Uniform Convergence Rates for Nonparametric Estimators Smoothed by the Beta Kernel*

Masayuki Hirukawa[†]
Ryukoku University

Irina Murtazashvili[‡]
Drexel University

Artem Prokhorov[§]
University of Sydney,
CEBA and CIREQ

August 6, 2021

Abstract

This paper provides a set of uniform consistency results with rates for nonparametric density and regression estimators smoothed by the beta kernel having support on the unit interval. Weak and strong uniform convergence is explored on the basis of expanding compact sets and general sequences of smoothing parameters. The results in this paper are useful for asymptotic analysis of two-step semiparametric estimation using a first-step kernel estimate as a plug-in. We provide simulations and a real data example illustrating attractive properties of the estimator.

JEL Classification Codes: C13; C14.

MSC 2000 Codes: 62G7; 62G8; 62G20.

Keywords: Beta kernel; boundary bias; nonparametric density estimation; nonparametric regression estimation; rates of convergence.

*The authors would like to thank the editor, an associate editor, two anonymous referees, and Bruce Hansen for their constructive comments and suggestions. Financial support through grants from Japan Society for the Promotion of Science (M. Hirukawa, Project No. 19K01595) and the Russian Science Foundation (A. Prokhorov, Project No. 20-18-00365) for various and non-overlapping parts of this research is gratefully acknowledged.

[†]e-mail: hirukawa@econ.ryukoku.ac.jp.

[‡]e-mail: im99@drexel.edu.

[§]e-mail: artem.prokhorov@sydney.edu.au.

1 Introduction

It is well known that using standard symmetric kernels to estimate unknown curves on a bounded support (e.g., on a half real line or on a compact set) leads to the so-called boundary bias in the vicinity of the boundary. While many boundary correction methods have been proposed since early works on boundary adapted kernel estimation by Müller (1991) and Jones (1993), smoothing by a nonstandard, asymmetric kernel function emerges as a viable alternative. Because an asymmetric kernel is based on a probability density function (pdf) having the same support as that of the curves, it is free of the boundary bias by construction. In addition, the shape of the kernel varies according to the position at which smoothing is done; in other words, the amount of smoothing changes in an adaptive manner. More than two decades have passed since the advent of asymmetric kernels, and a number of articles have reported favorable evidence from applying them to empirical models in economics and finance;see Hirukawa (2018) for more details on asymmetric kernels and examples of their applications. However, little is known about uniform consistency and convergence rates of asymmetric kernel estimators.

This paper studies uniform convergence for nonparametric estimators on a compact set smoothed by an asymmetric kernel. Compactness of the support often arises either by construction of the data or as a theoretical requirement. For the former, economic and financial variables defined as shares or proportions are typically bounded from above and below. Examples include expenditure and budget shares, unemployment rates, target zone exchange rates, and default and recovery rates, to name a few. For the latter, compactness is imposed, for instance, on the support of nonparametric copulas (see, e.g., Sancetta and Satchell, 2004), of the nonparametric part of the partial linear regressions (see, e.g., Yatchew, 1997) and of the covariates used for

nearest-neighbor matching (see, e.g., Abadie and Imbens, 2006). Furthermore, the fully nonparametric estimator of first-price auctions is also built on compactness of supports of the distributions of private values and observed bids (see, e.g., Guerre, Perrigne and Vuong, 2000).

Throughout this paper it is assumed, without loss of generality, that the compact set is a p-dimensional unit hypercube $[0,1]^p$. Among all asymmetric kernels, our particular focus is on the beta kernel by Chen (1999). The kernel takes the form

$$K_{B(x,b)}(u) = \frac{u^{x/b}(1-u)^{(1-x)/b}}{B\{x/b+1,(1-x)/b+1\}}\mathbf{1}\{u \in [0,1]\}$$

for a data point $u \in [0,1]$, a design point $x \in [0,1]$ and a smoothing parameter $b > 0$, where $B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}dy$ for $\alpha,\beta > 0$ is the beta function, and $\mathbf{1}\{\cdot\}$ denotes an indicator function. To cope with multivariate problems, we construct a tensor product kernel

$$\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u}) = \prod_{j=1}^{p} K_{B(x_j,b_j)}(u_j) = \prod_{j=1}^{p} \frac{u_j^{x_j/b_j}(1-u_j)^{(1-x_j)/b_j}}{B\{x_j/b_j+1,(1-x_j)/b_j+1\}}\mathbf{1}\{u_j \in [0,1]\},$$

where $\mathbf{u} := (u_1,\ldots,u_p)^\top \in [0,1]^p$, $\mathbf{x} := (x_1,\ldots,x_p)^\top \in [0,1]^p$ and $\mathbf{b} := (b_1,\ldots,b_p)^\top \in \mathbb{R}^p_{++}$ are p-dimensional vectors of data points, design points and smoothing parameters, respectively.

Our analysis is built on estimating $g(\mathbf{x}) := m(\mathbf{x})f(\mathbf{x})$, where $m(\mathbf{x}) := E(Y|\mathbf{X}=\mathbf{x})$ and $f(\mathbf{x})$ is the marginal pdf of $\mathbf{X}$. Given n *i.i.d.* observations $\{(Y_i,\mathbf{X}_i)\}_{i=1}^n \in \mathbb{R}\times[0,1]^p$, we consider the sample average estimator

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

and demonstrate its weak and strong uniform convergences with rates. On the one hand, most of multivariate beta kernel estimators such as the joint density estimator and the Nadaraya-Watson (or local constant) and local linear regression estimators can be expressed in this form. Consequently, our convergence results can be applied to each of these estimators. On the other hand, our focus in this paper is on uniform

consistency of $\hat{g}_B(\mathbf{x})$ on a p-dimensional hyperrectangle inside $[0,1]^p$ that is either fixed or expanding to $[0,1]^p$ at a suitable rate. It may be possible to demonstrate weak and strong uniform convergences of $\hat{g}_B(\mathbf{x})$ on $(0,1)^p$ after suitable modifications of regularity conditions but we do not pursue this here. In this case, an inferior variance convergence rate would dominates (as implied by pointwise convergence discussed in Section 4.1.1), and thus the results corresponding to Theorems 7 and 8 of this paper cannot be readily used in analyses of the type considered in our companion paper (Hirukawa, Murtazashvili and Prokhorov, 2021). It is also difficult to relate the final forms of the results with the optimal global rates for nonnegative kernel estimators derived by Stone (1982, 1983). We return to this point in the simulation section.

This paper has several contributions to the existing literature. First, uniform convergence for asymmetric kernel estimators on expanding sets has not been formally established in the literature known to us. There is rich literature investigating uniform convergence on expanding or unbounded sets for estimators based on standard symmetric kernels, including multivariate frameworks (e.g., Hansen, 2008; Kristensen, 2009). However, the univariate beta kernel cannot be expressed in the form $(1/b) K\{(u-x)/b\}$ suitable for symmetric kernels. Therefore, it is worth emphasizing that although our proofs take the same steps as in Hansen (2008), they rely on different techniques from those for symmetric kernels.

Second, our results hold under weaker regularity conditions than the existing ones in the literature. While some uniform consistency results are available for asymmetric kernel based estimators, they have important limitations. Bouezmarni and Rolin (2003) show weak and strong uniform consistency of the univariate beta kernel density estimator on $[0,1]$, but convergence rates are not provided explicitly. As the most relevant results to ours, Shi and Song (2016) and Koul and Song (2013) demonstrate strong uniform consistency with rates for univariate density and Nadaraya-Watson regression estimators smoothed by the gamma kernel (Chen, 2000) with support on

$\mathbb{R}_+$ and its variant, respectively. However, their focus is exclusively on uniform convergences on a fixed compact interval.

Third, our uniform results cover multivariate nonparametric regression estimation using asymmetric kernels. There are only a few works on asymmetric kernel estimation for two or more dimensional cases. Examples include Bouezmarni and Rombouts (2010) and Funke and Kawka (2015), and they deal exclusively with joint density estimation. We could find no work on asymmetric kernel regression estimation with multiple regressors.

Fourth, the beta kernel estimators concerned in this paper are allowed to employ different smoothing parameters for different dimensions. Hansen (2008) and Kristensen (2009) also derive uniform convergence results of sample average estimators using product kernels. While they establish the results on expanding sets like ours, they apply a single bandwidth parameter to all dimensions.

It appears that so far, applications of the beta kernel have been limited to purely nonparametric estimation problems like density and regression estimations. As suggested by Kanaya and Bhattacharya (2017), a lack of uniform convergence results for beta kernel estimators may have precluded researchers from employing the beta kernel to a wide variety of estimation problems. On the other hand, several authors report that beta kernel estimators exhibit attractive finite-sample properties for diverse applications. Examples include Renault and Scaillet (2004) and Hagmann, Renault and Scaillet (2005) for recovery rate density estimation, Kristensen (2010) for realized integrated volatility estimation, and our companion paper (Hirukawa, Murtazashvili and Prokhorov, 2021) for two-step two-sample semiparametric regression estimation.

The results in this paper may encourage application of the beta kernel in empirical economics and finance. For example, the beta kernel is readily applicable to uniform inference on nonparametric density and regression estimations such as those for first-price auctions and Lorenz curves. It can be also employed for the two-step

semiparametric estimation with a first-step nonparametric kernel-based plug-in estimate. Such a setting has been used by Robinson (1988), Newey (1994), Rilstone (1996), and Stengos and Yan (2001). Hirukawa, Murtazashvili and Prokhorov (2021) directly apply Theorem 7 as a building block of their two-step two-sample semiparametric regression estimator using the product beta kernel for the continuous part of the first-step nonparametric estimate.

The remainder of this paper is organized as follows. Section 2 delivers weak and strong uniform convergence results for the sample average estimator $\hat{g}_B(\mathbf{x})$ on a compact set that is either fixed or expanding to a unit hypercube. Section 3 applies the results to density and regression estimators. In the end, our analysis is extended to the nonparametric regression estimator with mixed categorical and continuous regressors à la Racine and Li (2004). Section 4 conducts Monte Carlo simulations and provides a real data application of beta kernel estimators. Section 5 concludes. All proofs are given in the Appendix.

The paper adopts the following notational conventions: for $a > 0$, $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)\, dt$ is the gamma function; for $\mathbf{z} \in \mathbb{R}^d$, $\nabla \{h(\mathbf{z})\}$ signifies a d-column vector of the first-order derivatives of a function $h(\mathbf{z})$; and "*a.s.*" denotes "almost surely". The expression '$X \overset{d}{=} Y$' reads "A random variable X obeys the distribution Y." Finally, we mean by $\|\mathbf{A}\|$ the Frobenius norm of matrix $\mathbf{A}$, i.e., $\|\mathbf{A}\| = \left\{\text{tr}\left(\mathbf{A}^\top \mathbf{A}\right)\right\}^{1/2}$.

2 Main Results

2.1 Weak Uniform Convergence of the Sample Average Estimator

Our analysis starts from demonstrating weak uniform consistency with rates of the sample average estimator $\hat{g}_B(\mathbf{x})$ for $g(\mathbf{x})$ on a p-hyperrectangle

$$\mathbb{S}_{\mathbf{X}} = \mathbb{S}_{\mathbf{X}}(\eta) := \prod_{j=1}^{p} [\eta_j, 1-\eta_j] \subseteq [0,1]^p,$$

where the boundary parameters $\eta := (\eta_1, \ldots, \eta_p)^\top$ either are fixed or shrink to zero at a suitable rate. To deliver the results, we impose the following regularity conditions.

Assumption 1. $\{(Y_i, \mathbf{X}_i)\}_{i=1}^n \in \mathbb{R} \times [0,1]^p$ are *i.i.d.* random variables.

Assumption 2. The second-order derivatives of $f(\mathbf{x})$ and $g(\mathbf{x})$ are continuous on $\mathbf{x} \in (0,1)^p$.

