International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 5, October 2024, pp. 6016~6022
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i5.pp6016-6022

Estimation of kernel density function using Kapur entropy

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 Apr 13, 2024
Revised Jul 11, 2024
Accepted Jul 17, 2024

Keywords:

Entropy estimator
Information theoretic measure
Kapur entropy
Kernel density estimation
Non-parametric estimator

ABSTRACT

Information-theoretic measures play a vital role in training learning systems. Many researchers proposed non-parametric entropy estimators that have applications in adaptive systems. In this work, a kernel density estimator using Kapur entropy of order α and type β has been proposed and discussed with the help of theorems and properties. From the results, it has been observed that the proposed density measure is consistent, minimum, and smooth for the probability density function (PDF) underlying given conditions and validated with the help of theorems and properties. The objective of the paper is to understand the theoretical viewpoint 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.set@mriu.edu.in

1. INTRODUCTION

Entropy estimation plays a significant role across the disciplines of science and technology such as engineering [1]–[3], biology [4], and physics [5]. Mathematical generalization of non-parametric entropy in terms of continuous random variables has been proposed by many researchers. Univariate and multivariate probability density functions (PDFs)-based Shannon entropy expressions were discussed in [6], [7]. In a simple parametric family, the changing PDF of the data may not exist. Therefore, it is necessary to estimate non-parametric entropy. These estimates are obtained by introducing a density estimator of the data in the entropy expression instead of the actual PDF. PDFs are required to estimate entropy, which can effectively evaluate entropy using kernel density methods, a well-studied area of research. Wegman and Davies [8] introduced a recursive density estimator to estimate time series and spatial data parameters. Kernel density estimators are widely used for the estimation of entropy [9] because they are computationally faster, easy to incorporate, and simple to understand [10].

2. PLUG-IN ESTIMATORS

Plug-in estimators are used to estimate a feature of probability distribution. Different plug-in estimators have been used for density estimation, such as integral estimates, resubstituting estimates, splitting data estimates, and cross-validation estimates. A brief introduction to the different types of plug-in estimators is discussed as follows.

- Integral estimates: approximate infinite integrals presented in the entropy expression. Entropy measures have been used to estimate the exact evaluation of the integral. Dmitriev and Tarasenko [11] proposed a similar estimation for Shannon entropy. In multivariate cases, Joe [12] approximated the integral estimate of Shannon entropy using kernel estimation of PDF and investigated that integral approximation becomes complicated in many cases of study.

- Re-substitution estimates: the expectation operator has been approximated in the entropy expression with the sample mean. In 1976, Ahmad and Lin [10] proposed an estimate for Shannon entropy using a kernel and investigated the mean-square consistency of the proposed estimate. In multivariate situations, Joe [12] discussed the re-substitution estimate of Shannon entropy with kernel PDF and obtained that the number of samples are increasing with the dimensionality of the data. In electrical engineering problems, entropy estimates can be obtained using spectral estimation based on polynomial type PDF [13]. Still, from the literature, it is revealed that re-substitution estimates are used to estimate Shannon entropy. Kapur's entropy of order α and type β is also part of the re-substitution class. Depending on the kind of application, researchers customize entropy estimation according to the requirement of the algorithm.
- Spilling data estimate is similar to re-substitution estimate but has a different methodology. In this estimation, a data sample is fragmented into two parts; the first part is utilized for density estimation, and the other part is used for sample mean [14]–[16].
- Cross-validation estimate is the generalization of re-substitution estimate based on the principle of leave-one-out. The estimate is obtained by taking the mean of the leave-one-out re-substitution estimates of the given data set. Ivanov and Rozhkova [17] proposed a cross-validation entropy estimator for Shannon entropy using kernel PDF.

3. USEFUL ESTIMATES

A sample of observations drawn from the given distribution. Plug-in estimate is a feature that can be approximated by the same feature of the empirical distribution. Other than plug-in estimates, more estimates have been discussed in the literature, proposed by various researchers under different situations.

3.1. Estimates based on sample spacing

Based on sample differences, a density estimate is constructed. In univariate cases, PDF can be estimated by ordering the samples from the smallest to the largest and separating the samples by defining *m*-spacing between the samples and then substituting the PDF estimate in the entropy expression (like re-substitution estimates). However, these density estimates need to be more consistent, and their generalization in multivariate cases is not trivial [18]–[20].

