Annals of the Institute of Statistical Mathematics (2025) 77:459–489
https://doi.org/10.1007/s10463-024-00923-8

On uniform consistency of nonparametric estimators smoothed by the gamma kernel

Benedikt Funke[1] · **Masayuki Hirukawa**[2]

Received: 31 May 2024 / Revised: 30 September 2024 / Accepted: 20 November 2024 /
Published online: 17 February 2025

Abstract
This paper documents a set of uniform consistency results with rates for nonparametric density and regression estimators smoothed by the gamma kernel having support on the nonnegative real line. It is known that this kernel can well calibrate the shapes of 'cost' distributions that are characterized by a sharp peak in the vicinity of the origin and a long right tail. In this paper, weak and strong uniform consistency and corresponding convergence rates of gamma kernel estimators are explored in a multivariate framework. Our analysis is built on compact sets expanding to the nonnegative orthant and general sequences of smoothing parameters. The results are useful for asymptotic analysis of two-step semiparametric estimation using a first-step kernel estimate as a plug-in.

Keywords Boundary bias · Density derivative estimation · Density estimation · Gamma kernel · Nonparametric regression estimation · Uniform convergence

1 Introduction

Researchers and policy-makers are often interested in the distributions of nonnegative economic and financial variables including incomes, wages, consumption expenditures, short-term interest rates, and actuarial losses. These variables are also examples of cost variables. Distributions of costs have two features in common. One is the existence of a natural boundary at the origin due to their

✉ Benedikt Funke
benedikt.funke@th-koeln.de

Masayuki Hirukawa
hirukawa@econ.ryukoku.ac.jp

[1] Institute for Insurance Studies, TH Köln - University of Applied Sciences, Gustav-Heinemann-Ufer 54, 50968 Köln, Germany

[2] Faculty of Economics, Ryukoku University, 67 Tsukamoto-cho, Fukakusa, Fushimi-ku, Kyoto 612-8577, Japan

nonnegativity. The other is that cost distributions are highly right-skewed with a concentration of observations in the vicinity of the origin and a long tail with sparse data. These features can be also observed in distributions of non-cost variables such as quantities demanded and transaction volumes.

Consider the problem of estimating densities of cost distributions nonparametrically via kernel smoothing. Then, two problems will arise when standard symmetric kernels are used. The first issue is the so-called boundary bias in the vicinity of the origin. Second, different amounts of smoothing should be made at different locations; to be more precise, while a short bandwidth is appropriate for the region near the boundary in order not to miss out the peak, a longer bandwidth is required to capture the shape of the tail part well. Accordingly, two distinct modifications must be made simultaneously for standard smoothing techniques. Boundary correction methods can handle the boundary issue; see, for instance, Sect. 3 of Karunamuni and Alberts (2005) for a concise review of the methods. To calibrate the density curves of cost variables well, we may resort to adaptive smoothing such as variable bandwidth methods by Abramson (1982) and Terrell and Scott (1992).

It is quite cumbersome to modify standard smoothing techniques in two different directions. Then, smoothing by nonstandard asymmetric kernel functions has emerged as a viable alternative. Asymmetric kernels have two properties in common. First, the support of an asymmetric kernel matches that of the underlying probability density function (pdf). Therefore, it is free of the boundary bias by construction. Second, shapes of the kernel vary according to the locations at which smoothing is made; in other words, the amount of smoothing changes in an adaptive manner. It is worth emphasizing that a single value of the smoothing parameter can generate a variety of shapes of an asymmetric kernel. From empirical standpoints, this is a clear advantage over the variable bandwidth methods for symmetric kernels, which require different bandwidth values for different design points.

Although asymmetric kernels possess aforementioned appealing properties, uniform consistency of asymmetric kernel estimators is yet to be fully explored. This paper aims at delivering weak and strong uniform convergence rates for nonparametric estimators with support on $\mathbb{R}_+$ smoothed by an asymmetric kernel. While a number of asymmetric kernels have been proposed for the last two decades, we specialize our analysis in one of the pioneered but still most popular asymmetric kernels -- the gamma kernel by Chen (2000). Our particular choice is grounded on three reasons. First, as presented in Table 1.1 of Hirukawa (2018), the gamma kernel is frequently applied to empirical models in economics and finance due to its favorable evidence. Second, this kernel has been discussed as an example of boundary kernels in textbooks (e.g., Racine, 2019), as well as the beta kernel by Chen (1999), which is yet another popular asymmetric kernel defined on the unit interval. Third, as will be discussed again shortly, asymmetric kernels cannot be expressed in the location-scale form that symmetric kernels take. Inevitably, kernel-specific arguments are required for detailed asymptotic analyses of asymmetric kernel estimators, and analytical tractability of the estimators is a key issue. The gamma function is a building block for the gamma kernel, and there is rich literature on approximation techniques to the gamma and related functions as they are actively studied.

For a data point $u \in \mathbb{R}_+$, a design point $x \in \mathbb{R}_+$ and a smoothing parameter $b > 0$, the gamma kernel is defined as

$$K_{G(x,b)}(u) = \frac{u^{x/b} \exp(-u/b)}{b^{x/b+1}\Gamma(x/b+1)} \mathbf{1}\{u \geq 0\},$$

where $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)dt$ for $a > 0$ is the gamma function, and $\mathbf{1}\{\cdot\}$ denotes an indicator function. A nonnegative random variable Z is said to obey the gamma distribution having the shape parameter $\alpha > 0$ and the scale parameter $\beta > 0$, which is denoted as $G(\alpha, \beta)$ in shorthand notation hereinafter, if its pdf is given by $f(z) = z^{\alpha-1} \exp(-z/\beta)\mathbf{1}\{z \geq 0\}/\{\beta^\alpha \Gamma(\alpha)\}$. Observe that the gamma kernel can be interpreted as the pdf of $G(x/b + 1, b)$.

To cope with multivariate problems, we construct a d-dimensional tensor product kernel

$$\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{u}) = \prod_{j=1}^{d} K_{G(x_j,b_j)}(u_j) = \prod_{j=1}^{d} \frac{u^{x_j/b_j} \exp\left(-u_j/b_j\right)}{b_j^{x_j/b_j+1}\Gamma\left(x_j/b_j+1\right)} \mathbf{1}\{u_j \geq 0\},$$

where $\mathbf{u} := \left(u_1, \ldots, u_d\right)^\top \in \mathbb{R}_+^d$, $\mathbf{x} := \left(x_1, \ldots, x_d\right)^\top \in \mathbb{R}_+^d$ and $\mathbf{b} := \left(b_1, \ldots, b_d\right)^\top \in \mathbb{R}_+^d$ are d-dimensional vectors of data points, design points and smoothing parameters, respectively. It is worth emphasizing that as in Hirukawa et al. (2022), different smoothing parameter values are allowed for different dimensions. This is a sharp contrast to Hansen (2008) and Kristensen (2009), who provide uniform convergence results of sample average functionals using product symmetric kernels and a single bandwidth value to all dimensions.

Our analysis takes the same procedure as in Hansen (2008), Kristensen (2009) and Hirukawa et al. (2022). The common approach across these authors is to start from examining a nonparametric estimator of

$$g(\mathbf{x}) := m(\mathbf{x})f(\mathbf{x}), \tag{1}$$

where $m(\mathbf{x}) := E(Y|\mathbf{X} = \mathbf{x})$ and $f(\mathbf{x})$ is the marginal pdf of $\mathbf{X}$. Based on this idea, the outline of our analysis can be described as follows. Given n *i.i.d.* observations $\left\{\left(Y_i, \mathbf{X}_i\right)\right\}_{i=1}^n \in \mathbb{R} \times \mathbb{R}_+^d$, uniform consistency of the sample average functional of the form

$$\hat{g}_G(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)$$

for $g(\mathbf{x})$ is investigated. Most of multivariate gamma kernel estimators such as the joint density estimator and the Nadaraya-Watson (NW) and local linear (LL) regression estimators can be expressed in this form. Therefore, weak and strong uniform convergence results of $\hat{g}_G(\mathbf{x})$ can directly apply to these estimators. The remaining question is how to determine the interval on which weak and strong uniform convergence rates of $\hat{g}_G(\mathbf{x})$ are derived. Our focus is on its uniform consistency on a d-dimensional hyperrectangle inside $\mathbb{R}_+^d$ that is either fixed or expanding to $\mathbb{R}_+^d$ at a

suitable rate. This framework enables us to employ Stirling's approximation to the gamma function as a workhorse in technical proofs.

This paper contributes to the literature in various aspects. First, Hansen (2008) and Kristensen (2009) prove uniform consistencies with rates of nonparametric curve estimators using standard symmetric kernels on expanding or unbounded sets in multivariate frameworks. As mentioned above, unlike symmetric kernels, the univariate gamma kernel cannot be expressed in the location-scale form $(1/b)K\{(u-x)/b\}$. Understandably, our proof strategies differ from those taken by Hansen (2008) and Kristensen (2009) substantially.

Second, this paper can be positioned as a complement to Hirukawa et al. (2022), who demonstrate weak and strong uniform convergences with rates for nonparametric estimators using the beta kernel. The results therein are applied to the multistep estimation for two-sample regression models by Hirukawa et al. (2023). So far applications of the gamma kernel have been concentrated on purely nonparametric estimation problems like density and regression estimations. It is anticipated that the results in this paper can serve as theoretical justification when the gamma estimates are employed as the first-step nonparametric estimate for two-step semiparametric estimation. Examples of such estimation procedures include Robinson (1988), Newey (1994), Rilstone (1996), and Stengos and Yan (2001), to name a few.

Third, our uniform convergence results are established in a multivariate framework, as in Hirukawa et al. (2022). Recently the literature on joint density estimation using asymmetric kernels has been growing. Examples include Bouezmarni and Rombouts (2010), Funke and Kawka (2015), Ouimet (2021, 2022), Ouimet and Tolosana-Delgado (2022), and Bertin et al. (2023), as well as Hirukawa et al. (2022). In particular, Ouimet (2022) extends the scope of asymmetric kernels to the one built on the pdf of the Wishart distribution, and he explores convergence properties of the joint density estimator using the Wishart asymmetric kernel. Notice that the Wishart distribution is a matrix-variate analogue of the gamma distribution. In contrast, the study on asymmetric kernel regression estimation with multiple regressors is still scarce. Examples other than Hirukawa et al. (2022) include Bouzebda et al. (2024) and Genest and Ouimet (2024), who examine the NW and LL regression estimators on the simplex using the Dirichlet kernel, respectively.

Last but not least, our companion paper (Funke and Hirukawa, 2024b) demonstrates uniform consistency of the first-order density derivative estimators smoothed by the product gamma and beta kernels. These estimators are defined as a multivariate extension of the one proposed by Funke and Hirukawa (2024a). A challenge lies in deriving uniform convergence rates of density derivative estimators when using the sample average functional-based approach. Consequently, we document the uniform convergence results in a separate paper. Interested readers may consult Funke and Hirukawa (2024b) for more details.

The remainder of this paper is organized as follows. Section 2 delivers weak and strong uniform convergence results for the sample average functional $\hat{g}_G(\mathbf{x})$ on a compact set that is either fixed or expanding to the d-dimensional nonnegative orthant. Subsequently, in Sect. 3, these results are applied to nonparametric density and regression estimators. Sect. 4 conducts Monte Carlo simulations. Inspired by Liu and Li (2023), this simulation study is designed to empirically confirm optimal

uniform convergence rates of gamma density and regression estimators. Sect. 5 concludes. All proofs are given in the Appendix.