Assumption 3. There are some constants $\delta > 0$ and $C_1 \in [1, \infty)$ so that $E|Y|^{2+\delta} < \infty$ and

$$\sup_{\mathbf{x} \in (0,1)^p} E\left(|Y|^{2+\delta} \middle| \mathbf{X} = \mathbf{x} \right) f(\mathbf{x}) \leq C_1. \tag{1}$$

Assumption 4W. $b_j (= b_j(n) > 0)$ and $\eta_j (= \eta_j(n) > 0)$ for $j = 1, \ldots, p$ satisfy $b_j, \eta_j \to 0$, $b_j/\eta_j \to 0$ and $\ln n / \left(n \sqrt{\prod_{j=1}^p b_j \eta_j} \right) \to 0$ as $n \to \infty$.

Assumption 2 implies that there is some constant $C_0 \in [1, \infty)$ so that

$$\sup_{\mathbf{x} \in (0,1)^p} f(\mathbf{x}) \leq C_0. \tag{2}$$

The uniform boundedness condition (1) in Assumption 3 implies that $E\left(|Y|^{2+\delta} \middle| \mathbf{X} = \mathbf{x} \right)$ is allowed to diverge at boundaries but no faster than $\{f(\mathbf{x})\}^{-1}$. A similar condition can be found, for instance, in Hansen (2008, Assumption 2) and Kristensen (2009, Assumption A3). The conditions on η_j in Assumption 4W are intended for the case of an expanding set. In particular, the condition $b_j/\eta_j \to 0$ means that the boundary parameter η_j must shrink to zero at a slower rate than b_j. As can be seen in the Appendix, this is crucial for Stirling's approximation to the gamma function. To a large degree, both η_j and r_n (to be introduced in Assumption 5 shortly) are theoretical devices used in the novel proof of the convergence results. A specific rule for choosing them would have limited practical use but any intermediate order sequence with regards to the bandwidth would be appropriate.

In the context of copula density estimation, this condition is known to be violated for many copulas. For such copulas, it is not uncommon to exclude the boundaries of the hypercube from the analysis (e.g., Sancetta and Satchell, 2004) or use trimming/weighting schemes at the edges (e.g., Hill and Prokhorov, 2016). Controlling the upper bounds of the density $f(\mathbf{x})$ and its derivatives is a key for deriving uniform variance and bias convergence rates, respectively. However, unboundedness of the density at boundaries leads to unboundedness of its derivatives in the same region, which induces extra complexity in these exercises. We leave these extensions for future work.

Below we document weak uniform consistency of $\hat{g}_B(\mathbf{x})$ for $g(\mathbf{x})$ on the expanding p-hyperrectangle $\mathbb{S}_{\mathbf{X}} \to [0,1]^p$ as $n \to \infty$.

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

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

2.2 Strong Uniform Convergence of the Sample Average Estimator

Next, we demonstrate strong uniform consistency with rates of $\hat{g}_B(\mathbf{x})$. Before doing so, the assumption on smoothing parameters must be suitably strengthened.

Assumption 4S. $b_j\,(=b_j(n)>0)$ and $\eta_j\,(=\eta_j(n)>0)$ for $j=1,\ldots,p$ satisfy $b_j,\eta_j \to 0$, $b_j/\eta_j \to 0$ and

$$\frac{\ln n}{n\sqrt{\prod_{j=1}^{p} b_j\eta_j}}\left(\sum_{j=1}^{p}\frac{1}{b_j^2}\right)^{1-\kappa} = O(1) \tag{3}$$

for some constant $\kappa \in [0,1)$, as $n \to \infty$.

The condition (3) is stronger than $\ln n/\left(n\sqrt{\prod_{j=1}^{p} b_j\eta_j}\right) \to 0$ in Assumption 4W

in that the former implies the latter. Under this condition, the statement in Theorem 1 can be strengthened to almost sure convergence.

Theorem 2. *If Assumptions 1-3 and 4S hold, then, as* $n \to \infty$*,*

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|\hat{g}_B(\mathbf{x}) - g(\mathbf{x})\right| = O\left(\sum_{j=1}^{p} b_j + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{p} b_j \eta_j}}}\right), \ a.s.$$

3 Applications

3.1 Density Estimation

This section considers a variety of applications of Theorems 1 and 2 to nonparametric estimators using the product beta kernel. We start from presenting two theorems on weak and strong uniform convergence for the joint density estimator

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

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

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

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.*

Theorems 3 and 4 lead to the optimal uniform convergence rates of $\hat{f}_B(\mathbf{x})$ when $S_{\mathbf{X}}$ is fixed and a single smoothing parameter b is employed for each dimension. Let $\eta_1, \ldots, \eta_p$ be fixed and $b_1, \ldots, b_p \propto b$, where b satisfies, as $n \to \infty$: (i) $b + \ln n / \left(nb^{p/2}\right) \to 0$ for weak uniform convergence; or (ii) $b \to 0$ and $\ln n / \left\{nb^{p/2+2(1-\kappa)}\right\} = O(1)$ with $\kappa \in [0, 1)$ for strong uniform convergence. Observe that $b = O\left\{(\ln n/n)^{2/(4+p)}\right\}$, which yields the optimal convergence rates below, satisfies each rate requirement.

For such b, the optimal weak and strong uniform convergence rates of $\hat{f}_B(\mathbf{x})$ are both $(\ln n/n)^{2/(4+p)}$. This rate coincides with Stone's (1983) optimal global rate for nonparametric density estimation.

3.2 Regression Estimation Using Continuous Data

We proceed to investigating regression estimation. We consider two most popular kernel regression estimators, namely, the Nadaraya-Watson and local linear regression estimators. The Nadaraya-Watson regression estimator smoothed by the product beta kernel is defined as

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

On the other hand, the local linear regression estimator of $m(\mathbf{x})$ and the estimator of its first-order derivative $\nabla\{m(\mathbf{x})\}$ are given by the minimizer of the local least squares problem

$$\sum_{i=1}^n \left\{Y_i - \alpha - \beta^\top (\mathbf{X}_i - \mathbf{x})\right\}^2 \mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i).$$

Let $\left(\tilde{\alpha}(\mathbf{x}), \tilde{\beta}(\mathbf{x})\right)$ be the minimizer for a given $\mathbf{x}$. Then, the local linear estimator $\tilde{m}_B(\mathbf{x}) = \tilde{\alpha}(\mathbf{x})$ smoothed by the product beta kernel has the closed form of

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

where

$$\begin{aligned}
\mathbf{S}_1(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n (\mathbf{X}_i - \mathbf{x}) \mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i), \\
\mathbf{S}_2(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n (\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^\top \mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i), \text{ and} \\
\mathbf{T}_1(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n Y_i (\mathbf{X}_i - \mathbf{x}) \mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i).
\end{aligned}$$

It may be the case that $f(\mathbf{x}) \to 0$ as $x_j \to 0, 1$, for some $1 \leq j \leq p$. To deal with this case, we follow Hansen (2008) and impose an additional condition. Subsequently, we deliver two theorems on uniform convergence of $\hat{m}_B(\mathbf{x})$ and $\tilde{m}_B(\mathbf{x})$.

Assumption 5. Let $r_n := \inf_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} f(\mathbf{x}) > 0$. As $n \to \infty$, $r_n \to 0$ and the following statements hold:

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

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

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

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} |\hat{m}_B(\mathbf{x}) - m(\mathbf{x})| = O_p\left\{r_n^{-1}\left(\sum_{j=1}^{p} b_j + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{p} b_j\eta_j}}}\right)\right\}, \text{ and} \tag{4}$$

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} |\tilde{m}_B(\mathbf{x}) - m(\mathbf{x})| = O_p\left\{r_n^{-2}\left(\sum_{j=1}^{p} \frac{b_j}{\eta_j} + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{p} b_j\eta_j}}}\right)\right\}. \tag{5}$$

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.*

It is of importance and interest to compare Theorems 5 and 6 with Theorems 8-11 of Hansen (2008). Taking into account that the marginal density $f(\mathbf{x})$ tends to shrink to zero at the rate r_n in tail regions, Hansen (2008, Theorems 8 and 9) demonstrates that when the regression surface $m(\mathbf{x})$ is estimated over the entire Euclidean space by the Nadaraya-Watson estimator using multivariate symmetric kernels, its weak and strong uniform convergence rates slow down from those of the corresponding sample average estimator by a factor of the additional penalty term r_n^{-1}. The statement (4) indicates that the result continues to hold after replacing multivariate symmetric kernels with the product beta kernel defined on a unit hypercube.

Theorems 10 and 11 of Hansen (2008) document that the penalty term is strengthened to r_n^{-2} for the local linear estimator smoothed by multivariate symmetric kernels. The statement (5) argues that the expanding compact set $\mathbb{S}_\mathbf{X}$ influences the uniform

convergence rate of the beta local linear estimator from two different angles. More specifically, in addition to the more stringent penalty term r_n^{-2}, the bias convergence also decelerates from $O\left(\sum_{j=1}^{p} b_j\right)$ to $O\left\{\sum_{j=1}^{p} (b_j/\eta_j)\right\}$ due to the edge effect of $\mathbb{S}_{\mathbf{X}}$.

We can also obtain the optimal uniform convergence rates of $\hat{m}_B(\mathbf{x})$ and $\tilde{m}_B(\mathbf{x})$ for a fixed $\mathbb{S}_{\mathbf{X}}$ and a single smoothing parameter b; see Section 3.1 for detailed rate requirements for b. In this scenario, $f(\mathbf{x})$ is bounded away from zero uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$. Then, the optimal weak and strong uniform convergence rates of $\hat{m}_B(\mathbf{x})$ and $\tilde{m}_B(\mathbf{x})$ are both $(\ln n/n)^{2/(4+p)}$. The rates agree with Stone's (1982) optimal global rate for nonparametric regression estimation.

3.3 Regression Estimation Using Mixed Categorical and Continuous Data

3.3.1 A Product Kernel for Mixed Data

In this section our analysis is further extended to nonparametric regression estimation using both categorical (or discrete) and continuous data. Racine and Li (2004) originally propose to use a product kernel constructed from a univariate symmetric kernel for the continuous part of the regression estimator. Li and Ouyang (2005) demonstrate strong uniform consistency of that estimator on a compact interval. Our aim is to establish weak and strong convergence of the regression estimator in which the product beta kernel is employed for the continuous part.

