa_{\lambda_0}{}^0\,(x^0-x_i{}^0)$
− Theorem 5.3: given that the $\kappa_\lambda(.)$ is continuous, differentiable, symmetric, and unimodal, the global minimum of the entropy estimator (4.3) is smooth.
Proof: $\widehat{H_\alpha}(x)=\frac{1}{1-\alpha}\left(\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}-1\right)$
Let is consider an argument with variable is defined as: $\widehat{P_\alpha}=\frac{1}{N^\alpha}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\kappa_\lambda(x_j-x_i)\right)^{\alpha-1}$
To prove that the proposed estimator is at global minimum, it is sufficient to show that $\kappa_\lambda(0)$ have semi-definite Hessian matrix with zero gradient. The gradient and the expressions of hessian matrix of the proposed entropy estimator are presented as, 0.

$$\frac{\partial \widehat{H_\alpha}}{\partial x_k} = \frac{1}{1-\alpha}\frac{\partial \widehat{P_\alpha}}{\partial x_k};\ \frac{\partial^2 \widehat{H_\alpha}}{\partial x_\ell \partial x_k} = \frac{1}{1-\alpha}\frac{\partial^2 \widehat{P_\alpha}}{\partial x_\ell \partial x_k}$$

Evaluating this expression at $\overline{x} = 0$, we get $\widehat{P_\alpha}\big|_{\overline{x}=0} = \kappa_\lambda^{\alpha-1}(0)$,

$$\left.\frac{\partial \widehat{P_\alpha}}{\partial x_k}\right|_{\overline{x}=0} = \frac{(\alpha-1)}{N^\alpha}\left[N^{\alpha-1}\kappa_\lambda^{\alpha-2}(0)\kappa'(0) - N^{\alpha-1}\kappa_\lambda^{\alpha-2}(0)\kappa'(0)\right] = 0$$

$$\left.\frac{\partial^2 \widehat{P_\alpha}}{\partial x_k^2}\right|_{\overline{x}=0} = \frac{(\alpha-1)(N-1)\kappa_\lambda^{\alpha-3}(0)}{N^2}\left[(\alpha-2){\kappa'}^2(0) + 2\kappa(0)\kappa''(0)\right]$$

$$\left.\frac{\partial^2 \widehat{P_\alpha}}{\partial x_\ell \partial x_k}\right|_{\overline{x}=0} = -\frac{(\alpha-1)\kappa_\lambda^{\alpha-3}(0)}{N^2}\left[(\alpha-2){\kappa'}^2(0) + 2\kappa(0)\kappa''(0)\right] \text{ which shows that,}$$

$$\left.\frac{\partial^2 \widehat{H_\alpha}}{\partial x_\ell \partial x_k}\right|_{\overline{x}=0} = \begin{cases} -(N-1)\kappa_\lambda^{\alpha-3}(0)\left[(\alpha-2){\kappa'}^2(0)+2\kappa(0)\kappa''(0)\right]/N^2, \ell=k \\ \kappa_\lambda^{\alpha-3}(0)\left[(\alpha-2){\kappa'}^2(0)+2\kappa(0)\kappa''(0)\right]/N^2, \ell\neq k \end{cases}$$

The eigen-pairs of the hessian matrix are,

$$\{0, [1, \ldots, 1]\},\ \left\{\frac{kN}{N-1}, [1, -1, 0, \ldots, 0]^T\right\}, \left\{\frac{kN}{N-1}, [1, 0, -1, 0, \ldots, 0]^T\right\}, \ldots..$$

where k and l denoting the diagonal and off-diagonal entries of the matrix.

There is one eigenvector corresponding to some eigenvalue that changes the mean of the data. Therefore, the Hessian matrix is a semi-definite. From the results, it is concluded that:

a. For given $\kappa_\lambda(.)$, non-zero eigenvalue has a multiplicity of $N-1$.
b. Since $\kappa_\lambda(0) > 0$, the eigenvalue is positive for $N > 1$
c. Also, $\kappa_\lambda{}'(.) = 0$ and $\kappa_\lambda{}''(.) < 0$, the hessian matrix is positive semi-definite.

Thus, the entropy estimator is at the global minimum which shows that the proposed estimator is suitable for entropy minimization adaptive systems. The properties of the kernel function underlying the proposed entropy measure are discussed as:

Property 5.3.1: as $\kappa_\lambda(.) \to \infty$, the proposed quadratic version of the entropy estimator approaches to the negative of the scaling and the biasedness of the sample variance.