This paper adopts the following notational conventions: '$a_n = o(b_n)$' signifies that a_n/b_n converges to 0; '$a_n = O(b_n)$' means that a_n/b_n is bounded; we say that '$a_n \asymp b_n$' if there exist constants $0 < c_1 < c_2 < \infty$ so that $c_1 a_n \le b_n \le c_2 a_n$; $\mathsf{D}_{\mathbf{x}} = \partial/\partial\mathbf{x} = (\partial/\partial x_1, \ldots, \partial/\partial x_d)^\top$ signifies the d-dimensional first-order partial derivative (or gradient) operator; $h_p^{(1)}(\mathbf{x}) = \frac{\partial h(\mathbf{x})}{\partial x_p}$ and $h_{pq}^{(2)}(\mathbf{x}) = \frac{\partial^2 h(\mathbf{x})}{\partial x_p \partial x_q}$ denote the first- and second-order partial derivatives of a function $h(\mathbf{x})$, respectively; $\Psi(x) = d\ln\Gamma(x)/dx = \Gamma^{(1)}(x)/\Gamma(x)$ is the digamma function; '*a.s.*' means "almost surely"; $\|\mathbf{A}\|$ is the Frobenius norm of matrix $\mathbf{A}$, i.e., $\|\mathbf{A}\| = \{\mathrm{tr}\,(\mathbf{A}^\top\mathbf{A})\}^{1/2}$; and the expression '$X \overset{d}{=} Y$' reads "A random variable X obeys the distribution Y."

2 Main results

2.1 Weak uniform convergence of the sample average functional

Our analysis starts from demonstrating weak uniform consistency with a rate of the sample average functional $\hat{g}_G(\mathbf{x})$ for $g(\mathbf{x})$ on a d-hyperrectangle

$$\mathbb{S}_{\mathbf{X}} = \mathbb{S}_{\mathbf{X}}(\boldsymbol{\eta}) := \prod_{j=1}^{d} \left[\eta_j, \eta_j^{-1}\right] \subseteq \mathbb{R}_+^d,$$

where the boundary parameters $\boldsymbol{\eta} := (\eta_1, \ldots, \eta_d)^\top$ either are fixed or shrink to zero at a suitable rate. Observe that in the latter scenario, $\mathbb{S}_{\mathbf{X}}$ expands to the d-dimensional nonnegative orthant.

To deliver the result, we impose the following regularity conditions.

Assumption 1 $\{(Y_i, \mathbf{X}_i)\}_{i=1}^n \in \mathbb{R} \times \mathbb{R}_+^d$ are *i.i.d.* random vectors.

Assumption 2 Let $h(\mathbf{x})$ be either $f(\mathbf{x})$ or $g(\mathbf{x})$. Then, there are constants $L_1, L_2, L_3, \gamma > 0$ that satisfy the followings.

(i) $\left|h_{jk}^{(2)}(\mathbf{x})\right| \le L_1\left[L_2^{-(2+\gamma)}\mathbf{1}\{x_k < L_2\} + x_k^{-(2+\gamma)}\mathbf{1}\{x_k \ge L_2\}\right]$ for all $\mathbf{x} \in \mathbb{R}_+^d$ and for all $j, k \in \{1, \ldots, d\}$.

(ii) $\left|h_{jk}^{(2)}(\mathbf{x}) - h_{jk}^{(2)}(\mathbf{x}')\right| \le L_3\|\mathbf{x} - \mathbf{x}'\|$ for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}_+^d$ and for all $j, k \in \{1, \ldots, d\}$.

Assumption 3 There are constants $\delta > 0$ and $C_1 \ge 1$ that satisfy $E|Y|^{2+\delta} < \infty$ and

$$\sup_{\mathbf{x} \in \mathbb{R}_+^d} E\left(|Y|^{2+\delta}\middle|\mathbf{X} = \mathbf{x}\right) f(\mathbf{x}) \le C_1. \tag{2}$$

Assumption 4W
Sequences $b_j\big(=b_j(n)\big), \eta_j\big(=\eta_j(n)\big) > 0$ satisfy the followings as $n \to \infty$.

(i) $b_j, \eta_j \to 0$ for all $j \in \{1, \ldots, d\}$.
(ii) There is a sequence $\rho(=\rho(n)) > 0$ that satisfies $b_j/\eta_j \asymp \rho$ for all $j \in \{1, \ldots, d\}$, $\rho \to 0$ and $\rho = o\Big(\min_{1\le j\le d} \eta_j^2\Big)$.
(iii) $\ln n \Big/ \left(n\sqrt{\prod_{j=1}^d b_j\eta_j}\right) \to 0$.

We begin our discussion on these regularity conditions by making a few remarks on Assumption 2. First, this assumption implies that second-order partial derivatives of $f(\mathbf{x})$ and $g(\mathbf{x})$ are Lipschitz continuous and uniformly bounded on $\mathbb{R}_+^d$. The condition (i) in Assumption 2 also regulates tail decay rates of second-order partial derivatives of $f(\mathbf{x})$ and $g(\mathbf{x})$. As given in Lemma 1 in the Appendix, second- and higher-order moments of the univariate gamma kernel around the design point x depend on x, which is unbounded from above. This condition helps control the order of magnitude in the leading bias term of $\hat{g}_G(\mathbf{x})$, because it ensures uniform boundedness of $\left|h_j^{(1)}(\mathbf{x})\right|$ and $\left|h_{jj}^{(2)}(\mathbf{x})x_j\right|$ on $\mathbb{R}_+^d$. Notice that an equivalent of the condition (i) cannot be found in the study of uniform convergence of beta kernel estimators by Hirukawa et al. (2022). While leading bias coefficients of univariate beta kernel estimators also depend on the design point, it is confined within the unit interval and thus this type of condition is unnecessary. Second, Assumption 2(i) also establishes uniform boundedness of $f(\mathbf{x})$; in other words, there is a constant $C_0 \ge 1$ so that

$$\sup_{\mathbf{x}\in\mathbb{R}_+^d} f(\mathbf{x}) \le C_0. \tag{3}$$

Third, a sufficient condition for Assumption 2 that $f(\mathbf{x})$ fulfills is that its second-order partial derivatives decay exponentially in all dimensions, in addition to their Lipschitz continuity and uniform boundedness. Examples of such densities include those of multivariate gamma (e.g., Das and Dey, 2010), multivariate truncated normal (e.g., Horrace, 2005) and multivariate folded normal distributions (e.g., Chakraborty and Chatterjee, 2013), to name a few.

It follows from (2) in Assumption 3 that $E\Big(|Y|^{2+\delta}\Big|\mathbf{X}=\mathbf{x}\Big)$ is allowed to diverge in the right tail but no faster than $\{f(\mathbf{x})\}^{-1}$. A similar condition can be found, for instance, in Hansen (2008, Assumption 2), Kristensen (2009, Assumption A3) and Hirukawa et al. (2022, Assumption 3).

Three conditions on the boundary parameter η_j in Assumption 4W are intended for the case of an expanding set. Conditions (i) and (ii) jointly mean that η_j shrinks to zero more slowly than b_j does. As will be revealed in the Appendix, this is crucial for Stirling's approximation to the gamma function. In addition, $b_j/\eta_j \asymp \rho$ in the condition (ii) merely indicates that the shrinkage rate of the ratio of b_j to η_j is identical across j. This does not automatically guarantee that the ratio b_j/η_j itself, the numerator b_j, or the denominator η_j are identical across j. Moreover,

$\rho = o\left(\min_{1\le j\le d}\eta_j^2\right)$ is a technical requirement for controlling orders of magnitude in the remainder terms of the bias.

Theorem 1 below documents weak uniform consistency of $\hat{g}_G(\mathbf{x})$ for $g(\mathbf{x})$ on the slowly expanding d-hyperrectangle $\mathbb{S}_\mathbf{X} \to \mathbb{R}_+^d$ as $n \to \infty$. Observe that the weak uniform convergence rate in this theorem is the same as what is obtained for the sample average functional smoothed by the product beta kernel in Hirukawa et al. (2022).

Theorem 1 *If Assumptions 1-3 and 4W hold, then, as* $n \to \infty$,

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} \left|\hat{g}_G(\mathbf{x}) - g(\mathbf{x})\right| = O_p\left(\sum_{j=1}^d b_j + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^d b_j\eta_j}}}\right).$$

2.2 Strong uniform convergence of the sample average functional

Next, we demonstrate strong uniform consistency of $\hat{g}_G(\mathbf{x})$ for $g(\mathbf{x})$ when $\mathbb{S}_\mathbf{X}$ is allowed to expand slowly to $\mathbb{R}_+^d$. Before proceeding, the assumption on tuning parameters must be suitably strengthened.

Assumption 4S

Sequences $b_j\big(=b_j(n)\big), \eta_j\big(=\eta_j(n)\big) > 0$ satisfy the followings as $n \to \infty$.

(i) $b_j, \eta_j \to 0$ for all $j \in \{1, \dots, d\}$.

(ii) There is a sequence $\rho(=\rho(n)) > 0$ that satisfies $b_j/\eta_j \asymp \rho$ for all $j \in \{1, \dots, d\}$, $\rho \to 0$ and $\rho = o\left(\min_{1\le j\le d}\eta_j^2\right)$.

(iii) There is a constant $\kappa \in [0, 1)$ that satisfies

$$\frac{\ln n}{n\sqrt{\prod_{j=1}^d b_j\eta_j}}\left(\sum_{j=1}^d \frac{1}{b_j\eta_j}\right)^{1-\kappa} = O(1). \tag{4}$$

Observe that the condition (iii) in Assumption 4W is replaced by a stronger condition (4). Under this condition, the convergence mode in Theorem 1 can be strengthened to almost sure convergence. Again in this case, the strong uniform convergence rate becomes the same as that of the sample average functional using the product beta kernel in Hirukawa et al. (2022).

Theorem 2 *If Assumptions 1-3 and 4S hold, then, as* $n \to \infty$, *the statement in Theorem 1 can be strengthened to almost sure convergence.*

This section concludes by deriving the optimal weak and strong uniform convergence rates of $\hat{g}_G(\mathbf{x})$ on the slowly expanding $\mathbb{S}_\mathbf{X}$. We concentrate on the optimal rates under the following sufficient conditions for Assumptions 4W and 4S. Let

sequences $b(=b(n)), \eta(=\eta(n)) > 0$ satisfy $b_1, \ldots, b_d \asymp b$, $\eta_1, \ldots, \eta_d \asymp \eta$, and either (i) $b + \eta + b/\eta^3 + \ln n/\left\{n(b\eta)^{d/2}\right\} \to 0$ (for weak uniform convergence) or (ii) $b + \eta + b/\eta^3 \to 0$ and $\ln n/\left\{n(b\eta)^{d/2+1-\kappa}\right\} = O(1)$ for some $\kappa \in [0, 1)$ (for strong uniform convergence) as $n \to \infty$. It follows from $b/\eta^3 \to 0$ that we may take $\eta \asymp b^{(1-\lambda)/3}$ for an arbitrarily small $\lambda > 0$. Then, by Theorems 1 and 2, each of weak and strong uniform convergence rates of $\hat{g}_G(\mathbf{x})$ reduces to $b + \sqrt{\ln n/\left\{nb^{d(4-\lambda)/6}\right\}}$. It is easy to see that two terms are balanced by $b^\dagger \asymp (\ln n/n)^{6/\{12+d(4-\lambda)\}}$, which yields the optimal uniform convergence rate $(\ln n/n)^{6/\{12+d(4-\lambda)\}}$, as well as the optimal boundary parameter $\eta^\dagger \asymp (\ln n/n)^{2(1-\lambda)/\{12+d(4-\lambda)\}}$.