3.2. Estimates based on nearest neighbor distances

In non-parametric statistics, nearest neighbor methods are considered a classical approach. In the multivariate cases, the density estimate is constructed as the sample mean of the logarithm of the normalized nearest neighbor distances with a constant. In [21]–[23], the proposed nearest neighbor estimates have various forms of consistency under certain conditions.

3.3. Entropy estimation for learning

The literature shows that the Shannon measure of entropy drew a lot of attention from researchers to propose algorithms for various learning systems. Estimating Shannon’s entropy has applied to Renyi’s entropy, a generalization of Shannon's entropy. This work proposes simple entropy estimators for training learning systems using Kapur entropy of order α and type β that are continuous and differentiable in terms of samples. The main objective of the work is to understand the mathematics behind the concept.

4. PRELIMINARIES

The terms like entropy and the measures of entropy are not covered as these were discussed by many researchers. Some of the required preliminaries are defined as follows.

4.1. Window function

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

$$\Phi(y) = \begin{Bmatrix} 1, & |u_j| \leq \dfrac{1}{2} \\ 0, & otherwise \end{Bmatrix}; (j = 1,2,\ldots.,d)$$

The generalization of the window function is given by Hassan *et al.* [24], known as the Parzen Window, a technique to estimate density function. This is a non-parametric density estimation technique, defined as

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

where $m,\ h,\ \varphi$ and $p(y)$ are the numbers of elements, dimension, window function, and probability density of y. Window width and kernel are the two critical parameters of Parzen Window. Let $\{y_1, \dots, y_N\}$ be the samples N drawn from the random variable. These samples are independent and identically distributed (i.i.d). The kernel function (arbitrary) $\kappa_\sigma(.)$ estimate of the PDF is given by Parzen [25] and is defined as:

$$\hat{f}_Y(y) = \frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(y - y_i)$$

4.2. Kernel density estimator

In the estimation of probability distribution, kernel density estimator is used when there are more data points in a sample around a location, the likelihood of occurrence for an observation is higher at that location. For smoothing parameter h and non-zero kernel function κ, the general expression of kernel density estimator is defined as:

$$\widehat{P_h}(y) = \frac{1}{m}\sum_{i=1}^{m}\kappa_h(y_i - y) = \frac{1}{m}\sum_{i=1}^{m}\kappa\left(\frac{y - y_i}{h}\right)$$

4.3. Kapur's Measure of entropy

The continuous and differentiable form of Kapur entropy of order α and type β is given as Kapur [26]:

$$H^K_{\alpha,\beta}(p) = \frac{1}{\beta - \alpha}\ log\ \left(\frac{\sum_{i=1}^{n} p_i^\alpha}{\left(\sum_{i=1}^{n} p_i^\beta\right)}\right);\ where\ \sum_{i=1}^{n} p_i \leq 1;\ \alpha \neq \beta;\ \alpha,\beta > 0 \quad (1)$$

Kapur entropy with expectation operator is written as,

$$H^K_{\alpha,\beta}(Y) = \frac{1}{\beta - \alpha}\ log\ \left(\frac{\int_{-\infty}^{\infty} f_Y^\alpha(y)\ dy}{\int_{-\infty}^{\infty} f_Y^\beta(y)\ dy}\right) = \frac{1}{\beta - \alpha}\ log\ \left(\frac{E_Y[f_Y^{\alpha-1}(y)]}{E_Y[f_Y^{\beta-1}(y)]}\right)$$

Entropy approximation with sample mean is written as,

$$H^K_{\alpha,\beta}(Y) = \frac{1}{\beta - \alpha}\ log\frac{1}{N}\left(\frac{\sum_{j=1}^{N} f_Y^{\alpha-1}(y_j)}{\sum_{j=1}^{N} f_Y^{\beta-1}(y_j)}\right)$$

Using Parzen Window estimator, the non-parametric estimator of Kapur entropy (1) is given as,

$$H^K_{\alpha,\beta}(Y) = \frac{1}{\beta - \alpha}\ log\frac{1}{N}\left(\frac{\left(\sum_{j=1}^{N}\frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(y_j - y_i)\right)^{\alpha-1}}{\left(\sum_{j=1}^{N}\frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(y_j - y_i)\right)^{\beta-1}}\right)$$

5. MAIN RESULTS

The proposed non-parametric entropy (1) is general. Therefore, it can estimate entropy on a learning system that employs a performance index to approximately defined weights. The results are presented in the form of theorems and properties for the proposed entropy estimation (1) and are discussed as follows.

