



On the convergence rate of d -dimensional fourth-order beta polynomial kernels

Benson Ade Eniola Afere*

Prince Abubakar Audu University, Anyigba, and 272102, Nigeria

ARTICLE INFO

Article history:

Received: 04 September 2023

Received in revised form: 25 November 2023

Accepted: 27 November 2023

Available online: 29 November 2023

Keywords: Kernel density estimation, d -dimensional fourth-order kernels, beta polynomial kernels, AMISE, rate of convergence

DOI: 10.61298/rans.2023.1.2.29

ABSTRACT

This article focuses on formulating the Asymptotic Mean Integrated Squared Error (AMISE) scheme for d -dimensional fourth-order beta polynomial kernels in the context of kernel density estimation. The primary objective is to assess how this scheme influences the convergence rate, which directly impacts the speed at which the estimated density converges to the true density, contributing to bias reduction. The article employs AMISE as a metric to quantify the overall dissimilarity between the estimated density \hat{f} and the true density f . Quantitatively, the proposed convergence scheme is compared to existing ones by Deheuvels and Jones *et al.* across various scenarios of sample sizes and dimensionalities. The study's findings provide compelling evidence that fourth-order beta polynomial kernels exhibit a significantly faster convergence rate compared to the rates documented in the literature. This accelerated convergence rate implies a substantial improvement in bias reduction capabilities. The results underscore the potential effectiveness of fourth-order beta polynomial kernels as a powerful tool for enhancing the accuracy of kernel density estimation tasks.

© 2023 The Author(s). Production and Hosting by FLAYOO Publishing House LTD on Behalf of the Nigerian Society of Physical Sciences (NSPS). Peer review under the responsibility of NSPS. This is an open access article under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. INTRODUCTION

Density estimation, a fundamental aspect of statistical analysis, involves deriving insights into probability density functions based on observed data [1]. Two primary approaches in this field are parametric and nonparametric methods. Nonparametric methods, particularly kernel density estimation (KDE), have gained prominence for their versatility in avoiding assumptions about specific density forms and their applicability across various scenarios, including multivariate analysis [2].

Significant research spanning from the 1950s onwards has focused on optimizing bandwidth selection for univariate KDE [3–5]. The test graph method introduced by Silverman [6], characterized by swift convergence and consistent performance, has been influential. Scott and Factor [7] explored the bandwidth's impact on bias, while Rudemo [8] focused on minimizing the mean integrated squared error (MISE) to identify optimal bandwidth choices.

Multivariate Kernel Density Estimation (MKDE) has witnessed noteworthy contributions, encompassing diverse parameterizations and optimal bandwidths for bivariate density estimation, as outlined by Wand and Jones [9]. Adaptive techniques, tailoring smoothing based on local data features, have emerged through the work of Bowman and Foster [10]. Wand and Jones [2] extended KDE to higher dimensions, deriving asymptotic and exact mean integrated squared error results. Duong and Hazelton [11] explored plug-in bandwidth matrices for bivariate estimation.

To address potential bias, researchers have explored higher-

*Corresponding author: Tel.: +234-913-5398-886

e-mail: afere.bae@ksu.edu.ng (Benson Ade Eniola Afere)

order kernels in KDE [12, 13]. Additionally, investigations into convergence rates and optimal bandwidth performance have been conducted by Wand and Jones [2], Jones and Signorini [13], Chacón [14], Duong and Hazelton [15], Hall and Marron [16], Hall *et al.* [17], Kimari *et al.* [18], and Sheather and Jones [19].

Recent research by Siloko *et al.* [20] has focused on reducing the AMISE in KDE. Their exploration involved techniques related to kernel density derivatives and kernel boosting, revealing that an increase in kernel derivatives and the number of boosting steps can effectively decrease AMISE. KDE plays a pivotal role in exploratory data analysis and statistical visualization due to its simplicity, interpretability, and wide applicability, making it a valuable tool for data analysis. Ongoing endeavors aim to develop generalized convergence strategies that reduce AMISE while preserving the statistical characteristics of real-world observations.

Within the context of KDE, for a given random sample X_1, X_2, \dots, X_n drawn from a common density f , the d -dimensional kernel density estimator is defined as follows:

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|} \sum_{i=1}^n K(\mathbf{H}(\mathbf{x} - \mathbf{X}_i)). \quad (1)$$

Here, $x = (x_1, x_2, \dots, x_d)^T$ and $X_i = (X_{i1}, X_{i2}, \dots, X_{id})^T$, $i = 1, \dots, n$. \mathbf{H} is a symmetric positive definite $d \times d$ matrix called the bandwidth matrix, and K is a d -variate kernel function satisfying $\int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} = 1$. In this article, we adopt the parameterizations $\mathbf{H} = h^2\mathbf{I}_d$ proposed by Cacoullos [21] for selecting the bandwidths. This choice allows for closed-form expressions for both the optimal bandwidth and AMISE [17–19, 22, 23]. The d -dimensional kernel density estimator, using the chosen parameterisations as discussed by Wand and Jones [2], can be expressed as follows:

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right). \quad (2)$$

The central objective of this study is to explore and elucidate the convergence rates within the domain of kernel density estimation. Convergence rates play a pivotal role in delineating how rapidly the estimated density, denoted as $\hat{f}(\mathbf{x})$, approaches the true density, represented by $f(\mathbf{x})$. This exploration is particularly significant as it extends observations from previous research conducted in univariate settings by Sliverman [1] and Afere [12] to the more complex and intricate landscape of higher-order d -dimensional kernels.