3 Applications

3.1 Density estimation

In this section, Theorems 1 and 2 are applied to two nonparametric estimation problems. We start from presenting two theorems on weak and strong uniform convergence for the joint density estimator using the product gamma kernel

$$\hat{f}_G(\mathbf{x}) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i).$$

Theorem 3 *If Assumptions 1-3 and 4W hold, then, as $n \to \infty$,*

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} \left|\hat{f}_G(\mathbf{x}) - f(\mathbf{x})\right| = O_p\left(\sum_{j=1}^{d} b_j + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{d} b_j\eta_j}}}\right).$$

Theorem 4 *If Assumptions 1-3 and 4S hold, then, as $n \to \infty$, the statement in Theorem 3 can be strengthened to almost sure convergence.*

From Theorems 3 and 4, in particular, we can obtain the optimal weak and strong uniform convergence rates of $\hat{f}_G(\mathbf{x})$ when $\mathbb{S}_\mathbf{X}$ is fixed and a single smoothing parameter b is employed for each dimension. Let $\eta_1, \ldots, \eta_d$ be fixed. Also consider a sequence $b(=b(n)) > 0$ that satisfies $b_1, \ldots, b_d \asymp b$ and either (i) $b + \ln n/\left(nb^{d/2}\right) \to 0$ (for weak uniform convergence) or (ii) $b \to 0$ and $\ln n/\left(nb^{d/2+1-\kappa}\right) = O(1)$ for some $\kappa \in [0, 1)$ (for strong uniform convergence) as $n \to \infty$. In this setup, each of the weak and strong uniform convergence rates reduces to $b + \sqrt{\ln n/\left(nb^{d/2}\right)}$, and it can be found that $b^* \asymp (\ln n/n)^{2/(4+d)}$ balances two terms. Under this b^*, the optimal uniform convergence rate becomes

$(\ln n/n)^{2/(4+d)}$. This rate coincides with Stone's (1983) optimal global rate for nonparametric density estimation. It is also straightforward to check that the $(\ln n/n)^{2/(4+d)}$ rate is faster than $(\ln n/n)^{6/\{12+d(4-\lambda)\}}$ (= the optimal uniform convergence rate when $\mathbb{S}_{\mathbf{X}}$ slowly expands to $\mathbb{R}_+^d$) in the previous section.

3.2 Regression estimation

Next, Theorems 1 and 2 are extended to the cases of nonparametric regression estimation. We consider two most popular kernel regression estimators, namely, the NW and LL regression estimators. The NW estimator smoothed by the product gamma kernel is defined as

$$\hat{m}_G(\mathbf{x}) := \frac{\sum_{i=1}^n Y_i \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)}{\sum_{i=1}^n \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)} = \frac{\hat{g}_G(\mathbf{x})}{\hat{f}_G(\mathbf{x})}.$$

The LL estimator of $m(\mathbf{x})$ and estimators of its first-order partial derivatives $\mathsf{D}_{\mathbf{x}} m(\mathbf{x})$ are given as the joint minimizer of the local least squares problem

$$\big(\tilde{\alpha}(\mathbf{x}), \tilde{\boldsymbol{\beta}}(\mathbf{x})\big) := \arg\min_{(\alpha,\boldsymbol{\beta})} \sum_{i=1}^n \big\{Y_i - \alpha - \boldsymbol{\beta}^\top(\mathbf{X}_i - \mathbf{x})\big\}^2 \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i).$$

The LL estimator $\tilde{m}_G(\mathbf{x}) = \tilde{\alpha}(\mathbf{x})$ can be also expressed in a closed form as

$$\tilde{m}_G(\mathbf{x}) := \frac{\hat{g}_G(\mathbf{x}) - \mathbf{S}_1(\mathbf{x})^\top \mathbf{S}_2(\mathbf{x})^{-1} \mathbf{T}_1(\mathbf{x})}{\hat{f}_G(\mathbf{x}) - \mathbf{S}_1(\mathbf{x})^\top \mathbf{S}_2(\mathbf{x})^{-1} \mathbf{S}_1(\mathbf{x})},$$

where

$$\mathbf{S}_1(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n (\mathbf{X}_i - \mathbf{x}) \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i),$$

$$\mathbf{S}_2(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n (\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^\top \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i), \text{ and}$$

$$\mathbf{T}_1(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n Y_i (\mathbf{X}_i - \mathbf{x}) \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i).$$

This expression results from equation (2.1) of Ruppert and Wand (1994) and also appears in the proof of Theorem 10 in Hansen (2008).

When dealing with regression estimation, we must take care of the cases in which the marginal density $f(\mathbf{x})$ becomes (close to) zero. Singularity of $f(\mathbf{x})$ does occur as $x_j \to \infty$ for some j. It may be also the case that $f(\mathbf{x}) \to 0$ as $x_j \to 0$ for some j. Then,

we make an additional assumption as in Hansen (2008) and Hirukawa et al. (2022) before stating two uniform convergence results for the kernel regression estimators.

Assumption 4 Let $r_n := \inf_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} f(\mathbf{x}) > 0$. Then, as $n \to \infty$, $r_n \to 0$ and the followings also hold true.

$$r_n^{-1}\left(\sum_{j=1}^{d} b_j + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{d} b_j \eta_j}}}\right) \to 0 \text{ for } \hat{m}_G(\mathbf{x}).$$

$$r_n^{-2}\left(\rho + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{d} b_j \eta_j}}}\right) \to 0 \text{ for } \tilde{m}_G(\mathbf{x}).$$

In reality, $f(\mathbf{x})$ is unknown, so is r_n. Therefore, we must estimate r_n in order to verify the above conditions in practice. Now two theorems on uniform convergence of $\hat{m}_G(\mathbf{x})$ and $\tilde{m}_G(\mathbf{x})$ are delivered.

Theorem 5 *If Assumptions 1-3, 4W, and 5 hold, then, as $n \to \infty$,*

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} \left|\hat{m}_G(\mathbf{x}) - m(\mathbf{x})\right| = O_p\left\{r_n^{-1}\left(\sum_{j=1}^{d} b_j + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{d} b_j \eta_j}}}\right)\right\}, \text{ and} \tag{5}$$

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} \left|\tilde{m}_G(\mathbf{x}) - m(\mathbf{x})\right| = O_p\left\{r_n^{-2}\left(\rho + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{d} b_j \eta_j}}}\right)\right\}. \tag{6}$$

Theorem 6 *If Assumptions 1-3, 4S, and 5 hold, then, as $n \to \infty$, the statements in Theorem 5 can be strengthened to almost sure convergence.*

Several remarks are in order. First, recognizing that $\sum_{j=1}^{d} b_j/\eta_j = O(\rho)$, we can find that uniform convergence rates of $\hat{m}_G(\mathbf{x})$ and $\tilde{m}_G(\mathbf{x})$ are the same as those for product beta kernel regression estimators in Theorems 5 and 6 of Hirukawa et al. (2022). A rationale is that since the gamma function is a common key building block for the gamma and beta kernels, similar structure appears in various aspects of asymptotic analysis for regression estimators smoothed by these kernels; see Chen (2002), for instance, for more details.

Second, the results for $\hat{m}_G(\mathbf{x})$ in Theorems 5 and 6 are comparable with those in Theorems 8 and 9 of Hansen (2008), who derives weak and strong uniform convergence rates of the NW estimator using multivariate symmetric kernels while

allowing $f(\mathbf{x})$ to shrink to zero at the rate r_n in tail regions of the entire Euclidean space. Hansen (2008) demonstrates that the uniform convergence rates slow down from those of the corresponding sample average functional by a factor of the additional penalty term r_n^{-1}. It follows from (5) that Hansen's (2008) results continue to hold after replacing multivariate symmetric kernels with the product gamma kernel defined on the nonnegative orthant.

Third, there are some similarities and differences between the results for $\tilde{m}_G(\mathbf{x})$ in Theorems 5 and 6 and those in Theorems 10 and 11 of Hansen (2008). Hansen (2008) documents in these theorems that the penalty term is strengthened to r_n^{-2} for the LL estimator smoothed by multivariate symmetric kernels. It can be found in (6) that expanding $\mathbb{S}_{\mathbf{X}}$ has two distinct effects on the uniform convergence rate of the product gamma LL estimator. Not only does $\mathbb{S}_{\mathbf{X}}$ assign a more stringent penalty r_n^{-2} in comparison with the NW case but also the bias convergence decelerates from $\sum_{j=1}^{d} b_j$ to ρ because of its edge effect.

Fourth, as in the previous section, it is possible to obtain the optimal weak and strong uniform convergence rates of $\hat{m}_G(\mathbf{x})$ and $\tilde{m}_G(\mathbf{x})$ for a fixed $\mathbb{S}_{\mathbf{X}}$ and a single smoothing parameter b. In this case, it may be safely assumed that $f(\mathbf{x})$ is bounded away from zero uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$ so that r_n is also fixed. In this setup, weak and strong uniform convergence rates of $\hat{m}_G(\mathbf{x})$ and $\tilde{m}_G(\mathbf{x})$ once again reduce to $b + \sqrt{\ln n/\left(nb^{d/2}\right)}$. Accordingly, the optimal uniform convergence rates of $\hat{m}_G(\mathbf{x})$ and $\tilde{m}_G(\mathbf{x})$ are both $(\ln n/n)^{2/(4+d)}$. The rates agree with Stone's (1982) optimal global rate for nonparametric regression estimation.

4 Monte Carlo study

In this section, we conduct a simulation study to empirically confirm optimal uniform convergence rates of gamma density and regression estimators. Our Monte Carlo design is largely inspired by the one in Liu and Li (2023). The sample size n varies from 100 to 500 such that $n = n_j = 100 + 20(j-1)$ for $j \in \{1, 2, \ldots, 21\}$. For each sample size $n = n_j$, the number of replications is 100. The Monte Carlo sample $\left\{\left(X_i, Y_i\right)\right\}_{i=1}^{n}$ is generated by $X_i \overset{d}{=} G(2,1)$ and $Y_i = m\left(X_i\right) + \epsilon_i$, where $m(x) = \ln(x+1)$, $\epsilon_i \overset{d}{=} N\left(0, 0.05^2\right)$ and $X_i \perp\!\!\!\perp \epsilon_i$. Using the Monte Carlo sample, we compute the gamma density estimate $\hat{f}_G(x)$ for the marginal distribution of X and the gamma NW regression estimate $\hat{m}_G(x)$. The design points are chosen as $x = x_k = 0.5 + 0.01(k-1)$ for $k \in \{1, 2, \ldots, 501\}$ so that all these points lie within the fixed interval $\mathbb{S}_{\mathbf{X}} = [0.5, 5.5]$. Since $\mathbb{S}_{\mathbf{x}}$ is fixed and the pdf of $G(2,1)$ is bounded away from zero on $\mathbb{S}_{\mathbf{X}}$, weak and strong optimal uniform convergence

rates of $\hat{f}_G(x)$ and $\hat{m}_G(x)$ become $(\ln n/n)^{2/5}$ under the optimal smoothing parameter $b^* \asymp (\ln n/n)^{2/5}$.

To implement $\hat{f}_G(x)$ and $\hat{m}_G(x)$, we consider two estimators of b^*, namely, the rule-of-thumb (ROT) and cross-validation (CV) type ones. The ROT smoothing parameter is common across density and regression estimates and defined as $\hat{b}_{ROT} := \hat{\sigma}_X(\ln n/n)^{2/5}$, where $\hat{\sigma}_X$ is the sample standard deviation of X. While the CV type ones differ between density and regression estimates, each is designed to minimize the corresponding uniform error bound (EB). The CV type smoothing parameter for $\hat{f}_G(x)$ is defined as

$$\hat{b}_{CV,f} := \arg\min_{b\in B} CV_f(b) := \arg\min_{b\in B}\left\{\max_{i:X_i\in\mathbb{S}_{\mathbf{X}}}\left|\hat{f}_{G,\hat{b}_{ROT}}(X_i) - \hat{f}_{G,b,-i}(X_i)\right|\right\},$$

where $\hat{f}_{G,\hat{b}_{ROT}}(x)$ is the density estimate using $\hat{b}_{ROT}$ as the pilot smoothing parameter, and $\hat{f}_{G,b,-i}(x) := \{1/(n-1)\}\sum_{\ell=1,\ell\neq i}^{n} K_{G(x,b)}(X_\ell)$ is the density estimate using the smoothing parameter b and the sample with the ith observation eliminated. A pilot simulation study confirms that a smaller smoothing parameter value than $\hat{b}_{ROT}$ tends to work well in our Monte Carlo design. Then, we choose the set $B = \left\{c\hat{\sigma}_X(\ln n/n)^{2/5} : c \in \{0.1, 0.2, \ldots, 1.0\}\right\}$. The CV type smoothing parameter for $\hat{m}_G(x)$ is also defined as

$$\hat{b}_{CV,m} := \arg\min_{b\in B} CV_m(b) := \arg\min_{b\in B}\left\{\max_{i:X_i\in\mathbb{S}_{\mathbf{X}}}\left|Y_i - \hat{m}_{G,-i,b}(X_i)\right|\right\},$$

where the set B is the same as above, and $\hat{m}_{G,-i,b}(x)$ is the NW estimate using the smoothing parameter b and the sample with the ith observation eliminated.