5.1. Theorem 5.1

Given that for consistent Parzen Windowing and sample mean, Kapur entropy of order α and type β entropy estimator (1) is consistent for the probability density function of linearly independent samples.
Proof: According to Parzen [25], in the estimation of the PDF, the sample mean converges to the population mean, which is the direct implication of the consistent Parzen Window estimator (1).

5.2. Property 5.2

Entropy estimator (1) using Parzen Windowing and mean approximation is the limiting case of Shannon entropy for $(\alpha, \beta \to 1)$ both in continuous and discrete case.
Proof: It should be noted that Kapur entropy of order α and type β (1) is discontinuous at $\alpha = \beta$. For $(\alpha, \beta \to 1)$, shannon measure of entropy is obtained.
In continuous case:

$$\begin{aligned} H_{\alpha,\beta}(Y) &= \frac{1}{\beta-\alpha} log\left[\frac{\int {f_Y}^{\alpha}(y)\ dy}{\int {f_Y}^{\beta}(y)\ dy}\right] \\ &= \lim_{\substack{\alpha\to1\\ \beta\to1}}\left\{\frac{1}{\beta-\alpha}\left[log\left(\frac{\int {f_Y}^{\alpha}(y)\ dy}{\int {f_Y}^{\beta}(y)\ dy}\right)\right]\right\} \\ &= -\int f_Y(y)\, log\ \ f_Y(y)\ dy\ -\ \int f_Y(y)\ \ log\, f_Y\,(y)\ dy \\ &= 2H_S(Y) \end{aligned}$$

In discrete case:

$$\begin{aligned} \lim_{\substack{\alpha\to1\\ \beta\to1}} \widehat{H}_{\alpha,\beta}(Y) &= \lim_{\substack{\alpha\to1\\ \beta\to1}} \frac{1}{\beta-\alpha} log\left[\frac{\frac{1}{N}\sum_j\left(\frac{1}{N}\sum_i \kappa_\lambda(y_j-y_i)\right)^{\alpha-1}}{\frac{1}{N}\sum_j\left(\frac{1}{N}\sum_i \kappa_\lambda(y_j-y_i)\right)^{\beta-1}}\right] \\ &= \lim_{\alpha\to1} -\frac{1}{N}\sum_j log\left(\frac{1}{N}\sum_i \kappa_\lambda(y_j-y_i)\right) - \lim_{\beta\to1}\frac{1}{N}\sum_j log\left(\frac{1}{N}\sum_i \kappa_\lambda(y_j-y_i)\right) \\ &= \ \ -2\ \widehat{H}_S(Y) \end{aligned}$$

Therefore, it is proved that the proposed entropy measure (1) approaches Shannon entropy.

5.3. Property 5.3

Scaling property of entropy estimator (1). Let $\{y_1, \ldots., y_N\}$ be the samples of random variable Y, estimated with kernel size λ. To estimate samples $\{cy_1, \ldots., cy_N\}$ of a random variable $c\ Y$, the size of the kernel will be scaled to $|c|$, *i.e.*, new kernel size is $|c|\lambda$.
Proof: Consider that for a random variable $c\ Y$, the PDF is $f_Y(y/c)/|c|$
In continuous case:

$$\begin{aligned} H_{\alpha,\beta}(Y) &= \frac{1}{\beta-\alpha} log\left[\frac{\int {f_Y}^{\alpha}(y)\ dy}{\int {f_Y}^{\beta}(y)\ dy}\right] \\ &= \frac{1}{\beta-\alpha}\left[log\int_{-\infty}^{\infty}\frac{1}{|c|}{f_Y}^{\alpha}\left(\frac{y}{c}\right)\ dy - log\int_{-\infty}^{\infty}\frac{1}{|c|}{f_Y}^{\beta}\left(\frac{y}{c}\right)\ dy\right] \\ &= \frac{1}{\beta-\alpha}\big[(1-\alpha)H_\alpha(Y) - (1-\beta)H_\beta(Y)\big] + log|c| \\ \Rightarrow H_{\alpha,\beta}(cY) &= \left\{\begin{matrix} H_\alpha(Y)\ +\ log|c|, & \beta=1 \\ H_\beta(Y)\ +\ log|c|, & \alpha=1 \end{matrix}\right\} \end{aligned}$$

