On performance of multivariate product kernel density estimation consistency

Ejakpovi S. U.[1]* and Siloko I. U.[2]

[1]College of Education, Warri, Department of Mathematics, P. M. B. 1251, Delta State, Nigeria.
[2]Department of Mathematics and Computer Science, Edo University Iyamho, P. M. B. 04, Auchi, Nigeria.
*Corresponding author. E-mail: sisraelgreat@gmail.com, siloko.israel@edouniversity.edu.ng.

Accepted 6th October, 2020

This paper examines the consistency that surrounds the Multivariate Quiweight product kernel function in achieving a better curve estimate and the target density. In revealing its features in the area of Multivariate Kernel Density Estimation of data, we adopted the Multivariate KDE product estimator to ascertain its generalized consistency expressions. This was done via the Taylor series expansion for all its higher order forms of m dimensions. The expressions that were established are subject to real life data simulations of different sample sizes based on the dimensions that emanated from the Multivariate Quiweight product kernel functions; the data analysis showed that it has a smaller bias and variance which influences its global performance as the Multivariate Quiweight product kernel function order and m dimensions increase.

Key words: AMISE; Beta Kernel, Consistency, Smoothing Parameter, Pentaweight, Product Kernel.

INTRODUCTION

The theoretical emphasis of kernel density estimation (KDE) focused on the various areas of estimating features that are related to the kernel functions in decades of its inception (Rosenblatt, 1956). In recent research studies, it has become the hub of nonparametric density estimations due to the giant strides applications in Biological Sciences, Econometric, Statistical Engineering, Statistical modeling, kernel regressions etc. The estimator exists in both univariate and multivariate forms with distinguished factors of the smoothing parameter, sample size compositions and the kernel functions, wherein the smoothing parameter is a unifying factor that determines the contour of the data in both forms (Silverman, 1986). The implication is that, the univariate form is an extension to several dimensions to birth the multivariate kernel form which ranges from the Bivariate; Trivariate; Quadrivariate etc dimensional data for visualization projections (Scott, 1992).

The multivariate density estimation has helped to reduce the problem of curse-of-dimensionality in the one dimensional univariate kernel density estimation whose statistical properties deteriorate very fast especially in higher order dimensions (Peracchi, 2014). Let $\{Z_i\}_{i=1}^{n}$ be a sample from the independent and identical distribution of a random m-vector Z with density function $f(Z) = f(Z_1, \dots, Z_m)$, then the multivariate generalization of the univariate kernel density estimation is given as:

$$\hat{f}(Z) = \frac{1}{nh_1, \dots, h_m} \sum_{i=1}^{n} K_m \left(\frac{Z_1 - Z_{i1}}{h_1}, \dots, \frac{Z_m - Z_{im}}{h_m} \right) \tag{1}$$

where $K_m: \mathbb{R}^m \to \mathbb{R}$ is a multivariate kernel function and h is called the smoothing parameter, Z_i is a sequence of continuous random m-vector ,which has a continuously differentiable density f twice (Deheuvels, 1977; Hensen, 2003; Doung and Hazelton, 2003). This multivariate kernel density estimator in compact form is expressed as:

$$\hat{f}(Z) = \frac{1}{nh_n^d}\sum_{i=1}^{n} K_m\left(\frac{Z_m - Z_{im}}{h_n}\right) \quad (2)$$

where the sequence of the smoothing parameter $\{h_m\}$ with the multivariate kernel $K_m: \mathbb{R}^m \to \mathbb{R}$ satisfies the following axioms:

$$\begin{cases} h_n \to 0 \\ nh_n^m \to \infty \\ \int K_m(u)du = 1, \int uK_m(u)du = 0 \text{ } and \int uu^T K_m(u)du = \mu_2(K)I_d \end{cases} \quad (3)$$

And the I_d denotes the dXd identity matrix while the multivariate kernel norm is $\|K_m\|_2^2 = \int K_m^2(u)du$ which adopts the smoothing parameter parameterization $H = h^2 I_d$ (Oyegue and Ogbonmwan, 2014). The estimate of $\hat{f}(Z)$ is measured by the asymptotic mean integrated squared error (AMISE) using the Taylor's series expansion that is composed of the asymptotic integrated squared bias (AISB) and the asymptotic integrated variance (AIV) given as:

$$\begin{cases} AIV = \dfrac{R(K)}{nh_n^d} \\ AISB = \dfrac{h_n^4}{(2!)^2}\mu_2(K)^2 R(f_{XX}) \end{cases} \quad (4)$$

where $R(K)$ represents the roughness of the multivariate kernel function, $\mu_2(K)^2$ is the variance of the kernel function and $R(f_{xx})$ is the roughness of the unknown probability density function (Guidoum, 2015; Siloko et al. 2019). The conjugation of both terms in Equation (4) will produce the estimate of the asymptotic mean integrated squared error (AMISE). This leads to a closed solution form when the AMISE takes the objective function whose minimization with respect to the smoothing parameter (h_n) will yield optimal smoothing parameter expressed as:

$$h_{opt} = \left(\frac{dR(K)}{n\mu_2(K)^2 \int f_{XX}(X)^2 dX}\right)^{\left(\frac{1}{4+d}\right)} \quad (5)$$

The optimal smoothing parameter in Equation 5 decreases at order $(-1/4 + d)$ and the optimal AMISE decreases at order $(-4/4 + d)$.