Finally, for $h \in \{f\ m\}$ and $\hat{h}_G \in \left\{\hat{f}_G, \hat{m}_G\right\}$, the uniform error bound

$$EB\left(\hat{h}_{G,n_j}\right) = \sup_{x\in\mathbb{S}_{\mathbf{X}}}\left|\hat{h}_G(x) - h(x)\right| \approx \max_{1\le k\le 501}\left|\hat{h}_G(x_k) - h(x_k)\right|$$

is calculated, where $\hat{h}_{G,n_j}$ signifies the dependence of the estimate $\hat{h}_G$ on the sample size n_j. Then, we take an average of error bounds over 100 Monte Carlo samples. The average error bound is denoted as $\overline{EB}\left(\hat{h}_{G,n_j}\right)$.

If the gamma density or regression estimator is able to achieve its optimal uniform convergence rate, then the scatter plot of $\left\{\left(\ln\left(n_j/\ln n_j\right), \ln\overline{EB}\left(\hat{h}_{G,n_j}\right)\right)\right\}_{j=1}^{21}$ for $h \in \{f\ m\}$ should be close enough to form a straight line $\ln\overline{EB}\left(\hat{h}_{G,n_j}\right) = (\text{constant}) + (-2/5)\ln\left(n_j/\ln n_j\right)$. Each panel of Fig. 1 presents the scatter plot and its corresponding reference line. The reference line in blue has slope $-2/5$, and its intercept is determined by ordinary least squares (OLS) after the slope is fixed at $-2/5$. It

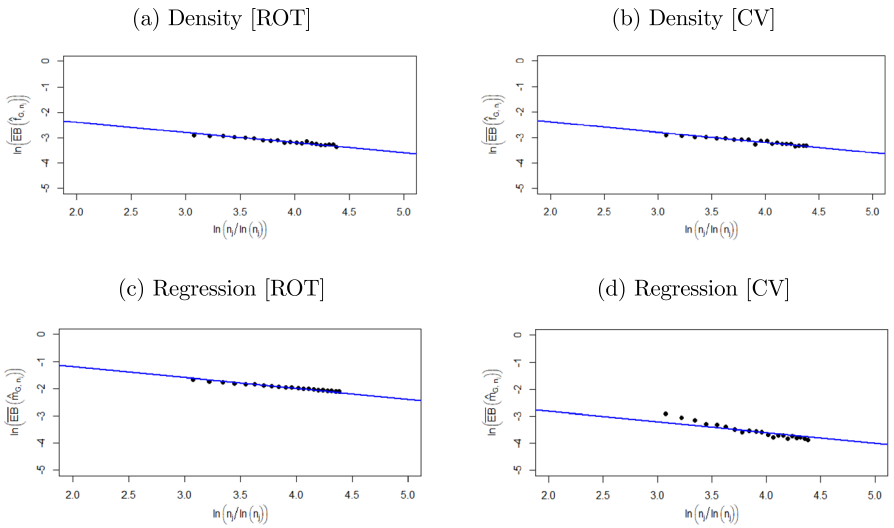

Fig. 1 ln-ln Plots for Gamma Density and Regression Estimators

can be found that the points are distributed around the line. While at first glance the scatter plot for Regression [CV] is slightly off the reference line, a closer look reveals that it occurs for first few sample sizes and the deviation disappears as the sample size increases. For reference, OLS slope estimates implied by scatter plots are −0.33 for Density [ROT], −0.35 for Density [CV], −0.33 for Regression [ROT], and −0.69 for Regression [CV]. Monte Carlo results indicate that all in all, uniform error bounds of gamma density and regression estimators agree with their theoretical optimal convergence rates.

5 Conclusion

In this paper, we have demonstrated weak and strong uniform convergence rates of joint density and regression estimators smoothed by the product gamma kernel. All such convergence results are established on a compact set that is either fixed within or expanding to the d-dimensional nonnegative orthant. Uniform convergences of density and regression estimators using the product gamma kernel can be obtained as applications of a sample average functional. The optimal uniform convergence rates on a fixed set of all these estimators concur with the corresponding best possible global convergence rates provided by Stone (1982, 1983). Finally, simulation results confirm optimal uniform convergence rates of the gamma density and regression estimators.

Acknowledgements The authors would like to thank two anonymous referees for their constructive comments and suggestions, which helped improve this paper substantially.

Appendix

A.1 Proof of Theorem 1

This proof requires the following lemmata. Lemma 1 presents moments of the univariate gamma kernel around the design point x. $E\left(\xi_x - x\right)^3$ and $E\left(\xi_x - x\right)^5$ are not provided because these are not used in the proofs. Lemma 2(i) is a uniform version of Stirling's approximation to the gamma function, whereas Lemma 2(ii) refers to a uniform approximation to the digamma function. Lemma 3 documents uniform bounds of the univariate gamma kernel and its first-order derivative with respect to the design point x. Lemma 4 states Bernstein's inequality.

Lemma 1 *Let* $\xi_x \overset{d}{=} G(x/b+1, b)$. *Then,*

$$\begin{aligned} E\left(\xi_x - x\right) &= b, \\ E\left(\xi_x - x\right)^2 &= xb + 2b^2, \\ E\left(\xi_x - x\right)^4 &= 3x^2b^2 + 26xb^3 + 24b^4, \text{ and} \\ E\left(\xi_x - x\right)^6 &= 15x^3b^3 + 340x^2b^4 + 1044xb^5 + 720b^6. \end{aligned}$$

Lemma 2 *Suppose that sequences* $b(=b(n)), \eta(=\eta(n)) > 0$ *satisfy* $b, \eta \to 0$ *and* $b/\eta \to 0$ *as* $n \to \infty$. *Then, the followings hold true as* $n \to \infty$.

$$\begin{aligned} (i) \quad \sup_{x \in [\eta, \eta^{-1}]} \left| \frac{\Gamma(x/b+1)}{\sqrt{2\pi}(x/b)^{x/b+1/2} \exp(-x/b)} - 1 \right| &= O\left(\frac{b}{\eta}\right). \\ (ii) \quad \sup_{x \in [\eta, \eta^{-1}]} \left| \frac{\Psi(x/b+1) - \ln(x/b)}{b/(2x)} - 1 \right| &= O\left(\frac{b}{\eta}\right). \end{aligned}$$

Lemma 3 *Under the same condition as in Lemma 2, the followings hold true as* $n \to \infty$.

$$\begin{aligned} (i) \quad \sup_{(x,u) \in [\eta, \eta^{-1}] \times \mathbb{R}_+} K_{G(x,b)}(u) &\leq \sqrt{\frac{2}{\pi}} b^{-\frac{1}{2}} \eta^{-\frac{1}{2}}. \\ (ii) \quad \sup_{(x,u) \in [\eta, \eta^{-1}] \times \mathbb{R}_+} \left| \frac{\partial K_{G(x,b)}(u)}{\partial x} \right| &\leq 4\sqrt{\frac{2}{\pi}} b^{-\frac{3}{2}} \eta^{-\frac{3}{2}}. \end{aligned}$$

Lemma 4 (*Van der Vaart and Wellner*, 1996, *Lemma* 2.2.9). *Let* $X_1, \ldots, X_n$ *be independent random variables with bounded ranges* $[-M, M]$ *and zero means. Then,*

$$\Pr\left(\left|\sum_{i=1}^{n} X_i\right| > x\right) \leq 2\exp\left\{-\frac{x^2}{2(v+Mx/3)}\right\}$$

for all x *and* $v \geq Var\left(\sum_{i=1}^{n} X_i\right)$.

A.1.1 Proof of Lemma 1

The formulae can be obtained by computing non-central moments of the gamma random variable. It is easiest and fastest to verify the formulae with the aid of Maple™ or Mathematica®. □

A.1.2 Proof of Lemma 2

By the double inequality for the gamma function in Theorem of Alzer (2003),

$$\left|\frac{\Gamma(x/b+1)}{\sqrt{2\pi}(x/b)^{x/b+1/2}\exp(-x/b)} - 1\right| \leq O\left(\frac{b}{x}\right).$$

Combining the double inequality for the digamma function in Theorem 5 of Gordon (1994) with with a recursive formula $\Psi(z+1) = \Psi(z) + 1/z$ for $z > 0$ also yields

$$\left|\frac{\Psi(x/b+1) - \ln(x/b)}{b/(2x)} - 1\right| \leq O\left(\frac{b}{x}\right).$$

The results immediately follow from $x \geq \eta$. □

A.1.3 Proof of Lemma 3

Proof of (i) Recognize that $u^{x/b}\exp(-u/b)$ is maximized at $u = x$, i.e., x is the mode of the pdf of $G(x/b+1, b)$. Hence, $u^{x/b}\exp(-u/b) \leq x^{x/b}\exp(-x/b)$. It follows from Lemma 2(i) and $b/\eta = o(1)$ that

$$K_{G(x,b)}(u) \leq \frac{x^{x/b}\exp(-x/b)}{b^{x/b+1}\sqrt{2\pi}(x/b)^{x/b+1/2}\exp(-x/b)\{1+o(1)\}} = \frac{b^{-1/2}x^{-1/2}\{1+o(1)\}}{\sqrt{2\pi}}.$$

The result is established by recognizing that the $o(1)$ term (in absolute value) is no greater than 1 for a sufficiently large n and that $x \geq \eta$.

Proof of (ii) We consider the cases of $u = 0$ and $u > 0$ separately. For $u = 0$, it suffices to show that $\lim_{u\uparrow 0} \partial K_{G(x,b)}(u)/\partial x = \lim_{u\downarrow 0} \partial K_{G(x,b)}(u)/\partial x = 0$. If this is the case, then $\partial K_{G(x,b)}(0)/\partial x = 0$ and the result trivially holds. The zero left limit can be immediately established by $K_{G(x,b)}(u) \equiv 0$ for $u < 0$. To evaluate the right limit, observe that for $u > 0$,

$$\begin{aligned}\frac{\partial K_{G(x,b)}(u)}{\partial x} &= \frac{1}{b}\left\{\ln u - \ln b - \Psi\left(\frac{x}{b}+1\right)\right\}K_{G(x,b)}(u)\\ &= \frac{1}{b}\left[(\ln u)K_{G(x,b)}(u) - \left\{\ln b + \Psi\left(\frac{x}{b}+1\right)\right\}K_{G(x,b)}(u)\right].\end{aligned} \tag{7}$$

Because $x/b \geq \eta/b > 0$ is the case for the exponent for u in $K_{G(x,b)}(u)$, we have $\lim_{u\downarrow 0} K_{G(x,b)}(u) = 0$. By L'Hôpital's rule, $\lim_{u\downarrow 0}(\ln u)K_{G(x,b)}(u) = 0$ also holds. Hence, the right limit is shown to be zero.