A variety of discrete kernels can be applied for the categorical part; see Harfouche et al. (2018) for a non-exhaustive list of such kernels. Among all discrete kernels, our focuses are on those given by Aitchison and Aitken (1976), and in what follows a product of their discrete kernels is exclusively considered. Suppose that a discrete random variable Z takes $c\,(\geq 2)$ different values, i.e., $Z \in \{0, 1, \ldots, c-1\}$. The variable can be further classified into either unordered or ordered, because the kernels employed for the two types of categorical variables differ slightly. The univariate

discrete kernel for an *unordered* variable is

$$l(v; z, \lambda) := \begin{cases} 1-\lambda & \text{if } v = z \\ \lambda/(c-1) & \text{if } v \neq z \end{cases},$$

where v, z and $\lambda \in (0,1)$ is the data point, the design point and the bandwidth, respectively. Given the same notations, the univariate discrete kernel for an *ordered* variable is in the form of

$$\ell(v; z, \lambda) := \binom{c}{|v-z|} (1-\lambda)^{c-|v-z|} \lambda^{|v-z|}.$$

The product discrete kernel can be built on these univariate discrete kernels. If $q_1 (\leq q)$ out of q discrete variables are unordered, then the product discrete kernel becomes

$$\mathbb{L}(\mathbf{v}; \mathbf{z}, \lambda) = \left\{ \prod_{k=1}^{q_1} l(v_k; z_k, \lambda_k) \right\} \left\{ \prod_{k=q_1+1}^{q} \ell(v_k; z_k, \lambda_k) \right\},$$

where $v := (v_1, \ldots, v_q)^\top$, $z := (z_1, \ldots, z_q)^\top$ and $\lambda := (\lambda_1, \ldots, \lambda_q)^\top$. Combining this with the product beta kernel $\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u})$ finally leads to the product kernel for the mixed categorical and continuous data

$$\mathbb{W}(\mathbf{u}, \mathbf{v}; \mathbf{x}, \mathbf{z}, \mathbf{b}, \lambda) := \mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u}) \mathbb{L}(\mathbf{v}; \mathbf{z}, \lambda).$$

Given this kernel and n *i.i.d.* observations $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i)\}_{i=1}^n \in \mathbb{R} \times [0,1]^p \times \mathbb{S}_{\mathbf{Z}}$, where $\mathbb{S}_{\mathbf{Z}} := \prod_{k=1}^q \{0, 1, \ldots, c_k - 1\}$, we consider a Nadayara-Watson-type regression estimator of the conditional mean $m(\mathbf{x}, \mathbf{z}) := E(Y | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})$. It is in the form of

$$\hat{m}_W(\mathbf{x}, \mathbf{z}) := \frac{\sum_{i=1}^n Y_i \mathbb{W}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x}, \mathbf{z}, \mathbf{b}, \lambda)}{\sum_{i=1}^n \mathbb{W}(\mathbf{X}_i, \mathbf{Z}_i; \mathbf{x}, \mathbf{z}, \mathbf{b}, \lambda)}.$$

3.3.2 Weak and Strong Uniform Convergence of the Estimator

Below we demonstrate weak and strong uniform consistency of $\hat{m}_W(\mathbf{x}, \mathbf{z})$. To explore the uniform convergence results, we modify Assumptions 1-3, 4W, 4S, and 5 as follows.

Assumption 1'. $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i)\}_{i=1}^n \in \mathbb{R} \times [0,1]^p \times \mathbb{S}_{\mathbf{Z}}$ are *i.i.d.* random variables.

Assumption 2'. Let $f(\mathbf{x},\mathbf{z})$ be the joint pdf of $(\mathbf{X},\mathbf{Z})$. Then, the second-order derivatives of $f(\mathbf{x},\mathbf{z})$ and $g(\mathbf{x},\mathbf{z}) := m(\mathbf{x},\mathbf{z}) f(\mathbf{x},\mathbf{z})$ with respect to $\mathbf{x}$ are continuous on $\mathbf{x} \in (0,1)^p$.

Assumption 3'. There are some constants $\delta > 0$ and $C_1 \in [1,\infty)$ so that $E|Y|^{2+\delta} < \infty$ and

$$\sup_{(\mathbf{x},\mathbf{z})\in(0,1)^p\times\mathbb{S}_{\mathbf{Z}}} E\left(|Y|^{2+\delta}\middle|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}\right) f(\mathbf{x},\mathbf{z}) \leq C_1.$$

Assumption 4W'. b_j and η_j for $j = 1,\ldots,p$ and $\lambda_k\,(=\lambda_k(n)\in(0,1))$ for $k = 1,\ldots,q$ satisfy $b_j,\eta_j,\lambda_k \to 0$, $b_j/\eta_j \to 0$ and $\ln n/\left(n\sqrt{\prod_{j=1}^p b_j\eta_j}\right) \to 0$ as $n\to\infty$.

Assumption 4S'. b_j and η_j for $j = 1,\ldots,p$ and $\lambda_k\,(=\lambda_k(n)\in(0,1))$ for $k = 1,\ldots,q$ satisfy $b_j,\eta_j,\lambda_k \to 0$, $b_j/\eta_j \to 0$ and

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

for some constant $\kappa\in[0,1)$, as $n\to\infty$.

Assumption 5'. As $n\to\infty$, $r_n \to 0$ and

$$r_n^{-1}\left(\sum_{j=1}^p b_j + \sum_{k=1}^q \lambda_k + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^p b_j\eta_j}}}\right) \to 0.$$

Let $\mathbb{S} := \mathbb{S}_{\mathbf{X}}\times\mathbb{S}_{\mathbf{Z}}$. The following theorems document weak and strong uniform convergence of $\hat{m}_W(\mathbf{x},\mathbf{z})$.

Theorem 7. *If Assumptions 1'-3', 4W', and 5' hold, then, as $n\to\infty$,*

$$\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{S}} |\hat{m}_W(\mathbf{x},\mathbf{z}) - m(\mathbf{x},\mathbf{z})| = O_p\left\{r_n^{-1}\left(\sum_{j=1}^p b_j + \sum_{k=1}^q \lambda_k + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^p b_j\eta_j}}}\right)\right\}.$$

Theorem 8. *If Assumptions 1'-3', 4S', and 5' hold, then, as $n \to \infty$, the statement in Theorem 7 can be strengthened to almost sure convergence.*

It is possible to obtain the optimal uniform convergence rates of $\hat{m}_W(\mathbf{x}, \mathbf{z})$ when $\eta_1, \ldots, \eta_p$ are fixed (so is $\mathbb{S}_{\mathbf{X}}$), a single smoothing parameter b is employed for the continuous part, and all bandwidths for the discrete part are set no greater than b in orders of magnitude. Detailed rate requirements for b are the same as in Section 3.1. Then, the optimal weak and strong uniform convergence rates of $\hat{m}_W(\mathbf{x}, \mathbf{z})$ are both $(\ln n/n)^{2/(4+p)}$. Again, the rates correspond with what Stone (1982) derives as the optimal global rate for nonparametric regression estimation.

4 Data Analysis

So far uniform convergence results of beta kernel estimators have been explored from a theoretical point of view. In this section, we turn to practical aspects of the estimators and conduct Monte Carlo simulations and a real data analysis. For illustrative purposes, we investigate univariate density estimation and nonparametric regression estimation with a univariate regressor.

4.1 Monte Carlo Simulations

4.1.1 Case #1: Density Estimation

Our simulation study starts from kernel density estimation. Two density functions with support on $[0, 1]$ are considered. One is the logit normal distribution by Johnson (1949) with density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x(1-x)} \exp\left[-\frac{\left\{\ln\left(\frac{x}{1-x}\right) - \mu\right\}^2}{2\sigma^2}\right], \ (\mu, \sigma) = (-0.25, 1.75).$$

The other is the truncated normal distribution with density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \Big/ \left\{\Phi\left(\frac{1-\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)\right\}, \ (\mu, \sigma) = (0.2, 0.4),$$

where $\Phi(\cdot)$ is the standard normal distribution function. Shapes of these densities are given in Figure 1. For each distribution, 1000 data sets of sample size $n \in \{200, 400\}$ are simulated.

The density estimators compared are: (i) the density estimator using the Epanechnikov kernel [KDE-E]; (ii) the local linear density estimator by Lejeune and Sarda (1992) and Jones (1993) using the Epanechnikov kernel [LLDE-E]; and (iii) the density estimator using the beta kernel [KDE-B]. The definition of LLDE-E for the unit interval is given in Chen (1999, p.138). LLDE-E is boundary-adaptive in the sense that pointwise $O(h^2)$ bias and $O\left\{(nh)^{-1}\right\}$ variance convergences are preserved near boundaries as well as in the interior, where h is the bandwidth. In contrast, the beta kernel slows down its pointwise variance convergence from $O\left\{\left(nb^{1/2}\right)^{-1}\right\}$ to $O\left\{(nb)^{-1}\right\}$ near boundaries (although its bias convergence is $O(b)$ uniformly on $[0, 1]$). Moreover, while LLDE-E may generate negative estimates, KDE-B necessarily generates nonnegative estimates because the beta kernel is nonnegative everywhere.

As performance measures of an estimator $\bar{f}(\cdot)$, we adopt the root integrated squared error (RISE) and the integrated absolute deviation (IAD), where

$$RISE\left(\bar{f}\right) = \sqrt{\int_0^1 \left\{\bar{f}(x) - f(x)\right\}^2 dx},$$
$$IAD\left(\bar{f}\right) = \int_0^1 \left|\bar{f}(x) - f(x)\right| dx,$$

and each integral is approximated by the trapezoidal rule on an equally-spaced grid on $[0, 1]$.

Choosing the tuning parameter $\alpha \in \{h, b\}$ is an important practical issue. Two alternative choice rules are compared. One is the 'best case' analysis. For each simulated sample the RISE is computed over a grid of α, and then a minimizer of the RISE is obtained. The tuning parameter chosen in this way is called "Oracle" hereinafter. The other is the data-driven, leave-one-out cross-validation (CV) method.

The CV criterion function is

$$CV_D(\alpha) = \int_0^1 \left\{\bar{f}_\alpha(x)\right\}^2 dx - \frac{2}{n}\sum_{i=1}^{n} \bar{f}_{-i,\alpha}(X_i),$$

where $\bar{f}_\alpha(\cdot)$ signifies the dependence of the estimator on α, and $\bar{f}_{-i,\alpha}(\cdot)$ is the density estimate using the sample with the ith observation eliminated. The minimizer of $CV_D(\alpha)$ is defined as the CV tuning parameter. Bouezmarni and Rombouts (2010, Theorem 4) demonstrate asymptotic optimality of the CV in terms of the mean integrated squared error (MISE).[1]

FIGURE 1 AND TABLE 1 ABOUT HERE

Table 1 reports averages and standard deviations of the two performance measures and tuning parameter values over 1000 Monte Carlo samples. The results suggest that KDE-B performs best in both Oracle and CV cases. Despite its theoretical superiority, LLDE-E looks inferior to KDE-E, which is subject to boundary effects. Superior performance of KDE-B over LLDE-E agrees with simulation results in Chen (1999). It also appears that performances of all three estimators are not affected much by switching the smoothing parameter selection method from Oracle to CV.

4.1.2 Case #2: Regression Estimation

Next, finite-sample properties of beta regression estimators are examined. The data are generated from the regression model $Y = m(X) + \epsilon$, where the regressor X is generated as absolute values of $N(0, 0.3^2)$ random variables truncated on $[-1, 1]$, and the error ϵ is drawn from $N(0, 0.05^2)$, independently of X. Observe that the design

[1]The MISE of KDE-B is given in equation (4.2) of Chen (1999), and an extension to the case of p-dimensional joint density estimation is straightforward. Suppose that we are interested in finding a single smoothing parameter b that minimizes the MISE for the p-dimensional case. As the MISE is $O\left\{b^2 + \left(nb^{p/2}\right)^{-1}\right\}$, the MISE-optimal smoothing parameter is $b^* = O\left(n^{-2/(4+p)}\right)$. Observe that b^* differs from $b = O\left\{\left(\ln n/n\right)^{2/(4+p)}\right\}$, which yields Stone's (1983) optimal global rate for nonparametric density estimation $(\ln n/n)^{2/(4+p)}$; in other words, the former does not attain the optimal global rate.

density is relatively sparse toward the right boundary. In addition, the true regression curve $m(x) = 2/3 - (x - 2/3)^2$ is concave quadratic. These aspects reflect the real data analysis below. As before, 1000 data sets of sample size $n \in \{200, 400\}$ are simulated.

The regression estimators compared are: (i) the Nadaraya-Watson estimator using the Epanechnikov kernel [NW-E]; (ii) the local linear estimator using the Epanechnikov kernel [LL-E]; (iii) the Nadaraya-Watson estimator using the beta kernel [NW-B]; and (iv) the local linear estimator using the beta kernel [LL-B]. The tuning parameter α for each estimator is chosen either in the Oracle manner or by the leave-one-out CV. The CV criterion function is

$$CV_R(\alpha) = \sum_{i=1}^{n} \{Y_i - \bar{m}_{-i,\alpha}(X_i)\}^2,$$

where $\bar{m}_{-i,\alpha}(\cdot)$ is the regression estimate using tuning parameter α and a sample with the ith observation eliminated. Again the minimizer of $CV_R(\alpha)$ is defined as the CV tuning parameter.[2] Finally, RISE and IAD are considered as performance measures.