The Multivariate Product Kernel Functions

Two approaches are used for the transformation of a univariate kernel function into a multivariate kernel functions: Product and Radial multivariate kernel (Scott, 1992; Wand and Jones, 1995); the Multivariate product kernel density approach is given as:

$$\begin{cases} \hat{f}(Z) = \dfrac{1}{nh^m}\sum_{i=1}^{n} K_m\left(\dfrac{Z_1 - Z_{i1}}{h}, \dots, \dfrac{Z_m - Z_{im}}{h}\right) \\ \hat{f}(Z) = \dfrac{1}{nh^m}\sum_{i=1}^{n}\left\{\prod_{j=1}^{m} K\left(\dfrac{Z_j - Z_{ij}}{h}\right)\right\} \end{cases} \quad (6)$$

In generating the univariate classical kernel functions and subsequent higher order kernel forms we adopt the modified construction rule of Ejakpovi et al. (2019), where the kernel order denoted by α takes 3, 4, 5 and 6 called Triweight, Quadriweight, Quiweight and Hexweight, respectively. And, their generalized univariate higher order kernel function of the Multivariate product is formed, as:

$$\begin{cases} K(t) = \begin{cases} \frac{35}{32}(1-t^2)^3, |t| \le 1 \\ 0, otherwise \end{cases} \\ K(t) = \begin{cases} \frac{315}{256}(1-t^2)^4, |t| \le 1 \\ 0, otherwise \end{cases} \\ K(t) = \begin{cases} \frac{693}{512}(1-t^2)^5, |t| \le 1 \\ 0, otherwise \end{cases} \\ K(t) = \begin{cases} \frac{3003}{2048}(1-t^2)^6, |t| \le 1 \\ 0, otherwise \end{cases} \end{cases} \quad (7)$$

And

$$K_{m\alpha}^{p}(t) = \begin{cases} \left(\frac{\left(\frac{1}{2}\right)_{\alpha+1}}{2\alpha!}\right)^{m} \prod_{i=1}^{m}\left((1-t_i^2)^{\alpha-1}(3-\{3+2\alpha\}t_i^2)\right), |t_i| \leq 1 \\ 0, otherwise \end{cases} \quad (8)$$

where $K_{m\alpha}^{p}(t)$ denotes the m-dimensions of the Multivariate product kernel function for any classical kernel function using Equation 7 and $\left(\frac{1}{2}\right)_{\alpha+1}$ is the normalization factor.

Generalized L_2-consistency asymptotic expressions

The consistency theory of the Multivariate estimator helps to achieve the discrepancy criterion that gives the measure of performance of the estimator in Equation 2. The Mathematical tractability of the generalized consistency asymptotic expressions seeks the Taylor's series expansion of orders $(2j+2)!$ for the asymptotic integrated squared bias and asymptotic integrated variance of the Multivariate product estimator of Equation 6.

Its combination yields the generalized asymptotic mean integrated squared error (AMISE) expression. The minimization when the AMISE is considered as the objective function with respect to the smoothing parameter gives a differential equation whose closed solution is the optimal asymptotic smoothing parameter for the Multivariate product kernel functions. When the asymptotic smoothing parameter whose value is plugged into the $AMISE\left(\hat{f}(.)\right)$ expression to produce the generalized L2-consistency of the Multivariate product estimator to any m dimensions with the best convergence rate of the $AMISE\left(\hat{f}(.,h)\right)$ is $O\left(n^{-(4j+4)/_{m+4j+4}}\right)$, then the mean of global discrepancy of the Multivariate product estimator is expressed as:

$$AMISE = \delta\left(\frac{\left(K_{(2j+2)}^{p}.I_m\right)^2(4j+4)\left(\left\|\gamma^{2j+2}f_{XX}\right\|_2^2\right)}{\left((2j+2)!\right)^2}\right)^{\frac{m}{m+4j+4}}\left(\frac{m\left\|K_{mQ}^{p}(t)\right\|_2^2}{n}\right)^{\frac{4j+4}{m+4j+4}} \quad (9)$$

where δ is an arbitrarily quotient factor of the Multivariate Quiweight kernel function in Equation 8 when $\alpha = 5$ for m dimensions; $K_{(2j+2)}^{p} = \int\sum_{i=1}^{m} t_i^{2j+2} K_{mQ}^{p}(t)dt$; $\left\|\mathrm{K}_{\mathrm{mQ}}^{\mathrm{p}}(\mathrm{t})\right\|_2^2 = \int\left(\mathrm{K}_{\mathrm{mQ}}^{\mathrm{p}}(\mathrm{t})\right)^2 \mathrm{dt}$ and $\boldsymbol{\gamma}^{\mathbf{2j+2}}f_{XX} = \sum_{i=1}^{m}\frac{\partial^{2j+2}}{\partial x_i}f(X)$ which is achieved from the data distribution.