In discrete case:

$$\begin{aligned} \widehat{H}_{\alpha,\beta}(cY) &= \frac{1}{\beta-\alpha}\left[\begin{matrix} log\frac{1}{N^\alpha}\sum_j\left(\sum_i \kappa_{|c|\lambda}(y_j-y_i)\right)^{\alpha-1} - \\ log\frac{1}{N^\beta}\sum_j\left(\sum_i \kappa_{|c|\lambda}(y_j-y_i)\right)^{\beta-1} \end{matrix}\right] \\ &= \frac{1}{\beta-\alpha}\big[(1-\alpha)\widehat{H}_\alpha(Y) - (1-\beta)\widehat{H}_\beta(Y) + (\beta-\alpha)\ log|c|\big] \\ \Rightarrow \widehat{H}_{\alpha,\beta}(cY) &= \left\{\begin{matrix} \widehat{H}_\alpha(Y)\ +\ log|c|, & \beta=1 \\ \widehat{H}_\beta(Y)\ +\ log|c|, & \alpha=1 \end{matrix}\right\} \end{aligned}$$

It is concluded that the kernel size affects the width of the kernel function linearly.
Remarks: The property (5.3) has applications when the entropy cost function is near the global extremum and is obtained by scaling the entropy cost function.

5.4. Theorem 5.4

According to Chawla [27], for equal samples, $y_1 = y_2 = \dots = y_N = c$ (say) with maximum value of kernel $\kappa_\lambda(0)$, the proposed entropy estimator (1) is minimum.
Proof: Since, $y_1 = y_2 = \dots = y_N = c$ (say), the entropy estimator (1) obtained the value $-\log K_\lambda(0)$. We have to show that

$$\frac{1}{\beta-\alpha}\left[\log\frac{1}{N^\alpha}\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\alpha-1} - \log\frac{1}{N^\beta}\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\beta-1}\right] \geq -\log\kappa_\lambda(0)$$

$$\Rightarrow \left\{\frac{\frac{1}{N^\alpha}\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\alpha-1}}{\frac{1}{N^\beta}\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\beta-1}}\right\} \leq {\kappa_\lambda}^{\alpha-\beta}(0)$$

$$\Rightarrow \frac{\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\alpha-1}}{\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\beta-1}} \leq N^{\alpha-\beta}{\kappa_\lambda}^{\alpha-\beta}(0)$$

replacing with the upper bounds

$$= \frac{\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\alpha-1}}{\sum_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\beta-1}} \leq \frac{N\max_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\alpha-1}}{N\max_j\left(\sum_i \kappa_\lambda(y_j-y_i)\right)^{\beta-1}}$$

$$\leq \frac{\max_j\left(N^{\alpha-1}\max_i\left[{\kappa_\lambda}^{\alpha-1}(y_j-y_i)\right]\right)}{\max_j\left(N^{\beta-1}\max_i\left[{\kappa_\lambda}^{\beta-1}(y_j-y_i)\right]\right)}$$

$$= N^{\alpha-\beta}\max_{i,j}{\kappa_\lambda}^{\alpha-\beta}(y_j-y_i) = N^{\alpha-\beta}{\kappa_\lambda}^{\alpha-\beta}(0)$$

It is evident that entropy cost function reaches its global minimum, when all the error samples are zero and used to train kernels in supervised learning.

5.5. Property 5.5

For joint entropy estimation of a random variable, all orthonormal matrices R, satisfies the condition $\kappa_\Sigma(\vartheta) = \kappa_\Sigma(R^{-1}\vartheta)$, the proposed entropy estimator (1) is invariant under rotation with $\kappa_\Sigma(\vartheta)$ as the multi-dimensional kernel function.
Proof: For n-dimensional random vectors Y and $\overline{Y}$, $\exists\ \overline{Y} = RY$ in which $[R]_{n\times n}$ real orthonormal matrix. In continuous case:

$$H_{\alpha,\beta}(\overline{Y}) = \frac{1}{\beta-\alpha}\log\left[\frac{\int {f_{\overline{Y}}}^{\alpha}(\overline{y})\ d\overline{y}}{\int {f_{\overline{Y}}}^{\beta}(\overline{y})\ d\overline{y}}\right]$$