The phenomenon of convergence rates serves as a crucial indicator, offering insights into the efficiency and speed of the estimation process. Faster convergence rates imply a swifter approximation of the estimated density to the underlying true density. Such insights are pivotal in understanding and improving the accuracy of kernel density estimation, especially when dealing with multivariate scenarios.

The subsequent sections of this article delve into a detailed examination of the convergence properties and behavior specifically related to fourth-order multivariate kernels. Building upon the well-established convergence schemes for second-order kernels documented in the existing literature [1, 2], our focus pivots towards the nuances and intricacies associated with fourth-order kernels. This shift in focus is motivated by the desire to

enhance our understanding of convergence dynamics in higher-dimensional spaces.

The structure of this work unfolds as follows: Section 2 lays the foundation for this article, discussing the framework and convergence schemes for fourth-order multivariate kernels. Furthermore, Section 3 is dedicated to the derivation of a generalized fourth-order convergence scheme for AMISE. This involves a rigorous mathematical exploration, providing explicit formulas that capture the essence of the convergence behavior in this specific context.

In Section 4, we delve into the detailed quantitative validation of the outcomes derived in our study. This section is instrumental in providing a rigorous assessment of the reliability, accuracy, and applicability of the proposed convergence schemes for fourth-order multivariate kernels. The validation process involves a meticulous comparison with the outcomes reported by Deheuvels (see Scott [24] pg 189) and Jones *et al.* (see Chacón and Duong [25] pg 70), encompassing various sample sizes and dimensionalities.

Finally, in Section 5, we embark on a meticulous examination of the results obtained from our study. This examination critically assesses the implications and importance of the observed convergence rates in kernel density estimation. Conclusions drawn from this analysis contribute to a broader understanding of the efficiency of the proposed convergence rate.

2. CONVERGENCE SCHEMES FOR THE FOURTH-ORDER KERNELS

In the landscape of this research domain, it is a prevalent practice to gauge the effectiveness of the density function f by employing the Mean Integrated Squared Error (MISE). However, the derivation of MISE in closed form is often unattainable, necessitating the adoption of an asymptotic approximation as a pragmatic solution. This approximation serves as an estimate for the MISE, providing a means to evaluate the performance of the density function under consideration. The asymptotic approximation takes the following form:

$$\text{AMISE}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \text{AISB}\hat{f}_{\mathbf{H}}(\mathbf{x}) + \text{AIV}\hat{f}_{\mathbf{H}}(\mathbf{x}), \quad (3)$$

where

$$\text{AISB}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \int_{\mathbb{R}^d} \text{bias}^2\hat{f}_{\mathbf{H}}(\mathbf{x})d\mathbf{x}, \quad (4)$$

and

$$\text{AIV}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \int_{\mathbb{R}^d} \text{var}\hat{f}_{\mathbf{H}}(\mathbf{x})d\mathbf{x}. \quad (5)$$

The basic assumptions for any multivariate second-order symmetric kernels are:

$$\left. \begin{array}{l} i. \int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} = 1 \\ ii. \int_{\mathbb{R}^d} \mathbf{t}K(\mathbf{t})d\mathbf{t} = 0_d \\ iii. \int_{\mathbb{R}^d} (\mathbf{t}^T\mathbf{t})K(\mathbf{t})d\mathbf{t} = K_2\mathbf{I}_d \neq 0. \end{array} \right\} \quad (6)$$

However, in 1962, Parzen [26] introduced the concept of using kernels that can take both negative and positive values. Consequently, for the purpose of achieving a fourth-order estimation,

we relax the assumptions outlined in Eq.(6) and obtain the following expression:

$$\left. \begin{aligned} i. & \int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} = 1 \\ ii. & \int_{\mathbb{R}^d} (\mathbf{t}^T \mathbf{t})K(\mathbf{t})d\mathbf{t} = 0_d \\ iii. & \int_{\mathbb{R}^d} (\mathbf{t}^T \mathbf{t})^2 K(\mathbf{t})d\mathbf{t} = K_4 \mathbf{I}_d \neq 0. \end{aligned} \right\} \quad (7)$$

Now, if the Taylor series expansion of $f(\mathbf{x} - \mathbf{t}\mathbf{H}^{\frac{1}{2}})$ up to the fifth term is used, and on application of the assumptions in Eq.(6) with the parametrisation $\mathbf{H} = h^2 \mathbf{I}_d$, the AMISE in Eq.(3) results to:

$$AMISE_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = (4!)^{-2} h^8 (K_4 \mathbf{I}_d)^2 \|\nabla^4 f(\mathbf{x})\|_2^2 + (nh^d)^{-1} \|K\|_2^2. \quad (8)$$

On differentiating Eq.(8) with respect to h and minimizing accordingly, we obtain h as:

$$h_{AMISE_{\hat{f}_{\mathbf{H}}}} = ((4!)^2 d 8^{-1} \|K\|_2^2 (K_4 \mathbf{I}_d)^{-2} \times (\|\nabla^4 f(\mathbf{x})\|_2^2)^{-1})^{\frac{1}{d+8}} n^{-\frac{1}{d+8}}. \quad (9)$$