Data implementation

Here, we shall examine the performance of the generalized Multivariate product estimator for the Multivariate Quiweight product kernel function whose Mathematical expressions are in section 3. Bivariate data of a sample size of two hundred and were used for the simulations; they represent skin elasticity measurement and Strength Data for Polymerization Process Study (Mason, Gunst and Hess, 2003) under a Mathematical software using Equation 9:

1-Dimension	2-Dimesion	3-Dimension	4-Dimension	5-Dimension	6-Dimension
0.0000165	0.0003120	0.0033254	0.0139049	0.120404	0.48589
0.0000135	0.0002773	0.0036886	0.0194347	0.250790	1.43582
0.0000110	0.0002164	0.0030267	0.0171614	0.268503	1.84114
0.0000091	0.0001473	0.0016481	0.0072556	0.081755	0.25602
0.0000050	0.0001023	0.0012346	0.0051621	0.045692	0.17215

DISCUSSIONS

The results displayed in data implementations vividly indicate the performance of the Multivariate Quiweight product kernel function on the bivariate data based on the homoscedascity of the data. The simulations showed that as the dimensions increase, the Multivariate Quiweight product functions performances decrease, except in dimensions five and six when their estimates tend to increase at some orders. The implications of this to the data is that in attempting to get a closer estimate of Skin Elasticity and Strength Data for polymerization process, the Multivariate Quiweight product kernel unction dimension one and two gave a better estimate of the Skin Elasticity measurement and Strength Data for polymerization process. This is because their estimates have a minimum discrepancy values from their true or underlying density. Though dimensions three, four, five and six also project the estimates but with little higher values compared to dimensions one and two. This implies that the Multivariate Quiweight product kernel function estimates with little higher values which clearly indicate that the Skin Elasticity measurement and Strength Data for polymerization process cannot be truly relied on. This is due to breakage or inelasticity of Skin or inadequate Strength of Data for polymerization process at these dimensions. These now give credence to dimensions one and two.

Conclusion

This paper examines the consistency theory of the Multivariate Quiweight product kernel function at generalized perspectives; its performance helps in determining the level of dimensions for better estimates of Skin Elasticity measurement and Strength Data for polymerization process as shown in the data visualization. Then, it becomes obvious that this consistency theory should be applied to areas of reliability study of events, possible forecasting of events in the field of regressions and otherwise. Also, the need to extend this generalized L_2-Consistency theory to the Multivariate Quiweight radial kernel function and its applications to real life situations is a grey area for further research study.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

Cacoullos, T. (1966). Estimation of a Multivariate Density. Annals of the Institute of Statistical Mathematics 18, 179 – 189.

Deheuvels, P. (1977). Estimation Nonparametriani de la densite par histogrammes generalizes II'. Publications de l'insitute de statistique de l'universite de Paris. 22, 1 – 25. Derivatives. The kedd R package version 1.0.3. Department of Probabilities and Statistics. University of Science and Technology, Houari Boumediene, Algeria.

Duong, T and Hazelton, M. L. (2003). Plug-In Bandwidth Matrices for Bivariate Kernel

Density Estimation. Nonparametric Statistics. 15 (1):17–30.

Ejakpovi, S. U., Ozobokeme, J. K and Uveruveh, F.O (2019). Generalized Efficiency of Kernel Density Derivative Estimation (KDDE): A Univariate Perspective. Abacus; Journal of the Mathematical Association of Nigeria, 46 (1): 460-464

Guidoum, A. C. (2015). Kernel Estimator and Bandwidth Selection for Density and its Derivatives: The kedd package, version 1.0.3. Department of Probabilities and Statistics, University of Science and Technology, Houari Boumediene, Algeria.

Hensen, B. E (2003). Exact mean integrated squared error for higher order kernel estimators. University of Wisconsin, Madeson. W153706, USA.

Mason, R. L, Gunst, R.F and Hess, J. L (2003). Statistical Design and Analysis of Experiments with Applications to Engineering and Sciences. Second Edition. Wiley, New York.

Oyegue, F. O and Ogbonmwan, S.M (2014). The Efficiency of Product Multivariate Kernels. Proceedings of the World Congress on Engineering and Computer Science 2014, Volume II. San Francisco, USA

Peracchi, F (2014). Applied Nonparametric Methods. http//www.nonparametricmethods.pdf

Rosenblatt, M (1956). Remarks on some nonparametric estimates of a Density function. Annals of Mathematical Sciences, 27, 832-837

Scott, D. W. (1992). Multivariate Density Estimation. Theory, Practice and Visualisation. Wiley, New York.

Siloko, I.U., Ikpotokin, O and Siloko, E.A (2019). A Key Note on Perforance of Smoothing Parameterizations in Kernel Density Estimation. Tanzania Journal of Science, 45 (1): 1-8

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Wand, M. P. and Jones, M. C. (1995). Kernel Smoothing. Chapman and Hall, London.