$$= \frac{1}{\beta-\alpha}\left[\log\int_{-\infty}^{\infty}\frac{1}{|R|^\alpha}{f_Y}^{\alpha}(R^{-1}\overline{y})\ d\overline{y} - \log\int_{-\infty}^{\infty}\frac{1}{|R|^\beta}{f_Y}^{\beta}(R^{-1}\overline{y})\ d\overline{y}\right]$$

$$= \frac{1}{\beta-\alpha}\left[\log|R|^{1-\alpha}\int_{-\infty}^{\infty}{f_Y}^{\alpha}(y)\ dy - \log|R|^{1-\beta}\int_{-\infty}^{\infty}{f_Y}^{\beta}(y)\ dy\right]$$

$$= \frac{1}{\beta-\alpha}\left[(1-\alpha)H_\alpha(Y) - (1-\beta)H_\beta(Y) + (\beta-\alpha)\log|R|\right]$$

$$H_{\alpha,\beta}(\overline{Y}) = \begin{cases} H_\alpha(Y) + \log|R|, & \beta = 1 \\ H_\beta(Y) + \log|R|, & \alpha = 1 \end{cases}$$

In discrete case:

$$\widehat{H}_{\alpha,\beta}(\overline{Y}) = \frac{1}{\beta-\alpha} log\left[\frac{\frac{1}{N^{\alpha}}\sum_j\left(\sum_i \kappa_{\Sigma}(Ry_j - Ry_i)\right)^{\alpha-1}}{\frac{1}{N^{\beta}}\sum_j\left(\sum_i \kappa_{\Sigma}(Ry_j - Ry_i)\right)^{\beta-1}}\right]$$
$$= \frac{1}{\beta-\alpha}\left[log\frac{1}{N^{\alpha}}\sum_j\left(\sum_i \frac{1}{|R|}\kappa_{\Sigma}\left(R^{-1}(Ry_j - Ry_i)\right)\right)^{\alpha-1} - log\frac{1}{N^{\beta}}\sum_j\left(\sum_i \frac{1}{|R|}\kappa_{\Sigma}\left(R^{-1}(Ry_j - Ry_i)\right)\right)^{\beta-1}\right]$$
$$= \frac{1}{\beta-\alpha}\left[log|R|^{\alpha-1}\frac{1}{N^{\alpha}}\sum_j\left(\sum_i \kappa_{\Sigma}(y_j - y_i)\right)^{\alpha-1} - log|R|^{\beta-1}\frac{1}{N^{\beta}}\sum_j\left(\sum_i \kappa_{\Sigma}(y_j - y_i)\right)^{\beta-1}\right]$$
$$= \frac{1}{\beta-\alpha}\left[(1-\alpha)\widehat{H}_{\alpha}(Y) - (1-\beta)\widehat{H}_{\beta}(Y)\right]$$
$$\widehat{H}_{\alpha,\beta}(\overline{Y}) = \begin{Bmatrix}\widehat{H}_{\alpha}(Y), & \beta = 1\\ \widehat{H}_{\beta}(Y), & \alpha = 1\end{Bmatrix}$$

6. CONCLUSION

In this work, general properties of the entropy estimator (1) that have applications in learning systems have been addressed. The results were obtained in mathematical validations, which can be seen in theorems 5.1 and 5.4, and properties 5.2, 5.3, 5.5 under certain conditions for choosing the kernel function. The proposed entropy estimator could be the necessary building block to train supervised learning systems.