Substituting Eq.(9) into Eq.(8), we have:

$$AMISE_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \frac{d+8}{8d} \left(\frac{8}{(4!)^2}\right)^{\frac{d}{d+8}} (d(K_4 \mathbf{I}_d)^2 \times \|\nabla^4 f(\mathbf{x})\|_2^2 \|K\|_2^2)^{\frac{8}{d+8}} n^{-\frac{8}{d+8}}. \quad (10)$$

The main significance from this is that the convergence rate $n^{-\frac{4}{d+4}}$ as proposed by Deheuvels in 1977 (see Scott [24] pg 189) decreases to $n^{-\frac{8}{d+8}}$ proposed by Jones *et al* when kernels meeting the conditions in Eq.(7) are applied (see Chacón and Duong [25] pg 70). Section 3 below presents our theorem which is an improvement to the result in Eq.(10) if fourth-order kernels is used.

3. THE PROPOSED GENERALIZED CONVERGENCE SCHEME

In this section, we introduce a comprehensive framework that builds upon the advancements made in Eqs.(9) and (10). We present Theorem 1, which encapsulates the enhanced outcomes achieved through this scheme.

Theorem 1. *If $K(\mathbf{t})$ is any differentiable d -dimensional fourth-order kernel that is parameterized by $\mathbf{H} = h^2 \mathbf{I}_d$ satisfying the conditions in Eq.(7), and $K_{2m+2} \mathbf{I}_d$ represents the $(2m+2)^{th}$ -moment of this kernel, with $\|\nabla^{2m+2} f(\mathbf{x})\|_2^2$ being the generalized L_2 -norm of d -dimensional Gaussian distribution with mean $\mu = (\mu_1, \mu_2, \dots, \mu_d)^T$, then the generalized asymptotic mean integrated squared error is given as:*

$$AMISE^{2m} \cong \left(\frac{d+4m+4}{d(4m+4)}\right) \left(\frac{4m+4}{((2m+2)!)^2}\right)^{\frac{d}{d+4m+4}} (d \|K\|_2^2)^{\frac{4m+4}{d+4m+4}} \times ((K_{2m+2} \cdot \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} (\|\nabla^{2m+2} f(\mathbf{x})\|_2^2)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}.$$

Proof. The bias term is given by:

$$bias_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = E\hat{f}_{\mathbf{H}}(\mathbf{x}) - f(\mathbf{x}). \quad (11)$$

If we substitute Eq.(2) into Eq.(11) and simplify, we have:

$$bias_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \int_{\mathbb{R}^d} K(\mathbf{t})f(\mathbf{x} - \mathbf{t}\mathbf{H}^{\frac{1}{2}})d\mathbf{t} - f(\mathbf{x}). \quad (12)$$

Now, on using the multivariate Taylor’s series expansion up to $(m+1)^{st}$ -times on Eq.(12), we have:

$$\begin{aligned} bias_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) &= \int_{\mathbb{R}^d} K(\mathbf{t})(f(\mathbf{x}) - \text{tr}\{\mathbf{t}^T D_f \mathbf{H}^{\frac{1}{2}}\} + \\ &\frac{1}{2!} \text{tr}\{(\mathbf{t}^T \mathbf{t}) \mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}} - \frac{1}{3!} \text{tr}\{\mathbf{t}^T D_f \mathbf{H}^{\frac{1}{2}} \times \\ &(\text{tr}\{(\mathbf{t}^T \mathbf{t}) \mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}}\})\} + \dots - \frac{1}{(2m+1)!} \text{tr}\{\mathbf{t}^T D_f \mathbf{H}^{\frac{1}{2}} \times \\ &\underbrace{[(\mathbf{t}^T \mathbf{t}) \mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}}] \dots (\mathbf{t}^T \mathbf{t}) \mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}}]}_{m \text{ times}}) + \frac{1}{(2m+2)!} \text{tr}\{\mathbf{t}^T [\\ &\underbrace{(\mathbf{t}^T \mathbf{t}) \mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}}] \dots (\mathbf{t}^T \mathbf{t}) \mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}}]}_{(m+1) \text{ times}}) - \dots) d\mathbf{t} - f(\mathbf{x}), \end{aligned}$$

where, $D_f(\mathbf{x})$ is the vector of first-order partial derivatives of f and $\mathcal{H}_f(\mathbf{x})$ is the Hessian matrix of the $d \times d$ matrix having (i, j) entry equal to $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = 0, i \neq j$.

On expanding the last equation further, we have:

$$\begin{aligned} Bias_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) &= f(\mathbf{x}) \int_{\mathbb{R}^d} K(\mathbf{t})d\mathbf{t} - \text{tr}\{(\int_{\mathbb{R}^d} \mathbf{t}^T K(\mathbf{t})d\mathbf{t})(D_f \mathbf{H}^{\frac{1}{2}})\} \\ &+ \frac{1}{2!} \text{tr}\{(\int_{\mathbb{R}^d} (\mathbf{t}^T \mathbf{t})K(\mathbf{t})d\mathbf{t})(\mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}})\} \\ &- \frac{1}{3!} \text{tr}\{(\int_{\mathbb{R}^d} \mathbf{t}^T (\mathbf{t}^T \mathbf{t})K(\mathbf{t})d\mathbf{t})(D_f \mathbf{H}^{\frac{1}{2}})(\mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}})\} + \dots \\ &- \frac{1}{(2m+1)!} \text{tr}\{(\int_{\mathbb{R}^d} \mathbf{t}^T [(\mathbf{t}^T \mathbf{t}) \dots (\mathbf{t}^T \mathbf{t})]K(\mathbf{t})d\mathbf{t}) \times \\ &(D_f \mathbf{H}^{\frac{1}{2}}) \underbrace{[(\mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}}) \dots (\mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}})]}_{(m \text{ times})}\} + \\ &\frac{1}{(2m+2)!} \text{tr}\{(\int_{\mathbb{R}^d} [(\mathbf{t}^T \mathbf{t}) \dots (\mathbf{t}^T \mathbf{t})](\mathbf{t}^T \mathbf{t})K(\mathbf{t})d\mathbf{t}) \times \\ &\underbrace{[(\mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}}) \dots (\mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}})]}_{(m+1) \text{ times}}) - \dots + \dots - f(\mathbf{x}). \end{aligned} \quad (13)$$

Imposition of the moment conditions of Eq.(7) and the $(2m+2)$ -th moment of the statement of the theorem in Eq.(13), Eq.(13) reduces to:

$$Bias_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \frac{1}{(2m+2)!} \text{tr}\{(\int_{\mathbb{R}^d} (\mathbf{t}^T \mathbf{t})^{m+1} K(\mathbf{t})d\mathbf{t}) \times (\mathbf{H}^{\frac{1}{2}} \mathcal{H}_f \mathbf{H}^{\frac{1}{2}})^{m+1}\} + o((\mathbf{t}^T \mathbf{t}) \mathbf{H}^{\frac{1}{2}})^{m+1}. \quad (14)$$

The trace of the Hessian matrix $(\mathcal{H}_f(\mathbf{x}))$ in Eq.(14) is defined by $\nabla^2 f(\mathbf{x})$. With this definition, the bias can be expressed as:

$$Bias_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \frac{1}{(2m+2)!} \int_{\mathbb{R}^d} (\mathbf{t}^T \mathbf{t})^{m+1} K(\mathbf{t})d\mathbf{t} \times [\nabla^2 f(\mathbf{x})]^{m+1} [\text{tr}(\mathbf{H})]^{m+1} + o((\mathbf{t}^T \mathbf{t}) \mathbf{H})^{m+1}. \quad (15)$$

Squaring both sides of Eq.(15) and using the bandwidth matrix $\mathbf{H} = h^2 \mathbf{I}_d$ and then substituting into Eq.(4), the AISB is obtained as:

$$AISB_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \frac{1}{((2m+2)!)^2} \left[\int_{\mathbb{R}^d} (\mathbf{t}^T \mathbf{t})^{m+1} K(\mathbf{t})d\mathbf{t} \right]^2 \times [\nabla^{2m+2} f(\mathbf{x})]^2 [h^{4m+4}]. \quad (16)$$

Also, on using the $(2m + 2)$ -th moment in the statement of the Theorem, Eq.(16) becomes:

$$\text{AISB}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \frac{1}{((2m + 2)!)^2} (K_{2m+2} \cdot \mathbf{I}_d)^2 \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^2 h^{4m+4}. \quad (17)$$

The variance term is given by:

$$\text{Var}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = E^2 \hat{f}_{\mathbf{H}}(\mathbf{x}) - [E \hat{f}_{\mathbf{H}}(\mathbf{x})]^2. \quad (18)$$

On substituting Eq.(18) into Eq.(5) and using the necessary assumptions as in the case of bias, we have:

$$\text{Var}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = (n\mathbf{H}^{\frac{1}{2}})^{-1} \int_{\mathbb{R}^d} K^2(\mathbf{t}) f(\mathbf{x} - \mathbf{t}\mathbf{H}^{\frac{1}{2}}) d\mathbf{t} - ((n\mathbf{H}^{\frac{1}{2}})^{-1} \int_{\mathbb{R}^d} K(\mathbf{t}) f(\mathbf{x} - \mathbf{t}\mathbf{H}^{\frac{1}{2}}) d\mathbf{t})^2. \quad (19)$$

Hence, following similar algebraic variable substitution and Taylor series expansion argument as in the case of bias and using all the necessary assumptions, Eq.(19) becomes:

$$\text{Var}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = (n|\mathbf{H}|^{\frac{1}{2}}) \|K\|_2^2 f(\mathbf{x}) + o(n|\mathbf{H}|^{\frac{1}{2}})^{-1}, \quad (20)$$

where $\|K\|_2^2 = \int_{\mathbb{R}^d} K^2(\mathbf{t}) d\mathbf{t}$ is the d -dimensional L_2 -norm of $K(\mathbf{t})$. On substituting Eq.(20) into Eq.(5), the resulting equation becomes:

$$\text{AIV}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \int_{\mathbb{R}^d} \text{Var}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) d\mathbf{x} = (n\mathbf{H}^{\frac{1}{2}})^{-1} \|K\|_2^2. \quad (21)$$

Also, on using the parametrisation $\mathbf{H} = h^2 \mathbf{I}_d$, Eq.(21) becomes:

$$\text{AIV}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = (nh^d)^{-1} \|K\|_2^2 \quad (22)$$