For $u > 0$, it follows from (7) that

$$\begin{aligned}b\left|\frac{\partial K_{G(x,b)}(u)}{\partial x}\right| &\leq |\ln u| K_{G(x,b)}(u) + \left|\ln b + \Psi\left(\frac{x}{b}+1\right)\right| K_{G(x,b)}(u)\\ &= A_1 + A_2 \text{ (say)}.\end{aligned} \tag{8}$$

We find the bound for A_2 first. By Lemma 2(ii), $b/\eta = o(1)$, $x \in \left[\eta, \eta^{-1}\right]$ and $|\ln z| \leq \max\left\{z, z^{-1}\right\}$ for $z > 0$,

$$\left|\ln b + \Psi\left(\frac{x}{b}+1\right)\right| \leq |\ln x| + o(1) \leq \max\left\{x, x^{-1}\right\} + o(1) \leq \eta^{-1}\{1 + o(\eta)\}.$$

Putting $o(\eta) \leq 1$ for a sufficiently large n and using part (i) of this lemma, we have

$$A_2 \leq 2\sqrt{\frac{2}{\pi}} b^{-1/2}\eta^{-3/2}. \tag{9}$$

Next, we work on A_1. Suppose that $|\ln u| \leq \max\left\{u, u^{-1}\right\} = u^{-1}$. In this case, $u^{-1}u^{x/b}\exp(-u/b)$ is maximized at $u = x - b (\geq \eta - b > 0$ for a sufficiently large $n)$. Then, by Lemma 2(i) and $(1 - b/x)^{x/b} = e^{-1}\{1 + o(1)\}$,

$$\begin{aligned}u^{-1}K_{G(x,b)}(u) &\leq \frac{(x-b)^{x/b-1}\exp\{-(x-b)/b\}}{b^{x/b+1}\sqrt{2\pi}(x/b)^{x/b+1/2}\exp(-x/b)\{1+o(1)\}}\\ &= \frac{b^{-1/2}x^{-3/2}\{1+o(1)\}}{\sqrt{2\pi}(1-b/x)}.\end{aligned}$$

Since $x^{-3/2} \leq \eta^{-3/2}$, $(1 - b/x)^{-1} \leq (1 - b/\eta)^{-1}$ and we may pick $o(1) \leq 1$ and $b/\eta \leq 1/2$ for a sufficiently large n, we finally have $u^{-1}K_{G(x,b)}(u) \leq 2\sqrt{2/\pi}b^{-1/2}\eta^{-3/2}$. Alternatively, when $|\ln u| \leq \max\left\{u, u^{-1}\right\} = u$, a similar procedure also yields $uK_{G(x,b)}(u) \leq 2\sqrt{2/\pi}b^{-1/2}\eta^{-3/2}$, and consequently,

$$A_1 \leq \max\left\{u^{-1}K_{G(x,b)}(u), uK_{G(x,b)}(u)\right\} \leq 2\sqrt{\frac{2}{\pi}} b^{-\frac{1}{2}}\eta^{-\frac{3}{2}}. \tag{10}$$

The proof is completed by substituting (9) and (10) into (8). □

A.1.4 Proof of Theorem 1

For ease of exposition, we additionally introduce the notations

$$
\begin{aligned}
a_n &= \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{d} b_j \eta_j}}},\\
\varsigma_{in}(\mathbf{x}) &= \frac{1}{n}\left[Y_i \mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right) - E\left\{Y_i \mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)\right\}\right],\\
\tau_n &= a_n^{-1/(1+\delta)},\\
\hat{Y}_i &= Y_i \mathbf{1}\left\{\left|Y_i\right| \leq \tau_n\right\},\\
\hat{\varsigma}_{in}(\mathbf{x}) &= \frac{1}{n}\left[\hat{Y}_i \mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right) - E\left\{\hat{Y}_i \mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)\right\}\right],\\
R_n(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n} Y_i \mathbf{1}\left\{\left|Y_i\right| > \tau_n\right\} \mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right), \text{ and}\\
N_n &= a_n^{-\left(1+\frac{1}{1+\delta}\right)} \left(\prod_{j=1}^{d} b_j \eta_j\right)^{-\frac{1}{2}} \left(\sum_{j=1}^{d} \frac{1}{b_j \eta_j}\right).
\end{aligned}
$$

The meaning of each notation will be clarified shortly.

Consider that

$$
\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|\hat{g}_G(\mathbf{x}) - g(\mathbf{x})\right| \leq \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|E\left\{\hat{g}_G(\mathbf{x})\right\} - g(\mathbf{x})\right| + \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|\hat{g}_G(\mathbf{x}) - E\left\{\hat{g}_G(\mathbf{x})\right\}\right|.
$$

The proof is completed if the following two statements are demonstrated.

$$
\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|E\left\{\hat{g}_G(\mathbf{x})\right\} - g(\mathbf{x})\right| = O\left(\sum_{j=1}^{d} b_j\right). \tag{11}
$$

$$
\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|\hat{g}_G(\mathbf{x}) - E\left\{\hat{g}_G(\mathbf{x})\right\}\right| = O_p\left(a_n\right). \tag{12}
$$

Proof of (11) Observe that

$$
\begin{aligned}
E\left\{\hat{g}_G(\mathbf{x})\right\} &= E\left\{E\left(Y_i|\mathbf{X}_i\right)\mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)\right\} = E\left\{m\left(\mathbf{X}_i\right)\mathbb{K}_{G(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)\right\}\\
&= \int_{\mathbb{R}_+^d} m(\mathbf{u}) f(\mathbf{u}) \mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{u}) \mathbf{du} = E\left\{g\left(\boldsymbol{\xi}_{\mathbf{x}}\right)\right\},
\end{aligned} \tag{13}
$$

where the final equality comes from (1) and the gamma random vector $\boldsymbol{\xi}_{\mathbf{x}} := \left(\xi_{x_1}, \ldots, \xi_{x_d}\right)^{\top}$ with $\xi_{x_j} \overset{d}{=} G\left(x_j/b_j + 1, b_j\right)$ and $\xi_{x_j} \perp\!\!\!\perp \xi_{x_k}$ for all $j \neq k$. Then, by a second-order Taylor expansion of $g\left(\boldsymbol{\xi}_{\mathbf{x}}\right)$ around $\boldsymbol{\xi}_{\mathbf{x}} = \mathbf{x}$,

$$
\begin{aligned}
E\{g(\boldsymbol{\xi}_{\mathbf{x}})\} &= g(\mathbf{x}) + \sum_{j=1}^{d} g_j^{(1)}(\mathbf{x}) E\left(\xi_{x_j} - x_j\right) + \frac{1}{2}\sum_{j=1}^{d} g_{jj}^{(2)}(\mathbf{x}) E\left(\xi_{x_j} - x_j\right)^2 \\
&\quad + \sum_{j=1}^{d}\sum_{k=1,k\neq j}^{d} g_{jk}^{(2)}(\mathbf{x}) E\left\{\left(\xi_{x_j} - x_j\right)\left(\xi_{x_k} - x_k\right)\right\} \\
&\quad + \frac{1}{2}\sum_{j=1}^{d} E\left[\left\{g_{jj}^{(2)}(\bar{\mathbf{x}}) - g_{jj}^{(2)}(\mathbf{x})\right\}\left(\xi_{x_j} - x_j\right)^2\right] \\
&\quad + \sum_{j=1}^{d}\sum_{k=1,k\neq j}^{d} E\left[\left\{g_{jj}^{(2)}(\bar{\mathbf{x}}) - g_{jj}^{(2)}(\mathbf{x})\right\}\left(\xi_{x_j} - x_j\right)\left(\xi_{x_k} - x_k\right)\right] \\
&= g(\mathbf{x}) + B_1 + B_2 + B_3 + B_4 + B_5 \text{ (say).}
\end{aligned}
$$

for some $\bar{\mathbf{x}}$ on the line segment joining $\boldsymbol{\xi}_{\mathbf{x}}$ and $\mathbf{x}$.

Using $\left|g_j^{(1)}(\mathbf{x})\right|, \left|g_{jj}^{(2)}(\mathbf{x})x_j\right|, \left|g_{jk}^{(2)}(\mathbf{x})\right| < \infty$ and Lemma 1, we have $|B_1| = O\left(\sum_{j=1}^{d} b_j\right)$, $|B_2| = O\left(\sum_{j=1}^{d} b_j\right)$ and $|B_3| = o\left(\sum_{j=1}^{d} b_j\right)$ uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$. Furthermore, by Assumption 2(ii) and the Cauchy-Schwarz inequality, orders of magnitude in $|B_4|$ and $|B_5|$ are determined by $\sum_{j=1}^{d}\left[E\left\{\left\|\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right\|^2\left(\xi_{x_j} - x_j\right)^4\right\}\right]^{1/2}$ and $\sum_{j=1}^{d}\sum_{k=1,k\neq j}^{d}\left[E\left\{\left\|\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right\|^2\left(\xi_{x_j} - x_j\right)^2\left(\xi_{x_k} - x_k\right)^2\right\}\right]^{1/2}$, respectively. Now, by Lemma 1 and $x_\ell \leq \eta_\ell^{-1}$,

$$
\begin{aligned}
&\left[E\left\{\left\|\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right\|^2\left(\xi_{x_j} - x_j\right)^4\right\}\right]^{\frac{1}{2}} \\
&= \left\{\sum_{\ell=1,\ell\neq j}^{d} E\left(\xi_{x_\ell} - x_\ell\right)^2 E\left(\xi_{x_j} - x_j\right)^4 + E\left(\xi_{x_j} - x_j\right)^6\right\}^{\frac{1}{2}} \\
&= \left[\left\{\sum_{\ell=1,\ell\neq j}^{d} O\left(x_\ell b_\ell\right)\right\} O\left(x_j^2 b_j^2\right) + O\left(x_j^3 b_j^3\right)\right]^{\frac{1}{2}} \\
&= O\left\{\left(x_j b_j\right)^2\left(\sum_{\ell=1}^{d} x_\ell b_\ell\right)\right\}^{\frac{1}{2}} \\
&\leq O\left\{\left(\frac{b_j}{\eta_j}\right)\left(\sum_{\ell=1}^{d}\left(\frac{b_\ell}{\eta_\ell}\right)\right)^{\frac{1}{2}}\right\},
\end{aligned}
$$

and as a result,

$$\sum_{j=1}^{d}\left[E\left\{\left\|\boldsymbol{\xi}_{\mathbf{x}}-\mathbf{x}\right\|^{2}\left(\xi_{x_j}-x_j\right)^{4}\right\}\right]^{\frac{1}{2}}$$
$$\leq O\left\{\sum_{j=1}^{d}\left(\frac{b_j}{\eta_j}\right)\left(\sum_{\ell=1}^{d}\left(\frac{b_\ell}{\eta_\ell}\right)\right)^{\frac{1}{2}}\right\}.$$

Similarly, for $k \neq j$,

$$\left[E\left\{\left\|\boldsymbol{\xi}_{\mathbf{x}}-\mathbf{x}\right\|^{2}\left(\xi_{x_j}-x_j\right)^{2}\left(\xi_{x_k}-x_k\right)^{2}\right\}\right]^{\frac{1}{2}}$$
$$= O\left\{\left(x_j b_j\right)\left(x_k b_k\right)\left(\sum_{\ell=1}^{d} x_\ell b_\ell\right)\right\}^{\frac{1}{2}}$$
$$\leq O\left\{\left(\frac{b_j}{\eta_j}\right)^{\frac{1}{2}}\left(\frac{b_k}{\eta_k}\right)^{\frac{1}{2}}\left(\sum_{\ell=1}^{d}\left(\frac{b_\ell}{\eta_\ell}\right)\right)^{\frac{1}{2}}\right\},$$

and thus

$$\sum_{j=1}^{d}\sum_{k=1,k\neq j}^{d}\left[E\left\{\left\|\boldsymbol{\xi}_{\mathbf{x}}-\mathbf{x}\right\|^{2}\left(\xi_{x_j}-x_j\right)^{2}\left(\xi_{x_k}-x_k\right)^{2}\right\}\right]^{\frac{1}{2}}$$
$$\leq O\left\{\sum_{j=1}^{d}\left(\frac{b_j}{\eta_j}\right)^{\frac{1}{2}}\sum_{k=1,k\neq j}^{d}\left(\frac{b_k}{\eta_k}\right)^{\frac{1}{2}}\left(\sum_{\ell=1}^{d}\left(\frac{b_\ell}{\eta_\ell}\right)\right)^{\frac{1}{2}}\right\}.$$

Finally, Assumption 4W(ii) leads to

$$\frac{b_j}{\eta_j}\left\{\sum_{\ell=1}^{d}\left(\frac{b_\ell}{\eta_\ell}\right)\right\}^{\frac{1}{2}}$$
$$= O\left\{b_j\left(\frac{\rho}{\eta_j^2}\right)^{\frac{1}{2}}\right\} \le O\left\{b_j\left(\frac{\rho}{\min_{1\le\ell\le d}\eta_\ell^2}\right)^{\frac{1}{2}}\right\} = o\big(b_j\big), \text{ and}$$
$$\left(\frac{b_j}{\eta_j}\right)^{\frac{1}{2}}\sum_{k=1,k\neq j}^{d}\left(\frac{b_k}{\eta_k}\right)^{\frac{1}{2}}\left\{\sum_{\ell=1}^{d}\left(\frac{b_\ell}{\eta_\ell}\right)\right\}^{\frac{1}{2}}$$
$$= O\left\{\frac{b_j}{\eta_j}\left(\sum_{\ell=1}^{d}\left(\frac{b_\ell}{\eta_\ell}\right)\right)^{\frac{1}{2}}\right\} = o\big(b_j\big),$$

and thus $|B_4|, |B_5| = o\left(\sum_{j=1}^{d} b_j\right)$ uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$. Therefore, (11) is established.