TABLE 2 ABOUT HERE

Results are provided in Table 2. In theory, LL-E should work under this design. However, the results indicate that it is outperformed even by NW-E in terms of RISE for both Oracle and CV cases. LL-B performs best and NW-B follows, except the CV case with $n = 200$. It is also worth remarking that two beta estimators are more stable than two Epanechnikov estimators after the smoothing parameter selection method is switched from Oracle to CV.

[2]The MISE of LL-B, for instance, is given in Section 3 of Chen (2002), and an extension to the case of multiple p regressors is again straightforward. As before, the MISE-optimal smoothing parameter for the p-dimensional case $b^* = O\left(n^{-2/(4+p)}\right)$ does not attain Stone's (1982) optimal global rate for nonparametric regression estimation $(\ln n/n)^{2/(4+p)}$.

4.2 A Real Data Example

We also apply beta density and regression estimations to a real data. The data set is extracted from the 1972 wave of the Panel Study of Income Dynamics (PSID) for our companion paper (Hirukawa, Murtazashvili and Prokhorov, 2021). Inspired by Henderson and Souto (2018), we estimate a nonparametric regression of $\ln(earnings)$ on *experience* using the white-male sub-sample, where *earnings* and *experience* are family head's total annual labor income (in US dollars) and work experience since 18 years old (in years), respectively. The sample size is 1977.

Applying beta kernel smoothing requires a transformation of the regressor to the unit scale. Because this variable ranges from 0 to 68, we first conduct beta kernel regression smoothing for the pair of $(\ln(earnings), experience/68)$ and then back-transform the estimation results to the original scale. Regression curves from the beta Nadaraya-Watson and local linear regression estimators are considered. Furthermore, $\ln(earnings)$ is thought to be a concave function of *experience* empirically, and the regression is popularly modelled parametrically as a quadratic function of *experience*. From this viewpoint, the least-squares predicted values from the quadratic regression model are also computed.

FIGURE 2 ABOUT HERE

Figure 2 presents regression curve estimates. Smoothing parameters of two beta estimators are chosen via the CV, and the values in the unit scale for the Nadaraya-Watson and local linear estimators are 0.017 and 0.066, respectively. For reference, the beta density estimate of *experience* is also provided, where the smoothing parameter is again chosen by the CV and its value in the unit scale is 0.006. As the real underlying curve is unknown, it is hard to judge among three regression estimators. However, it is safe to say that the curve from the local linear estimator is considerably close to the plot of predicted values from the quadratic regression model.

The Nadaraya-Watson estimate is also similar to the predicted value plot up until the middle part, but it becomes wavy on the right-tail part, reflecting sparseness of observations in this region.

5 Conclusion

In this paper, we have derived weak and strong uniform convergence rates of the sample average estimator smoothed by the product beta kernel on a compact set that is either fixed within or expanding to a p-dimensional unit hypercube. The results are then applied to nonparametric density and regression estimators using the product beta kernel. It is demonstrated that the optimal weak and strong convergence rates of the density and regression estimators are both $(\ln n/n)^{2/(p+4)}$, which coincides with the best possible uniform rate for nonnegative kernel estimators. For practical considerations, we apply beta kernel smoothing to simulations and an empirical data analysis.

A Appendix

A.1 Useful Lemmata

Before proceeding, we present a few lemmata, all of which are key building blocks for the technical proofs below. Throughout θ_{x_j} denotes a beta random variable so that $\theta_{x_j} \overset{d}{=} Beta\left\{x_j/b_j + 1, (1-x_j)/b_j + 1\right\}$. Also notice that Lemma A4 is Bernstein's inequality.

Lemma A1. *Let θ_{x_j} and θ_{x_k} be independent for $j \neq k$. Then, as $n \to \infty$,*

$$\sup_{x_j \in (0,1)} E\left(\theta_{x_j} - x_j\right) = O\left(b_j\right), \text{ and}$$

$$\sup_{x_j, x_k \in (0,1)} E\left\{\left(\theta_{x_j} - x_j\right)\left(\theta_{x_k} - x_k\right)\right\} = \begin{cases} O\left(b_j\right) & \text{for } j = k \\ O\left(b_j b_k\right) & \text{for } j \neq k \end{cases}.$$

Lemma A2. *Suppose that* $b\,(=b\,(n)>0)$ *and* $\eta\,(=\eta\,(n)>0)$ *satisfy* $b,\eta\to 0$ *and* $b/\eta\to 0$ *as* $n\to\infty$. *Then, as* $n\to\infty$,

$$\sup_{(x,u)\in[\eta,1-\eta]\times[0,1]} K_{B(x,b)}\,(u)\leq\left(\frac{9}{4\sqrt{\pi}}\right)b^{-1/2}\eta^{-1/2}.$$

Lemma A3. *Under the same condition as in Lemma A2, as* $n\to\infty$,

$$\sup_{(x,u)\in[\eta,1-\eta]\times[0,1]}\left|\frac{\partial K_{B(x,b)}\,(u)}{\partial x}\right|\leq\left\{\left(\frac{9}{4\sqrt{\pi}}\right)\left(\gamma+\frac{\pi^2}{6}\right)+1\right\}b^{-(2+1/2)}\eta^{-1/2},$$

where $\gamma=0.5772\ldots$ *is Euler's constant.*

Lemma A4. (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* $v\geq Var\,(\sum_{i=1}^{n}X_i)$.

A.1.1 Proof of Lemma A1

By the property of a beta random variable, $0<x_j<1$, $b_j>0$, and $b_j\leq 1$ for a sufficiently large n,

$$\left|E\left(\theta_{x_j}-x_j\right)\right|=\left|\frac{b_j\,(1-2x_j)}{1+2b_j}\right|\leq\frac{b_j\,(1+2)}{1+0}=3b_j,\text{ and}$$
$$\left|E\left(\theta_{x_j}-x_j\right)^2\right|=\left|\frac{b_j\left\{x_j\,(1-x_j)+2\left(3x_j^2-3x_j+1\right)b_j\right\}}{(1+2b_j)\,(1+3b)}\right|\leq\frac{b_j\,(1/2+2\cdot 1\cdot 1)}{(1+0)\,(1+0)}=\frac{5}{2}b_j.$$

Then, the results immediately follow. ■

A.1.2 Proof of Lemma A2

Recognize that $u^{x/b}\,(1-u)^{(1-x)/b}$ is maximized at $u=x$, i.e., x is the mode of the pdf of $Beta\left\{x/b+1,(1-x)\,/b+1\right\}$. Hence,

$$u^{x/b}\,(1-u)^{(1-x)/b}\leq x^{x/b}\,(1-x)^{(1-x)/b}\,. \tag{A1}$$

It also follows from $b/\eta \to 0$ that for a given $x \in [\eta, 1-\eta]$, $x/b, (1-x)/b \to \infty$ holds as $n \to \infty$. Then, we may apply Stirling's approximation to three gamma functions in

$$\frac{1}{B\{x/b+1, (1-x)/b+1\}} = \frac{(1/b+1)\,\Gamma(1/b+1)}{\Gamma(x/b+1)\,\Gamma\{(1-x)/b+1\}}.$$

Specifically, because both $O(b/x)$ and $O\{b/(1-x)\}$ are $o(1)$ as $n \to \infty$, uniformly on $x \in [\eta, 1-\eta]$, we have

$$\begin{aligned}
&\frac{(1/b+1)\,\Gamma(1/b+1)}{\Gamma(x/b+1)\,\Gamma\{(1-x)/b+1\}} \\
&= \left(\frac{1+b}{b}\right)\sqrt{2\pi}\left(\frac{1}{b}\right)^{\frac{1}{b}+\frac{1}{2}}\exp\left(-\frac{1}{b}\right)\{1+O(b)\} \\
&\times \left[\sqrt{2\pi}\left(\frac{x}{b}\right)^{\frac{x}{b}+\frac{1}{2}}\exp\left(-\frac{x}{b}\right)\left\{1+O\left(\frac{b}{x}\right)\right\}\right]^{-1} \\
&\times \left[\sqrt{2\pi}\left(\frac{1-x}{b}\right)^{\frac{1-x}{b}+\frac{1}{2}}\exp\left(-\frac{1-x}{b}\right)\left\{1+O\left(\frac{b}{1-x}\right)\right\}\right]^{-1} \\
&= \frac{(1+b)\,b^{-1/2}\{1+o(1)\}}{x^{x/b}(1-x)^{(1-x)/b}\sqrt{2\pi}\sqrt{x(1-x)}}. \qquad \text{(A2)}
\end{aligned}$$

Therefore, by (A1) and (A2), we have, for a given $x \in [\eta, 1-\eta]$,

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

Furthermore, for a sufficiently large n, we can make each of b, η and the $o(1)$ term no greater than $1/2$. Then, the right-hand side is bounded by

$$\frac{b^{-1/2}(1+1/2)^2}{\sqrt{2\pi}\sqrt{\eta(1-1/2)}} \leq \left(\frac{9}{4\sqrt{\pi}}\right) b^{-1/2}\eta^{-1/2},$$

which completes the proof. ■

A.1.3 Proof of Lemma A3

We consider the cases of $u = 0, 1$ and $0 < u < 1$ separately. For $u = 0, 1$, because $x \in [\eta, 1-\eta]$, we have $x/b, (1-x)/b > 0$ so that $K_{B(x,b)}(0) = K_{B(x,b)}(1) = 0$. Then, $\partial K_{B(x,b)}(0)/\partial x = \partial K_{B(x,b)}(1)/\partial x = 0$, and the result trivially holds.

For $0 < u < 1$, recognize that

$$\frac{\partial K_{B(x,b)}(u)}{\partial x} = \left[\frac{\partial \ln\left\{K_{B(x,b)}(u)\right\}}{\partial x}\right] K_{B(x,b)}(u).$$

Because

$$\begin{aligned}\ln\left\{K_{B(x,b)}(u)\right\} = \ln\Gamma\left(\frac{1}{b}+2\right) + \left(\frac{x}{b}\right)\ln u + \left(\frac{1-x}{b}\right)\ln(1-u) \\ - \ln\Gamma\left(\frac{x}{b}+1\right) - \ln\Gamma\left(\frac{1-x}{b}+1\right),\end{aligned}$$

we have

$$\frac{\partial \ln\left\{K_{B(x,b)}(u)\right\}}{\partial x} = \frac{1}{b}\left[\ln u - \ln(1-u) - \frac{\Gamma'(x/b+1)}{\Gamma(x/b+1)} + \frac{\Gamma'\{(1-x)/b+1\}}{\Gamma\{(1-x)/b+1\}}\right]$$

so that

$$\begin{aligned}\left|\frac{\partial K_{B(x,b)}(u)}{\partial x}\right| &\leq \frac{1}{b}\left\{|\ln u| + |\ln(1-u)|\right\} K_{B(x,b)}(u) \\ &\quad + \frac{1}{b}\left\{\left|\frac{\Gamma'(x/b+1)}{\Gamma(x/b+1)}\right| + \left|\frac{\Gamma'\{(1-x)/b+1\}}{\Gamma\{(1-x)/b+1\}}\right|\right\} K_{B(x,b)}(u) \\ &= G_1 + G_2 \ \text{(say)}.\end{aligned}$$

We find the bound for G_2 first. Differentiating both sides of $\ln\Gamma(z+1) = \ln z + \ln\Gamma(z)$ for $z > 0$ yields $\Gamma'(z+1)/\Gamma(z+1) = 1/z + \Gamma'(z)/\Gamma(z)$. Because the digamma function $\Gamma'(z)/\Gamma(z)$ admits the series expansion

$$\frac{\Gamma'(z)}{\Gamma(z)} = -\gamma - \frac{1}{z} + z\sum_{m=1}^{\infty}\frac{1}{m(z+m)},$$

where $\gamma = 0.5772\ldots$ is Euler's constant, we have

$$\left|\frac{\Gamma'(z+1)}{\Gamma(z+1)}\right| = \left|-\gamma + z\sum_{m=1}^{\infty}\frac{1}{m(z+m)}\right| \leq \gamma + z\sum_{m=1}^{\infty}\frac{1}{m^2} = \gamma + \frac{\pi^2}{6}z.$$

Combining this with Lemma A2, we find that

$$\begin{aligned}G_2 &\leq b^{-1}\left(2\gamma + \frac{\pi^2}{6}b^{-1}\right)\left(\frac{9}{4\sqrt{\pi}}\right) b^{-1/2}\eta^{-1/2} \\ &= \left(\frac{9}{4\sqrt{\pi}}\right)\left(2\gamma b + \frac{\pi^2}{6}\right) b^{-(2+1/2)}\eta^{-1/2} \\ &\leq \left(\frac{9}{4\sqrt{\pi}}\right)\left(\gamma + \frac{\pi^2}{6}\right) b^{-(2+1/2)}\eta^{-1/2} \qquad \text{(A3)}\end{aligned}$$

by putting $b \leq 1/2$ for a sufficiently large n.