Plugging back Eqs.(17) and (22) into Eq.(3), we have:

$$\text{AMISE}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) = \frac{h^{4m+4}}{((2m + 2)!)^2} [K_{2m+2} \cdot \mathbf{I}_d]^2 \times \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 + (nh^d)^{-1} \|K\|_2^2 \right). \quad (23)$$

On differentiating Eq.(23) with respect to h , we have:

$$\frac{\partial \text{AMISE}_{\hat{f}_{\mathbf{H}}}(\mathbf{x})}{\partial h} = (4m + 4) \frac{h^{4m+3}}{((2m + 2)!)^2} [K_{2m+2} \mathbf{I}_d]^2 \times \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 - dn^{-1} h^{-d-1} \|K\|_2^2 \right).$$

But at the minimum or maximum point, $\frac{\partial \text{AMISE}_{\hat{f}_{\mathbf{H}}}(\mathbf{x})}{\partial h} = 0$. Therefore,

$$\begin{aligned} (4m + 4) \frac{h^{4m+3}}{((2m + 2)!)^2} [K_{2m+2} \mathbf{I}_d]^2 \|\nabla^{2m+2} f(\mathbf{x})\|_2^2 &= dn^{-1} h^{-d-1} \|K\|_2^2 \\ \Rightarrow (4m + 4) \frac{(h^{4m+3})(h^{d+1})}{((2m + 2)!)^2} [K_{2m+2} \mathbf{I}_d]^2 \|\nabla^{2m+2} f(\mathbf{x})\|_2^2 &= dn^{-1} \|K\|_2^2 \\ \Rightarrow h^{d+4m+4} &= n^{-1} \frac{((2m + 2)!)^2}{(4m + 4)} \|K\|_2^2 \frac{d}{[K_{2m+2} \mathbf{I}_d]^2 \|\nabla^{2m+2} f(\mathbf{x})\|_2^2}. \end{aligned}$$

Hence, the approximate fourth-order optimal bandwidth, in the sense of minimizing Eq.(23) with respect to h , is of the form:

$$h = \left(\frac{((2m + 2)!)^2}{(4m + 4) \|K\|_2^2} \right)^{\frac{1}{d+4m+4}} \times \left(\frac{d}{[K_{2m+2} \mathbf{I}_d]^2 \|\nabla^{2m+2} f(\mathbf{x})\|_2^2} \right)^{\frac{1}{d+4m+4}} n^{-\frac{1}{d+4m+4}}. \quad (24)$$

On substituting Eq.(24) into Eq.(23), we have:

$$\begin{aligned} \text{AMISE}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) &= \frac{1}{((2m + 2)!)^2} \left[\frac{((2m + 2)!)^2}{4m + 4} \frac{\|K\|_2^2}{(K_{2m+2} \mathbf{I}_d)^2} \times \frac{d}{\|\nabla^{2m+2} f(\mathbf{x})\|_2^2} \right]^{\frac{4m+4}{d+4m+4}} (K_{2m+2} \mathbf{I}_d)^2 \|\nabla^{2m+2} f(\mathbf{x})\|_2^2 n^{-\frac{4m+4}{d+4m+4}} + \\ &n^{-1} \left(\frac{((2m + 2)!)^2}{(4m + 4)} \|K\|_2^2 \frac{d}{[K_{2m+2} \mathbf{I}_d]^2 \|\nabla^{2m+2} f(\mathbf{x})\|_2^2} \right)^{\frac{1}{d+4m+4}} \times \\ &\left(n^{-\frac{1}{d+4m+4}} \right)^{-d} \|K\|_2^2. \\ &= \frac{1}{((2m + 2)!)^2} (4m + 4)^{-\frac{4m+4}{d+4m+4}} d^{\frac{4m+4}{d+4m+4}} (\|K\|_2^2)^{\frac{4m+4}{d+4m+4}} \\ &((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} (\|K\|_2^2)^{\frac{d}{d+4m+4}} \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}} + \\ &(4m + 4)^{\frac{d}{d+4m+4}} \left(\frac{((2m + 2)!)^2}{(4m + 4)} \right)^{-\frac{d}{d+4m+4}} d^{-\frac{d}{d+4m+4}} (\|K\|_2^2)^{\frac{4m+4}{d+4m+4}} \\ &((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}. \\ &= \left[\frac{1}{((2m + 2)!)^2} \left(\frac{d + (4m + 4)}{d} \right)^{\frac{d+4m+4}{d+4m+4}} \right] \times \\ &(\|K\|_2^2)^{\frac{4m+4}{d+4m+4}} ((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}. \\ &= \left[\frac{1}{((2m + 2)!)^2} \left(\frac{d + 4m + 4}{d} \right)^{\frac{d}{d+4m+4}} \right] \times \\ &(\|K\|_2^2)^{\frac{4m+4}{d+4m+4}} ((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}. \\ &= \left(\frac{d + 4m + 4}{(4m + 4)} \right) \left[\frac{(4m + 4)}{d} \frac{1}{((2m + 2)!)^2} \right]^{\frac{d}{d+4m+4}} (\|K\|_2^2)^{\frac{4m+4}{d+4m+4}} \times \\ &((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}. \\ &= \frac{1}{d} \left(\frac{d + 4m + 4}{(4m + 4)} \right) \left[\frac{(4m + 4)}{((2m + 2)!)^2} \right]^{\frac{d}{d+4m+4}} d^{1-\frac{d}{d+4m+4}} (\|K\|_2^2)^{\frac{4m+4}{d+4m+4}} \times \\ &((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}. \\ &= \left(\frac{d + 4m + 4}{d(4m + 4)} \right) \left[\frac{(4m + 4)}{((2m + 2)!)^2} \right]^{\frac{d}{d+4m+4}} (d \|K\|_2^2)^{\frac{4m+4}{d+4m+4}} \times \\ &((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}. \end{aligned}$$