REFERENCES

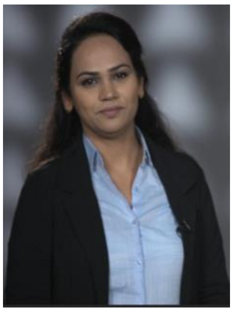
[1] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.
[2] S. Haykin, *Unsupervised adaptive filtering*. Wiley-Interscience, 2000.
[3] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. Urbana, 1964.
[4] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological sequence analysis*. Cambridge University Press, 1998, doi: 10.1017/CBO9780511790492.
[5] C. Beck and F. Schögl, *Thermodynamics of chaotic systems.* Cambridge University Press, 1993, doi: 10.1017/CBO9780511524585.
[6] A. V Lazo and P. Rathie, "On the entropy of continuous probability distributions (Corresp.)," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 120–122, Jan. 1978, doi: 10.1109/TIT.1978.1055832.
[7] N. A. Ahmed and D. V Gokhale, "Entropy expressions and their estimators for multivariate distributions," *IEEE Transactions on Information Theory*, vol. 35, no. 3, pp. 688–692, May 1989, doi: 10.1109/18.30996.
[8] E. J. Wegman and H. I. Davies, "Remarks on some recursive estimators of a probability density," *The Annals of Statistics*, vol. 7, no. 2, pp. 316–327, 1979.
[9] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. Van der Meulen, "Nonparametric entropy estimation: an overview," *International Journal of Mathematical and Statistical Sciences*, vol. 6, no. 1, pp. 17–39, 1997.
[10] I. Ahmad and Pi-Erh Lin, "A nonparametric estimation of the entropy for absolutely continuous distributions (Corresp.)," *IEEE Transactions on Information Theory*, vol. 22, no. 3, pp. 372–375, May 1976, doi: 10.1109/TIT.1976.1055550.
[11] 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, Jun. 1974, doi: 10.1137/1118083.
[12] 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, Dec. 1989, doi: 10.1007/BF00057735.
[13] J.-F. Bercher and C. Vignat, "Estimating the entropy of a signal with applications," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1687–1694, Jun. 2000, doi: 10.1109/78.845926.
[14] L. Györfi and E. C. van der Meulen, "Density-free convergence properties of various estimators of entropy," *Computational Statistics & Data Analysis*, vol. 5, no. 4, pp. 425–436, Sep. 1987, doi: 10.1016/0167-9473(87)90065-X.
[15] L. Györfi and E. C. Meulen, "On the nonparametric estimation of the entropy functional," in *Nonparametric Functional Estimation and Related Topics*, Dordrecht: Springer Netherlands, 1991, pp. 81–95, doi: 10.1007/978-94-011-3222-0_6.
[16] L. Györfi and E. C. van der Meulen, "An entropy estimate based on a Kernel density estimation," *Limit Theorems in Probability and Statistics*, pp. 229–240, 1989.
[17] A. V Ivanov and M. N. Rozhkova, "Properties of the statistical estimate of the entropy of a random vector with a probability density," *Prob. Info. Trans.*, vol. 17, no. 3, pp. 171–177, 1982.
[18] 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, Jun. 1985, doi: 10.1016/0047-259X(85)90023-5.
[19] P. Hall, "Limit theorems for sums of general functions of m-spacings," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Nov. 1984, vol. 96, no. 3, pp. 517–532, doi: 10.1017/S0305004100062459.
[20] 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," in *Proceedings of the IEEE*, 1968, vol. 56, no. 11, pp. 2052–2053, doi: 10.1109/PROC.1968.6784.
[21] L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," *Problemy Peredachi Informatsii*, vol. 23, no. 2, pp. 9–16, 1987.
[22] A. B. Tsybakov and E. C. Van der Meulen, "Root-n consistent estimators of entropy for densities with unbounded support," *Scandinavian Journal of Statistics*, pp. 75–83, 1996.
[23] 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*, pp. 185–214, 1983.
[24] 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, Aug. 2022, doi: 10.1007/s00202-021-01447-7.

[25] E. Parzen, "On estimation of a probability density function and mode," *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
[26] J. N. Kapur, "Generalized entropy of order α and type β," *The Math. Seminar*, vol. 4, pp. 78–82, 1967.
[27] L. Chawla, V. Kumar, and A. Saxena, "Kernel density estimation of Tsalli's entropy with applications in adaptive system training," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 2, pp. 2247–2253, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp2247-2253.

BIOGRAPHIES OF AUTHORS

Leena Chawla is a research scholar in the 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 for graduates and postgraduates. She has more than 4 research articles published in journals. She was awarded her M.Phil. in 2010 from Chaudhary Devi Lal University, Sirsa, Haryana, India. She can be contacted at leenabharat14@gmail.com.

Vijay Kumar is working as a professor of Mathematics, Manav Rachna International Institute of Research and Studies, Faridabad Haryana INDIA. His research interests include 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 on events of national and international reputation. Prof. Vijay Kumar serves as an editorial board member for several international/national journals and a reviewer for many national and international journals. He can be contacted at drvijaykumarsudan@gmail.com.



Arti Saxena is working as an associate professor in the 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 five research books published in her account and nine chapters in edited books. She got a young scientist award at 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 artisaxena.set@mriu.edu.in.