Next

$$G_1 = -\frac{1}{b}\frac{\{\ln u + \ln(1-u)\}\, u^{x/b}\,(1-u)^{(1-x)/b}}{B\{x/b+1,(1-x)/b+1\}}$$

is examined. We start from arguing that $\psi(u) := -\ln u\,(1-u)\,u^{x/b}\,(1-u)^{(1-x)/b}$ has the maximum on $(0,1)$ for a sufficiently small $b>0$. Let $\xi(u) := (u-x)\ln u\,(1-u) - b\,(1-2u)$ so that $\psi'(u) = b^{-1}u^{x/b-1}\,(1-u)^{(1-x)/b-1}\,\xi(u)$. Clearly, $\lim_{u\downarrow 0}\xi(u) \to \infty$ and $\lim_{u\uparrow 1}\xi(u) \to -\infty$ for $x \in (0,1)$ and $b \approx 0$. In addition, $\xi'(u) = \ln u\,(1-u) + (u-x)\,(1-2u)/\{u\,(1-u)\} + 2b$. A tedious but straightforward calculation yields $\ln u\,(1-u) \leq -\ln 4$ and $(u-x)\,(1-2u)/\{u\,(1-u)\} \leq 1 - 2\sqrt{x\,(1-x)}$. Therefore, $\xi'(u) \leq -\left\{\ln 4 + 2\sqrt{x\,(1-x)}\right\} + 1 + 2b < 0$ for $b \approx 0$, and thus $\xi(u)$ is shown to be monotone decreasing.

Next, let $u^* \in (0,1)$ maximize $\psi(u)$. Heuristically, $\xi(u) \approx (u-x)\ln u\,(1-u)$ for $b \approx 0$, and thus $u^* \approx x$ is the case. Therefore, $u^* \to x$ as $b \to 0$, but we even conjecture that $u^* = x + \alpha b$ for some constant $|\alpha| \in (0,\infty)$. Now this conjecture is shown to be true. Because u^* solves $\xi(u^*) = 0$, we have

$$|\alpha| = \left|\frac{u^*-x}{b}\right| = \left|\frac{1-2u^*}{\ln u^*(1-u^*)}\right| \leq \left|\frac{1-2x}{\ln x\,(1-x)}\right|\{1+o(1)\},$$

where the inequality is implied by $u^* = x + o(1)$. Observe that for $x \in (0,1)$, $|1-2x| \leq 1$ and $|\ln x\,(1-x)| \geq \ln 4$. It is also possible to make the $o(1)$ term no greater than 1 for a sufficiently large n, and thus the right-hand side is bounded by $1/\ln 2$. Therefore, $|\alpha| \in (0,\infty)$ is established.

For such u^*, $-\ln u^*(1-u^*) = -\ln x\,(1-x) + O(b)$, and

$$\begin{aligned}
&u^{*x/b}\,(1-u^*)^{(1-x)/b} \\
&= x^{x/b}\,(1-x)^{(1-x)/b}\left(1+\frac{\alpha b}{x}\right)^{x/b}\left(1-\frac{\alpha b}{1-x}\right)^{(1-x)/b} \\
&= x^{x/b}\,(1-x)^{(1-x)/b}\left\{\exp(\alpha)+O\left(\frac{b}{x}\right)\right\}\left\{\exp(-\alpha)+O\left(\frac{b}{1-x}\right)\right\} \\
&= x^{x/b}\,(1-x)^{(1-x)/b}\left\{1+O\left(\frac{b}{x}\right)+O\left(\frac{b}{1-x}\right)\right\}.
\end{aligned}$$

It follows from (A2) that

$$\begin{aligned} G_1 &\leq \frac{b^{-1}\psi\left(u^*\right)}{B\left\{x/b+1,(1-x)/b+1\right\}} \\ &\leq \frac{b^{-(1+1/2)}\left(1+b\right)\left\{1+o\left(1\right)\right\}}{\sqrt{2\pi}\sqrt{x\left(1-x\right)}} \\ &\times\left[-\left\{\ln x+\ln\left(1-x\right)\right\}+O\left(b\right)\right]\left\{1+O\left(\frac{b}{x}\right)+O\left(\frac{b}{1-x}\right)\right\} \\ &\leq O\left\{b^{-(1+1/2)}\eta^{-1/2}\left(-\ln\eta\right)\right\}, \end{aligned} \tag{A4}$$

uniformly on $x\in[\eta,1-\eta]$.

By (A3) and (A4), we finally have

$$\left|\frac{\partial K_{B(x,b)}\left(u\right)}{\partial x}\right|\leq\left\{\left(\frac{9}{4\sqrt{\pi}}\right)\left(\gamma+\frac{\pi^2}{6}\right)+O\left\{b\left(-\ln\eta\right)\right\}\right\}b^{-(2+1/2)}\eta^{-1/2}.$$

The lemma is established by making the $O\left\{b\left(-\ln\eta\right)\right\}$ term no greater than 1 for a sufficiently large n. ■

A.2 Proof of Theorem 1

For ease of exposition, we additionally introduce the following notations:

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

The meaning of each notation will be revealed shortly.

Consider that

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

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

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|E\left\{\hat{g}_B\left(\mathbf{x}\right)\right\} - g\left(\mathbf{x}\right)\right| = O\left(\sum_{j=1}^{p} b_j\right); \text{ and} \tag{A5}$$

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}} \left|\hat{g}_B\left(\mathbf{x}\right) - E\left\{\hat{g}_B\left(\mathbf{x}\right)\right\}\right| = O_p\left(a_n\right). \tag{A6}$$

For (A5); Observe that

$$\begin{aligned} E\left\{\hat{g}_B\left(\mathbf{x}\right)\right\} &= E\left\{E\left(Y_i|\,\mathbf{X}_i\right)\mathbb{K}_{B(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)\right\} \\ &= E\left\{m\left(\mathbf{X}_i\right)\mathbb{K}_{B(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)\right\} \\ &= \int_{[0,1]^p} m\left(\mathbf{u}\right) f\left(\mathbf{u}\right)\mathbb{K}_{B(\mathbf{x},\mathbf{b})}\left(\mathbf{u}\right)\mathbf{du} \\ &= E\left\{g\left(\theta_{\mathbf{x}}\right)\right\}, \end{aligned}$$

where $\theta_{\mathbf{x}} := \left(\theta_{x_1},\ldots,\theta_{x_p}\right)^{\top}$. By a second-order Taylor expansion around $\theta_{\mathbf{x}} = \mathbf{x}$,

$$\begin{aligned} E\left\{g\left(\theta_{\mathbf{x}}\right)\right\} = g\left(\mathbf{x}\right) &+ \sum_{j=1}^{p}\frac{\partial g\left(\mathbf{x}\right)}{\partial x_j}E\left(\theta_{x_j}-x_j\right) + \frac{1}{2}\sum_{j=1}^{p}\frac{\partial^2 g\left(\bar{\mathbf{x}}\right)}{\partial x_j^2}E\left(\theta_{x_j}-x_j\right)^2 \\ &+ \sum_{j=1}^{p}\sum_{k=1,k\neq j}^{p}\frac{\partial^2 g\left(\bar{\mathbf{x}}\right)}{\partial x_j\partial x_k}E\left\{\left(\theta_{x_j}-x_j\right)\left(\theta_{x_k}-x_k\right)\right\} \end{aligned}$$

for some $\bar{\mathbf{x}}$ joining $\theta_{\mathbf{x}}$ and $\mathbf{x}$. Then, Assumption 2 and Lemma A1 lead to (A5).

For (A6); As in the proof of Theorem 2 in Hansen (2008), 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 $O_p\left(a_n\right)$ uniformly on $\mathbf{x}\in\mathbb{S}_{\mathbf{X}}$;

2. Split each edge of the p-hyperrectangle $\mathbb{S}_{\mathbf{X}}$ into N_n equally-spaced grids to create N_n^p identical sub-hyperrectangles, and replace the supremum with a maximization over the finite N_n^p sub-hyperrectangles; and

3. Employ Lemma A4 (Bernstein's inequality) to bound the remainder term.

Step 1. Let $R_n(\mathbf{x}) := (1/n)\sum_{i=1}^n Y_i \mathbf{1}\{|Y_i| > \tau_n\}\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)$. Then, $\hat{g}_B(\mathbf{x}) - E\{\hat{g}_B(\mathbf{x})\} = \sum_{i=1}^n \varsigma_{in}(\mathbf{x})$ and $\sum_{i=1}^n \varsigma_{in}(\mathbf{x}) - \sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x}) = R_n(\mathbf{x}) - E\{R_n(\mathbf{x})\}$. Now, for $|Y_i| > \tau_n$, $(|Y_i|/\tau_n)^{1+\delta} > 1$ is the case. It follows that

$$\begin{aligned}
|E\{R_n(\mathbf{x})\}| &\le E\left\{|Y_i|\,\mathbf{1}\{|Y_i| > \tau_n\}\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\} \\
&\le E\left\{|Y_i|\left(\frac{|Y_i|}{\tau_n}\right)^{1+\delta}\mathbf{1}\{|Y_i| > \tau_n\}\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\} \\
&\le \tau_n^{-(1+\delta)} E\left\{|Y_i|^{2+\delta}\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\}, \qquad \text{(A7)}
\end{aligned}$$

where, by Assumption 3 and the fact that $\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\cdot)$ is the pdf of the product of p independent beta random variables $\theta_{x_1}, \ldots, \theta_{x_p}$,

$$\begin{aligned}
E\left\{|Y_i|^{2+\delta}\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\} &= E\left\{E\left(|Y_i|^{2+\delta}\middle|\mathbf{X}_i\right)\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i)\right\} \\
&= \int_{[0,1]^p} E\left(|Y|^{2+\delta}\middle|\mathbf{X} = \mathbf{u}\right) f(\mathbf{u})\,\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u})\,\mathbf{du} \le C_1. \qquad \text{(A8)}
\end{aligned}$$

Therefore, by the definition of τ_n, $|E\{R_n(\mathbf{x})\}| \le O(a_n)$ uniformly on $\mathbf{x} \in \mathbb{S}_{\mathbf{X}}$, and as a result of Markov's inequality,

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left|\sum_{i=1}^n \varsigma_{in}(\mathbf{x}) - \sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x})\right| = \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}|R_n(\mathbf{x}) - E\{R_n(\mathbf{x})\}| = O_p(a_n). \qquad \text{(A9)}$$