Thus, the expression for the generalized AMISE, independent of the optimal bandwidth (h), can be formulated as follows:

$$\begin{aligned} \text{AMISE}_{\hat{f}_{\mathbf{H}}}(\mathbf{x}) &\cong \left(\frac{d + 4m + 4}{d(4m + 4)} \right) \left[\frac{(4m + 4)}{((2m + 2)!)^2} \right]^{\frac{d}{d+4m+4}} \times \\ &(d \|K\|_2^2)^{\frac{4m+4}{d+4m+4}} ((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \times \\ &\left(\|\nabla^{2m+2} f(\mathbf{x})\|_2^2 \right)^{\frac{d}{d+4m+4}} \cdot n^{-\frac{4m+4}{d+4m+4}}. \end{aligned} \quad (25)$$

Adopting the nomenclature $AMISE^{2m}$ for $AMISE_{f_H}^{\hat{H}}(\mathbf{x})$, Eq.(25) becomes:

$$AMISE^{2m} \cong \left(\frac{d+4m+4}{d(4m+4)}\right) \left[\frac{(4m+4)}{((2m+2)!)^2}\right]^{\frac{d}{d+4m+4}} \times (d \|K\|_2^2)^{\frac{4m+4}{d+4m+4}} ((K_{2m+2} \mathbf{I}_d)^2)^{\frac{d}{d+4m+4}} \times (\|\nabla^{2m+2} f(\mathbf{x})\|_2^2)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}. \tag{26}$$

□

By comparing Eqs.(10) and (26), we observe a significant improvement in the convergence rate of the global error, also known as AMISE. The rate has transitioned from $n^{-\frac{8}{d+8}}$ to $n^{-\frac{4m+4}{d+4m+4}}$, showcasing a more favorable convergence behavior. This improvement has been made possible due to the regularity conditions outlined in the statement of the theorem (that is, $\int_{\mathbb{R}^d} (\mathbf{t}^T \mathbf{t})^{m+1} K(\mathbf{t}) d\mathbf{t} = K_{2m+2} \mathbf{I}_d$). The regularity conditions not only contribute to the reduction in global error but also play a crucial role in establishing the generalized convergence schemes for any d -dimensional fourth-order polynomial symmetric kernels. Thus, these conditions provide valuable insights and enable us to achieve improved convergence performance in the estimation process.

4. QUANTITATIVE VALIDATION OF RESULTS

In this section, our primary aim is to validate and reinforce the findings delineated in Section 3 through a comprehensive comparison with the outcomes reported by Deheuvels (see Scott [24] pg 189) and Jones *et al.* (see Chacón and Duong [25] pg 70). The objective of this quantitative validation is twofold. Firstly, it aims to affirm the robustness and precision of the obtained results by scrutinizing their consistency with established findings in the literature. Secondly, it seeks to elucidate the generalizability of our proposed convergence scheme across diverse scenarios of sample sizes and dimensionality. This thorough examination is crucial for bolstering the credibility and broader applicability of our research findings.

The detailed tabulated results presented in Tables 1 through 3 serve as the foundation for this quantitative validation. These tables provide a comprehensive comparison between the proposed convergence rate (PG) and those presented by Jones *et al.* (JMH) [25] and, Deheuvels (PD) [24]. The examination spans various sample sizes and dimensionalities, allowing for a nuanced understanding of how well our proposed schemes perform across a spectrum of scenarios.

By subjecting our results to this quantitative scrutiny, we aim to fortify the scientific rigor of our study and provide a clear assessment of the effectiveness and reliability of the proposed convergence scheme for fourth-order multivariate kernels. This process contributes to the comprehensive evaluation of our research findings and enhances their value in the broader context of kernel density estimation.

Table 1. Comparing the proposed convergence rate (PG) with the convergence rates presented by Jones *et al.* (JMH) [25] and, Deheuvels (PD) [24] across various sample sizes (n) when $d = 1$.

n	PG	JMH	PD
10	0.1003750000	0.1291550000	0.1584890000
25	0.0402118000	0.0571988000	0.0761462000
300	0.0033656200	0.0062822400	0.0104304000
1,000	0.0010119300	0.0021544300	0.0039810700
100,000	0.0000102135	0.0000359381	0.0001000000

In Table 1, a comparative analysis of the proposed convergence rate (PG) against Jones *et al.* (JMH) [25] and, Deheuvels (PD) [24] is presented. The comparison is conducted across various sample sizes (n) while maintaining a fixed dimensionality ($d = 1$). As the sample size increases, all approaches exhibit a decline in convergence rates. Notably, PG consistently demonstrates the fastest convergence, followed by JMH and PD, which generally exhibit higher rates than JMH. This observation underscores the superior performance of PG in achieving faster convergence under the specified conditions.