Proof of (12) As in the proofs of Theorem 2 in Hansen (2008) and Theorem 1 in Hirukawa et al. (2022), our proof takes the following three steps.

1. Demonstrate that the error bound from replacing Y_i with its truncated version $\hat{Y}_i$ is $O_p\big(a_n\big)$ uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$.
2. Split each edge of the d-hyperrectangle $\mathbb{S}_{\mathbf{X}}$ into N_n equally-spaced grids to create N_n^d sub-hyperrectangles, and replace the supremum with a maximization over the finite N_n^d sub-hyperrectangles.
3. Employ Lemma 4 (Bernstein's inequality) to bound the remainder term.

Step 1. A similar argument to the one in Step 1 for the proof of Theorem 1 in Hirukawa et al. (2022) applies here. Recognize that $\hat{g}_G(\mathbf{x}) - E\big\{\hat{g}_G(\mathbf{x})\big\} = \sum_{i=1}^{n} \varsigma_{in}(\mathbf{x})$ and that $\sum_{i=1}^{n} \varsigma_{in}(\mathbf{x}) - \sum_{i=1}^{n} \hat{\varsigma}_{in}(\mathbf{x}) = R_n(\mathbf{x}) - E\big\{R_n(\mathbf{x})\big\}$. It also follows from $|Y_i| > \tau_n$ that $\big(|Y_i|/\tau_n\big)^{1+\delta} > 1$ is the case, and thus

$$\begin{aligned}\left|E\{R_n(\mathbf{x})\}\right| &\le E\{|Y_i|\mathbf{1}\{|Y_i|>\tau_n\}\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\}\\ &\le E\left\{|Y_i|\left(\frac{|Y_i|}{\tau_n}\right)^{1+\delta}\mathbf{1}\{|Y_i|>\tau_n\}\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\}\\ &\le \tau_n^{-(1+\delta)}E\left\{|Y_i|^{2+\delta}\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\}\\ &= \tau_n^{-(1+\delta)}E\left\{E\left(|Y_i|^{2+\delta}\middle|\mathbf{X}_i\right)\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\}\\ &= \tau_n^{-(1+\delta)}\int_{\mathbb{R}_+^d}E\left(|Y|^{2+\delta}\middle|\mathbf{X}=\mathbf{u}\right)f(\mathbf{u})\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{u})\mathbf{du}.\end{aligned}$$

Observe that the right-hand side is bounded by $\tau_n^{-(1+\delta)}C_1$ by Assumption 3 and $\int_{\mathbb{R}_+^d}\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{u})\mathbf{du}=1$. Then, by the definition of τ_n, $\left|E\{R_n(\mathbf{x})\}\right|\le O(a_n)$ uniformly on $\mathbf{x}\in\mathbb{S}_{\mathbf{X}}$. Finally,

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left|\sum_{i=1}^n\varsigma_{in}(\mathbf{x})-\sum_{i=1}^n\hat{\varsigma}_{in}(\mathbf{x})\right|=\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left|R_n(\mathbf{x})-E\{R_n(\mathbf{x})\}\right|=O_p(a_n) \tag{14}$$

can be obtained by Markov's inequality.

Step 2. Let $\mathbf{A}_h$ be the hth sub-hyperrectangle for $h\in\{1,\dots,N_n^d\}$. Also let $\mathbf{x}_h$ be the most distant point from the origin in $\mathbf{A}_h$, i.e., $\mathbf{x}_h:=\arg\max_{\mathbf{x}\in\mathbf{A}_h}\|\mathbf{x}\|$. Suppose that the design point $\mathbf{x}$ falls into $\mathbf{A}_h$. Then, the order of magnitude in $\sup_{\mathbf{x}\in\mathbf{A}_h}\left|\sum_{i=1}^n\hat{\varsigma}_{in}(\mathbf{x})-\sum_{i=1}^n\hat{\varsigma}_{in}(\mathbf{x}_h)\right|$ is determined by $\left|\hat{Y}_i\right|\left|\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)-\mathbb{K}_{G(\mathbf{x}_h,\mathbf{b})}(\mathbf{X}_i)\right|$. Now, by the mean-value theorem,

$$\begin{aligned}\left|\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{u})-\mathbb{K}_{G(\mathbf{x}_h,\mathbf{b})}(\mathbf{u})\right| &= \left|\left\{\mathsf{D}_{\mathbf{x}}\mathbb{K}_{G(\tilde{\mathbf{x}},\mathbf{b})}(\mathbf{u})\right\}^{\top}(\mathbf{x}-\mathbf{x}_h)\right|\\ &\le \sup_{(\mathbf{x},\mathbf{u})\in\mathbf{A}_h\times\mathbb{R}_+^d}\left\|\mathsf{D}_{\mathbf{x}}\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{u})\right\|\sup_{\mathbf{x}\in\mathbf{A}_h}\|\mathbf{x}-\mathbf{x}_h\|\end{aligned}$$

for some $\tilde{\mathbf{x}}$ on the line segment joining $\mathbf{x}$ and $\mathbf{x}_h$. Furthermore, by Lemma 3,

$$\begin{aligned}\sup_{(\mathbf{x},\mathbf{u})\in\mathbf{A}_h\times\mathbb{R}_+^d}\left\|\mathsf{D}_{\mathbf{x}}\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{u})\right\| &= O\left\{\left(\prod_{j=1}^d b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^d\frac{1}{b_j^2\eta_j^2}\right)^{\frac{1}{2}}\right\}\\ &\le O\left\{\left(\prod_{j=1}^d b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^d\frac{1}{b_j\eta_j}\right)\right\}.\end{aligned}$$

It follows from $\sup_{\mathbf{x}\in\mathbf{A}_h}\|\mathbf{x}-\mathbf{x}_h\| = O(N_n^{-1})$ and the definitions of τ_n and N_n that uniformly on $(\mathbf{x},\mathbf{u}) \in \mathbf{A}_h \times \mathbb{R}_+^d$,

$$\begin{aligned}&\left|\hat{Y}_i\right|\left|\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i) - \mathbb{K}_{G(\mathbf{x}_h,\mathbf{b})}(\mathbf{X}_i)\right| \\ &\le O\left\{\tau_n N_n^{-1}\left(\prod_{j=1}^d b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^d \frac{1}{b_j\eta_j}\right)\right\} = O(a_n).\end{aligned}$$

Therefore,

$$\max_{1\le h\le N_n^d}\sup_{\mathbf{x}\in\mathbf{A}_h}\left|\sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x}) - \sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x}_h)\right| = O(a_n). \tag{15}$$

Step 3. Before employing Bernstein's inequality in Lemma 4, we must find two bounds M and v. First, by $\left|\hat{Y}_i\right| \le \tau_n$, Lemma 3 and $\int_{\mathbb{R}_+^d} f(\mathbf{u})\mathbf{du} = 1$,

$$\left|\hat{Y}_i\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right| \le \left(\frac{2}{\pi}\right)^{\frac{d}{2}}\tau_n\left(\prod_{j=1}^d b_j\eta_j\right)^{-\frac{1}{2}}, \text{ and}$$

$$\left|E\left\{\hat{Y}_i\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\}\right| \le E\left\{\left|\hat{Y}_i\right|\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\} \le \left(\frac{2}{\pi}\right)^{\frac{d}{2}}\tau_n\left(\prod_{j=1}^d b_j\eta_j\right)^{-\frac{1}{2}}.$$

It follows from the definitions of a_n and τ_n that

$$\left|\hat{\varsigma}_{in}(\mathbf{x})\right| \le \frac{1}{n}\left[\left|\hat{Y}_i\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right| + \left|E\left\{\hat{Y}_i\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\}\right|\right] \le 2\left(\frac{2}{\pi}\right)^{\frac{d}{2}}\frac{a_n^{2-\frac{1}{1+\delta}}}{\ln n} =: M.$$

Second,

$$\begin{aligned}Var\left\{\sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x})\right\} &= \sum_{i=1}^n Var\left\{\hat{\varsigma}_{in}(\mathbf{x})\right\} \\ &\le \frac{1}{n^2}\sum_{i=1}^n E\left\{\left|\hat{Y}_i\right|\mathbb{K}_{G(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\}^2 \\ &\le \frac{1}{n}\int_{\mathbb{R}_+^d} E\left(|Y|^2\middle|\mathbf{X}=\mathbf{u}\right)f(\mathbf{u})\mathbb{K}_{G(\mathbf{x},\mathbf{b})}^2(\mathbf{u})\mathbf{du}.\end{aligned}$$

By Lyapunov's inequality, (2), (3), and $C_0, C_1 \geq 1$,

$$E\Big(|Y|^2\Big|\mathbf{X}=\mathbf{u}\Big)f(\mathbf{u}) \leq \Big\{E\Big(|Y|^{2+\delta}\Big|\mathbf{X}=\mathbf{u}\Big)f(\mathbf{u})\Big\}^{\frac{2}{2+\delta}}\{f(\mathbf{u})\}^{\frac{\delta}{2+\delta}}$$
$$\leq C_1^{\frac{2}{2+\delta}}C_0^{\frac{\delta}{2+\delta}} \leq C_0C_1.$$

Moreover, by equation (3.2) of Chen (2000),

$$K^2_{G(x,b)}(u) := B_b(x)\frac{u^{2x/b}\exp(-2u/b)}{(b/2)^{2x/b+1}\Gamma(2x/b+1)}\mathbf{1}\{u \geq 0\}, \tag{16}$$

where

$$B_b(x) = \frac{b^{-1}\Gamma(2x/b+1)}{2^{2x/b+1}\Gamma^2(x/b+1)} \leq \frac{b^{-1/2}x^{-1/2}}{2\sqrt{\pi}} \leq \frac{b^{-1/2}\eta^{-1/2}}{2\sqrt{\pi}} \tag{17}$$

by equation (3.4) of Chen (2000) and $x \in \left[\eta, \eta^{-1}\right]$, and the remaining part of (16) is the pdf of $G(2x/b+1, b/2)$. Therefore, by the definition of a_n,

$$Var\left\{\sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x})\right\} \leq \frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}\frac{1}{n\sqrt{\prod_{j=1}^d b_j\eta_j}} = \frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}\frac{a_n^2}{\ln n} =: v.$$

Lemma 4 establishes that for such M and v and an arbitrarily chosen $K > 0$,

$$\Pr\left\{\left|\sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x})\right| > K\sqrt{\frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}a_n}\right\} \leq 2\exp\left[-\frac{K^2\ln n}{2\left\{1+\left(\frac{2}{3}\right)\left(\frac{4}{\sqrt{\pi}}\right)^{d/2}\frac{Ka_n^{1-1/(1+\delta)}}{\sqrt{C_0C_1}}\right\}}\right].$$