Step 2. Let $\mathbf{A}_h, h = 1, \ldots, N_n^p$ be the hth sub-hyperrectangle. Also let $\mathbf{x}_h$ be the most distant point in $\mathbf{A}_h$ from the origin, 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 rate of $\sup_{\mathbf{x}\in\mathbf{A}_h}|\sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x}) - \sum_{i=1}^n \hat{\varsigma}_{in}(\mathbf{x}_h)|$ is determined by $\left|\hat{Y}_i\right|\left|\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{X}_i) - \mathbb{K}_{B(\mathbf{x}_h,\mathbf{b})}(\mathbf{X}_i)\right|$. By the mean-value theorem,

$$\left|\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u}) - \mathbb{K}_{B(\mathbf{x}_h,\mathbf{b})}(\mathbf{u})\right| \le \sup_{(\mathbf{x},\mathbf{u})\in\mathbf{A}_h\times[0,1]^p}\left\|\nabla\left\{\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u})\right\}\right\| \sup_{\mathbf{x}\in\mathbf{A}_h}\|\mathbf{x} - \mathbf{x}_h\|$$

for some $\tilde{\mathbf{x}}$ joining $\mathbf{x}$ and $\mathbf{x}_h$. Furthermore, by Lemmata A2 and A3, for $k = 1, \ldots, p$,

$$\left|\frac{\partial\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u})}{\partial x_k}\right| \le \left\{\prod_{j=1,j\ne k}^p K_{B(x_j,b_j)}(u_j)\right\}\left|\frac{\partial K_{B(x_k,b_k)}(u_k)}{\partial x_k}\right| = O\left\{\left(\prod_{j=1}^p b_j\eta_j\right)^{-\frac{1}{2}}\frac{1}{b_k^2}\right\}$$

uniformly on $(\mathbf{x},\mathbf{u}) \in \mathbf{A}_h \times [0,1]^p$, and thus

$$\sup_{(\mathbf{x},\mathbf{u})\in\mathbf{A}_h\times[0,1]^p}\left\|\nabla\left\{\mathbb{K}_{B(\mathbf{x},\mathbf{b})}(\mathbf{u})\right\}\right\| = O\left\{\left(\prod_{j=1}^p b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^p\frac{1}{b_j^2}\right)\right\}.$$

It follows from $\sup_{\mathbf{x}\in\mathbf{A}_h}\|\mathbf{x}-\mathbf{x}_h\| = O\left(N_n^{-1}\right)$ that uniformly on $(\mathbf{x},\mathbf{u})\in\mathbf{A}_h\times[0,1]^p$,

$$\left|\hat{Y}_i\right|\left|\mathbb{K}_{B(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)-\mathbb{K}_{B(\mathbf{x}_h,\mathbf{b})}\left(\mathbf{X}_i\right)\right| \leq O\left\{\tau_n N_n^{-1}\left(\prod_{j=1}^p b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^p\frac{1}{b_j^2}\right)\right\} = O\left(a_n\right).$$

Therefore,

$$\max_{1\leq h\leq N_n^p}\sup_{\mathbf{x}\in\mathbf{A}_h}\left|\sum_{i=1}^n\hat{\varsigma}_{in}\left(\mathbf{x}\right)-\sum_{i=1}^n\hat{\varsigma}_{in}\left(\mathbf{x}_h\right)\right| = O\left(a_n\right). \tag{A10}$$

Step 3. Before employing Bernstein's inequality in Lemma A4, we must determine the values of M and v. First, by Lemma A2, for a sufficiently large n,

$$|\hat{\varsigma}_{in}\left(\mathbf{x}\right)| \leq 2\left(\frac{9}{4\sqrt{\pi}}\right)^p\frac{\tau_n}{n\sqrt{\prod_{j=1}^p b_j\eta_j}} = 2\left(\frac{9}{4\sqrt{\pi}}\right)^p\frac{a_n^{2-1/(1+\delta)}}{\ln n} =: M.$$

Second,

$$\begin{aligned}
Var\left\{\sum_{i=1}^n\hat{\varsigma}_{in}\left(\mathbf{x}\right)\right\} &= \sum_{i=1}^n Var\left\{\hat{\varsigma}_{in}\left(\mathbf{x}\right)\right\}\\
&\leq \frac{1}{n^2}\sum_{i=1}^n E\left\{\left|\hat{Y}_i\right|\mathbb{K}_{B(\mathbf{x},\mathbf{b})}\left(\mathbf{X}_i\right)\right\}^2\\
&= \frac{1}{n}\int_{[0,1]^p} E\left(\left|Y\right|^2\middle|\mathbf{X}=\mathbf{u}\right)f\left(\mathbf{u}\right)\mathbb{K}^2_{B(\mathbf{x},\mathbf{b})}\left(\mathbf{u}\right)\mathbf{du}.
\end{aligned}$$

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

$$\begin{aligned}
E\left(\left|Y\right|^2\middle|\mathbf{X}=\mathbf{u}\right)f\left(\mathbf{u}\right) &\leq \left\{E\left(\left|Y\right|^{2+\delta}\middle|\mathbf{X}=\mathbf{u}\right)f\left(\mathbf{u}\right)\right\}^{2/(2+\delta)}\left\{f\left(\mathbf{u}\right)\right\}^{\delta/(2+\delta)}\\
&\leq C_1^{2/(2+\delta)}C_0^{\delta/(2+\delta)} \leq C_0C_1.
\end{aligned}$$

Moreover,

$$\begin{aligned}
K^2_{B(x,b)}\left(u\right) = &\frac{B\left\{2x/b+1, 2\left(1-x\right)/b+1\right\}}{B^2\left\{x/b+1,\left(1-x\right)/b+1\right\}}\\
&\times\frac{u^{2x/b}\left(1-u\right)^{2(1-x)/b}}{B\left\{2x/b+1,2\left(1-x\right)/b+1\right\}}\mathbf{1}\left\{u\in[0,1]\right\}.
\end{aligned}$$

By Lemma of Chen (1999), the first term is bounded by $b^{-1/2}\left(1+b\right)^{3/2}/\left\{2\sqrt{\pi}\sqrt{x\left(1-x\right)}\right\}$ for a sufficiently large n. The second term is the pdf of $Beta\left\{2x/b+1,2\left(1-x\right)/b+1\right\}$. Therefore,

$$Var\left\{\sum_{i=1}^n\hat{\varsigma}_{in}\left(\mathbf{x}\right)\right\} \leq \frac{1}{n}C_0C_1\prod_{j=1}^p\frac{b_j^{-1/2}\left(1+b_j\right)^{3/2}}{2\sqrt{\pi}\sqrt{x_j\left(1-x_j\right)}} \leq \frac{1}{n}C_0C_1\prod_{j=1}^p\frac{b_j^{-1/2}\left(1+b_j\right)^{3/2}}{2\sqrt{\pi}\sqrt{\eta_j\left(1-\eta_j\right)}}.$$

For a sufficiently large n, $b_1, \ldots, b_p$ and $\eta_1, \ldots, \eta_p$ are no greater than $1/2$, and thus

$$Var\left\{\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}\right)\right\} \leq \frac{1}{n\sqrt{\prod_{j=1}^{p} b_j\eta_j}} C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p = \frac{a_n^2}{\ln n} C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p =: v.$$

Lemma A4 establishes that for such M and v and an arbitrarily chosen $K > 0$,

$$\begin{aligned}
&\Pr\left\{\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}\right)\right| > K\sqrt{C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p a_n}\right\} \\
&\leq 2\exp\left[-\frac{K^2\ln n}{2\left\{1+\frac{2}{3}\left(\frac{9}{4\sqrt{\pi}}\right)^p K a_n^{1-1/(1+\delta)}/\sqrt{C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p}\right\}}\right],
\end{aligned}$$

By $a_n = o\left(1\right)$, $\left(2/3\right)\left\{9/\left(4\sqrt{\pi}\right)\right\}^p K a_n^{1-1/(1+\delta)}/\sqrt{C_0C_1\left\{\left(3/4\right)\sqrt{3/\pi}\right\}^p} \leq 1$ for a sufficiently large n. Then,

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

In the end,

$$\begin{aligned}
&\Pr\left\{\max_{1\leq h\leq N_n^p}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}_h\right)\right| > K\sqrt{C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p a_n}\right\} \\
&\leq \sum_{h=1}^{N_n^p}\Pr\left\{\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}_h\right)\right| > K\sqrt{C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p a_n}\right\} \\
&\leq N_n^p \times \max_{1\leq h\leq N_n^p}\Pr\left\{\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}_h\right)\right| > K\sqrt{C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p a_n}\right\} \\
&= O\left(N_n^p n^{-K^2/4}\right). \qquad \text{(A11)}
\end{aligned}$$

Pick $K = 2\sqrt{5p}$. Then, by the definitions of N_n and a_n,

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

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

$$\max_{1 \leq h \leq N_n^p} \left| \sum_{i=1}^{n} \hat{\varsigma}_{in} (\mathbf{x}_h) \right| = O_p (a_n) . \tag{A12}$$

The statement (A6) follows from (A9), (A10) and (A12). This completes the proof. ■

A.3 Proof of Theorem 2

We keep using the notations in the proof of Theorem 1, whereas we redefine τ_n and N_n. Let

$$\tau_n := n^{\frac{1+\epsilon}{2+\delta}} \text{ and } N_n := n^{1+\epsilon} \left(\prod_{j=1}^{p} b_j \eta_j \right)^{-\frac{1}{2}} \left(\sum_{j=1}^{p} \frac{1}{b_j^2} \right)$$

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

The proof again follows three steps of that of Theorem 1. First, it follows from (A7) and (A8) that

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

Also by Markov's inequality and Assumption 3,

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

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

$$|R_n (\mathbf{x}) - E \{R_n (\mathbf{x})\}| = \left| \sum_{i=1}^{n} \varsigma_{in} (\mathbf{x}) - \sum_{i=1}^{n} \hat{\varsigma}_{in} (\mathbf{x}) \right| = O (a_n) , \text{ a.s.}, \tag{A13}$$

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

Second, observe that

$$\tau_n N_n^{-1} \left(\prod_{j=1}^{p} b_j \eta_j \right)^{-\frac{1}{2}} \left(\sum_{j=1}^{p} \frac{1}{b_j^2} \right) = n^{-(1+\epsilon)\left(\frac{1+\delta}{2+\delta}\right)} \leq O (a_n) .$$

Hence,

$$\max_{1\leq h\leq N_n^p}\sup_{\mathbf{x}\in\mathbf{A}_h}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}\right)-\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}_h\right)\right|$$
$$=O\left\{\tau_n N_n^{-1}\left(\prod_{j=1}^{p}b_j\eta_j\right)^{-\frac{1}{2}}\left(\sum_{j=1}^{p}\frac{1}{b_j^2}\right)\right\}=O\left(a_n\right). \tag{A14}$$

Third, (A11) holds for a sufficiently large n. In addition, (3) implies that

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

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

$$\sum_{n=1}^{\infty}\Pr\left\{\max_{1\leq h\leq N_n^p}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}_h\right)\right|>K\sqrt{C_0C_1\left(\frac{3}{4}\sqrt{\frac{3}{\pi}}\right)^p a_n}\right\}\leq\sum_{n=1}^{\infty}O\left(\frac{1}{n^{1+\epsilon}}\right)<\infty.$$

Therefore, by the Borel-Cantelli lemma,

$$\max_{1\leq h\leq N_n^p}\left|\sum_{i=1}^{n}\hat{\varsigma}_{in}\left(\mathbf{x}_h\right)\right|=O\left(a_n\right),\ \text{a.s.} \tag{A15}$$

The stated result is established by (A13), (A14) and (A15). ■

A.4 Proof of Theorem 3

This theorem can be established by putting $Y_i\equiv 1$ in Theorem 1. ■

A.5 Proof of Theorem 4

The argument is the same as for Theorem 3, except that Theorem 2 is employed so that the convergence holds almost surely. ■

A.6 Proof of Theorem 5

For (4); The proof closely follows that of Theorem 8 in Hansen (2008). Let

$$a_n^*:=\sum_{j=1}^{p}b_j+\sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{p}b_j\eta_j}}}.$$

Then, by Theorems 1 and 3 and Assumption 5,

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left|\frac{\hat{g}_B(\mathbf{x})}{f(\mathbf{x})}-m(\mathbf{x})\right|\leq\frac{\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}|\hat{g}_B(\mathbf{x})-g(\mathbf{x})|}{\inf_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}f(\mathbf{x})}=O_p\left(r_n^{-1}a_n^*\right),\text{ and}\tag{A16}$$

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left|\frac{\hat{f}_B(\mathbf{x})}{f(\mathbf{x})}-1\right|\leq\frac{\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left|\hat{f}_B(\mathbf{x})-f(\mathbf{x})\right|}{\inf_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}f(\mathbf{x})}=O_p\left(r_n^{-1}a_n^*\right).\tag{A17}$$

The result is established by recognizing that $\hat{m}_B(\mathbf{x})=\{\hat{g}_B(\mathbf{x})/f(\mathbf{x})\}/\left\{\hat{f}_B(\mathbf{x})/f(\mathbf{x})\right\}$.