Proof: at $\alpha = 2$;

$$\widehat{H}_2(X) = -\frac{1}{N^2}\sum_j \sum_i \left(\kappa_\lambda(0) + \frac{\kappa_\lambda''(0)}{2}\ (x_j - x_i)^2\right)\ + 1$$

$$= -\left[\kappa_\lambda(0) + \frac{\kappa_\lambda''(0)}{2}.\frac{1}{N^2}\sum_{j=1}^N \sum_{i=1}^N (x_j - x_i)^2\right] + 1$$

$$= -\left[[\kappa_\lambda(0) - 1] + \frac{(\overline{x^2} - \overline{x}^2)}{2}.\kappa_\lambda''(0)\right]$$

Property 5.3.2: in the case of joint entropy estimation, for all orthonormal matrices R, the multi-dimensional kernel function satisfies the condition $\kappa_\Sigma(\vartheta) = \kappa_\Sigma(R^{-1}\vartheta)$, then the proposed entropy estimator is invariant under rotation as the actual entropy estimator of a random variable.

Proof,

$$\widehat{H}_\alpha(\overline{X}) = \frac{1}{1-\alpha}\frac{1}{N^\alpha}\sum_j \left(\sum_i \kappa_\Sigma\ \ (Rx_j - Rx_i)\right)^{\alpha-1} - \frac{1}{1-\alpha}$$

$$= \frac{1}{1-\alpha}\frac{1}{N^\alpha}\sum_j \left(\sum_i \frac{1}{|R|}\kappa_\Sigma\left(R^{-1}(Rx_j - Rx_i)\right)\right)^{\alpha-1} - \frac{1}{1-\alpha}$$

$$\widehat{H}_\alpha(\overline{X}) = \frac{1}{1-\alpha}|R|^{1-\alpha}\frac{1}{N^\alpha}\sum_j \left(\sum_i \kappa_\Sigma(x_j - x_i)\right)^{\alpha-1} - \frac{1}{1-\alpha} = |R|^{1-\alpha}\left[\widehat{H}_\alpha(x) + 1\right] - \frac{1}{1-\alpha}$$

$$H_\alpha(\overline{X}) = \frac{1}{1-\alpha}\left[\int_{-\infty}^{\infty} f_{\overline{X}}{}^\alpha(\overline{x})d\overline{x} - 1\right] = \frac{1}{1-\alpha}\left[\int_{-\infty}^{\infty}\frac{1}{|R|^\alpha} f_{\overline{X}}{}^\alpha(R^{-1}\overline{x})d\overline{x} - 1\right]$$

$$= \frac{1}{1-\alpha}\left[|R|^{1-\alpha}\int_{-\infty}^{\infty} f_X{}^\alpha(x)\,dx + |R|^{1-\alpha} - |R|^{1-\alpha}\right] - \frac{1}{1-\alpha} = |R|^{1-\alpha}[H_\alpha(X) + 1] - \frac{1}{1-\alpha}$$

6. CONCLUSION

An entropy estimator using Tsalli's entropy has been proposed and the results have been discussed in the theorems (5.1-5.3). It has been noticed that the proposed entropy estimator is consistent, differentiable, and smooth under the given conditions. Some important properties and propositions have been established with the help of the proposed Tsallis-estimator. The proposed estimator has applications to the problems of dimensionality reduction, feature extraction, blind source separation, and power system risk assessment. In multivariate densities, the estimator can be used to test the consistency under certain conditions and have applications in image analysis and computer vision.

REFERENCES

[1] P. Gacs, "The Boltzmann entropy and randomness tests," in *Proceedings Workshop on Physics and Computation. PhysComp '94*, 1994, pp. 209–216, doi: 10.1109/PHYCMP.1994.363679.

[2] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948, doi: 10.1002/j.1538-7305.1948.tb00917.x.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, USA: John Wiley & Sons, 2005.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York, USA: Cambridge University Press, 2011.

[5] R. M. Fano and D. Hawkins, "Transmission of Information: A Statistical Theory of Communications," *American Journal of Physics*, vol. 29, no. 11, pp. 793–794, 1961, doi: 10.1119/1.1937609.

[6] S. Kullback, *Information theory and statistics*. Mineola, New York: Dover Publication, 1997.

[7] A. Renyi, "Some Fundamental Questions of Information Theory," *Selected Papers of Alfred Renyi, Academia Kiado, Budapest*, vol. 2, pp. 526–552, 1976.

[8] A. Renyi, *A Diary on Information Theory*. Hungary: Akadémiai Kiadó, 1984.

[9] P. Sahoo, C. Wilkins, and J. Yeager, "Threshold selection using Renyi's entropy," *Pattern Recognition*, vol. 30, no. 1, pp. 71–84, 1997, doi: 10.1016/S0031-3203(96)00065-9.

[10] D. P. Mittal, "On Additive and Non-Additive Entropies," *Kybernetika*, vol. 11, no. 4, pp. 271–276, 1975.

[11] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, no. 4, pp. 423–429, 1965, doi: 10.1016/S0019-9958(65)90332-3.