Because $a_n = o(1)$ by Assumption 4W(iii), it holds that $(2/3)\left(4/\sqrt{\pi}\right)^{d/2}Ka_n^{1-1/(1+\delta)}/\sqrt{C_0C_1} \leq 1$ for a sufficiently large n. Then,

$$\Pr\left\{\left|\sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x})\right| > K\sqrt{\frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}a_n}\right\} \leq 2\exp\left\{-\frac{K^2\ln n}{2(1+1)}\right\} = 2n^{-\frac{K^2}{4}}.$$

In the end,

$$
\begin{aligned}
&\Pr\left\{\max_{1\le h\le N_n^d}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| > K\sqrt{\frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}}a_n\right\}\\
&\le \sum_{h=1}^{N_n^d}\Pr\left\{\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| > K\sqrt{\frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}}a_n\right\} \qquad (18)\\
&\le N_n^d \times \max_{1\le h\le N_n^d}\Pr\left\{\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| > K\sqrt{\frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}}a_n\right\}\\
&= O\left(N_n^d n^{-K^2/4}\right).
\end{aligned}
$$

Pick $K = 2\sqrt{5d}$. Then, by the definition of N_n,

$$
N_n^d n^{-\frac{K^2}{4}} = \left\{a_n^{-\left(1+\frac{1}{1+\delta}\right)}\left(\prod_{j=1}^{d} b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{d}\frac{1}{b_j\eta_j}\right)n^{-5}\right\}^d. \qquad (19)
$$

It also follows from the definition of a_n that $n^{-1} = (\ln n)^{-1}a_n^2\left(\prod_{j=1}^{d} b_j\eta_j\right)^{1/2}$. Substituting this into the right-hand side of (19) finally yields

$$
N_n^d n^{-\frac{K^2}{4}} = \left[\frac{a_n^{8+\frac{\delta}{1+\delta}}}{\ln^5 n}\left(\prod_{j=1}^{d} b_j\eta_j\right)\left\{\sum_{j=1}^{d}\left(\prod_{k=1,k\neq j}^{d} b_k\eta_k\right)\right\}\right]^d \to 0
$$

as $n \to \infty$, which demonstrates that

$$
\max_{1\le h\le N_n^d}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| = O_p(a_n). \qquad (20)
$$

Now, (14), (15) and (20) establish (12). This completes the proof. □

A.2 Proof of Theorem 2

All notations in the proof of Theorem 1 are maintained, except that the definitions of τ_n and N_n are changed to $\tau_n = n^{(1+\epsilon)/(2+\delta)}$ and

$$N_n = n^{1+\epsilon}\left(\prod_{j=1}^{d} b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{d}\frac{1}{b_j\eta_j}\right)$$

for an arbitrarily small $\epsilon > 0$. Then, the proof is boiled down to demonstrating that $\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}}\left|\hat{g}_G(\mathbf{x}) - E\{\hat{g}_G(\mathbf{x})\}\right| = O(a_n)\, a.s.$

This proof also takes three steps as in the proof of (12). First, by the argument in Step 1 for the proof of (12) and the definitions of τ_n and a_n,

$$\left|E\{R_n(\mathbf{x})\}\right| \le \tau_n^{-(1+\delta)}C_1 = n^{-(1+\epsilon)\left(\frac{1+\delta}{2+\delta}\right)}C_1 \le O\left(n^{-1/2}\right) \le o(a_n) \le O(a_n). \quad (21)$$

Also by Markov's inequality and Assumption 3,

$$\sum_{n=1}^{\infty}\Pr\left(|Y_n| > \tau_n\right) < \sum_{n=1}^{\infty}\frac{E|Y|^{2+\delta}}{\tau_n^{2+\delta}} = E|Y|^{2+\delta}\sum_{n=1}^{\infty}\frac{1}{n^{1+\epsilon}} < \infty.$$

Then, by the Borel-Cantelli lemma, for a sufficiently large n, $|Y_n| \le \tau_n$ with probability 1. This implies that $|Y_i| \le \tau_n$ for any $i \le n$ with probability 1 for a sufficiently large n. It follows that $R_n(\mathbf{x}) = 0$ with probability 1, i.e.,

$$\left|R_n(\mathbf{x}) - E\{R_n(\mathbf{x})\}\right| = \left|\sum_{i=1}^{n}\varsigma_{in}(\mathbf{x}) - \sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x})\right| = O(a_n)a.s. \quad (22)$$

uniformly on $\mathbf{x} \in \mathbb{S}_\mathbf{X}$.

Second, it follows from the argument in Step 2 for the proof of (12) and the definitions of τ_n and N_n that

$$\begin{aligned}&\max_{1\le h\le N_n^d}\sup_{\mathbf{x}\in\mathbf{A}_h}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}) - \sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| \\ &= O\left\{\tau_n N_n^{-1}\left(\prod_{j=1}^{d} b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{d}\frac{1}{b_j\eta_j}\right)\right\} = O\left\{n^{-(1+\epsilon)\left(\frac{1+\delta}{2+\delta}\right)}\right\}.\end{aligned}$$

Then, using (21) yields

$$\max_{1\le h\le N_n^d}\sup_{\mathbf{x}\in\mathbf{A}_h}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}) - \sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| = O(a_n). \quad (23)$$

Third, (18) holds for a sufficiently large n. In addition, (4) can be rearranged to yield

$$\frac{\ln n}{n\left(\prod_{j=1}^{d} b_j\eta_j\right)^{\frac{\kappa}{2}}}\left\{\left(\prod_{j=1}^{d} b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{d}\frac{1}{b_j\eta_j}\right)\right\}^{1-\kappa} = O(1).$$

Accordingly,

$$\left(\prod_{j=1}^{d} b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{d}\frac{1}{b_j\eta_j}\right) = O\left\{n^{\frac{1}{1-\kappa}}\left(\frac{\left(\prod_{j=1}^{d} b_j\eta_j\right)^{\frac{\kappa}{2}}}{\ln n}\right)^{\frac{1}{1-\kappa}}\right\} \leq O\left(n^{\frac{1}{1-\kappa}}\right),$$

where the last inequality holds because $\left(\prod_{j=1}^{d} b_j\eta_j\right)^{\kappa/2}/\ln n$ is convergent. Then, picking $K = 2\sqrt{(d+1)(1+\epsilon)+d/(1-\kappa)}$ yields $N_n^d n^{-K^2/4} = O\left\{n^{-(1+\epsilon)}\right\}$ so that

$$\sum_{n=1}^{\infty}\Pr\left\{\max_{1\leq h\leq N_n^d}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| > K\sqrt{\frac{C_0C_1}{\left(2\sqrt{\pi}\right)^d}a_n}\right\} \leq \sum_{n=1}^{\infty} O\left\{n^{-(1+\epsilon)}\right\} < \infty.$$

Therefore, by the Borel-Cantelli lemma,

$$\max_{1\leq h\leq N_n^d}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}(\mathbf{x}_h)\right| = O(a_n)a.s. \tag{24}$$

The stated result is established by (22), (23) and (24). This completes the proof. □

A.3 Proofs of Theorem 3 and 4

Put $Y_i \equiv 1$ in Theorems 1 and 2, respectively. □

A.4 Proofs of Theorem 5 and 6

Below we concentrate on the proof of Theorem 5. Switching the arguments based on Theorem 1 to those on Theorem 2 can immediately establish Theorem 6, and thus details are omitted.

Proof of (5) By Theorem 1,

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{x}}\left|\frac{\hat{g}_G(\mathbf{x})}{f(\mathbf{x})} - m(\mathbf{x})\right| \leq \frac{\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{x}}\left|\hat{g}_G(\mathbf{x}) - g(\mathbf{x})\right|}{\inf_{\mathbf{x}\in\mathbb{S}_\mathbf{x}} f(\mathbf{x})} = O_p\left\{r_n^{-1}\left(\sum_{j=1}^{d} b_j + a_n\right)\right\} \tag{25}$$

and

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{x}}\left|\frac{\hat{f}_G(\mathbf{x})}{f(\mathbf{x})} - 1\right| \leq \frac{\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{x}}\left|\hat{f}_G(\mathbf{x}) - f(\mathbf{x})\right|}{\inf_{\mathbf{x}\in\mathbb{S}_\mathbf{x}} f(\mathbf{x})} = O_p\left\{r_n^{-1}\left(\sum_{j=1}^{d} b_j + a_n\right)\right\} \tag{26}$$

for a_n defined in the proof of Theorem 1. Therefore,

$$\begin{aligned}\hat{m}_G(\mathbf{x}) &= \frac{\hat{g}_G(\mathbf{x})/f(\mathbf{x})}{\hat{f}_G(\mathbf{x})/f(\mathbf{x})} \\ &= \frac{m(\mathbf{x}) + O_p\left\{r_n^{-1}\left(\sum_{j=1}^d b_j + a_n\right)\right\}}{1 + O_p\left\{r_n^{-1}\left(\sum_{j=1}^d b_j + a_n\right)\right\}} \\ &= m(\mathbf{x}) + O_p\left\{r_n^{-1}\left(\sum_{j=1}^d b_j + a_n\right)\right\}\end{aligned}$$

uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$.

Proof of (6) The proof requires to find uniform bounds of $\left\|\mathbf{S}_1(\mathbf{x})/f(\mathbf{x})\right\|$, $\left\|\mathbf{S}_2(\mathbf{x})/f(\mathbf{x})\right\|$ and $\left\|\mathbf{T}_1(\mathbf{x})/f(\mathbf{x})\right\|$. Our derivation starts from the third one. A similar argument to (13) yields $E\left\{\mathbf{T}_1(\mathbf{x})\right\} = E\left\{\left(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right)g\left(\boldsymbol{\xi}_{\mathbf{x}}\right)\right\}$ for $\boldsymbol{\xi}_{\mathbf{x}}$ defined therein. Then, by a first-order Taylor expansion of $g\left(\boldsymbol{\xi}_{\mathbf{x}}\right)$ around $\boldsymbol{\xi}_{\mathbf{x}} = \mathbf{x}$, we have $g\left(\boldsymbol{\xi}_{\mathbf{x}}\right) = g(\mathbf{x}) + \sum_{j=1}^d g_j^{(1)}(\check{\mathbf{x}})\left(\xi_{x_j} - x_j\right)$ for some $\check{\mathbf{x}}$ on the line segment joining $\boldsymbol{\xi}_{\mathbf{x}}$ and $\mathbf{x}$, and thus

$$E\left\{\mathbf{T}_1(\mathbf{x})\right\} = g(\mathbf{x})E\left(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right) + \sum_{j=1}^d E\left\{g_j^{(1)}(\check{\mathbf{x}})\left(\xi_{x_j} - x_j\right)\left(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right)\right\}.$$

By (2) and Lyapunov's inequality, $|g(\mathbf{x})| \leq \sup_{\mathbf{x}\in\mathbb{R}_+^d} E(|Y||\mathbf{X} = \mathbf{x})f(\mathbf{x}) < \infty$. It also follows from Lemma 1 and $x \in \left[\eta, \eta^{-1}\right]$

$$\left\|E\left(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right)\right\| \leq O\left\{\left(\sum_{j=1}^d b_j^2\right)^{\frac{1}{2}}\right\} = O\left\{\rho\left(\sum_{j=1}^d \eta_j^2\right)^{\frac{1}{2}}\right\} = o(\rho),$$

so that $\left\|g(\mathbf{x})E\left(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right)\right\| = o(\rho)$ uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$. Also by $\left|g_j^{(1)}(\mathbf{x})\right| < \infty$, the Cauchy-Schwarz inequality, Lemma 1, and Assumption 4W(ii),

$$\begin{aligned}&\left\|\sum_{j=1}^d E\left\{g_j^{(1)}(\check{\mathbf{x}})\left(\xi_{x_j} - x_j\right)\left(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right)\right\}\right\| \\ &\leq O\left\{\sum_{j=1}^d \left(E\left(\left(\xi_{x_j} - x_j\right)^2\left\|\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right\|^2\right)\right)^{\frac{1}{2}}\right\} = O(\rho)\end{aligned}$$

uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$. Therefore, $\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|E\left\{\mathbf{T}_1(\mathbf{x})\right\}\right\| = O(\rho)$. Moreover, choosing M and v in Lemma 4 suitably, we can show that $\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\mathbf{T}_1(\mathbf{x}) - E\left\{\mathbf{T}_1(\mathbf{x})\right\}\right\| = O_p\left(\rho^{1/2}a_n\right)$. Hence,