For (5); The proof requires to approximate the bias terms of $\mathbf{S}_1(\mathbf{x})$, $\mathbf{S}_2(\mathbf{x})$ and $\mathbf{T}_1(\mathbf{x})$. The approximations are based on the following facts:

1. It follows from the proof of Lemma A1 that $E\left(\theta_{x_j}-x_j\right)=(1-2x_j)b_j+O\left(b_j^2\right)$ and $E\left(\theta_{x_j}-x_j\right)^2=x_j(1-x_j)b_j+O\left(b_j^2\right)$, where the $O\left(b_j^2\right)$ terms are uniform on $x_j\in(0,1)$.

2. The property of a beta random variable also implies that $\sup_{x_j\in(0,1)}E\left(\theta_{x_j}-x_j\right)^m=O\left(b_j^2\right)$ for $m\geq 3$.

Using the notations a_n and a_n^*, which are defined in the proofs of Theorems 1 and 4, respectively, we start from approximating $\mathbf{T}_1(\mathbf{x})=E\{\mathbf{T}_1(\mathbf{x})\}+[\mathbf{T}_1(\mathbf{x})-E\{\mathbf{T}_1(\mathbf{x})\}]$. Observe that

$$E\{\mathbf{T}_1(\mathbf{x})\}=E\{(\theta_{\mathbf{x}}-\mathbf{x})g(\theta_{\mathbf{x}})\}:=\mathbf{B}_{\mathbf{T}_1}(\mathbf{x})+O\left(\sum_{j=1}^p b_j\right)^2,$$

where $\mathbf{B}_{\mathbf{T}_1}(\mathbf{x})$ is a $(p\times 1)$ vector of the leading term with its jth element being equal to $[(1-2x_j)g(\mathbf{x})+x_j(1-x_j)\{\partial g(\mathbf{x})/\partial x_j\}]b_j$, and the second term is uniform on $[0,1]^p$. It can be also shown that

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\|\mathbf{T}_1(\mathbf{x})-E\{\mathbf{T}_1(\mathbf{x})\}\|=O_p\left\{\left(\sum_{j=1}^p b_j\right)a_n\right\},$$

and thus

$$\mathbf{T}_1(\mathbf{x})=\mathbf{B}_{\mathbf{T}_1}(\mathbf{x})+O_p\left\{\left(\sum_{j=1}^p b_j\right)a_n^*\right\}.$$

Then, by uniform boundedness of derivatives of g and $a_n^* = o(1)$,

$$\begin{aligned}\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}}\left\|\frac{\mathbf{T}_1(\mathbf{x})}{f(\mathbf{x})}\right\| &\leq \frac{\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}}\|\mathbf{T}_1(\mathbf{x})\|}{\inf_{\mathbf{x}\in\mathbb{S}_\mathbf{X}} f(\mathbf{x})} \\ &= O\left(r_n^{-1}\sum_{j=1}^p b_j\right) + O_p\left\{r_n^{-1}\left(\sum_{j=1}^p b_j\right)a_n^*\right\} \\ &= O_p\left(r_n^{-1}\sum_{j=1}^p b_j\right).\end{aligned}$$

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

$$\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}}\left\|\frac{\mathbf{S}_1(\mathbf{x})}{f(\mathbf{x})}\right\| \leq O_p\left(r_n^{-1}\sum_{j=1}^p b_j\right).$$

Moreover, $\mathbf{S}_2(\mathbf{x})$ admits the expansion

$$\mathbf{S}_2(\mathbf{x}) = \mathbf{B}_{\mathbf{S}_2}(\mathbf{x}) + O\left(\sum_{j=1}^p b_j\right)^2 + O_p\left\{\left(\sum_{j=1}^p b_j\right)^2 a_n^*\right\},$$

where $\mathbf{B}_{\mathbf{S}_2}(\mathbf{x})$ is a $(p\times p)$ diagonal matrix of the leading term in $E\{\mathbf{S}_2(\mathbf{x})\}$ with its jth diagonal element being equal to $x_j(1-x_j)f(\mathbf{x})b_j$, and the second and third terms are uniform on $\mathbb{S}_\mathbf{X}$. It follows that

$$\begin{aligned}\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}}\left\|\frac{\mathbf{S}_2(\mathbf{x})}{f(\mathbf{x})}\right\|^{-1} &\leq \frac{1}{\inf_{\mathbf{x}\in\mathbb{S}_\mathbf{X}}\left\{\|\mathbf{B}_{\mathbf{S}_2}(\mathbf{x})\|/f(\mathbf{x}) + O_p\left(\sum_{j=1}^p b_j\right)^2/f(\mathbf{x})\right\}} \\ &= \frac{1}{O\left(\sum_{j=1}^p b_j\eta_j\right) + O_p\left(\sum_{j=1}^p b_j\right)^2}.\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\left(\sum_{j=1}^p b_j\right)^2 = \left(\sum_{j=1}^p \sqrt{b_j\eta_j}\sqrt{\frac{b_j}{\eta_j}}\right)^2 \leq \left(\sum_{j=1}^p b_j\eta_j\right)\left(\sum_{j=1}^p \frac{b_j}{\eta_j}\right),$$

and thus

$$\begin{aligned}\sup_{\mathbf{x}\in\mathbb{S}_\mathbf{X}}\left\|\frac{\mathbf{S}_2(\mathbf{x})}{f(\mathbf{x})}\right\|^{-1} &\leq \frac{1}{O\left\{\left(\sum_{j=1}^p b_j\right)^2/\left(\sum_{j=1}^p (b_j/\eta_j)\right)\right\} + O_p\left(\sum_{j=1}^p b_j\right)^2} \\ &\leq O_p\left\{\frac{\sum_{j=1}^p (b_j/\eta_j)}{\left(\sum_{j=1}^p b_j\right)^2}\right\}.\end{aligned}$$

Therefore,

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{T}_1(\mathbf{x})}{f(\mathbf{x})}\right\| \leq \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{S}_1(\mathbf{x})}{f(\mathbf{x})}\right\| \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{S}_2(\mathbf{x})}{f(\mathbf{x})}\right\|^{-1} \sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{T}_1(\mathbf{x})}{f(\mathbf{x})}\right\|$$
$$\leq O_p\left(r_n^{-2}\sum_{j=1}^{p}\frac{b_j}{\eta_j}\right),$$

and

$$\sup_{\mathbf{x}\in\mathbb{S}_{\mathbf{X}}}\left\|\frac{\mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{S}_1(\mathbf{x})}{f(\mathbf{x})}\right\| \leq O_p\left(r_n^{-2}\sum_{j=1}^{p}\frac{b_j}{\eta_j}\right).$$

Using (A16) and (A17) finally yields

$$\begin{aligned}\tilde{m}_B(\mathbf{x}) &= \frac{\hat{g}_B(\mathbf{x})/f(\mathbf{x}) - \mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{T}_1(\mathbf{x})/f(\mathbf{x})}{\hat{f}_B(\mathbf{x})/f(\mathbf{x}) - \mathbf{S}_1(\mathbf{x})^\top\mathbf{S}_2(\mathbf{x})^{-1}\mathbf{S}_1(\mathbf{x})/f(\mathbf{x})} \\ &= \frac{m(\mathbf{x}) + O_p(r_n^{-1}a_n^*) + O_p\left\{r_n^{-2}\sum_{j=1}^{p}(b_j/\eta_j)\right\}}{1 + O_p(r_n^{-1}a_n^*) + O_p\left\{r_n^{-2}\sum_{j=1}^{p}(b_j/\eta_j)\right\}} \\ &= m(\mathbf{x}) + O_p\left\{r_n^{-2}\left(\sum_{j=1}^{p}\frac{b_j}{\eta_j} + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{p}b_j\eta_j}}}\right)\right\}\end{aligned}$$

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

A.7 Proof of Theorem 6

The arguments are the same as for Theorem 5, except that Theorem 2 is employed so that the convergence holds almost surely. ■

A.8 Proof of Theorem 7

For the sample average estimator

$$\hat{g}_W(\mathbf{x},\mathbf{z}) := \frac{1}{n}\sum_{i=1}^{n}Y_i\mathbb{W}(\mathbf{X}_i,\mathbf{Z}_i;\mathbf{x},\mathbf{z},\mathbf{b},\lambda),$$

consider that

$$\begin{aligned}&\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{S}}|\hat{g}_W(\mathbf{x},\mathbf{z}) - g(\mathbf{x},\mathbf{z})| \\ &\leq \sup_{(\mathbf{x},\mathbf{z})\in\mathbb{S}}|E\{\hat{g}_W(\mathbf{x},\mathbf{z})\} - g(\mathbf{x},\mathbf{z})| + \sup_{(\mathbf{x},\mathbf{z})\in\mathbb{S}}|\hat{g}_W(\mathbf{x},\mathbf{z}) - E\{\hat{g}_W(\mathbf{x},\mathbf{z})\}|.\end{aligned}$$

Incorporating the arguments as in Lemmata A1 and A2 of Li and Ouyang (2005) into the proof of Theorem 1 yields

$$\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{S}} \left| E\left\{\hat{g}_W\left(\mathbf{x},\mathbf{z}\right)\right\} - g\left(\mathbf{x},\mathbf{z}\right)\right| = O\left(\sum_{j=1}^{p} b_j + \sum_{k=1}^{q} \lambda_k\right) \text{ and}$$

$$\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{S}} \left| \hat{g}_W\left(\mathbf{x},\mathbf{z}\right) - E\left\{\hat{g}_W\left(\mathbf{x},\mathbf{z}\right)\right\}\right| = O_p\left(\sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{p} b_j \eta_j}}}\right),$$

respectively. Therefore,

$$\sup_{(\mathbf{x},\mathbf{z})\in\mathbb{S}} \left| \hat{g}_W\left(\mathbf{x},\mathbf{z}\right) - g\left(\mathbf{x},\mathbf{z}\right)\right| = O_p\left(\sum_{j=1}^{p} b_j + \sum_{k=1}^{q} \lambda_k + \sqrt{\frac{\ln n}{n\sqrt{\prod_{j=1}^{p} b_j \eta_j}}}\right).$$

Combining this with an argument in the proof for (4) of Theorem 5 leads to the desired result. ■

A.9 Proof of Theorem 8

The argument is the same as for Theorem 7, except that Theorem 2 is employed so that the convergence holds almost surely. ■

References

[1] Abadie, A., and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235 - 267.

[2] Aitchison, J., and C. G. G. Aitken (1976): "Multivariate Binary Discrimination by the Kernel Method," *Biometrika*, 63, 413 - 420.

[3] Bouezmarni, T., and J.-M. Rolin (2003): "Consistency of the Beta Kernel Density Function Estimator," *Canadian Journal of Statistics*, 31, 89 - 98.