Table 2. Comparing the proposed convergence rate (PG) with the convergence rates presented by Jones *et al.* (JMH) [25] and, Deheuvels (PD) [24] across various sample sizes (n) when $d = 2$.

n	PG	JMH	PD
10	0.1007500000	0.1584890000	0.2154430000
25	0.0404287000	0.0761462000	0.1169610000
300	0.0034009200	0.0104304000	0.0223144000
1,000	0.0010254600	0.0039810700	0.0100000000
100,000	0.0000105033	0.0001000000	0.0004641590

Convergence rates (PG, JMH, and PD) are compared for various sample sizes (n) and fixed dimensionality ($d = 2$) in Table 2. All approaches see a drop in convergence rates as sample sizes increase. Compared to JMH and PD, PG regularly shows faster convergence; the differences get more noticeable as the sample size grows.

Table 3. Comparing the proposed convergence rate (PG) with the convergence rates presented by Jones *et al.* (JMH) [25] and, Deheuvels (PD) [24] across various sample sizes (n) when $d = 3$.

n	PG	JMH	PD
10	0.1011260000	0.1873820000	0.2682700000
25	0.0406497000	0.0962313000	0.1589190000
300	0.0034392900	0.0157927000	0.0384151000
1,000	0.0010407700	0.0065793300	0.0193070000
100,000	0.0000109125	0.0002310130	0.0013895000

Table 3 illustrates the variation in convergence rates (PG, JMH, and PD) with changes in sample sizes. Larger sample sizes are associated with decreased convergence rates, and notably, PG consistently demonstrates better performance compared to JMH and PD.

5. RESULTS AND CONCLUSION

Upon comparing Eqs. (10) and (26), a clear improvement in the convergence rate is observed for the global error of any fourth-order polynomial symmetric kernel. The transition from $n^{-\frac{8}{d+8}}$ (for $d = 1, 2, \dots, < \infty$) in Eq. (10) to $n^{-\frac{4m+4}{d+4m+4}}$ (for

$m = 1, 2, \dots, < \infty$; $d = 1, 2, \dots, < \infty$) in Eq. (26) signifies a notable enhancement in convergence behavior across varying dimensions and orders.

Moreover, obtaining a closed-form solution for the bandwidth and minimizing the generalized AMISE expression for fourth-order kernels yields noteworthy insights. Initially, the introduction of negative kernels proved to be a pivotal factor in enhancing MISE. This deliberate integration of negative kernels represents a substantial advancement in the overall efficiency of the estimation process, contributing significantly to the precision and accuracy of our methodology. Additionally, the AMISE of fourth-order kernels exhibits a swifter convergence rate compared to second-order kernels. However, it is crucial to recognize that this accelerated rate remains slower than the univariate case ($d = 1$) owing to the impact of the curse of dimensionality on the convergence rate, Scott [24].

Quantitatively validating these results, Table 1 compares the proposed convergence rate (PG) with Jones *et al.* (JMH) and, Deheuvels (PD) across various sample sizes (n) when the dimensionality (d) is fixed at 1. Larger sample sizes contribute to decreased convergence rates for all methods, with PG consistently exhibiting the lowest values, indicating the fastest convergence rate. The effectiveness of PG in achieving faster convergence remains consistent across all sample sizes, underscoring the importance of ample data for accurate estimates in scenarios with $d = 1$.

Table 2 extends this comparison to a dimensionality of 2. Increasing sample sizes leads to decreased convergence rates for all methods, and PG consistently demonstrates lower values compared to JMH and PD, with differences becoming more pronounced as sample size increases. The reduction in convergence rates with larger sample sizes emphasizes the importance of sufficient data for accurate estimation, particularly in a higher-dimensional space ($d = 2$). The observed differences between PG and other methods are more substantial, indicating a potentially more significant practical impact.

Table 3 further extends the analysis to a dimensionality of 3. Larger sample sizes again result in decreased convergence rates for all methods, and PG consistently exhibits lower values compared to JMH and PD, with differences becoming more pronounced with increasing sample size. The reduction in convergence rates with larger sample sizes underscores the importance of adequate data for accurate estimation, particularly in a higher-dimensional space ($d = 3$). The substantial differences between PG and other methods also highlight the potential practical impact of PG in this higher-dimensional scenario.

In summary, the tables collectively suggest that the proposed method (PG) consistently outperforms existing methods (JMH, PD) across varying dimensionalities. PG's superiority in achieving a faster convergence rate is particularly evident in scenarios with larger sample sizes, emphasizing its potential advantages in practical applications.

In conclusion, our extensive investigation, encompassing both analytical and quantitative analyses, has revealed the superior performance of the new convergence rate compared to its predecessors. The new convergence scheme for fourth-order kernels not only reduces bias but also improves convergence rates, especially in higher dimensions. So, using this generalized conver-

gence scheme for fourth-order kernels helps address challenges related to increasing dimensionality during the estimation procedure.