[12] M. Ben-Bassat and J. Raviv, "Renyi's Entropy and the Probability of Error," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 324–331, 1978, doi: 10.1109/TIT.1978.1055890.

[13] C. Cachin, "Smooth Entropy and Renyi Entropy," in *Advances in Cryptology — EUROCRYPT '97*, Berlin, Heidelberg: Springer, 1997, pp. 193–208.

[14] J. W. Fisher, "Nonlinear extensions to the minimum average correlation energy filter," University of Florida, 1997.

[15] J. C. Principe, D. Xu, and J. W. Fisher, "Information Theoretic Learning," in *Unsupervised Adaptive Filtering*, New York: Springer, 2000, pp. 265–319.

[16] J. C. Principe, D. Xu, Q. Zhao, and J. W. Fisher, "Learning from examples with information theoretic criteria," *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 26, no. 1, pp. 61–77, 2000, doi: 10.1023/a:1008143417156.

[17] D. Xu, J. C. Principe, J. Fisher, and H.-C. Wu, "A novel measure for independent component analysis (ICA)," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, 1998, pp. 1161–1164, doi: 10.1109/ICASSP.1998.675476.

[18] D. Xu, "Energy, Entropy and Information Potential for Neural Computation," University of Florida, 1999.

[19] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis*. Cambridge: Cambridge University Press, 1998.

[20] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, USA: University od Illinois Press, 1964.

[21] C. Beck and F. Schögl, *Thermodynamics of Chaotic Systems*. Cambridge: Cambridge University Press, 1993.

[22] S. Haykin, *Unsupervised Adaptive Filtering*. New York: John Wiley & Sons, 2000.

[23] J. Beirlant, E. Dudewicz, L. Györfi, and E. Van der Meulen, "Nonparametric entropy estimation: An overview," *International Journal of Mathematical and Statistical Sciences*, vol. 6, no. 1, pp. 17–39, 1997.

[24] Y. G. Dmitriev and F. P. Tarasenko, "On the Estimation of Functionals of the Probability Density and Its Derivatives," *Theory of Probability & Its Applications*, vol. 18, no. 3, pp. 628–633, 1974, doi: 10.1137/1118083.

[25] H. Joe, "Estimation of entropy and other functionals of a multivariate density," *Annals of the Institute of Statistical Mathematics*, vol. 41, no. 4, pp. 683–697, 1989, doi: 10.1007/BF00057735.

[26] I. A. Ahmad and P. E. Lin, "A Nonparametric Estimation of the Entropy for Absolutely Continuous Distributions," *IEEE Transactions on Information Theory*, vol. 22, no. 3, pp. 372–375, 1976, doi: 10.1109/TIT.1976.1055550.

[27] J. F. Bercher, "Estimating the entropy of a signal with applications," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1687–1694, 2000, doi: 10.1109/78.845926.

[28] R. Boscolo, H. Pan, and V. P. Roychowdhury, "Non-Parametric ICA," in *Proceedings of the Third International Conference on Independent Component Analysis and Blind Source Separation (ICA 2001)*, 2001, pp. 13–18.

[29] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[30] P. Viola, N. N. Schraudolph, and T. J. Sejnowski, "Empirical Entropy Manipulation for Real-World Problems," in *NIPS 1995: Proceedings of the 8th International Conference on Neural Information Processing Systems*, 1995, pp. 851–857.

[31] H. H. Yang and S. Amari, "Adaptive Online Learning Algorithms for Blind Separation: Maximum Entropy and Minimum Mutual Information," *Neural Computation*, vol. 9, no. 7, pp. 1457–1482, 1997, doi: 10.1162/neco.1997.9.7.1457.

[32] L. Györfi and E. C. van der Meulen, "Density-free convergence properties of various estimators of entropy," *Computational Statistics and Data Analysis*, vol. 5, no. 4, pp. 425–436, 1987, doi: 10.1016/0167-9473(87)90065-X.

[33] L. Györfi and E. C. van der Meulen, "An entropy estimation based on kernel density estimation," in *Limit theorems in probability and statistics*, 1989, pp. 229–240.

[34] L. Györfi and E. C. Meulen, "On the Nonparametric Estimation of the Entropy Functional," in *Nonparametric Functional Estimation and Related Topics*, Dordrecht, Netherlands: Springer, 1991, pp. 81–95.

[35] A. V Ivanov and M. N. Rozhkova, "Properties of the statistical estimate of the entropy of a random vector with a probability density," *Problems of Information Transmission*, vol. 17, pp. 171–178, 1981.