[4] Bouezmarni, T., and J. V. K. Rombouts (2010): "Nonparametric Density Estimation for Multivariate Bounded Data," *Journal of Statistical Planning and Inference*, 140, 139 - 152.

[5] Chen, S. X. (1999): "Beta Kernel Estimators for Density Functions," *Computational Statistics & Data Analysis*, 31, 131 - 145.

[6] Chen, S. X. (2000): "Probability Density Function Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471 - 480.

[7] Chen, S. X. (2002): "Local Linear Smoothers Using Asymmetric Kernels," *Annals of the Institute of Statistical Mathematics*, 54, 312 - 323.

[8] Funke, B., and R. Kawka (2015): "Nonparametric Density Estimation for Multivariate Bounded Data Using Two Non-negative Multiplicative Bias Correction Methods," *Computational Statistics and Data Analysis*, 92, 148 - 162.

[9] Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525 - 574.

[10] Hagmann, M., O. Renault, and O. Scaillet (2005): "Estimation of Recovery Rate Densities: Non-parametric and Semi-parametric Approaches versus Industry Practice," in E. I. Altman, A. Resti, and A. Sironi (eds.), *Recovery Risk: The Next Challenge in Credit Risk Management.* London: Risk Books, 323 - 346.

[11] Hansen, B. E. (2008): "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory*, 24, 726 - 748.

[12] Harfouche, L., S. Adjabi, N. Zougab, and B. Funke (2018): "Multiplicative Bias Correction for Discrete Kernels," *Statistical Methods & Applications*, 27, 253 - 276.

[13] Henderson, D. J., and A.-C. Souto (2018): "An Introduction to Nonparametric Regression for Labor Economists," *Journal of Labor Research*, 39, 355 - 382.

[14] Hill, J., and A. Prokhorov (2016): "GEL Estimation for Heavy-Tailed GARCH Models with Robust Empirical Likelihood Inference," *Journal of Econometrics*, 190, 18 - 45.

[15] Hirukawa, M. (2018): *Asymmetric Kernel Smoothing: Theory and Applications in Economics and Finance*. Singapore: Springer.

[16] Hirukawa, M., I. Murtazashvili, and A. Prokhorov (2021): "Yet Another Look at the Omitted Variable Bias," Working Paper.

[17] Johnson, N. L. (1949): "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36, 149 - 176.

[18] Jones, M. C. (1993): "Simple Boundary Correction for Kernel Density Estimation," *Statistics and Computing*, 3, 135 - 146.

[19] Kanaya, S., and D. Bhattacharya (2017): "Uniform Convergence of Smoothed Distribution Functions with an Application to Delta Method for the Lorenz Curve," Cambridge Working Papers in Economics 1760.

[20] Koul, H. L., and W. Song (2013): "Large Sample Results for Varying Kernel Regression Estimates," *Journal of Nonparametric Statistics*, 25, 829 - 853.

[21] Kristensen, D. (2009): "Uniform Convergence Rates of Kernel Estimators with Heterogenous Dependent Data," *Econometric Theory*, 25, 1433 - 1445.

[22] Kristensen, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach," *Econometric Theory*, 26, 60 - 93.

[23] Lejeune, M., and P. Sarda (1992): "Smooth Estimators of Distribution and Density Functions," *Computational Statistics & Data Analysis*, 14, 457 - 471.

[24] Li, Q., and D. Ouyang (2005): "Uniform Convergence Rate of Kernel Estimation with Mixed Categorical and Continuous Data," *Economics Letters*, 86, 291 - 296.

[25] Müller, H.-G. (1991): "Smooth Optimum Kernel Estimators Near Endpoints," *Biometrika*, 78, 521 - 530.

[26] Newey, W. K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233 - 253.

[27] Racine, J. S., and Q. Li (2004): "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics*, 119, 99 - 130.

[28] Renault, O., and O. Scaillet (2004): "On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities," *Journal of Banking and Finance*, 28, 2915 - 2931.

[29] Rilstone, P. (1996): "Nonparametric Estimation of Models with Generated Regressors," *International Economic Review*, 37, 299 - 313.

[30] Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931 - 954.

[31] Sancetta, A., and S. Satchell (2004): "The Bernstein Copula and Its Applications to Modeling and Approximations of Multivariate Distributions," *Econometric Theory*, 20, 535 - 562.

[32] Shi, J., and W. Song (2016): "Asymptotic Results in Gamma Kernel Regression," *Communications in Statistics - Theory and Methods*, 45, 3489 - 3509.

[33] Stengos, T., and B. Yan (2001): "Double Kernel Nonparametric Estimation in Semiparametric Econometric Models," *Journal of Nonparametric Statistics*, 13, 883 - 906.

[34] Stone, C. J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040 - 1053.

[35] Stone, C. J. (1983): "Optimal Uniform Rate of Convergence for Nonparametric Estimators of a Density Function or Its Derivatives," in M. H. Rizvi, J. S. Rustagi, and D. Siegmund (eds.), *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday.* New York: Academic Press, 393 - 406.

[36] Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer-Verlag.

[37] Yatchew, A. (1997): "An Elementary Estimator of the Partial Linear Model," *Economics Letters*, 57, 135 - 143.

Table 1: Simulation Results for Density Estimation

	Oracle			CV		
Estimator	RISE	IAD	h or b	RISE	IAD	h or b
			Logit Normal ($n = 200$)			
KDE-E	0.1997	0.1528	0.1460	0.2287	0.1783	0.1209
	(0.0434)	(0.0368)	(0.0409)	(0.0598)	(0.0510)	(0.0495)
LLDE-E	0.3365	0.2505	0.0395	0.3587	0.2478	0.0466
	(0.0524)	(0.0429)	(0.0090)	(0.0622)	(0.0479)	(0.0111)
KDE-B	0.1217	0.0918	0.2326	0.1409	0.1085	0.1851
	(0.0353)	(0.0319)	(0.0420)	(0.0524)	(0.0445)	(0.0947)
			Logit Normal ($n = 400$)			
KDE-E	0.1817	0.1378	0.1142	0.2014	0.1561	0.0901
	(0.0345)	(0.0299)	(0.0316)	(0.0444)	(0.0382)	(0.0335)
LLDE-E	0.2741	0.2066	0.0326	0.2930	0.2011	0.0409
	(0.0384)	(0.0315)	(0.0056)	(0.0470)	(0.0348)	(0.0075)
KDE-B	0.1169	0.0861	0.2278	0.1219	0.0914	0.2223
	(0.0268)	(0.0247)	(0.0482)	(0.0314)	(0.0282)	(0.0635)
			Truncated Normal ($n = 200$)			
KDE-E	0.1464	0.1064	0.1631	0.1780	0.1309	0.1400
	(0.0341)	(0.0288)	(0.0456)	(0.0605)	(0.0511)	(0.0513)
LLDE-E	0.2799	0.2003	0.0467	0.3072	0.1995	0.0533
	(0.0504)	(0.0377)	(0.0131)	(0.0602)	(0.0419)	(0.0123)
KDE-B	0.0975	0.0822	0.0754	0.1408	0.1227	0.1843
	(0.0388)	(0.0356)	(0.0339)	(0.0370)	(0.0342)	(0.0418)
			Truncated Normal ($n = 400$)			
KDE-E	0.1261	0.0896	0.1269	0.1480	0.1079	0.1042
	(0.0262)	(0.0207)	(0.0400)	(0.0412)	(0.0356)	(0.0369)
LLDE-E	0.2180	0.1571	0.0377	0.2428	0.1511	0.0472
	(0.0350)	(0.0269)	(0.0083)	(0.0437)	(0.0283)	(0.0085)
KDE-B	0.0785	0.0655	0.0556	0.1134	0.0991	0.1386
	(0.0285)	(0.0258)	(0.0228)	(0.0281)	(0.0260)	(0.0275)

Note: "Oracle" and "CV" indicate that the tuning parameter is chosen as the direct minimizer of the RISE and via the cross-validation, respectively. For each case, simulation averages and standard deviations (in parentheses) of two performance measures ("RISE" and "IAD") and tuning parameter values ("h or b") are reported.

Table 2: Simulation Results for Regression Estimation

	Oracle			CV		
Estimator	RISE	IAD	h or b	RISE	IAD	h or b
			$m(x) = 2/3 - (x - 2/3)^2$ $(n = 200)$			
NW-E	0.0253	0.0194	0.1460	0.1382	0.0591	0.0871
	(0.0117)	(0.0091)	(0.0574)	(0.0918)	(0.0411)	(0.0410)
LL-E	0.0541	0.0253	0.2123	0.2552	0.0789	0.1720
	(0.0632)	(0.0296)	(0.0681)	(1.9524)	(0.3703)	(0.0453)
NW-B	0.0231	0.0168	0.0192	0.0258	0.0171	0.0078
	(0.0091)	(0.0066)	(0.0142)	(0.0108)	(0.0060)	(0.0027)
LL-B	0.0192	0.0133	0.0724	0.0279	0.0171	0.0482
	(0.0133)	(0.0074)	(0.0459)	(0.0187)	(0.0091)	(0.0163)
			$m(x) = 2/3 - (x - 2/3)^2$ $(n = 400)$			
NW-E	0.0208	0.0155	0.1284	0.1047	0.0403	0.0715
	(0.0083)	(0.0063)	(0.0409)	(0.0786)	(0.0301)	(0.0300)
LL-E	0.0253	0.0144	0.2130	0.1700	0.0457	0.1530
	(0.0282)	(0.0113)	(0.0463)	(1.0894)	(0.2158)	(0.0452)
NW-B	0.0188	0.0135	0.0160	0.0217	0.0138	0.0057
	(0.0074)	(0.0048)	(0.0101)	(0.0090)	(0.0045)	(0.0017)
LL-B	0.0146	0.0103	0.0642	0.0209	0.0127	0.0390
	(0.0088)	(0.0048)	(0.0361)	(0.0134)	(0.0060)	(0.0094)

Note: "Oracle" and "CV" indicate that the tuning parameter is chosen as the direct minimizer of the RISE and via the cross-validation, respectively. For each case, simulation averages and standard deviations (in parentheses) of two performance measures ("RISE" and "IAD") and tuning parameter values ("h or b") are reported.

Figure 1: **True Densities for Simulations**

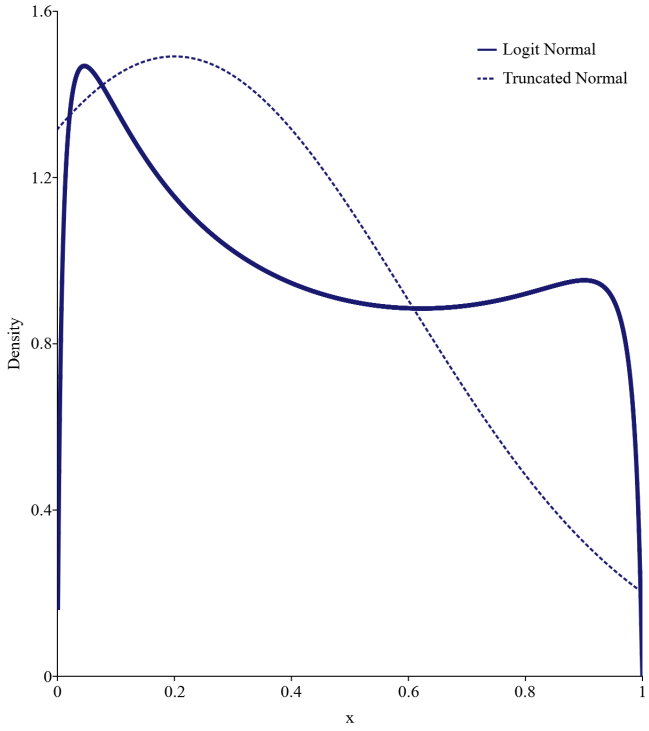

Figure 2: **Estimated Density and Regression Curves for the PSID Data**