References

- [1] B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman & Hall, London, 1986. <https://www.routledge.com/Density-Estimation-for-Statistics-and-Data-Analysis/Silverman/p/book/9780412246203>.
- [2] M. P. Wand & M. C. Jones, *Kernel smoothing*, Chapman & Hall, London, 1995. <https://www.taylorfrancis.com/books/mono/10.1201/b14876/kernel-smoothing-wand-jones>.
- [3] H. Akaike, "An approximation to the density functions ", *Annals of the Institute of Statistical Mathematics* **6** (1954) 127. <https://doi.org/10.1007/BF02900741>.
- [4] E. Fix & J. L. Hodges, "Discriminatory analysis: nonparametric discrimination consistency properties ", Report No. 4, Project No. **21.29.004**, USAF School of Aviation Medicine, Randolph Field, Texas, 1951, pp. 238-247. <https://sci-hub.se/10.2307/1403797>.
- [5] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function ", *Annals of Mathematical Statistics* **27** (1956) 832. <https://doi.org/10.1007/978-1-4419-8339-8-13>.
- [6] B. W. Silverman, "Density ratios, empirical likelihood, and cot death ", *Applications of Statistics* **27** (1978) 26. <https://doi.org/10.2307/2346222>.
- [7] D. W. Scott & L. E. Factor, "Monte carlo study of three data-based nonparametric probability density estimators ", *Journal of the American Statistical Association* **76** (1981) 9. <https://www.jstor.org/stable/2287033>.
- [8] M. Rudemo, "Empirical choice of histograms and kernel density estimators ", *Scandinavian Journal of Statistics* **9**(2) (1982) 65. <https://www.jstor.org/stable/4615859>.
- [9] M. P. Wand & M. C. Jones, "Comparison of smoothing parameterizations in bivariate kernel density estimation ", *Journal of the American Statistical Association* **88**(422) (1993) 520. <https://doi.org/10.2307/2290332>.
- [10] A. W. Bowman & P. J. Foster, "Adaptive smoothing and density-based tests of multivariate normality ", *Journal of the American Statistical Association* **88** (1993) 529. <https://doi.org/10.2307/2290333>.
- [11] T. Duong, & M. L. Hazelton, "Plug-in bandwidth matrices for bivariate kernel density estimation ", *Journal of Nonparametric Statistics* **15** (2003) 17. <https://www.mvstat.net/duong/research/publications/duong-hazelton-2003-jns.pdf>.
- [12] B. A. Afere, "A family of multivariate higher-order hybrid polynomial kernels in kernel density estimation (Doctoral dissertation)", University of Benin, Benin City, Nigeria, 2010.
- [13] M. C. Jones & D. F. Signorini, "A comparison of higher-order bias kernel density estimation ", *Journal of the American Statistical Association* **92** (1997) 1063. <https://doi.org/10.1080/01621459.1997.10474062>.
- [14] J. E. Chacón, "Data-driven choice of the smoothing parametrization for kernel density estimators ", *Canadian Journal of Statistics* **37** (2009) 249. <https://doi.org/10.1002/cjs.10016>.
- [15] T. Duong & M. L. Hazelton, "Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimator ", *Journal of Multivariate Analysis* **15** (2004) 17. <https://doi.org/10.1016/j.jmva.2004.04.004>.
- [16] P. Hall & J. S. Marron, "Lower bounds for bandwidth selection in density estimation ", *Probability Theory and Related Fields* **90** (1991) 149. <https://link.springer.com/article/10.1007/BF01192160>.
- [17] P. Hall, S. J. Sheather, M. C. Jones & J. S. Marron, "On optimal data-based bandwidth selection in kernel density estimation ", *Biometrika* **78** (1991) 263. <https://doi.org/10.2307/2337251>.
- [18] F. Kimari, A. Adem & L. Kiti, "Efficiency of various bandwidth selection methods across different kernels ", *IOSR Journal of Mathematics* **15** (2019) 55. <https://doi.org/10.9790/5728-1503015562>.
- [19] S. J. Sheather & M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation ", *Journal of the Royal Statistical Society, Series B (Methodological)* **53** (1991) 683. <https://www.jstor.org/stable/2345597>.
- [20] I. U. Siloko, E. A. Siloko, O. Ikpotokin, C. C. Ishiekwene & B. A. Afere, "On asymptotic mean integrated squared error's reduction techniques in kernel density estimation ", *International Journal of Computational and Theoretical Statistics*, **6** (2019) 89. <https://journal.uob.edu.bh/handle/123456789/3479>.
- [21] T. Cacoulous, "Estimation of a multivariate density ", *Annals of the In-*

- stitute of Statistical Mathematics, **18** (1966) 179. <https://doi.org/10.1007/BF02869528>.
- [22] A. W. Bowman, "An alternative method of cross-validation for the smoothing of density estimation ", *Biometrika*, **71**(2) (1984) 353. <https://doi.org/10.1093/biomet/71.2.353>.
- [23] M. Jiang & S. P. Provost, "A hybrid bandwidth selection methodology for kernel density estimation", *Journal of Statistical Computation and Simulation* **84** (2014) 614. <https://doi.org/10.1080/00949655.2012.721366>.
- [24] D. W. Scott, *Multivariate Density Estimation: Theory, practice and visualization*, John Wiley & Sons Inc., New York, 1992. <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316849>.
- [25] J. E. Chacón & T. Duong, *Multivariate Kernel Smoothing and Its Applications*, Chapman & Hall, London, 2018. <https://www.routledge.com/Multivariate-Kernel-Smoothing-and-Its-Applications/Chacon-Duong/p/book/9780367571733>.
- [26] E. Parzen, "On the estimation of a probability density function and the mode ", *Annals of Mathematical Statistics*, **33** (1962) 1065. <https://sci-hub.st/10.1214/aoms/1177704472>.