[36] J. Beirlant and M. C. A. van Zuijlen, “The empirical distribution function and strong laws for functions of order statistics of uniform spacings,” *Journal of Multivariate Analysis*, vol. 16, no. 3, pp. 300–317, 1985, doi: 10.1016/0047-259X(85)90023-5.
[37] P. Hall, “Limit theorems for sums of general functions of m-spacings,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 96, no. 3, pp. 517–532, 1984, doi: 10.1017/S0305004100062459.
[38] F. P. Tarasenko, “On the Evaluation of an Unknown Probability Density Function, the Direct Estimation of the Entropy from Independent Observations of a Continuous Random Variable, and the Distribution-Free Entropy Test of Goodness-of-Fit,” *Proceedings of the IEEE*, vol. 56, no. 11, pp. 2052–2053, 1968, doi: 10.1109/PROC.1968.6784.
[39] L. F. Kozachenko and N. N. Leonenko, “Sample Estimate of the Entropy of a Random Vector.,” *Problems of information transmission*, vol. 23, no. 2, pp. 95–101, 1987.
[40] A. B. Tsybakov and E. C. van der Meulen, “Root-N consistent estimators of entropy for densities with unbounded support,” *Proceedings of 1994 Workshop on Information Theory and Statistics*, vol. 23, no. 1, pp. 75–83, 1996.
[41] P. J. Bickel and L. Breiman, “Sums of Functions of Nearest Neighbor Distances, Moment Bounds, Limit Theorems and a Goodness of Fit Test,” *The Annals of Probability*, vol. 11, no. 1, pp. 185–214, 1983, doi: 10.1214/aop/1176993668.
[42] R. Ul Hassan, J. Yan, and Y. Liu, “Security risk assessment of wind integrated power system using Parzen window density estimation,” *Electrical Engineering*, vol. 104, no. 4, pp. 1997–2008, 2022, doi: 10.1007/s00202-021-01447-7.
[43] H. S. Bakouch, O. A. Elsamadony, and C. Chesneau, “A note on density estimation via the hyperbolic secant kernel estimator,” *Journal of Information and Optimization Sciences*, vol. 43, no. 8, pp. 2007–2019, 2022, doi: 10.1080/02522667.2022.2084244.
[44] N. Aruga and M. Tanaka, “Information theoretic learning with maximizing Tsallis entropy,” in *ITC-CSCC :International Technical Conference on Circuits Systems, Computers and Communications*, 2002, pp. 810–813.
[45] J. Li, Y. Xie, L. Dang, C. Song, and B. Chen, “Fixed-Point Minimum Error Entropy with Sparsity Penalty Constraints,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 8, pp. 2997–3001, 2021, doi: 10.1109/TCSII.2021.3063371.
[46] Abhishek, R. Kumar Yadav, and S. Verma, “Parzen Window Approximation on Riemannian Manifold,” *Pattern Recognition*, vol. 134, 2023, doi: 10.1016/j.patcog.2022.109081.
[47] F. Xiong, J. Zhang, Y. Ling, and Z. Zhang, “A Novel Image Thresholding Method Combining Entropy With Parzen Window Estimation,” *Computer Journal*, vol. 65, no. 8, pp. 2231–2244, 2022, doi: 10.1093/comjnl/bxab182.
[48] E. Parzen, “On estimation of a probability density function and mode,” *The Annals of Mathematical Statistics.*, vol. 33, no. 3, pp. 1065–1076, 1962.
[49] D. Erdogmus and J. C. Principe, “Convergence analysis of the information potential criterion in Adaline training,” in *Neural Networks for Signal Processing XI: Proceedings of the 2001 IEEE Signal Processing Society Workshop (IEEE Cat. No.01TH8584)*, 2001, pp. 123–132, doi: 10.1109/NNSP.2001.943117.

BIOGRAPHIES OF AUTHORS

Leena Chawla is research scholar in Department of Applied Sciences, Faculty of Engineering and Technology, Manav Rachna International Institute of Research and Studies, Faridabad. She has more than 10 years of teaching experience of graduates and postgraduates. She has more than 4 research articles published in journals. She awarded her M.Phil. in 2010 from Chaudhary Devi Lal University, Sirsa, Haryana, India. She can be contacted at email: leenabharat14@gmail.com.

Dr. Vijay Kumar is working as a Professor of Mathematics, Manav Rachna International Institute of Research & Studies, Faridabad, Haryana, India. His research interest includes decision making and machine learning. He has authored more than 60 research publications including research papers, conference papers, and chapters in books. He has authored 3 books, out of which 2 are research books. He gave several invited talks in the events of national and international repute. He serves as editorial board member for several international/national journals and reviewer for many national and international journals. He can be contacted at email: vijaykumar.fet@mriu.edu.in.

Dr. Arti Saxena is working as associate professor in Department of Applied Sciences, Faculty of Engineering and Technology, Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India. She has more than 40 research articles in various national and international journals of repute and conference proceedings. She has four research books published to her account and four chapters in edited books. She has got young scientist award in an international conference for best paper presentation in 2007. Her areas of interest are mathematical modeling, fixed point theory, and machine learning. She can be contacted at email: artisaxena.fet@mriu.edu.in.