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\mathbf{T}_1(\mathbf{x})\right\| = O(\rho) + O_p\left(\rho^{\frac{1}{2}} a_n\right) = O_p\left\{\rho^{\frac{1}{2}}\left(\rho^{\frac{1}{2}} + a_n\right)\right\}.$$

As a result,

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\frac{\mathbf{T}_1(\mathbf{x})}{f(\mathbf{x})}\right\| \leq \frac{\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\mathbf{T}_1(\mathbf{x})\right\|}{\inf_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} f(\mathbf{x})} = O_p\left\{r_n^{-1}\rho^{\frac{1}{2}}\left(\rho^{\frac{1}{2}} + a_n\right)\right\}. \tag{27}$$

Replacing g in $E\{\mathbf{T}_1(\mathbf{x})\}$ with f also yields

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\frac{\mathbf{S}_1(\mathbf{x})}{f(\mathbf{x})}\right\| = O_p\left\{r_n^{-1}\rho^{\frac{1}{2}}\left(\rho^{\frac{1}{2}} + a_n\right)\right\}. \tag{28}$$

Furthermore, we may write $E\{\mathbf{S}_2(\mathbf{x})\} = E\left\{(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})^\top f(\boldsymbol{\xi}_{\mathbf{x}})\right\}$. Then, by a first-order Taylor expansion of $f(\boldsymbol{\xi}_{\mathbf{x}})$ around $\boldsymbol{\xi}_{\mathbf{x}} = \mathbf{x}$,

$$\begin{aligned} E\{\mathbf{S}_2(\mathbf{x})\} =& f(\mathbf{x})\Big[E\left\{(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})^\top\right\} \\ &+ \frac{1}{f(\mathbf{x})}\sum_{j=1}^{d} E\left\{f_j^{(1)}(\check{\mathbf{x}})(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})^\top\left(\xi_{x_j} - x_j\right)\right\}\Big] \end{aligned}$$

for some $\check{\mathbf{x}}$ on the line segment joining $\boldsymbol{\xi}_{\mathbf{x}}$ and $\mathbf{x}$. Again, by $\left|f_j^{(1)}(\mathbf{x})\right| < \infty$, the Cauchy-Schwarz inequality, Lemma 1, and Assumption 4W(ii),

$$\begin{aligned} &\left\|\sum_{j=1}^{d} E\left\{f_j^{(1)}(\check{\mathbf{x}})(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})(\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x})^\top\left(\xi_{x_j} - x_j\right)\right\}\right\| \\ &\leq O\left\{\sum_{j=1}^{d}\left(E\left(\left\|\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right\|^4\left(\xi_{x_j} - x_j\right)^2\right)\right)^{\frac{1}{2}}\right\} = O\left(\rho^{\frac{3}{2}}\right) \end{aligned}$$

uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$. Using (3), $\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} E\left\|\boldsymbol{\xi}_{\mathbf{x}} - \mathbf{x}\right\|^2 = O(\rho)$ and $1/f(\mathbf{x}) \leq 1/\inf_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} f(\mathbf{x}) = r_n^{-1}$, we have

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\frac{E\{\mathbf{S}_2(\mathbf{x})\}}{f(\mathbf{x})}\right\| = O(\rho) + O\left(r_n^{-1}\rho^{\frac{3}{2}}\right).$$

Appropriate choices of M and ν in Lemma 4 also yield $\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\mathbf{S}_2(\mathbf{x}) - E\{\mathbf{S}_2(\mathbf{x})\}\right\| = O_p(\rho a_n)$, and thus

$$\begin{aligned} \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\frac{\mathbf{S}_2(\mathbf{x})}{f(\mathbf{x})}\right\| &\leq \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\frac{E\{\mathbf{S}_2(\mathbf{x})\}}{f(\mathbf{x})}\right\| + \frac{\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left\|\mathbf{S}_2(\mathbf{x}) - E\{\mathbf{S}_2(\mathbf{x})\}\right\|}{\inf_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} f(\mathbf{x})} \\ &= O(\rho) + O\left(r_n^{-1}\rho^{3/2}\right) + O_p\left(r_n^{-1}\rho a_n\right). \end{aligned}$$

Assumption 4 implies that $r_n^{-1}\rho^{1/2} = \left(r_n^{-2}\rho\right)^{1/2} = o(1)$, and thus $r_n^{-1}\rho^{3/2} = o(\rho)$. Since $r_n^{-1}a_n = \left(r_n^{-2}a_n\right)^{1/2}a_n^{1/2} = o(1)$, we also have $r_n^{-1}\rho a_n = o(\rho)$. As a consequence,

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{S}_2(\mathbf{x})}{f(\mathbf{x})}\right\| = O_p(\rho). \tag{29}$$

It follows from (27), (28) and (29) that

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{T}_1(\mathbf{x})}{f(\mathbf{x})}\right\| = O_p\left\{r_n^{-2}\left(\rho^{\frac{1}{2}}+a_n\right)^2\right\} = O_p\left\{r_n^{-2}\left(\rho+a_n\right)\right\},$$

and

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{S}_1(\mathbf{x})}{f(\mathbf{x})}\right\| = O_p\left\{r_n^{-2}\left(\rho+a_n\right)\right\}.$$

Using (25) and (26) and recognizing that $r_n^{-1}\left(\sum_{j=1}^d b_j + a_n\right) = o\left\{r_n^{-2}\left(\rho+a_n\right)\right\}$, we may conclude that

$$\begin{aligned}\tilde{m}_G(\mathbf{x}) &= \frac{\hat{g}_G(\mathbf{x})/f(\mathbf{x}) - \mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{T}_1(\mathbf{x})/f(\mathbf{x})}{\hat{f}_G(\mathbf{x})/f(\mathbf{x}) - \mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{S}_1(\mathbf{x})/f(\mathbf{x})}\\ &= \frac{m(\mathbf{x}) + O_p\left\{r_n^{-1}\left(\sum_{j=1}^d b_j + a_n\right)\right\} + O_p\left\{r_n^{-2}\left(\rho+a_n\right)\right\}}{1 + O_p\left\{r_n^{-1}\left(\sum_{j=1}^d b_j + a_n\right)\right\} + O_p\left\{r_n^{-2}\left(\rho+a_n\right)\right\}}\\ &= m(\mathbf{x}) + O_p\left\{r_n^{-2}\left(\rho+a_n\right)\right\}\end{aligned}$$

uniformly on $\mathbf{x}\in\mathbb{S}_{\mathbf{X}}$. □

Funding This work was supported by the Japan Society for the Promotion of Science under the grant number 23K01340 (M. Hirukawa).

Declaration of Interest

Conflict of interest The authors report that there are no Conflict of interest to declare.

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates - a square root law. *Annals of Statistics, 10*, 1217–1223.

Alzer, H. (2003). On Ramanujan's double inequality for the gamma function. *Bulletin of the London Mathematical Society, 35*, 601–607.

Bertin, K., Genest, C., Klutchnikoff, N. and Ouimet, N. (2023). Minimax properties of Dirichlet kernel density estimators, *Journal of Multivariate Analysis*, *195*, Article No.105158.

Bouezmarni, T., Rombouts, J. V. K. (2010). Nonparametric density estimation for multivariate bounded data. *Journal of Statistical Planning and Inference, 140*, 139–152.
Bouzebda, S., Nezzal, A., Elhattab, I. (2024). Limit theorems for nonparametric conditional U-statistics smoothed by asymmetric kernels. *AIMS Mathematics, 9*, 26195–26282.
Chakraborty, A. K., Chatterjee, M. (2013). On multivariate folded normal distribution, *Sankhyā: the indian journal of statistics*, *Series B, 75*, 1–15.
Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis, 31*, 131–145.
Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics, 52*, 471–480.
Chen, S. X. (2002). Local linear smoothers using asymmetric kernels. *Annals of the Institute of Statistical Mathematics, 54*, 312–323.
Das, S., Dey, D. K. (2010). On bayesian inference for generalized multivariate gamma distribution. *Statistics & Probability Letters, 80*, 1492–1499.
Funke, B., Hirukawa, M. (2024a). Density derivative estimation using asymmetric kernels. *Journal of Nonparametric Statistics, 36*, 994–1017.
Funke, B., Hirukawa, M. (2024b). Uniform Convergence Rates for Density Derivative Estimators Using Asymmetric Kernels, Working Paper.
Funke, B., Kawka, R. (2015). Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods, *Computational Statistics & Data Analysis*, 92, 148-162.
Genest, C., Ouimet, F. (2024). Local linear smoothing for regression surfaces on the simplex using dirichlet kernels. arXiv preprint arXiv:2408.07209v1
Gordon, L. (1994). A stochastic approach to the gamma function. *American Mathematical Monthly, 101*, 858–865.
Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory, 24*, 726–748.
Hirukawa, M. (2018). *Asymmetric kernel smoothing: theory and applications in economics and finance*. Singapore: Springer.
Hirukawa, M., Murtazashvili, I., Prokhorov, A. (2022). Uniform convergence rates for nonparametric estimators smoothed by the beta kernel. *Scandinavian Journal of Statistics, 49*, 1353–1382.
Hirukawa, M., Murtazashvili, I., Prokhorov, A. (2023). Yet another look at the omitted variable bias. *Econometric Reviews, 42*, 1–27.
Horrace, W. C. (2005). Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis, 94*, 209–221.
Karunamuni, R. J., Alberts, T. (2005). A generalized reflection method of boundary correction in kernel density estimation. *Canadian Journal of Statistics, 33*, 497–509.
Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogenous dependent data. *Econometric Theory, 25*, 1433–1445.
Liu, Z., Li, M. (2023). On the estimation of derivatives using plug-in kernel ridge regression estimators. *Journal of Machine Learning Research, 24*, 1–37.
Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory, 10*, 233–253.
Ouimet, F. (2021). Asymptotic properties of Bernstein estimators on the simplex, *Journal of Multivariate Analysis*, 185, Article No.104874.
Ouimet, F. (2022). A symmetric matrix-variate normal local approximation for the Wishart distribution and some applications, *Journal of Multivariate Analysis*, 189, Article No.104923.
Ouimet, F., Tolosana-Delgado, R. (2022). Asymptotic properties of Dirichlet kernel density estimators, *Journal of Multivariate Analysis*, 187, Article No.104832.
Racine, J. S. (2019). *An introduction to the advanced theory of nonparametric econometrics: A replicable approach using R*. Cambridge, UK: Cambridge University Press.
Rilstone, P. (1996). Nonparametric estimation of models with generated regressors. *International Economic Review, 37*, 299–313.
Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica, 56*, 931–954.
Ruppert, D., Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics, 22*, 1346–1370.
Stengos, T., Yan, B. (2001). Double kernel nonparametric estimation in semiparametric econometric models. *Journal of Nonparametric Statistics, 13*, 883–906.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics, 10*, 1040–1053.

Stone, C. J. (1983). Optimal Global Rates of Convergence for Nonparametric Regression. In M. H. Rizvi, J. S. Rustagi, D. Siegmund (eds.), *Recent advances in statistics: Papers in honor of Herman Chernoff on his sixtieth birthday* (pp. 393–406). New York: Academic Press.

Terrell, G. R., Scott, D. W. (1992). Variable kernel density estimation. *Annals of Statistics, 20*, 1236–1265.

Van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics*. New York: Springer-Verlag.